HURDLE NEGATIVE BINOMIAL MODEL FOR MOTOR VEHICLE CRASH INJURIES IN NAMIBIA

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

NDILIMEKE SHIYUKA

[200717421]

APRIL 2018

MAIN SUPERVISOR: DR LILLIAN PAZVAKAWAMBWA

ABSTRACT

Injuries caused by Motor Vehicle crashes were ranked 10th among the leading causes of death and 9th among the leading cause of disability worldwide (World Health Organization, 2013). In developing countries 90% of disabilities are caused by road traffic crashes. In Namibia, The Motor Vehicle Accident Fund spends approximately N$ 22,6 million monthly towards medical expenses of people injured in crashes (Tjihenuna, 2015), causing a concern, to the victims, their families and affects the country's economy at large. Efforts have been put in place by stakeholders to reduce road crashes, injuries and fatalities, despite these efforts the trend of crashes keeps on increasing yearly. The objectives of the study were to explore generalised linear models to establish risk factors associated with road traffic injuries in Namibia and develop strategies to guide policy on reduction of road traffic injuries.

The study was based on a quantitative cross sectional research design for all road crash injuries recorded from 2011 -2016 of secondary data from the MVA Fund database (21869), with number of injured persons per crash as the dependent variable.

Using the *MASS* and *pscl* packages in R version 3.3.2, six Generalised linear were explored: Poisson, Negative binomial, Zero inflated Poisson, Zero- inflated Negative Binomial, the Hurdle Poisson and the Hurdle negative binomial. The Akaike Information Criterion (AIC) and the Vuong's test showed that the Hurdle Negative Binomial was the best.

The probability of road traffic injuries (RTI) was inferred by the following variables: the crash type with vehicle to vehicle (OR=0.5, $p<0.001$), Cause of crash with driver behaviour (OR=0.1, $p<0.001$), fatalities with deaths (OR= -1.8, $p<0.001$) and Time of crash with peak time (OR=0.3, $p<0.001$) being higher probability. The intensity of RTI was inferred by: Months of crashes with holiday month (RR=0.2, $p<0.001$) Day of crash with Weekend (RR =0.1, $p<0.001$) Region with Northern regions (RR= 0.3, $p<0.001$), Type of crash with vehicle by vehicle (RR =1.6, $p<0,001$), Crash cause with driver behaviour (RR= 0.2,$p<0.001$), fatalities with deaths (RR = 1.0, $p<0.001$) number of vehicles with single vehicle (RR= -0.2, $p<0.001$) having higher probabilities when compared to others.

This thesis suggests that emphasis should be placed on driver behaviour, since it is evident that a higher risk of injuries presents itself among crashes that occurred as a result of driver behaviour. Additionally, campaigns should focus more on school holiday months and weekends, due to the fact that a large number of crashes occur more on weekends and school holiday months.

Table of contents

ACKNOWLEDGEMENTS

I would like to thank my supervisor Dr. Lilian Pazvakawambwa, who has been instrumental and guiding me with my studies and if it was not for her, I would not have acquired knowledge of this magnitude, as I have now developed into a better student, with this knowledge I would now plough it back to my community so that they also acquire the same knowledge.

I want to thank the University of Namibia (UNAM) for granting me this opportunity to study for my masters of science in applied statistics and demography at UNAM.

DEDICATION

I would like to dedicate this paper to my parents Lameck Shiyuka and Nancy Sheyavali and the rest of my family who helped me a lot in finalizing this report within the limited time frame. I am really thankful to them and the rest of the community that has contributed to my educational dream.

DECLARATION

I, Ndilimeke Shiyuka, hereby declare that this study is my own work and is a true reflection of my research, and that this work, or any part thereof has not been submitted for a degree at any other institution.

No part of this thesis/dissertation may be reproduced, stored in any retrieval system, or transmitted in any form, or by means (e.g. electronic, mechanical, photocopying, recording or otherwise) without the prior permission of the author, or The University of Namibia in that behalf.

I, Ndilimeke Shiyuka, 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.

………………………………………… …………………. ….…………

Name of Student Signature Date

LIST OF ABBREVIATIONS ACRONYMS

ANOVA	Analysis of Variance
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
DIC	Deviance Information Criterion
CI	Confidence Interval
CZI	Crashes with Zero Injuries
IGE1	Injuries Greater or Equal to 1
IDRE	Institute for Digital Research Education
GLM	Generalized Linear Models
HP	Hurdle Poisson
HNB	Hurdle Negative Binomial
MASS	Morden Applied Statistics with S
MVA Fund	Motor Vehicle Accident Fund
MLE	Maximum Likelihood Estimator
Nampol	Namibian Police
NB	Negative Binomial
NRSC	National Road Safety Council
NSA	Namibia Statistics Agency
OR	Odds Ratio
OLS	Ordinary Least Square
pscl	Political Science Computational Laboratory
RR	Relative Ratio

RT	Road Traffic
SPSS	Statistical Package for Social Sciences
UN	United Nations
WHO	World Health Organization
ZIP	Zero- Inflated Poisson
ZINB	Zero- Inflated Negative Binomial

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Background

Namibia is located in the South Western region of Africa, bordered by South Africa, Angola, Botswana and Zambia. The total population was estimated around 2.3 Million in 2015 according to (Namibia Statistics agency, 2012). Roads Authority in Namibia registered 353 805 vehicles country wide in 2015, (Motor Vehicle Accident Fund, 2016) and 300 046 in 2013 (Motor Vehicle Accident Fund, 2014) an increase in ownership of vehicles has led to high traffic flows and congestions which is in direct proportion with the rapid increase of road crashes and Injuries in Namibia that has occurred in the past half a decade.

Globally, it is estimated that injuries as a result of road traffic crashes results in approximately 1.25 million deaths annually and another 20 to 50 million people sustain injuries (World Health Organisation, 2015). The World Health Organisation (WHO) further states that Road traffic injuries are a leading cause of death, and the main cause of death among those aged 15–29 years (WHO, 2015), making this age group an accident high risk. Additionally, the MVA fund has recorded that 31% of road crash fatalities were people aged 16- 30. Also, a significant number of deaths due to road crash injuries are road users that are least protected- the non-vehicle occupants that are still at risk of being involve in a road traffic crash such as: motorcyclists, cyclists, pedestrians and people on vehicles driven by animals. By end of February 2016, MVA Fund recorded 93 lives have been lost in car accidents, this includes the 15 people who died on Oshivelo-Omuthiya Road in Oshikoto region in a horrific crash, and nearly 11 years after 27 people lost their lives outside Grootfontein in Otjozondjupa region in a bus crash in 2005. These crashes culminate in the MVA Fund spending more than N$ 22.6 million a month and over N$1 million daily towards medical expenses, injury grants, loss of support, as well as loss of income and funeral claims (Uugwanga, 2016).

Current trends suggest that by 2030 road traffic deaths will become the fifth leading causes of deaths unless urgent action is taken (World Health Organization, 2013). In fact for Namibia we have already reached this mile stone according to a study done by (MVA Fund, 2015), Motor vehicle crashes were the 5th cause of death in Namibia. A study on fatal injuries of United States of America citizens abroad indicated that: The highest age adjusted Proportional Mortality Ratios where highest in Africa. Also, the death rate in Africa due to road crashes is at 24.1 per 100 000 inhabitants (Guse, 2007), in 2013 Namibia stood at 29.9 deaths per 100 000 population (Motor VehicleAccident Fund, 2013) .

Of all the systems that people have to deal with on a daily basis, road transport is the most complex and the most dangerous (WHO, 2004). Humans have become largely dependent on the road transport system to get to fulfil our daily chores; it is the only way to get to work, visit acquaintances, go to school, and going to the hospital one has to do it through road transportation. Making the world a global village has brought us nearer to each other but in very dangerous ways. It does not matter whether you are a pedestrian, driver, passenger, regardless of your social class, age or sex, we are all at risk of being injured or killed in a road traffic accident. The road transport system is one that is shared by different types of people and animals, with different attitudes, moods and daily frustrations.

Motor vehicle crash injuries cause emotional, physical and economic mischief, they consume huge financial resources to the community. Nantulya & Reich (2003) have reported that poor people in developing countries have the highest burden of injuries and fatalities due to traffic accidents. In a similar study Mohan, Tiwari, Meleckidzedeck, & Fredrick, (2006), surveyed deaths of heads of households due to road crashes in Bangladesh, reviled that 32% of deaths occurred to heads of households. Additionally, Over 70% of households reported that their household income, food consumption and food production had decreased after a death or injury of a head of household due to a road crash. Road traffic injuries or deaths post significant pressure on families of everyone killed or injured: Loss of income, medical bills and funeral arrangements negatively affect the finances. Many families are driven into poverty Mohan et. al. (2006), increasing the number of orphans and vulnerable children. As of 2008 road accidents are the second largest contributor to the death toll after HIV/AIDS in Botswana (Mphela 2011), Botswana a developing economy like Namibia face challenges managing the impact of road accidents on the lives of its citizens. The same source further adds that, substantial amounts of money go into compensation for the injured, lives lost and maintenance of damaged cars and third party assets, for Namibia the financial burden is similar. Moreover, low and middle income countries have a high proportion of low income people, have motor vehicles and therefore have more of their population exposed to risk of crashes (Nantulya & Reich, 2003), as pedestrians, or motor cyclists. This corresponds with the finding in 44 countries that the higher gross national incomes, the lower the pedestrian volume in the traffic system, and the lower the pedestrian fatalities (Paulozzi, Ryan, Espitia-Hardeman, & Xi, 2007)

The Motor Vehicle Accident Fund (2016) reported that 51% of injured persons where between the ages of 16 and 35 in 2015 and 56% injuries among the same group in 2014, the age group between16 and 35 represents the most productive age; students in secondary or tertiary institutions, employees entrepreneurs and people that are at the foundation of formation of families.

Road crash survivors and acquaintances suffer social, psychological and economic affects, as a result, loss of family bread winners or extra funds for to care for persons with disabilities which all contribute to the negative equation.

Namibia’s Motor Vehicle Accident Fund (MVA Fund) (my secondary source of data) is the statutory body mandated to design, promote and implement crash and injury prevention measures. The company also provides assistance and benefits to all people injured and dependants of those killed in motor vehicle crashes in accordance with the (Motor Vehicle Accident Fund Act, 2007). The MVA Fund is the only organization in Namibia that collects crash data as they occur, MVA

Fund (2015) alludes that the MVA Fund only records statistics from crashes that resulted in injuries and / or fatalities. Not considering crashes that only caused damage to properties.

The MVA Fund and the NRSC have collaborated to reduce road carnage through public education campaigns. Several campaigns are run mostly during the festive season and the Easter weekend, Andima (2014) justified that there is high traffic flow during this time resulting causing more road crashes, injuries and deaths. (Iipinge & Owusu-Afriyies, 2011) studied the assessment of the effectiveness of road safety programmes in Namibia and suggested that very few people are aware of the road safety education or information.

In 2010 the United Nations General Assembly resolution declared a “Decade of Action for Road Safety” launched in May 2011 and commenced immediately from 2011 to 2020, with the aim of reducing road traffic casualties. Organizations like, National Road Safety Council (NRSC) and the MVA Fund in Namibia collaborated with the UN to adopt methods to get them working. In 2009, member countries of the United Nations ratified the Decade of Action for Road Safety 2011-2020, with the objective of stabilizing and then reducing the number of road crashes by 2020. The UN Road Safety Collaboration developed a Global Plan for the Decade of Action (Do A) for Road Safety 2011-2020 to fulfil the goals of the UN resolution. Namibia signed up and aligned its road safety undertakings to the objectives of the DoA in May 2011. Namibia through the MVA Fund is dedicated to; road safety management, safer roads and mobility, safer vehicles, safer road users and post- crash response.

1.2 Statement of the problem

Motor Vehicle Crash injuries are a major cause of death in the country, but neglected by most stakeholders and public health sectors in Namibia. Large amounts of dollars a spent on post- crash activities and only a minimal amount on reducing the number of crashes. Over a five year period between 2011 and 2015, the number of injured persons has increased by 30%, it was reported that there was 3% increase between 2012, 2013, 18% increase between 2013 and 2014, and a further 6 % increase between 2014 and 2015 (MVA , 2016).

Road crash injuries keeps on changing from time to time according to MVA the number of injured person was 5652 in 2012, 5845 in 2013, 6314 in 2014 and 7333 in 2015, indicating 347 injuries per 100, 000 populations in 2015. On average 20 people are injured daily in Namibia due to road crashes, the injuries vary from slight to severe. The MVA Fund and the National Road Safety Council (NRSC) do not have the capacity or the financial means to do a comprehensive study on the preventative measures of crashes or injuries. There are very few inferential studies that have been done to address this issue that are specific to Namibia. The research aims to analyse the different variables that are associated with the increase of crashes to close this gap and alert organisations, public and private sectors on the seriousness of road fatalities to vender resources in the subject. Especially resources in research to further scrutinize and understand crash causes and causes of injuries. It is essential that analysis is done with the available data, to study the relations among variables to determine causes of crashes

1.3 Objective of the study

The main objectives of the study are to explore generalised linear models to establish risk factors associated with road traffic injuries in Namibia and develop strategies to guide policy on reduction of road traffic injuries.

- Determine risk factors associated to road traffic injuries in Namibia.
- Explore generalised linear models to examine associations and relationships between the number of injured persons and regions, crash types, crash causes, number of vehicles involved and time occurred.
- Come up with possible recommendations on ways to reduce injuries as a result of motor vehicle crashes.

It’s expected that this study will highlight the variables that directly increase road crash injuries and fatalities, occurring in Namibia and will open the eyes of development planners, policy makers and fund allocators and donors to steer the resources in the right direction to curb this public health issue.

1.4 Significance of the study

The study is significant to the MVA Fund, National Road Safety Counsel of Namibia, insurance companies, policy makers, funders and researchers. Injuries of any kind have an impact on the increase of persons with disabilities, increase in unemployment, and a high physical and financial dependency. Road crash injuries do not only contribute to a high dependency ratio, but also to a low productivity rate in the economic sector.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Namibia Statistics Agency (2012), population projections indicated that there are 2.3 million people living in Namibia. The Khomas region in which the capital city Windhoek is situated has the largest population of 16.19% people. Windhoek has the highest occurrences of crashes (38%) of whichis and 2420, the number of people that sustained injuries in 2015 MVAFund (2016), this could be attributed to the fact that 45% of the registered vehicles by Roads Authority in 2015 were registered in the capital MVAFund (2016).

In the year 2000 it was estimated that globally US$518 billion is spent towards road crash fatalities (World Health Organisation, 2013). In Namibia the MVA spends more than N$157 million approximately US$ 14,2 million a year for financial assistance in funeral grants, medical cost, Loss of income and Loss of support and claims[1]. According to the (World Health Organisation,2013), 91% of the world's fatalities on the road occurs in low-income and middle- income countries of the African and East Mediterranean regions, even though these countries have just approximately half of the world's vehicle population. Langarde (2007) stated that developing countries already account for more than 85% of all road traffic deaths in the world, the rise in the number of vehicles per inhabitant will result in an anticipated 80% increase in injury and mortality rates between 2000 and 2020. In addition, in Africa it has been estimated that 59 000 people lost their lives in road traffic crashes in 1990 and this figure will increase to 144 000 by 2020, a 144% increase. The number of vehicles per inhabitant in Africa is still less than one licensed vehicle per 100 inhabitants in low-income countries in Africa versus one licensed vehicle per 60 inhabitants in high income countries (Langarde, 2007). In South Africa there were already 17 licensed vehicles per 100 inhabitants in 2005, and no decline in road traffic deaths has been observed so far. Fleet growth leads to increased road insecurity in developing countries, this explains, for example, the reported 400% increase in road crash deaths in Nigeria between the 1960s and 1980. Langarde (2007), further states that available historical data from developed countries show that it is only when a development threshold is achieved that mortalities due to the road crashes starts to decrease.

2.1.1 Developing countries and road crash, injuries and fatalities

Over 80% of traffic Fatalities occur in developing and emerging countries, even though these countries account only about $\frac{1}{3}$ of total motor vehicle fleet (Garg & Hyder, 2006). Accident rates in developing countries are often 10-70 times higher than in developed countries. The escalating road safety problem in developing world thus represents serious health, social and economic disaster. Developing countries suffer staggering annual loss exceeding US$ 100 billion for accidents, which is nearly equivalent to double of all developing assistances (Garg & Hyder, 2006). In 1998, more than 85% of deaths and 90% of disability adjusted life years lost worldwide due to road traffic accidents occurred in developing countries. The vast majority of traffic accidents in developing countries comprises vulnerable road users (i.e., pedestrians, bicyclists and motorcyclists) and are most prevalent in urban areas. On the contrary, according to Afukaar, Antwi, & Ofosu-Amaah, (2010), the 1994-1998 police data in Ghana, road traffic crashes were a leading cause of death and injuries in Ghana. The majority of road traffic fatalities (61.2%) and injuries (52.3%) occurred on roads in rural areas. About 58% more people died on roads in the rural areas than in urban areas, and generally more severe crashes occurred on rural roads compared with urban areas.

According to WHO (2011) data, traffic deaths have risen from approximately 999,000 in 1998 to just over 1.1 million in 2002, an increase of around 10%. Low-income and middle-income countries account for majorities of these increases. Reductions in traffic fatalities in high-income countries are attributed largely to the implementation of a wide range of road safety measures, including seat-belt use, vehicle crash protection, traffic-calming interventions and traffic law enforcement. However, the reduction in the reported statistics for the road traffic injury does not necessarily mean an improvement in road safety for all countries. Furthermore, road traffic deaths are predicted to increase by 83% in low income and middle income countries and to decrease by 27% in high income countries (WHO, 2011).

(Naci, Chisholm, & Baker, 2009) found that in developing countries, roads are poorly built and are poorly maintained. As a result, the roads have become death traps. Vehicles are poorly maintained due to poverty, ignorance and corruption among enforcement agents. Similarly a study by Kenny (2007), in which road casualties in four developing countries were inter-reviewed, found clear evidence that poorer sectors of community were much more likely to be involved in road crashes than those who were better educated and with higher personal or household incomes.

Currently, road traffic crashes in developed countries show a declining trend due to ongoing investment in safety programs and counter measures. However, the literature on pedestrian crash risk in developing countries is at an early stage as the implementation of road safety interventions has only begun recently and rapid motorisation is taking place (Naci, Chisholm, & Baker, 2009). Moreover, the focus of road safety interventions has generally been restricted to improving the safety of motorists rather than pedestrians. A considerable effort is required to understand the unique features of pedestrian crash risk in developing countries. For instance, pedestrian fatalities globally were estimated to total more than 400,000 per year, out of which 55.3% and 39.2% occur in low- and middle income countries respectively per year (Naci, Chisholm, & Baker, 2009). Although pedestrian crash risk is a major concern in developing countries' road traffic crashes, the problem has not been sufficiently investigated. This neglect has stimulated a call by the World Health Organization (WHO) for a global focus on pedestrians, particularly in developing countries (WHO, 2013b). Most of the increases in traffic deaths have occurred in low and middle income countries, especially in Africa. African countries account for more than 38% of total road traffic deaths (WHO, 2013b). In the case of Ethiopia, pedestrians account for 55% of fatal crashes per annum (WHO 2009), although the comparative levels of exposure to risk are not known with certainty. In spite of these high numbers, policy makers in developing countries have failed to remediate the growing scale of pedestrian crash risk, which is exacerbated by the rapid motorisation in these countries. Given these trends it is worthwhile to scrutinise the problem of pedestrian crash risk in developing countries by providing a comprehensive synthesis of the factors which expose pedestrians to crash risk.

2.2 Conceptual framework

The actual causes of road crashes in Namibia remain un-established to this day, the knowledge of contributing factors and the prediction of accidents risks are vital to develop the next generations' prevention technologies if we are to save our children from this social vice.

Movement of people and goods on the road is necessary for social, economic and political reasons, this locomotion results in a risk of road crash injuries or even fatalities. WHO (2014) classifies the main risk factor contributing to road traffic crashes as a function of four elements: first exposure- the amount of movement or travel within the system by different users or a given population density, secondly the underlying probability of a crash, given a particular exposure, thirdly the probability of injury given a crash and finally the outcome of the injury.

WHO (2015) Summarized the risk factors for road traffic injuries into 5 factors namely: Factors influencing exposure to risk, risk factors influencing crash involvement, risk factors influence crash severity, Post-crash injuries and lastly preventative measure. Detailed explanations of the five factors are stipulated in subtitles 2.2.1 to 2.2.5 respectively:

2.2.1 Factors influencing exposure to risk

Economic factors, including social deprivation, demographic factors, land use planning practices which influence the length of a trip or travel mode choice, mixture of high speed motorized traffic with vulnerable road users, insufficient attention to integration of road function with decisions about speed limits, road layout and design.

Kim, Ramjan and Mark (2016), predicted vehicle crashes using drivers' characteristics such as driver's age, gender, district, vehicle choice and driving experience versus past traffic violation. In a similar study Awadzi, Classen, Hall, Duncan & Garvan (2008) suggested that time of crash, gender, non-seat belt use, roll over crash, vehicle body type, number of lanes and non-airbag deployments and age were predictors of injuries among younger and older adults in fatal motor vehicle crashes. Both of these studies show that the mentioned variables are good predictors of vehicle injuries.

Kim et al (2016), concluded that young females, with fewer years of experience and middle sized vehicles were significantly associated with higher likelihood of vehicle crashes. Contrary to that, (MVA, 2016) reported that 64% of injured persons in 2014 were males, specifically, in 2015 young males between the ages of 16 and 35 years accounted for 47% of the total males while only 32% of the total number of females between the age of 16 and 35 were killed as a result of road crash injuries. Leaving young males more at the risk of dying or being injured in a road crash comparing to young females, female drivers had a lower accident risk than males, while older (aged over 50) and younger drivers (18-21 years old) had an increased risk compared to middle-aged driver (Mohammad et. al,.2012). Shope & Bingham (2008) and Mohammad et. al. (2012) argues that, less driving experience, personality factors, lack of traffic safety awareness, serious deficiencies in road infrastructure and limited funding, overloaded and oversized trucks and using failed inspection vehicles in rural areas have been identified as major factors causing fatal accidents in China.

.

2.2.2 Risk factors influencing crash involvement

Inappropriate or excessive speed, intoxication of alcohol, use of certain medical or recreational drugs, fatigue, being a young person, being a vulnerable road user in urban and residential areas, vehicle factors- such as braking, handling and maintenance of extra appliances, defect in road design, layout and maintenance which can also lead to unsafe road user behaviour, inadequate visibility due to environmental factors (making it hard to detect vehicle and other road user) are risk factors influencing crash involvement (WHO, 2012). Mohammad et al (2012) demonstrated a study in Australia that revealed that after adjusting for confounding factors, that vehicles made before 1984 had a 2.88 time greater chance of being involved in an injury compared to those made after 1994. Later versions of vehicles have been improved with extra safety features that the older makes may not have had. Also, the crash risk increases for every year that the vehicle's age increases. Driving long distance alone, when compared to driving with passengers, increases the risk of an accident. Furthermore, Mohammadi et al (2012) and the (Road traffic training manual, 2007) both agree that a population-based study in America showed that the strongest fatal crash initiation predictor was alcohol. Impairments by alcohol is an important factor influencing both the risk of a road crash as well as the severity of the injuries that result from the crash.

Other risk factors contributing to fatal crashes were; driving without a valid licence, not wearing a seat belt, experiencing a crash in the previous year, lack of driving experience and the lack of traffic safety awareness. A study on temporal patterns of animal-related traffic accidents in Eastern Cape, South Africa was conducted by (Eloff &Van Niekerk, 2008). The study analysed road accident data for five years period between two roads and revealed that there is significant difference between the monthly distributions of animal related accidents and non- animal related accidents. The same study also, unveiled that the months when most accidents occurred coincided with major school holidays because of more than normal traffic volumes.

According to Manges (2010), the other risk factor is the use of electronic devices while driving. In Namibia driving while directly on the phone is against the law. A traffic fine for using a phone in your hand while driving is N$ 2000 (approximately USD 147) or six month imprisonment Although, provision has been made so that the driver may still communicate while driving by making use of headsets or the Bluetooth device installed in the vehicle. Caird, Willness, Steel & Scialfa (2008) disagrees with the decision of communicating with a device while driving, even if it is through a Bluetooth device or head set. Card et. al. presents their study saying that, handheld and hands-free phones produced similar Reactionary Time (RT) decrements. Overall, a mean increase in RT of 0.25 seconds was found to all types of phone-related tasks, drivers using either phone type do not appreciably compensate by giving greater headway or reducing speed, that reactionary time is the same when driving regardless of whether the phone is handheld or otherwise.

(Zheng, Qu, Zhang, Ge, 2016) revealed that there is a significant negative correlation between intentional bias scores and several indicators of dangerous driving. Similarly, drivers with few dangerous driving behaviours showed greater intentional bias safety, attention is divided and the driver may not focus fully on the driving. It is prohibition of using a two way communication device that will decrease the risk factor, because whether the device is hand held or not, the driver's attention is divided, causing a risk.

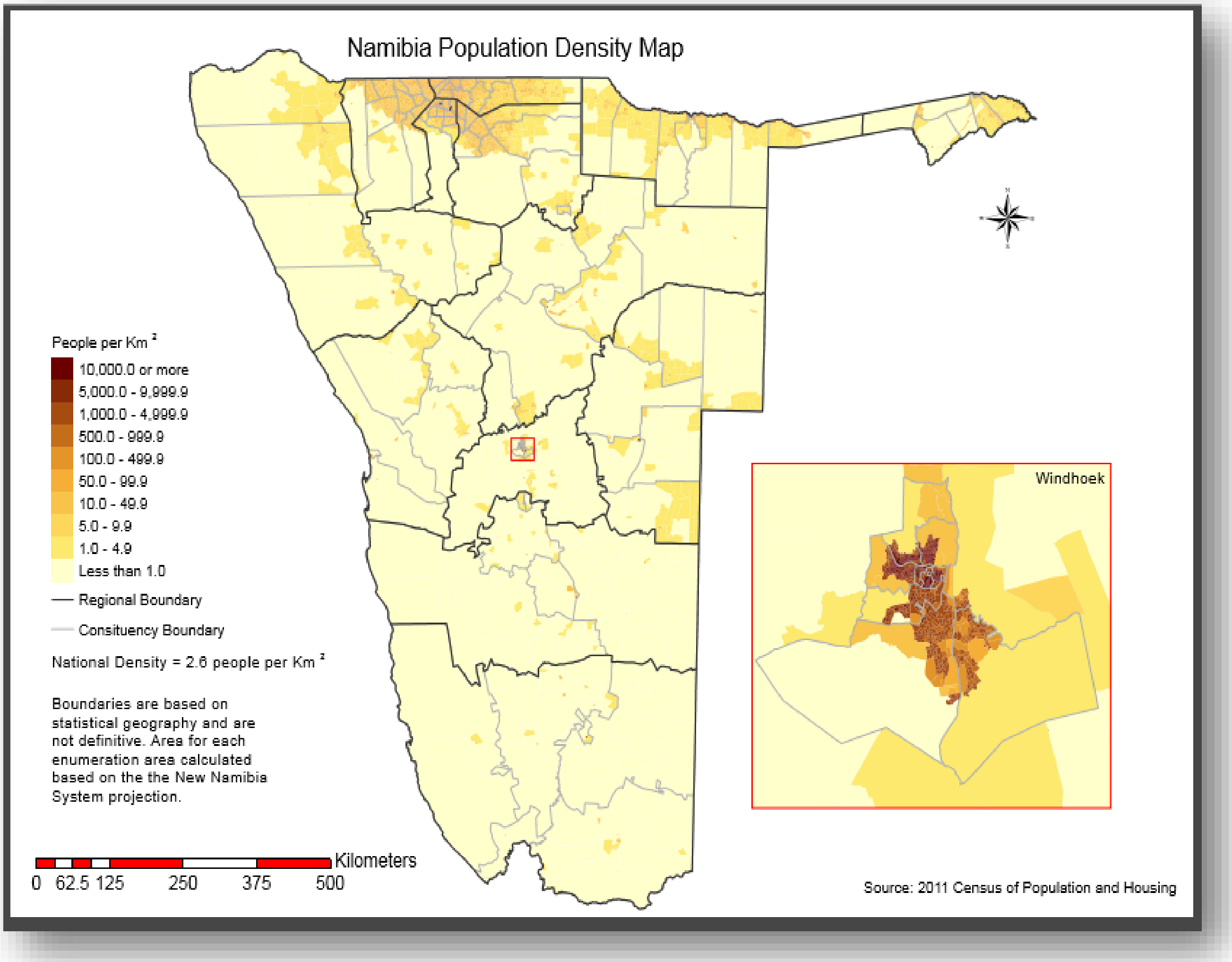
2.2.3 Risk factors influence crash severity

Human tolerance factors, inappropriate or excessive speed, seat-belt and child restraints not used, helmet not worn by users of two-wheeled vehicles, roadside objects not crash protective insufficient vehicle crash protection for occupants and for those hit by vehicles, presence of alcohol and drugs (Mohammad et al, 2012). A rapid increase in the number of motor vehicles, mixed traffic flows, inadequate infrastructural safety features- road designed were fit for a certain number of vehicles, when there is an increase of vehicles and the road infrastructure does not fit the current number of vehicles. On the other hand, Kim, Ulfarsson, Shankar, Mannering (2010) and Reynolds, Harris, Teschke, Cripton & Winters (2009) ascertain that cyclist and pedestrian safety is limited by the incomplete range of facilities, studies suggests that infrastructure influences injuries and crash risk. They found that multi-lanes can significantly increase risk to bicyclists and pedestrians unless a separated track is included in the design. Major roads are hazardous than minor roads, and presence of Pedestrian and bicycle facilities (bicycle only lanes, pedestrians only lane and sufficient bicycle and pedestrian crossings) was associated with the lowest risks. Pedestrian crash risk at night is higher than in the daytime due to the lower conspicuity of pedestrians (Schneider, Grembek, & Braughton 2012), this is exacerbated by the tendency for pedestrians to judge themselves as being more visible than they actually are at night. (Balk, Tyrrell, Brook & Carpenter, 2008), wearing darker clothing without wearing retroreflective clothing makes it worse to be detectable. Most pedestrians in developed countries can afford to buy retroreflective clothing which is both available and has been demonstrated to enhance visibility at night (Wood, et al. 2012).One study shows that drivers' ability in recognising pedestrians at night is degraded (Wood, et al. 2012) such that pedestrian fatalities may rise seven times higher at night than daytime (Sullivan & Flannagan, 2007). In addition, most locations in developed countries with high pedestrian traffic have sufficient street lighting to facilitate the visibility of pedestrians at night and thereby reduce road crashes (Beyer & Ker 2009), whereas the same does not apply as widely in developing countries. Researchers in Ghana, for example, have found that the night-time pedestrian crash rate is higher than the daytime rate since many built-up areas have not been provided with sufficient street lighting (Damsere-Derry, Ebel, Mock, Afukaar, & Donkor).

2.2.4 Post- crash injuries risk factors that influence crash severity

These are post-crash injuries that may have not been severe at the time the crash occurred, but become worse due to a lack of adequate resources. Figure 2.1 shows the population density of Namibia, the brown colour indicting the density by constituency. It is evident that the country is large in size with a very small population per area. The central has the highest number of people per km^2, followed by the Northern regions. Due to the fact that Namibia is a vast empty land with very many physical remote areas, many roads that connect towns are roads with little or no telecommunication network coverage and are rarely travelled. So that, if a crash occurs on one of these roads it can go undetected for hours or even days, meaning, even if an injury maybe minor the severity will increase because of the time it took for the victim to get medical care. The Economist (2012) narrates a story of two couples who had a crash on one of the remote areas, and could only receive medical assistance days later, when help came one of them had already died, there was no health facility or emergency services nearby for the injured person to get medical attention on time. This is a norm on Namibian roads, other familiar stories is when a crash result is a fire, or when people are trapped in a vehicle and the jaw of life is in another. A lack of appropriate pre-hospital care in emergency rooms, inadequate resources in the hospital and also on rescue services may increase the Post-crash injury risk factor that influence injury severity.

Figure 2.1 Namibia’s population density



Namibia Population Density Map

National Density = 2.6 people per Km ^{2}

Boundaries are based on statistical geography and are not definitive. Area for each enumeration area calculated based on the the New Namibia System projection.

Source: 2011 Census of Population and Housing

2.2.5 Preventative Measures

In Iran, Mohammad et al (2012), comprehensive road safety programs were initiated in 2005 to enforce three interventions: seatbelts, motorcycle helmet laws and general traffic law enforcement, like speed cameras, traffic patrol and mass media education campaigns on national roads. Similarly, the point system where negative points are collected for any traffic law violated by a driver, when the number of points exceed a set maximum, the driving licence is seized for a defined period of time. In a study by Novoa et. al. (2010)

Showed that the introduction of Penalty Points system in Spain reduced both the number of drivers involved in injury collisions and the number of people injured in collisions. Novoa et. al. further states that a large reduction was observed among the numbers of serious or fatal injuries.

2.3 Generalised linear models for Count Data

Count data are distributed as non-negative integers, are intrinsically heteroskedastic, right skewed and have a variance that increases with the mean (Chimba, Sando, Kwigizile & Kutela, 2014). Count data have possible zeros, for example, crashes where there were no injuries recorded for any victim. In theory, the potential outcome of a count variable range between zero and infinity, the observed numbers from real study sample are always finite because of the limited sample size. The traditional methods for analysing these data are test, analysis of Variance (ANOVA), or regression by ordinary least square (OLS) (Nussbaum, Elsadat & Khago, 2007).

In a study by Lord and Mannering (2010), some methodological challenges encountered in the current crash frequency studies were summarized, including temporal and spatial correlation, time-varying explanatory variables, and omitted-variables bias. They continue to say, most of the existing crash frequency modelling methods use aggregated data in extended time scales (e.g., yearly or monthly), instead of fine time scales (e.g., hourly, daily) containing detailed time-varying information. That the extended scales and aggregated variables lead to some limitations. For example, some important explanatory variables in crash frequency models sometimes change quickly over time, such as weather, road surface conditions and traffic flow (Lord and Mannering (2010). Furthermore, when data on smaller time scales (e.g., hours, days, rather than years) is adopted in crash frequency models, two major associated methodological challenges arise: (1) time-specific heterogeneity (e.g., micro-climates such as a snow storm, or temporal effects such as the weekend, which may influence the crash risk of the road segments in the same area at the same time), and (2) the preponderant portion of non-crash observations (more zero observations under fine time scales). Lord and Mannering (2010) continue to say: specifically, when the scales get smaller, the same road segment and/or the same time period may have multiple observations, which may be sharing unobserved effects because of the correlation over time and/or space. Also, short-term data has excess zeros, because of the fine time scales; there may be tremendous observations with no crash which indicate a big portion of zero counts. Appropriate models have to be taken into account for proper modelling of excess zeros.

2.3.1 Poisson Regression Model

Count data are often highly skewed, making it more appropriate to use the Poisson methods to analyse the variables specifically measuring degree of fit to a Poisson distribution and using Poisson regression. The Poisson is determined by one parameter, µ, which is both the mean and variance of the distribution. The Poisson distribution model is expressed as followed:

$$\mathrm{P}(y_i) = \frac{(e^{-\mu}{\mu_i}^{y_i})}{y_{i!}} \;\; y_i = 0{,}1{,}2 \text{ ... and } \mu > 0$$

The mean Parameter

$$\mathrm{E}\left[\frac{y_i}{X_i}\right] = \mu = e^{(X_i\beta)}, \text{ Variance} = \mu,$$

Where $y_i = a$ is a depended variable

X_i = Vector of explanatory variable

β = Coefficient of the corresponding factor

Chimba et al. (2014) compared model fit using the Poisson and multivariate linear regression models, concluding that the Poisson out performs the multivariate linear regression model due largely to its more appropriate statistical properties for describing non-negative discrete data like crashes. Nevertheless, they noted that if the crash data reveal significant over-dispersion around the estimated mean, the Poisson model becomes inadequate and more general distributional models, such as the negative binomial (NB), are needed.

2.3.2 Negative Binomial Regression Model

Although the Poisson regression model is commonly used for modelling count variables, it is also very restrictive. Under this model, the mean and variance are the same. However, in practice it often happens that the variance exceeds the mean, a phenomenon known as over-dispersion (Tang, He & Tu, 2012). The negative binomial also known as the Poisson-gamma model is an extension of the Poisson model that can overcome over-dispersion in the data. An alternative approach to modelling over-dispersion count data is to start from a Poisson regression model and add a multiplicative random effect to represent unobserved heterogeneity, and becomes a binomial regression model (Rodriguez, 2013).

Lord and Mannering (2010), states that the negative binomial assumes that the Poisson parameter follows a gamma probability distribution.

The negative binomial is derived by rewriting the Poisson parameter for each observation i as $\mu_i = e^{\beta X_i + \varepsilon_i}$ where e^{ε_i} is a gamma- distribution error term with mean 1 and varianceα. The addition of this term allows the variance to differ from the mean as:

$$Var[y_i] = E[y_i][1 + \alpha E[y_i]] = E[y_i] + \alpha E[y_i]^2$$

The Poisson regression model is a limiting model of the negative binomial regression model as α approaches zero, which means that selection between these two models is dependent upon the value of α (Lord and Mannering, 2010). The parameter α is the over-dispersion parameter. But, the Negative Binomial has limitations too, most notably its inability to handle under- dispersed data, dispersed- parameter estimation problems when the data are characterised by the low sample mean values and small sample size (Lord& Mannering 2010). The negative binomial probability density function is:

$$f(y_i; \mu; \theta) = \frac{\Gamma(y_i+\theta)}{\Gamma(\theta)\cdot y_i!} \cdot \frac{\mu^{y_i} \cdot \theta^{\theta}}{(\mu+\theta)^{yi+\theta}} ,$$

With mean μ or the expected value of the distribution, θ is the over- dispersion parameter; Γ (·) is the gamma function. The maximum likelihood function for the negative binomial model.

$$L(\beta|y, X) = \prod_{i=1}^{N} P(y_i|\mu_i) = \prod_{i=1}^{N} \frac{\Gamma(y_i + \theta^{-1})}{y!\,\Gamma(\theta^{-1})} \left(\frac{\theta^{-1}}{\theta^{-1} + \mu_i}\right)^{\theta-1} \left(\frac{\mu_i}{\theta^{-1} + \mu_i}\right)^{y_i}$$

Nussbaum et al (2007) further quotes that: The binomial formula calculates the probability of a certain number of successful trials for example (K), given the number of trials (n) and the probability to success (π), the formula is:

$$\frac{n!}{K!(n-K)!}e^{\pi}(1-\pi)^{k}$$

The negative binomial (NB) model can be expressed as:

$$p(y) = \left(\frac{\Gamma(y + \propto^{-1})}{\left(\Gamma(\propto^{-1})\Gamma(y+1)\right)}\right)\left(\frac{1}{(1+\propto\mu)}\right)^{\frac{1}{\propto}}\left(\frac{\propto\mu}{(1+\propto\mu)}\right)^{y}$$

Where the mean $\mu = E(y) = Var(\, e^{x\beta})$

The corresponding variance is $Var(y) = \mu + \propto \mu^{2}$. Similar extension to the negative binomial model are considered, including the zero-inflated Negative Binomial with constant and mean dependent split parameters, and the mean dependent over- dispersion factor Chimba et al.(2014).

Count data with extra zeros are common in many medical data and to data applications of data analysis and situational analysis. Chimba et al. (2014) challenges this assumption by arguing that the occurrence of crashes is in fact a binomial process, which can be approximated by a Poisson process when the number of trials (e.g., traffic exposure) is large with a small likelihood (risk) of crashes.

2.3.3 Zero-inflated Poisson Model

Wang, Yau, Lee, McLachlam (2007) agrees with Lord and Mannering,(2010) that Zero-inflated Poisson (ZIP) regression model is useful to analyse count data. He continues to say that for hierarchical or correlated count data where observations are either cluster or represent repeated outcomes from individual subjects, a class of ZIP mixed regression models may be appropriate. However, the ZIP parameter estimate can be severely biased if the non-zero counts are over dispersed in relation to the Poisson distribution.

Chimba et al.(2014) acknowledges that zero-inflated models are mainly used for modelling excessive zero count data. Furthermore, defines zero counts to be the situations where the likelihood of an event occurring is extremely rare. For example, not every crash will be recorded with an injury, so the number of injuries variable will have some zeros. Chimba et al. (2014) support Zero-Inflated Models by stating that, zero inflated models have surfaced as a plausible approach for use in crash analysis. Moreover, the use of zero-inflated models have been justified from the fact that Poisson and Negative Binomial (NB) models with or without their extensions as well as several variations seem to model non-negative discrete response variables, with over-dispersion and the underlying assumption that the occurrence of crashes observed at a given time and space scale follows a Poisson process.

The zero-inflated Poisson model, assumes that the events $y_i = (y_1 \ldots y_N)$ are independent and the model is

$$Pr[y_i = 0] = \phi_i + (1 - \phi_i)e^{-\mu_i}$$

$$Pr[y_i = r] = (1 - \phi)\left(\frac{(e^{-\mu_i}\,{\mu_i}^r)}{r!}\right), r = 1, 2 \ldots n$$

Where ϕ = proportions of zeros

Maximum likelihood estimates are used to estimate the parameters of the Zero-inflated Poisson model and confidence intervals are constructed by likelihood ratio tests. Count data may include both a Poisson and negative binomial models but data may still show more zeros than would be expected under either model. Zero-inflated Poisson (ZIP) regression is a model for count data with excess zeros. It assumes that with probability p the only possible observation is 0, and with probability 1 – p, a Poisson (μ) random variable is observed. Lord and Mannering (2010) argues that Zero inflated models have been developed to handle data characterized by a significant amount of zeros or more zeros than expected in a traditional Poisson or negative binomial. Zero-inflated models operate on the principle that the excess zero density that cannot be accommodated by the ordinary count structure is accounted for. Sometimes p and μ are unrelated; other times p is a simple function of μ such as $p = \frac{l}{1+\mu^T}$ for an unknown constant T (Lambert, 2012). In addition, the maximum likelihood estimates (MLE's) are approximately normal in large samples, and confidence intervals can be constructed by inverting likelihood ratio tests or using the approximate normality of the MLE's. Lambert (2012), goes on to say that the confidence intervals based on likelihood ratio tests are better. The Zero inflated model takes this form (Ridout, Demetrio, Hinde 1998):

$$P(Y = y) = \begin{cases} p + (1-p)e^{-T}, & y = 0, \\ (1-p)e^{-T}\dfrac{T^y}{y!}, & y > 0 \end{cases}$$

T is an unknown constant, we shall assume that $0 \leq p < 1$ in this study.

The Zero-inflated Poisson distribution takes the form:

$$E(Y) = (1-p)T = \mu,$$

$$Var(Y) = \mu + (\frac{p}{1-p})\mu^2,$$

2.3.4 Zero-inflated Negative Binomial Regression Model

Hu, Pavlicova & Nunes (2011), articulates that zero-inflated models assumes that the zero observations have two different origins: “structural” and “sampling”. The sampling zeros are due to the usual Poisson (or negative binomial) distribution, which assumes that those zero observations happened by chance. Zero-inflated models assume that some zeros are observed due to some specific structure in the data (Hu et al., 2011). For example, if the count is for people that have been injured in road crash accidents is the outcome, some people may be involved in an accident that was so slight that there couldn’t be an injury of an sort: these are what Hu calls the structural zeros because of the slightness of the crash people would not be in a situation to sustain any injuries. Other victims have been in a sever crash were even some people might have died, but some did not sustain any injuries, but they were at great risk of being severely injured or even dead. This Hu et al., (2011) assumes that it is a Poisson or negative binomial distribution that includes both zero (the “sampling zeros”) and non-zero counts.

2.3.5 Hurdle Poisson Model

On the other hand, Hosseinpour, Prasetijo, Yahaya & Ghadiri (2012) in their study of pedestrian- vehicle crashes observed that, the presence of over- dispersion in the pedestrian crashes is due to excess zero rather than variability in the crash data. To handle the issue of over- dispersion, the Hurdle Poisson models was found to be the best model among the considered models in terms of comparative measures. In contrast to the structural zero assumption presuming an inherently safe condition with no crashes, sampling zero assumption implies that all segments have crash potential and the zero state does not remain permanently on any road segment Hosseinpour, et al (2012). They further state that existing crash studies using hurdle models were primarily developed based on cross-sectional data (as opposed to panel data) and did not consider random effects.

The Other approach to excess zeros is to use a logit model to distinguish counts of zeros from large counts, effectively collapsing the count distribution into two categories, and then use the a truncated Poisson model, meaning a Poisson distribution where zero has been excluded, for the positive counts (Rodriguez, 2013). The term “hurdle” is evocative of a threshold that must be exceeded before events occur, with a process determining the number of events (Rodriguez 2013). The Hurdle model combines a count data model $f_{count}(y; x; \beta)$ (left truncated at y=1) and a zero hurdle model $f_{zero}(y; z; \gamma)$(right-censored at y=1) (Zeileis, Kleiber, Jackman, 2008):

$$f_{hurdle}(y, x, z, \beta, \gamma) = \begin{cases} f_{zero}(0; z; \gamma) & if\ y = 0, \\ 1 - f_{zero}(0; z; \gamma)\ \cdot \frac{f_{count}(y;x;\beta)}{f_{count}(0;x;\beta)} & if\ y > 0 \end{cases}$$

The model parameter $\beta, \gamma,$ and potentially one or two additional dispersion parameters θ are estimated by maximum likelihood, where the specification of the likelihood has the advantage that the count and the hurdle component can be maximized separately. According to Zeileis, Kleiber, Jackman (2008) the Hurdle count data can be fitted with the hurdle () function from the pscl package.

Positive integer based on truncated count data ($Y_i > 0$ are called the Poisson hurdle model when they are modelled using the Poisson distribution. Supposed y_i number of injured persons per crash and we consider the probability of $y_i = 0$ is $1 - p(x)$

2.3.6 Hurdle Negative Binomial Model

The Hurdle Negative binomial is considered for count models that are over dispersed. Hurdle Negative binomial models is mixed by a binary outcome of the count being below or above the hurdle (the selection variable), with a truncated model for outcomes above the hurdle (Saffari, Adnan & Greene, 2012). The hurdle negative binomial can handle excess zeros and the analysis of under- dispersion and over- dispersion. Similarly, Hu et al (2011) says that a hurdle model assumes that all zero data are from one “structural” source. The positive (non-zero) data have “sampling” origin, following either truncated Poisson or truncated negative- binomial distribution. For instance if we are doing a study on alcoholics in which the secondary outcome is the number of beers drank last month. One can assume that only non-drinkers will drink zero beers during the last month, people that drink and alcoholics will have a positive non-zero number of beers drank. This means that zeros can only come from one structural source of non-drinkers.

$$\text{Let } p_r(Y_i = y_i) = \begin{cases} \gamma_{0,} & y_i = 0, \\ (1-\gamma_0)\frac{\Gamma(y_i+\alpha^{-1})}{\Gamma(y_i+1)\Gamma(\alpha^{-1})}\frac{(1+\alpha\mu_i)^{-\alpha^{-1}-y_i}\alpha^{y_i}\mu_i^{y_i}}{1-(1+\alpha\mu_i)^{-\alpha^{-1}}}, & y_i > 0, \end{cases}$$

Where $\alpha(\geq 0)$ is a dispersion parameter that is assumed not to depend on covariates, and we suppose $0 < \gamma_0 < 1$

2.4 Testing Models

Before selecting the best model for the study, specific models will be performed to see the model that best fit the data looking at the level of zeros. The models are tested to choose the best that will maximize the results. This section will discuss few models testing and selection criteria.

2.4.1 Tests for Dispersion

Over-dispersion is a presence of greater variability in a data set than would be expected based on a given statistical model. If this parameter is θ, and test if θ is significantly different from zero. The null hypothesis tests if θ equals 0 against it's alternative that θ is not equal to zero.

H_0: $\theta = 0$ or H_1 : $\theta \neq 0$

The outcome of the results will be interpreted as: when $\theta = 0$ there is no dispersion in the data; when $\theta > 0$, the data is over-dispersion and when $\theta < 0$ the data is under-dispersed which is not very common in data.

Chimba et al., (2014), lists the following as the source of Over- dispersion:

- When some important independent variables are omitted from the model;
- When the data contains a lot of outliers resulting either from unreliable data collection or mistake and errors during data recording
- When the model fails to include a sufficient number of interaction terms
- When the variable itself is not appropriate and it needs transformation;
- If the distribution assumed is quite different from the real distribution which relates the data e.g., using linear model instead of a quadratic

Hu et al (2011) Further lists the sources of over- dispersion as:

- some covariates may be omitted and/or may not have a uniform effect on all subjects so that population heterogeneity has not been accounted for
- an excess number of zero events occurred compared to the Poisson distribution
- For the excessive zeros situation, it could be assumed that a sample is collected from two different sub-populations; one population always produces zero, or no event, while the other behaves like a Poisson distribution.

A considerable amount of statistical methodology has been developed to deal with over-dispersed data arising from excessive zero- count data (Chimba et al, 2014). Applications for the zero- inflated models can be found in several papers. Yet, continues Chimba, using these alternatives to the Poisson model seems to be a relatively a new approach among many researchers in applications. This is partly because once a statistical method becomes widely used in published literature, alterations to its usage are slow.

Zaninotto & Falaschetti (2010), did a study on comparison of methods for modelling a count outcome with excess zeros: using Application to Activities of Daily Living. They used the Likelihood Ratio test of over- dispersion, the Vuong’s test and geographical methods. In a similar study Khan, Ullah & Nitz (2011) also agree with Zaninotto & Falaschetti (2010), that the Vuong’s test can be used to test for over- dispersion, also, the Monte Carlo Simulation goodness- of- fit.

Furthermore, Chimba et al. (2014) says: the appropriateness of using the zero inflated model is that rather than the traditional Poisson or Negative Binomial model, the ZIP can be tested. The common known test statistic is through Vuong's value, estimated as shown below, Chimba et al.(2014):

$$m_i = \ln\left\{\frac{\left(\left(f_1\left(\frac{y_i}{X_i}\right)\right)\right)}{\left(\left(f_2\left(\frac{y_i}{X_i}\right)\right)\right)}\right\}$$

Where $f_1\left(\frac{y_i}{X_i}\right)$ is the probability density function for one model, say Zero-Inflated Negative Binomial, and $f_2\left(\frac{y_i}{X_i}\right)$ is the probability density for comparison model for example the standard Negative Binomial.

Hence, the Vouong's test for the hypothesis $E(m_i) = 0$ is given by:

$$V = \frac{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} m_i\right)}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(m_i - \overline{m})^2}}1$$

Under the null hypothesis, the Vuong's statistic is asymptotically normally distributed. At 5% significance level.

The Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) are a goodness- of- fit criteria used for model selection Yesilova, Kaydan, & Kaya (2010). Many Mote-Carlo simulation indicates that the BIC and the AIC selection criteria need to be used together, they are described as follows (Yesilova et al, 2010):

$$AIC = -2L + 2_p$$

And

$$AIC = -2L + p\ln(n)$$

In the Equations above, *L* indicates the log likelihood value, *p* is the parameter number and *n* is the sample size.

2.4.2 Goodness- of-fit tests

The likelihood ratio test is a statistical test used for comparing the goodness- of- fit of two models, one of which is the null model and the alternative other. It is more correct to use a model that has been tested and proves to fit the data well. Still, goodness of fit alone may not be sufficient as a criterion for model selection. The test is based on the likelihood ratio which expresses how many times more likely the data are under one model than the other. The logarithm of the obtained likelihood ratio, can then be used to compute a *p*-value, or compared to a critical value to decide whether to reject the null model in favor of the alternative model. When the logarithm of the likelihood ratio is used, the statistic is known as a log-likelihood ratio statistic, and the probability distribution of this test statistic, assuming that the null model is true, can be approximated using Wilks' theorem.

When two nested models are compared, goodness of fit will always favour the broader and more complex one since it is the one that is more favourable to the data. Thus, model selection based on such a criterion may result in models that over fit the data by paying too much attention to the noise, and as power decreases as the number of parameters grows, we may fail to find any significant association when applying over fitted models (Tang et. al., 2012). According to Box, Jenkins & Reinsel (2008), the goal for model selection is to find a comparatively simple model that adequately represent the data.

2.4.3 Akaike Information Criterion (AIC)

As a measure of Goodness-of- fit considering the influence of parameters for estimated models, AIC defined as: $AIC = -2logl + 2p$ where l is the maximized value of the likelihood function for the estimated model, and p is the number of parameters in the statistical model. The AIC attempts to select the model that best explains the data with minimum parameters (Zou, 2012). The Lower the absolute value of the AIC the better the model. For AIC the penalization is only imposed on the number of parameters in the model. On the other hand, if the difference in AIC of different models is of less than two, the models are reported as equally good models.

The AIC takes both the goodness-of-fit and model complexity into consideration, and enables us to compare two models nested or not. The disadvantage of the AIC is that it is not consistent selection criterion in that the probability of selecting the true model among the candidates does not approach one as the sample size goes to infinity says (Tang et. al., 2012).

2.4.4 Bayesian Information Criterion

The log of the likelihood function is analysed until it converges, the BIC is calculated as: $-2ln(L) + kln(n)$,

Where n = the number of data points in the dataset, the number of observations or the sample size, k = the number of free parameters to be estimated and l = the maximized value of the likelihood function for the estimated model (Zou, 2012). When given a number of estimated models, the model with the lower value in BIC is the one to be preferred. The BIC method can be used to test any types of models, the model interpretations are as follows for absolute values: 0-2 weak, 2-6 positive, 6-10 strong and greater than 10 very strong sample Size. The BIC combines model fit and model complexity, it's between the maximized likelihood representing lack of fit and the penalty term. Furthermore, if the difference in BIC between any two models is lower than 2, then there is no significance difference between the two model fit.

2.4.5 Deviance Information Criterion (DIC)

DIC is based on the same principle as the AIC and is also called the Schwarz information criteria. It is used when the fitted model is assumed to have a good representation of central location by the posterior mean in describing the posterior distribution of the estimated parameters. The DIC uses the effective number of parameters, according to Best and Richardson (2009) the DIC takes the form:

$DIC = D\bar{\theta} + 2pD$. Where pD is the effective number of parameters in the model, $D\bar{\theta} = 2\ lnl(x|\theta)$. Since the parameters are estimated from the sample drawn from the posterior distribution the deviance evaluated at the posterior mean of the estimated parameters is $D\bar{\theta}$. The effective number of parameters pD is therefore the difference between posterior mean deviance $D\bar{\theta}$ and the deviance evaluated. At the posterior mean $D\bar{\theta}$ of the estimated parameters, $pD = \overline{D\theta} + D\bar{\theta}$ resulting in:

$$\begin{aligned} DIC &= \overline{D(\theta)} - pD + 2pD \\ &= \overline{D(\theta)} + pD \end{aligned}$$

A change in the deviation with respect to the deviation of the preceding model to be significant difference if the value is at least 4.

It is important to point out that it is not the exact value of AIC or BIC that is of interest, but rather the change of the index across the different model that is informative for ranking models to select the best among the competing alternatives. Furthermore, the comparison of AIC and BIC is only sensible when the models are applied to the same dataset (Tang at. al. 2012).

2.4.6 The Vuong's Test

The Vuong's test is used as an alternative to the AIC and BIC, for example when the AICs and BIC is used, the model with lower information criteria is preferred. However, with the AIC and BIC we cannot determined whether model A is better that model B probabilistically (Desmarais & Harden, 2013). The Vuong non-nested test is based on a comparison of the predicted probabilities of two models that do not nest (Wilson, 2015). The formula for Vuong statistics:

$$\text{Vuong Statistics} = \frac{\text{LR(model1,model2)}-\text{C}}{\sqrt{NxV}} \sim \text{N (0,1)}$$

Where LR () is the summation of individual log likelihood ratio between 2 models. "C" is the correction term for the difference of Degrees of Freedom between 2 models.

For example zero-inflated Poisson versus ordinary Poisson, or zero-inflated negative-binomial versus ordinary negative-binomial). A large, positive test statistic provides evidence of the superiority of model 1 over model 2, while a large, negative test statistic is evidence of the superiority of model 2 over model 1. Under the null that the models are indistinguishable, the test statistic is asymptotically distributed standard normal (IDRE, 2015).

Wilson (2015) further elaborates that the null hypothesis of Vuong's test for non-nested models is that the expected value of their log-likelihood ratios equals zero, this implies that under the null hypothesis both models are "equally far away" from the data that is being modelled. If we temporarily ignore the issue of whether zero-inflated models and their non-zero-inflated counterparts are non-nested or otherwise, and consider them non-nested. To appropriately simulate the distribution of the log-likelihood ratios it would be necessary to re- sample from data that was somehow equidistant from zero-inflated and non-zero-inflated data, it is difficult to envisage the nature of such data. More importantly, non-rejection of the null hypothesis of Vuong's test for non-nested models, where the (supposedly) non-nested models are, say, the zero-inflated Poisson and standard Poisson model would mean that there is no evidence to conclude that either model fits the data better than the other, not that there is no evidence to support zero-inflation, and its rejection simply implies that either the zero-inflated Poisson model fits the data better than the Poisson model, or vice-versa, not that zero-inflation is present or absent.

For illustration: Let $f_1\ (y_i|\theta_1)$ and $f_2\ (y_i|\theta_2)$ denote the distribution function of two models. Under the classic testing paradigm, the form of the correct distribution is given and only the true vector of parameters in known. Under Vuong's setup, the form of the distribution is also not set up. So, it is possible that neither f_1 nor f_2 is the correct model for the data (Tang et. al., 2012). The idea of Vuong test is to compare the likelihood function under the two competing models. If the two models fit the data equally well, then their likelihood function would be identical. Vuong's test is to compare the best likelihood function that may be achieved between two models, Meaning:

$$E[\log(f_1(y_1|\theta_1^*))] - E\left[\log\left(f_2(y_1|\theta_2^*)\right)\right] = E\left[log\left(\frac{f_1(y_1|\theta_1^*)}{f_2(y_1|\theta_2^*)}\right)\right].$$

If the absolute value |V| is small, the corresponding p-value is bigger than a pre-specified critical value such as 0.05, then we will say that the two models fit the data equally well with no preference given to either model. But, if |V| yields a p-value smaller than the thresholds 0.05, then one of the models is better.

2.5 Conclusion

Hu,et al. (2011), utters that, a zero-inflated model assumes that the zero observations have two different origins: “structural” and “sampling”. The sampling zeros are due to the usual Poisson (or negative binomial) distribution, which assumes that those zero observations happened by chance. Zero-inflated models assume that some zeros are observed due to some specific structure in the data (Hu et al., 2011). For example, if the count for people that have been injured in road crash accidents is the outcome, some people may be involved in an accident that was so slight that there couldn’t be an injury of any sort: these are what Hu calls the structural zeros, because of the slightness of the crash, people would not be in a situation to sustain any injury. Other victims have been in a severe crash were even some people might have died, but some did not sustain any injury, but they were at great risk of being severely injured or even dead. This Hu et al (2011) assumes that it is a Poisson or negative binomial distribution that includes both zero (the “sampling zeros”) and non-zero counts. In contrast, Hu et al (2011) a hurdle model assumes that all zero data are from one “structural” source. The positive (i.e., non-zero) data have “sampling” origin, following either truncated Poisson or truncated negative-binomial distribution.

Literature has shown that the models to consider when modeling count data are: Poisson, the Negative Binomial, the Zero- Inflated Poisson, the Zero-Inflated Negative Binomial models, the Hurdle Poisson models and the Hurdle Negative binomial models. The tests for dispersion, and goodness-of-fit determined the appropriate model for this data, and therefore this study. Also, variables such as: time of crash, gender, district, vehicle choice, driving experience can be used to predict and model road crashes, Injuries and fatalities data.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

The previous chapters introduced the objectives, research problem and several research purposes that would direct the data analysis. Additionally, the review of literature established the background, similar studies by other researchers and a brief presentation of road crashes in general. The different possible methods to explore for the study have been stated. The research is in an effort to create useful models that look at the injured persons in Namibia. The chapter discusses the research design, Population, Data collection methods and data analysis procedure.

3.2 Research design

The study was a quantitative cross sectional research design for all road crash injuries recorded between 2011 and 2016. The data from the MVA Fund crash and claims web based system was retrieved and exported to Micro soft excel, the analysis of this study has been done in SPSS and R statistical packages. The study analysed injuries from crashes where one or more injuries or a deaths has occurred, for crashes recorded on the MVA Fund database before 12 February 2017 (when the data was requested by the researcher) but occurred between 2011 and 2016 . All injuries that occurred and were reported during the period of review and are part of the database will be part of the study. No sampling was necessary since all injuries that occurred between 2011 and 2016 and are recorded on the MVA Fund database and will be used for analysis.

3.3 Procedure

Secondary data retrieved from the MVA fund database was used for analysis; the main source of data is the MVA Fund Call Centre, where crashes are reported, through the toll-free Accident Response Number +264 819682. Information collected and recorded by the Call Centre is verified with the Namibian Police, Emergency Medical Rescue Services - Paramedics and Health Officials from public and private hospitals throughout the country.

The data is captured by regions and by towns, per one crash, the crash may involve more than one vehicle and more than one person, one crash may have non-injured, slight injuries, severe injuries and fatalities. This data is collected at the call centre where an operator gets direct information from any one reporting the crash through the toll free number cited before. After the crash has been verified, the operator enters this information on the database and any other operator is allowed to edit on that individual’s file if additional information is provided, the call centre operators records this information in the system based questionnaire. The Fund adopted the WHO’s method of recording road deaths since 2009, aligning to its standard definition of road fatality as “any person killed immediately or dying within 30 days as a result of a road crash” (WHO, 2013).

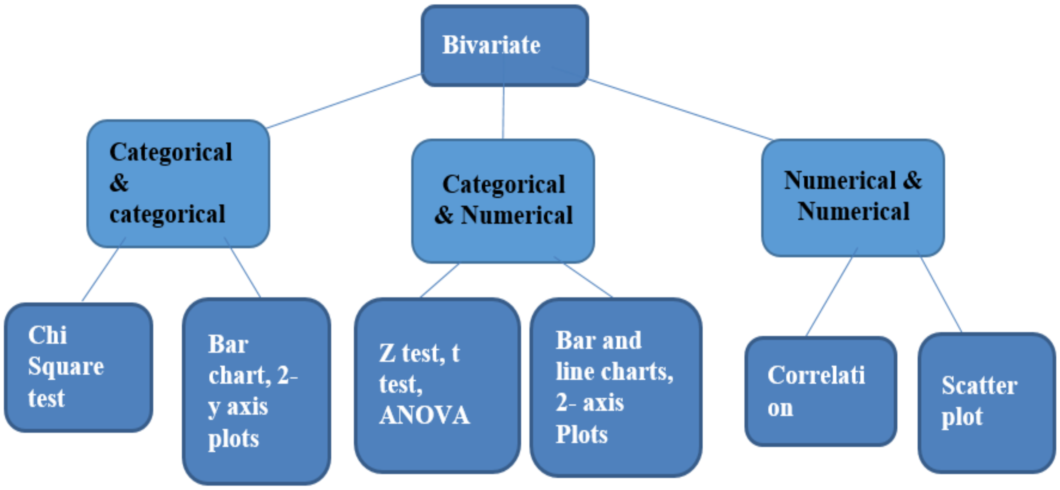
It should be noted that the MVA Fund only records statistics from crashes that resulted in injuries and / or fatalities. In essence, all crashes that resulted in property damage only are duly excluded in the data presented in this report.

3.4 Data Analysis

The researcher employed Micro soft excel for the outlook of the tables and figures, SPSS for data verification, removal of duplicates, and all basic data cleaning, validations and deriving of variables. The Generalised linear models were performed in the R version 3.4.0 and 3.2.0 software with additional packages like “MASS, “pscl” and stats for successful modelling. One of the objectives of this paper is to test association between different variables to know which variables have a relationship with the dependent variable.

Figure 3.1 shows a flow diagram for selecting the appropriate test statistics that need to be employed when selecting an appropriate test for association between types of variables – association between the independent variables and dependent variables. The diagram stipulates that if both the dependent and the independent variables are categorical, the chi- square method should be employed to test for association and a bar chart to present the data; but, if one of the variables categorical and the other numeric, the Z test, t test and the ANOVA test are the best methods to test for association and a bar chart or a line chart to present the data; and, if the two variables are both numerical, the r^2 correlation should be performed to test for association between the two variables using the scatterplot to present the data.

Figure 3.1 test for association flow chart is as follows



3.4.1 Descriptive analysis

Descriptive Analysis is a form of measure of centrality and dispersion, frequencies and cross tabulations of the injuries by region, month, type of crash, cause of crash, year crash occurred, number of people involved, number of vehicles involved. The dependent variable is the number of injuries per crash occurred, and the potential predictors are: month, region, towns, types of crashes, causes of crashes, time of day and number of vehicles involved. Also, variables such as: time of crash, gender, district, vehicle choice, driving experience can be used to predict and model road crashes, injuries and fatalities data.

3.4.2 Generalised linear models

Hu, et. al. (2011) established that count data follow a Poisson distribution; however, in practice such data often display greater heterogeneity in the form of excess zeros (zero-inflation) or greater spread in the values (over- dispersion) or both. Models that have been developed to handle over-dispersion are: (negative binomial (NB) model) or zero-inflation (zero-inflated Poisson (ZIP) and Poisson hurdle (PH) models) or both (zero-inflated negative binomial (ZINB) and negative binomial hurdle (NBH) models). Lee, Han Fulp & Giuliano (2011) agrees with Hu et al (2011) that models that are attributed to excessive zeros are Poisson, negative binomial, zero-inflated Poisson and zero-inflated negative binomial models. Hu et al (2011) further states that the models were compared in terms of covariate estimates along with their statistical inferences. Akaike's Information Criterion (AIC) values were used to consider the relative model fitting for the models as a goodness- of- fit statistic.

Lee, et. al., (2011) does not disagree with the previous authors that Poisson model can be applied to the count data of events occurring within a specific time period. That, the main feature of the Poisson model is the assumption that the mean and variance of the count data are equal. The mean is not always equal to the variance; it hardly happens in real life data. In fact, continues Hu et al (2011), in most cases, the observed variance is larger than the assumed variance, which is called over-dispersion. Furthermore, when the observed data involve excessive zero counts, the problem of over- dispersion results in underestimating the variance of the estimated parameter, and thus produces a misleading conclusion.

Literature has shown that the models to consider when modeling count data are: Poisson, negative binomial, Zero-Inflated Poisson, Zero-Inflated Negative Binomial models, and hurdle models. The tests for dispersion, and goodness-of-fit are used to determine the appropriate model for this data, and therefore this study.

Firstly, numerical coding was to be applied to categorical variables, and numerical variables. The latter were grouped using the SPSS package. Secondly, descriptive analysis was employed for all variables to better simplify the amounts of data in a more sensible form; to show a simpler picture of how the data is, using SPSS and excel for the aesthetics of the tables and figures. Thirdly, data was exported into the R 3.4.0 package to tests between the independent variable (Persons injured) and the dependent categorical variables (Day of crash, month, region, crash type, crash cause, time of crash) and numerical variables (Number of persons Injures, fatalities and number of vehicles involved). . Lastly, before the analysis could begin, there was a need to cluster the data into binary form. In order to run Zero-inflated and Hurdle models the data needed to be in forms of 0s and 1s. Also, R version 3.4.0 could not accept the pscl package within R to do the specific analysis using Zero- inflated and Hurdle models, So,, the researcher adopted version 3.2.2 of R, installed package pscl and analysis was successfully completed. Best models were observed using the Vuong's test and AIC as indicated in the literature review.

3.5 Ethical consideration

This particular researches involve human beings that may have been traumatized. Data from the MVA Fund did not include unique identification facts like names or ID numbers. Data provided by the MVA Fund are for this research purpose only, it will be treated with strict confidentiality and will not be share with any other party unless it is for contribution to this study only. Hence, an oath of secrecy was taken between the researcher and the MVA Fund.

CHAPTER 4

DATA ANALYSIS AND RESULTS

4.1 Data Management

The data was extracted from the database within the MVA Fund as is, it had numerous duplicates, prank call and non-crash related incidences. Due to the fact that the MVA Fund toll free number is free, some people call the number pretending to have been involved in a crash while it is not true resulting in a prank call. On other occasions another misfortune can occur in the society for example, domestic abuse, other illnesses or even accidents that are not motor vehicle related and therefore are non-crash related and should not be part of the dataset. Since all these records are captured in the MVA Fund database, when the data is requested and received, all the non- crash incidents forms part of it. Thus, it was the responsibility of the author to analyse each case to make sure that it was a legit crash.

Every crash has it's unique code "crash number" that constitutes the date, region letter, and a number of the crash that occurred that day in that region, this variable made it easy to identify duplicates and remove them using "remove duplicates" function in excel, the prank calls and non-crash incidences were removed manually using excel. Variables like year of occurrence, month, day were derived from the "Crash number" variable. Time, Number of persons injured, Number of persons involved and number of vehicles involved were grouped accordingly.

4.2 Description of variables

Table 4.1 shows the variables used for modelling. Categorical variables are String alphanumeric or numerical variables that use numeric codes to represent categories or qualitative data. In the dataset the Categorical variables are: Crash Number, Month, Day of crash, Region, Crash Type and Crash Cause. Whereas, Numerical variables are number variables that are used to measure of count, these are: Date Occurred, Year of occurrence, Crash time, Persons Involved, Persons Injured, Fatality, Vehicles Involved and G_time.

Table 4.1 Description of key variables

Variables	Description
Persons Injured	The number of people injured per crash
Year of occurrence	This is the number of crashes per year
Month	The number of crashes per month (1 = January, 2 = February, 3 = March, 4 = April, 5 = May, 6 = June, 7 = July, 8 = August, 9 = September, 10 = October, 11 = November, 12 = December)
Day of Crash	Number of crashes per that day (1 = Sunday, 2 = Monday, 3 = Tuesday, 4 = Wednesday, 5 = Thurday, 6 = Friday, 7 = Saturday
Region	Number of injred people per region over the years(1 = Caprivi, 2 = Erongo, 3 = Hardap, 4 = Karas, 5 = Kavango, 6 = Khomas, 7 = Kunene, 8 = Ohangwena, 9 = Omaheke, 10 = Omusati, 11 = Oshana, 12 = Oshikoto, 13 = Otjozondjupa)
Crash Type	Types of crashes (1 = Collision with animals, 2 = Collision with object, 3 = collision with other vehicles, 4 = Fell from moving vehicle, 5 = Motor Cycle/Bicycle, 6 = Other, 7 = Pedestrian, 8 = Roll over, 9 = Unknown

Table 4.1 Description of key variables continues…….

Variables	Description
Crash Cause	The potential cause of the crash (1 = Animal, 2 = Fell from moving vehicle, 3 = Intoxicated, 4 = Mechanical Failure, 5 = Motor cycle/bicycle, 6 = Other, 7 = Overloading, 8 = Pedestrian, 9 = Poor road design, 10 = Poor Visibility, 11 = Poor weather condition, Reckless and Negligent Driving, 13 = Speed, 14 = Tyre Burst,99 = Unknown)
Persons Involved	The total number of people that were involved per crash (1 = 1 to 5, 2 = 6 to 10, 3 = 11 to 15, 4 = 16 to 20, 5 = 21+)
Fatality	Number of people that died per crash (0 = 0, 1 = 1 to 5, 2 = 6 to 10)
Vehicles Involved	Number of vehicles involved per crash (1 = 1, 2 = 2, 3 = 3, 4 = 4+)
G_time	time the crash occurre (1 = 00:00 -1:59, 2 = 02:00- 03:59 3 = 04:00-05:59, 4 = 06:00-07:59, 5 = 08:00-09:59, 6 = 10:00-11:59, 7 = 12:00-13:59, 8 = 14:00-15:59, 9 = 16:00-17:59, 10 = 18:00-19:59, 11 = 20:00-21:59, 12 = 22:00-23:59

4.3 Data Analysis

The data analysis is performed using two statistical software simultaneously; The Statistical Program for Social Sciences (SPSS) and R. SPSS was mainly for data cleaning (identifying and removing duplicates, managing missing data, renaming and recoding variables) deriving of variables and running frequencies. Whereas the R package was used for model estimation criterions and the actual modelling of the data.

4.3.1 Descriptive data analysis

Table 4.1 shows the frequency of crashes that occurred per year from 2011 to 2016, the crashes show a tremendous increase especially from 2011-2016. There is a substantial increase in the number of crashes per year with a slight decline in 2016.

Figure 4.1 Number of crashes per year 2011-2016

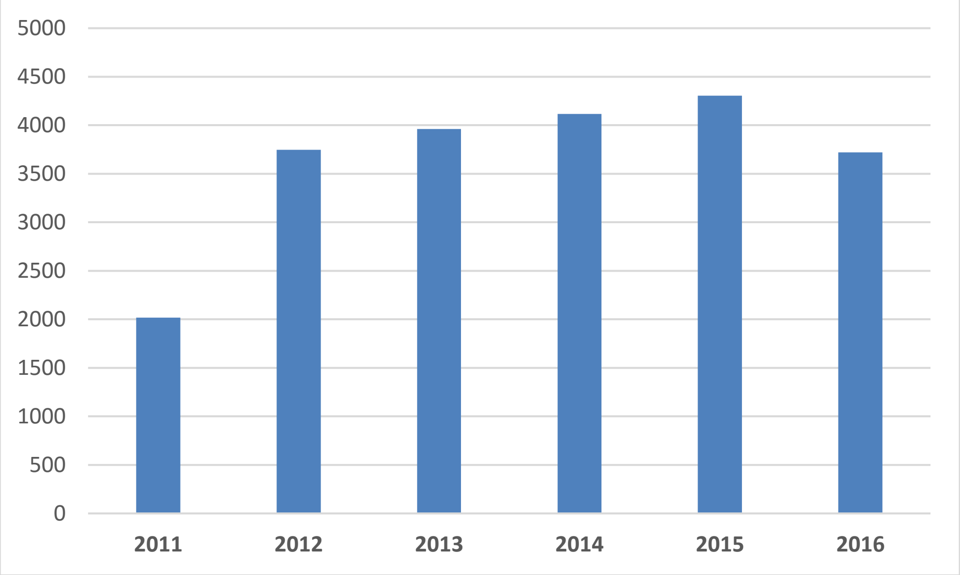


Table 4.2 shows the distribution of crashes by month and year of occurrence. There are relatively more crashes occurring on schools and public holiday months and towards the end of the year like March, April, July, August, October and December. December has the most number of crashes and January has the least accept, in 2013 and 2015 where July and August had the highest occurrence of crashes respectively. The year 2011 had a lot of missing data, which could have contributed to fewer crashes recorded in that year.

Table 4.2 Number of crashes per month and by year

	2011	**2012**	**2013**	**2014**	**2015**	**2016**	**Total**
January	26	243	269	269	320	30	1157
February	25	301	286	318	307	174	1411
March	30	317	353	367	391	378	1836
April	44	316	286	318	353	396	1713
May	68	303	308	373	360	332	1744
June	125	318	333	343	348	335	1802
July	141	326	403	337	368	395	1970
August	174	331	359	390	401	368	2023
September	365	320	337	296	332	330	1980
October	349	315	328	341	383	320	2036
November	301	281	333	369	353	262	1899
December	368	377	366	395	390	402	2298
Total	**2016**	**3748**	**3961**	**4116**	**4306**	**3722**	

It is evident in figure 4.2 that there are more crashes occurring during the weekend as compared to other days of the week. Saturday has the highest occurrence of crashes followed by Friday and Sunday.

Figure 4.2 Number Of crashes by day of the week

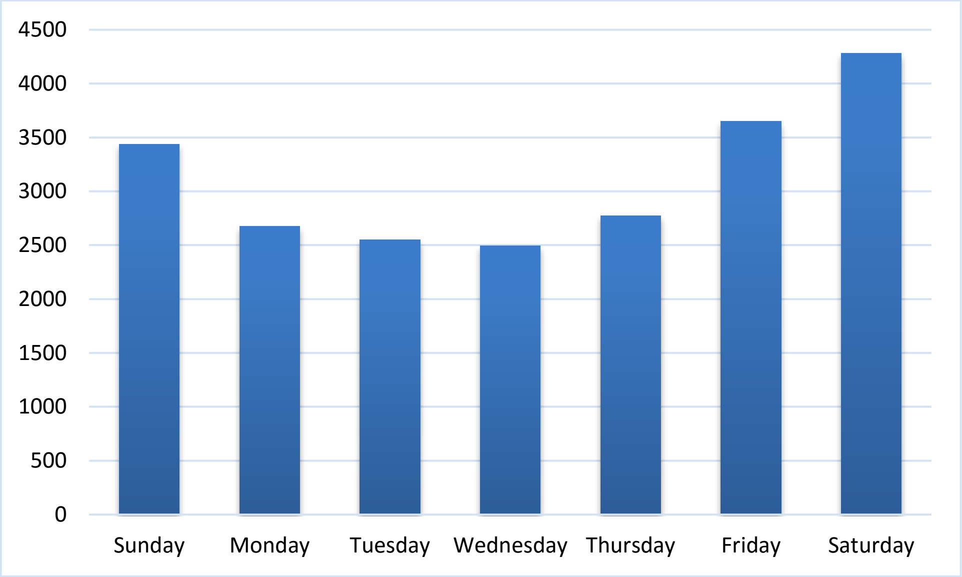


Figure 4.3 illustrates to us the number of crashes by region, the capital Khomas had the highest number of crashes followed by the Oshana, Otjozondjupa and the Erongo regions. The regions with the least crashes are Kunene, Omaheke and Caprivi regions.

Figure 4.3 Number of crashes by region

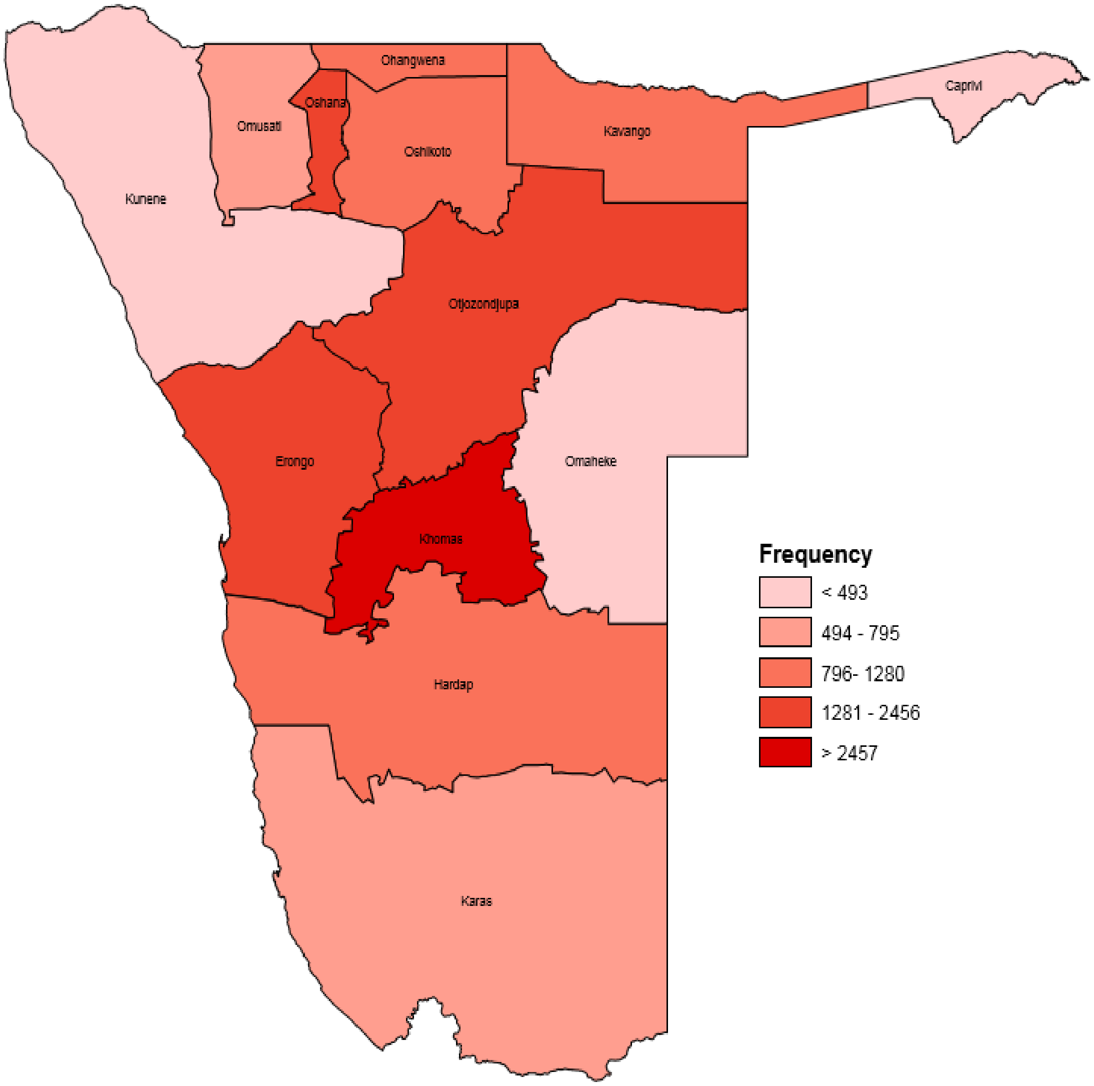


Table 4.3 stipulates that the leading type of crashes is the pedestrian, a pedestrian crash occurs when a vehicle collides with a human, this type of crash may causes serious injuries and death because the impact is on the un protected human body. Roll overs and collisions with other vehicles are relatively high types of crashes. These vehicles roll over because the forces the tires produce in the sideways (lateral) direction are strong enough to roll the vehicle over (Wielenga,1999).

Table 4.3 Number of crashes by type

	Frequency	**Percent**
Collision with Animal	787	3.6
Collision with object	1 250	5.7
Collision with other vehicles	5 502	25.2
Fell from moving vehicle	461	2.1
Motor cycle/Bicycle	773	3.5
Other	607	2.8
Pedestrian	6 128	28.0
Roll over	5 922	27.1
Unknown	439	2.0
Total	**21 869**	**100.0**

Due to inadequate resources in the crash investigations in Namibia, precise causes of crashes are not determined, therefore, the international classification of causes of death (ICD) could not yet be met for Namibia and hence for this study. Nevertheless, the accident report form by Namibia Police Force (Nampol) gives a space to tick the causes of crashes based either on the story of witnesses or what the investigators speculates might be the cause. Usually if an animal or blood is found on the vehicle, it is assumed that the animal was the cause of the crash, overlooking what the driver's behaviour might have been at the time. Some investigations take long before the actual results are out, consequently, some crashes remain "under investigation" especially where homicide is speculated.

Table 4.4 shows the classification of causes of crashes as indicated by the crash investigator. The highest number is other which contains a large number of other causes that the research decided to group together since individually the numbers were so small, for example: trying to avoid an obstruction on the road, saw an animal and tried to employ brakes, fell from moving vehicle. A large number of causes is due to recklessness and negligence of the road users this includes being intoxicated, using external devices while driving, not obeying traffic rules, driving without a lawfully obtained driving licence driving while exhausted and overtaking at blind spots. The author deliberately decided to analyse speed as a cause on its own and not be part of the recklessness and negligence category for emphasis and to light it on its own. Mphela (2011) established that: In Saudi Arabia 50% of crashes occur because of speed. This table proofs that crashes on the road are due to driver behaviour and quite little with other factors.

Table 4.4 Possible causes of crashes

	Frequencies	Percentage
Other	6256	28.6
Reckless and Negligent Driving	4408	19.9
Speed	3216	14.9
Pedestrian	2734	12.5
Missing	1387	6.3
Unknown	1361	6.2
Tyre Burst	664	3.0
Animal	486	2.2
Motor cycle/bicycle	314	1.4
Mechanical Failure	260	1.2
Fell from moving Vehicle	212	1.0
Poor Visibility	211	1.0
Poor road design	204	.9
Poor Weather condition	111	.5
Overloading	41	.2
Animal	4	.0
Total	21869	100.0

Table 4.5 demonstrates the number of deaths that occur per crash, the table indicates that an immense number of crashes occur with few fatalities as a result. But, 11.9% of the time the crash will cause a fatality, pedestrian crashes are the ones highly likely to have 1 person

died. There are reports also where you find 6 or more people have died in one crash. 99 is coded to represent the missing data of this variable.

Table 4.5 Number of fatalities

	Frequency	Percent
0	19 240	88.0
1 to 5	2 600	11.9
6 to 10	23	.1
99.0	6	.0
Total	21 869	100.0

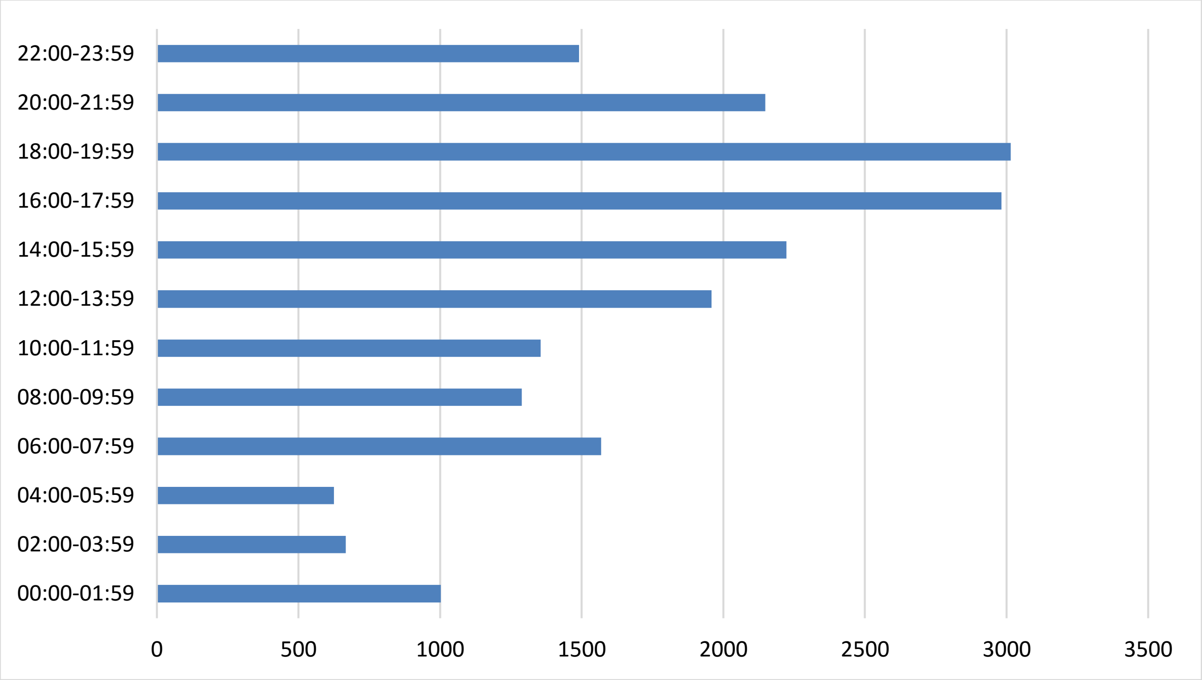
Table 4.6 demonstrates the number of vehicles involved in a crash, it is evident that for a crash to occur a vehicle has to have been involved, 71% of all crashes involved one vehicle, almost 25% involved 2 vehicles . Single vehicle crashes are those crashes that involve Pedestrians, roll overs or collision with fixed objects.

Table 4.6 Number of motor vehicles involved in a crash

	Frequency	**Percent**
1	15 727	71.9
2	5 422	24.8
3	597	2.7
4+	123	.6
Total	**21869**	**100.0**

Figure 4.4 illustrates the number of crashes by days of the week, it is evident that most of the crashes occur at 18:00 – 20:00 Hours of the day followed by the peak time of 16 -18 hours. The evening and afternoon peak hours displays more crashes compared to the peak hours of the morning.

Figure 4.4 Number of crashes by day of the week



The histogram displayed in figure 4.5 is the frequency distribution of the number of people injured per crash. Crashes in which 1 person has been injured has the highest frequency indicating that there are more crashes that occurs and one person has been injured, followed by crashes where no person gets injured. The graph indicates that the data is skewed to the right, this indicates that the dependent variable is not normally distributed. Meaning that the Poisson distribution is to be used to model the data.

Figure 4.5 Number of injured persons per crash

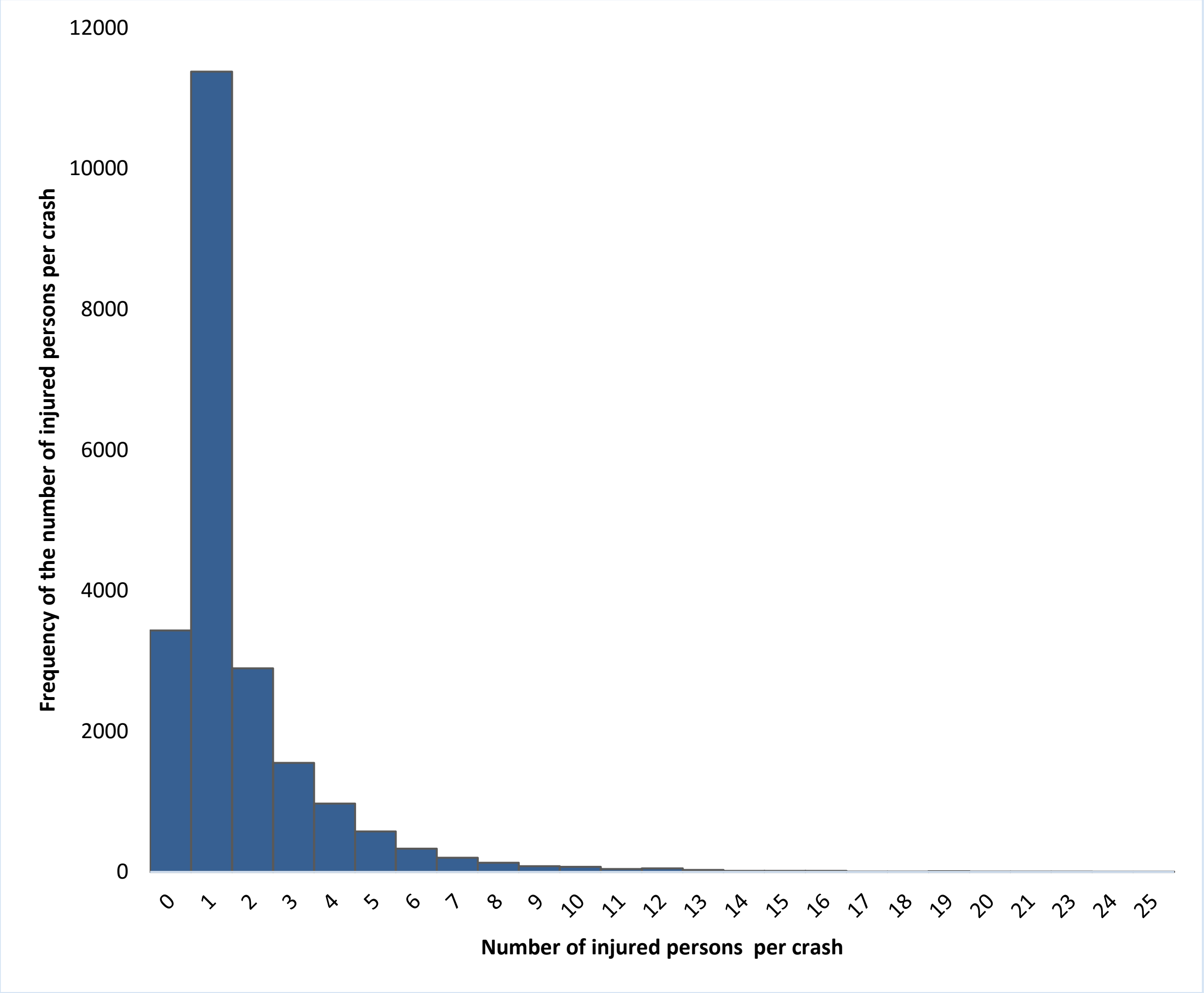


Table 4.7 indicates the minimum, maximum, mean, standard Deviation, skewness and kurtosis of the numeric variable of the data. The standard deviation measures the spread out of the data, the larger the standard deviation the more spread-out the observation. Variables; Number of persons involved, Number of persons injured, fatalities and number of vehicles involved in table 4.7 are all positively skewed, none of the variables are normally distributed. For the purpose of this research there will be no further normality

tests to be performed. However, the histogram for the dependent is displayed in the figure 4.5 to further demonstrate its Skewness.

In addition, below is the descriptive statistics of the numerical variables that are part of this study, on average 1.135 people are involved, 0.911 injured, 0.121 deaths and 1.32 vehicles involved in each and every crash that occurred. There is evidence that the data is not normally distributed and that it is positively skewed (skewness of normally distributed values is equal to 0). Kurtosis is very high for all the variables of interest again this is another indication that the data is not normally distributed. The mean is very much lower than the variance which means that there is over dispersion in the data. Indicating that generalised linear models are to be considered for this data.

Table 4.7 Descriptive data numeric variables

	Minimum	**Maximum**	**Mean**	**Std. Deviation**	**Skewness**	**Kurtosis**
Number of Persons Involved	1	5	1.135	0.4674	4.573	25.666
Number of Persons Injured	0	6	0.911	0.5375	2.264	20.526
Fatality	0	2	0.121	0.3294	2.412	4.163
Number of Vehicles Involved	1	4	1.319	0.5529	1.731	3.087

4.3.2 Tests for association

As part of the objectives of the paper, tests for associations between the dependent variable and the independent variables is required, since the research paper has both numeric and categorical variables different tests are performed. The variable Persons injured was recoded into categorical variable in order to carry out a chi- Squared test, between persons injured and the categorical variables (Month, Day of crash, Region, Crash type, Crash cause and Grouped time).

Table 4.8 below indicates the test results between Persons injured and categorical variables; the chi-square test for independence or Pearson's chi-square test is used to determine if there is association or a relationship between two categorical variables. At 5% level of significance the P value for all the variables is less than 0.05 this means that there is enough evidence to prove that there is association between the Persons Injured and month, Day of crash, Region, Crash type, Crash cause and time the crash has occurred. There is a high correlation of 0.874 between the Crash cause and the dependent variable, this is the same for crash type as well. Although there is correlation between month and persons injured, the relationship is close to zero.

Table 4.8 Test for association Persons Injured~ categorical variables

	Chisquare test	**P Value**	**Significance**	**R^2**
Month	156.820	<0.001	***	0.085
Day of crash	130.620	<0.001	***	0.077
Region	1013.120	<0.001	***	0.215
Crash Type	4919.140	<0.001	***	0.474
Crash Cause	2745.520	<0.001	***	0.874
Grouped time	664.710	<0.001	***	0.354

***** Variable is significant**

Table 4.9 test for correlation between the dependent variable (Persons injured) and the numerical variables (Number of persons involved, Number of vehicles involved and faatalities). Some at 5% significant level and others at 1% significant level. The numeric variables are strongly correlated with the number of injured persons as compared to the categorical variables. The largest correlation is of 0.039 for persons injured.

Table 4.9 Test for correlation between PersonsInjured~ numerical variables

Variable	Correlation coefficient
# Persons Involved	.039**
Fatality	.015*
# Vehicles Involved	.023**

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

4.3.3 Statistical modelling

Regression analysis treats all independent variables in the same analysis as numerical, with interval or ration scales whose values are directly comparable (Skrivanek, 2009). Often the researcher might want to analyse subgroups of the variables to distinguish different treatment groups, dummy variables are then used to tell the regression algorithm that the numbers are only used to indicate the levels within the variable and they do not have an inherent meaning. In simple cases numeric binary values are used where 0 is the control group and 1 is the treatment group, dummy variables are used as single regression equations that represent multiple groups. Table 4.10 shows how the levels were defined for the predictor variables.

Table 4.10 Description of dummy variables

Variables	Description
Persons Injured	Dependent variable
Month	Holiday months refers to the time when school children are on school Holiday Months (1 = January, April, May, August, December) Non-Holiday months (0 = February, March, June, July, September, October, Novembe)r

Day of Crash	Weekend = (1=Friday, Saturday, Sunday)
	Non-Weekend = (0 = Monday, Tuesday, Wednesday, Thursday)

Region	Northern regions refer to the regions in the northern of Namibia Northern (1= Oshana, Omusati, Ohangwena, Oshikoto, Okavango, Otjozondjupa) Non-Northern (0 = Caprivi, Erongo, Hardap, Omaheke, Kunene, Karas, Khomas)
Crash Type	Vehicle by vehicle means, crashes that do not involve other factors like trees, people or animal only vehicles were 1 or more vehicle were involved Vehicle by vehicle = (1 = Collision with other Vehicles, Roll over) Non Vehicle by vehicle = (0 = Collision with animal, collision with object, fell from moving vehicle, motor cycle/bicycle, other, Unknown)
Fatality	Deaths refers to crashes where 1 and more deaths have occurred, no deaths just means zero deaths occurred. Deaths (1= 1 to 5, 6 to 10), noDeaths (0=0)

Table 4.10 Description of dummy variables continue…

Variables	Description
Vehicles Involved	Single vehicles are crashes that involved only one motor vehicle, none single is when 2 and more vehicle where involved Single vehicle = 1=1 non single vehicle =0= 2,3, 4, 4+
G_time	Peak time is time when traffic is more on the road the time when , off peak there is few vehicles allowing for speed, Peak time (1 =06:00-07:59, 12:00-13:59, 16:00-17:59) off peak time (0 = 00:00 -1:59, 02:00- 03:59 04:00-05:59, 08:00-09:59, 10:00-11:59, 14:00-15:59, 18:00-19:59, 20:00-21:59, 22:00-23:59)
Crash Cause	Driver crash causes that are due to driver reaction or condition Driver = (1= Poor visibility, Reckless and negligence driving, Speed, Intoxicated) NonDriver (0= Animal, Poor Weather condition, Tyre Burst, Fell from moving vehicle, Mechanical failure, Motor Cyclist/ bicycle, other, overloading, pedestrian, poor road design, Unknown)

4.4 Results

4.4.1 Model selection

Table 4.11 show the different generalised linear models used by the author in this study, the Akaike Information Criteria (AIC) and Bayesian Information Criterion (BIC) are goodness of criteria used for model selection the fact that the goodness of fit statistics is greater than 1 shows that there was over dispersion in the data (Yesilova, Kaydan & Kaya, 2010). The Poisson Regression (PR), Negative Binomial (NB), Zero- Inflated Poisson (ZIP), Zero- Inflated Negative Binomial (ZINB), Hurdle Poisson (HP), and Hurdle Negative Binomial (HNB) are given in table 4.11 that produced widely different results. The model with the lowest AIC is the HNB which is bolded out in table 4.11. Therefore, the Hurdle Negative Binomial was chosen as the best model.

Table 4.11 AIC comparison model

	Model Name	**Log-likelihood**	**AIC**
Model 1	Poisson		79890.16
Model 2	Negative binomial	-73812.24	73832.24
Model 3	Zero-inflated Poisson model	-3.89e+04	77838.93
Model 4	Zero-inflated negative binomial	-3.639e+04	72815.07
Model 5	Hurdle Poisson	-3.733e+04	74695.09
Model 6	**Hurdle negative binomial**	**-3.203e+04**	**64089.84**

Table 4.12 gives a summary of the models comparisons based on the Vuong's statistics for the six regression models explored. The rankings of the model are as follows: Poisson <Negative binomial < Zero- Inflated Negative Binomial < Zero- Inflated Poisson < Hurdle Negative Binomial< Hurdle Poisson. Tang et. al. (2012) states that if the corresponding p-value is bigger than a pre-specified critical value such as 0.05, then one can conclude that the two models fit the data equally well with no preference given to either model. But, if |V| yields a p-value smaller than the thresholds 0.05, then one of the models is better.

The Vuong test is developed to test a hypothesis for comparing two models (Tang et. al.2012). A hypothesis test was done with all 6 models combinations to check for the best, results yielded in summarizing the tests between the HNB and the rest of the models. In this case the |V| is greater than the p- value which is less than 0.05 making model two the better model in comparison to all the other five models. Besides, table 4.11 tells us to focus on the HNB, when hypothesis were done to compare the HNB with the other models, the HBN, performed best, which led in its selection as the best model. Hence, we can confidently say the Hurdle Negative Binomial is the best models with evidence shown from the Vuong's test and the AIC comparison model.

Table 4.12 Model comparisons based on the Vuong's statistics

	Hypothesis Testing Against the HNB		Vuong Z-Statistic |V|	P-value
Poisson	(P , HNB)	Model2 > model1	15.254	p = 2.222e-16
Negative Binomial	(NB , HNB)	Model2 > model1	36.765	p = 2.222e-16
Zero-inflated Poisson	(ZP , HNB)	Model2 > model1	13.745	p = 2.222e-16
Zero-inflated Negative binomial	(ZN , HNB)	Model2 > model1	32.204	p = 2.222e-16
Hurdle Poisson	(HP , HNB)	Model2 > model1	10.155	p = 2.222e-16
Hurdle Negative Binomial	(HNB , HNB)	-	NA	-

4.4.2 Model estimates under the Hurdle Negative Binomial Model

Table 4.13 shows the results of the Hurdle negative Binomial regression. The HNB generates two separate models, one that predicts zero outcomes and another for none zeros. The Ordinary ratio (OR) shows model estimates for crashes with zero injuries or the zero outcomes while the Relative Ratio (RR) shows model estimates for crashes that had one or more injuries. When the p- value of any models shows to be more than the assigned level of significance, we reject the null hypothesis with the implication that there is no significance different between the two variables of test. The data is analysed based on the two separate models produced by the HNB; the Model 1 Ordinary ratio which focus on crashes where zero persons are injured, and model 2 the Relative Ratio where one or more people have been injured.

Model 1 (Ordinary Ratio)

There was no significant difference between holiday months and non-holiday months for crashes that did not have any injuries (p=0.66143). Likewise, for weekends OR results indicate that whether it is weekend or not there will still be crashes where no injuries were observed with (p= 0.0786). Also, there is no significance difference (p=0.6788) for northern regions in comparison with other regions in the rest of the country. On the other hand, for types of crashes, types that involved vehicles by vehicle where significantly high, thus if you increase vehicle by vehicle crashes by one point the odds that there would be zero injuries would increase by 0.4 if other variables are held constant (OR = 0.487, 97.5% CI: 0.3948, 0.5789).The causes of crashes that occurred as a result of driver behaviour are 0.147 more than those that have nothing to do with the driver (OR= 0.147, 97.2% CI: 0.0534.0.2409). For crashes where fatalities were recorded, the crashes that had fatalities were 1.835 times way less compared to those that had none, in other words for crashes with more fatalities decreases the number of injuries (OR= -1.835, 97.5% CI: -1.9269, -1.7433). With regards to the number of vehicles involved there was no significance difference between the number of vehicles involved, whether one or more vehicles are involved the probability of have zero injuries is there (p= 0.7885). On the other hand when we look at time, the non-injuries where 0.3 times higher during peak time as compared to off peak time (OR = 0.326, 97.5% CI: 0.2371, 0.4152).

Model 2 Relative Ratio

Results from table 4.13 also, showed that the intensity of injured person's due to crashes is significantly associated with the month, (crashes that had one or more persons injured). During holiday months it is 0.2 times more likely to have a crash with injured people as compared to months with no school holidays, if you are to increase the number of holiday months by one unit, the number of injuries will increase with 0.2 (RR = 0.209, 97.5%, CI: 0.1460, 0.2715). This is similar to days of the week, weekends are 0.1 times more likely to have injuries intensity as compared to other days of the week (RR = 0.106, 97.5% CI: 0.0434, 0.1682). When it comes to the five Northern regions, they are significantly high in the number of injuries when contrasted with other regions in the country (RR = 0.295, 97.5% CI: 0.2305, 0.3595). The model indicates that there is significance difference in the types of crashes, if vehicle to vehicle crashes increase by one unit, the number of injuries will increase by 1.6 while holding other variables constant, accordingly, the higher the number of vehicle by vehicle crashes, the higher the number of injuries (RR = 1.582, 97.2% CI: 1.5094, 1.6546). With the causes of crashes that had something to do with driver behaviour, they were significantly higher than those with no environmental or vehicle factors with (RR = 0.224, 97.5% CI: 0.1562, 0.2915). With regard to fatalities, high injury crashes tent to involve 1.0 times more deaths per crash as compared to those without fatalities (RR = 1.01, 97.5% CI: 0.8974. 1.1227). With the number of vehicles involved, single vehicle crashes involve less injuries compared to crashes with more than one vehicle (RR = -0.216, 97.5% CI: -2.2860, -0.1454). Moreover, the time the crash occurs is insignificant to the number of injuries that occur, at 97.5% confidence interval peak time is not significantly different from off peak time in terms of the number of injuries (p= 0.4054) unlike with zero injuries were the peak time yielded less injuries..

Table 4.13 Regression Estimates from the Hurdle Negative Binomial Regression Model

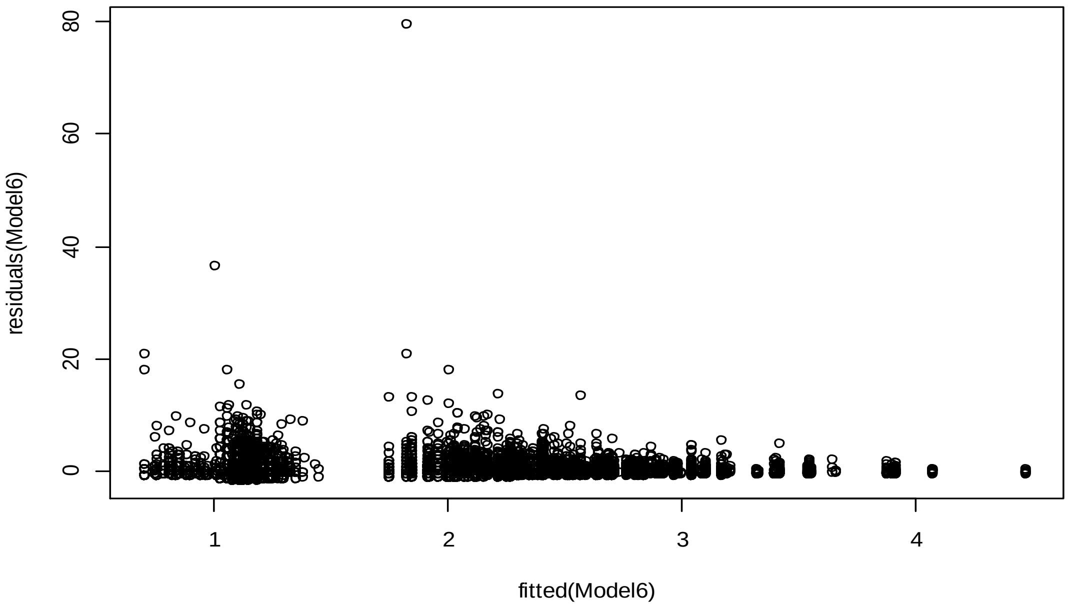
Variable	Hurdle Negative Binomial regression model			
	Injured persons probability (Zero people injured in a crash)		Injured Persons Intensity(# of Injures Persons >= 1)	
	OR	97.5% CI	RR	97.5% CI
Number of persons injured percrash	1.596***	(1.4718, 1.72049)	-4.427***	(-6.4427, -2.4113)
Month				
None Holiday month	1		1	
Holiday Month	-0.017	(-0.0955, 0.0606)	0.209***	(0.1460, 0.2715)
Day of Crash				
Non-Weekend	1		1	
Weekend	0.069	(-0.0079, 0.1469)	0.106***	(0.0434, 0.1682)
Region				
Non-Northern	1		1	
Northern	0.017	(-0.0635, 0.0976)	0.295***	(0.2305, 0.3595)

Table 4.13 Regression Estimates from the Hurdle Negative Binomial Regression Model continues…

Variable	Hurdle Negative Binomial regression model			
	Injured persons probability (Zero people injured in a crash)		Injured Persons Intensity(# of Injures Persons >= 1)	
	OR	97.5% CI	RR	97.5% CI
Crash Type				
Vehicle and other factors	1		1	
Vehicle by Vehicle	0.487***	(0.3948, 0.5789)	1.582***	(1.5094, 1.6546)
Crash Cause				
Non Driver	1		1	
Driver behaviour	0.147**	(0.0534, 0.2409)	0.224***	(0.1562, 0.2915)
Fatalities				
NoDeaths	1		1	
Deaths	-1.835***	(-1.9269, -1.7433)	1.01***	(0.8974, 1.1227)
Vehicles Involved				
More than one	1		1	
Single Vehicle	0.015	(-0.0875, 0.1153)	-0.216***	(-0.2860, -0.1454)
Time				
Off Peak time	1		1	
Peak time	0.326***	(0.2371, 0.4152)	0.029	(-0.0387 0.0959)

Figure 4.6 shows the residual plot and the fitted model plots for the Hurdle Negative Binomial, the rule is that when the residuals on the x- axis is zero, the predictions by the model is accurate, if the values are less than zero then the predictions were too high, but if the values are greater than zero then the predictions are too low. Figure 4.6 shows that the predicted values where correct since most of the values are close to zero, but, slightly low, with an exceptional outlier, there were no predictions that were too high.

Figure 4.6 Residual and fitted model for the HNB



CHAPTER 5

DISCUSSIONS, CONCLUSIONS AND RECOMMENDATION

5.1 Discussions

The number of crashes has increased relatively between 2011 and 2015 with an average of 4% increase every year; however, a 13% decline was experienced in 2016. December has the highest occurrences of crashes among all month while February has the least. When looking at the number of crashes per week, Figure 4.3, shows that 52% of the total crashes occur during the weekend, of which Saturday has the highest occurrence , followed by Saturday then Friday. The dataset had a distribution of crashes by region, the results indicated that 38% crashes were observed in the Khomas region and 11% in Erongo region respectively. Analysis is done looking at the types of crashes results yielded from pedestrian crashes, as the highest type of crash with 28%. Roll overs and collision with other vehicles followed with 27.1% and 25.2 respectively. The causes of crashes on the other hand, reckless and negligent driving, speed and pedestrians with 20%, 15% and 13% respectively. Also, when separately looking at the number of fatalities that occur at the frequency of crashes results it show that 88% of crashes do not result in deaths. Analysis were performed looking at the time when a crash occurs, more crashes happen during the rush hours of the late afternoon between 18:00 – 19:59 and between 16:00 – 17:59. In most crashes there is a high likely that one person will be injured, with a large number of pedestrian crashes this fact is expected. It is a very small probability for a crash to occur and no one injured, the data also showed a minimal number of mass casualties.

The chi-square and the Pearson R^2 test was used to test for association between the predicted and the predictor variables, of which all were found significant at 5% level of significance. All the variables had an R^2 less than 0.5 except for the causes of crashes that had $R^2 = 0.874$ making it the best predictor for road crash Injuries. Therefore, it is reasonable to conclude that the cause of crash can predict that the crash will have injuries.

Exploring the two part model is best, because it gives result of both the zero respondents and those of more than one responses. Poisson, Negative Binomial, Zero-Inflated Poisson, Zero-inflated negative binomial, Poisson Hurdle, and Hurdle negative binomial models were each fitted with “MASS” and “pscl” in R 3.3.2 packages using the glm, nb, zeroinfl, hurdle functions to fit all the models in order to choose the best.

Goodness – of – fit test were performed among the 6 models in order to choose the best one. It is not comprehensible to have only one test for the best model, hence, two tests; Vuong’s test and the AIC were implied. The results for both test indicated that the best model was the HNB, the hypothesis for Vuong revealed the HNB to be superior to the rest of the six models; leading to the decision of selecting the output for HNB to conclude results and the outcomes of the study.

The two-way model looked at crashes that had zero injuries (RO) and those with one or more injuries (RR), on the risk factors of the two there were some similarities found. For example, for CZI (Crashes with zero injuries) there was no significant difference between school holiday months and other months of the year, whereas crashes that had Injuries Greater than or Equal to one (IGE1) school holiday months (which also coincides with moths that have public holidays) had more crashes, concluding that, January, April, May, August and December who consist of school holidays and public holidays increase the number of injuries in a crash. Similarly, Sukhai, Jones, Love & Hayne (2011) declares that South Africa has a significant December peak in Road traffic crash fatalities. The peak is largely explained by traffic flow factors, increased alcohol consumption during the holiday may also be a risk factor. Also, weekends (Fridays, Saturdays and Sundays) proved to have had more IGE1s comparing to weekdays, Elliot (2009), said 52% of injuries that occurs during weekends with most crashes happening on Saturdays and Sundays.

One very interesting fact of the data is that peak hours indicate many CZI, meaning that although there is many crashes during the peak hours they are so minor that people do not get injured or so severe that people die. Kingham, Sabel & Bartie (2011) report their findings that crash rates are not occurring at a uniform rate throughout the day, with comparative increases in crash rates occurring during morning rush hour, and during the ’school runs’. However this study confirms that, crashes that occur during peak time they cause very few injuries, in layman terms, if a crash occurs at pick time, the probability that no one gets injured is high. But, for crashes where injuries have occurred it does not matter the time of day that the crash occurs, whether peak hours, or not, an injury can occur any time. It is worth noting that at this point we are only looking at injuries as our predicted variable, the results may yield differently if deaths were the variable of interest.

When looking in figure 2.1 of page 13 we see that the population density at the Northern regions of Kavango, Ohangwena, Omusati, Oshana and Oshikoto is concentrated, it is logical to assume that more crashes occur in those areas. The study indicates that the Northern regions are prompted to IGE1s more than the CZI which are not different. The mentioned figure also shows how vast the country is, this makes it difficult to tell the exact regions the crash has occurred, some of the very severe crashes occurred between roads that are connecting towns (MVAFund, 2016).

Another fascinating fact is the relationship between the number of none injured persons and fatalities by crash; crashes with fatalities decrease the number of none injured persons denoting that if the number of people that died increase, the number of none injured persons will decrease. This is due to the fact that, if there are fatalities in a crash it indicates the severity of the crash, and the more severe the crash the higher the chances of injuries and the chances of deaths in that crash. By the same token, as the number of deaths increases so does the number of injuries, so the number of fatalities work in conjunction with the number of injuries.

Crashes that occurred as a result of driver behaviour such as, Reckless and Negligence driving, speed and being intoxicated compared to issues like tyre burst, mechanical failure, poor road design, pedestrian, animal, poor weather condition indicated that the number of injured persons is predicted by the cause of crash and that injuries will even be more for crashes that occurred as a result of driver behaviour or circumstances. Ultimately, the behaviour and attitudes of the driver will determine how severe the crash will be, should a crash occur. The other observation was among the types of crashes, crashes resulted in vehicles crash only such as collisions with other vehicles and roll over exhibited crashes with more injuries with a gigantic difference for both the zero injury (CZI) crashes and with crashes that have had 1 and more injuries (IGE1), that is to say the type of crash does predict injuries, and if a crash involves only vehicles for instance collisions with other vehicles and roll overs increase the number of injuries more than the other types of crashes (collision with animals, collision objects, pedestrians, falling from moving vehicle, cyclist and motorcyclists). In contrast, Single vehicle crashes were significantly lower among IGE1 and insignificantly more with CZI, what this means is that crashes that involved one vehicle have the potential to reduce injuries, the fact is that with a single vehicle it is usually that the number of casualties are few for example: in a pedestrian crash, only one person gets either injured or killed, and most vehicles have a maximum of 5 occupants.

5.2 Conclusions

The risk factors associated with road traffic injuries were clearly stipulated in this study: Month, Day of crash, Region, Crash type, crash cause, number of vehicles involved, and the time the crash occurred. The Hurdle Negative Binomial generalised linear model, and chi- square test were used to determine association between the number of persons injured and the dependent variables. The chi – square test showed significant association between the dependent and independent variables with Pearson R^2 of less than 0.5 except for crash cause. Meaning that whatever has caused the crash will determine the number of injuries in that crash, it was further noted with the HNB that driver behaviour causing more injuries.

The Risk factors that influence crashes have been determined and found as follows: The types of crashes for vehicle to vehicle collisions and roll overs posted a greater probability to injuries as compared to pedestrians, collision with fixed objects, fall from moving vehicles and collisions with animals. Causes of crashes were another estimator of injuries; the study found that driver behaviour is a much larger contributor to injuries than other factors such as reckless and negligent driving, speed, poor visibility and being intoxicated. Other causes like the tyre bursts, mechanical failures, weather conditions and road designs are not the huge contributors to injuries. The month in which a crash has occurred is significant when specifically looking at school holiday month which are also months in which most public holidays fall (January, April, May, August and December), these holidays showed a great concern in the number of injuries in a crash. The weekend (Friday, Saturday and Sunday) also proved to increase the number of injuries per crash. Finally, the geography of the crash plays a role in the number of injured persons per crash. The Oshana, Omusati, Ohangwena, Oshikoto, Kavango and Otjozondjupa regions contribute more to the number of injuries compared to other regions. Crashes with fatalities are also good predictors of road crash injuries.

There challenge is that there is a gap in the orientation of the data, there is many missing variables for example, for the causes of crashes more records where for category other, either because they are under investigation or some other reason. Because of this, demographics of the victims could not be analysed. It would be of great advantage to look at the age of drivers, their gender and years of driving experience. Nevertheless, the classification of the levels of the variables can be explored and broken down further, to find specific causes of crashes. The two way models are best to model occurrences because they show you which factors post a greater risk than the other.

5.3 Recommendations

- Policy makers should concentrate on crashes reduction mechanisms that focuses on driver behaviour, issues of speed and being intoxicated post a greater risk to injuries. Since speed is the second cause of crash after reckless and negligent driving, laws can be passed that regulates speed; for instance, all vehicles in the country should be limited to a certain speed limit where it does not matter how the driver accelerates the vehicle will not pass the maximum limit. With an exception of emergency vehicles at least. Or introduce the arrest on the scene mechanism, where if a driver is found driving beyond the speed limit, should be arrested and should appear in court before continuing with their destination.
- The introduction of the penalty points system in Namibia may just be the answer, seeing that driver behaviour poses a huge risk to the increase of crashes. Novoa (2010), did a study to assess the effectiveness of the penalty points system introduced in some European countries showed that in Italy, Ireland and Spain indicated a reduction in the number of people injured by 19%, 36% and 12% respectively.
- School holiday's months coincided with public holiday's weekend's revealed high intensity of injuries. Campaigns should be concentrated during this time periods to check for speed and whether the driver is intoxicated while operating a vehicle. This interactions should be set up any time of the day, the study showed that peak time are not the dangerous times to drive early morning hours and very late at night is a time to explorer.

- As suggested by Reynolds, et al., (2009) a construction of a bicycle and pedestrian only lane plus sufficient pedestrian and cyclist crossing in towns may be the key to reducing pedestrian and cyclist types of crashes that contribute 28% to crashes. Furthermore, almost half of all deaths on the world's roads are road users with the least protection – motorcyclists, cyclists and pedestrians (WHO, 2015).
- Researches should perform a similar study that focuses on the demographics of the road users, to explore the age group, gender, economic status of people to see how they affect road safety.

REFERENCES

Afukaar, F. K., Antwi, P., & Ofosu-Amaah, S. (2010). Pattern of road traffic injuries in Ghana: implications for control. *Injury control and safety promotion*, 10(1-2), 69-76.

Andima, J. (2014, November 14). MVA Fund launches road safety campaign. *The Namibian*, p. 2.

Awadzi, K. D., Classen, S., Hall, A., Duncan, R. P., & Garvan, C. W. (2008). Predictors of injury among younger and older adults in fatal motor vehicle crashes. *Accident Analysis & Prevention*, *40*(6), 1804-1810.

Balk, S. A., Tyrrell, R. A., Brooks, J. O., & Carpenter, T. L. (2008). Highlighting human form and motion information enhances the conspicuity of pedestrians at night. *Perception*, 37(8), 1276-1284.

Beyer, F. R., & Ker, K. (2009). Street lighting for preventing road traffic injuries. *The Cochrane Library*.

Box, G., Jenkins G., & Reinsel G. (2008). Time series analysis: Forecasting and control 4th ed. Hoboken, NJ: John Wiley & Sons.

Caird, J. K., Willness, C. R., Steel, P., & Scialfa, C. (2008). A meta-analysis of the effects of cell phones on driver performance. *Accident Analysis & Prevention*, *40*(4), 1282-1293.

Chimba, D., Sando, T., Kwigizile, V., & Kutela, B. (2014). Modeling school bus crashes using zero-inflated model. *Journal of transportation and statistics*, *10*(1), 3-11.

Damsere-Derry, J., Ebel, B. E., Mock, C. N., Afukaar, F., & Donkor, P. (2010). Pedestrians' injury patterns in Ghana. *Accident Analysis & Prevention*, *42*(4), 1080-1088.

Desmarais, B. A., & Harden, J. J. (2013). Testing for zero inflation in count models: Bias correction for the Vuong test. *The Stata Journal*, *13*(4), 810-835.

Economist (2012, June 29). Sesriem clinic treats all the sick. *Economist.* Retrieved from: https://economist.com.na/2036/general-news/sesriem-clinic-treats-all-sick/Elliot Hannah. (2009, 1 21). *Fashion, Cars and Culture*. Retrieved from Forbes: https://www.forbes.com/2009/01/21/car-accident-times-forbeslife-cx_he_0121driving.html

Eloff, P., & Niekerk, A. V. (2008). *Temporal patterns of animal-related traffic accidents in the Eastern Cape* (Masters thesis), Stellenbosch University, South African.

Guse, C. E., Cortés, L. M., Hargarten, S. W., & Hennes, H. M. (2007). Fatal injuries of US citizens abroad. *Journal of travel medicine*, *14*(5), 279-287.

Garg, N., & Hyder, A. A. (2006). Exploring the relationship between development and road traffic injuries: a case study from India. *The European Journal of Public Health*, *16*(5), 487-491.

Hosseinpour M., Prasetijo J., Yahaya A. S, & Ghadiri S. M. R., (2012), A comparative Study of count Models: Application to Pedestrian-Vehicle Crashes Along Malaysia Federal Roads. *Traffic Injury prevention Journal*, 16 (6), 630 – 638 Retrieved from: http://dx.doi.org/10.1080/1539588.2012.736649.

Hu, M. C., Pavlicova, M., & Nunes, E. V. (2011). Zero-inflated and hurdle models of count data with extra zeros: examples from an HIV-risk reduction intervention trial. *The American journal of drug and alcohol abuse*, *37*(5), 367-375.

IDRE (2016) Introduction to SAS. UCLA: Statistical Consulting Group. From: https://stats.idre.ucla.edu/sas/modules/sas-learning-moduleintroduction-to-the-features-of-sas/ (accessed August 22, 2016).

Iipinge, M., & Owusu-Afriyies, P. (2011). Assessment of the Effective of Road Safety Programmesin Namibia: Leaners perspective. *Journal of Emerging Trends in Economic and Management Science*.

Kenny, C. (2009). Measuring Corruption in Infrastruture: Evidence from transition and developing countries. The journal of Development Studies, 45(3),3140 332.Retrieved from:https://doi.org/10.1080/00220380802265066

Khan, A., Ullah, S., & Nitz, J. (2011). Statistical modelling of falls count data with excess zeros. *Injury prevention*, *17*(4), 266-270.

Kim, D. H., Ramjan, L. M., & Mak, K. K. (2016). Prediction of vehicle crashes by drivers' characteristics and past traffic violations in Korea using a zero-inflated negative binomial model. *Traffic injury prevention*, *17*(1), 86-90.

Kim, J. K., Ulfarsson, G. F., Shankar, V. N., & Mannering, F. L. (2010). A note on modeling pedestrian-injury severity in motor-vehicle crashes with the mixed logit model. *Accident Analysis & Prevention*, *42*(6), 1751-1758.

Kingham, S., Sabel, C. E., & Bartie, P. (2011). The impact of the ‘school run’on road traffic accidents: A spatio-temporal analysis. *Journal of transport geography*, *19*(4), 705-711.

Lambert, D, (2012) Zero-Inflated Poisson Regression, With an Application to Defects in Manufacturing. *Technimetrics*, *34* (1), 1-14,

Langarde E (2007) Road traffic Injury is an Escalating Burden in African and Deserves Proportionate Efforts. *PLOS medicine Journal* 4(6): 170.doi:10.1371/journal.pmed.0040170

Lee, H. J., Han, G., Fulp, W. J., & Giuliano, A.R.,(2011) Analysis of overdispersed count data: application to the Human Papillomavirus Infection in Men (HIM)

Study. *Epidemiol Infect journal*, 140(6), 1087-1094. Doi: 10.1017/S095026881100166X

Lord, D., & Mannering, F. (2010). The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. *Transportation Research Part A: Policy and Practice*, *44*(5), 291-305.

Menges, W. (2010, December 21) Heavy new traffic fines now in force. The Namibian,p.5. Retrieved: from http://www.namibia.com.na/index.php?id=74380&page=archive-read

Mohammadi, M., Imani,M., Tajari,F., Akbari, F., Rashedi, F., Ghasemi, A., Moghaddam, A.(2012). Human and vehicle factors in motor vehicle crashes and severity of related injuries in South East Iran. *Health Scope*, 1(2): 61-65. , DOI: 10.5812/jhs.6838

Mohan, D., Tiwari, G., Meleckidzedeck, K., & Fredrick, M. N. Road traffic injury prevention training manual. 2006. *Geneva: World Health Organization and Indian Institute of Technlogy Delhi Google Scholar*.

Motor Vehicle Accident Fund (2015) *2014 crash and claim report* Windhoek. MVAFund

Motor Vehicle Accident Fund (2016) *2015 crash and claim report* Windhoek. MVAFund

Motor Vehicle Accident Fund Act, No. 10. (2007) Retrieved from: http://lac.org.na/laws/2007/3970.PDF

Mphela, T. (2011). The impact of traffic law enforcement on road accident fatalities in Botswana. *Journal of Transport and Supply Chain Management*, *5*(1), 264-277.

Naci, H., Chisholm, D., & Baker, T. D. (2009). Distribution of road traffic deaths by road user group: a global comparison. *Injury prevention*, *15*(1), 55-59.

Namibia statistics Agency (2012, December). *Namibia 2011 Population and Housing Census : main report.* Retrieved from: http//cms.my.na/assets/documents/p19dmn58guram30ttun89rdrp1.pdf

Nussbaum, M.E., Elsadat,S., & Khago A. H.(2007) Best Practice in Analyzing Count Data: *Poisson Regression* (printing house) P. 306-323

Nantulya, V. M., & Reich, M. R. (2003). The neglected epidemic: road traffic injuries in Developing countries. *Bmj*, *324*(7346), 1139-1141.

Novoa, M. A., Perez, K. Santamarina-Rubio, E., Mari-Dell'Olmo, M., Ferrando, J., Peiro, R., Tobias, A., Zori, P., & Borrell, C. (2010). The Impact of the Penalty Points System on road traffic injuries inSpain: A time series study. *American Journal of Public health.* 100(11): 2220-2227 DOI: 10.2105/ajph.2010.192104

Paulozzi, L. J., Ryan, G. W., Espitia-Hardeman, V. E. & XI, Y. (2007) Economic development's Effect on Road Transport-related Mortality among Different Types of Road Users: A Cross-sectional International Study. *Accident Analysis and Prevention*, 39(1): 606-617.

Reynolds, C. C., Harris, M. A., Teschke, K., Cripton, P. A., & Winters, M. (2009). The impact of transportation infrastructure on bicycling injuries and crashes: a review of the literature. *Environmental health*, *8*(1), 47.

Ridout, M., Demétrio, C. G., & Hinde, J. (1998, December). Models for count data with many zeros. In *Proceedings of the XIXth international biometric conference* (Vol. 19, pp. 179-192).

Rodriguez G. (2013) November 6, Models for count data with over-dispersion.

Saffari, S. E., Adnan, R., & Greene, W. (2012). Hurdle negative binomial regression model with right censored count data. *SORT*, *36*(2), 181-194.

Schneider, R. J., Grembek, O., & Braughton, M. (2012, November). Pedestrian crash risk on boundary roadways: A University campus case study. In *Transportation Research Board 92nd Annual Meeting*.

Skrivanek, S. (2009). The use of Dummy Variables in regression Analysis. *More Steam*. Retrieved from: https://www.moresteam.com/whitepapers/download/dummy-variables.pdf

Shope, J. T., & Bingham, C. R. (2008). Teen driving: motor-vehicle crashes and factors that contribute. *American Journal of Preventive Medicine*, *35*(3), S261-S271.

Sukhai, A., Jones, A. P., Love, B. S., & Haynes, R. (2011). Temporal variations in road traffic fatalities in South Africa. *Accident Analysis & Prevention*, *43*(1), 421-428.

Sullivan, J. M., & Flannagan, M. J. (2007). Determining the potential safety benefit of improved lighting in three pedestrian crash scenarios. Accident Analysis & Prevention, 39(3), 638-647.

Tang, W., He, H., & Tu, X. M. (2012). *Applied categorical and count data analysis*. CRC Press.

Tjihenuna, T.(2015, June 24). Road accident cost N$1 million a day. *The Namibia*. Retrieved from https://www.namibian.com.na

Uugwanga , M. (2016, March 3). Accident deaths to surpass HIV. *Informante*. Retrieved from: https://www.informante.web.na/accidents-deaths-supass-hiv .

Wang, K., Yau, K.K., Lee, A. H., McLachlan, G.J. (2007). Multilevel survival modelling of recurrent urinary tract infection. *Pubmed*, 87(3), 225-229.

Wielenga, T. J. (1999). *A Method for Reducing On-Road Rollovers--Anti-Rollover Braking* (No. 1999-01-0123). SAE Technical Paper.

Wilson, P. (2015). The misuse of the Vuong test for non-nested models to test for zero-inflation. *Economics Letters*, *127*, 51-53.

Wood, J. M., Tyrrell, R. A., Chaparro, A., Marszalek, R. P., Carberry, T. P., & Chu, B. S. (2012). Even moderate visual impairments degrade drivers' ability to see pedestrians at night. *Investigative ophthalmology & visual science*, 53(6), 2586-2592.

World Health Organization (2009) *Global status report on road safety: Time for action.* Geneva: World Health Organisation

World Health Organization (2011) *Global Plan for the Decade of Action for Road Safety 2011-2020*. Geneva: World Health Organization.

World health organization and the world bank (2004) *World report on road traffic injury prevention* Who.int/violence_injury_prevention/publication/road_traffic/world report/ chapter http://who.int/violence_injury_prevention/road_traffic/activities/roadsafety_training_manual_unit_2.pdf?ua=1

World Health Organization. (2013). *Global Status Report on Road safety*. WHO.

World Health Organization (2013a) Global Status Report on Road Safey 2013 Supporting a Decade of Action. Geneva: World Health Organisation.

World Health Organization. (2015). *Global Status Report on Road safety*. WHO.

Yesilova, A., Kaydan, M. B., & Kaya, Y. (2010). Modeling insect-egg data with excess zeros using zero-inflated regression models. *Hacettepe Journal of Mathematics and Statistics*, *39*(2), 273-282.

Zaninotto, P., & Falaschetti, E. (2010). Comparison of methods for modelling a count outcome with excess zeros: application to Activities of Daily Living (ADL-s). *Journal of Epidemiology & Community Health*, jech-2008.Khan, A., Ullah, S., & Nitz, J. (2011). Statistical modelling of falls count data with excess zeros. *Injury prevention*, *17*(4), 266-270.

Zeileis, A., Kleiber, C., & Jackman, S. (2008). Regression models for count data in R. *Journal of statistical software*, *27*(8), 1-25.

Zheng, T., Qu, W., Zhang, K., Ge,Y. (2016) The relationship between attentional bias toward safety and driving behavior. *Accident Analysis Prevention,* 22(8),96, doi: 10.1016/j.aap.2016.07.034

Zou, Y. (2012). *Over- and under-dispersed crash data: comparing the Conway-Maxwell-Poisson and double-Poisson distributions* (Published Master thesis) Texas A&M University, Texas, USA.

ANNEXURE A: SPSS data cleaning code

```
DATASET ACTIVATE DataSet1.
FREQUENCIES VARIABLES=DateOccurred Yearofoccurrence Month DayOfCrash
CrashTime Region CrashType
   CrashCause CrashDescription @#PersonsInvolved @#PersonsInjured Fatality
@#VehiclesInvolved


#renaming variable names so they have one exact identical name, many are the same
with just different characters#
IF (CrashType EQ 'collision with animal')    CrashType ='Collision with Animal'.
IF (CrashType EQ 'Collision with animal')    CrashType ='Collision with Animal'.
IF (CrashType EQ 'Collision with other vehicle')     CrashType ='Collision with other
vehicles'.
IF (CrashType EQ 'Motor cycle/bicycle')      CrashType ='Motor cycle/Bicycle'.
IF (CrashType EQ 'other')      CrashType ='Other'.
IF (CrashType EQ 'unknown')          CrashType ='Unknown'.
IF (CrashCause  EQ   'Author of own misfortune')CrashCause=      'Reckless and
Negligent Driving'.
IF (CrashCause  EQ   'Collision with Animal          ')CrashCause= 'Animal'.
IF (CrashCause  EQ   'Animal')CrashCause=          'Animal'.
IF (CrashCause  EQ   'Collision with other vehicles ')CrashCause= 'Reckless and
Negligent Driving'.
IF (CrashCause  EQ   'Coollision with other vehicles          ')CrashCause= 'Reckless and
Negligent Driving'.
```

```
IF (CrashCause  EQ   'Cyclist          ')CrashCause= 'Motor cycle/bicycle'.
IF (CrashCause  EQ   'Driving without licence        ')CrashCause= 'Reckless and Negligent Driving'.
IF (CrashCause  EQ   'Failed to Indicate       ')CrashCause= 'Reckless and Negligent Driving'.
IF (CrashCause  EQ   'Falling Object')CrashCause= 'Reckless and Negligent Driving'.
IF (CrashCause  EQ   'Fell from moving vehicle       ')CrashCause= 'Fell from moving Vehicle'.
IF (CrashCause  EQ   'Ignoring road signs    ')CrashCause= 'Reckless and Negligent Driving'.
IF (CrashCause  EQ   'Intoxicated     ')CrashCause= 'Intoxicated'.
IF (CrashCause  EQ   'Jumped from moving vehicle')CrashCause= 'Fell from moving Vehicle'.
IF (CrashCause  EQ   'Lost Control (not overturn)    ')CrashCause= 'Reckless and Negligent Driving'.
IF (CrashCause  EQ   'Mechanical Failure    ')CrashCause= 'Mechanical Failure'.
IF (CrashCause  EQ   'motor cycle/bicycle    ')CrashCause= 'Motor cycle/bicycle'.
IF (CrashCause  EQ   'Motor cycle/bicycle    ')CrashCause= 'Motor cycle/bicycle'.
IF (CrashCause  EQ   'MotorCyclist  ')CrashCause= 'Motor cycle/bicycle'.
IF (CrashCause  EQ   'No Licence,Speeding,Intoxicate        ')CrashCause= 'Speed'.
IF (CrashCause  EQ   'not adhering to trafic rules     ')CrashCause= 'Reckless and Negligent Driving'.
IF (CrashCause  EQ   'Other   ')CrashCause= 'Other'.
IF (CrashCause  EQ   'Other/Unknown(specify)       ')CrashCause= 'Other'.
IF (CrashCause  EQ   'Overloading   ')CrashCause= 'Overloading'.
IF (CrashCause  EQ   'Overtaking     ')CrashCause= 'Speed'.
IF (CrashCause  EQ   'Pedestrian     ')CrashCause= 'Pedestrian'.
IF (CrashCause  EQ   'Poor Road Conditions/Design          ')CrashCause= 'Poor road
design'.
IF (CrashCause  EQ   'Poor visibility ')CrashCause= 'Poor Visibility'.
IF (CrashCause  EQ   'Quad bike      ')CrashCause= 'Reckless and Negligent Driving'.
IF (CrashCause  EQ   'Ran Accross the Road          ')CrashCause= 'Reckless and
Negligent Driving'.
IF (CrashCause  EQ   'Reckless and Negligent Driving       ')CrashCause= 'Reckless and
Negligent Driving'.
IF (CrashCause  EQ   'Roll over       ')CrashCause= 'Speed'.
IF (CrashCause  EQ   'Single vehicle overturned      ')CrashCause= 'Speed'.
IF (CrashCause  EQ   'Speedeing      ')CrashCause= 'Speed'.
IF (CrashCause  EQ   'Speeding       ')CrashCause= 'Speed'.
IF (CrashCause  EQ   'Sudden Mechanical Failure   ')CrashCause= 'Mechanical Failure'.
IF (CrashCause  EQ   'Tire burst      ')CrashCause= 'Tyre Burst'.
IF (CrashCause  EQ   'Tyre Burst      ')CrashCause= 'Tyre Burst'.
IF (CrashCause  EQ   'Under Investigation   ')CrashCause= 'Unknown'.
IF (CrashCause  EQ   'Unknown        ')CrashCause= 'Unknown'.
IF (CrashCause  EQ   'Weather condition     ')CrashCause= 'Poor Weather condition'.
IF (CrashCause  EQ   'Weather Condition    ')CrashCause= 'Poor Weather condition'.
IF (CrashCause  EQ   'Weight Shifting         ')CrashCause= 'Overloading'.
IF (CrashCause  EQ   'With Animal(Domestic)        ')CrashCause= 'Animal'.
IF (CrashCause  EQ    'With Animal(Wild)    ')CrashCause= 'Animal'.
IF (CrashCause  EQ    'With fixed object(specify)     ')CrashCause= 'Reckless and
Negligent Driving'.
IF (CrashCause  EQ    'With pedestrian          ')CrashCause= 'Pedestrian'.
EXECUTE
```

#Grouping the time and renaming the variable CrashTime to G_time, the time variable appeared in the data as single time e,g 13:30#

```
IF  (CrashTime  >= TIME.HMs(0,00,00) & CrashTime <= TIME.HMs(1,59,00))
G_time=1.
IF  (CrashTime  >= TIME.HMs(2,00,00) & CrashTime <= TIME.HMs(3,59,00))
G_time=2.
IF  (CrashTime  >= TIME.HMs(4,00,00) & CrashTime <= TIME.HMs(5,59,00))
G_time=3.
IF  (CrashTime  >= TIME.HMs(6,00,00) & CrashTime <= TIME.HMs(7,59,00))
G_time=4.
IF  (CrashTime  >= TIME.HMs(8,00,00) & CrashTime <= TIME.HMs(9,59,00))
G_time=5.
IF  (CrashTime  >= TIME.HMs(10,00,00) & CrashTime <=
TIME.HMs(11,59,00))G_time=6.
IF  (CrashTime  >= TIME.HMs(12,00,00) & CrashTime <= TIME.HMs(13,59,00))
G_time=7.
IF  (CrashTime  >= TIME.HMs(14,00,00) & CrashTime <= TIME.HMs(15,59,00))
G_time=8.
```

```
IF  (CrashTime  >= TIME.HMs(16,00,00) & CrashTime <= TIME.HMs(17,59,00))
G_time=9.
IF  (CrashTime  >= TIME.HMs(18,00,00) & CrashTime <= TIME.HMs(19,59,00))
G_time=10.
IF  (CrashTime  >= TIME.HMs(20,00,00) & CrashTime <= TIME.HMs(21,59,00))
G_time=11.
IF  (CrashTime  >= TIME.HMs(22,00,00) & CrashTime <= TIME.HMs(23,59,00))
G_time=12.

EXECUTE
## Recoding variables into zeros and 1s
SPSSINC CREATE DUMMIES VARIABLE=MonthB
ROOTNAME1=MonthAB ROOTNAME2=Holiday
/OPTIONS ORDER=A USEVALUELABELS=NO USEML=YES OMITFIRST=NO
MACRONAME1="Month AB" MACRONAME2="nonHoliday".

DATASET ACTIVATE DataSet1.
RECODE Month ('January'=1) ('April'=1) ('May'=1) ('August'=1) ('December'=1)
(ELSE=0) INTO MonthB.
VARIABLE LABELS  MonthB 'MonthB'.
EXECUTE.
RECODE DayOfCrash ('Friday'=1) ('Saturday'=1) ('Sunday'=1) (ELSE=0) INTO DayB.
VARIABLE LABELS  DayB 'DayB'.
EXECUTE.
```

```
RECODE Region ('Kavango'=1) ('Ohangwena'=1) ('Omusati'=1) ('Oshana'=1)
('Oshikoto'=1)
   ('Otjozondjupa'=1) (ELSE=0) INTO RegionB.
VARIABLE LABELS  RegionB 'RegionB'.
EXECUTE.

RECODE Region ('Kavango'=1) ('Ohangwena'=1) ('Omusati'=1) ('Oshana'=1)
('Oshikoto'=1)
   ('Otjozondjupa'=1) (ELSE=0) INTO RegionB.
VARIABLE LABELS  RegionB 'RegionB'.
EXECUTE.

RECODE CrashType ('Collision with other vehicles'=1) ('Roll over'=1) ('Fell from
moving vehicle'=1)
   (ELSE=0) INTO CrashTypeB.
VARIABLE LABELS  CrashTypeB 'Crash TypeB'.
EXECUTE.

RECODE CrashCause ('Speed'=1) ('Intoxicated'=1) ('Reckless and Negligent
Driving'=1) (ELSE=0) INTO
   CauseB.
VARIABLE LABELS  CauseB 'Cause B'.
EXECUTE.
```

```
RECODE Time (4=1) (7=1) (9=1) (ELSE=0) INTO TimeB.

VARIABLE LABELS  TimeB 'Time B'.

EXECUTE.


RECODE VehiclesInvolved (1=1) (ELSE=0) INTO VehicleB.

VARIABLE LABELS  VehicleB 'Vehicle B'.

EXECUTE.


RECODE Fatality (0=0) (ELSE=1) INTO FatalityB.

VARIABLE LABELS  FatalityB 'Fatality B'.

EXECUTE.
```

ANNEXURE B: R code

```
> library(foreign) # intructing R that I would like to inport a file  from a foreign package
> DatasetB<-data.frame(read.spss("DatasetB.sav")) # calling the dataset from the file location to R, and giving it a a name in R. the dataset and the R workspase should be in the same directory
> DatasetB = read.spss("C:\\My Stuff\\2017 Thesis\\Analysis\\Dataset1.sav") # the location where the dataset is, also were R is
> names (DatasetB) # this is just to call the names of variables in R to make sure that the data has imported correctly
> install.packages (MASS) # loading the package in R that is responsible for analysing GLMs
>library (MASS) # calling the package
>Model1<-glm(PersonsInjured~MonthB+DayB+RegionB+CrashTypeB+CauseB+TimeB+VehicleB+FatalityB,data=DatasetB, family=poisson) # the Poisson regression model
> summary(Model1)

>Model2<-glm.nb(PersonsInjured~MonthB+DayB+RegionB+CrashTypeB+CauseB+TimeB+VehicleB+FatalityB,data=DatasetB) #  The   negative binomial models
> summary (Model2)
```

```
>Model3<-
zeroinfl(PersonsInjured~MonthB+DayB+RegionB+CrashTypeB+CauseB+TimeB+Vehi
cleB+FatalityB,data=DatasetB, dist="poisson") # the  Zero- Inflated Poisson Model
> summary(Model3)

>Model4<-
zeroinfl(PersonsInjured~MonthB+DayB+RegionB+CrashTypeB+CauseB+TimeB+Vehi
cleB+FatalityB,data=DatasetB, dist="negbin") # Zero- inflated  Negative  Binomial
> summary(Model4)

>Model5<-
hurdle(PersonsInjured~MonthB+DayB+RegionB+CrashTypeB+CauseB+TimeB+Vehic
leB+FatalityB,data=DatasetB, dist="poisson")  # Hurdle Poisson Model
> summary(Model5)

>Model6<-
hurdle(PersonsInjured~MonthB+DayB+RegionB+CrashTypeB+CauseB+TimeB+Vehic
leB+FatalityB,data=DatasetB, dist="negbin") #  Hurdle  Negative Binomial
> summary(Model6)

> AIC(Model1,Model2,Model3,Model4,Model5,Model6)
> BIC(Model1,Model2,Model3,Model4,Model5,Model6)
> residuals(Model6)
> par(mfrow=c(2,2))
> plot(residuals(Model6))

> plot(fitted(Model6))

> plot(residuals(Model6))~plot(fitted(Model6))

> plot(residuals(Model6)~fitted(Model6))

> qqplot(residuals(Model6, main = "Q-Q plot of Residuals"))
```