UNIVERSITY FOR DEVELOPMENT STUDIES

FORCASTING THE NUMBER OF DEATHS THROUGH ROAD TRAFFIC ACCIDENTS IN THE UPPER WEST REGION OF GHANA

ERIC WORDI

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

2013

UNIVERSITY FOR DEVELOPMENT STUDIES

FORCASTING THE NUMBER OF DEATHS THROUGH ROAD TRAFFIC ACCIDENTS IN THE UPPER WEST REGION OF GHANA

BY

ERIC WORDI (BSc APPLIED STATISTICS)

(UDS/MAS/0020/11)

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

AUGUST, 2013

DECLARATION

Student

I hereby declare that this thesis is the result of my original work and that no part of it has been presented for another degree in this University or elsewhere. Related works by others which served as a source of knowledge has been duly referenced and acknowledged.

Candidate's Signature..Date 13 - 08 - 13

Name: Wordi Eric

Supervisor

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

Supervisor's Signature..Date 15-08-13

Name: Dr. A. Y. Omari-Sasu

ABSTRACT

Many lives are lost yearly due to road traffic accidents worldwide. Apart from the psychological trauma that comes with these loses, economic and educational activities are also adversely affected. This work investigated the number of deaths through road traffic accident in the Upper West Region of Ghana by adopting a survey research method. Autoregressive Integrated Moving average (ARIMA) models were employed. The target population comprised of stakeholders including the Police service; Motor Transport and Traffic Unit from 2008 to 2012. The study variables based on the police report include the number of dead cases, age, gender and type of vehicles. The analysis was done using MINITAB. Data collected were subjected to statistical analysis and modeling, such as trend analysis and other descriptive analysis. The time series plot and the quadratic trend model revealed that there was an increasing trend in the death rate through road traffic accidents, number of deaths depend on time-period; ARIMA (2,1,3) fit the number of deaths through the road traffic accidents recorded. Furthermore, the forecast of this model showed a fluctuating trend. Based on these findings, it was recommended that; the vehicle inspection officers (VIO) should increase their inspections on road worthiness of many vehicles and the motor traffic and transport unit (MTTU) of the Ghana Police Service (GPS) should curb excessive speeding on our roads.

ACKNOWLEDGEMENTS

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

I would like to express my heartfelt appreciation to my supervisor, Dr. A. Y. Omari-Sasu, who read, scrutinized and provided the necessary support, encouragement and guidance which propelled me to complete this research.

To Dr. O.O Olaewe, Dr. Albert Lugutera and Mr. Richard Annang, I say thank you very much.

I appreciate immense contributions of Benjamin Nana Odum and Kugoriba James Baba Anekwoa in my quest to fulfill my academic dreams.

Also, I want to extend my sincere gratitude to officials of Ghana Police Service, Motor Transport and Traffic Unit, Wa in the Upper West Region of Ghana for releasing the data for this research work. May the Almighty God bless you all.

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

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

DEDICATION

This thesis is dedicated to my mother, Beatrice Hommey and the entire Wordi family.

TABLE OF CONTENTS

CHAPTER THREE: RESEARCH METHODOLOGY

CHAPTER FOUR: ANALYSIS AND DISCUSSIONS

CHAPTER FIVE: CONCLUSION AND RECOMMENDATIONS

APPENDICES

LIST OF TABLES

LIST OF FIGURES

ACRONYMS

ARIMA Autoregressive Integrated Moving Average

PACF Partial Autocorrelation Function

ACF Autocorrelation Function

KPSS Kwiatkowski- Phillips- Schmidt-Shin

BAC Blood Alcohol Concentration

TRL Transport Research Laboratory

WHO World Health Organization

CSA Case Study Approach

LDCs Less Developed Countries

GNP Gross National Product

GDP Gross Domestic Product

OECD Organisation for Economic Cooperation and Development

RTIs Road Traffic Injuries

GNA Ghana News Agency

AIC Akaike Information Criterion

MTTU Motor Transport and Traffic Unit

USA United States of America

UK United Kingdom

NRSC National Road Safety Commission

GRSP Global Road Safety Partnership

YLL Years of Life Lost

DALYs Disability Adjusted Life Years

BPD Business Partners for Development

DFID Department for International Development

BAC Blood Alcohol Concentration

DVLA Driver and Vehicle Licensing Authority

GPRTU Ghana Private Road Transport Union

ADF Augmented Dicky-Fuller Test

VIO Vehicle Inspection Officers

CHAPTER ONE

INTRODUCTION

1.1 Background of the Study

Road traffic accident in Ghana has become one of the growing concerns to most citizens in recent times. The British Medical Journal of 11th May 2002 indicated that more people die through road traffic accidents than from malaria worldwide; and that traffic accidents cause about 1.2 million deaths and injured 10 to 15 million people every year. Many researchers have come out with the causes, effects and recommendations to road traffic accidents. These cause include drink driving, machine failure and over speeding, Fosser, et al (1997). The mere increase in the number of accidents is not enough for one to conclude that there is an increase in the number of deaths. Hence there is need to analyze the accidents data statistically to determine whether there is any evidence of increasing road accidents which results to large number of people losing their lives.

Statistical projections show that between 2000 and 2020, the number of deaths related to traffic accidents will decrease with about 30% in high income countries. The reverse is expected in developing countries, where traffic accidents are expected to increase at a fast rate in the years to come. Road accident is defined as any activity which distracts the normal trajectory of a moving vehicle, in a manner that causes instability in the free flow of the vehicle. The accident is that the vehicle(s) involved veers off the road, collide, run over, vehicle on fire, etc. Road accidents in Ghana have been identified as

CHAPTER ONE

INTRODUCTION

1.1 Background of the Study

Road traffic accident in Ghana has become one of the growing concerns to most citizens in recent times. The British Medical Journal of 11th May 2002 indicated that more people die through road traffic accidents than from malaria worldwide; and that traffic accidents cause about 1.2 million deaths and injured 10 to 15 million people every year. Many researchers have come out with the causes, effects and recommendations to road traffic accidents. These cause include drink driving, machine failure and over speeding, Fosser, et al (1997). The mere increase in the number of accidents is not enough for one to conclude that there is an increase in the number of deaths. Hence there is need to analyze the accidents data statistically to determine whether there is any evidence of increasing road accidents which results to large number of people losing their lives.

Statistical projections show that between 2000 and 2020, the number of deaths related to traffic accidents will decrease with about 30% in high income countries. The reverse is expected in developing countries, where traffic accidents are expected to increase at a fast rate in the years to come. Road accident is defined as any activity which distracts the normal trajectory of a moving vehicle, in a manner that causes instability in the free flow of the vehicle. The accident is that the vehicle(s) involved veers off the road, collide, run over, vehicle on fire, etc. Road accidents in Ghana have been identified as one of the major causes of deaths in the country. It is classified as the second major cause of death in the country following malaria. According to the Building and Road Research Institute, there were 12,299 road accidents for the year 2009, Afukaar et al (2009). A total of 8,476 casualties were recorded with 2,237 of them losing their lives, while 6,242 sustained injuries. This revealed that there was an average of 6 deaths everyday in Ghana which was caused by road accidents. The National Road Safety Commission describes casualties as persons killed, seriously injured or slightly injured. The word casualty in this thesis shall be referred to any person who is killed through road traffic accident.

The most dangerous part is that most of the people who are killed by road accidents are those in the age group that constitute the work force of this nation. It is in this regard that, more attention needs to be placed on the research in to analysing the number of deaths through road traffic accidents in the Upper West Region of Ghana. According to Harrison and Berry (2008), it is estimated that about 100 million families worldwide are trying to cope with the death or disability of a family member involved in a road crash. The impact in terms of emotional and financial stress is enormous. Poverty, depression, physical illness and suicide are common consequences. Apart from the direct physical and psychological effects of injury on victims of road crashes, there are substantial impacts on their families, friends and on the community in general. The fear of traffic accidents can lead to reduced social interaction and cohesion as people remain indoors. In many countries, it has resulted in more sedentary lifestyles, with consequent health effects such as obesity and cardiovascular diseases.

The improvement in awareness is reflected in the recent establishment of the Global Road Safety Partnership (GRSP) by the World Bank, the International Federation of the Red Cross and Red Crescent Societies, bilateral aid agencies and other interested parties under the framework of the World Bank's Business Partners for Development (BPD) Programme. A steering committee for GRSP is now in place with the aim of creating a global information network that aims to produce solid evidence of the positive impact of partnerships – both the development impact and the business benefits. Two important aspects of GRSP are the involvement of the private sector in funding road safety projects and the promotion of greater awareness of road safety worldwide.

The introduction of the motor-vehicle into Africa is the single most important factor for change seen in Africa in the twentieth century. However, while Africa's share is only a small percentage of the world's vehicle population, the number of fatalities caused by road traffic is huge. According to Organization for Economic Cooperation and Development (OECD), a motor vehicle is over a hundred times more likely to be involved in a fatal road crash in an Africa country than in the UK or USA, Yoshitsugu and Hirakazu (2011)

Developing countries in Sub-Saharan Africa have the highest frequency of various accidents worldwide Peden et al. (2004). Although the implication of this is that the risk environment in countries needs further empirical attention, few studies have investigated how people in those societies perceive risk. This scenario calls for developing countries to put more effort toward the control and prevention of road traffic accident and their consequences. This can be achieved through multidisciplinary approach and research.

Such examination should be undertaken, because traffic accidents have negative impacts on social and economical improvements in developing countries. In this problem there are many agents:

i. The police who are interested in legal enforcement
ii. The insurance companies and vehicle owners in the monetary cost of road accidents
iii. The accident victims and their relatives in those disability and related cost of medical care
iv. The health care system and medical personnel who are responsible for emergency treatment and life savings of accident victims, Asogwa (1992).

Crash records indicate that in Ghana, over 1,600 people die annually in road traffic crashes and that more than 40 percent of the road traffic fatalities are occupants of cars, buses, and trucks (National Road Safety Commission, 2010). Most of these occupants would not have died if they had worn seat belts. This is because vehicle occupants who do not wear seat belts are more likely to be injured or killed in a crash than seat belt wearers (Dee, 1998; Reinfurt et al, 1996; Stewart, 1993).

Statistics on road accidents in the Upper West Region of Ghana as reported by (GNA, 2013), showed that the number of reported cases of road accidents had increased from 187 in 2011 to 196 in 2012, while the number of vehicles involved also recorded an increase from 276 in 2011 to 287 in 2012. In 2011, 60 of all vehicles involved in road accidents were commercial while 90 were private and 126 were motor bikes whereas in 2012, 52 were commercial, 97 private and 138 were motor bikes. These resulted in 82 deaths in 2012 as against 65 deaths in 2011. Indeed, accidents caused by motor bikes have since 2005 been on the ascendancy in the Wa Municipality, with the number of deaths rising from 9 in 2007 to 26 deaths in 2012.

These accidents were largely attributed to the lack of respect for road traffic regulations, lack of defensive driving, inexperienced and impatient driving, overloading and over speeding, drunkenness as well as refusal to obey road traffic regulations on the part of both drivers and motor bike riders.

Road traffic accident has always been attributed to human errors such as high alcohol content in the blood stream of the driver, over speeding, wrong overtaken among others. It has also been linked to poor road network, poor surfacing of the roads, witchcraft and the death –dying nature of some of the vehicles which ply the roads. There are numerous suggested solutions, various interventions by government, non-governmental organizations and other stakeholders to curtail the number of death through road traffic accidents, it could be possible that the factors such as type and nature of the road, age and sex contribute to the number of deaths through road traffic accidents.It is therefore very important to statistically analyse the accident data to ascertain the truth. When it is confirmed, then it will be prudent to apply some statistical model such as time series analysis to fit a model to the accident data for better prediction for decision making.

1.2 Problem Statement

Beside the first and second world wars (1914-1918), (1939-1945), road traffic accident is the next consumer of human lives. Its victims cut across the ages and gender. According to the National Road Safety Commission (2012), a total of 12,008 accident cases were reported nationwide, involving 16,132 vehicles with a total number of 10,024 people injured whilst a total number of 12,333 people lost their lives. Apart from the monumental damages of road accidents on human lives, it constitutes the cog in the wheel of economic progress of the country as well as educational well- being of the nation. It is so disturbing that millions of lives are painfully taken away yearly, apart from psychological trauma caused the loved ones who lose their next- of – kins. Therefore, this research is investigating statistical analysis of number of deaths through road traffic accident in the Upper West Region of Ghana.

1.3 Research Questions

In order to achieve the purpose of the research, the following guiding questions are devised;

i) What kind of motor vehicle casualties and non motorised casualties occurred in the Upper West Region in the period of 2008 to 2012?

ii) which factors are assessed as associated with the causes of traffic accidents in the Region?

iii) what kind of safety measures have been taken or implemented by local authorities to prevent road traffic accidents in the Upper West Region?

1.4 Objectives of the study

Main Objectives

To analyse the number of deaths in the Upper West Region through road traffic accidents.

Specific Objectives

The specific objectives of the study are as follows;

i) To determine whether the number of deaths has seasonal variation.

ii) To determine the trend of the deaths in the Upper West Region.

iii) To develop a model for the number of deaths through road traffic accidents in the Upper West Region.

iv) To forecast the future trend of the number of deaths

1.5 Significance of the Study

It is important to research in to the road traffic accident data in the Upper West Region of Ghana to come out with the reality on the ground so that, lgthe Vehicle inspection Officers (VIO) will be able to use the model and outcomes of the study to increase their inspections on road worthiness of many vehicles. Also, the Motor Traffic and Transport Unit (MTTU) of the Ghana Police Service (GPS) will use the information in this study to curb excessive speeding on our roads.

Furthermore, the judicial division will make use of recommendation in the study in imposing sanctions/punishment on the erred reckless and careless drivers. Drivers and License Authority (DVLA) will use the information in this research to conduct objective examination to drivers whenever they want to renew drivers' licence and to create a platform for future studies into the analysis of the number of deaths through road traffic accidents in the Upper West Region and any of the regions in Ghana.

CHAPTER TWO

LITERATURE REVIEW

2.0 Introduction

This chapter reviews the related literature. It will assess the existing mathematical and statistical models to project in to the future using Time series analysis. This will guide the direction of this study and aid in exploring the unknown.

2.1 The Root Causes of Road Traffic Accidents

The most common known causes of road traffic accidents in Ghana include indiscipline on the roads, over-loading, fatigue driving, drunk driving and over-speeding. Statistics showed that 60% of road accidents are caused by drunk driving and over-speeding. The latter alone constitutes about 50% of road accidents in the country. The poor nature of some of the roads, poor maintenance of vehicles, disregard for traffic regulations by most drivers and indiscriminate use of the road by some pedestrians are some of the other causes of motor accidents in the country (Feature Article, 2011).

2.2 The State of Road Accident in Ghana

Road traffic accidents is one of the growing concerns to most Ghanaians in recent times. A research conducted by Salim and Salimah (2005) indicated that road accident was the major cause of death in low middle income countries and predicted that road accidents was going to be the third major cause of deaths by 2020 if the trend of road traffic accidents continued.

Media reports revealed that there are high road accidents in Ghana, compared with other developing countries. In 2001, Ghana was ranked the second highest road accident-prone nation among six West African countries with 73 deaths per 1000 accidents, (Akongbota, 2011). The Ghanaian Times news reported on the 16th day of November, 2011 that a total of 1,986 lives were lost in the country through road traffic accidents from January to October, 2011, The Ghanaian Times.

Heidi (2006) reported that 1.2 million people in the world lose their lives through road accidents every year. The number has rising to 1.3 million globally every year and between 20 and 50 million sustaine various forms of injuries as a result of road accidents. The most affected of age bracket is 15 and 29. Road traffic accidents cost the world an amount of USS518 billion annually. It is estimated that if nothing is done globally to curtail the rampant nature of road accidents and most especially the causes of deaths of casualties before they are sent to hospitals then by the year 2020, 1.9 million people will be killed by road accidents in the world, (World Health Organisation, 2011).

2.3.1 Pedestrian Alcohol Consumption

Despite alcohol having an obvious effect on pedestrian casualties, comparatively, little research has been conducted on the issue, possibly because of difficulties involved in modifying legislation and behaviours of pedestrians. National data for Great Britain showed that the incidence of alcohol amongst fatally injured adult pedestrians was increasing: 46% of fatally injured pedestrians had blood alcohol concentration (BAC) in excess of 9mg/100ml in 1997 compared with 39% a decade earlier (DETR, 1999).There is evidence to suggest that alcohol has an effect on pedestrian collisions, though some of the research found on this subject dates back to forty years with relatively few studies conducted in the last decade. Despite the research being more than a decade ago, the problem is likely to have only got more widespread as the binge drinking culture in Britain has got worse in recent years (National Centre for Social Research, 2011).

2.3.2 Driver Alcohol Consumption

Impairment by alcohol is an important factor influencing both the risk of a road crash as well as the severity of the injuries that result from crashes. The frequency of drinking and driving varies between countries but it is almost universally a major risk factor for road traffic accidents. In many high-income countries, about 20% of drivers who died through road accident have excess alcohol in their blood. Studies in low-income countries have shown alcohol to be present in between 33% and 69% of drivers who died in road traffic accidents.

2.3.3 Driver's Attitude

Jorgenson and Abane (1999) noted that, one can distinguish between driving skills (Knowledge and training) and driving style which reflected attitudes and traffic risk perception.Training of drivers increases their skills. A Study done by Asogwa in Nigeria revealed that a sizeable proportion of drivers who possesses driving licenses never showed up in any driving school or went through a driving test but simply bought their licenses. Untrained drivers, not unexpectedly, often result in high accident rates (Asogwa, 1992a). In emergency conditions, stopping distance is also important. However, this depends very much on the driver's reaction time, speed of the vehicles, quality of tires, and the condition of the road (Lemming, 1969).

Studies done on drivers after being involved in motor accidents reported that although alcohol is the most prevalent source of driver's impairment, other drugs or substance abuse can also contribute to the problem (Violent et al, 1996; Kayombo, 1995; Broughton, 1991; Leon, 1996; Shibata and Fukubu ,1994). Driving under the influence of alcohol or other drug abuse is known to impair the driver's ability to judge and control the vehicle (Orsay et al, 1994). Furthermore, fatigue- related crashes occur more frequently on weekends than weekdays and typically occur in early morning (Asogwa, 1978).

Some medical conditions are also mentioned to be risk factors for driving (Hayes, 1972). For example, diabetes and epilepsy have been identified as factors that are associated with increased risk if a person is allowed to drive (Odero et al., 1997, Redelmeier and Tibshirani (1997) and Lave et al., 1993). Violant and Marshall (1996) reported that the frequency of road accidents involving epileptics and diabetics is double the normal and for a heart patient it is 60% higher. Fatigue or sleepiness is associated with a range of factors. Some of these factors are long-distance driving and sleep deprivation. Three high risk groups have been identified:

i. Young people, particularly males, aged 16–29 years;

ii. Shift workers whose sleep is disrupted by working at night or working long, irregular hours;

iii. People with untreated sleep apnea syndrome or narcolepsy.

Factors that substantially increase the risk of a fatal crash or a crash with serious injuries are:

i. Driving while feeling sleepy;

ii. Driving after five hours of sleep;

iii. Driving between 02:00am and 05:00am

2.3.4 Age for Safe Driving

The driver's age is known to be an important factor contributing to road accidents. Available literature shows that adolescents or young drivers are frequently involved in traffic accidents than other age groups, (Leon,1996) have shown that young drivers are more frequently involved in accidents caused by inappropriate speed and loss of control of the vehicle compared to other age group of drivers. The study by Graham (1993) reported that motor accidents were prevalent in certain age group and they occurred at certain hours of the day and week and at certain locations. Some people are more susceptible than others and susceptibility is increased by the use of alcohol and other drugs as well as other physiologic states such as fatigue (Graham, 1993).

Massie et al, (1995) in their study have also reported that old drivers (70 years and over) have the highest rate of fatal accident involvement while young drivers have highest rates of injurious involvement.

2.3.5 Human Behaviour Approaches

A recent European study observed that 80% of drivers involved in motor vehicle accidents believed that the other party could have done something to prevent the accident. A miniscule 5% admitted that they were the only one at fault. Surveys consistently revealed that the majority considered themselves more skillful and safer than the average driver. Some mistakes occur when a driver becomes distracted, perhaps by a cell phone call or a spilled cup of coffee. Very few accidents result from an 'Act of God,' like a tree falling on a vehicle (Mercy, 1993).

Driving is a complex system in which a large number of variables are interacting with each other but also with varying degree of dependence. Accidents may occur due to judgment errors, ignorance, incompetence, rule violation, lapses or carelessness, all of which are human errors (Lemming, 1969). The human factor or arrow contributes to the majority of road traffic accidents. A study done in Kenya reported that human factors were responsible for 85% of all causes (Odero, 1995).

2.4 Worldwide Trends in Road Traffic Death

In today's society, travelling by car is a common and a necessary activity (Cunill, 2004). The benefits of motor vehicles are well known and include: rapid movement of people and merchandise, convenience, relatively low cost, etc. However, motor vehicle use also has a number of undesirable consequences, such as pollution, noise and traffic accidents (Cunill et al., 2004). Traffic accidents are one of the main public health problems facing modern society (Parada et al., 2001; Cunill et al., 2004).

2.5 Road Traffic Accident Situation in Ghana

The escalating incidence of road accidents in Ghana is no news to a reasonable Ghanaian of ordinary intelligence. Despite increased road safety campaigns, the rate at which accidents occur on our roads is very alarming. It is a truism that one of the major challenges that this country is still battling with is motor accidents. Indeed, this conquerable foe is devouring our human and economic resources. Precious lives are lost thereby dwindling our scarce labour force in the country, causing a stir in our Gross Domestic Product (GDP). Continual media reports reveal that Ghana's road accident is high among developing countries.

According to the officials of Motor Traffic and Transport Unit of the police service during this period, there were 78 deaths, 373 serious injury cases and 966 minor accidents among others. The figure of road accidents in the country sharply went up to 12, 164 in 2004 but decreased to 11, 305 in 2005. Ghana records about 10,000 fatal road traffic accidents every year, out of which up to 1,600 people perish while 15,000 are seriously injured, such persons may die or become incapacitated, denying them the ability to contribute to the nation's development meaningfully. Besides, Ghana loses an amount of GH¢165, 000, representing 1.6% of GDP in solving road accident situations such as medical expenses of victims, damage to vehicles and insurance cost among others. In all, Ghana loses about 2% of her GDP annually due to road accidents (Chronicle, Aug 2008).

2.6 Solving Road Accidents in Ghana

"Road accidents cost the nation US$288 million in 2008". This was the chilling story as told by the Chairman of the Board of Directors of the Driver and Vehicle Licensing Authority (DVLA) in Ghana and reported by Chronicle newspaper in Ghana recently (Feature Article, Sep, 2011). This amount did not include the direct and indirect cost of road accidents to relatives of victims – wives, husbands, mothers, fathers, children, etc - such as funeral costs and the cost of having to do without loved ones and breadwinners. Through road accidents, many children and other dependants of victims have become destitute. Companies, businesses, state institutions (departments, hospitals, schools and colleges, universities etc) have lost many experienced and competent employees through road accidents. What makes Ghana's situation worse is that majority of the people do not have life and disability insurance policies from which they or their dependants can get some financial support when they are involved in accidents.

The above distressing situation calls for drastic intervention by the government and its agencies like the DVLA and the Police and indeed, all citizens to play their role in ensuring drastic reduction in road accidents in Ghana. It is very heartening that the DVLA is at long last taking the initiatives to address the problem – undoubtedly one of the biggest concerns and fears of every Ghanaian. Let's hope that the new initiatives will make our roads safer. Much has been written already in this forum about road accidents in Ghana and how to curb it. This article is intended to reinforce the importance of the subject matter and offer some suggestions on long term solutions to road accidents. The first major cause of road accidents in Ghana is poor driving skills. Other causes are the normal ones we are aware of:

i. Drivers with poor eye sights,
ii. Driver fatigue (dozing behind the steering wheel),
iii. Drunkenness,
iv. Over-speeding,
v. Defective vehicles
vi. Overloading,
vii. Poor roads,
viii. Non-existent road markings and signs

2.7 Effects of Gender on Road Accidents

It has been found that males compared to females have a higher risk of experiencing fatal crashes, while women have higher rates of involvement in injury crashes (Massie et al, 1995). Rivara and Bender (1985) have also reported that among the drivers of motor vehicles that struck victims, 69% of them were males and 31% females, controlled for gender exposure level. It appears that males are more at risk than females for all age groups. Odero at al, (1997) found out that in developing countries men are more at risk than women of being injured in crashes. The preponderance of males may be attributed due to their greater exposure to traffic and other associated factors. Concerning drivers, the relevance of gender to road safety has long been recognised and it is the contribution of male drivers to accidents which has attracted much attention (Dopson et al., 1999). This is because driving as a profession is mostly dominated by men.

2.8 Non-use of Crash Helmets

The main risk factor for motorized two-wheeler users is the non-use of crash helmets. The lack or inappropriate use of helmets has been shown to increase the risk of fatalities and injuries resulting from road crashes involving motorized two wheelers. Head injuries are a major cause of death, injury and disability among users of motorized two wheel vehicles. Many of these head injuries could have been prevented or their severity reduced through the use of simple and inexpensive helmets (Krug, 2000).

CHAPTER THREE

METHODOLOGY

3.0 Introduction

This chapter deals with research design, population of study, data collection procedure, modeling procedure and statistical analysis.

3.1 Target population

The target population for this study comprised of all road traffic victims of the Upper West Region of Ghana. However, because of time and cost implications, the sample is limited to death cases through road traffic accidents between 2008 and 2012 in Wa. Altogether, 251 of death cases constituted the sample for this study.

3.2 Source of Data and Data Analysis

The data were mainly secondary from the Upper West Regional Police Station; Motor Transport and Traffic Unit covering a period of five (5) years from 2008 to 2012. MINITAB statistical package and R-Consol will be used for the analysis. Time series will be used to determine the seasonal effects on death and model the death rate of road traffic accidents in the region.

3.3 Time Series

A time series is a sequence of observations which are ordered in time (or space). If observations are made on some phenomenon throughout time, it is most sensible to display the data in the order in which they arose, particularly since successive observations will probably be dependent. Time series are best displayed in a scatter plot. The series value X is plotted on the vertical axis and time t on the horizontal axis. Time is called the independent variable in this case however, something over which you have little control.

3.4 Modeling Procedure

The models that will be used to fit the data are explained in this section.

3.4.1 Simple Linear Regression

The first method used to analyze the data was a simple linear regression model, with number of injury cases as the response variable versus months as the predictor variable. The data were plotted in a scatter diagram and it was tentatively decided that there was an approximate linear association between the two variables.

The regression model is of the form:

$$Y=\beta_o + \beta_i X + \varepsilon_i \quad i= 1, 2 \qquad (3.1)$$

where

β_0 =y-intercept

β_i =slope of the line

ε_i =error term

X= Number of deaths.

3.4.2 Autoregressive Model

In statistics and signal processing, an autoregressive (AR) model is a type of random process which is often used to model and predict various types of natural phenomena. The autoregressive model is one of a group of linear prediction formulas that attempt to predict an output of a system based on the previous outputs.

The notation AR(p) indicates an autoregressive model of order p. The AR(p) model is defined as

$$X_t = C + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t \qquad (3.2)$$

where $\varphi_1 \ldots\ldots \varphi_p$ are the parameters of the model, C is a constant (often omitted for simplicity) and ε_t is white noise.

3.4.3 Autoregressive Integrated Moving Average

In statistics and econometrics, and in particular in time series analysis, an autoregressive integrated moving average (ARIMA) model is a generalization of an autoregressive moving average (ARMA) model. These models are fitted to time series data either to better understand the data or to predict future points in the series (forecasting). They are applied in some cases where data show evidence of non-stationarity, where an initial differencing step (corresponding to the "integrated" part of the model) can be applied to remove the non-stationarity.

The model is generally referred to as an ARIMA (p,d,q) model where p, d, and q are non-negative integers that refer to the order of the autoregressive, integrated, and moving average parts of the model respectively. ARIMA models form an important part of the Box-Jenkins approach to time-series modeling.

Given a time series of data X_t where t is an integer index and the X_t are real numbers, then an ARMA(p,q) model is given by:

$$\left(1-\sum_{i=1}^{p}\alpha_i B^i\right)Xt=\left(1+\sum_{i=1}^{q}\theta_i B^i\right)\varepsilon_t \qquad (3.3)$$

Where B is the lag operator, the α_I are the parameters of the autoregressive part of the model, the θi are the parameters of the moving average part and the ε_t are error terms. The error terms ε_t are generally assumed to be independent, identically distributed variables sampled from a normal distribution with zero mean.

Assume now that the polynomial $(1-\sum_{i=1}^{p}\alpha_i B^i)$ has a unitary root of multiplicity. Then it can be rewritten as:

$$(1-\sum_{i=1}^{p}\alpha_i B^i)=(1+\sum_{i=1}^{p-d}\varphi_i B^i)(1-B)^d \quad (3.4)$$

An ARIMA(*p*,*d*,*q*) process expresses this polynomial factorisation property, and is given by:

$$(1-\sum_{i=1}^{p}\alpha_i B^i)(1-B)^d X_t=(1+\sum_{i=1}^{q}\theta_i B^i)\,\varepsilon_t \quad (3.5)$$

and thus can be thought as a particular case of an ARMA(p+d,q) process having the auto-regressive polynomial with some roots in the unity. For this reason every ARIMA model with d>0 is not wide sense stationary.

Seasonal AutoRegressive Integrated Moving Average (SARIMA) model which is a generalization of the well known Box-Jenkins ARIMA model is to accommodate a data with both seasonal and non-seasonal features. The ARIMA model which is known to be a combination of the AutoRegressive (AR) and Moving Average (MA) models utilize past information of a given series in order to predict the future. The AR part of the model deals with the past observation of the series whiles the MA part deals with the past error of the series (see Hamilton, 1994; Pankratz, 1983).

The ARIMA model is applied in the case where the series has no seasonal features and also differenced stationary. This means that an initial differencing is required for the data to be stationary. The ARIMA model with its order is usually presented as ARIMA (p, d, q) model where p, d, and q are integers greater than or equal to zero and refer to the order of the autoregressive, integrated, and moving average parts of the model respectively. The first parameter prefers to the number of autoregressive lags, the second parameter drefers to the order of integration that makes the data stationary, and the third parameter qgives the number of moving average lags (Pankratz, 1983; Hurvich and Tsai, 1989; Hamilton, 1994; Kirchgässner and Wolters, 2007; Kleiber and Zeileis, 2008; Pfaff, 2008)

A process, $\{\varepsilon_t\}y_t$ is said to be ARIMA (p, d, q) if $\Delta^d y_t$ is described by a stationary ARMA(p, q) model.Δ Means differencing of y_t inorder to achieve stationarity. In general, we will write the ARIMA model as

$$\phi(B)(1-B)^d Y_t \theta(B)\varepsilon_t; \{\varepsilon_t\} \sim \mathrm{WN}(0, \delta^2 \qquad (3.6)$$

Where ε_t follows a white noise (WN) process. The autoregressive operator and moving average operator are defined as follows:

$$\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \ldots - \varphi_p B^p \qquad (3.7)$$

$$\theta(B) = 1 + \theta_1 B + \theta_2 B + \ldots. + \theta_q B^q \qquad (3.8)$$

φ(B)≠0 for ∅<1, the process {ε_t} y_t is stationary if and only if d=0, in which case it reduces to an ARMA(p, q) process.

The generalization of ARIMA model to the SARIMA model occurs when the series contains both seasonal and non-seasonal behavior. This behavior of the series makes the ARIMA model inefficient to be applied to the series. This is because it may not be able to capture the behavior along the seasonal part of the series and therefore mislead to a wrong order selection for non-seasonal component. The SARIMA model is sometimes called the multiplicative seasonal autoregressive integrated moving average model and is denoted by ARIMA (p, d, q) (P, D, Q) $_S$. This can be written in its lag form as (Halim & Bisono, 2008):

$$(B)\Phi(B^S)(1-B)^d(1-B^S)^D y_t = \theta(B)\Phi(B^S)\varepsilon_t \qquad (3.9)$$

$$\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p \qquad (3.10)$$

$$\Phi B^s = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_p B^{ps} \qquad (3.11)$$

$$\theta(B) = 1 + \theta_1 B + \theta_2 B + \dots + \theta_q B^q \qquad (3.12)$$

$$\Theta B^S = 1 - \Theta_1 B^S - \Theta_2 B^{2S} - \dots - \Theta_q B^{QS} \qquad (3.13)$$

Where p , d and q represents the order of non-seasonal AR, differencing and MA respectively. P, D and Q is the order of seasonal AR, differencing and MA respectively.

Y_t represents observable time series data at period t.

ε_t represents white noise1 error (random shock) at period t.

B represents backward shift operator ($B^k y_t = y_{t-k}$)

S represents seasonal order (e.g. s =4 for quarterly data and s =12 for monthly data).

3.4.4 Model Identification and Model Comparisons

In the identification stage of model building steps, we determine the possible ARIMA models that best fit the data under consideration. But before the search of the possible model for the data, the data under consideration must satisfy the condition of stationarity.

This is because the ARIMA model is appropriate for stationary time series data (i.e. the mean, variance, and autocorrelation are constant through time). If a time series is stationary then the mean of any major subset of the series does not differ significantly from the mean of any other major subset of the series. Also if a data series is stationary then the variance of any major subset of the series will differ from the variance of any other major subset only by chance ((Pankratz,1983). The stationarity condition ensures that the properties of the estimated parameters from the model are standard. That is the t statistic will asymptotically follow the usual t distribution. If this condition is assured then, the estimated model can be used for forecasting (Hamilton,1994). When the stationarity condition of the data is satisfied, the possible models suitable for the data can now be determined. The order of the model which AR, MA, SAR and SMA terms can be determined with the help of the ACF and the PACF plot of the stationary series. The ACF and PACF give more information about the behavior of the time series. The ACF gives information about the internal correlation between observations in a time series at different distances apart, usually expressed as a function of the time lag between observations. These two plots suggest the model we should build.

Checking the ACF and PACF plots, we should look at both the seasonal and non-seasonal lags. Usually the ACF and the PACF has spikes at lag k and cuts off after lag k at the non-seasonal level. Also the ACF and the PACF has spikes at lag k and cuts off after lag k at the seasonal level. The number of significant spikes suggests the order of the model. (Shumway and Stoffer, 2006). The ACF and PACF plot suggest the possible models that can be obtained for the data but it does not give the final model for the data. This means that for a given series, several possible models can be obtained. In other to select the best model among the possible models, the penalty function statistics such as Akaike Information Criterion (AIC or AICc) or Bayesian Information Criterion (BIC) can be used (Akaike, 1974; and Schwarz 1978). The AIC, AICc and BIC are a measure of the goodness of fit of an estimated statistical model. Given a data set, several competing models may be ranked according to their AIC, AICc or BIC with the one having the lowest information criterion value being the best. These information criterion judges a model by how close its fitted values tend to be to the true values, in terms of a certain expected value. The information criterion value assigned to a model is only meant to rank competing models and tell you which one is the best among the given alternatives.

$$AIC = 2k - 2\log(L) \text{ or } 2k + n\log(RSS/n) \quad (3.14)$$

k = the number of parameters in the statistical model, (p+q+P+Q+1)

L = the maximized value of the likelihood function for the estimated model.

RSS = the residual sum of squares of the estimated model.

n= the number of observation, or equivalently, the sample size

and AICc is AIC with a correction for finite sample sizes:

$$AICc = AIC + 2k(k+1)/n-k-1 \quad (3.15)$$

Where *n* denotes the sample size.

Thus, AICc is AIC with a greater penalty for extra parameters.

Burnham & Anderson (2002) strongly recommend using AICc, rather than AIC, if *n* is small or *k* is large. Since AICc converges to AIC as *n* gets large, AICc generally should be employed regardless. Using AIC, instead of AICc, when *n* is not many times larger than k^2, increases the probability of selecting models that have too many parameters, i.e. of over-fitting. The probability of AIC over-fitting can be substantial in some cases.

Brockwell and Davis, advise using AICc as the primary criterion in selecting the orders of an ARMA model for time series. McQuarrie and Tsai ground their high opinion of AICc on extensive simulation work with regression and time series.

AICc was first proposed by Hurvich and Tsai (1989). Different derivations of it are given by Brockwell and Davis, Burnham and Anderson, and Cavanaugh. All the derivations assume a univariate linear model with normally-distributed errors (conditional upon regressors); if that assumption does not hold, then the formula for AICc will usually change.

3.5 Lag (backshift) Operator

In time series analysis, the lag operator or backshift operator operates on an element of a time series to produce the previous element. For example, given some time series $X=\{X_1, X_2 \ldots\}$ Then

$$BX_t = X_{t-1} \quad \text{for all } t > 1$$

or equivalently

$X_t = BX_t$ for all $t \geq 1$ where B is the lag operator which can be raised to arbitrary integer powers so thata

$$B^{-1}X_t = X_{t+1} \tag{3.16}$$

and

$$B^k X_t = X_t \tag{3.17}$$

3.6 Ljung-Box Test

LJUNG-BOX Test is performed to check for randomness. There are a large number of tests of randomness (e.g., the runs tests). Autocorrelation plots are one common method test for randomness. The Ljung-Box test is based on the autocorrelation plot. However, instead of testing randomness at each distinct lag, it tests the "overall" randomness based on a number of lags. For this reason, it is often referred to as a "portmanteau" test. In this study, the test is therefore used to test whether the number of deaths vary by season.

More formally, the Ljung-Box test can be defined as follows.

H_0: The data are random H_1: The data are not random

Test Statistic: the test statistic is Q_{LB}=(n(n+2)$\sum_{j=1}^{h} r^2$(j)/n-j (3.18)

Significance level: α

Where n is the sample size, r(j) is the autocorrelation at lag j, and h is the number of lags being tested.

CHAPTER FOUR

ANALYSIS AND DISCUSSIONS

4.0 Introduction

This chapter mainly presents analysis of the various results obtained from the study. Minitab software package was used to obtain all the estimated parameters reported in the study. The analysis were categorized into Preliminary analysis, further analysis which consists of Model identification, Estimation of parameters, Model diagnostic and evaluating the accuracy of the forecast.

4.1 Descriptive Statistics

Table 4.1 Descriptive statistics for death through road traffic accidents

Var	N	N*	Mean	CoefVar	Min	Max	Sk	Kurt
Deaths	60	0.0	4.533	66.30	0.000	13.000	0.74	-0.04

From table 4.1 the number of deaths recorded through road traffic accident has a total observation to be 60 from 2008 to 2012 with no missing cases. It has an average record of deaths to be 4.533 for the four years under study. The minimum and maximum per month of the deaths recorded through road traffic accidents are 0.00 and 13.00 respectively. Also from the table it can be clearly seen that the coefficient of variation is 66.30 which explains the total variability relative to the mean. The series also shows a 0.74 skewness meaning that the series is positively skewed from the normal distribution and platykurtic in nature.

Table 4.2 Monthly Descriptive Statistics of deaths through road traffic accident

Var	Months	Mean	C. V	Min	Max	SK	Kurt
Deaths	Jan.	4.4	87.43	0	9	-0.07	-2.28
	Feb.	2.4	69.72	1	5	1.09	0.54
	Mar.	4.8	81.22	1	11	1.22	1.25
	April	6.4	81.64	2	13	0.59	-2.75
	May	4.8	86.40	0	10	0.24	-1.96
	June	4.4	57.05	2	7	0.20	-3.03
	July	4.8	57.81	2	8	-0.01	-2.70
	August	4.0	79.06	1	9	1.19	1.05
	Sept	5.0	34.64	4	8	1.92	3.67
	Oct.	4.0	58.63	2	7	0.58	-2.63
	Nov.	3.8	68.12	1	8	1.23	2.40
	Dec.	5.6	32.44	4	8	0.57	-2.23

Table 4.2 shows the various monthly distributions recorded through road traffic accident under the period of study.

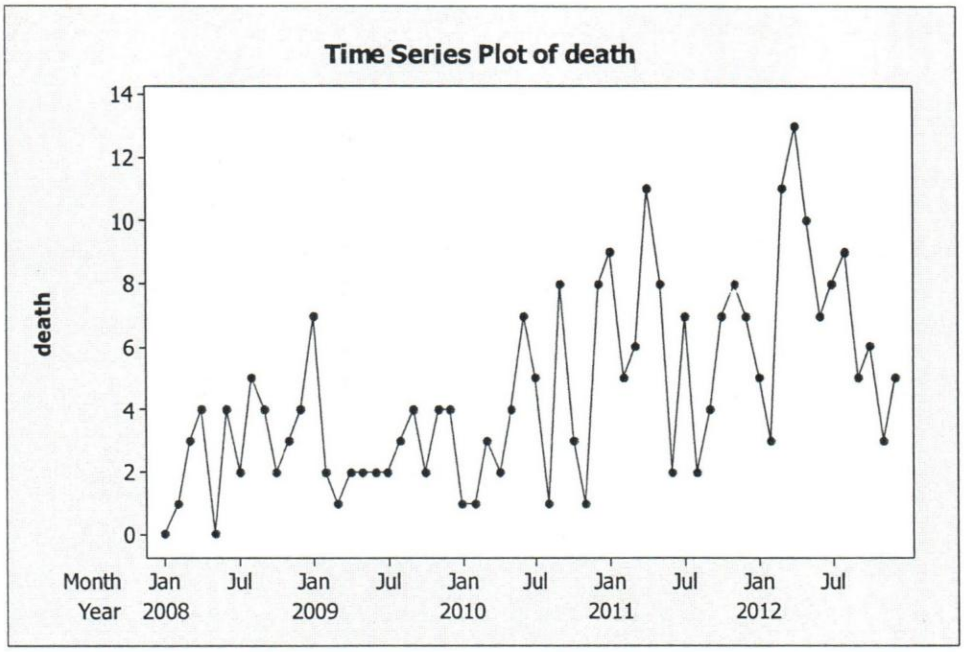


Figure 4.1: Time series plot for the Deaths

The time series plot shows the behavior of the data in an upward and downward trend from January 2008 to December 2012.

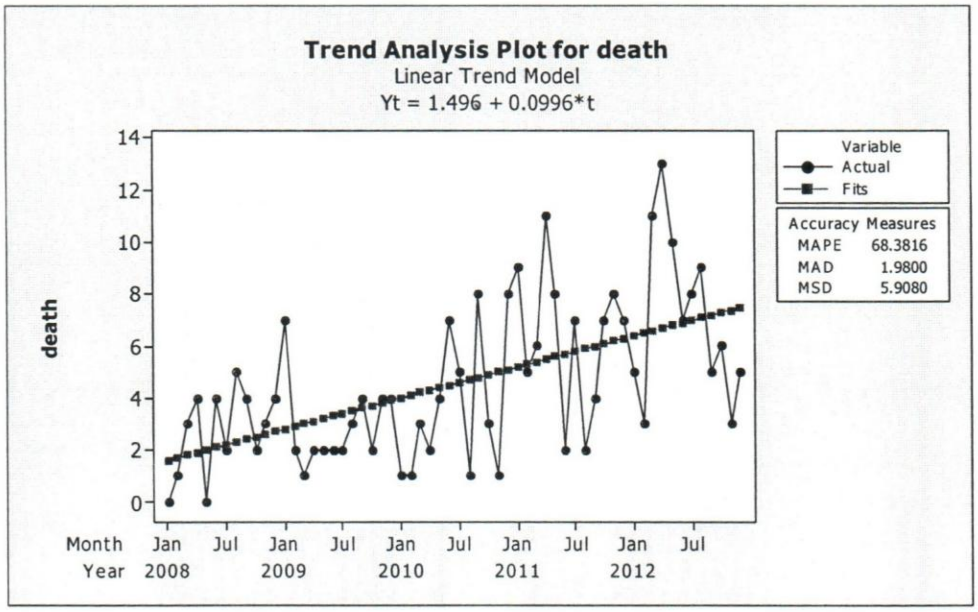


Figure 4.2: Plot of a linear trend model of the deaths

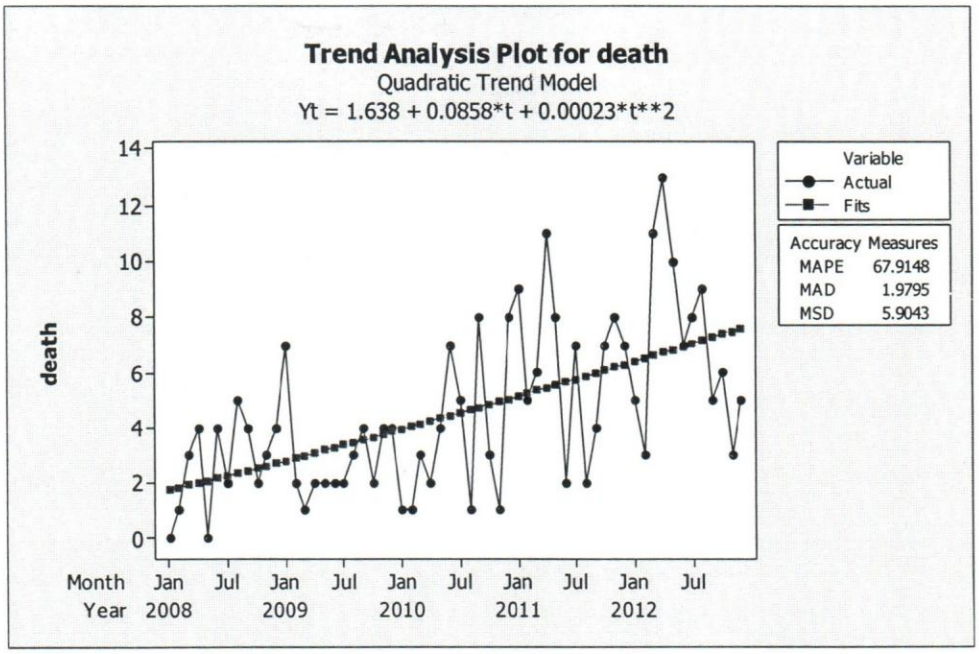


Figure 4.3: Plot of a quadratic trend model of the deaths.

Table 4.3 Trend Accuracy Measures

ACCURACY MEASURES			
MODEL	MAPE	MAD	MSD
LINEAR	68.3816	1.9800	5.9080
QUADRATIC	67.9148	1.9795	5.9043

Table 4.3 shows the accuracy measures of which quadratic trend model appear to have the least accuracy measures with a mean absolute percentage error (MAPE) of 67.9148, mean absolute deviation (MAD) of 1.9795 and a mean squared difference (MSD) of 5.9043. Therefore the data follows a quadratic trend model since it is the most excellent with the smallest measures of accuracy.

4.2 Test for stationarity

Hypothesis for KPSS Test

H_0 = The series is stationary.

H_1= The series is not stationary

Hypothesis for ADF Test

H_0 = The series is not stationary.

H_1= The series is stationary.

Table 4.4 Test for stationarity

TEST	T-Statistics	P-Value
KPSS	1.5077	0.01
ADF	-3.0723	0.1414

Table 4.4 shows the test for stationarity. Using Kwiatkowski-Phillips-Schmidt-Shin test revealed a p-value of 0.01 leading to a rejection of the null hypothesis that there is stationarity in the series, thus we fail to reject the alternative hypothesis that there is no stationarity in the series.

Also, Augmented Dickey-Fuller test revealed a test statistic of -3.0723 with a p-value of 0.1414 leading to a failure to reject the null hypothesis that there is no stationarity in the series.

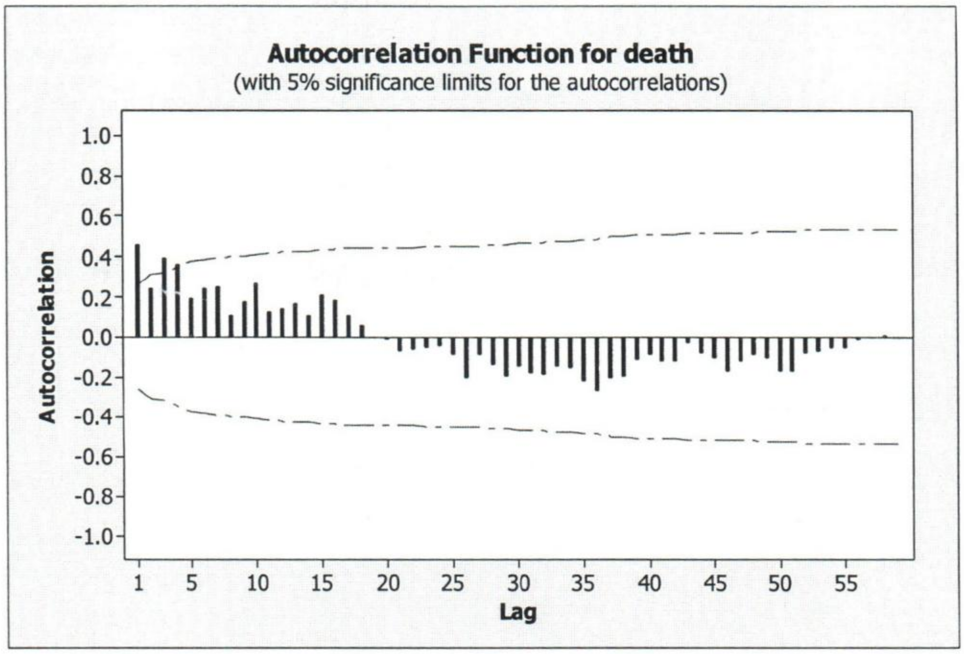


Figure 4.4 ACF plot for the deaths.

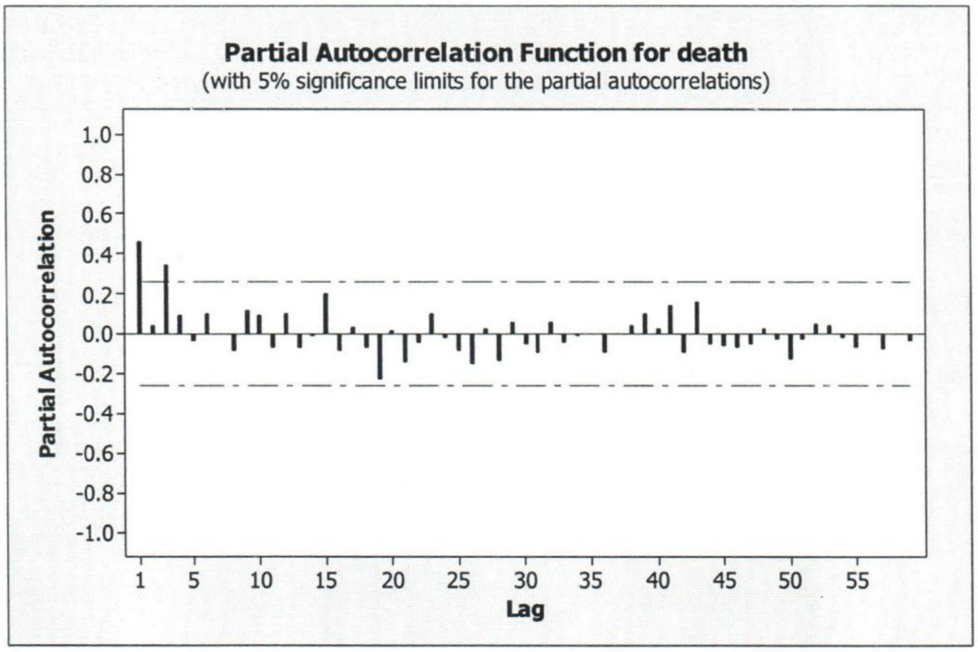


Figure 4.5 PACF plot for the deaths.

From figure 4.4, it can be seen that the ACF for the death data shows a large positive significant spike from lag 1 to lag 3, and the rest of the spikes decays rapidly in a wavelike form to zero line at lag 18. This shows that the data is not stationary with a pattern typical to a moving average (MA) process of order 3.

Also, from figure 4.5, it can be seen that the PACF for the deaths data shows a large positive significant spike at lag 1 and 2 and all the other spikes fluctuates within the 95% confidence limits. This pattern is typical to an autoregressive (AR) process of order 2.

Since the series is not stationary we therefore difference the series to attain stationarity.

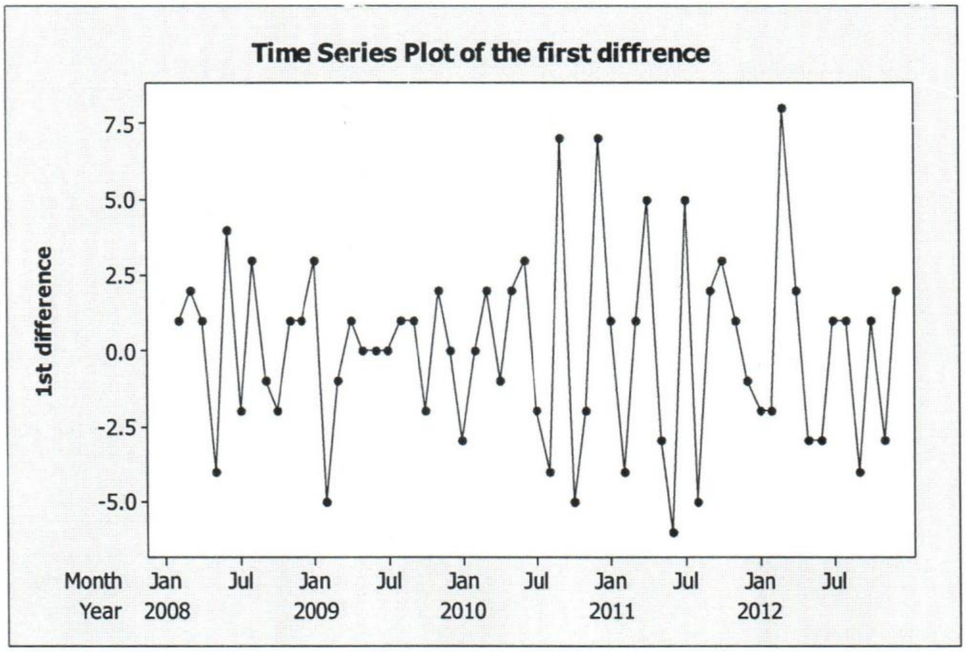


Figure 4.6 Time series plot after first difference

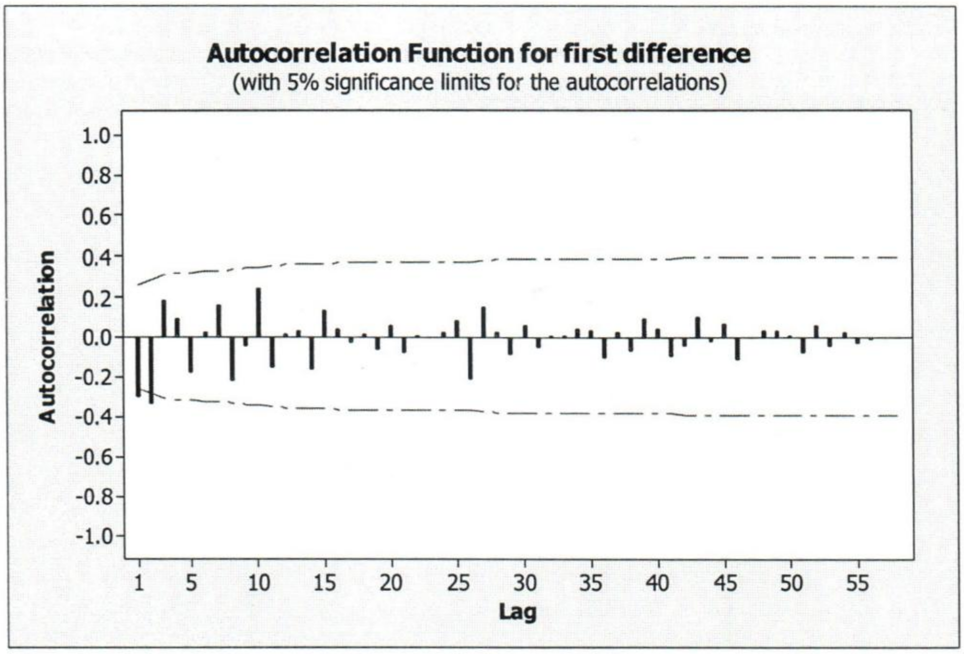


Figure 4.7 ACF graph for the first difference.

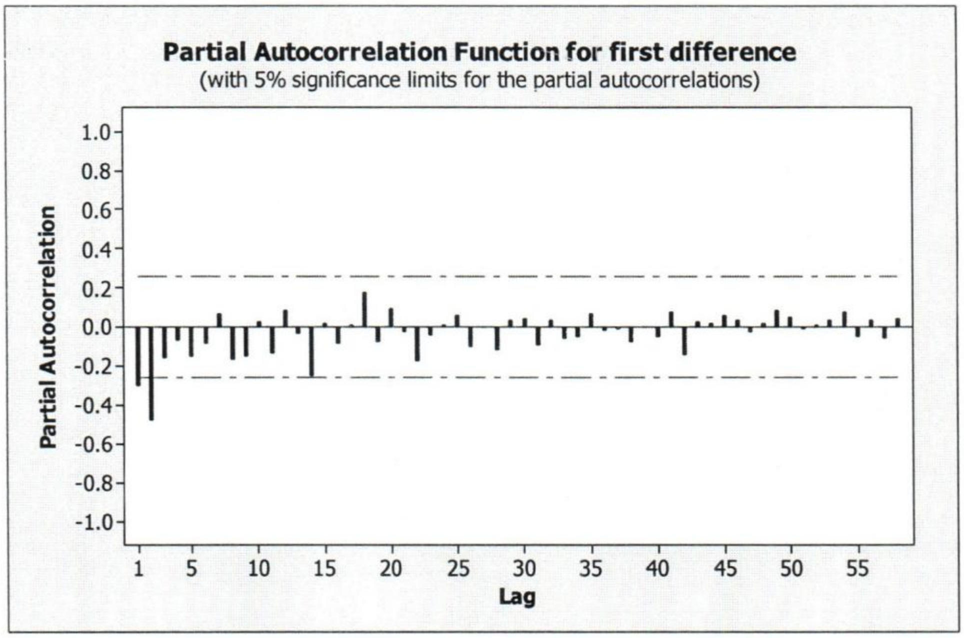


Figure 4.8 PACF graph for the first difference.

Hypothesis for KPSS Test

H_0 = The series is stationary.

H_1= The series is not stationary.

Hypothesis for ADF Test

H_0 = The series is not stationary.

H_1= The series is stationary.

Table 4.5 TESTS FOR STATIONARITY FOR FIRST DIFFERENCING

TESTS	T-Statistics	P-Value
KPSS	0.0265	0.1
ADF	-5.1331	0.01

From table 4.5,testing for stationarity after first differencing revealed that the series is now stationary with a KPSS test statistic of 0.0265 and ADF test statistic of -5.1331, Their respective p-value are 0.1 and 0.01 indicating that we fail to reject the null hypothesis for KPSS test that the series is stationary. We reject the null hypothesis for ADF test and conclude that the series is stationary with both tests.

Since the series is now stationary or the stationarity has been achieved we will therefore fit several combinations of ARIMA (p, d, q) model from which the best model that fit the data will be selected.

4.3 Model Estimation

The estimates of the various parameters were compared and the model with the least or smallest value of Sum of Square and Mean Square was selected as the best fit model, thus comparing the Sum of Squares and the Mean Square. The estimation of the parameters is of great important in the model building. The parameters obtained are estimated statistically by the method of least square (Table 4.6). A p-value is used to test the significance of the parameters. When a p-value of the parameters of the selected model is less or greater than the alpha value of 5 percent we include or exclude the parameter that is not significant. ARIMA (2,1,3) was selected as the best fit model since it has the minimum value of sum of square and mean square error in the table below.

Table 4.6 COMPARISON OF ARIMA MODEL

MODEL	MEAN SQUARE ERROR	SUM OF SQUARE ERROR
ARIMA (0,1,1)	6.424	366.187
ARIMA (0,1,2)	6.023	337.283
ARIMA (0,1,3)	6.025	331.365
ARIMA (0,1,4)	6.094	329.075
ARIMA (0,1,5)	6.167	323.663
ARIMA (1,1,0)	8.864	505.260
ARIMA (1,1,1)	6.277	351.520
ARIMA (1,1,2)	6.031	331.710
ARIMA (1,1,3)	6,245	337.219
ARIMA (1,1,4)	5.982	317.072

ARIMA (1,1,5)	6.224	323.636
ARIMA (2,1,0)	6.931	388.146
ARIMA (2,1,1)	6.043	332.361
ARIMA (2,1,2)	6.098	329.303
ARIMA (2,1,3)	5.233	277.347
ARIMA (2,1,4)	5.955	309.637
ARIMA (2,1,5)	5.828	297.212
ARIMA (3,1,0)	6.846	376.529
ARIMA (3,1,1)	7.189	388.214
ARIMA (3,1,2)	6.783	359.521
ARIMA (3,1,3)	5.271	274.113
ARIMA (3,1,4)	5.820	297.351
ARIMA (3,1,5)	6.261	313.026
ARIMA (4,1,0)	6.140	374.760
ARIMA (4,1,1)	7.101	376.353
ARIMA (4,1,2)	6.526	339.352
ARIMA (4,1,3)	5.261	268.309
ARIMA (4,1,4)	6.045	302.229
ARIMA (4,1,5)	6.599	323.375
ARIMA (5,1,0)	6.854	363.279
ARIMA (5,1,1)	7.092	368.802
ARIMA (5,1,2)	6.870	350.366
ARIMA (5,1,3)	5.882	294.109
ARIMA (5,1,5)	5.970	292.552

Since ARIMA (2,1,3) was considered as the best fit model, the general equation can be written as

$$X_t = \Phi_0 + \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3+} \varepsilon_t \qquad (4.1)$$

Where X_t represent the data 'bad loan', Φ_1 and Φ_2 represent the coefficient of the first and second order of the autoregressive process, θ_1, θ_2, and θ_3 represent the coefficient of the first, second, third, fourth and fifth order of the moving average and ε_t is the random walk process or White Noise Process.

Table 4.7 ARIMA (2,1,3) MODEL

TYPE	COEFFICIENT	P-VALUE
AR 1	-0.8219	0.000
AR 2	-0.8170	0.000
MA 1	-0.1888	0.084
MA 2	0.1471	0.184
MA 3	0.9449	0.000
Constant	0.25815	0.000

In fitting the model, since the p-value of AR 1, AR 2, MA 3 and the constant and are less than the alpha value of 0.05 and therefore must be included in the model but the p-value of MA 1 and MA 2 are greater than the alpha value of 0.05, therefore must be excluded from the model. The equation now becomes

$$X_t = \Phi_0 + \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \varepsilon_t + \theta_3 \qquad (4.2)$$

$$X_t = 0.25815 - 0.8219X_{t-1} - 0.8170X_{t-2} + \varepsilon_t + 0.9449\varepsilon_{t-3} \quad (4.3)$$

4.4 Diagnotic Checking

Since the parameters are estimated statistically, before forecasting it is necessary to check the adequacy of the identified model. The model is said to be adequate when the residual are random. To check the overall adequacy of the model, Ljung-Box Statistic is used which followed a chi-square distribution. The null hypothesis is either to reject or fail to reject based on the low or high p-value associated with the statistic. The residual of ARIMA (2,1,3) was checked for normality and randomness.

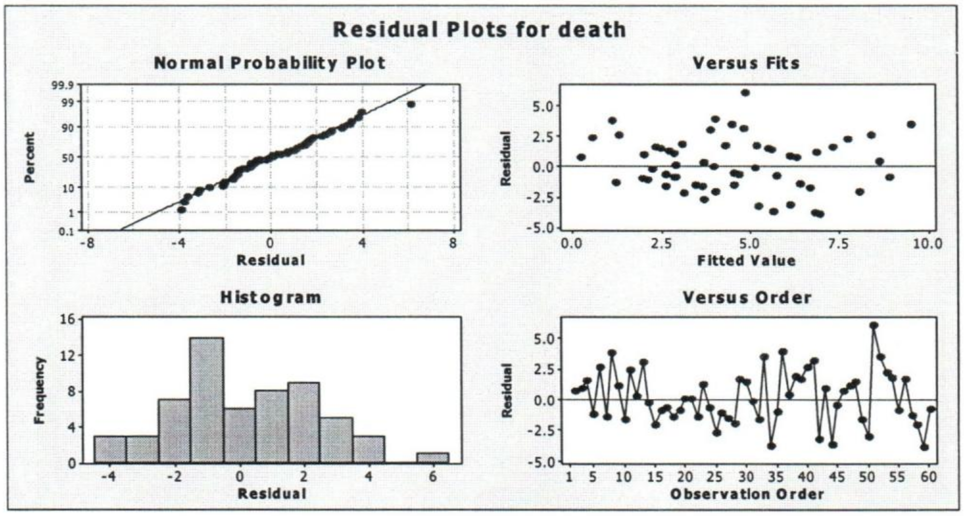


Figure 4.9 Four in one plot for normality and randomness

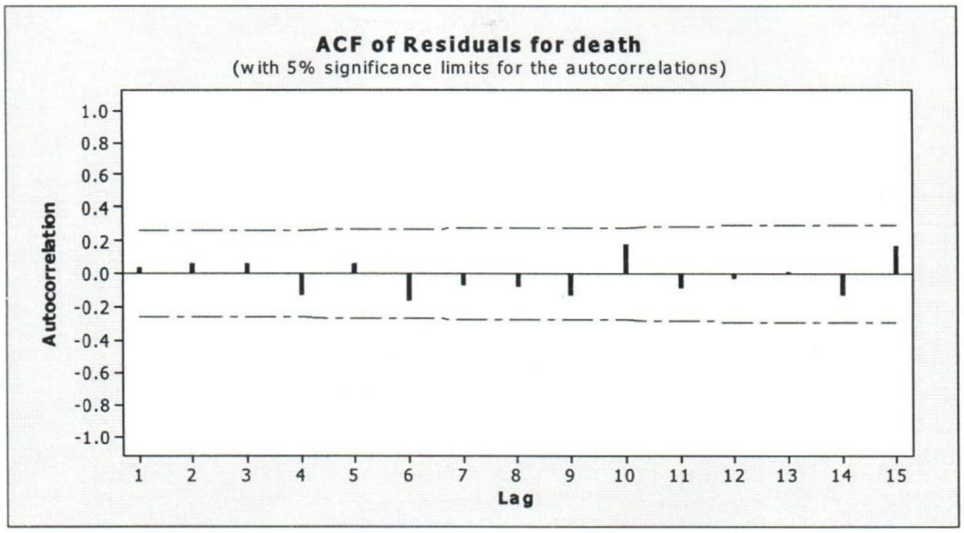


Figure 4.10 ACF plot for the residuals

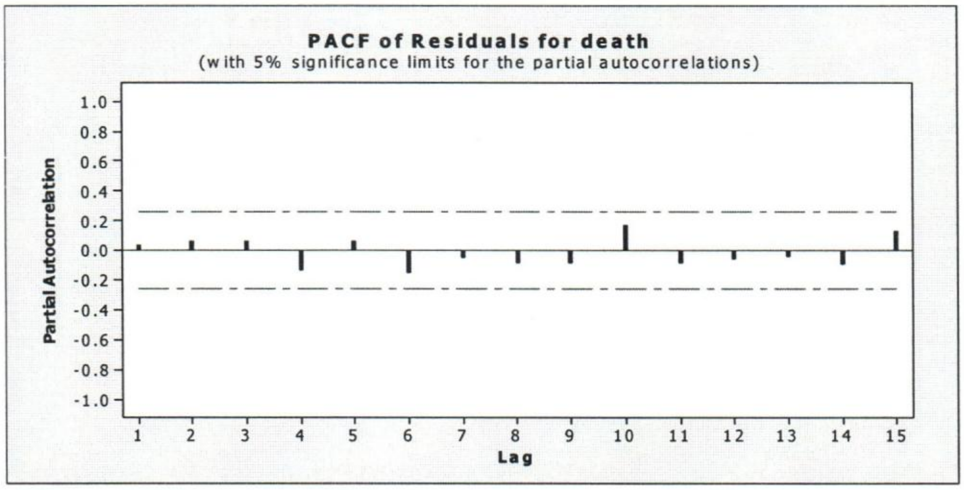


Figure 4.11 PACF plot for residuals

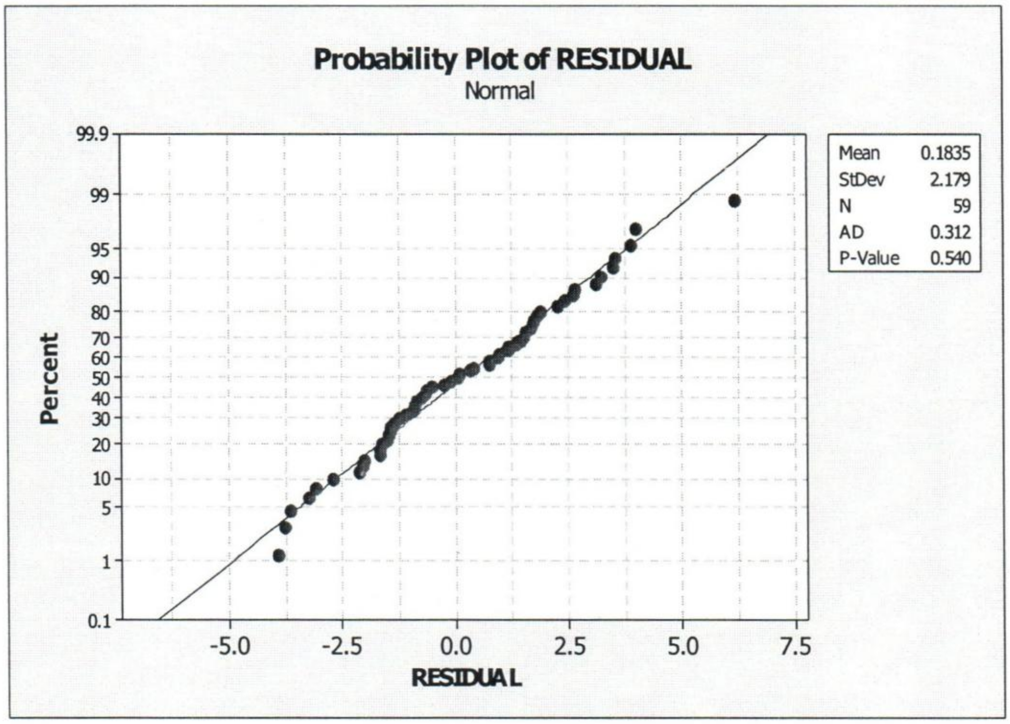


Figure 4.12 Anderson Darling Test for the residuals.

Since the parameters are estimated statistically, before forecasting it is necessary to check the adequacy of the identified model. The model is said to be adequate when the residual are random. To check the overall adequacy of the model, Ljung-Box Statistic is used which followed a chi-square distribution. The null hypothesis is either to reject or fail to reject based on the low or high p-value associated with the statistic. The residual of ARIMA (2,1,3) was checked for normality and randomness.

Figure 4.9 was used to check for the normality of the residual, the normal probability plot indicate that all the values of the residuals fails along the goodness of fit line although there are some outliers. The nature of the histogram also indicates tsshe normality of the residual.

The Kolmogorov-Smirnov Test, Anderson Darling Test and Ryan-Joiner Test were further carried out to affirm the normality for the residuals, with p.-values of 0.150, 0.540 and 0.100 respectively which is greater than the alpha at 5% leading us to accept the null hypothesis that the residuals are normally distributed against the alternative hypothesis that the residuals are not normally distributed.

Figure 4.10 and 4.11 show that all the residual spike of the ACF and the PACF graphs fall within the confidence interval which indicate that the residuals are random. Finally the overall adequacy of the model was checked using the Ljung-Box Statistics. Table 4.8 shows that all the p-value are greater than alpha level of 5%, which confirm that the ARIMA (2, 1, 3) model is adequate for forecasting.

Table 4.8 Forecast for Number of Deaths through Road Traffic Accident for 2013

Period	Forecast	95% Significant limits	
		Lower	Upper
61	8.4534	3.9689	12.9380
62	8.0648	3.2879	12.8417
63	6.5306	1.7413	11.3200
64	8.3672	3.5370	13.1973

65	8.3694	3.3848	13.3540
66	7.1252	2.1402	12.1102
67	8.4041	3.3923	13.4160
68	8.6277	3.5333	13.7221
69	7.6571	2.5616	12.7526
70	8.5303	3.4192	13.6415
71	8.8638	3.7082	14.0195
72	8.1344	2.9747	13.2942

The importance of the forecast is to examine what happens to the trend in the near future. That is, to check whether it will increase, decrease or remain constant in some years to come. The forecast of this model shows a fluctuating trend. This therefore implies that there is no particular pattern of the death rate regarding this data.

4.5 Discussion of the results

The results of the study clearly revealed that the deaths recorded through road traffic accident were asymmetric and more peaked in nature. The nature of the distribution of the deaths revealed that the deaths were closely distributed around the mean. The distribution of the deaths for each month of the year also revealed that there is stability of the deaths for each month of the year.

4.6. Trend analysis

The time series plot of the deaths through road accident in figure 4.1 shows the behavior of increase and decrease which finally indicate that there was an increasing trend in the data. The trend analyses were carried out and the measures of accuracy indicate that there was a quadratic trend in the bad loan data with Mean Absolute Percentage Error (MAPE) to be 67.9148, Mean Absolute Deviation of 1.9795 and Mean Square Deviation to be 5.9043 compared to the accuracy value of the linear trend with Mean Absolute Percentage Error (MAPE) to be 63.3816, Mean Absolute Deviation of 1.9800 and Mean Square Deviation to be 5.9043.

Therefore the equation of the quadratic trend is given as;

$$Y_t = 1.638 + 0.0858t + 0.00023t^2 \qquad (4.4)$$

This implies that as time increases the number of deaths also increase.

4.7 The Model

From figures 4.4 and 4.5, the Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF) graphs respectively showed that the data is non-stationary. This was because, the ACF graph decays slowly in a wavelike form towards zero, while the PACF graph shows two significant spike at lag one and two, which indicate that the data was non-stationary on a theoretical assumption. To affirm that the series was not stationary, the ADF Test and KPSS Test both revealed that the series was not stationary.

In order to attain stationarity, the data was difference once and the behavior of the time series plot gave a clear conclusion that it was now stationary. Also the ADF test and KPSS test was carried out after the first differencing which revealed that the series was now stationary. According to Makridakis (1998), the order of (p,d,q) are usually obtained from the combination of the order of the Autoregressive and that of the moving average. Different types of ARIMA (p,d,q) models were identified as ARIMA(0,1,1), ARIMA(0,1,2),ARIMA(0,1,3),ARIMA(0,1,4),ARIMA(0,1,5),ARIMA(1,1,0), ARIMA(1,1,1), ARIMA (1,1,2), ARIMA(1,1,3), ARIMA (1,1,4), ARIMA(1,1,5), ARIMA(2,1,0), ARIMA(2,1,1), ARIMA (2,1,2), ARIMA(2,1,3), ARIMA (2,1,4), ARIMA (2,1,5). ARIMA (3,1,0), ARIMA (3,1,1), ARIMA (3,1,2), ARIMA (3,1,3), ARIMA (3,1,4), ARIMA (3,1,5), ARIMA (4,1,0), ARIMA (4,1,1), ARIMA (4,1,2), ARIMA (4,1,3), ARIMA (4,1,4), ARIMA (4,1,5), ARIMA (5,1,0), ARIMA (5,1,1), ARIMA (5,1,2), ARIMA (5,1,3) and ARIMA (5,1,5).

The estimates of these various parameters were compared and the model with the least or smallest value of Sum of Square and Mean Square was selected as the best fit model, thus ARIMA (2, 1, 3) model was considered as the best fit model because it had the minimum value of sum of square and mean square error in the table 4.5. Since+ ARIMA (2, 1, 3) was considered as the best fit model. The general equation can be written as

$$X_t = \Phi_0 + \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3+} \varepsilon_t \qquad (4.3)$$

Where represent the data 'number of deaths', Φ_1 and Φ_2 represent the coefficient of the first and second order of the autoregressive process, θ_1, θ_2

and θ_3 represent the coefficient of the first, second and third order of the moving average and ε_t is the random walk process or White Noise Process.

From table 4.7, the P-Values of AR 1, AR 2 and MA 3 were less than the alpha value of 0.05 and therefore must be included in the model. Since the P-Value of MA 1 and MA 2 were greater than the alpha value, therefore must be excluded from the model. The equation now becomes

$$X_t = 0.25815 - 0.8219X_{t-1} - 0.8170X_{t-2} + 0.9449\varepsilon_{t-3} + \varepsilon_t \qquad (4.4)$$

4.7 Model Diagnostic

Figure 4.5 is a four in one graph which shows that the residuals are normally distributed, have constant variance, no outliers and presence of random fluctuation. The Anderson Darling Test for normality which was carried to confirm that the residuals are normally distributed reveal a p-value of 0.513 which is greater than the alpha at 5% level of significance, leading us to accept the null hypothesis 'the residuals are normally distributed against the alternative, the residuals is not normally distributed'. Figures 4.6 and 4.7 show that almost all the residual spike of the ACF and the PACF graphs fails within the confidence interval only one spike fail outside the confidence interval which was due to random influence that is 95% of the values of r fail within + or -1.96/n where n is the number of observation, which indicate that the residual are random. Finally, the table below shows that all the p-values are greater than the alpha level of 5%, which confirms that ARIMA (2, 1, 3) is the model.

CHAPTER FIVE

SUMMARY, CONCLUSION AND RECOMMENDATIONS

5.0 Introduction

This chapter covers the summary of all the findings of the research, the various conclusions drawn from the findings and suggested alternatives and recommendations to assist reduce the number of people who are killed through road traffic accidents in the Upper West Region of Ghana

5.1 Conclusions

This research work was carried out to determine whether the number of deaths has seasonal variations; the trend of deaths in the Upper West Region; to develop a model for the number of deaths through road traffic accidents in the Upper West Region and to forecast the future trend of number of deaths in the Upper West Region of Ghana. The monthly road traffic accidents in the Upper West Region were studied from January 2008 to December 2012.

From the results, the time series plot and the quadratic trend model both revealed that there was an increasing trend in the series thus the number of deaths recorded through traffic road accident. Also the descriptive analyses revealed that March, April and May recorded the highest accident cases for the respective years under study.

The autocorrelation function indicates that the moving average was of order three and also the partial autocorrelation function was of order two. The

Kwiatkowski-Phillips-Schmidt-Shin test and Augmented Dickey-Fuller test revealed that the series was not stationary and after the first difference the, both test revealed that the series was now stationary.

The study revealed that ARIMA (2,1,3) was the best model that fit the road traffic accident recorded. The normal probability plot, the histogram, the residual versus fitted value plot and the residual versus the observation order plot indicate that the residuals are normally distributed and random. The ACF and PACF graph for the residual revealed that the residual of the ARIMA (2,1,3) were random. Also the Kolmogorov-Smirnov Test, Anderson Darling Test and Ryan-Joiner Test confirmed that the residual of the ARIMA (2,1,3) model are normally distributed.

Finally the Ljung-Box test revealed that the ARIMA (2,1,3) model fitted was 100 percent good for forecasting.

Kwiatkowski-Phillips-Schmidt-Shin test and Augmented Dickey-Fuller test revealed that the series was not stationary and after the first difference the, both test revealed that the series was now stationary.

The study revealed that ARIMA (2,1,3) was the best model that fit the road traffic accident recorded. The normal probability plot, the histogram, the residual versus fitted value plot and the residual versus the observation order plot indicate that the residuals are normally distributed and random. The ACF and PACF graph for the residual revealed that the residual of the ARIMA (2,1,3) were random. Also the Kolmogorov-Smirnov Test, Anderson Darling Test and Ryan-Joiner Test confirmed that the residual of the ARIMA (2,1,3) model are normally distributed.

Finally the Ljung-Box test revealed that the ARIMA (2,1,3) model fitted was 100 percent good for forecasting.

5.2 Recommendations

Based on the state of deaths through road traffic accidents in the Upper West Region of Ghana, it is recommended that;

i) The Vehicle inspection Officers (VIO) should increase their inspections on road worthiness of many vehicles.

ii) The Motor Traffic and Transport Unit (MTTU) of the Ghana Police Service (GPS) should curb excessive speeding on our roads.

iii) The Judicial Division should impose harsher sanctions/punishment on reckless and careless drivers.

iv) The Drivers and License Authority (DVLA) should conduct objective examination to drivers whenever they want to renew their licenses.

v) Institutions that enforce road traffic regulations should apply the laws so that all perpetrators of road traffic offences are brought to book to deter others.

REFERENCES

Akaike, H. (1974). A New Look at the Statistical Model Identification. *Transactions on Automatic Control 19 (6): 716–723.*

Afukaar et al., (2006). Road traffic fatalities trends in the Ashanti region of Ghana. *Bi-Annual Journal of the Building and Road Research Institute (CSIR), Ghana. Vol. 10.*

Asogwa, S.E. (1978).Road traffic accidents: The doctor's point of view. *African Journal Medical Sciences; (7):29 35.*

Asogwa, S.E. (1992a). Road traffic accidents in Nigeria a review and reappraisal.*Accident Analysis Prevention; 24 (2):149-155.*

Asogwa, S.E. (1992b). Road traffic accidents in Nigeria: A review and a reappraisal, *Accident Analysis and Prevention: 23 (5), 343-435.*

Box, G.E.P. and Jenkins, G.M. (1976) Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco.

Broughton, J. (1991). Forecasting road accident casualties in Great Britain. *Accident Analysis and Prevention; 23(5):353-362.*

Burnham, K.P., and Anderson, D.R. (2002). Model Selection and Inference.Springer-Verlag, New York.

Cunill, M., Gras, M.E., Oliver A.S , Planes, M., & Sullamn, M.J.M. (2004). An investigation of factors reducing seat belt use among Spanish drivers and passengers on urban roads.*Accident Analysis and Prevention.36,439-445.*

Dee, T.S. (1998), Reconsidering the effects of seat belt laws and their enforcement status. *Accident Analysis and Prevention 30, 1-10.*

Department of the Environment, Transport and the Regions (1999). Road Accidents Great Britain 1998: Annual Report. London: DETR.

Dopson A. J, Baguley, C .J and Hills, B (1999). Road safety in Developing Countries; An overview, Overseas Centre. Transport and Road research Laboratory, Crowthorne, and Berkshire, United Kingdom.

Graham, J.D. (1993) Injuries from traffic crashes: meeting the challenge.Ann Rev Public Health; 14:515-543.

Ghana national road safety commission, Bolgatanga, 2012

Halim, S. and Bisono, I.N. (2008). Automatic Seasonal Autoregressive Moving Average Models and Unit Root Test Detection. International Journal of Management Science and Engineering Management, 3(4): 266-274.

Hamilton, J.D. (1994). Time Series Analysis.Princeton University Press, Princeton, NJ.

Harrison J.E and Berry J.G (2008). Serious injury due to landtransport accidents, Australia, 2005–06. Injury research and statistics series number 42. Cat. number.INJCAT 113. Adelaide: AIHW.

Hayes, M.M.M. (1972). Road accidents:Their cause and prevention,with particular reference to Rhodesia.*Central African Journal Med;23(4):71-74.*

Hurvich, C.M., and Tsai, C.L. (1989).Regression and Time Series Model Selection in Small Sample. Biometrika 76: 297-307.

Jorgenson, S. H, and Abane, A. M. (1999). A comparative study of urban traffic accidents in developing and developed countries: Empirical observations and problems from Trondheim (Norway) and Accra (Ghana). Bulletin of Ghana Geographical Association. No. 21, 113-128.

Kayombo, E.J. (1995). Motor traffic accidents in Dar es Salaam.Trop Geogr Med; 47(1): 37-39.

Kent, S.L.(1991). Epidemiology of Traffic accidents in A Swedish Municipality.*Accident Analysis Prevention; 23(6):509-519.*

Kirchgässner, G. and Wolters, J. (2007) Introduction to Modern Time Series Analysis. Springer-Verlag Berlin Heidelberg.

Kleiber, C. and Zeileis, A. (2008). Applied Econometrics with R. Springer Science +Business Media, LLC, NY, USA.

Krug EG, Sharma GK, Lozano R. *(2000).* The global burden of injuries. *American Journal of Public Health, , 90:523–526.*

Lave, L.B., Songer, T.J., La Porte, R.E. (1993). Should person with diabetes be licensed to drive trucks: Risk management. Risk Anal; 13(3):327-334.

Lemming J.J. (1969).Road accidents, Prevent or punish? Definitions, facts and figures. London, Cassell.Chapter 3:17-41.

Leon, S.R. (1996). Reducing death on the Road. The effects of minimum safety standards, Letter 78:17–25.

Massie, D.C., Campbell, K.L., Williams.A.F.(1995).Traffic accident involvement rates by driver's age and gender.*Accident Analysis and Prevention; 27(1):73-87.*

Mercy JA et al. Public health policy for preventing violence. Health Affairs, 1993:7–29.

National Road Safety Commission (NRSC). (2010). Road traffic Crashes Statistics.

Odero, W. (1995). Road Traffic Accidents in Kenya: An epidemiological appraisal. East Afr Med J; 72(5):295-305.

Odero, W., Garner, P. and Zwi, A. (1997). Road traffic injuries in the developing countries: a comprehensive review of epidemiological studies. Journal of Tropical Medicine and International Health. 2(5), 445-460.

Orsay, E.M., Dowan, W.L., Lewis, R., Rama, and K.V. (1994) .The impaired driver: Hospital and Police detection of alcohol and other drugs of abuse in motor vehicle crashes. *Ann Emerg Med; 24 (1):51-55.*

Pankratz, A. (1983). Forecasting with Univariate Box-Jenkins Models: Concepts and Cases.John Wily & Sons. Inc. USA.

Parada, M.A., Cohn, L.D., Gonzalez, E., Byrd, T., & Cortes, M. (2001). The validity of self-reported seat belt use: Hispanic and non-Hispanic drivers in El Paso. *Accident Analysis and Prevention.*

Peden, M., Scurfield, R., Sleet, D., Mohan, D., Hyder, A.A., & Mathers, C.eds. (2004). World Report on Road Traffic Injury Prevention. World Health Organization, Geneva.

Pfaff, B. (2008). Analysis of Integrated and Co integrated Time Series with R. 2Ed, Springer Science +Business Media, LLC, NY, USA.

Reinfurt, D., Williams, A.F., Wells, J., & Rodgman, E. (1996). Characteristics of drivers not using seat belts in a high use state. J. Saf. Res. 27, 209-215.

Rivara, F.P.and Barber, M. (1985).Demographic analysis of childhood pedestrian injuries.

Shibata, A. Fukuda, K.(1994). Risk factors of fatality in motor vehicle traffic accidents. *Accident Analysis Prevention; 26(3):391-397.*

Shumway, R.H. and Stoffer, D.S. (2006). Time Series Analysis and Its Applications With R Examples 2Ed, Springer Science +Business Media, LLC, NY, USA.

Stewart, J.R. (1993). Seat belt use and accident involvement: a comparison of driving behaviour before and after a seat belt law. *Accident Analysis and Prevention. 25, 757-763.*

The Chronicle (2011), Socio-Economic Mishaps.

World Health Organisation, The World Health Report 1999: Making a Difference, Geneva,1999

Violant, J.M., and Marshall, J.R. (1996).Cellular phones and traffic accidents: an Epidemiological Approach. Accid Anal Prev; 28(2):265-275.

Zhang, J., Fraser, S.,Lindsay, J.,Clarke, K,Mao,Y. (1998). Age specific pattern of factors Related to fatal motor vehicle traffic crashes: Focus on young and elderly drivers. Public Health; 112(5): 289-295.

APPENDICES

Appendix A: Model Estimation

Table A1: Final Estimates of Parameters of ARIMA (0,1,1)

Type	Coef	SE. Coef	T	P	SS	MS
MA 1	0.9574	0.0636	15.04	0.000	366.187	6.424
Constant	0.09602	0.02453	3.91	0.000		

Table A2: Modified Box-Pierce (Ljung-Box) Chi-Square statistic of ARIMA(0,1,1)

Lag	12	24	36	48
Chi-Square	12.0	19.8	25.8	38.1
DF	10	22	34	46
P-Value	0.287	0.593	0.843	0.789

Table A3: Final Estimates of Parameters of ARIMA(0,1,2)

Type	Coef	SE. Coef	T	P	SS	MS
MA 1	0.6134	0.1325	4.63	0.000	337.283	6.023
MA 2	0.3620	0.1334	2.71	0.009		
Constant	0.09532	0.02592	3.68	0.001		

Table A4: Modified Box-Pierce (Ljung -Box) Chi-Square statistic of ARIMA(0,1,2)

Lag	12	24	36	48
Chi-Square	11.7	17.7	27.0	37.4
DF	9	21	33	45
P-Value	0.229	0.666	0.761	0.782

Table A5: Final Estimates of Parameters of ARIMA(0,1,3)

Type	Coef	SE. Coef	T	P	SS	MS
MA 1	0.6588	0.1376	4.79	0.000	331.365	6.025
MA 2	0.4333	0.1517	2.86	0.006		
MA 3	-0.1264	0.1426	-0.89	0.379		
Constant	0.09852	0.02395	4.11	0.000		

Table A6: Modified Box-Pierce (Ljung-Box) Chi-Square statistic of ARIMA(0,1,3)

Lag	12	24	36	48
Chi-Square	8.1	14.8	23.7	35.2
DF	8	20	32	44
P-Value	0.421	0.788	0.854	0.826

Table A7: Final Estimates of Parameters of ARIMA(0,1,4)

Type	Coef	SE. Coef	T	P	SS	MS
MA 1	0.6748	0.1417	4.76	0.000	329.075	6.094

MA 2	0.4045	0.1634	2.48	0.016
MA 3	-0.1677	0.1644	-1.02	0.312
MA 4	0.0518	0.1458	0.36	0.724
Constant	0.10078	0.02505	4.02	0.000

Table A8: Modified Box-Pierce (Ljung-Box) Chi-Square statistic of ARIMA(0,1,4)

Lag	12	24	36	48
Chi-Square	7.2	13.5	21.8	33.8
DF	7	19	31	43
P-Value	0.413	0.814	0.888	0.840

ARIMA (0, 1, 5)

Table A9: Final Estimates of Parameters of ARIMA (0,1,5)

Type	Coef	SE. Coef	T	P	SS	MS
MA 1	0.6555	0.1446	4.53	0.000	323.653	6.107
MA 2	0.4146	0.1643	2.52	0.015		
MA 3	-0.1418	0.1750	-0.81	0.421		
MA 4	-0.0385	0.1670	-0.23	0.819		
MA 5	0.0882	0.1520	0.58	0.564		
Consatnt	0.10092	0.02468	4.09	0.000		

Table A10: Modified Box-Pierce (Ljung-Box) Chi-Square statistic of ARIMA (0, 1,5)

Lag	12	24	36	48
Chi-Square	7.5	15.0	25.7	37.0
DF	6	18	30	42
P-Value	0.274	0.660	0.691	0.690

ARIMA (1, 1, 0)

Table A11:Final Estimates of Parameters of ARIMA(1,1,0)

Type	Coef	SE. Coef	T	P	SS	MS
AR 1	-0.3047	0.1266	-2.41	0.019	505.260	8.864
Constant	0.0958	0.3876	0.25	0.806		

Table A12: Modified Box-Pierce (Ljung-Box) Chi-Square statistic of ARIMA(1,1,0)

Lag	12	24	36	48
Chi-Square	26.5	30.8	39.1	51.2
DF	10	22	34	46

P-Value	0.003	0.100	0.251	0.278

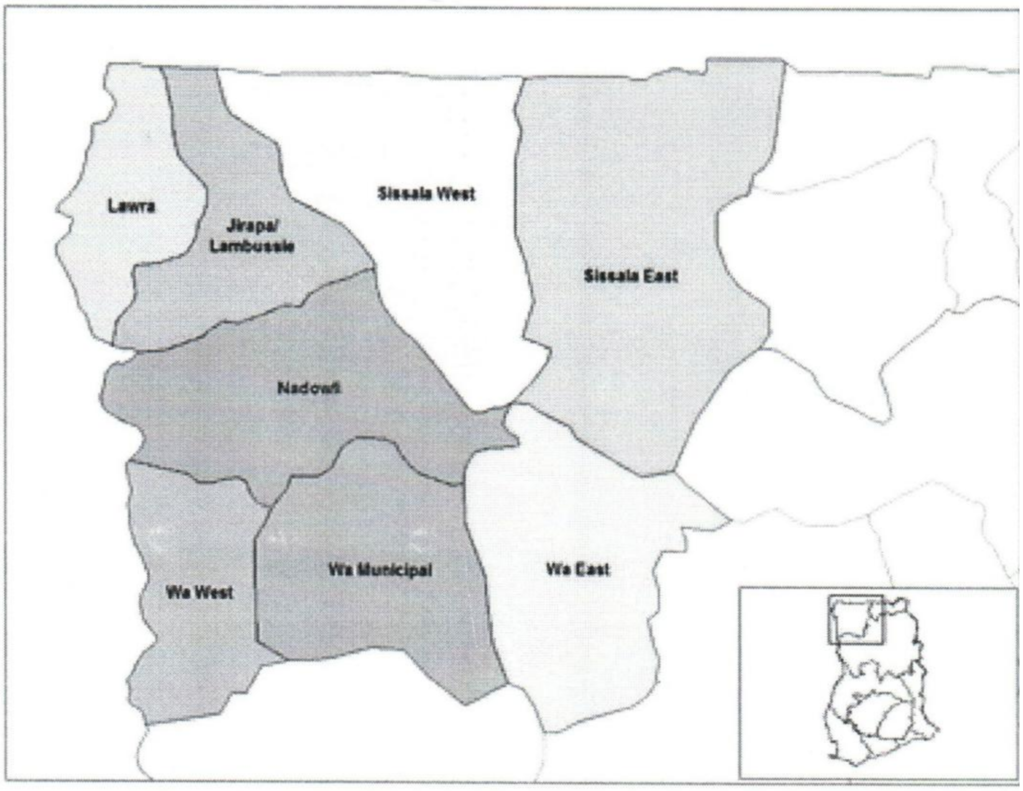
Appendix B

Table 4.8: Total Death Cases Reported Covering Five-Year.

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Tota
2008	0	1	3	4	0	3	2	4	4	2	3	3	30
2009	3	2	1	2	2	1	2	2	3	2	4	3	27
2010	1	1	2	2	4	7	4	1	4	3	1	4	34
2011	3	1	7	7	5	6	9	2	2	7	2	7	58
2012	4	7	7	6	3	11	11	10	5	6	3	5	74
Total	11	12	20	21	14	28	28	19	18	20	13	22	224

Source: MTTU-Wa

Appendix C



Source: Ghana national website

Figure 4.13: Map of Upper West Region