

2017-01-01

# Crash Rate as a Function of Access Point Density and Traffic Flow

George Pinal

*University of Texas at El Paso*, [georgepinal@yahoo.com](mailto:georgepinal@yahoo.com)

Follow this and additional works at: [https://digitalcommons.utep.edu/open\\_etd](https://digitalcommons.utep.edu/open_etd)



Part of the [Transportation Commons](#)

---

## Recommended Citation

Pinal, George, "Crash Rate as a Function of Access Point Density and Traffic Flow" (2017). *Open Access Theses & Dissertations*. 726.  
[https://digitalcommons.utep.edu/open\\_etd/726](https://digitalcommons.utep.edu/open_etd/726)

This is brought to you for free and open access by DigitalCommons@UTEP. It has been accepted for inclusion in Open Access Theses & Dissertations by an authorized administrator of DigitalCommons@UTEP. For more information, please contact [lweber@utep.edu](mailto:lweber@utep.edu).

# CRASH RATE AS A FUNCTION OF ACCESS POINT DENSITY AND TRAFFIC FLOW

GEORGE PINAL

Doctoral Program in Civil Engineering

APPROVED:

---

Ruey Long Cheu, Ph.D., Chair

---

Salvador Hernandez, Ph.D.

---

Soheil Nazarian, Ph.D.

---

Raed E. Al-Douri, Ph.D.

---

Luis Rene Contreras-Sapien, Ph.D

---

Charles H. Ambler, Ph.D.  
Dean of Graduate School

Copyright ©

by

George Pinal

2017

## **Dedication**

To my wonderful wife Sabine and children Alexander and Nicolás who have  
patiently supported me through all of my endeavors.

To my parents Alberto Pinal and Dolores Vizcaino who taught, encouraged and  
guided me to greater education and knowledge.

CRASH RATE AS A FUNCTION OF ACCESS POINT DENSITY AND TRAFFIC FLOW

by

GEORGE PINAL

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at El Paso

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

Department of Civil Engineering

THE UNIVERSITY OF TEXAS AT EL PASO

May 2017

## **Acknowledgements**

I would like to thank Dr. Kelvin Cheu, Dr. Salvador Hernandez , Dr. Soheil Nazarian, Dr. H.S. Oey, Dr. Luis Rene Contreras-Sapien and Dr. Raed Al-Douri for their technical recommendations, guidance, motivation and their highly valuable support throughout my studies.

My sincere gratitude to Dr. Carlos M. Chang and Dr. Lawrence E. Murr who assisted in the development and structure of this dissertation.

I wish to extend a special appreciation to the Texas Department of Transportation whose Crash Reporting and Analysis for Safer Highways system was instrumental for making this study possible.

I am especially thankful to the City of El Paso for their support of my studies. It was through this support that my endeavors were possible to enhance the livelihood and safety of our society. I want to thank Sergio Reyes, P.E. and Johanes Makahaube, P.E. at the City of El Paso for their guidance, knowledge and recommendations.

I want to thank wonderful wife Sabine for her patience, encouragement and tremendous assistance for the family. I am forever thankful to my father, Alberto, for his persistence to making great achievements in the world of mathematics and sciences and my mother, Dolores, for her encouragement and support throughout my endeavors and studies.

## **Abstract**

High crash rates on U.S. highways results in a considerable loss of human and economic resources. It is estimated that the U.S. loses about \$10 billion every year in motor vehicle crashes in terms of fatalities, injuries, and property damage. Demand for access points from highways to properties has steadily increased due to rapid growth in local economies creating thereby compromises between accessibility and mobility or capacity and safety. On the other hand, the availability of access is essential to residential and commercial developments and often occurs at the cost of safety and traffic operations. This often directs the need to negotiate a compromise between accessibility, mobility and safety. This research establishes a methodology that examines the best possible spacing between access points and geometric roadway factors that improves traffic flow and reduces traffic crash rates. Crash prediction models such as Tobit regression, exponential regression, Poisson, negative exponential model and cluster analysis, are assessed to look at easier access to entrances and exits into highways. These predictive modeling tools assess the factors causing crashes as well as where to target and prioritize future projects in terms of crash likelihood. Geographical Positioning Systems (GPS) enable a greater availability to acquire crash data such as the Crash Records Information System (CRIS) developed by the Texas Department of Transportation (TXDOT). The econometric methodology developed in this study shows that crash rates have a direct relationship with access density and flow. The model in this analysis provides a better understanding of the physical characteristics leading to the increase in crashes.

## Table of Contents

Acknowledgements.....	v
Abstract .....	vi
List of Tables .....	ix
List of Figures .....	x
1 Introduction.....	1
1.1 Background and Motivation.....	1
1.2 Research Objectives .....	2
2 Literature Review .....	3
2.1 Crash Prediction Models .....	3
2.1.1 Count Data Crash Models.....	3
2.1.2 Incident Clustering.....	7
2.1.3 Impact of Access driveway on Crash Rates on Multilane Highways .....	11
2.1.4 Incident Occurrence Models for Freeway Incident Management.....	13
2.1.5 Crash Prediction Models for Urban Unsignalized Intersections.....	15
2.1.6 Impact of Heavy Vehicles on Road Crashes.....	17
2.2 Crash Clustering and Hot Spot Analysis.....	19
2.2.1 Kernel Density Estimation (KDE) .....	19
2.2.2 Impact of Traffic on Road Crashes .....	26
2.3 Summary of Literature Review .....	29
3 Data.....	31
3.1 Multiple Vehicle Crashes.....	32
3.2 Model Variables .....	34
3.3 Police Crash Report Variables .....	38
3.4 Summary .....	39
4 Methodology.....	40
4.1 Intersection Distance to Crash Model .....	40
4.2 Regression Model Specifications .....	49
4.3 Model Goodness of Fit.....	53
4.4 Likelihood Ratio Test.....	54
4.5 Interpretation of Variables .....	54



5	Model Development .....	56
5.1	El Paso IH-10 Data – Tobit Model .....	56
5.2	San Antonio IH-410 Data – Tobit Model.....	57
5.3	El Paso IH-10 and San Antonio IH-410 Data Combined – Tobit Model .....	58
6	Model Results .....	59
6.1	Flow and Access Density Relationship.....	59
6.2	Roadway Characteristics .....	70
7	Conclusions.....	74
7.1	Research Summary.....	74
7.2	Research Limitations.....	75
7.3	Research Contributions .....	77
7.4	Further Research .....	77
	References.....	78
	Vita.....	81

## List of Tables

Table 3.1: Summary statistics of geometric and traffic variables.....	35
Table 6.1: Tobit estimation of crash rate on IH-10 and IH-410 .....	59
Table 6.2: Average Annual Daily Traffic (AADT) as a function of crash rate. ....	60
Table 6.3: Exponential regression estimation of access density on crash rate for IH-10 and IH-410. ....	60
Table 6.4: Exponential regression estimation on number of lanes on crash rate for IH-10 and IH-410. ....	61
Table 6.5: Tobit regression estimation of access density on crash rate for IH-10 and IH-410 ....	62
Table 6.6: Tobit regression estimation for number of lanes on crash rate for IH-10 and IH- 410.....	63
Table 6.7: Tobit regression estimation of access density and number of lanes on crash rate for IH-10 and IH-410. ....	65
Table 6.8: Tobit censored elasticity estimation of crash rate on IH-10 and IH-410.....	71

## List of Figures

Figure 2.1: Method for conducting critical incident analysis for a specific site (Hanowski et al., 2004). .....	8
Figure 2.2: Cluster Analysis (Hanowski et al., 2004).....	10
Figure 2.3: Crashes in upstream and downstream locations (Cullip et al., 1997). .....	13
Figure 2.4: Volume versus time of day (Cullip et al., 1997) .....	14
Figure 2.5: Occupancy versus time of day (Cullip et al., 1997). .....	14
Figure 2.6: Volume versus occupancy upstream (Cullip et al., 1997).....	15
Figure 2.7: Volume versus occupancy downstream. (Cullip et al., 1997).....	16
Figure 2.8: Spatial pattern of single vehicle crashes at different scale levels (Plug et al., 2011). .....	21
Figure 2.9: Spiderplots of all vehicle crashes versus single vehicle crashes in western Australia.(Plug et al., 2011) .....	22
Figure 2.10: Spider plots of all single vehicle crashes temporal distribution by related factors (Plug et al., 2011).....	23
Figure 2.11: Comap for single vehicle crashes in Perth (Plug et al., 2011).....	24
Figure 3.1: IH-10 in El Paso, Texas.....	33
Figure 3.2: IH-410 In San Antonio, Texas .....	34
Figure 3.3: Frequency histogram of crash rate per million vehicle miles traveled for IH-10 in El Paso Texas.....	37
Figure 3.4: Frequency histogram of crash rate per million vehicle miles traveled for IH-410 in San Antonio, Texas.....	38
Figure 4.1: Sample crash plot in El Paso on highways and expressways.....	40

Figure 4.2: Coordinate points of intersection along IH-10 in El Paso, Texas. ....	42
Figure 4.3: Coordinate points of Loop IH-410 in San Antonio, Texas. ....	43
Figure 4.4: Ranking and fitting crash coordinates among two corresponding intersection points on a straight or curved interstate highway. ....	45
Figure 4.5: Ranking and fitting crash coordinates among two corresponding intersection points on a looped interstate highway.....	48
Figure 6.1: Exponential relationship of number of lanes and access density on crash rate for IH-10 and IH-410.....	62
Figure 6.2: Tobit relationship of number of lanes and access density on crash rate for IH-10 and IH-410. ....	64
Figure 6.3: Tobit relationship of number of lanes and access density on crash rate for IH-10 and IH-410 3-dimensional. ....	66
Figure 6.4: IH-10 crash frequency versus distance range per intersection in El Paso, Texas. ....	67
Figure 6.5: IH-10 crash frequency divided by the number of intersections within distance range versus distance per intersection in El Paso, Texas.....	68
Figure 6.6: IH-410 crash frequency versus distance range between intersections in San Antonio, Texas.....	69
Figure 6.7: IH-410 crash frequency divided by number of intersections within distance range versus distance per intersection in San Antonio, Texas.....	70

# **1 Introduction**

## **1.1 Background and Motivation**

In the U.S., crashes are very costly. It is estimated that \$242 billion was lost in year 2010 due to fatalities, injuries and properties from motor vehicle crashes resulting in a considerable loss of human and economic resources (NHTSA, 2015). At this time, the amount in terms of motor vehicle crashes represents around 1.6% of the \$14.96 trillion real U.S. gross domestic product. In 2010, motor vehicle crashes resulted in 32,999 fatalities, 3.9 million injuries and 24 million vehicles that were wrecked. Included in the figure of the 242 billion, property damage accounted for \$76 billion and \$77 billion accounted for lost market and household productivity. Total medical costs summed up to \$23 billion. In 2010, motor vehicle crashes accounted for a cost of \$18 billion to all U.S. taxpayers. This is approximately 7% of all motor vehicle crash costs or \$156 of additional taxes for every household in the U.S. In crashes where drivers drove past the speed limit the cost was a total of \$52 billion in 2010. The amount of higher fuel consumption, vehicular congestion delay, and air pollution caused by motor vehicle crashes were estimated at \$28 billion in 2010 (NHTSA, 2015).

In addition, with the rapid growth of a local economy there is an increased demand for access points, entrances and exits into road facilities, particularly on multilane interstate highways. It is in those times that compromises between access availability to commercial or residential developments are made at the expense of traffic operations and safety. Furthermore, the need for accessibility and mobility versus the constraints of capacity and safety arises. As more access points are created so are more conflict points generated that increase the probability of crashes. However, is there a possibility to provide additional access points that procures

safety and improves traffic flow? Can a methodology be determined that provides the best possible spacing between access points while improving traffic flow and reducing crash rates?

## **1.2 Research Objectives**

Demand for access points has steadily increased due to rapid growth in local economies creating compromises between accessibility and mobility or capacity and safety.

The purpose of this research is to establish a methodology that provides the best possible spacing between access points that improves traffic flow and keeps traffic crashes to a minimum. Crash models can be used to help transportation planners and engineers assess the relevance of access when entering or exiting onto interstate highways. With the use of econometric models, different geometric and traffic characteristics can be assessed to determine future problem areas in terms of crash rates. Geometric characteristics such as number of lanes, median widths, roadbed widths and other exposure variables and their effects can further be observed through these modeling techniques.

The objective of this research is to conduct a statistical analysis of the impact of access points on crash rates and traffic flow to determine best possible spacing between access points and provides the most mobility without compromising safety or vice versa.

## 2 Literature Review

### 2.1 Crash Prediction Models

Crash prediction models (CPD) assist in the determination of how certain locations within a transportation network are more prone to crashes. This aids in the development of where to target and prioritize transportation projects in terms of safety. Future vehicular flows based on travel demand modeling can assist in crash prediction and determine future problem areas. Factors to consider are traffic volumes, volume to capacity ratios and vehicle miles travelled.

#### 2.1.1 Count Data Crash Models

Tarko et al. (1999) discussed that in any step a proper regression model must be selected. A widely used crash prediction model is the negative binomial distribution of this form:

$$A = kLQ^{\beta} e^{\sum \gamma_i X_i} \quad \text{Eq. (2.1)}$$

where:

$A$  = frequency of crashes (acc/5 years)

$L$  = length of the section

$Q$  = Average Annual Daily Traffic (AADT) on the section (1000 veh/24 h)

$k$  = slope parameter

$X_i$  = explanatory variable of factor  $i$

$\beta$  = coefficient of AADT

$\gamma$  = coefficient of the factor i

According to the Federal Highway Administration (1998), Poisson and negative binomial models are a prevalent method of modeling discrete rare events such as roadway crashes. The assumption being that crashes taking place on a specific road or at a specific intersection are not connected. In addition, a certain mean number of crashes per unit time are an attribute of the given location and other locations with similar features where the mean is presumed to depend on highway variables.

When investigating the safety effects of the dependent crash variable and assessing the regression coefficients ( $\gamma$ ), appropriate independent explanatory variables  $X_i$  should be incorporated. This can be achieved through appropriate statistical tests such as test of correlation coefficient, T-test, F-test, and maximum likelihood ratio test (Washington et al., 2011). Some independent variables that are known to have impacts on crash behavior are (Tarko et al. (1999)):

1. Segment length (expressed in meters),
2. Annual Average Daily Traffic (veh/day),
3. Number of lanes,
4. Lane width (m),
5. Median type (barrier type or not),
6. Median width (m),
7. Outside and inside shoulder widths (m),
8. Inside shoulder widths (m),



9. Paved outside shoulder
10. Presence of auxiliary lanes
11. Pavement serviceability index (PSI),
12. Number of curbs
13. Access control
14. Pavement material
15. Number of parking lanes
16. Access density number of access points kilometer or mile

When comparing the Poisson or negative binomial (NB) model to nonlinear regression results and analysis show greater effectiveness of Poisson and NB regression models compared to nonlinear regression model (Konduri et al., 2003).

Due to the manner in which crashes are collected or how the data is counted, crash models are typically of Poisson, exponential and generalized linear form (Washington et al., 2011). The number of crashes in a given space-time region is regarded as a random variable that takes integer values 0, 1, 2, ... with probabilities obeying the Poisson distribution. A characteristic feature of this distribution is that the variance, or mean squared deviation of this variable, is equal to its mean. The model falls under the heading of a generalized linear model. The exponential function guarantees that the mean is positive.

Negative binomial models, a variant of the Poisson model, are used in crash modeling. Such models generalize the Poisson form by allowing the variance to be over-dispersed, equal to

the mean with a term whose coefficient is the over-dispersion parameter. When this parameter is zero, the result is a Poisson model. When it is larger than zero, it represents variation in the independent variables present in the model. Such variation is due to crash-related factors pertaining to drivers, vehicles, and location not captured by the independent variables in question.

According to Washington et al. (2011), with regard to crashes occurring per year at various intersections in a city, the probability of intersection  $i$  having  $a_i$  crashes per year  $P(a_i)$  is:

$$P(a_i) = \frac{e^{(-\lambda_i)} \lambda_i^{a_i}}{a_i!} \quad \text{Eq. (2.2)}$$

and  $\lambda_i$  is the Poisson parameter indicating the expected number of events per period. Since the data are non-negative integers with the mean approximately equal to the variance, the Poisson regression approach can be used for estimating the parameter vector  $\beta$  such that:

$$\lambda_i = e^{(\beta X_i)} \text{ or } \text{LN}(\lambda_i) = \beta X_i \quad \text{Eq. (2.3)}$$

The exponential regression model is a continuous generalized linear model for nonnegative random variables (Greene, 2012). It is a single index model with probability density function:

$$f(y_i) = e^{(-\lambda_i X_i)}, y_i \geq 0, \quad \text{Eq. (2.4)}$$

$$\lambda_i = e^{(\beta' X_i)}, y_i \geq 0, \quad \text{Eq. (2.5)}$$

Where  $\beta$  is a vector having parameters that are estimated and  $X_i$  is a vector of descriptive variables. The conditional expected mean and variance are:

$$E\left[\frac{y_i}{x_i}\right] = \frac{1}{\lambda_i}, \quad \text{Eq. (2.6)}$$

$$VAR\left[\frac{y_i}{x_i}\right] = \frac{1}{\lambda_i^2} \quad \text{Eq. (2.7)}$$

### 2.1.2 Incident Clustering

Hanowski et al. (2004) developed a method called incident clustering. Approximately 45 to 75% of all roadway crashes is attributed to driver error. This is not well understood since specific errors that contribute to crashes and the characteristics of the driver's error are not easily specified. To clarify driver error, Hanowski et al. (2004) proposed a method that involves the grouping of incidents with similar characteristics that occur at the same location as well as the application of human-factor principles. The method improves on other approaches that rely only on traffic engineering principles. The human-factors principles allow critical crashes to be investigated from perspectives not determined through traditional physical engineering techniques. Examples include inappropriate behavior, inadequate knowledge, training or skill. The following is a method for conducting critical incident analysis for a specific site:

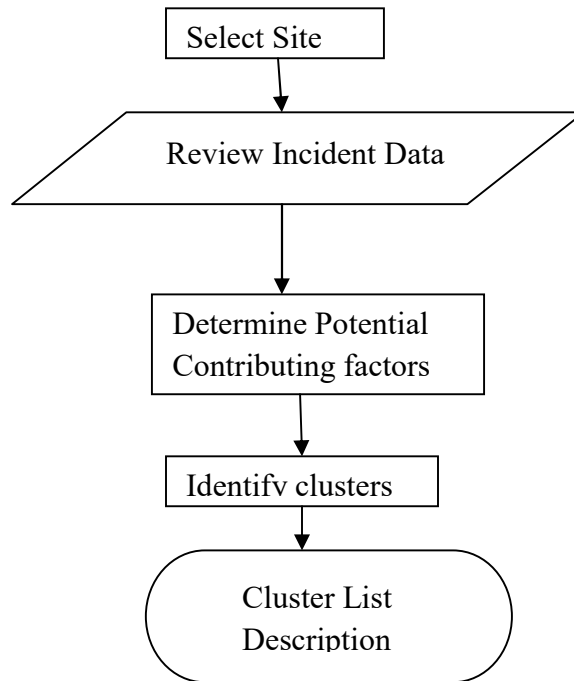


Figure 2.1: Method for conducting critical incident analysis for a specific site (Hanowski et al., 2004).

These five gathering techniques assisted in obtaining insight in terms of the crash location's geometry, control, surroundings and driver performance that may play a part in to traffic crashes. The following site evaluation method in Table 2.1 provides insight into potential infrastructure-related and non-infrastructure related causal factors associated with critical incidents analysis:

Table 2.1 1Steps to examine potential infrastructure-related and non-infrastructure related causal factors associated with critical incidents (Hanowski et al., 2004).

Step	Description
1. Site Survey	Conduct a site survey to ensure that a particular site is appropriate for in-depth investigation (i.e., is a problem site).
2. Site Drive-Through	Conduct a site drive-through with a car equipped with a video camera; drive through the site using all entrances and exits
3. Site Diagram and Description	Diagram the site in a bird's-eye (plan) view to provide insight into traffic conflicts that occur.
4. Videotape Surveillance	Conduct a videotape surveillance of the site to collect traffic event data. The video recordings can be conveniently analyzed in the lab at a later time.
5. Data Collection	Collect videotaped data at specified time intervals (e.g., dawn, peak hour, etc.).
6. Data Reduction and Extraction	Retrieve relevant information about the traffic events from the videotape and site-researchers' notes.
7. Incident Cluster Development	Indicate on the site diagram the precise location of each incident.
8. Conduct Appropriate Analyses	As appropriate and depending on the specific site problems, several analyses can be conducted to evaluate the site and lead to recommendations for infrastructure and non-infrastructure changes.

Critical incidences were assessed in terms of four classes (1) driver proficiency issues or problems with traffic law comprehension, vehicle kinematics, limitations, and driving skills; (2) driver impairment issues (i.e. alcohol, tiredness, health); (3) eager improper driving actions (i.e. road rage); (4) roadway infrastructure issues (i.e. geometrics, alignments, traffic signs, weather, delineation). These classifications led to incident cluster groups. The following, Figure 2.2 is a typical diagram that is developed in cluster analysis.

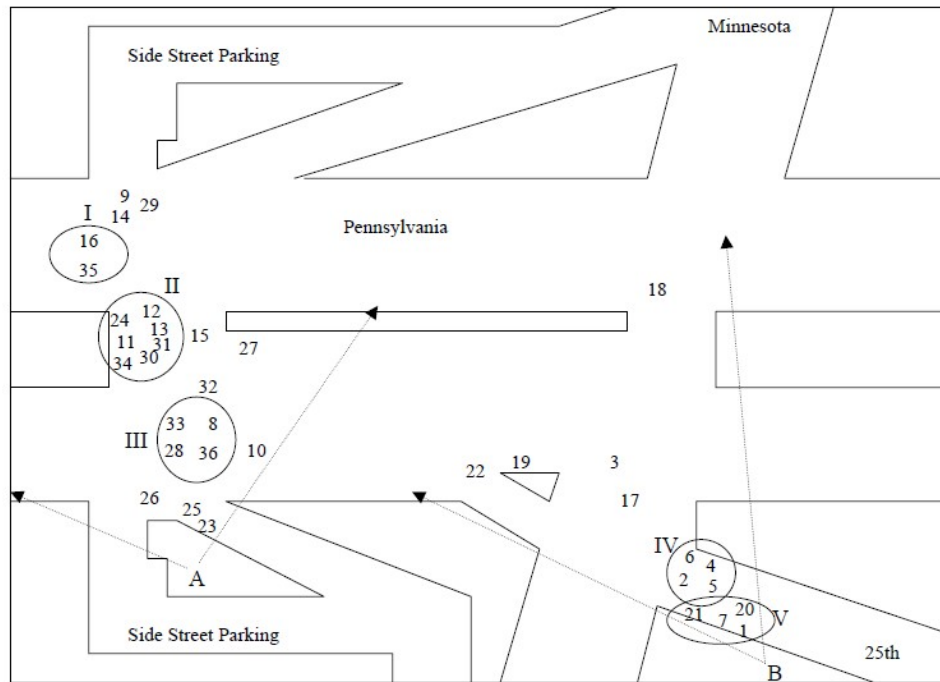


Figure 2.2: Cluster Analysis (Hanowski et al., 2004)

In Figure 2.2, each number refers to an incident reported event developed from videotaping of a crash in Washington, D.C. Potential contributing factors were assessed for each event such as driver proficiency, eager improper driving and other factors mentioned in preceding paragraph. The videotaping cameras locations are shown A and B in Figure 2.2. The viewing positions of cameras A and B are shown as arrows. Identified in Roman numerals are the groupings of the incidences into incident clusters shown in circles in Figure 2.2. In terms of cluster numerals IV and V it was hypothesized that those incidences were caused by potential infrastructure type issues. Clusters I, II and III did not entail any potential infrastructure but nevertheless showed a clustering of crashes. In cluster II an illegal U-turn was identified as eager improper driving since a posted no U-turn sign was physically present yet drivers continued to perform a U-turn. A possible contributing factor was identified as the sign's visibility.

However, Hanowski et al. (2004) describes a more plausible factor was the driver's need to make a prompt decision to find the simplest navigational route. The advantage of this approach is that it assists in foreseeing ahead of the noticeable causes of any crash incidence and not just defining it as driver error. Hanowski et al. (2004) mentions a potential solution for the U-turn situation is changing the infrastructure so drivers are not disobeying the law when performing the U-turn. An example of this could involve safe and legal modifications to enhance delineations or changes to traffic signal operations. Hanowski et al. (2004) proposed method contemplates further than just concluding a crash as a driver error. It re-assesses driver errors to determine a basis for developing better roadway design decision variables and looks further at potential design trade-offs.

### **2.1.3 Impact of Access driveway on Crash Rates on Multilane Highways**

In New Jersey, the highway system is regarded as one of the densest in the United States (Mouskos et al, 1999). A study was conducted that examined major geometric and traffic flow traits on crash rates at a macroscopic level. The main variables that were taken into account were AADT, number of lanes, shoulder existence, median presence, speed limit, and access points per mile. The study examined a relationship between the occurrences of crashes at signalized intersections and between intersections. Although not considered the only cause, access density is contributing factor of crashes between signalized intersections. Crash rates were found to be best modeled through a log-normal distribution rather than a normal distribution. Similar to Equation 2.3, the crash rate models derived by Mouskos et al. (1999) were determined as:

$$\text{LN}(\text{Crash rate}) = \beta_0 + \beta X_i \quad \text{Eq. (2.9)}$$

Characteristically if  $Y = LN(X)$  is normally (Gaussian) distributed, then the distribution of  $X$  is log-normal (Stahel et al., 2014). Stahel et al. (2014) describes in linear regression combinations that the means of normal variables are normally distributed and by central limit theorem means of non-normal variables are approximately normally distributed. According to Stahel (2014) regression models therefore presume normally distributed errors. In addition, if  $Y$  is normally distributed then it can be approximated by a Poisson distribution with  $\lambda = \mu$  and  $(\lambda)^{1/2} = \sigma$  (Disney et al., 1985). Disney (1985) shows that as the  $\lambda \rightarrow \infty$  the function  $Y = (Y - \lambda) / (\lambda)^{1/2}$  approaches the standard normal distribution.

The factors with significant impacts on crash rates were AADT, access density, shoulder, speed limit and intersection spacing. The number of lanes was determined to have less of significance. For the probability of equal means defined by method Kruskal – Wallis, the number of lanes for two and four lanes was significant at a value of 0.28. However, in a comparison between 2-lane and 4-lane highways, there was an indication that conclusions could not be arrived since variances were very large when compared to the matching means. Crashes with improper left turn and right turns movements at intersections were higher than crashes between intersections. Driver inattention was reported as greater reason for crashes at the section, the extent between intersections, than at intersections. No significant impacts were observed with rainy weather or wet conditions, however, when the road surface was covered with snow or ice there were differences between the proportions observed on the section crashes between two intersections as opposed to intersection crashes. Speed reduction, delay and the percentage of affected vehicles due to turning movements to and from access points were considered as significant variables in the estimation of access points on highway crashes.



### 2.1.4 Incident Occurrence Models for Freeway Incident Management

With regard to incident detections on arterial roadways, crash investigations are done by looking at crashes in upstream and downstream locations (Cullip et al., 1997). Figure 2.3 illustrates an example.

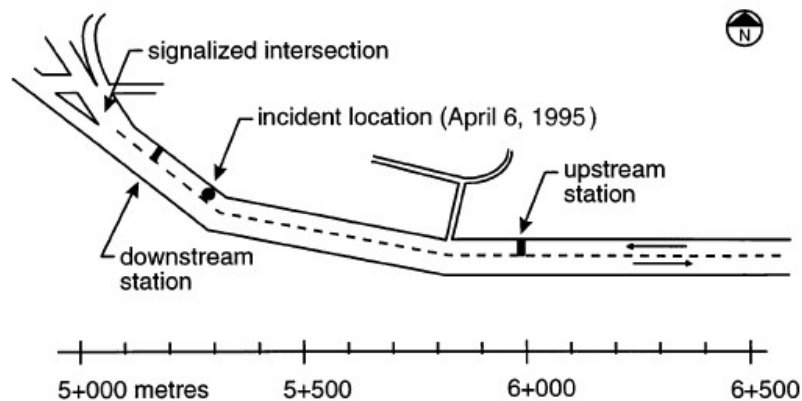


Figure 2.3: Crashes in upstream and downstream locations (Cullip et al., 1997).

An examination of the decrease in volume and occupancy, the percentage of time that there is a vehicle over the traffic response detector, are done by observing the data from downstream and upstream locations at specific times of the day on arterial roadways. With detectors such as loop, radar, infrared, ultrasonic or acoustic to name a few, it is possible to collect incident data and detect the presence of a crash or incident in a more timely manner than the time it takes a traffic operator to observe the crash. This is accomplished by analysis of traffic counts and percentage of occupancy provided by the detectors measuring traffic signal lengths and the development of an automatic incident detection algorithm. The algorithms use logged data to forecast future traffic measurements. An incident is acknowledged when the deviation between observed and forecasted data exceeds a predefined threshold. This can be shown in Figure 2.4 and 2.5. In a period of approximately 17 minutes, from 16.3 hours to 16.6

hours the statistical data from the detector shows a decrease in volume and percentage occupancy indicating the presence of an incident.

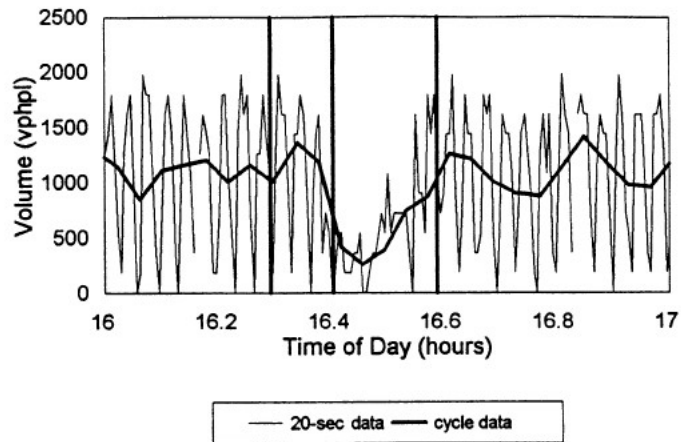


Figure 2.4: Volume versus time of day (Cullip et al., 1997)

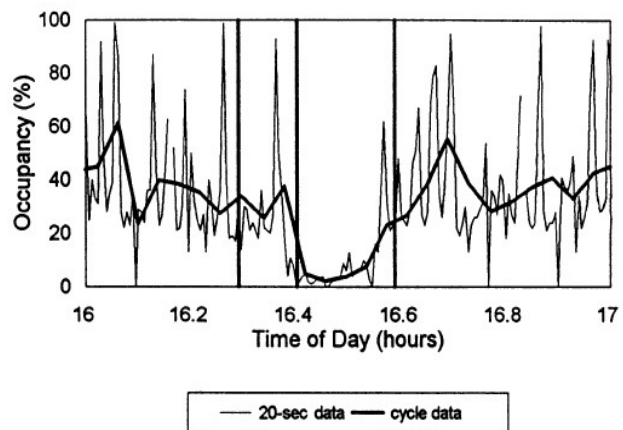


Figure 2.5: Occupancy versus time of day (Cullip et al., 1997).

### 2.1.5 Crash Prediction Models for Urban Unsignalized Intersections

Other characteristics that can be observed by Cullip et al. (1997) when examining Volume Vs. Occupancy model algorithms and automated incident detection (AID) logics at upstream and downstream locations of incident detection. One of these logics, as mentioned by Cullip et al. (1997), is the McMaster Freeway Incident Detection Algorithm and illustrated in Figures 2.6 and 2.7

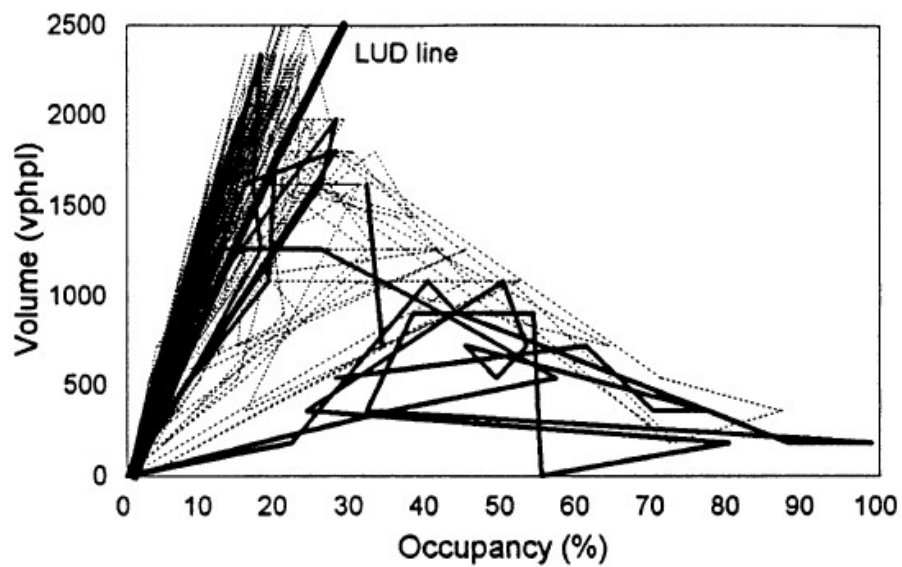


Figure 2.6: Volume versus occupancy upstream (Cullip et al., 1997)

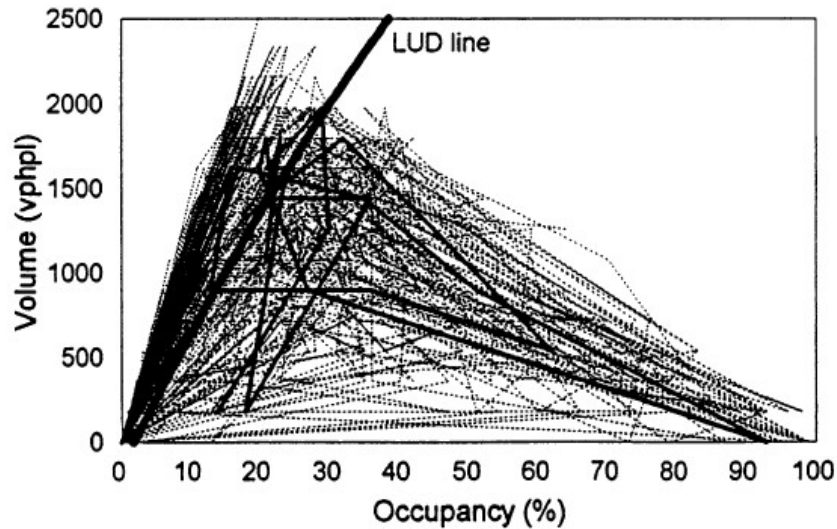


Figure 2.7: Volume versus occupancy downstream. (Cullip et al., 1997)

Figures 2.6 and 2.7 show a time connected volume occupancy plot for a lane of two stations where an incident was involved. The areas in Figures 2.5 and 2.6 are delineated by a lower bound of uncongested data (LUD) line considered a threshold. As described by Cullip et al. (1997) any data to the right is within a congested domain. Figure 2.6 shows the volume and occupancy at an upstream station with a reduced number of congested data points when compared to Figure 2.7. With more data points to the right of the LUD line, as presented in Figure 2.7, an incident is more apparent and predictable since it veers away from normal operating traffic conditions. Although accurate, Cullip et al. (1997) found this logic to present false alarms as there could be points unrelated to the incident. This prompted the investigation of other logics. This type of analysis can be used to determine the performance of access spacing and geometric features on interstate highways with more detectors positioned on these types of facilities.

According to Cullip et al. (1997), the logic producing the best results was an approach that compares the present cycle volume or occupancy with the average volume or occupancy

over the previous 3, 5, and 10 cycles. In one case this logic detected an incident 20 minutes before it was reported. This method demonstrated a competent way to spot an incident at different stations at a minimal time and no false alarms.

#### **2.1.6 Impact of Heavy Vehicles on Road Crashes**

Literature on heavy vehicle crashes is limited. To contribute in this area, Islam and Hernandez (2015) performed an analysis on fatality rates for heavy vehicle crashes. This approach utilized the random parameters Tobit model as it provides a more concrete perception of fatality rates per million truck-miles traveled. Statistically significant explanatory variables were searched for and determined to be temporal, spatial, road, driver, vehicular arrangements, passenger and environmental attributes. Data was collected through the Fatality Analysis Reporting Systems (FARS) database, a nationwide crash census system, from 2005 to 2008. Statistical analysis was performed through statistical software SAS and LIMDEP (NLOGIT 4.0). The number of occupants not fatally injured in a crash, number of vehicles involved in crash, effects of fatalities per ton-miles of freight were all found to be random and normally distributed. Number of vehicles in the crash and the number of persons not fatally injured were found to be statistically significant random parameters and the higher the number the higher the fatality rate. The study demonstrated the random parameters Tobit regression as a practical methodology in observing fatalities per million truck miles traveled and fatalities per ton-miles of freight.

Anderson and Hernandez (2016) investigated and compared two econometric methods: random parameter Tobit regression and latent class Tobit regression. This was performed to examine explanatory variables for heavy vehicle crashes per million-vehicle-miles traveled in Idaho. The study examined crash proportion by roadway classification. Crash rate distribution per principal arterials, major collectors, and interstates showed frequencies descriptive of a

distribution that has a large lower bound group of observations to the left similar to the representation of a negative exponential. Keeping this in mind it was pertinent to develop a statistical model that maintains linear assumptions essential for regression of a continuous dependent variable. Through the use of a random parameters approach, main factors by roadway classification were determined. These factors affect heavy vehicle crash rate. As part of a contributing factor for further research, this approach offered as well a structure to observe for heterogeneity. With respect to latent class Tobit modeling, latent class endeavors to obtain unobserved heterogeneity through allocating estimable parameters to differ with discrete distribution throughout unobserved groups of observations or classes. This approach is used to assess random parameter possibilities disadvantages for parameters to vary only throughout singular observations. By characterizing a finite number of points and assessing the mass probability of intervals between points heterogeneity can be checked. While distributions such as normal, uniform, triangular and lognormal were taken into account, the normal distribution was found to be the most statistically significant in terms of the mean and standard deviation. Log likelihood tests were employed to determine statistical significance through the use of chi square statistic with the degree of freedom being the number of estimated parameters. For principal arterials or interstates, the randomness of factors was found inconclusive and the fixed parameter model was a better fit than the model with no estimated parameters. For major collectors, variables were found to fit better a random parameter model. For Tobit regression for interstates in terms of t-statistics speed limit, AADT, passenger vehicle AADT, heavy vehicle AADT, curved horizontal geometrics, road configuration (2-way and divider), surface defects were considered and were all significant and correlated with heavy vehicle crashes per million-vehicle-miles traveled. For latent class Tobit regression in terms of t-statistics, it was observed

that for interstates with latent class 1, only surface defects were considered significant while for latent class 2 speed limit, total AADT, heavy vehicle AADT, and horizontal geometrics were considered significant. For each road classification, an increase in passenger vehicle AADT demonstrated a decrease in crash rates and decreases in passenger vehicle AADT showed increases in crash rates. Pearson product moment correlation coefficients show interstates to have a better model fit with the latent class approach than with the random parameter model.

## **2.2 Crash Clustering and Hot Spot Analysis**

The following sections describe spatial location methods to causal factors associated with the hotspots of crashes. There are various instances where crashes are attributed to the location where they occurred and when clustered together may indicate a point such as a traffic sign, a traffic signal, and a turning movement as a contributing crash factor.

### **2.2.1 Kernel Density Estimation (KDE)**

Plug et al. (2011) show that past investigations of single vehicle crashes have a propensity to group both spatially and temporally. However, little has been done in researching the correlation between the location of single vehicle crashes and the time they occur at various spatial extents. A limitation is the Getis-Ord hot spot analysis used in geographical information systems that require aggregation of data, rather than by means of individual crash locations. Temporal hotspots and times when crash frequencies are great have not been examined well or taken into consideration.

A solution to avoiding problems of aggregation is utilizing the kernel density estimation and performs analysis on a point dataset. A great advantage in applying the kernel density estimation method is that it helps in examining and clarifying the spread of risk of a crash. The

spread of risk is defined as the region around a distinct cluster with an increased probability for a crash to occur based on spatial dependency. The output is a raster layer with pixels assigned according to point density. Spiderplots were utilized to assess the temporal data and its chronological characteristics.

Visualization techniques such as a map animation, isosurface methods and comap were utilized. All vehicle crashes and single vehicle crashes were assessed at three different scales: regional, metropolitan and local areas. This can be seen in Figure 2.7. As shown in Figure 2.8, temporal data was assessed on weekday, weekends, on a daily and weekly basis.



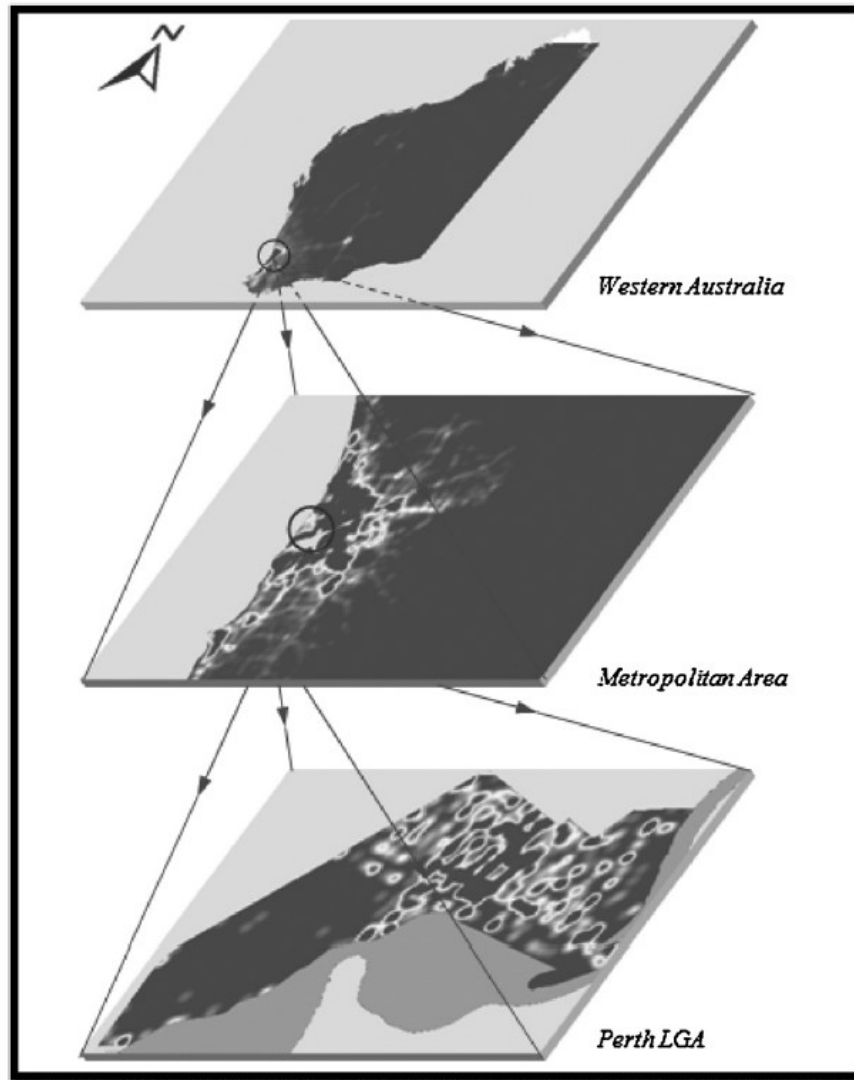


Figure 2.8: Spatial pattern of single vehicle crashes at different scale levels (Plug et al., 2011).

All vehicle crashes peaked between 8-9 a.m. and 3-6 p.m. or peak hours. Single vehicle crashes were more evenly distributed throughout the day with the highest frequency between 3:00 p.m. and midnight and light peak between 10:00 p.m. and midnight. Most single vehicle crashes happened during the night where lower traffic volumes are common and the likelihood for drivers to hit another car is low. Light or dark conditions were found uncertain. Shown in Figure 2.9, speeding was found to occur most frequently at night from 7:00 p.m. to 2:00 a.m.

with a peak at midnight. Fatigue crashes were found ordinarily in non-metropolitan areas around 1-2 a.m. and 2-4 a.m.

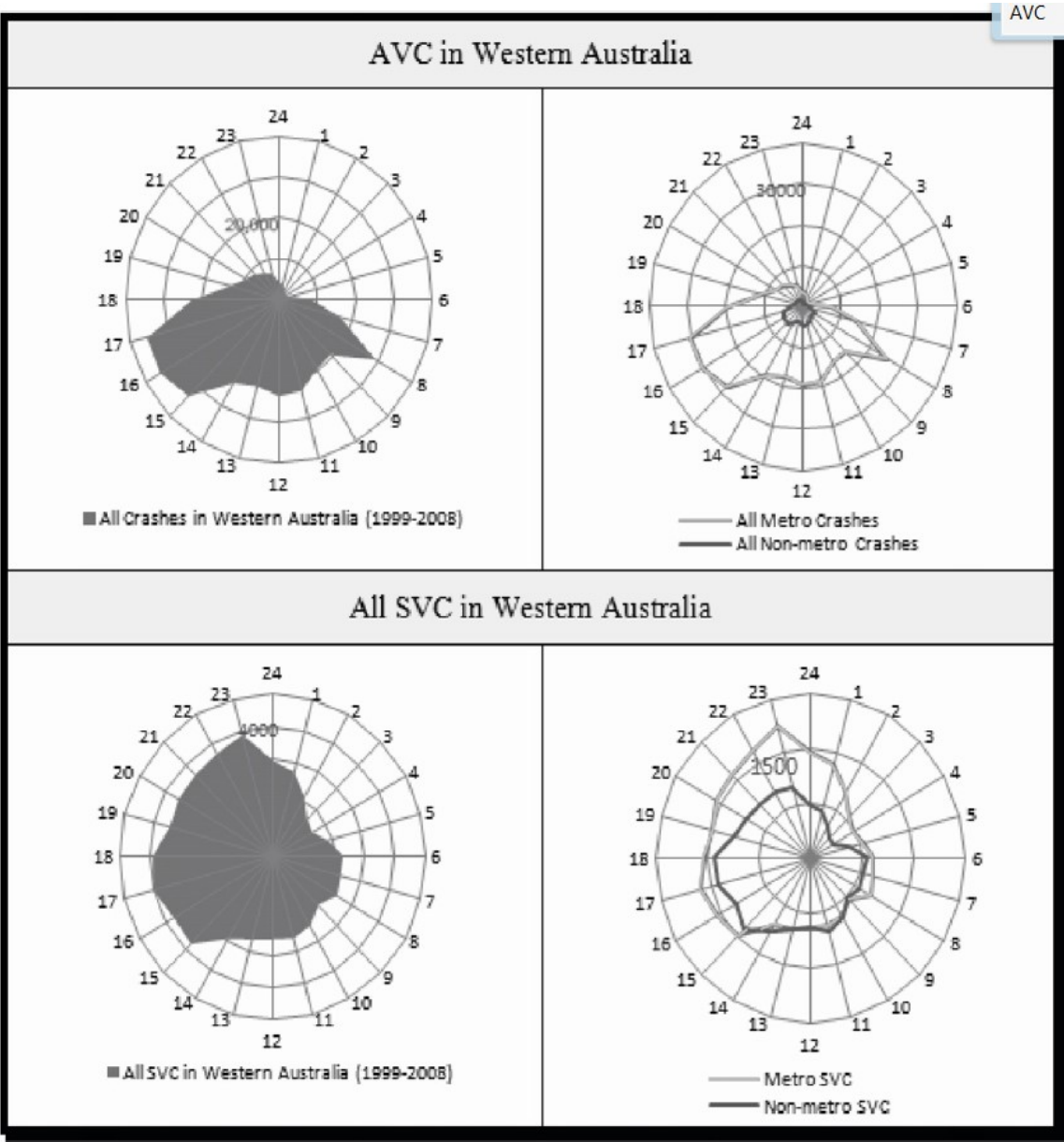


Figure 2.9: Spiderplots of all vehicle crashes versus single vehicle crashes in western Australia.(Plug et al., 2011)

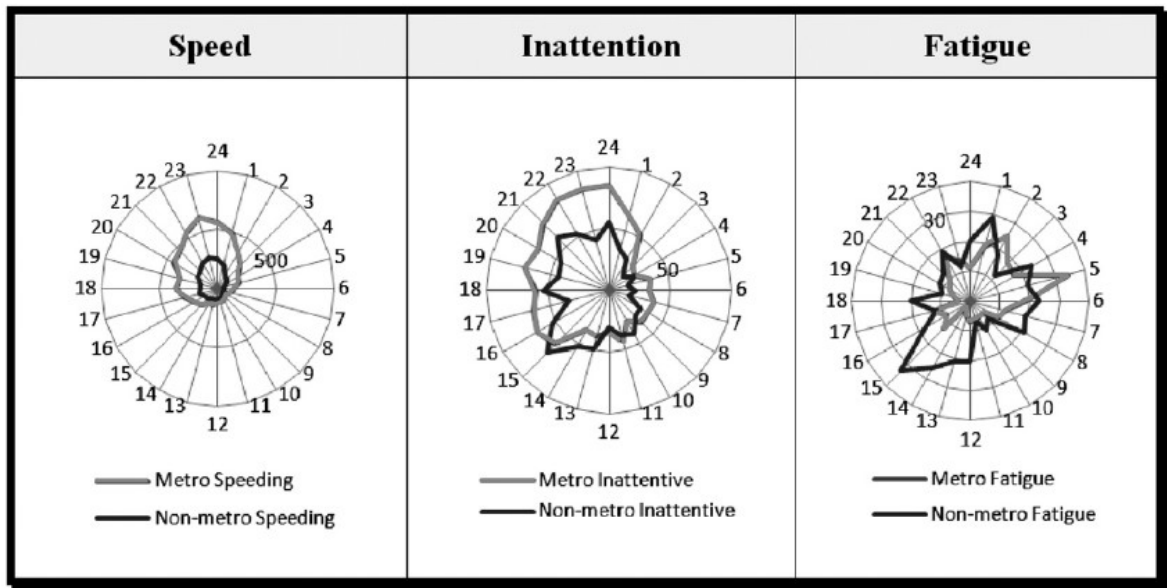


Figure 2.10: Spider plots of all single vehicle crashes temporal distribution by related factors (Plug et al., 2011).

Kernel density estimations showed that crashes are predominantly clustered around the metropolitan region as expected and more so in the center of the region. Crashes were clustered in the northern suburban parts on weekends and there were few crashes in those areas on weekdays. Weekday crashes occurred frequently in the central business district.

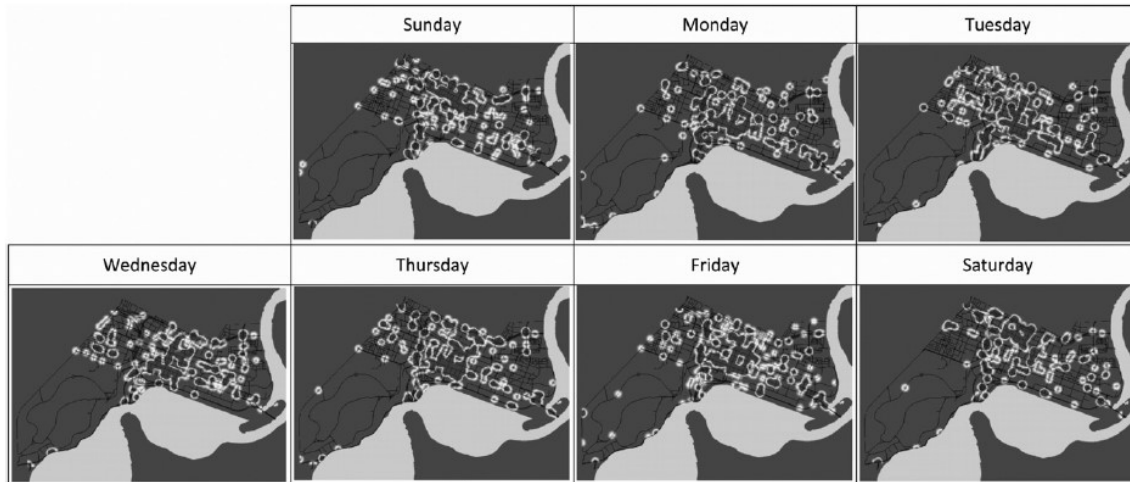


Figure 2.11: Comap for single vehicle crashes in Perth (Plug et al., 2011).

The spatial and temporal visualization techniques exposed vehicle crash patterns at certain areas and time periods as well as a revelation on the process of the crash patterns. As an example spatial and temporal hotspots exhibited by Kernel density estimation, spider graphs and comaps (Figure 2.10) show changes of patterns at certain timeframes or periods. This paper offers a way to examine spatial and temporal relationships to predict future crashes and how these clusters can determine best possible access density that reduce crashes.

According to Anderson (2008) identifying road crash hot spots plays a big role in allocating resources for safety improvements in order to reduce high density crash areas. Anderson explains (2008) the basis is that crashes occurring in resembling areas, not just road network or intersections, are spatially dependent due to a higher density of crashes in those particular areas as a result of a shared common cause. Despite human error or mechanical problems being typical causes of road crashes, spatial dependence should not be underestimated. Anderson (2008) mentions road crash hotspot analysis entails a complete comprehension vehicular crash process, the seriousness of injuries and awareness of the neighboring environment. Hotspots are based on existing data attributes concerning the crash such as vehicle

type, time of day, persons involved. The causes of exposed crash concentrations suggest a spatial dependence between crash incidences. Anderson (2008) uses KDE as a visualization of crashes based on density and clustering. This method according to Anderson (2008) assists in the examination of any environmental and land use factors that may cause crashes such as road length, traffic signals, speed cameras and signs. Anderson states (2008) the main advantage of KDE is in its establishing of the spread of the risk with an area around a defined cluster having a high chance of a crash to happen centered on its spatial dependency. Stats19 is a data base that was developed using crash data derived from reports involving local, central and police departments. Through Stats 19, Anderson (2008) compiled a collection of grid cells with high density. Anderson (2008) standardized the data by one of two variables, the number of crashes in each hotspot or the number of grid cells that compose each hotspot. The Stats19 information was consolidated by Anderson (2008) with the number of crashes within each hotspot and the road network data divided by the number of grid cells. In Anderson's (2008) study, various groups and clusters were identified and classified from five groups from A to E indicating A for pedestrians, B for high density vehicle damage, C danger to cyclists, D multiple main road crashes, E weekend risk takers and 15 clusters 1 to 15. The kernel method divides the region area under investigation into a predetermined number of grid cells. This method draws circular vicinity with a search radius (bandwidth) around each characteristic point such as a crash and a mathematical equation is applied from 1 at the position of the crash to zero at the perimeter of the vicinity. Anderson (2008) observed the KDE clustering process in this study noted a certain cluster type C10. Therefore, it identified 10 hotspots containing cyclists with an average number of grid cells per hot spot of 39.3 considered the highest in central London (specifically Westminster) with large traffic flows. The high index score for this cluster pointed out a high

number of crashes and a high number of cells. This shows that hotspots occurring in this cluster have a great number of crashes and are spatially widespread. In addition, the hotspots occurring in that particular Central London cluster demonstrated an encompassing sizeable area and a great number of crashes involving only one vehicle with the likelihood to result in a fatal injury. The result of this methodology provides a database of collision hotspots that share a common close spatial location implying a connecting casual factor. In this investigation KDE offers a method to assess the spread of accident risk or the risk of a crash occurring not just at a single point but over the estimated area. A shortcoming in need of further research for this approach is the establishing the statistical significance of the resulting clusters.

### **2.2.2 Impact of Traffic on Road Crashes**

According to Wang et al. (2009) traffic congestion and road crashes are two factors overlooked by transport policy makers. Their study mainly examined a hypothesis that an inverse relationship exists between those factors and whether or not the reduction of their impacts is simultaneously possible. Average speed of traffic is believed to be ordinarily high in a less congested road network resulting in more serious injuries and fatalities. In a congested road network where traffic is slower, it is conjectured there are less fatalities and injuries. Usually a positive correlation was observed between crashes and AADT. Thus, increased traffic volumes and congestion may lead to more crashes but in a less severe outcome. To transport policy makers the predicament would appear traffic congestion can improve road safety at the cost of a reduction in mobility and economic productivity. To examine these factors the study employed statistical models appropriate for count data such as Poisson-lognormal, Poisson-Gamma and a Poisson-lognormal model with Conditional AutoRegressive (CAR) from data collected on the M25 London Orbital Motorway disaggregated into 70 road segments.

The variable  $Y_i$  approximates the Poisson ( $\mu_i$ ) distribution as

$$\log(\mu_i) = \alpha + \beta X_i + v_i + u_i \quad \text{Eq. (2.9)}$$

Per Wang et al. (2009),  $Y_i$  is the observed number of crashes of road segment  $i$ ;  $\mu_i$  is the expected Poisson crash rate at road segment  $i$ ;  $\beta$  is the vector of coefficient to be estimated;  $v_i$  is the random term capturing heterogeneity effects on road segment  $i$ ;  $u_i$  is the spatially correlated effects for road segment  $i$ .

The data was obtained through the STATS 19 of the national road crash database. Other factors used in the assessment of each segment were traffic delay, traffic flow and average travel speed. For each of the four models, two scenarios looked into by Wang et al. (2009) were in terms of fatal, serious injury crashes and slight injury crashes. Logarithmic scale transformations were used to reduce variation on the independent variables such as AADT and the radius of curvature. The method of estimation used Markov Chain Monte Carlo (MCMC) with a complete hierarchical Bayesian structure. Statistical significant variables were found to be log of AADT, length of the segment and maximal vertical grade in all conditions for describing variation in the occurrence of slight injury accidents on the segments of M25 motorway. Number of lanes was significant under all conditions with the exception of Poisson-lognormal CAR. Standard deviation mean of the uncorrelated heterogeneity ( $v$ ) was found statistically significant for all models showing heterogeneity not present within the collected dataset. The standard deviation of the spatial correlation ( $u$ ) showed the road crashes as spatially correlated among road segments. The results did not confirm the congestion index as being significant in affecting road crashes. This was attributed to the basis of congestion varying from day to day and during the year having an influence on the effect of congestion on crashes. Wang et al. (2009) suggested that the Congestion Index would be captured by speed variance and traffic flow. This would

utilize measurements of acceleration noise at peak and off peak times and require a considerable amount of data because it would evaluate not only traffic conditions but also driver behavior.

Numerous studies depict an understanding of crashes and insight on roadway segments by method of non-negative count data modeling techniques such as Poisson and negative binomial (Anastasopoulos et al., 2007). However, there is a considerable interest to applying exposure based crash rates. An easily interpreted measure that has been standardized is the number of crashes per 100 million miles driven. Crash rates, a continuous variable, provide a better understanding of the safety of roadway segments rather than conceptualizing a non-negative integer value or number of crashes over a time period. According to Anastasopoulos et al.(2007) in finite time periods in mind and highway segments with chances of zero accidents within those periods, crash rates are better modeled through Tobit regression rather than standard ordinary least squares that would introduce biased and inconsistent parameter estimates. The Indiana study included exposure variables that were static in time and did not vary such as pavement, number of lanes and medians in contrast to weather or light conditions. These factors included pavement characteristics (friction, smooth pavement, rutting, pavement condition ratio), geometric characteristics (median width, median barrier presence, inside shoulder, outside shoulder), number of bridges (rumble strip presence, vertical curves per mile, ratio of curve length to roadway segment length, horizontal curve's degree of curvature, number of ramps), traffic characteristics (passenger AADT and average daily percent of combination trucks). Based on 325 observations, t-statistics showed all factors to be significant. Most factors showed a decrease in crash rates with the exception of pavement condition ratio, ratio of vertical curve length over the roadway segment length and number of ramps in the driving direction per lane mile. The number of ramps in the roadway segment per lane mile was noted to increase the



chances of a 100-million vehicle miles traveled crash rate by 16.87 % increase above zero. This was attributed to vehicle movements such as changing lanes, merging, turning, etc. This study showed that Tobit regression is a valid methodological approach for explanatory variables that contribute to crash rates.

## **2.3 Summary of Literature Review**

The literature presented is reflective of modeling techniques to assist in the prediction and analysis of crashes. Part of the literature investigated so far are regression and count models that assess statistically relevant geometric and traffic factors that either increase or decrease the number of crashes or crash rates. From data obtained from automatic incident detectors prediction models and algorithms assessing statistical flow characteristics such as volume and occupancy were developed that assist in the prompt detection of crashes. Incident clustering evaluated a method for the development of hotspots that assessed and classified the causes of crashes. Through KDE and clustering of crashes, a spatial and temporal analysis was performed to assess and predict the spread of risk as well as a method on how to target physical improvements to those crash areas.

Incident clustering and KDE offers a systematic approach to accurately target hotspots and assess locations where the risks of crashes are likely to occur on interstate highways. The limitation of this approach is that accurate police records are essential. If police records are inaccurate, this could present errors in terms of hotspot locations and lead to inaccuracies for making physical improvements to flow or access density at intersection segments. In terms of prediction models obtained through automatic incident detectors it offers a statistical way to assess flow characteristics though volumes and occupancy in the event of an incident. The automatic incident detection method is useful in the determination of physical improvements to

react quickly in avoiding congestion between intersections. Useful information that could be obtained from automatic incident detectors over a period of time are the identification of intersection segments containing the most number of incidents to the lowest number of incidents. Other useful data is the assessment of those incidents and the creation of a mean time between crash-failure of the various intersection segments to assess a preventive transportation infrastructure improvement or maintenance schedule to decrease crashes or incidences. However, a limitation is the amount of false alarms the method generates and potential errors that can arise when investigating physical improvements to reduce incidents or crashes for interstate highway intersection segments. Regression models can assess various factors that may or may not contribute to crashes and crash rates and are useful so long as the dependent factors are statistically valid. The advantage of regression models is that it does not rely too heavily on accurate police records. Statistical validation assessed in regression modeling errors determines the effectiveness of the factors obtained by police records and which factors contribute best to the crash model. In addition, the statistical validation from regression modeling filters police data that is not useful to the crash model. Regression modeling relies more on confirmed reported crashes rather than automated incident data that is subject to false crash alarms. A limitation is that regression modeling requires a substantial amount of data to ensure its validity that requires data processing resources however for the purposes of this study is the best choice.

### 3 Data

The Texas Department of Transportation (TxDOT) collects and analyzes crash data submitted by law enforcement officers on form CR-3, Texas Peace Officers Crash Report. A statewide automated database is maintained for all reported motor vehicle traffic crashes received by TxDOT. Crash records and reports are retained and approved by the Texas State Library and Archives Commission for the current calendar year. Statistics contained in these reports are generated from data provided by TxDOT's Crash Records Information System (CRIS).

The following are five files contained within this database:

- Crash Data – information about the crash, including conditions, location, counts, etc.
- Unit Data – information about vehicles involved in the crash including vehicle type, carrier motor vehicle (CMV) data, etc.
- Person Data – information about drivers, passengers, and non-vehicle persons (i.e. pedestrians and cyclists) involved in the crash. No personally identifying information will be included.
- Lookup Data – translation table for decoding ID values used in the other tables. Generally, the column name from the Crash, Vehicle/Driver, or Person Data file will match the given column name in the Lookup Data file. Where the names are different, the notes for that column indicate which lookup value to use.
- Citation Data – information about the citation identification (ID(s)) for the crash.

The purpose of the crash-access point research is to examine the crash predictive component of 6374 observations taken from Crash data of the TXDOT Crash Records Information System along Interstate Highway 10 in El Paso County and Interstate Highway 410 in San Antonio.

Using the latest crash statistics, years 2006, 2007, 2008, 2009, 2010, 2011, and 2012 was considered for this study. In terms of vehicle types this study considers multi-vehicle crashes from all vehicle types that involve large vehicles (i.e. heavy duty) or small vehicles (i.e. passenger vehicle, motorcycles).

### **3.1 Multiple Vehicle Crashes**

The scope of this analysis considered all type of vehicle crashes whether single or multiple according to crash reports. It is the intent of this study to examine the effects of single or multiple vehicular crashes as vehicles enter or exit to connect to their destined intersection or into the highway. In terms of access management where all sorts of vehicles weave one past another, all possibilities for vehicles to rear-end, strike at an angle or sideswipe other vehicles apply.

El Paso IH-10 was chosen for crash analysis since it is a vital transportation link mobilizing vehicular traffic bi-directionally east and west through the Franklin Mountains and typical of congested traffic. There are frequent crashes on IH-10 and mobility is challenging due to its geographical constraints as seen in Figure 3.1 El Paso is a city that is approximately parted in the middle by Mount Franklin dividing it into west and east halves. To the south of El Paso, the Rio Grande River is an international boundary between El Paso and Juarez, Mexico. Out of 3 vehicular mountain passages within the El Paso region, IH-10 is the most utilized transportation corridor narrowly passing between two geographic constraints, the Franklin Mountains and the southern border of El Paso to Juarez, Mexico. This section of IH-10 goes

through an area that is more densely urbanized areas while the areas to the far west and far east are less densely populated similar to rural areas. IH-10 has a total of 35 interstate segments or freeway sections between intersections. The segments vary in length from 0.32 miles to 6.7 miles.



Figure 3.1: IH-10 in El Paso, Texas.

San Antonio IH-410 was selected for crash analysis as it is as well a vital loop that circles the San Antonio metropolitan area as seen in Figure 3.2. It begins and ends at Interstate 35. IH-410, known as Connally Loop, has a northern half that facilitates traffic a densely urbanized area. It has up to five lanes in each direction. The southern half facilitates traffic in less dense populated areas comparable to rural areas with approximately two lanes in each direction. Apart from facilitating transportation to residential and commercial areas IH-410 provides access to airports, manufacturing, medical research facilities, etc. similar to El Paso's IH-10. IH-410 has a total of 49 interstate segments or freeway sections between intersections. The segments vary in length from 0.25 miles to 2.4 miles. Both IH-10 and IH-410 are considered acceptable

transportation facilities to study crash rates and traffic flow due to similarities in the variety of segment lengths and their service to rural and heavily urbanized areas.

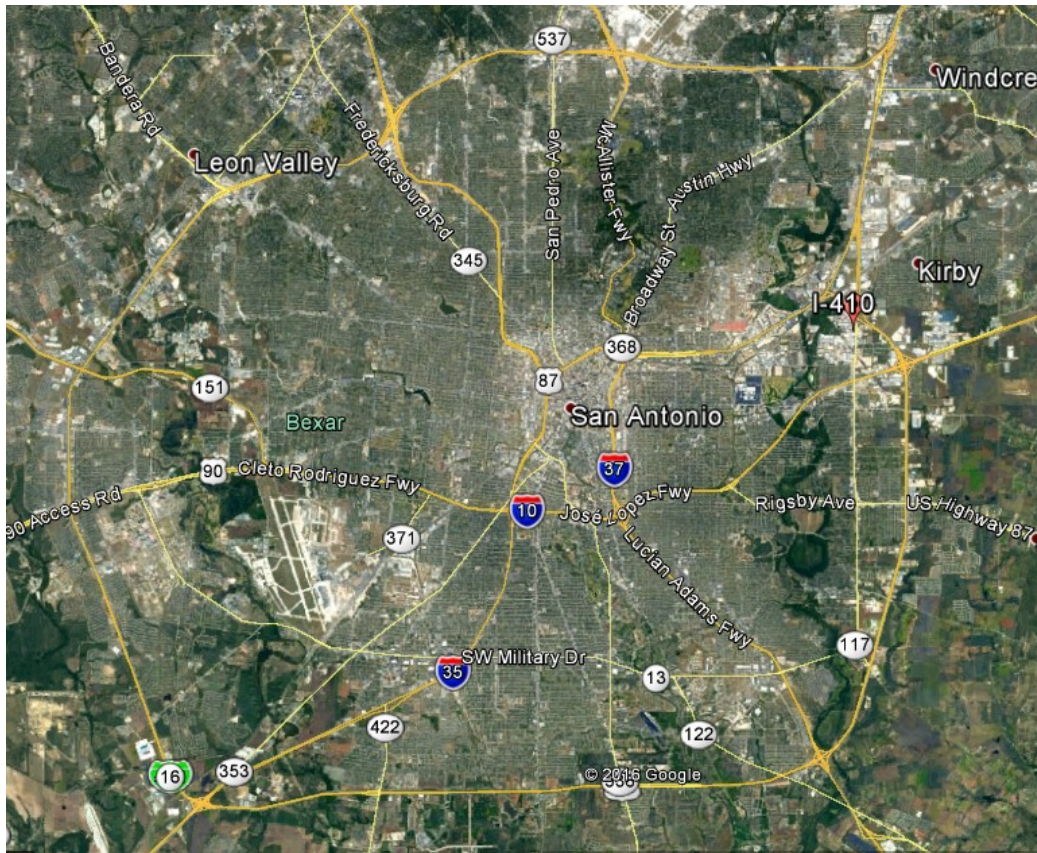


Figure 3.2: IH-410 In San Antonio, Texas

### 3.2 Model Variables

Vehicular crash data from interstate highways in Texas (IH-10 and IH-410) were collected for seven years (January 1, 2006 through December 31, 2012). Exposure variables, such as right shoulder length, presence of medians, number of access points, speed limits and roadway traits were examined on crash rates per 100 million vehicle miles traveled. Table 3.1 shows a summary statistics for geometric and traffic variables for both IH-10 and IH-410 combined.



Table 3.1: Summary statistics of geometric and traffic variables.

Description	Average	Standard Deviation
HP Shoulder RT (X8)- Width of outside shoulder on divided sections in feet	7.24	3.39
Auxiliary Lanes – (X23)- Lane to enter or exit between intersections ( 1 if Auxiliary lane present, 0 otherwise)	1.63	0.83
Number of Lanes observed (X24) – Quantity of lanes in the segment	3.24	1.17
Access Density (X25) – Number of access points per mile	4.68	3.23
Speed Limit (X26) – posted speed limit in mph	65.6	4.9

Examination of data show IH-10 in El Paso Texas and IH-410 to comparably have a high number of crashes, high traffic flow rates and many ramps and access points in various locations. A limitation in the CRIS dataset was that there was no ramp or access point information. Therefore, Google Earth was used to join I-10 access point information with the dataset using the mile-marker as a unique identifier to form an access point-crash relational database. As a result, the number of ramps was included as an additional field in the CRIS dataset for this analysis.

Since the information was provided in a raw format, a visual basic program in Excel was developed to count the number of crashes per day which was then included as an additional column. The data was collected or aggregated in a manner that identifies each data point that fits into the corresponding discreet segments bounded by two intersections where the crash occurred. Due to similarities in physical geometry on both directional sides, crashes were counted altogether within the segments. With the aggregation of data it was determined that only exposure variables or static variables that do not change with time (i.e. shoulder width, median width, number of lanes) were to be used. Non exposure random variables (i.e. light, weather,

and days of week) introduce higher errors within the model that require a greater sample size (Montgomery, 2012). This is attributed to the fact that crash rates on specific interstate segments are evaluated over some limited time. Consequently, the probability that no crashes are reported during the analysis period on many highway segments is likely. Thus, modeling crash rates by standard ordinary least squares would result in biased and inconsistent parameter estimates.

A dependent variable considered a key fundamental was access density defined as the number of access points per mile on a highway section in each direction. Dividing the number of access points with the corresponding section length is the method for determining access density (Mouskos et al., 1999). This measurement provides the distance between access points as shown below:

$$AD = \frac{N}{L} \quad \text{Eq. (3.1)}$$

where the variables are,

$AD$  = access density, #/mile

$N$  = number of access points

$L$  = length of the corresponding roadway section, in miles

Another key fundamental dependent variable is the crash rate that is defined as the number of crashes per million vehicle miles traveled on a highway section in each direction. The method for determining this variable is through dividing the number of crashes that took place in each direction with the AADT and length on a highway segment. It is determined by:



$$CR = \frac{M}{365 \cdot AADT \cdot L} \cdot 1 \times 10^6 \quad \text{Eq.(3.2)}$$

where the variables are,

$CR$  = crash rate, the number of crashes per Million vehicles traveled times the segment length

$M$  = number of crashes

$AADT$  = Annually Average Daily Traffic, in vehicles per day (vpd)

$L$  = length of the corresponding highway segment, in miles

In this study the crash rate is calculated by combining both directions of traffic. Figure 3.1 and Figure 3.2 show a frequency histogram of the crash rate per million vehicle miles traveled for El Paso IH-10 and San Antonio IH-410 respectively.

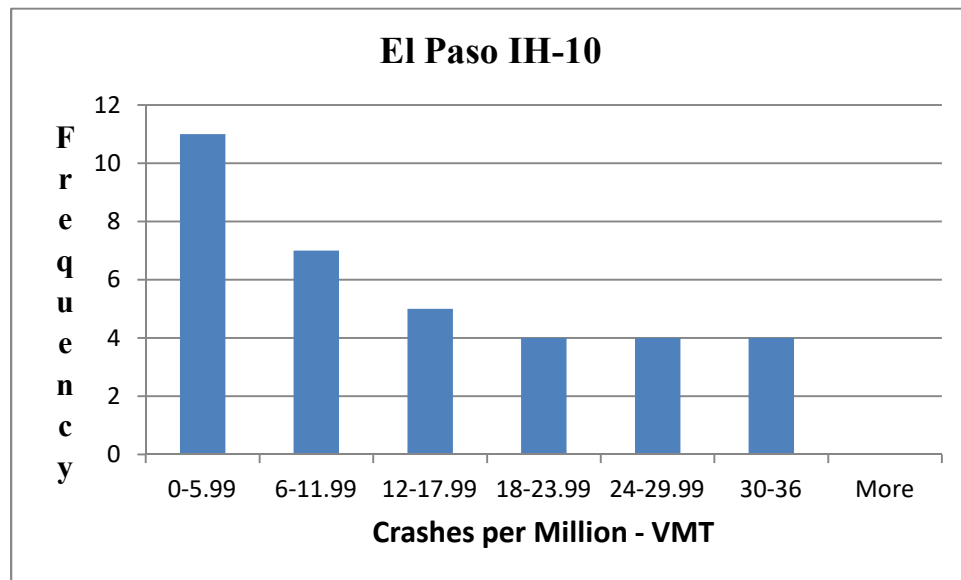


Figure 3.3: Frequency histogram of crash rate per million vehicle miles traveled for IH-10 in El Paso Texas.

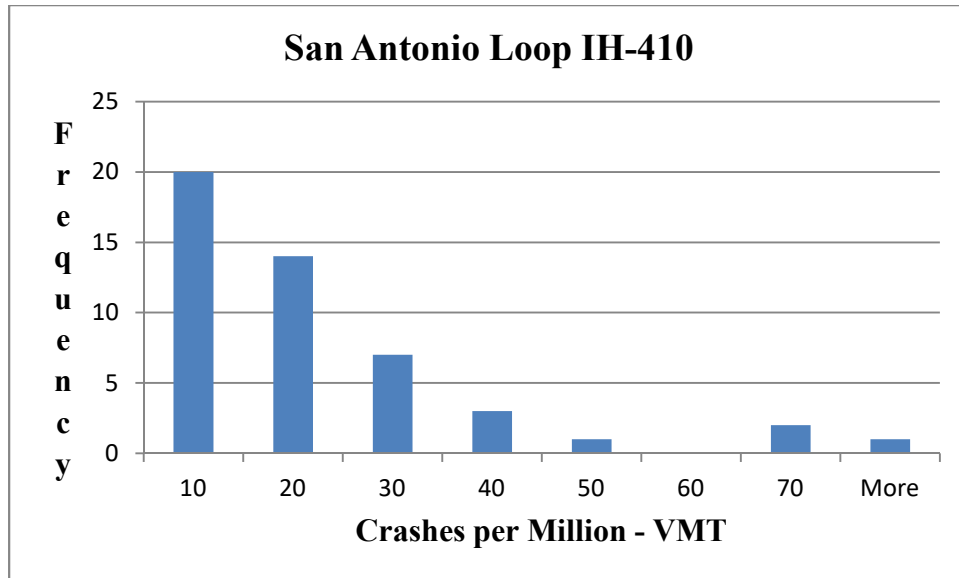


Figure 3.4: Frequency histogram of crash rate per million vehicle miles traveled for IH-410 in San Antonio, Texas.

The trend of Figures 3.3 and 3.4 have a leftward skew reminiscent of a negative exponential model. Similar to Anderson and Hernandez (2016), Figure 3.3 and 3.4 demonstrate the need to for a modeling approach that can explain the significant lower bound gathering of observations. In addition, the modeling approach must uphold linear assumptions for regression of a continuous dependent variable. The skew of both figures toward zero exhibits a likelihood of highway segments to have no crashes reported during the analysis period. This scenario renders the use of regression through ordinary least squares an impractical form of modeling since it can result in biased and inconsistent regression variable estimates. As pointed out by Anastasopoulos et al., (2007), Tobit regression is a suitable modeling approach since it constrains the dependent variable, crash rate, to be censored at zero.

### 3.3 Police Crash Report Variables

On site police information was collected in terms of location of the crash characteristics of roadway, vehicular, contributing environmental circumstances, geometrics, occupant injuries and type of vehicles. Although a wealth of valuable crash information was

compiled, there were police reports with missing or limited information to determine any worthy conclusions. There was a lack of consistent data for physical attributes such as horizontal and vertical curve data. This type of data can have a significant influence on crash rates as vehicles can veer off track due to inattention, weather, visibility, pavement conditions, curve design, etc. The lack of observations led to the discarding of variables that could contribute to the validity of a sensible crash model.

### **3.4 Summary**

It was through the TXDOT, Texas Peace officers and their resources that a vast crash information system (CRIS) was made available containing crash, unit, person, look up, and citation data, Crash statistics for this study were based on years from 2006 through 2012. Data was assessed in a manner to indentify physical attribute variables that can describe crash rates through their effects. Sites IH-10 and IH-410 were chosen due to similarities in the variety of segment lengths and their service to rural and heavily urbanized areas. Summary statistics were provided for exposure variables such as right shoulder length, presence of medians, number of access points, speed limits and roadway traits were examined on crash rates per 100 million vehicle miles traveled. Limitations in terms of data for access density information were presented. Crash rate and access density was defined as well a preliminary examination on what modeling approach would be suitable for the subsequent chapters of this study based histograms of the data presented.

## 4 Methodology

### 4.1 Intersection Distance to Crash Model

Like most cities in the United States apart from the Northeastern part of the U.S., El Paso and San Antonio was chosen as a case study since most of the population in both cities is in need of a vehicle due to urban sprawl. As a first step, a GIS plot of the crashes was performed on all major arterials and expressways in El Paso. Figure 4.1 shows a plot of crashes on all interstate highways in El Paso.

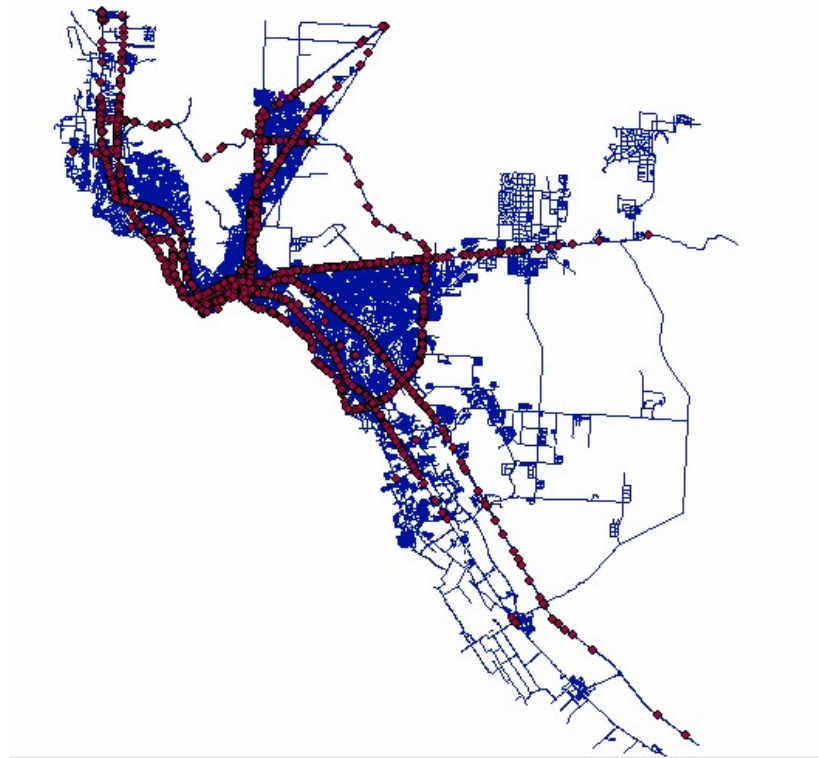


Figure 4.1: Sample crash plot in El Paso on highways and expressways.

In this study it was important to see how many crashes occurred between each intersection and the factors (i.e., geometrics, lane width, etc.) obtained from crash reports that contributed to those crashes within those intersections. Each crash coordinate in terms of latitude and longitude coordinates were fit among two corresponding intersection coordinate points. To do this, each analysis year containing all TXDOT CRIS Crash data in Texas to create a crash dataset was separated by city and county for both El Paso and San Antonio. With County and City information extracted, a new subset of that crash data was created to isolate and include only Interstate-10 for El Paso and Loop 410 for San Antonio. Intersection lat-long geographic NAD 83 coordinates for IH-10 and Loop-410 were identified from Google earth. A total of 36 coordinate points were determined for I-10 in El Paso to identify latitude and longitude of intersections along I-10.



Figure 4.2: Coordinate points of intersection along IH-10 in El Paso, Texas.

and 49 coordinate points were determined for IH-410 in San Antonio. The intersection coordinate points were taken and overlaid with the crash coordinate points. Corpscon 6 (U.S. Army Corp of Engineers, 2004) was used to convert all IH-10 intersection and crash coordinate points from Lat-Long to State plane coordinates NAD 83 for Central Texas 4203.



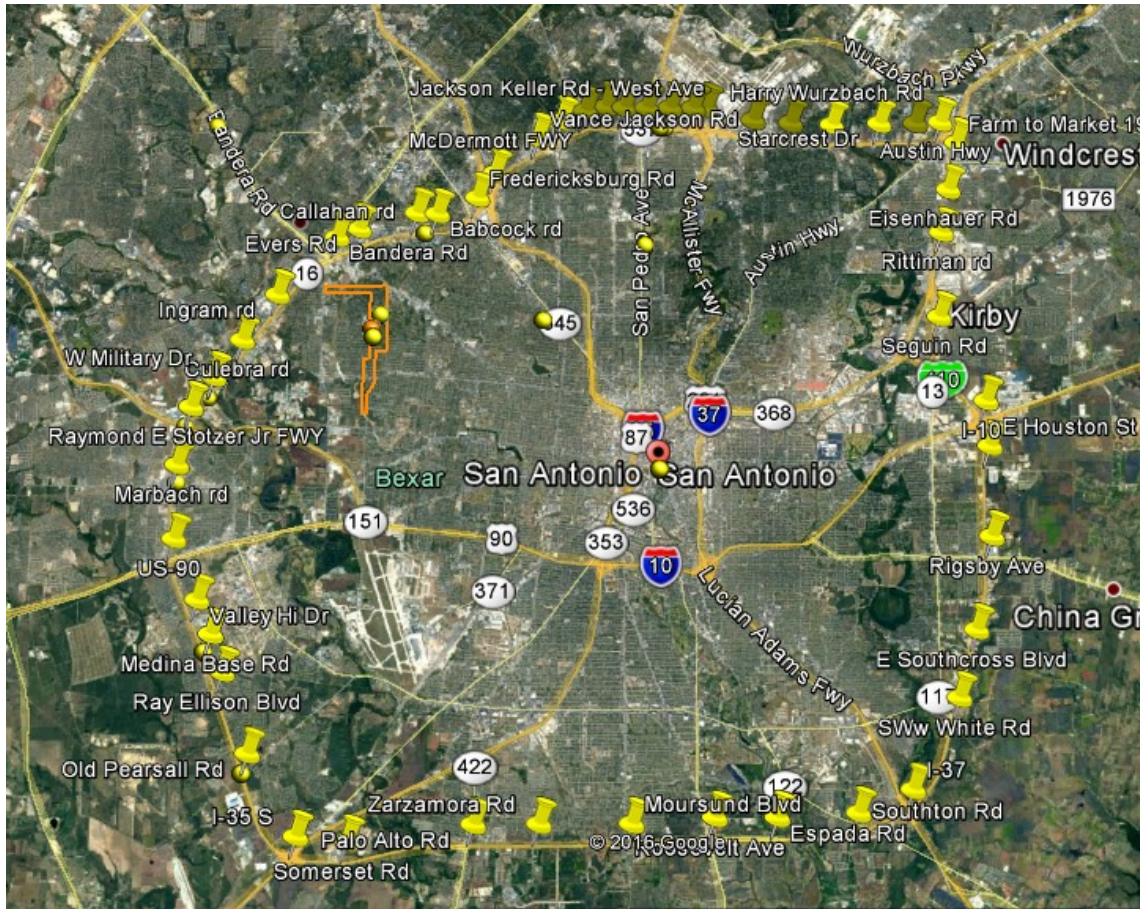


Figure 4.3: Coordinate points of Loop IH-410 in San Antonio, Texas.

This program is a MS-Windows-based program permitting clients to convert coordinates between Universal Transverse Mercator (UTM), State Plane, Geographic and the North American Datum of 1983 (NAD 83), High Accuracy Reference Networks (HARNs) and US National Grid systems on the North American Datum of 1927 (NAD 27). Corpscon draws on the National Geodetic Survey (NGS) program Nadcon to translate between NAD 27, NAD 83 and HARNs. Corpscon 6 provides vertical conversions to and from the National Geodetic Vertical Datum of 1929 (NGVD 29) and the North American Vertical Datum of 1988 (NAVD 88). Vertical conversions are based on the NGS program Vertcon and can be performed for the

continental U.S. only. Central Texas State plane coordinates are not only essential for geographical information analysis within West Texas but useful for obtaining distances in feet.

In order to fit crash coordinates among two corresponding intersection coordinate points these points are ranked. In the case of an Interstate Highway (IH) that follows a form of a line or curve, suppose a crash takes place, point 2, between two intersections 1 and 3 with respect to coordinates  $(X_1, Y_1)$ ,  $(X_2, Y_2)$  and  $(X_3, Y_3)$  in state plane coordinate system. As seen in Figure 4.2, the longitude and latitude coordinates of each crash report can be used to obtain a triangle between the crash or intersection coordinate  $(X_1, Y_1)$  and two reference points  $(R_{x1}, R_{y1})$  and  $(R_{x2}, R_{y2})$ . A datum reference line  $A$  is formed by the distance in feet between  $(R_{x1}, R_{y1})$  and  $(R_{x2}, R_{y2})$ . Line  $B_1$  is formed by the distance in feet between crash or intersection coordinate  $(X_1, Y_1)$  and reference point  $(R_{x2}, R_{y2})$ . Line  $C_1$  is formed by the distance in feet between crash or intersection coordinate  $(X_1, Y_1)$  and  $(R_{x1}, R_{y1})$ . Through the use of Law of Cosines angle  $\theta_1$  in degrees can be determined (Wooton et. al., 1966):

$$\theta_1(Degrees) = \cos^{-1}\left(\frac{C_1^2 + A^2 - B_1^2}{2 \cdot A \cdot C_1}\right) \quad \text{Eq.(4.1)}$$

The same procedure is followed to determine angle  $\theta_2$  in degrees where Line  $B_2$  is formed by the distance between crash or intersection coordinate  $(X_2, Y_2)$  and reference point  $(R_{x2}, R_{y2})$ . Line  $C_2$  is formed by the distance between crash or intersection coordinate  $(X_2, Y_2)$  and  $(R_{x1}, R_{y1})$ . Line  $A$  remains the same distance. Through the use of Law of Cosines angle  $\theta_2$  in degrees can be determined:

$$\theta_2(Degrees) = \cos^{-1}\left(\frac{C_2^2 + A^2 - B_2^2}{2 \cdot A \cdot C_2}\right) \quad \text{Eq.(4.2)}$$



To place the angle  $\theta_1$  and  $\theta_2$  in ranking order, the average of all latitude coordinates is obtained within all crash data for the city and highway under consideration. This average  $\bar{X}$  is used to determine the polarity of the angle such that:

$$Polarity = P = \frac{X_3 - \bar{X}}{|X_3 - \bar{X}|} \quad \text{Eq.(4.3)}$$

where,

$$Ranking \angle = 180^\circ - (P \cdot \theta_2) \quad \text{Eq.(4.4)}$$

Since the latitude  $X_i$  of crash or intersection coordinates of  $(X_1, Y_1)$  and  $(X_2, Y_2)$  are located left of the average latitude coordinate  $\bar{X}$ , Polarity  $(P) < 0$ . Angle  $\theta_1$  and  $\theta_2$  are therefore negative and:

$$Ranking \angle_1 = 180^\circ + (P \cdot \theta_1) \quad \text{Eq.(4.5)}$$

$$Ranking \angle_2 = 180^\circ + (P \cdot \theta_2) \quad \text{Eq.(4.6)}$$

where  $Ranking \angle_1 > Ranking \angle_2$

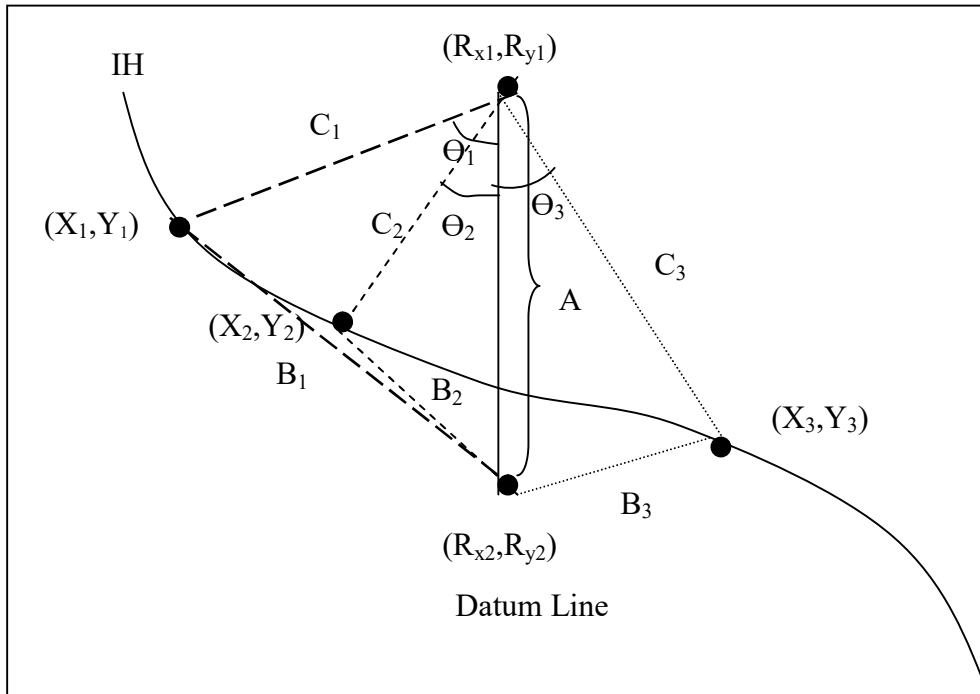


Figure 4.4: Ranking and fitting crash coordinates among two corresponding intersection points on a straight or curved interstate highway.

Through the use of Law of Cosines angle  $\theta_3$  in degrees can be determined:

$$\theta_3(\text{Degrees}) = \cos^{-1}\left(\frac{C_3^2 + A^2 - B_3^2}{2 \cdot A \cdot C_3}\right) \quad \text{Eq.(4.7)}$$

In the case of  $(X_3, Y_3)$ , the latitude coordinate  $X_3$  is located right of the average latitude coordinate  $\bar{X}$ , therefore the polarity  $(P) > 0$ . Angle  $\theta_3$  is positive and:

$$\text{Ranking } \angle_3 = 180^\circ - (P \cdot \theta_3) \quad \text{Eq.(4.8)}$$

where,  $\text{Ranking } \angle_1 > \text{Ranking } \angle_2 > \text{Ranking } \angle_3$

In the case of an Interstate Highway (IH) that follows a form of a loop or circle, suppose a crash takes place, point 2, between two intersections 1 and 3 with respect to coordinates  $(X_1, Y_1)$ ,  $(X_2, Y_2)$  and  $(X_3, Y_3)$  in state plane coordinate system (Coordinate Systems and Coordinate Transformations, 2004). As seen in Figure 4.5, similarly for a highway following a line or curve, the longitude and latitude coordinates of each crash report are taken to obtain a triangle between the crash or intersection coordinate  $(X_1, Y_1)$  and two reference points  $(R_{x1}, R_{y1})$  and  $(R_{x2}, R_{y2})$ . A datum reference line A is formed by the distance in feet between  $(R_{x1}, R_{y1})$  and  $(R_{x2}, R_{y2})$ . Line B<sub>1</sub> is formed by the distance in feet between crash or intersection coordinate  $(X_1, Y_1)$  and reference point  $(R_{x2}, R_{y2})$ . Line C<sub>1</sub> is formed by the distance in feet between crash or intersection coordinate  $(X_1, Y_1)$  and  $(R_{x1}, R_{y1})$ . Through the use of Law of Cosines angle  $\theta_1$  in degrees can be determined as in equation 4.1.

The same procedure is followed to determine angle  $\theta_2$  in degrees where Line B<sub>2</sub> is formed by the distance between crash or intersection coordinate  $(X_2, Y_2)$  and reference point  $(R_{x2}, R_{y2})$ . Line C<sub>2</sub> is formed by the distance between crash or intersection coordinate  $(X_2, Y_2)$  and

$(R_{xI}, R_{yI})$ . Line A remains the same distance. Through the use of Law of Cosines angle  $\theta_2$  in degrees can be determined as in equation 4.2.

To place the angle  $\theta_1$  and  $\theta_2$  in ranking order, the average of all latitude coordinates is obtained within all crash data for the city and highway under consideration. This average  $\bar{X}$  is used to determine the polarity as in equation 4.3:

where,

$$\text{Ranking } \angle_{Loop} = 360^\circ - (P \cdot \theta_2) \quad \text{Eq.(4.9)}$$

Since the latitude  $X_i$  crash or intersection coordinates of  $(X_1, Y_1)$  and  $(X_2, Y_2)$  are located left of the average latitude coordinate  $\bar{X}$ , Polarity  $(P) < 0$ . Angle  $\theta_1$  and  $\theta_2$  are therefore negative and:

$$\text{Ranking } \angle_{Loop,1} = 360^\circ + (P \cdot \theta_1) \quad \text{Eq.(4.10)}$$

$$\text{Ranking } \angle_{Loop,2} = 360^\circ + (P \cdot \theta_2) \quad \text{Eq.(4.11)}$$

where  $\text{Ranking } \angle 1 > \text{Ranking } \angle 2$  of the loop

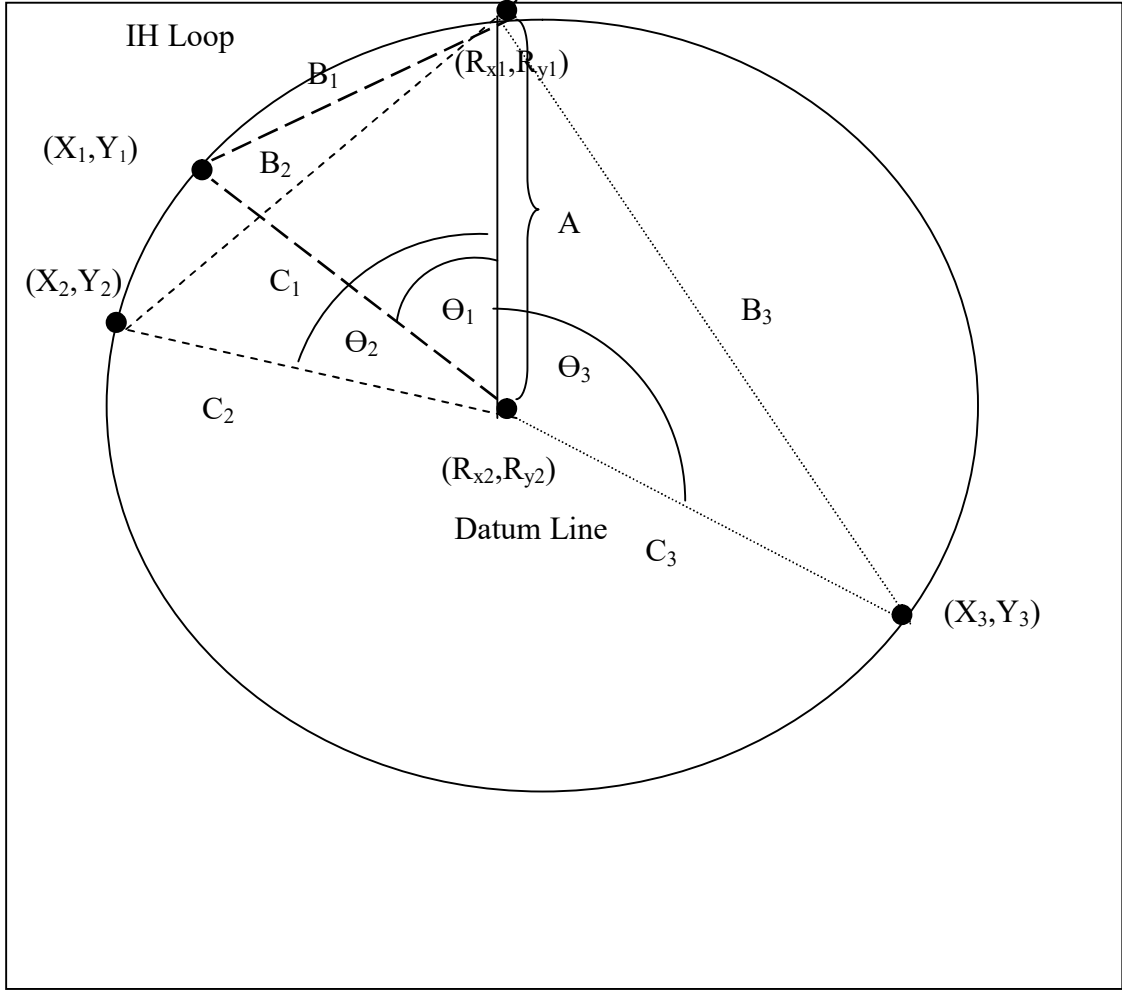


Figure 4.5: Ranking and fitting crash coordinates among two corresponding intersection points on a looped interstate highway.

Through the use of Law of Cosines angle  $\theta_3$  in degrees can be determined as in equation 4.7.

In the case of  $(X_3, Y_3)$ , the latitude coordinate  $X_3$  is located right of the average latitude coordinate  $\bar{X}$ , therefore the polarity  $(P) > 0$ . Angle  $\theta_3$  is positive and:

$$\text{Ranking } \angle_{Loop,3} = 360^\circ - (P \cdot \theta_3) \quad \text{Eq.(4.12)}$$

where, Ranking  $\angle 1 > \text{Ranking } \angle 2 > \text{Ranking } \angle 3$  of the loop.

All intersection and crash report information are ranked according to their ranking angle from  $360^{\circ}$  to  $0^{\circ}$ . Within this ranking order, segments of crashes between intersections are developed. A summation is performed of the number of crashes per intersection segment and crash report information is consolidated.

The data is then examined in two manners:

1. Examination of the frequency histogram of crashes to distance between intersections.
2. Examination of independent exposure variables that contribute to the crash rate in terms of crashes divided by AADT per distance between intersections.

## **4.2 Regression Model Specifications**

The Tobit model is a regression model in which the range of the dependent variable satisfies a constraint or is censored in a particular manner (Yoshimoto, 2008). Censoring occurs when there is data with a value at or above some threshold and much of the data concentrates on that value of the threshold. In this case, much data is available at that true value where it might be equal to the threshold, but data with higher values are plausible as well. An example is the 1980 federal law confining speed evaluations at nothing greater than 85 miles per hour. Despite vehicles with greater horsepower or better engine performance, the speed evaluations would be determined at 85 mph although vehicles may have moved at faster speeds. Therefore the only assurance is vehicles at the threshold were travelling at a minimum of 85 miles per hour (UCLA, 2016). This is considered right-censoring. The case is similar with speed data obtained by a speed gun that cannot detect speeds under 45 miles per hour on the Interstate which implies left censoring. (UCLA, 2016) In terms of crash rate data, left-censored information can occur with a collection at zero. Anastasopoulos et al., (2007) proposes this can be attributed, during the

time of observation, that crashes may have not been observed on all highway intersection segments. Given a roadway segment  $i$ , the Tobit model is articulated using a limit of zero as

$$Y_i^* = \beta X_i + \varepsilon_i \quad i = 1, 2, \dots, N, \quad \text{Eq. (4.13)}$$

where,

$$Y_i = Y_i^* \text{ if } Y_i^* > 0,$$

$$Y_i = 0 \text{ if } Y_i^* \leq 0$$

where  $Y_i$  is the dependent variable in crashes per 100 million vehicle miles traveled,  $\beta$  is a row vector of estimable parameters,  $X_i$  a column vector of independent variables in this case highway and traffic physical characteristics,  $\varepsilon_i$  is a normally and independently distributed error term with zero mean and constant variance  $\sigma^2$  and  $N$  is the number of observations. A hypothesized inherent latent variable equal to  $Y_i^*$  is observed only when positive.

The expected value  $E[Y]$  is

$$E[Y] = \beta X F(z) + \sigma f(z) \quad \text{Eq. (4.14)}$$

The term  $z = \beta X / \sigma$  is considered the  $z$ -score and a characteristic of normal distributions integral. The cumulative normal distribution function is  $F(z)$ . At a specific point the rate of change of the quantity of the normal curve is considered the standard normal density  $f(z)$ . The standard

deviation of the error term is shown as  $\sigma$ . The term  $E[Y]$  defines the relationship of the expected value of all observations with the expected value for cases above zero (uncensored),  $E[Y']$ , through the probability  $F(z)$  of values greater than zero as

$$E[Y] = F(z)E[Y'] \quad \text{Eq. (4.15)}$$

In terms of collision rates,  $Y'$  indicates observations greater than zero that are uncensored. As shown by Anastasopoulos et al. (2007), differentiation of Eq. (4.14) is used to ascertain the expected value and how it is affected by its associated independent variable. The row and column vector terms  $\beta X$  plus the expected value of the truncated normal error term provides the expected value of  $Y$  for observations greater than zero:

$$E[Y'] = E[Y|Y > 0] = E[Y|\varepsilon > -\beta X] = \beta X + \frac{\sigma f(z)}{F(z)} \quad \text{Eq. (4.16)}$$

Anastasopoulos et al. (2007) show for a particular individual variable  $X_k$  of Eq. (4.14) the first-order partial derivative is:

$$\frac{\delta E[Y]}{\delta X_k} = F(z) \left( \frac{\delta E[Y']}{\delta X_k} \right) + E[Y'] \left( \frac{\delta F(z)}{\delta X_k} \right) \quad \text{Eq. (4.17)}$$

The differential change in the total expected value defined as,  $\partial E[Y]/\partial X_k$ ; hence, the change in the expected value for cases above zero multiplied by the weighted probability of being above zero. The cumulative probability of the differential change of values greater than zero  $\partial F(z)/\partial X_k$  is multiplied by the weighted expected value of  $Y$  if above the limit and linked with an

independent variable (Anastasopoulos et al., 2007). The conditions in Eq. (4.17) can be assessed at an arbitrary value of  $\beta_k \mathbf{X}_k$ , usually at the mean of  $X_k$ . under an assumption  $\partial E[Y]/\partial X_k$  or  $\partial F[z]/\partial X_k$  has approximated values of  $\beta_k$  and  $\sigma$ .  $E[Y]$  is determined from Eq. (4.16), and  $F(z)$  through statistical procedures. Roncek's decomposition method (Anastasopoulos et al., 2007) shows the differential of the cumulative probability at zero with respect to the change in  $X_k$  as

$$\frac{\partial F(z)}{\partial X_k} = \frac{\beta_k f(z)}{\sigma} \quad \text{Eq. (4.18)}$$

The term  $\beta_k$  is the predicted Tobit parameter of variable  $k$ , and from Eq. (4.16), the differential of the expected value for cases greater than zero is

$$\frac{\delta E[Y]}{\delta X_k} = \beta_k \left( 1 - \frac{z f(z)}{F(z)} - \frac{f(z)^2}{F(z)^2} \right) \quad \text{Eq. (4.19)}$$

and the term  $z$  is the  $z$ -score for a standard normal probability density function (Anastasopoulos et al., 2007). As with a linear regression model, Tobit  $\beta_k$  should not be taken as the effect of  $X$  on  $Y_i$  and do not represent correct regression parameters for observations above the limit (Anastasopoulos et al., 2007).

Carson et al. (2007), considered instead of the condition for  $Y_i$  as described in Eq. 4.13 with the following condition:

$$Y_i^* = \beta \mathbf{X}_i + \varepsilon_i, i = 1, 2, \dots, N \quad \text{Eq. (4.20)}$$



where,

$$Y_i = Y_i^* \text{ if } Y_i^* > \gamma,$$

$$Y_i = 0 \text{ if } Y_i^* \leq \gamma$$

in which  $\gamma$  is a nonstochastic constant and as this constant is less than or equal to  $Y_i^*$ . Carson and Sun (2007) determined that by estimating an unknown threshold for the Tobit model through the minimum of uncensored  $Y_i$ 's, there is an estimator  $\hat{\gamma}$  of  $\gamma$  that is consistent and asymptotically exponentially distributed.

### 4.3 Model Goodness of Fit

A way to determine how a model fits generally and its effectiveness is through the Maddala Pseudo-  $R^2$  statistic defined by Veall et al. (1996) as :

$$pseudo R^2 = 1 - \exp \left[ \frac{2(LL(\beta) - LL(0))}{N} \right] \quad \text{Eq. (4.21)}$$

The term  $LL(\beta)$  is the log likelihood that has been maximized at convergence with the estimator parameter vector  $\beta$ , and  $LL(0)$  is the log likelihood when all parameters are equal to zero. The manner the Pseudo- $R^2$  statistic is maximized implying no increase can be achieved by altering any of the estimated parameters from zero then Pseudo- $R^2$  is zero. On the other hand if the likelihood function was estimated where every chosen parameter was sampled correctly, the estimated likelihood function  $R^2$  would be 1 (Veall et al.,1996). Per Veall et al. (1996), the range

of the Pseudo- $R^2$  is from zero to one and the closer it is to one suggests the model's results were predicted close to certainty.

#### 4.4 Likelihood Ratio Test

The equivalent likelihood function is:

$$L = \prod_0 [1 - \Phi \frac{\beta X}{\sigma}] \prod_1 \sigma^{-1} \phi (Y_i - \frac{\beta X}{\sigma}) \quad \text{Eq. (4.22)}$$

Where the Tobit model is modeled with censored observations (0) and positive non-censored observations (1) (Anastasopoulos et al., 2007). The term  $\bar{\phi}$  and  $\phi$  are the standard normal distribution function and the standard normal density function respectively.

#### 4.5 Interpretation of Variables

One way to examine a change in a parameter value on the change in the response variable is through their elasticity. Elasticity estimates the percent change on the dependent variable in terms of a one percent change in the explanatory variable (Pindyck and Rubinfeld, 1998). This can be determined as:

$$E_j = \hat{\beta}_j \frac{\bar{X}_j}{\bar{Y}} \approx \frac{\partial Y}{\bar{Y}} / \frac{X_j}{\bar{X}} \quad \text{Eq. (4.23)}$$

Where  $E_j$  is the elasticity and according to Pindyck and Rubinfeld (1998) measured at the point of the mean of each explanatory variable.  $\beta_j$  is the vector estimator for the  $j$ th observation,  $\bar{X}_j$  is the average value of the sample of the  $j$ th attribute and  $\bar{Y}$  is the average value of the dependent variable.

## 5 Model Development

As part of the objective two models were developed to determine the best possible spacing between access points that improves traffic flow and keeps traffic crashes to a minimum. In this chapter, an examination will be performed to see the similarities and differences of the data. Base models were developed using data on IH-10, IH-410 and the combined IH-10 and IH-410 data. These models were determined based on factors that are significant in predicting crash rates. Model Separation by Access Density and Flow Characteristics

### 5.1 El Paso IH-10 Data – Tobit Model

A comparison of the independent variables between an El Paso base model for crash rate data and a crash rate model as a function of access density and speed was performed. The comparison establishes if these independent variables parameters of the variables between an El Paso base model for crash rate data were estimated differently for the IH-10 interstate system. A likelihood ratio test was performed to examine if the crash rate model for access density and speed is not mis-specified when compared to the base model. Thus,

$$\chi^2 = -2[-LL(\beta_{\text{Crash Rate Model EP}}) - LL(\beta_{\text{Base}})] =$$

$$\chi^2 = -2[-21.98 - (-19.57)]$$

$$\chi^2 = 4.82$$

A  $\chi^2$  with at a 95% confidence level at 2 degrees of freedom is 5.99. Since  $\chi^2 = 4.82 < 5.99$ , this shows the El Paso Crash Rate model is mis-specified when compared to the Base model. However, it is important to note the model is approximate to being specified properly.

The corresponding Maddala Pseudo- $R^2$  statistic is:

$$Pseudo - R^2 = 1 - \exp \left[ \frac{2 \left( LL(\beta_{EP(CR)}) - LL(0) \right)}{N} \right]$$

$$Pseudo-R^2 = 1 - \exp [ (2(21.98-40.01))/36]$$

$$Pseudo-R^2 = 0.63$$

## 5.2 San Antonio IH-410 Data – Tobit Model

A comparison of the independent variables between a San Antonio base model for crash rate data and a crash rate model for access density and speed was performed. The comparison establishes if these independent variables parameters of the variables between a San Antonio base model for crash rate data were estimated differently for the IH-410 interstate system. A likelihood ratio test was performed to examine if the crash rate model for access density and speed is not mis-specified when compared to the base model. Thus,

$$\chi^2 = -2[-LL(\beta_{Crash Rate Model SA}) - LL(\beta_{Base})] =$$

$$\chi^2 = -2[-41.74 - (-39.4)]$$

$$\chi^2 = 4.68$$

A  $\chi^2$  with at a 95% confidence level at 2 degrees of freedom is 5.99. Since  $\chi^2 = 4.68 < 5.99$ , this shows the El Paso Crash Rate model is mis-specified when compared to the Base model. However, it is important to note the model is approximate to being specified properly.

The corresponding the Maddala Pseudo-  $R^2$  statistic is:

$$Pseudo - R^2 = 1 - \exp \left[ \frac{2 \left( LL(\beta_{SA(CR)}) - LL(0) \right)}{N} \right]$$

$$Pseudo- R^2 = 1 - \exp [ (2(41.74-45.2))/49]$$

$$Pseudo-R^2 = 0.13$$

### 5.3 El Paso IH-10 and San Antonio IH-410 Data Combined – Tobit Model

A comparison of the independent variables between an El Paso and San Antonio base model for crash rate data and a crash rate model for access density and number of lanes was performed. The comparison establishes that if these independent variable parameters of the variables between an El Paso-San Antonio base model for crash rate data were estimated differently for the IH-10-IH410 combination interstate system. A likelihood ratio test was performed to examine if the crash rate model for access density and speed is not misspecified when compared to the base model. Thus,

$$\begin{aligned}\chi^2 &= -2[-LL(\beta_{\text{Crash Rate Model SA-EP}}) - LL(\beta_{\text{Base}})] = \\ \chi^2 &= -2[-69.5 - (-61.6)] \\ \chi^2 &= 15.8\end{aligned}$$

A  $\chi^2$  with at a 95% confidence level at 2 degrees of freedom is 5.99. Since  $\chi^2 = 15.8 > 5.99$ , this shows the El Paso Crash Rate model when compared to the Base model is specified properly.

The corresponding the Maddala Pseudo-  $R^2$  statistic is:

$$\begin{aligned}Pseudo - R^2 &= 1 - \exp \left[ \frac{2 \left( LL(\beta_{EPSA(CR)}) - LL(0) \right)}{N} \right] \\ Pseudo-R^2 &= 1 - \exp[ (2(69.5-85.6))/49] \\ Pseudo-R^2 &= 0.48\end{aligned}$$

## 6 Model Results

### 6.1 Flow and Access Density Relationship

After an assessment of most variables within the above dataset in Table 6.1, the following are the independent variables that make up the equation of the model used for the dependent variable X20 Crash Rate:

Table 6.1: Tobit estimation of crash rate on IH-10 and IH-410

Independent Variable	Description	Estimated Coefficient	T-statistic	P value
X8	HP Shoulder RT	-0.24	-2.23	0.026
X23	Auxiliary Lanes –Merge or Present	-2.1	-2.18	0.029
X24	Number of lanes observed	1.59	2.24	0.025
X25	Access Density	0.21	1.26	0.200
X26	Speed Limit	-0.28	-3.14	0.002
One	Constant	23.0	3.30	0.001

The Tobit model was determined to be the model to provide the best representation on factors that affect the crash rate in terms of T-statistics. However, speed was noted to have a negative slope. The negative slope value specifies that the greater the speed the less the crash rate. This could be counter intuitive as most crashes are by nature attributed to high speed. Law enforcement officials typically enforce lower vehicular speed to avoid crashes. However, in a condition of non-congested free flow roads traffic speeds increase while congested roadways have the opposite effect in decreasing speeds. AADT has the tendency to increase crashes as can be seen in Table 6.2:

Table 6.2: Average Annual Daily Traffic (AADT) as a function of crash rate.

Independent Variable	Description	Estimated Coefficient	T-statistic	P value
X5	AADT	$0.35 \times 10^{-4}$	2.724	0.0064
One	Constant	4.14	5.662	0.0

Isolation of the independent variable of AADT on its own shows a positive coefficient in Table 6.2 and demonstrates that the greater the AADT the greater the crash rate. Table 6.3 shows the exponential regression estimation for the model that uses access density as the independent variable for the dependent variable X20 Crash Rate:

Table 6.3: Exponential regression estimation of access density on crash rate for IH-10 and IH-410.

Independent Variable	Description	Estimated Coefficient	T-statistic	P value
X25	Access Density	-0.04335	-1.525	0.1273
One	Constant	-2.513	-13.332	0.00

The T-statistic of access density is significant as -1.525 is greater than the test T-statistic of -1.2. The estimated coefficient of access density has a negative 4% slope where its product with a negative exponential shows a slight increase in the crash rate with an increase in access density

Table 6.4 shows the exponential regression estimation for the model that uses number of lanes as the independent variable for the dependent variable X20 Crash Rate:



The T-statistic of number of lanes is significant as -2.57 and it is greater than the test T-statistic of -1.2. The estimated coefficient of access density has a negative 27 % slope where its product with a negative exponential shows more of an increase in the crash rate with an increase in number of lanes.

Table 6.4: Exponential regression estimation on number of lanes on crash rate for IH-10 and IH-410.

Independent Variable	Description	Estimated Coefficient	T-statistic	P value
X24	Number of Lanes	-0.2784	-2.57	0.0102
One	Constant	-1.815	-4.898	0.0000

The following is the equation obtained for crash rate as a function of the number of lanes:

$$Y_{NL} = e^{-(0.2784X_{24}-1.815)} \quad \text{Eq.( 6.1)}$$

And the following equation was obtained for access density as the independent variable:

$$Y_A = e^{-(0.04335X_{25}-2.513)} \quad \text{Eq (6.2)}$$

The equation  $Y_{NL}$  for number of lanes indicates a positive relationship that shows the greater the number of lanes as a function of flow, the higher the crash rate. For access density there is a positive slope that shows that the greater the access density the higher the crash rate. The relationship is shown in Figure 6.1. This is partly attributed to the inverse relationship that length has with the crash rate response variable. This explains the lower crash rates as AADT was above 7,000 average vehicles per day for IH-10 in El Paso and IH-410 in San Antonio.

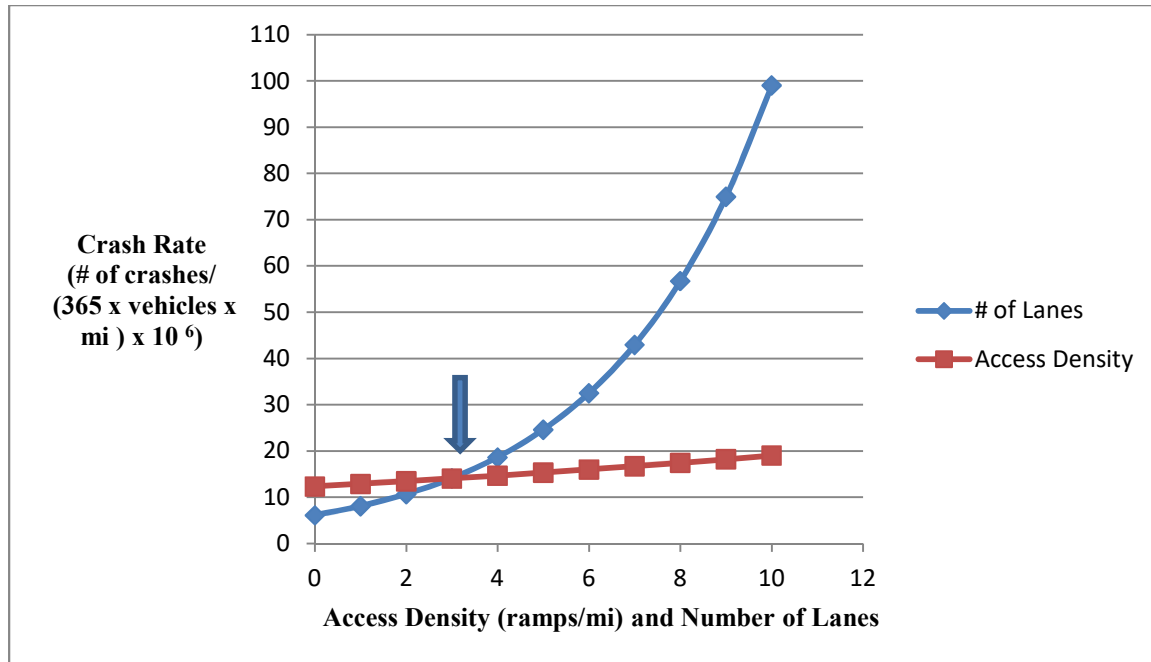


Figure 6.1: Exponential relationship of number of lanes and access density on crash rate for IH-10 and IH-410.

Figure 6.1 shows a minimum crash rate concentrated at a maximum of 3 lanes and 3 access points per mile (0.3 mi between access points).

Table 6.5: Tobit regression estimation of access density on crash rate for IH-10 and IH-410

Independent Variable	Description	Estimated Coefficient	T-statistic	P value
X25	Access Density	0.727	2.92	0.0035
One	Constant	4.96	5.49	0.00

This forms the regression equation of

$$Y_A = 0.727 X_{25} + 4.96 \quad \text{Eq. (6.3)}$$

Table 6.6 shows the Tobit regression estimation for number of lanes model used for the dependent variable X20 crash rate:

Table 6.6: Tobit regression estimation for number of lanes on crash rate for IH-10 and IH-410.

Independent Variable	Description	Estimated Coefficient	T-statistic	P value
X24	Number of Lanes	2.6	3.63	0.0003
One	Constant	0.375	0.226	0.82

Although the p-value for the constant is high, 0.82 in Table 6.6 and 0.9 in Table 6.7, it is important to note that the constant's purpose is to serve as the y-intercept. Frost (2017) explains that the p-value measures how compatible the data are with the null hypothesis that the constant equals zero. According to Frost (2017), an adequately low p-value for the constant proves the null hypothesis can be rejected and that the constant does not equal to zero where the regression line does not go through the origin. Furthermore, the high p-value indicates the null hypothesis cannot be rejected in that the constant can be zero. Frost (2017) demonstrates through a linear regression analysis example of height and weight of human beings that if the height were zero the linear regression equation would predict a negative weight. According to Frost (2017) this situation is impossible and if this situation were applied to a multiple regression analysis with various predictors it is improbable all predictors can be realistically set to zero. Frost (2017) makes clear that the constant in general has little meaning and determining if an insignificant value is different from zero has little meaning as well. In the case of the Table 6.6 and Table 6.7, it would seem plausible if the constant were zero since realistically if there were no number of lanes there would be no vehicular crashes. This would imply a crash rate of zero instead of 0.375 or 0.202 number of crashes per vehicle per mile, however per Frost's (2017) arguments there is little meaning of this constant and a road with zero lanes is no road.

From Table 6.6, the regression equation takes the form of

$$Y_{NL} = 2.6 X_{24} + 0.375 \quad \text{Eq. 6.4}$$

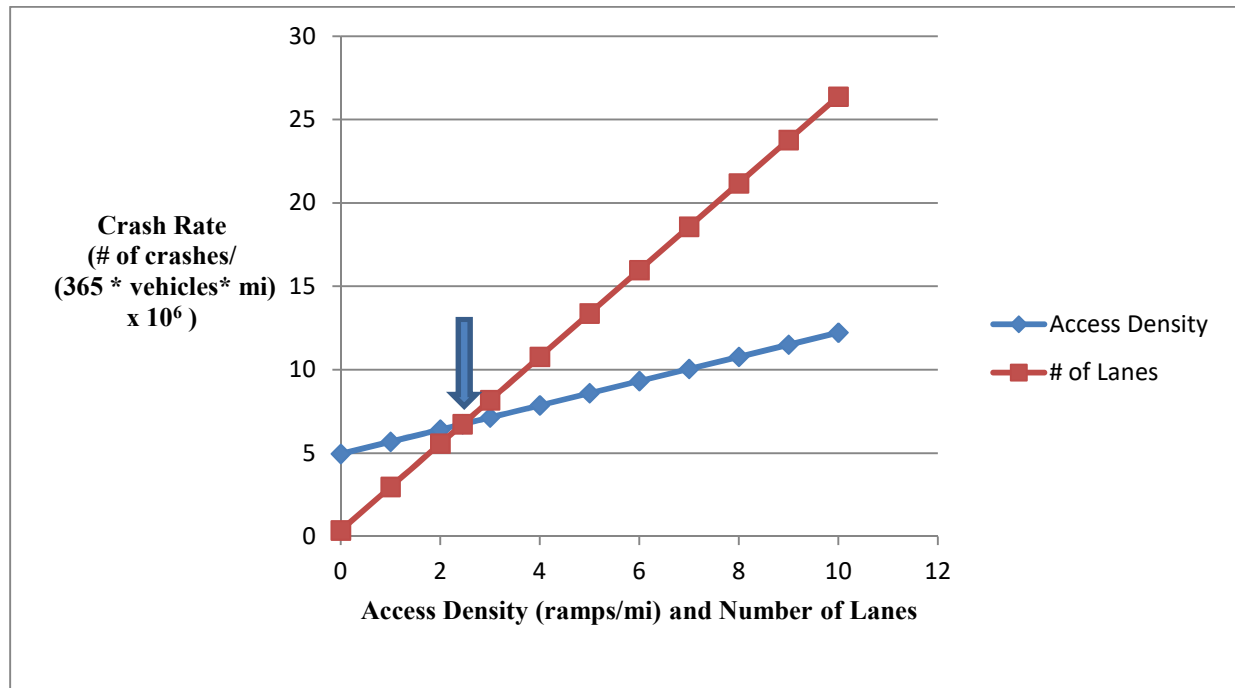


Figure 6.2: Tobit relationship of number of lanes and access density on crash rate for IH-10 and IH-410.

Figure 6.1 and Figure 6.2 shows almost exactly the same results between the Exponential regression model and the Tobit model. The minimum crash rate is concentrated at a maximum of 3 lanes and 3 access points per mile (0.3 mi between access points).

Shown in Table 6.7, the relationship between the number of lanes and access density was further examined through the development of regression equation with both factors:

Table 6.7: Tobit regression estimation of access density and number of lanes on crash rate for IH-10 and IH-410.

Independent Variable	Description	Estimated Coefficient	T-statistic	P value
X24	Number of Lanes	2.16	3.1	0.0019
X25	Access Density	0.352	1.56	0.1187
One	Constant	0.202	0.123	0.9

where,

$$Y_A = 2.16 X_{24} + 0.352 X_{25} + 0.202 \quad \text{Eq. 6.4}$$

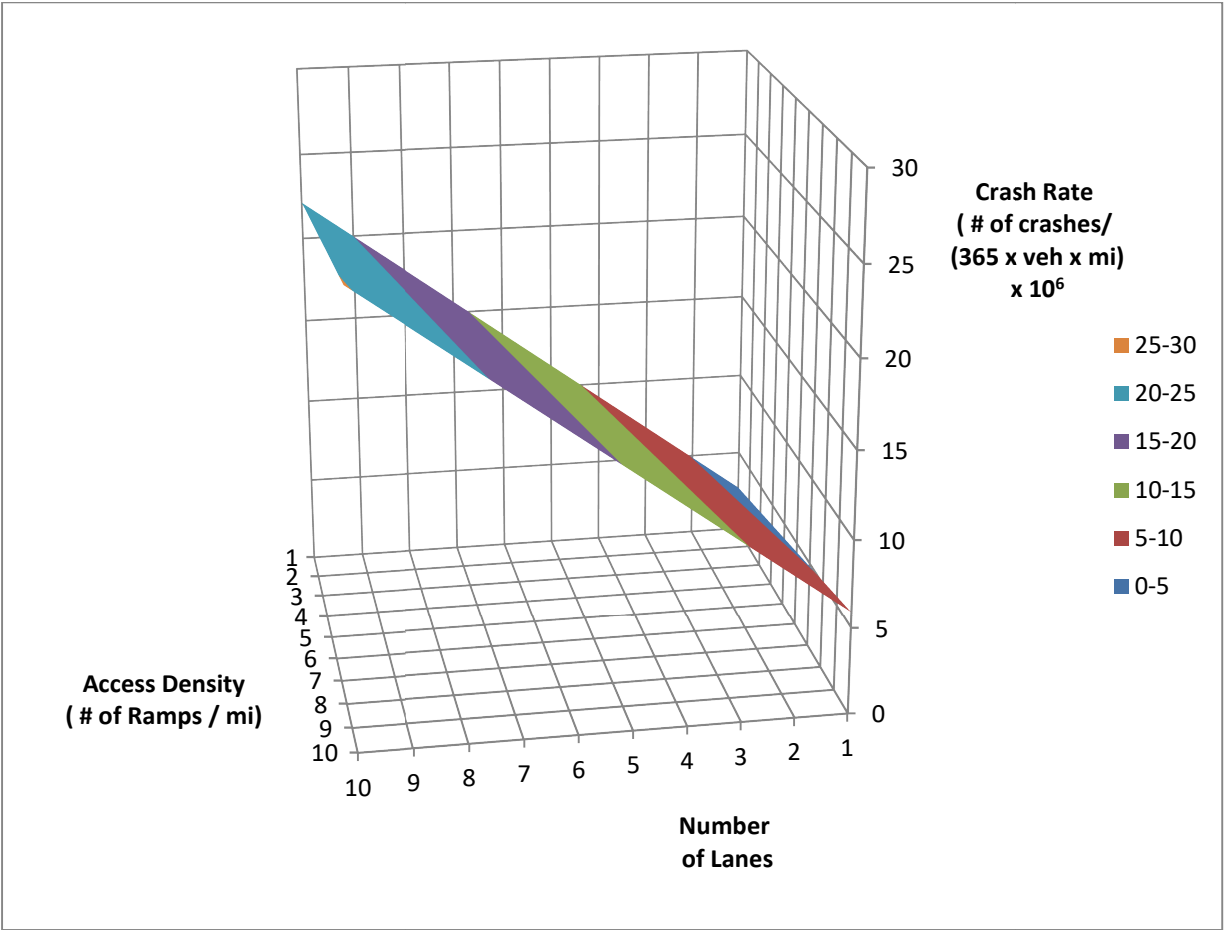


Figure 6.3: Tobit relationship of number of lanes and access density on crash rate for IH-10 and IH-410 3-dimensional.

Figure 6.3 demonstrates the relationship with both variables included in a model. It can be seen that as number of lanes increases the crash rate increases as well. Although the model shows that an increase in access density is matched with the increase in the crash rate, it is not as pronounced as the number of lanes and there is less impact with access density.

Figure 6.4 shows a comparison between crash frequency and the distance range per intersection for interstate highway 10 in El Paso. The crash frequency defines the number of crashes for all years that fall within the distance ranges or spacing of the intersections. It is apparent that as the spacing of the intersections increases the crash frequency appears to

decrease. A negative exponential appears to have a good fit. In this case, the numbers of crashes are increasing as the years proceed from 2006, 2007, 2008, 2009, 2010, 2011, 2012.

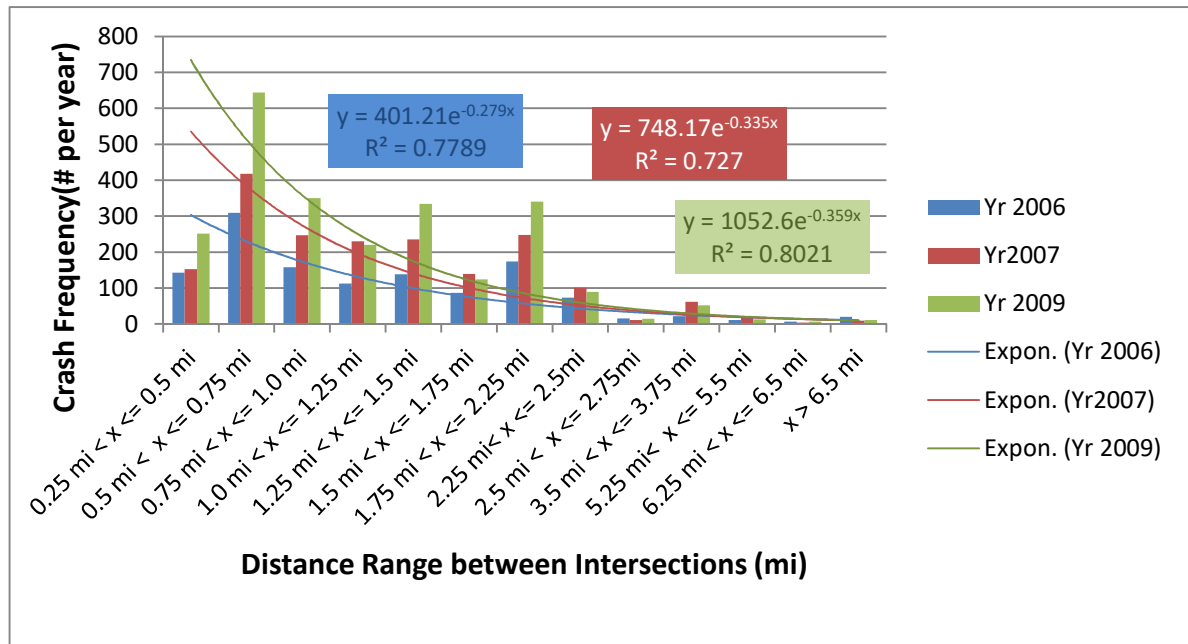


Figure 6.4: IH-10 crash frequency versus distance range per intersection in El Paso, Texas.

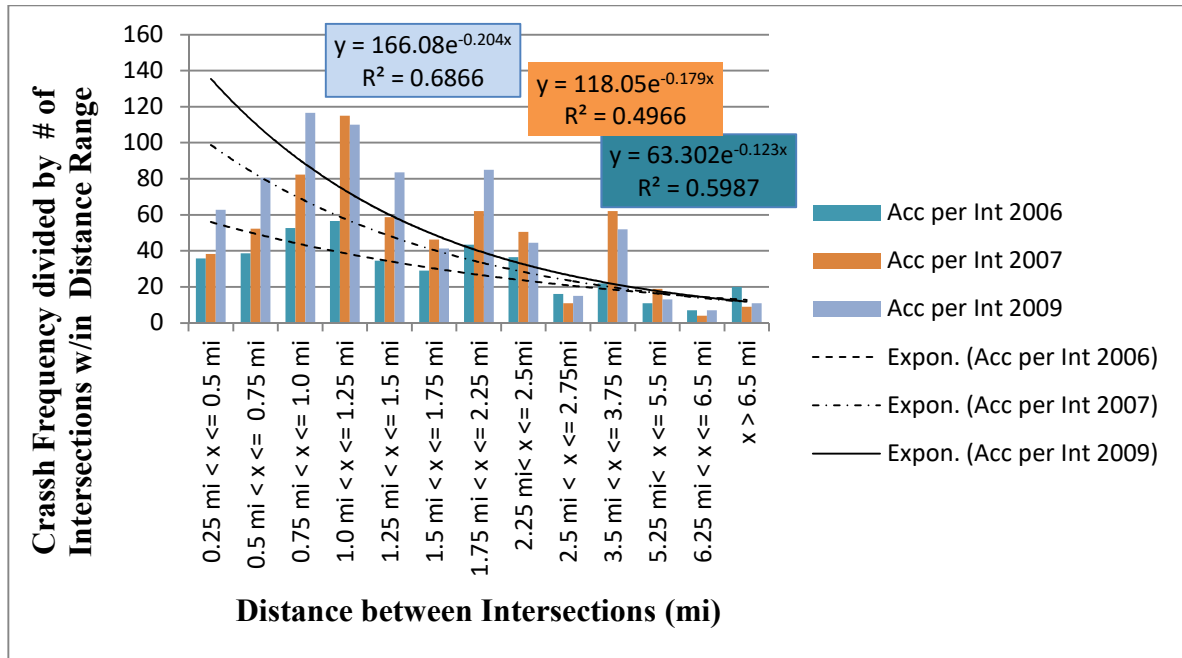


Figure 6.5: IH-10 crash frequency divided by the number of intersections within distance range versus distance per intersection in El Paso, Texas.

For every distance range the number of intersections (ramps) that fell within that range was counted. To further verify that the negative exponential model fits, the number of crashes per year was divided by the number of intersections in each segment. This provides the number of crashes per year per intersection defined as crash frequency in this section. Figure 6.5 shows the crash frequency versus the distance range. In this case the negative exponential provides a good fit with the lowest correlation coefficient of  $R^2$  equal to 0.50. This was the case for Figure 6.4 as well. The crash frequency shows an increase of crashes for every increasing year from 2006 through 2009.

In terms of San Antonio, the crash frequency is approximately twice as much as for El Paso as can be seen when comparing Figure 6.6 with Figure 6.4. However there continues to be a



good fit with a decreasing negative exponential effect as is the case for El Paso. One difference is that the number of reported crashes decreases as the years increase from 2006 through 2009.

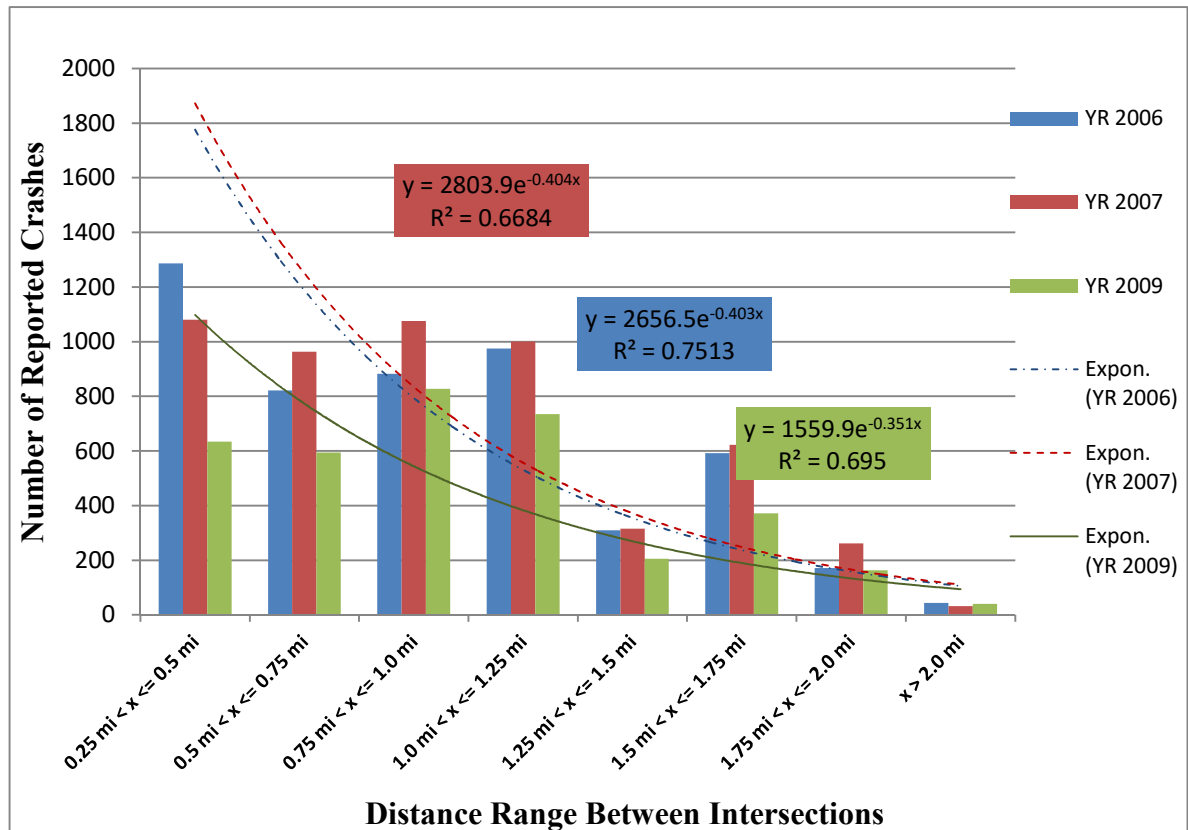


Figure 6.6: IH-410 crash frequency versus distance range between intersections in San Antonio, Texas.

A negative exponential shows a good fit (minimum  $R^2$  as 0.67) in Figure 6.6 for San Antonio IH-410 crash frequency versus the distance range between intersections.

Figure 6.7 shows the number of crashes per intersection per mile. In this case the negative exponential provides a good fit (minimum  $R^2$  as 0.51). This was the case for Figure 6.4 as well. Again, contrary to El Paso's case, the number of crashes per intersection shows a decrease of crashes for every year from 2006 through 2009.

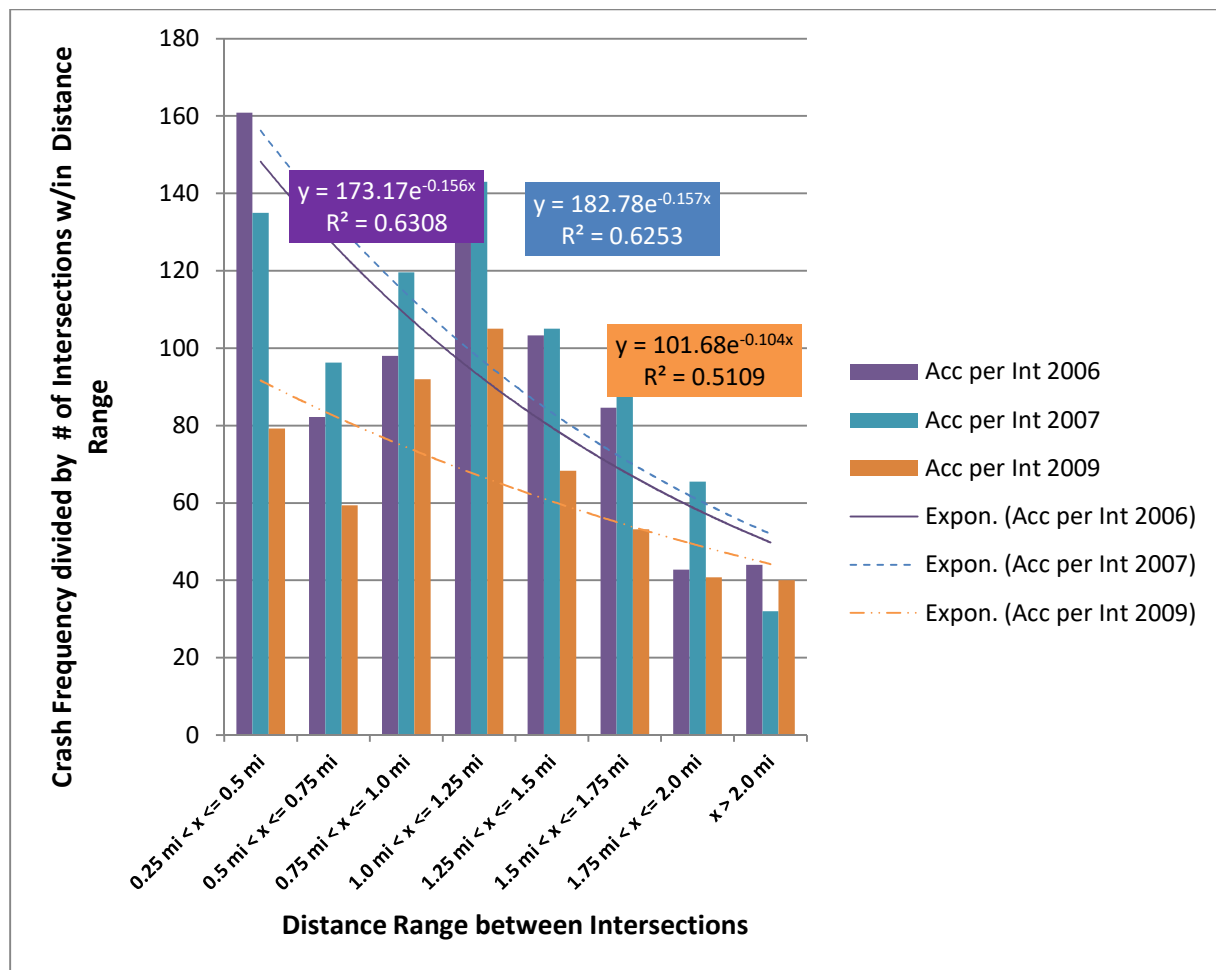


Figure 6.7: IH-410 crash frequency divided by number of intersections within distance range versus distance per intersection in San Antonio, Texas.

## 6.2 Roadway Characteristics

Table 6.5 Shows the Tobit base regression estimation of crash elasticities per 100 million VMT as follows:

Table 6.8: Tobit censored elasticity estimation of crash rate on IH-10 and IH-410.

Independent Variable	Description	Estimated Elasticity
X8	HP Shoulder RT	-0.12
X23	Auxiliary Lanes – Present or Not	-0.07
X24	Number of lanes observed	0.37
X25	Access Density	0.08
X26	Speed Limit	-1.29

Table 6.8 shows the Tobit base regression estimation of elasticities of crashes per 100 million VMT. The independent variable X8, right shoulder width in feet, was considered as part of the model since the presence of a shoulder and its width would have an effect to merge to the right shoulder especially if a driver experiences mechanical or compartment problems with the vehicle. The negative sign shows that a greater width is more likely it is to decrease crashes. The greater the width the greater the opportunity the driver feels he can drive to the right in the event he encounters a problem on the road and avoid a crash. Table 6.8 shows that for every one percent increase in right shoulder width there is a 0.12 percent decrease in the crash rate.

The independent variable X23, auxiliary lanes as present or non-present, was examined due to its effect on maintaining a lane that connects an exit ramp with the next entrance ramp without the ramp lane merging onto the main lanes. This provides vehicles with a longer time to transition onto the main lanes. The negative coefficient it has the effect of reducing crashes which is expected since a greater transition ramp length provides less of an opportunity for vehicles to collide as a vehicle exits the auxiliary lane and another exits the main lanes onto the

auxiliary lane. The driver does not have to react as fast to oncoming traffic. Table 6.8 shows that for every one percent increase in the presence of an auxiliary lane there is a 0.07 percent decrease in the crash rate.

The independent variable X24, the number of lanes, was chosen as independent variable since the number of lanes is believed to have an effect on crashes. In certain cases a greater number of lanes cause an increase in crashes. This is a contributing factor for vehicular flow. The greater the average annual daily traffic the more number of lanes are added to keep vehicles flowing as pointed out in Section 6.1 and Table 6.2. Flow rate and number of lanes play a role to a higher weaving of vehicles entering and exiting at different locations. In addition, a greater number of lanes cause a greater number of conflict points. Table 6.8 shows that for every one percent increase in the additional lanes there is a 0.37 percent increase in the crash rate. The positive value shows that the greater the number of lanes the greater the crash rate.

The independent variable X25, access density or the number of ramps per mile was observed to increase the number of crashes but at a gentle rate. As a manner of conjecture, access density was considered to contribute to more of a sharp increase in the number of crashes due to the degree of spacing between intersections. An inference made is the closer the spacing, the greater the difficulty for a vehicle to weave in and out of a highway system and the greater the spacing the less difficult for a vehicle to enter and exit the highway system. However, it is believed that in a saturated congested highway there is a tendency for vehicles to drive slower. This provides an opportunity for vehicles to enter and exit the freeway smoothly despite narrower vehicle spacing. Table 6.8 shows that for every one percent increase in access density there is a 0.08 percent increase in the crash rate. The positive value shows that the greater the access density the greater the crash rate.

The independent variable X26, speed limit was found to be negative where a decrease in the crash rate occurs at higher the speed limits. Table 6.8 shows that for every one percent increase in speed there is a 1.29 percent decrease in the crash rate. There is a difficulty in envisioning this relationship as higher speeds cause crashes due to drivers' greater likelihood of failing brake reaction times to objects or vehicles ahead. However, areas with less AADT tend to increase speeds and areas with more AADT decrease speeds. With fewer vehicles on the roadway, drivers have less chances of an involvement in a crash despite the higher speeds. With more vehicles on the roadway there are higher chances of being involved in a crash despite the lower speeds.

## 7 Conclusions

### 7.1 Research Summary

The statistical and econometric modeling as described in this study has examined the relationship between geometric features and their influence on crash rates. The econometric crash models developed helped to identify effects of traffic flow and access density on crashes. The analysis performed investigated various potential factors involved in crashes that were filtered through the econometric regression analysis. It was through examination of their combination and effects that a final model was established. Variables with exposure to crashes such as geometric static features were used since they do not change as intermittently over time. Variables such as environmental factors (i.e. light condition and weather) are erratic, change sporadically over time and would require greater sampling. These geometric characteristics such as access density, number of lanes, median widths, roadbed widths and other exposure variables and their effects were observed. The chief variables in question were access density and number of lanes, which are related to a flow characteristic. It was through this analysis and a demonstration of the use of the Tobit model and exponential regression that an approximate determined spacing was found where flow and access density could be at their best possible combination without greatly compromising the crash rate. A compilation of seven years of CRIS data was collected for IH-10 in El Paso and IH-410 in San Antonio. Of all crash information in Texas, the study focused on crash data that was filtered for only El Paso or San Antonio. Trigonometric functions and polar coordinates intersections were used to obtain distances of latitude and longitude coordinates for corresponding intersections and crash locations. The angles and distances were used to rank and place in order intersection and crash locations to be able to identify the number of crashes found between two intersection for a given interstate and

year. This was performed for all intersection in IH-10 in El Paso and Loop IH-410 in San Antonio. Each crash pertained to a unique identifier or case number that provided all CRIS environmental, geometric, state of the driver and other factors that were observed during the time of the crash. This information was used to statistically develop an econometric base model. Further examination and elimination of factors from the base model was carried out to assess the development of an effective crash rate model for El Paso and San Antonio Interstate systems. The base model with significant factors included right shoulder width, presence of auxiliary lanes, number of lanes, access density and speed. Tobit and exponential regression provided fitting models to ascertain crash rates given different factors based on the assemblage of crashes on IH-10 and IH-410 segments. The Tobit regression model through analysis using LIMDEP, statistical tests in terms of the T-test and Log –Likelihood test were used in proving the models’ validity. The development of the crash rate model proved useful to assess the effects of crash rates as a direct relationship with flow characteristics and access density. This model establishes a methodology for the assessment of access point spacing and traffic flow.

## **7.2 Research Limitations**

That data obtained for this study was collected through various crash reports provided by police officers across the State of Texas. It was observed that various officers collected information differently than others that introduce bias. Some information was entered fully where other information was left blank. Some information appeared to be entered with precision in terms of the appropriate latitude and longitude information while other information appeared to be estimated and clustered at the overpasses. This estimation was evident as many crash locations centered only at the location of the intersection and did not show an accurate dispersion among the segment. For this reason estimation determination in terms of how crash locations

varied within various distances between segments was not highly accurate. For instance, if out of 100% crashes, 100% crash locations were well identified in terms of latitude and longitude coordinates statistical patterns could be obtained to examine if certain geometric features had a definite effect on crashes. However, in many cases, the data showed many crashes all centered at the intersection and the remaining dispersed within the segment. This led to bias results in terms of crash locations providing inaccuracies to the geometric feature in question. This places an issue for applying methods such as Kernel Density Estimation that assists in determining accident hotspots as it may relate to environmental and landuse factors to identify crash causes. Despite this limitation in terms of inaccuracies, the CRIS crash location data provided a good indication of the effects of various geometric features on crash rates. A solution to providing more accurate police report crash location information is the development of a better standard automatic protocol that allows to quickly obtain information and limits greatly the amount of information that needs to be manually entered. Devices and software can be made available or reconstructed that can automatically obtain weather, road pavement indices, geographic locations, and other crucial crash information.

A more accurate model could be obtained through analysis performed in another Texas city or cities or cities in other states having a similar database to CRIS can help to further develop the validity of the model. Data that could be obtained from a city such as Houston can be included as part of the San Antonio-El Paso dataset. This city offers two interstate loops in urban and less population dense areas. The loops and road systems in Houston are subject to various geometric features and access spacing points that could offer greater T-values and higher likelihood values for various other factors not taken into consideration.



### **7.3 Research Contributions**

As part of a research contribution made is the analysis performed on how to assess the quantity of crashes that occurred between each intersection or segments and the factors (i.e., geometrics, lane width, etc.) obtained from crash reports that contributed to those crashes within those intersections. Each crash coordinate in terms of latitude and longitude coordinates were fit among two corresponding intersection coordinate points. A unique method for ranking was used by polar coordinates to fit the coordinate points within each intersection segment.

### **7.4 Further Research**

Since this study establishes a methodology to ease the transition of vehicles entering or exiting an interstate or highway system, it could be extended in the research of ramp metering or when vehicles can be allowed into a highway system through a traffic signal. Similar research and practical engineering designs were applied to synchronization gear similar to a gun synchronizer or interrupter used in World War I and II airplanes. Suppose the propeller moves similar to fast platoons of high speed traffic and the bullet shooting past the propeller is an entering or exiting vehicle. The interrupter of the plane would be similar to smart vehicle with vehicle sensing technology to allow the vehicle to move into high speed traffic without crashing other vehicles. This is similar to firing a bullet through a rotating propeller without the bullets colliding through the blades. With the advancement of smart vehicles, vehicles can interrupt the platoon-ing of traffic without causing accidents and significantly hindering the flow and velocity of traffic. Additional research in access density to traffic flow for ramp metering can further assist traffic to seamlessly enter into interstate or highway with less potential for vehicular encroaching activity leading to potential crashes.

## References

- Anastasopoulos, P.C., Tarko, A.P., Mannering, F.L. (2007). "Tobit analysis of vehicle accident rates on interstate highways." *Accident Analysis and Prevention*, 40(2008) 768-775.
- Anderson, J., and Hernandez, S. (2016). "Heavy Vehicle Crash Rate Analysis: A Comparison of Heterogeneity Methods Using Idaho Crash Data." *Transportation Research Record, Journal of the Transportation Research Board*. Accepted for publication.
- Anderson T.K. (2008). "Kernel density estimation and K-means clustering to profile road accident hotspots." *Accident Analysis and Prevention*, 41, 359-364.
- Carson, R.T., Sun, Y., (2007). "The Tobit Model with Non-zero Threshold." *Econometrics Journal*, Vol. (00).1-15.
- Coordinate Systems and Coordinate Transformations*. (2004). Retrieved from <http://ads.harvard.edu/books/1989fcm..book/Chapter2.pdf>
- Cullip, M.J., Hall, F.L. (1997). "Incident detection on an arterial roadway." *Transportation Research Record 1603*, Transportation Research Board. Washington DC. 112-118.
- Disney, R.L., Clark, B.A., (1985). *Probability and Random Processes: A first course with applications 2<sup>nd</sup> Ed*. John Wiley and Sons: New York.
- Federal Highway Administration. (1998). *Accident for two-lane rural roads: segment and intersections, FHWA-RD-98-133*.
- Frost, J. (2017). *Regression Analysis: How to Interpret the Constant (Y Intercept)*. Retrieved from <http://blog.minitab.com/blog/adventures-in-statistics-2/regression-analysis-how-to-interpret-the-constant-y-intercept>
- Greene, W.H. (2012). "NLOGIT Version 10.0 Econometric Modeling Guide". *Econometric Software Inc.*, New York.
- Hanowski, R.J., Medina, A.J., Wierwille, W.W., Lee, S.E. (2004). "Incident Clustering: Diagnostic Approach for Assessing Usability of Intersections and Other Road Sites." *Statistical Methods and Safety Data Analysis and Evaluation*, 697, 173-179.

- Islam, M.B., Hernandez, S. (2015). “Fatality rates for crashes involving heavy vehicles on highways: a random parameter Tobit regression approach.” *Journal of Transportation Safety and Security*, Taylor& Francis, London.
- Konduri, S., Labi, S., Sinha, K.C.(2003). “Incident Occurrence Models for Freeway Incident Management.” *Transportation Research Record 1856*, Transportation Research Board, Washington, DC, pp125-134.
- Montgomery, D.C. (2012). *Introduction to Statistical Quality Control 7<sup>th</sup> edition*. John Wiley and Sons: New York
- Mouskos, Kyriacos, C.,Sun, Wu, Qu Tao,.(1999). “Impact of Access Driveways on Accident Rates at Multilane Highways.” *NJ-99-008-NCTIP5*, New Jersey Department of Transportation.
- NHTSA.(2015). “The Economic and Societal of Motor Vehicle Crashes.” National Highway Traffic Safety Administration. Washington, D.C.
- Pindyck, R.S., and Rubinfeld, D.L. (1998). *Econometric Models and Economic Forecasts 4<sup>th</sup> Ed.* McGraw-Hill, Boston.
- Plug, C; Xia J., Caulfield, C. (2011), “Spatial and temporal visualization techniques for crash analysis.” *Accident Analysis and Prevention; Elsevier* 43(2011) 1937-1946.
- Sayed, T; Rodriguez, F.. (1999). “Accident Prediction Models for Urban Unsignalized Intersections in British Columbia, *Transportation Research Record 1665*, Transportation Research Board, Washington, D.C.
- Sinha, K.C.;Labi,S.; (2007) *Transportation Decision Making: Principles of Project and Programming*. John Wiley and Sons.
- Stahel, W., Limpert E., (2014) *The normal distribution is the log-normal distribution*. Retrieved from <http://stat.ethz.ch/~stahel/talks/lognormal.pdf>

- Tarko, A.P., Eranky, S., Sinha, K.C., Scinteie, R., (1999) "An Attempt to Develop Crash Reduction Factors Using Regression Technique", *78<sup>th</sup> Transportation Research Board Annual Meeting*, Transportation Research Board, pp1-19.
- University of California at Los Angeles (2017). "Tobit Analysis." Retrieved from <http://stats.idre.ucla.edu/stata/dae/tobit-analysis/>
- U.S Army Corp of Engineers. (2004). Corpscon 6.x Technical Documentation and Operating Instructions. Engineer Research and Development Center, Alexandria, VA.
- Wang, C., Quddus, M.A., Ison, S.G. (2009). "Impact of traffic congestion on road accidents: A spatial analysis of the M25 motorway in England." (2009). *Accident Analysis and Prevention*, 41, 798-808
- Washington, S.P;Karlaftis, M.G;Mannering, F.L (2011) *Statistical and Econometric Methods for Transportation Data Analysis*. Chapman and Hall/CRC. New York,
- Wooton, W., Beckenbach, E.F., Dolciani, M.P. (1966). *Modern Trigonometry*. Houghton Mifflin Co. Boston.
- Yoshimoto, H. (2008). "Tobit Models." Retrieved from [http://hisayoshimoto.bol.ucla.edu/203c/ta\\_note8\\_version2\\_student.pdf](http://hisayoshimoto.bol.ucla.edu/203c/ta_note8_version2_student.pdf)

## **Vita**

George Pinal was born on July 9, 1971 in Milwaukee, WI. He graduated from Coronado High School, El Paso, Texas in the spring of 1990 and entered The University of Texas at El Paso in the fall of that year. While pursuing a bachelor's degree in Industrial Engineering, he worked as an intern with the City of El Paso Metropolitan Planning Organization in 1995. In the fall of 1996 he earned a Masters in Civil Engineering degree with the support of the Dwight David Eisenhower Transportation Fellowship. After earning his degree, he obtained his Engineer-In-Training Certificate, worked as a quality engineer for Elcom Inc. and as a quality assurance superintendent for Epson, Inc. In 2005, George returned to the El Paso Metropolitan Planning Organization where he worked as a Transportation Manager. In 2006, he re-entered the Graduate School at The University of Texas at El Paso to pursue a doctorate degree in civil engineering. He has vigorously worked in the areas of planning and civil engineering.

Permanent address: 6209 Loma De Cristo  
El Paso, Texas, 79912

This thesis/dissertation was typed by George Pinal