

UNIVERSITY FOR DEVELOPMENT STUDIES

**A POISSON REGRESSION MODEL OF FATAL ROAD ACCIDENTS
ON SELECTED ROAD NETWORKS IN THE BRONG AHAFO
REGION, GHANA**

BY

LAWRENCE ACQUAH (BSc. MATHEMATICS/STATISTICS)

UDS/MAS/0009/09



**Thesis Submitted to the Department of Statistics in the
Faculty of Mathematical Sciences, University for Development Studies, in
Partial Fulfilment of the Requirements for the Award of Master of Science
Degree in Applied Statistics**

May, 2012

Dedication

This work is dedicated to my lovely wife Evelyn and children; Joseph and Nathaniel.



Declaration

Student

I hereby declare that this thesis is the result of my own original work and that no part of it has been presented for another degree in this University or elsewhere:

Signature.....



Date.....

18/10/12

Name: Acquah Lawrence

Supervisor

I hereby declare that the preparation and presentation of the thesis were supervised in accordance with guidelines on supervision of thesis laid down by the University for Development Studies.

Signature.....



Date.....

19/10/12

Dr. A. O. Adebajji



Abstract

Models for predicting fatal accidents on some selected road networks in the Brong Ahafo Region in Ghana were developed using Poisson regression model. The objectives of the study included both identification of factors contributing to the incidence of fatal accidents and the development of models to predict such occurrences on the selected road networks. The data on records of incidence of accidents in general in the years 2008 - 2010, generated by the Motor Traffic and Transport Unit (MTTU) of the Ghana Police Service, Brong Ahafo Region. Other information required for the study were taken or measured from the road networks selected for the study.

The research hypothesis states that, the occurrence of a fatal accident depends on at least a risk factor. The data was analysed using the SAS package version 9.2.2 and result that the most significant variables that contribute to the occurrence of fatal accidents include traffic volume and section length out of a total of seven variables identified besides the driver and vehicle characteristics which were not considered in this study. From the analysis, the Poisson regression model fitted the data well compared to other distributions such as the Gamma, Negative binomial, zero – inflated Poisson and the zero inflated Negative binomial models. Finally, the models developed were validated using the graphical approach of the residual analysis,



Acknowledgments

I would first and foremost thank the Almighty God so much for seeing me through to this stage. He has been the pillar and the reason behind this success story.

Next, I would like to thank my supervisor Dr. A. O. Adebajji for her guidance, insights, patience, gentle encouragement and advice which made this thesis possible. From her, I have learned a great deal and most importantly she has taught me how to research. Her enthusiasm and strives for excellence has made working with her a joy.

I am deeply grateful to the former Rector, Professor Sampson Agbozo; the current Registrar, Mr. V. Kumpeh; the Dean of the Business School, Mr. Mohammed Iddrisu Ngmenipuo; the former Head of Department, Alhaji Ibrahim – Abdul Fatawu of the Accountancy Studies and the current Head of Department, Bro. Remy Nyukorong all of Wa Polytechnic for not only approving my study leave but also helped me secure scholarship from the Ghana Education Trust Fund (GET Fund) to pursue the course, words alone cannot express my heartfelt gratitude to them, what I can only say is that may the Almighty God richly bless them.

Next, I would like to express my special thanks to Dr. Kazeem Gbolagade, the Dean of the Faculty of Mathematical Sciences for his encouraging words and taking the class through on how to conduct a meaningful research work. The knowledge acquired has helped me tremendously in carrying out this project.

I am deeply grateful to my able and very active Post Graduate Coordinator Dr. Seini for working administratively hard to ensure that we complete the course on time. On behalf of my colleagues I will say we are very much grateful.



I cannot end my acknowledgment without thanking the Motor Traffic and Transport Unit (MTTU) Officers of Ghana Police Service, Brong Ahafo Region for providing the data for the research. Without the data, this research would not have become a reality. I am very much grateful.

I would like to thank all my friends and colleagues who had helped in one way or the other in making this thesis possible. To my dear wife and children who were behind me throughout the study, I am humbled and blessed to have you by my side.





TABLE OF CONTENTS

Dedication	i
Declaration	ii
Abstract	iii
Acknowledgments	iv
Table of Contents	vi
List of Tables	xi
List of Figures	xiii
Some Definitions and Terms	xiv
Meaning of Initials and Acronyms	xiv
CHAPTER ONE INTRODUCTION	1
1.1 Background of the Study	1
1.2 Problem Statement	6
1.3 Motivation of the Study	7
1.4 Research Objectives	7
1.5 Research Hypothesis	8
1.6 Significance of the Study	8
1.7 Road Accidents in the Global View	9
1.8 Brief Historical Background of the Brong Ahafo Region	9
1.8.1 Creation of the Region	9
1.8.2 Physical Features	12
1.8.2.1 Area	12
1.8.2.2 Climate	12
1.8.2.3 Vegetation	12
1.8.3 Tourist Attraction Sites	13



1.84 Political and Administrative Structure	13
1.9 The Role of National Road Safety Commission (NRSC)	
In Curbing Road Accidents	15
1.10 Thesis Layout	17
CHAPTER 2 LITERATURE REVIEW	18
2.1 Introduction	18
2.2 Definition	18
2.3 Conceptual Facts on Factors Contributing to Accident	
Fatalities	18
2.3.1 Factors Contributing to Road Accidents as a Result	
of Human Behaviour	20
2.3.2 Factors Contributing to Road Accidents as a Result	
of Road Conditions	20
2.3.3 State of Vehicle as a Factor Contributing to	
Road Accident Fatalities	21
2.3.4 Occurrence of road accident as a result of road category	22
2.4 Review and Discussions on some of the Accident	
Prediction Models	23
2.4.1 Earlier Approaches on Accident Prediction Models	23
2.4.2 Accident Prediction Models used and Implications	25
2.4.3 Some of the Findings in Accident Prediction Studies	29
2.4.4 Recent Approaches to Analysing Accident Data	31
CHAPTER 3 RESEARCH METHODOLOGY	33
3.1 Introduction	33
3.2 Data Collection	33



3.3	Sources of Data	33
3.4	Some of the Data Captured on Traffic Accident	
	Fatality Form	34
3.4.1	Variables Considered in the Study	35
3.4.2	Description on how Some of the Primary Data	
	were Collected	35
3.4.2.1	Road Width	36
3.4.2.2	Traffic Volume	36
3.4.2.3	Road Signs and Other Road Geometrics	36
3.5	Sample	37
3.6	Statistical Methods and Analysis of Data	37
3.7	Description of Models Considered in the Study	37
3.7.1	Poisson Regression Model	37
3.7.2	Assumptions in Poisson Regression	39
3.7.3	Measure of Adequacy of Model	40
3.7.4	Over dispersion	41
3.7.5	Negative Binomial Regression	42
3.7.6	Poisson – Lognormal Model	44
3.7.7	Zero – Inflated Poisson and Negative	
	Binomial Model	44
3.7.8	Gamma model	45
3.8	Testing for Over dispersion in the Poisson	
	Regression	45
3.9	Model Development	46
3.9.1	Functional Form	46



3.9.2	Assessment of Individual Model Parameters	48
3.9.3	Goodness of Fit	49
3.9.3.1	Deviance and Pearson's Chi – Square	
	Statistic	49
3.9.4	Model selection	53
3.10	Model validation	53
3.10.1	Analysis is Residuals	53
CHAPTER 4 DATA PRESENTATION, ANALYSIS		
	AND RESULTS	55
4.1	Data Presentation	55
4.1.1	Tabular presentation of raw data	55
4.1.2	Graphical presentation of raw data	56
4.2	Univariate Analysis	59
4.3	Criteria for Choosing Model Fit Parameter	61
4.4	Testing for Over dispersion in Poisson	
	Regression	64
4.4.1	Testing for Over dispersion	65
4.5	Estimation and Significance Test of Parameters	
	from the Statistical Output	65
4.5.1	Analysis of Maximum Likelihood Parameter	
	Estimates (Techiman – Wenchi road)	66
4.5.2	Analysis of Maximum Likelihood Parameter	
	Estimates (Techiman – Tanoso road)	67
4.5.3	Analysis of Maximum Likelihood Parameter	
	Estimates (Sunyani – Brekum road)	69



4.6	Model Formulation	71
4.6.1	Formulation of Regression Models	72
4.6.2	Model Validation	73
4.7	Findings and Discussions	75
CHAPTER 5 DISCUSSIONS, SUMMARY, CONCLUSIONS AND		
RECOMMENDATIONS FOR FUTURE STUDIES		78
5.1	Discussions	78
5.2	Summary	79
5.3	Conclusions	80
5.4	Recommendation	81
5.4.1	Recommendations to the Transportation Industry	81
5.4.2	Recommendations for Further Studies	83
	References	84
	Appendices	91
	Appendix A Accident Data from Techiman – Wenchi	
	Road Network	91
	Appendix B Accident Data from Sunyani – Berekum	
	Road Network	92
	Appendix C Accident Data from Techiman – Tanoso	
	Road Network	93



List of Tables

1.1 List of Names of Municipalities and Districts in the Brong Ahafo Region	14
2.1 Summary on Some of the Previous Research and Model used in Accident Frequency data	28
2.2 Summary on Advantages and Disadvantages of Recent Models used for Analysing Accident frequency data	29
3.1 Summary of Formulas and Definitions for Model Fit Selection	52
4.1 Number of Accidents According to Accident Types on each Road Network	56
4.2 Purpose of use of Vehicle Involved in Accident Network	56
4.3 Descriptive Statistics of Fatal Accidents on Techiman – Wenchi Road	60
4.4 Descriptive Statistics of Fatal Accidents on Techiman – Tanoso Road	60
4.5 Descriptive Statistics of Fatal Accidents on Sunyani – Brekum Road	61
4.6 Criteria for assessing goodness of fit: Techiman – Wenchi road Network	62
4.7 Criteria for Assessing Goodness of Fit: Techiman – Tanoso Road Network	63
4.8 Criteria for assessing goodness of fit: Sunyani – Brekum road Network	63
4.9 Over Dispersion Test Result	65

4.10	Analysis of Data from the Techiman – Wenchi	
	Road Network	66
4.11	Type 1 Analysis Table	67
4.12	Analysis of Data from the Techiman – Tanoso	
	Road Network	68
4.13	Type 1 Analysis Table	68
4.14	Analysis of Data from the Sunyani – Brekum	
	Road Network	70
4.15	Type 1 Analysis Table	71
4.16	Parameter Estimates for the Significant Factors	71



List of Figures

1.1	Fatality Trend (1991 – 2010)	4
1.2	Brong ahafo Region Geographical Map	15
4.1	A cluster Bar Chart Showing the Number of Types of Accident in the Period under Consideration for each Road Section	57
4.2	A line graph showing the trend of accidents on Techiman – Wenchi Road	57
4.3	A Line Graph Showing the Trend of Accidents on Sunyani – Brekum Road	59
4.4	A Line Graph Showing the Trend of Accidents on Techiman – Tanoso Road	60
4.5	A Scatter Plot of Residuals Against Predicted Values for Techiman – Wenchi Road to Check Model Validation	73
4.6	A Scatter Plot of Residuals Against Predicted Values for Sunyani – Brekum Road to Check Model Validation	74
4.7	A Scatter Plot of Residuals Against Predicted Values for Techiman – Tanoso Road to Check Model Validation	74





DEFINITIONS AND TERMS

- **Fatality** occurs when a person involved in a road traffic crash dies within thirty days of the crash and as a result of the crash.
- **Fatality rate** refers to fatalities per 10,000 vehicles
- **Population risk** refers to fatalities per 100,000 population
- **Traffic Accident:** Accident in property damage, injury or death, and involves at least one vehicle on a public road.
- **Casualty** refers to a person injured or killed in a traffic crash. Thus the accident is the event whilst the casualty is the individual crash victim.
- **Fatal Accident** refers to an accident in which at least one casualty dies of injuries sustained within 30 days of occurrence.
- **Serious Injury Accident** refers to accident in which at least, one person is detained in hospital as in-patient for more than 24 hours.
- **Minor/Slight Injury Crash** refers to an accident in which the most severe injury sustained by casualty is only minor, requiring at most first-aid attention.
- **Damage-only Crash** refers to an accident which results only in vehicle or other material damage.

MEANING OF INITIALS AND ACRONYMS

- AADT – Annual Average Daily Traffic
- ADT – Average Daily Traffic
- ABS – Anti Break System
- APM – Accident Prediction Model
- BAC – Blood Alcohol Content

- BRRI - Building and Road Research Institute
- DVLA – Driver and Vehicle Licensing Authority
- EU – European Union
- GDP – Gross Domestic Product
- GLIM – Generalized Log - linear Model
- NCAP – New Car assessment Program
- NDC - National Democratic Congress
- NHTSA – National Highway Traffic Safety Administration
- NRSC – National Road Safety Commission
- OECD – Organization for Economic Cooperation and Development
- PIRE – Pacific Institute for Research and Evaluation
- WHO – World Health Organization
- WLSA – Weighted least Square Average



CHAPTER ONE

INTRODUCTION

1.1 Background of the Study

Road accidents like armed robbery have become the sword of Damocles hanging on the head of every Ghanaian, the young and the old are all at risk. Although statistics available reveal that pedestrians particularly men and children (between the ages of 4 to 10 years) are most vulnerable to road accidents, every person in Ghana stands the risk of this menace. Ghana is a hub of commercial activity in the West African sub – region and a major transit route for land locked countries like Mali, Niger and Burkina Faso. This means a constant increase in vehicular activity raises the stakes of safety on our roads.

A recent report released by the World Health Organization (WHO, 2011), reveals that road accidents kill about 1.27 million people a year globally, from which nearly half a million are pedestrians, cyclist and motorists.

It is noteworthy that Road – traffic accidents are the leading cause of death for people ages 10 to 24, and the 10th leading cause of death in the world, after heart disease, lower respiratory infections, HIV/AIDS and other conditions, according to WHO (2011). They also cause as many as 50 million nonfatal injuries a year, according to the report.

Irrespective of your social class or status, you can choose the type and model of vehicle to drive, the hospital you and your family want to attend, the school your children attend and many other things but you can never choose, even as the President of Ghana; the other person you share the road with.



Vehicle accidents are complex events involving the interactions of five major factors: the drivers, the traffic, the road, the vehicles, and the environment (e.g., weather and lighting conditions). Developing accident prediction model is a means of summarising these complicated interactive effects based on information contained in the data, as well as our engineering judgement and analytical assumptions about the accident process.

Conceptually, once the relationships between the variations of accident frequency and these five factors are established through models, one can use the models to devise cost – effective means or regulatory policies to effect or change some of these five factors in such a way that accidents will be reduced in the long run.

Road development is of high priority on the development agenda of the Government of Ghana. The Government has embarked on a massive road development programme because of the major role road transport plays in the movement of passengers and goods.

Over the years, the vehicle/population ratio in Ghana has been growing steadily. From a Vehicle / population ratio of around 31 vehicles per 1,000 population in 2002, the ratio has continued to climb to around 44 vehicles per 1,000 population in 2008. It will interest you to note that from 1991-2010, the estimated national population has grown by 63% whilst the estimated population of registered vehicles increased by 750% taking into account annual vehicle write-off (BRRI report, 2010).

In 2010, on an annual basis, the number of recorded road traffic crashes has surpassed 10,000. Within the 7-year period from 2002 to 2008, the number of people killed on Ghana's roads averaged 1,840 annually. Specifically, in 2008,



about 19 people per 10,000 vehicles were killed through road traffic crashes.

Within the same period, out of the total of 13,166 people who were killed through road traffic crashes on Ghana's road network, about 42% were pedestrians, 23% were passengers in buses and mini-buses, 12% were occupants of cars, 11% were occupants of Heavy Goods Vehicles (HGVs), about 5% were riders and passengers of bicycles, 4% were riders and passengers of motorcycles whilst 3% were occupants of pick-ups.

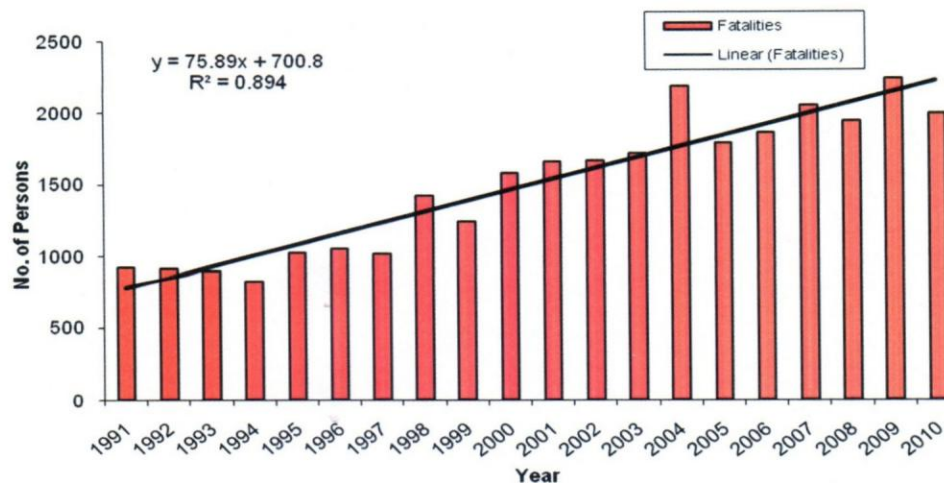
Figure 1.1 below shows accident fatality trend in Ghana (NRSC report, 2010) from 1991 to 2010.

Available statistics point to rising absolute fatalities. The existing scenario suggests the current rising trends in population growth and vehicle ownership could lead to further increases in the number of road traffic crashes if serious efforts are not made to reverse the upward trend in road traffic crashes.

In view of the magnitude of the problem of road traffic crashes and fatalities, the Government of Ghana, through the Ministry of Transportation, the National Road Safety Commission and with the collaboration of all stakeholders have been implementing data-led programmes and activities to address the road safety challenges of the country. Over the years, the interventions have achieved significant gains. Ghana's fatality rate of around 36 per 10,000 vehicles in 1996 has dropped to 18.76 per 10,000 vehicles in 2008. These gains could have been greater if there was a documented Road Safety Policy.



Figure 2a Fatality Trends (1991-2010)



SOURCE: BRRI REPORT (2010)

Fig 1.1: Fatality trends (1991 – 2010)

The number of road traffic crashes in 2010 (i.e. 11,506) represents a decrease of 6.4% over the 2009 figure and a marginal increase of 1.9% over the 2001 figure. The number of fatal crashes and their resulting fatalities in the past year also saw a decrease. Compared to the 2009 figures, fatal crashes decreased in 2010 by 5.8% and fatalities by 11.2%. In addition, there was a significant decrease of 8.6% in the overall number of casualties in 2010 compared with 2009. Relative to the year 2001 (i.e. NRSS I, 2001), the 2010 figures for fatal crashes, fatalities and overall casualties recorded corresponding increases of 34.1%, 19.6% and 13.9%, respectively.

The development of a transportation system has been the generative power for human beings to have the highest civilization above other creatures on the earth. However, this growth has caused safety problems that the transportation systems for our efficiency and comfort rob us of our lives, so efforts to overcome this issue have been and will be made by many persons in various fields. Especially, transportation engineers have made passionate researches to



analyze accidents and predict potential accidents in our systems. The motivation behind this study is to develop an accident prediction model in order to serve as a turnover point of these endeavours.

Research shows that, fatal accidents are preventable because the top contributing factors in most automobile accidents are human behaviour.

Referring to it as a *just killer*, road accident is no respecter of persons. In 2003, two members of Parliament belonging to the former president Kuffuor's administration unnaturally exited this wonderful world owing to road accidents. Media reports confirmed that former President Kuffuor's convoy was involved in many serious accidents during his tenure of office of which about six (6) of his security officers lost their lives. The former vice President in Kuffuor's regime also experienced deadly motor accidents one of which occurred somewhere in the Brong Ahafo Region of Ghana which claimed the life of a poor teenager.

The former President, Flt. Lt. Jerry John Rawlings was not immune from motor accidents. The infamous case in which a mini bus drove into his convoy, killing four (4) body guards on the spot has not been forgotten.

Some honourable members of Parliament such as Alhaji Kwadwo Mahama, Edward Osei Kwaku and John Achuliwor and others lost their lives to motor accidents. The passing away of Bishop Kwadwo Owusu and Ferdinand Ayim through road accidents still lingers in the mind of their loved ones. Referring to archives reminds us of Boniface Adams an aspiring deputy National propaganda secretary of the NDC who died through motor accident on 3rd January 2010.

Ghanaian celebrities such as Suzy Williams, a fast – rising actress, Terry Bonchaka, Kwame Owusu Ansah and others who died in motor accidents have indeed left a great vacuum in the Ghanaian Film and Music Industry. The death of three Urologists Prof. J.M.K Quartey, Dr. Isaac Bentsi and Dr. Benjamin Osei Wiafe of Korle – bu Teaching Hospital on Kumasi – Accra road on 27th August 2005 is not forgotten yet. The death of the children and some members of Rev. Eastwood Annaba of the Fountain Gate Chapel on good Friday in 2009 and in fact countless similar reported cases are still fresh in our memory.

1.2 Problem Statement

In developing countries, motorisation has been accompanied by rapid growth in road traffic injuries, which have become a leading cause of death and disability. According to the World Health Organisation (WHO, 2010), 1.3 million persons are killed and an additional 30-50 million are injured in road traffic crashes worldwide. Majority (over 85%) of the road traffic deaths occurred in low and middle income countries where 81% of the world's population live and own about 20% of the world's vehicles, costing the global community US\$518 billion annually.

Thus road accidents are as much a major threat to public health in Ghana as they are an enormous drain on the national economy. Improving safety on Ghana's roads is therefore a pressing national concern that has already found expression in the setting-up of the National Road Safety Commission (NRSC). Therefore there is the need to develop a working and effective accident predictive model using road geometries and other information peculiar to the terrain of the country that can well inform policy makers, road engineers and



other stakeholders on how best to reduce if not prevent the human resource and other valuable property from destruction through road accidents.

1.3 Motivation of the Study

The motivation here is the opportunity to design a crash prediction model using the Generalized Linear Model, to be precise, the Poisson Regression Model to help salvage one of the accident prone regions in the country in an attempt to reduce drastically the prevalence of road accidents. The use of Poisson Regression to fit the model was as a result of extensive use of other model fitting techniques by numerous researchers to fit accident prediction models and some of the weaknesses identified. Secondly, this model takes care of data that do not obey the conditions of normality. Models used by other Researchers are briefly discussed in **sections 2.2.2** under Literature Review.

1.4 Research Objectives

The first objective of this study is to identify road geometric variables and other factors influencing the occurrence of fatal accidents.

The second objective is to establish statistical valid model that can explain the variations in the occurrence of fatal accidents as a function of some factors and other road geometric variables.

Road Engineers and other Stakeholders in the Road Industry will be able to use these models and other findings to plan and take decisions that will improve safety measures on our roads.



1.5 Research Hypothesis

The hypotheses in this study are as stated below:

H_0 : the incidence of a fatal road accident is not as a result of a risk factor

H_1 : the occurrence of a fatal road accident depends on at least one risk factor

Mathematically, this can be stated as:

$H_0 : \beta_i = 0, \text{ where } i = 1, 2, 3, \dots, n$

$H_1 : \text{at least } \beta_j \neq 0, \text{ where } j = 1, 2, 3, \dots, n$

where β_i are defined as the coefficients of the explanatory variables (X's) in the model

1.6 Significance of the Study

The increasing annual number of road traffic accidents on the Ghanaian road network is a serious concern to the government and the citizenry as a whole. Accidents entail not only economic cost, but also the social cost of pain, grief and suffering of families of the victims. Establishing the cost of road traffic accidents thus creates the compelling need for investments in road safety measures in road design improvements, education campaigns, enforcement activities and research.

It is safe to say that practitioners (road traffic authorities, policy makers, their consultants) are interested in improving road safety and taking measures that will decrease the number of accidents on (their) roads. Therefore they want to know what the expected numbers of accidents will be in the near future. It is also likely that they are interested in measures that can prevent large numbers of accidents at low costs.



With the development of Accident Prediction Model (APM) one can estimate the expected number of accidents on a road (type) as a function of traffic volume and a set of risk factors (explanatory variables).

The purpose of developing this crash prediction models is to provide a realistic estimate of expected accident frequency as a function of traffic volume and roadway geometries over a highway segment. Development of such estimates is a critical component in the consideration of safety in highway planning and design.

1.7 Road Accidents in the Global View

While traffic death rates in many developed countries have stabilized or declined in recent years, they are relatively rising in many other parts of the world according to the report by WHO. More than 90% of traffic deaths occur in developing countries, even though those countries have fewer than half of the world's vehicles.

The highest death rates from road accidents are in Eastern Mediterranean and African countries, the report said. The lowest rates were found in the Netherlands, Sweden, the U.K and other developed countries.

1.8 Brief Historical Background of the Brong Ahafo Region

1.8.1 Creation of the Region

The Brong Ahafo Region was created on 4th April 1959 (by the Brong Ahafo Region Act No. 18 of 1959). The Act defined the area of the Brong Ahafo Region to consist of the Northern and the Western part of the then Ashanti Region and included the Prang and Yeji areas that before the enactment of the





Act formed part of the Northern Region. Before the Ashanti Empire was conquered by the British in 1900, the Brong and Ahafo states to the north and northwest of Kumasi (the capital of Ashanti Empire and the present Ashanti Region) were within the empire.

Nana Akumfi Ameyaw III traces his ancestry to King Akumfi Ameyaw I (1328-63), under whose reign the Brong Kingdom with its capital at Bono Manso grew to become the most powerful kingdom of its time. Indeed oral tradition has it that nearly all the different groups of the Akans, including the Asante, trace their origins to Bono after migrating from the “north”.

The first remembered King of the Bono Kingdom is King Asaman, who is credited with leading his Akan people from what may be present day Burkina Faso or even further north, to Bonoland (Buah, 1998). Later migrations led to the Asantes, Fantes, Denkyira and other Akans settling in their present locations.

Nana Akumfi Ameyaw is credited with the creation of gold dust as a currency and gold weights as a measure, later developed and adopted by all the other Akan groups, particularly the Asante. Legend has it that he even supported his yam shoots with sticks made of pure gold.

It was when King Opoku Ware of Asante defeated Bono in 1723 and destroyed Bono Manso that the capital moved to Techiman (Takyiman). Techiman and other Bono states therefore came under the Asante Empire until 1948 when Akumfi Ameyaw III led the secession of Bono from Asante, supported by other Bono states such as Dormaa.

The most significant change the British administration in Ashanti brought to the people of the Brong and Ahafo states until 1935 was that it made them



independent of Kumasi clan chiefs (Busia, 1951). The British administration worked out a strategy that severed the interference of the Kumasi clan chiefs with the internal affairs of the Brong and Ahafo states.

When the Ashanti Confederacy was restored in 1935 by the British administration, however, most of the Brong and Ahafo states saw that their independence from Ashanti was being threatened, because by restoring the Ashanti Confederacy, they were to revert to their former overlords in Kumasi. Though the Brong states joined the Ashanti Confederacy, most of them were not happy with the re-union because they felt their long historical association with Ashanti had brought them nothing. The opportune time came when in 1948 Nana Akumfi Ameyaw III, the Omanhene (paramount chief) of Techiman led Techiman to secede from the Ashanti confederacy (Austin, 1964). The secession of Techiman was supported by some of the Brong states and this led to the formation of the dynamic Brong political movement, Brong Kyempem Federation. The movement was formed in April 1951 at Dormaa Ahenkro under the auspices of the Dormaa State. The main objective of the movement was to struggle for a separate traditional council and a separate region for the Brong Ahafo states.

The name of the movement was later changed to the Brong Kyempem Council. In March 1955, the Prime Minister informed the National Assembly that the government was considering “the possibility of setting up a Brong Kyempem Council” to fulfil the desire of the Brongs for the establishment of a development committee for their area and that the government would “examine the case for the establishment of two administrative regions for Ashanti”.

In March 1959, the Brong Ahafo Bill was passed under a certificate of urgency by Parliament. The Brong Ahafo Region Act was enacted after receiving the Governor General's assent. Sunyani was made the capital of the new region.

1.8.2 Physical Features

1.8.2.1 Area

Brong Ahafo, with a territorial size of 39,557 square kilometres, is the second largest region in the country (16.6%). The region shares boundaries with the Northern Region to the north, the Volta and Eastern Regions to the south-east, Ashanti and Western Regions to the south, and Cote d'Ivoire to the west. The central point of the landmass of Ghana is in the region, at Kintampo.

1.8.2.2 Climate

The region has a tropical climate, with high temperatures averaging 23.9°C (750F) and a double maxima rainfall pattern. Rainfall ranges, from an average of 1000 millimetres in the northern parts to 1400 millimetres in the southern parts.

1.8.2.3 Vegetation

The region has two main vegetation types, the moist semi-deciduous forest, mostly in the southern and south eastern parts, and the guinea savannah woodland, which is predominant in the northern and north eastern parts of the region. The level of development and variations in economic activity are largely due to these two vegetation types. For example, the moist semi-deciduous forest zone is conducive for the production of cash crops, such as cocoa and cashew. Timber is also an important forest product, produced mainly in the Ahafo area around Mim, Goaso and Acherensua. Other cash crops grown





in the forest area are coffee, rubber and tobacco. The main food crops are maize, cassava, plantain, yam, cocoyam, rice and tomatoes.

1.8.3 Tourist Attraction sites

The ecology of the region has produced lots of tourist attractions. The Pumpum River falls 70 metres down some beautiful rocky steps to form the Kintampo Falls, as it continues its journey towards the Black Volta. The Fuller Falls, 7 kilometres west of Kintampo, also provides a scenic beauty as River Oyoko gently flows over a series of cascades towards the Black Volta. Another scenic site is the River Tano Pool which houses sacred fish that are jealously protected by the local community who live along the river near Techiman. There is also a pool on the Atweredaa River, which runs through the Techiman market.

Another type of tourist attractions are caves, sanctuaries and groves. The Buabeng-Fiema Monkey sanctuary, located 22 kilometres north of Nkoranza, covers a forest area of 4.4 square kilometres. It serves as home for black and white colobus and mona monkeys. Buoyem caves, which are hidden in a dry semi-deciduous forest, house a large colony of rosetta fruit bats. The Pinihini Amovi caves are also historic underground caves near Fiema.

The tourist attraction sites in the region cannot be complete without mention of the Tanoboase Sacred Grove. It is believed that the grove is the cradle of Brong civilization. The grove served as a hideout to the Brongs during the 18th century Brong-Ashanti wars.

1.8.4 Political and Administrative Structure

Brong Ahafo has seven (7) administrative municipalities and fifteen (15) districts, with Municipal Chief Executives (MCEs) and District Chief

Executives (DCEs) as the political heads. The MCEs and DCEs are assisted by Municipal / District Co-ordinating Directors (MCDs/DCDs) who are responsible for the day to day running of the districts.

The M/DCEs work under the Regional Minister (the political head of the region), while the M/DCDs are under the Regional Coordinating Director. Sunyani is the administrative headquarters of the region, where the Regional Minister resides.

Table 1.1 and Figure 1.2 below give the list of the names of the Municipalities, districts, and the geographical map of the region.

Table 1.1: List of Names of Municipalities and Districts in the Brong Ahafo Region

DISTRICT	MUNICIPAL
Asunafo South	Sunyani
Asutifi	Techiman
Atebubu – Amantin	Wenchi
Dormaa East	Asunafo North
Jaman North	Berekum
Jaman South	Dormaa
Kiintampo South	Kintampo
Nkoranza North	
Nkoranza South	
Pru	
Sene	
Sunyani West	
Tain	
Tano North	
Tano South	





Fig 1.2: Brong Ahafo Region Geographical Map

1.9 The Role of National Road Safety Commission (NRSC) in Curbing Road Accidents

Improving safety on Ghana's roads is a pressing national concern that has already found expression in the setting-up of the National Road Safety Commission (NRSC). The Commission has been tasked to initiate and oversee the implementation of a more proactive and structured programme of accident reduction. The initial efforts in this direction have culminated in the National Road Safety Strategy and Action Plan both of which underscore the need for *data-led interventions and innovative approaches* to understanding the occurrence mechanisms and *determining factors of road accidents* and



devising strategies to reduce the incidence of accidents on all manner of roads.

One of the strategies may be the design of accident predictive models using road geometries that can inform stakeholders in the sector.

Predictive models for accidents have been researched extensively in the world including Ghana and other African countries. The purpose of developing accident prediction models is to enable us provide a realistic estimate of expected accident frequency as a function of traffic volume and other roadway geometries over a highway segment. Development of such estimates is a critical component in the consideration of safety in highway planning and design.

Growth in urbanization and in the number of vehicles in many developing countries has led to the increase in traffic accidents on road networks which were never designed for the volumes and types of traffic which they are now required to carry. In addition, unplanned urban growth has led to incompatible land uses, with high levels of pedestrians and vehicle conflicts. The drifts from rural areas to urban centre often result in large numbers of new urban resident unused to such high traffic levels. Many developing countries continue to repeat the mistakes of the industrialized countries , many still permit linear development with direct access from frontage properties along major roads even though this is known to lead to safety problems. It is possible to identify hazardous sections of the road network so that appropriate remedial measures can be undertaken to reduce the likelihood and severity of accidents at those locations. This has proven to be one of the most cost-effective ways of improving road safety in industrialized countries.



1.10 Thesis Layout

The research is organized into five chapters. Chapter One contains the introduction. Chapter Two is devoted to literature review. Chapter Three presents the research methodology. Chapter Four contains data presentation and analysis. Chapter Five is devoted to summaries, conclusions and recommendations for future work.



CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

The intention of this literature review is to give a literature survey of some of the most important risk factors and concepts in road accidents as well as review of methods used by other researchers to analyse data on accidents. The literature for this study is focused on the principal risk factors that influence traffic accidents globally, extensive examination and discussion of approaches adopted by other researchers across the globe to analyse counts data such as accident data and development of accident prediction models.

2.2 Definition

Road traffic accident is a rare, random, multi factor event always preceded by a situation in which one or more road user have failed to cope with their environment, resulting in a collision on the public highway recorded by police (Radin, 2005). Transportation is an essential part of modern existence, linking the various activities in which people participate especially at home, work, school and go to shopping or recreation. The problem here is, if it is not properly and carefully used can become a devastating instrument for its users.

2.3 Conceptual Facts on Factors Contributing to Accident Fatalities

Many factors contribute directly or indirectly to the occurrence of road accidents some of which will be mentioned and discussed here. Nevertheless, at any given time, accident risk is affected by personal risk factors of the driver (e.g., hours of sleep the previous night, condition of eyesight), vehicle risk



factors (e.g., brake adjustment and conditions of tyres), environmental factors (e.g., weather and roadway features) and perhaps most important risks created by other drivers and traffic. Driver errors can be violations of rules, mistakes of judgment, inattention errors, or inexperience errors. Common driver errors resulting in accidents include recognition errors (failure to perceive a crash threat) and decision errors (risky driving behaviour such as tailgating), or poor decision-making in dynamic traffic situations - such as trying to cross a stream of traffic (Michael et al, 2004).

Dewer and Olson (2002) gave another common classification for driver errors resulting in accident as follows: Rule-based (failure to obey rules or regulations), Knowledge-based (failure to understand required safe behaviour), Skill-based (lack of proper skills to perform the task). Drivers can also make mistakes without obvious misbehaviours, such as failure to see another vehicle or misjudgement of a gap in the traffic stream. Red-light running may be regarded as a rule based misbehaviour if it is intentional, a skill-based mistake if it is not. (Michael S. et al, 2004).

Reason (1990) proposed three error categories: violations (deliberate deviations), mistakes (intended action with unintended consequences), and lapses/slips (execution of unintended action).

Rimmo (2002) has expanded this by splitting the lapses/slips category into inattention errors (unintended action resulting from recognition failure) and inexperience errors (unintended action resulting from lack of knowledge or skill). Rimmo's classification, with examples, follows: Violations (Deciding to drive when known to be very fatigued, deliberately exceeding speed limits, accelerating at green-to-yellow signal change).



Mistakes (Misjudging gap when crossing traffic, misjudging speed of oncoming vehicle, and misjudging stopping distance)

Inattention Errors (Failing to notice red light at intersection, failing to see that vehicle has stopped in lane ahead, and Failing to notice sign) Inexperience Errors (Having to check gear with hand, Driving in too low a gear and Switching on wrong appliance in truck).

2.3.1 Factors Contributing to Road Accident as a Result of Human Behaviour

Different studies confirm the fact that human factor (road users) is one of the major contributory factors to accidents. At the same time, any error in the system and on roads will lead to unsafe road user behaviour.

In Ghana, various studies have established that human error (which takes 70-93%) is the most cited contributory factor underlying the occurrence of road traffic accidents. Unacceptable road user behaviour, difficult working conditions for commercial vehicle drivers, deficiencies in the driver licensing system and inadequate enforcement of road safety laws and regulations have culminated into acts of impatience, recklessness, irresponsibility and non-compliance on the part of drivers.

2.3.2 Factors Contributing to Road Accidents as a Result of Road Conditions

There are many dangers lurking on our roadways when they are not properly maintained then unsuspecting drivers may be in jeopardy from these hazards. A study ordered by the Pacific Institute for Research and Evaluation (PIRE) has



found that poor road conditions are the leading reason for highway fatalities. Because of the poor condition and lack of maintenance of American roads, the roads contribute to approximately 31.4 percent of all accidents each year in America, according to the study by PIRE. Even if road conditions are not the sole cause of the accident, the conditions can cause crash injuries to be more severe. Poor road conditions greatly elevate the severity of an injury from minor to moderate or even fatal.

Motorways have the lowest risk on injury accidents compared to other types of roads because of the separation between vehicle movements according to their speed (no high speed variance). Elvik and Vaa (2004), shown that the rate of injury accidents per million vehicle kilometres of travel on motorways is about 25% of the average for all the public roads. Road surface conditions, poor road surface, defects in road design and maintenance contribute to an increase in the risk of accidents. Bester (2001) reported that countries with more paved roads will lead to lower fatality rates.

2.3.3 State of Vehicle as a Factor Contributing to Road Accident

Fatalities

Brand new cars tend to have more safety and protection features, such as air bags, anti-brake system (ABS), etc. Studies conducted by other researchers show that there is relation between vehicle age and risk of a car crash. One study (WHO, 2004) showed that occupants in cars manufactured before 1984 have almost three times the risk of new cars. Many developed countries improved vehicle crash worthiness and safety, which means the protection that a vehicle gives its passengers (and to the Vulnerable Road Users) from a crash.



Many countries in the European Union (EU) as well as USA have set out legislation for safety standards in motor vehicles, for instance the New Car Assessment Program (NCAP), where vehicle crash performance is evaluated by rating the vehicles models according to their safety level for occupant protection, child protection and pedestrian protection. Vehicle defects increase the risk of accident.

2.3.4 Occurrence of Road Accident as a Result of Road Category

Here, we define the road categories as either a rural or an urban road network. Urban roads will have more accidents and fewer fatalities or severity per kilometre travelled than rural roads, because of the density of vehicles and the lower speeds of travel. Hakkert and Braimaister (2002) have shown in one macro-study that countries with a high level of urbanisation will have higher population densities and they may experience lower levels of fatalities and serious injuries.

Page (2001) has found that the population who live in urban areas have fewer road accident fatalities than other places. Bester (2001) also reported similar results that countries with higher road densities will have fewer fatality rates. Shorter distances to medical services can explain this. In 2006, the fatality rate per 100 million vehicle miles travelled was reported as being 2.4 times higher in rural areas of the United States as compared to urban areas (NHTSA, 2008). Investigations that were carried out in the 1980s in several different states found that rural traffic injury fatality rates were higher than urban rates. Studies from other countries have documented higher occurrences of traffic crash fatalities in urban areas as well. A study in Ghana reported that the majority of traffic fatalities and injuries occurred in rural areas and that the crashes that



occurred on rural roads were generally more severe (Afukaar et al., 2003) and a study in Quebec, Canada found that severe crashes are more common in rural areas (Thouez et al., 1991).

2.4 Review and Discussions on Some of the Accident Prediction Models

Predictive models for accidents have been researched intensively in the world of which Ghana is not off the hook.

An Accident Prediction Model is a mathematical formula describing the relation between the safety level of existing roads (i.e. crashes, victims, injured, fatalities etc.) and variables that explain this level (road length, width, traffic volume etc)

2.4.1 Earlier Approaches on Accident Predicting Models

In 1949, R. J. Smeed developed a regression model (log-linear model) and he found an inverse (or negative) relationship between the traffic risk (fatality per motor vehicle) and the level of motorisation (number of vehicles per inhabitant). This regression represented the best estimates of the mean values of traffic risk for each given value of motorisation (what is called least square). This shows that with annually increasing traffic volume, fatalities per vehicle decrease. Smeed concluded that fatalities (F) in any country in a given year are related to the number of registered vehicles (V) and population (P) of that country by the following equation.

$$\frac{F}{V} = \alpha \left(\frac{V}{P} \right)^{-\beta} \quad (2.1)$$

where

F = number of fatalities in road accidents in the country

V = number of vehicles in the country

P = population



$$\alpha = 0.003, \beta = 2/3$$

This formula became popular and has been used in many studies. It is often referred to as Smeed's formula. This nonlinear relationship can be translated to a linear one by taking the logarithms of the two sides:

$$\log Y = \log \alpha - \beta \log X \quad (2.2)$$

where Y is $\frac{F}{V}$ and X is $\frac{V}{P}$.

The number of fatalities can be derived from Smeed's formula as:

$$F = c.V^\alpha.P^\beta \quad (2.3)$$

where c , α , β are parameters and they are estimated from data by using the least square method.

Some authors followed the equation of estimating the regression parameters (α , β) of the data by calculating the country's road safety performance in comparison to other countries; Jacobs and Hutchinson (1973), found that Smeed's formula can give a close estimation of the actual data and it can be applied to different sample sizes of countries and years with the use of different values of α and β . Some authors have tried to develop Smeed's formula and its accuracy further by including several socio-economic variables in the model.

Fieldwick (1987) has included speed limits in the same model. The number of registered vehicles has been replaced by the total vehicle kilometre driven in many studies (e.g. Silvak, 1983). This measure (vehicle kilometre driven) was not available at the time of Smeed's study.

At the same time, many studies have criticised Smeed's model because it only concentrates on the motorisation level of country and ignores the impact of other variables, (e.g Broughton, 1988), where according to Smeed's model, population and vehicles are the only country values, that influence the number



of fatalities. This means that road safety measures have no meaning because road fatalities can simply be predicted from population and vehicle numbers in any country and any year.

Adams (1987) criticised the model's accuracy because there would always be a decline in traffic risk for any increase in the number of vehicles, but generally in non-linear way. These models are not able to incorporate the effect of risk factors on accident involvement.

In a study to find the relationship between accidents and traffic flow as a measure of exposure, Tanner (1953), suggested that accident numbers were approximately proportional to the root of the product of the two-way major road traffic volume and turning flows from the minor road.

2.4.2 Accident Prediction Models Used and Implications

Oppe (1989) used multiple linear regression models. In these models the dependent variable (either number of accidents or accident rate) is a function of a series of independent variables such as speed or traffic volume. Accident occurrence in these models is assumed to be normally distributed. These models generally lack the distributional property that is necessary to describe adequately the random and discrete vehicle accident events on the road and they are inappropriate for making probabilistic statements about accident occurrence.

Generally, Multiple Linear Regression model has various shortcomings to use for predicting the number of accident. First, dependent variables are assumed to follow normal distribution in this model, but the number of accident is in counts. And it is assumed that there is no relation between error and



independent variable, but this assumption is not always true in case of accident in actuality. In addition, this model can predict a negative number that can not appear as the number of accident. Moreover, when accident did not happen in any spot, this method always predicts zero as the number of accident, and this result strains the truth that zero number means that spot absolutely safe.

Oppe, assumes that fatality rates follow a negative exponential learning function in relation to the number of vehicle kilometres and time. This method has been found to be most effective when the components describing the time series behave slowly over time as follows:

$$\ln\left(\frac{F_t}{V_t}\right) = \ln(R_t) = \alpha t + \beta \quad (2.4)$$

or equivalently $R_t = e^{\alpha t + \beta}$, Where the 'ln' function is the natural logarithm, F_t is the number of fatalities for some country in a year t, V_t is the number of vehicle kilometres travelled in that year, R_t is $\frac{F_t}{V_t}$ and α, β are constants. This

means that the logarithm of the fatality rate decreases (sign of improvement) if α is negative proportional with time. This model is called the negative exponential learning model, where β is supposed to be less than zero. Both α and β are the parameters to fit.

The traffic volume will increase quickly first and at the end it will reach its saturation level, which differs from country to country.

Saccomanno and Buyco (1988) and Blower et al (1993) used a Poisson log linear model to explain variations in accident rates. This Poisson regression model is especially suitable for handling data with large numbers of zero counts. However, this model could be inappropriate for road accident counts,





since it fails to account for extra-Poisson variation (the value of the variation could exceed the value of the mean) in the observed accidents counts.

Bernardo and Ivan (1997) studied the prediction of the number of crashes versus the crash rate using Poisson regression. They suggested that the Poisson distribution allows for the relationship between exposure and crashes to be more accurately modelled as opposed to the linear relationship assumed in crash rate prediction.

Poisson models deal with discrete data so that they have most of desirable characteristics to describe vehicle collisions of positive number and random attribution. However, these models may produce wrong coefficients and wrong standard errors if data has excessive variance, and it has been a problem to apply this model because variables that explain the number of accident were categorized data.

To solve this problem of extra-Poisson variation, several authors such as Miaou (1994) developed two types of negative binomial models, one using a maximum likelihood method and one using a method of moments. The maximum likelihood model was found to be more reliable than the Poisson regression model in predicting accidents where over dispersion is present.

Logistic regression (a type of regression where the dependent is a categorical as opposed to a numerical variable) has been the most popular technique in developing injury severity prediction models.

Lui and McGee (1998) used logistic regression to analyze the probability of fatal outcomes of accidents given that the crash has occurred.

In yet another study a logistic regression approach was used to examine the contribution of individual variable to the injury severity (Al-Ghamdi, 2002).

The study was limited to 560 accidents obtained from the police records in Riyadh, Saudi Arabia. According to the logistic regression results, out of nine independent variables used in this study, only two were found to be statistically significant with respect to the injury severity: location and cause of accident. For example, the odds of being in a fatal accident at a non-intersection location are 2.64 higher than those at an intersection, and the odds of severe injury increases on accidents caused by over-speeding and entering the wrong way traffic.

Table 2.1 below lists some of the Authors of previous researches conducted on traffic crash using some of the models relevant for analysing such data.

Table 2.1: Summary on some of the Previous Research and Model used in Accident -Frequency Data

Type of model	Previous Research using the model
Poisson	Jovanis and Chang (1986); Joshua and Garber (1990); Jones et. al. (1991); Miaou and Lum (1993); Miaou (1994)
Negative binomial/Poisson gamma	Maycock and Hall (1984); Hauer et al. (1988); Bonneson and McCoy (1993); Miaou (1994); Persaud (1994); Maher and Summersgill (1996); Mountain et al. (1996); Mountain et al. (1998); Turner and Nicholson (1998); Miaou and Lord (2003); Amoros et al. (2003);
Poisson-lognormal	Miaou et al. (2005); Lord and Miranda-Moreno (2008); Aquero Valverde and Jovanis (2008)
Zero-inflated Poisson and negative binomial	Miaou (1994); Shankar et al. (1997); Carson and Mannering (2001); Malyshkina and Mannering (2010a)
Conway-Maxwell-Poisson	Lord et al. (2008); Sellers and Shmueli (2010)
Gamma	Oh et al. (2006); Daniels et al. (2010)

Table 2.2 gives a summary of some of the weaknesses and strengths of the recent accident prediction models used by Researchers.



**Table 2.2: Summary on Advantages and Disadvantages of Recent Models
used for Analyzing Accident -Frequency Data**

Type of model	Advantages	Disadvantages
Poisson	Most basic model; easy to estimate	Cannot handle over- and under-dispersion; negatively influenced by the low sample mean and small sample size bias
Negative binomial/Poisson gamma	Easy to estimate can account for overdispersion	Cannot handle under-dispersion; can be adversely influenced by the low sample mean and small sample size bias
Poisson-lognormal	More flexible than the Poisson-gamma to handle over-dispersion	Cannot handle under-dispersion; can be adversely influenced by the low sample mean and small sample size bias (less than the Poisson-gamma); cannot estimate a varying dispersion parameter
Zero-inflated Poisson and negative binomial	Handles datasets that have a large number of zero-crash observations	Can create theoretical inconsistencies; zero inflated negative binomial can be adversely influenced by the low sample mean and small sample size bias
Conway-Maxwell-Poisson	Can handle under- and over-dispersion or combination of both using a variable dispersion (scaling) parameter	Could be negatively influenced by the low sample mean and small sample size bias; no multivariate extensions available to date
Gamma	Can handle under-dispersed data	Dual state model with one state having a long term mean equal to zero

2.4.3 Some of the Findings in Accident Prediction Studies

Zeeger (1998), based on data for two-lane of 5,000 miles of 7 states in the US, developed an accident model with subordinate variable of accident rates by accident types and independent variables of the whole width of shoulder, the width of lane, road vertical alignment, and average daily traffic volume. The result showed that accident rate was decreasing with smooth vertical



alignment, as less ADT and wider of lane, whole shoulder, as more less accidents.

In a study by Bonneson and McCoy (2001), they developed accident prediction model according to each condition by distinguishing separation and non separation of left-turn lane separating roads away from median separation facilities. As a result, they proposed that accidents were affected by AADT, length of roads, density, and land use and so on.

Mustakim, et al (2008), in their research into Black Spot study and Accident Prediction Model using Multiple Linear Regression showed that the percent accident reduction by changing the measures of each variables are, one access point per kilometre reduction can reduce accidents by 9.32 %, 5 kilometre per hour speed reduction can reduce accidents by 27.2%, 100 vehicle per hour volume reduction can reduce accidents by 4.33 % meanwhile an increment of one second in gap will reduce accident by 1.20%.

In his study (Ackaah, 2010) on crash prediction model for two – lane Rural Highways in the Ashanti Region of Ghana, reported that traffic flow, highway segment length, junction density, terrain type and presence of a village settlement within road segments were found to be statistically significant explanatory variables ($p < 0.05$).

In addition to his findings, adding one junction to a 1km section of road settlement was found to increase injury crashes by 32.0% and sections which had a village settlement within them were found to increase injury crashes by 60.3% compared with segments with no settlements.



According to the study of KimTae-wan (1996), on a model to predict road accidents he found that accidents were reducing as increase in the number of lanes.

Also, according to the study by Walton et al (1996), it has proposed that in case of establishing median barrier on the existing roads, median encroachment accidents could be reduced.

2.4.4 Recent Approaches to Analysing Accident Data

Classical least-squares (ordinary) regression techniques were used in developing some of the early accident predictive models (McGuigan, 1999). However, recent research has shown that ordinary least squares regression has some statistical properties that are undesirable for accident data analysis. These include the intrinsic assumption of homoscedascity (i.e. equal variance of the error terms for all values of the predictor variable) and the possibility of predicting accident frequency with negative values. In reality, accident counts are sporadic, discrete and non negative and their occurrence pattern would be more akin to a Poisson process, like any count data.

Incidentally, an attribute of the Poisson distribution, namely, the mean of the predicted variable is equal to its variance, does not usually hold when a substantial proportion of a database comprises zero accident counts, as is often the case in accident prediction modelling. With over-dispersed data (i.e. when the mean is less than the variance), Miaou and Lum observed that the Poisson model tends to produce inaccurate estimates. As a solution to this problem, the authors recommended the adoption of the Negative Binomial distribution, a more general probability distribution, which relaxes the constraints on the



mean and variance. In other more recent studies (Mountain, et al, 1996) and (Kulmala, 1995) the technique of “generalised linear models”, using the software package GLIM (McCullagh, and Nelder, 1989) has facilitated the use of more generalised probability distributions like the Negative Binomial. The GLIM approach is preferable because, among other things, it allows the representation of accident counts as coming from the family of exponential distributions, from which one can be chosen to correspond to the data used and it yields maximum likelihood estimates of parameters, i.e. values of parameters that are most likely to have given rise to the accident data.

Usually, Log linear models are considered as a basic method to analyze effects of categorized data. Basically, this model can be expressed as follows:

$$\ln Y_i = \beta X_i \quad i = 1, 2, \dots, n \quad 2.5$$

where, Y_i : Number of accidents for combination i

X_i : Independent variable i

β Parameters in the formula measure relation in X_i and display degrees of explanation power about accidents.

Reason why this model is widely used is that, elements affecting accidents are categorized data. Moreover, because accidents are discrete essentially, expression of difference about accident reaction is most efficient in expressed data system by categorized style. And log linear methods make it possible to test significance of categorized data as fixed quantity. In addition, it can handle positive characteristic of accident by Poisson distribution. However, weighted least squares method (WLSA) that uses exact calculation in this model usually displays high residual and needs large sample generally.



CHAPTER THREE

RESEARCH METHODOLOGY

3.1 Introduction

This chapter describes the method used to get the data required for the study. It also describes how the data was analysed to arrive at a solution to the objectives set. The description of the model adopted is also discussed under this chapter.

3.2 Data Collection

Selected number of accident prone roads in the Brong Ahafo Region connecting district capitals were considered in the study. A purely random sample of road networks for the study in the region, arguably, may not have guaranteed the inclusion of some key features and the adequate data required for the study. Specifically selected highway roads were used in this study. The road network considered include; Techiman – Tanoso road, Techiman – Wenchi road and Sunyani – Berekum road. The main reasons for selecting these road networks include the following:

- (i) They were accident prone road networks in the region and
- (ii) Availability of data on accidents within period for which the study was conducted.

3.3 Sources of Data

The data for this study was gotten from both primary (on the road network where accident occurred) and secondary sources (Police filled accident report form) of data.



The initial plan was to collect all information needed for the study from the Motor Traffic and Transport Unit (MTTU) of the Ghana Police Service in the Brong Ahafo Region. But it so happened that we could not obtain all the information needed. Hence information such as the Annual Daily Traffic Volume, nature of the road network and information on some of the road geometries were taken from the road network under consideration.

Information obtained from the Motor Traffic and Transport Unit (MTTU) apart from those taken from the field was treated as secondary data. Three years consecutive data, containing detailed information on road accidents – fatal, serious and minor crashes for the period, 2008 to 2010 that had occurred on any of the chosen road networks which was captured on the Police Accident Report, was used in this study. The period 2008 to 2010 was considered as the study period because prior to this period proper recording and management of data on accidents in almost all the MTTU offices visited were either not available or a lot of missing data. Hence the choice to consider the period 2008 to 2010.

3.4 Some of the Data Captured on Traffic Accident Fatality Form

The Traffic Accident Fatality Form of the MTTU unit collects data surrounding the occurrence of the traffic fatality. The following data are recorded on the form: Driver ID, Sex, Age, Date, Time, and address where the event occurred, Number of fatal victims in the same event; Type of traffic accident (including but not limited to a collision with a fixed or moving object, pedestrian struck, or a overturned vehicle); Role of the victim during the event (Including but not limited to: Driver of the vehicle, passenger, pedestrian); Use

of safety equipment; Type of vehicles involved in the event for both the victim and the counterpart involved (motorized vehicles, bicycle, motorcycle, pedestrian); and Alcohol level of the victim. Other variables include conditions, consequences, traffic and specific information about each "party" to the accident, age and type of driving license and vehicle characteristics.

3.4.1 Variables Considered in the Study

This study considered only these variables from both secondary and primary data:

- Date on which accident occurred.
- Accident severity (fatal, serious, minor or damage only).
- Number of victims who died or got injured.
- Place of accident (road linking two district capitals or towns).
- Number of vehicles involved in the accident.
- Purpose or use of the vehicle (private or commercial).
- Traffic volume (number of vehicles plying that road network per day).
- Number and type of road signs.
- Road geometries (length of road network, width of road, number and type of intersections (e.g. $\perp$ – junctions, Y – junctions, $\frac{\perp}{\perp}$ - junctions etc)).
- Road separation (median or no median).
- Road surface condition at the time of the accident (wet, dry, muddy etc).
- Road surface type (tarred good, tarred with potholes, gravels, earth with potholes, earth, etc).



3.4.2 Description on How Some of the Primary Data were Collected

Under this section, we give a vivid description on how some of the primary data in the study were collected.

3.4.2.1 Road Width

A road network was divided into five or more sections based on the data collected.

The width of a section (that is the distance between the left and the right edges of the road network which excludes the shoulders of the road) was taken at ten (10) different points/locations along the road section. It was ensured that the distances between these locations were evenly spaced and the average of the measurements taken as the width of that road section. We requested for the assistance of a policeman and one technical person on road construction to assist in the measurement of the width of the road network.

3.4.2.2 Traffic Volume

The road section traffic volume (number of vehicles plying on the road section per day) was taken each day for seven days for a particular road network by taking the traffic volume at three different locations in three different periods of the day (from 7.00a.m to 9.00am; from 12.00p.m to 2.00p.m; from 5.00p.m to 7.00p.m). The necessary conversions were made and the average found to represent the traffic volume for that particular road network.





3.4.2.3 Road Signs and Other Road Geometrics

The number of different types of road signs and other road geometries on identified road network sections were taken as we moved along such roads. In this case such features were observed and recorded if present.

3.5 Sample

As mentioned earlier, we selected three accident prone road networks in the Brong Ahafo Region for this study (Techiman – Wench, Techiman Tanoso and Sunyani – Berekum). Each road network was divided into five or more unequal sections based on the data of accident frequency on that stretch of road network.

3.6 Statistical Methods and Analysis of Data

From the literature surveyed, it appears most of the recent research works in accident prediction models were analysed using the Generalised Linear Model technique. Using the software package GLIM facilitates the use of more generalised probability distributions like the Negative Binomial Distributions. We used the SAS 9.2 version analysis tool in running the data collected and also excel version 2007 was used to draw some of the graphs required for the analysis.

3.7 Description of Models Considered in the Study

3.7.1 Poisson Regression Model

The Poisson regression is a member of a class of generalized linear models, which is an extension of traditional linear models that allows the mean of a

population to depend on a linear predictor through a nonlinear link function and allows the response probability distribution to be any member of an exponential family of distributions (McCullagh and Nelder, 1989).

Because accident frequency data are non-negative integers, the application of standard ordinary least-squares regression (which assumes a continuous dependent variable) is not appropriate. Given that the dependent variable is a non-negative integer, most of the recent thinking in the field has used the Poisson regression model as a starting point. In a Poisson regression model, the probability of roadway entity (segment, intersection, etc.) i having y_i accidents per some time period (where y_i is a non-negative integer) is given by:

$$P(y_i) = \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!}, \quad i = 0, 1, 2, \dots \quad 3.1$$

where $P(y_i)$ is the probability of roadway entity i having y_i crashes per time period and λ_i is the Poisson parameter for roadway entity i , which is equal to roadway entity i 's expected number of crashes per period, $E[y_i]$.

Under Poisson regression analysis we consider situations where the outcome variable is numeric but not normally distributed. In real life one often encounters situations where the outcome variable is numeric, but in the form of counts. Often it is a count of rare events such as the number of new cases of lung cancer occurring in a population over a certain period of time. The aim of regression analysis in such instances is to model the dependent variable Y as the estimate of outcome using some or all of the explanatory variables (in mathematical terminology estimating the outcome as a function of some explanatory variables).



When the response variable had a normal distribution we found that its mean could be linked to a set of explanatory variables using a linear function like

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n \quad (3.2)$$

In the case of **binary regression** the fact that probability lies between 0-1 imposes a constraint. The **normality assumption** of multiple linear regression is lost, and so also is the **assumption of constant variance**. Without these assumptions the F and t tests have no basis. The solution was to use the logistic transformation of the probability p or logit p , such that

$$\log_e \left(\frac{p}{1-p} \right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n \quad (3.3)$$

The β coefficients could now be interpreted as increasing or decreasing the log odds of an event, and $\exp\beta$ (the odds multiplier) could be used as the odds ratio for a unit increase or decrease in the explanatory variable. In survival analysis we use the natural logarithm of the hazard ratio, that is

$$\log_e \left(\frac{h(t)}{h_0(t)} \right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n \quad (3.4)$$

When the response variable is in the form of a count we face yet different constraint. Counts are all positive integers and for rare events the Poisson distribution (rather than the Normal) is more appropriate since the Poisson mean > 0 . So the logarithm of the response variable is linked to a linear function of explanatory variables such that

$$\log_e Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n \quad (3.5)$$

and so

$$Y = (e^{\beta_0})(e^{\beta_1 x_1})(e^{\beta_2 x_2}) \dots (e^{\beta_n x_n}) \quad (3.6)$$

In other words, the typical Poisson regression model expresses the log outcome rate as a linear function of a set of predictors.



3.7.2 Assumptions in Poisson Regression

The assumptions of Poisson regression include:

- i. Logarithm of the event rate changes linearly with equal increment increases in the exposure variable.
- ii. Changes in the rate from combined effects of different exposures or risk factors are multiplicative.
- iii. At each level of the covariates the number of cases has variance equal to the mean.
- iv. Observations are independent.

Methods to identify violations of assumption (iii) i.e. to determine whether variances are too large or too small include plots of residuals versus the mean at different levels of the predictor variable. Recall that in the case of normal linear regression, diagnostics of the model used plots of residuals against fits (fitted values). This means that the same diagnostics can be used in the case of Poisson Regression.

The typical Poisson Regression Model expresses the natural logarithm of the event or outcome of interest as a linear function of a set of predictors. The dependent variable is a count of the occurrences of interest e.g. the number of cases of fatal accidents that occur over a period of follow-up. Typically, one can estimate a rate ratio associated with a given predictor or exposure. A measure of the goodness of fit of the Poisson regression model is obtained by using the deviance statistic of a base-line model against a fuller model.



3.7.3 Measure of Adequacy of Model

The most widely used regression model for multivariate count data is the log-linear model (McCullough and Nelder, 1989):

$$\log[E(Y_i)] = \log t_i + \beta' x_i \quad (3.7)$$

where β is a vector of regression coefficients, x_i is a vector of covariates for subject i , so called offset variable $\log t_i$ is needed to account for possible different observation periods (t_i) for different subjects.

The popular measures of the adequacy of the model fit are deviance and Pearson Chi-Square (χ^2). If statistical model is correct then both quantities are asymptotically distributed as χ^2 statistics with $n-p$ degrees of freedom (df); where n is number of subjects and p is the number of fitted parameters.

Thus if the regression model is adequate, the expected value of both the deviance and Pearson Chi-Square is equal (or close) to $n-p$ (both the scaled deviance $\cong 1$ or the scaled Pearson Chi - Square: $X^2/df \cong 1$), otherwise there could be doubt about validity of the model.

One important characteristic of counts is that the variance tends to increase with the average size of the counts. The main feature of the Poisson model is that expected value of the random variable Y_i for subject i is equal to its variance: $\mu = E(Y_i) = \text{Var}(Y_i)$

3.7.4 Overdispersion

It is possible to account for over dispersion with respect to the Poisson model by introducing a dispersion parameter ϕ into the relationship between the variance and the mean; that is

$$\text{Var}(Y_i) = \phi \mu. \quad (3.8)$$





This method is based on a quasi-likelihood approach, which permits estimation of parameters and inferential testing without full knowledge of the probability distribution of the data (Wedderburn, 1974, McCullagh and Nelder, 1989). The scale parameter in the 'Analysis of Parameter Estimates' table of output is equal ϕ . When $\phi = 1$ we have the ordinary Poisson model and when $\phi > 1$ we have the overdispersed Poisson model. The introduction of the dispersion parameter, however, does not introduce a new probability distribution, but just gives a correction term for testing the parameter estimates under the Poisson model. The models are fit in the usual way, and the parameter estimates are not affected by the value of ϕ , but the estimated covariance matrix is inflated by this factor. This method produces an appropriate inference if overdispersion is modest (Cox, 1983) and it has become the conventional approach in Poisson regression analysis. McCullagh and Nelder (1989) suggested estimating the dispersion parameter ϕ as a ratio of the deviance or the Pearson

Chi-Square to its associated degrees of freedom. SAS implemented this approach by introducing an option SCALE= in the model statement of PROC GENMOD. One can estimate dispersion (scale) parameter by either specifying SCALE=DEVIANCE (=D, or just DSCALE) or SCALE=PEARSON (=P, or just PSCALE) and then appropriately adjust standard errors of regression coefficients.

Clearly, ignoring over-dispersion in the analysis would lead to underestimation of standard errors, and consequent over-statement of significance in hypothesis testing. Thus, concluding that using inappropriate model for count data can

dramatically change a statistical inference. The over dispersion must be accounted for by the analysis methods appropriate to the data.

3.7.5 Negative Binomial Regression

Another count data regression analysis model, which allows for overdispersion, is the negative binomial model (NB). The negative binomial distribution can be derived from the Poisson when the mean parameter is not identical for all members of the population, but itself is distributed with gamma distribution. In other words, the NB distribution provides one way of modelling heterogeneity in a population. The relationship between variance and mean for NB distribution has the form

$$\text{Var}(Y_i) = \mu + \phi \mu^2 \quad (3.9)$$

Where ϕ is an additional distribution parameter that must be estimated or set to a fixed value. The NB model is only an exponential family when ϕ is known.

The Negative Binomial (NB) distribution can be obtained from the mixture of Poisson and Gamma distribution and is expressed as

$$P(y_i / x_i) = \frac{\Gamma(y_i + \frac{1}{\phi})}{y_i! \Gamma(\frac{1}{\phi})} \left[\frac{1}{1 + \phi \lambda_i} \right]^{\frac{1}{\phi}} \left[\frac{\phi \lambda_i}{1 + \phi \lambda_i} \right]^{y_i} \quad (3.10)$$

for $y_i = 0, 1, 2, 3, \dots$

where y_i is the number of crashes for road segment i , λ_i is the expected number of crashes per period, which can be expressed as $\lambda_i = \exp(x_i' \beta)$

The mean and variance of negative binomial distribution are respectively,

$$E(y_i / x_i) = \lambda_i \text{ and } V(y_i / x_i) = \lambda_i [1 + \lambda_i \phi] > E(y_i / x_i)$$

One important characteristic of the NB distribution is that it naturally accounts for overdispersion due to its variance which is always greater ($\phi > 0$) than the variance of a Poisson distribution with the same mean, λ .



For this reason the NB model has greater flexibility in modelling the relationship between the expected value and the variance of y_i than the highly restrictive Poisson model. Note that, for small ϕ , the NB model approaches the Poisson model.

Although NB distribution is not in the exponential family, starting from version 7 of SAS, PROC GENMOD includes the possibility to run NB regression, by specifying option DIST=NB in the model statement. SAS is accounting for possible “residual” overdispersion by including additional scale parameter ϕ to the relationship $\text{Var}(y_i) = (\lambda + \phi \lambda^2)$.

3.7.6 Poisson-Lognormal Model

Recently, some researchers have proposed using the Poisson-lognormal model as an alternative to the negative binomial/Poisson-gamma model for modeling crash data (Miaou et al., 2003; Lord and Miranda-Moreno, 2008; Aquero-Valverde and Jovanis, 2008). The Poisson-lognormal model is similar to the negative binomial/Poisson-gamma model, but the $\exp(\varepsilon_i)$ term used to compute the Poisson parameter is lognormal rather than gamma distributed.

3.7.7 Zero-inflated Poisson and Negative Binomial Model

Zero-inflated models have been developed to handle data characterized by a significant amount of zeros or more zeros than one would expect in a traditional Poisson or negative binomial/Poisson-gamma model. Zero-inflated models operate on the principle that the excess zero density that cannot be accommodated by a traditional count structure is accounted for by a splitting



regime that models a crash-free versus a crash prone propensity of a roadway segment.

3.7.8 Gamma Model

The gamma model has been proposed by Oh et al. (2006) to analyze crash data exhibiting underdispersion (see also Cameron and Trivedi, 1998). This model can handle overdispersion and underdispersion and reduces to the Poisson model when the variance is roughly equal to the mean of the number of crashes. Although this model performs well statistically, it is still a dual-state model, with one of the states having a long-term mean equal to zero. The gamma model has seen limited use since it was first introduced by Oh et al. (2006).

3.8 Testing for overdispersion in the Poisson Regression

Deviance and Pearson Chi-Square divided by the degrees of freedom are used to detect overdispersion or underdispersion in the Poisson regression. Values greater than 1 indicate overdispersion, that is, the true variance is bigger than the mean, values smaller than 1 indicate underdispersion, the true variance is smaller than the mean. Evidence of underdispersion or overdispersion indicates inadequate fit of the Poisson model. We can test for overdispersion with a likelihood ratio test based on Poisson and negative binomial distributions. This test tests equality of the mean and the variance imposed by the Poisson distribution against the alternative that the variance exceeds the mean. For the negative binomial distribution, the variance = mean + ϕ mean² ($\phi > 0$, the negative binomial distribution reduces to Poisson when $\phi = 0$).



The null hypothesis is:

$$H_0: \phi = 0$$

And the alternative hypothesis is:

$$H_1: \phi > 0.$$

To carry out the test, we followed the steps below:

- Run the regression model using negative binomial distribution, record LL (log likelihood) value.
- Record LL (log – likelihood) for the Poisson model.
- Use the LR (likelihood ratio) test, that is, compute LR statistic, using the expression

$$-2(LL(\text{Poisson}) - LL(\text{negative binomial})).$$

The asymptotic distribution of the LR statistic has probability mass of one half at zero and one half – Chi-sq distribution with 1 df (Cameron, et. al, 1998). To test the null hypothesis at the significance level α , use the critical value of Chi-sq distribution corresponding to significance level 2α , that is reject H_0 if LR statistic $> \chi^2_{(1-2\alpha, 1 \text{ df})}$.

3.9 Model Development

The intention here is to develop a model that will relate the average accident frequency in a period to the best set of explanatory variables. Therefore a multiple regression approach was adopted within the framework of Generalised Linear Models (GLMs).

3.9.1 Functional Form

The theory of GLMs allows the variation in the dependent variable to be separated into the systematic and random parts (McCullagh et al. 1989). As a





consequence, it is possible to make structural and distributional assumptions, which describe these two types of variations respectively (Kulmala, 1995).

The structural assumption indicates that the expected value of the response variable can be related through a “link function” to a set of explanatory variables and their coefficients. On the other hand, random variation is described by a “random error term” associated with the model, which reflects the distributional properties of the response variable. The ordinary linear model tackles both the distributional and structural assumptions together and assumes the response variable to be Normally-distributed, quantitative and continuous and capable of taking any values. These run counter to the basic properties of accident counts, which are discrete, non-negative and generally governed by a non-stationary Poisson process (Jovanis and Chang, 1986).

Poisson regression models are estimated by specifying the Poisson parameter λ_i (the expected number of accidents per period) as a function of explanatory variables, the most common functional form being $\lambda_i = \exp(\beta \mathbf{X}_i)$, where $\mathbf{X}_i$ is a vector of explanatory variables and β is a vector of estimable parameters.

In general, the link between the Poisson mean and the covariates or independent variables is given as:

$$E(\lambda_i) = kQ^a \exp(\sum \beta_j x_{ij}) \quad (3.11)$$

Where $E(\lambda_i)$ is the expected number of the occurrence of the event (in a specified period)

Q – a measure of exposure.

k , β_j , and α – the model parameters or regression coefficients to be estimated (β_j represents the regression coefficient corresponding to the j -th explanatory variable other than traffic flow).

X_{ij} – the j -th explanatory variable other than measure of exposure for the i -th road section ($i = 1, n$; n being the total number of road sections in the study).

In accordance with the GLM framework, Equation (1) is transformed into the prediction mode using a log - link function as follows:

$$\ln[E(\lambda_i)] = \log k + \alpha \log Q + \sum \beta_j x_{ij} \quad (3.12)$$

Where, all parameters are as defined in Equation (3.11) above

By specifying the dependent variable, the model form, error distribution (in this case Poisson or Negative Binomial), the potential explanatory variables and the link function, the model is fitted, as the coefficients (model parameters) of the specified variables are estimated using the method of maximum likelihood.

3.9.2 Assessment of Individual Model Parameters

After developing the model, individual model parameters were generally assessed at two levels. The first test was to ensure that the estimated parameter coefficients were statistically significant. Thus, the ratio of the estimated coefficient to its standard error was required to pass the t -test at 5 per cent level of significance. The other aspect was to examine whether a parameter's contribution to the reduction in deviance is significant. That is, we found out whether the addition of the said parameter to the model increases the explanatory power of the model significantly. This assessment was made



simple using an output generated by the statistical software used to analyse the data upon a request for type1 analysis. SAS version 9.2.2 gives the likelihood ratio (LR) statistics upon request.

3.9.3 Goodness of Fit

3.9.3.1 Deviance and Pearson's Chi – Square Statistic

Two statistics that are helpful in assessing the goodness of fit of a given generalized linear model are the scaled deviance and Pearson's chi – square statistic.

(a) Deviance

The quality of the fit between the observed values (y) and predicted values ($\hat{y}$) can be measured by various test statistics, however, one of the useful statistic is called deviance and defined as:

$$D(y : \hat{y}) = -2[L(\hat{y}; y) - L(y; y)] \quad (3.13)$$

For a better model, one would expect smaller value of the $D(y : \hat{y})$ (McCullah and Nelder, 1987) and (Agresti, 2002).

Consequently, the scaled deviance is defined as:

$$\frac{D(y : \hat{y})}{\phi} \quad (3.14)$$

(b) Pearson's Chi – Square Statistic

Pearson's chi – square statistic is defined as

$$\chi^2 = \sum_i \frac{w_i(y_i - \mu_i)^2}{v(\mu_i)} \quad (3.15)$$

and the scaled Pearson's chi – square is χ^2 / ϕ

The scaled version of these statistics, under certain regularity conditions, has a limiting chi – square distribution, with degrees of freedom equal to the number



of observations minus the number of parameters estimated. The scaled version can be used as approximate guide to the goodness of fit of a given model.

3.9.4 Model Selection

(a) AIC/AICC

A commonly used measure of goodness of fit is the Akaike Information Criterion (AIC) (Akaike, 1973, 1974, 1976). AIC is a way of selecting a model from a set of models. The principle is to select a model that minimizes the negative likelihood penalized by the number of parameters; formula is indicated in table 3.0.

This construct is very useful in comparing and selecting non - nested model specifications, but we should not over rely exclusively on AIC criterion. Many authors have noted that the AIC has a strong bias towards models that over fit with extra parameters since the penalty component is obviously linear with increase in the number of explanatory variables, and the log likelihood often increases more rapidly (Carlin and Louis, 1996, p. 49; Neftci, 1982, p. 539; Sawa, 1978, p.1280).

However, a substantial benefit is that by including a penalty for increasing degrees of freedom, the AIC explicitly recognizes that basing model quality decisions on the value of likelihood alone is a poor strategy since the likelihood never decreases by adding more explanatory variables regardless of their inferential quality.

An alternative form of the AIC is the corrected AIC (AICC) which takes into account sample size by, especially, increasing the relative penalty for model complexity with small data sets. It is defined as:



$$AICC = -2(\ln(\text{likelihood})) + 2p^* \left(\frac{n}{(n-p-1)} \right) \quad (3.16)$$

Where n is the sample size. As n gets larger, AICC converges to AIC

($n-p-1 \rightarrow n$ as n gets much bigger than p , and so $\frac{n}{n-p-1}$ approaches 1),

and so there is no harm in using AICC regardless of the sample size (Burnham, 2002). A derived formula for calculating AICC is given in table 3.0

(b) BIC

Another commonly used measure of goodness of fit is that proposed by Schwartz (1978), referred to as both the Schwartz Criterion and the Bayesian Information Criterion (BIC). Even though it is derived from a different statistical perspective, the BIC resembles the AIC in calculation; formula is indicated in table 3.0.

Despite the strong visual similarity expressed between the AIC and the BIC, the two measures can indicate different model specifications from a set of alternatives as being optimal, with the AIC favouring more explanatory variables and a better fit and the BIC favouring fewer explanatory variables (parsimony) and a poorer fit (Koehler and Murphree, 1988, p. 188; Neftci, 1982, p. 537; Sawa, 1978, p. 1280). Since BIC exclusively includes sample size in the calculation, it is obviously more appropriate in model comparisons where sample size differs, and a model that can achieve a reasonable log likelihood fit with a smaller sample is penalized less than a comparable model with a larger sample. Whereas the AIC is just a convenient construction loosely derived from Maximum likelihood and negative entropy (Amemiya, 1985, p.



147; Grene, 1997, p. 401; Koehler and Murphree, 1988, p. 189), the BIC is strongly connected with Bayesian theory.

There are other competing model selection criteria as well as modifications of AIC and BIC that provide useful comparisons, but the basic Akaike and Schwartz constructs dominate the empirical work. Although it can be shown that nearly all of these measures are asymptotically equivalent (Zhang, 1992), Amemiya (1980) provides simulation evidence that the BIC finds the correct model more often than the AIC for small samples. However, Neftci's (1982) study found that the BIC was noticeably more sensitive to transformations on the data than the AIC.

Since these measures are only one part of assessment of model quality and neither has remarkably superior properties, the choice of which to use is primarily a function of personal preference. Amemiya states his preference on the AIC due to its simplicity (1981, p. 1505) and statistical software packages generally give the AIC as the default measure.

The table below gives some of the statistics used in model selection and their corresponding formulas.

Table 3.1: Summary of Formulas and Definitions for Model Fit Selection

Statistic	Definition or Formula
AIC	$AIC = n \ln \left(\frac{SSE}{n} \right) + 2p = -2LL + 2p$
AICC	$AICC = 1 + \ln \left(\frac{SSE}{n} \right) + \frac{2(p+1)}{n-p-2} = -2LL + 2p \frac{n}{n-p-1}$
BIC	$BIC = n \ln \left(\frac{SSE}{n} \right) + 2(p+2)q - 2q^2 = -2LL + p \log(n)$



where:

N = Number of observations

p = Number of parameters including the intercept

LL = Log likelihood estimated at the value of the estimated parameters.

AIC = Akaike's Information Criterion

AICC = Corrected Akaike's Information Criterion

BIC = Bayesian Information Criterion

SSE = Error Sum of Squares

$q = \frac{n\hat{\sigma}^2}{SSE}$ and $\hat{\sigma}^2$ = Estimate of pure error variance from fitting the full model.

3.10 Model Validation

3.10.1 Analysis of Residuals

The residuals from a fitted model are the differences between the responses observed at each combination values of the explanatory variables and the corresponding prediction of the response computed using the regression function. Mathematically, the definition of residual for the i^{th} observation in the data set is written

$$e_i = y_i - f(x_i; \hat{\beta}), \quad (3.17)$$

with y_i denoting the i^{th} response in the data set and x_i the vector of explanatory variables, each set at the corresponding values found in the i^{th} observation in the data set.

If the model fit to the data were correct, the residuals would approximate the random errors that make the relationship between the explanatory variables and the response variable a statistical relationship.



Therefore if the residuals appear to behave randomly, it suggests that the model fits the data well. On the other hand, if non – random structure is evident in the residuals, it is a clear sign that the model fits the data poorly.

There are many statistical tools for model validation, but the primary tool for most modelling applications is graphical residual analysis. Different types of plots of residuals from a fitted model provide information on the adequacy of different aspects of the model.

In this study, in an attempt to test the sufficiency of the functional part of the model used to fit the data, a scatter plot of residuals versus predictors was used.



CHAPTER FOUR

DATA PRESENTATION, ANALYSIS AND RESULTS

This chapter gives a presentation of the raw data, analysis of the data and some results as a result of the analysis made.

4.1 Data Presentation

The raw data taken from the MTTU at the District and the Municipal Police Offices were daily recordings of occurrence of accidents, but for the purpose of this study, the daily recordings were converted into six – monthly recordings. This means we will have six different periods over which data were taken for the three years duration (i.e. from 2008 to 2010). A period corresponds to six (6) months.

4.1.1 Tabular Presentation of Raw Data

Table 4.1 gives a summary of the total accidents that occurred on each of the road networks under consideration, categorized into the three types of accidents covering the period over which the study was conducted.

Table 4.1: Number of Accidents According to Accident Types on each Road Network

PERIOD	Techiman - Wenchi road			Sunyani - Berekum road			Techiman - Tanoso road		
	FATAL	SERIOUS	MINOR	FATAL	SERIOUS	MINOR	FATAL	SERIOUS	MINOR
Jan '08 – Jun '08	2	7	16	3	6	10	2	11	13
Jul '08 – Dec '08	1	10	13	2	14	12	3	12	40
Jan '09 – Jun '09	6	7	11	9	5	8	5	16	37
Jul '09 – Dec '09	3	10	11	6	9	10	5	15	28
Jan '10 – Jun '10	3	11	7	4	7	7	10	7	16
Jul '10 – Dec '10	3	11	11	2	7	13	4	6	11
TOTAL	18	56	68	26	48	60	29	67	145
% OF TOTAL	12.59	39.16	48.25	19.4	35.82	44.78	12.03	27.8	60.17



It also shows the percentage representation of each accident type on each of the road networks considered in the study. This table shows that, Techiman – Tanoso road network witnessed more accidents (241 incidents) as compared to the other two road networks. The corresponding graphical representation shows the same fact during period considered in the study.

Table 4.2 below shows the purpose of use of vehicle involved in accident during study period. It is obvious (using the raw data) that in all the three road networks considered, vehicles meant for commercial use were involved in accident more than those meant for private use. This raises alarm of the calibre of professional drivers we are training to take charge of our lives when we board such vehicles. It also raises questions on the credibility of issuance of license to drivers who use these commercial vehicles on our roads as well as the law enforcement agencies if they are up to the task of ensuring safety on our roads by doing the right thing.

Table 4.2: Purpose of use of Vehicle Involved in Accident

Road network	Purpose of use of vehicle		
	Commercial use	Private use	Total
Techiman - Wenchi	101	73	174
Sunyani - Berekum	109	70	179
Techiman - Tanoso	201	130	331

4.1.2 Graphical Presentation of Raw Data

The line graphs (Figure 4.2, 4.3 and 4.4) and the cluster bar graph (Figure 4.1) below portray the trend of each of the types of accidents on a particular road network and the dominating type of accident on the road networks considered in the Bron Ahafo Region respectively.



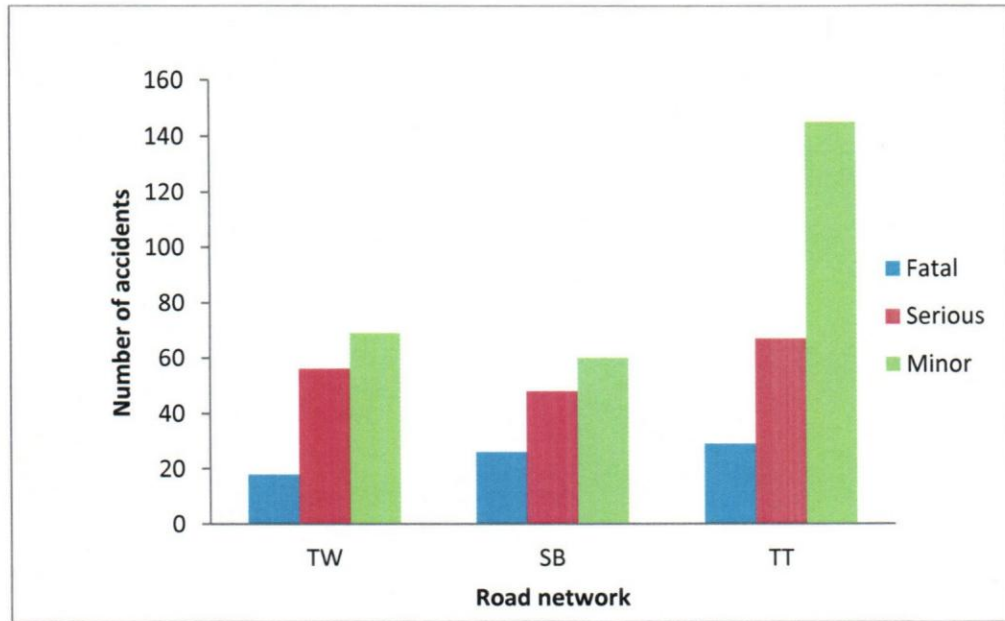


Fig 4.1: A cluster Bar Chart Showing the Number of Types of Accident in the Period Under Consideration for Each Road Section

Key: TW = Techiman – Wenchi road; SB = Sunyani – Berekum road; TT = Techiman – Tanoso road

Figures 4.2, 4.3 and 4.4 try to give the general trend of accidents occurring on each of the road networks from 2008 to 2010.

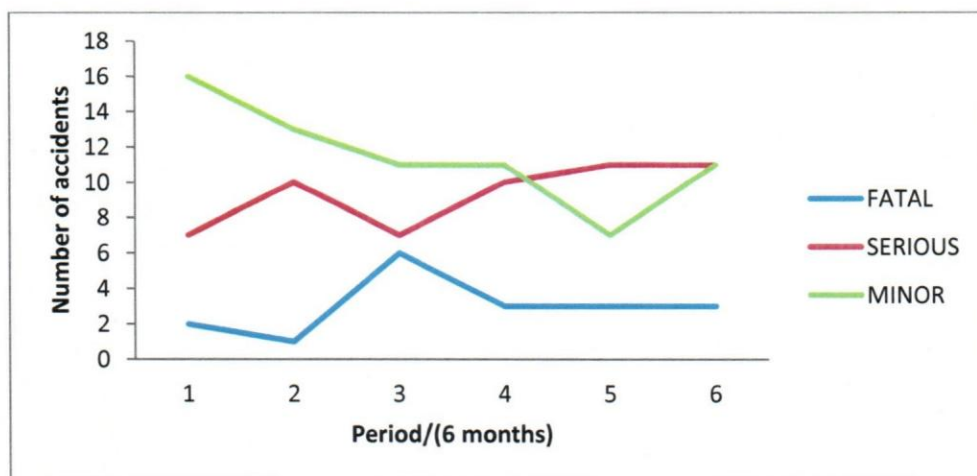


Fig 4.2: A Line Graph Showing the Trend of Accidents on Techiman - Wenchi Road

From figure 4.2, none of the line graphs representing the types of accidents shows a clear cut downward trend with the exception of the minor accident line graph which shows a relative reduction starting from 2008 down to the early part of 2010 but rises again towards the end of 2010.

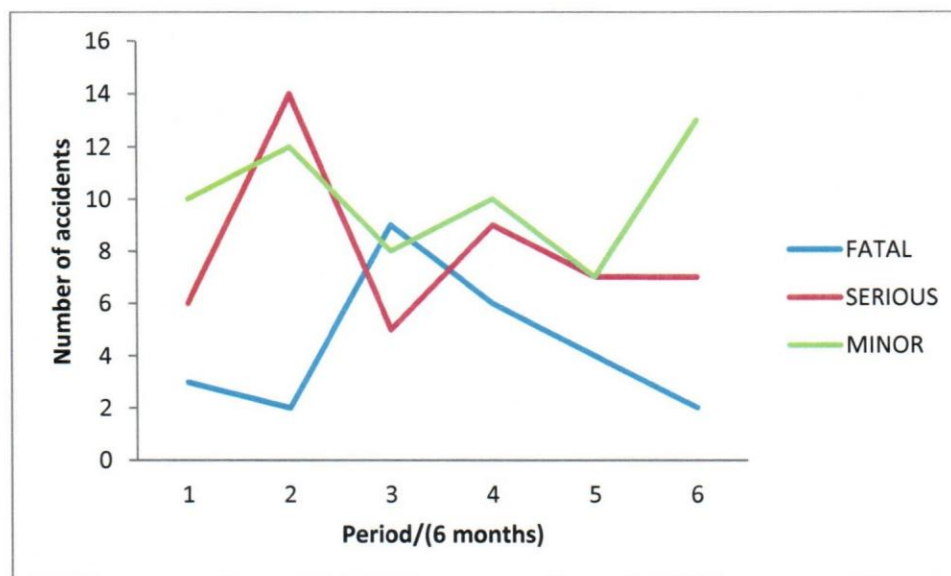


Fig 4.3: Aline Graph Showing the Trend of Accidents on the Sunyani - Berekum Road

On the Sunyani – Berekum road, the line graph (figure 4.3) shows a downward trend for the fatal accidents over the study period, but the serious and minor accident graphs rather show an increasing trend.



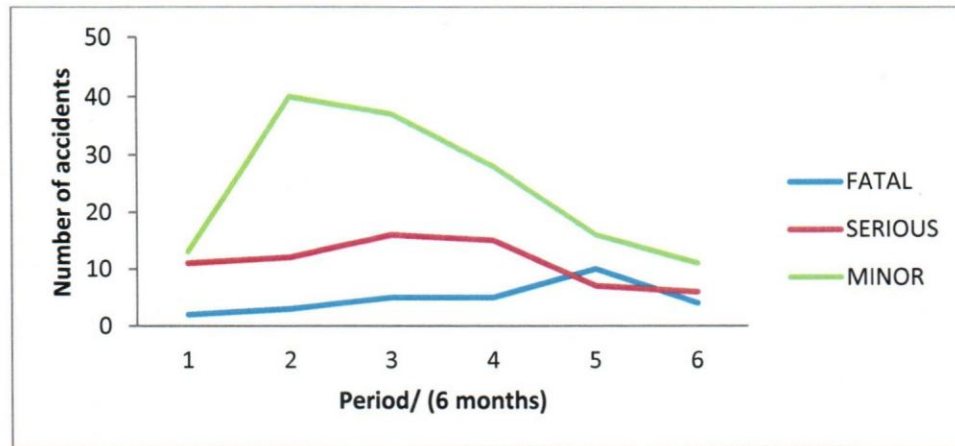


Fig 4.4: A line Graph showing the Trend of Accidents on Techiman - Tanoso Road

In the case of the Techiman – Tanoso road network, the line graphs (figure 4.4) show a general slow but steady downward trend for minor accidents after the first half of 2008 but the rest of the types of accidents do not show clear cut downward trend.

4.2 Univariate Analysis

The univariate analysis gives the descriptive measures or characteristics of the statistic or parameters of the population under consideration. The descriptive measures give an idea of the nature of data one is working with and most importantly an idea of the distribution it may follow before further analysis are done to accept the picture it portrays.

Tables 4.4, 4.5 and 4.6, give some of the descriptive measures of fatal accidents that occurred on the three road networks considered in the study in the periods under consideration (2008 – 2010).

One thing that runs through the descriptive analysis given in the tables that confirm the fact that the data we are dealing with is not normally distributed is



the measure of skewness which is positive and greater than zero. Hence the unknown distribution that fits this data is positively skewed.

Looking at the coefficient of variation for each of the road networks it appears that the variation in terms of the occurrence of fatal accidents is more pronounced on the Sunyani – Berekum followed by the Techiman – Wenchi and lastly the Techiman – Tanoso road.

**Table 4.3: Descriptive Statistics of Fatal Accidents on
Techiman - Wenchi Road**

Statistic / Parameter	value	Statistic / Parameter	value
N	36	Coefficient of variation	154.9193
Mean	0.5000	Uncorrected sum of squares	30
Variance	0.6000	Corrected sum of squares	21
Standard deviation	0.7746	Standard error mean	0.1291
Skewness	1.5622	Kurtosis	2.0244

**Table 4.4: Descriptive Statistics of Fatal Accidents on Techiman -
Tanoso Road**

Statistic / Parameter	value	Statistic / Parameter	value
N	36	Coefficient of variation	128.7327
Mean	0.8056	Uncorrected sum of squares	61



Table 4.4 Continuation

Variance	1.0754	Corrected sum of squares	37.6389
Standard deviation	1.0370	Standard error mean	0.1728
Skewness	0.9012	Kurtosis	-0.5698

**Table 4.5: Descriptive Statistics of Fatal Accidents on
Sunyani – Brekum Road**

Statistic/ Parameter	value	Statistic / Parameter	value
N	42	Coefficient of variation	159.366
Mean	0.61905	Uncorrected sum of Squares	56
Variance	0.97329	Corrected sum of Squares	39.905
Standard deviation	0.98655	Standard error mean	0.1522
Skewness	1.33162	Kurtosis	0.41720

4.3 Criteria for Choosing Model Fit Parameter

In an attempt to select a model that fits the data on each of the road network considered in the study, a lot of parameters were taken into consideration. The nature of the data itself is discrete and a count data which is very rare. We are talking about a model to predict fatal road accidents – although accidents may occur but fatal accidents do not occur so frequently. In this case continuous distributions were not considered as potential model fit distributions.

Also among the discrete distributions those that fit well with count data were selected and tested for use. The following distributions were considered in an attempt to fit the data to the most suitable model: Poisson, Negative binomial,



Zero – inflated Poisson, Zero – inflated Negative binomial, Poisson Log – normal, Geometric and Gamma.

Using model fit criteria such as AIC, AICC, BIC and others such as scaled deviance and the scaled Pearson chi – square measures, the Poisson and the negative binomial distributions were the best among the lot considered. The Poisson and the negative binomial distributions were also compared using same criteria and the Poisson distribution fits the data on each of the road networks considered better than the Negative binomial distribution.

Tables 4.6, 4.7, and 4.8 give details of the criteria used.

Table 4.6: Criteria for Assessing Goodness of Fit: Techiman – Wenchi

Road Network

Criteria	Poisson			Negative Binomial		
	df	Value	Value / df	df	Value	Value / df
Deviance	30	28.6339	0.9545	30	30.8238	1.0275
Scaled deviance	30	28.6339	0.9545	30	30.8238	1.0275
Pearson χ^2	30	29.2000	0.9733	30	31.2964	1.0432
Scaled Pearson χ^2	30	29.2000	0.9733	30	31.2964	1.0432
Log likelihood		-24.8623			-24.8201	
Full log likelihood		-28.7335			-28.6913	
AIC		69.4669			71.3824	
AICC		72.3635			75.3825	
BIC		78.9680			82.4672	



Table 4.7: Criteria for Assessing Goodness of Fit: Techiman – Tanoso

Road

	Poisson			Negative Binomial		
Criteria	df	Value	Value / df	df	Value	Value / df
Deviance	30	31.9510	1.0650	30	36.6739	1.2225
Scaled deviance	30	31.9510	1.0650	30	36.6739	1.2225
Pearson χ^2	30	28.0238	0.9341	30	32.6177	1.0873
Scaled Pearson χ^2	30	28.0238	0.9341	30	32.6177	1.0873
Log likelihood		-25.3839			-25.2500	
Full log likelihood		-35.6112			-35.4773	
AIC		83.2225			84.9546	
AICC		86.1190			88.9546	
BIC		92.7236			96.0392	

Table 4.8: Criteria for Assessing Goodness of Fit: Sunyani – Berekum

Road Network

	Poisson			Negative Binomial		
Criteria	df	Value	Value / df	df	Value	Value / df
Deviance	35	44.8525	1.2815	35	35.0596	1.0017
Scaled deviance	35	44.8525	1.2815	35	35.0596	1.0017
Pearson χ^2	35	39.3000	1.1229	35	29.2486	0.8357
Scaled Pearson χ^2	35	39.3000	1.1229	35	29.2486	0.8357
Log likelihood		-30.2210			-29.7272	
Full log likelihood		-39.7552			-39.2614	



Table 4.8 Continuation

AIC		93.5103			94.5227	
AICC		96.8044			98.8864	
BIC		105.6740			108.4241	

From tables 4.6, 4.7 and 4.8 the measured values for AIC, AICC and BIC under Poisson are smaller than its corresponding values under the Negative binomial hence the choice for the Poisson distribution to fit the data.

Information on the three tables above show that data on Techiman – Wenchi road fits best to the Poisson model, followed by the data on the Techiman – Tanoso road and finally the Sunyani – Berekum road using the AIC values.

Further analysis was conducted in section 4.4 to accept the fact that, the data was not overdispersed which would have required that we go for the Negative Binomial distribution instead of the Poisson to fit the data.

4.4 Testing for Overdispersion in the Poisson Regression

Although the ratios of the Pearson chi – square and deviance to the degrees of freedoms obtained under the criteria for assessing goodness of fit gave values very close to one (1) in tables 4.7, 4.8 and 4.9 as well as values obtained for AIC, AICC and BIC, implying Poisson model fit is better, we further test for overdispersion in the Poisson regression to confirm it or otherwise.

As already explained under section 3.7.0, this is a test of equality of the mean and variance imposed by the Poisson distribution against the alternative that the variance exceeds the mean



4.4.1 Testing for Overdispersion

The table below (Table 4.9) shows the result of the test of over dispersion done on each of the road networks.

Table 4.9: Overdispersion Test Results

Road Network	LR Statistic Value	Critical Value
Techiman - Wench	0.0844	2.7055
Sunyani - Berekum	0.9876	2.7055
Techiman - Tanoso	0.2678	2.7055

From table 4.9, since each LR statistic value for a particular road network is less than the critical value taken from the tables, it implies we failed to reject the null hypothesis for lack of evidence based on the data we are working with and conclude that the mean and the variance are equal.

4.5 Estimation and Significance Test of Parameters from the Statistical Output

As indicated earlier, SAS version 9.2.2 was used to run the data. Data for each road network was treated separately and separate output observed. The subsections below give detail information on output obtained for each of the road network considered in the study. For each of the road networks, information on parameter estimates and significance of variables is accompanied by a table giving log likelihood ratio test (Type 1 analysis) on variables considered in the study.



4.5.1 Analysis of Maximum Likelihood Parameter Estimates

(Techiman – Wenchi Road)

Table 4.10 gives summary of the output generated by the SAS 9.2.2 version upon a command issued for the analysis of the data on Techiman – Wenchi road network. The table gives among other things the parameter estimates and the significance of each parameter in the model.

(i) Maximum Likelihood Parameter Estimates and Tests of Significance

Table 4.10: Analysis of Data from the Techiman – Wenchi Road Network

Parameter	df	Estimate	S.E	Wald 95% CL		Wald χ^2	Pr> χ^2
Intercept	1	1241.298	254.8307	741.8307	1740.747	23.73	<.0000
Traffic volume	1	-0.0652	0.0114	-0.0875	-0.0428	32.66	<.0001
Section length	1	7.6142	1.6228	4.4337	10.7948	22.02	<.0001
Road width	1	-111.987	27.3252	-165.543	-58.4302	16.80	<.0001
Number of junctions	1	-6.1871	0.0000	-6.1871	-6.1871	.	.
Speed ramps	1	1.4271	2.1525	-2.7917	5.6460	0.44	0.5073
Zebra crossings	0	0.0000	0.0000	0.0000	0.0000	.	.
Speed limits	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale	0	1.0000	0.0000	1.0000	1.0000	.	.

Table 4.10 shows that out of the seven variables considered in the study, only three of them excluding the intercept were significant, namely, traffic volume, section length and road width. The type 1 analysis (Table 4.11) gave traffic volume and section length as the only variables that contribute significantly to the prediction power of the model specified.



(ii) Log likelihood Ratio Statistics for Type 1 Analysis

Table 4.11: Type 1 Analysis Table

Source	Deviance	Df	χ^2	Pr > χ^2
Intercept	39.8627	1		
Traffic volume	34.7066	1	5.16	0.0232
Section Length	30.3514	1	4.36	0.0369
Road width	30.0922	1	0.26	0.6106
No. of junctions	29.0771	1	1.02	0.3137
Speed ramps	28.6339	1	0.44	0.5056
Zebra crossing	28.6339	0	0.00	.
Speed limit	28.6339	0	0.00	.

Table 4.11 shows that only two of the variables, traffic volume and section length, contribute significantly to the reduction of deviance in the model.

4.5.2 Analysis of Maximum Likelihood Parameter Estimates

(Techiman – Tanoso Road)

Table 4.12 gives summary of the output generated by SAS 9.2.2 version upon a command issued for the analysis of the data on Techiman – Tanoso road network. The table gives among other things the parameter estimates and the significance of each parameter in the model. For each of the cells in the table, the value in the first row corresponds to the estimate of a parameter from the Poisson fitted model and the value in the second row corresponds to the estimate from the Negative binomial fitted model.



(i) **Maximum Likelihood Parameter Estimates and Tests of Significance**

Table 4.12: Analysis of Data From the Techiman – Tanoso Road Network

Parameter	df	Estimate	S.E	Wald 95% CL		Wald χ^2	Pr > χ^2
Intercept	1	-886.085	86.9071	-1056.42	-715.751	103.95	<.0001
Traffic volume	1	0.0246	0.0016	0.0215	0.0278	241.46	<.0001
Section length	1	4.5255	1.1719	2.2285	6.8225	14.91	0.0001
Road width	1	84.9135	8.9630	67.3464	102.4807	89.75	<.0001
Number of junctions	1	-2.7780	0.4725	-3.7040	-1.8520	34.57	<.0001
Speed ramps	0	-6.3828	0.0000	-6.3828	-6.3828	.	.
Zebra crossings	0	0.0000	0.0000	0.0000	0.0000	.	.
Speed limits	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale / Dispersion	0	1.0000	0.0000	1.0000	1.0000	.	.

Table 4.12 shows that out of the seven variables considered in the study, only four of them excluding the intercept were significant, namely, traffic volume, section length road width and number of junctions. The type 1 analysis (Table 4.13) gave traffic volume and section length as the only variables that significantly contribute to the prediction power of the model specified.

(ii) **Log likelihood Ratio statistics for Type 1 Analysis**

Table 4.13: Type 1 Analysis Table

Source	Deviance	Df	χ^2	Pr > χ^2
Intercept	51.7241	1		
Traffic volume	41.8294	1	9.89	0.0017



Table 4.13 continuation

Section Length	34.1086	1	7.72	0.0055
Road width	33.4968	1	0.61	0.4341
No. of junctions	33.3002	1	0.20	0.6575
Speed ramps	31.9510	1	1.35	0.2454
Zebra crossing	31.9510	0	0.00	.
Speed limit	31.9510	0	0.00	.

Table 4.13 shows that only traffic volume and road section contribute significantly to the reduction of deviance

4.5.3 Analysis of Maximum Likelihood Parameter Estimates (Sunyani - Berekum)

Table 4.14 gives summary of the output generated by SAS 9.2.2 version upon a command issued for the analysis of the data on Sunyani - Berekum road network. The table gives among other things the parameter estimates and the significance of each parameter in the model. For each of the cells in the table, the value in the first row corresponds to the estimate of a parameter from the Poisson fitted model and the value in the second row corresponds to the estimate from the Negative binomial fitted model.



(i) Maximum Likelihood Parameter Estimates and Tests of Significance

Table 4.14: Analysis of Data From the Sunyani - Berekum Road

Network

Parameter	df	Estimate	S.E	Wald 95% CL		Wald χ^2	Pr> χ^2
Intercept	1	642.7721	177.2588	295.3512	990.1929	13.15	<.0003
Traffic volume	1	-0.0240	0.0030	-0.0299	-0.0181	63.55	<.0001
Section length	1	8.6688	1.3221	6.0775	11.2601	42.99	<.0001
Road width	1	-63.3735	19.9466	-102.468	-24.2789	10.09	<.0015
Number of junctions	1	9.3100	1.7713	5.8384	12.7816	27.63	<.0001
Speed ramps	1	21.2201	2.0404	17.2210	25.2193	108.16	<.0001
Zebra crossings	0	-10.2663	0.0000	-10.2663	-10.2663	.	.
Speed limits	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale / Dispersion	0	1.0000	0.0000	1.0000	1.0000	.	.

Table 4.14 shows that out of the seven variables considered in the study, only five of them excluding the intercept were significant, namely, traffic volume, section length road width, number of junctions and Speed ramps. The type 1 analysis (Table 4.15) gave traffic volume and section length as the only variables that significantly contribute to the prediction power of the model specified.



4.6.1 Formulation of Regression Models

We now fit Poisson regression models to the road sections using model information provided on table 4.16.

From definition, a typical Poisson regression model expresses log outcome as a linear function of a set of predictors.

That is,

$$\log(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n \quad (4.1)$$

or

$$y = \exp(\sum x_i \beta_i) = e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n} \quad (4.2)$$

Where x_i 's define the explanatory variables and β_i 's are the coefficients of the explanatory variables and y gives the accident frequency per period.

Hence we have the various regression models given below:

- (i) Regression model for Techiman – Wenchi road section

$$y = e^{(13.7325 - 0.003 \text{Traffic volume} + 0.7178 \text{Section length})} \quad (4.3)$$

- (ii) Regression model for Techiman – Tanoso road section

$$y = e^{(3.5679 - 0.0009 \text{Traffic volume} + 0.85978 \text{Section length})} \quad (4.4)$$

- (iii) Regression model for Sunyani - Berekum road section

$$y = e^{(1.0362 - 0.0004 \text{Traffic volume} + 0.6001 \text{Section length})} \quad (4.5)$$



4.6.2 Model Validation

As already discussed in section 3.90, the residual analysis was selected to validate the models developed if they are good enough to make reliable predictions. Theory on the use of graphical residual analysis reveals that in a scatter plot of residuals against predicted values, if the residuals appear to behave randomly, it suggests that the model fits the data well. On the other hand, if non – random structure is evident in the residuals, it is a clear sign that the model fits the data poorly.

Scatter plots were done for each of the three models derived and the figures below (Figure 4.3, 4.4 and 4.5) give a picture of the nature of the scatter plot.

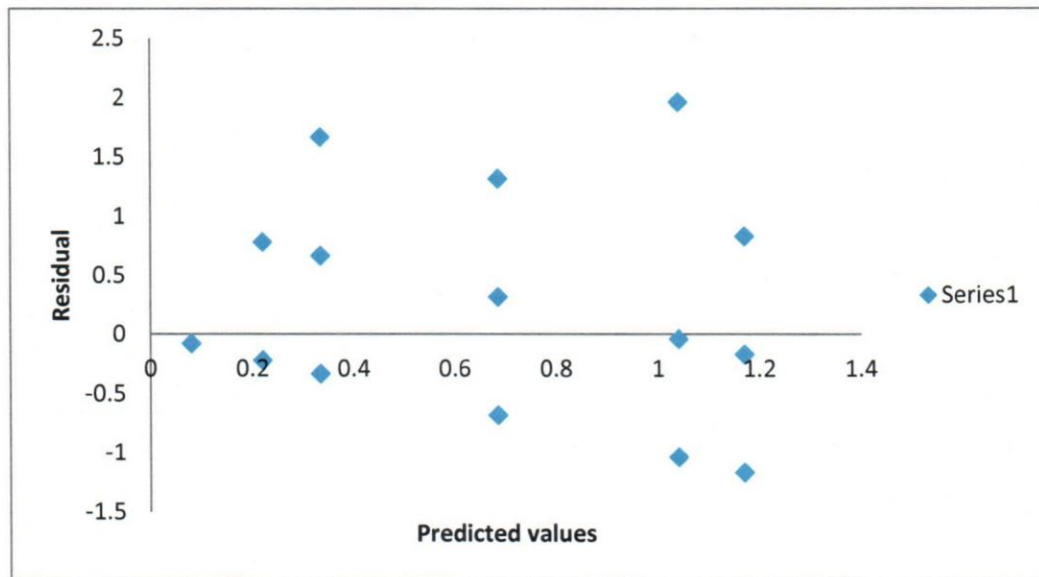
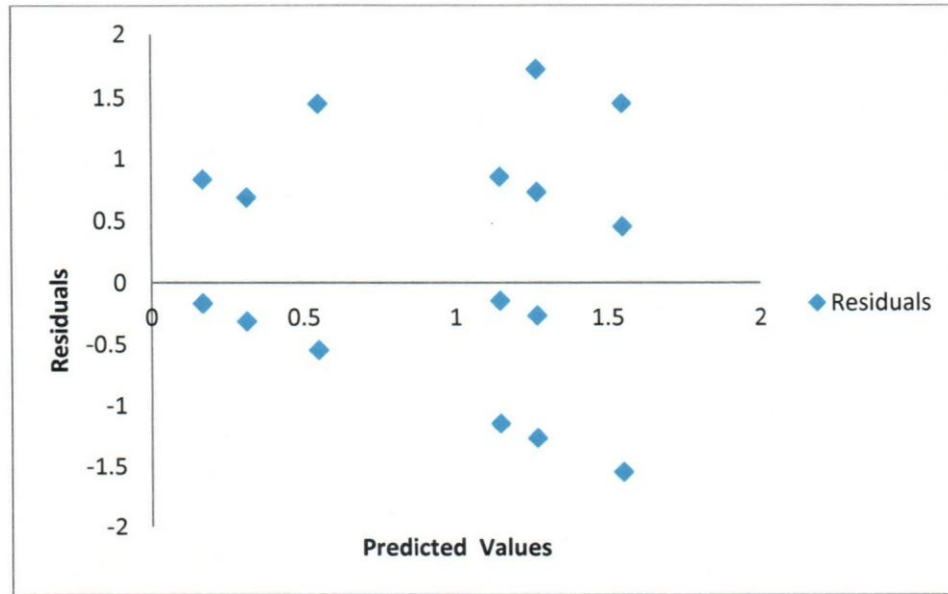


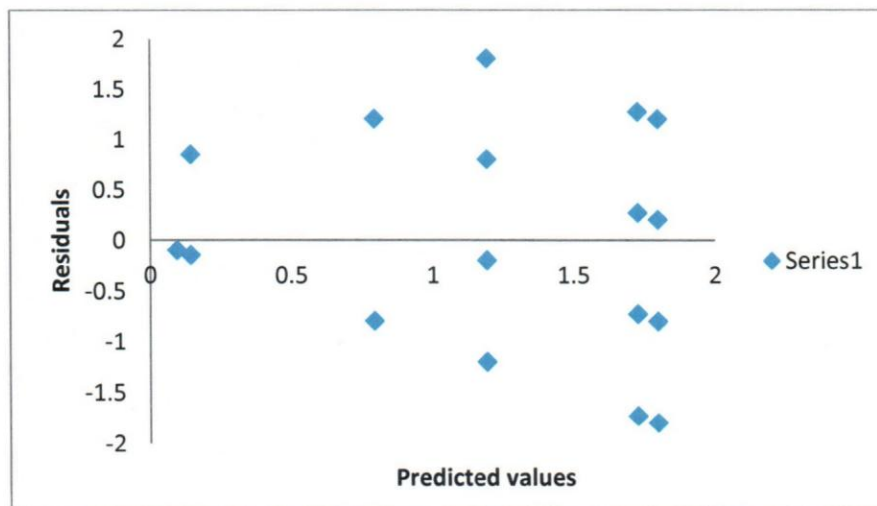
Fig 4.5: A Scatter plot of Residuals Against Predicted Values for

Techiman – Wench Road to Check Model Validation





**Fig 4.6: A Scatter Plot of Residuals Against Predicted Values for
Sunyani - Brekum Road to Check Model Validation**



**Fig 4.7: A Scatter Plot of Residuals Against Predicted Values for
Techiman - Tanoso Road to Check Model Validation**

From figure 4.3, it can be seen that the points in the scatter plot are randomly scattered between 2.0 and -1.5, implying the model fits the data well and can give approximately accurate predictions.



Similarly, figure 4.4 reveals a scatter plot with residual points randomly spread symmetrically around the origin between +2.0 and -2.0, which also gives an idea of a better fit of the model to the data.

A similar pattern was observed in figure 4.5, which gives scatter plot points falling randomly between +2.0 and -2.0, once again indicating a good fit to the data used to derive the model.

4.7 Findings and Discussions

The study seeks to identify possible factors that contribute to occurrence of fatal accidents on our roads and subsequently develop models to predict future occurrences of fatal accidents using the factors identified. Three specific road networks were considered in the Brong Ahafo Region, namely the Techiman – Wenchi road, Sunyani – Berekum road and the Techiman – Tanoso road.

One of the findings in the preliminary analysis shows that it appears more commercial vehicles were involved in accident (Table 4.3) as compared to private vehicles on each of the road networks. Although further tests were not carried out to confirm this finding but this definitely sends a strong wakeup call to the driver training schools and all stake holders in the professional driving business to set their priorities right.

The study also shows that occurrence of accidents on the three road network considered is not all that alarming although some of the accident types were on the slow pace ascendancy. This calls for no complacency but put our weight behind the National Road Safety Commission to work harder to reduce drastically the occurrence of accidents and for that matter fatal road accidents.



Also two out of the seven factors identified emerged as significant predictors in all the models developed. These were traffic volume and section length.

The initial analysis by the SAS output showed that other factors such as the Number of junctions, number of speed rumps and road width were significant for some of the road networks, but the Type 1 analysis found only two factors significant since the other factors do not contribute significantly to the reduction of deviance when included in the models.

Tables 4.10, 4.12 and 4.14, show that traffic volume ($Pr < 0.0001$) and section length ($pr < 0.0001$) were significant and this was confirmed by the Type 1 analysis on tables 4.11, 4.13 and 4.15.

Using only the significant factors, the data was run again, new estimates of parameter coefficients were obtained (Table 4.16) which was used to develop the fatal accident predictive model for each of the three road networks (Equations 4.3, 4.4 and 4.5).

In general, the deductions concerning the models developed and its relation to the statistically significant factors on all the three road sections showed the following:

- (i) Fatal accident frequency is directly related to the section length (measured in kilometres). Implying increase in section length of a particular road network directly increases the accident frequency.
- (ii) Fatal accident frequency is inversely proportional to traffic volume on all the road sections considered in the study.

These deductions conform to the findings of Mustakin, F et al (2008) in their research work conducted on Black spot study and Accident Prediction Model using multiple regressions.

The second deduction confirms the findings of most researchers in this area of study that road accidents are more pronounced in the rural highways (low traffic volume) compared to road networks located in the urban areas (high traffic volume).



CHAPTER FIVE

DISCUSSIONS, SUMMARY, CONCLUSIONS AND RECOMMENDATIONS FOR FURTHER STUDIES

This last chapter focuses on major conclusions as a result of the findings and other challenges encountered in the study, we begin with the summary. Finally, recommendations will be given to those who would want to do further studies in this area.

5.1 Discussions

One of the deductions made from the models established using data from each of the road sections shows that higher traffic volume on a road section reduces frequency of road accidents. This finding at first glance may sound controversial but let's look at it this way, the outcome of the occurrence of accident on rural road section and urban road section. We can say that traffic volume is relatively higher in the urban areas as compared to rural areas and the same can be said of the number and easy accessibility to quality health facility as well as qualified health personnel. Although we may have low traffic volume on rural roads but most of the accidents that occur on such roads result in fatality because seriously injured victims may not have easy access to a health care centre or a qualified first Aider which may lead to loss of life and eventually the accident classified as fatal. In the urban areas, the story is different; we may have higher traffic volume and relatively higher incidence of accidents but because victims of accidents have easy access to a health facility and qualified health personnel the incidence of fatality is lower.



5.2 Summary

The study was focused on identification of roadway factors that contributes to the occurrence of fatal accidents and the development of models to predict such occurrences.

It was based on accident data from the MTTU regional office of the Ghana Police Service in the Brong Ahafo. The data included information on general daily accident frequency and information on road network where the incidents took place from 2008 to 2010.

The data was finally fitted to the Poisson regression model after comparing it to other competing models that can fit well to count and rare data like data on fatal accident occurrence.

Literature shows that many factors have bearing on the occurrence of road accidents which includes the driver's characteristics, the characteristics of the vehicle and the road way environment. This study focuses on the road way environment and how it affects fatal accident frequency on the selected roads as mentioned earlier.

The study finally showed that traffic volume and road section length contributes significantly to the occurrence of fatal road accidents.

Other factors such as the number of road way junctions, number of speed ramps on the roadway and the width of road sections were also to some extent identified to be significant, but further analysis using the log likelihood ratio test showed that they do not contribute significantly to the reduction of deviance in the model developed.

Other information deduced from the models developed showed that increase in traffic volume on a particular road network reduces fatal accident frequency



whereas increase in section length has a direct relationship to the number of fatal accidents occurring on a particular road network.

5.3 Conclusions

The following conclusions can be drawn from the study of which accident data from the Ghana Police MTTU stations in the Brong Ahafo region were used based on three selected road networks, namely, Techiman – Wenchi road, Sunyani Berekum road and Techiman – Tanoso road.

The results showed that, more commercial vehicles were involved in accident over the study period compared to private vehicles although further tests were not conducted to confirm this outcome. This sends a caveat to all stake holders in the commercial driving business to ensure that peoples' lives entrusted into the hands of the professional drivers are handled with care and a sense of responsibility.

Further analysis on the data collected on the three road networks showed that traffic volume and section length contributes significantly to the occurrence of road accidents.

Other information deduced from the models developed showed that increase in traffic volume on a particular road network reduces fatal accident frequency whereas increase in section length has a direct relationship to the number of fatal accidents occurring on a particular road network.

The data fitted well to the Poisson regression model after comparing its results with other possible count data fitting models such as the negative binomial,



gamma, zero – inflated Poisson and the Zero – inflated negative binomial distributions.

5.4 Limitations of the Study

In the course of the study, it was realized that the culture of keeping records, especially records on very sensitive events like road accidents needs much to be desired. While some MTTU offices in Ghana Police Service are doing well and keeping abreast with the use of ICT in the 21st century, other offices are performing very poorly. Much emphasis should be placed on training most of the personnel who work on data of this nature and their offices equipped with current equipments or apparatus to enable them work effectively and efficiently.

The lengthy bureaucratic procedures contributed to delay in data collection for this study. The attitudes and some of the behaviours put up by some heads of departments when one go for information from them is not encouraging and motivating enough to whip your interest in the study you intend embarking on. To some people they just feel lazy getting the information for you and for others the information is not there, although they are supposed to keep records on that, whereas to some others they want you to promise giving them something before they give you what you need. All these negative practices should be put behind us if we are serious about moving the country forward and reducing accidents by 50% in the year 2015.



5.4 Recommendations

As mentioned earlier, this study focused on factors as a result of the road environment and features on the road network which contributes to the occurrence of fatal road accidents.

5.4.1 Recommendations to the Transportation Industry

Road Transportation Industry – by this we are referring to road engineers, road users, National Road Safety Commission, Driver Vehicle and Licensing Authority and the Driving Schools. From the descriptive analysis, it appears we have more accidents involving commercial drivers as compared to private drivers on the three road networks considered although no further statistical tests were conducted to confirm the finding. This finding sends a caveat to the Driver Vehicle and Licensing Authority, National Road Safety Commission, the Security Agencies and the Driving Schools to reconsider, enforce and maintain quality training of the professional drivers to help reduce the carnage on our roads.

The fact that accessibility to fast and good health facilities can help reduce fatality on our roads, the National Road Safety Commission and Policy Makers in the Transportation Industry must intensify training of more First Aiders in the Rural areas as well as the provision of adequate and well equipped health centres that can promptly attend to accident victims.



5.4.2 Recommendations for Further Studies

The following recommendations may be considered:

- (i) Any further study on this may include two or more of the major conditions that affect occurrence of accidents. Such as the vehicle conditions and the road environment or in addition to this two, one may include the driver characteristics as well. This may enhance the models derived and give better predictions.
- (ii) The study considered the case of only fatal accidents, other studies may also consider modelling for only serious or minor accidents or combinations of any two of the types of accidents.
- (iii) The study was conducted in the Brong Ahafo region and for only three of the road sections in the region, similar studies may be conducted for the other road networks in the region or on road sections in other regions in the country.
- (iv) Road sections of length longer than those considered in this study may also be considered to check if it will be significant in predicting the occurrence of road accidents.



REFERENCES

- Adams, J., (1987). *Smeed's law: some further thoughts*. Traffic Engineering and Control 10 (7), pp. 70-73. Vol. 33, pp. 663-672.
- Afukaar, F.K., Antwi P., and Ofosu-Amaah, S. (2003). Pattern of road traffic injuries in Ghana: Implications for control. *Injury control and safety promotion*, 10(12): 69-76.
- Agresti, A. (2002). *Categorical Data Analysis*. New York, Wiley.
- Aguero-Valverde, J., Jovanis, P.P. (2008). *Analysis of road crash frequency with spatial models*. Transportation Research Record 2061, 55-63.
- Akaike, H. (1973). *Information theory and an extension of the maximum likelihood principle*. In N. Petro & Csadki (Eds). Proceedings of the second International Symposium on Information Theory (pp. 716 – 723)
- Akaike, H. (1974). *A new look at statistical model identification*. IEEE Transactions on Automatic Control, AU – 19, pp. 716 - 722
- Akaike, H. (1976). *Canonical correlation analysis of time series and the use of an information criterion*. In R.K. Mehra & D.G. Lainiotis (Eds.), System identification: Advances and case studies (pp. 52 – 107)
- Amemiya, T. (1980). *Selection of regressors*. International Economic Review, (pp. 21, 331 – 354)
- Amemiya, T. (1981). *Qualitative response models: A survey*. Journal of Economic Literature, XIX, 1483 – 1536
- Amemiya, T. (1985). *Advanced econometrics*. Cambridge, M.A.: Harvard University press
- Armour M, Cinquegrana C. (1990). *Victorian study of single vehicle rural*





accidents. Proceedings of the 15th Australian Road Research Board Conference, Part 7, Australian Road Research Board, Melbourne, pp. 79-91.

Austin, C. (1964). *The migration of the Akans*. M.A.: Harvard University press

Bester, C.J. (2001). *Explaining national road fatalities*. Accident analysis and prevention. Vol. 33, pp. 663-672.

Bonneson J. A., McCoy P. T. (2001). *Effect of Median Treatment on Urban Arterial Safety and Accident prediction Model*, TRR 1581.

Broughton, J. (1988). *Predictive models of road accident fatalities*. Traffic Engineering and Control, May 1988, ISSN: 0041-0683, pp. 296-300.

Buah, F.K. (1998). *A history of Ghana*. Oxford Micmillan publishers. ISBN: 03336593141

Busia, K.A. The position of chief in the modern political system of the Ashanti London: Frank Cases & Co, 1951.

Cameron, A. C., Trivedi, P.K., (1998). *Regression analysis of count data*, Cambridge University Press

Carlin, B.P., Louis, T.A. (1996). *Bayes and empirical methods for data analysis*. New York: Chapman & Hall.

Cox D. R, (1983). "Some Remarks on Overdispersion". *Biometrika*, 70, 269-274.

Dewer, R. E. (2002). Individual differences. Chapter 5 in Dewer, R. E. and Olson, P. L., *Human Factors in Traffic Safety*, Lawyers and Judges Publishing Co., Tucson, AZ, ISBN 0-913875-47-3.

Elvik, R., Vaa, T., (2004). *The Handbook of Road Safety Measures*. Elsevier Amsterdam, ISBN: 0-08-044091-6, pp.66 and pp. 676-803.



- Elvik R. and Vaa, T. (1990). *Human factors, road accident data and information technology*. Oslo: Institute of Transport Economics. 155p
- Evans, L., (1991). *Traffic safety and the driver*, New York: Van Nostrand Reinhold, P25-43, pp. 60-95.
- Fieldwick, R. and Brown R.J. (1987). *The effect of speed limits on road casualties*. Traffic Engineering and Control, Vol. 28, pp 635-640
- Forjuoh, S.N. (2003). *Traffic-related Injury Prevention Interventions for Low-Income Countries*: Injury Control and Safety Promotion, 10.1-2: 109-18.
- Greene, W. (1997). *Economic Analysis* (3rd ed). New York: Prentice Hall
- Hakkert, A.S., and Braimaister, L., (2002). *The uses of exposure and risk in road safety studies*. SWOV Institute for Road Safety Research, the Netherlands.
- Jacobs, G.D., Hutchinson, P., (1973). *A study of accident rates in developing countries*. TRRL report LR 546. Transport and Road Research Laboratory, Crowthorne, Berkshire.
- Jovanis, P.P. and Chang, H-L, (1986). *Modelling the relationship of accidents to miles travelled*: Transportation Research Record 1068, Transportation Research Board, Washington D.C., pp42-51.
- Koehler, A.B., Murphree, E.S. (1988). *A comparison of the Akaike and Schwartz criterion for selecting model order*: Applied Statistics, pp. 187 - 195
- Kulmala, R., (1995). *Safety at rural three - and four-arm junctions, development and application of accident prediction models*: Technical Research Centre of Finland, VTT 233, 104p.
- Lord, D., Miranda-Moreno, L.F., (2008). *Effects of low sample mean values*

and small sample size on the estimation of the fixed dispersion parameter of Poisson-gamma models for modelling motor vehicle crashes: A Bayesian perspective. *Safety Science* 46(5), 751-770.

McCullagh, P. and Nelder, J.A., (1989). *Generalised Linear Models: Monographs on Statistics and Applied Probability*, Vol.37, Chapman and Hall.

McCullagh, P. and Nelder, J.A. (1989), *Generalized Linear Models*, Second Edition, Longon: Chapman and Hall.

McGuigan, D.R.D. (1999). *The use of relationships between road accidents and traffic flow in "black spot" identification*: *Traffic Engineering and Control*, Vol.22, Nos.8/9. Pp 448-453.

Miaou, S. and Lum, H., (1993). *Modelling vehicle accident and highway geometric design relationships*: *Accident analysis and prevention*, Vol.25, No.6, pp 689-709.

Michael S., Gabriel T., Seth, T.K., (2004), "Individual differences and the high risks of commercial driver"

Mohammad A. Hadi and A.S. Jaradat (1998), "Analysis of Commercial Minibus Accidents", *Accident Analysis and Prevention* vol.30 No.5

Neftci, S.N. (1982). *Specification of Economic time series models using Akaike's Criterion*. *Journal of the American Statistical Association* (pp. 77, 537 – 540)

NHTSA (2000). *Traffic safety facts: Older population*. DOT HS 809 328
National Highway Traffic Safety Administration, Washington, DC

NHTSA. (2007). *Graduated driver licensing system*. [Cited Dec. 31, 2009]

Retrieved from:



<http://www.nhtsa.dot.gov/people/injury/TSFLaws/PDFs/810727W.pdf>

NHTSA. (2008). *Traffic Safety Facts: Rural/Urban Comparison*.

[Cited: January 11, 2010]. Available from:<http://www>

nrd.nhtsa.dot.gov/Pubs/810812.PDF

Nyberg, A., and Gregersen, N.P. (2007). *Practicing for and performance on drivers license tests in relation to gender differences in crash involvement among novice drivers*. Journal of Safety Research, 38 (1):71-80.

Oh, J., Washington, S.P., Nam, D. (2006). *Accident prediction model for railway-highway interfaces*. Accident Analysis and Prevention 38(2), 346-56.

Oppe S., (1989). *Macroscopic models for traffic and traffic safety*. Accident Analysis and Prevention, Vol.21, pp. 225-232.

Page,Y., (2001). *A statistical model to compare road mortality in OECD countries*. Accident Analysis and Prevention, 33, pp. 371-385.

Radin, U. R. S., (2005); "Modelling of Traffic Accidents Along Major Trunk Roads in Malaysia", Journal IEM.vol. 62 No. 3 2005.

Reason, J. (1990). *Human Error*. Cambridge University Press, Cambridge, UK.

Rimmo, P. A., (2002). "Aberrant Driving Behaviour: Homogeneity of a Four-Factor Structure in Samples Differing in Age and Gender." Ergonomics, Vol. 45, No. 8, 569-582

Saccomanno F. F., Buyco C.,(1988). Generalized loglinear models of truck accident rates, "Transportation research record", Vol. 1172, 23-31.

Sawa, T. (1978). Information criterion for discriminating among alternative regression models. Econometrica 46, 1273 – 1291





Scwartz, G. (1978). Estimating the dimension of model. *Annals of statistics*, 6, 461-464

Shaw-Pin Miaou, Patricia S. Hu, Tommy Wright, Ajay K. Rath, and Stacy C. Davis, *Relationship between truck accidents and highway geometric design: A Poisson Regression Approach*, *Transportation Research Record*, Vol. 1376, 10-18.

Silvak, M., (1983). *Society's Aggression Level as a Predictor of Traffic Fatality Rate*. *Journal of Safety Research* 14, 93-99.

Smeed, R.J., (1949). *Some statistical aspects of road safety research*. *Journal of Royal Statistical Society Series A* 112, pp. 1-34.

Summersgill, I., Kennedy, J.V. and Baynes, D., (1996). *Accidents at three-arm priority junctions on urban single-carriageway roads*, TRL Report 184, Transport Research Laboratory, Crowthorne, p.132.

Tanner, J.C., (1953). *Accidents at rural 3-way junctions*, *Journal of Institution of Highway Engineers*, Vol.2, No.11, pp56-67.

Thouez, J.P., Van den Bossche, F., Wets, G. (1991). Geographical Variations of Motor-Vehicle Injuries in Quebec, 1983-1988. *Social Science and Medicine* 33.4: 415-21.

Web 1: Retrieved from:

<http://knowledgebase.findlaw.com/kb/2009/Dec/59238.html> - 27k
(last visited: 13/06/2012)

Wedderburn, R. W. M. (1974). "Quasi-Likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method," *Biometrika*, 61, 439-447.

WHO, (2004). *World Report on Road Traffic Injury Prevention*, Chapter 3

"Risk Factors", Geneva.

http://www.who.int/worldhealthday/2004/infomaterials/world_report/en/

(Retrieved on December 19, 2010).

Zegeer, Charles V., J. Hummer, D. Reinfurt, L. Herf, and W.W. Hunter (1986),

"Safety Effects of Cross-Section Design for Two- Lane Roads", Report

No. FHWA-RD-87-008, Federal Highway Administration, Washington,

D. C.

Zhang, P. (1992). On the distributional properties of model selection. Journal

of the American Statistical association, 87, 732 - 737



Appendix A: Accident Data from Techiman – Wenchi Road Network

Period	Fatal	Serious	Minor	Com Vehicles	Private Vehicles	Traffic Volume	Section Length	Road Width	No. of T-Junctions	Speed ramps	Zebra Crossing	Speed Limits	Nature Road
1	0	2	3	4	1	5900	2	8	1	3	3	2	1
2	0	1	2	2	1	5900	4	7.9	1	2	2	1	1
3	1	0	3	2	2	6000	6	7.85	3	1	1	1	1
4	1	2	1	3	1	6200	7	7.8	3	1	1	1	1
5	0	2	4	3	3	5900	5	7.9	2	1	2	1	1
6	0	0	3	2	1	5800	3	7.9	1	2	2	2	1
7	0	3	2	3	3	5900	2	8	1	3	3	2	1
8	0	2	1	3	1	5900	4	7.9	1	2	2	1	1
9	0	1	2	3	0	6000	6	7.85	3	1	1	1	1
10	1	0	4	2	4	6200	7	7.8	3	1	1	1	1
11	0	3	2	4	2	5900	5	7.9	2	1	2	1	1
12	0	1	2	3	1	5800	3	7.9	1	2	2	2	1
13	0	1	3	4	2	5900	2	8	1	3	3	2	1
14	0	2	1	2	2	5900	4	7.9	1	2	2	1	1
15	3	1	1	2	4	6000	6	7.85	3	1	1	1	1
16	0	2	1	2	3	6200	7	7.8	3	1	1	1	1
17	2	1	3	5	2	5900	5	7.9	2	1	2	1	1
18	1	0	2	3	1	5800	3	7.9	1	2	2	2	1
19	0	3	0	3	2	5900	2	8	1	3	3	2	1
20	2	1	1	4	2	5900	4	7.9	1	2	2	1	1
21	0	1	4	3	3	6000	6	7.85	3	1	1	1	1
22	1	2	2	4	2	6200	7	7.8	3	1	1	1	1
23	0	1	1	2	0	5900	5	7.9	2	1	2	1	1
24	0	2	3	4	2	5800	3	7.9	1	2	2	2	1
25	0	1	1	0	2	5900	2	8	1	3	3	2	1
26	0	3	1	3	3	5900	4	7.9	1	2	2	1	1
27	0	2	2	3	1	6000	6	7.85	3	1	1	1	1
28	2	2	1	4	1	6200	7	7.8	3	1	1	1	1
29	1	1	1	2	2	5900	5	7.9	2	1	2	1	1
30	0	2	1	2	2	5800	3	7.9	1	2	2	2	1
31	0	1	2	2	4	5900	2	8	1	3	3	2	1
32	1	3	2	2	4	5900	4	7.9	1	2	2	1	1
33	1	2	3	4	2	6000	6	7.85	3	1	1	1	1
34	1	2	2	3	2	6200	7	7.8	3	1	1	1	1
35	0	1	1	2	2	5900	5	7.9	2	1	2	1	1
36	0	2	1	2	3	5800	3	7.9	1	2	2	2	1



Appendix B: Accident Data from Sunyani – Berekum Road Network

Fatal	Serious	Minor	Com Vehicles	Private Vehicles	Traffic Volume	Section Length	Road Width	T - Junction	Speed Ramps	Zebra Crossing	Speed Limit	Nature Road
0	0	1	0	1	10080	2	8	1	4	3	2	1
0	1	0	2	0	10000	3	7.9	3	3	2	1	1
0	1	3	3	2	10100	4	7.8	4	2	2	1	1
0	2	1	4	0	11000	6	7.7	5	1	1	1	1
0	1	3	3	1	12000	7	7.6	6	1	1	1	1
2	1	0	2	2	10500	5.5	7.85	4	2	2	1	1
1	0	2	2	2	10800	2.5	7.9	2	5	3	2	1
0	4	1	5	2	10080	2	8	1	4	3	2	1
0	2	2	3	2	10000	3	7.9	3	3	2	1	1
0	0	4	3	1	10100	4	7.8	4	2	2	1	1
0	4	2	3	3	11000	6	7.7	5	1	1	1	1
2	0	2	3	3	12000	7	7.6	6	1	1	1	1
0	2	1	2	2	10500	5.5	7.85	4	2	2	1	1
0	2	0	2	1	10800	2.5	7.9	2	5	3	2	1
0	1	3	3	2	10080	2	8	1	4	3	2	1
1	2	0	3	1	10000	3	7.9	3	3	2	1	1
2	0	2	2	3	10100	4	7.8	4	2	2	1	1
1	0	1	1	1	11000	6	7.7	5	1	1	1	1
3	1	0	4	3	12000	7	7.6	6	1	1	1	1
2	0	2	4	2	10500	5.5	7.85	4	2	2	1	1
0	1	0	1	1	10800	2.5	7.9	2	5	3	2	1
0	3	1	4	1	10080	2	8	1	4	3	2	1
0	0	4	3	3	10000	3	7.9	3	3	2	1	1
0	2	3	3	4	10100	4	7.8	4	2	2	1	1
3	2	0	4	1	11000	6	7.7	5	1	1	1	1
3	0	1	3	1	12000	7	7.6	6	1	1	1	1
0	1	1	2	2	10500	5.5	7.85	4	2	2	1	1
0	1	0	1	1	10800	2.5	7.9	2	5	3	2	1
0	1	0	2	0	10080	2	8	1	4	3	2	1
1	0	2	4	0	10000	3	7.9	3	3	2	1	1
2	0	1	1	3	10100	4	7.8	4	2	2	1	1
0	3	1	4	2	11000	6	7.7	5	1	1	1	1
0	1	2	3	2	12000	7	7.6	6	1	1	1	1
1	1	0	2	0	10500	5.5	7.85	4	2	2	1	1
0	1	1	0	3	10800	2.5	7.9	2	5	3	2	1
0	2	3	4	2	10080	2	8	1	4	3	2	1
0	1	1	2	1	10000	3	7.9	3	3	2	1	1
0	1	1	3	0	10100	4	7.8	4	2	2	1	1
2	0	3	3	3	11000	6	7.7	5	1	1	1	1
0	0	2	2	2	12000	7	7.6	6	1	1	1	1
0	1	1	2	0	10500	5.5	7.85	4	2	2	1	1
0	2	2	2	4	10800	2.5	7.9	2	5	3	2	1



Appendix C: Accident Data from Techiman - Tanoso Road Network

d	Fatal	Serious	Minor	Com Vehicles	Private Vehicles	Traffic Volume	Section Length	Road Width	T - Junctions	Speed Ramps	Zebra Crossing	Speed Limit	Nature of Road
	0	2	1	3	2	8500	2	8	2	5	6	2	1
	0	0	3	3	1	9000	3	7.9	1	3	4	1	1
	0	3	4	5	3	9000	5	7.8	3	2	3	2	1
	0	1	2	4	1	9500	6	7.6	5	1	2	2	1
	2	2	0	2	4	11000	8	7.2	7	2	1	2	1
	0	3	3	4	4	10000	7	7.4	5	1	2	2	1
	0	1	4	6	2	8500	2	8	2	5	6	2	1
	0	4	8	11	6	9000	3	7.9	1	3	4	1	1
	2	1	8	9	4	9000	5	7.8	3	2	3	2	1
	0	0	8	7	5	9500	6	7.6	5	1	2	2	1
	0	4	7	9	5	11000	8	7.2	7	2	1	2	1
	1	2	5	6	3	10000	7	7.4	5	1	2	2	1
	0	2	4	4	4	8500	2	8	2	5	6	2	1
	0	2	8	8	4	9000	3	7.9	1	3	4	1	1
	0	5	9	8	7	9000	5	7.8	3	2	3	2	1
	1	2	5	8	3	9500	6	7.6	5	1	2	2	1
	2	1	5	7	3	11000	8	7.2	7	2	1	2	1
	2	4	6	8	5	10000	7	7.4	5	1	2	2	1
	0	1	3	3	3	8500	2	8	2	5	6	2	1
	1	2	6	6	5	9000	3	7.9	1	3	4	1	1
	0	5	3	7	6	9000	5	7.8	3	2	3	2	1
	3	0	5	6	4	9500	6	7.6	5	1	2	2	1
	1	4	7	8	6	11000	8	7.2	7	2	1	2	1
	0	3	4	6	7	10000	7	7.4	5	1	2	2	1
	0	3	5	7	4	8500	2	8	2	5	6	2	1
	0	2	3	4	1	9000	3	7.9	1	3	4	1	1
	2	0	4	7	5	9000	5	7.8	3	2	3	2	1
	2	1	0	3	3	9500	6	7.6	5	1	2	2	1
	3	0	4	7	3	11000	8	7.2	7	2	1	2	1
	3	1	0	6	3	10000	7	7.4	5	1	2	2	1
	0	2	5	5	3	8500	2	8	2	5	6	2	1
	0	0	2	2	2	9000	3	7.9	1	3	4	1	1
	0	1	0	1	1	9000	5	7.8	3	2	3	2	1
	2	0	0	4	1	9500	6	7.6	5	1	2	2	1
	1	3	1	3	4	11000	8	7.2	7	2	1	2	1
	1	0	3	4	3	10000	7	7.4	5	1	2	2	1

