INVESTIGATION OF CHARACTERISTICS AND ASSESSMENT OF CRASH SEVERITY FACTORS ASSOCIATED WITH TRUCK-RELATED CRASHES IN OHIO

Thesis
Submitted to
The School of Engineering of the UNIVERSITY OF DAYTON
In Partial Fulfillment of the Requirements for
The Degree of
Master of Science in Civil Engineering

By
Thaar Alqahtani
Dayton, Ohio
August, 2016
ABSTRACT

INVESTIGATION OF CHARACTERISTICS AND ASSESSMENT OF CRASH SEVERITY FACTORS ASSOCIATED WITH TRUCK-RELATED CRASHES IN OHIO

Name: Alqahtani, Thaar
University of Dayton

Advisor: Dr. Deogratias Eustace

Truck safety is a very crucial aspect of the overall safety of the transportation system. Statewide, there has been a significant increase in the probability of trucks being involved in crashes, primarily due to the fact that the total number of registered trucks, as well as the truck vehicle-miles traveled, have both increased within the last 10 years. Recognizing the substantial impact of truck-related crashes in the overall transportation safety, this study attempted to identify the contributing factors that influence the increase in truck-related crash severity, using truck-related crash data for the last two and half years (July 2013-December 2015) that were obtained from the Ohio Department of Public Safety Traffic (ODPS).

This thesis study used the classification tree model to investigate the important factors affecting injury and fatality related to truck crashes in Ohio. Eighteen independent variables that represent various driver, roadway, environmental and crash characteristics were tested in the classification tree
model of truck-related crash model. The dependent variable, crash severity was coded as a binary variable, with no injury and injury/fatal as its two crash severity levels. The classification tree model selected five independent variables as the only most significant factors influencing truck-related crash severity. These variables are crash type, posted speed limit, collision event, speed-related and road contour. Their significance is also in that order, with the crash type being the most significant, contributing about 55.8% to the model, posted speed limit contributing about 18.5%, collision event about 17.7%, speed-related about 6.0% and lastly road contour about 2.0%.
ACKNOWLEDGEMENTS

Above all else, I would like to express my most profound gratitude to the Almighty God for the unwavering provisions, enlightenment, and guidance which helped me to successfully carry out this endeavor.

My heartfelt appreciation also goes to my thesis advisor, Dr. Deogratias Eustace, who took the time to review, critique, and provide valuable feedback as well as the much-needed academic support that I needed to be able to accomplish this research. To Dr. Peter Hovey, I am deeply indebted for his immense contribution and support, specifically in helping me to ensure the accuracy of the statistical data applied in this research. It also gives me great pleasure to acknowledge Mr. Paul Goodhue for his very meticulous scrutiny of my thesis manuscript and his valuable technical advice. I consider myself blessed to have all of you providing the academic guidance that I needed to carry out this worthy task.

To all my friends and classmates who contributed in their own special ways to my success in this program, I offer my sincerest words of gratitude. Most importantly, to my family, both home and abroad, I cannot express how thankful I am for their continued support, prayers, contributions, and for bearing with me throughout this program of study.
# TABLE OF CONTENTS

ABSTRACT ................................................................................................................................... iii  
ACKNOWLEDGEMENTS ............................................................................................................ v  
LIST OF FIGURES ..................................................................................................................... viii  
LIST OF TABLES .......................................................................................................................... x  
CHAPTER I INTRODUCTION ..................................................................................................... 1  
  1.1 Introduction ..................................................................................................................... 1  
  1.2 Problem Statement ........................................................................................................ 2  
  1.3 Objective of Study ......................................................................................................... 4  
  1.4 Organization of the Thesis ......................................................................................... 5  
CHAPTER II LITERATURE REVIEW .......................................................................................... 6  
CHAPTER III METHODOLOGY AND DATA COLLECTION .................................................. 12  
  3.1 Methodology ................................................................................................................. 12  
  3.1.1 Introduction ............................................................................................................. 12  
  3.1.2 Classification Tree Modeling ................................................................................ 13  
  3.1.3 Validation ................................................................................................................ 16  
  3.1.4 The Criteria of Node Splitting ............................................................................. 16  
  3.1.5 Tests for Goodness-of-Fit .................................................................................... 17  
  3.2 Data Collection ............................................................................................................ 19  
  3.2.1 Crash Data .............................................................................................................. 19  
  3.2.2 Merging Data Files ............................................................................................... 20  
  3.2.3 Creating Truck Crashes Database ................................................................... 22  
  3.3 Description of Selected Variables ............................................................................ 22  
  3.3.1 Description of Dependent Variable ................................................................... 23  
  3.3.2 Description of Independent Variables ................................................................. 23  
CHAPTER IV RESULTS AND DISCUSSION ........................................................................... 30  
  4.1 Introduction .................................................................................................................... 30  
  4.1.1 Descriptive Results of Truck-Related Traffic Crashes ......................................... 30  
  4.2 Result of Classification Tree Modeling ....................................................................... 35
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.2.1 Introduction</td>
<td>35</td>
</tr>
<tr>
<td>4.2.2 The Ideal Size of Tree</td>
<td>36</td>
</tr>
<tr>
<td>4.2.3 Model Validation</td>
<td>38</td>
</tr>
<tr>
<td>4.2.4 The Significant Predictors</td>
<td>39</td>
</tr>
<tr>
<td>4.2.5 Final Classification Tree</td>
<td>40</td>
</tr>
<tr>
<td>4.3 Discussion</td>
<td>50</td>
</tr>
<tr>
<td>CHAPTER V CONCLUSION AND RECOMMENDATIONS</td>
<td>52</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>55</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure 1.1 Percent of truck-related crashes in Ohio by crash severity level, 2011-2015…………………………………………………………………………..4

Figure 3.1 An example of a flow-chart of a classification tree………………..15

Figure 3.2 An example of ROC curve………………………………………….19

Figure 3.3 The process of merging "crash-unit-people" file……………………21

Figure 3.4 Description of crash severity………………………………………23

Figure 4.1 Distribution of crash severity by the crash type…………………...31

Figure 4.2 Distribution of the crash severity by collision event………………32

Figure 4.3 Distribution of the crash severity by posted speed……………….33

Figure 4.4 Distribution of crash severity by road contour……………………34

Figure 4.5 Distribution of crash severity by alcohol-related…………………..35

Figure 4.6 Distribution of crash severity by drug related……………………35

Figure 4.7 Split history of validation and training subsets for full tree…………37

Figure 4.8 Split history of validation and training subsets for ideal size tree……38
Figure 4.9 Receiver-operating characteristics of training dataset..................39

Figure 4.10 Receiver-operating characteristics of validation dataset..............39

Figure 4.11 Final classification tree results..................................................42

Figure 4.12 The root node split by crash type..............................................43

Figure 4.13 Crash type (sidwip, NCMVT) node split by collision event.........44

Figure 4.14 Collision event (CMVT and NCE) node split by posted speed limit.................................................................45

Figure 4.15 The split of the node for posted speed limits 45 mi/h or higher and road contour.................................................................46

Figure 4.16 The crash types rear, head-on and angle node split based on the posted speed limit.................................................................47

Figure 4.17 The posted speed limit less than 45 mi/h node splitting..............48

Figure 4.18 Splitting of collision events (CMVT, NCE and CMObj) node by speed-related.................................................................49

Figure 4.19 Posted speed limits 45mi/h or higher node split by crash ..........49

Figure 4.20 Splitting of crash type (rear and angle) node by speed-related ......50
LIST OF TABLES

Table 1.1 Summary of truck-related crashes by severity in Ohio 2011-2015......3
Table 3.1 Description of human and driver variables...............................25
Table 3.2 Description of road and environmental variables........................27
Table 3.3 Description of vehicle and crash variables.................................29
Table 4.1 The column contributions report for selecting significant predictor
variables........................................................................................................40
CHAPTER I

INTRODUCTION

1.1 Introduction

Significant improvements in road and highway infrastructures and railway systems throughout the United States has contributed to the vast increase and ease of the movement of goods, services, and travelers across the nation. With the increase in the demand for moving heavy goods, it became essential to deploy large trucks in moving heavy cargo to various destinations.

According to Kotikalapudi and Dissanayake. (2013). a vehicle can be considered as a large truck if it has a gross vehicle weight rating (GVWR) of 10,000 pounds or greater. Trucks are deemed the more reasonable choice for moving heavy goods because they are cheaper compared to sea vessel or airline transportation. Furthermore, heavy truck transportation is considered more flexible compared to ships and airplanes (Mooren et al., 2014), in the sense that they can make stop-overs at any place where a station is available, unlike the airplane that only makes a stopover at an airport. Subsequently, large trucks can easily navigate remote areas that are inaccessible to ships or airplanes. Other forms of transportation that include ships and airplanes require the client to take his or her cargo to the harbor or airport respectively. For these reasons, most
Americans prefer transporting their cargo through the use of large trucks due to their ability to move fast, reach remote destinations, and most importantly due to their affordability.

Over the years, the United States has experienced a number of transportation-related fatalities occurring throughout her roadway network. According to National Highway Traffic Safety Administration (NHTSA), an estimated 342,000 large trucks were involved in police-reported traffic crashes in 2013 alone. There were 3,964 people killed, which means 10.8 deaths per day, and an estimated 95,000 people injured in crashes involving large trucks. In terms of severity, an alarming increase in truck-related crashes has been recorded, with more and more trucks being involved in accidents. Bezwada (2010) reports that in the United States large trucks are over-represented in fatal and serious traffic crashes whereby they make up only 3% of all registered vehicles and account for 7% of vehicle miles of travel but they are involved in 11% of all motor vehicle traffic fatalities.

1.2 Problem Statement

In Ohio, the increase in incidents involving truck crashes has raised flags among concerned citizens and traffic safety professionals and it has now been considered a major road safety issue. Based on the latest data, in 2015 trucks were involved in crashes every three hours. Table 1.1 shows that truck-related crashes have generally been increasing while they contributed 8.7% of all fatal crashes in 2011 but contributed 12.1% of fatal crashes in 2015. These data were
analyzed from the Ohio Department of Public Safety (ODPS) traffic crash database from 2011 to 2015. Figure 1.1, which displays the same data depicted in Table 1.1, clearly shows that truck-related fatal crashes contribute more to fatal crashes than injury and PDO truck crashes do to their respective injury levels. Due to the increase in the number of victims in truck-related crash incidents, it is important to identify factors associated with crash severities of these types of crashes, which may help traffic engineers in devising appropriate countermeasures that can reduce traffic-related deaths and injuries.

There are many different factors that contribute to the increase in truck crash severity. These factors can generally be classified into: human and driver-related, road geometric-related, crash-related and environmental-related factors. This thesis study aims to investigate the contributing factors and identify the most significant factors in crash severity related to large trucks in the state of Ohio.

Table 1.1 Summary of truck-related crashes by severity in Ohio 2011-2015

<table>
<thead>
<tr>
<th>Year</th>
<th>Fatal Crashes</th>
<th>Injury Crashes</th>
<th>Property Damage Only Crashes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Truck</td>
<td>All</td>
<td>%</td>
</tr>
<tr>
<td>2011</td>
<td>126</td>
<td>1452</td>
<td>8.7</td>
</tr>
<tr>
<td>2012</td>
<td>161</td>
<td>1588</td>
<td>10.1</td>
</tr>
<tr>
<td>2013</td>
<td>165</td>
<td>1495</td>
<td>11.0</td>
</tr>
<tr>
<td>2014</td>
<td>139</td>
<td>1430</td>
<td>9.7</td>
</tr>
<tr>
<td>2015</td>
<td>200</td>
<td>1647</td>
<td>12.1</td>
</tr>
</tbody>
</table>
1.3 **Objective of Study**

The main objective of this study was to determine the factors that contribute to fatal and injury-causing crashes associated with truck accidents that occurred in Ohio from July 2013 until December 2015. The data used in this study were obtained from the Ohio Department of Public Safety (ODPS). The objective of this research study was to investigate the main predictor factors that lead to large truck-related fatal or injury-causing crashes, by examining factors such as environmental factors, traffic factors, road geometry, and driver (human) factors, that may have caused an increase in truck-related crashes and their severity levels in Ohio. For purposes of clarification, the traffic crash severity is defined by the severity of the injury suffered by the person who was most injured in the crash. Traffic crash severity is typically categorized into three levels: fatal
crash, an injury crash, or property damage only crash (PDO). This study was designed to highlight factors that have an important effect in causing or contributing to truck-related crashes likely to cause fatalities and injuries in the state of Ohio.

1.4 Organization of the Thesis

This thesis is organized into five chapters. The first chapter introduces the extent of the problem associated with truck-crash severity; the second chapter presents the literature review on the topic; the third chapter describes the methodology applied and data collection processes involved, providing the data obtained and explaining the relevant variables. The fourth chapter contains the results and discussion; and the fifth chapter provides a summary and conclusion of the research findings.
CHAPTER II
LITERATURE REVIEW

Chang and Mannering (1999) examined the injury severity and vehicle occupancy relationship for large truck-involved crashes and non-truck-involved crashes utilizing data from the state of Washington. They developed separate nested logit models for trucked-related crashes and non-truck-related crashes. Based on their findings, the variables that significantly increased the danger risk for truck-involved crashes are high posted speed and type of collision especially for crashes occurring when a vehicle is making right or left turn and rear-end type of collisions. In addition, the severity of the most injured vehicle occupant increases if the crash involves a truck. The impacts of truck crashes are more severe in multi-vehicle crashes than single-vehicle crashes due to the impact the large vehicle causes to occupants of other smaller vehicles. Abdel-Aty and Abdelwahab (2004) developed artificial neural network (ANN) models to predict injury severity levels in traffic crashes. Their results show that gender, vehicle speed, seat belt use, type of vehicle, point of impact, and area type (rural/urban contribute to the likelihood of injury severity levels.

Khattak et al., (2003) developed binary probit binary models to examine truck rollover propensity and ordered probit models to evaluate injury severity of large trucks in single-vehicle crashes. They categorized higher risk factors in
single-vehicle truck crashes into three main themes: (1) dangerous truck-driver behaviors especially speeding, reckless driving, alcohol and drug use, non-use of restraints, and traffic control violations; (2) exposure to dangerous roadway geometry, predominantly more curves, and (3) trucks that transport hazardous materials and post-crash fires.

Lyman and Braver (2003) made investigative analysis of data for twenty-five (25) years on US nationwide crashes involving large truck by using exposure measures such as occupant fatalities per 100,000 population, per 10,000 licensed drivers, per 10,000 registered trucks and per 100 million vehicle-miles of travel (VMT). Among their findings, they note that although large truck involvement in fatal crashes has dropped substantially when measured per unit of travel, but the public health concern of large truck-related crashes, as measured by deaths per 100,000 population, has not improved over time because of the large increase in truck vehicle miles of travel.

USDOT (2006) prepared an investigative analysis directed to crash samples of trucks involved in fatal or injury crashes from 24 sites in 17 states. Results from that study indicate that 87% of the reported crashes were mainly due to actions of the driver and driving errors. On the other hand, 13% of the reported crashes were mainly due to roadway geometric characteristics or weather conditions.

Cantor et al. (2010) developed a model to investigate the contribution of driver factors on the number of crashes in which the driver was involved using a national-level database. Their findings suggest that significant factors related to
the likelihood of a crash occurrence include driver age, weight, height, gender, and employment stability as well as previous driver and vehicle violations and past crashes. Khorashadi et al. (2005) used four years of crash data from California and conducted a multinomial logit model to determine contributing factors of injury severity of drivers (of both large-trucks and passenger-vehicles) involved in large truck-related crashes. Their study report significant differences with regard to risk factors between urban and rural models developed. Elaborating on these differences, they show that for rural crashes involving tractor–trailer combinations, the probability of the driver sustaining severe or fatal injuries increased by about 26% compared to crashes involving single-unit trucks. On the other hand, for crashes in urban areas, this same probability increased by about 700%. Likewise, for crashes where alcohol or drug use was reported as a primary contributing factor for the crash, the probability of severe or fatal injury was much higher in urban areas than that in rural areas.

Nassiri and Edrissi (2006) developed logit and neural network models to predict the severity of truck-related crashes and to identify the related significant factors for crashes occurring on two-lane rural highways in Iran. They found out that a variety of variables related to roadways, vehicles, environment and drivers, specifically driver fatigue, head-on collisions and lack of vehicle control significantly contributed to the severity of truck-related crashes. The research work carried out by Chen and Chen (2011) investigated injury severities of truck drivers involved in multi-vehicle (MV) and single-vehicle (SV) crashes by using mixed logit models. Their results show that some variables were only significant
in the SV model or the MV model, but not both. Interestingly, they also note that some of the variables that were significant in both SV and MV models, they showed considerable difference in marginal effects in the two models; even some of them had opposite effects to SV and MV crashes.

Lemp et al. (2011) used heteroskedastic ordered probit models to study the impact of vehicle, occupant, driver, and environmental characteristics on injury effects for individuals involved in large truck-related crashes. Their results suggest that the probability of fatalities and severe injury is expected to be higher with the number of trailers, but lower with the truck length and gross vehicle weight rating (GVWR).

A study by Zhu and Srinivasan (2011) presents a panel, heteroskedastic ordered-probit model capable of simultaneously analyzing the injury severities of all persons involved in a crash by estimating models of large-truck crashes. Their study suggests that several driver behavior characteristics such as use of illegal drugs, driving under influence (DUI), and inattention are statistically significant predictors of injury severity. Additionally, they note that the availability of airbags and the use of seatbelts are associated with lower levels of injuries to car drivers and their passengers in the event of crashes with large trucks.

Shankar et al. (1996) applied a nested logit model in estimating the severity crash on rural freeways using a five-year crash data from a 61 km section of rural interstate in Washington State. Their estimation outcomes indicate that environmental conditions, crash type, vehicle attributes, and
highway design and driver characteristics have significant effects on crash severity on rural interstates.

Mulinazzi et al. (2009) analyzed heavy vehicle crashes on Interstate 70 (I-70) that involved strong winds. The data were analyzed to determine the correlations between the vehicle and freight characteristics, crash occurrences and weather conditions. They constructed a model to predict the probability of wind induced truck crashes on I-70 in the state of Kansas. The study found that wind speed was not a statistically significant factor in predicting such crashes.

Dissanayake and Bezwada (2010) analyzed characteristics and contributory factors related to fatal crashes involving large trucks in the United States utilizing 2003-2007 crash data obtained from FARS database. A multinomial logistic regression model was developed to predict the type of crash (truck or non-truck) as the dependent variable. Their study results indicate that driver-related factors such as cellular phone usage, failure to yield right-of-way, and inattentiveness, inadequate warning signs and poor shoulder conditions were found to be more significant contributing factors to truck crashes than to non-truck crashes.

A study carried out by Charbotel et al. (2003) assessed severity factors for injury severities sustained by truck drivers involved in traffic crashes by utilizing data from the trauma registry of road crash victims of the Rhône region, France for 1995-1995 time period. A number of descriptive characteristics of the victims (such as age, place of residence, etc.) and their crash characteristics (such as place, time, type of crash, seatbelt wearing, etc.) were used in that study. Their
results show that truck drivers were more seriously injured than car drivers and low rate of seatbelt use was a significant factor in explaining the difference in the injuries suffered by truck drivers compared with car drivers resulting in a higher rate of limb and abdominal injuries. In addition, their results identify that trucks are dangerous for other road users.
CHAPTER III
METHODOLOGY AND DATA COLLECTION

3.1 Methodology

3.1.1 Introduction

This chapter discusses data collection effort and the research methodology used in achieving the objectives of the present study.. In this thesis study, the classification tree model, also known as decision tree model, was used in investigating the characteristics of crash severity of large truck-related crashes. The classification tree model is a powerful multivariate technique that is used for both data exploration and prediction (Lavery, 2012). The classification tree model provides a graphical description of classifications that we make, events that are likely to occur, and the associated outcomes. This happens after a combination of events and classifications; the assigning of probabilities to events is determined for every outcome. The main aim of a classification tree is to come out with the best classifications. Structuring a classification tree would include concepts like nodes and branches, among others, depending on the applicability of the model.

Lavery (2012) explains that classification trees splits a data set (assigning observations in the dataset to groups) hierarchically (i.e., groups are then divided
into subgroups) based on the ability of independent variables (X’s), associated with the observations, to predict the dependent variable (Y).

According to Myles et al., (2004), classification tree modeling has been extensively used in analyzing exploratory data analyses and in predictive modeling applications because of its suitability in using determinant features of extracting patterns in large databases.

3.1.2 Classification Tree Modeling

The classification tree model is performed by dividing datasets into smaller and homogeneous subgroups by using a set of “if-statements.” The classification tree model is a hierarchical model that is composed of discriminant functions or classification (decision) rules that are applied recursively to partition the entire sample of dataset into pure and single class subsets. Classification trees use some statistical measurements in order to split the dataset into small and more homogeneous subgroups. The model divides the dataset based on the most predictive independent variable for the response variable. The classification tree model basically selects the appropriate measurements based on the type of the response variable.

The dependent and independent variables might be continuous or categorical. If the predictor variable is categorical (nominal or ordinal), the model splits the predictor variable into two groups of levels (categories). If the response variable is continuous, the model splits the predictor variable based on the cut-off
value into two partitions and the measure of the difference in the two groups is computed as the sum of squares of the differences between the means. Both the variable to be split at a given level and the cutting value for the split are determined by maximizing LogWorth, a quantity related to the p-value associated with the sum of squares based on the difference in means. For continuous response variable the fitted values are the means within the two groups but for the case of a categorical response the splits are computed by maximizing a LogWorth statistic, which is related to the likelihood ratio chi-square statistic, known as $G^2$, also called log-likelihood ratio or Deviance ($D$). For the categorical response the fitted values are the estimated proportions within the groups.

The classification tree model can best be explained by a simple flow chart shown in Figure 3.1. The classification tree model consists of two types of nodes, branch nodes (including the root node) and leaf (or terminal nodes). Reference to Figure 3.1, Node 1 (root node) in the classification tree contains the entire sample dataset. Each of the remaining nodes (2 through 5) contains a subset of the entire dataset. Each branch node is a “parent” to two “children” nodes. For example, node 1 is split to produce nodes 2 and 3 and thus node 1 becomes their parent and then node 2 is the parent to nodes 4 and 5.
Classifications in classification pathways are represented by linking lines between each parent node and its children nodes. Identically, the classification tree begins by splitting original datasets found at root nodes into two subsets based on a particular attribute value test. This particular process gets repeated at each derived subset (branch node) in a recursive manner known as recursive partitioning. This repetitive procedure stops when the subset at a node has all the same value of the response variable, or when splitting no longer improves the predictions and this node becomes a leaf or terminal node (e.g., nodes 3, 4, and 5 in Figure 3.1).
3.1.3 Validation

Validation is one of the common techniques used in examining how strong a predictor (independent) variable predicts the response variable. In view of this procedure, the whole dataset is divided into two sets: the training dataset and the validation dataset. The training dataset is used in building (developing) the model while the validation dataset is used to evaluate the performance of the model built by using the training dataset. The essence here is to test the model by using the dataset that was not at all used in creating the model.

3.1.4 The Criteria of Node Splitting

The LogWorth statistic is a parameter usually used to grow as well as prune the classification tree model. It is used to indicate whether a particular predictor variable is significant or not. The larger the LogWorth value, the more significant the predictor variable is. The model generally splits the node based on the larger LogWorth statistic and is computed as shown in Equation 3.1:

\[
\text{LogWorth} = -\log_{10} (p - \text{value}) \quad \text{............................................... (3.1)}
\]

Where the adjusted p-value is calculated by taking into account the number of different ways splits can be made.
For a categorical response variable, the log-likelihood-ratio chi-square, $G^2$, is computed. This is essentially twice the change in the entropy. This entropy is computed as shown in Equation 3.2:

$$G^2 = 2 \sum f_o \log \left( \frac{f_o}{f_e} \right) \tag{3.2}$$

Where

$f_o = \text{observed frequency in a node}$

$f_e = \text{expected frequency in a node}$

A candidate $G^2$ that has been chosen for splitting is computed as shown by Equation 3.3:

$$G_{test}^2 = G_{parent}^2 - (G_{right}^2 + G_{left}^2) \tag{3.3}$$

### 3.1.5 Tests for Goodness-of-Fit

The method used in assessing the goodness-of-fit of a classification tree model depends on the type of response variable in the dataset. A scatter plot of actual versus predicted values is used for a continuous response variable and the receiver operating characteristics (ROC) curve is usually used for categorical response variables. The ROC curve is a statistical tool and a graphical plot of the sensitivity (true positive rate) vs. 1–specificity (false positive rate) used to
examine categorical response variables. This tool offers a complete and visually attractive way to determine the power or accuracy of the projected data (Agresti, 2007).

Agresti (2007) says that the accurate predictive performance of data analysis is achieved by the area under the ROC curve. In other words, the area found under the receiver-operating curve is used in ensuring that the probabilities of the predictions, as well as the outcomes, are compatible. The curve takes values from a range of zero to one where a value of 0 means a perfectly inaccurate test, and a value of 1 means a perfectly accurate test (SAS, 2003). The ROC curve usually takes a shape of a concave curve connecting the points (0,0) on the left lowest corner and (1,1) on the right top most corner (SAS, 2003), as shown in Figure 3.2. A ROC curve having a value of 0.5 indicates that the projections are essentially based on a random guess. Acceptable values should be larger than 0.5 and the closer to 1, the better the prediction.
3.2 Data Collection

The crash data used for this study were obtained from the Ohio Department of Public Safety (ODPS). Dataset used for this thesis study contains data collected within the last two and half years from July 2013 to December 2015. This section discusses in detail the data collection methodology; crash variables, merging data files and creation of the truck-only dataset.

3.2.1 Crash Data

Traffic crash data for the last five years, from 2011-2015, were downloaded from the ODPS website. These crash files are organized in relational format into four related files with all records compiled together by
calendar year. These four files are: (1) crash records, (2) unit records, (3) people records, and (4) ODOT records. Each file contains one common variable known as document number (DCONO) that relates records in all database to their respective accidents. In addition, the unit records file and people records file have an additional variable called unit number (UNITNO). The UNITNO together with the DOCNO are used to relate all people involved in traffic crashes to their correct vehicles they were traveling in and their specific crash incidents they were involved in. Consequently, these are two very important variables that are used to combine the four files together into one file by properly linking all related records together.

3.2.2 Merging Data Files

Each of the four files introduced above is briefly explained below:

a) The “crash records” file contains information specific to each crash event, e.g., crash severity, vehicle in error, date of crash, time of crash, name of city, village or township where the crash occurred, FIPS place code, crash location, type of road, if alcohol or drug was involved, if speeding was involved, etc.

b) The “unit records” file contains information on each unit/vehicle involved in that crash event. Information recorded includes unit type (e.g., motor vehicle, motorcycle, bicycle, pedestrian, etc.), point of impact, number of occupants in the unit, etc.
c) The “people records” file contains information on each person involved in each crash event with the exception of hit-and-run cases where information is always not available. Information recorded in this file includes person type (e.g., driver, occupant, or pedestrian), age, gender, severity of injury sustained by an individual, safety equipment used, etc.

d) The “ODOT records” file contains information such as county code, route type, latitude, longitude, crash type, etc.

In this study, the crash records, unit records and people records were needed to achieve the objectives of the study. These files were merged together by using a single-to-many-merging technique in SPSS software (version 22). The crash records and unit records were joined together by the common variable DCONO to create a “crash-unit” file. Then the joint “crash-unit” file was merged with their corresponding people records by using the two common variables DOCNO and UNITNO to create a joint “crash-unit-people” file for each year. After that, the joint “crash-unit-people” files for two and half calendar years were combined together to create a dataset of traffic crash data from July 2013 to December 2015. Figure 3.3 shows the crash data merging process.

Figure 3.3 The process of merging “crash-unit-people” file
3.2.3 Creating Truck Crashes Database

The resulting merged file containing all information about the crashes that occurred in Ohio from July 1, 2013 to December 31, 2015 was used to create the truck only crash dataset. Split file technique in SPSS software (version 22) was used to create a crash dataset that includes large trucks only by using the unit type variable. The final file that trucks has 31,712 records, which contains drivers only.

The Ohio crash report manual (OH-1) divides the truck type variable into the following categories: (1) single unit truck or van with two or more axles with six tires, (2) single unit truck with three or more axles, (3) single unit truck/trailer, (4) truck/tractor (bobtail), (5) tractor/semi-trailer, (6) tractor/doubles, (7) tractor/triples, and (8) other med/heavy vehicle. However, in this study the truck types were recoded into three groups: the first group contains single unit truck or van with two or more axles with six tires; the second group includes single unit truck with three or more axles, single unit truck / trailer and other med/heavy vehicles; and the third group includes tractor vehicles (i.e. truck/tractor (bobtail), tractor/semi-trailer, tractor/doubles, tractor/triples).

3.3 Description of Selected Variables

This section discusses the characteristics of truck-related crashes that occurred on Ohio’s road networks from July 2013 to December 2015. The characteristics of dependent and independent variables selected are described.
3.3.1 Description of Dependent Variable

The dependent variables are those variables that are subject to change based on the outcomes of the independent variables. For this study, the dependent variable is crash severity, which was recoded into a binary variable with two categories: (1) fatal/injury (F&I); and (2) property damage only (PDO). As shown in Figure 3.4, the fatal and injury crash category constitutes 6,683 (21.1%) crash records, and the property damage only crash category contains 25,029 (78.9%) crash records. The objective of this section is to describe the characteristics of fatal and injury crashes.

![Crash Severity Description](image)

Figure 3.4 Description of crash severity

3.3.2 Description of Independent Variables

Independent variables can generally be divided into three major categories as depicted in Tables 3.1 through 3.3; which are human/driver
characteristics, roadway/environmental characteristics and crash/vehicle characteristics respectively.

### 3.3.2.1 Driver Characteristics

The human factors, as the term implies, refer to the factors that are determined by the drivers involved in the crash incidents. Specifically, these include gender, age, teen-related, alcohol-related, drug-related, and speed-related factors as shown in Table 3.1.

Table 3.1 shows variables pertaining to the driver as human factors. Age of the truck driver was recoded into age groups as follows: (1) younger than or equal to 25 years old; (2) 26-35 years old; (3) 36-45 years old; (4) 46-55 years old; and (5) older than 55 years old. Data show that 8.9% of the truck driver 25 years old or younger; 17.3% of the drivers were 26-35 years old; 22% of the drivers were 36-45 years old; 26.2% were 46-55 years old; and 22% of the drivers were older than 55 years.. The age of 3.7% of the drivers was unknown or missing data. The gender distribution of the truck drivers involved in crashes shows that, 92.2% of the drivers were males, 3.4% were females, and 4.4% of them were unidentified.

Similarly, data shows that 7.4% of the crashes were teen-related, i.e., one of the driver involved in the crash was a teenager driver. Alcohol was involved in about 1.2% of the truck-related crashes, and drug involvement was reported in
only 0.7% of the truck-related crashes. Speeding was reported be involved in 8.6% of the truck-related crashes.

Table 3.1 Description of human and driver variables

<table>
<thead>
<tr>
<th>Variables</th>
<th>Description</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGE</td>
<td>Age &lt;= 25</td>
<td>2810</td>
<td>8.9</td>
</tr>
<tr>
<td></td>
<td>25&lt; Age&lt;=35</td>
<td>5478</td>
<td>17.3</td>
</tr>
<tr>
<td></td>
<td>35&lt; Age&lt;=45</td>
<td>6963</td>
<td>22.0</td>
</tr>
<tr>
<td></td>
<td>45&lt; Age&lt;=55</td>
<td>8305</td>
<td>26.2</td>
</tr>
<tr>
<td></td>
<td>Age &gt;55</td>
<td>6972</td>
<td>22.0</td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>1184</td>
<td>3.7</td>
</tr>
<tr>
<td>GENDER</td>
<td>Male</td>
<td>29239</td>
<td>92.2</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>1091</td>
<td>3.4</td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>1382</td>
<td>4.4</td>
</tr>
<tr>
<td>TEEN RELATED</td>
<td>No</td>
<td>29367</td>
<td>92.6</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>2345</td>
<td>7.4</td>
</tr>
<tr>
<td>ALCOHOL RELATED</td>
<td>No</td>
<td>31316</td>
<td>98.8</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>396</td>
<td>1.2</td>
</tr>
<tr>
<td>DRUG RELATED</td>
<td>No</td>
<td>31493</td>
<td>99.3</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>219</td>
<td>0.7</td>
</tr>
<tr>
<td>SPEED RELATED</td>
<td>No</td>
<td>28985</td>
<td>91.4</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>2727</td>
<td>8.6</td>
</tr>
</tbody>
</table>

3.3.2.2 Road and Environmental Characteristics

Table 3.2 depicts variables that represent road and environmental characteristics. Data show that 16.1% of truck-related crashes occurred on roadways with posted speed limits of less than 35 mi/h; 20% of truck-related crashes happened on roadways with posted speed limits between 35-40 mi/h, and 11% of truck-related crashes happened on roadways with posted speed limits between 45-55 mi/h. In addition, 24.1% of truck-related crashes occurred on roadways with posted speed limits of 55 mi/h, and 25.7% of truck-related crashes happened on roadways with posted speed limits higher than or equal to
60 mi/h. About 2.8% of truck crashes occurred on roads with unknown or missing posted speed limits.

Roadway alignment (contour) was divided into four groups: straight level, straight grade, curve level and curve grade. Data shows that 71.8% of truck-crashes occurred on straight level segments and 17.3% of truck-related crashes occurred on straight grade segments. In addition, about 4.9% of truck-related crashes occurred on curve level segments and 5.8% of truck-related crashes occurred on curve grade segments. The roadway light conditions when truck-related crashes occurred show that most of the truck-related crashes (74.8%) occurred during daylight, and 3.6% occurred during dawn or dusk. Moreover, 8.6% and 11.8% of truck-related crashes occurred on dark-lighted and dark-not-lighted roadways, respectively. The percentages of intersection and work zone crashes associated with truck crashes were 24.6% and 8.6%, respectively. The road condition was divided into three categories: dry, wet and adverse condition (i.e. snow, ice, sand, slush, etc.). About 81.5% of the truck-related crashes happened on dry roadways, while only 17.8% occurred on wet roadways, and 9% occurred on adverse road conditions.

When it comes to the time when truck-related crashes happened, data show that 85.3% of crashes took place during the early morning/daytime period (04:00-18:59), 8.7% happened during the early night period (19:00 to 22:59), and 5.9% during late night period (23:00-03:59).
<table>
<thead>
<tr>
<th>Variables</th>
<th>Description</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>POSTED SPEED</td>
<td>PS&lt;35</td>
<td>5196</td>
<td>16.4</td>
</tr>
<tr>
<td></td>
<td>35≤PS&lt;45</td>
<td>6335</td>
<td>20.0</td>
</tr>
<tr>
<td></td>
<td>45≤PS&lt;55</td>
<td>3492</td>
<td>11.0</td>
</tr>
<tr>
<td></td>
<td>PS = 55</td>
<td>7647</td>
<td>24.1</td>
</tr>
<tr>
<td></td>
<td>PS≥60</td>
<td>8157</td>
<td>25.7</td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>885</td>
<td>2.8</td>
</tr>
<tr>
<td>ROAD ALIGNMENT</td>
<td>Straight Level (SLVL)</td>
<td>22761</td>
<td>71.8</td>
</tr>
<tr>
<td></td>
<td>Straight Grade (SGRD)</td>
<td>5495</td>
<td>17.3</td>
</tr>
<tr>
<td></td>
<td>Curve Level (CLVL)</td>
<td>1553</td>
<td>4.9</td>
</tr>
<tr>
<td></td>
<td>Curve Grade (CGAD)</td>
<td>1848</td>
<td>5.8</td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>55</td>
<td>0.2</td>
</tr>
<tr>
<td>LIGHT CONDITION</td>
<td>Daylight (DL)</td>
<td>23730</td>
<td>74.8</td>
</tr>
<tr>
<td></td>
<td>Dawn/ Dusk (DD)</td>
<td>1196</td>
<td>3.8</td>
</tr>
<tr>
<td></td>
<td>Dark Lighted Roadway (DLR)</td>
<td>2730</td>
<td>8.6</td>
</tr>
<tr>
<td></td>
<td>Dark Roadway Not Lighted (DNLR)</td>
<td>3730</td>
<td>11.8</td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>326</td>
<td>1.0</td>
</tr>
<tr>
<td>INTERSECTION</td>
<td>No</td>
<td>23925</td>
<td>75.4</td>
</tr>
<tr>
<td>RELATED</td>
<td>Yes</td>
<td>7787</td>
<td>24.6</td>
</tr>
<tr>
<td>WORK ZONE RELATED</td>
<td>No</td>
<td>28989</td>
<td>91.4</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>2723</td>
<td>8.6</td>
</tr>
<tr>
<td>ROAD CONDITION</td>
<td>Dry</td>
<td>25846</td>
<td>81.5</td>
</tr>
<tr>
<td></td>
<td>Wet</td>
<td>4769</td>
<td>15.0</td>
</tr>
<tr>
<td></td>
<td>Adverse (adv)</td>
<td>1097</td>
<td>3.5</td>
</tr>
<tr>
<td>TIME OF CRASH</td>
<td>Early morning/daytime (04:00-18:59)</td>
<td>27061</td>
<td>85.3</td>
</tr>
<tr>
<td></td>
<td>Early night (19:00-22:59)</td>
<td>2766</td>
<td>8.7</td>
</tr>
<tr>
<td></td>
<td>Late night (2300-0359)</td>
<td>1885</td>
<td>5.9</td>
</tr>
<tr>
<td>DAY OF WEEK</td>
<td>Week Day (WD)</td>
<td>28022</td>
<td>88.4</td>
</tr>
<tr>
<td></td>
<td>Weekends (WE)</td>
<td>3690</td>
<td>11.6</td>
</tr>
<tr>
<td>WEATHER</td>
<td>Clear</td>
<td>18567</td>
<td>58.5</td>
</tr>
<tr>
<td></td>
<td>Cloudy</td>
<td>8859</td>
<td>27.9</td>
</tr>
<tr>
<td></td>
<td>Rain</td>
<td>2875</td>
<td>9.1</td>
</tr>
<tr>
<td></td>
<td>Adverse (adv)</td>
<td>1411</td>
<td>4.4</td>
</tr>
</tbody>
</table>

The days when the truck-related crashes occurred were grouped into weekdays (Monday through Friday) and weekends (Saturday and Sunday).
data show that 88.4% of truck-related crashes occurred during weekdays 11.6% during weekends. When assessing the weather condition when truck-related crashes occurred, we observe that 58.5% of crashes took place when the weather was clear, 27.9% happened when the weather was cloudy, 9.1% occurred when it was raining, and 4.4% took place during adverse weather conditions (i.e., snow, fog, severe wind, sleet and blowing sand).

### 3.3.2.3 Vehicle and Crash Characteristics

Table 3.3 shows vehicle and crash related variables. About 22.4% of truck-related crashes involved single unit trucks with two axles, 26.1% involved single unit trucks with three or more axles, and 51.5% involved tractor-trailer combination trucks. Collision types were grouped into four categories: (1) non-collision event or single-vehicle collision; (2) collision with another motor vehicle in transport; (3) collision with a movable object such as a pedestrian or an animal; and (4) collision with a fixed object. Single-vehicle collisions made up 18.3% of the truck-related crashes. Collision with another vehicle in transport constituted 66.3%, and collision with movable and fixed object made up 5% and 10.1% of the truck-related crashes, respectively.

In terms of crash type, the non-collision between motor vehicle in transport made up 28.6% of the crashes. The rear crash type (which includes rear-end, rear-to-rear, and backing crashes) comprised 23.9% of truck-related crashes; the head-on and angle crash type made up 1.2% and 19% of truck-related crashes, respectively; and sideswipe crashes (which includes sideswipe in same direction and sideswipe in opposite direction) made up 26.3% of truck-related crashes.
### Table 3.3 Description of vehicle and crash variables

<table>
<thead>
<tr>
<th>Variables</th>
<th>Description</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>UNIT TYPE</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Single Unit Truck 2 Axles (SUT 2xls)</td>
<td>7103</td>
<td>22.4</td>
</tr>
<tr>
<td></td>
<td>Single Unit Truck 3 Axles or more (SUT 3xls)</td>
<td>8283</td>
<td>26.1</td>
</tr>
<tr>
<td></td>
<td>Tractor-Trailer Combinations (TTC)</td>
<td>16326</td>
<td>51.5</td>
</tr>
<tr>
<td><strong>COLLISION EVENT</strong></td>
<td>Non Collision Event (NCE)</td>
<td>5790</td>
<td>18.3</td>
</tr>
<tr>
<td></td>
<td>Collision with Motor Vehicle In Transport (CMVT)</td>
<td>21013</td>
<td>66.3</td>
</tr>
<tr>
<td></td>
<td>Collision with Unfixed/movable Object (CMObj)</td>
<td>1575</td>
<td>5.0</td>
</tr>
<tr>
<td></td>
<td>Collision with Fixed Object (CFObj)</td>
<td>3227</td>
<td>10.2</td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>107</td>
<td>0.3</td>
</tr>
<tr>
<td><strong>COLLISION TYPE</strong></td>
<td>Not Collision between Motor Vehicle In Transport (NC)</td>
<td>9067</td>
<td>28.6</td>
</tr>
<tr>
<td></td>
<td>Rear</td>
<td>7567</td>
<td>23.9</td>
</tr>
<tr>
<td></td>
<td>Head on (HON)</td>
<td>375</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>Angle (Ang)</td>
<td>6013</td>
<td>19.0</td>
</tr>
<tr>
<td></td>
<td>Sideswipe (sidwip)</td>
<td>8330</td>
<td>26.3</td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>360</td>
<td>1.1</td>
</tr>
</tbody>
</table>
CHAPTER IV

RESULTS AND DISCUSSION

4.1 Introduction

A number of different factors contribute to the increase in truck-related crash severity, and these factors can generally be categorized as human and driver-related; road geometric-related; crash-related; and environment-related factors. In this study, several factors were statistically analyzed in order to determine the most significant contributors. There were 18 independent (explanatory) variables that were selected and investigated in an attempt to identify their impact in crash severity levels of large truck-related crashes. The effects of these explanatory variables on large truck crash severity were estimated statistically by use of the classification tree model routines in JMP software package (version 12). The first part of this chapter presents the descriptive results and the second part discusses the results of the classification tree model.

4.1.1 Descriptive Results of Truck-Related Traffic Crashes

Figure 4.1 shows the number of fatal and injury crash severity level caused by the different crash types. From Figure 4.1, it is obvious that head-on type of crashes were the most dangerous where 65.1% of head-on truck-related
crashes resulted into fatalities and injuries. The second most dangerous are angle type crashes, for which 32.7% of them resulting into either fatalities or injuries, and followed closely by rear-related crashes with a fatality/injury rate of 28.8%.

Figure 4.1 Distribution of crash severity by the crash type

Figure 4.2 illustrates the percent of fatal and injury crashes based on collision events. The most hazardous collision event involved a collision with motor vehicle in transport with where 24.8% of these events were fatal or injury crashes. The next highest collision event that resulted into higher fatality and injury rates was the non-collision event with 20.9% of occurrences. The other collision types, collision with movable objects and collision with fixed objects had lower rates of fatal/injury occurrences, with 8.9% and 3.3%, respectively.
Figure 4.2 Distribution of the crash severity by collision event

Figure 4.3 shows the percentage of truck-related crashes that were fatal or injury crashes by posted speed limits. It is clear that roadways with posted speed limits above 40 mi/h have higher percentages of their crashes resulting into fatalities or injuries. Roadways with posted speed limits of 55 mi/h happened to have the highest rate of 26.5%, followed closely with roadways having posted speed limits between 45 mi/h and 50 mi/h with a rate of 24.9%. Roadways with posted speed limits 60 mi/h and above wind up the top three with a rate of 23.3% of fatal and injury crashes.
Figure 4.3 Distribution of the crash severity by posted speed

Figure 4.4 shows the rates of fatal and injury truck-related crashes by road alignment type. It can be seen that the highest rate of fatal and injury crashes of 25.2% occurred on curve grade segments, followed with crashes on straight grade segments with a rate of 22.1%. Although most of the truck-related crashes occurred in straight level road segments (refer to Table 3.2), but they have the lowest rates of fatal/injury crashes at 20.5%.
Figure 4.4 Distribution of crash severity by road contour

Figure 4.5 shows the effect of the influence of alcohol use to the rate of fatal and injury crashes. When alcohol use was involved in the truck-related crashes, the rate of fatal or injury crashes was 52.5% as compared to only 20.7% when alcohol was not involved. Again, even the incidents were alcohol use was very small (only 1.2%, refer to Table 3.1), but they were very dangerous as more than half of them resulted into fatalities or injuries. Figure 4.6 shows the effect drug use to the rate of fatal and injury crashes. When drug use was involved in truck-related crashes, the rate of fatal and injury crashes was 55.3%. Similar to alcohol use, drugs-related truck crashes made up of a tiny percentage of total crashes (0.7%, refer to Table 3.1), but they result into higher fatal and injury rates.
4.2 Result of Classification Tree Modeling

4.2.1 Introduction

The tree classification model was applied in the current study to investigate the relationship between the crash severity of truck-related crashes...
and a number of selected independent (predictor) variables. There were 19 variables selected from the ODPS datasets for this study. The crash severity was the dependent variable, which consists of two categories fatal/injury (F&I) and property damage only (PDO) as an ordinal scale, whereas, and the remaining variables were the predictors. The predictors have different model measurements; ten of them were polytomous and the others were binary. The polytomous variables include age, posted speed limit, road condition, road alignment, weather, light condition, unit type, time of crash, collision event and crash type. The binary variables include gender, intersection-related, alcohol-related, drug-related, teen-related, work zone-related and speed-related.

4.2.2 The Ideal Size of Tree

The dataset used for this study contained a total number of 31,712 complete data records. JMP software was to run the classification tree model by setting the crash severity variable as the target variable. In addition, the dataset was divided into two subsets; one for validation and the other for training. The validation subset was set to contain 30% of the observations, and the training sample was to have 70% of the observations, and the actual selection of which data records go into either of the two subsets was randomly selected by the software.

Figure 4.7 displays the history of tree splitting, which grows the tree based on the $R^2$ and number of splits, for both validation and training datasets. Clearly, the tree will stop growing up after 42 splits with overfitting issues. Overfitting of
the tree model creates more variables, which is not desirable. These issues can be avoided by reducing the split numbers. In Figure 4.7, it can be observed that the validation and training samples have almost the same $R^2$ after split number 10 and the curve remains fairly flat, which means there no significant improvement to the model predicting power after 10 splits. Based on that criterion, the ideal tree model should stop splitting after 10 splits to get the most significant predictors. Figure 4.8 shows the split history of validation and training subsets for ideal size tree.

![Split History](image)

Figure 4.7 Split history of validation and training subsets for full tree
4.2.3 Model Validation

As previously mentioned, the model can be tested by using the validation subset. The classification tree model creates two receiver-operating characteristics (ROC) curves for validation and training subsets. Figure 4.9 and 4.10 show the ROCs of the training and validation datasets respectively. It is obvious that all curves for levels of crash severity (F&I, PDO) are above the 45-degree diagonal line, which means that their rates are more than the random guess prediction point (0.5). Therefore, the classification tree model is satisfactory to investigate the relationship between the crash severity level and other predictors. In addition, the Figure 4.9 and 4.10 indicate the prediction degree of accuracy for F&I and PDO as 70.3% and 70.6% respectively.

Figure 4.8 Split history of validation and training subsets for ideal size tree
4.2.4 The Significant Predictors

There were eighteen predictors used in an attempt to identify the ones that statistically significant in predicting crash severity levels of truck-related crashes.
Table 4.1 shows the JMP’s column contributions report. The important predictors with significant $G^2$ statistic values were crash type, posted speed limit, collision event, speed-related and road alignment. Crash type is a single most significant predictor of truck-related crashes having the highest $G^2$ statistic value and predicting about 55.8% of the variation in truck-related crash severity.

![Table 4.1 The column contributions report for selecting significant predictor variables](image)

4.2.5 Final Classification Tree

Figure 4.11 displays the final classification tree of the training sample with 11 terminal (leaf) nodes. Each node shows the target level rates, data count, and $G^2$ values. The branch nodes also include LogWorth values, for which tree splitting is based on. The total number of splits was stopped at ten, which was determined to be ideal size for this tree according to the split history curve.
Therefore, crash type, collision event, posted speed limit, speed-related and roadway contour were determined to be significant predictor variables of truck-related crash severity by the classification tree.
Figure 4.11 Final classification tree results
Interpretation of the results of Figure 4.11 is somehow straightforward as introduced in Section 3.1.2. First of all, the root node, which contains the entire dataset, was split based on the variable crash type, which means that the crash type variable was the most significant predictor in predicting the crash severity of truck-related crashes. The classification tree model did split the crash types sideswipe and non-collision with motor vehicle in transport into the left child node of the main tree; and the rest of the crash types (i.e. rear, angle and head-on) to the right hand child node. It can be seen that the crash types rear, angle and head-on were more hazardous than the sideswipe and non-collision with motor vehicle in transport. Fatal and injury crashes made up 30.7% of the total crashes in the rear, angle and head-on crashes group versus 12.9% for the sideswipe and non-collision with motor vehicle in transport group as shown in Figure 4.12.

On the left side of the main tree, the collision types sideswipe (sidwip) and non-collision with motor vehicle in transport (NCMVT) node was split based on the collision event as shown in Figure 4.13. The collision events collision with fixed object (CFObj) and collision with movable objects (CMObj) (such as pedestrian, animal and other unfixed objects) formed the left side child node with
4.6% of its crashes being fatal or injury. The other collision event types in the right child node, not-collision event (NCE) and collision with motor vehicle in transport (CMVT), 15.5% of their crashes were either fatal or injury crashes. Therefore, non-collision and collision with motor vehicle in transport were more dangerous compared to collision with fixed object and collision with movable object events.

![Figure 4.13 Crash type (sidwip, NCMVT) node split by collision event](image)

After that, the node of collision event types non-collision event and collision with motor vehicle in transport was split based on the posted speed limit variable. When the posted speed limit was higher than or equal to 45 mi/h, the percent of fatal and injury went up to 17.3% whereas, when the speed limit was less than 45 mi/h, the percent of the fatal and injury severity was only 10.2%. Figure 4.14 shows how the collision event (CMVT and NCE) node was split. This shows that when the posted speed limit was 45 mi/h or higher, the probability of truck-related crashes involving collisions with motor vehicle in transport and non-collision events resulting into fatal or injury increased.
Figure 4.14 Collision event (CMVT and NCE) node split by posted speed limit

Figure 4.15 shows that the node for posted speed limits of 45 mi/h and higher was split by the variable road contour, with two child nodes where straight level (SLVL) and straight grade (SGRD) segments were placed into one child node and curve level (CLVL) and curve grade (CGRD) segments were placed into the other. For road segments with posted speed limits of 45 mi/h and higher; truck-related crashes on road contours SLVL and SGRD have lower chances of being fatal or injury (15.9%) as compared to road contours CLVL and CGRD whose percent of fatal and injury chances was 24.2%. Furthermore, the node for road contours with higher probability of injury and fatal crashes (CLVL, CGRD) was split by the variable crash type. Thus, non-collision (overturn and rollover) crash types that occurred on curve level and curve grade segments had higher chances of being fatal or injury (31.2%) than sideswipe crashes that occurred on the same type of segments with the probability of injury and fatal of 15.6%.
Figure 4.15 The split of the node for posted speed limits 45 mi/h or higher and road contour

On the right side of the main tree we find a node with crash types rear, head-on and angle, which resulted into more severe injuries after splitting the root node. Figure 4.16 shows that this node was further split by the variable posted speed limit. While posted speed limits less than 45 mi/h were grouped into a left child node, posted speed limits of 45 mi/h or higher were grouped into a right child node. Truck-related crashes that occurred on roadways with higher posted speed limits (i.e. above or equal to 45 mi/h) had a higher probability of fatal and injury severity of 38.1%. Whereas, truck-related crashes that took place on roadways with posted speed limits less than 45 mi/h had a lower chance (21.6%) of fatal or injury severity. As expected, rear, angle, and head-on crashes involving heavy trucks that occurred on roadways with high posted speed limits had a higher probability of having fatal and injury crashes.
Figure 4.16 The crash types rear, head-on and angle node split based on the posted speed limit

A node of lower posted speed limits was further split by the variable collision events as shown in Figure 4.17. The split created two child nodes: the collision with fixed object event (CFObj) in the left child node and the other collision events (i.e. collision with motor vehicle in transport (CMVT), not-collision event (NCE) and collision with unfixed/moving object (CMObj)) were grouped together into the right child node. It can be seen that the collision events NCE, CMVT and CMObj that occurred on roadways with lower posted speed limits had a higher likelihood of resulting into injuries and fatalities (22.9%) than the collisions with fixed objects, which had a much lower likelihood of injuries and fatalities of about 2.5%.
Collision events (i.e. collision with motor vehicle in transport (CMVT), not-collision event (NCE) and collision with unfixed/moving object (CMObj)) node was further split by the variable speed-related as shown in Figure 4.18. When speeding was involved for CMVT, NCE, and CMObj crash events involving heavy trucks, the probability of fatal and injury severity was higher at 52.4%. On the other hand, the probability of the fatal and injury was lower (21.4%) when speeding was not involved in such crashes. Therefore, this supports a commonsense notion that crashes will likely be more severe if speeding is involved in heavy truck crashes.
Figure 4.18 Splitting of collision events (CMVT, NCE and CMObj) node by speed-related

On the other hand, the higher posted speed limits node (45mi/h or higher) was split by the variable crash type as shown in Figure 4.19. It can be seen that it was divided into two nodes where crash types rear, and angle were placed into the left side child node and crash type head-on was placed in the right hand child node. The results show that the head-on crash type was more dangerous with a higher probability of fatal and injury crashes (73.5%) than the rear and angle crash types whose probability of fatal and injury crashes was only 37.0%.

Figure 4.19 Posted speed limits 45mi/h or higher node split by crash type
Figure 4.20 shows the results of speeding on crash types rear and angle occurring on high posted speed limits (i.e. 45 mi/h or higher). The crash types (rear and angle) node was split by the variable speed-related. When speeding was involved in rear and angle crashes that were truck-related, the probability of fatal and injury severity was 51.2%, which is much higher than when speeding was not involved whose probability of fatal and injury was 35.6%. When all leaf nodes become pure, i.e. the classification tree model does not find reasonable or significant differences in the data, then splitting ends.

Figure 4.20 Splitting of crash type (rear and angle) node by speed-related

4.3 Discussion

In this study, it was determined that crash type, collision event, posted speed limit, speed-related, and road contour were the significant factors contributing to injury and fatal severities of heavy truck-related crashes. The crash type was the most significant one because the root node was first split by this variable and the column contributions report shows that it explains about 55.8% of the variation in the dependent variable, crash severity. It was also a significant factor found in other earlier similar studies (e.g. Shankar et al., 1996;
Dissanayake and Kotikalapudi, 2012). Furthermore, roadway posted speed limit (18.5%) and collision event (17.7%) were also identified as significant factors in predicting the crash severity of truck-related crashes. Speed-related variable was the fourth significant variable, which contributed about 6.0% in predicting truck-related crash severity. Similarly, Chang and Mannering (1999) and Edrissi and Nassiri (2006) in their previous studies found speed-related variable to be a significant factor affecting truck-related crashes.
CHAPTER V
CONCLUSION AND RECOMMENDATIONS

A truck-related crash is a crash that occurs when a truck collides with another vehicle or motorcycle, or hits a pedestrian or in a single-vehicle crash where a truck leaves its designated traveled way and in the process it collides with a fixed object or simply overturns. The main objective of this thesis study was to determine the factors that contribute significantly to the levels of crash severity when truck-related crashes occur. Based on this objective, a two and half-year crash data from July 2013 to December 2015 obtained from the Ohio Department of Public Safety was used for this analysis. In this study, a classification tree model was used to investigate significant factors that contribute truck-related crash severity in Ohio.

The classification tree model was used because it has the ability to detect the important predictors for crash severity. The classification tree procedure identified only five factors, which are considered to explain a large amount of the variation in the response variable (i.e., crash severity levels) out of eighteen independent variables considered in the current study. These important predictors of crash severity include crash type, collision events, posted speed limit, speed-related, and road alignment only. The classification tree model
determined that the most severe truck-related crashes occurred when trucks get involved in head-on, angle, and rear-end type of collisions with other vehicles in transport. The presumable reason is that occupants of other vehicles, which are smaller in size, (i.e., passenger vehicles) are most likely to sustain severe injuries when they collide with trucks while both types of vehicles are traveling at highway speeds. The impact due to shear momentum of the truck vehicle is enough to cause severe damage to the other vehicle. In addition, the tree reveals that when these crashes occur on high posted speed limit roads, head-on types tend to increase more chances of injuries and fatalities than the other two types. The collision events (i.e., types) variable also shows truck-related crashes are more dangerous when trucks collide with another vehicle in transport (i.e., multi-vehicle collision) compared to single-vehicle collision (i.e., when the truck collides with a fixed object or none-collision event). Furthermore, the classification tree shows that truck-related crashes increase in severity when speeding is involved. A crash is considered speeding-related if the driver was charged with a speeding-related offense or if an investigating police officer indicates that racing, driving too fast for conditions, or exceeding the posted speed limit was a contributing factor in the crash. Lastly, the classification tree shows that road contour is another variable that affects the crash severity of truck-related crashes. Specifically, according to the tree when a truck–related crash occurs on medium to high speed limit roads (i.e., posted speed limits of 45 mi/h and above), with curve-grade and curve-level segments, they tend to increase the likelihood of causing fatal and injuries.
The current study recommends further studies to be done in the future to investigate the effect of Ohio’s legislature to raise the speed limits on rural interstates from 65 mi/h to 70 mi/h starting July 1, 2013. Although we had data from January 1, 2011, we did not want to include data before July 1, 2013 due to that major change in speed limits on rural Interstate roads and our datasets would have overlapped between these time periods. This may play a role in raising severity levels of truck-related crashes in the state due to increasing the chances of trucks colliding with other vehicles at relatively higher speeds than before.
REFERENCES


