Report No. K-TRAN: KSU-05-6
FINAL REPORT

DATA MINING THE KANSAS TRAFFIC-CRASH DATABASE

Yacoub Najjar, Ph.D., P.E.
Srinivas Mandavilli, Ph.D., P.E.
Kansas State University
Manhattan, Kansas

August 2009

A COOPERATIVE TRANSPORTATION RESEARCH PROGRAM BETWEEN:

KANSAS DEPARTMENT OF TRANSPORTATION
KANSAS STATE UNIVERSITY
UNIVERSITY OF KANSAS

KANSAS DEPARTMENT OF TRANSPORTATION

1 **Report No.** K-TRAN: KSU-05-6	2 **Government Accession No.**	3 **Recipient Catalog No.**
4 **Title and Subtitle** Data Mining the Kansas Traffic-Crash Database		5 **Report Date** August 2009 6 **Performing Organization Code**
7 **Author(s)** Yacoub Najjar, Ph.D., P.E., Srinivas Mandavilli, Ph.D., P.E.		8 **Performing Organization Report No.**
9 **Performing Organization Name and Address** Department of Civil Engineering Kansas State University 2118 Fiedler Hall Manhattan, Kansas 66506		10 **Work Unit No. (TRAIS)** 11 **Contract or Grant No.** C1532
12 **Sponsoring Agency Name and Address** Kansas Department of Transportation Bureau of Materials and Research 700 SW Harrison Street Topeka, Kansas 66603-3745		13 **Type of Report and Period Covered** Final Report February 2005 - July 2009 14 **Sponsoring Agency Code** RE-0372-01

15 **Supplementary Notes**
For more information write to address in block 9.

16 **Abstract**

Traffic crashes results from the interaction of different parameters which includes highway geometrics, traffic characteristics and human factors. Geometric variables include number of lanes, lane width, median width, shoulder width, roadway section length, and shoulder width while traffic characteristics include AADT, Percentage of Heavy Vehicles and Speed. The effect of these parameters can be correlated by crash prediction models that predict crash rates at particular roadway section.

Transportation Agencies and State Departments of Transportation are continuously faced with decisions concerning the safety of highways. The evaluation and comparison of alternative long-range highway plans should include the safety implications of respective plans. The commonly available models for safety analysis are crash prediction models. By performing an in-depth analysis of crash databases and developing crash rate prediction models, better decisions can be taken in regard to future traffic planning operations.

The main objective of this study is to utilize artificial neural network techniques and develop crash rate prediction models for Kansas road networks. Six networks have been studied and crash prediction models for each network have been developed.

The models developed for each of the road networks are unique and show that geometric variables and traffic have a significant impact on the crash behavior. The models developed in this study would be utilized by Kansas Department of Transportation in evaluating roadway design features, reconstruction impacts and to make decisions in regard to future traffic planning operations. Sensitivity analysis was performed on all the geometric variables in the models. It has been found that all the continuous variables have different effects on different networks. It is very difficult to generalize the behavior of a particular variable. The same results were observed for categorical variables, too.

Vehicle Type, Driver age and seat belt use by drivers have also been studied and it has been found that Driver Age Group (18-20) has the highest involvement in crashes on all road networks. Passenger cars have the highest crash involvement among vehicle types and among all vehicle types; bus drivers have the highest seat belt compliance for all networks.

This research serves as a starting point to demonstrate the use of artificial neural networks to develop crash rate prediction models that could present useful insight to the potential corresponding safety and traffic operation performance.

17 **Key Words** Traffic crash, crash rate, driver age		18 **Distribution Statement** No restrictions. This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161	
19 **Security Classification (of this report)** Unclassified	20 **Security Classification (of this page)** Unclassified	21 **No. of pages** 165	22 **Price**

Form DOT F 1700.7 (8-72)

DATA MINING THE KANSAS TRAFFIC-CRASH DATABASE

Final Report

Prepared by

Yacoub Najjar, Ph.D., P.E.
Srinivas Mandavilli, Ph.D., P.E.

Kansas State University
Department of Civil Engineering
2118 Fiedler Hall
Manhattan, Kansas 66506

A Report on Research Sponsored By

THE KANSAS DEPARTMENT OF TRANSPORTATION
TOPEKA, KANSAS

August 2009

© Copyright 2009, **Kansas Department of Transportation**

PREFACE

The Kansas Department of Transportation's (KDOT) Kansas Transportation Research and New-Developments (K-TRAN) Research Program funded this research project. It is an ongoing, cooperative and comprehensive research program addressing transportation needs of the state of Kansas utilizing academic and research resources from KDOT, Kansas State University and the University of Kansas. Transportation professionals in KDOT and the universities jointly develop the projects included in the research program.

NOTICE

The authors and the state of Kansas do not endorse products or manufacturers. Trade and manufacturers' names appear herein solely because they are considered essential to the object of this report.

This information is available in alternative accessible formats. To obtain an alternative format, contact the Office of Transportation Information, Kansas Department of Transportation, 700 SW Harrison, Topeka, Kansas 66603-3745 or phone (785) 296-3585 (Voice) (TDD).

DISCLAIMER

The contents of this report reflect the views of the authors who are responsible for the facts and accuracy of the data presented herein. The contents do not necessarily reflect the views or the policies of the state of Kansas. This report does not constitute a standard, specification or regulation.

ABSTRACT

Traffic crashes results from the interaction of different parameters which includes highway geometrics, traffic characteristics and human factors. Geometric variables include number of lanes, lane width, median width, shoulder width, roadway section length, and shoulder width while traffic characteristics include AADT, Percentage of Heavy Vehicles and Speed. The effect of these parameters can be correlated by crash prediction models that predict crash rates at particular roadway section.

Transportation Agencies and State Departments of Transportation are continuously faced with decisions concerning the safety of highways. The evaluation and comparison of alternative long-range highway plans should include the safety implications of respective plans. The commonly available models for safety analysis are crash prediction models. By performing an in-depth analysis of crash databases and developing crash rate prediction models, better decisions can be taken in regard to future traffic planning operations.

The main objective of this study is to utilize artificial neural network techniques and develop crash rate prediction models for Kansas road networks. Six networks have been studied and crash prediction models for each network have been developed. Four crash rate categories have been considered in this study. They are:

- Total Crash Rate (TCR) [Injury + Fatal + Property Damage Only]
- Injury Crash Rate (ICR) [Disabling Injury + Possible Injury + Non-Incapacitating Injury]
- Severe Injury Crash Rate (SICR) [Disabling + Fatal]
- Fatal Crash Rate (FCR) [Fatal]

The models developed for each of the road networks are unique and show that geometric variables and traffic have a significant impact on the crash behavior. The models developed in this study would be utilized by Kansas Department of Transportation in evaluating roadway design features, reconstruction impacts and to make decisions in regard to future traffic planning operations. Sensitivity analysis was performed on all the geometric variables in the models. It has been found that all the continuous variables have different effects on different networks. It is very difficult to generalize the behavior of a particular variable. Same results were observed for categorical variables, too.

Vehicle Type, Driver age and seat belt use by drivers have also been studied and it has been found that Driver Age Group (18-20) has the highest involvement in crashes on all road networks. Passenger cars have the highest crash involvement among vehicle types and among all vehicle types; bus drivers have the highest seat belt compliance for all networks.

This research serves as a starting point to demonstrate the use of artificial neural networks to develop crash rate prediction models that could present useful insight to the potential corresponding safety and traffic operation performance.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 - INTRODUCTION AND OBJECTIVES

1.1 Introduction

Transportation safety is a very important issue in United States (U.S) because about 94 percent of all transportation-related fatalities and injuries involve highway motor vehicle crashes (FHWA, 2001). According to National Highway Traffic Safety Association (NHTSA) there were an estimated 5,973,000 police-reported traffic crashes, in which 38,588 people were killed in highway crashes. Also highway crashes were the leading cause of deaths in U.S for ages 2 through 14 in the year 2006 (NHTSA, 2008). The highest price we pay for motor vehicle crashes is in the loss of human lives; however society also bears the brunt of the many costs associated with these crashes. The estimated cost of motor vehicle crashes in 2000 totaled $230.6 billion. It is equal to approximately $820 for every person living in the United States and 2.3% of the U.S. Gross Domestic Product (NHTSA, 2001). The fatality statistics on our highways (Table 1.1) portray the extent of the problem and the immediate need for corrective action.

Driving involves several complex interactions among drivers, vehicles and roadways. Crashes result when there is a conflict between at least two of these elements. Usually crashes are the result of bad decisions made by the driver in the driving environment designed by Traffic Engineers and Transportation Planners. So it is very important to have a thorough understanding of complex relationships in order to reduce the likelihood of conflicts among these elements.

The Transportation Department for every state is continuously faced with decisions concerning the safety of highways. The evaluation and comparison of alternative long-range highway plans for rural and urban areas should include the safety implications of respective plans. The commonly available models for safety analysis are crash prediction models. Crash prediction models are very important as they deal with quantifying the relationship between the crashes observed at a site and the existing traffic and geometric conditions. These prediction models can give us an idea of the important variables and the influence of each in causing crashes at a site.

This research focuses on the large historical traffic crash database maintained by the Kansas Department of Transportation (KDOT), for the state of Kansas. The database has records since 1990. This existing crash database can help transportation engineers in evaluating the safety measures that they undertake. Crash records and their analysis are an essential element in any traffic safety program for several reasons:

- They aid in locating high crash locations on the existing highway system
- The crash experience provides an evaluation of design features
- Efficient planning is based, in part, on traffic volumes and crash rates
- The analysis of crash records may have a direct influence on the budgeting for improvements

Currently, KDOT does not perform an in-depth analysis (i.e., mining) of their traffic-crash historical database. There is a lot of hidden information within this database and in order to efficiently study and extract the information, it is essential to use data mining techniques to mine this crash database. Such an in-depth analysis can be performed by mining the database using statistical and/or artificial neural network (S/ANN) approaches. The employed mining process usually yields new useful correlations between crashes and prevailing traffic and roadway characteristics. ANN's are one of the recently explored advanced technologies which show promise in the area of transportation engineering. However, in contrast to the availability of a large number of successful application demonstrations, it is hard to find studies in the literature that provide systematic examinations of the state-of-the-art, application domains, and the applicability of artificial neural networks in data mining traffic crash databases. Without such an in-depth mining process, these new correlations are most likely to stay hidden within the databases. Therefore, KDOT may never be able to capitalize on the richness of the available traffic-crash historical database in order to make better decisions in regard to future traffic planning operations. The extracted or discovered new correlations could aid KDOT in better understanding the interaction between crashes and prevailing traffic and roadway characteristics. Moreover, availability of such new correlations could aid KDOT in obtaining reliable estimates of anticipated number and/or type of crashes on specific highway system during the coming years.

This information could serve as an early warning if estimates are higher than expected. Consequently, this may allow KDOT to take appropriate actions such as the implementation of various safety measures on specific highway sections in order to reduce the anticipated traffic-related crashes and thereby helping KDOT prepare for things before they actually happen. Since the crash database is huge, there are no clearly defined expectations about the kind of patterns that are lying hidden inside.

1.2 Objectives

Traffic crashes result from the interaction of different parameters which includes highway geometrics, traffic characteristics and human factors. Geometric variables include number of lanes, lane width, median width, shoulder width, roadway section length, and shoulder width while traffic characteristics include AADT, Percentage of Heavy Vehicles and Speed. The effect of these parameters can be correlated by crash prediction models that predict crash rates at particular roadway section. There have been several studies conducted and modeling methodologies adopted in the past related to crash rates and most of the studies had drawbacks and problems with data. This research tries to avoid the drawbacks of earlier studies and utilizes the powerful capabilities of ANN's to data mine the crash database and develop crash rate prediction models.

The main objectives of this research are:

1. Examine the relationships between the occurrence of crashes and related causal factors including traffic and geometric variables and developing crash rate prediction models using Artificial Neural Networks for Kansas road networks.
2. Investigate the relationship between crash rates, driver age, vehicle type and seat belt use using data aggregation.

Additionally, this research is the first in the nation to utilize the Artificial Neural Network (ANN) mining approach to extract new and reliable traffic-crash correlations from historical databases.

Table 1.1: Fatality Statistics - NHTSA

Fatality Rates on U.S Highways						
Year	**Fatal Crashes**	**VMT (Billions)**	**Fatalities/100 Million VMT**	**Fatalities/100,000 Pop.**	**Fatalities/100,000 R.V.**	**Fatalities/100,000 L.D.**
2006	38588	3014	1.41	14.24	16.96	21.03
2005	39252	2989	1.46	14.67	17.71	21.70
2004	38253	2923	1.46	14.52	18.00	21.54
2003	38477	2891	1.48	14.75	18.58	21.86
2002	38491	2856	1.51	14.94	19.06	22.13
2001	37862	2797	1.51	14.80	19.07	22.06
2000	37526	2747	1.53	14.86	19.33	22.00
1999	37140	2691	1.55	15.30	19.61	22.29
1998	37107	2632	1.58	15.36	19.95	22.44
1997	37324	2562	1.64	15.69	20.64	22.99
1996	37494	2486	1.69	15.86	20.86	23.43
1995	37241	2423	1.73	15.91	21.22	23.68
1994	36254	2358	1.73	15.64	21.15	23.21

VMT-Vehicle Miles Traveled, Pop.-Population , R.V.-Registered Vehicles, L.D-Licensed Drivers

CHAPTER 2 - ARTIFICIAL NEURAL NETWORKS

2.1 Introduction

Artificial Neural Networks (ANN's) are computational systems whose architecture and operation are inspired from our knowledge about biological neural cells (neurons) in the brain. ANN's can be described either as mathematical and computational models for non-linear function approximation, data classification, clustering and non-parametric regression or as simulations of the behavior of collections of model biological neurons. These are not simulations of real neurons in the sense that they do not model exactly like the real neuron, but, model several aspects of the information combining and pattern recognition behavior of real neurons in a simple yet meaningful way. Neural network modeling has shown incredible capability for emulation, analysis, prediction, and association.

ANN's can be used in a variety of powerful ways to learn and reproduce rules or operations from given examples; to analyze and generalize from sample facts and make predictions from these; or to memorize characteristics and features of given data and to match or make associations from new data to the old data. ANN's are able to solve difficult problems in a way that resembles human intelligence. The uniqueness of neural networks is their ability to learn by example. Traditional artificial intelligence (AI) solutions rely on symbolic processing of the data, an approach which requires a prior human knowledge about the problem. Also neural networks techniques have an advantage over statistical methods of data classification because they are distribution-free and do not require a prior knowledge about the statistical distributions of the classes in the data sources in order to classify them.

2.2 ANN Architecture-General Features

As ANN's are models of biological neural structures, the starting point for any kind of neural network analysis is a model neuron whose behavior follows closely our understanding of how real neurons work. This model neuron is shown in Figure **2.2** where the neuron (i) has m input vector (x_m) and a single output (y_i). Each input signal is weighted, that is, it is multiplied with the weight value (w) of the corresponding input line (by analogy to the synaptic strength of the connections of real neurons). The neuron will combine these weighted inputs by forming their sum and, with reference to a threshold value and activation function it determines its output.

In mathematical terms, we may describe the neuron by writing Equation 2.1 and 2.2,

$$u = \sum_{i=1}^{N} w_i x_i \qquad \text{Equation 2.1}$$

$$y = f(u - \theta) \qquad \text{Equation 2.2}$$

where $x_1, x_2, x_3, ... x_n$ are the inputs, $w_1, w_2, w_3, ... w_n$ are the synaptic weights, u is the *activation potential* of the neuron, θ is the *threshold*, y is the output signal of the neuron, and $f(.)$ is the *activation function*. For notational convenience, Equations 2.1 and 2.2 may be reformulated into Equations 2.3 and 2.4 by setting $w_0 = \theta$ and $x_0 = -1$.

$$\sum_{i=1}^{N} w_i x_i - \theta = \sum_{i=1}^{N} w_i x_i \qquad \text{Equation 2.3}$$

$$y = f\left(\sum_{i=1}^{N} w_i x_i - \theta\right) \qquad \text{Equation 2.4}$$

The combination of a fixed input $x_0 = -1$ and of an extra input weight $w_0 = \theta$ accounts for what is known as a *bias* input. Note that the new notation has augmented any input vector $x \in \Re^N$ to the vector $(-1, x)^{\in \Re^{N+1}}$ and also the weight vector $w \in \Re^N$ of the neuron, to the vector $(w_0, w)^{\in \Re^{N+1}}$. The activation function, denoted by $f(.)$, defines the output of the neuron in terms of the activity level at its input. The most common form of activation function used in the construction of ANN's is the *sigmoid* function. An example of the sigmoid is the *logistic* function, defined by Equation 2.5,

$$f(u) = \frac{1}{1 + e^{(-au)}} \qquad \text{Equation 2.5}$$

where *a* is the slope parameter of the sigmoid function. By varying the parameter, we can obtain sigmoid functions of different slopes. In the limit, as the slope parameter approaches infinity, the sigmoid function becomes simply a threshold function. The threshold function however, can take only the values 0 or 1, whereas a sigmoid function assumes a continuous range of values from 0 to 1. Also the sigmoid function is differentiable, whereas the threshold function is not. Differentiability is an important feature of neural network theory and has a fundamental role in the learning process in ANN's.

The artificial neural network model is generally referred to as a multi-layer perceptron (MLP). This architecture consists mainly of three types of neuron layers, namely input layer, hidden layer(s) and an output layer. The nodes in an input layer are called input neurons or nodes; they encode the data presented to the network for processing. These nodes do not process the information, but simply distribute the information to other nodes in the next layer. The nodes in the middle layers, not directly visible from input or output, are called hidden nodes. These neurons provide the nonlinearities for the network and compute an internal representation of the data. The nodes in the output layer are referred to as output neurons: they encode possible desired values assigned to the input data. A typical multilayer perceptron neural network model is shown in Figure 2.3.

Neural networks are classified according to structure and learning method employed. The architecture (structure) of neural network represents the connectivity between the various neurons in the structure. The learning methods vary in relation to the way weights of the interconnections are updated during learning process. As for structure, neural networks are chiefly divided into *feedforward* networks and *recurrent* networks

2.3 The Training Technique

In this study a unique training technique has been adopted. The selection of a good network structure is the initial and most essential stage. There are various issues related to network structure that should be predetermined (or pre-adjusted) prior to training including the initial values of connections weights, number of hidden layers, maximum number of iterations allowed, stopping criteria, and minimum and maximum number of hidden nodes in each hidden layer.

The first step in training is to choose the algorithm to be used for training the desired ANN. The most commonly employed is the standard back-propagation algorithm which is the one adapted throughout this study. Before training, the connection weights are set to small random values as explained earlier, including the weights connecting the biases to the hidden and output layers. Then after each training step, a new set of connections is determined. To assess the potential success of this set of connections after each epoch (iteration), some statistical accuracy measures such as the overall *Coefficient of Determination Factor (*or simply *R^2), Mean Average Relative Error (MARE),* and the Averaged-Squared-Error *(ASE)* are evaluated. R^2 factor can be calculated using Equation 2.6.

$$R^2 = \frac{\sum_{j=1}^{J} \frac{\sigma^j_{xy} \times \sigma^j_{xy}}{\sigma^j_{xx} \times \sigma^j_{yy}}}{J} \qquad \text{Equation 2.6}$$

$$\sigma^j_{xy} = \sum_{n=1}^{P} x^j_n y^j_n - \frac{1}{P}\sum_{n=1}^{P} x^j_n . \sum_{n=1}^{P} y^j_n \qquad \text{Equation 2.7}$$

$$\sigma^j_{xx} = \sum_{n=1}^{P} x^j_n x^j_n - \frac{1}{P}\sum_{n=1}^{P} x^j_n . \sum_{n=1}^{P} x^j_n \qquad \text{Equation 2.8}$$

$$\sigma^j_{yy} = \sum_{n=1}^{P} y^j_n y^j_n - \frac{1}{P}\sum_{n=1}^{P} y^j_n . \sum_{n=1}^{P} y^j_n \qquad \text{Equation 2.9}$$

Where, x^j_n refers to the value of the actual (desired) output *j* parameter for pattern number n, while y^j_n refers to the associated ANN prediction. *P* is the total number of provided patterns and *J* is the total number of output parameters. MARE can be calculated by Equation 2.10.

$$\text{MARE}(\%) = \frac{100}{PJ} . \sum_{j=1}^{J} \sum_{n=1}^{P} \left| \frac{y_n^{\ j} - x_n^{\ j}}{x_n^{\ j}} \right| \qquad \text{Equation 2.10}$$

The resulting statistical accuracy measures can be used to assess the level of agreement between predicted and desirable output values. Therefore, they can be used to select the optimal (best performing) network by examines a plot of error versus the number of epochs (iterations). This plot, which is generally called the *"learning curve",* is used to illustrate the training-accuracy history for every trained network. ASE is calculated according to Equation 2.11.

$$ASE = \sum_{j=1}^{J} \sum_{n=1}^{P} (y_n^{\,j} - x_n^{\,j})^2 \qquad \text{Equation 2.11}$$

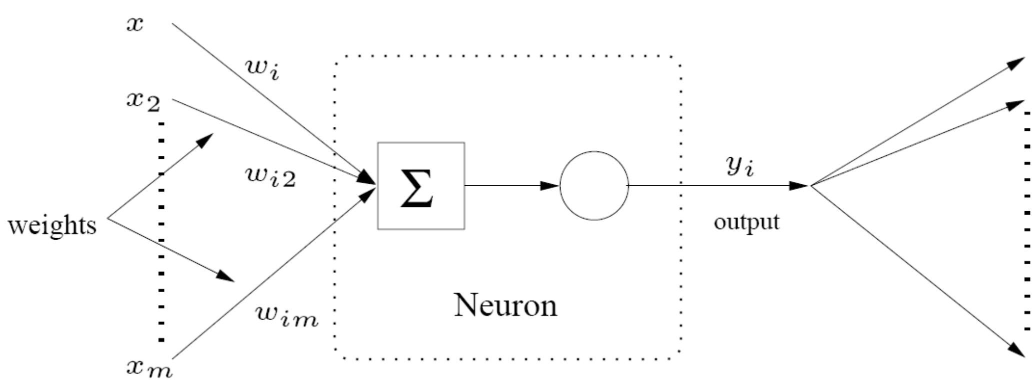


Figure 2.1: Artificial Neuron

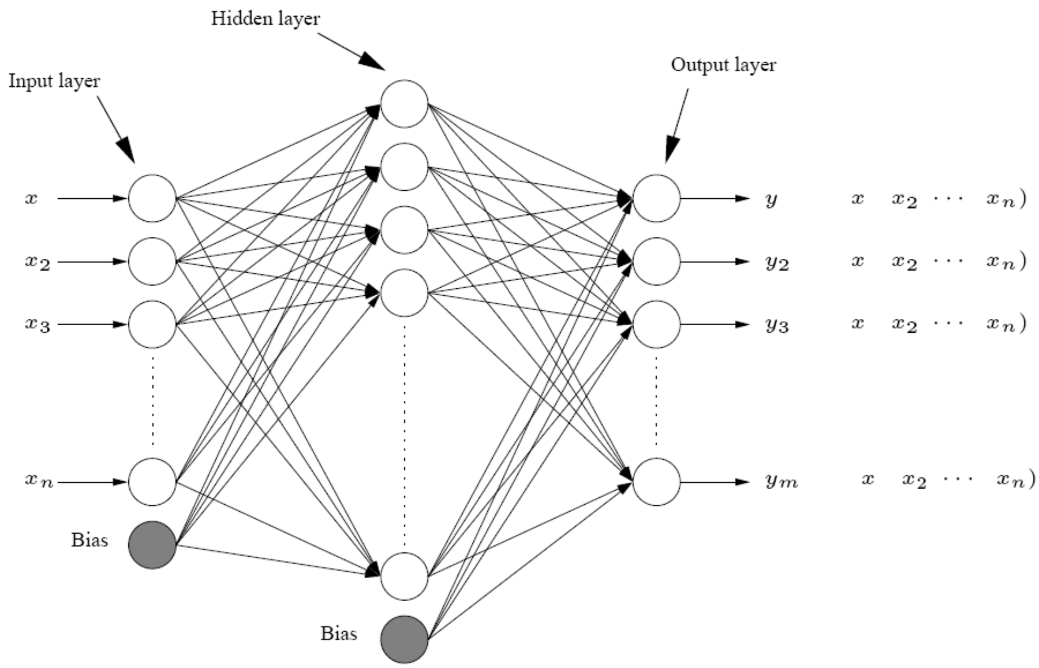


Figure 2.3: Typical Multilayer Perceptron Model

CHAPTER 3 - LITERATURE REVIEW

3.1 Introduction

There have been several studies conducted and modeling methodologies adopted in the past related to crash rates. The following section gives a summary of the studies and the benefits/drawbacks of the modeling methodologies adopted.

3.2 Crash Modeling

There are different types of modeling methodologies. Each modeling methodology is discussed below.

3.2.1 Single and Multivariate Deterministic Models

Single-variate and multivariate deterministic models explored relationships between crashes and the influencing factors. Many of those relationships were qualitative, which incorporated the expert knowledge and past experience. The modeling techniques were relatively primitive.

Zeeger et al., (1994) reviewed highway safety studies and developed relationships between vehicle safety and geometric variables like lane width, shoulder width, and shoulder type and found significant effects of these variables on highway vehicular safety.

Glennon, (1987) studied the effects of alignment and sight distance on highway safety and found that there was no clear effect of improved intersection sight distance on highway safety.

Garber and Ehrhart, (2000) studied the influence of causal factors on the occurrence of crashes and found that speed variance had a positive relationship with crash rates.

The relationships regarding the influence of variables like highway geometric parameters on the crash rates were not consistent due to the omission of influencing variables, lack of sufficient data or the inherent disadvantage of deterministic models. The complexity crashes made researchers use multivariate models in modeling crashes as these models accounted for the influences of multiple factors.

Mohamedshah et al., (1993) developed multivariate linear regression model using data from the Highway Safety Information System (HSIS) and achieved good results.

Garber and Ehrhart, (2000) also developed multivariate deterministic models of highway crashes using significant independent variables like mean speed, standard deviation of speed, flow per lane, lane width, shoulder width and found that speed variance had a significant influence on the crash rate.

3.2.2 Stochastic Models

Stochastic models assume that the occurrence of vehicular crashes is random and showed great potential in obtaining the true models of crashes. Okamoto et al., (1989) suggested that the occurrence of traffic crashes is stochastic. Joshua and Garber, (1992) first developed several Poisson regression models to describe the occurrence of crashes. Various studies further examined the goodness-of-fit of Poisson regression models. (Miaou, (1994), Miaou et al., (1992), Hadi et al., (1993), Miaou and Lum, (1993), Vogt and Bared, (1998), Ivan et al., (1999)). More stochastic models were proposed other than Poisson regression models, which included Zero Inflated Poisson (ZIP) (Miaou, 1994), Negative Binomial (Miaou et al., (1992), Miaou, (1994), Karlaftis and Tarko, (1998), Fridstrom and Ingebrigtsen, (1991)) and Extended Negative

Binomial regression models. While dependent variables in these models are stochastic, the link functions are deterministic. The link functions are used to connect the mean of crash counts with independent variables.

3.2.3 Multiple-Logistic Models

Multiple-Logistic models were also used in crash modeling and were designed to describe the probabilities of count variables. They follow S-shaped curve and the value of dependent variable improves significantly when the independent variable or the function of independent variables of the model reaches certain thresholds.

Joshua and Garber, (1992) used multiple logistic regression models to analyze the relationship between the probability of truck crash involvement and highway geometric and traffic variables.

Lin et al., (1993) applied time dependent logistic regression model to analyze the relationship between safety and truck driver service hours. The logistic regression models developed found that the driving time had a strong influence on safety performance.

3.2.4 Fault Tree Analysis

Fault Tree (FT) Analysis is another modeling approach used. The advantages of fault tree analysis are:

1 It can be used to identify the causal factors of crashes clearly and clarify the whole possible processes;

2 The probability of a crash can also be obtained; and

3 Effective strategies can be provided in accordance with the major and secondary factors.

Joshua and Garber, (1992) performed fault tree analysis to examine the major factors associated with crashes and the interactions among those factors. In a fault tree analysis the outcome (crash) is located at the top of the fault tree and all possible events or actions in different paths leading to the occurrence of crashes were defined as basic events. The probabilities of basic events in the fault tree were assessed from the crash data. The basic events were decided as major factors and the interactions between the major factors were accounted as secondary factors. The probability of a top event was determined through the probabilities of major and secondary factors

3.2.5 Classification and Regression Tree (CART) Analysis

Classification and Regression Tree (CART) Analysis is another modeling technique. CART's are non-parametric procedures for explaining and/or predicting either a categorical or continuous response. Hakkert et al., (1996) used CART as a preliminary tool to explain the relationships between independent variables and road crashes. CART was also used to identify significant variables for further analysis. It is adaptable in dealing with high dimensional and non-homogeneous data set. The tree structure is very helpful to clarify the relationships between independent variables and crash event and interactions among independent variables.

3.2.6 Artificial Intelligence Techniques

Artificial Intelligence Techniques are other modeling technique used. Fuzzy methods, Hybrid techniques and Artificial Neural Networks (ANN) belong to the paradigm of artificial intelligence. These techniques are gaining importance and are widely being applied in highway safety research.

Vaija, (1987) discussed fuzzy methods and applied them in the study of safety. Three different fuzzy methods were discussed in that study: the simple fuzzy expert system, fuzzy linear regression and fuzzified linear programming. Although the study was more about the process control and accident analysis, it was very helpful in the modeling of highway crashes since vagueness is common among all kinds of accident process. Fuzzy methods discussed in their study presented good modeling alternatives.

Hybrid techniques combine merits of different methods. (Awad and Jason, 1998) applied hybrid system using fuzzy logic and neural networks to predict crash frequency. The hybrid system took advantage of the properties and strengths of both fuzzy logic and neural networks. The authors concluded that ANN techniques were good choices in analyzing highway vehicular crashes because simple models could not represent the complex relationships between crashes and causal factors. Apparently, ANN techniques require more training data to obtain satisfactory results.

Artificial neural networks (ANN's) have been successfully applied in several transportation problems. ANN's do not require any prior information and can properly map the input patterns to the output patterns. Najjar et al., (2000) successfully applied ANN's to assess the impact of raising speed limits on Kansas Highways. Ali (2000) applied ANN's for various other Civil Engineering problems and found that ANN models were robust compared to all other conventional statistical models. The power of the neural networks stems from the fact that they try to resemble the capabilities of the human brain. ANN's possess certain features that have advantageous characteristics, which were thoroughly identified by (Haykin, 1994). Some studies (Subba et al., (1998), Faghri and Aneja, (1999)) compared several statistical methods to ANN, and found that the ANN's performance is better in calibration and prediction.

Al-Alwai et al., (1996) used an ANN model to estimate the number of car crashes as a function of several related variables such as population growth and gross domestic product. Results of the ANN were compared to a principle component analysis (PCA) regression technique and the results demonstrated that the ANN model provided better prediction than the PCA model. Mussone et al., (1999) applied artificial neural networks to analyze vehicular crashes that occurred at intersections in Milan, Italy. Their results showed that ANN are capable of extracting information, in terms of the factors that explain crashes and the factors contributing to a higher degree of danger.

Chong et al., (OSU) studied the National Automotive Sampling System General Estimates System automobile accident data from 1995 to 2000 and tried to model the severity of injury resulting from traffic accidents. They utilized trained neural networks using hybrid learning approaches, decision trees and a hybrid approach involving decision trees and neural networks. Their results revealed that in most cases the hybrid decision tree-neural network obtains higher accuracy of prediction than the individual approaches.

Abdelwahab and Aty, (2001) studied the 1997 accident data for the Central Florida area. The research focused on two-vehicle accidents that occurred at signalized intersections. They found out that their MLP neural network with Levenberg-Marquardt algorithm as training algorithm achieved 65.6% and 60.4% classification accuracy for the training and testing phases, respectively.

Yang et al., (1999) used neural network approach to detect safer driving patterns that have less chances of causing death and injury when a car crash occurs. They performed the Cramer's V Coefficient test to identify significant variables that cause injury, therefore, reduced the dimensions of the data for the analysis. Then, they applied data transformation method with a frequency-based scheme to transform categorical codes into numerical values. Using a back propagation neural network and the 1997 Alabama interstate alcohol-related data, they further studied the weights on the trained network to obtain a set of controllable cause variables that are likely causing the injury during a crash.

Dia and Rose, (1997) used 'real-world' data for developing a multi-layered MLP neural network freeway incident detection model. They compared the performance of the neural network model and the incident detection model in operation on Melbourne's freeways. Results showed that neural network model could provide faster and more reliable incident detection over the model that was in operation on Melbourne's freeways.

Ljubiè P. et al, (Slovenia) conducted an analysis of a large database containing UK personal injury traffic accident data. They use data mining techniques and found out some trends and patterns in the dynamic change of the number of traffic accidents during time.

Vogt and Bared, (1998) stated in their literature review that ANN would be a good alternative modeling method to the stochastic regression model.

3.3 Summary of Literature Review

From all the modeling techniques the following is a summary of the capabilities of each modeling technique based on literature review (Garber and Lei, 2001):

1. Single-variate and multivariate deterministic models explored qualitative relationships between crashes and the influencing factors and mainly incorporated the expert knowledge and past experience. Hence the modeling techniques were relatively primitive.
2. Stochastic regression models showed great potential in obtaining the true models of crashes. Though stochastic regression models were applied and their theoretical disadvantages compensated, past research did not pay enough attention to traffic flow parameters and the used data were subject to both sampling and non-sampling errors. The resulting models were of limited context without widespread applications to support their credibility. Hence the success of these models is still futuristic.
3. Fault tree analysis can clearly identify the causal factors and the whole process of crashes, but it is not feasible in large-scale data modeling. Fault tree analysis needs the probabilities of the occurrence of a certain number of crashes caused by various single factors which is hard to be determined because of the difficulty to separate the influences of different factors.
4. CART can be used preliminarily to analyze the relationships between independent variables and crashes and identify critical independent variables to be included in the models.
5. Fuzzy methods are consistent with the characteristics of crashes.

CHAPTER 4 - CRASH RATES FOR KANSAS ROAD NETWORKS

4.1 Introduction

It is a standard practice to review historical crash data for analysis of trends and documentation of probable causes of highway crashes. Crash rate trends are an effective tool to measure safety hazards on highways as they combine crash frequency with vehicle exposure. The main aim of this chapter is to look at the crash rate trends for Kansas Road Networks.

4.2 Database Development

The first step in this research is the database development. This is the most important phase of the entire research as quality of data controls the efficiency of the models being developed. Every crash has many attributes and detailed information for each attribute is stored in a separate database. An important feature of this research is that it interconnects databases that have not previously been used together systematically but that, when fused, create a rich environment for developing crash prediction models.

The main source of data is the Kansas Accident Recording System (KARS). KARS connects several smaller databases and forms datasets which contain general roadway information (e.g., lane width, surface type, and shoulder type) for each roadway. Crash records include data on vehicles, drivers, roadway conditions, and the severity of the crashes. Traffic Volume information is obtained from High Accident

Location System (HALS) database. KARS and HALS databases are combined to give the final datasets used for predicting crash rates on the Kansas Road Networks.

4.2.1 Kansas Accident Recording System (KARS)

KARS is a comprehensive database maintained by the KDOT having crash records since 1990. Only the police reported crashes are included in this system. Every crash in the system is identified by a unique accident key. KARS connects all the smaller databases based on this unique key to give detailed information for every crash.

4.2.2 High Accident Location System (HALS)

This database is mainly used for identifying the high accident locations for the state of Kansas. This database reflects the most up to date information on the traffic flow for each roadway section. To capitalize on latest traffic flow information for each of the sections, used in this research, this database is used.

4.3 Road Networks

The entire Kansas road network is divided into two major categories Rural and Urban. Within each category the roads are further classified into different types. A total of six road networks have been formed using the database. The networks are as follows:

1. Rural
    - Rural 2-lane (R2L)
    - Rural Expressways (RE)
    - Rural Freeways (RF)
    - Rural KTA (RKTA)
2. Urban

- Urban Freeways. (UF)
- Urban Expressways (UE)

The KDOT traffic-crash historical database (KARS) was be used to compile a yearly breakdown of the following crash-related categories:

i. Total Crash Rate (TCR) [Injury + Fatal+ Property Damage Only]
ii. Injury Crash Rate (ICR) [Disabling Injury + Possible Injury + Non-Incapacitating Injury]
iii. Severe Injury Crash Rate (SICR) [Disabling Injury + Fatal]
iv. Fatal Crash Rate (FCR) [Fatal]

4.4 Limitations and Definitions Adopted

The limitations and definitions adopted in this research are as follows:

- Only the state system roads have been included in this study because crashes occurring on the state system have accurate and up to date information on the traffic volumes and roadway geometric characteristics. All other roads do not have the necessary information. Without accurate information on traffic volumes it would not be possible to calculate crash rates. In order to have quality data sets for modeling, all the non-state system roads have been excluded.
- All the sections that have undergone major reconstruction changes have been excluded. This way only sections that have remained constant throughout the study period are present in the data. This again is a check for quality data.

- The entire road network is divided into smaller sections. So every accident is associated with a section. Accidents at the junction of two sections are always associated with the subsequent section (i.e. section starting next, in the direction of travel).
- In this research the following definitions have been adopted in order to identify freeways and expressways:
- **Freeways:** Multi-lane, divided arterials with full access control at grade-separated interchanges
- **Expressways:** Multi-lane, divided arterials with access limited primarily to grade-separations and at grade intersections.
- Both **Freeways** and **Expressways** are minimum two-lane facilities with posted speed limit >=45 mph.

4.5 Logic Used for Identifying Crashes

Four primary crash categories have been included in this study. There are three types of injury crashes in the database. They are Disabling Injury Crashes, Non-Incapacitating Injury Crashes and Possible Injury Crashes. All these three categories have been summed into one category- the Injury crashes. The Disabling Injury Crashes are summed up with the Fatal Crashes and are called Severe Injury Crashes. The Injury, Non-Injury and Fatal Crashes are summed into one category-Total Crashes.

The following logic has been used to identify the crashes while querying the database:

- **Fatal Crash**: where # of Fatalities >0
- **Disabling Injury crash**: where # of Fatalities =0 <u>and</u> # of Disabled >0

- **Non-Incapacitating Injury crash**: where # of Fatalities =0 and # of Disabled =0 and # of Non-incapacitating Injury >0
- **Possible Injury crash**: where # of Fatalities =0 and # of Disabled =0 and # of Non-incapacitating Injury =0 and # of Possible Injury >0
- **Sever Injury crash**: where # of Fatalities >0 OR # of Disabled >0

4.6 Crash Rates

All crash rates except fatal crash rate are expressed per million vehicle miles of travel. The fatal crash rate is expressed per 100 million vehicle miles of travel. The formulae used for calculating the crash rates are given in Equations 4.1, 4.2, 4.3 and 4.4.

Total Crash Rate (TCR) [Injury + Fatal+ PDO]

$$\text{TCR} = \left(\frac{1{,}000{,}000 * \text{Total Crashes}}{\Sigma(\text{AADT*Section Length* } 365)}\right) \qquad \text{Equation 4.1}$$

Injury Crash Rate (ICR) [Disabling + Possible + Non-Incapacitating]

$$\text{ICR} = \left(\frac{1{,}000{,}000 * \text{Injury Crashes}}{\Sigma(\text{AADT*Section Length* } 365)}\right) \qquad \text{Equation 4.2}$$

Severe Injury Crash Rate (SICR) [Disabling + Fatal]

$$\text{SICR} = \left(\frac{1{,}000{,}000 * \text{SevereInjury Crashes}}{\Sigma(\text{AADT*Section Length* } 365)}\right) \qquad \text{Equation 4.3}$$

Fatal Crash Rate (FCR) [Fatal]

$$FCR = \left(\frac{100{,}000{,}000 * \text{Fatal Crashes}}{\Sigma(\text{AADT*Section Length*365})} \right) \qquad \text{Equation 4.4}$$

The crash rates calculated using above equations are tabulated by network type and shown in Table 4.**1**. Table 4.2 shows the results by crash rate type. The results shown in these tables have been plotted (See Figures 4.3, 4.4 and 4.5) to clearly depict the trends in crash rates by network type and crash rate type. These trends would give a clear picture of trends on each network over the years and also would serve as a tool for preliminary estimate of crash rates for future years on a particular road network. The average crash rate values for all networks are given in Table 4.3.

4.7 Discussion of Results

The crash rate trends shown in Figures 4.3 and 4.4 give us an idea of crash rates for the rural and urban networks. The following are the observations:

- For all the networks except Urban Expressways it can be seen that the Total Crash Rate shows an increasing trend. All the other crash rates remain fairly constant all throughout.
- In Figure 4.5 it is slightly difficult to generalize the crash trends for each network. Hence best fit trend lines have been plotted for all networks and are shown in Figure 4.6 Figure 4.1.
- Figure 4.6 shows the plots by roadway category. The observations from the trends lines are:
    - o For all networks the TCR increases up to 1997. For all the rural networks the TCR remains constant from 1997 to 2001. For the

Urban networks the TCR decreases for urban expressways from 1997 to 2001 and remains constant for urban freeways.

- ICR increases slightly for Rural KTA and remains constant for the remaining rural networks and urban freeways. In case of urban expressways it remains constant until 1997 and then decreases.
- SICR and FCR remain constant for all road networks. It is hard to predict the trends from this figure. Figure 4.1 zooms on the crash rates and would give a clearer picture for SICR and FCR.

- Figure 4.1 shows the plots by crash rate type. The observations from the trends lines are:
  - For TCR, all the road networks show an increasing trend up to 1997 and then remain constant.
  - For Rural KTA ICR increases up to 1998 and then decreases slightly. For Urban Expressway ICR increases up to 1995 and then decreases. For all other networks ICR remains constant up to 1997 and then decreases slightly
  - SICR and FCR do not have any consistent pattern. The reason for this behavior is due to limited number of datasets. Though we see increasing decreasing patterns for the networks the amount of increase or decrease is very small. We can say that the there is no definite trend or they remain fairly constant
- Figure 4. shows the histogram for average crash rates for all road networks. The observations from the histogram are:

- For TCR, Rural 2 Lane has the highest average TCR in the rural network and Urban Expressways has the highest average TCR in the urban network. For all the networks combined Urban Expressways has the highest average TCR.
- For ICR, Rural 2 Lane and Rural KTA have the highest average ICR in the rural network and Urban Expressways has the highest average TCR in the urban network. For all the networks combined Urban Expressways has the highest average ICR.
- For SICR, Rural 2 Lane has the highest average SICR in the rural network and Urban Expressways has the highest average SICR in the urban network. For all the networks combined Urban Expressways has the highest average SICR
- For FCR, Rural 2 Lane has the highest average FCR in the rural network and Urban Expressways has the highest average FCR in the urban network. For all the networks combined Rural 2 Lane has the highest average FCR.

4.8 Conclusions

The following conclusions can be deduced based on earlier discussion.

- Based on observations from Figures 4.6, 4.5 and 4.6**,** it can be said that TCR for all road networks increases up to 1997 and then remains constant, ICR/SICR/FCR do not have a common trend for the entire network as a whole.

- Among the Rural Networks Rural 2 Lane has highest crash rates and among Urban Networks Urban Expressways has the highest crash rates.
- The crash rate trends are a good starting point for data mining the crash database as we have some preliminary values to expect.

Table 4.1: Crash rates by network type

Rate Type	Rural KTA										
Year-------------------->	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	Average
Total Crash Rate	0.491	1.017	0.798	0.883	0.993	0.952	1.125	0.999	0.943	0.951	**0.915**
Injury Crash Rate	0.132	0.244	0.222	0.242	0.286	0.260	0.294	0.269	0.251	0.254	**0.245**
Severe Injury Crash Rate	0.029	0.043	0.055	0.027	0.050	0.043	0.047	0.060	0.046	0.052	**0.045**
Fatal Crash Rate	0.005	0.008	0.007	0.004	0.007	0.001	0.001	0.012	0.007	0.005	**0.006**
	Rural 2 Lane										
Total Crash Rate	1.149	1.359	1.317	1.430	1.469	1.560	1.618	1.558	1.458	1.543	**1.446**
Injury Crash Rate	0.245	0.276	0.250	0.256	0.268	0.257	0.260	0.236	0.221	0.223	**0.249**
Severe Injury Crash Rate	0.088	0.092	0.081	0.086	0.092	0.088	0.088	0.084	0.075	0.075	**0.085**
Fatal Crash Rate	0.021	0.025	0.021	0.022	0.025	0.023	0.026	0.024	0.021	0.024	**0.023**
	Rural Expressways										
Total Crash Rate	0.553	0.632	0.682	0.929	1.135	0.637	0.806	0.834	0.821	0.930	**0.796**
Injury Crash Rate	0.150	0.157	0.163	0.196	0.277	0.123	0.164	0.165	0.150	0.162	**0.171**
Severe Injury Crash Rate	0.063	0.052	0.059	0.046	0.074	0.043	0.057	0.058	0.046	0.068	**0.057**
Fatal Crash Rate	0.014	0.003	0.014	0.008	0.005	0.011	0.016	0.013	0.015	0.022	**0.012**
	Rural Freeways										
Total Crash Rate	0.523	0.794	0.616	0.720	0.713	0.790	0.820	0.714	0.699	0.699	**0.709**
Injury Crash Rate	0.105	0.145	0.125	0.146	0.145	0.142	0.148	0.136	0.136	0.124	**0.135**
Severe Injury Crash Rate	0.034	0.044	0.052	0.045	0.044	0.042	0.044	0.042	0.041	0.039	**0.043**
Fatal Crash Rate	0.006	0.007	0.009	0.010	0.007	0.007	0.006	0.009	0.009	0.008	**0.008**
	Urban Expressways										
Total Crash Rate	1.502	1.630	1.481	1.547	1.744	1.822	1.925	1.778	1.525	1.556	**1.651**
Injury Crash Rate	0.364	0.438	0.417	0.415	0.486	0.448	0.428	0.334	0.331	0.230	**0.389**
Severe Injury Crash Rate	0.118	0.099	0.076	0.096	0.129	0.107	0.101	0.071	0.063	0.048	**0.091**
Fatal Crash Rate	0.013	0.025	0.006	0.013	0.030	0.020	0.010	0.031	0.018	0.009	**0.017**
	Urban Freeways										
Total Crash Rate	0.842	1.120	0.919	1.101	1.197	1.168	1.225	1.208	1.121	1.163	**1.106**
Injury Crash Rate	0.179	0.215	0.202	0.226	0.256	0.222	0.226	0.208	0.195	0.182	**0.211**
Severe Injury Crash Rate	0.034	0.046	0.041	0.047	0.050	0.045	0.049	0.047	0.037	0.033	**0.043**
Fatal Crash Rate	0.002	0.004	0.005	0.005	0.006	0.008	0.006	0.005	0.004	0.006	**0.005**

Table 4.2: Crash rates for each network by crash rate type

	Total Crash Rate										
	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	Average
Rural KTA	0.491	1.017	0.798	0.883	0.993	0.952	1.125	0.999	0.943	0.951	**0.915**
Rural 2 Lane	1.149	1.359	1.317	1.430	1.469	1.560	1.618	1.558	1.458	1.543	**1.446**
Rural Expressways	0.553	0.632	0.682	0.929	1.135	0.637	0.806	0.834	0.821	0.930	**0.796**
Rural Freeways	0.523	0.794	0.616	0.720	0.713	0.790	0.820	0.714	0.699	0.699	**0.709**
Urban Expressways	1.502	1.630	1.481	1.547	1.744	1.822	1.925	1.778	1.525	1.556	**1.651**
Urban Freeways	0.842	1.120	0.919	1.101	1.197	1.168	1.225	1.208	1.121	1.163	**1.106**
	Injury Crash Rate										
	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	Average
Rural KTA	0.132	0.244	0.222	0.242	0.286	0.260	0.294	0.269	0.251	0.254	**0.245**
Rural 2 Lane	0.245	0.276	0.250	0.256	0.268	0.257	0.260	0.236	0.221	0.223	**0.249**
Rural Expressways	0.150	0.157	0.163	0.196	0.277	0.123	0.164	0.165	0.150	0.162	**0.171**
Rural Freeways	0.105	0.145	0.125	0.146	0.145	0.142	0.148	0.136	0.136	0.124	**0.135**
Urban Expressways	0.364	0.438	0.417	0.415	0.486	0.448	0.428	0.334	0.331	0.230	**0.389**
Urban Freeways	0.179	0.215	0.202	0.226	0.256	0.222	0.226	0.208	0.195	0.182	**0.211**
	Severe Injury Crash Rate										
	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	Average
Rural KTA	0.029	0.043	0.055	0.027	0.050	0.043	0.047	0.060	0.046	0.052	**0.045**
Rural 2 Lane	0.088	0.092	0.081	0.086	0.092	0.088	0.088	0.084	0.075	0.075	**0.085**
Rural Expressways	0.063	0.052	0.059	0.046	0.074	0.043	0.057	0.058	0.046	0.068	**0.057**
Rural Freeways	0.034	0.044	0.052	0.045	0.044	0.042	0.044	0.042	0.041	0.039	**0.043**
Urban Expressways	0.063	0.052	0.059	0.046	0.074	0.043	0.057	0.058	0.046	0.068	**0.057**
Urban Freeways	0.034	0.046	0.041	0.047	0.050	0.045	0.049	0.047	0.037	0.033	**0.043**
	Fatal Crash Rate										
	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	Average
Rural KTA	0.005	0.008	0.007	0.004	0.007	0.001	0.001	0.012	0.007	0.005	**0.006**
Rural 2 Lane	0.021	0.025	0.021	0.022	0.025	0.023	0.026	0.024	0.021	0.024	**0.023**
Rural Expressways	0.014	0.003	0.014	0.008	0.005	0.011	0.016	0.013	0.015	0.022	**0.012**
Rural Freeways	0.006	0.007	0.009	0.010	0.007	0.007	0.006	0.009	0.009	0.008	**0.008**
Urban Expressways	0.013	0.025	0.006	0.013	0.030	0.020	0.010	0.031	0.018	0.009	**0.017**
Urban Freeways	0.002	0.004	0.005	0.005	0.006	0.008	0.006	0.005	0.004	0.006	**0.005**

Table 4.3:Average crash rate values for all road networks

	Total	Injury	Severe. Injury	Fatal
Rural KTA	0.9152	0.2454	0.0452	0.0057
Rural 2 Lane	1.4461	0.2492	0.0849	0.0232
Rural Expressways	0.7959	0.1707	0.0566	0.0121
Rural Freeways	0.7088	0.1352	0.0427	0.0078
Urban Expressways	1.651	0.3891	0.0908	0.0175
Urban Freeways	1.1064	0.2111	0.0429	0.0051

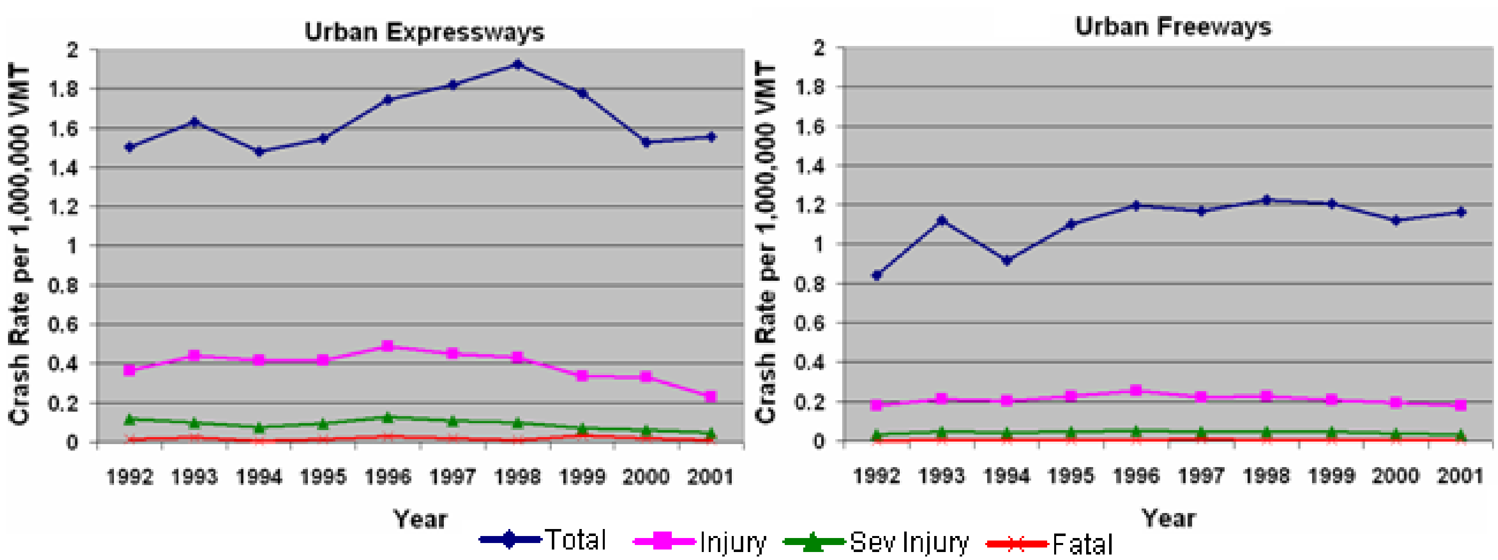


Figure 4.1: Crash rates for the urban network

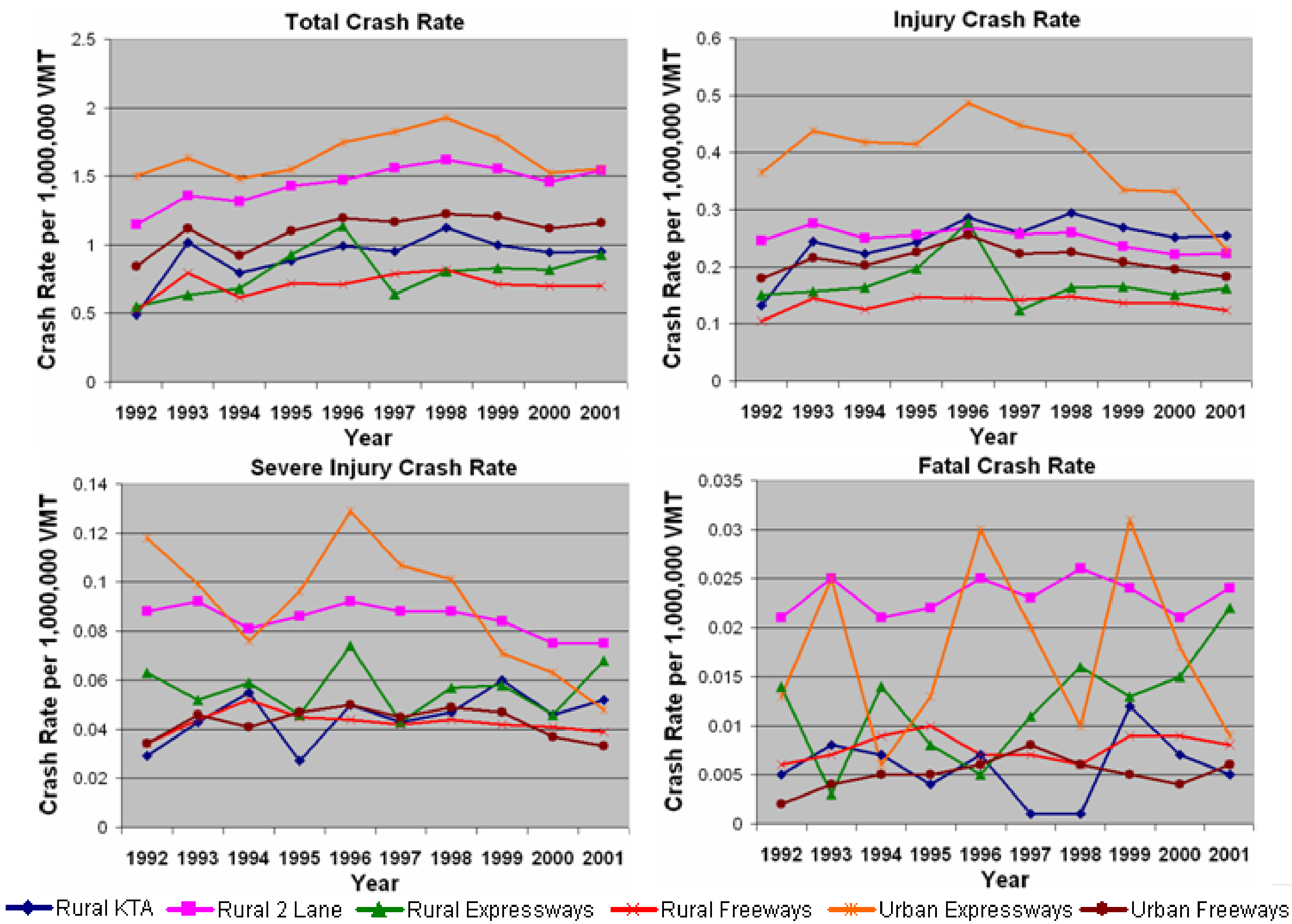


Figure 4.2: Crash rates for all networks by crash rate type

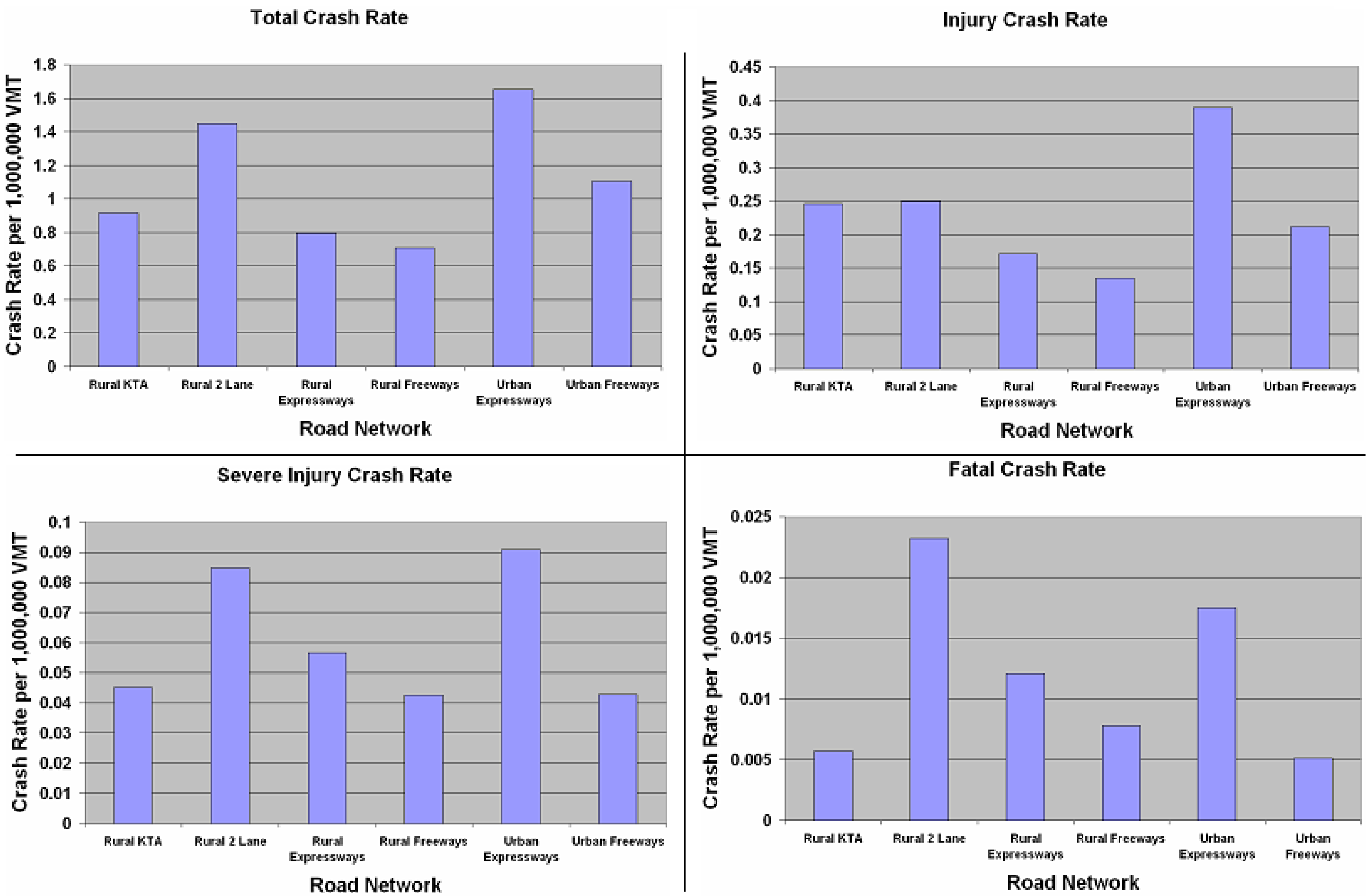


Figure 4.3: Average crash rates for all road networks

CHAPTER 5 - CRASH MODELING AND SENSITIVITY ANALYSIS

5.1 Introduction

The modeling methodology adopted for all networks considered in this study, is basically the same. Hence, only one network will be discussed in detail. For remaining networks, only the modeling results will be discussed. This chapter will concentrate on the details associated with the modeling process and its outcomes in relation to the Rural Expressways Network.

5.2 Rural Expressways Network

The KARS database is queried to separate all the accidents related to this network. The entire network is made up of small sections. Crash rates for each section are calculated and are used in modeling. The input variables consisted of two types (continuous variables and categorical variables). Table 5.2 shows all the input variables for the Rural Expressway network. The continuous variables have a numerical value and the categorical variables have a binary code. The input variables include geometric variables, traffic volume, speed and Heavy Vehicle (%).

5.3 Distributions of Variables

Histograms were plotted to give the distribution of categorical and continuous variables. The histograms are shown in Figures 5.1 and 5.2 With the help of these distributions, we would know the percentage representing each of the variables in the entire dataset. Each of categorical variables is made up of different types. A brief description of the different types is given below. It is to be noted that only the variables used in the final model are discussed herein.

5.3.1 Categorical Variables

There are four types of categorical variables.

1. Route Class

The classification of roadways into different categories (classes) is one of the basic steps of a sound planning process. The most obvious reason to stratify and distinguish between roads is the available funds, which usually are not sufficient to provide equally good service on all facilities. Federal funding eligibility is tied to a road's functional classification, thereby, increasing the need for classification. The KDOT classifies its roads into five classes. The higher classes account for a smaller portion of the total mileage but the greater part of the vehicle miles of travel. The guidelines for classification were based on national averages and do not necessarily meet the needs of individual states. The Route Class variable determines the class of the roadway. In all there are five classes of roads in the entire Kansas database. A Brief description of each class is given below:

a. **Route Class A** coincides with the Interstate System (National System of Interstate and Defense Highways). As its name implies, its primary civilian function is to serve interstate travel. This class occupies the highest position compared to any other class and serves large number of vehicles traveling long distances. The design and operations of this class is much superior compared to other classes.

b. **Route Class B** along with Route Class A serves the most important corridors of statewide and interstate auto as well as heavy truck travel. Nearly all cities with a population over 10,000 are within ten miles of these routes. Incorporation of a facility as a Route Class B implies a commitment to the high standards throughout its length, regardless of the possibility of lower volumes on part of the route.

c. **Route Class C** is also an important part of the statewide arterial system and is integrated with Classes A and B to provide efficient service to all areas of the State. While some interstate travel occurs on the Class C Routes, their principal function is to provide person and commodity movement between regions of the State. Because of a high number of commuter trips, some of the Class C Routes which radiate out from the metropolitan areas are among the highest volume routes in the State. Even though local traffic makes up a larger portion of the traffic stream, these are also important arterial routes. The higher volume segments of Class C Routes have the same design standards as Class B Routes.

d. **Route Class D** contains routes that serve both inter-county movement as well as routes that provide access to the arterial routes for county seats and other small urban areas that are not on a Class A, B or C Route. The importance of these routes for commercial service is related primarily to the small communities they serve. The percent of trucks on these routes should generally be low, but specific sections serving commercial traffic generators may carry a large number of trucks. Almost no interstate mobility is provided, except as access connectors to an interstate route; however, there may be trips to local industries involved in interstate activities. Many Class D state highways serve the same function as roads maintained by counties.

e. **Route Class E** is made up of stubs and routes whose service is limited almost exclusively to local travel. Truck traffic typically makes up a small percent of the total number of vehicles, but may be significantly higher if there is local industry generating large numbers of trips. The average trip length is generally short but may vary widely, depending on the nature of the local area served. The Class E system of roads generally serves traffic which originates from county or township roads.

Figure 5.1a shows the distribution plot for Route Class. It can be seen that Route Class C (48%) and Route Class B (40%) constitute the major portion of the Rural Expressway Network. Route Class D (9%) and Route Class E (3%) combined represent 12% of the network. This network does not have Class A roads. The network serves important corridors of statewide and interstate auto as well as heavy truck travel, via its Routes B and C.

2. Median Type

Rural multi-lane expressways include some type of median treatment. This median could be a variety of types, such as depressed median, raised median with turn lanes, or barrier. The medians usually are non-traversable type but in some situations as in the case of at-grade intersections, painted medians are traversable. Figure 5.1b shows the distribution plot for Median Types. Depressed Median or Median Type 6 (64%) constitutes the major portion of this network. This is followed by Raised Median + Turning Lanes or Median Type 4 (30%) and then followed by Barrier Median or Median Type 3 (6%).

For multi-lane expressways in most rural environments, a depressed median is the most preferred median treatment which is reflected from the histogram plot in Figure 5.1b. The depressed median allows flexibility on running independent grades, while providing a larger separation between travel directions. It is generally used on rural multi-lane expressways, where right of way is available. From literature review it can be seen that the depressed medians are much safer compared to other medians.

The raised median with turn lanes is usually preferred in rural developed areas such as rural communities and development centers. The presence of a raised median with turn lanes increases the chance for crashes because of vehicles slowing down to make left turns.

Barrier medians are provided in cases where there are right of way restrictions. Placing a barrier median largely eliminates the severe cross-median accidents. At the same time, the barrier becomes the target of collisions that would otherwise not occur. It will cause additional crashes by deflecting vehicles back into the traffic stream. In addition, for narrow medians, the barrier seems to cause increases in speed in the median lane, and changes in vehicle placement, which reduces the clearance between parallel streams. There are positive and negative effects associated with this type of median. The overall effect is dependent on the prevailing conditions and therefore varies from place to place.

3. Shoulder Type Inside

Roadways are provided with different types of shoulders and the mechanisms by which they impact safety are very diverse. Depending on the type of Shoulder they serve different purposes. For example gravel shoulders alert the stray driver, paved shoulders allow drivers to regain loss of control etc. Shoulder Type Inside is provided on the left side of the roadway. Figure 5.1c shows the distribution. Bituminous Base (Shoulder Type 11) constitutes a major portion (61%) of the network. The remainder sections have Curb and Gutter (Shoulder Type 19, 14%), Portland Cement Concrete

(P.C.C.) Shoulder (Shoulder Type 29, 6%) and Aggregate Base Stabilized (A.B.S) Shoulder (Shoulder Type 5, 19%)

4. Shoulder Type Outside

These also serve the same purpose as inside shoulders. Shoulder Type Outside is provided on the right side of the roadway. Figure 5.1d shows the distribution. Outside Curb and Gutter shoulders (Shoulder Type 19) constitutes a major portion (33%) of the network while other types [Bituminous Base (20%), Portland Cement Concrete (P.C.C.) (20%), and Aggregate Base Stabilized (A.B.S) (6%)] make for the remaining significant parts (i.e., each represents at least 5% of the total).

5.3.2 Continuous Variables

There are five types of continuous variables.

1. Surface Width

This variable is used as alternative for the number of lanes. Figure 5.2a shows the distribution plot for Surface Width. A majority of the network has a surface width of 24 feet (70%). The remainder of the network has surface widths 29 feet (18%), 27feet (7%), 26 feet (3%) and 28 feet (2%).

2. AADT

Figure 5.2b shows the distribution plot for AADT. 50% of the network sections have an AADT of (<=3000 vehicles), 29% of them have AADT between 3000 and 6000 vehicles and the remainder 21% have AADT between 6000 and 9000 vehicles.

3. Heavy Vehicles

Figure 5.2c shows the distribution plot for percentage of Heavy Vehicles (%HV) As can be noted, 49% of the sections have %HV between 5% and 10%, while 20% of the sections have %HV between 10% and 15%. 17% of the sections have %HV value of less than 5%. The remainder (14%) sections have HV percentages between 16% and 30% as indicated in Figure 5.2c.

4. Speed limit

Figure 5.2d depicts the distribution plot for Speed Limit. In this network, the 5-year average speed limits varied from 45 mph to 70 mph. The predominant speed limits were 55 mph and 65 mph, each accounting for almost 20% of the network's sections. Please note that the speed limits have been averaged over a 5 year period and hence we may have values that are not divisible by 5.

5. Median width

Figure 5.2e shows the distribution plot for the Median Width. In this case, the widths varied from 5 feet to 30 feet. The predominant width is 30 feet, which was used on about 27% of network's sections. A notable number of sections were built with median widths of 9, 15, 18 and 7 feet. Those sections represent about 50% of the networks' total sections. The remaining section (representing less than 5% of the network total) used various median widths ranging from 5 to 25 feet.

5.4 Training Methodology

Crash rates were calculated for each section in the network and then averaged over two time-frame periods. The first time-frame period is from 1992 to 1995, while the second time-frame period uses the data from 1996 to 2001. This procedure yielded two datasets for each section, one for each 5-year period. In so doing, for this network, this resulted in 49 sections and 98 datasets.

The maximum and minimum values of all the continuous variables in the input file were found. This was done to get the ranges for each variable. Based on the above values the maximum and minimum values for the neural network being trained are established. The categorical variables have binary values as their maximum and minimum. Once the values are fixed, the network makes sure that the limits are not exceeded. Table 5.2 gives the maximum and minimum values of all input variables. The entire data is then divided into training, testing and validation data sets. Care has been taken to include correct exemplars in the training, testing and validation data sets to ensure that the developed model will be used in an interpolation mode. Training utilized 49 datasets, testing and validation utilized 25 and 24 datasets each. All input/output parameters were normalized to numerical values between 0 and +1.0. This was done to improve generalization and reduce the large numerical variability among the values of input (or output) parameters. If this normalization is not performed, it could significantly affect the weight adjustment process during training, hence affecting the prediction accuracy of the network. The models were trained to map an input vector of dimensionality forty five (45) into an output vector of dimensionality four (i.e., a mapping from R^{45} to R^4, where 45 is the total initial number of input variables). The data is trained using a Multilayer Perceptron (MLP) neural network with standard back-propagation algorithm. The MLP program used in this research is TR-SEQ developed by Najjar and Ali (1998). The chosen MLP has an Input Layer, One Hidden Layer and an Output Layer. The basic schematic is shown in Figure 5.3

Before training, the connection weights are set to small random values as explained earlier, including the weights connecting the biases to the hidden and output layers. Then after each training step, a new set of connections is determined.

The model building process started with a full model that uses all input variables shown in Table 5.2. Then, reduced, but more efficient, models were obtained using a selected set of variables. The number of input nodes started at 45 nodes, representing the 45 input variables listed in Table 5.2. Iterations were carried out from 1 to 1000 with hidden nodes ranging from 1 to 10.

The structure of the MLP needs to be fixed using this iterative process. The first training helps us in determining the maximum number of hidden nodes and the iterations needed. To assess the potential success of this set of connections after each epoch (iteration), statistical accuracy measures such as the overall *Coefficient of Determination Factor* (or simply R^2), *Mean Average Relative Error (MARE),* and the *Average Squared Error (ASE)* are evaluated. Table 5.3 gives the results of the first training. Only the training datasets are used during this process. The testing datasets were then run to see how the model predicts them.

The best network was obtained at 800 iterations and with 3 hidden nodes. If we examine Table 5.3, it can be noted that the training R^2 (0.96986), and the testing R^2 (0.0127) are far apart. The *(MARE)* and the *(ASE)* values for training and testing are also not close. This suggests that the underlying logic from the training datasets has been well captured by the MLP network, but there is an additional logic associated with the testing datasets. In order to re-check this, a test run is made using the validation datasets to see how the network performed. The results are listed in Table 5.4. The results suggest that the network did not do well on the validation datasets. In order to extract the complete logic present in the database, a final run using all datasets (training, testing and validation) is performed and the connection weights are updated. The starting nodes, maximum number of hidden nodes and iterations are not altered in this process. This procedure allows us to extract extra logic without overtraining. Table 5.5 presents the final overall model performance results.

After this step, the maximum hidden nodes, maximum iterations and connection weights are fixed for the MLP. At this stage, all input variables are present. In the next phase some variables are dropped from the model and the network is re-trained. This re-training is performed to ensure that only significant variables are retained in the final model. The first variable to be dropped was the Subsection (SS) associated parameter.

From Table 5.6, it can be seen that when the Subsection (SS) parameter is dropped, an increase in R^2, a decrease in *MARE* and in *ASE are noted*. Hence, SS is dropped out of the model. Accordingly, the same process is repeated with other input variables as noted in Table 5.7 and Table 5.8. In each process, one variable is eliminated. This process, is discontinued when the *MARE* and the *ASE* could not be improved any further. As noted in Table 5.9, dropping any additional variables would negatively impact the *MARE,* and the *ASE values*. Hence, the input-variable elimination process is stopped. The final network structure, after dropping three variables (i.e., Subsection, Shoulder Width Outside, and Shoulder Width Inside), is given in Table 5.10.

Following this cycle of variable elimination and importance evaluation, all needed input variables are fixed. The final model was selected based on the highest accuracy on the testing and validation datasets. In all, 21 variables will be included in the final

ANN-based model. Subsequently, the entire process of training, testing and validation to obtain the final network is activated. The difference between training at this stage from that at the initial training stage is that, the network's optimal hidden nodes are known. The training is done to re-calculate the connections weights and improve the accuracy of the network. When some of the input variables are dropped, the "best network" obtained from initial training is no longer the "best". Hence the entire process is repeated with the new input layer. Tables 5.11, 5.612 and 5.713 list the results obtained. Table 5.14 gives the final list of input variables included in the final model.

The best overall network was obtained at 900 iterations and 3 hidden nodes. Hence the MLP final model has 21 input nodes, 3 hidden nodes, and 4 output nodes. The output layer of the MLP model consists of four neurons, representing the four crash rates. Figure 5.4 shows the structure of the final MLP neural network model. The structure of final model can be expressed in the following mathematical form (i.e., Equation 5.1).

$$(Y_{1,}\ Y_2, Y_3, Y_4) = ANN_{21\text{-}3\text{-}4}(X_1, X_2, X_3, X_4\ \ldots X_{21}) \qquad \text{Equation 5.1}$$

Where:

$$\text{Outputs} = \left\{ \begin{array}{l} Y_1 = \text{ Severe Injury Crash Rate,} \\ Y_2 = \text{ Injury Crash Rate,} \\ Y_3 = \text{ Fatal Crash Rate, and} \\ Y_4 = \text{ Total Crash Rate} \end{array} \right\}$$

$$\text{Inputs} = \begin{cases} X_1 = \text{Section Length}, \\ X_2 = \text{Surface Width}, \\ X_3 \text{ - } X_6 = \text{Route Class}, \\ X_7 \text{ - } X_{10} = \text{Shoulder Type Inside}, \\ X_{11} \text{ - } X_{14} = \text{Shoulder Type Outside}, \\ X_{15} \text{ - } X_{17} = \text{Median Type}, \\ X_{18} = \text{Median Width}, \\ X_{19} = \text{Average ADT}, \\ X_{20} = \text{Average Percentage of Heavy Vehicles}, \\ X_{21} = \text{Average Speed Limit} \end{cases}$$

Here, 21-3-4 represents 21 Inputs, 3 Hidden Nodes, and 4 Outputs. The R^2 values for the overall model and the each individual crash rate are: R^2 (Model) = 0.7034,

R^2 (SICR) = 0.7951, R^2 (ICR) = 0.7888, R^2 (FCR) = 0.4706, and R^2 (TCR) = 0.7592

As stated earlier, the described modeling methodology will also be used to model the crash rates for the remaining 5 networks. All results associated with these 5 networks are discussed in the next chapter.

5.5 Sensitivity Analysis

After the final structure is obtained, sensitivity analysis was performed on the input variables included in the final model. The sensitivity analysis would quantify, to some degree, the effect of each variable on the crash rates. To investigate the effect of continuous variables, each selected variable was given several values while keeping all other, un-related, variables stationary. Crash rates were then calculated and graphical trends are plotted. Once the sensitivity analysis was completed, the network's associated categorical variables were ranked from worst to best cases.

5.5.1 Development of Ranking Procedure

The ranking of categorical variables can easily be done if we are to rank them on the basis of a specific crash rate type. However, if the purpose is to determine an overall ranking based on all crash rates, this becomes a difficult task achieve. To overcome this dilemma, all categorical variables were first ranked for each crash rate type and the results were presented to a group of people to obtain an overall ranking. The individual ranking results were then compared in order to examine whether the rankings are consistent. As expected, the overall rankings were not consistent. For example (see results in Table 5.15) tables similar to Table 5.15 were given to the group and everyone was asked to give an overall ranking while taking into consideration individual ranking of each crash rate type. By examining the results given in Table 5.15, it can be seen that Route Class A is ranked the best while Route Class E is ranked the worst for three crash rate types. Positions in between varied between Route Classes B, C and D. When overall ranking was asked for, the study group Ranked A as best and E as worst. But the in-between ranking order was inconsistent. In order to resolve this overall ranking problem, a Combined Crash Potential Index (CCPI) was developed. In this case, it was decided to associate specific weight factors (i.e., degree of importance) to each crash rate. An iterative process was conducted to select the most appropriate weight factor, for each crash rate type, in order to establish consistent overall ranking process. The final expression for the CCPI, adopted in this study, is given in Equation 5.2.

$$CCPI = (TCR) + (2^{*}\,ICR) + (5^{*}\,SICR) + (10^{*}\,FCR) \qquad \text{Equation 5.2}$$

Where CCPI = Combined Crash Potential Index, TCR = Total Crash Rate, ICR = Injury Crash Rate, SICR = Severe Injury Crash Rate, and FCR = Fatal Crash Rate.

In the process of developing the weight factors, initial weights (based on the perceived importance of each crash rate type in comparison with the other 3 crash rates) were assigned and final overall rankings were obtained. Subsequently, the weights developed for one categorical variable were also used for other variables and then tested whether the overall ranking order obtained was consistent with the overall ranking given by the study group. The main goal of this iterative process was to select a set of weight factors that will be universally applicable for all associated categorical variables. After several iterations, the final set of weight factors were obtained as can be seen in Equation 5.2. Note that a weight factor, to some degree, reflects the degree of importance of its associated crash rate type in relative to other crash rates. For example, FCR is almost 10 times more significant than TCR.

Utilizing Equation 5.2, CCPI values associated with each categorical variable were obtained. Using the resulting CCPI values, all parameters within each categorical variable can be easily ranked. The resulting overall categorical variable rankings are given in Table 5.16.

5.5.2 Sensitivity of Categorical Variables

The overall rankings reported in this section are based on the CCPI values listed in Table 5.16.

1. Route Class

From crash potential point of view, Route Classes B and C are worst, Route Class D is moderate and Route Class E is best. The crash rates are expected to be higher on Route B and lower on Route E considering the facility type. The results are not in consistence with KDOT classification. The reasons for inconsistency might be due to changes in traffic patterns and flow, new developments or change of roadway features.

2. Median Type

Based on the calculated CCPI values for this case, Raised Median with Turning Lanes (Median Type 4) is the worst, Depressed Median (Median Type 6) is moderate and Barrier (Median Type 3) is the best. From the literature, one would expect Depressed Median to be the best, but in our case Barrier Median is the best. There might have been many left turn crashes on this network which has pushed Raised Median with Turn Lanes to the worst spot.

3. Shoulder Type Inside

Based on the calculated CCPI values listed in Table 5.16, Bituminous Base type shoulder is expected to be associated with the highest overall crash rates, P.C.C Shoulder is expected to be associated with moderate crash rates and, Aggregate Base Stabilized (A.B.S) and Curb and Gutter shoulders are expected to be associated with the least overall crash rates.

4. Shoulder Type Outside

As indicated by the CCPI values listed in Table 5.16, Bituminous Base is expected to be associated with the highest overall crash rates, Curb and Gutter is expected to be associated with moderate crash rates and, Aggregate Base Stabilized (A.B.S) and P.C.C Shoulders are expected to be associated with the least overall crash rates. From the literature, it is known that roads with paved shoulders are associated with fewer crashes than similar roads with unpaved shoulders. Once again, it is difficult to generalize the behavior of crash rates by shoulder type. It is very difficult to completely isolate the influence of the shoulder type on crash rates from other inter-related variables.

5.5.3 Sensitivity of Continuous Variables

Graphical results (trends) of the associated sensitivity analysis are presented in Figures 5.5.

1. Surface Width

Generally speaking, TCR/ICR/SICR, decrease with increase in surface width. The reason for this decrease can be explained by the fact that greater surface widths may provide more room for correction in near-accident circumstances. For example, if a vehicle is moving on a narrow lane, the slightest error or inattention can cause the vehicle to run over the edge-drop onto the shoulder. On the other hand, if the lane is wider, the same inattention will most likely allow the vehicle to stay on the road. In these near-accident circumstances, it will be difficult to separate between the effects of surface width, shoulder width, shoulder paving, edge-drops etc. It is likely that surface width plays a somewhat different role in single and multilane roads.

2. AADT

TCR/ICR increases up to 7000 vehicles and then decrease, while SICR remains constant up to 5000 vehicles and then increases. With increase in AADT, there is a higher chance for crashes as there are more vehicles occupying the facility. The severity of the crash would also be affected. For example, if the facility is completely saturated, then there is a higher chance for Property Damage Only (PDO) type crashes.

If the AADT reaches a certain threshold value for the facility, any addition beyond this point would most likely lower the crash rates since the movement of vehicles would be somewhat restricted.

3. Percentage of Heavy Vehicles

TCR decreases slightly with the increase in percent of heavy vehicles. On the other hand, ICR/SICR are statistically the same for the entire study range. Note that, with the increase in percentage of heavy vehicles, TCR is expected to decrease, since heavy vehicles constitute a large portion of the vehicle fleet. As a result, there is a lesser chance for speed differential, weaving and overtaking as compared to regular passenger cars. Beyond a certain threshold, if the percent of Heavy Vehicles increase, the crash rates might increase as well. In the case of this network, it seems that the threshold value is outside the study range, and hence this perceived behavior is not reflected in the trend depicted in Figure 5.2c.

4. Speed Limit

TCR/ICR/SICR remains unchanged with increase in speed limit. With the increase in speed limit there is no change in crash rates. Generally speaking, higher posted speed limits may lead to higher crash rates. On the other hand, these facilities, generally, have more lanes and better geometric configurations. The geometric configurations of this network are the main reason for the notable change in the (Total, Injury and Severe Injury) crash rates. In the sensitivity analysis, speed limits beyond 70 mph have not been considered. Hence, the noted trends would apply for speed limits up to 70 mph. If speed limits are increased beyond 70 mph, then the noted trends might be applicable.

5. Median Width

TCR/ICR remain almost constant up to 14 feet and then decrease. On the other hand, SICR remains almost unchanged for the entire study range. With the increase in median width, there is a lesser chance of opposite streams of vehicles to collide. As a result drivers are safer since reductions in (Total and Injury) crash rates are noted.

Table 5.1: Initial Input Variables Used in the Rural Expressways MLP Network

Node#	Input Variable	Type	Value
1	Section Length Miles	Continuous	Numerical
2	Surface Width	Continuous	Numerical
3-6	Route Class	Categorical [4]*	Binary
7-9	Surface Type	Categorical [3]*	Binary
10	Shoulder Width Outside	Continuous	Numerical
11	Shoulder Width Inside	Continuous	Numerical
12-20	Shoulder Type Outside	Categorical [9]*	Binary
21-29	Shoulder Type Inside	Categorical [9]*	Binary
30-35	Median Type	Categorical [6]*	Numerical
36	Additional Surface Width Inside	Continuous	Numerical
37	Additional Surface Width Outside	Continuous	Numerical
38	Median Width	Continuous	Numerical
39	Average Annual Daily Traffic (AADT)	Continuous	Numerical
40	Average % of Heavy Vehicles	Continuous	Numerical
41	Average Speed Limit	Continuous	Numerical
42-45	Sub Section	Categorical [4]*	Binary

* The number in the parenthesis refers to total number of categorical variables

Table 5.2: ANN Maximum & Minimum Values of Input Variables Used- Rural Expressways

Node #	Input Variable	ANN (Max)	ANN (Min)
1	Section Length Miles	14	0
2	Surface Width	36	11
3-6	Route Class	1	0
7-9	Surface Type	1	0
10	Shoulder Width Outside	15	0
11	Shoulder Width Inside	15	0
12-20	Shoulder Type Outside	1	0
21-29	Shoulder Type Inside	1	0
30-35	Median Type	1	0
36	Additional Surface Width Inside	15	0
37	Additional Surface Width Outside	15	0
38	Median Width	100	0
39	Average Annual Daily Traffic (AADT)	12000	200
40	Average % of Heavy Vehicles	50	0
41	Average Speed Limit	75	35
42-45	Sub Section	1	0

Table 5.3: Results of Training using Training/Testing Datasets

Initial Training using Training/Testing Datasets									
Start Node	Iterations	Hidden Nodes	MARE-Training*	MARE-Testing	R^2 Training	R^2 Testing	ASE-Training	ASE-Testing	ASE-Combined
1	300	3	497.423	452.297	0.87075	0.11654	0.001148	0.004962	0.00611
2	**800**	**3**	**491.406**	**436.964**	**0.96986**	**0.0127**	**0.001947**	**0.004818**	**0.006766**
3	800	3	493.14	457.594	0.81584	0.0154	0.002202	0.005109	0.007312
4	1000	5	480.216	443.655	0.94123	0.01302	0.001571	0.005047	0.006618
5	1000	6	505.78	472.782	0.85611	0.05216	0.001559	0.004996	0.006555
6	300	7	509.519	460.075	0.8177	0.03645	0.002535	0.004732	0.007267
7	200	8	508.485	481.648	0.80182	0.02414	0.002559	0.005326	0.007885
8	200	9	504.982	465.608	0.8349	0.05044	0.002634	0.005025	0.007659
9	900	9	498.139	474.368	0.73064	0.04617	0.002633	0.005081	0.007714
10	900	10	525.01	499.967	0.70503	0.0907	0.003199	0.004983	0.008182

*R2-Overall Coefficient of Determination Factor, MARE-Mean Average Relative Error, ASE-Averaged-Squared-Error

Table 5.4: Results of Training using Training/Validation Datasets

Initial Training using Training/Validation Datasets									
Start Node	Iterations	Hidden Nodes	MARE-Training	MARE-Validation	R^2 Training	R^2 Validation	ASE-Training	ASE-Validation	ASE-Combined
2	**800**	**3**	**491.406**	**468.763**	**0.96986**	**0.01682**	**0.001947**	**0.011185**	**0.013132**

Table 5.5: Results of Training using all datasets

All Training					
Start Node	Iterations	Hidden Nodes	MARE-Training	R^2 Training	ASE- Training
2	**800**	**3**	**496.656**	**0.57685**	**0.003204**

Table 5.6: Results of Training using all variables

Original					
Start Node	Iterations	Hidden Nodes	MARE-Training*	R^2 Training	ASE - Training
2	800	3	496.656	0.57685	0.003204
Variable Dropped	Iterations	Hidden Nodes	MARE- Training	R^2 Training	SSEN- Training
SS	800	3	464.68	0.8111	0.001733
Shoulder Type Out	800	3	489.655	0.6631	0.002054
Shoulder Type In	800	3	490.321	0.47284	0.004059
Shoulder Width Out	800	3	480.576	0.54737	0.003477
Shoulder Width In	800	3	481.578	0.66879	0.002135

Table 5.7: Results of Training after dropping SS

Original-SS Dropped					
Start Node	Iterations	Hidden Nodes	MARE-Training	R^2 Training	ASE- Training
2	800	3	464.68	0.8111	0.001733
Variable Dropped	Iterations	Hidden Nodes	MARE- Training	R^2 Training	SSEN- Training
Shoulder Type Out	800	3	494.951	0.64641	0.001899
Shoulder Type In	800	3	487.925	0.30159	0.005747
Shoulder Width Out	800	3	501.493	0.80163	0.001846
Shoulder Width In	800	3	509.831	0.75217	0.002996

Table 5.8: Results of Training after dropping SS and Shoulder Width Outside

Original-SS, Shoulder Width Out Dropped					
Start Node	Iterations	Hidden Nodes	MARE-Training	R^2 Training	ASE- Training
2	800	3	501.493	0.80163	0.001846
Variable Dropped	Iterations	Hidden Nodes	MARE- Training	R^2 Training	SSEN- Training
Shoulder Type Out	800	3	496.721	0.64658	0.00187
Shoulder Type In	800	3	520.706	0.74143	0.001804
Shoulder Width In	800	3	468.665	0.79932	0.001753

Table 5.9: Results of Training after dropping SS and Shoulder Widths Outside and Inside

Original-SS, Shoulder Width Out, Shoulder Width In Dropped					
Start Node	**Iterations**	**Hidden Nodes**	**MARE-Training***	**R^2 Training**	**ASE- Training**
2	800	3	468.665	0.79932	0.001753
Variable Dropped	**Iterations**	**Hidden Nodes**	**MARE-Training**	**R^2 Training**	**SSEN-Training**
Shoulder Type Out	800	3	495.535	0.71703	0.002765
Shoulder Type In	800	3	471.824	0.70211	0.002784

Table 5.4: Final Network Structure

Final Structure (after variables dropped)					
Start Node	**Iterations**	**Hidden Nodes**	**MARE-Training***	**R^2 Training**	**ASE- Training**
2	**800**	**3**	**468.665**	**0.79932**	**0.001753**

Table 5.5: Results of Final Training using Training/Testing Datasets

Training/Testing Datasets										
Start Node	**Iterations**	**Hidden Nodes**	**MARE-Training***	**MARE-Testing**	**R^2 Training**	**R^2 Testing**	**ASE-Training**	**ASE-Testing**	**ASE-Combined**
1	900	3	522.737	480.323	0.88089	0.50735	0.000197	0.004352	0.004549

Table 5.6: Results of Final Training using Training/Validation Datasets

Training/Validation Datasets									
Start Node	**Iterations**	**Hidden Nodes**	**MARE-Training**	**MARE-Validation**	**R^2 Training**	**R^2 Validation**	**ASE-Training**	**ASE-Validation**	**ASE-Combined**
1	900	3	522.737	515.188	0.88089	0.24007	0.000197	0.011506	0.011703

Table 5.7: Final Network

All Datasets					
Start Node	**Iterations**	**Hidden Nodes**	**MARE-Training***	**R^2 Training**	**ASE- Training**
1	**900**	**3**	**504.892**	**0.70347**	**0.001821**

---

*The number in the parenthesis refers to the total number of categorical variables

Table 5.8: Final Input Variables for Rural Expressways MLP Network

Node #	Input Variable	Type	Value
1	Section Length Miles	Continuous	Numerical
2	Surface Width	Continuous	Numerical
3-6	Route Class	Categorical [4]*	Binary
7-10	Shoulder Type Inside	Categorical [4]*	Binary
11-14	Shoulder Type Outside	Categorical [4]*	Binary
15-17	Median Type	Categorical [3]*	Binary
18	Median Width	Continuous	Numerical
19	Average ADT	Continuous	Numerical
20	Average % of Heavy Vehicles	Continuous	Numerical
21	5-year Average Speed Limit	Continuous	Numerical

Table 5.15: Sample Results for the Categorical Variable ranking procedure

Rankings from Best to Worst			
Total	**Injury**	**Severe Injury**	**Fatal**
Route Class A	Route Class A	Route Class A	Route Class B
Route Class B	Route Class C	Route Class D	Route Class A
Route Class C	Route Class B	Route Class B	Route Class C
Route Class D	Route Class D	Route Class C	Route Class E
Route Class E	Route Class E	Route Class E	Route Class D

Table 5.16: Final Overall Ranking of the Categorical Variables – Rural Expressways

Route Class Rankings (Worst to Best)
Route Class B (CCPI = 27.6)
Route Class C (CCPI = 15.9)
Route Class D (CCPI = 7.9)
Route Class E (CCPI = 3.3)
Median Type Rankings (Worst to Best)
Median Type 4: Raised Median + Turning Lanes (CCPI =24.1)
Median Type 6: Depressed Median (CCPI = 18.8)
Median Type 3: Barrier (CCPI = 2.7)
Shoulder Type Inside Ranking (Worst to Best)
STI(11): Bituminous Base (CCPI = 29.9)
STI(29): P.C.C. Shoulder (CCPI = 15.0)
STI(5): Aggregate Base Stabilized (A.B.S) (CCPI = 3.6)
STI(19): Curb and Gutter (CCPI = 3.5)
Shoulder Type Outside Ranking (Worst to Best)
STO(11): Bituminous Base (CCPI = 39.7)
STO(19): Curb and Gutter (CCPI = 7.5)
STO(5): Aggregate Base Stabilized (A.B.S) (CCPI = 3.7)
STO(29): P.C.C. Shoulder (CCPI = 2.5)

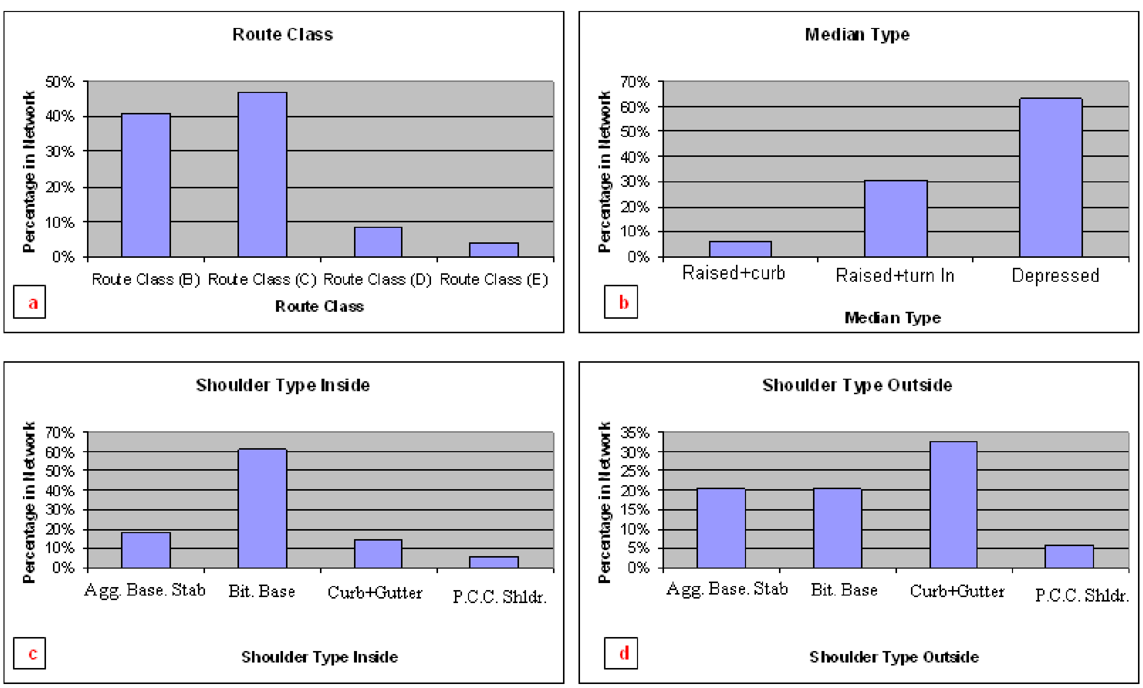


Figure 5.1: Categorical Variables Distributions- (Rural Expressway Network)

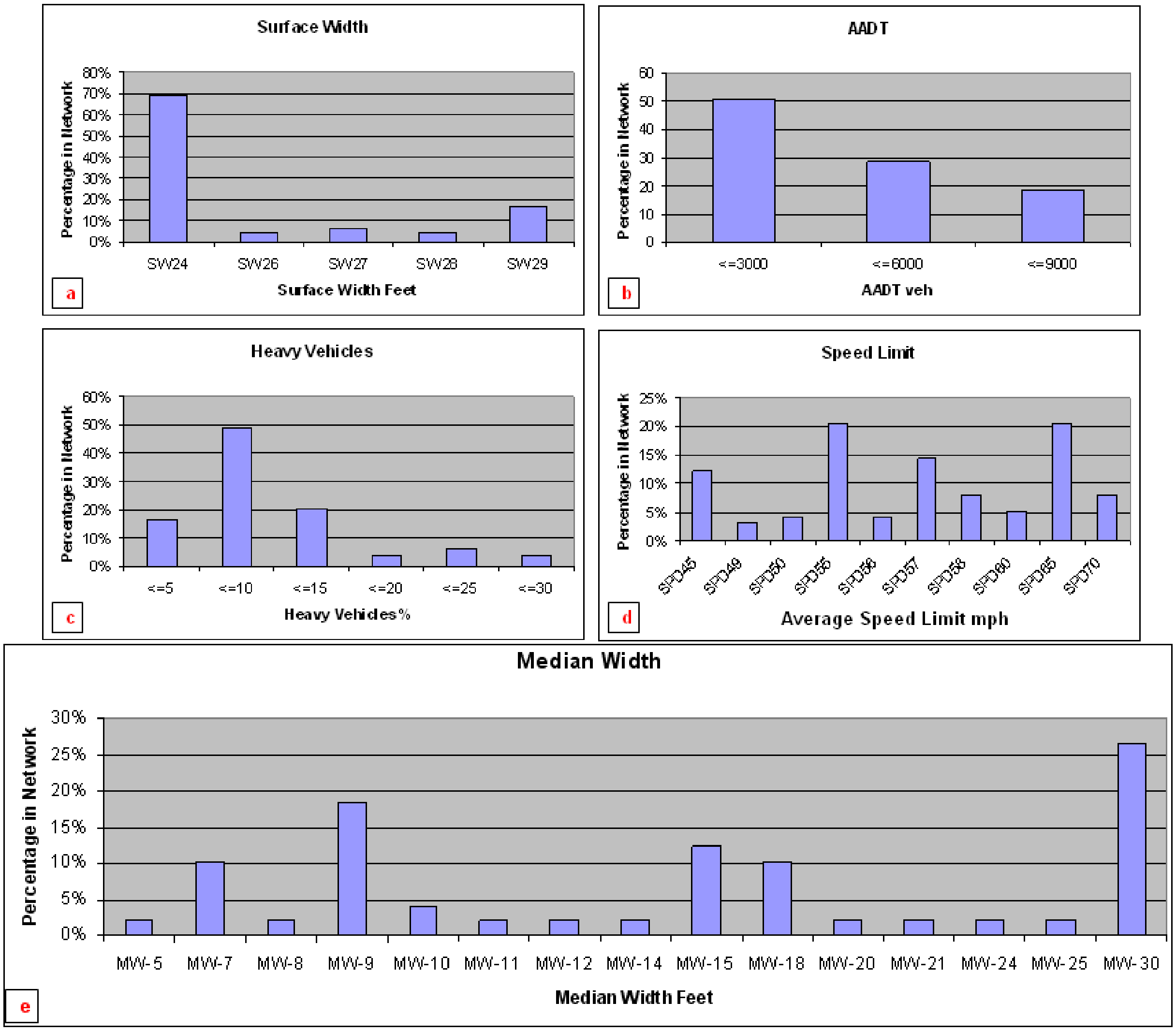


Figure 5.2: Continuous Variables Distributions- (Rural Expressway Network)

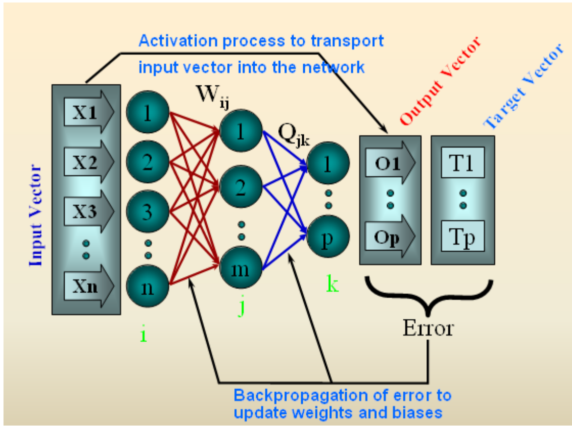


Figure 5.3:Schematic of MLP Neural Network

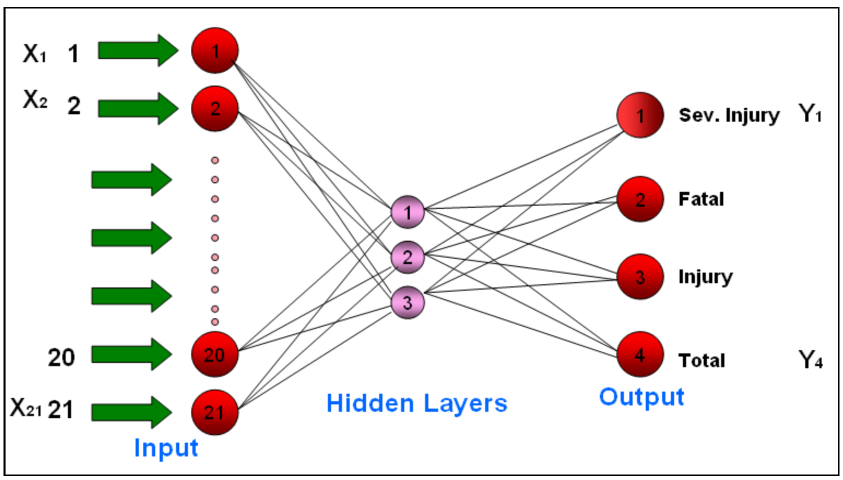


Figure 5.4:Final Network Structure- Rural Expressways

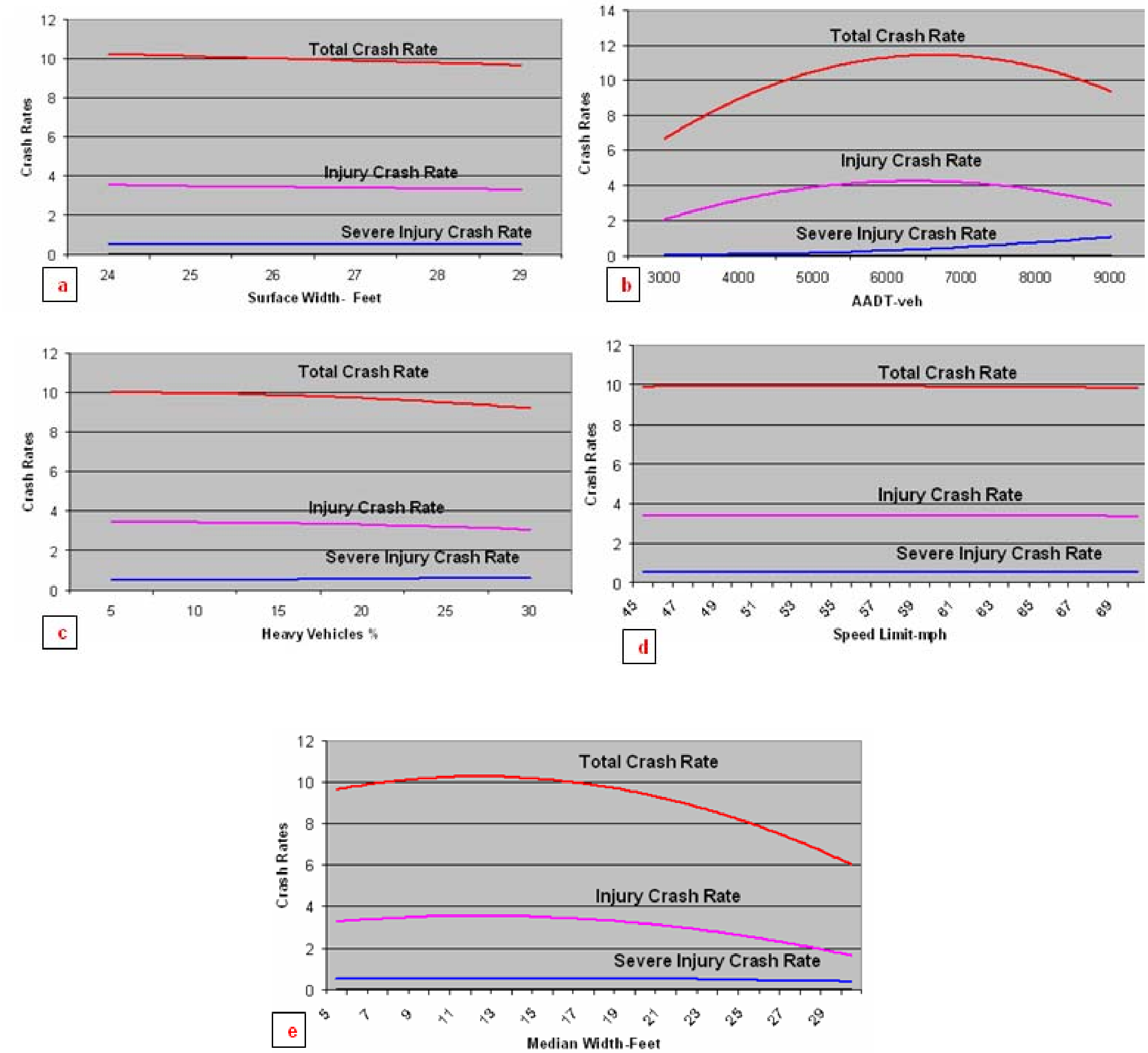


Figure 5.5:Sensitivity Analysis Results of Continuous Variables – (Rural Expressway Network)

CHAPTER 6 - MODELING AND SENSITIVITY ANALYSIS RESULTS

6.1 Introduction

This chapter presents the results of modeling and sensitivity analysis for remaining five networks. The modeling methodology is exactly the same for all networks.

6.3 Rural 2 Lane Network

This is the largest network in the database. This network has a total of 5114 datasets. Training utilized 2565 datasets, testing utilized 1282 datasets and validation utilized 1267datasets.

6.3 Distribution of Variables

The distribution of categorical and continuous variables for the rural 2 lane network are shown in Figures 6. and 6.

6.3.1 Categorical Variables

There are two types of categorical variables for the Rural 2 Lane Network.

1. **Route Class**

   Figure 6.a shows the distribution plot for Route Class. Route Class D (34%) and Route Class E (26%) constitute the major portion of the Rural 2 Lane Network. Route Class C (21%) and Route Class B (19%) constitute the remaining part of the network. This network is the major corridor in the entire state of Kansas.

2. **Shoulder Type Outside**

Figures 6.b shows the distribution plot for Shoulder Type Outside. There are 13 shoulders types for this network. Turf (Shoulder Type 2), covers (40%) of the network. The remaining 12 types combined represent (60%) of the network.

6.3.2 Continuous Variables

There are five types of continuous variables.

1. **Surface Width**

Figure 6.a shows the distribution plot for Surface Width. A majority of the network has a surface width of 24 feet (68%). The remainder of the network has surface widths of ranging from 21ft to 50ft and all of them combined represent (32%) of the network.

2. **AADT**

Figure 6.b shows the distribution plot for AADT. The AADT range for this network is 0-9000 vehicles. 36% of the network sections have an AADT of (<=1000 vehicles), 29% of the sections have AADT between 1001- 2000 vehicles and the remaining network (35%) has AADT between 2001 and 9000 vehicles.

3. **Heavy Vehicles**

Figure 6.c shows the distribution plot for percentage of Heavy Vehicles (%HV). The %HV range for this network is 0-45%. Almost half of the sections (49%) have %HV between 6% and 15%. The remaining sections (51%) have %HV ranging from 0 to 5% and 16-45%.

4. **Speed limit**

Figure 6.d depicts the distribution plot for Speed Limit. In this network, the 5-year average speed limits varied from 45 mph to 70 mph. The predominant speed limits were 55 mph (26%), 57 mph (20%) and 65 mph (21%). These three speeds cover 67% of network and the remainder 33% have speeds ranging from 20 mph to 63 mph. Please note that the speed limits have been averaged over a 5 year period and hence we may have values that are not divisible by 5.

5. **Shoulder Width Outside**

Figure 6.e shows the distribution plot for the Shoulder Width. The outside shoulder widths varied from 0 to 10 feet. The predominant width is 10feet, which was used on about 35% of network's sections. The remaining sections (65%) had widths of less than 10ft.

6.4 Training Methodology

The training methodology is exactly the same as outlined in Chapter 5. The initial number of input nodes started at 46 nodes (representing the 46 input variables) and iterations were carried out to identify the best possible network structure. The final structure was obtained at 700 iterations and 6 hidden nodes. The initial input variables are shown in Table 6.1 and the final input variables are shown in Table 6.2.

Figure 6. shows the structure of the final MLP neural network model and Equation 6.1 gives the structure of the final model.

$$(Y_{1,}\ Y_2, Y_3, Y_4) = ANN_{23\text{-}6\text{-}4}(X_1, X_2, X_3, X_4 \ldots X_{23}) \qquad \text{Equation 6.1}$$

Where:

$$\text{Outputs} = \begin{Bmatrix} Y_1 = \text{ Severe Injury Crash Rate,} \\ Y_2 \ = \text{ Injury Crash Rate,} \\ Y_3 \ = \text{ Fatal Crash Rate, and} \\ Y_4 \ = \text{ Total Crash Rate} \end{Bmatrix}$$

$$\text{Inputs} = \begin{Bmatrix} X_1 = \text{Section Length,} \\ X_2 \ = \text{ Surface Width,} \\ X_3 \text{ - } X_6 \ = \text{ Route Class,} \\ X_7 \ = \text{ Shoulder Width Outside} \\ X_8 \text{ - } X_{20} \ = \text{ Shoulder Type Outside,} \\ X_{21} \ = \text{ Average ADT,} \\ X_{22} \ = \text{ Average Percentage of Heavy Vehicles,} \\ X_{23} \ = \text{ Average Speed Limit} \end{Bmatrix}$$

23-3-4 represents 23 Inputs, 6 Hidden Nodes, and 4 Outputs. The R^2 values for the model and the individual crash rates are: R^2 (Model) = 0.4655, R^2 (SICR) = 0.8647, R^2 (ICR) = 0.0456, R^2 (FCR) = 0.7791, R^2 (TCR) = 0.1728

6.5 Sensitivity Analysis

After the final structure is obtained, sensitivity analysis was performed on the input variables in the final model. The results for categorical variables are shown in Table 6.3. The results for continuous variable are shown in Figure 6.4 a-e.

6.5.1 Sensitivity of Categorical Variables

The overall rankings reported herein for Rural 2 Lane network are based on the CCPI values listed in Table 6.3.

1. **Route Class**

   From crash potential point of view, Route Classes E and D are worst, Route Class C is moderate and Route Class B is best. The crash rates are expected to be higher on Route Class E and least on Route Class B. This observation is consistent with the KDOT classification of the Kansas 2-lane routes.

2. **Shoulder Type Outside**

   Based on the calculated CCPI values for this case, Shoulder types: Aggregate Base Stabilized (A.B.S) and Gutter, and Calcium Chloride with Limestone are expected to be associated with the highest overall crash rates. On the other hand, Shoulders having One Feet Bitumen with Remainder Turf are expected to be associated with the least overall crash rates.

6.5.2 Sensitivity of Continuous Variables

Graphical results (trends) of the associated sensitivity analysis are presented in Figure 6.4 a-e.

1. **Surface Width**

   TCR decreases with increase in surface width up to 26 feet. Further increase in surface width slightly increases TCR. ICR/SICR, decrease with increase in surface width. This network has only one travel lane in each direction. Greater surface widths would separate opposite streams of traffic and would reduce the likelihood of head on collisions. Beyond a certain point, the extra widths would give room for overtaking and speeding, which again might lead to an increase in crash rates.

2. **AADT**

TCR remains unchanged with increase in AADT up to 5000 vehicles and then increases. ICR/SICR remain statistically the same for the entire study range.

3. **Percentage of Heavy Vehicles**

TCR sharply decreases with increase in % heavy vehicles up to 30%. Beyond this point, TCR is remains unchanged. On the other hand, ICR/SICR remains statistically the same for the entire study range. The results are consistent with general behavior.

4. **Speed Limit**

TCR decreases with increase in speed limit. ICR/SICR remains unchanged with increase in speed limit. The maximum speed limit for this network is 65 mph. The observed trends hold good up to 65 mph. Beyond this speed, the crashes might go up. The trend is in conformance with the literature review.

5. **Shoulder Width**

In this case, TCR/ICR/SICR shows a decreasing trend with increase in shoulder width up to 6 feet and then start increasing. Providing additional shoulder widths beyond 6ft might be providing a false sense of security to the drivers thereby leading to an increase in crash rates. s the drivers with a false since of security.

6.6 Rural Freeways Network

This network has a total of 349 datasets. Training utilized 178 datasets, testing utilized 86 datasets and validation utilized 85 datasets.

6.7 Distribution of Variables

The distributions of categorical and continuous variables for the Rural Freeways network are shown in Figures 6.5 and 6..

6.7.1 Categorical Variables

There are three types of categorical variables for the Rural Freeways Network.

1. Route Class

Figure 6.5a shows the distribution plot for Route Class. Route Class A (82%) constitutes the major portion of the Rural Freeway Network. Route Class B (11%) and Route Class C (19%) constitute the remaining part of the network.

2. Median Type

Figure 6.5b shows the distribution plot for Median Type. Depressed Median (Median Type 6) constitutes the major portion (82%) of the network and Barrier (Median Type 8) covers the remaining (18%).

3. Shoulder Type Outside

Figure 6.5c shows the distribution plot for Shoulder Type Inside. Bituminous Base (Shoulder Type 11) constitutes the major portion (69%) of the network. P.C.C Shoulder (Shoulder Type 29) and Asphaltic Concrete Shoulder (Shoulder Type 30) cover the remaining (31%).

6.7.2 Continuous Variables

There are four types of continuous variables.

1. **Median Width**

Figure 6.a shows the distribution plot for Median Width. The median width's range for this network is from 10ft-30ft. The predominant median width is 30 feet, which was used on about (80%) of the sections and the remaining sections have widths ranging from 10ft-27ft.

2. **Heavy Vehicles**

Figure 6.b shows the distribution plot for percentage of Heavy Vehicles (%HV). 66% of the sections have %HV between 16% and 30%. Of this 66%, 24% of sections have %HV between 26 and 30, 23% of sections have %HV between 16 and 20, and 19% of sections have %HV between 16 and 20. The remaining sections have %HV ranging from 0 to 15 and 36 to 40.

3. **AADT**

Figure 6.c shows the distribution plot for AADT. The AADT range for this network is 0-15700 vehicles. 59% of the network sections have an AADT of (<=5000 vehicles), 39% of them have AADT between 5000 and 10000 vehicles and the remaining 2% of them have AADT greater than 10000 and 15700 vehicles.

4. **Speed limit**

Figure 6.d shows the distribution plot for Speed Limit. In this network, the 5-year average speed limits varied from 58 mph to 70 mph. The predominant speed limits were 66 mph (42%) and 70 mph (48%). The remaining (10%) sections have speeds ranging from 58 mph to 65 mph. Please note that the speed limits have

been averaged over a 5 year period and hence we may have values that are not divisible by 5.

6.8 Training Methodology

The initial number of input nodes started at 25 nodes (representing the 25 input variables) and iterations were carried out to identify the best possible network structure. The final structure was obtained at 200 iterations and 5 hidden nodes. The initial input variables are shown in Table 6.4 and the final input variables are shown in Table 6.5.

Figure 6. shows the structure of the final MLP neural network model and Equation 6.2 gives the structure of the final model.

$$(Y_{1,}\ Y_2, Y_3, Y_4) = ANN_{13\text{-}5\text{-}4}(X_1, X_2, X_3, X_4\ \ldots X_{13}) \qquad \text{Equation 6.2}$$

Where:

$$\text{Outputs} = \begin{Bmatrix} Y_1 = \text{ Severe Injury Crash Rate,} \\ Y_2 = \text{ Injury Crash Rate,} \\ Y_3 = \text{ Fatal Crash Rate, and} \\ Y_4 = \text{ Total Crash Rate} \end{Bmatrix}$$

$$\text{Inputs} = \begin{Bmatrix} X_1 = \text{Section Length,} \\ X_2\text{-}X_4 = \text{ Route Class,} \\ X_5 - X_7 = \text{ Shoulder Type Inside,} \\ X_8\text{-}X_9 = \text{ Median Type,} \\ X_{10} = \text{ Median Width,} \\ X_{11} = \text{ Average ADT,} \\ X_{12} = \text{ Average Percentage of Heavy Vehicles,} \\ X_{13} = \text{ Average Speed Limit} \end{Bmatrix}$$

13-5-4 represents 13 Inputs, 5 Hidden Nodes, and 4 Outputs. The R^2 values for the model and the individual crash rates are: R^2 (Model) = 0.2803, R^2 (SICR) = 0.5074, R^2 (ICR) = 0.3048, R^2 (FCR) = 0.0426, R^2 (TCR) = 0.2664

6.9 Sensitivity Analysis

After the final structure is obtained, sensitivity analysis was performed on the input variables in the final model and Table 6. gives the results for categorical variables. The results for continuous variable are shown in Figure 6. a-d.

6.9.1 Discussion- Sensitivity of Categorical Variables

The ranking reported herein are based on the CCPI values listed in Table 6.6.

1. Route Class

From crash potential point of view, Route Classes A is worst, Route Class C is moderate and Route Class B is best. The crash rates are expected to be higher on Route Class A and least on Route Class B.

2. Median Type

From crash potential point of view, Depressed Median (Median Type 8) is the worst, and Barrier (Median Type 6) is the best.

3. Shoulder Type Outside

Based on the calculated CCPI values for this case, Bituminous Base shoulder is expected to be associated with the highest overall crash rates, Asphaltic Concrete Shoulder is expected to be associated with moderate crash rates and P.C.C Shoulder is expected to be associated with the least overall crash rates.

6.9.2 Discussion- Sensitivity of Continuous Variables

1. AADT

TCR increases up to AADT value of 10000 and then decreases beyond this point. ICR remains constant up to 10000 and then decreases slightly. SICR remains statistically the same for the entire study range. The findings are consistent with the literature.

2. **Percentage of Heavy Vehicles**

   TCR remains constant up to 25% and then decreases slightly. On the other hand, ICR/SICR are statistically the same for the entire study range. The findings are consistent with other networks.

3. **Speed Limit**

   TCR remains constant up to 63mph and then increases slightly. ICR/SICR are statistically the same for the entire study range. The findings are consistent with the literature.

6.10 Rural KTA Network

This is the smallest network in the database. It has a total of 65 datasets. Training utilized 37 datasets, testing utilized 14 datasets and validation utilized 14 datasets.

6.11 Distribution of Variables

The distribution of categorical and continuous variables for the Rural KTA Network is shown in Figures 6. and 6..

6.11.1 Categorical Variables

There is only one categorical variable for this network.

Sub-Section (SS)

Figures 6. shows the distribution plot for Sub-Section. SS-I is the most predominant subsection type and covers 76% of the network. SS-II and SS-III cover the remaining 24%. There are no details provided by KDOT about the Sub-Section types and the use of each type.

6.11.2 Continuous Variables

There are three types of continuous variables.

1. **Heavy Vehicles**

   Figure 6.a shows the distribution plot for percentage of Heavy Vehicles (%HV).The percentage of HV for this network varied from 14-27. 20% of the sections have 18%HV and 17% of sections have 20%HV. The remaining sections (63%) have %HV ranging from 14 to 27%.

2. **AADT**

   Figure 6.0b shows the distribution plot for AADT. The AADT range for this network is 0-12000 vehicles. 23% of sections have AADT between 2000 and 3000 vehicles, 22% of sections have AADT between 5000 and 6000 vehicles and 19% of sections have AADT between 4000 and 5000 vehicles. The remaining sections (36%) have AADT between 2000 and 12000 vehicles.

3. **Speed limit**

   Figure 6.c shows the distribution plot for Speed Limit. There are only two speed limits (66 mph and 70 mph) for this network and each cover 50% of the network. Please note that the speed limits have been averaged over a 5 year period and hence we may have values that are not divisible by 5.

6.12 Training Methodology

The initial number of input nodes started at 17 nodes (representing the 17 input variables) and iterations were carried out to identify the best possible network structure. The final structure was obtained at 600 iterations and 3 hidden nodes. The initial input variables are shown in Table 6. and the final input variables are shown in Table 6..

Figure 6. shows the structure of the final MLP neural network model and Equation 6.3 gives the structure of the final model.

$$(Y_{1,}\ Y_2, Y_3, Y_4) = \mathrm{ANN}_{7\text{-}3\text{-}4}(X_1, X_2, X_3, X_4\ \ldots X_7) \qquad \text{Equation 6.3}$$

Where

$$\text{Outputs} = \left\{\begin{array}{l} Y_1 = \text{ Severe Injury Crash Rate,} \\ Y_2\ = \text{ Injury Crash Rate,} \\ Y_3\ = \text{ Fatal Crash Rate, and} \\ Y_4\ = \text{ Total Crash Rate} \end{array}\right\}$$

$$\text{Inputs} = \left\{\begin{array}{l} X_1 = \text{Section Length,} \\ X_2\text{-}X_4\ = \text{ Sub Section,} \\ X_5\ = \text{ Average ADT,} \\ X_6\ = \text{ Average Percentage of Heavy Vehicles,} \\ X_7\ = \text{ Average Speed Limit} \end{array}\right\}$$

7-3-4 represents 7 Inputs, 3 Hidden Nodes, 4 Outputs. The R^2 values for the model and the individual crash rates are: R^2 (Model) = 0.1201, R^2 (SICR) = 0.1458, R^2 (ICR) = 0.0693, R^2 (FCR) = 0.0598, R^2 (TCR) = 0.2057

6.13 Sensitivity Analysis

After the final structure is obtained, sensitivity analysis was performed on the input variables in the final model and Table 6. gives the results for categorical variables. Figure 6. a-c shows the results for continuous variables

6.13.1 Discussion- Sensitivity of Categorical Variables

The ranking reported herein are based on the CCPI values listed in Table 6.

SS (Sub-Section)

From crash potential point of view, SS-II is worst, SS-I is moderate and SS-III is best. SS refers to subsection and each network is divide into sub-sections based on the section geometric characteristics. It is clearly not stated in the KARS manual about the characteristics of each type and hence only the order of the sub-sections in terms of expected crash rates is given.

6.13.2 Discussion- Sensitivity of Continuous Variables

1. **Percentage of Heavy Vehicles**

   TCR remains constant up to 20% HV and then has a steep decrease. On the other hand, ICR/SICR remain statistically unchanged for the entire study range. The findings are consistent with other networks.

2. **AADT**

   TCR decreases slightly up to AADT value of 10000 and increases slightly. ICR/SICR remain statistically unchanged for the entire study range. The results are different from other networks. One reason for the difference in behavior can be attributed to the study volume range.

3. **Speed Limit**

   TCR/ICR/SICR, are statistically the same for the entire study range. The results are slightly different, compared to other networks. It is to be noted that Turnpike roads are slightly different compared to other roads. The maintenance and operation of these networks are done by a separate authority when compared to the regular road networks. The slightly different behavior of the traffic variables on this network may be attributed to this fact.

6.14 Urban Freeways Network

This is the largest urban network and had a total of 435 datasets. Training utilized 220 datasets, testing utilized 108 datasets and validation utilized 107 datasets.

6.15 Distribution of Variables

The distribution of categorical and continuous variables for the Rural Freeways network is shown in Figures 6.13 and 6.4.

6.15.1 Categorical Variables

There are three types of categorical variables for the Urban Freeways Network.

1. **Route Class**

   Figure 6.13a shows the distribution plot for Route Class. Route Class A (75%) constitutes the major portion of the network followed by Route Class B (15%), Route Class C (8%), Route Class D (1%) and Route Class E (1%).

2. **Median Type**

   Figure 6.b shows the distribution plot for Median Type. Depressed Median (Median Type 6) (55%) constitutes the major portion of the network, Barrier (Median Type 8) covers (42%) and Raised Median with Curbs (Median Type 3) covers (3%).

3. **Shoulder Type Outside**

   Figure 6.c shows the distribution plot for Shoulder Type Outside. P.C.C. Shoulder (Shoulder Type 29) (49%) constitutes the major portion of the network followed by Bituminous Base (Shoulder Type 11) (32%), Asphaltic Concrete Shoulder (Shoulder Type 30) (11%) and Curb and Gutter (Shoulder Type 19) (8%).

6.15.2 Continuous Variables

There are six types of continuous variables for the Urban Freeways Network.

1. **Surface Width**

Figure 6.a shows the distribution plot for Surface Width. The predominant surface width is 24 feet, which covers (55%) of the network. The remainder of the network has surface widths ranging from 26ft to 51 ft.

2. **Shoulder Width Inside**

Figure 6.b shows the distribution plot for the Shoulder Width Inside. The inside shoulder widths varied from 0-10ft. The most predominant width is 6ft, which was used on (50%) of the network's sections and the other widths cover the remaining 50% of sections.

3. **Median Width**

Figure 6.c shows the distribution plot for the Median Width. For this network the median widths varied from 8ft-55ft. The most predominant median width is 30ft and covers (58%) of the network's sections while other widths cover the remaining (42%) of sections.

4. **AADT**

Figure 6.d shows the distribution plot for AADT. The AADT range for this network is from 0 to 60000 vehicles. The most predominant AADT is between 5000 and 10000 vehicles and covers (24%) of the network.

5. **Heavy Vehicles**

Figure 6.e shows the distribution plot for percentage of Heavy Vehicles (%HV). 45% of the sections have %HV between 6% and 10%. The remaining sections have %HV ranging from 0-35%.

6. **Speed limit**

Figure 6.f depicts the distribution plot for Speed Limit. In this network, the 5-year average speed limits varied from 55mph to 70mph. The predominant speed limits were 55 mph (16%), 57mph (14%), 65mph (15%) and 70mph (17%). Please note that the speed limits have been averaged over a 5 year period and hence we may have values that are not divisible by 5.

6.16 Training Methodology

The initial number of input nodes started at 30 nodes (representing the 30 input variables) and iterations were carried out to identify the best possible network structure. The final structure was obtained at 700 iterations and 5 hidden nodes. The initial input variables are shown in Table 6. and the final input variables are shown in Table 6..

Figure 6. shows the structure of the final MLP neural network model and Equation 6.4 gives the structure of the final model.

$$(Y_{1,}\ Y_2, Y_3, Y_4) = ANN_{19\text{-}5\text{-}4}(X_1, X_2, X_3, X_4\ \ldots X_{19}) \qquad \text{Equation 6.4}$$

Where:

$$\text{Outputs} = \left\{ \begin{array}{l} Y_1 = \text{ Severe Injury Crash Rate,} \\ Y_2 \ = \text{ Injury Crash Rate,} \\ Y_3 \ = \text{ Fatal Crash Rate, and} \\ Y_4 \ = \text{ Total Crash Rate} \end{array} \right\}$$

$$\text{Inputs} = \left\{ \begin{array}{l} X_1 = \text{Section Length,} \\ X_2 = \text{Surface Width} \\ X_3\text{-}X_7 \ = \text{ Route Class,} \\ X_8 \ = \text{ Shoulder Width Inside,} \\ X_9\text{-}X_{12} \ = \text{ Shoulder Type Outside,} \\ X_{13}\text{-}X_{15} \ = \text{ Median Type,} \\ X_{16} \ = \text{ Median Width,} \\ X_{17} \ = \text{ Average ADT,} \\ X_{18} \ = \text{ Average Percentage of Heavy Vehicles,} \\ X_{19} \ = \text{ Average Speed Limit} \end{array} \right\}$$

19-5-4 represents 19 Inputs, 5 Hidden Nodes, 4 Outputs. The R^2 values for the model and the individual crash rates are:R^2 (Model) = 0.2766, R^2 (SICR) = 0.1007, R^2 (ICR) = 0.2648, R^2 (FCR) = 0.3310, R^2 (TCR) = 0.4101.

6.17 Sensitivity Analysis

After the final structure is obtained, sensitivity analysis was performed on the input variables in the final model and Table 6. gives the results for categorical variables. The results for continuous variable are shown in Figure 6. a-f.

6.17.1 Discussion- Sensitivity of Categorical Variables

The ranking reported herein are based on the CCPI values listed in Table 6..

1. **Route Class**

   From crash potential point of view, Route Classes E and C are worst, Route Class A is moderate and Route Classes D and B are best. The crash rates are expected to be higher on Route Classes E and C, moderate on Route Class A and low on Route Classes D and B.

2. **Median Type**

   From crash potential point of view, Depressed Median (Median Type 8) is the worst, Barrier (Median Type 6) is moderate and Raised Median with Curbs is the best. The findings are slightly different from literature.

3. **Shoulder Type Outside**

   Based on the calculated CCPI values for this case, Bituminous Base and Asphaltic Concrete Shoulders are expected to be associated with the highest overall crash rates, P.C.C Shoulder is expected to be associated with moderate crash rates and Curb and Gutter Shoulder is expected to be associated with the least overall crash rates.

6.17.2 Discussion- Sensitivity of Continuous Variables

1. **Surface Width**

   TCR/ICR decrease with increase in surface width up to 36ft and then increase. The addition of surface width beyond 36ft would give in extra room for drivers to overtake and speed and could be the reason for increase in crash rates.

2. **Shoulder Width Inside**

TCR/ICR decrease with increase in inside shoulder width up to 5ft and then increase. Beyond 5 feet if the inside shoulder width is increased the extra room would almost serve as an extra lane and drivers might use that extra width for overtaking maneuvers which in turn might increase the crash rates.

3. **Median Width**

TCR/ICR remains almost constant up to 28 feet and then increase.

4. **AADT**

TCR/ICR steeply decrease with increase in AADT. The results are consistent with literature.

5. **Percentage of Heavy Vehicles**

TCR increases slightly up to 20% and then decreases slightly. On the other hand, ICR is fairly constant for the entire study range. The findings are slightly different from the literature.

6. **Speed Limit**

TCR increases slightly up to 63mph and then decreases. ICR remains statistically same for the entire study range. The findings are slightly different from the literature.

It is to be noted that in all cases considered herein, FCR's have very limited datasets and therefore it is difficult to generalize their behavior.

6.18 Urban Expressways

This network has a total of 80 datasets. Training utilized 46 datasets, testing utilized 17 datasets and validation utilized 17 datasets.

6.19 Distribution of Variables

The distribution of categorical and continuous variables for the Urban Freeways network is shown in Figures 6.17 and 6.18.

6.19.1 Categorical Variables

There are two types of categorical variables for the Urban Freeways Network.

1. **Route Class**

   Figure 6.17a shows the distribution plot for Route Class. Route Class C (48%) constitutes the major portion of the network followed by Route Class B (40%), Route Class D (8%) and Route Class E (4%).

2. **Median Type**

   Figure 6.b shows the distribution plot for Median Type. Depressed Median (Median Type 6) (64%) constitutes the major portion of the network, Raised Median with Turning Lanes (Median Type 4) covers (30%) and Raised Median with Curbs (Median Type 3) covers (6%).

6.19.2 Continuous Variables

There are six types of continuous variables for the Urban Freeways Network.

1. **Surface Width**

   Figure 6.a shows the distribution plot for Surface Width. The most predominant surface width for this network is 24 feet, which was used on (98%) of the network's sections. The remaining 2% sections have a surface width of 26ft.

2. **Shoulder Width Outside**

   Figure 6.b shows the distribution plot for the Shoulder Width Outside. Almost the entire network (99%) has outside shoulder width of 10 feet. The remaining 1% of the network has no outside shoulder widths.

3. **Median Width**

   Figure 6.c shows the distribution plot for the Median Width. For this network the median widths varied from 2ft-30ft. The predominant widths are 18 ft (20%), 17ft (13%), and 10ft (13%). Other widths cover the remaining (54%) of the network sections.

4. **AADT**

   Figure 6.d shows the distribution plot for AADT. The AADT range for this network is from 0 to 20000 vehicles. The most predominant AADT ranges are 0 to 5000 vehicles (43%) and 6000 to 10000 vehicles (41%). The remaining sections (16%) have AADT between 10000 and 20000.

5. **Heavy Vehicles**

   Figure 6.e shows the distribution plot for percentage of Heavy Vehicles (%HV). 45% of the sections have %HV between 6% and 10%. The remaining sections have %HV ranging from 0-35%.

6. **Speed limit**

   Figure 6.f shows the distribution plot for Speed Limit. In this network, the 5-year average speed limits varied from 45 mph to 60 mph. The most predominant speed limit is 55 mph (60%). Please note that the speed limits have been averaged over a 5 year period and hence we may have values that are not divisible by 5.

6.20 Training Methodology

The initial number of input nodes started at 29 nodes (representing the 29 input variables) and iterations were carried out to identify the best possible network structure. The final structure was obtained at 1000 iterations and 4 hidden nodes. The initial input variables are shown in Table 6. and the final input variables are shown in Table 6.

Figure 6. shows the structure of the final MLP neural network model and Equation 6.5 gives the structure of the final model.

$$(Y_{1,}\ Y_2, Y_3, Y_4) = ANN_{14\text{-}4\text{-}4}(X_1, X_2, X_3, X_4\ \ldots X_{14}) \qquad \text{Equation 6.5}$$

Where:

$$\text{Outputs} = \left\{ \begin{array}{l} Y_1 = \text{ Severe Injury Crash Rate,} \\ Y_2 \ = \text{ Injury Crash Rate,} \\ Y_3 \ = \text{ Fatal Crash Rate, and} \\ Y_4 \ = \text{ Total Crash Rate} \end{array} \right\}$$

$$\text{Inputs} = \left\{ \begin{array}{l} X_1 = \text{Section Length,} \\ X_2\text{-}X_5 \ = \text{ Route Class,} \\ X_6 = \text{Surface Width,} \\ X_7 \ = \text{ Shoulder Width Outside,} \\ X_8\text{-}X_{10} \ = \text{ Median Type,} \\ X_{11} \ = \text{ Median Width,} \\ X_{12} \ = \text{ Average ADT,} \\ X_{13} \ = \text{ Average Percentage of Heavy Vehicles,} \\ X_{14} \ = \text{ Average Speed Limit} \end{array} \right\}$$

14-4-4 represents 14 Inputs, 4 Hidden Nodes, 4 Outputs. The R^2 values for the model and the individual crash rates are: R^2 (Model) = 0.6806, R^2 (SICR) = 0.8245, R^2 (ICR) = 0.9190, R^2 (FCR) = 0.0655, R^2 (TCR) = 0.9136

6.21 Sensitivity Analysis

After the final structure is obtained, sensitivity analysis was performed on the input variables in the final model and Table 6. gives the results for categorical variables. The results for continuous variable are shown in Figure 6. a-f.

6.21.1 Discussion- Sensitivity of Categorical Variables

The ranking reported herein are based on the CCPI values listed in Table 6..

1. **Route Class**

   From crash potential point of view, Route Class B is worst, Route Class C is moderate and Route Classes D and E are best. The crash rates are expected to be higher on Route Class B, moderate on Route Class C and low on Route Classes D and E.

2. **Median Type**

   From crash potential point of view, Barrier (Median Type 6) is the worst, Raised Median with Curbs (Medina Type 3) is moderate and Depressed Median (Median Type 8) is the best. The findings are not consistent with literature.

6.21.2 Discussion- Sensitivity of Continuous Variables

1. **Surface Width**

   TCR/ICR decreases steeply with increase in AADT. SICR remains statistically unchanged for the entire study range.

2. **Shoulder Width Outside**

   TCR decreases steeply with increase in shoulder width outside. ICR/SICR also decrease with increase in shoulder width, but the decrease is not as steep as TCR.

3. **Median Width**

TCR decreases steeply up to 20 feet and then starts increasing. ICR/SICR decrease up to 20 feet and then remain unchanged after that. The results are consistent with other networks.

4. **AADT**

   TCR/ICR increases up to 10000 vehicles and then decrease. The findings are not consistent with literature

5. **Percentage of Heavy Vehicles**

   TCR/ICR increase with increase in heavy vehicles up to 12% and then decrease. SICR remains statistically same for the entire study range. The findings are not consistent with literature.

6. **Speed Limit**

   TCR/ICR increase with increase in speed limit up to 56mph and then decrease. SICR remains statistically same for the entire study range. The findings are not consistent with literature.

It is to be noted that in all cases considered herein, FCR's have very limited datasets and therefore it is difficult to generalize their behavior.

6.22 Conclusions

The models developed for all the networks are quiet different from each other. This tells us that each network behaves differently and the crash rate patterns are not the same throughout. The sensitivity analysis for each network provides information about how each variable in the model influences the crash rates. This in turn may provide engineers with a preliminary idea of the variables to control/alter in case of problematic situations. Also, the crash prediction models would be very helpful in future planning operations and when undertaking any major reconstruction works.

Table 6.1: Initial Input Variables-Rural 2 Lane Network

Node #	Input Variable	Type	Value
1	Section Length Miles	Continuous	Numerical
2	Surface Width	Continuous	Numerical
3-6	Route Class	Categorical [4]*	Binary
7-9	Surface Type	Categorical [3]*	Binary
10	Shoulder Width Outside	Continuous	Numerical
11	Shoulder Width Inside	Continuous	Numerical
12-24	Shoulder Type Outside	Categorical [13]*	Binary
25-37	Shoulder Type Inside	Categorical [13]*	Binary
38	Additional Surface Width Inside	Continuous	Numerical
39	Additional Surface Width Outside	Continuous	Numerical
40	Average ADT	Continuous	Numerical
41	Average % of Heavy Vehicles	Continuous	Numerical
42	Average Speed Limit	Continuous	Numerical
43-46	Sub Section	Categorical [4]*	Binary

Table 6.2: Final Input Variables-Rural 2 Lane Network

Node #	Input Variable	Type	Value
1	Section Length Miles	Continuous	Numerical
2	Surface Width	Continuous	Numerical
3-6	Route Class	Categorical [4]*	Binary
7	Shoulder Width Outside	Continuous	Numerical
8-20	Shoulder Type Outside	Categorical	Binary
21	Average ADT	Continuous	Numerical
22	Average % of Heavy Vehicles	Continuous	Numerical
23	Average Speed Limit	Continuous	Numerical

* Number in parenthesis refers to the No. of categorical variables

Table 6.3: Ranking of Categorical Variables- Rural 2 Lane

Route Class Rankings (Worst to Best)
Route Class E (CCPI = 4.5)
Route Class D (CCPI = 4.2)
Route Class C (CCPI = 3.2)
Route Class B (CCPI = 2.8)
Shoulder Type Outside Ranking (Worst to Best)
Aggregate Base Stabilized (A.B.S) and Gutter (CCPI = 5)
Calcium Chloride with Limestone (CCPI = 4.8)
Three Feet Bitumen + Remainder Turf (CCPI = 4.6)
A.B.S (Wedge) + Remainder Turf (CCPI = 4.6)
P.C.C. Shoulder (CCPI = 4.3)
Two Feet Bitumen + Remainder Turf (CCPI = 4.1)
B.B and C/G (CCPI = 4.13)
Turf (CCPI = 4.04)
Aggregate 1" with CaCl2 (3R), LT 6" (CCPI = 3.92)
Three Feet Bitumen + Remainder Aggregate (CCPI = 3.87)
Two Feet Bitumen + Remainder Aggregate (CCPI = 3.78)
Four Feet Bitumen + Remainder Turf (CCPI = 3.79)
One Foot Bitumen + Remainder Turf (CCPI = 3.38)

Table 6.4: Initial Input Variables Used in the Rural Freeways MLP Network

Node #	Input Variable	Type	Value
1	Section Length Miles	Continuous	Numerical
2	Surface Width	Continuous	Numerical
3-5	Route Class	Categorical [3]*	Binary
6-8	Surface Type	Categorical [3]*	Binary
9	Shoulder Width Outside	Continuous	Numerical
10	Shoulder Width Inside	Continuous	Numerical
11-13	Shoulder Type Outside	Categorical [3]*	Binary
14-16	Shoulder Type Inside	Categorical [3]*	Binary
17	Additional Surface Width Inside	Continuous	Numerical
18	Additional Surface Width Outside	Continuous	Numerical
19	Average ADT	Continuous	Numerical
20	Average % of Heavy Vehicles	Continuous	Numerical
21	Average Speed Limit	Continuous	Numerical
22-25	Sub Section	Categorical [4]*	Binary

* Number in parenthesis refers to the number of categorical variables included

Table 6.5: Final Input Variables Used in the Rural Freeways MLP Network

Node #	Input Variable	Type	Value
1	Section Length Miles	Continuous	Numerical
2-4	Route Class	Categorical [3]*	Binary
5-7	Shoulder Type Inside	Categorical [3]*	Binary
8-9	Median Type	Categorical [2]*	Binary
10	Median Width	Continuous	Numerical
11	Average ADT	Continuous	Numerical
12	Average % of Heavy Vehicles	Continuous	Numerical
13	Average Speed Limit	Continuous	Numerical

Table 6.6: Ranking of Categorical Variables- Rural Freeways

Route Class Rankings (Worst to Best)
Route Class A (CCPI = 3.5224)
Route Class C (CCPI = 3.3255)
Route Class B (CCPI = 3.1635)
Median Type Rankings (Worst to Best)
Depressed Median (CCPI=4.033)
Barrier (CCPI=3.282)
Shoulder Type Outside Ranking (Worst to Best)
Bituminous Base (CCPI = 4.033)
Asphaltic Concrete Shoulder (CCPI = 3.296)
P.C.C. Shoulder (CCPI = 3.469)

Table 6.7: Initial Input Variables Used in the Rural KTA Network

Node #	Input Variable	Type	Value
1	Section Length Miles	Continuous	Numerical
2	Surface Width	Continuous	Numerical
3	Route Class	Categorical [1]*	Binary
4	Shoulder Width Outside	Continuous	Numerical
5	Shoulder Width Inside	Continuous	Numerical
6	Shoulder Type Outside	Categorical [1]*	Binary
7	Shoulder Type Inside	Categorical [1]*	Binary
8	Additional Surface Width Inside	Continuous	Numerical
9	Additional Surface Width Outside	Continuous	Numerical
10	Median Type	Categorical [1]*	Binary
11	Median Width	Continuous	Numerical
12	Average ADT	Continuous	Numerical
13	Average % of Heavy Vehicles	Continuous	Numerical
14	Average Speed Limit	Continuous	Numerical
15-17	Sub Section	Categorical [3]*	Binary

Table 6.8: Final Input Variables Used in the Rural KTA Network

Node #	Input Variable	Type	Value
1	Section Length Miles	Continuous	Numerical
2-4	Sub Section	Categorical [3]*	Binary
5	Average ADT	Continuous	Numerical
6	Average % of Heavy Vehicles	Continuous	Numerical
7	Average Speed Limit	Continuous	Numerical

Table 6.9: Ranking of Categorical Variables- Rural KTA Network

SS Rankings (Worst to Best)
SS (2) (CCPI = 4.329)
SS (1) (CCPI = 4.165)
SS (3) (CCPI = 4.050)

* Number in parenthesis refers to the number of categorical variables included

* Number in parenthesis refers to the number of categorical variables included

Table 6.10: Initial Input Variables Used in the Urban Freeways MLP Network

Node #	Input Variable	Type	Value
1	Section Length Miles	Continuous	Numerical
2	Surface Width	Continuous	Numerical
3-7	Route Class	Categorical [5]*	Binary
8	Shoulder Width Outside	Continuous	Numerical
9	Shoulder Width Inside	Continuous	Numerical
10-13	Shoulder Type Outside	Categorical [4]*	Binary
14-17	Shoulder Type Inside	Categorical [4]*	Binary
18	Additional Surface Width Inside	Continuous	Numerical
19	Additional Surface Width Outside	Continuous	Numerical
20-22	Median Type	Categorical [3]*	Binary
23	Median Width	Continuous	Numerical
24	Average ADT	Continuous	Numerical
25	Average % of Heavy Vehicles	Continuous	Numerical
26	Average Speed Limit	Continuous	Numerical
27-30	Sub Section	Categorical [4]*	Binary

Table 6.11: Final Input Variables Used in the Urban Freeways MLP Network

Node #	Input Variable	Type	Value
1	Section Length Miles	Continuous	Numerical
2	Surface Width	Continuous	Numerical
3-7	Route Class	Categorical [5]*	Binary
8	Shoulder Width Inside	Continuous	Numerical
9-12	Shoulder Type Outside	Categorical [4]*	Binary
13-15	Median Type	Categorical [3]*	Binary
16	Median Width	Continuous	Numerical
17	Average ADT	Continuous	Numerical
18	Average % of Heavy Vehicles	Continuous	Numerical
19	Average Speed Limit	Continuous	Numerical

* Number in parenthesis refers to the number of categorical variables included

Table 6.12: Ranking of Categorical Variables- Urban Freeways

Route Class Rankings (Worst to Best)
Route Class E (CCPI = 51.1)
Route Class C (CCPI = 39.3)
Route Class A (CCPI = 25.7)
Route Class D (CCPI = 4.47)
Route Class B (CCPI = 3.47)
Median Type Rankings (Worst to Best)
Depressed Median (CCPI=28.26)
Barrier (CCPI=14.87)
Raised Median With Curbs (CCPI=2.86)
Shoulder Type Outside Ranking (Worst to Best)
Bituminous Base (CCPI = 38.70)
Asphaltic Concrete Shoulder (CCPI = 34.07)
P.C.C. Shoulder (CCPI = 10.91)
Curb and Gutter (CCPI=4.6)

Table 6.13: Initial Input Variables Used in the Urban Expressways MLP Network

Node #	Input Variable	Type	Value
1	Section Length Miles	Continuous	Numerical
2	Surface Width	Continuous	Numerical
3-6	Route Class	Categorical [4]*	Binary
7	Shoulder Width Outside	Continuous	Numerical
8	Shoulder Width Inside	Continuous	Numerical
19-12	Shoulder Type Outside	Categorical [4]*	Binary
13-16	Shoulder Type Inside	Categorical [4]*	Binary
17	Additional Surface Width Inside	Continuous	Numerical
18	Additional Surface Width Outside	Continuous	Numerical
19-21	Median Type	Categorical [3]*	Binary
22	Median Width	Continuous	Numerical
23	Average ADT	Continuous	Numerical
24	Average % of Heavy Vehicles	Continuous	Numerical
25	Average Speed Limit	Continuous	Numerical
26-29	Sub Section	Categorical [4]*	Binary

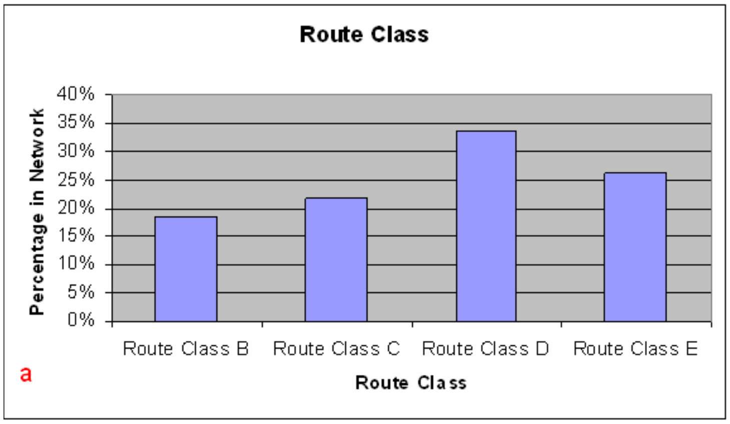
* Number in parenthesis refers to the number of categorical variables

Table 6.14: Final Input Variables Used in the Urban Expressways MLP Network

Node #	Input Variable	Type	Value
1	Section Length Miles	Continuous	Numerical
2	Surface Width	Continuous	Numerical
3-6	Route Class	Categorical [4]*	Binary
7	Shoulder Width Outside	Continuous	Numerical
8-10	Median Type	Categorical [3]*	Binary
11	Median Width	Continuous	Numerical
12	Average ADT	Continuous	Numerical
13	Average % of Heavy Vehicles	Continuous	Numerical
14	Average Speed Limit	Continuous	Numerical

Table 6.15: Ranking of Categorical Variables- Urban Expressways

Route Class Rankings (Worst to Best)
Route Class B (CCPI = 38.985)
Route Class C (CCPI = 11.815)
Route Class D (CCPI = 5.376)
Route Class E (CCPI = 5.061)
Median Type Rankings (Worst to Best)
Barrier (CCPI=26.004)
Raised Median With Curbs (CCPI=15.124)
Depressed Median (CCPI=11.856)



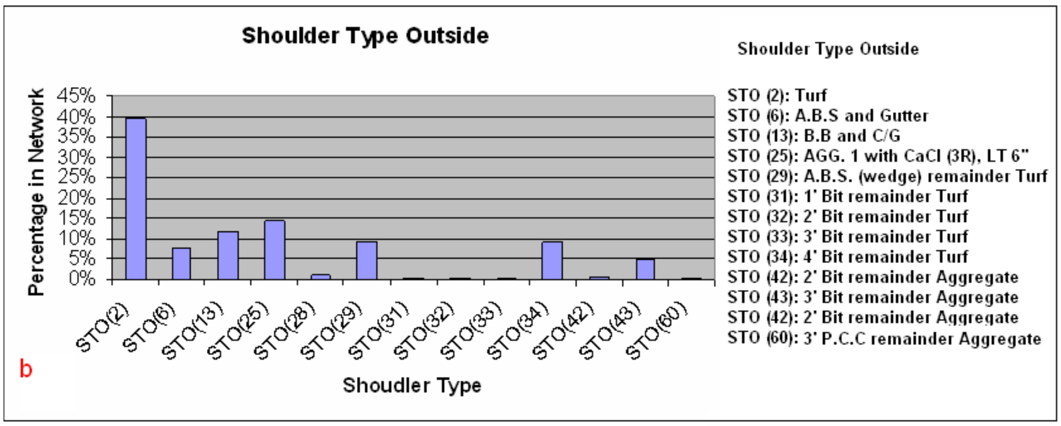


Figure 6.1: Categorical Variables Distribution-Rural 2 Lane Network

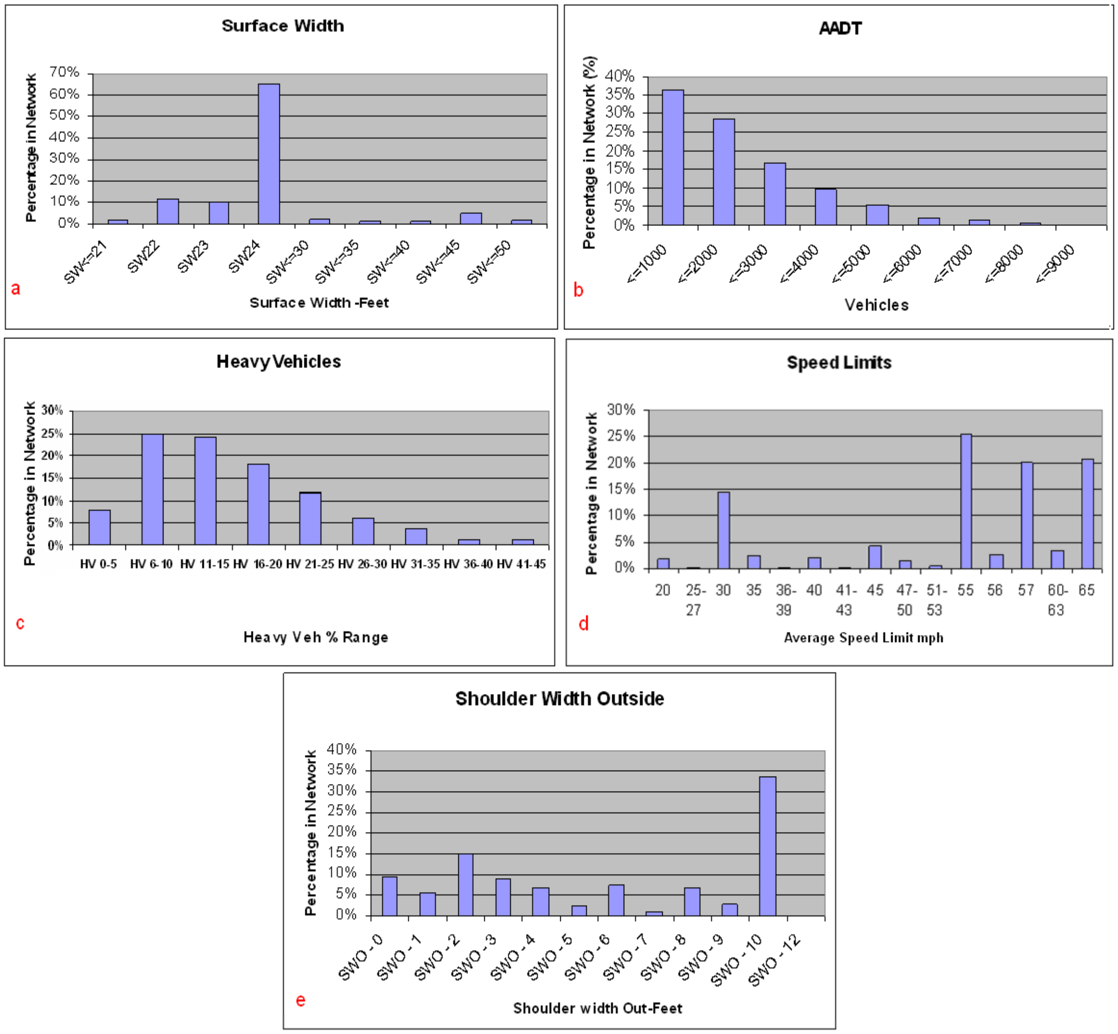


Figure 6.2: Continuous Variables Distribution-Rural 2 Lane Network

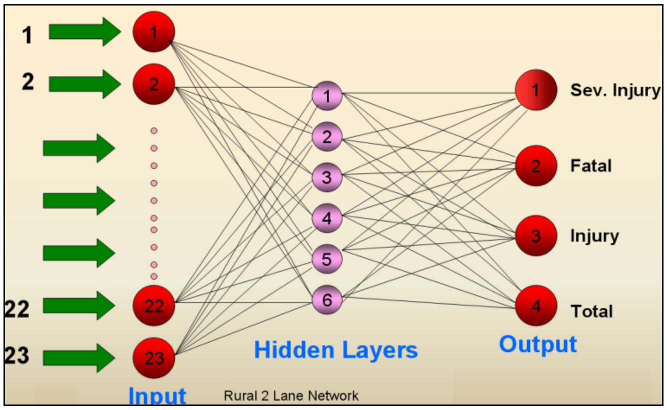


Figure 6.3: Final Network Structure-Rural 2 Lane

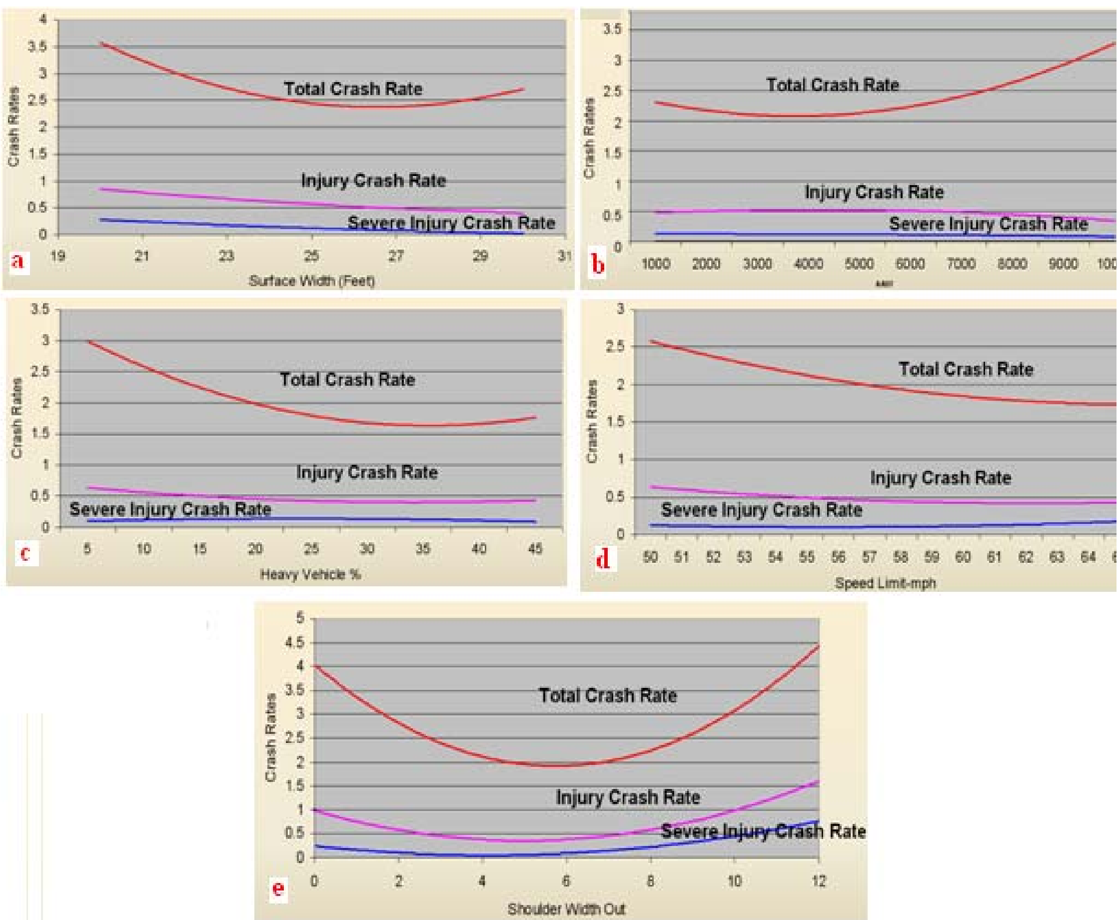


Figure 6.4: Sensitivity of Continuous Variables –I (Rural 2 Lane Network)

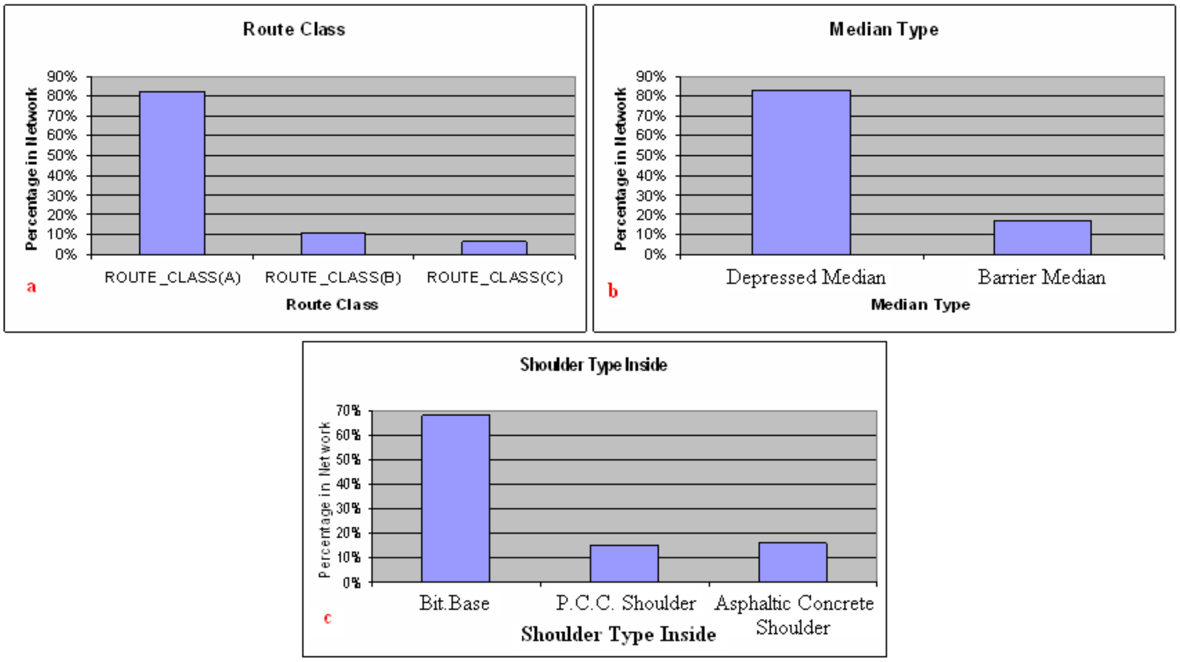


Figure 6.5: Categorical Variables Distribution-Rural Freeways Network

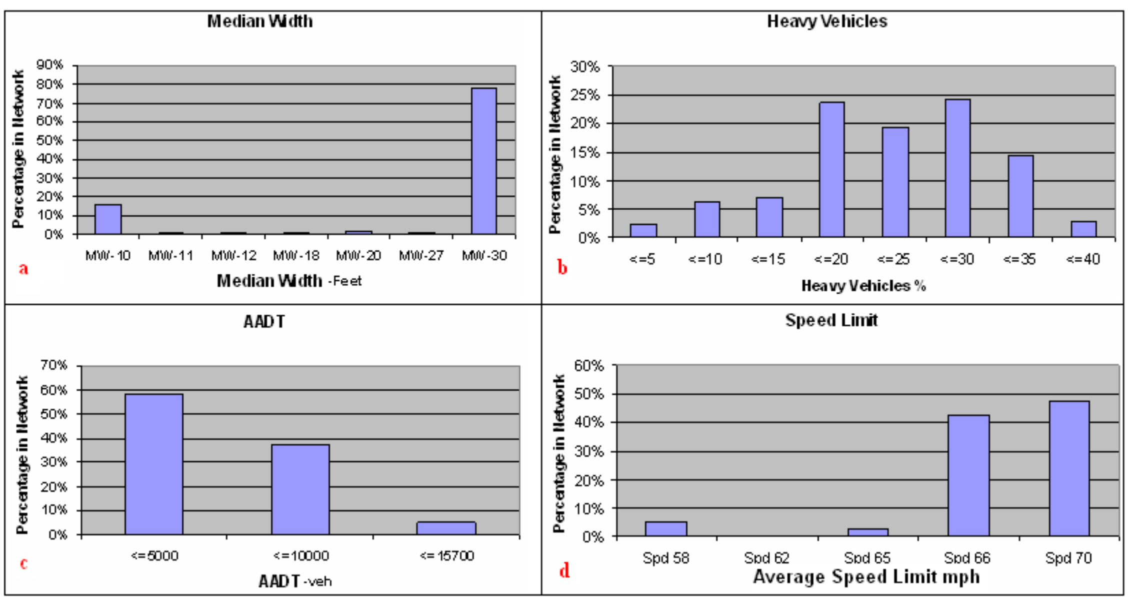


Figure 6.6: Continuous Variables Distribution-Rural Freeways Network

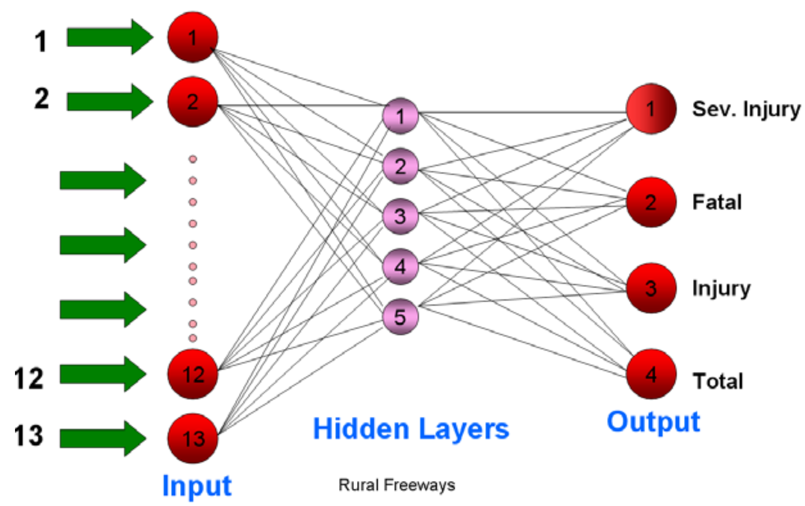


Figure 6.7: Final Network Structure- Rural Freeways

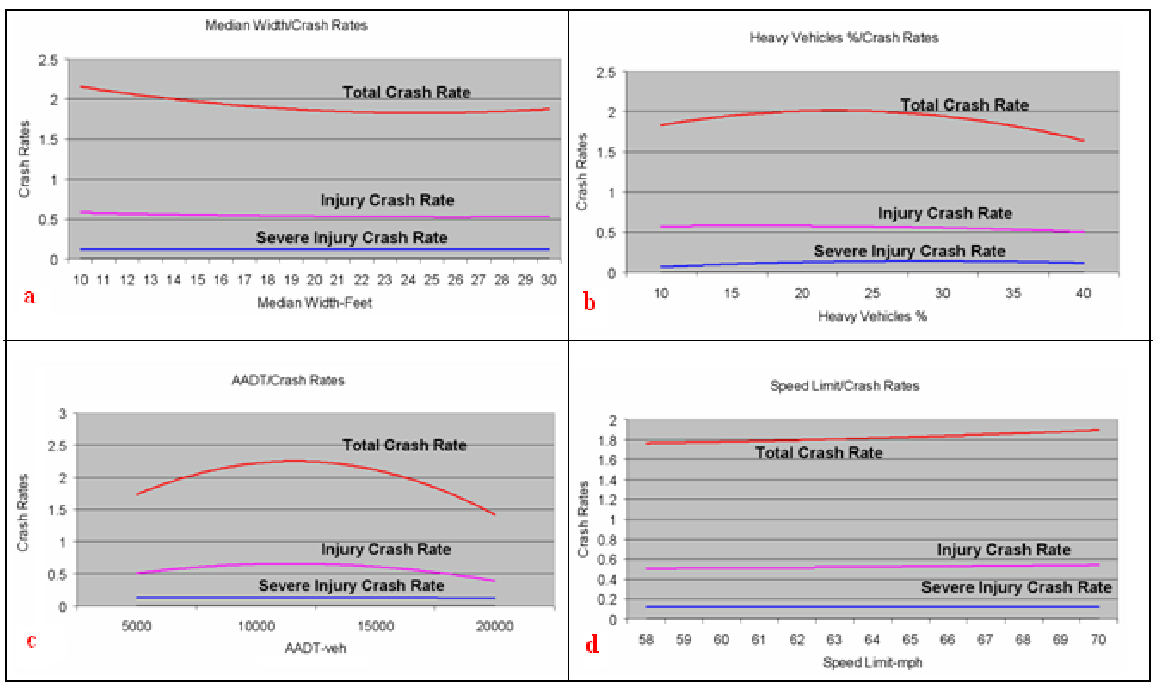


Figure 6.8: Sensitivity of Continuous Variables – (Rural Freeway Network)

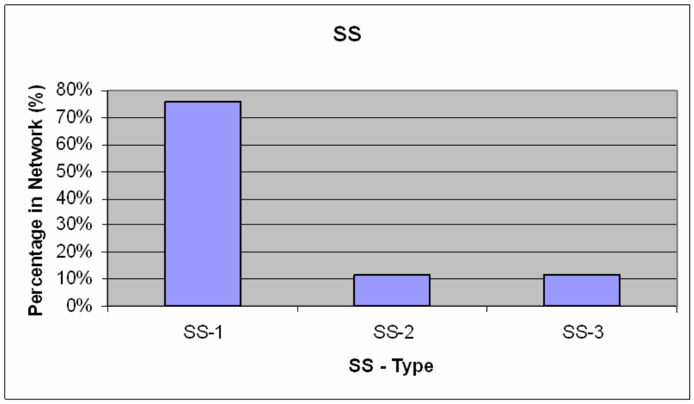


Figure 6.9: Categorical Variables Distribution- (Rural KTA Network)

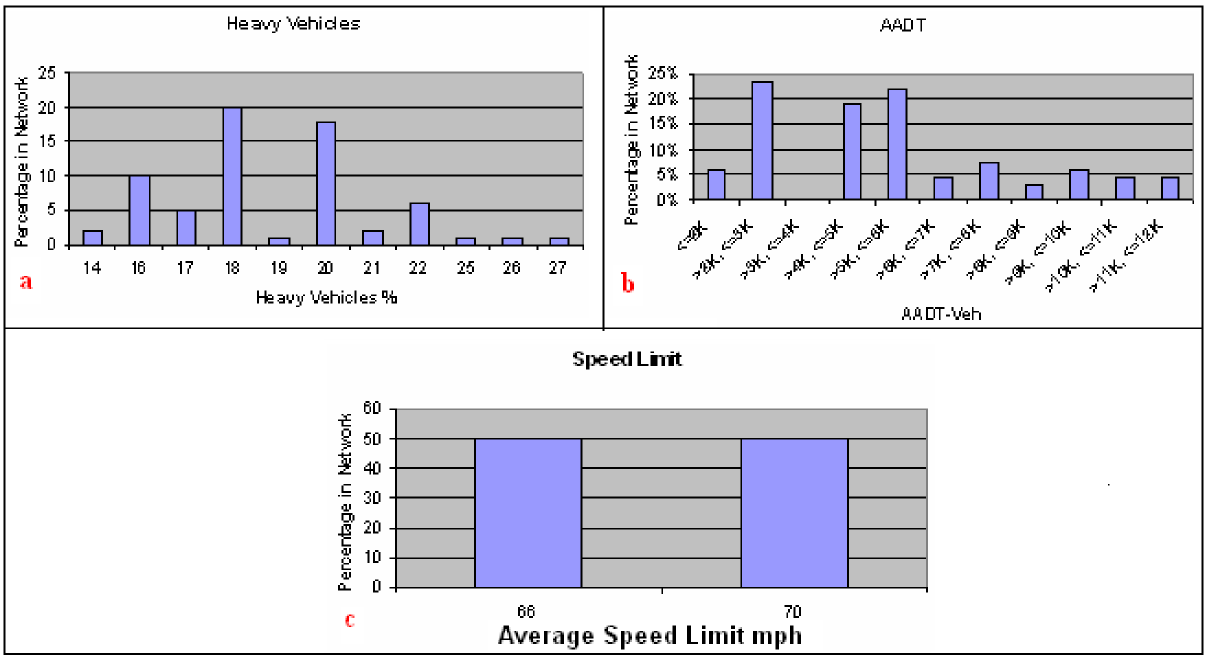


Figure 6.10: Continuous Variables Distribution- (Rural KTA Network)

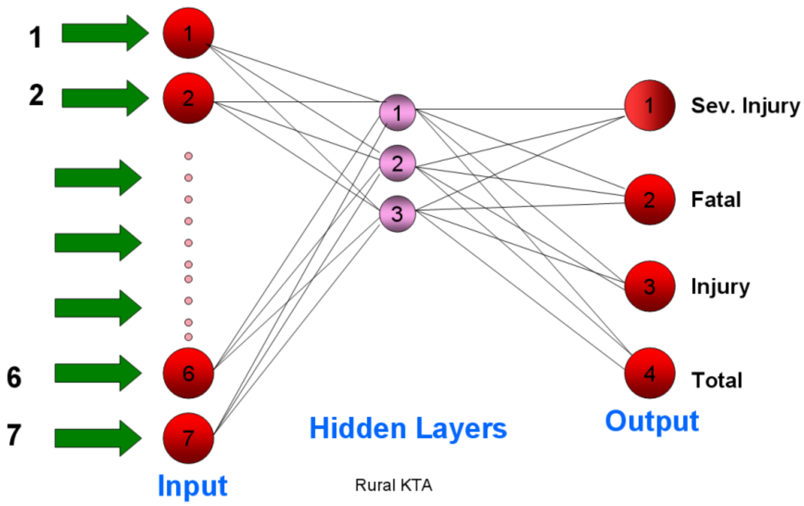


Figure 6.11: Final Network Structure-Rural KTA Network

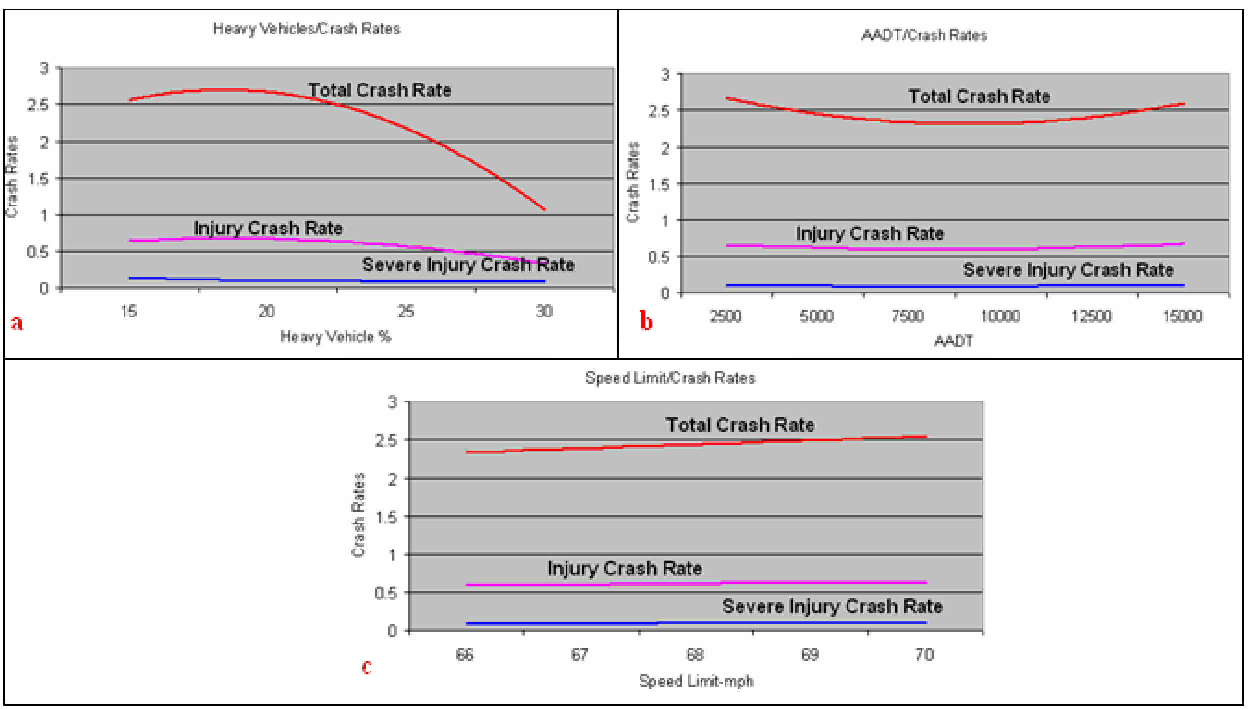


Figure 6.12: Sensitivity of Continuous Variables – (Rural KTA Network)

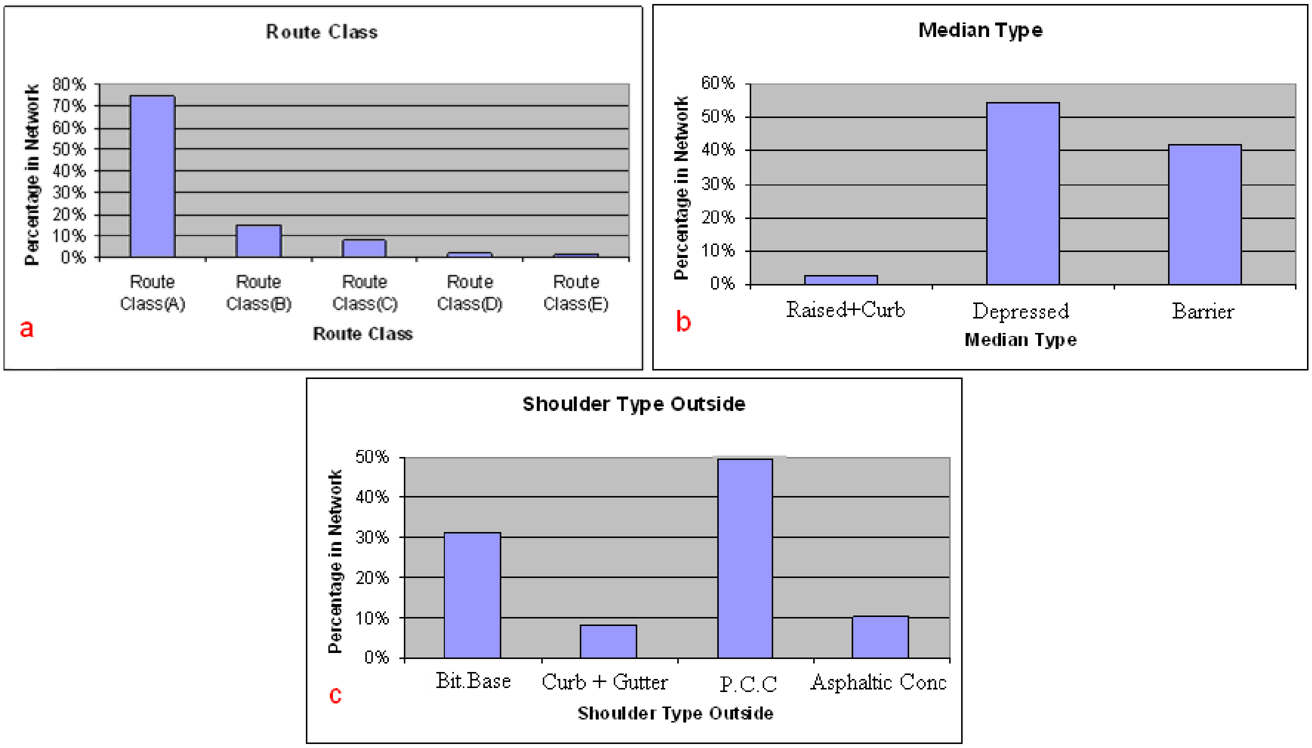


Figure 6.13: Categorical Variables Distribution- (Urban Freeways Network)

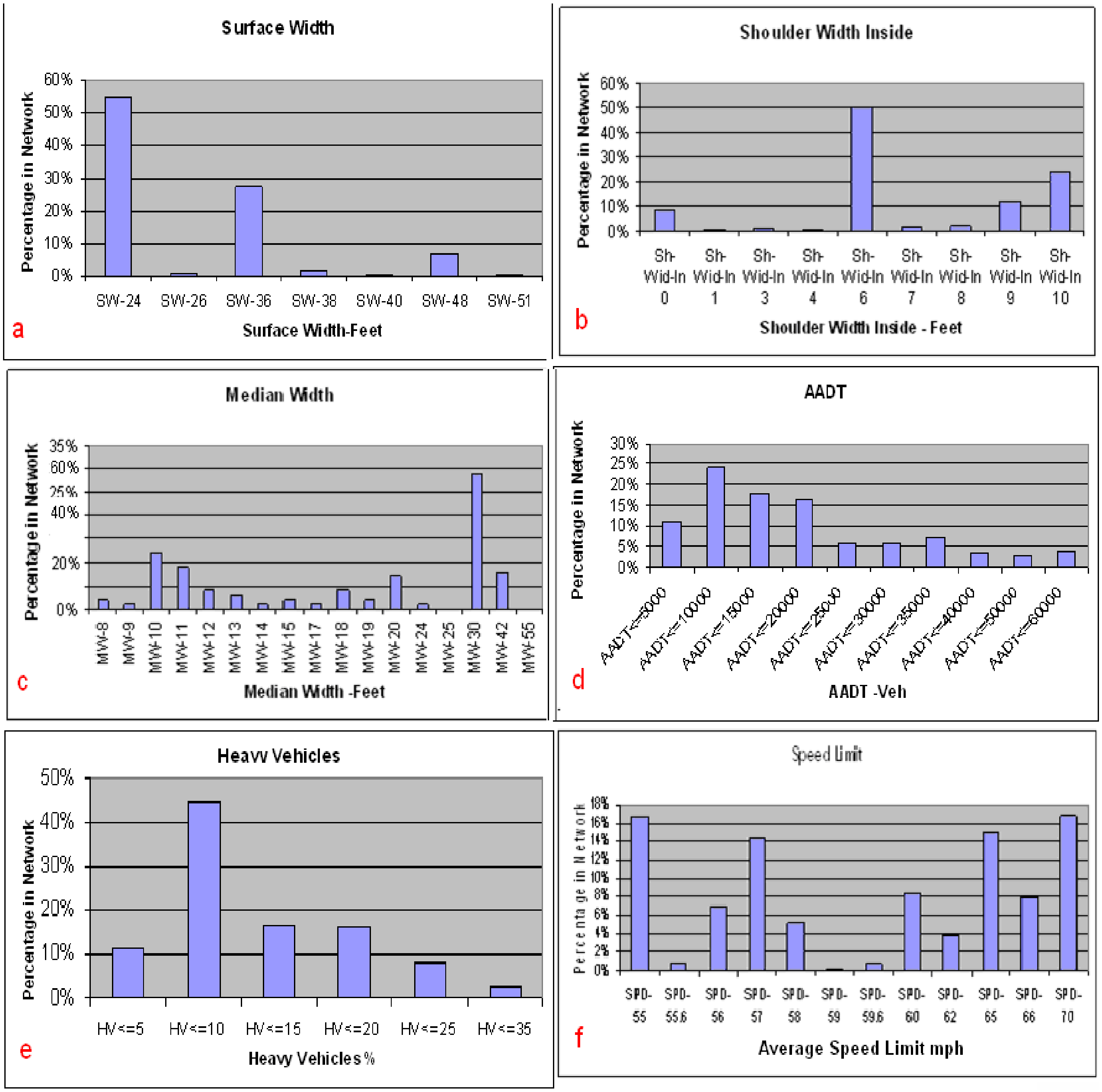


Figure 6.14: Continuous Variables Distribution- (Urban Freeways Network)

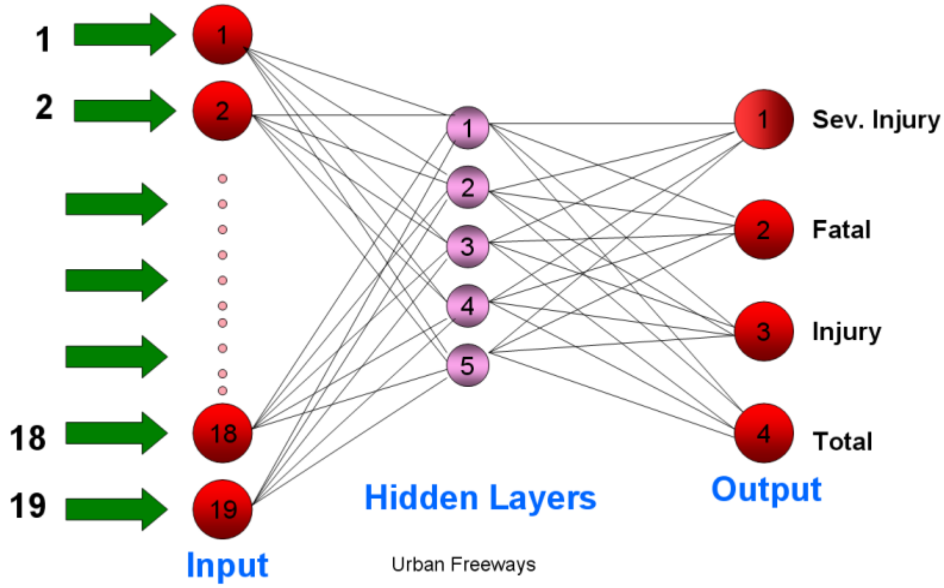


Figure 6.15: Final Network Structure-Urban Freeways Network

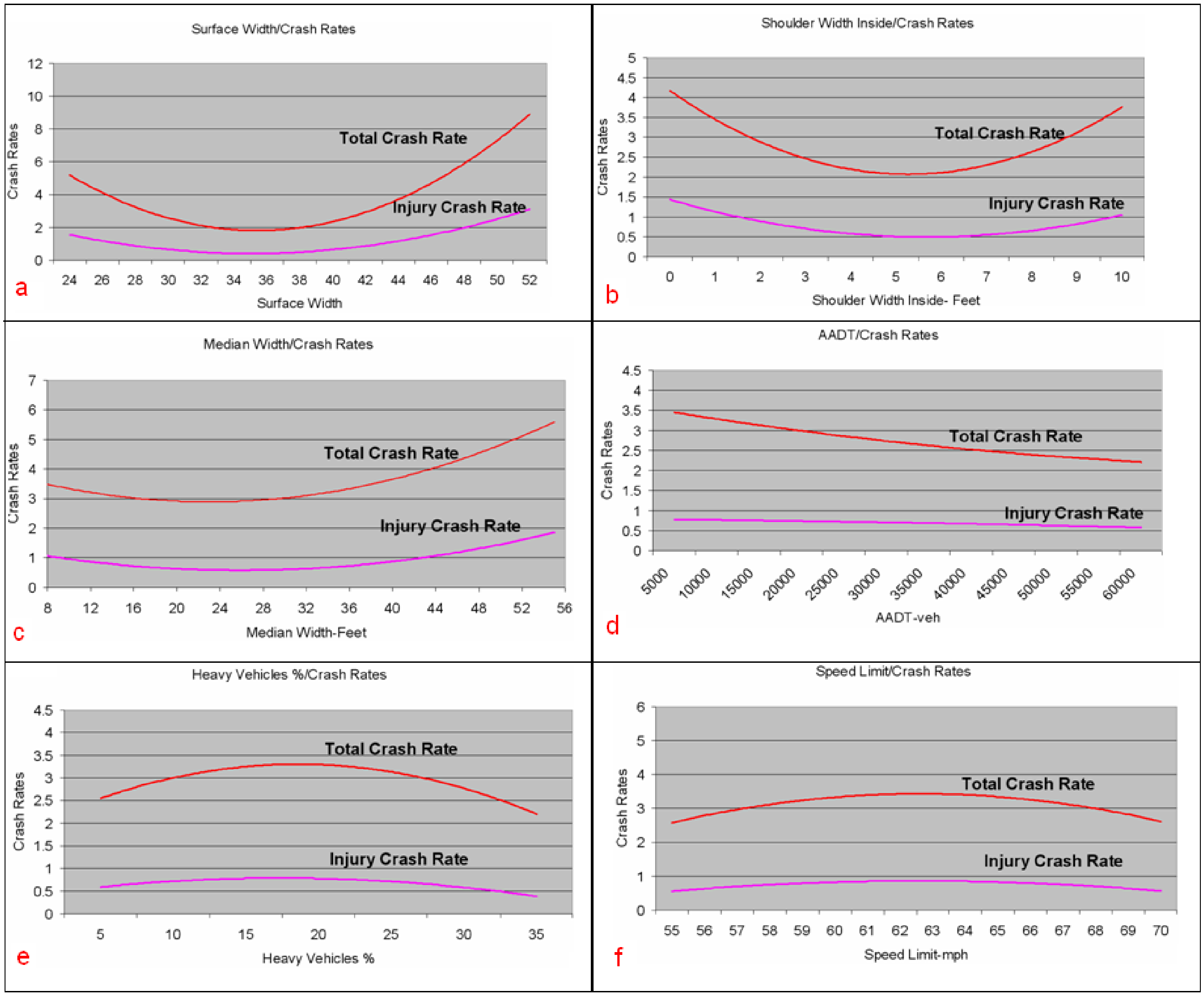


Figure 6.16: Sensitivity of Continuous Variables – (Urban Freeways Network)

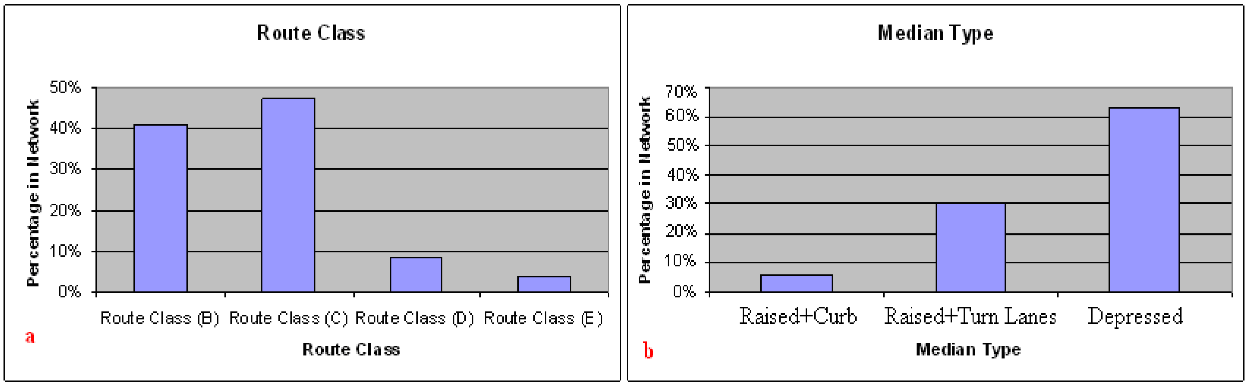


Figure 6.17: Categorical Variables Distribution- (Urban Expressways Network)

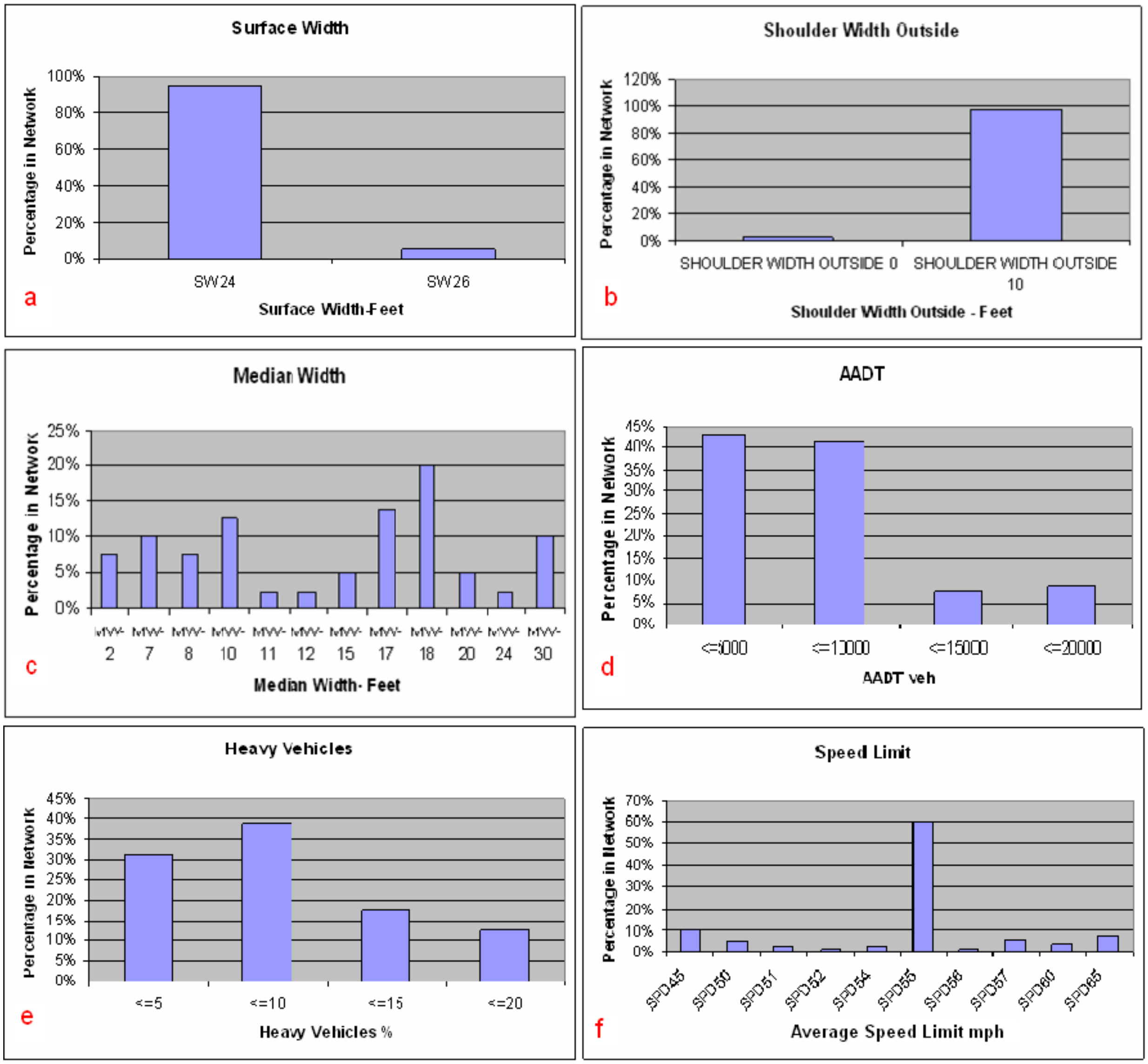


Figure 6.18:Continuous Variables Distribution- (Urban Expressways Network)

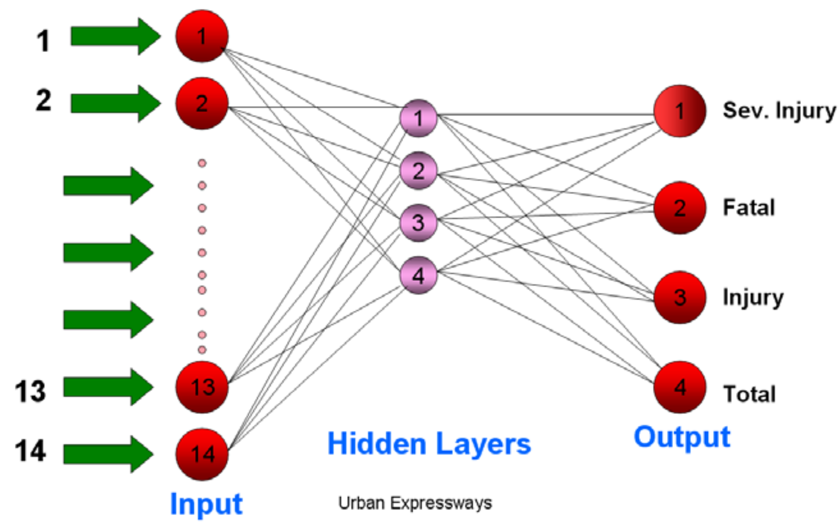


Figure 6.19: Final Network Structure-Urban Expressways Network

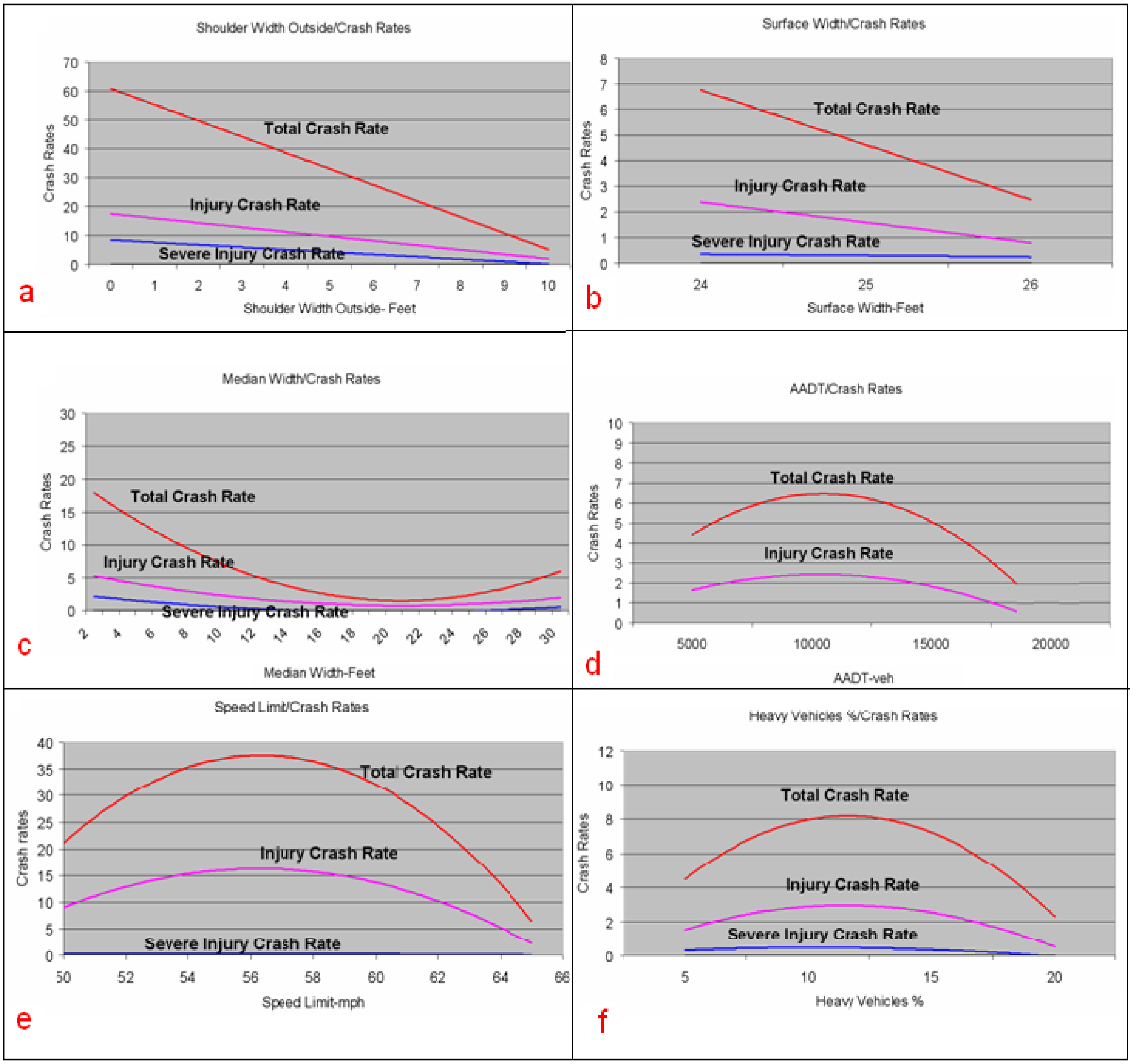


Figure 6.20: Sensitivity of Continuous Variables – (Urban Expressways Network)

CHAPTER 7 - SEAT BELT USE, DRIVE AGE, VEHICLE TYPES AND CRASH RATES

7.1 Introduction

In earlier chapters we have developed crash rate prediction models for different road networks. This chapter focuses on seat belt use, driver age, vehicle types and crash rates. The results presented in this chapter would give preliminary information on the statistics for Kansas road Networks.

7.2 Literature Review-Seat Belts

Seat belts are one of the most effective protective devices available to vehicle occupants. The National Highway Traffic Safety Administration (NHTSA) estimates that over the past 26 years, 135,000 fatalities and 3.8 million injuries in the U.S. have been prevented by seat belts. They also estimate that an additional 315,000 fatalities and 5.2 million injuries would have been prevented during this timeframe if all vehicle occupants had used seat belts. In 2001 alone, 13,274 lives were saved in crashes through the use of seat belts, and an estimated 7,334 lives could have been saved during the same time period, had all occupants used seat belts (NHTSA, 2003). As laws mandating their use have become more prevalent, seat belt usage has risen. However, according to figures complied by the NHTSA, of the 31,910 vehicle occupants killed in vehicle crashes in the United States in 2001, 60% were not wearing seat belts (NHTSA, 2003). All the above statistics suggest that the effectiveness of seat belts as a safety device is unquestionable.

7.2.1 Seat Belts-National Overview

Using seat belts is one of the most effective strategies available to the driving public for avoiding deaths and injuries in a crash (Dinh-Zarr et al., 2001). It is a federal law that all cars are equipped with seat belts. Even today, approximately one-quarter of U.S. drivers and front-seat passengers are still observed not to be buckled up (Glassbrenner, 2002). Nonusers tend to be involved in more crashes than belt users (Reinfurt et al., 1996), and belt use is lower (about 40 percent) for drivers involved in severe crashes (O'Neill, 2001). Moreover, at observed national belt use rates of 75 percent, the United States continues to lag far behind the 90 to 95 percent belt use rates achieved in Canada, Australia, and several northern European countries

7.2.2 Seat Belt Laws

To encourage seat belt use, all states except New Hampshire have seat belt laws, although laws vary from state to state (NHTSA, 2001a). Each state law falls into one of two categories: primary or secondary. States with primary enforcement laws allow officers to stop a vehicle for an observed belt violation. In states with secondary enforcement laws, an initial stop must be made for another violation before a belt citation can be issued (Ulmer et al., 1995). Several studies have noted the difference in seat belt use in primary states compared to secondary states. According to a 2002 National Occupant Protection Use Survey (NOPUS) study, the average seat belt usage rate for states with primary laws is 80% and the average rate for states with secondary laws is 69% (NHTSA, 2002). NOPUS is a Probability-based observational survey of belt use by drivers and front-seat occupants of passenger vehicles. This study has been conducted annually since 1994 and provides nationally representative data on observed belt use in passenger vehicles and some demographic detail, such as belt use by race, ethnicity, and gender (Glassbrenner 2002,)

Wells et al. (2001) studied seat belt use in four cities (two in states with primary laws and two in states with secondary laws) and found that the rates were higher in states with primary laws. Other studies have noted the change in belt use when a primary law replaces a secondary law. In January 1993, the state of California made the shift from secondary to primary enforcement. Studies conducted (Ulmer, Preusser, and Cosgrove (1995)) examined the impact of this change on seat belt usage in six communities. The percentage of drivers observed using seat belts increased from 58% before the change in the law to 76% following the change. In addition, drivers who were surveyed and had knowledge of the law said they were more likely to use seat belts than they were in the past. When Maryland, Oklahoma, and the District of Columbia changed their laws from secondary to primary, the increases in seat belt rates ranged from 9% to 14% (NHTSA, 2001a). Initial improvements, however, are not typically sustain. The rates drop after the initial surge but remain higher than the rates prior to the enactment of primary enforcement legislation (Ulmer et al. 1995, Eby and Vivoda 2001).

7.2.2 Characteristics Associated with Belt Use

Certain driver characteristics such as gender, age, education, and income have been linked to low seat belt use. Many studies have determined that females wear seat belts more often than males (NHTSA 2000, FDOT 2001, NHTSA 2003). Females have higher rates across all age groups, all vehicle types, all ethnic/racial groups, and in both primary and secondary states (Nelson et al., 1998, Ulmer et al., 1995, Eby and Vivoda

2001, Wells et al., 2002). Seat belt use has also been shown to increase with age, education, and income. Several studies have noted that older drivers tend to be more likely to wear their seat belts (NHTSA 2000, Ulmer et al., 1995, Williams et al., 1996, FDOT 2001). Seat belt usage is also affected by the type of vehicle. Usage rates in passenger cars, minivans, and SUVs are typically higher than use rates in pickup trucks. The 2002 NOPUS study (NHTSA, 2002) showed a 77% seat belt usage rate for passenger cars, 78% for vans and sport utility vehicles (SUVs), and 64% for pickup trucks. The 1998 Motor Vehicle Occupant Safety Survey (MVOSS)* (NHTSA, 2000) reported rates between 80–83% for cars, van/minivans, and SUVs, but 65% for pickup trucks (FDOT 2001, Ulmer et al. 1995, Eby and Vivoda 2001, Williams et al. 1996). Studies have also shown that seat belt rates tend to be slightly higher in urban areas than rural areas. About 80% of people in urban areas reported “all the time” use while about 77% in rural areas reported “all the time” use (NHTSA, 2000). According to the 1998 NOPUS study, as cited by the 1998 MVOSS (NHTSA, 2000), 74.5% drivers in urban areas were observed with belts compared with 67% in rural areas. A study by the Florida Department of Transportation (FDOT) (2001) determined that the rates in Florida were higher in urban counties than rural counties and slightly higher on urban roadways than rural roadways (Williams et al. 1996).

7.2.4 Reasons for Non-Use

If seat belts are so effective, why don’t more motorists buckle up? Unlike air bags or automatic restraint systems, manual belts require action on the part of drivers and passengers. Reasons for not using belts stem from a complex mix of situational, habitual, and attitudinal factors. Many drivers and vehicle occupants report that they would like to be wearing a seat belt in a crash but have not acquired the habit of buckling up on all trips. For this group (referred to hereafter as “part-time users”), belt use is situational; they tend to buckle up when the weather is poor or when they are taking longer trips on high-speed roads where they perceive driving as riskier. In surveys, these users report that the primary reasons for their not buckling up are driving short distances, forgetting, being in a hurry, or discomfort from the belt (Block 2001, v). In contrast, the much smaller group of motorists who never or rarely use their belts—the so-called “hard-core nonusers”—report negative attitudes toward seat belts as the primary reason for nonuse. These include discomfort, unfounded claims that belts are dangerous in a crash (e.g., could trap the driver in the vehicle), infringement of personal freedom and resentment of authority, and the attitude that they “just don’t feel like wearing them” (Block 2001, v). According to NHTSA’s most recent telephone survey on occupant restraint issues (Block 2001, 12), one-fifth of drivers can be characterized as part-time users, that is, they report using their belts most or some of the time, and about 4 percent as hard-core nonusers, those who report never or rarely using their belts. The latter group is small but has a high crash risk. Unbelted drivers have significantly more traffic violations, higher crash involvement rates, higher arrest rates, and higher alcohol consumption than those who buckle up all or part of the time (Reinfurt et al. 1996).

7.3 Driver Age-Literature Review

Driver age is a very important factor in crashes. Driver age mainly accounts for perceptual and cognitive skills and deficiencies of these general skills contribute to higher crash rates.

7.3.1 Young Drivers

Young drivers ages 16 to 24 constitute a high-risk driver group. According NHTSA, drivers under the age of 21 are approximately 2.5 times more likely to be killed in a crash than drivers ages 25 to 69. In 2004, 5,610 teenagers (16-19) died in the United States from crash injuries. The crash risk for young drivers is particularly high during the first years in which they are eligible for driver's licenses. They have very high rates of both fatal and nonfatal crashes compared with older drivers. This is true whether rates are based on the total number of young drivers, the number with licenses, or miles driven.

Young drivers have characteristics and capabilities that differ from those of mature drivers in several areas like cognition, motivation, attitudes, perception and control of risk, and visual search and attention.

Young drivers have been shown to be deficient in braking, steering and speed adjustment skills (Mayhew & Simpson, 1995). This is reflected in their performance on the road - they have difficulty maintaining proper lane position, in accelerating and decelerating smoothly, and in adjusting their speed to changing conditions and circumstances. Young drivers' skill deficiencies are also reflected in the quality of use of vehicle controls, including the amplitude, duration, velocity and acceleration of control movements (FORS, 1997).

Young drivers also have perceptual deficiencies relating to searching and scanning the environment and detecting hazards. Novice drivers take longer to perceive and respond to hazards, and they tend to misread the risks associated with specific situations (Mayhew & Simpson, 1995). Although young and novice drivers gain car control skills very quickly, within 20 hours of driving experience, the perception of risk or hazardous situations is acquired much more slowly (Mills, 1998).

7.3.2 Older Drivers

Older drivers are the most rapidly growing segment of the population in US. The safety problem of older drivers depends on how their performance is viewed (Transportation Research Board, 1988; Waller, 1991). Based upon numbers of licensed drivers, older drivers actually have fewer crashes and violations.

As such, based on driver records alone, older drivers appear to be safer than other age groups. However, when crash rates based on mileage driven are calculated, drivers above 75 years of age have the highest rates of any age group, including teenage drivers. Overall, older drivers have fewer numbers of crashes overall than younger drivers, but have more crashes per mile driven.

It is also the case that older drivers and older people in general, are more vulnerable to serious or fatal injury in a crash of specified dimensions (Evans, 1991). This increased vulnerability to injury begins around age 55 and increases with age (Evans, 1993; Evans, Garish, & Teheri, 1998; Partyka, 1983; Pike, 1989). Thus, even in relatively minor crashes, an older person is more susceptible to serious injury.

Older drivers are also more likely to suffer from medical disabilities that may impair their driving and they may use medications that could affect their driving performance (Leveille, et al., 1994; Neutel, 1995). Even in the absence of clear evidence of medical impairment due to the use of medications, with increasing age most older drivers experience some loss in visual perception ability (e.g., Bailey & Sheedy, 1988; Owsley & Sloane, 1990; Schieber, 1994a), decreases in cognitive functioning (e.g., Cerella, 1985; Denney & Palmer, 1981), and decreased psychomotor function (e.g., Kausler, 1991; Marottoli & Drickamer, 1993; Yee, 1985). Simply put, age takes its toll.

Older drivers, as a group, are aware of declining abilities, and take compensatory measures (Yee, 1985). They limit their driving to the safest times and places, they reduce their speed, and they may drive with a copilot (e.g., Eby & Kostyniuk, in press; Kostyniuk, Streff, & Eby, 1997; Persson, 1993). Older drivers are not characterized by risk-taking behaviors, such as speeding or reckless driving. Rather, their errors reflect impairment of perception and cognition. Yet, even though older drivers try to limit themselves to the safest times and places, their crash rate per mile driven rises dramatically with increasing age.

7.4 Current Study

In this study driver ages and vehicle types have been divided into groups. The driver ages have been separated into five groups: 14-17, 18-20, 21-34, 35-69 and 70 & greater. The vehicles have been divided into eight groups: Passenger Car (PC), Pick Up (PU), Sport utility Vehicles (UT), Minivans (VAN), Tractor Trailer (TT), Trucks (TRK), Bus (BUS) and Unknown Vehicles (UNKW).

The driver age involvement in crashes, vehicle types and crashes, seat belt use within each vehicle type, are plotted for each crash rate type and network. The ages of drivers involved in crashes are plotted as a continuous function of age and also by age group. While plotting the involvement by age group, the plots have been normalized. Normalizing the data within age groups would show the true picture and would account for the different numbers of drivers in each age group.

Since data aggregation has been used the plots would give a preliminary idea of the data for Kansas road networks.

7.5 Seat Belt Compliance Index

Seat Belt usage, by drivers, for each crash type would not give the overall picture for the network as a whole. To give the overall compliance within each network an index has been developed using the methodology followed for Investment Index discussed in Chapter 4. The weight factors for each crash rate type are same as the ones used earlier for Investment Index calculation. The Combined Seat Belt Compliance Index (CSBCI) is given in Equation 7.1.

$$CSBCI = 100 * \left[\frac{[(TSBCI) + (2 * ISBCI) + (5 * SISBCI) + (10 * FSBCI)])}{18} \right] \quad \text{Equation 7.1}$$

Where

CSBCI = Combined Seat Belt Compliance Index

TSBCI = Total Crash Rate Seat Belt Compliance Index,

ISBCI = Injury Crash Rate Seat Belt Compliance Index

SISBCI = Severe Injury Crash Rate Seat Belt Compliance Index,

FSBCI = Fatal Crash Rate Seat Belt Compliance Index

Based on the CSBCI values the seat belt compliance ranking for vehicle types is evaluated.

7.6 Discussion of Results

The plots for driver age, seat belt use and vehicle types for each network are shown in Figures 7.-7.. The results are discussed below for each network.

1. **Rural Expressways**

a. Driver Age: If we look at the continuous age distribution for the TCR (Figure 7.a), the highest values are observed only for (18-20) age group. The crash involvement decreases as the age increases. For ICR/SICR/FCR (Figures 7.a, 7.3a, and 7.a) the behavior is slightly different. This behavior is due to limited number of datasets. In order to get the exact picture, the driver age groups are normalized and plotted. For all crash rate types the driver age group (18-20) has the maximum involvement in crashes (Figures 7.b, 7.b, 7.3b and 7.b). This behavior is in conformance with the literature review. In case of ICR even the (14-17) age group has crash involvement equal to the (18-20) age group. Other age groups do not have a high crash involvement.

b. Vehicle Type: Passenger Cars have the highest crash involvement for all crash rate types. (Figures 7.c,7.c, 7.3c and 7.c)

c. Seat belt use and Vehicle Type: There is consistent pattern for seat belt compliance. (Figures 7.d,7.d, 7.3d and 7.d) In general buses and trucks have very high compliance. The overall rankings based on CSBCI values for Rural Expressway network is as follows: BUS-100%*,TRK-98%,VAN-83%,UNKW-82%,PU-71%,UT-69%,TT-67%, PC-60%.

---

* The number indicates percentage of drivers wearing seat belts within each vehicle class

2. **Rural 2 Lane**

a. **Driver Age:** If we look at the continuous age distributions, all the crash rate types have similar patterns. This network is the largest network and there are a large number of datasets. The distribution pattern for TCR and ICR (Figures 7.a and 7.a) are exactly similar. The distributions for SICR and FCR ((Figures 7.a and 7.a) Figures 7.a and 7.a) have slightly different trends. There are small variations but if we look overall they also show a decreasing trend with increase in age. The plots for normalized age groups show that (18-20) age group has highest crash involvement. (Figures 7.b,7.b,7.b and 7.b)

b. **Vehicle Type:** Passenger Cars have the highest crash involvement for all crash rate types. (Figures 7.5c,7.6c,7.7c,7.8c)

c. **Seat belt use and Vehicle Type:** The seat belt compliance results are shown in (Figures 7.d,7.d,7.d,7.d) The overall rankings based on CSBCI values for Rural 2 Lane Network is as follows: BUS-99%,TT-78%,VAN-68%,TRK-66%,PC-63%,UNKW-62%,UT-61%,PU-55%.

For all the other networks also the patterns are exactly similar. In all networks the (18-20) age groups have highest crash involvement. Among vehicle types Passenger cars have highest involvement. The seat belt compliance changes slightly for each network.

The rankings for each of the remaining networks are given below:

3. **Rural KTA:** BUS-100%,TT-99%,UT-98%,PC-97%,VAN-97%,TRK-96%,PU-89%,UNKW-68%.

4. **RuralFreeways:**BUS-100%,TT-86%,VAN-82%,TRK-78%,UT-72%,PC-65%,PU-45%,UNKW-33%
5. **UrbanExpressways:**BUS-100%,UNKW-100%,TRK-97%,VAN-90%,UT-85%,PC-74%,TT-71%,PU-52%
6. **Urban Freeways:** BUS-100%,UNKW-100%,TT-96%,PU-85%,PC-67%,TRK-63%,UT-36%,VAN-35%

7.7 Conclusions

The general conclusions on driver age, vehicle type and seat belt compliance are as follows:

1. Driver Age Group (18-20) has the highest number of crashes on all networks.
2. Passenger cars have the highest crash involvement among vehicle types.
3. Seat belt compliance is not completely consistent among all networks. Buses have the highest compliance among all networks.

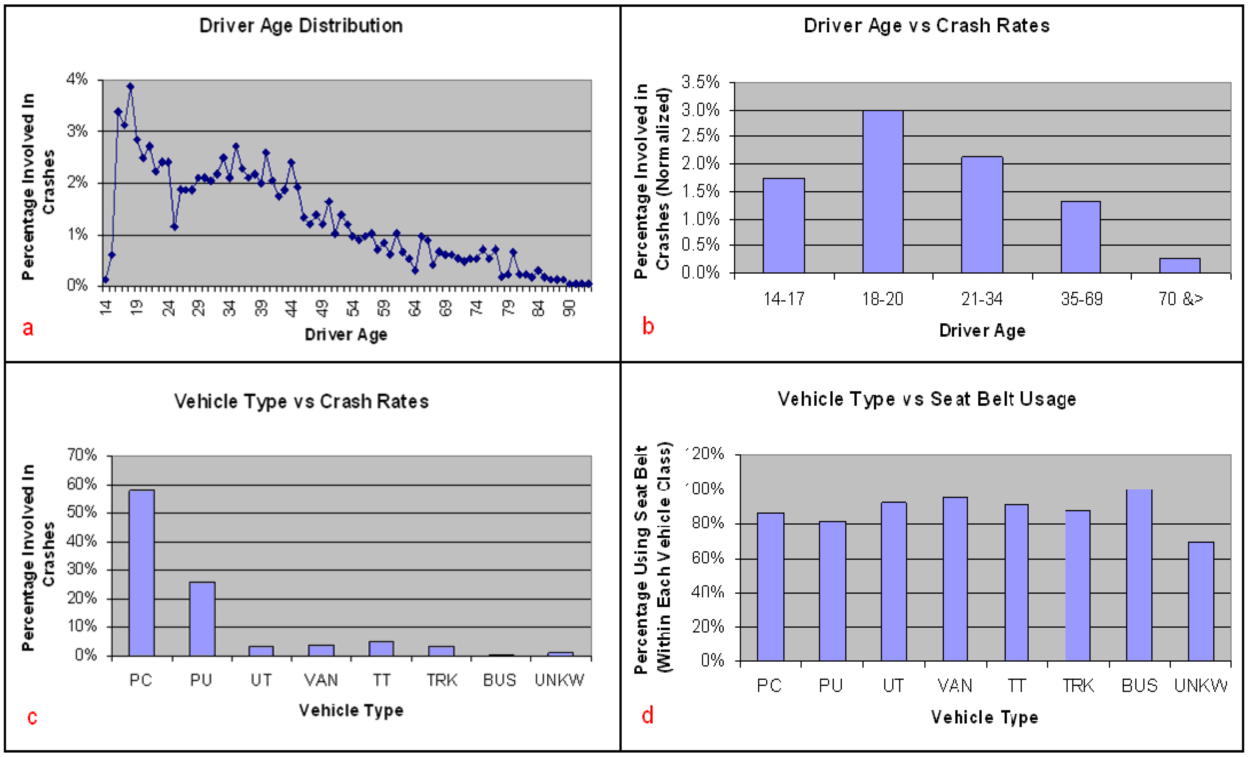


Figure 7.1: Driver Age, Seat Belt Use and Vehicle Type-Rural Expressways (TCR)

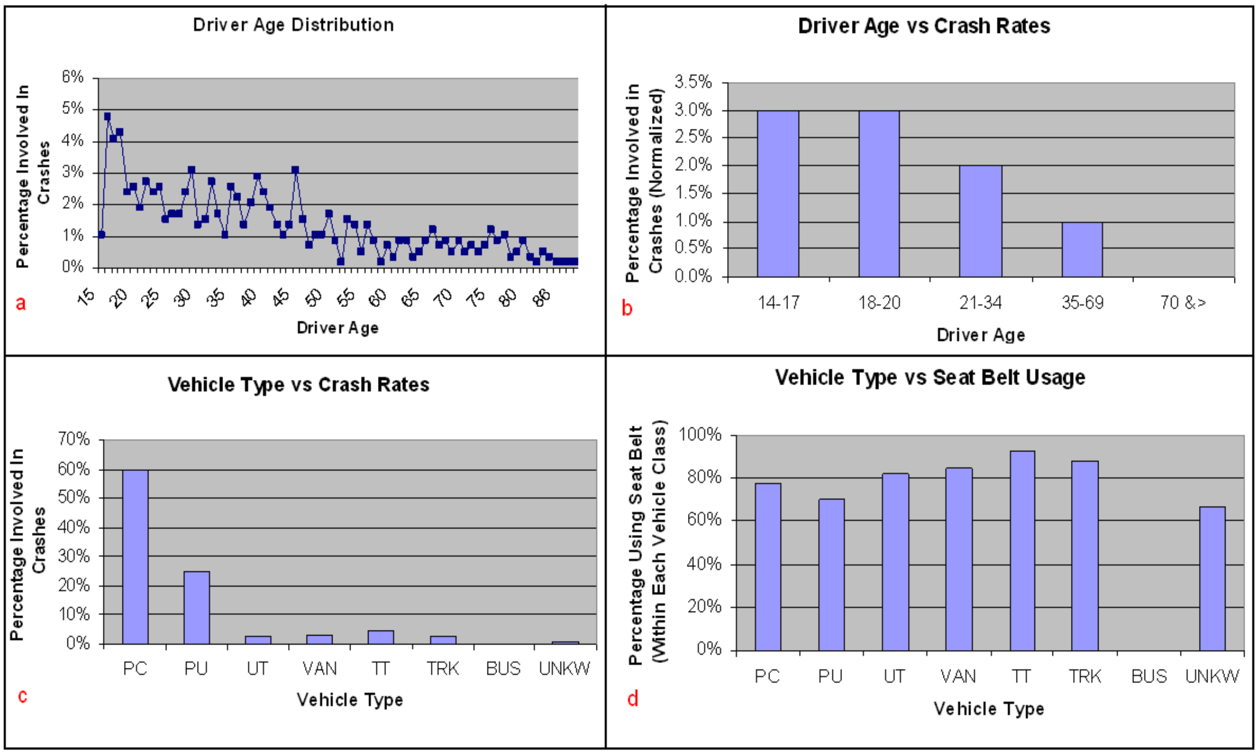


Figure 7.2: Driver Age, Seat Belt Use and Vehicle Type-Rural Expressways (ICR)

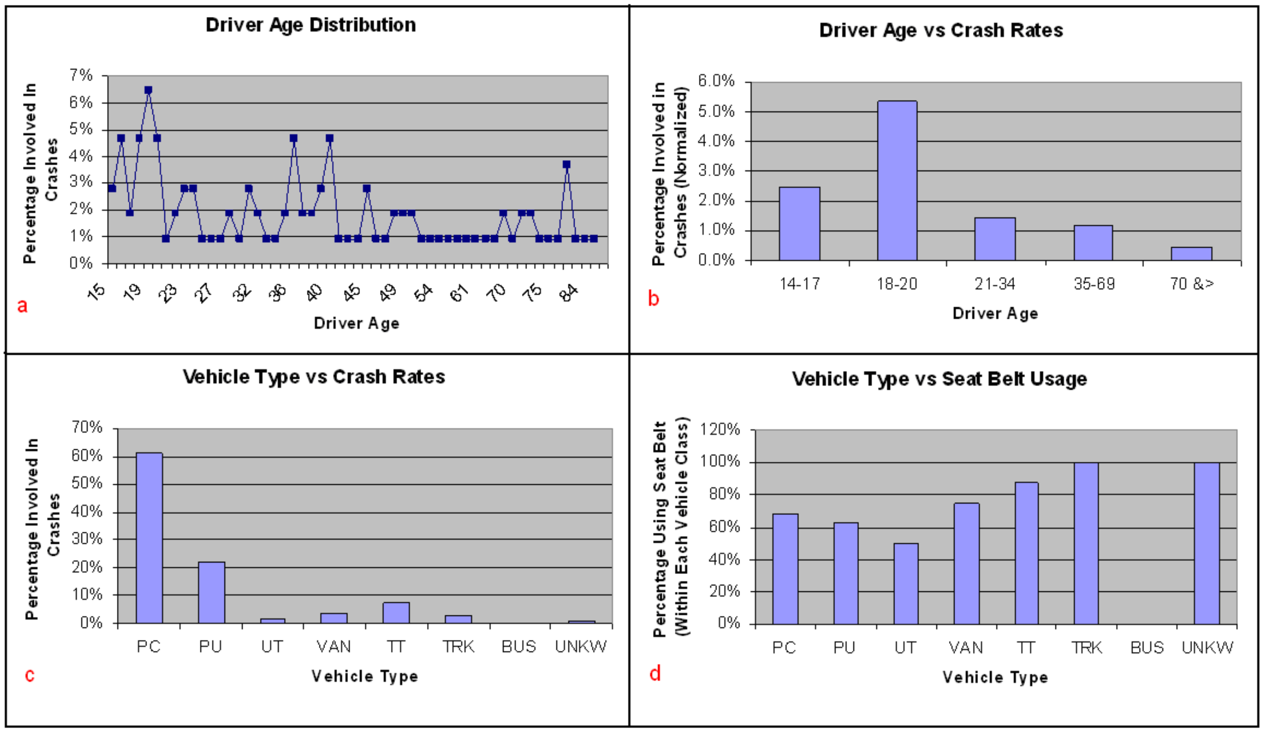


Figure 7.3: Driver Age, Seat Belt Use and Vehicle Type-Rural Expressways (SICR)

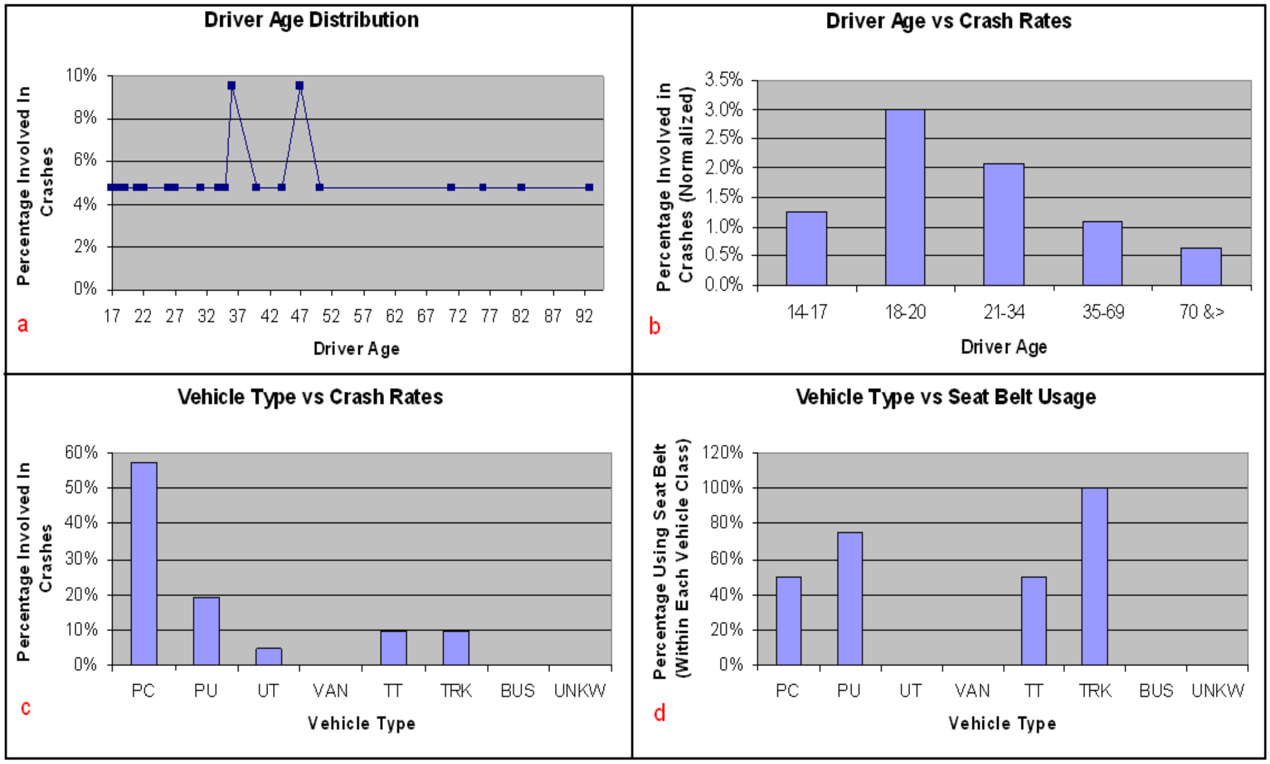


Figure 7.4: Driver Age, Seat Belt Use and Vehicle Type-Rural Expressways (FCR)

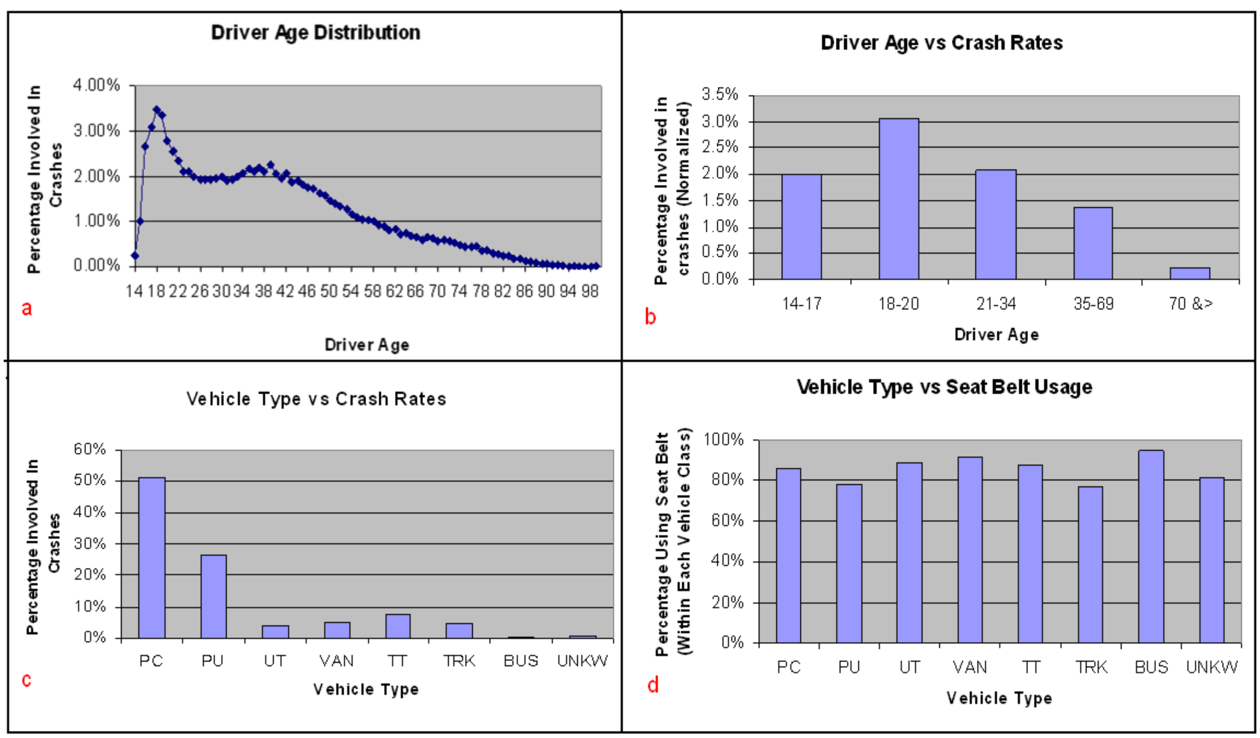


Figure 7.5: Driver Age, Seat Belt Use and Vehicle Type-Rural 2 Lane (TCR)

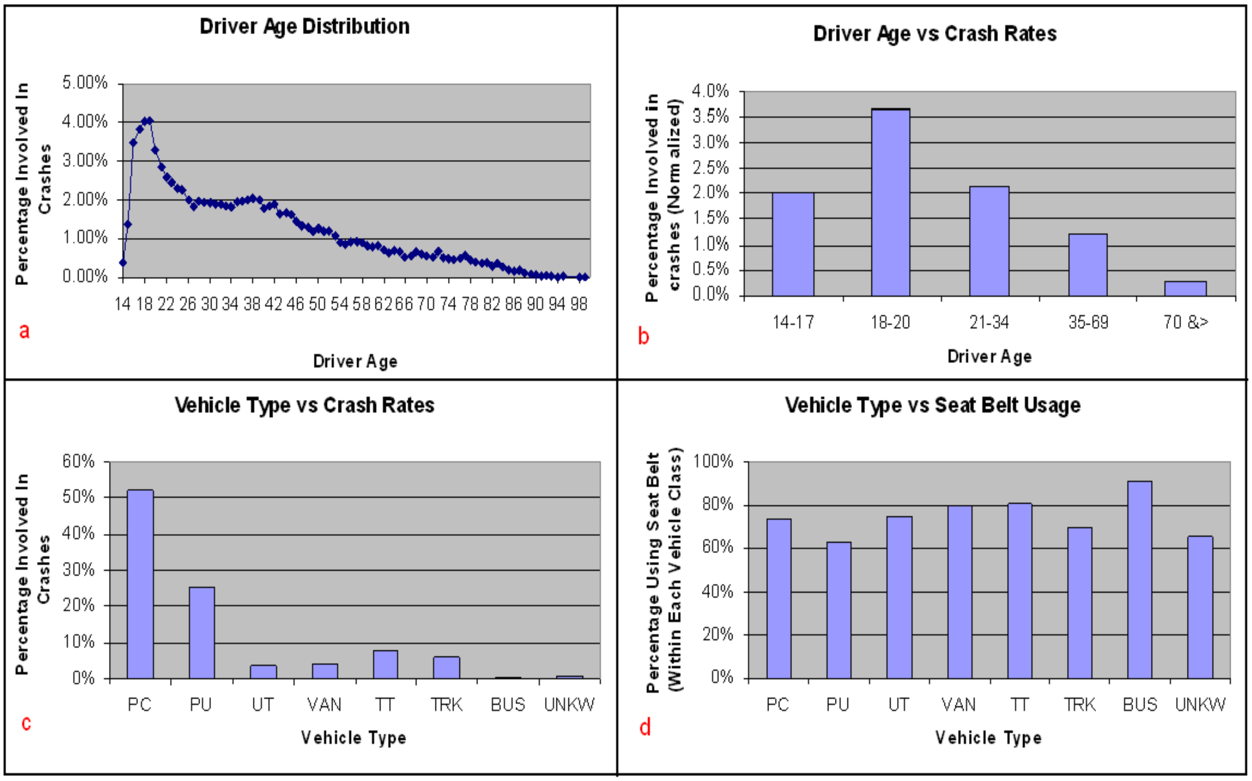


Figure 7.6: Driver Age, Seat Belt Use and Vehicle Type-Rural 2 Lane (ICR)

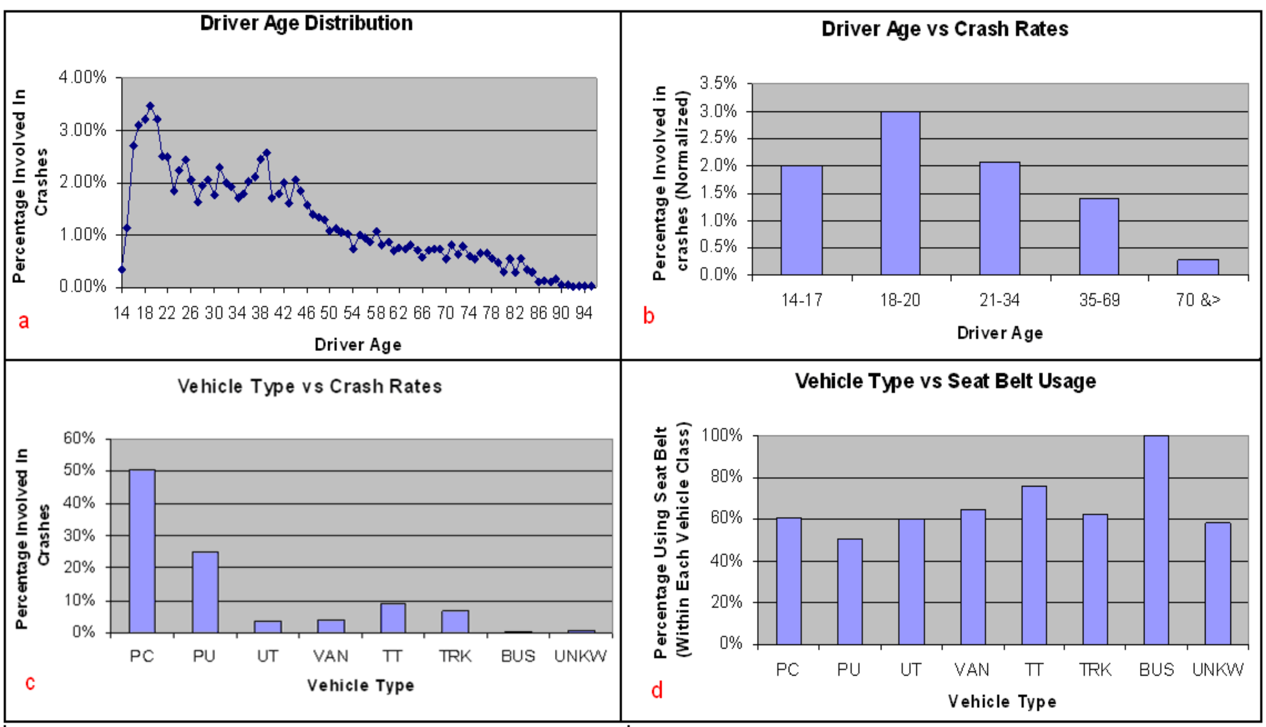


Figure 7.7: Driver Age, Seat Belt Use and Vehicle Type-Rural 2 Lane (SICR)

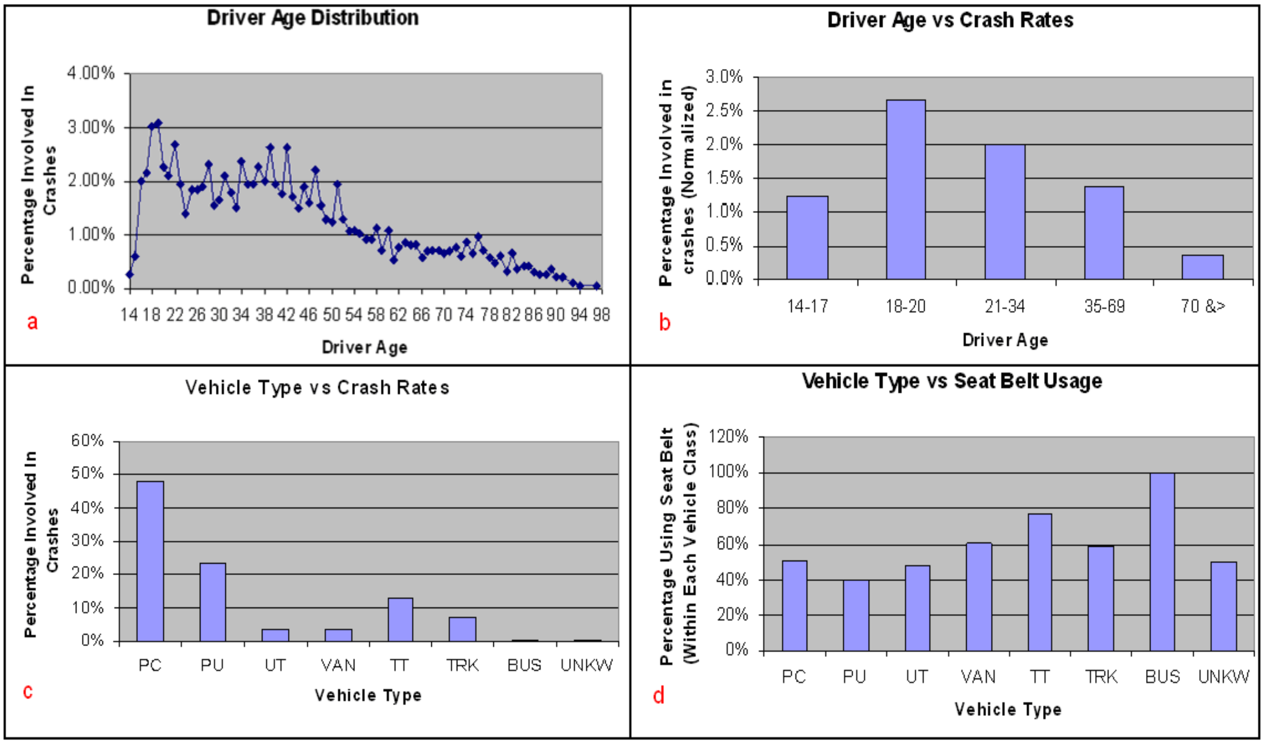


Figure 7.8: Driver Age, Seat Belt Use and Vehicle Type-Rural 2 Lane (FCR)

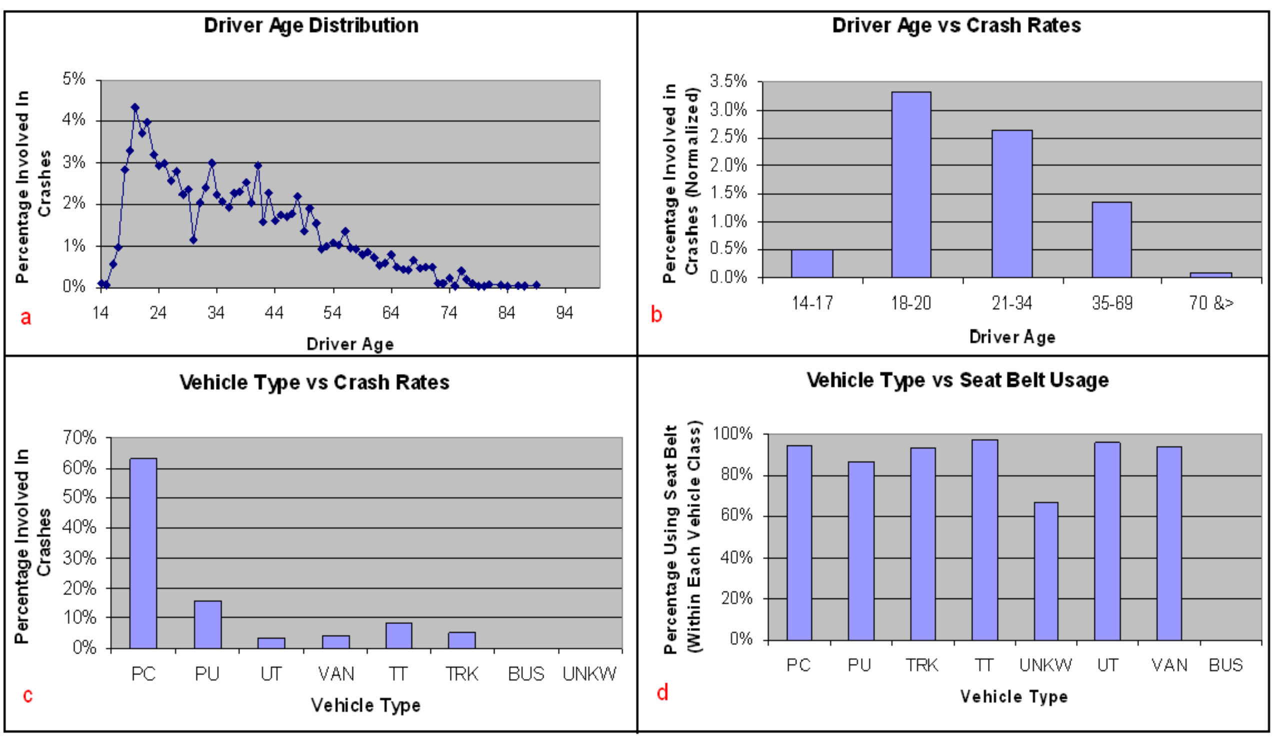


Figure 7.9: Driver Age, Seat Belt Use and Vehicle Type-Rural KTA (TCR)

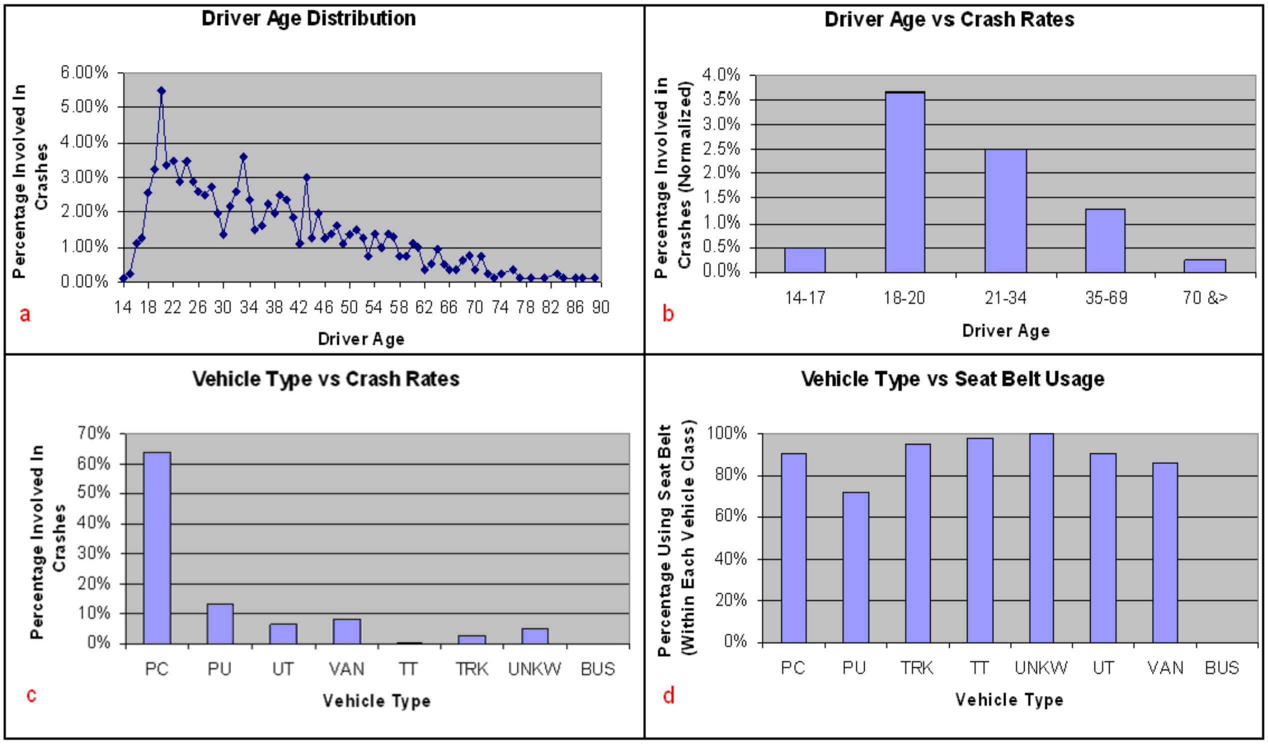


Figure 7.10: Driver Age, Seat Belt Use and Vehicle Type-Rural KTA (ICR)

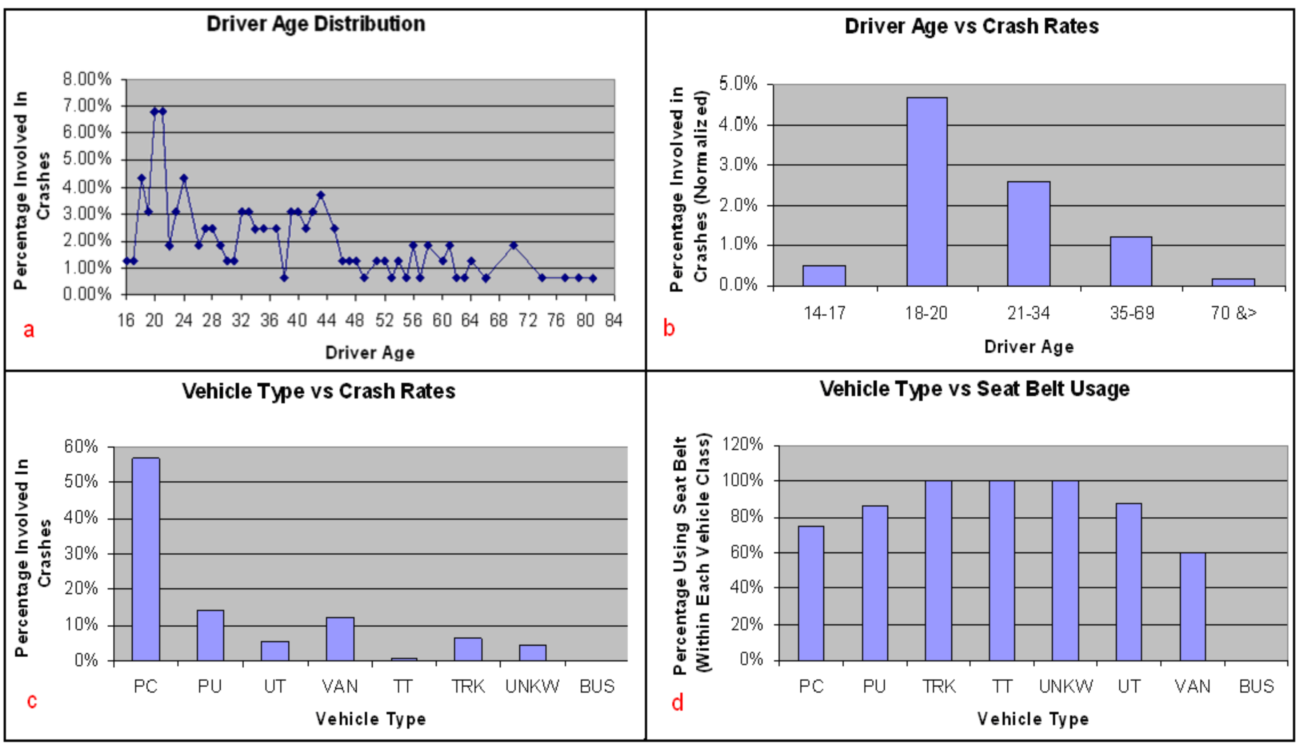


Figure 7.11: Driver Age, Seat Belt Use and Vehicle Type-Rural KTA (SICR)

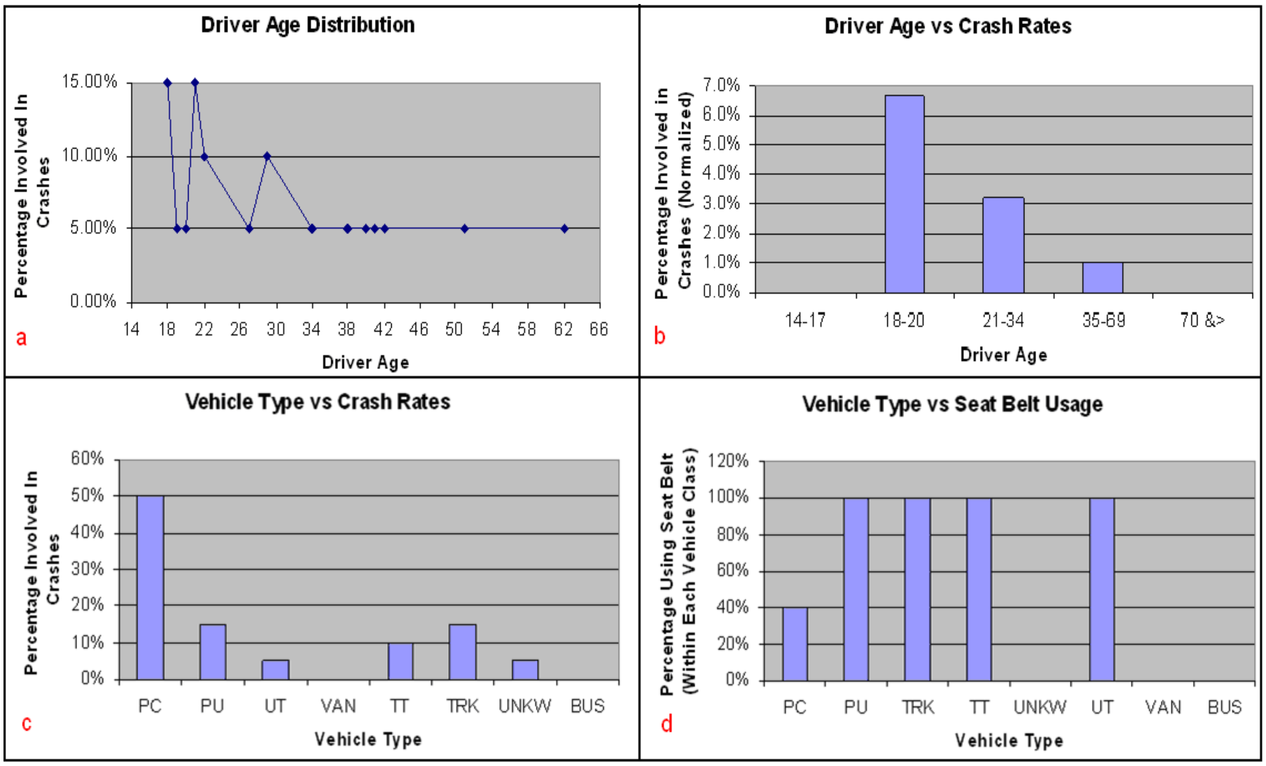


Figure 7.12: Driver Age, Seat Belt Use and Vehicle Type-Rural KTA (FCR)

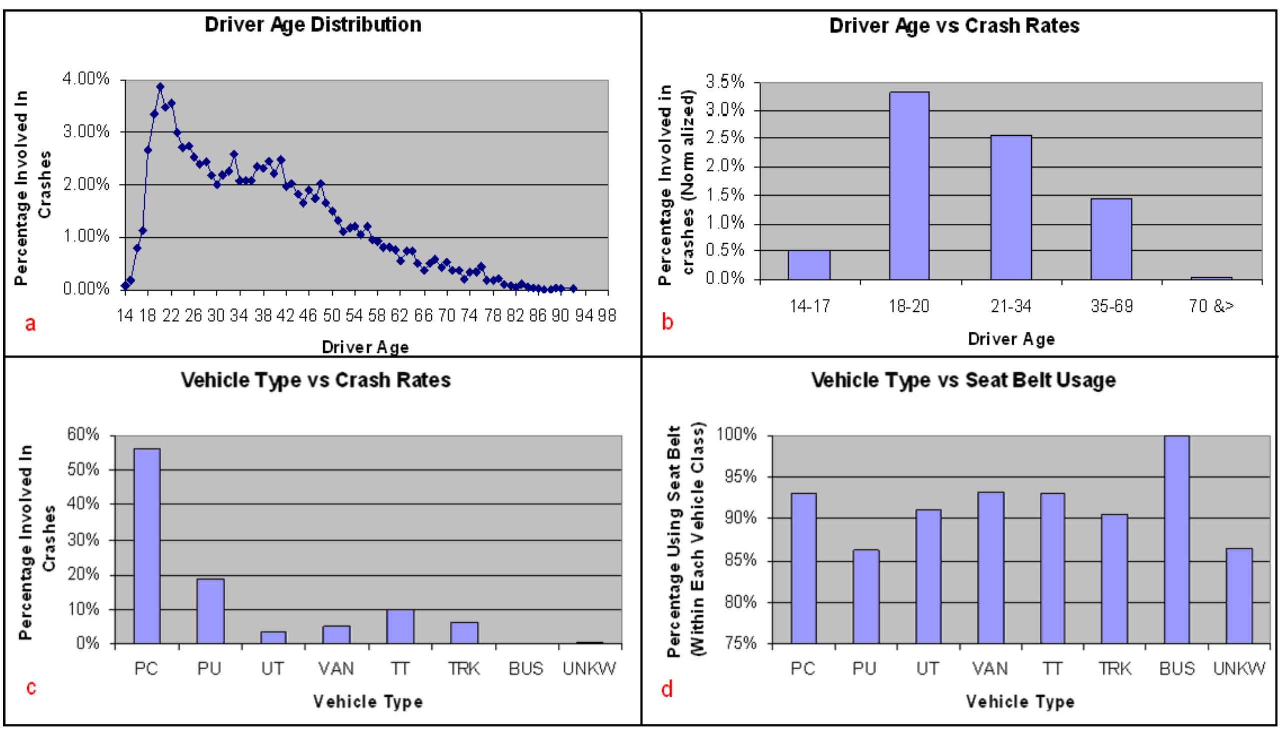


Figure 7.13: Driver Age, Seat Belt Use and Vehicle Type-Rural Freeways (TCR)

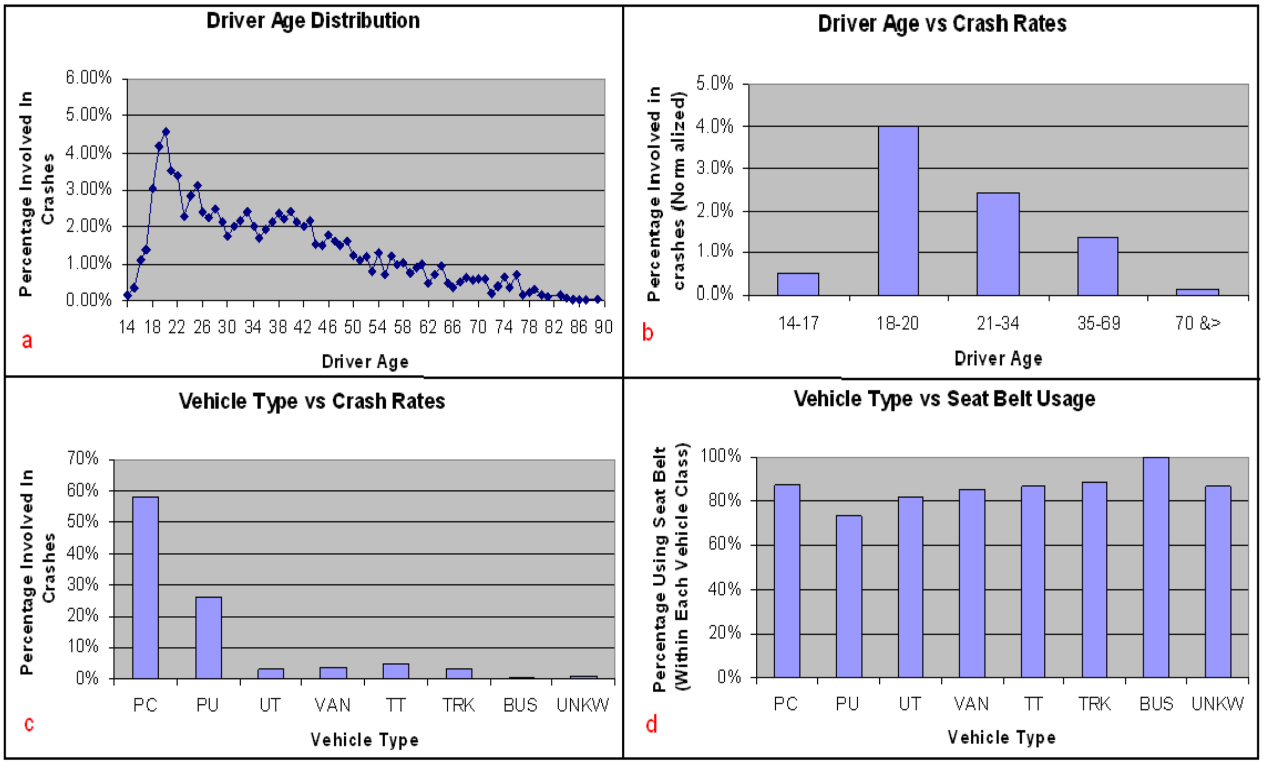


Figure 7.14: Driver Age, Seat Belt Use and Vehicle Type-Rural Freeways (ICR)

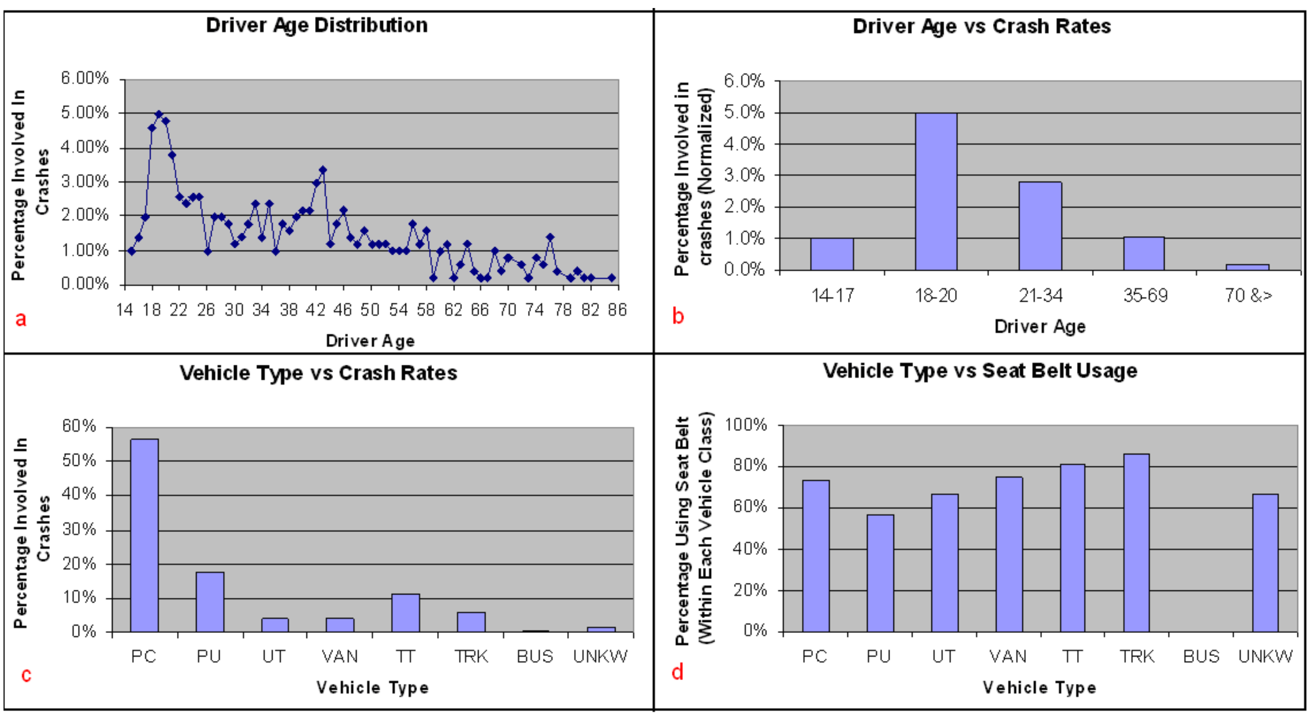


Figure 7.15: Driver Age, Seat Belt Use and Vehicle Type-Rural Freeways (SICR)

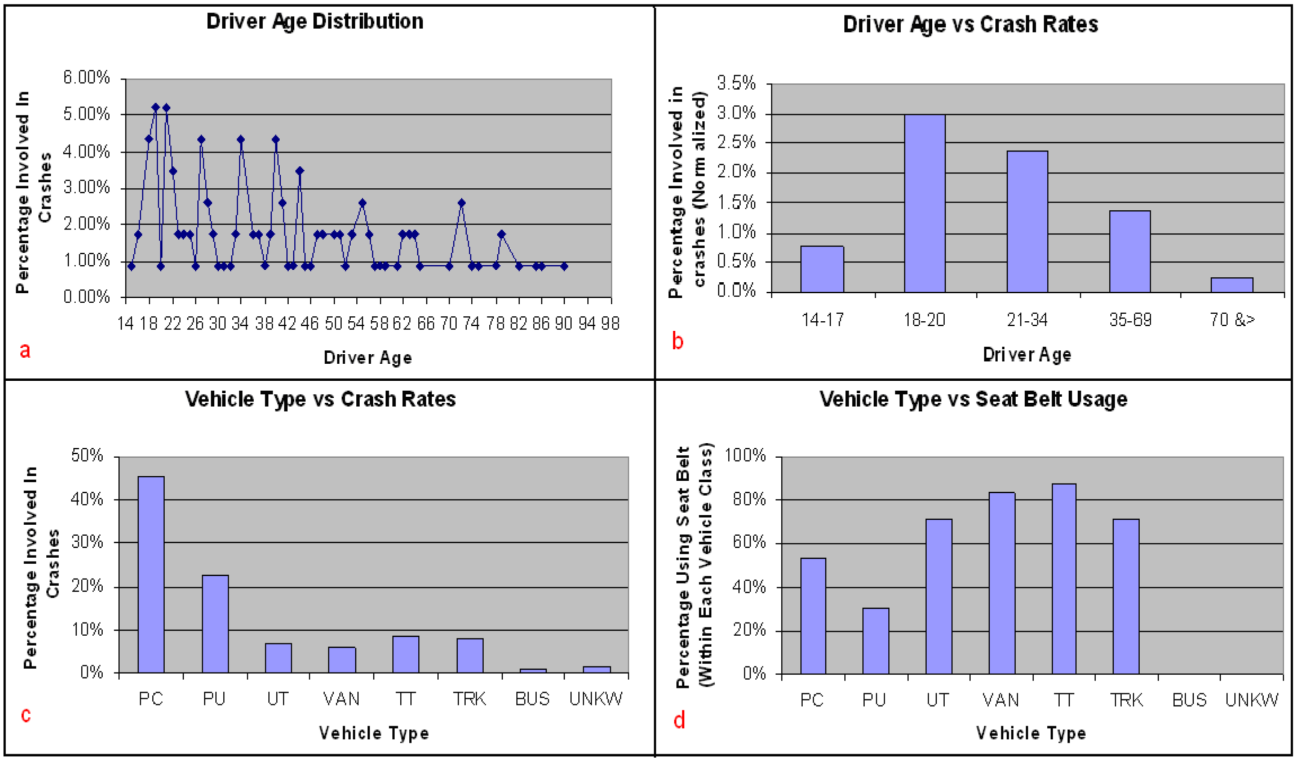


Figure 7.16: Driver Age, Seat Belt Use and Vehicle Type-Rural Freeways (FCR)

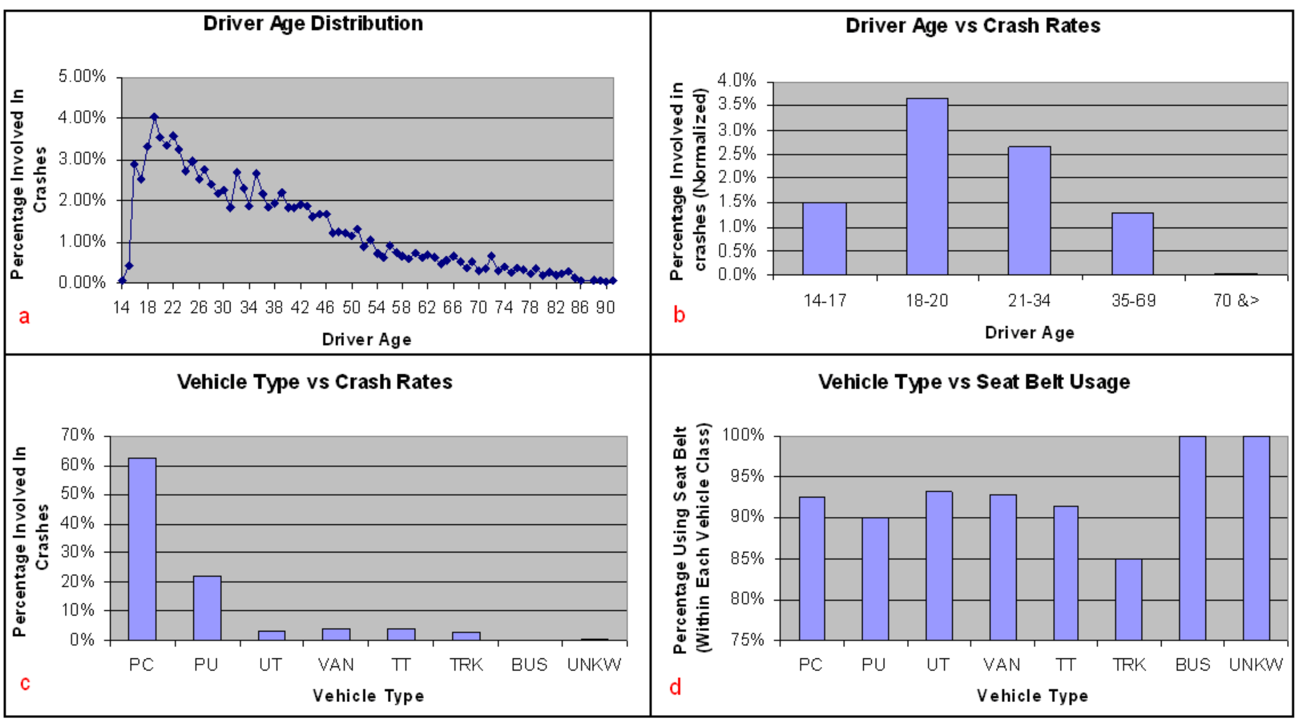


Figure 7.17: Driver Age, Seat Belt Use and Vehicle Type-Urban Expressways (TCR)

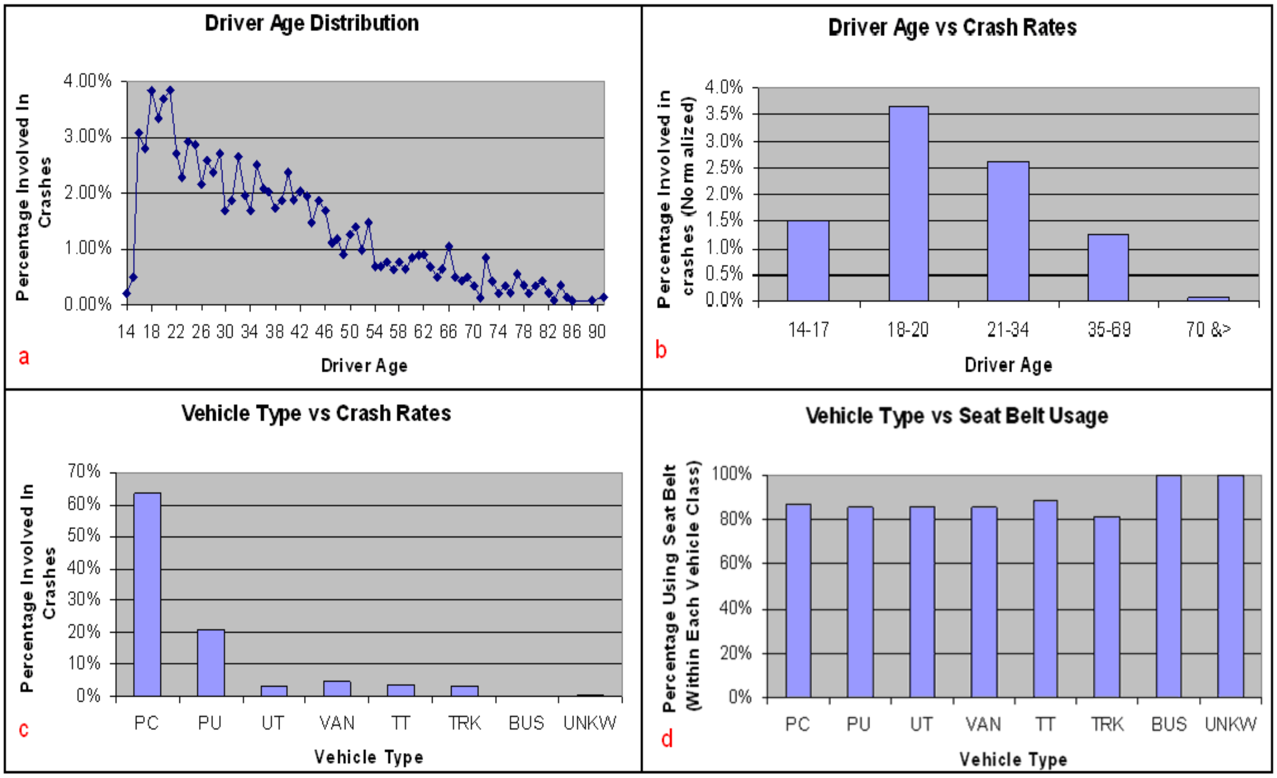


Figure 7.18: Driver Age, Seat Belt Use and Vehicle Type-Urban Expressways(ICR)

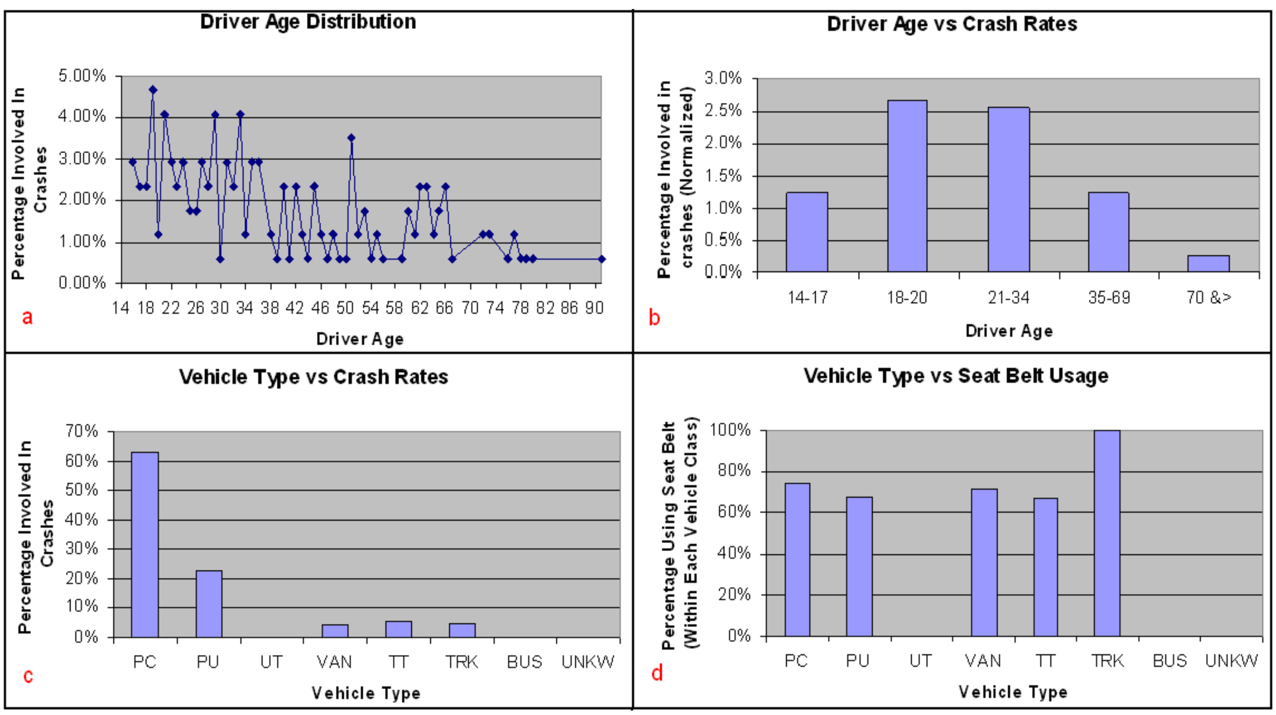


Figure 7.19: Driver Age, Seat Belt Use and Vehicle Type-Urban Expressways(SICR)

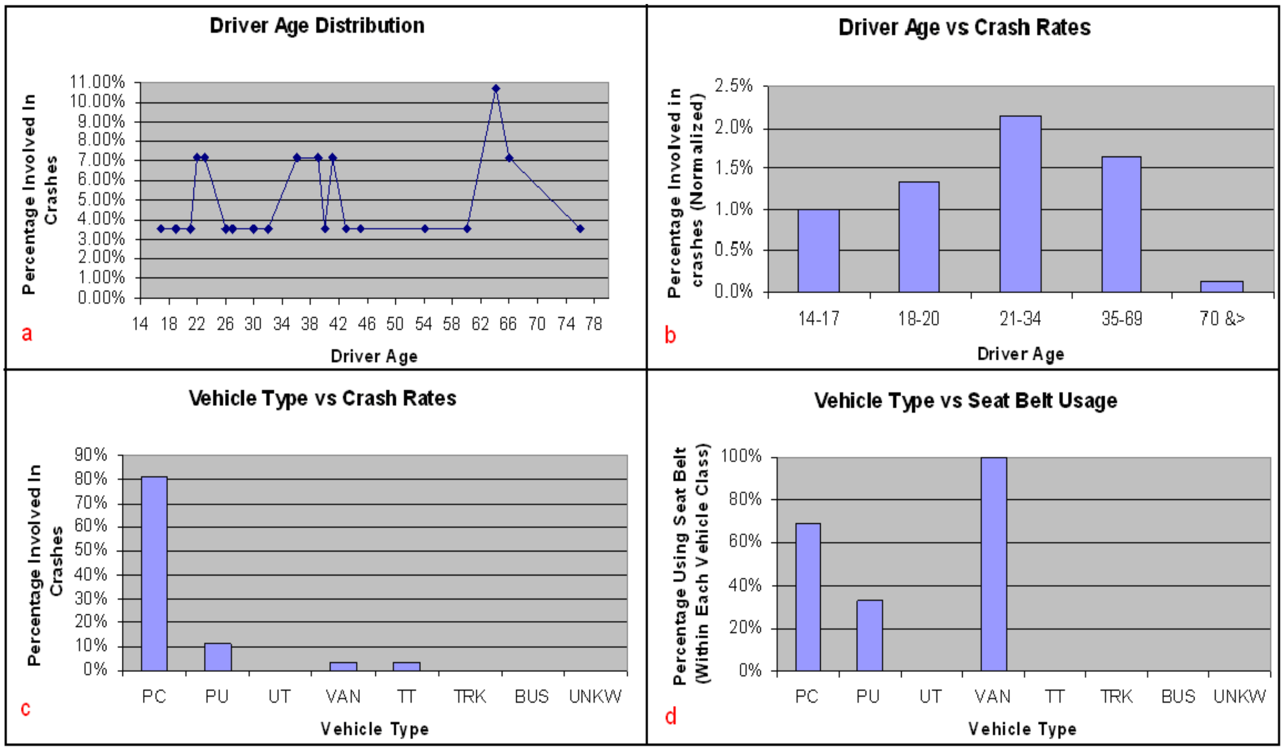


Figure 7.20: Driver Age, Seat Belt Use and Vehicle Type-Urban Expressways(FCR)

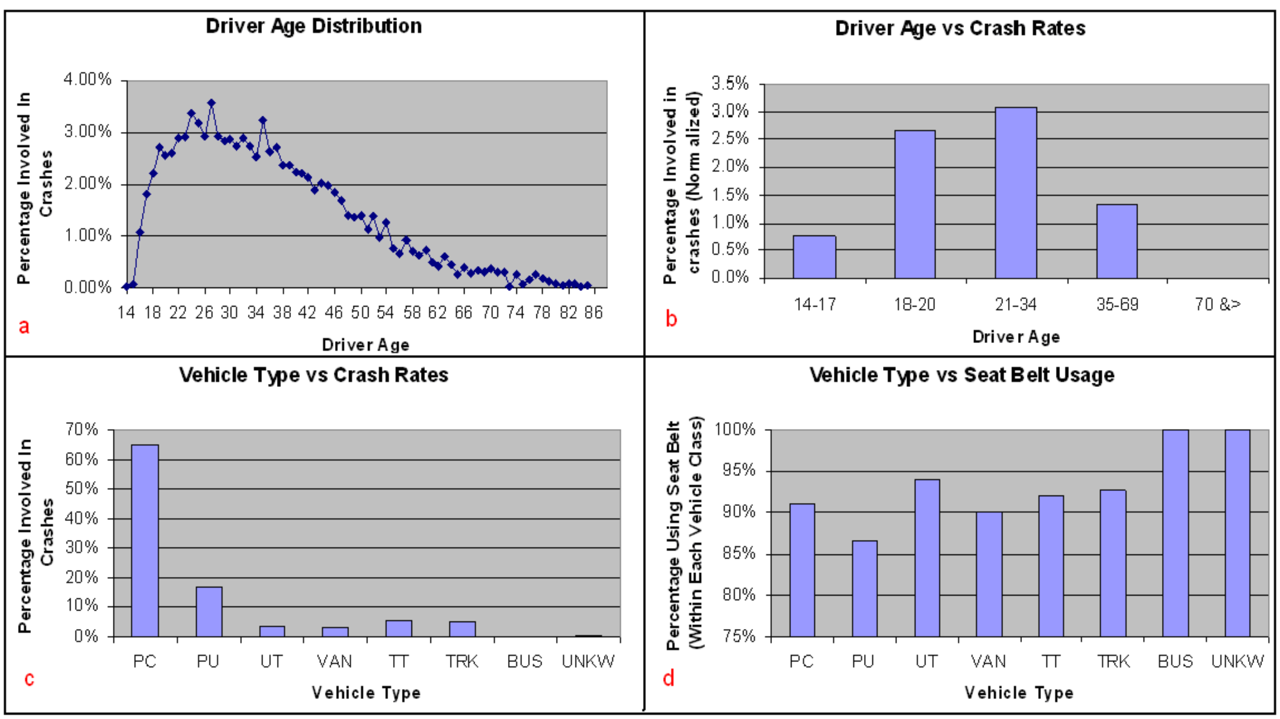


Figure 7.21: Driver Age, Seat Belt Use and Vehicle Type-Urban Freeways(TCR)

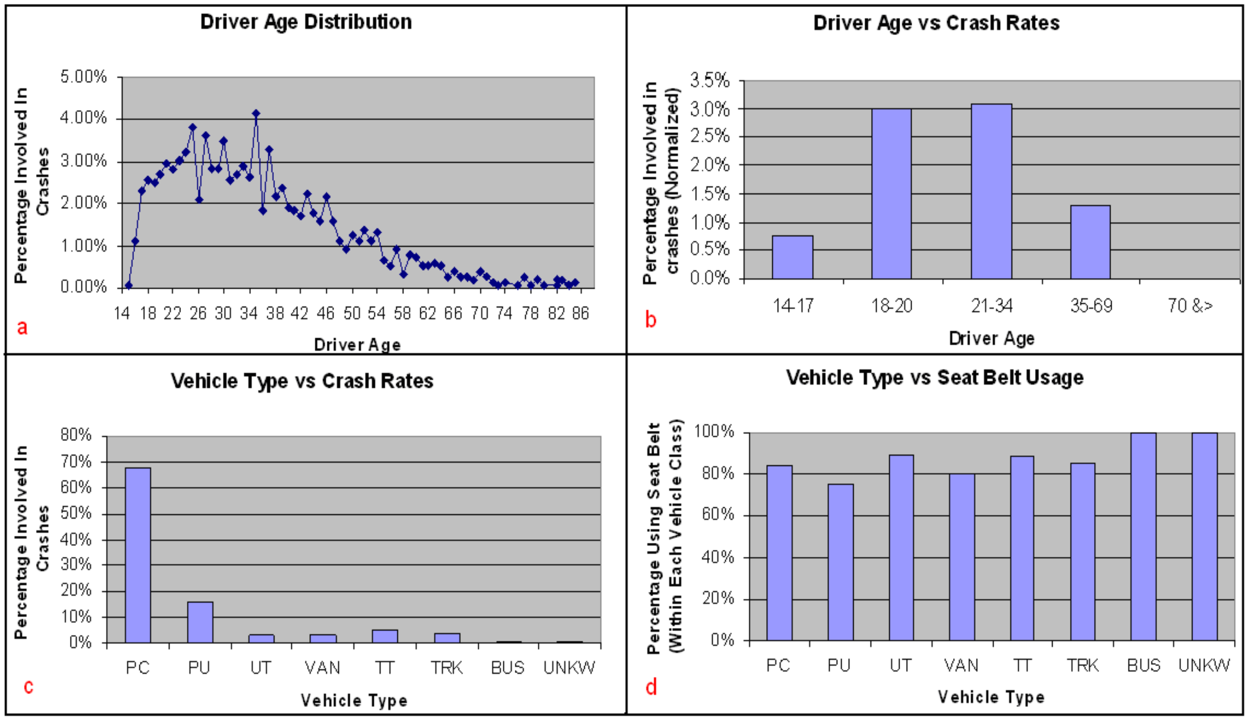


Figure 7.22: Driver Age, Seat Belt Use and Vehicle Type-Urban Freeways(ICR)

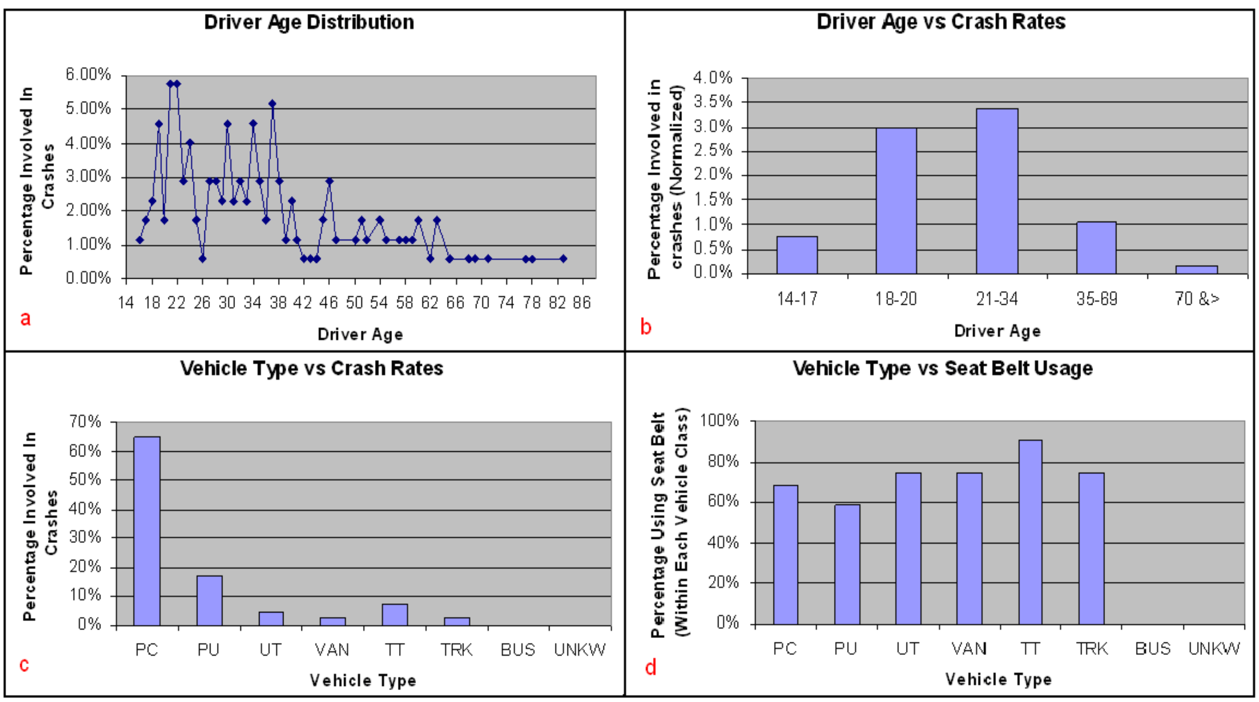


Figure 7.23: Driver Age, Seat Belt Use and Vehicle Type-Urban Freeways(SICR)

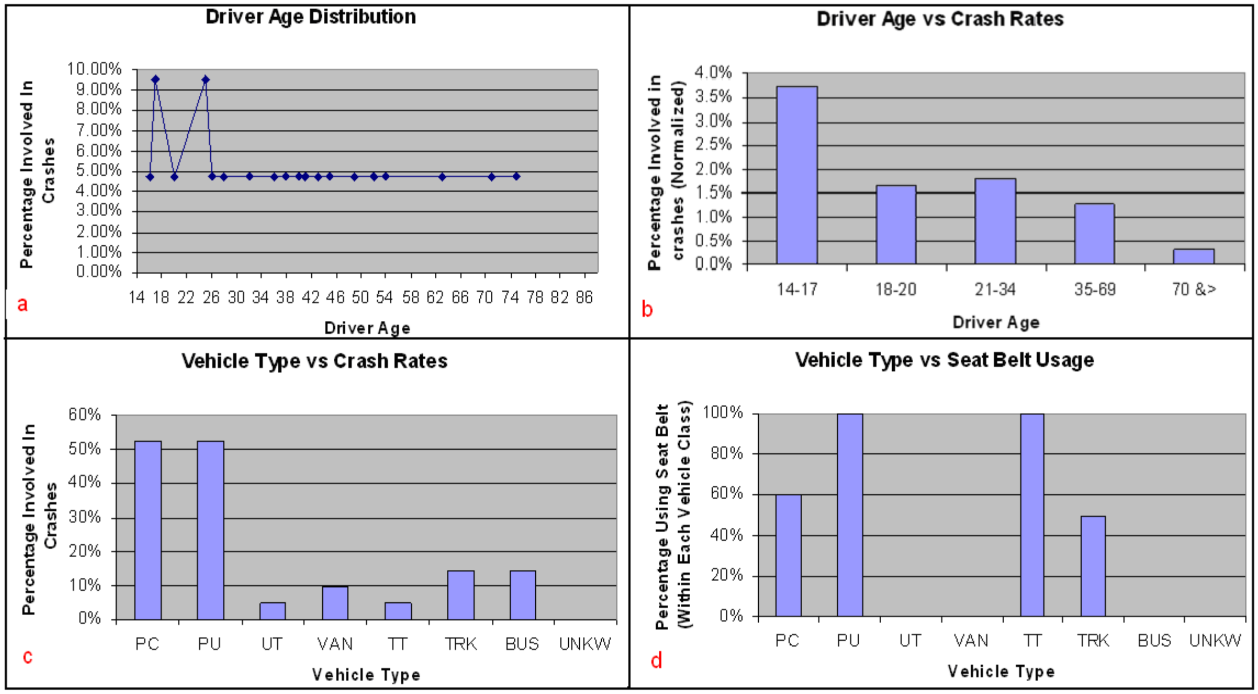


Figure 7.24:Driver Age, Seat Belt Use and Vehicle Type-Urban Freeways(FCR)

CHAPTER 8 - CONCLUSIONS

8.1 Introduction

Over the years modeling techniques shifted from conventional regression to stochastic regression and artificial intelligence network models. Reliable data is being available for modeling and the objective of modeling crashes was transferred from providing criteria and assessment for highway construction and maintenance to supporting advanced traffic management systems. Earlier research emphasized the relationships between highway geometric variables and crashes, while current research focuses more on exploring the relationships between traffic variables and crashes under a certain geometric characteristics.

8.2 Overall Conclusions

The following conclusions can be deduced based for Kansas Road networks:

- Crash rate trends are an effective tool to measure safety hazards on highways as they combine crash frequency with vehicle exposure.
- Rural 2 Lane Network and Urban Expressways network have the highest crash rates in rural and urban categories respectively.
- The average crash rate trends (all networks combined) show that TCR increased up to 1997 and then remained constant. ICR/SICR/FCR do not have any common trends.
- Crash rates and trends give a preliminary picture of the problems associated with each road network. Based on crash rate trend, high crash locations can be identified. If the expected values are much higher than the anticipated values, then proper countermeasures can be taken to improve safety.

- The crash models developed for each network would give engineers and planners a preliminary idea of the variables to control/alter in case of problematic situations and in future planning operations.
- The neural network models developed for crash rates need to be updated on a regular basis as the traffic conditions keep changing. The model updating process would not be as cumbersome as the development phase. Also if new variables need to be incorporated it can be done easily. This convenience is usually not available with conventional regression and statistical models.
- This study has shown the potential of Artificial Neural Network modeling. Further studies should be conducted on different crash databases in other states to and support the findings of this research.
- Driver Age Group (18-20) has the involvement in crashes on all road networks. Some of the recommendations to reduce the crash risk of teen drivers included driver improvement programs, driver education and training, special licensing programs for teens (provisional and graduated licensing), BAC limits, and curfew laws.
- Passenger cars have the highest crash involvement among vehicle types.
- Among all vehicle types, buses have the highest seat belt compliance for all networks. The State of Kansas should change its seat belt law from secondary to primary enforcement. This would enable all drivers to wear seat belts which in turn would reduce crash risk.

8.4 Future Research

One of the major objectives of crash modeling research is to support the traffic management including regular and real time management. Accurate and reliable relationships between the occurrence of crashes and highway geometric and traffic conditions under a certain environment could present useful insight to the potential corresponding safety and traffic operation performance. Therefore, more research should be performed to incorporate the models into traffic management systems.

REFERENCES

Abdelwahab, H. T. & Abdel-Aty, M. A., Development of Artificial Neural Network Models to Predict Driver Injury Severity in Traffic Accidents at Signalized Intersections. *Transportation Research Record* 1746, Paper No. 01-2234, Washington, D.C.

Al-Alawi, S., Ali, G., and Bakheit, C. (1996). ''A novel approach for traffic accident analysis and prediction using artificial neural networks,'' *Road Transport Research,* Vol 5, pp. 118–128.

Ali, Hossam., (2000). "Neuronet-Based liquefaction potential assessment and stress-strain behavior simulation of sandy soils," Doctoral dissertation, Kansas State University, 2000.

Awad, W., Jason, B., (1998). "Prediction models for truck accidents at freeway ramps in Washington state using regression and artificial intelligence techniques," *Transportation Research Record* 1635, pp.30-36.

Bailey, I.L. & Sheedy, J.E. (1988). Vision screening for driver licensure. In *Transportation in an Aging Society, Improving Mobility and Safety of Older Persons, Volume 2*. Washington, DC: National Academy Press.

Block, A. W. 2001. *2000 Motor Vehicle Occupant Safety Survey. Vol. 2, Seat Belt Report.* National Highway Traffic Safety Administration, U.S. Department of Transportation, Nov.

Bowie N., Walz M. (1991). "Data analysis of the speed-related crash issue," Proc. *13th International Technical Conference on Experimental Safety Vehicles,* Paris, 4-7 Nov

Carpenter G.A. and Grossberg S. (1998), "The ART of adaptive pattern recognition by a self-organizing neural network," *Computer*, March 1988, pp.77-88.

Cerella, J. (1985). Information processing rates in the elderly. *Psychological Bulletin*, 98, 67-83.

Chong M., Ajith A., and Paprzycki M., "Traffic Accident Analysis Using Hybrid Decision Tree- Neural Network Approach," Computer Science Department, Oklahoma State University.

Cirillo, J.A. (1968). Interstate System Accident Research—Study II—Interim Report II. *Public Roads,* Vol. 35, No. 3

Denney, N.W. & Palmer, M. (1981). Adult age differences on traditional and practical problem-solving measures. *Journal of Gerontology*, 36, 323-328.

Dia, H., & Rose, G., (1997). "Development and Evaluation of Neural Network Freeway Incident Detection Models Using Field Data," *Transportation Research Record*, Vol. 5, pp.313-331.

Dinh-Zarr, T. B., D. A. Sleet, R. A. Shults, S. Zaza, R. W. Elder, J. L. Nichols, R. S. Thompson, and D. M. Sosin. 2001. Reviews of Evidence Regarding Interventions to Increase the Use of Safety Belts. *American Journal of Preventive Medicine,* Vol. 21, No. 4S, Nov., pp. 48–65.

Eby, David W., Jonathon M. Vivoda. Standard Enforcement in Michigan: A One Year Followup and Review. The University of Michigan and Michigan Office of Highway Safety Planning, UMTRI-2001-22, May 2001.

Elman J.L. (1990), "Finding structure in time, " *Cognitive Science*, Vol. 14, pp. 179-211

Evans, L. (1991). *Traffic Safety and the Driver*. New York, NY: John Wiley.

Evans, L. (1993). How safe were today's older drivers when they were younger? *American Journal of Epidemiology*, 137, 769-775.

Evans, L., Garish, P.H., & Taheri, B. (1998). Epidemiology of the older driver: Some preliminary fingings from data through 1996. In *Proceedings of the 16th International Technical Conference of the Enhanced*

Faghri, A., and Aneja, S. (1999). ''Estimation of percentage of pass-by trips generated by a shopping center using artificial neural networks," *Transportation Planning and Technology,* Vol 22, pp.271–286.

Federal Office of Road Safety (1997). Young Driver Research Program - A review of information on young driver crashes. CR 128.

FHWA (2001) Serving Rural America, Federal Highway Administration, Washington, D.C, 2001, Publication No.: FHWA-EP-01-033 HEP-01/9-01(1M) E

Fildes, B.N., G. Rumbold, and A. Leening. 1991. *Speed Behavior and Drivers' Attitude to Speeding.* Report 16. Monash University Accident Research Center, Monash, Victoria, Australia, June.

Florida Department of Transportation. *Click It or Ticket Campaign: Follow-up Observational Survey of Safety Belt Use in Florida*. August 2001.

Fridstrom, L., Ingebrigtsen, S., (1991). “An aggregate accident model based on pooled, regional time-series data,” *Accident Analysis & Prevention*, Vol.23, pp. 363-378.

Garber, N., Ehrhart, A., (2000). “Effect of speed, flow, and geometric characteristics on crash frequency for two lane highways,” *Transportation Research Record 1717*

Garber, N and Lei Wu, (2001). “Stochastic models relating crash probabilities with geometric and corresponding traffic characteristics data,” Research Report No. UVACTS-5-15-74

Glennon, J., (1987). “Effect of alignment on highway safety. Relationship between safety and key highway features: state of the art report 6,” Transportation Research Board, Washington, DC.

Glassbrenner, D, (2002). *Safety Belt and Helmet Use in 2002—Overall Results.* DOT-HS-809-500. National Highway Traffic Safety Administration, U.S. Department of Transportation, Sept.

Goldberg D. (1989), Genetic algorithms in search, optimization, and machine learning, Addison-Wesley, Reading, MA

Hadi, M., Aruldhas, J., Chow, L., Miaou, S., Lum, H., (1993). “Statistical evaluation of the effects of highway geometric design on truck accident involvements,” Transportation Research Record 1407, pp. 11-23.

Hakkert, A., Hocherman, I., Mensah, A., (1996). “Levels of Safety on Interurban Roads,”. *Transportation Research Record* 1553, pp. 95-102.

Haykin, S. (1994). *“Neural Networks: A Comprehensive Foundation,”* IEEE Press

Holland J.H. (1975), Adaptation in natural and artificial systems, University of Michigan Press, Ann Arbor, MI

Hopfield J.J. (1982), “Neural networks and physical systems with emergent collective computational abilities, “ Proc. National Academy of Sciences, USA, Vol. 79, pp. 2554-2558.

Ivan, J., Pasupathy, R., Ossenbruggen, P., (1999). “Differences in causality factors for single and multi-vehicle crashes on two-lane roads,” *Accident Analysis and Prevention*, Vol.31, pp. 695-704.

Jain A.K., Mao J. and Mohiuddin K.M. (1996), “Artificial Neural Networks: A tutorial,” *Computer*, IEEE, March 1996, pp. 31-44.

Jordan M.I. (1986), “Attractor dynamics and parallelism in connectionist sequential machine,” Proc. *8th Annual conference of the Cognitive Science Society*, Amherst, MA, pp. 531-546

Joshua, S., Garber, N., (1992). “A causal analysis of large vehicle accidents through fault-tree analysis,” *Risk Analysis*, Vol 12, No. 2, pp. 173- 187.

Karlaftis, M., Tarko, A., (1998). “Heterogeneity considerations in accident modeling,” *Accident Analysis and Prevention*, Vol.30, pp. 425-433

Kausler, D.H. (1991). *Experimental Psychology, Cognition, and Human Aging*. New York, NY: Springer-Verlag.

Kohonen T. (1989), Self-organization and associative memory, 3rd edition, Springer-Verlac New York Inc

Kostyniuk, L.P. Streff, F.M., & Eby, D.W. (1997). *The Older Driver and Navigation Assistance Systems*. (Report No. UMTRI-94-47). Ann Arbor, MI: The University of Michigan Transportation Research Institute.

Leveille, S.G., Buchner, D.M., Koepsel, T.D., McCloskey, L.W., Wolf, M.E., & Wagner, E.H. (1994). Psychoactive medications and injurious motor vehicle collisions involving older drivers. *Epidemiology*, 5, 591-598.

Lin, T., Jovanis, P., Yang, C., (1993). “Modeling the safety of truck driver service hours using time-dependent logistic regression,” *Transportation Research Record* 1407, pp. 1-10.

Ljubiè P., Todorovski L., Lavraè N. and Bullas, “Time-Series Analysis of UK Traffic Accident Data,” Department of Intelligent Systems, Institute Jožef Stefan, Slovenia.

Mayhew, D.R. & Simpson, H.M. (1995). The Role of Driving Experience: Implications for the Training and Licensing of New Drivers. Toronto: Insurance Bureau of Canada.

Marottoli, R.A. & Drickamer, M.A. (1993). Psychomotor mobility and the elderly driver. *Clinics in Geriatric Medicine*, 9, 403-411.

Miaou, S., (1994). “The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions,” *Accident Analysis and Prevention*, 26, pp. 471-482.

Miaou, S., Hu, P., Wright, T., Rathi, A., Davis, S., (1992). “Relationship between truck accidents and highway geometric design: a Poisson regression approach,” *Transportation Research Record* 1376, pp. 10-18.

Miaou, S., Lum, H., (1993). “Modeling vehicle accidents and highway geometric design relationships,” *Accident Analysis and Prevention*, Vol.25, pp. 689-709.

Mills, J. (1998). New drivers: Entering the driver licensing system for the first time. In L.J. Faulks, K.B. Smith, & K.N. Smith (Eds). *Young Drivers*. (pp17-23).

Mohamedshah, Y., Paniati, J., Hobeika, A., (1993). "Truck accident models for interstates and two-lane rural roads," Transportation Research Record 1407, pp. 35-41.

Schalkoff, R.J. (1997). *Artificial Neural Networks*, McGraw-Hill Companies

Mussone, L., Ferrari, A., and Oneta, M. (1999). ''An analysis of urbancollisions using an artificial intelligence model.'' *Accident Analysis & Prevention,* 31, pp.705–718.

Najjar, Y., Stokes, R., and Russell, Eugene (2000) "Setting Speed Limits on Kansas Two-Lane Highways: A Neuronet Approach," *Transportation Research Board,* Washington, D.C., 2000.

Nelson, David E., Julie Bolen, and Marcie-Jo Kresnow. Trends in Safety Belt Use by Demographics and by Type of State Safety Belt Law, 1987 through 1993. *American Journal of Public Health*, Vol. 88, No. 2, February 1998, pp. 245–249.

Neutel, C.I. (1995). Risk of traffic accident injury after a prewscription for a benzodiazepine. *Annals of Epidemiology*, 5, 239-244.

NHTSA (2001), National Center for Statistics and Analysis, NHTSA, using NCHS (CDC) 2001 Mortality Data.

NHTSA (2001a), National Highway Traffic Safety Administration. Seat Belt Use Surges in Maryland, Oklahoma, and the District of Columbia with Primary Enforcement. *Traffic Tech*, No. 245, April 2001.

NHTSA (2002), National Highway Traffic Safety Administration. 2000 Motor Vehicle Occupant Safety Survey Shows Continued Gains in Seat Belt Use. Traffic Tech, No. 271, April 2002.

NHTSA (2005), National Highway Traffic Safety Administration, Traffic Safety Facts-2004 Data

Okamoto, H., Koshi, M., (1989). "A method to cope with the random errors of observed accident rates in regression analysis," *Accident Analysis and Prevention,*. Vol.21, 317-332.

O'Neill, B, (2001). Seat Belt Use: Where We've Been, Where We Are, and What's Next. In *2001 Seat Belt Summit,* Automotive Coalition for Traffic Safety, Inc., Jan. 11–13. www.actsinc.org/Acrobat/ SeatbeltSummit2000.pdf.

Owsley, C. & Sloane, M.E. (1990). Vision and aging. In F. Bolter & J. Grafman (Eds.), *Handbook of Neuropsychology, Volume 4*. Amsterdam, The Netherlands: Elsevier.

Partyka, S. (1983). *Comparison of Age of Drivers in Two Car Fatal Crashes*. Washington, DC: National Highway Traffic Safety Administration.

Persaud, B., Dzbik, L., (1992). "Relating freeway accidents to traffic, geometric and operational factors," *Compendium of Technical Papers. Institute of Transportation Engineers Annual Meeting,* Washington, DC, pp. 298-301.

Persson, D. (1993). The elderly driver: Deciding when to stop. *The Gerontologist*, 33, 88-91.

Pike, J.A. (1989). The quantitative effect of age on injury outcome. In *Proceedings of the Twelfth Experimental Safety Vehicle Conference*, Gothenburg, Sweden.

Reinfurt, D., A. Williams, J. Wells, and E. Rogman. (1996). Characteristics of Drivers Not Using Seat Belts in a High Belt Use State. *Journal of Safety Research,* Vol. 27, No. 4, pp. 209–215.

Rumelhart D.E. and McClleland J.L. (1986), Parallel distributed processing: explorations in the microstructure of cognition, MIT Press, Cambridge, MA

Safety of Vehicles. Washington, DC: US Department of Transportation.

Schieber, F. (1994a). *Recent Developments in Vision, Aging, and Driving: 1988-1994*. (Report No. UMTRI-94- 26). Ann Arbor, MI: The University of Michigan Transportation Research Institute.

Solomon, D. (1964). *Accidents on Main Rural Highways Related to Speed, Driver, and Vehicle*. Bureau of Public Roads, U.S. Department of Commerce, July, 44 pp

Subba, R., Sikdar, P., Krishna, K., and Dhingra, S., (1998). "Another insight into artificial neural networks through behavioral analysis of access mode choice," *Computers, Environment and Urban Systems,* pp. 22-5, 485–496.

Transportation Research Board (1988). Transportation in an Aging Society, Improving Mobility and Safety for Older Persons, Volume One. Washington, DC: National Academy Press.

Ulmer, Robert G., Carol W. Preusser, David F. Preusser, and Linda A. Cosgrove. Evaluation of California's Safety Belt Law Change from Secondary to Primary Enforcement. *Journal of Safety Research*, Vol. 26, No. 4, Winter 1995, pp. 213–220.

Vaija, P., (1987). "Application of fuzzy methods to process safety control and to accident analysis," *Acta Polytechnical Scandinavica, Chemical Technology and Metallurgy Series*. No. 180, Helsinki 1987, pp. 26

Vogt, A., Bared, J., (1998). “Accident Models for Two-lame Rural Segments and Intersections,” *Transportation Research Record* 1635, TRB, National Research Council, Washington, DC, pp. 18-29

Waller, P.F. (1991). The older driver. *Human Factors*, 33, 499-505.

Wells, JoAnn K., Allan F. Williams and Charles M. Farmer. Seat Belt Use Among African Americans, Hispanics, and Whites. *Accident Analysis and Prevention*, Vol. 34, No. 4, July 2002, pp. 523–529.

West, L.B., Jr., and J.W. Dunn. 1971. Accidents, Speed Deviation and Speed Limits. *Traffic Engineering*, Vol. 41, No. 10, pp. 52–55

Widrow B. and Hoff M.E. (1960), “Adaptive switching circuits,” Proc. *1960 IRE WESCON Convention Record*, Part 4, IRE, NEW York, pp. 96-104.

Williams, Allan F., Donald Reinfurt, and JoAnn K. Wells. Increasing Seat Belt Use in North Carolina. *Journal of Safety Research*, Vol. 27, No. 1, Spring 1996, pp. 33–41.

Yang, W.T., Chen, H. C., & Brown, D. B., (1999). “Detecting Safer Driving Patterns By A Neural Network Approach,” *ANNIE ’99 for the Proceedings of Smart Engineering System Design Neural Network, Evolutionary Programming, Complex Systems and Data Mining*, Vol. 9, pp 839-844

Yee, D. (1985). A survey of the traffic safety needs and problems of drivers age 55 and over. In J.W. Malfetti (Ed.), *Needs and Problems of Older Drivers: Survey Results and Recommendations*. Falls Church, VA: AAA Foundation for Traffic Safety.

Zegeer, C., Stewart, R., Council, F., Neuman, T., (1994). “Accident relationships of roadway width on low-volume roads,” *Transportation Research Record* 1445, pp. 160- 168.

K - TRAN

KANSAS TRANSPORTATION RESEARCH
AND
NEW - DEVELOPMENTS PROGRAM

A COOPERATIVE TRANSPORTATION RESEARCH PROGRAM BETWEEN:

KANSAS DEPARTMENT OF TRANSPORTATION

THE UNIVERSITY OF KANSAS

KANSAS STATE UNIVERSITY