University of Alberta

Investigating and Modeling Traffic Collision Frequency and Possibility for Edmonton

by

Gurjeet Singh Shaheed

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science
in
Transportation Engineering

Department of Civil and Environmental Engineering

DEDICATION

I would like to dedicate this dissertation to my parents for their inseparable support and prayers. My father, Darshan Singh Shaheed, in the first place is the person who put the fundament of my learning character, showing me the joy of intellectual pursuit ever since I was a child. My mother, Harvinder Kaur Shaheed, is the one who sincerely raised me with her care and love. Words fail me to express my appreciation to my brother Gurmeet Singh Shaheed. I would not have finished this degree without his love, companionship and encouragement. In my life, I personally feel very lucky to have so much love and support from near and dear ones. Thank you, Lord for always being there for me.

ABSTRACT

This study was conducted to investigate and model the high traffic collision frequencies in the City of Edmonton, Canada. Consistent collision spikes were observed on Fridays compared to the other days of the week. The first Negative Binomial model was formulated to establish a relation between the collision frequency and the independent variables. The second Multinomial logistic regression model was formulated to examine the probability of age categories and gender involved in collision for each day of week considering collision has happened.

The proposed collision prediction models were found good. They could provide a realistic estimate of expected collision frequency and properties of collision for a particular day as a function of number of hours of daylight, number of hours of snowfall, visibility, age and gender. It is hoped that predicted collision frequency will help the decision maker to quantify traffic safety of Edmonton and improve the scenario.

ACKNOWLEDGMENTS

I am deeply indebted to my supervisor, Dr. Zhi-Jun (Tony) Qiu, for the guidance and support that he so graciously offered me. I attribute the level of my Masters degree to his encouragement and effort and without him this thesis, too, would not have been completed or written. One simply could not wish for a better or friendlier supervisor. On more than one occasion, his commitment to the completion of my thesis ensured that I stayed the course.

I would also like to thank those who offered their expertise in model formulation, namely Md. Hadiuzzaman and Md. Ahsanul Karim. Klebert Machado provided good arguments about the logic and different causes of traffic collisions and model formulation. I am appreciative of all their cooperation and contribution to my understanding of their experiences.

Many thanks go in particular to Dr. Marwan El-Rich, Department of Civil and Environmental Engineering and Dr. Bonnie M. Dobbs, Department of Family Medicine for their critical feedback and participation in the supervisory committee process.

I would like to give thanks to Office of Traffic Safety group of the City of Edmonton for providing the Motor Vehicle Collision data for this project.

Lastly, I offer my regards and blessings to all of those who supported me in any respect during the completion of the project.

TABLE OF CONTENTS

LIST OF TABLES

List of Figures

CHAPTER 1. INTRODUCTION

This chapter illustrates the geographic information for the City of Edmonton, Canada. It provides information about the present collision status and motivation for the study. Traffic collision and collision prediction models are defined while presenting traffic collision statistics of different regions.

1.1 Background

The City of Edmonton is the capital of the province of Alberta, Canada. It is a hub of Canada's sixth largest census metropolitan area [1]. It encompasses a land area of 700 square kilometres, with a population of 782,439 on April 1, 2009, based on the municipal Census [2]. The map for the City of Edmonton is shown in Figure 1.1. The number of people driving in the city is increasing day-by-day due to the continuous growth of the city. A total of 3,100 kilometres of roads exist within the city to accommodate the 380,475 vehicles registered in Edmonton [3]. In addition to the increase in vehicular traffic, increased collision rates also have been observed each year in the city. The increase in traffic collisions results in challenges to planning, designing and monitoring the roadway network of the City of Edmonton, activities which are associated with the goals of fostering a safer and efficient road environment. The increase in traffic collisions underscores the importance of investigating and modeling the high traffic collisions for the City of Edmonton in order to meet these goals. Using these predicted traffic collisions, necessary countermeasures can be implemented in the city to improve the existing scenario.

Based on data from the Edmonton Police Service, there are seven high collision locations in Edmonton, one in North Edmonton, five in the Downtown area, nine in South East Edmonton, four in South the West Edmonton, and four in West part of the city [4].

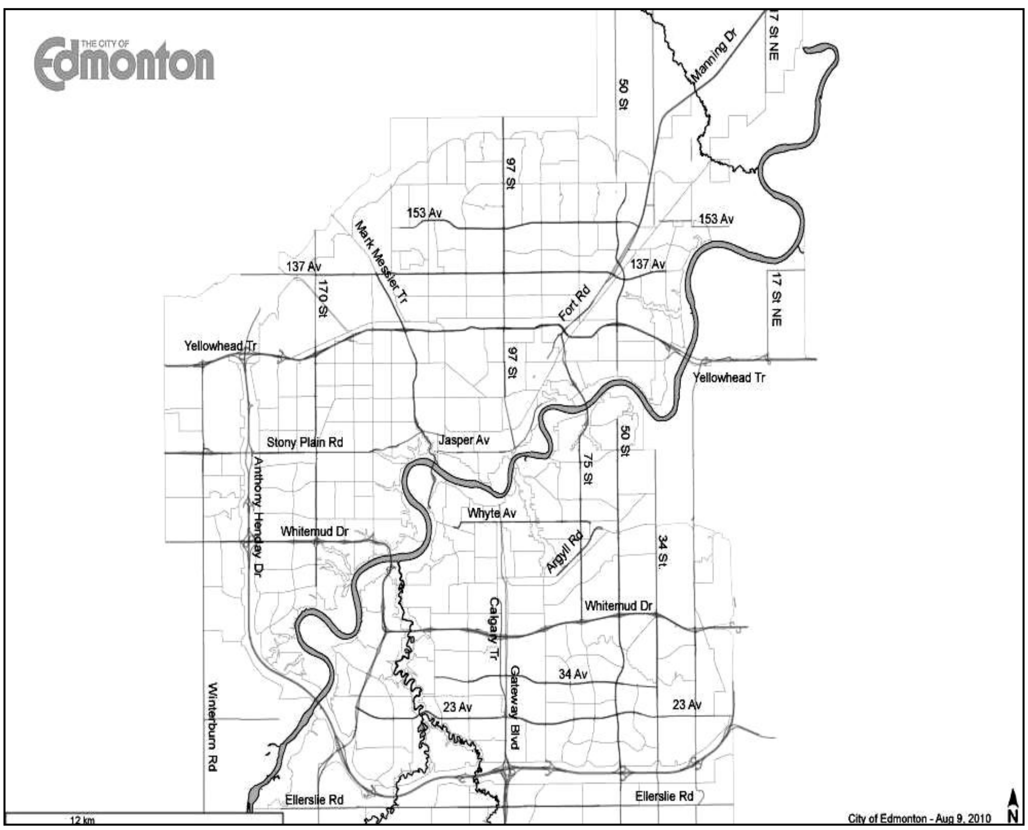


FIGURE 1.1: THE CITY OF EDMONTON MAP

Source: The City of Edmonton [5]

1.2 Traffic Collision and Safety

In general, an event involving two or more objects hitting each other from any direction is typically called as a *Collision*. According to the City of Edmonton's Office of Traffic Safety, a reportable traffic collision is defined as a collision which results in the property damage and/or injury of greater than or equal to $1000. All collisions occur on public roads (and do not occur on any private property) with at least one motorized vehicle involved in the collision occurred [6].

Traffic safety can be measured in terms of the number and rate of traffic collisions [7]. To improve traffic safety is a challenge for both traffic administrators and practitioners. Improved traffic safety can be achieved by upgrading planning, designing, construction and operations and maintenance of transportation networks. One of the methods used to improve traffic safety is the prediction of traffic collisions and taking necessary counter measures at the right time. Prediction of traffic collision is, however challenging. Also, to improve traffic safety three elements influencing traffic operations are considered: the driver, the vehicle and the roadway [8].

1.3 Traffic Collision Statistics

Traffic safety is a worldwide problem with over 500 million cars and trucks in use. Each year more than 500,000 people die in motor vehicle collisions, with approximately 15 million people are injured [9].

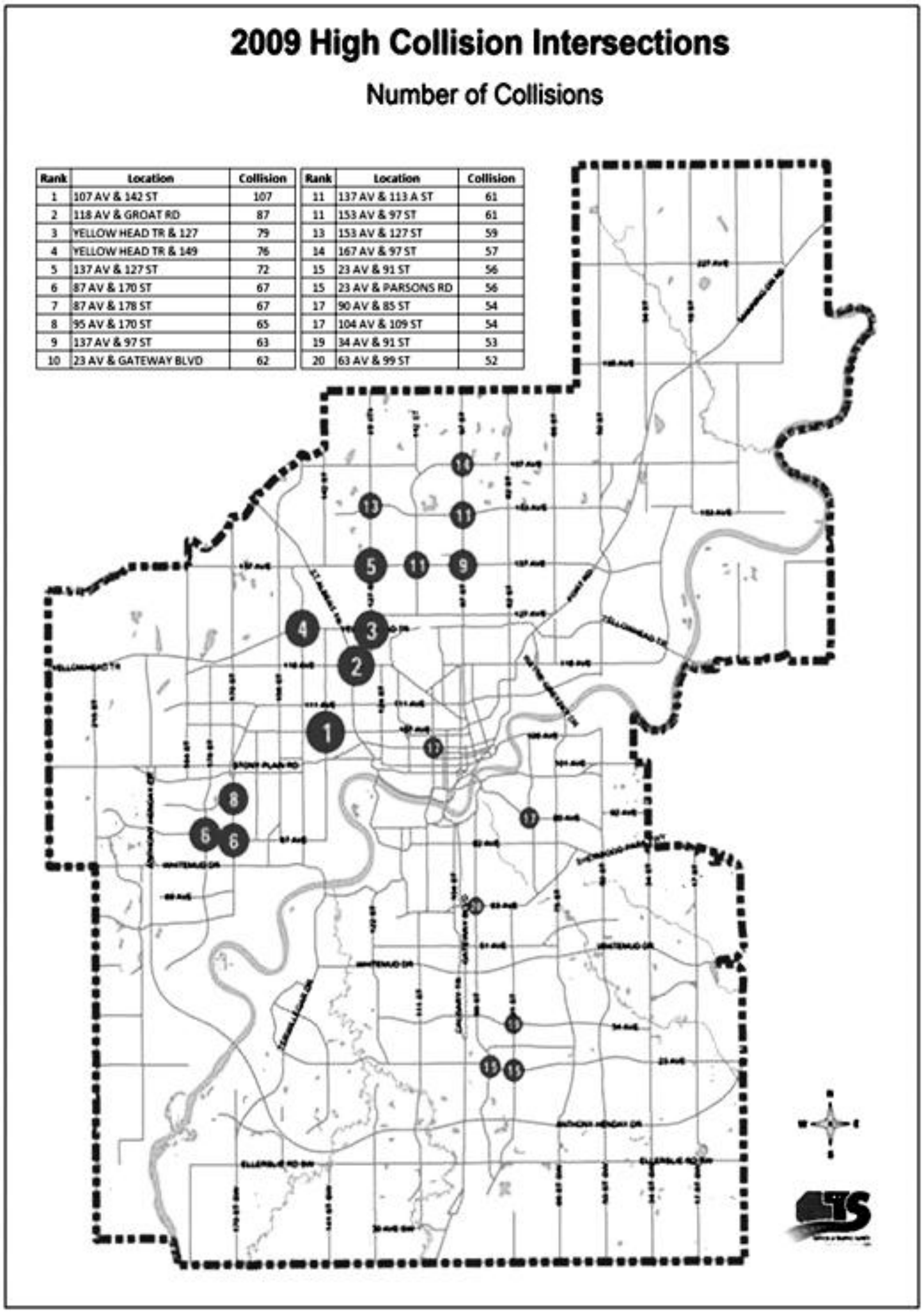
In the United States, after cardiac diseases and cancer, the most significant cause of years of potential life lost are *motor vehicle collisions*. Motor vehicle collisions are prominent cause of death for the people between the ages of 1 to 34 years [9]. According to the report on traffic safety facts from the United States Department of Transportation, there were an estimated 5,811,000 traffic collisions in 2008. An average of 102 people in the United States of America died each day in motor vehicle collisions (i.e. one in every minute) [10].

In Canada, road traffic fatalities have declined by 32.5 percent since 1987. The latest statistics indicates that in 2007, about 52.2 percent of those who died were motor vehicle drivers. The total number of people killed and injured on Canadian roads was found to be 2,469 and 138,470 respectively [11].

During 2008, 158,055 collisions were recorded on Alberta roadways, with an estimated fatal collisions 375 (0.2%) of those collisions reported [12]. Out of these 375 collisions, 26% of the fatal collisions involved unsafe speed and 22.5% of drivers had consumed alcohol before a fatal collision. According to Alberta Transportation, $4 billion is the estimated cost of traffic collision to society per year [13]. On average, 400 people are killed and 23,000 injured each year as a result of motor vehicle collisions in Alberta. Motor vehicle collisions are the leading cause of death for Alberta under the age of 30 [14].

From 2003 to 2009, there has been an increase in traffic collisions of approximately 23% [15]. A total of 21,350 annual collisions were observed in the City of Edmonton over the past few years. This has resulted results in 6,850 injury collisions approximately, 22 fatalities, and $78,000,000 in property damage per year [3]. Fifty-five percent approximately of these collisions have occurred at intersections, accounting for 66% of the total injuries and 42% of the total fatalities reported [3].

Figure 1.2 shows the high collision intersections in 2009 in the City of Edmonton as per the 2009 Motor Vehicle Collision report [15]. As shown in Figure 1.2, there are close to 20 high collision intersections. These intersections are ranked as per collision frequency on those particular intersections.



Rank	Location	Collision	Rank	Location	Collision
1	107 AV & 142 ST	107	11	137 AV & 113 A ST	61
2	118 AV & GROAT RD	87	11	153 AV & 97 ST	61
3	YELLOW HEAD TR & 127	79	13	153 AV & 127 ST	59
4	YELLOW HEAD TR & 149	76	14	167 AV & 97 ST	57
5	137 AV & 127 ST	72	15	23 AV & 91 ST	56
6	87 AV & 170 ST	67	15	23 AV & PARSONS RD	56
7	87 AV & 178 ST	67	17	90 AV & 85 ST	54
8	95 AV & 170 ST	65	17	104 AV & 109 ST	54
9	137 AV & 97 ST	63	19	34 AV & 91 ST	53
10	23 AV & GATEWAY BLVD	62	20	63 AV & 99 ST	52

FIGURE 1.2: THE HIGH COLLISION LOCATIONS IN CITY OF EDMONTON

Source: City of Edmonton [15]

1.4 Motivation for the Research

This research has been done to formulate the collision prediction models for the City of Edmonton using the past seven years (2003-2009) of collision data. Collision prediction models are statistically developed mathematical tools that relate the occurrence of traffic collisions to various traits or explanatory variables of locations and zones [16]. Preliminary analysis of the collision data shows a spike on Friday as compared to the other days of the week, a trend that is consistent from 2003 to 2009. As indicated in the literature review, there are many traffic collision contributing factors. This research was carried out to find out the most influential traffic collision contributing factors in the City of Edmonton. Results of the research will allow for the adoption of various counter measures to reduce traffic collisions.

1.5 Research Objectives

In order to make this research doable and deliverable at the end, the specific research objectives were finalized after the motivation of the research. The three major objectives of this research are to:

1. Investigate safety data pattern;
2. Identify and rank contributing factors;
3. Formulate collision prediction models.

1.6 Scope of Study

This study has used traffic collision data for the City of Edmonton as a whole. The time scope of this study was from 2003 to 2009. The collision data were provided by the Office of Traffic Safety of the City of Edmonton. This research has attempted to explain the consistently observed higher number of traffic collisions observed on Friday compared to other days of the week in the whole City of Edmonton.

1.7 Contents of Dissertation

Chapter 1 has provided an introduction to traffic collisions and the present collision scenario in North America. It also outlines the motivation for this research. The various types of collision prediction models used previously and their characteristics are discussed in Chapter 2. In Chapter 3, the different data sources used to formulate the collision prediction models and the detailed data analyses are described. Chapter 4 explains the detailed procedure used to scrutinize and model the daily traffic collision frequency. Chapter 5 outlines the different statistical measures used to examine the relative error between predicted collisions and the actual count of collisions for each year for both the models. Qualitative model accuracy test results are also incorporated in this chapter. The results and discussion of this work are summarized in Chapter 6.

CHAPTER 2. LITERATURE REVIEW

This chapter provides an overview of the different types of traffic collision prediction models used in the past. It incorporates the different patterns of traffic collisions such as holiday patterns and day of week patterns observed by different authors at different times.

2.1 Introduction

Prediction of motor vehicle collisions can significantly benefit traffic operations by taking appropriate counter measures at the right time. Several collision prediction models have been developed in the past for studying urban traffic safety at different locations. Different authors have used different collision prediction models based on their different study objectives.

A prediction model is formulated to develop a relationship between the collision frequency as the dependent variable and collision contributing factors as the independent variables. Historically, the most common independent variables used in prediction models are: section length, traffic volume, driver distribution, un-signalized intersection density, driveway density, pedestrian crosswalk density, number of traffic lanes, type of median and, type of land use.

2.2 Terminology and Key Definitions

- ✓ **COLLISION:** A collision is typically an event involving motor vehicles or other objects hitting each other. The City of Edmonton has defined a collision as an incident occurred on public roads involving fatality, injury and property damage only, with the cost of property damage incurred in the collision equal to or greater than $ 1000.

- ✓ **COLLISION PREDICTION MODEL:** It is a statistically developed mathematical tool that shows a relation between traffic collisions and different contributing factors.

- ✓ **CHRISTMAS:** The holiday which falls on December 25th.

- ✓ **FATAL TRAFFIC CRASH:** A collision in which one or more persons were killed as a result of the crash, with the death occurrence at the time of the incident or within 30 days of the incident.

- ✓ **HOLIDAY:** A term encompassing one of the 12 national and provincial holidays observed in Alberta, according to the Canadian Heritage Department, with Palm Sunday, Easter Sunday, Christmas Eve and New Year`s Eve considered for the data analysis.

- ✓ **INJURY:** It ranges severity from minimal (scrapes, bruises, complaint of pain) to major (treated and detained in hospital). If an injury is so severe that it might be the cause of death, it is called a fatal injury.

- ✓ **PROPERTY DAMAGE TRAFFIC COLLISION:** Any collision in which no person was killed or injured but property was damaged as a result of the incident.

- ✓ **PUBLIC ROADS:** As per the City of Edmonton, all urban and rural roads including City owned streets and provincially owned roads in the Edmonton area are termed as public roads.

- ✓ **SEVERITY:** This study considers fatality, injury and property damage only (PDO) as three measures of severity.

- ✓ **SUNRISE/SUNSET: S**unset and sunrise are the true rising and setting times of the sun computed by the NRC Herzberg Institute of Astrophysics (NRC-HIA).

2.3 Holiday Pattern of Previous Safety Data Analysis

A review of previous studies indicates that different travel patterns and traffic collision patterns are observed during the holiday periods. To study holiday collision patterns, a study was conducted by the Missouri traffic safety authorities. An annual report was published by the Missouri State Highway Patrol Statistical Analysis Centre and Public Information and Education Division in the year 2007. Each year, Missouri traffic safety authorities face problems due to an increase in the number of traffic collisions on holidays like Memorial Day, Fourth of July, Labor Day, Thanksgiving Day, Christmas and the New Year holiday [17]. Results indicate that Missouri State experiences different traffic collision statistics on holidays compared to the regular days.

The Missouri Traffic Safety Authority summarizes that most of the holiday periods start at the end of a full workday. Drivers take long trips, which result in fatigue due to lack of rest. This situation also results in reduced defensive driving abilities and increases the possibility of human error such as speeding, resulting in traffic collisions. Also, consumption of alcoholic beverages increases during holiday periods, increasing the chances of a traffic collision occurring.

During holiday periods, significant variations in traffic volumes were observed by Zhaobin Liu et al. 2008 [18]. However, this change in traffic volumes may be due to variation of road types, days of the week and holidays. If reconstruction is not reasonable to accommodate the peak volumes during short periods, an advanced traveler information system may be used to warn the public of the potential road congestion during holiday periods. The results of this study indicated that the peaking of holiday traffic has strong directional features. The authors also suggested the application of reversible lanes to accommodate the imbalanced peak holiday traffic volume. Unique holiday traffic characteristics also affect other traffic engineering aspects, such as traffic data imputation, signal timing, road safety and road maintenance.

2.4 Friday Pattern of Previous Safety Data Analysis

For the City of Montreal, Canada, the effects of rain, mean temperature and snow on traffic collisions were studied by Mircea-Paul et al. 1998 [19]. A significant impact of all the three variables on traffic collision was observed. This research was important to assist provinces and municipalities to reduce spending on winter snow cleaning allowing for a reduction in operating expenses. Their study was based on analysis by month, year and the entire study period from 1990 to 1992. Three time frames were used: monthly, annual and the entire study period.

This study noticed a significant number of collisions with increased snowfall and a spike on Friday with the lowest number of collisions on Sunday as compared to other days of the week. Similar to snowfall, rainfall was also perceived as positively correlated with the number of collisions. The analysis of traffic densities (available only for 1992) was carried out indicating a significant peak of traffic densities on Friday.

McGwin et al. 1999 also found a higher number of traffic collisions on Fridays, considering all Alabama State collision reports in 1996 [20]. The study analyzed the characteristics of traffic collisions among younger and older drivers. The results of this study suggested that the young drivers are more risk taking and their lack of skill contributes in traffic collisions. Older drivers have a natural tendency to avoid risk but their perception/reaction times are also slow due to illness which counterbalances each other. Similar trends of higher number of collisions on a particular day of week were found in King County, Washington State, 2005 [21]. The 2005 Traffic Safety Report of King County road services observed peaks of traffic collisions on Friday and Saturday. In 2006, the same peak was found to be only on Friday; however in 2007 the peak was shifted to Saturday. Also, the trend of a higher number of traffic collisions on Fridays and Saturdays was found in Ontario, Canada [22]. This study provides insight to the situational risks of young drivers considering the influence of passengers, time of day and day of week on collision rates. Results of this study indicate that 16-19 year old drivers have higher collision rates compared to 20-24 and 25-59 year old drivers.

Saskatchewan Government Insurance (SGI), Saskatoon in 2006, reported that the highest number of fatal collisions occurred in May, while the highest number of injury collisions occurred in March in 2006. The maximum number of collisions was found to be on Friday as compared to the other days of the week. The afternoon rush hour (3:00 pm to 6:00 pm) was reported as the most collision-prone time period. Drivers between 15 and 24 years of age were responsible for most of the collisions. It was observed that 35% of fatal collisions occurred due to drinking drivers compared to 14% of injury collisions. Another feature of the report was that collision victims who did not buckle up were 18 times more likely to be killed than those who buckled up. Some other human factors that contribute to casualty collisions in

Saskatchewan were driver inattention, inexperience/confusion, distraction, drinking and impairment [23].

2.5 Collision Prediction Models in Past Studies

Different modeling approaches have been proposed to predict traffic collision frequency. Most of them use mathematical models depending on their study objectives [24, 25 and 26]. Based on characteristics of traffic collision data, Poisson and Negative Binomial (NB) models to more complicated ones like Zero inflated models are used. Poisson and NB distribution or Poisson lognormal distributions have been used in past studies to account for the over dispersion in the collision data [27, 28 and 29]. These two models are reported to be good for statistical approximation of collision frequencies (providing a good fit to the data) as compared to the other models [30].

Nofal et al. 1997 studied the seasonal variation and weather effects on road traffic collisions in the city Riyadh for a period of 1989-1993 [31]. This study was based on time of day, lighting condition and prevalent weather conditions as major contributing factors for the seasonal variation. The majority of traffic collisions were found to be in the summer when there is more daylight. For periods of heavy traffic and intense sunlight (between 12 noon and 3 pm) collisions were more prominent. They found a positive correlation between traffic collisions and temperature. However, their study did not focus on daily variation of traffic collision.

Smith 2000 authored studies related to alcohol ingestion, driver fatigue and distraction from use of cellular phones. This research found that alcohol ingestion plays an important role in driver performance. To reduce collisions due to alcohol ingestion, the author suggests adjustment of the legal Blood Alcohol Content; discourage minors from purchasing alcohol, and use of sobriety checkpoints [32].

Sawalha and Sayed 2001 discussed the development of accident prediction models for estimating the safety performance of urban arterial roadways in Vancouver and Richmond, British Columbia, Canada. The study examined the effect of traffic and geometric variables on the safety of urban arterials. The results of the study indicated that traffic volume, section length, un-signalized intersection density, driveway density, pedestrian crosswalk density, number of traffic lanes, type of median and nature of land use have a significant impact on the occurrence of traffic collisions [33].

Miranda et al. 2005 investigated the relative performance of the three models: traditional negative binomial, Heterogeneous Negative Binomial (HNB), and Poisson lognormal model in the prediction of traffic collisions. They found that Poisson lognormal model and the HNB model produce better fits to the data than the traditional NB model. Limited data set is one of the limitations of this study [34].

Caliendo et al. 2007 compared three distributions, namely Poisson distribution, negative binomial distribution and negative multinomial distributions in the prediction of traffic collisions. The data set included traffic volumes and was limited to a four-lane median-divided motorway of Italy. The authors suggested that negative multinomial distribution is the most appropriate for modeling the longitudinal collision data. This research concluded that Poisson distribution is inappropriate for modeling random variation of the collision frequencies due to the over dispersion of the collision data. Increase in the power of explanation was found for both negative binomial and negative multinomial model when the over dispersion parameter in each section is assumed to be proportion to the section length [35].

2.6 Dummy Variables in Collision Prediction Models

For a specific day of the week dummy variables can be incorporated in the collision prediction model to capture the exposure of the days on traffic collision [36]. A dummy variable is a numerical variable used in the regression analysis to represent whether a set of conditions happens or not. These variables often are binary (0 or 1) variables. If the variable has a specific category, then the value of the dummy variable in the cell will be considered as 1, otherwise it will be considered as 0. It is observed from the previous studies that traffic conditions change each day of the week. Therefore, to capture the exposure of traffic conditions, dummy variables can be used in the collision prediction models. The dummy variable considers the value as 1 associated with specific day of the week to capture the exposure of that particular day; otherwise it takes a value of 0.

2.7 Summary

The development of the prediction models is a very challenging task due to the random nature of traffic collisions. In this chapter, studies are reviewed which indicate a similar pattern for collisions on holidays and on Friday. In this section, the different types of traffic collision prediction models with various collision contributing factors used in the previous studies are also outlined. Poission and Negative Binomial models are commonly used as collision prediction models. These two models have been found to be good for statistical approximation of collision frequencies. This chapter has provided insight into capturing exposure of the day using dummy variables.

CHAPTER 3. COLLISION DATA ANALYSIS FOR EDMONTON

This chapter introduces the different data sources (MVCIS, Sunrise/Sunset Data and National Climate Data and Information Archive) used to formulate collision prediction models. The data analysis shows the correlation between different factors considered for the model formulation. The investigation of the data encouraged us to go further and formulate the model using these three sources.

3.1 Introduction

To conduct this research, three different data sources were used, namely: 1) MVCIS (Motor Vehicle Collision Information System), 2) Sunrise/Sunset Data, 3) National Climate data and Information Achieve. All three data sets were categorized by the day of the year for a period of seven years (2003 to 2009). A combined master database was prepared to use in the model formulation. This chapter presents the detailed investigation and analysis of all three data sources.

3.2 MVCIS (Motor Vehicle Collision Information System)

In this research, data from the City of Edmonton's collision database known as MVCIS (Motor Vehicle Collision Information System) is used. This database is derived from police reports containing information on motor vehicle collisions occurring on public roads. The MVCIS database provides a complete story for all traffic collisions involving fatality, injury and property damage only. The cost of damage incurred in the collision is equal to or greater than $1000 [15]. The time scope of the study was January 01, 2003 to December 31, 2009. The MVCIS data are used to compute the collision frequency per day. Initially, the data were sorted on the day of the week, whether a holiday or not and long weekend for the analysis. The analyses are discussed later in this chapter.

One of the major collision contributing factors for different collision frequencies on different days of the week is Average Annual Weekday Traffic flow (AAWT). However, daily variation of traffic was not available for this research, limiting the daily pattern analysis. The MVCIS data detail the date and hour of each collision, but do not indicate whether it happened on a holiday or a long weekend. In this project, the pattern of the collisions on holidays, extended holidays and switched Mondays, was also investigated in comparison to the regular week days.

3.3 Sunrise/Sunset Data

The sunset/sunrise data are maintained by the National Research Council (NRC) Canada. These times and related data are computed by the NRC Herzberg Institute of Astrophysics (NRC-HIA). The number of daylight hours was calculated using sunrise/sunset timing. The Herzberg Institute of

Astrophysics uses the standard scientific formulae, adopted by the National Almanac Offices of the United States and the United Kingdom and used worldwide. Sunset and sunrise are the true rising and setting times of the sun. For each day, the hours of daylight gives information of daylight from sunrise to sunset on that particular day. The hours of illumination in the data provide information for each day the duration of daylight hours from sunrise to sunset, and the duration of civil twilight (summed for morning and evening), when sky illumination is present, and their total.

The appearance and disappearance of the upper limb of the sun, as observed at sea level on a refracted (apparent) sea horizon, represents the rising and setting times (respectively) of the sun in this data. Due to the irregularities of terrain, these theoretical times will only approximate the rising and setting times observed on land. Even on a perfect (sea) horizon, variations in the atmospheric temperature profile can cause the amount of atmospheric refraction of light to vary, such that observed rise and set times may deviate from the computed values by one or two minutes. The rise and set times are calculated to the nearest minute only, based on the average atmospheric refraction. If an event occurs twice on the same date, the program identifies only the first occurrence. Accuracy of the sunrise/sunset times are computed as ± 2 minutes [37].

3.4 National Climate Data and Information Archive

The data from National Climate Archives Online were downloaded as XML file through the *“Bulk Data”* tool located in the Navigation Options box. The national climate information is provided and maintained by the Environment Canada. The weather elements analyzed in this research were: rain, snow, visibility, and the state of the weather (e.g., thunderstorms, blowing snow, fog, smoke and haze, etc). The state of weather is considered to estimate the number of hours of snowfall and number of hours of rainfall on a particular day. This database corresponds to three basic sampling frequencies; a) Hourly; b) Daily and; c) Monthly.

In the hourly frequency of the data, if any of the 'state of weather' has snow in it we have considered it as a snowy hour. Using the hourly frequency of data, we computed the number of hours of snowfall per day. Considering the average value of a given 24 hourly value per day, we estimated the visibility corresponding to that particular day. Visibility refers to the distance that people can see through the atmosphere at ground level and clarity of the image they are looking at. The unit of visibility used by Environment Canada was kilometers [38].

3.5 Traffic Safety Data Analysis

The data for 2003 to 2009 were investigated and analyzed to determine the daily, weekly, monthly and yearly collision trend. The data were cleaned and filtered for each individual collision report.

3.5.1 Daily Collision Pattern

A chart (Figure 3.1) was prepared to study the collision trend between annual collision and days of the week from 2003 to 2009. Figure 3.1 shows variation of the total collision frequency by day of week (Monday through Sunday) per year from 2003 to 2009. It was observed from the Figure 3.1 that the number of collisions each year on Friday was higher as compared to the other days of the week. The collision frequency observed in year 2008 was reported as the highest from all the seven years. Consistently, the fewest number of collisions were on Sunday. On analyzing the data for all the seven years the following is evident:

- ✓ 2003 – For Mondays and Tuesdays, the number of collision remains about the same as (3349 and 3321 respectively). On Wednesday there is an increase observed from Monday and Tuesday with 3479 traffic collisions. A drop is recognized on Thursday with 3395 collisions, which is less than Wednesday but higher than Monday and Tuesday. A spike was witnessed on Friday with 3859 traffic collisions. The curve sharply decreases for the weekend with the minimum collision number seen on Sunday as 1985 and on Saturday as 2739 (much lower than other days of week).

✓ 2004 – The curve for 2004 follows exactly similar trend as of the 2003 curve. The only difference is that Thursday (3203) has higher number of traffic collisions as compared to the Monday (2962), Tuesday (2954) and Wednesday (3142). Friday found as highest with 3589 traffic collisions. The Saturday and Sunday follows a sharp decreasing trend with 2760 and 1978 traffic collisions respectively.

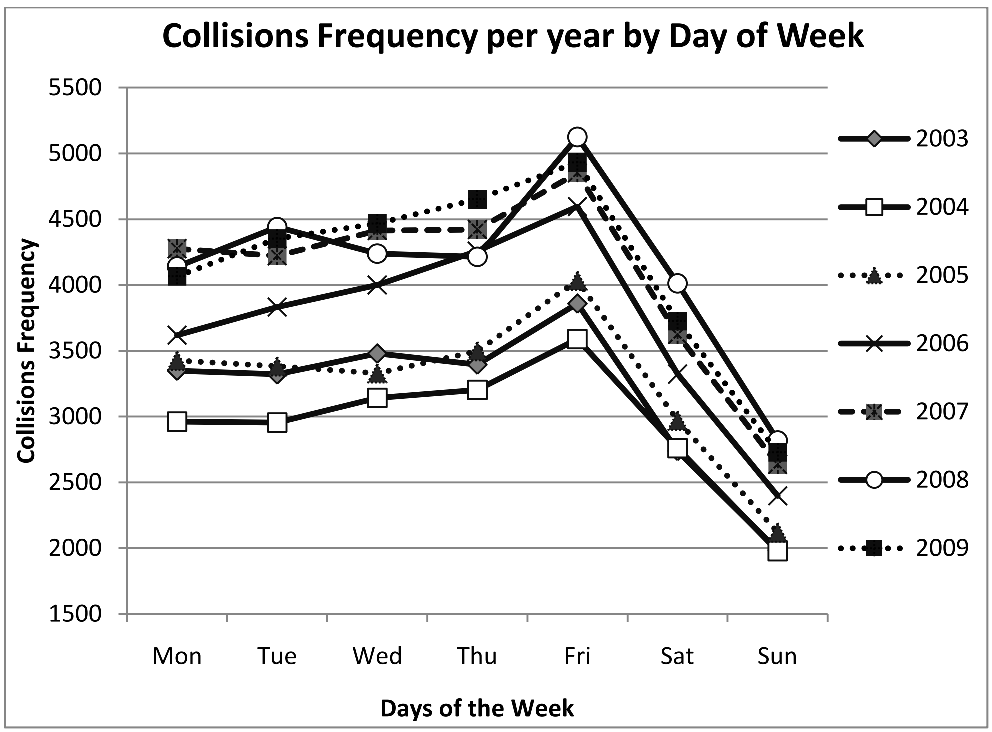


FIGURE 3.1: COLLISIONS FREQUENCY PER YEAR BY DAY OF THE WEEK

✓ 2005 – A decreasing trend was seen from Monday to Wednesday with 3426, 3383 and 3330 traffic collisions on Monday, Tuesday and

Wednesday respectively. The collision number increases on Thursday with 3498 and reaches maximum on Friday with 4034 traffic collisions. 2971 and 2121 traffic collisions were observed on Saturday and Sunday. Sunday was found to be the minimum compared to the other days of the week.

- ✓ 2006 – A direct increase was observed from Monday to Friday. 3619 traffic collisions on Monday, 3832 on Tuesday, 4001 on Wednesday, 4257 on Thursday, and 4596 on Friday were reported. Again, Friday was recorded as maximum in number compared with other days of the week. The lowest number of traffic collisions was found on Sunday with 2398 number. Saturday was reported with 3323 traffic collisions.

- ✓ 2007 – Tuesday is found lower than Monday with 4222 and 4276 traffic collisions respectively. Almost similar number of traffic collisions was observed on Wednesday and Thursday with 4413 and 4422 respectively. A peak was found on Friday with 4852 traffic collision. The number of traffic collisions decreases on Saturday and Sunday with 3624 and 2637 respectively.

- ✓ 2008 – Higher number of traffic collisions was observed on Tuesday as compared to Monday with 4441 and 4140 collision number respectively. Wednesday (4239) and Thursday (4214) were found almost similar in number. Friday was found to be most severe day with 5125 traffic collisions. Sunday was found least severe with 2819 traffic collision number. Saturday was reported with 4013 traffic collision in year 2008.

- ✓ 2009 – Traffic collisions reported on Monday, Tuesday, Wednesday, Thursday and Friday were 4066, 4352, 4468, 4651 and, 5125 respectively. An increasing trend was observed from Monday to Friday. Saturday and Sunday was found as 3725 and 2729. The differences of average traffic collisions from Monday to Thursday and Friday in percentage from 2003-2009 is 12.26%, 14.59%, 15.49%, 14.55%, 10.69%, 16.91%, 11.70% respectively.

TABLE 3.1: COLLISION FREQUENCY PER YEAR BY DAY OF THE WEEK

Day of Week	Year						
	2009	2008	2007	2006	2005	2004	2003
Mon	4066	4140	4276	3619	3426	2962	3349
Tue	4352	4441	4222	3832	3383	2954	3321
Wed	4468	4239	4413	4001	3330	3142	3479
Thu	4651	4214	4422	4257	3498	3203	3395
Fri	4931	5125	4852	4596	4034	3589	3859
Sat	3725	4013	3624	3323	2971	2760	2739
Sun	2729	2819	2637	2398	2121	1978	1985
Total	**28802**	**28991**	**28446**	**26026**	**22763**	**20588**	**22127**
Mon-Thu average	4354	4259	4333	3927	3409	3065	3386
Friday - Avg % diff.	11.70%	16.91%	10.69%	14.55%	15.49%	14.59%	12.26%

From 2003-2009, the average number of accidents per year from Monday to Thursday is 3819 and on Fridays is 4427 accidents per year, which are 13.72% greater than the average.

Figure 3.2 shows that collisions frequency on Friday is higher than average of collisions from Monday to Thursday consistently from 2003 to 2009. The value of chi-square test statistics for the difference of collisions on Friday and average of collisions from Monday to Thursday is significantly higher from the critical chi square value considering all the seven years ($p < 0.5$).

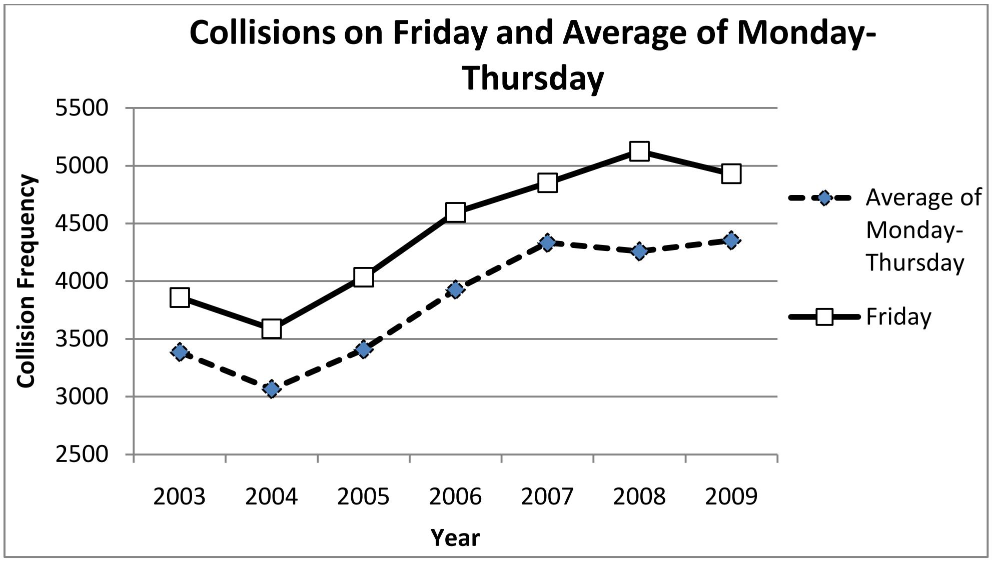


FIGURE 3.2: COLLISIONS FREQUENCY PER YEAR BY DAY OF THE WEEK

Traffic collisions on Friday were significantly higher at 95% confidence interval compared to the other days of the week ($X^2 = 12.6$, $p < 0.5$).

3.5.2 Weekly Collision Pattern

The collision data (MVCIS) were cleaned and filtered for every week to observe the variation of the collisions week by week. To see the weekly pattern of traffic collisions from 2003 to 2009, a chart was prepared for the average collision frequency each week. Each year the curve follows a u-shape from first week of the year to the last week of the year. It is observed from the Figure 3.3, that maximum number of collisions happened in the first and last few weeks of the year.

Fewer collisions were observed in the weeks of 25 to 30. Also, these weeks were observed as having higher number of daylight hours compared to the other weeks of the year. An increasing trend of traffic collisions from year 2003 to 2009 is also be perceived from the Figure 3.3.

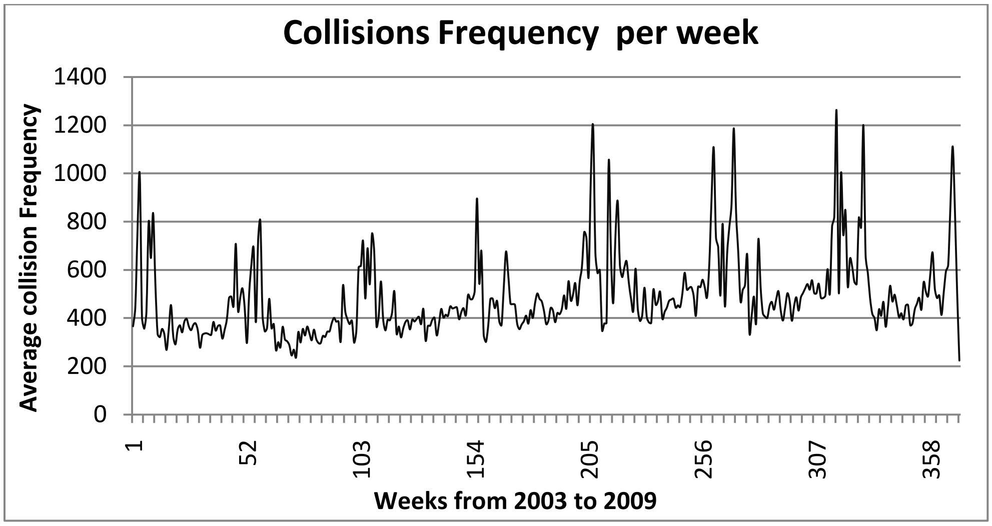


FIGURE 3.3: WEEKLY PATTERN OF COLLISION FREQUENCIES (2003-2009)

3.5.3 Monthly Collision Pattern

A similar trend was found from the monthly analysis of the collisions as shown in Figures 3.4 and 3.5. Two different graphs were used to present a clearer picture of the collision trends. Consistent spikes have been observed on Friday from January to December for all the years from 2003 to 2009. It was explored that month of January and December had significantly higher number of accidents as compared to the other months of the year. For all the months collision trend is similar to the one we found earlier (Figure 3.1) except for the month of April and December.

In December, high collision frequency is observed on Monday as compared to the other days of the week (Figure 3.5). In April, collisions were found more on Saturday instead of Friday. However, in the same months second most remarkable was Friday.

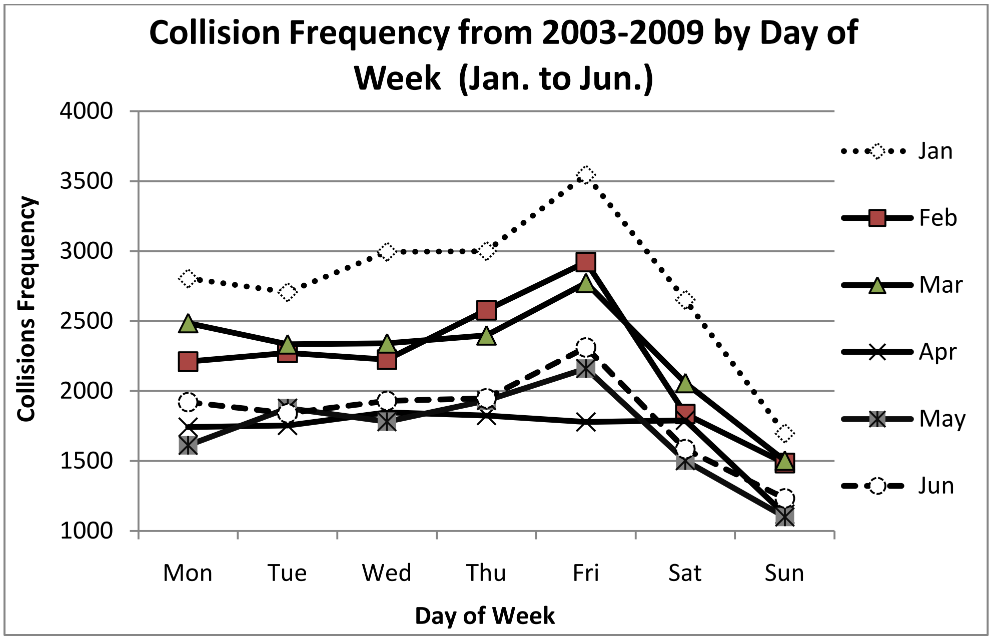


FIGURE 3.4: COLLISION FREQUENCIES BY DAY OF WEEK FROM JANUARY TO JUNE

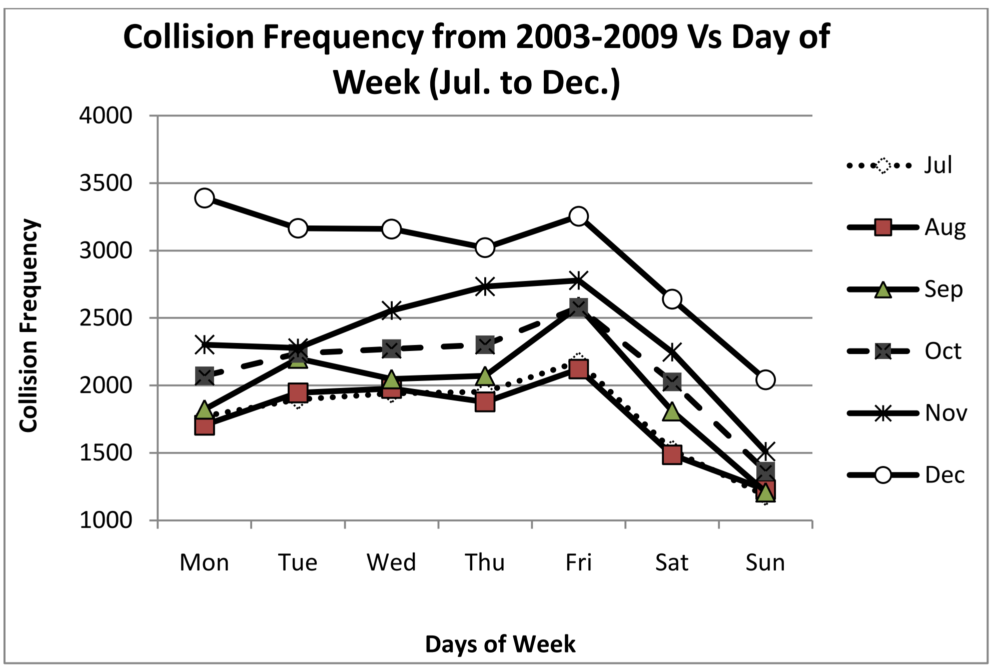


FIGURE 3.5: COLLISION FREQUENCIES BY DAY OF WEEK FROM JULY TO DECEMBER

For all the months the lowest number of traffic collisions was found to be on Sunday. On analyzing the MVCIS data following is perceived:

- ✓ January – Collision frequency observed in January is much higher than the other months of the year (Figure 3.4). Wednesday (2995) and Thursday (3000) was found to be almost similar. Mondays is higher than Tuesday with 2804 and 2706 traffic collisions respectively. Friday is identified as the most severe with 3545 collisions. The least number of traffic collisions was noticed on Sunday with 1697 traffic collisions. Saturday was noticed as the second lowest with 2652 collision number.

- ✓ February – Higher number of traffic collisions is noticed on Tuesday (2272) compared to Monday and Wednesday with 2210 and 2225 traffic collisions respectively. The highest traffic collisions were observed on Friday with 2921 and Thursday with 2578. Collisions drop on Saturday and Sunday with 1836 and 1485 traffic collision number respectively.

- ✓ March – Tuesday, Wednesday and Thursday were observed as resembling each other with 2334, 2341 and, 2398 collision number respectively. Monday was remarked as 2486 traffic collisions. Traffic collisions noticed on Friday, Saturday and Sunday were 2773, 2057 and 1502.

- ✓ April – A difference of only 10 collisions was noticed for Monday and Tuesday with 1742 and 1754 traffic collisions respectively. Wednesday was distinguished with 1847 collision number and the highest from Monday through Sunday. 1825 traffic collision were deduced from the collision data on Thursday. Friday and Saturday were not much different with 1778 and 1791 collisions. Sunday remains lowest with 1119 traffic collisions.

- ✓ May – An increasing trend was observed from Monday (1612) to Tuesday (1876). Again a decrease in collision number was identified on Wednesday with 1781 traffic collisions. Collisions increase till Friday (2160) from Thursday (1933) and then drop on Sunday (1100).

- ✓ June – Monday, Wednesday and Thursday were almost similar with 1920, 1929 and 1947 traffic collisions. 1841 collisions were noticed on Tuesday. A peak of 2312 traffic collisions was observed on Friday. Collisions decrease from Saturday to Sunday with 1585 and 1232 collisions respectively.

- ✓ July – An increasing trend was observed from Tuesday to Friday with 1895, 1942, 1952 and 2176 traffic collisions respectively. Monday was remarked with 1920 collisions. Collisions decrease from Saturday with 1522 to Sunday with 1181 traffic collisions.

- ✓ August – Traffic collisions increases from Monday to Wednesday with 1705 collisions on Monday, 1945 on Tuesday and 1976 on Wednesday. Then collisions drop on Thursday with 1878 traffic collisions. A peak of 2121 collisions was noticed on Friday. Least number of collisions (1226) was found on Sunday. Saturday was remarked with 1486 traffic collisions.

- ✓ September – Collisions decrease from Tuesday to Wednesday with 2200 and 2048 collisions. 1823 traffic collisions were recorded on Monday. Collision number perceived on Thursday and Friday were 2072 and 2586 respectively. Traffic collisions found on Saturday were 1810 and on Sunday were 1209

- ✓ October - Collision trend increases from Monday to Friday. Monday was noticed with 2071 collisions. Tuesday, Wednesday and Thursday were identified much the same with 2238, 2272 and 2302 traffic collisions. 2580 collision peak was remarked on Friday. 2027 and 1364 traffic collisions were recorded on Saturday and Sunday respectively.

- ✓ November – Monday and Tuesday were noticed to be almost similar with 2303 and 2278 traffic collisions respectively. The November curve follows a shape similar to bell. The collisions increase up to Friday with 2779 traffic collisions. Sunday is lowest with 1510 collisions.

TABLE 3.2: AVERAGE COLLISION FREQUENCY PER MONTH BY DAYS OF WEEK

Months	Days of Week						
	Mon	Tue	Wed	Thu	Fri	Sat	Sun
January	2804	2706	2995	3000	3545	2652	1697
February	2210	2272	2225	2578	2921	1836	1485
March	2486	2334	2341	2398	2773	2057	1502
April	1742	1754	1847	1825	1778	1791	1119
May	1612	1876	1781	1933	2160	1502	1100
June	1920	1841	1929	1947	2312	1585	1232
July	1772	1895	1942	1952	2176	1522	1181
August	1705	1945	1976	1878	2121	1486	1226
September	1823	2200	2048	2072	2586	1810	1209
October	2071	2238	2272	2302	2580	2027	1364
November	2303	2278	2556	2733	2779	2247	1510
December	3390	3166	3160	3022	3255	2640	2042

- ✓ December – A different collision pattern for December was realized from the data analysis. Traffic collision number was the highest on Monday. Friday is the second highest with 3255. Tuesday and Wednesday remains much the same with 3166 and 3160 traffic collisions respectively. 3022 collisions were noticed on Thursday. Collisions drop from Saturday with 2640 to Sunday with 2042 traffic collisions.

The data for daylight hours were extracted from Sunrise/Sunset data maintained by National Research Council - Herzberg Institute of Astrophysics, Canada. We aggregated this data from 2003 to 2009. After this, a chart (Figure 3.6) between the average number of hours of daylight and month of the year was plotted. The trend of average number of hours of daylight for different days of the week was noticed to be very similar by months of the year. It is found that the curves are almost overlapping each other. The data analysis shows 18 hours of daylight in June and 9 hours in January and December (i.e. almost 50% decreasing).

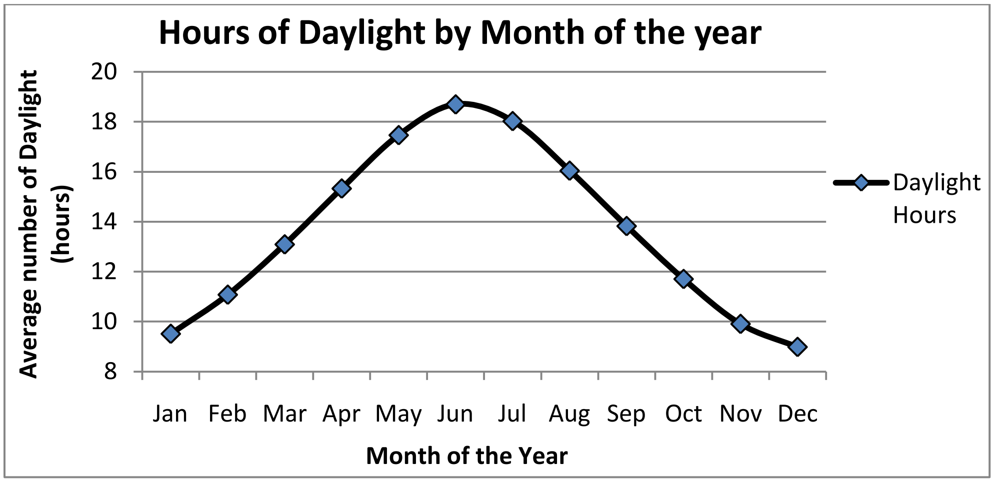


FIGURE 3.6: AVERAGE HOURS OF DAYLIGHT IN EACH MONTH (2003-2009)

Figure 3.7 illustrates the monthly variation of total number of collisions from 2003 to 2009 by the day of week. The data analysis shows that number of collisions follows a trend of U-shape from January to December. For the months of January and December the total numbers of collisions were greatest with significantly fewer in the months of April to August.

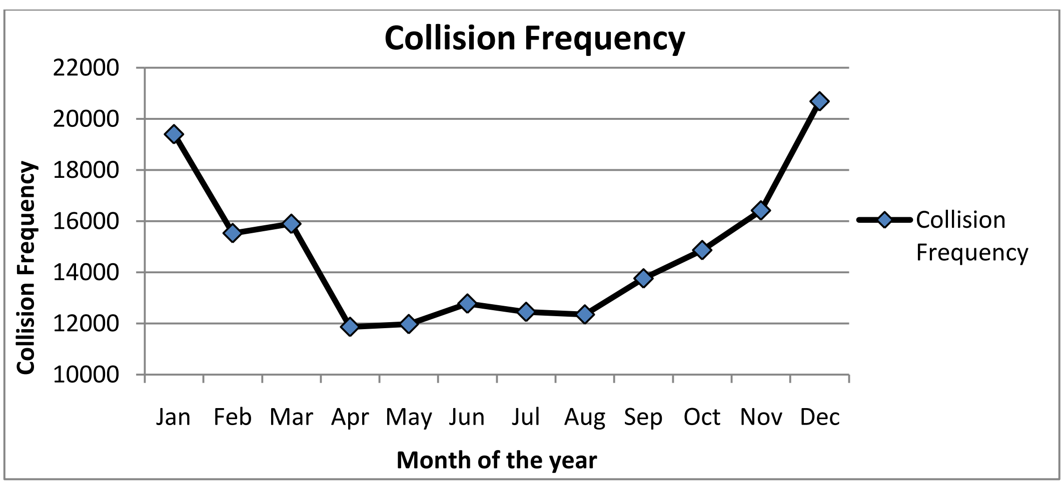


FIGURE 3.7: COLLISION FREQUENCIES IN EACH MONTH (2003-2009)

There is a strong negative correlation ($r = -0.88$, $p < 0.5$) between the total number of collisions and number of hours of the daylight as perceived from the Figures 3.6 and 3.7. The negative value of correlation coefficient indicates that there is a negative linear relationship between collision frequency and number of hours of daylight. The pattern of association indicates that if number of hours of daylight increase, collision frequency decreases. It motivates us to consider the number of hours of daylight as one of the variables in the collision prediction model.

3.5.4 Hourly Collision Pattern

The hourly variation of traffic collision was investigated to see the general trend of traffic collisions. Two curves were plotted to see the difference between the Wednesday and Friday. Wednesday was considered as a base day assuming that Wednesday is similar to the other weekdays. From Figure 3.8 it is perceived that traffic collisions in the afternoon increase significantly as compared to the morning. After that it decreases till evening. The Friday curve is observed to be higher than the Wednesday from afternoon till night. This shows that more number of collisions happened in the afternoon period on Friday as compared to the other days of the week. The Figure 3.8 shows traffic collision variation for the data from 2003 to 2009.

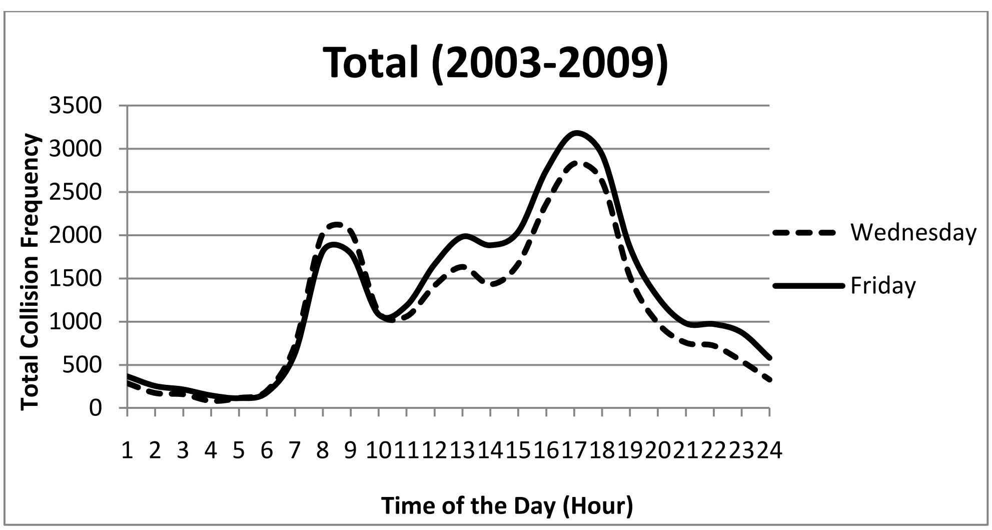


FIGURE 3.8: HOURLY TOTAL COLLISION FREQUENCIES (2003-2009)

The Figure 3.9 shows the general hourly variation of traffic flow. This figure shows three curves with minimum, maximum and average values. The general hourly traffic flow pattern is found to be very similar to the hourly variation of traffic collision in the City of Edmonton. It shows that number of traffic collisions is a function of traffic flow.

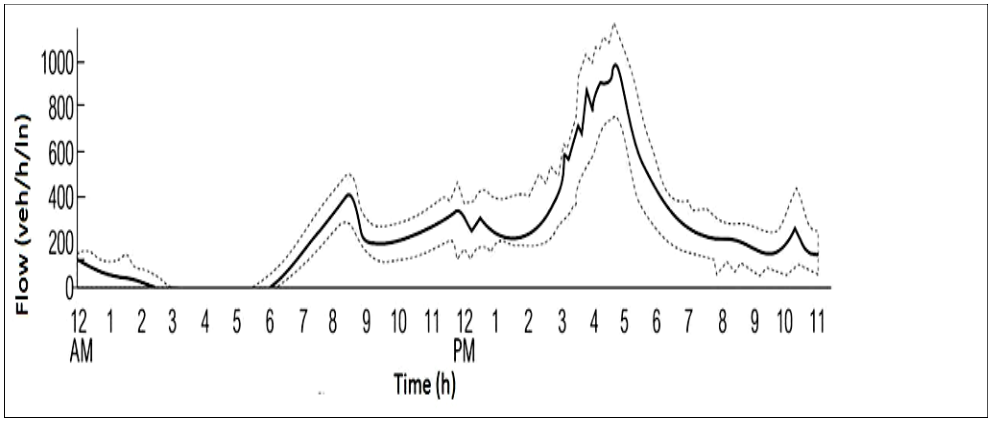


FIGURE 3.9: HOURLY TRAFFIC FLOW PROFILE

Source: Highway Capacity manual 2000. Copyright, national Academy of Sciences, Washington, D.C. Exhibit 8-7, p. 8-7.

3.5.5 Seasonal Collision Pattern

To study the seasonal variation of the collisions, we considered the following different ranges of the seasons from 2003-2009. Seasons of year winter, spring, summer, and fall (Table 3-3) were considered to study the seasonal variation of traffic collisions. The first day of the seasons is decided based on the position of the Sun. In Canada, the first day of the winter and summer are when the Sun is farthest south and farthest north respectively. Also, the first day of the spring and fall seasons are when the Sun crosses the celestial equator moving northward and southward respectively. Normally, December 21, March 20, June 20, September 22 are the first days of the winter, spring, and summer and fall seasons respectively.

Table 3.3: Adopted range of week and 1st day of the season

Season	1st day	Week Range
Winter	December, 21	1-11, 51-53
Spring	March, 20	12-24
Summer	June, 20	25-38
Fall	September, 22	39-50

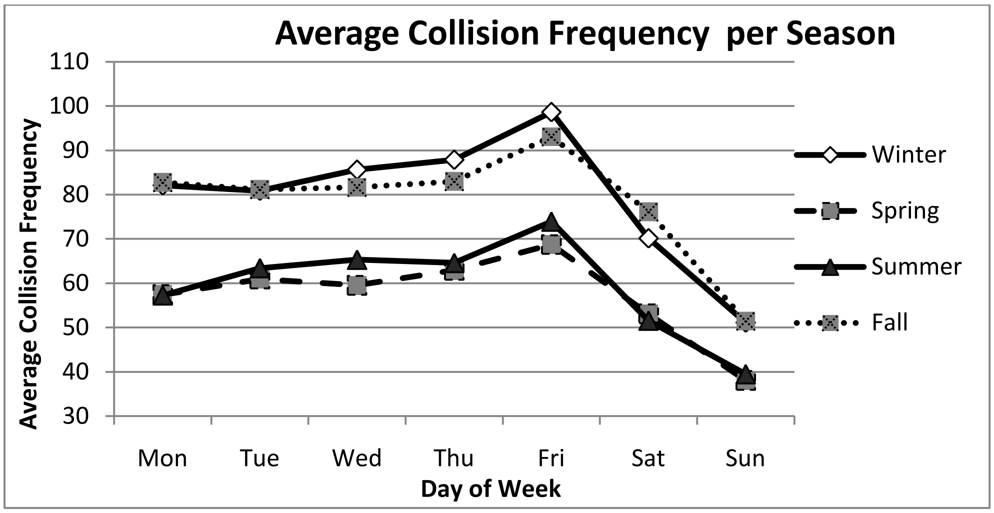


FIGURE 3.10: COLLISION FREQUENCIES PER SEASON BY DAY OF WEEK (2003-2009)

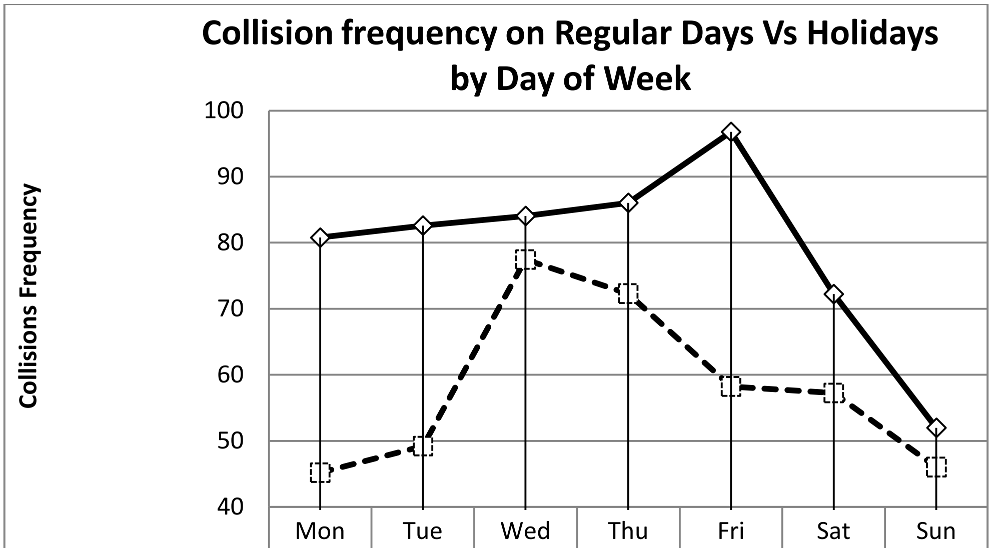
Hence these first days are considered in this research for the seasonal collision analysis. Week 1 to 11 and week 51 to 53 were considered in the winter season whereas, week 12 to 24 was assumed to reflect spring season. Summer was supposed to be from week 25 to 38 and fall season was assumed to be in the range of week 39 to 50. Average collision frequency per season from 2003-2009 (Figure 3.10) shows the similar trend as in Figure 3.1. The highest number of collisions on Friday and the least number of collisions on Sunday was noticed from seasonal data analysis as shown in Figure 3.10. Spring and summer seasons were found to have less number of collisions for all days of week as compared to the fall and winter.

3.5.6 Holiday Collision Pattern

It is very important for engineers, planners and managers to understand the local as well as statutory holidays and their dates for safe and efficient traffic management. It is necessary to identity which holiday occasions are observed in the province of Alberta in order to investigate the variation features of holiday traffic. The observing dates of the 12 national and provincial holidays observed in Alberta, according to the Canadian Heritage Department and Palm Sunday, Easter Sunday, Christmas Eve and New Year`s Eve were considered for the analysis. To capture the effect of the holidays on the collision trend, the **sixteen** most common Canadian Holidays were listed, from 2003 – 2009 as follows:-

1. New Year's Day, U.S & Canada
2. Family Day, Canada (Alberta)
3. Palm Sunday
4. Good Friday, Canada
5. Easter Sunday
6. Easter Monday, Canada
7. Victoria Day, Canada
8. Canada Day, Canada
9. Civic Holiday, Canada
10. Labor Day, Canada
11. Thanksgiving Day, Canada
12. Remembrance Day, Canada (Alberta)
13. Christmas Eve, U.S and Canada
14. Christmas Day, U.S and Canada
15. Boxing Day, Canada
16. New Year's Eve

All traffic collisions from 2003-2009 were normalized (Figure 3.11) considering the total number of accidents per year, multiplied by 29000 (total number of accidents on 2009). From the analysis of the holidays and regular days a different trend of traffic collisions was observed for holidays. It was found that holiday on Wednesday was more promising for the collision to occur as compared to other days of the week. Thursday was explored to be the next prominent after Wednesday.



	Mon	Tue	Wed	Thu	Fri	Sat	Sun
Regular Days	81	83	84	86	97	72	52
Holidays	45	49	77	72	58	57	46
Difference	-44.1%	-40.4%	-7.9%	-16.0%	-39.8%	-20.7%	-11.4%

FIGURE 3.11: NUMBER OF COLLISIONS FOR REGULAR DAYS VS. HOLIDAYS BY DAY OF WEEK

Monday is found to be the lowest for holiday. Therefore to capture the trend of the holidays, we incorporated the dummy variables in the prediction model. There is maximum difference of 44.1% for expected number of collisions on Monday between regular days and holidays.

The least difference is for Wednesday (7.9%) with Sunday as the second least (11.4%). The second and third highest difference was observed on Tuesday and Friday with 40.4% and 39.8 % difference respectively (Figure 3.11). This gave a clear picture as to how many accidents happen on holidays and long weekends. After that, all Sunday or Saturday holiday were switched to Monday (a crescent common practice in Alberta). On the switched Mondays, the average number of collision was 65. This observation also suffers from the same problem as discussed before, the statistical representativeness of the sample (Only 15 observations for 7 years of data). The average number of collisions on switched Mondays is between the extended holiday pattern and regular Mondays. The table was plotted between days of week and regular days, holidays, extended holidays and switched Mondays as shown.

TABLE 3.4: SAMPLE SIZE FOR EACH PATTERN ANALYSIS

Day of Week	Regular days	Holidays	Extended Holidays	Switched Mondays
Mon	365	47	5	15
Tue	365	7	0	0
Wed	367	9	1	0
Thu	366	8	1	0
Fri	365	14	4	0
Sat	365	7	57	0
Sun	365	21	51	0

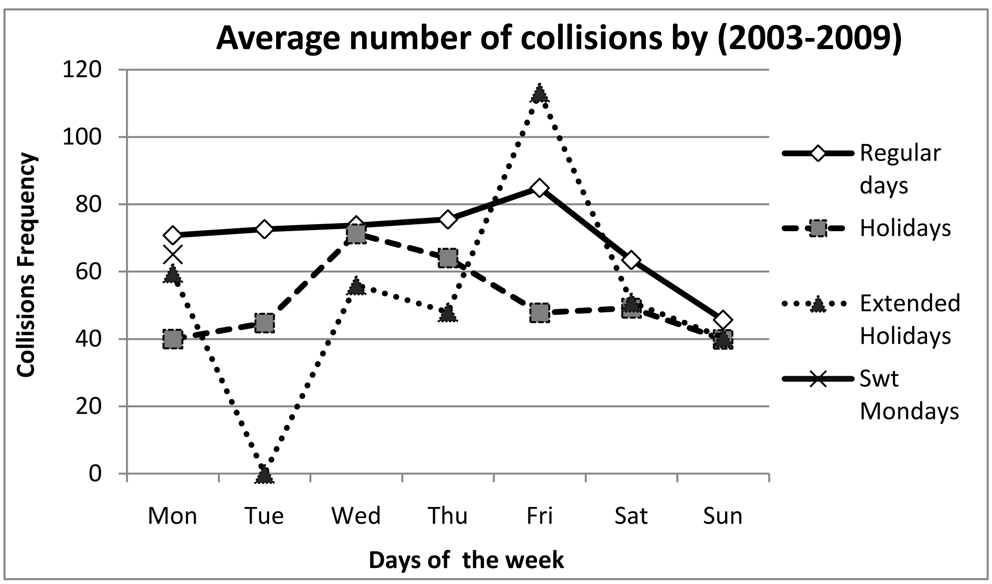


FIGURE 3.12: AVERAGE NUMBERS OF COLLISIONS BY DAY-OF-WEEK (2003-2009)

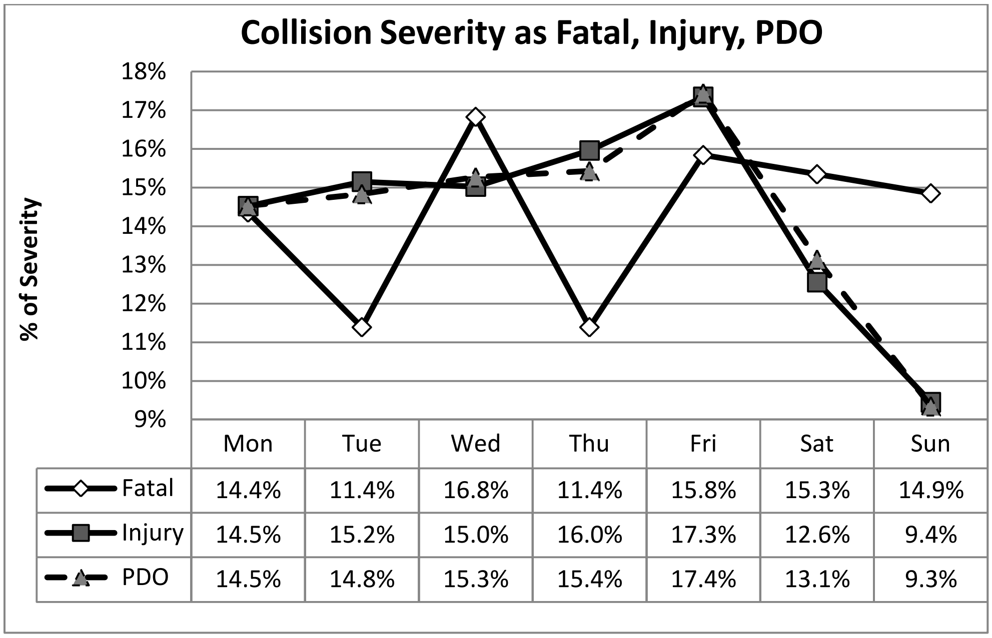
On Monday the highest number of holidays (47) was observed. For extended holidays Saturday was more prominent with value 57. From the analysis of the holidays, it was found that holiday on Monday is more promising for the collision to occur as compared to the holidays on the other days of the week.

When holidays fall on a Saturday or Sunday, the next working day is usually considered a legal holiday. Most of the time in the analysis of the data holidays were switched to the coming Mondays. These were called as Switched Holidays.

For regular days from Monday to Sunday the graph shows that maximum accidents occurs on Fridays (average = 85) and the least accidents occurs on Sundays (average = 46). There was only one data point for the switched Mondays in the graph. From Mondays to Wednesdays the expected number of collisions is almost same (71-74). On holidays the graph shows that from Mondays to Sundays the maximum accidents occurs on Wednesdays (71) and Thursdays (64). Sundays and Mondays have the least accidents (40). On extended holidays the maximum accidents occurs on Friday (113). Despite of that, little can be said because of the reduced number of observations (sample size) from Monday to Friday. Note that Saturday and Sunday on extended holidays show a similar pattern with Saturday and Sunday holidays.

3.5.7 Pattern by Collision Severity

The MVCIS data was filtered and analyzed on the basis of severity. This study consider fatal, injury and property damage only (PDO) as three measures of severity. Figure 3.13 was drawn to see the collision severity as fatal, injury and PDO. It was found from the analysis that percentage of injury and PDO collisions on Friday was higher compared to the other days as shown in Figure 3.13. Sunday was perceived as lowest severity in both injury and PDO. However, the trend of fatal collisions noticed was different from injury and PDO.



	Mon	Tue	Wed	Thu	Fri	Sat	Sun
Fatal	14.4%	11.4%	16.8%	11.4%	15.8%	15.3%	14.9%
Injury	14.5%	15.2%	15.0%	16.0%	17.3%	12.6%	9.4%
PDO	14.5%	14.8%	15.3%	15.4%	17.4%	13.1%	9.3%

FIGURE 3.13: COLLISION SEVERITY BY DAY OF WEEK (2003-2009)

The maximum percentage was found on Wednesday for fatal collisions with Friday as next highest. Tuesday and Thursday were exactly similar in percentage (11.4%) were lowest in comparison to other days. Saturday and Sunday were third and fourth highest in percentage with 15.3% and 14.9% respectively. Percentage of Monday (14.4%) was found lower than Sunday but higher than Tuesday.

3.5.8 Trend of Number of Hours of Snowfall

The national climate data maintained by the Environment Canada were analyzed to see the association between traffic collision frequency and the number of hours of snowfall. The state of weather column in the data was considered to estimate the number of hours of snowfall. If any state of weather (in the hourly frequency of the data) has snow in it, then it was considered as a snowy hour. Figure 3.14 shows the clear relationship between number of hours of snowfall and average collision frequency. An increasing trend was perceived from the chart as shown. The least collision frequency was observed when there was no snowfall on a particular day. In other words, with zero number of hours of snowfall a minimum collision frequency of 62 was observed. The maximum of average collision frequency with 144 was observed correspond to 12 hours of snowfall.

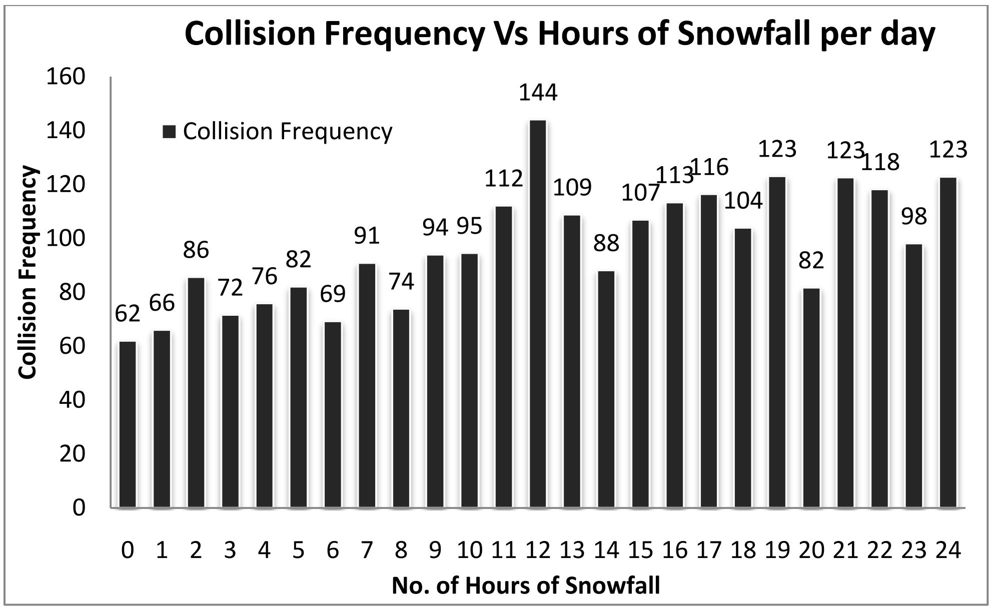


FIGURE 3.14: AVG. COLLISION FREQUENCY AND NO. OF HOURS OF SNOWFALL

The climate data were also used to calculate the number of rainfall hours on a particular day. If any state of weather (in the hourly frequency of the data) has rain in it, then it was considered as a rainy hour. There was no particular trend or pattern observed from this analysis. The data analysis did not show a good relationship between number of rainfall hours and the average collision frequency. The analysis of traffic collisions and snowfall shows a particular trend. However no such relationship was observed in the similar analysis with the rainfall. Therefore, the idea of considering the number of hours of rainfall as an independent variable in the collision prediction model was dropped. The parameter values of the variables will be different if rainfall is considered as an independent variable in the collision prediction model.

3.5.9 Trend of Hours of Daylight

For each day, the hours of daylight gives information of daylight from sunrise to sunset on that particular day. The appearance and disappearance of the upper limb of the sun as observed at sea level on a refracted (apparent) sea horizon represents the rising and setting times (respectively) of the sun in this data. The data analysis shows that the collision frequency follows the decreasing trend with increasing in number of daylight hours (Figure 3.15). As we discussed preciously, maximum 19 hours of daylight is observed in the months of summer and minimum 9 hours of daylight in winters was observed from the data. The maximum average collision frequency of 92 was observed with 9 hours of the daylight. The least collision frequency was found correspond to 17 hours of the daylight. The 18 and 19 hours of daylight show a small increase in the average collision frequency with 58 and 69 values respectively. Similar average collision frequency was found correspond to 15 and 16 hours of daylight. Also, the average collision frequency of 73 matches both 11 hours of daylight and 13 hours of daylight.

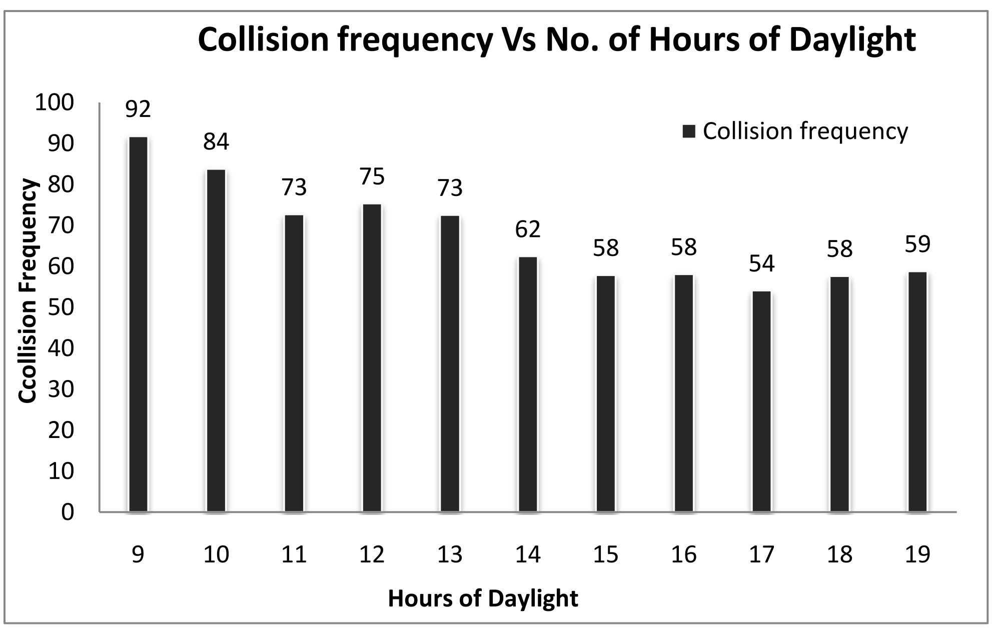


FIGURE 3.15: COLLISION FREQUENCY AND NO. OF HOURS OF DAYLIGHT

3.5.10 Surface Condition analysis

Figure 3.16 shows the daily variation of the total collision frequency for the different categories of the surface condition. The Figure shows five different surface conditions defined by the City of Edmonton MVCIS database. The five categories of the surface condition are 1) Dry, 2) Wet, 3) Loose sand/Dirt/Gravel, 4) Snowy/Icy and 5) Other. It is noticed from the figure that larger number of traffic collisions were observed on the Dry pavement condition as compared to the other four conditions of the pavement.

The Snowy/Icy condition is perceived to be the second most involved in traffic collisions after Dry condition. There were very few number of traffic collisions observed on Loose sand/Dirt/Gravel surface condition and Other Surface condition. The daily variation remains same for the dry condition showing Friday spikes. Consistently lower traffic collisions were observed on Saturday and Sunday.

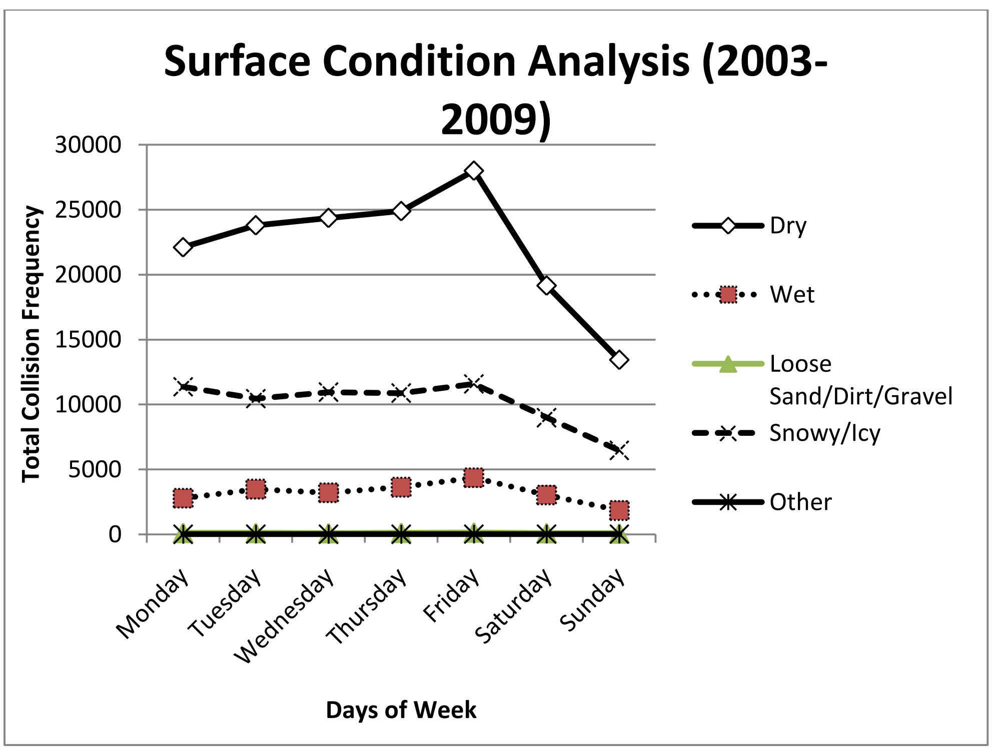


FIGURE 3.16: COLLISION FREQUENCY AT DIFFERENT SURFACE CONDITIONS

3.5.11 Age Analysis

The *Age* of Driver or Passenger for each collision record was categorized into 6 categories as below:

1 ≤ 20;
2 > 20 and ≤ 30;
3 > 30 and ≤ 40;
4 > 40 and ≤ 50;
5 > 50 and ≤ 60 and;
6 > 60

Figure 3.17 shows the daily variation of traffic collisions form Monday through Sunday of the age groups defined above. The age group 21-30, was found to be much involved in traffic collisions as compared to other age groups. The similar Friday spikes can also be observed from this chart. The collision frequencies observed in the age group 31-40 and 41-50 were found to be very similar. The age group considering the ages less than equal to 20 remains at fourth position after the age group of 41-50. The age group greater than 60 is observed to be least involved in traffic collisions for the entire analysis period (2003-2009).

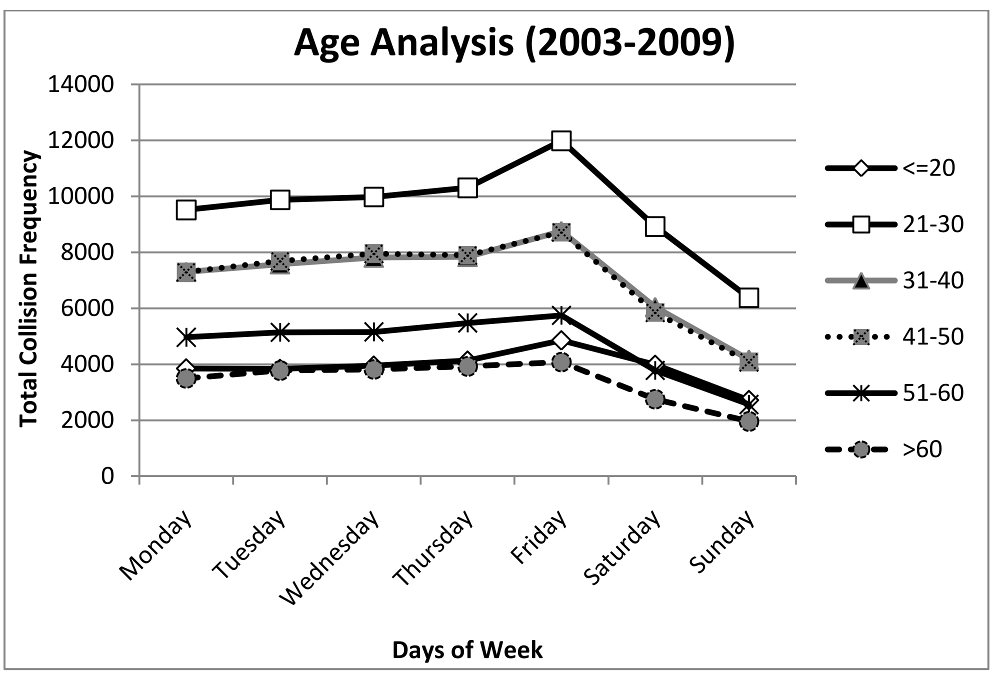


FIGURE 3.17: DAILY COLLISION FREQUENCY IN EACH AGE GROUP

3.5.12 Gender Analysis

The *Gender* of Driver or Passenger for each collision record was categorized as below:

1. Male and;
2. Female

Both male and female have higher probability to involving in collision on Friday. Males were perceived as more prone to traffic collisions as compared to the females for all the days including weekends. Lower collision frequencies were observed on both Saturday and Sunday as shown previously in the data analysis.

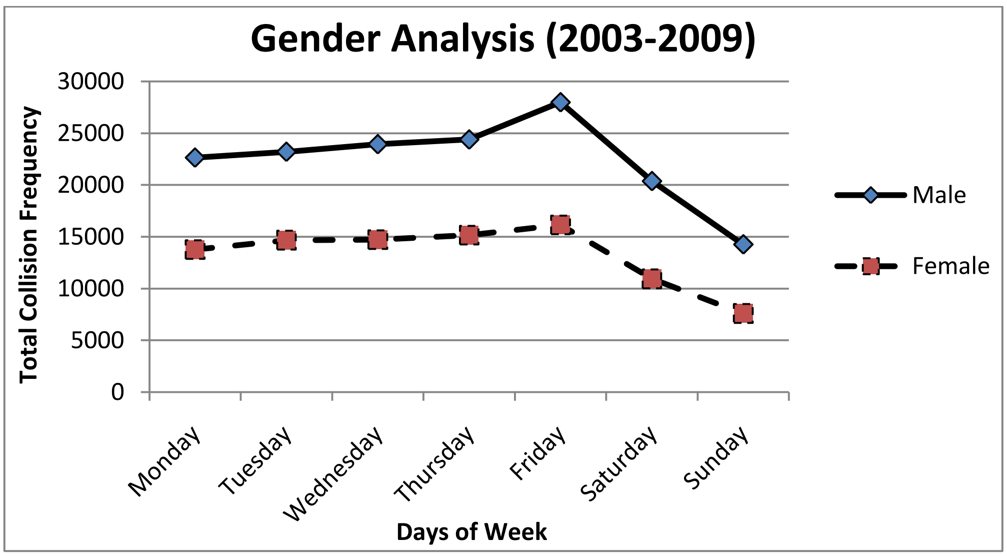


FIGURE 3.18: DAILY COLLISION FREQUENCY BY GENDER

3.6 Summary

The Data analysis shows that the collision frequency follows the decreasing trend with increasing in number of daylight hours. The collision data were derived from the police reports for collisions involving fatality, injury and property damage only equal to or greater than $1000. The study considers three major collision contributing factors: 1) Number of daylight hours, 2) Number of snowfall hours, and; 3) Visibility on daily basis. The dummy variables were incorporated in the model to capture the exposure of the day of week and the holidays. In this study three different data sources were compiled and used: 1) MVCIS (Motor Vehicle Collision Information System); 2) Sunrise/Sunset data; and 3) National Climate Data and Information Archive. The dry and snowy/icy surface condition, the age group 21-30 and the male gender group were observed to be more involved in traffic collisions.

CHAPTER 4. COLLISION PREDICTION MODEL FORMULATION

This chapter outlines the detailed procedure adopted to formulate the collision prediction model using the previously discussed data source. It includes the different independent variables considered for the model formulation. Model calibration is also discussed in detail in this chapter.

4.1 Introduction

For this study, after literature review, different collision contributing factors were finalized and ranked. All the factors discussed later in this chapter, contribute to traffic collisions. The variance was found significantly greater that the mean in the data analysis. Due to the over dispersion of the collision data a Negative Binomial (NB) model was formulated to establish a relation between the collision frequency and the independent variables. The generalized linear model (GLM) procedure in SPSS was adopted to estimate the parameters of the NB model. The model accuracy test results were very good.

The number of daylight hours, snowfall hours and visibility were found to have significant impact on traffic collisions. The statistical measures show that increased number of traffic collision is expected with the increase of number of hour of snowfall. The proposed negative binomial model was found good and this model can be applied in the cities having long winter conditions. It could provide a realistic estimate of expected collision frequency for a particular day as a function of number of hours of daylight, number of hours of snowfall and visibility.

4.2 First Collision Prediction Model

4.2.1 Model Variables

The contributing factors were categorized in three major groups: I. Road Environment related factors; II. Driver related factors and; III. Vehicle related factors. Further the road environment related factors were classified into four categories: a) Traffic Flow; b) Traffic Control; c) Infrastructure and; 4) Weather Condition. These categories were still needed to be narrowed down. So, we divided these categories further in different sub-categories. Driver related factors and vehicle related factors were classified into four and three sub-categories respectively. The structure of the collision contributing factors after categorization is shown in Figure 4.1.

All the factors listed in the Figure 4.1 contribute to traffic collisions. Average Annual Weekday Traffic flow (AAWT) was one of the main candidates to be selected as one of the major collision contributing factor. However, AAWT data were not available. Therefore, dummy variables were incorporated to capture the exposure of the daily traffic volumes. The information in the

MVCIS data was limited to answer the daily patterns of the collisions in the Edmonton. As a result, two new sources were used as discussed preciously. From these data sources, the number of daylight hours, the number of snowfall hours and visibility were used. The data analyses of these sources were described in the previous chapter.

The study objective was to evaluate the factors which have daily influence on traffic collision. Some of factors in MVCIS were location specific and some factors do not vary daily.

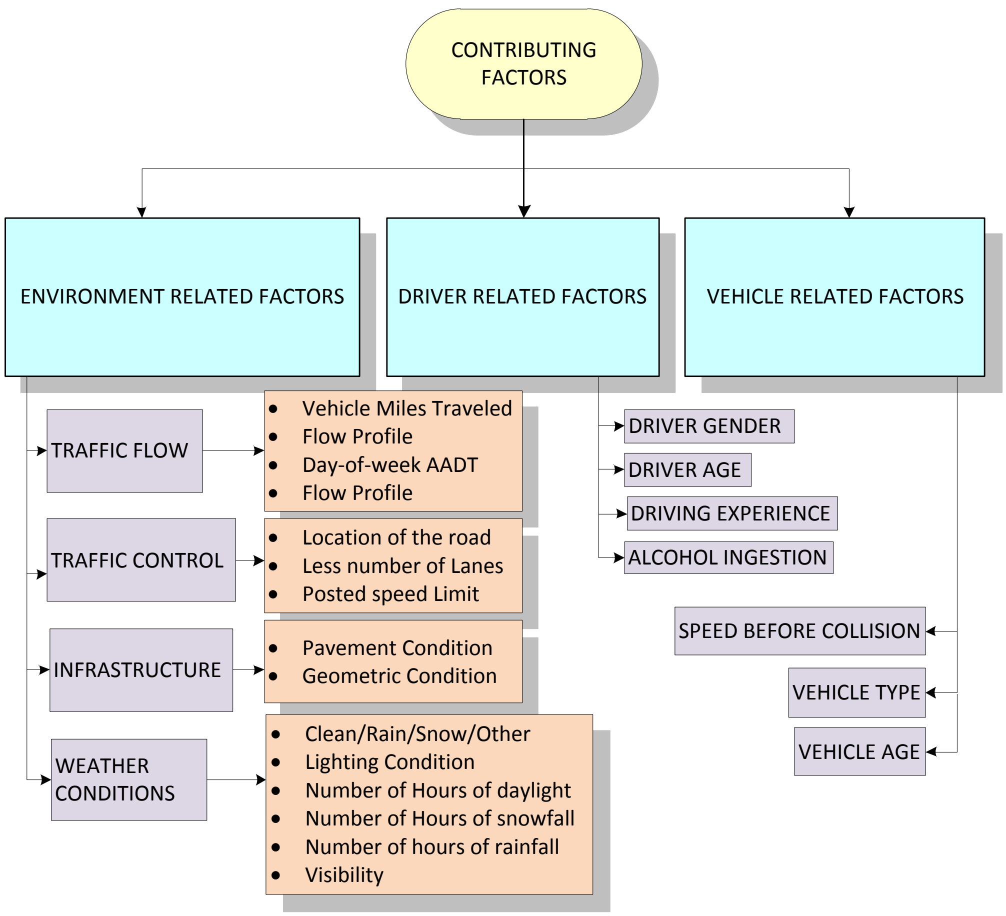


FIGURE 4.1: COLLISION CONTRIBUTING FACTORS

We could not find any specific pattern of rainfall that may be used to interpret the daily variation of traffic collisions. Finally, in the model we have considered three continuous variables (number of daylight hours, number of snowfall hours and visibility in km) and eight dummy variables to capture the daily exposure.

4.2.2 Proposed First Model

Collision prediction models usually assume that collision frequency is distributed as a Poisson or Negative Binomial (NB). The equality of the mean and variance of the Poisson distribution limits the ability of the Poisson model in modeling wider range of count data. However, the mean and variance of the NB distribution are two independent parameters. This allows the NB model to have greater flexibility than the Poisson model to fit the frequency patterns of the observed count data. For this study, we used the NB distribution to model count response data due to over-dispersion (i.e., the presence of greater variability in the data than would be expected under the Poisson model).

For modeling collision frequencies we adopted the regression analysis for count data. NB model was used in capturing observed and unobserved collision variations for each day along the year. This model can be written as, $Y_i \sim NB(\mu_i, \alpha)$, where Y_i represents the number of collision each day i (i=1,…365/366), μ_i stands for the mean collision frequency, and α is the over-dispersion parameter. One of the shortcomings of the NB model is the assumption of a constant over-dispersion parameter (α) for all observations. It is assumed that the mean collision frequency (μ_i) is a function of a set of covariates through the log link function. The functional form of the adopted regression model is:

$$\eta_i = \ln(\mu_i) = \beta_0 + \sum_{j=1}^{n} \beta_j x_{ij} \tag{4.1}$$

Where, μ_i = expected number of collision for each day i;

η_i = logarithm of μ_i;

β_0 = intercept;

β_j = coefficient for the jth explanatory variable (j = 1 . . . n);

x_{ij} = value of the jth explanatory variable for day i;

n = number of explanatory variables; (n=11)

Total eleven explanatory variables were considered in the model. Eight are dichotomous/dummy variables those capture exposure of each day including holiday. For the dummy variables from Monday to Sunday and holiday, the corresponding coefficients are β_1, β_2, β_3, β_4, β_5, β_6, β_7, β_8. The remaining three (Hour of daylight, Hour of Snowfall, Visibility) are continuous variables and their corresponding coefficients are β_9, β_{10}, β_{11}.

Two statistical measures for goodness of fit; 1) scaled deviance (SD) and 2) the Pearson $\chi 2$ statistic were commonly used in the past [39]. The scaled deviance is defined as the likelihood ratio test statistic which measures twice the difference between the log-likelihoods of the studied model and the full or saturated model. The full model has as many parameters as there are observations (each collision having its own collision parameter) so that the model fits the data perfectly. Therefore, the full model, which possesses the maximum log-likelihood achievable under the given data, provides a baseline for assessing the goodness of fit of an intermediate model with parameters [40].

4.2.3 First Model Calibration

The statistical software SPSS Version 15 was used to estimate the model parameters. 80% of the count data for 365/366 days observation was used to calibrate the model parameters for each year. For a specific day, the value of the corresponding dummy variable is considered 1 in the equation (1) and the remaining dummy variables values are 0. For determining the coefficients of dummy variables, Wednesday was considered as a base dummy variable. The estimated parameter values of the dummy variables will show the strength of collision frequency on a particular day compared to the base variable. Saturday and Sunday were not considered as a base as these are representing weekend. Traffic pattern is different on Weekend (Saturday and Sunday) from other days. Monday comes exactly after holidays of Saturday and Sunday. Both traffic pattern and collision pattern are different compared to other days so the idea of considering Monday as base dummy variable was dropped. Friday was not considered for base dummy variable as it is under observation. On Friday, the trends of parameters values were to be observed. So it was not taken as the base dummy variable. Tuesday, Wednesday, and Thursday are almost similar days of the week and any of these can be considered as base variable. The numbers of collisions observed on Wednesday were less than Thursday but higher than Tuesday. Therefore, in this study Wednesday was considered as the base dummy variable. If the base variable was some other day, let's say Thursday then the estimated parameter values would be different and the coefficients of the dummy variables would have shown the collision frequency compared to the base variable i.e. Thursday.

The parameters of the NB model were estimated by using the generalized linear model (GLM) procedure in SPSS. Tables 4.1 summarize the model parameter estimates and their associated statistics under the NB distribution. An examination of the tables indicates that the model parameter estimates for most of the variables are significant at the 95% confidence level.

Hour of daylight was found to have a statistically significant effect on collision frequency. The negative model coefficient also makes intuitive sense, as it suggests that reduced hour of daylight was associated with increased number of accidents. Hour of snowfall was found to have a statistically significant effect on collision frequency for all the model year. The coefficients for the hour of snowfall are positive, which indicates an increase in the mean accident frequency with snowfall and this is very intuitive.

The coefficient for visibility was found statistically significant for only the year 2004. The negative model coefficient suggests that reduced visibility was associated with increased number of collisions. Counter-intuitive sign was obtained for visibility distance for the year 2008. One possible reason might be due to extensive new road construction and rehabilitation work that was started in the city in year 2008. At many places vehicle movement was restricted due to roadway lane reduction and closures. Due to lack of familiarity to the altered road, it might cause higher collision probability even for the same visibility.

The positive coefficient for specific dummy variable shows relatively higher collision probability on that day compared to Wednesday in that year. The collision data analysis shows higher average collision number for Friday compared to any other days. The model estimated parameter compliance with that finding can be observed from the Table 4.1. For all the years, the obtained value of coefficient for Friday is positive and the strength of coefficient is significantly higher than all other day having positive coefficient. For some of the days of the week, the estimated model parameters for dummy variables are not statistically significant. It represents that day is not significantly different from Wednesday in term of collision frequency. Also, the negative coefficient values for some dummy variables were obtained and it shows that during those days collision probability is less than Wednesday.

The negative coefficient values of holidays from 2003 to 2009 show that it has less collision probability compared to the regular days. This finding is in compliance with the data analysis discussed previously.

TABLE 4.1: ESTIMATES OF MODEL PARAMETERS USING NB REGRESSION

Model Year	Parameter	β_1 (Mon)	β_2 (Tue)	β_3 (Wed)	β_4 (Thu)	β_5 (Fri)	β_6 (Sat)	β_7 (Sun)	β_0 (Constant)	β_8 (Hol)	β_9 (Hour. of Daylight)	β_{10} (Hour. of Snowfall)	β_{11} (Visibility)	t-value (95% Confidence Level)
2003	Coefficients	-0.091	-0.028	0.000	-0.046	0.115	-0.255	-0.549	4.907	-0.487	-0.039	0.021	-0.011	1.96
	t-value	1.398	0.444	-	0.734	1.840	4.008	8.480	26.924	5.493	7.367	4.339	1.595	
2004	Coefficients	0.013	0.001	0.000	0.036	0.170	-0.129	-0.382	5.188	-0.488	-0.050	0.016	-0.023	1.96
	t-value	0.178	0.001	-	0.526	2.457	1.864	5.273	28.255	5.310	8.472	3.266	3.598	
2005	Coefficients	0.096	0.036	0.000	0.046	0.214	-0.127	-0.442	4.876	-0.518	-0.038	0.013	-0.006	1.96
	t-value	1.705	0.654	-	0.825	3.880	2.287	7.732	35.567	6.240	8.014	3.188	1.239	
2006	Coefficients	-0.103	-0.039	0.000	-0.016	0.098	-0.222	-0.514	4.772	-0.384	-0.025	0.020	-0.005	1.96
	t-value	1.645	0.651	-	0.266	1.625	3.505	8.099	36.656	4.466	4.814	4.846	1.182	
2007	Coefficients	0.007	-0.034	0.000	0.003	0.128	-0.155	-0.480	5.045	-0.369	-0.038	0.020	-0.007	1.96
	t-value	0.118	0.564	-	0.044	2.139	2.589	7.888	28.395	4.079	7.785	4.376	1.071	
2008	Coefficients	-0.003	0.017	0.000	-0.004	0.182	-0.077	-0.392	4.610	-0.460	-0.038	0.040	0.010	1.96
	t-value	0.054	0.292	-	0.063	3.093	1.300	6.526	29.668	5.733	7.649	8.441	1.700	
2009	Coefficients	-0.083	-0.027	0.000	0.065	0.146	-0.143	-0.434	4.964	-0.480	-0.042	0.023	-0.001	1.96
	t-value	1.288	0.412	-	1.002	2.223	2.169	6.486	28.527	5.699	7.472	4.965	0.176	

4.3 Second Collision Prediction Model

4.3.1 Multinomial Logistic Regression

Multinomial Logistic Regression model is the second model built in the family of collision prediction models. This model is built for the collision probability of gender and different age groups for either driver or passenger involved in the collision for the City of Edmonton. As discussed previously, the negative binomial model was formulated to predict the collision frequency based on the number of hours of daylight, number of hours of snowfall and visibility. The stochastic nature of the collision prediction model was considered in the second collision prediction model. This model predicts the percentage of the different age groups and gender of both drivers and passengers involved in the collisions considering that the collision has happened.

In the previous studies the stochastic nature of the collision prediction models was built based on the type of collisions like front to end collision, end to end collision, right to end collision etc. In these studies they predict the probability of the different types of collision likely to happen. With the collision data if we have the cases like in some particular conditions the collision doesn't happen then we could build a model which defines the probability of the collision to happen or not. However we have the collision data which only shows that collision has happened on a particular day, at a particular time and at a particular location.

As we have only the cases of collisions happened, therefore, the main condition for this type of model considered is that collision has happened. We were interested to see the probability of the involvement of the different age groups and gender of both driver and passengers involved in the collision considering that the collision has happened. Multinomial logistic regression is used to analyze the relationships between a non-metric dependent variable and metric or dichotomous independent variables. The non-metric dependent variable is a kind of dichotomous variable which only includes dummy coded variables as a value like 1,2,3,4 and so on. However for non-metric case it will not capture any range of values. For example driver gender is a non-metric variable because it can be coded as 0 and 1 or 1 and 2. For these dummies 0's and 1's the non-metric variable will not take any range of value.

However, age can be a metric variable as it can be categorized in different age groups of different ranges like 20 to 30 or 30 40 and so on. For this model non-metric dependent variables were available as probability of collision to happen each day. The dependent variables were categorized for each day from 1 to 7 from Sunday to Saturday. There may be any number of values for these seven categories of the dependent variables with a minimum of 3. If there are only two categories for the dependent variables it is better to go for binomial regression which is more efficient than the multinomial logistic regression. Therefore, the multinomial logistic regression model was used due to the seven categories of the dependent variables.

Multinomial logistic regression compares multiple groups through a combination of binary logistic regressions. The group comparisons are equivalent to the comparisons for a dummy-coded dependent variable, with the group which has the highest numeric score used as the reference group. For categorical values it is not appropriate to use multinomial logistic regression model. For linear regression the response/dependent variable must be in the metric scale. It should be in a range and ratio scale. Ratio scale is a scale of measurement which tells us whether the value of the variable is better or worse than the previous one. Linear regression model is good for ratio scale. The dependent variables we have are collision report number, the day of week etc. The ratio of collision report number and the day of week mean nothing for calculating the ratio scale. If the dependent variables are polytomous i.e. the categorical observations is more than 2 and filter variable is also categorical then we can build a model using log-linear model.

Multinomial logistic regression provides a set of coefficients for each of the two comparisons. The coefficients for the reference group are all zeros, similar to the coefficients for the reference group for a dummy-coded variable. Thus, there are three equations, one for each group defined by the dependent variable. The three equations can be used to compute the probability that a subject is a member of each of the three groups. A case is predicted to belong to the group associated with the highest probability. The predicted group membership can be compared to actual group membership to obtain a measure of classification accuracy.

This study didn't consider the log-linear model as it has two disadvantages. Firstly, it has many more parameters and many of them are not filtered. It generates so many parameters for example; if we have 100 parameters and 100 observations, log linear model will capture all of them. However we don't need all these parameters. Secondly, the log-linear model is more complicated to interpret because the effect of the critical "X" variable on the response of the critical "Y" variable is discussed by the XY association. But we were more interested to see the probability of the individual age and gender category on a collision. We were not interested to see the association of the X and Y variables. So we have used the multinomial logistic regression model.

4.3.2 Level of Measurement

Multinomial logistic regression analysis requires that the dependent variable be non-metric. Dichotomous, nominal, and ordinal variables satisfy the level of measurement requirement. It is required for the logistic regression analysis that the dependent variable must be non-metric. It will not take any range of values only dichotomous values like 1, 2, 3, and so on. The independent variable must be dichotomous or metric. Multinomial logistic regression analysis requires that the independent variables be metric or dichotomous. Since SPSS will automatically dummy-code nominal level variables, they can be included since they will be dichotomized in the analysis. The statistical software SPSS was used to formulate this model. In SPSS, non-metric independent variables are included as “factors”. SPSS will dummy-code non-metric independent variables. In SPSS, metric independent variables are included as “covariates”. Covariate captures the range of values like for age group ≤ 20 was coded as 1. If an independent variable is ordinal, we will attach the usual caution.

4.3.3 Assumptions, Outliers and Requirements

The multinomial logistic regression model doesn’t assume normality, linearity and homogeneity of the variance. It is preferred to discriminate analysis when the data doesn’t satisfy these assumptions as it doesn’t impose the above mentioned requirements. So this is the flexibility of the model. SPSS doesn’t compute any diagnostic statistics for outliers. Outliers are the dependent variables that are distant from the rest of the data in the given range. To evaluate outliers, the advice is to run multiple binary logistic regressions and use those results to test the exclusion of outliers or influential cases. However, it is a very tedious job and time consuming. We can identify the outliers in a more simple way by computing Mahalanobis distance in the SPSS regression procedure discussed later in this chapter.

We have mainly two independent variables age and gender of both driver and passenger involved in collision. The minimum 10 number of cases per independent variable should be available, using a guideline provided by Hosmer and Lemeshow [41], one of the main resources for Logistic Regression. This is the sample size requirement.

Step 1: Define the Research Problem

In this stage, the following issues are addressed:

1. Relationship to be analyzed:

   The goal of this analysis is to examine the probability of different age category and gender in the involvement in a crash for different days of week. Again the assumption is that the collision has happened. This is basically logical type of model.

2. Specifying the dependent and independent variables:

a. The dependent variable:

The dependent variable is Day of week for each collision record and it was considered in 7 categories: Monday = 1, Tuesday = 2, Wednesday = 3, Thursday = 4, Friday = 5, Saturday = 6, Sunday = 7. These are non-metric dichotomous variables. The assumption is that different age and gender categories have different probabilities for involving in a collision for different days of the week.

SPSS uses the highest numbered choice by default (in this case is 7 representing Sunday), as the reference category. Because these are the dummy variables and in order to find the parameters of the model we need to consider a category as a reference. For the first collision prediction model discussed previously was built considering Wednesday as the reference category. So we have again considered Wednesday as a reference category in the second collision prediction model.

b. The independent variables:

There are two independent variables considered in this study. The first independent variable is *Age* of Driver or Passenger for each collision record and it was considered in 6 age categories as below:

1 ≤ 20;

2 > 20 and ≤ 30;

3 > 30 and ≤ 40;

4 > 40 and ≤ 50;
5 > 50 and ≤ 60 and;
6 > 60

The second independent variable is *Gender* of Driver + Passenger for each collision record and it was categorized as below:

1 1 - Male and;
2 0 - Female

These are the dichotomous variables. We can model the age both as a covariate and as a factor because it can take a range. For the factor there is another consideration that is independent variable has a direct influence on the dependent variable then it is called as covariate. However if gender change from male to female or vice-versa, we cannot say directly that probability of collision happen or not. Both variables were considered as a factor because one of them doesn't have direct influence on the independent variable.

3. Method for including independent variables:

   The only method for including variables multinomial logistic regression in SPSS is direct entry of all variables.

Step 2: Develop the Analysis Plan: Sample Size Issues

As discussed previously that at least minimum sample size of 10 cases per independent variable should be available. For the Missing data analysis if some information is missing in the data then it will generate error. There was information missing in the collision records. In this research a collision was eliminated from the data if that collision has some missing information. This research SPSS doesn't show anything about the missing data. So we bypassed any missing data analysis. This study doesn't consider any analysis for the missing data like what percentage of the missing data that was available in the MVCIS data. The comprehensive MVCIS data provided by the City of Edmonton was formatted as per requirement of the SPSS using various formulas and filters.

Step 3: Model Estimation: Logistic Regression and Assessing Overall Fit

Once we have the data in the appropriate format we decided to go for logistic regression model and choose the model form. Linear regression is not appropriate for situation in which there is no natural ordering to the values of the dependent variable. We have seven categories of the dependent variables as discussed before and we cannot order them naturally from 1 to 7. If dependent variable is high, low or medium then we can categorize it as 1, 2 and 3. However it is difficult to ordering the categorical variable. Due to this reason the multinomial logistic regression is the best alternative in this case.

For a dependent variable with K categories (we have 7 categories), consider the existence of K unobserved (there are some variables one cannot observe, but it has there) continuous variables, Z_1, ... Z_K, each of which can be thought of as the "propensity toward" a category. For each of seven categories we will have unobserved variables. If this value so called propensity toward collision (in statistical term) is higher, then the probability of that category in the collision is higher. In the case of a collision involvement in the different days of week, Z_k represents a driver's propensity toward involving in a crash for in the k^{th} day, with larger values of Z_k corresponding to greater probabilities of involving in a crash (assuming all other Z's remain the same).

4.3.4 Proposed Second Model

Mathematically, the relationship between the Z's and the probability of a particular outcome is described in this formula:

$$\pi_{ik} = \frac{e^{Z_{ik}}}{e^{Z_{i1}} + e^{Z_{i2}} + \dots + e^{Z_{iK}}} \tag{4.2}$$

Where, π_{ik} = Probability of the ith case falls in category k and;
Z_{ik} = Value of the kth unobserved continuous variable for the ith case

Mathematically, the relationship between the Z's and the probability of a particular outcome is described in this formula. In the model we have unobserved variable for each category. For these unobserved variables we had to make the observation based on the two predicted variables (age and gender). In order to capture the unobserved variables we have followed the equation below:

$$Z_{ik} = b_{k0} + b_{k1} X_{i1} + b_{k2} X_{i2} + \dots + b_{kJ} X_{iJ} \tag{4.3}$$

Where, X_{ij} = the j^{th} predictor of the ith case;
b_{Kj} = the j^{th} coefficient for the kth unobserved variable and;
J = the number of predictors;

Z_{iK} was also assumed to be linearly related to the parameters. Based on age and gender we can capture the unobserved variables from the above equation. Now once we have this equation, put this equation in the previous Equation 4.3. Here we don`t have any unobserved variables. Now the Equation 4.2 looks like as below:

$$\pi_{iK} = \frac{e^{b_{k0}+ b_{k1} X_{i1}+ \dots + b_{kJ} X_{iJ}}}{e^{b_{10}+ b_{11} X_{i1}+ \dots + b_{1J} X_{iJ}} + \dots + e^{b_{k0}+ b_{k1} X_{i1}+ \dots + b_{K0} X_{iJ}}} \tag{4.4}$$

It is notable that if Z_k were observable, one would simply fit a linear regression to each Z_k and be done. However, since Z_k is unobserved, one must relate the predictors to the probability of interest by substituting for Z_k.

Equation 4.4 is not a close form equation. So the challenge is how to identify the parameters of the equation. In order to find the parameters of the equation a constant term was incorporated over the equation and it will not change the ratio. As it stands, if you add a constant to each Z, then the outcome probability is unchanged. This is the problem of *non-identifiability*. To solve this problem, Z_K is (arbitrarily) set to 0. The K^{th} category is called the reference category, because all parameters in the model are interpreted in reference to it. It's a good idea (for convenience sake) to choose the reference category so that it is the "standard" category to which others would naturally be compared.

$$\pi_{ik}(\textit{with constants added to } z's) = \frac{e^{Z_{ik}+C}}{e^{Z_{i1}+C} + e^{Z_{i2}+C} + \ldots + e^{Z_{iK}+C}} \quad (4.5)$$

$$= \frac{e^{Z_{ik}}\, e^{C}}{e^{Z_{i1}}\, e^{C} + e^{Z_{i2}}\, e^{C} + \ldots + e^{Z_{iK}}\, e^{C}}$$

$$= \frac{e^{Z_{ik}}}{e^{Z_{i1}} + e^{Z_{i2}} + \ldots + e^{Z_{iK}}}$$

$$= \pi_{ik}$$

For Wednesday we have put $Z_K = 0$ as a reference category which solved our problem. Now, the parameters estimated will be compared to this reference category. The coefficients are estimated through an iterative Maximum Likelihood method.

4.4 Summary

The statistical software SPSS was used to estimate the model parameters. 80% of the count data for 365/366 days observation were used to calibrate the model parameters for each year. Due to over dispersion of the data, a Negative Binomial (NB) model was formulated to establish a relation between the collision frequency and the independent variables. The generalized linear model procedure in SPSS was adopted to estimate the parameters of the NB model.

Multinomial Logistic Regression model is the second model built in the family of collision prediction models. This model predicts the percentage of the different age groups and gender of both drivers and passengers involved in the collisions considering that the collision has happened**.** For Wednesday we have put $Z_K = 0$ as a reference category which solved our problem. In SPSS, metric independent variables are included as “covariates”. Covariate captures the range of values like for age group ≤ 20 coded as 1. If an independent variable is ordinal, we will attach the usual caution. The minimum ten cases per independent variable should be available. The dependent variable is Day of week for each collision record and it was considered in 7 categories: Monday = 1, Tuesday = 2, Wednesday = 3, Thursday = 4, Friday = 5, Saturday = 6, Sunday = 7. There are two independent variables considered in this study. The first independent variable is *Age* of Driver or Passenger for each collision record and it was considered in 6 age categories as below ≤ 20; > 20 and ≤ 30; > 30 and ≤ 40; > 40 and ≤ 50; > 50 and ≤ 60 and; > 60. The second independent variable is *Gender* of Driver or Passenger for each collision record and it was categorized as Male and Female.

CHAPTER 5. MODEL VALIDATION

This validation procedure of the collision prediction model formulated in the chapter 4 is discussed in this chapter. Two statistical measures were used to see the relative error between the actual and the predicted collision frequency.

5.1 Introduction

The variance was found significantly greater than the mean in traffic collision data analysis. Due to the over dispersion of the data a Negative Binomial (NB) model was formulated to establish a relation between the collision frequency and the independent variables. The generalized linear model (GLM) procedure in SPSS was adopted to estimate the parameters of the NB model. The model accuracy test results were very good. Two statistical measures namely Standard Error of Estimate and Root Mean Square Error (RMSE) values were calculated to see the relative error between predicted and actual count of collisions for each year. Scaled Deviance and Pearson Chi-Square are also outlined as a measure of goodness of fit. From the plots of observed and predicted collision data, it can be said that there is a very good agreement among those data.

5.2 First Model Validation

After calibrating the adopted negative binomial regression model, the remaining 20% of 365 days observations for each year were used in the model validation. Two statistical measures namely Standard Error of Estimate and Root Mean Square Error (RMSE) values were calculated to see the relative error between predicted and actual count of collisions for each year. Table 5.1 shows the measured statistical results. It is desirable that the value of RMSE should be close to 30% [42]. RMSE and Standard Error of Estimate were calculated using the following equations:

$$\mathrm{RMSE} = \frac{\sqrt{\frac{\sum_{j=1}^{n}(\mathrm{Model}_j - \mathrm{Count}_J)^2}{(\text{Number of Observation Point} - 1)}}}{\left(\frac{\sum_{j=1}^{n}\mathrm{Count}_j}{\text{Number of Observation Point}}\right)} \mathrm{X}100 \tag{5.1}$$

$$\text{Standard Error of Estimate} = \sqrt{\frac{\sum_{j=1}^{n}(\mathrm{Model}_j - \mathrm{Count}_J)^2}{\text{Number of Observation Point} - 1}} \tag{5.2}$$

TABLE 5.1: RMSE AND STANDARD ERROR OF ESTIMATE

Model	RMSE (%)	Standard Error of Estimate
2003	30.94	21.223
2004	21.29	11.132
2005	31.99	19.411
2006	26.38	17.913
2007	30.94	22.832
2008	29.51	22.653
2009	32.09	23.669

TABLE 5.2: FIRST MODEL ACCURACY TEST RESULTS

Model Year	Parameter Considered	Parameter Value	Critical Value
2003	*Scaled Deviance*	274.200	329.648
	Scaled Pearson Chi-Square (χ^2)	289.000	329.648
	Deviance/D.O.F	0.746	0.8 ~ 1.2
2004	*Scaled Deviance*	250.323	334.989
	Scaled Pearson Chi-Square (χ^2)	294.000	334.989
	Deviance/D.O.F	.817	0.8 ~ 1.2
2005	*Scaled Deviance*	274.210	308.254
	Scaled Pearson Chi-Square (χ^2)	305.885	308.254
	Deviance/D.O.F	1.019	0.8 ~ 1.2
2006	*Scaled Deviance*	266.763	333.922
	Scaled Pearson Chi-Square (χ^2)	293.000	333.922
	Deviance/D.O.F	1.002	0.8 ~ 1.2
2007	*Scaled Deviance*	251.455	309.325
	Scaled Pearson Chi-Square (χ^2)	270.000	309.325
	Deviance/D.O.F	1.046	0.8 ~ 1.2
2008	*Scaled Deviance*	269.874	332.854
	Scaled Pearson Chi-Square (χ^2)	292.00	332.854
	Deviance/D.O.F	1.105	0.8 ~ 1.2
2009	*Scaled Deviance*	255.299	324.305
	Scaled Pearson Chi-Square (χ^2)	284.000	324.305
	Deviance/D.O.F	0.973	0.8 ~ 1.2

Table 5-2 shows the model accuracy test results for the year 2003~2009. It is desirable that scaled deviance (SD) and the Pearson $\chi 2$ statistic should be less than critical $\chi^2_{0.05}$ value. Deviance/D.O.F should be close to 1 for NB distribution. For all the model years, we can see that the developed model has satisfied the preceding criterion. Basically it shows all the models for different years were fit the data well. Also, the qualitative model accuracy can be observed in Figure 5.1 to 5.7. If the data point lies on the 45^0 line, it means perfect correlation between observed and predicted collision frequency. The calibrated model can explain perfectly the variation in the observed collision data. However, it is not rational. As spatial, temporal, traffic, roadway, collision characteristics and many other variables influence both collision frequency and severity. We can hardly get the perfect information for all those variables. From the plot of observed and predicted collision data, it can be said that there is a very good agreement among those data. However, there are some outliers as can be observed in the Figure 5.1 to 5.7.

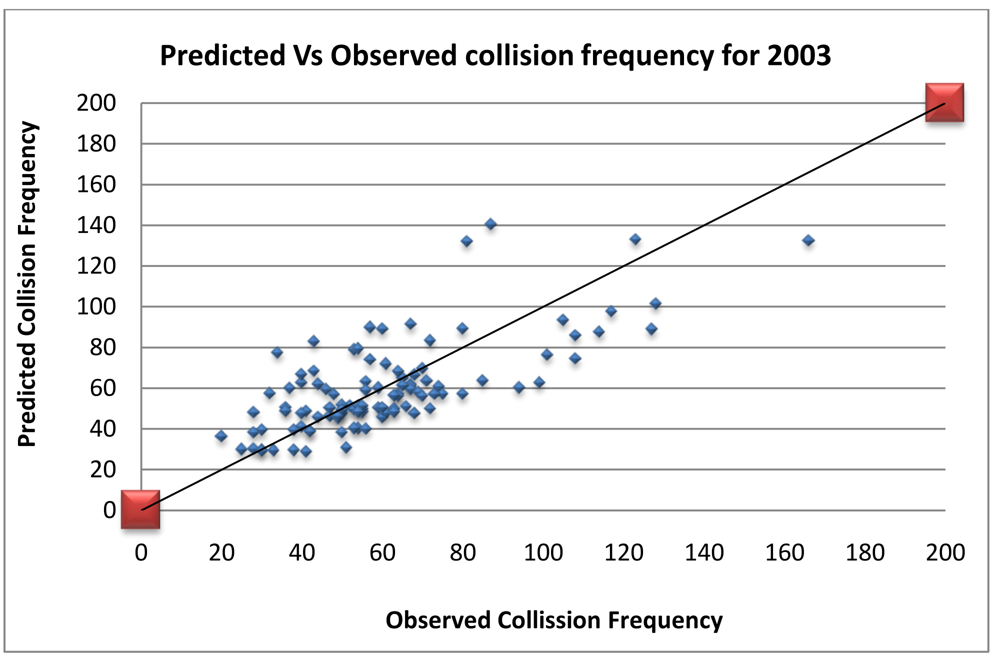


FIGURE 5.1: PREDICTED VS. OBSERVED COLLISION FREQUENCY (2003)

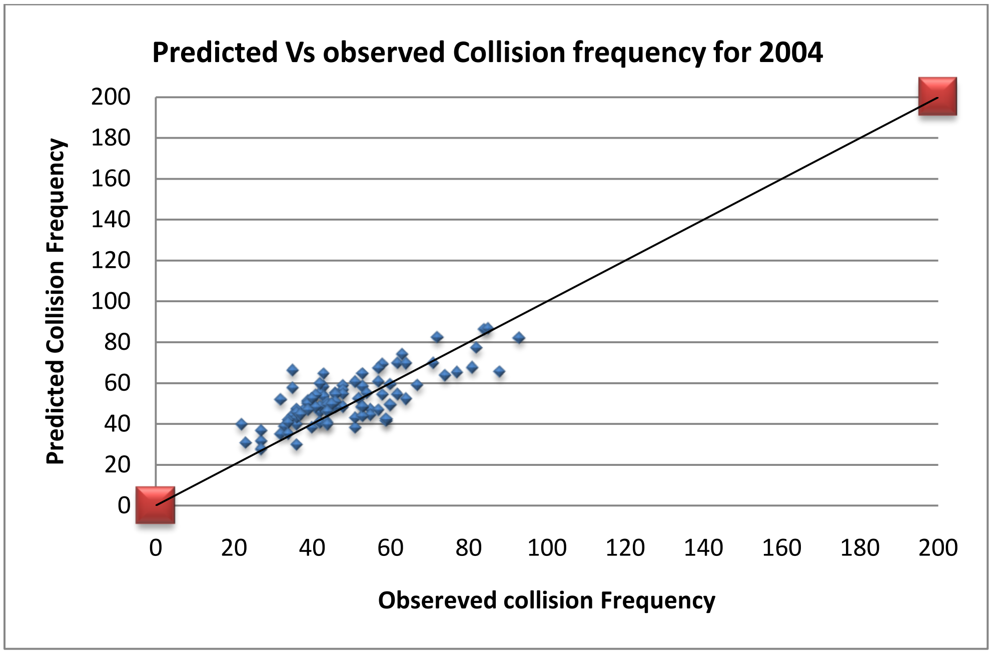


FIGURE 5.2: PREDICTED VS. OBSERVED COLLISION FREQUENCY (2004)

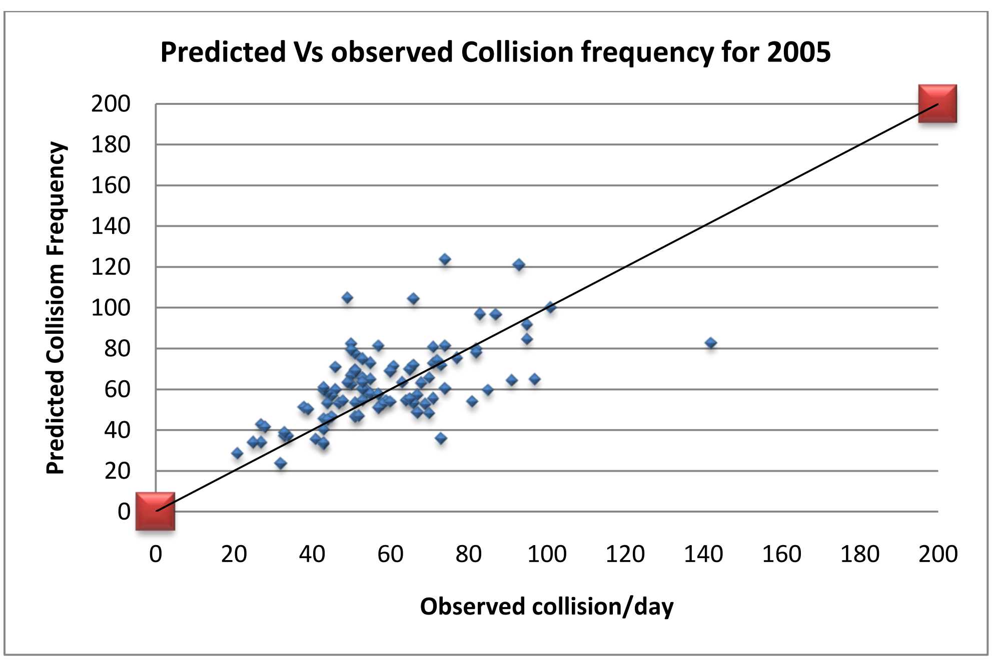


FIGURE 5.3: PREDICTED VS. OBSERVED COLLISION FREQUENCY (2005)

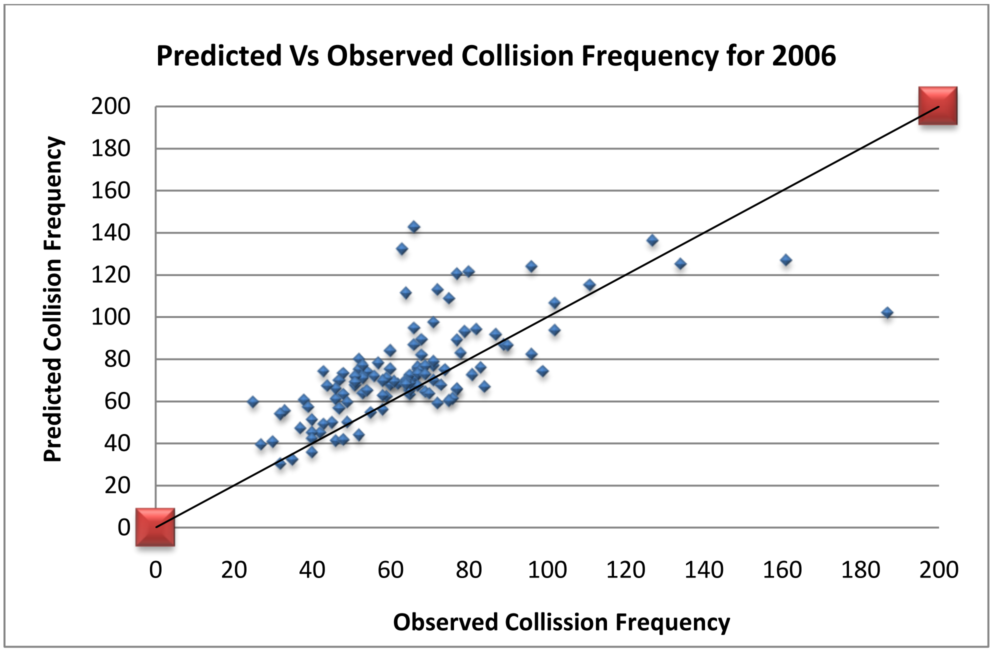


FIGURE 5.4: PREDICTED VS. OBSERVED COLLISION FREQUENCY (2006)

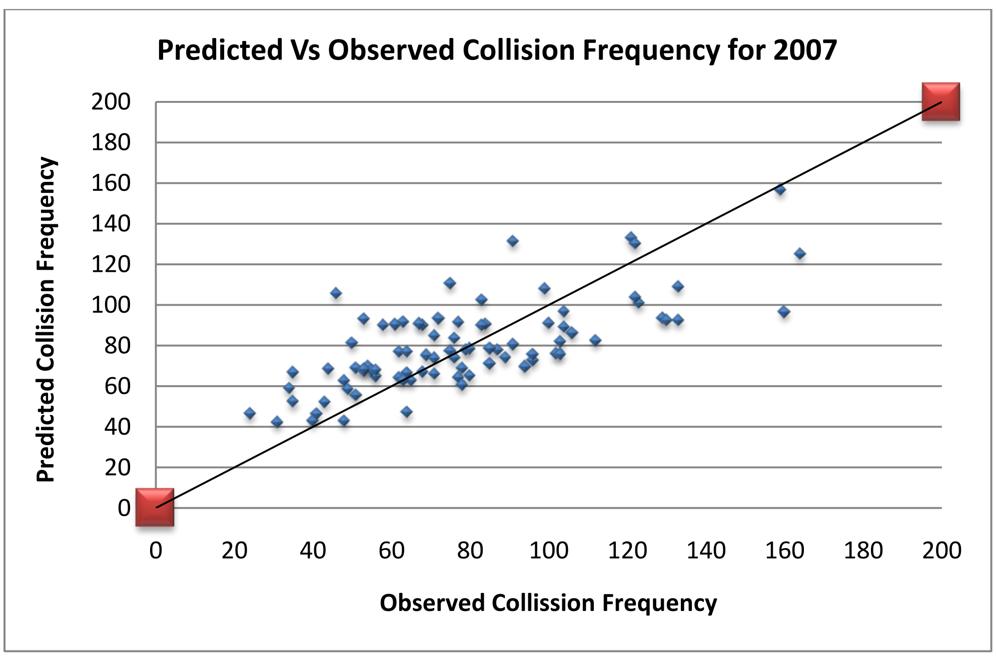


FIGURE 5.5: PREDICTED VS. OBSERVED COLLISION FREQUENCY (2007)

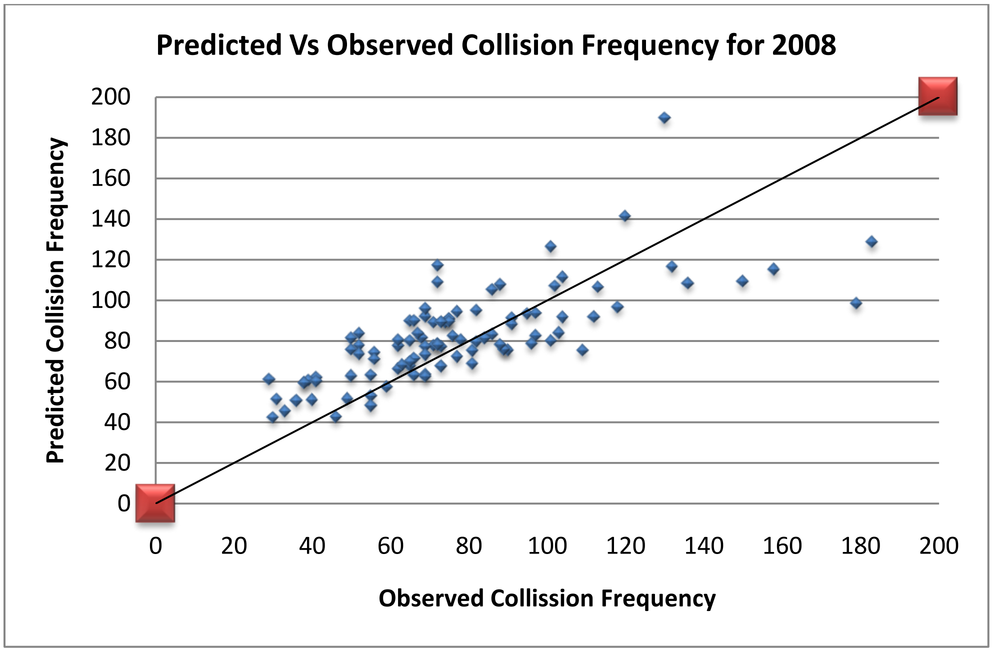


FIGURE 5.6: PREDICTED VS. OBSERVED COLLISION FREQUENCY (2008)

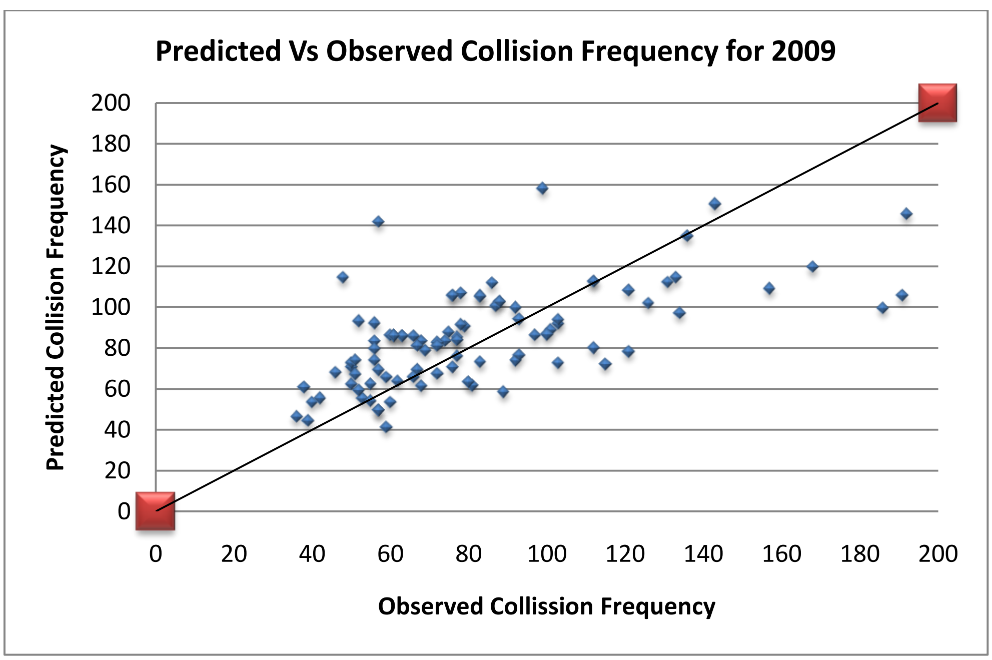


FIGURE 5.7: PREDICTED VS. OBSERVED COLLISION FREQUENCY (2009)

5.3 Trends of First Modal Parameters

One of the objectives of this research was to see the variation in the model parameters for all the seven years. The seven different models were formulated to see these trends and patterns of the model parameters. These trends of the parameter values will be used to predict the future collision frequency. The model parameters for the future years will be predicted based on these formulated model parameters. The Trend of Parameter values for continuous variables, show fluctuation form 2003 to 2009. For β_{10} the trend is decreasing from 2003 to 2005. However, no significant change was observed from 2006 to 2007.

Later on, in the year 2008, there was an increase in the value and then a decrease in the year 2009. The influence of β_{11} increases from 2003 to 2004 and it remains almost similar from 2005 to 2007.The value of β_{11} increases in year 2008 and then decreased in 2009. The value of β_9 dropped from 2003 to 2004 and then increased till 2006. A decreasing trend can be seen from 2007 to 2009.

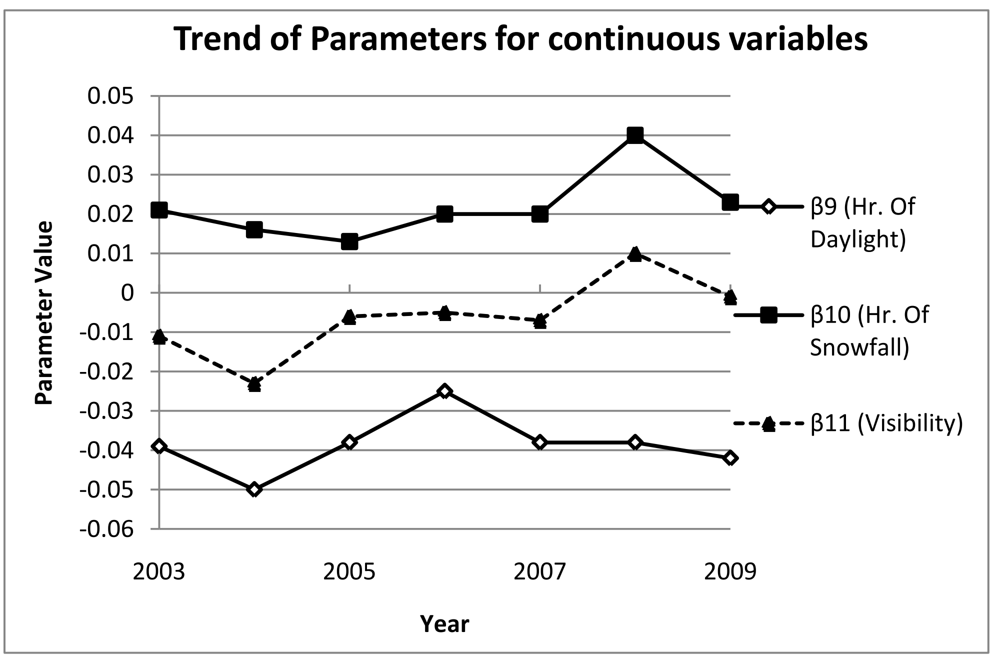


FIGURE 5.8: TREND OF CALIBRATED PARAMETERS FOR CONTINUOUS VARIABLES

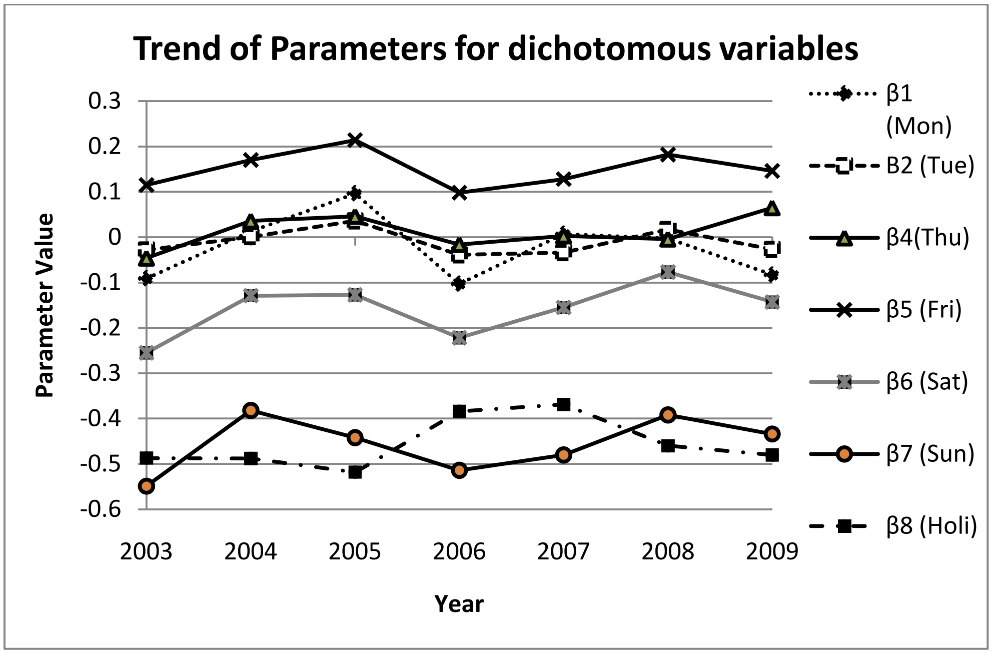


FIGURE 5.9: TREND OF CALIBRATED PARAMETERS FOR DICHOTOMOUS VARIABLES

The Trend of parameter values for dichotomous variable shows that β_6, β_7, β_8 are associated with the negative values, except for β_7 for 2009. Fluctuation of parameter values can be seen from 2003 to 2009 in Figure 5.5 and 5.9. β_2, β_5 β_6 follows the similar trend from 2003 to 2009 with different values. The calibrated parameter of β_7 compliance with the data analysis which shows a decreasing trends of traffic collision on Sunday.

5.4 Second Model Fit

Significance test of the model log likelihood

Logistic regression parameter was computed using maximum likelihood estimation method. The method of maximum likelihood selects the value of the model parameters that produces the distribution most likely to have result in the observed data. This is the principle considering peak stage of driver and underlying probability model. The observed data has a distribution and maximum likelihood finds the parameter. Based on these parameters the values were estimated and it tries to maximize the objective function so that the estimated parameter value distribution and the observed value distribution should match as close as possible. To maximize this objective function we need to maximize the log likelihood function.

The Initial Log Likelihood Function, (-2 Log Likelihood or -2LL) is a statistical measure like total sums of squares in regression. If our independent variables have a relationship to the dependent variable, we will improve our ability to predict the dependent variable accurately, and the log likelihood measure will decrease.

This test is analogous to the F-test for R^2 or change in R^2 value in multiple regressions which tests whether or not the improvement in the model associated with the additional variables is statistically significant. In this model the model Chi-Square values have a good significance. The results of the different model fitting information are provided in the Appendix. So we conclude that there is a significant relationship between the dependent variable and the set of independent variables.

Measure of Analogous to R^2

The next SPSS outputs indicate the strength of the relationship between the dependent variable and the independent variables, analogous to the R^2 measures in multiple regressions. The Cox and Snell R^2 measure operates like R^2, with higher values indicating greater model fit. However, this measure is limited so that it cannot reach the maximum value of 1, so Nagelkerke proposed a modification that had the range from 0 to 1. We will rely upon Nagelkerke's measure as indicating the strength of the relationship. If we applied our interpretive criteria to the Nagelkerke R^2, we would characterize the relationship as weak.

Check for Numerical Problems

There are several numerical problems that can occur in logistic regression that are not detected by SPSS or other statistical packages: multicolinearity among the independent variables. This problem produce large standard errors (over 2) for the variables included in the analysis and very often produce very large B coefficients as well. If we encounter large standard errors for the predictor variables, we should examine frequency tables, one-way ANOVAs, and correlations for the variables involved to try to identify the source of the problem. None of the standard errors or B coefficients is excessively large, so there is no evidence of a numeric problem with this analysis.

Presence of outliers

Multinomial logistic regression does not provide any output for detecting outliers. However, if we are concerned with outliers, we can identify outliers on the combination of independent variables by computing Mahalanobis distance in the SPSS regression procedure.

5.5 Second Model Results

Interpretation of the Results

In this section, we address the following issues:

a. Identifying the statistically significant predictor variables (1, 2)
b. Direction of relationship and contribution to dependent variable

a. Identifying the statistically significant predictor variables (1,2)

1 - There are two outputs related to the statistical significance of individual predictor variables: the Likelihood Ratio Tests and Parameter Estimates. The Likelihood Ratio Tests indicate the contribution of the variable to the overall relationship between the dependent variable and the individual independent variables. The Parameter Estimates focus on the role of each independent variable in differentiating between the

groups specified by the dependent variable. The likelihood ratio tests are a hypothesis test that the variable contributes to the reduction in error measured by the –2 log likelihood statistic. In this model, the variables age and gender are all significant contributors to explain differences in voting preference.

2 - The two equations in the table of Parameter Estimates are labeled by the group which contrasts to the reference group. The first equation is labeled “6 Saturday", and the second equation is labeled “7 Sunday". The coefficients for each logistic regression equation are found in the column labeled as B. The hypothesis that the coefficient is not zero, i.e. changes the odds of the dependent variable event, is tested with the Wald statistic, instead of the t-test as was done for the individual B coefficients in the multiple regression equation.

b. Direction of relationship and contribution to dependent variables

Interpretation of the independent variables is aided by the "Exp (B)" column which contains the odds ratio for each independent variable. We can state that for a particular age category and gender category the likelihood that driver would involve in an accident is increased, decreased or remains same.

TABLE 5.3: SECOND MODEL GOODNESS OF FIT

Year	Test	Chi-Square	Degree of Freedom	Sig.
2003	Pearson	38.157	30.000	0.146
	Deviance	38.101	30.000	0.147
2004	Pearson	20.353	30.000	0.907
	Deviance	20.366	30.000	0.907
2005	Pearson	49.757	30.000	0.013
	Deviance	49.683	30.000	0.013
2006	Pearson	26.833	30.000	0.632
	Deviance	26.827	30.000	0.632
2007	Pearson	33.557	30.000	0.299
	Deviance	33.479	30.000	0.302
2008	Pearson	17.612	30.000	0.964
	Deviance	17.665	30.000	0.964
2009	Pearson	20.353	30.000	0.907
	Deviance	20.366	30.000	0.907

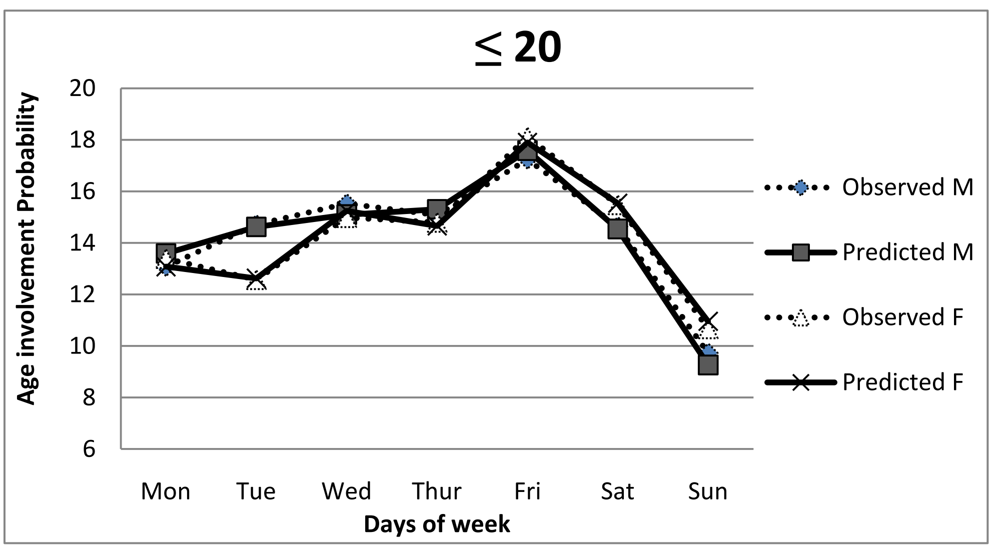


FIGURE 5.10: PREDICTED VS. OBSERVED COLLISION FREQUENCY FOR AGE ≤ 20 (2009)

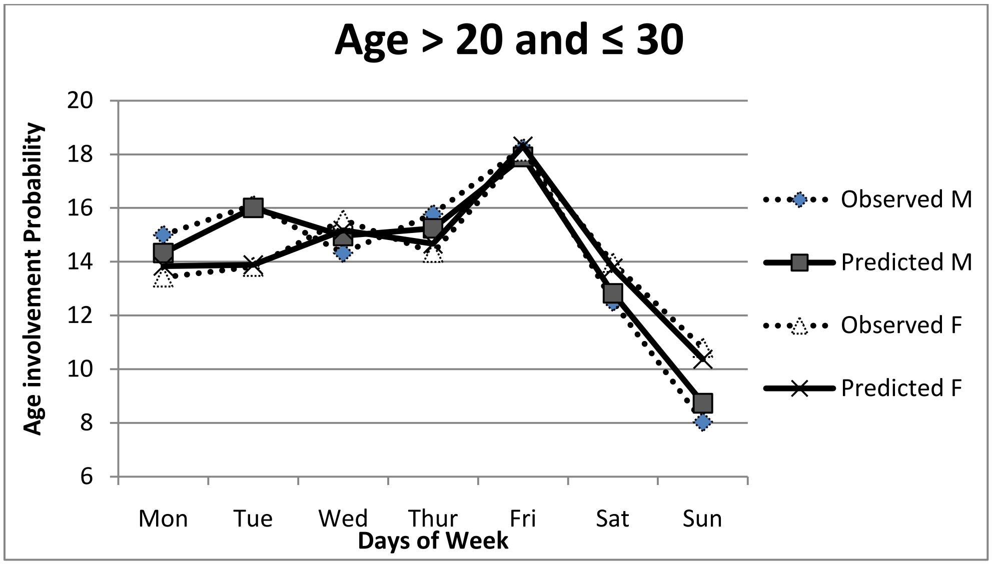


FIGURE 5.11: PREDICTED VS. OBSERVED COLLISION FREQUENCY FOR AGE > 20 AND ≤ 30 (2009)

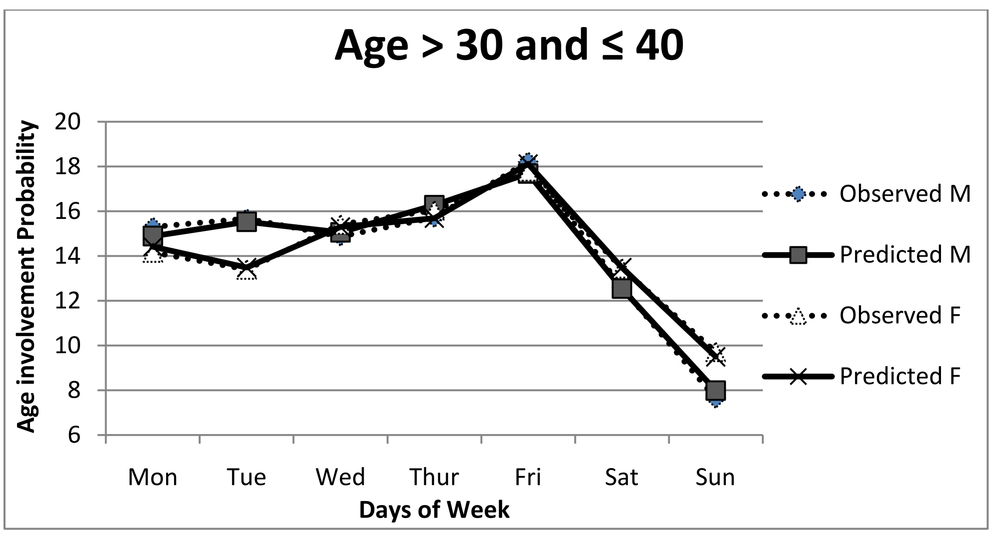


FIGURE 5.12: PREDICTED VS. OBSERVED COLLISION FREQUENCY FOR AGE > 30 AND ≤ 40 (2009)

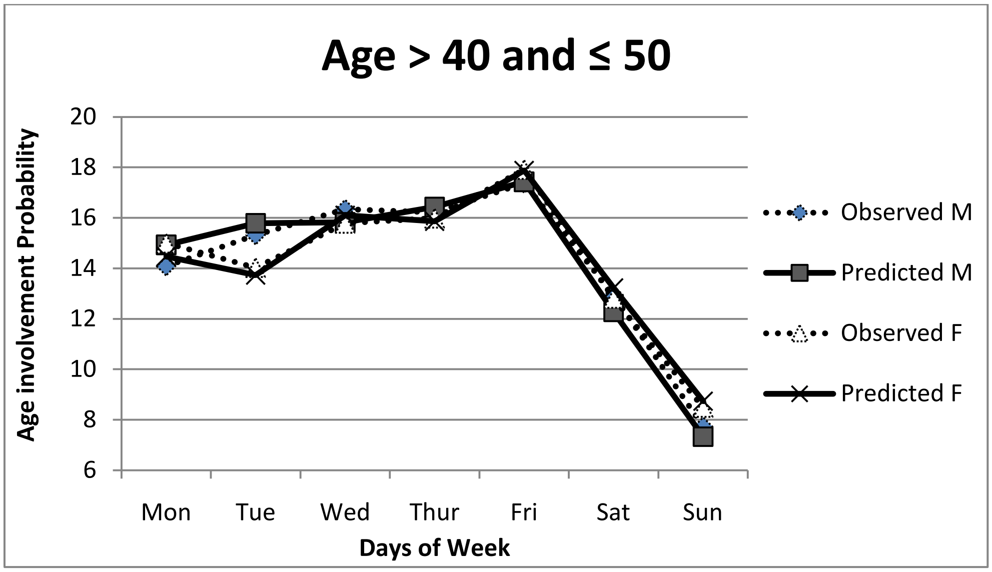


FIGURE 5.13: PREDICTED VS. OBSERVED COLLISION FREQUENCY FOR AGE > 40 AND ≤ 50 (2009)

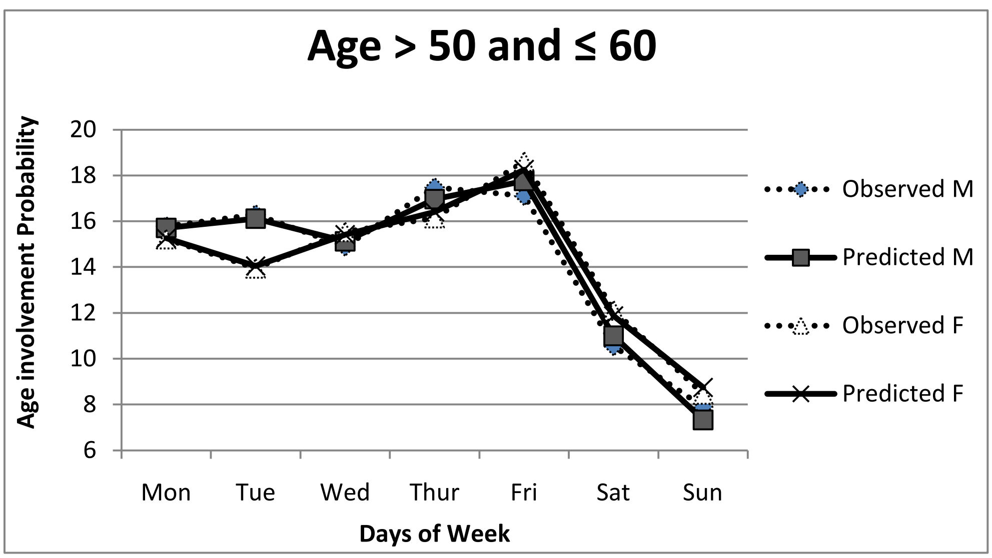


FIGURE 5.14: PREDICTED VS. OBSERVED COLLISION FREQUENCY FOR AGE > 50 AND ≤ 60 (2009)

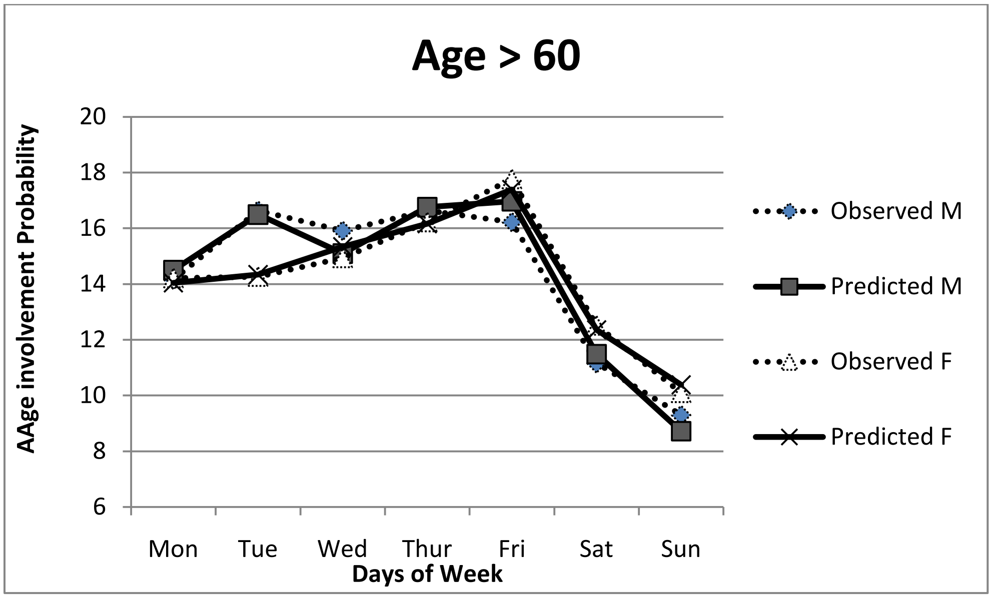


FIGURE 5.15: PREDICTED VS. OBSERVED COLLISION FREQUENCY FOR AGE > 20 (2009)

5.6 Summary

The model accuracy test results were very good for both models. From the statistical measures, the significant impact on traffic collisions was found by the number of snowfall hours and daylight hours in the first model. To check the model accuracy the observed and predicted collision frequencies were compared. The second collision prediction model is perceived to be very good from the model accuracy results. The statistical measures show that increase of both number of snowfall hours and number of daylight hours results in increased number of traffic collision. Proposed negative binomial model and multinomial regression model were found good. Age and Gender categories have different probabilities for involving in a collision for different days of the week. The first model could provide a realistic estimate of expected collision frequency for a particular day as a function of number of hours of daylight, number of hours of snowfall and visibility. The second model can be used in future for analysis to see the trend of the each age and gender category defined in the model.

CHAPTER 6. SUMMARY AND RECOMMENDATIONS

This report has outlined the development of two different collision prediction models. The results and achievements of the work performed are summarized in this chapter. The limitations and the future scope of work are also part of this chapter.

6.1 Summary

Several collision prediction models have been developed in the past depending on different research objectives. The main objective of this work was to develop the collision prediction models for estimating the safety potential for the City of Edmonton. The City of Edmonton consistently facing a problem of collision spikes on Friday. Consistently higher number of traffic collisions was observed on Friday as compared to other days of the week from 2003 to 2009 in Edmonton, Canada. This study was carried out to answer the unchanging Friday spikes in the city. The motivation of the research was to find out the probable causes of these Friday spikes and what counter measures can be implemented to improve the current traffic safety scenario of the City of Edmonton.

The collision data were derived from the police reports for collisions involving fatality, injury and property damage only equal to or greater than $1000. The data were investigated and two traffic collision prediction models were formulated. This study considers collision contributing factors: 1) Number of daylight hours, 2) Number of snowfall hours, and; 3) Visibility on daily basis. The dummy variables were incorporated in the first model to capture the exposure of the day of week and the holidays. In this study three data sources were compiled and used: 1) MVCIS (Motor Vehicle Collision Information System); 2) Sunrise/Sunset data; and 3) National Climate Data and Information Archive.

The variance was perceived significantly greater than the mean from the collision data. Therefore, due to over dispersion of the collision data, a Negative Binomial (NB) model was formulated to establish a relation between the collision frequency and the independent variables. The generalized linear model procedure in SPSS was adopted to estimate the parameters of the NB model. The model accuracy test results were very good. The significant impact on traffic collisions was found by the number of snowfall hours and daylight hours. Increase of number of snowfall hours results in increased number of traffic collision as shown by statistical measures. Proposed negative binomial model was found good and this model can be applied in the cities having long winter conditions. It could provide a realistic estimate of expected collision frequency for a particular day as a function of number of hours of daylight, number of hours of snowfall and visibility.

This research examined the effect of different factors on traffic safety of the City of Edmonton. To check the model accuracy for the second collision prediction mode, the observed and predicted collision frequencies were compared. The second collision prediction model is perceived to be very good. The model accuracy test results were very good. This model is built for the collision probability of gender and different age groups for either driver or passenger involved in the collision for the City of Edmonton. The stochastic nature of the collision prediction model was considered in the second collision prediction model. This model predicts the percentage of the different age groups and gender of both drivers and passengers involved in the collisions considering that the collision has happened. The minimum 10 number of cases per independent variable should be available. The dependent variable is Day of week for each collision record and it was considered in 7 categories: Monday = 1, Tuesday = 2, Wednesday = 3, Thursday = 4, Friday = 5, Saturday = 6, Sunday = 7. There are two independent variables considered in this study. The first independent variable is Age of Driver or Passenger for each collision record and it was considered in 6 age categories as below ≤ 20; > 20 and ≤ 30; > 30 and ≤ 40; > 40 and ≤ 50; > 50 and ≤ 60 and; > 60. The second independent variable is Gender of people involved in each collision record and it was categorized as Male and Female.

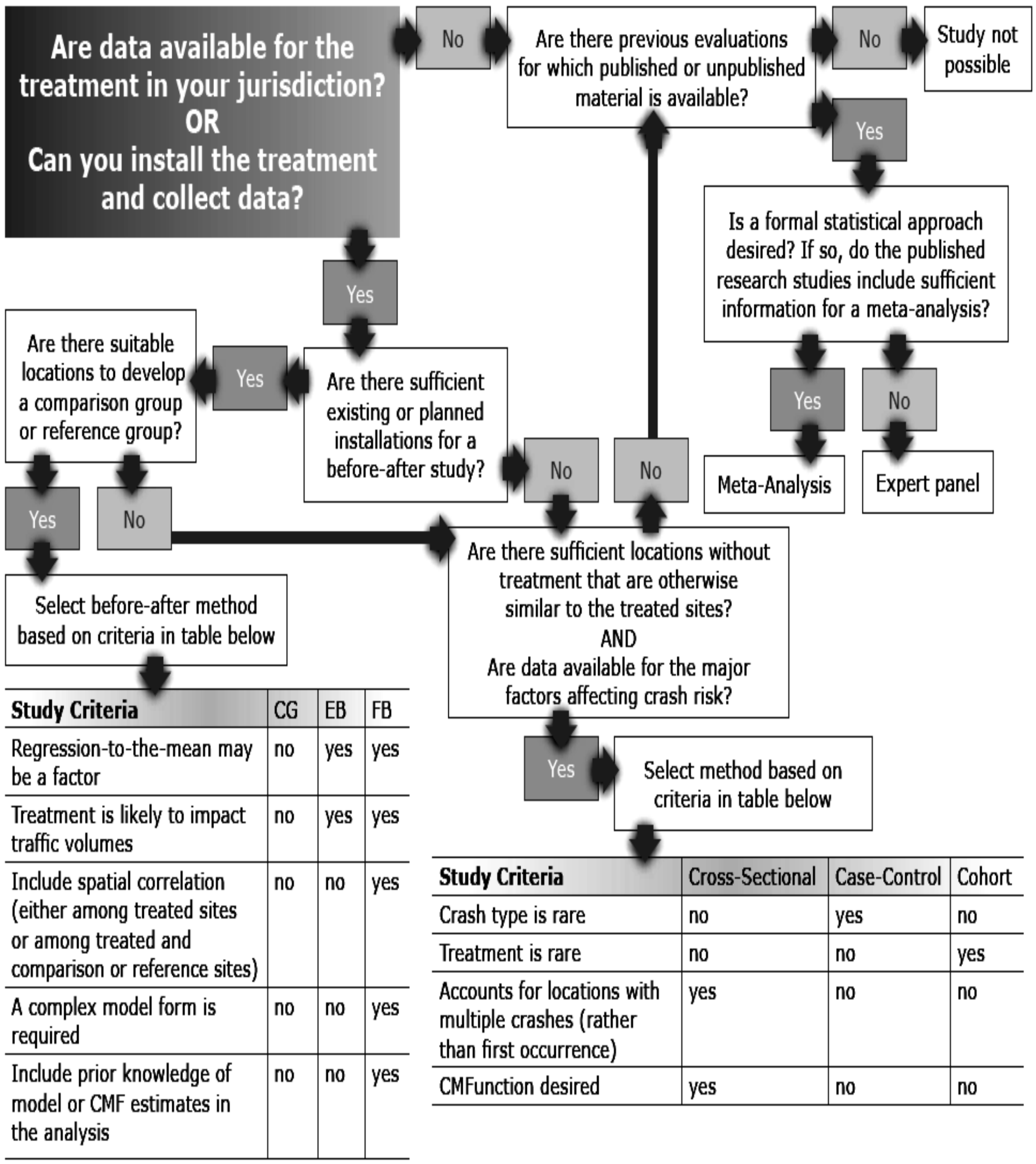
Principle for this model is that for a peak stage of driver and underlying probability model the method of maximum likelihood selects the value of the model parameters that produce the distribution most likely to have result in the observed data. The observed data has a distribution and maximum likelihood finds the parameter. In this model, the variables age and gender are all significant contributors to explaining differences in voting preference. We can state that for a particular age category and gender category the likelihood that driver would involve in an accident is increased, decreased or remains same.

The Age and Gender categories have different probabilities for involving in a collision for different days of the week. The second model can be used in future for analysis to see the trend of the each age and gender category defined in the model. It is hoped that predicted collision frequency will help the decision makers to quantify traffic safety of Edmonton and improving the existing scenario.

The implication of the work which has been done suggests that we can predict collision frequency if we know the environment factors such as number of hours of snowfall, number of hours of daylight and visibility on a particular day. Also the probability of a collision can be predicted if we know the age and gender of the people involved in the collision.

6.2 Study Design Procedure

Based on the availability and the project goals the preferred study design can be selected from the flowchart shown below (Figure 6.1). As shown in the flow chart the first step is to identify whether or not a collision based evaluation will be possible for the treatment of interests. It is required to check whether sufficient data available for the treatment or can you install and collect the data for the treatment. The answer to the first question will help to determine appropriate evaluation in the steps e.g., before-after evaluation, cross-sectional evaluation, etc. Also it is identified further if it will be necessary to develop a CMF (Crash Modification Factor) using meta-analysis or an expert panel. Several additional questions as shown in the flow chart will guide the user to identify the appropriate study design procedure. Through the different thought process it is required to identify an appropriate study design, alternative approach, or to conclude that it is not possible to develop a CMF at present. In the Flow Chart, EB, FB and CG represents Empirical Bayes, Full Bayes and Comparison Group respectively.



Study Criteria	CG	EB	FB
Regression-to-the-mean may be a factor	no	yes	yes
Treatment is likely to impact traffic volumes	no	yes	yes
Include spatial correlation (either among treated sites or among treated and comparison or reference sites)	no	no	yes
A complex model form is required	no	no	yes
Include prior knowledge of model or CMF estimates in the analysis	no	no	yes

Study Criteria	Cross-Sectional	Case-Control	Cohort
Crash type is rare	no	yes	no
Treatment is rare	no	no	yes
Accounts for locations with multiple crashes (rather than first occurrence)	yes	no	no
CMFunction desired	yes	no	no

FIGURE 6.1: FLOW CHART FOR STUDY DESIGN PROCEDURE

Source: Figure 5, Office of Safety, FHWA, 2010

6.3 Recommended Countermeasures

A CMF Clearinghouse provides information for the set of Crash Modification Factors (CMF) identified from the different studies. A five point quality rating serves as the primary method for identifying the quality of a CMF. Based on the descriptive data analysis and the literature review the following are the recommended counter measures finalized from this study:

TABLE 6.1: SNOWFALL RELATED COUNTERMEASURES

1	**Effects of snow fences and higher state of preparedness for the snow winter season**							
	CMF	**CRF (%)**	**Quality**	**Crash Type**	**Crash Severity**	**Roadway Type**	**Area Type**	**Reference**
	0.89	11		All	All	All	All	Elvik, R. and Vaa, T., 2004
2	**Increased pavement friction**							
	CMF	**CRF (%)**	**Quality**	**Crash Type**	**Crash Severity**	**Roadway Type**	**Area Type**	**Reference**
	0.76	24		All	All	All	All	Harkey et al., 2008
3	**Effects of use of salt (chemical de-icing) during the whole winter season (baseline = no salt)**							
	CMF	**CRF (%)**	**Quality**	**Crash Type**	**Crash Severity**	**Roadway Type**	**Area Type**	**Reference**
	0.85	15		All	Serious injury, Minor injury	All	All	Elvik, R. and Vaa, T., 2004
4	**Raise standard for winter maintenance**							
	CMF	**CRF (%)**	**Quality**	**Crash Type**	**Crash Severity**	**Roadway Type**	**Area Type**	**Reference**
	0.89 [B]	11		All	Serious injury, Minor injury	All	All	Elvik, R. and Vaa, T., 2004
5	**Short-term effects of all measures to control snow, slush or ice**							
	CMF	**CRF (%)**	**Quality**	**Crash Type**	**Crash Severity**	**Roadway Type**	**Area Type**	**Reference**
	0.5	50		All	Not specified	All	All	Elvik, R. and Vaa, T., 2004

TABLE 6.2: ILLUMINATION RELATED COUNTERMEASURES

1	**Provide Highway Lighting**							
	CMF	**CRF (%)**	**Quality**	**Crash Type**	**Crash Severity**	**Roadway Type**	**Area Type**	**Reference**
	0.37	64		All	Fatal	All	Urban	Elvik, R. and Vaa, T., 2004

TABLE 6.3: VISIBILITY OF SIGNS RELATED COUNTERMEASURES

1	**Install signs to conform to MUTCD**							
	CMF	**CRF (%)**	**Quality**	**Crash Type**	**Crash Severity**	**Roadway Type**	**Area Type**	**Reference**
	0.85	15		All	Serious injury, Minor injury	Local	Urban	Elvik, R. and Vaa, T., 2004
	0.93	7		All	Property Damage Only (PDO)	Local	Urban	Elvik, R. and Vaa, T., 2004
2	**Improve visibility of signal heads**							
	CMF	**CRF (%)**	**Quality**	**Crash Type**	**Crash Severity**	**Roadway Type**	**Area Type**	**Reference**
	0.93	7		Nighttimes	All	Not specified	Urban	Sayed et al., 2007

TABLE 6.4: SPEED MANAGEMENT RELATED COUNTERMEASURES

1	**Traffic calming**							
	CMF	**CRF (%)**	**Quality**	**Crash Type**	**Crash Severity**	**Roadway Type**	**Area Type**	**Reference**
	0.68	32		All	All	Minor Collector	Urban	Elvik, R. and Vaa, T., 2004
	0.67	33		All	Serious injury, Minor injury	Minor Collector	Urban	Elvik, R. and Vaa, T., 2004
	0.67	33		All	Serious injury, Minor injury	Minor Collector	Urban	Elvik, R. and Vaa, T., 2004
	0.75	25		All	Property Damage Only (PDO)	Minor Collector	Urban	Elvik, R. and Vaa, T., 2004

	Area-wide traffic calming							
	CMF	**CRF (%)**	**Quality**	**Crash Type**	**Crash Severity**	**Roadway Type**	**Area Type**	**Reference**
2	0.82 [B]	18	★★★☆☆	All	Serious injury, Minor injury	Local	Urban	Elvik, R. and Vaa, T., 2004
	0.89 [B]	11	★★★★★	All	Serious injury, Minor injury	All	Urban	Elvik, R. and Vaa, T., 2004
	0.94	6	★★★☆☆	All	Serious injury, Minor injury	Minor Collector	Urban	Elvik, R. and Vaa, T., 2004
	0.94	6	★★★☆☆	All	Property Damage Only (PDO)	Local	Urban	Elvik, R. and Vaa, T., 2004
	0.95	5	★★★☆☆	All	Property Damage Only (PDO)	All	Urban	Elvik, R. and Vaa, T., 2004
	0.97	3	★★★☆☆	All	Property Damage Only (PDO)	Minor Collector	Urban	Elvik, R. and Vaa, T., 2004
	Install transverse rumble strips as traffic calming device							
	CMF	**CRF (%)**	**Quality**	**Crash Type**	**Crash Severity**	**Roadway Type**	**Area Type**	**Reference**
3	0.64	36	★★★★☆	All	Serious injury, Minor injury	Local	Urban and Suburban	Elvik, R. and Vaa, T., 2004
	0.66	34	★★★★☆	All	All	Local	Urban and Suburban	Elvik, R. and Vaa, T., 2004
	0.73	28	★★☆☆☆	All	Property Damage Only (PDO)	Local	Urban and Suburban	Elvik, R. and Vaa, T., 2004

	CMF	CRF (%)	Quality	Crash Type	Crash Severity	Roadway Type	Area Type	Reference
	Install new fluorescent curve signs or upgrade existing curve signs to fluorescent sheeting							
4	0.82	18	★★★★☆	Non-intersection	All	All	Rural	Srinivasan et al., 2009
	0.82	18	★★★★☆	Head on, Non-intersection, Run off road, Sideswipe	All	All	Rural	Srinivasan et al., 2009
	0.75	25	★★★★☆	Non-intersection	Fatal, Serious injury, Minor injury	All	Rural	Srinivasan et al., 2009
	0.66	34	★★★★☆	Head on, Nighttimes, Non-intersection, Run off road, Sideswipe	All	All	Rural	Srinivasan et al., 2009
	0.65	35	★★★☆☆	Nighttimes, Non-intersection	All	All	Rural	Srinivasan et al., 2009

6.4 Limitation of Study

Data availability was limited. For example, the daily traffic volume data were not available for the analysis and model formulation. Previous studies show that the variation in the daily traffic volume is the major reason of the different collision rates by day of week. The daily traffic volume could be one of the major contributors to the spikes on Friday in the City of Edmonton.

The study considers the whole City of Edmonton to improve traffic safety of the city. The information for some collision records was unavailable. Therefore, not all the collisions were used to formulate the collision prediction model. If all information in the collision records was made available then the estimated model parameters could be different.

6.5 Future Scope

The future scope of this research is as follows:

A. IMPACT FOR TRAFFIC COLLISION RATE

1. The next phase of the research will focus on the high collision locations in the City of Edmonton. The specific intersections and corridors will be investigated to see the different types of the collisions occurring there. The study will focus on defining the patterns and trends of the collisions like rear end collisions, head to head collisions, and etc. Also, once the data was made available additional study must analyze variation of daily traffic. The next phase of the safety research in the City of Edmonton may focus on a macro-zone basis approach. The city is divided to different regions, so that the volume of traffic of each region could be estimated and used in the model formulation.

2. Considering the data availability, we may choose 97 Street and Whitemud drive corridors in the next phase to analyze the relationship between collision rates and traffic flow/speed. The Transportation research group of University of Alberta has calibrated the 2010 VISSIM model for Whitemud Drive in Edmonton. Whitemud Drive can be used in future in micro-simulation VISSIM model to validate the collision countermeasures. Variable speed limit can be based on local- or time-specific road conditions. It is typically implemented with enforcement.

3. The work or construction zones produce significant impact on traffic. Many construction activities were started in the year 2008 with a corresponding restriction in traffic movement. This might be one of the reasons of the increase in traffic collisions in year 2008. We are proposing to analyze the impact between collision occurrence and construction/work zone. Restricted transition areas should be as little as possible, recognizing that drivers will not reduce their speed unless essential.

B. COLLISION COUNTERMEASURES

1. The next phase of the research will focus to develop a process that allows identifying the best counter measures for traffic collisions and estimating the Crash Modification Factors in the location based research approach.

2. The next phase of the research will consider the Active Traffic Management strategies like Variable Speed Limit (VSL) to improve the safety on the selected corridors. The VSL is a promising control strategy to improve safety. The City of Edmonton has planned operational field tests for mobility from June to October in 2011. We propose to conduct a detailed before and after analysis for safety then. This research will identify the issues related to the mobility and safety in the City of Edmonton.

REFERENCES

1. The City of Edmonton (2008). "Annual Report." Finance and Treasury Department of the City of Edmonton.
2. The City of Edmonton. (2010) "Municipal Census." <http://www.edmonton.ca/city_government/municipal-census.aspx> (Aug 24, 2010).
3. Cebryk, G., and Bell, T. (2004). "Traffic Safety at Intersections: The Edmonton Experience." *Annual Conference of the Transportation Association of Canada*, 1-11.
4. Edmonton Police Service. (2010). "High Collision Locations." <http://www.edmontonpolice.ca/TrafficVehicles/TrafficCollisions/HighCollisionLocations.aspx > (October 21, 2010).
5. The City of Edmonton. (2010) *Edmonton Maps*, <http://maps.edmonton.ca/Scripts/main.asp> Accessed on Aug 09, 2010
6. The City of Edmonton. (2006). "Motor Vehicle Collisions Monitoring 2005." Transportation planning department, <http://www.edmonton.ca/transportation/RoadsTraffic/2005_MVCIS_Summary.pdf> (June 10, 2010).
7. Anielski, M. (2004). "Alberta Traffic Safety Progress Report - Key Indicators and Trends." *Anielski Management inc. and Alberta Motor Association*, <http://www.ama.ab.ca/images/images_pdf/traffic_progress_reportS04_3F.pdf > (June 02, 2010).

8. Roger P. R., Prassas E. S. and McShane W. R. (2010). "Traffic Engineering." Fourth edition, Upper Saddle River, NJ: Pearson, c2011.
9. Garber N. J. and Hoel L. A. (2010). "Traffic and Highway Engineering." Fourth edition, Cenage Learning, USA.
10. "Traffic Safety Facts - 2008 Data." US Department of Transportation. Published by National Highway Traffic Safety Administration's National Center for Statistics and Analysis, <http://www-nrd.nhtsa.dot.gov/Pubs/811162.PDF> (June 02, 2010)
11. Transport Canada. (2010). "Further drop in road fatalities: 2007 Canadian motor vehicle traffic collision statistics." No. H036/10, <http://www.tc.gc.ca/eng/mediaroom/releases-2010-h036e-5883.htm> (June 06, 2010).
12. Alberta Transportation. (2008). "Alberta Traffic Collision Statistics 2008." Published by Office of Traffic Safety, <http://www.transportation.alberta.ca/Content/docType47/Production/2008AR.pdf> (May 19, 2010).
13. Alberta Transportation. (2010). "Stats and Facts." Government of Alberta, <http://www.transportation.alberta.ca/Content/docType437/Production/StatsandFacts.pdf> (June 04, 2010).
14. Li. K., and Waters N. (2005). "Transportation Networks, Case-Based Reasoning and Traffic Collision Analysis: A Methodology for the 21st Century". *Advances in Spatial Science*, Part A, 63-92.
15. Office of Traffic Safety. (2010) "Motor Vehicle Collisions 2009 - Interim Report." Published by The City of Edmonton, April 2010. <http://www.edmonton.ca/transportation/2009AnnualCollisionReport.pdf> (May 21, 2010).
16. Khonakar, B., Sayed T., and Lovegrove G. (2010). "Transferability of Community-Based Collision Prediction Models for Use in Road Safety

Planning Applications." *Journal of Transportation Engineering*, 136(10), 871-880.

17. Emmel, P. (2007). "Missouri Holiday Crashes Report." Statistical Analysis Center, Missouri State Highway Patrol, <http://www.mshp.dps.missouri.gov/MSHPWeb/SAC/pdf/2007HolidayReport.pdf> (July 23, 2010).
18. Liu Z., Sharma, S., and Datla, S. (2008). "The Necessity of Understanding Temporal and Spatial Variations of Holiday Traffic." *ITE Journal on the Web*, 78(11), 69-74.
19. Andresscu, M., and Frost D. B. (1998). "Weather and Traffic Accidents in Montreal, Canada." *Climate Research*, 9(3), 225-230.
20. McGwin G. Jr., and Brown D. B. (1999). "Characteristics of Traffic Collisions among Young, Middle-aged, and Older drivers." *Accident Analysis and Prevention*, 31(3), 181-198.
21. King County Road Service. (2005 - 2007). "Traffic Safety reports." Department of Transportation, Washington state, <http://www.kingcounty.gov/transportation/kcdot/roads/trafficmaintenanceengineering/trafficsafetyreports.aspx> (May 10, 2010).
22. Doherty, S. T., Andrey J. C., and MacGrgor C. (1998). "The Situational Risks of Young Drivers: The Influence of Passengers, Time of Day and Day of Week on Accident Rates." *Accident Analysis and Prevention*, 30(1), 45-52.
23. "Saskatchewan Traffic Accident Facts 2006." Saskatchewan Government Insurance (2010), <http:// www.sgi.sk.ca> (June 08, 2010).
24. Xie, Y., Lord D., and Zhang Y. (2007). "Prediction Motor Vehicle Collisions Using Bayesian Neural Models." *Accident Analysis and Prevention*, 39(5), 922-233.
25. Staubach, M. (2009). "Factors Correlated with Traffic Accidents as a Basis for Evaluating Advanced Driver Assistance Systems." *Accident Analysis and Prevention*, 41(5), 1025-1033.

26. Jang, H., Lee S., and Kim S.W. (2010). "Bayesian Analysis for Zero-inflated Regression Models with the Power Prior: Applications to Road Safety Countermeasures." *Accident Analysis and Prevention*, 42(2), 540-547.
27. Lord, D., Washington S.P., and 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*. 37(1), 35-46.
28. Hauer, E. (1997). "Observational Before-After Studies in Road Safety: Estimating the Effect of Highway and Traffic Engineering Measures on Road Safety." *Oxford, England: Pergamon Press*, Elsevier Science Ltd.
29. Hinde, J., and Demetrio C. G. B. (1998). "Over-Dispersion: Model and Estimation." *Computational Statistics and Data Analysis*, 27(2), 151-70.
30. Kim, H., Sun D., and Tsutakawa R. K. (2002). "Lognormal vs. Gamma: Extra Variations." *Biometrical Journal*, 44(3), 305-323.
31. Nofal, FH., and Saeed, AAW. (1997). "Seasonal Variation and Weather Effects on Road Traffic Accidents in Riyadh City." *Public Health*, 111(1), 51-55.
32. Smith Daniel J. (2000). "Human Factors and Traffic Crashes." MTC Transportation Scholars Conference Ames, Iowa, 136-151.
33. Sawalha, Z., and Sayed, T. (2001). "Evaluating Safety of Urban Arterial Roadways." *Journal of Transportation Engineering*, 127(2), 151-158.
34. Miranda-Moreno, L. F., Fu. L., Saccomanno F. F. and Labbe A. (1995). "Alternative Risk Models for Ranking Locations for Safety Improvements." *Transportation Research Record: Journal of Transportation Research*, 1908, 1-8.
35. Caliendo, C., Guida M., and Parisi A. (2007). "A Collision Prediction Model for Multilane Roads." *Accident Analysis and Prevention*, 39(4), 657-670.

36. Brijs, T., Karlis D. and Wets G. (2008). "Studying the Effect of Weather Conditions on Daily Crash Counts Using a Discrete Time-series Model." *Accident Analysis and Prevention*, 40 (3), 1180-1190.
37. National Research Council. (2010). "Sunrise/Sunset data". Maintained by National Research Council - Herzberg Institute of Astrophysics, Canada, <http://www.nrc-cnrc.gc.ca/eng/services/hia/sunrise-sunset.html> (June 30, 2010).
38. National Climate Data and Information archive. (2010). Maintained by Environment Canada, <http://www.climate.weatheroffice.gc.ca/climateData/canada_e.html> (June 30, 2010).
39. McCullagh, P., and J. A. Nelder. (1989). "Generalized Linear Models." Chapman and Hall, New York.
40. El-Basyouncy K., and Sayed T. (2006). "Comparison of Two Negative Binomial Regression Techniques in Developing Accident Prediction Models." *Transportation Research Record: Journal of the Transportation Research Board*, 1950, 9-16.
41. Hosmer David W. and Stanley Lameshow. (2005). "Applied Logistic Regression." Second edition, John wiley and Sons, Inc.
42. Hasan, S. (2007) "Development of Travel Demand Model for Dhaka City." Master of Science Thesis, Bangladesh University of Engineering and Technology, Bangladesh.
43. "An Introduction to the Highway Safety Manual." *American Association of State Highway and Transportation Officials (AASHTO)*, Washington, DC. (2010).
44. "Highway Safety Manual Knowledge Base." *American Association of State Highway and Transportation Officials (AASHTO)*, Washington, DC. (2009).

45. Geni Bahar (2010). "Methodology for Development and Inclusion of Crash Modification Factors in the First Edition of the Highways Safety Manual." Transportation Research Board, Transportation Research Circular E-C142.
46. "Collision Modification Factors Clearinghouse" (2010). <http://www.cmfclearinghouse.org/index.cfm> (December 07, 2010)
47. Gross. F. et al. (2010) "A Guide to Developing Quality Crash Modification Factors." *US Department of Transportation*, Federal Highway Administration, Report for the FDOT, Dec. 2010.
48. Sayed T. and Paul de Leur. (2008). "Collision Modification Factors for British Columbia." Prepared for Engineering Branch BC Ministry of Transportation and Infrastructure.
49. Zegeer, C. V. et al. (2004). "Guidance for Implementation of the AASHTO Strategic Highway Safety Plan. Volume 12: A Guide for Reducing Collisions at Signalized Intersections." National Cooperative Highway Research Program. Report 500, Washington, D.C. 2004.

APPENDICES

OBSERVED AND PREDICTED FREQUENCIES 2003

Gender	Age	Day Of Week		Frequency Observed	Frequency Predicted	Frequency Pearson Residual	Percentage Observed	Percentage Predicted
0	1	1	Mon	236	236.512	-0.036	14.640	14.672
		2	Tue	260	237.661	1.569	16.129	14.743
		3	Wed	233	230.756	0.160	14.454	14.315
		4	Thur	247	248.653	-0.114	15.323	15.425
		5	Fri	262	267.615	-0.376	16.253	16.601
		6	Sat	225	225.317	-0.023	13.958	13.977
		7	Sun	149	165.486	-1.353	9.243	10.266
	2	1	Mon	486	483.922	0.103	15.351	15.285
		2	Tue	464	481.177	-0.850	14.656	15.198
		3	Wed	513	489.084	1.176	16.203	15.448
		4	Thur	499	498.518	0.024	15.761	15.746
		5	Fri	555	545.418	0.451	17.530	17.227
		6	Sat	341	353.280	-0.693	10.771	11.159
		7	Sun	308	314.600	-0.392	9.728	9.937
	3	1	Mon	431	438.357	-0.383	15.782	16.051
		2	Tue	458	449.209	0.454	16.770	16.449
		3	Wed	424	429.867	-0.308	15.525	15.740
		4	Thur	462	429.899	1.687	16.917	15.741
		5	Fri	475	486.007	-0.551	17.393	17.796
		6	Sat	275	288.460	-0.838	10.070	10.562
		7	Sun	206	209.203	-0.230	7.543	7.660
	4	1	Mon	467	452.933	0.725	17.451	16.926
		2	Tue	439	451.036	-0.622	16.405	16.855
		3	Wed	427	426.698	0.016	15.957	15.945
		4	Thur	386	416.043	-1.603	14.425	15.547
		5	Fri	424	434.104	-0.530	15.845	16.222

Gender	Age	Day Of Week		Frequency			Percentage	
				Observed	Predicted	Pearson Residual	Observed	Predicted
		6	Sat	296	279.857	1.020	11.061	10.458
		7	Sun	237	215.329	1.540	8.857	8.047
	5	1	Mon	227	224.644	0.171	15.775	15.611
		2	Tue	245	241.490	0.248	17.026	16.782
		3	Wed	192	213.150	-1.570	13.343	14.812
		4	Thur	234	243.882	-0.694	16.261	16.948
		5	Fri	275	255.059	1.377	19.110	17.725
		6	Sat	144	145.438	-0.126	10.007	10.107
		7	Sun	122	115.337	0.647	8.478	8.015
	6	1	Mon	131	141.632	-0.974	14.670	15.860
		2	Tue	142	147.427	-0.489	15.901	16.509
		3	Wed	152	151.444	0.050	17.021	16.959
		4	Thur	158	149.005	0.807	17.693	16.686
		5	Fri	141	143.798	-0.255	15.789	16.103
		6	Sat	99	87.648	1.277	11.086	9.815
		7	Sun	70	72.045	-0.251	7.839	8.068
1	1	1	Mon	363	362.488	0.029	13.956	13.936
		2	Tue	330	352.339	-1.280	12.687	13.546
		3	Wed	370	372.244	-0.126	14.225	14.312
		4	Thur	394	392.347	0.091	15.148	15.084
		5	Fri	451	445.385	0.292	17.339	17.124
		6	Sat	412	411.683	0.017	15.840	15.828
		7	Sun	281	264.514	1.069	10.804	10.170
	2	1	Mon	764	766.078	-0.081	14.541	14.581
		2	Tue	754	736.823	0.682	14.351	14.024
		3	Wed	791	814.916	-0.911	15.055	15.510
		4	Thur	812	812.482	-0.018	15.455	15.464
		5	Fri	928	937.582	-0.345	17.663	17.845
		6	Sat	679	666.720	0.509	12.923	12.690
		7	Sun	526	519.400	0.305	10.011	9.886
	3	1	Mon	676	668.643	0.309	15.508	15.339
		2	Tue	654	662.791	-0.371	15.003	15.205
		3	Wed	696	690.133	0.243	15.967	15.832
		4	Thur	643	675.101	-1.344	14.751	15.488
		5	Fri	816	804.993	0.430	18.720	18.467

Gender	Age	Day Of Week		Frequency			Percentage	
				Observed	Predicted	Pearson Residual	Observed	Predicted
		6	Sat	538	524.540	0.627	12.342	12.034
		7	Sun	336	332.797	0.183	7.708	7.635
	4	1	Mon	669	683.067	-0.588	15.864	16.198
		2	Tue	670	657.964	0.511	15.888	15.603
		3	Wed	677	677.302	-0.013	16.054	16.061
		4	Thur	676	645.957	1.285	16.030	15.318
		5	Fri	721	710.896	0.416	17.097	16.858
		6	Sat	487	503.143	-0.767	11.548	11.931
		7	Sun	317	338.671	-1.228	7.517	8.031
	5	1	Mon	378	380.356	-0.131	14.841	14.933
		2	Tue	392	395.510	-0.192	15.391	15.528
		3	Wed	401	379.850	1.176	15.744	14.914
		4	Thur	435	425.118	0.525	17.079	16.691
		5	Fri	449	468.941	-1.019	17.629	18.412
		6	Sat	295	293.562	0.089	11.582	11.526
		7	Sun	197	203.663	-0.487	7.735	7.996
	6	1	Mon	331	320.368	0.645	15.687	15.183
		2	Tue	328	322.573	0.328	15.545	15.288
		3	Wed	360	360.556	-0.032	17.062	17.088
		4	Thur	338	346.995	-0.528	16.019	16.445
		5	Fri	356	353.202	0.163	16.872	16.739
		6	Sat	225	236.352	-0.784	10.664	11.201
		7	Sun	172	169.955	0.164	8.152	8.055

OBSERVED AND PREDICTED FREQUENCIES 2004

Gender	Age	Day Of Week		Frequency			Percentage	
				Observed	Predicted	Pearson Residual	Observed	Predicted
0	1	1	Mon	176	182.941	-0.552	13.076	13.591
		2	Tue	198	196.840	0.089	14.710	14.624
		3	Wed	209	203.198	0.442	15.527	15.096
		4	Thur	203	205.995	-0.227	15.082	15.304
		5	Fri	232	236.656	-0.333	17.236	17.582
		6	Sat	197	195.681	0.102	14.636	14.538
		7	Sun	131	124.689	0.593	9.733	9.264
	2	1	Mon	446	425.851	1.055	15.002	14.324
		2	Tue	479	475.905	0.155	16.112	16.008
		3	Wed	426	444.856	-0.969	14.329	14.963
		4	Thur	469	453.181	0.807	15.775	15.243
		5	Fri	542	532.368	0.461	18.231	17.907
		6	Sat	372	381.310	-0.511	12.513	12.826
		7	Sun	239	259.528	-1.334	8.039	8.730
	3	1	Mon	367	357.348	0.553	15.298	14.896
		2	Tue	376	372.681	0.187	15.673	15.535
		3	Wed	357	361.425	-0.253	14.881	15.066
		4	Thur	377	390.575	-0.751	15.715	16.281
		5	Fri	437	424.505	0.668	18.216	17.695
		6	Sat	302	300.912	0.067	12.589	12.543
		7	Sun	183	191.554	-0.644	7.628	7.985
	4	1	Mon	343	363.510	-1.166	14.086	14.929
		2	Tue	373	384.460	-0.637	15.318	15.789
		3	Wed	398	385.369	0.701	16.345	15.826
		4	Thur	395	399.986	-0.273	16.222	16.427
		5	Fri	423	424.397	-0.075	17.372	17.429
		6	Sat	315	298.848	0.998	12.936	12.273
		7	Sun	188	178.431	0.744	7.721	7.328
	5	1	Mon	214	213.023	0.073	15.782	15.710
		2	Tue	221	218.485	0.186	16.298	16.112
		3	Wed	202	205.068	-0.233	14.897	15.123
		4	Thur	237	230.082	0.501	17.478	16.968

Gender	Age	Day Of Week		Frequency			Percentage	
				Observed	Predicted	Pearson Residual	Observed	Predicted
		5	Fri	232	240.824	-0.627	17.109	17.760
		6	Sat	143	149.131	-0.532	10.546	10.998
		7	Sun	107	99.388	0.793	7.891	7.329
	6	1	Mon	137	140.327	-0.304	14.153	14.497
		2	Tue	161	159.630	0.119	16.632	16.491
		3	Wed	154	146.084	0.711	15.909	15.091
		4	Thur	161	162.180	-0.102	16.632	16.754
		5	Fri	157	164.251	-0.621	16.219	16.968
		6	Sat	108	111.118	-0.314	11.157	11.479
		7	Sun	90	84.410	0.637	9.298	8.720
1	1	1	Mon	295	288.059	0.439	13.391	13.076
		2	Tue	277	278.160	-0.074	12.574	12.626
		3	Wed	330	335.802	-0.344	14.980	15.243
		4	Thur	326	323.005	0.180	14.798	14.662
		5	Fri	399	394.344	0.259	18.112	17.900
		6	Sat	341	342.319	-0.078	15.479	15.539
		7	Sun	235	241.311	-0.431	10.667	10.954
	2	1	Mon	644	664.149	-0.842	13.419	13.839
		2	Tue	663	666.095	-0.129	13.815	13.880
		3	Wed	747	728.144	0.759	15.566	15.173
		4	Thur	688	703.819	-0.645	14.336	14.666
		5	Fri	869	878.632	-0.360	18.108	18.309
		6	Sat	670	660.690	0.390	13.961	13.767
		7	Sun	518	497.472	0.972	10.794	10.366
	3	1	Mon	542	551.652	-0.444	14.163	14.415
		2	Tue	513	516.319	-0.157	13.405	13.491
		3	Wed	590	585.575	0.199	15.417	15.301
		4	Thur	614	600.425	0.603	16.044	15.689
		5	Fri	681	693.495	-0.524	17.795	18.121
		6	Sat	515	516.088	-0.051	13.457	13.485
		7	Sun	372	363.446	0.472	9.720	9.497
	4	1	Mon	590	569.490	0.929	14.994	14.472
		2	Tue	552	540.540	0.531	14.028	13.737
		3	Wed	621	633.631	-0.548	15.781	16.102

Gender	Age	Day Of Week		Frequency			Percentage	
				Observed	Predicted	Pearson Residual	Observed	Predicted
		4	Thur	629	624.014	0.218	15.985	15.858
		5	Fri	705	703.603	0.058	17.916	17.881
		6	Sat	504	520.152	-0.760	12.808	13.219
		7	Sun	334	343.569	-0.540	8.488	8.731
	5	1	Mon	382	382.977	-0.054	15.219	15.258
		2	Tue	350	352.515	-0.144	13.944	14.044
		3	Wed	390	386.932	0.170	15.538	15.416
		4	Thur	405	411.918	-0.373	16.135	16.411
		5	Fri	467	458.176	0.456	18.606	18.254
		6	Sat	304	297.869	0.378	12.112	11.867
		7	Sun	212	219.612	-0.538	8.446	8.749
	6	1	Mon	276	272.673	0.217	14.212	14.041
		2	Tue	277	278.370	-0.089	14.264	14.334
		3	Wed	290	297.916	-0.498	14.933	15.341
		4	Thur	315	313.820	0.073	16.220	16.160
		5	Fri	345	337.749	0.434	17.765	17.392
		6	Sat	243	239.882	0.215	12.513	12.352
		7	Sun	196	201.590	-0.416	10.093	10.381

OBSERVED AND PREDICTED FREQUENCIES 2005

Gender	Age	Day Of Week		Frequency			Percentage	
				Observed	Predicted	Pearson Residual	Observed	Predicted
0	1	1	Mon	217	200.509	1.261	15.863	14.657
		2	Tue	191	193.186	-0.170	13.962	14.122
		3	Wed	165	183.033	-1.432	12.061	13.380
		4	Thur	219	225.001	-0.438	16.009	16.447
		5	Fri	244	250.636	-0.464	17.836	18.321
		6	Sat	209	186.397	1.781	15.278	13.625
		7	Sun	123	129.238	-0.577	8.991	9.447
	2	1	Mon	537	499.140	1.840	16.322	15.171
		2	Tue	496	489.153	0.336	15.076	14.868
		3	Wed	464	471.102	-0.353	14.103	14.319
		4	Thur	493	493.323	-0.016	14.985	14.995
		5	Fri	598	605.293	-0.328	18.176	18.398
		6	Sat	407	429.939	-1.187	12.371	13.068
		7	Sun	295	302.050	-0.426	8.967	9.181
	3	1	Mon	338	348.576	-0.612	13.955	14.392
		2	Tue	348	374.492	-1.489	14.368	15.462
		3	Wed	401	384.635	0.910	16.557	15.881
		4	Thur	423	405.073	0.976	17.465	16.725
		5	Fri	409	422.942	-0.746	16.887	17.463
		6	Sat	304	303.199	0.049	12.552	12.519
		7	Sun	199	183.083	1.224	8.216	7.559
	4	1	Mon	339	364.619	-1.456	14.043	15.104
		2	Tue	368	347.955	1.162	15.244	14.414
		3	Wed	374	381.217	-0.403	15.493	15.792
		4	Thur	425	417.310	0.414	17.606	17.287
		5	Fri	451	435.411	0.825	18.683	18.037
		6	Sat	293	283.120	0.625	12.138	11.728
		7	Sun	164	184.367	-1.561	6.794	7.637
	5	1	Mon	237	246.304	-0.647	15.440	16.046
		2	Tue	230	237.429	-0.524	14.984	15.468
		3	Wed	235	236.817	-0.128	15.309	15.428
		4	Thur	248	254.427	-0.441	16.156	16.575

Gender	Age	Day Of Week		Frequency Observed	Frequency Predicted	Frequency Pearson Residual	Percentage Observed	Percentage Predicted
		5	Fri	285	264.857	1.361	18.567	17.255
		6	Sat	172	178.450	-0.514	11.205	11.625
		7	Sun	128	116.716	1.087	8.339	7.604
	6	1	Mon	138	146.853	-0.790	13.677	14.554
		2	Tue	159	149.784	0.816	15.758	14.845
		3	Wed	191	173.196	1.486	18.930	17.165
		4	Thur	158	170.865	-1.080	15.659	16.934
		5	Fri	165	172.861	-0.657	16.353	17.132
		6	Sat	111	114.895	-0.386	11.001	11.387
		7	Sun	87	80.546	0.750	8.622	7.983
1	1	1	Mon	336	352.491	-0.950	13.907	14.590
		2	Tue	343	340.814	0.128	14.197	14.107
		3	Wed	326	307.967	1.100	13.493	12.747
		4	Thur	372	365.999	0.341	15.397	15.149
		5	Fri	467	460.364	0.344	19.329	19.055
		6	Sat	325	347.603	-1.310	13.452	14.388
		7	Sun	247	240.762	0.424	10.224	9.965
	2	1	Mon	748	785.860	-1.466	14.371	15.098
		2	Tue	766	772.847	-0.267	14.717	14.848
		3	Wed	717	709.898	0.287	13.775	13.639
		4	Thur	719	718.677	0.013	13.814	13.807
		5	Fri	1003	995.707	0.257	19.270	19.130
		6	Sat	741	718.061	0.922	14.236	13.796
		7	Sun	511	503.950	0.330	9.817	9.682
	3	1	Mon	562	551.424	0.487	14.651	14.375
		2	Tue	621	594.508	1.182	16.189	15.498
		3	Wed	566	582.365	-0.736	14.755	15.182
		4	Thur	575	592.927	-0.801	14.990	15.457
		5	Fri	713	699.058	0.583	18.587	18.224
		6	Sat	508	508.801	-0.038	13.243	13.264
		7	Sun	291	306.917	-0.947	7.586	8.001
	4	1	Mon	624	598.381	1.137	15.742	15.095
		2	Tue	553	573.045	-0.905	13.951	14.456
		3	Wed	606	598.783	0.320	15.288	15.106

Gender	Age	Day Of Week		Frequency			Percentage	
				Observed	Predicted	Pearson Residual	Observed	Predicted
		4	Thur	626	633.690	-0.333	15.792	15.986
		5	Fri	731	746.589	-0.633	18.441	18.834
		6	Sat	483	492.880	-0.476	12.185	12.434
		7	Sun	341	320.633	1.186	8.602	8.089
	5	1	Mon	442	432.696	0.488	16.376	16.032
		2	Tue	426	418.571	0.395	15.784	15.508
		3	Wed	400	398.183	0.099	14.820	14.753
		4	Thur	420	413.573	0.343	15.561	15.323
		5	Fri	466	486.143	-1.009	17.266	18.012
		6	Sat	339	332.550	0.378	12.560	12.321
		7	Sun	206	217.284	-0.798	7.632	8.051
	6	1	Mon	309	300.147	0.553	14.985	14.556
		2	Tue	298	307.216	-0.570	14.452	14.899
		3	Wed	321	338.804	-1.058	15.567	16.431
		4	Thur	336	323.135	0.779	16.295	15.671
		5	Fri	377	369.139	0.452	18.283	17.902
		6	Sat	253	249.105	0.263	12.270	12.081
		7	Sun	168	174.454	-0.511	8.147	8.460

OBSERVED AND PREDICTED FREQUENCIES 2006

Gender	Age	Day Of Week		Frequency			Percentage	
				Observed	Predicted	Pearson Residual	Observed	Predicted
0	1	1	Mon	237	227.495	0.683	15.370	14.753
		2	Tue	225	233.518	-0.605	14.591	15.144
		3	Wed	236	231.558	0.317	15.305	15.017
		4	Thur	214	231.305	-1.234	13.878	15.000
		5	Fri	282	276.253	0.382	18.288	17.915
		6	Sat	211	212.568	-0.116	13.684	13.785
		7	Sun	137	129.304	0.707	8.885	8.385
	2	1	Mon	570	564.173	0.265	14.264	14.118
		2	Tue	601	601.816	-0.036	15.040	15.060
		3	Wed	635	621.345	0.596	15.891	15.549
		4	Thur	681	677.189	0.161	17.042	16.947
		5	Fri	683	681.013	0.084	17.092	17.042
		6	Sat	501	508.918	-0.376	12.538	12.736
		7	Sun	325	341.545	-0.936	8.133	8.547
	3	1	Mon	423	415.597	0.392	14.571	14.316
		2	Tue	454	452.084	0.098	15.639	15.573
		3	Wed	479	504.076	-1.229	16.500	17.364
		4	Thur	513	494.198	0.929	17.671	17.024
		5	Fri	496	490.282	0.283	17.086	16.889
		6	Sat	318	318.791	-0.047	10.954	10.981
		7	Sun	220	227.972	-0.550	7.578	7.853
	4	1	Mon	416	423.903	-0.414	13.789	14.050
		2	Tue	483	474.087	0.446	16.009	15.714
		3	Wed	527	523.628	0.162	17.468	17.356
		4	Thur	508	510.962	-0.144	16.838	16.936
		5	Fri	511	521.916	-0.525	16.937	17.299
		6	Sat	320	320.443	-0.026	10.607	10.621
		7	Sun	252	242.062	0.666	8.353	8.023
	5	1	Mon	275	280.596	-0.362	14.635	14.933
		2	Tue	294	297.097	-0.196	15.647	15.811
		3	Wed	313	327.323	-0.871	16.658	17.420
		4	Thur	347	334.984	0.724	18.467	17.828

Gender	Age	Day Of Week		Frequency			Percentage	
				Observed	Predicted	Pearson Residual	Observed	Predicted
		5	Fri	325	308.092	1.054	17.296	16.397
		6	Sat	189	194.415	-0.410	10.059	10.347
		7	Sun	136	136.493	-0.044	7.238	7.264
	6	1	Mon	171	180.237	-0.747	14.394	15.171
		2	Tue	193	191.398	0.126	16.246	16.111
		3	Wed	221	203.071	1.382	18.603	17.094
		4	Thur	186	200.361	-1.113	15.657	16.865
		5	Fri	174	193.445	-1.528	14.646	16.283
		6	Sat	145	128.865	1.505	12.205	10.847
		7	Sun	98	90.624	0.806	8.249	7.628
1	1	1	Mon	361	370.505	-0.533	13.731	14.093
		2	Tue	385	376.482	0.474	14.644	14.320
		3	Wed	355	359.442	-0.252	13.503	13.672
		4	Thur	392	374.695	0.965	14.911	14.252
		5	Fri	493	498.747	-0.286	18.752	18.971
		6	Sat	402	400.432	0.085	15.291	15.231
		7	Sun	241	248.696	-0.513	9.167	9.460
	2	1	Mon	876	881.827	-0.211	13.431	13.521
		2	Tue	932	931.184	0.029	14.290	14.278
		3	Wed	912	925.655	-0.485	13.983	14.193
		4	Thur	1049	1052.811	-0.128	16.084	16.142
		5	Fri	1178	1179.987	-0.064	18.062	18.092
		6	Sat	928	920.082	0.282	14.229	14.107
		7	Sun	647	630.455	0.693	9.920	9.667
	3	1	Mon	624	631.403	-0.317	13.616	13.777
		2	Tue	678	679.916	-0.080	14.794	14.836
		3	Wed	755	729.924	1.012	16.474	15.927
		4	Thur	728	746.802	-0.752	15.885	16.295
		5	Fri	820	825.718	-0.220	17.892	18.017
		6	Sat	561	560.209	0.036	12.241	12.224
		7	Sun	417	409.028	0.413	9.099	8.925
	4	1	Mon	641	633.097	0.338	13.688	13.519
		2	Tue	692	700.913	-0.365	14.777	14.967
		3	Wed	742	745.372	-0.135	15.845	15.917

Gender	Age	Day Of Week		Frequency			Percentage	
				Observed	Predicted	Pearson Residual	Observed	Predicted
		4	Thur	762	759.038	0.117	16.272	16.208
		5	Fri	875	864.084	0.411	18.685	18.452
		6	Sat	554	553.557	0.020	11.830	11.821
		7	Sun	417	426.938	-0.505	8.905	9.117
	5	1	Mon	473	467.404	0.280	14.581	14.408
		2	Tue	493	489.903	0.152	15.197	15.102
		3	Wed	534	519.677	0.686	16.461	16.020
		4	Thur	543	555.016	-0.560	16.739	17.109
		5	Fri	552	568.908	-0.781	17.016	17.537
		6	Sat	380	374.585	0.298	11.714	11.547
		7	Sun	269	268.507	0.031	8.292	8.277
	6	1	Mon	377	367.763	0.521	14.984	14.617
		2	Tue	385	386.602	-0.089	15.302	15.366
		3	Wed	377	394.929	-0.983	14.984	15.697
		4	Thur	421	406.639	0.778	16.733	16.162
		5	Fri	457	437.555	1.023	18.164	17.391
		6	Sat	288	304.135	-0.987	11.447	12.088
		7	Sun	211	218.376	-0.522	8.386	8.679

OBSERVED AND PREDICTED FREQUENCIES 2007

Gender	Age	Day of Week		Frequency			Percentage	
				Observed	Predicted	Pearson Residual	Observed	Predicted
0	1	1	Mon	200	209.788	-0.732	13.986	14.670
		2	Tue	191	205.553	-1.097	13.357	14.374
		3	Wed	227	220.788	0.455	15.874	15.440
		4	Thur	233	232.537	0.033	16.294	16.261
		5	Fri	244	240.153	0.272	17.063	16.794
		6	Sat	197	192.001	0.388	13.776	13.427
		7	Sun	138	129.181	0.814	9.650	9.034
	2	1	Mon	591	589.487	0.067	14.808	14.770
		2	Tue	618	622.976	-0.217	15.485	15.610
		3	Wed	634	637.691	-0.159	15.886	15.978
		4	Thur	610	607.574	0.107	15.284	15.224
		5	Fri	712	695.251	0.699	17.840	17.420
		6	Sat	485	489.269	-0.206	12.152	12.259
		7	Sun	341	348.750	-0.434	8.544	8.738
	3	1	Mon	477	467.391	0.483	15.711	15.395
		2	Tue	480	468.998	0.552	15.810	15.448
		3	Wed	504	497.582	0.315	16.601	16.389
		4	Thur	490	479.824	0.506	16.140	15.804
		5	Fri	511	531.200	-0.965	16.831	17.497
		6	Sat	337	338.773	-0.102	11.100	11.159
		7	Sun	237	252.232	-1.002	7.806	8.308
	4	1	Mon	445	447.621	-0.134	14.740	14.827
		2	Tue	506	500.516	0.268	16.761	16.579
		3	Wed	515	507.406	0.370	17.059	16.807
		4	Thur	461	477.323	-0.814	15.270	15.811
		5	Fri	556	550.191	0.274	18.417	18.224
		6	Sat	326	318.376	0.452	10.798	10.546
		7	Sun	210	217.566	-0.532	6.956	7.207
	5	1	Mon	268	294.961	-1.705	13.872	15.267
		2	Tue	323	306.801	1.008	16.718	15.880
		3	Wed	329	326.961	0.124	17.029	16.923
		4	Thur	353	341.083	0.711	18.271	17.654

Gender	Age	Day of Week		Frequency Observed	Frequency Predicted	Frequency Pearson Residual	Percentage Observed	Percentage Predicted
		5	Fri	314	323.923	-0.604	16.253	16.766
		6	Sat	188	200.091	-0.903	9.731	10.357
		7	Sun	157	138.181	1.661	8.126	7.152
	6	1	Mon	215	186.752	2.244	17.409	15.122
		2	Tue	196	209.155	-0.998	15.870	16.936
		3	Wed	179	197.572	-1.442	14.494	15.998
		4	Thur	194	202.658	-0.665	15.709	16.410
		5	Fri	205	201.283	0.286	16.599	16.298
		6	Sat	151	145.490	0.486	12.227	11.781
		7	Sun	95	92.090	0.315	7.692	7.457
1	1	1	Mon	409	399.212	0.530	14.889	14.533
		2	Tue	382	367.447	0.816	13.906	13.376
		3	Wed	408	414.212	-0.331	14.853	15.079
		4	Thur	428	428.463	-0.024	15.581	15.597
		5	Fri	459	462.847	-0.196	16.709	16.849
		6	Sat	397	401.999	-0.270	14.452	14.634
		7	Sun	264	272.819	-0.563	9.610	9.932
	2	1	Mon	1033	1034.513	-0.051	14.638	14.659
		2	Tue	1032	1027.024	0.168	14.624	14.553
		3	Wed	1107	1103.309	0.121	15.687	15.634
		4	Thur	1030	1032.426	-0.082	14.595	14.630
		5	Fri	1219	1235.749	-0.525	17.274	17.511
		6	Sat	949	944.731	0.149	13.448	13.387
		7	Sun	687	679.250	0.313	9.735	9.625
	3	1	Mon	733	742.609	-0.383	15.107	15.305
		2	Tue	689	700.002	-0.450	14.200	14.427
		3	Wed	773	779.418	-0.251	15.932	16.064
		4	Thur	728	738.176	-0.407	15.004	15.214
		5	Fri	875	854.800	0.761	18.034	17.617
		6	Sat	594	592.227	0.078	12.242	12.206
		7	Sun	460	444.768	0.758	9.481	9.167
	4	1	Mon	719	716.379	0.106	14.831	14.777
		2	Tue	747	752.484	-0.217	15.408	15.522
		3	Wed	793	800.594	-0.294	16.357	16.514

Gender	Age	Day of Week		Frequency Observed	Frequency Predicted	Frequency Pearson Residual	Percentage Observed	Percentage Predicted
		4	Thur	756	739.677	0.652	15.594	15.257
		5	Fri	886	891.809	-0.215	18.276	18.395
		6	Sat	553	560.624	-0.342	11.407	11.564
		7	Sun	394	386.434	0.401	8.127	7.971
	5	1	Mon	544	517.039	1.288	16.019	15.225
		2	Tue	489	505.199	-0.781	14.399	14.876
		3	Wed	563	565.039	-0.094	16.578	16.638
		4	Thur	567	578.917	-0.544	16.696	17.047
		5	Fri	585	575.077	0.454	17.226	16.934
		6	Sat	398	385.909	0.654	11.720	11.364
		7	Sun	250	268.819	-1.196	7.362	7.916
	6	1	Mon	348	376.248	-1.580	13.926	15.056
		2	Tue	409	395.845	0.721	16.367	15.840
		3	Wed	411	392.428	1.021	16.447	15.703
		4	Thur	404	395.342	0.475	16.166	15.820
		5	Fri	407	410.717	-0.201	16.287	16.435
		6	Sat	317	322.510	-0.329	12.685	12.906
		7	Sun	203	205.910	-0.212	8.123	8.240

OBSERVED AND PREDICTED FREQUENCIES 2008

Gender	Age	Day Of Week		Frequency			Percentage	
				Observed	Predicted	Pearson Residual	Observed	Predicted
0	1	1	Mon	190	196.608	-0.507	13.287	13.749
		2	Tue	219	213.868	0.381	15.315	14.956
		3	Wed	203	205.636	-0.199	14.196	14.380
		4	Thur	246	224.126	1.591	17.203	15.673
		5	Fri	248	257.541	-0.657	17.343	18.010
		6	Sat	193	201.013	-0.610	13.497	14.057
		7	Sun	131	131.208	-0.019	9.161	9.175
	2	1	Mon	526	535.196	-0.428	13.473	13.709
		2	Tue	590	600.011	-0.444	15.113	15.369
		3	Wed	588	584.627	0.151	15.061	14.975
		4	Thur	567	583.812	-0.754	14.524	14.954
		5	Fri	723	716.288	0.278	18.519	18.348
		6	Sat	536	525.533	0.491	13.730	13.461
		7	Sun	374	358.533	0.857	9.580	9.184
	3	1	Mon	394	388.153	0.321	14.598	14.381
		2	Tue	439	448.617	-0.497	16.265	16.622
		3	Wed	435	422.631	0.655	16.117	15.659
		4	Thur	401	410.677	-0.519	14.857	15.216
		5	Fri	477	477.784	-0.040	17.673	17.702
		6	Sat	322	324.819	-0.167	11.930	12.035
		7	Sun	231	226.319	0.325	8.559	8.385
	4	1	Mon	391	387.964	0.166	14.203	14.092
		2	Tue	480	471.353	0.437	17.436	17.121
		3	Wed	432	435.188	-0.167	15.692	15.808
		4	Thur	436	424.596	0.602	15.837	15.423
		5	Fri	492	484.592	0.371	17.871	17.602
		6	Sat	329	330.580	-0.093	11.951	12.008
		7	Sun	193	218.726	-1.813	7.011	7.945
	5	1	Mon	293	289.372	0.232	15.462	15.270
		2	Tue	323	313.998	0.556	17.045	16.570
		3	Wed	305	314.976	-0.616	16.095	16.621
		4	Thur	305	306.804	-0.112	16.095	16.190
		5	Fri	339	337.162	0.110	17.889	17.792

Gender	Age	Day Of Week		Frequency			Percentage	
				Observed	Predicted	Pearson Residual	Observed	Predicted
		6	Sat	206	209.949	-0.289	10.871	11.079
		7	Sun	124	122.740	0.118	6.544	6.477
	6	1	Mon	170	166.708	0.274	13.969	13.698
		2	Tue	205	208.153	-0.240	16.845	17.104
		3	Wed	204	203.942	0.004	16.763	16.758
		4	Thur	210	214.986	-0.375	17.256	17.665
		5	Fri	207	212.633	-0.425	17.009	17.472
		6	Sat	130	124.106	0.558	10.682	10.198
		7	Sun	91	86.473	0.505	7.477	7.105
1	1	1	Mon	342	335.392	0.388	13.869	13.601
		2	Tue	344	349.132	-0.296	13.950	14.158
		3	Wed	329	326.364	0.157	13.341	13.235
		4	Thur	337	358.874	-1.249	13.666	14.553
		5	Fri	457	447.459	0.499	18.532	18.145
		6	Sat	393	384.987	0.445	15.937	15.612
		7	Sun	264	263.792	0.014	10.706	10.697
	2	1	Mon	904	894.804	0.331	13.712	13.572
		2	Tue	970	959.989	0.350	14.713	14.561
		3	Wed	906	909.373	-0.120	13.742	13.793
		4	Thur	933	916.188	0.599	14.151	13.896
		5	Fri	1213	1219.712	-0.213	18.398	18.500
		6	Sat	976	986.467	-0.361	14.804	14.962
		7	Sun	691	706.467	-0.616	10.481	10.715
	3	1	Mon	650	655.847	-0.247	14.174	14.301
		2	Tue	735	725.383	0.389	16.027	15.817
		3	Wed	652	664.369	-0.519	14.217	14.487
		4	Thur	661	651.323	0.409	14.413	14.202
		5	Fri	823	822.216	0.030	17.946	17.929
		6	Sat	619	616.181	0.122	13.498	13.436
		7	Sun	446	450.681	-0.232	9.725	9.827
	4	1	Mon	634	637.036	-0.130	13.965	14.032
		2	Tue	732	740.647	-0.347	16.123	16.314
		3	Wed	668	664.812	0.134	14.714	14.643
		4	Thur	643	654.404	-0.482	14.163	14.414
		5	Fri	803	810.408	-0.287	17.687	17.850

Gender	Age	Day Of Week		Frequency			Percentage	
				Observed	Predicted	Pearson Residual	Observed	Predicted
		6	Sat	611	609.420	0.069	13.458	13.423
		7	Sun	449	423.274	1.313	9.890	9.323
	5	1	Mon	501	504.628	-0.175	15.163	15.273
		2	Tue	515	524.002	-0.429	15.587	15.860
		3	Wed	521	511.024	0.480	15.769	15.467
		4	Thur	504	502.196	0.087	15.254	15.200
		5	Fri	597	598.838	-0.083	18.069	18.125
		6	Sat	415	411.051	0.208	12.561	12.441
		7	Sun	251	252.260	-0.083	7.597	7.635
	6	1	Mon	328	331.292	-0.195	13.582	13.718
		2	Tue	399	395.847	0.173	16.522	16.391
		3	Wed	377	377.058	-0.003	15.611	15.613
		4	Thur	406	401.014	0.273	16.812	16.605
		5	Fri	436	430.367	0.300	18.054	17.821
		6	Sat	271	276.894	-0.376	11.222	11.466
		7	Sun	198	202.527	-0.332	8.199	8.386

OBSERVED AND PREDICTED FREQUENCIES 2009

Gender	Age	Day of Week		Frequency			Percentage	
				Observed	Predicted	Pearson Residual	Observed	Predicted
0	1	1	Mon	176	182.941	-0.552	13.076	13.591
		2	Tue	198	196.840	0.089	14.710	14.624
		3	Wed	209	203.198	0.442	15.527	15.096
		4	Thur	203	205.995	-0.227	15.082	15.304
		5	Fri	232	236.656	-0.333	17.236	17.582
		6	Sat	197	195.681	0.102	14.636	14.538
		7	Sun	131	124.689	0.593	9.733	9.264
	2	1	Mon	446	425.851	1.055	15.002	14.324
		2	Tue	479	475.905	0.155	16.112	16.008
		3	Wed	426	444.856	-0.969	14.329	14.963
		4	Thur	469	453.181	0.807	15.775	15.243
		5	Fri	542	532.368	0.461	18.231	17.907
		6	Sat	372	381.310	-0.511	12.513	12.826
		7	Sun	239	259.528	-1.334	8.039	8.730
	3	1	Mon	367	357.348	0.553	15.298	14.896
		2	Tue	376	372.681	0.187	15.673	15.535
		3	Wed	357	361.425	-0.253	14.881	15.066
		4	Thur	377	390.575	-0.751	15.715	16.281
		5	Fri	437	424.505	0.668	18.216	17.695
		6	Sat	302	300.912	0.067	12.589	12.543
		7	Sun	183	191.554	-0.644	7.628	7.985
	4	1	Mon	343	363.510	-1.166	14.086	14.929
		2	Tue	373	384.460	-0.637	15.318	15.789
		3	Wed	398	385.369	0.701	16.345	15.826
		4	Thur	395	399.986	-0.273	16.222	16.427
		5	Fri	423	424.397	-0.075	17.372	17.429
		6	Sat	315	298.848	0.998	12.936	12.273
		7	Sun	188	178.431	0.744	7.721	7.328
	5	1	Mon	214	213.023	0.073	15.782	15.710
		2	Tue	221	218.485	0.186	16.298	16.112
		3	Wed	202	205.068	-0.233	14.897	15.123
		4	Thur	237	230.082	0.501	17.478	16.968
		5	Fri	232	240.824	-0.627	17.109	17.760
		6	Sat	143	149.131	-0.532	10.546	10.998
		7	Sun	107	99.388	0.793	7.891	7.329

Gender	Age	Day of Week		Frequency			Percentage	
				Observed	Predicted	Pearson Residual	Observed	Predicted
	6	1	Mon	137	140.327	-0.304	14.153	14.497
		2	Tue	161	159.630	0.119	16.632	16.491
		3	Wed	154	146.084	0.711	15.909	15.091
		4	Thur	161	162.180	-0.102	16.632	16.754
		5	Fri	157	164.251	-0.621	16.219	16.968
		6	Sat	108	111.118	-0.314	11.157	11.479
		7	Sun	90	84.410	0.637	9.298	8.720
1	1	1	Mon	295	288.059	0.439	13.391	13.076
		2	Tue	277	278.160	-0.074	12.574	12.626
		3	Wed	330	335.802	-0.344	14.980	15.243
		4	Thur	326	323.005	0.180	14.798	14.662
		5	Fri	399	394.344	0.259	18.112	17.900
		6	Sat	341	342.319	-0.078	15.479	15.539
		7	Sun	235	241.311	-0.431	10.667	10.954
	2	1	Mon	644	664.149	-0.842	13.419	13.839
		2	Tue	663	666.095	-0.129	13.815	13.880
		3	Wed	747	728.144	0.759	15.566	15.173
		4	Thur	688	703.819	-0.645	14.336	14.666
		5	Fri	869	878.632	-0.360	18.108	18.309
		6	Sat	670	660.690	0.390	13.961	13.767
		7	Sun	518	497.472	0.972	10.794	10.366
	3	1	Mon	542	551.652	-0.444	14.163	14.415
		2	Tue	513	516.319	-0.157	13.405	13.491
		3	Wed	590	585.575	0.199	15.417	15.301
		4	Thur	614	600.425	0.603	16.044	15.689
		5	Fri	681	693.495	-0.524	17.795	18.121
		6	Sat	515	516.088	-0.051	13.457	13.485
		7	Sun	372	363.446	0.472	9.720	9.497
	4	1	Mon	590	569.490	0.929	14.994	14.472
		2	Tue	552	540.540	0.531	14.028	13.737
		3	Wed	621	633.631	-0.548	15.781	16.102
		4	Thur	629	624.014	0.218	15.985	15.858
		5	Fri	705	703.603	0.058	17.916	17.881
		6	Sat	504	520.152	-0.760	12.808	13.219
		7	Sun	334	343.569	-0.540	8.488	8.731
	5	1	Mon	382	382.977	-0.054	15.219	15.258
		2	Tue	350	352.515	-0.144	13.944	14.044

Gender	Age	Day of Week		Frequency Observed	Frequency Predicted	Frequency Pearson Residual	Percentage Observed	Percentage Predicted
		3	Wed	390	386.932	0.170	15.538	15.416
		4	Thur	405	411.918	-0.373	16.135	16.411
		5	Fri	467	458.176	0.456	18.606	18.254
		6	Sat	304	297.869	0.378	12.112	11.867
		7	Sun	212	219.612	-0.538	8.446	8.749
	6	1	Mon	276	272.673	0.217	14.212	14.041
		2	Tue	277	278.370	-0.089	14.264	14.334
		3	Wed	290	297.916	-0.498	14.933	15.341
		4	Thur	315	313.820	0.073	16.220	16.160
		5	Fri	345	337.749	0.434	17.765	17.392
		6	Sat	243	239.882	0.215	12.513	12.352
		7	Sun	196	201.590	-0.416	10.093	10.381

2ND MODEL FITTING INFORMATION (2003-2009)

Year	Model	Model Fitting Criteria	Likelihood Ratio Tests		
		-2 Log Likelihood	Chi-Square	Degree of Freedom	Significance
2003	Intercept Only	729.007			
	Final	568.615	160.393	36.000	0.000
2004	Intercept Only	658.582			
	Final	545.319	113.263	36.000	0.000
2005	Intercept Only	686.038			
	Final	577.838	108.200	36.000	0.000
2006	Intercept Only	719.397			
	Final	567.673	151.724	36.000	0.000
2007	Intercept Only	705.309			
	Final	576.059	129.251	36.000	0.000
2008	Intercept Only	748.277			
	Final	556.620	191.658	36.000	0.000
2009	Intercept Only	658.582			
	Final	545.319	113.263	36.000	0.000

2ND MODEL LIKELIHOOD RATIO TEST (2003-2009)

Year	Effect	Model Fitting Criteria	Likelihood Ratio Tests		
		-2 Log Likelihood of Reduced Model	Chi-Square	Degree of Freedom	Significance
2003	Intercept	568.615	0.000	0.000	
	Age	702.406	133.792	30.000	0.000
	Gender	596.349	27.734	6.000	0.000
2004	Intercept	545.319	0.000	0.000	
	Age	611.004	65.685	30.000	0.000
	Gender	593.976	48.657	6.000	0.000
2005	Intercept	577.838	0.000	0.000	
	Age	669.535	91.697	30.000	0.000
	Gender	595.222	17.385	6.000	0.008
2006	Intercept	567.673	0.000	0.000	
	Age	669.701	102.028	30.000	0.000
	Gender	618.232	50.559	6.000	0.000
2007	Intercept	576.059	0.000	0.000	
	Age	678.264	102.205	30.000	0.000
	Gender	602.007	25.948	6.000	0.000
2008	Intercept	556.620	0.000	0.000	
	Age	699.177	142.557	30.000	0.000
	Gender	607.866	51.246	6.000	0.000
2009	Intercept	545.319	0.000	0.000	
	Age	611.004	65.685	30.000	0.000
	Gender	593.976	48.657	6.000	0.000