BAYESIAN AND MULTILEVEL APPROACHES TO MODELLING ROAD TRAFFIC FATALITIES

By

CHRISTIAN AKRONG HESSE

(20115576)

A Thesis submitted to the Department of Mathematics,
Kwame Nkrumah University of Science and
Technology, Kumasi in fulfilment of
the requirements for the degree of
Doctor of Philosophy in Mathematical Statistics

August, 2016

DECLARATION

I hereby declare that this submission is my own work towards the PhD Mathematical Statistics and, to the best of my knowledge, it contains no material previously published by another person nor material which has been accepted for the award of any other degree of the University, except where due acknowledgement has been made in the text.

Name of Student: Christian Akrong Hesse ....
(PG No. 20115576) SIGNATURE DATE

CERTIFIED BY:
Name of Supervisor: **Dr. F. T. Oduro** ..... ……...
(SUPERVISOR) SIGNATURE DATE

CERTIFIED BY:
Name of Supervisor: **Prof. J. B. Ofosu** ..... ……...
(CO-SUPERVISOR) SIGNATURE DATE

CERTIFIED BY:
Name of Head of Department: **Dr. R. Avuglah** …….. ..……...
(HEAD OF DEPARTMENT) SIGNATURE DATE

ABSTRACT

Smeed, in 1949, provided a regression model for estimating road traffic fatalities (RTFs). In this study, a modified form of Smeed's model is proposed for which it was shown that the multiplicative error term is less than that of Smeed's original model for most situations. Based on this Modified Smeed's model, Bayesian and multilevel methods were developed to assess RTF risk across sub populations of a given geographical zone. These methods consider the parameters of the Smeed's model to be random variables and therefore make it possible to compute variances across space provided there is significant intercept variation of the regression equation across such regions. Using data from Ghana, the robustness of the Bayesian estimates was indicated at low sample sizes with respect to the Normal, Laplace and Cauchy prior distributions. Thus the Bayesian and Multilevel methods performed at least as well as the traditional method of estimating parameters and beyond this were able to assess risk differences through variability of these parameters in space.

ACKNOWLEDGEMENTS

I first and foremost render my sincere thanks to God for granting me the strength and wisdom to be able to complete this work successfully.

My appreciation also goes to my main supervisor, Dr. F. T. Oduro of the Department of Mathematics, KNUST, who assisted me greatly by spending much time in explaining the concept of Markov chain Monte Carlo (MCMC) in relation to Bayesian analysis.

I also wish to express special appreciation to Prof. J. B. Ofosu of the Department of Mathematics and Statistics, Methodist University College Ghana, for his contribution, comments, criticisms and suggestions which have made this thesis a success.

The project also could not continue without the advice and persistent encouragement of Prof. O. A. Y. Jackson of the Department of Mathematics and Statistics, Methodist University College Ghana.

My sincere gratitude goes to Prof. S. K. Amponsah of the Department of Mathematics, KNUST, for his moral support, encouragement and for providing professional guidelines.

I am also grateful to the following institutions for providing the data used in this study:

1. National Road Safety Commission (NRSC) of Ghana,
2. The Driver and Vehicle Licensing Authority (DVLA) of Ghana, 3. Ghana Police Motor Traffic and Transport Unit (MTTU)
4. Ghana Statistical Service.

DEDICATION

This thesis is dedicated to my wife, Mrs. Caroline Hesse, whose unceasing prayers kept me through to the successful completion of this publication. May the Lord grant her long life and good health to enjoy the fruits of her labour.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ABC	.....	Approximate Bayesian Computation
ABS	.....	Anti-lock Braking Systems
ADAS	.....	Advanced Driver Assistance Systems
AIDS	.....	Acquired Immune Deficiency Syndrome
CODA	.....	Convergence Diagnosis and Output Analysis
DOE	.....	Department of the Environment
DV	.....	Dependent Variable
DVLA	.....	Driver and Vehicle Licensing Authority
EEC	.....	European Economic Commission
EMS	.....	Emergency Medical Services
EP	.....	European Parliament
F. I.	.....	Fatality Index
GDP	.....	Gross Domestic Product
GSS	.....	Ghana Statistical Service
HGVs	.....	Heavy Goods Vehicles
HIV	.....	Human Immunodeficiency Virus

HMs	.....	Hierarchical Models
ICC	.....	Intraclass Correlation Coefficient
ILEA	.....	London Education Authority
IV	.....	Independent Variable
JSP	.....	Junior School Project
lme	.....	Linear Mixed Effects
LSD	.....	Least Significant Difference
LSR	.....	Least Square Regression
MCMC	.....	Markov Chain Monte Carlo
MEPs	.....	Members of the European Parliament
ML	.....	Maximum Likelihood
MLDA	.....	Minimum Legal Drinking Age
MLMs	.....	Multilevel Linear Models
MRC	.....	Multilevel Random Coefficient
NI	.....	Northern Ireland
nlme	.....	Linear & Nonlinear Mixed Effects
NRSC	.....	National Road Safety Commission
OLS	.....	Ordinary Least Squares
PDs	.....	Physical Disabilities
RELR	.....	Random-Effects Logistic Regression
REML	.....	Restricted Maximum Likelihood
RTAs	.....	Road Traffic Accidents
RTFs	.....	Road Traffic Fatalities
SEM	.....	Structural Equation Model
SMC	.....	Sequential Monte Carlo
SSA	.....	Sum of Squares Due to Road User Class,
SSB	.....	Sum of Squares Due to Geographical Region,
SSE	.....	Residual Sum of Squares

SST	.....	Corrected Sum of Squares
TB	.....	Tuberculosis
VC	.....	Variance-Components
WHO	.....	World Health Organization

[1]DEFINITIONS

Road Traffic Accident: Accident resulting in injury, death or property damage and which involves at least one vehicle on a public road.

Road Traffic Casualty: Any road traffic accident victim injured or killed within 30 days of the crash. Thus the crash is the event whilst the casualty is the individual crash victim.

Accident Severity: Severity of the most seriously injured casualty.

Fatal Accident: Road traffic accident in which, at least, one casualty dies of injuries sustained within 30 days of occurrence of the accident.

Serious Injury

[1] National Road Safety Commission of Ghana (2011). Building and Road Research Institute (BRRI), *Road Traffic Crashes in Ghana*, Statistics

Accident: Road traffic accident in which, at least, one person is detained in hospital as an in-patient for more than 24 hours.

Minor or Slight Injury Accident: Road traffic accident in which the most severe injury sustained by a casualty is only minor, requiring at most first-aid attention.

Damage-only Accident: Road traffic accident which results only in vehicle or other material damage.

Urban: City or municipal area characterized by a concentration of residential, commercial and industrial buildings and activities.

Non-urban: Areas not falling under the definition of urban.

Junction: Any place at which two or more roads meet, whatever the angle of intersection and parts of such roads lying within 25 meters of that place.

CHAPTER ONE INTRODUCTION

1.0 Background of the Study

There has been rapid economic growth in many countries in the world resulting in an enormous increase in the number of vehicles. As a result, Road Traffic Accidents (RTAs) have become a serious public health problem. RTAs are considered by World Health Organization (WHO, 2004) as a global health problem claiming approximately 1.2 million fatalities per annum. This may be due to the fact that when roads are newly constructed there is likelihood that drivers may be tempted to over speed which may lead to accidents. Another contributing factor is that smooth roads may cause a driver to lose concentration or even be lazy and hence doze off thereby skidding off the road while driving.

It should be pointed out that the European Economic Commission (EEC) and the World Health Organization (1979) have recommended a definition for road traffic accident fatalities which includes only deaths which occur within 30 days following the accident, since 93 – 97% of these fatalities take place within a one month period. A number of countries have not yet adopted this definition (see WHO, 1979). For example, in some countries, a road traffic fatality is recorded only if the victim dies at the site or is dead upon arrival at the hospital. In order to make comparison of accident statistics between countries reasonable, figures obtained from countries which have not adopted the 30-day fatality definition, should be properly adjusted. No adjustment is required for figures from countries such as Ghana, U.S.A and Great Britain, which have adopted the standard fatality definition.

Road traffic fatality rates of a country are known to depend upon factors such as population, the number of motor vehicles in use, the total length of roads, the population density and economic conditions among others. Two of these factors are of prime importance, namely, population (P) and the number of vehicles (N) (Smeed, 1964).

Smeed (1949) gave a regression model for estimating road traffic fatalities (RTFs). In his paper, Smeed showed that the formula

$$\frac{D}{N} = 0.0003 \left(\frac{N}{P}\right)^{2/3} \quad \text{.....(1.1)}$$

gave a fairly good fit to the data from 20 countries, including European countries, USA, Canada, Australia and New Zealand (D = Number of RTFs, P = population size and N = number of vehicles in use). The results obtained by Smeed in his study are consistent with other reported studies by Bener and Ofosu (1991), Jacobs and Bardsley (1977), Fouracre and Jecobs (1977), Ghee et al. (1997) in which an expression of the form

$$\frac{D}{N} = \alpha \left(\frac{N}{P}\right)^{\beta} \quad \text{.....(1.2)}$$

was used for the estimation of RTFs.

Ponnaluri (2012) used data from all states in India to develop seven different models for predicting road traffic fatalities (RTFs) and also examined if the individual models were more relevant for application. The seven models, including that of Smeed's, were tested for fit with the actual data. Smeed's model was found to be the best fit. He showed that the original Smeed formulation cannot simply be discounted due to reasons cited by many researchers. This is because Smeed's model is ***parsimonious in parameter usage*.** According Ponnaluri (2012), Smeed's model appears to be observation-driven, evidence-based, and logically valid in measuring the ***per vehicle fatality rate*.**

The dramatic increase in vehicle travel in developing countries called for the effective introduction of intervention/strategies that reduce traffic accidents. An important piece of information for such an introduction lay in the prediction of accidents and their fatalities, which was addressed in a paper published by Al-Matawah and Jadaan (2009). They stated that Smeed's model was originally developed for the prediction of traffic fatalities in both developed and developing countries.

Many authors tried to validate or update the Smeed formula based on newer data. The law was found to be valid with some changes in parameters (Adams, 1987). Fortunately, the increasing trend of the total number of fatalities started to change towards a decreasing trend in some countries, such as UK, from the 60s (Safe Speed, 2013).

It should be noted that, the predominant factors affecting RTFs are not the same as those of road traffic accidents (RTAs). Exposures to risk of RTFs (such as human error, vehicular speed, vehicular density, weather conditions, and nature of the roads and total length of roads) are predominant factors influencing road traffic accidents within a geographical region. However, the rate of RTFs is determined by vulnerability to risk (Such as accessibility, timeliness and appropriateness of emergency medical care as well as adequacy and enforcement of use of safety mechanisms in vehicles).

Exposure to risk of RTFs and vulnerability to risk of RTFs are not correlated. Thus, high exposure does not necessarily imply high vulnerability. For instance, according to National Road Safety Commission (NRSC)[2] of Ghana (2011) report, Greater Accra Region in Ghana, with the highest exposure to the risk of RTF (due to high population and vehicular densities), has the lowest RTF rate among all the other 9 regions in Ghana. Whilst the three Northern regions of Ghana, with the lowest population density have the highest rate of RTFs. [3]Nigeria and Ghana have almost the same vehicular density. However, inhabitants of Nigeria are more vulnerable to die as result of road traffic accidents. Developing countries, with only about 10% of the world motorization, account for about 85% of annual RTFs in the world ([3]WHO, 2004, 2009). Thus, developed countries, though have ***greater exposure to risk of RTFs*** due to high vehicular density, are however less vulnerable to RTFs compared to developing countries.

Predominant factors effecting of RTFs are categorized into two:

(1) Internal factors (Safety mechanism in vehicles such as such as anti-lock braking systems (ABS), air bags and seatbelts)

(2) External factors (Ambulance & Emergency Medical Services, pre-hospital care road traffic accident trauma patients)

One reason why developing countries are more vulnerable to risk of RTF is due to the fact that a large proportion of road traffic accident trauma patients in these regions do not have access to formal emergency medical services (Tiska, et al., 2002). Secondly, the ages of vehicles and

[2] National Road Safety Commission of Ghana (2011). Building and Road Research Institute (BRRI), *Road Traffic Crashes in Ghana*, Statistics

[3] http://www.nationmaster.com/country-info/stats/Transport/Road/Motor-vehicles-per-1000-people [3] World Health Organization (2004) World report on road traffic injury prevention, Geneva.
World Health Organization. (2009). Global Status Report on Road Safety, time for action. WHO, Geneva.

availability of modern safety mechanisms in vehicles plying the roads in these regions have significant effect on the consequences of road traffic accidents. It is obvious that if greater attention is paid on improving road safety mechanisms (such as anti-lock braking systems (ABS), air bags, better design of cars and increased wearing of seatbelts in cars) there could be substantial benefits in reducing injuries and fatalities with respect to road traffic accidents in developing countries.

Smeed's model is of the form

$$\frac{D}{N} = \alpha \left(\frac{N}{P}\right)^{\beta} e, \quad \text{.....(1.3)}$$

where D = Number of RTFs, P = population size, N = number of vehicles in use, e = multiplicative error term, and α & β are parameters to be estimated. Equation (1.3) can be expressed as

$$Y = \alpha X e^{\varepsilon}, \quad \text{.....(1.4)}$$

where, the predictor variable is $X = N/P$ *vehicular density* and the dependent variable is $Y = D/N$ *per vehicle fatality rate*.

The factors affecting RTAs correspond to exposure X while the factors affecting RTFs correspond to vulnerability given the same exposure. In Smeed's model exposure is measured by the variable X whereas vulnerability for a given X is captured by the parameters α and β.

Let X_1 (with $Y = Y_1$) and X_2 (with $Y = Y_2$) be two predictor variables of two geographical regions such that $X_1 = X_2$. If $Y_1 = Y_2$, then the different values of Y is not based on X but is due to the fact that α and β vary across the two geographical regions. It therefore follows that, the parameters of Smeed's model vary from one geographical region to another. Thus, one could use these parameters to assess variability of the risk of RTFs across geographical regions.

1.1 Problem Statement

From the above, parameters of Smeed's model vary from one geographical region to another. Thus, one could use these parameters to assess variability of risk of RTFs across geographical regions. Moreover, a road traffic intervention may be more effective in some geographical regions than others. Classical estimation based on information from a particular region can be essentially useless if the sample size is small in that region.

Although there is extensive and growing literature on Least Square Regression (LSR) models for the estimation of road traffic fatalities of a country, the same cannot be said with respect to the application of Bayesian and multilevel approaches in modeling road traffic fatalities of a country. Smeed's (1949) formulation and other related studies by Ponnaluri (2012), Ghee *et al*., (1997), Bener and Ofosu (1991), Jacobs and Bardsley (1977), Fouracre and Jacobs (1977) used LSR method to estimate the parameters.

However, the LSR approach assumes that the model parameters are constants and thus does not allow the variability of the parameters. Moreover, LSR method of estimation is *very sensitive* to violation of the normality assumption of the model.

Since the parameters of the Smeed's model vary from one geographical region to the other, we need an estimation procedure that

(1) is robust with respect to the assumptions of the model,

(2) could be used to estimate the variance of the parameters across geographical regions,

(3) enables us compare the risk of RTFs across the geographical regions,

1.2 Objectives of the Study

General Objective

The earlier discussion revealed that there appears to be a link between the parameters of Smeed's model and the risk of road traffic fatalities across geographical regions. As a general objective, therefore, this study aims at developing statistical methodology, based on Smeed's model, for assessing the risk of RTFs across sub-populations of a given geographical zone.

Specific Objectives

(1) The first objective of the study is to develop a modified Smeed's model which is more accurate.

(2) Based on the modified Smeed model, the study seeks to develop and use

- the Bayesian analysis approach to derive an estimator, based on a prior distribution that is robust with respect to the normality assumption,
- the multilevel analysis approach to compare the risk of RTFs across sub-populations of a given geographical zone.

(3) Finally, the study seeks to use data from Ghana to validate the developed methods.

1.3 Justification for the Study

The Bayesian and multilevel methods of estimation consider the parameters of the Smeed's model to be random variables and therefore make it possible to compare the risk of road traffic fatalities across geographical regions.

These are very powerful methods of analysis which has the ability to estimate the variance terms that reflects the degrees to which regions differ in terms of the parameters. This distinguishes Bayesian and Multilevel models from the conventional least squares regression method which contains only one variance term, usually denoted by $\square^2$, which reflects the degree to which the actual value of *y* differs from its predicted value within a specific region, which is also associated with the Bayesian and the multilevel models.

In a multilevel model, for instance, we use random variables to model the variation between regions. An alternative approach is to use an ordinary regression model, but to include a set of dummy variables to represent the differences between the regions. The multilevel analysis, for instance, offers several advantages over that of least squares method.

1. We can generalize to a wider population.

For example, one can say something about the growth rate of RTFs that is expected in the Greater Accra region from which the sample was selected.

2. Information can be shared between regions.

By assuming that the random effects come from a common distribution, a multilevel model can share information between regions. This can improve the precision of predictions for regions that have relatively little data.

3. Fewer parameters are needed.

   By contrast, the approach via dummy variables would require 20 parameters, two from each region, together with the variance of the dependent variable for each region. Multilevel analysis will require only 7 parameters to be estimated. This reduction in the number of parameters is particularly important with more complex models and a limited amount of data.

1.4 Significance of the Study

The significance of this study is that, a modified Smeed's model for assessing the risk of road traffic fatalities (RTFs) across sub-populations of a given geographical zone could assist in determining what policy interventions or safety mechanisms must be put in place to reduce or minimize the risk of RTFs.

Policy and planning interventions for minimizing risk of RTFs which focused on exposure to risk or directed towards regulating the behaviour road users are not likely to yield the desired results. Human behaviour, in a complex traffic environment, is uncertain and therefore effort to regulate human behaviour in an indiscipline traffic environment usually achieves little results in any geographical region in the world.

However, interventions directed toward enhancing accessible, timely and appropriate emergency medical care as well as enforcement of use of safety mechanisms in vehicles and ensuring the crashworthiness of vehicles may go a long way in minimizing vulnerability to risk of RTFs. Crashworthiness is the ability of a vehicle to protect its occupants during an impact. It is a measure of how well a vehicle performs during a collision. In the modified Smeed's model, vulnerability for a given exposure is captured by the parameters of the model.

For example, if the parameters of the modified Smeed's model across each sub-population of a given geographical zone are estimated, then the expected risk of RTFs for each region could be obtained and hence appropriate policy and planning interventions could be applied.

In the light of this, there is the need to establish Bayesian and Multilevel approaches for determining the important factors that influence risk of RTFs in a given geographical zone. Prediction using Bayesian and Multilevel models offer a more scientific and potentially better approach to minimizing the risk of RTFs. These modeling approaches will serve as a predictive tool that can be used to examine the distribution of risk across geographical regions. The analysis has the potential to reveal the future regional variations that are likely to exist in the incidence of road traffic fatalities in a geographical zone. It is also rich in terms of policy implications and economics.

1.5 Limitations of the study

The data from Ghana for the validation of the proposed Bayesian and multilevel methods used in this study were obtained from the following sources.

(a) The data on the number of road traffic fatalities were obtained from the National Road Safety Commission (NRSC) of Ghana

(b) The Driver and Vehicle Licensing Authority (DVLA) of Ghana provided the data on the number of registered vehicles in Ghana.

(c) The estimated population figures were obtained from Ghana Statistical Service (GSS) 2010 Population and Housing Census, Summary Report of Final Report.

It is believed that not all accidents are reported to the police for records to be made on them. Also, it is possible that the police might not have filled the accident report form for all accidents which might have been reported to them. It is therefore imperative to admit that the data provided by National Road Safety Commission (NRSC) of Ghana might be under recorded. Also, population censuses in Ghana are conducted every 10 years. Thus, the population figures used in this study where estimated based on the population growth rate of the 2000 and 2010 population censuses. However, since these sources are state institutions with appropriately qualified and train personnel the study assumes that these data are reliable and representative.

1.6 Outline of the thesis

The thesis is organized in six chapters that are linked to the general objective of this study. It also includes information from various sources relating to the study. Chapter One gives the background of the study, problem statement and states the main objective of the study. It also highlights the justification for the study as well as the significance of the study.

Chapter Two reviews the various literature related to the topic under consideration in order to uncover critical facts and findings which have already been identified by previous researchers. There are two sections in this chapter. First it discusses the relevant concepts and theoretical framework of the study (the concepts of Road Traffic Fatalities RTFs). Secondly, the chapter reviews studies of what the researcher found as important contribution to application of Bayesian and multilevel methods.

Chapter Three reviews the methodology used in the study. First it put forward the derivation of a modified Smeed's model and also determines how accurate the proposed modified model of this study is. Based on the modified Smeed's model, the chapter also developed the methodology of two Bayesian approaches for estimating the regression coefficients. Finally, the chapter developed the methodology of the multilevel method which can be used to assess the risk of road traffic fatalities across sub-population of a given geographical zone.

Chapter Four presents some preliminary investigations on some characteristics of road traffic accidents and particularly road traffic fatalities in Ghana which are of general interest and have a certain bearing on the main results of this study. There are five sections, the first is on the epidemiology of RTAs and focusses on the demographic aspects of fatalities, the second deals with the regional distribution of RTFs and is related to the main results, the third deals with RTF characteristics of types of road users while the forth section deals with the effect of age on road traffic fatality index in Ghana. The final section of chapter 4 derives a logistic regression model for predicting the annual distribution of the proportion of road traffic casualties who die as a result of road traffic accidents in Ghana.

In Chapter five, the study uses data from Ghana to validate the methodology discussed in Chapter 3. The Bayesian method is applied to estimate the regression parameters using the 'conjugate prior' and maximum a posteriori approaches. The Chapter also illustrates the estimation

of multilevel random coefficient using data from the 10 geographical regions in Ghana and examined the risk of road traffic fatality across these regions.

Chapter Six contains the final discussion, conclusions and recommendations.

CHAPTER TWO REVIEW OF RELATED LITERATURE 2.0 Introduction

There are two sections in this chapter. First it discusses the relevant concepts and theoretical framework of the study (the concepts of Road Traffic Fatalities RTF). The concepts help to assess causes and socio-economic consequences of RTAs induced Physical Disabilities (PDs) on livelihoods and well-being of the victims and their households. Theoretical framework is defined as a „conceptual model of how one theorizes or makes logical sense of the relationships among several factors that have been identified as important to the problem" (Sakaran 2003: 19). The theories guide and direct identification of literature sources that suit the research questions and help as a tool for analysis of the findings.

Secondly, the chapter reviews studies of what the researcher found as important contribution to application of Bayesian analyses of hierarchical (multilevel) models and Markov Chain Monte Carlo (MCMC) methods. It also discusses the reviews of recent history of Bayesian statistical softwares. This review is also reflected in the findings and the concluding chapter as it helps to identify critical areas that need intervention.

2.1 Theoretical framework and Concepts

2.1.1 Theoretical framework

Elvik (2006) proposed a framework for a rational analysis of road safety problems. This starts with the definition of a road safety problem as "*Any factor that contributes to the occurrence of accidents or the severity of injuries."* It further defines objectives of rational road safety analysis as "*the identification of those problems that make the greatest contribution to accidents or injuries and that are amenable to treatment*". The taxonomy, a corner stone of the researcher"s rational analysis of road safety problems, aims at providing categorization of road safety problems and has two inseparable parts: Analysis of the size or importance of problem (quantification) and a concept of the amenability of problems to treatment (amenability). Road safety problems are considered having several dimensions such as magnitude, complexity, territoriality, dynamics, severity, inequity, perception and amenability to treatment.

Lu and Wevers (2005) presented a conceptual model for the effects of road traffic safety measures, based on a breakdown in underlying components of road traffic safety (probability and consequence), and five (speed and conflict related) variables that influence these components, and are influenced by traffic safety measures. The model allows estimating relative effects, and together with available data on absolute effects of infrastructure measures, to estimate absolute effects for Advanced Driver Assistance Systems (ADAS) based measures. It may in general help to improve insight in the mechanisms between traffic safety measures and their effects. The model is illustrated by a case study concerning rural roads in the Netherlands.

Since in developing countries long series of traffic volume data are absent, a model for the fit and prediction of road traffic fatalities for developing countries was developed by Koornstra (2007). This model was based on the relationships of income level per capita with road traffic mortality.

Moutari et al. (2005) introduced a macroscopic model for road traffic accidents along highways‟ sections. The researchers the motivation and the derivation of a such model, and presented its mathematical properties. The results are presented by means of examples where a section of a one-way crowded highway contains in the middle a cluster of drivers whose dynamics is prone to road traffic accidents. The coupling conditions was discussed and presented some existential results of weak solutions to the associated Riemann Problems. Furthermore, some features of the proposed model were illustrated through some numerical simulations.

In Blum and Gaudry (2000), household income is used as an economic indicator. A rise in the income results in an increasing vehicle ownership, which in turn increases road use demand and the number of accidents. Also the current rate of interest affects road use, accidents and victims. Unemployment has a small but highly significant negative effect on road use, but only a moderate significant effect on casualties. Tegnér et al. (2000) show that an increase in the number of employed results in a higher number of vehicle-kilometres.

Hakim et al. (1991) published a very comprehensive comparison of different macro models. The objectives of these authors are the identification and establishment of the significance of policy and socio-economic variables affecting the level of road accidents, and the identification of the variables associated with effective policies and interventions to enable decision makers to improve the level of road safety. In general, the form of a macro model can be written as $Y = F(X)$, where Y is the number of accidents or an accident rate, and X is the vector of explanatory variables (driving,

demographic or economic parameters). Sometimes intervention variables are introduced, to show the effect of an intervention or policy change. In this section, several important aspects of macro models are reviewed.

In Germany, an improved version (SNUS-2.5) of the SNUS-1 model (Gaudry and Blum 1993) was developed. The main difficulties faced by the authors have to do with the specification of the employment activity variable and with the role of vehicle stocks in road demand models. Further, several specific aspects characterize the German situation, namely the absence of general speed limits on motorways, the large size of the country, with high car ownership and an important car industry, and the poly-central infrastructure and structural breaks in the data series caused by the unification of the country.

Fridstrom developed the TRULS-1 model in Norway as part of the researcher‟s PhD thesis. It is the successor of the generalised Poisson regression models estimated in Fridstrom and Ingebrigtsen (1991). In the TRULS-1 model, the assumption is made that casualty counts follow a generalised Poisson distribution. The main attempt of the model is to explain exposure. In this study, the traffic safety in the Stockholm Region has been investigated (Tegnér et al., 2000). The model of traffic demand for Stockholm is estimated on aggregate time-series data for the Stockholm County. The objective is to explain traffic volumes (in vehicle-kilometres) and road accidents, using a spectrum of monthly explanatory variables.

TRACS-CA involves the development and the estimation of a structural aggregate model of highway safety, based upon historical time series data (1981 – 1989) from California. The model is consistent with Gaudry‟s DRAG multi-equation approach. McCarthy (2000) generalizes a previous version of the model by refining empirical specifications in traffic exposure and crash frequency, and by including additional models for crash mortality and morbidity.

Fournier and Simard (2000) report an inverse relationship between gasoline car accidents and gasoline prices, the average price of gasoline and the number of non-work-related trips. Note that the effect of gasoline price changes on accidents is not direct. The price of gasoline determines its demand, which in turn affects the number of accidents. Also the number of kilometres driven decreases with a rise in gasoline prices (Tegnér et al., 2000). McCarthy (2000), however, did not find any impact of the real gasoline price on the demand for travel. This is explained by relating the gasoline price to the opportunity cost of travel, which is a generalised cost, including monetary

and time costs. In the same model, the number of accidents does decrease with a rise in the gasoline price.

Variables that may be used to express economic and/or social stress are rates of net outmigration, levels of violent and property crimes, police calls for domestic disputes, rates of suicide and worker strikes. Note that these variables are much wider than aggressive behaviour in traffic. Sivak (1983) found that, as violence and aggressiveness rise, the number of injuries in road accidents increases.

Young drivers are considered as a high-risk group, having a higher probability of involvement in car accidents with injuries. Also in Fournier and Simard (2000), an increase in the number of young drivers, between 16 and 24 years old, results in a rise in the number of road accidents. Quite often, the topic of young drivers has been related to the effect of the minimum legal drinking age (MLDA) on accidents.

Several authors (e.g. Blum and Gaudry, 2000) showed that reducing the speed limits appears to be related to a reduction in fatalities. Also the severity of injuries is positively related to the allowed speed. According to McCarthy (2000), however, increased speed limits slightly reduce risk exposure. In his model for California, higher speeds have no effects on fatal crashes, but a strong positive impact on the frequency of non-fatal injury crashes. Keeping all else constant, there were fewer fatalities per fatal crash and fewer injuries per non-fatal injury crash after the increased speed limits law. But if a crash occurs, a higher speed results in more serious injuries.

According to Zlatoper (1984) and Loeb (1987), the periodic inspection of motor vehicles reduces the number of road fatalities. White (1986) showed that the probability of accident involvement increased with the length of time between inspections. Schroer and Peyton (1979) stated that the inspected vehicles have a lower accident rate than the uninspected vehicles. Also, the accident rate of inspected vehicles decreases after inspection. Poor mechanical condition is a significant factor in motor-vehicle accidents. Fournier and Simard (2000) include in their model an index of maintenance costs for vehicles. According to this model, an increase in vehicle maintenance costs would result in a decrease in the distance travelled and in the number of road accidents. Crain (1980), on the other hand, concluded that vehicle inspection programs have minor impact on highway safety. Random inspections are more effective and less expensive than periodic inspections.

2.1.2 Concepts

Data provided by Peden et al. (2004) indicated that African Region had the highest mortality rate, with 28.3 deaths per 100 000 population. This was followed closely by the low-income and middle-income countries of the Eastern Mediterranean Region, at 26.4 per 100 000 population. Countries in the Western Pacific Region and the South-East Asia Region accounted for more than half of all road traffic deaths in the world. The report indicated that there are notable differences in the way different road users are affected by road traffic collisions as summarized below:

- More than half of all global road traffic deaths occur among young adults between 15 and 44 years of age.
- 73% of all global road traffic fatalities are males.
- Vulnerable road users (pedestrians, cyclists and motorcyclists) account for a much greater proportion of road traffic collisions in low-income and middle-income countries than in high-income countries.

According to research work conducted by Afukaar et al. (2003), majority of road traffic fatalities (61.2%) and injuries (52.3%) occurred on roads in rural areas. About 58% more people died on roads in the rural areas than in urban areas, and generally more severe crashes occurred on rural roads compared with urban areas. Pedestrians accounted for 46.2% of all road traffic fatalities. The majority of these (66.8%) occurred in urban areas. The second leading population of road users affected was riders in passenger-ferrying buses, minibuses and trucks. The majority of these (42.8%) were killed on roads that pass through rural areas. Pedestrian casualties were overrepresented (nearly 90%) in five regions located in the southern half of the country. The study recommended that efforts to tackle pedestrian safety should focus on the five regions of the country where most pedestrian fatalities occur in urban areas. Policies are also needed to protect passengers in commercially operated passenger-ferrying buses, minibuses and trucks because these vehicles carry a higher risk of being involved in fatal crashes.

According to the World Health Organisation (WHO, 2004), approximately 16,000 people die everyday worldwide from all types of injuries. Injuries represent about 12% of the global burden of diseases, making injuries the third most important cause of overall mortality. Deaths from traffic injury are a very significant part of the problem accounting for 25% of all deaths from injury.

There have been several reviews of the costs to society of road traffic injury. A major review was presented in 1994 by the European Commission: "Socio-economic cost of road accidents, final report of action COST 313" (Alfaro et al., 1994). This report is now more than 10 years old. A more recent survey was made as part of the ROSEBUD-project (de Blaeij et al., 2004). This survey first considered methods used in estimating the costs to society of traffic injury, then presented recent cost estimates for selected countries. As far as methods for estimating costs are concerned, the typology shown in Figure. 2.1 was developed in COST-313.

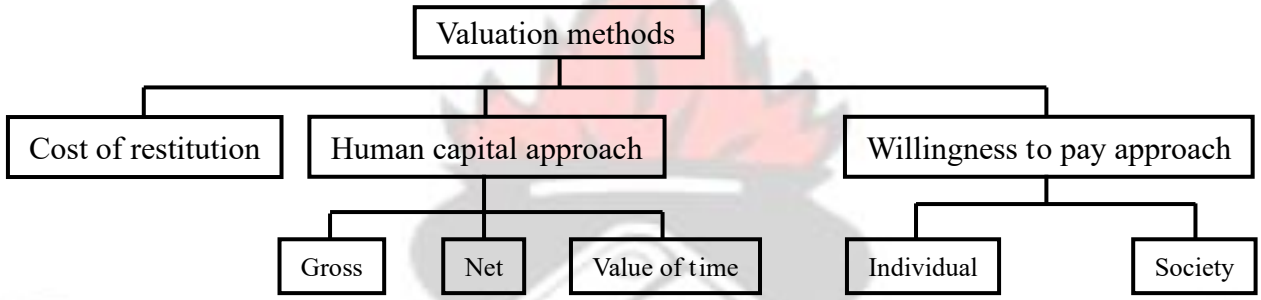


Figure. 2.1: Methods for estimating costs of traffic injury

The costs of restitution are the direct costs generated by road accidents (for example, medical costs, property damage or administrative costs). Generally speaking, the human capital approach is used to estimate the value of lost productive capacity due to a traffic death, whereas the willingness-to-pay approach is used to estimate the value of lost quality of life. Two varieties of the willingness-to-pay approach are normally used: the individual willingness-to-pay and the social willingness-to-pay approaches. According to the former, information about willingness-topay is obtained from individuals, either by studying behaviour in situations where reduced risk must be traded off against other commodities or by means of questionnaires. According to the latter, society"s willingness-to-pay for reduced risk is inferred from the valuation implicit in public decisions like setting speed limits. More information on the different costing methods is given by Trawén et al. (2001), Wesemann (2000) and de Blaeij et al. (2004).

Adekunle (2010) examined the effect of road traffic accident deaths on socio-economic development of Nigeria and suggested measures to improve safety on Nigerian roads. Twelve socio-economic variables were initially selected and the application of principal component analysis on the variables resulted in the emergence of five components which could be used to

describe the pattern of socio-economic development in the country. Road Traffic Accident deaths were assessed through the identification of road traffic accident fatality for the 36 states and the federal capital in the country. Multiple regression method was then used to assess the relationship between the road traffic accident fatalities and the five components of socioeconomic development in the country. The results showed that the five components namely urbanization, security, Universal Basic Education, Public Utility and Unemployment could be used to explain the pattern of road traffic accident deaths and socio-economic development in the country.

Several schools of thought have arisen in an attempt to describe the causes of Road Traffic Accidents (RTAs). One of such is the concept known as the epidemiological model of road accident. An important relationship exists between the concept of risk and accident. The concept of cost is inextricably linked to epidemiological and anthropological notions of risk. Epidemiologists and Clinicians have generally divided risk factors into three categories when addressing the issue of road traffic accidents. The three categories are – human, vehicle and physical/social environmental factors (Mishra et al., 2010)

Five „human" factors have been identified as area where clinically based interventions may have positive outcomes – use of alcohol, use of drugs, morbidity, use of occupant restraints (seat belts) and advanced age. Epidemiological research has tended to focus on human risk factors because they are most relevant to the search for preventive measures and because they have been identified as the most frequent cause of crashes (Polen and Friedman, 1988).

A strong positive association between increasing blood alcohol concentration and the risk of road traffic accident involvement has been documented by researchers for many decades. Alcohol use is generally seen as contributing to traffic injuries by impairing driving capabilities and thus increasing the risk of crash involvement (Pludemmann et al., 2004). Although alcohol is generally thought to be the most important risk factors among all drugs, some evidence has also linked the use of minor tranquilizers such as benzodiazepines to increase risk of crash involvement (Gururaj, 2004).

Studies have also linked certain chronic medical conditions to elevated risks of crash involvement while other studies have presented evidence suggesting that those medical conditions represent a negligible risk in reference to automobile injuries or fatalities (Barbone et al., 1998). There is some evidence however that drivers with diabetes, epilepsy, cardiovascular disease or

mental illness experience higher crash and violation rates (Mishra et al., 2010) but there is an equal number of studies indicating that neither chronic medical conditions nor disabilities among automobile drivers put them at greater risk of road traffic accidents (Mohan, 2002).

Also, much research has demonstrated the efficacy of occupant restraint systems (seat belts) in reducing injuries and preventing deaths in road traffic accidents. Investigations include laboratory studies, post-crash comparisons of injuries sustained by restrained and unrestrained occupants and post-crash judgements by crash analysts regarding the probable effects of restraints had been used. Estimates of effectiveness vary depending on the restraint system being investigated, the type of crash, the size of the vehicle (Zlatoper, 1984) and other factors but tend to cluster between 40 per cent and 60 per cent meaning risk of injury or death due to a road accident is reduced 40 – 60 per cent by using seat belts.

Descriptive studies also suggest that the risk of death from automobile crashes is elevated in older individuals. Fatality rates per passenger/km of travel are relatively high among the older than 70 years of age (Sagberg, 1999). Data from a Northern Ohio Trauma study indicate that motor vehicle deaths per 100,000 population sharply increase among men at about age 70 years and the proportion of injuries that were fatal rose dramatically in those older than 60 years of age.

Although, AUSTROADS (1994) has pointed out that road accidents occur as a result of one or more than one of the following factors: human factor, vehicle factor, road factor and environmental factor. Road traffic accidents in developing countries are largely due to the human factors (approximately 80 – 90 per cent). In fact, recklessness or negligence of drivers, excessive speed, inattention, confusion and lack of judgement are listed as the main human causes of road traffic accidents in Nigeria (Pratte, 1998).

The geographical approach to the study of traffic accidents relates the concept of place, time and environment to accident occurrence. It is believed that land use, road element, width of the road, bending of the road, hilly area, topography and regional distribution in occurrence of road traffic accident are factors to be considered. According to Cutter (1993), geographical scale is important for impacts and their reduction. Land use pattern, types of road network, local business and activity pattern will influence the system risk in an area (Komba, 2006). There is also ruralurban differences. In urban areas, there are more accidents, lower degree of injury while in rural areas there are lower accident levels but more serious fatalities (Astrom et al., 2006).

2.2 General Review on Road Traffic-Accident Fatalities

Quite a large number of the existing studies are first of all interested in exploring the link between income and road traffic fatalities at a cross-country level (see e.g. Wintermute, 1985; Jacobs and Cutting, 1986; S¨oderland and Zwi, 1995; Van Beeck, Borsboom and Mackenbach, 2000; Kopits and Cropper, 2005; Anbarci, Escaleras and Register, 2006; Bishai, Quresh, James et al., 2006; Paulozzi, Ryan, Espitia-Hardeman et al., 2007). Researchers more or less find that at very low levels of income road traffic fatalities per (100,000) person(s) increase with income (because motorization goes up) up to a certain threshold, after which countries seem to be able to invest in safety measures (including safer cars) and possibly behavioral changes that bring traffic fatalities again down (e.g. separate tracks, preventive measures). This inverted U-shaped pattern first has been explicitly pointed out by Van Beeck et al., (2000). Kopits and Cropper (2005) tried, in addition, to relate traffic accidents to environmental externalities. The researchers suspect that the per capita income at which traffic fatalities begin to decline is in the range of incomes at which other externalities such as air pollution begin to decline as well. However, most of these studies offer surprisingly little discussion of how the income effect should be interpreted, for instance, whether this should be seen as a direct effect of income on road traffic crashes or whether income is first of all a proxy for the quality of the road network, the degree of motorization, the implementation and enforcement of safety measures and many other factors. Moreover, even after controlling for income, the residual is usually quite important. Indeed, Japan, for instance experiences a road mortality rate of 5.18 deaths per 100,000 inhabitants, while in the United States this rate is almost three times as high (13.94). Van Beeck et al. (2000) highlight that Greece and Spain are cases with particular by high fatality rates in Europe, even after income is controlled. Wintermute (1985), in an earlier study, emphasized that the analysis of road traffic crash fatalities need to account for a much broader set of determinants than just income. The researcher mentions determinants such as geography, rules and regulations, urbanization, nature of traffic mix, infrastructure development, availability of medical services and culture, but does not examine these empirically.

Yet a few studies do indeed analyze the conditional and unconditional effect of income by including in the analysis additional determinants of this type (Jacobs and Cutting (1986); S¨oderland and Zwi (1995); Anbarci et al. (2006); and Bishai et al. (2006)).

Jacobs and Cutting (1986) use a cross-sectional analysis to examine the link between fatality rates and social, economic and physical characteristics of selected developing countries. These include, besides GDP per capita, the number of circulating vehicles, road density (standardized by land size), vehicle density (per kilometer of road), population per physician and population per hospital bed. The study show that fatality rates are not only related to GDP per capita, but also to vehicle density and population per hospital bed. Nevertheless, vehicle ownership is the only remaining statistically significant variable in a regression in which the effects of all these determinants are examined simultaneously. Using the fatality index (proportion of all persons injured that die) as the dependent variable, the only significant variable in the multiple regression analysis is population per physician (probably partly due to the limited sample size and multicollinearity problems).

S¨oderland and Zwi (1995), also performed a multiple regression analysis. The researchers use data from 83 countries for the year 1990. As the dependent variable the researchers considered alternatively the crude traffic-related death rate per 100,000 persons per year, traffic related deaths per 1,000 registered four-wheeled vehicles per year, the ratio of mid-age to total population mortality, the ratio of the male/female mortality rate and fatal injuries as a proportion of total injuries. The researchers introduce different explanatory variables according to the outcome analyze: number of vehicles per capita, road density (km of road per square km), total surface area, GDP per capita, health expenditure as percentage of GDP and population density. The authors found that GDP per capita is positively correlated with traffic-related deaths per 100,000 population, but negatively correlated with traffic deaths per 1,000 registered cars, suggesting that in per vehicle terms, income reduces road crash fatalities. Moreover, the number of road traffic accident related deaths among youth and elderly people is directly linked to population density. Finally, the study showed that GDP per capita and health expenditure as a share of GDP are associated with a declining rate of fatal injuries among road victims.

Bishai et al. (2006) have a particular focus on the transmission channel between income and road crashes. The authors use data from 41 countries for the period 1992 to 1996. Considered

outcome variables include road traffic crashes, injuries, and fatalities. As channel variables, the researchers used the number of vehicles, kilometers of roadway, fuel consumption, alcohol consumption (available for a single year only), and population density. Fixed effects regression models were used to control for time-constant unobservable heterogeneity across countries. The authors found that with a GDP of $1500 to $8000 PPP per capita, controlling for the other variables, further national income growth does not bring about a further increase in road traffic crash fatalities, although the number of crashes and injuries continue to rise. Bishai et al. (2006) underlines that GDP has probably to be seen as a „proxy" for a set of relevant but hard-tomeasure factors such as urbanization, vehicle mix, road quality and health services. However, the authors do not investigate these channels further.

Other studies exploit the within-country variance in road traffic fatalities (see e.g. Garg and Hyder (2006), La Torre (2007), Van Beeck (2000) and Traynor (2008)). Traynor (2008), for instance, analyzes the relationship between income and fatalities (per vehicle and miles traveled) across counties in the U.S. State of Ohio. The researcher introduces various explanatory variables in addition to income such as population density, the incidence of alcohol abuse and the share of teenage drivers. He finds that the county population density, the presence of interstate highways in rural counties, the prevalence of severe alcohol abuse, the proportion of teen drivers and the presence of a large college student population all have statistically significant relationships with county fatality rates; while for most counties the correlation between per capita income and road-related deaths is not statistically significant. La Torre et al. (2007) identify in addition the employment rate and alcohol consumption as important determinants of road traffic fatalities (both are positively associated with fatalities). However, both studies suffer from a possible omitted variable bias since they do not control for regional or county-fixed effects.

Only very few studies analyze the determinants of road traffic fatalities at the individual level, such as for instance the relation between individual income and driving behavior and how different types of individuals respond to different laws and forms of enforcement. The understanding of why individuals engage in risky behavior such as drink and drive, excess speed, infringement of traffic rules etc. might be particulary important to design and target effective policies.

Fosgerau (2005) uses a large cross-sectional dataset from the Danish National Travel Survey (1996 – 2001) and shows that speed decreases with age, men drive faster than women, singles

drive slightly faster than married individuals and speed decreases with urbanization. He also shows that the effect of income on speed is positive and highly statistically significant. He argues that a higher income increases the perceived value of time and decreases the „real cost" of fines and other speed dependent user costs (noise for instance), which are independent of own income; thus, so the hypothesis, higher income leads to higher speed.

Factor et al. (2008) emphasize the importance of social and cultural characteristics. They use a data set which merges Israeli census data with road traffic accident records. Estimating a logistic regression where the dependant (latent) variable is the probability of drivers from different social groups to be involved in a fatal or a severe road accident, they show that Muslims, separated and widowed people, males, young people, low-skilled workers and less educated individuals have a higher chance to be involved. The authors conclude that traffic accidents may in part be socially generated. They refer to the different habits, skills and styles of each sub-group, which may imply different risk-taking levels.

Only a few authors have made an attempt to explicitly model the behavior of drivers. Exceptions are for instance Blomquist (1986), Boyer and Dionne (1987) and Bishai et al. (2006). In the models suggested in that literature drivers typically are confronted with accidents of a certain probability of occurrence, that depends on own safety efforts (use of safety belt, speed, vehicle quality) and other drivers" driving behavior as well as exogenous safety measures. The „music" in these models comes from the assumption that on the one hand own safety efforts and exogenous safety measures create a disutility because they involve time costs, discomfort, energy and money, on the other hand, in case of an accident the driver has to bear the costs of the accident such as car repair and medical services. Drivers are assumed to weigh these costs against the benefits in order to maximize their expected utility. It is easy to show that in such a setting, one may find that drivers decrease own safety efforts in response to an increase in exogenous safety measures. For instance the introduction of safety belts may lead to higher speed. Keeler (1994) found some evidence for this kind of behavior using panel data for the US. Similar findings exist in the area of HIV/AIDS prevention policies. A study based on a randomized controlled trial found for instance that a higher prevalence of condom use following an information campaign was accompanied by more risky sexual behavior (Kajubi et al., 2005)

This finding leads to another interesting question. Does the effectiveness of campaigns which inform about the consequences of risky behavior depend on riskaversion or even change the risk attitude? Kenkel (1991) shows, for instance, that smoking behavior is responsive to health knowledge. Nevertheless, the author also stresses that formal education still has an impact on health behavior even if health knowledge is controlled for. The expected interactive effects of schooling and health knowledge on alcohol consumption or exercise are not found. According to the author, the differences in the respective stigma attached to these activities across socioeconomic groups may explain this result. Cook and Bellis (2001) find based on (a rather unrepresentative) survey among students that behavior and knowledge about the risks are uncorrelated. But they also find that the perception of risk is related to risk aversion. They identify being male, being younger, having parents in white-collar occupations, belief in God and early exposure to risk as factors that reduce risk aversion. The authors conclude that an effective provision of health information does not only need to transmit the knowledge but requires also an intimate understanding of the media, culture and public perception.

The understanding of individual behavior is crucial to understand and assess the effectiveness of other interventions as well, not just information campaigns. Lave (1985), for example, examines traffic fatalities in conjunction with speed limit legislations across US states and concludes that speed limits are not an adequate policy. He argues that not the average speed but the variance of speed (absence of coordination) causes traffic fatalities. A problem of that argument is of course that the variance of speed is rising in the average speed, so the results do not rule out the fact that a lower speed leads to less fatalities holding constant the variation in speed. It is hard to believe that both the occurrence and the severity of car crashes are independent of the level of speed at which crashes happen and just depend on the difference in speed. This and other studies do also not properly control for the degree of the enforcement of speed limits.

Using panel data from 46 Japanese prefectures for the years 1988 to 2000, Yamamura empirically examines the role of social norms (reinforced by social capital and social structures) for drivers‟ attitudes, in particular dangerous driving. Social capital and social structures are proxied by the number of community centers in the prefecture, the share of emigrants to other prefectures, and the share of immigrants from other prefectures. The study controls for a number of variables, including the number of policemen, which are seen as a proxy for „formal‟ deterrents.

Given that this variable is likely to be endogenous, it is instrumented using income. The involved exclusion restriction can of course be doubted since income itself has to be seen, as shown above, as a determinant of driving behavior. However, taken together the study finds that formal deterrents hardly affect dangerous driving behavior, whereas informal deterrence prevents drivers from driving dangerously (but does not necessarily enhance attentive driving). Similar to the studies cited above, this study also finds that mandated safety inspections induce drivers to drive less attentively (the „off setting effect").

Carpenter (2004) uses the US American Behavioral Risk Factor Surveillance System (BRFSS), which is a large state representative telephone survey collecting information on alcohol consumption and drunk-driving behavior for young adults, 18 years and older to assess the effects of the "Zero Tolerance" policies. The empirical model accounts for unobserved state, year and seasonal fixed effects. The author also introduces in his regression other control variables such as drink-driving laws, the state unemployment rate, the state beer tax and the state minimum legal drinking age. The author finds strong evidence that the main effect of the "Zero Tolerance" policy was to reduce heavy episodic drinking by males aged 18 to 20. An increase of the beer tax or changes in the minimum drinking age are shown to be less effective as they tax all levels of drinking instead of those that lead directly to the alcohol-related traffic fatalities.

Another, again small, strand of the literature investigates the determinants of the involvement of road users and the health and economic burden victims have to carry. As one can expect, the typical profile of victims varies a lot between low and high income countries. Whereas in low income countries pedestrians and (motor) cyclists are the most vulnerable road users, car occupants dominate in high income countries (see e.g. Jacobs et al., 2000; Ansari et al., 2000; Montazeri, 2004; Regional Health Forum of South-East Asia, 2004; Paulozzi et al., 2007 and WHO, 2009). Again, not much is known about within country variation, i.e. whether poorer population groups are systematically more affected than richer groups. But a study in Bangalore, India seems to provide some evidence for such a negative gradient in income (Aeron-Thomas et al., 2004). This correlation is mainly driven by the fact that different income groups use different transport means. The study by Factor et al. (2008), cited above, also suggested for the case of Israel that among the most vulnerable road users are minorities, low-skilled workers and individuals with low education.

Given the limited space of this note, we have certainly not given justice to all the work which has been done on the economic aspects of road traffic crashes, but nevertheless to us it seems to be, in particular compared to other health problems, an under-researched area. This is in particular true concerning the determinants of driving behavior. A good starting point to make further progress in this field would be to elaborate a rigorous conceptual framework of driving behavior and to test such a model using an experimental design. The theoretical side would certainly benefit if it addressed the interaction between risk attitude and time preference (see e.g.

Van der Pol and Ruggeri, 2008).

2.3 Review on Regression Models

Wedagama (2010) did a study to investigate the influence of accident related factors on motorcycle fatal accidents in the city of Denpasar during the period 2006 – 2008 using a logistic regression model. The study found that the fatality of collision with pedestrians and right angle accidents were respectively about 0.44 and 0.40 times lower than collision with other vehicles and accidents due to other factors. In contrast, the odds that a motorcycle accident will be fatal due to collision with heavy and light vehicles were 1.67 times more likely than with other motorcycles. Collision with pedestrians, right angle accidents, and heavy and light vehicles were respectively accounted for 31%, 29%, and 63% of motorcycle fatal accidents.

Goswami and Sonowal (2011) did statistical analysis of road traffic accident data for the year 2009 in Dibrugarh city, Assam, India. Data interpretation was done using Degree of freedom, Chi-square test for goodness of fit, χ^2-test for independence of attributes and Kruskal-

Wallis test. They found that human characteristics (rush and negligence) make 95.38% of the total RTAs. 60% of the accidents were recorded during day time (6 AM to 6 PM). The peak time was between 12 PM to 6 PM (38.46%). The highest numbers of accidents (32.30%) were observed in the heavy rainy season during the months of July – September.

A study was done by Fujita and Shibata (2006) to clarify the relation between alcohol use and traffic fatalities in accidents involving motor vehicles in Japan between 1987 and 1996. Multiple logistic regression models were used to assess the effect of alcohol use on the risk of traffic-accident deaths. The data showed that 58,421 male drivers were involved in traffic accidents during

the 10-year study period, and that 271 of these were killed as a result of the accident. Alcohol use was significantly associated with speed, seat belt use, time, and road form. Among male motorcar drivers, the odds ratio of alcohol use before driving, after adjusting for age, calendar year, time, and road form, was 4.08 (95% confidence interval, 3.08–5.40), which means that about 75% of fatalities (attributable risk percent among exposed) might have been prevented if drivers had not drunk before driving.

A descriptive analysis of road traffic accidents (RTAs) and injury data in Kenya was done using routine accident reports, official statistical abstracts, published and unpublished surveys. The characteristics of injury - producing accidents examined included trends, distribution patterns, risk factors, types of vehicles involved, and road-users injured or killed between 1962 and 1992. It was found that fatality rate per 10,000 vehicles increased from 50.7 to 64.2, while fatality per 100,000 populations ranged between 7.3 and 8.6. 66% of the accidents occurred during daytime. 60% of the reported RTAs occurred on rural roads and had a higher case fatality rate (CFR) of 16% compared to those occurring in urban areas (11%). Human factors were responsible for 85% of all causes. Vehicle-pedestrian collisions were most severe and had the highest CFR of 24%, while only 12% of injuries resulting from vehicle-vehicle accidents were fatal. Utility vehicles and buses were involved in 62% of the injury producing accidents. Of all traffic fatalities reported, pedestrians comprised 42%, passengers 38%, drivers 12%, and cyclists 8%.

Mohammad (2009) conducted statistical analysis for road traffic accidents and associated casualties in Bangladesh. An exploration was undertaken using the averages (per annum) of rates of fatal casualty, accident and involved vehicles applying Bar-charts. Annual time series data were also investigated using trend lines. Time series, Mann-Whitney, Kruskal-Wallis tests were used as well as modeling of two/ three-way data was conducted using the frequencies of fatal casualty, fatal accident and involved vehicles applying Poisson regression. The research found out that pedestrians are highly involved in the casualty figures. Fatal hit pedestrian is the main collision type accident. Maximum fatal accidents occur at out of junction. Cities have higher accident and casualty rates than that for non-cities (divisions/ districts, excluding cities). National highways are the main venues of accidents and casualties. Heavy vehicles including buses and trucks are predominantly involved in casualty accident.

Ahmad et al. (2012) used regression model analysis for the analysis of accidents using SPSS. The frequency of accidents has been established and trend of frequency of involvement in the road accident by the registered vehicles and population has been statistically formulated. Here dependent variable is number of accident and independent variables are registered Vehicle and population. Finally, it is found that regression model data is closed to the collected accident data

Boakye et al. (2013) showed statistical evidence of relationship between road traffic accidents and population growth in Ghana in order to ascertain additional information in contributing to previous researches that have emerged in dealing with this menace. Time series data on yearly road traffic accidents and population values for Ghana covering the period 1990 to 2012 were used. The results from the analysis shows three key findings: a systematic visible pattern of growth in both road traffic accidents and population over the period; evidence of statistical relationship between road traffic accidents and population growth in Ghana as given by the correlation coefficient (r) of 0.854, with a corresponding coefficient of determination
(r-square) of 72.9% indicating that for the period under study based on the available data, the population is able to account for 72.9% of the changes in accidents in Ghana; and finally a regression model developed for the purposes of estimating and forecasting on the basis of the analysis, specifically based on test of hypothesis and model validation.

According to research results of Tortum et al. (2012) in Turkey, driver, pedestrian, vehicle, passenger and road failures in main traffic accidents failure list were ranked based on their effectiveness. It is found that road failures such as road pits, wheel trace, soft shoulders, loose material, permanent wave, deficiency of road signs and road settlements have an important effect on traffic accidents. To reduce road failures, road projects must be planned carefully to meet human needs. In addition to this, road infrastructure must be built according to specific road projects and standards.

In Jordan, Abojaradeh (2013) developed traffic accidents regression prediction models in Amman Greater Area, . These models relate accident numbers, as a dependent variable, with possible causes of accidents that are related to driver behavior, as independent variables. Also, to propose effective counter measures to reduce the frequency and severity of traffic accidents in Jordan. Accident data were collected from the General Security Directorate and from the Jordan Traffic Institute for the selected areas inside Greater Amman Area in Jordan. These data were

analyzed and used in the regression models. Several regression prediction models were formed and the best models were chosen. The intersections and road segments, under this study, were arranged according to the traffic accidents severity. The most dangerous and hazardous streets and intersections were located in the study areas. Proper treatments and improvements are needed to reduce the number and severity of accidents in these areas. Preventive counter measures were recommended to enhance traffic safety in Jordan specially Amman Area.

2.4 Review of Bayesian Analysis

2.4.1 The Origin of Bayesian Analysis

The term *Bayesian* refers to Thomas Bayes (1702–1761), who proved a special case of what is now called Bayes' theorem in a paper titled "An Essay towards solving a Problem in the Doctrine of Chances**".** In that special case, the prior and posterior distributions were Beta distributions and the data came from Bernoulli trials. It was Pierre-Simon Laplace (1749–1827) who introduced a general version of the theorem and used it to approach problems in celestial mechanics, medical statistics, reliability, and jurisprudence. Early Bayesian inference, which used uniform priors following Laplace's principle of insufficient reason, was called **"**inverse probability**"** (because it infers backwards from observations to parameters, or from effects to causes). After the 1920s, "inverse probability" was largely supplanted by a collection of methods that came to be called frequentist statistics.

In the 20th century, the ideas of Laplace were further developed in two different directions, giving rise to *objective* and *subjective* currents in Bayesian practice. In the objectivist stream, the statistical analysis depends on only the model assumed and the data analysed. No subjective decisions need to be involved. In contrast, "subjectivist" statisticians deny the possibility of fully objective analysis for the general case.

In the 1980s, there was a dramatic growth in research and applications of Bayesian methods, mostly attributed to the discovery of Markov Chain Monte Carlo methods, which removed many of the computational problems, and an increasing interest in nonstandard, complex applications. Despite the growth of Bayesian research, most undergraduate teaching is still based on frequentist statistics. Nonetheless, Bayesian methods are widely accepted and used, such as in the fields of machine learning and talent analytics.

2.4.2 Review of Bayesian Analyses of Hierarchical (Multilevel) Models

Bayesian analyses of hierarchical (multilevel) linear models have been considered for at least forty years (Hill, 1965) and have remained a topic of theoretical and applied interest. On the theoretical side, hierarchical models allow a more "objective" approach to inference by estimating the parameters of prior distributions from data rather than requiring them to be specified using subjective information (see Efron and Morris, 1975). At a practical level, hierarchical models are flexible tools for combining information and partial pooling of inferences (see Gelman et al., 2003).

Browne and Draper (2005) reviewed much of the extensive literature in the course of comparing Bayesian and non-Bayesian inference for hierarchical models. As part of their article, Browne and Draper considered some different prior distributions for variance parameters. They used simulation studies, whose design is realistic for educational and medical research to compare Bayesian and likelihood-based methods for fitting variance-components (VC) and random-effects logistic regression (RELR) models. The likelihood approaches they examined are based on the methods most widely used in current applied multilevel (hierarchical) analyses: maximum likelihood (ML) and restricted ML (REML) for Gaussian outcomes, and marginal and penalized quasi-likelihood (MQL and PQL) for Bernoulli outcomes. Their Bayesian methods applied Markov chain Monte Carlo (MCMC) estimation, with adaptive hybrid Metropolis-Gibbs sampling for RELR models, and several diffuse prior distributions.

The Junior School Project (JSP; Mortimore et al., 1988; Woodhouse et al., 1995) was a longitudinal study of about 2,000 pupils from 50 primary schools chosen randomly from the 636 Inner London Education Authority (ILEA) schools in 1980. A variety of measurements were made on the students during the four years of the study, including background variables (such as gender, age at entry, ethnicity, and social class) and measures of educational outcomes such as mathematics test scores (on a scale from 0 to 40) at year 3 (math3) and year 5 (math5). A principal goal of the study was to establish whether some schools were more effective than others in promoting pupils' learning and development, after adjusting for background differences.

The 1987 Guatemalan National Survey of Maternal and Child Health (Pebley and Goldman,

1992) was based on a multistage cluster sample of 5,160 women aged 15 – 44 years living in 240 communities, with the goal of increased understanding of the determinants of health for mothers and children in the period during and after pregnancy. The data have a three-level structure – births within mothers within communities – and one analysis of particular interest estimated the probability of receiving modern (physician or trained nurse) prenatal care as a function of covariates at all three levels. Rodriguez and Goldman (1995) studied a subsample of 2,449 births by 1,558 women who (a) lived in the 161 communities with accurate cluster-level information and (b) had some form of prenatal care during pregnancy.

An article by Jeong-Hun (2007) explored the performance of a Bayesian application of spatial voting models to the roll calls of the Fifth European Parliament (EP). Focusing on two distinct voting behaviours of members of the EP (MEPs) – high absenteeism and the defection from majority alternatives caused by the influence of national parties – it shows that the Bayesian method is complementary to the standard NOMINATE method. In general, the two methods produced very similar estimates and work as robustness checks for the results from each other. However, the Bayesian method enabled the researcher to measure the uncertainty of estimates resulting from the estimation with a large number of missing data and some randomappearing roll calls. In this way, it helps us draw more confident inferences about MEPs‟ voting behaviour.

Patrick (2001) demonstrated the application of multilevel modeling to one of the most common issues that confront institutional researchers: that of student attrition, where the response variable is typically binary rather than continuous. Comparisons are made with a traditional logistic regression approach. The data pertain to one large university. The techniques illustrated may be extended to the analysis of data sets encompassing many institutions, making meaningful inter-institutional comparisons of performance feasible even when there is hierarchical clustering present in the data.

An analytical and software advances can be used to demonstrate that a broad class of Multilevel linear models (MLMs) may be estimated as structural equation models (Bauer, 2003). Moreover, within the structural equation model (SEM) approach it is possible to include measurement models for predictors or outcomes, and to estimate the mediational pathways among predictors explicitly, tasks which are currently difficult with the conventional approach to multilevel modeling. The equivalency of the SEM approach with conventional methods for estimating MLMs is illustrated

using empirical examples, including an example involving both multiple indicator latent factors for the outcomes and a causal chain for the predictors. The limitations of this approach for estimating MLMs are discussed and alternative approaches are considered.

Meta-analysis is formulated as a special case of a multilevel (hierarchical data) model in which the highest level is that of the study and the lowest level is that of an observation on an individual respondent. Studies can be combined within a single model where the responses occur at different levels of the data hierarchy and efficient estimates are obtained. An example is given by Goldstein and Yang (2000) from the studies of class sizes and achievement in schools, where study data are available at the aggregate level in terms of overall mean values for classes of different sizes, and also at the student level.

The extent and nature of contextual effects on juvenile offending are frequent subjects of current research, mainly in the USA. Oberwittler (2004) presented empirical results of a new study which hints at the existence of neighbourhood contextual effects on serious offending by adolescents. The study is based on three types of cross-sectional data on 61 neighbourhoods in two German cities and a rural area: a self-report survey of students aged about 13 to 16, a separate survey of residents in the survey neighbourhoods, and census and administrative data on the same neighbourhoods. Multilevel analysis was applied to identify and explain the neighbourhood-level variance of self-reported serious juvenile offending. Hypotheses from both main traditions of theoretical reasoning about contextual effects on juvenile delinquency – sub-cultural and disorganization theories – are supported by the empirical findings. The spatial concentration of adolescents with attitudes typical of delinquent subcultures increases the likelihood of serious offending net of relevant individual predictors, whereas the social capital of neighbourhoods (as measured by the independent survey of residents) reduces it.

Bayesian methods have become widespread in marketing literature. We review the essence of the Bayesian approach and explain why it is particularly useful for marketing problems. While the appeal of the Bayesian approach has long been noted by researchers, recent developments in computational methods and expanded availability of detailed marketplace data has fueled the growth in application of Bayesian methods in marketing. Rossi and Allenby (2003), in their paper titled Bayesian Statistics and Marketing, emphasized the modularity and flexibility of modern Bayesian approaches. The usefulness of Bayesian methods in situations in which there is limited

information about a large number of units or where the information comes from different sources is noted.

Walker et al. (2007) reviewed the concepts and methods of Bayesian statistical analysis, which can offer innovative and powerful solutions to some challenging analytical problems that characterize developmental research. In the article, the researchers demonstrated the utility of Bayesian analysis, explained its unique adeptness in some circumstances, addressed some concerns and misconceptions about the approach, and illustrated some applications of Bayesian analysis to issues that frequently arise in developmental research. The illustrations of the approach used reflect several important issues within the domain of moral reasoning development (such as assessing patterns of stage change over time); however, the methods are readily applicable across content areas in developmental research.

In applications of hierarchical models (HMs), a potential weakness of empirical Bayes estimation approach is that they do not to take into account uncertainty in the estimation of the variance components (see, e.g., Dempster et al., 1987). One possible solution entails employing a fully Bayesian approach, which involves specifying a prior probability distribution for the variance components and then integrating over the variance components as well as other unknowns in the HM to obtain a marginal posterior distribution of interest (see, e.g., Draper, 1995). Though the required integrations are often exceedingly complex, Markov-chain Monte Carlo techniques (e.g., the Gibbs sampler) provide a viable means of obtaining marginal posteriors of interest in many complex settings. Seltzer et al. (1996) in their article, fully generalized the Gibbs sampling algorithms presented in Seltzer (1993) to a broad range of settings in which vectors of random regression parameters in the HM (e.g., school means and slopes) are assumed multivariate normally or multivariate *t*-distributed across groups. Through analyses of the data from an innovative mathematics curriculum, the researchers examined when and why it becomes important to employ a fully Bayesian approach and discuss the need to study the sensitivity of results to alternative prior distributional assumptions for the variance components and for the random regression parameters.

2.4.3 Review of Markov Chain Monte Carlo methodology

Markov chain Monte Carlo (MCMC), which originated from Metropolis et. al. (1953), is a generic method for approximate sampling from an arbitrary distribution. The idea of MCMC is that for an arbitrary distribution of interest, one can generate a Markov chain whose limiting distribution is equal to the desired distribution. In its simplest form, the Monte Carlo method is nothing more than a computer-based exploitation of the Law of Large Numbers to estimate a certain probability or expectation.

Markov Chain Monte Carlo is useful because it is often much easier to construct a Markov chain with a specified stationary distribution than it is to work directly with the distribution itself.

At the heart of any Monte Carlo method is a random number generator: a procedure that produces an infinite stream

$$U_1, U_2, \ldots$$

of random variables that are independent and identically distributed (i.i.d.) according to some probability distribution. When this distribution is the uniform distribution on the interval (0, 1) (that is, Dist = U(0, 1)), the generator is said to be a **uniform random number generator**. Computer-generated random numbers are sometimes called pseudorandom numbers, we will refer to them simply as random numbers. In MATLAB, for example, this is provided by the **rand** function. The user is typically requested only to input an initial number, called the **seed**, and upon invocation the random number generator produces a sequence of independent uniform random variables on the interval (0, 1).

The most prominent MCMC algorithms are:

(1) The *Metropolis-Hastings* algorithm and in particular the independence sampler and random walk sampler;

(2) The *Gibbs sampler*, which is particularly useful in Bayesian analysis;

(3) The *hit-and-run* sampler – commonly used in Bayesian settings with a highly constrained parameter space and for generic rare-event simulation problems;

(4) The *shake-and-bake* algorithm – a practical approach for generating points uniformly distributed on the surface of a polytope;

(5) *Metropolis – Gibbs* hybrids and the *multiple-try Metropolis-Hastings* method, in which ideas from different MCMC algorithms are combined;

(6) *Auxiliary variable samplers* such as the slice sampler and the *Swendsen-Wang algorithm*;

(7) The *reversible-jump* sampler, which has applications in Bayesian model selection.

Markov Chain Monte Carlo (MCMC) algorithms – such as the Metropolis-Hastings algorithm (Metropolis et al., 1953 and Hastings, 1970) and the Gibbs sampler (Geman and Geman 1984, Gelfand and Smith, 1990) – have been an extremely popular tool in statistics [see for example the recent reviews Smith and Roberts (1993), Tierney (1994), Gilks et al. (1996a)]. In addition to the large body of applied work which uses these mothods, there has been a substantial amount of progress on the theoretical aspects of these algorithms. To the applied user, it is often unclear what lessons (if any) can be learned from these theoretical results.

Brooks (1998) provided a simple, comprehensive and tutorial review of some of the most common areas of research in this field. The article discussed how MCMC algorithms can be constructed from standard building-blocks to produce Markov chains with the desired stationary distribution. It also discussed more complex ideas that have been proposed in the literature, such as continuous time and dimension jumping methods. Some implementational issues associated with MCMC methods were also mentioned. The paper looked at the arguments for and against multiple replications, considered how long chains should be run for and how to determine suitable starting points. Graphical models and how graphical approaches can be used to simplify MCMC implementation was discussed. Finally, the study presented a couple of examples, which were used as case-studies to highlight some of the points made earlier in the text. In particular, he used a simple change-point model to illustrate how to tackle a typical Bayesian modelling problem via the MCMC method, before using mixture model problems to provide illustrations of good sampler output and of the implementation of a reversible jump MCMC algorithm.

Statistical methods of inference typically require the likelihood function to be computable in a reasonable amount of time. The class of “likelihood-free" methods termed Approximate Bayesian Computation (ABC) are able to eliminate this requirement, replacing the evaluation of the likelihood with simulation from it. Likelihood-free methods have gained in efficiency and popularity in the past few years, following their integration with Markov Chain Monte Carlo (MCMC) and Sequential Monte Carlo (SMC) in order to better explore the parameter space. They have been applied primarily to estimating the parameters of a given model, but can also be used to compare models. Didelot et al. (2011) presented novel likelihood-free approaches to model

comparison, based upon the independent estimation of the evidence of each model under study. Key advantages of these approaches over previous techniques are that they allow the exploitation of MCMC or SMC algorithms for exploring the parameter space, and that they do not require a sampler able to mix between models. They validated the proposed methods using a simple exponential family problem before providing a realistic problem from human population genetics: the comparison of different demographic models based upon genetic data from the Y chromosome.

Geochemical signatures deposited in otoliths are a potentially powerful means of identifying the origin and dispersal history of fish. However, current analytical methods for assigning natural origins of fish in mixed-stock analyses require knowledge of the number of potential sources and their characteristic geochemical signatures. Such baseline data are difficult or impossible to obtain for many species. A new approach to this problem can be found in iterative Markov Chain Monte Carlo (MCMC) algorithms that simultaneously estimate population parameters and assign individuals to groups. MCMC procedures only require an estimate of the number of source populations, and post hoc model selection based on the deviance information criterion can be used to infer the correct number of chemically distinct sources. White (2008) et al described the basics of the MCMC approach and outlined the specific decisions required when implementing the technique with otolith geochemical data. The researchers also illustrated the use of the MCMC approach on simulated data and empirical geochemical signatures in otoliths from young-of-the-year and adult weakfish, Cynoscion regalis, from the U.S. Atlantic coast. While the study described how investigators can use MCMC to complement existing analytical tools for use with otolith geochemical data, the MCMC approach is suitable for any mixed-stock problem with a continuous, multivariate data.

2.4.4 Recent History of Bayesian Statistical Software

In the last 20 years, Bayesian statistical software has emerged from humble beginnings to the powerful applications that we have today. There is no doubt that this trend will continue. In the near future, we expect to see new theoretical advances, better software and faster hardware. Sparapani and Laud (2008) reviewed some of the history of the Bayesian statistical software.

The era of modern Bayesian statistical computation can be said to begin with the paper by Gelfand and Smith (1990). The construction of general purpose computational software, however,

begins with three seminal papers by Gilks et al. (1995) and Spiegelhalter et al. (1995). These advances culminated in the release of the free software package BUGS (Bayesian inference Using Gibbs Sampling, Spiegelhalter et al., 1995). BUGS software had two components: a model specification language, and a command language that could be utilized either interactively by the command line or in batch via a script file. BUGS was available for many Unix platforms as well as Linux and MS-DOS. BUGS relied on other software like the free software R to create input data file(s) and to analyze its output data files such as the R packages CODA (Convergence Diagnosis and Output Analysis for MCMC: Plummer at al. 2006) or BOA (Bayesian Output Analysis: Smith, 2005). BUGS although still available, is no longer maintained. BUGS was succeeded by the free software package WinBUGS (Lunn et al., 2000).

WinBUGS is only available for MS Windows and is based on the BlackBox Component Builder developed by Oberon microsystems, a component-based development environment for the programming language Component Pascal. The model specification language is largely the same as that of BUGS, and WinBUGS still relies on other software to create input data file(s) as before. Interactive use is handled by the GUI of WinBUGS. Batch processing is handled by a new WinBUGS command language which is not the same as the BUGS command language. WinBUGS also provides its own convergence diagnostics via the Gelman-Rubin Statistics (1992) while still allowing you to create output data files to analyze as in the past. The R package R2WinBUGS (Sturtz et al., 2005) is a work-in-progress that manages the whole process from R: submitting the data and model file to WinBUGS, batch processing the MCMC sampling in WinBUGS and returning the samples to R. Although WinBUGS is an MS Windows application, it is currently possible to run it on other x86 platforms, like Unix/Linux and Mac OS X, via Wine, a free software, open source implementation of the MS Windows API for X11/OpenGL (and R2WinBUGS can take advantage of Wine as well).

WinBUGS is considered to be stable, but it will be phased out in the future. Current development is based on OpenBUGS (Thomas et al., 2006), an open source version of WinBUGS that runs on MS Windows, Linux and as an R package. Although, OpenBUGS is in its early stages, OpenBUGS for MS Windows is quite robust and where new WinBUGS features are appearing. OpenBUGS shares much with WinBUGS including most of what has been described above like convergence diagnostics, R2WinBUGS and Wine. One di_erence is that OpenBUGS does not

share the WinBUGS command language for batch processing, but instead has its own command language which is also not the same as the original BUGS command language. An advantage of OpenBUGS is you don't have to register it annually, something that was a minor irritant with WinBUGS. From here on out, the phrase OpenBUGS will refer to the MS Windows version and all comments will apply equally well to WinBUGS.

SAS ® started out as a statistical analysis software package at a time when there were few options. Over time, SAS also built in capabilities that would facilitate data operations such as capture, management and manipulation. And, it is in this unique framework that SAS has prospered as one of the few annual fee software packages: you don't buy SAS, you “rent” it. SAS combines two levels of data programming: a low-level called the DATASTEP and a high-level known as Procedures or PROCs. SAS also provides the user with the SAS Macro Language: a facility for creating reusable SAS scripts called macros that can also provide high-level Procedure-like functionality.SAS provides two SAS macros (Westfall, 1999), bayestests and bayesintervals, for multiple testing and simultaneous intervals from the posterior sample. Also, with SAS you can perform Bayesian Variance Component analysis (Wolfinger, 2000). And very recently SAS has made available three Bayesian-capable PROCs: GENMOD, PHREG and LIFEREG which will be included in the next release of SAS. These PROCs are available as an experimental download on the MS Windows platform with the names BGENMOD, BPHREG and BLIFEREG. In addition, the user's manual (SAS Institute Inc. 2006) contains a nice introduction to Bayesian statistics. The 40 pages of material, including 7 pages of references, is worth reading for all who are interested in Bayesian statistics, whether they plan on using SAS or not, and whether they are novices or more advanced.

Currently, many Bayesian statisticians use R or SAS for its powerful data manipulation, and penBUGS for the statistical analysis. In this manner, the modeling and inference exibility of OpenBUGS can be combined with the data manipulation and graphical power of R or SAS to explore the Markov chain Monte Carlo samples obtained from OpenBUGS.

For those using SAS, this process is facilitated by the free software, open source SAS macro library called RASmacro (Sparapani, 2004). It is a library of middle-level SAS macros that are the building blocks for high-level SAS macros. RASmacro provides two SAS macros, _lexport and _sexport, to create input data for OpenBUGS. _lexport takes a list of SAS dataset variables and

creates an input data _le of scalars referred to as a “list” data _le. _sexport takes a list of SAS dataset variables and creates an input data _le of vectors referred to as a “structure” data _le. RASmacro also provides two SAS macros, _decoda and _debugs, to process OpenBUGS output _les. _decoda creates a SAS dataset from the OpenBUGS text output _les: the index _le and chain _le(s). _debugs generates posterior statistics and plots, histograms and trace _les (_decoda will call _debugs if statistics and graphics are requested).

2.5 Conclusion

It is more evident from the literature review that regression analysis is a powerful way of developing a model for predicting road traffic fatalities. This study seeks to modify Smeed‟s (1949) formula in order to derive a model for predicting road traffic fatalities in Ghana. The Literature review also revealed that Bayesian and multilevel analyses has not been applied, as predictive model, to study the trend and effect of road traffic accidents.

CHAPTER THREE

METHODOLOGY

3.0 Introduction

This chapter will put forward the derivation of a modified Smeed’s model and also determine how accurate the proposed modified model of this study is. The question to be addressed here is:
how does the modified Smeed’s model compare to that of Smeed (1949) in their performance?

The chapter also explored the concepts and techniques for analyzing and making use of the linear relationship between two variables. This analysis may lead to an equation that can be used to predict the value of a dependent variable given the value of an independent variable.

Based on the modified Smeed’s model, the chapter seeks to develop the methodology of two Bayesian approaches for estimating the regression coefficients. The two methods are Conjugate prior and Maximum a Posteriori.

Finally, using this modified Smeed’s model, a multilevel method is developed to assess the risk of road traffic fatalities (RTFs) across sub-populations of a given geographical zone.

3.1 A Modification of Smeed"s Model

Smeed's model, which is of the form

$$\frac{D}{N} = \alpha \left(\frac{N}{P}\right)^{\beta} e, \qquad \text{.....................................(3.1)}$$

measures the per vehicle fatality rate, D/N, of a geographical region, where D = Number of RTFs, P = population size, N = number of vehicles in use, e = multiplicative error term, and α & β are parameters to be estimated. It was shown, in Chapter One of this study, that the α and β vary from one geographical region to another and thus, could be used to assess variability of risk of RTFs across sub-populations of a given geographical zone. In this section, the study derives a modified Smeed's model.

Multiplying both sides of (3.1) by N/P, we obtain

$$\frac{D}{P} = \alpha \left(\frac{N}{P}\right)^{\beta} \left(\frac{N}{P} e\right). \qquad \text{.....................................(3.2)}$$

The modified Smeed's model of this study, which estimates the ***per capita fatality rate*** (also called [4]***public health risk indicator***), is of the form

$$\frac{D}{P} = \alpha \left(\frac{N}{P}\right)^{\beta} u, \qquad \text{.....................................(3.3)}$$

where $u = \left(\frac{N}{P} e\right) \approx e$ provided $N \approx P$. Table 3.1 is an extract from the list of countries with ranks based on the number of road motor vehicles per 1,000 inhabitants. For every country in the

[4] National Road Safety Commission of Ghana (2011). Building and Road Research Institute (BRRI), *Road Traffic Crashes in Ghana*, Statistics

world, except San Marino, the number of registered vehicles in use, *N*, is less than the population size, *P*.

Table 3.1: List of Countries by the Number of Road Motor Vehicles per 1,000 Population

Rank	Country	$\frac{N}{P}$ 1 000	Rank	Country	$\frac{N}{P}$ 1 000
1	San Marino	1,263	85	South Africa	165
2	Monaco	899	143	Nigeria	31
3	United States	797	145	Ghana	30
31	United Kingdom	519	192	Togo	2

Source: NationMaster: Transport > Road > Motor vehicle per 1000 people

Since $N \approx P$ for most situations, it follows that the multiplicative error term u in the modified Smeed's model of this study is less than that of Smeed's original model, making the modified Smeed's model preferred.

The modified Smeed's model is nonlinear but can be transformed to linear model by using special transformation. Such nonlinear model is called ***intrinsically linear***. Daniel and Wood (1980), Montgomery and Peck (1992), and Myers (1990) give several nonlinear models that are intrinsically linear. Thus, Equation (3.3) can be transformed to a linear model by a logarithmic transformation of the form

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i, \quad i = 1, 2, \cdots, n \quad \text{.....(3.4)}$$

where k is a positive integer.

For example, Equation (3.3) can be written in the form

$$\ln D = \ln \beta + \beta_1 \ln N + \beta_2 \ln P + \ln u \quad \text{.....(3.5)}$$

Alternatively we may write this as

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_{i2} + \varepsilon_i, \quad i = 1, 2, \cdots, n \quad \text{.....(3.6)}$$

where $k = 2$, $y_i = \ln D$, $x_{i1} = \ln N$, $x_{i2} = \ln P$, $\beta_0 = \ln$, β_1 , $\beta_2 = \ln(1-\beta)$ and $\varepsilon_i = \ln u_i$, $i = 1, 2, \ldots, n$. This transformation requires that $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ are normally and independently distributed with mean 0 and variance σ^2[5]. In Equation (3.6), we have introduced an additive random error term ε_i. However, if we refer back to the original Equation (3.3), we see that this is equivalent to assuming a ***multiplicative error*** term u.

Another possible linear transformation of Equation (3.3) is of the form

$$\ln\left(\frac{D}{P}\right) = \beta_0 + \ln\left(\frac{N}{P}\right)^{\beta_1} + \ln u \quad \text{.....(3.7)}$$

which can be expressed as

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad \text{.....(3.8)}$$

where, $k = 1$, $\beta_0 = \ln$, $\beta_1 = \beta_1$, $x_i = \ln\left(\frac{N}{P}\right)$, $y_i = \ln\left(\frac{D}{P}\right)$ and $\varepsilon_i = \ln u_i$, $i = 1, 2, \ldots, n$.

The linear transformation in Equation (3.7) is preferred to that of Equation (3.5) because of the following reason. Since D/P is a risk indicator (known as [2]***Public Health Risk indicator***) used in epidemiological studies, it follows that any one-to-one relation of this indicator, such as $Y = \ln(D/P)$, can also be used as risk indicator of RTF. This is in sync with the general objective of this studies.

3.2 The multiple linear regression model

In the multiple linear regression model, we assume that a linear relationship exists between a variable Y, which we call the ***dependent variable***, and k independent variables, $X_1, X_2, \ldots, X_k$. The independent variables are sometimes referred to as ***explanatory variables*** because of their use in explaining the variation in Y . They are also called ***predictor variables***, because of their use in predicting Y.

[5] National Road Safety Commission of Ghana (2011). Building and Road Research Institute (BRRI), *Road Traffic Crashes in Ghana*, Statistics

3.2.1 The model equation

The multiple linear regression model can be expressed as given in Equation (3.4), where $\beta_0, \beta_1, \ldots, \beta_k$ are called ***partial regression coefficients*** and where, it is assumed that $E(\varepsilon) = 0$ and $V(\varepsilon) = \sigma^2$. The parameter β_j represents the expected change in the response Y per unit change in x_j when all the remaining independent variables x_i $(i \neq j)$ are held constant. The term linear is used because the Equation (3.4) is a linear function of the unknown parameters $\beta_0, \beta_1, \ldots, \beta_k$.

3.2.2 Least squares estimation of parameters

The method of least squares can be used to estimate the regression coefficients in the multiple linear regression model. Suppose $n > k$ observations are available, and let x_{ij} denote the i^{th} observation of variable X_j. The observations are $(x_{i1}, x_{i2}, \ldots, x_{ik}, y_i)$, $i = 1, 2, \ldots, n$. It is customary to present the data for a multiple linear regression in a table such as Table 3.2.

Table 3.2: Data for a multiple linear regression

y	x_1	x_2	…	x_k
y_1	x_{11}	x_{12}	…	x_{1k}
y_2	x_{21}	x_{22}	…	x_{2k}
.	.	.	…	.
.	.	.	…	.
.	.	.	…	.
y_n	x_{n1}	x_{n2}	…	x_{nk}

Each set of observations $(x_{i1}, x_{i2}, \ldots, x_{ik}, y_i)$ satisfies the model in Equation (3.5), or

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i, \quad i = 1, \ldots, n$$

$$= \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + \varepsilon_i, \quad i = 1, \ldots, n \quad \ldots\ldots\ldots\ldots\ldots(3.9)$$

The least squares estimates, $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$, of $\beta_0, \beta_1, \ldots,$ and β_k, are the values of

$\beta_0, \beta_1, \ldots, \beta_k$ which minimize

$$Q = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \ldots - \beta_k x_{ik})^2,$$

the sum of the squares of the deviations of the points from any proposed hyperplane. Thus, the least squares estimates of $\beta_0, \beta_1, \ldots, \beta_k$ must satisfy

$$\left.\frac{\partial Q}{\partial \beta_0}\right|_{\hat\beta_0, \hat\beta_1, \ldots, \hat\beta_k} = -2\sum_{i=1}^{n}\left(y_i - \hat\beta_0 - \sum_{j=1}^{k}\hat\beta_j x_{ij}\right) = 0 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots..(3.10)$$

and

$$\left.\frac{\partial Q}{\partial \beta_j}\right|_{\hat\beta_0, \hat\beta_1, \ldots, \hat\beta_k} = -2\sum_{i=1}^{n}\left(y_i - \hat\beta_0 - \sum_{j=1}^{k}\hat\beta_j x_{ij}\right)x_{ij} = 0, \quad j = 1, 2, \ldots, k \quad \ldots\ldots..\ldots\ldots..(3.11)$$

Simplifying Equations (3.7) and (3.8), we obtain the least squares normal equations

$$\left.\begin{aligned}
n\hat\beta_0 + \hat\beta_1\sum_{i=1}^{n} x_{i1} + \hat\beta_2\sum_{i=1}^{n} x_{i2} + \cdots + \hat\beta_k\sum_{i=1}^{n} x_{ik} &= \sum_{i=1}^{n} y_i \\
\hat\beta_0\sum_{i=1}^{n} x_{i1} + \hat\beta_1\sum_{i=1}^{n} x_{i1}^2 + \hat\beta_2\sum_{i=1}^{n} x_{i1}x_{i2} + \cdots + \hat\beta_k\sum_{i=1}^{n} x_{i1}x_{ik} &= \sum_{i=1}^{n} x_{i1}y_i \\
\vdots \qquad\qquad \vdots \qquad\qquad \vdots \qquad\qquad & \quad \vdots \\
\hat\beta_0\sum_{i=1}^{n} x_{ik} + \hat\beta_1\sum_{i=1}^{n} x_{ik}x_{i1} + \hat\beta_2\sum_{i=1}^{n} x_{ik}x_{i2} + \cdots + \hat\beta_k\sum_{i=1}^{n} x_{ik}^2 &= \sum_{i=1}^{n} x_{ik}y_i
\end{aligned}\right\} \ldots\ldots\ldots\ldots(3.12)$$

The fitted regression plane is then

$$\hat y = \hat\beta_0 + \hat\beta_1 x_1 + \hat\beta_2 x_2 + \cdots + \hat\beta_k x_k \,. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(3.13)$$

Notice that there are $p = k + 1$ normal equations, one for each of the unknown regression coefficients. The solution to the normal equations will be the least squares estimates of the regression coefficients, $\hat\beta_0, \hat\beta_1, \ldots,$ and $\hat\beta_k$. The normal equations can be solved by any method

appropriate for solving a system of linear equations. Snedecor and Cochran (1989) and Steel and Torrie (1979) give numerical examples for four variables and Anderson and Bancroft (1952) illustrate the calculations involved when there are five variables.

3.2.3 The matrix approach to multiple linear regression

In fitting a multiple linear regression model, particularly when the number of variables exceeds two, a knowledge of matrix theory can facilitate the mathematical manipulation considerably.

Suppose that the experimenter has k regressor variables and n observations, $(x_{i1}, \ldots, x_{ik}, y_i)$, $i = 1, 2, \ldots, n$, and the model relating the regressors to the response is given by Equation (3.6). The model is a system of n equations that can be expressed in matrix notation as

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{Y}$ is an $n \times 1$ vector, $\boldsymbol{X}$ is an $n \times (k+1)$ matrix, $\boldsymbol{\beta}$ is a $(k+1)$-dimensional column vector, and $\boldsymbol{\varepsilon}$ is an n-dimensional column vector. That is,

$$\boldsymbol{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \text{ and } \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}.$$

It is assumed that $E(\boldsymbol{\varepsilon}) = \mathbf{0}$, $V(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{I}$, where $V(\boldsymbol{\varepsilon})$ denotes the variance-covariance matrix of $\boldsymbol{\varepsilon}$.

Notice that the elements of the vector $\boldsymbol{Y}$ are the observed responses, while the elements of the matrix $\boldsymbol{X}$ are the observed values of the explanatory variables. The following theorem gives important results for finding the least squares estimator of $\boldsymbol{\beta}$ and for proving Theorem 12.3. A proof of the theorem is given by Rao (1973).

Theorem 3.1

$\frac{\partial}{\partial X}(B'X) = B$, $\frac{\partial}{\partial X}(X'AX) = 2AX$, $V(A'W) = A'V(W)A$, where B is a constant vector, A is a constant matrix and $V(W)$ is the variance-covariance matrix of the random variables in the vector W (see Rao, 1973).

Least squares estimates of parameters

The method of least squares seeks an estimate of β that minimizes the sum of squared deviations between the fitted and observed responses. Thus, we wish to find the value of β that minimizes

$$Q(\beta) = \sum_{i=1}^{n} \varepsilon_i^2 = \varepsilon'\varepsilon = (Y - X\beta)'(Y - X\beta).$$

The least squares estimate, $\hat{\beta}$, of β is therefore the solution of β in the equation

$$\frac{\partial Q}{\partial \beta} = \mathbf{0}. \quad \text{……………………………………………………………………………….(3.14)}$$

Note that $Q(\beta)$ can be expressed as

$$Q(\beta) = Y'Y - \beta'X'Y - Y'X\beta + \beta'X'X\beta$$
$$= Y'Y - 2Y'X\beta + \beta'X'X\beta,$$

since $\beta'X'Y$ is a 1×1 matrix or a scalar, and its transpose $(\beta'X'Y)' = Y'X\beta$ is the same scalar.

The least squares estimate of β must satisfy.

$$\left.\frac{\partial Q}{\partial \beta}\right|_{\hat{\beta}} = -2X'Y + 2X'X\hat{\beta} = 0,$$

which simplifies to

$$\boldsymbol{X}\quad\boldsymbol{X}\quad\hat{\quad}\quad\quad\boldsymbol{X}\quad\boldsymbol{Y}.$$

……(3.15)

Equations (3.12) are the least squares normal equations in matrix form. They are identical to the scalar form of the normal equations given earlier in Equations (3.9). To solve the normal equations, pre-multiply both sides of Equations (3.12) by the inverse of $\boldsymbol{X}\boldsymbol{X}$. This gives the least squares estimate of as

$$\hat{} = (\boldsymbol{X}\boldsymbol{X})^{-1}\boldsymbol{X}\boldsymbol{Y}.$$

……(3.16)

The solution assumes that the matrix $\boldsymbol{X}\boldsymbol{X}$ is nonsingular and so $(\boldsymbol{X}\boldsymbol{X})^{-1}$exists. The inverse, $(\boldsymbol{X}\boldsymbol{X})^{-1}$, exists if the regressors are ***linearly independent***, that is, if no column of $\boldsymbol{X}$ is a linear combination of the other columns.

Techniques for finding $(\boldsymbol{X}\boldsymbol{X})^{-1}$ are explained in many textbooks on elementary determinants and matrices. There are also many high-speed computer packages available for multiple regression.

It is easy to see that the matrix form of the normal equations is identical to the scalar form. Writing out Equations (3.12) in detail, we obtain

$$\begin{bmatrix} n & \sum x_{i1} & \sum x_{i2} & \cdots & \sum x_{ik} \\ \sum x_{i1} & \sum x_{i1}^2 & \sum x_{i1}x_{i2} & \cdots & \sum x_{i1}x_{ik} \\ \vdots & \vdots & \vdots & & \vdots \\ \sum x_{ik} & \sum x_{ik}x_{i1} & \sum x_{ik}x_{i2} & \cdots & \sum x_{ik}^2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_{i1}y_i \\ \vdots \\ \sum x_{ik}y_i \end{bmatrix}.$$

If the indicated matrix multiplication is performed, the scalar form of the normal equations (that is, Equations (3.9)) will result. In this form, it is easy to see that $\boldsymbol{X}\boldsymbol{X}$ is a $p \times p$ symmetric matrix and $\boldsymbol{X}\boldsymbol{Y}$ is a $p \times 1$ column vector, where $p = k + 1$. Note the special structure of the matrix $\boldsymbol{X}\boldsymbol{X}$. The diagonal elements are the sums of squares of the elements in the columns of $\boldsymbol{X}$, and the off-diagonal elements are the sums of cross-products of the elements of the

columns of $\boldsymbol{X}$. Furthermore, the elements of $\boldsymbol{X'Y}$ are the sums of the cross products of the columns of $\boldsymbol{X}$ and $\boldsymbol{Y}$.

The fitted regression model is

$$\hat{y}_i = \hat{\beta}_0 + \sum_{j=1}^{k} \hat{\beta}_j x_{ij}, \quad i = 1, 2, \ldots, n \qquad \text{.....(3.17)}$$

In matrix notation, the fitted model is

$$\boldsymbol{\hat{Y} = X\hat{\beta}} = \boldsymbol{X(X'X)^{-1}X'Y.} \qquad \text{.....(3.18)}$$

The difference between the observation y_i and the fitted values $\hat{y}_i$, is a residual, say $e_i = y_i - \hat{y}_i$. The $(n \times 1)$ vector of residuals is denoted by

$$\boldsymbol{e = Y - \hat{Y}} \qquad \text{.....(3.19)}$$

The statistical properties of the least squares estimators $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ may be easily found, under certain assumptions on the error terms $\beta_1, \beta_2, \ldots, \beta_k$, in the regression model.

Theorem 3.2

If $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ then $\hat{\boldsymbol{\beta}}$ is an unbiased estimator of $\boldsymbol{\beta}$ (see Ofosu et al. 2014).

Theorem 3.3 (The covariance of $\hat{\beta}$)

If $V(\boldsymbol{\varepsilon}) = \sigma^2\boldsymbol{I}$, where $\boldsymbol{I}$ is a $(k+1) \times (k+1)$ identity matrix, then $V(\hat{\boldsymbol{\beta}}) = \sigma^2(\boldsymbol{X'X})^{-1}$.

Therefore, the diagonal elements of $\sigma^2(\boldsymbol{X'X})^{-1}$ are the variances of $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ and the offdiagonal elements of the matrix are the covariances (see Ofosu et al. 2014).

Theorem 3.4 (The Gauss-Markov theorem)

The least squares estimator of β, $\hat{\beta} = (X'X)^{-1} X'Y$, is the best linear unbiased estimator (BLUE) of β (see Montgomery et al., 2006).

3.2.4 Polynomial regression

The linear regression model, $Y = X\beta + \varepsilon$, is a general model that can be used to fit any relationship that is linear in the unknown parameters, β_0, β_1, . . ., and β_k. This includes the important class of polynomial regression models. For example, the second-degree polynomial in one variable,

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon,$$

is a linear regression model. In this section, we consider polynomial regression models. In general, the k^{th} order polynomial regression model in one variable is

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_k x^k + \varepsilon. \quad \text{……………………………………………(3.20)}$$

Polynomial regression models are widely used when the response is curvilinear, because the general principles of multiple regression can be applied. For example, in equation (3.20), if we set $x_j = x^j$, $j = 1, 2, \ldots, k$, then (3.20) becomes a multiple linear regression model. Confusion sometimes arises when we speak of a polynomial regression model as a linear model. However, statisticians normally refer to a linear regression model as one in which the parameters occur linearly, regardless of how the independent variables occur.

Quadratic regression

Here, we assume that

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon, \quad \text{……………………………………………(3.21)}$$

where $E(\varepsilon_i) = 0$. Given a set of data consisting of n points (x_i, y_i), $i = 1, 2, \ldots, n$, we obtain the least squares estimates of β_0, β_1 and β_2 by minimizing

$$Q = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i - \beta_2 x_i^2)^2.$$

Now,

$$\frac{\partial Q}{\partial \beta_0} = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i - \beta_2 x_i^2),$$

$$\frac{\partial Q}{\partial \beta_1} = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i - \beta_2 x_i^2) x_i,$$

$$\frac{\partial Q}{\partial \beta_2} = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i - \beta_2 x_i^2) x_i^2.$$

Equating the partial derivatives to zero and replacing β_0, β_1 and β_2 by $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$, we obtain the least squares normal equations as

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_i + \hat{\beta}_2 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i,$$

$$\hat{\beta}_0 \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 + \hat{\beta}_2 \sum_{i=1}^{n} x_i^3 = \sum_{i=1}^{n} y_i x_i,$$

$$\hat{\beta}_0 \sum_{i=1}^{n} x_{i2} + \hat{\beta}_1 \sum_{i=1}^{n} x_{i3} + \hat{\beta}_2 \sum_{i=1}^{n} x_{i4} + \sum_{i=1}^{n} x_{i2} y$$

$$x_{i2}. \quad i=1 \quad i=1 \quad i=1 \quad i=1$$

These equations can be solved by means of a scientific calculator.

Cubic and higher order polynomial regression

A cubic equation involves one extra term $\beta_3 x^3$ in the model defined by Equation (3.21) which now becomes

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \varepsilon \quad \text{.....(3.22)}$$

The least squares estimates of β_0, β_1, β_2 and β_3 can be found by minimizing

$$Q = \sum_{i=1}^{n} \left(y_i - \beta_0 - \beta_1 x_i - \beta_2 x_i^2 - \beta_3 x_i^3 \right)^2 \quad \text{.....(3.23)}$$

with respect to β_0, β_1, β_2 and β_3.

In a similar manner to the above, higher order equations can be fitted. The work involved in solving the normal equations increases considerably as the number of terms in the model increases. It is therefore highly desirable to have a computer package available to fit the required model.

If a computer is not being used, the fitting of a high order polynomial can be facilitated by making use of orthogonal polynomials. For details of the use of orthogonal polynomials, see Johnson and Leone (1976).

3.3 Bayesian Approach

3.3.1 Introduction

Thus far, the researcher has assumed that the regression coefficients β_0, β_1, and β_k are fixed unknown parameters which lies in a parameter space Ω. Based on the sample, inferences can be made about $\boldsymbol{\beta}$.

In this section, the study considers the situation where, before a sample is taken, some information about **β** is known. It is assumed that this knowledge about **β** can be expressed in the form of a probability distribution over □. The multiple linear regression model, with *k* predictor variables, in Equation (3.4), can be expressed as

$y_i = \beta \boldsymbol{x}_i + \varepsilon_i, \quad i = 1, 2, ..., n$ .....(3.24) where

$\boldsymbol{x}_i = (1, x_{1i}, x_{2i}, ..., x_{ki})$. It is assumed that the unknown parameter vector $\boldsymbol{\beta} = (\beta_0, \beta_1, ... \beta_k)$ is a value of some multivariate random variable with a multivariate prior distribution.

The range of possible values that the regression coefficients $\beta_0, \beta_1, ... \beta_k$ can take is $-\infty$ to $+\infty$. Thus, the largest possible domain of the prior distribution is the set of all real numbers. This limits us to distribution which can take both negative and positive values. Therefore, the most suitable prior distributions are the bivariate Normal, Laplace and Cauchy distributions.

Two Bayesian methods were used in estimating the parameters in Equation (3.24) These are the 'conjugate prior' method and the maximum a posteriori method which are discussed in the following sequel.

3.3.2 „Conjugate Prior" Method

In this section, we assume that the random variable *Y*, with components y_i, in Equation (3.24), has the normal distribution with mean $\boldsymbol{\beta}\boldsymbol{x}$ and variance σ^2. Thus, the likelihood function will also follow a normal distribution. Since the normal distribution is conjugate to itself (or *selfconjugate*) with respect to a normal likelihood function, choosing a bivariate normal prior over **β** will ensure that the posterior distribution is also normal. The conditional p.d.f. of *Y* is then given by

$$f_Y(y_i \mid \boldsymbol{\beta}, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(y_i - \boldsymbol{\beta}\boldsymbol{x})^2\right], \quad y_i = 0. \quad \text{.....(3.25)}$$

The likelihood function is given by

$$f_{\mathbf{Y}}(\mathbf{y}|\boldsymbol{\beta}) = (2\pi)^{-n/2}\exp\left(-\frac{1}{2}\sum_{i=1}^{n}(y_i-\boldsymbol{\beta}'\mathbf{x}_i)^2\right),$$

$$y_i = 0.1\exp(2.2i+1)+\varepsilon_i$$

$$= (2\pi)^{-n/2}\exp\left(-\frac{1}{2}\sum_{i=1}^{n}(y_i-\boldsymbol{\beta}'\mathbf{x}_i)^2\right)$$

$$= k_1\exp\left(-\frac{1}{2}(\mathbf{y}-\boldsymbol{\beta}\mathbf{x})'(\mathbf{y}-\boldsymbol{\beta}\mathbf{x})\right), \quad \mathbf{y} = (y_1, y_2, ..., y_n). \quad \text{.....(3.26)}$$

where $k_1 = (2\pi)^{-n/2}$. It is assumed that $\boldsymbol{\beta}$ has a multivariate normal distribution with mean vector $\boldsymbol{\mu} = (0, 1, ..., k)$ and covariance matrix $\boldsymbol{\Sigma}$. Thus, the p.d.f. of $\boldsymbol{\beta}$ is

$$p(\boldsymbol{\beta}) = (2\pi)^{-k/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\left(-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu})\right)$$

.....(3.27)

where $\boldsymbol{\Sigma} = \begin{pmatrix} a_{00} & a_{01} & a_{02} & \cdots & a_{0k} \\ a_{10} & a_{11} & a_{12} & \cdots & a_{1k} \\ a_{20} & a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ a_{k0} & a_{k1} & a_{k2} & \cdots & a_{kk} \end{pmatrix}$. Thus, Equation (3.27) can expressed as

$$p(\boldsymbol{\beta}) = k_2\exp\left(-\frac{1}{2}(\beta_0-\mu_0, \beta_1-\mu_1, ..., \beta_k-\mu_k)\begin{pmatrix} a_{00} & a_{01} & a_{02} & \cdots & a_{0k} \\ a_{10} & a_{11} & a_{12} & \cdots & a_{1k} \\ a_{20} & a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ a_{k0} & a_{k1} & a_{k2} & \cdots & a_{kk} \end{pmatrix}\begin{pmatrix} \beta_0-\mu_0 \\ \beta_1-\mu_1 \\ \vdots \\ \beta_k-\mu_k \end{pmatrix}\right)$$

□ □

□□ □□a_{k0} a_{k1} a_{k2} a_{kk} □□□□ □□$_{kk}$ □□□

□

$= k$ □ □ k $_2$ k□1 k □□□

$_2$ exp□□□□$^{1}_{2}$□□□/□□0a_{ll} □□ □□$_{ll}$ □ □ 2 $_j$□□ □ □0$_s$ $_j$□ 1a_{js} □□ □□$_j$ $_j$ □□□ □□$_s$

$_s$ □□□□□□□, ...(3.28)

where k_2 □ $_2^{1}$□ Σ^{-1}□$_2$. The joint pdf of $\boldsymbol{Y}$ □ (Y Y_1, $_2$, ..., Y_n) and □ is

$$f(\boldsymbol{y}, \boldsymbol{\beta}) = f(\boldsymbol{y}|\boldsymbol{\beta})\, p(\boldsymbol{\beta}). \quad \text{.....(3.29)}$$

The marginal p.d.f. of $\boldsymbol{Y}$ □ (Y Y_1, $_2$, ..., Y_n) is given by

$$f_{\boldsymbol{Y}}(y_1, y_2, \ldots, y_n) = \int f(\boldsymbol{y}|\boldsymbol{\beta})\, p(\boldsymbol{\beta})\, d\boldsymbol{\beta} \quad \text{.....(3.30)}$$

□

The posterior distribution is the conditional p.d.f. of □$_j$ given $\boldsymbol{Y_j}$ □ $\boldsymbol{y_j}$ which is given by

$$p(\boldsymbol{\beta}|\boldsymbol{y}) = \frac{f(\boldsymbol{y}, \boldsymbol{\beta})}{f_{\boldsymbol{Y}}(y_1, y_2, \ldots, y_n)} = \frac{f(\boldsymbol{y}|\boldsymbol{\beta})\, p(\boldsymbol{\beta})}{f_{\boldsymbol{Y}}(y_1, y_2, \ldots, y_n)}$$

$$= k f_3(\boldsymbol{y}|\boldsymbol{\beta})\, p(\boldsymbol{\beta}). \quad \text{.....(3.31)}$$

where the value of k_3 does not depend on $\boldsymbol{\beta}$. From Equations (3.26) and (3.28), the posterior distribution can therefore be expressed as

$$p(\boldsymbol{\beta}|\boldsymbol{y}) = k_3 \, k_1 \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \boldsymbol{\beta}\boldsymbol{x}_i)^2\right] k_2 \exp\left[-\frac{1}{2} (\boldsymbol{\beta} - \boldsymbol{\mu}_0)' \boldsymbol{\Sigma}_0^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu}_0)\right]$$

$$= k_4 \exp\left[-\frac{1}{2} \left\{ \frac{1}{\sigma^2} (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) + (\boldsymbol{\beta} - \boldsymbol{\mu}_0)' \boldsymbol{\Sigma}_0^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu}_0) \right\}\right], \quad \text{.....(3.32)}$$

= kexp□□□□ □1 2 □□n □□□ k □2 1□ □k all □□ □□l l □2□ 2k□ □□ □ □□1 k ajs □□ □□j j □□□ □□s s □□□□□□……..(3.32) yi

□□ □0 □ □jxji □ □

□□ 2 i 1□ j□1 □□ 2□□□l□0 j 0s j 1 □□□□

where k □ k1□k2 □k3. Now let

2

v = □1 2 i□□n1□□□□ yi □□ □0 j□k□1□ jxji □□□□

= □1 2 □n yi2 □ □n2 □ □02 □1 2 □ □k n □2 2jxji □ 22 □□□□0i□□n1 yi □ j□

□k□1i□n1□ jx yji i □□□□ i□1 j□1i□1 □ □

k□1 k n

□ 2/2 □ □ □ □ □j s x x ji si ……………………………………………………….(3.33)

□ j□0 s j□ □1i□1

k 2 k□1 k

w = □ a jj □□ □□j j □ □ 2 □ □ a js □□ □□j j □□□ □□s s □ l□0 j□0s j□ □1

k

= □ ajj □□ □ □ □ □□2j 2 j j2j □□ 2k□□1 □k ajs □□ □ □□

□ □□ □ □□ □j s s j j s j s □

j□0 j□0s j□ □1

= □k a jj□ □2j 2 □k ajj □ □ □ j j □k ajj□ □2j 2k□□1 □k ajs □ □ □ j

s 2k□□1 □k ajs s □ □ j j□0 j□0 j□0 j□0s j□ □1 j□0s j□ □1

k□1 k k□1 k

□2 □ □ a js□ □ □j s 2 □ □ a js□ □j s

…………………………………………(3.34) j□0 s j□ □1 j□0s j□ □1

Thus, $v'w = \sum_{i=1}^{n} a_{00} - 2\sum_{j=1}^{k} a_{j0}\sum_{i=1}^{n} x_{ji} + \sum_{j=1}^{k}\sum_{l=1}^{k} a_{jl}\sum_{i=1}^{n} x_{ji}x_{li} + 2\sum_{j=1}^{k}\sum_{i=1}^{n} x_{ji}y_i - 2\sum_{l=1}^{k} a_{l0}\sum_{i=1}^{n} x_{li} + \sum_{i=1}^{n} y_i^2$

$$= \sum_{j=1}^{k}\sum_{l=1}^{k} a_{jl}\sum_{i=1}^{n} x_{ji}x_{li} - 2\sum_{j=1}^{k} a_{j0}\sum_{i=1}^{n} x_{ji} + 2\sum_{j=1}^{k}\sum_{i=1}^{n} x_{ji}y_i + \sum_{i=1}^{n} y_i^2 \quad r \quad \text{....(3.35)}$$

where r is a constant term, independent of β_j. It can be seen that $Q(\boldsymbol{\beta}) - v'w$ is a quadratic form of the matrix $\boldsymbol{\Sigma}_{\beta}^{-1}$ in $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots \beta_k)$. Therefore the posterior distribution of $\boldsymbol{\beta}$ is of the form

$$p(\boldsymbol{\beta}|y) = ke^{-\frac{1}{2}Q(\boldsymbol{\beta})} \quad . \qquad \text{.....(3.36)}$$

Hence Equation (3.33) follows the multivariate normal distribution with mean vector given by

$$\boldsymbol{\mu} = \boldsymbol{\Sigma}_{\beta}\boldsymbol{C}, \qquad \text{.....(3.37)}$$

where $\boldsymbol{\Sigma}_{\beta}$ is a $(k+1)\times(k+1)$ matrix with inverse $\boldsymbol{\Sigma}_{\beta}^{-1} = (m_{ij})$ whose elements are given as

$$m_{00} = \frac{n}{2} - a_{00},$$

$$m_{j0} = \frac{1}{2}\sum_{i=1}^{n} x_{ji} - a_{j0}, \quad j = 1,2,\ldots k$$

$$m_{0j} = \frac{1}{2}\sum_{i=1}^{n} x_{ji} - a_{0j}, \quad j = 1,2,\ldots k \qquad \text{.....(3.38)}$$

$$m_{jl} = \frac{1}{2}\sum_{i=1}^{n} x_{ij}x_{il} - a_{ij}, \quad i \neq j,$$

$$m_{ii} = \frac{1}{2}\sum_{i=1}^{n} x_{ij}^2 - a_{ii}, \quad j = 1,2,\ldots k$$

$\square_{i\square 1}$ $\square\square$

$\boldsymbol{C}$ is a column vector of order $(k + 1)$ with elements given as

$C0 \square \square 2\square\square\square\square \square \underline{1}2\ i\square\square n1\ yi \square\ j\square k\square 1a0\ j\square\ j\ \square\square\square\square \quad \square\square\square$

$\square$ ……(3.39)

$\square \quad n \quad k$

$Ci \square \square 2\square\square\square \square \underline{1}2\ i\square\square 1\ y\ xi\ li \square\ j\square\square 1aij\square\ j\ \square\square\square\square, \quad l \square 1,2,...,k\square\square\square$

Estimation of μ and Σ

To estimate the parameters of the prior distribution of the regression parameters, we used the jackknife sample as follows:

Let $\boldsymbol{\beta}\hat{}_l \square \square \square\square\hat{}_0 l, \hat{}_1 l, ..., \square\hat{}_{kl}\square; l \square 1, 2, 3, ...,$

n be the l^{th} jackknife estimate of the regression.

Then the estimate of the mean vector $\boldsymbol{\mu}$ of the random vector $\boldsymbol{\beta} \square \square\square_0, \square_1, ..., \quad \square_k\square$ is given as

$\boldsymbol{\mu}\hat{} \square \square\square\hat{}_0, \square\hat{}_1,...,\square\hat{}_k\square\square$, where

$$\square\square\hat{}_j \qquad \frac{1}{n}\square\square_{ji}, \quad j \square 0, 1, ..., .k \qquad \text{.....(3.40)}$$

$i\square 1$

and an estimate of the covariance matrix of $\boldsymbol{\beta}$ is given by

$n\square \quad \square\square \quad \square\square a\hat{}_{ij}\square$. ……..(3.41)

$$\boldsymbol{\Sigma}\hat{} \square \ n\ \frac{1}{\square 1}\square\ \boldsymbol{\beta}\hat{}_j \square\ \boldsymbol{\mu}\hat{}_j \quad \boldsymbol{\beta}\hat{}_i \square\ \boldsymbol{\mu}\hat{}_j \ \square\ \square$$

$j\square 1$

The estimate of the standard error of the i^{th} coefficient based on the Bayesian estimate is the square root of the i^{th} diagonal elements of $\boldsymbol{\Sigma}\hat{}_{\beta}$.

Loss Function

Let $\boldsymbol{t}$ denote an estimator of $\boldsymbol{\beta}$. The ***loss function***, $L(\boldsymbol{\beta}, \boldsymbol{t})$, is defined to be a real-valued function satisfying

(i) $L(\boldsymbol{\beta}, \boldsymbol{t}) \geq 0$ for all possible estimators $\boldsymbol{t}$ and for all $\boldsymbol{\beta}$ in the parameter space Ω.

(ii) $L(\boldsymbol{\beta}, \boldsymbol{t}) = 0$.

Some possible loss functions are (see Box and Tio (1973))

(i) $L_1(\boldsymbol{\beta}, \boldsymbol{t}) = (\boldsymbol{t} - \boldsymbol{\beta})^2$, (ii) $L_2(\boldsymbol{\beta}, \boldsymbol{t}) = |\boldsymbol{t} - \boldsymbol{\beta}|$, (iii) $L_3(\boldsymbol{\beta}, \boldsymbol{t}) = \begin{cases} 0, & \text{if } \boldsymbol{t} = \boldsymbol{\beta} \\ 1, & \text{if } \boldsymbol{t} \neq \boldsymbol{\beta} \end{cases}$.,

$L_1(\boldsymbol{\beta}, \boldsymbol{t})$ is called the ***quadratic loss function*** (or the ***square error loss function***). $L_2(\boldsymbol{\beta}, \boldsymbol{t})$ is called **"*the absolute value*" *loss function*** while $L_3(\boldsymbol{\beta}, \boldsymbol{t})$ is called the **"zero-one" *loss function*.**

Notice that both L_1 and L_2 increase as the error $(\boldsymbol{\beta} - \boldsymbol{t})$ increases in magnitude.

The loss function, $L(\boldsymbol{\beta}, \boldsymbol{t})$, is a random variable. The expected value of $L(\boldsymbol{\beta}, \boldsymbol{t})$ with respect to the joint distribution of $Y_1, Y_2, ..., Y_n$, is called the ***risk function*** of t, denoted by $R_{\boldsymbol{t}}(\boldsymbol{\beta})$.

Thus, if Y is continuous, then (see Lindley (1965)),

$$R_t(\boldsymbol{\beta}) = \int_{\Omega} L(\boldsymbol{\beta}, \boldsymbol{t}) f(\boldsymbol{y}; \boldsymbol{\beta}) dy, \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.42)$$

where Ω is the sample space. $R_t(\boldsymbol{\beta})$ is a function of $\boldsymbol{\beta}$ and represents the expected loss of using $\boldsymbol{t}$ as an estimator of $\boldsymbol{\beta}$. Two or more estimators could be compared by looking at their respective risk functions, preference being given to that estimator with the minimum risk function.

The ***Bayes risk*** of the estimator $\boldsymbol{t}$ is denoted by $r(\boldsymbol{t})$ and is given by (see Lindley (1965))

$$r(\boldsymbol{t}) = \int_{\Omega} R_{\boldsymbol{t}}(\boldsymbol{\beta})\, p(\boldsymbol{\beta})\, d\boldsymbol{\beta} \quad \ldots\ldots(3.43)$$

The ***Bayes estimator*** of a parameter, with respect to a given loss function and prior distribution, is defined to be the estimator with the smallest Bayes risk (see Lindley (1965)).

Returning to the form of the Bayes risk, $r(\boldsymbol{t})$, we see that (provided it is mathematically justifiable to reverse the order of integration – an assumption we shall always make unless otherwise stated),

$$\begin{aligned} r(\boldsymbol{t}) &= \int_{\Omega}\int_{\mathcal{Y}} L(\boldsymbol{\beta}, \boldsymbol{t})\, f(\boldsymbol{y}|\boldsymbol{\beta})\, d\boldsymbol{y}\, p(\boldsymbol{\beta})\, d\boldsymbol{\beta} \\ &= \int_{\Omega}\int_{\mathcal{Y}} L(\boldsymbol{\beta}, \boldsymbol{t})\, f(\boldsymbol{y}|\boldsymbol{\beta})\, p(\boldsymbol{\beta})\, d\boldsymbol{y}\, d\boldsymbol{\beta}. \\ &= \int_{\Omega}\int_{\mathcal{Y}} L(\boldsymbol{\beta}, \boldsymbol{t})\, p(\boldsymbol{\beta}|\boldsymbol{y})\, f(\boldsymbol{y})\, d\boldsymbol{y}\, d\boldsymbol{\beta} \\ &= \int_{\mathcal{Y}}\left\{\int_{\Omega} L(\boldsymbol{\beta}, \boldsymbol{t})\, p(\boldsymbol{\beta}|\boldsymbol{y})\, d\boldsymbol{\beta}\right\} f(\boldsymbol{y})\, d\boldsymbol{y}. \\ &= \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(3.44) \end{aligned}$$

Since the integration is non-negative, the double integration can be minimized if the expression within the braces is minimized for each $y_1, y_2, \ldots, y_n$. Thus, choosing $\boldsymbol{t}$ to minimize $r(\boldsymbol{t})$ is equivalent to choosing $\boldsymbol{t}$ to minimize

$$h(\boldsymbol{t}) = \int_{\Omega} L(\boldsymbol{\beta}, \boldsymbol{t})\, p(\boldsymbol{\beta}|\boldsymbol{y})\, d\boldsymbol{\beta}, \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(3.45)$$

the posterior risk. Thus, the Bayes estimator of $h(\boldsymbol{\beta})$ with respect to a loss function and a prior distribution, is that estimator which minimizes the posterior risk.

Theorem 3.5

The mean of the posterior distribution is the Bayes estimator of β with respect to the quadratic loss function.

Proof

The Bayes estimator of β with respect to the quadratic loss function, is the value of t which minimizes the posterior risk

$$h(t) = \int (t-\beta)^2 p(\beta|y)\, d\beta.$$

Differentiating $h(t)$ with respect to t and equating $h'(t)$ to zero, we obtain

$$t \int p(\beta|y)\, d\beta = \int \beta p(\beta|y)\, d\beta$$

i.e. $t = \int \beta p(\beta|y)\, d\beta$ or $t = \int \beta p(\beta|y)\, d\beta$.

Therefore the Bayes estimate of β with respect to the quadratic loss function, is the mean of the posterior distribution of β. (Notice again that $\int p(\beta|y)\, d\beta = 1$, since $p(\beta|y)$ is a probability density function).

Theorem 3.2

The median of the posterior distribution is the Bayes estimator of β with respect to the absolute value loss function.

Proof

We consider the case $\beta \in R$. We want to choose t to minimize the posterior expected loss

$$\int_{-\infty}^{t} p(\beta \mid y)\, d\beta = \int_{t}^{\infty} p(\beta \mid y)\, d\beta \Rightarrow \int_{-\infty}^{t} p(\beta \mid y)\, d\beta - \int_{t}^{\infty} p(\beta \mid y)\, d\beta = 0.$$

Differentiating with respect to t and equating to zero, we obtain

$$\int_{-\infty}^{t} p(\beta \mid x)\, d\beta = \int_{t}^{\infty} p(\beta \mid x)\, d\beta,$$

that is,

$$2\int_{-\infty}^{t} p(\beta \mid y)\, d\beta = \int_{-\infty}^{t} p(\beta \mid y)\, d\beta + \int_{t}^{\infty} p(\beta \mid y)\, d\beta$$

$$= \int_{-\infty}^{\infty} p(\beta \mid y)\, d\beta = 1.$$

Thus,

$$\int_{-\infty}^{t} p(\beta \mid y)\, d\beta = \frac{1}{2},$$

and so t is the median of the posterior distribution of β.

Theorem 3.3

The mode of the posterior distribution is the Bayes estimator of β with respect to the zero-one loss function.

3.3.3 Maximum a Posteriori Method

In Bayesian data analysis, one way to apply a model to data is to find the maximum a posteriori (MAP) parameter values. The goal here is to find the parameter estimates that maximize the posterior probability of the parameters given the data. In other words, we find the mode of the posterior distribution. This corresponds to (Steyvers (2011)):

$$\beta_{MAP} = \underset{\beta}{\operatorname{argmax}}\, p(\beta \mid y)$$

$$= \underset{\beta}{\operatorname{argmax}}\, f(y \mid \beta)\, p(\beta)$$

$$= \underset{\boldsymbol{\beta}}{\operatorname{argmax}} f(\boldsymbol{y}\mid\boldsymbol{\beta})\, p(\boldsymbol{\beta}) \quad \text{.....(3.46)}$$

The Bayesian approach that we are really interested in is posterior sampling. With the MAP approach, we get a single set of parameter values for a model. Therefore, we are characterizing the posterior distribution with the mode of this distribution.

In a more comprehensive Bayesian approach, the goal is to characterize the full posterior distribution and not to simply find the mode of this distribution. In some cases, we might be able to find an analytic expression for the posterior distribution. However, in many cases, we have to resort to sampling techniques, such as Markov chain Monte Carlo (MCMC), to get samples from the posterior distribution. These samples can be used to calculate a number of things, such as means, variances and other moments of the distribution. We can also check whether there are any correlations between parameters.

The MCMC can be used to draw samples from a distribution, you should realize that MCMC can be used to get samples from the posterior distribution (Steyvers (2011)). We will start by illustrating the simplest of all MCMC methods: the Metropolis sampler. This is a special case of the Metropolis-Hastings sampler discussed in (2). Suppose our goal is to sample from the target density $p(\boldsymbol{\beta})$, with $\boldsymbol{\beta}$. The Metropolis sampler creates a Markov chain that produces a sequence of values:

$$\boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)}, \ldots, \boldsymbol{\beta}^{(t)}, \ldots$$

where $\boldsymbol{\beta}^{(t)}$ represents the state of a Markov chain at iteration t. The samples from the chain, after burning, reflect samples from the target distribution $p(\boldsymbol{\beta})$.

In the Metropolis procedure, we initialize the first state, $\boldsymbol{\beta}^{(1)}$ to some initial value. We then use a proposal distribution $q(\boldsymbol{\beta}^{*} \mid \boldsymbol{\beta}^{(t-1)})$ to generate a candidate point $\boldsymbol{\beta}^{*}$ that is conditional on the previous state of the sampler. The next step is to either accept the proposal or reject it. The probability of accepting the proposal is:

$$\alpha = \min\left(1, \frac{p(\boldsymbol{\beta}^{*} \mid \mathbf{y})}{p(\boldsymbol{\beta}^{(t-1)} \mid \mathbf{y})}\right). \quad \text{.....(3.47)}$$

To make a decision on whether to actually accept or reject the proposal, we generate a uniform deviate u. If $u \leq \alpha$, we accept the proposal and the next state is set equal to the proposal: $\boldsymbol{\beta}^{(t)} = \boldsymbol{\beta}^{*}$. If $u > \alpha$, we reject the proposal, and the next state is set equal to the old state:

$\boldsymbol{\beta}^{(t)} = \boldsymbol{\beta}^{(t-1)}$. We continue generating new proposals conditional on the current state of the sampler, and either accept or reject the proposals. This procedure continues until the sampler reaches convergence. At this point, the samples $\boldsymbol{\beta}^{(t)}$ reflect samples from the target distribution $p(\boldsymbol{\beta} \mid \mathbf{y})$. Here is a summary of the steps of the Metropolis sampler (Steyvers (2011)):

1. Set $t = 1$
2. Generate an initial value for $\beta_j \sim U(\theta_{1j}, \theta_{2j})$, $j = 0, 1, ..., k$
3. Repeat

   $t = t + 1$

   Do a Metropolis Hastings step on $\beta_j, j = 0, 1, ..., k$:

   Generate a proposal $\beta_j^{*} \sim N(\beta_j, \sigma_j^2)$;

   Evaluate the acceptance probability $\alpha = \min\left(1, \frac{p(\boldsymbol{\beta}^{*} \mid \mathbf{y})}{p(\boldsymbol{\beta}^{(t-1)} \mid \mathbf{y})}\right)$;

   Generate a u from a Uniform(0, 1) distribution

   If $u \leq \alpha$, accept the proposal and set $\beta_j^{(t)} = \beta_j^{*}, j = 0, 1, ..., k$
4. Until $t = T$.

3.4 Multilevel Approach

3.4.1 Introduction

In this Section, the study seeks to develop a Multilevel Analysis approach to estimate the regional distribution of parameters based on the modified Smeed's model and use them to compare the risk of RTFs across geographical regions.

3.4.2 Multilevel Model Specification

Assuming the population (geographical zone) is stratified into J sub-populations with n_j observations in the j^{th} sub-population. Equation (3.24) therefore becomes $y_{ij} = \beta_{0j} + \beta_{1j} x_{1ij} + \beta_{2j} x_{2ij} + \cdots + \beta_{kj} x_{kij} + \varepsilon_{ij}$,

$$= \beta_{0j} + \sum_{l=1}^{k} \beta_{lj} x_{lij} + \varepsilon_{ij}, \quad i_j = 1, 2, \ldots, n_j, \quad j = 1, 2, \ldots, J \qquad \text{.....(3.48)}$$

Across all geographical regions, $\beta_j = (\beta_{0j}, \beta_{1j}, \ldots, \beta_{kj})$ are assumed to have multivariate normal distribution (Hox, 2010). Thus, each β_{lj} ($l = 0, 1, 2, \ldots, k$) can be modeled as

$$\beta_{0j} = \gamma_{00} + \gamma_{01} z_j + u_{0j} \qquad \text{.....(3.49)}$$

$$\beta_{lj} = \gamma_{l0} + \gamma_{l1} z_j + u_{lj}, \quad l = 1, \ldots, k \text{ and } j = 1, 2, \ldots, J. \qquad \text{.....(3.50)}$$

In equations (3.31) and (3.32) the regression coefficients γ_{00}, γ_{01}, γ_{l0} and γ_{l1} ($l = 1, \ldots, k$) are not assumed to vary across geographical regions. They are therefore referred to as fixed coefficients. Substituting equations (3.49) and (3.50) into equation (3.48) yields the single equation model:

$$y_{ij} = \gamma_{00} + \gamma_{01} z_j + u_{0j} + \sum_{l=1}^{k} \left(\gamma_{l0} + \gamma_{l1} z_j + u_{lj} \right) x_{lij} + \varepsilon_{ij},$$

$$= \beta_{00} + \beta_{01} z_j + \sum_{l=1}^{k} \beta_{l0} x_{lij} + \sum_{l=1}^{k} \beta_{l1} x_{lij} z_j + u_{0j} + \sum_{l=1}^{k} u_{lj} x_{lij} + \varepsilon_{ij}, \quad i_j = 1, 2, ..., 2, ..., n_J \quad .....(3.51)$$

$u_{lj} \sim N(0, \sigma_l)$, $l = 0,1,...,$ and k $\varepsilon_{ij} \sim N(0, \sigma^2)$. Y has the normal distribution with mean

$$\beta_{00} + \beta_{01} z_j + \sum_{l=1}^{k} \beta_{l0} + \beta_{l1} z_j x_{lij} \quad \text{……...……...…...(3.52)}$$

and variance

$$v = \beta_{0}^{k} + 2 \sum x_{lij} x_{rij} \sigma_{lr} + \sigma^2. \quad \text{……..…..(3.53)}$$

The parameters to be estimated are β_{l0}, β_{l1}, σ_l, $\sigma_{lr} (l \neq r)$ and σ^2, $l = 0, 1, \ldots, k$.

If β_0 differs significantly from 0, then the parameters of the modified Smeed's model can be used to compare the risk of RTFs across the J geographical regions.

Equating the partial derivatives of the likelihood function to zero, we obtain the maximum likelihood estimators of the parameters β_{l0}, β_{l1}, σ_l, $\sigma_{lr} (l \neq r)$ and σ^2 as $\hat{\beta}_{l0}$, $\hat{\beta}_{l1}$, $\hat{\sigma}_l$, $\hat{\sigma}_{lr} (l \neq r)$ and $\hat{\sigma}^2$ respectively.

The segment $\beta_{00} + \beta_{01} z_j + \sum_{l=1}^{k} \beta_{l0} x_{lij} + \sum_{l=1}^{k} \beta_{l1} x_{lij} z_j + u_{0j}$ in equation (3.48) contains the fixed coefficients. It is often called the fixed (or deterministic) part of the model. The segment $u_{0j} + \sum_{l=1}^{k} u_{lj} x_{lij} + \varepsilon_{ij}$ in equation (3.48) contains the random error terms, and it is often called

the random (or stochastic) part of the model. The term $\boldsymbol{x}_{ij}z_j$ is an interaction term that appears in the model as a consequence of modeling the parameter β_j of zonal-level variable $\boldsymbol{x}_{ij}$ with the regional-level variable z_j. Thus the moderator effect of z on the relationship between the dependent variable y and the predictor x, is expressed in the single equation version of the model as *cross-level interaction*. The random error term e_{2j} is connected to $\boldsymbol{X}_{ij}$. Since the explanatory variable $\boldsymbol{x}_{ij}$ and the error term u_j are multiplied, the resulting total error will be different for different values of $\boldsymbol{x}_{ij}$, a situation that in ordinary multiple regression analysis is called '*heteroscendasticity*'.

Multilevel model is needed for this kind of analysis because the pattern of occurrence of road traffic fatalities in the same region in each year is generally more similar than the observations from different regions, which violates the assumption of independence of all observations. This lack of independence can be expressed as a correlation coefficient: the intra-regional correlation. The multilevel regression model can be used to estimate the intraregional correlation. The model use for this purpose is a model that constrains no explanatory variables at all. This is called *intercept-only* model. If there are no explanatory variables at the lowest level and the highest level, Equations (3.48) and (3.49), respectively, reduces to:

$$y_{ij} = \beta_{0j} + r_{ij}, \quad \text{.....(3.54)}$$

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad \text{.....(3.55)}$$

Substituting (3.55) into (3.54), we obtain the single equation model

$$y_{ij} = \gamma_{00} + u_{0j} + r_{ij}, \quad i = 1, 2, ..., n_j; \; j = 1, 2, ..., J \quad \text{.....(3.56)}$$

The model of Equation (3.56) does not explain any variance, it only decomposes the variance into two independent components: σ^2, which is the variance of the lowest level errors r_{ij}, and τ_0, which is the variance of the highest level errors u_{0j}. Using this model we can estimate the intra-regional correlation ρ by the equation

$$\rho = \frac{\sigma_0^2}{\sigma_0^2 + \sigma^2}. \quad \text{.....(3.57)}$$

The intra-regional correlation ρ is a population estimate of the variance explained by the population structure. Equation (3.57) simply states that the intra-regional correlation is equal to the estimated proportion of regional level variance compared to the estimated total variance.

CHAPTER FOUR

PRELIMINARY INVESTIGATIONS USING DATA FROM GHANA

4.0 Introduction

In this chapter, some preliminary investigations on some characteristics of road traffic accidents are performed and particularly road traffic fatalities in Ghana which are of general interest and have a certain bearing on the main results of this study. There are four sections, the first is on the epidemiology of RTAs and focusses on the demographic aspects of fatalities, the second deals with the regional distribution of RTFs, the third deals with RTF characteristics of types of road users and the final section examined the effect of age on road traffic fatality index in Ghana.

4.1 Epidemiology of Road Traffic Accidents in Ghana

4.1.1 Introduction

The methods developed and adopted in the field of public health for the study and control of epidemic diseases provide a useful framework for the study and control of road traffic accidents. Accidents may be interpreted as resulting from the total forces involved in the competition between man and his environment (Gordon, 1949), and the epidemiology method thus offers a scientific approach to the prevention of road traffic accidents.

The first study of global patterns of death among people aged between 10 – 24 years of age has found that road traffic accidents, complications during pregnancy and child birth, suicide, violence, HIV/AIDS and tuberculosis (TB) are the major causes of mortality. Many causes of death of young people are preventable and treatable. The study, which was supported by the World Health Organization (WHO) and published in the Lancet Medical Journal (Lozano, et al. 2012), found

that 2.6 million young people are dying each year, with 97% of these deaths taking place in low- and middle-income countries.

In this section, morbidity and mortality data from road traffic accidents (RTAs) as known in Ghana and other epidemiological variables of RTAs are studied. Since the predominant factors affecting road traffic fatalities in Ghana are population size and the number of registered vehicles, which are subject to rapid changes, the degree and direction of change are likely to determine the magnitude of the effect of RTAs. Thus, the study in this section, is conducted with the objective of:

1. analysing the patterns of road traffic accidents, injuries and fatalities in Ghana;
2. determining the magnitude of RTAs in Ghana;
3. identifying some current and pertinent factors in the aetiology of RTAs in Ghana.

Based on the above, we make some suggestions and recommendations on how to prevent this serious public health problem.

In a similar study, Odero et al. (1997) reviewed the epidemiological studies of road traffic injury in developing countries and examined the evidence for association with alcohol. The study revealed that, about three-quarters of road traffic deaths in the world occur in developing countries and about 80% of the casualties are men. According to a similar research work conducted by Nilambar et al. (2004), in South India, there were 83% male and 17% female accident victims. Labourers were the highest (29.9%) in number among the victims. The highest number of accidents took place in the month of January (12.9%) and on Sundays (17.1%). The occupants of the various vehicles constituted the large (45%) group of the victims. Among the motorized vehicles, two wheeler drivers were more (31.1%) involved in accidents. Out of 254 drivers, 14.9% were found to have consumed alcohol. Being knocked down was the commonest mode of accidents.

The data used in this study were obtained from the following sources.

(a) The data on the number of road traffic fatalities were obtained from the National Road Safety Commission (NRSC) of Ghana.

(b) The Driver and Vehicle Licensing Authority (DVLA) of Ghana provided the data on the number of registered vehicles in Ghana.

(c) The estimated population figures were obtained from the Ghana Statistical Service 2010 Population and Housing Census data

4.1.2 Population and RTA Pattern in Ghana

Table A1, in the appendix, shows the magnitude of RTAs over a period of 19 years, (from 1991 to 2009) in Ghana. During the period, 27,819 died in 189,172 road traffic accidents. The average incidence of the morbidity and mortality patterns from RTAs during the period were 62.2 and 7.4 per 100 000 population, respectively. The morbidity pattern was similar throughout the period with a mean of 1.2 per accident.

Changes in the index of the Public Health Risk (PHR) of road traffic accidents however give cause for concern. Since 1997, there has generally been a gradual upward trend, as shown in Table A1. Although, the 9.2 fatalities/100 000 in 2009 population is relatively low by international standards, it still points to the fact that more and more people as a proportion of the population are being killed through road traffic accidents. It means that, the public health significance of road traffic accidents is growing, and that should serve as a trigger for early action to forestall a serious national health problem.

Between 1991 and 2009, mortality rate per 100 accidents increased from 11.0 to 18.2. This represents an increase of 65.5% during the period. The risk indicator, which measures the chance of one death in a RTA, has increased by more than 70% during the 19-year period. Improved trauma care interventions would help save some lives from RTAs. For the year 2009, for instance, one person was killed in every five road traffic accidents that occurred.

Although the number of accidents increased during the period 1991 to 2009, the number of fatal and injurious accidents per 100 road traffic accidents remained almost constant during the period, with an average of 14.4 and 61.5, respectively. Thus, about 14 of every 100 road traffic accidents during the period were fatal, whilst 62 out of every 100 RTAs resulted in an injury. These figures showed that RTAs still pose a major public health problem, threatening the quality of life in Ghana.

4.1.3 Distribution of Road Traffic Fatalities by Age Group and Gender

Table A1, in the appendix, shows that, unlike many fatal diseases, road traffic accidents kill people from all age groups. A cumulative total of 23 697 fatalities were recorded during the 19year period. The highest fatalities during the period, were in the 26 – 35 year old age group. The table also shows that the active age group, 16 – 45 years, were the most vulnerable in road traffic fatalities, representing more than 60% of the total fatalities in the 19-year period.

Table A2 gives the annual distribution of male/female ratio of road traffic fatalities. It can be seen that, during the 19-year period, road traffic accidents are responsible for a far higher rate of death among males, by an approximate ratio of 3:1. Similar proportions apply to all the years. In the 19-year period, 73.7% of the road traffic fatalities were males while 26.3% were females. Male predominance in road traffic fatalities in Ghana may be due to the fact that men spend substantially more time in moving vehicles than women. Men are also more likely to be employed as drivers and mechanics of cars and trucks, including drivers of long haul vehicles which may mean spending several days and nights in a vehicle. Males, therefore, have a higher exposure to the risk of road traffic injuries.

4.1.4 The Distribution of Months and Days During Which Persons were Killed or Injured in RTAs

Table A3, shows the monthly distribution of road traffic injuries and fatalities in Ghana, in 2010 and 2011. In 2011, the highest incidence of 260 road traffic fatalities was recorded in the month of November. This represents 11.8% of the road traffic fatalities that year. In 2010, the highest incidence of 11.9% was recorded in the month of October. In 2011, February and June have the lowest incidence of 6.5% and 6.7% of road traffic fatalities, respectively.

The trend where the Christmas season and activities preceding it were associated with many fatal RTAs, seemed to have marginally disappeared, since, in 2011, November happened to be the worst month, as shown in Table A3.

Table A4 shows the days occurrences of road traffic accidents, in 2010 and 2011. It can be seen that, between January 2010 and December 2011, there was significant variation in the number of road traffic fatalities and the number of persons injured per day. Saturday stood out as the "problem day", during which most road traffic fatalities occurred. This may be due to the fact that,

in Ghana, most funerals, all-night parties and other social activities are on Saturdays. Many people return from these activities intoxicated with alcohol. The role of alcohol intoxication in the causation of RTAs should therefore not be underestimated.

In the year 2011, the highest number of road traffic fatalities (398; 18.1%) occurred on Saturdays and in the year 2011, the lowest number of road traffic fatalities occurred on Wednesdays. Surprisingly, in the year 2011, Mondays (14.7%) and Thursdays (14.5%) recorded more fatalities than Fridays (14.2%) and Sundays (13.6%), which, according to NRSC of Ghana, are known to be associated with high fatalities. This will have to be studied for at least two more years before any conclusion can be drawn.

4.1.5 Road User Class Involved in Deaths and Injuries

Table A5 shows the various descriptions of road users at risk from January 1991 to December 2009, as far as the effects of road traffic accidents are concerned.

It can be seen that, during the 19-year period, pedestrians were more likely to be injured or killed in RTAs than other road users. This may be due to the fact that, in Ghana, separating cars and pedestrians on the road by providing pavements, is very often not done. Speed limits of 30 km/h in shared-space residential areas are commonly not implemented. Car and bus fronts, as generally designed, do not provide protection for pedestrians against injury at collision speeds of 30 km/h or greater. During the 19-year period, more than 40% of those who were killed through road traffic accidents were pedestrians, followed by bus passengers (20.7%), car occupants (11.8%) and Heavy Goods Vehicles (10.4%) in that order.

Buses, in particular, have high number of occupants and are therefore always likely to produce casualties (fatalities) far more than the number of registered buses when they get involved in accidents. The number of Heavy Goods Vehicles (HGVs) occupants killed in road traffic accidents, is unacceptable, considering the fact that they are not required to carry passengers.

In terms of strategy, isolating buses and HGVs for road safety interventions, would be consistent with the recommendation by the National Road Safety Commission, since most bus and HGV fatalities are recorded on the trunk roads. Ensuring the use of seat-belts in cars and buses will significantly save lives of some cars and bus occupants. Again, the inappropriate use of HGVs to ferry passengers should be stopped. Cutting down the overall pedestrian fatalities would require

active speed management on all categories of road users. A comprehensive traffic calming programme and speed controls may also be imperative. This may buttress the need to rationalize the National Highway System so as to bypass major settlements. This will be in keeping with the mobility functional requirements of the National Highways. Given the continuing high casualties among public transport buses and HGVs, it is rather urgent that, in addition to providing speed management measures on trunk roads passing through settlements, these categories of vehicles should be subjected to operational speed restrictions in the interest of the travelling public. These recommendations are in line with that of the National Road Safety Commission road safety report for 2011. The very direct link between speed at the time of collision and injury outcomes does not need to be over-emphasized.

Also of significance to note is the type of vehicles involved in fatal accidents. Cars constituted about 48% of vehicles involved in accidents (see Table A5). The involvement of buses, HGVs and pick-up utility vehicles also still trail car-involvement in that order. Of all the vehicle types, it is the HGVs and buses that are over-represented in their crash involvement relative to their proportion in the overall national vehicle mix. But even more worrying is that these classes of vehicle, accounted for higher proportions of involvement in fatal accidents.

The magnitude of road traffic accidents (RTAs) in Ghana over the past two decades is borne out by the fact that, averagely, about 72 persons out of every 100 000 population, suffered from grievous bodily injury and close to 8 persons of the same population died from RTAs. More than 60% of road traffic fatalities occur in children and young persons under 35 years of age. Many of these victims are likely to be pedestrians and young adults who were either drivers or passengers. About 75% of road traffic accident victims were males since more males than females own and drive vehicles in Ghana.

4.1.6 Conclusion

This study has shown that, during the period 1991 to 2009, males were more at risk than females in being injured in road traffic accidents. The preponderance of males may be attributed to their greater exposure to traffic and other associated factors. Mondal et al. (2011) and Odero et al. (1997) gave similar conclusions which are well documented. Male dominance in road traffic fatalities in

Ghana may be due to the fact that men spend substantially more time in moving vehicles than women.

The findings that the active age group, 16 – 45 years, was the most vulnerable in road traffic fatalities, representing more than 60% of the total fatalities in the 19-year period, is well documented in this paper. This has important economic significance as these are people in their most economically productive years.

This section has, also, given sufficient evidence of relatively high incidence of road traffic casualties on Saturdays. This may be due to the fact that, in Ghana, most funerals, all-night parties and other social activities are held on Saturdays. A good number of people most probably return from these activities intoxicated with alcohol.

Road traffic accidents in Ghana have not received the attention warranted, considering the magnitude of the problem. There is the need to view road traffic accidents as an issue that needs urgent attention aimed at reducing the health, social and economic impacts.

4.2 Comparative Analysis of Regional Distribution of the Risk of Road Traffic Fatalities in Ghana

4.2.1 Introduction

According to National Road Safety Commission (NRSC) of Ghana 2011 report, one key national Road Traffic Fatality index required for characterization and comparison of the extent and risk of traffic fatality among the ten regions of Ghana is fatalities per 100 accidents. Table A6, in the appendix, shows the distribution of the rate of road traffic fatalities per 100 accidents by region from 1991 – 2009. We wish to determine if there are significant differences between the mean regional fatalities per 100 accidents. Table 4.1 shows the mean road traffic fatalities per 100 accidents for each region from 1991 – 2009 as well as that over the national data.

Table 4.1: Regional mean fatalities per 100 accidents from 1991 to 2009

Regions	Greater Accra	Ashanti	Western	Eastern	Central	Volta	Northern	Upper East	Upper West	Brong-Ahafo	National
Total	107.8	338	320.8	378.4	414.8	447.7	777.5	518.6	538.5	543.5	274.4
Mean	5.67	17.79	16.88	19.92	21.83	23.56	40.92	27.29	28.34	28.61	14.44

4.2.2 Normality test

The one-way analysis of variance model assumes that the observations are normally and independently distributed with the same variance for each region or factor level (see Ofosu et al. (2014)). In practice, these assumptions will usually not hold exactly. We therefore check the normality assumption, using Shapiro-Wilk *W* test. The null hypothesis is

H_0: observations under each region are normally distributed against

the alternative hypothesis

H_1: observations under each region are not from a normally distributed population The value of the Shapiro-Wilk *W* test statistic for each of the ten regions is given in Table 4.2 below.

Table 4.2: Observed values of the *W* test statistic

Test Statistic	Greater Accra	Ashanti	Western	Eastern	Central	Volta	Northern	Upper East	Upper West	Brong-Ahafo
W_o	0.9708	0.9596	0.9622	0.9712	0.9408	0.9592	0.9466	0.8849	0.9323	0.9367

H_0 is rejected at the 5% level of significance if the computed value of *W* is less than 0.901, the tabulated 5% point. Since, for the Upper East region, the observed value of the test statistic W_o □ 0.8849 □ 0.901, H_0 is rejected at the 5% level and conclude that the sample is from a nonnormally distributed population. For each of the remaining 9 regions, we fail to reject H_0 and therefore conclude that there is not enough evidence of non-normality of these samples.

4.2.2 Test for homogeneity of variances

Levene's test (Levene 1960) is used to test if 10 samples have equal variances. We wish to test

H_0: □ □ □ □$_1^2$ $_2^2$ ...□ □$_{10}^2$ against

H_0: □ □□$_i^2$□$_j^2$ for at least one pair (i, j).

In Table A6, let y_{ij} represent the i^{th} observation taken under the j^{th} region and

1910 19

$y_{.j} = \sum y_{ij}$, $\bar{y}_j = y_j / 19$, ($j = 1, 2, ..., 10$), $y_{..} = \sum\sum y_{ij}$, $\bar{y}_{..} = y_{..}/190$. $i=1 j=1$

t = number of treatments = 10

n_i = number of observations from treatment (region) i

$N = n_1 + n_2 + ... + n_{10}$ = overall size of combined samples = 190,

$D_{ij} = |y_{ij} - \bar{y}_i|$ absolute deviation of observation j from treatment i mean

$\bar{D}_i$ = average of the n_i absolute deviations from treatment i

$\bar{D}$ = average of all N absolute deviations

The Levene's test statistic is given by

$$F_{Levene} = \frac{\sum_{i=1}^{10} n_i (\bar{D}_i - \bar{D})^2 / 9}{\sum_{i=1}^{10} \sum_{j} (D_{ij} - \bar{D}_i)^2 / 180} \quad \text{.....(4.1)}$$

When H_0 is true, F_{Levene} has the F-distribution with 9 and 180 degrees of freedom. H_0 is rejected at significance level 0.05 when the observed value of F_{Levene} is greater than $F_{0.05, 9,180} = 1.9322$. Since the observed F-ratio, 8.741, is greater than the critical F-value, 1.9322, the null hypothesis is rejected at the 0.05 level of significance and conclude that there are significant differences among the ten variances.

4.2.3 Kruskal-Wallis Test

Since there is a good reason to believe that the homogeneity of variance assumption has been violated, the Kruskal-Wallis test is used to determine if there are significant differences between the mean regional rates of fatalities per 100 accidents.

We wish to the test the hypotheses

H_0: There are no differences in the rate of fatalities per 100 accidents across regions, H_1: There are differences in the rate of fatalities per 100 accidents across regions.

Since the 10 samples are independently drawn from the populations of interest, the random variables under study are continuous, the measurement scale used is at least ordinal and each of the 10 sample sizes is at least 5, large approximation is used. Thus, the test statistic is

$$H = \frac{12}{N(N+1)} \sum_{j=1}^{10} \frac{R_j^2}{n_j} - 3(N+1), \quad \text{.....(4.2)}$$

where R_j = rank sum for sample j, where the rank of each measurement is computed according to its relative magnitude in the totality of the data for the 10 samples. In the presence of many ties, the test statistic H can be corrected using (4.2) and (4.3).

$$C = 1 - \frac{\sum_{i=1}^{r} (t_i^3 - t_i)}{N^3 - N} \quad \text{.....(4.3)}$$

where t_i is the number of ties of the i^{th} group ties.

$$H^* = \frac{H}{C}. \quad \text{.....(4.4)}$$

When H_0 is true, H^* has the chi-square distribution with 9 degrees of freedom. We reject H_0 at 5% level of significance if the computed value of H^* is greater than $\chi^2_{0.05,\,9} = 16.92$.

From the data, the computed value of H^* is 105.038. Since the observed value of H^* is greater than the critical value, 16.92, H_0 is rejected at the 0.05 level of significance and conclude that there is sufficient evidence that indicates that there are significance differences in the rate of fatalities per 100 accidents across regions.

4.2.4 Multiple comparison tests

Since the Kruskal-Wallis test indicates that the null hypothesis should be rejected, it implies that there are differences among the fatality rates for the 10 regions. But as to which of the regions are significantly different, the analysis does not specify. Obviously, in such a situation, a different method for comparing individual regional rates is warranted.

A popular nonparametric test to compare fatality rates between two regions is the Mann Whitney *U* test. Some investigators interpret this test as comparing the medians between the two populations. The hypotheses to be tested are

H_0: There are no differences in the rate of fatalities between regions *i* and *j*, against H_1: There are differences in the rate of fatalities between regions *i* and *j*.

The test statistic for the Mann Whitney *U* Test is the smaller of U_1 and U_2, defined below.

$$U^1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 \quad \Big\} \quad \text{.....(4.5)}$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$$

Thus, the test statistic is

$$U = \min(U_1, U_2). \quad \text{.....(4.6)}$$

The values of the observed Mann Whitney *U* test statistics of the 45 pairs of regions are given in Table 4.3. If the observed test statistic is less than the critical value (which in all cases is 113), we conclude that the fatality rates of the two regions are different at the 0.05 level of significance. Pairs of regions with rates of fatalities significantly different are highlighted in Table 4.3. For example, the observed test statistic between the rates of fatality of Greater Accra region and that of Western region is 3. Thus, we can conclude that there is significant difference between the fatality rate of Greater Accra and Western regions.

Table 4.3: Values of the Mann Whitney *U* test statistics

		1	2	3	4	5	6	7	8	9	10
		Greater Accra	**Ashanti**	**Western**	**Eastern**	**Central**	**Volta**	**Northern**	**Upper East**	**Upper West**	**Brong-Ahafo**
1	**Greater Accra**		**4.5**	**3**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
2	**Ashanti**			162	135	**100**	**86**	**3**	**105**	**106**	**37.5**
3	**Western**				**110**	**72**	**80**	**0**	**89**	**91.5**	**28.5**
4	**Eastern**					**110.5**	120	**0**	124	132	**34**
5	**Central**						153	**0**	162	156	251
6	**Volta**							**24**	169	152	**110**
7	**Northern**								**77**	**84.5**	**52**

8	**Upper East**									179	144.5
9	**Upper West**										154.5
10	**Brong-Ahafo**										

4.2.4 Conclusion

It is obvious the average road traffic fatality rates per 100 accidents in Greater Accra are significantly lower than that of the remaining 9 regions. This implies that the risk of dying as a result of a road traffic fatality in Greater Accra is relatively low, recording an average rate of 5.7 road traffic fatalities per 100 accidents (see Table 4.1). Thus, out of every 100 road traffic accidents in the Greater Accra, about 6 of the victims are likely to die. Also, the average rate of fatality at the Northern region is significantly higher than any other region of Ghana. This result points to the fact that more and more people as a proportion of the recorded number of accidents are being killed through road traffic accidents in the Northern region.

4.3 The Effect on Road Traffic Fatality Index of Road Users in Ghana

4.3.1 Introduction

In the previous section, it was concluded that there are significance differences in the rate of fatalities per 100 accidents among the 10 geographical regions of Ghana. Another Road Traffic Fatality index required for characterization and comparison of the extent and risk of traffic fatality between regions and also among various road users of Ghana is the number road traffic fatality per 100 casualties. Road users in Ghana can be classified under the following eight categories: pedestrians, car occupants, goods vehicle occupants, bus/mini-bus occupants, motorcyclists, pick-up occupants, cyclists and others.

In this section, a two-factor factorial design and analysis of variance is used to determine if there are significant differences in road traffic fatality index rates among road user classes and also that, if there are significant differences in fatality index rates among the 10 geographical regions of Ghana. The interaction between road user class and geographical region in Ghana shall also be considered. Analysis of variance is well covered in several books, including those of Cochran and Cox (1957), Cox (1958), Fisher (1966), Kempthorne (1952), Montgomery (2001) and Ofosu et al. (2014).

Table A7 shows the distribution of road user class by fatality, casualty and fatality indices (F. I.) from 2010 to 2013, where F. I. refers to the number of road traffic fatalities per 100 casualties.

4.3.2 Method

An experiment involving the following 2 factors is considered:

A: The effect of a road user class on F. I. in Ghana.

B: The effect of a geographical region on F. I. in Ghana.

It is assumed assume that there is an interaction effect between the two factors. This means that the effect of road user class depends on the level of geographical region and vice versa. Our main objective is to determine if there are significant differences among the various road users in Ghana and to investigate if there is significant interaction between the factors *A* and *B*. Road user class is investigated at 7 levels, while geographical region is investigated at 10 levels. The experiment is replicated 4 times (4 years).

Table 4.4, extracted from Table A7, gives the data arrangement for a two-factor factorial experiment, with observations in each cell being the F. I. of a road user class in a specified geographical region over the 4 year interval (2010 to 2013).

We wish to test the following hypothesis:

1. there are no differences in road traffic fatality index among the road user classes;
2. geographical region has no significant effect on fatality index;
3. the road user class and region do not interact.

using a 0.05 level of significance.

Table 4.4: Data arrangement for the two-factor factorial experiment

		Geographical region										
		Greater Accra 1	Ashanti 2	Northern 3	Brong Ahafo 4	Upper East 5	Upper West 6	Central 7	Eastern 8	Volta 9	Western 10	Total
Road User Class	Pedestrian 1	16.8 81.1 20.8 20.4 23.1	35.4 113.6 25.4 27.2 25.6	41.5 177.2 50 50 35.7	38.3 148.2 37.1 35.5 37.3	62.5 206.1 23 62.5 58.1	34.8 182.2 45.2 64.7 37.5	29.2 110.5 29.5 23.2 28.6	28 104.2 23 24.9 28.3	35.2 126.9 25.8 31 34.9	25.5 105.9 25.2 25.4 29.8	1355.9
	Car Occupant 2	3.2	6.5	3.4	13.1	7.4 29 7.5	0 25 20 0	6.2	9	7.2	13	322.7

	22.7 5.4 6.7 7.4	33.3 8 6.7 12.1	48.2 17.1 12.8 14.9	41.1 10.6 11.7 5.7	7.4 6.7	5	26.2 4 8.9 7.1	31.7 8.1 7.8 6.8	32.8 8.6 9.3 7.7	32.7 7.4 8.2 4.1	
Goods Veh. Ocpts 3	17.6 48.7 10.8 12.3 8	18.4 59.5 16.7 11.1 13.3	12.4 51.1 12.7 8.4 17.6	19.5 73 14.5 16.6 22.4	0 28.2 8.3 11.6 8.3	20.3 74.7 13.3 28.6 12.5	15.7 59 18.4 12 12.9	16.5 58.7 19.7 8.9 13.6	6.8 41.5 13.6 10.8 10.3	26.3 78.4 23.2 15.3 13.6	**572.8**
Bus/Mini-Bus 4	4.8 20.5 6 5.6 4.1	8.4 28.8 6.1 6.7 7.6	16 59 5.7 19.5 17.8	11.7 41.6 4.5 11.1 14.3	4.8 15.2 5.6 4.8 0	17.8 66.1 7.1 21.7 19.5	8.8 44.6 5 8 22.8	3.7 23.5 5.3 9.8 4.7	8 33.2 6.3 11.8 7.1	8.9 29.5 4.9 8.3 7.4	**362**
Motorcyclists 5	10.5 40.7 10.9 6.7 12.6	24.3 81.2 21.1 20.8 15	19.4 100 36.6 17.3 26.7	20.3 86.2 20.1 19.2 26.6	32.6 125.8 38.8 15.6 38.8	31.2 135.1 26.7 43.9 33.3	18.4 63.2 20.9 9.3 14.6	12.8 63.7 15.4 16 19.5	22.9 67.4 10.3 19.9 14.3	23.1 99.1 30 19.1 26.9	**862.4**
Pick-Up Ocpts 9	6.4 44.2 5.8 13.3 18.7	7.1 26.4 9.1 3.4 6.8	10.6 163.3 24.4 70 58.3	5.4 52.8 1.7 21.7 24	6.5 45.9 7.9 25 6.5	0 22.3 3.7 9.5 9.1	4.2 34.6 3.9 15 11.5	5.9 20.8 4.1 5.9 4.9	26.5 41.1 0 3.6 11	4 33.6 8.6 11 10	**485**
Cyclist 7	13.3 76.2 18.7 25.8 18.4	30.4 104.6 20.4 20.5 33.3	234.8 70 58.3 42.9 63.6	106.2 21.7 24 40.5 20	20.8 144.8 46.7 46.7 30.6	33.3 211.2 57.1 33.3 87.5	103.5 45 21.4 16 21.1	26.1 103 24.2 25.8 26.9	18.8 64.4 0 14.3 31.3	122.3 8 37.9 36.4 40	**1271**
Total	**334.1**	**447.4**	**833.6**	**549.1**	**595**	**716.6**	**441.6**	**405.6**	**407.3**	**501.5**	**5231.8**

The model for this experiment is $i = 1, 2, ..., 7$

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk} \quad j = 1, 2, ..., 10 \quad \text{.....(4.7)}$$

$$k = 1, 2, ..., 4$$

where y_{ijk} is a typical observation (road traffic fatality index), μ is the overall mean effect, α_i is the effect due to the i^{th} road user class, β_j is the effect due to the j^{th} geographical region, $(\alpha\beta)_{ij}$ is the joint effect of the road user class at i^{th} level and the effect of geographical region at j^{th} level, ε_{ijk} are independent $N(0, \sigma^2)$.

$$\sum_{j=1}^{7}\sum_{i=1}^{10}(\alpha\beta)_{ij} = 0. \quad \text{.....(4.8)}$$

Total uncorrected sum of squares = 156 267.56. The computation for the various sums of squares are then given by

$$SST = \sum_{i=1}^{7}\sum_{j=1}^{10}\sum_{k=1}^{4} y_{ijk}^2 - 280\bar{y}...^2 = 156\,267.56 - 5231.8280^2 = 58\,511.38,$$

$$SSA = \frac{1}{40}\sum_{i=1}^{7} y_{i..}^2 - 280\bar{y}...^2 = \frac{1}{40}(1355.9^2 + 322.7^2 + ... 1271^2) - 5231.8280^2 = 27\,147.41,$$

$$SSB = \frac{1}{28}\sum_{j=1}^{10} y_{.j.}^2 - 280\bar{y}...^2 = \frac{1}{28}(334.1^2 + 447.4^2 + ... + 501.5^2) - 5231.8280^2 = 27\,147.41,$$

$$SSE = \sum_{i=1}^{7}\sum_{j=1}^{10}\sum_{k=1}^{4} y_{ijk}^2 - \frac{1}{4}\sum_{ij} t_{ij}^2 = 156\,267.56 - \frac{1}{4}(81.1^2 + 113.6^2 + ... 122.3^2) = 13\,535.19,$$

$$SSAB = SST - SSA - SSB - SSE = 10\,133.26,$$

where *SST* is the corrected sum of squares, *SSA* is the sum of squares due to road user class, *SSB* is the sum of squares due to geographical region, *SSAB* is the sum of squares due to the interaction between road user class and geographical region, and *SSE* is the residual sum of squares. The computations are summarized in Table 4.5.

1. Let C_i denote the effect of the i^{th} level of road user class on road traffic fatality index. The hypothesis to be tested is $H_0 : C_1 = C_2 = ... = C_7 = 0$ against H_1: at least one $C_i \neq 0$. The

test statistic is $F = \frac{\text{road user class mean square}}{\text{residual mean square}}$. F has the F-distribution with 6 and 210 degrees

of freedom when H_0 is true. H_0 is rejected at significance level 0.05 when the computed value of F is greater than $F_{0.05,\ 6,\ 210} = 2.10$. Since $70.20 > 2.10$, H_0 is rejected at 0.05 level of significance and therefore conclude that different road user classes have different effects on the road traffic fatality index.

Table 4.5: ANOVA table for the effects of factors A and B on F. I.

Source of variation	Sum of squares	Degrees of freedom	Mean square	F
Road user class	27147.41	6	4524.57	70.20
Region	7695.51	9	855.06	13.27
Interaction	10133.26	54	187.65	2.91
Residual	13535.19	210	64.45	
Total	58511.38	279		

2. Let R_j denote the effect of the j^{th} level of geographical region on the road traffic fatality index $j = 1, 2, ..., 10$. The hypothesis to be tested is H_0: $R_1 = R_2 = ... R_{10} = 0$ against

H_1: at least one $R_j \neq 0$. The test statistic is $F = \frac{\text{geographical region mean square}}{\text{residual mean square}}$. F has the Fdistribution with 9 and 210 degrees of freedom. H_0 is rejected at significance level 0.05 when the computed value of F is greater than $F_{0.05,\ 9,\ 210} = 1.88$. Since 13.27, the computed value of F, is greater than 1.88, H_0 is rejected at the 5% level and therefore conclude that different regions have different effects on the road traffic fatality index.

3. Here, the null hypothesis is H_0: all $(CR)_{ij} = 0$ against H_1: at least one $(CR)_{ij} \neq 0$. The test

statistic is $F = \frac{\text{interaction mean square}}{\text{residual mean square}}$. F has the F-distribution with 54 and 218 degrees of freedom when H_0 is true. Reject H_0 at significance level 0.05 when the

computed value of F is greater than $F_{0.05,\,54,\,210} \approx 1$. Since 2.91, the computed value of F, is greater than 1,

H_0 is rejected at the 5% level and therefore conclude that there is interaction between road user class and the type of region.

4.3.3 Multiple comparisons

Since the analysis of variance indicates that road user class means differ significantly, it is of interest to make comparisons between the individual road user class means to discover the specific differences.

Over the years, several methods for making multiple comparisons have been suggested. Duncan (1951, 1952, 1955) has contributed a considerable amount of research to the subject of multiple comparisons. Other multiple comparison methods in use are those proposed by Tukey (1949, 1953), Newman (1939), Keuls (1952), and Scheffé (1953, 1959). The advantages and disadvantages of the various multiple comparison methods are discussed by Bancroft (1968), O"Neill and Wetherill (1971), Daniel and Coogler (1975), Winer (1971) and Ofosu et al. (2014). Daniel (1980) has prepared a bibliography on multiple comparison procedures.

Tukey"s (1953) test, which is usually referred to as the ***honestly significant difference*** (*HSD*) test, makes use of a single value against which all differences can be compared. This value, called the *HSD*, is given by

$$HSD = q_{\alpha,\,k,\,\nu}\sqrt{MSE/n}, \quad \text{.....(4.9)}$$

Suppose we let $\alpha = 0.05$. Entering Table A11, in the Appendix, with $\alpha = 0.05$, $k = 7$ and $\nu = 210$, we obtain $q_{0.05,\,7,\,210} \approx 4.17$. In Table 4.5, we have $n = 40$ and $MSE = 64.45$. Hence, from Equation (4.9),

$$HSD = 4.17\sqrt{64.45/40} = 5.293.$$

Table 4.6 shows the mean fatality index for each of the road user class in Ghana.

Table 4.6: Mean road traffic fatality index for road user classes in Ghana

Road User Class	Pedestrian	Car Occupants	Goods Veh. Occupants	Bus/MiniBus	Motorcyclist	Pick-Up Occupants	Cyclist

Mean fatality index	33.90	8.07	14.32	9.05	21.56	12.13	31.78

The observed difference between each pair of means is compared to the *HSD*. If the observed difference is greater than 5.293, then the road traffic fatality indices of the two road user classes are significantly different. The values of the observed differences between pair of means of the 7 road user classes are given in Table 4.7. Pairs of road user classes with rates of fatalities per hundred casualties not significantly different are highlighted in Table 4.7.

Table 4.7: Observed differences between pair of means of road user classes

		1	2	3	4	5	6	7
		Pedestrian	**Car Occupants**	**Goods Veh. Occupants**	**Bus/MiniBus**	**Motorcyclist**	**Pick-Up Occupants**	**Cyclist**
1	**Pedestrian**		25.83	19.58	24.85	12.34	21.77	**2.12**
2	**Car Occupants**			6.25	**0.98**	13.49	**4.06**	23.71
3	**Goods Veh. Occupants**				**5.27**	7.24	**2.19**	17.46
4	**Bus/MiniBus**					12.51	**3.08**	22.73
5	**Motorcyclist**						9.43	10.22
6	**Pick-Up Occupants**							19.65
7	**Cyclist**							

For example, from Table 4.7, it can be seen that, the observed difference between the mean fatality indices for pedestrian and motorcyclist is 12.34. Since 12.34 is greater than 5.293, it follows that there is a significant difference between the road traffic fatality indices for pedestrian and motorcyclists. It is obvious that the road traffic fatality index for pedestrians is significantly greater than that of other road users except for cyclists. This means that, the risk of dying in a road traffic accident among pedestrians and cyclists are both significantly higher than those of other road users, recording an average rate of 33.9 and 31.78 deaths per 100 casualties, respectively.

4.3.4 Conclusion

In the previous section, it has been found that, there are significant differences in road traffic fatality indices (fatality per 100 casualties) among various road users and also among the ten geographical regions of Ghana. The risk of dying in a road traffic accident among pedestrians and cyclists are both significantly higher than those of other road users. This points to the fact that more and more people as a proportion of the recorded number of casualties, are being killed through road traffic accidents among these two categories of road users.

The encroachment on pedestrian walkways and footbridges and some roadways has limited pedestrian space along the corridor and the linked roads and thereby increasing the risk of pedestrians being injured or killed in road traffic accidents. Many storeowners make use of the space in front of their stores, including the pedestrian walkways, to showcase their stock. The lack of safeguarded pedestrian space on sidewalks along major roads and the lack of safe zebra crossings have also aggravated this risk.

There is the need for pedestrian-friendly flyovers to aid in crossing major highways. Adanu (2004) asserts that, in order for Accra to develop a sustainable transport system, it must increase its use of public transit (metro buses), and Non-Motorized Transport (walkways for pedestrians and cycling ways for bicyclists).

Bicycling as a form of transport is environmentally friendly and relatively cheap compared with other forms of transport. It also promotes healthy exercise. Reports demonstrate that a sizeable portion of Accra‟s population utilize this form of transport. Ghana‟s *National Transport Policy*, in 2008, recognizes the need for a strong Non-Motorized Transport component to the country‟s overall transportation development, highlighting these reasons. We need to develop the appropriate infrastructure (such as bicycle paths, free and open sidewalk) and safety measures (including motorists‟ recognition and respect for pedestrians and bicyclists) and legal protections for non-motorized transport. As a country, there is the need to formulate policies that will

1. foster a safer regime for use of non-motorised transport,
2. create better conditions for pedestrians,
3. foster greater use of bicycles.

An extensive study on bicycle use among the urban poor in Nima and Jamestown of Accra (Turner et al, 1995) highlighted the general negative attitudes within certain communities toward cyclists.

Healthy transport, as described by Banister (2008), requires separating people and traffic, with separate routes and space for pedestrians and cyclists. Investment in separate, dedicated infrastructure for cyclists could reduce these negative attitudes and the risk environment for cyclists. As well, promoting bicycle use as a transport mode requires addressing the cultural and community perceptions of bicycling use in different ethnic communities (Turner et al, 1995).

4.4 The Effect of Age on Road Traffic Fatality Index in Ghana

4.4.1 Introduction

Casualties of road traffic accidents in Ghana by age groups, from 2009 – 2013, are given in Table 4.8. Unlike many fatal diseases, road traffic accidents kill people from all age groups, including young and middle-aged people in their active years. A cumulative total of 10 555 fatalities is recorded over the 5-year period. The highest fatalities during the period were in the 26 – 35 year old. Table 4.8 also shows that the active age group, 16 – 45 years, was the most vulnerable in road traffic fatalities, representing 63.2% of the total fatalities in the 5-year period. According to the National Road Safety Commission (NRSC) of Ghana 2013 annual report, one key national Road Traffic Fatality index (F. I.) required for characterization and comparison of the extent and risk of road traffic fatality is fatalities per 100 casualties (see Hesse and Ofosu, 2015). In Table 4.8, the distribution of the rate of road traffic fatalities per 100 accidents by age groups from 2009 – 2013 are also computed.

Table 4.8: Age distributions of fatalities and injuries from road traffic accidents from 2010 to 2013

		Casualties											
		Persons Killed						Persons Injured					
		2013	2012	2011	2010	2009	Total	2013	2012	2011	2010	2009	Total
Age groups (Years)	0 – 5	97	113	126	136	130	602	214	241	276	389	401	1521
	6 – 15	148	170	212	217	250	997	529	789	846	962	1112	4238
	16 – 25	315	335	365	269	388	1672	2172	2509	2723	3110	3245	13759
	26 – 35	531	661	658	577	609	3036	3871	4458	5070	5297	5861	24557
	36 – 45	359	441	400	379	383	1962	2162	2753	3009	2932	3138	13994
	46 – 55	188	236	209	184	222	1039	1001	1334	1374	1399	1512	6620
	56 – 65	149	159	126	129	141	704	472	621	493	563	618	2767
	Over 65	111	125	103	95	109	543	190	296	229	266	246	1227

Total	**1898**	**2240**	**2199**	**1986**	**2232**	**10555**	**10611**	**13001**	**14020**	**14918**	**16133**	**3994**

It can be seen, from Table 4.9, that the F. I. increased from 24.5 to 31.2 among children under 6 years from year 2009 to 2013, whilst that of the „over 65" age groups increased marginally from 30.7 to 36.9 over the same period. In very simple terms, these changes imply that the chance of at least one casualty dying as a result of road traffic accident has increased over the period. It can be observed that, over the 5 year period, the „over 65" continues to be the age group with the highest national fatality rate. For instance, in 2013, about 37% of all road traffic casualties who were over 65 years lost their lives while 31% of casualties who were 5 years old or less died as a result of road traffic accidents.

Table 4.9: Rate of fatalities per 100 casualties (fatality indices)

		0 – 5	**6 – 15**	**16 – 25**	**26 – 35**	**36 – 45**	**46 – 55**	**56 – 65**	**Over 65**
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
2013	**1**	31.2	21.9	12.7	12.1	14.2	15.8	24.0	36.9
2012	**2**	31.9	17.7	11.8	12.9	13.8	15.0	20.4	29.7
2011	**3**	31.3	20.0	11.8	11.5	11.7	13.2	20.4	31.0
2010	**4**	25.9	18.4	8.0	9.8	11.4	11.6	18.6	26.3
2009	**5**	24.5	18.4	10.7	9.4	10.9	12.8	18.6	30.7
mean		**29.0**	**19.3**	**11.0**	**11.1**	**12.4**	**13.7**	**20.4**	**30.9**

The number of road traffic fatality victims in Ghana can be classified according to two criteria, of a set of entities, namely casualty and age group. Casualty has 2 levels (i.e. fatalities and injured) while age group has 8 levels. These form a 2 □ 8 contingency table as shown in Table 4.10.

Table 4.10: Road traffic accidents victims from 2010 to 2013

		Age Group								
		0 – 5	**6 – 15**	**16 – 25**	**26 – 35**	**36 – 45**	**46 – 55**	**56 – 65**	**Over 65**	**Total**
casualty	**Fatalities**	602	997	1672	3036	1962	1039	704	543	10555
	Injured	1521	4238	5759	24557	13994	38551	2767	1227	92614
	Total	2123	5235	7431	27593	15956	39590	3471	1770	103169

In this study, we wish to know whether road traffic casualty and age group are independent. If they are independent, then we would expect to find the same proportion of fatalities across various age groups. We also propose the use of the completely randomized single factor experiment to determine if there are significant differences in road traffic fatality index rates among the various age groups.

4.4.2 Method

Table 4 shows an $r \times c$ contingency table where O_{ij} is the observed frequency for level i of the first method of classification and level j of the second method of classification, where

$R_i = \sum_{j=1}^{c} O_{ij}$ is the *marginal total* for row i and $C_j = \sum_{i=1}^{r} O_{ij}$ is the *marginal total* for column j.

Note that $\sum_{i=1}^{r} R_i = \sum_{j=1}^{c} C_j = n$, where n is the total sample size.

Table 4.11: An $r \times c$ contingency table

	Columns				
	1	2	…	c	Total
1	O_{11}	O_{12}	…	O_{1c}	R_1
2	O_{21}	O_{22}	…	O_{2c}	R_2
Rows.	.	.	.	.	.
	.	.	.	.	.
	.	.	.	.	.
r	O_{r1}	O_{r2}	…	O_{rc}	R_r
Total	C_1	C_2	…	C_c	n

We are interested in testing the null hypothesis

H_0: the row-and-column methods of classification are independent

against the alternative hypothesis

H_1: the row-and-column methods of classification are not independent.

The test statistic is given by (see Cramér (1946) and Birch (1964)).

$$H = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij} - E_{ij})^2}{E_{ij}}, \quad \text{.....(4.10)}$$

were E_{ij} is the expected cell frequency for the $(ij)^{th}$ cell. It can be shown that, if H_0 is true, then:

$$E_{ij} = \frac{(\text{row total}) \times (\text{column total})}{\text{grand total}}. \quad \text{.....(4.11)}$$

It can also be shown that, for large n, the statistic H has an approximate chi-square distribution with $(r - 1)(c - 1)$ degrees of freedom if H_0 is true (see Ofosu and Hesse (2011)). Therefore, we would reject the hypothesis of independence if the observed value of the test statistic H is greater than the critical value $\chi^2_{(r-1)(c-1)}$, where α is the size of the test. An extensive treatment of the chi-square distribution can be found in the book by Lancaster (1969).

If we reject the null hypothesis, we conclude that there is some interaction between the two criteria of classification.

4.4.3 Results

1. Test of independence

The null and the alternative hypotheses are:

H_0: Casualty is independent of age group.

H_1: Casualty is not independent of age group.

We first find the expected cell frequencies. These are calculated by using Equation (2). Table 4.12 shows the expected cell frequencies of Table 4.10 using Equation (2). For example, $E_{11} = \frac{10555 \times 103169}{2123} = 217.200$.

Table 4.12: Expected cell frequencies of Table 4.10

		Age Group								
		0 – 5	**6 – 15**	**16 – 25**	**26 – 35**	**36 – 45**	**46 – 55**	**56 – 65**	**Over 65**	**Total**
casualty	**Fatalities**	217.200	535.582	760.250	2822.981	1632.424	4050.368	355.111	181.085	10555
	Injured	1905.800	4699.418	6670.750	24770.019	14323.576	35539.632	3115.889	1588.915	92614
	Total	2123	5235	7431	27593	15956	39590	3471	1770	103169

Note that the expected frequencies in any row or column add up to the appropriate marginal total. The test statistic is

$$H = \sum_{i=1}^{2} \sum_{j=1}^{8} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}.$$

When H_0 is true, H has the chi-square distribution with 7 [i.e. (2 – 1)(8 – 1)] degrees of freedom. We reject H_0 at 0.05 level of significance when the computed value of the test statistic is greater than $\chi^2_{0.05,7} = 14.07$. Substituting both the observed values in Table 3 and their corresponding expected values in Table 4.12 into $\chi_{ij} = \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$, we obtain the cells in Table 4.13.

Table 4.13: Calculations of the observed test statistic

	1	2	3	4	5	6	7	8	Total
χ_{1j}	245.97	213.55	497.18	14.95	55.36	8727.95	172.90	241.22	10169.08
χ_{2j}	97.35	50.24	144.35	1.85	7.76	235.23	43.99	106.75	687.52
Total	343.32	263.79	641.53	16.79	63.12	8963.18	216.89	347.97	10856.59

Thus, the observed value of the test statistic is

$$\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{8}\frac{(O_{ij} - E_{ij})^2}{E_{ij}} = 10856.59.$$

Since $10856.59 > 14.07$, we reject the hypothesis of independence and conclude that casualty is not independent of age group.

2. Completely randomized single factor experiment

Table 2 is a typical data of a single-factor experiment with 8 levels (age groups) of the factor, where the factor is the effect of age on F. I. We wish to determine if there are significant differences between the average F. I. across the 8 age groups. In Table 4.9, let y_{ij} represent the i^{th} observation taken under the j^{th} age group and

$$y_{.j} = \sum_{i=1}^{5} y_{ij}, \quad \bar{y}_{.j} = y_{.j}/5, \quad (j = 1, 2, ..., 8), \quad y_{..} = \sum_{i=1}^{5}\sum_{j=1}^{8} y_{ij}, \quad \bar{y}_{..} = y_{..}/40.$$

Let μ_j represent the true mean of the j^{th} age group and ε_{ij} the experimental error. The model for the completely randomized single factor experiment is $y_{ij} = \mu + \mu_i + \varepsilon_{ij}$, $(j = 1, 2, ..., 8, i = 1, 2, ..., 5)$. …..…….....(4.12)

The one-way analysis of variance model assumes that the observations are normally and independently distributed with the same variance for each region or factor level (see Ofosu et al. (2014)).

Validation of normality and homogeneity of variances assumptions

We check the normality assumption, using the Shapiro-Wilk W test. The null hypothesis is

H_0: observations under each region are normally distributed

against the alternative hypothesis

H_1: observations under each region are not from a normally distributed population

The value of the Shapiro-Wilk W test statistic for each of the eight age groups is given in Table 4.14 below.

Table 4.14: Observed values of the *W* test statistic

Test Statistic	0 – 5	6 – 15	16 – 25	26 – 35	36 – 45	46 – 55	56 – 65	Over 65
W_o	0.802	0.883	0.864	0.930	0.871	0.951	0.836	0.925

H_0 is rejected at the 5% level of significance if the computed value of W is less than 0.762, the tabulated 5% point of the distribution of the Shapiro-Wilk test statistic. For each of the 8 age groups, we fail to reject H_0 and therefore conclude that there is not enough evidence of nonnormality of these samples.

Levene's test (Levene 1960) is used to test if 8 samples have equal variances. We wish to test

$$H_0: \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_8^2 \quad \text{against}$$

$$H_0: \sigma_i^2 \neq \sigma_j^2 \quad \text{for at least one pair } (i, j).$$

In Table 4.10, let y_{ij} represent the i^{th} observation taken under the j^{th} age group and

$$y_{.j} = \sum_{i=1}^{5} y_{ij}, \quad \bar{y}_{.j} = y_{.j}/5, \quad (j = 1, 2, \ldots, 8), \quad y_{..} = \sum_{j=1}^{8}\sum_{i=1}^{5} y_{ij}, \quad \bar{y}_{..} = y_{..}/40.$$

t = number of treatments = 8

n_i = number of observations from treatment (region) i

$N = n_1 + n_2 + \ldots + n_9$ = overall size of combined samples = 40,

$D_{ij} = \left| y_{ij} - \bar{y}_i \right|$ = absolute deviation of observation j from treatment i mean

$\bar{D}_i$ = average of the n_i absolute deviations from treatment i

$\bar{D}$ = average of all N absolute deviations

The Levene‟s test statistic is given by

$$F_{Levene} = \frac{\sum_{i=1}^{8} n_i \left(\bar{D}_i - \bar{D}\right)^2 / 7}{\sum_{i=1}^{8} n_i \left(D_{ij} - \bar{D}_i\right)^2 / 32} \quad \ldots\ldots(4.13)$$

When H_0 is true, F_{Levene} has the F-distribution with 4 and 40 degrees of freedom. H_0 is rejected at significance level 0.05 when the observed value of F_{Levene} is greater than $F_{0.05, 7, 32} = 2.33$. Since the observed F-ratio, 1.332, is less than the critical F-value, 2.33, we fail to reject the null hypothesis at the 0.05 level of significance and conclude that there are no significant differences among the ten variances.

One-way analysis of variance

Since the normality and homogeneity of variances assumptions are validated, we can use the one-way analysis of variance to determine if the fatality indices across age groups vary significantly. We wish to test the hypothesis

H_0: The mean fatality indices are the same across the 8 categories of age groups, against the alternative hypothesis

H_1: The mean fatality indices are not the same for at least 2 of age groups.

The total corrected sum of squares is given by

$$SST = \sum_{j=1}^{8}\sum_{i=1}^{5} y_{ij}^2 - 40\bar{y}_{..}^2 = 2374.360. \quad \text{.....(4.14)}$$

The sum of squares among treatments is

$$SSA = \sum_{j=1}^{8} 5\bar{y}_{.j}^2 - 40\bar{y}_{..}^2 = 2193.712. \quad \text{.....(4.15)}$$

The within treatment sum of squares, SSW, can be obtained from the equation

$$SSW = SST - SSA = 180.648. \quad \text{.....(4.16)}$$

The analysis of variance results, based on the data in Table 2, are summarized in Table 8 below.

Table 4.15: Analysis of variance table

Source of variation	Sum of squares	Degrees of freedom	Mean square	F-ratio
Among treatments	2193.712	7	313.387	55.513

Within treatments	180.648	32	5.645	
Total	2374.360	39		

The test statistic is

$$F = \frac{\text{among treatments mean square}}{\text{within treatments mean square}}.$$

When H_0 is true, F has the F-distribution with 7 and 32 degrees of freedom. We reject H_0 at significance level 0.05 when the observed value of F is greater than $F_{0.05,\,7,32} = 2.33$. From Table 8, the computed value of F is 55.513. Since the observed F-ratio, 55.513, is greater than the critical F-value, 2.33, we reject the null hypothesis at the 0.05 level of significance and conclude that there are significant differences among the fatality indices across the 8 age groups.

4.4.4 Discussion

1. Multiple comparison method

Since the analysis of variance indicates that the null hypothesis should be rejected, it means that there are differences among the 8 treatment means. But as to which of the means are significantly different, the analysis does not specify. Obviously, in such a situation, we need a different method for comparing individual treatment means. One such method is the multiple comparison test.

Over the years, several methods for making multiple comparison tests have been suggested. Duncan (1951, 1952, 1955) has contributed a considerable amount of research to the subject of multiple comparisons. Other multiple comparison methods in use are those proposed by Tukey (1949, 1953), Newman (1939), Keuls (1952), and Scheffé (1953, 1959). The advantages and disadvantages of the various multiple comparison methods are discussed by Bancroft (1968), O"Neill and Wetherill (1971), Daniel and Coogler (1975), Winer (1971) and Ofosu et al. (2014). Daniel (1980) has prepared a bibliography on multiple comparison procedures.

The oldest multiple comparison method, and perhaps the most widely used, is the least significant difference method of Fisher, who first discussed it in the 1935 edition of his book

"The design of experiments" (see Ofosu et al. (2014)). To use this method, we first calculate the least significant difference, (LSD), for the given data. This is given by

$$LSD = t_{\frac{1}{2},\,N-k}\sqrt{\frac{2MSW}{n}}, \quad \text{.....(4.17)}$$

where the level of significance □ □ 0.05, N □ 40, n □5, k □8 and MSW□5.645. This gives LSD = 3.068.

The observed difference between each pair of means is compared to the LSD. If the observed numerical difference is greater than 3.068, then the road traffic fatality indices of the two age groups are significantly different. The values of the observed numerical differences between pairs of means of the 8 age groups are given in Table 9. Pairs of age groups with fatality indices not significantly different are highlighted in Table 9.

Table 4.16: Observed numerical differences between pair of means of road user classes

		0 – 5	6 – 15	16 – 25	26 – 35	36 – 45	46 – 55	56 – 65	Over 65
		29.0	19.3	11.0	11.1	12.4	13.7	20.4	30.9
0 – 5	29.0		9.7	18.0	17.9	16.6	15.3	8.6	**1.9**
6 – 15	19.3			8.3	8.2	6.9	5.6	**1.1**	11.6
16 – 25	11.0				**0.1**	**1.4**	**2.7**	9.4	19.9
26 – 35	11.1					**1.3**	**2.6**	9.3	19.8
36 – 45	12.4						**1.3**	8.0	18.5
46 – 55	13.7							6.7	17.2
56 – 65	20.4								10.5
Over 65	30.9								

For example, from Table 9, it can be seen that, the observed numerical difference between the mean fatality indices for the age groups „0 – 5“ and „26 – 35“ is 17.9. Since 17.9 is greater than 3.068, it follows that there is a significant difference between the two age groups with respect to F. I. It is obvious that the road traffic fatality index for „0 – 5“ age group is significantly higher than that of other age groups except for „Over 65“. This means that, the risk of dying in a road traffic accident among „0 – 5“ and „Over 65“ are both significantly higher than those of other age groups, recording an average rate of 29.0 and 30.9 deaths per 100 casualties, respectively.

4.4.5 Conclusion

We‟ve shown that road traffic casualty level depends on age group of victims involved using a 2 □ 8 contingency analysis.

The analysis of variance revealed that there are significant differences in road traffic fatality indices (fatality per 100 casualties) among various age groups in Ghana. The risks of dying in a road traffic accident among children under 6 years and older population who are over 65 years are both significantly higher than those of other age groups. This points to the fact that, although smaller number of children under 6 years and older population who are over 65 years die in road traffic accidents each year, more and more people as a proportion of the recorded number of casualties, are being killed through road traffic accidents among these two categories of age groups. Thus, the probability of being killed in a fatal road traffic accident is significantly high in each of these two age groups. This may be due to higher fragility of children and older population of road users.

These findings are consistent with a related study by Loughran et al. (2007), in which they reported that older drivers are more than twice as likely as middle-aged drivers to cause an accident. The research revealed that drivers and passengers riding in cars driven by older drivers are nearly seven times likelier to die in an auto accident than are passengers and drivers riding in cars driven by middle-aged drivers. This statistic suggests that older individuals are much likelier than middle-aged individuals to die in a car accident. Given these trends, the research suggests that public policy should focus more on improving the safety of automobile travel for older drivers and less on screening out older drivers whose driving abilities have deteriorated unacceptably.

4.5. Logistic Regression Approach to Modelling Road Traffic Casualties in Ghana

4.5.1 Introduction

Table 4.17, adapted from the National Road Safety Commission (NRSC) of Ghana, shows the annual distribution of road traffic injuries and fatalities in Ghana, from 1991 and 2013. The road traffic accident statistics in 2013 represent a reduction of 15.3% in fatalities over the 2012 figure. The fatality figure of 1 898 in 2013 is the lowest since year 2007. Relative to the year 2001, the

2013 figure for fatalities (1 898) recorded an increase of 14.3%, indicating an upward trend. A cumulative total of 316 669 casualties were recorded over the 23-years period, where fatalities formed 11.4% of this figure.

Table 4.17: Annual distribution of road traffic fatalities and injuries in Ghana from 1991 to 2013

		Casualty					Casualty		
i	Year	Fatality	Injury	Total	*i*	Year	Fatality	Injury	Total
1	1991	920	8773	9693	13	2003	1716	14469	16185
2	1992	914	9116	10030	14	2004	2186	16259	18445
3	1993	901	7677	8578	15	2005	1776	14034	15810
4	1994	824	7664	8488	16	2006	1856	14492	16348
5	1995	1026	9106	10132	17	2007	2043	14373	16416
6	1996	1049	9903	10952	18	2008	1938	14531	16469
7	1997	1015	10433	11448	19	2009	2237	16259	18496
8	1998	1419	11786	13205	20	2010	1986	14918	16904
9	1999	1237	10202	11439	21	2011	2199	14020	16219
10	2000	1437	12310	13747	22	2012	2240	13001	15241
11	2001	1660	13178	14838	23	2013	1898	10611	12509
12	2002	1665	13412	15077		**Total**	**36142**	**280527**	**316669**
						Percentage	**11.41**	**88.59**	**100.00**

According to NRSC of Ghana report, the number of road traffic crashes in 2013 (i.e. 9 200) represents a decrease of 23.9% and 18% over the 2012 and 2001 figures, respectively. The number of fatal crashes and their resulting fatalities in the previous year also saw a decrease. Compared to the 2012 figures, fatal crashes decreased in 2013 by 17% and fatalities by 15.3%. There was also a decrease of 17.9% in the overall number of casualties in 2013 compared with 2012. Relative to the year 2001, the 2013 figures for fatal accidents and fatalities recorded corresponding increases of 24.7% and 44.5% respectively, whilst overall casualties recorded a decrease of 15.6%.

In the logistic regression analysis of this data, road traffic casualty is considered as the response or dependent variable of interest and year as predictors. The response has two categories: fatality and injury. The general objective of this analysis is to describe the way in which casualty distribution of road traffic fatalities varies by year and use this variation to predict future distribution. Logistic regression was proposed, as an alternative to ordinary least squares, in the

late 1960s and early 1970s (Cabrera, 1994), and it became routinely available in statistical packages in the early 1980s. Since that time, the use of logistic regression has increased in the social sciences (e.g., Chuang, 1997; Janik & Kravitz, 1994; Tolman & Weisz, 1995) and in educational research, especially in higher education (Austin et al., 1992).

Other studies have been conducted in the area of road traffic casualties in Ghana. Hesse et al. (2014c) derived a Bayesian model for predicting the annual regional distribution of the number of road traffic fatalities in Ghana. The study showed that population and number of registered vehicles are predominant factors affecting road traffic fatalities in Ghana. Similar conclusions were arrived at when a least square regression method (see Hesse et al. (2014d)) and multilevel random coefficient method (see Hesse et al. (2014e)) were used to derive models for predicting road traffic fatalities in Ghana.

4.5.2 Methods

Let n_i denote the number of road traffic casualties in the ith year in Ghana and let y_i denote the number of road traffic fatalities (RTFs) in the ith year in Ghana. We view y_i as a value of a random variable Y_i that takes the values 0, 1, …, n_i. If we assume the n_i observations for each year are independent, and they all have the same probability p_i of dying as a result of RTAs, then Y_i has the binomial distribution with parameters p_i and n_i [i.e. $Y_i \; B \; n(\; i, p_i)$]. The probability mass function of Y_i is given by:

$$f_{y(\;i)} \; \square \; \square \; n_{yii} \; \square \; p_{i}^{y_i}(1 \square \; p_{i})^{n_i \square y_i}, \; y_{i} \; \square \; 0, 1, ..., n_{i} \quad \text{.....(4.18)}$$

It can be shown that the expected value and variance of Y_i are (Ofosu and Hesse, 2010):

$$E\; Y_{(\;)i} \; \square \; n\; p_{i\; i} \quad \text{and} \quad \text{var}(\;)_{i} \; \square \; n\; p_{i\; i}(1 \square \; p_{i}). \quad \text{.....(4.19)}$$

The odds$_i$ is the ratio of the probability to its complement, or the ratio of favourable to unfavourable cases. Thus,

$$\text{odds}_i = \frac{p_i}{1 - p_i} \quad \text{.....(4.20)}$$

We take logarithms, calculating the logit or log-odds

$$\beta_i = \text{logit}(p_i) = \ln\left(\frac{p_i}{1 - p_i}\right), \quad \text{.....(4.21)}$$

If the logit of the underlying probability p_i is a linear function of the predictors, then we can write

$$\text{logit}(p_i) = \ln\left(\frac{p_i}{1 - p_i}\right) = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik}$$

$$= \boldsymbol{x}_i \boldsymbol{\beta}, \quad i = 0, 1, \ldots, k \quad \text{.....(4.22)}$$ where $\boldsymbol{x}_i = (1, x_{i1}, x_{i2}, \ldots, x_{ik})$ and $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \ldots, \beta_k)$. Exponentiating Equation (4.22) we find that the odds for the ith unit are given by:

$$\frac{p_i}{1 - p_i} = \exp(\boldsymbol{x}_i \boldsymbol{\beta}). \quad \text{.....(4.23)}$$

Solving for the probability p_i in the logit model gives

$$p_i = \frac{1}{1 + \exp(-\boldsymbol{x}_i \boldsymbol{\beta})} = \frac{\exp(\boldsymbol{x}_i \boldsymbol{\beta})}{1 + \exp(\boldsymbol{x}_i \boldsymbol{\beta})}. \quad \text{.....(4.24)}$$

Maximum likelihood estimation

The p.d.f. of Y_i is:

$$f(y_i) = \binom{n_i}{y_i} p_i^{y_i} (1 - p_i)^{n_i - y_i}, y_i = 0, 1, \ldots, n_i$$

The likelihood function is given by

$$k\ n_i\ p_{yi}\ \ l(\ )\square\ =$$

$$_{i\square\square 0}\square_{yi}\square\ \ _{i}(1\square\ p_i)_{n_i\square_{yi}}$$

$$\square\ \ \square_k\square\ \ \overline{\qquad}\ \ _{p_i}\square_{yi}(1\square\ p_i)_{n_i}$$

$$_{i\square 1\ 1\ \square\ p_i}$$

$$k \qquad\qquad n$$

$$\overline{\qquad\qquad}$$

$$\square\ _{i\square\square 0}e^{x_i\square\beta_{yi}}\square\square\square 1\ \square\ 1e^{x_i\beta}\square\square\square\ _{i}\square\ _{i\square\square k0}e^{x_i\square\beta_{yi}}\square 1\ \square\ e^{x_i\square\beta}\square_{\square n_i}$$

……(4.25)

The maximum likelihood estimates of $\square_0, \square_1,..., \square_k$ are the values of $\square_0, \square_1,..., \square_k$ which maximize the likelihood function. They are also the values of $\square_0, \square_1,..., \square_k$ which maximize:

$$L(\ )\boldsymbol{\beta}\ \square\ \ln\ (\ )l\ \boldsymbol{\beta}\ \square\ \square^k\ \boldsymbol{x}_i\square\boldsymbol{\beta}y_i\ \square\ \square^k\ n_i\ \ln\ 1\square\ \ \square\ e^{x_i\square\beta}\square.$$

$$_{i\square 0}\qquad _{i\square 0}$$

……(4.26)

The first derivative of $\boldsymbol{x_i}\square\boldsymbol{\beta}$ with respect to $\square_j$ is $x_{ij}\square$, thus:

$$\underline{\square L} = \square_k\ y\ x_{i\ ij}\square\ \square\ \square_k\ n_i\ \square\overline{\square e^{x_i\beta}x_i\beta}\ \square\square\square x_{ij}\square\ \square\ _{i\square\square k0}\ y\ x_{i\ ij}\square\ \square_{i\square\square k0}n\ p x_{i\ i\ ij}\square$$

$$\square\square_j \qquad i\square 0 i\square 0 \qquad \square 1\ \square\ e\ kk$$

$$k$$

$$= \qquad\qquad \square\ y\ x_{i\ ij}\square\ \square\ \square\ \square_{i\ ijx}\square\ \square\ \square\ \square y_i\ \square\ \square_i\ \square x_{ij}\square$$

……...(4.27)

$$i\square 0 \qquad i\square 0 \qquad i\square 0$$

where $\square\ \square_i\ E\ Y(\ )_i\ \square\ n\ p_{i\ i}$, p_i depends on the covariates x_i and $\square$ is a vector of $(k\ \square 1)$ parameters. Setting each partial derivative in (4.27) to zero, and replacing $\square_0, \square_1, ..., \square_k$ by $\square\ \square\hat{}_0, \hat{}_1, ..., \square\hat{}_k$ we obtain the maximum likelihood estimates of $\square_0, \square_1, ..., \square_k$. The methods of solution are iterative

in nature and have been programmed into logistic regression software. The interested reader may consult the text by McCullagh and Nelder (1989) for a general discussion of the methods used by most programs. The second derivatives used in computing the standard errors of the parameter estimates, $\hat{\beta}$, are

$$\frac{\partial^2 L}{\partial \beta_j \partial \beta_l} = -\sum_{i=0}^{k} n_i x_{ij} x_{il} \frac{e^{x_i\beta}}{\left(1+e^{x_i\beta}\right)^2}$$

$$= -\sum_{i=0}^{k} n_i p_i (1-p_i) x_{ij} x_{il} .$$

The comparison of observed to predicted values using the likelihood function is based on the following expression:

$$D = -2\ln\left[\frac{(\text{likelihood of the fitted model})}{(\text{likelihood of the saturated model})}\right]$$

$= 2[\ln(\text{likelihood of the saturated model}) - \ln(\text{likelihood of the fitted model})]$..(4.28) The log-likelihood of the fitted model can be written as:

$$L(\hat{\boldsymbol{\beta}}) = \sum_{i=0}^{k}\left[y_i \ln \hat{p}_i + (n_i - y_i)\ln(1-\hat{p}_i)\right]$$

$$= \sum_{i=0}^{k}\left[y_i \ln\left(\frac{\hat{y}_i}{n_i}\right) + (n_i - y_i)\ln\left(\frac{n_i - \hat{y}_i}{n_i}\right)\right]$$

……………………………………(4.29)

For the saturated model, we replace $\hat{y}_i$ in Equation (12) by y_i. Equation (4.28) then becomes

$$D = 2\sum_{i=0}^{k}\left[y_i \ln\left(\frac{y_i}{\hat{y}_i}\right) + (n_i - y_i)\ln\left(\frac{n_i - y_i}{n_i - \hat{y}_i}\right)\right],$$

………………………………………(4.30)

where, y_i = the observed and $\hat{y}_i$ = the fitted value for the i^{th} observation.

In particular, to assess the significance of an independent variable, we compare the value of D with and without the independent variable in the equation. The change in D due to the inclusion of the independent variable in the model is:

G = D(model without the variable) − D(model with the variable) = −2ln [(likelihood of the model without variable)/(likelihood of the fitted model)].

It can be shown that, when the variable is not in the model, the maximum likelihood estimate of β_0 is $\ln[m_1/m_0]$, where $m_1 = \sum_{i=0}^{k} y_i$ and $m_0 = \sum_{i=0}^{k} (n_i - y_i)$.

$$\hat{\pi} = m_1/n, \quad 1 - \hat{\pi} = m_0/n$$

$$G = -2\left[\sum_{i=0}^{k} y_i \ln\left(\frac{n_i \hat{\pi}}{y_i}\right) + \sum_{i=0}^{k} (n_i - y_i)\ln\left(\frac{n_i(1-\hat{\pi})}{n_i - y_i}\right)\right]$$

$$= 2\left[\sum_{i=0}^{k} y_i \ln \hat{p}_i + \sum_{i=0}^{k} (n_i - y_i)\ln(1 - \hat{p}_i)\right] - 2[m_1 \ln m_1 + m_0 \ln m_0 - n \ln n], \quad \text{……..(4.31)}$$

where $n = m_1 + m_0$. If the hypothesis that $\beta_j = 0$, $i = 1, 2, ..., k$ is true, then G has the chi-square distribution with k degrees of freedom (see Hosmer et al. (1989)).

4.5.3 Results

In this section, we illustrate the use of statistical packages in *R* to fit logistic regression models as a special case of a generalized linear model with family binomial and link logit. We first begin the analysis using `nlme` package in *R*. First, the data set, on road traffic casualties from 1991 to 2001

in Ghana, is loaded for analysis as shown in Listing (A9) in the appendix. Listing (A10) shows the fit of the logistic regression model to the data using the glm() function in *R*.

The results of the application of the *R* function „`summary(logistic)`", which presents the parameter estimate and standard errors for the model, are simplified in Table 4.18.

Table 4.18: Parameter estimates for logistic model of road traffic fatalities in Ghana from 1991 to 2001

j	0	1	2	3	4	5	6	7	8	9	10
Years	1991 Intercept	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001
Estimates $\hat{\beta}_j$	-2.25506	-0.04490	0.11258	0.02494	0.07179	0.01006	-0.07502	0.13810	0.14517	0.10721	0.18333
Standard errors	0.03465	0.04904	0.04941	0.05045	0.04781	0.04749	0.04777	0.04462	0.04591	0.04448	0.04335
Odds_i	0.10487	0.10026	0.11736	0.10752	0.11267	0.10593	0.09729	0.12040	0.12125	0.11673	0.12597
$\hat{p}i$	0.09491	0.09113	0.10504	0.09708	0.10126	0.09578	0.08866	0.10746	0.10814	0.10453	0.11188
z	-65.07200	-0.91600	2.27900	0.49400	1.50200	0.21200	-1.57100	3.09500	3.16200	2.41000	4.22900
p-value	2×10^{-16}	0.35992	0.02269	0.62102	0.13315	0.83223	0.11629	0.00197	0.00157	0.01593	2.35×10^{-5}

The fitted logistic equation, for the i^{th} year, is therefore given by $\ln\left(\frac{\hat{p}_i}{1-\hat{p}_i}\right) = -2.25506 - 0.04490x_{i1} - 0.11258x_{i2} \ldots - 0.18333x_{i10}$, ……(4.32)

where $x_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{otherwise,} \end{cases}$ which gives the odds for the i^{th} year as

$$\frac{\hat{p}_i}{1-\hat{p}_i} = \exp(-2.25506 - 0.04490x_{i1} - 0.11258x_{i2} \ldots - 0.18333x_{i10}).$$

Thus,

$$\hat{p}_i = \frac{1}{1+\exp(-2.25506 - 0.04490x_{i1} - 0.11258x_{i2} \ldots - 0.18333x_{i10})}. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(4.33)$$

For instance, when $i = 0$, $\hat{p}_0 = \frac{1}{1+\exp(-2.25506)} = 0.09491$,

which gives the estimate of the proportion of road traffic casualties who died in the year 1991. Note that, in computing for the value of $\hat{p}_0$, $x_{i1} = x_{i2} = \ldots = x_{i10} = 0$. Note further that, in computing for $\hat{p}_i$, $i = 0$, the predictor x_{ij} takes the value one (1) for $i = j$ while the remaining 9 predictors assume the value zero (0). Thus, from Table 2, when $i = 5$,

$$\hat{p}_5 = \frac{1}{1+\exp(-2.25506 - 0.04490(0) - 0.04490(0) - \ldots - 0.01006(1) - 0.01006(1) - \ldots - 0.18333(0) - 0.18333(0))}$$

$$= \frac{\exp(-2.25506 - 0.01006)}{1+\exp(-2.25506 - 0.01006)} = 0.09578.$$

The remaining values of $\hat{p}_i$ are given in Table 4.18. The method for specifying the design variables involves setting all of them equal to 0 for the reference year (1991), and then setting a single design variable equal to 1 for each of the other groups.

The significance of the logistic regression relationship can be assessed by using the null deviance to test the hypotheses

H_0: $\beta_j = 0$, $j = 1, 2, ..., 10$ against

H_1: not all the $\beta_i = 0$

at 0.05 level of significance. The test statistic is

$$G = 2\sum_{i=0}^{10} y_i \ln \hat{p}_i + (n_i - y_i)\ln(1 - \hat{p}_i)$$

$$= 12402 \ln(12402) + 110148 \ln(110148) + 12284 \ln(122550) = 74.182.$$

When H_0 is true, G has the chi-square distribution with 10 degrees of freedom (see Hosmer et al. (1989)). We reject H_0 at significance level 0.05 if the computed value of G is greater than $\chi^2_{0.05,10} = 18.31$. From the *R* function „summary(logistic)", the value of the test statistic is $G_0 = 74.182$. Since 74.182, the calculated value of G, is greater than 18.31, the test is significant at the 5% level. We therefore reject the null hypothesis in this case and conclude that at least one or more of the 10 coefficients is different from zero.

Since the analysis indicates that the null hypothesis should be rejected at the 5% level, it means that some of the coefficients are significantly different from zero. But as to which of the coefficients are significantly different, the analysis does not specify. Before concluding that any or all of the coefficients are nonzero, we may look at the univariable Wald test statistics (see Hosmer et al. (1989)),

$$W_j = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \quad \text{.....(4.34)}$$

These are shown in the seventh column, labeled z, in Table 2. Under the hypothesis that the i^{th} coefficient is zero, W_i has the standard normal distribution. The p-values computed under this

hypothesis are shown in the eighth column of Table 2. If we use a level of significance of 0.05, then we would conclude that the coefficients for the years 1993, 1998, 1999, 2000 and 2001 are significantly different from zero, while that of the remaining year are not significant. That is, there is little statistical justification for including coefficients for the years 1992, 1994, 1995, 1996 and 1998 in the model. However, according to Hosmer et al. (1989), we must not base our models entirely on tests of statistical significance. There are numerous other considerations that influence our decision to include variables in a model. This is based on the fact that it is possible for individual variables not to exhibit strong confounding, but when taken collectively, considerable confounding can be present in the data; see Rothman et al. (2008), Maldonado and Greenland (1993), Greenland (1989), and Miettinen (1976).

The purpose of analysing these data is not the determination of the parameters. Interest is centered on how good the model is in estimating future road traffic fatality values using these estimates. At this stage, we wish to use the model in Equation (15) to estimate the number of road traffic fatalities from the year 2002 to 2011, a period of ten years. To do this, a single design variable x_{ij}, for year i, is set equal to 2 when $i = j$ and then all remaining variables are set equal to 0, where i represents any of the years from 2002 to 2011. We use $x_i1, x_i2, ..., x_i10$ in Equation (15) as our design variables for the year 2002, 2003, …, 2011, respectively. For instance, in year 3 (i.e. the year 2004), the design variables together with the corresponding parameter estimates are given in Table 4.19.

Table 4.19: Design variables for year 2004

Year	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Variables for 2004	$x_{31} = 0$	$x_{32} = 0$	$x_{33} = 2$	$x_{34} = 0$	$x_{35} = 0$	$x_{36} = 0$	$x_{37} = 0$	$x_{38} = 0$	$x_{39} = 0$	$x_{3,10} = 0$
$\hat{\beta}_j$	-0.04490	0.11258	0.02494	0.07179	0.01006	-0.07502	0.13810	0.14517	0.10721	0.18333

Thus, a point estimate of the proportion of road traffic casualties who died in 2004 is given by

(see Equation (16)) $\hat{p}_{2004} = \frac{1}{1 + \exp(2.25506) \times \exp(2.25506 + 0.02494(2) + 0.02494(2))} = 0.09929.$

From Table 1, the total number of road traffic casualties in 2004 is 18 445. Thus a point estimate of the total number of road traffic fatalities in 2004 is (to the nearest whole number)

$$\hat{D}_{2004} = 18\,445 \times 0.09929 = 1831.$$

The actual road traffic fatalities D together with the values of $\hat{D}$ calculated from Equation (4.33) are given in Table 4.20. The percentage differences between the calculated and actual values are also given in Table 4.20.

It can be seen that all the calculated figures, $\hat{D}$, corresponding to the coefficients, β_j, that were significantly different from 0, are within 10% of the actual figure and only one (i.e. 0.07179) of the five coefficients, that were not significantly different from 0, estimated D within 10% (i.e. 3.9%) of its actual value.

Table 4.20: Comparison of actual fatalities and fatalities estimated from Equation (4.33)

j	1	2	3	4	5	6	7	8	9	10
Year	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
$\hat{\beta}_j$	-0.04490	**0.11258**	0.02494	0.07179	0.01006	-0.07502	**0.13810**	**0.14517**	**0.10721**	**0.18333**
D	1665	1716	2186	1776	1856	2043	1938	2237	1986	2199
$\hat{D}$	1318.9	1879.1	1831.3	1707.3	1580.1	1359.0	2000.0	2274.2	1944.0	2131.6
Error	346.1	-163.1	354.7	68.7	275.9	684.0	-62.0	-37.2	42.0	67.4
Error %	20.8	**9.5**	16.2	3.9	14.9	33.5	**3.2**	**1.7**	**2.1**	**3.1**

4.5.4 Conclusion

Logistic regression analysis of road traffic fatalities in Ghana has been performed using road traffic accident data from the National Road Safety Commission. The data span from 1991 to 2001. The formula for predicting the proportion of road traffic casualties who die in the i^{th} year using a logistic regression approach is

$$\ln\left(\frac{\hat{p}_i}{1-\hat{p}_i}\right) = \beta_0 + \beta_1 x_i + \ldots + \beta_{10} x_{i10}$$, where the values of the parameters $\beta_0, \beta_1, \ldots, \beta_{10}$ are given in Table 2. Using the model to estimate the number of road traffic fatalities from 2002 to 2011 in the Ghana, it was noted that of the 10 calculated figures, 6 are within 10% of the actual figure.

CHAPTER FIVE

VALIDATION OF BAYESIAN AND MULTILEVEL METHODS USING DATA FROM GHANA

5.0 Introduction

In Section 3.1, a modified Smeed's model

$$\frac{D}{P} = \alpha \left(\frac{N}{P}\right)^{\beta} u, \quad \text{.....(5.1)}$$

was developed, where D = Number of RTFs, P = population size, N = number of vehicles in use, u = multiplicative error term, and α & β are parameters to be estimated. It was shown that the parameters of the model can vary from one geographical region to another and hence, can be used to assess variability of risk of road traffic fatalities across geographical regions of a given geographical zone. It was shown that Equation (5.1), being intrinsically linear, can be transformed by a logarithmic transformation given as

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad \text{.....(5.2)}$$

where $\beta_0 = \ln\alpha$, $\beta_1 = \beta$, $x_i = \ln(N/P)$, $y_i = \ln(D/P)$ and $\varepsilon_i = \ln u_i$, $i = 1, 2, \ldots, n$

This Chapter seek to use Ghana data to validate the methods developed in Chapter 3 of this study.

For comparative purposes, Section 5.1 of this Chapter uses the least squares regression method to estimate the parameters of the modified Smeed's model.

In Section 5.2, the study seeks to use data from Ghana to validate the Bayesian method and to assess the robustness of the model. The questions to be addressed are:

- How accurate is the Bayesian method for estimating parameters of the modified Smeed's model compare to that of least squares regression method?
- How accurate is the proposed modified Smeed's model of this study, in Ghana and how does the modified model compare to that of Smeed (1949) in their performance?

The final section of this chapter uses data from Ghana to validate the multilevel method and compare the risk of RTFs across the 10 geographical regions in Ghana.

Table 5.1 gives the estimated population size and the number of motor vehicles and road traffic fatalities in Ghana (1991 – 2012).

Table 5.1: Estimated Population and the number of motor vehicles, fatalities and casualties in Ghana (1991-2012)

No.	Year	Population (P)	Motor Vehicles (N)	Fatalities (D)	$x_i = \ln\left(\frac{N}{P}\right)$	$y_i = \ln\left(\frac{D}{P}\right)$
1	1991	14821000	132051	920	-4.7206	-9.68718
2	1992	15222000	137966	914	-4.7035	-9.72042
3	1993	15634000	157782	901	-4.5960	-9.76145
4	1994	16056000	193198	824	-4.4201	-9.87742
5	1995	16491000	234962	1026	-4.2511	-9.6849
6	1996	16937000	297475	1049	-4.0419	-9.68942
7	1997	17395000	340913	1015	-3.9323	-9.74905
8	1998	17865000	393225	1419	-3.8162	-9.44065
9	1999	18349000	458182	1237	-3.6901	-9.60464
10	2000	18845000	511063	1437	-3.6075	-9.48145
11	2001	19328000	567780	1660	-3.5276	-9.36249
12	2002	19811000	613153	1665	-3.4754	-9.38417
13	2003	20508000	643824	1716	-3.4611	-9.38857
14	2004	21093000	703372	2186	-3.4008	-9.17462
15	2005	21694000	767067	1776	-3.3422	-9.41043
16	2006	22294000	841314	1856	-3.2771	-9.39365
17	2007	22911000	922748	2043	-3.2120	-9.32495
18	2008	23544000	942000	1938	-3.2186	-9.40497
19	2009	24196000	1030000	2237	-3.1566	-9.28881
20	2010	24223000	1122700	1986	-3.0716	-9.40894
21	2011	25099000	1225754	2199	-3.0193	-9.34258
22	2012	25726000	1328808	2249	-2.9632	-9.34477

5.1 A Least Squares Regression Method

5.1.1 Estimation of Regression Parameters

This transformation in (5.2) requires that $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{19}$ are normally and independently distributed with mean 0 and variance σ^2. The least squares estimates of β_0 and β_1 are the values of β_0 and β_1 which minimize

$$Q = \sum_{i=1}^{19} \left[y_i - (\beta_0 + \beta_1 x_i) \right]^2 . \quad \ldots\ldots(5.3)$$

Equating these partial derivatives, $\frac{\partial Q}{\partial \beta_0}$ and $\frac{\partial Q}{\partial \beta_1}$, to zero (because the partial derivatives are equal to zero at the minimum point) and replacing β_0 and β_1 by $\hat{\beta}_0$ and $\hat{\beta}_1$, we obtain

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{19} x_i y_i - \frac{1}{19}\left(\sum_{i=1}^{19} x_i\right)\left(\sum_{i=1}^{19} y_i\right)}{\sum_{i=1}^{19} x_i^2 - \frac{1}{19}\left(\sum_{i=1}^{19} x_i\right)^2} = \frac{S_{xy}}{S_{xx}} \quad \ldots\ldots(5.4)$$

and

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}. \quad \ldots\ldots(5.5)$$

The values of $\hat{\beta}_0$ and $\hat{\beta}_1$ are determined using Table A8, in the appendix, which gives the values of $y_i = \ln(D/P)$ and $x_i = \ln(N/P)$ for the 19-year period 1991 – 2009. From Table A10, $S_{xx} = 5.03764$ and $S_{xy} = 1.60457$. Using these results in Equations (5.4) and (5.5), we obtain

$$\hat{\beta}_1 = 0.318516 \text{ and } \hat{\beta}_0 = -8.31275.$$

5.1.2 Validation of Regression Relation

The significance of the regression relationship can be assessed by using analysis of variance techniques to test the null hypothesis H_0: $\beta_1 = 0$ against the alternative hypothesis H_1: $\beta_1 \neq 0$ at 0.05 level of significance. The sum of squares due to linear regression is given by S_2

$$SSR = S_{xy}^2/S_{xx} = (1.60457)^2/5.037642 = 0.5110806.$$

The total corrected sum of squares is given by

$$SST = S_{yy} = \sum_{i=1}^{19} y_i^2 - \left(\sum_{i=1}^{19} y_i\right)^2/19 = 0.6887237.$$

Therefore, the residual sum of squares is $SSE = SST - SSR = 0.034558938$.
The calculations can be summarized in the following ANOVA table.

Table 5.2: Analysis of Variance table

Source of variation	Sum of squares	Degrees of freedom	Mean square	F
Linear regression	0.5110806	1	0.511081	48.90913
Residual	0.1776431	17	0.01045	
Total	0.6887237	18		

The test statistic for testing H_0 against H_1 is $F = \frac{\text{regression mean square}}{\text{residual mean square}}$. When H_0 is true, F has the F-distribution with 1 and 17 degrees of freedom. We reject H_0 at significance level 0.05 if the computed value of F is greater than $F_{0.05,\,1,\,17} = 4.45$. Since 48.9, the calculated value of F, is greater than 4.45, the test is significant at the 5% level. There is enough evidence to conclude that there is a linear relationship between the expected value of $y = \ln(D/P)$ and $x = \ln(N/P)$. The coefficient of determination (R^2) value from the LSM is

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{0.177767}{0.689854} = 0.742069135.$$

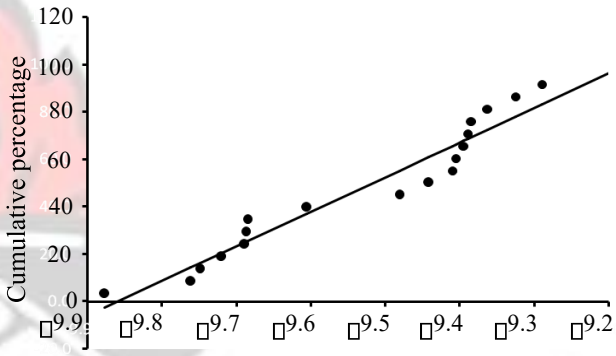
0.7423 indicating 74.23% of the variability in the response data is explained by the predictor variables.

5.1.3 Validation of the normality assumption

The observations y_i, $i = 1, 2, ..., 19$, are first ordered from the smallest to largest. The ranked values $y_{(i)}$, $i = 1, ..., 19$, the corresponding percentage cumulative $p_i = 100(i - 3/8)/(n + 1/4)$ and standardized residuals $d_i = e_i/\sqrt{\sigma^2}$, $i = 1, 2, ..., 19$ are given in Table A9, where $e_i = y_i - \hat{y}_i$, $i = 1, 2, ..., 19$ and $\sigma^2 = \hat{\sigma}^2 = 0.01045$ is the estimate of the population variance.

Fig. 5.1 shows the probability plot of the observations. It can be seen from Figure 5.1 that the observations are closed to normal because of how well the points follow the line.

It can be seen from Table A9, in the appendix, that of the 19 calculated standardized residuals, 18 are within the interval $(-1.96, 1.96)$ which represents about 95%.



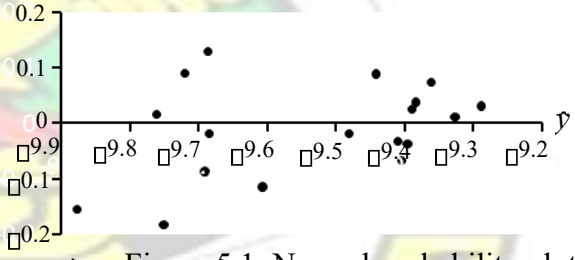


Figure 5.1: Normal probability plot

It is frequently helpful to plot residuals against $\hat{y}_i$ as shown in Figure 5.2. Pattern in the $_i$ plot represents the ideal situation satisfactory for normality (see Ofosu et al., 2013).

Figure 5.2 Pattern for the residual plot There is strong evidence to conclude that, at sample size of 19, the errors are normally distributed. Thus, given $\hat{\beta}_0 = e^{-8.31275} = 0.000245$ and $\hat{\beta}_1 = 0.318516$, Equation (5.1) becomes

$$D/P = 0.000245 \times N/P^{0.318516}. \quad \text{.....(5.6)}$$

5.2 A Bayesian Method

5.2.1 Introduction

In this Section, the study wishes to determine the Bayesian estimates of the parameters, β_0 and β_1, using the conjugate prior and maximum a posteriori approaches and also test the robustness of the Bayesian method with respect to the normality assumption of the model.

The ***prior distribution*** is a key part of Bayesian inference and represents the information about an uncertain parameter β that is combined with the probability distribution of new data to yield the ***posterior distribution*** which in turn is used for future inferences and decisions involving β. The range of possible value that the regression coefficients β_0 and β_1 can take is from $-\infty$ to $+\infty$. Thus, the largest possible domain of the prior distribution is the set of all real numbers. This limits us to distribution which can take both negative and positive values. Therefore, the most suitable prior distributions are the bivariate Normal, Laplace and Cauchy distributions. The maximum a posteriori method is used with respect to the Laplace and Cauchy prior distributions while for the bivariate normal prior both methods are used.

5.2.2 Conjugate Prior Method

Based on Equation (3.25) and the linear regression model of Equation (5.3), the conditional p.d.f. of Y is

$$f_Y(y_i|\beta_0, \beta_1) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(y_i - \beta_0 - \beta_1 x_2 - \cdots)^2\right], \quad y_i = 0. \quad \text{.....(5.8)}$$

Thus, the likelihood function is

$$f_{\mathbf{Y}}(\mathbf{y}|\beta_0, \beta_1) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2\right), \quad y > 0 \quad \text{.....(5.9)}$$

From Equation (3.28), the p.d.f. of the bivariate normal prior distribution of $\boldsymbol{\beta} = (\beta_0, \beta_1)$, with mean vector $\boldsymbol{\mu} = (\mu_0, \mu_1)$ and covariance matrix Σ is

$$p(\beta_0, \beta_1) = k_2 \exp\left(-\frac{1}{2}(a_{00}(\beta_0 - \mu_0)^2 + a_{11}(\beta_1 - \mu_1)^2 + 2a_{01}(\beta_0 - \mu_0)(\beta_1 - \mu_1)\right), \quad \text{.....(5.10)}$$

where $a_{11} = a_{00}$

$a_{01} = a_{10}$.

Thus, the posterior distribution in Equation (3.32) can be expressed as

$$p(\beta_0, \beta_1|\mathbf{y}) = k\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{19}(y_i - \beta_0 - \beta_1 x_i)^2 - \frac{1}{2}(a_{00}(\beta_0 - \mu_0)^2 + a_{11}(\beta_1 - \mu_1)^2 + 2a_{01}(\beta_0 - \mu_0)(\beta_1 - \mu_1)\right), \quad \text{.....(5.11)}$$

Equation (3.35), with a sample size of 19, therefore becomes

$$Q(\boldsymbol{\beta}) = -\frac{19}{2}(a_{00}\beta_0^2 + 2a_{01}\beta_0\beta_1 + a_{11}\beta_1^2) + \beta_0\sum_{i=1}^{19}y_i + \beta_1\sum_{i=1}^{19}x_i y_i - (a_{00}\mu_0 + a_{01}\mu_1)\beta_0 - (a_{01}\mu_0 + a_{11}\mu_1)\beta_1$$

$$-2\beta_0\sum_{i=1}^{19}x_i - \beta_1\sum_{i=1}^{19}x_i^2 - a_{01}\mu_0\beta_1 - a_{01}\mu_1\beta_0 + 2\beta_0\sum_{i=1}^{19}x_i + a_{01}\mu_0\beta_1 + a_{01}\mu_1\beta_0$$

$$\sum_{i=1}^{19} \frac{1}{2}\left(y_i^2 - a_{00} - \beta_0^2 a_{11} - \beta_1^2 2^{a_{01}} 0 1\right) \quad \text{.....(5.12)}$$

Hence, Equation (3.37) becomes

$$\boldsymbol{\mu}_{\beta} = -\frac{1}{2}\boldsymbol{\Sigma}_{\beta}\boldsymbol{C}, \quad \text{.....(5.13)}$$

where $\boldsymbol{\Sigma}_{\beta}$ is a 2 2 matrix with inverse $\boldsymbol{\Sigma}_{\beta}^{-1} = m_{ij}$ whose elements are given as

$$\left.\begin{aligned} m_{00} &= \frac{19}{2} - a_{00}, \\ m_{01} &= \frac{1}{2}\sum_{i=1}^{19} x_i, \\ 01, & \end{aligned}\right\}$$

$$19 \quad \text{.....(5.14)} \quad m_{10} = \frac{1}{2}\sum_{i=1}^{19} x_i - 10,$$

$$m_{22} = \frac{1}{2}\sum_{i=1}^{19} x_{i2} - a_{11}.$$

$\boldsymbol{C}$ is a column vector of order (2 1) with element given as

$$C_1 = 2 \frac{1}{2} i19 1 y_i - a_{00} 0 a_{01} 1 \quad \text{.....(5.15)}$$

$$C_2 = 2 \frac{1}{2} i19 1 x y_i i - a_{11} 1 - a_{01} 0$$

The 19 jackknife sample estimates of β_0 and β_1, based on the national data, derived from the values of y_i and x_i in Table 5.1 are given in Table 5.3.

Table 5.3: Jackknife estimates of β_0 and β_1

Parameters	1	2	3	4	5	6	7	8	9	10
β_0	-8.2025	-8.2373	-8.3012	-8.3889	-8.3177	-8.3254	-8.3236	-8.3143	-8.2968	-8.308
β_1	0.35003	0.34012	0.32186	0.29589	0.31694	0.31386	0.31296	0.31942	0.32103	0.31951
Parameters	11	12	13	14	15	16	17	18	19	**Mean**
β_0	-8.331	-8.3226	-8.3202	-8.3933	-8.2978	-8.2939	-8.3177	-8.2757	-8.3294	**-8.3105**
β_1	0.3148	0.31646	0.31696	0.30058	0.32197	0.32294	0.31741	0.32727	0.31461	**0.31916**

Based on Equations (3.40) and (3.41), jackknife estimate of the mean vector and covariance of the random vector **β** is computed as follows

$$\hat{\boldsymbol{\mu}} = (-8.3105,\ 0.3192) \quad \text{and} \quad \hat{\boldsymbol{\Sigma}} = \begin{pmatrix} 0.001860 & 0.000504 \\ 0.000504 & 0.000139 \end{pmatrix}.$$

Based on Equations (5.14) and (5.15),

$$\hat{\boldsymbol{\Sigma}}_{\beta} = \begin{pmatrix} 0.0004712 & 0.0017421 \\ 0.0001297 & 0.0004712 \end{pmatrix} \quad \text{and} \quad \hat{C} = \begin{pmatrix} 646278.208 \\ 2353969.324 \end{pmatrix}.$$

Thus, from Equation (5.13) the posterior Bayes estimate of **β** is given by

$$\hat{\boldsymbol{\mu}}_{\beta} = \tfrac{1}{2}\boldsymbol{\Sigma}_{\beta} C = \begin{pmatrix} -8.31048 \\ 0.319162 \end{pmatrix}. \qquad \text{...(5.16)}$$

The sum of square due to error (*SSE*) of the conjugate prior method is computed as

$$SSE = \sum_{i=1}^{19} \left(y_i - \hat{y}_i\right)^2 = 0.177767.$$

The total sum of squares is given by

19 2

$$SST = \sum_{i=1}^{} (y_i - \bar{y}_i)^2 = 0.689854.$$

The coefficient of determination of the conjugate prior method is given by

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{0.177767}{0.689854} = 0.74231135.$$

Table 5.4 shows the coefficients estimates and the corresponding standard errors for the least square and the conjugate prior methods.

Table 5.4: Comparison of Coefficients of Least Square and Conjugate Prior Methods

	Model			
	Least Squares		**Conjugate Prior**	
	Coefficient	Standard Error	Coefficient	Standard Error
β_0 = intercept	–8.31179	0.17386	–8.31048	0.04174
β_1 = coefficient of *x*	0.31879	0.04555	0.31916	0.01139
Coefficient of determination	0.7423		0.7423	

It can be seen from Table 5.4, that the estimated coefficients β_0 and β_1, are almost the same for the least squares and the conjugate prior methods. Both methods also reported the same coefficient of determination R^2. The conjugate prior estimates recorded comparatively very small standard errors; making the conjugate prior method preferred.

5.2.3 Maximum a posteriori method

In Bayesian data analysis, one way to apply a model to data is to find the maximum a posteriori (MAP) parameter values. Our objective here is to determine the parameter estimates that maximize the posterior distribution given the data with respect to the bivariate Normal, Laplace and Cauchy prior distributions.

Bivariate Normal prior distribution

The prior distribution in Equation (3.25) can be written in terms of β as

$$f(\beta_0, \beta_1) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2}q}, \quad -\infty < \beta_0 < \infty, \quad -\infty < \beta_1 < \infty, \text{..................(5.17)}$$

where

$$q = \frac{1}{1-\rho^2}\left[\left(\frac{\beta_0-\mu_0}{\sigma_1}\right)^2 - 2\rho\left(\frac{\beta_0-\mu_0}{\sigma_1}\right)\left(\frac{\beta_1-\mu_1}{\sigma_2}\right) + \left(\frac{\beta_1-\mu_1}{\sigma_2}\right)^2\right], \quad \sigma_1^2 = \text{var}(\beta_0), \ \sigma_2^2 = \text{var}(\beta_1),$$

$$\rho = \frac{\text{cov}(\beta_0, \beta_1)}{\sigma_1\sigma_2}, \quad -\infty < \mu_0 < \infty, \ -\infty < \mu_1 < \infty.$$

Thus, the posterior distribution can be expressed as

$$p(\beta_0, \beta_1 \mid \boldsymbol{y}) = k \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{19}\left(y_i - \beta_0 - \beta_1 x_i\right)^2 - \frac{1}{2}q\right\}. \text{..........................(5.18)}$$

where $k = \left[\left(2\pi\sigma^2\right)^{\frac{19}{2}} 2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}\right]^{-1}$

To find the maximum a posteriori (MAP) parameter values, we can resort to Markov chain Monte Carlo (MCMC) sampling techniques to get samples from the posterior distribution. The maximum a posteriori (MAP) parameter correspond to

$$\boldsymbol{\beta}_{MAP} = \underset{\boldsymbol{\beta}}{\text{argmax}}\ p(\beta_0, \beta_1 \mid \boldsymbol{y}). \text{...(5.19)}$$

Suitable MCMC approach is the Metropolis Hastings (MH) sampler as discussed in Section (3.3.3) of Chapter 3. The following is the description algorithm for the procedure: 1. Set $t = 1$

2. Generate an initial value for $\beta_0 \sim U(10, 4)$ and $\beta_1 \sim U(0, 1)$.

3. Repeat

   $t = t + 1$

   Do a MH step on β_0

   Generate a proposal $\beta_0^* \sim N(\beta_0, 0.002)$;

   $$p(\beta_0^*, \beta_1 | y)$$

   Evaluate the acceptance probability $a = \min 1,$;

   $$p(0 1, y)$$

   Generate a u from a Uniform(0, 1) distribution

   If $u \leq a$, accept the proposal and set $\beta_0 = \beta_0^*$,

   Do a MH step on β_1,

   Generate a proposal $\beta_1^* \sim N(\beta_1, 0.0001)$;

   $$p(\beta_0, \beta_1^* | y)$$

   Evaluate the acceptance probability $a = \min 1,$;

   $$p(0 1, y)$$

   Generate a u from a Uniform(0, 1) distribution

   If $u \leq a$, accept the proposal and set $\beta_1 = \beta_1^*$,

4. Until $t = 5000$.

The MATLAB code for the implementation of component-wise Metropolis sampler for the posterior distribution is as given in Listings A1 and A2 in the appendix.

Table 5.5 shows estimated values of β_0 and β_1 based on least squares, conjugate prior and maximum a posteriori methods. The results show that the estimated coefficients of β_0 and β_1 are almost the same for the least squares, conjugate prior and maximum a posteriori methods of estimates.

Table 5.5: Comparison of Coefficients of Least Squares, Conjugate Prior and maximum a Posteriori Methods

	Methods		
	Least Square	**Conjugate Prior**	**Maximum a Posteriori**
□0 (**Standard error**)	–8.31179 (**0.17386**)	–8.31048 (**0.04174**)	–8.29094 (**0.03978**)
□1 (**Standard error**)	0.31879 (**0.04555**)	0.31916 (**0.01139**)	0.32460 (**0.01098**)

Laplace Prior Distribution

It is assumed that □□□β , β0 1□ has a bivariate Laplace distribution with mean vector $\boldsymbol{\mu}$ □ □ □□ 0, 0□. The joint p.d.f. is given by

$$f(\beta_0, \beta_1) = \frac{1}{4b_1b_2} e^{-\left(\frac{|\beta_0 - \mu_0|}{b_1} + \frac{|\beta_1 - \mu_1|}{b_2}\right)}, \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(5.20)$$

□□□□ □□ □□□□ □□0 , 1 , b_1□ 0, b_2 □ 0. Thus, the posterior distribution can be expressed as

$$p(\beta_0, \beta_1 | \boldsymbol{y}) = k \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2 - \frac{1}{b_1}|\beta_0 - \mu_0| - \frac{1}{b_2}|\beta_1 - \mu_1|\right], \quad \ldots\ldots\ldots(5.21)$$

□□ □□ □ □0 , □□ □□ □ □1 , b_1 □ 0, b_2 □ 0. It can be show that $E(\beta_0) = \mu_0$, $E(\beta_1) = \mu_1$, var( )□ □ $2b_1^2$ and var( )□ □ $2b_2^2$ (see Norton 1984). Given n independent and identically distributed sample □01, □02, ..., □0n and □11, □12, ..., □1n, the maximum likelihood estimators of □1 and □2 are given by

□ □^0 median of (□01, □02, ..., □0n) and □ □^1 median of (□11, □12, ..., □1n), ……(5.22) and the maximum likelihood estimators of b_1 and b_2 are (see Norton 1984)

$$\frac{n}{n}$$

$$\hat{b}_0 = \frac{1}{n}\sum_{i=1}^{n}\left|\beta_0 - \hat{\beta}_0\right| \quad \text{and} \quad \hat{b}_1 = \frac{1}{n}\sum_{i=1}^{n}\left|\beta_1 - \hat{\beta}_1\right| \quad \text{.....(5.23)}$$

Using the 19 jackknife sample estimates of β_0 and β_1 in Table 5.3, the maximum likelihood estimates of β_0 and β_1 are $\hat{\beta}_0$ 8.3177 and $\hat{\beta}_0$ 0.31741 while that of b_0 and b_1 are $\hat{b}_0$ 0.02761 and $\hat{b}_1$ 0.007424 respectively. Thus, the posterior distribution can be expressed as

$$p(\beta_0, \beta_1, \mathbf{y}) = k \exp\left[-\frac{1}{19}\sum_{i}\left|y_i - (\beta_0 + \beta_1 x_i)\right|^2 - \frac{1}{\hat{b}_1}\left|\beta_1 - \hat{\beta}_0\right| - \frac{1}{\hat{b}_2}\left|\beta_0 - \hat{\beta}_1\right|\right] \quad \text{.....(5.24)}$$

Since Equation (5.12) does not correspond to any analytic expression, the bivariate componentwise Metropolis-Hastings sampler can be used to determine the maximum a posteriori Bayesian estimates. The implementation of component-wise Metropolis-Hastings sampler for the posterior distribution, using MATLAB codes similar to that of Listings A1 and A2 in the appendix, gave a maximum a posteriori Bayesian estimates of β_0 and β_1 to be –8.320085 and 0.317051 with standard errors of 0.039047 and 0.010450 respectively.

Cauchy Prior Distribution

The bivariate random variable (β_0, β_1) has the Cauchy distribution if the p.d.f. can be expressed in the form given in the following form

$$f(\beta_0, \beta_1) = \frac{1}{2\pi}\left[\frac{a_1}{(\beta_0 - \alpha_1)^2 + a_1^2}\right]\left[\frac{a_2}{(\beta_1 - \alpha_2)^2 + a_2^2}\right], \quad -\infty < \beta_0 < \infty, \quad -\infty < \beta_1 < \infty \quad \text{.....(5.25)}$$

The moment estimators of β_0 and β_1 do not exist while the maximum likelihood estimates tend to be complicated by the fact that this requires finding the roots of high degree polynomial. One simple method is to take median value of the samples from the random variables β_0 and β_1 as an

estimator of the parameters a and b. Thus, from Table 5.9, the estimates of a and b are $\hat{a} = -8.3177$ and $\hat{b} = 0.31741$. Thus, the posterior distribution can be expressed as

$$p(\beta_0, \beta_1 | \mathbf{y}) = k \exp\left[-\frac{1}{2}\sum_{i=1}^{19}(y_i - \beta_0 - \beta_1 x_i)^2\right] \frac{1}{\sqrt{2\pi}(1+\hat{b})^2} \frac{1}{\sqrt{2\pi}}. \quad \ldots\ldots(5.26)$$

The component-wise Metropolis-Hastings sampler for the posterior distribution based on the MATLAB codes, gave maximum a posteriori estimates of β and $\beta_{0\,1}$ to be –8.312857 and 0.317400, respectively.

The resulting posterior Bayesian estimates for the Normal, Laplace and Cauchy prior distributions are summarized in the Table 5.6. Given a sample size 19, the posterior Bayes estimate is reasonably consistent for the Normal, Laplace and Cauchy prior distributions.

Table 5.6: Posterior Bayesian estimates for different priors with a sample size of 19

	Prior distribution				
	Normal		**Laplace**		**Cauchy**
	Estimate	Standard Error	Estimate	Standard Error	
β_0	–8.31048	0.04174	–8.32009	0.039047	–8.31286
β_1	0.31916	0.01139	0.31705	0.010450	0.31740

Table 5.7 shows the posterior Bayesian estimates of β_0 and β_1 at four different sample sizes (5, 10, 15 and 19) using the Normal, Laplace and Cauchy prior distributions.

Table 5.7: Bayesian estimates with respect to sample size and prior distribution

	Prior distribution					
Sample size	**Normal**		**Laplace**		**Cauchy**	
n	β_0	β_1	β_0	β_1	β_0	β_1
5 **(Standard error)**	–5.99608 **(0.67355)**	0. 99041 **(0.02767)**	–8.30608 **(0.61978)**	0. 31923 **(0.02195)**	–5.13317	1.01961

10 (**Standard error**)	–8.29381 (**0.44057**)	0.32272 (**0.01629**)	–8.29637 (**0.43884**)	0.32863 (**0.01596**)	–7.72230	0.46478
15 (**Standard error**)	–8.31195 (**0.36057**)	0.31647 (**0.01328**)	–8.29288 (**0.35747**)	0.32266 (**0.01298**)	–8.31034	0.31694
19 (**Standard error**)	–8.31048 (**0.31916**)	0.31916 (**0.01139**)	–8.32009 (**0.31705**)	0.31705 (**0.01045**)	–8.31286	0.31740

It can be seen that, at sample sizes of 5 and 10, the posterior Bayesian estimates of β_0 and β_1 are not consistent across the three prior distributions used. Thus, the estimated values of β_0 and β_1 are said to be sensitive with respect to the prior distribution. At a sample size of 15 or more, the model becomes insensitive to the prior distribution. The relative influence of the prior distribution decreases while that of the data increases with a sample size of 15 or more. It can also be seen that the posterior Bayesian estimate is reasonably consistent for the Laplace prior distribution across all four sample sizes used. Even at a sample size of 5 where the normality assumption was violated, the estimates based on the Laplace prior distribution was robust. Thus, the Laplace prior distribution is preferred when the sample size is small.

Comparison of actual fatalities and estimated fatalities from Smeed's equation

Smeed's equation was used to calculate the number of fatalities for each year in the 19-year period 1991 – 2009, using the data given in Table 5.1. The results of this application are given in Table 5.7, where: D = annual road deaths and $D\hat{}$ = estimate of D from Smeed's equation.

Table 5.7: Comparison of actual fatalities and estimated fatalities from Smeed's equation

No.	Year	D	$D\hat{}$	Error	Error %	No.	Year	D	$D\hat{}$	Error	Error %
1	1991	920	922	2	0.2	**11**	2001	1660	1789	129	7.8
2	1992	914	952	38	4.2	**12**	2002	1665	1866	201	12.1
3	1993	901	1014	113	12.5	**13**	2003	1716	1941	225	13.1
4	1994	824	1104	280	34.0	**14**	2004	2186	2037	-149	6.8
5	1995	1026	1199	173	16.9	**15**	2005	1776	2136	360	20.3
6	1996	1049	1321	272	25.9	**16**	2006	1856	2243	387	20.9

7	1997	1015	1407	392	38.6	**17**	2007	2043	2356	313	15.3
8	1998	1419	1502	83	5.8	**18**	2008	1938	2416	478	24.7
9	1999	1237	1609	372	30.1	**19**	2009	2237	2535	298	13.3
10	2000	1437	1699	262	18.2		**Total**	**27819**	**32048**		

It can be seen from Table 5.7 that, the application of Smeed's equation leads to over-estimation of the number of road traffic fatalities (RTFs) in Ghana. The result shows that on average, the expected fatalities as estimated by Smeed's formula exceeded the observed by 17%. The paired *t*-test is used to test the null hypothesis that there is no difference between actual RTFs and the estimated RTFs. The value of the test statistic for the two equal-tail test is 6.154 with a *p*-value of 0.0000082. Since the *p*-value is less than 0.05, we reject the null hypothesis at the 5% level. We therefore conclude that there is a significant difference between the observed RTFs and those estimated from Smeed's formula, at the 5% level. A Levene's test (Levene 1960) shows that the variances of the two sets of data are homogeneous. Given the relatively large deviations between observed and expected values of *D* in Table 5.7, Smeed's equation has proved to be an imperfect predictive tool of RTFs in Ghana.

Comparison of actual fatalities and fatalities estimated from Modified Smeed's Equation

The modified Smeed's equation is given by

$$D/P = 0.000245(N/P)^{0.318516}. \quad \text{.....(5.27)}$$

The actual fatalities *D* together with the values of $\hat{D}$ calculated from Equation (5.14) are given in Table 5.8. The differences between the calculated and actual values are also given.

Table 5.8: Comparison of actual fatalities and fatalities estimated from Equation (5.14)

No.	Year	*D*	$\hat{D}$	Error	Error %	No.	Year	*D*	$\hat{D}$	Error	Error %
1	1991	920	807	113	12.2	**12**	2002	1665	1604	61	3.6
2	1992	914	834	80	8.8	**13**	2003	1716	1668	48	2.8
3	1993	901	886	15	1.7	**14**	2004	2186	1749	437	20.0

4	1994	824	962	-138	16.8	**15**	2005	1776	1833	-57	3.2
5	1995	1026	1043	-17	1.7	**16**	2006	1856	1923	-67	3.6
6	1996	1049	1145	-96	9.2	**17**	2007	2043	2018	25	1.2
7	1997	1015	1218	-203	20.0	**18**	2008	1938	2069	-131	6.8
8	1998	1419	1298	121	8.5	**19**	2009	2237	2169	68	3.0
9	1999	1237	1388	-151	12.2	**20**	2010	1986	223	-245	12.3
10	2000	1437	1463	-26	1.8	**21**	2011	2199	2351	-152	6.9
11	2001	1660	1540	120	7.3	**22**	2012	2249	2453	-204	9.1

It can be seen that of the 22 calculated figures, 16 are within 10% of the actual figure, 21 are within 20% and one is in error by 20.2% of its actual value. Thus, the modified regression model is relatively more accurate in estimating road traffic fatalities in Ghana than the Smeed (1994) equation. Averagely, estimates from the modified regression model exceeded the observed by 7.8% compared to 17% from Smeed's equation.

The paired *t*-test can be used to determine if there are significant differences between the observed fatalities and those from the proposed model. Let X_i and Y_i denote the observed and estimated number of road traffic fatalities in the i^{th} year respectively. We assume that X_i is

$N(\mu_1, \sigma_1^2)$ and Y_i is $N(\mu_2, \sigma_2^2)$, $i = 1, 2, ..., 22$.

From Table 5.8, the observed Levene *F*-ratio (Levene 1960), 0.117, is less than the critical *F*-value, $F_{0.05,1,36} = 0.824$. We therefore conclude that there is no significance difference between σ_1^2 and σ_2^2.

We wish to test H_0: $\mu_1 = \mu_2$ against H_1: $\mu_1 \neq \mu_2$. Let $D_i = Y_i - X_i$ $(i = 1, 2, \ldots, 22)$, and $\mu_D = \mu_1 - \mu_2$. H_0 and H_1 can be expressed in the form H_0: $\mu_D = 0$ and H_0: $\mu_D \neq 0$. The test statistic is

$$T = \frac{\bar{D}}{S_D/\sqrt{n}}$$

S_D 10

and S_D □ — □□ —

where $\bar{D}$ □ 221_i22□□1D_i 2211 □□□i22□1D_{i2} □ 221 □□□i22□□1D_i □□□2 □□□□.

T has the t-distribution with 21 degrees of freedom when H_o is true. Let t denote the computed value of T. We reject H_0 at significance level 0.05 if t □$t_{0.025,21}$ □ 2.080 or t □□$t_{0.025,21}$ □□2.080. From the data, the value of the test statistic is t □ 0.850. Since 0.850 is less than 2.080, we fail to reject H_0 at the 5% level of significance. We conclude that there is no significant difference between the observed and the estimated road traffic fatalities.

5.2.4 Estimates of Alpha and Beta - Monte Carlo Simulation

A full Bayesian approach to modeling requires the specification of probability distributions for both the data and the unknown parameters. The Bayesian method requires that before a sample is taken, some information about □ and □ must be known. It is assumed that this knowledge about □ and □ can be expressed in the form of a probability distribution over the parameter space □.
This prior p.d.f. summarizes what is known about □ prior to taking a random sample. The question is: How can this additional information be obtained and used in estimation?

Ever since the original scheme was proposed by Rev. Thomas Bayes (1763), a crucial problem has been the prior distribution $p(\ )$.□ How does one select a prior distribution $p(\ )$□ , to express the uncertainty about the unknown parameter □? There may be some empirical evidence obtained through earlier experiments which would help us decide on the prior distribution $p(\ )$.□ On the other hand, one can decide on $p(\ )$□ , or at least a class of priors $p(\ )$□ , on a subjective basis or in a normal way. Various rules have been suggested and it appears that there is no neat solution to the problem.

One of the assumptions of the least squares road traffic fatality model is that, given □ and □, the distribution of the dependent variable is normally distributed. The question is: What if these

observations are not from a normally distributed population? The maximum likelihood method, used commonly in the estimation of unknown parameters, are asymptotic; that is, the distribution of the dependent variable is approximately normal when n is large. In this section, we use simulation methods to help us decide on the prior distribution with respect to the sample size under a specified distribution.

The values of N, P and D are simulated from a random variable with specified distributions and parameters given in Table A10, in the Appendix. The simulation is set to initially take a sample of size 15 from each distribution for 1000 iterations. In each case the values of □ and □ are computed and the means values are recorded. The process is repeated using samples of sizes 20, 25, 30, … in that order. In each case, the mean values of □ and □ are recorded for the selected distributions and sample sizes, and the results are given in Table A11, in the Appendix.

The R codes for the implementation of these simulations are given in Listing (A3), in the Appendix.

Table 5.9 shows the actual road traffic fatalities, D, together with the estimated fatalities using the values of □ and □ estimated from the Exponential, Log-Normal, Uniform and Gamma distributions for selected sample sizes from 1991 to 2012. The percentage differences between the calculated and actual values are given in Table A12, in the Appendix.

Table 5.9: Expected road traffic fatalities from simulation of *N*, *P* and *D*

		Exponential				**LogNormal**				**Uniform**				**Gamma**			
Year	***D***	**Sample Size**				**Sample Size**				**Sample Size**				**Sample Size**			
		15	**50**	**100**	**150**	**15**	**50**	**100**	**150**	**15**	**50**	**100**	**150**	**15**	**50**	**100**	**150**
1991	920	351	348	339	350	653	654	658	649	488	492	486	488	30370	30396	30289	30354
1992	914	367	363	354	365	677	678	681	673	507	510	505	506	31288	31315	31206	31273
1993	901	418	414	404	417	734	735	739	730	556	560	554	555	32768	32793	32685	32756
1994	824	510	504	493	508	824	826	829	821	636	640	633	635	34744	34765	34661	34738
1995	1026	617	610	598	615	921	924	928	920	724	728	721	723	36797	36812	36714	36797

1996	1049	776	768	754	775	1052	1055	1059	1052	846	849	842	845	39254	39263	39173	39264
1997	1015	887	877	862	886	1142	1146	1149	1143	929	932	925	928	41125	41131	41045	41142
1998	1419	1020	1008	992	1019	1244	1248	1251	1246	1024	1027	1020	1023	43136	43136	43056	43159
1999	1237	1184	1170	1153	1183	1361	1367	1370	1365	1137	1138	1131	1135	45330	45326	45251	45361
2000	1437	1318	1301	1284	1317	1458	1464	1466	1463	1228	1229	1222	1226	47258	47250	47179	47295
2001	1660	1461	1442	1424	1460	1557	1563	1566	1563	1323	1323	1316	1321	49178	49165	49099	49220
2002	1665	1575	1555	1536	1575	1638	1645	1648	1646	1400	1400	1393	1398	50886	50871	50807	50933
2003	1716	1653	1632	1613	1653	1708	1716	1718	1716	1462	1462	1455	1460	52813	52796	52731	52862
2004	2186	1803	1780	1760	1803	1811	1819	1822	1821	1560	1560	1552	1558	54917	54897	54836	54973
2005	1776	1963	1937	1917	1964	1919	1928	1930	1930	1663	1662	1655	1661	57086	57061	57004	57147
2006	1856	2149	2121	2100	2151	2038	2047	2049	2050	1779	1777	1770	1776	59362	59332	59280	59430
2007	2043	2353	2321	2301	2355	2164	2175	2176	2179	1903	1900	1893	1900	61730	61695	61648	61805
2008	1938	2402	2370	2349	2405	2217	2227	2229	2231	1947	1944	1937	1945	63359	63324	63275	63436
2009	2237	2622	2587	2565	2625	2350	2362	2364	2367	2079	2075	2068	2076	65850	65810	65766	65935
2010	1986	2851	2812	2792	2856	2456	2469	2470	2476	2192	2187	2181	2189	66949	66903	66869	67042
2011	2199	3108	3065	3045	3114	2613	2627	2628	2635	2346	2339	2333	2342	70032	69980	69951	70133
2012	2249	3364	3317	3298	3371	2756	2770	2770	2780	2488	2481	2475	2485	72515	72457	72435	72625

It's obvious from Table 5.9 that, the expected road traffic fatalities with and estimated using the Gamma distribution leads to over-estimation of the number of road traffic fatalities in Ghana, with percentage differences exceeding 2000 for each year over all sample sizes. Thus, given the extremely large deviations between observed and expected values of *D* in Table 5.9, Equation (5.27) has proved to be an imperfect predictive tool of road traffic fatalities in Ghana when *N*, *P* and *D* follow the gamma distribution.

For each distribution, the completely randomized single factor experiment of the effect of sample size on the expected road traffic fatalities is conducted with 4 levels of the factor. The observed *F*-ratio and the corresponding *P*-value for each distribution are 1.0 and 2.7, respectively. The conclusion is that: the estimated road traffic fatalities don't change significantly with increasing sample size under each distribution. Prior to that, Levene's test (Levene 1960) for variance shows that, for each distribution, the variances are homogeneous.

The paired *t*-test is also conducted to determine if there is significant difference between the observed RTFs from the National Road Safety Commission of Ghana and estimated RTFs under each distribution. Table 5.10 shows observed two tail tests and the corresponding *p*-values for testing if there is significance difference between the observed and the estimated fatalities under each distribution.

Table 5.10: T-statistics and P-values of the paired t-test

Exponential		LogNormal		Uniform		Gamma	
*t*statistic	*p*-value	*t*statistic	*p*-value	*t*statistic	*p*-value	*t*statistic	*p*-value
0.224	0.825	1.069	0.297	4.121	0.000	17.976	0.000

Since, from Table 5.10, that *p*-value is greater than 0.05 for exponential and log-normal distributions, we conclude that there are no significance differences between the observed and the estimated fatalities for these distributions. Thus, when *N*, *P* and *D* are from these two distributions, the proposed model of this study can be applied. The *p*-values in Table 5.6, show that the proposed regression model is not a perfect predictive tool for estimating RTFs when the variables are from the uniform and gamma distributions.

The values of *N* and *P* were simulated from the normal distribution while that of *D* was simulated from the exponential distribution for 1000 iterations. The process was repeated for 30 different sample sizes. The values of and for each of these sample sizes are as given in Table A13, in the Appendix. The Montecarlo R codes for the implementation of the iterations and the estimations of and are given in Listing (A4) in the Appendix.

Table 5.11 shows the actual RTFs together with estimated values of RTAs using the values of and in Table A13 for selected sample sizes. A completely randomized single factor experiment conducted, with four selected sample sizes being the factor levels, shows that there is no significant difference between the estimated RTFs with respect to the sample sizes.

Table 5.11: Comparison of actual fatalities and estimated fatalities from the proposed distributions

		Years	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001
		Actual Fatalities	920	914	901	824	1026	1049	1015	1419	1237	1437	1660
Estimated Fatalities	Sample size	15	636	665	760	931	1132	1434	1643	1895	2208	2463	2737
		50	638	667	763	934	1136	1438	1648	1901	2215	2471	2745
		100	640	668	764	936	1138	1441	1651	1905	2219	2476	2750
		150	635	664	759	929	1130	1431	1640	1892	2204	2458	2731
		Years	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
		Actual Fatalities	1665	1716	2186	1776	1856	2043	1938	2237	1986	2199	2249
Estimated Fatalities	Sample size	15	2955	3103	3390	3697	4055	4448	4540	4965	5411	5908	6405
		50	2964	3112	3400	3708	4067	4461	4554	4979	5428	5926	6424
		100	2970	3119	3407	3716	4075	4470	4563	4989	5438	5938	6437
		150	2949	3097	3383	3690	4047	4439	4531	4955	5401	5896	6392

To compare the difference between the actual RTFs to the estimated RTFs for a sample size of 150, a paired *t*-test was conducted. The value of the test statistic obtained was $t = 5.035$ with a p-value of 0.000055. Since 0.000055 is less than 0.05, we conclude that there is a significance difference between the actual RTFs and the estimated RTFs at the 5% level. Thus, if N and P are both from a normally distributed population and D is exponential distribution, then the proposed model is not a perfect predictive tool for estimating RTFs in Ghana.

5.3 A Multilevel Approach

5.3.1 Introduction

Similar to a Bayesian model, where the parameters are considered as random variables, this section seeks to develop a Multilevel Random Coefficient (MRC) model for predicting road traffic fatalities in Ghana. In this model, the number of road traffic fatalities and the regional groups are conceptualized as a hierarchical system of road traffic fatalities and geographical regions of Ghana, with fatalities and regions defined at separate levels of this hierarchical system. One can think of MRC models as ordinary regression models that have additional variance terms for handling non-independence, due to group membership (Hox, 1998). This class of models is also often referred to as mixed-effects models (Snijders & Bosker, 1999).

Instead of estimating a separate regression equation for each of the 10 regions in Ghana, a multilevel regression analysis is applied to estimate the values of the regression coefficients for each region, based on the data for the region. The key to understanding MRC models is to understand how nesting fatalities within geographical regions can produce additional sources of variance (non-independence) in data (Hox, 1998).

The modified Smeed's model for the j^{th} geographical region in Ghana is given by

$$D_{ij}/P_{ij} = \alpha_j (N_{ij}/P_{ij})^{\beta_j} u_{ij},\ j = 1, 2, ..., 10, \quad \text{.....(5.28)}$$

where α_j and β_j are parameters to be estimated. In Equation (5.29), N_{ij} is the number of registered vehicles in the i^{th} year recorded in the j^{th} region, P_{ij} is the population size in the i^{th} year recorded in the j^{th} region and the multiplicative error term, u_{ij}, is such that $\varepsilon_{ij} = \ln u_{ij}$ is $N(0, \sigma^2_j)$. Thus, Equation (5.3) for the j^{th} geographical region in Ghana becomes $y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + \varepsilon_{ij},\ j = 1, 2, ..., 10, \quad \text{.....(5.29)}$ where $\beta_{0j} = \ln \alpha_j$, $\beta_{1j} = \beta_j$, $x_{ij} = \ln(N P_{ij})$, $y_{ij} = \ln(D P_{ij})$ and $\varepsilon_{ij} = \ln u_{ij}$, $i = 1, 2, \ldots, 19$. In this regression equation, β_{0j} is the usual intercept, β_{1j} is the usual regression coefficient (regression slope) and ε_{ij} is the usual residual error term. Since the parameters of

equation (5.29) assumed to vary across the various regions, they are considered to be random variables, referred to as random coefficients. Across all regions, the parameters have a distribution with some mean and variance. Thus, β_{0j} and β_{1j} can be modeled as

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\bar{x}_j + u_{0j}, \qquad \beta_{1j} = \gamma_{10} + u_{1j}. \quad \ldots\ldots(5.30)$$

The terms u_{0j} and u_{1j} are the residual errors at the regional-level, where $\bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}$. From Equations (5.29) and (5.30), we have $y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + \gamma_{01}\bar{x}_j + u_{1j}x_{ij} + u_{0j} + \varepsilon_{ij}$, $j = 1, 2, ..., 10$. ……………..…….(5.31)

The first variance term that distinguishes a MRC model from a regression model, is a term that reflects the degree to which regions differ in their intercepts. A significant variance term,

$$\tau_0 = \text{var}(u_{0j}), \quad \ldots\ldots(5.32)$$

indicates that regions significantly differ in terms of the dependent variable (DV). Significant regional-level variance further suggests that it may be useful to include regional-level variables as predictors. Regional-level variables (or level-2 variables) differ across regions, but are consistent within-regions.

The second variance term that distinguishes a MRC model from typical regression, reflects the degree to which slopes between independent and dependent variables vary across regions is

$$\tau_1 = \text{var}(u_{1j}). \quad \ldots\ldots(5.33)$$

Single-level regression models generally assume that the relationship between the independent variable (IV) and dependent variable (DV) is constant across regions. In contrast, MRC models permit one to test whether the slope varies among regions. If slopes significantly vary, one can attempt to explain the variation as a function of regional differences.

A third variance term is common to both MRC and regression models. This variance term,

$$\sigma^2 = \text{var}(\varepsilon_{ij}), \quad \ldots\ldots(5.34)$$

reflects the degree to which the actual value of y differs from its predicted value within a specific region. One can think of σ^2 as an estimate of within-group variance. One uses fatality-level or level-1 variables to predict within-region variance, σ^2. Level-1 variables differ among members of the same region.

It is assumed that the regional-level residuals u_{0j} and u_{1j} as well as the national-level residuals R_{ij} have mean 0, given the value of the explanatory variable X. Thus, γ_{10} is the average regression coefficient just as γ_{00} is the average intercept. The first part of Equation (5.31), $\gamma_{00} + \gamma_{10} x_{ij} + \gamma_{01} z_j$, is called the fixed part of the model. The second part $u_{1j} x_{ij} + u_{0j} + R_{ij}$, is called the random part.

The term $u_{1j} x_{ij}$ can be regarded as random interaction between group (region) and x. This model implied that the regions are characterized by two random effects: their intercept and their slope. Thus, x has a random coefficient. These two regional effects are usually correlated. The assumption is that, for different regions, the pairs of random effect (u_{0j}, u_{1j}) are independent and identically distributed, that they are independent of the national-level residuals R_{ij}, and that all R_{ij} are independent and identically distributed. The covariance of the regional-level residuals (u_{0j}, u_{1j}) is given by

$$\text{cov}(u_{0j}, u_{1j}) = \tau_{01}. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(5.35)$$

Thus, from Equations (5.31), (5.32), (5.33), (5.34) and (5.35),

$$V_j = \text{var}(Y_{ij} \mid x_{ij}) = \tau_0^2 + 2\tau_{01} x_{ij} + \tau_1^2 x_{ij}^2 + \sigma^2, \quad \ldots\ldots\ldots\ldots\ldots\ldots(5.36)$$

and, for two different year i and i' ($i \neq i'$),

$$\text{cov}(Y_{ij}, Y_{i'j} \mid x_{ij}, x_{i'j}) = \tau_0^2 + \tau_{01}(x_{ij} + x_{i'j}) + \tau_1^2 x_{ij} x_{i'j}. \quad \ldots\ldots\ldots\ldots\ldots(5.37)$$

$\{\square_1, \square_2, \ldots, \square_{10}\}$ is assumed to be a random sample of size 10 taken from a population whose distribution depend on the parameters $\square_{00}$, $\square_{01}$, $\square_{10}$, $\square_0$, $\square_1$, $\square_{01}$ and $\square^2$, where $\square_j = (\square_{0j}, \square_{1j})$, j = 1, 2, …, 10 (Pinhiero & Bates, 2000). Thus, y_{ij} is a value of the random variable Y whose distribution depends on the unknown parameters $\square_{00}$, $\square_{01}$, $\square_{10}$, $\square_0$, $\square_1$, $\square_{01}$ and $\square^2$. Y has the normal distribution with mean $\square 00 \square \square 10^{x}{}_{ij} + \square 01^{\not x}{}_{j}$ and variance V_j. Thus, the probability density function (p.d.f.) of Y given $X \square {}^{x}{}_{ij}$ is

$$f_{Yij}\left(y_{ij} \mid X \square x_{ij}\right) \square \frac{\square}{\sqrt{\square}} 2V1_j \square \exp\left(\square\square\square 2V1_j \square y_{ij} \square \square 00 \square \square 10x_{ij} \square \square 01\not x_j \square 2\square\square\right). \quad \ldots\ldots\ldots(5.38)$$

The likelihood function, for 19 years data, in the j^{th} region is given by

$$l\left(\square \square \square 0\ 1\ 0, \sqrt{\square}, \ldots, V_j \square\right) = \prod_{i\square 1}^{19} \frac{\square}{2V1_j \square} \exp\left(\square\square\square 2V1_j \square y_{ij} \square \square 00 \square \square 10x_{ij} \square \square 01\not x_j \square\right) \square\square$$

$$= \left(\square\square\square \frac{1}{2V1_j i19\square\square 1}\right) \square \exp\left(\square\square\square 2V1_j i19\square\square 1 \square y_{ij} \square\square 00 \square\square 10x_{ij} \square\square 01\not x_j \square 2\square\square\right). \quad \ldots\ldots\ldots(5.39)$$

The maximum likelihood estimates of the seven parameters $\square 00$, $\square 01$, $\square 10$, $\square 0$, $\square 1$, $\square 01$ and $\square^2$ are the values of $\square 00\ 01\ 10, \square, \square, \square\ \square\ \square 0, 1, 01$ and $\square^2$ which maximize the likelihood function.

They are also the values of $\square 00\ 01\ 10, \square, \square, \square\ \square\ \square 0, 1, 01$ and $\square^2$ which maximize

$$L\left(\square \square \square 0, 1, 0, V_j \square\right) = \ln l\left(\square \square \square 0, 1, 0, V_j \square\right)$$

n

$$= \square \underline{19}2 \ln 2\square\square \underline{19}2 \ln V_j \square 2V\underline{1}_j i\square\square 1\square y_{ij} \square\square \square\square 00 \qquad 10x_{ij} \square\square 01\neq j$$

$$\square_2, \ldots\ldots (5.40)$$

where $V_j \square \square \square \square 0\ 2\ 01x_{ij} \square \square 1x_{ij}^2 \square \square^2$. The partial derivatives of L with respect to $\square \square 0\ 1,\ ,$

$\square 0, \square 0, \square 1, \square 01$ and $\square^2$ are given by

$$\square\square\square L00 \square V\underline{1}_j i\square n1\square y_{ij} \square\square \square\square 0010x_{ij} \square\square 01\neq j \square, \qquad \square\square\square$$

$$\square\square\square L01 \square V\underline{1}_j i\square n1\neq j \square y_{ij} \square\square \square\square 0010x_{ij} \square\square 01\neq j \square, \square\square\square$$

$$\square\square\square \underline{L}_{10} \square V\underline{1}_j i\square\square n1x\ y_{ij} \square\ _{ij} \square\square \square\square 0010x_{ij} \square\square 01\neq j \square, \quad \square\square\square$$

$$\square\square\square \underline{L}_0 \square \square\ \underline{19} \square \underline{12} \square n \square y_{ij} \square\square \square\square 00 \qquad 10x_{ij} \square\square 01\neq j \square 2, \square\square\square$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (5.41)$$

$$2V_j \qquad 2V_j$$

$$i\square 1 \qquad \square$$

$$\square\square\square \underline{L}_1 \square \square 192Vx_{ijj2} \square 2xVij2j2\ i\square\square n1\square y_{ij} \square\square \square\square 00 \qquad 10x_{ij} \square\square 01\neq j \square 2, \square\square\square$$

$$\square\ \square\ \overline{19x} \square x \square n \square \qquad \square_2 \square\square$$

$$\square\square\square \underline{L}_{01} \_\_\_\_\_ {}_{ij} \quad {}_{ij} \quad y_{ij} \square\square \square\square 0010x_{ij} \square\square 01\neq j \ ,\square$$

$$V_j \quad V_{j2}\ i\square 1 \square \quad n \quad 2 \square$$

$$\frac{\partial L_2}{\partial \beta_{00}} = \frac{1}{2\sigma_v^2}\sum_{i=1}^{n}\sum_{j=1}^{m}(y_{ij} - \beta_{00} - \beta_{01}x_{ij} - u_j)$$

Equating these partial derivatives to zero (because derivative is zero at the minimum point) and replacing β_{00}, β_{01}, β_{10}, β_0, β_1, β_{01} and σ^2 by $\hat{\beta}_{00}$, $\hat{\beta}_{01}$, $\hat{\beta}_{10}$, $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_{01}$ and $\hat{\sigma}^2$ we obtain the maximum likelihood estimates for the parameters.

Computing the maximum likelihood estimates requires an iterative procedure. At the start, the computer program generates reasonable starting values for the various parameters.

Because multilevel modeling involves predicting variance at different levels, one often begins a multilevel analysis by determining the levels at which significant variation exist. The following sub-sections illustrate the estimation of the regression coefficient using the Linear & Nonlinear Mixed Effects (nlme) package in *R* (Pinhiero & Bates, 2000).

Regional data

Table A14, in the appendix, shows the observable number, D_{ij}, of road traffic fatality in the i^{th} year recorded in the j^{th} region, the number, N_{ij}, of registered vehicles in the i^{th} year recorded in the j^{th} region and P_{ij}, the population size in the i^{th} year recorded in the j^{th} region of Ghana from 1991 to 2009. Table A15 shows the values of $x_{ij} = \ln(N_{ij}/P_{ij})$ and the corresponding values of $y_{ij} = \ln(D_{ij}/P_{ij})$ for the ten regions of Ghana from 1991 to 2009.

5.3.2 The unconditional means model, M_0

In this step, the study examines if there will be significant intercept variation (β_0). In this case, the general assumption is that, there is significant variation in σ^2 (Bryk & Raudenbush, 1992). If β_0 does not differ significantly from 0, there may be little reason to use random coefficient modeling since simpler Ordinary Least Squares (OLS) modeling will suffice. Note that if slopes randomly

vary even if intercepts do not, there may still be reason to estimate random coefficient models (Snijders & Bosker, 1999).

An unconditional means model does not contain any predictors, but includes a random intercept variance term for groups. Thus, the analysis, in this case, is similar to the randomized single factor experiment of analysis of variance. Thus, the study seek to determine if there is significance variation between the mean $\bar{Y}_{.j}$ across the ten regions in Ghana. The model is:

$$Y_{ij} = \beta_{0j} + r_{ij}$$

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad \text{.....(5.42)}$$

In combined form, the model is:

$$y_{ij} = \gamma_{00} + u_{0j} + r_{ij}. \quad \text{.....(5.43)}$$

We wish to determine two estimates of variance;

1. τ_0 associated with $e_{\tau j}$ reflecting the variance in how much each groups' intercept varies from the overall intercept (γ_0),

2. σ^2 associated with r_{ij} reflecting how much each individuals' score differs from the group mean.

The observed variance within region j is given by

$$s^2_j = \frac{1}{18} \sum_{i=1}^{19} (y_{ij} - \bar{y}_j)^2, \quad \text{.....(5.44)}$$

where $\bar{y}_{.j}$ is the mean of the j^{th} region. The observed within-region variance, or pooled within-region variance is

$$MSW = s^2_{within} = \frac{1}{10+10+\cdots+10} (18s^2_{2j} + 10s^2_{2j} + \cdots + s^2_{2j}) \quad \text{.....(5.45)}$$

$j=1\ i=1$ $j=1$ $j=1$

If the model (5.43) holds, then the expectation of S_{within}^2 is equal to σ^2. That is,

$$E\left(S_{within}^2\right) = \sigma^2. \quad \text{.....(5.46)}$$

Thus,

$$\hat{\sigma}^2 = S_{within}^2 \quad \text{.....(5.47)}$$

The observed between-region variance (variance of the group means) is given by

$$S_{between}^2 = \frac{1}{9}\sum_{j=1}^{10}\left(\bar{y}_j - \bar{y}_{..}\right)^2 \quad \text{.....(5.53)}$$

where $\bar{y}_{..}$ is the overall mean. The total observed variance is

$$MST = S_{total}^2 = \frac{1}{189}\sum_{j=1}^{10}\sum_{i=1}^{19}\left(y_{ij} - \bar{y}_{..}\right)^2 \quad \text{.....(5.54)}$$

It can be shown that

$$MST = MSW + MSA, \quad \text{.....(5.55)}$$

where $MSA = 19S_{between}^2$. The expectation of the between-region variance is given by

$$E\left(S_{between}^2\right) = \sigma_0^2 + \frac{\sigma^2}{19} \quad \text{.....(5.55) Thus,}$$

$$\hat{\sigma}_0^2 = S_{between}^2 - \frac{\hat{\sigma}^2}{19} \quad \text{.....(5.56)}$$

$$= \frac{MSA - MSW}{19}.$$

The unconditional means model and all other random coefficient models that the study considers are estimated using the `lme` (linear mixed effects) function in the `nlme` (Linear & Nonlinear Mixed Effects) package of *R* (Pinheiro & Bates, 2000).

The researcher begins the analysis using nlme package in *R*. First, the data set, i.e. the regional distribution of road traffic fatalities in Table A15, is copied on the clipboard in the format as shown in Table A16 (see Appendix) and loaded for analysis as shown in Listing (5.1) in the Appendix.

```
> fatalities<-read.table(file="clipboard",sep="\t",header=T)
```
….Listing (5.1) In the model, the fixed formula is *y*~1 as applied in Listing (A4). This states that the only predictor of *y* is an intercept term.

```
> Null.Model<-lme(y~1,random=~1|Regions,data=fatalities,
   +control=list(opt="optim"))
```
……...Listing (5.2)

The `VarCorr` function provides estimates of variance for an `lme` object as shown in Listing (A5)

Intraclass Correlation Coefficient (ICC)

As with the completely randomized single-factor experiments, it is useful to determine how much of the total variance is between-groups. This can be accomplished by calculating the Intraclass Correlation Coefficient (ICC). Using this model, we can estimate the ICC value ρ by using the equation (Hox, 2010 and Snijders & Bosker, 1999)

$$\hat{\rho} = \frac{\hat{\tau}_0^2}{\hat{\tau}_0^2 + \hat{\sigma}^2}. \quad \ldots\ldots(5.57)$$

where $\hat{\tau}_0$ and $\hat{\sigma}^2$ are point estimates of τ and σ^2, respectively. The standard error of this estimator, when $n = 19$ and $a = 10$, is given by (Hox, 2010)

$$SE.\hat{\rho} = \sqrt{(1-\rho)^2\{1+(n-1)\rho\}^2 \frac{2}{an(n-1)(a-1)}}. \quad \ldots\ldots(5.58)$$

The purpose of the unconditional means model is to estimate the between-group and withingroup variance in the form of τ_0 and σ^2 respectively. Thus, from Listing (A5), the estimate of

τ_0 and σ^2 are $\hat{\tau}_0$ 0.1891104 and $\hat{\sigma}^2$ 0.1389485. Using these values in Equation (5.57), we obtain,

$$\hat{\rho} = \frac{0.1891104}{} = 0.5764526.$$

0.1891104 0.1389485□

The estimate of the ICC value can also be computed from an ANOVA model, given in Listing (5.2).

```
> tmod<-aov(y~as.factor(Regions),data=fatalities.1
```
.....Listing (5.2)

The ICC has values that lie in the range [0, 1]. It describes how strongly observations between regions resemble each other. If there is full agreement in every region, then □ □^{2} 0 and the ICC = 1. If there is no agreement, then □ □0 0, and the ICC = 0. The closer the ICC value to 1, the stronger the resemblance of observations between regions.

Estimating group-mean reliability

When exploring the properties of the outcome variable, it can also be of interest to examine the reliability of the group mean. The reliability of group means often affects one's ability to detect emergent phenomena. In other words, a prerequisite for detecting emergent relationships at the aggregate level is to have reliable group means (Bliese 1998). By convention, estimates around 0.70 are considered reliable. Group mean reliability estimates are a function of the ICC and group size (see Bliese, 2000; Bryk & Raudenbush, 1992). ICC(2) is among regions variance (*MSA*) minus within regions variance (MSW) over among regions variance (*MSA*).

$$ICC(2) = \frac{MSA}{MSA} \square \frac{MSW}{} \quad \text{.....(5.58)}$$

$$= \frac{3.73205 \;\; 0.138953.73205\square}{} \square \; 0.96277.$$

The `GmeanRel` function from the multilevel package in R calculates the ICC, the group size, and the group mean reliability for each group. When we apply the `GmeanRel` function to our `Null.Model`, as shown in Listing (A6), based on the 10 regions in the `fatalities` data set, we are interested in two things. First, we are interested in the average reliability of the 10 regions. Second, we are interested in determining whether or not there are specific regions that have particularly low reliability. Notice that the overall group-mean reliability is acceptable at 0.96277.

Determining whether τ_0 is significant.

If it is assumed that the within-region deviations ε_{ij} are normally distributed, then we can the hypothesis that ICC is 0, which is the same as the null hypothesis that there are no regional differences, or the true between-region variance is 0. The test statistic is

$$F = MSW/MSA,$$

which has the F-distribution with 9 and 180 degrees of freedom when the null hypothesis is true. We reject the null hypothesis at 0.05 level of significance if the observed F value is greater than $F_{0.05,9,180} = 1.9322$. From Table 5.14, since the observed value of the test statistic, 26.8592, is greater than 1.9322, we reject the null hypothesis and conclude that the intercept variance, τ_0, is significantly different from zero.

In summary, we would conclude that there is significant intercept variation in terms of y scores across the 10 regions. We also estimate that about 58% of the variation in y score is a function of the region to which it is observed. Thus, a model that allows for random variation in y among regions is better than a model that does not allow for this random variation.

5.3.3 Random intercept model: M_1

At this point of the analysis, there are two sources of variation that we can attempt to explain in subsequent modeling – within-region variation (σ^2) and between-region intercept variation (τ_0). In this section, we begin to build a model that predicts these two sources of variation. A first step towards modeling between-group variability is to let the intercept vary between regions. This reflects that some groups tend to have, on average, higher responses Y and others tend to have lower responses. The form of the model is:

$$y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + \varepsilon_{ij},$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01} x_j + u_{0j} \quad \text{.....(5.59)}$$

$\beta_{1j} = \beta_{10}$’ $\beta_{\square}$

The first row of Equation (5.59) states that the *y* score is a function of the intercept for a region plus a component that reflects the linear effect of the observed *x* value plus some random error. The second line states that each region’s intercept is a function of some common intercept (β_{00}) plus a component that reflects the linear effect of regional average of *x* values plus some random between-regions error. The third line states that the slope between *x* and *y* is fixed – it is not allowed to randomly vary across groups. The three rows are combine into a single equation

$$Y_{ij} = \beta_{00} \beta_{\square}\beta_{10}x_{ij} \beta_{\square}\beta_{01}\bar{x}_j \beta_{\square}u_{0j} \beta_{\square}\beta_{ij}, \quad j \beta_{\square}1, 2, ...,10. \quad \text{.....(5.60)}$$

Essential assumptions are that all residuals, u_{0j} and β_{ij}, are mutually independent and have zero means given the values x_{ij} of the explanatory variable. For the u_{0j}, just as for the β_{ij}, it is assumed that they are drawn from normally distributed populations. The population variance of the national-level residuals β_{ij} is assumed to be constant across the regions, and denoted by $\beta_{\square}^2$; the population variance of the regional-level residuals u_{0j} is denoted by β_0.

Using the data in Table A16 (see Appendix), model M_1 is specified in the *R* package `lme` as shown in Listing (5.10).

```
Model.1<-lme(y~x+G.x,random=~1|Regions,data
=fatalities,control=list(opt="optim"))
```
}.....Listing (5.3)

Listing (A7) gives the summary of the results of Listing (5.3).

Table 5.12 presents the parameter estimate and standard errors for both models (M_0 and M_1). In this table, the intercept-only model estimated the intercept as β9.688842, which is simply the average of the *y* values of all regions and fatalities. The variance of the fatality-level residual error, symbolized by $\beta_{\square}^2$, is estimated as 0.1389485. The variance of the regional-level residual errors, symbolized by β_0 is estimated as 0.1891104. All parameter estimates are much larger than the corresponding standard errors, and calculation of the Z-test shows that they are all significant at *p*

< 0.05. The deviance reported in Table 5.12 is a measure of model misfit; when we add explanatory variable to the model, the deviance is expected to go down.

Table 5.12: Intercept-only model and model with explanatory variables

Model	M_0: intercept only		M_1: with predictor	
Fixed effect	Coefficient	Standard Error	Coefficient	Standard Error
γ_{00} Intercept	−9.688842	0.1401	−10.075599	0.7426
γ_{10} coffiecient of x_{ij}			0.459058	0.0374
γ_{01} coefficient of $\bar{x}_j$			−0.544840	0.1658
Random part	Parameters	Standard Error	Parameter	Standard Error
τ_0 var(u_{0j})	0.1891104	0.2085	0.20938573	0.1447
σ^2 var(ε_{ij})	0.1389485	0.0855	0.07586128	0.0632
Deviance	198.201		94.554	

In the second model, where the explanatory variable was included, the regression coefficients for all three variables are significantly different. Notice that the *x*-scores are significantly positively related to the *y*-scores. Furthermore after controlling the fatality-level relationship, average *x*scores are negatively related to the average *y*-score in a region. The interpretation of this model indicates that the slope at the regional-level significantly differs from the slope at the fatalitylevel. A unit increase in $\bar{x}$ is associated with a –0.085 (–0.545 + 0.460) decrease in average *y*score. The coefficient of –0.545 reflects the degree of difference between the two slopes.

The within-region and between-region regression coefficients would be equal if, in Equation (5.38), the coefficient of $\bar{x}$ would be 0, i.e. $\gamma_{10} = 0$. This null hypothesis can be tested using the test statistics

$$T = \frac{\text{estimate}}{\text{standard error}},$$

which has the *t*-distribution with 9 degrees of freedom. The value of *T* based on the given data is $t = -0.544840 / 0.1658 = -3.286$, which is significant at the 0.05 level.

The within-region deviation about this regression equation, ε_{ij}, have a variance of 0.0759 (standard deviation 0.2755). Within each region, the effect (regression coefficient) of x_{ij} is

0.459, so the regression lines are parallel. Regions differ in two ways; they may have different mean x-values, which affects the expected results y_{ij} through the term $0.545\bar{x}_j$; this is an explained difference between the regions; and they have randomly different values for u_{0j}, which is an unexplained difference. These two ingredients contribute to the region-dependent intercept, given by $\beta_0 = 10.076 + u_{0j} + 0.545\bar{x}_j$.

The estimate of regional-level residual $\hat{u}_{0j}$ and the corresponding values of β_0 and β_1 for each region are given together with the values of $\bar{x}_j$ are given in Table 5.13.

Table 5.13: Estimate of the values of u_{0j}, β_{0j}, β_j and β_j for each region

Regions	Greater Accra	Ashanti	Western	Eastern	Central	Volta	Northern	Upper East	Upper West	Brong Ahafo
$\hat{u}_{0j}$	0.43873	0.24502	0.00278	0.58195	0.36494	-0.14103	-0.59011	-0.42436	-0.59757	0.11865
$\bar{x}_j$	-2.35474	-3.92579	-4.85053	-5.24053	-5.41474	-5.53579	-4.59368	-4.24947	-3.91211	-4.99790
$\hat{\beta}_{0j}$	-8.35385	-7.69156	-7.42995	-6.63827	-6.76036	-7.20038	-8.16278	-8.18457	-8.54161	-7.23378
$\hat{\beta}_j$	0.45906	0.45906	0.45906	0.45906	0.45906	0.45906	0.45906	0.45906	0.45906	0.45906
$\hat{\beta}_j$	0.0002355	0.0004567	0.0005932	0.0013093	0.0011588	0.0007463	0.0002851	0.0002789	0.0001952	0.0007218

The estimated values of β_0 and β_1 can be used to estimate the number of road traffic fatalities in each region. For instance, in Greater Accra region, where $\bar{x} - \beta_0 = 2.35474$, the estimated values for β_{0j} and β_{1j} are $\beta_0 = \beta_{01} = 10.076 - 0.43873 - 0.545\bar{x} - \beta_0 = 8.35385$ and $\beta_1 = \beta_{11} = 0.45906$, respectively. Therefore, the estimate for β_1 is

$$\beta_1 = \hat{\beta}_1 = e^{-8.35385} = 0.0002355. \quad \text{.....(5.61)}$$

Equation (4.12), for Greater Accra region, therefore becomes

$$D_{i1}/P_{i1} = 0.0002355\left[N_{i1}/P_{i1}\right]^{0.45906}, \quad \text{.....(5.62)}$$

where D_{i1} is the number of road traffic fatalities in the i^{th} year, N_{i1} is the number of registered vehicles in the i^{th} year and P_{i1} is the estimated population size in the i^{th} year, for Greater Accra region.

5.3.4 Random slope model M_2

In the random intercept model of M_1, the regions differ with respect to the average value of the dependent variable: the only random group is the random intercept. But the relation between explanatory and dependent variables can differ between regions in more ways. The study, therefore, continue the analysis by trying to explain the third source of variation, namely, variation in the slope, β_1. The model that we test is:

$$y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + e_{ij}, \quad e_{ij} \sim N(0, \sigma^2)$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01} z_j + u_{0j} \quad \text{.....(5.63)}$$

$$\beta_{1j} = \gamma_{10} + u_{1j} \quad u_{1j} \sim N(0, \tau_1)$$

The intercepts β_{0j} as well as the regression coefficients, or slopes, β_{1j} are region-dependent.

When we combine the three rows into a single equation in the form $y_{ij} = \gamma_{00} + \gamma_{10} x_{ij} + \gamma_{01} z_j + u_{1j} x_{ij} + u_{0j} + e_{ij}$, $j = 1, 2, ..., 10$. .....(5.64) The slope β_{1j} is normally distributed with mean γ_{10} and variance τ_1. The variance term associated with u_{1j} is τ_1. Since 95% of the probability of a normal distribution is within two standard deviations from the mean, it follows that approximately 95% of the regions have slopes between $\gamma_{10} - 2\sqrt{\tau_1}$ and $\gamma_{10} + 2\sqrt{\tau_1}$. Fig. 5.3 presents 10 regression lines for the 10 regions of Ghana using the data in Table A15. The figure demonstrates regression lines that characterize, according to this model, the population of geographical regions in Ghana.

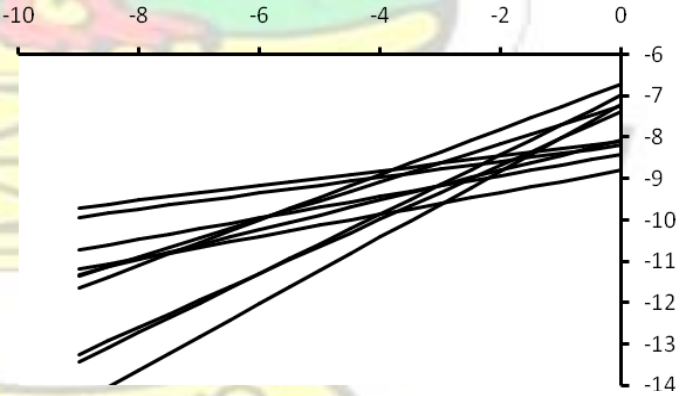


Figure 5.3: Ten random regression lines from Table A15

In *R* this model is designated as shown in Listing (5.4).

```
>Model.2<-lme(y~x+G.x,random=~x|Regions,
    data=fatalities,control=list(opt="optim"))
```
.....Listing (5.4)

The summary of the results Listing (5.4) are given in Listing (A8), in the appendix. The `VarCorr` function, as applied in Listing (A9), provides estimates of variance for an `lme` object.

Table 5.14 presents the parameter estimate and standard errors for the models M_0, M_1 and M_2. The within-region regression in model M_2 is 0.4459 and between-region regression coefficient is -0.3384 + 0.4459 = 0.1075. All the standard errors of the estimated parameters in model M_2 are smaller than the corresponding values of model M_1. Moreover, the deviance, which measures the model misfit, is much lower in M_2 as compared to that of M_1. Thus, the estimated parameters based on model M_2 is preferred.

Table 5.14: Comparison of models M_0, M_1 and M_2

Model	M_0: intercept only		M_1: with predictor		M_2: with predictor	
Fixed effect	Coefficient	Standard Error	Coefficient	Standard Error	Coefficient	Standard Error
□ □00 Intercept	-9.6888	0.1401	-10.0756	0.7426	-9.2341	0.2065
□10 □ coffiecient of x_{ij}			0.4591	0.0374	0.4459	0.0707
□01 □ coefficient of $\bar{x}_j$			-0.5448	0.1658	-0.3384	0.0516
Random part	Parameter	Standard Error	Parameter	Standard Error	Parameter	Standard Error
□ □0 var(u_{0j})	0.1891	0.2085	0.2094	0.1447	0.1545	0.1243
□ □1 var(u_{1j})					0.0382	0.0618
□01 □ cov(u_{0j}, u_{1j})					0.0766	
$\square^2$ □ var(□$_{ij}$)	0.1389	0.0855	0.0759	0.0632	0.0630	0.0576
Deviance	198.201		94.554		64.749	

In the null model M_0, the variance estimate from the within-region residual, $□^2$, is 0.1389. and the variance estimate for the intercept, $□_0$, is 0.1891. The variance estimates from the model M_2, with one predictor, are □ $\hat{□}^2$ 0.0630 and □ $\hat{□}_0$ 0.1545. That is, the variance of the within-region residuals decreased from 0.1389 to 0.0630 and the variance of the between-region intercepts decreased from 0.1891 to 0.1545.

$$\text{Variance explained} \; □ \; 1 \; □ \; \frac{\text{varince with predictor}}{\text{variace without predictor}} \quad \text{...(5.65)}$$

The y-values explained is 1 – (0.0630/0.1389) or 0.55 (55%) of the within-region variance in 2, and regional-mean values x explained is 1 – (0.1545/0.1891) or 0.18 (18%) of the between-

□

region intercept variance $□_0$.

Should the value of 0.0382 for the random slope variance be considered to be high? The slope standard deviation is $\sqrt{0.0382}$ □ 0.195, and the average slope is □ $□_0$ 0.4459. The values of ‘average slope □ two standard deviations’ range from 0.0559 to 0.8359. This implies that the effect of x is clearly positive in all regions. Table 5.15 gives the slope of the least square regression line for each of the 10 regions of Ghana based on the data in Table A15.

Table 5.15: Slope of the least square regression line for each region in Ghana

Greater Accra	Ashanti	Western	Eastern	Central	Volta	Northern	Upper East	Upper West	Brong-Ahafo
0.267	0.360	0.258	0.179	0.193	0.549	0.717	0.654	0.804	0.458

It can be seen from Table 5.15 that all the 10 regions have slopes between 0.0559 and 0.8359. Thus, the normality assumption of the slope is validated. The correlation between random slope and random intercept is □ □ $\frac{0.0766}{\sqrt{0.1545 \;\; 0.0382}}$ □ 0.997. □□

The standard deviation of the x-values is about 1.05, and the mean is –4.5. Hence fatalities with x values among the bottom fewer percent or the top few percent have x values of about – 6.6 and – 2.4 respectively. Substituting these values in the contribution of the random effect give $u_{0j} \square 6.6u_{1j}$ and $u_{0j} \square 2.4u_{1j}$. It follows from Equations (5.42) and (5.43) that when $x \square \square 6.6$,

$\square 2.4$, $\text{var}\square Y x_{ij} ij \square\square 6.6\square \square 0.1545\square \square 2\ 0.0766\ (\ 6.6)\square \square \square 0.0382\ (\ 6.6)\square \square \square^{2} \square 0.0630 \square$

0.8704, $\text{cov}\square Y Y x_{ij}, {}_{ij}\square ij \square\square 6.6, x_{ij}\square \square\square 2.4\square \square 0.1545\square 0.0766(\ 6.6\square \square 2.4)\square 0.0382\square \square \square 6.6$

$2.4\ 0.0702$ $\text{var}\square Y x_{ij\ i\ j\square\square} \square\square^{2.4}\square \square 0.1545 \square \square 2\ 0.0766\ (\ 2.4)\ \square \square^{2} \square 0.0382\ (\ 2.4\square \square\ ) \square$

$0.0630 \square 0.0699$. and therefore

$\square \square Y_{ij}, Y_{i\ j}\square \mid x_{ij} \square\square 6.6, x_{i\ j}\square \square\square 2.4\square \square \sqrt{} \square 0.8704\ \ 0.06990.0702\square \square 0.2846.$

Thus, the highest value of x and the least value of x in the same region are positively correlated over the population of regions. The positive correlation corresponds to the result that the value of x for which the variance given by (5.42) is minimal, is outside the range from -6.6 to -2.4. For the estimates in Table 5.17, this variance is

$$\text{var}\square Y x_{ijij} \square x\square \square 0.1545 \square 0.1532\ x \square 0.0382x^{2} \square \square^{2}.$$

Equating the derivative with respect to x to 0, shows that the variance is minimal when $x \square\square 0.1532$ $0.0382 \square\square 4.01$, which is within the range – 6.6 to –2.4.

In Table 5.14, the model M_2 represents within each region, denoted j, a linear regression equation

$Y_{ij} = \square 9.2341 \square 0.4459x_{ij} \square 0.3384\cancel{x_j} + u_{0j}\ ijx \square\ u_{1j} \square \square_{ij}$, ……..…(5.66) where u_{0j} and u_{1j} are region-dependent deviations each with mean 0 and variances 0.1545 and 0.0630 respectively. The application of the R code `coef(Model.2)` gives the intercept and the coefficients of x and $\cancel{x}$ as shown in Table 5.16.

Table 5.16: Intercept and coefficients of x and $\varkappa$

No.	Regions	Intercept	x	$\varkappa$
1	Greater Accra	-9.506844	0.3083572	-0.3384525
2	Ashanti	-9.402255	0.3614688	-0.3384525
3	Western	-9.319224	0.4053849	-0.3384525
4	Eastern	-9.704008	0.2109577	-0.3384525
5	Central	-9.575697	0.2758323	-0.3384525
6	Volta	-9.270846	0.4259363	-0.3384525
7	Northern	-8.806641	0.6594775	-0.3384525
8	Upper East	-8.839118	0.6439825	-0.3384525
9	Upper West	-8.530726	0.7993004	-0.3384525
10	Brong Ahafo	-9.385768	0.3686119	-0.3384525

The estimate of regional-level residuals $u\hat{}0j$ and $u\hat{}1j$ and the corresponding values of β_{0j} and β_{1j} for each region are given in Table 5.17.

Table 5.17: Estimate of regional-level residuals and the values of β and β

Regions	$u\hat{}0j$	$u\hat{}1j$	$\beta\hat{}_{0j}$	$\beta\hat{}_j$	$\beta\hat{}_j + e_{0j}$
Greater Accra	-0.273	-0.138	-8.709877	0.3083572	0.0001649
Ashanti	-0.168	-0.084	-8.073562	0.3614688	0.0003117
Western	-0.085	-0.041	-7.677551	0.4053849	0.0004631
Eastern	-0.470	-0.235	-7.930339	0.2109577	0.0003597
Central	-0.342	-0.170	-7.743066	0.2758323	0.0004337
Volta	-0.037	-0.020	-7.397244	0.4259363	0.0006129
Northern	0.427	0.214	-7.251897	0.6594775	0.0007088
Upper East	0.395	0.198	-7.400873	0.6439825	0.0006107
Upper West	0.703	0.353	-7.206664	0.7993004	0.0007416
Brong Ahafo	-0.152	-0.077	-7.694218	0.3686119	0.0004555

Based on Table 5.17, the estimate of the number of road traffic fatalities, $D\hat{}_{ij}$, of the j^{th} geographical region of Ghana in the i^{th} year, can be obtained from the formula

$$\hat{D}_{ij}/P_{ij} = \hat{\alpha}_j \left(N_{ij}/P_{ij}\right)^{\hat{\beta}_j}, \quad j = 1, 2, ..., 10 \quad \text{.....(5.67)}$$

N_{ij} is the number of registered vehicles in the i^{th} year recorded in the j^{th} region while P_{ij} represents the population size in the i^{th} year recorded in the j^{th} region.

For instance, in Greater Accra region, where $\beta = -2.35474$, the estimated values for α_{0j} and α_{1j} are

$$\hat{\alpha}_{01} = 9.2341 - 0.33845(-2.35474) - 0.273(8.710),$$

$$\hat{\alpha}_{11} = 0.4459 - 0.138 = 0.308.$$

Therefore, the estimate for α_j is

$$\hat{\alpha}_1 = e^{-8.710} = 0.0001649. \quad \text{.....(5.68)}$$

Equation (5.67), for Greater Accra region, therefore becomes

$$D_{i1}/P_{i1} = 0.000164948\left(N_{i1}/P_{i1}\right)^{0.3083572}. \quad \text{.....(5.69)}$$

The actual road traffic fatalities for Greater Accra, D_{i1}, from 1991 to 2012, together with the corresponding values of $\hat{D}_{i1}$ calculated from Equation (5.69), are given in Table 5.18. The percentage differences between the calculated and actual values are also given. It can be seen that, from 2001 to 2012, in the Greater Accra region, all the 12 calculated figures are within 10% of the actual figure. Out of the 22 calculated figures, from 1991 to 2012, 15 are within 10% of the actual figure and 19 are within 20% of the actual value. The paired t-test statistic for comparing the actual RTFs and the estimated RTFs is 0.484 with a p-value of 0.344. Since 0.344 is greater than 0.05, we conclude that there is no significant difference between the actual RTFs and the estimated RTFs for Greater Accra, at the 5% level of significance. The Levene's test shows that the variances are homogeneous.

Table 5.18: Comparison of actual fatalities and fatalities estimated from Equation (5.68) for Greater Accra region

i	Year	*Di*1	*D*ˆ*i*1	Error	Error %	*i*	Year	*Di*1	*D*ˆ*i*1	Error	Error %
1	1991	126	120.1	5.9	4.7	**12**	2002	239	262.5	-23.5	9.8
2	1992	164	125.4	38.6	23.5	**13**	2003	240	262.6	-22.6	9.4
3	1993	115	134.7	-19.7	17.1	**14**	2004	299	290.1	8.9	3.0
4	1994	155	147.7	7.3	4.7	**15**	2005	306	304.3	1.7	0.6
5	1995	190	161.6	28.4	14.9	**16**	2006	325	319.8	5.2	1.6
6	1996	191	179.1	11.9	6.2	**17**	2007	370	336.0	34.0	9.2
7	1997	174	192.4	-18.4	10.6	**18**	2008	385	350.6	34.4	8.9
8	1998	258	207.1	50.9	19.7	**19**	2009	420	378.5	41.5	9.9
9	1999	172	223.7	-51.7	30.1	**20**	2010	424	384.8	39.2	9.3
10	2000	196	241.6	-45.6	23.2	**21**	2011	425	403.8	21.2	5.0
11	2001	239	249.1	-10.1	4.2	**22**	2012	435	442.4	-7.4	1.7

5.4 Road Traffic Fatality risk Indicators

In this Section, the study seeks to use Ghana data to establish the assertion that the parameter estimates of the modified Smeed's model can be used as risk indicators of road traffic fatalities across the ten geographical regions in Ghana.

Road traffic fatality indices such as road traffic fatalities (RTFs) per 100 accidents and road traffic fatalities (RTFs) per 100 casualties are used by National Road Safety Commission (NRSC) of Ghana and World Health Organization (WHO) as risk indicators to characterized and compare the extent and risk of traffic fatalities across geographical regions. These indices became very useful measures of risk to compare risk of dying as result of road traffic accidents across the 10 geographical regions in Ghana.

In Table 5.18, the average fatality indices from 1991 to 2009, with respect to *RTFs per 100* Accident and *RTFs per 100* Casualties, across the 10 geographical regions in Ghana are given in

the 3rd and 4th columns. The first two columns give the parameter estimates of the modified Smeed's model across the 10 geographical regions. The study aims at determining if there is positive correlation between the parameter estimates of this studies and the fatality indices based on NRSC and WHO definition of risk.

Table 5.18: Parameter estimates and Fatality indices

Regions	□□ˆ 10^5	□□ˆ 10^2	*RTF per 100* Accident	*RTF per 100* Casualties
Greater Accra	16.5	30.836	5.7	7.7
Ashanti	31.2	36.147	17.8	12.2
Western	46.3	40.538	16.9	10.7
Eastern	36.0	21.096	19.9	9.7
Central	43.4	27.583	21.8	11.4
Volta	61.3	42.594	23.6	11.2
Northern	70.9	65.948	40.9	18.1
Upper East	61.1	64.398	27.3	17.0
Upper West	74.2	79.930	28.3	14.6
Brong-Ahafo	45.6	36.861	28.6	14.5

Table 5.19 shows the correlation coefficients between the parameter estimates of the modified Smeed's model and the fatality indices. The corresponding *p*-values for the test are in parenthesis. Since for each pair the *p*-value is less than 0.05, we conclude that there is strong correlation between the parameter estimates of this studies and the fatality indices based on

NRSC definition of risk. Thus, the parameter estimates □ˆ and □ˆ in Equation (5.67) can be used as risk indicators of RTFs in Ghana.

Table 5.19: Correlations coefficients

	□ˆ	□ˆ	*RTF* per *100* Accident	*RTF* per *100* Casualties
□ˆ	1			

□ˆ	0.8312 (0.003)	1		
RTF per *100* Accident	0.8424 (0.002)	0.6341 (0.049)	1	
RTF per *100* Casualties	0.7708 (0.009)	0.7610 (0.010)	0.9011 (0.000)	1

CHAPTER SIX SUMMARY, CONCLUSION AND RECOMMENDATIONS

6.1 Summary

Smeed (1949) proposed a model for estimating road traffic fatalities (RTFs). Smeed's model gave a fairly good fit to the data from 20 countries, including European countries, USA, Canada, Australia and New Zealand. The results obtained by Smeed in his study was consistent with other reported studies by Bener and Ofosu (1991), Jacobs and Bardsley (1977), Fouracre and Jecobs (1977), Ghee et al. (1997). Ponnaluri (2012) showed that Smeed's model is parsimonious in parameter usage. These related studies point to the fact that Smeed's model appears to be observation-driven, evidence-based, and logically valid in measuring the ***per vehicle fatality rate*.**

This study put forward the derivation of a modified Smeed's model and also determine how accurate the proposed modified model of this study is. The question addressed here was: how does the modified Smeed's model compare to that of Smeed (1949) in their performance?

It was shown that the predominant factors affecting road traffic fatalities (RTFs) are not the same as that of road traffic accidents (RTAs). Exposures to risk of RTFs (such as human error, vehicular speed, vehicular density, weather conditions, nature of the roads and total length of roads) are predominant factors influencing road traffic accidents within a geographical region. However, the rate of RTFs is determined by vulnerability to risk (Such as accessibility, timeliness and appropriateness of emergency medical care as well as adequacy and enforcement of use of safety mechanisms in vehicles). Exposure to risk RTFs and vulnerability to RTFs are not necessarily correlated. The factors affecting RTAs correspond to exposure X while factors affecting RTFs correspond to vulnerability given the same exposure X. In Smeed's model exposure is measured by the variable X whereas vulnerability for a given X is captured by the parameters

It was also demonstrated that the parameters of Smeed's model vary from one geographical region to another and hence could be use to assess variability of risk of RTFs across geographical regions. Thus, the study proposed more robust Bayesian and Multilevel estimation procedures that allow the study to estimate the variance of the parameters across geographical regions and hence enables us compare the risk of RTFs across the geographical regions.

The study, therefore, focused on developing statistical methodology, based on Smeed's model, for assessing the risk of RTFs across sub-populations of a given geographical zone. To achieve this general objective, the study first developed a modified Smeed's model and uses it to develop a Bayesian and multilevel methods to compare the risk of RTFs across sub-populations of a given geographical zone.

Some preliminary investigations on some characteristics of road traffic accidents were also performed and particularly road traffic fatalities in Ghana which are of general interest and have a certain bearing on the main results of this study.

Finally, the study used Ghana data to validate the developed Bayesian and Multilevel methods and also used the parameter estimates to assess the risk of road traffic fatalities across the ten geographical regions in Ghana.

6.2 Conclusion

A modified Smeed's model,

$$\frac{D}{P} = \sqrt{N P u},$$

has been developed. The multiplicative error term u in the modified Smeed's model of this study was found to be less than that of Smeed's, making the modified Smeed's model preferred. Using data from Ghana, it was confirmed that the modified Smeed's model for this studies, is relatively more accurate in estimating RTFs in Ghana than the Smeed's equation.

Based on this modified Smeed's model, Bayesian and multilevel methods were developed to assess RTF risk across sub populations of a given geographical zone. These methods consider the parameters of the Smeed's model to be random variables and therefore make it possible to compute

variances across space provided there is significant intercept variation of the regression equation across such regions.

Using data from Ghana, the robustness of the Bayesian estimates was indicated at low sample sizes with respect to the Normal, Laplace and Cauchy prior distributions. Thus, the Bayesian and Multilevel methods performed at least as well as the traditional method of estimating parameters and beyond this were able to assess risk differences through variability of these parameters in space.

The study has shown that population and number of registered vehicles are predominant factors affecting road traffic fatalities. The effect of other additional factors on road traffic fatality such as human (the driver, passenger and pedestrian), vehicle (its condition and maintenance), environmental/weather and nature of the road cannot be ruled out.

Using data from Ghana, the result seems to suggest that road safety efforts by the National Road Safety Commission of Ghana, are not yielding the desired results of reducing the number of road traffic deaths. This is due to lack of accessible, timely and appropriate emergency medical care. A large proportion of road traffic accident trauma patients in Ghana do not have access to formal Emergency Medical Services. Another reason is inadequate safety mechanisms in vehicles as well as improper enforcement of the use of these mechanisms. The age of vehicles and availability of modern safety mechanisms in vehicles plying the roads of Ghana have significant effect on consequences of road traffic accidents.

6.3 Recommendations

(a) Modern Safety Mechanisms

Since the availability of modern safety mechanisms in vehicles have significant effect road traffic fatalities, greater attention must be paid on improving road safety mechanisms in cars such as anti-lock braking systems (ABS), air bags, better design of cars and increased wearing of seatbelts. The enforcement of the use of road safety mechanisms in cars could substantial benefit in reducing injuries and fatalities with respect to road traffic accidents.

In Ghana, the preventive measures of the National Road Safety Commission of Ghana are predominantly directed towards regulating the behaviour road users. However, human behaviour,

in a complex traffic environment, is uncertain and therefore effort to regulate human behaviour in an indiscipline traffic environment usually achieves little results. Vehicle engineering measures must therefore be integrated to have maximum effect in reducing the high spate of road traffic fatalities in Ghana. Enforcement of seat-belt wearing by bus and car occupants, standard crash helmet wearing by motor-cycle riders and passengers and ensuring the crashworthiness of vehicles, must be strictly pursued and sustained. Crashworthiness is the ability of a vehicle to protect its occupants during an impact. It is a measure of how well a vehicle performs during a collision.

One country that has been successful in enforcing crashworthiness regulation of vehicles plying its roads is [6]Northern Ireland (NI). The Department of the Environment (DOE) reports that the number of people killed on Northern Ireland's roads in 2010 was the lowest since records began in 1931. The figures reported show that the number of people killed in accidents in NI fell from 115 in 2009 to 55 in 2010, representing a 50% fall in fatalities and a 20% reduction in serious injuries. Of the 55 people killed in 2010, 10 were pedestrians, 10 on motorcycles and the rest in other vehicles. This success, among other things, was attributed to the Crashworthiness of vehicles plying the road of NI.

According to the world report on road traffic injury prevention (2004), for car occupants, wearing seat-belts in well-designed cars can provide protection to a maximum of 70 km/h in frontal impacts and 50 km/h in side impacts. Higher speeds could be tolerated if the interface between the road infrastructure and vehicle were to be well-designed and crash-protective – for example, by the provision of crash cushions on sharp ends of roadside barriers. However, most infrastructure and speed limits in existence today allow much higher speeds without the presence of crash-protective interfaces between vehicle and roadside objects, and without significant use of seat-belts. This is particularly the case in many low-income and middle-income countries.

(b) Emergency Medical Services

The absence of formal emergency medical care in most developing countries, necessitates that innovative and low cost solutions be devised to meet the growing need for pre-hospital trauma

[6] Northern Ireland's road safety strategy to 2020. Annual Report on Northern Ireland 's Road Safety Strategy to 2020 (the Strategy) and covers the period 1 January 2011 to 31 December 2011

care. The relatively high road traffic mortality rates in most developing countries implies relatively poor health facilities in these regions. It is obvious that improving the health facilities in these regions to make it comparable to that of developed countries would drastically reduce the road traffic fatality rate there.

(d) Training of commercial drivers

In most developing countries, majority of injured persons are transported to the hospital by some type of commercial vehicle. It has also been reported, in Ghana, that taxi and bus drivers regularly arrive at traffic crash sites while either injured vehicle occupants or pedestrians are still present, and usually participate in the care and/or transport of such casualties (Tiska, et al., 2002). As commercial drivers play such a prominent part in the transport and care of road traffic accident casualties, it follows that if properly trained, these drivers could significantly improve pre-hospital trauma care. This suggests that improvements in pre-hospital care among commercial vehicle drivers, could potentially have an important impact on decreasing the mortality of critically injured road traffic casualties.

6.4 Information gain for future research

Since availability of Emergency Medical Services (EMS) to road traffic casualties is one of the key factors that determines the road traffic fatality rate, there is the need of a study into the allocation of ambulance services across sub-populations of a given geographical zone. EMS managers face the redeployment problem of relocating available ambulances to the potential location sites when calls are received. Thus, the objective of such a study is to formulate a model for an ambulance location problem in the geographical zone. The locations of emergency service stations such as ambulances and hospitals are of paramount importance in order to achieve an effective and reliable emergency response system. The fatalities and disabilities caused by road traffic accidents may be significantly reduced through an effective planning of the locations of these stations. To this end a Geographical Information System (GIS) implementation of the method of the current study would be quite useful.

Secondly, since the age of vehicles and availability of modern safety mechanisms in vehicles plying the roads of a particular geographical region have significant effect on consequences of road

traffic accidents, there is the need for a research to investigate the effect of age and availability of modern safety mechanism of vehicles on the rate of road traffic fatalities.

References

Abojaradeh, M. (2013). Traffic accidents prediction models to improve traffic safety in Greater Amman area. *Journal of Civil and Environmental Research*, **3** (**2**), 87 – 101.

Adams, J. G. U. (1987). Smeed‟s law: Some further thoughts. *Traffic Engineering and Control*, Available: http://john-adams.co.uk/wpcontent/uploads/2006/smeed's%20law.pdf**.** pp. 7073 [online, accessed December 2013].

Adekunle, J. A. (2010). Road traffic accident deaths and socio-economic development in echerc. *International Review of Business and Social Sciences*, **1** (**5**), 47 – 60.

Aeron-Thomas, A., Jacobs, G. D., Sexton, B., Gururaj, G. and Rahman, F. (2004). The involvement and impact of road crashes on the poor: Bangladesh and India case studies. Transport Research Laboratory, Published Project Report 010.

Afshartous, D., and de Leeuw, J. (2005). Prediction in Multilevel Models. *Journal of Educational and Behavioral Statistics*, **30** (**1**), 1 – 30.

Afukaar F. K. et al (2003). Pattern of road traffic injuries in Ghana: Implications for control. *Injury Control and Safety Promotion*, **10,** 69 – 76.

Ahmad, H. N., Jahir, B. A., Syeda, Z. F., and Abu, Z. (2012). Study on frequency analysis of Sylhet City‟s road accident. *International Journal of Engineering and Technology*, **2** (**4**), 608 – 615

Aitken, A. C. (1935). On least squares and linear combination of observations. *Proc. Roy. Soc. Edin.* **55,** 42 – 8.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & B. F. Csaki (Eds.), Second International Symposium on Information Theory, (pp. 267-281). Academiai Kiado: Budapest

Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control, AC-19, 716-723.

Alfaro, J. L., Chapuis, M., Fabre, F. (Eds). (1994). Socio-economic cost of road accidents: final report of action COST 313. Commission of the European Community, Brussels.

Al-Matawah, J. and Jadaan, K. (2009). Application of prediction techniques to road safety in developing countries. International Journal of Applied Science and Engineering. 7 (**2**), 169.

Anbarci, N., Escaleras, M. and Register, C. (2006). Traffic Fatalities and Public Sector Corruption. KYKLOS, **59** (**3**), 327 – 344.

Anderson, R. L. and Bancroft, T. A. (1952). Statistical Theory and Research. *McGraw-Hill, New York.*

Ansari, S., Akhdar, F. Mandoorah, M. and Moutaery, K. (2000). Causes and effects of road traffic accidents in Saudi Arabia. *Public Health*, **114**, 37 – 39.

Astrom, J. S., Kert, M. P. and Jovin, R. D. (2006). Signatures of four generations of Road Safety Planning in Nairobi City, Kenya. *Journal of Eastern African Research and Development*, **120**, 186 – 201.

Austin, J. T., Yaffee, R. A., & Hinkle, D. E. (1992). Logistic regression for research in higher education. *Higher Education: Handbook of Theory and Research, 8,* 379410.

AUSTROADS, 1994. Road Safety Audit. Sydney

Balogun JA, Abereoje OK 1992. Pattern of road traffic accident cases in a Nigerian University teaching hospital between 1987 and 1990. J Trop Med Hyg, 95(1): 2 3 – 29

Bancroft, T. A. (1968). Topics in intermediate statistical methods. *Iowa State university Press, Ames, Iowa*.

Banister, D. (2008). "The sustainable mobility paradigm," *Transport Policy,* 15, pp. 73-80.

Barbone, F., McMahon, A. D., Davey, P. G., Morris, A. D., Reid, I. C., McDevitt, D. G., Mac Donald, T. M. (1998). Association of road-traffic accidents with benzodiazepine use. *Lancet*, **352** (**9137**), 1331 – 1336.

Bartko, J. J. (1976). On various intraclass correlation reliability coefficients. *Psychological Bulletin, 83*, 762-765.

Bassiri, D. (1988). Large and small sample properties of maximum likelihood estimates for the hierarchical linear model. Unpublished doctoral dissertation, Department of Coun-seling, Educational Psychology and Special Education, Michigan State University.

Bauer, D. J. (2003). Estimating Multilevel Linear Models as Structural Equation Models. *Journal of Educational and Behavioral Statistics*, **28** (**2**), 135 – 167.

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Phil. Trans*., **53**, 370 – 418.

Bener, A. and Ofosu, J. B. (1991). Road traffic fatalities in Saudi Arabia. *Journal of the International Association of Traffic and Safety Sciences*, **15**, 35 – 38.

Berkhof, J., and Snijders, T. A. B. (2001). Variance component testing in multilevel models. *Educational and Behavioral Statistics*, **26**, 133 – 152.

Birch, M. W. (1964). A new proof of the Fisher-Pearson theorem. *Annals of Mathematical Statistics*, 35, 817 – 824.

Bishai, D., Quresh, James, A. P. and Ghaffar, A. (2006). National road casualties and economic development. *Health Economics*, **15**, 65 – 81.

Blaeij, A., de., Koetse, M., Tseng, Y., Rietveld, P., Verhoef, E. (2004). Valuation of safety, time, air pollution, climate change and noise; methods and estimates for various countries. Report prepared for ROSEBUD. Department of Spatial Economics, Vrije Universiteit, Amsterdam.

Bliese, P. D. (2000). Within-group agreement, non-independence, and reliability: Implications for data aggregation and Analysis. In K. J. Klein & S. W. Kozlowski (Eds.), *Multilevel Theory, Research, and Methods in Organizations* (pp. 349-381). San Francisco, CA: Jossey-Bass, Inc.

Blomquist, G. (1986). A Utility Maximization Model of Driver Traffic Safety Behavior. Accident Analysis and Prevention, **18** (**5**), 371 – 375.

Blum, U., Gaudry, M. (2000). The SNUS-2.5 Model for Germany. In: *Structural Road Accident Models: The International DRAG Family* (Gaudry, M. and Lassarre, S. Eds.), Chap. 3, 67 – 96, Elsevier Science, Oxford.

Boakye, A., Abledu, G. K., and Semevoh, R. (2013). Regression Analysis of Road Traffic Accidents and Population Growth in Ghana. *International journal of business and social research*, **3** (**10**), 41 – 47.

Box, G. E. P. and Tio, G. C. (1973). Bayesian Inference in Statistical Analysis. *AddisonWesley*.

Boyer, M. and Dionne, G. (1987). The Economics of Road Safety. *Transportation Research Part B: Methodological*, **21B** (**5**), 413 – 431.

Brooks, S. P. (1998). Markov chain Monte Carlo method and its application. *Journal of the Royal Statistical Society, Series D (The Statistician)*, 47 (1), 69 – 100.

Browne, W. J., and Draper, D. (2006). A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Analysis*, 1 (3), 473 – 514.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models*. Newbury Park, CA: Sage.

Burger, M. and Repiský, J. (2002). Problems of linear least square regression and approaches to handle them. *Advanced Research in Scientific Areas*, **3** (**7**) 257 – 262.

Burnham, K. P., and Anderson, D. R. (2001). Kullback-Leibler information as a basis for strong inference in ecological studies. Wildlife Research **28** :111-119.

Burnham, K. P., and Anderson, D. R. (2002). Model Selection and Multimodel Inference: a practical information-theoretic approach, 2nd edition. Springer-Verlag, New York.

Busing, F. (1993). Distribution characteristics of variance estimates in two-level models (Tech. Rep. No. PRM 93-04). Leiden, The Netherlands: Department of Psychometrics and Research Methodology, University of Leiden

Cabrera, A. F. (1994). Logistic regression analysis in higher education: An applied perspective. *Higher Education: Handbook of Theory and Research, Vol. 10,* 225-256.

Carpenter, C. (2004). How do zero tolerance drunk driving laws work? Journal of Health Economics, **23**, 61 – 83.

Chen, G. (2010). Road Traffic Safety in African Countries – status, trend, contributing factors, counter measures and challenges, *International Journal of Injury Control and Safety Promotion,* **17** (**4**), 247 – 255.

Chuang, H. L. (1997). High school youth‟s dropout and re-enrollment behaviour. *Economics of Educaution Review, 16(2),* 171-186.

Cochran, W. G. and Cox, G. M. (1957). *Experimental Designs*, 2nd Ed. *John Wiley and Sons, New York.*

Cook, P.A. and Bellis, M.A. (2001). Knowing the risk: relationships between risk behaviour and health knowledge. *Public Health*, **115**, 54 – 61.

Cox, D. R. (1958). Planning of Experiments. *John Wiley and Sons, New York*.

Crain, M. W. (1980). Vehicle safety inspection systems: How effective*?* Washington, DC, American Enterprise Institute for Public Policy Research.

Cramér, H. (1946). Mathematical Methods of Statistics. Princeton University Press, Princeton, N. J.

Cutter, S.L. (1993). Living with Risks: Geography of Technological Hard. Great Britain: Edward Arnold.

Daniel, W. W. (1980). Multiple comparison procedures. A selected Bibiography. Vance Bibliographies, Monticello, III, 1980.

Daniel, C., and Wood F. (1980). Fitting Equations to Data. 2nd ed., *John Wiley & Sons, New York.*

Daniel W. W. and Coegler, C. E. (1975), Beyond Analysis of variance. A comparison of some multiple comparison procedures. *Physical Therapy*, JJ, 114 – 15.

David, H . A., Hartleyh, O. & Pearsone, S. (1954). The distribution of the ratio, in a single normal sample, of range to standard deviation. *Biometrika,* **41,** 482 - 483.

DeGroot, M. H. (1970). Optimal Statistical Decisions. McGraw-Hill Book Company, New York

De Leeuw, J. and Kreft, I. (1995). Questioning multilevel models. *Journal of the Educational and Behavioral Statistics*, 20, 171 – 189.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1980). Iteratively reweighted least squares. In P. R. Krishnaiah (Ed.), Multivariate analysis V, 35 – 57. Amsterdam: North-Holland.

Didelot, X., Richard, G., Everitt, R., Johansen A., Daniel J. and Lawson D. (2011). Likelihoodfree estimation of model evidence. *Bayesian Analysis*, **6** (1), 49 – 76.

Draper, D. (1995). Inference and hierarchical odelling in the social sciences. *Journal of Educational and Behavioral Statistics*, **20**, 115 – 147.

Duncan, D. B. (1951). A significance test for difference between ranked treatments in an Analysis of variance. *Verginal Journal of science*, (2), 171 – 789.

Duncan, D. B. (1952). "On the properties of the multiple comparison tests. *Verginal Journal of Science,* (3), 50-67.

Duncan, D. B. (1955). Multiple Range and Multiple F. Tests. *Biometrics*, **11**, 1 – 42.

Draperi, N. R. and Smith, H. (1981) Applied Regression Analysis, Second Edition. *John Wiley & Sons, New York.*

Efron, B., and Morris, C. (1975). Data analysis using Stein‟s estimator and its generalizations. *Journal of the American Statistical Association*, **70**, 311 – 319.

Efron, E., & Morris, C. (1971). Limiting the risk of Bayes and empirical Bayes estimators, Part I: The Bayes case. *Journal of the American Statistical Association*, **66**, 807-815 Eliason, E. R. (1993). Maximum likelihood estimation. New York: Chapman & Hall.

Elvik, R. (2006A). Laws of accident causation. Accident Analysis and Prevention, **38**, 742 – 747.

Factor, R., Mahalel, D. and Yair, G. (2008). Inter-group differences in road traffic crash involvement. Accident analysis and Prevention, **40**, 2000 – 2007.

Fisher, R. A. (1966). The Design of Experiments, 8th Ed. Hafner Publishing Company, New York.

Fosgerau, M. (2005). Speed and Income. Journal of Transport Economics and Policy, **39** (**2**), 225 – 240.

Fouad, A. G. (1994). Application of Smeed‟s formula to access development of traffic safety in Jordan. Accid Anal, 6 (**6**).

Fouracre, P. and Jecobs, G. D. (1977). Further research on road accident rate in developing countries. TRRL report LR 270. Transport and Road Research Laboratory, Crowthorne, Berkshire.

Fournier, F. and Simard, R. (2000). The DRAG-2 Model for Quebec. In: *Structural Road Accident Models: The International DRAG Family* (Gaudry, M. and Lassarre, S. Eds.), Chap. 2, 37 – 66, Elsevier Science, Oxford.

Fridstrom, L. and Ingebrigtsen, S. (1991). An aggregate accident model based on combined close-sectional and time-series data. Working paper E-1523. Oslo, Norvay, Institute of Transport Economics.

Fujita, Y. and Shibata, A. (2006). Relationship between traffic fatalities and drunk driving in Japan. *Traffic Inj Prev*., 7, 325 – 327.

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

Gaudry, M. (1993). Le Modèle DRAG: echerch echerché au monde du travail – une expertise exploratoire. Publication CRT-948, Centre de echerché sur les transports, Université de Montréal.

Geisser, S. (1979). A predictive approach to model selection. *Journal of the American Statistical Association*, **74**, 153 – 160.

Gelfand, A. and Smith, A. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the Americal Statistical Association*, **85**, 398 – 409.

Gelman, A. and Rubin, D. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, **7**, 457 – 472.

Gelman, A. (2003). Bugs.R: functions for calling Bugs from R. www.stat.columbia.edu/_gelman/bugsR/

Geman, S., and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. IEEE Trans. *Pattern Anal. Machine intell*., **6**, 721 – 741.

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

Gilks, W., Bes,t N., and Tan, K. (1995). Adaptive rejection Metropolis sampling with Gibbs sampling. *Applied Statistics*, **44**, 455 – 472.

Gilks, W. R., Roberts, G. O., and Sahu, S. K. (1996a). Adaptive Markov chain Monte Carlo. Preprint.

Goldstein, H. (2003). Multilevel statistical models (3rd ed.) London Edward Anold.

Goldstein, H. and Yang, M. (2000). Meta-analysis using multilevel models with an application to the study of class size effect. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, 49 (3), 399 – 412.

Gordon, J. E. (1949). The epidemiology of accidents. *Amer. J. Public Health*, **39** (**4**) 504 – 515.

Goswami and Sonowal. (2011). Statistical Analysis of road traffic accident data for the year 2009 in Dibrugarh city, Assam, India. http:interstat.statjournals.net/YEAR/2011/articles/

Gray, J., Goldstein, H., Thomas, S. (2001). Predicting the future: The role of past perfor- mance in determining trends in institutional effectiveness at A level. *British Educational Research Journal*, **27**, 391 – 406.

Greenland, S. (1989). Modelling variable selection in epidemiologic analysis. *American Journal of Public Health*, 79, 340–349.

Gulliford, M. C., Ukoumunne, O. C., and Chinn, S. (1999). Components of variance and intraclass correlations for the design of community-based surveys and intervention studies. *American Journal of Epidemiology*, **149**, 876 – 883.

Gururaj, G. (2004). Alcohol and Road Traffic Injuries in South Asia: Challenges for prevention, *JCPSP* **14** (**12**), 713 – 718.

Hakim, S., Shefer, D. and Hakkert, A. (1991). A Critical Review of Macro Models for Road Accidents. *Accident Analysis and Prevention*, **23** (**5**), 379 – 400.

Hakkert, A., Livneh, M. and Mahalel, D. (1976). Levels of safety in accidents studies. A safety in accidents studies a Saftey index. Australian Road Research Board Proc.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. Biometrika, **57**, 97 – 109.

Hilden-Minton, J. (1995). Multilevel diagnostics for mixed and hierarchical linear models. Unpublished doctoral dissertation, UCLA.

Hill, B. M. (1965). Inference about variance components in the one-way model. *Journal of the American Statistical Association*, **60**, 806 – 825.

Hogg, V. & Craig,"AJ.. (1956). S~lfficienst tatistics in elementary distribution theory. *Sankyd,* **17,** 209 – 216.

Hosmer, D. W., Lemeshow, S. and Sturdivant, R. X. (1989). *Applied Logistic Regression*. John Wiley & Sons, Inc., Hoboken, New Jersey.

Hox, J. J. (1998). Multilevel modeling. When and why. In I. Balderjahn, R. Mathar and M Schader (Eds.), (classification, data analysis, and data highways), 147 – 154. New York: Springer Verlag.

Hox, J. J. (2010). Multilevel analysis: Techniques and applications 2nd edition. Routledge Taylor and Francis Group.

Jacobs, G. and Aeron-Thomas, A. (2000). *Africa road safety review final report.* Washington, DC: US Department of Transportation, Federal Highway Administration.

Jacobs, G. D. and Cutting, C. A. (1986). Further research on accident rates in developing countries. Accident Analysis and Prevention, **18** (**2**), 119 – 127.

Jacobs, G. D., Aeron-Thomas, A. and Astrop, A. (2000). Estimating global road fatalities. Transport Research Laboratory, TRL Report 445.

Jacobs, G., and Bardsley, M. (1977). Research on road accidents in developing countries. Traffic engineering & control.

Jeong-Hun H. (2007). Analysing Roll Calls of the European Parliament: A Bayesian Application. *European Union Politics*, **8** (**4**), 479 – 507.

Johnson, N. L and Leone, F. C. (1976). Statistics and Experimental Design in Engineering and Physical Sciences. *John Wiley and Sons, New York*.

Janik, J., & Kravitz, H. M. (1994). Linking work and domestic problems with police suicide. *Suicide and* Life *Threatening Behavior;* 24(3), 267-274.

Kajubi, P., Kamya, M. R. Kamya, S., Chen, S McFarland, W. and Hearst, N. (2005). Increasing Condom Use Without Reducing HIV Risk: Results of a Controlled Community Trial in Uganda. JAIDS Journal of Acquired Immune Deficiency Syndromes, **40** (**1**), 77 – 82.

Kass, R. E., and Raftery, A. E. (1995), "Bayes Factors and Model Uncer- tainty," Journal of the American Statistical Association, 90, 773-795.

Keeler, T. E. (1994). Highway Safety, Economic Behavior, and Driving Environment. American Economic Review, 84 (3), 684-693.

Kempthorne, O . (1952). The design and analysis of experiments, *John Wiley and Sons*, New York.

Kendall, M. G. & Stuart, A. (1973). The Advanced Theory of Statistics. Vol. 2, Inference and Relationship (3rd edn). *Griffin*, London.

Kenkel, D.S. (1991). Health behavior, health knowledge and schooling. The Journal of Political Economy, **99** (**2**), 287 – 305.

Keuls, M. (1952). The use of the studentized range in connection with an Analysis of variance. Euphytica, **1**, 112 – 122.

Kim, K. S. (1990). Multilevel data analysis: A comparative examination of analytical alternatives. Unpublished doctoral dissertation, Department of Education, University of California, Los Angeles.

Komba, D.D. (2006). Risk Factors and Road Traffic Accidents in Tanzania: A case study of Kibaha District. Master Thesis, Department of Geography, Norwegian University of Science and Technology (NTNU), Trodheim

Koornstra, M. J. (2007). Prediction of traffic fatalities and prospects for mobility becoming sustainable-safe. *Sadhna*, **32** (**4**), 365 – 395.

Kopits, E. and Cropper, M. (2005). Traffic fatalities and economic growth. Accident Analysis and Prevention, **37**, 169 – 178.

Kreft, I. G. G. (1996). Are multilevel techniques necessary? An overview, including simulation studies. Unpublished manuscript, California State University at Los Angeles. Retrieved July 6, 2005 from http://www.calstatela.edu/faculty/ikreft/quarterly.html

Kullback, S., and Leibler, R. A. (1951). On information and sufficiency. Annals of Mathematical Statistics **22** :79-86.

Lancaster, H. O. (1969). The chi-squared distribution. John Wiley and Sons, New York.

La-Torre, Van Beeck, G., Quaranta, E. G., Mannocci, A. and W. Ricciardi. (2007). Determinants of within-country variation in traffic accident mortality: a geographical analysis. *International Journal of Health Geographics*, **6**, 49.

Lagarde, E. (2007). Road Traffic Injury is an Escalating Burden in Africa and Deserves Proportionate Research Efforts. *PLOS Med*, **4** (**6**), 967 – 971.

Laplace, *A*. (1902). *Philosophical Essay*, New York, 107–108. (Translation in this paragraph of article is from Hahn)

Latinpoulou, M. P. (1982). Application of Smeed‟s equation for road accidents in Greece.Greece accident analysis.

Lave, C. A. (1985). Coordination, and the 55MPH limit. *American Economic Review*, **75** (**5**), 1159 – 1164.

Levene, H. (1960). In *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, I. Olkin et al. eds., Stanford University Press, pp. 278-292.

Lindley, D. V. (1965). Introduction to probability and statistics from a Bayesian viewpoint, Part 2, Inference. *Cambridge University Press*.

Lloyd, E. H. (1952). Least squares estimation of location and scale parameters using order statistics. *Biometrika,* **39,** 88 – 95.

Loeb, P. D. (1987). The determinants of automobile fatalities. *Journal of Transport Economics and Policy*, **21**, 279 – 287.

Loughran, D. S., Seabury, S. A. and Zakaras, L (2007). *Regulating Older Drivers: Are New Policies Needed?* Santa Monica, Calif.: RAND Corporation, OP-189-ICJ.

Lozano, R., Naghavi, M., Foreman, K., Lim, S., Aboyans, V., and others (2012). Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010**.** *Lancet medical journal*, 380 (**9859**), 2095 – 2128.

Lu M., Wevers K. (2005). Modelling and assessing the effects of road traffic safety measures. 18th ICTCT workshop, Poster session.

Lunn D, Thomas A, Best N, and Spiegelhalter D. (2000). WinBUGS – a Bayesian odeling framework: concepts, structure, and extensibility, *Statistics and Computing*, **10**, 325 – 337.

Maas C. J. M. and Hox J. J., (2005). Sufficient Sample Sizes for Multilevel Modeling. *Hogrefe & Huber Publishers*, 1 (3), 86 – 92.

Maddala, G. S. (1977). Econometrics. Singapore: McGrav-Hill.

Maldonado, G., and Greenland, S. (1993). Interpreting model coefficients when the true model form is unknown. *Epidemiology*, 4, 310–318.

Martinez, W. L. and Martinez, A. R. (2002). Computational statistics handbook with MATLAB. Chapman & Hall/CRC, Boca Raton London New York Washington, D.C.

Mishra, B, Sinha, N. D, Sukkla, S. K, Sinha, A. K (2010). Epidemiological Study of Road Traffic Accident Cases from W estern Nepal, Indian J. Community Med. 35(1):115-121 Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, **21** (**6**), 1087 – 1092.

McCarthy, P. (2000). The TRACS-CA Model for California. In: *Structural Road Accident Models: The International DRAG Family* (Gaudry, M. and Lassarre, S. Eds.), Chap. 7, 185 – 204, Elsevier Science, Oxford.

McCullagh, P., and Nelder, J. A. (1989). Generalized Linear Models, 2nd edition. Chapman & Hall, New York, USA.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E. (1953). Equations of state calculations by fast computing machines. *J. Chem. Phys*., **21**, 1087 – 1091.

Miettinen, O. S. (1976). Stratification by multivariate confounder score. *American Journal of Epidemiology*, 104, 609–620.

Mishra, B., Sinha, N.D., Sukkla, S.K., Sinha, A.K (2010). Epidemiological Study of Road Traffic Accident Cases from Western Nepal, *Indian J. Community Med.*, **35** (**1**), 115–121

Mohammad, S. (2009) A Statistical Analysis of Road Traffic Accidents and Casualties in Bangladesh. http://researchrepository.napier.ac.uk/2753/ Access date: 13/07/2012

Mohan, D. (2002). Road safety in less-motorised environments: Future concerns. *Int. J. Epidemiol,* **31**, 327 – 532.

Mok, M. (1995, June). Sample sizes for 2-level designs in educational research. Multi- level Modeling Newsletter.

Mondal, P., Abhishek, K., Bhangale, U., and Dinesh, T. (2011). A silent tsunami on Indian road: A comprehensive analysis of epidemiological aspects of road traffic accidents. *British Journal of Medicine & Medical Research*, 1 (**1**), 14 – 23.

Montgomery, D. C., and Peck, E. A. (1992). Introduction to linear regression analysis, 2nd ed., *John Wiley & Sons*, New York.

Montgomery, D. C. (2001). Design and analysis of experiments, 5th edition. *John Wiley and Sons*, New York.

Montgomery, D, Peck, E. A, Vining, G. G. (2006). Introduction to linear regression analysis. *Wiley Interscience, New York*.

Montazeri, A. (2004). Road-traffic-related mortality in Iran: descriptive study. *Public Health*, **118**, 110 – 113.

Mooney, C. Z. and Duval R. D. (1993). Bootstrapping, A nonparametric approach to statistical inference. Newbery Park.CA: Sage.

Mortimore P, Sammons P, Stoll L, Lewis D, Ecob R (1988). School Matters. Wells: Open Books.

Moutari S. Herty M. Klein A. Oeser M. Schleper V. Steinauer B. (2005). Modeling road traffic accidents using macroscopic second-order models of traffic flow. *IMA Journal of Applied Mathematics*, **1**, 1 – 23.

Murray, C. and Lopez, A. (eds.). (1996). The Global Burden of Disease. Cambridge, M A: Harvard Press.

Myers, R. H. (1990). Classical and Modern Regression Applications, 2nd ed. *PWS-Kents*, Boston.

National Road Safety Commission of Ghana (2013). Building and Road Research Institute (BRRI), *Road Traffic Crashes in Ghana*, Statistics.

Nilambar, J., Srinivasa, D., Gautam, R., and Jagdish, S. (2004). Epidemiological study of road traffic accident cases: a study from South India. *Indian Journal of Community Medicine*, **29** (**1**), 20 – 24.

Newman, D. (1939). The distribution of the range in samples from a normal population in terms of an independent estimate of standard deviation. *Biometrica*, 31, 20 – 30.

Norton, R. M. (May 1984). *"The Double Exponential Distribution: Using Calculus to Find a Maximum Likelihood Estimator"*. The American Statistician (American Statistical Association) **38** (**2**): 135–136.

Oberwittler, D. (2004). A Multilevel Analysis of Neighbourhood Contextual Effects on Serious Juvenile Offending: The Role of Subcultural Values and Social Disorganization. *European Journal of Criminology.* 1: 201, DOI: 10.1177/1477370804041248.

Odero, W., Garner, P., and Zwi, A. (1997). Road Traffic Injuries in Developing Countries: A comprehensive Review of Epidemiological Studies. Tropical Medicine and International Health, **2** (**5**), 445 – 460.

OECD (Organisation for Economic Coopration and Development), 2006. Young Drivers: The Road Safety. <http://www.cemet.org>. Muhlrad, N; Lassare, S. (2005). *Systems approach to injury control,* New Delhi: Macmillan India Ltd.

Ofosu, J. B., & Hesse, C. A. (2011). Elementary Statistical Methods. *EPP Books Services, Accra*.

Ofosu, J. B., Hesse, C. A., and Otchere, F. (2014). Intermediate Statistical Methods. *EPP Books Services, Accra*.

Ofosu, J. B., & Hesse, C. A. (2010). Introduction to Probability and Probability Distributions. *EPP Books Services, Accra*.

O‟Neill, R. and Wetherill, G. B. (1971). The Present State of Multiple Comparison Methods. *Journal of the Royal statistical Society*, B, 33, 218 – 241.

Patrick W. J. (2001). Estimating First-Year Student Attrition Rates: An Application of Multilevel Modeling Using Categorical Variables. *Research in Higher Education*, **42** (**2**), 151 – 170.

Patton, G. C., Coffey, C., Sawyer, S. M., Viner, R. M., Haller, D. M., Bose, K., Vos, T., Ferguson, J., and Mathers, C. D. (2009). Global patterns of mortality in young people. *Lancet*, **374** (**9693**), 881 – 892.

Paulozzi, L. J., Ryan, W. R., Espitia-Hardeman, V. E. and Xi, Y. (2007). Economic development‟s effect on road transport-related mortality among different types of road users: a cross-sectional international study. Accident Analysis and Prevention, **39**, 606 – 617.

Pearsone,. S. and Stephensm, . A. (1964). The ratio of range to standard deviation in the same normal sample. *Biometrika,* **51,** 484 – 487.

Pebley AR, Goldman N (1992). Family, community, ethnic identity, and the use of formal health care services in Guatemala. Working Paper 92{12, Princeton NJ: O_ce of Population Research.

Peden, M., Scurfield, R., Sleet, D., Mohan, D., Jyder, A., Jarawan, E., and Mathers, C. (2004). World Report on Road Traffic Injury Prevention. Geneva: World Health Organization.

Peltzer, K. & Renner, W. (2004). Psychosocial correlates of the impact of road traffic accidents among South African drivers and passengers. *Accident Analysis and Prevention,* **36**, 367 – 374.

Pinheiro, J. C. & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. New York: Springer-Verlag.

Pitt, M. A., & Myung, I. J. (2002). When a good fit can be bad. *Trends in Cognitive Sciences*, *6*, 421–425.

Pluddenmann, A., Parry, C.D.H., Donson, H., Sukhai, A. (2004). Alcohol use and trauma in Cape Town, Durban and Port Elizabeth, South Africa: 1999-2001, *Injury Control and Safety Promotion*, **11** (**4**), 265 – 267.

Plummer M, Best N, Cowles K, and Vines K. (2006). CODA: Convergence Diagnosis and Output Analysis for MCMC, http://CRAN.R-project.org/doc/Rnews/Rnews_2006-1.pdf.

Polen, M. R., Friedman, G. D. (1988). Automobile injury – Selected Risk Factors and Prevention in the Health Care Setting. *Journal of the American Medical Association*, **259** (**1**), 77 – 81.

Ponnaluri, R. V. (2012). Modeling road traffic fatalities in India: Smeed's law, time invariance and regional specificity. International Association of Traffic and Safety Sciences, **36**, 75 – 82.

Pratte, D. (1998). Road to Ruin: Road Traffic Accidents in the Developing World. *NEXUS*, **13**, 46 – 62.

Raiffa, H. and Schlaifer, R. (1961). *Applied Statistical Decision Theory*. Division of Research, Graduate School of Business Administration, Harvard University.

Rao, C. R. (1973). Linear Statistical Inference and its applications, 2nd edition. *John Wiley and Sons Ltd., New York.*

Rodriguez G, Goldman N (1995). An assessment of estimation procedures for multilevel models with binary responses. *Journal of the Royal Statistical Society, Series A*, **158**, 73 – 89.

Rossi, P. E. and Allenby, G. M. (2003). Bayesian Statistics and Marketing. Marketing Science, **22** (**3**), 304 – 328.

Rothman, K. J., Greenland, S., and Lash, T. L. (2008). *Modern Epidemiology*, *Third Edition*, Lippincott-Raven, Philadelphia.

Rubin, D. (1980). Using empirical Bayes techniques in the Law School Validity Studies. Journal of the American Statistical Association, **75**, 801 – 827.

Sachs L (1997). Angewandte Statistik. 8 edition. Springer, Berlin.

Schwarz, G. (1978). Estimating the dimension of a model. Annals of Statistics 6, 461{464 Kass, R. and Raftery, A. (1995). Bayes Factors. Journal of the American Statistical Association 90, 773(795).

Safe Speed (SS) .(2013). Smeed and beyond: predicting road deaths, updated: http://www.safespeed.org.uk/smeed.html, [online, accessed December. 4, 2013].

Sagberg, F., Glad, A. (1999). Traffic safety for the elderly: Literature study, risk analyses and assessment of safety measures, Osho Institute of Transport Economics.

Sakaran, U (2003) „*Research Methods of Business.* New York: John Wiley & Sons Ink.

Sarhana, . E. & Greenbergb,. G. (1956). Estimation of location and scale parameters by order statistics from singly and double censored samples. Part I. *Ann. Math. Statist.* **27,** 427-51.

SAS Institute Inc. 2006, Preliminary Capabilitites for Bayesian Analysis in SAS/STAT Software, SAS Institute Inc., Cary, NC, USA.

Scheffé H. (1953). A method for judging all contrasts in the Scheffé, H. (1950) Analysis of variance. *Biometrika*, **40**, 87 – 104.

Scheffé, H. (1959). Analysis of variance. *Wiley and Sons*. New York.

Schroer, B. J. and PEYTON, W. (1979). The effects of automobile inspections on accident rates. *Accident Analysis and Prevention*, **11**, 61-68.

Seltzer, M. H., Wong, W. H., Anthony, S. and Bryk, A. S. (1996). Bayesian Analysis in Applications of Hierarchical Models: Issues and Methods. *Journal of Educational and Behavioral Statistics*, **21** (**2**), 131 – 167.

Seltzer, M. (1993). Sensitivity analysis for fixed effects in the hierarchical model: A Gibbs sampling approach. *Journal of Educational Statistics*, **18**, 207 – 235.

Shapiro, S. S. & Wilk, M. B. (1965a). Testing the normality of several samples. (Unpublished manuscript).

Shapiro, S. S. (1964). An analysis of variance test for normality (complete samples). Unpublished Ph.D. Thesis, Rutgers-The State University

Shirley M. (2006). Road Traffic Accidents – A Challenging Epidemic. *Sultan Qaboos University Medical Journal*, **6** (**1**), 3 – 5.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420-428.

Sivak, M. (1983). Society‟s aggression level as a predictor of traffic fatality rate. *Journal of Safety Research*, **14**, 93 – 99.

Smeed, R. J. (1949). Some statistical aspects of road safety research. *J. Roy Stats. Soc. Series-A* **12** (**1**), 1 – 23.

Smeed, R. J. (1964). Methods available to reduce the number of road casualties. *Traff. Eng. And Control*, 6, 509 – 515.

Smith B. (2005). Bayesian Output Analysis Program (BOA), Version 1.1.5, The University of Iowa, http://www.public-health.uiowa.edu/boa.

Smith, A. F. M., and Roberts, G. O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods (with discussion). *J. Roy. Statist. Soc. Ser. B*, **55**, 3 – 24.

Snedecor, G. and Cochran, W. G. (1989). Statistical Methods, 8th ed. Ames. Iowa: *The Iowa State University Press.*

Snijders, T. A. B. & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage Publications.

Söderlund N, Zwi A. B. (1995) Traffic-related mortality in industrialized and less developed countries. *Bulletin of the World Health Organization*, **73**, 175 – 182.

Sparapani A. and Laud W. (2008). A Recent History of Bayesian Statistical Software. Division of Biostatistics Medical College of Wisconsin.

Sparapani, R. (2004). Some SAS for BUGS Data, http://ww2.mcw.edu/pcor/bugs/tr049.pdf.

Spiegelhalter D, Thomas A, and Best N. (1995a). Computation on Bayesian graphical models. In Bayesian Statistics 5.

Spiegelhalter D, Thomas A, Best N, and Gilks W. (1995b). BUGS: Bayesian inference using Gibbs sampling, Version 0.50, MRC Biostatistics Unit, Cambridge, UK.

Steel, R. G. D., and Torrie, J. H. (1979). Principles and Procedures of Statistics, 2nd ed. *McGraw-Hill Book Company, New York.*

Steyvers, M. (2011). Computational statistics with MATLAB. University of California, Irvine, psiexp.ss.uci.edu/research/teachingP205C/205C.pd.

Sturtz, S, Ligges, U. and Gelman, A. (2005). R2WinBUGS: A Package for Running WinBUGS from R.

Tegner, G., Holmberg, I., Loncar-Lucassi, V. And Nilsson, C. (2000). The DRAG-Stockholm Model. In: *Structural Road Accident Models: The International DRAG Family* (Gaudry, M. and Lassarre, S. Eds.), Chap. 5, 127 – 156, Elsevier Science, Oxford.

Thomas, A., O"Hara, B., Ligges, U. and Sturtz, S. (2006). Making BUGS Open, http://CRAN.R-project. Org/doc/Rnews/Rnews_2006-1.pdf.

Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion). *Ann. Statist*., **22**, 1701 – 1762.

Tiska, M. A., Adu-Ampofo, M., Boakye G., Tuuli L., and Mock, C. N. (2002). A model of pre-hospital trauma training for lay persons devised in Africa. *Emergency Medical Journal*, **21**, 237–239.

Tolman, R. M., & Weisz, A. (1995). Coordinated community intervention for domestic violence: The effects of arrest and prosecution on recidivism of woman abuse perpetrators. *Crime and Delinquency,* 41(4), 481495.

Tortum, A., Çodur, M. Y., and Kilinç, B. (2012). Modeling Traffic Accidents in Turkey Using Regression Analysis. Iğdır University Journal of the Institute of Science and Technology, **2** (**3**), 69 – 78.

Trawén, A., Maraste, P. and Persson, U. (2001). Methods for estimating road accident costs – A comparison of costs for a fatal casualty in different countries. Paper to Traffic Safety on Three Continents, International Conference in Moscow, 19 – 21.

Traynor, T. L. (2008). Regional economic conditions and crash fatality rates: a cross county analysis. Journal of Safety Research, **39**, 33 – 39.

Turner, J.M., Greico, M. and Kwakye, E. A. (1995). *Subverting Sustainability? Infrastructural and Cultural Barriers to Cycle Use in Accra*. Seventh World Conference on Transport Research, Sydney, Australia.

Tukey, J. W. (1949). Comparing Individual Mean in the Analysis of variance. Biometrics, 5, 99 – 114.

Tukey, J. W. (1953). The problem of multiple comparisons. *Ditto*, Princeton University.

Van Beeck, E. F., Borsboom, G .J . J. and Mackenbach, J . P. (2000). Economic development and traffic accident mortality in the industrialized world, 1962 – 1990. *International Journal of Epidemiology*, **29**, 503 – 509.

Van der Leeden, R., and Busing, F. (1994). *First iteration versus IGLS RIGLS estimates in twolevel models: A Monte Carlo study with ML3*. Unpublished manuscript, Leiden University, the Netherlands.

Van der Leeden, R., Busing, F., and Meijer, E. (1997). *Applications of bootstrap methods for two-level models*. Paper presented at the Multilevel Conference, Amsterdam.

Van der Pol, M. and Ruggeri, M. (2008). Is risk attitude outcome specific within health domain? Journal of Health Economics, **27**, 706 – 717.

Walker, L. J., Gustafson P., and Jeremy A. Frimer J. A. (2007). The application of Bayesian analysis to issues in developmental research. International Journal of Behavioral Development, **31** (**4**), 366–373.

Walpole, R. E., Myers, R. H. and Myers, S. L. (1998). Probability and Statistics for Engineers and Scientists, 6th Edition. *Prentice Hall International*, Inc. New Jersey.

Wedagama, D. M. P. (2012). Estimating the Influence of Accident Related Factors on Motorcycle Fatal Accidents using Logistic Regression. http://www.researchgate.net/publication/45825669_Estimating_the_Influence_of_Accident_Related_Factors_on_Motorcycle_Fatal_Accidents_using_Logistic_Regression_(Case_Study_DenpasarBali). Access date: 17/04/2012.

Wesemann, P. (2000). Economic evaluation of road safety measures. Contribution to the 117th Round Table, 26 and 27 October 2000, Paris. SWOV Publication D-2000-16E. SWOV Institute for Road Safety Research, The Netherlands.

Westfall, Tobias, Rom, Wolfinger, and Hochberg, (1999). Multiple Comparisons and Multiple Tests Using the SAS ® System, SAS Institute Inc., Cary, NC, USA. ISBN 1-58025-397-0.

White, J., Standish, D., Thorrold, R., and Warner, R. (2008). Markov chain monte carlo methods for assigning larvae to natal sites using natural geochemical tags. Ecological Applications, **18** (**8**), 1901 – 1913.

White, W. T. (1986). Does vehicle inspection prevent accidents? *Accident Analysis and Prevention*, **7**, 281 – 288.

Williams, A.F. (2003). Teenage Drivers: Patterns of risk. *Journal of Safety Research*, **34**, 5-15.

Winer, B. J. (1971). Statistical Principles in Experimental Design. 2nd edition, McGraw-Hill, New York.

Wintemute, G. J. (1985). Is motor vehicle-related a disease of development?, Accident Analysis and Prevention, **17** (**3**), 223 – 237.

Wolfinger, R. (2000). Nonconjugate Bayesian analysis of variance component models. *Biometrics*, **56**, 768 – 774.

Woodhouse, G., Rasbash, J., Goldstein, H., Yang, M., Howarth, J. and Plewis, I. (1995). A Guide to MLn for New Users. London: Institute of Education, University of London.

World Health Organization (2004) World report on road traffic injury prevention, Geneva.

World Health Organization. (2009). Global Status Report on Road Safety, time for action. WHO, Geneva.

World Health Organization (1979). Road Traffic Accidents Statistics. Copenhagen, Regional Office for Europe.

Yamamura, E. (2008). Impact of formal and informal deterrents on driving behavior. The Journal of Socio-Economics, **37**, 2505 – 2512.

Zlatoper, T.J. (1984). Regression analysis of time series data on motor vehicle deaths in the United States, *Journal of Transport Economics and Policy,* **18** (**3**), 263 – 274.

Appendix

Table A1: Population and RTA pattern in Ghana during the period 1991 to 2011

Year	Population × 10^3	No. of RTAs	No. injured from RTAs	Mortality from RTAs	Injury accidents	Fatal Accidents	No. injured per 100 000 population	No. of persons injured per accident	Death rate per 100 000 population	Death rate per 100 accidents	No. of injurious accidents per 100 RTAs	No. of fatal accidents per 100 RTAs
1991	14821	8370	8773	920	4866	724	59.2	1.0	6.2	11.0	58.1	8.6
1992	15222	6922	9116	914	4515	717	59.9	1.3	6.0	13.2	65.2	10.4
1993	15634	6467	7677	901	4119	704	49.1	1.2	5.8	13.9	63.7	10.9
1994	16056	6584	7664	824	4088	632	47.7	1.2	5.1	12.5	62.1	9.6

1995	16491	8313	9106	1026	4897	813	55.2	1.1	6.2	12.3	58.9	9.8
1996	16937	8488	9903	1049	4964	830	58.5	1.2	6.2	12.4	58.5	9.8
1997	17395	9918	10433	1015	5638	864	60.0	1.1	5.8	10.2	56.8	8.7
1998	17865	10996	11786	1419	6370	1127	66.0	1.1	7.9	12.9	57.9	10.2
1999	18349	8763	10202	1237	5303	979	55.6	1.2	6.7	14.1	60.5	11.2
2000	18845	11087	12310	1437	6429	1092	65.3	1.1	7.6	13.0	58.0	9.8
2001	19328	11293	13178	1660	6831	1257	68.2	1.2	8.6	14.7	60.5	11.1
2002	19811	10715	13412	1665	6593	1245	67.7	1.3	8.4	15.5	61.5	11.6
2003	20508	10542	14469	1716	6849	1327	70.6	1.4	8.4	16.3	65.0	12.6
2004	21093	12175	16259	2186	7852	1600	77.1	1.3	10.4	18.0	64.5	13.1
2005	21694	11320	14034	1776	7025	1388	64.7	1.2	8.2	15.7	62.1	12.3
2006	22294	11668	14492	1856	7137	1419	65.0	1.2	8.3	15.9	61.2	12.2
2007	22911	12038	14373	2043	7533	1622	62.7	1.2	8.9	17.0	62.6	13.5
2008	23544	11214	14531	1938	7309	1647	61.7	1.3	8.2	17.3	65.2	14.7
2009	24196	12299	16259	2237	8188	1790	67.2	1.3	9.2	18.2	66.6	14.6
Total	**362994**	**189172**	**227977**	**27819**	**116506**	**21777**	**1181.4**	**22.9**	**142.1**	**274.1**	**1168.9**	
Mean	**19104.9**	**9956.4**	**11998.8**	**1464.2**	**6131.9**	**1146.2**	**62.2**	**1.2**	**7.5**	**14.4**	**61.5**	

Table A2: Annual distribution of road traffic fatalities by gender

Year	Fatalities		Male/female ratio	Year	Fatalities		Male/female ratio
	Male	Female			Male	Female	
1991	642	273	2.4	2002	1175	480	2.4
1992	647	253	2.6	2003	1280	437	2.9
1993	662	210	3.2	2004	1568	587	2.7
1994	616	196	3.1	2005	1292	463	2.8
1995	708	290	2.4	2006	1348	492	2.7
1996	744	280	2.7	2007	1554	489	3.2
1997	728	273	2.7	2008	1448	490	3.0
1998	1013	381	2.7	2009	1655	582	2.8
1999	887	315	2.8	**Total**	**21762**	**7848**	**2.8**

2000	1091	441	2.5	**(%)**	**73.3**	**26.7**	
2001	1193	441	2.7				

Table A3: Months during which persons were killed or injured in RTAs, in 2010 and 2011

	2010				2011			
	Fatalities		Persons injured		Fatalities		Persons injured	
Month	**Number**	**%**	**Number**	**%**	**Number**	**%**	**Number**	**%**
January	124	6.2	1316	8.8	176	8.0	1103	7.9
February	139	7.0	975	8.5	142	6.5	934	6.7
March	112	5.6	1211	8.1	187	8.5	1138	8.1
April	181	9.1	1120	7.5	178	8.1	1192	8.5
May	167	8.4	1405	9.4	190	8.6	1212	8.6
June	143	7.2	1091	7.3	148	6.7	1055	7.5
July	170	8.6	1008	6.8	177	8.0	1069	7.6
August	129	6.5	1170	7.8	174	7.9	1173	8.4
September	163	8.2	1413	9.5	199	9.0	1296	9.2
October	237	11.9	1430	9.6	160	7.3	1143	8.2
November	188	9.5	1336	9.0	260	11.8	1376	9.8
December	233	11.7	1443	9.7	208	9.5	1329	9.5
Total	**1986**	**100**	**14918**	**100**	**2199**	**100**	**14020**	**100**

Table A4: Day of occurrence of road traffic accidents, from January 2010 to December 2011

	2010				2011			
	Fatalities		Persons injured		Fatalities		Persons injured	
Day	**Number**	**%**	**Number**	**%**	**Number**	**%**	**Number**	**%**
Monday	258	13.0	2061	13.8	323	14.7	1794	12.8
Tuesday	249	12.5	1901	12.7	282	12.8	1750	12.5
Wednesday	218	11.0	1866	12.5	267	12.1	1966	14.0
Thursday	245	12.3	1930	12.9	318	14.5	1778	12.7
Friday	297	15.0	2300	15.3	312	14.2	2218	15.8
Saturday	403	20.3	2583	17.3	398	18.1	2503	17.9
Sunday	316	15.9	2300	15.4	299	13.6	2011	14.3
Total	**1986**	**100**	**14918**	**100**	**2199**	**100**	**14020**	**100**

Table A5: Road user class involved in deaths and injuries

	Road User Class and Vehicle type in accidents																					
	Pedestrian		Car			Heavy Goods Vehicles (HGVs)			Bus/Mini Bus			Motor cycle			Pick-up			Bicycle			Other		
Year	No. Killed	No. Injured	No. Killed	No. Injured	No. of Vehicles	No. Killed	No. Injured	No. of Vehicles	No. Killed	No. Injured	No. of Vehicles	No. Killed	No. Injured	No. of Vehicles	No. Killed	No. Injured	No. of Vehicles	No. Killed	No. Injured	No. of Vehicles	No. Killed	No. Injured	No. of Vehicles
1991	423	2250	85	1852	6544	106	759	1283	177	2529	288	16	242	311	45	638	795	28	258	442	17	72	134
1992	388	1971	126	1883	4921	83	613	1081	215	3339	2381	18	211	258	23	625	731	43	267	402	12	42	113
1993	404	1806	93	1625	4721	118	494	976	186	2736	2356	11	228	279	29	372	637	35	248	359	18	44	114
1994	367	1826	81	1602	4728	91	488	1116	180	2733	2585	18	203	260	41	431	708	22	227	305	17	61	115
1995	488	2266	95	1733	6410	87	671	1440	232	3325	3145	21	221	288	34	454	929	40	263	359	19	45	123
1996	461	2408	115	1711	6485	130	872	1418	197	2661	3419	15	262	337	47	540	1004	44	254	358	32	84	157
1997	491	2569	107	1912	7258	111	608	1741	181	3982	4291	28	310	435	48	566	1154	30	298	388	10	72	152
1998	630	2777	137	2001	8011	150	743	1772	328	4597	4839	29	376	470	55	787	1335	63	331	491	24	74	178
1999	528	2165	142	1798	6146	111	738	1522	281	4263	3708	35	343	436	50	492	1046	60	281	426	10	63	165
2000	662	2965	207	2679	9270	189	932	1853	314	4886	4705	42	414	539	72	682	1208	62	332	498	13	62	225
2001	757	2899	182	2783	8852	146	959	1740	399	5089	4607	44	402	518	41	512	1175	59	357	470	31	131	262
2002	681	2757	202	2783	8314	171	1079	2089	421	5577	4312	48	380	469	57	454	1082	69	334	478	16	46	114
2003	724	2784	218	2874	7696	228	1335	2193	341	6144	4326	53	496	616	47	454	986	91	360	562	16	82	154
2004	869	3146	246	3153	8904	235	1427	2598	556	6749	4849	100	685	792	53	519	1172	100	421	613	14	79	163
2005	733	2890	242	2679	8277	200	1111	2283	317	5809	4410	109	595	860	76	527	1181	92	363	562	13	57	153
2006	770	3117	206	2643	8391	270	1315	2636	382	5790	4696	94	619	828	34	484	1137	84	384	559	16	141	403
2007	880	3059	212	2913	8809	213	1074	2610	414	5575	4777	182	805	1063	36	531	1267	85	339	487	16	59	128
2008	855	2779	274	2988	7932	184	1587	2648	282	5269	4305	170	965	1210	45	561	1145	111	305	449	13	54	239
2009	938	3118	283	3616	9145	193	1247	2662	466	6290	4772	192	1055	1345	53	615	1334	92	252	373	20	50	232
Total	12049	49552	3253	45228	140814	3016	18052	35661	5869	87343	72771	1225	8812	11314	886	10244	20026	1210	5874	8581	327	1318	3324
%	43.1	21.5	11.8	20.3	48.0	10.4	7.9	12.4	20.7	38.4	24.5	5.5	4.4	4.4	3.1	4.5	6.8	4.3	2.5	2.8	1.1	0.6	1.2

Table A6: Rate of road traffic fatalities per 100 accidents by region

Year	Greater Accra	Ashanti	Western	Eastern	Central	Volta	Northern	Upper East	Upper West	Brong-Ahafo	National
1991	3.4	17.9	8.9	15.7	18.1	25.1	27.7	17.8	20.0	19.5	11.0
1992	7.8	12.0	12.2	15.5	19.3	14.5	35.7	23.2	11.8	26.3	13.2
1993	5.3	15.2	14.0	19.3	19.8	15.4	50.0	10.9	30.8	40.5	13.9
1994	6.7	12.7	7.1	17.3	20.5	14.1	43.1	22.7	9.4	26.4	12.5
1995	5.2	14.3	14.4	20.8	23.7	19.0	27.2	16.2	15.3	25.1	12.3
1996	5.1	6.7	13.7	18.6	23.5	29.8	47.9	21.2	20.0	31.2	12.4
1997	4.1	10.3	14.5	17.2	19.1	16.2	26.9	14.6	12.0	20.2	10.2
1998	5.2	13.1	14.2	22.6	22.2	28.9	53.0	12.3	16.7	27.0	12.9
1999	5.0	14.6	12.6	24.8	21.7	19.8	44.4	19.5	28.2	21.0	14.1
2000	4.5	18.3	20.0	19.1	21.7	17.5	31.9	50.3	17.5	22.4	13.5
2001	4.8	22.6	19.7	20.0	21.6	25.6	29.3	19.7	19.8	30.8	14.7
2002	4.0	20.2	15.0	23.6	25.9	23.8	36.8	21.1	30.3	32.3	15.5
2003	5.6	19.7	18.3	19.0	20.7	29.4	68.0	23.6	54.7	24.9	16.1
2004	6.5	28.3	19.8	19.1	22.8	24.5	40.6	32.5	33.3	29.2	18.0
2005	6.3	18.8	25.9	20.7	20.0	21.5	43.3	43.6	36.6	29.3	15.7
2006	6.1	22.7	21.1	16.0	20.8	32.4	42.1	35.2	33.9	39.3	15.9
2007	6.9	23.4	26.4	20.8	26.8	29.3	41.2	50.7	37.0	38.3	17.0
2008	7.6	23.4	25.8	22.7	19.8	35.6	37.0	38.1	45.6	22.4	17.3
2009	7.7	23.8	17.2	25.6	26.8	25.3	51.4	45.4	65.6	37.4	18.2
Total	**107.8**	**338**	**320.8**	**378.4**	**414.8**	**447.7**	**777.5**	**518.6**	**538.5**	**543.5**	**274.4**
Mean	**5.67**	**17.79**	**16.88**	**19.92**	**21.83**	**23.56**	**40.92**	**27.29**	**28.34**	**28.61**	**14.44**

Table

A7: Road User Class by Fatality, Casualty and Fatality Index for 2010 &2011

	Persons Killed	Casualties	Fatality Index	Persons Killed	Casualties	Fatality Index	Persons Killed	Casualties	Fatality Index	Persons Killed	Casualties	Fatality Index	Persons Killed	Casualties	Fatality Index	Persons Killed	Casualties	Fatality Index
	Greater Accra						Northern						Brong Ahafo					
	2011			2010			2011			2010			2011			2010		
Pedestrian	245	1176	20.8	231	1388	16.6	17	41	41.5	20	40	50.0	82	231	35.5	57	153	37.3
Car Occupant	30	931	3.2	56	1032	5.4	2	58	3.4	13	76	17.1	45	344	13.1	33	311	10.6
Goods Veh. Ocpts	13	163	8.0	19	155	12.3	13	74	17.6	21	250	8.4	38	170	22.4	29	175	16.6
Bus/Mini-Bus	46	951	4.8	60	1008	6.0	57	357	16.0	15	263	5.7	52	444	11.7	21	464	4.5
Motorcyclist	49	389	12.6	31	460	6.7	20	75	26.7	13	75	17.3	57	214	26.6	20	104	19.2
Pick-Up Ocpts	7	110	6.4	8	139	5.8	7	66	10.6	22	90	24.4	3	56	5.4	1	59	1.7
Cyclist	16	62	25.8	18	98	18.4	6	14	42.9	7	11	63.6	17	42	40.5	5	25	20.0
Other	1	11	9.1	1	13	7.7	1	4	25.0	3	7	42.9	3	13	23.1	3	11	27.3
Total	407	3793	10.7	424	4293	9.9	474	3614	13.1	20	40	50.0	297	1514	19.6	169	1302	13.0
	Ashanti						Upper East						Upper West					
	2011			2010			2011			2010			2011			2010		
Pedestrian	211	776	27.2	193	753	25.6	15	24	62.5	13	26.8	23.0	6	16	37.5	11	17	64.7
Car Occupant	45	692	6.5	58	721	8.0	2	27	7.4	3	7.1	7.5	0	23	0.0	5	25	20.0
Goods Veh. Ocpts	38	286	13.3	32	289	11.1	1	12	8.3	4	14.0	11.6	1	8	12.5	2	7	28.6
Bus/Mini-Bus	120	1428	8.4	97	1587	6.1	1	21	4.8	1	7.8	5.6	8	45	17.8	3	42	7.1
Motorcyclist	40	267	15.0	38	183	20.8	26	67	38.8	13	19.9	15.6	32	96	33.3	25	57	43.9
Pick-Up Ocpts	8	113	7.1	15	164	9.1	2	31	6.5	0	7.4	7.9	0	13	0.0	1	27	3.7
Cyclist	8	39	20.5	11	33	33.3	7	15	46.7	11	28.6	30.6	3	9	33.3	7	8	87.5
Other	4	13	30.8	10	22	45.5	0	0	-	0	20.0	26.6	0	0	-	0	0	0.0
Total	474	3614	13.1	454	3752	12.1	54	197	27.4	13	13.6	11.7	50	210	23.8	11	17	64.7
	Central						Eastern						National Fatality Index					
	2011			2010			2011			2010			2011			2010		
Pedestrian	101	346	29.2	98	422	23.2	98	346	28.3	109	438	24.9	26.8			23.0		
Car Occupant	28	450	6.2	14	347	4.0	44	488	9.0	50	614	8.1	7.1			7.5		
Goods Veh. Ocpts	12	93	12.9	11	92	12.0	33	242	13.6	20	224	8.9	14.0			11.6		
Bus/Mini-Bus	44	502	8.8	29	575	5.0	37	1006	3.7	54	1017	5.3	7.8			5.6		
Motorcyclist	12	82	14.6	8	86	9.3	22	113	19.5	13	81	16.0	19.9			15.6		
Pick-Up Ocpts	2	48	4.2	2	51	3.9	6	102	5.9	3	73	4.1	7.4			7.9		
Cyclist	4	25	16.0	4	19	21.1	8	31	25.8	7	26	26.9	28.6			30.6		
Other	0	4	0.0	1	3	33.3	0	3	0.0	3	10	30.0	20.0			26.6		
Total	203	1550	13.1	167	1595	10.5	248	2331	10.6	259	2483	10.4	13.6			11.7		
	Volta						Western						National Fatality Index					
	2011			2010			2011			2010			2011			2010		
Pedestrian	45	128	35.2	50	194	25.8	78	262	29.8	71	280	25.4	26.8			23.0		
Car Occupant	14	194	7.2	19	222	8.6	41	315	13.0	20	269	7.4	7.1			7.5		
Goods Veh. Ocpts	7	68	10.3	8	74	10.8	12	88	13.6	13	85	15.3	14.0			11.6		
Bus/Mini-Bus	37	460	8.0	24	381	6.3	25	280	8.9	21	427	4.9	7.8			5.6		

Table

Motorcyclist	20	140	14.3	32	161	19.9	35	130	26.9	17	89	19.1	19.9			15.6		
Pick-Up Ocpts	9	34	26.5	0	53	0.0	2	50	4.0	3	35	8.6	7.4			7.9		
Cyclist	3	21	14.3	10	32	31.3	8	22	36.4	12	30	40.0	28.6			30.6		
Other	4	22	18.2	0	6	0.0	2	5	40.0	0	0	0.0	20.0			26.6		
Total	139	1067	13.0	50	194	25.8	203	1152	17.6	71	280	25.4	13.6			11.7		

A7: (Cont):Road User Class by Fatality, Casualty and Fatality Index for 2012 &2013

	Persons Killed	**Casualties**	**Fatality Index**	**Persons Killed**	**Casualties**	**Fatality Index**	**Persons Killed**	**Casualties**	**Fatality Index**	**Persons Killed**	**Casualties**	**Fatality Index**	**Persons Killed**	**Casualties**	**Fatality Index**	**Persons Killed**	**Casualties**	**Fatality Index**
	Greater Accra						**Northern**						**Brong Ahafo**					
	2013			**2012**			**2013**			**2012**			**2013**			**2012**		
Pedestrian	196	960	20.4	333	1441	23.1	26	52	50.0	5	14	35.7	64	167	38.3	66	178	37.1
Car Occupant	65	880	7.4	54	808	6.7	5	39	12.8	7	47	14.9	16	279	5.7	30	256	11.7
Goods Veh. Ocpts	13	74	17.6	15	139	10.8	27	218	12.4	10	79	12.7	25	128	19.5	17	117	14.5
Bus/Mini-Bus	28	675	4.1	55	986	5.6	52	292	17.8	42	215	19.5	38	266	14.3	55	496	11.1
Motorcyclist	43	411	10.5	56	512	10.9	20	103	19.4	26	71	36.6	42	207	20.3	40	199	20.1
Pick-Up Ocpts	7	69	10.1	8	114	7.0	2	46	4.3	2	26	7.7	1	44	2.3	3	30	10.0
Cyclist	8	60	13.3	14	75	18.7	7	10	70.0	7	12	58.3	5	23	21.7	6	25	24.0
Other	3	40	7.5	0	15	0.0	1	26	3.8	0	0		10	25	40.0	4	7	57.1
Total	363	3169	11.5	535	4090	13.1	140	786	17.8	99	464	21.3	201	1139	17.6	221	1308	16.9
	Ashanti						**Upper East**						**Upper West**					
	2013			**2012**			**2013**			**2012**			**2013**			**2012**		
Pedestrian	172	486	35.4	188	739	25.4	18	31	58.1	15	24	62.5	8	23	34.8	19	42	45.2
Car Occupant	56	463	12.1	39	581	6.7	1	15	6.7	2	27	7.4	1	20	5.0	0	18	0.0
Goods Veh. Ocpts	34	185	18.4	43	258	16.7	0	2	0.0	1	12	8.3	12	59	20.3	2	15	13.3
Bus/Mini-Bus	67	883	7.6	89	1323	6.7	0	2	0.0	1	21	4.8	8	41	19.5	13	60	21.7
Motorcyclist	54	222	24.3	56	266	21.1	31	95	32.6	26	67	38.8	39	125	31.2	24	90	26.7
Pick-Up Ocpts	2	59	3.4	5	73	6.8	2	8	25.0	2	31	6.5	2	21	9.5	1	11	9.1
Cyclist	7	23	30.4	10	49	20.4	5	24	20.8	7	15	46.7	2	6	33.3	12	21	57.1
Other	14	51	27.5	2	9	22.2	0	0	-	0	0	-	0	1	0.0	0	0	-
Total	406	2372	17.1	432	3298	13.1	57	177	32.2	54	197	27.4	72	296	24.3	71	257	27.6
	Central						**Eastern**						**National Fatality Index**					
	2013			**2012**			**2013**			**2012**			**2013**			**2012**		
Pedestrian	64	217	29.5	90	315	28.6	78	279	28.0	107	453	23.6	28.5			25.8		
Car Occupant	21	296	7.1	35	395	8.9	32	473	6.8	37	477	7.8	7.7			7.9		
Goods Veh. Ocpts	11	70	15.7	21	114	18.4	17	103	16.5	46	234	19.7	16.9			16.6		
Bus/Mini-Bus	70	307	22.8	33	415	8.0	33	705	4.7	89	910	9.8	9.1			8.6		
Motorcyclist	19	103	18.4	18	86	20.9	16	125	12.8	24	156	15.4	19.6			19.0		
Pick-Up Ocpts	6	40	15.0	6	52	11.5	3	51	5.9	4	81	4.9	7.3			7.6		
Cyclist	9	20	45.0	3	14	21.4	6	23	26.1	8	33	24.2	25.5			28.1		
Other	0	3	0.0	1	3	33.3	12	54	22.2	1	1	100	20.3			22.0		
Total	200	1056	18.9	207	1394	14.8	197	1813	10.9	316	2345	13.5	15.2			14.6		

Table

	Volta						Western						National Fatality Index	
	2013			2012			2013			2012			2013	2012
Pedestrian	58	166	34.9	39	126	31.0	42	165	25.5	63	250	25.2	28.5	25.8
Car Occupant	13	168	7.7	13	140	9.3	8	196	4.1	27	331	8.2	7.7	7.9
Goods Veh. Ocpts	5	74	6.8	3	22	13.6	30	114	26.3	22	95	23.2	16.9	16.6
Bus/Mini-Bus	23	322	7.1	25	211	11.8	10	136	7.4	25	301	8.3	9.1	8.6
Motorcyclist	35	153	22.9	15	146	10.3	24	104	23.1	48	160	30.0	19.6	19.0
Pick-Up Ocpts	1	28	3.6	1	19	5.3	3	31	11.0	5	50	10.0	7.3	7.6
Cyclist	3	16	18.8	0	5	0.0	3	11	8.0	11	29	37.9	25.5	28.1
Other	4	9	44.4	0	2	0.0	0	8	0.0	1	4	25.0	20.3	22.0
Total	142	936	15.2	96	671	14.3	120	765	15.7	202	1220	16.6	15.2	14.6

A8: Value of $y_i = \log(D/P)$ and $x_i = \log(N/P)$ from 1991 – 2009

i	Year	y_i	x_i	$x\,y_i\,i$	$xi2$	$yi2$	i	Year	y_i	x_i	$x\,y_i\,i$	$xi2$	$yi2$
1	1991	-9.69	-4.72	45.73	22.28	93.84	**11**	2001	-9.36	-3.53	33.03	12.44	87.66
2	1992	-9.72	-4.70	45.72	22.12	94.49	**12**	2002	-9.38	-3.48	32.61	12.08	88.06
3	1993	-9.76	-4.60	44.86	21.12	95.29	**13**	2003	-9.39	-3.46	32.50	11.98	88.15
4	1994	-9.88	-4.42	43.66	19.54	97.56	**14**	2004	-9.18	-3.40	31.20	11.57	84.17
5	1995	-9.69	-4.25	41.17	18.07	93.80	**15**	2005	-9.41	-3.34	31.45	11.17	88.56
6	1996	-9.69	-4.04	39.16	16.34	93.89	**16**	2006	-9.39	-3.28	30.78	10.74	88.24
7	1997	-9.75	-3.93	38.34	15.46	95.04	**17**	2007	-9.33	-3.21	29.95	10.32	86.96
8	1998	-9.44	-3.82	36.03	14.56	89.13	**18**	2008	-9.41	-3.22	30.27	10.36	88.45
9	1999	-9.61	-3.69	35.44	13.62	92.25	**19**	2009	-9.29	-3.16	29.32	9.96	86.28
10	2000	-9.48	-3.61	34.20	13.01	89.90		**Total**	**-180.84**	**-71.85**	**685.44**	**276.75**	**1721.70**

Table A9: Data for probability plot and residual analysis

i	$y(\,)i$	$\hat{y}i$	p_i	e_i	d_i	i	$y(\,)i$	$\hat{y}i$	p_i	e_i	d_i
1	-9.877	-9.721	3.247	-0.156	-1.574	11	-9.410	-9.377	55.195	-0.033	-0.330

Table

2	-9.761	-9.777	8.442	0.016	0.158	12	-9.405	-9.338	60.390	-0.067	-0.674
3	-9.749	-9.565	13.636	-0.184	-1.851	13	-9.394	-9.357	65.584	-0.037	-0.377
4	-9.720	-9.811	18.831	0.091	0.913	14	-9.389	-9.415	70.779	0.026	0.263
5	-9.689	-9.600	24.026	-0.089	-0.894	15	-9.384	-9.420	75.974	0.036	0.358
6	-9.687	-9.816	29.221	0.129	1.303	16	-9.362	-9.436	81.169	0.074	0.750
7	-9.685	-9.667	34.416	-0.018	-0.184	17	-9.325	-9.336	86.364	0.011	0.109
8	-9.605	-9.488	39.610	-0.117	-1.177	18	-9.289	-9.318	91.558	0.029	0.295
9	-9.481	-9.462	44.805	-0.019	-0.192	19	-9.175	-9.396	96.753	0.221	2.225
10	-9.441	-9.528	50.000	0.087	0.878						

Table A10: Distribution and Parameters for Monte Carlo Simulation

Exponential parameters	**LogNormal Parameters**
P = rlnorm(n,(1/20000000))	P = rlnorm(n,16.79254,0.173165)
N = rlnorm(n,(1/620000))	N = rlnorm(n,13.11504,0.727722)
D = rlnorm(n,(1/1600))	D = rlnorm(n,7.295927,0.346973)
Uniform Parameters runif(n, min = 14821000, max = 30726000) runif(n, min = 102051, max = 1728808) runif(n, min = 700, max = 2900)	**Gamma Distribution** P = rgamma(n,90,70)*10000000 N = rgamma(n,2,66)*10000000 D = rgamma(n,90,300)*100000

Table A11: Mean values of and for selected distributions and sample sizes

Count	Sample Size	Exponential		LogNormal		Uniform		Gamma	
		α	β	α	β	α	β	α	β
1	15	-6.063	0.972	-7.644	0.505	-7.427	0.613	-5.334	0.181
2	20	-6.028	0.981	-7.630	0.506	-7.434	0.612	-5.321	0.185
3	25	-6.032	0.984	-7.613	0.512	-7.414	0.618	-5.333	0.182
4	30	-6.018	0.986	-7.604	0.515	-7.430	0.613	-5.339	0.181
5	35	-6.033	0.984	-7.631	0.508	-7.417	0.619	-5.326	0.184
6	40	-6.056	0.978	-7.619	0.512	-7.433	0.610	-5.338	0.181
7	45	-6.091	0.971	-7.630	0.509	-7.427	0.612	-5.330	0.182
8	50	-6.084	0.969	-7.633	0.507	-7.448	0.607	-5.337	0.181
9	55	-6.038	0.978	-7.605	0.514	-7.439	0.610	-5.336	0.182
10	60	-6.000	0.993	-7.612	0.513	-7.422	0.618	-5.327	0.183
11	65	-6.044	0.975	-7.630	0.507	-7.456	0.605	-5.339	0.180
12	70	-6.054	0.977	-7.624	0.508	-7.453	0.607	-5.333	0.182
13	75	-6.039	0.981	-7.611	0.513	-7.426	0.615	-5.333	0.182
14	80	-6.026	0.985	-7.608	0.513	-7.431	0.614	-5.324	0.184
15	85	-6.076	0.973	-7.593	0.517	-7.428	0.614	-5.331	0.182
16	90	-6.026	0.979	-7.621	0.511	-7.433	0.611	-5.340	0.180
17	95	-6.069	0.973	-7.614	0.511	-7.430	0.613	-5.334	0.181
18	100	-6.055	0.981	-7.642	0.504	-7.435	0.612	-5.332	0.182
19	105	-6.026	0.984	-7.619	0.511	-7.426	0.614	-5.337	0.181
20	110	-6.075	0.973	-7.623	0.510	-7.465	0.603	-5.326	0.184
21	115	-6.010	0.987	-7.596	0.516	-7.438	0.612	-5.326	0.184
22	120	-6.043	0.982	-7.618	0.511	-7.431	0.612	-5.340	0.180
23	125	-6.045	0.975	-7.617	0.512	-7.423	0.616	-5.330	0.182
24	130	-6.065	0.975	-7.618	0.510	-7.436	0.610	-5.337	0.180
25	135	-6.037	0.984	-7.607	0.512	-7.437	0.612	-5.335	0.181
26	140	-6.051	0.977	-7.628	0.507	-7.438	0.611	-5.337	0.181
27	145	-6.031	0.981	-7.621	0.510	-7.420	0.615	-5.325	0.184

Table

28	150	-6.049	0.976	-7.611	0.514	-7.429	0.613	-5.329	0.183
29	155	-6.043	0.972	-7.636	0.507	-7.432	0.611	-5.327	0.183
30	160	-6.050	0.973	-7.611	0.513	-7.432	0.612	-5.329	0.183

Table A12: Expected percentage differences between the observed and expected road traffic fatalities given on Table 5.9

		Exponential				LogNormal				Uniform				Gamma			
		Sample Size				Sample Size				Sample Size				Sample Size			
Year	*D*	15	50	100	150	15	50	100	150	15	50	100	150	15	50	100	150
1991	920	62	62	63	62	29	29	28	29	47	47	47	47	3201	3204	3192	3199
1992	914	60	60	61	60	26	26	25	26	45	44	45	45	3323	3326	3314	3322
1993	901	54	54	55	54	19	18	18	19	38	38	39	38	3537	3540	3528	3536
1994	824	38	39	40	38	0	0	1	0	23	22	23	23	4117	4119	4106	4116
1995	1026	40	41	42	40	10	10	10	10	29	29	30	29	3486	3488	3478	3486
1996	1049	26	27	28	26	0	1	1	0	19	19	20	19	3642	3643	3634	3643
1997	1015	13	14	15	13	12	13	13	13	8	8	9	9	3952	3952	3944	3953
1998	1419	28	29	30	28	12	12	12	12	28	28	28	28	2940	2940	2934	2941
1999	1237	4	5	7	4	10	10	11	10	8	8	9	8	3565	3564	3558	3567
2000	1437	8	9	11	8	1	2	2	2	15	14	15	15	3189	3188	3183	3191
2001	1660	12	13	14	12	6	6	6	6	20	20	21	20	2863	2862	2858	2865
2002	1665	5	7	8	5	2	1	1	1	16	16	16	16	2956	2955	2951	2959
2003	1716	4	5	6	4	0	0	0	0	15	15	15	15	2978	2977	2973	2981
2004	2186	18	19	19	17	17	17	17	17	29	29	29	29	2412	2411	2408	2415
2005	1776	11	9	8	11	8	9	9	9	6	6	7	6	3114	3113	3110	3118
2006	1856	16	14	13	16	10	10	10	10	4	4	5	4	3098	3097	3094	3102
2007	2043	15	14	13	15	6	6	7	7	7	7	7	7	2922	2920	2918	2925
2008	1938	24	22	21	24	14	15	15	15	0	0	0	0	3169	3167	3165	3173
2009	2237	17	16	15	17	5	6	6	6	7	7	8	7	2844	2842	2840	2847
2010	1986	44	42	41	44	24	24	24	25	10	10	10	10	3271	3269	3267	3276
2011	2199	41	39	38	42	19	19	19	20	7	6	6	7	3085	3082	3081	3089
2012	2249	50	47	47	50	23	23	23	24	11	10	10	10	3124	3122	3121	3129

Table A13: The values of □ and □ for each of the 30 different sample sizes

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Sample Size	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85
□	-5.335	-5.334	-5.338	-5.337	-5.337	-5.335	-5.334	-5.332	-5.333	-5.336	-5.341	-5.333	-5.335	-5.331	-5.343
□	0.181	0.181	0.181	0.181	0.181	0.182	0.182	0.182	0.181	0.181	0.180	0.182	0.181	0.182	0.179
	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Sample Size	90	95	100	105	110	115	120	125	130	135	140	145	150	155	160
□	-5.325	-5.334	-5.330	-5.336	-5.342	-5.334	-5.323	-5.331	-5.329	-5.336	-5.344	-5.340	-5.337	-5.336	-5.336
□	0.183	0.182	0.182	0.181	0.179	0.182	0.184	0.182	0.183	0.181	0.179	0.180	0.181	0.181	0.181

A14:Regional distribution of the number of road traffic fatalities, registered vehicles and estimated population size from 1991 to 2009

	Greater Accra 1			Ashanti 2			Western 3			Eastern 4			Central 5		
Year	D_{i1}	N_{i1}	P_{i1}	D_{i2}	N_{i2}	P_{i2}	D_{i3}	N_{i3}	P_{i3}	D_{i4}	N_{i4}	P_{i4}	D_{i5}	N_{i5}	P_{i5}
1991	126	81382	1934520	183	21394	2641258	65	4485	1443424	183	3476	1852699	98	2226	1321216
1992	164	85027	2019639	153	22353	2731061	90	4686	1489614	204	3632	1878637	122	2326	1348961
1993	115	97240	2108503	168	25563	2823917	108	5359	1537282	207	4153	1904938	97	2660	1377289
1994	155	119066	2201277	161	31301	2919930	49	6562	1586475	186	5086	1931607	123	3257	1406212
1995	190	144805	2298133	174	38068	3019208	104	7981	1637242	192	6185	1958650	128	3961	1435743
1996	191	183331	2399251	175	48196	3121861	105	10104	1689634	196	7830	1986071	130	5014	1465893
1997	174	210101	2504818	220	55233	3228004	111	11580	1743702	181	8974	2013876	131	5747	1496677
1998	258	242341	2615030	283	63709	3337756	127	13356	1799500	291	10351	2042070	146	6628	1528107
1999	172	282373	2730091	178	74233	3451240	104	15563	1857084	294	12061	2070659	165	7723	1560198
2000	196	314963	2905726	280	82800	3612950	111	17359	1924577	295	13453	2106696	185	8615	1593823
2001	239	349917	2995804	350	91989	3710500	146	19285	1963069	296	14946	2150937	206	9571	1643232
2002	239	377880	3088673	351	99341	3810683	146	20827	2002330	297	16140	2196106	207	10336	1694172
2003	240	396783	3184422	360	104310	3913572	146	21868	2042377	298	16947	2242225	208	10853	1746691
2004	299	433482	3283139	565	113957	4019238	158	23891	2083224	325	18515	2289311	234	11857	1800838
2005	306	472736	3384917	314	124277	4127757	154	26054	2124889	299	20191	2337387	183	12930	1856664
2006	325	518494	3489849	340	136306	4239207	155	28576	2167386	305	22146	2386472	190	14182	1914221
2007	370	568681	3598034	376	149500	4353665	156	31342	2210734	305	24289	2436588	190	15555	1973562
2008	385	580546	3709574	416	152619	4471214	169	31996	2254949	294	24796	2487756	150	15879	2034742
2009	420	634779	3824570	440	166876	4591937	180	34985	2300048	320	27112	2539999	220	17362	2097819

	Volta 6			Northern 7			Upper East 8			Upper West 9			Brong-Ahafo 10		
Year	D_{i6}	N_{i6}	P_{i6}	D_{i7}	N_{i7}	P_{i7}	D_{i8}	N_{i8}	P_{i8}	D_{i9}	N_{i9}	P_{i9}	D_{i10}	N_{i10}	P_{i10}
1991	92	2008	1382575	41	5653	1412935	23	4037	834245	13	3651	513584	96	3738	1444102
1992	50	2098	1408844	30	5906	1452497	32	4218	843422	8	3814	525396	61	3906	1481648
1993	59	2399	1435612	17	6755	1493167	14	4824	852700	16	4362	537481	100	4467	1520171
1994	27	2938	1462888	31	8271	1534976	20	5907	862079	3	5341	549843	69	5469	1559695
1995	80	3573	1490683	38	10059	1577955	21	7184	871562	13	6496	562489	86	6652	1600248
1996	85	4524	1519006	40	12735	1622138	26	9095	881149	14	8224	575426	87	8422	1641854
1997	43	5184	1547867	35	14594	1667558	14	10423	890842	6	9425	588661	100	9651	1684542
1998	91	5980	1577277	61	16834	1714250	26	12023	900641	16	10871	602200	120	11132	1728340
1999	72	6968	1607245	76	19615	1762249	30	14009	910548	22	12667	616051	124	12971	1773277

Table

2000	89	7772	1635421	78	21878	1820806	48	15625	920089	25	14129	576583	130	14468	1815408
2001	135	8634	1676307	79	24306	1873609	34	17360	931130	26	15697	587538	149	16074	1857162
2002	135	9324	1718214	80	26249	1927944	34	18747	942304	26	16951	598701	150	17359	1899877
2003	140	9791	1761170	90	27562	1983854	45	19685	953611	35	17799	610077	154	18227	1943574
2004	167	10696	1805199	131	30111	2041386	68	21505	965055	37	19446	621668	202	19913	1988277
2005	122	11665	1850329	97	32838	2100586	79	23453	976635	30	21207	633480	192	21716	2034007
2006	169	12794	1896587	112	36016	2161503	82	25723	988355	34	23259	645516	244	23818	2080789
2007	170	14032	1944002	113	39502	2224187	83	28213	1000215	35	25511	657781	245	26123	2128647
2008	179	14325	1992602	95	40327	2288688	59	28801	1012218	36	26043	670279	155	26668	2177606
2009	180	15663	2042417	113	44094	2355060	65	31492	1024364	40	28476	683014	259	29160	2227691

Table A15: Value of $y_{ij} = \ln(D_{ij}/P_{ij})$ and $x_{ij} = \ln(N/P_{ijij})$ from 1991 – 2009

	Greater Accra 1		Ashanti 2		Western 3		Eastern 4		Central 5		Volta 6		Northern 7		Upper East 8		Upper West 9		Brong Ahafo 10	
Year	$xi1$	$yi1$	$xi2$	$yi2$	$xi3$	$yi3$	$xi4$	$yi4$	$xi5$	$yi5$	$xi6$	$yi6$	$xi7$	$yi7$	$xi8$	$yi8$	$xi9$	$yi9$	$xi10$	$yi10$
1991	-3.17	-9.64	-4.82	-9.58	-5.77	-10.01	-6.28	-9.22	-6.39	-9.51	-6.53	-9.62	-5.52	-10.45	-5.33	-10.50	-4.95	-10.58	-5.96	-9.62
1992	-3.17	-9.42	-4.81	-9.79	-5.76	-9.71	-6.25	-9.13	-6.36	-9.31	-6.51	-10.25	-5.51	-10.79	-5.30	-10.18	-4.93	-11.09	-5.94	-10.10
1993	-3.08	-9.82	-4.70	-9.73	-5.66	-9.56	-6.13	-9.13	-6.25	-9.56	-6.39	-10.10	-5.40	-11.38	-5.17	-11.02	-4.81	-10.42	-5.83	-9.63
1994	-2.92	-9.56	-4.54	-9.81	-5.49	-10.39	-5.94	-9.25	-6.07	-9.34	-6.21	-10.90	-5.22	-10.81	-4.98	-10.67	-4.63	-12.12	-5.65	-10.03
1995	-2.76	-9.40	-4.37	-9.76	-5.32	-9.66	-5.76	-9.23	-5.89	-9.33	-6.03	-9.83	-5.06	-10.63	-4.80	-10.63	-4.46	-10.68	-5.48	-9.83
1996	-2.57	-9.44	-4.17	-9.79	-5.12	-9.69	-5.54	-9.22	-5.68	-9.33	-5.82	-9.79	-4.85	-10.61	-4.57	-10.43	-4.25	-10.62	-5.27	-9.85
1997	-2.48	-9.57	-4.07	-9.59	-5.01	-9.66	-5.41	-9.32	-5.56	-9.34	-5.70	-10.49	-4.74	-10.77	-4.45	-11.06	-4.13	-11.49	-5.16	-9.73
1998	-2.38	-9.22	-3.96	-9.38	-4.90	-9.56	-5.28	-8.86	-5.44	-9.26	-5.58	-9.76	-4.62	-10.24	-4.32	-10.45	-4.01	-10.54	-5.05	-9.58
1999	-2.27	-9.67	-3.84	-9.87	-4.78	-9.79	-5.15	-8.86	-5.31	-9.15	-5.44	-10.01	-4.50	-10.05	-4.17	-10.32	-3.88	-10.24	-4.92	-9.57
2000	-2.22	-9.60	-3.78	-9.47	-4.71	-9.76	-5.05	-8.87	-5.22	-9.06	-5.35	-9.82	-4.42	-10.06	-4.08	-9.86	-3.71	-10.05	-4.83	-9.54
2001	-2.15	-9.44	-3.70	-9.27	-4.62	-9.51	-4.97	-8.89	-5.15	-8.98	-5.27	-9.43	-4.34	-10.07	-3.98	-10.22	-3.62	-10.03	-4.75	-9.43

2002	-2.10	-9.47	-3.65	-9.29	-4.57	-9.53	-4.91	-8.91	-5.10	-9.01	-5.22	-9.45	-4.30	-10.09	-3.92	-10.23	-3.56	-10.04	-4.70	-9.45
2003	-2.08	-9.49	-3.62	-9.29	-4.54	-9.55	-4.89	-8.93	-5.08	-9.04	-5.19	-9.44	-4.28	-10.00	-3.88	-9.96	-3.53	-9.77	-4.67	-9.44
2004	-2.02	-9.30	-3.56	-8.87	-4.47	-9.49	-4.82	-8.86	-5.02	-8.95	-5.13	-9.29	-4.22	-9.65	-3.80	-9.56	-3.46	-9.73	-4.60	-9.19
2005	-1.97	-9.31	-3.50	-9.48	-4.40	-9.53	-4.75	-8.96	-4.97	-9.22	-5.07	-9.63	-4.16	-9.98	-3.73	-9.42	-3.40	-9.96	-4.54	-9.27
2006	-1.91	-9.28	-3.44	-9.43	-4.33	-9.55	-4.68	-8.97	-4.91	-9.22	-5.00	-9.33	-4.09	-9.87	-3.65	-9.40	-3.32	-9.85	-4.47	-9.05
2007	-1.84	-9.18	-3.37	-9.36	-4.26	-9.56	-4.61	-8.99	-4.84	-9.25	-4.93	-9.34	-4.03	-9.89	-3.57	-9.40	-3.25	-9.84	-4.40	-9.07
2008	-1.85	-9.17	-3.38	-9.28	-4.26	-9.50	-4.61	-9.04	-4.85	-9.52	-4.94	-9.32	-4.04	-10.09	-3.56	-9.75	-3.25	-9.83	-4.40	-9.55
2009	-1.80	-9.12	-3.31	-9.25	-4.19	-9.46	-4.54	-8.98	-4.79	-9.16	-4.87	-9.34	-3.98	-9.94	-3.48	-9.67	-3.18	-9.75	-4.34	-9.06

A16: Regional distribution of the estimated road traffic fatalities, from 1991 to 2012

	Regions	*x*	*y*	G.*x*	G.*y*	Year		Regions	*x*	*y*	G.*x*	G.*y*	Year
1	Ashanti	-4.82	-9.58	-3.926	-9.489	1991	49	Central	-5.15	-8.98	-5.415	-9.239	2001
2	Ashanti	-4.81	-9.79	-3.926	-9.489	1992	50	Central	-5.1	-9.01	-5.415	-9.239	2002
3	Ashanti	-4.7	-9.73	-3.926	-9.489	1993	51	Central	-5.08	-9.04	-5.415	-9.239	2003
4	Ashanti	-4.54	-9.81	-3.926	-9.489	1994	52	Central	-5.02	-8.95	-5.415	-9.239	2004
5	Ashanti	-4.37	-9.76	-3.926	-9.489	1995	53	Central	-4.97	-9.22	-5.415	-9.239	2005
6	Ashanti	-4.17	-9.79	-3.926	-9.489	1996	54	Central	-4.91	-9.22	-5.415	-9.239	2006
7	Ashanti	-4.07	-9.59	-3.926	-9.489	1997	55	Central	-4.84	-9.25	-5.415	-9.239	2007
8	Ashanti	-3.96	-9.38	-3.926	-9.489	1998	56	Central	-4.85	-9.52	-5.415	-9.239	2008
9	Ashanti	-3.84	-9.87	-3.926	-9.489	1999	57	Central	-4.79	-9.16	-5.415	-9.239	2009
10	Ashanti	-3.78	-9.47	-3.926	-9.489	2000	58	Eastern	-6.28	-9.22	-5.241	-9.033	1991
11	Ashanti	-3.7	-9.27	-3.926	-9.489	2001	59	Eastern	-6.25	-9.13	-5.241	-9.033	1992
12	Ashanti	-3.65	-9.29	-3.926	-9.489	2002	60	Eastern	-6.13	-9.13	-5.241	-9.033	1993
13	Ashanti	-3.62	-9.29	-3.926	-9.489	2003	61	Eastern	-5.94	-9.25	-5.241	-9.033	1994

Table

14	Ashanti	-3.56	-8.87	-3.926	-9.489	2004	62	Eastern	-5.76	-9.23	-5.241	-9.033	1995
15	Ashanti	-3.5	-9.48	-3.926	-9.489	2005	63	Eastern	-5.54	-9.22	-5.241	-9.033	1996
16	Ashanti	-3.44	-9.43	-3.926	-9.489	2006	64	Eastern	-5.41	-9.32	-5.241	-9.033	1997
17	Ashanti	-3.37	-9.36	-3.926	-9.489	2007	65	Eastern	-5.28	-8.86	-5.241	-9.033	1998
18	Ashanti	-3.38	-9.28	-3.926	-9.489	2008	66	Eastern	-5.15	-8.86	-5.241	-9.033	1999
19	Ashanti	-3.31	-9.25	-3.926	-9.489	2009	67	Eastern	-5.05	-8.87	-5.241	-9.033	2000
20	Brong Ahafo	-5.96	-9.62	-4.998	-9.526	1991	68	Eastern	-4.97	-8.89	-5.241	-9.033	2001
21	Brong Ahafo	-5.94	-10.1	-4.998	-9.526	1992	69	Eastern	-4.91	-8.91	-5.241	-9.033	2002
22	Brong Ahafo	-5.83	-9.63	-4.998	-9.526	1993	70	Eastern	-4.89	-8.93	-5.241	-9.033	2003
23	Brong Ahafo	-5.65	-10.03	-4.998	-9.526	1994	71	Eastern	-4.82	-8.86	-5.241	-9.033	2004
24	Brong Ahafo	-5.48	-9.83	-4.998	-9.526	1995	72	Eastern	-4.75	-8.96	-5.241	-9.033	2005
25	Brong Ahafo	-5.27	-9.85	-4.998	-9.526	1996	73	Eastern	-4.68	-8.97	-5.241	-9.033	2006
26	Brong Ahafo	-5.16	-9.73	-4.998	-9.526	1997	74	Eastern	-4.61	-8.99	-5.241	-9.033	2007
27	Brong Ahafo	-5.05	-9.58	-4.998	-9.526	1998	75	Eastern	-4.61	-9.04	-5.241	-9.033	2008
28	Brong Ahafo	-4.92	-9.57	-4.998	-9.526	1999	76	Eastern	-4.54	-8.98	-5.241	-9.033	2009
29	Brong Ahafo	-4.83	-9.54	-4.998	-9.526	2000	77	Greater Accra	-3.17	-9.64	-2.355	-9.426	1991
30	Brong Ahafo	-4.75	-9.43	-4.998	-9.526	2001	78	Greater Accra	-3.17	-9.42	-2.355	-9.426	1992
31	Brong Ahafo	-4.7	-9.45	-4.998	-9.526	2002	79	Greater Accra	-3.08	-9.82	-2.355	-9.426	1993
32	Brong Ahafo	-4.67	-9.44	-4.998	-9.526	2003	80	Greater Accra	-2.92	-9.56	-2.355	-9.426	1994
33	Brong Ahafo	-4.6	-9.19	-4.998	-9.526	2004	81	Greater Accra	-2.76	-9.4	-2.355	-9.426	1995
34	Brong Ahafo	-4.54	-9.27	-4.998	-9.526	2005	82	Greater Accra	-2.57	-9.44	-2.355	-9.426	1996
35	Brong Ahafo	-4.47	-9.05	-4.998	-9.526	2006	83	Greater Accra	-2.48	-9.57	-2.355	-9.426	1997
36	Brong Ahafo	-4.4	-9.07	-4.998	-9.526	2007	84	Greater Accra	-2.38	-9.22	-2.355	-9.426	1998
37	Brong Ahafo	-4.4	-9.55	-4.998	-9.526	2008	85	Greater Accra	-2.27	-9.67	-2.355	-9.426	1999
38	Brong Ahafo	-4.34	-9.06	-4.998	-9.526	2009	86	Greater Accra	-2.22	-9.6	-2.355	-9.426	2000
39	Central	-6.39	-9.51	-5.415	-9.239	1991	87	Greater Accra	-2.15	-9.44	-2.355	-9.426	2001
40	Central	-6.36	-9.31	-5.415	-9.239	1992	88	Greater Accra	-2.1	-9.47	-2.355	-9.426	2002
41	Central	-6.25	-9.56	-5.415	-9.239	1993	89	Greater Accra	-2.08	-9.49	-2.355	-9.426	2003
42	Central	-6.07	-9.34	-5.415	-9.239	1994	90	Greater Accra	-2.02	-9.3	-2.355	-9.426	2004
43	Central	-5.89	-9.33	-5.415	-9.239	1995	91	Greater Accra	-1.97	-9.31	-2.355	-9.426	2005
44	Central	-5.68	-9.33	-5.415	-9.239	1996	92	Greater Accra	-1.91	-9.28	-2.355	-9.426	2006
45	Central	-5.56	-9.34	-5.415	-9.239	1997	93	Greater Accra	-1.84	-9.18	-2.355	-9.426	2007
46	Central	-5.44	-9.26	-5.415	-9.239	1998	94	Greater Accra	-1.85	-9.17	-2.355	-9.426	2008
47	Central	-5.31	-9.15	-5.415	-9.239	1999	95	Greater Accra	-1.8	-9.12	-2.355	-9.426	2009
48	Central	-5.22	-9.06	-5.415	-9.239	2000							

Table *A*16 (Cont.): Regional distribution of the estimated road traffic fatalities, from 1991 to 2012

	Regions	*x*	*y*	G.*x*	G.*y*	Year		Regions	*x*	*y*	G.*x*	G.*y*	Year
96	Northern	-5.52	-10.45	-4.594	-10.283	1991	144	Upper West	-3.62	-10.03	-3.912	-10.349	2001
97	Northern	-5.51	-10.79	-4.594	-10.283	1992	145	Upper West	-3.56	-10.04	-3.912	-10.349	2002
98	Northern	-5.4	-11.38	-4.594	-10.283	1993	146	Upper West	-3.53	-9.77	-3.912	-10.349	2003
99	Northern	-5.22	-10.81	-4.594	-10.283	1994	147	Upper West	-3.46	-9.73	-3.912	-10.349	2004
100	Northern	-5.06	-10.63	-4.594	-10.283	1995	148	Upper West	-3.4	-9.96	-3.912	-10.349	2005
101	Northern	-4.85	-10.61	-4.594	-10.283	1996	149	Upper West	-3.32	-9.85	-3.912	-10.349	2006
102	Northern	-4.74	-10.77	-4.594	-10.283	1997	150	Upper West	-3.25	-9.84	-3.912	-10.349	2007

103	Northern	-4.62	-10.24	-4.594	-10.283	1998	151	Upper West	-3.25	-9.83	-3.912	-10.349	2008
104	Northern	-4.5	-10.05	-4.594	-10.283	1999	152	Upper West	-3.18	-9.75	-3.912	-10.349	2009
105	Northern	-4.42	-10.06	-4.594	-10.283	2000	153	Volta	-6.53	-9.62	-5.536	-9.744	1991
106	Northern	-4.34	-10.07	-4.594	-10.283	2001	154	Volta	-6.51	-10.25	-5.536	-9.744	1992
107	Northern	-4.3	-10.09	-4.594	-10.283	2002	155	Volta	-6.39	-10.1	-5.536	-9.744	1993
108	Northern	-4.28	-10	-4.594	-10.283	2003	156	Volta	-6.21	-10.9	-5.536	-9.744	1994
109	Northern	-4.22	-9.65	-4.594	-10.283	2004	157	Volta	-6.03	-9.83	-5.536	-9.744	1995
110	Northern	-4.16	-9.98	-4.594	-10.283	2005	158	Volta	-5.82	-9.79	-5.536	-9.744	1996
111	Northern	-4.09	-9.87	-4.594	-10.283	2006	159	Volta	-5.7	-10.49	-5.536	-9.744	1997
112	Northern	-4.03	-9.89	-4.594	-10.283	2007	160	Volta	-5.58	-9.76	-5.536	-9.744	1998
113	Northern	-4.04	-10.09	-4.594	-10.283	2008	161	Volta	-5.44	-10.01	-5.536	-9.744	1999
114	Northern	-3.98	-9.94	-4.594	-10.283	2009	162	Volta	-5.35	-9.82	-5.536	-9.744	2000
115	Upper East	-5.33	-10.5	-4.249	-10.144	1991	163	Volta	-5.27	-9.43	-5.536	-9.744	2001
116	Upper East	-5.3	-10.18	-4.249	-10.144	1992	164	Volta	-5.22	-9.45	-5.536	-9.744	2002
117	Upper East	-5.17	-11.02	-4.249	-10.144	1993	165	Volta	-5.19	-9.44	-5.536	-9.744	2003
118	Upper East	-4.98	-10.67	-4.249	-10.144	1994	166	Volta	-5.13	-9.29	-5.536	-9.744	2004
119	Upper East	-4.8	-10.63	-4.249	-10.144	1995	167	Volta	-5.07	-9.63	-5.536	-9.744	2005
120	Upper East	-4.57	-10.43	-4.249	-10.144	1996	168	Volta	-5	-9.33	-5.536	-9.744	2006
121	Upper East	-4.45	-11.06	-4.249	-10.144	1997	169	Volta	-4.93	-9.34	-5.536	-9.744	2007
122	Upper East	-4.32	-10.45	-4.249	-10.144	1998	170	Volta	-4.94	-9.32	-5.536	-9.744	2008
123	Upper East	-4.17	-10.32	-4.249	-10.144	1999	171	Volta	-4.87	-9.34	-5.536	-9.744	2009
124	Upper East	-4.08	-9.86	-4.249	-10.144	2000	172	Western	-5.77	-10.01	-4.851	-9.656	1991
125	Upper East	-3.98	-10.22	-4.249	-10.144	2001	173	Western	-5.76	-9.71	-4.851	-9.656	1992
126	Upper East	-3.92	-10.23	-4.249	-10.144	2002	174	Western	-5.66	-9.56	-4.851	-9.656	1993
127	Upper East	-3.88	-9.96	-4.249	-10.144	2003	175	Western	-5.49	-10.39	-4.851	-9.656	1994
128	Upper East	-3.8	-9.56	-4.249	-10.144	2004	176	Western	-5.32	-9.66	-4.851	-9.656	1995
129	Upper East	-3.73	-9.42	-4.249	-10.144	2005	177	Western	-5.12	-9.69	-4.851	-9.656	1996
130	Upper East	-3.65	-9.4	-4.249	-10.144	2006	178	Western	-5.01	-9.66	-4.851	-9.656	1997
131	Upper East	-3.57	-9.4	-4.249	-10.144	2007	179	Western	-4.9	-9.56	-4.851	-9.656	1998
132	Upper East	-3.56	-9.75	-4.249	-10.144	2008	180	Western	-4.78	-9.79	-4.851	-9.656	1999
133	Upper East	-3.48	-9.67	-4.249	-10.144	2009	181	Western	-4.71	-9.76	-4.851	-9.656	2000
134	Upper West	-4.95	-10.58	-3.912	-10.349	1991	182	Western	-4.62	-9.51	-4.851	-9.656	2001
135	Upper West	-4.93	-11.09	-3.912	-10.349	1992	183	Western	-4.57	-9.53	-4.851	-9.656	2002
136	Upper West	-4.81	-10.42	-3.912	-10.349	1993	184	Western	-4.54	-9.55	-4.851	-9.656	2003
137	Upper West	-4.63	-12.12	-3.912	-10.349	1994	185	Western	-4.47	-9.49	-4.851	-9.656	2004
138	Upper West	-4.46	-10.68	-3.912	-10.349	1995	186	Western	-4.4	-9.53	-4.851	-9.656	2005
139	Upper West	-4.25	-10.62	-3.912	-10.349	1996	187	Western	-4.33	-9.55	-4.851	-9.656	2006
140	Upper West	-4.13	-11.49	-3.912	-10.349	1997	188	Western	-4.26	-9.56	-4.851	-9.656	2007
141	Upper West	-4.01	-10.54	-3.912	-10.349	1998	189	Western	-4.26	-9.5	-4.851	-9.656	2008
142	Upper West	-3.88	-10.24	-3.912	-10.349	1999	190	Western	-4.19	-9.46	-4.851	-9.656	2009
143	Upper West	-3.71	-10.05	-3.912	-10.349	2000							

A17: Computation of the *W* test statistics

i	y_i	y_{20-i}	$y_{20-i} - y_i$	a_{20-i}	$a_{20-i}(y_{20-i} - y_i)$
1	-9.8166	-9.1167	0.6999	0.4808	0.3365
2	-9.6724	-9.1732	0.4992	0.3232	0.1613

Table

3	-9.6391	-9.1824	0.4567	0.2561	0.1170
4	-9.6041	-9.2238	0.3803	0.2059	0.0783
5	-9.5747	-9.2815	0.2932	0.1641	0.0481
6	-9.5611	-9.3039	0.2572	0.1271	0.0327
7	-9.4931	-9.3113	0.1818	0.0932	0.0169
8	-9.4668	-9.4006	0.0662	0.0612	0.0041
9	-9.4384	-9.4186	0.0198	0.0303	0.0006
10	-9.4363	-9.4363	0.0000	0.0000	0.0000
					b □ 0.7955

Table *A*18: Coefficients □$^{a}_{n\ i\square\square1}$□ for *W* test for normality

n / *i*	2	3	4	5	6	7	8	9	10
1	0.7071	0.7071	0.6872	0.6646	0.6431	0.6233	0.6052	0.5888	0.5739
2		0.0000	0.1677	0.2413	0.2806	0.3031	0.3164	0.3244	0.3291
3				0.0000	0.0875	0.1401	0.1743	0.1976	0.2141
4						0.0000	0.0561	0.0947	0.1224
5								0.0000	0.0399

Table *A*18 (Cont.): Coefficients □$^{a}_{n\ i\square\square1}$□ for *W* test for normality	***n***	**11**	**12**	**12**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
i											
1		0.5601	0.5475	0.5359	0.5259	0.5150	0.5056	0.4968	0.4886	0.4808	0.4734
2		0.3315	0.3325	0.3325	0.3318	0.3306	0.3290	0.3273	0.3253	0.3232	0.3211
3		0.2260	0.2347	0.2412	0.2460	0.2495	0.2521	0.2540	0.2553	0.2561	0.2565
4		0.1429	0.1586	0.1707	0.1802	0.1878	0.1939	0.1988	0.2027	0.2059	0.2085

	0.1099	0.1240	0.1353	0.1447	0.1524	0.1587	0.1641	
	0.0539	0.0727	0.0880	0.1005	0.1109	0.1197	0.1271	
5					0.0695	0.0922	0.1686	
6					0.0000	0.0303	0.1334	
7					0.0000	0.0240	0.0433	0.0593
	0.0725	0.0837	0.0932	0.1013				
8					0.0000	0.0196	0.0359	0.0496
	0.0612	0.0711						
9					0.0000	0.0163	0.0303	0.0422
10					0.0000	0.0140		

Table *A*18 (Cont.): Coefficients $a_{n\,i}$ for *W* test for normality

n \ *i*	21	22	23	24	25	26	27	28	29	30
1	0.4643	0.4590	0.4542	0.4493	0.4450	0.4407	0.4366	0.4328	0.4291	0.4254
2	0.3185	0.3156	0.3126	0.3098	0.3069	0.3043	0.3018	0.2992	0.2968	0.2944
3	0.2578	0.2571	0.2563	0.2554	0.2543	0.2533	0.2522	0.2410	0.2499	0.2487
4	0.2119	0.2131	0.2139	0.2145	0.2148	0.2151	0.2152	0.2151	0.2150	0.2148
5	0.1736	0.1764	0.1787	0.1807	0.1822	0.1836	0.1848	0.1857	0.1864	0.1870
6	0.1399	0.1443	0.1480	0.1512	0.1539	0.1563	0.1584	0.1601	0.1616	0.1630
7	0.1092	0.1150	0.1201	0.1245	0.1283	0.1316	0.1346	0.1372	0.1392	0.1415
8	0.0804	0.0878	0.0941	0.0997	0.1046	0.1086	0.1128	0.1162	0.1192	0.1219

Table

9			0.0530	0.0618	0.0696	0.0764
			0.0823	0.0876	0.0923	0.1965
			0.1002	0.1036		
10	0.0459		0.0263	0.0368	0.0523	0.0610
			0.0672	0.0728	0.0862	
	0.0228			0.0598	0.0650	
11			0.0000	0.0122	0.0321	0.0403
			0.0476	0.0540	0.0697	
12			0.0000	0.0107	0.0200	0.0284
			0.0358	0.0424	0.0483	0.0537
13			0.0000	0.0094	0.0178	0.0253
			0.0320	0.0381		
14			0.0000	0.0084	0.0159	0.0227
15			0.0000	0.0076		

Coefficients $a_{n,i}$ for *W* test for normality — ***n* 31 32 33 34 35 36 37 38 39 40**

i										
1									0.4220	0.4188
									0.4156	0.4127
									0.4096	0.4068
									0.4040	0.4015
									0.3989	0.3964
2									0.2921	0.2898
									0.2876	0.2854
									0.2834	0.2813
									0.2794	0.2774
									0.2755	0.2737
3									0.2475	0.2463
									0.2451	0.2439
									0.2427	0.2415
									0.2403	0.2391
									0.2380	0.2368
4									0.2145	0.2141
									0.2137	0.2132
									0.2127	0.2121
									0.2116	0.2110
									0.2104	0.2098
5									0.1874	0.1878
									0.1882	0.1883
	0.1880								0.1881	0.1880
	0.1660								0.1686	0.1689

					0.1883 0.1883 0.1878
6					0.1641 0.1651 0.1667 0.1673 0.1678 0.1683 0.1691
7					0.1433 0.1449 0.1463 0.1475 0.1487 0.1496 0.1505 0.1513 0.1520 0.1526
8					0.1243 0.1265 0.1284 0.1301 0.1317 0.1331 0.1344 0.1356 0.1366 0.1376
9					0.1066 0.1093 0.1118 0.1140 0.1160 0.1179 0.1196 0.1211 1225 0.1237
10	0.0981		56	0.1075	0.0892 0.0931 0.0988 0.1013
	0.0844			0.0947	0.0967 0.1036 0.10 0.1108
11					0.0739 0.0777 0.0844 0.0873 0.0900 0.0924 0.0986
12					0.0585 0.0629 0.0669 0.0706 0.0739 0.0770 0.0798 0.0824 0.0848 0.0870
13					0.0435 0.0485 0.0530 0.0572 0.0610 0.0645

Table

			0.0677 0.0706
			0.0733 0.0759
14			0.0289 0.0344
			0.0395 0.0441
			0.0484 0.0523
			0.0559 0.0592
			0.0622 0.0651
15		0.0481	0.0545 0.0206
		0.0372	0.0409 0.0314
			0.0361 0.0404
			0.0444 0.0546
16			0.0000 0.0068
			0.0131 0.0187
			0.0239 0.0287
			0.0331 0.0444
17			0.0000 0.0062
			0.0119 0.0172
			0.0220 0.0264
			0.0305 0.0343
18			0.0000 0.0057
			0.0110 0.0158
			0.0203 0.0244
19			0.0000 0.0053
			0.0101 0.0146
20			0.0000 0.0049

Table

i	**A18 (Cont.): Coefficients $a_{n\,i}$ for *W* test for normality**										***n*** **41 42 43 44 45 46 47 48 49 50**
1	0.3940	0.3917	0.3894	0.3872	0.3850	0.3830	0.3808	0.3789	0.3770	0.3751	
2	0.2719	0.2701	0.2684	0.2667	0.2651	0.2635	0.2620	0.2604	0.2589	0.2574	
3	0.2357	0.2345	0.2334	0.2323	0.2313	0.2302	0.2291	0.2281	0.2271	0.2260	
4	0.2091	0.2085	0.2078	0.2072	0.2065	0.2058	0.2052	0.2045	0.2038	0.2032	
5	0.1876	0.1874	0.1868	0.1865	0.1862	0.1859	0.1847				
6	0.1693	0.1694	0.1695	0.1695	0.1695	0.1695	0.1691				
7	0.1531	0.1535	0.1539	0.1542	0.1545	0.1548	0.1550	0.1551	0.1553	0.1554	
8	0.1384	0.1392	0.1398	0.1405	0.1410	0.1415	0.1420	0.1423	0.1427	0.1430	
9	0.1249	0.1259	0.1269	0.1278	0.1286	0.1293	0.1300	0.1306	0.1312	0.1317	
10	0.1123	0.1136	0.1160	0.1170	0.1180	0.1189	0.1212				
11	0.1004	0.1020	0.1049	0.1062	0.1073	0.1085	0.1113				
	0.1879		0.1897	0.1255							
12	0.1695		0.1693	0.1692	0.0891	0.0909	0.0927	0.0943	0.0959	0.0972	0.0986 0.0998 0.1010 0.1020

Table

13					0.0782	0.0804	0.0824	0.0842	0.0860 0.0876 0.0892 0.0906 0.0919 0.0932
14					0.0677	0.0701	0.0724	0.0745	0.0765 0.0783 0.0801 0.0817 0.0832 0.0846
15			0.0628		0.0575	0.0621	0.0663	0.0701	0.0736 0.0768 0.0798 0.0825 0.0851 0.0875 0.0898 0.0919 0.0939 0.0958 0.0976 0.0993
16			0.0534		0.0476	0.0506	0.0560	0.0584	0.0607 0.0648 0.0685
17					0.0379	0.0411	0.0442	0.0471	0.0497 0.0522 0.0546 0.0568 0.0588 0.0608
18					0.0283	0.0318	0.0352	0.0383	0.0412 0.0439 0.0465 0.0489 0.0511 0.0532
19					0.0188	0.0227	0.0263	0.0296	0.0328 0.0357 0.0385 0.0411 0.0436 0.0459
20					0.0094	0.0136	0.0176	0.0211	0.0245 0.0277 0.0307 0.0336 0.0363 0.0388 0.0412 0.0435
21					0.0000	0.0045	0.0087	0.0126	0.0163 0.0197 0.0229 0.0259 0.0288 0.0314 0.0339 0.0362
22					0.0000	0.0042	0.0081	0.0118	0.0153 0.0185 0.0215 0.0244 0.0271 0.0297 0.0321 0.0344
23					0.0000	0.0039	0.0076	0.0111	0.0143 0.0174 0.0203 0.0230 0.0256 0.0280 0.0303 0.0325
24					0.0000	0.0037	0.0071	0.0104	0.0135 0.0164 0.0192 0.0218 0.0243 0.0267 0.0289 0.0310
25					0.0000	0.0035			

Percentage points of the *W* test* for *n* = 3(1)50 Level *A19:*

	Level								
n	**0.01**	**0.02**	**0.05**	**0.10**	**0.50**	**0.90**	**0.95**	**0.98**	**0.99**
3	0.753	0.756	0.767	0.789	0.959	0.998	0.999	1.000	1.000
4	0.687	0.707	0.748	0.792	0.935	0.987	0.992	0.996	0.997
5	0.686	0.715	0.762	0.806	0.927	0.979			0.993

Table

6	0.713	0.748	0.788	0.826	0.927	0.974	0.986	0.991	0.989
							0.981	0.086	
7	0.730	0.760	0.803	0.838	0.928	0.972	0.979	0.985	0.988
8	0.749	0.778	0.818	0.851	0.932	0.972	0.978	0.984	0.987
9	0.764	0.791	0.829	0.859	0.935	0.972	0.978	0.984	0.986
10	0.781	0.806	0.842	0.869	0.938	0.972	0.978	0.983	0.986
11	0.792	0.817	0.850	0.876	0.940	0.973	0.979	0.984	0.986
12	0.805	0.828	0.859	0.883	0.943	0.973	0.979	0.984	0.986
13	0.814	0.837	0.866	0.889	0.945	0.974	0.979	0.984	0.986
14	0.825	0.846	0.874	0.895	0.947	0.975	0.980	0.984	0.986
15	0.835	0.855	0.881	0.901	0.950	0.975	0.980	0.984	0.987
16	0.844	0.863	0.887	0.906	0.952	0.976	0.981	0.985	0.987
17	0.851	0.869	0.892	0.910	0.954	0.977	0.981	0.985	0.987
18	0.858	0.874	0.897	0.914	0.956	0.978	0.982	0.986	0.988
19	0.863	0.879	0.901	0.917	0.957	0.978	0.982	0.986	0.988
20	0.868	0.884	0.905	0.920	0.959	0.979	0.983	0.986	0.988
21	0.873	0.888	0.908	0.923	0.960	0.980	0.983	0.987	0.989
22	0.878	0.892	0.911	0.926	0.961	0.980	0.984	0.987	0.989
23	0.881	0.895	0.914	0.928	0.962	0.981	0.984	0.987	0.989
24	0.884	0.898	0.916	0.930	0.963	0.981	0.984	0.987	0.989
25	0.888	0.901	0.918	0.931	0.964	0.981	0.985	0.988	0.989
26	0.891	0.904	0.920	0.933	0.965	0.982	0.985	0.988	0.989
27	0.894	0.906	0.923	0.935	0.965	0.982	0.985	0.988	0.990
28	0.896	0.908	0.924	0.936	0.966	0.982	0.985	0.988	0.990
29	0.898	0.910	0.926	0.937	0.966	0.982	0.985	0.988	0.990
30	0.900	0.912	0.927	0.939	0.967	0.983	0.985	0.988	0.900

***A*19 (Cont.): Percentage points of the *W* test* for *n* = 3(1)50 Level**

	Level		
n	**0.01**	**0.95**	**0.99**

Table								0.98	
31	0.902	**0.02**	**0.05**	**0.10**	**0.50**	**0.90**	0.986	0.988	0.990
		0.914	0.929	0.940	0.967	0.983			
32	0.904	0.915	0.930	0.941	0.968	0.983	0.986	0.988	0.990
33	0.906	0.917	0.931	0.942	0.968	0.983	0.986	0.989	0.990
34	0.908	0.919	0.933	0.943	0.969	0.983	0.986	0.989	0.990
35	0.910	0.920	0.934	0.944	0.969	0.984	0.986	0.989	0.990
36	0.912	0.922	0.935	0.945	0.970	0.984	0.986	0.989	0.990
37	0.914	0.924	0.936	0.946	0.970	0.984	0.987	0.989	0.990
38	0.916	0.925	0.938	0.947	0.971	0.984	0.987	0.989	0.990
39	0.917	0.927	0.939	0.948	0.971	0.984	0.987	0.989	0.991
40	0.919	0.928	0.940	0.949	0.972	0.985	0.987	0.989	0.991
41	0.920	0.929	0.941	0.950	0.972	0.985	0.987	0.989	0.991
42	0.922	0.930	0.942	0.951	0.972	0.985	0.987	0.989	0.991
43	0.923	0.932	0.943	0.951	0.973	0.985	0.987	0.990	0.991
44	0.924	0.933	0.944	0.952	0.973	0.985	0.987	0.990	0.991
45	0.926	0.934	0.945	0.953	0.973	0.985	0.988	0.990	0.991
46	0.927	0.935	0.945	0.953	0.974	0.985	0.988	0.990	0.991
47	0.928	0.936	0.946	0.954	0.974	0.985	0.988	0.990	0.991
48	0.929	0.937	0.947	0.954	0.974	0.985	0.988	0.990	0.991
49	0.929	0.937	0.947	0.955	0.974	0.985	0.988	0.990	0.991
50	0.930	0.938	0.947	0.955	0.974	0.985	0.988	0.990	0.991

Listing A1: Matlab code for function to evaluate the posterior distribution

```
1. function y = post(alpha,beta,sigma,sigma1,sigma2,mu1,mu2,rho,x1,y1)

2. y=exp(f-0.5*((sum(y1-alpha-beta*x1)^2)/sigma^2+(((alpha-mu1)/sigma1)^2
```


Listing A2: Implementation of component-wise Metropolis sampler for the posterior distribution

```
1.   % % Metropolis procedure to sample from the posterior distribution
2.   % Component-wise updating. Use a normal proposal distribution  3. %
% Initialize the Metropolis sampler 4.

5.6. T=5000;  propsigma=[0.02,0.004]; % Set the maximum number of iteration%
standard deviation of  proposal distribution
7.  thetamin=[-10,0];  % define minimum for alpha and beta

8.  thetamax=[seed=1;rand(-4,1];  'state'% define maximum for alpha and
    beta,seed);randn('state',seed);  % set the random seed
9.
10. state=zeros(2,T);   % storage space for the state of the sampler
11. alpha=unifrnd(thetamin(1),thetamax(1)); % Start value for alpha 12.
    beta=unifrnd(thetamin(2),thetamax(2)); % Start value for beta 13.

14.  t=1;  state(1,t)=alpha;  % initialize
iteration at 1% save the current state
15.  state(2,t)=beta;  16.  % % Start sampling 17.
18.   while t<T;  % Iterate until we have T samples
19.   t=t+1;
20.   % % Propose a new value for alpha
21.   new_alpha=normrnd(alpha,propsigma(1));
22.   pratio=post(new_alpha,beta,sigma,sigma1,sigma2,mu1,mu2,rho,x1,y1,f)/p
      ost(alpha,beta,sigma,sigma1,sigma2,mu1,mu2,rho,x1,y1,f);

23.24.     a=min([1 pratio]);  u=rand;  % Draw a uniform deviate from [0 1]%
Calculate the acceptance ratio
25.   if u<a;  % Do we accept this proposal?
26.27.        End alpha=new_alpha;   % proposal becomes new value for alpha

28.
29. % % Propose a new value for beta 30.
new_beta=normrnd(beta,propsigma(2));
31.


pratio=post(alpha,new_beta,sigma,sigma1,sigma2,mu1,mu2,rho,x1,y1,f)/post(alpha
,beta,sigma,sigma1,sigma2,mu1,mu2,rho,x1,y1,f);
32.   a=min([1 pratio]);  % Calculate the acceptance ratio

33.34.       u=rand;  if u<a  % Do we accept this proposal?% Draw a uniform
deviate from [0  1]

35.      beta=new_beta;  % proposal becomes new value for theta2
36.      End
37.
38.   % %  Save state
```

```
39.    state(1,t)=alpha;
40.    state(2,t)=beta;
41.    end
```


Exponential parameters

```
P = rlnorm(n,(1/20000000))
N = rlnorm(n,(1/620000))
D = rlnorm(n,(1/1600))
```

LogNormal Parameters

```
P = rlnorm(n,16.79254,0.173165)
```

```
N = rlnorm(n,13.11504,0.727722)
D = rlnorm(n,7.295927,0.346973)
```

……Listing (A3)

Uniform

```
runif(n, min = 14821000, max =
30726000) runif(n, min = 102051, max
= 1728808) runif(n, min = 700, max =
2900)
```

Gamma Distribution

```
P = rgamma(n,90,70)*10000000
N = rgamma(n,2,66)*10000000
D = rgamma(n,90,300)*100000
```

```
simm <- function(n){

  #simulated P, N, D, X and Y, (an
example)
  N = rlnorm(n,13.11504,0.727722)
  P = rlnorm(n,16.79254,0.173165)
  D = rexp(n,(1/1600))
  X = sort(log(N/P))    Y = sort(log(D/P))    m
= data.frame(Y,X)

  #regression model to generate alpha and beta
(A4)   regmodel<- summary(lm(Y~X, data=m))

  #extract alpha and beta and save
  gradcoefficient<-
as.matrix(regmodel$coefficients)   alpha<-
gradcoefficient[1,1]    beta<-
gradcoefficient[2,1]

   output<- (c(alpha,beta))    output}
```

……Listing regmodel

```
    >VarCorr(Null.Model)
  Regions = pdLogChol(1)
          Variance                StdDev
     (Intercept)  0.1891104  0.4348683
     Residual     0.1389485  0.3727579
```

……(Listing A5)

```
> Null.Model<-lme(y~1,random=~1|Regions,data=
  +fatalities,control=list(opt="optim"))
> GREL.DAT<-GmeanRel(Null.Model) > names(GREL.DAT)
 [1] "ICC"     "Group"   "GrpSize" "MeanRel"

> GREL.DAT$ICC #ICC estimate
[1] 0.5764526
> GREL.DAT$MeanRel
[1] 0.9627688 0.9627688 0.9627688 0.9627688 0.9627688
[6] 0.9627688 0.9627688 0.9627688 0.9627688 0.9627688
> mean(GREL.DAT$MeanRel) #Average group-mean
reliability [1] 0.9627688
```

..... Listing (A6)

```
> summary(Model.1)
  Linear mixed-effects model fit by REML
 Data: fatalities
       AIC       BIC     logLik
  104.5536 120.7091 -47.27679

  Random effects:
 Formula: ~1 | Regions
            (Intercept) Residual
  StdDev: 0.4575869   0.2754293

  Fixed effects: y ~ x + G.x
                  Value Std.Error DF t-value p-value
 (Intercept) -10.075599 0.7426214 179 -13.567611 0.0000 x  0.459058
 0.0373957 179 12.275704 0.0000 G.x -0.544840
 0.1657983 8 -3.286159 0.0111

  Correlation:
    (Intr) x      x 0.000     G.x 0.955 -0.226

  Standardized Within-Group Residuals:
      Min       Q1       Med       Q3      Max
 -5.2750037 -0.5117580 0.1294992 0.5301295 2.0982667

  Number of Observations: 190 Number of
  Groups: 10
```

Listing (A7)

```
> summary(Model.2)
  Linear mixed-effects model fit by REML
 Data: fatalities
   AIC      BIC      logLik
   78.74943 101.3672 -32.37471

  Random effects:
```

```
    Formula: ~x | Regions
  Structure: General positive-definite,
             Log-Cholesky parametrization
                StdDev     Corr
    (Intercept) 0.3930480 (Intr)
    x            0.1955038 0.997
    Residual     0.2510588

   Fixed effects: y ~ x + G.x
                  Value           Std.Error DF t-value p-value
    (Intercept) -9.234113 0.20646101 178 -44.72570 0 x
    0.445931 0.07067286 178 6.30979 0
    G.x            -0.338452 0.05155597 178 -6.56476 0

  Correlation:
      (Intr)      x x   0.544
    G.x       0.594      -0.300

   Standardized Within-Group Residuals:
    Min          Q1           Med          Q3          Max
   -4.82969980 -0.38141416 0.05199466 0.49544405 2.52586513
Number of Observations: 190
Number of Groups: 10
```

……Listing (A8)

```
   > VarCorr(Model.2)
      Regions = pdLogChol(x)
             Variance        StdDev          Corr
      (Intercept)  0.15448674   0.3930480  (Intr)  x
      0.03822173  0.1955038  0.997
      Residual    0.06303053  0.2510588
```

.....Listing (5.14)

Listing (A9):

```
> rtf<-data.frame(matrix(c(1,920,8773,2,914,9116,3,901,7677,4,824,7664,5,
  1026,9106,6,1049,9903,7,1015,10433,8,1419,11786,9,1237,10202,10,1437,12310,11,1660,1
  3178),11,3,byrow=TRUE))
> names(rtf)<-c("year","Fatality","Injury")
>  rtf$Casualty<-rtf$Fatality+rtf$Injury
>  rtf$year<-factor(rtf$year,labels=c("1991","1992","1993","1994","1995",
 "1996","1997","1998","1999","2000","2001"))
> rtf$Y<-cbind(rtf$Fatality,rtf$Injury)
>  rtf

    Year Fatality Injury Casualty Y.1             Y.2
1      1991 920      8773   9693   920    8773
2      1992 914      9116   10030  914    9116
3      1993 901      7677   8578   901    7677
4      1994 824      7664   8488   824    7664
5      1995 1026     9106   10132  1026   9106
6      1996 1049     9903   10952  1049   9903
7      1997 1015     10433 11448   1015   10433
8      1998 1419     11786 13205   1419   11786
9      1999 1237     10202 11439   1237   10202
10     2000 1437     12310 13747   1437   12310
11     2001 1660     13178 14838   1660   13178
```

Listing (10):

```
>  logistic< glm(Y~year,family=binomial,data=rtf)
```