



# Addis Ababa University Office of Graduate Study

## Faculty of Science Department of Statistics

### **Analysis of Factors Associated with Vehicle Collisions Resulting in Property Damage in Traffic Accidents (The Case of Addis Ababa)**

By Kasahun Hailemariam

A THESIS SUBMITTED TO THE OFFICE GRADUATE STUDIES OF ADDIS  
ABABA UNIVERSITY, IN PARTIAL FULFILLMENT FOR THE AWARD OF  
MASTER OF SCIENCE IN STATISTICS

December 2008  
Addis Ababa

Addis Ababa University Office of Graduate Study  
Faculty of Science  
Department of Statistics

**Analysis of Factors Associated with Car Collisions  
Resulting in Property Damage in Traffic Accidents  
(The Case of Addis Ababa)**

By Kasahun Hailemariam Beyene

Approved by the Board of Examiners:-

Butte Gotu (Dr.)

Advisor

-----  
Signature

M.K Sharma (PhD)

Examiner

-----  
Signature

<b>Content</b>	<b>page</b>
Table of Content.....	i
Acronyms/List of Abbreviation.....	iii
Acknowledgment.....	iv
Abstract.....	vi
Chapter One: Introduction.....	1
1.1. Background.....	1
1.2. Road Traffic Accident Risk and Motorization.....	3
1.3. Road Safety Measures in Ethiopia.....	5
1.4. Mode of Transport System and Traffic accident in Ethiopia.....	8
1.5. Factors contributing to Road Accidents.....	9
1.6. The Study Area.....	10
1.7. Statement of the Problem.....	11
1.8. Objectives.....	13
1.8.1. General Objectives.....	13
1.8.2. Specific Objectives.....	13
1.9. Significance of the Study.....	13
1.10. Limitation of the Study.....	13
1.11. Organization of the Paper.....	14
Chapter Two: Literature Review.....	16
Chapter Three: Methodology.....	24
3.1. Data.....	24
3.2. Variables.....	24
3.3. The Model.....	27
3.3.1. The Gamma Count Model.....	28
3.4. Parameter Estimation and Test Statistics.....	34
3.4.1. Parameter Estimation.....	34
3.4.1.1. The Dispersion Parameter.....	34
3.4.1.2. Estimation of Coefficients of Explanatory Variables.....	38
3.4.2. Statistical Test of Parameters.....	39
3.5. Chi Square Test of Independence .....	41

Chapter Four: Result and Discussion.....	44
4.1. The Goodness of Fit.....	44
4.2. Variable Selection.....	45
4.3. Descriptive Statistics.....	47
4.3.1. Vehicle Related Factors.....	48
4.3.2. Site Factors.....	51
4.3.3. Road Factors.....	51
4.4. Gamma Regression .....	52
4.4.1. Saturated Versus Reduced Model.....	53
4.4.2. Interpretation of Coefficients of Gamma Regression.....	53
4.4.2.1.Vehicle Related Factors.....	54
4.4.2.2.Site Factors.....	55
4.4.2.3.Road Factors.....	55
4.4.3. Prediction Using Estimated parameters.....	55
4.4.3.1.Mean of the Model.....	55
4.4.3.2.Variance of the Model.....	57
Chapter Five: Conclusion and Recommendation.....	58
5.1. Conclusion.....	58
5.2. Recommendation.....	59
References.....	60
Annex1: Descriptive Statistics and Histogram of Response Variable.....	65
Annex2: Poisson Regression Model.....	67
Annex 3: Saturated Gamma Regression Model.....	74
Annex 4: Reduced Gamma Regression First Model.....	81
Annex 5: Reduced Gamma Regression Second Model.....	85
Annex6: Definition of Design matrix.....	89
Annex 7: SAS Codes.....	94
Annex 8: Descriptive Statistics for Significant Variables.....	95

## Acronyms/List of Abbreviations

<b>AATCID:</b>	Addis Ababa Traffic Control and Investigation Department
<b>AA:</b>	Addis Ababa
<b>ADT:</b>	Average Daily Traffic
<b>AU:</b>	African Union
<b>BNN:</b>	Bayesian Neural Network
<b>BPNN:</b>	Backward Propagation Neural Network
<b>CSA:</b>	Central Statistical Authority
<b>E.C:</b>	Ethiopian Calendar
<b>EB:</b>	Empirical Bayes
<b>ERA:</b>	Ethiopian Road Authority
<b>FTCID:</b>	Federal Traffic Control and Investigation Department
<b>GDP:</b>	Gross Domestic Products
<b>GLM:</b>	Generalized Linear Model
<b>HBM:</b>	Hierarchical Bayes Models
<b>LMP:</b>	Low Mean Problem
<b>ML: -</b>	Maximum Likelihood
<b>MLE:</b>	Maximum Likelihood Estimator
<b>NB:</b>	Negative Binomial
<b>NR:</b>	Newton Raphson
<b>NRSCO:</b>	National Road Safety Coordination Office
<b>OLS:-</b>	Ordinary Least Square
<b>PA:</b>	Paved Shoulder
<b>SAS:-</b>	Statistical Analysis Software
<b>SPF:</b>	Safety Performance Function
<b>TRADS:</b>	Traffic Accidents Data System
<b>TRL:</b>	Transport Research Laboratory
<b>UK:</b>	United Kingdom
<b>UN:</b>	United Nations
<b>UNECA:</b>	United Nation Economic Commission for Africa
<b>USA:</b>	United States of America

## Acknowledgements

First, and foremost, I thank God for giving me the opportunity to pursue my graduate study at Science Faculty Department of Statistics, Addis Ababa University.

I am very grateful to Dr Butte Gotu, my thesis research advisor, for his invaluable comments and suggestions, which contributed to the successful realization of the study. I have benefited a lot from his advice. My deep appreciation and thanks goes to him for his critical reviewing of my work, read it tirelessly, commented and guided my research work.

My special thanks also go to all staff of the Department of Statistics especially to Department Head, Ato Sileshi Fanta and to my classmates, for all the encouragements and support that they have provided me during my study years.

This research work is made possible through the generous support and wholeheartedly encouragement of various institutions and individuals. Among these, I would like to express my gratitude Addis Ababa Traffic Investigation and Controlling Department and National Safety Council Office specially Ato Abebe Asrat and Habtamu Abuye for the provision of summarized data and useful reference materials.

My profound appreciations also go to Ato Abbas Kedir and his wife W/ro Amina Aman, without whom the whole process of undertaking this research paper would have not been possible. He helped me in reviewing the paper, in SAS software, and data cleaning. His kindness, encouragement and moral support and advice in every aspect will be unforgettable.

My profound appreciation also goes to Kinfe Hailemariam, Bible Mulate, Dr. Semu Mitiku, Girma Tadese and Comander Melaku Belew for their valuable advices and moral support in all conditions.

On final, I am very much indebted to my parents especially to my wife Kibenesh Mechesa, and all my parents without whom encouragement, my work would not have been reached this stage.

*Kasahun Hailemariam Beyene*

## Abstract

The numbers of vehicles crashed due to traffic accidents have a great impact on the economy of the societies. There are several approaches that researchers have employed to study this problem. This study will attempt to address factors that affect the number of vehicles crashed resulting in property damage due to road traffic accidents in Addis Ababa.

This study is based on a secondary data obtained from Addis Ababa Traffic Control and Investigation Department in 18 month period. This paper is focused on traffic accidents resulted in property damage only. For analyzing this data we use Gamma regression model because this is count data and also under dispersed.

In order to establish the required relationships we use 22 explanatory variables with a total of 132 levels. Among the 22 explanatory variables only five variables with a total of 36 levels that have significant association with response variable at 10%  $\alpha$  level significance remains in the model for analysis.

Findings of this study have shown all levels of light condition, all types of vehicles, all vehicle services years expect vehicles with service years between 2-5 years, all types of collisions expect collision with non-living objects, and all types of road junctions have significant effect on the number of vehicles crashed that resulted in property damage. The result also shows that the coefficients of all vehicle types have negative sign. This indicates that type of vehicles change the mean the response variables; i.e. the number of vehicles crashed per incidence of accidents changes by a value lies between 0 and 1. However, coefficients of levels other variables have positive sign and hence change the mean of the response variables in more than 1. Since the coefficients of level vehicles with service years between 2-5 years and collision with non-living objects are not significantly different from zero, they have no effects on the mean of the response variable.

# Chapter one: Introduction

## 1.1. Background

Rapidly growing population and urbanization result in a significant increase in number of vehicles and traffic accidents as well. Unfortunately, nearly one million people lose their lives and about ten millions face different injuries worldwide in various traffic accidents every year. Additionally, millions of dollars are lost due to accidents and a significant amount of money is spent for health treatment and rehabilitation. All these have made transportation planners and transportation engineers to consider road safety more seriously. Designing safer roads and/or rehabilitation of current road geometries are among important factors in reducing traffic accidents.

The direct global average cost of road traffic crashes over the last five years is about US\$500 billion annually, while the estimated cost to developing nations is US\$65 billion, almost about eight times the total amount of development assistance sent to such nations every year. Indirect costs to victims, families and governments, such as potential income and societal contributions lost, are not included in these calculations. The average cost of crashes in low- and middle-income countries has been between 1% and 1.5% of their GDP (Lawren et. al., 2005).

Road traffic injuries can be considered as a global epidemic and the situation is getting worse. A strong case, therefore, can be made for reducing road traffic accidents in these countries on economic grounds alone, as they consume massive financial resources that the countries concerned can ill-afford to lose.

In fact, road crashes are both predictable and preventable. The causes of road traffic accidents have been established: excessive speed, consumption of drugs

and alcohol, failure to use protective measures such as seatbelt and helmets, poor vehicle impact protection, poor road design etc.

Road traffic accident is the result of the interaction between humans, vehicles and road infrastructure as well as subject to traffic regulations. In this process, the interaction between humans is the key element as nearly all traffic accidents are due to human error (Lu 2006).

Whenever one drives a motor vehicle, there is always some degree of risk of being involved in a crash. In some crashes, a vehicle problem or an environmental condition may significantly contribute to a crash. However, the transport research laboratory indicates that human error was the sole factor in road traffic vehicle collision. According to Chandrartna et al. (2006), drivers who had been repeatedly involved in crash accidents have an ever greater likelihood of being in a crash accident than others. Many researchers have attempted to determine which variables are the best crash predictors and revealed that a number of demographic (e.g. Age, Sex, Educational background etc), psychological (e.g. aggression), Situation (e.g. city size) and behavioral (e.g. risky driving practices, road violations) factors increase an individual's risk for crashes in general.

Comparisons of international road accident fatality rates (expressed as deaths per 10,000 vehicle licensed) show that African countries, particularly Ethiopia, have the highest rates worldwide. For example, Great Britain has about 3500 road deaths each year with some 25 million motor vehicles licensed; this gives a fatality rate of about 1.4. Conversely, in Ethiopia in 2007, there were 2500 reported deaths but with only 220,000 licensed vehicles. This shows that the rate of fatality in Ethiopia was 114, giving a fatality rate about 81.17 times greater than Great Britain.

The numbers of vehicles crashed due to traffic accidents have a great impact on the economy of the societies. In recent years, researchers have paid increasing attention to determine factors that significantly affect severity of damages caused by traffic accidents. There are several approaches that researchers have employed to study this problem: number of injuries, number of deaths etc.

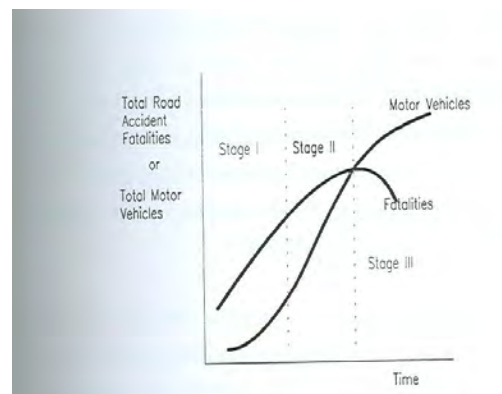
## **1.2. Road Traffic Accident Risk and Motorization**

As experienced in many developed countries, the development of motorization generally follows a typical trend of the S-shaped curve as shown in Figure 1-1 even though the development could be interrupted for some period now and then due to transient economic difficulties in different countries. This overall trend has been characterized into three distinctive stages of development (Berhanu, 2000), namely: developing motorization, exploding motorization and saturation motorization. Along with these stages of development, the issue of road traffic safety has been given different levels of importance and consequently road accident fatalities tend to follow a typical trend shown in Figure 1-1.

These two curves in Figure 1-1 show the general relationship between motorization and safety at a given development stage of a country. In the initial stage of motorization, the population has little knowledge of how to safely integrate the motor traffic into the system. Safety is not considered as a major issue in the design of roads and in vehicles. There is little experience of comprehensive traffic regulation. Road users are also less aware of how to act safely in the motor-traffic system. During this stage, road accident fatalities, injuries and property damage rise with the increase in motorization and population. In the second stage, as the traffic hazard increases with high rate of motorization, awareness and concern develops. It is at this time both the

government and the public should start to mobilize human and economic resources to work against the road-traffic hazard.

Improvements in the areas of road and vehicle design as well as influencing human behaviour by means such as education, legislation, and enforcement start to be implemented. Most important, at this stage, there is an increased understanding of the man-vehicle-road-regulation system. In terms of road safety, this stage of development is characterized by the decrease in the rate of results of traffic accidents per vehicle. The third phase of development starts when road-traffic accident peaks and begins to decline. This is the historical achievement of a number of industrialized countries in the early 1970s (Berhanu 2000).



*Figure 1-1: Typical trend for the development of motorization and traffic accidents*

In the third development stage, traffic accident fatalities, injuries and property damages continue declining with the implementations of the proven safety improvements on black spot road sections, drinking and driving, speeding, and vehicle safety. To date, promising new developments are being achieved in many countries through targeted road safety programmes. The concept emerged in the United Kingdom for accident prevention, safety auditing, is being practiced in a number of developed countries such as the UK, the United States, Australia,

New Zealand, Norway and Denmark. The demand of the society in the developed countries is now for a sustainable and safe traffic environment through:

- a road infrastructure where the design is adapted to the limitations of human capacity;
- vehicles which minimize the task of the road users are constructed to protect human beings effectively; and
- A road user, who is educated adequately, informed, and where necessary controlled.

As a matter of fact, there cannot be any distinct levels of motorization to classify the development into these phases applicable to all countries. The speed of motorization mainly depends on the economic situation of a country, while the status of the traffic safety depends on the country's resources, attitudes, and efforts to combat the problem. For more perception, however, Jorgensen (1996) suggests typical levels of motorization for these stages. Thus, the first stage represents early motorization up to 50-100 vehicles per 1000 inhabitants, the second stage goes up to 400 vehicles per 1000 inhabitants, and the third stage is over 400 vehicles per 1000 inhabitants.

A number of developing countries are at present in the first stage. Ethiopia with a level of motorization of 30 vehicles per 10,000 inhabitants in 2007/8 is among them. The development of motorization, as described above, has been very low due to unaffordable economic scale of the country.

### **1.3. Road Safety measures in Ethiopia**

Each year, road traffic accidents cause a great loss of human and economic resources in Ethiopia. This problem is increasing from year to year at alarming

rate accompanying rapid increase of population and vehicles. Compared with international risk figures, the country is one of the worst examples in terms of fatality rate per vehicle (See Annex 1 Figure3)

Roads are generally planned and built to move people and goods from one place to another. Within these traffic functions, however, individual roads may serve parts of the total travel process in particular. Some roads cope with long distance travel only; others play a role as a distributor, and some roads just serve as direct access to properties alongside or parking of vehicles. To ease the contradictions between these functions and enable roads to fulfill their roles satisfactorily, roads are generally classified based on flow function (rapid processing through traffic), distributor function (making district and regions accessible) and access function (allowing properties to be reached).

Considerations of these functions are very important during planning, designing and operation of roads. Preventing misuse of roads and streets makes the traffic safer. One of the main problems of road transport system in Ethiopia today is, however, that roads and streets are expected to fulfill incompatible functions at the same time (Berhanu 2000).

According to Berhanu (2000), intercity highways are purposely planned to pass through towns and villages mainly for commercial benefits for these places. These roads, however, also carry high-speed through traffic without speed reduction features and cause serious traffic safety problems in such towns and villages. The prior victims of this situation are unaware pedestrians and operators of animal drawn carts.

All of these factors and others are the results of poor road planning and the substandard geometry and facilities of roads in the country. In dimensioning and aligning the design of roads, proper safety considerations are not taken. Roads

and their environment could, thus, be considered as the most important factors for the traffic accidents in Ethiopia. In many countries, design engineers traditionally consider several criteria during the design stage of roads. Designing a road is predominantly an art of giving the right weight to these criteria in order to find the most satisfactory solution. Safety is among the criteria that is taken into account most implicitly. The designer is supposed to take decisions with safety in mind at every step in the design process and decisions are rarely taken exclusively for the sake of safety. A new tool known as the safety audit for considering road safety issues as explicitly as possible during the planning phase is, however, now in operation in some developed countries.

In most countries, road standards have been set in order to help engineers consider road design criteria appropriately in the processes of road design. One of the major advantages of road standards is the possibility of incorporating safety issues uniformly and appropriately in the design of roads for a given traffic situation. Moreover road standards ensure uniformity between different designs, which makes traffic situations and road user behaviour more predictable for safe traffic operations.

The following are some of the inadequacies with respect to safety in the general road design and traffic operation practices, which can be found in Ethiopia:

- Sharp horizontal curves placed after long tangents without transitional curves on major high speed roads;
- Long steep grades on major roads carrying a high volume of heavy goods vehicles without dedicated climbing lanes, thus creating high speed differentials;
- Inadequate sight distance on major roads;
- Inadequately designed and .controlled road junctions for the efficient and

- safe movements of the prevailing volume and characteristics of traffic;
- Absence of pedestrian crossing facilities and inadequate footpaths which segregate pedestrians from the motor traffic;
  - Narrow bridges without protected footpaths for pedestrians and animals on the bridge, and bridges located on sharp curves; and
  - Absence of road accessories, which support efficient and safe use; such as road signs, marking, channelization, safety fences and barriers, and reflectors

These and other inadequacies of roads and their environment, generally, result in inefficient and unsafe traffic with a lot of uncertainties among road users leaving them with complex-decision making process. This brings us to the issue of the imperative need for understanding the influence of road and traffic factors for improving the existing roads and establishing comprehensive standards of planning and design for new roads. Incorporating safety in the road network along with better understanding of the characteristics of the prevailing road traffic is a scientific concepts leading towards sustainable and safe road traffic.

#### **1.4. Mode of Transport System and Traffic Accidents in Ethiopia**

Understandably, the modern transport system in the country is not well developed to give adequate service and stimulate development. Generally, in Ethiopia there are three main modes of transport system namely land, water and air. The land road transport sector covers 95% of passenger and 99% of freight cargo movement. Therefore, road transport system remains the only mode of modern transport that Ethiopia heavily relies on for both domestic as well as for international transport services in the country. Ethiopia has one of the most successful airlines in Africa, but air transport represents only a small fraction of local out of the total transport mainly due to the lack of airport infrastructure and

unaffordable airfare to most Ethiopians. Oldest Addis Ababa-Djibouti railway is the only major rail link in Ethiopia today connecting the capital to Djibouti seaport, which passes through two important cities in the country, Dire Dawa and Adama (Nazareth). At the moment it is not regularly operating.

### **1.5. Factors Contributing to Road Accidents**

Basically, traffic accidents are connected to the three principal factors in traffic: the road user, the vehicle, and the road. The extent of the effects of these factors, however, is more domestic in nature and can only be identified more precisely by on-the-spot accident investigation. Such in-depth accident investigation, in general, requires skilled manpower and resources. Its application, therefore, is limited to some developed countries. In Ethiopia, investigations are based on general observations and the reports by the police.

The traffic police usually report what they consider the primary causes of accidents. The causes reported are not really as accurate as those identified in on-the-spot accident investigations. Normally, the reports by the police indicate a single cause of each accident and do not list all factors involved as in on-the-spot accident investigations. Furthermore, the effects of some factors could be underestimated. Nevertheless, the police reports do provide some insight into the major causes for road accidents.

Road safety researchers worldwide agree that minimal progress can be achieved in reducing traffic accidents unless the problem itself is clearly defined. This indicates the need for detailed traffic accident recording and storing of relevant accident information in accident databanks over a long period. In general, the traffic safety situation in any country can primarily be determined by conditions of the following four elements:

- The general level of traffic-safety awareness and behaviour of road users;
- The safety standard of the road and its environment;
- The quality of the vehicle fleet; and
- The availability of comprehensive and adequate legislation and enforcement.

The complex nature of these components and the sensitive interaction between them necessitates that they should be dealt with in a systematic and proper manner to improve road safety. Unfortunately, the unfavorable circumstances connected with these elements individually and collectively are contributing to the fundamental issues that collectively constitute the problematic traffic-safety situation in Ethiopia (Berhanu 2000).

## **1.6. The Study Area**

Addis Ababa is the capital and a primary city of Ethiopia. It extends over 540 Sq. Km and is located at an altitude of about 2,408m above sea level at 9.02° N 38.44° E. Administratively the city is identified into 10 sub-cities and 99 kebeles. Addis Ababa has an estimated total population (for July 2006) of 2,973,000, consisting of 1,428,000 male and 1,545,000 female. It is estimated that presently there are no rural parts to the city, so 100% of the inhabitants are considered as urban dwellers; 24% of all urban dwellers in Ethiopia are in Addis Ababa. The City has an estimated density of 5,607.96 people per square kilometer (CSA, 2005).

Addis Ababa is growing and changing very fast. It is on the threshold of becoming an international city. Physical infrastructural systems are critical in sustaining and promoting growth and change. Presently Addis Ababa serves as a transport hub of the nation. It is catered by an international airport and a

reasonable extent of road network. But the city transport system suffers from many inadequacies. The primary road of the city includes two east-west and north-south axial, a newly built ring road and a number of other roads.

Addis Ababa is the headquarters of many international organizations including the UNECA and the AU . Public transportation is through public buses or midi-buses or blue and white share taxis , locally known as "blue donkeys". The taxis are usually minibuses that can sit at least twelve people. Two people are responsible for each taxi, the driver and his assistant who collects fares and calls out the taxi's destination (Wikipedia, 2006). Non-motorized transport is also an important component. It consists mainly of walking. Large shares of person trips are estimated to be performed by walk. Animals are used for carrying loads but their number is few (ERA 2005).

### **1.7. Statement of the Problem**

According to Transport Research Laboratory (TRL 2001), the most recent data available show that every month around 400 people are killed or hospitalized by road traffic accidents in Ethiopia. Road traffic accidents are continuing to rise even though fatalities due to road accidents have decreased slightly by 2% since 1995. According to this report there are 13302 number of road traffic accidents. Out of this 7783 led to property damage only. The number of accidents in Addis comprises 58.5% of the total traffic accidents in Ethiopia.

From Ethiopian economic perspective the road traffic accident cost has been estimated to be in between 350-430 million Birr loss annually. This situation is worsening and likely to be even more severe than official statistics indicates, due to under reporting of road traffic casualties.

As per the National Road Safety Coordination Office (NRSCO), there were at least 114 fatalities per 10,000 motor vehicles, almost 2500 lives lost, 8000 reported injuries and 10170 resulting in property damages only, on average, every year over the ten years through out the country.

The tragedy is more or less similar in Addis Ababa. The rate of traffic accidents in Addis Ababa goes up together with the increase of motor vehicles and population size. The rise in automobile ownership together with the poor condition of the roads has resulted in the high level of traffic safety and congestion problems. The share of Addis Ababa City in the total number of accidents was 60 percent in 1989 with annual average traffic accident growth of 31.4 percent. In 2005, 700 accidents per month were reported (Tesema et al., 2005). As per the recent (2006/7) records the share of Addis Ababa is about 53% of the total traffic accident of the country.

At all levels, whether at national or international level, road traffic accidents continue to be a growing problem. In connection with this, according to a World Health organization/World Bank Report, deaths from non-communicable diseases are expected to grow from 28.1 million a year in 1990 to 49.7 million by 2020, which is an increase in absolute number of 77%. Traffic accidents are the main cause of this rise. Road traffic injures are expected to take higher place in the rank order of disease burden in the near future (Mekonnen, T. 2007).

The consequences of traffic accidents are categorized as: Deaths, Heavy Injuries, Light Injuries and Property damage only. Property damage is a type of accident that does not include any injury on human life (like deaths, heavy injuries and light injuries). The focus of this paper is on traffic accidents resulting in property damage only.

This study concentrates on the number of vehicles crashed per accident and will attempt to address the following questions:

- What are the variables that mainly determine the number of vehicles crashed in accidents resulting in property damage only?
- What are the impacts of significant variables on the number of vehicles crashed in accidents resulting in property damage only?

## **1.8. OBJECTIVES**

### **1.8.1. General Objectives**

The general objective of this study is to identify the major factors that are associated with vehicle crash and determine the number of vehicles crashed per traffic accident resulting in property damage only.

### **1.8.2. Specific Objectives:**

- To identify and describe the major factors that contributes to vehicle crashes in traffic accidents resulting in property damage only.
- To analyze the impact of demographic, geographic and vehicle related variables.
- To fit appropriate model to estimate the number of vehicles crashed per accident resulting in property damage only.

## **1.9. Significance of the Study**

To find ways of minimizing the risk of accident that result in property damage and design appropriate safety measures.

### **1.10. Limitations of the Study**

The content of the road accident records in AATCID missed relevant details of an accident report required for any road safety improvement work. The reporting form, in the daily report book, is not designed to include details of each vehicle and road user involved in an accident. The report, further, does not contain details of the road section and the precise location of an accident. The location of an accident is reported broadly by Kebele and Woredas or the name of the surroundings. Besides, because plain paper is used on the spot, person in charge may not remember the required accident details and as a result the form available at the local traffic police office is usually incomplete.

This study is based on a secondary data obtained from AATCID. The data obtained from AATCID contains records of road traffic accidents, number of vehicles crashed and the number of injuries (which includes fatalities, serious and light injuries) in each accident. The records show that there were 17147 traffic accidents in 1999 E.C. (2006/7) at national level out of which 10170 resulted in property damage only. Out of 10170 accidents that resulted in property damage 6953 (68.37%) took place in Addis Ababa. Since some of the records of accidents in Addis Ababa are incomplete, only 4163 records of the 1999(2006/7) accidents were considered in this study. In addition, out of 4167 records of accidents in 1998 E.C. only 2813 records were considered. Therefore, a total of 6976 records were considered for analysis.

### **1.11. Organization of the Paper**

The paper is presented in five chapters. The first chapter, introduction, deals with general background of the study, road traffic accident risk and safety measures in Ethiopia. The statement of the problem, objectives, the study area,

significance of the research and limitation of the study are also addressed in this chapter.

The second chapter reviews works related to the major problem being investigated.

In the third chapter, the method of analysis; namely the variables considered, model fitting and model diagnostics are discussed. In addition, the assumptions of the model under consideration are discussed.

Chapter four is devoted to the result and discussion of the estimated parameters of the model. Here, the contribution of each random variable, test of parameters and model based results are described.

The last chapter, Chapter five, gives conclusions drawn from the result of study.

## Chapter two: Literature Review

The characteristics of road traffic are domestic in nature reflecting the interactions of three components: road, road users, and vehicles. This, in turn, suggests the domestic nature of the characteristics of traffic accidents derived from the safety level of these components in the traffic system. The summary statistics on traffic accidents indicating the magnitude of the total problem do not identify the sources of failures in the system in detail and therefore, possible remedies can not be introduced from overall figures. For detailed analysis, it is necessary to establish typical characteristics of traffic accidents at various specific conditions.

Generalized linear models (GLM) and hierarchical Bayes models (HBM) have been the most common types of model favored by transportation safety analysts. Over the last few years, researchers have proposed the back-propagation neural network (BPNN) model for modeling the phenomenon under study. Compared to GLM and HBM, BPNN have received much less attention in highway safety modeling. The reasons are attributed to the complexity of estimating the model as well as the problem related to “over-fitting” the data. To circumvent the over-fitting problem, some statisticians have proposed the use of Bayesian neural network (BNN) models. These models have been shown to perform better than BPNN models while at the same time reducing the difficulty associated with over-fitting the data (Xie, et al, 2007).

Moreover, the influence of traffic volume and road geometry elements on accident rate has been a subject of various scientific studies. There is an interesting relationship between the number of accidents and traffic volume such

that accident rate increases for both low and high traffic volumes. In the low traffic volume, mostly individual vehicles are involved in accidents, while in the higher traffic volume, group of vehicles are involved in accidents in general. Road design parameters including lane and shoulder widths and numbers, sight distance, and curve characteristics, etc. can play a significant role on the type of accidents mentioned previously (Akgungor et al. 2007).

Among others, Akgungor et al. (2007) undertook research using “change one-factor-at-a-time” approach regarding sensitivity of accident rate on road geometry elements such as lane width, shoulder width and type, and median using data obtained from Federal Highway Administration Department of Transportation, USA. The result of the research indicated that increasing lane width from 9 to 12 ft decreases fatal accident rate by 40.5%, while non fatal accident rate such as light, serious and property damage decreases by 19.8%. On the other hand, decreasing total lane width from 48 to 40 ft increases accident rate by 25%. Using previous study results Akgungor et al. (2006) estimated that increasing each lane width by 0.25 m decreases accident rate by 2 to 2.5%. As quoted by Akgungor, a number of studies investigating the effect of lane width on accident rate were done by Zegeer et al, (1987), Hadi et al.(1995), Zegeer and Council (1995), Miaou (1996) and Wang et al. (1998). Overall, these studies demonstrate that lane width has a significant influence on accident rate.

Generally, the sensitivity of a selected accident prediction model and its model parameters was analyzed by the fractional factorial analysis method. In the study, the accident prediction model proposed by Zegeer et al. (1987) was used since it incorporates both traffic and road geometry parameters in addition to terrain characteristics. The evaluation of sensitivity analysis revealed that main parameters average daily traffic (ADT), width (W), width of paved shoulder (PA) and median (H), and two-parameter interactions ADT-W, ADT-PA and ADT-H

have significant effects on number of accidents. Based on the absolute value of parameter effects at the  $3\sigma$  and  $2\sigma$  thresholds, ADT was found to be of primary importance, while the remaining identified parameters and parameter interactions seemed to be of secondary importance. This is in agreement with the fact that ADT is an effective parameter in determining road geometry and hence is directly related to number of accidents. The sensitivity analysis shows that the fractional factorial design method is an efficient way of examining the relative importance of the selected accident prediction model parameters (Akgungor et al. 2006).

Regarding crashes, Lord (2006) analyzes motor vehicle crashes using Poisson-Gamma model for road traffic accidents data collected in Toronto, Ontario. In this study the effect of low sample means values and a small sample size on the estimation of the fixed dispersion parameter is examined. It was pointed that crash data collected for safety studies often have unusual attributes of being characterized by low sample mean values. Previous studies have shown that the goodness-of-fit of statistical models produced from such datasets can be significantly affected. This issue has been defined as the “low mean problem” (LMP). The study indicated that the dispersion parameter plays an important role in many types of safety studies and should, therefore, be reliably estimated.

To this end a series of Poisson-gamma distributions were simulated using different values describing the mean, the dispersion parameter, and the sample size. Three estimators commonly used by transportation safety modelers for estimating the dispersion parameter of Poisson-gamma models were evaluated: the method of moments, the weighted regression, and the maximum likelihood method. In an attempt to complement the outcome of the simulation study, Poisson-gamma models were fitted to crash data collected in Toronto, Ontario which was characterized by a low sample mean and small sample size. The study

shows that a low sample mean combined with a small sample size can seriously affect the estimation of the dispersion parameter, no matter which estimator is used within the estimation process. The probability the dispersion parameter becomes unreliably estimated increases significantly as the sample mean and sample size decrease. Consequently, the results show that an unreliably estimated dispersion parameter can significantly undermine empirical Bayes (EB) estimates as well as the estimation of confidence intervals for the gamma mean and predicted response.

According to Xie et al, 2007, a series of models was fitted using data collected on rural frontage roads in Texas. Three types of models were compared: BPNN, BNN and the negative binomial (NB) regression models. The results of this study show that, in general, both types of neural network models perform better than the NB regression model in terms of data prediction. Although the BPNN model can occasionally provide better or approximately equivalent prediction performance compared to the BNN model, in most cases its prediction performance is worse than the BNN model. In addition, the data fitting performance of the BPNN model is consistently worse than the BNN model, which suggests that the BNN model has better generalization abilities than the BPNN model and can effectively alleviate the over-fitting problem without significantly compromising the nonlinear approximation ability. The results also show that BNNs could be used for other useful analyses in highway safety, including the development of accident modification factors and for improving the prediction capabilities for evaluating different highway design alternatives.

The study also pointed out that most traffic crash prediction models (sometimes referred to as safety performance function or SPF) have been based on statistical regression techniques. Initially, linear regression models were the model of choice for modeling motor vehicle collisions. Linear models, when not corrected

for unequal variance, assume that the number of crashes follows a normal distribution. It has been reported in the literature that these models are inadequate for modeling count data, since crash data exhibit non-constant variance.

Due to the inadequacy of linear regression models for analyzing discrete, nonnegative, sporadic, and asymmetrically distributed random events, GLMs have been proposed for modeling crash frequency. Among GLMs, several types of models have been used by researchers, including Poisson regression, Poisson-gamma or NB regression, Gamma regression model and other variations of the NB regression model. It is generally agreed that if the sample variance is significantly greater than sample mean, then NB models should be used in lieu of Poisson regression models. On the other hand if the sample variance is significantly smaller than the sample mean, which is defined as under dispersion, gamma models are the models of choice (Oh et al., 2006). Zero-inflated models (Poisson and NB) have been proposed for modeling crash data with an apparent excess of zero observations, but their application has been discredited when the characteristics and the nature of the data do not warrant the application of such models (Lord et al., 2005b, 2007; Warton, 2005). More recently, Bayesian multivariate generalized linear mixed model (Song et al., 2006), and multivariate Poisson regression model (Miaou and Song, 2005; Ma and Kockelman, 2006; Park and Lord, 2006) have also been proposed for modeling motor vehicle collisions.

According to Yau et al (2006), 'Multiple-vehicle traffic accident' refers to a crash between two or more moving objects. Variables such as drivers, road type, speed limit and number of vehicles involved in the accident are expected to play a much more important role in association with injury severity in multiple-vehicle accidents. To study the factors influencing injury severity of multiple-vehicle traffic accidents, a population-based study was conducted on the data obtained

from the Traffic Accident Data System (TRADS), which was developed by the Transport Department, Police Force and Information Technology Services Department, Hong Kong. Multiple-vehicle traffic accidents occurring during the 2-year period 1999 and 2000 were considered. Potential risk factors such as district, human, vehicle, safety, environmental and site factors were examined. Categorizing injury severity into “fatal/serious” and “slight”, a stepwise logistic regression model was applied to the population data set. The district board, time of the accident, driver’s gender, vehicle type, road type, speed limit and the number of vehicles involved are significant factors influencing the injury severity. Identification of risk factors for severe traffic accidents provides valuable information to help with new and improved road safety control measures.

Moreover, Yau (2004) undertook the same study on the severity of single vehicle traffic accidents in Hong Kong by considering single vehicle data of three major vehicle type namely private vehicles, goods vehicles and motorcycles, using the same model and two year data. The effect of district, Human, vehicle, safety, environmental and site factors on injury severity was examined. The result of the study shows that private vehicle, district board, gender of driver, age of vehicle, time of the accident and straight light condition are significant; for goods vehicle seat- belt usage and weekday occurrence are the only significant factors; and for motorcycles age of vehicle, weekday and time of accident were significant for injury severity.

In Ethiopia, according to the National Road Safety Coordination Office, the number of deaths due to traffic accident is estimated to be 114 per 10,000 vehicles. Of those killed over half are pedestrians, of which 30% are children. It was noted that one out of five people injured dies as a result of road accident, which again is a very high figure mainly due to poor situation of the emergency medical services (National Road Safety Coordination Office (NRSCO), 2007).

Based on a five-year average records NRSCO further states that, of the personal injury accidents, 81 % are caused due to drivers' error, 5% due to vehicle defect, 4% due to pedestrian error, 1 % due to road defects and 9% due to other problems. These figures show that the majority of personal injury accidents are caused as a result of drivers' error. The study further shows that professional drivers are involved in 88% of the fatal accidents. Special purpose vehicles and motor bicycles cause 8% of such accidents. On the other hand automobile drivers have very good safety records with only 4% of the fatal accidents, which is equivalent to a rate of 12 fatal accidents per 10,000 vehicles. Automobiles share is around 43% of the total vehicle number of the country.

The study indicates that fatal accident types were 68% pedestrian strike, 13% overturn, 6% fall from vehicles, 3% animal or cart strike and 10% for all the other remaining crash types. The underling reasons for these accidents are noted to be:

-

- 1) Improper behavior or low skill of drivers (Over speeding, not respecting pedestrian priority...etc.)
- 2) Poor vehicle technical conditions
- 3) Animals and carts using the highways
- 4) Pedestrians not taking proper precautions
- 5) Poor traffic law enforcement
- 6) Poor emergency medical services and
- 7) Safety consideration not sufficiently given in roads developments

Berhanu, (2004) conducted a study on traffic safety with road environment and traffic flows based on the data collected on arterial roads in Addis Ababa. Poisson and negative binomial regression methods were used to relate the discrete accident data with the road and traffic flow explanatory variables.

The results show that the existing inadequate road infrastructure and poor road traffic operations are the potential contributors of the ever-growing challenge of the road transport in Addis Ababa. The results also indicate that improvements in roadway width, pedestrian facilities, geometry of road and access management are effective in reducing road traffic accidents.

Most of the studies conducted on traffic accidents indicate that factors that contribute to the occurrence of traffic accident can be categorized as per their relation to Vehicle, Driver or Road condition. Others like visibility, weather condition, population density, overturning, type of collisions...etc are also additional factors that may be taken into consideration.

Even though many studies have been conducted giving priority to the severity of injuries due to traffic accidents and the factors associated with injuries traffic accident that resulted in property damages and the factors associated with property damage have not been given attention. Based on the fact that traffic accidents resulting in property damage have considerable negative impact on national and global economic, social and environment status this study will attempt to identify and describe major factors that contribute vehicle crashes in traffic accidents resulting in property damage and propose ways of minimizing the risk of accidents.

## Chapter Three: Methodology

### 3.1. Data

This study is based on a secondary data obtained from Addis Ababa Traffic Control and Investigation Department (AATCID). The accident database of the whole Addis Ababa (AA) created for this study contains all accidents reported by the police in the study period, February 2006 to September 2007. A total of 10,079 accidents were reported of which 2620 were injury accidents and the rest 7459 accidents were property damage only. This paper focuses on traffic accidents resulted in property damage only. Due to incompleteness of some of the records a total of 6976 accidents that resulted in property damage only were considered.

### 3.2. Variables

Variables to be included in the study are the following:

**Response Variable:** For the formulation of predictive models in this study, we consider a set of 6976 traffic accident incidents resulted in property damage only. The response variable ( $Y_i$ ) is defined as a discrete random variable representing the number of vehicles crashed per incident  $i$  of traffic accident that resulted in property damage only during a defined study period, where  $i=1, 2, \dots, n$ . The actual observation of  $Y_i$  is denoted by  $y_i$ .

**Predictor Variables:** These variables are sometimes called independent or exogenous or explanatory variables. For this data analysis, the following categories of predictor variables are considered: Demographic, Road, Site and Vehicle related

- **Demographic Factors**: - are factors related to drivers involved in traffic accidents that resulted in property damage. These factors are categorized as follows:
  - Drivers Sex: Male or female
  - Drivers age category:- Below 18 years, 18-30 years, 31-50 years and above 51 years
  - Educational Level of drivers:- Illiterate, Adult education, Primary, Junior, High School and above High School
  - Vehicle-driver relation: - owner, hired, others.
  - Driver Experience:- No license, less than a year, 1-2 years, 2-5 years, 5-10 years, above 10 years.
  - License Grade:- First, Second, Third, Fourth, Fifth and Special
- **Site factors**: - indicate the location, weather condition etc and are classified as follows:-
  - Sub city:- is the administrative autonomous category of the city according to locations by local names like Arada, Kirkos, Bole, Yeka etc
  - Place of Accidents:- The levels of this variable includes Residence, School, Market, Office, Hospital, Factory, Recreational Area, Religious Places
  - Light Condition: Day Light, During Sunrise, During Sunset, Dark with Road Light, Dark with Poor Road Light
  - Weather Condition:- Good Weather, Cold Weather, Heavy Rain, Slight Rain and Others
- **Road Factors**:- are factors related to road and road map and include
  - Road type:- Good Asphalt, Damaged Asphalt, Concrete and Paved
  - Road Direction:- Straight and Flat, Straight with Slight inclination, Down steep, Slightly twisted, Stepped, Straight with more Inclination
  - Road Junction:- Cross Shaped Joint, T-Shaped Joint, Y-Shaped Joint,

Circled Island, No Junction

- Road classification:- One Way Road, Two Way Road, Road Divided by Solid Paints, Road Divided by Islands, Road Divided by Broken Paints
- Road condition:- Wet, Muddy, Dry

**Vehicle Related Factors:** - are factors that aggravate road traffic accidents due to mechanical and /or technical problems. These factors include:

- **Registration Code:** - is the identification code of plate of a vehicle and indicates to which category it belongs. These include Code1, Code2, Code3, Code4 and Code 5.
- **Vehicle Service year:** indicates the years of service of a vehicle: Less than a year, 1-2 Years, 2-5 Years, 3-10 Years, 10 years and above.
- **Ownership:** - this describes the owner of a vehicle. The categories includes; Private, Government, Public Organizations, International Organizations, UN, Defense & Police, Core Diplomat and Others
- **Vehicle type:** - is the type of vehicle that causes accidents. It is classified into the following categories:
  - **Motor Cycle:** - are a two wheel propelled by mechanical engine.
  - **Automobile:** is a private vehicle that has a capacity of holding up to eight seats.
  - **Station Wagon:** - is a four-wheal drive vehicle that has up to 12 seats.
  - **Taxi (TX):**- is a city transport that has a capacity of holding 12 passengers.
  - **Public Bus with 12 seats:** - is a passenger transport that has a capacity of holding 12 passengers.
  - **Public Bus with 13-45:-** is a passenger transport that has a capacity of holding 13 to 45 passengers.

- **Public Bus with more than 45 Seats:** - is a long distance travel buses that has a capacity of holding above 46 passengers.
- **Pick Up:** a vehicle that has a capacity of loading up to 1000Kgs.
- **Freight:** is a vehicle that has a capacity of loading 1100 to 4000 Kgs.
- **Freight with 41-100Q:** is a vehicle that has a capacity of loading 4100-10000 Kgs.
- **Freight with Trailer:-** is any type of dry cargo with trailer.
- **Liquid Freight:-** is a vehicle that is used to transport liquids materials.
- **Special Vehicles:-** is a vehicle used for construction, agriculture etc.
- **Status of Vehicle Movement:-** like while moving straight, while turning left, right, while moving back etc
- **Type of collision:** - Head-To-Head, Head-To-Tail, Head-To-Side, Side-To-Side, Collision with Non-living Object(s), Collision with parked vehicle(s), Collision of Pedestrian, Overturning

### 3.3. The Model

Statistical or crash prediction models have frequently been used in highway safety studies mainly to identify major contributing factors or establish relationships between crashes and explanatory variables, such as traffic flows, types of traffic control, and highway geometric variables, with the aim that effective countermeasures could be implemented to reduce the number and severity of motor vehicle collisions occurring on different types of highway entities. The models can also be used to predict crash frequencies on sites that have not been used for estimating the original models or with different traffic flow and highway geometric conditions (Xie et al. 2007).

As discussed in chapter two, standard linear regression models are inadequate for analyzing discrete, nonnegative, sporadic, and asymmetrically distributed

random events. Previous studies that documented the development and application of crash prediction models have usually focused on statistical techniques that are based on the generalized linear modeling (GLM) framework. GLMs have been very popular because they have explicit theoretical foundations and can produce coefficients for each explanatory variable included in the model (Xie et al 2007).

Among GLMs, Gamma regression model is used for analyzing data whose mean is greater than the variance i.e. under dispersion. The summary statistics for the number of vehicles crashed per incidence of accidents for the data considered in this study given in Table 1 of Annex 1, shows that the mean is greater than the variance. This leads us to use gamma regression model for the analysis of accident data of Addis Ababa.

### 3.3.1. The Gamma Count Model

The Gamma count model can be used to model data sets that show under dispersion (Winkelmann 1995). To apply the gamma model proposed by Winkelmann on the observed discrete random variable  $y_i$  for  $i=1, 2, \dots, n$  we assume that the  $y_i$ 's are identically and independently gamma distributed (Oh et.al 2006). It is assumed that the number of vehicles crashed per incidence  $i$  is identical and independent of each other. This means the occurrence of vehicles crashed do not depend either on the previous or on the next accident happened. According to Winkelmann (1995), the count gamma model is derived from continuous gamma model given by:-

$$f(y, \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y} \quad \text{where } \alpha, \beta \in \mathfrak{R}^+ \dots \dots \dots (1)$$

for  $y > 0$ . It has mean  $E(y) = \frac{\alpha}{\beta}$  and variance  $Var(y) = \frac{\alpha}{\beta^2}$  for continuous data. The gamma probability model for count data can be expressed as:

$$Pr [y_i=j] = \text{Gam}(\alpha j, \lambda_i) - \text{Gam}(\alpha(j+1), \lambda_i) \dots \dots \dots (2)$$

where:  $\lambda_i = \exp(\beta X_i)$

$$\text{Gam}(\alpha j, \lambda_i) = \begin{cases} 1 & \text{if } j = 0 \\ \frac{1}{\Gamma(\alpha j)} \int_0^\infty u^{\alpha j - 1} e^{-u} du & \text{if } j > 0 \text{ and } j = 1, 2, 3, \dots \end{cases}$$

According to this model, if the dispersion parameter,  $\alpha$ , is greater than 1; then the model is said to be under dispersed, if the value of  $\alpha < 1$ , then the model is said to be over dispersed. Otherwise, if the value of  $\alpha = 1$ , then model is said to be equi-dispersed, which in turn reduces the gamma probability model to the Poisson model. Due to the relative scarcity of the gamma probability count model in the transportation literature, the conditional mean function given by the following formula is used.

$$E[y_i | X_i] = \sum_{j=1}^{\infty} j(\alpha j, \lambda_i) \dots \dots \dots (3)$$

And the cumulative distribution function of Gamma distributed random variable is:

$$\begin{aligned} F(Y | \alpha, \lambda_i) &= \int_0^Y \frac{\lambda_i^{\alpha j}}{\Gamma(\alpha j)} u^{\alpha j - 1} e^{-\lambda_i u} du, \alpha > 0, \lambda_i > 0 \\ &= \frac{1}{\Gamma(\alpha j)} \int_0^Y u^{\alpha j - 1} e^{-u} du, \quad j = 0, 1, \dots \dots \dots (4) \\ &= \text{Gam}(\alpha j, \lambda_i Y) \end{aligned}$$

Winkelmann (1995) has developed a more generalized gamma probability model given in (1) for count data. To apply this model to count data, consider n random events of the number of vehicles crashed per incidents of accident each with gamma distributed. Let the sum of these events be  $v_n$ . Then  $v_n$  can be written as:

$$v_n = y_1 + \dots + y_n, \quad n = 1, 2, \dots \quad (5)$$

Where  $\{y_i\}$  are series of i.i.d. Gamma distributed. As quoted by Winkelmann (1995) Johnson and Kotz, 1970 show that the reproductive property of the gamma distribution implies that  $v_n$  is also gamma distributed with density function:

$$f_n(v, \alpha, \beta) = \frac{\beta^{n\alpha}}{\Gamma(n\alpha)} v^{n\alpha-1} e^{-\beta v} \quad \dots \quad (6)$$

To derive the new count data distribution, we have to evaluate the cumulative distribution function

$$F_n(Y) = \frac{1}{\Gamma(n\alpha)} \int_0^{\infty} u^{n\alpha-1} e^{-u} du, \quad n = 1, 2, \dots \quad (7)$$

where the integral is the incomplete gamma function. The right hand side will be denoted as  $G(\alpha n, \beta Y)$ . This coincides with the expression given in equation (2)

Integrating (7) by parts gives (see in the Last Annex)

$$G(n, \lambda_i) = 1 - e^{-\lambda_i} \left( 1 + \lambda_i + \frac{(\lambda_i)^2}{2!} + \dots + \frac{(\lambda_i)^{\alpha n - 1}}{(\alpha n - 1)!} \right) \quad \dots \quad (8)$$

Hence,

$$\begin{aligned} \Pr [y_i = n] &= G(\alpha n, \lambda_i) - G(\alpha n + \alpha, \lambda_i) \\ &= e^{-\lambda_i} \sum_{i=1}^{\alpha-1} \frac{(\lambda_i)^{\alpha n + i}}{(\alpha n + i)!}, \quad n = 0, 1, 2, \dots \quad (9) \end{aligned}$$

For  $\alpha = 1$ ,  $f(y)$  is the exponential density and (9) simplifies to the Poisson distribution. It was noted that the Poisson distribution is characterized by independently exponentially distributed random variable for the events of occurrence (number of vehicles crashed per incident) .

For non-integer  $\alpha$ , no closed-form expression is available for  $G(\alpha n, \lambda_i)$  (and thus for  $P\{y_i=n\}$ ). As quoted by Winkelmann 1995, Abramowitz and Stegun (1964), and Bowman and Shenton (1988) show that the numerical evaluations of the integral is based on asymptotic expansions of the normal distributions when  $\alpha$  is non-integer.

According to Winkelmann (1995), for a given  $\{Y_i\}$  sequence of independent, positive, identically distributed random variables and  $Y(n)$  the number of incidence of accidents between 0 and  $n$ , then the numeral value of  $\alpha$  for dependent variable i.e. the number vehicles crashed, causes over-(under) dispersion of the distribution of  $Y(n)$  as  $n \rightarrow \infty$ . If  $n$  is very large  $Y(n)$  is asymptotically normally distributed with

$$Y(n) \overset{asy}{\sim} normal\left(\frac{n}{\mu}, \frac{\sigma^2 n}{\mu^3}\right) \dots \dots \dots (10)$$

This indicates that  $Y(n)$  is asymptotically normally distributed as:

$$Y(n) \overset{asy}{\sim} normal\left(\frac{\lambda_i}{\alpha}, \frac{\lambda_i}{\alpha^2}\right) \dots \dots \dots (11)$$

The ratio of variance to mean of the limiting distribution is given by

$$\frac{Variance}{Mean} = \frac{\sigma^2 n \mu}{\mu^3 n} = \frac{\sigma^2}{\mu^2} = \left(\frac{\sigma}{\mu}\right)^2 \dots \dots \dots (12)$$

Thus, the variance mean ratio is greater (less) than 1 if and only if the coefficient of variation of the number of vehicles crashed  $v = \sigma/\mu$  is greater (less) than 1. If  $v < 1$  then the count distribution is under dispersed. However, if  $v > 1$  then the count distribution is over dispersed. The exponential distributions having a coefficient of variation,  $v=1$ , leads to equi- dispersion. The distribution simplifies to Poison distribution.

The limiting variance-mean ratio equals a constant  $1/\alpha$ . It follows that the gamma count distribution function (2) displays over dispersion for  $0 < \alpha < 1$  and under dispersion for  $\alpha > 1$ .

To obtain a gamma count regression model the assumption of a homogeneous population is relaxed by formulating a conditional model in which the parameters depend on a vector of individual covariates  $X_i$ . Assuming that the number of vehicles crashed is identical for all observations  $N$  may be set to unity with out loss of generality as long as an intercept is included. Assume that the mean for Gamma Model holds according to equation (1) i.e.

$$\mu_i = \frac{\alpha}{\beta_i} = \lambda_i = \exp(X_i\beta) \dots\dots\dots (13)$$

To make the estimated coefficients comparable to those obtained from the negative binomial or Poisson regression models, a simple strategy is to hold all explanatory variables constant at their means, and to compute  $\Delta \hat{Y} / \Delta x$  where  $x$  is the remaining explanatory variable, and the change is defined by a unit increase at the mean value (in the case of continuous variables), or by a change from 0 to 1 in the case of dummy variables. This measures literally the estimated effect of that explanatory variable on the dependent variable, holding other variables constant.

Given a sample of independent observations  $(y_1, \dots, y_n)$ , estimates,  $\hat{\alpha}$  and  $\hat{\beta}$  can be obtained by maximizing the log-likelihood of gamma function which is given by

$$\ell(y_i; x_i, \alpha, \beta) = \sum \ln \{ G(\alpha y_i, \alpha \exp(x_i' \beta)) - G(\alpha y_i + \alpha, \alpha \exp(x_i' \beta)) \} \dots\dots\dots (14)$$

Since, the function is non-linear in  $\alpha$  and  $\beta$ , an iterative solution algorithm is needed for calculating the indicated parameters. According to Winkelmann (1995), the algorithm described in Berndt, Hall and Hausman(1974) with numerical forward differenced gradients was used to solve for parameters iteratively.

According to King (1989), an event count regression model is specified by letting

the expected count,  $E(Y_i) = \lambda_i$  in a Poisson regression equations vary over kmj observations according to a specific function of a vector of explanatory variables. This may be written in general form as  $\lambda_i = \lambda_i(X_i, \beta)$ , where  $X_i$  is a vector of  $k$  exogenous variables;  $\lambda_i$  is as defined above and  $\beta$  is a  $k \times 1$  parameter vector.

The assumptions that the unobserved process generated the observed counts may lead to different variance in the event count regression. The variance of  $Y_i$  under the Poisson distribution is equal to its expected value:  $V(Y_i) = E(Y_i) = \lambda_i$ . However, this result at the level of the counts relies on micro level assumptions about the process generating the counts-independence and homogeneity. If these assumptions do not apply, then the variance is not equal to the mean and hence the Poisson distribution can not be applied. In this case, even though the log likelihood functions yield consistent estimates, the estimates will be inefficient, and the standard errors may be inconsistent. Therefore, using other appropriate model like Gamma and Negative Binomial models are important in cases where variance and mean of the considered random variables are not equal.

More generally, let

$$V(Y_i) = \lambda_i \phi \dots\dots\dots(15)$$

for  $\lambda_i > 0$  and  $\phi > 0$ ; where  $\phi$  is the dispersion parameter. When individual events are independent with homogeneous rates of occurrence, then  $\phi=1$  and  $V(Y_i) = \lambda_i$ , indicating the case of Poisson dispersion. Alternative assumptions may lead to other values for  $\phi$ . For example, if  $\phi > 1$  then the data are said to be over dispersed. However, if  $0 < \phi < 1$ , the data are considered to be under dispersed. In general,  $\phi$  is the inverse of the scale parameter  $v$ .

After re-parameterizing the binomial distribution along the lines of the Poisson and negative binomial distributions the following model could be developed:

$$f_b(y_i | \lambda_i, \sigma^2) = \frac{\Gamma\left(\frac{-\lambda_i}{\alpha^2 - 1} + 1\right)}{y_i! \Gamma\left(\frac{-\lambda_i}{\alpha^2 - 1} - y_i + 1\right)} (1 - \sigma^2)^{y_i} (\sigma^2)^{\frac{-\lambda_i}{\alpha^2 - 1} - y_i} \dots\dots\dots(16)$$

and where  $E(Y_i) = \lambda_i = \exp(X_i\beta) > 0$ ,  $V(Y_i) = \lambda_i \sigma^2$ , and  $0 < \sigma^2 < 1$ . This expression directly addresses the situation of under dispersion.

### 3.4. Parameter Estimation and test statistics

#### 3.4.1. Parameter Estimation

##### 3.4.1.1. The Dispersion Parameter

The estimation of dispersion parameter  $\phi$  varies based on whether the parameter is known or unknown (SAS 1999). In cases, where the dispersion parameter is not known, an estimate can be used to obtain an approximation to the scaled deviance and Pearson's chi-square statistic. One strategy is to fit a model that contains a sufficient number of parameters so that all systematic variations are removed. In this case estimate  $\phi$  from this model, and then use this estimate in computing the scaled deviance of sub-models. The deviance or Pearson's chi-square divided by its degrees of freedom is sometimes used as an estimate of the dispersion parameter  $\phi$ . For example, since the limiting chi-square distribution of the scaled deviance is  $D^* = \frac{D}{\phi}$  has  $n-p$  degrees of freedom, where  $n$  is the number of observations and  $p$  the number of parameters, equating  $D^*$  to its mean and solving for  $\phi$  yields  $\hat{\phi} = \frac{D}{n-p}$ . Similarly, an estimate of  $\phi$  based on Pearson's chi-square is  $\hat{\phi} = \frac{\chi^2}{n-p}$ . Alternatively, a maximum likelihood estimate of  $\phi$  can be computed by the procedure, or one can estimate the dispersion parameter from the scale parameter.

There are several options available for handling the exponential distribution dispersion parameter. One way in which the SAS procedure calculates the dispersion parameter of the model is by dividing the estimated deviance by its degrees of freedom or dividing the estimated Pearson's chi-square statistic by its degrees of freedom as shown above. If the estimate of dispersion after fitting divided by the degrees of freedom is not near 1, then the data is not equi-dispersed. In general the data is *over dispersed* if the dispersion estimate is greater than 1 and *under dispersed* if the dispersion estimate is less than 1. A simple way to model this situation is to allow the variance functions of these distributions to have a multiplicative over dispersion factor  $\phi$  i.e. Poisson:  $V(\mu)=\phi\mu$

Lord (2006), stated that although numerous estimators have been proposed in many literature, the three most common ones used by transportation safety modelers are given below. Some of these estimators are used by existing statistical software programs; while others have been proposed for re-calibrating Poisson-gamma models developed using crash data.

#### **A. Estimator 1**

The first estimator consists of computing the dispersion parameter using the method of moments. For Estimator 1, the analyst is required to use the output of the regression analysis. Once the value of the dispersion parameter is estimated, the analyst puts the new value into the regression model and performs a new regression analysis. This iteration is performed until all values converge (i.e., dispersion parameter, coefficients, etc.). This estimator usually converges after a single iteration. Estimator 1 has been tested and used extensively in various fields of research (Gourieroux et al., 1984a,b; Lawless, 1987; Hauer, 1997; McCullagh and Nelder, 1989; Cameron and Trivedi, 1998). FHWA (2003) proposed the idea of re-calibrating statistical models of motor vehicle crashes.

The estimator is given by the following equation

$$\hat{\alpha} = \frac{1}{n-p} \sum_{i=1}^n \frac{\{(y_i - \hat{\mu}_i)^2 - \hat{\mu}_i\}}{\hat{\mu}_i^2} \dots\dots\dots (17)$$

In Eq.(17), p refers to the number of parameters included in the model and n is the total number of observations. The confidence intervals for estimating the uncertainty associated with this estimator are usually not calculated due to the complexity of mathematics involved for computing the intervals. Although the delta method can be used for building confidence intervals, its derivation can be very cumbersome (Cameron, 2005). Therefore, it is suggested to use a bootstrapping method for estimating the confidence intervals for Estimator1 (Cameron and Trivedi, 1998; Cameron, 2005).

**B. Estimator 2**

The second estimator has been proposed by Cameron and Trivedi (1986). For Estimator 2, the dispersion parameter is estimated using a weighted regression analysis. The second alternative estimator is given by the following equation

$$\frac{(y_i - \hat{\mu}_i)^2 - y_i}{\hat{\mu}_i} = \alpha \hat{\mu}_i + \varepsilon \dots\dots\dots (18)$$

where  $\varepsilon$  the error term obtained by subtracting the predicted values by the model from the observed values.

In essence, the functional form of this estimator is very similar to Estimator 1, but the actual crash count is subtracted from the square of the difference between the observed and predicted values. This estimator uses the same iterative procedure as Estimator 1. According to Cameron and Trivedi (1986), this estimator provides a more rational way for estimating the dispersion parameter since the variance of

the left-hand of the equation is asymptotically distributed. Lord and Bonneson (2005) proposed this estimator for re-calibrating Poisson-gamma models developed using motor vehicle crashes. The confidence intervals can be computed directly from the output of the regression analysis.

### C. Estimator 3

The third estimator was originally proposed by Fisher (1941) and later improved by Lawless (1987). This estimator consists of estimating the dispersion parameter using the maximum likelihood method. The log-likelihood function of Poisson-gamma model is given by the following equation (Lord, 2006):

$$\ell(\alpha, \hat{\mu}_i) = \sum \left( \ln \left\{ \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(\alpha^{-1})} \right\} + y_i \ln \{ \hat{\mu}_i \} - (y_i + \alpha^{-1}) \ln \{ 1 + \alpha \hat{\mu}_i \} \right) \dots \dots \dots (19)$$

Considering the fact that (see Lord 2006)

$$\ln \left\{ \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(\alpha^{-1})} \right\} = \sum_{j=0}^{y_i-1} \ln \left\{ \frac{1 + \alpha j}{\alpha} \right\} = \sum_{j=0}^{y_i-1} \ln(1 + \alpha j) \dots \dots \dots (20)$$

and inserting equation (20) into (19) gives the following log-likelihood function

$$\ell(\alpha, \hat{\mu}_i) = \sum_{i=1}^n \left( \ln(1 + \alpha j) + y_i \ln \{ \hat{\mu}_i \} - (y_i + \alpha^{-1}) \ln \{ 1 + \alpha \hat{\mu}_i \} \right) \dots \dots \dots (21)$$

The Newton-Raphson (NR) scoring algorithm can be used to find the values of the log-likelihood function through the maximum likelihood method (Fletcher 1970; Walsh, 1975). According to Lawless (1987), this estimator is valid only if the parameters  $\beta$ s (the coefficients of the model) and  $\alpha$  (Dispersion parameter) are asymptotically independent and normally distributed. The dispersion parameter of the Poisson-gamma model is estimated using the approach proposed by

Lawless (1987). SAS (1999) also uses the same approach for estimating the dispersion parameter.

The NR algorithm can be used for building the confidence intervals associated with the estimator (Lawless, 1987; Payne, 2000). In fact, many statistical software programs now provide confidence intervals for the estimated dispersion parameter.

### 3.4.1.2. Estimation of Coefficient of Explanatory Variables

In order to estimate  $\beta$  the effect of the explanatory variables on the dependent variable the method of maximum likelihood is used (King 1989). By assuming the absence of autocorrelation, the likelihood function may be written as

$$L(\beta | y) = \prod_{i=1}^n f_p(y_i | \lambda_i) = \prod_{i=1}^n \frac{e^{-\lambda_i} (\lambda_i)^{y_i}}{y_i!} \dots\dots\dots (22)$$

where  $\lambda_i = \exp(X_i\beta)$ . The basic idea of maximum likelihood is that the data are taken as given, and we choose the value of  $p$  that maximizes this function and hence the likelihood of having generated the data. The log-likelihood function, reduced to sufficient statistics, is then given by

$$L(\beta | y) = \sum_{i=1}^n \{-\exp(X_i\beta) + y_i(X_i\beta)\}$$

Standard numerical maximization methods can easily be applied to this globally concave function by using one of many available computer programs.

Based on equations (21) and (22) and the assumption that the observations  $(y_i | x_i)$  are independent, it is considered that the most appropriate estimator is the maximum likelihood estimator (MLE). The log-likelihood function is therefore:-

$$\ln L(\beta) = \sum_{i=1}^n \{y_i X_i' \beta - \exp(X_i' \beta) - \ln y_i!\} \dots\dots\dots (23)$$

The Poisson MLE, denoted  $\hat{\beta}_p$ , is the solution to k linear equations corresponding to the first-order condition for maximum likelihood,

$$\sum_{i=1}^n (y_i - \exp(X_i' \beta)) X_i = 0 \dots\dots\dots (24)$$

If  $X_i$  includes a constant term then the residuals  $y_i - \exp(X_i' \beta)$  sum to zero (Equation 24). The Log-likelihood function is globally concave; hence solving these equations by Gauss-Newton or Newton-Raphson iterative algorithm yields unique parameters estimates.

By standard maximum likelihood theory of correctly specified models, the estimator  $\hat{\beta}_p$  is consistent for  $\beta$  and asymptotically normal with the sample covariance matrix:-

$$V(\hat{\beta}_p) = \left( \sum \mu_i X_i X_i' \right)^{-1} \dots\dots\dots (25)$$

where  $\mu_i$  is of the exponential form of (2).

### 3.4.2. Statistical Test of Parameters

A statistical test of under dispersion is highly desirable after running both Poisson and Gamma regression models. Most count models with under dispersion are specified according to equation (2) where  $\phi$  is an unknown parameter. It is assumed that under both null and alternative hypotheses the mean is correctly specified as, for example,  $\exp((X_i' \beta))$ , while under the null hypothesis  $\phi=1$  so that  $V[y_i | x_i] = \mu_i$ . A simple test statistic for  $H_0: \phi=1$  versus  $H_1: \phi \neq 1$  or  $H_1: \phi < 1$  can be

computed by estimating the Poisson model, constructing fitted values  $\hat{\mu}_i = \exp(X_i' \hat{\beta})$  and running the OLS regression with or without constant based on the data. From the reported test-statistic for  $\phi$ , if the null hypothesis is not rejected the Poisson Regression model will be employed; otherwise Gamma Regression will be used in analysis of the data.

Whether Poisson or Gamma regression will be used to analyze the data, can be induced from the effect of explanatory variables on the response variable. Generally, the significance of  $\beta$  shows the relevance of an explicative variable on computing the expected value of the response variables  $Y_i$ . It is well known fact that a given random variable whether it is normal or asymptotic normal, the hypothesis  $H_0: \beta=0$  is carried out by t-test for sufficiently large continuous data sets. According to Agristi (1996), however, there are three methods for performing significance test of the hypothesis  $H_0: \beta_i = 0$  versus  $H_1: \beta_i \neq 0$  about the parameters in GLMs. These are the Wald test, the Likelihood ratio test and the normality test for large sample sizes. The simplest uses the large-sample normality of ML estimates. The test statistic is given by  $z = \frac{\hat{\beta}}{ASE}$ , where ASE is asymptotic standard error and has an approximate standard normal distribution when  $\beta = 0$ . One refers Z distribution tables for the values of the test statistics.

The other test statistics is ratio test which is best for count data. The most appropriate test statistics for count data is the likelihood ratio test statistic. The likelihood ratio test statistic is given by  $-2\log(\ell_0 / \ell_1) = -2[\log \ell_0 - \log \ell_1] = -2(\ell_0 - \ell_1)$ ; where  $\ell_0$  denotes the maximized value for the simpler model representing the null hypothesis and  $\ell_1$  denotes the maximized value of the likelihood for the full model. Among these test statistics, we use the Likelihood ratio test for the analysis of data obtained from Addis.

According to the SAS output given in Annex 6 the final value of scale parameter is 13.2579. Hence, the inverse of this value is 0.075. The estimate of standard error of the dispersion parameter is 0.0062. Therefore, the calculated value of the t i.e.

$$t_{\text{cal}} = \frac{\hat{\phi} - \phi}{SE(\hat{\phi})} = \frac{0.075 - 1}{0.0062} = -149.19 \quad \text{with P value } 0.004. \quad \text{Therefore, } \phi \text{ is significantly}$$

different from 1 and hence we can use Gamma distribution.

By analyzing the data for both the Poisson model ( $\ell_0$ ) and Gamma model ( $\ell_1$ ) using SAS, the result of the maximum likelihood ratios are -5067.3 and -4662.43 respectively (Summary 1 of Annex 2 and Annex 3 respectively). Therefore, the value of likelihood ratio test statistic is  $-2(-5067.3 + 4662.43) = 809.76$ . This value is compared with  $\chi^2$  of degree of freedom of 1. This is significant with p-value  $< 0.0001$  at any level of  $\alpha$  significance. In addition to this, the calculated value of the dispersion parameter using Deviance and Pearson Chi-Square divided by degree of freedom are 0.0758 and 0.0712 respectively. These results show that the dispersion parameter is very far from one. This can lead to the assumption that the dispersion parameter is different from one.

Thus, we can conclude that, there is statistically significant under dispersion in the data that can lead to the rejection of the assumption of Poisson distribution. Consequently, the analysis of data in this study can be done by applying Gamma Regression Techniques.

### 3.5. Chi-Squared Tests of Independence

Consider categorical data that consist of frequency counts of observations occurring in the response categories. Let X and Y denote two categorical variables, X having i levels and Y having j levels. Let the ij possible combinations of outcomes are displayed in a two-way table. Furthermore, let  $\pi_{ij} = P(X=i, Y=j)$

denote the probability that (X,Y) falls in the cell in row i and column j. In addition, let the probabilities  $\{\pi_{ij}\}$  form the joint distribution of X and Y, and the marginal distributions are the row and column totals of the joint probabilities and they are denoted by  $\{\pi_{i+}\}$  and  $\{\pi_{+j}\}$  for the row and column variable respectively.

When two variables are independent, the probability of any particular column response j is the same in each row. In two-way contingency tables the null hypothesis of statistical independence of two responses has the form (Agresti,1996).

Ho:  $\pi_{ij}=\pi_{i+}\pi_{+j}$ , for all i and j.

To test the Ho, we identify  $\mu_{ij}=n\pi_{ij}=n\pi_{i+}\pi_{+j}$  as the expected frequency. Here  $\mu_{ij}$  is the expected value of  $n_{ij}$ (sample cell counts) assuming independence. And n is the sample size. Since  $\{\pi_{i+}\}$  and  $\{\pi_{+j}\}$  are unknown, they are estimated by substituting sample proportions for the unknown probabilities, such that

$$\hat{\mu}_{ij} = np_{i+}p_{+j} = n \frac{n_{i+}}{n} \frac{n_{+j}}{n} = \frac{n_{i+}n_{+j}}{n}$$

The  $\{\hat{\mu}_{ij}\}$  are called expected frequencies.

For testing independence in I x J contingency tables, the Pearson and likelihood ratio statistics are given by

$$\chi^2 = \sum \sum \frac{(n_{ij} - \hat{\mu}_{ij})^2}{\hat{\mu}_{ij}} \text{ and } G^2 = 2 \sum \sum n_{ij} \log \left( \frac{n_{ij}}{\hat{\mu}_{ij}} \right) \text{ respectively.}$$

Their large sample chi-squared distributions have  $df=(I-1)(J-1)$ . This means, under Ho,  $\{\pi_{+j}\}$  and  $\{\pi_{i+}\}$  determine the cell probabilities.

Thus, chi-squared test, type 1 and type3 test statistics are used to identify and select explanatory variables which have statistically significant association with

the dependent variable. To identify whether there is a significant association between explanatory variables and the response variable stated in chapter 3, the specified chi-squared statistics was employed.

## Chapter Four: Result and Discussion

### 4.1. The Goodness of fit

The 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. For a fixed value of the dispersion parameter, the scaled deviance is twice the difference between the maximum achievable log likelihood and the log likelihood at the maximum likelihood estimates of the regression parameters. But Pearson's chi-square statistic is defined as the sum of the products of the weight of the variable and the square of the difference between actual and predicted values divided by the variance of the predicted value.

The scaled version of both 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 an approximate guide to the goodness of fit of a given model (SAS 1999). However, Agresti, (1996) states that for many GLMs, the deviance has approximately a chi-squared distribution. If the Chi-Square statistics shows that the deviance is significant at the specified  $\alpha$  level of significance then the specified model explains the variables well. According to SAS output given in Annex 3 Summary 1, the scaled deviance for the Gamma model is 7062.2495 with 6865 degrees of freedom. The p-values for chi-square at 6865 df is 0.046. This shows that the specified model significantly explains the variables in the model. Therefore, Gamma Model is the model we use in the analysis of the number of vehicles crashed as described in chapter three.

To check whether there is a departure from the ideal plot of gamma distributed random variable or not, the summary statistics for the number of vehicles

crashed is plotted (Figure1). The actual plot is similar to the ideal plot. This also supports the application of the proposed model for analyzing accident data.

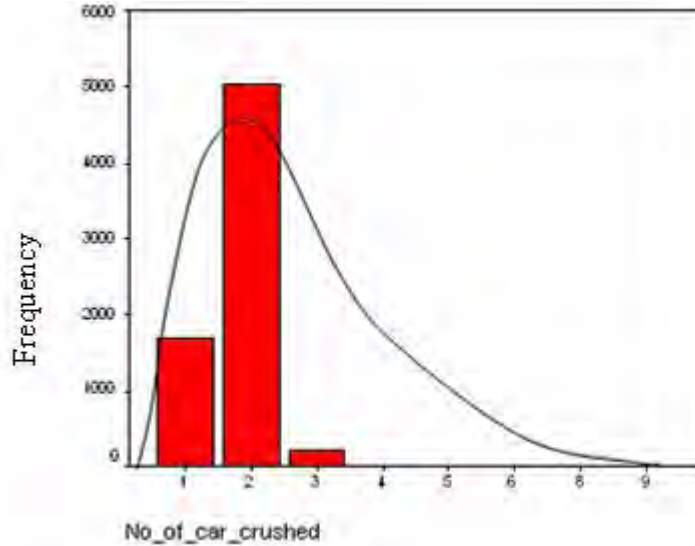


Figure 1:-Plots of Number of Cars crashed

According to Agresti (1996) components of the deviance, called deviance residuals, provide diagnostic measures of goodness of fit for individual observations. They are alternatives to Pearson and adjusted residuals. Like the Pearson residuals, standardized deviance residuals are approximately normally distributed with minor deviation from the standard normal variates. Whether this hypothesis is satisfied or not the standardized deviance residual for the normal plot is given in Annex1; Figure 2. The plots indicate that the normality assumption is valid.

## 4.2. Variable Selection

In order to establish the required relationships between the numbers of vehicles crashed (dependent variable); per incidence of road accidents and different explanatory variables, we have to select explanatory variables that have significant association with the dependent variable. For this purpose we have

employed SAS Software. As selection criteria, Type1, Type3 and P-values of the chi-square test statistics were used. Primarily there are 22 variables with a total of 132 levels. In this variable selection, the analysis was started by fitting a sequence of models, beginning with a simple model with only an intercept term, and continuing through a model of specified complexity, fitting one additional effect on each step. Likelihood ratio statistics, that is, twice the difference of the log likelihoods, are computed between successive models. The asymptotic distribution of the likelihood ratio statistics, under the hypothesis that the additional parameters included in the model are equal to 0, is a chi-square with degrees of freedom equal to the difference in the number of parameters estimated in the successive models. Thus, these statistics can be used in a test of hypothesis of the significance of each additional variables included in the model. This clearly indicates that forward regression analysis was used.

As per Type1 and type 3 test statistics of the SAS output 15 variables namely: date of the week, drivers' sex, drivers' age, drivers' educational background, drivers' experience, license grade, sub-city, place of accident, ownership, road classification, type of road, weather condition, registration type, road direction and road condition are not significantly associated with response variable at 10%  $\alpha$  level of significance.

For the second time the model is re-fitted by using the remaining 7 significant variables after excluding all those non-significant variables. The output of the model at this step shows that all the 7 variables have significant association with response variable by both type1 and 3 test statistics. However, the p-value of the coefficients of all levels of the two variables, namely; driver relation with vehicle and status of vehicle movements, are not significantly different from zero at 10%  $\alpha$  level of significance. Therefore, we have excluded them and left with

only five variables with a total of 36 levels that have significant association with the response variable. The detail SAS output is given in Annex 3, 4 and 5.

The result shows that all variables under demographic factors are insignificant.

Among the site factors only light condition is the variable that has strong association with the number of vehicles crashed that resulted in property damage. The rest site factors have no significant association with the response variable (Number of vehicles crashed) even at 10% $\alpha$  level of significance.

Among the road factors only road junction has strong associated with the response variable. The remaining road factors are insignificant. This means road type, road classification and road condition have no significant effect on the outcome of the traffic accidents that resulted in property damage.

Among Vehicle related factors all vehicle related variables have strong associated except ownership state of the vehicle.

Therefore, analysis and discussion is based on the selected five variables that have significant relationship with the response variable for both descriptive statistics and gamma regression model.

### **4.3. Descriptive Statistics**

Traffic accidents that involve more than one vehicle have higher cost of damage than single vehicle crashes. Therefore, analyzing crashes between single vehicle and multiple crashes that resulted in property damage is important. Hence, the dependent variable was converted to a categorical variable in order to compare the impact of each variable on single vehicle crash (crashes involving only one vehicle) and multiple vehicle crashes (crashes that involve more than one vehicle) that resulted in property damage only.

Table1: Cars crashed per incidence of Accidents resulted in property damage

Crash Category	Frequency	Percent
Multiple Vehicle	5296	75.9%
Single Vehicle	1680	24.1%
Total	6976	100.0%

The number of vehicles crashed per accident that resulted in property damage ranges from 1 to 9. According to table 1 given above, 75.9% multiple vehicle crashes, while the remaining 24.1% is single vehicle crashes.

In order to compare the share of each variable, descriptive statistics is given on the total number of vehicles crashed, single vehicle crash and multiple vehicle crashes that resulted in property damage only. The mean and the variance of the number of vehicles crashed with respect to the levels of each variable were also calculated. The three indicators are:

- Number of Accidents (by each level of a variable)
- Number of Cars Crashed (by each level of a variable)
- Average Number of Cars Crashed per incidence of accidents (by each level of a variable).

### 4.3.1. Vehicle Related Factors

#### A. Vehicle Service year

Table 2: Accidents by vehicle service year

Categories	Multiple Vehicle				Single Vehicle		Overall			
	Number of Cars Crashed	% of cars Crashed	Number of accidents	Cars crashed /accidents	Number of Cars Crashed	% of cars Crashed	Total No of Cars crashed	% of cars Crashed	Cars crashed /accidents (Mean)	Variance
Less than 1 year	575	5.3	280		115	6.8	690	5.5		0.266
1-2 years	226	2.1	109		54	3.2	280	2.2		0.303
2-5 years	1805	16.6	874		282	16.8	2087	16.6		0.308
5-10 years	4909	45.1	2403		627	37.3	5536	44.1		0.218
Above 10 years	3375	31	1630		602	35.8	3977	31.6		0.310
<b>Total</b>	<b>10890</b>	<b>100</b>	<b>5296</b>	<b>2.1</b>	<b>1680</b>	<b>100</b>	<b>12570</b>	<b>100</b>	<b>1.8</b>	<b>0.268</b>

Among the five categories of service years of the vehicle, vehicles with service above 5 years have 75.7% share of the total number of vehicles crashed that resulted in property damage only. The average number of vehicles crashed per incidence in multiple crashes is 2.1 while the overall figure is 1.8. In addition to this, while vehicles with service above 5 years have 76.1% and 73.1% of the total multiple and single vehicle crashes respectively. It is to be noted that 88% of the total vehicles in the country have already served for 5 years or above.

## B. Vehicle type

Table 3:- Accident by vehicle type

Categories	Multiple Vehicle				Single Vehicle		Overall			
	Number of Cars Crashed	% of cars Crashed	Number of accidents	Cars crashed /accidents	Number of Cars Crashed	% of cars Crashed	Total No of Cars crashed	% of cars Crashed	Cars crashed /accidents (Mean)	Variance
41-100 Quintals	467	4.3	228		143	8.5	610	4.9		0.295
Automobile	3445	31.6	1682		485	28.9	3930	31.3		0.236
Bicycle	2	0	1		2	0.1	4	0		0.000
Liquid Freighters	27	0.2	13		8	0.5	35	0.3		0.359
Freight 11-40 Quintals	718	6.6	345		145	8.6	863	6.9		0.305
Freight with trailer	63	0.6	29		31	1.8	94	0.7		0.419
Motor Cycle	44	0.4	21		9	0.5	53	0.4		0.268
Pickup up to 10 Quintals	1154	10.6	567		147	8.8	1301	10.4		0.211
Public 12 seats	371	3.4	180		50	3	421	3.3		0.303
Public 13-45 seats	435	4	210		52	3.1	487	3.9		0.283
Public above 46 seats	543	5	263		117	7	660	5.3		0.374
Special Vehicles	23	0.2	8		4	0.2	27	0.2		4.750
Station Wagon	1429	13.1	695		237	14.1	1666	13.3		0.258
Taxi	2169	19.9	1054		250	14.9	2419	19.2		0.233
<b>Total</b>	<b>10890</b>	<b>100</b>	<b>5296</b>	<b>2.1</b>	<b>1680</b>	<b>100</b>	<b>12570</b>	<b>100</b>	<b>1.8</b>	<b>0.268</b>

This variable has thirteen categories. Among the thirteen categories, Automobiles Taxies and Station Wagon were involved in 31.3%, 19.2% and 13.3% (a total of 63.8%) of the accidents respectively.

While Automobiles, Taxies and Station wagon were involved in 31.6% 19.9% and 13.1% of the multiple crashes. The figures for single vehicle crash are 28.9%, 14.9% and 14.1% respectively. It is to be noted that 38.4%, 7.2 % and 6.9% of the total vehicles in the country are Automobile, Taxi and station Wagons respectively.

### C. Type of Collision

Table 4:- Accidents by type of collision

Categories	Multiple Vehicle				Single Vehicle		Overall			
	Number of Cars Crashed	% of vehicles of Crashed	Number of accidents	Cars crashed /accidents	Number of Cars Crashed	% of vehicles Crashed	Total No of Cars crashed	% of vehicles Crashed	Cars crashed /accidents (Mean)	Variance
Collision with Pedestrian	12	0.1	6		2	0.1	14	0.1		0.214
Collision with non living Objects	63	0.6	28		371	22.1	434	3.5		0.191
Collision with parked Vehicle	115	1.1	54		22	1.3	137	1.1		0.427
Head to Head collision	301	2.8	147		13	0.8	314	2.5		0.124
Head to side Collision	3607	33.1	1768		422	25.1	4029	32.1		0.212
Head to tail Collision	4775	43.8	2297		502	29.9	5277	42.0		0.264
Overtuning		0.0			39	2.3	39	0.3		0.000
Side to side Collision	2017	18.5	996		309	18.4	2326	18.5		0.213
<b>Total</b>	<b>10890</b>	<b>100.0</b>	<b>5296</b>	<b>2.1</b>	<b>1680</b>	<b>100.0</b>	<b>12570</b>	<b>100.0</b>	<b>1.8</b>	<b>0.268</b>

To take corrective measures and reduce the risk of vehicle crashes, identifying the type of collision is also important. Based on the observed data the types of collision can be categorized into eight. Among these categories; Head-to-Tail and Head-to-Side collisions have 74.1% share of the total accidents resulting in property damage only.

Out of the total multiple crashes 43.8% are Head-to-Tail collision while 33.1% are Head-to-Side collision. However, the share these categories for single vehicle crashes are 29.9% and 25.1% respectively. This shows that the shares of these categories for single vehicle crash are less than the case for multiple vehicle crashes. Moreover, the share of vehicles crashed due to collision with non living objects in case of single crash is 22.1%. This category is ranked third among the categories that aggravate single vehicle crashes.

### 4.3.2. Site factors

#### Light Condition

Table 5:- Accident by light condition

Categories	Multiple Vehicle				Single Vehicle		Overall			
	Number of Cars Crashed	% of vehicles Crashed	Number of accidents	Cars crashed/accidents	Number of Cars Crashed	% of vehicles Crashed	Total No of Cars crashed	% of vehicles Crashed	Cars crashed/accidents (Mean)	Variance
Dark with poor road light	8	0.1	4		6	0.4	16	0.1		0.242
Dark with road light	1018	9.3	494		228	13.6	1246	9.9		0.310
Day Light	9722	89.3	4728		1417	84.3	11139	88.6		0.262
During Sun rise	57	0.5	28		12	0.7	69	0.5		0.256
During Sun set	85	0.8	42		15	0.9	100	0.8		0.224
<b>Total</b>	<b>10890</b>	<b>100.0</b>	<b>5296</b>	<b>2.1</b>	<b>1680</b>	<b>100.0</b>	<b>12570</b>	<b>100.0</b>	<b>1.8</b>	<b>0.268</b>

It was observed that 88.6% of the total accidents that resulted in property damage took place during day time. The figure for multiple and single crashes are 89.3% and 84.3% respectively.

### 4.3.3. Road Factors

#### Road Junction

Table 6:- Accident by road junction

Categories	Multiple Vehicle				Single Vehicle		Overall			
	Number of Cars Crashed	% of vehicles Crashed	Number of accidents	Cars crashed/accidents	Number of Cars Crashed	% of vehicles Crashed	Total No of Cars crashed	% of vehicles Crashed	Cars crashed/accidents (Mean)	Variance
+ - Shape	994	9.1	486		117	7.0	1111	8.8		0.223
Circle island	532	4.9	262		70	4.2	602	4.8		0.201
No junction	7060	64.8	3424		1188	70.7	8248	65.6		0.288
T-Shape	2243	20.6	1094		283	16.8	2526	20.1		0.231
Y-Shape	61	0.6	30		22	1.3	83	0.7		0.285
<b>Total</b>	<b>10890</b>	<b>100.0</b>	<b>5296</b>	<b>2.1</b>	<b>1680</b>	<b>100.0</b>	<b>12570</b>	<b>100.0</b>	<b>1.8</b>	<b>0.268</b>

Results in this table 6 shows that 65.6% of the total crashes took place on roads with no junctions.

## 4.4. Gamma Regression

### 4.4.1. Saturated versus Reduced Models

As mentioned in section 3.3.1 of Chapter 3, Gamma Count Regression is the selected method for analyzing the accident data. A saturated model that includes all possible indicator variables (22 variables with a total of 132 indicator variables) is constructed. A reduced model that involves variables that are significant at 10%  $\alpha$  level of significance among the 22 variables used in the saturated model is then constructed. Accordingly, only 5 variables with a total of 36 levels are considered in the final model.

To check whether there is no difference between the final reduced model and saturated model, the likelihood ratio test of the deviance of the two models are used. Let  $L_S$  denotes the maximized log-likelihood values for the most complex model which is called saturated and let  $L_M$  denotes the maximized log-likelihood value for the reduced model of interest. The deviance of a model is defined as  $\text{Deviance} = -2[L_M - L_S]$  and is used as a test statistics for testing the hypothesis that all parameters that are in the saturated model but not in the reduced equal to zero. For two models, suppose that  $M_0$  is the reduced model which is the special case of the complex model i.e.  $M_1$ . Given that the more complex model holds, the likelihood ratio statistics for testing that the simpler model holds is (Agresti 1996):

$$-2[L_0 - L_1] = -2[L_0 - L_S] - \{-2[L_1 - L_S]\} = \text{Deviance}_0 - \text{Deviance}_1.$$

One can compare the models by using their deviances. For large samples, the deviance has approximately a chi-squared distribution. Therefore, the difference between the deviances of the two models is tested by chi-square test. The degree of freedom for the chi-square test of the differences of the deviances is equal to

the difference between the numbers of parameters of saturated model and reduced model.

As shown in Annex 3 and Annex 5, the deviance of the saturated model is 524.16 while the deviance of the reduced model is 532.79 respectively. The saturated model has a total of 133 parameters including the intercept. The reduced model with fewer variables has a total of 37 parameters including the intercept of the model. The difference between the two parameters of the respective model, which is considered as the df, is therefore  $133-37=96$ . Moreover, the difference between the deviances of the two models is  $532.87-524.16=8.71$ . This is not significant since the  $\chi^2(96)$  is 124.3 at 5% level of significance with p-value = 1.000.

Thus, it can be concluded that there is no significant advantage that can be gained by considering the more complex equation i.e. saturated model. Therefore, reduced model is selected as the better regression equation for analyzing the data.

#### **4.4.2. Interpretation of coefficients of Gamma Regression**

The Gamma Distribution has a positive mean. Though one can model the Gamma mean in GLMs using the identity link, it is more common to model the exponents,  $\text{Exp}$ , of the mean. Like the linear predictor, the exponent of the mean can take any real value. Hence, the log mean is the natural parameter for Gamma distribution, and the log link is the canonical link for a GLM with the random components.

As described in previous chapter, let  $\mu=\lambda=\exp(\mathbf{X}\boldsymbol{\beta})$  denotes the expected value for a regression variate  $Y$ , and let  $X$  denotes the matrix of explanatory variables, then the natural logarithm of the expected gamma model has a form of  $\mathbf{X}\boldsymbol{\beta}$ . From this model, the mean satisfies the exponential relationship:

$$\mu = \lambda = \exp(\mathbf{X}\boldsymbol{\beta}) = e^{X\boldsymbol{\beta}} = e^{\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k} = e^{\sum_{i=1}^{36} \beta_i X_i} = \prod_{i=1}^{36} e^{\beta_i X_i} .$$

The expression shows that if  $\beta_i=0$ , then  $e^{\beta_i} = e^0 = 1$  and the multiplicative factor is 1; that is, the indicator variable  $i$  has no any impact on the mean of the response variable. However, if  $\beta_i > 0$ , then  $e^{\beta_i} > 1$ , and the indicator variable  $i$  contribute to the mean of the response variable at a rate greater than one. On the contrary; if  $\beta_i < 0$ , then  $e^{\beta_i} < 1$  and the rate at which the mean of the response variable changes due to that indicator variable lies between 0 and 1. Coefficient Estimation, Standard Error and Test Statistics of each category of indicator variables using Gamma Regression Model of selected variables is given in Annex 5.

#### **4.4.2.1. Vehicle related factors**

##### **A. Vehicle Service year**

Among the five categories of this variable only the coefficient of the variables representing vehicles with services 2-5 years is not significantly different from zero at 10%  $\alpha$  level of significance, but the estimates of others significant at 10%  $\alpha$  level of significance with positive coefficients. This indicates that all categories of this variable contribute values at a rate greater than one.

##### **B. Vehicle type**

The coefficients of all categories of types of vehicles are significantly different from the zero. However, the sign of the coefficients are negative. This means all categories of type of vehicle contribute value at a rate between 0 and 1.

##### **C. Type of Collision**

All categories of types of collision expect collision with non living objects, are significantly different from 0. Regarding the sign of the significant coefficients, all of them are positive. This implies that the significant categories contribute values at a rate greater than one.

#### **4.4.2.2. Site factors**

Among factors of site, only light condition is the remaining variable in the reduced model. As shown in Annex 5 all categories of this variable are significant. The sign of these categories are positive indicating that the categories contribute values at a rate greater than one.

#### **4.4.2.3. Road factors**

Among the road factors only road junction is considered in the reduced model. The coefficients of all five levels of this variable are significant at 5% level of significance with positive sign. Hence, we can say that all levels of road junction contribute a value at a rate greater than one.

### **4.4.3. Prediction Using Estimated Parameters**

#### **4.4.3.1. Mean of the Model**

The mean,  $\mu_i$ , of the response in the  $i^{\text{th}}$  observation is related to a linear predictor through exponential function, Exp. In model specification we have stated that the number of vehicles crashed per incidence of accidents has a Gamma probability distribution and the mean,  $\mu_i$ , is related to light condition(LC), type of vehicle(VT), vehicle service year (VSY), type of collision (CT) and road junction(RJ) as indicated in the reduced model. Therefore, the mean can be derived from the following equations:-

$$\ln(\bar{Y}) = \alpha + \sum_{j=1}^4 b_{1j} LC_j + \sum_{k=1}^{12} b_{2k} VT_k + \sum_{l=1}^4 b_{3l} VSY_l + \sum_{m=1}^7 b_{4m} CT_m + \sum_{h=1}^4 b_{5h} RJ_h$$

Where

$\bar{Y}$  = estimated mean number of vehicles crashed per incidence of accident that in property damage

$\alpha$  = the intercept of the model

$LC_j$  = Light Condition of level  $j$

$VT_k$  = Type of Vehicle of level  $k$

$VSY_l$  = Vehicle Service Year of level  $l$

$CT_m$  = Type of Collision of level  $m$

$RJ_h$  = Road Junction of level  $h$

$LC_j = 1$  for Light Condition of level  $j$  and 0 for other levels  $j=1,2,3,4$

$VT_k = 1$  for Type of Vehicle of level  $k$  and for other level  $s=k=1,2,\dots,12$

$VSY_l = 1$  for Vehicle Service Year of level  $l$  and for other levels  $l=1,2,\dots,4$

$CT_m = 1$  for Type of Collision of level  $m$  and 0 for other levels  $m=1,2,\dots,7$

$RJ_h = 1$  for Road Junction of level  $h$  and 0 for other levels  $h=1,2,\dots,4$

Using the estimated coefficients obtained (Annex 5) the above equation can be expressed as

$$\log(\bar{Y}) = -0.1119 + 0.2031LC_1 + \dots + 0.1941LC_4 + -0.1980VT_1 + \dots + -0.2777VT_{12} + 0.0311 VSY_1 + \dots + 0.0491VSY_4 + 0.6534CT_1 + \dots + 0.5511CT_7 + 0.0969RJ_1 + \dots + 0.1019RJ_4$$

Using this formula we can estimate the mean of number of vehicles crashed per incidence of accident by taking different level of the categories of those significant indicator variables. For instance if we observe a combination of LC2, VT9, VSY3, CT6 and RJ1 then the mean of the number of vehicles crashed is estimated to be 0.6134. Therefore, the exponent of 0.6134 i.e.  $\exp(0.6134) = 1.847$ .

This is almost equal to the mean of the response variable of the population given in Annex 1.

#### 4.4.3.2. Variance of the Model

Probability distributions of the response  $Y$  in generalized linear models are usually parameterized in terms of the mean,  $\mu$ , and dispersion parameter  $\phi$ . The probability distributions that are available in the GENMOD procedure for Gamma distribution, scale parameter and the variance of  $Y$  are also given below (SAS 1999).

From Annex 5 the scale parameter  $\nu = 13.2559$ . Therefore, the variance for the incidence of accident given above is  $V(Y) = \frac{\mu^2}{\nu} = \frac{1.847^2}{13.2559} = 0.257$  and the dispersion parameter  $\phi = \nu^{-1} = 13.2559^{-1} = 0.0754$ . This shows that the variance of the incidence per event is approximately equal to the variance of the population. Regarding the dispersion parameter the result obtained indicates that the value is very far from 1. Hence, using Gamma Distribution for this data analysis is appropriate.

## **Chapter Five: Conclusion and Recommendation**

### **5.1. Conclusion**

The gamma regression model reveals that the coefficients of vehicle related variables can be considered as major factors for vehicle crashes. Positive parameter estimates of the coefficients of the indicator variables imply that the rate of contribution of the variables greater than one while parameters with negative sign indicates that the rate of contribution of the variables lie between 0 and 1. Accordingly almost all variables under the categories vehicle service year and type of collision contribute values at a rate greater than one.

Further more, the results show that the number of vehicles crashed per incidence of accident that resulted in property damage is mainly determined by five variables namely: type of vehicles, vehicle service years, type of collision, light condition and road junction among the 22 variables considered in the model. Therefore, actions to be taken in view of reducing the number of vehicle crashed per incidence of accident should mainly focus on these five variables.

Most of accidents took place on straight road sections than on junctions. In addition, high number of vehicle crashes occurred during day light, the time when there is a high traffic movement.

The overall mean and variance of the number of vehicles crashed per incidence of accident are 1.8 and 0.268 respectively.

Vehicles with service years greater than 5 were involved in 75.7% of the total number of vehicles crashed. Among the categories of vehicle type, Automobiles, Taxes and Station Wagons were involved in 31.3%, 19.2% and 13.3% of the accidents.

Regarding the type of collision 43.8% are Head-to- collision while 33.1% are Head-to-Side collisions.

## **5.2. Recommendations**

Despite having a very low road network density and relatively small number of vehicle, Addis Ababa has a relatively high accident records which in turn showed a negative impact on the national economy. Due to this fact there is a need for putting in place sustainable solutions without delay in addition to the efforts being made to reduce the frequency of traffic accidents in general. It is believed that special attention should be given to the following points:

- Strict control and management of vehicle movement on the roads that have no junctions in the City especially in high traffic volume
- Road engineering solution is required in order to minimize the risk of vehicles on different road junction.
- The drivers and executive agency should focus on ways of improving movements of vehicles during day time through e.g. patrol traffic police control.
- Creation of intensive awareness for road user.
- Further studies can be made on the area of traffic accidents by considering detail and accurate information on various variables. For example if the causes and consequences of an accident are recorded in detail instead of broad categories results could be more accurate and efficient.

## I. REFERENCES

- Abramowitz, M., and Stegun, I.A (1964) Handbook of Mathematical Function, National Bureau of Standards, Applied Mathematics Series Nr. 55, Washington D.C.
- Agresti, A., 1996. An Introduction to Categorical Data Analysis. Wiley, New York
- Akgungor, A. P. and Yildiz, O. (2007): Sensitivity analysis of an accident prediction model by the fractional factorial method: *Accident Analysis and Prevention* 39, 63-68.
- Berhanu, G. (2004): Models relating traffic safety with road environment and traffic flows on arterial roads in Addis Ababa: *Accident Analysis and Prevention*, 36, 697–704
- Berhanu, G., 2000. Effects of Road and Traffic Factors on Road Safety in Ethiopia. Dr.-Ing. Dissertation, Norwegian University of Science and Technology, Trondheim, Norway
- Berndt, E., Hall, R., and Hausman, J. (1974) “ Estimation and Inference in non-linear structural models,” *Annual of Economic and Social Measurement*, 3/ 4 653-666
- Bowman, K.O., and Shenton, L.R. (1988), Properties of Estimators for the Gamma Distribution, New York: Marcel Dekker.
- Cameron, A.C., 2005. Personal communication. College Station, TX
- Cameron, A.C., Trivedi, P.K., 1986: Econometric models based on count data: Comparisons and Applications of Some Estimators and Tests. *Journal of Applied Econometric*. 1, 29–53.
- Cameron, A.C., Trivedi, P.K., 1998. Regression Analysis of Count Data. Cambridge University Press, Cambridge, UK.
- Chandraratna,S. Stamatiadis, N. and Stromberg, A (2006): Crash involvement of drivers with multiple crashes: *Accident Analysis and Prevention* 38, 532-541.
- Ethiopian Road Authority (ERA) 2005: Urban Transport Study and Preparation of Pilot project for Addis Ababa, 1, 1-12. (Unpublished)
- Federal Democratic Republic of Ethiopia Central Statistics Agency (CSA), (2005): *Statistical Abstract*

- Federal Highway Administration, 2003. Module 1 – Network Screening. SafetyAnalyst: Software Tools for Safety Management of Specific Highway Sites. U.S. Department of Transportation, Washington, DC (web page: <http://www.safetyanalyst.org/whitepapers/module1.pdf> , accessed on June 10, 2005).
- Fisher, R.A., 1941. The Negative Binomial Distribution. *Ann. Eugenics* 11, 182–187.
- Fletcher, R., 1970. A new approach to variable metric algorithms. *Comput. J.* 13, 317–322.
- Gourieroux, C., Monfort, A., Trognon, A., 1984a. Pseudo Maximum Likelihood methods: theory. *Econometrika* 52, 681–700.
- Gourieroux, C., Monfort, A., Trognon, A., 1984b. Pseudo maximum likelihood methods: application. *Econometrika* 52, 701–720.
- Hadi, M.A., Aruldas, J., Chow, L.-., Wattleworth, J.A., 1995. Estimating safety effects of cross-section design for various highway types using negative binomial regression. *Transport. Res. Rec.* 1500, 169–177
- Hauer, E., 1997. *Observational Before–After Studies in Road Safety: Estimating the Effect of Highway and Traffic Engineering Measures on Road Safety*. Elsevier Science Ltd., Oxford.
- Hauer, E., 1997. *Observational Before–After Studies in Road Safety: Estimating the Effect of Highway and Traffic Engineering Measures on Road Safety*. Elsevier Science Ltd., Oxford.
- Jorgensen, N.O. (1996): *Traffic Safety*. In Bent Thagenen (ed.), *Highway and Traffic Engineering in Developing countries*. E. & FN Spon, UK
- King, G., (1989): Variance Specification in Event Count Models: From Restrictive Assumptions to a Generalized Estimator: *American Journal of Political Science*, 33, (3) 762-84
- Lauren P. et al (2005): *Road Traffic Injuries- Can We Stop A Global Epidemic?* Retrieved December 8, 2006 from the World Wide Web [http://www.thedoctorwillseeyounow.com/articles/other/road\\_33/](http://www.thedoctorwillseeyounow.com/articles/other/road_33/).

- Lawless, J.F., 1987. Negative binomial and mixed Poisson regression: *Canada Journal of Statistics*. 15 (3), 209–225.
- Lord, D., Miranda-Moreno, L.F., 2006. Effects of Low Sample Mean Values and Small Sample Size on the Estimation of the Dispersion Parameter of Modeling motor vehicle crashes using Poisson-gamma models: Examining the effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter: *Accident Analysis and Prevention* ,38, 751–766
- Lord, D. (2006): Modeling motor vehicle crashes using Poisson-gamma models: Examining the effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter: *Accident Analysis and Prevention*, 38, 751–766
- Lord, D., Washington, S. and Ivan, J.N (2007): Further notes on the application of zero-inflated models in highway safety: *Accident Analysis and Prevention* 39, 53–57
- Lord, D., Washington, S.P., Ivan, J.N., 2005b. Poisson, Poisson-gamma and zero inflated regression models of motor vehicle crashes: balancing statistical fit and theory. *Accident Analysis and Prevention*. 37 (1), 35–46.
- Lu, M. (2006) :Modeling the effects of road traffic safety measures: *Accident Analysis and Prevention* 38, 507–517
- Ma, J., Kockelman, K.M., 2006. Bayesian multivariate Poisson regression for models of injury count by severity. In: *Proceedings of the Transportation Research Board 85th Annual Meeting*, Washington, DC (CD-ROM).
- McCullagh, P. and Nelder, J.A (1989): *Generalized Linear Models*, Second Edition, Chapman and Hall, Chicago.
- Mekonnon, T. (2007): Empirical analysis on traffic accidents Involving human injuries: the case of Addis Ababa (Unpublished)
- Miaou, S. P., 1996. Measuring the Goodness-of-Fit of Accident Prediction Models. Federal Highway Administration FHWA-RD-96-040. U.S. Department of Research Council, Washington, D.C., USA

- Miaou, S.-P., Song, J.J., 2005. Bayesian ranking of sites for engineering safety improvements: Decision parameter, treatability concept, statistical criterion and spatial dependence. *Accident Analysis and Prevention*. 37 (4), 699–720.
- Oh, J., Simon P. Washington, S.P., and Nam, D. (2006): Accident prediction Model for railway-highway interference. *Accident Analysis and Prevention*, 38, 346-356.
- Park, E.S., Lord, D., 2006. Multivariate Poisson-lognormal models for jointly modeling crash frequency by severity. In: Paper presented at the 86th Annual Meeting of the Transportation Research Board, Washington, DC.
- Payne, R.W. (Ed.), 2000. The Guide to Genstat. Lawes Agricultural Trust, Rothamsted Experimental Station, Oxford, UK
- SAS 1999 A User Guide Manual by SAS Institute Inc., Cary, NC, USA
- Song, J.J., Gosh, M., Miaou, S.-P., Malik, B., 2006. Bayesian multivariate spatial models for roadway traffic crash mapping. *Journal of Multivariate Analysis*. 97 (1), 246–273.
- Tesema, T. et al (2005): Rule Mining And Classification Of Road Traffic Accidents Using Adaptive Regression Trees. *International Journal of Simulation Systems. Science & Technology Special Issue on Soft Computing for Modeling and Simulation*, 6, 10-11.
- Transport Research Laboratory (2001): Study Report for a sectoral Road safety program in Ethiopia, 2, 1-30.
- Walsh, G.R., 1975. Methods of Optimization. Wiley Publishing Co., London, UK.
- Wang, J., Hughes, W.E., Stewart, R., 1998. Safety Effects of Cross-section Design on Rural Multi-lane Highways. Federal Highway Administration, FHWA-RD-98-071. U.S. Department of Transportation, USA.
- Warton, D.I., 2005. Many zeros does not mean zero inflation: comparing the goodness-of-fit of parametric models to multivariate abundance data. *Environmetrics* 16 (2), 275–289.
- Wikipedia: The Free Encyclopedia. Retrieved December 10, 2006 from the World Wide Web [www.wikipedia.org](http://www.wikipedia.org)

- Winkelmann, R. 1995: Duration dependence and dispersion in count data models. *Journal of Business & Economic Statistics* 1995, 13(4): 467-474.
- Winkelmann, R., 1995: Recent development in count data modeling: theory and applications. *Journal of Economic Survey*, 9, 1-24.
- Xie , Y., Lord, D., Zhang, Y 2007: Predicting motor vehicle collisions using Bayesian neural network models: An empirical analysis: *Accident Analysis and Prevention* 39, 922–933
- Yau, K.K.W., 2004. Risk factors affecting the severity of single vehicle traffic accidents in Hong Kong. *Accident Analysis and Prevention*. 36, 333–340.
- Zegeer, C.V., Council, F.M., 1995. Safety Relationship Associated with Cross-sectional Roadway Elements. *Transportation Research Record* 1512, Washington, D.C., USA, pp. 29–35.
- Zegeer, C.V., Deacon, J.A., 1987. Effect of lane width, shoulder width, and shoulder type on highway safety. In: *Relationship Between Safety and Key Highway Features*. State of the Art Report ,Transportation Research Board, Washington, D.C., USA.
- Zegeer, C.V., Reinfurt, D., Hummer, J., Herf, L., Hunter, W., 1987. Safety Effect of Cross-section Design for Two-lane Roads. vols. 1–2. Report FHWA-RD87/008 and 009 Federal Highway Administration, Department of Transportation, USA.

**Annex 1:** Descriptive Statistics and Histogram of the response variable

<i>No of vehicles crushed</i>	
Mean	1.801892
Standard Error	0.006193
Median	2
Mode	2
Standard Deviation	0.51726
Sample Variance	0.267558
Kurtosis	10.65195
Skewness	0.687405
Range	8
Minimum	1
Maximum	9
Sum	12570
Count	6976
Confidence Level (95.0%)	0.01214

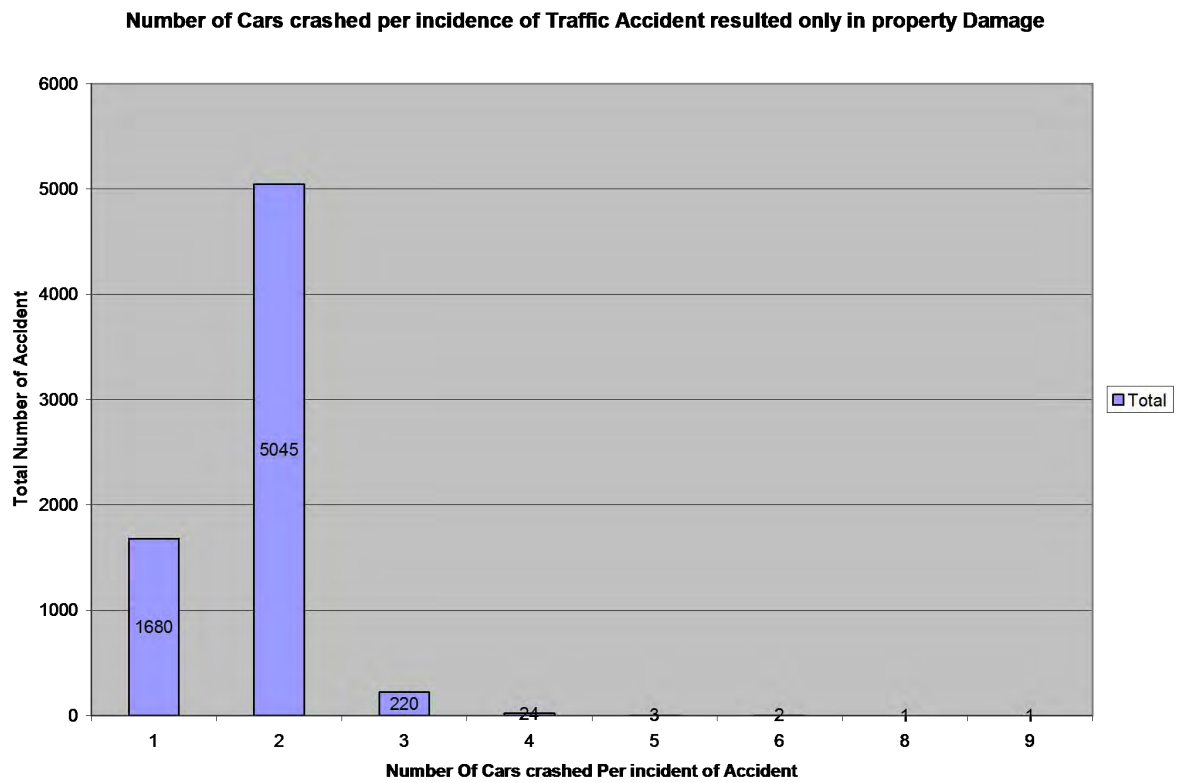


Figure 1:- Number of Cars crashed

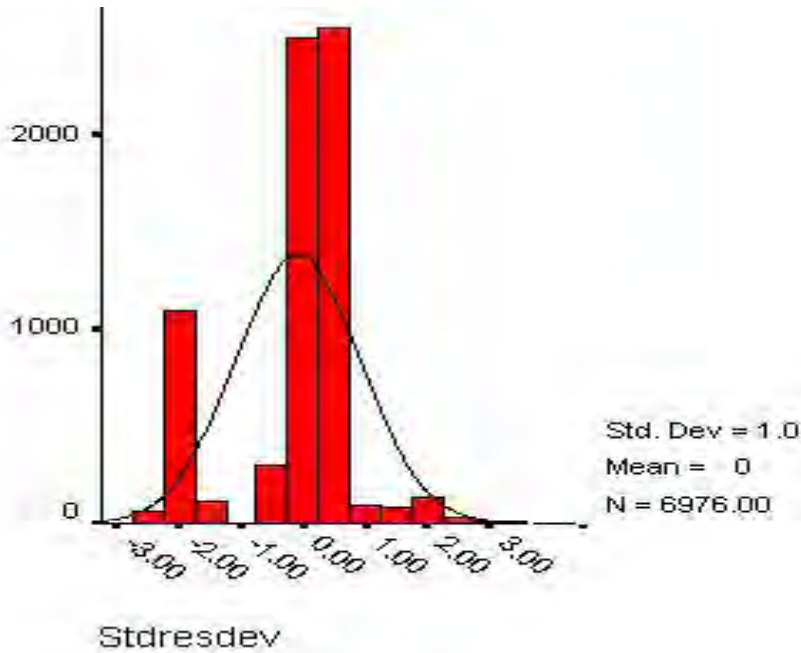
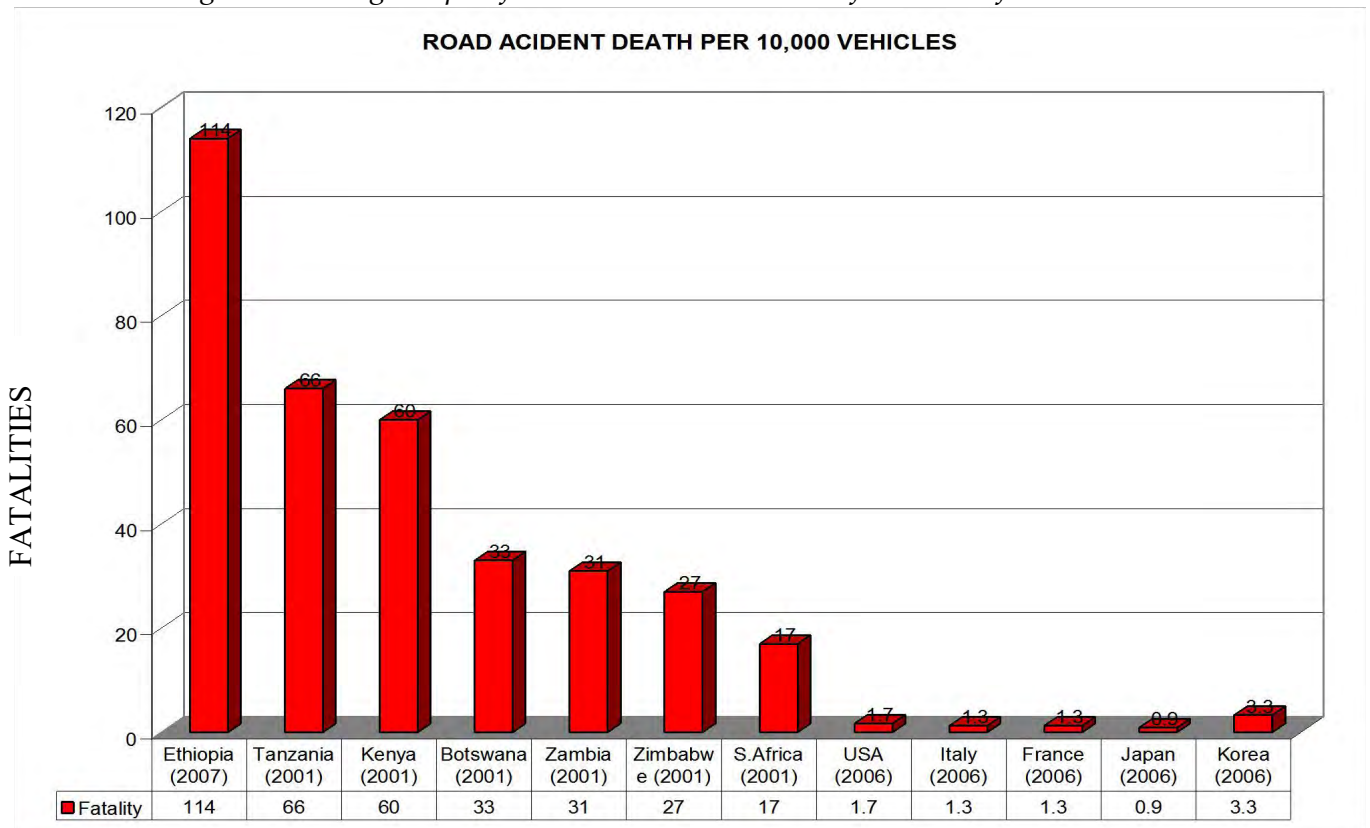


Figure 2: Histogram plot for standardized deviance of Number of Cars crashed



## Annex 2: Poisson Regression Model

Summary1: Summary information for numbers of vehicles crashed including all factors using Poisson Regression Model

The SAS System  
The GENMOD Procedure

### Model Information

Data Set	WORK.CARCRASH
Distribution	Poisson
Link Function	Log
Dependent Variable	Noofcarscrushed
Observations Used	6976

### Class Level Information

Class	Levels	Values
DateWk	7	1 2 3 4 5 6 7
Gender	2	1 2
DAgeCode	4	1 2 3 4
Educcode	6	1 2 3 4 5 6
Relationcode	3	1 2 3
DrExperCode	6	0 1 2 3 4 5
LiGracode	7	1 2 3 4 5 6 7
Subcitycode	9	1 2 3 4 5 6 7 8 9
Placecode	8	1 2 3 4 5 6 7 8
Licocode	5	1 2 3 4 5
WeCoCode	4	1 2 3 4
Platecode	5	1 2 3 4 5
ownercode	7	1 2 3 4 5 6 7
Vtypecode	13	1 2 3 4 5 6 7 8 9 10 11 12 13
SeryearCode	5	0 1 2 3 4
AcVeMoCode	11	1 2 3 4 5 6 7 8 9 10 11
TypCollCode	8	1 2 3 4 5 6 7 8
Rdclscode	5	1 2 3 4 5
RoadDrcode	6	1 2 3 4 5 6
RJCode	5	1 2 3 4 5
TyRoCode	3	1 2 3
RoCocode	3	1 2 3

Criteria For Assessing Goodness Of Fit

<i>Criterion</i>	<i>DF</i>	<i>Value</i>	<i>Value/DF</i>
<i>Deviance</i>	<i>6865</i>	<i>871.6570</i>	<i>0.1270</i>
<i>Scaled Deviance</i>	<i>6865</i>	<i>871.6570</i>	<i>0.1270</i>
<i>Pearson Chi-Square</i>	<i>6865</i>	<i>870.2673</i>	<i>0.1268</i>
<i>Scaled Pearson X2</i>	<i>6865</i>	<i>870.2673</i>	<i>0.1268</i>
<i>Log Likelihood</i>		<i>-5071.7518</i>	

*Algorithm converged.*

Summary2: Coefficient Estimation, Standard Error and Test Statistics of each category of indicator variables using Poisson Regression Model

The SAS System

The GENMOD Procedure  
Analysis of Parameter Estimates

Parameter	Levels	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
					Lower	Upper		
Intercept		1	-0.1905	0.7056	-1.5734	1.1924	0.07	0.7871
Monday	1	1	-0.0046	0.0402	-0.0835	0.0742	0.01	0.9080
Tuesday	2	1	-0.0096	0.0401	-0.0881	0.0689	0.06	0.8107
Wednesday	3	1	0.0007	0.0404	-0.0786	0.0800	0.00	0.9864
Thursday	4	1	0.0025	0.0398	-0.0756	0.0806	0.00	0.9506
Friday	5	1	-0.0090	0.0399	-0.0871	0.0691	0.05	0.8212
Saturday	6	1	-0.0004	0.0406	-0.0800	0.0792	0.00	0.9916
Sunday	7	0	0.0000	0.0000	0.0000	0.0000	.	.
Male	1	1	-0.0206	0.0494	-0.1173	0.0762	0.17	0.6773
Female	2	0	0.0000	0.0000	0.0000	0.0000	.	.
< 18 Year	1	1	-0.0661	0.2045	-0.4670	0.3347	0.10	0.7464
18-30 Years	2	1	-0.0023	0.0337	-0.0684	0.0639	0.00	0.9460
30-50 Years	3	1	-0.0103	0.0307	-0.0705	0.0499	0.11	0.7374
> 50 Years	4	0	0.0000	0.0000	0.0000	0.0000	.	.
Illiterate	1	1	0.0057	0.4196	-0.8167	0.8281	0.00	0.9892
Adult Education	2	1	0.0685	0.1143	-0.1555	0.2925	0.36	0.5491
Primary	3	1	-0.0099	0.0369	-0.0823	0.0625	0.07	0.7883
Junior	4	1	0.0171	0.0350	-0.0515	0.0856	0.24	0.6260

Secondary	5	1	-0.0054	0.0266	- 0.0576	0.0467	0.04	0.8380
Above Secondary	6	0	0.0000	0.0000	0.0000	0.0000	.	.
Owner	1	1	-0.0001	0.0335	- 0.0658	0.0657	0.00	0.9979
Hired	2	1	0.0219	0.0311	- 0.0390	0.0828	0.50	0.4816
Others	3	0	0.0000	0.0000	0.0000	0.0000	.	.
No License	0	1	0.0408	0.1247	- 0.2035	0.2851	0.11	0.7435
< a year	1	1	-0.0003	0.0501	- 0.0985	0.0978	0.00	0.9947
1-2 Years Exp	2	1	0.0015	0.0333	- 0.0639	0.0668	0.00	0.9649
2-5 Years EXP	3	1	-0.0070	0.0271	- 0.0601	0.0461	0.07	0.7962
5-10 Years Exp	4	1	0.0080	0.0259	- 0.0429	0.0588	0.09	0.7591
> 10 years EXP	5	0	0.0000	0.0000	0.0000	0.0000	.	.
1 <sup>st</sup> Grade Lice	1	1	-0.0485	0.2394	- 0.5176	0.4207	0.04	0.8394
2 <sup>nd</sup> Grade Lice	2	1	-0.0351	0.1096	- 0.2499	0.1797	0.10	0.7486
3 <sup>rd</sup> Grade Lice	3	1	-0.0256	0.1083	- 0.2379	0.1867	0.06	0.8135
4 <sup>th</sup> Grade Lice	4	1	-0.0280	0.1090	- 0.2417	0.1856	0.07	0.7970
5 <sup>th</sup> Grade Lice	5	1	-0.0525	0.1097	- 0.2675	0.1626	0.23	0.6327
No License	6	1	0.0151	0.1457	- 0.2704	0.3007	0.01	0.9173
Special Lice	7	0	0.0000	0.0000	0.0000	0.0000	.	.
Addis Ketema	1	1	-0.0233	0.0473	- 0.1160	0.0694	0.24	0.6228
Akaki	2	1	0.0715	0.5045	- 0.9172	1.0603	0.02	0.8873
Areeda	3	1	0.0036	0.0377	- 0.0704	0.0775	0.01	0.9244
Bole	4	1	-0.0185	0.0340	- 0.0850	0.0481	0.30	0.5865
Kirkos	5	1	-0.0038	0.0327	- 0.0678	0.0602	0.01	0.9073
Kolfie	6	1	0.0022	0.1501	-	0.2963	0.00	0.9884

					0.2920				
Lideta	7	1	-0.0090	0.0392	-	0.0678	0.05	0.8176	
					0.0859				
Nifas Silk	8	1	-0.1461	0.2849	-	0.4123	0.26	0.6081	
					0.7044				
Yeka	9	0	0.0000	0.0000	0.0000	0.0000	.	.	
Residence	1	1	0.0226	0.0638	-	0.1476	0.13	0.7232	
					0.1024				
School	2	1	0.0045	0.0643	-	0.1215	0.00	0.9441	
					0.1305				
Market Area	3	1	0.0036	0.0536	-	0.1087	0.00	0.9458	
					0.1014				
Around Office	4	1	0.0030	0.0478	-	0.0966	0.00	0.9498	
					0.0906				
Hospital Area	5	1	0.0496	0.0878	-	0.2218	0.32	0.5719	
					0.1225				
Factory Area	6	1	0.0619	0.3594	-	0.7663	0.03	0.8633	
					0.6425				
Recreation Area	7	1	-0.0143	0.0545	-	0.0925	0.07	0.7932	
					0.1211				
Religious Area	8	0	0.0000	0.0000	0.0000	0.0000	.	.	
Day Light	1	1	0.2099	0.2522	-0.2845	0.7043	0.69	0.4053	
During Sunrise	2	1	0.1731	0.2794	-0.3745	0.7207	0.38	0.5356	
During Sunset	3	1	0.1686	0.2714	-0.3633	0.7006	0.39	0.5343	
Dark Withroad Li	4	1	0.2048	0.2535	-0.2919	0.7016	0.65	0.4190	
Dark noRoadLight	5	0	0.0000	0.0000	0.0000	0.0000	.	.	
Good Weather	1	1	0.0507	0.1475	-0.2384	0.3397	0.12	0.7312	
Cold Weather	2	1	0.0430	0.1498	-0.2506	0.3367	0.08	0.7740	
Heavy rain	3	1	0.0568	0.1647	-0.2660	0.3796	0.12	0.7302	
Others	4	0	0.0000	0.0000	0.0000	0.0000	.	.	
Taxi	1	1	0.0933	0.1154	-0.1328	0.3194	0.65	0.4186	
Private	2	1	0.0199	0.0575	-0.0928	0.1327	0.12	0.7288	
Commercial	3	1	0.0216	0.0541	-0.0845	0.1277	0.16	0.6897	
Government	4	1	0.0369	0.0677	-0.0958	0.1696	0.30	0.5861	
Others	5	0	0.0000	0.0000	0.0000	0.0000	.	.	
Private	1	1	-0.0173	0.0981	-0.2096	0.1751	0.03	0.8604	
Government	2	1	-0.0404	0.1027	-0.2417	0.1608	0.16	0.6938	
Public Org	3	1	-0.0297	0.0994	-0.2246	0.1652	0.09	0.7652	
International	4	1	-0.0256	0.1502	-0.3200	0.2687	0.03	0.8644	
UN Agencies	5	1	0.0124	0.1295	-0.2415	0.2663	0.01	0.9238	
Police & Def	6	1	-0.0107	0.1122	-0.2306	0.2092	0.01	0.9239	
Core Diplomat	7	0	0.0000	0.0000	0.0000	0.0000	.	.	
Taxi	1	1	-0.2598	0.2357	-0.7216	0.2021	1.22	0.2703	

Automobile	2	1	-0.1743	0.2125	-0.5908	0.2421	0.67	0.4119
Motor Cycle	3	1	-0.2250	0.2667	-0.7478	0.2978	0.71	0.3990
Station Wagon	4	1	-0.1945	0.2121	-0.6102	0.2212	0.84	0.3591
Pickup 10 Qin	5	1	-0.1899	0.2128	-0.6070	0.2272	0.80	0.3723
Freight 11-40Q	6	1	-0.2226	0.2149	-0.6437	0.1986	1.07	0.3003
Freight41-100Q	7	1	-0.2663	0.2152	-0.6882	0.1555	1.53	0.2159
Public 12 Seats	8	1	-0.1925	0.2169	-0.6176	0.2326	0.79	0.3748
Public13-45Seats	9	1	-0.1633	0.2167	-0.5880	0.2614	0.57	0.4510
Public >45 seat	10	1	-0.2166	0.2167	-0.6414	0.2082	1.00	0.3177
Liquid Freight	11	1	-0.2968	0.2720	-0.8300	0.2364	1.19	0.2753
Freight with T	12	1	-0.2608	0.2363	-0.7240	0.2023	1.22	0.2697
Special Vehicle	13	0	0.0000	0.0000	0.0000	0.0000	.	.
1-2 years Ser	1	1	-0.0300	0.0423	-0.1129	0.0528	0.50	0.4777
2-5 years Ser	2	1	-0.0463	0.0628	-0.1693	0.0768	0.54	0.4612
5-10 Years Ser	3	1	0.0050	0.0279	-0.0496	0.0596	0.03	0.8582
> 10 Years ser	0	1	0.0203	0.0213	-0.0215	0.0621	0.91	0.3408
< a year	4	0	0.0000	0.0000	0.0000	0.0000	.	.
Move in Strt	1	1	0.0884	0.1844	-0.2731	0.4498	0.23	0.6318
Move Back	2	1	0.0271	0.1869	-0.3392	0.3934	0.02	0.8849
Enter Junction	3	1	0.0983	0.1941	-0.2821	0.4787	0.26	0.6126
Enter Crossed R	4	1	0.0715	0.1938	-0.3084	0.4515	0.14	0.7121
Overtaking	5	1	0.0436	0.1923	-0.3332	0.4205	0.05	0.8204
Going out	6	1	-0.0047	0.2464	-0.4876	0.4781	0.00	0.9847
U-Shaped turn	7	1	0.0425	0.2559	-0.4590	0.5440	0.03	0.8680
Turning Left	8	1	0.0824	0.1860	-0.2821	0.4469	0.20	0.6577
Turning Right	9	1	0.0921	0.1899	-0.2800	0.4643	0.24	0.6275
While Stopping	10	1	0.1366	0.2292	-0.3127	0.5859	0.36	0.5511
Opening Door	11	0	0.0000	0.0000	0.0000	0.0000	.	.
Head-to-Head	1	1	0.6575	0.1710	0.3223	0.9927	14.78	0.0001
Head-to-Tail	2	1	0.6211	0.1619	0.3037	0.9384	14.71	0.0001
Head-to-Side	3	1	0.5943	0.1623	0.2763	0.9123	13.41	0.0002
Side-to-Side	4	1	0.5669	0.1627	0.2480	0.8858	12.14	0.0005
Non-Living ob	5	1	0.0697	0.1682	-0.2600	0.3994	0.17	0.6785
Stopped Vehi	6	1	0.5973	0.1831	0.2384	0.9562	10.64	0.0011
Pedestrian	7	1	0.5439	0.3136	-0.0708	1.1586	3.01	0.0829
Overtuning	8	0	0.0000	0.0000	0.0000	0.0000	.	.
One Way Road	1	1	0.0525	0.1967	-0.3331	0.4380	0.07	0.7897
Two way Road	2	1	0.0500	0.1792	-0.3013	0.4013	0.08	0.7802
Solid paints	3	1	0.0683	0.1820	-0.2884	0.4250	0.14	0.7074
Divided by Isl	4	1	0.0554	0.1783	-0.2939	0.4048	0.10	0.7558
Divided by Bro	5	0	0.0000	0.0000	0.0000	0.0000	.	.
Straight Road	1	1	0.1196	0.3364	-0.5398	0.7790	0.13	0.7222
Some Inclined	2	1	0.1267	0.3450	-0.5494	0.8029	0.13	0.7134

Down Steep	3	1	0.1872	0.3509	-0.5005	0.8748	0.28	0.5937
Twisted Road	4	1	0.1128	0.3870	-0.6457	0.8713	0.08	0.7707
Steeped Road	5	1	0.0484	0.3570	-0.6512	0.7481	0.02	0.8921
St with moreinc	6	0	0.0000	0.0000	0.0000	0.0000	.	.
No junction	1	1	0.0127	0.0331	-0.0522	0.0775	0.15	0.7018
+ Cross shaped	2	1	0.0122	0.0240	-0.0348	0.0593	0.26	0.6101
T-Shaped Road	3	1	-0.1013	0.1114	-0.3197	0.1171	0.83	0.3632
Y-Shaped Road	4	1	0.0127	0.0442	-0.0740	0.0994	0.08	0.7742
Circled	5	0	0.0000	0.0000	0.0000	0.0000	.	.
Good Asphalt	1	1	-0.0412	0.3415	-0.7106	0.6282	0.01	0.9040
Damaged Asphalt	2	1	0.0004	0.3550	-0.6953	0.6961	0.00	0.9990
Concrete/Paved	3	0	0.0000	0.0000	0.0000	0.0000	.	.
Dry	1	1	-0.0392	0.1976	-0.4264	0.3480	0.04	0.8428
Wet	2	1	-0.0085	0.1968	-0.3942	0.3772	0.00	0.9655
Muddy Road	3	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale		1	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

Summary3: Type 1 Coefficient Estimation and Test Statistics of indicator variables using Poisson Regression Model

The SAS System  
The GENMOD Procedure  
LR Statistics For Type 1 Analysis

Source	Deviance	DF	Chi-Square	Pr > ChiSq
Intercept	1064.7827			
DateWk	1064.2601	6	0.52	0.9976
Gender	1064.1469	1	0.11	0.7365
DAgeCode	1063.3224	3	0.82	0.8436
Educcode	1062.3356	5	0.99	0.9636
Relationcode	1061.4734	2	0.86	0.6498
DrExperCode	1060.9014	5	0.57	0.9892
LiGracode	1056.5623	6	4.34	0.6309
Subcitycode	1052.9777	8	3.58	0.8925
Placecode	1051.6945	7	1.28	0.9889
Licocode	1047.9752	4	3.72	0.4453
WeCoCode	1047.0329	3	0.94	0.8152
Platecode	1045.0605	4	1.97	0.7408
ownercode	1044.8975	6	0.16	0.9999
Vtypecode	1035.8906	12	9.01	0.7023
SeryearCode	1033.2959	4	2.59	0.6278
AcVeMoCode	1029.8266	10	3.47	0.9681
TypCollCode	874.7920	7	155.03	<.0001
Rdclscode	874.5050	4	0.29	0.9906
RoadDrcode	873.5729	5	0.93	0.9679
RJCode	872.2618	4	1.31	0.8595
TyRoCode	872.0347	2	0.23	0.8926
RoCocode	871.6570	2	0.38	0.8279

Summary4: Type 3 Coefficient Estimation and Test Statistics of indicator variables using Poisson Regression Model

LR Statistics For Type 3 Analysis

Source	Chi-		
	DF	Square	Pr > ChiSq
DateWk	6	0.27	0.9996
Gender	1	0.17	0.6781
DAgeCode	3	0.30	0.9602
Educcode	5	1.17	0.9473
Relationcode	2	0.78	0.6769
DrExperCode	5	0.49	0.9927
LiGracode	6	1.06	0.9832
Subcitycode	8	1.09	0.9976
Placecode	7	1.01	0.9947
Licocode	4	1.00	0.9096
WeCoCode	3	0.14	0.9867
Platecode	4	0.80	0.9380
ownercode	6	0.54	0.9974
Vtypecode	12	6.20	0.9055
SeryearCode	4	2.82	0.5884
AcVeMoCode	10	4.73	0.9083
TypCollCode	7	153.14	<.0001
Rdclscode	4	0.27	0.9914
RoadDrcode	5	0.95	0.9666
RJCode	4	1.31	0.8599
TyRoCode	2	0.19	0.9110
RoCocode	2	0.38	0.8279

## Annex 3: Saturated Gamma Regression Model

Summary1: Summary information for numbers of vehicles crashed including all factors using Gamma Regression Model

The GENMOD Procedure

Model Information

Data Set	WORK.CARCRASH
Distribution	Gamma
Link Function	Log
Dependent Variable	No of vehicles crashed
Observations Used	6976

Class Level Information

Class	Levels	Values
Date Wk	7	1 2 3 4 5 6 7
Gender	2	1 2
DAgeCode	4	1 2 3 4
Educcode	6	1 2 3 4 5 6
Relationcode	3	1 2 3
DrExperCode	6	0 1 2 3 4 5
LiGracode	7	1 2 3 4 5 6 7
Subcitycode	9	1 2 3 4 5 6 7 8 9
Placecode	8	1 2 3 4 5 6 7 8
Licocode	5	1 2 3 4 5
WeCoCode	4	1 2 3 4
Platecode	5	1 2 3 4 5
ownercode	7	1 2 3 4 5 6 7
Vtypecode	13	1 2 3 4 5 6 7 8 9 10 11 12 13
SeryearCode	5	0 1 2 3 4
AcVeMoCode	11	1 2 3 4 5 6 7 8 9 10 11
TypCollCode	8	1 2 3 4 5 6 7 8
Rdclscode	5	1 2 3 4 5
RoadDrcode	6	1 2 3 4 5 6
RJCode	5	1 2 3 4 5
TyRoCode	3	1 2 3
RoCocode	3	1 2 3

Criteria For Assessing Goodness Of Fit

<i>Criterion</i>	<i>DF</i>	<i>Value</i>	<i>Value/DF</i>
<i>Deviance</i>	<i>6865</i>	<i>524.1595</i>	<i>0.0764</i>
<i>Scaled Deviance</i>	<i>6865</i>	<i>7062.2495</i>	<i>1.0287</i>
<i>Pearson Chi-Square</i>	<i>6865</i>	<i>495.9998</i>	<i>0.0723</i>
<i>Scaled Pearson X2</i>	<i>6865</i>	<i>6682.8407</i>	<i>0.9735</i>
<i>Log Likelihood</i>		<i>-4698.0392</i>	

*Algorithm converged.*

Summary2: Coefficient Estimation, Standard Error and Test Statistics of each category of indicator variables using Gamma Regression with full Model

Analysis of Parameter Estimates

Parameter	Levels	DF	Estimate	Standard Error	Likelihood Ratio 95% Confidence Limit		Chi-Square	Pr > ChiSq
					Lower	Upper		
					Intercept	1		
Monday	1	-0.0048	0.0147	-0.0336	0.0239	0.11	0.7447	
Tuesday	2	-0.0089	0.0146	-0.0375	0.0197	0.37	0.5424	
Wednesday	3	0.0017	0.0147	-0.0273	0.0305	0.01	0.9101	
Thursday	4	0.006	0.0145	-0.0225	0.0343	0.17	0.6816	
Friday	5	-0.0099	0.0145	-0.0383	0.0184	0.47	0.4938	
Saturday	6	-0.0007	0.0148	-0.0297	0.0283	0	0.9632	
Sunday	7	0	0	0	0	.	.	
Male	1	-0.0181	0.0182	-0.0539	0.0174	0.99	0.3203	
Female	2	0	0	0	0	.	.	
< 18 Year	1	-0.0628	0.0743	-0.2053	0.0863	0.71	0.3986	
18-30 Years	2	-0.002	0.0124	-0.0263	0.0222	0.03	0.871	
30-50 Years	3	-0.0093	0.0112	-0.0314	0.0127	0.69	0.4057	
> 50 Years	4	0	0	0	0	.	.	
Illiterate	1	0.0168	0.1613	-0.285	0.3488	0.01	0.9171	
Adult Education	2	0.0669	0.0427	-0.0159	0.1517	2.45	0.1176	
Primary	3	-0.0107	0.0135	-0.0371	0.0158	0.63	0.4278	
Junior	4	0.0165	0.0129	-0.0087	0.0417	1.64	0.1997	
Secondary	5	-0.0057	0.0097	-0.0248	0.0134	0.34	0.5585	
Above Secondary	6	0	0	0	0	.	.	
Owner	1	-0.0004	0.0122	-0.0243	0.0235	0	0.9745	
Hired	2	0.0208	0.0113	-0.0013	0.043	3.4	0.0651	
Others	3	0	0	0	0	.	.	
No License	0	0.0334	0.046	-0.0567	0.1238	0.53	0.4679	
< a year	1	-0.0014	0.0183	-0.037	0.0346	0.01	0.9396	
1-2 Years Exp	2	0.0018	0.0122	-0.022	0.0258	0.02	0.8801	
2-5 Years EXP	3	-0.0065	0.0099	-0.0258	0.0129	0.43	0.5125	
5-10 Years Exp	4	0.008	0.0095	-0.0098	0.0273	0.85	0.3561	

			7					
> 10 years Exp	5	0	0	0	0	0	.	.
1 <sup>st</sup> Grade Lice	1	1	-0.0373	0.0875	-0.208	0.1351	0.18	0.6695
2 <sup>nd</sup> Grade Lice	2	1	-0.0335	0.0403	-0.1133	0.0446	0.69	0.4058
3 <sup>rd</sup> Grade Lice	3	1	-0.0235	0.0398	-0.1025	0.0537	0.35	0.5544
4 <sup>th</sup> Grade Lice	4	1	-0.0252	0.0401	-0.1046	0.0525	0.4	0.5293
5 <sup>th</sup> Grade Lice	5	1	-0.0518	0.0403	-0.1316	0.0264	1.65	0.1988
No License	6	1	0.014	0.0539	-0.0915	0.1198	0.07	0.7905
			3					
Special Lice	7	0	0	0	0	0	.	.
Addis Ketema	1	1	-0.0208	0.0172	-0.0545	0.013	1.46	0.2267
Akaki	2	1	0.079	0.1942	-0.2789	0.4854	0.17	0.6812
			8					
Areeda	3	1	0.006	0.0139	-0.0212	0.0332	0.19	0.6643
Bole	4	1	-0.0177	0.0124	-0.042	0.0066	2.04	0.1537
Kirkos	5	1	-0.0016	0.012	-0.0251	0.0218	0.02	0.892
Kolfie	6	1	0.008	0.055	-0.0974	0.1182	0.02	0.8744
			7					
Lideta	7	1	-0.005	0.0144	-0.0332	0.0231	0.12	0.7267
Nifas Silk	8	1	-0.147	0.0985	-0.3346	0.052	2.23	0.1357
Yeka	9	0	0	0	0	0	.	.
Residence	1	1	0.020	0.0231	-0.0247	0.0659	0.8	0.3712
			7					
School	2	1	-0.0109	0.0234	-0.0567	0.0349	0.22	0.6417
Market Area	3	1	0.0017	0.0195	-0.0367	0.0398	0.01	0.9318
Around Office	4	1	0.0001	0.0174	-0.0341	0.034	0	0.9966
Hospital Area	5	1	0.073	0.0327	0.0093	0.1374	4.99	0.0254
Factory Area	6	1	0.0576	0.1382	-0.2023	0.3407	0.17	0.6769
Recreation Area	7	1	-0.017	0.0198	-0.0559	0.0216	0.74	0.3895
Religious Area	8	0	0	0	0	0	.	.
Day Light	1	1	0.1956	0.0795	0.0358	0.3476	6.06	0.0138
During Sunrise	2	1	0.1548	0.0904	-0.0253	0.3294	2.93	0.0869
During Sunset	3	1	0.1547	0.0873	-0.0196	0.3228	3.14	0.0764
Dark Withroad Li	4	1	0.1882	0.08	0.0276	0.3412	5.54	0.0186
Dark noRoadLight	5	0	0	0	0	0	.	.
Good Weather	1	1	0.0461	0.0543	-0.0619	0.1511	0.72	0.3961
Cold Weather	2	1	0.0372	0.0551	-0.0724	0.1438	0.45	0.5003
Heavy rain	3	1	0.0535	0.0608	-0.0668	0.1717	0.77	0.3789
Others	4	0	0	0	0	0	.	.
Taxi	1	1	0.0988	0.0426	0.0154	0.1824	5.38	0.0203
Private	2	1	0.0169	0.0208	-0.024	0.0576	0.66	0.4158
Commercial	3	1	0.0191	0.0196	-0.0194	0.0574	0.95	0.3292
Government	4	1	0.0385	0.0246	-0.0098	0.0866	2.45	0.1178

Others	5	0	0	0	0	0	.	.
Private	1	1	-0.0121	0.0359	-0.0828	0.0578	0.11	0.7354
Government	2	1	-0.0382	0.0375	-0.1121	0.0349	1.04	0.308
Public Org	3	1	-0.024	0.0363	-0.0956	0.0467	0.44	0.5077
International	4	1	-0.0232	0.0553	-0.1309	0.0859	0.18	0.6744
UN Agencies	5	1	0.0149	0.0476	-0.0782	0.1085	0.1	0.7544
Police & Def	6	1	-0.0086	0.041	-0.0891	0.0716	0.04	0.8336
Core Diplomat	7	0	0	0	0	0	.	.
Taxi	1	1	-0.2527	0.0933	-0.4386	-0.0725	7.33	0.0068
Automobile	2	1	-0.1597	0.0854	-0.3305	0.0044	3.5	0.0613
Motor Cycle	3	1	-0.2163	0.1037	-0.4213	-0.0146	4.35	0.037
Station Wagon	4	1	-0.1799	0.0853	-0.3505	-0.016	4.45	0.0349
Pickup 10 Qin	5	1	-0.1748	0.0855	-0.3458	-0.0104	4.18	0.041
Freight 11-40Q	6	1	-0.2072	0.086	-0.3791	-0.0418	5.8	0.016
Freight41-100Q	7	1	-0.2501	0.0861	-0.4222	-0.0846	8.45	0.0037
Public 12 Seats	8	1	-0.1788	0.0868	-0.3523	-0.0119	4.25	0.0393
Public13-45Seats	9	1	-0.1422	0.0867	-0.3154	0.0246	2.69	0.1009
Public >45 seat	10	1	-0.1921	0.0866	-0.3652	-0.0254	4.92	0.0266
Liquid Freight	11	1	-0.2823	0.1042	-0.4882	-0.0793	7.34	0.0068
Freight with T	12	1	-0.2385	0.0922	-0.4221	-0.0603	6.69	0.0097
Special Vehicle	13	0	0	0	0	0	.	.
1-2 years Ser	1	1	-0.0201	0.0257	-0.0702	0.0305	0.61	0.4347
2-5 years Ser	2	1	0.0319	0.0162	0.0001	0.0635	3.89	0.0485
5-10 Years Ser	3	1	0.0476	0.0149	0.0183	0.0767	10.22	0.0014
> 10 Years ser	0	1	0.0287	0.0153	-0.0013	0.0585	3.55	0.0596
< a year	4	0	0	0	0	0	0	.
Move in Strt	1	1	0.0887	0.0649	-0.0412	0.2134	1.87	0.172
Move Back	2	1	0.0317	0.0658	-0.0999	0.1583	0.23	0.6297
Enter Junction	3	1	0.1014	0.0687	-0.0357	0.2339	2.17	0.1403
Enter Crossed R	4	1	0.0717	0.0685	-0.0649	0.2038	1.1	0.2951
Overtaking	5	1	0.0436	0.0678	-0.0916	0.1742	0.41	0.5198
Going out	6	1	-0.01	0.0866	-0.1804	0.1593	0.01	0.9083
U-Shaped turn	7	1	0.0384	0.0914	-0.141	0.2177	0.18	0.6747
Turning Left	8	1	0.0844	0.0655	-0.0467	0.2104	1.66	0.1978
Turning Right	9	1	0.0923	0.067	-0.0415	0.2212	1.9	0.1684
While Stopping	10	1	0.1399	0.0824	-0.0227	0.3005	2.88	0.0895
Opening Door	11	0	0	0	0	0	.	.
Head-to-Head	1	1	0.6564	0.0493	0.558	0.7522	177.19	<.0001
Head-to-Tail	2	1	0.618	0.0446	0.529	0.7043	192.15	<.0001
Head-to-Side	3	1	0.5915	0.0447	0.502	0.678	174.72	<.0001

Side-to-Side	4	1	0.5634	0.045	0.474	0.6504	156.96	<.0001
Non-Living ob	5	1	0.069	0.0463	-0.0228	0.1587	2.23	0.1357
Stopped Vehi	6	1	0.5976	0.0544	0.490	0.7038	120.63	<.0001
					4			
Pedestrian	7	1	0.5426	0.1066	0.338	0.7564	25.91	<.0001
					1			
Overtuning	8	0	0	0	0	0	.	.
One Way Road	1	1	0.0537	0.0715	-0.0885	0.1918	0.56	0.4528
Two way Road	2	1	0.0571	0.0651	-0.0732	0.1823	0.77	0.3808
Solid paints	3	1	0.0724	0.0662	-0.0599	0.1998	1.2	0.2739
Divided by Isl	4	1	0.0601	0.0648	-0.0695	0.1846	0.86	0.3535
Divided by Bro	5	0	0	0	0	0	.	.
Straight Road	1	1	0.0943	0.1126	-0.1345	0.3075	0.7	0.4026
Some Inclined	2	1	0.1031	0.1161	-0.132	0.3236	0.79	0.3746
Down Steep	3	1	0.1647	0.1186	-0.075	0.3906	1.93	0.165
Twisted Road	4	1	0.0734	0.1321	-0.1908	0.3279	0.31	0.5786
Steeped Road	5	1	0.0189	0.1199	-0.2232	0.2475	0.02	0.8751
St with moreinc	6	0	0	0	0	0	.	.
No junction	1	1	0.0997	0.0385	0.023	0.1743	6.7	0.0096
					3			
+ Cross shaped	2	1	0.1111	0.04	0.032	0.1887	7.74	0.0054
T-Shaped Road	3	1	0.1117	0.039	0.034	0.1873	8.2	0.0042
					4			
Y-Shaped Road	4	1	0.1112	0.0413	0.029	0.1915	7.25	0.0071
					5			
Circled	5	0	0	0	0	0	.	.
Good Asphalt	1	1	-0.0333	0.1238	-0.2856	0.2005	0.07	0.7878
Damaged Asphalt	2	1	0.0142	0.1289	-0.2474	0.2588	0.01	0.9125
Concrete/Paved	3	0	0	0	0	0	.	.
Dry	1	1	-0.0396	0.0746	-0.1891	0.1034	0.28	0.5953
Wet	2	1	-0.0112	0.0744	-0.1603	0.1314	0.02	0.88
Muddy Road	3	0	0	0	0	0	.	.
Scale		1	13.4735	0.2254	13.0366	13.920		
					1			

NOTE: The scale parameter was estimated by maximum likelihood.

Summary3: Type 1 Coefficient Estimation and Test Statistics of indicator variables using Gamma Regression Full Model

LR Statistics For Type 1 Analysis

Source	2*Log Likelihood	DF	Chi- Square	Pr > ChiSq
Intercept	-10877.336			
DateWk	-10874.154	6	3.18	0.7858
Gender	-10873.476	1	0.68	0.4102
DAgeCode	-10868.454	3	5.02	0.1702
Educcode	-10862.506	5	5.95	0.3114
Relationcode	-10857.250	2	5.26	0.0722
DrExperCode	-10853.760	5	3.49	0.6249
LiGracode	-10827.207	6	26.55	0.0002
Subcitycode	-10804.947	8	22.26	0.0045
Placecode	-10796.845	7	8.10	0.3237
Licocode	-10773.444	4	23.40	0.0001
WeCoCode	-10767.633	3	5.81	0.1212
Platecode	-10755.278	4	12.35	0.0149
ownercode	-10754.348	6	0.93	0.9881
Vtypecode	-10697.674	12	56.67	<.0001
SeryearCode	-10681.664	4	16.01	0.0030
AcVeMoCode	-10659.870	10	21.79	0.0162
TypCollCode	-9419.4732	7	1240.40	<.0001
Rdclscode	-9417.6849	4	1.79	0.7746
RoadDrcode	-9410.1134	5	7.57	0.1815
RJCode	-9400.5269	4	9.59	0.0480
TyRoCode	-9398.4751	2	2.05	0.3585
RoCocode	-9396.0784	2	2.40	0.3017

Summary4: Type 3 Coefficient Estimation and Test Statistics of indicator variables using Gamma Regression Full Model

LR Statistics For Type 3 Analysis  
Chi-

Source	DF	Square	Pr > ChiSq
DateWk	6	2.88	0.8233
Gender	1	0.99	0.3192
DAgeCode	3	1.90	0.5936
Educcode	5	8.63	0.1248
Relationcode	2	5.39	0.0675
DrExperCode	5	3.46	0.6299
LiGracode	6	8.59	0.1980
Subcitycode	8	8.67	0.3711
Placecode	7	11.99	0.1007
Licocode	4	7.95	0.0936
WeCoCode	3	0.91	0.8228
Platecode	4	7.02	0.1350
ownercode	6	4.39	0.6235
Vtypecode	12	46.87	<.0001
SeryearCode	4	19.87	0.0005
AcVeMoCode	10	32.84	0.0003
TypCollCode	7	1230.43	<.0001
Rdclscode	4	1.77	0.7773
RoadDrcode	5	7.71	0.1727
RJCode	4	9.56	0.0485
TyRoCode	2	1.75	0.4175
RoCocode	2	2.40	0.3017

## Annex 4: Reduced Gamma Regression (First Model)

Summary1: Summary information for numbers of vehicles crashed using Reduced Gamma Regression (First Model)

The GENMOD Procedure

Model Information

Data Set	WORK.CARCRASH
Distribution	Gamma
Link Function	Log
Dependent Variable	Noofcarscrushed
Observations Used	6976

Class Level Information

Class	Levels	Values
Relationcode	3	1 2 3
Licocode	5	1 2 3 4 5
Vtypecode	13	1 2 3 4 5 6 7 8 9 10 11 12 13
SeryearCode	5	0 1 2 3 4
AcVeMoCode	11	1 2 3 4 5 6 7 8 9 10 11
TypCollCode	8	1 2 3 4 5 6 7 8
RJCode	5	1 2 3 4 5

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	6932	529.9920	0.0765
Scaled Deviance	6932	7063.1927	1.0189
Pearson Chi-Square	6932	510.4081	0.0736
Scaled Pearson X2	6932	6802.1992	0.9813
Log Likelihood		-4737.1165	

Algorithm converged.



Summary2: Coefficient Estimation, Standard Error and Test Statistics of each category of indicator variables using Reduced Gamma Regression (First Model)

Analysis Of Parameter Estimates

Parameter	Levels	DF	Estimate	Standard Error	Likelihood Ratio 95% Confidence Limit		Chi-Square	Pr > ChiSq
					Lower	Upper		
					Intercept			
Owner	1	1	-0.0013	0.0121	-0.0251	0.0225	0.01	0.9165
Hired	2	1	0.017	0.0106	-0.0035	0.038	2.66	0.103
Others	3	0	0	0	0	0	.	.
Day Light	1	1	0.203	0.0795	0.0433	0.3552	6.54	0.0106
During Sunrise	2	1	0.174	0.0905	-0.0057	0.3492	3.72	0.0537
During Sunset	3	1	0.165	0.0873	-0.0095	0.3332	3.57	0.0589
Dark Withroad Li	4	1	0.196	0.08	0.0358	0.3497	6.04	0.014
Dark noRoadLight	5	0	0	0	0	0	.	.
Taxi	1	1	-0.1861	0.0796	-0.3462	-0.0338	5.46	0.0194
Automobile	2	1	-0.1818	0.0795	-0.3418	-0.0297	5.22	0.0223
Motor Cycle	3	1	-0.2404	0.0926	-0.4245	-0.0614	6.75	0.0094
Station Wagon	4	1	-0.2038	0.0797	-0.3641	-0.0513	6.53	0.0106
Pickup 10 Qin	5	1	-0.1994	0.0799	-0.3601	-0.0465	6.22	0.0126
Freight 11-40Q	6	1	-0.229	0.0802	-0.3902	-0.0756	8.16	0.0043
Freight41-100Q	7	1	-0.2762	0.0805	-0.438	-0.1221	11.76	0.0006
Public 12 Seats	8	1	-0.1942	0.0813	-0.3574	-0.0386	5.71	0.0169
Public 13-45 S	9	1	-0.1666	0.081	-0.3293	-0.0114	4.23	0.0398
Public >45 seat	10	1	-0.2336	0.0805	-0.3953	-0.0796	8.42	0.0037
Liquid Freight	11	1	-0.2911	0.0993	-0.4876	-0.098	8.6	0.0034
Freight with T	12	1	-0.2766	0.0869	-0.4502	-0.1093	10.12	0.0015
Special Vehicle	13	0	0	0	0	0	.	.
> 10 Years ser	0	1	0.033	0.0151	0.004	0.0633	4.97	0.0259
1-2 years Ser	1	1	-0.0164	0.0256	-0.0664	0.034	0.41	0.523
2-5 years Ser	2	1	0.034	0.0161	0.0024	0.0655	4.47	0.0344
5-10 Years Ser	3	1	0.051	0.0148	0.0225	0.0805	12.17	0.0005
< a year	4	0	0	0	0	0	.	.
Move in Strt	1	1	0.075	0.0649	-0.0543	0.2003	1.36	0.244

			6					
Move Back	2	1	0.017 9	0.0658	-0.1137	0.1443	0.07	0.7861
Enter Junction	3	1	0.091 2	0.0688	-0.046	0.2237	1.76	0.185
Enter Crossed	4	1	0.059 3	0.0685	-0.0774	0.1913	0.75	0.3869
Overtaking	5	1	0.032 5	0.0678	-0.1028	0.1631	0.23	0.6315
Going out	6	1	-0.0136	0.0866	-0.184	0.1557	0.02	0.8756
U-Shaped turn	7	1	0.027 1	0.0915	-0.1525	0.2067	0.09	0.7673
Turning Left	8	1	0.072 7	0.0655	-0.0583	0.1986	1.23	0.2669
Turning Right	9	1	0.081 9	0.0669	-0.0519	0.2107	1.5	0.2214
While Stopping	10	1	0.127 3	0.0823	-0.0352	0.2878	2.39	0.122
Opening Door	11	0	0	0	0	0	.	.
Head-to-Head	1	1	0.658 9	0.0492	0.5616	0.7545	179.41	<.0001
Head-to-Tail	2	1	0.621 8	0.0445	0.5334	0.7078	195.47	<.0001
Head-to-Side	3	1	0.593 8	0.0446	0.5051	0.6801	176.89	<.0001
Side-to-Side	4	1	0.565	0.0449	0.4758	0.6518	158.49	<.0001
Non-Living ob	5	1	0.078 6	0.0462	-0.013	0.168	2.9	0.0885
Stopped Vehi	6	1	0.6	0.0544	0.4928	0.706	121.82	<.0001
Pedestrian	7	1	0.561 7	0.1066	0.3572	0.7757	27.75	<.0001
Overtuning	8	0	0	0	0	0	.	.
No junction	1	1	0.100 7	0.0384	0.0246	0.1751	6.89	0.0087
+ Cross shaped	2	1	0.111 3	0.0398	0.0325	0.1884	7.83	0.0051
T-Shaped Road	3	1	0.112 1	0.0389	0.035	0.1875	8.32	0.0039
Y-Shaped Road	4	1	0.109 6	0.0411	0.0282	0.1894	7.11	0.0077
Circled	5	0	0	0	0	0	.	.
Scale		1	13.327	0.2229	12.894 9	13.768 7		

*NOTE: The scale parameter was estimated by maximum likelihood.*

Summary4: Type 1 Coefficient Estimation and Test Statistics of indicator variables using  
Reduced Gamma Regression First Model

LR Statistics For Type 1 Analysis

Source	2*Log Likelihood	DF	Chi- Square	Pr > ChiSq
Intercept	-10877.336			
Relationcode	-10872.185	2	5.15	0.0761
Licocode	-10844.658	4	27.53	<.0001
Vtypecode	-10753.172	12	91.49	<.0001
SeryearCode	-10735.445	4	17.73	0.0014
AcVeMoCode	-10713.264	10	22.18	0.0142
TypCollCode	-9483.6292	7	1229.64	<.0001
RJCode	-9474.2330	4	9.40	0.0519

Summary4: Type 3 Coefficient Estimation and Test Statistics of indicator variables using  
Reduced Gamma Regression First Model

LR Statistics For Type 3 Analysis

Source	DF	Chi- Square	Pr > ChiSq
Relationcode	2	4.99	0.0825
Licocode	4	7.81	0.0987
Vtypecode	12	65.33	<.0001
SeryearCode	4	21.70	0.0002
AcVeMoCode	10	33.33	0.0002
TypCollCode	7	1221.44	<.0001
RJCode	4	9.40	0.0519

## Annex 5: Using Reduced Gamma Regression Second Model

### Summary1: Summary information for numbers of vehicles crashed using Reduced Gamma Regression Second Model

The GENMOD Procedure

#### Model Information

Data Set	WORK.CARCRASH	
Distribution	Gamma	
Link Function	Log	
Dependent Variable	Noofcarscrushed	Noofcarscrushed
Observations Used	6976	

#### Class Level Information

Class	Levels	Values
Licocode	5	1 2 3 4 5
Vtypecode	13	1 2 3 4 5 6 7 8 9 10 11 12 13
SeryearCode	5	0 1 2 3 4
TypCollCode	8	1 2 3 4 5 6 7 8
RJCode	5	1 2 3 4 5

#### Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	6944	532.8684	0.0767
Scaled Deviance	6944	7063.6596	1.0172
Pearson Chi-Square	6944	508.4698	0.0732
Scaled Pearson X2	6944	6740.2335	0.9707
Log Likelihood		-4756.2322	

Algorithm converged.

Summary2: Coefficient Estimation, Standard Error and Test Statistics of each category of indicator variables using Reduced Gamma Regression Second Model

Analysis Of Parameter Estimates

Parameter	Levels	DF	Estimate	Standard Error	Likelihood Ratio 95% Confidence Limit		Chi-Square	Pr > ChiSq
					Lower	Upper		
Intercept		1	-0.1119	0.1262	-0.3559	0.1390	0.79	0.3753
Day Light	1	1	0.2031	0.0797	0.0429	0.3554	6.50	0.0108
During Sunrise	2	1	0.1727	0.0906	-0.0078	0.3477	3.63	0.0567
During Sunset	3	1	0.1578	0.0875	-0.0170	0.3263	3.25	0.0713
Dark Withroad Li	4	1	0.1941	0.0802	0.0329	0.3474	5.86	0.0155
Dark noRoadLight	5	0	0.0000	0.0000	0.0000	0.0000	.	.
Taxi	1	1	-0.1890	0.0798	-0.3495	-0.0365	5.61	0.0179
Automobile	2	1	-0.1907	0.0797	-0.3510	-0.0385	5.73	0.0167
Motor Cycle	3	1	-0.2382	0.0928	-0.4227	-0.0588	6.59	0.0102
Station Wagon	4	1	-0.2102	0.0799	-0.3709	-0.0573	6.91	0.0086
Pickup 10 Qin	5	1	-0.2037	0.0801	-0.3648	-0.0505	6.47	0.0110
Freight 11-40Q	6	1	-0.2314	0.0804	-0.3930	-0.0776	8.29	0.0040
Freight41-100Q	7	1	-0.2771	0.0807	-0.4393	-0.1227	11.79	0.0006
Public 12 Seats	8	1	-0.1970	0.0815	-0.3606	-0.0410	5.85	0.0156
Public 13-45 S	9	1	-0.1694	0.0812	-0.3326	-0.0139	4.35	0.0370
Public >45 seat	10	1	-0.2383	0.0807	-0.4004	-0.0839	8.72	0.0031
Liquid Freight	11	1	-0.2888	0.0995	-0.4858	-0.0952	8.42	0.0037
Freight with T	12	1	-0.2777	0.0871	-0.4518	-0.1101	10.17	0.0014
Special Vehicle	13	0	0.0000	0.0000	0.0000	0.0000	.	.

> 10 Years ser	0	1	0.0311	0.0152	0.001 3	0.0608	4.22	0.0399
1-2 years Ser	1	1	-0.0167	0.0257	- 0.0668	0.0338	0.42	0.5152
2-5 years Ser	2	1	0.0334	0.0161	0.001 7	0.0649	4.30	0.0381
5-10 Years Ser	3	1	0.0491	0.0148	0.019 9	0.0781	10.98	0.0009
< a year	4	0	0.0000	0.0000	0.000 0	0.0000	.	.
Head-to-Head	1	1	0.6534	0.0493	0.555 9	0.7491	175.83	<.0001
Head-to-Tail	2	1	0.6125	0.0446	0.523 9	0.6986	188.97	<.0001
Head-to-Side	3	1	0.5845	0.0447	0.495 7	0.6709	171.14	<.0001
Side-to-Side	4	1	0.5570	0.0449	0.467 7	0.6439	153.55	<.0001
Non-Living ob	5	1	0.0702	0.0462	- 0.0216	0.1598	2.30	0.1291
Stopped Vehi	6	1	0.5760	0.0543	0.468 9	0.6819	112.44	<.0001
Pedestrian	7	1	0.5511	0.1069	0.346 2	0.7656	26.59	<.0001
Overturning	8	0	0.0000	0.0000	0.000 0	0.0000	.	.
No junction	1	1	0.0969	0.0384	0.020 7	0.1713	6.37	0.0116
+ Cross shaped	2	1	0.1095	0.0398	0.030 6	0.1867	7.57	0.0059
T-Shaped Road	3	1	0.1089	0.0389	0.031 7	0.1843	7.83	0.0051
Y-Shaped Road	4	1	0.1089	0.0411	0.027 6	0.1886	7.03	0.0080
Circled	5	0	0.0000	0.0000	0.000 0	0.0000	.	.
Scale		1	13.2559	0.2217	12.826 2	13.695 2		

NOTE: The scale parameter was estimated by maximum likelihood.

Summary4: Type 1 Coefficient Estimation and Test Statistics of indicator variables using Reduced Gamma Regression Second Model

LR Statistics For Type 1 Analysis

Source	2*Log Likelihood	DF	Chi-Square	Pr > ChiSq
Intercept	-10877.336			
Licocode	-10848.256	4	29.08	<.0001
Vtypecode	-10764.059	12	84.20	<.0001
SeryearCode	-10745.877	4	18.18	0.0011
TypCollCode	-9522.0828	7	1223.79	<.0001
RJCode	-9512.4643	4	9.62	0.0474

Summary4: Type 3 Coefficient Estimation and Test Statistics of indicator variables using Reduced Gamma Regression Second Model

LR Statistics For Type 3 Analysis

Source	DF	Chi-Square	Pr > ChiSq
Licocode	4	8.51	0.0745
Vtypecode	12	62.84	<.0001
SeryearCode	4	20.03	0.0005
TypCollCode	7	1213.92	<.0001
RJCode	4	9.62	0.0474

## Annex 6: Definition of design matrix

Definition of design matrix used in both saturated and reduced model of Gamma Regression Model

Factors	Variable	Variable Representation	Description of Dummy variables
Demographic Factors	Date of the Week	X <sub>i1</sub>	1 If accidents of vehicles crashed occurred on Monday, 0 otherwise
		X <sub>i2</sub>	1 If accidents of vehicles crashed occurred on Tuesday 0 otherwise
		X <sub>i3</sub>	1 If accidents of vehicles crashed occurred on Wednesday, 0 otherwise
		X <sub>i4</sub>	1 If accidents of vehicles crashed occurred on Thursday, 0 otherwise
		X <sub>i5</sub>	1 If accidents of vehicles crashed occurred on Friday, 0 otherwise
		X <sub>i6</sub>	1 If accidents of vehicles crashed occurred on Saturday 0 otherwise
		X <sub>i7</sub>	1 If accidents of vehicles crashed occurred on Sunday, 0 otherwise
	Sex of driver	X <sub>i8</sub>	1 If drivers of accused vehicle is Male,0 otherwise
		X <sub>i9</sub>	1 If drivers of accused vehicle is Female,0 otherwise
	Drivers age categories	X <sub>i10</sub>	1 If drivers age category is less than 18 years, 0 Otherwise
		X <sub>i11</sub>	1 If drivers age category is between 18 and 30 years, 0 Otherwise
		X <sub>i12</sub>	1 If drivers age category is Between 31 and 50 years, 0 Otherwise
		X <sub>i13</sub>	1 If drivers age category is above 51 years, 0 Otherwise
	Drivers Educational Status	X <sub>i14</sub>	1 if educational status of the driver is Illiterate; 0 otherwise
		X <sub>i15</sub>	1 if educational status of the driver is only Adult Education; 0 otherwise
		X <sub>i16</sub>	1 if educational status of the driver is Primary School, 0 otherwise
		X <sub>i17</sub>	1 if educational status of the driver is Junior School, 0 otherwise
		X <sub>i18</sub>	1 if educational status of the driver is Secondary School, 0 otherwise
		X <sub>i19</sub>	1 if educational status of the driver is Above High School, 0 otherwise
	Relation with Vehicle	X <sub>i20</sub>	1 if the driver of the vehicle is the owner of the vehicle, 0 otherwise
		X <sub>i21</sub>	1 if the driver of the vehicle is the Hired driver, 0 otherwise
		X <sub>i22</sub>	1 if the driver of the vehicle is the others, 0 otherwise
	Drivers Experience	X <sub>i23</sub>	1 if drivers have no experience/license, 0 otherwise
		X <sub>i24</sub>	1 if drivers experience is less than a years, 0 otherwise
		X <sub>i25</sub>	1 if drivers experience is between 1 and 2 years, 0 otherwise
		X <sub>i26</sub>	1 if drivers experience is between 2 and 5 years, 0 otherwise
		X <sub>i27</sub>	1 if drivers experience is between 5 and 10 years, 0 otherwise

		X <sub>i28</sub>	1 if drivers experience is above 10 years, 0 otherwise
	License Grade	X <sub>i29</sub>	1 if drivers have no license, 0 otherwise
		X <sub>i30</sub>	1 if drivers have first grade license, 0 otherwise
		X <sub>i31</sub>	1 if drivers have Second grade license, 0 otherwise
		X <sub>i32</sub>	1 if drivers have Third grade license, 0 otherwise
		X <sub>i33</sub>	1 if drivers have Fourth grade license, 0 otherwise
		X <sub>i34</sub>	1 if drivers have Fifth grade license, 0 otherwise
		X <sub>i35</sub>	1 if drivers have Special license, 0 otherwise
Site Factors		Sub City	X <sub>i36</sub>
	X <sub>i37</sub>		1 if accident is happened in Akaki Sub-City, 0 otherwise
	X <sub>i38</sub>		1 if accident is happened in Areda Sub-City, 0 otherwise
	X <sub>i39</sub>		1 if accident is happened in Bole Sub-City, 0 otherwise
	X <sub>i40</sub>		1 if accident is happened in Kirkos Sub-City, 0 otherwise
	X <sub>i41</sub>		1 if accident is happened in Kolfe Kereniyo Sub-City, 0 otherwise
	X <sub>i42</sub>		1 if accident is happened in Lideta Sub-City, 0 otherwise
	X <sub>i43</sub>		1 if accident is happened in Nifas Silk Lefto Sub-City, 0 otherwise
	X <sub>i44</sub>		1 if accident is happened in Yeka Sub-City, 0 otherwise
	Place of Accident	X <sub>i45</sub>	1 if incident of accident is happened in Residence area, 0 otherwise
		X <sub>i46</sub>	1 if incident of accident is happened around School, 0 otherwise
		X <sub>i47</sub>	1 if incident of accident is happened around Market area, 0 otherwise
		X <sub>i48</sub>	1 if incident of accident is happened around Office, 0 otherwise
		X <sub>i49</sub>	1 if incident of accident is happened around Hospital, 0 otherwise
		X <sub>i50</sub>	1 if incident of accident is happened around Factory, 0 otherwise
		X <sub>i51</sub>	1 if incident of accident is happened around Recreational area, 0 otherwise
		X <sub>i52</sub>	1 if incident of accident is happened around Religious Area, 0 otherwise
	Light Condition	X <sub>i53</sub>	1 if incident of accident is happened during day light, 0 otherwise
		X <sub>i54</sub>	1 if incident of accident is happened during Sunrises, 0 otherwise
X <sub>i55</sub>		1 if incident of accident is happened during Sunset, 0 otherwise	
X <sub>i56</sub>		1 if incident of accident is happened during Dark with road light, 0 otherwise	
X <sub>i57</sub>		1 if incident of accident is happened Dark with poor road light, 0 otherwise	
Weather condition	X <sub>i58</sub>	1 if the weather condition of the accident area Good Weather, 0 otherwise	
	X <sub>i59</sub>	1 if the weather condition of the accident area Cold Weather, 0 otherwise	

		$X_{i60}$	1 if the weather condition of the accident area Slight Rain, 0 otherwise
		$X_{i61}$	1 if the weather condition of the accident area Other (Hot, Cold, etc), 0 otherwise
Vehicle Related Factors	Registration type	$X_{i62}$	1 if plate code of accused vehicle is 1, 0 Otherwise
		$X_{i63}$	1 if plate code of accused vehicle is 2, 0 Otherwise
		$X_{i64}$	1 if vehicle plate code of accused vehicle is 3, 0 Otherwise
		$X_{i65}$	1 if plate code of accused vehicle is 4, 0 Otherwise
		$X_{i66}$	1 if plate code of accused vehicle is 5, 0 Otherwise
	Ownership	$X_{i67}$	1 if vehicle crashed belongs to Private owner, 0 Otherwise
		$X_{i68}$	1 if vehicle crashed belongs to Government, 0 Otherwise
		$X_{i69}$	1 if vehicle crashed belongs to Public Organization, 0 otherwise
		$X_{i70}$	1 if vehicle crashed belongs to International Organization, 0 otherwise
		$X_{i71}$	1 if vehicle crashed belongs to UN Agencies, 0 otherwise
		$X_{i72}$	1 if vehicle crashed belongs to Defense and Police owned vehicles, 0 otherwise
		$X_{i73}$	1 if vehicle crashed belongs to Core Diplomat and Others, 0 otherwise
	Vehicle Type	$X_{i74}$	1 if the vehicle type that resulted accident is Tax, 0 otherwise
		$X_{i75}$	1 if the vehicle type that resulted accident is Automobile, 0 otherwise
		$X_{i76}$	1 if the vehicle type that resulted accident is Motor Cycle and Bicycle, 0 otherwise
		$X_{i77}$	1 if the vehicle type that resulted accident is Station Wagon, 0 otherwise
		$X_{i78}$	1 if the vehicle type that resulted accident is Pickup up to 10 Quintals, 0 otherwise
		$X_{i79}$	1 if the vehicle type that resulted accident is Freighter 11-40 Quintals, 0 otherwise
		$X_{i80}$	1 if the vehicle type that resulted accident is Freighter 41-100 Quintals, 0 otherwise
		$X_{i81}$	1 if the vehicle type that resulted accident is Public Bus with 12 seats, 0 otherwise
$X_{i82}$		1 if the vehicle type that resulted accident is Public Bus with 13-44 seats, 0 otherwise	
$X_{i83}$		1 if the vehicle type that resulted accident is Public Bus $\geq 45$ seats, 0 otherwise	
$X_{i84}$		1 if the vehicle type that resulted accident is Liquid Freighters, 0 otherwise	
$X_{i85}$		1 if the vehicle type that resulted accident is Freight with trailers, 0 otherwise	
$X_{i86}$		1 if the vehicle type that resulted accident is Special Vehicles, 0 otherwise	
Vehicle	$X_{i87}$	1 if Service Year of the vehicle is 1-2 years, 0 Otherwise	

	<b>Service Year</b>	X <sub>i88</sub>	1 if Service Year of the vehicle is 2-5 Years, 0 Otherwise
		X <sub>i89</sub>	1 if Service Year of the vehicle is 5-10 Years, 0 Otherwise
		X <sub>i90</sub>	1 if Service Year of the vehicle is above 10 year, 0 Otherwise
	<b>Vehicle Movement</b>	X <sub>i91</sub>	1 if vehicle is crashed while moving in straight line, 0 otherwise
		X <sub>i92</sub>	1 if vehicle is crashed while moving back ward, 0 otherwise
		X <sub>i93</sub>	1 if vehicle is crashed while Entering Road Junction, 0 otherwise
		X <sub>i94</sub>	1 if vehicle is crashed while Entering Crossed road, 0 otherwise
		X <sub>i95</sub>	1 if vehicle is crashed while Over taking, 0 otherwise
		X <sub>i96</sub>	1 if vehicle is crashed while Coming off from House, 0 otherwise
		X <sub>i97</sub>	1 if vehicle is crashed while U-Shaped turning, 0 otherwise
		X <sub>i98</sub>	1 if vehicle is crashed while Turning to Left, 0 otherwise
		X <sub>i99</sub>	1 if vehicle is crashed while Turning to Right, 0 otherwise
		X <sub>i100</sub>	1 if vehicle is crashed while stopping, 0 otherwise
		X <sub>i101</sub>	1 if vehicle is crashed while Opening the Door, 0 otherwise
	<b>Type of collision</b>	X <sub>i102</sub>	1 if type of collision is Head-to-Head, 0 Otherwise
		X <sub>i103</sub>	1 if type of collision is Head-to-Tail, 0 Otherwise
		X <sub>i104</sub>	1 if type of collision is Head-to-Side, 0 Otherwise
		X <sub>i105</sub>	1 if type of collision is Side-to-Side, 0 Otherwise
		X <sub>i106</sub>	1 if type of collision is with non-living objects, 0 Otherwise
X <sub>i107</sub>		1 if type of collision is with Stopped Vehicle, 0 Otherwise	
X <sub>i108</sub>		1 if type of collision is Collision with pedestrian, 0 Otherwise	
X <sub>i109</sub>		1 if type of collision is Overturning, 0 Otherwise	
<b>Road Factors</b>	<b>Road Classification</b>	X <sub>i110</sub>	1 if road on which accidents was happened is One Way road, 0 Otherwise
		X <sub>i111</sub>	1 if road on which accidents was happened is Two Way road, 0 Otherwise
		X <sub>i112</sub>	1 if road on which accidents was happened is divided by Solid Paints, 0 Otherwise
		X <sub>i113</sub>	1 if road on which accidents was happened is Divided by Islands, 0 Otherwise
		X <sub>i114</sub>	1 if road on which accidents was happened is Divided by Broken Paint, 0 Otherwise
		<b>Road Direction</b>	X <sub>i115</sub>
	X <sub>i116</sub>		1 if road on which accidents was happened is Straight and Flat, 0 Otherwise
	X <sub>i117</sub>		1 if road on which accidents was happened is Straight with Slight inclination, 0 Otherwise
	X <sub>i118</sub>	1 if road on which accidents was happened is Down Steep, 0 Otherwise	

	$X_{i119}$	1 if road on which accidents was happened is Slightly Twisted, 0 Otherwise
	$X_{i120}$	1 if road on which accidents was happened is Steep Road, 0 Otherwise
	$X_{i121}$	1 if road on which accidents was happened is Straight with More Inclination, 0 Otherwise
<b>Road Junction</b>	$X_{i122}$	1 if road on which accidents was happened is Cross Shaped Road(+), 0 Otherwise
	$X_{i123}$	1 if road on which accidents was happened is T-Shaped Road, 0 Otherwise
	$X_{i124}$	1 if road on which accidents was happened is Y-Shaped Road, 0 Otherwise
	$X_{i125}$	1 if road on which accidents was happened is Circled Island, 0 Otherwise
	$X_{i126}$	1 if road on which accidents was happened has no Junction, 0 Otherwise
<b>Road Type</b>	$X_{i127}$	1 if road on which accidents was happened is Good Asphalt, 0 Otherwise
	$X_{i128}$	1 if road on which accidents was happened is damaged Asphalt, 0 Otherwise
	$X_{i129}$	1 if road on which accidents was happened is Others(Concrete, Pebble), 0 Otherwise
<b>Road Condition</b>	$X_{i130}$	1 if road on which accidents was happened is Dry Road, 0 Otherwise
	$X_{i131}$	1 if road on which accidents was happened is Wet Road, 0 Otherwise
	$X_{i132}$	1 if road on which accidents was happened is Muddy Road, 0 Otherwise

## Annex 7: SAS Codes for the Models

```
Data carcrash;
set kasahun.thesis4;
proc genmod data=carcrash ;
class DateWk      Gender      DAgeCode      Educcode      Relationcode
      DrExperCode LiGracode   Subcitycode  Placecode     Licocode
      WeCoCode    Platecode   ownercode   Vtypecode     SeryearCode
      AcVeMoCode  TypCollCode Rdclscode   RoadDrcode   RJCode
      TyRoCode    RoCocode;
model Noofcarscrushed=DateWk  Gender      DAgeCode      Educcode
      Relationcode      DrExperCode LiGracode   Subcitycode  Placecode
      Licocode          WeCoCode    Platecode   ownercode   Vtypecode
      SeryearCode AcVeMoCode  TypCollCode Rdclscode   RoadDrcode
      RJCode          TyRoCode    RoCocode/  dist=Poison  type1
      type3 link=log;
run;
```

---

```
proc genmod data=carcrash ;/* step 1*/
class DateWk      Gender      DAgeCode      Educcode      Relationcode
      DrExperCode LiGracode   Subcitycode  Placecode     Licocode
      WeCoCode    Platecode   ownercode   Vtypecode     SeryearCode
      AcVeMoCode  TypCollCode Rdclscode   RoadDrcode   RJCode
      TyRoCode    RoCocode;
model Noofcarscrushed=DateWk  Gender      DAgeCode      Educcode
      Relationcode      DrExperCode LiGracode   Subcitycode  Placecode
      Licocode          WeCoCode    Platecode   ownercode   Vtypecode
      SeryearCode AcVeMoCode  TypCollCode Rdclscode   RoadDrcode
      RJCode          TyRoCode    RoCocode/  dist=gamma  type1 type3 link=log
lrci;
run;
```

---

```
proc genmod data=carcrash ;/* step2*/
class Relationcode Licocode Vtypecode      SeryearCode AcVeMoCode
      TypCollCode RJCode;
model Noofcarscrushed=Relationcode Licocode Vtypecode SeryearCode
      AcVeMoCode  TypCollCode RJCode/  dist=gamma  type1 type3 link=log
lrci;
run;
```

---

```
proc genmod data=carcrash ;/* step 3*/
class Licocode Vtypecode      SeryearCode TypCollCode RJCode;
model Noofcarscrushed= Licocode Vtypecode SeryearCode TypCollCode
      RJCode/  dist=gamma  type1 type3 link=log lrci;
run;
```

### Annex 8: Descriptive Statistics for Significant variables.

Significant Factors	Variable Name	Categories	Multiple Vehicle				Single Vehicle		Overall			
			Number of Cars Crashed	% of vehicles Crashed	Number of accidents	Cars crashed per accidents	Number of Cars Crashed	% of vehicles Crashed	Total Number of Cars crashed	% of vehicles Crashed	Total Number of Accidents	Cars crashed per accidents
Demographic	Relation With Vehicle	Owner	1785	16.4	689	2.6	290	17.3	2075	16.5	1159	1.8
		Hired	7698	70.7	3742	2.1	1137	67.7	8835	70.3	4879	1.8
		Others	1407	12.9	685	2.1	253	15.1	1660	13.2	938	1.8
		<b>Total</b>	<b>10890</b>	<b>100</b>	<b>5296</b>	<b>2.1</b>	<b>1680</b>	<b>100</b>	<b>12570</b>	<b>100</b>	<b>6976</b>	<b>1.8</b>
	License Grade	No License	165	1.5	77	2.1	37	2.3	202	1.6	114	1.8
		First Grade	29	0.3	14	2.1	6	0.4	35	0.3	20	1.8
		Second Grade	2746	25.6	1341	2.0	420	25.6	3166	25.2	1761	1.8
		Third Grade	4606	42.9	2236	2.1	616	37.5	5222	41.5	2852	1.8
		Fourth Grade	1927	18.0	938	2.1	308	18.7	2235	17.8	1246	1.8
		Fifth Grade	1331	12.4	651	2.0	276	16.8	1607	12.8	927	1.7
		Special License	86	0.8	39	2.2	17	1.0	103	0.8	56	1.8
	<b>Total</b>	<b>10725</b>	<b>100.0</b>	<b>5296</b>	<b>2.0</b>	<b>1643</b>	<b>100</b>	<b>12570</b>	<b>100.0</b>	<b>6976</b>	<b>1.8</b>	
	Vehicle Related Factors	Registration type	Code 1	2193	20.1	1066	2.1	245	14.6	2438	19.4	1311
Code 2			3267	30.0	1590	2.1	507	30.2	3774	30.0	2097	1.8
Code 3			3893	35.7	1885	2.1	702	41.8	4595	36.6	2587	1.8
Code 4			817	7.5	403	2.0	112	6.7	929	7.4	515	1.8
Code 5			720	6.6	352	2.0	114	6.8	834	6.6	466	1.8
<b>Total</b>			<b>10890</b>	<b>100.0</b>	<b>5296</b>	<b>2.1</b>	<b>1680</b>	<b>100</b>	<b>12570</b>	<b>100.0</b>	<b>6976</b>	<b>1.8</b>
Service Year of Vehicle		Less than 1 year	575	5.3	280	2.1	115	6.8	690	5.5	395	1.7
		1-2 years	226	2.1	109	2.1	54	3.2	280	2.2	163	1.7
		2-5 years	1805	16.6	874	2.1	282	16.8	2087	16.6	1156	1.8
		5-10 years	4909	45.1	2403	2	627	37.3	5536	44	3030	1.8

	Above 10 years	3375	31	1630	2.1	602	35.8	3977	31.6	2232	1.8
	<b>Total</b>	<b>10890</b>	<b>100</b>	<b>5296</b>	<b>2.1</b>	<b>1680</b>	<b>100</b>	<b>12570</b>	<b>100</b>	<b>6976</b>	<b>1.8</b>
<b>Vehicle Type</b>	41-100 Quintals	467	4.3	228	2	143	8.5	610	4.9	371	1.6
	Automobile	3445	31.6	1682	2	485	28.9	3930	31.3	2167	1.8
	Bicycle	2	0	1	2	2	0.1	4	0	3	1.3
	Liquid Freighters	27	0.2	13	2.1	8	0.5	35	0.3	21	1.7
	Freight 11-40 Quintals	718	6.6	345	2.1	145	8.6	863	6.9	490	1.8
	Freight with trailer	63	0.6	29	2.2	31	1.8	94	0.7	60	1.6
	Motor Cycle	44	0.4	21	2.1	9	0.5	53	0.4	30	1.8
	Pickup up to 10 Quintals	1154	10.6	567	2	147	8.8	1301	10.4	714	1.8
	Public 12 seats	371	3.4	180	2.1	50	3	421	3.3	230	1.8
	Public 13-45 seats	435	4	210	2.1	52	3.1	487	3.9	262	1.9
	Public above 46 seats	543	5	263	2.1	117	7	660	5.3	380	1.7
	Special Vehicles	23	0.2	8	2.9	4	0.2	27	0.2	12	2.3
	Station Wagon	1429	13.1	695	2.1	237	14.1	1666	13.3	932	1.8
	Taxi	2169	19.9	1054	2.1	250	14.9	2419	19.2	1304	1.9
	<b>Total</b>	<b>10890</b>	<b>100</b>	<b>5296</b>	<b>2.1</b>	<b>1680</b>	<b>100</b>	<b>12570</b>	<b>100</b>	<b>6976</b>	<b>1.8</b>
<b>Vehicle Movement</b>	While Entering Crossed Road	252	2.3	125	2.0	30	1.8	282	2.2	155	1.8
	While Entering to Road Junction	254	2.3	126	2.0	21	1.3	275	2.2	147	1.9
	While Coming off from home	30	0.3	15	2.0	8	0.5	38	0.3	23	1.7
	While	270	2.5	134	2.0	51	3.0	321	2.6	185	1.7

	Overtaking										
	While Stopping	49	0.4	24	2.0	5	0.3	54	0.4	29	1.9
	While U-Shape turning	28	0.3	14	2.0	4	0.2	32	0.3	18	1.8
	While Turn to left	1442	13.2	707	2.0	179	10.7	1621	12.9	886	1.8
	While Turn to right	409	3.8	201	2.0	61	3.6	470	3.7	262	1.8
	While moving back	885	8.1	435	2.0	183	10.9	1068	8.5	618	1.7
	While moving straight	7248	66.6	3504	2.1	1131	67.3	8379	66.7	4635	1.8
	While opening door	23	0.2	11	2.1	7	0.4	30	0.2	18	1.7
	<b>Total</b>	<b>10890</b>	<b>100.0</b>	<b>5296</b>	<b>2.1</b>	<b>1680</b>	<b>100.0</b>	<b>12570</b>	<b>100.0</b>	<b>6976</b>	<b>1.8</b>
<b>Type of Collision</b>	Collision of Pedestrian	12	0.1	6	2.0	2	0.1	14	0.1	8	1.8
	Collision with non living Objects	63	0.6	28	2.3	371	22.1	434	3.5	399	1.1
	collision with Stopped Vehicle	115	1.1	54	2.1	22	1.3	137	1.1	76	1.8
	Head to Head collision	301	2.8	147	2.0	13	0.8	314	2.5	160	2.0
	Head to side Collision	3607	33.1	1768	2.0	422	25.1	4029	32.1	2190	1.8
	Head to tail Collision	4775	43.8	2297	2.1	502	29.9	5277	42.0	2799	1.9
	Overtaking		0.0			39	2.3	39	0.3	39	1.0
	Side to side Collision	2017	18.5	996	2.0	309	18.4	2326	18.5	1305	1.8

		<b>Total</b>	<b>10890</b>	<b>100.0</b>	<b>5296</b>	<b>2.1</b>	<b>1680</b>	<b>100.0</b>	<b>12570</b>	<b>100.0</b>	<b>6976</b>	<b>1.8</b>	
Site Factors	Sub City	Addis ketema	649	6	319	2	110	6.5	759	6	429	1.8	
		Akaki	4	0	2	2		0	4	0	2	2	
		Arada	1418	13	685	2.1	202	12	1620	12.9	887	1.8	
		Bole	2257	20.7	1096	2.1	448	26.7	2705	21.5	1544	1.8	
		Kirkos	4133	38	2012	2.1	549	32.7	4682	37.2	2561	1.8	
		Kolfie	42	0.4	21	2	5	0.3	47	0.4	26	1.8	
		Lideta	1176	10.8	572	2.1	176	10.5	1352	10.8	748	1.8	
		Nifas Silk	10	0.1	5	2	3	0.2	13	0.1	8	1.6	
		Yeka	1201	11	584	2.1	187	11.1	1388	11	771	1.8	
		<b>Total</b>	<b>10890</b>	<b>100</b>	<b>5296</b>	<b>2.1</b>	<b>1680</b>	<b>100</b>	<b>12570</b>	<b>100</b>	<b>6976</b>	<b>1.8</b>	
	Light Condition	Dark with no road light		0.0				2	0.1	2	0.0	2	1.0
		Dark with poor road light	8	0.1	4	2.0	6	0.4	14	0.1	10	1.4	
		Dark with road light	1018	9.3	494	2.1	228	13.6	1246	9.9	722	1.7	
		Day Light	9722	89.3	4728	2.1	1417	84.3	11139	88.6	6145	1.8	
		During Sun rise	57	0.5	28	2.0	12	0.7	69	0.5	40	1.7	
		During Sun set	85	0.8	42	2.0	15	0.9	100	0.8	57	1.8	
		<b>Total</b>	<b>10890</b>	<b>100.0</b>	<b>5296</b>	<b>2.1</b>	<b>1680</b>	<b>100.0</b>	<b>12570</b>	<b>100.0</b>	<b>6976</b>	<b>1.8</b>	
	Road Factor	Road Junction	+ - Shape	994	9.1	486	2.0	117	7.0	1111	8.8	603	1.8
			Circle island	532	4.9	262	2.0	70	4.2	602	4.8	332	1.8
No junction			7060	64.8	3424	2.1	1188	70.7	8248	65.6	4612	1.8	
T-Shape			2243	20.6	1094	2.1	283	16.8	2526	20.1	1377	1.8	
Y-Shape			61	0.6	30	2.0	22	1.3	83	0.7	52	1.6	
<b>Total</b>			<b>10890</b>	<b>100.0</b>	<b>5296</b>	<b>2.1</b>	<b>1680</b>	<b>100.0</b>	<b>12570</b>	<b>100.0</b>	<b>6976</b>	<b>1.8</b>	

Integrating by parts

$$\Pr [y_i=n] = G(\alpha n, \lambda_i) - G(\alpha n + \alpha, \lambda_i)$$

$$= \frac{1}{\Gamma(\alpha j)} \int_0^\infty u^{\alpha j - 1} e^{-u} du - \frac{1}{\Gamma(\alpha j + \alpha)} \int_0^\infty u^{\alpha j + \alpha - 1} e^{-u} du \quad \text{if } j > 0 \text{ and } j = 1, 2, 3, \dots$$

$$= \frac{1}{\Gamma(\alpha j)} - \frac{1}{\Gamma(\alpha j + \alpha)} \left( \int_0^\infty u^{\alpha j - 1} e^{-u} du - \int_0^\infty u^{\alpha j + \alpha - 1} e^{-u} du \right) \quad \text{if } j > 0 \text{ and } j = 1, 2, 3, \dots$$

$$= \frac{1}{\Gamma(\alpha j)} - \frac{1}{\alpha(\alpha j)} \left( \int_0^\infty u^{\alpha j - 1} e^{-u} du - \int_0^\infty u^\alpha u^{\alpha j - 1} e^{-u} du \right) \quad \text{if } j > 0 \text{ and } j = 1, 2, 3, \dots$$

$$= \frac{1}{\Gamma(\alpha j)} \left( 1 - \frac{1}{\alpha} \underbrace{\left( \int_0^\infty (1 - u^\alpha) u^{\alpha j - 1} e^{-u} du \right)} \right)$$

Integrating part by this part gives

$$1 - e^{-\lambda_i} \left( 1 + \lambda_i + \frac{(\lambda_i)^2}{2!} + \dots + \frac{(\lambda_i)^{\alpha n - 1}}{(\alpha n - 1)!} \right)$$

$$e^{\lambda_i} \sum_{i=1}^{\alpha-1} \frac{(\lambda_i)^{\alpha n + i}}{(\alpha n + i)!}, \quad n = 0, 1, 2, \dots$$

I, the undersigned, declare that the thesis is my original work, has not been presented for degrees in any other university and all sources of material used for the thesis have been duly acknowledged.

Name: Kassahun Hailemariam Beyene

Signature: .....

Place: Faculty of Science, Addis Ababa University

Date: December, 2008

This thesis has been submitted for examination with my approval as a University Advisor.

.....

.....

Dr. Butte Gotu