STATISTICAL ANALYSIS OF ROAD ACCIDENTS FATALITY IN GHANA USING POISSON REGRESSION

By

Oppong, Richard Asumadu

A Thesis submitted to the Department of Mathematics, Kwame Nkrumah University of Science and Technology in partial fulfillment of the requirements for the degree of

MASTER OF PHILOSOPHY

COLLEGE OF SCIENCE

NOVEMBER, 2012

DECLARATION

I hereby declare that this submission is my own work towards the award of the Master of Philosophy degree and that, to the best of my knowledge, it contains no material previously used or published by another person or material which had been accepted for the award of similar or any other degree of the university, except where due acknowledgement had been made in the text.

Oppong Richard Asumadu	………………………	………………………
(PG5071510)		
Student Name & ID	Signature	Date

Certified by:

Mr. F.K Darkwah	………………………	………………………
Supervisor's Name	Signature	Date

Certified by:

Mr. F.K Darkwah	………………………	………………………
Head of Dept. Name	Signature	Date

ACKNOWLEDGMENT

I wish to express my sincere gratitude to the Almighty God for His protection I have enjoyed throughout my life, most especially for sailing me through this program successfully.

I would like to express my heartfelt appreciation to my supervisor, Mr. F.K. Darkwah, who read, scrutinized and provided the necessary support, encouragement and guidance which propel me to complete this research.

To Dr.Kofi Z. Batse, Dr. Mrs. A. O. Adebanji, Dr. S. K. Amponsah and Nana Kena Frimpong, I say thank you very much.

I appreciate immense contributions of Kofi Agygarko Ababio, Eric Nimako Aidoo, Owusu Foster, Abdul Wahab and Bashiru Imoro Ibn Saeed in my quest to fulfill my academic dreams.

To friends and loved ones who contributed in diverse ways to my success in this program, I say a big thank you and may God richly bless you.

Finally, my thanks go to my family, for their prayers, support and inspirations throughout this program of study.

DEDICATION

This thesis is dedicated to the Almighty God, the Oppong family of Nkukua Buoho, my supervisor, Mr. F.K Darkwah, and anybody doing anything to reduce road accidents in Africa.

ABSTRACT

Road accident in this country is known to be the second major cause of death after malaria and it is reported that there is an average of 1909 people who are killed by road accidents annually. The most dangerous part of it is that about 60% of these people who die through road accidents are within the ages of 16 and 45 years-the labour force of the country. The ultimate goal of this thesis is to use Poisson regression to fit a model to the secondary data which was obtained from the Building and Road Research Institute of the Council for Scientific and Industrial Research on the number of people killed by road accidents in Ghana from 2001-2010, given the type of vehicle which was involved in the accident, the ages of those who were killed, the day that the accident which killed the people occurred and time (in years). The results of the Poisson analysis showed that there was over dispersion in the data. Negative binomial regression analysis was therefore used to validate the Poisson regression model. It was clear that the negative binomial regression model was the best fit for the data. It was observed from the result that in general the number of people killed in road accidents gradually increases with time (as the years go by, the number of people killed by road accidents increases). The type of vehicle involved in an accident was found to be associated with the number of people expected to be killed in the accident with cars and buses as the type that killed most people and heavy duty vehicles such as bulldozer and trucks which were classified as “others” killed the least people in accidents. The day an accident occurred affected the expected number of people killed in that accident because it was identified that Saturday had the highest number of people killed in road accidents. Finally, the age of a person involved in road accident could determine as to whether one would be killed. The results showed that people in the age group of 16-25 were mostly killed in road accidents.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

1.0 Introduction

Road accidents in Ghana have been identified as one of the major causes of deaths in the country. It is classified as the second major cause of death in the country following malaria. According to the road traffic crashes in Ghana statistics for the year 2009 by Building and Road Research Institute, shows that there were 12,299 road accidents for the year 2009, Afukaar et al (2009). There were total of 18,496 casualties with 2,237 of them losing their lives, while 6,242 sustained serious injuries. This reveals that there was an average of 6 deaths everyday in Ghana which was caused by road accidents. The most dangerous part of it all is that most of the people who are killed by road accidents are those in the age group that constitute the work force of this nation. It is in this regard that more attention needs to be placed on the research into road accident and its impact on human lives and properties in Ghana.

1.1 Background of The Study

Vehicular accident in this country has become one of the growing concerns to most Ghanaians in recent times. This is as a result of the tremendous effect of accidents on human lives, properties and the environment. Many researchers have come out with the causes, effects and recommendations to vehicular accidents. These causes include drink driving, machine failure and over speeding, Sagberg, Fosser, and Saetermo(1997), National Road Safety Commission (2009) and Adams (1982). Yet every year the road safety commission, Ghana Statistical Service and other organizations would report an increase in vehicular accidents; Annual Report, National Road Safety Commission, Ghana (2009). The mere increase in the number of accidents is not enough for one to conclude that really there is an increase in vehicular accidents; hence the need to analyze the accidents data statistically to check whether there is any evidence of increasing road accidents as years go by resulting to large number of people losing their lives.

This research will consider road accidents and its impact on human lives and properties in the entire ten regions of Ghana. It is estimated that the population of Ghana stands at 24,223,431and there are over 1,030,000 registered vehicles which ply the various road network in the country. Apart from the Ghanaian registered vehicles there are other vehicles from the neighboring countries which moves in and out of the Ghana due to the various ports we have in the country. There are not less than 10,000 recorded road accidents annually in Ghana. These accidents cause over 1,600 people to lose their lives and more than 15,000 getting injured.

Researchers have been modeling vehicular accidents with crash prevention models in various parts of the world. However, it is extremely difficult to just apply models which have worked somewhere to data obtained from different country due to the variations in the various factors pertaining in different countries, Fletcher et al (2006).

There has not been much statistical research in the field of road accident in Ghana. This might have been as a result of the inadequate information available on road accidents and its impact on human live and properties in the country.

These road accidents have killed a lot of people in this country and as such it is described as one of the major causes of death in Ghana. The causes of death of casualties in road accidents have been associated with secondary collision, improper handling of casualties and inadequate emergency services in the country.

The Building and Road Research Institute of the Council for Scientific and Industrial Research has being publishing the descriptive statistics of road crashes in Ghana for the past ten years.

In these documents are certain factors which could be possible contributors of the death of casualties in road accidents in Ghana. These factors include location of the accident, the age and sex of the casualty, the type of vehicle involved in the accident, nature of the road, weather conditions, the days of the week, the time of the day, and the ownership of the vehicle involved in the accident.

Salifu (2004) has developed a forecasting model for traffic crashes for unsignalised urban junction, Afukaar and Debrah (2007) have also model traffic crashes for signalized urban junction in Ghana and Ackaah (2011) has modeled traffic crashes on rural highways in the Ashanti region.

It is however surprising that in spite of the numerous factors identified by researches as the causes of road accidents in Ghana and its consequences on human lives and properties; nobody has modeled the causes of road accidents and its contributions to the death and survival of casualties in Ghana to authenticate the contributions of each of these factors to the casualties' death. Furthermore, there has not been any work on the likelihood of a casualty surviving in road accident or the possibility of a casualty surviving in a particular type of vehicle getting involved in a road accident.

Many researchers prefer the use of traffic accident to road accident but for the purposes this thesis the two would be used interchangeably.

Road accident is defined as any activity which distracts the normal trajectory of a moving vehicle(s), in a manner that causes instability in the free flow of the vehicle. The accident is that the vehicle(s) involved veer(s) off the road, collide, run over, vehicle on fire, etc.

The national road safety commission describes casualties as persons killed, seriously injured and slightly injured. The word casualty in this thesis shall be referred to any person involved in road accident. This will be grouped as casualties who survived and those who are killed in the road accident.

Vehicular accident has always been attributed to human errors such as high alcoholic content in the blood stream of the driver, over speeding, wrong overtaken among others. It has also been linked to poor road network, poor surfacing of the roads, witchcraft and the death-dying nature of some of the vehicles which ply the roads. There are numerous suggested solutions, various interventions by government, nongovernmental organizations and other road stakeholders to curtail road accident and its replica effects on human lives and properties, it could be possible that these factors such as type and nature of the road, age and sex of casualty contribute casualty survival in road accidents in the country and have still not been considered. It is in view of this that this research seeks to identify if there is any relationship between the casualty survival in road accident and these factors.

It is therefore important to statistically analyze the accident data to ascertain the truth or otherwise of these possibilities.

When it is confirmed that there is a relationship, then it will be prudent to apply some mathematical and statistical models such as Poisson regression and/or negative binomials to fit a model to the accident data for better prediction for decision making.

1.2 Problem Statement

The road accidents in this country seem to be in ascendency. These accidents have been categorized by the National Road Safety Commission as fatal, serious and minor.

This classification is based on the extent of damage to human lives and properties.

The root causes of road accidents and its effects on human lives and properties have been associated with human errors and superstition. The purpose of this thesis is to find out the contributions of other factors such as types of vehicles which ply the road, road description and surface type, days of the week in which the accident occurred, the age of the casualties in the road accidents.

1.3 Objective of The Research

This research seeks to achieve the following objectives;

1. To perform descriptive analysis of the data.
2. To model road accidents fatality in Ghana using Poisson regression.
3. To validate the models with negative binomial regression.

1.4 Methodology

The research combined both quantitative and categorical data for the analysis. Secondary data was taken from the Council for Scientific and Industrial Research; Building and Road Research Institute. Poisson regression and/or negative binomial regression would be the main tools for data analysis. Data analysis would be done using Microsoft Excel Minitab and R statistical packages.

1.5 Significance of The Study

It is imperative to research into the vehicular accident data in Ghana to come out with the reality on the ground so that;

1. Policy makers could come out with strategies to reduce the numerous deaths caused by vehicular accidents to the barest minimum in the country.
2. To fit a model to accident data in the country for better prediction of number of people who are killed in road accidents in order to plan for future occurrence.
3. To create a platform for future studies into vehicular accident and its effects on human lives and properties in Ghana.

1.6 Organisation of The Thesis

This thesis contains five main chapters but prelude to these chapters are the Abstract which summarizes the whole research, the Table of content, List of Figures, List of Abbreviations, Dedications and Acknowledgement.

The references and the appendixes could also be found after the last chapter

Chapter one of this thesis contains the Background Information; the objectives of the study, organization of the thesis, the limitations and the scope of the research. Chapter Two is made up of the Literature Review, Chapter Three discusses the various methods used for the research, and Chapter Four presents the Analysis and modeling of the data and the discussions of findings, and the last Chapter deals with summary of finding, conclusions and recommendations.

1.7 Limitations and The Scope of The Research

The availability of road accident data in this country is usually difficult to come by and even if it obtained, the information on it is normally scanty. It is believed that not all accidents are reported to the police for records to be made on them due the human nature and Ghanaian hospitality attitude. Also, it is possible that the police might not have filled the accident report form for all accidents which might have been reported to them. It is therefore imperative to admit that the data provided by Building and Road Research Institute might be under recorded. However, there is enough evidence from the various researchers who have used road accident data from Building and Road Research Institute that their data is reliable and representative.

CHAPTER 2

LITERATURE REVIEW

2.0 Introduction

This aspect of the study reviews the various literatures related to the topic under consideration in order to uncover critical facts and findings which have already been identified by previous researchers and numerous studies in and around the causes, effects and economic implications of road accidents with particular reference to causes of deaths of road accident casualties. It will also assess the existing mathematical and statistical models for casualty survival in road accidents using Poisson and Negative Binomial regression and their applications in the real life situations. This will guide the direction of this study and aid in exploring the unknown.

2.1 The State of Road Accident

Vehicular accident in this country has become one of the growing concerns to most Ghanaians in recent times. This is as a result of the tremendous effect of road accidents on human lives, properties and the environment.

Heidi (2006) reported that 1.2 million people in the world lose their lives through road accidents every year. This number has rising to 1.3 million people who lose their lives globally every year and between 20 and 50 million people sustain various forms of injuries annually as a result of road accident. The most affected of these consequences of road accidents is the people in the age bracket of 15 and 29. Road accidents cost the world an amount of US$518 billion annually. It is estimated that if nothing is done globally to curtail the rampant nature of road accidents and most especially the causes of deaths of casualties before they are sent to hospitals then by the year

2020, 1.9 million people will be killed by road accidents in the world, (World Health Organisation, 2011).

A research conducted by Salim and Salimah (2005) also indicated that road accident was the ninth major cause of death in low-middle income countries and predicted that road accident was going to be the third major cause of deaths in these countries by 2020 if the trend of vehicular accident was to be allowed to continue.

Media reports reveal that there is a high road accident in Ghana, when compared with other developing countries. In 2001, Ghana was ranked as the second highest road accident-prone nation among six West African countries with 73 deaths per 1000 accidents, (Akongbota, 2011).

The Ghanaian Times news paper reported on the 16^{th} day of November, 2011 that a total of 1,986 lives were lost in the country through road accidents from January to October, 2011, The Ghanaian Times (16^{th} November, 2011).

2.2 Causes of Road Accident

Many researchers have come out with the causes, effects and recommendations to vehicular accidents in Ghana and elsewhere. For instance, Ayeboo (2009), identified that the numerous accidents on our road networks have been linked to various causes which include over speeding, drink driving, wrong over taking, poor road network and the rickety vehicles which ply on our roads.

Furthermore, the National Road Safety Commission (NRSC) has identified over twenty causes of road accidents in Ghana which include unnecessary speeding, lack of proper judgment of drivers, inadequate experience, carelessness, wrong overtaking, recklessness, intoxication, over loading, machine failure, dazzling and defective light, boredom, unwillingness to alight from motion objects (vehicles, motor cycles, human being and uncontrolled animals), skid and road surface defect, level crossing and obstruction. Other factors are inadequate enforcement of road laws and traffic regulations, use of mobile phones when driving, failure to buckle the seat belt and corruption, (National Road Safety Commission, 2007).

In spite of all these factors, Ocansey (2011) observed that poor vision of drivers could also be a major contributory factor to road accidents. It was obvious that the actual factors which may be influencing the traffic crashes in Ghana have not been identified since most of the factors stated above have not yet been tested with any mathematical and statistical tool to ascertain the truth or otherwise of their contributions.

Elsewhere, the causes of road accidents have also been linked to one or combination of the following four factors, equipment failure, road design, drivers' behavior and poor road maintenance. However, studies have shown that over 95% of all road crashes are caused by the behavior of the driver and the combination of one or more of the other three factors, (Driving guidelines, no date).

According to the country report on Road Safety in Cambodia, road accident is caused by human factors (road users), road defects and vehicle defects. It was found in the report that road accident in Cambodia was increased by 50% in five years while the fatality rate was doubled. To help reduce the rate of road accident it was suggested that Road accidents Safety Committee was set up, accident data system was established, accident evaluation policy and driver training measures were to be put in place, Ung Chun (2007).

In spite of all these facts, some Ghanaians still associate some of the road accidents in Ghana to superstitions, witchcraft and evil forces, are accidents caused by witches or irresponsible government policies? It is therefore believed that as a result of these spiritual activities, most people die in road accidents so that more blood would be obtained by the witches, wizards and the evil forces for their spiritual activities, Okyere (2006).

Some researchers have also attributed the escalating number of carnage on our roads especially in sub Saharan Africa to bribery and corruption. In a study conducted in Russia to find out the contribution of corruption to road toll, it was found out that people were paying as much as US800.00 to obtain driving license without going through any form of driving school ("Russia" Today,2010).

There is enough evidence in South Africa that the government uses over R500 million annually from the Road Safety Fund to fight fraud, bribery and corruption (Arrive Alive, n.d)

According to Chitere and Kibua (2004) the transport industry of Kenya is so much fragmented with the transport ministry, office of the president and other agencies playing conflicting roles which create bureaucracy, bribery and corruption in the industry since security personnel (police) fail to check and introduce transport laws.

Also research by Khayesi (1997) and Lamba et al (1986) shown that most workers of public transports are employed on the bases of relational ties. This practice has not given room to qualify and competent people to work in the transport industry leading to rampant road accidents in Kenya.

There is substantial evidence to prove that the higher the number of road accidents which occur in given time period, the higher the number of casualties who die in the accident.

According to Afukaar et al (2009) in their report presented to the National Road Safety Commission, there was a total of 11320 road accident which killed 1779 people in 2005. The number of road accidents increased to 12038 in 2007 and killed 2024 people. At the end of 2009, there were 12299 road accidents in the country and 2237 lives were lost. However, the report did not fit any model which could be used to estimate the likely road accidents in subsequent years, vis-à-vis the number of casualties who are likely to lose their lives in such accidents.

Interestingly, in a study conducted in South Delhi by Kumar et al (2008), it was found out that most fatal accidents occurred on Saturday but in a study at Nepal, the highest number of road accidents occurred on Sunday and the least number on Monday, Jha and Agrawal (2004).

Coincidentally, it was found in a study at South Africa that most people died through road accidents which occurred on Saturday (20.8%) followed by Sunday with 17.1%, (Injury Mortality Surveillance System, 2005)

Kumar et al (2008) identified November as the month with the highest number of fatal accidents in Delhi, 11.04% of all fatal accidents in Delhi occurred in November. This finding contradicted the result obtained in Nepal by Jha and Agrawal (2004) who suggested that July was the month in which most fatal accidents occurred in Nepal.

In a research conducted in Delhi by Mehta (1968) and Ghosh (1992) found that most people were killed in road accidents which occurred in January but National Crime Record Bureau (2005) reported higher incidence of road accidents with much victims in May and March in India.

These varying results from various researchers in different countries indicate that it will be difficult to use what prevail in one country to estimate for another country since conditions associated with road accidents may vary from country to country.

2.4 Causes of Casualty's Death in Road Accident

The cause of death of casualties has been associated with many factors such as secondary collision, failure of drivers and vehicle occupants to put on seat belt and riders failing to put on helmet, Afukaar et al (2009).

Studies have shown that sleep related accidents tend to be more severe and as such most people are killed. This situation is as a result of the driver's inability to prevent and stop certain actions such as applying the brakes before collision and steering onto the main road if the vehicle veers off the road. The research identified that in order to reduce the risk of drowsy driving and its related crashes, drivers are advised to have sufficient sleep, drivers are to avoid drinking especially when feeling sleepy and reduce driving between midnight and 6 :00 am, Strohl et al (1998).

Homes and Reyner (1995) suggested that due to the inactive nature of the sleeping driver to control the vehicle prior to the accident, sleep related accidents have high risk of death as compared with the other forms of road accidents. Furthermore, in a research conducted in the North Carolina, sleep related accident was found to the most severe accidents among all other types of road accidents, Allan et al (1995).

Also, Zomer (1990) identified that the number of casualties in sleep related road accidents is 50% higher than all accidents. It had three times fatalities and doubles the seriously injured as compared with non sleeping related road accidents. The sleep related accidents are normally more severe and kills a lot of people because there is no control on the part of those involved in the accident, particularly, the driver. In this vain, there are certain circumstances which might have been avoided to reduce the number of casualties but due to the driver's inability to control the vehicle the people suffer the consequences.

The age of the vehicle involved in an accident cannot be ruled out of the contributors when one is assessing the cause of death of casualties in road accidents. Broughton (2007) identified that when two vehicles collide, the driver and occupants of the older vehicle are usually at more risk of being killed than those in the newer vehicle. In that study, it was estimated that the mean risk of death of drivers of vehicles which were registered in 2000 to 2003 was less than half of the risk for the drivers of vehicles which were registered in 1998 to 1999, Broughton (2007).

This phenomenon may be due to weaker nature of the various parts of the older vehicles and probably the improvement and modernization in the manufacturing of newer vehicles.

However, casualty rate increases in collision with more modern cars on non-built-up roads where speeds are higher as compared with that of older cars.

The size of a vehicle has also been found to contribute to the death of road users in traffic crashes. From the findings of Broughton (2007) in his study into road accidents data from 2001 to 2005, it was revealed that the driver casualty rate increases with the size of the other vehicle in collision. The question now is, for the past 30 years, the weight and size of vehicles have been improved by 30% yet the number of casualties' deaths have not decrease in accordance with that. The fact still remains that people end up relying so much on the strength of their vehicles and take undue risk especially the youth, Broughton (2007).

Studies have shown that young drivers and young passengers die more in road traffic crashes than their older counterparts, Broughton (2005).

In a research conducted in Britain and Wales to assess the death pattern of various age groups and their sexes within the period of 2000 to 2002, it was found out that 40% of males and 30% of females' drivers who died in road accidents were in the age bracket of 16 and 19 years.

This number had risen to 44% for males and 38% for females by the end of 2005, (Department of Transport, 2006b).

However, it is interesting to note that this pattern changes with age, as the road users grow then the number of females who die through accidents become more than that of males. From 1994 to 2004, there were 13% deaths for men above 30 years and 30% for females in that same age group, (Department of Transport, 2006b). These sudden changes may be due to the fact that women tend to accumulate more fatigue than men as they grow. Further, women are known to travel more often than men at old age to visit their children and other relatives.

Also, Kumar et al (2008) found out in their research in South Delhi that with all the people who were killed in road accidents, 88.2% of them were males. This result actually confirmed the studies by earlier researchers as Salgado and Colombaje (1998), Shadev (1994), and Henriksson (2001), all of whom proposed and substantiated that more males are killed in road accidents than females.

Drink-driving is another factor which was identified by Clarke et al (2007) as a contributor to death of casualties in road accidents. The reason for this could be link to the inability of the drunk driver to control the vehicle as a result of sleeping, Zomer et al (1990).

Aside the drunk drivers, passengers and other road users who are drunk may even not be aware of what could be going on around them before, during and after the accident in order to take caution to avoid serious injuries and deaths in situation where they could have done so. In addition to this, when passengers are drunk, then it becomes extremely difficult for drivers to take their advice even if they are right. The end result is that drivers do their own things and end up causing accidents which kill people.

There were 1106 car drivers who were killed in road traffic crashes in 2005 at Britain and Wales and a study into this data by Clarke et al (2007) showed that 40% of those who died worn no seat belts and most of them were people between the age 17and 29 years. It was further identified that the desire for buckling the seat belt increases as one grows beyond 30 years, Clarke et al (2007).

Broughton and Walter (2007) also found out that drivers and vehicle occupants tend to avoid the use of seat belt in the night and as a result casualties' death in road accidents is higher in the night.

One of the commonest thing identified by researchers as the cause of death in road traffic crashes is anoxia-loss of oxygen supply –which cause a blockage in the air ways of the casualties and if immediate aid is not to the casualty, he/she dies after a short while due to inadequate supply of oxygen (British Red Cross, 1997).

Although, there are certain forms of accident which cannot be prevented, it is evidently true that pre-hospital death of road traffic crashes' victims can be prevented when timely and proper first aid measures are put in place, Redmond (1994).

Hussein and Redmond (1994) in their study conducted in Staffordshire in pre-hospital deaths in road accidents, they came out with the result that 39% to 85% are preventable and these deaths are caused by airway obstruction.

Studies have ascertained the medical assertion that for any accident, there is a 'golden hour' which exists for casualties after the accident. Within this period, road accident victims have a greater chance of surviving else they lose their lives, British Red Cross (1997). It is therefore imperative that immediate first aid is provided to road accident victims before they are rush to the hospital.

2.5 Road accident models.

The fatality rate over the years has been used to compare road accident incidence in large number of countries. Fatality rate is defined as the number of deaths which occurred through road accidents with respect to some measure of the use of road system. However, Fatality rate has been defined by several authors to suit the needs of their researches. Ghee et al (1997) stated that fatality rate is defined as the number of injury accidents occurring per annum per million vehicle kilometer travelled. But since there is no much reliable accident data base in developing countries and much information required to compute this type of fatality rate, Ghee et al (1997) defined the fatality rate for road accidents in a given country to be measured in respect of the number of persons killed through road accidents per 10,000 licensed vehicles in a country. As population increases and the number of licensed vehicles in developing countries are rising rapidly, Rajesh (2006) suggested that fatality rate is defined as number of road accident deaths per 100,000 licensed vehicles.

However, Jacobs (1980) suggested that this index cannot be used to compare accident fatality rates of different countries since the countries may vary in terms of population and total vehicles which ply their roads. He then proposed a model which assessed the relationship between fatalities, population and motorization of the country. This model supported the Smeed Formula for international comparisons of accident fatalities, Smeed (1938, 1968). Smeed in 1938 used accident data from different countries and proposed the formula $\frac{D}{N} = 0.0003\left(\frac{N}{P}\right)^{-0.67}$, where D is annual number of fatalities from road accidents, N is number of vehicles in use and P is population. This model was confirmed in 1968, Smeed (1968).

Silyanov (1973) used the idea of Smeed and modeled accident data from different countries and had similar results as that of Smeed except that there were some variations in the constant terms in the models. In all these models none of the researchers made an attempt of using the model to predict accident fatalities in the countries concerned but as basis for only comparison.

Also, this assertion of comparing the accident fatality rate of different countries using deaths via road accidents, population and number of registered vehicles may not be applicable to many countries, Andreassen (1985) such as Ghana because there are vehicles which ply our roads and are not registered at all or are registered in foreign countries' numbers.

For instance, vehicles from the nearby countries like Burkina Faso and Mali import most of their goods via Ghanaian ports and as a result of the trade relationship among the West Africa Sub-Region, vehicles from these countries ply our roads to convey their goods.

Furthermore, the population of a given country in a particular time is an estimate since census is done mostly in ten years intervals and the results of the census is not published in the same year it is conducted.

In order to obtain true and accurate model that can fit into accident data for comparison and prediction, other researchers upon the identification of flaws in previous models tried different means by including more factors in the accident data analysis, Livneh and Hakkert (1972) researched into road accidents in Israel using employment and population data. Susan and Partyka (1984) also modeled road accident using employment and population data.

Andreassen (1985) raised serious objection to the use of death per vehicles licensed in order to make international comparison of road accident fatalities because it was found out that the two parameters were not linearly related over time. He then came out with a general formula, $D = constant * N^{m1} * P^{m2}$, which could also be used to predict the number of deaths in road accidents.

Where *D* is the number of deaths in road accidents, *N* is number of vehicles in use, *P* is the population of the country and *m1, m2* are variables of interest. The difficulty in applying the equation by Andreassen is how to determine the constant term and the indices which might vary from country to country.

Time series analysis was used by Mekky (1985) to study the effect of rapid increase in the motorization levels on the rate of fatalities in some developing countries. Many researchers have dived into the investigation of traffic crash patterns in different countries in order to understand its relationship with the fatality rate of road accident. Among such researcher are Dinesh (1985), who investigated crash patterns in Delhi, Emanalo et al (1987) developed the trend curves for road accidents, casualties and other vital quantities in Zambia, Pramada and Sarkar (1993) studied the variations in the pattern of road accidents in various States and Union Territories of India, Johnson (1997) studied the change in the number of accidents between before-year and after-year, Thole'n (1999) and Velin et al (2002) have also compared the variations in the change in the number of accidents in all the control sites of public roads in the Region West Sweden which were not surfaced during the study period.

Tanner (1953) proposed a model of accidents that occurs in 3-way junction part of the road in a given period and concluded that it is approximately proportional to the root of the product of the two way major road traffic volume and the turning flows from the minor road.

In 1978, Hakkert and Mahalel (1978) in their investigations to estimate the number of accidents which occur at intersections of the road came out with the ‘’product of intersection flow’’ model. Also, Leong (1973) suggested a model which states that ‘’product of flow’’ each raised to the power less than one and this model was confirmed by Hauer et al (1989). McGuigan also investigated the root product flow model after Tanner (1953) and tested the sum of inflows relationships and concluded that the use of the root product flows' model better model fits accident data as compared with the sum of inflows model which cannot be justified universally.

However, Mountain and Fawaz (1996) identified that the sum of flow model has some inconsistencies in relation to its ability to predict more than zero accidents between conflicting streams of traffic even when one of inflows is zero. Furthermore, the sum of inflows' model has the possibility of predicting equal numbers of accident for a given value of total inflows with no regards to the distribution of flows between the major and minor arms of the junctions, Salifu (2004).

In correcting these errors associated with the sum of inflows model and the product of flows model, Salifu (2004) reported that the cross product of flows model, sum of crossing flow product and the sum of encounter flow product models produce much more better fit to accident data than the simple flow models such as the total junction traffic inflow model. He stated that the most influential traffic exposure model for X-junction is sum of crossing flow product model, Salifu (2004)

Jacobs and Cuttings (1986) investigated into the previous accident models developed by earlier researchers to improve upon them. A study was conducted to find the effects of speed limits on road accidents by Fieldwick (1987) who identified that speed limits have significant impact on road safety and severity in both rural and urban roads.

John and Adams (1987) assessed Smeed's law and provided more insight in the analysis of road accident data using Smeed's model. Minter (1987) also discussed the applications of two accident models which were developed by Wright and Towel for road safety problems and came out finally with a new model for estimating road accidents in United Kingdom.

Pramada and Sarkar (1997) used road length as an additional parameter and established a model for road accidents with the length of road covered by the vehicle as a factor. Jamal and Jamil (2001) presented a general model to predict road accident fatalities in Yemen. Pramada (2004) used road accident data to compare the models developed by Smeed and Andreassen and confirmed that the two models worked well.

2.6 Poisson Regression

Poisson regression analysis is a technique used to model dependent variables that describe count data (Cameron et al, 1998). It is often applied to study the occurrence of small number of counts as a function of a set of predictor variables, in experimental and observational study in many disciplines, including Economy, Demography, Psychology, Biology and Medicine (Gardener et al, 1995). The Poisson regression model may be used as an alternative to the Cox model for survival analysis, when hazard rates are approximately constant during the observation period and the risk of the event under study is small (e.g., incidence of road accidents).

Poisson regression model usually replaces Cox model, which cannot be easily applied to aggregated data. Furthermore, using rates from an external population selected as a referent, Poisson regression model has often been applied to estimate standardized mortality and incidence ratios in cohort studies and in ecological investigations (Breslow et al 1987).

Finally, some variants of the Poisson regression model have been proposed to take into account the extra-variability (over dispersion) observed in actual data, mainly due to the presence of spatial clusters or other sources of autocorrelation (Trivedi et al, 1998)

Besides medical studies, the Poisson regression model has been used in different fields of research, ranging from herd management assessment to animal health in domestic and wild animals and control of infectious diseases in different animal species. The Poisson model has also been applied to data analysis in a multidisciplinary study on cancer incidence in veterinary and other workers of veterinary industry compared to that of other part of active population in Sweden (Travier et al, 2003).

Miauo and Lum (1993) assessed the statistical nature of the two conventional linear regression model which have been used by most researchers to develop road accident relationships and found out that these two linear regression models fail to consider and describe the distributional characteristics of road accidents as in its randomness, non-negative nature of road accidents, the discrete and count properties of the event of road accident. Miauo (1994) compared Poisson and Negative Binomial regressions since they all cater for the distributional properties of accident data. Many other researchers have also assessed the use of linear regression models for road accident models and confirmed the limitations in such models, Kim et al (2005) used generalized log-linear models and Garber and Wu (2001) applied stochastic models in fitting models to road accidents data. However, one should be more cautious of the use the Poisson and the negative binomials since the estimation of the various parameters could be misleading, Miauo (1994).

In Ghana much work has not been done in this regard but researchers such as Salifu (2004) used the generalized linear model to predict road accidents in unsignalised urban junctions, Afukaar et al (2007) and Ackaah and Salifu (2011) applied Poisson, Negative Binomial and Log-Linear regression models to accident data in Ghana.

CHAPTER 3

METHODOLOGY

3.0 Introduction

This chapter deals with a detailed description of the methods used for this research and explained the theory behind the various distributions and models for the analysis. The chapter addressed the possible probability distributions of accident data (count data) and their likely regression models which may include the Poisson and Negative Binomial distributions, the Generalized Linear models such as Poisson and Negative-Binomial regression models and the mixed effect model. It also provided the description of the software packages used for the analysis and modeling.

3.1 Generalized Linear Models (GLM)

Generalized linear model (GLM) was first introduced by Nelder and Wedderburn (1972). It provided a unified framework to study various regression models, rather than a separate study for each individual regression. Generalized linear model (GLM) is an extension of the classical linear models. It includes linear regression models, analysis of variance models, Logistic regression models, Poisson regression models, Zero-inflated Poisson regression models, Negative Binomial regression models, log-linear models, as well as many other models. The above models share a number of unique properties, such as linearity and a common method for parameter estimation. A generalized linear model consists of three components:

1. A *random component*, which specify the conditional distribution of the response variable, Y_i, given the explanatory variables, x_{ij}

2. A linear function of the regression variables, called the *linear predictor*,

$$\eta_i = \alpha + \beta_1 X_{i1} + \cdots + \beta_\kappa X_{ik} = x_i^{'}\beta \qquad 3.1$$

on which the expected value μ_i of Y_i depends.

3. An invertible *link function*, $g(\mu_i) = \eta_i$ 3.2

   which transforms the expectation of the response to the linear predictor. The inverse of the link function is sometimes called the *mean function*:$g^{-1}(\eta_i) = \mu_i$ 3.3

For traditional linear models in which the random component consists of the assumption that the response variable follows the Normal distribution, the canonical link function is the identity link. The identity link specifies that the expected mean of the response variable is identical to the linear predictor, rather than to a non-linear function of the linear predictor.

That is, for the normal linear model, the link function $g(u_i) = u_i$ 3.4

The Generalized Linear Model is an extension of the Linear Model to include response variables that follow any probability distribution in the exponential family of distributions. Many commonly used distributions in the exponential family are the normal, binomial, Poisson, exponential, gamma and inverse Gaussian distributions. In addition, several other distributions are in the exponential family and they include the beta, multinomial, Dirichlet, and Pareto. There are other several distributions which are not in the exponential family but are used for statistical modeling and they include the student's *t* and uniform distributions.

3.2 The Exponential Family

GLMs may be used to model variables following distributions in the exponential family with probability density function $f(y;\theta,\varphi) = exp\left\{\frac{y\theta - b(\theta)}{a(\varphi)} + c(y;\varphi)\right\}$ 3.5

or,

$$\log \mathrm{f}(y;\theta,\varphi) = \frac{y\theta - b(\theta)}{a(\varphi)} + c(y;\varphi) \quad 3.6$$

Where φ is a dispersion parameter and $a(\varphi)$, $b(\theta)$ and $c(y;\varphi)$ are known functions.

For distributions in the exponential families, the conditional variance of Y is a function of the mean, μ together with a dispersion parameter, φ

That is, $E(Y_i) = \mu_i = b^{'}(\theta)$ 3.7

and $var(Y_i) = \sigma_i^2 = b^{''}(\theta)a(\varphi)$ 3.8

Where $b^{'}(\theta)$ and $b^{''}(\theta)$ are the first and second derivatives of $b(\theta)$. The dispersion parameter is usually fixed to one for some distributions.

3.3 The Link Function

In theory, link functions $\eta_i = g(\mu_i)$ can be any monotonic, differentiable function. In practice, only a small set of link functions are actually utilized. In particular, links are chosen such that the *inverse link* $\mu_i = g^{-1}(\eta_i)$ is easily computed, and so that g^{-1} maps from $X_i\beta = \eta_i \in \Theta$ into the set of admissible values for μ_i. A log link is usually used for the Poisson model, since while $\mu_i = g(\mu_i) \in \Theta$, because Y_i is a count, we have $\mu_i \in 0,1,\ldots$. For binomial data, the link function maps from $0 < \mu_i < 1$ to $\mu_i \in \Theta$.

The Table 3.1 shows some link functions in the exponential family with their expectations and variances.

Table 3.1 Exponential Family and their link functions

Distribution	**Link function** θ	**Canonical link** $a(\theta)$	**Dispersion** ϕ	**Expectation** $E(y)$	**Variance** $Var(\mu) = \frac{\text{var}(y)}{\phi}$
$B(n,\pi)$	$\ln \frac{\pi}{1-\pi}$	$n \ln(1+e^{\theta})$	1	$n\pi$	$n\pi(1-\pi)$
$P(\mu)$	$\ln \mu$	e^{θ}	1	μ	μ
$N(\mu,\sigma^2)$	μ	$\frac{1}{2}\theta^2$	σ^2	μ	1
$G(\mu,v)$	$\frac{-1}{\mu}$	$-\ln(-\theta)$	σ^2	μ	μ^2
$IG(\mu,\sigma^2)$	$\frac{-1}{2\mu^2}$	$-\sqrt{-2\theta}$	σ^2	μ	μ^3
$NB(\mu,k)$	$\ln \frac{k\mu}{1+k\mu}$	$-\frac{1}{k}\ln(1-ke^{\theta})$	1	μ	$\mu(1+k\mu)$

3.4 The Statistical Model

The canonical treatment of GLMs is from McCullagh and Nelder (1989), and this review closely follows their notation and approach.

Begin by considering the familiar linear regression model, $Y_i = X_i\beta + \varepsilon_i$, 3.9

where $i = 1, 2, \dots, \dots, n$: Y_i is a dependent variable, X_i is a vector of k independent variables or predictors, β is a k-by-1 vector of unknown parameters and the ε_i are zero-mean stochastic disturbances. Typically, the ε_i are assumed to be independent across observations with constant variance σ_i^2, and distributed normally. That is, the normal linear regression model is characterized by the following features:

1. **Stochastic component:** the Y_i are usually assumed to have independent normal distributions with $E(Y_i) = \mu_i$, with constant variance σ^2, or $Y_i \widetilde{iid} N(\mu, \sigma^2)$

   If it is not normally distributed then $y_{i \sim} P(u_i)$

2. **Systematic component:** specifies the explanatory or the independent variables for the model:

   $$\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$$

   The covariates x_i combine linearly with the coefficients to form the linear predictor

   $$\eta_i = X_i \beta \qquad 3.10$$

3. **Link between the random and systematic components:** the linear predictor $X_i\beta = \eta_i$ is a function of the mean parameter μ_i via a *link* function, $g(\mu_i)$.

   It should be noted that for the normal linear model g is an identity.

3.5 The Log-Linear Models

Suppose that we have a sample of n observations $y_{1,}y_{2,\dots,}y_n$ which can be treated as realizations of independent Poisson random variables, with, $Y_i \sim P(\mu_i)$, and suppose that we want to let the mean μ_i (and therefore the variance) depend on a vector of explanatory variables x_i .

We could entertain a simple linear model of the form

$$\mu_i = x_i'\beta \qquad 3.11$$

But this model has the disadvantage that the linear predictor on the right hand side can assume any real value, whereas the Poisson mean on the left hand side, which represents an expected count, has to be non-negative.

A straightforward solution to this problem is to use the model of logarithm of the mean instead of using a linear model. Thus, we take logs calculating

$$\eta_i = \log(\mu_i) \qquad 3.12$$

and assume that the transformed mean follows a linear model

$$\eta_i = x_i^{'}\beta \qquad 3.13$$

Thus, we consider a generalized linear model with link log. Combining these two steps in one we can write the log-linear model as

$$\log(\mu_i) = x_i^{'}\beta \qquad 3.14$$

In this model the regression coefficient β_j represents the expected change in the log of the mean per unit change in the predictor x_j. In other words increasing x_j by one unit is associated with an increase of β_j in the log of the mean.

Exponentiating Equation we obtain a multiplicative model for the mean itself:

$$\mu_i = \exp(x_i^{'}\beta) \qquad 3.15$$

In this model, an exponentiated regression coefficient $\exp(\beta_j)$ represents a multiplicative effect of the j-th predictor on the mean. Increasing x_j by one unit multiplies the mean by a factor $\exp(\beta_j)$.

A further advantage of using the log link stems from the empirical observation that with count data the effects of predictors are often multiplicative rather than additive. That is, one typically observes small effects for small counts, and large effects for large counts. If the effect is in fact proportional to the count, working in the log scale leads to a much simpler model.

3.6 Fisher Scoring in Log - Linear Models

Fisher scoring algorithm is a form of Newton-Raphson method used in statistics to solve maximum likelihood equations numerically. Nelder and Wedderburn (1972) applied Fisher scoring algorithm to estimate $\hat{\beta}$ in generalized linear models. The Fisher scoring algorithm for Poisson regression models with canonical link would be considered, where it would be modelled as:

$$g(\eta_i) = \log(\mu_i) \qquad 3.16$$

The derivative of the link is easily seen to be

$$g' = \frac{1}{\mu_i} \qquad 3.17$$

Specifically, given an initial estimate β, the algorithms update it to β^{new} by

$$\beta^{new} = \beta + \{E(-\frac{\partial L}{\partial\beta\partial\,\beta^T})\}^{-1}\frac{\partial L}{\partial\beta} \qquad 3.18$$

Where both derivatives are evaluated at β, and the expectation is evaluated as if β were the true parameter values.

β is then replaced by β^{new} and the updating is repeated until convergence.

It can be shown that for a GLM, the updating equation can be rewritten as

$$\beta^{new} = \beta + (X^TWX)^{-1}X^TW_z \qquad 3.19$$

Where, z is the n-vector with ith component

$$z_i = (Y_i - \mu_i)y'(\mu_i) \qquad 3.20$$

And, W is the $n \times n$ diagonal matrix with

$$W_i = \{g'(\mu_i)^2 b''(\theta_i)\}^{-1} \qquad 3.21$$

$$W_i = (\mu_i . \frac{1}{\mu_i^2})^{-1} \quad 3.22$$

And this simplifies to

$$W_i = \mu_i \quad 3.23$$

It is noted that the weight is inversely proportional to the variance of the working dependent variable.

3.7 The Poisson Distribution

The Poisson distribution (pronounced [pwasɔ̃]) (or Poisson law of small numbers) is a discrete probability distribution that expresses the probability of a number of events occurring in a fixed period of time if these events occur with a known average rate and each count occur independently of the time since the last event. The Poisson distribution can also be used for the number of events in other specified intervals such as distance, area or volume.

The Poisson regression model is a technique used to describe count data as a function of a set of predictor variables. In the last two decades it has been extensively used both in human and in veterinary epidemiological studies to investigate the incidence and mortality of chronic diseases. Also Poisson regression has been applied in the analysis of accident data for modelling traffic crashes in different parts of the world. Among its numerous applications, Poisson regression has been mainly applied to compare exposed and unexposed cohorts and to evaluate the causes of road traffic accidents.

The distribution was first introduced by Simeon-Denis Poisson (1781–1840) and published in 1838 in his probability theory. The work focused on certain random variables N that count, among other things, the number of discrete occurrences (sometimes called “arrivals”) that take place during a time-interval of given length, Poisson (1838).

If the expected number of occurrences in this interval is λ, then the probability that there are exactly *k* occurrences

(*k* being a non-negative integer, *k* = 0, 1, 2, ...) is equal to

$$f(k,\lambda) = \frac{\lambda^k e^{-\lambda}}{k!} \qquad 3.24$$

where

• *k* is the number of occurrences of an event - the probability of which is given by the function $f(k,\lambda)$

• λ is a positive real number, equal to the expected number of occurrences that occur during the given interval.

For instance, if the events occur on average rate of 5 times per minute, and one is interested in probability for k times of events occurring in a 12 minute interval, one would use as the model a Poisson distribution with λ = 12×5 = 60.

The parameter λ is not only the *mean* number of occurrences,(k) but also its variance

$$\sigma_k^2 = E(k^2) - [E(k)]^2 = \lambda. \qquad 3.25$$

Thus, the number of observed occurrences fluctuates about its mean λ with a standard deviation

$$\sigma_k = \sqrt{\lambda}. \qquad 3.26$$

As a function of *k*, this is the discrete probability mass function. The Poisson distribution can be derived as a limiting case of the binomial distribution. The Poisson distribution can be applied to systems with a large number of possible events, each of which is rare. A classic example is the nuclear decay of atoms.

The Poisson distribution is sometimes called a Poissonian, analogous to the term Gaussian for a Gauss or normal distribution.

Assumptions of Poisson distribution are:

- Observations are independent.
- Probability of occurrence in a short interval is proportional to the length of the interval.
- Probability of another occurrence in such a short interval is zero.

Verification of Poisson distribution as a member of exponential family.

The Poisson distribution belongs to the exponential family as defined by Nelder and Wedderburn (1972).

By taking logarithm of the Poisson distribution function, we obtain

$$logf_i(y_i) = y_i \log(u_i) - u_i - \log(y_i!) \qquad 3.27$$

where

y_i is the number of occurrences of the count or event- the probability of which is $f(y_i)$

u_i is the expected number of occurrences that occur during the given interval.

Looking at the coefficient of y_i we observe immediately that the link function is $\log(u_i)$ and the canonical parameter (θ_i) is given by

$$\theta_i = \log(u_i) \qquad 3.28$$

Therefore, the canonical link is the logarithm.

Solving for u_i we obtain the inverse link

$$u_i = e^{\theta_i} \qquad 3.29$$

and we see that we can write the second term in equation 3.12 as

$$\mathrm{b}(\theta_i) = e^{\theta_i} \qquad 3.30$$

The last term is a function of y_i only, so we identify from equation 3.27 that

$$c(y_i, \varphi) = \log(y_i!) \qquad 3.31$$

Finally, it should be noted that we can take the dispersion parameter $(\varphi) = 1$, just as it is in the binomial case and we verify that Poisson distribution belongs to the exponential family.

Verification of equal Mean and Variance

Differentiating the cumulant function $b(\theta_i)$ we have

$$u_i = b^{'}(\theta_i) = e^{\theta_i} = u_i \qquad 3.32$$

And differentiating again we have

$$\sigma^2 = b^{''}(\theta_i) = e^{\theta_i} = u_i \qquad 3.33$$

Hence the mean is equal to the variance

3.8 Poisson Regression

In spite of its recent wide application, Poisson regression model remains partly poorly known, especially if compared with other regression techniques, like linear, logistic and Cox regression models.

The Poisson regression model assumes that the sample of n observations, y_i are observations on independent Poisson variables Y_i with mean u_i.

If this model is correct, the equal variance assumption of classic linear regression is violated, since the Y_i have means equal to their variances.

So we fit the generalized linear model,

$$\log(u_i) = x_i^{'}\beta \qquad 3.34$$

We say that the Poisson regression model is a generalized linear model with Poisson error and a log link, so that

$$u_i = \exp(x_i^{'}\beta) \qquad 3.35$$

This implies that one unit increases in an x_i are associated with a multiplication of u_i by $\exp(\beta_i)$.

3.9 Model Specification

The primary equation of the model is $P(Y_i = y_i) = \frac{e^{-\mu}\mu^{y_i}}{y_i!}, y_i = 0,1,2, \ldots$ 3.36

The most common formulation of this model is the log-linear specification as in equation

$$\log(\mu_i) = x_i^{'}\beta \qquad 3.37$$

The expected number of events per period is given by

$$E(y_i|x_i) = \mu_i = e^{x_i^{'}\beta} \qquad 3.38$$

Thus:

$$\frac{dE(y_i|x_i)}{dx_i} = \beta e^{x_i^{'}\beta} = \beta_i\mu_i \qquad 3.39$$

The major assumption of Poisson model is

$$E(y_i|x_i) = \mu_i = e^{x_i'\beta} = Var(y_i|x_i) \qquad 3.40$$

This assumption would be tested later on. If $Var(y_i|x_i) > E(y_i|x_i)$ then there is over-dispersion. If, $Var(y_i|x_i) < E(y_i|x_i)$ then under-dispersion has occurred.

3.10 Dispersion and the Negative Binomial Model

The major assumption of the Poisson model is equation 3.40

$$E[y_i|x_i] = \lambda_i = e^{x_i'\beta} = Var[y_i|x_i]$$

Implying that the conditional mean function equate the condition variance function.

This is very restrictive. If $E[\, y_i|x_i\,] < Var[\, y_i|x_i\,]$ then we speak about overdispersion, and when $E[\, y_i|x_i\,] > Var[\, y_i|x_i\,]$ we say we have underdispersion. The Poisson model does not allow for over or underdispersion. A richer model is obtained by using the negative binomial distribution instead of the Poisson distribution.

Instead of equation

$$P[\, Y_i = y_i\,] = \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}, \qquad 3.41$$

we then use

$$P\left(Y_i = \frac{y_i}{\beta}, x_i\right) = \frac{\Gamma(\theta+y_i)}{\Gamma(y_i+1)\Gamma(\theta)}\left(\frac{\lambda_i}{\lambda_i+\theta}\right)^{y_i}\left(1-\frac{\lambda_i}{\lambda_i+\theta}\right)^{\theta} \qquad 3.42$$

This negative binomial distribution can be shown to have conditional mean λ_i and conditional variance $\lambda_i(1+\eta^2\lambda_i)$ with $\eta^2 \coloneqq \frac{1}{\theta}$. Note that the parameter η^2 is not allowed to vary over the observations. As before, the conditional mean function is modeled as

$$E[\, y_i|x_i] = \lambda_i = e^{x_i'\beta} \qquad 3.43$$

The conditional variance function is then given by

$$Var[y_i|x_i] = e^{x_i'\beta}(1+\eta^2 e^{x_i'\beta}) \qquad 3.44$$

Using maximum likelihood, we can then estimate the regression parameter β, and also the extra parameter η. The parameter η measures the degree of over (or under)dispersion. The limit case $\eta = 0$ corresponds to the Poisson model.

3.11 Parameter Estimation

Estimation involves estimating the regression parameters specifically using the maximum likelihood estimation

3.11.1 Maximum Likelihood Estimation

The likelihood function for n independent Poisson observations is a product of probabilities given by $L(\theta \,|\, X, Y) = \prod_{i=1}^{n} \frac{e^{y_{i(\theta' x)}} e^{-e^{\theta' x}}}{y!}$ 3.45

If $prob(y_i) = \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}$ is the probability function of Poisson distribution

Taking logarithms of equation 3.45 and ignoring the constant involving $\log(y_i!)$, we find that the log-likelihood function as

$$\log L(\beta) = \sum_{i=0}^{n}[-\lambda_i + y_i x_i'\beta] \qquad 3.46$$

$$= \sum_{i=1}^{n}[-e^{x_i'\beta} + y_i x_i'\beta] \qquad 3.47$$

where $\lambda_i = \mu_i = e^{x_i'\beta}$

The parameters of this equation can be estimated using maximum likelihood method

$$\frac{\partial L}{\partial \beta} = \sum_{i=1}^{n} \left(y_i - e^{x_i'\beta} \right) x_i = 0 \quad 3.48$$

and

$$\frac{\partial^2 L}{\partial \beta \partial \beta'} = -\sum_{i=1}^{n} \left[e^{x_i'\beta} x_i' x_i \right], \quad 3.49$$

which is the Hessian of the function and with typical element

$$\frac{\partial^2 L}{\partial \beta \partial \beta'} = -\sum_{i=1}^{n} [e^{x_i'\beta} x_{ij} x_{il}] \;\; ; \; j, l = 1,2, \ldots, \ldots p. \quad 3.50$$

As $\frac{\partial^2 L}{\partial \beta_j \partial \beta_i} = -\sum_{i=1}^{n} [e^{x_i'\beta} x_{ij} x_{il}]$ does not involve the *y* data 3.51

$$k_{jl} = E\left(\frac{\partial^2 L}{\partial \beta_j \partial \beta_i} \right) = -\sum_{i=1}^{n} [e^{x_i'\beta} x_{ij} x_{il}]; j, l = 1,2, \ldots, \ldots, p. \quad 3.52$$

And the information matrix is

$$K = \sum_{i=1}^{n} [e^{x_i'\beta} x_i' x_i] \quad 3.53$$

There is no closed form solution to, $\frac{\partial L}{\partial \beta} = \sum_{i=1}^{n} (y_i - e^{x_i'\beta}) x_i = 0$ 3.54

so the MLE for β must be obtained numerically. However, as the Hessian is negative definite for all x and β, the MLE ($\hat{\beta}$) is unique, if it exists.

From $\frac{\partial^2 L}{\partial \beta_j \partial \beta_l} = -\sum_{i=1}^{n} [e^{x_i'\beta} x_{ij} x_{il}]$ and: $k_{jl} = E\left(\frac{\partial^2 L}{\partial \beta_j \partial \beta_l} \right) = -\sum_{i=1}^{n} [e^{x_i'\beta} x_{ij} x_{il}]$ 3.55

$$k_{jlr} = \left(\frac{\partial^3 L}{\partial\beta_j \partial\beta_l \partial\beta_r}\right) = -\sum_{i=1}^{n}[e^{x_i'\beta} x_{ij} x_{il} x_{ir}] \quad 3.56$$

And

$$k_{jl}^{(r)} = \frac{\partial k_{jl}}{\partial\beta_r} = -\sum_{i=1}^{n}[e^{x_i'\beta} x_{ij} x_{il} x_{ir}] \; , j, l, r = 1,2,3, \dots, p. \quad 3.57$$

To make matters more transparent, consider the case of a single covariate and an intercept. Then x_i is a scalar observation and

$$L = \sum_{I=1}^{N}[-\lambda_i + y_i(\beta_1 + \beta_2 x_i) - \log(y_i)] \quad 3.58$$

Where $\lambda_i = \exp(\beta_1 + \beta_2)$, for $= 1,2., \dots, n$.

The first order conditions, $\frac{\partial L}{\partial\beta} = 0$ yield a system of K equations (one for each β) of the form

$$\sum_{i=1}^{n}(y_i - e^{x_i'\beta})x_i = 0 \quad 3.59$$

Where $\hat{y}_i = e^{x_i'\hat{\beta}}$ is the fitted value of y_i. The predicted/fitted value has as usual been taken as the estimated value of $(y_i|x_i)$.

This first order condition tells us that the vector of residual is r orthogonal to the vectors of explicative variables.

3.12 Tests of Hypotheses

Likelihood ratio tests for log-linear models can easily be constructed in terms of deviances. In general, the difference in deviances between two nested models has approximately in large samples a chi-squared distribution with degrees of freedom equal to the difference in the number of parameters between the models, under the assumption that the smaller model is correct. One can also construct Wald tests, based on the fact that the maximum likelihood estimator $\hat{\beta}$ has approximately in large samples a multivariate normal distribution with mean equal to the true parameter value β and variance-covariance matrix, $var(\hat{\beta}) = X'WX$ 3.60

where X is the model matrix and W is the diagonal matrix of estimation weights.

3.13 Likelihood Ratio Test

A simple test on the overall fit of the model, as an analogue to the F-test in the classical regression model is a Likelihood Ratio test on the "slopes".

The model with only the intercept is nothing but the mean of the counts, or

$\lambda_1 = \bar{y}$ 3.61

Where $\bar{y} = \sum_{i=1}^{n} \frac{y_i}{n}$ 3.62

The corresponding log-likelihood is:

$L_R = n\bar{y} + \log(\bar{y}) \left(\sum_{i=1}^{n} y_i\right) - \sum_{i=1}^{n} log y_i!$ 3.63

where the R stands for the "restricted" model, as opposed to the "unrestricted" model with $K - 1$ slope parameters. The last term in $\sum_{i=1}^{n} log y_i$! can be dropped, as long as it is also dropped in the calculation of the maximized likelihood $L_u = \sum_{i=1}^{n} [-e^{x'_i \beta} + y_i x'_i \beta - log y_i!]$ 3.64

for the unrestricted model, L_U using $L = e^{x'_i \hat{\beta}_t}$ 3.65

The Likelihood Ratio test is then:

$LR = 2(L_U - L_R)$ 3.66

Which follows a χ^2 distribution with K-1 degrees of freedom.

3.14 Goodness of Fit Test

In order to assess the adequacy of the Poisson regression model you should first look at the basic descriptive statistics for the event count data. If the count mean and variance are significantly different (equivalent in a Poisson distribution) then the model is likely to be over-dispersed or under-dispersed. The model analysis option gives a scale parameter (sp) as a measure of over-dispersion; this is equal to the Pearson chi-square statistic divided by the number of observations minus the number of parameters (covariates and intercept). Under-dispersion is very uncommon to various forms of count data especially with accident data.

The variances of the coefficients can be adjusted by multiplying by sp. The goodness of fit test statistics and residuals can be adjusted by dividing by sp. Using a quasi-likelihood approach sp could be integrated with the regression, but this would assume a known fixed value for sp, which is seldom the case. A better approach to over-dispersed Poisson models is to use a parametric alternative model, the negative binomial.

The deviance (likelihood ratio) test statistic, G^2, is the most useful summary of the adequacy of the fitted model. It represents the change in deviance between the fitted model and the model with a constant term and no covariates; therefore G^2 is not calculated if no constant is specified. If this test is significant then the covariates contribute significantly to the model.

The deviance goodness of fit test reflects the fit of the data to a Poisson distribution in the regression. If this test is significant then a red asterisk is shown by the P value, and you should consider other covariates and/or other error distributions such as negative binomial.

Technical validation:

The deviance function is:

$$Deviance = 2\sum_{i=1}^{n} y_i \ln\left[\frac{y_i}{\hat{\mu}_i}\right] - (y_i - \hat{\mu}_i) \qquad 3.67$$

Where y is the number of events, n is the number of observations and $\hat{\mu}_i$ is the fitted Poisson mean. The first term is identical to the binomial deviance, representing `twice' a sum of observed times log of observed over fitted'. The second term, a sum of differences between observed and fitted values, is usually zero, because MLE's in Poisson models have the property of reproducing marginal totals, as noted above.

The log-likelihood function is:

$$L = \sum_{i=1}^{n} y_i \ln(\hat{\mu}_i) - \hat{\mu}_i - \ln(y_i!) \quad 3.68$$

The maximum likelihood regression proceeds by iteratively re-weighted least squares, using singular value decomposition to solve the linear system at each iteration, until the change in deviance is within the specified accuracy.

The Pearson chi-square residual is:

$$r_p = \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i} \quad 3.69$$

For large samples the distribution of the deviance is approximately a chi- squared with $n - p$ degrees of freedom, where n is the number of observations and p the number of parameters. Thus, the deviance can be used directly to test the goodness of fit of the model. An alternative measure of goodness of fit is Pearson's chi-squared statistic, which is defined as

The Pearson goodness of fit test statistic is:

$$\chi^2 = \sum_{i=1}^{n} \frac{y_i - \mu_i}{\sqrt{\hat{\mu}_i}} \quad 3.70$$

The deviance residual is (Cook and Weisberg, 1982):

$$r_d = sign(y_i - \hat{\mu}_i)\sqrt{deviance(y_i, \hat{\mu}_i)} \quad 3.71$$

The Freeman-Tukey, variance stabilized, residual is (Freeman and Tukey, 1950):

$$r_{ft} = \sqrt{y_i} + \sqrt{y_i + 1} - \sqrt{4\hat{\mu}_i + 1} \quad 3.72$$

The standardized residual is:

$$r_s = \frac{y_i - \mu_i}{\sqrt{1 - h_i}} \quad 3.73$$

where h is the leverage (diagonal of the Hat matrix).

3.16 Akaike Information Criterion (AIC)

The Akaike Information Criterion (AIC) is a way of selecting a model from a set of models. The chosen model is the one that minimizes the Kullback-Leibler distance between the model and the truth. It's based on information theory, but a heuristic way to think about it is as a criterion that seeks a model that has a good fit to the truth but few parameters. It is defined as:

AIC = -2 (ln (likelihood)) + 2 K   3.74

where likelihood is the probability of the data given a model and K is the number of free parameters in the model. AIC scores are often shown as ΔAIC scores, or difference between the best model (smallest AIC) and each model (so the best model has a ΔAIC of zero).

The second order information criterion, often called AICc, takes into account sample size by, essentially, increasing the relative penalty for model complexity with small data sets. It is defined as:

AICc = -2 (ln (likelihood)) + 2 K * (n / (n - K - 1)) 3.75

where n is the sample size. As n gets larger, AICc converges to AIC (n - K -1) → n as n gets much bigger than K, and so (n / (n - K - 1)) approaches 1), and so there's really no harm in always using AICs regardless of sample size. In model selection in comparative methods, sample size often refers to the number of taxa (Butler and King, 2004; O'Meara et al., 2006).

3.17 Software (Microsoft Excel, Minitab and R)

The preliminary and exploratory analysis of the data would be done using Microsoft Excel and Minitab 14 statistical packages. In the Poisson mean analysis, Minitab 14 computes the overall mean of the data using $\bar{c} = \sum_{i=1}^{k} \frac{c_i}{k}$, 3.76

$\bar{c}$ is the mean of the count data, c_i is the ith count observation and k number of observations.

It then sets upper and lower decision limits which are used as basis for comparison with individual means at a given significance level as

U/LDL= $\bar{c} \pm h_\alpha s \times \sqrt{\left(\frac{k-1}{k}\right)}$, 3.77

where h_α = inverse cumulative probability of α_2 for the standard normal distribution, where $\alpha_2 = 1 - \alpha / (2\times k)$.

R Statistical Software (R Development Core Team, 2008), uses the concept of maximum likelihood in analyzing GLMs. It applies built in model functions glm() to model Poisson regression (Chambers and Hastie, 1992) in the stats package and glm.nb() in the MASS pack for negative binomial regression(Venables and Ripley, 2002) along with associated methods for diagnostics and inference. In each of these the link function is also specified for R to fit the preferred model.

For instance, to model the number of people who visited the Bank after every five minutes given their gender is as shown below for both Poisson;

```
> ill=read.csv('C://Users/my pc/Documents/illustration.csv')
> attach(ill)
> names(ill)
[1] "time"     "gender"    "attendants"
> ill
   time gender attendants
1    1     1        5
2    2     1        3
3    3     1        4
4    4     2        4
5    5     2        3
6    6     2        5
7    7     1        4
8    8     1        6
9    9     1        3
10  10     1        2
> gender=factor(gender,levels=c(1,2))
>  mod1<-glm(attendants~time+gender,family=poisson(link="log"))
> summary(mod1)

Call:
glm(formula = attendants ~ time + gender, family = poisson(link = "log"))

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-0.86462  -0.46918  -0.04625   0.28567   1.12556
```

Coefficients:

```
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.47866    0.36161   4.089 4.33e-05 ***
time        -0.02306    0.05616  -0.411    0.681
gender2      0.02275    0.34793   0.065    0.948
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for Poisson family taken to be 1)

    Null deviance: 3.3524  on 9  degrees of freedom
Residual deviance: 3.1732  on 7  degrees of freedom
AIC: 41.178
Number of Fisher Scoring iterations: 4
```

Model for the number of people killed by road accidents in Ghana from 2001-2010 will be analyzed using R software with the number of people killed in road accidents being the dependent variable and, age and yearly time, vehicle type and yearly time and the day the accident that killed the people occurred as the independent variables. Total percentage of variation in the dependent variable explained by these factors will be analyzed and discussed.

CHAPTER 4

ANALYSIS

4.0 Introduction

This chapter focuses on the preliminary analysis of data on the number of people killed by road accident in Ghana and the detailed analysis of the data using various statistical tools such as Poisson regression and Negative Binomial regression to fit a model to the data.

4.1 Data Source

The data for this thesis was a secondly data obtained from the Building and Road Research Institute of the Council for Scientific and Industrial Research. The data was originally collected using accident report form by the Motor Traffic and Transport Unit of the Ghana Police Service. This study considered accident data for ten year period from 2001 to 2010. The number of people killed by road accident was used as the response variable in all models and the other variables such as age of casualty, the day the accident occurred which resulted in the death of the people, the time people were killed, vehicle type and road user class as the explanatory variables.

4.2 The Number of People Killed by Road Accidents in Ghana

There were 114,770 road accidents which occurred in Ghana from 2001 to 2010 which killed 19,088 people. This shows that on the average, 11,477 road accidents occurred every year and 1,909 lives are lost through these accidents.

4.2.1 Age Distribution of People Killed by Road Accidents in Ghana.

The figure 4.1 below shows the distribution of the number of people who were killed by road accidents and their age groupings. It is clear from the figure that the youth is the most vulnerable to road accident. The age group 26-35 recorded 4469 casualties from 2001-2010, the highest number of deaths. This was followed by 16-25 which recorded 2982, a little above those killed in the age group 36-45 which had 2949. This result is not surprising since research has shown that most people who are at risk in road accidents are in the ages between 15 and 44 years, Margie et al (2002). The age group which recorded the least number of deaths is people of over 65 years which recorded 824 people who died in road accidents within the ten year period. It should be noted however that people of over 65 years are pensioners and are mostly not in active service and therefore do not travel regularly. This result is clearly shown in the figure 4.1 below.

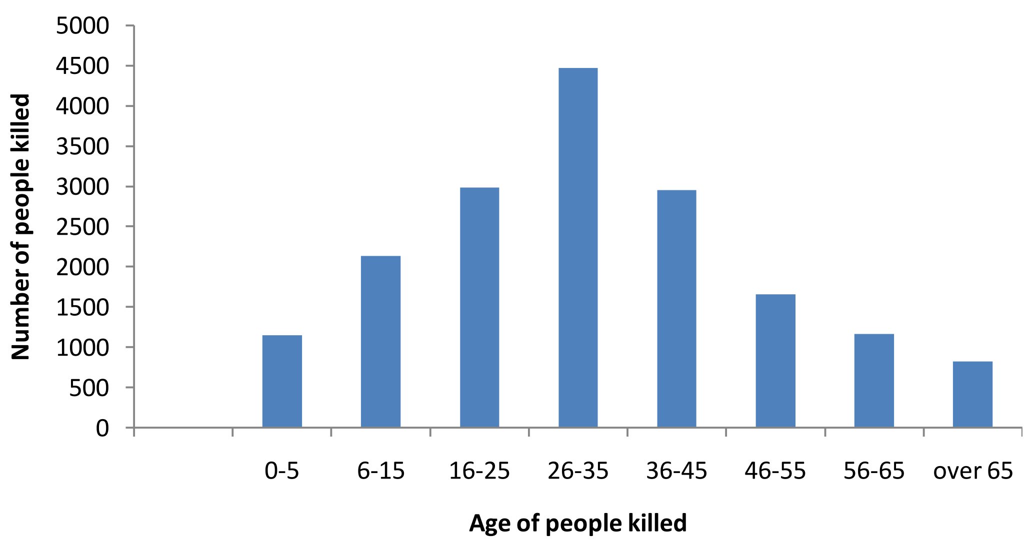

Figure 4.1 A bar graph illustrating the average number of people killed by road accident in Ghana from 2001 t0 2010 for various age groupings.

It is interesting to note from the graph of figure 4.1 that the risk of one getting killed by road accident in Ghana increases from infancy till one gets to the early adulthood. Thus, at the age of 26 to 35 one is at the peak of dying through road accident but as one grows past the 35 years, the risk level begins to reduce gradually till 65 years and over where the risk of dying in road accident decreases drastically. In fact the count distribution from figure 4.1 could be assumed to be following the normal distribution since almost half of the data lies to right of the Age 26-35 and the other half of the data lays the left of the age 26-35.

4.2.2 Annual Distribution of People Killed by Road Accidents in Ghana.

The Table 4.1 below shows the time in years for which accidents that killed the people occurred. It also presents the total number of people killed in road accident annually from 2001-2010.

Table 4.1 Total Number of people killed by Road Accidents from 2001 to 2010

Year	Total No. of People Killed
2001	1660
2002	1666
2003	1734
2004	2185
2005	1784
2006	1856
2007	2043
2008	1937
2009	2237
2010	1986

The most significant feature of Table 4.1 is that the number of people who were killed by road accidents in Ghana seems to be increasing as years go by. In 2001 there were 1660 people who were killed in road accidents, this was increased to 1666 in 2002 and in 2003 the number rose to 1734. By 2010, the number had risen to 1986. However, there were sharp increases in 2004, 2007 and 2009 with the number of who were killed in road accidents being 2185, 2043 and 2237 respectively. This observation support what Odoom (2010) stated in his publication that as years go by, the number of vehicle in Ghana will be increasing and the number of traffic accident increases accordingly and as such the number of people who are likely to be killed in road accidents will surely increase.

4.2.3 Days in which People were Killed by Road Accidents in Ghana.

The number of people who were killed in each of day of the week is presented in table 4.2 below from 2001-2010. It also contains the percentage number of people killed in each day for the ten year period and the arithmetic means of the number of people killed in each day for every year for the ten years.

Table 4.2 Number of people who were killed in the days of the week by road accidents in Ghana from 2001-2010.

Day	Number of People Killed	Percentage Killed	Mean
Monday	2686	13.74335	268.6
Tuesday	2473	12.6535	247.3
Wednesday	2164	11.07245	216.4
Thursday	2965	15.1709	296.5
Friday	3410	17.44781	341
Saturday	2923	14.956	292.3
Sunday	2923	14.956	292.3

From the Table 4.2 one observes that Friday has the highest number of people who were killed by road accidents from 2001 to 2010 in Ghana. There were 3,410 people which constitute 17.4% of the total number of those killed by road accidents. This was followed by Thursday which had 2,965 people constituting 15.2% of those killed by road accidents. Saturday and Sunday recorded the same number of people who were killed in road accidents within the 10 year period. Wednesday recorded 2,164 which was the least number of people who were killed in road accident.

The table also shows the mean number of people who were killed in each day for every year for 2001-2010 by road accidents in Ghana which shows that most people were killed on Friday and Wednesday recorded the least number of people who were killed through road accidents on the average. However, the fact that the mean numbers of people killed by road accidents differ in numerical terms from day to day is not enough to conclude that there are significant differences in the means. To ascertain the truth or otherwise of this, Poisson mean analysis was conducted using Minitab software and the result is as shown in the figure 4.2 below.

The Minitab package version 14 computes the overall expectation of the count data using $\overline{c} = \sum_{i=1}^{k} \frac{c_i}{k}$, $\overline{c}$ is the mean of the count data, c_i is the ith count observation and k number of observations.

It computes the standard deviation (s) of the data as $s = \sqrt{\overline{c}}$ and uses it to determine the upper and lower decision lines (U/LDL) at a given confidence level (α)using

U/LDL$= \overline{c} \pm h_{\alpha} s \times \sqrt{\left(\frac{k-1}{k}\right)}$, where h_{α} = inverse cumulative probability of $\alpha_2$for the standard normal distribution, where α_2 = $1 - \alpha$ / (2$\times$ k).

Minitab package 14 finally plots the count data with the decision lines which could then be used to compare the means at the significant level stated.

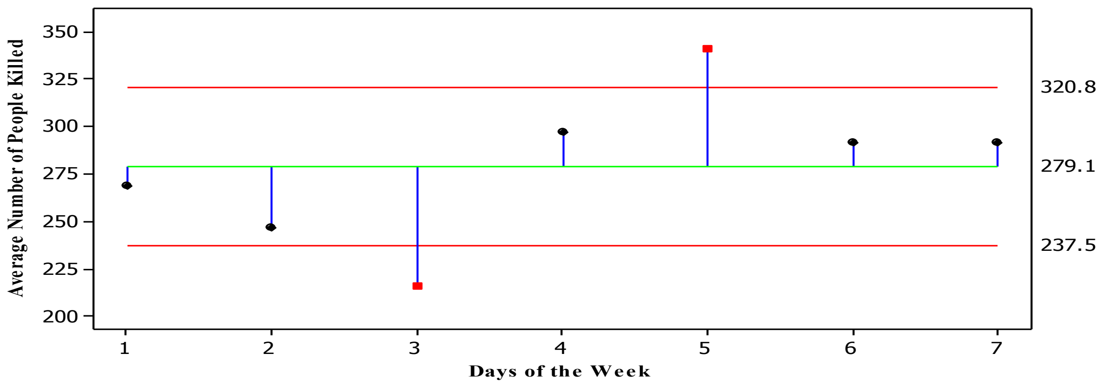

Figure 4.2 A plot showing the difference in the mean number of People killed by road accident for the days of the week.

From the figure 4.2 1, 2,…, 7 represent Monday, Tuesday,…,Sunday respectively. The overall mean of the number of people killed in each day of the week was found to be 279, the upper decision limit was computed as 320.8 and the lower decision limit was 237.7.

It is obvious from the figure 4.2 that there is no significant differences among the mean numbers of people who killed by road accident for all the days of the week except Wednesday and Friday. This is because they are all lying within the decision limits. The number of people killed on Wednesday is significantly smaller than all the other days at 0.05 α level while Friday has more people been killed on the average than the rest of the days at 95% significant level.

3.2.4 Type of Vehicle that Killed People in Road Accidents in Ghana.

The type of vehicle involved in road accident cannot be ruled out as a contributory factor to the number of people who are killed in that accident. The type of vehicle involved in the road accident which killed the people, the number of people killed by the type of vehicle, the percentage number of people killed and the average number of people killed by type of vehicle for every year are presented in the Table 4.3 below.

Table 4.3 The number of people who were killed by road accidents through different types of vehicles in Ghana from 2001 to 2010.

Vehicle type	Total number of people killed	Percentage killed	Ave killed
Car	5527	29.2063	553
Goods vehicle	3675	19.41978	368
Bus/mini bus	5650	29.85627	565
motor cycle	1495	7.900021	150
Pick up	1033	5.458677	103
Cycle	979	5.173325	98
Others	565	2.985627	57

From Table 4.5 above, it could be seen that Bus/Mini bus killed 5,650 people from 2001 to 2010 which constitute 29.2% of the total number of people killed via road accident in the same period.

This was followed closely by cars which killed 5,527 representing 29.9% of those who were killed by road accidents. Goods vehicle was the third on the list of type of vehicles which kill most people in accidents with 3,675 people who were killed for the ten year period. This figure represents 19.4 % of the total deaths through road accidents in Ghana from 2001 to 2010. The type of vehicle which killed the least number of people for the years under consideration is others. Only 565 people were killed by other type of vehicles such as tractor, bulldozer, tipper, mixer and loading box which constitute just 3% of the total number of people killed by various types of vehicles through road accidents. The average number of people killed by various types of vehicles in Ghana is as shown in the figure 4.3 below.

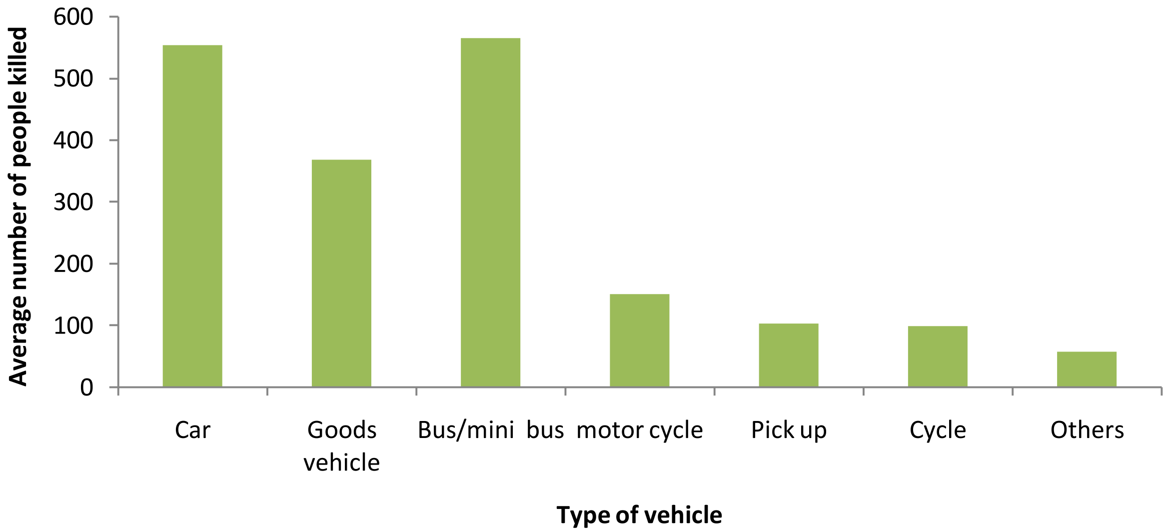

Figure 4.3 A bar chart showing the average number of people killed by different types of vehicles in Ghana from 2001 to 2010.

The chart shows clearly that cars, mini buses/buses and goods only vehicles are the three major killers of people of Ghana through road accidents. The nature of figure 4.3 shows as if the number of people a type of vehicle can carry may influence the number of people that vehicle can kill in accidents. The buses and mini buses are known to be the major carriers of people in Ghana, followed by cars, goods vehicles, motor cycles, pick up, cycle and the heavy duty vehicles like the truck and bulldozer in that orders as the number of people the kill in road accidents.

4.3 Modeling the Number of People Killed by Road Accidents in Ghana.

In order to model the number of people who are killed by road accidents in Ghana, R statistical software version 2.12.0 was used. The Generalized Linear Model (glm) procedure with Poisson as the main distribution specified using the Log link function. The Negative Binomial distribution was used to correct the error of over dispersion in the data in situations where the result of the Poisson regression model shows over dispersion.

The various models obtained when number of people killed by road accidents was regressed on factors such as age, the day the accident which killed the people occurred, the year and type of vehicle which was involved in the accident that killed the people are presented.

4.3.1 Type of Vehicle involved in the Accident that Killed the People.

The number of people killed in road accidents by different type of vehicles was modeled using Poisson regression and the results are presented in the table 4.4 below. The table contains various models generated and their AIC's

Table 4.4 The Poisson Regression Models for of number of People Killed different Types of Vehicle in Road Accidents from 2001to 2010 with their AIC's

Models	AIC's
1. $\log(mean\_killed) = \alpha_1 + \beta_i Vehicle\ type, i = 1,2, \ldots 8$	1427.2
2. $\log(mean\_killed) = \alpha_2 + \beta_i Year, i = 1,2, \ldots 10$	12059
3. $\log(mean\_killed) = \alpha_1 + \beta_i Vehicle\ type + \beta_j, Year\ i = 1,2, \ldots 7\ and\ j = 1,2, \ldots 10$	1096.3

The Table 4.4 shows that the best model which fit the type of vehicle involved in the accidents that killed people in Ghana from 2001 to 2010 is model 3 because it has the smallest AIC. The AIC of model 3 was found to be 1096.3 with a null deviance of 11894.55 on 69 degrees of freedom and residual deviance of 570.97on 54 degrees of freedom following the chi-square distribution (χ^2) on one degree of freedom.

The parameter and their estimates, the standard errors, the Chi-squares with their associated p values are presented in table 4.5 below.

Table 4.5 The Parameter Estimates of Selected Poisson Model for the number of people who were Killed by different types of vehicles in Road Accident from 2001- 2010 in Ghana.

Parameter	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	6.09040	0.02810	216.761	< 2e-16
Veh.type2	-0.40809	0.02128	-19.173	< 2e-16
Veh.type3	0.02201	0.01892	1.163	0.244662
Veh.type4	-1.30752	0.02915	-44.852	< 2e-16
Veh.type5	-1.67718	0.03390	-49.479	< 2e-16
Veh.type6	-1.73087	0.03468	-49.916	< 2e-16
Veh.type7	-2.28057	0.04417	-51.634	< 2e-16
year2002	0.03892	0.03602	1.080	0.279984
year2003	0.10893	0.03542	3.075	0.002103
year2004	0.30245	0.03391	8.918	< 2e-16
year2005	0.13411	0.03521	3.809	0.000140
year2006	0.20499	0.03464	5.917	3.28e-09
year2007	0.30099	0.03392	8.872	< 2e-16
year2008	0.24512	0.03434	7.139	9.41e-13
year2009	0.39036	0.03330	11.722	< 2e-16
year2010	0.42554	0.0330	12.869	< 2e-16

The Table 4.5 presents the parameter estimates of the selected model for the number of people who were killed by different types of vehicles in road accidents. The AIC of this model was 1096.3; the null deviance was 11894.55 on 69 degrees of freedom and residual deviance of 570.97 on 54 degrees of freedom following the chi-square distribution (χ^2) with one degree of freedom. The dispersion parameter was found to be 10.74019 and P-value of 0.00 which indicates that the model is significant at 5% $\alpha -$level. However, the assumption of equal variance to the mean in Poisson distribution has been violated since the dispersion parameter is not approximately equal to1. The dispersion parameter of the above model is 10.74 which is far greater than 1, an indication of over dispersion in the data. This means that the parameters of the model have been over estimated and the standard errors have been under estimated which will not give a true reflection of model which could provide appropriate mean number of people who were killed by different types of vehicles in road accidents from 2001 to 2010 in Ghana. To address this error, Negative Binomial regression was used to modify the model to nullify the effect of over dispersion in the data and the result is shown in the table 4.6 below.

Table 4.6 The Parameter Estimates of Negative Binomial Regression Model for Number of People who were Killed by Different Types of Vehicles in Road Accident from 2001-2010 in Ghana

Coefficient	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	6.07980	0.11210	54.237	< 2e-16
Veh.type2	-0.39209	0.10169	-3.856	0.000115
Veh.type3	0.02487	0.10123	0.246	0.805906
Veh.type4	-1.36794	0.10388	-13.169	< 2e-16
Veh.type5	-1.67558	0.10512	-15.940	< 2e-16
Veh.type6	-1.72659	0.10536	-16.387	< 2e-16
Veh.type7	-2.31388	0.10919	-21.191	< 2e-16
year2002	-0.04041	0.12883	-0.314	0.753797
year2003	0.05038	0.12843	0.392	0.694892
year2004	0.27526	0.12758	2.158	0.030955
year2005	0.13912	0.12808	1.086	0.277370
year2006	0.21209	0.12780	1.660	0.097011
year2007	0.32977	0.12739	2.589	0.009635
year2008	0.34928	0.12733	2.743	0.006086
year2009	0.43932	0.12705	3.458	0.000544
year2010	0.53288	0.12678	4.203	2.63e-05

From the Table 4.6 it is observed that the parameter estimates have reduced and the standard errors have also increased. The parametric analysis for the comparison between the Poisson and Negative Binomial regression for goodness of fit test of the model is shown in table 4.7.

Table 4.7 Parametric comparison between Poisson and Negative Binomial Regression for Goodness of Fit Test

Assessment parameter	Poisson Regression Model	Negative Binomial Regression Model
Null Deviance	11894.55	988.852
Degree of Freedom	69	69
Residual Deviance	570.97	66.666
Degrees of Freedom	54	54
Dispersion Parameter	10.57352	1.234556
AIC	1096.3	769.1

Looking at the results presented in Table 4.7, it is clear that the negative binomial regression model is actually the best model which fit the vehicular type data because the dispersion parameter has reduced from 10.57 which was giving by the Poisson regression model to 1.23.

The AIC of the Poisson model also reduced from 1093.3 to 769.1 in the negative Binomial model.

R package takes the first category in a data as the base level by default and as such Veh.type1 (cars) and the year 2001 were picked as the base levels for comparison in the analysis of the parameter estimates in the negative binomial regression model. The intercept was found to be 6.07980 which was very significant at 95% significant level with p-value of < 2e-16 following a Pearson Chi-square χ^2 distribution with 54 degrees of freedom as shown in table 4.6 above. With the exception of veh.type3 (bus/mini bus) which was not significantly different from the cars in the model, the rest of the vehicles were all significantly smaller than the base level in the model at 5% α-level for every year. For instance, veh.type2 (goods vehicle) was found to have parameter estimate of -0.39209 less than the logarithm of the expected number of people who were killed by cars for every year. It could also be said from table 4.6 that the expected number of people who were killed by veh.type7 (others) was $e^{-2.31388}$ =0.09887686 times less than that of cars for every year.

The table 4.6 further reveals that the expected number of people who were killed by different types of vehicles for the years 2002, 2003, 2005 and 2006 were not significantly different from 2001 for all types of vehicles in the model giving that the year 2001 is the base level. 2010 was found to be the year which had most people killed by road accident for all types of vehicles in Ghana. It was found that 2010 had $e^{0.53288}$ =1.703832 times more than the expected number of people killed in 2001 for all types of vehicles in Ghana. The model for the above table is presented in equation 4.1 below.

$$\log(mean\_killed) = 6.07980 - 0.39209V_1 + 0.02487V_2 - 1.36794V_3 - 1.67558V_4 - 1.72659V_5 - 2.31388V_6 - 0.04041Y_1 + 0.05038Y_2 + 0.27526Y_3 + 0.13912Y_4 + 0.21209Y_5 + 0.32977Y_6 + 0.34928Y_7 + 0.43932Y_8 + 0.53288Y_9 \qquad 4.1$$

where $\boldsymbol{V_1}$, $V_{\boldsymbol{2}}$,… $\boldsymbol{V}_6$ represent Veh.type2 (Goods vehicle), Veh.type3 (Bus/mini bus),…, Veh.type6 (others) respectively and $\boldsymbol{Y_1}$, $\boldsymbol{Y_2}$,…, $\boldsymbol{Y_9}$, denotes 2002,2003,…,2010 in that order.

4.3.2 The Days in which People were Killed in Road Accidents.

The days in which people were killed in road accident from 2001 to 2010 was modeled using Poisson regression and the various models with their AIC's presented in table 4.8.

Table 4.8 The Poisson Regression Models for number of People Killed in Road Accidents in the Days of the Week in Year with their AIC's

Models	AIC's
1. $\log(mean\_killed) = \alpha_1 + \beta_i Day, i = 1,2, \ldots 7$	854.97
2. $\log(mean\_killed) = \alpha_2 + \beta_i Year, i = 1,2, \ldots 10$	1032.2
3. $\log(mean\_killed) = \alpha_1 + \beta_i Day + \beta_j Year, i = 1,2, \ldots 7 \; and \; j = 1,2, \ldots 10$	676.08

The Table 4.8 shows that the best model which fit the number of people killed by road accidents in the days of the week from 2001-2010 in Ghana is model 3 because it has 676.08 which is the smallest AIC. Further, it has a null deviance of 689.08 on 69 degrees of freedom and residual deviance of 124.05 on 54 degrees of freedom following the chi-square distribution (χ_1^2). The AIC's of model 1 and 2 were 854.97 and 1032.2 respectively which do not represent a good model. The parameters with their corresponding estimates are shown in the table 4.9, it contains the standard errors and the Chi-squares with their associated p values.

Table 4.9 The Parameter Estimates of the Selected Poisson Model

Coefficients	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	5.453566	0.030370	179.572	< 2e-16
Day2	-0.082621	0.027869	-2.965	0.003030
Day3	-0.216095	0.028886	-7.481	7.38e-14
Day4	-0.085050	0.027886	-3.050	0.002289
Day5	0.098824	0.026638	3.710	0.000207
Day6	0.238659	0.025798	9.251	< 2e-16
Day7	0.084557	0.026729	3.164	0.001559
year2002	0.003608	0.034679	0.104	0.917140
year2003	0.043613	0.034338	1.270	0.204045
year2004	0.274798	0.032559	8.440	< 2e-16
year2005	0.072040	0.034102	2.112	0.034644
year2006	0.111606	0.033782	3.304	0.000954
year2007	0.207602	0.033044	6.283	3.33e-10
year2008	0.154323	0.033447	4.614	3.95e-06
year2009	0.298318	0.032395	9.209	< 2e-16
year2010	0.179305	0.033256	5.392	6.98e-08

The Table 4.9 shows the parameter estimates of the selected model. The AIC of this model was 676.08; the null deviance was 689.08 on 69 degrees of freedom and residual deviance of 124.05 on 54 degrees of freedom following the chi-square distribution (χ^2) with one degree of freedom. The dispersion parameter was found to be 2.297222 and P-value of 1.959054e-07 which indicates that the model is significant. However, one assumption of Poisson distribution which is the equality of the mean and variance which means that the dispersion parameter should always be closer to 1 has been violated. The dispersion parameter of the above model is far greater than 1, an indication of over dispersion in the data. This means that the parameters of the model have been over estimated and will not give a true reflection of number of people likely to be killed through road accidents in a given day of the week for a particular year. To eliminate this error, Negative Binomial regression was used to validate the model and the result is shown in the table 4.10 below.

Table 4.10 Parameter Estimates of Negative Binomial Regression for the Selected Model.

Parameter	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	5.452302	0.039726	137.248	< 2e-16
Day2	-0.081113	0.036716	-2.209	0.027162
Day3	-0.214166	0.037492	-5.712	1.11e-08
Day4	-0.084739	0.036736	-2.307	0.021072
Day5	0.096618	0.035807	2.698	0.006970
Day6	0.238497	0.035176	6.780	1.20e-11
Day7	0.087156	0.035852	2.431	0.015058
year2002	0.008467	0.044964	0.188	0.850638
year2003	0.046434	0.044715	1.038	0.299064
year2004	0.278236	0.043358	6.417	1.39e-10
year2005	0.070783	0.044559	1.589	0.112169
year2006	0.115135	0.044284	2.600	0.009324
year2007	0.208437	0.043738	4.766	1.88e-06
year2008	0.153238	0.044056	3.478	0.000505
year2009	0.295600	0.043267	6.832	8.37e-12
year2010	0.176739	0.043918	4.024	5.72e-05

The table below shows the assessment criteria for Poisson and Negative Binomial regression models.

Table 4.11 Parametric comparison between Poisson and Negative Binomial Regression for Goodness of Fit Test

Assessment parameter	Poisson Regression Model	Negative Binomial Regression Model
Null Deviance	689.08	384.79
Degrees of Freedom	69	69
Residual Deviance	124.05	69.64
Degrees of Freedom	54	54
Dispersion Parameter	2.297222	1.289629
AIC	676.08	663.6

The dispersion parameter for the negative binomial model was 1.289629 which was closer to 1 and a good sign of reduced over dispersion in the data if not totally eliminated as compared with 2.297222 for the Poisson regression model as shown in the Table 4.11.

Also, it could be observed from the table 4.11 that the AIC of the negative binomial regression model is 663.6 which is smaller than that of Poisson regression model of AIC 676 and indication of better model from the negative binomial regression.

The R package takes the first category in a data as the base level by default and as such Monday and 2001 were picked as the base levels for comparison in the analysis of the parameter estimates in the model. From table 4.9 the intercept was found to be 5.45302 which was very significant at 95% significant level. All the days were significant in the model at 5% α-level but the most significant day was found Saturday with coefficient of 0.238497. This means that the expected number of people killed through accident on Saturday is $e^{0.238497}$ =1.2693340 more than that of Monday for every year. The day with the least significance in the model was Tuesday which had a parameter estimate of -0.081113indicating a reduction of $e^{-0.08113}$=0.922089 which means that the expected number of people who were killed on Tuesday was 7.8% lower for each year. From the table 4.10 the expected number of people killed in road accident for the year 2001 on Mondays is $e^{5.45302}$ = 233.4645. This means that the expected number of people who were killed in 2001 on Mondays only is 233.

From the Table 4.10 the selected model for the accident data using day and year can be formulated as shown in equation 4.2 below.

$$\log(mean\_killed) = 5.45302 - 0.081113D_1 - 0.214166D_2 - 0.084739D_3 + 0.096618D_4 + 0.238497D_5 + 0.087156D_6 + 0.008467Y_1 + 0.046434Y_2 + 0.278236Y_3 + 0.070783Y_4 + 0.115135Y_5 + 0.208437Y_6 + 0.153238Y_7 + 0.295600Y_8 + 0.176739Y_9 \qquad 4.2$$

where $\boldsymbol{D_1}$, $D_{\mathbf{2}}$,… D_6 represent Day2 (Tuesday), Day3 (Wednesday),…, Day6(Sunday) respectively and $\boldsymbol{Y_1}$, $\boldsymbol{Y_2}$,…, $\boldsymbol{Y_9}$, denotes 2002,2003,…,2010 in that order.

To obtain the expected number of people killed in any day in a given year, the day and the year are represented by 1 in the equation and the rest are represented by 0. For example the expected number of people killed on Tuesday in the year 2003 is given by $e^{5.45302 - 0.081113\,(1) + 0.278236\,(1)}$ =284.3350. That is, the expected number of people killed on Tuesdays in the year 2003 was 284.

4.3.3 The Ages of People Killed in Road Accidents in Ghana.

The ages of people who were killed in road accidents in Ghana from 2001-2010 were treated as categorical variable and were therefore put into groups. The results of the various models generated with their AIC's for the age groups of people killed in road accidents from 2001-2010 if illustrated in table 4.11 below

Table 4.12 The Poisson Regression Models for different categories age groups of number of People Killed in Road Accidents in Years with their AIC's

Models	AIC's
1. $\log(mean\_killed) = \alpha_1 + \beta_i Age\ group, i = 1,2, \ldots 8$	1185.9
2. $\log(mean\_killed) = \alpha_2 + \beta_i Year, i = 1,2, \ldots 10$	5487.3
3. $\log(mean\_killed) = \alpha_1 + \beta_i Age\ group + \beta_j Year, i = 1,2, \ldots 7\ and\ j = 1,2, \ldots 10$	753.77

From Table 4.12 the AIC of model 1 was found to be 1185.9, that of model 2 was 5487.3 and for model 3 it was 753.77. It is therefore obvious that model 3is chosen as the best model since it has the smallest AIC.

The parameter and their estimates, the standard errors, the Chi-squares with their associated p values are presented in table 4.13 below.

Table 4.13 The Parameter Estimates of Selected Poisson Model for Age Group of those who were killed by Road Accident in Ghana.

Coefficients	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	4.533842	0.039043	116.123	< 2e-16
Age.grp2	0.620979	0.036579	16.976	< 2e-16
Age.grp3	0.953702	0.034723	27.466	< 2e-16
Age.grp4	1.358273	0.033077	41.064	< 2e-16
Age.grp5	0.942574	0.034777	27.104	< 2e-16
Age.grp6	0.366117	0.038390	9.537	< 2e-16
Age.grp7	0.017257	0.041542	0.415	0.67785
Age.grp8	-0.332477	0.045650	-7.283	3.26e-13
ye2002	0.003552	0.037696	0.094	0.92492
ye2003	0.039768	0.037360	1.064	0.28712
ye2004	0.288571	0.035286	8.178	2.88e-16
ye2005	0.097572	0.036842	2.648	0.00809
ye2006	0.090446	0.036904	2.451	0.01425
ye2007	0.373892	0.034662	10.787	< 2e-16
ye2008	0.321619	0.035039	9.179	< 2e-16
ye2009	0.462861	0.034055	13.591	< 2e-16
ye2010	0.294421	0.035242	8.354	< 2e-16

The Table 4.13 presents the parameter estimates of the selected model. The AIC of this model was 753.77, the null deviance was 5352.66 on 79 degrees of freedom and residual deviance of 155.09 on 63 degrees of freedom following the chi-square distribution (χ^2) with one degree of freedom. The dispersion parameter was found to be 2.461746 and P-value of 9.951746e-10 which indicates that the model is significant at 5% α −level. However, the assumption of equal variance to the mean in Poisson distribution has been violated since the dispersion parameter is not approximately equal to1. The dispersion parameter of the above model is 2.46 which is far greater than 1, an indication of over dispersion in the data. This means that the parameters of the model have been over estimated and the standard errors have been under estimated and will not give a true reflection of the model for the age groups of people who were killed via road accidents in Ghana from 2001 to 2010. To address this error, Negative Binomial regression was used to modify the model to nullify the effect of over dispersion and the result is shown in the table 4.14 below.

Table 4.14 The Parameter Estimates of Negative Binomial Regression Model for Age Group of those who were killed by Road Accident in Ghana

Coefficients	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	4.544991	0.050543	89.923	< 2e-16
Age.group2	0.622941	0.047270	13.178	< 2e-16
Age.group3	0.951368	0.045862	20.744	< 2e-16
Age.group4	1.347528	0.044641	30.186	< 2e-16
Age.group5	0.941337	0.045899	20.509	< 2e-16
Age.group6	0.362424	0.048720	7.439	1.02e-13
Age.group7	0.016055	0.051231	0.313	0.7540
Age.group8	-0.334953	0.054632	-6.131	8.73e-10
year12002	0.005785	0.051913	0.111	0.9113
year12003	0.034622	0.051711	0.670	0.5032
year12004	0.287376	0.050125	5.733	9.85e-09
year12005	0.097097	0.051287	1.893	0.0583
year12006	0.094011	0.051308	1.832	0.0669
year12007	0.339428	0.049837	6.811	9.71e-12
year12008	0.314554	0.049973	6.294	3.08e-10
year12009	0.444804	0.049292	9.024	< 2e-16
year12010	0.280971	0.050161	5.601	2.13e-08

Table 4.15 Parametric comparison between Poisson and Negative Binomial Regression for Goodness of Fit Test

Assessment parameter	Poisson Regression Model	Negative Binomial Regression Model
Null Deviance	5352.66	2688.924
Degree of Freedom	79	79
Residual Deviance	155.09	72.754
Degrees of Freedom	63	63
Dispersion Parameter	2.461746	1.154825
AIC	753.77	724.51

From the Table 4.15 it is observed that the dispersion parameter for the negative binomial regression model is 1.154825 which is approximately equal to1, an indication of a good sign of reduced over dispersion in the data if not totally eliminated.

By default, R package takes the first category in a data as the base level and as such people from 0-5 years and the year 2001 were picked as the base levels for comparison in the analysis of the parameter estimates in the model as shown in table 4.13. The intercept was found to be 4.544991and very significant at 95% significant level. With the exception of people from 56-65 years group which was not significant in the model, the rest were all significant in the model at 5% α-level. It is observed that people above 65 years who were killed via road accidents was significantly smaller by -0.334953 in logarithmic count than those younger than 65years. The most significant age group in the model was found to be 26-35 years with parameter estimate of 1.347528. This means that the expected number of people killed through accident in the age group of 26-35 is $e^{1.347528}$ =3.847902 times more than that of 0-5years for every year. It could also be seen from the model that those in the age bracket of 16-25 years are the second most affected in death through road accidents in Ghana. The parameter estimate for 16-25 was found to be 0.951368 which implies that people who are killed by road accidents and are in the ages from 16-25 years are $e^{0.951368}$=2.589249 times more than those who were killed and are in the age group from 0-5 years for every year. 36-45 years who were killed via road accidents was 0.941337 more than 0-5 years in logarithmic counts for every year.

The table 4.18 further reveals that the years 2002, 2002, 2005 and 2006 were not significant in the model giving that the year 2001 is the base level. 2009 was found to be the year which had most people killed by road accident for all ages in Ghana. It was found that 2009 had $e^{-0.444804}$ =1.560184 times more people killed than 2001 for all age groups in Ghana. The model for the above table is presented in equation 4.3 below.

$$\log(mean\_killed) = 4.54499 + 0.622941A_1 + 0.951368A_2 + 1.347528A_3 + 0.941337A_4 + 0.362424A_5 + 0.016055A_6 - 0.334953A_7 + 0.005785Y_1 + 0.034622Y_2 + 0.287376Y_3 + 0.097097Y_4 + 0.094011Y_5 + 0.229428Y_6 + 0.314554Y_7 + 0.444804Y_8 + 0.280971Y_9 \qquad 4.3$$

where $\boldsymbol{A_1}$, $\boldsymbol{A_2}$,… $\boldsymbol{A_7}$ represent age group 2 (6-15 years), age group 3 (16-25 years),…, age group 7 (65+ years) respectively and $\boldsymbol{Y_1}$, $\boldsymbol{Y_2}$,…, $\boldsymbol{Y_9}$, denotes 2002,2003,…,2010 in that order.

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

5.0 Introduction

This chapter covers the summary of all the findings of this research, the various conclusions drawn from the findings and suggested alternatives and recommendations to assist reduce the number of people who are killed by road accidents in Ghana.

5.1 Conclusions

This research aimed at modeling the number of people who are killed by road accidents in Ghana.

The results also revealed that the most affected people who die through road accidents in Ghana are the youth. Out of 1909 people who are killed via road accidents on the average, 447 of them are people in the age group of 26-35, 298 are in 16-25 and 295 are those of the ages from 36-45.

This initial result was confirmed by the model 4.2 which indicated that the expected number of people who are killed in road accident and are in the ages from 26-35 is approximately four times that of the base level for every year. Those in the age groups of 16-25 and 36-45 have expected number killed to be three times that of the base level for every year. This gives us a cause to worry since the work force of the country is being killed by road accidents.

The preliminary analysis of the number of people who are killed at any given day in the week showed that most people are killed on Fridays in road accidents and Wednesday is the day which recorded the least number of people killed by road accident. It was found that on the average 341 people are killed on Friday every year and 216 are killed on Wednesday. It was found out from the Poisson mean analysis that the expected number of people who are killed in road accidents on Monday, Tuesday, Thursday, Saturday and Sunday not different from are the grand mean of the number of people killed by road accident every year.

However, the results of the Poisson regression analysis showed that Saturday has the biggest expected number of people who are killed by road accidents when time (in years) was included in the model. It was realized that the number of people killed on Saturday by road accidents is $e^{0.238497}$ times more than the base level for every year. Surprisingly, Tuesday was also found as the day which recorded the least number of casualties death in road accidents in contradiction to the Wednesday as shown by the preliminary analysis. The parameter estimate of Tuesday was -0.081113 and that of Wednesday was -0.214166 which indicates 92% and 81% reduction from the base level respectively for Tuesday and Wednesday for every year.

In the investigation of the number of people who are killed in different types of vehicles through road accidents, it was identified that buses and mini buses killed more people in road accident. This was followed by cars. The model also confirmed this through the parameter estimates for the various types of vehicles. It was observed from the model that bus/mini bus was 1.025182 times more than the base level for every year, while the rest of the vehicles were all less than the base level.

5.2 Recommendation

Looking at the state of road accidents and the number of people who are killed via road accidents in Ghana, it is recommended that;

1. Education on road accidents should be intensified especially among the youth

2. Social activities such as funerals, marriage ceremonies and festivals which are always performed at the week endings should be reduced to help minimize the number of accidents at the week endings.
3. Since the type of vehicle involved in the accident affects the number of people expected to be killed, drivers of vehicles such as cars and buses should be given special training to be able to avoid preventable accidents.
4. The accident data base of the country should be expanded to include more variable so that researcher could really determine the actual factors contributing the casualties' death in road accidents.
5. Finally, institutions that enforce road traffic regulations should do well to apply the law especially on Fridays so that all perpetrators of traffic offences shall be brought to book to deter others from repeating such offences.

REFERENCES

Afukaar F.K., Agyemang W., and Ackaar W. (2006) Estimation of levels of under –reporting of road traffic accident in Ghana. Ministry of transportation: Nation Accident management project, report. Final Report.

Afukaar F.K., Agyamang W., Debrah E.K., and Ackaah W. (2006).Estimation of the cost of road traffic accident in Ghana. Ministry of Transport: Nation management project. Final report.

Afukaar F.K., Agyemang W, and Most I. (2008) Accident statistics 2007. Building and Road Research Institution council for scientific and industrial Research, Kumasi.

Afukaar F.K., and Debrah E.K., (2007). Accident prediction for signalized intersections in Ghana. Ministry of transportation: Nation Accident management project, report. Final report

Andreassen D. (1985). Linking deaths with vehicles and population. Traffic Engineering and Control. 26:547-549.

British red cross (1997). Anyone can save a life. Road accidents and first aid. European Transport Safety Council, a strategy and road plan for the European Union, Brussels.

Broughton J. (2007). Casualty rate by type of car. TRL Report No. PPR 203. Crowthorne: TRL Limited.

Broughton J., and Walter L. (2007).Trends in fatal car accidents: Analysis of CCIS Data. TRL Report No. PPR 172. Crowthorne: TRL Limited.

Cameron A.C., and Trivedi P.K. (1998). Regression Analysis of Count Data, Cambridge University Press, Cambridge, U.K.

Cameron A.C., and Trivedi P.K. (1998). Regression Analysis of Count Data, econometric Society Monographs, No. 30, Cambridge University Press.

Chambers J.M., and Hastie T.J. eds. (1992). Statistical Models in S. Chapman & Hall, New York. This is also called the “White Book”.

Clarke D., Ward P., Truman W., and Bartle C. (2007). Fatal vehicle-occupant collisions: An in-depth study. Road Safety Report No. 75. London: Department for Transport.

Department for Transport (2006b). Road accidents in Great Britain. London: Department for Transport.

Department for Transport (2007). Trends in fatal car-occupants accidents. Road Safety Research Report No.7. London: Department for Transport.

Dinesh M. (1985) An analysis of road traffic fatalities in Delhi, India. Accident Analysis and Prevention. 17:33-45.

Emenalo M., Pusteli A., Ciampi, and Joshi H.P. (1987). Analysis of road accidents data in Zambia. Traffic Engineering and Control. 28:635-640.

Fieldwick R., and Brown R.J. (1987). The effect of speed limits on road casualties. Traffic Engineering and Control. 28:635-640.

Fletcher J.P., Baguley C.J., sexton B., and Done S. (2006) Road accident modeling for highway development counties. Main Report Trials in India and Tanzania. Report No PPRO95.

Freeman R., Hans J.P., Seizer M., Staff H.B., Marshall L., and Shack ford S., (1988). Injuries related to all vehicular accidents : A close look at head and neck frame. Laryngoscope. 98

Gardner H. (1995). Reflections on multiple intelligences. Phi Delta Kappan, 77 200-Generalized Linear Models (with J.A. Nelder),(1983), Chapman and Hall, London.

Ghee C., Silcock D., Astrop A., and Jacobs G. (1997). Socio-economic aspects of road accidents in developing countries, Transport Research Laboratory Report 247:29.

Hakkert A.S., and Mahalel D. (1978). Estimating the number of accidents at intersections from the knowledge of traffic flows from the approaches. Accident Analysis and Prevention. 10:69-79

Hauer E., Ng J.C.N., and Lovell J. (1989). Estimation of safety at signalized intersections. Transportation Research Record 1185. Transportation Research Board. Washington D. C.

Heidi W. (2006). Road accidents increase dramatically worldwide. On the web, http://www.prb.org

Home J.A., Zomer J., and Lavie P. (1990). Sleep related automobile accidents-when and who? Horne J.A. ed. Sleep 90. Buchum: Pontenagel press. 448-51

Hussain L.M., and Redmond A.D. (1997). Are pre-hospital deaths from accidental injury preventable? 308: 119-127

Jacobs G.D., and Cutting C.A. (1986). Further research on accident rates in developing countries. Accident Analysis and Prevention. 18:119-127.

Jamal R.M.A., and Jamil A.N. (2001). Casual models for road accident fatalities in Yemen. Accident Analysis and Prevention. 33:547-561.

James H.S. (2009). Poisson Regression: Multilevel Regression Modeling. Department of Psychology and Human Development: Vanderbilt University.

John G.U., and Adams (1987). Smeed's law: Some further thoughts. Traffic Engineering and Control. 28:70-73.

Kim S.H., Chung S.B., and Song K.S. (2005). Development of an accident prevention model using GLM (generalized log -lirear model) and EB Method a case of seal. Journal of the Eastern Asia society for Transportation studies 6:3669-36682

Kumar A.M.D., Sanjeau L.M.D., Agawam D.M.C.H., Kava R.M.D., and Dora T.D.M.D. (2008). Fatal road accidents and their relationship with head injuries. An epidemiological survey of five years. Indian journal of Neurotrauma (IJNT) 5(2):63-67

Livneh M., and Hakkert A.S. (1972). Some factors affecting the increase in road accidents. Accident Analysis and Prevention. 4:117-133.

MC Guiana D.R.D. (2009).The use of relationships between road accident and traffic flow in "black spot" identification. Traffic engineering and control. (22/8/9):448-453.Leong HJW (1973) Relationship between accident traffic volumes at urban Australian Road research.5(3):72-90.

McCullagh P., and Nelder J.A. (1983). Generalized Linear Models. London: Chapman and Hall.

McCullagh P., and Nelder J.A. (1989). Generalized Linear Models (Second edn). New York: Chapman and Hall.

Mekky A. (1985). Effects of rapid motorization levels on road fatalities in some rich developing countries. Accident Analysis and Prevention. 17:101-109.

Miaou S.P. (1994). The relation between track accident and geometric design of road section: Poisson vesrus negative binomial regression. Accident Analysis and Prevention. 26(4) :471-482.

Miaou S.P., and Lum H. (1993a). Modelindg vehicle accidents and highway geometric design relationships. Accident Analysis and Prevention. 25:689-1993

Miaou S.P., and Lum H. (1993b). A statistical evaluation of the effects of highway geometric design on truck accident involvements. Transportation Research Records. 1407:11-23.

Minter A.L. (1987) Road casualties improvement by learning processes. Traffic Engineering and Control. 28:74-79.

Montazeri A. (2004). Road traffic – relation mortality in Iran. A descriptive study: Pupil Heath 118: 110 – 3

Mountain, and Fatwa B. (1996). Estimating accidents at junction using routinely available input data. Traffic Engineering and control. 37 (11):624-628.

National centre on sleep disorder research and national road safety administration (1998). Drowsy driving and automobile crashes. Report HS808707. NHTSAA website. On the web http://www.nhtsa.dot.gov.

National Road Safety Commision (2007). Annual Report. National Road Safety Commission, Ghana.

Nelder J.A., Wedderburn, and RWM (1972). Generalized linear models, Journal of the statistical society.

Nelder J.A., Wedderburn, and RWM (1972). Generalized linear models. Journal of the Royal Statistical Society, Series A 135(3), 370–384.

Pack A.I., Pack A.M., Rodman E., Cucchiara A., Dinges E.F., and Schwab C.W. (1995). Characteristics of crashes attributed to driver fallen asleep. Accident Analysis and Prevention 27(6):769-75

Piet de Jong, and Gillian Z.H. (2008). Generalized Linear Models for Insurance Data, Cambridge University Press, New York

Pramada V.P., and Sarkar P.K. (1993). Variation in the pattern of road accidents in different states and union territories in India. Proceedings of the third national conference on transportation systems studies: Analysis and Policy. 1X-5 to 1X-9.

Pramada V.P., and Sarkar P.K. (1997). Models for road accidents in India. Highway Research Bulletin, Indian Road Congress. 56:1-11.

Redmond A.D. (1994). Letters. British Medical Journal. 309:57

Sagberg F., and Saetermo (1997). An investigation of behavioural adaptation to airbags and antilog brakes among taxi drivers (29 ed). Accident Analysis and Prevention. 293-302.

Sealift M. (2003).Accident prediction model for unsignalised urban junctions in Ghana. International Association of Traffic and safety (IATSS) Research 28(1).

Selifu M., Mosi I., Addae –Boafah K., and Sarpong K. (2006). Accident holistic 2005. Building and road institute. Council for scientific and industial Research, kumasi.

Smeed R.J. (1968). Variation in the pattern of accident rates in different countries and their causes. Traffic Engineering and Control. 10:364-371.

Susan C., and Partyka (1984). Simple model of fatality trends using employment and population data. Accident Analysis and Prevention. 16:21-222.

Ung C.H. (2007). Royal Government of Cambodia. Ministry of public works and transport, country report on road safety in Cambodia.

Venables W.N., Smith D.M., and R Development Team (2010). An introduction to R, Note on R: A programming Environment for Data Analysis and Graphics. Version 2.12.0 (2010-10-15).