AN ANALYSIS OF ROAD TRAFFIC FACTORS AND ROAD SAFETY STRATEGIES THAT PREDICT ROAD FATALITIES OVER TIME ACROSS FIFTY STATES IN USA

by

Benzaman

A thesis submitted in partial fulfillment
of the requirements for the degree

of

Master of Science

in

Mechanical and Industrial Engineering

MONTANA STATE UNIVERSITY
Bozeman, Montana

April 2017

DEDICATION

To my dearest mother, Naseem Jahan Eva and my dearest father, Shahiduzzaman who always motivated me and encouraged me to dream big and pursue greatness.

ACKNOWLEDGEMENTS

I would like to whole heartedly acknowledge and thank my advisor, Dr. Nicholas J. Ward for his tremendous support, guidance and motivation. Additionally, I would also like to thank my committee members, Dr. Steven Swinford, Dr. David Claudio and Dr. William J. Schell for their valuable comments and guidance throughout the progress of this thesis. Without the support of everyone mentioned above this thesis would not have been possible.

TABLE OF CONTENTS

TABLE OF CONTENTS – CONTINUED

LIST OF TABLES

LIST OF TABLES – CONTINUED

Table Page

LIST OF TABLES – CONTINUED

Table Page

LIST OF FIGURES

LIST OF FIGURES – CONTINUED

Figure Page

LIST OF FIGURES – CONTINUED

Figure Page

LIST OF FIGURES – CONTINUED

Figure Page

LIST OF FIGURES – CONTINUED

Figure Page

LIST OF FIGURES – CONTINUED

Figure Page

LIST OF FIGURES – CONTINUED

Figure Page

LIST OF FIGURES – CONTINUED

Figure Page

LIST OF FIGURES – CONTINUED

Figure Page

LIST OF FIGURES – CONTINUED

Figure Page

LIST OF FIGURES – CONTINUED

Figure Page

ABSTRACT

Strategizing to decrease statewide road fatalities is an important aspect in road safety research in the United States. But obtaining information on a variety of variables, such as economic, socio-cultural, demographic and political factors, at the state level can be a difficult task. The public databases sometimes do not provide full information on these variables due to missing data. If these variables are neglected from the analysis because of missing data points. valuable information is lost in the process. Therefore, analyzing missing data has been considered as an additional step towards variable selection process in this thesis study. In order to impute the missing data, multiple imputation method was chosen. After the data imputation, the significant variables associated with road fatalities in 50 states were identified. This was done using a linear regression model which revealed that the top reasons for road fatalities are drunk driving, distracted driving and unemployment. In the process of linear regression modelling 48 predictive models were obtained.

During the process of data collection, it was observed that data sources did not offer necessary information on road safety culture, behaviors, norms, attitudes and beliefs related to road fatality. This study offers two solutions for inferring a road safety culture and understanding its effects. The first solution was to analyze residuals from random effects two-way panel regression model and to generate performance indicator of inferred road safety culture. From the value of the indicators it was clear which state was the safest twenty years ago and which state is the safest now. Through the change in the value of the indicator, a state's progress in terms of safety culture was also measured. The second solution was to use people's political views on the democratic party and the republican party as a proxy for the road safety culture. This resulted in a significant increase in the goodness of fit for the linear regression model. This thesis provides prediction models, significant factors, and performance indicators of a road safety culture which can be used in state level road safety strategy development and policy making.

CHAPTER ONE – INTRODUCTION

1.1. Road Fatalities, Analytical Models & Factors

Road traffic crashes (RTC) and the associated fatalities are considered one of the most important socio-economic issues of our time. On a worldwide scale, more than 1.2 million people die of road crashes each year and with another 50 million injured (Hughes et al. 2015). RTCs are one of the top ten leading causes of death globally and the top cause of death for people between the ages of 15-29 (Lozano et al. 2013). A recent comparative analysis of 17 countries concluded that the national social cost of road crashes has a significant impact on the gross domestic product (GDP) (Wijnen and Stipdonk 2016). This social cost of traffic fatalities is emphasized by international organizations such as WHO, World Bank, ADB, and European Commission (EC). Notably, international assessments reveal that the United States has fallen behind in the development of road traffic safety compared to France, Canada & Australia, even though they all had the same fatality rates per 100,000 persons (Naumann et al. 2010). While the fatality rates in these countries decreased by 60%, the United states has only been able to reduce the rate by 35%. Compared to the European countries, 15-16 fatalities per 100,000 person is still high and a concern (Organization 2009).

Even though a 35 percent reduction is a tremendous feat, identifying the factors that cause road crashes is a challenging task. This is because some of factors are still unknown and unmeasured. Moreover, the understanding of traffic system safety may be further obscured by the fact that road fatalities or crash events are discrete and random events (Chin and Quddus 2003). The degree to which these factors differ depends on

states, societies and regions (Norman and Organization 1962). Due to these varying locational, unobservable and immeasurable factors, strategizing road safety with the goal of decreasing road fatality is considered to be complex task (Dupont et al. 2014).

The application of different analytical models is one of the most useful methods for identifying the significant factors related to human, vehicle, socio-economic, road infrastructure, land-use, and the environment (Quddus 2008). These analytical models can include regression modelling, time series modelling, correlation analysis, spatial analysis etc. But these analytical models should be based on a theoretical framework. In this case, a framework can dictate and identify the potential causes of road fatalities and help policy makers and researchers with countermeasures. Hakim et al. (1991) proposed that any analytical model deriving from an acceptable framework is expected to have four basic characteristics:

1. Description of the phenomenon,
2. Explanation,
3. Prediction, and
4. Incorporation of policy variables

This thesis has identified that the incidence of road fatalities across United States as a phenomenon that needs to be analyzed thoroughly. It will emphasize analytical modelling of road fatalities across all 50 states in the United States and then investigate which variables are significant using regression modeling.

1.2. Objective of the Thesis

The primary objective of this thesis is to identify factors predictive of traffic fatalities within all 50 states in the USA. These factors include socio-economic, political, legal and legislation, and healthcare aspects, as well as weather, road infrastructure, vehicles, road users, collective driving behavior etc. Two different analytic methods were used to identify which of these factors are predictive based on 21 years of state-level traffic fatality data. These analytic methods include an approach with linear regression models and a panel data regression model.

To obtain better parameter estimates associated with the analytical models mentioned above, the missing data analysis method was considered as an important step in the variable selection process. This was done by generating several datasets with imputed data and analyzing them separately. Finally, the results were pooled to obtain the final selected variables. The steps associated with the missing data analysis and the justification for the variable selection procedure are also discussed in this thesis.

1.3. Thesis Structure

This thesis is structured in seven chapters including the introduction in which a general overview is expressed. Chapter Two reviews the literature and previous research work on theoretical frameworks of road fatalities and proposes a new conceptual framework, regression based modelling techniques, and finally, the missing data analysis techniques involving multiple method. Chapter three explains the definition of the selected variables, data mapping on the new theoretical framework, as well as the

limitation of data according to new conceptual framework. Chapter Four describes the analytical method of missing data analysis and generates linear regression models for the 50 states. It also highlights the variables that explain the variation in the fatality trend and specifies significant factors as found through statistical analysis. Analysis of the panel regression model approach combining the data from all 50 states data will be presented in Chapter Five. Chapter Five also discusses residual transformation to obtain performance indicators, provides a critique to a previous study, and outlines the effects of proxy cultural factors on regression model parameters. A comparative discussion of the analyzed results obtained from the two different methods is presented in Chapter Six, along with views on counter intuitive results, relative comparison of two modelling techniques, and a discussion on safe states and dangerous states for driving. Chapter Seven provides recommendations on the possible impact of this thesis on the government policy-making for road fatality reduction at the state level as well as federal level. Chapter Eight includes a conclusion and recommendations for future research scopes.

CHAPTER TWO – LITERATURE REVIEW

To identify the causes of road fatality on a state level or on a national level, various analytical models are utilized. These analytical models provide a rationale for selecting variables for this thesis. However, before an analytical model is formulized, a theoretical framework is required. A theoretical framework provides a better understanding of real world systems by abstraction or simplification in order to integrate knowledge across different disciplines (Heemskerk, Wilson, and Pavao-Zuckerman 2003). The framework will enable one to look for variables that may differ from country to country or state to state, depending on how the area of the different geography is defined. This literature review will discuss each of these fundamental steps as well as an approach towards handling missing data. This chapter is divided into three sections and will discuss:

i. Theoretical frameworks of road fatality,
ii. Analytical Methods to Test Models, and
iii. Missing data analysis techniques and the multiple imputation method.

2.1. Theoretical Frameworks of Road Fatality

From a very broad perspective, the defining elements within a certain geographic boundary are the human geography and the physical geography (Bonnett 2008). On one hand, human geography consists of human, political, cultural, social and economic aspects. On the other hand, physical geography addresses climatology, meteorology, environmental management and other aspects of earth science. Considering that there are

some fields (hydrology, oceanography, pedology, paleogeography, glaciology, geodesy etc.) which do not have a distinct relationship with road fatalities, if the scope is narrowed a bit further, certain aspects of physical geography, especially meteorology, have been extensively reviewed in relation to their effect on road fatality (Eisenberg and Warner 2005, Fridstrøm et al. 1995, Qiu and Nixon 2008).

Although discrete and random in occurrence, a road fatality incident occurs within a certain location, i.e. within a city or a region. In almost all road safety research, a certain boundary or area - whether it is a country, state, region, or a city - is considered. So, the geographic boundary is considered as the top level factor. As the overall aspect of geographic boundary can be fully described by the human geography and physical geography, these will be considered as the next level.

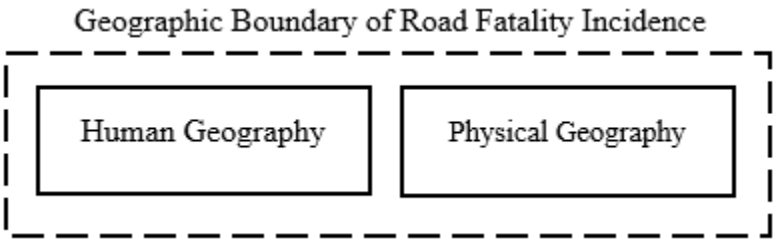

Figure 2.1. Defining geographic boundary at the top level

The contexts related to human geography are extensively reviewed and researched as factors such as:

i. Human - population, population growth rate, population density, population groups (according to age, gender, race, location, social status) etc.

ii. Economic conditions – GDP per capita, per capita income growth, motorization, unemployment rate, infrastructure development, prices of fuel, capital investment on infrastructure, industrialization etc.,

iii. Social and cultural conditions - ethics, norms, values, religious views, level of education, literacy rate, views towards laws and legislations, status of healthcare system, behavior, attitude etc., and

iv. Political conditions – political views of people, governance quality and its effectiveness, appliance of law and effectiveness of the judiciary body, law and enforcement capability & development, political interventions etc.

The next level illustrates the different aspects of human geography and physical geography as the framework explores cultural, social, economic, political, human aspects, meteorology and climatology, which have more direct relation with road fatalities.

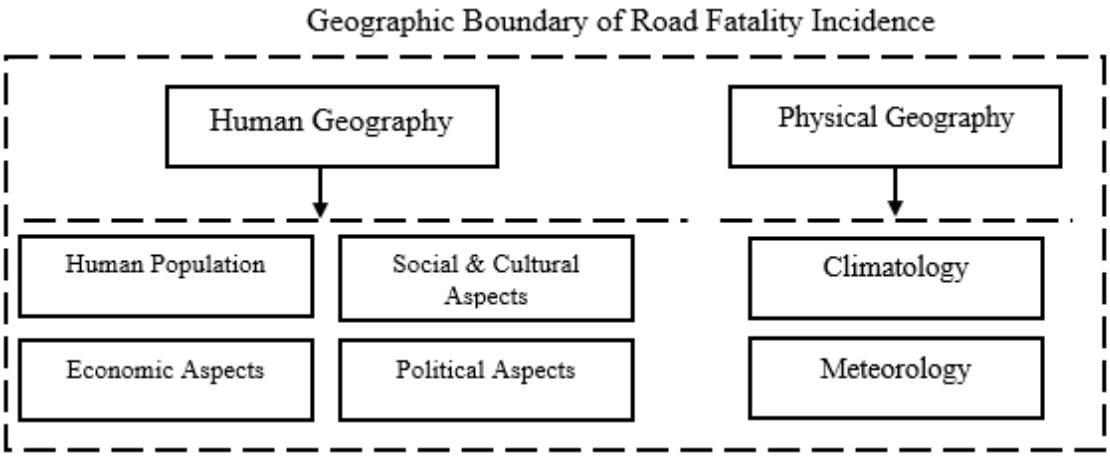

Figure 2.2. Defining the second level of the integrated framework

In a comparative study conducted by Eiksund (2009) to better understand risk-taking behavior among youth in rural and urban areas in Norway, it was emphasized that

road accidents are a result of a series of unfortunate events. These events are triggered by system risk – which is comprised of road systems, vehicles and weather conditions and the extent of risk taking in the culture – which is shaped by norms, values, beliefs and perceptions. The study focused attention on the fact that values, perception, norms and beliefs affect the process of gain or loss mentality in risk-taking behaviors exhibited by road users. However, the key elements that influence system risk and risk taking culture are socio-economic and socio-cultural environment.

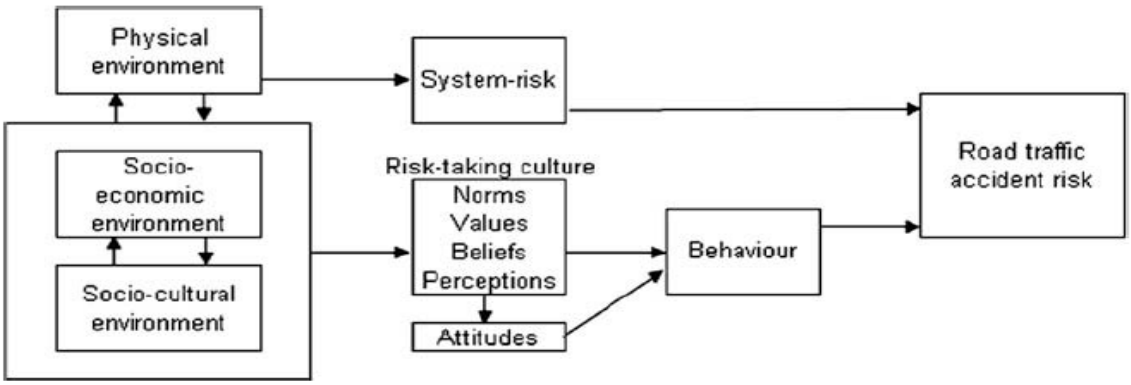

Figure 2.3. Illustration of framework proposed by the study Eiksund

There is a distinctive similarity between the features provided by Eiksund and the branches of human geography. The features of socio-economic environment and socio-cultural environment fall under the socio-cultural, economic and political aspects, all branches of human geography. The system risk has been defined as the risk factors originating from the vehicles, road infrastructure and the weather. In a similar study, the road risks were modelled by three safety pillars: road users, vehicles and infrastructure (Othman, Thomson, and Lannér 2009).

Schepers et al. (2014) investigated the relationship between risk to exposure, crash risk, and consequences of crashes and injuries in a conceptual framework. The study presented factors such as travel behavior for exposure to risk, and road infrastructure, road user and vehicles for risk. The travel behavior is considered a cultural factor, which in terms of the authors is influenced by location of activities, resistances (transportation cost) and needs, and opportunities and abilities (NOA). It should be pointed out that the three safety pillars are also illustrated as interacting factors with crash risk and injury risk. In a similar approach, Rumar (1999) described the road safety problem (number of human injuries and fatalities) as a function of the product of exposure, crash risk and injury consequences. The framework utilized by Schepers is given below:

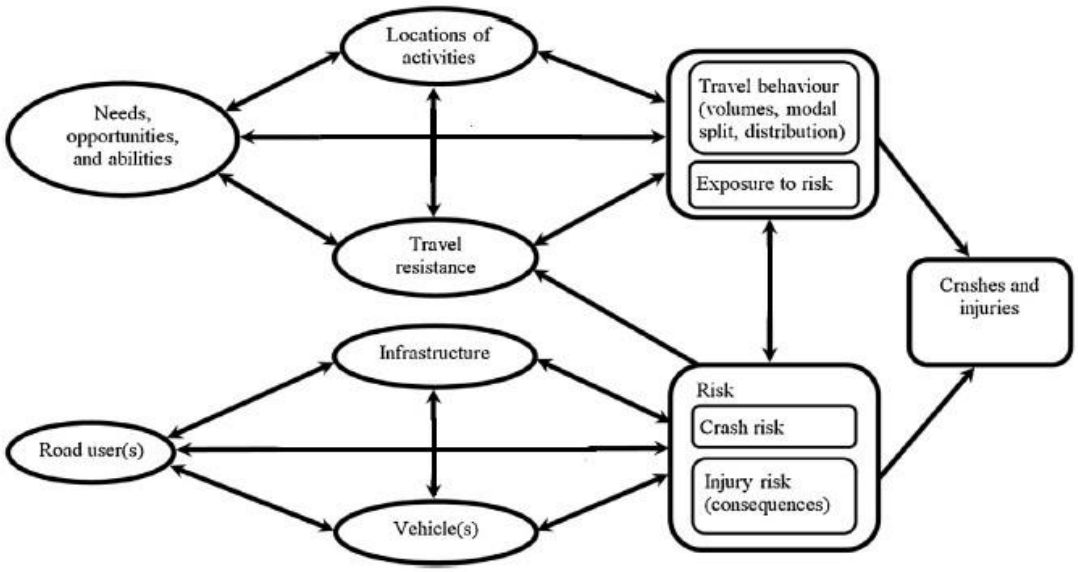

Figure 2.4. Conceptual model depicting exposure, risk and consequences

In another study, Haddon (1980) described that motor vehicle fatality and injury prevention is based upon analyzing the sequence of phases, known as pre-crash phase, crash phase and post-crash phase. They also describe the interaction between four factors which are human, vehicles and equipment, environment and roadway, and socio-economic environment. Pre-crash phase relates with economic and social contexts as well as transport and land-use related factors; however, crash occurrence, phase factors include road infrastructure, vehicles and human components. The post-crash situation or the consequences of the crash depend on the response system. Factors of the natural environment, such as topography, remain constant, and the study determined them to influence the pre-crash phase, while meteorological factors, such as precipitation and sunlight, affect the crash phase. Based on the insight proposed by Haddon, a range of road fatality reduction factors have been proposed, researched and revised. Peden et al. (2004) illustrated the Haddon Matrix as below:

		FACTORS		
PHASE		HUMAN	VEHICLES AND EQUIPMENT	ENVIRONMENT
Pre-crash	Crash prevention	Information Attitudes Impairment Police enforcement	Roadworthiness Lighting Braking Handling Speed management	Road design and road layout Speed limits Pedestrian facilities
Crash	Injury prevention during the crash	Use of restraints Impairment	Occupant restraints Other safety devices Crash-protective design	Crash-protective roadside objects
Post-crash	Life sustaining	First-aid skill Access to medics	Ease of access Fire risk	Rescue facilities Congestion

Figure 2.5. The Haddon matrix (Source: World Report on Road Traffic Injury Prevention 2004)

Certain aspects from the Haddon Matrix as well as the three pillars of road safety have been captured in the integrated model. The Haddon Matrix clearly shows the role of

the environment on road fatality issues. Just as some aspects of physical geography have been left out of the framework, it is important to understand that not all the aspects of human geography are related with road fatality. The three most important factors which are highly related with road fatality and have been extensively researched and reviewed are road users, vehicles and infrastructure. These three pillars were also captured in the framework proposed by Schepers et al (2014).

Road users are highly correlated with the population within a geographic boundary and can be represented as a subset of the population. Road users can be segregated by gender, age group, race, etc. if the data associated with such driver attributes are available.

Similarly, vehicles and infrastructure growth can be associated with economic development. Economic aspects such as GDP per capita, total civilian labor force, number of employed and unemployed people, unemployment rate, inflation, etc. indicates economic situation.

The integrated framework captures these features as well as cultural, political and economic aspects. The social and cultural factors within a certain boundary shape an individual's behaviors, attitude, values and norms. Similarly, the political environment is comprised of political bodies such as government, judiciary branches and related features regarding effectiveness, transparency, capability, etc.

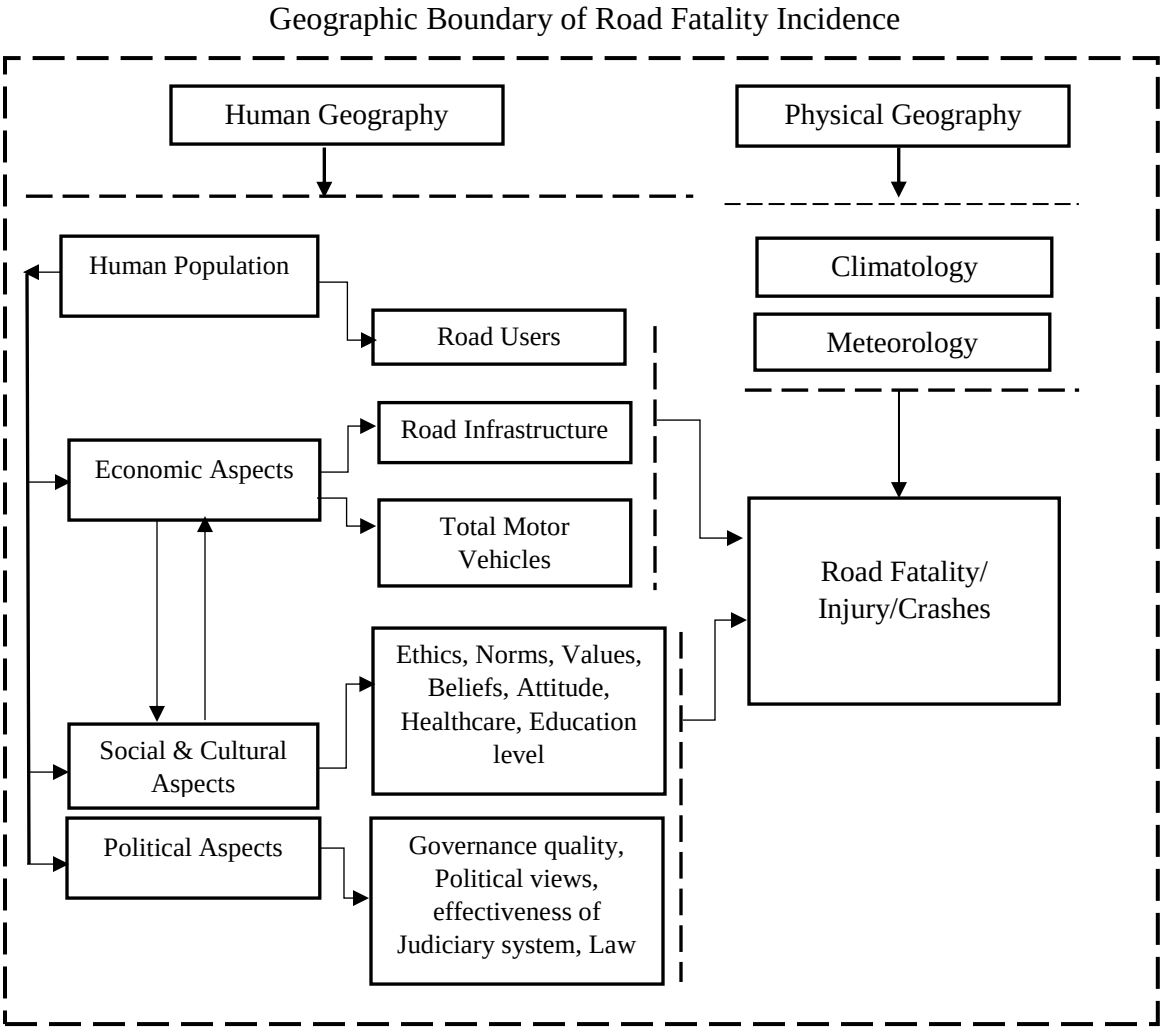

Figure 2.6. Features of the Integrated Framework at the third level

The hypothesis is based on the conceptual model presented here, any variables associated with road fatalities can be captured and analytical models can be formulated if empirical or experimental data are available, since this conceptual framework provides a definitive and logical rationale for selecting the variables.

2.2. Analytical Methods to Test Models

It is important to consider the attributes and challenges associated with modelling and parameter estimation. On that point, Hakim et al. (1991) discussed methodological issues related with analytical models and pointed out that aspects, such as collinearity among economic and demographic variables, variables related with driving aspects and the structure of the error term plays a significant role since dependence in the error terms results in biased estimates. In the study, it was also debated whether, when considering cross sectional analysis, time invariant factors such as topography, climate, and life-style should be assumed to undergo changes. Even though cross sectional analysis presents the researchers with lot of data, incorrectly specifying the nature of the time invariant variables may create ambiguity in cross sectional results. Under such circumstances, it is important to analyze the data for random effects which reflect the influence of the assumed time invariant variables on the result parameters & estimations.

In addition, when such datasets are analyzed, data points within the scope of the time interval are found to be missing. In most cases, these variables with their incomplete sets of data and information are sidelined from the analysis, which not only makes the data analysis approach ineffective, but also creates bias in the studies. Table 2.1 reviews previous research on regression analysis and missing data analysis, primarily focusing on studies in the USA and in some cases, comparative studies with other countries. Table 2.1 is given below:

Table 2.1. Literature Review

Sl no	Author	Topic/Objective	Geographic Boundary	Data Set Specifics	Methodology	Findings	Conclusion	Future Research
1	Zlatoper (1991)	Identifying the determinants of motor vehicle deaths	United States	1987 state level data, Cross sectional in nature	Least squares regression	Income, ratio of urban to rural driving, expenditures on highway and patrols, seat belt laws are inversely related with the motor vehicle death. Variables such as driving speed, density, alcohol consumption, temperature etc. are directly related		
2	Leigh and Waldon (1991)	Analyzing the effect of unemployment	All the states and District of Columbia	1976 - 1980	Regression	1% increase in the unemployment rate corresponding 5% reduction in fatality rate	Worsening unemployment leads to higher fatality rates (keeping number of miles driven constant). More unemployment leads to lesser driving	Increase in fatality due to worsening of unemployment, may be due to stress

Table 2.1. Literature Review – continued

Sl no	Author	Topic/Objective	Geographic Boundary	Data Set Specifics	Methodology	Findings	Conclusion	Future Research
3	Farmer (1997)	Analyzing & modelling monthly motor-vehicle fatalities and identifying significant variables	United States	Monthly data, from January 1975 - September 1995	Linear Regression	VMT and number of unemployed person were highly significant. Each are negatively associated	Short term changes in traffic fatality can be predicted by the number of unemployed persons	Hypothesized that an improved economy may lead to more recreational driving
4	van Beeck, Borsboom, and Mackenbach (2000)	The effects of economic prosperity	21 OECD countries	1962 to 1990	No mention of regression analysis, Longitudinal and cross sectional relationship	A non-linear relationship is found.	In early years, economic development lead to a higher fatality but later had a reversal effect	
5	Whetten-Goldstein et al. (2000)	Analyzing the effects of different alcohol policies, civil liability and other factors (Unemployment rate, Percentage of Miles Driven on Rural Roads, Number of Miles Driven Per Population, etc)	All the states	1984 to 1995	Logit chi square method and fixed effects to create a quasi-time series analysis	Laws related to suing bars for drunken behaviors were found highly significant. Dram shop liability policy is strongly associated with reducing fatalities.	Effects of social host liability, zero tolerance laws, and open container laws should be monitored. Higher deaths among adults where higher concentration of bars are present.	

Table 2.1. Literature Review – continued

Sl no	Author	Topic/Objective	Geographic Boundary	Data Set Specifics	Methodology	Findings	Conclusion	Future Research
6	Calkins and Zlatoper (2001)	Effectiveness of the seat belt usage	All the states	1988 and 1997 data	Cross sectional regression, 3 functional forms, 12 regression model in total	Primary and secondary seat belt law show positive association with the total and non-occupant deaths	Theory of offsetting behavior is supported - which means drivers tend to drive recklessly when they feel safe	Recognizing the probability of the offsetting behavior & precautions with increase in police monitoring to mitigate effects
7	Page (2001)	Draw comparative inference on road safety	21 OECD countries including United States	1980 to 1994	Time series cross sectional regression, Fuller and Battese Method	Seven significant variables, progression indicators of countries in 15 years	European countries are ahead of USA in terms road safety development	Better statistical modelling and more factors should be included
8	Cohen and Einav (2003)	Effects of mandatory seat belt law on driving behaviors and road fatality. Other variables were urban & rural traffic density unemployment, income, VMT, percentage, Demographics, crime rates.	All the states and District of Columbia	1983 to 1997	Simple linear regression and log-log regressions	1% increase in seat belt usage will have 136 fatalities reduction (using a linear specification), and that a 1% increase in usage will reduce 0.13% annual fatalities (using a log-log specification)	The theory of compensating behavior by Peltzman (1975) was refuted. Most of the control variable lost significance once state effects were imposed.	

Table 2.1. Literature Review – continued

Sl no	Author	Topic/Objective	Geographic Boundary	Data Set Specifics	Methodology	Findings	Conclusion	Future Research
9	Noland (2003)	Analyzing effects of improvement in road infrastructure on traffic related fatalities and injuries	All states	1984 to 1997	Fixed effect negative binomial regression	Road improvements was not significant. Age mix in population, increased seat belt use, reduced per capita alcohol consumption, increased seat belt usage, and advancement in medical technology were found significant	Results strongly refute the hypothesis that infrastructure improvements have been effective at reducing total fatalities and injuries. Control changes in exogenous factors and other interventions	Exercising newer statistical models for processing random components
10	Gaygısız (2010)	Firstly, effects of Cultural variables (Hofstede's cultural dimension, WGI, Schwartz value dimension scores) on road fatalities and secondly moderator role of WGI between cultural variables and road fatalities	46 countries including United States	From 2002	Correlation analysis, principal component analysis, ANOVA analysis	WGI were negatively correlated. “Power distance” dimension was positively correlated. “embeddedness”, “hierarchy” and “mastery” were positively related	Governance quality of a country has significant influence on national traffic safety, GDP negatively related with traffic fatality	Quality of police enforcement, driver education, and highway engineering can be included in the model when the data is available

Table 2.1. Literature Review – continued

Sl no	Author	Topic/Objective	Geographic Boundary	Data Set Specifics	Methodology	Findings	Conclusion	Future Research
11	Law, Noland, and Evans (2013)	Analyzed the effect of political stability, education level, per capita income, democracy, and more equitable income distribution on the enactment of motorcycle helmet law and seat belt law	31 countries including USA, an international comparison based study	Panel data, 1962 to 2003	Random effects binary logit model	Improvements in democracy, education levels, per capita income, political stability, and in-come distribution within a country are associated with the enactment of helmet law & seat belt law	Findings support Kuznets effect - that when income level rises sufficiently it induces demand for better road safety and laws	Effectivity of laws related to - speed, training of young drivers, and driving under influence should be checked
12	Antoniou et al. (2016)	Analyzing Effect of GDP	30 countries from Europe	1975 to 2012	Panel regression of log scales on both dependent and independent variables	The long-term elasticity is significant and positive for 10 countries (+0.63).	Outlines conditions to approach panel time series in a systematic way, capturing differences in integration between the series, and optimal selection and grouping of countries for robust estimates	Exploration of error correction models to simultaneously capture short-term and long-term effects of GDP

2.3. Missing Data Analysis Techniques and Multiple Imputation Method

One of the distinctive challenges to data analysis is processing and sorting raw data, which in most cases are incomplete, imperfect and erroneous. This poses a challenge in data processing, since just ignoring such information leads to definite loss of valuable information and to some extent bias in the study. This could prevent discovery of important factors or variables which may be considered highly significant in that field of research, and thus lead to invalid conclusions. According to Barnard and Meng (1999), the concerns surrounding missing data lead to three conclusions: (1) bias due to the differences between observed and unobserved data, (2) loss of efficiency, and (3) complication in data handling and analysis.

Determining the significant factors behind road crashes or road fatalities and injuries poses a challenge since the quality of data depends on the data site and data capturing effectiveness. However, recent developments in data analysis and data mining techniques have mitigated some of these missing data challenges by means of data imputation methods. Little and Rubin (2014) have categorized the data imputation methods into four groups. They are provided below:

i. Procedures based on completely recorded units,
ii. Data imputation based procedures,
iii. Weighting procedures, and
iv. Model-based procedures.

Depending on the structure of the data and whether the missing data scope is univariate or multivariate in nature, missing data can be imputed by likelihood-based

methods and estimated by inferences, Bayesian and multiple imputations for multivariate variables, bootstrapping, simulation based techniques and expectation-maximization (EM) algorithm techniques (He 2006). Based on the missing nature of the data, Rubin (1976) stated that every data point has some likelihood of missing, and classified missing data problems into three categories. The categories are given below:

i. Missing completely at random (MCAR) – when the nature of the missing data is unrelated to the data,
ii. Missing at random (MAR) – when the nature of the missing data is related to some sub-groups within the data,
iii. Missing not at random (MNAR) – when the nature of the missing data is related to the data but the reasons behind the missing is unknown.

Based on this distinction, the methods for data imputation change and Rubin (1976) described the process to be the *missing data model* or *response model*. It was implied these response models have certain conditions under which statistical inferences can be provided and estimations can be made. This thesis chose the multiple imputation(MI) model for missing data analysis purposes mainly because of the multivariate nature of the missing data as well as the performance capability under MAR or MNAR situations(Rubin 2004). It is currently also the universally accepted model for imputing missing data. A table is given below which summarizes some safety research where multiple imputation method was used:

Table 2.2. Review of research associated with multiple imputation missing data

Sl no	Author	Topic/ Objective	Geographic Boundary	Data Set Specifics	Methodology	Number of Imputation	Conclusion	Future Research
1	Heitjan and Little (1991)	Imputing missing data regarding seat belt use and blood alcohol content of the driver	FARS database, United States	BAC - 57.1%, Alcohol Involvement - 29.2%, seat belt use - 23.3%	Multiple Imputation based on predictive mean matching. Monte Carlo Simulation to explore the frequency properties	Five values were imputed for each missing data point	Method is effective in imputing missing BAC values if supporting data of police reported alcohol involvement is available. Monte Carlo results provided improved results than single imputation methods	
2	Subramanian and Utter (1998)	Imputing missing BAC Level Values in FARS database	FARS database, United States	58% data were missing	Multiple Imputation	Ten values were imputed for each missing data point	Missing data were revised back to 1982, to provide better estimates in alcohol involvement trend analysis	

Table 2.2. Review of research associated with multiple imputation missing data – continued

Sl no	Author	Topic/ Objective	Geographic Boundary	Data Set Specifics	Methodology	Number of Imputation	Conclusion	Future Research
3	Zador, Krawchuk, and Voas (2000)	Updating the estimates for alcohol related relative risk of driver involvement in fatal crashes	FARS database, National Roadside Survey United States	Weekend nights in 48 states, from 1 a.m. to 3 a.m.	Logistic Regression, Multiple Imputation	10 imputed data for each missing point	Novel study which estimated relative risk of higher BAC for six groups of drivers involved in fatal crashes. Study revealed that a driver with a BAC under 0.10% pose highly elevated risk both to them and other.	Further research support is needed to confirm the findings. (The BAC level has been already readjusted to 0.08%)
4	Tefft (2012)	Estimation of the proportion of crashes that involved drowsy driver	NASS CDS, 1999 to 2008	45% missing data on driver drowsiness was addressed	Multivariate imputation	10 imputed data for each missing points.	Results suggest that the prevalence of fatal crashes that involve a drowsy driver is over 350% greater than has been reported previously.	Further checking of the validity of the imputed data to reduce bias

CHAPTER THREE – VARIABLE SELECTION, MAPPING AND DATA LIMITATION

The chapter will highlight on the variables which have been selected for the analysis along with the definitions. Additionally, important discussions on data sites from which the data have been collected, variable mapping onto the new framework, lack of data on cultural factors due to reliable data site will also be addressed in this chapter.

3.1. The Selected Variables

Based on the integrated framework, a range of variables were selected based on human, economic, political and environmental factors. Since it is very challenging to identify each and every variable each represents the overall aspect of the factors defined by human geography and physical geography. Apart from the cultural factors, a wide range of variables have been collected from various public datasets (accessed online) for the time frame from 1994 to 2014. There are an overall of 46 variables including the dependent variable which is the total number of fatalities, among which there are 38 numerical and eight categorical variables. A summary of source information as well as factors associated with the new framework is given below:

Table 3.1. Data collection Sources for the Variables

Factors or Variables	Definition	Source Description
Fatalities	Number of fatalities from motor vehicle crashes involving vehicle occupants and non-occupants	https://www-fars.nhtsa.dot.gov/Main/index.aspx (Accessed at September 2016)
Speeding	Fraction of 'driving too fast for conditions or more than posted maximum' of all the errors and violations by drivers that led to fatal crashes	
Distraction Driving	Fraction of 'Inattentive driving and cellphone use while driving" of all the errors and violations by drivers that led to fatal crashes	
Drunk Driving	Fraction of 'under the influence of drugs, alcohol or medication while driving' of all the errors and violations by drivers that led to fatal crashes	
Total Labor Force	Total number of civilian labor force	https://www.bls.gov/ (Accessed at September 2016)
Employed	People employed within the labor force	
Unemployed	People unemployed with the labor force	
Unemployment Rate	Percentage of unemployed people	
GDP per capita	Average income per person in dollars	https://www.bea.gov/ (Accessed at September 2016)
Road Improvement Capital Expenditure	Total State wise capital outlay for interstate, arterial and collector systems	https://www.nhtsa.gov/ (Accessed at September 2016)
Diesel Consumption	Diesel consumption in thousands of gallons	
Gasoline Consumption	Gasoline consumption in thousands of gallons	
Resident Population	Total state level population in thousands	
Urban Population	State level population in thousands in urban areas	
Rural Population	State level population in thousands in rural areas	

Table 3.1. Data collection Sources for the Variables (Continued)

Factors or Variables	**Definition**	**Source Description**
Total Road Length	Total road length in miles	https://www.nhtsa.gov/ (Accessed at September 2016)
Rural Road	Total road length in miles in rural areas	
Total Number of Vehicles	Total registered motor vehicles	
Motorcycles	Number of motorcycles owned privately & commercially	
Heavy Trucks/Tractors	Total number of truck tractors registered	
Total Road Users	Total registered and licensed drivers	
Young Road Users	Registered and licensed driver from age 15 to 20	
Male Driver	Percentage of male driver of the total road users	
Female Driver	Percentage of female driver of the total road users	
Million VMT	Annual travel by motor vehicles in million miles	
Temperature	Annual temperature in degree Fahrenheit	https://www.ncdc.noaa.gov/ (Accessed at October 2016)
Precipitation	Annual rainfall in inches	
Alcohol Consumption per Capita	Per capita consumption of all beverages (wine, beer and spirits) in gallons	https://www.niaaa.nih.gov/ (Accessed at October 2016)
Diesel Price	Annual average diesel price per gallon	https://www.eia.gov/ (Accessed at September 2016)
Gasoline Price	Annual average gasoline price per gallon	
Total Number of Hospitals	Total number of community hospitals in states (federal hospitals, psychiatric hospitals, long term care hospitals are excluded)	https://www.cdc.gov/nchs/hus/previous.htm (Accessed at September 2016)
Beds per 1000 people	Beds per 1000 person in community hospitals in states (Beds in federal hospitals, psychiatric hospitals, long term care hospitals are excluded)	
Physicians	Active non-federal physicians per 10000 civilians	http://kff.org/ (Accessed at September 2016)

Table 3.1. Data collection Sources for the Variables (Continued)

Factors or Variables	**Definition**	**Source Description**
Number of Nurses	Total number of registered and licensed nurses	https://www.bls.gov/ (Accessed at September 2016)
Emergency Response Team	Total number of people of emergency medic team and police, fire and ambulance dispatchers	
Seat Belt law	Primary enforcement of seat belt law (presence/absence)	http://www.iihs.org/iihs/ratings (Accessed at November 2016)
Child Restraint Law	Primary enforcement of child restraint law (presence/absence)	
Helmet Law	Primary enforcement of helmet law (presence/absence)	Chase (2014)
Texting While Driving Law	Texting on the cellphone or smartphone while driving law (presence/absence)	
BAC Level Law	Blood alcohol concentration (≤ 0.08) while driving level law (presence/absence)	https://www.rita.dot.gov/ (Accessed at November 2016)
Hand Held Devices While Driving Law	Primary enforcement of hand held devices while driving law (presence/absence)	
Cellphone use while driving for novice and teen driver Law	Cellphone use while driving for novice and teen drivers law (presence/absence)	McCartt, Kidd, and Teoh (2014)
DemOrRepublican	Candidate representing which party during presidential election (Dem =1, Rep = 0)	http://uselectionatlas.org/ (Accessed at February 2017)
Dem	Vote percentage obtained by Democratic party candidate in presidential election	
Rep	Vote percentage obtained by Republican party candidate in presidential election	
VotePer	Percentage of vote casted from voter population	

For the analysis, the law based variables have been considered as binary. After primary enforcement of these laws – a value of one was considered but for secondary enforcement or absence of law – zero was considered. In addition to that, the variable 'DemorRep' showed the results in the US presidential election, by defining a state to be democratic or republican with one and zero respectively.

The variable - emergency response team was defined to be consist of emergency medic team and the police, fire & ambulance dispatchers. 'Number of nurses' – is a variable that consists the total number of nurses in that state (licensed practical nurses and registered nurses), since accurate data on total number of active nurses was not obtained. Physicians per 10000 civilians was obtained from *Health, United States* report and it is considered as a proxy variable for the total number of physicians per 10000 civilians. This variable only represents active non-federal physicians in each state since the data on federal physicians were not found. As accurate data on state wise Beds for 1000 people and total number of hospitals were not found, the data on community hospital on these two aspects were taken as a proxy.

The definition of community hospital did not include federal hospitals, long term care hospitals, psychiatric hospitals, alcoholism related hospitals, and institutions for mentally changed. The data provided on the community hospitals represented 85% of the total hospitals (Source: 1999 - 2015 AHA Annual Survey, Copyright 2016 by Health Forum, LLC, an affiliate of the American Hospital Association).

Young drivers have been defined based on the age group from 15 to 20 years (Source: Traffic Safety Facts-DOT HS 811 218-November 2009, "*Fatal Crashes Involving Young Drivers"*).

3.2. Mapping the Variables on The Integrated Framework

The main goal of this sub section is to map the collected variable to the integrated framework. The illustration is given below:

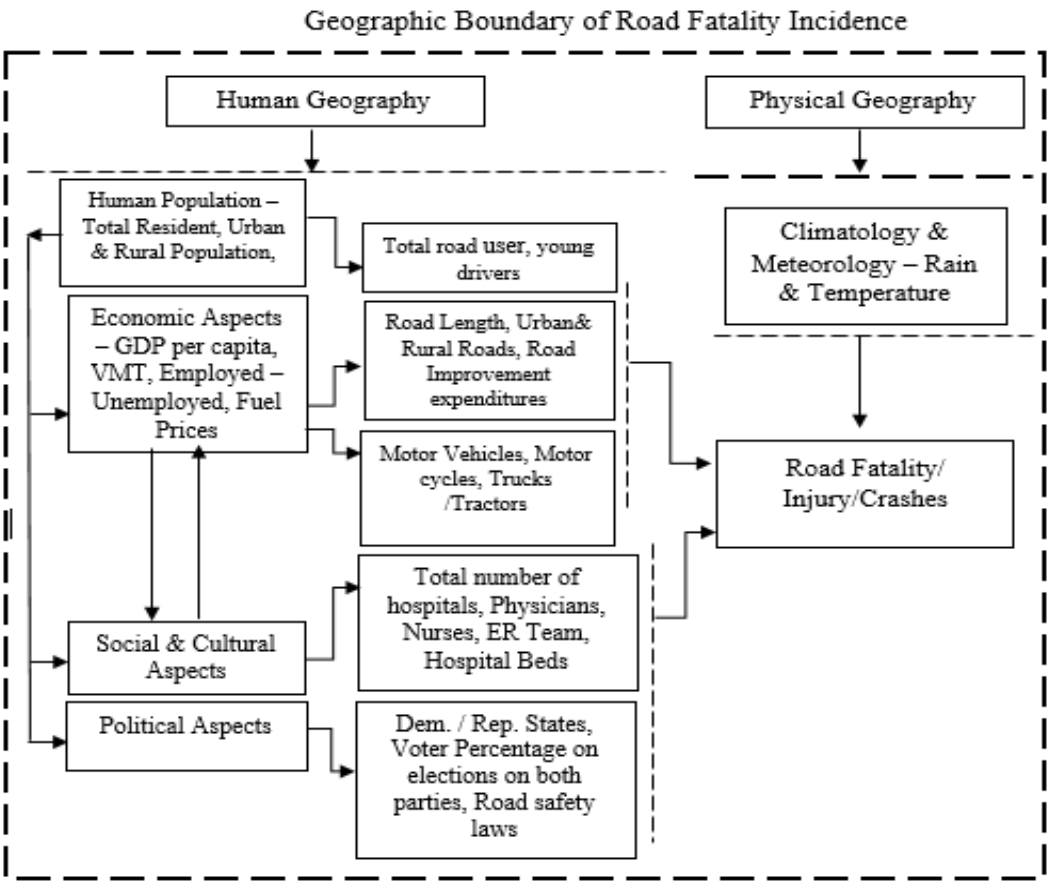

Figure 3.1. Data mapping of the Integrated Model at the third level

3.3. Data Limitation

It is noticeable from the list of variables that the cultural aspects within a certain geographic boundary were not captured. This has been identified as a limitation of this study, since to the authors best knowledge database associated with measures of culture, governance quality, driving behavior, norms, attitude do not exist. Unlike European Social Survey *(ESS)* and SARTRE *(Social Attitudes to Road Traffic Risk in Europe)* which capture attitudes, beliefs and behavior patterns of the various populations in Europe, there is not any robust & detailed database for United States that captures state level information. In a study by Cestac, Kraïem, and Assailly (2016), the above mentioned databases were used to explore links between cultural values and perceived drunk driving of one's peers & self-reported behavior. These databases have been used in numerous studies linking road fatalities with social norms, attitudes, culture, beliefs, and perception (Nabi et al. 2007, Quimby, House, and Ride 2005, Vanlaar and Yannis 2006, Vereeck and Vrolix 2007).

In order to draw some inferences about culture and attitude, governance quality and other behavioral traits, a performance indicator was created based on replicating the study of Page (2001). The inferences were drawn from analyzing the residuals which captured the time invariant factors as well as factors which were not considered in the model (exogenous variables). This study will be discussed in chapter five.

Four political factors were taken as a proxy for both the political condition and cultural mindset of people to draw some form of inference in the first and second method of analysis.

CHAPTER FOUR – MISSING DATA ANALYSIS AND LINEAR REGRESSION MODELS

It has been discussed in the literature review that concerns with missing data include biasing due to the differences between observed and unobserved data, efficiency loss in terms of information loss, and data handling complications. Multiple Imputation is one of the many ways of handling this missing data issue regardless of the nature of the missingness. This chapter will highlight reasons behind choosing multiple imputation method to treat missing data, describe the steps associated with multiple imputation, and finally, summarize the pooled estimates and parameters for linear regression models for the 50 states. Finally, a prediction model (regression model for Montana) was used to obtain predicted fatalities to validate with the actual fatality data.

Among the selected variables, ones that have missing data within the time frame were considered for the thesis. Amount of missing data points associated with these variables among 21 data points (1994-2014) are summarized below:

Table 4.1. Summarized information on the missing data for each state

Variables	Number of Missing Data	Which year the data is missing
Urban Population	9	1994 - 97, 2009, 2011 – 14
Rural Population	9	1994 - 97, 2009, 2011 – 14
Number of Hospitals	5	1994 - 1998
Number of Nurses	3	1994 - 1996
Number of Emergency Response Team	3	1994 - 1996
Annual Price of Diesel	3	1994 – 96 (17 states only)
Number of Beds per 1000 person	3	1995, 1999 & 2014
Number of Physicians per 1000 people	2	1999 & 2014
Percentage of Male Driver	1	1994
Percentage of Male Driver	1	1994
Expenditures of Road Improvement	1	2011

As per the discussion in the literature review, multiple imputation was chosen because of its performance capability under MAR or MNAR situations. Morris, White, and Royston (2014) stated that multiple imputation can provide valid inference on MCAR, MAR, and MNAR. However, it is important to specify the imputation model correctly so that Rubin's rules estimate consistent parameters and confidence intervals that incorporate uncertainty due to missing data (Schafer 1999).

4.1. Reasons Behind Choosing Multiple Imputation

The rationale behind considering missing data analysis for this study originated because a significant amount of data was missing for healthcare data variables, which fell under socio-cultural factors, and demographic variables, which fell under human population factors. For urban and rural population variable, 42.85% of the data was missing, and for the number of hospitals 23.81% was missing. For the variable 'number of nurses', 'emergency response team' and 'number of beds per 1000 person' - 14.28% data was missing. For 17 states, the annual price of diesel was missing (14.28%). 9.5% data for the 'physicians for 10000 persons' and 4.76% data for the percentage of male driver, percentage of female driver and road improvement capital expenditures were missing.

In terms of affecting the reliability and the validity of the results, missing data can cause decrease in statistical power and bias parameter estimates (Roth 1994). Statistical power refers to the ability to discover a relationship in a set of data. In addition, ignoring missing data causes loss of information and increases standard errors associated with

parameters (Peng et al. 2006). The matter which needs to be considered is the proportion of missing data as it is directly related with the statistical inferences. Dong and Peng (2013)asserted that a threshold for an acceptable percentage of missing data has not been reviewed yet. However, Schafer (1999) pointed out that a missing data below 5% is inconsequential. Tabachnick, Fidell, and Osterlind (2001) stated that the way missing data is handled and its pattern have a big impact on the result. The pros and cons of different missing data mechanisms is discussed in the next paragraph. The pattern which exhibited in the dataset was an arbitrary missing pattern.

Before choosing an imputation method for missing data analysis, it was important consider the nature of missingness within the dataset. Simple methods such as list-wise and pairwise deletion, mean and random imputation, regression and stochastic regression imputation, last observation carried forward (LOCF) and baseline observation carried forward (BOCF) exhibit different types of disadvantages. A summary is given below:

Methods	Disadvantages
Listwise Deletion	If the data are not MCAR, the process can severly bias estimates of means, regression coefficients, and correlations.
Pairwise Deletion	Estimates become biased if the data are not MCAR and correlations outside the range [-1,+1] can occur.
Mean Imputation	Distorts the distribution of the dataset, underestimates the variance, disturb the relations between variables when data are not MCAR
Regression Imputation	Yields deterministic data but ignores predictive uncertainty
Stochastic Regression Imputation	Problems arise with negative values, does not cover the high end of data distribution
LOCF & BOFC	To be used to Monotone structured missing datasets, can yield biased estimates under MCAR situations

Figure 4.1. Missing data mechanisms conditional demerits (Source: Van Buuren (2012), Page 16, Table 1.1)

4.2. Steps Involved in Multiple Imputation Procedure

The first step in any data imputation procedure involves assessing with the incomplete dataset. In this thesiss R software was used as a platform and the MICE package was used to impute the incomplete dataset. The multiple Imputation method substitutes the missing value in a dataset for each of the variables, with plausible solutions to create a complete dataset. Therefore, a single imputed dataset does not incorporate the predictive uncertainty.

4.2.1. Number of Imputation and Method Selection

It is important to note that, there is a significant relationship between the amount of missing data and the number of imputations (m) of the incomplete dataset. Rubin (2004) illustrated that the efficiency of an estimate based on m imputations has a close relationship relative to an estimate based on an infinite number of imputations. The relation is given by:

$$T_m = (1+\gamma/m) * T_\infty, \qquad \text{(Equation 4.1)}$$

In equation 4.1, m = the number of imputations,

γ = the true population fraction of missing information,

T_m = the total variance for 'm' number imputations, and

T_∞ = the total variance for ∞ number of imputations.

When m = ∞, the total variance of the complete dataset and the pooled variance of an infinite number of imputed dataset are the same, since the infinite number of imputation causes all error to disappear (Van Buuren 2012). In this thesis, the maximum

number of missing data for a single variable was nine among 21 data points. So, the γ value is 9/21 or 0.42857. The number of imputation taken for this thesis was 10, based on the analysis that the calculated variance for T_m would be 1.0428 times or (4.28%) larger than the ideal variance. The calculation is shown below:

$$T_m = (1 + 0.42857/10) * T_\infty = 1.0428 * T_\infty$$

The confidence interval is thus $\sqrt{1.0428} = 1.0211$ (i.e. 2.11), the maximum confidence interval associated with this thesis. Schafer and Graham (2002) showed that with 42.86% missing information, 10 imputations is 100/ (1 + .042857) or 95.89% efficient for the maximum number of missing data points for a single variable. For the other variables with lesser number of missing values, the overall efficiency would be higher. Although a higher value of m yields better results, in terms of variance & confidence intervals, additional imputation analysis produces little advantage over lesser number of imputation analysis (Schafer and Olsen 1998).

For creating m number of complete datasets with imputed data points, this thesis used R software and MICE package. The default method which is given by *"pmm"* or *predictive mean method*, was not chosen for the non-monotone structure of the dataset; *fast pmm* was not chosen for the same reason. Also, since the data structure was not hierarchical or nested, methods such as *2l. norm, 2l.pan, 2l only.mean, 2l only.norm* or *2l only.pmm* were not chosen. Methods such as *log.reg, log reg.boot, polyreg, Ida* and *polyr* are related with imputing categorical variables or factor level variables and there were no categorical or factor level variables which required imputation. A tree model regression based approach, such as *cart*, was considered for the analysis method and R_f was not

considered. Among the remaining methods, such as *norm, norm.nob, norm.boot, mean* was not chosen since it distorts the data distribution, *sample* was not chosen since imputed value is confined within current non-missing data sample, instead the *norm* method which is based on Bayesian linear regression analysis was chosen. To the authors best knowledge, the relative comparison between such methods within MICE package, has not been reviewed or researched.

4.2.2. Multiple Imputed Dataset and Variable Selection

After the imputation step is accomplished, there are now m number of complete datasets with imputed data points. The second step requires the estimation of parameters from these m imputed datasets. This means that each of these datasets needs to be individually analyzed to obtain parameters of interest such as intercepts, standard errors, the confidence interval, etc. The difference in results can be explained by the difference in imputed data which incorporates the uncertainty.

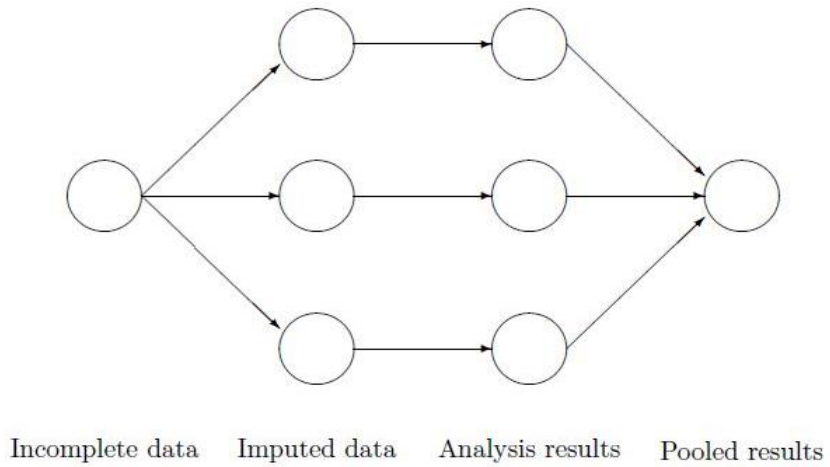

Figure 4.2. The steps involved in multiple imputation analysis

As previously discussed in the literature review, in analytical models the presence of multi-collinearity among variables overestimates the parameters if variables are present in the regression equation. Because of this, a step-wise elimination of variables was made in R software based on the extremely high variance inflation factor (VIF) value. O'Brien (2007)stated that for measuring the multi-collinearity of the *i*th independent variable with other variables, both VIF and tolerance are widely used and accepted.

While analyzing the results from these different datasets, it was observed that the variables differed from dataset to dataset after a screening procedure with the VIF threshold as 5 for strictly low multi-collinearity. With ten new completed datasets obtained after the imputation, the maximum number of times a variable could appear was ten. For example, analyzing the multiple imputed datasets for the state *Montana* revealed the presence of the variables illustrated below:

Table 4.2. Variable appearance for ten newly imputed datasets for Montana

Name of the Variable	Number of Appearance
Precipitation	10
Speeding	10
Temperature	10
Total number of hospitals	6
Emergency Response Team	8
Rural Population	10
BAC Level Law	7
Distracting Driving	8
Physicians	5
Male Driver	6
Female Driver, Road Improvement Expenditure	4
Unemployed, Number of Nurses	3
Drunk Driving	2
Beds per 1000 persons	1

The variation in datasets caused the regression estimates and coefficients to vary significantly. Positive (green arrow) and negative (red arrow) coefficients appeared in the regression model for the same variable according to the dataset. For the state of Montana, this is illustrated below:

Model Number and Variable Description	Model 1 Estimate	Model 2 Estimate	Model 3 Estimate	Model 4 Estimate	Model 5 Estimate	Model 6 Estimate	Model 7 Estimate	Model 8 Estimate	Model 9 Estimate	Model 10 Estimate
(Intercept)	1388.35	-1853.6	-231.5	-2411.4	-40.336	2770.34	1851.29	-2082.1	523.908	871.641
Temperature	-2.252	4.892	4.823	7.892	5.169	-0.496	1.682	2.669	4.418	5.163
Precipitation	0.195	3.269	0.722	3.615	3.58	2.411	0.626	3.578	2.275	1,551
BAC Level Law	-0.329	-16.226	4.306	-12.575			2.499		-9.123	-9.6
Speeding	-11.133	1.247	20.903	11.415	87.167	100.161	-23.092	112.118	18.834	18.983
Distracted Driving	91.622	248.467	229.961	260.696			182.64		264.069	276.66
Total Number of Hospitals	0.994		6.104		3.55		0.119	2.596	2.959	
Emergency Response Team	-0.073	-0.114		-0.097	0.012	-0.003	-0.103		-0.088	-0.131
Rural Population	0.262	0.198		0.414	0.397	0.019	-0.203	0.73	-0.127	-0.344
Male Driver	-22.279			41.445	-5.886	-45.78			-10.334	
Unemployed	-0.001			0.001		-0.002				
Physicians		-0.265	-0.813	0.891				-2.117		2.005
Female Driver		37.409					-28.459	33.686		-13.08
Road Improvement Capital Expenditures			-1E-05	-6E-06	-5E-06			-1E-05		
Number of Nurses					-0.013	-0.022	-0.01			
Drunk Driving					95.609	62.412		146.774		
Beds per 1000 persons									-2.469	

Figure 4.3. Parameter estimates of ten regression models for dependent variable fatalities

One of the challenges regarding the multiple imputation is the rigorous process of computation. For ten imputed datasets, the results varied significantly, as it can be seen from the table illustrated above. An important question that needs clarification is that which variables to select and procedure for pooling all the results.

The variable selection process involve doing stepwise model selection separately for each imputed dataset and then creating another model containing the variables which appeared at least 50% of the time (Brand 1999). This way the selected variables do not

appear to have seen selected accidentally. Also, a backward elimination of variables is applied to the new set of variables by checking *p*-value from the pooled likelihood ratio. It is generalized that if the *p*-value is larger than 0.05, the corresponding variable should be removed from the super model, but it should be selected if the *p*-value is smaller than 0.05. This procedure is repeated until it has been applied to all the variables.

Studies conducted by Vergouwe et al. (2010) and Wood, White, and Royston (2008) distinguished three different approaches for selecting variables:

1. Majority – selecting variables which appear appearing in half of the models;
2. Stack – assigning a fixed weight to each dataset after stacking m datasets into one super dataset; and
3. Wald Test – selecting variables based on Wald statistics

Evaluating the three approaches, the Majority method is identical to step 1 of Brand (1999), whereas the Wald Test method is similar to Brand's step 2, with the difference being the application of the Wald Test instead of the likelihood ratio test. The Wald Test is a recommended procedure since it is well established and follows Rubin's rules, but the Majority and Stack methods fail to consider the uncertainty caused by missing data. This thesis has used a combinational approach in selecting variables based on the Majority method and the Wald Test method. For this study, the Majority method was applied to the primary selection of the variables for the state of Montana, and the variables which appeared in at least half of the regression models remained to be analyzed using the Wald's test.

Table 4.3. After applying the majority rule of variable selection

Name of the Variable	Number of Appearance
Precipitation	10
Temperature	10
Speeding	10
BAC Level Law	7
Distracted Driving	8
Total number of hospitals	6
Rural Population	10
Emergency Response Team	8
Male Driver	5
Physicians	5

The variables which appeared consistently in all ten models were selected to be the final variables. Buuren and Groothuis-Oudshoorn (2011) have shown that in R software platform after the imputation of ten datasets, the MICE package has a special class object called *mids (multiple imputed dataset)* in which the new datasets can be are stored. A snapshot is given from the R software platform:

```
                        est         se           t        df  Pr(>|t|)        lo 95      hi 95 nmis       fmi     lambda
(Intercept)      -195.6839480 328.295791 -0.59605988 10.337992 0.5639576 -923.9437780 532.575882   NA 0.3461366 0.23079640
Temperature         5.6233766   4.851860  1.15901458 13.861963 0.2660270   -4.7925584  16.039312    0 0.1453404 0.03032743
Precipitation      -1.3312752   2.997919 -0.44406649 12.943649 0.6643279   -7.8107522   5.148202    0 0.2007444 0.08610364
Speeding            7.1309266  73.062630  0.09760019 11.824088 0.9238847 -152.3219369 166.583790    0 0.2639665 0.14917732
Rural.Population    0.4357778   0.463869  0.93944149  6.032267 0.3835811   -0.6977986   1.569354    9 0.6072823 0.49559202
> pool.r.squared(fit)
          est       lo 95     hi 95       fmi
R^2 0.2975858 0.001019421 0.6908935 0.3837044
> pool.r.squared(fit,adjusted = TRUE)
        est lo 95 hi 95 fmi
adj R^2 NaN   NaN   NaN  NA
```

Figure 4.4. Initial stage towards selecting the final variables for pooled estimates

In the Figure 4.4, the dependent variable ‘fatalities’ is a function of five independent variables – precipitation, speeding, rural population, and number of unemployed. *Montanaz* is a *mids* object in which 10 imputed datasets are stored and is

initiated to capture all the pooled estimates. The next steps are to checking whether other variables pass the Wald's test.

```
### Second Step - Distracted.Driving/Emergency.Response.Team/factor(BAC.Level.Law)/Total.Number.of.Hospitals
fit1<-with(Montanaz, lm(Fatalities ~ Temperature+Precipitation+Speeding+Rural.Population+Distracted.Driving))
est1<-pool(fit1)
summary(est1)
pool.r.squared(fit1)
pool.r.squared(fit1,adjusted = TRUE)
pool.compare(fit1,fit,method = "Wald")$pvalue [,1]
```

Figure 4.5. Checking Wald's test statistic for possible significance of the variables

In the illustration, the new variable *Distracted Driving* is added to the model. To check whether it is significant, in the last line of coding, the *pool.compare()* function is utilized to compare the two nested models (fit1 and fit) on multiple imputed datasets *(mids)* Montanaz. The *pool.compare()* function with the "Wald" method pools the *p*-values for comparing the nested models and checks the overall significance using the method proposed by Li et al. (Li et al. 1991).The Wald test revealed that, the *p*-value in this case was 0.029. So, the variable *Distracted Driving* was selected for the final model.

```
> pool.compare(fit1,fit, method = "Wald")$pvalue [,1]
[1] 0.02943201
>
```

Figure 4.6. Wald's test revealing distracted driving factor is non-significant

The Wald's test was similarly carried out for several other two variables: number of emergency response team, BAC level laws, total number of hospitals, number of physicians, and male driver. The results revealed that the variable – total number of hospitals was significant and thus needed to be included in the final model.

```
fit2<-with(Montanaz, lm(Fatalities ~ Temperature+Precipitation+Speeding+Rural.Population+Distracted.Driving
                        +Total.Number.of.Hospitals))
est2<-pool(fit2)
summary(est2)
pool.r.squared(fit2)
pool.r.squared(fit2,adjusted = TRUE)
pool.compare(fit2,fit1,method = "Wald")$pvalue [,1]
```

Figure 4.7. Total number of hospitals is included in the model for Wald's test

```
> pool.compare(fit2,fit1,method = "Wald")$pvalue [,1]
[1] 0.04444589
>
```

Figure 4.8. Wald's test indicating total number of hospital variable is significant

After including the variable – total number of hospital, in the model, the process was repeated with the other variables – emergency response team, BAC level law, male driver, and physician to see whether the variable was also significant or not.

```
fit3<-with(Montanaz, lm(Fatalities ~ Temperature+Precipitation+Speeding+Rural.Population+Distracted.Driving
                        +Total.Number.of.Hospitals+Physicians))
est3<-pool(fit3)
summary(est3)
pool.r.squared(fit3)
pool.r.squared(fit3,adjusted = TRUE)
pool.compare(fit3,fit2,method = "Wald")$pvalue [,1]
```

Figure 4.9. Checking the significance of the variable Physicians

It was found that, the variable – 'physicians' is not a significant variable for the analysis. Wald's test statistic revealed a *p*-value of 0.7715. It is important to note that in the *pool.compare ()* function, the two nested models are now fit2 & fit1, since total number of hospital variable was included in the model, all the other models will now have to be compared with fit1. Since there are not anymore variables, fit1 is the final model. The variables which are finalized for the pooled estimation stage are given below:

Table 4.4. Variable selection finalization for pooled estimation

Name of the Variable	Number of Appearance
Precipitation	10
Speeding	10
Temperature	10
Total number of hospitals	6
Rural Population	10
Distracted Driving	8

4.2.3. Rubin's Rule and Pooled Estimation

The final step of missing data analysis requires pooling m parameter estimates obtained from the m different models into one estimate and associating a variance with the estimate. To calculate the variance correctly, the conventional sampling variance or within-imputation variance and between imputation variance caused by the missing data are needed. Rubin (1987) developed a simple method for combining the results of separate estimates and standard errors from m datasets into a pooled estimate with standard error, confidence intervals and *p*-values. In the MICE package, these pooled estimation can be obtained and implemented by using *pool ()* and *pool.scalar ()* codes.

To put it simply, if an inference is to be made from the ten imputed datasets, regression coefficients should be obtained at first for all the variables which were finalized. Assuming that the regression coefficients for the *i*-th variable are $\hat{\beta_1}, \hat{\beta_2}, \ldots \hat{\beta_m}$ and the standard errors are $s_1, s_2, \ldots s_m$. the point estimate for the regression coefficient is given by:

$$\hat{\beta} = \frac{1}{m} \sum_{m=1}^{M} \hat{\beta_m} \qquad \text{(Equation 4.2)}$$

The variance within imputations is given by the equation below:

$$W = \frac{1}{m}\sum_{m=1}^{M} s_m^2 \qquad \text{(Equation 4.3)}$$

The variance between imputations is given by the equation below:

$$B = \frac{1}{m-1}\sum_{m=1}^{M}(\hat{\beta}_m - \hat{\beta})^2 \qquad \text{(Equation 4.4)}$$

The total variance estimate V_β is given by the equation below:

$$V_\beta = W + \left(1 + \frac{1}{m}\right) * B \qquad \text{(Equation 4.5)}$$

It is important to note that, the square root of V_β will ultimately result in the calculation of the overall standard error. All the estimates from the individual regression models were captured for the state of Montana. A summary of the parameter estimates (intercept β_0 and coefficients $\hat{\beta}_m$) and standard errors are given below in a tabular form:

Model Number and Variable Description	Model 1 Estimate	Model 2 Estimate	Model 3 Estimate	Model 4 Estimate	Model 5 Estimate	Model 6 Estimate	Model 7 Estimate	Model 8 Estimate	Model 9 Estimate	Model 10 Estimate
(Intercept)	-321	-420	-273	-411	-415	-205	-128	-489	-182	-132.5
Precipitation	0.546	1.793	0.74	1.312	1.302	-0.05	0.225	1.657	0.544	-0.473
Speeding	11.65	-43.3	0.517	-101	-48.4	-15.8	-56.7	-10.9	-38.6	-49.27
Distracted.Driving	122.7	203.8	230	198.1	181.4	247.7	218.1	157	238.5	216.32
Rural.Population	0.301	0.259	-0.09	0.488	0.317	0.038	-0.04	0.609	-0.14	-0.071
Temperature	2.34	3.404	4.355	3.869	4.52	6.033	2.22	3.487	3.472	2.979
Total.Number.of.Hospitals	5.615	6.617	6.376	4.257	5.274	2.786	5.343	4.581	6.039	5.329

Figure 4.10. Summary of coefficients from ten imputed datasets

the pooled coefficients $\hat{\beta}$, W, B & V_β have been calculated in accordance with Rubin's rule for the pooled estimation procedure and the equations 4.2, 4.3, 4.4 and 4.5. A summary is given below:

Table 4.5. Summary of Pooled Estimates

Model Parameters	Avg. Estimate	Variation Within Imputation	Between imputation	Total variance	Standard Errors
(Intercept)	-297.4854	67482.2	17459.332	84941.528	291.4472988
Precipitation	0.76	5.746416575	0.557685333	6.3041	2.510796686
Speeding	-35.1382	4293.048	1071.239509	5364.288	73.24129982
Distracted Driving	201.3557	8629.832	1492.937763	10122.77454	100.612
Rural Population	0.1681	0.0882102	0.068705211	0.15691454	0.3961244
Temperature	3.6679	14.612624	1.256375656	15.689	3.960934233
Total.Number.of.Hospitals	5.2217	5.268704	1.267186011	6.53589	2.556538676

The values for the average point estimates and standard errors were very similar to what had been obtained from the *pool ()* function after the analysis was made. A snapshot of that result is given below for better understanding:

```
                              est          se          t        df   Pr(>|t|)        lo 95      hi 95 nmis
(Intercept)          -297.4854238 291.4473702 -1.0207175  9.067562 0.33384787 -956.0371101 361.066263   NA
Temperature             3.6678797   3.9619933  0.9257663 11.156983 0.37415871   -5.0374622  12.373222    0
Precipitation           0.7600228   2.5108146  0.3026997 11.021601 0.76775511   -4.7649218   6.284967    0
Speeding              -35.1383402  73.2413752 -0.4797608  9.165688 0.64263889 -200.3664117 130.089731    0
Rural.Population        0.1681234   0.3961244  0.4244208  5.497493 0.68736160   -0.8230451   1.159292    9
Distracted.Driving    201.3557283 100.6128459  2.0012924 10.044968 0.07310283  -22.6877049 425.399162    0
Total.Number.of.Hospitals   5.2216852   2.5565495  2.0424738  9.263975 0.07059013   -0.5366081  10.979979    5
>
```

Figure 4.11. Pooled estimates illustrating regression coefficients and standard errors

From the analysis, the most significant variable which were *distracted driving* and *total number of hospitals* variable at a significance level of 0.01. If the missing data approach had not been taken, the variable – *total number of hospitals* would not have been considered and the model fit would have been less. A comparison is given below for better understanding:

Table 4.6. Comparison of Results illustrating the importance of missing variables

Description of Comparison	Fit Value	est	lo 95	hi 95
Without - Total number of Hospital	$R^{\wedge 2}$	0.49296	0.09694	0.7668
	$R^{\wedge 2}_{\text{Adj}}$	0.3224	0.02165	0.6635
With - Total number of Hospital	$R^{\wedge 2}$	0.63501	0.27895	0.84735
	$R^{\wedge 2}_{\text{Adj}}$	0.47702	0.10301	0.77039

Even though there are some non-significant variables present in the pooled results, a stepwise regression analysis was utilized to obtain the significant variables for the final prediction model. A stepwise analysis approach revealed that for the states of Montana, distracted driving and total number of hospitals were the two most significant factors ($p<0.05$). The final regression model can be expressed as:

*Fatalities = -79.007 -197.3895*Distracted Driving + 5.77 * Total Number of Hospitals*

(Equation 4.6)

The model fit of the regression equation was obtained as $R^{\wedge 2}$ at 56.25 % and $R^{\wedge 2}_{\text{adj}}$ at 51.35%. An illustration is given below which shows the pooled estimates and statistical significances:

```
                              est         se          t       df   Pr(>|t|)        lo 95     hi 95
(Intercept)            -79.007520 111.619581 -0.7078285  9.57378 0.49592183 -329.2204641 171.20542
Distracted.Driving     197.389549  75.442701  2.6164168 12.45991 0.02195366   33.6844235 361.09467
Total.Number.of.Hospitals   5.771651   2.214121  2.6067461  9.43803 0.02737791    0.7981736  10.74513
> pool.r.squared(fit2)
          est     lo 95    hi 95       fmi
R^2 0.5625774 0.1885625 0.812995 0.1783332
> pool.r.squared(fit2,adjusted = TRUE)
             est     lo 95   hi 95       fmi
adj R^2 0.5135047 0.1367028 0.78881 0.1937063
```

Figure 4.12. Final Regression Model for Fatality Prediction for the States of Montana

4.2.4 Summary of the Prediction Models for All 50 States

The prediction models for the rest of the states were obtained using the same analysis procedure. A summary of the results is given below, including the prediction models for all the 50 states as well as any significant factors found from the regression parameter estimations. The details of these results have been attached in the appendix (Appendix B).

Table 4.7. Prediction Model for Fatalities of 50 states

State	Prediction Model for Fatalities	Significant Factors
AL	Fatality = 502.872 – 0.0212 * Unemployed + 7.187 * Total number of hospitals	Unemployed, Total number of hospitals
AK	Fatality = 86.25 - 20.36*Child Restraint Law	Child Restraint Law
AZ	The data does not fit well and no significant variables were found	
AR	Fatality = 215.0203+ 1789.14*Drunk Driving	Drunk Driving
CA	Fatality = 1833.19 + 10409.26*Drunk Driving +6864.81* Distracted Driving - 0.000667* Unemployed	Drunk Driving, Distracted Driving, Unemployed
CO	Fatality = 342.31 + 2623.15 * Distracted Driving	Distracted Driving
CT	Fatality = 473.52 – 559.1010 * Drunk Driving	Drunk Driving
DE	Fatality = 139.04 - 0.0000932*Road Improvement Expenditures	Road Improvement Expenditures
FL	Fatality = 2209.58 – 403.18 * Texting while driving law + 4694.18 * Drunk Driving + 1648.93 * Distracted Driving – 0.00118 * Unemployed	Texting while driving law, Drunk Driving, Distraced Driving, Unemployed
GA	Fatality = 1664.75 – 18607.27 * Distracted Driving	Distracted Driving
HI	Fatality = -376.58 + 7.18 * Temperature – 32.98 * Handheld phones while driving law – 0.00094* Unemployed	Temperature, Handheld phones while driving law, Unemployed
ID	Fatality = 313.14 -45*Texting while driving law- 0.00157 * Unemployed	Texting while driving law, Unemployed
IL	Fatality = 2056.18 – 1689.09 * Speeding – 6027.48 * Distracted Driving – 0.0191 * Emergency Response Team – 226.86 * Cellphone use while driving for novice or teen driver law	Speeding, Distracted Driving, ER Team, Cellphone law
IN	Fatality = 780.87 + 1140.59 * Drunk Driving – 0.000703 * Unemployed – 71.4 * BAC level law	Drunk Driving, Unemployed, BAC level law
IA	Fatality = 582.09 - 0.000274 * Road Improvement Expenditures	Road Improvement Expenditures
KS	Fatality = 726.63 – 0.0787 * Emergency Response Team	ER Team
KY	Fatality = 514.669 - 117.42*Cellphone use while driving for Novice and Teen Driver Law + 1452.703* Drunk Driving	Cellphone law, Drunk Driving

Table 4.7. Prediction Model for Fatalities of 50 states – continued

State	Prediction Model for Fatalities	Significant Factors
LA	Fatality = 1068.85 – 0.00165 * Unemployed + 0.0000881 * Road Improvement Expenditures – 198.17 * Cellphone use while driving for novice or teen drivers law	Unemployed, Road Improvement Expenditures, Cellphone law
ME	Fatality = 136.77 + 262.405 * Distracted Driving	Distracted Driving
MD	Fatality = 622.375 – 142.575 * Handheld phones while driving law	Handheld phones while driving law
MA	Fatality = 429.81 – 73.21 * Cellphone use while driving for novice or teen drivers law	Cellphone law
MI	Fatality = 1654.61 – 252.75 * Cellphone use while driving for novice or teen drivers law- 4765.99 * Distracted Driving – 0.001127 * Unemployed	Cellphone law, Distracted Driving, Unemployed
MN	Fatality = -19.72 + 2291.8 * Drunk Driving	Drunk Driving
MS	Fatality = 740.956 – 137.31 * Texting while driving law + 504.67 * Distracted Driving	Texting while driving law, Distracted Driving
MO	Fatality = 6016.523 – 86.349 * Female Driver – 0.0602 * Emergency Response Team	Female Driver, ER Team
MT	Fatality = -79.007 + 197.39 * Distracted Driving + 5.77 * Total number of hospitals	Distracted Driving, Total Number of Hospitals
NE	Fatality = 349.745 – 538.86 * Speeding – 66.33 * Cellphone use while driving for novice or teen drivers law	Speeding, Cellphone law
NV	Fatality = 154.405 - 37.88 * BAC Level Law + 796.39 * Drunk Driving – 376.44 * Speeding	BAC Level Law, Drunk Driving, Speeding
NH	Fatality = 72.725 + 208.94 * Drunk Driving	Drunk Driving
NJ	Fatality = 713.197 + 740.37 * Drunk Driving – 272.95 * Distracted Driving – 0.0000962 * Road Improvement Expenditures	Drunk Driving, Distracted Driving, Road Improvement Expenditures
NM	Fatality = 605.89+ 8.203 * Precipitation + 551.43* Speeding - 0.0177 * Million VMT	Precipitation, Speeding, Million VMT
NY	Fatality = 2659.48 – 1527.34 * Speeding – 2473.308 * Drunk Driving – 0.000132 * Road Improvement Expenditures	Speeding, Drunk Driving, Road Improvement Expenditures
NC	Fatality = 733.698 + 3294.31 * Drunk Driving – 101.0008 * Cellphone use while driving for novice or teen drivers law	Drunk Driving, Cellphone law
ND	Fatality = -73.638 + 0.0193 * Number of Nurses	Number of Nurses
OH	Fatality = 1889.19 – 226.45 * Cellphone use while driving for novice or teen drivers law – 6.29 * Precipitation – 0.000998 * Unemployed	Cellphone law, Precipitation, Unemployed
OK	The data does not fit well and no significant variables were found	
OR	Fatality = 305.4 + 4.46 * Precipitation + 628.14 * Drunk Driving – 0.0001678 * Road Improvement Expenditures – 0.0007044 * Unemployed	Precipitation, Drunk Driving, Road Improvement Expenditures, Unemployed
PA	Fatality = 1077.956 + 2013.005 * Distracted Driving – 937.52 * Speeding + 2148.202 * Drunk Driving	Distracted Driving, Speeding, Drunk Driving
RI	Fatality = 100.06 – 59.05 * Speeding - 0.0000864* Road Improvement Expenditures	Speeding, Road Improvement Expenditures
SC	Fatality = 779.90 + 1376.74 * Drunk Driving – 0.00182 * Unemployed	Drunk Driving, Unemployed
SD	Fatality = -4660.32 + 96.39 * Female Driver	Female Driver

Table 4.7. Prediction Model for Fatalities of 50 states – continued

State	Prediction Model for Fatalities	Significant Factors
TN	Fatality = 1512.96 – 0.1022 * Rural Population – 118.44 * Cellphone use while driving for novice or teen drivers law – 1153.8 * Speeding	Cellphone law, Speeding, Rural Population
TX	Fatality = 3869.877 – 7931.225 * Distracted Driving	Distracted Driving
UT	Fatality = 174.43 + 769.633 * Drunk Driving	Drunk Driving
VT	Fatality = -464.73 – 0.844 * Precipitation – 73.15 * Speeding + 11.84 * Male. Driver	Precipitation, Speeding, Male Driver
VA	Fatality = 7282.8 – 0.000304 * Road Improvement Expenditures – 119.49 * Female Driver	Road Improvement Expenditures, Female Driver
WA	Fatality = 719.22 – 0.000422 * Unemployed – 130.47 * Handheld phones while driving law	Unemployed, Handheld phones while driving law
WV	Fatality = 481.97 – 58.117 * Handheld phones while driving law – 0.00198 * Unemployed	Unemployed, Handheld phones while driving law
WI	Fatality = 15421.75 – 123.15 * Cellphone use while driving for novice or teen drivers law – 0.0013 * Unemployed – 292.74 * Male Driver + 652.81 * Drunk Driving	Cellphone law, Unemployed, Male Driver, Drunk Driving
WY	Fatality = 56.77 + 286.065 * Drunk Driving + 345.007 * Distracted Driving	Drunk Driving, Distracted Driving

It was observed that variables relating to risky behavior speeding, distracted driving, and drunk driving were highly significant in some cases and were among the top reasons. The total number of unemployed people as well as the laws regarding road safety were also highly significant in explaining the variation for road fatalities. An illustration is given below that highlights the significant variables and the frequency as well as how they are associated with the dependent variable:

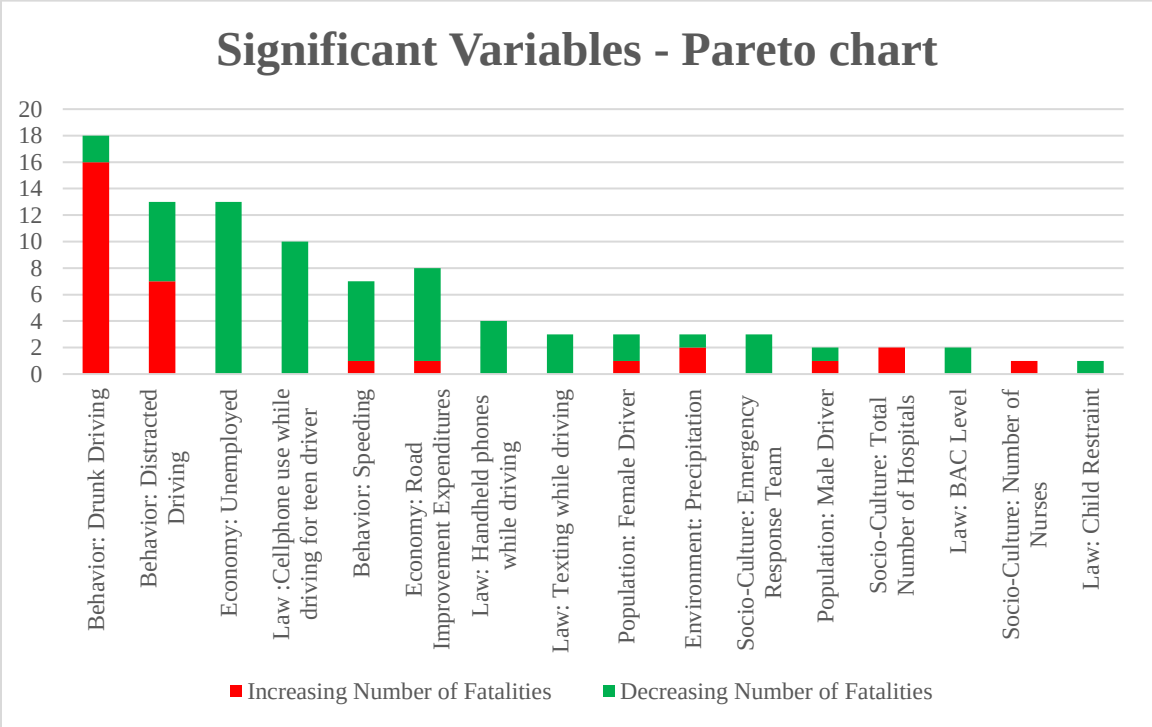

Figure 4.13. Pareto analysis of significant variables explaining fatalities overall

Depending on the regression models obtained it can be observed that some states require more than one variable to capture the variation of the dependent variable. The maximum number of variables which were found to be significant in a regression model is four. An illustration is give below to represent the frequency of the states versus the number of variables present in their respective regression models:

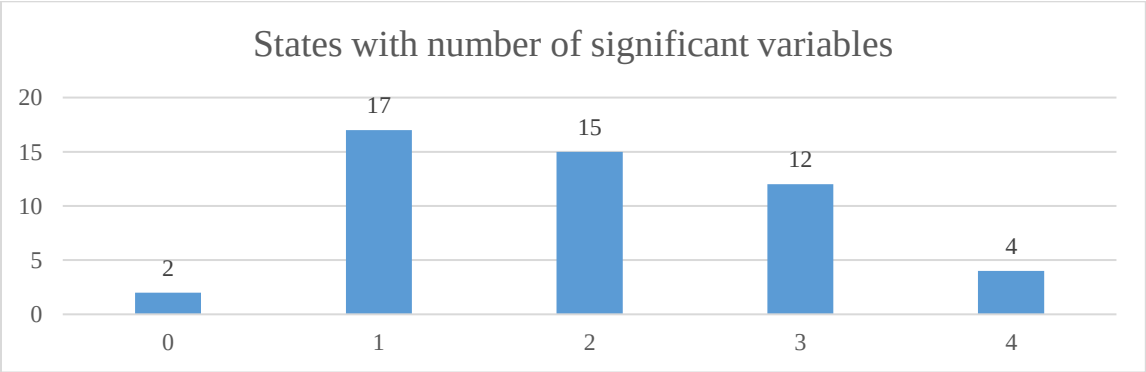

Figure 4.14. Frequency distribution of states with number of significant variables

4.2.5. Estimate Fatality from Prediction Model & Validation

After obtaining the prediction model for all 50 states by pooling estimates and parameters from 10 different datasets, the fatalities were estimated. In this case, the number of fatality was calculated for all 10 datasets. The average point estimate for the number of fatality along with the standard deviation were also calculated for better explanation purposes. A self-explanatory comparative illustration is given below for the state of Montana:

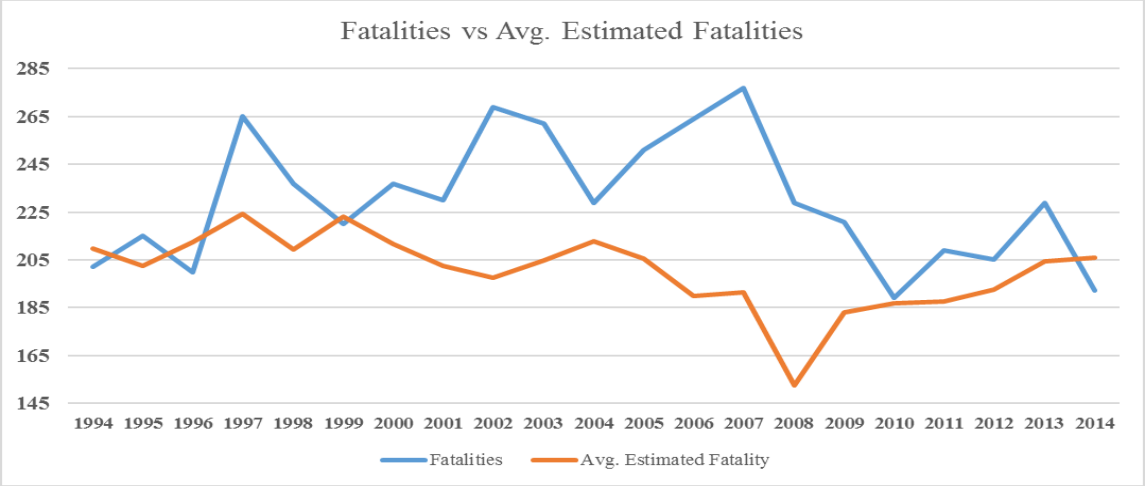

Figure 4.15. Comparing actual fatality with estimated avg. fatality from imputed datasets

A summary of the estimated values from the prediction model is given below for a comprehensive understanding of the overall measures. The major variation in estimation occurred where the missing data was imputed.

Fatalities	Predicted y1	Predicted y2	Predicted y3	Predicted y4	Predicted y5	Predicted y6	Predicted y7	Predicted y8	Predicted y9	Predicted y10	Avg. Estimated Fatality	Standard dev
202	197.64	207.02	216.52	192.59	212.60	245.63	204.74	222.04	204.95	194.84	209.86	15.68
215	208.31	190.73	217.34	163.16	193.78	244.99	188.57	202.35	203.28	213.23	202.57	21.43
200	200.44	208.60	218.44	193.83	219.73	237.89	204.57	226.02	209.35	204.38	212.32	13.18
265	230.92	221.69	231.95	220.33	225.58	231.32	216.18	227.99	221.68	214.48	224.21	6.31
237	210.75	212.73	210.60	194.47	206.77	209.38	214.75	224.70	203.06	204.88	209.21	7.94
220	223.15	223.15	223.15	223.15	223.15	223.15	223.15	223.15	223.15	223.15	223.15	0.00
237	211.77	211.77	211.77	211.77	211.77	211.77	211.77	211.77	211.77	211.77	211.77	0.00
230	202.68	202.68	202.68	202.68	202.68	202.68	202.68	202.68	202.68	202.68	202.68	0.00
269	197.36	197.36	197.36	197.36	197.36	197.36	197.36	197.36	197.36	197.36	197.36	0.00
262	204.79	204.79	204.79	204.79	204.79	204.79	204.79	204.79	204.79	204.79	204.79	0.00
229	213.00	213.00	213.00	213.00	213.00	213.00	213.00	213.00	213.00	213.00	213.00	0.00
251	205.58	205.58	205.58	205.58	205.58	205.58	205.58	205.58	205.58	205.58	205.58	0.00
264	189.78	189.78	189.78	189.78	189.78	189.78	189.78	189.78	189.78	189.78	189.78	0.00
277	191.50	191.50	191.50	191.50	191.50	191.50	191.50	191.50	191.50	191.50	191.50	0.00
229	152.33	152.33	152.33	152.33	152.33	152.33	152.33	152.33	152.33	152.33	152.33	0.00
221	182.88	182.88	182.88	182.88	182.88	182.88	182.88	182.88	182.88	182.88	182.88	0.00
189	186.99	186.99	186.99	186.99	186.99	186.99	186.99	186.99	186.99	186.99	186.99	0.00
209	187.54	187.54	187.54	187.54	187.54	187.54	187.54	187.54	187.54	187.54	187.54	0.00
205	192.42	192.42	192.42	192.42	192.42	192.42	192.42	192.42	192.42	192.42	192.42	0.00
229	204.30	204.30	204.30	204.30	204.30	204.30	204.30	204.30	204.30	204.30	204.30	0.00
192	205.98	205.98	205.98	205.98	205.98	205.98	205.98	205.98	205.98	205.98	205.98	0.00

Figure 4.16. Illustrating all the estimated predictions for road fatalities of Montana

Using the last five data points from the imputed datasets, predicted fatalities were estimated and validated. A summary table is given below:

Table 4.8. Validating estimated results against the actual fatalities

Year	Fatalities	Avg. Estimated Fatalities	Residual
2010	189	186.99	2.01
2011	209	187.54	21.46
2012	205	192.42	12.58
2013	229	204.30	24.70
2014	192	205.98	-13.98

As discussed earlier, the overall adjusted goodness of model fit statistic (${R^2}_{adj}$) is 51.35%. After analyzing the actual fatality versus the average estimated predicted fatality

from the correlation analysis, it was found that the result is not significant. A summary is given below:

Table 4.9. Correlation Analysis between Actual Fatality vs Predicted Fatality

Correlation Analysis	Correlation value	p-value
Pearson Correlation	0.053	0.821
Spearman Correlation	0.039	0.867

Even though for the states of Montana, the correlation analysis did not provide significant results, for the states of Illinois, the correlation result was highly significant. A graphical illustration is given below, where the predicted fatality was compared with actual fatality:

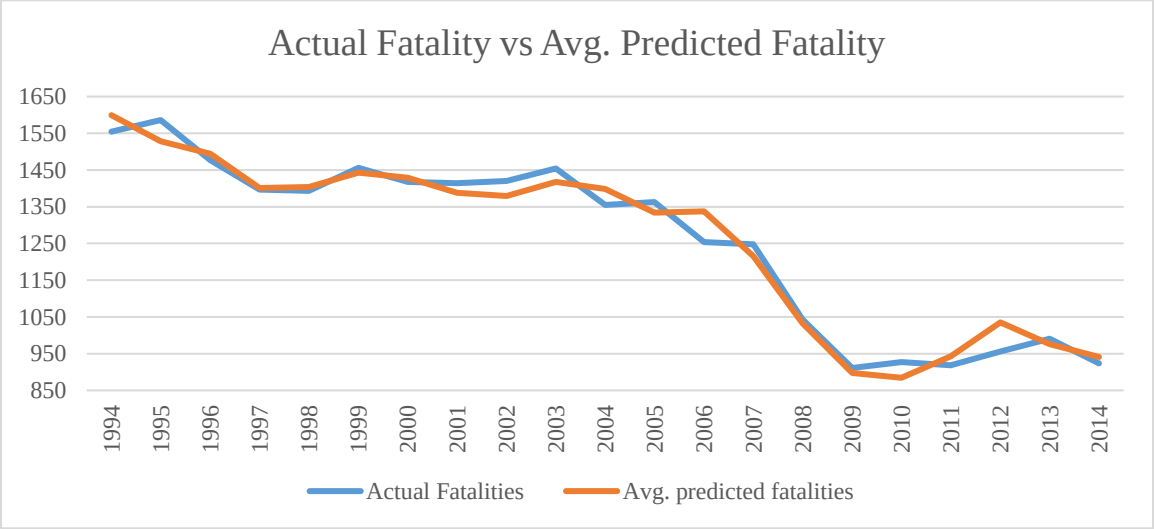

Figure 4.17. Comparative illustration of actual and estimated fatalities for Illinois

The correlation analysis between actual fatalities and estimated fatalities for the state of Illinois was 0.977 and p-value was < 0.0001. The correlation analysis result for the states of Illinois is shown below:

Correlation: Fatality, Avg. fatality (hat)

```
Pearson correlation of Fatality and Avg. fatality (hat) = 0.977
P-Value = 0.000
```

Figure 4.18. Significant correlation result for Illinois state

It was revealed that in terms of correlation between actual number of fatality and predicted fatality 47 states were significant. But the relationship between predicted and actual fatalities for Montana was not correlated.

CHAPTER FIVE – PANEL DATA REGRESSION MODEL ANALYSIS

Chapter three discussed the limitations of US data sources regarding reliable information on state level people's behavior, attitude, norm, belief, culture, and governance quality. This is a drawback as conclusive evidence of their relationship to road fatalities cannot be formulated, in contrast to European road safety research where sources such as ESS and SARTRE are used quite often. Page (2001) in his study, obtained performance indicators of 21 countries of inferred cultural factors of road safety to capture and compare level of safety culture. After a careful review of the literature, no such efforts to measure the cultural progress in terms of road safety culture were found that applies to the 50 states.

Thus, this study offers two solutions. One, infer the progress of road safety culture, norms, attitudes, government policies, road safety performance and other exogenous variables by utilizing the dataset of 50 states ranging 21 years of time frame. And secondly, using four political variables as a proxy for culture, in terms of people's attitude towards politics. These political variables were associated with two major political parties and how people responded towards them during each election.

This chapter is divided into five subsections. The first three sections will describe the modelling approach, analysis & result compilation and state wise comparative illustration of the progression of inferred safety culture. The fourth subsection provides a critique to the study of Page. Finally, the fifth subsection provides the effect of the four political variables as a surrogate for culture by illustrating better goodness of fit statistics without them.

5.1. Panel Data Modelling

Antoniou et al. (2016) provided a systematic classification of different data analysis approach for a dataset that spans cross-sectional number (N) to 30 and possesses a sufficiently long time series. One of the most significant aspect of panel data analysis is to first analyze whether the dataset is a balanced panel data set or an unbalanced panel data set. A balanced data set is defined as a set that contains all elements observed in all time frames and on the other hand, unbalanced data is a set of data where in certain years, the data category is not observed (Baltagi 2008). PROC MI program was used to impute the missing data points and since the number of time series is constant for all the variables over the cross section, the data set is considered a balanced data. The basic structure of panel data regression model is given as such:

$$Y_{it} = \beta_0 + \sum \beta\, X_{it} + u_{it} \qquad \text{(Equation 5.1)}$$

In the equation, i = 1, 2… N (spatial width, represented number of countries, region, state, individual, etc.) & t = 1, 2… T (the time index, in this case t = 21). Y_{it} is the dependent variable or the response variable of state i at time t. β_0 is the regression model intercept and X_{it}& β are described by independent or explanatory variables and regression coefficients associated with its respective explanatory variable. The residual of the model is given by u_{it} and based the nature of assumption being placed on the error term the panel modelling approach is divided into two major categories:

1. Fixed effects(FE) model, and
2. Random effects(RE) model.

Based on the study by Page (2001), this study replicated his methodology by analyzing the data of 50 states. Unobserved variables related to culture, human behavior, and other known-unknowns of which the data is not available, inference was drawn from residuals obtained from the random effects model approach. Stock and Watson (2003) illustrated in FE approach, the dependent variable depends only on the chosen independent variables. Unobserved exogenous variables are assumed to be fixed, and the individual specific effect is correlated with the independent variables. On the other hand, in random effects model, the variation across entities is assumed to be random and uncorrelated with independent variables.

Hausman (1978) presented a test for panel data, which checks whether the errors are correlated with the independent variables or not. The Hausman test distinguishes random effects model from fixed effects model and was used before initiating the analysis procedure, to support the claim of using random effects model. Hausman effect showed *p*-value less than 0.05. This meant testing parameters rejected the null-hypothesis of errors being uncorrelated. This will be presented in the following section.

The international comparative study on road fatality prediction by Page (2001) was done based on Fuller and Battese (1974) method, in which the residuals were separated into three terms, error component for cross sections (v_i), error component for time series (e_t), and idiosyncratic error component between cross sections & time sections (ε_{it}). The expression is given below:

$$u_{it} = v_i + e_t + \varepsilon_{it} \qquad \text{(Equation 5.2)}$$

Where, $v_i, e_t, and\ \varepsilon_{it}$ are independent, with zero mean and constant variance.

5.2. Result Analysis and Compilation

In order to obtain the results for panel data analysis, SAS (v.9.4) software was used in this thesis. Similar to previous analysis in chapter four, the dependent variable was taken to be ‘number of fatalities’. *PROC MI* & *PROC MIANALYZE* were utilized to generate multiple imputed datasets and pooling the estimation, while the *PROC REG*, & *PROC PANEL* commands were used to obtain the necessary results for the stepwise analysis of VIF, model fit, diagnostics results, estimates, and other parameters.

Since the threshold for VIF was taken to be a strict value of five, the variable selected from the preliminary analysis showed that, there were some variables even though had passed the VIF threshold were non-significant to the result. In a stepwise manner, such variables were taken out.

From one of the imputed dataset, the initial analysis with proc panel revealed a moderate fit with the selected variables which include economic, political, and geographical factors with global F-statistic showing p value < 0.0001. This means hypothesis of the overall panel regression model could only be described by the intercept is rejected. An illustration is given that shows preliminary variable selection process based on VIF values:

Parameter Estimates							
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Tolerance	Variance Inflation
Intercept	1	-542.50620	309.43037	-1.75	0.0799	.	0
GDP	1	-0.00413	0.00088562	-4.66	<.0001	0.36960	2.70566
UnempRate	1	-6.84279	3.71549	-1.84	0.0658	0.59452	1.68204
RuralRoad	1	0.00146	0.00019974	7.30	<.0001	0.46386	2.15580
Temp	1	13.42556	0.87827	15.29	<.0001	0.48111	2.07851
AlCons	1	112.74367	13.20337	8.54	<.0001	0.71445	1.39968
SBLaw	1	-24.67365	13.36561	-1.85	0.0652	0.66248	1.50949
TextingLaw	1	-98.30185	20.83724	-4.72	<.0001	0.37479	2.66814
CellLaw	1	-37.25132	19.42063	-1.92	0.0554	0.44796	2.23233
HandheldLaw	1	-129.37028	25.56448	-5.06	<.0001	0.73369	1.36298
DisDriving	1	-167.77233	70.30531	-2.39	0.0172	0.94279	1.06068
Beds	1	-2.80802	7.12654	-0.39	0.6936	0.61422	1.62808
MilVMT	1	0.01226	0.00014063	87.15	<.0001	0.45170	2.21385
MD	1	7.94873	5.15565	1.54	0.1234	0.67461	1.48233
DemorRep	1	-48.88707	19.72263	-2.48	0.0133	0.30292	3.30125
Dem	1	-6.42471	1.39074	-4.62	<.0001	0.20146	4.96385
Rep	1	-2.72175	1.20678	-2.26	0.0243	0.21823	4.58241
VotePer	1	-148.70493	62.48216	-2.38	0.0175	0.70233	1.42384

Figure 5.1. Stepwise elimination of variable illustrating VIF values within threshold

Since multi-collinearity is associated with independent variable and they remain identical *PROC REG* command can be used to evaluate VIF values, however the fit values or estimates do not reflect the actual estimates. In order to find the actual estimates for panel data, *proc panel* command was used SAS. The initial model fit showed a moderate fit of 48.42%. However, this included non-significant variables as well. Eliminating those variables, the model fit became 45.65%.

It needs to be pointed out that the Hausman test revealed the presence of random effects in the panel model since *p*-value is less than 0.05. The variance component estimates will be required when safety culture performance indicators are analyzed and compared. From the panel result it was found that, GDP, unemployment rate (UnempRate),

rural road length (RuralRoad), temperature (temp), alcohol consumption per capita (AlCons), texting while driving law (TextingLaw), handheld devices while driving law (HandheldLaw), million vehicle miles travelled (MilVMT), male driver (MD), and voter appearance percentage (VotePer) were all found to be highly significant. Even though they have been found to be significant in one analysis, all of them might not be significant after pooling all the estimates from the imputed dataset. The result one of the imputed dataset from SAS illustrating parameter estimates for these variables are given below:

Parameter Estimates					
Variable	DF	Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-1529.04	327.5	-4.67	<.0001
GDP	1	-0.00369	0.00107	-3.44	0.0006
UnempRate	1	-29.211	3.7326	-7.83	<.0001
RuralRoad	1	0.004724	0.000583	8.11	<.0001
Temp	1	16.005	2.3765	6.73	<.0001
AlCons	1	180.5722	30.9901	5.83	<.0001
TextingLaw	1	-74.6919	15.2809	-4.89	<.0001
HandheldLaw	1	-56.562	19.4826	-2.90	0.0038
MilVMT	1	0.0079	0.000395	20.00	<.0001
MD	1	15.8585	5.5945	2.83	0.0047
VotePer	1	-244.4	85.4504	-2.86	0.0043

Figure 5.2. Estimates of significant variables from the panel data analysis

However, analyzing the distribution of the residuals it was found that, the residuals were heteroscedastic in nature when regressing dependent variable – number of fatality. The existence of heteroscedasticity is a major concern as it can propel the regression analysis to deduce correlated and non-uniform residuals – with fluctuating

variance with the effects being modelled. An illustration is given below which shows clear form of pattern in the Q-Q plot and skewed distribution:

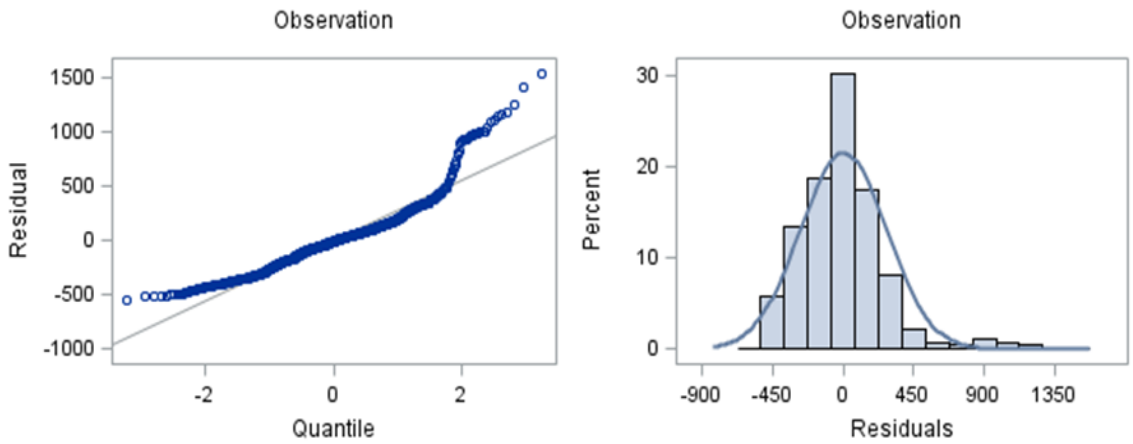

Figure 5.3. Skewed distribution of residuals and presence of heteroscedasticity

It also needs to be pointed out that, if the indicators are created based on the number of fatalities then there is a high probability that the comparative analysis from the residuals will lead to incorrect understanding. It is supported by assumption that the quality of the road, road safety signs, guiderails, and other surrounding factors are same. This is not the case, as the roads in Montana might be more dangerous than roads in California, even though the number of fatality is higher in the latter. Since road fatality and the level of exposure is highly correlated, the road fatality per unit of exposure can clearly distinguish the states based on their road safety performance.

Basing upon this notion and to reduce the heteroscedasticity implication on the model, the thesis took fatality per billion vehicle mile travelled as the new dependent variable. The procedure for selecting the variables follows the same process discussed previously. Steps are:

1. Primarily check the threshold of the VIF whether it is strictly less than 5 in all the imputed datasets obtained from *PROC MI*.
2. 18 variables are shown to have strict VIF values. Analysis reveales 18 variables have VIF value strictly below 5. The result illustrates two additional variables - population, and road improvement capital expenditures (RoadExp) which have VIF values < 5 and thus included in the model.
3. The variable illustrated in this analysis showed the selected variables that did not cross the VIF threshold value in any of the imputed dataset.

Parameter Estimates							
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Tolerance	Variance Inflation
Intercept	1	-0.48310	4.84667	-0.10	0.9206	.	0
GDP	1	-0.00006164	0.00001411	-4.37	<.0001	0.35313	2.83179
UnempRate	1	0.26974	0.05834	4.62	<.0001	0.58513	1.70903
RuralRoad	1	0.00000743	0.00000308	2.42	0.0159	0.47421	2.10878
Temp	1	0.16411	0.01360	12.07	<.0001	0.48692	2.05372
AlCons	1	1.11762	0.20672	5.41	<.0001	0.70730	1.41382
SBLaw	1	-0.51063	0.20940	-2.44	0.0149	0.65499	1.52673
TextingLaw	1	-2.33204	0.33162	-7.03	<.0001	0.35911	2.78463
CellLaw	1	-0.60282	0.30490	-1.98	0.0483	0.44106	2.26729
HandheldLaw	1	0.24892	0.40277	0.62	0.5367	0.71732	1.39408
DisDriving	1	-0.25537	1.09641	-0.23	0.8159	0.94077	1.06296
Beds	1	1.03659	0.11324	9.15	<.0001	0.62102	1.61026
Population	1	-0.00009062	0.00002704	-3.35	0.0008	0.23812	4.19965
MD	1	0.34152	0.08072	4.23	<.0001	0.66780	1.49744
DemorRep	1	-0.53902	0.30797	-1.75	0.0804	0.30149	3.31681
Dem	1	-0.18728	0.02168	-8.64	<.0001	0.20125	4.96894
Rep	1	-0.05387	0.01895	-2.84	0.0046	0.21471	4.65749
VotePer	1	-4.60339	0.97522	-4.72	<.0001	0.69965	1.42929
RoadExp	1	-2.50753E-7	1.602165E-7	-1.57	0.1179	0.26455	3.78004

Figure 5.4. Selected variables after primary screening of variance inflation factor

The overall process of imputation was similar to what had been described in chapter four, as well as the process of pooling the estimates also followed Rubin's rule. The major differences were associated with in built SAS programs of difference imputation methods which are markov chain monte carlo (MCMC) method, and fully conditional specification (FCS). Depending on the structure of missing data patterns, the monotone missing data pattern based analysis were excluded from the scope of this thesis, since the missing data pattern was arbitrary. This thesis will focus on the MCMC method with the assumption of multivariate normal distribution of the data.

The estimates were pooled for these variables with PROC MIANALYZE program in SAS which is given below:

Parameter Estimates										
Parameter	Estimate	Std Error	95% Confidence Limits		DF	Minimum	Maximum	Theta0	t for H0: Parameter=Theta0	Pr > |t|
Intercept	13.983974	4.813531	4.54962	23.41833	2.45E6	13.718110	14.323269	0	2.91	0.0037
GDP	0.000033187	0.000019011	-0.00000	0.00007	683848	0.000031724	0.000034543	0	1.75	0.0809
UnempRate	-0.154383	0.056487	-0.26510	-0.04367	253305	-0.160410	-0.146064	0	-2.73	0.0063
RuralRoad	0.000022816	0.000008352	0.00001	0.00004	2.04E6	0.000022131	0.000023366	0	2.73	0.0063
Temp	0.213398	0.034016	0.14673	0.28007	257291	0.209405	0.217543	0	6.27	<.0001
AlCons	2.157722	0.420896	1.33278	2.98266	931619	2.124002	2.189735	0	5.13	<.0001
SBLaw	-0.506487	0.169378	-0.83846	-0.17451	615751	-0.518308	-0.493262	0	-2.99	0.0028
CellLaw	0.482310	0.170556	0.14803	0.81659	400669	0.467785	0.499952	0	2.83	0.0047
DisDriving	2.179935	0.829430	0.55428	3.80559	6.16E6	2.134168	2.214452	0	2.63	0.0086
Beds	0.264283	0.113809	0.04088	0.48769	783.84	0.210521	0.320551	0	2.32	0.0205
Population	-0.000226	0.000058912	-0.00034	-0.00011	3.97E6	-0.000230	-0.000223	0	-3.84	0.0001
MD	-0.170696	0.071557	-0.31095	-0.03045	1.42E6	-0.176108	-0.166744	0	-2.39	0.0171
Dem	-0.072756	0.022730	-0.11731	-0.02821	466459	-0.075330	-0.070217	0	-3.20	0.0014
Rep	-0.090589	0.022629	-0.13494	-0.04624	581400	-0.092501	-0.088008	0	-4.00	<.0001
VotePer	-1.994250	1.732421	-5.38973	1.40123	4.8E6	-2.117765	-1.929953	0	-1.15	0.2497
RoadExp	0.000000206	0.000000120	-0.00000	0.00000	986333	0.000000193	0.000000213	0	1.72	0.0851

Figure 5.5. The pooled parameter estimates initial stage

The results indicated that most of the variables were significant. Removing the non-significant factors – GDP per capita, VotePer and RoadExp, the new pooled estimated parameters revealed:

Parameter Estimates										
Parameter	Estimate	Std Error	95% Confidence Limits		DF	Minimum	Maximum	Theta0	t for H0: Parameter=Theta0	Pr > |t|
Intercept	11.869799	4.529718	2.99171	20.74789	2.34E6	11.594835	12.176695	0	2.62	0.0088
UnempRate	-0.186438	0.053373	-0.29105	-0.08183	545818	-0.191228	-0.180456	0	-3.49	0.0005
RuralRoad	0.000020217	0.000008131	0.00000	0.00004	2.39E6	0.000019559	0.000020648	0	2.49	0.0129
Temp	0.208918	0.033078	0.14409	0.27375	450446	0.205426	0.212299	0	6.32	<.0001
AlCons	2.373939	0.400011	1.58993	3.15795	763238	2.339318	2.406590	0	5.93	<.0001
SBLaw	-0.509472	0.168904	-0.84052	-0.17843	1.11E6	-0.519505	-0.498254	0	-3.02	0.0026
CellLaw	0.562354	0.165726	0.23754	0.88717	566154	0.547872	0.575458	0	3.39	0.0007
DisDriving	2.128774	0.828871	0.50422	3.75333	1.2E7	2.089857	2.158047	0	2.57	0.0102
Beds	0.228691	0.111633	0.00966	0.44772	1133.1	0.185987	0.284182	0	2.05	0.0407
Population	-0.000158	0.000050111	-0.00026	-0.00006	1.18E6	-0.000163	-0.000154	0	-3.15	0.0017
MD	-0.167514	0.071240	-0.30714	-0.02789	2.52E6	-0.172669	-0.164105	0	-2.35	0.0187
Dem	-0.050612	0.021341	-0.09244	-0.00878	1.52E6	-0.052057	-0.048968	0	-2.37	0.0177
Rep	-0.064000	0.020824	-0.10481	-0.02319	1.26E6	-0.066063	-0.062378	0	-3.07	0.0021

Figure 5.6. Pooled parameter estimates of the significant variables

According to the Rubin's rule formula, the within variance and between variance were calculated. The illustration is given below shows the total variance, degree of freedom as well as relative increase in variance:

Parameter	Variance			DF	Relative Increase in Variance
	Between	Within	Total		
Intercept	0.036613	20.478069	20.518343	2.34E6	0.001967
UnempRate	0.000010516	0.002837	0.002849	545818	0.004077
RuralRoad	1.165866E-13	6.598013E-11	6.610838E-11	2.39E6	0.001944
Temp	0.000004446	0.001089	0.001094	450446	0.004490
AlCons	0.000500	0.159460	0.160009	763238	0.003446
SBLaw	0.000073962	0.028447	0.028529	1.11E6	0.002860
CellLaw	0.000099550	0.027356	0.027465	566154	0.004003
DisDriving	0.000540	0.686433	0.687027	1.2E7	0.000866
Beds	0.001010	0.011351	0.012462	1133.1	0.097841
Population	6.314863E-12	2.5041448E-9	2.5110911E-9	1.18E6	0.002774
MD	0.000008711	0.005066	0.005075	2.52E6	0.001892
Dem	0.000001007	0.000454	0.000455	1.52E6	0.002438
Rep	0.000001054	0.000432	0.000434	1.26E6	0.002681

Figure 5.7. Illustrating relative change in variance due to multiple imputed dataset

The Hausman test revealed that a *p*-value less than 0.05 with degrees of freedom(df) being 4, which suggested that it is a random effect model. In addition to that, the overall model fit (R-square) value and SSE were estimated to be 13.81% & 2075.93 by obtaining an average estimate of 10 imputed datasets. A summary is given below of the obtained goodness of fit from the datasets:

Imputation Dataset	R Square Value	SSE	MSE	RMSE	Variance component for cross section	Variance component for time series	Variance component for error
1	13.69	2076.69	2.0026	1.415	5.903	1.200	1.789
2	13.85	2077.21	2.0031	1.415	5.845	1.175	1.790
3	13.96	2072.21	1.9983	1.414	5.871	1.185	1.787
4	13.86	2071.96	1.9980	1.414	5.912	1.201	1.788
5	13.8	2076.34	2.0023	1.415	5.917	1.177	1.790
6	13.77	2077.79	2.0037	1.416	5.902	1.174	1.791
7	13.8	2075.41	2.0014	1.415	5.889	1.192	1.791
8	13.67	2079.71	2.0055	1.416	5.895	1.183	1.792
9	13.79	2076.62	2.0025	1.415	5.833	1.196	1.790
10	13.91	2075.37	2.0013	1.415	5.766	1.196	1.789
Average Estimates	13.81	2075.93	2.0019	1.415	5.873	1.188	1.790

Figure 5.8. Model fit & diagnostics for 10 imputed datasets

The equation which will be used for the predicting and residual analysis is formulized using the pooled estimates. The equation is given below:

$$\widehat{Fatality\ per\ billion\ VMT} = 11.87 - 0.1864 * UnempRate + 0.00002 * RuralRoad + 0.209 * Temp + 2.3739 * AlCons - 0.5094 * SBLaw + 0.5623 * CellLaw + 2.1287 * DisDriving + 0.2286 * Beds - 0.000158 * Population - 0.1675 * MD - 0.0506 * Dem - 0.064 * Rep \quad \text{(Equation 5.3)}$$

5.3. Residual Analysis and Indicators

The estimated values of 'fatality per billion VMT' included certain amount of errors or residuals. The residuals in case of two way random effects modelling have three components embedded into it as previously discussed in section 5.1. In the study of Page (2001), a log link function was used to model the road fatality, although it was not properly reasoned. According to his analysis, the overall estimation of error comprises of two major parts – error related with cross sections and error related with time series. Assuming the error is represented by R_{it}, an empirical Bayesian estimate of this error can be given be the following equation (5.4):

$$R_{it}\hat{} = \frac{\sigma^2(v)+\sigma^2(e)}{\sigma^2(v)+\sigma^2(e)+\sigma^2(\varepsilon)} * [\ F_{it} - \beta_0 - \sum_{j=1}^{k} \beta_j X_{ijt}\ ] \qquad \text{(Equation 5.4)}$$

Where, $\sigma^2(v)$ is the variance component for cross sections, $\sigma^2(e)$ is the variance component for the time series, and $\sigma^2(\varepsilon)$ is the variance component for errors which were obtained and estimated from the imputed results. β_0 , and β_j are pooled intercept & pooled coefficient obtained from the result analysis (Figure 5.8.). The obtained residuals are multiplied by the ratio of the variance components, to obtain $R_{it}\hat{}$. As there are 50 states and 10 data sets imputed, an overall of 10,500 estimates of $R_{it}\hat{}$ were obtained. For each imputed dataset, the residuals for year 1994, 1995, and 1996 were averaged to indicate the state of the time invariant factors at the year 1994. Similar approach was taken to indicate the state of time invariant factors at the year 2014 (averaging residuals of year 2012, 2013 & 2014). A final average of the two indicators were taken since there are 10 imputed datasets, and in doing so, the uncertainty of the missing data – due to the

variable “Beds” have been incorporated. One of the differences with the study of the Page’s is the formulation of the indicator (I_{it}), as it depends on the structure of the residual itself. In the study done by Page, the error restructure was given by $e^{R_{it}\hat{}}$ due to the log link function. However, in this study, the formula for obtaining the I_{it} is given as:

I_{it}= 100 - $R_{it}\hat{}$ / $actual$ fatality per billion VMT * 100 (Equation 5.5)

Similar to the study by Page (2001), the value of $R_{it}\hat{}$ was negative the estimated fatality was higher than the actual value. That is, safety performance is better than expected. In this case (based on equation 5.5), I_{it} will be calculated to be > 100. This result implies a “protective” safety culture captured as part of the unobserved variables represented by the residual from the predictive model. Similarly, if the estimated residuals are positive, it means the actual number of fatalities is higher than predicted, resulting in I_{it} to be calculated < 100. This will imply that for that year, the inferred road safety culture (as part of the unobserved variables) was “risky” because performance was worse than predicted. A score of 100 implies that safety culture was unchanged. A summary is given below in a tabular format:

Table 5.1. Relationship between the performance indicator and inferred road safety culture

Residual ($R_{it}\hat{}$)	Performance Indicator (I_{it})	Traffic Safety Culture
Positive	Less than 100	Risky
Negative	More than 100	Protective

As an example, a step by step approach to obtain the performance indicators for the states of Montana is described below. From the performance indicator value, traffic

safety culture was inferred for all the states. The steps associated with transforming residuals to performance indicators are:

- Finding the predicted value of fatality per billion VMT for 10 datasets
- Finding the difference between actual fatality per billion VMT and predicted fatality per billion VMT
- Finding $R_{it}{}^{\wedge}$ from equation 5.4
- Obtaining average $R_{it}{}^{\wedge}$ from 10 datasets
- Averaging 1994 – 1996 $R_{it}{}^{\wedge}$ values to set a single average $R_{it}{}^{\wedge}$ value explaining the unobserved exogenous variables and time invariant variables for the year 1994
- Averaging 2012 – 2014 $R_{it}{}^{\wedge}$ values to set a single average $R_{it}{}^{\wedge}$ value explaining the unobserved exogenous variables and time invariant variables for the year 2014
- Finding the performance indicators using equation 5.5 for the year 1994 & 2014.

According to the steps mentioned above, to find the predicted fatality per billion VMT, equation 5.3 was utilized and for the states of Montana the predicted fatality per billion VMT (FPBVMT)values are provided below:

Year	FPBVMT	Predicted FPBVMT (1)	Predicted FPBVMT (2)	Predicted FPBVMT (3)	Predicted FPBVMT (4)	Predicted FPBVMT (5)	Predicted FPBVMT (6)	Predicted FPBVMT (7)	Predicted FPBVMT (8)	Predicted FPBVMT (9)	Predicted FPBVMT (10)
1994	22.1588	15.5551	15.5551	15.5551	15.5551	15.5551	15.5551	15.5551	15.5551	15.5551	15.5551
1995	22.8748	14.6130	14.9305	14.8463	15.0562	14.7866	14.8940	14.8475	15.1546	15.3192	15.0004
1996	21.1730	14.3775	14.3775	14.3775	14.3775	14.3775	14.3775	14.3775	14.3775	14.3775	14.3775
1997	28.2155	14.3873	14.3873	14.3873	14.3873	14.3873	14.3873	14.3873	14.3873	14.3873	14.3873
1998	24.7158	14.7778	14.7778	14.7778	14.7778	14.7778	14.7778	14.7778	14.7778	14.7778	14.7778
1999	22.3691	15.0518	14.5825	14.6557	14.7791	14.6563	14.8602	15.1093	14.9590	14.7419	15.1666
2000	23.9830	14.6859	14.6859	14.6859	14.6859	14.6859	14.6859	14.6859	14.6859	14.6859	14.6859
2001	22.9747	14.6180	14.6180	14.6180	14.6180	14.6180	14.6180	14.6180	14.6180	14.6180	14.6180
2002	25.8778	14.4018	14.4018	14.4018	14.4018	14.4018	14.4018	14.4018	14.4018	14.4018	14.4018
2003	24.0942	14.6672	14.6672	14.6672	14.6672	14.6672	14.6672	14.6672	14.6672	14.6672	14.6672
2004	20.4337	14.6123	14.6123	14.6123	14.6123	14.6123	14.6123	14.6123	14.6123	14.6123	14.6123
2005	22.5598	14.5451	14.5451	14.5451	14.5451	14.5451	14.5451	14.5451	14.5451	14.5451	14.5451
2006	23.4354	15.3592	15.3592	15.3592	15.3592	15.3592	15.3592	15.3592	15.3592	15.3592	15.3592
2007	24.4981	15.4425	15.4425	15.4425	15.4425	15.4425	15.4425	15.4425	15.4425	15.4425	15.4425
2008	21.1802	14.6118	14.6118	14.6118	14.6118	14.6118	14.6118	14.6118	14.6118	14.6118	14.6118
2009	20.0708	14.3628	14.3628	14.3628	14.3628	14.3628	14.3628	14.3628	14.3628	14.3628	14.3628
2010	16.8904	14.4643	14.4643	14.4643	14.4643	14.4643	14.4643	14.4643	14.4643	14.4643	14.4643
2011	17.9251	14.4373	14.4373	14.4373	14.4373	14.4373	14.4373	14.4373	14.4373	14.4373	14.4373
2012	17.2485	15.5542	15.5542	15.5542	15.5542	15.5542	15.5542	15.5542	15.5542	15.5542	15.5542
2013	19.0314	15.0919	15.0919	15.0919	15.0919	15.0919	15.0919	15.0919	15.0919	15.0919	15.0919
2014	15.7932	15.4013	15.1514	15.3954	14.8145	14.9749	14.8310	14.9778	14.8146	15.1465	15.1996

Figure 5.9. Predicted fatality per billion VMT from 10 datasets

After obtaining the predicted fatality per billion VMT from 10 datasets, they were subtracted from the actual fatality per billion VMT. The subtracted values obtained for Montana are given in the illustration below:

Residuals (1)	Residuals (2)	Residuals (3)	Residuals (4)	Residuals (5)	Residuals (6)	Residuals (7)	Residuals (8)	Residuals (9)	Residuals (10)
6.6038	6.6038	6.6038	6.6038	6.6038	6.6038	6.6038	6.6038	6.6038	6.6038
8.2618	7.9443	8.0285	7.8186	8.0882	7.9808	8.0272	7.7202	7.5555	7.8744
6.7955	6.7955	6.7955	6.7955	6.7955	6.7955	6.7955	6.7955	6.7955	6.7955
13.8282	13.8282	13.8282	13.8282	13.8282	13.8282	13.8282	13.8282	13.8282	13.8282
9.9380	9.9380	9.9380	9.9380	9.9380	9.9380	9.9380	9.9380	9.9380	9.9380
7.3173	7.7866	7.7134	7.5900	7.7128	7.5089	7.2598	7.4101	7.6271	7.2025
9.2971	9.2971	9.2971	9.2971	9.2971	9.2971	9.2971	9.2971	9.2971	9.2971
8.3568	8.3568	8.3568	8.3568	8.3568	8.3568	8.3568	8.3568	8.3568	8.3568
11.4760	11.4760	11.4760	11.4760	11.4760	11.4760	11.4760	11.4760	11.4760	11.4760
9.4269	9.4269	9.4269	9.4269	9.4269	9.4269	9.4269	9.4269	9.4269	9.4269
5.8214	5.8214	5.8214	5.8214	5.8214	5.8214	5.8214	5.8214	5.8214	5.8214
8.0147	8.0147	8.0147	8.0147	8.0147	8.0147	8.0147	8.0147	8.0147	8.0147
8.0762	8.0762	8.0762	8.0762	8.0762	8.0762	8.0762	8.0762	8.0762	8.0762
9.0556	9.0556	9.0556	9.0556	9.0556	9.0556	9.0556	9.0556	9.0556	9.0556
6.5683	6.5683	6.5683	6.5683	6.5683	6.5683	6.5683	6.5683	6.5683	6.5683
5.7080	5.7080	5.7080	5.7080	5.7080	5.7080	5.7080	5.7080	5.7080	5.7080
2.4261	2.4261	2.4261	2.4261	2.4261	2.4261	2.4261	2.4261	2.4261	2.4261
3.4878	3.4878	3.4878	3.4878	3.4878	3.4878	3.4878	3.4878	3.4878	3.4878
1.6943	1.6943	1.6943	1.6943	1.6943	1.6943	1.6943	1.6943	1.6943	1.6943
3.9396	3.9396	3.9396	3.9396	3.9396	3.9396	3.9396	3.9396	3.9396	3.9396
0.3919	0.6418	0.3978	0.9787	0.8183	0.9622	0.8154	0.9786	0.6467	0.5936

Figure 5.10. Subtracted value of predicted and actual fatality per billion VMT

By using the variance component factors for cross sections, time series and error from figure 5.8. and utilizing the equation 5.4, $R_{it}{}^{\wedge}$ were obtained. After obtaining these values from 10 datasets, they were averaged. An illustration is given below:

Year	Rit (1)	Rit (2)	Rit (3)	Rit (4)	Rit (5)	Rit (6)	Rit (7)	Rit (8)	Rit (9)	Rit (10)	Average (Rit)
1994	5.2685	5.2685	5.2685	5.2685	5.2685	5.2685	5.2685	5.2685	5.2685	5.2685	5.2685
1995	6.5913	6.3380	6.4052	6.2377	6.4528	6.3671	6.4042	6.1592	6.0278	6.2822	6.3266
1996	5.4214	5.4214	5.4214	5.4214	5.4214	5.4214	5.4214	5.4214	5.4214	5.4214	5.4214
1997	11.0322	11.0322	11.0322	11.0322	11.0322	11.0322	11.0322	11.0322	11.0322	11.0322	11.0322
1998	7.9286	7.9286	7.9286	7.9286	7.9286	7.9286	7.9286	7.9286	7.9286	7.9286	7.9286
1999	5.8378	6.2122	6.1538	6.0554	6.1533	5.9906	5.7919	5.9118	6.0850	5.7462	5.9938
2000	7.4173	7.4173	7.4173	7.4173	7.4173	7.4173	7.4173	7.4173	7.4173	7.4173	7.4173
2001	6.6671	6.6671	6.6671	6.6671	6.6671	6.6671	6.6671	6.6671	6.6671	6.6671	6.6671
2002	9.1556	9.1556	9.1556	9.1556	9.1556	9.1556	9.1556	9.1556	9.1556	9.1556	9.1556
2003	7.5209	7.5209	7.5209	7.5209	7.5209	7.5209	7.5209	7.5209	7.5209	7.5209	7.5209
2004	4.6443	4.6443	4.6443	4.6443	4.6443	4.6443	4.6443	4.6443	4.6443	4.6443	4.6443
2005	6.3942	6.3942	6.3942	6.3942	6.3942	6.3942	6.3942	6.3942	6.3942	6.3942	6.3942
2006	6.4432	6.4432	6.4432	6.4432	6.4432	6.4432	6.4432	6.4432	6.4432	6.4432	6.4432
2007	7.2246	7.2246	7.2246	7.2246	7.2246	7.2246	7.2246	7.2246	7.2246	7.2246	7.2246
2008	5.2403	5.2403	5.2403	5.2403	5.2403	5.2403	5.2403	5.2403	5.2403	5.2403	5.2403
2009	4.5539	4.5539	4.5539	4.5539	4.5539	4.5539	4.5539	4.5539	4.5539	4.5539	4.5539
2010	1.9356	1.9356	1.9356	1.9356	1.9356	1.9356	1.9356	1.9356	1.9356	1.9356	1.9356
2011	2.7826	2.7826	2.7826	2.7826	2.7826	2.7826	2.7826	2.7826	2.7826	2.7826	2.7826
2012	1.3517	1.3517	1.3517	1.3517	1.3517	1.3517	1.3517	1.3517	1.3517	1.3517	1.3517
2013	3.1430	3.1430	3.1430	3.1430	3.1430	3.1430	3.1430	3.1430	3.1430	3.1430	3.1430
2014	0.3126	0.5120	0.3173	0.7808	0.6529	0.7676	0.6505	0.7807	0.5159	0.4736	0.5764

Figure 5.11. Obtaining the $R_{it}{}^{\wedge}$ values for each dataset and the average $R_{it}{}^{\wedge}$

To obtain average $R_{it}{}^{\wedge}$ for the year 1994 and 2014, the average residuals from 1994 to 1996 and from 2012 to 2014 were taken to be average.

$R_{it}{}^{\wedge}(1994)$ = (5.2685 + 6.3266 + 5.4214) / 3 = 5.672

$R_{it}{}^{\wedge}(2014)$ = (1.3517+ 3.1430 + 0.5764) / 3 = 1.6903

By using the equation 5.5 the performance indicator was found to be 74.4. The detailed calculation is shown below:

$I_{it}(1994)$= 100 - $5.672/22.1588$ * 100 = 74.40 (approximated)

Similarly, $I_{it}(2014)$ was found to be 89.297 (approximated)

By following the same procedure, performance indicators of the other states were obtained. The performance indicators provide a level of inference of road safety culture, behavior, and other unobserved components. However, they cannot be segregated to find the specific factors behind the progress. The performance indicators for all the states are given below along with the amount to progression in 21 years:

Table 5.2. Variation in estimated safety culture across states (I_{it})

State	I(1994)	I(2014)	Difference between 21 years	State	I(1994)	I(2014)	Difference between 21 years
Alabama	80.722	122.116	41.394	Montana	74.40	89.297	14.897
Alaska	60.282	77.841	17.559	Nebraska	96.325	129.996	33.671
Arizona	77.522	112.257	34.736	Nevada	86.464	132.576	46.112
Arkansas	75.662	103.904	28.242	New Hampshire	144.458	194.508	50.050
California	83.013	112.532	29.519	New Jersey	106.221	160.589	54.368
Colorado	91.250	135.188	43.937	New Mexico	75.323	107.479	32.156
Connecticut	114.019	141.620	27.601	New York	76.122	108.424	32.302
Delaware	104.752	140.961	36.209	North Carolina	85.203	114.461	29.258
Florida	88.665	121.722	33.058	North Dakota	115.197	110.891	-4.306
Georgia	101.585	136.496	34.911	Ohio	101.119	136.805	35.686
Hawaii	109.168	163.566	54.398	Oklahoma	94.568	109.293	14.724
Idaho	73.960	111.965	38.004	Oregon	85.241	134.156	48.915
Illinois	95.313	139.937	44.624	Pennsylvania	86.974	101.811	14.837
Indiana	96.955	129.000	32.045	Rhode Island	141.444	161.390	19.946
Iowa	82.499	136.636	54.137	South Carolina	79.791	95.253	15.462
Kansas	96.169	125.050	28.881	South Dakota	79.528	107.996	28.468
Kentucky	79.976	98.739	18.763	Tennessee	77.868	107.416	29.549
Louisiana	82.783	114.733	31.950	Texas	100.004	113.540	13.537
Maine	93.738	124.954	31.216	Utah	79.875	126.797	46.922
Maryland	97.224	150.811	53.588	Vermont	96.392	161.688	65.296
Massachusetts	142.654	192.986	50.332	Virginia	112.916	149.735	36.819
Michigan	82.586	120.048	37.461	Washington	101.028	145.591	44.563
Minnesota	109.157	192.766	83.609	West Virginia	70.920	84.756	13.835
Mississippi	71.750	105.857	34.107	Wisconsin	106.583	145.037	38.454
Missouri	90.021	131.551	41.530	Wyoming	72.795	109.811	37.015

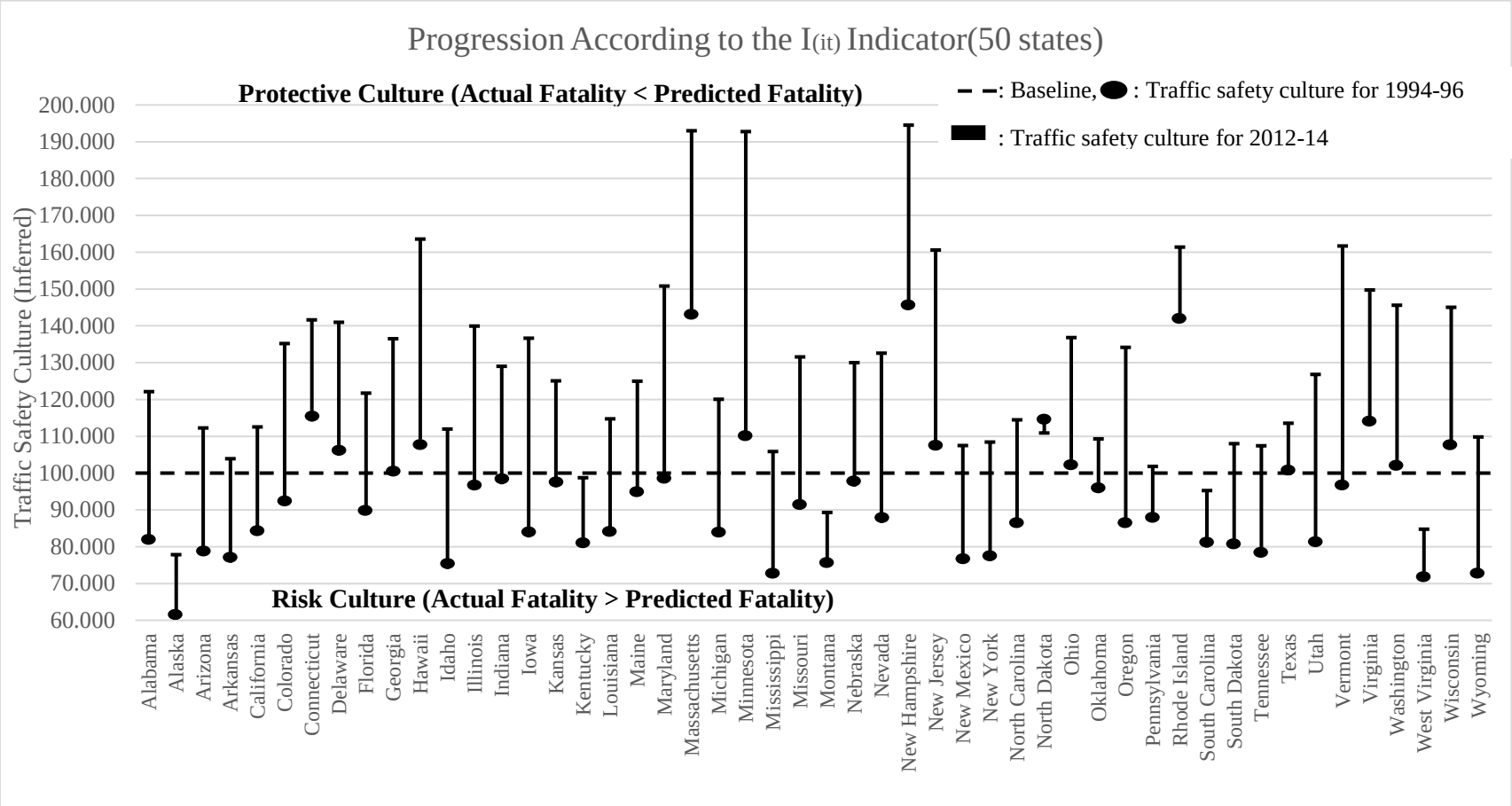

Figure 5.12. Progression of 50 states according to I_{it} indicator with inferred safety culture

The progression graph illustrates the fact that almost all the states made a good progression in terms of strengthening traffic safety culture from 1994 to 2014 except North Dakota as. The general increase in inferred safety culture is an indication that the road safety culture, road safety policies and other safety interventions have a positive effect on the states. The analyzed results show that the best state in terms of road fatality per billion vehicles million travelled is New Hampshire in both I (1994) & I (2014). The states of Minnesota, Vermont and Hawaii progressed the most in terms of inferred safety culture performance indicator. Since the overall model fit is only 13.85%, it can be generalized that the unexplained variations are contained in the residuals can be explained by aspects of human culture, norms, attitude, advancement in vehicle technology, driving behavior culture – such as seat belt wearing rates, helmet wearing rates, speeding nature etc.

Figure 5.3 and table 5.3 illustrate that, 21 years ago, 34 states had risky traffic culture, as the inferred safety culture were below the baseline. Among them Maryland, Vermont, and Colorado improved their road safety culture significantly in 21 years. In terms of progress, the states of Minnesota progressed the most, from a traffic safety culture value of 109.16 in 1994 to 192.77 in 2014. According to the inferred traffic safety culture indicators proposed in this study, New Hampshire was found to be the safest state with protective traffic culture while Alaska has the riskiest traffic culture all throughout 1994 to 2014. Even though the states of North Dakota possess a protective traffic safety culture, it deteriorated in the last 21 years.

Table 5.3. Ranking of states according to performance indicators

Rank	State	I(1994)	Rank	State	I(2014)	Rank	State	Difference between 21 years
1	New Hampshire	144.46	1	New Hampshire	194.51	1	Minnesota	83.609
2	Massachusetts	142.65	2	Massachusetts	192.99	2	Vermont	65.296
3	Rhode Island	141.44	3	Minnesota	192.77	3	Hawaii	54.398
4	North Dakota	115.20	4	Hawaii	163.57	4	New Jersey	54.368
5	Connecticut	114.02	5	Vermont	161.69	5	Iowa	54.137
6	Virginia	112.92	6	Rhode Island	161.39	6	Maryland	53.588
7	Hawaii	109.17	7	New Jersey	160.59	7	Massachusetts	50.332
8	Minnesota	109.16	8	Maryland	150.81	8	New Hampshire	50.050
9	Wisconsin	106.58	9	Virginia	149.74	9	Oregon	48.915
10	New Jersey	106.22	10	Washington	145.59	10	Utah	46.922
11	Delaware	104.75	11	Wisconsin	145.04	11	Nevada	46.112
12	Georgia	101.58	12	Connecticut	141.62	12	Illinois	44.624
13	Ohio	101.12	13	Delaware	140.96	13	Washington	44.563
14	Washington	101.03	14	Illinois	139.94	14	Colorado	43.937
15	Texas	100.00	15	Ohio	136.81	15	Missouri	41.530
16	Maryland	97.22	16	Iowa	136.64	16	Alabama	41.394
17	Indiana	96.96	17	Georgia	136.50	17	Wisconsin	38.454
18	Vermont	96.39	18	Colorado	135.19	18	Idaho	38.004
19	Nebraska	96.33	19	Oregon	134.16	19	Michigan	37.461
20	Kansas	96.17	20	Nevada	132.58	20	Wyoming	37.015
21	Illinois	95.31	21	Missouri	131.55	21	Virginia	36.819
22	Oklahoma	94.57	22	Nebraska	130.00	22	Delaware	36.209
23	Maine	93.74	23	Indiana	129.00	23	Ohio	35.686
24	Colorado	91.25	24	Utah	126.80	24	Georgia	34.911
25	Missouri	90.02	25	Kansas	125.05	25	Arizona	34.736

Table 5.3. Ranking of states according to performance indicators – continued

Rank	State	I(1994)	Rank	State	I(2014)	Rank	State	Difference between 21 years
26	Florida	88.66	26	Maine	124.95	26	Mississippi	34.107
27	Pennsylvania	86.97	27	Alabama	122.12	27	Nebraska	33.671
28	Nevada	86.46	28	Florida	121.72	28	Florida	33.058
29	Oregon	85.24	29	Michigan	120.05	29	New York	32.302
30	North Carolina	85.20	30	Louisiana	114.73	30	New Mexico	32.156
31	California	83.01	31	North Carolina	114.46	31	Indiana	32.045
32	Louisiana	82.78	32	Texas	113.54	32	Louisiana	31.950
33	Michigan	82.59	33	California	112.53	33	Maine	31.216
34	Iowa	82.50	34	Arizona	112.26	34	Tennessee	29.549
35	Alabama	80.72	35	Idaho	111.96	35	California	29.519
36	Kentucky	79.98	36	North Dakota	110.89	36	North Carolina	29.258
37	Utah	79.87	37	Wyoming	109.81	37	Kansas	28.881
38	South Carolina	79.79	38	Oklahoma	109.29	38	South Dakota	28.468
39	South Dakota	79.53	39	New York	108.42	39	Arkansas	28.242
40	Tennessee	77.87	40	South Dakota	108.00	40	Connecticut	27.601
41	Arizona	77.52	41	New Mexico	107.48	41	Rhode Island	19.946
42	New York	76.12	42	Tennessee	107.42	42	Kentucky	18.763
43	Arkansas	75.66	43	Mississippi	105.86	43	Alaska	17.559
44	New Mexico	75.32	44	Arkansas	103.90	44	Montana	14.897
45	Montana	74.40	45	Pennsylvania	101.81	45	South Carolina	15.462
46	Idaho	73.96	46	Kentucky	98.74	46	Pennsylvania	14.837
47	Wyoming	72.80	47	South Carolina	95.25	47	Oklahoma	14.724
48	Mississippi	71.75	48	Montana	89.30	48	West Virginia	13.835
49	West Virginia	70.92	49	West Virginia	84.76	49	Texas	13.537
50	Alaska	60.28	50	Alaska	77.84	50	North Dakota	-4.306

5.4. Critiquing the Study by Page

The study done by Page highlighted the progression of 21 OECD countries including United States. From the illustration given below, it can be visible that the progression of USA was not properly captured. This can be related to the fact that the variables which was selected for the study could not properly captured the progression of the safety culture, behaviors, road safety performance and other safety interventions in USA.

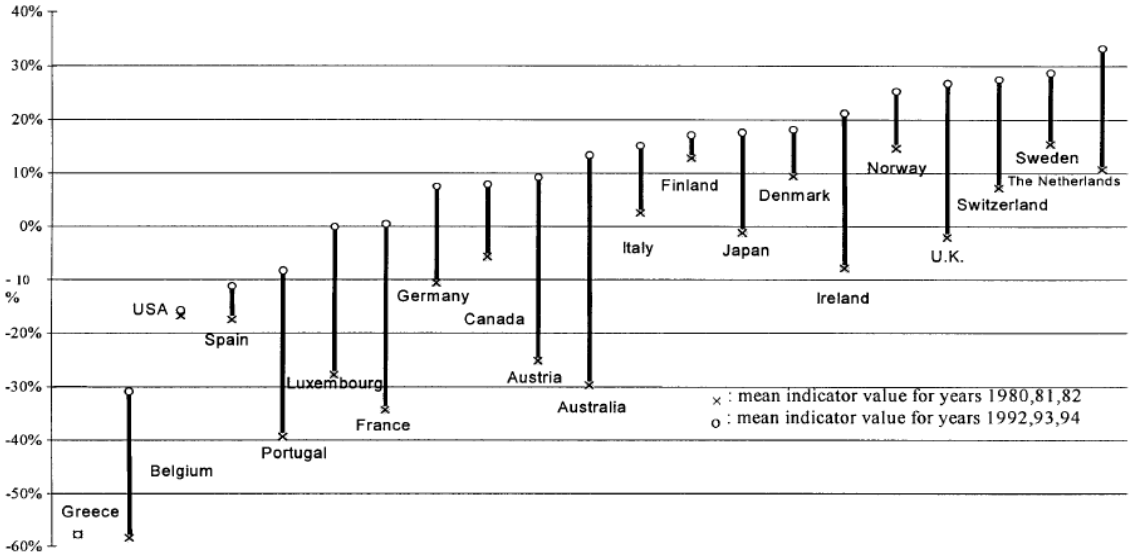

Figure 5.13. Comparative analysis of progression from Page's Study

One of the major criticism that can be drawn from the study of Page, is related with handling missing data. In cases, where a few data points were found to be missing linear regression, interpolation, or smoothing technique methods were used primarily. From reviewing the literature, it was understood that such imputation techniques provide deterministic data points, which in return causes certain bias in parameter estimation.

Also, it is not well understood from his study the regarding the proportion of missing data and the pattern they exhibited also. This was important since handling the missing data depends on both these aspects.

From his analysis, it can be directly inferred that the variables which were selected reflected mostly the human population and economic indicators. It was also mentioned in the study that, variables such as safety interventions, road engineering & routing, and meteorology need to be included. This study has included all these factors according to the integrated model. The findings showed 12 variables to be significant in this study. All these variables can be mapped into the integrated model which proves the hypothesis of this study regarding the framework.

It was not evident from the study of Page, whether the heteroscedasticity in the model was corrected since no graphical representation of error distribution was presented in the study. Even though the graphical representation showed errors are normally distributed across the 50 states to come with standard errors which is consistent with heteroscedasticity and autocorrelation, this study considered Newey-West standard errors to be a solution. Newey and West (1986) developed a way of obtaining heteroscedasticity and autocorrelation (up to some lag) consistent standard errors. Although initially Newey West standard errors was proposed for time series data use only, it now also be applied to panel data models. The overall Newey-West standard errors obtained from 10 imputed datasets were pooled together. The results are shown below:

Parameter Estimates										
Parameter	Estimate	Std Error	95% Confidence Limits		DF	Minimum	Maximum	Theta0	t for H0: Parameter=Theta0	Pr > |t|
Intercept	11.869799	7.166778	-2.17683	25.91643	1.46E7	11.594835	12.176695	0	1.66	0.0977
UnempRate	-0.186438	0.067327	-0.31840	-0.05448	1.38E6	-0.191228	-0.180456	0	-2.77	0.0056
RuralRoad	0.000020217	0.000010178	0.00000	0.00004	5.87E6	0.000019559	0.000020648	0	1.99	0.0470
Temp	0.208918	0.044801	0.12111	0.29673	1.52E6	0.205426	0.212299	0	4.66	<.0001
AlCons	2.373939	0.595823	1.20615	3.54173	3.76E6	2.339318	2.406590	0	3.98	<.0001
SBLaw	-0.509472	0.207567	-0.91630	-0.10265	2.52E6	-0.519505	-0.498254	0	-2.45	0.0141
CellLaw	0.562354	0.212193	0.14646	0.97825	1.52E6	0.547872	0.575458	0	2.65	0.0080
DisDriving	2.128774	1.226852	-0.27581	4.53336	5.77E7	2.089857	2.158047	0	1.74	0.0827
Beds	0.228691	0.147537	-0.06058	0.51796	3457.1	0.185987	0.284182	0	1.55	0.1212
Population	-0.000158	0.000053304	-0.00026	-0.00005	1.51E6	-0.000163	-0.000154	0	-2.96	0.0031
MD	-0.167514	0.098734	-0.36103	0.02600	9.32E6	-0.172669	-0.164105	0	-1.70	0.0898
Dem	-0.050612	0.034152	-0.11755	0.01632	9.98E6	-0.052057	-0.048968	0	-1.48	0.1383
Rep	-0.064000	0.031870	-0.12646	-0.00154	6.91E6	-0.066063	-0.062378	0	-2.01	0.0446

Figure 5.14. Newey-West standard errors varying from previous standard error

The difference from the previous result and the result given right above is in the standard errors and in their confidence intervals. Due to this fact, the p-value also changed indicating the model has been over estimated. The overall variance information suggests that the total variance is larger with Newey-West standard errors. The new variance results are given below:

Parameter	Variance			DF	Relative Increase in Variance	Fraction Missing Information
	Between	Within	Total			
Intercept	0.036613	51.322427	51.362702	1.46E7	0.000785	0.000784
UnempRate	0.000010516	0.004521	0.004533	1.38E6	0.002558	0.002553
RuralRoad	1.165866E-13	1.034733E-10	1.036015E-10	5.87E6	0.001239	0.001238
Temp	0.000004446	0.002002	0.002007	1.52E6	0.002443	0.002438
AlCons	0.000500	0.354455	0.355005	3.76E6	0.001550	0.001548
SBLaw	0.000073962	0.043003	0.043084	2.52E6	0.001892	0.001889
CellLaw	0.000099550	0.044917	0.045026	1.52E6	0.002438	0.002433
DisDriving	0.000540	1.504571	1.505165	5.77E7	0.000395	0.000395
Beds	0.001010	0.020657	0.021767	3457.1	0.053766	0.051571
Population	6.314863E-12	2.834343E-9	2.8412894E-9	1.51E6	0.002451	0.002446
MD	0.000008711	0.009739	0.009748	9.32E6	0.000984	0.000983
Dem	0.000001007	0.001165	0.001166	9.98E6	0.000950	0.000950
Rep	0.000001054	0.001015	0.001016	6.91E6	0.001143	0.001142

Figure 5.15. Total variance of the parameter after Newey-West standard error

Mishkin (1990) described that for finite sample size and data points correction of heteroscedasticity may help produce valid standard errors as shown by the Newey-West method, but it may not help to improve statistical inferences. So, the regression model has been kept the same as suggested in section 5.3.

5.5. Effect of Proxy Cultural Variables on Simple Linear Regression

In chapter three, it was discussed about the limitations of a reliable source from which data on culture, attitude, behavior and norms associated with road safety and performance can be obtained. For that, this study considered four political factors, which had been considered as proxy for cultural factors as well as a representation of the political factors. The main objective here was to capture the effects of culture into the regression analysis and to show whether there is any significant impact on the overall statistical goodness of fit. In chapter five states – Florida, Georgia, Indiana, Iowa and North Carolina were chosen randomly to check the significance of these factors.

Table: 5.4. Description of political factors

Description of the variable	Abbreviated form in the analysis
Is the state democratic or republican	DemorRep
Democrat candidate voter percentage	Dem
Republican candidate voter percentage	Rep
Percentage of Voters Voted	VotePer

The political variables used in this thesis are not an actual reflection of the political aspects relating to the road fatalities within certain geographic boundary but as proxy variables. From the overall stepwise analysis from variable selection procedure to final selection of significant variables, it is shown in the pooled estimates and regression

models that some form of significance is present in all the cases for these five states. The prediction models follow as such:

Table 5.5. Prediction model incorporating political factors

State	Prediction Model for Fatalities	Significant Factors
FL	Fatality = 664.98 + 976.83 * Distracted Driving - 0.00127 * Unemployed - 521.29* factor(DemorRep) + 52.91 * Dem + 1048.99 * VotePer	Distracted Driving, Unemployed, Factor (Dem or Rep), Dem, VotePer
GA	Fatality = 3222.88 – 13804.62 * Distracted Driving – 36.31 * Dem	Distracted Driving, Dem
IN	Fatality = 829.98 + 1357.85 * Drunk Driving – 8.1 * Dem + 1123.99 * Distracted Driving	Drunk Driving, Dem, Distracted Driving
IA	Fatality = 1207.44 – 448.33 * Drunk Driving + 1801.089 * Distracted Driving – 0.12 * Emergency Response Team – 7.072 *Dem	Drunk Driving, Distracted Driving, ER Team, Dem
NC	Fatality = -3443.59 + 8.98 * Rep - 0.0004811 * Unemployed + 92.502 * Male Driver	Rep, Unemployed, Male Driver

It was found that by including political factors into the prediction model made the results better in terms of goodness of fit and error terms. A summary of comparative analysis between the two results is given below:

Table 5.6. Comparison between model fit diagnostics after inclusion of political factors

Model Description	Prediction Model Fit Diagnostics					
	Without Political Factors			With Political Factors		
	R^2	R^2 $_{Adj}$	SSE	R^2	R^2 $_{Adj}$	SSE
Florida	0.76937	0.71171	916278.9	0.87414	0.83218	281679.4
Georgia	0.66414	0.6464	222234.7	0.78529	0.7614	142091.6
Indiana	0.82154	0.79005	32886.25	0.82719	0.7967	31848.88
Iowa	0.5622	0.5387	25181.06	0.8962	0.8703	5574.041
North Carolina	0.73467	0.70519	83931.14	0.84637	0.81924	46057.63

The difference in result in terms of decreasing SSE and improving R^2 $_{Adj}$ values suggest the fact that the four variables used as a proxy for culture does have a significant effect on road fatality.

CHAPTER SIX – DISCUSSIONS

This chapter will focus on discussing the results from the two different types of analysis illustrated in Chapter Four as well as Chapter Five. The chapter is divided into four sections. The first section will focus on comparative analysis. The second section will discuss findings from the models which were counterintuitive. Relative advantages and limitations of two models will be discussed in third section, and finally, the fourth section will make inferences on which states are safer and which are dangerous for driving.

6.1. Comparative Analysis of the Findings

Common variables found using both types of analysis were distracted driving and cell phone use for novice and teen drivers law. In both analyses, the coefficients of the unemployment factors (number of unemployed and unemployment rate) in relation to the number of fatalities were found to be negatively associated (Table 4.7. and Figure 5.6.). It needs to be highlighted that in the variable – number of unemployed was the only economic variable found to be a significant variable during linear regression models. Similarly, in the panel regression model, the only economic variable was found to be unemployment rate. Even though ‘unemployed’ was also found to be significant in the stepwise variable selection, it was eliminated due to the high VIF value. The findings associated with this variable were consistent with the prior literature (Leigh and Waldon 1991). In addition, Partyka (1984) suggested that the negative coefficient for the unemployment implies the traveler’s inability to pay for the travel, which in effect

reduces non-work based travel and the level of fatalities. Although the effects of the laws against cellphone use for novice and teen drivers negatively related in the linear regression model, the panel regression model offered a positive effect. A table has been provided below that summarizes the findings from both models and findings that were not present in them:

Table 6.1. Comparison of findings of variables from two different analytical models

Variables Obtained in Both Analytical Models	**Variables not present in the linear regression model but was present in panel regression**	**Variables not present in the panel regression model but was present in linear regression**
Distracted Driving and Cellphone use for novice & teen drivers and male driver (a few states in linear regression model)	Population, seat belt law, Beds per 1000 people, Rural Roads, Alcohol consumption per capita, Temperature	Female drivers, Precipitation, Handheld device while driving law, Texting while driving law, Emergency response team, Total number of hospitals, Number of nurses

There are some differences in terms of formulating the two models. The linear regression model was formulated with the dependent variable being ‘number of fatalities’. In contrast, in the panel regression analysis, the dependent variable was ‘number of fatalities per billion vehicle miles’. This was done due to the heteroskedastic residuals found from the analysis. The differences in dependent variables resulted in different findings from the two types of analysis, specifically in terms of identifying the significant variables.

In line with the integrated framework, the panel regression model identified ‘population’ and ‘male driver’ were significantly related. However, from the linear

regression analysis it was seen that population-based factors only appeared a small number of times. It was interesting to see that the total number of young drivers was not found to be significant in any of the statewide regression models. One of possible reasons behind this is multicollinearity as several variables such as human population, young driver population, total number of motor vehicles, and number of motorcycles were found to be significantly positively correlated. With a strict VIF threshold, these variables were eliminated in the stepwise process.

In terms of political factors, laws against using handheld devices while driving and against cellphone usage by novice and teen while driving were found to be significant in a few states. The regression model showed that after enforcing these laws the number of fatality decrease. However, in the panel data regression model the result showed seat belt law and cellphone law for novice and teen drivers were highly significant. One explanation that can be given with respect to the negative relation is the offsetting behavior theory. Peltzman (1975) suggested that even though safety regulations have tended to decrease the risk associated with crash, drivers have offset this by choosing to take greater crash risk, which in turn increase the chance of road fatality.

6.2. Explanation Behind the Counter Intuitive Results

The analyses from both the linear regression model and the panel regression model have revealed counterintuitive results. In general, speeding and distracted driving are risk factors which should be positively associated with the number of fatalities. However, some of the regression models yielded negative coefficients which were

counterintuitive and unexpected. In general, a regression model tries to fit the data in the best possible manner to explain the variation of the dependent variable through the independent variable or variables. One of the underlying assumption behind this is the fact that the dataset is correct. Considering this fact, data associated with these risk factors for the states in which such results were obtained was checked. For the states of Tennessee, the scatter plot diagram revealed a negative correlation. It was seen that, for higher number of fatalities the percentage of speeding risky behavior was low. The illustration is given below:

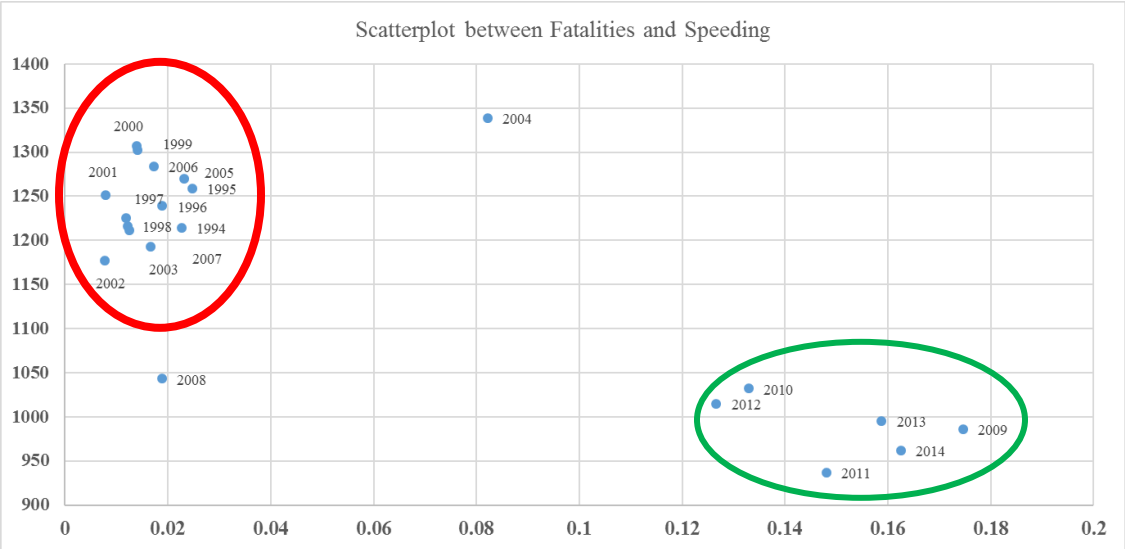

Figure 6.1. Scatter plot diagram of the number of fatalities vs speeding

The scatter plot shows that there are two clusters of data. The main distinction between these two clusters are the reporting year. From this illustration, two speculations can be made. First, the two apparent clusters in the data may create a spurious correlation based on linear line calculated in the analysis between these two “point” clusters. Second, the different time periods it is apparent that the two clusters represent different periods

represented by these clusters might imply that methods of data collection and reporting are different. This may suggest that the meaning of this variable is different between these time periods such that a negative correlation is artificially created. Similar clustering of information was also found for the datasets in Nevada, Rhode Island and Vermont.

There can be numerous reasons behind road fatalities or road crashes, such as drunk driving, speeding and distracted driving. These variables can be attributed to drivers as it accounts for 94% +/- 2.2% of the critical reasons behind the pre-crash event (Singh 2015). According the IIHS site (Source: http://www.iihs.org) recent years have shown a decline in road fatalities associated with speeding and drunk driving. Research shows that there is an increase in fatalities due to distracted driving happening in United States (Wilson and Stimpson 2010). Eby and Vivoda (2003) highlighted that with the increase usage of cellphones or smartphones, the distracted driving has gone up.

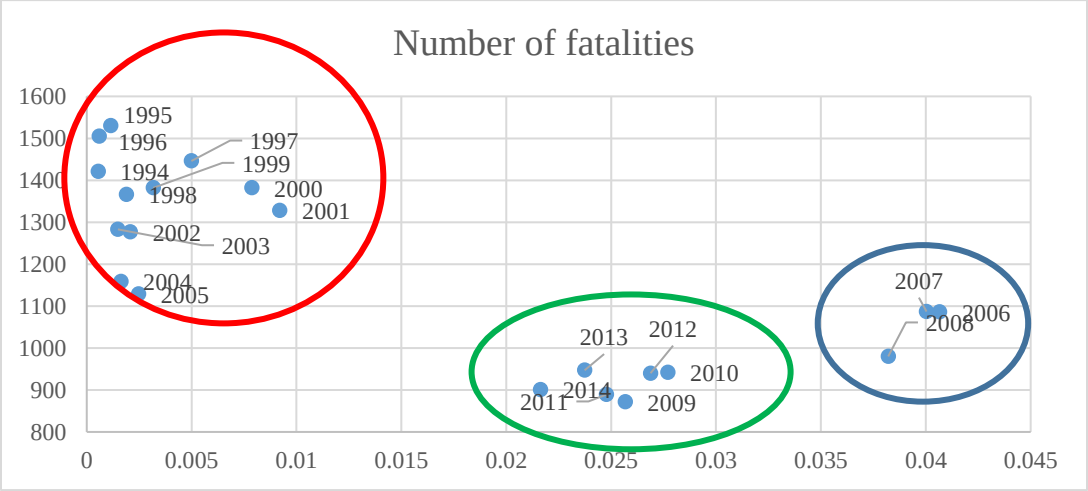

Figure 6.2. Scatter plot diagram of number of fatalities vs distracted driving

The scatter plot diagram also supported that the distracted driving had increased in the recent years. However, the scatter plot did not segregate fatalities according to the cause attributed to the fatalities. In the past, speeding and drunk driving were the dominant factors of road fatalities, while distracted driving was not. However, IIHS data showed that the recently the fatalities associated with speeding and drunk driving decreased. Even though fatalities due to distracted driving increased, the overall net fatalities decreased, predominantly because of a decrease in fatalities due to speeding and drunk driving. As the net number of fatalities decreased, the regression model suggested a positive affect with increasing distracted driving.

In the regression models, variables such as ‘total number of hospitals’ and ‘number of nurses’ were found to be positively related with the number of fatalities, which is counter intuitive. The apparent positive relationship between the number of hospitals and fatalities is perhaps a function of the direction of causation. With increases in injuries from traffic crashes, there is a need for more hospitals. Some of those injuries will result in death. So, the number of hospitals appears to correlate with increased fatalities. Given that nurses are employed at hospitals, more hospitals equate to more nurses. And so, more nurses also appear to correlate with fatalities. But in both cases, it is logical that both these variables reflect increased crash injuries (fatalities), rather than determinants of them.

Finally, from panel data regression model, the coefficient associated with population variable was found to be negative Within a geographic boundary can also be related with population density. If the area of each states is considered time invariant the

population density will gradually rise. Clark and Cushing (2004) studied urban and rural traffic fatalities in different states and found a negative correlation. A similar study by Clark (2003) showed that mortality was higher in locations with population less than 25,000 and lower in urban areas. The association from panel regression model analysis is consistent with these research findings.

6.3. Relative Advantages and Weaknesses of Each Method

A significant advantage that panel regression model holds over the linear regression model is the number of data points. In the linear regression, the total number of data points were 21 whereas for panel data regression modelling the total number of data points were 1050. As panel regression data contain more degrees of freedom and statistical variability than linear regression data more accurate inference of model parameter is obtained (Hsiao, Mountain, and Illman 1995). However, for linear regression model, the study obtained 48 state wise linear regression model which can be readily used for specific purposes and to directly obtain predictive number of fatality.

For the panel data regression modelling, the goodness of fit is low around 13.81%. Also the coefficients of the variables are the same throughout all the states. Thus state-wise magnitude of effects of these variables cannot be understood. However, with linear regression modelling the magnitude of the effect easily distinguished. Apart from two instances (Arizona and Oklahoma) the goodness of fit statistics in terms of adjusted r – squared value were quite good, going up to 95.06% for the states of Illinois.

One of the significant benefits of using panel data regression analysis is that it generates accurate predictions. However, compared to linear regression models, policy evaluations such as law or effect of categorical variables may be misleading (Hsiao 2007). This hypothesis was not tested in the analysis, since the forms of the dependent variable were different in analytical models. While comparing with linear regression models, the challenge of panel regression methodology was impact of heteroscedasticity, which had been illustrated by figure 5.3. Hsiao (2007) also suggested that to obtain valid inferences of coefficients proper analytical modelling approach is required. This was done in this study while the dependent variable was changed to 'fatalities per billion VMT" and it showed that heteroscedasticity residual pattern was greatly reduced. To obtain correct estimations and inferences from the parameters proper selection of panel regression model is considered very important. Compared to linear regression models, panel regression modelling is complicated because of these added steps.

6.4. Inferring the Safe States and Dangerous States for Driving

This study has suggested which states possesses protective culture and which states possesses risky culture in terms of road safety culture and other behavioral aspects. In order to compare our results with the most recent '2014 status of state-wise fatality per 100 million vehicle miles travelled', this thesis referred to the results given by 'Insurance Institute for Highway Safety'(IIHS) (Source: http://www.iihs.org/iihs/topics/t/general-statistics/fatalityfacts/state-by-state-overview, Accessed at March, 2017). A comparative

table is given below that shows the ranking of states by fatality per 100 million VMT and by traffic safety culture indicator:

Table 6.2. Ranking of traffic safety culture I(2014) and fatality per 100 mil. VMT

Rank	State	Fatality per 100 million VMT	Rank	State	I(2014)
1	Massachusetts	0.57	1	New Hampshire	194.51
2	Vermont	0.62	2	Massachusetts	192.99
3	Minnesota	0.63	3	Minnesota	192.77
4	Rhode Island	0.68	4	Hawaii	163.57
5	New Hampshire	0.73	5	Vermont	161.69
6	New Jersey	0.74	6	Rhode Island	161.39
7	Maryland	0.78	7	New Jersey	160.59
8	Connecticut	0.8	8	Maryland	150.81
9	New York	0.8	9	Virginia	149.74
10	Washington	0.8	10	Washington	145.59
11	Wisconsin	0.84	11	Wisconsin	145.04
12	Virginia	0.87	12	Connecticut	141.62
13	Illinois	0.88	13	Delaware	140.96
14	Ohio	0.89	14	Illinois	139.94
15	California	0.92	15	Ohio	136.81
16	Maine	0.92	16	Iowa	136.64
17	Hawaii	0.93	17	Georgia	136.50
18	Michigan	0.93	18	Colorado	135.19
19	Utah	0.93	19	Oregon	134.16
20	Indiana	0.94	20	Nevada	132.58
21	Colorado	1.00	21	Missouri	131.55
22	Iowa	1.02	22	Nebraska	130.00
23	Oregon	1.03	23	Indiana	129.00
24	Georgia	1.04	24	Utah	126.80
25	Missouri	1.08	25	Kansas	125.05
26	Idaho	1.15	26	Maine	124.95
27	Nebraska	1.15	27	Alabama	122.12
28	Nevada	1.15	28	Florida	121.72
29	North Carolina	1.19	29	Michigan	120.05
30	Pennsylvania	1.20	30	Louisiana	114.73

Table 6.2. Ranking of traffic safety culture I(2014) and fatality per 100 mil. VMT – continued

Rank	State	Fatality per 100 million VMT	Rank	State	I(2014)
31	Arizona	1.23	31	North Carolina	114.46
32	Florida	1.24	32	Texas	113.54
33	Alabama	1.25	33	California	112.53
34	Kansas	1.25	34	Arizona	112.26
35	Delaware	1.26	35	Idaho	111.96
36	North Dakota	1.28	36	North Dakota	110.89
37	Tennessee	1.33	37	Wyoming	109.81
38	Arkansas	1.37	38	Oklahoma	109.29
39	Kentucky	1.4	39	New York	108.42
40	Oklahoma	1.4	40	South Dakota	108.00
41	West Virginia	1.42	41	New Mexico	107.48
42	Texas	1.46	42	Tennessee	107.42
43	South Dakota	1.47	43	Mississippi	105.86
44	Alaska	1.50	44	Arkansas	103.90
45	New Mexico	1.51	45	Pennsylvania	101.81
46	Louisiana	1.53	46	Kentucky	98.74
47	Mississippi	1.54	47	South Carolina	95.25
48	Montana	1.58	48	Montana	90.67
49	Wyoming	1.59	49	West Virginia	84.76
50	South Carolina	1.65	50	Alaska	77.84

A correlation analysis between the fatality per 100 million VMT and the performance indicators obtained from 2014 revealed a strong negative correlation. This implies that when the traffic safety culture is lower, the fatality per 100 million VMT will be higher. An illustration is given below:

Correlation: I(2014), Fatality per 100 million VMT

```
Pearson correlation of I(2014) and Fatality per 100 million VMT = -0.821
P-Value = 0.000
```

Figure 6.3. Correlation analysis revealing a strong negative correlation

It is not evident from the performance indicators that which factors contributed to the improvement, however, it can be inferred in general that the enforcement of laws played a vital role in terms of reducing the state-wise fatalities. It can be speculated that the roads and highways became safer in terms of proper signage, visibility, guiderails, and traffic signals at vital intersections. The vehicles in terms of safety features has increased the likelihood of safer travel and driving experiences by implementing collision alert features, navigations, braking systems etc.

CHAPTER SEVEN – RECOMMENDATIONS

The impact of this thesis is tremendous due to the comprehensiveness of the study, the number of independent variables being considered, as well as the scope of the study which encompassed all 50 states. The thesis focused on various aspects which can be helpful for future research purposes in developing government policy, road safety strategies, and budgeting decisions concerning road safety improvement projects.

The findings of this thesis can be helpful for state-level governments or administrations if more focus were to be given to the significant factors obtained from the linear regression model. For example, it was found from the linear regression model that distracted driving is a significant factor contributing towards a number of road fatalities in Montana. Laws associated with phones or handheld devices which have been discussed in this thesis are -

1. Use of handheld devices while driving,
2. Texting while driving and
3. Cellphone use while driving for novice or teen drivers

These activities are legal in Montana roads as there is no law currently enforced (Chase 2014). This is a significant correlation since the enforcement of these laws can reduce the amount of distracted driving. A similar approach can be taken in terms of identifying variables associated with 'total number of hospitals' such as location, facility equipment, skill level, responsiveness of the emergency response team etc.

The second aspect of the thesis which could provide insight to the state-level administrations as well as road safety researchers is the performance indicators. Although

it should be noted that the formulization process of the indicators can be improved. It will prove very effective in terms of gathering information about the states which are lagging in indicator ranking so that it can replicate road safety culture, laws or legislations from the states which did better in terms of a performance indicator.

The inferred safety culture indicator does not give clear indication as to which specific unobserved factors are contributing to road fatalities. Extensive and elaborative research needs to be carried out in order to uncover these factors by defining specific geographic boundaries. These geographic boundaries can be within a state, city or a county which in turn shall have influence over the factors associated with human population, socio-cultural, economic and meteorological aspects. Rather than grouping everything that was done in this thesis into one performance group, overall unobserved safety culture performance can be divided into different subgroups. These subgroups are:

- Safety improvement policies
- Safety cultural
- Driving behavior and culture
- Effectiveness criteria of road improvement and safer vehicles
- Political factors
- Geographical factors

Based on the initial approach suggested by this thesis, road safety researchers can strive towards achieving results associated with urban and rural studies, within specific state level study. For specific state level analysis, the linear regression model and the significant variables associated with it can be a good starting point. However, for such

deep level inferences proper variables on culture, attitude, behaviors and norms related with road safety should be included.

This thesis has included people's view point on politics as a proxy for culture. The discussion in chapter 5 revealed that the inclusion of these variables has brought significant change in terms of improving the goodness of fit in both modelling approaches. Since it is difficult to infer something about culture, this thesis study has offered two suggestions – first to infer road safety culture from performance indicators, and secondly, to use using political viewpoints as a proxy for culture.

CHAPTER EIGHT – CONCLUSION AND FUTURE RESEARCH

As after reviewing the literature it was understood that a comprehensive analysis that encompasses all 50 states was missing, it was necessary to come up with a structural approach to explore the causes behind road fatalities. It was also noticed during the literature review process that the selection procedure of the factors concerning road fatality was biased by elimination of the factors with missing data. Therefore, after selecting the variables, the missing data points were imputed and analyzed.

Data on 47 exogenous variables were collected state-wise from 1994 to 2014 in the process, among which 9 variables had missing data points within this time frame. With multiple imputation analysis the missing data points were obtained and estimated. 10 imputed datasets were obtained. Two modelling approaches were considered – linear regression and panel data regression.

After analyzing the state wise datasets, 48 linear regression models were obtained except for Arizona and Oklahoma. The overall procedure of variable selection and data modelling was done in R software by using several of R-packages. Drunk Driving, distracted driving and number of unemployed people were found as top three significant factors.

Panel data regression analysis procedure with random effects two-way modelling was done in a similar approach to the study by Page. The data analysis process was done in SAS environment. The main purpose of this section was to define the progression of road safety cultures and other exogenous variables which were not included in regression analysis. Another way this thesis incorporated cultural aspects is by taking political

variables which represents people's view point on political parties. This was done as an approach to draw out some inferences regarding culture. And the results showed significant improvement in model fit results.

Considering the overall comprehensiveness of the study there are some limitations which needs to be addressed in future researches.

1. However, there were some limitations with the datasets for the variables which had been associated with risky driving behavior (speeding, drunk driving and distracted driving) and healthcare related variables (hospitals, nurses, physicians and beds per 1000 people). In an ideal situation, the risky driving behavior should be captured for all types of crashes rather than only for fatal crashes. An ideal dataset associated with road traffic safety research should include all types of crashes per their level of severity. Similarly, the additional data associated with federal hospitals, long term care hospital, psychiatric hospitals and other form of nursing cares would have been helpful in terms to correctly determining the estimates as well as their relation to the fatality.
2. According to the conceptual framework presented in this thesis, cultural factors were not included since to the best knowledge of the author no such reliable source of information were found which ranges the time frame and covers all states. It is very significant to either obtain such information on driving culture, behavior, norms or to formulate different dimensions of driving culture or culture as a general which will differentiate each state.

3. The dependent variable in this thesis was considered to be fatalities and fatalities per billion vehicle miles travelled. The dependent variables can be stated in other form of raw number such as fatalities per thousand/million inhabitant or fatalities per unit income etc. In such cases, the independent variables also can be formed in such manner where each variable are considered from a certain benchmark value such as precipitation, temperature, snow level depth etc. Economic indicators such as GDP per capita, unemployment can be considered as a growth of increasing or decreasing percentage basing upon the previous year value.
4. The data imputation structure has been based on the assumption that the missing data is of missing at random (MAR) category. It was not explored whether this assumption is valid, since it would require tremendous amount of background information. If the missingness is not MAR, then there are two strategies which can be followed. Firstly, identifying additional information to make the data from MAR and secondly sensitivity analysis.
5. The basic linear regression models have been considered for formulizing the relationship between dependent and independent variables. Better statistical modelling approaches can be exercised. As road fatalities are discrete, countable and random events with non – zero mean or variance, different type of Poisson regression model can be exercised with a log link function.
6. The value obtained for the indicators for each state, and their progression should not be considered definitive since many exogenous variables were left out from the

model. Future studies should include these variables, and thus update the statistical indicators here and explore new procedures to create such indicators.

Even though the time span in this thesis was taken to be 21 years, within one-year difference between each time point, the number of data-points can be extended by taking month wise data points. However, it might be a rigorous task to collect continuous variables with such extensive detail information. That being said obtaining the effect of categorical variables such as laws will be effected, since they can be precisely modelled with the time of its enforcement. According to the matrix, the factors such as economic recessions, and other form of interventions can also be modelled if the starting and ending of its effect are clearly defined.

REFERENCES CITED

Antoniou, Constantinos, George Yannis, Eleonora Papadimitriou, and Sylvain Lassarre. 2016. "Relating traffic fatalities to GDP in Europe on the long term." *Accident Analysis & Prevention* 92:89-96.

Baltagi, Badi. 2008. *Econometric analysis of panel data*: John Wiley & Sons.

Barnard, John, and Xiao-Li Meng. 1999. "Applications of multiple imputation in medical studies: from AIDS to NHANES." *Statistical methods in medical research* 8 (1):17-36.

Bonnett, Alastair. 2008. *What is geography?*: Sage.

Brand, Jaap. 1999. *Development, implementation and evaluation of multiple imputation strategies for the statistical analysis of incomplete data sets*.

Buuren, Stef, and Karin Groothuis-Oudshoorn. 2011. "mice: Multivariate imputation by chained equations in R." *Journal of statistical software* 45 (3).

Calkins, Lindsay Noble, and Thomas J Zlatoper. 2001. "The effects of mandatory seat belt laws on motor vehicle fatalities in the United States." *Social Science Quarterly* 82 (4):716-732.

Cestac, Julien, Sami Kraïem, and Jean-Pascal Assailly. 2016. "Cultural values and random breath tests as moderators of the social influence on drunk driving in 15 countries." *Journal of safety research* 56:89-96.

Chase, JD Catherine. 2014. "US state and federal laws targeting distracted driving." *Annals of advances in automotive medicine* 58:84.

Chin, Hoong Chor, and Mohammed Abdul Quddus. 2003. "Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections." *Accident Analysis & Prevention* 35 (2):253-259.

Clark, David E. 2003. "Effect of population density on mortality after motor vehicle collisions." *Accident Analysis & Prevention* 35 (6):965-971.

Clark, David E, and Brad M Cushing. 2004. "Rural and urban traffic fatalities, vehicle miles, and population density." *Accident Analysis & Prevention* 36 (6):967-972.

Cohen, Alma, and Liran Einav. 2003. "The effects of mandatory seat belt laws on driving behavior and traffic fatalities." *Review of Economics and Statistics* 85 (4):828-843.

Dong, Yiran, and Chao-Ying Joanne Peng. 2013. "Principled missing data methods for researchers." *SpringerPlus* 2 (1):222.

Dupont, Emmanuelle, Jacques JF Commandeur, Sylvain Lassarre, Frits Bijleveld, Heike Martensen, Constantinos Antoniou, Eleonora Papadimitriou, George Yannis, Elke Hermans, and Katherine Pérez. 2014. "Latent risk and trend models for the evolution of annual fatality numbers in 30 European countries." *Accident Analysis & Prevention* 71:327-336.

Eby, David W, and Jonathon M Vivoda. 2003. "Driver hand-held mobile phone use and safety belt use." *Accident Analysis & Prevention* 35 (6):893-895.

Eiksund, Sveinung. 2009. "A geographical perspective on driving attitudes and behaviour among young adults in urban and rural Norway." *Safety Science* 47 (4):529-536.

Eisenberg, Daniel, and Kenneth E Warner. 2005. "Effects of snowfalls on motor vehicle collisions, injuries, and fatalities." *American journal of public health* 95 (1):120-124.

Farmer, Charles M. 1997. "Trends in motor vehicle fatalities." *Journal of Safety Research* 28 (1):37-48.

Fridstrøm, Lasse, Jan Ifver, Siv Ingebrigtsen, Risto Kulmala, and Lars Krogsgård Thomsen. 1995. "Measuring the contribution of randomness, exposure, weather, and daylight to the variation in road accident counts." *Accident Analysis & Prevention* 27 (1):1-20.

Fuller, Wayne A, and George E Battese. 1974. "Estimation of linear models with crossed-error structure." *Journal of Econometrics* 2 (1):67-78.

Gaygısız, Esma. 2010. "Cultural values and governance quality as correlates of road traffic fatalities: A nation level analysis." *Accident Analysis & Prevention* 42 (6):1894-1901.

Haddon, William. 1980. "Options for the prevention of motor vehicle crash injury." *Israel journal of medical sciences* 16 (1):45-65.

Hakim, Simon, Daniel Shefer, Alfred-Shalom Hakkert, and Irit Hocherman. 1991. "A critical review of macro models for road accidents." *Accident Analysis & Prevention* 23 (5):379-400.

Hausman, Jerry A. 1978. "Specification tests in econometrics." *Econometrica: Journal of the Econometric Society*:1251-1271.

He, Yan. 2006. "Missing data imputation for tree-based models." University OF California Los Angeles.

Heemskerk, Marieke, Karen Wilson, and Mitchell Pavao-Zuckerman. 2003. "Conceptual models as tools for communication across disciplines." *Conservation Ecology* 7 (3).

Heitjan, Daniel F, and Roderick JA Little. 1991. "Multiple imputation for the fatal accident reporting system." *Applied Statistics*:13-29.

Hsiao, Cheng. 2007. "Panel data analysis—advantages and challenges." *Test* 16 (1):1-22.

Hsiao, Cheng, Dean C Mountain, and Kathleen Ho Illman. 1995. "A Bayesian integration of end-use metering and conditional-demand analysis." *Journal of Business & Economic Statistics* 13 (3):315-326.

Hughes, BP, S Newstead, Anna Anund, CC Shu, and Torbjörn Falkmer. 2015. "A review of models relevant to road safety." *Accident Analysis & Prevention* 74:250-270.

Law, Teik Hua, Robert B Noland, and Andrew W Evans. 2013. "Factors associated with the enactment of safety belt and motorcycle helmet laws." *Risk analysis* 33 (7):1367-1378.

Leigh, J Paul, and Harold M Waldon. 1991. "Unemployment and highway fatalities." *Journal of health politics, policy and law* 16 (1):135-156.

Li, Kim-Hung, Xiao-Li Meng, Trivellore E Raghunathan, and Donald B Rubin. 1991. "Significance levels from repeated p-values with multiply-imputed data." *Statistica Sinica*:65-92.

Little, Roderick JA, and Donald B Rubin. 2014. *Statistical analysis with missing data*: John Wiley & Sons.

Lozano, Rafael, Mohsen Naghavi, Kyle Foreman, Stephen Lim, Kenji Shibuya, Victor Aboyans, Jerry Abraham, Timothy Adair, Rakesh Aggarwal, and Stephanie Y Ahn. 2013. "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." *The Lancet* 380 (9859):2095-2128.

McCartt, Anne T, David G Kidd, and Eric R Teoh. 2014. "Driver cellphone and texting bans in the United States: evidence of effectiveness." *Annals of Advances in Automotive Medicine* 58:99-114.

Mishkin, Frederic S. 1990. "Does correcting for heteroscedasticity help?" *Economics Letters* 34 (4):351-356.

Morris, Tim P, Ian R White, and Patrick Royston. 2014. "Tuning multiple imputation by predictive mean matching and local residual draws." *BMC medical research methodology* 14 (1):75.

Nabi, Hermann, Louis Rachid Salmi, Sylviane Lafont, Mireille Chiron, Marie Zins, and Emmanuel Lagarde. 2007. "Attitudes associated with behavioral predictors of serious road traffic crashes: results from the GAZEL cohort." *Injury Prevention* 13 (1):26-31.

Naumann, Rebecca B, Ann M Dellinger, Eduard Zaloshnja, Bruce A Lawrence, and Ted R Miller. 2010. "Incidence and total lifetime costs of motor vehicle–related fatal and nonfatal injury by road user type, United States, 2005." *Traffic injury prevention* 11 (4):353-360.

Newey, Whitney K, and Kenneth D West. 1986. A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix. National Bureau of Economic Research Cambridge, Mass., USA.

Noland, Robert B. 2003. "Traffic fatalities and injuries: the effect of changes in infrastructure and other trends." *Accident Analysis & Prevention* 35 (4):599-611.

Norman, Leslie George, and World Health Organization. 1962. "Road traffic accidents: epidemiology, control, and prevention."

O'brien, Robert M. 2007. "A caution regarding rules of thumb for variance inflation factors." *Quality & Quantity* 41 (5):673-690.

Organization, World Health. 2009. *Global status report on road safety: time for action*: World Health Organization.

Othman, Sarbaz, Robert Thomson, and Gunnar Lannér. 2009. "Identifying critical road geometry parameters affecting crash rate and crash type." *Annals of advances in automotive medicine* 53.

Page, Yves. 2001. "A statistical model to compare road mortality in OECD countries." *Accident Analysis & Prevention* 33 (3):371-385.

Partyka, Susan C. 1984. "Simple models of fatality trends using employment and population data." *Accident Analysis & Prevention* 16 (3):211-222.

Peden, Margie, Richard Scurfield, David Sleet, Dinesh Mohan, Adnan A Hyder, Eva Jarawan, and Colin D Mathers. 2004. World report on road traffic injury prevention. World Health Organization Geneva.

Peltzman, Sam. 1975. "The effects of automobile safety regulation." *Journal of political Economy* 83 (4):677-725.

Peng, Chao-Ying Joanne, Michael Harwell, Show-Mann Liou, and Lee H Ehman. 2006. "Advances in missing data methods and implications for educational research." *Real data analysis*:31-78.

Qiu, Lin, and Wilfrid Nixon. 2008. "Effects of adverse weather on traffic crashes: systematic review and meta-analysis." *Transportation Research Record: Journal of the Transportation Research Board* (2055):139-146.

Quddus, Mohammed A. 2008. "Time series count data models: an empirical application to traffic accidents." *Accident Analysis & Prevention* 40 (5):1732-1741.

Quimby, AR, Crowthorne House, and Nine Mile Ride. 2005. "Comparing UK and European drivers on speed and speeding issues: some results from SARTRE 3 survey." Behavioural Research in Road Safety: Fifteenth Seminar. London: Department for Transport.

Roth, Philip L. 1994. "Missing data: A conceptual review for applied psychologists." *Personnel psychology* 47 (3):537-560.

Rubin, Donald B. 1976. "Inference and missing data." *Biometrika*:581-592.

Rubin, Donald B. 2004. *Multiple imputation for nonresponse in surveys*. Vol. 81: John Wiley & Sons.

Rumar, Kåre. 1999. "Transport safety visions, targets and strategies: beyond 2000." *1st European Transport Safety Lecture. European Transport Safety Council, Brussels, Tech. Rep*.

Schafer, Joseph L. 1999. "Multiple imputation: a primer." *Statistical methods in medical research* 8 (1):3-15.

Schafer, Joseph L, and John W Graham. 2002. "Missing data: our view of the state of the art." *Psychological methods* 7 (2):147.

Schafer, Joseph L, and Maren K Olsen. 1998. "Multiple imputation for multivariate missing-data problems: A data analyst's perspective." *Multivariate behavioral research* 33 (4):545-571.

Schepers, Paul, Marjan Hagenzieker, Rob Methorst, Bert Van Wee, and Fred Wegman. 2014. "A conceptual framework for road safety and mobility applied to cycling safety." *Accident Analysis & Prevention* 62:331-340.

Singh, Santokh. 2015. Critical reasons for crashes investigated in the national motor vehicle crash causation survey.

Stock, JH, and MW Watson. 2003. "Introduction to econometrics. Boston: Addison-Wesley."

Subramanian, Rajesh, and Dennis Utter. 1998. "Multiple imputation of missing blood alcohol concentration (BAC) values in FARS." In *NHTSA Technical Report DOT HS 808 816*.

Tabachnick, Barbara G, Linda S Fidell, and Steven J Osterlind. 2001. "Using multivariate statistics."

Tefft, Brian Christopher. 2012. "Prevalence of motor vehicle crashes involving drowsy drivers, United States, 1999–2008." *Accident Analysis & Prevention* 45:180-186.

van Beeck, Eduard F, Gerard JJ Borsboom, and Johan P Mackenbach. 2000. "Economic development and traffic accident mortality in the industrialized world, 1962–1990." *International journal of epidemiology* 29 (3):503-509.

Van Buuren, Stef. 2012. *Flexible imputation of missing data*: CRC press.

Vanlaar, Ward, and George Yannis. 2006. "Perception of road accident causes." *Accident Analysis & Prevention* 38 (1):155-161.

Vereeck, Lode, and Klara Vrolix. 2007. "The social willingness to comply with the law: The effect of social attitudes on traffic fatalities." *International Review of Law and Economics* 27 (4):385-408.

Vergouwe, Yvonne, Patrick Royston, Karel GM Moons, and Douglas G Altman. 2010. "Development and validation of a prediction model with missing predictor data: a practical approach." *Journal of clinical epidemiology* 63 (2):205-214.

Whetten-Goldstein, Kathryn, Frank A Sloan, Emily Stout, and Lan Liang. 2000. "Civil liability, criminal law, and other policies and alcohol-related motor vehicle fatalities in the United States: 1984–1995." *Accident Analysis & Prevention* 32 (6):723-733.

Wijnen, Wim, and Henk Stipdonk. 2016. "Social costs of road crashes: An international analysis." *Accident Analysis & Prevention* 94:97-106.

Wood, Angela M, Ian R White, and Patrick Royston. 2008. "How should variable selection be performed with multiply imputed data?" *Statistics in medicine* 27 (17):3227-3246.

Zador, Paul L, Sheila A Krawchuk, and Robert B Voas. 2000. "Alcohol-related relative risk of driver fatalities and driver involvement in fatal crashes in relation to driver age and gender: an update using 1996 data." *Journal of studies on alcohol* 61 (3):387-395.

Zlatoper, Thomas J. 1991. "Determinants of motor vehicle deaths in the United States: a cross-sectional analysis." *Accident Analysis & Prevention* 23 (5):431-436.

APPENDICES

APPENDIX A

LINEAR REGRESSION WITH POOLED ESTIMATES

```
                              est          se        t       df     Pr(>|t|)        lo 95        hi 95 nmis      fmi     lambda
(Intercept)            502.87249334 279.86904976  1.796814  7.762353 0.1112328600 -145.96215362 1151.70714030   NA 0.5375137 0.43195161
Unemployed              -0.02124549   0.00512867 -4.142495 15.639571 0.0007999271   -0.03213818   -0.01035279    0 0.1406345 0.03734273
Total.Number.of.Hospitals  7.18756846   2.35501578  3.052026  7.085105 0.0182580868    1.63237227   12.74276466    5 0.5769378 0.47228550
> pool.r.squared(fit3)
        est     lo 95     hi 95      fmi
R^2 0.722894 0.4108246 0.8870808 0.1402273
> pool.r.squared(fit3,adjusted = TRUE)
             est     lo 95     hi 95      fmi
adj R^2 0.6919944 0.3615892 0.8733635 0.1465332
>
```

Figure A.1. Road fatality prediction model for Alabama

```
                                  est      se        t     df     Pr(>|t|)     lo 95     hi 95 nmis        fmi lambda
(Intercept)                  86.25000 3.083017 27.975841 17.271 8.881784e-16  79.75317  92.74683   NA 0.09866312      0
factor(Child.Restraint.Law)1 -20.36111 4.709387 -4.323517 17.271 4.459889e-04 -30.28519 -10.43704   NA 0.09866312      0
> pool.r.squared(fit3)
         est     lo 95     hi 95         fmi
R^2 0.4959249 0.1533222 0.7588052 2.000442e-05
> pool.r.squared(fit3,adjusted = TRUE)
             est     lo 95   hi 95         fmi
adj R^2 0.4693946 0.1295029 0.74287 2.000442e-05
>
```

Figure A.2. Road fatality prediction model for Alaska

```
                    est       se        t     df    Pr(>|t|)     lo 95     hi 95 nmis        fmi lambda
(Intercept)    215.0203 112.4407 1.912300 17.271 0.072570627 -21.92559  451.9662   NA 0.09866312      0
Drunk.Driving 1789.1378 508.0068 3.521878 17.271 0.002564926 718.61689 2859.6587    0 0.09866312      0
> pool.r.squared(fitark)
          est      lo 95     hi 95         fmi
R^2 0.3949743 0.07292138 0.6954168 2.000442e-05
> pool.r.squared(fitark,adjusted = TRUE)
              est      lo 95     hi 95         fmi
adj R^2 0.3631308 0.05336228 0.6736789 2.000442e-05
>
```

Figure A.3. Road fatality prediction model for Arkansas

```
                         est          se        t       df     Pr(>|t|)         lo 95         hi 95 nmis       fmi lambda
(Intercept)        1.833188e+03 8.796991e+02  2.083881 15.29847 0.0543355660 -3.866507e+01  3.705041e+03   NA 0.1092988      0
Drunk.Driving      1.040926e+04 3.160980e+03  3.293050 15.29847 0.0048203207  3.683227e+03  1.713530e+04    0 0.1092988      0
Distracted.Driving 6.864806e+03 2.385118e+03  2.878183 15.29847 0.0113133531  1.789673e+03  1.193994e+04    0 0.1092988      0
Unemployed        -6.678485e-04 1.505712e-04 -4.435434 15.29847 0.0004602365 -9.882389e-04 -3.474582e-04    0 0.1092988      0
> pool.r.squared(fitcal)
         est     lo 95     hi 95         fmi
R^2 0.7947695 0.5587158 0.9129304 2.000442e-05
> pool.r.squared(fitcal,adjusted = TRUE)
             est     lo 95     hi 95         fmi
adj R^2 0.7585524 0.4954979 0.8962402 2.000442e-05
>
```

Figure A.4. Road fatality prediction model for California

```
                      est        se         t     df     Pr(>|t|)     lo 95     hi 95 nmis        fmi lambda
(Intercept)        342.3192  28.78153 11.893710 17.271 9.557215e-10  281.6679  402.9704   NA 0.09866312      0
Distracted.Driving 2623.1551 298.00604  8.802355 17.271 8.546779e-08 1995.1680 3251.1422    0 0.09866312      0
> pool.r.squared(fitcol)
        est    lo 95     hi 95          fmi
R^2 0.803071 0.573792 0.9166937 2.000442e-05
> pool.r.squared(fitcol,adjusted = TRUE)
             est     lo 95     hi 95          fmi
adj R^2 0.7927063 0.5550029 0.9119916 2.000442e-05
>
```

Figure A.5. Road fatality prediction model for Colorado

```
                   est        se         t     df     Pr(>|t|)     lo 95     hi 95 nmis        fmi lambda
(Intercept)    473.5294  55.83384  8.481047 17.271 1.448689e-07  355.8710  591.1879   NA 0.09866312      0
Drunk.Driving -559.1010 175.55152 -3.184826 17.271 5.338696e-03 -929.0401 -189.1619    0 0.09866312      0
> pool.r.squared(fitconn)
         est      lo 95    hi 95          fmi
R^2 0.348045 0.04508729 0.663035 2.000442e-05
> pool.r.squared(fitconn,adjusted = TRUE)
              est      lo 95     hi 95          fmi
adj R^2 0.3137316 0.02868947 0.6379039 2.000442e-05
>
```

Figure A.6. Road fatality prediction model for Connecticut

```
                                          est          se        t       df     Pr(>|t|)        lo 95        hi 95 nmis      fmi
(Intercept)                      1.390411e+02 8.848708e+00 15.713154 12.82631 9.346519e-10  1.198983e+02  1.581839e+02   NA 0.3084536
Road.Improvement.Capital.Expenditures -9.329028e-05 4.220447e-05 -2.210436 12.82690 4.587746e-02 -1.845927e-04 -1.987827e-06    1 0.3084270
                                        lambda
(Intercept)                          0.2084202
Road.Improvement.Capital.Expenditures 0.2083940
> pool.r.squared(fitdel)
         est        lo 95     hi 95       fmi
R^2 0.2449886 0.0007955429 0.6154682 0.2000814
> pool.r.squared(fitdel,adjusted = TRUE)
             est       lo 95     hi 95       fmi
adj R^2 0.2020207 0.002417322 0.5909804 0.2596047
>
```

Figure A.7. Road fatality prediction model for Delaware

```
                                   est          se        t       df     Pr(>|t|)        lo 95        hi 95 nmis      fmi lambda
(Intercept)                  2.209586e+03 4.238373e+02  5.213288 14.31436 0.0001222244  1.302415e+03  3.116757e+03   NA 0.1155111      0
factor(Texting.while.driving)1 -4.031819e+02 1.343440e+02 -3.001115 14.31436 0.0093382548 -6.907286e+02 -1.156351e+02   NA 0.1155111      0
Drunk.Driving                4.694187e+03 1.563847e+03  3.001693 14.31436 0.0093274690  1.346967e+03  8.041407e+03    0 0.1155111      0
Distracted.Driving           1.648932e+03 5.015524e+02  3.287656 14.31436 0.0052563883  5.754208e+02  2.722442e+03    0 0.1155111      0
Unemployed                  -1.188031e-03 2.330403e-04 -5.097964 14.31436 0.0001514875 -1.686825e-03 -6.892372e-04    0 0.1155111      0
> pool.r.squared(fitflo)
         est     lo 95     hi 95          fmi
R^2 0.7693702 0.5139488 0.9012728 2.000442e-05
> pool.r.squared(fitflo,adjusted = TRUE)
             est     lo 95   hi 95         fmi
adj R^2 0.7117127 0.4197657 0.87396 2.000442e-05
```

Figure A.8. Road fatality prediction model for Florida

```
                          est         se         t     df     Pr(>|t|)      lo 95      hi 95 nmis        fmi lambda
(Intercept)          1664.751   39.10018 42.576565 17.271 0.000000e+00   1582.356   1747.147   NA 0.09866312      0
Distracted.Driving -18607.274 3035.67443 -6.129536 17.271 1.037525e-05 -25004.341 -12210.208    0 0.09866312      0
> pool.r.squared(fit)
          est     lo 95     hi 95          fmi
R^2 0.6641401 0.3495959 0.8504588 2.000442e-05
> pool.r.squared(fit,adjusted = TRUE)
              est     lo 95     hi 95          fmi
adj R^2 0.6464633 0.3252135 0.8414847 2.000442e-05
```

Figure A.9. Road fatality prediction model for Georgia

```
                                          est          se        t       df    Pr(>|t|)         lo 95         hi 95 nmis       fmi lambda
(Intercept)                      -3.765881e+02 2.325968e+02 -1.619060 15.29847 0.125861514 -8.715154e+02  1.183392e+02   NA 0.1092988      0
Temperature                       7.189460e+00 3.108758e+00  2.312647 15.29847 0.035045545  5.745417e-01  1.380438e+01    0 0.1092988      0
factor(Handheld.phones.while.driving)1 -3.298629e+01 8.991048e+00 -3.668793 15.29847 0.002214937 -5.211775e+01 -1.385484e+01   NA 0.1092988      0
Unemployed                       -9.422509e-04 3.089268e-04 -3.050078 15.29847 0.007955786 -1.599596e-03 -2.849061e-04    0 0.1092988      0
> pool.r.squared(fit1)
         est     lo 95     hi 95         fmi
R^2 0.6303521 0.3037759 0.8331848 2.000442e-05
> pool.r.squared(fit1,adjusted = TRUE)
             est     lo 95     hi 95         fmi
adj R^2 0.5651201 0.2245008 0.7983108 2.000442e-05
```

Figure A.10. Road fatality prediction model for Hawaii

```
                                        est          se        t       df     Pr(>|t|)         lo 95         hi 95 nmis       fmi lambda
(Intercept)                   313.140807056 1.385257e+01 22.605255 16.28409 9.969803e-14 283.816259010  3.424654e+02   NA 0.1037125      0
factor(Texting.while.driving)1 -45.000405659 1.252060e+01 -3.594109 16.28409 2.371408e-03 -71.505312604 -1.849550e+01   NA 0.1037125      0
Unemployed                      -0.001578997 3.423675e-04 -4.611996 16.28409 2.762853e-04  -0.002303756 -8.542385e-04    0 0.1037125      0
> pool.r.squared(fit1)
          est     lo 95     hi 95         fmi
R^2 0.7101003 0.4172773 0.8731783 2.000442e-05
> pool.r.squared(fit1,adjusted = TRUE)
             est     lo 95     hi 95         fmi
adj R^2 0.6778892 0.3691897 0.8573463 2.000442e-05
>
```

Figure A.11. Road fatality prediction model for Idaho

```
                                                                  est          se        t       df     Pr(>|t|)         lo 95         hi 95 nmis
(Intercept)                                                 2056.1868837 1.122344e+02 18.320467 10.373182 3.082846e-09  1.807327e+03  2.305047e+03   NA
Speeding                                                   -1689.0917311 4.691515e+02 -3.600312 12.915038 3.261622e-03 -2.703310e+03 -6.748734e+02    0
Distracted.Driving                                         -6027.4849577 2.262168e+03 -2.664473 13.328712 1.913146e-02 -1.090238e+04 -1.152592e+03    0
Emergency.Response.Team                                       -0.0196611 6.461564e-03 -3.042777  8.837937 1.424750e-02 -3.431913e-02 -5.003065e-03    3
factor(Cellphone.use.while.driving.for.Novice.or.Teen.Driver)1  -226.8605928 4.225179e+01 -5.369254 12.331443 1.531815e-04 -3.186457e+02 -1.350755e+02   NA
                                                                     fmi     lambda
(Intercept)                                                 0.3441802 0.22885279
Speeding                                                    0.2024019 0.08776360
Distracted.Driving                                          0.1780993 0.06337851
Emergency.Response.Team                                     0.4312928 0.31567770
factor(Cellphone.use.while.driving.for.Novice.or.Teen.Driver)1 0.2356521 0.12098367
> pool.r.squared(fit3)
         est     lo 95     hi 95      fmi
R^2 0.9605137 0.8918967 0.9859104 0.223988
> pool.r.squared(fit3,adjusted = TRUE)
             est     lo 95    hi 95      fmi
adj R^2 0.9506357 0.8659325 0.982348 0.2258505
```

Figure A.12. Road fatality prediction model for Illinois

```
                                  est          se        t       df     Pr(>|t|)        lo 95        hi 95 nmis      fmi lambda
(Intercept)             7.808783e+02 1.067057e+02  7.318056 15.29847 2.254940e-06  5.538264e+02  1.007930e+03   NA 0.1092988      0
Drunk.Driving           1.140595e+03 4.115555e+02  2.771425 15.29847 1.405763e-02  2.648736e+02  2.016317e+03    0 0.1092988      0
Unemployed             -7.036743e-04 1.927863e-04 -3.650022 15.29847 2.302635e-03 -1.113891e-03 -2.934571e-04    0 0.1092988      0
factor(BAC.Level.Law)1 -7.143347e+01 2.729830e+01 -2.616774 15.29847 1.920458e-02 -1.295197e+02 -1.334725e+01   NA 0.1092988      0
> pool.r.squared(fit4)
         est     lo 95     hi 95         fmi
R^2 0.8215481 0.6081458 0.9249895 2.000442e-05
> pool.r.squared(fit4,adjusted = TRUE)
             est     lo 95     hi 95         fmi
adj R^2 0.7900566 0.5502546 0.9107838 2.000442e-05
```

Figure A.13. Road fatality prediction model for Indiana

```
                                          est          se        t       df     Pr(>|t|)        lo 95        hi 95 nmis      fmi    lambda
(Intercept)                        5.820938e+02 3.955005e+01 14.717905 11.23874 1.076559e-08  4.952698e+02  6.689178e+02   NA 0.3804271 0.2791793
Road.Improvement.Capital.Expenditures -2.740364e-04 6.511437e-05 -4.208539 11.75077 1.269344e-03 -4.162428e-04 -1.318299e-04    1 0.3568949 0.2560218
> pool.r.squared(fit1)
        est     lo 95     hi 95      fmi
R^2 0.5622708 0.1520952 0.8299353 0.3364951
> pool.r.squared(fit1,adjusted = TRUE)
            est     lo 95     hi 95      fmi
adj R^2 0.5387554 0.1266125 0.8201385 0.3478174
```

Figure A.14. Road fatality prediction model for Iowa

```
                              est          se        t       df     Pr(>|t|)       lo 95        hi 95 nmis       fmi    lambda
(Intercept)            726.63517254 115.96562367  6.265953 10.49147 7.544867e-05 469.8790069 983.391338160   NA 0.4155424 0.3138221
Emergency.Response.Team  -0.07874663   0.03193464 -2.465868 10.66390 3.197412e-02  -0.1493055  -0.008187724    3 0.4073479 0.3057263
> pool.r.squared(fit)
         est      lo 95     hi 95       fmi
R^2 0.3128713 0.01596202 0.6614847 0.1675879
> pool.r.squared(fit,adjusted = TRUE)
             est       lo 95     hi 95       fmi
adj R^2 0.2758756 0.005363947 0.6371761 0.1858907
```

Figure A.15. Road fatality prediction model for Kansas

```
                                                          est       se        t       df     Pr(>|t|)     lo 95     hi 95 nmis      fmi
(Intercept)                                          514.6690 125.25914  4.108834 16.28409 0.0007942525  249.5074  779.83048   NA 0.1037125
factor(Cellphone.use.while.driving.for.Novice.or.Teen.Driver)1 -117.4285  33.97914 -3.455901 16.28409 0.0031825446 -189.3591  -45.49799   NA 0.1037125
Drunk.Driving                                       1452.7030 525.25891  2.765689 16.28409 0.0136160602  340.7806 2564.62547    0 0.1037125
                                                    lambda
(Intercept)                                              0
factor(Cellphone.use.while.driving.for.Novice.or.Teen.Driver)1      0
Drunk.Driving                                            0
> pool.r.squared(fit)
         est     lo 95     hi 95         fmi
R^2 0.6875153 0.3832381 0.8621214 2.000442e-05
> pool.r.squared(fit,adjusted = TRUE)
             est     lo 95     hi 95         fmi
adj R^2 0.6527947 0.3338427 0.8447147 2.000442e-05
```

Figure A.16. Road fatality prediction model for Kentucky

```
                                                          est          se        t      df     Pr(>|t|)        lo 95        hi 95 nmis
(Intercept)                                      1.068858e+03 6.019816e+01 17.755666 14.92460 1.909051e-11  9.404926e+02  1.197224e+03   NA
Unemployed                                      -1.658940e-03 4.152577e-04 -3.994964 15.01916 1.168544e-03 -2.543942e-03 -7.739371e-04    0
Road.Improvement.Capital.Expenditures            8.817651e-05 2.839253e-05  3.105624 15.01995 7.224283e-03  2.766627e-05  1.486868e-04    1
factor(Cellphone.use.while.driving.for.Novice.or.Teen.Driver)1 -1.981739e+02 2.831354e+01 -6.999262 15.23758 3.929855e-06 -2.584409e+02 -1.379069e+02   NA
                                                     fmi      lambda
(Intercept)                                    0.1325735 0.023631821
Unemployed                                     0.1268477 0.017834193
Road.Improvement.Capital.Expenditures          0.1267993 0.017785122
factor(Cellphone.use.while.driving.for.Novice.or.Teen.Driver)1 0.1132711 0.004051833
> pool.r.squared(fit3)
         est     lo 95    hi 95       fmi
R^2 0.9077111 0.7820212 0.962552 0.01473759
> pool.r.squared(fit3,adjusted = TRUE)
            est    lo 95     hi 95       fmi
adj R^2 0.8914234 0.746757 0.9557151 0.01501158
```

Figure A.17. Road fatality prediction model for Louisiana

```
                      est        se         t     df     Pr(>|t|)    lo 95    hi 95 nmis        fmi lambda
(Intercept)      136.7761  7.687072 17.793002 17.271 1.513234e-12 120.5771 152.9750   NA 0.09866312      0
Distracted.Driving 262.4053 48.088880  5.456673 17.271 4.034654e-05 161.0677 363.7428    0 0.09866312      0
> pool.r.squared(fit1)
          est     lo 95     hi 95          fmi
R^2 0.6104584 0.2783283 0.8227709 2.000442e-05
> pool.r.squared(fit1,adjusted = TRUE)
              est    lo 95    hi 95          fmi
adj R^2 0.5899562 0.253274 0.811839 2.000442e-05
```

Figure A.18. Road fatality prediction model for Maine

```
                                        est       se        t     df    Pr(>|t|)     lo 95     hi 95 nmis        fmi lambda
(Intercept)                         622.375  8.179031 76.093979 17.271 0.00000e+00  605.1394  639.6106   NA 0.09866312      0
factor(Handheld.phones.while.driving)1 -142.575 16.762025 -8.505834 17.271 1.39027e-07 -177.8976 -107.2524   NA 0.09866312      0
> pool.r.squared(fit1)
         est     lo 95    hi 95          fmi
R^2 0.792007 0.5537475 0.911673 2.000442e-05
> pool.r.squared(fit1,adjusted = TRUE)
              est    lo 95     hi 95          fmi
adj R^2 0.7810599 0.534299 0.9066653 2.000442e-05
```

Figure A.19. Road fatality prediction model for Maryland

```
                                                          est      se       t     df     Pr(>|t|)     lo 95     hi 95 nmis        fmi lambda
(Intercept)                                          429.8125  8.53623 50.35156 17.271 0.0000000000  411.8241 447.80087   NA 0.09866312      0
factor(Cellphone.use.while.driving.for.Novice.or.Teen.Driver)1 -73.2125 17.49407 -4.18499 17.271 0.0006026149 -110.0777 -36.34731   NA 0.09866312      0
> pool.r.squared(fit2)
         est     lo 95     hi 95         fmi
R^2 0.4796537 0.1384865 0.7490881 2.000442e-05
> pool.r.squared(fit2,adjusted = TRUE)
            est     lo 95     hi 95         fmi
adj R^2 0.4522671 0.1151427 0.7323247 2.000442e-05
```

Figure A.20. Road fatality prediction model for Massachusetts

```
                                                                  est          se        t      df    Pr(>|t|)        lo 95        hi 95 nmis
(Intercept)                                                1.654611e+03 6.579117e+01 25.149438 15.29847 7.305268e-14  1.514618e+03  1.794604e+03   NA
factor(Cellphone.use.while.driving.for.Novice.or.Teen.Driver)1 -2.527576e+02 7.377171e+01 -3.426213 15.29847 3.659517e-03 -4.097315e+02 -9.578370e+01   NA
Distracted.Driving                                        -4.765999e+03 2.029763e+03 -2.348057 15.29847 3.270719e-02 -9.084996e+03 -4.470026e+02    0
Unemployed                                                -1.127566e-03 2.108585e-04 -5.347500 15.29847 7.611161e-05 -1.576238e-03 -6.788941e-04    0
                                                                 fmi lambda
(Intercept)                                               0.1092988      0
factor(Cellphone.use.while.driving.for.Novice.or.Teen.Driver)1 0.1092988      0
Distracted.Driving                                        0.1092988      0
Unemployed                                                0.1092988      0
> pool.r.squared(fit1)
         est     lo 95     hi 95         fmi
R^2 0.8337305 0.6314045 0.9303997 2.000442e-05
> pool.r.squared(fit1,adjusted = TRUE)
            est     lo 95     hi 95         fmi
adj R^2 0.8043888 0.5762058 0.9172891 2.000442e-05
```

Figure A.21. Road fatality prediction model for Michigan

```
                    est       se          t     df    Pr(>|t|)     lo 95     hi 95 nmis        fmi lambda
(Intercept)    -19.72655 166.8129 -0.1182555 17.271 0.907229693 -371.2509  331.7978   NA 0.09866312      0
Drunk.Driving 2291.79956 691.1844  3.3157570 17.271 0.004019054  835.2690 3748.3301    0 0.09866312      0
> pool.r.squared(fit)
          est     lo 95     hi 95          fmi
R^2 0.3665451 0.0553241 0.6760559 2.000442e-05
> pool.r.squared(fit,adjusted = TRUE)
             est      lo 95   hi 95          fmi
adj R^2 0.3332054 0.03757854 0.65233 2.000442e-05
```

Figure A.22. Road fatality prediction model for Minnesota

```
                                    est       se        t      df     Pr(>|t|)     lo 95     hi 95 nmis       fmi lambda
(Intercept)                    740.9562  38.81127 19.091267 16.28409 1.419975e-12  658.7965 823.11596   NA 0.1037125      0
factor(Texting.while.driving)1 -137.3193  50.22737 -2.733954 16.28409 1.453944e-02 -243.6458 -30.99281   NA 0.1037125      0
Distracted.Driving             504.6740 163.59972  3.084810 16.28409 6.987741e-03  158.3492 850.99885    0 0.1037125      0
> pool.r.squared(fit2)
         est     lo 95     hi 95          fmi
R^2 0.7427881 0.4692602 0.8888319 2.000442e-05
> pool.r.squared(fit2,adjusted = TRUE)
            est     lo 95     hi 95          fmi
adj R^2 0.714209 0.4236336 0.8751682 2.000442e-05
```

Figure A.23. Road fatality prediction model for Mississippi

```
                            est          se        t      df    Pr(>|t|)        lo 95        hi 95 nmis       fmi    lambda
(Intercept)          6016.52354905 1.797929e+03  3.346363 10.63391 0.006820052 2042.6236308  9.990423e+03   NA 0.3858314 0.2802489
Female.Driver         -86.34948766 3.511388e+01 -2.459127 11.25565 0.031278264 -163.4210002 -9.277975e+00    1 0.3553702 0.2501730
Emergency.Response.Team  -0.06028601 2.399786e-02 -2.512142 12.27401 0.026917134   -0.1124437 -8.128341e-03    3 0.3064449 0.2019467
> pool.r.squared(fit1)
         est    lo 95    hi 95      fmi
R^2 0.5644565 0.199422 0.809746 0.1341722
> pool.r.squared(fit1,adjusted = TRUE)
            est     lo 95     hi 95      fmi
adj R^2 0.5157572 0.1478394 0.7847457 0.1450045
```

Figure A.24. Road fatality prediction model for Missouri

```
                            est         se         t       df   Pr(>|t|)        lo 95     hi 95 nmis       fmi    lambda
(Intercept)          -79.007520 111.619581 -0.7078285  9.57378 0.49592183 -329.2204641 171.20542   NA 0.4393781 0.3333381
Distracted.Driving   197.389549  75.442701  2.6164168 12.45991 0.02195366   33.6844235 361.09467    0 0.2975896 0.1932189
Total.Number.of.Hospitals   5.771651   2.214121  2.6067461  9.43803 0.02737791    0.7981736  10.74513    5 0.4464138 0.3403428
> pool.r.squared(fit2)
          est     lo 95    hi 95      fmi
R^2 0.5625774 0.1885625 0.812995 0.1783332
> pool.r.squared(fit2,adjusted = TRUE)
              est     lo 95   hi 95      fmi
adj R^2 0.5135047 0.1367028 0.78881 0.1937063
>
```

Figure A.25. Road fatality prediction model for Montana

```
                                                                  est        se        t      df     Pr(>|t|)      lo 95     hi 95 nmis      fmi
(Intercept)                                                  349.74523  18.626971 18.776280 16.28409 1.841194e-12  310.31372  389.17673   NA 0.1037125
Speeding                                                    -538.86416 140.947099 -3.823166 16.28409 1.456255e-03 -837.23558 -240.49275    0 0.1037125
factor(Cellphone.use.while.driving.for.Novice.or.Teen.Driver)1  -66.33974   7.639273 -8.684038 16.28409 1.642533e-07  -82.51134  -50.16813   NA 0.1037125
                                                            lambda
(Intercept)                                                      0
Speeding                                                         0
factor(Cellphone.use.while.driving.for.Novice.or.Teen.Driver)1   0
> pool.r.squared(fit1)
         est     lo 95     hi 95         fmi
R^2 0.8379223 0.6395202 0.9322506 2.000442e-05
> pool.r.squared(fit1,adjusted = TRUE)
             est     lo 95     hi 95         fmi
adj R^2 0.8199136 0.6050623 0.9242601 2.000442e-05
>
```

Figure A.26. Road fatality prediction model for Nebraska

```
                            est        se        t       df    Pr(>|t|)       lo 95      hi 95 nmis       fmi lambda
(Intercept)            154.40509  78.91632  1.956567 15.29847 0.068906706  -13.515697  322.32588   NA 0.1092988      0
factor(BAC.Level.Law)1  37.88603  15.71563  2.410723 15.29847 0.028924973    4.445788   71.32627   NA 0.1092988      0
Drunk.Driving          796.39079 274.52436  2.900984 15.29847 0.010798827  212.248681 1380.53289    0 0.1092988      0
Speeding              -376.44247 120.08117 -3.134900 15.29847 0.006681395 -631.955195 -120.92974    0 0.1092988      0
> pool.r.squared(fit1)
          est     lo 95     hi 95         fmi
R^2 0.7518145 0.4841896 0.8930847 2.000442e-05
> pool.r.squared(fit1,adjusted = TRUE)
              est     lo 95     hi 95         fmi
adj R^2 0.7080171 0.4140738 0.8721669 2.000442e-05
>
```

Figure A.27. Road fatality prediction model for Nevada

```
                    est       se        t     df    Pr(>|t|)    lo 95    hi 95 nmis        fmi lambda
(Intercept)    72.72515 20.81433 3.493993 17.271 0.002725897 28.86317 116.5871   NA 0.09866312      0
Drunk.Driving 208.94920 77.80828 2.685437 17.271 0.015492324 44.98408 372.9143    0 0.09866312      0
> pool.r.squared(fit)
          est      lo 95     hi 95         fmi
R^2 0.2751293 0.01440193 0.6078936 2.000442e-05
> pool.r.squared(fit,adjusted = TRUE)
              est       lo 95     hi 95         fmi
adj R^2 0.2369782 0.004868301 0.5760731 2.000442e-05
>
```

Figure A.28. Road fatality prediction model for New Hampshire

```
                                          est          se        t      df     Pr(>|t|)        lo 95        hi 95 nmis      fmi     lambda
(Intercept)                        7.131970e+02 7.635603e+01  9.340415 14.37389 1.737419e-07  5.498283e+02  8.765657e+02   NA 0.1645184 0.055830042
Drunk.Driving                      7.403761e+02 2.803214e+02  2.641168 14.66048 1.878764e-02  1.416777e+02  1.339074e+03    0 0.1481610 0.039372692
Distracted.Driving                -2.729588e+02 8.869836e+01 -3.077383 15.27428 7.532635e-03 -4.617197e+02 -8.419802e+01    0 0.1109362 0.001676215
Road.Improvement.Capital.Expenditures -9.621729e-05 1.585782e-05 -6.067498 14.86970 2.236624e-05 -1.300433e-04 -6.239133e-05    1 0.1358604 0.026956092
> pool.r.squared(fit1)
         est    lo 95     hi 95       fmi
R^2 0.7776882 0.520939 0.9070481 0.04567408
> pool.r.squared(fit1,adjusted = TRUE)
             est    lo 95     hi 95       fmi
adj R^2 0.7384251 0.453753 0.8891776 0.04818141
>
```

Figure A.29. Road fatality prediction model for New Jersey

```
                    est           se         t       df     Pr(>|t|)        lo 95         hi 95 nmis       fmi lambda
(Intercept)   605.89612851 1.383714e+02  4.378766 15.29847 0.0005161816 311.46480366  9.003275e+02   NA 0.1092988      0
Precipitation   8.20379034 3.583537e+00  2.289299 15.29847 0.0366722094   0.57861967  1.582896e+01    0 0.1092988      0
Speeding      551.43167607 2.321650e+02  2.375171 15.29847 0.0310168038  57.42315488  1.045440e+03    0 0.1092988      0
Million.VMT    -0.01770168 4.710629e-03 -3.757816 15.29847 0.0018426277  -0.02772511 -7.678244e-03    0 0.1092988      0
> pool.r.squared(fit1)
          est    lo 95     hi 95          fmi
R^2 0.6154762 0.284641 0.8254152 2.000442e-05
> pool.r.squared(fit1,adjusted = TRUE)
             est     lo 95     hi 95          fmi
adj R^2 0.5476191 0.2052544 0.7885804 2.000442e-05
>
```

Figure A.30. Road fatality prediction model for New Mexico

```
                                          est          se        t      df     Pr(>|t|)        lo 95        hi 95 nmis      fmi     lambda
(Intercept)                        2.659489e+03 2.319737e+02 11.464616 15.06840 7.659625e-09  2.165244e+03  3.153734e+03   NA 0.1238317 0.01477687
Speeding                          -1.527346e+03 6.756378e+02 -2.260599 14.80062 3.930135e-02 -2.969125e+03 -8.556653e+01    0 0.1399600 0.03109847
Drunk.Driving                     -2.473308e+03 9.897285e+02 -2.498976 14.91803 2.462899e-02 -4.583874e+03 -3.627411e+02    0 0.1329685 0.02403143
Road.Improvement.Capital.Expenditures -1.325560e-04 4.934317e-05 -2.686410 14.64206 1.718848e-02 -2.379528e-04 -2.715914e-05    1 0.1492284 0.04044844
> pool.r.squared(fit1)
        est     lo 95     hi 95       fmi
R^2 0.816848 0.5892952 0.9252253 0.06674115
> pool.r.squared(fit1,adjusted = TRUE)
             est     lo 95     hi 95       fmi
adj R^2 0.7844987 0.5290415 0.9111103 0.06932811
>
```

Figure A.31. Road fatality prediction model for New York

```
                                                                  est      se        t      df     Pr(>|t|)     lo 95     hi 95 nmis      fmi lambda
(Intercept)                                                  733.6984 156.5935  4.685369 16.28409 0.0002372674  402.2051 1065.19171   NA 0.1037125      0
Drunk.Driving                                               3294.3157 647.2151  5.089986 16.28409 0.0001034898 1924.2237 4664.40779    0 0.1037125      0
factor(Cellphone.use.while.driving.for.Novice.or.Teen.Driver)1 -101.0008  31.4551 -3.210950 16.28409 0.0053532390 -167.5882  -34.41334   NA 0.1037125      0
> pool.r.squared(fit1)
         est     lo 95     hi 95         fmi
R^2 0.7346725 0.4560508 0.8849829 2.000442e-05
> pool.r.squared(fit1,adjusted = TRUE)
             est     lo 95     hi 95         fmi
adj R^2 0.7051916 0.4097497 0.8707924 2.000442e-05
>
```

Figure A.32. Road fatality prediction model for North Carolina

```
                      est          se        t       df    Pr(>|t|)        lo 95       hi 95 nmis       fmi    lambda
(Intercept)     -73.63813269 42.779122065 -1.721357 9.706162 0.116838795 -1.693485e+02 22.07228000   NA 0.4536720 0.3516133
Number.of.Nurses  0.01933219  0.004405119  4.388574 9.853629 0.001407795  9.497176e-03  0.02916721    3 0.4464037 0.3443924
> pool.r.squared(fit1)
         est     lo 95     hi 95       fmi
R^2 0.6131141 0.2252685 0.8467355 0.2710655
> pool.r.squared(fit1,adjusted = TRUE)
             est     lo 95     hi 95       fmi
adj R^2 0.5924567 0.1985332 0.8380189 0.2808675
>
```

Figure A.33. Road fatality prediction model for North Dakota

```
                                                          est          se        t       df    Pr(>|t|)        lo 95        hi 95 nmis
(Intercept)                                        1.899190e+03 1.076755e+02 17.638084 15.29847 1.396705e-11  1.670075e+03  2.128306e+03   NA
factor(Cellphone.use.while.driving.for.Novice.or.Teen.Driver)1 -2.264557e+02 3.615098e+01 -6.264164 15.29847 1.386390e-05 -3.033790e+02 -1.495324e+02   NA
Precipitation                                     -6.298372e+00 2.590197e+00 -2.431619 15.29847 2.775850e-02 -1.180988e+01 -7.868636e-01    0
Unemployed                                        -9.987174e-04 1.209818e-04 -8.255105 15.29847 5.074791e-07 -1.256147e-03 -7.412883e-04    0
                                                       fmi lambda
(Intercept)                                        0.1092988      0
factor(Cellphone.use.while.driving.for.Novice.or.Teen.Driver)1 0.1092988      0
Precipitation                                      0.1092988      0
Unemployed                                         0.1092988      0
> pool.r.squared(fit2)
         est     lo 95     hi 95         fmi
R^2 0.8888149 0.7427521 0.9543019 2.000442e-05
> pool.r.squared(fit2,adjusted = TRUE)
            est     lo 95     hi 95         fmi
adj R^2 0.869194 0.701914 0.9458906 2.000442e-05
>
```

Figure A.34. Road fatality prediction model for Ohio

```
                                        est          se        t       df   Pr(>|t|)        lo 95        hi 95 nmis       fmi      lambda
(Intercept)                      3.054009e+02 1.167174e+02  2.616585 14.24042 0.02009048 55.4629431341  5.553389e+02   NA 0.1206227 0.005222028
Precipitation                    4.464996e+00 1.648789e+00  2.708046 14.29733 0.01673845  0.9355820387  7.994410e+00    0 0.1167628 0.001286782
Drunk.Driving                    6.281469e+02 2.579439e+02  2.435208 14.26203 0.02857220 75.8644119274  1.180429e+03    0 0.1191623 0.003733596
Road.Improvement.Capital.Expenditures -1.678923e-04 5.681566e-05 -2.955036 12.50874 0.01156892 -0.0002911269 -4.465770e-05    1 0.2256480 0.111003368
Unemployed                      -7.044869e-04 3.252694e-04 -2.165857 12.45935 0.05037027 -0.0014103010  1.327256e-06    0 0.2284423 0.113792147
> pool.r.squared(fit)
         est     lo 95     hi 95       fmi
R^2 0.7679099 0.5044849 0.9024494 0.04195748
> pool.r.squared(fit,adjusted = TRUE)
           est     lo 95     hi 95       fmi
adj R^2 0.70984 0.4087809 0.8755738 0.04532621
>
```

Figure A.35. Road fatality prediction model for Oregon

```
                    est       se         t       df     Pr(>|t|)      lo 95     hi 95 nmis       fmi lambda
(Intercept)        1077.956 189.4894  5.688740 15.29847 3.984752e-05   674.7539 1481.1575   NA 0.1092988      0
Distracted.Driving 2013.005 454.0264  4.433675 15.29847 4.618768e-04  1046.9128 2979.0978    0 0.1092988      0
Speeding           -937.522 291.7356 -3.213602 15.29847 5.680207e-03 -1558.2867 -316.7573    0 0.1092988      0
Drunk.Driving      2148.202 584.0083  3.678375 15.29847 2.171472e-03   905.5293 3390.8742    0 0.1092988      0
> pool.r.squared(fit2)
          est     lo 95     hi 95          fmi
R^2 0.7948304 0.5588256 0.9129581 2.000442e-05
> pool.r.squared(fit2,adjusted = TRUE)
             est     lo 95     hi 95          fmi
adj R^2 0.758624 0.4956189 0.8962736 2.000442e-05
>
```

Figure A.36. Road fatality prediction model for Pennsylvania

```
                                          est          se        t      df     Pr(>|t|)        lo 95        hi 95 nmis      fmi     lambda
(Intercept)                      1.000618e+02 6.311369e+00 15.854214 14.78743 1.095961e-10  8.659257e+01  1.135310e+02   NA 0.1851168 0.08188490
Speeding                        -5.905777e+01 1.960767e+01 -3.011973 15.99337 8.273269e-03 -1.006256e+02 -1.748997e+01    0 0.1208867 0.01742126
Road.Improvement.Capital.Expenditures -8.648919e-05 3.338435e-05 -2.590711 14.24410 2.113101e-02 -1.579765e-04 -1.500183e-05    1 0.2120709 0.10869589
> pool.r.squared(fit)
         est     lo 95     hi 95       fmi
R^2 0.5590843 0.2099812 0.7988566 0.04524181
> pool.r.squared(fit,adjusted = TRUE)
            est     lo 95     hi 95       fmi
adj R^2 0.5099889 0.1587704 0.7718304 0.04963597
>
```

Figure A.37. Road fatality prediction model for Rhode Island

```
                     est          se        t      df     Pr(>|t|)        lo 95        hi 95 nmis      fmi lambda
(Intercept)    7.799015e+02 9.685335e+01  8.052395 16.28409 4.508760e-07 574.872283655  9.849307e+02   NA 0.1037125      0
Drunk.Driving  1.376740e+03 3.661685e+02  3.759855 16.28409 1.666257e-03 601.596806576  2.151883e+03    0 0.1037125      0
Unemployed    -1.829811e-03 3.580679e-04 -5.110235 16.28409 9.932742e-05  -0.002587806 -1.071816e-03    0 0.1037125      0
> pool.r.squared(fit2)
         est     lo 95    hi 95         fmi
R^2 0.6056792 0.2723821 0.820241 2.000442e-05
> pool.r.squared(fit2,adjusted = TRUE)
             est     lo 95     hi 95         fmi
adj R^2 0.5618658 0.2208578 0.7965141 2.000442e-05
>
```

Figure A.38. Road fatality prediction model for South Carolina

```
                    est         se        t      df    Pr(>|t|)       lo 95     hi 95 nmis      fmi     lambda
(Intercept)   -4660.32325 1258.99198 -3.701631 15.31297 0.002066573 -7339.03208 -1981.6144   NA 0.1970086 0.09856035
Female.Driver    96.39556   25.19456  3.826047 15.30661 0.001599114    42.78817   150.0029    1 0.1973024 0.09885184
> pool.r.squared(fit1)
          est     lo 95     hi 95       fmi
R^2 0.4602639 0.1140042 0.7429913 0.05375202
> pool.r.squared(fit1,adjusted = TRUE)
             est      lo 95     hi 95       fmi
adj R^2 0.4317528 0.09134552 0.7258715 0.05747872
>
```

Figure A.39. Road fatality prediction model for South Dakota

```
                                                          est          se        t      df     Pr(>|t|)       lo 95        hi 95 nmis
(Intercept)                                     1512.9618118 104.80623397 14.435800 10.01977 4.941301e-08  1279.50142  1.746422e+03   NA
Rural.Population                                  -0.1022746   0.04326014 -2.364176 10.59110 3.835666e-02    -0.19794 -6.609247e-03    9
factor(Cellphone.use.while.driving.for.Novice.or.Teen.Driver)1  -118.4470576  42.11073175 -2.812752 12.29842 1.535043e-02  -209.95215 -2.694197e+01   NA
Speeding                                       -1153.8007331 331.18484184 -3.483857 11.36948 4.875264e-03 -1879.85393 -4.277475e+02    0
                                                     fmi    lambda
(Intercept)                                    0.3920111 0.2816660
Rural.Population                               0.3619925 0.2519068
factor(Cellphone.use.while.driving.for.Novice.or.Teen.Driver)1 0.2742378 0.1650877
Speeding                                       0.3217759 0.2121150
> pool.r.squared(fit2)
         est     lo 95     hi 95       fmi
R^2 0.8040798 0.5604119 0.9207825 0.09584592
> pool.r.squared(fit2,adjusted = TRUE)
             est     lo 95     hi 95       fmi
adj R^2 0.7694579 0.4969581 0.9057832 0.09967597
>
```

Figure A.40. Road fatality prediction model for Tennessee

```
                      est        se         t     df     Pr(>|t|)      lo 95     hi 95 nmis        fmi lambda
(Intercept)        3869.877  166.9053 23.186071 17.271 1.865175e-14   3518.158  4221.596   NA 0.09866312      0
Distracted.Driving -7931.255 3261.0669 -2.432105 17.271 2.616198e-02 -14803.290 -1059.220    0 0.09866312      0
> pool.r.squared(fit1)
          est      lo 95     hi 95         fmi
R^2 0.2374112 0.004949572 0.5764481 2.000442e-05
> pool.r.squared(fit1,adjusted = TRUE)
             est        lo 95     hi 95         fmi
adj R^2 0.197275 0.0002377544 0.5401283 2.000442e-05
>
```

Figure A.41. Road fatality prediction model for Texas

```
                   est        se        t     df    Pr(>|t|)    lo 95    hi 95 nmis        fmi lambda
(Intercept)   174.430  45.36052 3.845414 17.271 0.001264455  78.8419  270.018   NA 0.09866312      0
Drunk.Driving 769.633 279.05676 2.757980 17.271 0.013301698 181.5777 1357.688    0 0.09866312      0
> pool.r.squared(fit)
          est      lo 95     hi 95         fmi
R^2 0.2858876 0.01792839 0.6164605 2.000442e-05
> pool.r.squared(fit,adjusted = TRUE)
             est      lo 95     hi 95         fmi
adj R^2 0.2483027 0.00720135 0.5857693 2.000442e-05
>
```

Figure A.42. Road fatality prediction model for Utah

```
                     est          se        t       df   Pr(>|t|)       lo 95        hi 95 nmis       fmi     lambda
(Intercept)   -464.7366474 244.0009812 -1.904651 13.16372 0.07890848 -991.203068  61.72977352   NA 0.2296283 0.12084733
Precipitation   -0.8445193   0.3777031 -2.235934 15.20767 0.04075521   -1.648618  -0.04042063    0 0.1151622 0.00597467
Speeding       -73.1575638  29.3944717 -2.488821 14.90088 0.02514215 -135.846715 -10.46841291    0 0.1339967 0.02507151
Male.Driver     11.8427035   4.7234408  2.507220 13.05822 0.02615912    1.642953  22.04245412    1 0.2351167 0.12630015
> pool.r.squared(fit1)
          est     lo 95     hi 95       fmi
R^2 0.7161218 0.4071364 0.8819583 0.1053593
> pool.r.squared(fit1,adjusted = TRUE)
             est    lo 95     hi 95       fmi
adj R^2 0.6658845 0.330554 0.8588227 0.1132264
>
```

Figure A.43. Road fatality prediction model for Vermont

```
                                        est          se        t       df    Pr(>|t|)        lo 95        hi 95 nmis       fmi    lambda
(Intercept)                       7.282803e+03 1.745229e+03  4.172978 13.242964 0.001052153  3.519485e+03  1.104612e+04   NA 0.2603193 0.1564532
Road.Improvement.Capital.Expenditures -3.040577e-04 1.252119e-04 -2.428345  4.365496 0.066801164 -6.405183e-04  3.240285e-05    1 0.7523981 0.6601040
Female.Driver                    -1.194943e+02 3.435133e+01 -3.478595 12.749289 0.004188858 -1.938545e+02 -4.513419e+01    1 0.2838218 0.1796451
> pool.r.squared(fit2)
          est     lo 95     hi 95       fmi
R^2 0.6290495 0.1685899 0.8803255 0.5095215
> pool.r.squared(fit2,adjusted = TRUE)
             est     lo 95     hi 95       fmi
adj R^2 0.5862833 0.1143816 0.8671951 0.5326611
>
```

Figure A.44. Road fatality prediction model for Virginia

```
                                          est          se        t       df     Pr(>|t|)        lo 95        hi 95 nmis      fmi lambda
(Intercept)                      7.192239e+02 32.096183965 22.408393 16.28409 1.145750e-13  6.512794e+02  7.871684e+02   NA 0.1037125      0
Unemployed                      -4.220415e-04  0.000170413 -2.476581 16.28409 2.459422e-02 -7.827893e-04 -6.129363e-05    0 0.1037125      0
factor(Handheld.phones.while.driving)1 -1.304705e+02 20.813820008 -6.268455 16.28409 1.033886e-05 -1.745313e+02 -8.640965e+01   NA 0.1037125      0
> pool.r.squared(fit1)
         est     lo 95     hi 95         fmi
R^2 0.8647028 0.6927509 0.9439496 2.000442e-05
> pool.r.squared(fit1,adjusted = TRUE)
            est     lo 95     hi 95         fmi
adj R^2 0.8496698 0.6625751 0.9374091 2.000442e-05
>
```

Figure A.45. Road fatality prediction model for Washington

```
                                          est          se        t       df     Pr(>|t|)        lo 95        hi 95 nmis      fmi lambda
(Intercept)                    481.971609428 29.823576253 16.160758 16.28409 1.882450e-11 418.837974691  5.451052e+02   NA 0.1037125      0
factor(Handheld.phones.while.driving)1 -58.117540714 16.582397358 -3.504773 16.28409 2.868194e-03 -93.220876711 -2.301420e+01   NA 0.1037125      0
Unemployed                      -0.001985674  0.000569257 -3.488185 16.28409 2.971261e-03  -0.003190736 -7.806118e-04    0 0.1037125      0
> pool.r.squared(fit1)
         est     lo 95     hi 95         fmi
R^2 0.6107751 0.2787247 0.8229381 2.000442e-05
> pool.r.squared(fit1,adjusted = TRUE)
            est     lo 95     hi 95         fmi
adj R^2 0.5675279 0.2272151 0.7996365 2.000442e-05
>
```

Figure A.46. Road fatality prediction model for West Virginia

```
                                                                  est          se        t       df     Pr(>|t|)        lo 95        hi 95 nmis
(Intercept)                                                1.542175e+04 4.675358e+03  3.298518 13.73196 0.0053992040  5.375715e+03  2.546779e+04   NA
factor(Cellphone.use.while.driving.for.Novice.or.Teen.Driver)1 -1.231512e+02 4.175846e+01 -2.949130 14.05614 0.0105268895 -2.126806e+02 -3.362170e+01   NA
Unemployed                                                -1.309550e-03 2.680951e-04 -4.884647 13.95394 0.0002433912 -1.884735e-03 -7.343651e-04    0
Male.Driver                                               -2.927490e+02 9.255330e+01 -3.163031 13.73621 0.0070509093 -4.916143e+02 -9.388368e+01    1
Drunk.Driving                                              6.528118e+02 2.401843e+02  2.717963 14.10941 0.0165680247  1.380423e+02  1.167581e+03    0
                                                                fmi     lambda
(Intercept)                                                0.1535151 0.03859697
factor(Cellphone.use.while.driving.for.Novice.or.Teen.Driver)1 0.1328479 0.01765878
Unemployed                                                 0.1394681 0.02437715
Male.Driver                                                0.1532503 0.03832928
Drunk.Driving                                              0.1293541 0.01410854
> pool.r.squared(fit4)
         est     lo 95     hi 95       fmi
R^2 0.8122165 0.5890501 0.9212005 0.01100344
> pool.r.squared(fit4,adjusted = TRUE)
            est     lo 95     hi 95       fmi
adj R^2 0.7652631 0.5049907 0.8998844 0.01171188
>
```

Figure A.47. Road fatality prediction model for Wisconsin

```
                          est        se        t       df     Pr(>|t|)     lo 95    hi 95 nmis       fmi lambda
(Intercept)          56.77185  21.34301 2.659974 16.28409 0.0169318253  11.59076 101.9530   NA 0.1037125      0
Drunk.Driving       286.06529  66.14078 4.325097 16.28409 0.0005033321 146.05164 426.0789    0 0.1037125      0
Distracted.Driving 345.00724 125.18660 2.755944 16.28409 0.0138934261  79.99929 610.0152    0 0.1037125      0
> pool.r.squared(fit)
        est     lo 95     hi 95          fmi
R^2 0.54446 0.2018703 0.7868059 2.000442e-05
> pool.r.squared(fit,adjusted = TRUE)
              est     lo 95     hi 95          fmi
adj R^2 0.4938444 0.1513851 0.7575724 2.000442e-05
>
```

Figure A.48. Road fatality prediction model for Wyoming

APPENDIX B

PREDICTED RESULTS FROM IMPUTED DATASETS

Year	Fatality	Predicted fatality1	Predicted fatality2	Predicted fatality3	Predicted fatality4	Predicted fatality5	Predicted fatality6	Predicted fatality7	Predicted fatality8	Predicted fatality9	Predicted fatality10	Avg. fatality (hat)
1994	1083	1106.11	1076.82	1046.85	1202.44	1076.65	1062.97	1123.36	1119.04	1150.57	1134.29	1109.91
1995	1114	1130.71	1089.35	941.38	1113.34	1179.81	1061.56	1115.28	1191.55	1100.35	1210.80	1113.41
1996	1146	1146.81	1101.86	1197.87	1070.37	1261.42	1138.01	1122.39	1158.76	1171.04	1286.15	1165.47
1997	1192	1186.75	1196.45	1239.86	1039.55	1236.34	1188.86	1110.37	1204.69	1170.73	1257.35	1183.09
1998	1071	1051.95	1037.07	1046.74	1068.74	1141.80	1026.13	1060.14	1091.58	1087.36	1113.09	1072.46
1999	1138	1029.52	1029.52	1029.52	1029.52	1029.52	1029.52	1029.52	1029.52	1029.52	1029.52	1029.52
2000	996	1059.12	1059.12	1059.12	1059.12	1059.12	1059.12	1059.12	1059.12	1059.12	1059.12	1059.12
2001	991	1054.22	1054.22	1054.22	1054.22	1054.22	1054.22	1054.22	1054.22	1054.22	1054.22	1054.22
2002	1038	1033.61	1033.61	1033.61	1033.61	1033.61	1033.61	1033.61	1033.61	1033.61	1033.61	1033.61
2003	1004	1025.15	1025.15	1025.15	1025.15	1025.15	1025.15	1025.15	1025.15	1025.15	1025.15	1025.15
2004	1154	1057.50	1057.50	1057.50	1057.50	1057.50	1057.50	1057.50	1057.50	1057.50	1057.50	1057.50
2005	1148	1077.00	1077.00	1077.00	1077.00	1077.00	1077.00	1077.00	1077.00	1077.00	1077.00	1077.00
2006	1207	1097.38	1097.38	1097.38	1097.38	1097.38	1097.38	1097.38	1097.38	1097.38	1097.38	1097.38
2007	1110	1112.65	1112.65	1112.65	1112.65	1112.65	1112.65	1112.65	1112.65	1112.65	1112.65	1112.65
2008	969	1095.29	1095.29	1095.29	1095.29	1095.29	1095.29	1095.29	1095.29	1095.29	1095.29	1095.29
2009	848	875.56	875.56	875.56	875.56	875.56	875.56	875.56	875.56	875.56	875.56	875.56
2010	862	842.58	842.58	842.58	842.58	842.58	842.58	842.58	842.58	842.58	842.58	842.58
2011	895	860.46	860.46	860.46	860.46	860.46	860.46	860.46	860.46	860.46	860.46	860.46
2012	865	854.23	854.23	854.23	854.23	854.23	854.23	854.23	854.23	854.23	854.23	854.23
2013	853	896.84	896.84	896.84	896.84	896.84	896.84	896.84	896.84	896.84	896.84	896.84
2014	820	923.85	923.85	923.85	923.85	923.85	923.85	923.85	923.85	923.85	923.85	923.85

Figure B.1. Alabama fatality prediction according to the prediction model

Correlation: Fatality, Avg. fatality (hat)

```
Pearson correlation of Fatality and Avg. fatality (hat) = 0.868
P-Value = 0.000
```

Figure B.2. Correlation result indicating predictive capability of the regression model

Year	Fatalities	Avg. Fatality(hat)
1994	85	86.25
1995	87	86.25
1996	81	86.25
1997	77	86.25
1998	70	86.25
1999	79	86.25
2000	106	86.25
2001	89	86.25
2002	89	86.25
2003	98	86.25
2004	101	86.25
2005	73	86.25
2006	74	65.89
2007	82	65.89
2008	62	65.89
2009	64	65.89
2010	56	65.89
2011	72	65.89
2012	59	65.89
2013	51	65.89
2014	73	65.89

Figure B.3. Alaska fatality prediction according to the prediction model

Correlation: Fatality, Avg. fatality (hat)

```
Pearson correlation of Fatality and Avg. fatality (hat) = 0.704
P-Value = 0.000
```

Figure B.4. Correlation result indicating predictive capability of the regression model of Alaska

Table B.1. Arkansas fatality prediction results from regression model

Year	Fatalities	Avg. fatality (hat)
1994	609	622.47
1995	631	601.07
1996	615	629.27
1997	660	589.37
1998	625	607.38
1999	604	611.72
2000	652	584.84
2001	611	571.19
2002	640	639.69
2003	640	649.38
2004	703	655.50
2005	654	634.23
2006	665	668.10
2007	649	640.52
2008	600	652.67
2009	596	577.59
2010	571	597.56
2011	551	580.93
2012	560	554.83
2013	498	538.50
2014	466	593.21

Correlation: Fatality, Avg. fatality (hat)

```
Pearson correlation of Fatality and Avg. fatality (hat) = 0.628
P-Value = 0.002
```

Figure B.5. Correlation result indicating predictive capability of the regression model of Arkansas

Table B.2. California fatality prediction results from regression model

Year	Fatalities	Avg. fatality(hat)
1994	4232	3762.41
1995	4192	3921.33
1996	3989	3742.64
1997	3688	3660.31
1998	3494	3819.27
1999	3559	3883.06
2000	3753	3880.22
2001	3956	3773.79
2002	4088	3742.22
2003	4224	4389.88
2004	4120	4224.44
2005	4333	4199.30
2006	4240	4081.53
2007	3995	3967.64
2008	3434	3850.97
2009	3090	3003.07
2010	2720	2666.23
2011	2816	2807.20
2012	2966	3049.78
2013	3107	3261.48
2014	3074	3406.24

Correlation: Fatality, Avg. fatality (hat)

```
Pearson correlation of Fatality and Avg. fatality (hat) = 0.891
P-Value = 0.000
```

Figure B.6. Correlation result indicating predictive capability of the regression model of California

Table B.3. Colorado fatality prediction results from regression model

Year	Fatalities	Avg. fatality(hat)
1994	586	596.88
1995	645	566.57
1996	617	578.50
1997	613	575.04
1998	628	632.12
1999	626	685.42
2000	681	695.49
2001	741	709.31
2002	743	734.61
2003	642	618.00
2004	667	644.82
2005	606	574.17
2006	535	607.17
2007	554	559.22
2008	548	528.98
2009	465	551.98
2010	450	479.75
2011	447	471.26
2012	474	499.19
2013	482	479.01
2014	488	450.30

Correlation: Fatality, Avg. fatality (hat)

```
Pearson correlation of Fatality and Avg. fatality (hat) = 0.896
P-Value = 0.000
```

Figure B.7. Correlation result indicating predictive capability of the regression model of Colorado

Table B.4. Connecticut fatality prediction results from regression model

Year	Fatalities	Avg. fatality (hat)
1994	310	312.32
1995	317	303.42
1996	310	303.72
1997	339	321.88
1998	329	317.40
1999	301	321.32
2000	341	309.15
2001	318	313.39
2002	325	329.28
2003	298	308.47
2004	294	279.13
2005	278	280.45
2006	311	261.51
2007	296	279.55
2008	302	285.04
2009	224	280.80
2010	320	296.65
2011	221	263.30
2012	264	288.43
2013	286	274.43
2014	248	302.18

Correlation: Fatality, Avg. fatality (hat)

```
Pearson correlation of Fatality and Avg. fatality (hat) = 0.590
P-Value = 0.005
```

Figure B.8. Correlation result indicating predictive capability of the regression model of Connecticut

Year	Fatality	Predicted fatality1	Predicted fatality2	Predicted fatality3	Predicted fatality4	Predicted fatality5	Predicted fatality6	Predicted fatality7	Predicted fatality8	Predicted fatality9	Predicted fatality10	Avg. fatality (hat)
1994	112	122.56	122.56	122.56	122.56	122.56	122.56	122.56	122.56	122.56	122.56	122.56
1995	121	125.46	125.46	125.46	125.46	125.46	125.46	125.46	125.46	125.46	125.46	125.46
1996	116	126.81	126.81	126.81	126.81	126.81	126.81	126.81	126.81	126.81	126.81	126.81
1997	143	126.83	126.83	126.83	126.83	126.83	126.83	126.83	126.83	126.83	126.83	126.83
1998	115	122.14	122.14	122.14	122.14	122.14	122.14	122.14	122.14	122.14	122.14	122.14
1999	100	127.82	127.82	127.82	127.82	127.82	127.82	127.82	127.82	127.82	127.82	127.82
2000	123	121.91	121.91	121.91	121.91	121.91	121.91	121.91	121.91	121.91	121.91	121.91
2001	136	122.05	122.05	122.05	122.05	122.05	122.05	122.05	122.05	122.05	122.05	122.05
2002	124	123.40	123.40	123.40	123.40	123.40	123.40	123.40	123.40	123.40	123.40	123.40
2003	142	125.70	125.70	125.70	125.70	125.70	125.70	125.70	125.70	125.70	125.70	125.70
2004	134	124.85	124.85	124.85	124.85	124.85	124.85	124.85	124.85	124.85	124.85	124.85
2005	133	122.96	122.96	122.96	122.96	122.96	122.96	122.96	122.96	122.96	122.96	122.96
2006	148	128.32	128.32	128.32	128.32	128.32	128.32	128.32	128.32	128.32	128.32	128.32
2007	117	123.31	123.31	123.31	123.31	123.31	123.31	123.31	123.31	123.31	123.31	123.31
2008	121	120.47	120.47	120.47	120.47	120.47	120.47	120.47	120.47	120.47	120.47	120.47
2009	116	113.42	113.42	113.42	113.42	113.42	113.42	113.42	113.42	113.42	113.42	113.42
2010	101	109.14	109.14	109.14	109.14	109.14	109.14	109.14	109.14	109.14	109.14	109.14
2011	99	127.63	128.37	111.91	123.95	116.98	113.98	117.44	123.04	120.24	120.34	120.39
2012	114	102.58	102.58	102.58	102.58	102.58	102.58	102.58	102.58	102.58	102.58	102.58
2013	99	106.94	106.94	106.94	106.94	106.94	106.94	106.94	106.94	106.94	106.94	106.94
2014	121	118.56	118.56	118.56	118.56	118.56	118.56	118.56	118.56	118.56	118.56	118.56

Figure B.9. Delaware fatality prediction according to the prediction model

Correlation: Fatality, Avg. fatality (hat)

```
Pearson correlation of Fatality and Avg. fatality (hat) = 0.511
P-Value = 0.018
```

Figure B.10. Correlation result indicating predictive capability of the regression model of Delaware

Table B.5. Florida fatality prediction results from regression model

Year	Fatalities	Avg. fatality (hat)
1994	2687	3060.41
1995	2805	3050.82
1996	2753	2985.49
1997	2785	2946.62
1998	2825	2967.75
1999	2920	3036.76
2000	2999	3052.51
2001	3012	2970.87
2002	3136	2992.35
2003	3169	3005.59
2004	3244	3059.90
2005	3518	3600.84
2006	3357	3446.94
2007	3213	3326.98
2008	2980	3106.69
2009	2560	2902.21
2010	2444	2481.36
2011	2400	2779.40
2012	2431	2844.79
2013	2403	2516.85
2014	2494	2654.66

Correlation: Fatality, Avg. fatality (hat)

```
Pearson correlation of Fatality and Avg. fatality (hat) = 0.863
P-Value = 0.000
```

Figure B.11. Correlation result indicating predictive capability of the regression model of Florida

Table B.6. Georgia fatality prediction results from regression model

Year	Fatalities	Avg. y(hat)
1994	1425	1544.63
1995	1488	1561.25
1996	1573	1515.21
1997	1577	1544.28
1998	1568	1553.99
1999	1508	1511.75
2000	1541	1588.36
2001	1647	1558.36
2002	1524	1596.30
2003	1603	1412.84
2004	1634	1598.53
2005	1729	1624.71
2006	1693	1574.62
2007	1641	1527.23
2008	1495	1523.16
2009	1292	1498.83
2010	1247	1169.36
2011	1226	1282.08
2012	1192	1425.51
2013	1180	1170.75
2014	1164	1165.23

Correlation: Fatality, Avg. fatality (hat)

```
Pearson correlation of Fatality and Avg. fatality (hat) = 0.815
P-Value = 0.000
```

Figure B.12. Correlation result indicating predictive capability of the regression model of Georgia

Table B.7. Hawaii fatality prediction results from regression model

Year	Fatalities	Avg. fatality(hat)
1994	122	130.21
1995	130	133.77
1996	148	128.33
1997	131	127.53
1998	120	113.14
1999	98	111.25
2000	132	125.52
2001	140	124.75
2002	119	131.49
2003	133	137.22
2004	142	144.56
2005	140	142.92
2006	161	138.24
2007	138	139.46
2008	107	126.49
2009	109	105.11
2010	113	111.08
2011	100	111.18
2012	125	113.92
2013	102	95.71
2014	95	100.05

Correlation: Fatality, Avg. fatality (hat)

```
Pearson correlation of Fatality and Avg. fatality (hat) = 0.794
P-Value = 0.000
```

Figure B.13. Correlation result indicating predictive capability of the regression model of Hawaii

Table B.8. Idaho fatality prediction results from regression model

Year	Fatalities	Avg. fatality(hat)
1994	250	263.31
1995	262	262.07
1996	258	261.81
1997	259	261.94
1998	265	261.28
1999	278	262.93
2000	276	264.74
2001	259	258.90
2002	264	253.43
2003	293	252.30
2004	260	258.91
2005	275	266.93
2006	267	272.91
2007	252	276.87
2008	232	251.95
2009	226	208.17
2010	209	206.08
2011	167	213.63
2012	184	181.39
2013	214	194.60
2014	186	209.27

Correlation: Fatality, Avg. fatality (hat)

```
Pearson correlation of Fatality and Avg. fatality (hat) = 0.843
P-Value = 0.000
```

Figure B.14. Correlation result indicating predictive capability of the regression model of Idaho

Year	Fatalities	Predicted fatality1	Predicted fatality2	Predicted fatality3	Predicted fatality4	Predicted fatality5	Predicted fatality6	Predicted fatality7	Predicted fatality8	Predicted fatality9	Predicted fatality10	Avg. fatality(hat)
1994	1554	1667.89	1601.08	1475.18	1576.51	1514.54	1527.36	1529.81	1559.14	1576.08	1554.80	1558.24
1995	1586	1529.00	1486.56	1451.92	1491.18	1424.87	1509.80	1441.39	1414.54	1515.75	1518.40	1478.34
1996	1477	1392.10	1436.32	1480.29	1453.10	1437.60	1444.29	1427.32	1476.00	1449.15	1487.60	1448.38
1997	1397	1428.92	1428.92	1428.92	1428.92	1428.92	1428.92	1428.92	1428.92	1428.92	1428.92	1428.92
1998	1393	1437.04	1437.04	1437.04	1437.04	1437.04	1437.04	1437.04	1437.04	1437.04	1437.04	1437.04
1999	1456	1473.89	1473.89	1473.89	1473.89	1473.89	1473.89	1473.89	1473.89	1473.89	1473.89	1473.89
2000	1418	1460.02	1460.02	1460.02	1460.02	1460.02	1460.02	1460.02	1460.02	1460.02	1460.02	1460.02
2001	1414	1418.21	1418.21	1418.21	1418.21	1418.21	1418.21	1418.21	1418.21	1418.21	1418.21	1418.21
2002	1420	1388.09	1388.09	1388.09	1388.09	1388.09	1388.09	1388.09	1388.09	1388.09	1388.09	1388.09
2003	1454	1431.71	1431.71	1431.71	1431.71	1431.71	1431.71	1431.71	1431.71	1431.71	1431.71	1431.71
2004	1355	1416.93	1416.93	1416.93	1416.93	1416.93	1416.93	1416.93	1416.93	1416.93	1416.93	1416.93
2005	1363	1329.48	1329.48	1329.48	1329.48	1329.48	1329.48	1329.48	1329.48	1329.48	1329.48	1329.48
2006	1254	1342.61	1342.61	1342.61	1342.61	1342.61	1342.61	1342.61	1342.61	1342.61	1342.61	1342.61
2007	1248	1220.02	1220.02	1220.02	1220.02	1220.02	1220.02	1220.02	1220.02	1220.02	1220.02	1220.02
2008	1043	1045.08	1045.08	1045.08	1045.08	1045.08	1045.08	1045.08	1045.08	1045.08	1045.08	1045.08
2009	911	899.56	899.56	899.56	899.56	899.56	899.56	899.56	899.56	899.56	899.56	899.56
2010	927	882.43	882.43	882.43	882.43	882.43	882.43	882.43	882.43	882.43	882.43	882.43
2011	918	926.03	926.03	926.03	926.03	926.03	926.03	926.03	926.03	926.03	926.03	926.03
2012	956	1021.25	1021.25	1021.25	1021.25	1021.25	1021.25	1021.25	1021.25	1021.25	1021.25	1021.25
2013	991	1005.87	1005.87	1005.87	1005.87	1005.87	1005.87	1005.87	1005.87	1005.87	1005.87	1005.87
2014	924	986.74	986.74	986.74	986.74	986.74	986.74	986.74	986.74	986.74	986.74	986.74

Figure B.15. Illinois fatality prediction according to the prediction model

Correlation: Fatality, Avg. fatality (hat)

```
Pearson correlation of Fatality and Avg. fatality (hat) = 0.977
P-Value = 0.000
```

Figure B.16. Correlation result indicating predictive capability of the regression model of Illinois

Table B.9. Indiana fatality prediction results from regression model

Year	Fatalities	Avg. fatality (hat)
1994	971	947.08
1995	960	947.97
1996	984	952.32
1997	935	971.89
1998	982	1001.22
1999	1020	966.12
2000	886	951.88
2001	909	862.96
2002	792	832.50
2003	833	846.10
2004	947	880.54
2005	938	903.48
2006	902	927.62
2007	898	897.58
2008	820	850.10
2009	693	750.17
2010	754	706.73
2011	751	749.28
2012	781	772.43
2013	784	756.06
2014	746	814.78

Correlation: Fatality, Avg. fatality (hat)

```
Pearson correlation of Fatality and Avg. fatality (hat) = 0.906
P-Value = 0.000
```

Figure B.17. Correlation result indicating predictive capability of the regression model of Indiana

Year	Fatalities	Predicted fatality1	Predicted fatality2	Predicted fatality3	Predicted fatality4	Predicted fatality5	Predicted fatality6	Predicted fatality7	Predicted fatality8	Predicted fatality9	Predicted fatality10	Avg. fatality(hat)
1994	478	468.81	468.81	468.81	468.81	468.81	468.81	468.81	468.81	468.81	468.81	468.81
1995	527	465.32	465.32	465.32	465.32	465.32	465.32	465.32	465.32	465.32	465.32	465.32
1996	465	463.33	463.33	463.33	463.33	463.33	463.33	463.33	463.33	463.33	463.33	463.33
1997	468	458.94	458.94	458.94	458.94	458.94	458.94	458.94	458.94	458.94	458.94	458.94
1998	449	456.15	456.15	456.15	456.15	456.15	456.15	456.15	456.15	456.15	456.15	456.15
1999	490	449.01	449.01	449.01	449.01	449.01	449.01	449.01	449.01	449.01	449.01	449.01
2000	445	393.19	393.19	393.19	393.19	393.19	393.19	393.19	393.19	393.19	393.19	393.19
2001	446	418.93	418.93	418.93	418.93	418.93	418.93	418.93	418.93	418.93	418.93	418.93
2002	405	418.66	418.66	418.66	418.66	418.66	418.66	418.66	418.66	418.66	418.66	418.66
2003	443	414.70	414.70	414.70	414.70	414.70	414.70	414.70	414.70	414.70	414.70	414.70
2004	388	447.82	447.82	447.82	447.82	447.82	447.82	447.82	447.82	447.82	447.82	447.82
2005	450	460.57	460.57	460.57	460.57	460.57	460.57	460.57	460.57	460.57	460.57	460.57
2006	439	436.88	436.88	436.88	436.88	436.88	436.88	436.88	436.88	436.88	436.88	436.88
2007	446	427.82	427.82	427.82	427.82	427.82	427.82	427.82	427.82	427.82	427.82	427.82
2008	412	440.09	440.09	440.09	440.09	440.09	440.09	440.09	440.09	440.09	440.09	440.09
2009	371	401.41	401.41	401.41	401.41	401.41	401.41	401.41	401.41	401.41	401.41	401.41
2010	390	373.64	373.64	373.64	373.64	373.64	373.64	373.64	373.64	373.64	373.64	373.64
2011	360	439.86	430.90	492.68	397.66	409.27	409.43	474.31	340.68	494.30	449.03	433.81
2012	365	338.99	338.99	338.99	338.99	338.99	338.99	338.99	338.99	338.99	338.99	338.99
2013	317	365.01	365.01	365.01	365.01	365.01	365.01	365.01	365.01	365.01	365.01	365.01
2014	321	346.18	346.18	346.18	346.18	346.18	346.18	346.18	346.18	346.18	346.18	346.18

Figure B.18. Iowa fatality prediction according to the prediction model

Correlation: Fatality, Avg. fatality (hat)

```
Pearson correlation of Fatality and Avg. fatality (hat) = 0.753
P-Value = 0.000
```

Figure B.19. Correlation result indicating predictive capability of the regression model of Iowa

Year	Fatalities	Predicted fatality1	Predicted fatality2	Predicted fatality3	Predicted fatality4	Predicted fatality5	Predicted fatality6	Predicted fatality7	Predicted fatality8	Predicted fatality9	Predicted fatality10	Avg. fatality(hat)
1994	442	523.05	503.37	459.20	500.48	479.14	483.06	527.61	462.90	504.74	494.85	493.84
1995	442	511.52	428.97	474.65	478.58	453.67	469.54	504.82	450.51	474.47	463.78	471.05
1996	490	475.90	469.37	453.02	441.10	462.57	398.17	481.73	470.52	464.44	424.79	454.16
1997	482	442.52	442.52	442.52	442.52	442.52	442.52	442.52	442.52	442.52	442.52	442.52
1998	492	459.84	459.84	459.84	459.84	459.84	459.84	459.84	459.84	459.84	459.84	459.84
1999	540	443.31	443.31	443.31	443.31	443.31	443.31	443.31	443.31	443.31	443.31	443.31
2000	461	473.22	473.22	473.22	473.22	473.22	473.22	473.22	473.22	473.22	473.22	473.22
2001	494	426.00	426.00	426.00	426.00	426.00	426.00	426.00	426.00	426.00	426.00	426.00
2002	507	466.13	466.13	466.13	466.13	466.13	466.13	466.13	466.13	466.13	466.13	466.13
2003	469	462.99	462.99	462.99	462.99	462.99	462.99	462.99	462.99	462.99	462.99	462.99
2004	459	438.59	438.59	438.59	438.59	438.59	438.59	438.59	438.59	438.59	438.59	438.59
2005	428	451.97	451.97	451.97	451.97	451.97	451.97	451.97	451.97	451.97	451.97	451.97
2006	468	484.23	484.23	484.23	484.23	484.23	484.23	484.23	484.23	484.23	484.23	484.23
2007	416	446.46	446.46	446.46	446.46	446.46	446.46	446.46	446.46	446.46	446.46	446.46
2008	384	433.87	433.87	433.87	433.87	433.87	433.87	433.87	433.87	433.87	433.87	433.87
2009	386	403.17	403.17	403.17	403.17	403.17	403.17	403.17	403.17	403.17	403.17	403.17
2010	431	422.06	422.06	422.06	422.06	422.06	422.06	422.06	422.06	422.06	422.06	422.06
2011	386	415.77	415.77	415.77	415.77	415.77	415.77	415.77	415.77	415.77	415.77	415.77
2012	405	425.21	425.21	425.21	425.21	425.21	425.21	425.21	425.21	425.21	425.21	425.21
2013	350	386.65	386.65	386.65	386.65	386.65	386.65	386.65	386.65	386.65	386.65	386.65
2014	385	426.00	426.00	426.00	426.00	426.00	426.00	426.00	426.00	426.00	426.00	426.00

Figure B.20. Kansas fatality prediction according to the prediction model

Correlation: Fatality, Avg. fatality (hat)

```
Pearson correlation of Fatality and Avg. fatality (hat) = 0.575
P-Value = 0.006
```

Figure B.21. Correlation result indicating predictive capability of the regression model of Kansas

Table B.10. Kentucky fatality prediction results from regression model

Year	Fatalities	Avg. fatality(hat)
1994	778	879.09
1995	849	856.95
1996	842	864.68
1997	857	850.37
1998	858	854.76
1999	814	863.76
2000	820	849.78
2001	845	815.84
2002	915	838.15
2003	928	873.90
2004	964	939.31
2005	985	904.50
2006	913	866.84
2007	864	884.71
2008	825	864.36
2009	791	820.06
2010	760	780.94
2011	720	693.30
2012	746	667.80
2013	638	714.17
2014	672	700.77

Correlation: Fatality, Avg. fatality (hat)

```
Pearson correlation of Fatality and Avg. fatality (hat) = 0.829
P-Value = 0.000
```

Figure B.22. Correlation result indicating predictive capability of the regression model of Kentucky

Year	Fatalities	Predicted fatality1	Predicted fatality2	Predicted fatality3	Predicted fatality4	Predicted fatality5	Predicted fatality6	Predicted fatality7	Predicted fatality8	Predicted fatality9	Predicted fatality10	Avg. fatality(hat)
1994	843	869.88	869.88	869.88	869.88	869.88	869.88	869.88	869.88	869.88	869.88	869.88
1995	894	890.74	890.74	890.74	890.74	890.74	890.74	890.74	890.74	890.74	890.74	890.74
1996	902	896.88	896.88	896.88	896.88	896.88	896.88	896.88	896.88	896.88	896.88	896.88
1997	931	907.08	907.08	907.08	907.08	907.08	907.08	907.08	907.08	907.08	907.08	907.08
1998	926	929.55	929.55	929.55	929.55	929.55	929.55	929.55	929.55	929.55	929.55	929.55
1999	938	955.62	955.62	955.62	955.62	955.62	955.62	955.62	955.62	955.62	955.62	955.62
2000	938	953.80	953.80	953.80	953.80	953.80	953.80	953.80	953.80	953.80	953.80	953.80
2001	952	920.66	920.66	920.66	920.66	920.66	920.66	920.66	920.66	920.66	920.66	920.66
2002	907	917.17	917.17	917.17	917.17	917.17	917.17	917.17	917.17	917.17	917.17	917.17
2003	940	922.06	922.06	922.06	922.06	922.06	922.06	922.06	922.06	922.06	922.06	922.06
2004	927	942.44	942.44	942.44	942.44	942.44	942.44	942.44	942.44	942.44	942.44	942.44
2005	963	900.34	900.34	900.34	900.34	900.34	900.34	900.34	900.34	900.34	900.34	900.34
2006	987	1016.92	1016.92	1016.92	1016.92	1016.92	1016.92	1016.92	1016.92	1016.92	1016.92	1016.92
2007	993	1032.01	1032.01	1032.01	1032.01	1032.01	1032.01	1032.01	1032.01	1032.01	1032.01	1032.01
2008	916	842.99	842.99	842.99	842.99	842.99	842.99	842.99	842.99	842.99	842.99	842.99
2009	824	811.01	811.01	811.01	811.01	811.01	811.01	811.01	811.01	811.01	811.01	811.01
2010	721	736.26	736.26	736.26	736.26	736.26	736.26	736.26	736.26	736.26	736.26	736.26
2011	680	698.09	693.38	705.25	729.74	738.92	709.54	701.28	722.57	704.29	700.16	710.32
2012	723	747.37	747.37	747.37	747.37	747.37	747.37	747.37	747.37	747.37	747.37	747.37
2013	703	738.38	738.38	738.38	738.38	738.38	738.38	738.38	738.38	738.38	738.38	738.38
2014	737	726.29	726.29	726.29	726.29	726.29	726.29	726.29	726.29	726.29	726.29	726.29

Figure B.23. Louisiana fatality prediction according to the prediction model

Correlation: Fatality, Avg. fatality (hat)

```
Pearson correlation of Fatality and Avg. fatality (hat) = 0.953
P-Value = 0.000
```

Figure B.24. Correlation result indicating predictive capability of the regression model of Louisiana

Table B.11. Maine fatality prediction results from regression model

Year	Fatalities	Avg. fatality(hat)
1994	188	180.94
1995	187	186.13
1996	169	171.94
1997	192	187.79
1998	192	183.63
1999	181	168.01
2000	169	191.72
2001	192	192.50
2002	216	200.03
2003	207	199.49
2004	194	191.86
2005	169	179.15
2006	188	165.29
2007	183	161.88
2008	155	177.40
2009	159	171.52
2010	161	180.02
2011	136	151.45
2012	164	146.31
2013	144	149.65
2014	131	140.16

Correlation: Fatality, Avg. fatality (hat)

```
Pearson correlation of Fatality and Avg. fatality (hat) = 0.781
P-Value = 0.000
```

Figure B.25. Correlation result indicating predictive capability of the regression model of Maine

Table: B.12. Maryland fatality prediction results from regression model

Year	Fatalities	Predicted fatality
1994	651	622.38
1995	671	622.38
1996	608	622.38
1997	611	622.38
1998	606	622.38
1999	590	622.38
2000	588	622.38
2001	659	622.38
2002	661	622.38
2003	650	622.38
2004	643	622.38
2005	614	479.80
2006	652	479.80
2007	614	479.80
2008	591	479.80
2009	549	479.80
2010	496	479.80
2011	485	479.80
2012	511	479.80
2013	465	479.80
2014	442	479.80

Correlation: Fatality, Avg. fatality (hat)

```
Pearson correlation of Fatality and Avg. fatality (hat) = 0.650
P-Value = 0.001
```

Figure B.26. Correlation result indicating predictive capability of the regression model of Maryland

Table B.13. Massachusetts fatality prediction results from regression model

Year	Fatalities	Predicted fatality
1994	440	429.81
1995	444	429.81
1996	417	429.81
1997	441	429.81
1998	406	429.81
1999	414	429.81
2000	433	429.81
2001	477	429.81
2002	459	429.81
2003	462	429.81
2004	476	429.81
2005	441	429.81
2006	429	429.81
2007	434	429.81
2008	364	429.81
2009	340	429.81
2010	347	356.60
2011	374	356.60
2012	383	356.60
2013	351	356.60
2014	328	356.60

Correlation: Fatality, Avg. fatality (hat)

```
Pearson correlation of Fatality and Avg. fatality (hat) = 0.693
P-Value = 0.001
```

Figure B.27. Correlation result indicating predictive capability of the regression model of Massachusetts

Table B.14. Michigan fatality prediction results from regression model

Year	Fatalities	Predicted fatality
1994	1421	1331.53
1995	1530	1364.38
1996	1505	1383.44
1997	1446	1393.03
1998	1366	1421.89
1999	1382	1426.14
2000	1382	1404.17
2001	1328	1310.19
2002	1277	1288.36
2003	1283	1241.46
2004	1159	1246.72
2005	1129	1256.60
2006	1086	1059.55
2007	1087	1063.15
2008	980	1052.75
2009	872	770.37
2010	942	854.07
2011	889	1043.00
2012	940	1096.86
2013	947	880.84
2014	901	967.59

Correlation: Fatality, Avg. fatality (hat)

```
Pearson correlation of Fatality and Avg. fatality (hat) = 0.913
P-Value = 0.000
```

Figure B.28. Correlation result indicating predictive capability of the regression model of Michigan

Table B.15. Minnesota fatality prediction results from regression model

Year	Fatalities	Predicted fatality
1994	646	588.69
1995	597	605.92
1996	576	532.27
1997	600	476.09
1998	650	571.81
1999	626	490.29
2000	625	591.84
2001	568	538.13
2002	657	544.79
2003	655	570.04
2004	567	528.65
2005	559	578.32
2006	494	595.15
2007	510	637.40
2008	455	546.49
2009	421	439.59
2010	411	458.46
2011	368	501.80
2012	395	500.92
2013	387	414.26
2014	361	417.26

Correlation: Fatality, Avg. fatality (hat)

```
Pearson correlation of Fatality and Avg. fatality (hat) = 0.605
P-Value = 0.004
```

Figure B.29. Correlation result indicating predictive capability of the regression model of Minnesota

Table B.16. Mississippi fatality prediction results from regression model

Year	Fatalities	Predicted Fatality
1994	791	806.78
1995	868	818.36
1996	811	849.88
1997	861	837.48
1998	948	879.44
1999	927	847.93
2000	949	869.29
2001	784	912.56
2002	885	913.55
2003	872	941.39
2004	900	905.96
2005	931	863.18
2006	911	830.04
2007	884	817.56
2008	783	833.57
2009	700	773.61
2010	641	745.38
2011	630	603.65
2012	582	609.70
2013	613	607.39
2014	607	611.30

Correlation: Fatality, Avg. fatality (hat)

```
Pearson correlation of Fatality and Avg. fatality (hat) = 0.862
P-Value = 0.000
```

Figure B.30. Correlation result indicating predictive capability of the regression model of Mississippi

Year	Fatalities	Predicted fatality1	Predicted fatality2	Predicted fatality3	Predicted fatality4	Predicted fatality5	Predicted fatality6	Predicted fatality7	Predicted fatality8	Predicted fatality9	Predicted fatality10	Avg. Predicted fatality
1994	1089	1587.84	1574.47	1610.89	1588.68	1450.90	1482.99	1517.32	1568.64	1537.36	1646.45	1556.55
1995	1109	1459.95	1428.75	1434.08	1430.64	1472.90	1436.69	1413.55	1430.65	1405.56	1479.23	1439.20
1996	1148	1467.09	1433.57	1422.73	1406.35	1401.48	1374.06	1421.05	1378.10	1403.15	1449.07	1415.67
1997	1192	1402.42	1402.42	1402.42	1402.42	1402.42	1402.42	1402.42	1402.42	1402.42	1402.42	1402.42
1998	1169	1416.06	1416.06	1416.06	1416.06	1416.06	1416.06	1416.06	1416.06	1416.06	1416.06	1416.06
1999	1094	1383.00	1383.00	1383.00	1383.00	1383.00	1383.00	1383.00	1383.00	1383.00	1383.00	1383.00
2000	1157	1234.51	1234.51	1234.51	1234.51	1234.51	1234.51	1234.51	1234.51	1234.51	1234.51	1234.51
2001	1098	1358.67	1358.67	1358.67	1358.67	1358.67	1358.67	1358.67	1358.67	1358.67	1358.67	1358.67
2002	1208	1353.46	1353.46	1353.46	1353.46	1353.46	1353.46	1353.46	1353.46	1353.46	1353.46	1353.46
2003	1232	1362.52	1362.52	1362.52	1362.52	1362.52	1362.52	1362.52	1362.52	1362.52	1362.52	1362.52
2004	1130	1364.78	1364.78	1364.78	1364.78	1364.78	1364.78	1364.78	1364.78	1364.78	1364.78	1364.78
2005	1257	1621.65	1621.65	1621.65	1621.65	1621.65	1621.65	1621.65	1621.65	1621.65	1621.65	1621.65
2006	1096	1345.57	1345.57	1345.57	1345.57	1345.57	1345.57	1345.57	1345.57	1345.57	1345.57	1345.57
2007	992	1345.62	1345.62	1345.62	1345.62	1345.62	1345.62	1345.62	1345.62	1345.62	1345.62	1345.62
2008	960	1321.76	1321.76	1321.76	1321.76	1321.76	1321.76	1321.76	1321.76	1321.76	1321.76	1321.76
2009	878	1327.67	1327.67	1327.67	1327.67	1327.67	1327.67	1327.67	1327.67	1327.67	1327.67	1327.67
2010	821	1307.96	1307.96	1307.96	1307.96	1307.96	1307.96	1307.96	1307.96	1307.96	1307.96	1307.96
2011	786	1308.01	1308.01	1308.01	1308.01	1308.01	1308.01	1308.01	1308.01	1308.01	1308.01	1308.01
2012	826	1271.82	1271.82	1271.82	1271.82	1271.82	1271.82	1271.82	1271.82	1271.82	1271.82	1271.82
2013	757	1269.59	1269.59	1269.59	1269.59	1269.59	1269.59	1269.59	1269.59	1269.59	1269.59	1269.59
2014	766	1262.95	1262.95	1262.95	1262.95	1262.95	1262.95	1262.95	1262.95	1262.95	1262.95	1262.95

Figure B.31. Missouri fatality prediction according to the prediction model

Correlation: Fatality, Avg. fatality (hat)

```
Pearson correlation of Fatality and Avg. fatality (hat) = 0.584
P-Value = 0.005
```

Figure B.32. Correlation result indicating predictive capability of the regression model of Missouri

Table B.17. Nebraska fatality prediction results from regression model

Year	Fatalities	Predicted fatality
1994	271	272.51
1995	254	291.34
1996	293	277.17
1997	302	281.34
1998	315	307.80
1999	295	288.65
2000	276	260.23
2001	246	249.67
2002	307	308.06
2003	293	291.78
2004	254	272.76
2005	276	272.76
2006	269	276.93
2007	256	228.62
2008	208	225.68
2009	223	224.05
2010	190	204.78
2011	181	203.87
2012	212	208.23
2013	211	208.46
2014	225	202.40

Correlation: Fatality, Avg. fatality (hat)

```
Pearson correlation of Fatality and Avg. fatality (hat) = 0.915
P-Value = 0.000
```

Figure B.33. Correlation result indicating predictive capability of the regression model of Nebraska

Table B.18. Nevada fatality prediction results from regression model

Year	Fatalities	Predicted fatality
1994	294	361.13
1995	313	308.16
1996	348	321.76
1997	347	335.82
1998	361	337.32
1999	350	339.07
2000	323	363.07
2001	314	336.70
2002	381	327.96
2003	368	318.76
2004	395	311.33
2005	427	337.14
2006	431	328.11
2007	373	264.93
2008	324	262.70
2009	243	154.89
2010	257	191.85
2011	246	207.34
2012	261	197.73
2013	266	195.28
2014	290	201.75

Correlation: Fatality, Avg. fatality (hat)

```
Pearson correlation of Fatality and Avg. fatality (hat) = 0.710
P-Value = 0.000
```

Figure B.34. Correlation result indicating predictive capability of the regression model of Nevada

Table B.19. New Hampshire fatality prediction results from regression model

Year	Fatalities	Predicted fatality
1994	119	130.68
1995	118	123.38
1996	134	121.15
1997	125	135.14
1998	128	124.20
1999	140	123.72
2000	126	132.61
2001	142	127.23
2002	127	126.30
2003	127	136.23
2004	171	142.37
2005	166	138.78
2006	127	144.50
2007	129	121.79
2008	138	144.03
2009	110	134.04
2010	128	117.58
2011	90	114.21
2012	108	118.28
2013	135	112.48
2014	95	114.25

Correlation: Fatality, Avg. fatality (hat)

```
Pearson correlation of Fatality and Avg. fatality (hat) = 0.525
P-Value = 0.015
```

Figure B.35. Correlation result indicating predictive capability of the regression model of New Hampshire

Year	Fatalities	Predicted fatality1	Predicted fatality2	Predicted fatality3	Predicted fatality4	Predicted fatality5	Predicted fatality6	Predicted fatality7	Predicted fatality8	Predicted fatality9	Predicted fatality10	Predicted Avg. fatality
1994	761	786.24	786.24	786.24	786.24	786.24	786.24	786.24	786.24	786.24	786.24	786.24
1995	774	771.99	771.99	771.99	771.99	771.99	771.99	771.99	771.99	771.99	771.99	771.99
1996	814	760.10	760.10	760.10	760.10	760.10	760.10	760.10	760.10	760.10	760.10	760.10
1997	775	722.90	722.90	722.90	722.90	722.90	722.90	722.90	722.90	722.90	722.90	722.90
1998	741	769.84	769.84	769.84	769.84	769.84	769.84	769.84	769.84	769.84	769.84	769.84
1999	726	754.18	754.18	754.18	754.18	754.18	754.18	754.18	754.18	754.18	754.18	754.18
2000	731	744.49	744.49	744.49	744.49	744.49	744.49	744.49	744.49	744.49	744.49	744.49
2001	745	762.82	762.82	762.82	762.82	762.82	762.82	762.82	762.82	762.82	762.82	762.82
2002	771	702.07	702.07	702.07	702.07	702.07	702.07	702.07	702.07	702.07	702.07	702.07
2003	733	699.65	699.65	699.65	699.65	699.65	699.65	699.65	699.65	699.65	699.65	699.65
2004	723	751.54	751.54	751.54	751.54	751.54	751.54	751.54	751.54	751.54	751.54	751.54
2005	747	749.78	749.78	749.78	749.78	749.78	749.78	749.78	749.78	749.78	749.78	749.78
2006	771	731.16	731.16	731.16	731.16	731.16	731.16	731.16	731.16	731.16	731.16	731.16
2007	724	652.95	652.95	652.95	652.95	652.95	652.95	652.95	652.95	652.95	652.95	652.95
2008	590	666.17	666.17	666.17	666.17	666.17	666.17	666.17	666.17	666.17	666.17	666.17
2009	584	634.59	634.59	634.59	634.59	634.59	634.59	634.59	634.59	634.59	634.59	634.59
2010	556	595.27	595.27	595.27	595.27	595.27	595.27	595.27	595.27	595.27	595.27	595.27
2011	627	661.81	666.52	606.19	677.26	719.06	671.22	604.49	578.77	723.14	632.41	654.09
2012	589	589.54	589.54	589.54	589.54	589.54	589.54	589.54	589.54	589.54	589.54	589.54
2013	542	550.62	550.62	550.62	550.62	550.62	550.62	550.62	550.62	550.62	550.62	550.62
2014	556	530.92	530.92	530.92	530.92	530.92	530.92	530.92	530.92	530.92	530.92	530.92

Figure B.36. New Jersey fatality prediction according to the prediction model

Correlation: Fatality, Avg. fatality (hat)

```
Pearson correlation of Fatality and Avg. fatality (hat) = 0.887
P-Value = 0.000
```

Figure B.37. Correlation result indicating predictive capability of the regression model of New Jersey

Table B.20. New Mexico fatality prediction results from regression model

Year	Fatalities	Predicted fatality
1994	447	464.04
1995	485	466.00
1996	485	472.04
1997	484	491.63
1998	424	449.11
1999	460	474.01
2000	432	446.61
2001	464	393.19
2002	449	450.02
2003	439	462.10
2004	521	488.16
2005	488	437.94
2006	484	432.33
2007	413	407.30
2008	366	353.45
2009	361	372.51
2010	349	409.99
2011	350	348.21
2012	366	347.86
2013	311	401.64
2014	383	393.35

Correlation: Fatality, Avg. fatality (hat)

```
Pearson correlation of Fatality and Avg. fatality (hat) = 0.785
P-Value = 0.000
```

Figure B.38. Correlation result indicating predictive capability of the regression model of New Mexico

Year	Fatalities	Predicted fatality1	Predicted fatality2	Predicted fatality3	Predicted fatality4	Predicted fatality5	Predicted fatality6	Predicted fatality7	Predicted fatality8	Predicted fatality9	Predicted fatality10	Avg. predicted fatality
1994	1678	1608.67	1608.67	1608.67	1608.67	1608.67	1608.67	1608.67	1608.67	1608.67	1608.67	1608.67
1995	1679	1597.65	1597.65	1597.65	1597.65	1597.65	1597.65	1597.65	1597.65	1597.65	1597.65	1597.65
1996	1593	1599.64	1599.64	1599.64	1599.64	1599.64	1599.64	1599.64	1599.64	1599.64	1599.64	1599.64
1997	1652	1574.24	1574.24	1574.24	1574.24	1574.24	1574.24	1574.24	1574.24	1574.24	1574.24	1574.24
1998	1514	1537.19	1537.19	1537.19	1537.19	1537.19	1537.19	1537.19	1537.19	1537.19	1537.19	1537.19
1999	1599	1530.93	1530.93	1530.93	1530.93	1530.93	1530.93	1530.93	1530.93	1530.93	1530.93	1530.93
2000	1460	1546.80	1546.80	1546.80	1546.80	1546.80	1546.80	1546.80	1546.80	1546.80	1546.80	1546.80
2001	1564	1536.74	1536.74	1536.74	1536.74	1536.74	1536.74	1536.74	1536.74	1536.74	1536.74	1536.74
2002	1530	1526.12	1526.12	1526.12	1526.12	1526.12	1526.12	1526.12	1526.12	1526.12	1526.12	1526.12
2003	1493	1496.27	1496.27	1496.27	1496.27	1496.27	1496.27	1496.27	1496.27	1496.27	1496.27	1496.27
2004	1495	1348.12	1348.12	1348.12	1348.12	1348.12	1348.12	1348.12	1348.12	1348.12	1348.12	1348.12
2005	1434	1419.91	1419.91	1419.91	1419.91	1419.91	1419.91	1419.91	1419.91	1419.91	1419.91	1419.91
2006	1454	1458.70	1458.70	1458.70	1458.70	1458.70	1458.70	1458.70	1458.70	1458.70	1458.70	1458.70
2007	1332	1347.05	1347.05	1347.05	1347.05	1347.05	1347.05	1347.05	1347.05	1347.05	1347.05	1347.05
2008	1238	1400.86	1400.86	1400.86	1400.86	1400.86	1400.86	1400.86	1400.86	1400.86	1400.86	1400.86
2009	1158	1257.63	1257.63	1257.63	1257.63	1257.63	1257.63	1257.63	1257.63	1257.63	1257.63	1257.63
2010	1201	1270.83	1270.83	1270.83	1270.83	1270.83	1270.83	1270.83	1270.83	1270.83	1270.83	1270.83
2011	1171	1227.70	1370.41	1323.01	1329.58	1264.53	1264.53	1363.01	1374.87	1375.79	1259.71	1315.31
2012	1180	1248.71	1248.71	1248.71	1248.71	1248.71	1248.71	1248.71	1248.71	1248.71	1248.71	1248.71
2013	1202	1179.75	1179.75	1179.75	1179.75	1179.75	1179.75	1179.75	1179.75	1179.75	1179.75	1179.75
2014	1039	899.77	899.77	899.77	899.77	899.77	899.77	899.77	899.77	899.77	899.77	899.77

Figure B.39. New York fatality prediction according to the prediction model

Correlation: Fatality, Avg. fatality (hat)

```
Pearson correlation of Fatality and Avg. fatality (hat) = 0.918
P-Value = 0.000
```

Figure B.40. Correlation result indicating predictive capability of the regression model of New York

Table B.21. North Carolina fatality prediction results from regression model

Year	Fatalities	Avg. y(hat)
1994	1431	1522.89
1995	1448	1481.58
1996	1494	1514.38
1997	1483	1515.85
1998	1596	1474.20
1999	1505	1511.29
2000	1557	1571.04
2001	1530	1481.75
2002	1576	1526.95
2003	1553	1558.41
2004	1573	1561.59
2005	1547	1573.05
2006	1554	1483.75
2007	1676	1533.12
2008	1428	1557.50
2009	1313	1288.03
2010	1320	1294.12
2011	1230	1284.30
2012	1299	1353.90
2013	1290	1338.17
2014	1284	1261.11

Correlation: Fatality, Avg. fatality (hat)

```
Pearson correlation of Fatality and Avg. fatality (hat) = 0.857
P-Value = 0.000
```

Figure B.41. Correlation result indicating predictive capability of the regression model of North Carolina

Year	Fatalities	Predicted fatality1	Predicted fatality2	Predicted fatality3	Predicted fatality4	Predicted fatality5	Predicted fatality6	Predicted fatality7	Predicted fatality8	Predicted fatality9	Predicted fatality10	Avg. predicted fatality
1994	88	83.75	97.77	56.93	71.79	78.80	91.04	65.84	86.10	81.97	79.07	79.31
1995	74	76.96	101.29	133.03	106.66	100.26	106.13	94.97	60.19	95.74	79.82	95.51
1996	85	96.99	93.71	85.32	100.87	100.03	86.56	81.69	71.50	105.06	77.99	89.97
1997	105	105.08	105.08	105.08	105.08	105.08	105.08	105.08	105.08	105.08	105.08	105.08
1998	92	104.12	104.12	104.12	104.12	104.12	104.12	104.12	104.12	104.12	104.12	104.12
1999	119	102.96	102.96	102.96	102.96	102.96	102.96	102.96	102.96	102.96	102.96	102.96
2000	86	105.85	105.85	105.85	105.85	105.85	105.85	105.85	105.85	105.85	105.85	105.85
2001	105	101.61	101.61	101.61	101.61	101.61	101.61	101.61	101.61	101.61	101.61	101.61
2002	97	97.17	97.17	97.17	97.17	97.17	97.17	97.17	97.17	97.17	97.17	97.17
2003	105	101.61	101.61	101.61	101.61	101.61	101.61	101.61	101.61	101.61	101.61	101.61
2004	100	102.38	102.38	102.38	102.38	102.38	102.38	102.38	102.38	102.38	102.38	102.38
2005	123	109.91	109.91	109.91	109.91	109.91	109.91	109.91	109.91	109.91	109.91	109.91
2006	111	115.89	115.89	115.89	115.89	115.89	115.89	115.89	115.89	115.89	115.89	115.89
2007	111	117.63	117.63	117.63	117.63	117.63	117.63	117.63	117.63	117.63	117.63	117.63
2008	104	111.45	111.45	111.45	111.45	111.45	111.45	111.45	111.45	111.45	111.45	111.45
2009	140	110.87	110.87	110.87	110.87	110.87	110.87	110.87	110.87	110.87	110.87	110.87
2010	105	122.45	122.45	122.45	122.45	122.45	122.45	122.45	122.45	122.45	122.45	122.45
2011	148	158.73	158.73	158.73	158.73	158.73	158.73	158.73	158.73	158.73	158.73	158.73
2012	170	135.96	135.96	135.96	135.96	135.96	135.96	135.96	135.96	135.96	135.96	135.96
2013	148	138.08	138.08	138.08	138.08	138.08	138.08	138.08	138.08	138.08	138.08	138.08
2014	135	136.93	136.93	136.93	136.93	136.93	136.93	136.93	136.93	136.93	136.93	136.93

Figure B.42. North Dakota fatality prediction according to the prediction model

Correlation: Fatality, Avg. fatality (hat)

```
Pearson correlation of Fatality and Avg. fatality (hat) = 0.826
P-Value = 0.000
```

Figure B.43. Correlation result indicating predictive capability of the regression model of North Dakota

Table B.22. Ohio fatality prediction results from regression model

Year	Fatalities	Predicted fatality
1994	1370	1357.59
1995	1360	1378.41
1996	1391	1313.23
1997	1441	1385.44
1998	1422	1392.22
1999	1430	1434.51
2000	1366	1401.25
2001	1378	1403.24
2002	1418	1310.03
2003	1274	1231.61
2004	1286	1237.40
2005	1321	1290.73
2006	1238	1293.53
2007	1255	1296.53
2008	1191	1218.88
2009	1022	1047.47
2010	1080	1066.85
2011	1017	1031.76
2012	1121	1008.71
2013	989	982.79
2014	1006	1096.33

Correlation: Fatality, Avg. fatality (hat)

```
Pearson correlation of Fatality and Avg. fatality (hat) = 0.943
P-Value = 0.000
```

Figure B.44. Correlation result indicating predictive capability of the regression model of Ohio

Table B.23. Oregon fatality prediction results from regression model

Year	Fatalities	Predicted fatality
1994	494	468.94
1995	574	532.60
1996	526	547.50
1997	524	486.64
1998	538	519.16
1999	414	482.85
2000	451	462.69
2001	488	442.66
2002	436	441.46
2003	512	461.34
2004	456	480.43
2005	487	475.07
2006	478	490.52
2007	455	450.42
2008	416	426.99
2009	377	323.70
2010	317	286.81
2011	331	411.48
2012	337	404.22
2013	313	362.70
2014	357	356.59

Correlation: Fatality, Avg. fatality (hat)

```
Pearson correlation of Fatality and Avg. fatality (hat) = 0.858
P-Value = 0.000
```

Figure B.45. Correlation result indicating predictive capability of the regression model of Oregon

Table B.24. Pennsylvania fatality prediction results from regression model

Year	Fatalities	Predicted fatality
1994	1441	1536.22
1995	1480	1471.99
1996	1469	1468.96
1997	1557	1526.79
1998	1481	1552.97
1999	1549	1493.40
2000	1520	1513.31
2001	1532	1494.50
2002	1614	1513.53
2003	1577	1578.83
2004	1490	1545.08
2005	1616	1541.54
2006	1525	1602.15
2007	1491	1433.84
2008	1468	1433.91
2009	1256	1251.06
2010	1324	1268.87
2011	1286	1307.15
2012	1310	1276.05
2013	1210	1285.98
2014	1195	1294.88

Correlation: Fatality, Avg. fatality (hat)

```
Pearson correlation of Fatality and Avg. fatality (hat) = 0.892
P-Value = 0.000
```

Figure B.46. Correlation result indicating predictive capability of the regression model of Pennsylvania

Year	Fatalities	Predicted fatality1	Predicted fatality2	Predicted fatality3	Predicted fatality4	Predicted fatality5	Predicted fatality6	Predicted fatality7	Predicted fatality8	Predicted fatality9	Predicted fatality10	Avg. predicted fatality
1994	63	66.69	66.69	66.69	66.69	66.69	66.69	66.69	66.69	66.69	66.69	66.69
1995	69	66.42	66.42	66.42	66.42	66.42	66.42	66.42	66.42	66.42	66.42	66.42
1996	69	75.39	75.39	75.39	75.39	75.39	75.39	75.39	75.39	75.39	75.39	75.39
1997	75	79.76	79.76	79.76	79.76	79.76	79.76	79.76	79.76	79.76	79.76	79.76
1998	74	72.07	72.07	72.07	72.07	72.07	72.07	72.07	72.07	72.07	72.07	72.07
1999	88	85.49	85.49	85.49	85.49	85.49	85.49	85.49	85.49	85.49	85.49	85.49
2000	80	84.54	84.54	84.54	84.54	84.54	84.54	84.54	84.54	84.54	84.54	84.54
2001	81	84.61	84.61	84.61	84.61	84.61	84.61	84.61	84.61	84.61	84.61	84.61
2002	84	82.69	82.69	82.69	82.69	82.69	82.69	82.69	82.69	82.69	82.69	82.69
2003	104	87.79	87.79	87.79	87.79	87.79	87.79	87.79	87.79	87.79	87.79	87.79
2004	83	84.51	84.51	84.51	84.51	84.51	84.51	84.51	84.51	84.51	84.51	84.51
2005	87	80.85	80.85	80.85	80.85	80.85	80.85	80.85	80.85	80.85	80.85	80.85
2006	81	77.55	77.55	77.55	77.55	77.55	77.55	77.55	77.55	77.55	77.55	77.55
2007	69	68.85	68.85	68.85	68.85	68.85	68.85	68.85	68.85	68.85	68.85	68.85
2008	65	75.49	75.49	75.49	75.49	75.49	75.49	75.49	75.49	75.49	75.49	75.49
2009	83	66.78	66.78	66.78	66.78	66.78	66.78	66.78	66.78	66.78	66.78	66.78
2010	67	63.65	63.65	63.65	63.65	63.65	63.65	63.65	63.65	63.65	63.65	63.65
2011	66	66.76	73.30	75.30	72.84	83.58	78.92	80.38	72.42	70.20	72.42	74.61
2012	64	57.48	57.48	57.48	57.48	57.48	57.48	57.48	57.48	57.48	57.48	57.48
2013	65	67.22	67.22	67.22	67.22	67.22	67.22	67.22	67.22	67.22	67.22	67.22
2014	52	66.95	66.95	66.95	66.95	66.95	66.95	66.95	66.95	66.95	66.95	66.95

Figure B.47. Rhode Island fatality prediction according to the prediction model

Correlation: Fatality, Avg. fatality (hat)

```
Pearson correlation of Fatality and Avg. fatality (hat) = 0.771
P-Value = 0.000
```

Figure B.48. Correlation result indicating predictive capability of the regression model of Rhode Island

Table B.25. South Carolina fatality prediction results from regression model

Year	Fatalities	Predicted Fatality
1994	847	850.40
1995	881	931.57
1996	930	954.85
1997	903	967.38
1998	1002	972.59
1999	1065	957.29
2000	1065	1050.90
2001	1060	1043.64
2002	1053	1020.86
2003	969	964.85
2004	1046	958.77
2005	1094	1008.22
2006	1045	1025.37
2007	1077	1086.62
2008	921	1030.02
2009	894	804.07
2010	809	822.66
2011	828	820.46
2012	863	870.78
2013	768	914.26
2014	824	916.78

Correlation: Fatality, Avg. fatality (hat)

```
Pearson correlation of Fatality and Avg. fatality (hat) = 0.778
P-Value = 0.000
```

Figure B.49. Correlation result indicating predictive capability of the regression model of South Carolina

Year	Fatalities	Predicted fatality1	Predicted fatality2	Predicted fatality3	Predicted fatality4	Predicted fatality5	Predicted fatality6	Predicted fatality7	Predicted fatality8	Predicted fatality9	Predicted fatality10	Avg. Predicted fatality
1994	154	161.64	196.50	176.67	137.89	132.49	150.40	177.16	154.93	186.58	154.50	162.88
1995	158	163.04	163.04	163.04	163.04	163.04	163.04	163.04	163.04	163.04	163.04	163.04
1996	175	173.64	173.64	173.64	173.64	173.64	173.64	173.64	173.64	173.64	173.64	173.64
1997	148	174.60	174.60	174.60	174.60	174.60	174.60	174.60	174.60	174.60	174.60	174.60
1998	165	148.58	148.58	148.58	148.58	148.58	148.58	148.58	148.58	148.58	148.58	148.58
1999	150	168.82	168.82	168.82	168.82	168.82	168.82	168.82	168.82	168.82	168.82	168.82
2000	173	170.75	170.75	170.75	170.75	170.75	170.75	170.75	170.75	170.75	170.75	170.75
2001	171	167.86	167.86	167.86	167.86	167.86	167.86	167.86	167.86	167.86	167.86	167.86
2002	180	168.82	168.82	168.82	168.82	168.82	168.82	168.82	168.82	168.82	168.82	168.82
2003	203	169.78	169.78	169.78	169.78	169.78	169.78	169.78	169.78	169.78	169.78	169.78
2004	197	173.64	173.64	173.64	173.64	173.64	173.64	173.64	173.64	173.64	173.64	173.64
2005	186	172.67	172.67	172.67	172.67	172.67	172.67	172.67	172.67	172.67	172.67	172.67
2006	191	164.00	164.00	164.00	164.00	164.00	164.00	164.00	164.00	164.00	164.00	164.00
2007	146	157.25	157.25	157.25	157.25	157.25	157.25	157.25	157.25	157.25	157.25	157.25
2008	121	154.36	154.36	154.36	154.36	154.36	154.36	154.36	154.36	154.36	154.36	154.36
2009	131	150.50	150.50	150.50	150.50	150.50	150.50	150.50	150.50	150.50	150.50	150.50
2010	140	146.65	146.65	146.65	146.65	146.65	146.65	146.65	146.65	146.65	146.65	146.65
2011	111	137.01	137.01	137.01	137.01	137.01	137.01	137.01	137.01	137.01	137.01	137.01
2012	133	127.37	127.37	127.37	127.37	127.37	127.37	127.37	127.37	127.37	127.37	127.37
2013	135	125.44	125.44	125.44	125.44	125.44	125.44	125.44	125.44	125.44	125.44	125.44
2014	136	121.59	121.59	121.59	121.59	121.59	121.59	121.59	121.59	121.59	121.59	121.59

Figure B.50. South Dakota fatality prediction according to the prediction model

Correlation: Fatality, Avg. fatality (hat)

```
Pearson correlation of Fatality and Avg. fatality (hat) = 0.699
P-Value = 0.000
```

Figure B.51. Correlation result indicating predictive capability of the regression model of South Dakota

Year	Fatalities	Predicted fatality1	Predicted fatality2	Predicted fatality3	Predicted fatality4	Predicted fatality5	Predicted fatality6	Predicted fatality7	Predicted fatality8	Predicted fatality9	Predicted fatality10	Avg. predicted fatality
1994	1214	1253.56	1142.05	1237.56	1244.45	1317.86	1279.64	1183.27	1173.36	1188.09	1222.00	1224.18
1995	1259	1287.29	1169.56	1231.24	1225.44	1282.23	1251.14	1228.59	1208.19	1109.76	1185.07	1217.85
1996	1239	1283.14	1268.16	1196.12	1254.23	1312.16	1354.86	1225.99	1213.31	1217.95	1234.21	1256.01
1997	1225	1274.13	1280.72	1283.35	1331.32	1241.77	1226.40	1199.17	1265.26	1222.34	1254.82	1257.93
1998	1216	1274.04	1274.04	1274.04	1274.04	1274.04	1274.04	1274.04	1274.04	1274.04	1274.04	1274.04
1999	1302	1271.80	1271.80	1271.80	1271.80	1271.80	1271.80	1271.80	1271.80	1271.80	1271.80	1271.80
2000	1307	1226.64	1226.64	1226.64	1226.64	1226.64	1226.64	1226.64	1226.64	1226.64	1226.64	1226.64
2001	1251	1230.78	1230.78	1230.78	1230.78	1230.78	1230.78	1230.78	1230.78	1230.78	1230.78	1230.78
2002	1177	1239.33	1239.33	1239.33	1239.33	1239.33	1239.33	1239.33	1239.33	1239.33	1239.33	1239.33
2003	1193	1277.39	1277.39	1277.39	1277.39	1277.39	1277.39	1277.39	1277.39	1277.39	1277.39	1277.39
2004	1339	1254.59	1254.59	1254.59	1254.59	1254.59	1254.59	1254.59	1254.59	1254.59	1254.59	1254.59
2005	1270	1205.66	1205.66	1205.66	1205.66	1205.66	1205.66	1205.66	1205.66	1205.66	1205.66	1205.66
2006	1284	1205.06	1205.06	1205.06	1205.06	1205.06	1205.06	1205.06	1205.06	1205.06	1205.06	1205.06
2007	1211	1207.42	1207.42	1207.42	1207.42	1207.42	1207.42	1207.42	1207.42	1207.42	1207.42	1207.42
2008	1043	1132.72	1132.72	1132.72	1132.72	1132.72	1132.72	1132.72	1132.72	1132.72	1132.72	1132.72
2009	986	903.29	951.97	1002.52	1011.36	1045.16	902.55	949.44	985.47	987.94	945.51	968.52
2010	1032	1023.15	1023.15	1023.15	1023.15	1023.15	1023.15	1023.15	1023.15	1023.15	1023.15	1023.15
2011	937	992.48	1013.65	954.14	1061.37	1003.40	975.62	994.71	1006.65	965.26	938.77	990.60
2012	1015	1111.15	979.75	1036.44	1013.69	1098.80	1040.34	1059.87	1074.58	1033.07	1051.38	1049.91
2013	995	958.33	952.61	953.72	941.20	1010.52	971.75	962.46	1052.36	1019.34	980.05	980.23
2014	962	897.09	980.79	963.95	960.98	1032.39	958.22	1013.83	957.84	1006.97	957.25	972.93

Figure B.52.Tennessee fatality prediction according to the prediction model

Correlation: Fatality, Avg. fatality (hat)

```
Pearson correlation of Fatality and Avg. fatality (hat) = 0.885
P-Value = 0.000
```

Figure B.53. Correlation result indicating predictive capability of the regression model of Tennessee

Table B.26. Texas fatality prediction results from regression model

Year	Fatalities	Predicted fatality
1994	3187	3592.60
1995	3183	3594.30
1996	3742	3542.57
1997	3513	3525.77
1998	3586	3550.47
1999	3522	3568.02
2000	3779	3592.62
2001	3736	3451.63
2002	3823	3522.96
2003	3821	3593.08
2004	3699	3507.83
2005	3536	3611.21
2006	3531	3570.93
2007	3466	3496.34
2008	3476	3540.70
2009	3104	3404.83
2010	3023	3220.01
2011	3054	3302.96
2012	3408	3333.37
2013	3389	3311.57
2014	3538	3282.26

Correlation: Fatality, Avg. fatality (hat)

```
Pearson correlation of Fatality and Avg. fatality (hat) = 0.487
P-Value = 0.025
```

Figure B.54. Correlation result indicating predictive capability of the regression model of Texas

Table B.27. Utah fatality prediction results from regression model

Year	Fatalities	Predicted fatality
1994	343	316.52
1995	325	331.85
1996	321	314.13
1997	366	282.66
1998	350	280.25
1999	360	313.55
2000	373	323.79
2001	291	302.25
2002	328	312.57
2003	309	265.98
2004	296	333.04
2005	282	285.40
2006	287	322.56
2007	299	302.25
2008	276	327.08
2009	244	262.79
2010	253	268.62
2011	243	289.87
2012	217	253.50
2013	220	265.27
2014	256	285.05

Correlation: Fatality, Avg. fatality (hat)

```
Pearson correlation of Fatality and Avg. fatality (hat) = 0.535
P-Value = 0.013
```

Figure B.55. Correlation result indicating predictive capability of the regression model of Utah

Year	Fatalities	Predicted fatality1	Predicted fatality2	Predicted fatality3	Predicted fatality4	Predicted fatality5	Predicted fatality6	Predicted fatality7	Predicted fatality8	Predicted fatality9	Predicted fatality10	Avg. predicted fatality
1994	77	100.65	94.04	92.24	77.49	88.09	83.75	85.13	78.65	80.14	88.32	86.85
1995	106	95.61	95.61	95.61	95.61	95.61	95.61	95.61	95.61	95.61	95.61	95.61
1996	88	81.97	81.97	81.97	81.97	81.97	81.97	81.97	81.97	81.97	81.97	81.97
1997	96	93.84	93.84	93.84	93.84	93.84	93.84	93.84	93.84	93.84	93.84	93.84
1998	104	97.41	97.41	97.41	97.41	97.41	97.41	97.41	97.41	97.41	97.41	97.41
1999	90	89.07	89.07	89.07	89.07	89.07	89.07	89.07	89.07	89.07	89.07	89.07
2000	76	82.61	82.61	82.61	82.61	82.61	82.61	82.61	82.61	82.61	82.61	82.61
2001	92	94.39	94.39	94.39	94.39	94.39	94.39	94.39	94.39	94.39	94.39	94.39
2002	78	76.60	76.60	76.60	76.60	76.60	76.60	76.60	76.60	76.60	76.60	76.60
2003	69	72.88	72.88	72.88	72.88	72.88	72.88	72.88	72.88	72.88	72.88	72.88
2004	98	86.80	86.80	86.80	86.80	86.80	86.80	86.80	86.80	86.80	86.80	86.80
2005	73	71.79	71.79	71.79	71.79	71.79	71.79	71.79	71.79	71.79	71.79	71.79
2006	87	72.42	72.42	72.42	72.42	72.42	72.42	72.42	72.42	72.42	72.42	72.42
2007	66	77.44	77.44	77.44	77.44	77.44	77.44	77.44	77.44	77.44	77.44	77.44
2008	73	71.86	71.86	71.86	71.86	71.86	71.86	71.86	71.86	71.86	71.86	71.86
2009	74	67.22	67.22	67.22	67.22	67.22	67.22	67.22	67.22	67.22	67.22	67.22
2010	71	61.04	61.04	61.04	61.04	61.04	61.04	61.04	61.04	61.04	61.04	61.04
2011	55	57.31	57.31	57.31	57.31	57.31	57.31	57.31	57.31	57.31	57.31	57.31
2012	77	64.47	64.47	64.47	64.47	64.47	64.47	64.47	64.47	64.47	64.47	64.47
2013	69	59.70	59.70	59.70	59.70	59.70	59.70	59.70	59.70	59.70	59.70	59.70
2014	44	58.23	58.23	58.23	58.23	58.23	58.23	58.23	58.23	58.23	58.23	58.23

Figure B.56. Vermont fatality prediction according to the prediction model

Correlation: Fatality, Avg. fatality (hat)

```
Pearson correlation of Fatality and Avg. fatality (hat) = 0.791
P-Value = 0.000
```

Figure B.57. Correlation result indicating predictive capability of the regression model of Vermont

Year	Fatalities	Predicted fatality1	Predicted fatality2	Predicted fatality3	Predicted fatality4	Predicted fatality5	Predicted fatality6	Predicted fatality7	Predicted fatality8	Predicted fatality9	Predicted fatality10	Avg.predicted fatality
1994	930	915.38	892.13	947.03	754.92	901.30	975.32	890.02	1027.39	921.02	1002.81	922.73
1995	900	877.20	877.20	877.20	877.20	877.20	877.20	877.20	877.20	877.20	877.20	877.20
1996	877	834.83	834.83	834.83	834.83	834.83	834.83	834.83	834.83	834.83	834.83	834.83
1997	984	973.64	973.64	973.64	973.64	973.64	973.64	973.64	973.64	973.64	973.64	973.64
1998	935	933.14	933.14	933.14	933.14	933.14	933.14	933.14	933.14	933.14	933.14	933.14
1999	878	892.82	892.82	892.82	892.82	892.82	892.82	892.82	892.82	892.82	892.82	892.82
2000	929	942.01	942.01	942.01	942.01	942.01	942.01	942.01	942.01	942.01	942.01	942.01
2001	935	929.74	929.74	929.74	929.74	929.74	929.74	929.74	929.74	929.74	929.74	929.74
2002	914	882.26	882.26	882.26	882.26	882.26	882.26	882.26	882.26	882.26	882.26	882.26
2003	943	937.75	937.75	937.75	937.75	937.75	937.75	937.75	937.75	937.75	937.75	937.75
2004	922	953.60	953.60	953.60	953.60	953.60	953.60	953.60	953.60	953.60	953.60	953.60
2005	947	901.81	901.81	901.81	901.81	901.81	901.81	901.81	901.81	901.81	901.81	901.81
2006	962	938.23	938.23	938.23	938.23	938.23	938.23	938.23	938.23	938.23	938.23	938.23
2007	1027	894.22	894.22	894.22	894.22	894.22	894.22	894.22	894.22	894.22	894.22	894.22
2008	825	795.77	795.77	795.77	795.77	795.77	795.77	795.77	795.77	795.77	795.77	795.77
2009	758	834.29	834.29	834.29	834.29	834.29	834.29	834.29	834.29	834.29	834.29	834.29
2010	740	828.88	828.88	828.88	828.88	828.88	828.88	828.88	828.88	828.88	828.88	828.88
2011	764	1009.83	1019.47	858.03	771.50	768.51	853.08	805.26	863.66	1072.66	755.47	877.75
2012	776	742.81	742.81	742.81	742.81	742.81	742.81	742.81	742.81	742.81	742.81	742.81
2013	740	748.58	748.58	748.58	748.58	748.58	748.58	748.58	748.58	748.58	748.58	748.58
2014	703	779.71	779.71	779.71	779.71	779.71	779.71	779.71	779.71	779.71	779.71	779.71

Figure B.58. Virginia fatality prediction according to the prediction model

Correlation: Fatality, Avg. fatality (hat)

```
Pearson correlation of Fatality and Avg. fatality (hat) = 0.642
P-Value = 0.002
```

Figure B.59. Correlation result indicating predictive capability of the regression model of Virginia

Table B.28. Washington fatality prediction results from regression model

Year	Fatalities	Predicted fatality
1994	640	643.70
1995	653	642.79
1996	712	645.67
1997	674	657.41
1998	662	657.77
1999	637	655.59
2000	631	652.72
2001	649	638.18
2002	658	622.97
2003	600	621.50
2004	567	635.07
2005	649	643.11
2006	633	648.57
2007	571	520.93
2008	521	507.92
2009	492	451.59
2010	460	441.69
2011	454	454.80
2012	438	470.87
2013	436	487.31
2014	462	498.94

Correlation: Fatality, Avg. fatality (hat)

```
Pearson correlation of Fatality and Avg. fatality (hat) = 0.930
P-Value = 0.000
```

Figure B.60. Correlation result indicating predictive capability of the regression model of Washington

Table B.29. West Virginia fatality prediction results from regression model

Year	Fatalities	Predicted fatality
1994	356	345.51
1995	376	359.37
1996	348	364.06
1997	381	374.09
1998	354	377.66
1999	395	379.32
2000	411	394.10
2001	376	401.37
2002	439	388.21
2003	394	387.64
2004	410	399.36
2005	374	402.34
2006	410	403.04
2007	432	407.72
2008	378	410.16
2009	357	356.61
2010	315	344.11
2011	338	354.52
2012	339	304.92
2013	332	318.13
2014	272	320.88

Correlation: Fatality, Avg. fatality (hat)

```
Pearson correlation of Fatality and Avg. fatality (hat) = 0.782
P-Value = 0.000
```

Figure B.61. Correlation result indicating predictive capability of the regression model of West Virginia

Year	Fatalities	Predicted fatality1	Predicted fatality2	Predicted fatality3	Predicted fatality4	Predicted fatality5	Predicted fatality6	Predicted fatality7	Predicted fatality8	Predicted fatality9	Predicted fatality10	Avg. predicted fatality
1994	712	739.69	748.15	756.91	753.33	696.51	751.71	799.76	740.01	674.23	669.40	732.97
1995	745	684.70	684.70	684.70	684.70	684.70	684.70	684.70	684.70	684.70	684.70	684.70
1996	761	751.27	751.27	751.27	751.27	751.27	751.27	751.27	751.27	751.27	751.27	751.27
1997	725	733.10	733.10	733.10	733.10	733.10	733.10	733.10	733.10	733.10	733.10	733.10
1998	714	749.50	749.50	749.50	749.50	749.50	749.50	749.50	749.50	749.50	749.50	749.50
1999	745	780.00	780.00	780.00	780.00	780.00	780.00	780.00	780.00	780.00	780.00	780.00
2000	799	762.49	762.49	762.49	762.49	762.49	762.49	762.49	762.49	762.49	762.49	762.49
2001	763	753.56	753.56	753.56	753.56	753.56	753.56	753.56	753.56	753.56	753.56	753.56
2002	803	752.32	752.32	752.32	752.32	752.32	752.32	752.32	752.32	752.32	752.32	752.32
2003	848	860.83	860.83	860.83	860.83	860.83	860.83	860.83	860.83	860.83	860.83	860.83
2004	792	770.49	770.49	770.49	770.49	770.49	770.49	770.49	770.49	770.49	770.49	770.49
2005	815	754.84	754.84	754.84	754.84	754.84	754.84	754.84	754.84	754.84	754.84	754.84
2006	724	758.05	758.05	758.05	758.05	758.05	758.05	758.05	758.05	758.05	758.05	758.05
2007	756	749.16	749.16	749.16	749.16	749.16	749.16	749.16	749.16	749.16	749.16	749.16
2008	605	713.87	713.87	713.87	713.87	713.87	713.87	713.87	713.87	713.87	713.87	713.87
2009	561	582.16	582.16	582.16	582.16	582.16	582.16	582.16	582.16	582.16	582.16	582.16
2010	572	565.62	565.62	565.62	565.62	565.62	565.62	565.62	565.62	565.62	565.62	565.62
2011	582	601.86	601.86	601.86	601.86	601.86	601.86	601.86	601.86	601.86	601.86	601.86
2012	615	551.30	551.30	551.30	551.30	551.30	551.30	551.30	551.30	551.30	551.30	551.30
2013	543	543.10	543.10	543.10	543.10	543.10	543.10	543.10	543.10	543.10	543.10	543.10
2014	507	577.53	577.53	577.53	577.53	577.53	577.53	577.53	577.53	577.53	577.53	577.53

Figure B.62. Wisconsin fatality prediction according to the prediction model

Correlation: Fatality, Avg. fatality (hat)

```
Pearson correlation of Fatality and Avg. fatality (hat) = 0.904
P-Value = 0.000
```

Figure B.63. Correlation result indicating predictive capability of the regression model of Wisconsin

Table B.30. Wyoming fatality prediction results from regression model

Year	Fatalities	Predicted fatality
1994	144	171.94
1995	170	173.81
1996	143	156.35
1997	137	146.79
1998	154	158.69
1999	189	160.32
2000	152	157.94
2001	186	162.27
2002	176	169.99
2003	165	158.64
2004	164	167.27
2005	170	180.46
2006	195	158.64
2007	150	159.95
2008	159	174.29
2009	134	127.40
2010	155	139.90
2011	135	130.71
2012	123	130.48
2013	87	112.03
2014	150	140.11

Correlation: Fatality, Avg. fatality (hat)

```
Pearson correlation of Fatality and Avg. fatality (hat) = 0.738
P-Value = 0.000
```

Figure B.64. Correlation result indicating predictive capability of the regression model of Wyoming

APPENDIX C

POOLED ESTIMATES OF PROXY CULTURE REGRESSION MODELS

```
                          est           se         t     df     Pr(>|t|)         lo 95         hi 95 nmis      fmi lambda
(Intercept)        664.98579818 6.169433e+02  1.077872 13.332 3.002079e-01 -6.644730e+02  1.994445e+03   NA 0.122459      0
Distracted.Driving 976.83587595 4.345347e+02  2.248004 13.332 4.209049e-02  4.045152e+01  1.913220e+03    0 0.122459      0
Unemployed          -0.00127528 2.100239e-04 -6.072068 13.332 3.549449e-05 -1.727863e-03 -8.226964e-04    0 0.122459      0
factor(DemorRep)1  -521.29040183 8.315304e+01 -6.269048 13.332 2.571847e-05 -7.004779e+02 -3.421029e+02   NA 0.122459      0
Dem                 52.91212821 1.126936e+01  4.695219 13.332 3.917186e-04  2.862764e+01  7.719662e+01    0 0.122459      0
VotePer           1048.98970589 5.131185e+02  2.044342 13.332 6.119211e-02 -5.673588e+01  2.154715e+03    0 0.122459      0
> pool.r.squared(fitflop)
         est     lo 95     hi 95          fmi
R^2 0.8741401 0.7120844 0.9480214 2.000442e-05
> pool.r.squared(fitflop,adjusted = TRUE)
             est     lo 95     hi 95          fmi
adj R^2 0.8321868 0.6284302 0.9297167 2.000442e-05
>
```

Figure C.1. Pooled estimates and model fit for Florida

```
                          est         se         t       df     Pr(>|t|)        lo 95       hi 95 nmis       fmi lambda
(Intercept)          3222.88108  490.06488  6.576438 16.28409 5.837908e-06   2185.46100  4260.30116   NA 0.1037125      0
Distracted.Driving -13804.62165 2913.98560 -4.737368 16.28409 2.130669e-04 -19973.24815 -7635.99515    0 0.1037125      0
Dem                  -36.31951   11.39869 -3.186287 16.28409 5.640030e-03    -60.44945   -12.18958    0 0.1037125      0
> pool.r.squared(fitgpol)
         est     lo 95    hi 95          fmi
R^2 0.7852591 0.5417142 0.908591 2.000442e-05
> pool.r.squared(fitgpol,adjusted = TRUE)
            est     lo 95     hi 95          fmi
adj R^2 0.761399 0.5003177 0.8975684 2.000442e-05
>
```

Figure C.2. Pooled estimates and model fit for Georgia

```
                          est         se         t       df     Pr(>|t|)     lo 95       hi 95 nmis       fmi lambda
(Intercept)         829.981047 211.969988  3.915559 15.29847 0.0013308121 378.94422 1281.017868   NA 0.1092988      0
Drunk.Driving      1357.850597 457.573342  2.967504 15.29847 0.0094248885 384.21076 2331.490439    0 0.1092988      0
Dem                  -8.109994   2.973001 -2.727882 15.29847 0.0153535515 -14.43605   -1.783944    0 0.1092988      0
Distracted.Driving 1123.993902 234.986306  4.783231 15.29847 0.0002291772 623.98219 1624.005614    0 0.1092988      0
> pool.r.squared(fitpol)
          est     lo 95     hi 95          fmi
R^2 0.8271952 0.6188668 0.9275032 2.000442e-05
> pool.r.squared(fitpol,adjusted = TRUE)
             est     lo 95     hi 95          fmi
adj R^2 0.7967003 0.5622025 0.9138077 2.000442e-05
>
```

Figure C.3. Pooled estimates and model fit for Indiana

```
                            est           se         t        df     Pr(>|t|)        lo 95        hi 95 nmis       fmi    lambda
(Intercept)          1207.4417915 106.85621312 11.299687  7.194253 7.739334e-06  956.1434866 1458.7400964   NA 0.5310072 0.4165385
Drunk.Driving         -448.3356317 131.81320914 -3.401295 11.836444 5.354689e-03 -735.9727023 -160.6985611    0 0.2632810 0.1484955
Distracted.Driving    1801.0897439 507.33469027  3.550102 11.998291 3.996738e-03  695.6849496 2906.4945382    0 0.2542865 0.1395463
Emergency.Response.Team  -0.1206136   0.03498424 -3.447655  8.315881 8.224934e-03   -0.2007571   -0.0404701    3 0.4620736 0.3465877
Dem                     -7.0720640   2.90208831 -2.436888  7.451298 4.291734e-02  -13.8510493   -0.2930786    0 0.5148339 0.4000193
> pool.r.squared(fitlowp)
         est     lo 95     hi 95       fmi
R^2 0.8962828 0.7304433 0.9624887 0.2374489
> pool.r.squared(fitlowp,adjusted = TRUE)
             est     lo 95   hi 95       fmi
adj R^2 0.8703002 0.6700396 0.95283 0.2431964
>
```

Figure C.4. Pooled estimates and model fit for Iowa

```
                    est           se         t       df   Pr(>|t|)         lo 95         hi 95 nmis       fmi     lambda
(Intercept) -3.443590e+03 1.666884e+03 -2.065885 12.05778 0.06102157 -7.073490e+03  1.863089e+02   NA 0.2865478 0.17727158
Rep          8.983886e+00 3.509383e+00  2.559961 11.00972 0.02650422  1.260617e+00  1.670715e+01    0 0.3402880 0.23042495
Unemployed  -4.811690e-04 1.627468e-04 -2.956550 13.95814 0.01043776 -8.303243e-04 -1.320136e-04    0 0.1874672 0.07882627
Male.Driver  9.250522e+01 3.505226e+01  2.639066 11.61282 0.02212859  1.584954e+01  1.691609e+02    1 0.3093041 0.19978121
> pool.r.squared(fit1)
          est    lo 95     hi 95        fmi
R^2 0.8463726 0.649263 0.9374488 0.05098351
> pool.r.squared(fit1,adjusted = TRUE)
             est     lo 95     hi 95        fmi
adj R^2 0.8192474 0.5960577 0.9257697 0.05260823
>
```

Figure C.5. Pooled estimates and model fit for North Carolina