2013

Interaction effects of prevailing weather conditions and crash characteristics on crash severity: A case study on two corridors in Iowa

Dimitrios Vasilios Bilionis
Iowa State University

Recommended Citation
Bilionis, Dimitrios Vasilios, "Interaction effects of prevailing weather conditions and crash characteristics on crash severity: A case study on two corridors in Iowa" (2013). *Graduate Theses and Dissertations*. 13182.
https://lib.dr.iastate.edu/etd/13182

Interaction effects of prevailing weather conditions and crash characteristics on crash severity: A case study on two corridors in Iowa

by

Dimitrios Vasilios Bilionis

A thesis submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of
MASTER OF SCIENCE

Major: Civil Engineering (Transportation Engineering)

Program of Study Committee:
Konstantina Gkritza, Co-Major Professor
Omar Smadi, Co-Major Professor
Peter Sherman

Iowa State University
Ames, Iowa
2013

DEDICATION

This thesis is dedicated to my wonderful mother Nicoletta Bilioni. Her love, patience and support have been essential parts of my life and serve as inspirations in order to achieve my personal goals.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

First of all, I would like to thank my major adviser Dr. Konstantina Gkritza for her support and contribution which have been priceless components of my graduate studies.

I would also like to thank Dr. Omar Smadi and Dr. Peter Sherman who served as my Program of Study Committee Members, especially for their useful comments and recommendations which provided me with valuable insights for my thesis.

Furthermore, I would like to thank Tina Greenfield and Brian Carlson from the Iowa Department of Transportation for providing me with data useful in my analysis. I also appreciate Zach Hans and Inya Nlenaya from the Institute for Transportation (InTrans) for their support, especially in data processing.

I am finally grateful to Dr. Konstantina Gkritza and Dr. Omar Smadi for funding my graduate studies and thus giving me the once in a lifetime opportunity to study in a prestigious institution as the Iowa State University.

Last but not least, I would like to thank Dr. Basak Aldemir-Bektas for her support and our excellent collaboration with regards to my duties as a graduate research assistant at InTrans and also my fellow students who have been valuable "co-travelers" in the two-year journey of my graduate studies.

ABSTRACT

It has been widely accepted that weather has a significant impact on road safety. The large body of literature points out that weather is an environmental factor that affects both frequency and severity of crashes. Research has shown that especially adverse weather conditions are associated with increase in the numbers of crashes and rates. Furthermore, the prevailing weather conditions may influence the severity outcome of a crash. However, weather may be a factor that modifies crash conditions and not the major cause of a crash. Thus, any interactions between the prevailing weather conditions and other crash specific variables especially on crash severity should be taken into account.

In view of the above, the purpose of this thesis is to investigate the interaction effects of the prevailing weather conditions in combination with other crash characteristics on crash severity. To do so, a study on two different corridors in Iowa was conducted. Specifically, road segments from two different facilities, an Interstate route (I-80) and a US route (US-34), were selected and the corresponding crash severity was examined by estimating discrete outcome models.

The estimation results show that adverse weather conditions when interacting with other crash attributes influence crash severity. Among the weather conditions, temperature below freezing, precipitation (type and amount) and wind (speed and direction) were found to contribute to the severity outcome of crash. However, the combination of the prevailing weather conditions and route classification may have diverse effects on crash severity. For instance, adverse weather was found to reduce the probability of very severe crashes on the interstate study corridor, while the opposite effect was observed on the US route corridor.

The results of this thesis could provide transportation agencies with useful insights about the maintenance and operation activities that should be undertaken on different roadway facilities, especially under adverse weather conditions. Finally, the findings of this study can have potential implications in driver education as well as informing road users about the various effects of weather on safety with an emphasis on safe driving under inclement weather.

CHAPTER 1. INTRODUCTION

1.1 Background Summary and Problem Statement

Road crashes currently constitute one of the world's leading causes of death. According to the World Health Organization (WHO), approximately 1.24 million people die and another 20 to 50 million are involved in non-fatal injury crashes globally on an annual basis. These findings rank road crashes as the eighth leading cause of death in today's world (WHO, 2013). In the United States (U.S.), 32,367 people died and around 2.22 million were injured in 2011 (NHTSA, 2012). Only in the state of Iowa, 360 people lost their lives and 28,396 were injured in a total of 48,713 crashes that occurred on the state's roadway network in 2011 (Iowa Department of Transportation, 2012).

It has been well established that the factors that contribute to a road crash can be categorized into three major groups: a) driver conditions and behavior; b) roadway design and environment; and c) vehicle. Weather is one of the factors (typically classified into group b) that have a significant effect on road safety, since weather conditions partly determine the road conditions and driver's behavior (SWOV, 2012). Specifically, weather can affect both the collision and casualty rates by affecting road surface and environment conditions (for example, reduction in pavement friction, impaired visibility, etc.) and drivers' behavior (for example, difficulty in vehicle steering and handling, lost control, etc.). The previous effects are also more intense in case of adverse weather conditions which can be considered as a chronic hazard for road users (Andrey et al., 2003), especially in countries with long periods of winter such as Canada or the north and central states of the U.S.

Based on the above context, the influence of weather conditions on road safety has attracted the attention of several researchers in recent years. Specifically, numerous papers have been published on the effect of weather parameters (such as temperature, precipitation, etc.) on the frequency and severity of crashes. These studies are described in Chapter 2 of this thesis. Nevertheless, most of those studies examined the effect of weather conditions as a single variable among all the other factors that can affect a crash. However, one should not ignore that weather conditions (such as temperature) may be a modifier factor of crash conditions and not a major cause (see Andreescu and Frost, 1998). Thus, any interaction effects between the weather conditions and other crash specific variables (such as type of collision, vehicle speed, road classification, etc.) may be ignored when considering the weather conditions as a single variable. Therefore, there is a need to study the interaction effects of weather conditions along with other crash specific factors on crash severity on different types of facilities in a bid to fully investigate the combination of factors influencing crash severity.

The need for such study in Iowa is of particular importance as winter weather related crashes in the state are very frequent. Historical data shows that during the winters of 1995/1996 to 2004/2005 approximately one-third of all crashes that occurred on rural, state-maintained highways in Iowa, were winter weather-related. Moreover, approximately half of the rural Interstate crashes were winter-weather related (Hans, et al., 2011). Furthermore, the Iowa Department of Transportation (Iowa DOT) spends significant amount of its budget on winter maintenance operations and also, regularly invests funds to examine the safety and mobility impacts of winter weather.

1.2 Research Objectives and Tasks

The main objective of this thesis is to investigate the interaction effects of weather conditions and other crash-specific factors on crash severity. Moreover, an objective of this thesis is to constitute a case study of different roadway facilities of the Iowa network. Thus, this study aims to provide with results that could be used to better understand the influence of the prevailing weather conditions in combination with other crash attributes on crash severity, based also on the type of facility where the crash occurred. These results could give useful insights and recommendations, firstly to road users in adjusting their driving behavior and secondly to agencies (e.g. the local DOT) in forming their maintenance and operational activities according to the prevailing weather conditions, especially during periods of inclement weather.

The following presents the main tasks of this thesis:

Task 1: Literature Review

Past work on the impact of weather on road safety is reviewed and synthesized. The main focus is placed on two major areas. First, studies that investigated the effect of the weather conditions on crash frequency and severity are examined. Thus, a summary and critical synthesis of the findings is performed. Secondly, a review of the different methodologies and data collection techniques that have been applied in past road safety studies and especially in those related to weather effects is conducted. Thus, a list of applicable approaches and methodologies to accomplish the research objectives of this thesis is created.

Task 2: Selection of the Study Area and Period, Data Collection and Integration

The area of study is selected based on the characteristics of the corridors and the number of crashes (sample size). Specifically, two corridors: a 4-lane divided facility (I-80) and a 2-lane undivided facility (US-34) are selected. The analysis period is from 2009 to 2011. Then, four types of information are collected. Information regarding the crashes (crash data) is collected by the Iowa DOT crash datasets. Information regarding the prevailing weather conditions (weather data) is obtained from the nearest RWIS stations. Roadway and Traffic Data is acquired from the Iowa DOT GIMS files and the records of ATR stations. After all the (separate) datasets are analyzed and processed, the integration of all data is performed. Eventually, a single dataset for each corridor is created. Those datasets are the inputs of the statistical analysis.

Task 3: Selection of Methodology

The most suitable statistical methodology is selected based on the review and synthesis that will be performed in Task 1. Specifically, discrete outcome probability models are selected as the most appropriate method in order to approach the thesis objective. After the selection, the mathematical background and properties of these methods are studied.

Task 4: Statistical Analysis of Data

Two different types of discrete outcome probability models are created for each corridor of study, namely a binary probit model and a multinomial logit model. The probabilities of different crash severity levels are considered as dependent variables. The models are evaluated based on the signs and significance of their explanatory variables and their overall fit. Finally, the findings are interpreted based on the sign of the coefficients and their elasticities.

Task 5: Conclusions, Limitations and Recommendations

In Task 5, conclusions that could be useful to road users and traffic agencies are drawn. Furthermore, the limitations of this study are summarized and critically viewed. Finally, recommendations for future research are offered.

CHAPTER 2. LITERATURE REVIEW

This chapter presents a synthesis of the literature on the impact of weather on road safety. First, a general overview of the relationship between weather and safety is provided. The second section discusses the impact of the most common weather elements (i.e., precipitation, temperature, etc.) on safety. The third section of this chapter reviews different methodologies and data collection techniques that have been adopted by researchers in the past. Finally, the review concludes with a summary and critical discussion of the reviewed findings of the published research in the area.

2.1 Overview

It is widely accepted that weather has an influence on road safety since the weather conditions partly determine the road conditions and driver's behavior (SWOV, 2012). There is a large body of literature on the impact of weather conditions on road safety. Relative references go as back as in 1950's (Hermans et al., 2006). It is also noteworthy that recent studies have examined the interaction between weather and road safety within the recent climate change debate and have studied the corresponding countermeasures that should be established (Koetse & Rietveld, 2009; Andersson & Chapman, 2011)

Weather is an environmental factor that affects collision and casualty rates. Various weather conditions can be safety threats, such as reduced road friction, which leads to more slippery roads, limited visibility and other adversities that can make vehicle handling very difficult and dangerous. Such situations are more frequent during adverse weather conditions, such as heavy rainfalls and snowstorms.

Weather conditions can influence traffic as well. Empirical findings suggest that traffic volume is usually lower during inclement weather than during “normal” conditions. Also, the presence or expectation (based on weather forecasts) of unfavorable weather conditions may affect the mode choice and driver behavior, something which can consequently affect road safety. Furthermore, in cases of adverse weather conditions vehicle speeds are typically lower, while congestion may be also observed. For example, in cases of low visibility while people usually drive slower, but keep shorter space headways, which can increase the risk of crashes (SWOV, 2012). However, it has been shown that under such conditions crashes are more frequent but less severe (Khattak et al., 1998; Koetse & Rietveld, 2009).

The majority of papers in the literature examine the effect of adverse weather on the frequency and the severity of crashes in various types of facilities (Khattak et al., 1998; Knapp et al. 2000). Qiu & Nixon (2008) presented a systematic review and meta-analysis on the effect of adverse weather on road crashes. The major finding of that work was that most precipitation events were associated with a considerable increase in both crashes and crash rates, with snow having a greater effect than rain.

A number of recent studies have also addressed the issues of road maintenance over the winter period and mainly the appropriate activities during events of adverse weather (e.g. snowstorms). The objective of such studies was to assess the effectiveness of the (winter or other) maintenance policies examining crash frequency (Usman et al., 2010, 2012a). Furthermore, an ongoing study sponsored by the Iowa Department of Transportation aims to “identify locations of possible interest systematically with respect to winter weather-related safety performance based on crash history” (Hans et al., 2011).

Based on the above context, in the next section the author briefly presents the major research findings regarding the impact of the various weather elements, such as precipitation and temperature on road safety.

2.2 The Impact of Various Weather Elements on Road Safety

2.2.1 Precipitation

Precipitation is the most cited weather parameter in the road safety literature. Past research mainly considers rainfall and snowfall as precipitation. In almost all studies, precipitation was found to have a significant effect on the frequency and severity of crashes. For a systematic review and meta-analysis of studies associated with precipitation refer to Qiu & Nixon (2008).

Andrey (2010) noted that “empirical investigations indicate that collision rates usually increase during precipitation by 50-100% relative to normal seasonal conditions. As well, collision rate increases tend to be higher during snowfall than rainfall, although snow-related collisions tend to be associated with fewer fatalities”.

In addition, a consistent finding among a number of studies is that increases in crash risk are often most elevated during rainfalls following dry spells, during freezing rain and in cases of the first snowfalls of the season. Apart from the amount of precipitation, intensity (amount of precipitation over time) seems to have an effect (Andrey & Yagar, 2003; Eisenberg, 2004; Eisenberg & Warner, 2005; Keay & Simmonds, 2006).

The following two subsections address the major findings regarding the impact of rainfall and snowfall on safety.

2.2.1.1 Rainfall

Rainfall constitutes a driving hazard for a number of reasons. First, during rainfalls road users are likely to face problems with visibility. This phenomenon is also more intense during night, since the reflection of lights on a road with accumulated water makes the detection of the road and the objects near to it more difficult (Brodsky & Hakkert, 1988).

However, the most important factor related to rainfall is that of aquaplaning. In other words, the more the rainfall the less the friction of road surface is. This can lead to dynamic aquaplaning, which constitutes a very serious threat for safety. Indeed, when the road has been dry for a long period, even a drizzle can lead to viscous aquaplaning if drops of oil and dust, together with water, produce a thin liquid film on the road surface (SWOV, 2012). Nevertheless, the chance of aquaplaning is lower when rain gets heavier or during the last of a series of rain events. This is because the surface is swept clean after a significant amount of water has been fallen.

Considering the above, a number of studies have addressed these lagged effects of rainfall. For example, Eisenberg (2004) concluded that the amount of rain on a previous day affects the number of crashes on a given day. Whereas, crash risk is greater when there is a long dry spell between two events of rainfall. The latter is also supported by Keay & Simmonds (2006).

Despite the previous specific patterns regarding crash frequency and risk, the effect of rainfall on crash severity is more controversial. As mentioned previously, adverse weather may lead to more but less severe crashes. However, this finding is more obvious in cases of snow (which will be discussed in a following subsection) than in rainfall. For example, Qiu

& Nixon (2008) in a meta-analysis of published work argued that most precipitation events (including rainfalls) are associated with “considerable increased crash risk, a somewhat lesser increase in injury rates and minor increase in fatal risk”. On the other hand, Edwards (1998) found that during rainfall, crashes with minor injuries were relatively more frequent than crashes with serious injuries. According to the author, average speeds are lower during rainfall, and as a result the outcome of a crash is less serious.

Eisenberg (2004) argued that the risk of fatal crashes during rainfall decreases. In other words, rainfall has a negative association with the number of fatal crashes. However, this was the case when monthly data was analyzed. When the analysis was extended to daily data, the effect of rainfall in fatal crashes was positive.

2.2.1.2 Snowfall

The second type of precipitation which has been addressed in numerous studies is snowfall. A lot of emphasis regarding this weather parameter has been given especially in regions with heavy winters, such as the Nordic countries (Peltola & Kantonen, 1987; Schendersson, 1988), Canada (Andrey, 2010) and some states in the United States (Knapp et al., 2000).

Research has shown that driving during a snowstorm is difficult, since the visibility is worse, and also the accumulated snow may be frozen on the road, a fact that makes the road surface slippery and thus vehicle handling difficult. Furthermore, when wind is present during a snowstorm the situation may be even worse, since wind can cause blowing snow effect or impair the visibility of drivers (Usman et al., 2012a).

Regarding crash frequency during snowfall, a Finnish study (Peltola & Kantonen, 1987), led to the conclusion that around one third of crashes occurred when there was snow (or ice) on the road. Similar findings were revealed by a study in Sweden (Schendersson, 1988), which showed crash risk increased rapidly during light or moderate snowfall rates. In North America, Andreescu & Frost (1998) argued that even for days of light snowfall with less than 1 inch of snow, there was an increase in the mean number of crashes in comparison with dry days in Montreal, Canada, while Ahmed et al. (2012) suggested that crash frequency during snow season months (October-April) was more than double than during dry season months (May-September) in Colorado. Finally Eisenberg (2004) suggested that snow exhibited an inverted U-relationship with respect to crash risk. In other words, crash rates appear to peak around median level of snow and then decrease for heavier snowfall.

Turning to the effect of snowfall on crash severity Qiu & Nixon (2008) stated that “there is a debate on whether injury rate decreases during snowfall”. Indeed, Frindstorm et al. (1995) found negative coefficients when they examined the monthly number of days with snow for different crash types. While this finding is contradictory to intuitive sense, possible explanations might be the reduced exposure, the increased visibility at night and the adaptation of driving habits to such weather conditions (Hermans et al., 2006).

Following Eisenberg (2004), it seems that snow has a positive effect on non-fatal (injury) and property-damage-only (PDO) crashes, but negative effect on fatal crashes. Likewise, Andrey (2010) argued that the risk of minimal or minor injury is 89% higher during snowfall as compared to seasonal dry conditions, when he studied Canadian data for a period of nineteen years (1984-2002). The lower risk for a fatality during snowfall could be attributed to lower speeds, the more careful driving during snowstorms and winter maintenance activities which are taken by agencies.

2.2.2 Temperature

While there is a large number of published studies addressing the effect of precipitation (rainfall and snowfall) on road safety, there are only a few studies that have examined the effect of temperature on road crashes.

High temperatures may have a psychological effect on drivers (SWOV, 2012). According to a German study (DVR, 2000), emotions rise with temperature, people are more irritable to others, they get tired, they lose their concentration and their reaction time increases. All the aforementioned factors can affect road safety. For example, a French study by Laaidi & Laaidi (2002) as cited by SWOV (2012) found an increased number of crashes during heat waves. The authors argued that people possibly drive at other times of the day than they use to and that they sleep shorter or less deeply because of the high night-time temperatures.

In line with past research, Stern & Zehavi (1989), conducted a study in Israel, a country with Mediterranean climate and hot and long summer periods, and found that the possibility of a crash is higher for higher temperatures. Moreover, the results showed that single-vehicle crashes are more likely to occur. Similar findings have been reported in two studies in Greece (Yannis et al., 2008; Karlaftis & Yannis, 2010). However, it should be noted that the previous studies (in Israel and Greece) used data which was from the 1980’s and 1990’s, when a significant percentage of vehicles had no air conditioning systems, a fact which may bias the results.

Note that while interpreting the effect of temperature on road safety one should take into consideration the fact that the mobility (exposure) is higher (e.g. more recreational trips etc.), during periods with good weather and high temperatures (spring, summer). Furthermore, the frequency of crashes may increase due to the higher mobility of more vulnerable groups of road users, such as riders and pedestrians.

Finally, it should be noted that a review of studies that focused on winter weather crashes can reveal results contradictory to the above findings. For example, Andreescu & Frost (1998) found that temperature had a negative effect on winter road crashes in Canada. However, the authors noted that this relationship was inverted during summer. Similar results are reported in other Canadian studies (Karim et al., 2012; Usman et al., 2012a).

2.2.3 Wind

Research has shown that wind can have significant effect on road safety. Edwards (1994) argued that high winds can significantly increase crash risk. Similarly, Laaidi & Laaidi (2002) found a positive relationship between wind variation and the total number of crashes. Moreover, wind is a serious hazard for large vehicles. For example gusts of wind can push high vehicles such as vans, trucks and buses off course and, under extreme conditions can even cause them to roll over (SWOV, 2012). Baker & Reynolds (1992) examined the wind-induced crashes that occurred during a specific storm event on the 25th day of January 1990, in the UK. They found that the 47% of those crashes accounted for rollovers, the 19% for ran-off-road, while the 16% for collision with trees. What is more, the 66% of the observed 400 fatal or injury crashes was associated with heavy vehicles. Extensive research has also investigated the effect of wind direction (and especially of cross-winds) on vehicle movement and road safety (Baker, 1986; Coleman & Baker, 1990; Baker, 1993). Finally, wind can magnify adverse weather conditions, such as snowstorms (Usman et al., 2012a).

2.2.4 Visibility (Fog)

Visibility plays an important role in road safety. Reduced visibility can impair driver's vision and make driving difficult and dangerous. Perry & Symons (1991) argued that among all adverse conditions drivers experience more fear in fog, which affects visibility. Fog leads to a reduction in visibility because light is diffused by the fog droplets. When this happens, people generally drive slower, but also keep shorter space headways. This, in combination with the decreased field of vision, increases the risk of crashes (SWOV, 2012).

2.3 Data Collection Techniques and Methodological Approaches

This section of the chapter presents the main types of data and the corresponding sources, which are cited in the road safety literature and in particular, in studies which examine the effect of weather conditions on crashes. Finally, the section concludes with a brief review of the most common methodological approaches that researchers have utilized in an effort to investigate the impact of various factors on road safety.

2.3.1 Data Types, Sources and Level of Aggregation

Road safety is affected by various factors that can be classified into three major categories: infrastructure and environment, vehicle and user. For this reason, past studies have typically used the following types of data: crash data, roadway data (road geometry, classification, etc.), weather and exposure (traffic volume) data.

Crash data can be obtained by datasets that authorities process and maintain. For each reported crash, officials should fill a detailed report describing the circumstances under which the crash occurred (e.g. crash time and location, severity and type of crash, type and number of involved vehicles, driver attributes etc.).

Data associated with the roadway characteristics (e.g. geometry, functional classification, number of lanes etc.) are usually available in the archives of the corresponding agencies, which own the facility. State agencies collect and inventory such data.

Weather data can be obtained by multiple sources. In the majority of the studies, weather data was provided by weather stations which were installed on specific locations, such as airports (Automated Surface Observing Systems - ASOS or Automated Weather Observing Systems - AWOS) throughout a country. Furthermore, historical data can be found in the records of local weather offices or services. In the United States, possible sources of weather data are the National Weather Service (NWS), the National Oceanic and Atmospheric Administration (NOAA) and especially for the state of Iowa, the Iowa Environmental Mesonet, which is administered by the Department of Agronomy of the Iowa State University.

Weather data can be acquired from Road Weather Information Systems (RWIS) as well. RWIS are installed on specific locations of the roadway network and monitor air and surface conditions, such as temperature, precipitation, wind etc. A number of studies have employed data from RWIS (Usman et al., 2010, 2012a, b; Knapp et al., 2000). RWIS can provide relevant and real-time data, since these stations are installed on or near the roadway and have as a sole purpose the collection of real-time information about weather conditions (Ahmed et al., 2012).

Traffic data, such as volumes (AADT, VMT, etc.), speeds and vehicle classification is usually obtained from Automatic Traffic Recorders (ATR). ATRs, similarly to RWIS, are installed on specific locations of the network and are equipped with loop detectors, which can monitor the traffic conditions. Many studies have used data from ATRs (Knapp et al., 2000; Stout & Souleyrette, 2005). Apart from ATRs, Automatic Vehicle Identification (AVI) systems can also provide traffic information. Recent studies have employed traffic data from AVIs (Ahmed et al., 2012; Abdel-Aty et al., 2012).

However, real traffic data may not be always available. For this reason, proxies for exposure have been used in studies. For example, El-Basyouni & Kwon (2012) used the annual number of registered passenger vehicles as a proxy in order to investigate the impact of weather factors on collision frequency and severity in Edmonton Canada. Yannis et al. (2008) obtained traffic information collected from a toll station in the Athens region in a similar study. Finally, a number of studies (Brijs et al., 2008; Karlaftis & Yannis, 2010; Karim et al., 2012) have employed dummy variables associated with months, days of the week or holidays in an effort to capture seasonality and potential time-effects but also in an effort to control for exposure. In specific, Brijs et al. (2008) argued that “day-of-the-week dummies can be seen as some kind of proxy variable when real traffic exposure information is missing” after their investigation of the effect of weather on daily collision counts in the Netherlands.

The level of data aggregation is also a significant factor that may influence the final outcomes of a study. Following Hermans et al. (2006), three levels of aggregation can be discerned: the macro-level (one observation each year), the meso-level (one observation each day) and the micro-level (one observation each fraction of day). Although all levels of aggregation have their advantages and disadvantages, the higher the level of aggregation is the more information may be lost. For example, Hermans et al. (2006) argued that while monthly studies can capture the seasonal influence, they cannot capture the traffic volume patterns which are mainly daily or hourly related. As such, these studies may not be appropriate for measuring weather influences due to oversimplification. Usman et al. (2012a) have also confirmed the above statements when quantifying the safety effect of winter road maintenance.

However, it should be noted that there is no consensus regarding the analysis periods among all studies. In other words, there are studies which used monthly data (Fridstorm et al., 1995; Shankar et al. 1995), while others used daily data (Keay & Simmonds, 2006; Brijs et al., 2008) or hourly data (Hermans et al., 2006). Furthermore, Eisenberg (2004) analyzed both monthly and daily data and found that data format may lead to different results regarding the effect of precipitation on road crashes.

Finally, some studies have distinguished the analysis period into seasons, such as dry versus snowy or winter versus non-winter (Ahmed et al., 2012), while others, especially those focused on the effect of adverse weather, considered specific events (e.g. a storm) as one analysis period (Knapp et al., 2000; Usman et al., 2012a; Jung et al., 2012).

2.3.2 Methodological Approaches

This subsection presents the most common methodological approaches that have been used in road safety research and especially in studies of the effect of weather on road crashes. However, it should be noted that the various methodologies along with the studies where these have been employed are discussed in brief. The reader can refer to each study for more details about the corresponding methodology.

2.3.2.1 Count Data Models

Road crashes seem to occur randomly in time and space. Furthermore, crashes are assumed to be Poisson distributed (Lord et al., 2005; Hermans et al., 2006). Thus, Poisson regression models have been widely used in order to examine road crashes. However, Poisson distribution is a one-parameter distribution and assumes that the variance equals the mean. This property though is not always fulfilled, especially when crashes are studied. This is due to the overdispersion (i.e., the presence of extra variation) of data.

Another widely used methodological approach in road safety is the Negative Binomial regression model. This method assumes that crashes follow the Negative Binomial distribution, which is a generalization of the Poisson distribution, but it does not assume the mean to be equal to the variance. Thus, any overdispersion issues are addressed. In addition, other types of distribution, such as the Negative Multinomial Distribution or the Poisson Lognormal Distribution have been employed in some studies in order to handle overdispersed data (Miranda-Moreno et al., 1995; Caliendo et al., 2007). Furthermore, when examining road crashes it is very often that there are study observation periods (e.g. days, hours or storm events) with no crashes. In these cases, Zero-Inflated models have been used in order to address the preponderance of zeros (Shankar et al., 2004; Hermans et al., 2006).

2.3.2.2 Multivariate and Multilevel Models

A number of studies have used multivariate or multilevel models. For example, El-Basyouni and Kwon (2012) developed a Multivariate Poisson Lognormal (MVPL) model in order to assess time and weather effect on collision frequency, while Usman et al. (2012b) used three different multilevel structures: a Multilevel Multinomial Logit (MML), a Multilevel Nested Binary Logit (MBL) and a Multilevel Ordered Logit (MOL) in order to study conditional probabilities of collisions. Moreover, some studies (Ahmed et al., 2012; El-Basyouny & Kwon, 2012) utilized the Bayesian methodology and its properties in order to estimate the parameters of the models.

The use of different modeling frameworks and the comparison of their efficiency in a study is also very common in the literature. For example, Miranda-Moreno et al. (1995) compared the performance of three models: a traditional Negative Binomial, a Heterogeneous Negative Binomial and a Poisson Lognormal model and found that the Poisson Lognormal and the Heterogeneous Negative Binomial models had better fit than the traditional Negative Binomial. Similarly, Caliendo et al. (2007) compared three different distributions: the Poisson distribution, the Negative Binomial distribution and the Negative Multinomial distribution. The authors eventually argued that the Negative Multinomial distribution was the most appropriate one for modeling longitudinal collision data. On the other hand, Hermans et al. (2006) argued that a Negative Binomial Regression model was better than a Poisson Regression model, a Zero-Inflated Poisson model and a Zero-Inflated Negative Binomial model. Finally, Usman et al. (2012b) suggested that among the three different model structures (as mentioned earlier), the Multilevel Multinomial Logit (MML) models provided better predictions.

2.3.2.3 Time-series Analyses

In a large body of literature safety data is available on a time-series dimension, i.e., the variables examined are available over a (long) period of time. A time-series of count data is a sequence of non-negative integer observations over time (Karlaftis & Yannis, 2010). Several models for the analysis of time-series of count data is available, but the Integer Autoregressive Moving Average (INARMA) class of models has found wide application in many studies. For a list of studies along with the methodologies incorporated in the analysis of time-series data the reader can refer to Karlaftis & Yannis (2010).

Based on the above context, Brijs et al. (2008) elaborated Integer Value Autoregressive models in order to investigate the effect of weather on daily crash counts. The authors also performed correction for first order serial autocorrelation. Finally, they compared the results of the INAR models with other “traditional models”, such as Poisson Regression models and Negative Binomial Regression models and found that the INAR models were better than the Negative Binomial models, while the Poisson Regression models were the worst of all. Furthermore, the authors argued that the autocorrelation present in the data was significant.

A similar methodological approach was also adopted by Karlaftis and Yannis (2010), when studied daily crash data related to Athens, Greece. The authors also used lagged variables, which are variables with values related to a previous day of a given day, in order to address any lagged effect of the explanatory variables. This approach was also adopted by Eisenberg (2004).

2.3.2.4 Discrete (Ordered or Unordered) Data Models

Ordered probit modeling has been widely used for investigating the impact of various factors on the severity of crashes (for example, Khattak et al., 1998; Deng et al., 2012; Jung et al., 2012).This method is usually selected because it captures the ordinal nature of the severity. For example, police officers typically use the KABCO scale when reporting the severity of a crash at the scene of a crash. KABCO scale categorizes crashes based on the severity as follows: Fatal Injury crash (K), Incapacitating (Major) Injury crash (A), Non-incapacitating (Minor) Injury crash (B), Possible Injury and No Injury (C) or Property Damage Only (PDO) crash (O). On the other hand, the Abbreviate Injury Scale (AIS) is widely used in hospital records. AIS represents the “threat to life” associated with an injury. AIS scale uses numbers (0-6) to code crash severity, with number 6 to correspond to a fatal crash, 5 through 1 to an injury crash and 0 to a PDO crash (Sinha & Labi, 2007). In either case, analysis techniques for ordered data can be applied.

However, the use of ordered probability models may raise some issues. This is because ordered models do not have the flexibility to capture the effect of the explanatory variables on the interior category probabilities, for example A, B or C severity levels (based on the KABCO scale). For this reason, Washington et al. (2011) argued that one should be cautious in the selection of ordered models when examines crash severities. Unordered discrete data models, such as multinomial or nested logit models do not place any such restrictions and are preferred for modeling crash severity (for example, see Usman et al., 2012b).

Table 2.1 provides a summary of the data and methodological approaches that have been applied in previous studies listed in chronological order. which examine the effect of weather conditions on the frequency and severity of crashes.

2.4 Summary

This chapter first reviewed previous work on the effect of weather on the frequency and severity of crashes. It was found that weather elements, such as precipitation and temperature affect road safety in various ways. In the majority of the studies, weather elements associated with adverse conditions were found to have a negative effect mainly on the frequency of crashes, but they seemed to result in a less severe injury outcome (i.e., crashes are more frequent but less severe under adverse conditions). A number of studies also concluded that (high) temperature has positive effect on road crashes, especially during summer. Nevertheless, the effect of temperature on road crashes may be negative during winter periods. Other weather elements such as wind and visibility were found to have a negative effect on road safety, especially when they interact with precipitation (such as a snowfall event accompanied with heavy wind).

Next, this chapter reviewed the types of data and methodological approaches that have been used in past studies. In the majority of the studies, data was obtained from multiple sources, such as appropriate datasets (crash and roadway data) and records of specific stations (weather and traffic data). In cases where data was not available in the desired format, proxies were used. Some studies have also addressed the potential impacts that the level of data aggregation may have on the final outcomes of a study. Turning to methodology, different approaches have been used. These approaches range from simple regression models to multilevel frameworks or advanced techniques associated with time-series analysis (such as INAR models). On the other hand, the severity of crashes has been analyzed by ordered probability models which address the ordinal nature of crash severity.

The next chapter will present the different datasets (crash, roadway, weather, traffic) that used in the analysis along with their corresponding sources. Furthermore, the procedures of data processing and integration will be described.

Table 2. 1: Summary of Data and Methods used in Previous Studies

Study	Dependent		Weather Data		Exposure Data			Level of Data Aggregation						Method(s)					
	Frequency	Severity	Historical Data from Weather Stations	RWIS	Traffic Data from Transportation Agencies	ATR or AVI	Proxies	Hourly	Daily	Monthly	Seasonal	Event	Other	Poisson	Negative Binomial	Zero-Inflated Models	INAR	Multilevel Logit Models	Bayes Methodology
Shankar et al. (1995)	●		●							●					●				
Frindstorm et al. (1995)	●	●	●				●			●				●					
Knapp et al. (2000)	●			●		●						●		●					
Eisenberg (2004)	●		●				●		●	●					●				
Shankar et al. (2004)	●		●		●					●	●					●			
Keay and Simmonds (2006)	●		●		●				●					●					
Hermans et al. (2006)	●		●					●						●	●	●			
Caliendo et al. (2007)	●		●				●		●					●	●				
Yannis et al. (2008)	●		●				●		●	●				●					
Brijs et al. (2008)	●		●				●		●								●		
Karlaftis & Yannis (2010)	●		●				●		●								●		
Usman et al. (2010)	●			●		●						●			●	●			
Usman et al. (2012b)		●	●		●								●					●	
Ahmed et al. (2012)	●			●		●					●								●
El-Basyouny & Kwon (2012)	●	●	●				●		●					●					●

CHAPTER 3. DATA DESCRIPTION AND INTEGRATION

This chapter describes the data that were utilized in this thesis. In particular, four different types of data were incorporated in the analysis: Crash Data, Weather Data, Roadway and Traffic Data. The area of study and the main sources for each type of data will be described, followed by an extensive description of the integration process that was carried out in order to create the final comprehensive dataset.

3.1 Study Area

Two different corridors were selected as the study area of this thesis. The first corridor is located on Interstate 80 (I-80). The analysis section starts at the intersection of I-80 with the IA-117 and ends at the intersection of I-80 with IA-149. It has a total length of 64.61 miles (103.97 Km). The second corridor is located on US-34, starting at the intersection of US-34 with US-71 and ending at the intersection of US-34 with I-35. It has a total length of 65.32 miles (105.13 Km). Tables 3.1 and 3.2 show the basic roadway and traffic attributes of the two corridors as obtained from the Geographic Information Management System (GIMS) files of the Iowa Department of Transportation (DOT) for each year of the analysis period (see also Section 3.2.3). It should be mentioned that the tables present the weighted means and standard deviations (based on the length of each segment) of the values of all the road segments that constitute the aforementioned corridors.

Table 3. 1: Roadway and Traffic Variables of I-80 corridor

Attribute	Mean (Standard Deviation) or Percentage		
	2009	**2010**	**2011**
Number of Lanes 4/5/6	91.4/3.9/4.7	91.4/3.9/4.8	90.7/3.4/5.9
Median Width (ft.)	49.96 (0.65)	49.96 (0.65)	49.96 (0.65)
Road Width (shoulders not included) (ft.)	24.71 (2.84)	24.72 (2.86)	24.79 (2.98)
Right Shoulder Width (ft.)	9.98 (0.30)	9.98 (0.30)	9.98 (0.30)
Left Shoulder Width (ft.)	6.00 (0.11)	6.00 (0.11)	6.00 (0.11)
Speed Limit (MPH)	70.00 (0.00)	70.00 (0.00)	70.00 (0.00)
PSI rating	1.45 (1.44)	1.45 (1.44)	1.45 (1.44)
Slope	-0.68 (0.73)	-0.68 (0.73)	-0.68 (0.73)
IRI (in/mi)	79.84 (33.92)	79.84 (33.92)	79.85 (33.98)
AADT (veh/day)	26,442.63 (12,18.30)	27,764.46 (1,623.48)	27,669.90 (1,602.19)
Truck AADT (veh/day)	9,053.63 (181.86)	8,663.65 (287.15)	8,764.71 (167.20)

Table 3. 2: Roadway and Traffic Variables of US-34 corridor

Attribute	Mean (Standard Deviation) or Percentage		
	2009	**2010**	**2011**
Number of Lanes 2/3/4/5/6/7	76.8/15.0/6.7/1.1/0.3/0.2	76.8/15.0/6.7/1.1/0.3/0.2	76.8/15.0/6.7/1.1/0.3/0.2
Median Width (ft.)	N/A	N/A	N/A
Road Width (shoulders not included) (ft.)	27.15 (6.49)	27.15 (6.49)	27.15 (6.49)
Right Shoulder Width (ft.)	9.19 (1.88)	9.19 (1.88)	9.31 (1.79)
Left Shoulder Width (ft.)	9.03 (2.01)	9.03 (2.01)	9.15 (1.94)
Speed Limit (MPH)	54.41 (3.36)	54.41 (3.36)	54.41 (3.36)
PSI rating	2.55 (1.52)	2.55 (1.52)	2.55 (1.52)
Slope	-0.96 (1.12)	-0.96 (1.12)	-0.96 (1.12)
IRI (in/mi)	77.47 (33.34)	77.47 (33.34)	77.47 (33.34)
AADT (veh/day)	3,102.26 (1,377.74)	3,156.88 (1,401.00)	3,148.14 (1,385.99)
Truck AADT (veh/day)	518.27 (100.89)	527.37 (101.99)	525.67 (101.08)

As shown in Table 3.1, the I-80 corridor represents a divided facility with an average median width of 49.96 feet. In addition, it has four lanes along the 91.4% of its length. Based on this information this corridor is considered as a four-lane divided facility. The posted speed limit of the route is 70 mph, while its total Annual Average Daily Traffic (AADT) is greater than 26,000 veh/day, in all three years of the analysis period. Table 3.1 also provides additional information about the slope of the corridor and the pavement condition (Pavement Service Index –PSI and International Roughness Index –IRI).

On the other hand, the US-34 corridor has no median and two lanes along the 76.8% of its total length (shown in Table 3.2). Thus, it is considered as a two-lane undivided facility for the purpose of this thesis. Furthermore, it has an average posted speed limit of 54.44 mph. Finally, the average total AADT is greater than 3,000 veh/day.

Considering the above, the two facilities have different characteristics in terms of number of lanes, presence of median, speed limit and traffic conditions. One of the objectives of this thesis is to investigate whether these differences in the geometric and operational characteristics contribute to the severity outcome of crashes.

3.2 Data Description

3.2.1 Crash Data

The Iowa DOT collects information regarding the crashes that occur on all public roads of the State. According to the Iowa Accident Report Form (2010) as shown in Appendix A, all crashes that resulted in fatalities or injuries and the property damage only crashes with a value of more than $1,500 should be recorded. The records are stored in comprehensive datasets maintained by the agency. Those datasets can provide information for each crash such as, but not limited to: location/time, severity level and crash type, environmental and roadway conditions, driver and vehicle characteristics. Furthermore, crashes are geo-coded and the crash locations are saved in GIS format.

A number of 1,036 crashes occurred in the study area over the period 2009-2011. Specifically, 828 of those were located on I-80 corridor (average crash rate 0.429 crashes per million VMT) and 208 on the US-34 corridor (average crash rate 0.930 crashes per million VMT). The following tables show the descriptive statistics of various attributes of those crashes. This information was acquired by the aforementioned crash datasets and will be considered for further analysis.

Table 3. 3: Descriptive Statistics of the Attributes of Crashes on I-80

Attribute	**Percentage**
CRASH SEVERITY	
Fatal/Major Injury/Minor Injury/Possible-Unknown Injury/PDO	1.3/2.8/7.0/10.6/78.3
MONTH	
Jan/Feb/Mar/Apr/May/Jun/Jul/Aug/Sep/Oct/Nov/Dec	17.8/10.3/4.8/5.0/7.5/8.2/6.0/5.0/5.0/6.3/8.3/15.9
YEAR	
2009/2010/2011	35.2/37.2/27.5
DAY OF WEEK	
Sun/Mon/Tue/Wed/Thur/Fri/Sat	14.3/15.0/14.5/12.8/12.0/15.6/15.9
LIGHTING	
Daylight/Darkness/Dawn/Dusk	54.2/39.4/3.6/2.7
LOCATION OF FIRST HARMFUL EVENT	
On Roadway/Shoulder/Median/Roadside/Outside Traffic Way/Unknown/NR	60.3/7.7/8.5/5.4/0.5/4.5/13.2
RURAL OR URBAN ROAD	
Rural/Urban	92.8/7.2
NUMBER OF VEHICLES INVOLVED	
1/2/3/5/6	66.0/29.7/2.7/1.0/0.3/0.2
CONTRIBUTING CIRCUMSTANCES-ENVIRONMENT	
None Apparent/Weather Conditions/Physical Obstruction/Pedestrian Action/Glare/Animal/Previous Accident/Other/Unknown/NR	44.9/33.2/0.5/0.1/0.2/4.2/1.7/0.8/0.1/14.1

Table 3.3 (continued)

Attribute	**Percentage**
LIGHT	
Daylight/Dusk/Dawn/Dark(roadway lighted)/Dark(roadway unlighted)/Dark(unknown lighting conditions)/Unknown/NR	50.4/2.3/2.7/2.1/26.8/0.8/0.1/14.6
SURFACE CONDITIONS	
Dry/Wet/Ice/Snow/Slush/Water/Other/Unknown/NR	37.9/10.1/22.1/13.0/1.0/0.1/0.2/0.7/14.6
FIRST HARMFUL EVENT	
No Collision Event: Overturn-rollover/Jackknife/Other non-collision event	13.5/3.0/5.7
Collision with: vehicle in traffic/vehicle in or from other roadway/Parked Motor Vehicle/Animal/Other	28.0/1.1/1.1/19.0/2.7
Collision with fixed object: Bridge-Overpass/Underpass/Culvert/Guardrail/Concrete Barrier/Tree/Pole/Sign Post/Ditch/Curb-Island-Median/Tree/Pole/Other fixed object/Unknown	3.7/0.2/0.2/6.7/0.1/0.7/0.4/1.8/2.1/0.2
MANNER OF CRASH/COLLISION	
Non-collision/Head-on/Rear-end/Broadside/Sideswipe, same direction/Sideswipe, opposite direction/Unknown/ NR	55.4/1.4/11.5/0.6/15.7/1.9/1.1/12.3
MAJOR CAUSE	
Animal/Crossed Center Line	18.9/3.3
FTYROW: From Parked Position/Other	0.1/1.4
Driving Too Fast For Conditions/Exceeded Authorized Speed/Made Improper Turn/Followed Too Close/Operating the Vehicle in an Inappropriate Manner/Swerving-Evasive Action/Over-correcting, Over-steering/Downhill Roadway/Equipment Failure	20.4/0.2/0.1/2.7/1.0/14.7/1.6/0.1/1.3/
Ran off road: Right/Straight/Left	12.2/0.1/9.5
Lost Control/Inattentive or Distracted Driver/Vision Obstructed/Oversized Load, Vehicle/Other Improper Action/Other No Improper Action/Unknown	4.6/0.8/0.2/0.2/2.4/2.1/1.6
VISION OBSCUREMENT	
Not Obscured/Hillcrest/Moving Vehicles/Blinded by Sun or Headlights/Blowing Snow/Fog, smoke, dust/Other/Unknown/NR	77.4/0.1/0.2/0.1/3.3/0.1/1.3/2.1/15.2

Table 3. 4: Descriptive Statistics of the Attributes of Crashes on US-34

Attribute	**Percentages**
CRASH SEVERITY	
Fatal/Major Injury/Minor Injury/Possible-Unknown Injury/PDO	1.4/4.8/5.3/17.8/70.7
MONTH	
Jan/Feb/Mar/Apr/May/Jun/Jul/Aug/Sep/Oct/Nov/Dec	7.7/9.1/6.7/3.4/9.1/5.8/9.6/6.3/11.1/8.2/13.0/10.1

Table 3.4 (continued)

Attribute	**Percentages**
YEAR	
2009/2010/2011	30.3/38.0/31.7
DAY OF WEEK	
Sun/Mon/Tue/Wed/Thur/Fri/Sat	7.7/18.8/16.3/15.4/14.9/15.9/11.1
LIGHTING	
Daylight/Darkness/Dawn/Dusk	63.9/31.7/1.0/3.4
LOCATION OF FIRST HARMFUL EVENT	
On Roadway/Shoulder/Roadside/Outside Traffic Way/Unknown/NR	82.2/1.4/0.5/2.4/1.0/12.5
RURAL OR URBAN ROAD	
Rural/Urban	67.3/32.7
NUMBER OF VEHICLES INVOLVED	
1/2/3	39.4/57.7/2.9
CONTRIBUTING CIRCUMSTANCES-ENVIRONMENT	
None Apparent/Weather Conditions/Physical Obstruction/Glare/Animal/Other/NR	58.1/11.5/1.4/1/10.6/1/16.3
LIGHT	
Daylight/Dusk/Dawn/Dark(roadway lighted)/Dark(roadway unlighted)/Dark(unknown lighting conditions)/NR	58.2/3.8/1.9/5.3/14.4/1.4/14.9
SURFACE CONDITIONS	
Dry/Wet/Ice/Snow/Slush/Other/Unknown/NR	62.0/9.1/5.8/6.3/0.5/0.5/1/14.9
FIRST HARMFUL EVENT	
No Collision Event: Overturn-rollover/Other non-collision event	3.8/0.5
Collision with: vehicle in traffic/vehicle in or from other roadway/Animal/Other	52.9/6.3/28.4/0.5
Collision with fixed object: Bridge-Overpass/Ditch/Curb-Island-Median/Tree/Pole/Other fixed object/Unknown	0.5/3.8/0.5/1.0/1.0/0.5/0.5
MANNER OF CRASH/COLLISION	
Non-collision/Head-on/Rear-end/Angle, oncoming left turn/Broadside/Sideswipe, same direction/Sideswipe, opposite direction/Unknown/ NR	28.4/2.9/16.3/7.2/17.8/8.7/6.3/4.8/12.0
MAJOR CAUSE	
Animal/Ran Traffic Signal/Ran Stop Light/Crossed Center Line	28.4/0.5/4.8/9.1
FTYROW: From Stop Sign/Making Left Turn/From Driveway/Other	5.8/5.8/1.4/2.9
Travelling Wrong Way/Driving Too Fast For Conditions/Made Improper Turn/Followed Too Close/Operating the Vehicle in an Inappropriate Manner/Swerving-Evasive Action/Over-correcting, Over-steering/Equipment Failure	0.5/1.9/2.4/9.1/0.5/4.3/1.4/1.0
Ran off road: Right/Straight/Left	2.4/0.5/4.8
Lost Control/Inattentive or Distracted Driver/Vision Obstructed/Unknown	3.4/1.0/1.0/1.0
VISION OBSCUREMENT	
Not Obscured/Moving Vehicles/Frosted Windows-Windshield/Blowing Snow/Fog, smoke, dust/Unknown/NR	77.9/1.4/0.5/1.0/0.5/2.9/15.9

3.2.2 Weather Data

The weather information that was used in this thesis was acquired from the Road Weather Information Systems (RWIS), which are installed along the two study corridors. A total number of 62 RWIS are installed on specific locations of the roadway network in Iowa (mainly on primary roads).

As mentioned in Chapter 2, RWIS monitor air and surface conditions such as, temperature, precipitation, wind etc. Moreover, RWIS have the ability to automatically update their records each time a change in the weather conditions occurs (e.g. a drop in temperature, the beginning of a precipitation event etc.). This ability combined with the proximity to the roadway can provide real time information about the current conditions of the road. All the recorded information is collected by the Iowa DOT and is stored in comprehensive datasets.

For the purpose of this thesis, the author utilized information coming from six RWIS (Figure 3.1): three along the I-80 (located in Colfax, Grinnell and Williamsburg) and three along the US-34 (located in Red Oak, Creston and Osceola) over the analysis period 2009-2011. The information that was extracted and included in the analysis as variables is the following: Temperature, Dew Point Temperature, Relative Humidity (RH), Precipitation Type, Precipitation Rate (from which the Accumulated Amount of Precipitation was derived), Average Wind Speed, Wind Speed Gust, Average Wind Direction and Average Wind Speed Gust. Tables 3.5 and 3.6 provide the descriptive statistics of the weather variables over the analysis period related to corridors I-80 and US-34, respectively.

Figure 3. 1: Locations of RWIS in the Area of Study

Table 3. 5: Descriptive Statistics of the Weather Variables related to Crashes on I-80

Weather Parameter	**Mean (Standard Deviation) or Percentage**	**Min**	**Max**
Temperature (C)	5.27 (13.14)	-27.58	36.38
Dew Point Temperature (C)	1.25 (11.87)	-30.52	30.70
Relative Humidity (%)	77.71 (17.37)	15.50	99.09
Type of Precipitation			
No Precipitation/Rain/Snow	61.9/11.8/26.3	N/A	N/A
Total Accumulated Amount of Precipitation	2.43 (10.28)	0.00	128.79
Wind Speed (Km/hr)	15.42 (9.57)	0.08	47.83
Wind Speed Gust(Km/hr)	22.11 (13.18)	1.00	64.08
Wind Direction			
N/NE/E/SE/S/SW/W/NW	5.9/11.4/12.3/12.2/13.5/13.7/12.2/18.9	N/A	N/A
Wind Gust Direction			
N/NE/E/SE/S/SW/W/NW	5.4/12.5/11.7/12.0/12.3/12.0/13.4/20.7	N/A	N/A

N/A: not applicable

Table 3. 6: Descriptive Statistics of the Weather Parameters related to Crashes on US-34

Weather Variable	**Mean (Standard Deviation) or Percentage**	**Min**	**Max**
Temperature (C)	10.18 (12.56)	-20.01	33.11
Dew Point Temperature (C)	4.13 (10.88)	-26.41	25.00
Relative Humidity (%)	69.69 (20.49)	28.67	99.08
Type of Precipitation			
No Precipitation/Rain/Snow	82.6/6.4/11.0	N/A	N/A
Total Accumulated Amount of Precipitation	0.80 (4.79)	0.00	48.36
Wind Speed (Km/hr)	14.81 (10.81)	0.00	71.00
Wind Speed Gust(Km/hr)	23.22(15.49)	0.00	122.89
Wind Direction			
N/NE/E/SE/S/SW/W/NW	6.1/12.7/13.3/16.0/21.0/9.4/7.7/13.8	N/A	N/A
Wind Gust Direction			
N/NE/E/SE/S/SW/W/NW	3.9/11.0/16.0/13.3/19.3/11.6/7.2/16.0	N/A	N/A

N/A: not applicable

3.2.3 Roadway Data

As mentioned in section 3.1, information about the roadway and traffic conditions was acquired from the GIMS files of the Iowa DOT. GIMS files provide detailed roadway information about all Iowa roads, such as segment ID, road classification, geometric characteristics (such as median type and width), speed limit, AADT and so forth. In the GIMS files all the Iowa roads are divided into segments with similar characteristics. A new segment starts wherever there is a change in any of the road conditions or geometry (e.g. AADT, speed limit, median width etc.). For instance, a specific corridor (e.g. I-80) is divided into a finite number of segments which share common characteristics and when a change occurs (e.g. a change in the AADT after an interchange) then a new segment (with a new ID) begins. A detailed description of the information that is included in the GIMS file can be found in the *"Base Record Road and Structure Data Manual"* provided by the Office of Transportation Data of the Iowa DOT (Iowa DOT, 2001).

The Iowa DOT updates the GIMS files every year in order to keep track of any changes occur in the roadway system. Furthermore, GIMS files are in GIS format (shapefiles) and thus they can be used and processed in any software with a GIS interface (e.g. ArcMap, TransCAD, etc.). For the purpose of this thesis, the author utilized information about the traffic conditions (AADT and Truck AADT) recorded in the GIMS files for the period 2009-2011. Figure 3.2 shows a map with a visual representation of the variations in the AADT along the two corridors of study as of year 2011.

Figure 3. 2: Variations in the AADT along the two study corridors (as of 2011)

3.2.4 Traffic Data

Automatic Traffic Recorders (ATRs) collect information related to traffic conditions, such as speeds, volumes (AADT, Vehicle Miles Travelled (VMT), etc.) and vehicle classification. These recorders are installed on specific locations of the network and are equipped with loop detectors which can monitor traffic conditions. The Iowa DOT has installed a number of more than 100 ATRs on the roadways of the State.

The Office of Transportation Data of the Iowa DOT in cooperation with the Federal Highway Administration (FHWA) releases an annual ATR report at the beginning of each year. The report contains traffic information that has been collected from ATRs during the previous year. Apart from the numerical information (i.e., AADT, etc.) the report includes a number of graphs which show the hourly, daily and monthly variations, as percentages of the AADT. These variations are given for roads of different classification (interstate, primary and secondary) and environment (rural or municipal). A typical form of these graphs is presented in Figure 3.3. It is noteworthy that the traffic is significantly higher during the summer months (percentages of AADT larger than 100%).

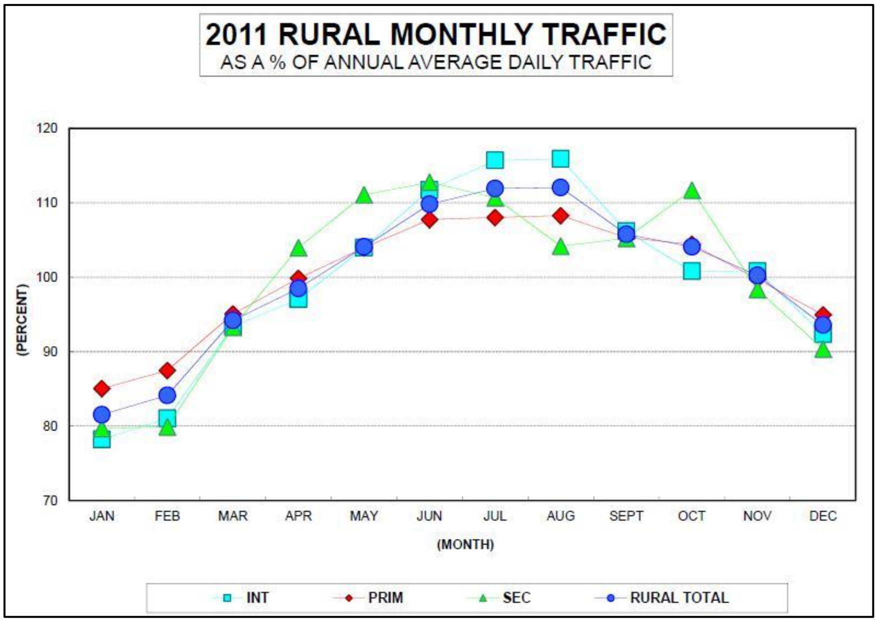


Figure 3. 3: Typical Form of a Graph with the Variations of AADT

In this thesis the author utilized the information that was provided by the ATR reports and referred to the analysis period, in order to capture any seasonal variations of the AADT. Specifically, the author considered the monthly variations (percentages of the AADT) for the Interstate (I-80) and Primary (US-34) roads and adjusted accordingly the values of the AADT and Truck AADT that were acquired from the GIMS files. For more information on this process, the reader may refer to the section 3.3.4.

3.3 Data Integration

According to FHWA, data integration is the method by which multiple datasets coming from various sources can be combined or linked together and can be applied to solve problems (FHWA, 2010). In this section, the author is going to present the process of data integration in order to create the final dataset with all the variables that were used in the analysis.

For the purpose of this thesis a number of different datasets coming from various sources (as described previously) had to be integrated in order all the required information to be put together and constitute a unique dataset which was the final input of the analysis. To do so, couple of different software (ArcMap 2010, MS Excel 2010, MS Access 2010) was used during the various stages of the process. Figures 3.4 and 3.5 present the detailed flowchart of the stages of the process. The color in each box of the chart indicates the software that was used in each stage.

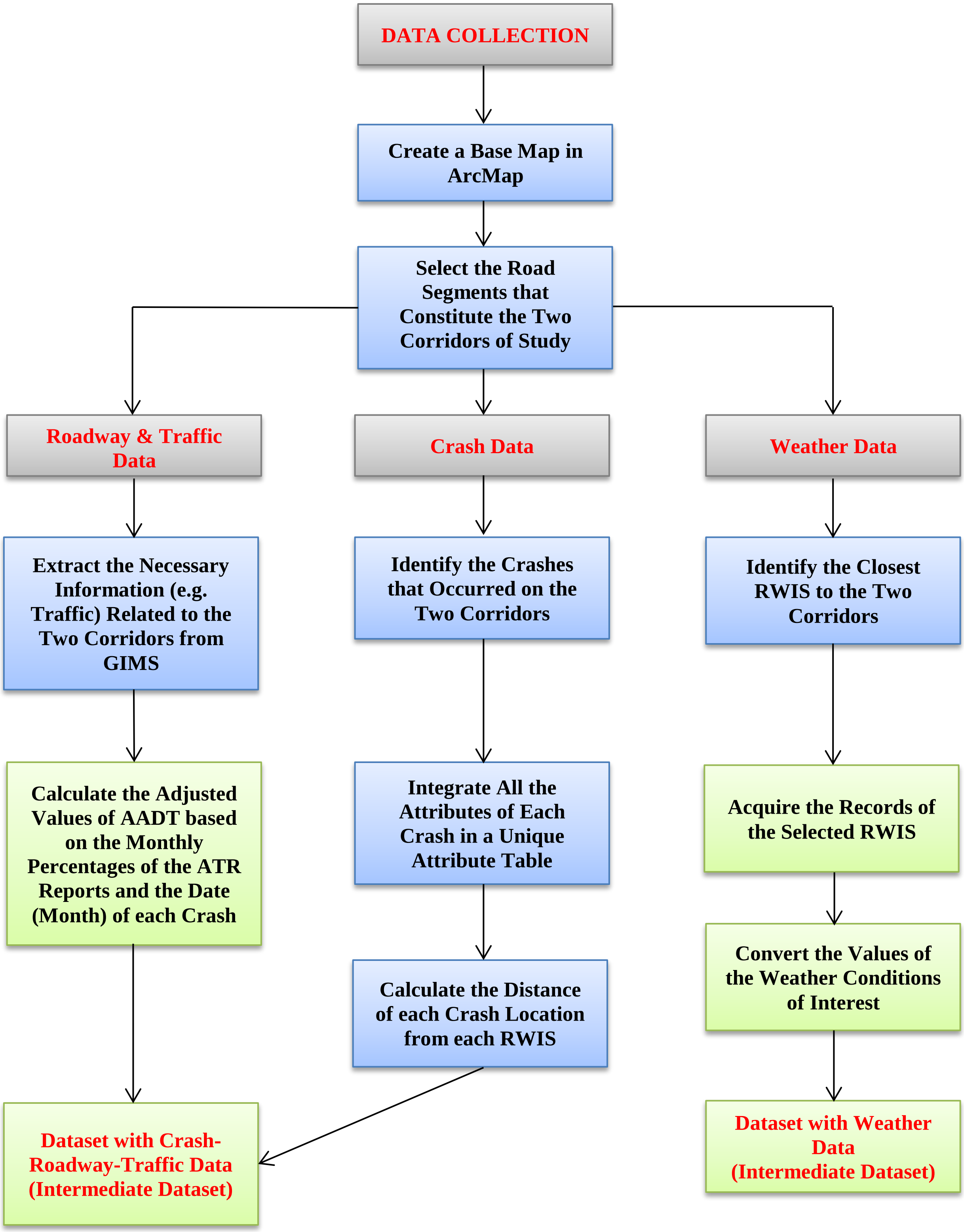


Figure 3. 4: Process of Data Integration (Initial Datasets to Intermediate Datasets)

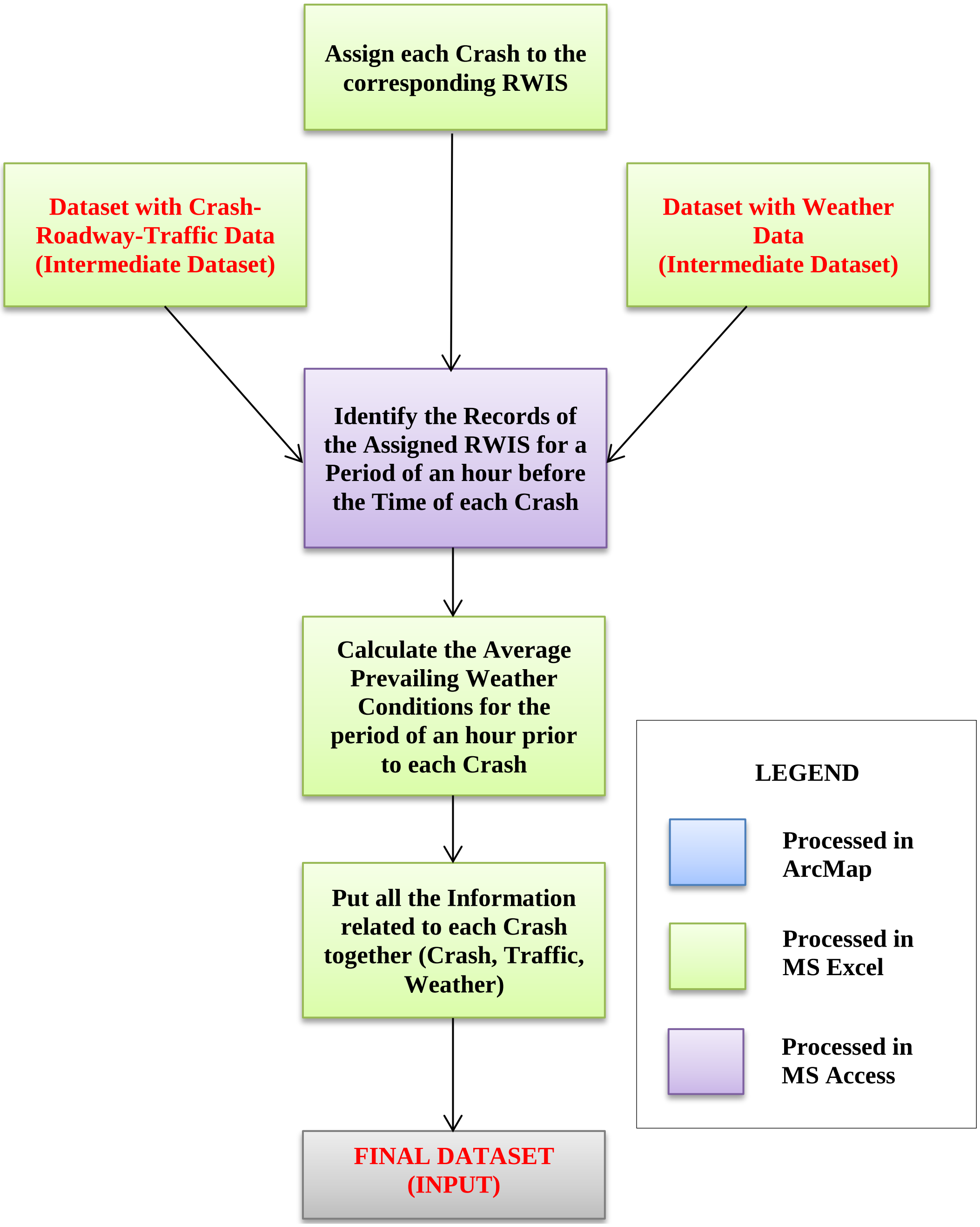


Figure 3. 5: Process of Data Integration (Intermediate Datasets to Final Dataset)

3.3.1 Creating a Base Map

The base map with the two corridors of study was created in ArcMap 2010. To do so, the author first imported a basemap layer from Bing Maps and focused on the state of Iowa. Then, a layer of the GIMS files which contains the whole roadway network of Iowa with the corresponding information for each road segment (as mentioned previously) was created. Furthermore, the layer with the RWIS locations was imported. The next step was to select the road segments (as coded in the GIMS files) for the two corridors of study (I-80 and US-34). In this task, the tool of the "*Interactive Selection Method"* was used. This process created two new layers: one for I-80 and one for US-34. It is noteworthy that this was done for each year of the analysis period (2009-2011). Thus, three base maps were created. Figure 3.6 presents as screenshot of an ArcMap file (base map).

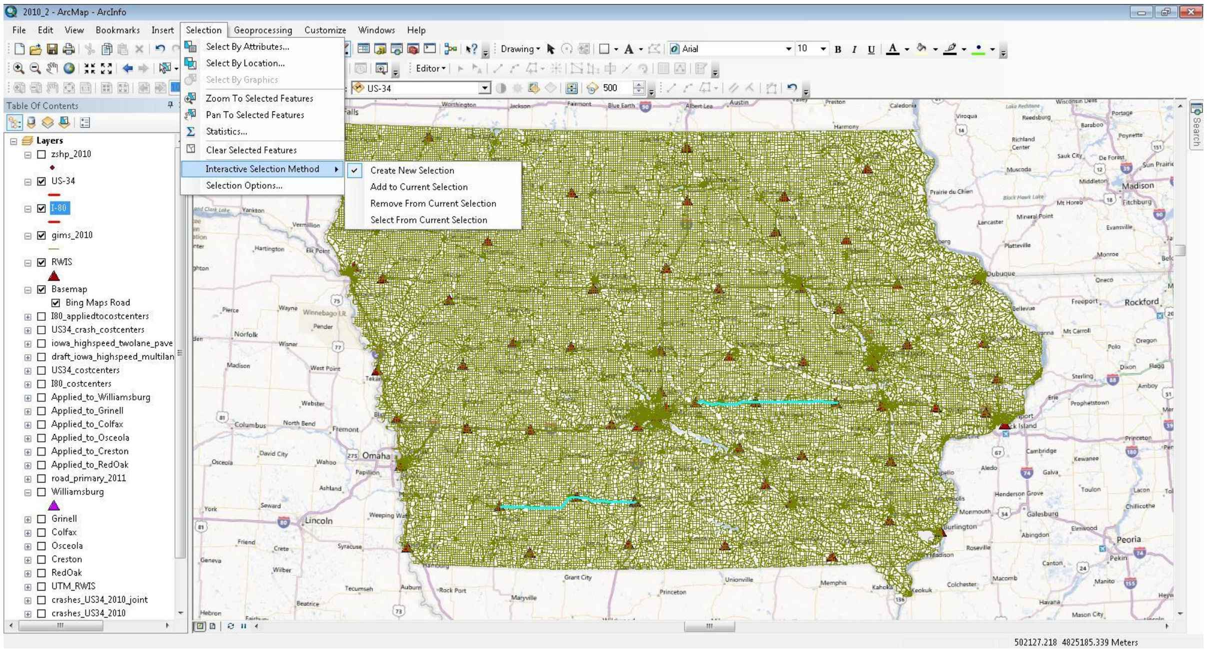
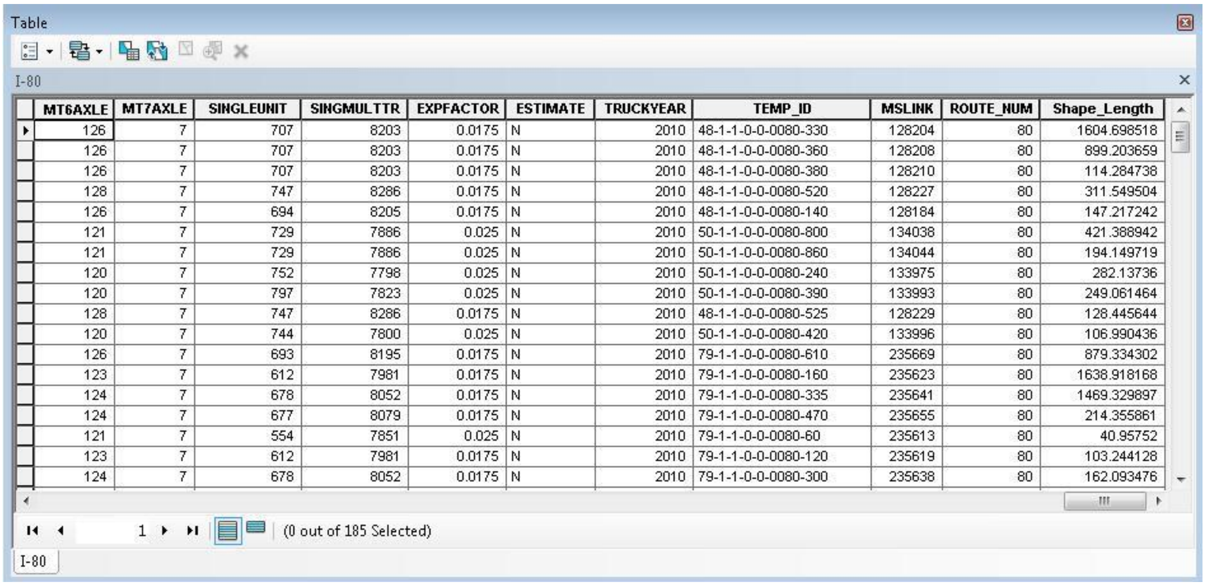


Figure 3. 6: Base Map and Selection of the Two Corridors

3.3.2 Roadway and Traffic Data

All the GIMS information related to the two corridors of study (such as geometric characteristics, AADT etc.) is saved in the attribute tables of the corresponding layers. A screenshot of a typical attribute table of a GIMS file is shown in Figure 3.7. An attribute table can be also exported in a dbf format file and then opened and easily processed in MS Excel.



MT6AXLE	MT7AXLE	SINGLEUNIT	SINGMULTTR	EXPFACTOR	ESTIMATE	TRUCKYEAR	TEMP_ID	MSLINK	ROUTE_NUM	Shape_Length
126	7	707	8203	0.0175	N	2010	48-1-1-0-0-0080-330	128204	80	1604.698518
126	7	707	8203	0.0175	N	2010	48-1-1-0-0-0080-360	128208	80	899.203659
126	7	707	8203	0.0175	N	2010	48-1-1-0-0-0080-380	128210	80	114.284738
128	7	747	8286	0.0175	N	2010	48-1-1-0-0-0080-520	128227	80	311.549504
126	7	694	8205	0.0175	N	2010	48-1-1-0-0-0080-140	128184	80	147.217242
121	7	729	7886	0.025	N	2010	50-1-1-0-0-0080-800	134038	80	421.388942
121	7	729	7886	0.025	N	2010	50-1-1-0-0-0080-860	134044	80	194.149719
120	7	752	7798	0.025	N	2010	50-1-1-0-0-0080-240	133975	80	282.13736
120	7	797	7823	0.025	N	2010	50-1-1-0-0-0080-390	133993	80	249.061464
128	7	747	8286	0.0175	N	2010	48-1-1-0-0-0080-525	128229	80	128.445644
120	7	744	7800	0.025	N	2010	50-1-1-0-0-0080-420	133996	80	106.990436
126	7	693	8195	0.0175	N	2010	79-1-1-0-0-0080-610	235669	80	879.334302
123	7	612	7981	0.0175	N	2010	79-1-1-0-0-0080-160	235623	80	1638.918168
124	7	678	8052	0.0175	N	2010	79-1-1-0-0-0080-335	235641	80	1469.329897
124	7	677	8079	0.0175	N	2010	79-1-1-0-0-0080-470	235655	80	214.355861
121	7	554	7851	0.025	N	2010	79-1-1-0-0-0080-60	235613	80	40.95752
123	7	612	7981	0.0175	N	2010	79-1-1-0-0-0080-120	235619	80	103.244128
124	7	678	8052	0.0175	N	2010	79-1-1-0-0-0080-300	235638	80	162.093476

1 (0 out of 185 Selected)

Figure 3. 7: A Typical Attribute Table of a GIMS file

3.3.3 Crash Data Integration

The first task of the crash data integration was to identify those crashes that occurred along the two corridors during the analysis period. As mentioned previously, the location (along with the other attributes) of each recorded crash is saved in GIS format. Thus, the crash points can be projected on each year's base map by importing the corresponding shapefile with the geographic coordinates of crash locations. Figure 3.8 presents a map which shows the crash points of the year of 2010.

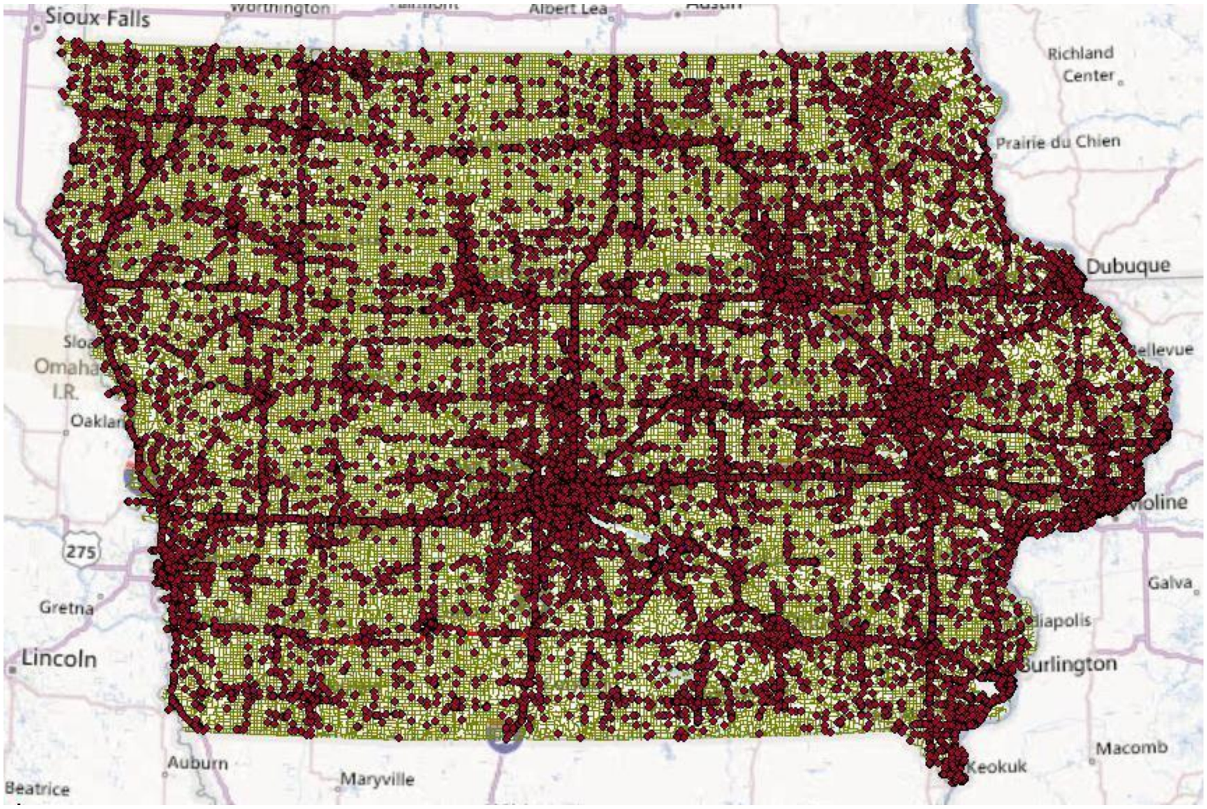
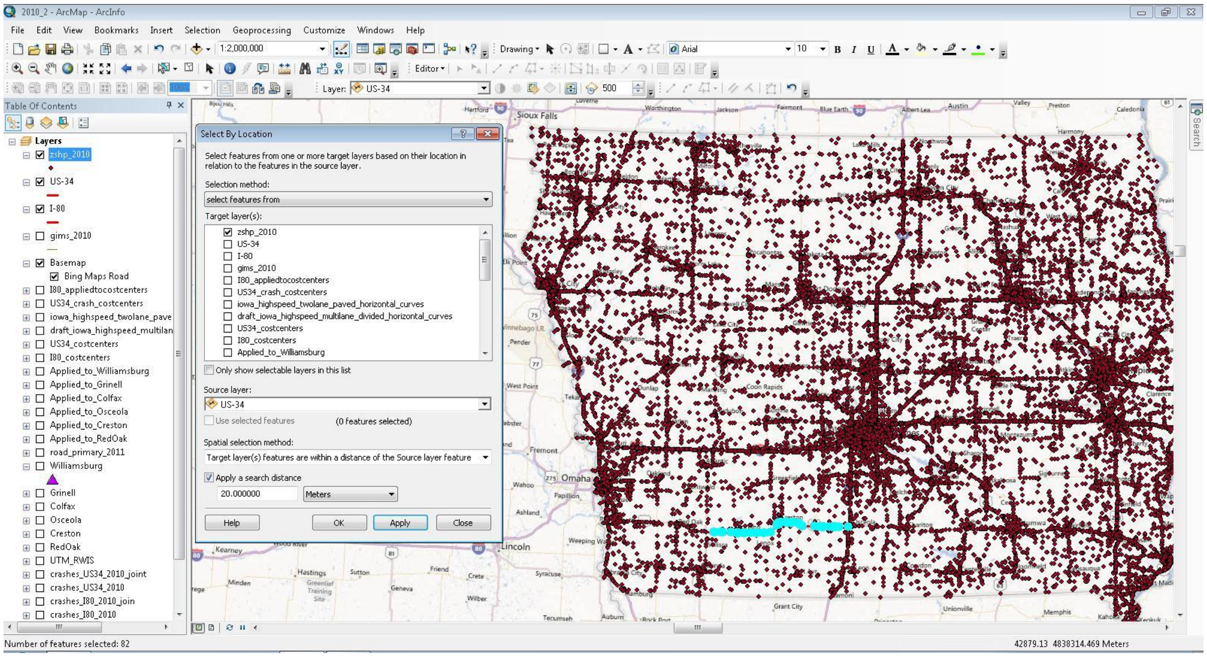


Figure 3. 8: Crash Points of the Year 2010

The previous map shows all crashes that happened in Iowa during 2010. From those crashes, the crashes located along the two corridors of interest were selected. To do this, the *"Select by Location"* method in ArcMap was used. Specifically, the author selected the features (crashes) from the layer zshp_2010 (layer with crash locations) with a source layer, the layer of the corridor I-80 (or US-34). As a spatial selection method, the method which selects features that are within a specific distance (offset distance) of the source layer feature was used. In this particular case, the offset distance was set to be at 20 meters. The author did so, in order to capture any cases of crashes that occurred on the corridor of interest, but their exact location was not recorded correctly due to several reasons: 1) vehicle run off the road after crash; 2) GPS device accuracy; 3) changes in the road systems (or cartography); and 4) cloud cover (Gao, 2012). Figure 3.9 presents a screenshot of the method in ArcGIS.



Select By Location

Select features from one or more target layers based on their location in relation to the features in the source layer.

Selection method:
select features from

Target layer(s):
☑ zshp_2010
☐ US-34
☐ I-80
☐ gims_2010
☐ I80_appliedtocostcenters
☐ US34_crash_costcenters
☐ iowa_highspeed_twolane_paved_horizontal_curves
☐ draft_iowa_highspeed_multilane_divided_horizontal_curves
☐ US34_costcenters
☐ I80_costcenters
☐ Applied_to_Williamsburg

☐ Only show selectable layers in this list

Source layer:
US-34

☐ Use selected features (0 features selected)

Spatial selection method:
Target layer(s) features are within a distance of the Source layer feature

☑ Apply a search distance
20.000000 Meters

Help OK Apply Close

Number of features selected: 82

42879.13 4838314.469 Meters

Figure 3. 9: Selection of Crashes with the *"Select by Location"* Method in ArcGIS

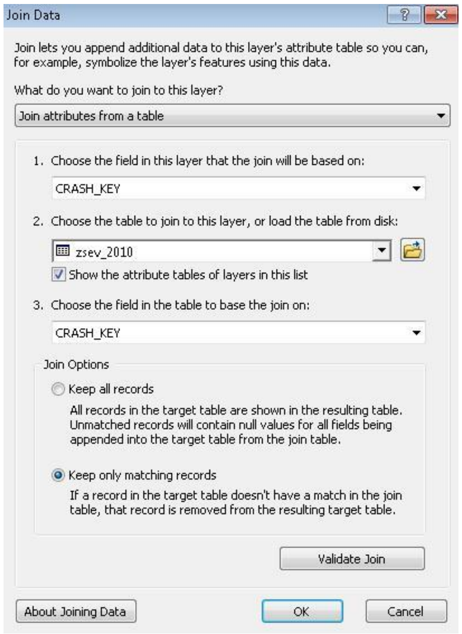
Once the crashes for each corridor were selected and the corresponding layers were created (one for I-80 and one for US-34), the next step was to integrate all the attributes for each crash in a unique table. Furthermore, the information for each crash should be joined with the information (from the GIMS file) of the corresponding road segment on which the crash was assigned. The first task was done by using the *"Join by Attributes"* method of ArcMap. For the second task, the *"Spatial Join"* method was used.

All the parameters of each recorded crash are saved in separate tables (files) according to the kind of information that they provide. For the purpose of this thesis, the author utilized information from tables that provide the following information:

- Crash Point Parameters (table zshp)
- Location/Time Crash Parameters (table zltp)
- Severity Level Crash Parameters (table zsev)
- Environmental Crash Parameters (table zenv)
- Roadway Crash Parameters 1 (table zrda)
- Roadway Crash Parameters 2 (table zrdb)
- Crash Type Parameters 1 (table zcta)
- Driver Crash Parameters (table zdrv)
- Vehicle Crash Parameters (table zveh)

The parameter that all the aforementioned tables share is the *"CRASH KEY"*. The *"CRASH KEY"* is a "unique identifier" for each crash. Thus, all the information from each separate table can be joined together based on the *"CRASH KEY"*.

The task of joining the separate tables and creating a comprehensive dataset with the crash attributes of interest of the selected crashes (for each year of the analysis period) was done in ArcMap. Specifically, the necessary tables into the already created ArcMap file were imported and the *"Join by Attributes"* method was used, as mentioned previously. Of course, the join was based on the *"CRASH KEY"*, which is the common field in each table. Figure 3.10 presents the "pop-up" window of this method in ArcMap.



Join lets you append additional data to this layer's attribute table so you can, for example, symbolize the layer's features using this data.

What do you want to join to this layer?

Join attributes from a table

1. Choose the field in this layer that the join will be based on:

CRASH_KEY

2. Choose the table to join to this layer, or load the table from disk:

zsev_2010

☑ Show the attribute tables of layers in this list

3. Choose the field in the table to base the join on:

CRASH_KEY

Join Options

○ Keep all records

All records in the target table are shown in the resulting table. Unmatched records will contain null values for all fields being appended into the target table from the join table.

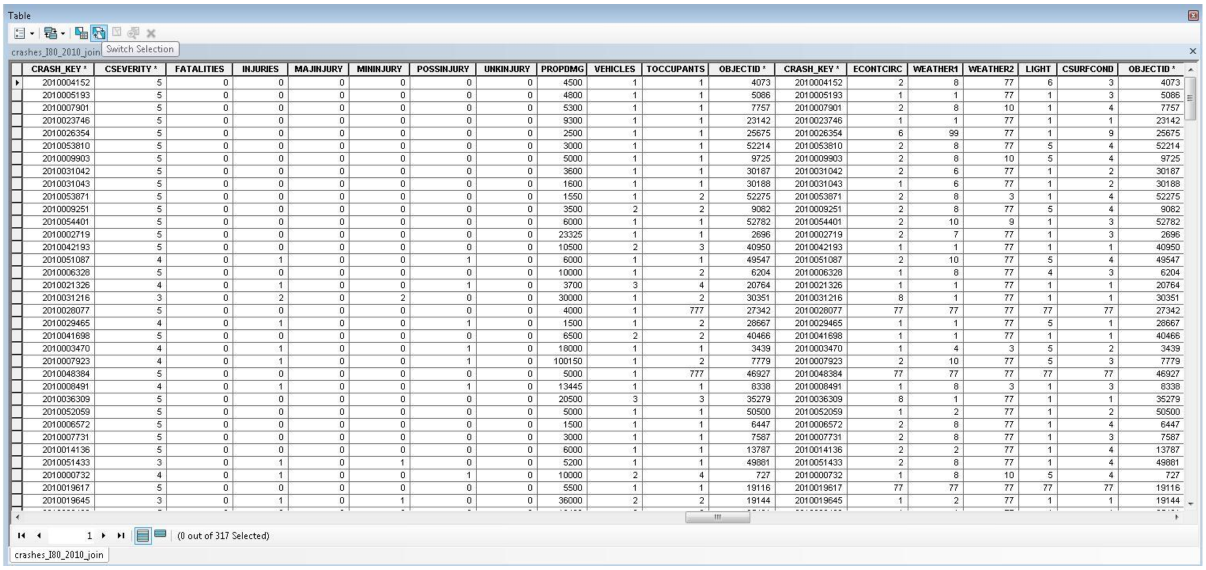
● Keep only matching records

If a record in the target table doesn't have a match in the join table, that record is removed from the resulting target table.

Figure 3. 10: *"Join by Attributes"* Method Window in ArcMap

It should be noted that all the tables were joined to the layers with the selected crashes that had been created in a previous step. Thus, the attribute tables of the crash layers had all the information of interest for each selected crash. A screenshot of an attribute table of a crash layer is shown in Figure 3.11.

The next step was to integrate the crash attributes with the information of the corresponding road segments (e.g. AADT) to which the selected crashes were assigned. This was done by using the *"Spatial Join"* Method of ArcMap. This method joins data from two layers based on spatial location. Moreover, in the resulting attribute table the distance between the joined features is given. In this case, lines (road segments) were joined to points (crashes). Thus, at the end of the process the distance of each crash from each road segment was calculated. That was actually the offset distance which was mentioned previously. Figure 3.12 presents a screenshot of the method's window in ArcMap.



crashes_I80_2010_join

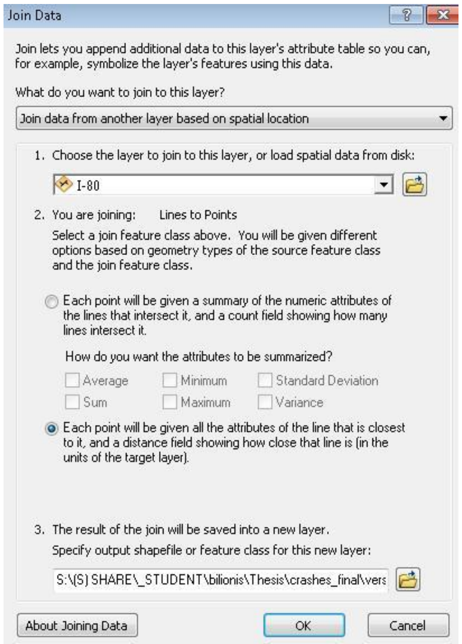
CRASH_KEY *	CSEVERITY *	FATALITIES	INJURIES	MAJINJURY	MININJURY	POSSINJURY	UNKINJURY	PROPDMG	VEHICLES	TOCCUPANTS	OBJECTID *	CRASH_KEY *	ECONTCIRC	WEATHER1	WEATHER2	LIGHT	CSURFCOND	OBJECTID *
2010004152	5	0	0	0	0	0	0	4500	1	1	4073	2010004152	2	8	77	6	3	4073
2010005193	5	0	0	0	0	0	0	4800	1	1	5086	2010005193	1	1	77	1	3	5086
2010007901	5	0	0	0	0	0	0	5300	1	1	7757	2010007901	2	8	10	1	4	7757
2010023746	5	0	0	0	0	0	0	9300	1	1	23142	2010023746	1	1	77	1	1	23142
2010026354	5	0	0	0	0	0	0	2500	1	1	25675	2010026354	6	99	77	1	9	25675
2010053810	5	0	0	0	0	0	0	3000	1	1	52214	2010053810	2	8	77	5	4	52214
2010009903	5	0	0	0	0	0	0	5000	1	1	9725	2010009903	2	8	10	5	4	9725
2010031042	5	0	0	0	0	0	0	3600	1	1	30187	2010031042	2	6	77	1	2	30187
2010031043	5	0	0	0	0	0	0	1600	1	1	30188	2010031043	1	6	77	1	2	30188
2010053871	5	0	0	0	0	0	0	1550	1	2	52275	2010053871	2	8	3	1	4	52275
2010009251	5	0	0	0	0	0	0	3500	2	2	9082	2010009251	2	8	77	5	4	9082
2010054401	5	0	0	0	0	0	0	6000	1	1	52782	2010054401	2	10	9	1	3	52782
2010002719	5	0	0	0	0	0	0	23325	1	1	2696	2010002719	2	7	77	1	3	2696
2010042193	5	0	0	0	0	0	0	10500	2	3	40950	2010042193	1	1	77	1	1	40950
2010051087	4	0	1	0	0	1	0	6000	1	1	49547	2010051087	2	10	77	5	4	49547
2010006328	5	0	0	0	0	0	0	10000	1	2	6204	2010006328	1	8	77	4	3	6204
2010021326	4	0	1	0	0	1	0	3700	3	4	20764	2010021326	1	1	77	1	1	20764
2010031216	3	0	2	0	2	0	0	30000	1	2	30351	2010031216	8	1	77	1	1	30351
2010028077	5	0	0	0	0	0	0	4000	1	777	27342	2010028077	77	77	77	77	77	27342
2010029465	4	0	1	0	0	1	0	1500	1	2	28667	2010029465	1	1	77	5	1	28667
2010041698	5	0	0	0	0	0	0	6500	2	2	40466	2010041698	1	1	77	1	1	40466
2010003470	4	0	1	0	0	1	0	18000	1	1	3439	2010003470	1	4	3	5	2	3439
2010007923	4	0	1	0	0	1	0	100150	1	2	7779	2010007923	2	10	77	5	3	7779
2010048384	5	0	0	0	0	0	0	5000	1	777	46927	2010048384	77	77	77	77	77	46927
2010008491	4	0	1	0	0	1	0	13445	1	1	8338	2010008491	1	8	3	1	3	8338
2010036309	5	0	0	0	0	0	0	20500	3	3	35279	2010036309	8	1	77	1	1	35279
2010052059	5	0	0	0	0	0	0	5000	1	1	50500	2010052059	1	2	77	1	2	50500
2010006572	5	0	0	0	0	0	0	1500	1	1	6447	2010006572	2	8	77	1	4	6447
2010007731	5	0	0	0	0	0	0	3000	1	1	7587	2010007731	2	8	77	1	3	7587
2010014136	5	0	0	0	0	0	0	6000	1	1	13787	2010014136	2	2	77	1	4	13787
2010051433	3	0	1	0	1	0	0	5200	1	1	49881	2010051433	2	8	77	1	4	49881
2010000732	4	0	1	0	0	1	0	10000	2	4	727	2010000732	1	8	10	5	4	727
2010019617	5	0	0	0	0	0	0	5500	1	1	19116	2010019617	77	77	77	77	77	19116
2010019645	3	0	1	0	1	0	0	36000	2	2	19144	2010019645	1	2	77	1	1	19144

1 (0 out of 317 Selected)

crashes_I80_2010_join

Figure 3. 11: Sample of an Attribute Table of a Crash Layer

.



Join lets you append additional data to this layer's attribute table so you can, for example, symbolize the layer's features using this data.

What do you want to join to this layer?

Join data from another layer based on spatial location

1. Choose the layer to join to this layer, or load spatial data from disk:

I-80

2. You are joining: Lines to Points

Select a join feature class above. You will be given different options based on geometry types of the source feature class and the join feature class.

○ Each point will be given a summary of the numeric attributes of the lines that intersect it, and a count field showing how many lines intersect it.

How do you want the attributes to be summarized?

☐ Average ☐ Minimum ☐ Standard Deviation

☐ Sum ☐ Maximum ☐ Variance

● Each point will be given all the attributes of the line that is closest to it, and a distance field showing how close that line is (in the units of the target layer).

3. The result of the join will be saved into a new layer.

Specify output shapefile or feature class for this new layer:

S:\(S) SHARE_STUDENT\bilionis\Thesis\crashes_final\vers

Figure 3. 12: *"Spatial Join"* Method Window in ArcMap

The author also noted that the distance between the features of the joined layers was provided and used the *"Spatial Join"* method in order to calculate the distance of each crash from each RWIS (by joining the crash layers to the RWIS layers). That distance was used in a following step of the weather data processing.

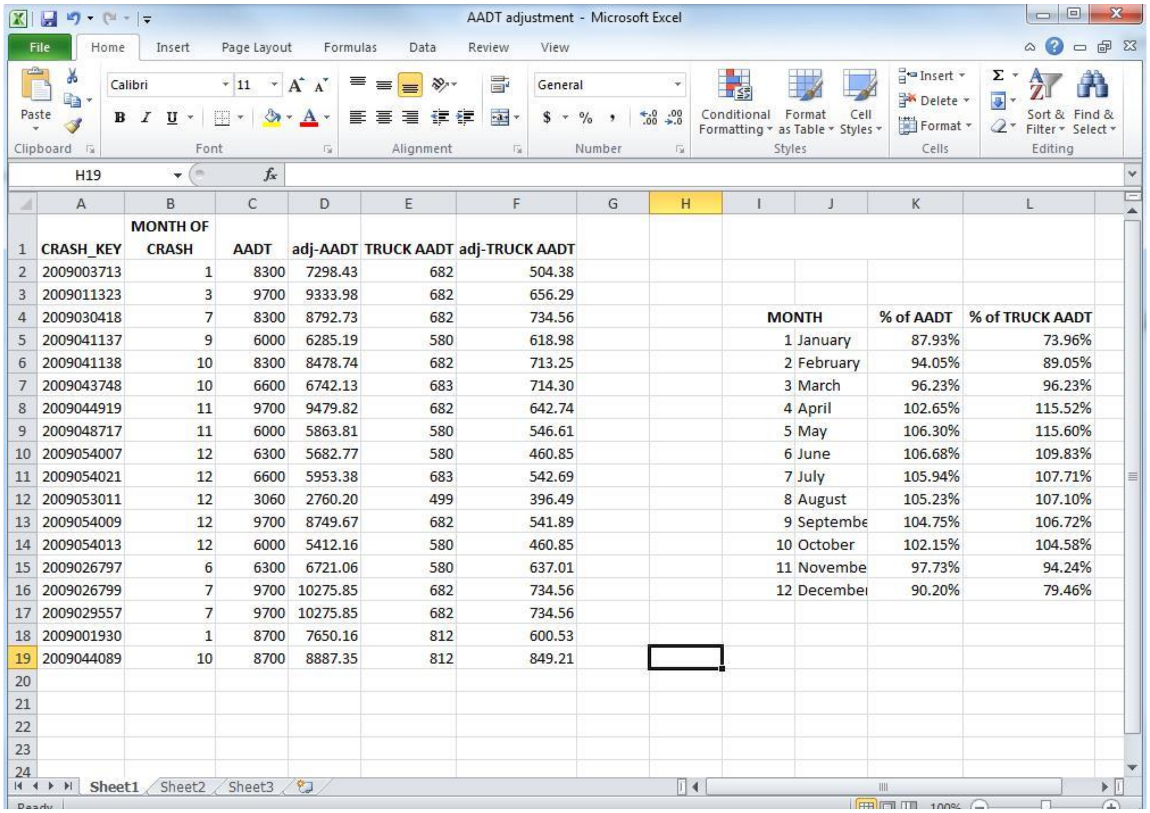
Following the previous process comprehensive datasets with the crash and road segment attributes of the crashes that occurred on the two corridors of study during the period 2009-2011 were created.

The next step was to export these datasets to MS Excel in order to start the numerical processing and create the intermediate dataset with the Crash-Roadway-Traffic Data (see Figure 3.4). At this point, it should be noted that a very careful post-screening procedure was followed in order to eliminate any crashes which were included in the datasets due to errors in the selection and join process. Finally, it should be mentioned that only the crashes that were recorded to have occurred on the mainline of the two corridors were included in the (intermediate and final) datasets. Crashes that were recorded on the ramps were eliminated from the next steps of the analysis.

3.3.4 Intermediate Dataset with Crash-Roadway-Traffic Data

The datasets that were created in ArcMap included all the required crash, roadway and traffic information. Traffic information included the AADT (total and truck) of the road segments as recorded in the GIMS files. However, any seasonal variations of traffic during the year are not captured by AADT. For this reason, the author decided to adjust the values of the AADT that GIMS files provide based on the monthly percentages of variation that were reported in the annual ATR reports (see Section 3.2.4).

The adjustment process was simple. Specifically, the annual value of the AADT was multiplied by the corresponding percentage of the month when each crash happened. For instance, if a crash happened in January of 2009, then the adjusted AADT is the AADT of the road segment of the crash (as reported in GIMS 2009) multiplied by the percentage for January 2009 (as reported in the ATR report of 2009) for the specific classification (interstate or primary) and environment (rural or urban) of the road segment. The calculations were done in MS Excel 2010. Figure 3.13 presents a sample spreadsheet of the process.



	A	B	C	D	E	F
1	CRASH_KEY	MONTH OF CRASH	AADT	adj-AADT	TRUCK AADT	adj-TRUCK AADT
2	2009003713	1	8300	7298.43	682	504.38
3	2009011323	3	9700	9333.98	682	656.29
4	2009030418	7	8300	8792.73	682	734.56
5	2009041137	9	6000	6285.19	580	618.98
6	2009041138	10	8300	8478.74	682	713.25
7	2009043748	10	6600	6742.13	683	714.30
8	2009044919	11	9700	9479.82	682	642.74
9	2009048717	11	6000	5863.81	580	546.61
10	2009054007	12	6300	5682.77	580	460.85
11	2009054021	12	6600	5953.38	683	542.69
12	2009053011	12	3060	2760.20	499	396.49
13	2009054009	12	9700	8749.67	682	541.89
14	2009054013	12	6000	5412.16	580	460.85
15	2009026797	6	6300	6721.06	580	637.01
16	2009026799	7	9700	10275.85	682	734.56
17	2009029557	7	9700	10275.85	682	734.56
18	2009001930	1	8700	7650.16	812	600.53
19	2009044089	10	8700	8887.35	812	849.21

MONTH		% of AADT	% of TRUCK AADT
1	January	87.93%	73.96%
2	February	94.05%	89.05%
3	March	96.23%	96.23%
4	April	102.65%	115.52%
5	May	106.30%	115.60%
6	June	106.68%	109.83%
7	July	105.94%	107.71%
8	August	105.23%	107.10%
9	Septembe	104.75%	106.72%
10	October	102.15%	104.58%
11	Novembe	97.73%	94.24%
12	Decembe	90.20%	79.46%

Figure 3. 13: Sample of Excel Spreadsheet of the AADT Adjustment Process

The previous process for all the selected crashes during the period of study resulted in an intermediate crash-roadway-traffic dataset. That dataset was finally integrated with the other intermediate dataset of the weather data, in order to get the comprehensive dataset with all the information that was used as input in the analysis.

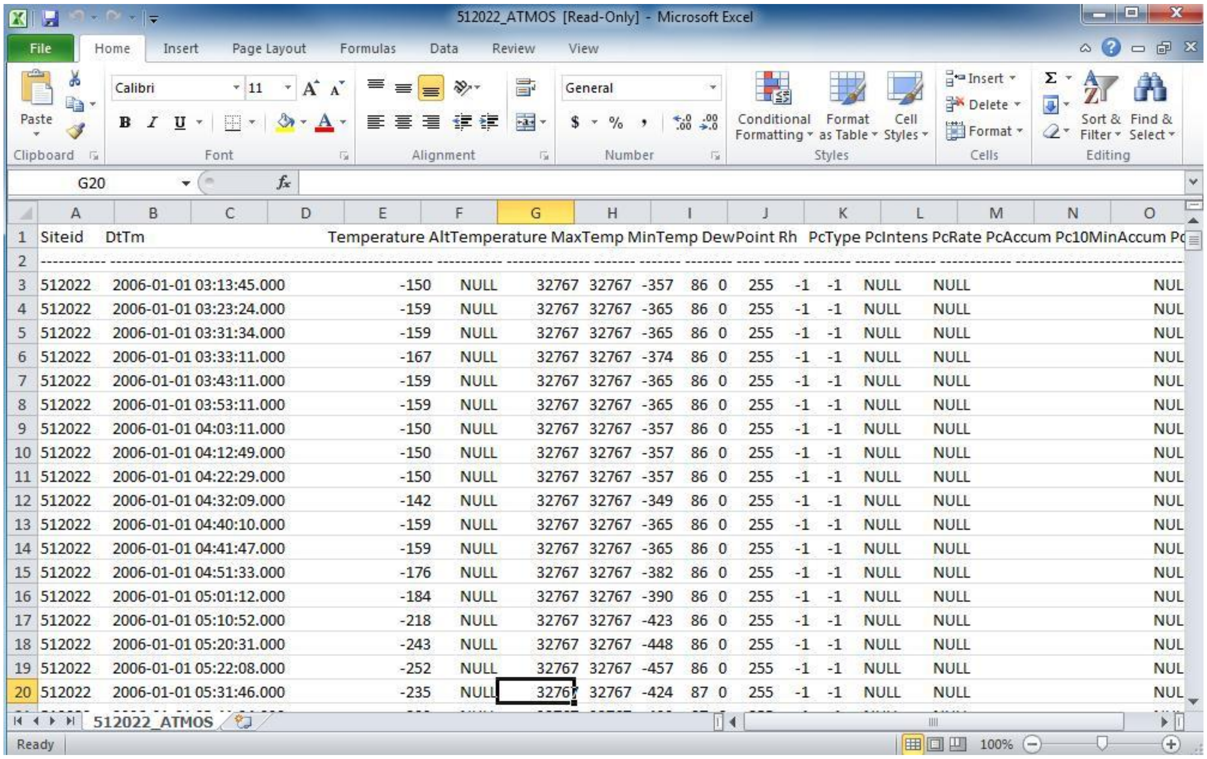
3.3.5 Weather Data Processing

The next step was to create the second intermediate dataset with the weather data. First of all, once the two corridors of study were selected (in ArcMap), the closest RWIS to whose records would be incorporated in the analysis were identified. As already mentioned, records from six RWIS were used. Specifically, the closest RWIS to the I-80 corridor were located at Colfax (sysid: 512053), Grinnell (sysid: 512022) and Williamsburg (sysid: 512048). On the other hand, the closest RWIS to the US-34 corridor were those of Red Oak (sysid: 512038), Creston (sysid: 512013) and Osceola (sysid: 512035).

The Iowa DOT provided the author with raw records of the selected RWIS for the period 2009-2011. Figure 3.14 presents the typical form of a raw dataset from RWIS (as is opened in MS Excel 2010). For this reason, extensive data processing was required.

The main steps of the weather data processing were as follows:

i. Organizing the raw datasets of each RWIS (e.g. cleaning and sorting) by year.
ii. Converting the raw data into the appropriate format (e.g. convert GMT to local time, convert temperature to degrees of C or F, etc.).
iii. Creating the intermediate datasets (for each year and RWIS) with the weather records from the first available day to the last available day of the year (sometimes the records of some days are missing due to instrument malfunctions).



```
   A      B                        C   D   E           F              G        H       I        J   K      L        M      N       O
1  Siteid DtTm                             Temperature AltTemperature MaxTemp MinTemp DewPoint Rh  PcType PcIntens PcRate PcAccum Pc10MinAccum Pc
2  ---------- ---------------------------------------------- ---------- ------------- ------- ------- -------- ---- ------ -------- ------ ------- ----------- ------------------------
3  512022 2006-01-01 03:13:45.000   -150  NULL  32767  32767  -357  86  0  255  -1  -1  NULL  NULL  NUL
4  512022 2006-01-01 03:23:24.000   -159  NULL  32767  32767  -365  86  0  255  -1  -1  NULL  NULL  NUL
5  512022 2006-01-01 03:31:34.000   -159  NULL  32767  32767  -365  86  0  255  -1  -1  NULL  NULL  NUL
6  512022 2006-01-01 03:33:11.000   -167  NULL  32767  32767  -374  86  0  255  -1  -1  NULL  NULL  NUL
7  512022 2006-01-01 03:43:11.000   -159  NULL  32767  32767  -365  86  0  255  -1  -1  NULL  NULL  NUL
8  512022 2006-01-01 03:53:11.000   -159  NULL  32767  32767  -365  86  0  255  -1  -1  NULL  NULL  NUL
9  512022 2006-01-01 04:03:11.000   -150  NULL  32767  32767  -357  86  0  255  -1  -1  NULL  NULL  NUL
10 512022 2006-01-01 04:12:49.000   -150  NULL  32767  32767  -357  86  0  255  -1  -1  NULL  NULL  NUL
11 512022 2006-01-01 04:22:29.000   -150  NULL  32767  32767  -357  86  0  255  -1  -1  NULL  NULL  NUL
12 512022 2006-01-01 04:32:09.000   -142  NULL  32767  32767  -349  86  0  255  -1  -1  NULL  NULL  NUL
13 512022 2006-01-01 04:40:10.000   -159  NULL  32767  32767  -365  86  0  255  -1  -1  NULL  NULL  NUL
14 512022 2006-01-01 04:41:47.000   -159  NULL  32767  32767  -365  86  0  255  -1  -1  NULL  NULL  NUL
15 512022 2006-01-01 04:51:33.000   -176  NULL  32767  32767  -382  86  0  255  -1  -1  NULL  NULL  NUL
16 512022 2006-01-01 05:01:12.000   -184  NULL  32767  32767  -390  86  0  255  -1  -1  NULL  NULL  NUL
17 512022 2006-01-01 05:10:52.000   -218  NULL  32767  32767  -423  86  0  255  -1  -1  NULL  NULL  NUL
18 512022 2006-01-01 05:20:31.000   -243  NULL  32767  32767  -448  86  0  255  -1  -1  NULL  NULL  NUL
19 512022 2006-01-01 05:22:08.000   -252  NULL  32767  32767  -457  86  0  255  -1  -1  NULL  NULL  NUL
20 512022 2006-01-01 05:31:46.000   -235  NULL  32767  32767  -424  87  0  255  -1  -1  NULL  NULL  NUL
```

Figure 3. 14: Typical Form of a Raw Dataset from RWIS

3.3.5.1 Organizing the Raw Datasets

An initial processing and organizing of the raw datasets was required, since the datasets were in a raw format, as it is shown in Figure 3.14. During the process of organizing, all the records (rows of the datasets) were sorted by time (oldest to newest). Furthermore, the all the information (columns) which would not be used in the analysis or were incomplete was removed. Finally, it is noteworthy that the number of rows in each spreadsheet was really large (in some cases was larger than 500,000) for the three years of the analysis period. This was due to the fact that (as mentioned in section 3.2.2) a RWIS has the ability to automatically update its records each time a change in the weather conditions occurs. Thus in some cases a new record (row in the spreadsheet) was observed every two (or even less than one) minutes, whenever rapid changes in the weather conditions occurred. For this reason, the records were split by year for each RWIS. That also made the MS Excel files smaller in size and easier in processing.

3.3.5.2 Converting the Raw Data

The data still remained in a format which was not appropriate for statistical analysis, even after the first step of processing. For instance, the temperature records were not in degrees C (or F) but in 0.01 degrees C, while the precipitation rate was in 0.025 mm/hr, and not in mm/hr (or in/hr). Furthermore, the reference time of each record was the GMT and not the local time. Thus, the records could not be related directly to each crash and integrated with the dataset of the Crash-Roadway-Traffic Data. For this reason, the appropriate conversions had to be done.

First the time of each record was converted from GMT to local. In other words, the time was actually “moved” 6 hours before (or 5 hours during the daylight saving period – summer time). For the values of the weather conditions, the conversion table of Figure 3.15 which was provided by the Iowa DOT was used.

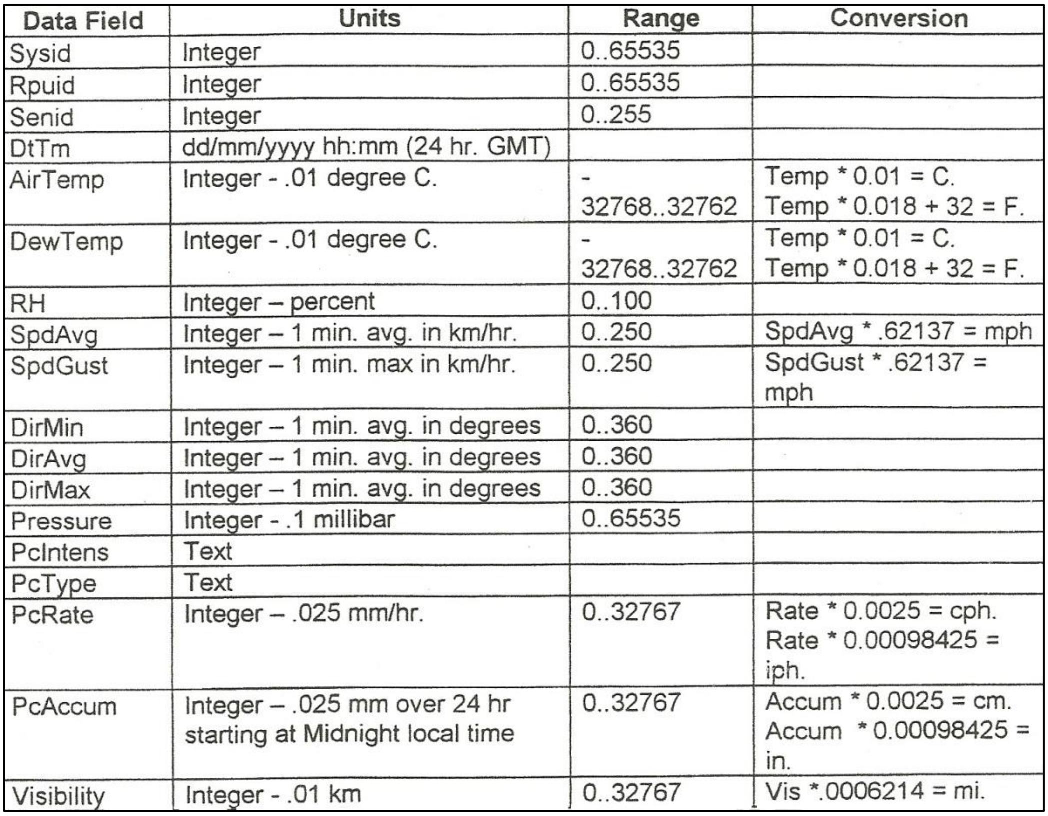
3.3.5.3 Intermediate Dataset with Weather Data

After the conversion process was finished, the author was able to create the intermediate dataset with the weather data. Specifically, an intermediate dataset for each year and each RWIS was created. These datasets included all the weather variables that were processed in the previous step and would be used in the analysis.

In particular, the weather variables were:

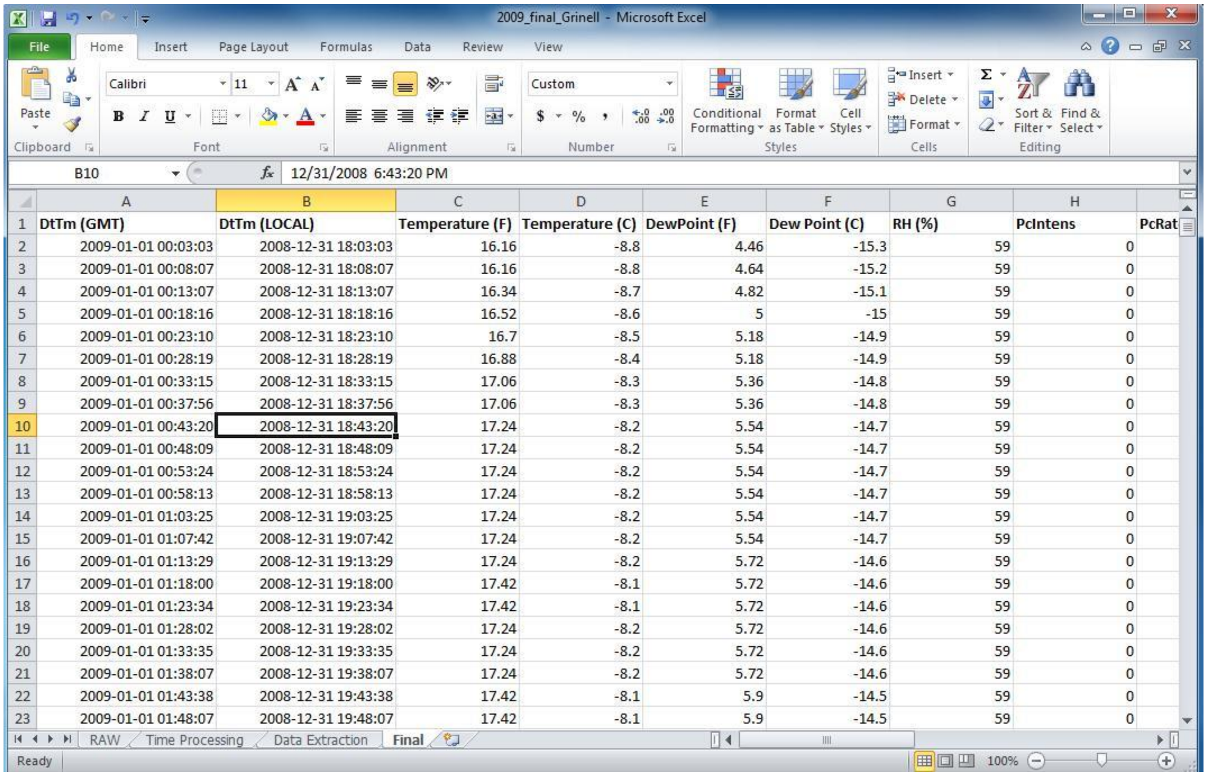
1. Temperature (in degrees C and F)
2. Dew Point Temperature (in degrees C and F)
3. Relative Humidity (RH %)
4. Precipitation Intensity (categorical)
5. Precipitation Rate (in mm/hr and in/hr)
6. Average Wind Speed (in Km/hr and mi/hr)
7. Wind Speed Gust (in Km/hr and mi/hr)
8. Wind Speed Average Direction (in degrees)
9. Wind Speed Gust Average Direction (in degrees)

A sample of an intermediate dataset with the weather data is presented in Figure 3.16.



Data Field	Units	Range	Conversion
Sysid	Integer	0..65535	
Rpuid	Integer	0..65535	
Senid	Integer	0..255	
DtTm	dd/mm/yyyy hh:mm (24 hr. GMT)		
AirTemp	Integer - .01 degree C.	- 32768..32762	Temp * 0.01 = C. Temp * 0.018 + 32 = F.
DewTemp	Integer - .01 degree C.	- 32768..32762	Temp * 0.01 = C. Temp * 0.018 + 32 = F.
RH	Integer – percent	0..100	
SpdAvg	Integer – 1 min. avg. in km/hr.	0..250	SpdAvg * .62137 = mph
SpdGust	Integer – 1 min. max in km/hr.	0..250	SpdGust * .62137 = mph
DirMin	Integer – 1 min. avg. in degrees	0..360	
DirAvg	Integer – 1 min. avg. in degrees	0..360	
DirMax	Integer – 1 min. avg. in degrees	0..360	
Pressure	Integer - .1 millibar	0..65535	
PcIntens	Text		
PcType	Text		
PcRate	Integer – .025 mm/hr.	0..32767	Rate * 0.0025 = cph. Rate * 0.00098425 = iph.
PcAccum	Integer – .025 mm over 24 hr starting at Midnight local time	0..32767	Accum * 0.0025 = cm. Accum * 0.00098425 = in.
Visibility	Integer - .01 km	0..32767	Vis *.0006214 = mi.

Figure 3. 15: Conversion Table (Source: Iowa DOT)



B10 12/31/2008 6:43:20 PM

	A	B	C	D	E	F	G	H	
1	DtTm (GMT)	DtTm (LOCAL)	Temperature (F)	Temperature (C)	DewPoint (F)	Dew Point (C)	RH (%)	PcIntens	PcRat
2	2009-01-01 00:03:03	2008-12-31 18:03:03	16.16	-8.8	4.46	-15.3	59	0	
3	2009-01-01 00:08:07	2008-12-31 18:08:07	16.16	-8.8	4.64	-15.2	59	0	
4	2009-01-01 00:13:07	2008-12-31 18:13:07	16.34	-8.7	4.82	-15.1	59	0	
5	2009-01-01 00:18:16	2008-12-31 18:18:16	16.52	-8.6	5	-15	59	0	
6	2009-01-01 00:23:10	2008-12-31 18:23:10	16.7	-8.5	5.18	-14.9	59	0	
7	2009-01-01 00:28:19	2008-12-31 18:28:19	16.88	-8.4	5.18	-14.9	59	0	
8	2009-01-01 00:33:15	2008-12-31 18:33:15	17.06	-8.3	5.36	-14.8	59	0	
9	2009-01-01 00:37:56	2008-12-31 18:37:56	17.06	-8.3	5.36	-14.8	59	0	
10	2009-01-01 00:43:20	2008-12-31 18:43:20	17.24	-8.2	5.54	-14.7	59	0	
11	2009-01-01 00:48:09	2008-12-31 18:48:09	17.24	-8.2	5.54	-14.7	59	0	
12	2009-01-01 00:53:24	2008-12-31 18:53:24	17.24	-8.2	5.54	-14.7	59	0	
13	2009-01-01 00:58:13	2008-12-31 18:58:13	17.24	-8.2	5.54	-14.7	59	0	
14	2009-01-01 01:03:25	2008-12-31 19:03:25	17.24	-8.2	5.54	-14.7	59	0	
15	2009-01-01 01:07:42	2008-12-31 19:07:42	17.24	-8.2	5.54	-14.7	59	0	
16	2009-01-01 01:13:29	2008-12-31 19:13:29	17.24	-8.2	5.72	-14.6	59	0	
17	2009-01-01 01:18:00	2008-12-31 19:18:00	17.42	-8.1	5.72	-14.6	59	0	
18	2009-01-01 01:23:34	2008-12-31 19:23:34	17.42	-8.1	5.72	-14.6	59	0	
19	2009-01-01 01:28:02	2008-12-31 19:28:02	17.24	-8.2	5.72	-14.6	59	0	
20	2009-01-01 01:33:35	2008-12-31 19:33:35	17.24	-8.2	5.72	-14.6	59	0	
21	2009-01-01 01:38:07	2008-12-31 19:38:07	17.24	-8.2	5.72	-14.6	59	0	
22	2009-01-01 01:43:38	2008-12-31 19:43:38	17.42	-8.1	5.9	-14.5	59	0	
23	2009-01-01 01:48:07	2008-12-31 19:48:07	17.42	-8.1	5.9	-14.5	59	0	

Figure 3. 16: Sample of an Intermediate Dataset with Weather Data

3.3.6 Integrating the Two Intermediate Datasets

The last task of the data integration process was to integrate the information of the two intermediate datasets. As base for this integration the time of each crash was used. The main steps of this process (as also shown in Figure 3.5) were:

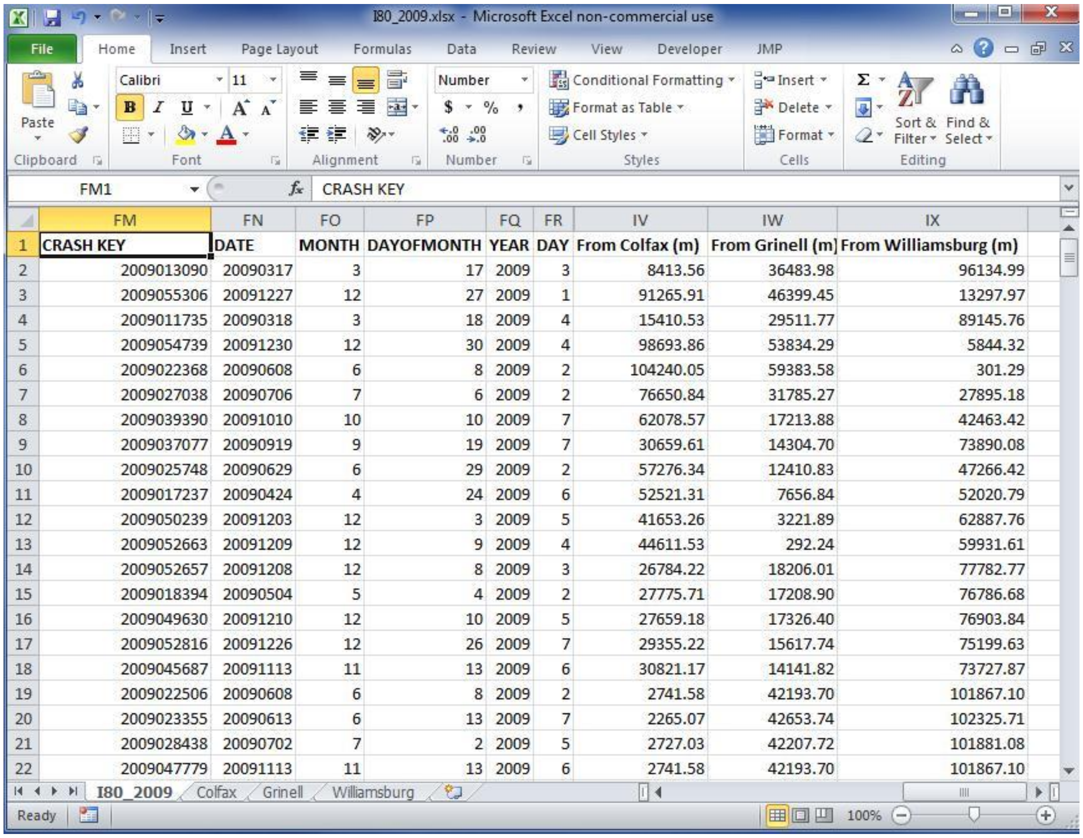
i. Assigning each crash to the nearest RWIS

ii. Identifying in the datasets of the weather data, the records for a period of an hour before the time of each crash

iii. Deriving the average values for the prevailing weather conditions for a period of an hour prior to each crash (i.e., hourly weather conditions)

3.3.6.1 Assigning Each Crash to the Nearest RWIS

The process of assigning the crashes to their nearest RWIS was based on the distance between the location of each crash (crash point) and the location of the corresponding RWIS on the corridor of the crash. As mentioned in a previous step, the distance between each crash point to each RWIS was calculated in ArcMap. These distances were included in the dataset with the crash data (Figure 3.17).



	FM	FN	FO	FP	FQ	FR	IV	IW	IX
1	CRASH KEY	DATE	MONTH	DAYOFMONTH	YEAR	DAY	From Colfax (m)	From Grinell (m)	From Williamsburg (m)
2	2009013090	20090317	3	17	2009	3	8413.56	36483.98	96134.99
3	2009055306	20091227	12	27	2009	1	91265.91	46399.45	13297.97
4	2009011735	20090318	3	18	2009	4	15410.53	29511.77	89145.76
5	2009054739	20091230	12	30	2009	4	98693.86	53834.29	5844.32
6	2009022368	20090608	6	8	2009	2	104240.05	59383.58	301.29
7	2009027038	20090706	7	6	2009	2	76650.84	31785.27	27895.18
8	2009039390	20091010	10	10	2009	7	62078.57	17213.88	42463.42
9	2009037077	20090919	9	19	2009	7	30659.61	14304.70	73890.08
10	2009025748	20090629	6	29	2009	2	57276.34	12410.83	47266.42
11	2009017237	20090424	4	24	2009	6	52521.31	7656.84	52020.79
12	2009050239	20091203	12	3	2009	5	41653.26	3221.89	62887.76
13	2009052663	20091209	12	9	2009	4	44611.53	292.24	59931.61
14	2009052657	20091208	12	8	2009	3	26784.22	18206.01	77782.77
15	2009018394	20090504	5	4	2009	2	27775.71	17208.90	76786.68
16	2009049630	20091210	12	10	2009	5	27659.18	17326.40	76903.84
17	2009052816	20091226	12	26	2009	7	29355.22	15617.74	75199.63
18	2009045687	20091113	11	13	2009	6	30821.17	14141.82	73727.87
19	2009022506	20090608	6	8	2009	2	2741.58	42193.70	101867.10
20	2009023355	20090613	6	13	2009	7	2265.07	42653.74	102325.71
21	2009028438	20090702	7	2	2009	5	2727.03	42207.72	101881.08
22	2009047779	20091113	11	13	2009	6	2741.58	42193.70	101867.10

Figure 3. 17: Fields of the Computed Distances between Crash Points and RWIS locations

The rule that was used for the assigning was the following:

- Each crash was assigned to the nearest RWIS if its distance from that was less than 10 miles (16 Km).
- If a crash point had a distance of more than 10 miles from each of the two nearest RWIS, then the crash was assigned to both RWIS. In this case, for the derivation of the average weather values, a weighted average was used.

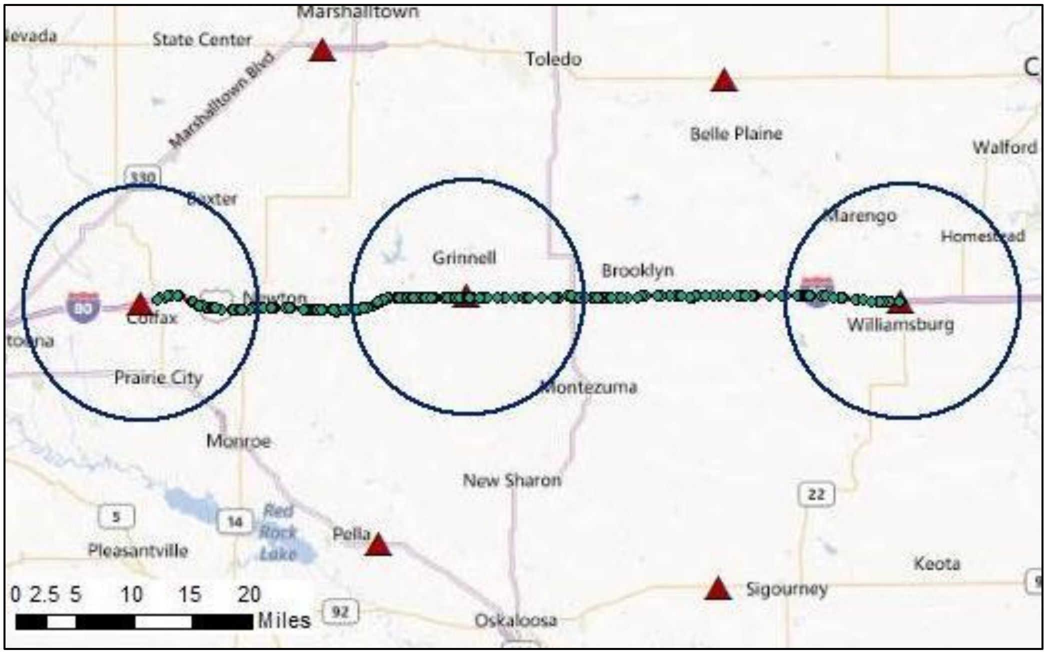


Figure 3. 18: Assigning Rule (Graphical Illustration)

The previous rule is depicted graphically in Figure 3.18. In other words, the crashes (green points) that were located inside the circles (radius = 10 miles) were assigned to one RWIS (that of the center of the circle). The crashes located outside of the circles were assigned to the two nearest RWIS.

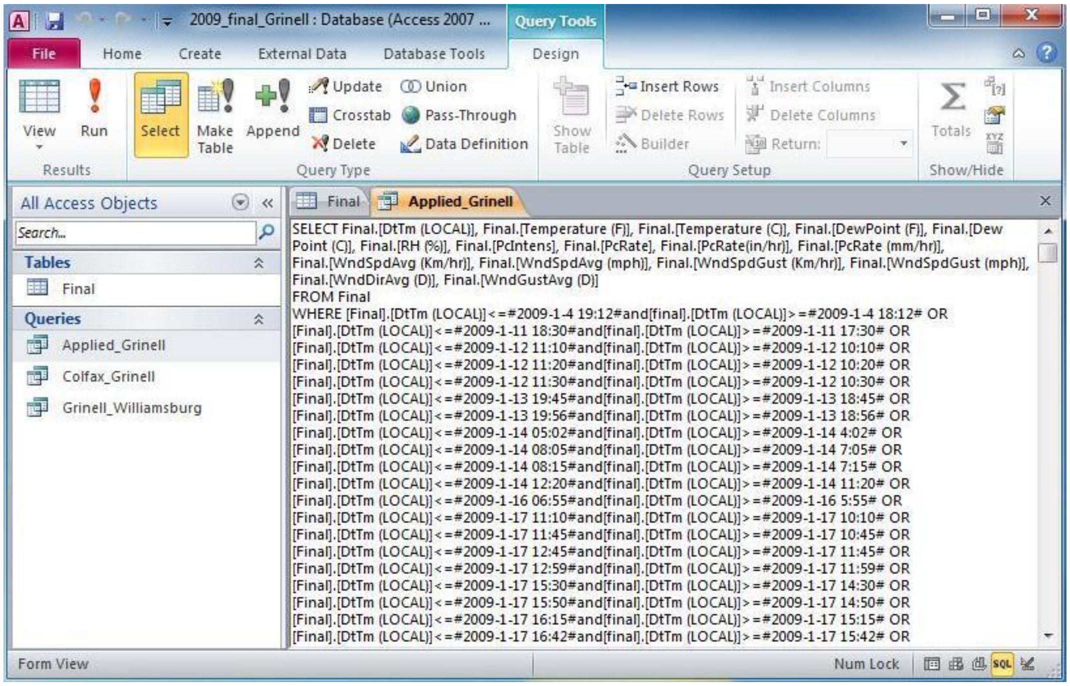
Note that there is not a standard rule for assigning crash locations to weather stations in the literature, but rather researchers have selected a rule that was convenient based on the availability of the weather stations, the area and purpose of the study. For instance, Hermans et al. (2006) assigned each crash to the nearest weather station in a Dutch nationwide study. In a similar way, Shankar et al. (2004) assigned each of the one-mile segments of their study area to the nearest weather station. On the other hand, some authors prefer to combine reports from different weather stations. Usman et al. (2012a, b) used the arithmetic mean of the values of the weather parameters that were recorded by different weather stations which were located within an (arbitrarily) defined 60 Km buffering zone around each route of their study. Finally, Jung et al. (2012) tried to approximate real-time weather values by applying an inverse squared distance rule to the weather records of the nearest three stations to each crash location.

In this thesis, the aforementioned rule (Figure 3.18) was decided, based on the spatial distribution of the RWIS and the local climate conditions.

3.3.6.2 Identifying the Records for a Period of an Hour prior to each Crash

The next step was to identify the records in the datasets with the weather data from which the average (hourly) weather conditions would be derived for a period of an hour prior to each crash. This task was performed in MS Access 2010.

During this process, the datasets with the weather data were first imported into MS Access (for each year and RWIS). Then, having already assigned each crash to the corresponding RWIS, the time of each crash was used in order to create queries. Specifically, SQL Queries which selected all the records in the dataset for a period of an hour prior to each crash were used. Figure 3.19 presents a sample code for a SQL query that was used. It should be noted that for the crashes which had been assigned to two RWIS, the query was performed in the datasets of both RWIS. At the end of the process all the required records in order to derive the average prevailing weather conditions had been extracted.



```
SELECT Final.[DtTm (LOCAL)], Final.[Temperature (F)], Final.[Temperature (C)], Final.[DewPoint (F)], Final.[Dew
Point (C)], Final.[RH (%)], Final.[PcIntens], Final.[PcRate], Final.[PcRate(in/hr)], Final.[PcRate (mm/hr)],
Final.[WndSpdAvg (Km/hr)], Final.[WndSpdAvg (mph)], Final.[WndSpdGust (Km/hr)], Final.[WndSpdGust (mph)],
Final.[WndDirAvg (D)], Final.[WndGustAvg (D)]
FROM Final
WHERE [Final].[DtTm (LOCAL)]<=#2009-1-4 19:12#and[final].[DtTm (LOCAL)]>=#2009-1-4 18:12# OR
[Final].[DtTm (LOCAL)]<=#2009-1-11 18:30#and[final].[DtTm (LOCAL)]>=#2009-1-11 17:30# OR
[Final].[DtTm (LOCAL)]<=#2009-1-12 11:10#and[final].[DtTm (LOCAL)]>=#2009-1-12 10:10# OR
[Final].[DtTm (LOCAL)]<=#2009-1-12 11:20#and[final].[DtTm (LOCAL)]>=#2009-1-12 10:20# OR
[Final].[DtTm (LOCAL)]<=#2009-1-12 11:30#and[final].[DtTm (LOCAL)]>=#2009-1-12 10:30# OR
[Final].[DtTm (LOCAL)]<=#2009-1-13 19:45#and[final].[DtTm (LOCAL)]>=#2009-1-13 18:45# OR
[Final].[DtTm (LOCAL)]<=#2009-1-13 19:56#and[final].[DtTm (LOCAL)]>=#2009-1-13 18:56# OR
[Final].[DtTm (LOCAL)]<=#2009-1-14 05:02#and[final].[DtTm (LOCAL)]>=#2009-1-14 4:02# OR
[Final].[DtTm (LOCAL)]<=#2009-1-14 08:05#and[final].[DtTm (LOCAL)]>=#2009-1-14 7:05# OR
[Final].[DtTm (LOCAL)]<=#2009-1-14 08:15#and[final].[DtTm (LOCAL)]>=#2009-1-14 7:15# OR
[Final].[DtTm (LOCAL)]<=#2009-1-14 12:20#and[final].[DtTm (LOCAL)]>=#2009-1-14 11:20# OR
[Final].[DtTm (LOCAL)]<=#2009-1-16 06:55#and[final].[DtTm (LOCAL)]>=#2009-1-16 5:55# OR
[Final].[DtTm (LOCAL)]<=#2009-1-17 11:10#and[final].[DtTm (LOCAL)]>=#2009-1-17 10:10# OR
[Final].[DtTm (LOCAL)]<=#2009-1-17 11:45#and[final].[DtTm (LOCAL)]>=#2009-1-17 10:45# OR
[Final].[DtTm (LOCAL)]<=#2009-1-17 12:45#and[final].[DtTm (LOCAL)]>=#2009-1-17 11:45# OR
[Final].[DtTm (LOCAL)]<=#2009-1-17 12:59#and[final].[DtTm (LOCAL)]>=#2009-1-17 11:59# OR
[Final].[DtTm (LOCAL)]<=#2009-1-17 15:30#and[final].[DtTm (LOCAL)]>=#2009-1-17 14:30# OR
[Final].[DtTm (LOCAL)]<=#2009-1-17 15:50#and[final].[DtTm (LOCAL)]>=#2009-1-17 14:50# OR
[Final].[DtTm (LOCAL)]<=#2009-1-17 16:15#and[final].[DtTm (LOCAL)]>=#2009-1-17 15:15# OR
[Final].[DtTm (LOCAL)]<=#2009-1-17 16:42#and[final].[DtTm (LOCAL)]>=#2009-1-17 15:42# OR
```

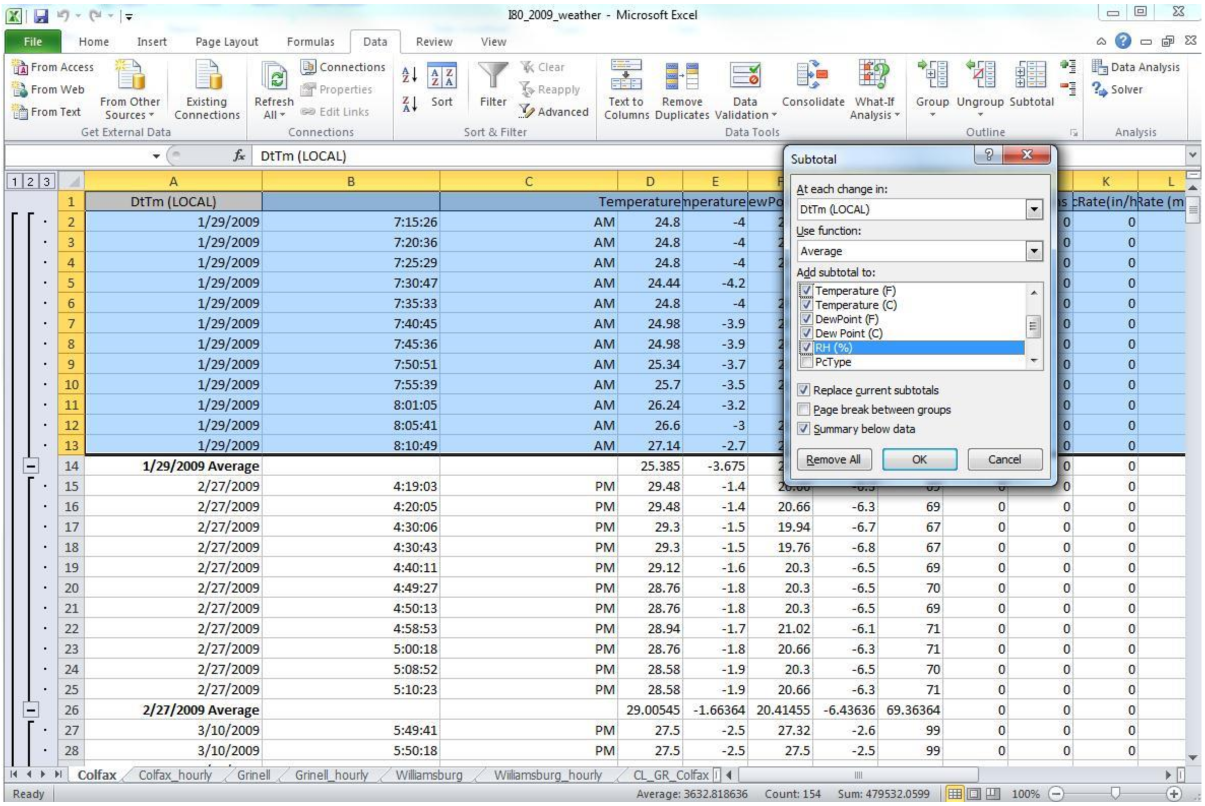
Figure 3. 19: Sample Code for SQL Query in MS Access 2010

3.3.6.3 Deriving the Average Prevailing Weather Conditions

The datasets that were created in MS Access were opened in MS Excel in order to derive (i.e., calculate) the average values of the prevailing weather conditions for an hour prior to each crash. For the calculations the *"subtotal"* tool of Excel was used (see Figure 3.20).

The *"subtotal"* tool can be found on the *"Data"* ribbon of Excel. This option allows the user to calculate various values (such as averages, summaries, maximum, minimum, etc.)

related to specific parts (subtotals) of a spreadsheet. In this particular case, each subtotal included all the weather records during the period of an hour prior to each crash. Thus, unless some records were missing, in each spreadsheet there were as many subtotals as the number of crashes assigned to the specific RWIS for the specific year.



DtTm (LOCAL)

	A	B	C	D	E	F	G	H	I	J	K	L
1	DtTm (LOCAL)			Temperature	nperature	ewPo				s	cRate(in/h	Rate (m
2	1/29/2009	7:15:26	AM	24.8	-4					0	0	
3	1/29/2009	7:20:36	AM	24.8	-4					0	0	
4	1/29/2009	7:25:29	AM	24.8	-4					0	0	
5	1/29/2009	7:30:47	AM	24.44	-4.2					0	0	
6	1/29/2009	7:35:33	AM	24.8	-4					0	0	
7	1/29/2009	7:40:45	AM	24.98	-3.9					0	0	
8	1/29/2009	7:45:36	AM	24.98	-3.9					0	0	
9	1/29/2009	7:50:51	AM	25.34	-3.7					0	0	
10	1/29/2009	7:55:39	AM	25.7	-3.5					0	0	
11	1/29/2009	8:01:05	AM	26.24	-3.2					0	0	
12	1/29/2009	8:05:41	AM	26.6	-3					0	0	
13	1/29/2009	8:10:49	AM	27.14	-2.7					0	0	
14	**1/29/2009 Average**			25.385	-3.675					0	0	
15	2/27/2009	4:19:03	PM	29.48	-1.4					0	0	
16	2/27/2009	4:20:05	PM	29.48	-1.4	20.66	-6.3	69	0	0	0	
17	2/27/2009	4:30:06	PM	29.3	-1.5	19.94	-6.7	67	0	0	0	
18	2/27/2009	4:30:43	PM	29.3	-1.5	19.76	-6.8	67	0	0	0	
19	2/27/2009	4:40:11	PM	29.12	-1.6	20.3	-6.5	69	0	0	0	
20	2/27/2009	4:49:27	PM	28.76	-1.8	20.3	-6.5	70	0	0	0	
21	2/27/2009	4:50:13	PM	28.76	-1.8	20.3	-6.5	69	0	0	0	
22	2/27/2009	4:58:53	PM	28.94	-1.7	21.02	-6.1	71	0	0	0	
23	2/27/2009	5:00:18	PM	28.76	-1.8	20.66	-6.3	71	0	0	0	
24	2/27/2009	5:08:52	PM	28.58	-1.9	20.3	-6.5	70	0	0	0	
25	2/27/2009	5:10:23	PM	28.58	-1.9	20.66	-6.3	71	0	0	0	
26	**2/27/2009 Average**			29.00545	-1.66364	20.41455	-6.43636	69.36364	0	0	0	
27	3/10/2009	5:49:41	PM	27.5	-2.5	27.32	-2.6	99	0	0	0	
28	3/10/2009	5:50:18	PM	27.5	-2.5	27.5	-2.5	99	0	0	0	

Subtotal

At each change in:
DtTm (LOCAL)

Use function:
Average

Add subtotal to:
☑ Temperature (F)
☑ Temperature (C)
☑ DewPoint (F)
☑ Dew Point (C)
☑ RH (%)
☐ PcType

☑ Replace current subtotals
☐ Page break between groups
☑ Summary below data

Remove All | OK | Cancel

Average: 3632.818636 Count: 154 Sum: 479532.0599

Figure 3. 20: The “Subtotal” Tool of MS Excel 2010

Weather Conditions (Variables)

From this process the hourly values for the following weather conditions (that used as variables in the analysis) were derived:

- Average Temperature (C)
- Average Dew Point Temperature (C)

- Average Relative Humidity (RH)
- Type of Precipitation
- Total Accumulated Amount (Height) of Precipitation (mm)
- Average Wind Speed (Km/hr)
- Average Wind Speed Gust (Km/hr)
- Average Wind Direction (Km/hr)
- Average Wind Gust Direction (Km/hr)

The *"average option"* of the *"subtotal"* tool for the weather conditions was used, for the variables where average values could be used (temperature, wind speed etc.). The precipitation though should be treated in a different manner. As far as the precipitation type is concerned, there were cases where during a period of an hour, two or even three different types of precipitation were observed. For instance, the records might show no precipitation during the first quarter of an hour, rain during the following 10 minutes and snow for the rest of the hour! In such cases, the most prevailing type of precipitation was assumed (e.g. snow in the previous example).

Regarding the amount of precipitation, the author decided to use the total accumulated height of precipitation that fell during the hourly period. As mentioned previously, the Precipitation Rate was obtained from the RWIS records. That variable was in mm/hr (after the necessary conversions). Based on the values of this variable, the precipitation height was calculated. For the calculation, the precipitation rate was multiplied by the duration of the event (i.e., the period between two consecutive records in my dataset). In other words, the following formula was used:

$$PcHeight = PcRate\frac{PcDuration}{60} \tag{3.1}$$

where: $PcHeight$ = Amount (Height) of Precipitation during the last hour (mm)

$PcRate$ = Rate of Precipitation (mm/hr)

$PcDuration$ = Duration of Precipitation Event (min)

From the previous calculation a new column was created in the spreadsheets. Then, by applying the *"subtotal"* tool, but using the *"sum option"* at this time, the total amount (height) of precipitation that fell during the period of an hour prior to each crash was computed.

It should be mentioned here that the previous average values correspond to each crash and thus could be integrated with the rest of information for each crash. However, that applied only to the crashes which were assigned to one RWIS. For the crashes which were assigned to two RWIS the weighted average of those values (based on distance) should be calculated. In order to do that, the following formula was used:

$$W.V. = Value_A\frac{TotDist. - CrashDist._A}{TotDist.} + Value_B\frac{TotDist. - CrashDist._B}{TotDist.} \tag{3.2}$$

where: $W.V.$ = Weighted Value of the Weather Parameter

$Value_A$ = Value of the Parameter as Reported in RWIS A

$CrashDist._A$ = Distance between the Crash Location and RWIS A

$Value_B$ = Value of the Parameter as Reported in RWIS B

$CrashDist_B$ = Distance between the Crash Location and RWIS B

$TotDist$ = Distance between RWIS A and B

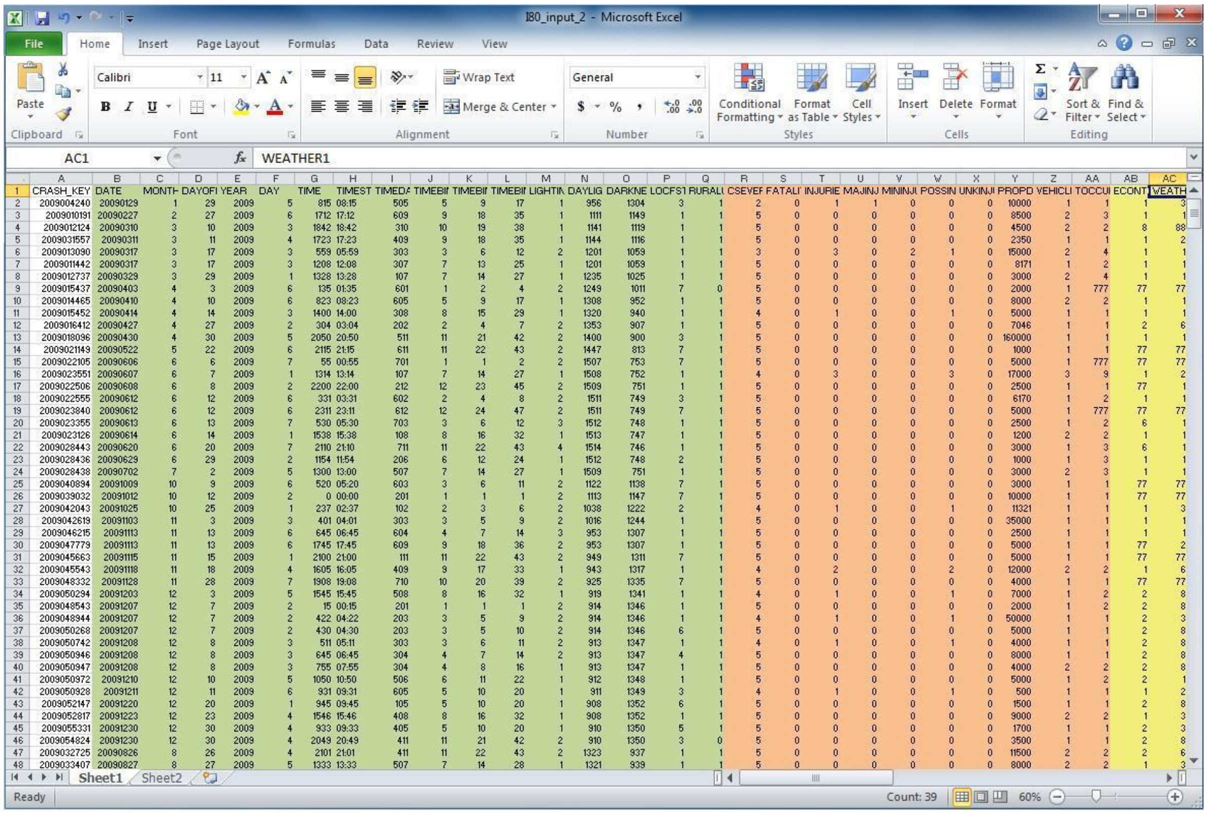
Finally, it should be noted that for the cases where the records from only one of the two RWIS were available, only the available values were used.

3.3.7 Creating the Final Dataset – Input for the Analysis

Once the processing of the weather data was finished, all the required parameters for each crash had been determined. These parameters were related to crash attributes, traffic conditions and weather conditions of the roadway. Thus, the final dataset which constituted the input for the statistical analysis could be created.

In order to create the final dataset, all the data from the different spreadsheets was integrated by using the *"CRASH KEY"* of each crash as base for the integration. Eventually, two datasets were created, one for the I-80 corridor and one for the US-34 corridor, which contained all the crashes along with their corresponding parameters over the period 2009-2011. Figure 3.21 presents a part of the dataset of I-80 corridor. Each row in the dataset refers to a crash and each column to a specific crash parameter. The columns are colored according to the type of their parameters (e.g., location/ time, traffic, weather, etc.).

The aforementioned datasets were then utilized in the statistical analysis that will be the subject of the following chapter.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC
1	CRASH_KEY	DATE	MONTH	DAYOFI	YEAR	DAY	TIME	TIMEST	TIMEDA	TIMEBI	TIMEBI	TIMEBI	LIGHTIN	DAYLIG	DARKNE	LOCFST	RURALU	CSEVER	FATALI	INJURIE	MAJINJ	MININJ	POSSIN	UNKINJ	PROPD	VEHICL	TOCCU	ECONT	WEATH
2	2009004240	20090129	1	29	2009	5	815	08:15	505	5	9	17	1	956	1304	3	1	2	0	1	1	0	0	0	10000	1	1	1	3
3	2009010191	20090227	2	27	2009	6	1712	17:12	609	9	18	35	1	1111	1149	1	1	5	0	0	0	0	0	0	8500	2	3	1	1
4	2009012124	20090310	3	10	2009	3	1842	18:42	310	10	19	38	1	1141	1119	1	1	5	0	0	0	0	0	0	4500	2	2	8	88
5	2009031557	20090311	3	11	2009	4	1723	17:23	409	9	18	35	1	1144	1116	1	1	5	0	0	0	0	0	0	2350	1	1	1	2
6	2009013090	20090317	3	17	2009	3	559	05:59	303	3	6	12	2	1201	1059	1	1	3	0	3	0	2	1	0	15000	2	4	1	1
7	2009011442	20090317	3	17	2009	3	1208	12:08	307	7	13	25	1	1201	1059	1	1	5	0	0	0	0	0	0	8171	1	2	1	1
8	2009012737	20090329	3	29	2009	1	1328	13:28	107	7	14	27	1	1235	1025	1	1	5	0	0	0	0	0	0	3000	2	4	1	1
9	2009015437	20090403	4	3	2009	6	135	01:35	601	1	2	4	2	1249	1011	7	0	5	0	0	0	0	0	0	2000	1	777	77	77
10	2009014465	20090410	4	10	2009	6	823	08:23	605	5	9	17	1	1308	952	1	1	5	0	0	0	0	0	0	8000	2	2	1	1
11	2009015452	20090414	4	14	2009	3	1400	14:00	308	8	15	29	1	1320	940	1	1	4	0	1	0	0	1	0	5000	1	1	1	1
12	2009016412	20090427	4	27	2009	2	304	03:04	202	2	4	7	2	1353	907	1	1	5	0	0	0	0	0	0	7046	1	1	2	6
13	2009018096	20090430	4	30	2009	5	2050	20:50	511	11	21	42	2	1400	900	3	1	5	0	0	0	0	0	0	160000	1	1	1	1
14	2009021149	20090522	5	22	2009	6	2115	21:15	611	11	22	43	2	1447	813	7	1	5	0	0	0	0	0	0	1000	1	1	77	77
15	2009022105	20090606	6	6	2009	7	55	00:55	701	1	1	2	2	1507	753	7	1	5	0	0	0	0	0	0	5000	1	777	77	77
16	2009023551	20090607	6	7	2009	1	1314	13:14	107	7	14	27	1	1508	752	1	1	4	0	3	0	0	3	0	17000	3	9	1	2
17	2009022506	20090608	6	8	2009	2	2200	22:00	212	12	23	45	2	1509	751	1	1	5	0	0	0	0	0	0	2500	1	1	77	1
18	2009022555	20090612	6	12	2009	6	331	03:31	602	2	4	8	2	1511	749	3	1	5	0	0	0	0	0	0	6170	1	2	1	1
19	2009023840	20090612	6	12	2009	6	2311	23:11	612	12	24	47	2	1511	749	7	1	5	0	0	0	0	0	0	5000	1	777	77	77
20	2009023355	20090613	6	13	2009	7	530	05:30	703	3	6	12	3	1512	748	1	1	5	0	0	0	0	0	0	2500	1	2	6	1
21	2009023126	20090614	6	14	2009	1	1538	15:38	108	8	16	32	1	1513	747	1	1	5	0	0	0	0	0	0	1200	2	2	1	1
22	2009028443	20090620	6	20	2009	7	2110	21:10	711	11	22	43	4	1514	746	1	1	5	0	0	0	0	0	0	3000	1	3	6	1
23	2009028436	20090629	6	29	2009	2	1154	11:54	206	6	12	24	1	1512	748	2	1	5	0	0	0	0	0	0	1000	1	3	1	1
24	2009028438	20090702	7	2	2009	5	1300	13:00	507	7	14	27	1	1509	751	1	1	5	0	0	0	0	0	0	3000	2	3	1	1
25	2009040894	20091009	10	9	2009	6	520	05:20	603	3	6	11	2	1122	1138	7	1	5	0	0	0	0	0	0	3000	1	1	77	77
26	2009039032	20091012	10	12	2009	2	0	00:00	201	1	1	1	2	1113	1147	7	1	5	0	0	0	0	0	0	10000	1	1	77	77
27	2009042043	20091025	10	25	2009	1	237	02:37	102	2	3	6	2	1038	1222	2	1	4	0	1	0	0	1	0	11321	1	1	1	3
28	2009042619	20091103	11	3	2009	3	401	04:01	303	3	5	9	2	1016	1244	1	1	5	0	0	0	0	0	0	35000	1	1	1	1
29	2009046215	20091113	11	13	2009	6	645	06:45	604	4	7	14	3	953	1307	1	1	5	0	0	0	0	0	0	2500	1	1	1	1
30	2009047779	20091113	11	13	2009	6	1745	17:45	609	9	18	36	2	953	1307	1	1	5	0	0	0	0	0	0	5000	1	1	77	2
31	2009045663	20091115	11	15	2009	1	2100	21:00	111	11	22	43	2	949	1311	7	1	5	0	0	0	0	0	0	5000	1	1	77	77
32	2009045543	20091118	11	18	2009	4	1605	16:05	409	9	17	33	1	943	1317	1	1	4	0	2	0	0	2	0	12000	2	2	1	6
33	2009048332	20091128	11	28	2009	7	1908	19:08	710	10	20	39	2	925	1335	7	1	5	0	0	0	0	0	0	4000	1	1	77	77
34	2009050294	20091203	12	3	2009	5	1545	15:45	508	8	16	32	1	919	1341	1	1	4	0	1	0	0	1	0	7000	1	2	2	8
35	2009048543	20091207	12	7	2009	2	15	00:15	201	1	1	1	2	914	1346	1	1	5	0	0	0	0	0	0	2000	1	2	2	8
36	2009048944	20091207	12	7	2009	2	422	04:22	203	3	5	9	2	914	1346	1	1	4	0	1	0	0	1	0	50000	1	1	2	3
37	2009050268	20091207	12	7	2009	2	430	04:30	203	3	5	10	2	914	1346	6	1	5	0	0	0	0	0	0	5000	1	1	2	8
38	2009050742	20091208	12	8	2009	3	511	05:11	303	3	6	11	2	913	1347	1	1	4	0	1	0	0	1	0	4000	1	1	2	8
39	2009050946	20091208	12	8	2009	3	645	06:45	304	4	7	14	2	913	1347	4	1	5	0	0	0	0	0	0	8000	1	1	2	8
40	2009050947	20091208	12	8	2009	3	755	07:55	304	4	8	16	1	913	1347	1	1	5	0	0	0	0	0	0	4000	2	2	2	8
41	2009050972	20091210	12	10	2009	5	1050	10:50	506	6	11	22	1	912	1348	1	1	5	0	0	0	0	0	0	5000	1	2	2	1
42	2009050928	20091211	12	11	2009	6	931	09:31	605	5	10	20	1	911	1349	3	1	4	0	1	0	0	1	0	500	1	1	1	2
43	2009052147	20091220	12	20	2009	1	945	09:45	105	5	10	20	1	908	1352	6	1	5	0	0	0	0	0	0	1500	1	1	2	8
44	2009052817	20091223	12	23	2009	4	1546	15:46	408	8	16	32	1	908	1352	1	1	5	0	0	0	0	0	0	9000	2	2	1	3
45	2009055331	20091230	12	30	2009	4	933	09:33	405	5	10	20	1	910	1350	5	1	5	0	0	0	0	0	0	1700	1	1	2	3
46	2009054824	20091230	12	30	2009	4	2049	20:49	411	11	21	42	2	910	1350	3	0	5	0	0	0	0	0	0	3500	1	1	2	8
47	2009032725	20090826	8	26	2009	4	2101	21:01	411	11	22	43	2	1323	937	1	1	5	0	0	0	0	0	0	11500	2	2	2	6
48	2009033407	20090827	8	27	2009	5	1333	13:33	507	7	14	28	1	1321	939	1	1	5	0	0	0	0	0	0	8000	2	2	1	3

Figure 3. 21: Part of the Final Dataset (Input)

3.4 Summary

In most of the transportation related studies, large amounts of data have to be processed and integrated in order the required datasets for the analysis to be created. A common problem is that the data come from different sources and can be in different format. In this thesis, four different types of data are considered for two Iowa corridors: Crash Data, Weather Data, Roadway and Traffic Data.

This chapter described the four different types of data and the procedure of integrating them in a final comprehensive dataset. The final dataset will be further analyzed using statistical techniques. The next chapter will present the statistical methods that will be applied to address the thesis’ objectives.

CHAPTER 4. METHODOLOGY

This chapter describes the statistical methodology that was adopted in this thesis.

4.1 Overview

A short review of the methodological approaches that have been applied to estimate crash frequency and severity in past studies associated was presented in Chapter 2. For a more comprehensive review of the used methodologies on the estimation of crash frequency the reader may refer to Lord & Mannering (2010).

The main objective of this thesis (as outlined in Chapter 1) is to investigate the effect of the prevailing weather conditions along with other crash attributes on the severity of crashes. As already mentioned in Chapter 2 (Section 2.3.2), discrete (ordered or unordered) data models have been widely used in assessing the impact of various factors on crash severity. As explained in that section, crash severity can be analyzed as a discrete variable that takes values based on the level of the severity outcome (i.e., fatal injury crash, major injury crash, etc.). Thus, discrete outcome models are suitable for examining crash severity. However, it should be noted that while ordered models are more appealing and predominant in the literature of crash severity because of the inherent ordered nature of severity, one should be cautious in their use since they do not have the flexibility to capture the effect of the explanatory variables on the interior category (i.e., that of minor injury crashes) probabilities. On the other hand, unordered discrete models do not pose any such restrictions and thus may be preferable (Washington et al., 2011).

Considering the above, two different types of unordered discrete outcome models were selected for the purpose of this thesis. Specifically, a binary probit model and a multinomial logit model (MNL) were estimated in order to investigate the effect of the weather conditions along with other crash attributes on the severity of crashes (injury/no-injury in the case of the binary model, and three injury severity outcome in the case of the MNL) on each of the two study corridors during the analysis period 2009-2011.

4.2 Discrete Outcome Models

In modeling (unordered) discrete outcomes it is convenient to define a linear function of covariates (i.e., explanatory variables) that affect specific discrete outcomes (as the various levels of crash severity in the current case).

Following Washington et al. (2011), let T_{in} be a linear function that determines the discrete outcome i (i.e., severity level) for the observation n (i.e., crash), such that:

$$T_{in} = \beta_i X_{in} + \varepsilon_{in} \qquad (4.1)$$

where: β_i = a vector of estimable parameters for outcome i

X_{in} = the vector of the observable characteristics (i.e., crash attributes) that determine the outcome of observation n

ε_{in} = an error term that accounts for any unobserved effects (e.g. omitted variables)

If now I denotes all possible discrete outcomes for observation n, the probability $P_n(i)$ of the observation n to have a specific discrete outcome i $(i \in I)$ is given by the equation:

$$P_n(i) = P(T_{in} \geq T_{In}) \ \forall \ I \neq i \tag{4.2}$$

Equation (4.2) can be modified by substituting Equation (4.1) as:

$$P_n(i) = P(\beta_i X_{in} - \beta_I X_{In} \geq \varepsilon_{In} - \varepsilon_{in}) \ \forall \ I \neq i \tag{4.3}$$

In the case of this thesis, $P_n(i)$ actually denotes the probability of a specific crash n to have a specific severity outcome i, based on the crash attributes X_{in} and their estimated parameters β_i.

4.2.1 Binary Probit Model

A model that considers only two discrete outcomes is called binary model. Furthermore, if the error term ε_{in} (see Equations 4.1 and 4.3) is assumed to be normally distributed, then the model is a binary probit model. In this case, Equation 4.3 can be written as:

$$P_n(0) = P(\beta_0 X_{0n} - \beta_1 X_{1n} \geq \varepsilon_{1n} - \varepsilon_{0n}) \tag{4.4}$$

where 0 and 1 denote the two outcomes.

In other words, Equation 4.4 estimates the probability of outcome 0 occurring for observation n, where ε_{0n} and ε_{1n} are normally distributed with mean equal to zero, variances σ_0^2 and σ_1^2 respectively, and covariance σ_{01}. However, since the subtraction of two normally distributed variates produces a normally distributed variate, it can be assumed that the difference ε_{0n} - ε_{1n} follows a normal distribution with mean zero and variance $\sigma^2 = \sigma_0^2 + \sigma_1^2 - \sigma_{01}$ (Washington et al., 2011).

Thus, the probability $P_n(0)$ is given by the formula:

$$P_n(0) = \Phi\left(\frac{\beta_0 X_{0n} - \beta_1 X_{1n}}{\sigma}\right) \tag{4.5}$$

where Φ( ) the standardized cumulative normal distribution.

The vector of parameters β is estimated by using maximum likelihood methods. For a binary probit model the log-likelihood is:

$$LL = \sum_{n=1}^{N}(\delta_{0n}\, LN\Phi\left(\frac{\beta_0 X_{0n} - \beta_1 X_{1n}}{\sigma}\right) - (1 - \delta_{0n})LN\Phi\left(\frac{\beta_0 X_{0n} - \beta_1 X_{1n}}{\sigma}\right)) \tag{4.6}$$

where δ_{0n} is equal to 1 if the observed discrete outcome for observation n is 0 and zero otherwise.

4.2.2 Multinomial Logit Model

In cases where more than two discrete outcomes are considered, multinomial models can be applied. Multinomial Logit Models (MNL) is a type of discrete models in which the errors in the equations associated with each discrete outcome follow a Gumbel distribution.

Following Washington et al. (2011), in a multinomial case the term $\beta_I X_{In} + \varepsilon_{In}$ in Equation 4.3 can be replaced with the highest value (maximum) of all other $\beta_I X_{In} \neq 1$. Thus, Equation 4.3 can be rewritten as:

$$P_n(i) = P\left(\beta_i X_{in} + \varepsilon_{in} \geq \max_{\forall\, I \neq i}(\beta_I X_{In} - \varepsilon_{In})\right) \tag{4.7}$$

If all ε_{In} are independently and identically Gumbel distributed random variates with modes ω_{In} and scale parameter (common) η, the maximum of $\beta_I X_{In} + \varepsilon_{In}$ follows a Gumbel distribution with mode:

$$\frac{1}{\eta} LN \sum EXP(\eta\, \beta_I X_{In}) \tag{4.8}$$

If ε'_n is a disturbance term associated with the maximum of all possible discrete outcomes ≠ i and has mode zero and scale parameter η then the maximum of all possible discrete outcomes ≠ i is associated with the parameter and covariate product

$$\beta' X'_n = \frac{1}{\eta} LN \sum EXP(\eta\, \beta_I X_{In}) \tag{4.9}$$

It can be then proven (see Johnson and Kotz (1970) as cited by Washington et al. (2011)) that Equation 4.9 can be written as:

$$P_n(i) = P(\beta' X'_n + \varepsilon'_n - \beta_i X_{in} + \varepsilon_{in} \leq 0\ ) \tag{4.10}$$

where $P_n(i)$ is the probability of observation n to have a specific discrete outcome i.

Since the difference between two independently distributed Gumbel variates with common scale parameter η is logistic distributed, Equation 4.10 can be written as:

$$P_n(i) = \frac{1}{1 + EXP[\eta(\beta' X'_n - \beta_i X_{in})]} \tag{4.11}$$

Substituting with Equation 4.9 and rearranging the terms Equation 4.11 gives:

$$P_n(i) = \frac{EXP[\eta(\beta_i X_{in})]}{EXP[\eta(\beta_i X_{in})] + EXP[LN \sum EXP(\eta\, \beta_I X_{In})\,]} \tag{4.12}$$

After all, the standard multinomial logit (MNL) formulation (where $\eta = 1$) is written as:

$$P_n(i) = \frac{EXP(\beta_i X_{in})}{\sum_{\forall I} EXP(\beta_I X_{In})} \tag{4.13}$$

For the estimation of the vector of parameters β, the log-likelihood function is now:

$$LL = \sum_{n=1}^{N} \left(\sum_{i=1}^{I} \delta_{in} \ \left[\beta_i X_{in} - LN \sum\nolimits_{\forall I} EXP(\beta_I X_{In}) \right] \right) \tag{4.14}$$

where I is the total number of outcomes and δ_{in} is equal to 1 if the observed discrete outcome for observation n is i and zero otherwise.

Finally, as Washington et al. (2011) argue the choice of Gumbel distribution has to do with computational convenience. In fact, Gumbel distribution is pretty similar to normal distribution.

4.2.3 Statistical Evaluation

The statistical significance of each parameter in a discrete outcome model (either a binary probit or a MNL model) can be assessed by considering a one-tailed *t*-test. This test examines if the estimated parameter β for a specific variable is significantly different from zero. The test statistic is assumed to follow a *t*-distribution and its value is:

$$t = \frac{\beta - 0}{S.E.(\beta)} \tag{4.15}$$

where $S.E.\ (\beta)$ is the standard error of the parameter.

It is noteworthy though that for the case of MNL models, the use of t-statistics is not precisely correct, since the MNL model is derived from a Gumbel distribution and not a normal. However, in practice the results of *t*-test can provide a reliable approximation of the true significance (Washington et al., 2011).

The overall fit of a discrete outcome model is usually assessed by the ρ^2 statistic (also known as McFadden ρ^2 statistic). The ρ^2 statistic is:

$$\rho^2 = 1 - \frac{LL(\beta)}{LL(0)} \tag{4.16}$$

where: $LL(\beta)$ = the log-likelihood at convergence with parameter vector β

$LL(0)$ = the initial log-likelihood with all parameters set to zero

Since a perfect model has likelihood function of one (thus $LL(\beta)=0$), a ρ^2 of one means that the model predicts the outcomes with certainty (i.e., a perfect model). In fact, ρ^2 takes values between zero and one, with closer to one values to be desirable (similarly to R^2 of the linear regression).

Finally, a disadvantage of ρ^2 is that improves as additional parameters are included in the model (Washington et al., 2011). For this reason, an adjusted (corrected) ρ^2 is used that takes into account the number K of the model's parameters:

$$adjusted - \rho^2 = 1 - \frac{LL(\beta) - K}{LL(0)} \tag{4.17}$$

4.2.4 Interpretation of Findings

4.2.4.1 Elasticity

Elasticity is used to assess the impact of a specific variable in the outcome probabilities. Following Washington et al. (2011), elasticity is computed from the partial derivative of each observation n (n subscripting omitted):

$$E_{x_{ki}}^{P(i)} = \frac{\partial P(i)}{\partial x_{ki}} \times \frac{x_{ki}}{P(i)} \tag{4.18}$$

where: $P(i)$ = the probability of outcome i

x_{ki} = the value of the variable k for outcome i

For a MNL model Equation 4.18 can be written as:

$$E_{x_{ki}}^{P(i)} = [1 - P(i)]\,\beta_{ki}x_{ki} \tag{4.19}$$

Elasticity values give the percent effect that a 1% change in x_{ki} (when x_{ki} is a continuous variable) has on the outcome probability $P(i)$. If the value of elasticity is less than one, then the variable is said to be inelastic. In that case, a 1% change in x_{ki} will result in less than 1% change in probability $P(i)$. On the other hand, a variable with an elasticity value greater than one is considered elastic. A 1% change of an elastic variable causes a greater than 1% change in the outcome probability.

The previous apply to cases of small changes in continuous variables and thus these elasticities are also called point elastictities. In cases of large changes (e.g. doubling of the value of the variable) non-negligible errors may occur.

When considering an indicator variable (i.e., a variable that takes on the values 0 or 1) a pseudoelasticity can be computed (Washington et al. 2011):

$$E_{x_{ki}}^{P(i)} = \frac{\mathrm{EXP}[\Delta(\beta_i X_i)] \sum_{\forall I} EXP(\beta_{kI} X_{kI})}{\mathrm{EXP}[\Delta(\beta_i X_i)] \sum_{\forall I_n} EXP(\beta_{kI} X_{kI}) + \sum_{\forall I \neq I_n} EXP(\beta_{kI} X_{kI})} - 1 \qquad (4.20)$$

where: I_n = is the set of alternate outcomes with x_{ki} in the function determining the outcome

I = the set of all possible outcomes

The interpretation of the pseudoelasticity value is similar to that of elasticity, namely it is the percent effect that a change from zero to one in the indicator variable has on the outcome probability *P(i)*.

Finally, the aforementioned elasticities are direct elasticities, since they capture the impact that a change in a variable which determines the likelihood of an outcome *i* has on the probability of the same outcome *i*. If the effect on the probability of another outcome different than *i* is of interest (let *j* be that outcome), then a cross elasticity should be computed:

$$E_{x_{ki}}^{P(j)} = -P(i)\, \beta_{ki} x_{ki} \qquad (4.21)$$

It should be noted that in case of a variable that is included in more than one utility functions, then the net effect of the variable can be determined by considering both direct and cross elasticities that are estimated for the variable of interest.

4.2.4.2 Marginal Effects

Another way to assess the impact of a specific variable on the outcome probabilities is by estimating marginal effects. Marginal effects measure the effect of a “one unit” change in x_{ki} (when x_{ki} is a continuous variable) has on the outcome probability $P(i)$, as shown in Equation 4.22. Marginal effects are easier in the interpretation, especially for indicator or integer variables (Washington et al. 2011). For instance, in case of an indicator variable, marginal effects measure the impact on the outcome probabilities of a change in the variable’s value from zero to one.

The marginal effect on the probability of an outcome *i* is given by the formula:

$$ME_{x_{ki}}^{P(i)} = \frac{\partial P(i)}{\partial x_{ki}} \tag{4.22}$$

4.3 Summary

This chapter presented the mathematical background of the statistical methodology (discrete outcome models) that was used for analyzing the data collected for this thesis. The estimation results and major findings along with their interpretation will be provided in the next chapter.

CHAPTER 5. STATISTICAL DATA ANALYSIS

This chapter presents the estimation results of the two types of crash severity models (binary probit and multinomial logit) that were developed for each study corridor. The discussion includes a description (and review of the theoretical background) of the different types of variables that were utilized in this thesis, the statistical evaluation of each model, and the interpretation of findings.

5.1 Types of Variables

Following Miller and Miller (2004) as a random variable (or simple variable) X is called a real-valued function defined over the elements of a sample space S. Under a more tangible perspective, one could say that a variable is actually an action of measuring, recording or computing a number. For instance, if variable X corresponds to the 1-hr average temperature prior to a crash that happened on the I-80 corridor of this study, it could be stated that:

X = the act or recording (or more precisely computing) the 1-hr average temperature prior to a crash.

All the possible numbers that this act can produce (i.e., all the possible values that this variable can take) constitute the sample space of X and are denoted with S_X. For instance, regarding the case of temperature the sample space could include any real number or $S_X = (-\infty, +\infty)$. In fact, it is intuitive that the temperature under normal conditions cannot be greater (or lower) than a specific value, e.g. greater than 50 ^oC (or lower than -40 ^oC).

5.1.1 Continuous Variables

In statistics, a variable is considered continuous when it can take on any value within a range of values (Washington et al., 2011). For instance, the 1-hr average temperature prior to a crash of the previous example, the 1-hr average wind speed prior to a crash, or the adjusted Truck ADT of the road segment where the crash occurred, are continuous variables. Most of the variables associated with the prevailing weather conditions during a crash (i.e., temperature, precipitation amount, wind speed, etc.) and were used in this thesis are continuous variables. For describing continuous variables statistics as the mean, the standard deviation, minimum, maximum etc. are commonly used.

5.1.2 Discrete Variables

A discrete variable is a variable that can take only on specific integer values or else have a discrete sample space (Miller & Miller, 2004). In other words, the act associated with a discrete variable can produce only integer values (i.e., -1, 0, 1, 2, …).

In this thesis, discrete variables constitute the majority of the used variables. For instance if: Y is the act or recording the severity level of a crash, then Y is a discrete random variable, then its sample space could be $S_Y = \{1, 2, 3, 4, 5\}$, where:

1: corresponds to a fatal injury crash

2: corresponds to a major injury crash

3: corresponds to a minor injury crash

4: corresponds to a possible/unknown injury crash

5: corresponds to a property damage only (PDO) crash

Moreover, variable Y is also called ordered variable since its possible values reveal an order. For instance, as the value of Y increases, the severity level of the crash is lower.

Another type of discrete variable that was used in this thesis is unordered variables. For instance if Z is associated with the act of recording the manner of crash/collision, then it could be $S_Z = \{1, 2, 3, 4\}$, where:

1: corresponds to a non-collision crash (i.e., overturn/rollover)

2: corresponds to a head-on collision

3: corresponds to a rear-end collision, etc.

In this case, the values of the sample space are not associated with any order, but they just denote a specific attribute (e.g., the manner of collision). Finally, these variables are also called categorical since their values are associated with a specific category (i.e., a rear-end crash).

It might not be meaningful to compute the mean or standard deviation for discrete variables (especially in case of an unordered variable). Instead the frequency or relative frequency (i.e., percentage) of each value can be used in describing a discrete variable.

5.1.3 Count Variables

This category includes the variables which are produced by a counting action. For instance, as a count variable could be considered the act of recording the number of vehicles that were involved in a crash. Certainly, a count variable can take only non-negative integer values, thus its sample space should be S = {0, 1, 2, …}. In this thesis, count variables were the number of vehicles that were involved in a crash or the ADT of the road segment in vehicles per day (note though that the adjusted ADT based on the month is a continuous variable and not a count). Although mean or standard deviation can be computed, they might not always be meaningful for a count variable, thus frequencies can be also used for the description of this type of variables.

5.1.4 Indicator Variables

An indicator (also known as dummy) variable is a binary variable since it can take on the values 0 or 1, thus its sample space is S = {0, 1}. Indicator variables are usually associated with the presence of a specific condition or event, or the validity of a statement. In other words, an indicator variable could represent a response as Yes/No or True/False.

Let consider a variable named "PDO" which is associated with whether a crash had a property damage only outcome. In that case, if 1: accounts for a PDO crash and 0: for an injury crash, then if "PDO=1", the crash of interest had a property damage only outcome. The sample space of that variable is $S_{PDO} = \{0, 1\}$.

All the types of the aforementioned variables (continuous, discrete, count) can produce indicator variables. For instance, an indicator variable could be associated with whether the 1-hr average temperature prior to a crash was lower than $0^{o}C$ (the 1-hr average temperature is a continuous variable) or whether the type of the crash was rear-end (type of crash is a discrete variable) or whether two or more vehicles were involved in the crash (the number of vehicles is a count variable).

5.1.5 Interaction Variables

The main objective of this thesis was to investigate the interaction effects of weather conditions and other crash parameters on crash severity. For this reason, interaction variables were created.

An interaction variable is based on the combination of two or more variables. This combination is usually made by the multiplication of the values of the variables of interest. For instance, if it is of interest whether a crash occurred under snowfall and dark lighting conditions, the corresponding interaction variable can be created by multiplying the (indicator) variable associated with snowfall by the (indicator) variable associated with dark lighting conditions. In that case, if the resulting interaction variable takes the value of 1, then the specific crash occurred under snowfall and dark lighting conditions. It should be noted that in the previous case both indicator variables should take the value of 1 (i.e., both conditions should be true). Actually, one could say that interaction variables are associated with joint events.

An indicator variable can be created by the combination of any type of variables. For instance, an indicator variable (e.g., rural road) could be multiplied by a continuous variable (e.g., 1-hr precipitation amount) and give an interaction variable associated with the 1-hr precipitation amount that had fallen prior to crash which occurred on a rural road. Thus, an interaction variable could be either binary or continuous or even discrete.

The main variables that were used in this thesis along with their sample spaces and main descriptive statistics are listed in Appendix B.

5.2 Estimation Results

Two different types of discrete outcome models were developed during the statistical analysis of data. Specifically, a binary probit and a multinomial logit model were estimated for each of the corridors of study. In both models, different levels of crash severity were considered as discrete outcomes. Moreover, various types of variables (see Section 5.1) related to the prevailing weather conditions (for a period of an hour prior to each crash) and other crash attributes were used as explanatory variables. All models were estimated using the statistical software NLOGIT 4.0. It should also be noted that the models were estimated based on the complete set of observations. In other words, any observations with missing data (i.e., weather variables or crash attributes) were skipped.

The following present the estimation results and the interpretation of the key findings for each model. For the exact outputs of the software the reader may refer to Appendix C.

5.2.1 I-80 Corridor

5.2.1.1 Binary Probit Model

A binary probit model for the I-80 corridor was estimated first. That model considered two discrete outcomes: 0: if the crash had a Property Damage Only (PDO) outcome and 1: if the crash had an injury outcome (i.e., fatal, major, minor or possible/unknown injury). It should be noted that the PDO crashes constituted the 78.3% of the sample, while 21.7% of crashes were injury crashes (see also Table 3.3).

Table 5.1 presents the estimation results of the binary probit model for crash severity on I-80 corridor.

Table 5. 1: Binary Probit Model Estimation Results for Crash Severity on I-80 Corridor

Variable	**Variable Type**	**Estimated Parameter**	**t-statistic**	**Marginal Effects**
Constant		-0.483	-4.425	
Overturn/Rollover	Indicator	0.880	5.610	0.291
Driving Too Fast for Conditions and Wind of Non-Parallel Direction to the Direction of Vehicle Movement	Indicator	0.338	2.301	0.099
Crash during a Snowfall Event in December	Indicator	0.323	1.624	0.096
February	Indicator	0.317	1.782	0.094
Single Vehicle Crash under Temperature below 0°C	Indicator	-0.568	-3.954	-0.136
Crash during a Rainfall Event between 5:00 pm to 10:00 pm	Indicator	-0.801	-2.156	-0.145
Snowfall Event and Wind of Cross Direction to the Direction of Vehicle Movement	Indicator	-0.454	-1.737	-0.099
Collision with Animal	Indicator	-0.992	-5.127	-0.194
Average Wind Speed	Continuous	-0.015	-2.578	-0.004
Number of Observations		795		
Log-Likelihood at convergence		-368.287		
Restricted Log-likelihood		-415.190		
adj-ρ^2		0.089		

The binary probit model included one continuous variable and eight indicator variables (six of those were interaction variables). All the variables were significant at the 95% confidence level apart from the indicator variable related to a crash during a snowfall event, which was significant at the 90% confidence level (as indicated by the values of the t-statistic). Variables with a positive parameter (coefficient) increase the probability of an injury outcome, while those with a negative coefficient increase the probability of a PDO outcome (or equivalently decrease the probability of an injury outcome).

Overall, four of the explanatory variables were found to increase the probability of an injury outcome (i.e., had positive parameters). Specifically, the probability of an injury outcome increases (by 0.291 according to the marginal effects) in crashes involving an overturn or rollover, regardless of the prevailing weather conditions. This finding makes an intuitive sense by taking into account the intensity of such collisions. A crash whose reported major cause was driving too fast for conditions and also occurred under wind direction non-parallel to the vehicle movement may lead to an injury outcome. This finding could be attributed to the effect which may have the combination of the vehicle speed with the wind direction on the aerodynamic resistance and consequently, to the manner of crash (i.e., sideswipe, overturn etc.). However, more investigation is necessary in order to draw any safe conclusion. Finally, a crash under snowfall conditions that occurred in December and any crash occurred in February (regardless of the weather conditions) seem to lead to injury outcomes as well. These findings are likely picking up the adverse weather conditions during winter on road safety.

On the other hand, five variables were found to increase the probability of a PDO outcome (i.e., had negative parameters). Three of them were interaction variables associated with adverse weather conditions (e.g., temperature below 0^oC, rainfall or snowfall events). This finding is in line with past literature (cite some studies) and can be attributed to the increased alertness of the drivers during adverse weather conditions. A collision with animal has an increased probability (by 0.194) of resulting in a PDO outcome. This can make intuitive sense if one considers that animals are sat a disadvantage when colliding with vehicles. Finally, an interesting finding is that as the average wind speed increases the probability of a PDO crash increases as well. This is the only continuous variable of the model and has an elasticity equal to - 0.18. Thus a 1% increase in wind speed leads to a 0.18% decrease in the probability of an injury crash (or equivalently 0.18% increase in the probability of a PDO crash). This effect may be attributed to the aerodynamics resistance (similarly with the effect of the wind direction).

5.2.1.2 Multinomial Logit Model

Three different levels of crash severity were considered in the multinomial logit (MNL) model: property damage only (PDO), possible/unknown injury and fatal/major/minor injury. The advantage of the MNL model is that it can consider more than two outcomes. Thus, more individual levels of severity can be investigated. In this case, three levels were selected based on the distribution of crash severity. Specifically, 78.3% of the observed crashes were PDOs, 10.6% were Possible/Unknown injuries and the 11.1% were Fatal/Major/Minor Injuries (see Table 3.3).

After the final model was specified a comprehensive analysis and interpretation of findings were performed by considering the signs of the parameters and the elasticities of the variables.

Table 5.2 presents the estimation results of the MNL model for crash severity on I-80 corridor, while Table 5.3 presents the values of the elasticity for each variable included in the model.

Table 5. 2: Multinomial Logit Model Estimation Results for Crash Severity on I-80 Corridor

Variable	**Variable Type**	**Estimated Parameter**	**t-statistic**
PDO Function			
Single Vehicle Crash under Temperature below 0^oC	Indicator	1.288	3.622
Average Wind Speed (Km/hr)	Continuous	0.023	2.218

Table 5.2 (continued)

Variable	**Variable Type**	**Estimated Parameter**	**t-statistic**
Snowfall Event and Wind of Cross Direction to the Direction of Vehicle Movement	Indicator	0.823	1.586
Crash during Rainfall between 5:00 pm to 10:00 pm	Indicator	1.539	2.028
Possible/Unknown Injury Function			
Constant		-1.581	-7.349
Collision with Animal	Indicator	-2.316	-3.174
Single Vehicle Crash under Temperature below 0°C	Indicator	1.005	2.366
Overturn/Rollover	Indicator	1.312	4.002
Driving Too Fast for Conditions and under Wind Speed between 13.9 and 24.5 Km/hr	Indicator	0.662	1.700
Fatal/Major/Minor Injury Function			
Constant		-1.568	-7.574
Overturn/Rollover	Indicator	1.744	5.035
Rural Road and 1-hr Precipitation Amount (in mm)	Continuous	-0.069	-1.465
Driving Too Fast for Conditions and Wind of Non-Parallel Direction to the Direction of Vehicle Movement	Indicator	1.318	4.185
November	Indicator	-1.105	-1.804
October	Indicator	-1.164	-1.576
Number of Observations		770	
Log-Likelihood at convergence		-467.744	
Restricted Log-likelihood		-516.326	
adj-ρ^2		0.085	

Table 5. 3: Direct and Cross Elasticities of the Variables included in the MNL Model for I-80 Corridor

Variable	**Elasticity (%)**		
	PDO	**Possible/Unknown Injury**	**Fatal/Major/Minor Injury**
PDO Function			
Single Vehicle Crash under Temperature below 0°C	31.49*	-63.73	-63.73

Table 5. 3 (continued)

Variable	Elasticity (%)		
	PDO	**Possible/Unknown Injury**	**Fatal/Major/Minor Injury**
Average Wind Speed (Km/hr)	0.27*	-0.07	-0.07
Snowfall Event and Wind of Cross Direction to the Direction of Vehicle Movement	15.11*	-49.45	-49.45
Crash during Rainfall between 5:00 pm to 10:00 pm	23.43*	-73.51	-73.51
Possible/Unknown Injury Function			
Collision with Animal	13.08	-88.84*	13.08
Single Vehicle Crash under Temperature below 0^{o}C	-11.74	141.11*	-11.74
Overturn/Rollover	-17.70	205.61*	-17.70
Driving Too Fast for Conditions and under Wind Speed between 13.9 and 24.5 Km/hr	-8.22	77.92*	-8.22
Fatal/Major/Minor Injury Function			
Overturn/Rollover	-27.11	-27.11	316.92*
Rural Road and 1-hr Precipitation Amount (in mm)	0.01	0.01	-0.11*
Driving Too Fast for Conditions and Wind of Non-Parallel Direction to the Direction of Vehicle Movement	-18.64	-18.64	203.96*
November	9.07	9.07	-63.88*
October	9.22	9.22	-65.90*

* Direct Elasticity

Three (utility) functions (associated with each possible outcome) of the form of Equation 4.1 were estimated in this model. Eventually, the function associated with the PDO outcome included four explanatory variables (three indicator variables and a continuous one). The function associated with the possible/unknown injury outcome included four indicator variables, while the function of the fatal/major/minor injury outcome included five variables (four indicator and a continuous one). Similarly to binary probit models, a positive sign of a parameter indicates an increase in the probability of observing the outcome whose function includes that variable, while a negative sign indicates a decrease in that probability. However, apart from the signs one should also consider any net effects on the probabilities in cases of variables that are included in more than one function.

According to Table 5.2, the variable related to a single vehicle crash occurred under temperature below 0^oC was significant for the PDO and possible/unknown injury outcomes. The positive parameter shows that, while temperature below 0^oC may pose a driving hazard, the severity outcome of a single vehicle crash occurred under those conditions is not very severe. This could be attributed to the alertness of the driver and the lower vehicle speed due to the adverse conditions (such as snow or ice on the road) that are associated with such low temperatures. Finally, this variable has inelastic effects on the probability of the various severity levels, since a change on its value from 0 to1 leads to changes in the probabilities of the severity outcomes lower than 100% (see Table 3.3).

The effect of wind on road safety has been established in a number of papers (Baker & Reynolds, 1992; Edwards, 1994; SWOV, 2012). Higher wind speeds constitute a driving hazard especially for specific types of vehicles such as trucks, buses, or motorcycles. The effect of wind should be mainly associated with the aerodynamics of the vehicle movement. Furthermore, Usman et al. (2012a) argued that severe winds can magnify the adverse weather conditions, such as a snowstorm. In the MNL model (as also in the binary probit) the effect of wind speed was found to be significant. Specifically, higher wind speeds were found to increase the probability of a PDO crash. However, this effect was inelastic, since an increase in the average wind speed by 1% leads to a 0.27% increase in the probability of a PDO outcome and 0.07% decrease in the probability of injury outcomes. Moreover, vehicle speed in combination with wind speed may contribute to the severity outcome of a crash. Specifically, a crash that occurred under conditions of high wind speed and its cause (as stated in crash report) was driving too fast for conditions was found to result in a possible/unknown injury outcome (with a 77.92 % increase in the corresponding probability).

Apart from speed, the direction of wind seems to be significant as well. A number of studies have been published on the effect of wind direction (especially related to cross winds) on the behavior of vehicles (Coleman & Baker, 1990; Baker, 1993). As also in the binary probit model, winds of cross direction during a snowfall event were found to increase the probability of a PDO crash. For the case of I-80, a wind of direction N-S (or S-N) was considered to be a wind of cross direction, since the major direction of the corridor is E-W (or W-E). This finding can be attributed to the negative effect of the wind direction on the movement of vehicle; for example, cross winds (especially sudden gusts) may cause deviations from the direction of movement. Furthermore, cross winds during precipitation events (such as snowfall) may affect visibility. The effect of this variable is inelastic since a change in the variable’s value leads to changes in the probabilities lower than 100%.

Nevertheless, a crash with reported cause driving too fast for conditions in combination with wind of non-parallel direction to the direction of movement (i.e., of any direction different from E-W or W-E) is more likely to have a fatal/major/minor injury outcome. It is also noteworthy that the effect of this variable is elastic since the corresponding elasticity is equal to 203.96. This finding could be attributed to the type (and the intensity) of the harmful event (such just an overturn/rollover or a sideswipe) that these conditions may cause.

Rainfall events in combination with time of the day may contribute to the severity outcome of a crash. Specifically, a crash that occurred during an evening rainfall event (between 5:00 pm and 10:00 pm) is more likely to be a PDO crash (with a 23.43% increase in probability). As cited in the literature, rainfall events during night constitute a driving hazard, since the reflection of lights on the accumulated water makes the detection of the road and the objects in near vicinity more difficult (Brodsky & Hakkert, 1988). However, even in this case the severity outcome of a crash under rainfall during evening hours seems to be less severe as with most cases of a crash during inclement weather. In addition, an interesting finding was that as the 1-hr precipitation amount on a rural road increases, the probability of a fatal/major/minor injury outcome decreases. Specifically, for an increase in the precipitation amount by 1% on a rural road the probability of a fatal/major/minor injury crash decreases by 0.11%. This finding can make an intuitive sense if the adjustment of speed and the alertness of driver are taken into consideration.

In the binary probit model, a collision with animal was found to reduce the probability of an injury crash, regardless of the weather conditions. In the MNL model, the variable associated with a collision with animal was found to reduce the probability of a possible/unknown injury outcome. Specifically, a collision with animal reduces the probability of a possible/unknown injury outcome by 88.84%. An obvious explanation is that animals are normally more vulnerable than vehicles.

An event of an overturn or rollover is associated with injury outcomes, regardless of the weather conditions. Specifically, the corresponding variable has positive parameters in the functions of possible/unknown injury and fatal/major/minor injury outcomes. It is also noteworthy that the effect of this variable is elastic with changes in the outcome probabilities greater than 100%. This probably has to do with the intensity of this type of events (i.e., overturns). However, more investigation on adequate samples of crashes associated with this type of event in combination with the prevailing weather conditions is recommended.

Finally, weather conditions are associated with the month of the year. Thus, indicator variables related to months were considered in this model. Specifically, October and November were found to reduce the probability of a crash with a highly severe outcome. This finding could be attributed to the fact that the weather conditions change during that period (e.g. temperature drops, snowfalls start) and the duration of the day (and thus of the daylight) is getting shorter. The aforementioned changes probably make the drivers more careful, thus crashes of high severity are less likely. Also, it should be mentioned that road users who are more vulnerable to severe crashes, such as motorcyclists or pedestrians, make fewer trips during that period of year.

In conclusion, both the binary probit and multinomial logit models for crash severity on I-80 corridor lead to similar findings. However, the multinomial logit model maybe preferable for the analysis of crash severity because it provides with the flexibility to consider more than two severity outcomes. Thus, a multinomial logit model can better investigate the effect of specific factors on the interior categories of severity (Washington et al., 2011), as the possible/unknown injury outcome in this thesis. This advantage outweighs binary probit models which can consider only two outcomes (e.g. injury / no injury).

Overall, the major finding is that adverse weather conditions such as temperature below 0^{o}C, rainfall and snowfall events were found to be associated with crashes of lower severity, such as PDO. This finding is in alignment with existing literature. Furthermore, wind speed and direction was found to play a role in the severity outcome of a crash, especially when the cause of the crash was reported to be due to inappropriate vehicle speed (driving too fast for conditions). Moreover, variables associated with specific types of crashes such as overturn/rollover or collision with animal were found to be significant regardless of the weather conditions under they occurred. Finally, specific months (October and November) were found to lead to less severe crash probably because of the change in road users' habits.

5.2.2 US-34 Corridor

The corresponding models for the US-34 corridor were estimated following exactly the same process as in the estimation of the models for the I-80 corridor. The results of the estimation process and the key findings are presented in the following.

5.2.2.1 Binary Probit Model

The binary probit model was estimated before the MNL model (as in the case of I-80 corridor). On the US-34 corridor, PDO crashes constituted the 70.7% of total crashes, while injury crashes accounted for the 29.3% (see Table 3.4).

Table 5.4 presents the results of the binary probit model for crash severity on US-34 corridor.

Table 5. 4: Binary Probit Model Estimation Results for Crash Severity on US-34 Corridor

Variable	**Variable Type**	**Estimated Parameter**	**t-statistic**	**Marginal Effects**
Constant		-1.463	-4.848	
Crash during a Snowfall Event and Dark Lighting Conditions	Indicator	-1.024	-1.744	-0.249

Table 5.4 (continued)

Variable	Variable Type	Estimated Parameter	t-statistic	Marginal Effects
Crash with 2 or more Vehicles Involved under Temperature below 0^{o}C	Indicator	0.652	2.313	0.242
Rural Road and Average Wind Speed (Km/hr)	Continuous	0.025	2.610	0.009
Wind of Non-Parallel Direction to the Direction of Vehicle Movement	Indicator	0.537	1.811	0.166
Crash between 6:00 am and 9:00 am	Indicator	0.729	2.419	0.275
August	Indicator	1.566	3.719	0.562
Number of Observations		172		
Log-Likelihood at convergence		-91.002		
Restricted Log-likelihood		-107.791		
adj-ρ^2		0.091		

The binary probit model of the US-34 corridor included six variables. Five of those were indicator and one was continuous. All the parameters were significant at the 95% confidence level. From those variables, only one was found to decrease the probability of an injury outcome. Specifically, a crash during a snowfall event and under dark lighting conditions is more likely to have a PDO outcome. Moreover, based on the marginal effects the decrease in the probability of an injury crash under these conditions is equal to 0.249.

The aforementioned variable was the only variable related with adverse weather conditions that was found to reduce crash severity. All the other variables associated with inclement weather were found to increase the probability of injury outcomes. Specifically, crashes with two or more vehicles involved under temperature below 0^{o}C were found to lead to injury outcomes, with an increase of 0.242 in the corresponding probability. In addition, an increase in the 1-hr average wind speed in combination with rural environment was found to increase the probability of an injury crash. That effect though is inelastic (elasticity = 0.14).

Moreover, wind direction seems to influence crash severity. Specifically, winds of direction non-parallel to the direction of vehicle movement were found to lead to severe outcomes, with an increase of 0166 in the corresponding probability. Finally, crashes that occurred during morning hours and crashes that occurred in August were found to have injury outcomes.

5.2.2.2 Multinomial Logit Model

The estimation of the binary probit model was succeeded by the estimation of the multinomial logit (MNL) Model. As in the case of I-80 corridor, the purpose of the estimation of the MNL model was the investigation of more than two severity outcomes. Specifically, the same three severity outcomes with I-80 were considered: property damage only (PDO), possible/unknown injury and fatal/major/minor injury. As shown in Table 3.4, 70.7% of the observed crashes had a PDO outcome, 17.8% had a possible/unknown outcome and 11.5% had fatal/major/minor injury outcome.

Table 5.5 presents the estimation results of the MNL model for crash severity on US-34 corridor, while Table 5.6 presents the values of the elasticity for each variable included in the model.

Table 5. 5: Multinomial Logit Model Estimation Results for Crash Severity on US-34 Corridor

Variable	**Variable Type**	**Estimated Parameter**	**t-statistic**
PDO Function			
August	Indicator	-2.377	-3.573
Crash between 6:00 am and 9:00 am	Indicator	-1.314	-2.599
Rural Road and Average Wind Speed (Km/hr)	Continuous	-0.029	-1.836

Table 5. 5 (continued)

Variable	**Variable Type**	**Estimated Parameter**	**t-statistic**
Possible/Unknown Injury Function			
Constant		-2.300	-7.270
Crash under Temperature below 0^{o}C	Indicator	0.913	2.309
Fatal/Major/Minor Injury Function			
Constant		-2.796	-6.836
Crash under Temperature below 0^{o}C	Indicator	0.913	2.309
Logarithm of Truck-AADT of the road segment where crash occurred and Wind of Non-Parallel Direction to the Direction of Vehicle Movement	Continuous	0.205	2.617
Crash under Dark Lighting Conditions	Indicator	-1.864	-2.370
Number of Observations		181	
Log-Likelihood at convergence		-128.348	
Restricted Log-likelihood		-148.169	
adj-ρ^2		0.114	

Table 5. 6: Multinomial Logit Model Estimation Results for Crash Severity on US-34 Corridor

Variable	**Elasticity on**		
	PDO	**Possible/Unknown Injury**	**Fatal/Major/Minor Injury**
PDO Function			
August	-68.12*	243.40	243.40
Crash between 6:00 am and 9:00 am	-45.89*	101.33	101.33
Rural Road and Average Wind Speed (Km/hr)	-0.11*	0.16	0.16
Possible/Unknown Injury Function			
Crash under Temperature below 0^{o}C	-16.91	107.05*	-16.91
Fatal/Major/Minor Injury Function			
Crash under Temperature below 0^{o}C	-11.67	-11.67	120.10*

Table 5.6 (continued)

Variable	**Elasticity on**		
	PDO	**Possible/Unknown Injury**	**Fatal/Major/Minor Injury**
Logarithm of Truck-AADT of the road segment where crash occurred and Wind of Non-Parallel Direction to the Direction of Vehicle Movement	-0.08	-0.08	0.27*
Crash under Dark Lighting Conditions	18.19	18.19	-81.67*

* Direct Elasticity

The results of the multinomial logit (MNL) model were almost similar to the results of the binary probit model. The main differences were: the consideration of all the crashes that occurred under temperature below 0^oC (and not only multivehicle crashes as in the binary probit model), the interaction variable of Truck ADT with the wind of non-parallel direction to the direction of vehicle movement that was found significant for the fatal/major/minor injury utility function and finally the finding that a crash under dark lighting conditions is less likely to have a fatal/major/minor injury outcome.

In the MNL model any crash that occurred under temperature below 0^oC was more likely to be of an injury outcome. Moreover, the effect of this variable to the fatal/major/minor injury outcome was found elastic with a net elasticity (direct and cross elasticity combined) greater than 100%. On the other hand, the net elasticity of the possible/unknown injury outcome was less than 100% (inelastic effect). This finding is contradictory to the finding of I-80 corridor regarding the temperature. Nevertheless, it should be noted that in the case of the I-80 corridor only single vehicle crashes were considered in the models. In the case of US-34 corridor though, any crash was considered in the MNL model. However, the majority of crashes on US-34 corridor (around 60%) were multivehicle crashes. Thus, one could speculate that a multivehicle crash under temperature below 0^oC on an undivided facility (as US-34) is more likely to be of high severity. This could be attributed to the absence of median and the higher possibility of crashes between vehicles travelled on different directions. Especially, in cases of temperature below 0^oC where the possibility of presence of snow or ice on the road surface is high, such kind of crashes (e.g. head-on crashes) mainly caused by the loss of vehicle control are expected to be higher severity. Furthermore, speculations about the efficiency of (different) maintenance policies that are applied on routes of different classification (i.e., more emphasis on interstate routes) could also be made.

The Truck-AADT of the road segment in combination with winds of non-parallel direction was found to have a negative effect on crash severity. Specifically, as the logarithm of the Truck-AADT increases the probability of a Fatal/Major/Minor outcome increases as well (elasticity 0.27). This finding could be attributed to the negative effect that wind direction (and especially cross winds) may have on the safety of large vehicles. Also that type of winds may cause specific types of crashes (e.g. overturns) which usually have more severe outcomes.

Finally, crashes under dark lighting conditions tend to be of lower severity. This could be explained by the increased alertness of the drivers, especially in cases of no lighted road segments.

The rest of the variables of the MNL model were also included in the binary probit model. The most interesting finding is that the increase of wind speed on a rural road reduces the probability of a PDO outcome (elasticity = -0.11). This finding is contradictory to the finding about wind speed of the I-80 corridor. Moreover, crashes occurred in August and crashes in the morning hours were found to have a lower probability of a PDO outcome. It is also noteworthy that both variables had elastic effects, with the variable related to August to have an elasticity equal to 243.40.

Overall, especially for the case of US-34 corridor, both models gave similar results. Thus, one could argue that the simpler binary probit model could suffice. However, it should be noted that apart from the flexibility of the investigation of more than two outcomes, the MNL model had also a better fit (i.e., larger adjusted-ρ^2) in the case of US-34 corridor. Finally, it should be noted that the models of US-34 were estimated on smaller sample than the models of I-80. Thus, one could say that the results of US-34 may be less reliable.

5.3 Summary

This chapter presented the results of the statistical data analysis of this thesis. First, the background of the different types of variables that were used in the analysis was provided. Then, the results of the discrete outcome models that were estimated for each corridor of study were discussed.

In general, both types of models gave similar results for the same corridor. As far as the I-80 corridor is concerned, adverse weather conditions were found to increase the probability of low severity crashes. Furthermore, wind (speed and direction) was proven as a significant weather factor that has multiple (interaction) effects on crash severity. On the other hand, the results of the US-34 corridor were somewhat contradictory to those of the I-

80 corridor. However, the number of observations on US-34 was smaller than that of I-80, thus the results of US-34 corridor are less reliable.

Based on the aforementioned results, conclusions and recommendations about the implication of the study’s findings and future research were made. Those conclusions and recommendations along with the limitations of this study will be discussed in the following chapter.

CHAPTER 6. CONCLUSIONS, LIMITATIONS, AND RECOMMENDATIONS

6.1 Summary

This thesis investigated the interaction effects of weather conditions and other crash-specific factors on crash severity. For this purpose, crashes that occurred on two different Iowa corridors of similar lengths were selected. The two corridors represented a four-lane divided facility (I-80) and a two-lane undivided facility (US-34). The analysis period covered the years from 2009 to 2011. The selection was based on the number of crashes that occurred on those corridors in order to constitute an adequate analysis sample and also their proximity to Road Weather Information Systems (RWIS) for obtaining weather-related data.

Four different types of data were utilized in the analysis: crash data (from the Iowa DOT crash datasets), weather data (from RWIS), roadway data (from the Iowa DOT GIMS files) and traffic data (from ATRs). The data was processed and integrated in a unique dataset for each corridor that was the input for the statistical analysis. Furthermore, interaction variables were created based on that data. Those variables related crash-specific factors to the prevailing weather conditions for a period of one hour prior to each crash. Discrete outcome models (a binary probit model and a multinomial logit model) were estimated to investigate the interaction effects of weather conditions and other crash- specific factors on different levels of crash severity for each of the two corridors. The estimation results of those models were analyzed and interpreted and finally conclusions were drawn.

6.2 Key Findings

The following summarize the major findings of this thesis:

I-80 Corridor

Adverse weather conditions influence crash severity, especially due to their interaction with other crash variables. In most of the cases though, the outcome of a crash under inclement weather is more likely to be of low severity. For instance, single vehicle crashes under temperature below 0^{o}C are more like to result in a PDO or possible/unknown injury outcome. Furthermore, precipitation events, especially when they are associated with impaired visibility, may lead to less severe crashes. Specifically, crashes under snowfall conditions concurrently with wind of cross-direction or crashes during rainfall events at night (or late afternoon) were found to lead to a PDO outcome. Finally, an increase in the precipitation amount on a rural road is associated with a decrease in the probability of fatal/major/minor injury outcomes. The previous findings are also supported by existing literature and can be attributed mainly to the increased alertness of the drivers and the adjustment of vehicle speed during adverse weather conditions.

A very interesting finding is the effect of wind on crash severity. Both wind speed and direction, especially when they interact with other factors seem to affect crash severity. As the 1-hr average wind speed increases the probability of a PDO outcome increases as well. However, a crash whose reported cause was inappropriate speed (i.e., driving too fast for conditions) in combination with high wind speed is more likely to result in a possible/unknown injury outcome.

Turning to wind direction, winds of non-parallel direction to direction of vehicle’s movement were found to interact with other weather conditions such as snowfall, or other crash factors such as vehicle speed. Specifically, winds of cross direction during snowfall may increase the probability of a PDO outcome, while winds of non-parallel direction in combination with inappropriate vehicle speed may lead to fatal/major/minor injury outcomes.

Finally, specific types of crash were found to affect crash severity regardless of the prevailing weather conditions. For instance, a crash resulted in an overturn or rollover is more likely to be of high severity (i.e., possible/unknown injury or fatal/major/minor injury). On the other hand, collisions with animal lead to less severe crash outcomes.

US-34 Corridor

The models of US-34 corridor included fewer weather-related variables than those of the I-80. Thus, it is more difficult to draw many inferences in terms of the weather effects on crash severity for that corridor. Nevertheless, variables associated with adverse weather conditions were found to lead to severe outcomes (injuries). This finding is contradictory to the major findings of the I-80 models.

The effect of wind (speed and direction) was found significant as well. However, the 1-hr average wind speed especially in a rural environment has now a negative effect on the probability of a PDO outcome. Specifically, as the average wind speed on a rural road increases, the probability of a PDO crash decreases. This finding is also contradictory to the findings of the I-80 models. In terms of wind direction, winds of non-parallel direction to the direction of the vehicle movement in combination with the logarithm of truck ADT of the road segment were found to increase the probability of fatal/major/minor injury outcomes.

Finally, crashes occurred during morning hours and crashes occurred in August were less likely to be of PDO outcome, regardless of the prevailing weather conditions.

In conclusion, the findings of the models were different on the two corridors, since different variables were found significant in each case. However, it is noteworthy that the models of the US-34 corridor included fewer interaction variables (and fewer variables in general). This fact may suggest that different factors affect the severity of crashes on each study corridor. The main difference though is that adverse weather conditions lead to injury outcomes on US-34, a finding which is contradictory to I-80 corridor and the existing literature. Thus, once could say that the combination of adverse weather conditions and route classification may influence crash severity in different ways. This difference could be attributed to the geometric and roadway characteristics of the corridors and particularly the absence of median on the US route. In addition, one could speculate that different maintenance measures and policies, especially during adverse weather conditions, are applied to routes based on their functional classification and traffic volumes. In other words, an interstate route (with higher traffic volumes) may gather more and early attention (e.g., snow plowing, better enforcement) during inclement weather than routes of other classification.

Finally, it should be pointed out that based on the estimation results, multinomial logit (MNL) models are preferable to binary probit models due to their flexibility to investigate more than two outcomes. Furthermore, creating interaction variables is useful for investigating the effects that a combination of two or more different variables may have on specific outcomes.

6.3 Study Limitations

One of the main limitations of this thesis is the small area of study. Only two corridors of specific length and classification of the Iowa roadway network were considered in the analysis. Although those corridors were selected based on the adequate sample of cases that they could provide and their proximity to RWIS, any generalization of the results of this study to the entire state network might not be accurate. Thus, the author recommends this study to be evaluated as a case study of two different corridors and not as a statewide analysis.

Furthermore, the author did not consider human factor effects (e.g., gender and age of drivers) or vehicle characteristics (e.g., vehicle classification). These factors were not included since the focus of this thesis was on roadway factors and their interaction with weather conditions. It is anticipated that including human or vehicle factors would involve a more computationally expensive estimation process but the overall model fit would improve. In that case, mixed logit models (with random parameters) could be used to take into account the variability in human-factors (i.e., multiple age ranges, many different vehicle classifications, etc.).

In addition to the aforementioned issues, some limitations are also associated with the data and especially the weather information. This study attempted to incorporate real-time information about the prevailing weather conditions at the location and time of a crash. However, as in similar studies, real-time information about weather conditions is not available for every single point of a corridor. Despite the fact that weather records from the nearest RWIS were utilized, the recorded weather conditions at the RWIS location at the time of crash (or actually the 1-hr period prior to that time) may be quite different from the exact weather conditions at crash location, especially when the distance between the RWIS and the crash point is large. This is mainly true for the records reporting wind speed, direction, etc.

The weather data also suffered from additional issues. First, the records of the RWIS were obtained in extremely raw format and in many cases were incomplete (see also Section 3.3). Thus, the corresponding weather information was missing for some crash observations (especially information related to precipitation). Those observations were skipped during the statistical analysis since interaction variables associated with weather conditions could not be created. As a result, the sample size was reduced and thus the models were developed on fewer observations than the original number of observations. It should be noted that the problem of missing data was more serious on US-34 corridor. Thus, the models of that corridor were estimated based on a small sample, a fact that needs to be considered when reviewing the corresponding results and conclusions. Finally, another limitation was the lack of visibility information for most of the weather records during the analysis period. As such, it was not possible to examine the significance of this variable on crash severity as indicated in many similar previous studies.

Lastly, real-time information about actual vehicle speeds was not collected. That was due to the difficulty in collecting and processing raw speed data by Automatic Traffic Recorders (ATR) at the location and time of each crash. For this reason, any variables and inferences related to speed (e.g., driving too fast for conditions) were based only on the speed information provided by the crash reports. However, this information is subjective to the person who filled in the report and may be subject to errors or omissions.

6.4 Recommendations

As already mentioned, this thesis constitutes a case study of two specific corridors. However, the findings of this study can have general implications on the improvement of road safety especially during adverse weather conditions. The results of this study could be of interest to transportation agencies, driving education and license providers, and road users. Furthermore, the study limitations can provide the groundwork for future research and also improvements in data collection and maintenance.

First, transportation agencies, such as the Iowa Department of Transportation, can take a number of measures in order to improve and ensure safety on roadways, especially in periods of inclement weather. For instance, appropriate maintenance and operation activities should be performed during events of adverse weather (such as when temperature is below freezing or under precipitation events). The need of effective maintenance of corridors of lower classification and traffic volumes should not be underestimated. It is noteworthy that findings of this study suggest that although adverse weather may be associated with less severe crashes on interstates, this is opposite for the case of US routes. Besides, except for the models' results, this argument is also supported by a simple analysis and test of the difference in the probabilities of a severe crash (such as fatal or major injury crash) during adverse weather conditions among the two corridors (refer to Appendix D for more details). This finding could be attributed to the priority (and the larger share of funds) that might be given to the maintenance of interstate routes.

Nevertheless, if more attention and funds were allocated to routes of lower classification (such as US routes), the DOTs and the states in general could benefit from responding to less severe crashes and incur lower costs associated with them. For instance, let consider the effect of the temperature below freezing on a US route. Based on Equation 4.13, the probability of a fatal/major/minor injury outcome for a crash that occurred under temperature below 0^{o}C is 0.1 or 10% (if all the other variables are assumed to be equal to 0). However, if the crash was avoided under such conditions, the probability of that crash to be of a fatal/major/minor injury outcome would be 5%. Thus, the probability is reduced by 50% (refer to Appendix E for the calculation).

Weather conditions were also found to interact with vehicle speed and that interaction seems to contribute to crash severity. In specific, driving too fast for conditions in combination with high wind speed and/or non-parallel direction was found to increase the probability of severe outcomes (possible/unknown injury or fatal/major/minor injury) on the Interstate corridor. Moreover, the increase in wind speed on a rural environment and a US corridor was found to be associated with an increase in the probability of a severe outcome. Thus, the use of adjusted speed limits according to the prevailing weather conditions or at least with a seasonal effect (i.e., during winter period) could be a potential beneficial measure. Another promising measure might be the introduction of electronic signs installed on specific spots on the highway network which would provide the road users with real-time information about the prevailing (or future) weather conditions (e.g., temperature, wind speed and direction, chance of precipitation, etc.) in order to properly adjust their driving speed and increase their alertness.

In addition with the aforementioned measures, higher levels of enforcement may be necessary to ensure that road users comply with the speed limits and other safe driving rules. Enforcement is necessary not only during adverse weather conditions but also under other circumstances such as morning peak hours or during the night, and especially on corridors of lower classification. For instance, according to the US corridor model, morning periods between 6:00 am to 9:00 am and dark lighting conditions (probably at night) are associated with increase in the probability of severe outcomes. Finally, although more investigation of the effect of wind is needed, a discussion about whether specific types of vehicles (such as large trucks or motorcycles) should be allowed (or not) to travel during events of severe winds might be useful. Besides, as found by the US corridor model, winds (especially of non-parallel direction) in association with Truck AADT seem to lead to severe outcomes.

Apart from policy measures, emphasis on the driver education is essential. Based on the results of this study, appropriate driving (for example, selection of speeds) according to the prevailing weather conditions may avoid serious crash outcomes. For this reason, driver education should provide future drivers with skills on safe driving under adverse conditions. These skills could also be tested during the driving license exams. This however requires the adjustment of the driver’s manual as well. For instance, specific reference to the effect of weather conditions (such as the negative effect of wind or the temperature below freezing) could be included in the manual. Also, driving sessions under events of inclement weather (such as rainfall or snowfall) could be organized by driver education providers. Finally, information campaigns about safe driving during winter and under events of inclement weather may contribute to the reduction of crashes, or at least the severe ones.

Finally, based on the aforementioned data limitations, the author would like to point out the need of comprehensive and accurate data. Special also emphasis should be given to the installation and maintenance of recording stations (as RWIS), especially on routes of lower classification.

Recommendations for future research should be mainly associated with the need of generalization of the current findings to other routes and road networks. For this purpose, similar studies on larger samples from different routes and in larger scale (e.g. statewide) are necessary. Furthermore, a more systematic investigation of the interaction effects of weather conditions along with other crash parameters (by using interaction variables) is highly recommended. Finally, the incorporation of human and vehicle factors in the analysis is recommended for a more comprehensive analysis of this phenomenon.

REFERENCES

Abdel-Aty, M. A., Hassan, H. M., Ahmed, M. & Al-Ghamdi, A. S., 2012. Real-Time Prediction of Visibility Related Crashes. *Transportation Research Part C* , Volume 24, pp. 288-298.

Abdel-Aty, M. A. & Pemmanaboina, R., 2006. Calibrating a Real-Time Traffic Crash-Prediction Model Using Archived Weather and ITS Traffic Data. *IEEE Transactions on Intelligect Transportation Systems*, Volume 7, pp. 167-174.

Agarwal, M., Maze, T. & Souleyrette, R., 2005. *Impact of Weather on Urban Freeway Traffic Flow Characteristics and Facility Capacity*, Ames, IA: Institute for Transportation, Iowa State University.

Ahmed, M., Abdel-Aty, M. A. & Yu, R., 2012. *Assessment of the Interaction between Crash Occurrence, Mountainous Freeway Geometry, Real-Time Weather and AVI Traffic Data*. Washington D.C.: Presented at the 91st TRB Annual Meeting, Transportation Research Board, National Reserach Council.

Andersson, A. K. & Chapman, L., 2011. The Impact of Climate Change on Winter Road Maintenance and Traffic Accidents in West Midlands, UK. *Accident Analysis and Prevention*, Volume 43, pp. 284-289.

Andreescu, M. & Frost, D. B., 1998. Weather and Traffic Accidents in Montreal, Canada. *Climate Research*, Volume 9, pp. 225-230.

Andrey, J., 2010. Long-term Trends in Weather Related Crashes. *Journal of Transport Geography*, Volume 18, pp. 247-258.

Andrey, J., Mills, B., Leahy, M. & Suggett, J., 2003. Weather as a Chronic Hazard for Road Transportation in Canadian Cities. *Natural Hazards*, Volume 28, pp. 319-343.

Andrey, J. & Yagar, S., 1993. A Temporal Analysis of Rain-Related Crash Risk. *Accident Analysis and Prevention*, Volume 25, pp. 465-472.

Baker, C. J., 1993. The Behaviour of Road Vehicles in Unsteady Cross Winds. *Journal of Wind Engineering and Industrial Aerodynamics*, Volume 49, pp. 439-448.

Baker, C. J. & Reynolds, S., 1992. Wind-Induced Accidents of Road Vehicles. *Accident Analysis and Prevention*, Volume 24, pp. 559-575.

Brijs, T., Karlis, D. & Wets, G., 2008. Studying the Effect of Weather Conditions on Daily Crash Counts Using a Descrete Time-Series Model. *Accident Analysis and Prevention*, Volume 40, pp. 1180-1190.

Brodsky, H. & Hakkert, A., 1988. Risk of a Road Accident in Rainy Weather. *Accident Analysis and Prevention,* Volume 20, pp. 161-176.

Caliendo, C., Guida , M. & Parisi, A., 2007. A Crash Prediction Model for Multilane Roads. *Accident Analysis and Prevention,* Volume 39, pp. 657-670.

Coleman, S. A. & Baker, C. J., 1990. High Sided Road Vehicles in Cross Winds. *Journal of Wind Engineering and Industrial Aeorodynamics,* Volume 36, pp. 1383-1392.

Deng, Z., Ivan , J. & Garder, P., 2006. Analysis of Factors Affecting the Severity of Head-on Crashes. Two-Lane Rural Highways in Connecticut. *In Transportation Research Record: Journal of Transportation Research Board,* No 1953. Transportation Research Board, National Research Council, Washington D.C., pp. 137-146.

DVR, 2000. *Wind und Wetter. Wie das Wetter auf den Menschen wirkt,* Bonn, Germany: DVR-Report 2000, No. 4, 8-11. Deutscher Verkehrssicherheitsrat e.V.

Edwards, J. B., 1994. Wind-related Road Accidents in England and Wales 1980-1990. *Journal of Wind Engineering and Industrial Aerodynamics,* Volume 52, pp. 293-303.

Edwards, J. B., 1998. The Relationship between Road Accident Severity and Recorded Weather. *Journal of Safety Research,* Volume 29, pp. 249-262.

Eisenberg, D., 2004. The Mixed Effects of Precipitation. *Accident Analysis and Prevention,* Volume 36, pp. 637-647.

Eisenberg, D. & Warner, K. E., 2005. Effects of Snowfalls on Motor Vehicle Crashes, Injuries, and Fatalities. *American Journal of Public Health,* Volume 95, pp. 120-124.

El-Basyouny, K. & Kwon, D.-W., 2012. *Assessing Time and Weather Effects on Collision Frequency by Severity in Edmonton Using Multivariate Safety Performance Functions.* Washington D.C., Presented at the 91st TRB Annual Meeting, Transportation Research Board, National Research Council.

Fridstrom, L. et al., 1995. Measuring the Contribution of Randomness, Exposure, Weather and Daylight to the Variation in Road Accident Counts. *Accident Analysis and Prevention,* Volume 27, pp. 785-796.

Gao, J., 2012. *A Study on the Relationship Between Asset Condition and Safety on Primary Roads in Iowa.* Ames, IA: Master's Thesis, Iowa State University.

Hans, Z., Hawkins, N. & Nlenanya, I., 2011. *Safety and Mobility Impacts of Winter Weather - Phase I,* Ames, IA: Institute for Transportation, Iowa State University http://www.intrans.iastate.edu/reports/winter_safety_mobility_phI_w_cvr2.pdf.

Hermans, E., Brijs, T., Stiers, T. & Offermans, C., 2006. *The Impact of Weather Conditions on Road Safety Investigated on an Hourly Basis.* Washington D.C., Presented at the 85th TRB Annual Meeting, Transportation Research Board, National Research Council.

Iowa Department of Transportation, 2001. *Base Record and Structure Data Manual.* Ames, Iowa: Office of Transportation Data. Iowa Department of Transportation.

Iowa Department of Transportation, 2012. *Automatic Traffic Recorder Report.* [Online] Available at: http://www.iowadot.gov/maps//atr.htm[Accessed 20 July 2012].

Iowa Department of Transportation, 2012. *Iowa Motor Vehicle Crashes - 1925 to 2011.* [Online] Available at: http://www.iowadot.gov/mvd/ods/stats/crashhistory.pdf [Accessed 20 March 2013]

Jung, S., Qin, X. & Noyce, D., 2012. Injury Severity in Multilevel Crash in Rainy Weather. *Journal of Transportation Engineering, ASCE,* Volume 138, pp. 50-59.

Karim, M. A., Hadiuzzaman, M. & Qiu, T. Z., 2012. *Developing Collision Prediction Models with Weather and Driver Characteristics Related Variables.* Washington D.C., Presented at the 91st TRB Annual Meeting, Transportation Research Board, National Research Council.

Karlaftis, M. G. & Yannis, G., 2010. *Weather Effects on Daily Traffic Accidents and Fatalities: A Time Series Count Data Approach.* Washington D.C., Presented at the 89th TRB Annual Meeting, Transportation Research Board, National Research Council.

Keay, K. & Simmonds, I., 2004. The Association of Rainfall and Other Weather Variables with Road Traffic Volume in Melbourne, Australia. *Accident Analysis and Prevention,* Volume 37, pp. 109-124.

Keay, K. & Simmonds, I., 2006. Road Accidents and Rainfall in a Large Australian City. *Accident Analysis and Prevention,* Volume 38, pp. 445-454.

Khattak, A. J., Kantor, P. & Council, F. M., 1998. Role of Adverse Weather in Key Crash Types on Limited-Access Roadways. Implications for Advanced Weather Systems. *In Transportation Research Record: Journal of Transportation Research Board*, No 1621. Transportation Research Board, National Research Council, Washington D.C., pp. 10-19.

Knapp, K., Kroeger, D. & Giese, K., 2000. *Mobility and Safety Impacts of Winter Storm Events in a Freeway Environment,* Ames, IA: Institute for Transportation, Iowa State University http://www.intrans.iastate.edu/reports/winstorm.pdf.

Koetse, M. J. & Rietveld, P., 2009. The Impact of Climate Change and Weather on Transport: An Overview of Empirical Findings. *Transportation Research Part D,* Volume 14, pp. 205-221.

Laaidi, K. & Laaidi, M., 2002. Météorologie et Sécurité Routière. *Via Secura,* Volume 56, pp. 22-24.

Lord, D. & Mannering, F., 2010. The Statistical Analysis of Crash-Frequency Data: A Review and Assessment of Methodological Alternatives. *Transportation Research Part A,* Volume 44, pp. 291-305.

Lord, D., Washington, S. P. & Ivan, J. N., 2005. Poisson, Poisson-Gamma and Zero-Inflated Regression Models of Motor Vehicle Crashes: Balancing Statistical Fit and Theory. *Accident Analysis and Prevention,* Volume 37, pp. 35-46.

Miller, I. & Miller, M., 2004. *John E. Freund's Mathematical Statistics with Applications.* 7th ed. Upper Saddle River, NJ: Pearson Prentice Hall.

Miranda-Moreno, L. F., Fu, L., Saccomanno, F. & Labbe, A., 1995. Alternative Risk Models for Ranking Locations for Safety Improvements. *In Transportation Research Record: Journal of Transportation Research Board,* No 1908. Transportation Research Board, National Research Council, Washington D.C., pp. 1-8.

National Highway Traffic Safety Administration, 2012. *2011 Motor Vehicle Crashes: Overwiew,* Washington D.C.: United States Department of Transportation.

National Research Council (U.S.), 2000. *Highway Capacity Manual 2000.* Washington D.C.: Transportation Research Board.

Ott, R. L. & Longnecker, M., 2010. *An Introduction to Statistical Methods and Data Analysis.* 6th ed. Belmont, CA: Brooks/Cole, Cengage Learing.

Peltola, H. & Kantonen, J., 1987. Experience on Winter Time Road Accidents. *Tie Ja Liikenne,* Volume 57, pp. 11-14.

Perry, A. H. & Symons, L. J., 1991. *Highway Meteorology.* New York, NY: E & FN Spon.

Qiu, L. & Nixon, W. A., 2008. Effects of Adverse Weather on Traffic Crashes. *In Transportation Research Record: Journal of Transportation Research Board,* No 2055. Transportation Research Board, National Research Council, Washington D.C., pp. 139-146.

Schendersson, R., 1988. *Relationships between Traffic Accidents, Road Surface Conditionsand Winter Maintenance Measures. Accident Risks for Different Quantities of Snowfall,* s.l.: Statents vaeg-och Traffikistitut.

Shankar, V., Mannering, F. & Barfiled, W., 1995. Effect of Roadway Geometrics and Enviromental Factors on Rural Freeway Accidents Frequencies. *Accident Analysis and Prevention,* Volume 27, pp. 371-389.

Shankar, V. N., Chayanan, S., Sittikariya, S., Shyu, M.-B., Juvva, N. K. & Milton, J. C., 2004. Marginal Impacts of Design, Traffic, Weather, and Related Interactions on Roadside Crashes.. *In Transportation Research Record: Journal of Transportation Research Board,* No 1897. Transportation Research Board, National Research Council, Washington D.C., pp. 156-163.

Shankar, V. N., Mannering, F. & Barfield, W., 1996. Statistical Analysis of Accident Severity on Rural Freeways.. *Accident Analysis and Prevention,* Volume 28, pp. 391-401.

Sinha, K. C. & Labi, S., 2007. *Transportation Decision Making.* 1st ed. Hoboken, NJ: John Wiley & Sons.

Stern, E. & Zehavi, Y., 1990. Road Safety and Hot Weather: A Study in Applied Transport Geography. *Transactions of the Institute of British Geographers,* Volume 15, pp. 102-111.

Stipdonk, H. L., 2008. *Time Series Application on Road Safety. Developments in Europe,* Brussels, Belgium: Deliverable D7.10 of the EU FP6 Project SafeNet. European Commission.

Stout, T. B. & Souleyrette, R. R., 2005. *Analyzing Crash Risk Using Automatic Traffic Recorder Speed Data.* Ames, IA, Proceeding of the 2005 Mid-Continent Transportation Research Symposium.

SWOV, 2012. *The Influence of Weather on Road Safety,* Leindschendam, the Netherlands: SWOV.

Usman, T., Fu, L. & Miranda-Moreno, L. F., 2010. Quantifying Safety Benefit of Winter Road Maintenance: Accident Frequency Modeling. *Accident Analysis and Prevention,* Volume 42, pp. 1878-1887.

Usman, T., Fu, L. & Miranda-Moreno , L. F., 2012a. A Disaggregate Model for Quantifying the Safety Effects of Winter Road Maintenance. *Accident Analysis and Prevention,* Volume 48, pp. 368-378.

Usman, T., Miranda-Moreno, L. F. & Fu, L., 2012b. *Analysis of Injury Severity Outcomes of Highway Winter Crashes: A Multi-Level Modeling Approach.* Washington D.C., Presented at the 91s TRB Annual Meeting, Transportation Research Board, National Research Council .

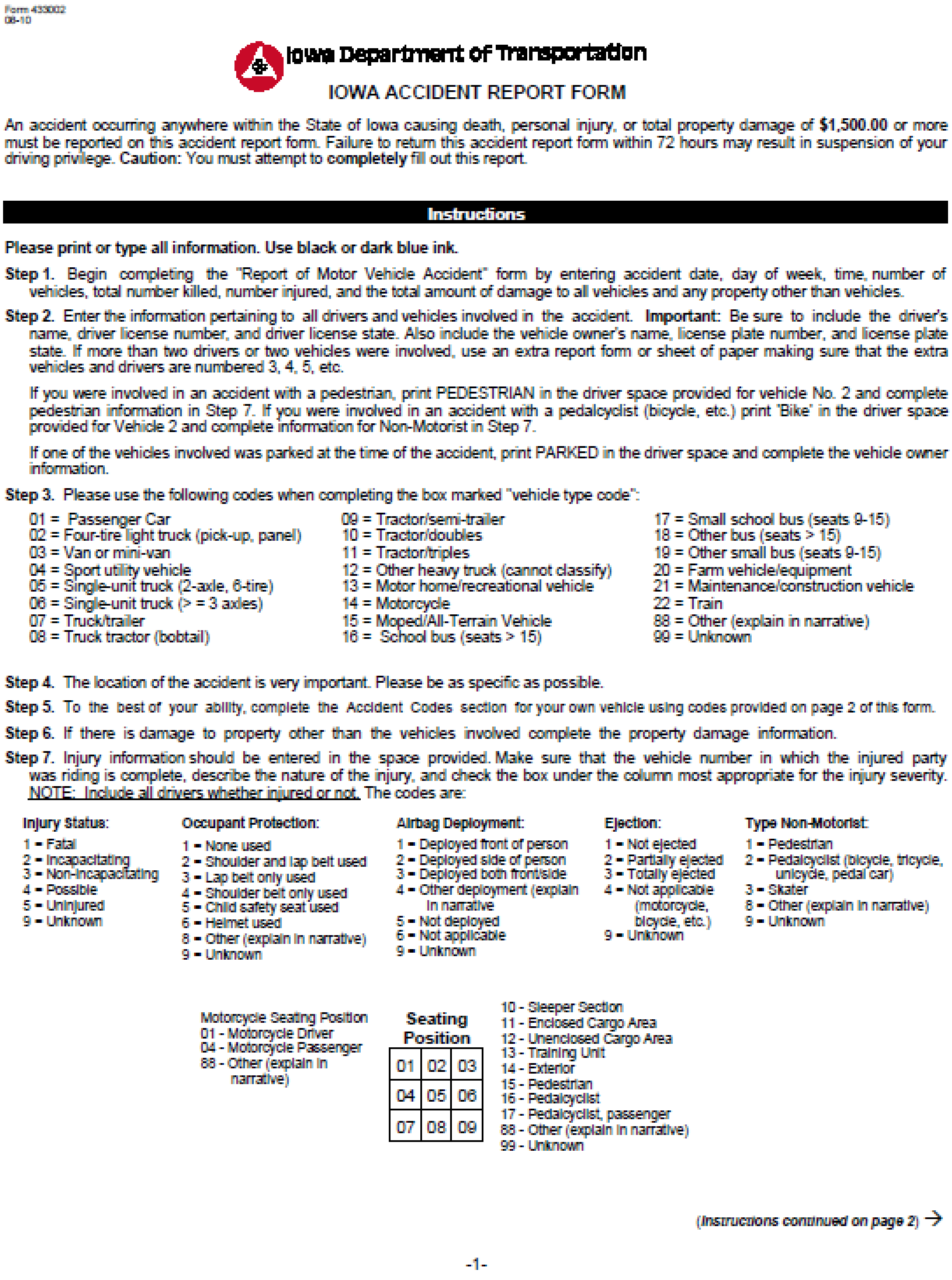
Washington, S. P., Karlaftis, M. G. & Mannering, F. L., 2011. *Statistical and Econometric Methods for Transportation Data Analysis.* 2nd ed. Boca Raton, FL: Chapman & Hall/CRC.

World Health Organization, 2013. *Global Status Report On Road Safety 2013,* Geneva, Switzerland: World Health Organization.

Yannis, G., Antoniou, C., Katsochis, D. & Papadimitriou, E., 2008. *Impact of Meteorological Variables on the Number of Injury Accidents in the Athens Region,* Brussels, Belgium: In Deliverable D7.10 of the EU FP6 Project SafetyNet, European Commision.

APPENDIX A

IOWA ACCIDENT REPORT FORM



Form 433002
08-10

Iowa Department of Transportation

IOWA ACCIDENT REPORT FORM

An accident occurring anywhere within the State of Iowa causing death, personal injury, or total property damage of **$1,500.00** or more must be reported on this accident report form. Failure to return this accident report form within 72 hours may result in suspension of your driving privilege. **Caution:** You must attempt to **completely** fill out this report.

Instructions

Please print or type all information. Use black or dark blue ink.

Step 1. Begin completing the "Report of Motor Vehicle Accident" form by entering accident date, day of week, time, number of vehicles, total number killed, number injured, and the total amount of damage to all vehicles and any property other than vehicles.

Step 2. Enter the information pertaining to all drivers and vehicles involved in the accident. **Important:** Be sure to include the driver's name, driver license number, and driver license state. Also include the vehicle owner's name, license plate number, and license plate state. If more than two drivers or two vehicles were involved, use an extra report form or sheet of paper making sure that the extra vehicles and drivers are numbered 3, 4, 5, etc.

If you were involved in an accident with a pedestrian, print PEDESTRIAN in the driver space provided for vehicle No. 2 and complete pedestrian information in Step 7. If you were involved in an accident with a pedalcyclist (bicycle, etc.) print 'Bike' in the driver space provided for Vehicle 2 and complete information for Non-Motorist in Step 7.

If one of the vehicles involved was parked at the time of the accident, print PARKED in the driver space and complete the vehicle owner information.

Step 3. Please use the following codes when completing the box marked "vehicle type code":

01 = Passenger Car
02 = Four-tire light truck (pick-up, panel)
03 = Van or mini-van
04 = Sport utility vehicle
05 = Single-unit truck (2-axle, 6-tire)
06 = Single-unit truck (> = 3 axles)
07 = Truck/trailer
08 = Truck tractor (bobtail)
09 = Tractor/semi-trailer
10 = Tractor/doubles
11 = Tractor/triples
12 = Other heavy truck (cannot classify)
13 = Motor home/recreational vehicle
14 = Motorcycle
15 = Moped/All-Terrain Vehicle
16 = School bus (seats > 15)
17 = Small school bus (seats 9-15)
18 = Other bus (seats > 15)
19 = Other small bus (seats 9-15)
20 = Farm vehicle/equipment
21 = Maintenance/construction vehicle
22 = Train
88 = Other (explain in narrative)
99 = Unknown

Step 4. The location of the accident is very important. Please be as specific as possible.

Step 5. To the best of your ability, complete the Accident Codes section for your own vehicle using codes provided on page 2 of this form.

Step 6. If there is damage to property other than the vehicles involved complete the property damage information.

Step 7. Injury information should be entered in the space provided. Make sure that the vehicle number in which the injured party was riding is complete, describe the nature of the injury, and check the box under the column most appropriate for the injury severity. <u>NOTE: Include all drivers whether injured or not.</u> The codes are:

Injury Status:
1 = Fatal
2 = Incapacitating
3 = Non-Incapacitating
4 = Possible
5 = Uninjured
9 = Unknown

Occupant Protection:
1 = None used
2 = Shoulder and lap belt used
3 = Lap belt only used
4 = Shoulder belt only used
5 = Child safety seat used
6 = Helmet used
8 = Other (explain in narrative)
9 = Unknown

Airbag Deployment:
1 = Deployed front of person
2 = Deployed side of person
3 = Deployed both front/side
4 = Other deployment (explain in narrative
5 = Not deployed
6 = Not applicable
9 = Unknown

Ejection:
1 = Not ejected
2 = Partially ejected
3 = Totally ejected
4 = Not applicable (motorcycle, bicycle, etc.)
9 = Unknown

Type Non-Motorist:
1 = Pedestrian
2 = Pedalcyclist (bicycle, tricycle, unicycle, pedal car)
3 = Skater
8 = Other (explain in narrative)
9 = Unknown

Motorcycle Seating Position
01 - Motorcycle Driver
04 - Motorcycle Passenger
88 - Other (explain in narrative)

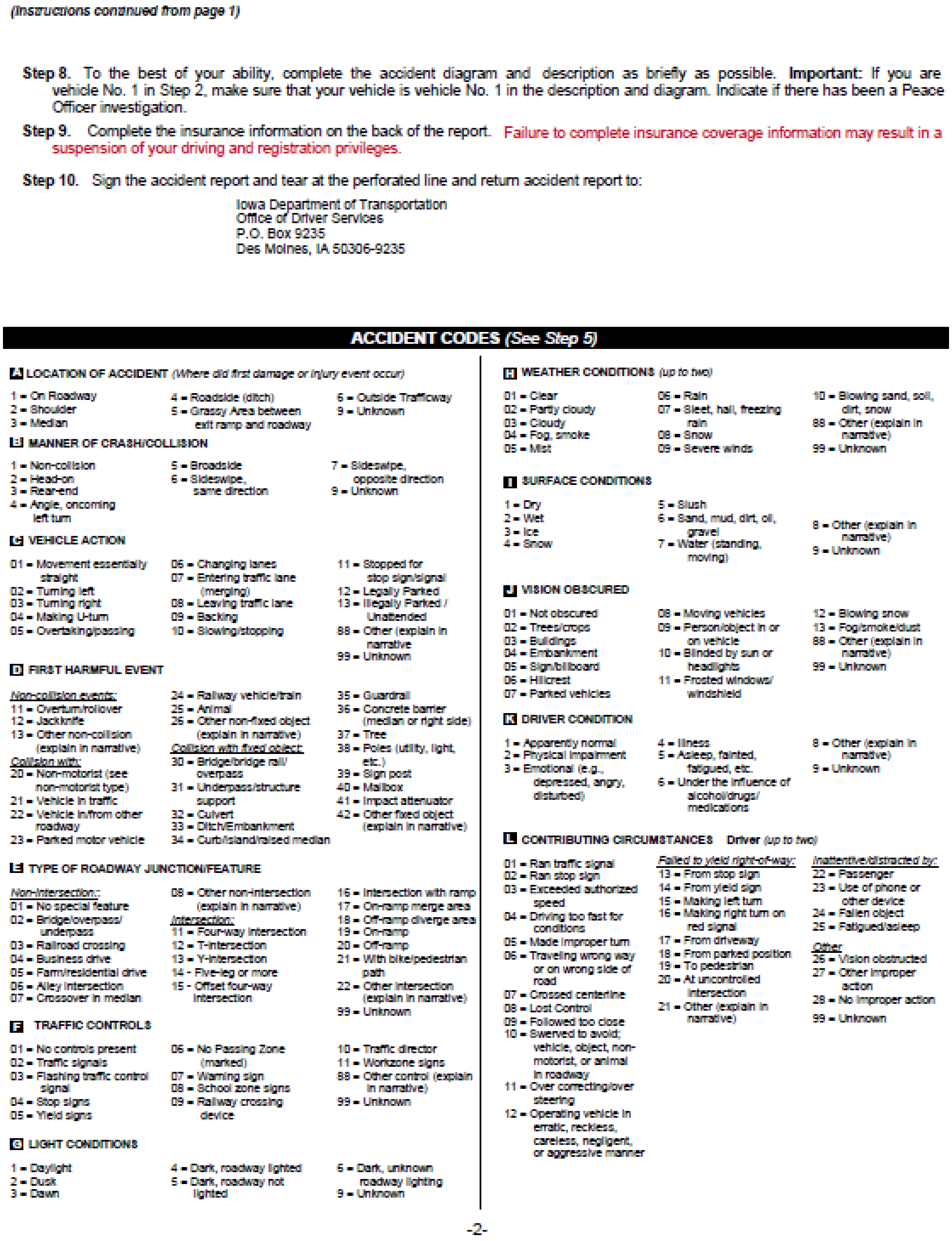
Seating Position

01	02	03
04	05	06
07	08	09

10 - Sleeper Section
11 - Enclosed Cargo Area
12 - Unenclosed Cargo Area
13 - Training Unit
14 - Exterior
15 - Pedestrian
16 - Pedalcyclist
17 - Pedalcyclist, passenger
88 - Other (explain in narrative)
99 - Unknown

(Instructions continued on page 2) →

-1-



(Instructions continued from page 1)

Step 8. To the best of your ability, complete the accident diagram and description as briefly as possible. **Important:** If you are vehicle No. 1 in Step 2, make sure that your vehicle is vehicle No. 1 in the description and diagram. Indicate if there has been a Peace Officer investigation.

Step 9. Complete the insurance information on the back of the report. Failure to complete insurance coverage information may result in a suspension of your driving and registration privileges.

Step 10. Sign the accident report and tear at the perforated line and return accident report to:

Iowa Department of Transportation
Office of Driver Services
P.O. Box 9235
Des Moines, IA 50306-9235

ACCIDENT CODES *(See Step 5)*

A LOCATION OF ACCIDENT *(Where did first damage or injury event occur)*

1 = On Roadway
2 = Shoulder
3 = Median
4 = Roadside (ditch)
5 = Grassy Area between exit ramp and roadway
6 = Outside Trafficway
9 = Unknown

B MANNER OF CRASH/COLLISION

1 = Non-collision
2 = Head-on
3 = Rear-end
4 = Angle, oncoming left turn
5 = Broadside
6 = Sideswipe, same direction
7 = Sideswipe, opposite direction
9 = Unknown

C VEHICLE ACTION

01 = Movement essentially straight
02 = Turning left
03 = Turning right
04 = Making U-turn
05 = Overtaking/passing
06 = Changing lanes
07 = Entering traffic lane (merging)
08 = Leaving traffic lane
09 = Backing
10 = Slowing/stopping
11 = Stopped for stop sign/signal
12 = Legally Parked
13 = Illegally Parked / Unattended
88 = Other (explain in narrative
99 = Unknown

D FIRST HARMFUL EVENT

Non-collision events:
11 = Overturn/rollover
12 = Jackknife
13 = Other non-collision (explain in narrative)
Collision with:
20 = Non-motorist (see non-motorist type)
21 = Vehicle in traffic
22 = Vehicle in/from other roadway
23 = Parked motor vehicle
24 = Railway vehicle/train
25 = Animal
26 = Other non-fixed object (explain in narrative)
Collision with fixed object:
30 = Bridge/bridge rail/ overpass
31 = Underpass/structure support
32 = Culvert
33 = Ditch/Embankment
34 = Curb/island/raised median
35 = Guardrail
36 = Concrete barrier (median or right side)
37 = Tree
38 = Poles (utility, light, etc.)
39 = Sign post
40 = Mailbox
41 = Impact attenuator
42 = Other fixed object (explain in narrative)

E TYPE OF ROADWAY JUNCTION/FEATURE

Non-intersection::
01 = No special feature
02 = Bridge/overpass/ underpass
03 = Railroad crossing
04 = Business drive
05 = Farm/residential drive
06 = Alley intersection
07 = Crossover in median
08 = Other non-intersection (explain in narrative)
Intersection:
11 = Four-way intersection
12 = T-intersection
13 = Y-intersection
14 - Five-leg or more
15 - Offset four-way intersection
16 = Intersection with ramp
17 = On-ramp merge area
18 = Off-ramp diverge area
19 = On-ramp
20 = Off-ramp
21 = With bike/pedestrian path
22 = Other intersection (explain in narrative)
99 = Unknown

F TRAFFIC CONTROLS

01 = No controls present
02 = Traffic signals
03 = Flashing traffic control signal
04 = Stop signs
05 = Yield signs
06 = No Passing Zone (marked)
07 = Warning sign
08 = School zone signs
09 = Railway crossing device
10 = Traffic director
11 = Workzone signs
88 = Other control (explain in narrative)
99 = Unknown

G LIGHT CONDITIONS

1 = Daylight
2 = Dusk
3 = Dawn
4 = Dark, roadway lighted
5 = Dark, roadway not lighted
6 = Dark, unknown roadway lighting
9 = Unknown

H WEATHER CONDITIONS *(up to two)*

01 = Clear
02 = Partly cloudy
03 = Cloudy
04 = Fog, smoke
05 = Mist
06 = Rain
07 = Sleet, hail, freezing rain
08 = Snow
09 = Severe winds
10 = Blowing sand, soil, dirt, snow
88 = Other (explain in narrative)
99 = Unknown

I SURFACE CONDITIONS

1 = Dry
2 = Wet
3 = Ice
4 = Snow
5 = Slush
6 = Sand, mud, dirt, oil, gravel
7 = Water (standing, moving)
8 = Other (explain in narrative)
9 = Unknown

J VISION OBSCURED

01 = Not obscured
02 = Trees/crops
03 = Buildings
04 = Embankment
05 = Sign/billboard
06 = Hillcrest
07 = Parked vehicles
08 = Moving vehicles
09 = Person/object in or on vehicle
10 = Blinded by sun or headlights
11 = Frosted windows/ windshield
12 = Blowing snow
13 = Fog/smoke/dust
88 = Other (explain in narrative)
99 = Unknown

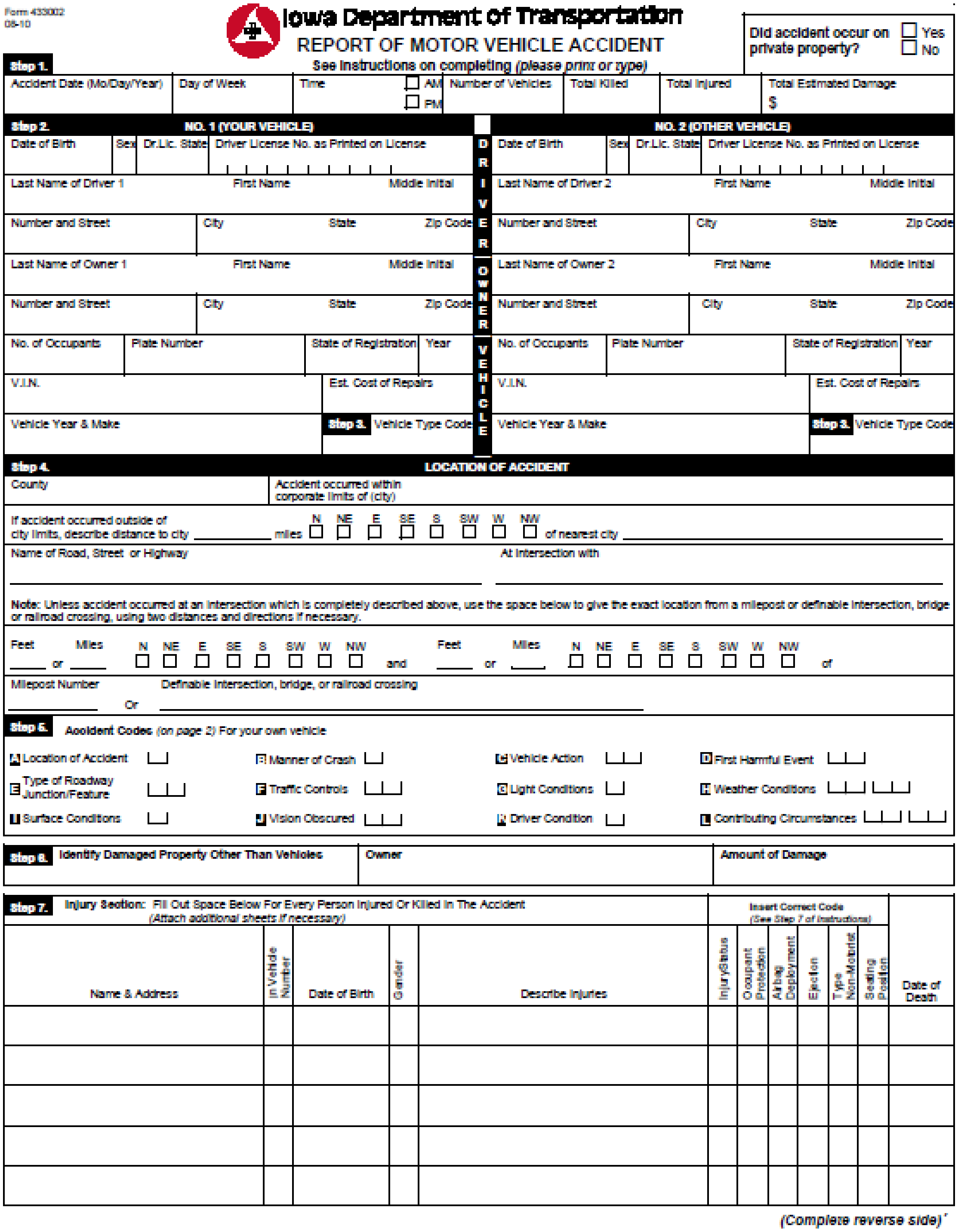
K DRIVER CONDITION

1 = Apparently normal
2 = Physical impairment
3 = Emotional (e.g., depressed, angry, disturbed)
4 = Illness
5 = Asleep, fainted, fatigued, etc.
6 = Under the influence of alcohol/drugs/ medications
8 = Other (explain in narrative)
9 = Unknown

L CONTRIBUTING CIRCUMSTANCES **Driver** *(up to two)*

01 = Ran traffic signal
02 = Ran stop sign
03 = Exceeded authorized speed
04 = Driving too fast for conditions
05 = Made improper turn
06 = Traveling wrong way or on wrong side of road
07 = Crossed centerline
08 = Lost Control
09 = Followed too close
10 = Swerved to avoid; vehicle, object, non-motorist, or animal in roadway
11 = Over correcting/over steering
12 = Operating vehicle in erratic, reckless, careless, negligent, or aggressive manner
Failed to yield right-of-way:
13 = From stop sign
14 = From yield sign
15 = Making left turn
16 = Making right turn on red signal
17 = From driveway
18 = From parked position
19 = To pedestrian
20 = At uncontrolled intersection
21 = Other (explain in narrative)
Inattentive/distracted by:
22 = Passenger
23 = Use of phone or other device
24 = Fallen object
25 = Fatigued/asleep
Other
26 = Vision obstructed
27 = Other improper action
28 = No improper action
99 = Unknown

-2-



Form 433002
08-10

Iowa Department of Transportation
REPORT OF MOTOR VEHICLE ACCIDENT

See instructions on completing *(please print or type)*

Did accident occur on private property? ☐ Yes ☐ No

Step 1.

Accident Date (Mo/Day/Year)	Day of Week	Time	☐ AM ☐ PM	Number of Vehicles	Total Killed	Total Injured	Total Estimated Damage $

Step 2.

NO. 1 (YOUR VEHICLE)		**NO. 2 (OTHER VEHICLE)**
Date of Birth · Sex · Dr.Lic. State · Driver License No. as Printed on License	D	Date of Birth · Sex · Dr.Lic. State · Driver License No. as Printed on License
Last Name of Driver 1 · First Name · Middle Initial	R I V	Last Name of Driver 2 · First Name · Middle Initial
Number and Street · City · State · Zip Code	E R	Number and Street · City · State · Zip Code
Last Name of Owner 1 · First Name · Middle Initial	O W	Last Name of Owner 2 · First Name · Middle Initial
Number and Street · City · State · Zip Code	N E R	Number and Street · City · State · Zip Code
No. of Occupants · Plate Number · State of Registration · Year	V E	No. of Occupants · Plate Number · State of Registration · Year
V.I.N. · Est. Cost of Repairs	H I C	V.I.N. · Est. Cost of Repairs
Vehicle Year & Make · **Step 3.** Vehicle Type Code	L E	Vehicle Year & Make · **Step 3.** Vehicle Type Code

Step 4. **LOCATION OF ACCIDENT**

County

Accident occurred within corporate limits of (city)

If accident occurred outside of city limits, describe distance to city ________ miles ☐ N ☐ NE ☐ E ☐ SE ☐ S ☐ SW ☐ W ☐ NW of nearest city ________________

Name of Road, Street or Highway ________________ At intersection with ________________

Note: Unless accident occurred at an intersection which is completely described above, use the space below to give the exact location from a milepost or definable intersection, bridge or railroad crossing, using two distances and directions if necessary.

Feet ____ or Miles ____ ☐ N ☐ NE ☐ E ☐ SE ☐ S ☐ SW ☐ W ☐ NW and Feet ____ or Miles ____ ☐ N ☐ NE ☐ E ☐ SE ☐ S ☐ SW ☐ W ☐ NW of

Milepost Number ____________ Or Definable Intersection, bridge, or railroad crossing ________________

Step 5. **Accident Codes** *(on page 2)* For your own vehicle

A Location of Accident ⎣_⎦	B Manner of Crash ⎣_⎦	C Vehicle Action ⎣_⎣_⎦	D First Harmful Event ⎣_⎣_⎦
E Type of Roadway Junction/Feature ⎣_⎣_⎦	F Traffic Controls ⎣_⎣_⎦	G Light Conditions ⎣_⎦	H Weather Conditions ⎣_⎣_⎦ ⎣_⎣_⎦
I Surface Conditions ⎣_⎦	J Vision Obscured ⎣_⎣_⎦	K Driver Condition ⎣_⎦	L Contributing Circumstances ⎣_⎣_⎦ ⎣_⎣_⎦

Step 6. **Identify Damaged Property Other Than Vehicles** · **Owner** · **Amount of Damage**

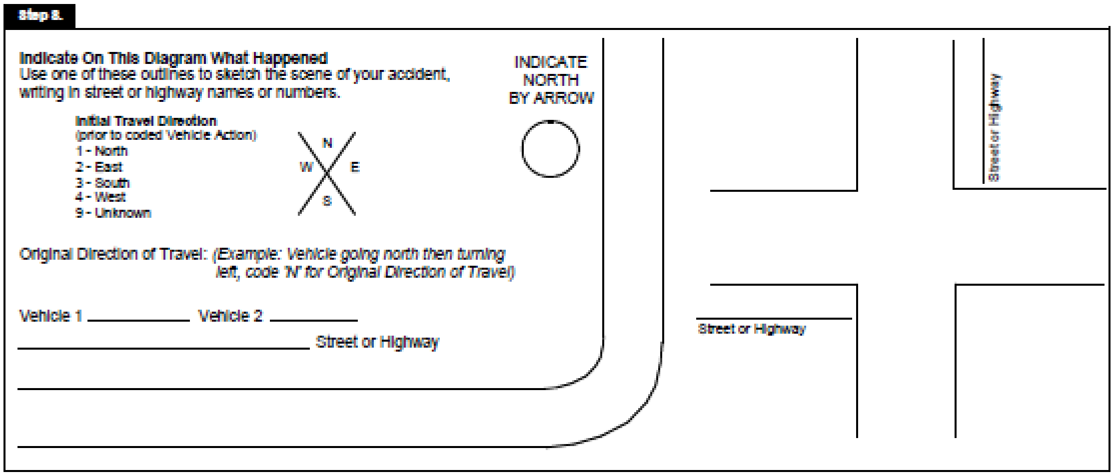
Step 7. **Injury Section:** Fill Out Space Below For Every Person Injured Or Killed In The Accident
(Attach additional sheets if necessary)

					Insert Correct Code *(See Step 7 of Instructions)*						
Name & Address	In Vehicle Number	Date of Birth	Gender	Describe Injuries	InjuryStatus	Occupant Protection	Airbag Deployment	Ejection	Type Non-Motorist	Seating Position	Date of Death

(Complete reverse side)

-3-

Step 8.



Indicate On This Diagram What Happened
Use one of these outlines to sketch the scene of your accident, writing in street or highway names or numbers.

Initial Travel Direction
(prior to coded Vehicle Action)
1 - North
2 - East
3 - South
4 - West
9 - Unknown

INDICATE NORTH BY ARROW

Original Direction of Travel: *(Example: Vehicle going north then turning left, code 'N' for Original Direction of Travel)*

Vehicle 1 ____________ Vehicle 2 ____________

________________________________ Street or Highway

Description
____
____
____
____
____
____
____

Did Peace Officer investigate? ☐ Yes ☐ No Department ________________________

If you did not have automobile liability insurance coverage for this accident, please check this box ☐.

If you had automobile liability insurance coverage for this accident, please complete insurance information below:

Failure To Complete Insurance Coverage Information Requested Below May Result In A Suspension Of Your Driving And/Or Registration Privileges.

Step 9.

Name of Insurance Company **(Not Agent)** Providing Insurance To Cover Your Liability For Damage Or Injury To Others:

____

Name of Agent Who Sold Policy ____

Agent Address ____

Policy No. ____________________ Policy Period: From ____________________ To ____________________

V.I.N. No.____

Name of Driver ____

Name of Owner ____

Name of Policyholder ____

Step 10.

Date	Signature of Driver of Vehicle No. 1	If Signed By Person Other Than Driver, Give Reason

IMPORTANT: This accident should also be reported directly to your insurance company. Failure to report may jeopardize your automobile liability insurance.

APPENDIX B

LIST OF VARIABLES USED IN THE MODELS

Table B. 1: Variables Used in the Models of I-80 Corridor

VARIABLE DESCRIPTION	**VARIABLE MNEMONIC**	**VARIABLE TYPE**	**VARIABLE SAMPLE SPACE**	**MEAN (STD. DEVIATION) OR PRECENTAGE (of observations equal to 1)**
Temperature (^{o}C)	X45	Continuous	[-27.58, 36.37]	5.27 (13.14)
1-hr Precipitation Amount (mm)	X52	Continuous	[0, 128.79]	2.43 (10.28)
1-hr Avg. Wind Speed (Km/hr)	X53	Continuous	[0, 47.83]	15.42 (9.57)
Adjusted segment AADT	X61	Continuous	[6658.37, 10753.70]	26036.40 (4165.58)
Adjusted segment Truck-AADT	X63	Continuous	[6658.37,10753.70]	8497.19 (847.31)
January	JAN	Indicator	{0,1}	17.75
February	FEB	Indicator	{0,1}	10.27
March	MAR	Indicator	{0,1}	4.83
April	APR	Indicator	{0,1}	4.95
May	MAY	Indicator	{0,1}	7.49
June	JUN	Indicator	{0,1}	8.21
July	JUL	Indicator	{0,1}	6.04
August	AUG	Indicator	{0,1}	4.95
September	SEP	Indicator	{0,1}	4.95
October	OCT	Indicator	{0,1}	6.28
November	NOV	Indicator	{0,1}	8.33
December	DEC	Indicator	{0,1}	15.94
Year 2009	D2009	Indicator	{0,1}	35.27
Year 2010	D2010	Indicator	{0,1}	37.20
Year 2011	D2011	Indicator	{0,1}	27.54
Daylight Conditions	DAYLGHT	Indicator	{0,1}	54.23
Dark Conditions	DARK	Indicator	{0,1}	39.49
Rural Road	RURAL	Indicator	{0,1}	92.75
Single Vehicle Crash	VEH1	Indicator	{0,1}	66.06
Collision with Vehicle	COLVEH	Indicator	{0,1}	28.02

Table B. 1 (continued)

VARIABLE DESCRIPTION	VARIABLE MNEMONIC	VARIABLE TYPE	VARIABLE SAMPLE SPACE	MEAN (STD. DEVIATION) OR PRECENTAGE (of observations equal to 1)
Collision with Guardrail	COLANI	Indicator	{0,1}	18.96
Collision with Animal	COLGUAD	Indicator	{0,1}	9.42
Non-collision event	NONCOL	Indicator	{0,1}	63.22
Rear-End Collision	REAREND	Indicator	{0,1}	13.09
Sideswipe	SSWIPE	Indicator	{0,1}	17.91
Overturn/Rollover	OVERTRN	Indicator	{0,1}	13.53
Driving too fast for conditions	TOOFAST	Indicator	{0,1}	20.41
Driving too close	TOOCLOSE	Indicator	{0,1}	2.66
Swerving/Evasive Action	SWEREV	Indicator	{0,1}	14.73
Lost Control	LOSTCON	Indicator	{0,1}	4.59
Rainfall Event	PTRAIN	Indicator	{0,1}	11.79
Snowfall Event	PTSNOW	Indicator	{0,1}	26.35
Wind Direction: North-East	NRES	Indicator	{0,1}	11.42
Wind Direction: East	ES	Indicator	{0,1}	12.30
Wind Direction: South-East	STES	Indicator	{0,1}	12.05
Wind Direction: South	ST	Indicator	{0,1}	13.55
Wind Direction: South-West	STWS	Indicator	{0,1}	13.68
Wind Direction: West	WS	Indicator	{0,1}	12.17
Wind Direction: North-West	NRWS	Indicator	{0,1}	18.95
Wind Direction: North	NR	Indicator	{0,1}	5.90
Logarithm of Adjusted segment AADT	LOGADT	Continuous	[9.75, 10.52]	10.15 (0.16)
Logarithm of Adjusted segment Truck-AADT	LOGTRADT	Continuous	[8.80, 9.28]	9.04 (0.10)
Precipitation Event	PRECIP	Indicator	{0,1}	35.71
Temperature below 0oC	BELOWZ	Indicator	{0,1}	46.93

Table B. 1 (continued)

VARIABLE DESCRIPTION	VARIABLE MNEMONIC	VARIABLE TYPE	VARIABLE SAMPLE SPACE	MEAN (STD. DEVIATION) OR PRECENTAGE (of observations equal to 1)
Driving Too Fast for conditions & Temp. below 0^{o}C	BZTOOFST	Interaction-Indicator	{0,1}	19.32
Single Vehicle Crash & Temp. below 0^{o}C	BZVEH1M	Interaction-Indicator	{0,1}	30.24
Collision with Vehicle & Temp. below 0^{o}C	COLVHBZ	Interaction-Indicator	{0,1}	13.93
Rural Road & 1-hr Precipitation Amount	RPREC	Interaction-Continuous	[0, 128.79]	1.61 (7.47)
1-hr Avg. Wind Speed (lower than 13.9 Km/hr)	WINSA1	Indicator	{0,1}	51.19
1-hr Avg. Wind Speed (between 13.9 and 24.5 Km/hr)	WINSA3	Indicator	{0,1}	17.31
1-hr Avg. Wind Speed (greater than 24.5 Km/hr)	WINSA2	Indicator	{0,1}	31.49
Rural Road & 1-hr Avg Wind Speed	RWINSA	Interaction-Continuous	[0, 64.08]	20.53 (13.98)
Sideswipe & 1-hr Avg. Wind Speed	ORTRWSA	Interaction-Continuous	[0, 61.00]	3.59 (10.51)
Overturn/Rollover & 1-hr Avg. Wind Speed	SSWWSA	Interaction-Continuous	[0, 60.75]	4.12 (10.49)
Driving Too Fast for conditions & 1-hr Avg. Wind Speed (Km/hr)	TFWSA	Interaction-Continuous	[0, 64.08]	5.62 (12.78)
Driving Too Fast for conditions & 1-hr Avg. Wind Speed (lower than 13.9 Km/hr)	TFWSA1	Interaction-Indicator	{0,1}	6.52
Driving Too Fast for conditions & 1-hr Avg. Wind Speed (between 13.9 and 24.5 Km/hr)	TFWSA2	Interaction-Indicator	{0,1}	7.28

Table B. 1 (continued)

VARIABLE DESCRIPTION	VARIABLE MNEMONIC	VARIABLE TYPE	VARIABLE SAMPLE SPACE	MEAN (STD. DEVIATION) OR PRECENTAGE (of observations equal to 1)
Driving Too Fast for conditions & 1-hr Avg. Wind Speed (greater than 24.5 Km/hr)	TFWSA3	Interaction-Indicator	{0,1}	7.03
Daylight Conditions & Temperature (^{o}C)	TDAY	Interaction-Continuous	[-24.45, 35.64]	3.30 (11.14)
Daylight Conditions & Temp. below 0^{o}C	BZDAY	Interaction-Indicator	{0,1}	27.35
Darkness Conditions & Temperature (^{o}C)	TDARK	Interaction-Continuous	[-27.58, 36.37]	1.58 (7.30)
Darkness Conditions & Temp. below 0^{o}C	BZDARK	Interaction-Indicator	{0,1}	17.06
Daylight Conditions & Snowfall Event	SDAY	Interaction-Indicator	{0,1}	17.57
Daylight Conditions & Precipitation Event	PRDAY	Interaction-Indicator	{0,1}	23.12
Daylight Conditions & 1-hr Precipitation Amount (mm)	PHDAY	Interaction-Continuous	[0, 128.79]	19.92 (9.57)
Darkness Conditions & 1-hr Precipitation Amount (mm)	PHDARK	Interaction-Continuous	[0, 73.98]	0.45 (3.86)
Snowfall Event in December	SDEC	Interaction-Indicator	{0,1}	7.15
Snowfall Event in January	SJAN	Interaction-Indicator	{0,1}	10.29
Snowfall Event in February	SFEB	Interaction-Indicator	{0,1}	7.03
Wind of Parallel Direction to the Direction of Vehicle Movement	HORWIN	Indicator	{0,1}	24.47

Table B. 1 (continued)

VARIABLE DESCRIPTION	**VARIABLE MNEMONIC**	**VARIABLE TYPE**	**VARIABLE SAMPLE SPACE**	**MEAN (STD. DEVIATION) OR PRECENTAGE (of observations equal to 1)**
Wind of Cross Direction to the Direction of Vehicle Movement	VERWIN	Indicator	{0,1}	19.45
Wind of Non-Parallel Direction to the Direction of Vehicle Movement	NPWIN	Indicator	{0,1}	75.53
Precipitation Event & Wind of Non-Parallel Direction	PRECIPNP	Interaction-Indicator	{0,1}	27.66
Snowfall Event & Wind of Non-Parallel Direction	SNOWNP	Interaction-Indicator	{0,1}	20.33
Driving Too Fast for Conditions & Wind of Non-Parallel Direction	TOOFSTNP	Interaction-Indicator	{0,1}	16.69
Time of Crash 10:00 pm - 4:00 am	TB1	Indicator	{0,1}	15.70
Time of Crash 6:00 am - 9:00 am	TB3	Indicator	{0,1}	12.92
Time of Crash 9:00 am - 4:00 pm	TB4	Indicator	{0,1}	31.52
Time of Crash 7:00 pm - 10:00 pm	TB5	Indicator	{0,1}	19.08
Time of Crash 4:00 pm - 7:00 pm	TB6	Indicator	{0,1}	14.61
Time of Crash 5:00 pm - 10:00 pm	TB7	Indicator	{0,1}	26.57
Time of Crash 4:00 pm - 10:00 pm	TB8	Indicator	{0,1}	33.70
Time of Crash 7:00 pm - 4:00 am	TB9	Indicator	{0,1}	30.31
Time of Crash 4:00 pm - 10:00 pm & Precipitation Event	TB8PR	Interaction-Indicator	{0,1}	13.51

Table B. 1 (continued)

VARIABLE DESCRIPTION	VARIABLE MNEMONIC	VARIABLE TYPE	VARIABLE SAMPLE SPACE	MEAN (STD. DEVIATION) OR PRECENTAGE (of observations equal to 1)
Time of Crash 9:00 am - 4:00 pm & Snowfall Event	TB4SN	Interaction-Indicator	{0,1}	10.79
Time of Crash 7:00 pm - 10:00 pm & Snowfall Event	TB5SN	Interaction-Indicator	{0,1}	6.02
Time of Crash 4:00 pm - 7:00 pm & Snowfall Event	TB6SN	Interaction-Indicator	{0,1}	2.89
Time of Crash 5:00 pm - 10:00 pm & Snowfall Event	TB7SN	Interaction-Indicator	{0,1}	6.40
Time of Crash 4:00 pm - 10:00 pm & Snowfall Event	TB8SN	Interaction-Indicator	{0,1}	8.91
Time of Crash 7:00 pm - 4:00 am & Snowfall Event	TB9SN	Interaction-Indicator	{0,1}	4.89
Time of Crash 9:00 am - 4:00 pm & Rainfall Event	TB4RN	Interaction-Indicator	{0,1}	3.89
Time of Crash 7:00 pm - 10:00 pm & Rainfall Event	TB5RN	Interaction-Indicator	{0,1}	3.14
Time of Crash 4:00 pm - 7:00 pm & Rainfall Event	TB6RN	Interaction-Indicator	{0,1}	1.63
Time of Crash 5:00 pm - 10:00 pm & Rainfall Event	TB7RN	Interaction-Indicator	{0,1}	3.89
Time of Crash 4:00 pm - 10:00 pm & Rainfall Event	TB8RN	Interaction-Indicator	{0,1}	4.77
Time of Crash 7:00 pm - 4:00 am & Rainfall Event	TB9RN	Interaction-Indicator	{0,1}	3.39

Table B. 1 (continued)

VARIABLE DESCRIPTION	VARIABLE MNEMONIC	VARIABLE TYPE	VARIABLE SAMPLE SPACE	MEAN (STD. DEVIATION) OR PRECENTAGE (of observations equal to 1)
Time of Crash 10:00 pm - 4:00 am & Temp. below 0oC	TB1BZ	Interaction-Indicator	{0,1}	5.02
Time of Crash 6:00 am - 9:00 am & Temp. below 0oC	TB3BZ	Interaction-Indicator	{0,1}	8.03
Time of Crash 9:00 am - 4:00 pm & Temp. below 0oC	TB4BZ	Interaction-Indicator	{0,1}	17.57
Time of Crash 7:00 pm - 10:00 pm & Temp. below 0^{o}C	TB5BZ	Interaction-Indicator	{0,1}	8.28
Time of Crash 4:00 pm - 7:00 pm & Temp. below 0^{o}C	TB6BZ	Interaction-Indicator	{0,1}	5.90
Time of Crash 5:00 pm - 10:00 pm & Temp. below 0^{o}C	TB7BZ	Interaction-Indicator	{0,1}	11.17
Time of Crash 4:00 pm - 10:00 pm & Temp. below 0^{o}C	TB8BZ	Interaction-Indicator	{0,1}	14.18
Time of Crash 7:00 pm - 4:00 am & Temp. below 0^{o}C	TB9BZ	Interaction-Indicator	{0,1}	10.92
Driving Too Fast for conditions & Non-Parallel Wind	TFSTNP	Interaction-Indicator	{0,1}	16.69
Overturn/Rollover & Snowfall Event	ORTRNSN	Interaction-Indicator	{0,1}	7.65

Table B. 2: Variables Used in the Models of US-34 Corridor

VARIABLE DESCRIPTION	VARIABLE MNEMONIC	VARIABLE TYPE	VARIABLE SAMPLE SPACE	MEAN (STD. DEVIATION) OR PRECENTAGES (of observations equal to 1)
Temperature (^{o}C)	X45	Continuous	[-27.58, 36.37]	5.27 (13.14)
1-hr Precipitation Amount (mm)	X52	Continuous	[0, 128.79]	2.43 (10.28)
1-hr Avg. Wind Speed (Km/hr)	X53	Continuous	[0, 47.83]	15.42 (9.57)
Adjusted segment AADT	X61	Continuous	[6658.37, 10753.70]	26036.40 (4165.58)
Adjusted segment Truck-AADT	X63	Continuous	[6658.37,10753.70]	8497.19 (847.31)
January	JAN	Indicator	{0,1}	17.75
February	FEB	Indicator	{0,1}	10.27
March	MAR	Indicator	{0,1}	4.83
April	APR	Indicator	{0,1}	4.95
May	MAY	Indicator	{0,1}	7.49
June	JUN	Indicator	{0,1}	8.21
July	JUL	Indicator	{0,1}	6.04
August	AUG	Indicator	{0,1}	4.95
September	SEP	Indicator	{0,1}	4.95
October	OCT	Indicator	{0,1}	6.28
November	NOV	Indicator	{0,1}	8.33
December	DEC	Indicator	{0,1}	15.94
Sunday	SUN	Indicator	{0,1}	14.25
Monday	MON	Indicator	{0,1}	14.98
Tuesday	TUE	Indicator	{0,1}	14.49
Wednesday	WED	Indicator	{0,1}	12.80
Thursday	THU	Indicator	{0,1}	11.96
Friday	FRI	Indicator	{0,1}	15.58
Saturday	SAT	Indicator	{0,1}	15.94
Year 2009	D2009	Indicator	{0,1}	35.27
Year 2010	D2010	Indicator	{0,1}	37.20
Year 2011	D2011	Indicator	{0,1}	27.54
Daylight Conditions	DAYLGHT	Indicator	{0,1}	54.23
Dark Conditions	DARK	Indicator	{0,1}	39.49
Rural Road	RURAL	Indicator	{0,1}	92.75

Table B. 2 (continued)

VARIABLE DESCRIPTION	VARIABLE MNEMONIC	VARIABLE TYPE	VARIABLE SAMPLE SPACE	MEAN (STD. DEVIATION) OR PRECENTAGES (of observations equal to 1)
Single Vehicle Crash	VEH1M	Indicator	{0,1}	66.06
Collision with Vehicle	COLVEH	Indicator	{0,1}	28.02
Collision with Guardrail	COLANI	Indicator	{0,1}	18.96
Collision with Animal	COLGUAD	Indicator	{0,1}	9.42
Non-collision event	NONCOL	Indicator	{0,1}	63.22
Rear-End Collision	REAREND	Indicator	{0,1}	13.09
Sideswipe	SSWIPE	Indicator	{0,1}	17.91
Overturn/Rollover	OVERTRN	Indicator	{0,1}	13.53
Driving too fast for conditions	TOOFAST	Indicator	{0,1}	20.41
Driving too close	TOOCLOSE	Indicator	{0,1}	2.66
Swerving/Evasive Action	SWEREV	Indicator	{0,1}	14.73
Lost Control	LOSTCON	Indicator	{0,1}	4.59
Rainfall Event	PTRAIN	Indicator	{0,1}	11.79
Snowfall Event	PTSNOW	Indicator	{0,1}	26.35
Wind Direction: North-East	NRES	Indicator	{0,1}	11.42
Wind Direction: East	ES	Indicator	{0,1}	12.30
Wind Direction: South-East	STES	Indicator	{0,1}	12.05
Wind Direction: South	ST	Indicator	{0,1}	13.55
Wind Direction: South-West	STWS	Indicator	{0,1}	13.68
Wind Direction: West	WS	Indicator	{0,1}	12.17
Wind Direction: North-West	NRWS	Indicator	{0,1}	18.95
Wind Direction: North	NR	Indicator	{0,1}	5.90
Logarithm of Adjusted segment AADT	LOGADT	Continuous	[9.75, 10.52]	10.15 (0.16)
Logarithm of Adjusted segment Truck-AADT	LOGTRADT	Continuous	[8.80, 9.28]	9.04 (0.10)

Table B. 2 (continued)

VARIABLE DESCRIPTION	VARIABLE MNEMONIC	VARIABLE TYPE	VARIABLE SAMPLE SPACE	MEAN (STD. DEVIATION) OR PRECENTAGES (of observations equal to 1)
Precipitation Event	PRECIP	Indicator	{0,1}	35.71
Temperature below 0^oC	BELOWZ	Indicator	{0,1}	46.93
Driving Too Fast for conditions & Temp. below 0^oC	BZTOOFST	Interaction-Indicator	{0,1}	19.32
Single Vehicle Crash & Temp. below 0^oC	BZVEH1M	Interaction-Indicator	{0,1}	30.24
Collision with Vehicle & Temp. below 0^oC	COLVHBZ	Interaction-Indicator	{0,1}	13.93
Rural Road & 1-hr Precipitation Amount	RPREC	Interaction-Continuous	[0, 128.79]	1.61 (7.47)
1-hr Avg. Wind Speed (lower than 13.9 Km/hr)	WINSA1	Indicator	{0,1}	51.19
1-hr Avg. Wind Speed (between 13.9 and 24.5 Km/hr)	WINSA3	Indicator	{0,1}	17.31
1-hr Avg. Wind Speed (greater than 24.5 Km/hr)	WINSA2	Indicator	{0,1}	31.49
Rural Road & 1-hr Avg Wind Speed	RWINSA	Interaction-Continuous	[0, 64.08]	20.53 (13.98)
Sideswipe & 1-hr Avg. Wind Speed	ORTRWSA	Interaction-Continuous	[0, 61.00]	3.59 (10.51)
Overturn/Rollover & 1-hr Avg. Wind Speed	SSWWSA	Interaction-Continuous	[0, 60.75]	4.12 (10.49)
Driving Too Fast for conditions & 1-hr Avg. Wind Speed (Km/hr)	TFWSA	Interaction-Continuous	[0, 64.08]	5.62 (12.78)
Driving Too Fast for conditions & 1-hr Avg. Wind Speed (lower than 13.9 Km/hr)	TFWSA1	Interaction-Indicator	{0,1}	6.52

Table B. 2 (continued)

VARIABLE DESCRIPTION	VARIABLE MNEMONIC	VARIABLE TYPE	VARIABLE SAMPLE SPACE	MEAN (STD. DEVIATION) OR PRECENTAGES (of observations equal to 1)
Driving Too Fast for conditions & 1-hr Avg. Wind Speed (between 13.9 and 24.5 Km/hr)	TFWSA2	Interaction-Indicator	{0,1}	7.28
Driving Too Fast for conditions & 1-hr Avg. Wind Speed (greater than 24.5 Km/hr)	TFWSA3	Interaction-Indicator	{0,1}	7.03
Daylight Conditions & Temperature (^{o}C)	TDAY	Interaction-Continuous	[-24.45, 35.64]	3.30 (11.14)
Daylight Conditions & Temp. below 0^{o}C	BZDAY	Interaction-Indicator	{0,1}	27.35
Darkness Conditions & Temperature (^{o}C)	TDARK	Interaction-Continuous	[-27.58, 36.37]	1.58 (7.30)
Darkness Conditions & Temp. below 0^{o}C	BZDARK	Interaction-Indicator	{0,1}	17.06
Daylight Conditions & Snowfall Event	SDAY	Interaction-Indicator	{0,1}	17.57
Daylight Conditions & Precipitation Event	PRDAY	Interaction-Indicator	{0,1}	23.12
Daylight Conditions & 1-hr Precipitation Amount (mm)	PHDAY	Interaction-Continuous	[0, 128.79]	19.92 (9.57)
Darkness Conditions & 1-hr Precipitation Amount (mm)	PHDARK	Interaction-Continuous	[0, 73.98]	0.45 (3.86)
Snowfall Event in December	SDEC	Interaction-Indicator	{0,1}	7.15
Snowfall Event in January	SJAN	Interaction-Indicator	{0,1}	10.29
Snowfall Event in February	SFEB	Interaction-Indicator	{0,1}	7.03

Table B. 2 (continued)

VARIABLE DESCRIPTION	VARIABLE MNEMONIC	VARIABLE TYPE	VARIABLE SAMPLE SPACE	MEAN (STD. DEVIATION) OR PRECENTAGES (of observations equal to 1)
Wind of Parallel Direction to the Direction of Vehicle Movement	HORWIN	Indicator	{0,1}	24.47
Wind of Cross Direction to the Direction of Vehicle Movement	VERWIN	Indicator	{0,1}	19.45
Wind of Non-Parallel Direction to the Direction of Vehicle Movement	NPWIN	Indicator	{0,1}	75.53
Precipitation Event & Wind of Non-Parallel Direction	PRECIPNP	Interaction-Indicator	{0,1}	27.66
Snowfall Event & Wind of Non-Parallel Direction	SNOWNP	Interaction-Indicator	{0,1}	20.33
Driving Too Fast for Conditions & Wind of Non-Parallel Direction	TOOFSTNP	Interaction-Indicator	{0,1}	16.69
Time of Crash 10:00 pm - 4:00 am	TB1	Indicator	{0,1}	15.70
Time of Crash 6:00 am - 9:00 am	TB3	Indicator	{0,1}	12.92
Time of Crash 9:00 am - 4:00 pm	TB4	Indicator	{0,1}	31.52
Time of Crash 7:00 pm - 10:00 pm	TB5	Indicator	{0,1}	19.08
Time of Crash 4:00 pm - 7:00 pm	TB6	Indicator	{0,1}	14.61
Time of Crash 5:00 pm - 10:00 pm	TB7	Indicator	{0,1}	26.57
Time of Crash 4:00 pm - 10:00 pm	TB8	Indicator	{0,1}	33.70
Time of Crash 7:00 pm - 4:00 am	TB9	Indicator	{0,1}	30.31

Table B. 2 (continued)

VARIABLE DESCRIPTION	**VARIABLE MNEMONIC**	**VARIABLE TYPE**	**VARIABLE SAMPLE SPACE**	**MEAN (STD. DEVIATION) OR PRECENTAGES (of observations equal to 1)**
Time of Crash 4:00 pm - 10:00 pm & Precipitation Event	TB8PR	Interaction-Indicator	{0,1}	13.51
Time of Crash 9:00 am - 4:00 pm & Snowfall Event	TB4SN	Interaction-Indicator	{0,1}	10.79
Time of Crash 7:00 pm - 10:00 pm & Snowfall Event	TB5SN	Interaction-Indicator	{0,1}	6.02
Time of Crash 4:00 pm - 7:00 pm & Snowfall Event	TB6SN	Interaction-Indicator	{0,1}	2.89
Time of Crash 5:00 pm - 10:00 pm & Snowfall Event	TB7SN	Interaction-Indicator	{0,1}	6.40
Time of Crash 4:00 pm - 10:00 pm & Snowfall Event	TB8SN	Interaction-Indicator	{0,1}	8.91
Time of Crash 7:00 pm - 4:00 am & Snowfall Event	TB9SN	Interaction-Indicator	{0,1}	4.89
Time of Crash 9:00 am - 4:00 pm & Rainfall Event	TB4RN	Interaction-Indicator	{0,1}	3.89
Time of Crash 7:00 pm - 10:00 pm & Rainfall Event	TB5RN	Interaction-Indicator	{0,1}	3.14
Time of Crash 4:00 pm - 7:00 pm & Rainfall Event	TB6RN	Interaction-Indicator	{0,1}	1.63
Time of Crash 5:00 pm - 10:00 pm & Rainfall Event	TB7RN	Interaction-Indicator	{0,1}	3.89
Time of Crash 4:00 pm - 10:00 pm & Rainfall Event	TB8RN	Interaction-Indicator	{0,1}	4.77
Time of Crash 7:00 pm - 4:00 am & Rainfall Event	TB9RN	Interaction-Indicator	{0,1}	3.39

Table B. 2 (continued)

VARIABLE DESCRIPTION	VARIABLE MNEMONIC	VARIABLE TYPE	VARIABLE SAMPLE SPACE	MEAN (STD. DEVIATION) OR PRECENTAGES (of observations equal to 1)
Time of Crash 10:00 pm - 4:00 am & Temp. below 0^oC	TB1BZ	Interaction-Indicator	{0,1}	5.02
Time of Crash 6:00 am - 9:00 am & Temp. below 0^oC	TB3BZ	Interaction-Indicator	{0,1}	8.03
Time of Crash 9:00 am - 4:00 pm & Temp. below 0^oC	TB4BZ	Interaction-Indicator	{0,1}	17.57
Time of Crash 7:00 pm - 10:00 pm & Temp. below 0^oC	TB5BZ	Interaction-Indicator	{0,1}	8.28
Time of Crash 4:00 pm - 7:00 pm & Temp. below 0^oC	TB6BZ	Interaction-Indicator	{0,1}	5.90
Time of Crash 5:00 pm - 10:00 pm & Temp. below 0^oC	TB7BZ	Interaction-Indicator	{0,1}	11.17
Time of Crash 4:00 pm - 10:00 pm & Temp. below 0^oC	TB8BZ	Interaction-Indicator	{0,1}	14.18
Time of Crash 7:00 pm - 4:00 am & Temp. below 0^oC	TB9BZ	Interaction-Indicator	{0,1}	10.92
Driving Too Fast for conditions & Non-Parallel Wind	TFSTNP	Interaction-Indicator	{0,1}	16.69
Overturn/Rollover & Snowfall Event	ORTRNSN	Interaction-Indicator	{0,1}	7.65

APPENDIX C

NLOGIT OUTPUTS

I-80 Corridor

Binary Probit Model

```
--> probit; lhs=x70;rhs=one,bzveh1,x54,colani,overtrn,feb,sdec,tb7rn,snowvr,t...

 ************************************************************************
 * NOTE: Deleted     33 observations with missing data. N is now    795 *
 ************************************************************************

Normal exit from iterations. Exit status=0.

+---------------------------------------------+
| Binomial Probit Model                       |
| Maximum Likelihood Estimates                |
| Model estimated: Mar 18, 2013 at 02:19:13AM.|
| Dependent variable                  X70     |
| Weighting variable                 None     |
| Number of observations              795     |
| Iterations completed                  6     |
| Log likelihood function       -368.2874     |
| Number of parameters                 10     |
| Info. Criterion: AIC =           .95167     |
|   Finite Sample: AIC =           .95202     |
| Info. Criterion: BIC =          1.01051     |
| Info. Criterion:HQIC =           .97428     |
| Restricted log likelihood     -415.1904     |
| McFadden Pseudo R-squared      .1129675     |
| Chi squared                    93.80607     |
| Degrees of freedom                    9     |
| Prob[ChiSqd > value] =         .0000000     |
| Hosmer-Lemeshow chi-squared =   5.25686     |
| P-value=  .72980 with deg.fr. =       8     |
+---------------------------------------------+
+---------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+---------+--------------+----------------+--------+--------+----------+
---------+Index function for probability
 Constant|     -.48332436       .10921425    -4.425   .0000
 BZVEH1  |     -.56763920       .14355768    -3.954   .0001    .30314465
 X54     |     -.01508953       .00585361    -2.578   .0099   15.4296651
 COLANI  |     -.99152531       .19337475    -5.127   .0000    .18993711
 OVERTRN |      .88035052       .15693446     5.610   .0000    .13584906
 FEB     |      .31743282       .17815881     1.782   .0748    .10566038
 SDEC    |      .32307277       .19889981     1.624   .1043    .07169811
 TB7RN   |     -.80113995       .37151493    -2.156   .0311    .03899371
 SNOWVR  |     -.45390713       .26127040    -1.737   .0823    .05157233
 TOOFSTNP|      .33755568       .14672789     2.301   .0214    .16729560

+-------------------------------------------+
| Partial derivatives of E[y] = F[*]   with |
```

```
| respect to the vector of characteristics. |
| They are computed at the means of the Xs. |
| Observations used for means are All Obs.  |
+-------------------------------------------+
+---------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient   | Standard Error |b/St.Er.|P[|Z|>z]|Elasticity|
+---------+--------------+----------------+--------+--------+----------+
---------+
 Constant|    -.15455378       .03845727    -4.019   .0001
---------+Marginal effect for dummy variable is P|1 - P|0.
 BZVEH1  |    -.13597667       .03058812    -4.445   .0000    -.22344539
 X54     |    -.00402076       .00155961    -2.578   .0099    -.33629616
---------+Marginal effect for dummy variable is P|1 - P|0.
 COLANI  |    -.19440698       .02491558    -7.803   .0000    -.20016068
---------+Marginal effect for dummy variable is P|1 - P|0.
 OVERTRN |     .29091532       .05836430     4.984   .0000     .21423001
---------+Marginal effect for dummy variable is P|1 - P|0.
 FEB     |     .09375395       .05742638     1.633   .1026     .05369809
---------+Marginal effect for dummy variable is P|1 - P|0.
 SDEC    |     .09637569       .06522616     1.478   .1395     .03745694
---------+Marginal effect for dummy variable is P|1 - P|0.
 TB7RN   |    -.14529316       .03935012    -3.692   .0002    -.03071120
---------+Marginal effect for dummy variable is P|1 - P|0.
 SNOWVR  |    -.09884505       .04458722    -2.217   .0266    -.02763304
---------+Marginal effect for dummy variable is P|1 - P|0.
 TOOFSTNP|     .09868931       .04645296     2.125   .0336     .08949766

+----------------------------------------+
| Fit Measures for Binomial Choice Model |
| Probit   model for variable X70        |
+----------------------------------------+
| Proportions P0= .783648   P1= .216352  |
| N =      795 N0=      623   N1=      172  |
| LogL=      -368.287 LogL0=     -415.190  |
| Estrella = 1-(L/L0)^(-2L0/n) = .11769  |
+----------------------------------------+
|     Efron |  McFadden  |  Ben./Lerman  |
|    .11914 |    .11297  |       .69981  |
|    Cramer | Veall/Zim. |      Rsqrd_ML  |
|    .11573 |    .20659  |       .11130  |
+----------------------------------------+
| Information  Akaike I.C. Schwarz I.C.  |
| Criteria          .95167       1.01051  |
+----------------------------------------+
+---------------------------------------------------------+
|Predictions for Binary Choice Model.  Predicted value is |
|1 when probability is greater than  .500000, 0 otherwise.|
|Note, column or row total percentages may not sum to     |
|100% because of rounding. Percentages are of full sample.|
+------+---------------------------------+----------------+
|Actual|         Predicted Value         |                |
|Value |       0                1        | Total Actual   |
+------+----------------+----------------+----------------+
|  0   |    612 ( 77.0%)|     11 (  1.4%)|    623 ( 78.4%)|
|  1   |    147 ( 18.5%)|     25 (  3.1%)|    172 ( 21.6%)|
+------+----------------+----------------+----------------+
|Total |    759 ( 95.5%)|     36 (  4.5%)|    795 (100.0%)|
+------+----------------+----------------+----------------+

=======================================================================
Analysis of Binary Choice Model Predictions Based on Threshold =  .5000
-----------------------------------------------------------------------
Prediction Success
```

```
-----------------------------------------------------------------------
Sensitivity = actual 1s correctly predicted                     14.535%
Specificity = actual 0s correctly predicted                     98.234%
Positive predictive value = predicted 1s that were actual 1s    69.444%
Negative predictive value = predicted 0s that were actual 0s    80.632%
Correct prediction = actual 1s and 0s correctly predicted       80.126%
-----------------------------------------------------------------------
Prediction Failure
-----------------------------------------------------------------------
False pos. for true neg. = actual 0s predicted as 1s             1.766%
False neg. for true pos. = actual 1s predicted as 0s            85.465%
False pos. for predicted pos. = predicted 1s actual 0s          30.556%
False neg. for predicted neg. = predicted 0s actual 1s          19.368%
False predictions = actual 1s and 0s incorrectly predicted      19.874%
```

```
--> DSTAT; rhs = bzveh1,x54,colani,overtrn,feb,sdec,tb7rn,snowvr,toofstnp;out...
Descriptive Statistics
All results based on nonmissing observations.
===============================================================================
Variable     Mean       Std.Dev.     Minimum      Maximum        Cases Missing
===============================================================================
------------------------------------------------------------------------------
All observations in current sample
------------------------------------------------------------------------------
BZVEH1  |  .302384      .459579      .000000      1.00000          797      31
X54     |  15.4204      9.56645      .833333E-01  47.8333          798      30
COLANI  |  .190073      .392596      .000000      1.00000          826       2
OVERTRN |  .135593      .342564      .000000      1.00000          826       2
FEB     |  .102657      .303694      .000000      1.00000          828       0
SDEC    |  .714286E-01  .257701      .000000      1.00000          798      30
TB7RN   |  .388471E-01  .193352      .000000      1.00000          798      30
SNOWVR  |  .513784E-01  .220907      .000000      1.00000          798      30
TOOFSTNP|  .166667      .372912      .000000      1.00000          798      30

Correlation Matrix for Listed Variables

           BZVEH1      X54   COLANI  OVERTRN      FEB     SDEC    TB7RN   SNOWVR
  BZVEH1  1.00000   .22649  -.23565   .38548   .31637   .13494  -.13286   .08131
     X54   .22649  1.00000  -.24862   .16747   .13571   .07055   .00196   .07461
  COLANI  -.23565  -.24862  1.00000  -.19199  -.15601  -.12214   .05155  -.11292
 OVERTRN   .38548   .16747  -.19199  1.00000   .10257   .01788  -.02297  -.02606
     FEB   .31637   .13571  -.15601   .10257  1.00000  -.09552  -.06924   .14187
    SDEC   .13494   .07055  -.12214   .01788  -.09552  1.00000  -.05598   .15565
   TB7RN  -.13286   .00196   .05155  -.02297  -.06924  -.05598  1.00000  -.04697
  SNOWVR   .08131   .07461  -.11292  -.02606   .14187   .15565  -.04697  1.00000

           BZVEH1      X54   COLANI  OVERTRN      FEB     SDEC    TB7RN   SNOWVR
TOOFSTNP   .29096   .24787  -.21704   .13704   .27350   .03219  -.05547   .21548

         TOOFSTNP
TOOFSTNP  1.00000
```

Multinomial Logit Model

```
--> nlogit;lhs=x1;choices=PDO,POSUN,FINJ;model:
```

```
    u(PDO)=bzvh1pd*bzveh2m+winspd*x53+snwvrpd*snowvr
    +tb7rnpd*tb7rn/
    u(POSUN)=POSUN*one+colanips*colani+bzvh1ps*bzveh2m
    +overps*overtrn+tfwsa2ps*tfwsa2/
    u(FINJ)=FINJ*one+overf*overtrn+rprecf*rprec
    +tfstnpf*tfstnp+novf*nov+octf*oct$
+---------------------------------------------+
| Discrete choice and multinomial logit models|
+---------------------------------------------+
Normal exit from iterations. Exit status=0.
+---------------------------------------------+
| Discrete choice (multinomial logit) model   |
| Maximum Likelihood Estimates                |
| Model estimated: Mar 16, 2013 at 02:14:51AM.|
| Dependent variable                Choice    |
| Weighting variable                  None    |
| Number of observations               770    |
| Iterations completed                   7    |
| Log likelihood function       -467.7436    |
| Number of parameters                  15    |
| Info. Criterion: AIC =           1.25388    |
|   Finite Sample: AIC =           1.25471    |
| Info. Criterion: BIC =           1.34439    |
| Info. Criterion:HQIC =           1.28871    |
| R2=1-LogL/LogL*  Log-L fncn  R-sqrd  RsqAdj |
| Constants only    -516.3259  .09409  .08518 |
| Chi-squared[13]           =     97.16453    |
| Prob [ chi squared > value ] =    .00000    |
| Response data are given as ind. choice.     |
| Number of obs.=   828, skipped  58 bad obs. |
+---------------------------------------------+

+---------------------------------------------+
| Notes No coefficients=> P(i,j)=1/J(i).      |
|       Constants only => P(i,j) uses ASCs    |
|         only. N(j)/N if fixed choice set.   |
|         N(j) = total sample frequency for j |
|         N    = total sample frequency.      |
|       These 2 models are simple MNL models. |
|       R-sqrd = 1 - LogL(model)/logL(other)  |
|       RsqAdj=1-[nJ/(nJ-nparm)]*(1-R-sqrd)   |
|         nJ   = sum over i, choice set sizes |
+---------------------------------------------+
+--------+--------------+----------------+--------+--------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]|
+--------+--------------+----------------+--------+--------+
 BZVH1PD |    1.28839965       .35567475     3.622   .0003
 WINSPD  |     .02341733       .01055855     2.218   .0266
 SNWVRPD |     .82321804       .51920644     1.586   .1128
 TB7RNPD |    1.53874634       .75858378     2.028   .0425
 POSUN   |   -1.58112783       .21515759    -7.349   .0000
 COLANIPS|   -2.31609335       .72970767    -3.174   .0015
 BZVH1PS |    1.00492942       .42475914     2.366   .0180
 OVERPS  |    1.31208310       .32782405     4.002   .0001
 TFWSA2PS|     .66161076       .38911882     1.700   .0891
 FINJ    |   -1.56782585       .20699644    -7.574   .0000
 OVERF   |    1.74407186       .34640654     5.035   .0000
 RPRECF  |    -.06903149       .04712546    -1.465   .1430
 TFSTNPF |    1.31765276       .31486770     4.185   .0000
 NOVF    |   -1.10493814       .61246453    -1.804   .0712
 OCTF    |   -1.16444082       .73889393    -1.576   .1150

--> dstat; rhs = bzveh2m, x53, snowvr, tb7rn, colani, tfwsa2, overtrn, rprec,...
```

```
Descriptive Statistics
All results based on nonmissing observations.
===============================================================================
Variable      Mean         Std.Dev.      Minimum       Maximum        Cases Missing
===============================================================================
-------------------------------------------------------------------------------
All observations in current sample
-------------------------------------------------------------------------------
BZVEH2M |  .302384       .459387       .000000       1.00000          2391       93
X53     |  15.4190       9.56837       .833333E-01   47.8333          2391       93
SNOWVR  |  .514429E-01   .220946       .000000       1.00000          2391       93
TB7RN   |  .388959E-01   .193387       .000000       1.00000          2391       93
COLANI  |  .189614       .392074       .000000       1.00000          2484        0
TFWSA2  |  .727729E-01   .259818       .000000       1.00000          2391       93
OVERTRN |  .135266       .342076       .000000       1.00000          2484        0
RPREC   |  1.61154       7.47527       .000000       128.786          2310      174
TFSTNP  |  .166876       .372943       .000000       1.00000          2391       93
NOV     |  .833333E-01   .276441       .000000       1.00000          2484        0
OCT     |  .628019E-01   .242655       .000000       1.00000          2484        0

Correlation Matrix for Listed Variables

          BZVEH2M      X53   SNOWVR    TB7RN   COLANI   TFWSA2  OVERTRN    RPREC
 BZVEH2M  1.00000   .20021   .10196  -.13119  -.22993   .16654   .38671   .01238
     X53   .20021  1.00000   .08936   .01046  -.24037   .13499   .16916   .02472
  SNOWVR   .10196   .08936  1.00000  -.04536  -.10938   .08190  -.01474   .04621
   TB7RN  -.13119   .01046  -.04536  1.00000   .04860  -.05681  -.02226   .05901
  COLANI  -.22993  -.24037  -.10938   .04860  1.00000  -.13698  -.19409  -.09547
  TFWSA2   .16654   .13499   .08190  -.05681  -.13698  1.00000   .08359   .10692
 OVERTRN   .38671   .16916  -.01474  -.02226  -.19409   .08359  1.00000  -.01389
   RPREC   .01238   .02472   .04621   .05901  -.09547   .10692  -.01389  1.00000

          BZVEH2M      X53   SNOWVR    TB7RN   COLANI   TFWSA2  OVERTRN    RPREC
  TFSTNP   .30861   .25798   .24968  -.05542  -.21950   .44743   .12348   .14722
     NOV  -.15090  -.03952  -.06948   .00514   .23440  -.08701  -.09657   .00787
     OCT  -.10841  -.10358  -.05773  -.02633   .13923  -.05165  -.08681  -.03918

           TFSTNP      NOV      OCT
  TFSTNP  1.00000  -.11493  -.10152
     NOV  -.11493  1.00000  -.08179
     OCT  -.10152  -.08179  1.00000
```

US-34 Corridor

Binary Probit Model

```
--> probit; lhs = x70; rhs= one, aug, tb3, rwinsa, sdark,vh2bz, npwin$

 ************************************************************************
 * NOTE: Deleted     36 observations with missing data. N is now    172 *
 ************************************************************************

Normal exit from iterations. Exit status=0.

+---------------------------------------------+
| Binomial Probit Model                       |
| Maximum Likelihood Estimates                |
```

```
| Model estimated: Mar 20, 2013 at 10:15:37AM.|
| Dependent variable                  X70     |
| Weighting variable                 None     |
| Number of observations              172     |
| Iterations completed                  5     |
| Log likelihood function       -91.00207     |
| Number of parameters                  7     |
| Info. Criterion: AIC =          1.13956     |
|   Finite Sample: AIC =          1.14353     |
| Info. Criterion: BIC =          1.26765     |
| Info. Criterion:HQIC =          1.19153     |
| Restricted log likelihood     -107.7914     |
| McFadden Pseudo R-squared      .1557574     |
| Chi squared                    33.57861     |
| Degrees of freedom                    6     |
| Prob[ChiSqd > value] =         .8111896E-05 |
| Hosmer-Lemeshow chi-squared =   5.79974     |
| P-value=  .66965 with deg.fr. =       8     |
+---------------------------------------------+
+---------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient   | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+---------+--------------+----------------+--------+--------+----------+
---------+Index function for probability
 Constant|   -1.46255615       .30165320    -4.848   .0000
 AUG     |    1.56575759       .42106641     3.719   .0002    .06976744
 TB3     |     .72911091       .30135576     2.419   .0155    .12209302
 RWINSA  |     .02541666       .00973812     2.610   .0091   9.27200281
 SDARK   |   -1.02440976       .58751475    -1.744   .0812    .05813953
 VH2BZ   |     .65187442       .28182120     2.313   .0207    .18023256
 NPWIN   |     .53693934       .29647457     1.811   .0701    .79651163

+-------------------------------------------+
| Partial derivatives of E[y] = F[*]   with |
| respect to the vector of characteristics. |
| They are computed at the means of the Xs. |
| Observations used for means are All Obs.  |
+-------------------------------------------+
+---------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient   | Standard Error |b/St.Er.|P[|Z|>z]|Elasticity|
+---------+--------------+----------------+--------+--------+----------+
---------+Index function for probability
 Constant|    -.50349003       .09237290    -5.451   .0000
---------+Marginal effect for dummy variable is P|1 - P|0.
 AUG     |     .56239310       .11292186     4.980   .0000    .13366011
---------+Marginal effect for dummy variable is P|1 - P|0.
 TB3     |     .27498422       .11661665     2.358   .0184    .11436882
 RWINSA  |     .00874977       .00334707     2.614   .0089    .27636269
---------+Marginal effect for dummy variable is P|1 - P|0.
 SDARK   |    -.24858903       .08338262    -2.981   .0029   -.04923371
---------+Marginal effect for dummy variable is P|1 - P|0.
 VH2BZ   |     .24209126       .10808309     2.240   .0251    .14863510
---------+Marginal effect for dummy variable is P|1 - P|0.
 NPWIN   |     .16638179       .08006176     2.078   .0377    .45144718

+----------------------------------------+
| Fit Measures for Binomial Choice Model |
| Probit   model for variable X70        |
+----------------------------------------+
| Proportions P0= .680233   P1= .319767  |
| N =      172 N0=     117   N1=      55  |
| LogL=      -91.002 LogL0=    -107.791  |
| Estrella = 1-(L/L0)^(-2L0/n) = .19121  |
+----------------------------------------+
```

```
|     Efron |  McFadden   |  Ben./Lerman  |
|    .18985 |    .15576   |       .64796  |
|    Cramer | Veall/Zim.  |     Rsqrd_ML  |
|    .18812 |    .29365   |       .17735  |
+----------------------------------------+
| Information  Akaike I.C. Schwarz I.C.  |
| Criteria        1.13956       1.26765  |
+----------------------------------------+
+----------------------------------------------------------+
|Predictions for Binary Choice Model.  Predicted value is |
|1 when probability is greater than  .500000, 0 otherwise.|
|Note, column or row total percentages may not sum to     |
|100% because of rounding. Percentages are of full sample.|
+------+---------------------------------+----------------+
|Actual|         Predicted Value         |                |
|Value |       0                1        | Total Actual   |
+------+----------------+----------------+----------------+
|  0   |    106 ( 61.6%)|     11 (   6.4%)|    117 ( 68.0%)|
|  1   |     34 ( 19.8%)|     21 ( 12.2%)|     55 ( 32.0%)|
+------+----------------+----------------+----------------+
|Total |    140 ( 81.4%)|     32 ( 18.6%)|    172 (100.0%)|
+------+----------------+----------------+----------------+

=======================================================================
Analysis of Binary Choice Model Predictions Based on Threshold =  .5000
-----------------------------------------------------------------------
Prediction Success
-----------------------------------------------------------------------
Sensitivity = actual 1s correctly predicted                     38.182%
Specificity = actual 0s correctly predicted                     90.598%
Positive predictive value = predicted 1s that were actual 1s    65.625%
Negative predictive value = predicted 0s that were actual 0s    75.714%
Correct prediction = actual 1s and 0s correctly predicted       73.837%
-----------------------------------------------------------------------
Prediction Failure
-----------------------------------------------------------------------
False pos. for true neg. = actual 0s predicted as 1s             9.402%
False neg. for true pos. = actual 1s predicted as 0s            61.818%
False pos. for predicted pos. = predicted 1s actual 0s          34.375%
False neg. for predicted neg. = predicted 0s actual 1s          24.286%
False predictions = actual 1s and 0s incorrectly predicted      26.163%
=======================================================================
```

```
--> dstat; rhs = aug, tb3, rwinsa, sdark,vh2bz, npwin; output =2 $
Descriptive Statistics
All results based on nonmissing observations.
===============================================================================
Variable      Mean        Std.Dev.      Minimum       Maximum        Cases Missing
===============================================================================
-------------------------------------------------------------------------------
All observations in current sample
-------------------------------------------------------------------------------
AUG      |  .625000E-01  .242645      .000000      1.00000          208       0
TB3      |  .134615      .342136      .000000      1.00000          208       0
RWINSA   |  9.33797      10.8166      .000000      71.0000          181      27
SDARK    |  .581395E-01  .234690      .000000      1.00000          172      36
VH2BZ    |  .171271      .377790      .000000      1.00000          181      27
NPWIN    |  .790055      .408399      .000000      1.00000          181      27

Correlation Matrix for Listed Variables

              AUG      TB3    RWINSA    SDARK    VH2BZ    NPWIN
     AUG  1.00000  -.10213  -.05405  -.06804  -.12841  -.03164
```

```
     TB3  -.10213  1.00000   .06842  -.01677   .00994   .01205
  RWINSA  -.05405   .06842  1.00000   .21531   .09788   .03398
   SDARK  -.06804  -.01677   .21531  1.00000   .27132   .06387
   VH2BZ  -.12841   .00994   .09788   .27132  1.00000   .08672
   NPWIN  -.03164   .01205   .03398   .06387   .08672  1.00000
```

```
--> create; elasrwsa = (1-proute)*0.025*rwinsa $
--> dstat; rhs =elasrwsa$
Descriptive Statistics
All results based on nonmissing observations.
===============================================================================
Variable     Mean        Std.Dev.      Minimum      Maximum        Cases Missing
===============================================================================
------------------------------------------------------------------------------
All observations in current sample
------------------------------------------------------------------------------
ELASRWSA|  .135723      .145550      .000000      .657624          172      36
```

Multinomial Logit Model

```
--> nlogit;lhs=x1;choices=PDO,POSUN,FINJ;model:
    u(PDO)=augpd*aug+tb3pd*tb3+rwinsapd*rwinsa/
    u(POSUN)=POSUN*one+belowz*belowz/
    u(FINJ)=FINJ*one+logTRvrf*logtrvr+darkf*dark+belowz*belowz
    ;effects: aug(PDO)/tb3(PDO)/rwinsa(PDO)/
    belowz(POSUN)/belowz(FINJ)/
    logtrvr(FINJ)/dark(FINJ)$
+----------------------------------------------+
| Discrete choice and multinomial logit models|
+----------------------------------------------+

+-------------------------------------------------------+
|WARNING:   Bad observations were found in the sample. |
|Found  27 bad observations among     208 individuals. |
|You can use ;CheckData to get a list of these points. |
+-------------------------------------------------------+

Normal exit from iterations. Exit status=0.
+----------------------------------------------+
| Discrete choice (multinomial logit) model    |
| Maximum Likelihood Estimates                 |
| Model estimated: Mar 20, 2013 at 10:24:58AM.|
| Dependent variable              Choice       |
| Weighting variable                None       |
| Number of observations             181       |
| Iterations completed                 6       |
| Log likelihood function      -128.3480       |
| Number of parameters                 8       |
| Info. Criterion: AIC =         1.50661       |
|   Finite Sample: AIC =         1.51123       |
| Info. Criterion: BIC =         1.64798       |
| Info. Criterion:HQIC =         1.56392       |
| R2=1-LogL/LogL*  Log-L fncn  R-sqrd  RsqAdj |
| Constants only    -148.1688  .13377  .11420 |
| Chi-squared[ 6]          =     39.64143      |
```

```
| Prob [ chi squared > value ] =   .00000     |
| Response data are given as ind. choice.     |
| Number of obs.=   208, skipped  27 bad obs. |
+---------------------------------------------+

+---------------------------------------------+
| Notes No coefficients=> P(i,j)=1/J(i).      |
|       Constants only => P(i,j) uses ASCs    |
|         only. N(j)/N if fixed choice set.   |
|         N(j) = total sample frequency for j |
|         N    = total sample frequency.      |
|       These 2 models are simple MNL models. |
|       R-sqrd = 1 - LogL(model)/logL(other)  |
|       RsqAdj=1-[nJ/(nJ-nparm)]*(1-R-sqrd)   |
|         nJ   = sum over i, choice set sizes |
+---------------------------------------------+
+--------+--------------+----------------+--------+--------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]|
+--------+--------------+----------------+--------+--------+
 AUGPD   |   -2.37723186       .66526106    -3.573   .0004
 TB3PD   |   -1.31388699       .50544268    -2.599   .0093
 RWINSAPD|    -.02919346       .01590340    -1.836   .0664
 POSUN   |   -2.29950051       .31630729    -7.270   .0000
 BELOWZ  |     .91284405       .39539160     2.309   .0210
 FINJ    |   -2.79569939       .40896325    -6.836   .0000
 LOGTRVRF|     .20479711       .07825527     2.617   .0089
 DARKF   |   -1.86398707       .78633663    -2.370   .0178
```

```
--> dstat; rhs = aug, tb3, rwinsa, belowz, dark, logTRvr; output =2 $
Descriptive Statistics
All results based on nonmissing observations.
===============================================================================
Variable     Mean        Std.Dev.     Minimum      Maximum       Cases Missing
===============================================================================
------------------------------------------------------------------------------
All observations in current sample
------------------------------------------------------------------------------
AUG      |  .625000E-01  .242256      .000000      1.00000          624       0
TB3      |  .134615      .341586      .000000      1.00000          624       0
RWINSA   |  9.33797      10.7966      .000000      71.0000          543      81
BELOWZ   |  .276243      .447551      .000000      1.00000          543      81
DARK     |  .317308      .465802      .000000      1.00000          624       0
LOGTRVR  |  1.70004      2.79506      .000000      6.63379          543      81

Correlation Matrix for Listed Variables

             AUG       TB3    RWINSA   BELOWZ     DARK   LOGTRVR
     AUG  1.00000  -.10078  -.03892  -.17186  -.00102   .03103
     TB3  -.10078  1.00000   .06520  -.03091  -.09320   .01001
  RWINSA  -.03892   .06520  1.00000   .21865   .04104   .08110
  BELOWZ  -.17186  -.03091   .21865  1.00000   .09437  -.02210
    DARK  -.00102  -.09320   .04104   .09437  1.00000   .04058
 LOGTRVR   .03103   .01001   .08110  -.02210   .04058  1.00000
```

APPENDIX D

TEST OF DIFERRENCE IN THE PROBABILITY OF SEVERE CRASHES

According to a descriptive analysis of the used data, 470 crashes occurred on the I-80 corridor under adverse weather conditions (e.g. temperature below freezing or precipitation events). Of those crashes, only 18 resulted in a fatal or major injury (severe) outcome. This corresponds to a probability of $\hat{p}_{80} = \frac{18}{470} = 0.038.$

On the other hand, 60 crashes occurred on the US-34 corridor under adverse weather conditions. Of those crashes only 4 resulted in a fatal or major injury (severe) outcome. Thus the corresponding probability is $\hat{p}_{34} = \frac{4}{60} = 0.067.$

However, if *X* is the random binary variable (takes on values 0 or 1) associated with whether a crash under adverse weather conditions had a severe outcome, then *X* can be considered as a Bernoulli variable, since it can be associated with a Bernoulli trial of two possible outcomes, such as fatal or major injury outcome or outcome of lower severity (Miller & Miller, 2004). Let denote the p-value of this Bernoulli variable as p_X. Note also that the p-value of a Bernoulli random variable is actually the probability that the variable equals one.

Based on the above context, one could say that in this study, there are 470 Bernoulli variables associated with a severe outcome under adverse weather conditions on the I-80 corridor and 60 Bernoulli variables on the US-34 corridor.

An estimator of p_X for n Bernoulli variables can be given by the formula: $\hat{p}_X = \frac{1}{n}\sum_{k=1}^{n} X_k$. Thus, for the I-80 corridor: $\hat{p}_{80} = \frac{1}{470}\sum_{k=1}^{470} X_k = 0.038$ and for the US-34 corridor: $\hat{p}_{34} = \frac{1}{60}\sum_{k=1}^{60} X_k = 0.067.$

Moreover, it can be proven by incorporating the Central Limit Theorem that the mean and variance of $\hat{p}_X$ are $\mu_{\hat{p}_X} = E(\hat{p}_X) = \frac{1}{n}\sum_{k=1}^{n} X_k$ and $\sigma^2_{\hat{p}_X} = \frac{\sigma^2_X}{n}$, where: $\sigma^2_X = p_X(1 - p_X)$.

In the case of the two corridors, n is large enough (greater than 30) in order to invoke the Central Limit Theorem, thus:

$$\mu_{\hat{p}_X80} = 0.038 \text{ and } \sigma^2_{\hat{p}_X80} = \frac{p_X(1-p_X)}{n} = \frac{0.038(1-0.038)}{470} = 0.000007$$

$$\mu_{\hat{p}_X34} = 0.067 \text{ and } \sigma^2_{\hat{p}_X34} = \frac{p_X(1-p_X)}{n} = \frac{0.067(1-0.067)}{60} = 0.001$$

In order to examine that the probability of a severe crash during adverse weather conditions is significantly larger on the US-34 corridor than the I-80 corridor, one could perform a simple hypothesis test, such as:

$$H_o\text{: } \mu_{\hat{p}_X80} - \mu_{\hat{p}_X34} \geq 0 \qquad \text{vs.} \qquad H_\alpha\text{: } \mu_{\hat{p}_X80} - \mu_{\hat{p}_X34} < 0$$

An appropriate test for the aforementioned hypotheses is the Welch's t test, which is an approximation to the t-test (Otto & Longnecker, 2010). The test is performed as follows:

Test Statistic: $$t' = \frac{\left(\mu_{\hat{p}_X80} - \mu_{\hat{p}_X34}\right) - 0}{\sqrt{\frac{\sigma^2_{\hat{p}_X80}}{n_{80}} + \frac{\sigma^2_{\hat{p}_X34}}{n_{34}}}}$$

Rejection Region: reject H$_o$ if $t' \leq -t_{\alpha}$, with degrees of freedom equal to:

$$df = \frac{(n_{80} - n_{34})}{(1-c)^2(n_{80}-1)+c^2(n_{34}-1)}, \text{ where: } c = \frac{\frac{\sigma^2_{\hat{p}X80}}{n_{80}}}{\frac{\sigma^2_{\hat{p}X80}}{n_{80}}+\frac{\sigma^2_{\hat{p}X34}}{n_{34}}}$$

By applying the corresponding numbers of the I-80 and US-34 corridors, the values of the previous formulas are: $t' = -7$ and $df = 61$. By using a t-distribution table it can be inferred that H$_o$ should be rejected at any acceptable α-level (p-value < 0.0005).

In conclusion, the probability of having a fatal or major injury crash under adverse weather conditions for a US route is significantly larger than the corresponding probability for an Interstate route at any acceptable α-level. Equivalently, adverse weather conditions seem to be associated with lower probability of a severe crash on an Interstate than on a route of lower classification.

APPENDIX E

CALCULATION OF THE PROBALITY OF A FATAL CRASH

The probability of a crash that occurred under temperature below 0^{o}C on a US route to result in a fatal injury outcome is given by Equation 4.13. Based on Table 5.5, the values of the utility functions for each of the three outcomes are (assuming that the values of all the other variables are equal to 0):

$$V_{PDO} = \beta_{\iota,PDO} X_{\iota,PDO} = 0$$

$$V_{POSUN} = \beta_{\iota,POSUN} X_{\iota,POSUN} = -2.300 + 0.913 = -1.387$$

$$V_{F} = \beta_{\iota,F} X_{\iota,F} = -2.796 + 0.913 = -1.883$$

Thus, from Equation 4.13:

$$P(F) = \frac{EXP(-1.883)}{EXP(0) + EXP(-1.387) + EXP(-1.883)} = 0.108$$

If the crash had occurred under different weather conditions (i.e., temperature higher than 0^{o}C), then the values of the utility functions would be:

$$V_{PDO} = \beta_{\iota,PDO} X_{\iota,PDO} = 0$$

$$V_{POSUN} = \beta_{\iota,POSUN} X_{\iota,POSUN} = -2.300$$

$$V_{F} = \beta_{\iota,F} X_{\iota,F} = -2.796$$

And thus the probability to be of a fatal outcome would be:

$$P(F) = \frac{EXP(-2.796)}{EXP(0) + EXP(-2.300) + EXP(-2.796)} = 0.053$$