

**FORMULATION AND ANALYSIS OF PEDESTRIAN SAFETY PROBLEMS  
USING BAYESIAN NETWORK MODEL**

by

Rudolph Tetteh Bedeley

A thesis submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Master of Civil Engineering

Fall 2010

Copyright 2010 Rudolph Tetteh Bedeley  
All Rights Reserved

**FORMULATION AND ANALYSIS OF PEDESTRIAN SAFETY PROBLEMS  
USING BAYESIAN NETWORK**

by

Rudolph Tetteh Bedeley

Approved: \_\_\_\_\_  
Earl E. Lee II, Ph.D.  
Professor in charge of thesis on behalf of the Advisory Committee

Approved: \_\_\_\_\_  
Harry Shenton III, Ph.D.  
Chair of the Department of Civil & Environmental Engineering

Approved: \_\_\_\_\_  
Michael J. Chajes, Ph.D.  
Dean of the College of Engineering

Approved: \_\_\_\_\_  
Charles G. Riordan, Ph.D.  
Vice Provost for Graduate and Professional Education

## **ACKNOWLEDGMENTS**

I would like to express my profound gratitude to God Almighty for His awesome grace and constant mercy shown unto me throughout these challenging times. The contributions and encouragement from all love ones is also greatly appreciated.

Specifically, I want to sincerely thank Professor Nii Okine Atto-Okine for his unprecedented fatherly guidance, support and encouragement that have brought about the success of this work. I also owe a big gratitude to my advisor Professor Earl ‘Rusty’ Lee for being so supportive.

My appreciation also goes to the entire faculty of the University of Delaware for the wealth of knowledge they have imparted in me. And also to my parents and friends especially Yaw O. Adu-Gyamfi who have always been there for me I remain indebted for your steadfast support and faith in my ability.

## TABLE OF CONTENTS

<b>LIST OF TABLES.....</b>	<b>vi</b>
<b>LIST OF FIGURES.....</b>	<b>vii</b>
<b>ABSTRACT .....</b>	<b>ix</b>

### Chapter

<b>1</b>	<b>INTRODUCTION .....</b>	<b>1</b>
	1.1 General.....	1
	1.2 Statement of the Problem .....	3
	1.3 Research Objectives .....	5
	1.4 Research Approach.....	6
<b>2</b>	<b>BACKGROUND AND LITERATURE REVIEW .....</b>	<b>8</b>
	2.1 General.....	8
	2.2 Pedestrian Safety .....	10
	2.3 Pedestrian General Behavior .....	15
	2.3.1 Pedestrian Behavior at Signalized Intersection .....	17
	2.3.2 Pedestrian Behavior at Unsignalized Intersection .....	19
<b>3</b>	<b>DEVELOPMENT OF THE METHODOLOGY.....</b>	<b>21</b>
	3.1 Site Selection .....	21
	3.2 Data Collection .....	23
<b>4</b>	<b>BAYESIAN NETWORK.....</b>	<b>27</b>
	4.1 Introduction .....	27
	4.2 Probability Distribution .....	31
	4.2.1 Conditional Probability Distribution .....	32
	4.3 Pedestrian Crossing Behavior Model .....	35
	4.4 Model Development .....	37
	4.4.1 Graphical Level .....	38
	4.4.2 Information (Qualitative) Level.....	41
	4.4.3 Numerical (Quantitative) Level.....	42
<b>5</b>	<b>DATA ANALYSIS AND DISCUSSION OF RESULTS.....</b>	<b>49</b>
	5.1 Data Analysis.....	49
	5.2 Discussion of Results .....	56

5.3	Sensitivity Analysis .....	58
5.3.1	Season Sensitivity Analysis.....	59
5.3.2	Pedestrian Signal Length Sensitivity.....	64
5.3.3	Pedestrian Crossing Behavior Sensitivity.....	70
<b>6</b>	<b>CONCLUSION AND RECOMMENDATION.....</b>	<b>76</b>
6.1	General.....	76
6.2	Conclusion.....	76
6.3	Recommendation and Future Research .....	79
	<b>REFERENCES .....</b>	<b>80</b>
	<b>APPENDIX I.....</b>	<b>84</b>
	<b>APPENDIX II .....</b>	<b>90</b>

## LIST OF TABLES

Table 1:	Data Collection Schedule and Pedestrian Flow.....	25
Table 2:	Model's node names and their respective states.....	40
Table 3:	Observed and elicited subjective probabilities for all parent nodes. ....	43
Table 4:	Conditional probability table (CPT) for <i>Weather Condition</i> node .....	45
Table 5:	Conditional probability table for Vehicle Speed node .....	46
Table 6:	Conditional probability table for Roadway Environment node.....	46
Table 7:	Conditional probability table for Pedestrian's Motive node.....	47
Table 8:	Conditional probability table for Ped. Crossing Behavior node.....	47
Table 9:	Conditional probability table for <i>pedestrian crossing behavior</i> node .....	54

## LIST OF FIGURES

Figure 1:	Pedestrians killed between the periods of 1998-2008 in the U.S. (NHTSA, 2009).....	11
Figure 2:	Pedestrian, Pedalcyclist and Other/Unknown non-motorists death in 2008 (NHTSA, 2009).....	12
Figure 3:	Pedestrians disregarding vehicle while crossing the road. ....	14
Figure 4:	Study area spots on East Main Street & Delaware Avenue Corridors.....	22
Figure 5:	Schematic diagram of a simple Bayesian network structure indicating cause-and-effect relationship between parent node $S$ and children nodes $X_1, \dots, X_i$ .....	30
Figure 6:	Schematic structure of pedestrian crossing behavior model illustrating causal relationships between variables. Arcs are drawn from cause to effect. ....	37
Figure 7:	Bayesian network model in its graphical and information level of model development. ....	39
Figure 8:	Bar chart showing pedestrian crossing behavior at each of the six (6) study locations during spring season. ....	51
Figure 9:	Bar chart showing pedestrian crossing behavior at each of the six (6) study locations during summer season. ....	53
Figure 10:	Compiled Bayesian network model indicating final results for each of the states after the network has been calibrated with data. ....	55
Figure 11:	Bayesian network model showing the results of varying the states of the season node (summer=100%; spring=0%).....	59
Figure 12:	Bayesian network model showing the results of varying the states of the season node (summer=100%; spring=0%).....	61

Figure 13:	Sensitivity plots for Summer and Spring Season. ....	62
Figure 14:	Bayesian network model showing the results of varying the states of pedestrian signal cycle node (> 2mins = 100%; < 2mins = 0%).....	65
Figure 15:	Bayesian network model showing the results of varying the states of the peds. signal cycle node (< 2mins = 100%; > 2mins = 0%).....	66
Figure 16:	Sensitivity plots for pedestrian signals. ....	68
Figure 17:	Bayesian network model showing the results of varying the states of the pedestrian crossing behavior node (rational = 100%).....	70
Figure 18:	Bayesian network model showing the results of varying the states of pedestrian crossing behavior node (irrational = 100%). ....	72
Figure 19:	Sensitivity plots for pedestrian crossing behavior.....	73



## **ABSTRACT**

Causes of pedestrian road accident have been a major concern to transportation engineers and other road safety professionals despite all efforts being applied to alleviate this problem. Although studies have aimed at modeling and analyzing the causes of pedestrian road accidents, the bulk of these studies have been found to be too stochastically oriented and more macroscopic than it is necessary. Consequently, the existing models seldom incorporate the interactions between pedestrians and their immediate environment. In this study, pedestrian crossing behavior during spring and summer season has been thoroughly investigated using Bayesian network modeling technique. The model was constructed with variables known to influence pedestrian crossing behavior either directly or indirectly. Stages of the model building process including Graphical Level (GL), Information or Qualitative Level (IL) and Quantitative Level (QL) have been discussed and implemented to extract useful information from both observed data and data elicited from stakeholders' opinion as well as experts' experience. The robustness of the Bayesian network model is compared based on its ability to produce physically meaningful results that truly reflects realistic behavior of a system. The model's results show that pedestrians often exhibit rational crossing behavior than they do irrationally and such an attitude is found to be influenced mostly by their own motives and less by external factors even though roadway environment did not favored them. Also, a sensitivity analysis carried out revealed that signal timing phase length is the most influential parameter that affects pedestrian crossing behavior.

**Keywords:** Bayesian network model; Graphical Level; Information/Qualitative Level; Quantitative Level Stochastic, Macroscopic.

# Chapter 1

## INTRODUCTION

### 1.1 General

Walking is a basic form of locomotion which human beings undertake daily to maintain social dynamism. People will naturally choose to walk everywhere they can for various fundamental objectives depending on whether the environment is safe, comfortable and inviting. Among some of the numerous reasons why people voluntarily decide to walk are; for exercise, to run errand, to go to work/school, or sometimes exploring around on foot just to have fun by meeting people and seeing new places. Everyone moves in their own way, with their unique mental and physical distinctive peculiarities and motivations. However, human behavior is based on a number of shared decision-making parameters as well as spatial accessibility restrictions imposed by the environment where it occurs.

From a safety point of view, most pedestrian injuries tend to occur when people are crossing traffic lanes (de Lavalette et al., 2009). However, road safety regulations impose few restrictions on pedestrians. Unlike drivers, who have to keep to a certain route, pedestrians are free to move around more. According to Jian et al. (2005), pedestrians are more flexible and more intelligent than vehicles: they can set their course within the urban space as they think best, without or rarely being subject to the types of police checks that other users are. It turns out that pedestrians, as users, are not really controlled by regulations, and their behavior is often unpredictable.

Crossing behavior models concern pedestrians' decision making as regards the time and/or location of road crossings. These appear to be largely governed by either the gap acceptance theory, according to which each pedestrian has a critical gap to cross the road, or utility theory, according to which the utility of each alternative is a latent concept which is modeled as a random variable depending on the attributes of the alternative and the characteristics of the decision-maker. An important number of studies have been published, examining different aspects of road crossing at various locations and in different conditions.

The system of rules of roadways is about all of the precepts relating to user behavior. For pedestrians, these rules are not very extensive compared with, for example, the set of rules that govern driver behavior. They instruct pedestrians to walk on the pavement, to cross roads in designated places (pedestrian crossings), and also to comply with traffic signals. The system of rules for pedestrians thus involves how to use a certain location, a pedestrian crossing, involves also a length of time, the period that the pedestrian phase of the traffic signals allows them to cross, and is made up of only a few simple rules. However, we frequently observe that these rules are often violated. This is due to significant differences arising between behaviors anticipated by those responsible for road safety and actual behaviors observed daily on the roads. This disparity between prescribed and actual behavior is not simply a question of knowledge-base, because we can hypothesize that this disrespect for rules would not be observed when pedestrians' primary task was to comply with rules whilst street crossing as a secondary task.

In order to explain this lack of respect for what is relatively a minor number of regulations affecting pedestrians, one needs to view the crossing task as a

secondary task which is subordinate to the often more demanding primary task, the journey itself (Tijus, 2001). Moreover, one can consider that the journey, just like any task, is accomplished by seeking to reduce its level of difficulty and complexity, whilst at the same time maintaining a satisfactory level of safety. Therefore it is not a question of either obeying or disobeying regulations, but rather of weighing up the demands of the task against the gains made in the activity in question. In my view, investigating the way the environment in which the primary and secondary tasks take place affects the pedestrian's decision to cross, should reveal one of the factors that may help explain the lack of adherence to road signals.

## **1.2 Statement of the Problem**

Previous research on pedestrians' movement in urban environment is extensive and ranges from pedestrian flow modeling to individual pedestrians' behavior. In order to model pedestrian movement, it is necessary to consider the activity agenda of pedestrians and incorporate the interactions between pedestrians and their environment (roadway, traffic and crowd). A complicated decision making process is involved, in which pedestrians perceive and assess their environment, decide their strategy and adapt it accordingly if necessary.

However, pedestrians' behavior may not always be based on a simple stimulus-response process, but may also be strongly related to human factors. Moreover, in contrast to vehicles flows, which are distributed along fixed corridors of the road environment and are subject to specific traffic rules, pedestrian flows are characterized by a significant degree of randomness, so that one could consider that

each individual's trip is unique. Consequently, pedestrians' behavior may be far more flexible and adaptable than motorists.

Existing research on pedestrian movement and behavior models focuses on two separate aspects of pedestrian behavior: route choice and crossing behavior. Route or itinerary choice models concern pedestrians' decision making process as regards the optimal path between an origin and a (fixed or not) destination, among a number of alternatives, under some constraints. This area has attracted much attention from researchers and catapulted quite a number of research studies worldwide. Although the state of the art offers convenient numerical analysis for pedestrian behavioral evaluation and prediction, they suffer from a few limitations enumerated below:

- Inability to capture pedestrian crossing decision based on prevailing conditions faced by pedestrian. For instance, if the person is late to catch a bus; and no vehicle is in the vicinity and traffic signal turns red on pedestrian, what are the chances that s/he will cross the intersection on red. This calls for a methodology that is capable of showing how pedestrian crossing behavior can be influence by both human and prevailing environmental conditions.

- The need to properly characterize our degree of belief about pedestrian crossing behavior. Such a characterization could relieve a whole bunch of doubt and uncertainties faced by researchers when it comes to predicting human behavior.

- The need to extract useful and valuable information for traffic safety innovations and operations.

Valuable information mentioned above is lacking in the researches carried out so far and this consequently influence pedestrian safety improvement goals and

objectives. An appropriate pedestrian crossing behavior modeling tool should provide the quality of information needed to aid pedestrian safety improvement policy. The Bayesian Network model is capable of providing further insight into the random nature of pedestrian behavioral pattern. The procedure is able to identify causal relationships of variable, their interaction and the influence on a given target of interest.

### **1.3 Research Objectives**

Outline in the following are the basic objective of this research;

- To model pedestrian crossing attitude using Bayesian Networks based on human and environmental/physical situations that influence their crossing decision. This will enable us to identify most influential and uncertain parameters that impact the average pedestrian crossing decision.
- To be able to ascertain which of the two seasons (summer and spring) do pedestrian often demonstrate more eccentric crossing behavior by willfully violating traffic laws. Thus, we want to be able to predict with some degree of certainty how pedestrian crossing behavior changes with changes in the season.
- To design a classification process that identifies pedestrians into certain categories based on certain behavioral traits that they exhibit.

## **1.4 Research Approach**

Exhaustive review of current trend in pedestrian safety research is carried out in order to be abreast with state of the art approach and also justify the rationale for improving upon these methods. The various pedestrian crossing behavior models will be thoroughly reviewed and information extracted will be combined as an input to Bayesian networks to help model and explicitly analyze pedestrian crossing behavior pattern. The robustness of the Bayesian Believe functions methodologies is tested through its ability to use conditional probability theory to accurately formulate, delineate and predict the degree of believe of an event happening contingent on information from past experience. Basically, the development of this thesis involves the following outline:

1. A general introduction to pedestrian safety problems facing traffic engineers of our days; emphasizing on the study objectives and the discrepancies created by classical safety analysis methods.

2. Step two will comprise of an overview of pedestrian behavioral analysis and how it has taken a center stage over time. A comprehensive review of the state of the art (including research and analysis methods employed) is also presented and used as background information for introducing the prowess of the Bayesian network as an alternative approach for behavioral analysis.

3. The third step subsumes a thorough review of the methodologies employed by various researchers into pedestrian safety and behavioral analysis, clearly expounding the underlying assumptions and possible challenges of those assumptions used.

4. Task for step four of this study encompasses the analysis and interpretation of our results.

5. The last task for this study covers the main conclusions based on entire study done so far and directions into the future of this study.



## Chapter 2

### BACKGROUND AND LITERATURE REVIEW

#### 2.1 General

Pedestrians face a variety of challenges when they walk along and across streets with motor vehicles. In an attempt to minimize these challenges, policy makers and community leaders are asking for help to slow traffic down, make it safer to cross the street, and make the street more inviting to pedestrians. People of almost all ages, both sexes and in all walks of life, walk, set against a background of steadily increasing vehicle numbers and traffic levels. Most individual trips, whatever the primary mode used, begin and/or finish with a walk section, so that walking is a fundamental component of all travel (AUSTROADS, 1996). All road users are pedestrians in some point in time.

Inevitably, walking almost involves crossing a road, where the desire line of the pedestrian conflicts with the higher speed and lesser vulnerability of motor vehicles. Where speeds and/or flows are high, this can result in either delay or risk for the pedestrian. Recent decades have seen an overall fall in pedestrian injuries due to several factors, including: (i) effective speed management, (ii) reduction in walking, (iii) better vehicle design, and (iv) traffic rerouting among others (Martin, 2006).

However, pedestrian/motor vehicle crashes are still a serious problem throughout the world and the United States has a particular problem with pedestrian

deaths and injuries. Each year the US records more than 8,700 pedestrian deaths and another 70,000 disabling injuries as a result of traffic collisions (Rank, 1989).

According to 2008 study carried out by the National Highway Traffic Safety Administration (NHTSA) in conjunction with the Federal Highway Administration (FHWA), the following statistics reveals deaths recorded in the U.S that is attributed to the crossing behavior of pedestrians. The study shows that about 842 (19.2%) of the pedestrian deaths occurred as a results of pedestrian walking, playing, working, etc, in the roadway; 831 (19%) pedestrians killed as a result of improper crossing of the roadway or intersection; 741 (16.9%) pedestrian died as a result of failure to yield right of way; 584 (13.3%) pedestrian deaths recorded as a result of being under the influence of alcohol, drugs or medication; 480 (11%) pedestrians killed due to darting or running into the road; 479 (10.9%) pedestrians died as a result of poor visibility. Moreover, 87 (2.0%) pedestrian deaths was also recorded as a result of pedestrian not being attentive (probably talking, eating, etc); 74 (1.7%) pedestrians killed as a result of failure to obey traffic signs, signals, or officer; 33 (0.8%) pedestrian died due to physical impairment; 20 (0.5%) pedestrian deaths recorded as a result of emotional problems (e.g. depression, anger, disturbed); 15 (0.3%) pedestrians killed as a result of pedestrian getting on/off/in/out of transport vehicle among others.

Numerous researches have been carried out into the behavior and movement of pedestrians at junctions and/or at other crossing locations. An important part of these researches concerns the evaluation of roadway design, traffic control features and road safety treatments by means of before-and-after studies on pedestrians observed behavior and safety.

## **2.2 Pedestrian Safety**

More than half of all pedestrian deaths and disabling injuries happen when entering or crossing the street. Pedestrian traffic collisions do not vary regionally by accident type; they are nearly the same east to west, north to south. Pedestrian safety is a major area of traffic safety. Although specific traffic safety programs, such as mandatory usage of safety belts, tougher drinking-and-driving measures, enhanced roadway delineation, and elimination of roadside hazards, are reducing overall traffic injuries and fatalities, the numbers of pedestrian/traffic collisions, injuries, and deaths have remained a major safety concern.

Since 1970, the National Highway Traffic Safety Administration had carried out a series of research projects that identified pedestrian accident types by category. New pedestrian safety countermeasures were developed to address specific pedestrian accident types and were field tested. To directly address the primary cause of pedestrian/traffic collisions—the behavior and actions of pedestrians themselves—safety education messages spelling out how to be a safe pedestrian were developed and refined through field testing. Basic pedestrian safety messages, such as Stop at the curb; Look left, right, and left again before crossing; Look over your shoulder at intersections for turning vehicles; Move out to where you can see; and Wait until new green signal (or walk signal) were incorporated into the traffic safety educational programs of several major U.S. cities. Significant reductions in pedestrian/traffic collisions were noted in these cities, especially for children. The following graph shows a significant reduction in pedestrian killed between the periods of 1998-2008 according to the NHTSA Fatality Analysis Report (FARs) published in 2009.

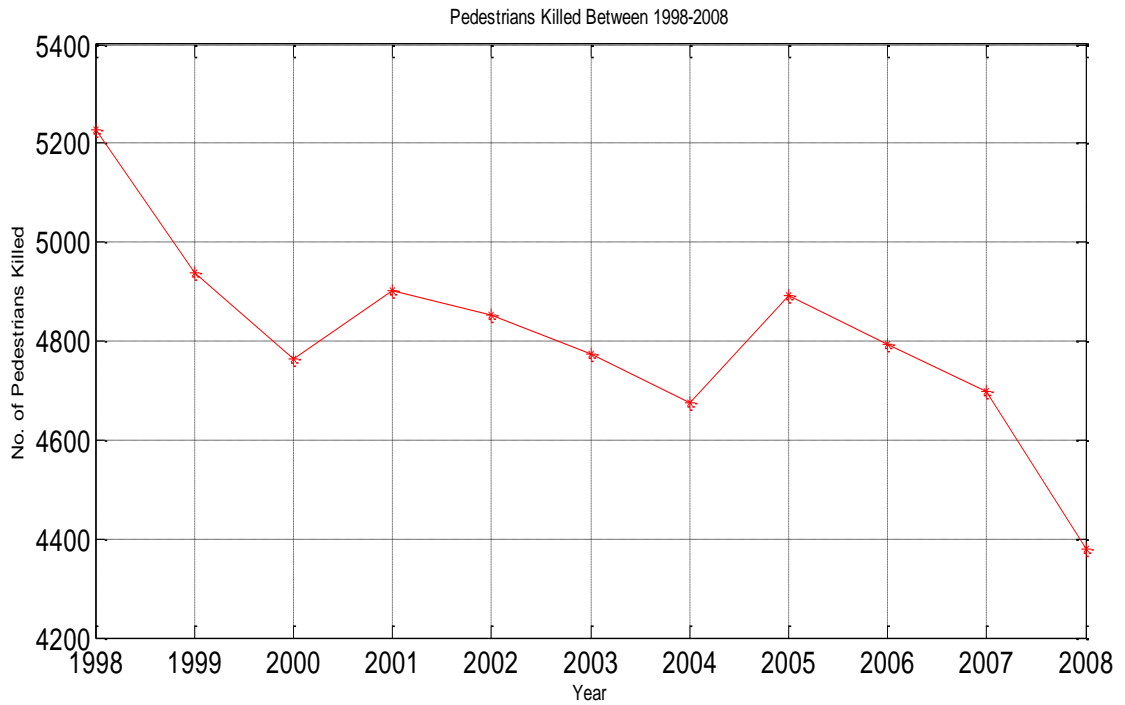


Figure 1: Pedestrians killed between the periods of 1998-2008 in the U.S. (NHTSA, 2009).

Even though there has been a significant decline in pedestrian deaths over the last decades, pedestrians still face the highest risk when compared to pedalcyclist and other/unknown non motorist traffic risk. The graph below shows pedestrians killed as compared to pedalcyclist and other/unknown non motorist killed in 2008 according to NHTSA Fatality Analysis Report published in 2009.

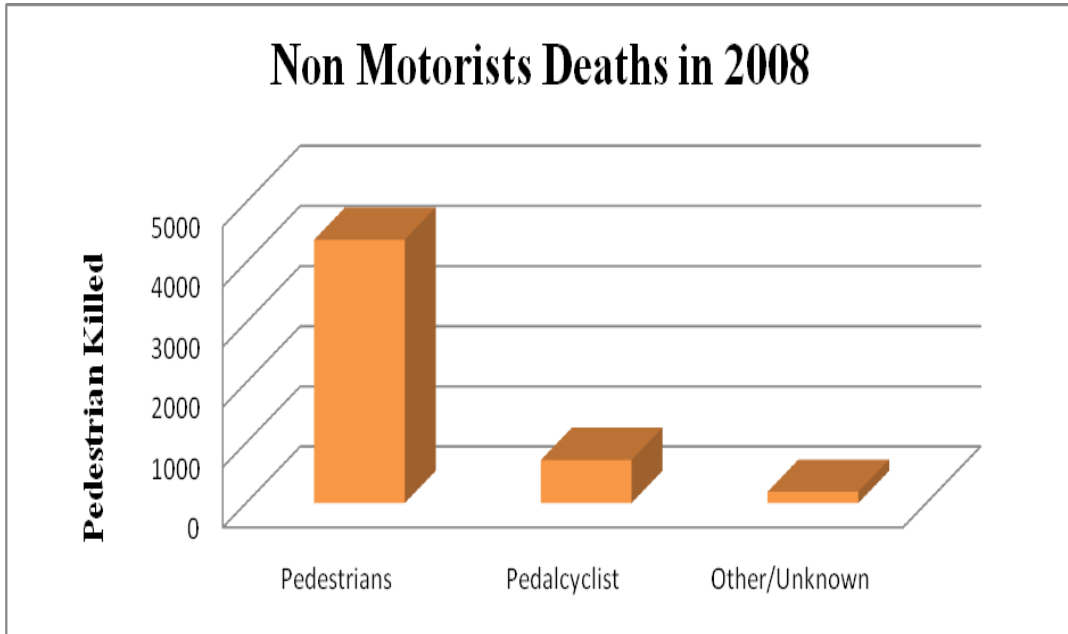


Figure 2: Pedestrian, Pedalcyclist and Other/Unknown non-motorists death in 2008 (NHTSA, 2009).

In 2008, about 73,378 pedestrians were involved in a crash out of which 4,378 deaths and 69,000 injuries were recorded. Of the crashes recorded, however, about 1,050 pedestrians representing 24% were killed and 33,000 pedestrian injuries representing 47.6% were recorded at traffic intersections according to Traffic Safety Facts published in 2009 by NHTSA. Nearly one in every five traffic fatalities involves a pedestrian. More people are killed each year in pedestrian/traffic collisions than by drowning or by fires or by poisoning. The annual cost of pedestrian/traffic collisions to society exceeds one billion dollars (Rank 1989).

A number of pedestrian mobility and safety treatments have been evaluated with positive results, including the implementation of measures prompting motorists to yield for pedestrians (Koenig & Wu, 1994; Nasar, 2003), the construction of speed humps downstream uncontrolled pedestrian crosswalks (Dixon, Alvarez, Rodriguez, & Jacko, 1997), the implementation of fluorescent strong yellow–green pedestrian warning signs at mid-block locations (Clark, Hummer, & Dutt, 1996), the construction of a refuge island (Nee & Hallenberg, 2003), the use of waiting countdown timers at traffic controlled junctions (Keegan & O’Mahony, 2003), the implementation of systems for detecting pedestrians near the crosswalk zone and for warning drivers (Hakkert, Gitelman, & Ben-Shabat, 2002) or providing an earlier activation or an extension of the pedestrian stage (Carsten, Sherborne, & Rothengatter, 1998).

Moreover, pedestrian road crossing is often incorporated in multi-modal level-of-service analyses (Winters, Cleland, Mierzewski, & Tucker, 2001). In several related researches, measures of effectiveness for crossing at junctions were proposed, difficulty to cross was proposed as a measure of effectiveness for mid-block locations (Baltes & Chu, 2002) and pedestrians’ road crossing options were seen as measures of accessibility to transit.

Although these results are very useful from a traffic engineering and policy viewpoint, they incorporate behavioral elements in a macroscopic way only. Several authors argue that, despite the improvements of the road and traffic features creating a safer environment, the unsafe behavior of pedestrians is less affected (Hakkert et al., 2002; Nee & Hallenberg, 2003). It is therefore necessary to further analyze the behavior of pedestrians itself, in order to better integrate it into the traffic

features evaluations. Within this context, an important number of researches deal with the behavior of pedestrians in terms of crossing decisions and the related determinants.

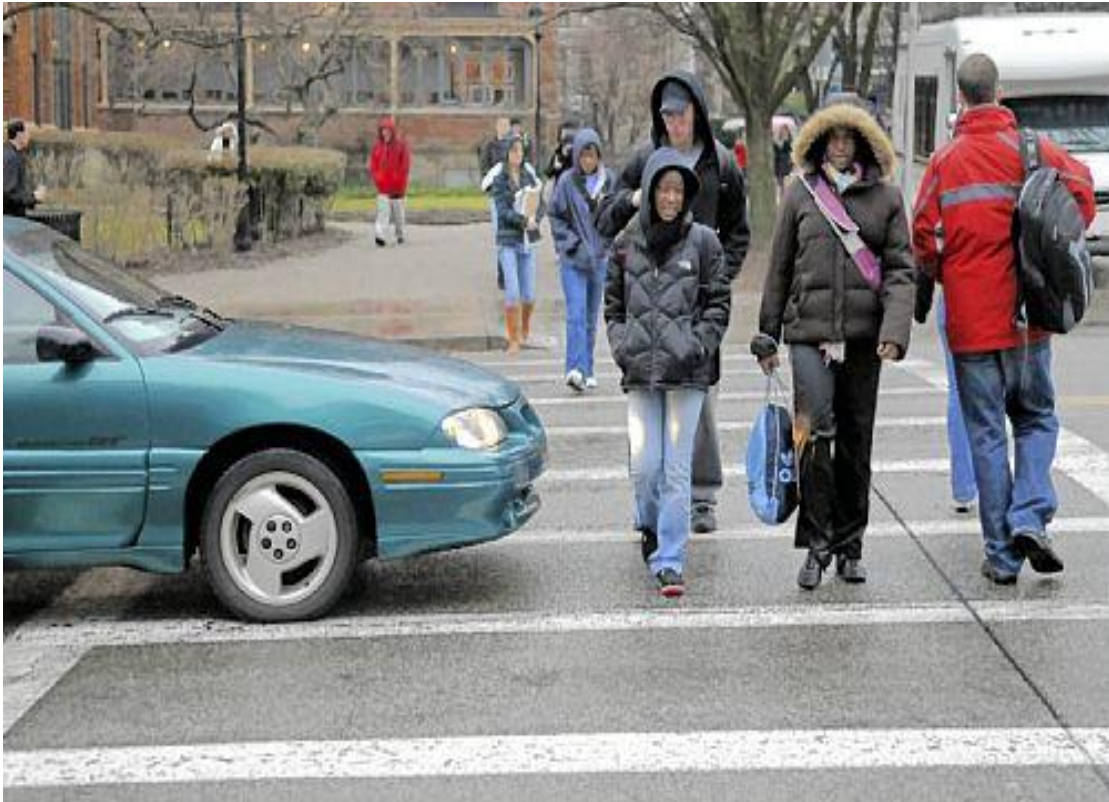


Figure 3: Pedestrians disregarding vehicle while crossing the road.

Planning and creating a safer roadway environment for pedestrians has been a combined responsibility of both traffic engineers and planners. The design and construction of pedestrian facilities on existing and new streets and highways is

necessary to ensure the safety of pedestrians. An engineering or physical facility change to the roadway environment is often the most appropriate solution to a pedestrian safety hazard. Engineering measures, such as sidewalks, one-way streets, lighting, physical barriers, grade separations, traffic signals, signs, and pedestrian signals'; marking of crosswalks; school zone improvements; safety islands; parking design; and pedestrian malls have been demonstrated to improve pedestrian safety.' Recently, significant new engineering information has been developed for the application of three of these measures: sidewalks, the marking of crosswalks, and the use of standard-timed pedestrian signals.

### **2.3 Pedestrian General Behavior**

Pedestrian crossing behavior is determined partly by the task to be accomplished, and partly by the physical environment in which the crossing takes place. The task to be accomplished is to cross the road to get somewhere, whilst avoiding the paths of vehicles moving within a physical environment that possesses a particular configuration of road features: the color of the traffic lights, the presence of vehicles, the actions of other pedestrians, etc. These variables can be considered as a set of constraints that determine the behavioral outcome: deciding whether to cross the road or to wait according to the Constraints model built-up by Richard et al. (1993) for problem-solving investigations.

The Constraint Model was developed for problems that involve changing the state of a situation until the desired end state is obtained. This is the case when a person moves from location x to location y. By considering constraints as rules which limit possible actions, the model formalizes the idea that problem-solving is a



compromise between constraints which can prove to be contradictory: one can achieve task objectives by using knowledge one has about the situation and by using one's interpretations of objects and their properties and adaptation heuristics arising within daily life. These objectives, knowledge, interpretations and heuristics are all constraints that add up. This build up of restrictions can lead to impasses, that is to say situations in which nothing can be done, given the constraints; in order to be able to act, certain constraints have to be sacrificed. "Cross at once a street to take the bus which gets ready to leave" a constraint is for example which can, for the current situation; show itself contradicted, for the same current situation, by force "not to cross when the light is green". The constraints having a weight and being able to organize itself in a list, the resolution of the impasse takes place by deleting the constraint that has the weakest weight.

The Constraint Model can therefore be applied particularly successfully to the task of a pedestrian crossing a road. Studying a pedestrian's behavior in terms of variables, which can act as constraints, would enable us to better understand their decision-making, to make environmental planning suggestions to facilitate natural behavior which is not dangerous and which involves fewer rule violations, and to obtain an improved match between the individual and the system.

It is noted that additional observational studies of pedestrians behavior in specific urban areas are available e.g. by national or local transport planning or road safety authorities. In this review, only studies that explicitly deal with road crossing behavior and include some statistical analysis or modelling are examined, while priority is given to scientific publications and other research reports. A useful review

of (mostly earlier) studies on pedestrians crossing behavior in terms of gap or delay acceptance can be found in Ishaque and Noland (2007).

### ***2.3.1 Pedestrian Behavior at Signalized Intersection***

An important number of studies have provided insight into several aspects of pedestrians crossing behavior and have also contributed in the quantification of the related determinants, although mainly focusing on a particular set of determinants in each case. Most importantly, crossing behavior is examined at or around specific locations, e.g. junctions or other crossing facilities or other uncontrolled locations.

Pedestrian moving along a road segment is faced with a number of crossing alternatives, from which he or she shall select crossing locations. This selection is affected by characteristics of the trip (e.g. the origin and destination, the complexity and the length of the route), characteristics of the infrastructure (e.g. pedestrian facilities, road geometry and traffic conditions), as well as individual characteristics (e.g. age and gender, risk proneness, delay acceptance etc.). The selected crossing options shall therefore reflect the combined assessment of the above features under specific conditions. Certainly, one should allow a degree of randomness in crossing behavior. However, it would not be reasonable to assume that pedestrians crossing decisions are independent of the number and type of crossing alternatives along the trip and the other internal and external factors.

A study into illegal pedestrian crossing at signalized intersection: incidence and relative risk was carried out by Mark J. King, David Soole, and Ameneh Ghafourian (2009). The results from the study provided evidence that illegal crossing behaviors are associated with an increased crash risk. The authors further

recommended based on other implication for the promotion of walking for health and environmental reasons that the emphasis on the provision of better pedestrian facilities should be supplemented by drawing attention to the fact that crossing a road is one part of walking which is formally regulated, and that this is because of the risk involved.

Oxley, Fildes, Ihssen, Charlton, and Days (2005) investigated age differences in the ability to choose safe time gaps in traffic as well as some of the factors involved in such judgments in a simulated road-crossing task. Participant decision times were compared by means of ANOVA. A logistic regression model was then developed for gap selection, in relation to walking time, age group, time (or distance) gap and vehicle speed. Results showed that a large proportion of the elderly pedestrians opted for unsafe traffic gaps, given their walking times. Moreover, all participants crossing decisions appeared to be based primarily on the distance of oncoming vehicles and to a lesser extent on time of arrival.

Another study of pedestrian crashes at crossing facilities in New South Wales and Victoria (Austroads, 2000a) found that illegal pedestrian movements featured in 32–44% of pedestrian crashes at signalised intersections and 45% at pedestrian operated signals (i.e. not at a signalised intersection). In a more recent study, violation of traffic laws by the victim was found to be one of the “predominant contributing factors” in all pedestrian categories examined.

Being cognizance of pedestrian rules does not seem to be the issue; rather, pedestrians want to cross where it is convenient for them, and with as little delay as possible. Enforcement of the rules by police is infrequent, and considered by the public to be unwarranted. Engineering measures also tend to be ignored, with measures such

as overpasses and underpasses and pedestrian barriers having little effect on illegal crossing behavior.

### ***2.3.2 Pedestrian Behavior at Unsignalized Intersection***

To be able to model and understand pedestrian movement, it is necessary to consider the activity agenda of pedestrians and incorporate the interactions between pedestrians and their environment (roadway, traffic and crowd). A complicated decision making process is involved, in which pedestrians perceive and assess their environment, decide their strategy and adapt it accordingly if necessary (Papadimitriou et al., 2009). However, pedestrians' behavior may not always be based on a simple stimulus-response process, but may also be strongly related to human factors.

In contrast to vehicles flows, which are distributed along fixed corridors of the road environment and are subject to specific traffic rules, pedestrian flows are characterized by a significant degree of randomness, so that one could consider that each individual's trip is unique. Consequently, pedestrians' behavior may be far more flexible and adaptable than motorists'.

Oxley, Fildes, Ihsen, Charlton, and Days (1997) examined the crossing behavior of elderly pedestrians at mid-block locations by measuring a number of indicators such as kerb delay, gap acceptance, crossing time, time-of-arrival, minimum safety margin and crossing style (non-interactive vs. interactive). Measurements for elderly pedestrians were compared to those of younger ones by means of *t-tests*. Results showed that elderly pedestrians present increased kerb delay, and accept larger gaps; however they also frequently adopt unsafe interactive crossing styles. A related

study (Bernhoft & Carstensen, 2008) revealed that older pedestrians appreciate sidewalks and crossing facilities much more than younger pedestrians.

Rosenbloom, Ben-Eliyahu, and Nemrodov (2008) used a similar method to examine the crossing behavior of children and found that not looking was the most prevalent unsafe behavior, followed by the combination of not looking and not stopping, and not stopping before crossing. They also found that children accompanied by an adult committed more unsafe behaviors, especially when not holding hands with the adult.

Another relevant research (Hamed, 2001) deals with modeling pedestrian crossing behavior at mid-block locations on divided and undivided roads, using data collected in the city of Amman, Jordan. First, pedestrian kerb waiting time is modeled as a survival model i.e. a risk function giving the instantaneous failure rate (ceasing the waiting time) assuming the pedestrian has not successfully crossed at a given time point. A time-dependent baseline risk is compared to an exogenous variables generated risk. Additionally, the number of crossing attempts was modeled under Poisson and Negative Binomial assumptions in relation to waiting time. Explanatory variables included gender, age, crossing frequency, number of people in the group, access to private vehicle, destination, home location and previous accident involvement; surprisingly, traffic parameters were not found to be statistically significant. The model's results suggest that pedestrians waiting times and number of crossing attempts are strongly related. Moreover, it was shown that, in divided roads, pedestrians behave differently from one side to the median, than from median to the other side of the road.

## Chapter 3

### DEVELOPMENT OF THE METHODOLOGY

#### 3.1 Site Selection

This research was conducted at six (6) different intersections in Newark, a small city in Delaware which is known for its college town (university of Delaware) charm as well as other industrial and commercial diversity. The city's good transportation infrastructure system as well as its natural geographical location in the Mid Atlantic region of the country has attracted a huge number of vehicles on its major roadways. Tons of vehicles floods the city's Main Street and the Delaware Avenue, which are the two main corridors within the city that traverse to the regional main highway (I-95). Trips originated from New York or Pennsylvania via the city to Maryland or DC use East Main Street corridor whiles opposite trips goes through the town on Delaware Avenue corridor.

East Main Street is a one-way, two-lane street with vehicle commuting in a uni-directional flow (i.e. east-west bound). Being the main street that traverses through the city's downtown, there are several attractions along this stretch of roadway such as shops, restaurant, bars, good sidewalks etc that appeals to pedestrians. On the other hand, the Delaware Avenue corridor is also one-way, two lane roadway that traverse through University of Delaware's main campus. Traffic flow on Delaware Avenue roadway is also unidirectional (i.e. west-east bound).

For the purpose of this study, we selected a total of six intersections, three on each of the two corridors, which are located within the downtown area of the city. As depicted on the map in figure 4, the three important intersections on East Main Street as far as this study is concerned are identified with numbers 1, 2 & 3 as East Main Street & South College Ave, East Main Street & Academy Street, and East Main Street & South Chapel Street respectively. Also on the Delaware Avenue Corridor, the intersections selected for this study are identified with numbers 4, 5 & 6 as Delaware Avenue & South College Avenue, Delaware Avenue & Academy Street, and Delaware Avenue & South Chapel Street respectively.

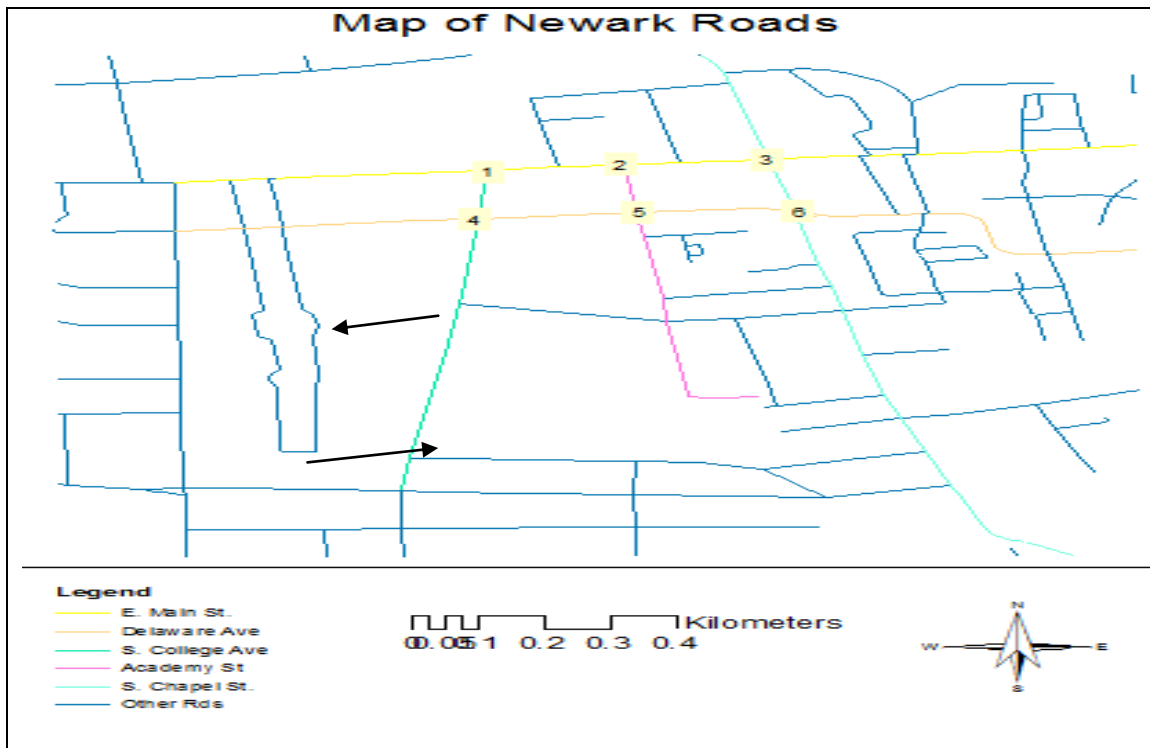


Figure 4: Study area spots on East Main Street & Delaware Avenue Corridors.

On the East Main Street corridor, these intersections were selected basically because they are very close to activity areas such as restaurants, bars, shopping centers, banks among others where pedestrian often frequent for one reason or the other thereby making it inevitable for them to cross the road at one point in time. However, the reason for selecting the intersections on Delaware Avenue corridor is mainly because that corridor is often flooded by students as well as both teaching and non teaching faculty members who either study or work at the University, thereby making them cross the road to go to lectures, offices, dormitories and lunch.

Intersections 1 & 2 on East Main Street are T-Junction intersections with two pedestrian crosswalks on each intersection, one across the Main Street and the other across the streets joining the Main Street (Academy St., S. College Ave.) respectively. In addition, the two intersections have both pedestrian and traffic signals simultaneously directing the movement of both pedestrians and vehicle traffic respectively. The remaining four intersections such as 3, 4, 5 & 6 are all four-way intersections with four pedestrian cross walks.

### **3.2 Data Collection**

As mentioned earlier, this research was undertaken in the city of Newark, Delaware, during the first two months of spring season and the last two months of summer season, 2010. The city is relatively a small area with population estimate of about 30,000 people according to the U.S. Census Bureau 2006 data. The city is home to the State's premier and largest university (University of Delaware), which has about 20,000 students population, 4,000 faculty and staff members. Based on this huge number which constitute about more than two-thirds of the city's entire population, the



study locations were selected such that they are in close proximity to the university campus. There are a significant number of individuals in the city that patronize non-motorized means of transportation to commute to work or school. As student safety is major concern, many of the roadways close to or within the university campus are specially designed in such a way that they are pedestrian-friendly to ensure maximum safety of pedestrians who constitute the largest number of road users in the city.

Field observation was one of the methods employed by this study where observers collected pedestrian data with tally sheets as pedestrians cross an intersection. Data for the spring season were taken during the times the rain was about to fall or at the time it was actually falling. There was no specific time designated for counts during the spring season data collection since the rain didn't follow irregular pattern. However, data for the first two months of the summer season were taken continuously for two hours between 11:00 am and 1:00 pm. Data on how pedestrians crossed the roadway based on whether they violated the legal traffic regulations as well as conditions about the prevailing traffic situations were simultaneously recorded by observers.

Also, data from an interview of a couple dozen college students and stakeholders such as working adults were collected to bolster the field observed data. Questions about whether their crossing behavior would change while about to make a crossing decision given unforeseen circumstances such as an imminent rain, heavy traffic, long wait time etc were spontaneously answered in the course of the interview. Experts' opinion in the form of findings from previous research by seasoned and distinguished scholars was also helpful.

The study area intersections had different pedestrian flow rate, with values varying between 11 and 37 pedestrian crossings per hour for spring season and 70 and 464 pedestrian crossings per hour for summer season, as shown in Table 1.

Table 1: Data Collection Schedule and Pedestrian Flow

<i>Spring</i>				
<b>Intersection</b>	<b>Date</b>	<b>Volume (ped)</b>	<b>Period (hours)</b>	<b>Flow (ped/hour)</b>
E. Main & S. College Ave.	3/19/10	57	2	29
E. Main & Academy St	3/26/10	22	2	11
E. Main & S. Chapel St	4/2/10	48	2	22
Del. Ave. & S. College Ave.	4/14/10	74	2	37
Del Ave. & Academy St	4/20/10	66	2	33
Del Ave. & S. Chapel St.	4/26/10	35	2	18
<i>Summer</i>				
E. Main & S. College Ave.	8/31/10	246	2	123
E. Main & Academy St.	9/3/10	741	2	371
E. Main & S. Chapel St.	9/8/10	158	2	79
Del. Ave. & S. College Ave.	9/14/10	927	2	464
Del. Ave. & Academy St.	9/23/10	442	2	221
Del. Ave. & S. Chapel St.	7/28/10	139	2	70

Parameters of interest observed during the data collection period for this research are initially categorized into two main factors that influence pedestrian crossing against traffic signal. These include: (1) human instincts/motives; and (2) roadway

environment. For the human factors, three main variables that mostly influence pedestrian crossing behavior were considered for our analysis – (a) presence of other pedestrians waiting to cross the intersection, (b) weather condition, and (c) signal timing. Also for the roadway environment, three important variables were taken into considerations and these are: (a) presence of vehicles, (b) speed of approaching vehicles, and (c) pedestrian signal cycle length.

During the field data collection exercise, the following directives were strictly observed to ensure counting mistakes are highly minimized since they preclude the scope of this study.

- The field observer must note any problem or interruption in the data collection, such as break or lack of attention for any reason. These interruptions are important since the main objective was to compare the accuracy of the methods.

- The field observer must count only pedestrians who cross at an intersection and not those who cross at midblock. Bicycle riders were not to be counted unless they are walking their bicycle across the intersection.

Field data were collated in Microsoft Excel spreadsheet database for preliminary processing and initial analysis of its statistical significance before it was finally used to calibrate the Bayesian network pedestrian crossing model. For quality control purposes, all database tables were compared with the original field data sheets.

## Chapter 4

### BAYESIAN NETWORK

#### 4.1 Introduction

Bayesian networks (or belief networks) are probabilistic graphical models representing a set of variables and their probabilistic dependencies. Over the last decade, Bayesian network has become a popular representation for encoding uncertain expert knowledge in expert systems (Heckerman et al., 1995a). More recently, researchers have developed methods for learning Bayesian networks from data. The techniques that have been developed are new and still evolving, but they have been shown to be remarkably effective for some data-analysis problems. The Bayesian networks are useful modeling tools in statistics and artificial intelligent fields. Their popularity can hugely be attributed to the intuitive graphical representation of the interdependencies between variables, and to the savings in computational independence assumptions.

Bayesian networks were originally developed to allow the impact of uncertainty about management systems to be accounted for in the decision making process. This means that decision makers can balance the desirability of an outcome against the chance that the management option selected may fail to achieve it. This facility is particularly important for environmental management where the complexity of the natural world means that it is rarely possible to predict the exact impact of any management intervention. In an uncertain world, Bayesian networks allow users to

estimate the chance that a management intervention will have a particular effect and then investigate the consequences of their uncertainty (Cain, 2001).

Bayesian network consists of a qualitative part, a directed graph, or network,  $G$  that does not include directed cycles, called an acyclic directed graph, and a quantitative part that corresponds to a joint probability distribution  $P$ . The nodes of the network correspond to the variables in the joint probability distribution. The structure of the network reflects conditional independence assumptions that must be satisfied by the associated probability distribution. These assumptions are called Markov properties.

By applying the chain rule of probability theory, taking into the Markov properties, the joint probability distribution  $P$  can be obtained in the form as shown in equation:

$$P(X_1 \dots X_n) = \prod_{i=1}^n P(X_i | \text{Parents}_G(X_i)) \quad 1$$

where  $\text{Parents}_G(X_i)$  represents the set of parents of the vertices corresponding to  $X_i$  in graph  $G$  (Gavai, 2009). Using the Markov properties, it can also be deduced that any variable  $X$  is conditionally independent of all the other variables given its parents, children, and children's parents. The parents, children, and children's parents are called the Markov blanket of the variable.

Bayesian network for a specific problem can be constructed in myriad of ways. For instance, if one knows the interaction of a system very well, one could even construct a hypothetical network manually, based on knowledge extracted from literature. Each of the entities of the system in the network would then be associated with probabilities contained in conditional probability table often referred to as *prior probabilities* or one's *degree of belief*. Tampering with prior probabilities of one node

allows us to explore and update probability values over the rest of the nodes. The graphical nature of the network combined with probability theory permits one to do data analysis in an intuitive way. If there happen to be feedback loops involved in a domain, then this situation will be better modelled using *dynamic* Bayesian networks (Murphy, 2002).

Relationships between variables of interest are established from data using partial correlations as a measure to identify (conditional) independences between variables from the available dataset. Once constructed, direct and indirect relationships can be identified easily just by looking at the Bayesian networks. There are two aspects for representing data using this technique namely qualitative and quantitative. The qualitative aspect involves representation of data using nodes and edges where their relationships are quantified using a conditional probability distribution. The nodes represent variables and the edges represent causal or influential relationships between variables. Bayesian networks therefore enable us to infer the relationships, and reason about them efficiently, from cause to effect, yielding a kind of *predictive reasoning*, and conversely, from effect to cause, yielding a kind of *diagnostic reasoning*, and eventually, inter-causal relationship, often called *intercausal reasoning*.

There are two different ways by which one can find the relationship between variables, otherwise described as learning the network from data. These include: (1) using search-and-score-based methods, and (2) using constraint-based methods. The search-and-score method has been extensively explored and applied by researchers of various fields including transportation engineering. However, a major problem identified with this approach is that they are data demanding and computationally inefficient. The search space that needs to be explored is

superexponential in the number of variables, and therefore, normally heuristic or approximation methods are used (Husmeier et al., 2005). However, this does not resolve the large requirements with respect to the amount of data. Constraint-based method, where local structure is obtained by carrying out conditional independence tests on sets of variables is then more feasible (Gavai, 2009). One other advantage of using these two methods, however, is that background knowledge can be readily incorporated into the learning process.

Because Knowledge about relationships between entities of interest is not always available, one is left with no choice than to find these relationships from the experimental data. Situation like this is very common especially when the number of variables is large and there is little or no knowledge available or accessible of the underlying process. Besides, it can be very cumbersome a task to construct networks of several hundred nodes just by hand. In view of this, a considerable number of research has been carried out in this area to apply unsupervised learning of conditional dependencies and independencies relationships from data, hence the evolution of learning Bayesian network as a typical example of such research.

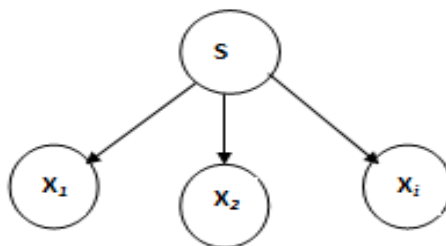


Figure 5: Schematic diagram of a simple Bayesian network structure indicating cause-and-effect relationship between parent node  $S$  and children nodes  $X_1, \dots, X_i$ .

A good number of software packages have been developed to construct, implement and manipulate Bayesian networks due to its burgeoning popularity. The networks constructed in this thesis have been implemented through the commercially available product called Netica<sup>(R)</sup> software. We begin by introducing the concept of probabilistic models and continue with the construction process of our pedestrian crossing behavior Bayesian networks model.

## 4.2 Probability Distribution

To be able to fully understand the underlying principles behind Bayesian network, this section provides a quick, brief and summarized review of the fundamental concepts used in probability theory (Gavi, 2009). First, we have to understand the meaning of *Events* and *Variable* as they are extensively used in the field of statistics to convey a message.

An event ( $E$ ) is often used to describe the existing state of our world in which we live. Several events may occur either alternatively or concurrently, and such joint events are often described using Boolean operators such as conjunction ( $(E \wedge E')$  or  $(E \cap E')$  or  $(E, E')$ ), disjunction ( $(E \vee E')$  or  $(E \cup E')$ ) and negation ( $\neg E$ , also denoted  $E^-$ ). All events are defined as a subset of a *sample space* ( $S$ ), which consists of all possible outcomes of an experiment. It is often a common practice and also more convenient to decompose a sample space into several, possibly disjointed, subsets, which introduces the notion of a *variable*.

Variables may take elements of their associated domain as value, represented as  $X = x$ , where  $X$  denotes the variable, or sets variables, and  $x$  is the value



the variables take on. In this thesis, we represent each node in our model with a variable such as *season*, *weather condition*, *vehicle speed* etc. Hence, the state of our variable *season* has two distinct values which are represented in our model as *summer* and *fall*.

Probability distribution expresses *uncertainty* about the occurrence of events, for instance, due to measured errors. Given a set of variables  $X$  with associated domain  $D(X)$ , it is stated that values of  $D(X)$  are exhaustively, and mutually exclusive. This implies that it holds for a set of discrete variables that  $P(X=x, X=x') = 0$ , where  $x \neq x'$ , are two different values of  $X$ , and in addition to that,  $\sum_{x \in D(X)} P(X = x) = 1$ .

In theoretical sense, it generally doesn't make any difference whether  $X$  is a single variable, or singleton set, or a non-singleton set of variables. A probability distribution  $P(X)$  where  $X$  is a set of variables is called a *joint* or *multivariate* probability distribution. Special approaches have been developed to cope with joint probability distributions including several variables. Coping with statistical independence information become particularly relevant when representing joint probability distributions as the size of a discrete joint probability distribution that is exponential in the number of variable.

#### ***4.2.1 Conditional Probability Distribution***

Given two distinct events,  $X$  and  $Y$ , a conditional probability is denoted as  $P(X | Y)$ , which can be interpreted as finding the probability of an event  $X$  knowing for sure that event  $Y$  has already occurred or existed. Thus, by mathematical definition:

$$P(X | Y) = \frac{P(X \cap Y)}{P(Y)} \quad 2$$

where  $P(Y) > 0$ . This equation gives us a clear understanding about the fact that the uncertainty with respect to event  $Y$  is divided out in the uncertainty in the joint event  $(X \cap Y)$  and for that matter, makes only event  $X$  remains uncertain.

To further explain this point, let's supposed we had an experiment and observed that the event  $Y$  shows up 8 times after certain trials. Assuming of those 8 that shows up, event  $X$  is also expressed 5 times, then given the two events are mutually exclusive,  $P(X | Y)$  would be estimated as  $5/8 \approx 0.63$  or 63%. Thus, the probability that event  $X$  is expressed given the fact that the event  $Y$  has already occurred is estimated to be 63% by using a method that involves counting the occurrence of the events, which is then translated into relative frequencies of the two events  $X$  and  $Y$ .

If  $X$  is likely to occur when  $Y$  occurs, then knowing that  $Y$  has already occurred raises ones confidence in assigning higher probability to  $X$ 's occurrence than in situation where there is no information about the occurrence of  $Y$ . We can explicitly express this notion by comparing the probability distribution  $P(Y | X)$  and the prior distribution  $P(X)$ . In general, if it is known that  $X$  and  $Y$  systematically co-vary in some way, then  $P(X | Y)$  will not equal to  $P(X)$ . On the other hand, if  $X$  and  $Y$  are independent events, then  $P(X | Y)$  would be expected to equal  $P(X)$ . The need to compute conditional probability arises whenever we think the occurrence of some event has a bearing on the probability of another event's occurring.

The most basic and intuitive method for computing  $P(X | Y)$  is “the set enumeration method”. According to this method,  $P(X | Y)$  can be computed by counting the number of times  $X$  and  $Y$  occur together,  $N_{X,Y}$ , and dividing by the number to times  $Y$ ,  $N_Y$ , occurs which is mathematically expressed below:

$$P(X | Y) = \frac{N_{X,Y}}{N_Y} \quad 3$$

yielding a popular relative frequency definition of the probabilities according to Laplace. Following the definition of conditional probability distribution from above, the very famous Bayes’ theorem can be derived. It follows from Equation (1) above that:

$$P(X \cap Y) = P(X | Y)P(Y) \quad 4$$

Likewise, we also can write that:

$$P(Y \cap X) = P(Y | X)P(X) \quad 5$$

However, the left hand side expressions of Equations (3) and (4) are commutative and holds that  $P(X \cap Y) = P(Y \cap X)$ , hence we combine the two equations as follows:

$$P(X | Y)P(Y) = P(Y | X)P(X) \quad 6$$

After a little algebraic manipulation, we arrive at *Bayes’ Theorem*:

$$P(X | Y) = \frac{P(Y | X)P(X)}{P(Y)} \quad 7$$

where  $P(Y) > 0$ . It can be inferred that Bayes formula and the formula for computing conditional probability differ only in the numerator where in the original conditional probability formula, the numerator  $P(X \cap Y)$  is replaced by  $P(Y | X)P(X)$  according to Bayes theorem. Moreover, Equation (3) is sometimes referred to as the factorization theorem and it is unique in its recursive application, as exemplified in the following expression:

$$P(X, Y, Z) = P(X | Y, Z)P(Y, Z) = P(X | Y, Z)P(Y | Z)P(Z)$$

This provides special technique that can be applied to decompose any joint probability distribution into factors.

### **4.3 Pedestrian Crossing Behavior Model**

We begin building pedestrian crossing behavior model by first identifying the variables that will be represented in the model as nodes. These variables includes: season, weather condition, presence of other pedestrians, pedestrian signal cycle phase length, and presence of vehicles. The remaining variables used to construct the model encompasses: speed of oncoming vehicles, pedestrian motive, roadway condition and pedestrian crossing behavior.

Nodes that have direct influence on other nodes are linked by arrows which emanate from the influential node and terminate on the influenced node. The node from which arrows originate is referred to as the parent node whiles the node on which an arrow terminates is referred to as the child node. All parent nodes such as season, presence of other vehicles, and presence of other pedestrians are described as independent nodes because their existence is not influenced by any other parent node.

Conversely, the remaining nodes in the model are classified as dependent nodes because they are directly influence by their respective parent variables.

According to this model, arrows drawn from left node (parent node) to right node (child node) implies the left nodes influences the right nodes or the right nodes are contingent on or influence by the left node. For instance, the *season* node directly influences two other variables (weather condition and vehicle speed) as depicted in our model. However, pedestrian motive node is influenced by three different variables (signal length, weather condition, and other's presence) just as roadway environment node is also influenced by another three of the variables (vehicles presence, vehicle speed, and signal length). Eventually, pedestrian motive node and roadway condition node together influence pedestrian behavior node.

Since left nodes in the network influence the nodes directly on the right hand side which link them by arrows, there exists a conditional relationship as one goes from right node to left node in an opposite direction of the arrows. This conditionality between variables is what the Bayesian network uses to model reality which is then used for prediction and diagnoses. In our case, we are concerned with how seasonal variations affect behavior of pedestrians at intersections and use conditional probability to analyze the season with most rational or eccentric pedestrian crossing behavior. Figure 6 below is a brief graphical display of the Bayesian network pedestrian crossing model.

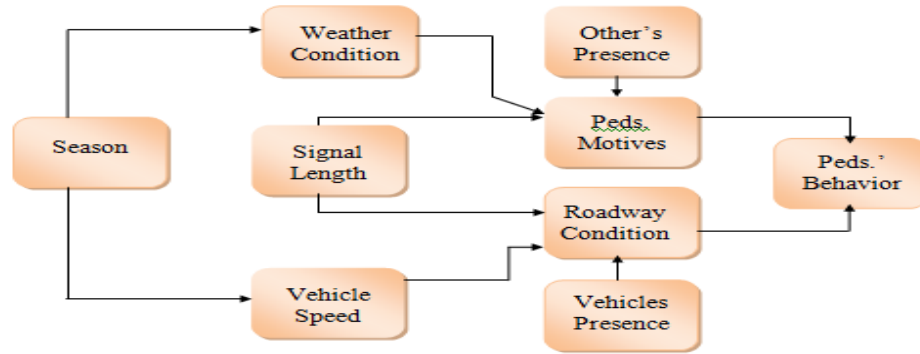


Figure 6: Schematic structure of pedestrian crossing behavior model illustrating causal relationships between variables. Arcs are drawn from cause to effect.

#### 4.4 Model Development

A statistical commercial Bayesian network modeling software called Netica<sup>(R)</sup> was used to model factors that critically influence pedestrian crossing behavior. After constructing the model, the results obtained enable us to determine which of the two seasons (summer, spring) do pedestrians behave more erratically when it comes to crossing a roadway intersection. The model building process is begun by first breaking down its development methodology into three main levels/stages: (1) Graphical level, (2) Information level, and (3) Numerical level. A detail exposition is given into each of the levels of the model development process in order to forestall any potential ambiguity that might arise in the minds of all stakeholders who were not necessarily involved in its construction.

#### ***4.4.1 Graphical Level***

The graphical level is considered as the fundamental level during the construction process of the Bayesian network model. At this level, we begin by identifying our nodes with their appropriate names that will enable us communicate our conceptual ideas of the environment we are modeling. We first of all identified our nodes according to these node names: season, weather condition, pedestrian signal cycle length, speed of vehicles, presence of other pedestrians, presence of vehicles at intersection, roadway environment, human motives and pedestrian crossing attitude. The variables were selected based on the fact that exhaustive literature search discloses them as the most significant factors that influence pedestrian safety.

The nodes are linked with arrows according to cause-and-effect relationship to eventually make up a complete graphical Bayesian network structure. Arrows in the network originate from parent node and terminate on child node. The position of each node in the network structure is determined by whether that node is an: (i) implementing factor, (ii) intervention, (iii) intermediate factor, or (iv) objective. An implementing factor is usually the node that initiates the entire network structure and determines whether an intervention can be implemented both immediately and in the future. The season node according to our model is a typical example of an implementing factor or variable.

Intervention nodes symbolize things one wish to implement in order to achieve their objectives. Nodes such as weather condition, presence of others, signal cycle length, vehicle speed, and presence of vehicles all constitute intervention factors in our model. Intermediate factors are those nodes that link objectives and interventions. Human motive node and roadway environment nodes are the two typical examples of intermediate nodes according to our model. Finally, objective nodes are

usually the last node on which the network terminates and such nodes often provide a vague idea about the main purpose for which the model has been built. For instance, the objective node for our model is the pedestrian crossing behavior node because not only is it the last node on which the entire network terminates, but also because it is the central part of the model.

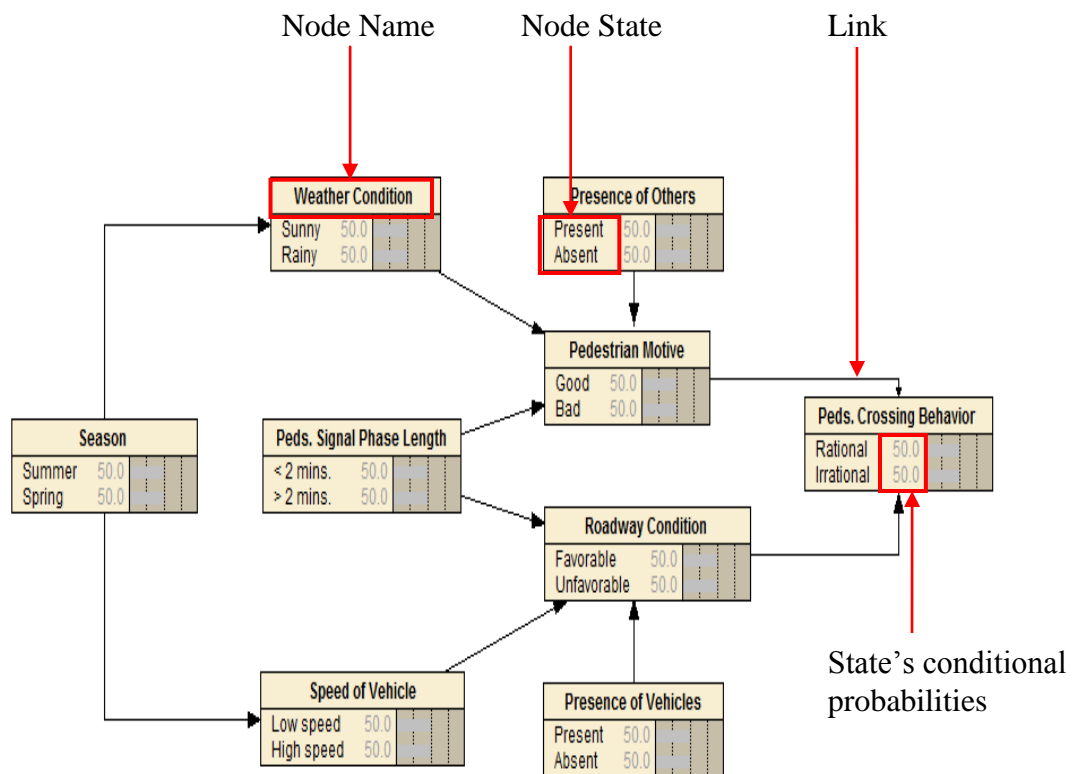


Figure 7: Bayesian network model in its graphical and information level of model development.



Table 2: Model's node names and their respective states

#	NODE NAME	STATES
1	Season	1. Summer 2. Fall
2	Weather Condition	1. Sunny 2. Rainy
3	Peds. Signal Cycle Length	1. < 2mins. 2. > 2mins.
4	Speed of Vehicle	1. Low speed 2. High Speed
5	Presence of Vehicles	1. Present 2. Absent
6	Presence of Other Pedestrians	1. Present 2. Absent
7	Roadway Environment	1. Favorable 2. Unfavorable
8	Human Instincts/Motives	1. Good 2. Bad
9	Pedestrian Crossing Attitude	1. Rational 2. Irrational

#### 4.4.2 Information (Qualitative) Level

This is often considered as the qualitative part of the Bayesian network structure where every aspect of the network is explicitly explained. Thinking about what states to give a node is an excellent check on whether the variable you have chosen properly represents the idea you are trying to capture. The states of each of the nodes of our model were cautiously chosen to ensure that they fit in with the logic of our Bayesian network model of pedestrian crossing behavior as a whole.

Basically, each of the variables used in this model as illustrated in figure 7 above has two states. All the nodes in the model have a discrete binary state. Thus, a node is identified to be in either one state or the other at any given time. For instance, the state of the season node is said to be either *summer* or *spring*, the state of weather condition node is given to be either *favorable* or *unfavourable*, the state of presence of other pedestrian node is given to be either *present* or *absent*. Also, the state of signal timing length is given to be either *less than 2 minutes* or *greater than 2 minutes*, the state of presence of other vehicles node is given to be either *present* or *absent*, the state of vehicle speed is given to be either *high* or *low* speed, the state of pedestrian motive node is given to be either *good* or *bad*, the state of roadway condition node is given to be *favorable* or *unfavourable*, and the state of pedestrian crossing behavior node is given to be either *rational* or *irrational*.

Choosing states is fairly straightforward while one is trying to represent the basic ideas but becomes more difficult when you start to fill in the conditional probability tables CPTs (Cain, 2001). This is because you will often need to define exact values for the states you have chosen. To begin with, however, one does not have to restrict themselves by worrying about quantifying the states s/he chooses. It is more important to make sure the Bayesian network is logical and expresses all the

necessary ideas. If one needs to do it anyhow, it will be possible to adapt it later to help fill in the CPTs.

#### ***4.4.3 Numerical (Quantitative) Level***

This level which is also described as the quantitative phase is considered as the final part of the model building process. The CPTs of each of the nodes are filled with appropriate data and after compiling the network, a graph in a form of belief bars is displayed against the state of each node. Beside each belief bar in the model's output are numerical values (range: 0-100) that represent the probability of each variable being in either of the states of a given variable. The values can also be described as a numerical representation of one's degree of belief (independent or conditional probability values) for each of the states of every node. Analysis for the model is therefore based on these numbers (probabilities) that are displayed against the state of the nodes.

The parent nodes in the network are all independent variables with the probability of their respective states originating from either direct field observation, elicitation from stakeholder's opinion, expert knowledge or reasonable hypothesis. Table 3 below summarizes the probability or one's degree of believe of the states of each parent node that were used as direct inputs of the model. It must however be noted that the state probability values for the season and pedestrian signal timing length nodes presented in the Table 2 are both hypothetical values that were initially conjectured for the purpose of this study. Because of the assumptions used to obtain such subjective probability values, it was logically reasonable that equal probability values (50%) was used for each state of such variables. Conversely, the state

probability values for presence of other pedestrians and presence of vehicles' node presented in the table below are directly observed data that were obtained from field observation.

Table 3: Observed and elicited subjective probabilities for all parent nodes.

PARENT NODE NAME	STATE NAME	STATE PROBABILITY
Season	Summer	0.50
	Spring	0.50
Peds. Signal Timing Length	< 2 minutes	0.50
	> 2minutes	0.50
Presence of Other Peds	Present	0.60
	Absent	0.40
Presence of Vehicles	Present	0.90
	Absent	0.10

The Netica software applies Baye's theorem to compute the conditional probabilities for the states of all the child nodes in the model. A Bayesian network for a set of variables  $X = \{X_1, \dots, X_n\}$  consists of (1) a network structure S that encodes

a set of conditional independence assertions about variables in  $X$ , and (2) a set  $P$  of local probability distributions associated with each variable. Together, these components define the joint probability distribution for  $X$ . The network structure  $S$  is a directed acyclic graph. The nodes in  $S$  are in one-to-one correspondence with the variables  $X$ . We use  $X_i$  to denote both the variable and its corresponding node, and  $Pa_i$  to denote the parents of node  $X_i$  in  $S$  as well as the variables corresponding to those parents. The lack of possible arcs in  $S$  encodes conditional independencies. In particular, given structure  $S$ , the joint probability distribution for  $X$  is given by

$$p(x) = \prod_{i=1}^n p(x_i | pa_i) \quad 8$$

The local probability distributions  $P$  are the distributions corresponding to the terms in the product of equation (7). Consequently, the pair  $(S; P)$  encodes the joint distribution  $p(x)$ . The probabilities encoded by a Bayesian network may be Bayesian or physical. When building Bayesian networks from prior knowledge alone, the probabilities will be Bayesian. When learning these networks from data, the probabilities will be physical (and their values may be uncertain).

In the next phase of Bayesian-network construction, we build a directed acyclic graph that encodes assertions of conditional independence. One approach for doing so is based on the following observations. From the chain rule of probability, we have

$$p(x) = \prod_{i=1}^n p(x_i | x_1, \dots, x_{i-1}) \quad 9$$

Now for every  $X_i$ , there will be some subset  $\Pi_i \subseteq \{X_1, \dots, X_{i-1}\}$  such that  $X_i$  and  $\{X_1, \dots, X_{i-1}\} \setminus \Pi_i$  are conditionally independent given  $\Pi_i$ . Thus, for any  $x$ ,

$$p(x_i | x_1, \dots, x_{i-1}) = p(x_i | \pi_i) \quad 10$$

Combining equations (8) and (9) yields the following equation

$$p(x) = \prod_{i=1}^n p(x_i | \pi_i) \quad 11$$

Tabulated below are the results of the computed conditional probabilities for each of the states of all child nodes. These conditional probability values are what is displayed in our model's output as ones degree of belief that a given child node is in a particular state or the other.

Table 4: Conditional probability table (CPT) for *Weather Condition* node

	<i>Weather Condition:</i>	
	Sunny	Rainy
<b>Season</b>		
Summer	0.95	0.05
Spring	0.40	0.60

Table 5: Contional probability table for Vehicl Speed node

	<i>Vehicle Speed:</i>	Low	High
<b>Season</b>			
Summer		0.95	0.80
Spring		0.70	0.30

Table 6: Conditional probability table for Roadway Environment node

		<i>Roadway Env.:</i>	Favorable	Unfavorable
<b>Veh. Present</b>	<b>Vehicle Speed</b>	<b>Peds. Sign.</b>		
Present	Low	> 2mins.	0.10	0.90
Present	Low	<= 2mins	0.90	0.10
Present	High	> 2mins.	0.05	0.95
Present	High	<= 2mins.	0.20	0.80
Absent	Low	> 2mins.	0.60	0.40
Absent	Low	<= 2mins.	0.95	0.05
Absent	High	> 2mins.	0.20	0.80
Absent	High	<= 2mins	0.30	0.70

Table 7: Conditonal probability table for Pedestrian’s Motive node

			<i>Peds.’ Motive:</i>	
			Good	Bad
<b>Other’s Pre.</b>	<b>Weather</b>	<b>Peds. Sign.</b>		
Present	Sunny	> 2mins.	0.80	0.20
Present	Sunny	<= 2mins	0.95	0.05
Present	Rainy	> 2mins.	0.10	0.90
Present	Rainy	<= 2mins.	0.30	0.70
Absent	Sunny	> 2mins.	0.50	0.50
Absent	Sunny	<= 2mins.	0.90	0.10
Absent	Rainy	> 2mins.	0.05	0.95
Absent	Rainy	<= 2mins	0.10	0.90

Table 8: Conditional probability table for Ped. Crossing Behavior node

		<b>Crossing Behavior:</b>	
		Rational	Irrational
<b>Human Motives</b>	<b>Roadway Env.</b>		
Good	Favorable	0.95	0.05
Good	Unfavourable	0.70	0.30
Bad	Favorable	0.40	0.60
Bad	Unfavourable	0.50	0.95



The conditional probability values as presented in the above tables are the direct input values used by the Netica software to compute the belief values displayed against each state in the network. It is however imperative to underscore the fact that these values may have some inherent biasness since it is realistically complex to generalize the actions and perceptions of entire human race. However, the values herein obtained have proven to be statistically reasonable from a preliminary test performed such as ANOVA test which are further elaborated in Appendix II. A detailed exposition and analysis of the data and subsequently the results obtained from the model was carried out as explain in chapter 5 of this study.

## Chapter 5

### DATA ANALYSIS AND DISCUSSION OF RESULTS

#### 5.1 Data Analysis

Field data comprising of a total of 400 pedestrians for spring season and 1200 pedestrian for summer season as observed at all the six different intersections was used for this analysis. During the spring season, it was observed that 280 (70%) of pedestrians crossed the intersection unlawfully (i.e. crossed against stop signal), 20 (5%) waited to cross legally because of others presence whiles the remaining 100 (25%) crossed lawfully. However, it was also observed during the summer season that 300 (25%) pedestrians crossed the intersection illegally (crossed against stop signal), 420 (35%) waited for their legal turn because of presence of others whiles the remaining 480 (40%) crossed legally irrespective of traffic or others presence.

Stakeholders and experts were also contacted for their opinion which together with a confirmation from results obtained from field observation helped in the estimation of the values for the remaining parameters such as weather condition, traffic situation, speed of vehicles during the two seasons, presence of vehicles, and presence of other pedestrians.

The data was statistically collated after initial normalization and preprocessing to determine the strength of their correlation and statistical significance. In order to properly conceptualize and characterize the behavior of the data, initial

plots were carried out. From ANOVA and t-test analysis, the measured variables proved to have a strong correlation and statistically significant. Thus, preliminary analysis of data yielded a 95% confidence interval results which is very satisfactory and suitable according to the scope of this study to be used to calibrate and analyze the Bayesian network model. Attached in Appendix II is an illustration of the ANOVA and p-values analysis for the field data.

Preliminary analysis of the data was carried out to help show the plots of the parameters observed at each of the six different locations where data was collected. The empirical data that was collected for each of the two seasons was used to generate bar charts to be able to graphically delineate the pattern of behavior that is contained in the data. As mentioned earlier, the behavior of pedestrians as they crossed each of the six study locations were categorized into three main classes namely: those who crossed the road illegally, those who waited because of other people waiting to cross and those who just waited for their legal turn to cross regardless of traffic or other peoples' presence.

The bar graph generated for the spring season as depicted in figure 8 reveals illegal crossing behavior was highest among all six of the study locations while those who waited to cross because of other's presences turn out to be the least for all the study locations. Of all the six intersections where illegal crossing behavior was observed, Delaware Avenue and South College Avenue crossing location recorded the highest lawlessness in pedestrian crossing behavior. East Main Street and North Chapel Street intersection recorded the least illegal pedestrian crossing behavior. Conversely, for the same locations where pedestrians crossed legally, East Main Street and North Chapel Street surprisingly recorded the highest of law abiding

pedestrians whereas East Main Street and South Chapel recorded the least results for pedestrians that crossed lawfully.

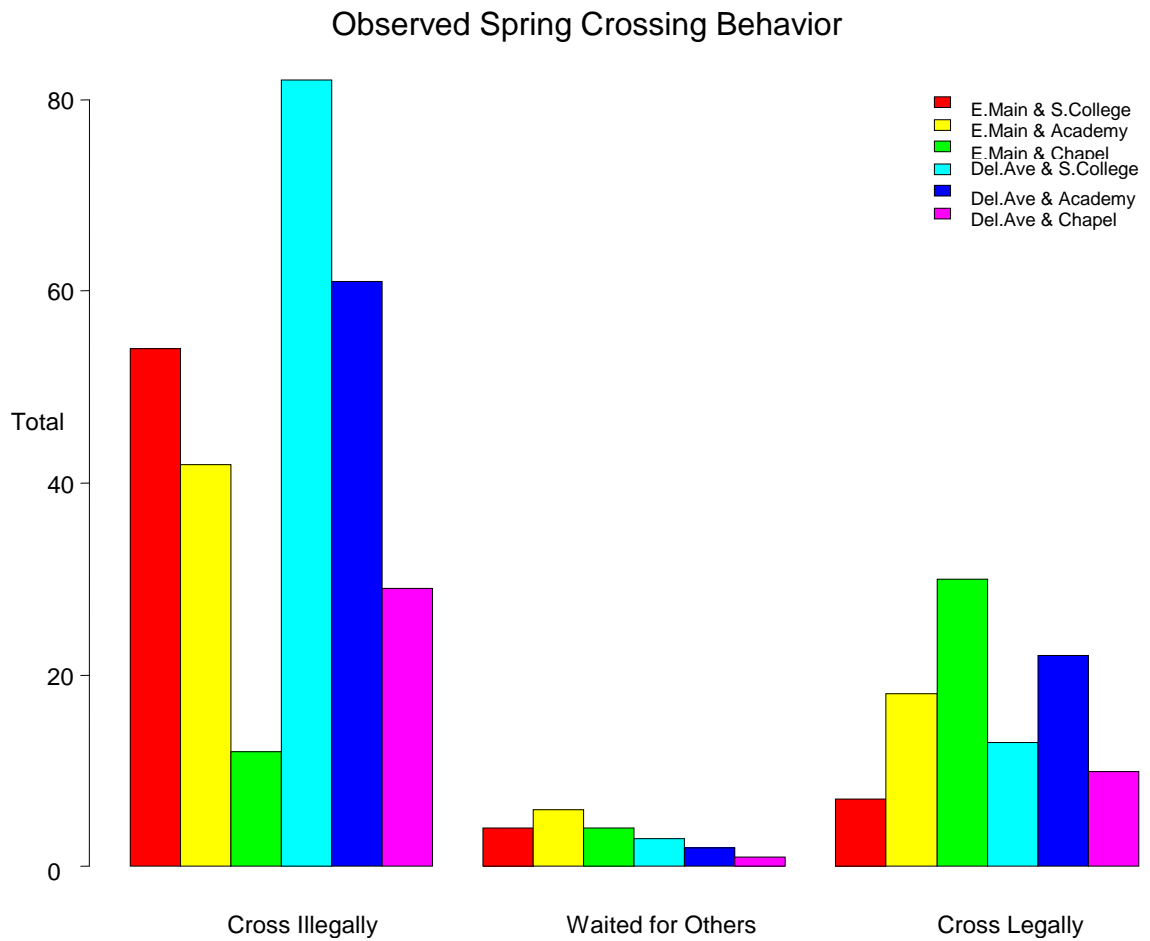


Figure 8: Bar chart showing pedestrian crossing behavior at each of the six (6) study locations during spring season.

In the same manner, bar plot for summer season was generated using empirical data collected during the summer time. The graphical output from the summer season data was absolutely different compared to that of the spring season. The output of the two graphs attests to the fact that pedestrian roadway crossing behavior changes over time and at different locations. Unlike the spring season, the summer season plot reveals that pedestrians crossed the road lawfully for most of the six study locations than they did unlawfully during last spring season. The graph as depicted in figure 9 shows that the location with the highest observed legal crossing behavior of pedestrians is East Main Street and Academy Street whereas the location with the least observed pedestrians showing law abiding crossing behavior is Delaware Avenue and Academy Street. However, for those who crossed the intersection unlawfully, East Main Street and Academy Street surprisingly recorded the highest whereas Delaware Avenue and North Chapel Street recorded the least.

### Observed Summer Crossing Behavior

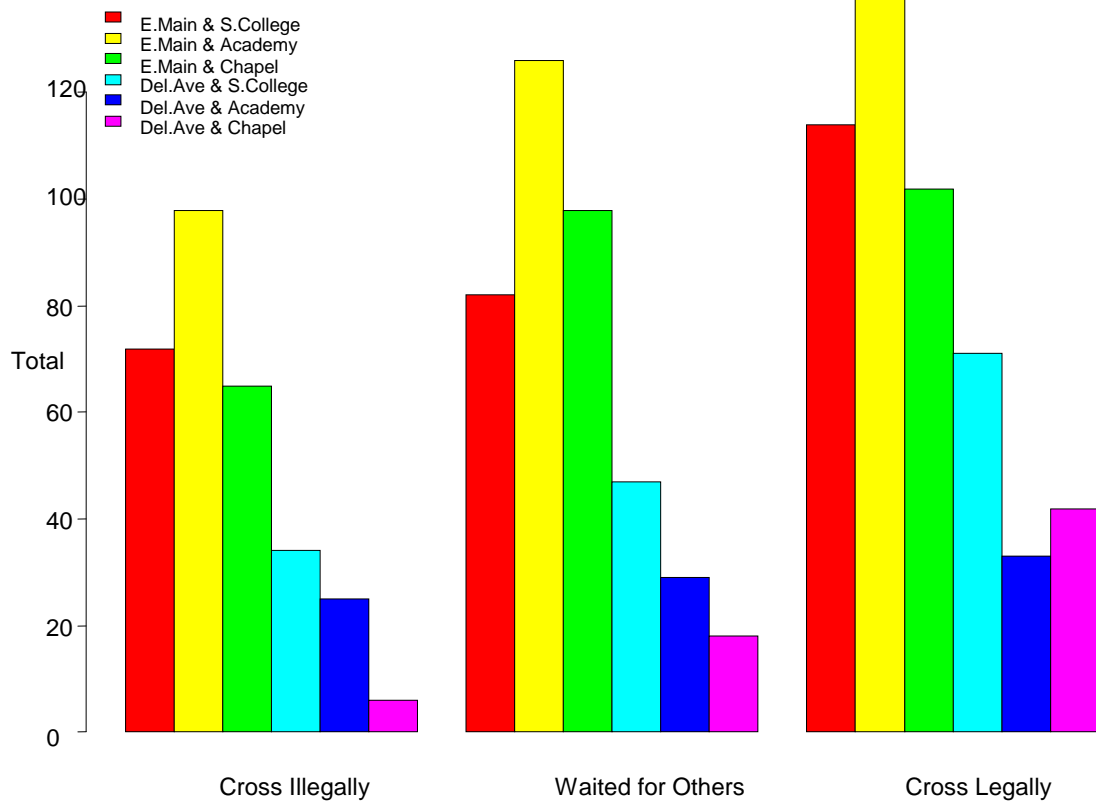


Figure 9: Bar chart showing pedestrian crossing behavior at each of the six (6) study locations during summer season.

Calibration of the Bayesian network model of pedestrian crossing behavior was then carried out after the preliminary data preprocessing and analysis. Underlying each node in the Bayesian network model are either unconditional or conditional probability tables (CPTs) depending on whether the node is a parent or child. The

probability of states of all parent nodes are directly interred into unconditional probability table associated with each parent node whiles observed state probabilities are interred into their respective conditional probability tables associated with each child node. Table 9 below is an example of a CPT describing the relationship between pedestrian crossing behavior (child node) and human motive/instincts as well as roadway environment (the two parent nodes). CPT contains entries for every possible combination of the states of the parents.

Table 9: Conditional probability table for *pedestrian crossing behavior* node

		<b>Crossing Behavior:</b>	
		Rational	Irrational
<b>Human Motives</b>	<b>Roadway Env.</b>		
Good	Favorable	0.95	0.05
Good	Unfavourable	0.70	0.30
Bad	Favorable	0.40	0.60
Bad	Unfavourable	0.50	0.95

The first row of the above table can be interpreted as: “if human motives is good and roadway environment is favorable, then there is 95% chance that pedestrian crossing behavior will be rational, and 5% chance that pedestrian crossing behavior will be irrational.” The model’s final output after all its conditional and unconditional probability tables has been filled with data and compiled is presented in figure 10 below.

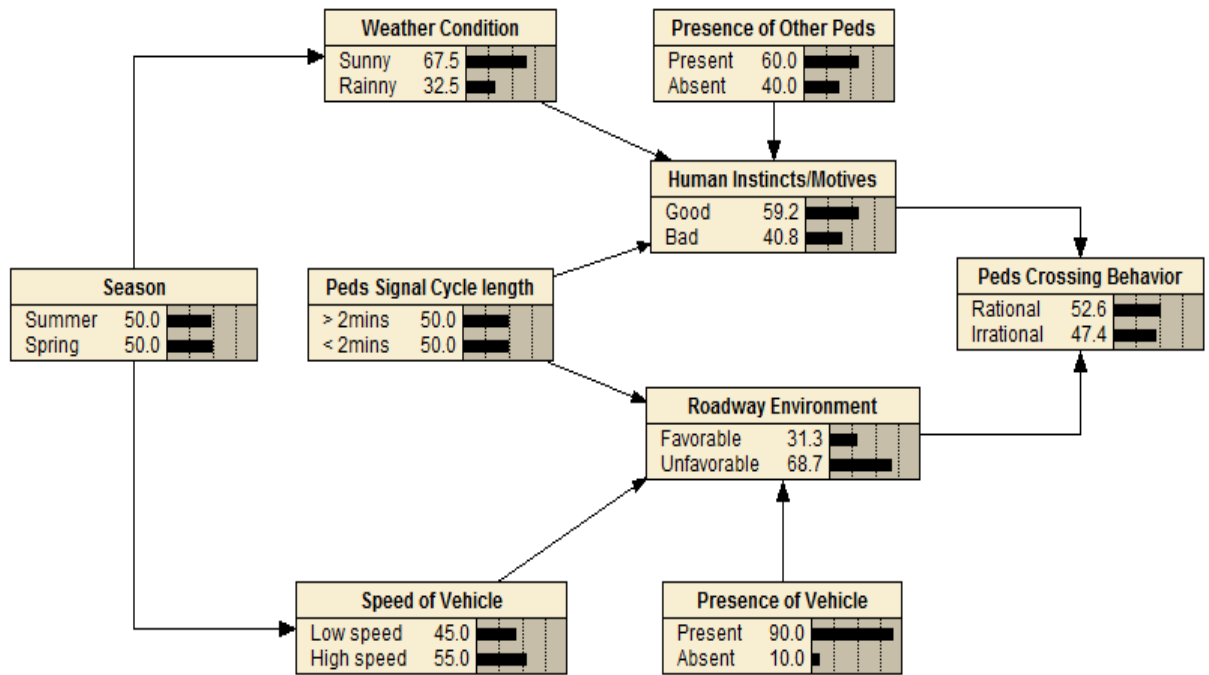


Figure 10: Compiled Bayesian network model indicating final results for each of the states after the network has been calibrated with data.

Once all the CPT for each node is completed with data, the network is eventually compiled and used for further analysis. In a general sense, this is performed by altering the states of some nodes of interest while observing the effect this has on corresponding nodes of interest. Since the model is serially connected to form a chain network, the impact arising from varying the state of any node is instantaneously transmitted through the network in accordance with the relationships expressed by the CPTs. Changes in any node simply arise from the combined effect of changes in all the nodes linked to it either directly or indirectly. This idea justifies earlier claim that



Bayesian network encodes a joint probability distribution over all the nodes. Each time a change is effected to the state of a node, the joint distribution is automatically altered through the iterative application of Bayes' theorem. Changes in the Bayesian network are observed as changes in the chance that a node is in a particular state. Uncertainty in the CPT often makes it rare for a node to definitely be in one state or another and it is far more common for probability distributions across all the states of a node to be observed.

## 5.2 Discussion of Results

The model's output depicted in figure 10 is realistically meaningful as observed from field data collection as well as elicitation from experts' judgment and stakeholders opinion. An initial equal state probability of 50% was assumed for each of the states of the season variable because of the equal time window that was used to collect data for each of these two states. This therefore explains the reason why the belief bars is showing the same probability (50% each) for *summer* and *spring* season. However, during the data collection period, it was observed that there were relatively more sunny days in the summer season {i.e  $p(\text{Sunny} | \text{Summer} = 0.95)$ } than there were rainy days {i.e.  $p(\text{Rainy} | \text{Summer} = 0.05)$ } and conversely, there were rather a little more rainy days in spring season {i.e  $p(\text{Rainy} | \text{Spring} = 0.60)$ } than there were sunny days {i.e  $p(\text{Sunny} | \text{Spring} = 0.40)$ }. Consequently, the model generated the degree of belief for each of the states of weather condition node (Sunny = 67.5%, Rainy = 32.5%) based on Bayes' theory of conditional probability.

The model generated the degree of belief for the vehicle speed node (low speed = 45%, high speed = 55%) from observed field data based on conditional

probability approach {i.e  $p(\text{low speed} \mid \text{summer} = 0.2)$ ,  $p(\text{high speed} \mid \text{summer} = 0.8)$ ;  $p(\text{low speed} \mid \text{fall} = 0.7)$ ,  $p(\text{high speed} \mid \text{fall} = 0.30)$ }. However, it was observed that about one half of our study area location has pedestrian signal timing to be less than 2 minutes while the other half has signal timing to be greater than 2 minutes depending on the engineering and planning design of each of the intersection. Therefore, since the signal timing node is not conditioned on any parent node, there is no conditional probability associated with this node and so its states has equal degree of belief ( $>2\text{mins.} = 50\%$ ,  $<2\text{mins} = 50\%$ ).

It can also be inferred from the model that 60% of the time of data collection period there were other pedestrians present at the intersection waiting for their legal crossing time and 40% otherwise. Also, for each of the intersections where data was collected, it was observed that 90% of the time there were vehicles present and 10% otherwise.

Using information gathered from experts' knowledge as well as stakeholders' opinion, actual probability for the states of both human instincts node and roadway environment node were elicited. However, these two nodes according to our model are each conditioned on three other parent nodes thereby making their states take on the conditional probability values instead of the actual field observed values. Thus, for the human instincts variable, there is 59.2% belief that pedestrian motives were good and 40.8% belief that their motives were bad given the conditionality related to the states of the other three parent nodes. Also for roadway environment node, there is 31.3% belief that this is favourable all the time and 68.7% belief that it is always unfavourable given the conditional relation that exist between the state of that node and that of three other parent nodes that link to it.

For the target node, the model displays the beliefs for its states that were rather computed by the software as a result of a conditional relationship that exist between the node and other two parent nodes. Hence, node's state indicates that 52.6% of the time pedestrian crossing behavior will be rational and 47.4% of the time they exhibited irrational crossing behavior. The model's result for the pedestrian crossing behavior node was compared with actual field observed results (rational=53%; irrational=47%). The two results turned out to be almost the same with model's acceptable root mean square error of 0.02m. This therefore confirms the accuracy of the model and makes it an effective and acceptable tool for such behavioral studies.

### **5.3 Sensitivity Analysis**

Since the primary objective of this research is to investigate the effect of the two seasons (summer & spring) on pedestrian crossing behavior, and also understand how signal cycle length influences crossing behavior, a sensitivity analysis was performed on two variables. The analysis was first carried out using the season variable followed by the signal cycle length variable. The essence of the sensitivity analysis for these two variables was basically to enable us be able to predict how pedestrian crossing behavior varies from season to season and from shorter pedestrian waiting time to a longer one. However, a backward sensitivity analysis was also investigated for pedestrian crossing behavior variable in order to be able to understand which of the two crossing behavior types is more common during summer and spring season.

### 5.3.1 Season Sensitivity Analysis

The two states of the season variable (summer and spring) were initially varied simultaneously in order to ascertain the magnitude of their influence on the target variable (peds. crossing behavior). Initially, a 50% degree of belief for each of the two states of the season variable was used as a base case scenario and their subsequent influence on the crossing behavior variable showed that 52.6% of the time pedestrians behave rationally in crossing a roadway intersection and 47.4% of the time they behave otherwise.

However, an initial assumption of 0% and then 100% degree of belief for each of the two states of the season variable was used alternatively and their corresponding effect on pedestrian crossing behavior was closely observed. The outcome of this sensitivity is depicted in figure 11 below.

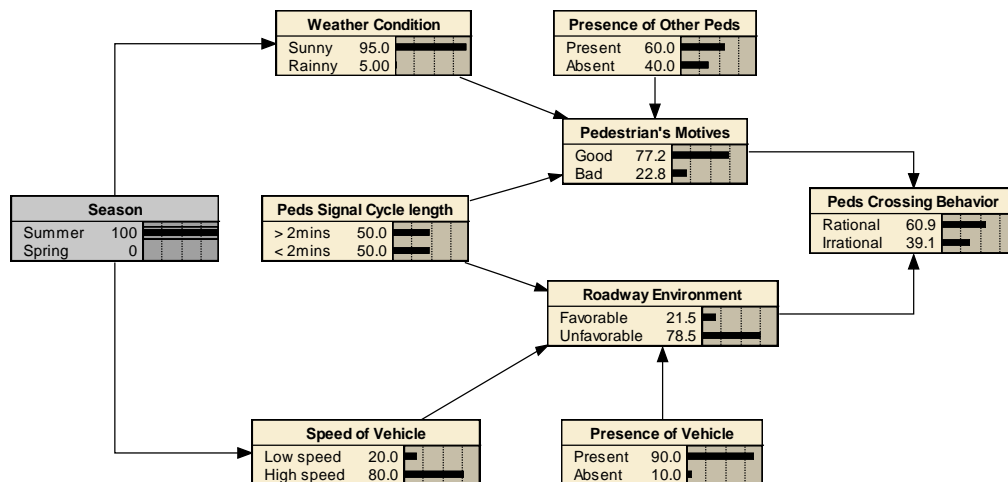


Figure 11: Bayesian network model showing the results of varying the states of the season node (summer=100%; spring=0%).

There was a significant change in all the nodes that are connected to the season node when the sensitivity was carried out. However, the effect on the pedestrian crossing behavior variable is what the analysis of this sensitivity is based. The belief that pedestrian crossing behavior will be rational given that it is summer season showed a substantial increase from 52.6% to 60.9%, (percentage increase = 15.8%). Conversely, the belief that pedestrian will exhibit irrational behavior given that it is summer decrease significantly from 47.4% to 39.1 by the same amount of percentage decrease value of 15.8%.

This result is realistically meaningful as pedestrian are more cautious in crossing an intersection during summer season because of the pool of vehicles that flood the roadways leading to congestion and lots of reckless driving. Also, because the general weather condition during summer is always nice, good visibility as well as normal roadway condition induces drivers to sometimes drive above normal speed limit. This however makes the entire roadway environment more dangerous and consequently unfavourable to pedestrians. As a result of this, pedestrian naturally becomes more cautious (good motive) about the high traffic risk and try to always play it safe by obeying traffic rules. Hence, they tend to behave more rationally and cautiously during the summer season. The sensitivity for the spring season on the other hand shows a rather opposite story as depicted in figure 12 below.

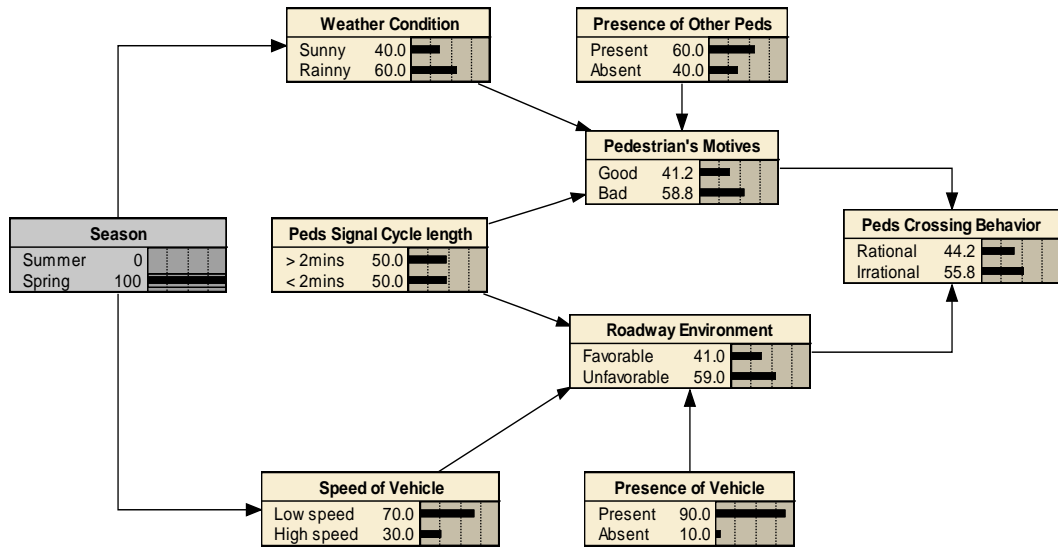


Figure 12: Bayesian network model showing the results of varying the states of the season node (summer=100%; spring=0%).

Given 100% evidence that we are in the spring season, again the model shows a significant change in all the nodes that are connected to the season (parent) variable. The consequence of the sensitivity on pedestrian crossing behavior shows substantial increase in belief from 47.4% to 55.8%, that pedestrians behave more irrationally during spring season (percentage increase = 17.7%). Conversely, the belief that pedestrian will exhibit rational behavior given that it is spring season significantly decreased from 52.6% to 44.2% by a percentage decrease of approximately 16%.

The result is realistically convincing for spring season as pedestrians become more careless but rather drivers become more careful. Due to the frequent downpour of rain during spring season, drivers tend to exercise more caution whiles

driving to avoid accident by slowing down their speed. Pedestrians on the other hand abhor standing in the rain waiting for their legal turn to cross the road and so they gradually lose their composure (bad motive) and eventually end up disregarding traffic signals in such circumstance. Hence, they tend to behave more erratically and carelessly trying to avoid being sodden by the downpour of rain.

Further sensitivity analysis was carried out by varying the belief of the states of the season variable over equal interval of 10 units and corresponding changes in the pedestrian crossing behavior variable was noted. A linear graph was plotted to ascertain the effect of varying the beliefs of the states of the season variable on pedestrian crossing behavior.

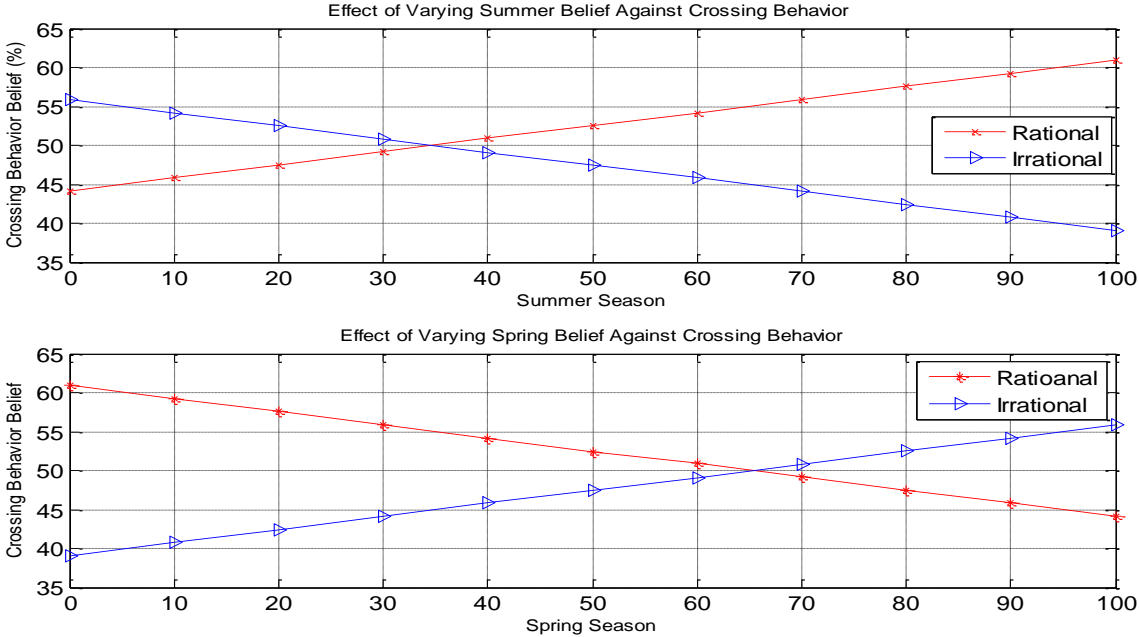


Figure 13: Sensitivity plots for Summer and Spring Season.

The two sensitivity graphs are direct opposite of each other as they both depict how the two states of pedestrian crossing behavior (rational & irrational) are influenced by changes in the beliefs of the states of the season variable. The first graph in figure 13 shows the general trend of pedestrian behavior during summer season. It can be inferred from the graph that as belief about summer season increases, the probability that pedestrians exhibit rational crossing behavior increases linearly as represented by a red line with a positive gradient of 0.168. However, the probability that pedestrians behave irrationally is rather decreasing linearly with an increase in belief that it is summer season. This is represented on the graph by a blue line and shows a negative gradient of 0.168. It is also important to note that the two graphs intersect at a summer belief point of 35% and crossing behavior belief of 50%. Pedestrian crossing behavior at this point is believed to be in equilibrium and it is known to be the point at which one is unable to predict whether a pedestrian is going to behave rationally or irrationally when crossing the roadway.

On the other hand, the second graph of the same figure 13 shows a sensitivity result for spring season that turns out to be an antithesis to the summer season sensitivity. It can be concluded from the graph that as belief in spring season increases with an assumption of arrival of more evidence, there is a substantial increase in probability that a pedestrian would behave more irrationally and a decrease in probability that their crossing behavior would rather be rational. In view of this however, the irrational crossing behavior line yields a positive gradient of 0.167 while the rational crossing behavior line produces a negative gradient of 0.168. The two lines also intersect at an important equilibrium point of spring belief value (65%) and crossing behavior value (50%). At this crucial intersecting point it becomes virtually



impossible to predict whether pedestrian crossing behavior would be rational or irrational.

### ***5.3.2 Pedestrian Signal Length Sensitivity***

The two states of the pedestrian signal cycle variable were also varied alternatively in order to ascertain the magnitude of their influence on the target variable (peds. crossing behavior). Initially, a 50% degree of belief for each of the two states of the signal phase variable was used as a base case scenario and their subsequent combined influence with other variables on the crossing behavior variable showed that 52.6% of the time pedestrians behave rationally in crossing a roadway intersection and 47.4% of the time they behave otherwise.

However, an initial assumption of 0% and then 100% degree of belief for each of the two states of the season variable was used alternatively and their corresponding effect on pedestrian crossing behavior was closely observed. The outcome of this sensitivity is depicted in figure 14 below.

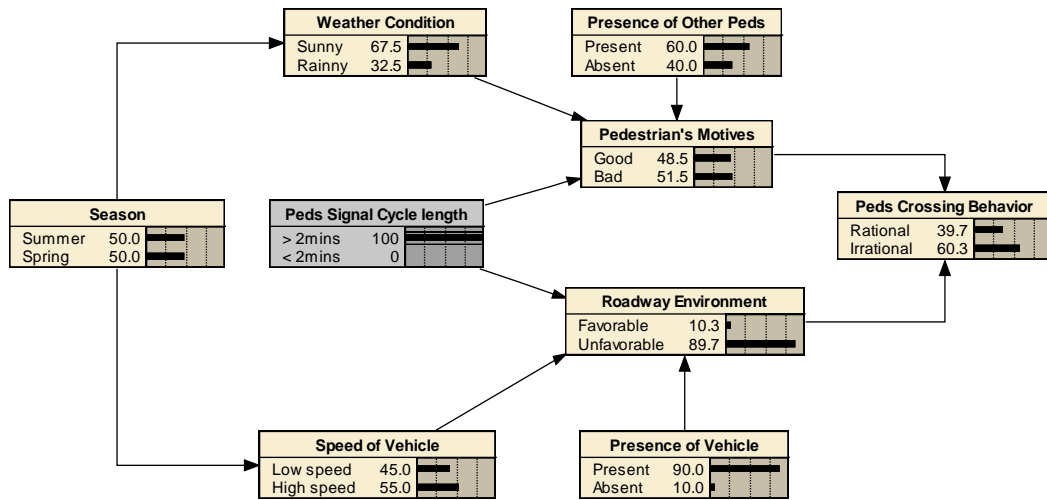


Figure 14: Bayesian network model showing the results of varying the states of pedestrian signal cycle node (> 2mins = 100%; < 2mins = 0%).

Analyzing the effect on pedestrian crossing behavior variable, it can be inferred from figure 14 that pedestrian crossing behavior will be rational given signal length of greater than 2 minutes showed a substantial decrease from 52.6% to 39.79%, (% decrease = 24.5). Conversely, the belief that pedestrian will exhibit irrational behavior given signal length greater than 2 minutes rather increased significantly from 47.4% to 60.3 with a percentage increase of 27.2%.

The result is indeed a clear manifestation of reality and truly meaningful as pedestrian motive gets bad when they begin to have the impression that a particular roadway design characteristics is more favorable to vehicles than it is to them. They however become impatient, intolerant and eventually disregard traffic laws by behaving more irrationally and often carelessly. The sensitivity for the signal cycle

length of less than 2 minutes on the other hand shows a rather opposite story as depicted in figure 15 below.

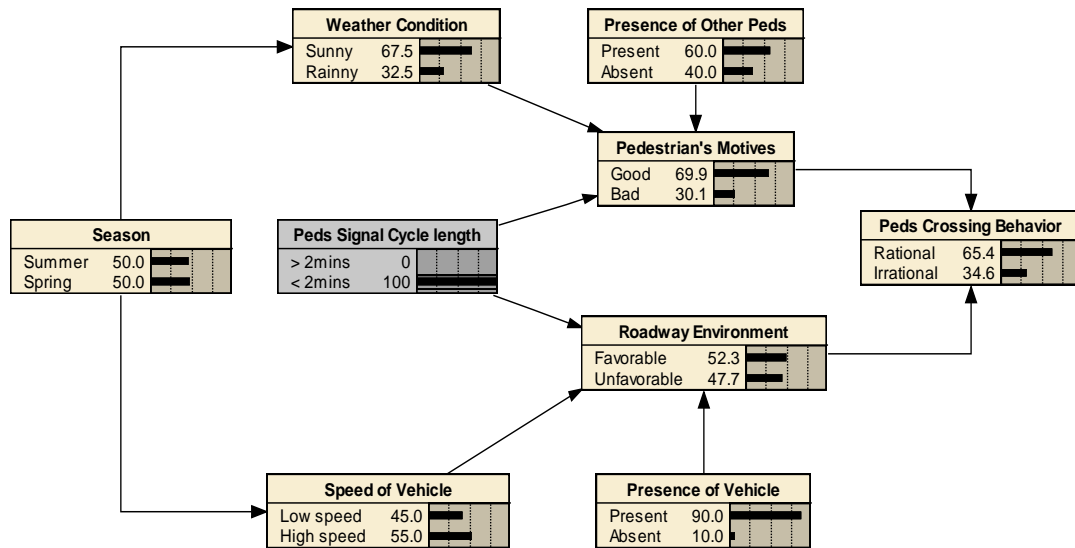


Figure 15: Bayesian network model showing the results of varying the states of the peds. signal cycle node (< 2mins = 100%; > 2mins = 0%).

Assuming pedestrian signal length for all the study sites to be less than 2 minutes (signals < 2 mins. = 100%), again the model shows a significant change in all the nodes that are connected to the signal cycle (parent) variable. The consequence of the sensitivity on pedestrian crossing behavior shows substantial increase in belief from 52.6% to 65.4% that pedestrians behave more rationally given all signal cycle length of less than 2 minutes (percentage increase = 24.3%). Conversely, the belief

that pedestrian put up irrational behavior assuming signal cycle length of all the sites were less than 2 minutes significantly decreased from 47.4% to 34.6% by a percentage decrease of approximately 27%.

The result from this sensitivity is realistically meaningful in the sense that the less the waiting time for pedestrian about to cross an intersection, the more forbearance and composed they would be. Furthermore, the more favorable and friendly a roadway environment is to pedestrians, the more patient and tolerant they get. Consequently, pedestrian motive and conscience prior to crossing an intersection with such characteristics will be good and eventually make them behave more rationally and extremely scrupulous.

Further sensitivity analysis was carried out by varying the belief of the states of the signal cycle length variable over equal interval of 10 units and corresponding changes in the pedestrian crossing behavior variable was noted. A linear graph was plotted to ascertain the effect of varying the beliefs of the states of this variable on pedestrian crossing behavior over a constant interval.

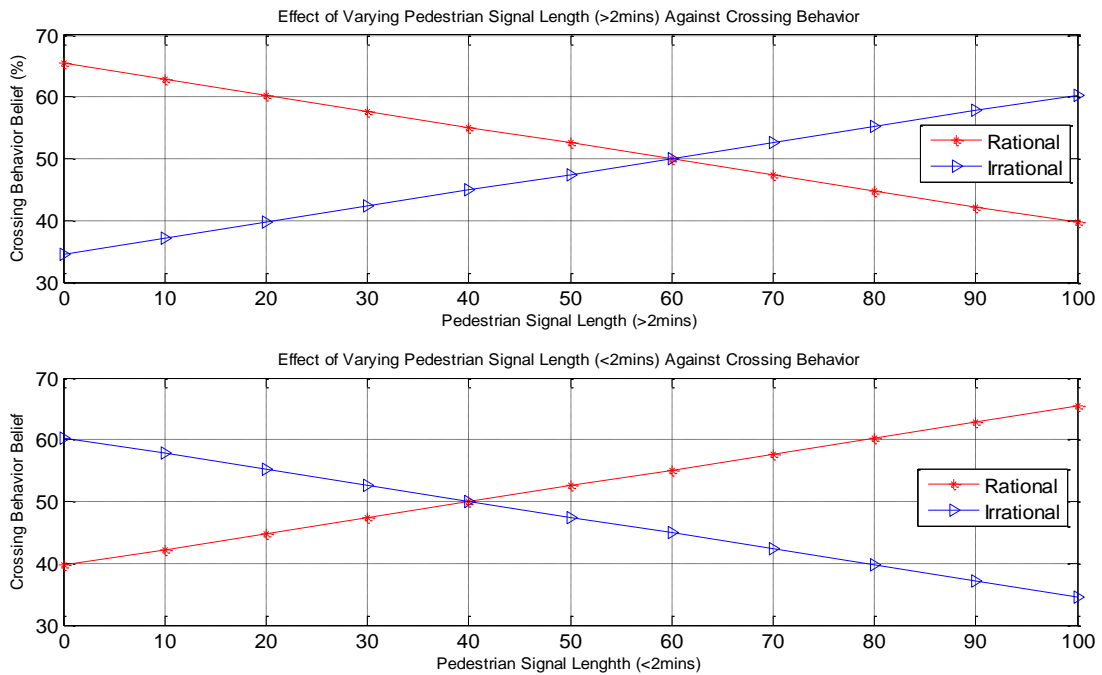


Figure 16: Sensivity plots for pedestrian signals.

The two sensitivity graphs as depicted in figure 16 are direct opposite of each other as they both display how the two states of pedestrian crossing behavior (rational & irrational) are influenced by changes in the beliefs of the states of signal cycle length variable. The first graph in figure 16 shows the general trend of pedestrian behavior relative to signal length being greater than 2 minutes. It can be concluded from the graph that as belief about signal cycle lengths of all the study sites being greater than 2 minutes increases, the probability that pedestrians exhibit rational crossing behavior decreases linearly as represented by a red line with a negative gradient of 0.258. However, the belief that pedestrian behave irrationally rather increases linearly

with an increase in belief that signal cycle length is greater than 2 minutes. This irrational behavior is represented on the graph by a blue line and shows a positive gradient of 0.258. An equilibrium point is reached according to the graph where belief that signal length greater than two minutes is at point 60% and crossing behavior belief point of 50%. This is the critical point at which one is unable to predict with absolute certainty that pedestrian would behave rationally or otherwise when crossing a roadway.

On the other hand, the second graph of the same figure 16 is illustrating a sensitivity result for pedestrian signal cycle length of not more than 2 minutes and this shows a result that is an antithesis to the cycle length greater than 2 minutes sensitivity. It can be concluded from the graph that as belief in cycle length up to 2 minutes increases with an assumption of arrival of some evidence, there is a substantial increase in probability that pedestrian would behave more rationally and a decrease in probability that their crossing behavior would rather be irrational. As a result of this, the rational crossing behavior line shows a positive gradient of 0.257, while the irrational crossing behavior line shows a negative gradient of 0.259. Moreover, the two lines intersect at a point where the belief that signal cycle length is up to 2 minutes is at the 40% point and crossing behavior belief of 50%. This is a very crucial point on the graph where the two lines are said to be in equilibrium at that point of intersection and is believed to be the point where it becomes virtually impossible to predict whether pedestrian crossing behavior would be rational or irrational.

### 5.3.3 Pedestrian Crossing Behavior Sensitivity

In order to be able to ascertain which of the two seasons we often observe more rational and/or irrational pedestrian crossing behavior, a sensitivity analysis was carried out on pedestrian crossing variable in the model. This was done by alternatively assuming 100% belief for each of the two states of the crossing behavior variables and then later used an increment of 10 units for each of the two states. Figure 17 below displays the result of the sensitivity for assuming 100% belief that pedestrians behave rationally.

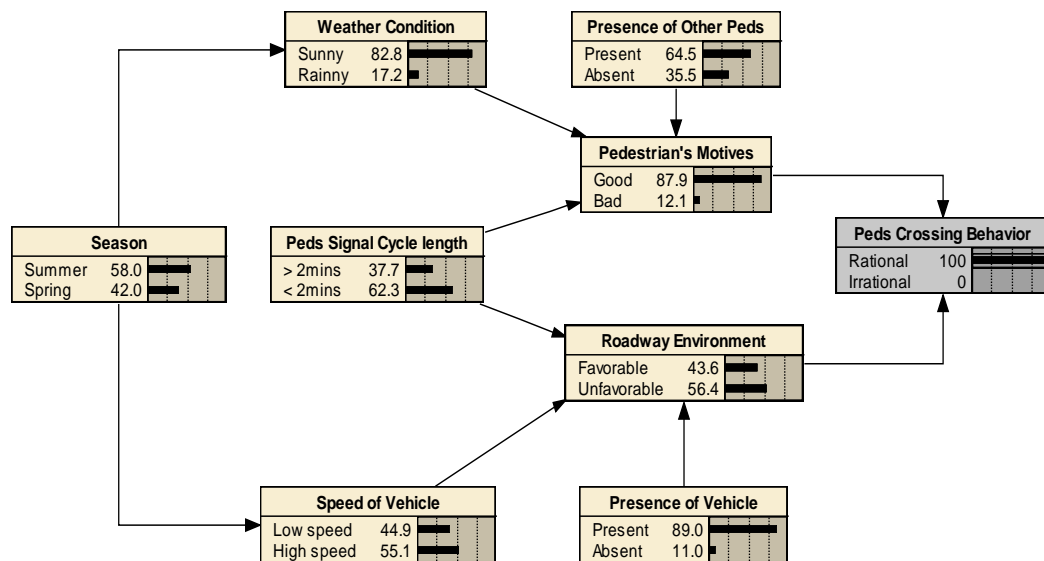


Figure 17: Bayesian network model showing the results of varying the states of the pedestrian crossing behavior node (rational = 100%).

By assuming a 100% evidence for rational behavior, the model's results shows that 58% of the time such attitude or crossing behavior put up by pedestrians is mostly exhibited during the summer season and 42% of the time such behavior is exhibited during the spring season. The result obtained however, can be attributed to so many contributing factors such as higher probability of pedestrian motive being good even though roadway environment may not have been all that favorable, high probability of other pedestrian always patiently waiting for their legal turn to cross the road, and high probability of good weather condition because of high chance of sunny days.

Also, because the model shows a high chance that a lot of vehicles would be on the roadway during this time with potentially high chance of overspeeding and also because pedestrian signal cycle length has a high probability to be less than 2 minutes according to the model's results, the final result that summer season shows a higher degree of belief given pedestrian makes rational crossing decision is indeed a true reflection of reality and makes a lot of sense. On the contrary, the alternative results obtained by assuming 100% evidence for an irrational crossing behavior is directly opposite to the former according to model as depicted graphically below.



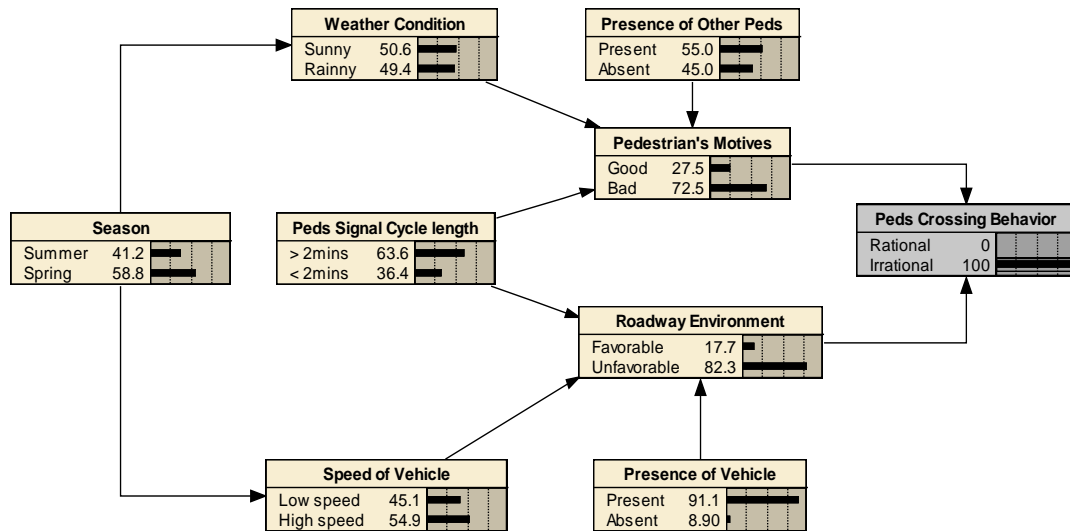


Figure 18: Bayesian network model showing the results of varying the states of pedestrian crossing behavior node (irrational = 100%).

The sensitivity result as shown in figure 18 proves the fact that by assuming a 100% evidence that pedestrian crossing behavior would be irrational, the model displays a drastic decrease from 50% to 41.2 % in belief to confirm our hypothesis that such irrational behavior is attributed to summer season while it rather shows a significant increase in belief from 50% to 58.8% to establish the fact that an irrational behavior of this sort is more evident and attributable to spring season.

Again, the result is logically convincing and reflective of reality as supported by attributes that is explained by the intermediate nodes in the model. Given that pedestrians exhibit irrational crossing behavior, it can be inferred from the model that there is a high conditional probability that pedestrian motive would be bad and

roadway environment would be unfavourable because of high chance of presence of vehicles and long waiting time. The result of this sensitivity is further explained by varying the state of the pedestrian crossing behavior variables over an equal interval of 10 units to help throw more lights on the effects of the season variable. The following plots show the results.

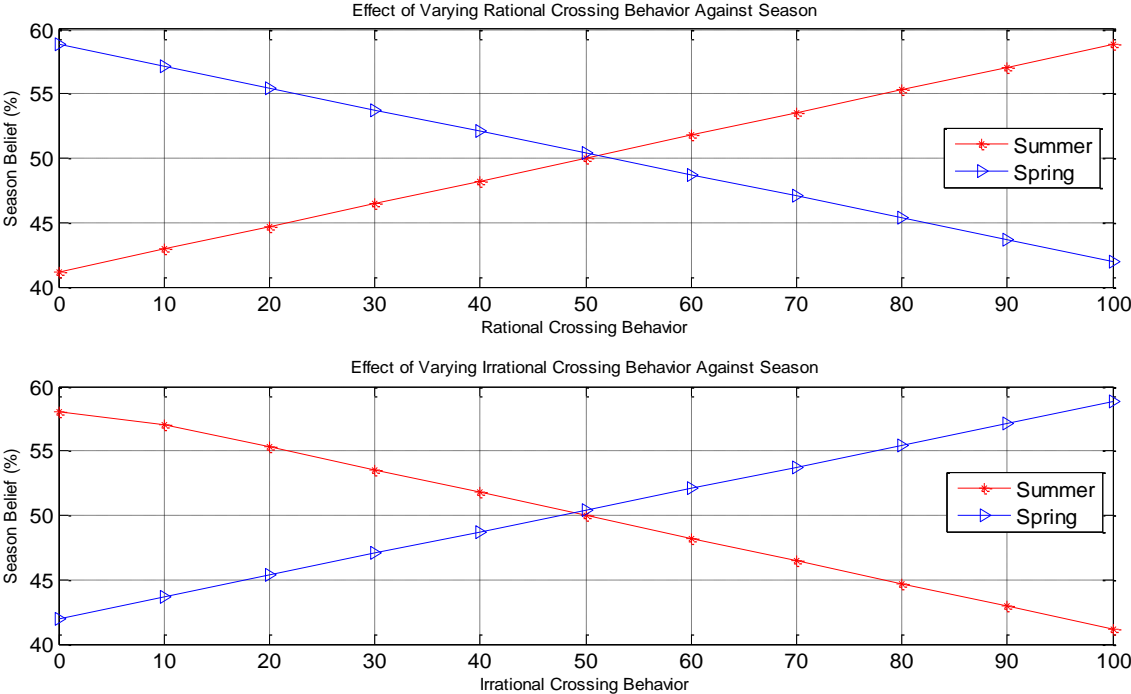


Figure 19: Sensitivity plots for pedestrian crossing behavior.

Again the two sensitivity graphs show directly opposite results of varying the states of pedestrian crossing behavior (rational & irrational) and their effects on

each of the states of the season variable. The first graph in figure 19 specifically illustrates the effects of varying the belief of rational crossing behavior between 0% and 100% for every 10 units increment in evidence. It can however be inferred from the graph that as the belief about pedestrian acting rationally increases, the probability that such behavior is attributable to summer season increases linearly as represented by a red line with a positive gradient of 0.176. Meanwhile, the belief that such behavior is related to spring season is rather decreasing linearly as represented on the graph by a blue line with a negative gradient of 0.168. The two lines also intersect at a crucial equilibrium point on the graph where it becomes apparently impossible to predict which of the two seasons would pedestrian crossing behavior be more rational or irrational. This critical point is marked on the graph at the intersection of rational behavior belief point of 50% and season belief point of 50%.

On the other hand, the second graph of the same figure 19 is illustrating a sensitivity result of varying the belief of irrational crossing behavior over the same range for every 10 unit increment in evidence. It can be concluded from the graph that as the belief about pedestrian acting irrationally increases, the corresponding probability that such behavior is evident in spring season increases linearly as represented on the graph by a blue line with a positive gradient of 0.168. Contrarily, the belief that such behavior is a characteristic of summer season is linearly decreasing as depicted on the graph by a red line and a negative gradient of 0.176. Moreover, the two lines on the graph have a common point of intersection which is marked by irrational crossing behavior point of 50% and season belief point of 50%. This crucial intersecting point is what is described as the point of equilibrium where it becomes

apparently impossible to predict which of the two season pedestrians would behave more rationally or otherwise.

Sensitivity analysis was also carried out for the remaining two independent variables of the model such as *presence of others* and *presence of vehicles* just to understand their behavior on pedestrian's crossing behavior. The graphs illustrating their results are included in the Appendix II even though this study ignores the analysis of their impact because of the insignificance of their influence on pedestrian crossing behavior.

## **Chapter 6**

### **CONCLUSION AND RECOMMENDATION**

#### **6.1 General**

This study has explored the capabilities of using the Bayesian network to formulate and analyze key factors that influence pedestrian behavior in relation to crossing a roadway. Although myriads of studies of this sort has been carried out in the past using various forms of approaches including computer simulation, statistical modeling, etc, the prowess of Bayesian network approach was not being exploited by existing research to clearly capture the causal relationship leading to certain behavior. A review of the existing dominating processes and methods for pedestrian safety and behavioral analysis was used as a background for introducing the efficacy of the Bayesian network modeling approach.

#### **6.2 Conclusion**

Owing to the analysis of the results presented in this thesis, the following conclusions and recommendations can justifiably be drawn.

The Bayesian network approach for modeling (formulating & analyzing) pedestrian behavior is highly innovative and well-suited for such a behavioral studies because it clearly provides an insight into conditions that significantly influence pedestrian crossing decision. A trend in pedestrian crossing behavior has been revealed

using the Bayesian network modeling technique and useful information extracted as well about why pedestrian behave the way they do when it comes to making an intersection crossing decision. Nonetheless, although the Bayesian network modeling approach used for this study may not be ideally precise, the output of the model when compared with that of reality was found to be highly accurate as confirmed by an acceptable computed root mean square error of 0.02m. It can however be concluded from this study that pedestrian crossing behavior which is believed to be a random process is greatly influenced by the combined effects of both human and environmental factors. Obtaining an empirically reasonable trend which realistically characterizes information pertaining to behavioral pattern requires that the effect of other relevant influential parameters on pedestrian motive and roadway environment showing causal relationship be identified and a general influential trend accurately analyzed.

The models results however reveal that rational crossing behavior outstrips irrational crossing behavior in terms of belief in general pedestrian crossing behavior which is contingent on pedestrian motive and roadway environment. An extended development that was discovered from this study was that, the roadway environment variable didn't have much influence compared to pedestrian motive variable although it shows a high degree of belief that it is always unfavourable to pedestrians about to cross the roadway. Instead, pedestrian motive is found to have a relatively greater influence on pedestrian making a rational or irrational crossing decision. Thus, the ability to be able to predict pedestrian crossing behavior is more dependent on one's previous knowledge about whether pedestrian motive prior to crossing the roadway is

good or bad as opposed to whether roadway environment is generally favorable or otherwise.

Moreover, the study underscores the fact that majority of pedestrians always behave more rationally than they do irrationally because of their cognizance about the potentially high risk associated with crossing a roadway. Such general scrupulous behavior exhibited by pedestrians was found to be less dependent on type of season (summer or spring) as it turned out that majority of pedestrians were law abiding people who are more concerned about their safety. A sensitivity analysis became imperative to enable us obtain a more physically meaningful information from some selected variables of interest which are believed to have a significant impact on pedestrian crossing behavior. The presented sensitivity analysis revealed that pedestrian crossing behavior is mostly influenced by the length of pedestrian signal timing phase. Where pedestrian are made to wait a little longer for their turn to legally cross an intersection, they begin to grow impatient and look for the least possible opportunity to act against the law.

The next most influential variable that came up from the sensitivity analysis was the season variable and it showed that pedestrian always behaved more rationally during the summer and otherwise during the spring season. By carrying out a sensitivity analysis on the pedestrian crossing behavior, it was useful and physically convincing to confirm our finding that rational crossing behavior is mostly observed during summer season whiles irrational crossing behavior is also mostly observed during spring season.

### **6.3 Recommendation and Future Research**

- The implication and usefulness of Bayesian network for modeling behavior at different analytical levels ought to be investigated further to increase the level of accuracy and precision in estimating states conditional probabilities for a given variable.
- The number of study area in subsequent study should be increased in order to obtain a more general normal distribution of the population that will guarantee a higher accuracy and precision in terms of results to be generated from the model.
- Further research should be targeted at investigating how pedestrian crossing behavior varies over the four different seasons (winter, spring, summer & fall) instead of just the two seasons considered in this study because of time constraint.



## REFERENCES

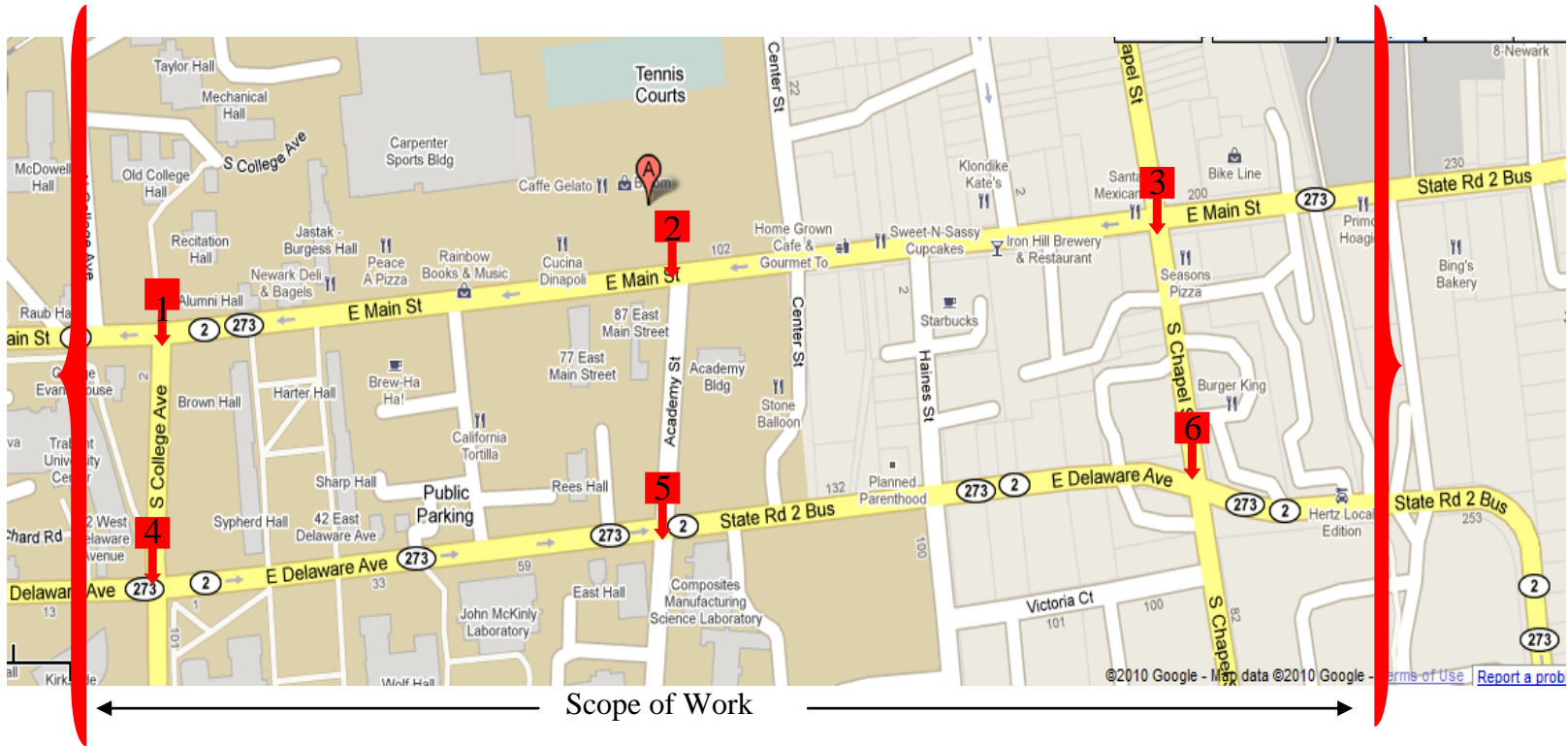
- [1] Anowar, S., Yasmin, S., Tay, R. (2010). "A Logistic Model for Examining Pedestrian Serious Injury Risk in Traffic Crashes." *Proceedings of the 20<sup>th</sup> Canadian Multidisciplinary Road Safety Conference*, Niagara Falls, Ontario.
- [2] AUSTRROADS (1996). "Urban Speed Management in Australia." AP 118. Austroads, Sydney.
- [3] Cain, J. (2001), "Planning improvements in natural resources management: Guidelines for using Bayesian networks to support the planning and management of development programs in the water sector and beyond, *Centre for Ecology and Hydrology*, Wallingford, UK.. Retrieved June 22, 2010 from [www.nwl.ac.uk/ih/www/products/bbnguidelines.html](http://www.nwl.ac.uk/ih/www/products/bbnguidelines.html)
- [4] Cynecki, M.J., Sparks, J.W., Grote, J.L. (1993). "Rumble Strips and Pedestrian Safety." *Institute of Transportation Engineers, ITE Journal*; Vol. 63(8); pp. 18-24.
- [5] Das, B. (1999). "Representing Uncertainties Using Bayesian Networks." *DSTO Electronics and Surveillance Research Laboratory*, DSTO-TR-0918, Salisbury, South Australia.
- [6] Das, S., Manski, C.F., Manuszak, M.D. (2005). "Walk or Wait? An Empirical Analysis of Street Crossing Decision." *Journal of Applied Econometrics, J. Appl. Econ.*; Vol. 20(4); pp. 529-548.
- [7] De Lavalette, B.C., Tijus, C., Poitrenaud, S., Leproux, C., Bergeron, J., Thouez, J.P (2009). "Pedestrian Crossing Decision: A Situational and Behavioral Approach." *Safety Science*, Vol. 47 (11); pp. 1248-1253.
- [8] FARS [Fatality Analysis Reporting System]. (2009). FARS encyclopedia. Retrieved June 15, 2010, from <http://www-fars.nhtsa.dot.gov>.

- [9] Fitzpatrick, K., Turner, S., Brewer, M. (2007). "Improving Pedestrian Safety at Unsignalized Intersections." *Institute of Transportation Engineers, ITE Journal*; Vol. 77(5); pp. 34-41
- [10] Harrell, W.A. (1991). "Factors Influencing Pedestrian Cautiousness in Crossing Streets." *The Journal of Social Psychology*, 131(3); 367-372.
- [11] Heckerman, D. (1996). "A Tutorial on Learning with Bayesian Networks." *Technical Report*, MSR-TR-95-06.
- [12] Heckerman, D., Geiger, D. & Chickering, D. M. (1995), 'Learning Bayesian networks: The combination of knowledge and statistical data', *Machine Learning*. Vol. 20(3); pp. 197–243.
- [13] Hubbard, S.M.L., Bullock, Darcy M., Thai, J.H. (2008). "Trial Implementation of a Leading Pedestrian Interval: Lessons Learned." *Institute of Transportation Engineers, ITE Journal*; Vol. 78(10); pp. 32
- [14] Jones, T.L., Tomcheck, P. (2000). "Pedestrian Accident in Marked and Unmarked Crosswalks: A Quantitative Study." *Institute of Transportation Engineers, ITE Journal*, Vol. 70(9); pp. 42.
- [15] Li, Y., Fernie, G. (2010). "Pedestrian Behavior and Safety on A Two-stage Crossing With A Center Refuge Island and The Effect of Winter Weather on Pedestrian Compliance Rate." *Accident Analysis and Prevention*, 42(4); 1156-63.
- [16] Martin, A. (2006). "Factors Influencing Pedestrian Safety: A Literature Review." *Unpublished Project Report*, UPR SE/199/05.
- [17] O'Mara, P.J. (2000). "Queens Boulevard Pedestrian Safety Study." *Institute of Transportation Engineers, ITE Journal*; Vol. 70(10); pp. 39.
- [18] Oxley, J., Corben, B., Fildes, B., O'Hare, M. (2004). "Older Vulnerable Road Users – Measures To Reduce Crash And Injury Risk." *Monash University Accident Research Center*, Clayton, Victoria.

- [19] Oxley, J.A., Ihsen, E., Fildes, B.N., Charlton, J.L., Day, R.H. (2005). "Crossing Roads Safely: An Experimental Study of Age Differences in Gap Selection by Pedestrian." *Accident Analysis and Prevention*, Vol. 37(5); pp. 962-971.
- [20] Redmon, T. (2003). "Assessing the Attitude and Behaviors of Pedestrians and Drivers in Traffic Situations." *Institute of Transportation Engineers, ITE Journal*; Vol. 73(4); pp. 26.
- [21] Retting, R.A., Nitzburg, M.S., Farmer, C.M., Knoblauch, R.L. (2002). "Field Evaluation of Two Methods for Restricting Right Turn on Red to Promote Safety." *Institute of Transportation Engineers, ITE Journal*; Vol. 72(1); pp. 32.
- [22] Retting, R.A., Van Houten, R., Malenfant, L., Van Houten, J., Farmer, C.M. (1996). "Special Signs and Pavement Markings Improve Pedestrian Safety." *Institute of Transportation Engineers, ITE Journal*; Vol. 66(12); pp. 28.
- [23] Sze., N.N., Wong, S.C. (2007). "Diagnostic Analysis of the Logistic Model for Pedestrian Injury Severity in Traffic Crashes." *Accident Analysis and Prevention*, Vol. 39(6); pp. 1267-1278.
- [24] Van Houten, R., Retting, R.A., Van Houten, J., Farmer, C.M., Malenfant., L. J.E. (1999). "Use of Animation in LED Pedestrian Signals To Improve Pedestrian Safety." *Institute of Transportation Engineers, ITE Journal*; Vol. 69(2); pp. 30.
- [25] Wu, F., Sun, P.R., Ke, H.K., Huang, H.H., Chiang, Y. (2009). "A Study of Pre-Decision Evaluation Using Influence Diagram: An Estimation of The Benefits of Influenza Vaccination." *Journal of Medical Systems*, Vol. 33(1); pp. 19-25.
- [26] Yagil, D. (2000). "Beliefs, Motives and Situational Factors Related to Pedestrians' Self-reported Behavior at Signal-controlled Crossing." *Transportation Research Part F*, Vol. 3, pp. 1-13.
- [28] Yannis, G., Kanellaidis, G., Dimitropoulos, J., Muhrad, N. (2007). "Assessment of Pedestrian Safety Measures in Europe." *Institute of Transportation Engineers, ITE Journal*; Vol. 77(12); pp. 40-47.

[27] Zegeer, C.V., Esse, C.T., Stewart, J.R., Huang, H.F., Lagerwey, P. (2004). "Safety Analysis of Marked Versus Unmarked Crosswalks in 30 Cities." *Institute of Transportation Engineers, ITE Journal*; Vol. 74(1); pp. 34-41.

**APPENDIX I**  
**LOCATION AND PEDESTRIAN DATA COLLECTION**



Legend

 Location of Data collection Site

Data Collection Sheet Sample

Site: \_\_\_\_\_ Date: \_\_\_\_\_ Start: \_\_\_\_\_ Stop: \_\_\_\_\_

Crossing Behavior	<u>Spring Season</u>		<u>Summer Season</u>	
	Sunny	Rainy	Sunny	Rainy
Crossing against stop signal (Jaywalking)				
Waited for legal turn because of influence of others presence				
Crossed legally without the influence of other's presence				
<b>Remarks:</b>				

Field Data Collection Sample

Site: F. Main & Academy Date: 4/19 8/31/10 Start: 10:00 am Stop: 12:50 pm

Crossing Behavior	<u>Spring Season</u>		<u>Summer Season</u>	
	Sunny	Rainy	Sunny	Rainy
Crossing against stop signal (Jaywalking)	IIII IIII IIII III = 18	IIII IIII IIII IIII IIII IIII IIII IIII II = 42	IIII IIII IIII = 14	IIII IIII IIII = 9
Waited for legal turn because of influence of others presence	IIII IIII IIII = 15	IIII IIII = 3	IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII = 53	IIII IIII IIII = 7
Crossed legally without the influence of other's presence	IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII = 62	IIII IIII IIII = 13	IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII = 92	IIII IIII IIII = 4
<b>Remarks:</b> <u>huge volum of cars observed during Summer Season than Spring Season.</u>				

Location 1: East Main Street & South College Avenue



Data Collection Spot

Location 2: East Main Street & Academy Street



Data Collection Spot

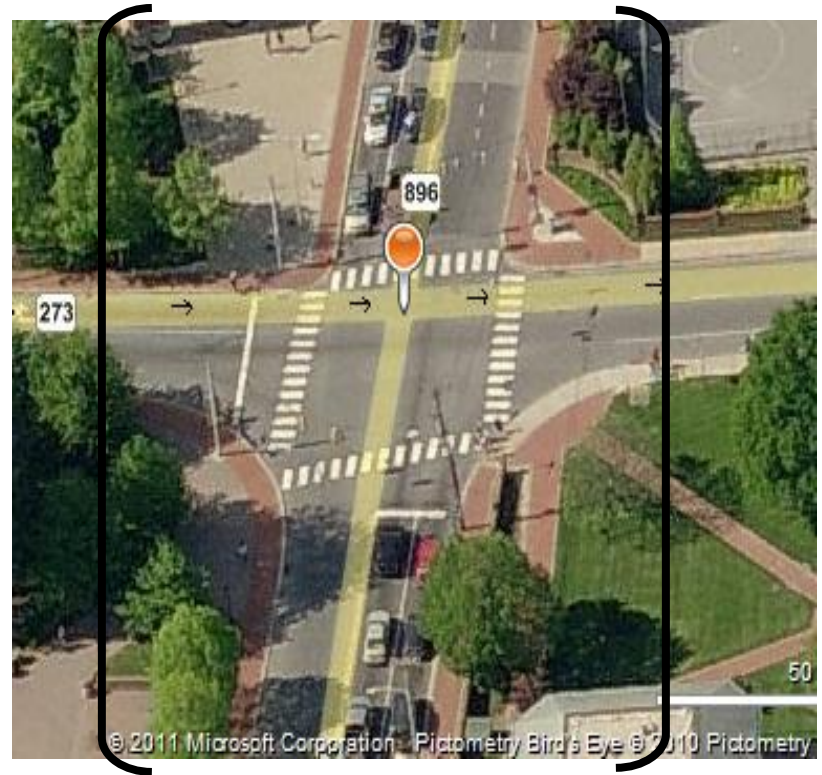


Location 3: East Main Street & North/South Chapel Avenue



Data Collection Spot

Location 4: Delaware Avenue & South College Avenue



Data Collection Spot

Location 5: Delaware Avenue & Academy Street



Data Collection Spot

Location 6: Delaware Avenue & South Chapel Street



Data Collection Spot

## APPENDIX II

### COMPUTATIONS AND PRELIMINARY ANALYSIS OF DATA

Peds. Behavior	Spring Season	Summer Season
Crossing illegally	280	300
Waited because of others presence	20	420
Crossed legally without influence	100	480
<b>Total</b>	<b>400</b>	<b>1200</b>

#### Analysis of Variance (ANOVA)

Source	Deg. of Freedom	Sum of Squares	Mean Square	F
Regression	$p - 1$	$SS_{reg}$	$SS_{reg}/p - 1$	F
Residual	$n - p$	RSS	$RSS/(n - p)$	
Total	$n - 1$	SYY		

Where

$$F = \frac{(SYY - RSS)/(p - 1)}{RSS/(n - p)}$$

RSS - Residual Sum of Squares

$SS_{reg}$  - Regression Sum of Squares

$SSY$  – Total of Residual and Regression Sum of Squares

$p$  – probability values (p-value)

$n$  – number of variables (observation)

$F_{p-1, n-p}$  are referred to for critical value or a p-value. Large values of  $F$  would indicate rejection of the null. This  $F$  test is widely used in regression and analysis of variance. Information in the above test is traditionally presented in an *analysis of variance* table and most computer packages produce a variant on this.

ANOVA results computed in R programming environment

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Ind	1	106667	106667	8.1633	<b>0.04606</b>
Residuals	4	52267	13067	-	-

The p-value of 0.04606 tells us that there is almost a **95.394%** probability that the between-sample variance is significantly larger than the within-sample variance; thus, we have strong evidence that there is a difference between the mean results for the two seasons.

### 95 % Confidence Interval

Confidence intervals provide an alternative way of expressing the uncertainty in our estimates. For the confidence intervals and regions that we will consider here, the following relationship holds. For a  $100(1 - \alpha)\%$  confidence region, any point that lies within the region represents a null hypothesis that would not be rejected at the  $100\alpha\%$  level while every point outside represents a null hypothesis that would be rejected.

As with testing, we start with the simultaneous regions. Some results from multivariate analysis show that

$$\frac{(\beta - \beta)^T X^T X (\beta - \beta)}{\sigma} \sim \chi_p^2$$

and

$$\frac{(n-p)\delta^2}{\sigma} \sim \chi_{n-p}^2$$

And these two quantities are independent. Hence

$$\frac{(\beta - \beta)^T X^T X (\beta - \beta)}{p\delta^2} \sim \frac{\chi_p^2/p}{\chi_{n-p}^2/(n-p)} \equiv F_{p,n-p}$$

So to form a  $100(1-\alpha)\%$  confidence region for  $\beta$ , take  $\beta$  such that

$$(\beta - \beta)^T X^T X (\beta - \beta) \leq p\delta^2 F_{p,n-p}^{(\alpha)}$$

These regions are ellipsoidally shaped and because these ellipsoids live in higher dimensions, they can not easily be visualized (Faraway, 2002). Alternatively, one could consider each parameter individually which leads to confidence interval that take the general form of

Estimates  $\pm$  critical value  $\times$  standard error of estimate

Or specifically in this case:

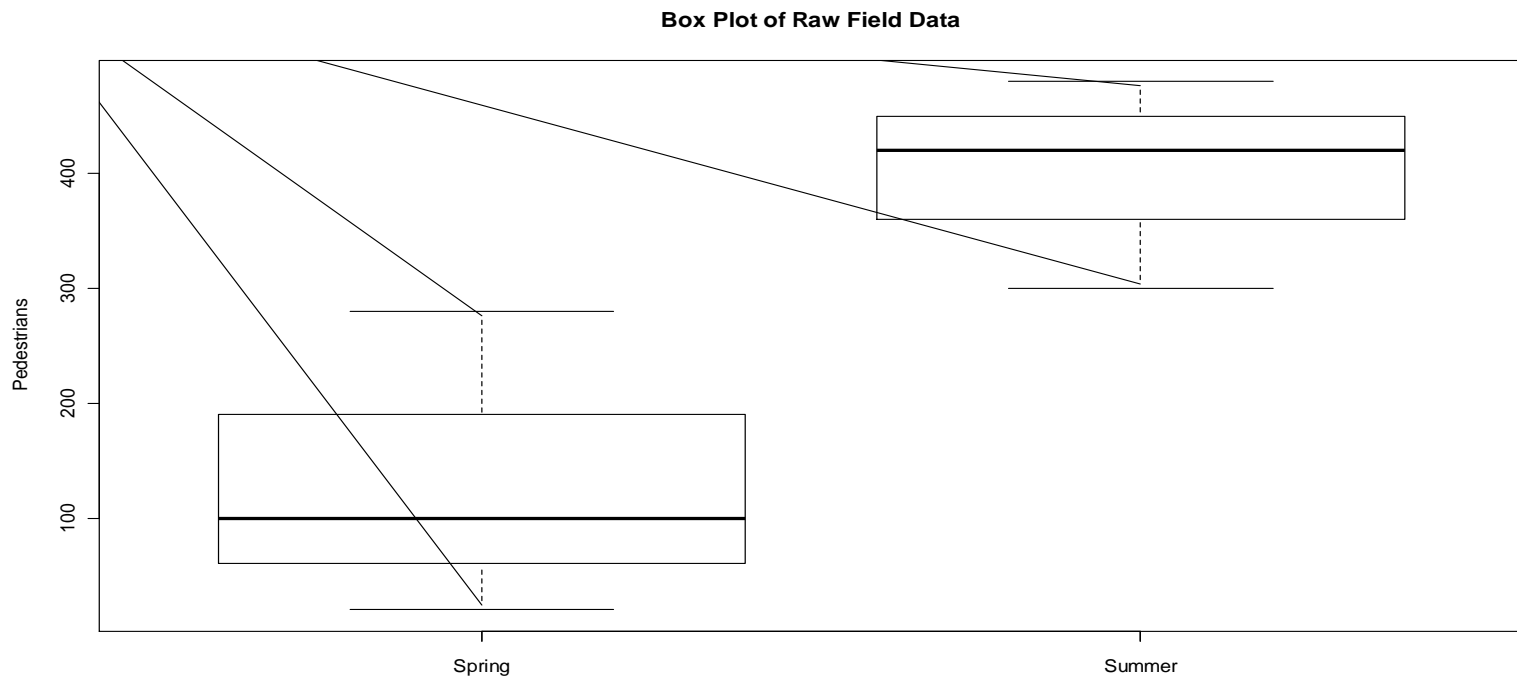
$$\beta_i \pm t_{n-p}^{(\alpha/2)} \delta \sqrt{(X^T X)^{-1}_{ii}}$$

Computed 95% Confidence Interval using R programming Language

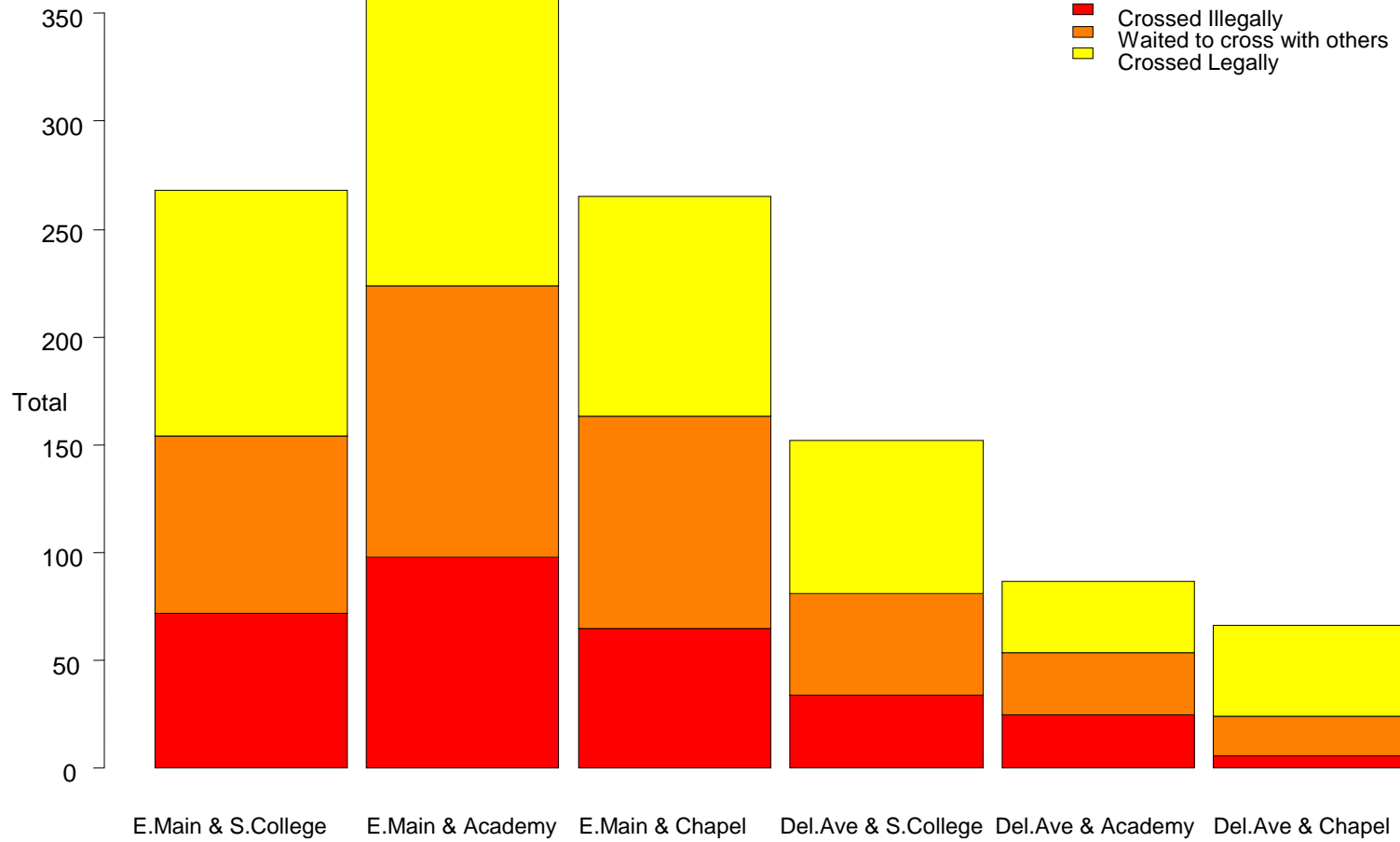
	Difference	Lower	Upper	p adj
Spring-summer	266.6667	7.530666	525.8027	<b>0.04606</b>

The table shows the actual differences between the mean values for the two seasons and the lower and upper boundaries for this difference. If the lower and upper boundaries includes a difference of zero (0), then there is no evidence for a significant difference between the means; if the range does not include zero, then the difference is significant. In this case we find that differences between spring and summer season does not include zero and so it is highly significant. The values of p adjustment (p adj) give the probability level at which the difference is significant; that is, the difference between spring and summer is significant at the **95.39%** confidence level.

## PLOTS OF RAW DATA AND SENSITIVITY

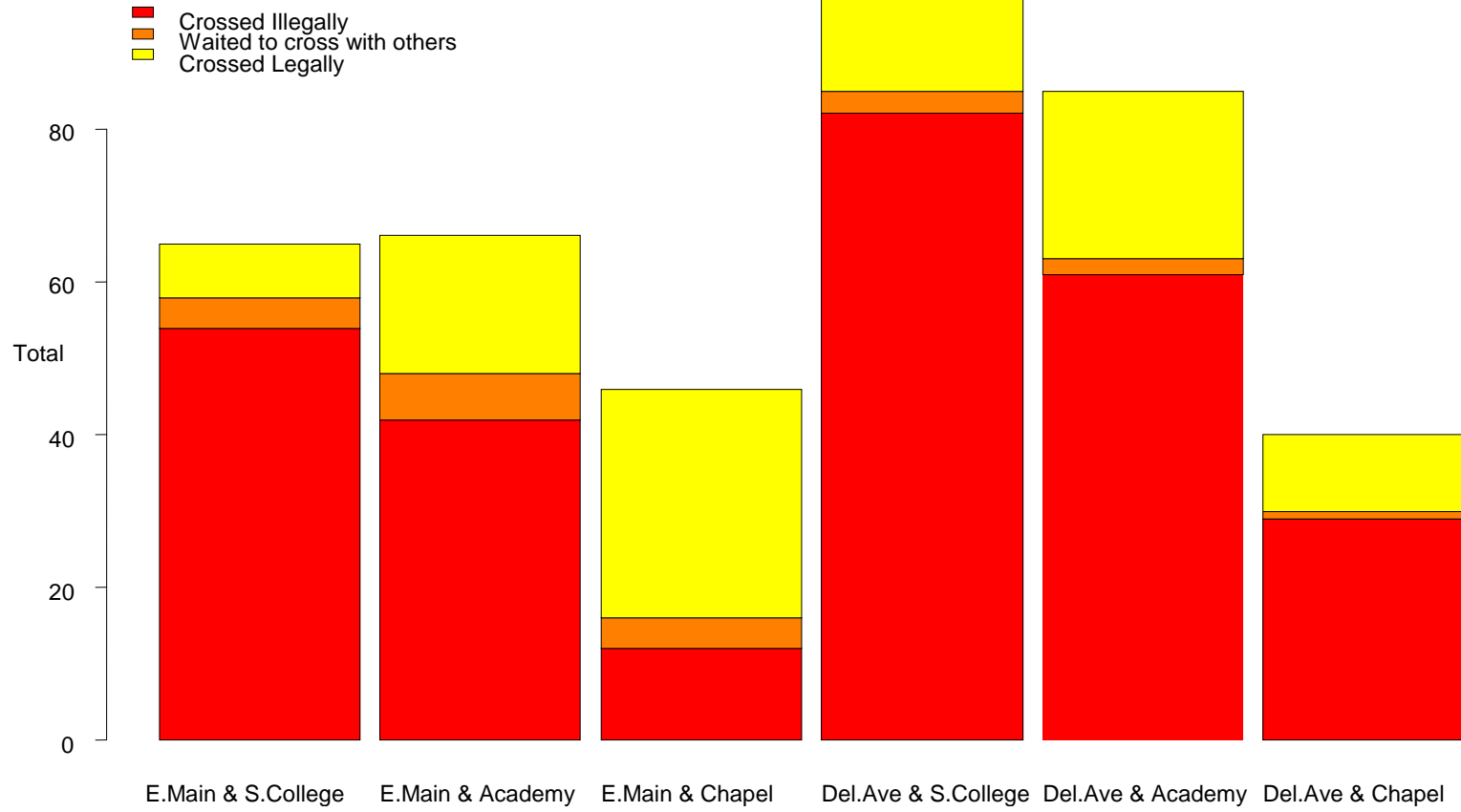


### Summer Season Crossing Behavior





### Spring Season Crossing Behavior



## Sensitivity Plots

