# The relationship between pedestrian crash rates and injuries sustained and the location, time of day, traffic volume and the hypothetical travel speed at the time of the crash on the Gauteng Freeways

**By:**

**Janneke Snijder**

*The thesis presented in fulfilment of the requirements for the degree of Master of Engineering in the Faculty of Civil Engineering at Stellenbosch University*

**Supervisor: Prof. Marion Sinclair**

**December 2019**

**Declaration**

By submitting this thesis/dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

**Date: December 2019**

**Abstract**

Informal settlements are situated adjacent to the freeway, resulting in pedestrians that cross and walk along the South African freeways regularly. This phenomenon is due to the historical and current demographics, where certain racial groups were forced to move out to the city perimeter. The current economic situation results in informal settlements that grow rapidly. The low-income areas are not provided with proper transport facilities, forcing the residents of the settlements to walk to their destinations. Research indicates that pedestrians tend to walk the shortest path to their destination, resulting in pedestrians that cross the freeway to get to their destinations. The aim of this study is to determine the impact of the land use surrounding pedestrian crashes, the day and time on which the pedestrian crash occurred, the traffic volume and vehicle speed, the vehicle type involved in the crash as well as the number of lanes to cross have on the severity as well as the crash rate of the pedestrian crashes on the Gauteng FMS network.

The injuries sustained in the pedestrian crashes was modelled using a binomial logistic model. The crash rate of the pedestrian crashes was modelled using a multinomial logistic model. It was determined that the independent variables for the two models are different, and a clear relationship between the two models was found. The traffic volume on the road, the average travel speed, the day of the crash, the vehicle type involved in the crash as well as the number of lanes to cross had an impact on the probability to sustain fatal injuries in a pedestrian crash. The probability of sustaining fatal injuries increase during the weekend, during the off and night peaks, when the travel speed increase, when the number of lanes to cross increase as well as for the case when a heavy vehicle is involved in the crash. The probability of sustaining fatal injuries decreases as the number of vehicles on the road decreased. The crash rate, on the other hand, was dependent on the traffic volume, vehicle speed, land use, day of the crash, as well as the peak hour in which the crash occurred. It was, however, also determined that the land use attribute might cause this model to predict biased results, and this model was therefore not regarded as an accurate model. The injury model was regarded as a good model, and it was concluded that this model could be applied in other locations in South Africa.

## Opsomming

Informele nedersettings is geleë aangrensend aan die snelweg, wat lei tot voetgangers wat op ‘n gereelde basis die Suid-Afrikaanse snelweë oorsteek en langs die snelweë loop. Dit is as gevolg van die historiese en huidige demografie waar sekere rassegroepe gedwing is om na die stad grense te skuif. Die huidige ekonomiese situasie lei tot informele nedersettings wat vinnig groei. Die lae-inkomste gebiede is nie voorsien met 'n goeie vervoer fasiliteite nie, en dwing die inwoners van die nedersettings om na hul bestemmings te loop. Navorsing dui daarop dat voetgangers geneig is om die kortste pad na hul bestemming te loop, wat lei tot voetgangers wat die snelweg oorsteek om hul bestemmings te bereik. Die doel van hierdie studie is om te bepaal wat die impak van die grondgebruik rondom voetgangerongelukke, die dag en tyd waarop die voetgangerongeluk plaasgevind het, die verkeersvolume en voertuig se spoed, die tipe voertuig wat betrokke is in die ongeluk asook die aantal lane om oor te steek op die erns van die voetgangerongeluk asook op die ongelukskoers van die voetgangerongelukke op die Gauteng FMS netwerk het.

‘n Binomiale logistiese model is ontwikkel vir die beserings wat opgedoen is in die voetgangerongelukke en 'n multinomiale logistiese model is ontwikkel vir die ongelukskoers model. Daar is vasgestel dat die onafhanklike veranderlikes vir die twee modelle verskil en dit het tot die gevolgtrekking gelei dat daar nie 'n duidelike verband tussen die twee modelle is nie. Die verkeersvolume op die pad, die gemiddelde reisspoed, die dag van die ongeluk, die tipe voertuig wat betrokke is in die ongeluk asook die aantal lane om oor te steek het 'n impak op die waarskynlikheid om noodlottige beserings in 'n voetgangerongeluk op te doen. Die waarskynlikheid vir noodlottige beserings verhoog tydens die naweek, tydens die af en nag pieke, wanneer die voertuig se spoed toeneem, wanneer die aantal lane om oor te steek verhoog sowel as vir die geval wanneer 'n swaar voertuig betrokke is in die ongeluk. Die waarskynlikheid vir noodlottige beserings verminder as die aantal voertuie op die pad afgeneem. Die ongelukskoers, aan die ander kant, was afhanklik van die verkeersvolume, spoed van die voertuig, grondgebruik, dag van die ongeluk, sowel as die piek uur waarin die ongeluk plaasgevind het. Dit is egter ook bepaal dat die grondgebruik kan veroorsaak dat hierdie model onakkurate resultate voorspel en hierdie model is dus nie as 'n akkurate model beskou nie. Die beseringsmodel was beskou as 'n goeie model en dit is tot die gevolgtrekking gekom dat hierdie model in ander plekke in Suid-Afrika toegepas kan word.

## Contents

**List of Figures**



**List of Tables**

**List of Graphs**

# 1. INTRODUCTION

Walking forms part of all trips that are made by any transport user. Regardless of the other transport modes that will form part of the trip, each person has to walk from their origin to their mode of transport to reach their destination. According to the World Health Organisation (2018), road traffic deaths are the eighth highest cause of death for people of all ages and the main leading cause of death for children and young adults. It is estimated that more than half of the road traffic deaths are amongst vulnerable road users, namely pedestrians, cyclists and motorcyclists. A total number of 40% of the road traffic deaths in Africa are pedestrians, as indicated in Figure 1 (WHO, 2018).

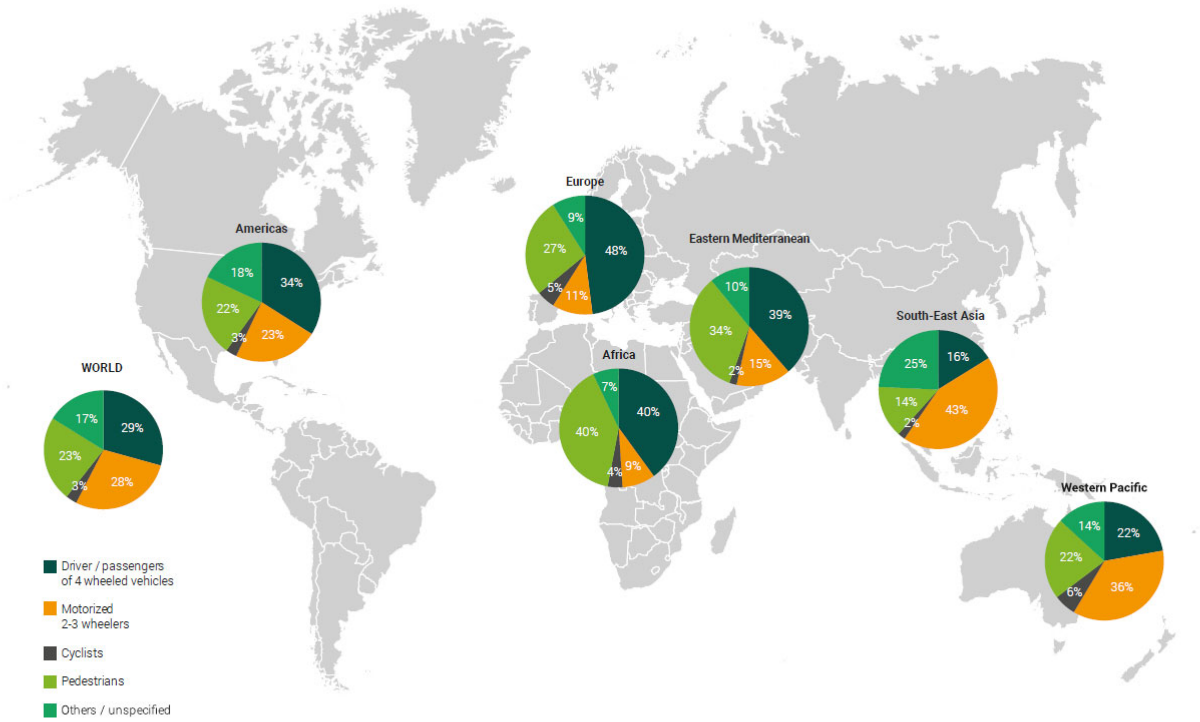



*Figure 1: Percentage split of road traffic deaths (Source: WHO 2018)*

It is therefore evident that pedestrians are at high risk on the African roads. Elements such as traffic volume, travel speed, the height of the pedestrian, vehicle type, and pedestrian accommodation along the road play a significant role in the injuries sustained in pedestrian crashes.

Road traffic deaths are in the top three causes of deaths for people aged 5 to 44 years. These are the years at which one get educated and start to earn an income. These people play an essential role in the country's economy, and it is, therefore, important to address this problem. These deaths do not only have a negative impact on the future economy and social growth of the country, but it also cost the country a lot of money in the present.

According to a study conducted by Sinclair and Zuidgeest, (2016), South Africa is one of the only countries where there is high pedestrian activity along the freeways. This phenomenon is partly due to the Apartheid regulations where non-white residents were forced to live outside the white suburban areas and therefore had to move out to the city perimeters. Urbanisation and poverty also play a role in growing informal settlements located alongside the South African freeways.

Research indicated that the impact speed of the vehicle that crashes with a pedestrian plays a significant role in the injuries sustained. The freeway is not designed to accommodate pedestrians, and this makes it an unsafe environment for pedestrians. Pedestrians that walk along or cross the freeway regularly are expected to be at high risk of being involved in a serious or fatal pedestrian crash. The higher the travel speed, the longer time a vehicle needs to come to a complete stop. High travel speeds in areas where there is major pedestrian activity can result in a high number of pedestrian crashes since the vehicle will not have sufficient time and distance to stop when there is a potential conflict between the vehicle and a pedestrian. In addition to this, at higher travel speeds and multi-lane flows on freeways, it becomes more difficult for pedestrians to determine a safe gap in traffic.

The South African National Roads Agency Limited (SANRAL) aims to improve the safety of the Freeway Management System (FMS) network by using intelligent transport solutions. These measures are designed to improve congestion on the freeway, to improve the road safety on the network and also to improve the incident response timelines. This will all have a positive influence on the pedestrian crashes that happen on the freeway; namely, the crashes will be detected and reacted to in a shorter time, which is sometimes critical for an injured person. The crashes will be reduced when road safety has improved, and when the congestion levels of the network have improved.

Real-time traffic data is obtained by using CCTV cameras that are located at strategic points on the road network. These cameras are used as a surveillance tool to survey the road network for any safety and security or any other road incident (Bester, G. 2015). These measures are expected to help improve the safety and traffic congestion on the freeways, however, without a complete understanding of the human behaviour and the factors that play a role in the pedestrian crashes along the freeway, pedestrian safety and crash factors will remain a problem.

Pedestrians are the most vulnerable road users on the road, especially along the freeways where no pedestrian facilities are provided and where vehicles travel at high speeds. It is therefore seen as an important topic to address and do research on.

# 2. LITERATURE REVIEW

## 2.1 INTRODUCTION

Historic crash data indicates that the number of crashes and the number of road traffic deaths have increased in the past few years in South Africa (StatsSA, 2009). The increased number of road traffic deaths are partly due to a constant need for transportation, and it is expected that this need will increase in the future.

Crash statistics from the Road Traffic Management Corporation (RTMC, 2017) indicated that there had been a decrease of 0.4% in pedestrian fatalities from 2016 to 2017. Pedestrian fatalities, however, still contribute to a tremendous 38% of all road traffic fatalities.

Various factors were determined to influence pedestrian crashes. These factors include the pedestrian and traffic volume on the road segment, the cross-section of the road, the speed at which the vehicle is travelling, the land use in the surrounding area as well as the pedestrian behaviour, (Oxley *et al.*, 2018).

There is a growth in traffic volumes on the freeways and in the urban areas increase every year. The number and size of the informal settlements also increase, resulting in an increase in the number of people waiting at and walking to public transport facilities. This results in an increase in the risk of pedestrian crashes on the roads in and surrounding cities and towns.

The risk of pedestrian crashes that occur on the side and shoulders of a road increases when the speed limit increases, with higher traffic volumes as well as when insufficient shoulders and sidewalks are provided along the road (Zegeer and Bushell, 2012).  The severity of pedestrian crash injuries depends on factors such as vehicle type, pedestrian age and impact speed. Figure 2 gives an indication of the points of impact for an adult pedestrian when hit by a vehicle (Zegeer and Bushell, 2012, Arrive Alive, 2019).

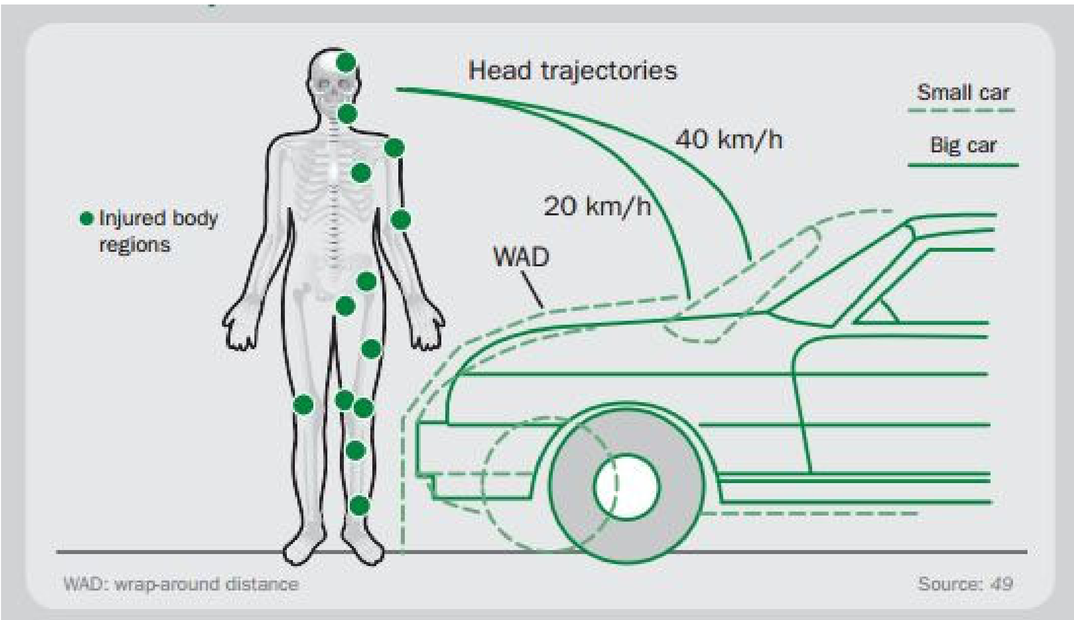



*Figure 2: Impact points of an adult human body in a pedestrian crash [Source: Arrive Alive]*

The different parts of a human body can sustain different types of impact. The injuries can be related to the severity of the injuries sustained in a pedestrian crash. This threshold is categorised into four categories (Kröyer, 2015; Martin and Wu, 2018):

1. Properties of the pedestrian: age, weight, height, the direction in which pedestrian was hit;
2. Features of the vehicle: shape, weight, the height of the vehicle;
3. Impact speed affecting the impact at which a vehicle hit the pedestrian, and the pedestrian hit the ground; and
4. Force sustained by the pedestrian body when it hit the ground or another object.

## 2.2 HUMAN BEHAVIOUR

Sinclair and Zuidgeest, (2016) and Dada, Zuidgeest and Hess, (2019) studied human behaviour to understand why pedestrians cross the freeway at the place they do. Factors such as gender, age, safety, comfort, and most direct path were found to play essential roles in the pedestrian's decision on where and how they want to cross the freeway. The safety risk is a factor that plays a significant role in the pedestrian's choice of where to cross the freeway. This safety factor includes both criminal activities in the case when they have the option to cross the freeway via a pedestrian bridge as well as the risk of being hit by a fast-moving vehicle when they have to cross the road by taking a gap in the traffic stream.

Sinclair and Zuidgeest (2016) determined that the majority of the trips that were surveyed as part of their study was home-based-work trips. The home-based-work trips then result in the majority of the trips made in the mornings and afternoons when the pedestrians are on their way to or from their workplaces. Pedestrian bridges and viaducts are implemented along the freeways in South Africa to enable pedestrians to cross the freeway safely. Pedestrians, however, do not always make use of these facilities, due to safety reasons as already mentioned above, the design of the facility and the locality of the facility (Sinclair and Zuidgeest, 2016). Pedestrians would often rather take a chance by crossing the freeway illegally than to cross using the pedestrian crossing facilities that have been implemented and stand the chance of being robbed on the pedestrian bridge. The illegal crossing behaviour increases the risk of being involved in a pedestrian crash on the freeways.

Extensive research has been done on the relationship that gender and age have on pedestrian behaviour. It was determined that males tend to adopt riskier behaviour than females (Zhang, Chen, and Wei, 2019, Classen, Awadzi and Mkanta, 2008; Kashani and Besharati, 2017). Statistics proof this, where 75-77% of the fatal crashes were males, resulting in females comprising of the minority of the fatal injuries, (RTMC, 2017).

People gain knowledge and experience by age. Children typically do not have the experience and the ability to distinguish between safe and unsafe situations. This trend can be seen in the number of children who sustain severe to fatal injuries in pedestrian crashes. The older people get, the slower they walk, and the longer reaction times they require to make a decision and act upon it. Elderly pedestrians therefore also face a higher risk of being involved in a pedestrian crash as well as to sustain more severe injuries than pedestrians who are in the middle-aged group (Classen, Awadzi and Mkanta, 2008, Kashani and Besharati, 2017). Research conducted by Zhang, Chen and Wei, (2019) and Jang *et al.*, (2010) confirmed that children and elderly pedestrians are at higher risk to be involved in pedestrian crashes and to sustain more severe injuries than pedestrians who are middle-aged.

Pedestrians can cross a road using three different methods (Zhang, Chen, and Wei, 2019). These crossing methods have different levels of safety. These methods are specifically based on urban environments where formal medians, of sufficient width for refugee, are provided.

- <u>Single-stage crossing</u>
  The single-stage crossing method refers to crossing the roadway at once. This method is the method where both sides of the road are crossed without waiting on the median for another gap to cross the second part of the road. A requirement for this crossing method

is that there are no vehicles that pass the pedestrian while the pedestrian is busy to cross the road. This crossing method can only be used if the traffic volume on the road is low enough that a long enough gap is available to cross the entire section of the road at once. There are, therefore, no vehicle-pedestrian conflicts for this crossing method.

- <u>Two-stage crossing</u>
  The two-stage crossing method refers to pedestrians that cross the road, and while they are crossing the road, a vehicle passes the pedestrian on the other direction of the road. This vehicle, therefore, did not pose any safety hazard to the pedestrian. The pedestrian, therefore, do not have to stop and wait for another gap, but there are potential vehicle-pedestrian conflicts for this crossing method.
- <u>Rolling-gap crossing</u>
  For this method, pedestrians have to cross the road one direction at a time. They have to wait for a gap to cross the one direction, wait on the median island for a gap in the other direction.

The rolling-gap crossing method is the method that has proved to have the most conflict between vehicles and pedestrians. In the study conducted by Zhang, Chen, and Wei, (2019), the majority of pedestrian-vehicle crashes were recorded for this crossing method. The rolling-gap crossing method is the crossing method that is used for the higher class roads, i.e., the roads that are typically wider and that have higher traffic volume than the lower class roads. The high traffic volume increases the risk of pedestrian crashes since there are more conflict points and fewer gaps that can be taken to cross the road.

Pedestrians who cross the road have to take a sufficient gap to allow them to cross the road without being hit by a vehicle. A portion of the pedestrian crashes occurs due to pedestrians that misjudge the gap they take (Miao, Yang, and Liang, 2016). Pedestrians also tend to take a gap which is big enough to cross the first few lanes of the road but tends to allow smaller gaps to cross the last few lanes of a road. Zhang, Chen, and Wei (2019) explained that this might be because the pedestrian only focuses on the lanes and the traffic nearest to him. Another reason might be that since the pedestrians are almost done crossing the road, they tend to take more chances by taking smaller gaps to cross the last few lanes.

Miao, Yang, and Liang, (2016) conducted a study to determine the portion of pedestrian crashes where the pedestrian was at fault versus the portion of pedestrian crashes where the driver of the vehicle was at fault. The following factors were listed as the main contributors to pedestrian crashes:

- Crossing the freeway without right of way;
- Crossing the road quickly in front of a big/heavy vehicle;
- Crossing the roadway in poor visibility conditions, with pedestrian wearing dark clothes, making visibility worse;
- Suddenly walking backwards when crossing the road;
- Unattended children on and along the road;
- Walking close to large vehicles, especially when they move fast; and
- Staying on the lane divided line at night.

Pedestrians that walk alone when they are drunk were also one of the factors that influenced the probability of being involved in a pedestrian crash (Miao, Yang, and Liang, 2016). According to Parmet, Lynm and Glass, (2002a) and Chen *et al.*, (2016) pedestrian safety is influenced by the pedestrian and drivers that are under the influence of alcohol, no or poor lighting and pedestrian activity in rural areas. The number of child pedestrian crashes during the night peak decreases since the majority of the children are not on the roads during the night. The risk, however, increases for adult pedestrians.

Alcohol has a negative influence on people's behaviour and reactions. People who are intoxicated with alcohol tend to act more aggressively, have blurry vision, slower reaction times and struggles to make clear decisions, (Kashani and Besharati, 2017, Classen, Awadzi and Mkanta, 2008, Jang *et al.*, 2010, Miao, Yang and Liang, 2016, Moore *et al.*, 2011). Pedestrians that are under the influence of alcohol might suddenly enter the roadway, without looking at whether it is safe to cross or not, (Jang *et al.*, 2010). Drivers who are under the influence of alcohol might not see pedestrians crossing the road or might have a too slow reaction time to avoid the pedestrian crash.

Pedestrians have the option to walk on the side of the road facing the oncoming traffic or to walk in the direction of the traffic flow. It was determined that the direction in which the pedestrian walks, has an influence on the injury severity and on the risk of being involved in a pedestrian crash. Research indicated that the risk of being involved in a pedestrian crash increases when pedestrians walk in the direction of the traffic flow. It was also determined that the number of pedestrians that walk against the traffic flow, i.e., facing the on-coming traffic, increases when the pedestrian is under the influence of alcohol, (Luoma and Peltola, 2013). Research has indicated that the risk for pedestrian crashes increases when the pedestrian is under the influence of alcohol. It was also determined that the risk of pedestrian crashes increases when pedestrians walk in the direction of the traffic flow, and it can be concluded that the pedestrian crashes where the pedestrian is under the influence of alcohol, does not increase due to the direction in which the pedestrian walks on the road.

Pedestrians cross the road in groups or individually. It is expected that a group of pedestrians who cross the road are safer than pedestrians that cross the road individually. The crossing behaviour in groups is expected to be safer because there are more eyes to check for dangerous situations and gaps in the traffic stream. It is, therefore, reasonable to expect that pedestrians that cross in groups will take safer gaps than a pedestrian that has to take a gap on his own, (Amoh-Gyimah *et al.*, 2017). A group of pedestrians is more visible than a pedestrian that is standing on the side of the road on his own, which improves the safety of the pedestrians in a group environment. It was, found, however, that fewer conflicts occur when pedestrians cross the road individually than when a group crosses the road at the same time. This results in a contradictory conclusion that the risk of being involved in a pedestrian crash when a pedestrian crosses the road one at a time decreases (Zhang, Chen, and Wei, 2019).

Cinnamon, Schuurman, and Hameed, (2011) determined that drivers of vehicles tend to pay more attention to their rear and side mirrors as well as to the vehicle instruments than to what is located on the side of the road. Drivers, therefore, do not notice pedestrians on the side of the road. This increases the risk of pedestrian crashes when the vehicle has to swerve into the shoulder to avoid a crash with another vehicle. The pedestrian on the side of the road might also think that a driver in an oncoming vehicle see them and will slow down when they cross the road. This might not be the case, and this also increases the risk of pedestrian crashes.

## 2.3 SPEED CONE

Every person has a certain angle at which they can see. This is called the peripheral angle. The width of this peripheral angle decreases when a person is travelling at a high speed. The average person, for example, has a peripheral angle of 100 degrees when they travel at a speed of 40 km/h. Table 1 below gives a summary of the peripheral angles at which an average person can see at different travel speeds (Forbes, 2012).

*Table 1: Peripheral angles at different speed limits*

| Speed | Peripheral Angle |
|---|---|
| **40km/h** | 100 degrees |
| **70km/h** | 65 degrees |
| **100km/h** | <40 degrees |

This phenomenon is often referred to as the speed cone. The term speed cone means that drivers can see fewer objects or pedestrians on the side of the road when they travel at high speeds.

This does not mean that drivers are unable to see objects on the side of the road, but the probability of overlooking a pedestrian or any other object increases as the travel speed of the vehicle increases (Forbes, 2012).

Drivers, therefore, do not always see pedestrians waiting on the side of the road to cross since the pedestrian are standing outside the peripheral angle. Drivers only see the pedestrian when they enter the driver's peripheral angle, which results in shorter reaction times to swerve out of the way to avoid a crash with the pedestrian.

It is not only the peripheral angle of the driver of a vehicle that has an impact on pedestrian crashes, it is also the peripheral angle of the pedestrian that crosses the road that has an impact on the pedestrian crashes. In studies done by Askin *et al.*, (1991) and David *et al.*, (1986) it was determined that the peripheral angle of children are significantly smaller than the peripheral angles of adults. This can cause an increase in child pedestrian crashes since child pedestrians will more easily overlook an on-coming vehicle due to their smaller peripheral angles.

## 2.4 FREE-FLOW AND CONGESTION

The crashes that occur on the freeway are influenced by many factors. One of these factors is the traffic flow at the time of the crash. As the traffic volume on the road increases, the spacing between the vehicles decreases. This results in a shorter stopping distance and an increase in potential conflict points between the vehicles on the road. The probability of a crash is, therefore, higher as the traffic volume increases (Yeo *et al.*, 2013). It is not only the potential conflict points between the vehicles that increase, but the potential conflict points between the vehicles and pedestrians will also increase as congestion on the road increases. The probability for a pedestrian crash during more congested times will, therefore, be higher than for free-flow conditions, (Zheng, Ahn, and Monsere, 2010).

This phenomenon was confirmed by Yeo *et al.*, (2013), who investigated the crash frequency on freeways using four different traffic states, namely free-flow state, bottleneck front state, congested state and back of queue state. It was confirmed that the crash frequency for the bottleneck, congested and back of queue states was significantly higher than the crash frequency on free-flow sections. This was expected since there are less potential conflict points for the free-flow state; however, the severity of the crashes in relation to the traffic state in which a road operate were not part of the study and no relationship between the severity of the injuries and traffic state was made.

Yeon, Hernandez, and Elefteriadou, (2009) studied the freeway capacity by day of the week, time of the day and segment type and concluded that the vehicle flow on the freeway is higher during the AM and PM peak hours than for the off-peak hours. It was determined that the flow of the traffic volume on the freeway per day of the week seem to follow the same pattern, i.e. the same vehicle flow patterns are observed every Monday, and every Tuesday appears to follow the same pattern and so on. The travel speed of the vehicles follows an inverse pattern to the capacity flow, i.e. as the flow increase and the freeway get more congested the speed drops dramatically, which is expected. Drivers who were caught in the congested sections of the road experienced different levels of frustration. Some drivers tend to drive aggressively due to their frustration. Other drivers tend to speed when they are past the congested areas in order to make up for the time lost while being stuck in congestion, which increases the risk of being involved in a pedestrian crash. This phenomenon was researched by Huang, Sun, and Zhang (2018). They determined that the frustration levels of the drivers were dependent on the type of congestion that drivers were driving in.

There are two different types of congestion. The one type is recurrent congestion, which is the type of congestion that occurs on a regular basis. Recurrent congestion is the congestion that is generally experienced during the AM and PM peaks, when the people are on their way to work or when they are travelling back after the workday. People, therefore, expect this congestion, and they plan their trip and travel time by taking the delays experienced due to the recurrent congestion into account. The frustration levels experienced in recurrent congestion was determined to be lower than for non-recurrent congestion (Huang, Sun and Zhang, 2018).

The second type of congestion is non-recurrent congestion. This is the type of congestion that resulted in the highest levels of frustration since the drivers did not expect to be stuck in congestion. Non-recurrent congestion is when a section of road that is usually not congested, is unexpectedly congested (Huang, Sun and Zhang, 2018).

The crash rate for different types of crashes is expected to differ during the day (Martin, 2002). The crash rate for single vehicle crashes was determined to be the highest during low traffic conditions. As the traffic volume on the road increases, the single vehicle crash rate decreased, while the multivehicle crash rate started to increase. The highest crash rates for multi-vehicle crashes were observed during high traffic conditions and decreased as the traffic volumes decreased. Qin, Ivan and Ravishanker, (2004) and Martin, (2002) determined that the crashes that occur in high traffic conditions normally occur on weekdays, while the majority of the crashes that occur in light traffic conditions occur during the weekend.

## 2.5 ENVIRONMENTAL FACTORS

According to the RTMC (2017), the majority of all the fatal crashes occur during the weekends. Amoh-Gyimah *et al.* (2017) found that the crashes that happen during weekends result in more severe injuries. Weekend activities usually are more social and leisure-based than during the week, resulting in higher use of alcohol during the weekend than during weekdays (Amoh-Gyimah *et al.*, 2017). Peden *et al.* (1996) found that 62.1% of the pedestrians that were involved in pedestrian crashes were under the influence of alcohol. Since it is known that alcohol use is higher during weekends, it is also expected that a large portion of pedestrians and drivers who are involved in the pedestrian crashes during the weekend and in night time are under the influence of alcohol. This assumption was confirmed in studies conducted by Miao, Yang and Liang (2016), Classen, Awadzi and Mkanta (2008), Kashani and Besharati (2017), Lin *et al.* (2019), and Peden *et al.* (1996). Bianco (2017) made contradicting findings of the day of the week on which the majority of the crashes occurs. Contradicting results have been obtained in research about the crash rates during the week and weekends. Research has indicated that the majority of the pedestrian crashes occur on weekends; however, Bianco (2017) conducted a study in which they concluded that the majority of the pedestrian crashes occur during normal weekdays.

Various studies conducted by Hine and Russell (1993), Amoh-Gyimah *et al.* (2017), Lee *et al.* (2016), Cinnamon, Schuurman and Hameed (2011), Verster and Fourie (2018), and Papadimitriou, Lassarre and Yannis (2016), proved that various factors have an influence on the crossing behaviour of a pedestrian, as well as on the risk of being involved in a pedestrian crash. These factors include the number of lanes to cross, the road type, travel speed, traffic volume, and the crossing method. Median divided roads resulted in lower crash risk as well as reduced injury severity. The pedestrian is exposed to the traffic stream for a shorter period of time on dual carriageway lanes, resulting in shorter time periods in which they are exposed to the traffic volume on the road. They also have to take a gap in a traffic stream that travels in one direction only, which simplifies the crossing method significantly. Multi-lane roads with raised medians have a lower risk of fatal injury (Li and Fernie, 2010; Campbell *et al.*, 2004). A contradictory result was obtained by Al-Ghamdi (2002), who determined that there is an increase in fatal injuries on two-lane roads that are provided with a median. Wider roads result in longer walking time, since pedestrians have to cross more lanes, and are therefore exposed to more conflict points. Higher crash rates and the risk of sustaining fatal injuries are therefore associated with long crossing distances. The width of the shoulder along a road has an impact on the number of pedestrian crashes that happen on the road. Research conducted in Ghana indicated that the risk to be involved in a pedestrian crash increases when the shoulder width decreases, (Amoh-Gyimah *et al.*, 2017).

Crashes that occur on straight roads are less likely to result in serious injuries. According to a study conducted in Ghana, the locations where the shoulders along the road are in poor condition or where there are no shoulders resulted in reduced risk of being in a crash as well as to sustain fatal injuries. There is a fear in Ghana of dangerous reptiles in the grass and overgrown shoulders, resulting in fewer pedestrians walking there. An interesting finding of this study is that gravel shoulders on the sides of the road resulted in less severe injuries. It was determined that the majority of the crashes occur at locations where the shoulders are in good condition. It was observed that the majority of the pedestrians prefer to use the roads with proper shoulders as their walking route, resulting in higher pedestrian volumes and therefore a higher risk of being involved in a pedestrian crash, (Al-Ghamdi, 2002).

The percentage of pedestrians that change their path and speed increase as the number of lanes on the road increases. An increase in the number of lanes results in an increase in the percentage of pedestrians that cross the road using the rolling-gap method. A road with more lanes is often a road on with higher traffic volumes and vehicle speeds, which result in fewer gaps that can be taken to cross the road. This results in a decrease in single-stage crossing opportunities (Xiaoyun *et al.*, 2011). Another factor that changes the crossing behaviour of pedestrians is when there is an increase in the travel speed on the road. Roads with high speeds give pedestrians the feeling that they have to cross the road faster. This result in pedestrians that tend to run over the road to safely cross, but now have the risk of falling and getting hit by a vehicle, (Xiaoyun *et al.*, 2011).

It was found that the pedestrian-vehicle conflicts increase with an increase in the number of lanes on the road. It was found that more conflict between the vehicles and pedestrians occurred at the lanes that the pedestrians had to cross last. A reason for this phenomenon might be that pedestrians tend to wait for a long gap when they start to cross the road, but when they approach the safe location (i.e., the other side of the road), they tend to take chances and cross the last lanes by taking smaller gaps. Another possible reason is that the pedestrians misjudge the length of the gap they need to cross the road. It should be noted that this study was done in an urban environment and that the speed for each lane is therefore similar, (Xiaoyun *et al.*, 2011).

The number of pedestrians that are present on the road as well as the traffic volume on the road influence the risk of being in a pedestrian crash. The number of potential conflict points increases as the pedestrian and traffic volume increases, resulting in a higher probability of a pedestrian crash (Amoh-Gyimah *et al.*, 2017; Verster and Fourie, 2018).

The probability of being involved in a pedestrian crash during night time is significantly higher than the probability of being involved in a pedestrian crash during day time (Bianco, 2017). It was also determined that the speed travelled at the time of a crash has an impact on the injuries sustained in crashes, (Huang, Sun and Zhang, 2018, Lin *et al.*, 2019, Papadimitriou, Lassarre and Yannis, 2016). The crashes that occur in congested traffic periods (i.e. during congested speed or breakdown speed) therefore tend to be less severe than the crashes that happen in free-flow conditions, which is associated with higher travel speeds. The risk of being in a crash, however, increases as the traffic flow increases. There are, therefore, a higher occurrence of severe crashes in the off and night peak periods than in the AM and PM peaks, which are related to higher traffic volumes. Drivers tend to speed in free-flow conditions, increasing the risk of being involved in a pedestrian crash. This will typically occur in the off peak periods, which also result in a higher risk of pedestrian crashes and also the injuries sustained in the crash. The visibility in the dusk and night time, i.e. in the night peak, decreases which results in a higher risk of crashes. Fatigue and stress also play a role in the concentration of a driver or pedestrian that crosses the road. Fatigue results in lower concentration levels, which in turn increase the reaction times. Fatigue, therefore, has a negative impact on the pedestrian crashes on the freeway. The main contributing factors in night pedestrian crashes were found to be due to speeding, fatigue, visibility, stress. Kononov *et al.* (2013), and Amoh-Gyimah *et al.* (2017) determined that the number of pedestrian crashes in the PM peak periods is higher than in the AM peak periods. This is due to the drivers and the pedestrians that are experiencing higher levels of fatigue when the workday is done than before work. They also experience higher stress levels and frustration in the PM peak periods than in the AM peak periods. This results in people paying less concentration on the road, which increase the risk of pedestrian crashes.

It was found that speed has the most significant impact on the injuries sustained in crashes during fine weather conditions, with the elderly pedestrians that sustained the most severe injuries. Research indicates that a greater portion of pedestrian crashes that occur in un-lit, dark areas results in more severe injuries than pedestrian crashes that happened in daylight conditions or in areas where street lighting is provided, (Li *et al.*, 2017, Bianco, 2017). Griffith (1994) and Chen *et al.* (2016) confirmed the conclusions made in research. The authors determined that the majority of crashes occur during night time. It was indicated that the risk of being involved in a crash on sections of road without street lighting is higher than for sections of road where the roadway was well lit. Bianco (2017) found that there is a higher risk of being involved in a fatal pedestrian crash during night time conditions. The risk of sustaining fatal injuries increases even more in adverse weather conditions or when the pedestrian or driver of the vehicle was intoxicated.

The pavement of a road is designed such that it provides sufficient friction between the tyres of the vehicle and the pavement so that the vehicle will not lose control when it has to brake unexpectedly. When the road is wet, the friction coefficient is smaller, which results in vehicles that lose control more easily than in dry conditions (Brodsky and Hakkert, 1988). The visibility on the road reduces in bad weather conditions, especially in night time conditions where the glare of the streetlights and the headlights of the vehicles reduce the visibility even further. This results in more time that is required for a vehicle to react and come to a complete stop. People tend to drive more slowly and to increase their following distance when it rains, which counteract the friction problem as well as the problems faced with visibility and reaction times (Jang *et al.*, 2010, Li *et al.*, 2017, Luoma and Peltola, 2013).

Research has indicated that the crash rates on roads vary between the seasons in a year. Ballesteros, Dischinger, and Langenberg (2004) determined that the majority of the crashes occur during the winter and autumn months. This is due to the low light conditions, restricting the visibility of obstructions, other vehicles, and pedestrians in the road (Ballesteros, Dischinger and Langenberg, 2004). Low light conditions are known to have a negative impact on the number and severity of crashes. It is not only the time of year that has an impact on the number of crashes that occur on a section of road. The time of day, as well as the weather conditions, plays a significant role in the number and severity of the pedestrian crashes (Kim *et al.*, 2010). According to Kim *et al.* (2010), there is a decrease in the risk of sustaining fatal injuries when being in a pedestrian crash when it is raining, or when there is fog on the road. It was, however, also determined that there is a risk of sustaining serious injuries increase under these conditions. Mayr *et al.* (2018) confirm this trend. Mayr *et al.* (2018) investigated the severity of child pedestrian injuries, and he concluded that the majority of the child pedestrian crashes occurred during sunny and dry weather conditions. A reason for this trend might be due to the fact that children are prone to play and walk outside during the day and in fine weather conditions (Mayr *et al.*, 2018). They tend to stay indoors when it is dark, and when the weather conditions are bad, resulting in less child pedestrian crashes on rainy days. Adults, however, who are forced to go outside to go to their jobs still face the probability to be involved in a pedestrian crash. Mayr *et al.* (2018) determined that the lack of visibility of the vehicle (for the pedestrian) and lack of visibility of the pedestrian (for the driver) was the leading cause of the crash.

The pedestrian volumes on the road decrease during adverse weather conditions. Pedestrians, however, tend to run over a road more quickly in adverse weather conditions than in pleasant weather conditions. A reason why pedestrians run over the road in bad weather conditions is to reduce the time that they have to spend outside on the road. This results in a higher risk of falling in the road and also the risk of being in a crash. Pedestrians also tend to take more risks and taking shorter gaps to cross the road, which again increases the risk of being in a crash. Risky crossing behaviour, combined with the visibility constraint, increase the risk of being in a pedestrian crash during bad weather conditions (Bianco, 2017, Yeo *et al.*, 2013, Amoh-Gyimah *et al.*, 2017, Kashani and Besharati, 2017). Research has been carried out to determine whether the significance of the injuries sustained in a pedestrian crash is influenced by weather conditions or not. The studies indicated some contradicting results, with some studies suggesting that bad weather conditions reduce the severity of the crash due to lower speeds and increased following distances, while other studies suggested that the severity of a pedestrian crash in bad weather conditions increases (Brodsky and Hakkert, 1988, Yeo *et al.*, 2013, Amoh-Gyimah *et al.*, 2017, Kashani and Besharati, 2017).

Pedestrians and drivers, who are using their cell phones while walking or driving, cause safety hazards for both themselves as well as for the other road users. The person who is using his cell phone does not pay full attention to what is happening on the road and might, therefore, react too late and cause a crash. Drivers tend to serve out of their lanes, and pedestrians tend to divert off their walking path when they do not pay full attention to the road. This phenomenon was investigated by other researchers and confirmed that the use of cell phones increases the risk of being involved in a pedestrian crash (Thompson *et al.*, 2013, Jang *et al.*, 2010, Bianco, 2017). Cell phones are only one of a few factors that can distract the drivers and pedestrians travelling on the road. Other factors such as eating, drinking, having a conversation with other people as well as listening to music also have an impact in pedestrian crashes, (Poó, Ledesma and Trujillo, 2018, Thompson *et al.*, 2013).

## 2.6 TRAVEL SPEED

Different speed limits are applicable to different road classes. Higher class roads are typically associated with higher speed limits. Research has indicated that speed has a major impact on the number of crashes on the road as well as on the injuries sustained in the crash, where more severe injuries are sustained at higher impact speeds (Kim *et al.*, 2010, Moore *et al.*, 2011, Constant and Lagarde, 2010). A study conducted by Ossiander and Cummings (2002) investigated what the impact of changing the speed limit on a road would have on the number of fatal injuries in road traffic crashes. A significant increase in fatal crashes was observed when the speed limit of a rural highway was increased with 15%. The crash rate of the road did not change significantly, indicating that the number of crashes on the road did not change significantly. This leads to the conclusion that it is either only the injuries sustained in crashes that changed significantly or that the number of crashes increased proportionally to the number of re-distributed traffic from other roads to the road with a higher speed limit (Ossiander and Cummings, 2002). This is an interesting result since research indicates that when the speed on a road increase, the crash rate also increases. This phenomenon is predicted by Nilssons' predictive model (Cameron and Elvik, 2010).

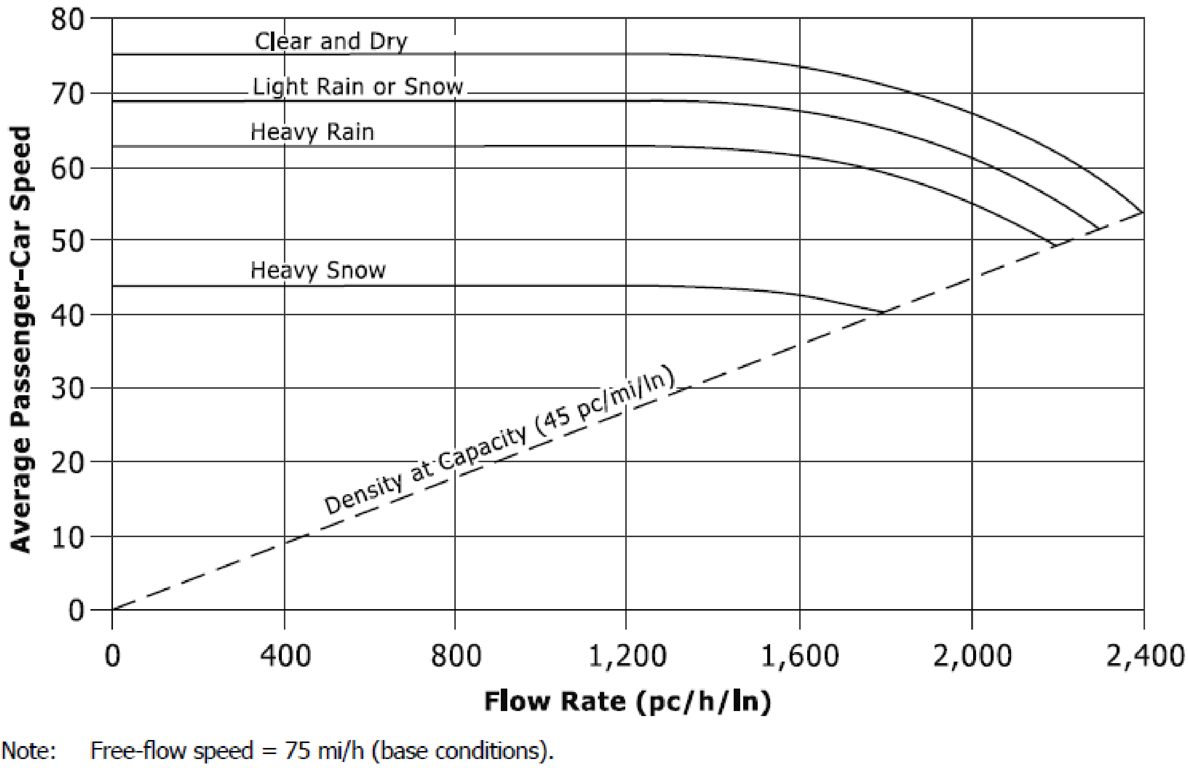



*Figure 3: Vehicle flow versus speed travelled (Source: HCM2010)*

It is evident from Figure 3 that the speed travel on a freeway stays constant for traffic volumes between 0 and 1400 passenger car units per hour per lane in clear weather conditions. It is only when the traffic volumes on the freeway approach congestion levels that the speed travelled on the road decrease dramatically. In wet or extreme weather conditions, the point where the speed starts to decrease rapidly is reached earlier than in fine weather conditions. A reason for this is due to the visibility that decreases in unfavourable weather conditions.

Extensive research has been done on the injuries sustained by pedestrians at different impact speeds. It has been determined that the risk of a fatality at speeds below 55km/h is low (Kröyer, Jonsson and Várhelyi, 2014, Kröyer, 2015, Martin and Wu, 2018, Tefft, 2013, Rosén and Sander, 2009). The travel speeds on freeways fluctuate between free-flow speed and congestion or breakdown speed during the different peak hours of the day (Yeon, Hernandez and Elefteriadou, 2009).

The faster a vehicle drive, the longer it requires to stop, resulting in longer reaction times required to avoid a pedestrian crash. The kinetic energy at which a vehicle travels increases quadratically as the speed increases. This results in a high probability of fatal injuries at high speeds. It is, however, also determined that the number of conflict points increases as the traffic volume increases and the probability of a crash is, therefore, higher (Tanishita and van Wee, 2017).

For a long time, it was believed that there is a u-shaped relationship between speed and crash rates. Recent research, however, indicated that this is not the case (Forbes, 2012). Other research, however, still indicate that the u-shaped relationship between the crash rates and impact speed can be found (Kononov *et al.*, 2013). In a study conducted by Kononov *et al.*, (2013) it was determined that the number of crashes on a road segment increase slowly as the traffic volume on the road increases until it reaches a certain point from which it increases more rapidly to an equilibrium point.

Golob, Recker, and Alvarez, (2004) have found that there is a u-shaped function between the crashes and the traffic flow on the road, with a high number of single vehicle crashes when the traffic volume is low, and a high number of multiple vehicle crashes when the traffic volume is high. It was observed that the highest number of crashes do not occur during the AM and PM peak periods but occur in the off and night peak periods. Golob, Recker, and Alvarez (2004) determined that it is not only the mean speed that has an influence on the safety of a road but rather a combination of speed and traffic volume. The variation between speed and the traffic volume on the road was found to have a negative impact on the safety of the road. It was determined that the crash rate is dependent on the mean speed as well as the variation in speed. It was, however, determined that there are three attributes that contribute to the safety of a road, namely the speed, density, and the traffic volume.

The mean speed travelled on a road increases when the traffic volume decreases and on the other hand decreases when the traffic volume decreases. The speed variation was determined to have a negative influence on the number of crashes that occur on the road, i.e., as the speed variation increases, the number of crashes also increases. It was also determined that the mean speed, as well as the traffic flow, has an impact on the number of crashes. It was determined that it is not only the mean speed but also the changes in the mean speed on a road segment that has an influence on the number of crashes (Parmet, Lynm and Glass, 2002a).

Speed is seen as improving mobility, but also has road safety implications, for all road users (Forbes, 2012). From the study conducted by Parmet, Lynm, and Glass (2002a), it was determined that the injuries sustained at higher speeds were more severe than at lower impact speeds. They found that vehicles that hit a child pedestrian while braking caused less severe injuries than vehicles that did not attempt to brake before impact. Parmet, Lynm and Glass (2002a) determined that the relationship between pedestrian injuries sustained and speed have an exponential relationship, where the injuries sustained increases as the impact speed increase.

Bianco (2017) determined that the speed travelled on the road has a negative impact on the risk of being involved in a pedestrian crash. It is, however, not only the speed travelled on the road that has an influence on the occurrence of pedestrian crashes, but also the gender of the pedestrian involved in the crash, the land uses surrounding the crash site and the income level of the pedestrian involved in the crash (Bianco, 2017).

According to Kononov *et al.* (2013), the time at which the majority of the pedestrian crashes occur was determined to be outside the peak periods, i.e., during the off peak and night peak periods. The night time peak period is the time where the probability of drivers falling asleep, drivers being under the influence of alcohol or experience fatigue is the greatest, resulting in a higher probability of crashes. Another reason for this is due to the lower traffic volumes, which result in vehicles that can travel at free-flow speed and not at congested speed as in the AM and PM peak periods. Clifton, Burnier and Akar (2009) determined that the number of lanes that has to be crossed together with the speed travelled on the road has a negative impact on the probability to sustain fatal injuries in pedestrian crashes. The higher the speed travelled and the more lanes to be crossed, the higher the probability of sustaining fatal injuries in a pedestrian crash.

When improvement measures are installed on the road or at the intersection, the way in which the reduction or increase in crashes is measured is called the crash modification factor. This factor is calculated by raising the division of the end speed with the start speed to a factor of a statistical constant. It is, therefore, evident that the factor will increase if the speed after implementation increases. This relationship makes it clear that the crashes will probably increase as the travel speed increases. The equation is given below.

$$CMF = (\frac{V_a}{V_b})^x \qquad \text{(Eq. 1)}$$

The differential between the speed that a pedestrian and vehicle travel at is big, especially on the freeways. The distance that a vehicle needs to come to a complete stop increases as the travel speed increases. The probability of losing control when a vehicle swerves out of the way for an object increases as the travel speed increases (Forbes, 2012). When the equations of motion are studied, it is clear that the distance and time required to come to a complete stop increases as the travel speed increases (van As et al., 2002). Drivers do not always have enough time and roadway to react and avoid a pedestrian crash on high-speed roads.

It is engineering practice to set the speed limit of a road to the 85th percentile speed travelled on that road. Another important viewpoint in traffic engineering is that the speed variance has a significant impact on the crashes and injuries sustained in these crashes. It would, therefore, be argued that it would be beneficial to set the speed limit such that the variance on the road is as small as possible. In cases where the 85th percentile speed is higher than the speed limit, it would be recommended that the speed limit should be increased to reduce the variance between the vehicles travelling at the speed limit and the vehicles travelling at the 85th percentile speed. It was, however, determined that if the speed limit is changed, the variance does not change. The reason for this is because the traffic on the road all either increase or decrease their travel speed to adjust to the speed limit, leaving the difference between the speed limit and the 85th percentile speed the same (Ossiander and Cummings, 2002).

## 2.7 VEHICLE TYPE AND SECONDARY IMPACT FORCE

There are discrepancies in the research about the impact that the vehicle type and weight have on the severity of the injuries sustained in a pedestrian crash. Martin and Wu (2018) and Tefft (2013) indicated that there is a relationship between the shape and weight of the vehicle and the injuries sustained. Pedestrian crashes involving light delivery vehicles (LDV), sport utility vehicles (SUV) as well as heavy vehicles (HV) are related to more severe injuries than a pedestrian crash with a passenger vehicle. Kashani and Besharati (2017) determined that the majority of vehicles that are involved in pedestrian crashes are passenger vehicles. The other vehicle types, e.g. heavy vehicles and buses, do not stand as high a risk of being involved in a pedestrian crash. An explanation of this phenomenon might be due to passenger vehicles that make up the highest portion of the modal split of a road. This increases the risk for passenger vehicles of being involved in a pedestrian crash. Clifton, Burnier, and Akar (2009) conducted a study to determine whether the vehicle type has an impact on the injuries sustained in a pedestrian crash. It was determined that, even though passenger vehicles pose the highest risk of being involved in a pedestrian crash, the heavier vehicles contribute to more severe injuries.

When a pedestrian gets hit by a vehicle, the impact point between the vehicle and the pedestrian is determined by the height of the pedestrian as well as the shape of the vehicle. Schmucker *et* *al.* (2010) determined that the shape of a vehicle has a significant impact on the severity of the injuries sustained. It was also concluded that high energy pedestrian crashes resulted in severe head injuries. This is due to the kinetic energy of an object, where the kinetic energy is equal to half the mass of the object times the speed squared. When this is taken into account, it is clear why the heavier vehicles might result in more severe injuries, since these vehicles have higher kinetic energies that lighter vehicles.

Kröyer, Jonsson and Várhelyi (2014) reasoned that since the difference between the weight of the pedestrian and the vehicle's weight is already so big, it is not the weight that has an influence on the severity of the injury, but it is rather the front structure of the vehicle that has an impact on the severity of the injuries.

Ballesteros, Dischinger, and Langenberg (2004) conducted a study to determine the influence that different types of vehicles have on pedestrian injuries. Ballesteros, Dischinger and Langenberg (2004) investigated the influence that the type of vehicle, the speed at which the vehicle drove as well as the weight of the vehicle has on pedestrian injuries. It was determined that the pedestrians that were involved in a crash with a SUV or a pick-up van sustained more severe injuries and was at a higher risk to sustain fatal injuries than pedestrians that were hit by smaller, or passenger vehicles. The vehicle weight had a direct relationship on the injuries, i.e. the heaviest vehicles caused the most severe injuries, and pedestrians that were hit by lighter vehicles sustained less severe injuries compared to the heavier vehicles. Interestingly enough, Ballesteros, Dischinger, and Langenberg (2004) found that the type of vehicle did not have an influence on the injuries sustained for the scenario where the speed and the weight of the vehicles were controlled. They concluded that the vehicle speed, shape, and size of the vehicle had an influence on the injuries sustained at lower speed limits; however, his study was based on incidents at a speed limit of 48km/h (30mph) and lower. The impact points where larger vehicles, e.g., SUVs, hit pedestrians, are situated higher than for smaller vehicles. This means that the injuries sustained by pedestrians that are hit by SUV’s are located higher up the body than for smaller vehicles. Pedestrians that are hit by SUV’s are twice as likely to sustain brain injuries as pedestrians hit by smaller vehicles (Crocetta *et al.*, 2015). A pedestrian that is involved in a crash with a SUV typically hit the vehicle with their heads on the bonnet, while head injuries with smaller vehicles are associated with windscreen injuries. It is therefore concluded that more severe injuries are sustained at low-speed limits (40km/h or lower) for SUV's and large vehicles (Ballesteros, Dischinger and Langenberg, 2004, Crocetta *et al.*, 2015).

The majority of the research on the influence that the vehicle type has on injuries sustained only focussed on adult pedestrians. DiMaggio, Durkin, and Richardson (2006), however, determined that the same relationship can be seen in child pedestrian crashes. It was determined that children who are hit by heavy vehicles sustain more severe injuries than the children that are hit by a passenger vehicle. They also confirmed that the shape of the vehicle plays a significant role in the injuries sustained by child pedestrians.

Adults that are hit by a vehicle normally land on the bonnet of the vehicle or are transferred over the vehicle. Children, however, are typically thrown forward and will land in front of the vehicle instead of on the vehicle. The reason for this is due to the size of the children compared to the vehicle (Parmet, Lynm and Glass, 2002a).

Crocetta *et al.*, (2015) related the injuries sustained by pedestrians that were involved in pedestrian crashes by looking at the impact at which they hit the ground in relation to the vehicle speed and size. Crocetta *et al.*, (2015) determined that the impact at which the pedestrian hit the ground after they were hit by a light vehicle, was the main reason for severe head injuries. They concluded that pedestrians tend to sustain more severe head injuries due to ground impact at lower speeds when hit by an SUV than for smaller vehicles.

## 2.8 LAND USE SURROUNDING THE AREA

Kim *et al.* (2010) determined that the probability of pedestrian crashes increase in areas where there are commercial and retail land uses. The injury type of these crashes increases in a u-form. This results in a probability of an increase in the number of fatal pedestrian crashes, but that there might also be an increase in the number of pedestrian crashes where no injuries are sustained. In short, according to Kim *et al.* (2010), the number of pedestrian crashes increases at the locations where commercial land use are located.

The traffic volume increases as vacant land are developed. It is known that every type of development generates traffic. Different trip generation factors are linked to different types of land-uses, with the majority of the additional trips being generated in the AM and PM peak hours on weekdays (South Africa Committee of Transport Officials, 2013). With the urban sprawl that occurs around the cities, the traffic on the freeway also increases. Developments, including retail and industrial areas, arise along the freeways, generating additional traffic. This traffic does not necessarily travel on the freeways; however, people that live in the informal settlements along the freeway can easily access the developments from the side of the freeway. Since these people often do not have access to public or private transport modes, they can only rely on walking as their mode of transport. People tend to take the shortest route from an origin to a destination (Sinclair and Zuidgeest, 2016). This, therefore, means that people who want to access the developments along the freeway cross the freeway in order to access the retail or industrial areas.

Research has indicated that certain land uses have a negative impact on pedestrian crashes (i.e., an increasing number of pedestrian crashes). Land uses, such as residential densification, retail, and industrial, are known to increase the number of pedestrian crashes. It has been determined by Priyantha Wedagama, Bird and Metcalfe (2006) that retail land use has in influence in pedestrian crashes during working hours, although, this study was based on urban areas, and can, therefore, be different than for areas outside a city. Lin *et al.* (2019), and Osama and Sayed (2017) also confirmed that the land uses listed above had an impact on pedestrian crashes.

In a study conducted by Miranda-Moreno, Morency, and El-Geneidy (2011), it was also determined that the land use on the side of a road has an impact on the pedestrian crash rate. Bianco (2017) confirmed this phenomenon. The number of crashes increased in the areas where the majority of the developments are industrial and commercial developments. A reason for this might be because these types of land uses provide a high number of job opportunities and will, therefore, attract a high number of people. Clifton, Burnier, and Akar (2009), however, conducted a study to determine the relationship that the gender of the pedestrian and the land use in the surrounding area have on pedestrian crashes. They concluded that there is a relationship between the gender of a pedestrian and the risk of being in a pedestrian crash, but that there is not a relationship between the land use in the surrounding area and the risk of being involved in a pedestrian crash. This phenomenon might be explained by looking at the types of trips that the specific genders make. There is a possibility that the two attributes might be collinear, and that the land use attribute will, therefore, be included in the gender attribute. In a study conducted by Graham, Glaister, and Anderson (2005), another observation was made. Graham, Glaister, and Anderson (2005) determined that the land use in the surrounding area have a negative impact on the risk of sustaining fatal injuries in a pedestrian crash. The child pedestrian crashes were more likely to occur in the residential areas, whereas the adult pedestrian crashes decreased in these areas and increased at the places where the majority of the job opportunities were found.

## 2.9 INCOME LEVEL

Pedestrians from low-income households are more vulnerable to be involved in a pedestrian crash since they typically have to walk long distances since and because pedestrian facilities are provided in low-income areas. In South Africa, low-income households are typically located outside the city area, adjacent to the freeway, (Sinclair and Zuidgeest, 2016). Osama and Sayed (2017) conducted a study to evaluate what the impact of different factors, namely socio-economics; land use, built environment, and road facility have on pedestrian crashes and injuries. They determined that pedestrians from lower-income households are at higher risk to be involved in a pedestrian crash. Political and economic factors are a reason for the many settlements along the South African freeways (Sinclair and Zuidgeest, 2016). These areas are not provided with appropriate transport infrastructure, forcing the people to walk to the nearest public transport facility, which usually is along the freeway or at the ramps of interchanges. Public transport facilities are not necessarily provided at the interchanges along the freeway, which force pedestrians to walk to their destinations. A portion of the pedestrians in South Africa are therefore forced to cross the freeway in order to get to their destinations. The pedestrians typically have to leave their homes early in the morning and only come back late in the afternoons. This means that they have to undertake a part of their trip, including walking, in the dark, which also increases the risk of being involved in a pedestrian crash (Zegeer and Bushell, 2012).

The children that grow up in low income areas do not have playgrounds or a fenced garden in which they can play. These children are therefore forced to play outside on the roads, which increase the risk of being hit by a vehicle. Children are not educated in road safety and will not be on the lookout for vehicles when they are playing in the road reserve. This increases the risk for children to be involved in a pedestrian crash (Parmet, Lynm and Glass, 2002a).

Noland, Klein, and Tulach (2013) conducted a study to determine whether income level and land use in the area has an impact on pedestrian fatalities. A relationship between the land use in the area and pedestrian fatalities were made; however, he could not determine whether income level has a significant impact on the pedestrian fatalities or not. Bianco (2017), on the other hand, found a relationship between the income level and pedestrian fatalities. The findings that were made by Bianco were in line with the research and findings made by authors such as Parmet, Lynm, and Glass (2002b), Graham, Glaister, and Anderson (2005), and Lin *et al.*, (2019). Pedestrians living in low income areas often do not have access to vehicles and are therefore forced to use walking as their primary mode of transport. Low income residential areas are often located close to main roads and arterials. These roads are associated with high traffic volumes. This, therefore, results in high pedestrian volumes at locations associated with high traffic volumes and therefore, also an increase in the risk of pedestrian crashes (Lin *et al.*, 2019).

## 2.10 CONCLUSION

When the research above is taken into account, it can be concluded that there is a close relationship between the travel speed and injuries sustained by pedestrians in pedestrian crashes. Factors that should also be considered is the time of day, the impact force at which the pedestrian hit the ground or any other secondary object as well as the force at which the pedestrian is struck. This force is dependent on vehicle mass and acceleration.

It is challenging to model pedestrian behaviour, and there is limited research done on the relationship that different travel speeds on the freeway have on pedestrian crashes and the injuries sustained in these crashes. Many factors, such as speed, land use, and income levels, influence pedestrian crashes, as already mentioned above. Many of these factors only have been researched in urban areas and on lower speed roads. It is, however, important to understand what the impact on the combination of all these factors has on pedestrian safety along freeways.

Almost no data was found on whether the travel speed of the vehicles plays a role in the decision to cross the freeway at-grade of via the pedestrian facilities. Limited research was found that relate the pedestrian crash rate with the capacity flow and the vehicle speed along freeways.

From the literature, it is clear that various factors play a role in the risk of being in a pedestrian crash as well as the risk of sustaining fatal injuries in the pedestrian crash. There is, however, limited research done to determine the impact that all these variables together have on the pedestrian crash rate and the severity of the injuries sustained in these crashes.

Freeways are not designed to accommodate pedestrians. The high travel speeds, number of lanes to cross as well as high traffic volumes all contributes to unsafe pedestrian environments (Clifton, Burnier and Akar, 2009). It is not abnormal to observe pedestrians crossing or walking along the freeways in South Africa. The informal settlements are located along the freeways, which are typically median divided roads with more than three lanes per direction. These roads are associated with high traffic volumes and speed limits. When this is taking into account, it is clear that there is a need to investigate the what the combination of the latter factors have on the crash rate and injuries sustained in the pedestrian crashes that happen on the South African freeways.

# 3. METHODOLOGY

## 3.1 INTRODUCTION

The research aimed to determine what influence a list of external factors in the traffic environment had on pedestrian injuries and the pedestrian crash rates on the Gauteng FMS Network. The relationship between these factors and the injuries sustained in a pedestrian crash were determined using a multinomial logistic (MNL) model, as further discussed in Section 3.5. The investigated topics are listed below:

1. What impact does the time at which a pedestrian crash occurs have on the injury severity and the pedestrian crash rate?
2. What is the relationship between the average traffic volume and hypothetical travel speed on the injuries sustained in pedestrian crashes?
3. Do the average traffic volume and the hypothetical travel speed have an influence on the crash rate in the study area?
4. Does the land use surrounding the crash locations play a role in the number of pedestrian crashes and the injury severity of these crashes?
5. What impact does the vehicle type involved in the crash have on the risk of being in a pedestrian crash as well as on the pedestrian's injuries?
6. Is there a trend between the number of crashes and the injuries sustained in these crashes and the day on which these crashes happen?
7. What influence does the number of lanes on the road have on the pedestrian crash rate and injuries sustained in the crash?

## 3.2 STUDY AREA

The ring road consisting of the N1 Western bypass, N12 Southern bypass and N3 Eastern bypass was chosen as the study area for this study. The study area was determined using pedestrian crash data of the past six years (2013-2018). A total of 38% of all the pedestrian crashes occurred on the Gauteng FMS Network in the analysis period happened on these three sections of road. The study area is indicated in Figure 4.

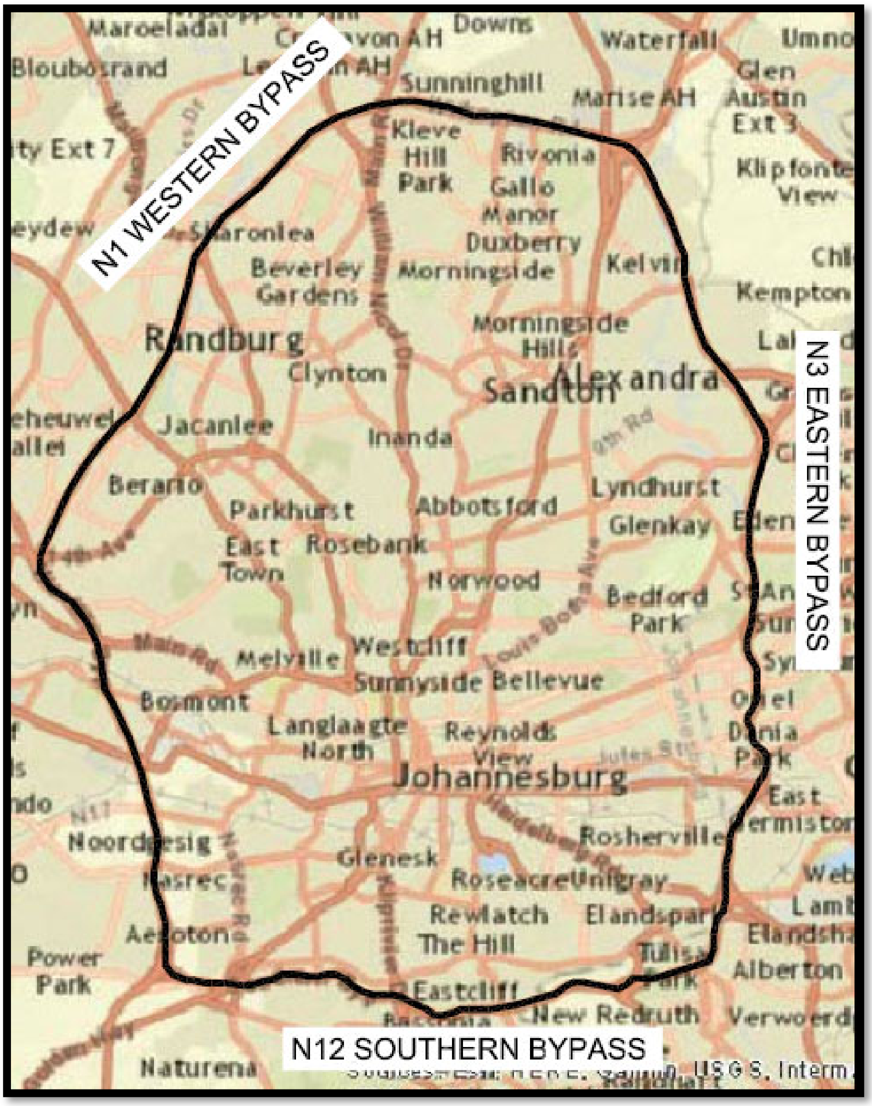



*Figure 4: Study area*

A total of 177 pedestrian crashes occurred in the five year analysis period in the study area. For the majority of these crashes, only one pedestrian was involved; however, there were two or three pedestrians involved in some of these crashes. In the cases where more than one pedestrian was involved in the crash, the data line was duplicated and changed so that every data line relates to only one injury type. If there were a crash where one pedestrian sustained serious injuries and another pedestrian who sustained slight injuries, the data line was copied such that there is a line that corresponds to only the serious injury as well as a line with only the slight injury as reference injury. After separating all these injuries, a total number of 188 data lines were recorded, which correspond to 188 pedestrians that were involved in the 177 pedestrian crashes.

**Geometry and Cross-section**

The Gauteng FMS Network was upgraded in 2010 as part of the Gauteng Infrastructure Framework Plan (GFIP). The entire section of the Gauteng FMS Network that is included in the study area was upgraded to a dual carriageway, with between three and five lanes per direction. Jersey barriers are installed along the majority of the study area, with w-beam barriers that are installed on isolated sections on the N1 Western bypass and N12 Southern bypass. These sections of the freeway are provided with a median with the w-beam barriers that are installed on both sides of the median. There are two sections on the N12 Southern bypass which is only provided with a median island with vegetation to serve as a barrier. The entire study area is also provided with street lighting. The street lights are implemented on the barriers or on the medians that are provided on the road network. The locations of the different barriers and median locations are indicated in Figure 5.

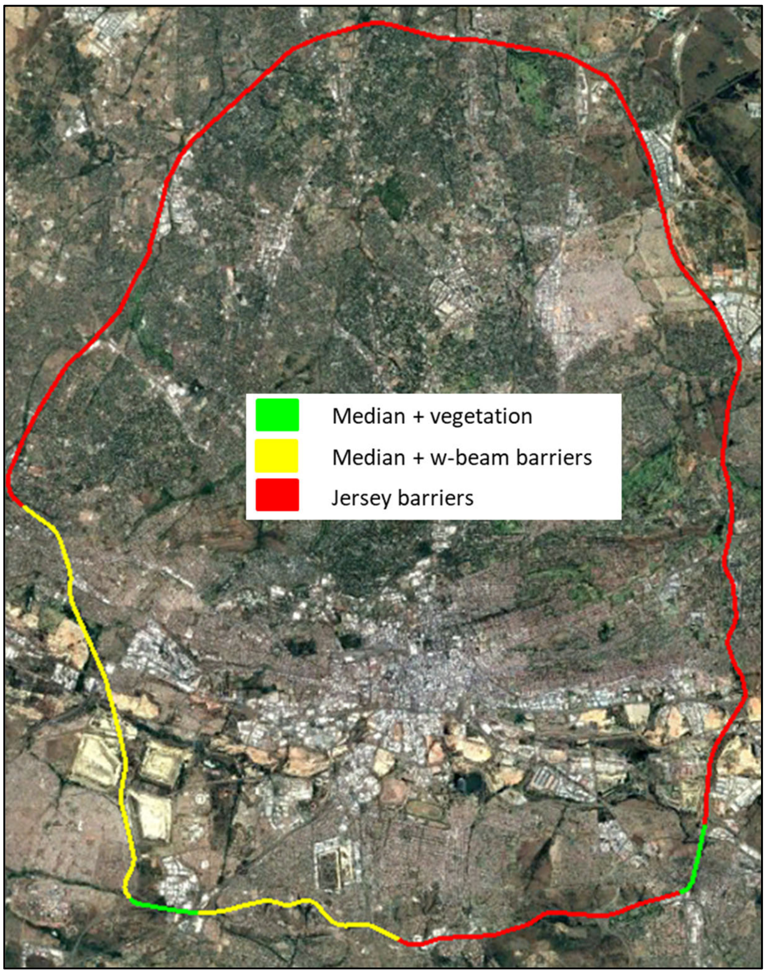



*Figure 5: Location of different barriers in the study area*

The study area was divided into 26 sub-sections to simplify the analyses and the developing of the MNL model. These sub-sections are typically the section of road between two consecutive interchanges. This was done to simplify the analyses of the data by ensuring that every subsection has the same traffic volume and travel speed. The cross-section of each sub-section is summarised in Table 2.

*Table 2: Summary of the sub-area cross-sections*

| Subsection | Road name | Number of lanes | | | | |
|---|---|---|---|---|---|---|
| | | 2 | 3 | 4 | 5 | 6 |
| 1 | N1 Western Bypass | | | | x | |
| 2 | | | | x | x | |
| 3 | | | | x | | |
| 4 | | | | x | | |
| 5 | | | | x | | |
| 6 | | | | x | | |
| 7 | | | | x | | |
| 8 | | | | x | | |
| 9 | | | | x | | |
| 10 | | | | x | | |
| 11 | N12 Southern Bypass | | | x | x | |
| 12 | | | | x | | |
| 13 | | | | x | | |
| 14 | | | | x | | |
| 15 | | | | x | | |
| 16 | | | | x | | |
| 17 | | | | x | x | |
| 18 | N3 Eastern Bypass | | | | x | |
| 19 | | | | x | x | x |
| 20 | | | | x | x | |
| 21 | | | x | | x | |
| 22 | | | x | | x | |
| 23 | | | | x | | |
| 24 | | | | x | | |
| 25 | | | | x | | |
| 26 | | | | x | | |

**Non-motorised and Public Transport Facilities**

Non-motorised transport (NMT) and public transport (PT) facilities have been provided along the FMS Network. A summary of the non-motorised transport and public transport facilities are provided in Figure 6.

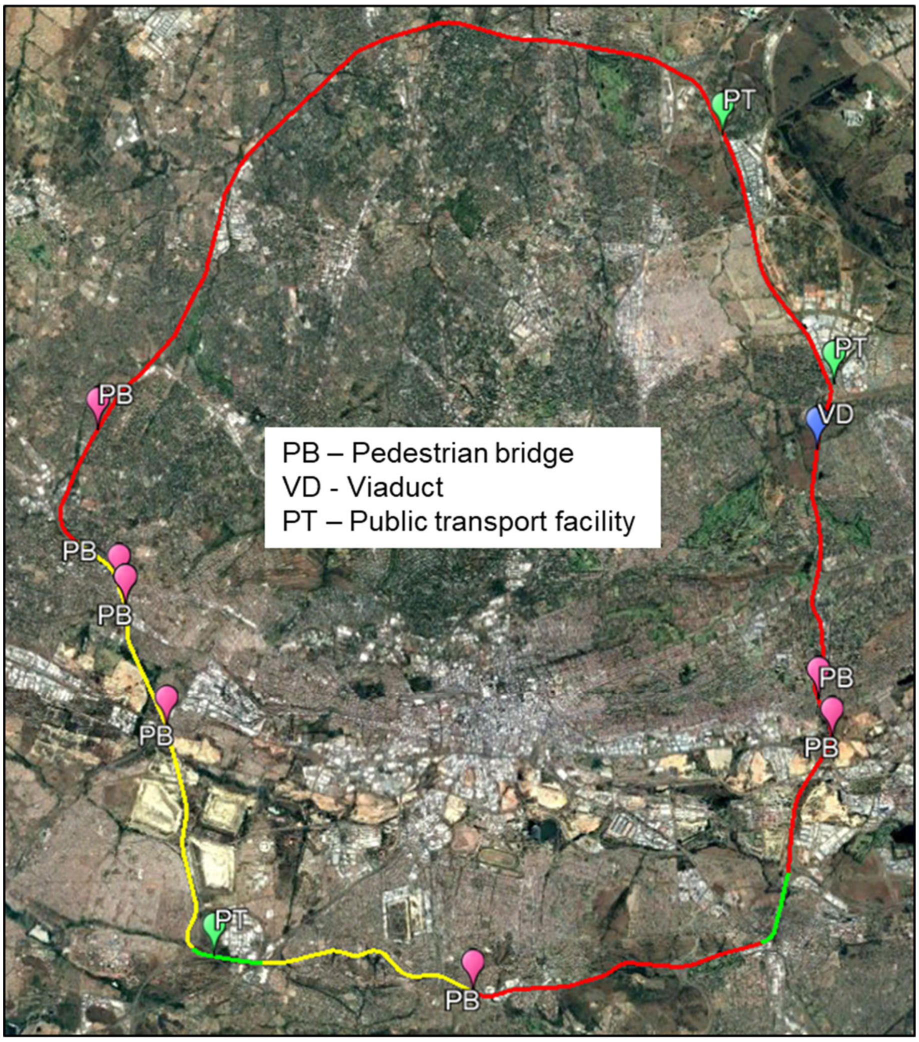



*Figure 6: Public Transport facilities and medians provided in the study area*

There are numerous PRASA railway stations in close vicinity of the study area, as indicated in Figure 7. It is known that public transport corridors attract pedestrians. It is also known that pedestrians tend to take the shortest route to their destinations (Sinclair and Zuidgeest, 2016). Pedestrians, therefore, cross the freeway to get to the public transport facilities for the cases where the shortest route is across the freeway. The public transport layby located at the N12 interchange also resulted in pedestrians that crossed the interchange ramps and the freeway to get to the layby.

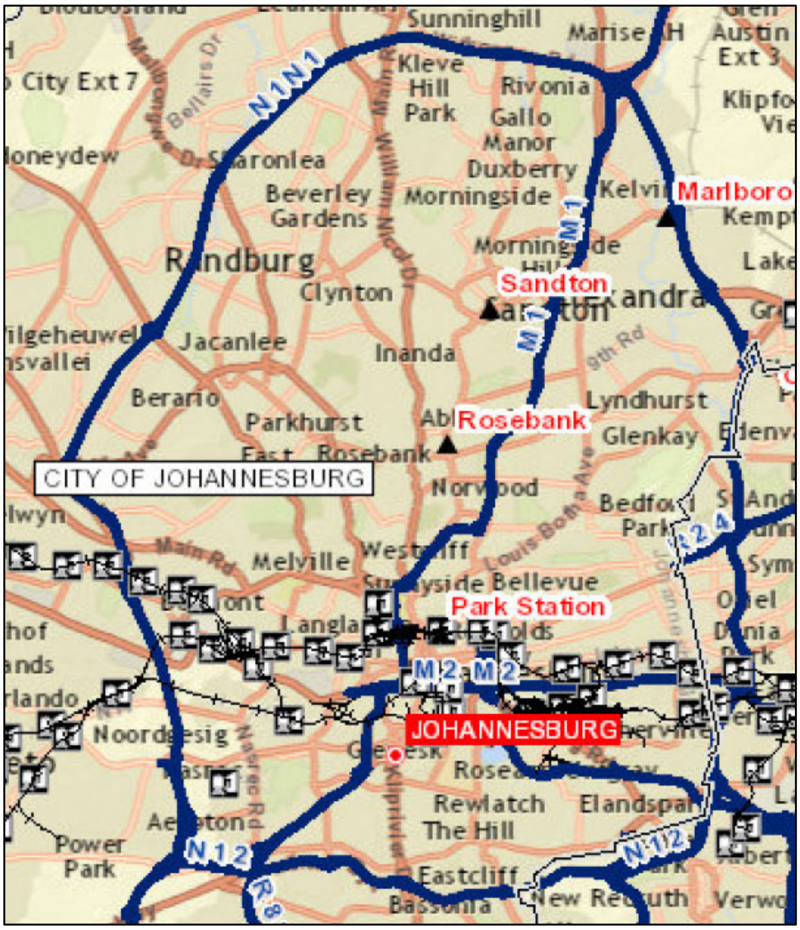



*Figure 7: Prasa railway stations and railway lines*

The pedestrian bridges are located at strategic positions; however, the five bridges do not cater for the full pedestrian crossing demand across the entire study area. Many pedestrians will not make use of the pedestrian bridge if it is not close to the location where they have to cross the freeway and as determined in various studies, pedestrians may also tend to avoid pedestrian bridges due to security reasons (Sinclair and Zuidgeest, 2016, Özkan, Porter and Lajunen, 2016).

Informal pedestrian footpaths are present along the freeway included in the study area. Some of the footpaths are located on one side of the road, which leads to the assumption that the pedestrians are either picked up at this location on the freeway, dropped off from where they walk further to their destinations or that the pedestrian walk on the shoulder of the freeway for a while before they divert away from the freeway to the built-up areas. It is also possible that the pedestrians cross the first part of the freeway (one direction), walk along the barrier for a certain time and distance before they cross the second half of the freeway (second direction).

## 3.3 RESEARCH DESIGN

The data collection for this research was based on different data collection methods. The data was obtained from SANRAL, Comprehensive Traffic Observation (CTO) stations, visual observations, and public websites and portals.

*i. SANRAL Crash Data*

Every crash that occurs on the roads under the SANRAL jurisdiction in South Africa is recorded and logged by SANRAL. This data contains information such as the location of the crash, cause of the crash, vehicles involved, and the day on which the crash occurred.

The historical crash data of pedestrian crashes that occurred on the Gauteng FMS Network between the years 2013 - 2018 was obtained from Innovative Transport Solutions (Pty) Ltd. Permission from SANRAL to use this data was obtained. The crash data was plotted on Google Earth to determine the locations where a high density of pedestrian crashes occurred.

Due to the process used to log the pedestrian crashes, it was assumed that there are errors in the data. The detection time might, for instance, be regarded as the incident time; however, a few minutes or maybe even more time might have passed after the occurrence of the incident before the incident was detected and reported. The difference in the incident time and the detection time would normally not differ by much, and it was assumed that the difference between the incident and detection time is small enough to fall within the same peak period, e.g. the AM peak period. The location at which the crash occurred is another factor that can be logged in different ways. The pedestrian might have been trajected into a different lane due to the impact force between the vehicle and the pedestrian, especially if the vehicle travelled at a high speed at the point of impact. The location that was logged might therefore not be the exact location where the primary impact point occurred, but will instead be at the location where the second impact point, i.e. where the pedestrian hit the ground, occurred.

It was, however, assumed that the logging of the data was done with a reasonable degree of accuracy and that the SANRAL data will, therefore, be sufficient for the analysis purposes of this study.

*ii. CTO Stations*

CTO Stations are located along the entire FMS Network. These stations are typically located at the ramps of interchanges and record information such as the speed that was driven on the road and the traffic volumes that were present on the road. Syntell is the company that manages these stations on behalf of SANRAL. CTO station data for 50 stations were requested to determine the traffic and speed data for different times of day and days of the week for the study area. The detailed data for these stations were not available for the year 2013, but the data from 2014 to 2018 was provided for the analyses purposes. The SANRAL Yearbook of 2013 was used to obtain the necessary speed and traffic data for the year 2013.

The 26 sub-sections in which the study area was sub-divided were associated with a CTO station, one in each direction. The CTO stations located at the on- and off-ramps of the interchanges were used to extract information from in order to calculate the speed and traffic volume per time of day on every subsection as well as on the ramps of the interchanges.

*iii. Visual Observations*

Visual observations were carried out to determine pedestrian crossing behaviour, land use in the vicinity of the study area and the number of lanes that has to be crossed to get from the one side of the road to the other side. These observations were made using a combination of Google Earth, the GRCO ESRI platform, a visit to the TMC, live camera feed which is available on the iTraffic website as well as by conducting physical site visits.

## 3.4 METHODS

### 3.4.1 Pedestrian Crossing Behaviour

The pedestrians who walk along the freeway and those who cross the freeway and on- and off-ramps were observed to determine the typical crossing behaviour of the people living in the vicinity of the freeway. A site visit was conducted on 19 April 2019 and 15 and 17 June 2019, the TMC was visited in the AM peak hour of 20 May 2019 and the CCTV camera feed on the traffic website was observed between 1 May 2019 and 17 May 2019 to determine how pedestrians behave along the section of freeway that is included in the study area.

During the site visits and the visit to the TMC, it was determined that the majority of the pedestrians walk along the freeway and either crossed at the ramps of the freeway or at the interchange itself. During the observations, it was established that less than 5% of the pedestrians crossed the freeway midblock.

A pedestrian count was conducted on the N3 Eastern bypass between the London Road and Marlboro Road interchanges, which are located adjacent to the Alexandra informal settlement. This settlement is one of the biggest informal settlements in the study area, and it was therefore assumed that the data obtained from the pedestrian count, would give a reasonable estimation of how pedestrians behave along the freeway and would also give an indication of the locations where they prefer to cross. It was assumed that the pedestrian behaviour on all the sections of the freeway included in the study area would be similar to the pedestrian behaviour at the London Road interchange. Table 3 gives a summary of the pedestrian behaviour during the AM peak hour of a typical day:

*Table 3: Pedestrian count summary*

| No | Pedestrian activity | Pedestrians per hour |
|---|---|---|
| 1 | Pedestrians walking on the informal walkways on the eastern side of the N3 (Marlboro Road to London Road) | 20 |
| 2 | Pedestrians walking on the informal walkways on the eastern side of the N3 (London Road to Marlboro Road) | 215 |
| 3 | Pedestrians walking on the informal walkways on the western side of the N3 (Marlboro Road to London Road) | 85 |
| 4 | Pedestrians walking on the informal walkways on the western side of the N3 (London Road to Marlboro Road) | 200 |
| 5 | Pedestrians crossing at the ramps of the Marlboro Road interchange (all directions) | 505 |
| 6 | Pedestrians crossing at the ramps of the London Road interchange (all directions) | 145 |
| 7 | Pedestrians crossing the freeway at any location between the London Road and Marlboro Road interchanges | 5 |

It is therefore evident that the majority of the pedestrians walk along the freeway to cross either at the ramps of the interchanges or walk along the ramps to cross at the ramp terminal. It was observed that pedestrians waited and were picked up along the freeway as well as on the ramps of the interchange, as indicated in Figure 8.

*Figure 8: Pedestrians waiting to be picked up by public transport vehicles*

It is also evident from the pedestrian count that the number of pedestrians that cross the freeway is significantly lower than the number of pedestrians that cross the ramps of the interchanges, which is consistent with the observations made during the site visits and the visit to the TMC. It should, however, be noted that pedestrians tend to present a different behaviour when they realise that they are observed. The pedestrians that would cross the freeway when they are not observed might, therefore, be different from the number of pedestrians that crossed the freeway when the pedestrian count was done.

Due to the barriers and median that is installed on the entire section of the FMS Network included in this study, the pedestrians that have to cross the freeway are forced to cross in a rolling-gap manner. The pedestrians cannot cross both traffic streams using a single-stage crossing method. This reduces the number of conflict points that pedestrians have to anticipate when they cross the freeway since it is only necessary to observe the oncoming traffic in one direction and not in both directions. The rolling-gap crossing method also decreases the gap that has to be taken to safely cross the freeway since the pedestrians only have to cross one direction at a time. An inside shoulder is provided on the entire section of the freeway, which provides refugee for the pedestrians to climb over the barrier and wait for a gap in the oncoming traffic stream to cross the second half of the freeway. It is noted that the shoulder is not implemented to be used as a walkway or as a refuge island for the pedestrians. It is; however, safer for the pedestrians to be able to wait on the shoulder than to have to sit on the barrier while they wait for another gap in the traffic stream to cross the second half of the freeway.

Pedestrian footpaths were observed along the freeway, indicating that pedestrians walk along the shoulders of the freeway. This was confirmed during the site visit as well as on the footage that was seen in the TMC. The majority of the pedestrians that were observed were walking along the freeway or waiting at the ramps or on the shoulder of the freeway for their transport.

The following crossing behaviours were observed:

1. Pedestrians walk on the side of the road or on the outside shoulder of the freeway to get to their destinations without crossing the freeway.
2. Pedestrians walk along the freeway towards the ramps of an interchange, cross only the ramps and continue walking along the freeway
3. Pedestrians walk along the ramp towards the ramp terminal where they cross and continue walking on the internal roads leading to and from the freeway.
4. Pedestrians wait on the outside shoulder for a sufficient gap to cross the first half of the freeway, climb over the barrier and wait for a second gap to cross the second half of the freeway.
5. Pedestrians wait on the outside shoulder for a sufficient gap to cross the first half of the freeway, climb over the barrier walk along the barrier for some time before waiting for a second gap to cross the second half of the freeway

It was observed that pedestrians tend to walk along the barrier on the freeway when one of the sides of the freeway is fenced off. The pedestrian, therefore, crosses the first part of the freeway, but cannot cross the second half of the freeway due to fencing on the side of the freeway or due to the gradient on the other side of the freeway which prevents them from leaving the freeway. They are therefore forced to walk on the median or along the barrier until there is a gap in the fencing or until they reach an interchange where they can cross and walk along the ramp to exit the freeway.

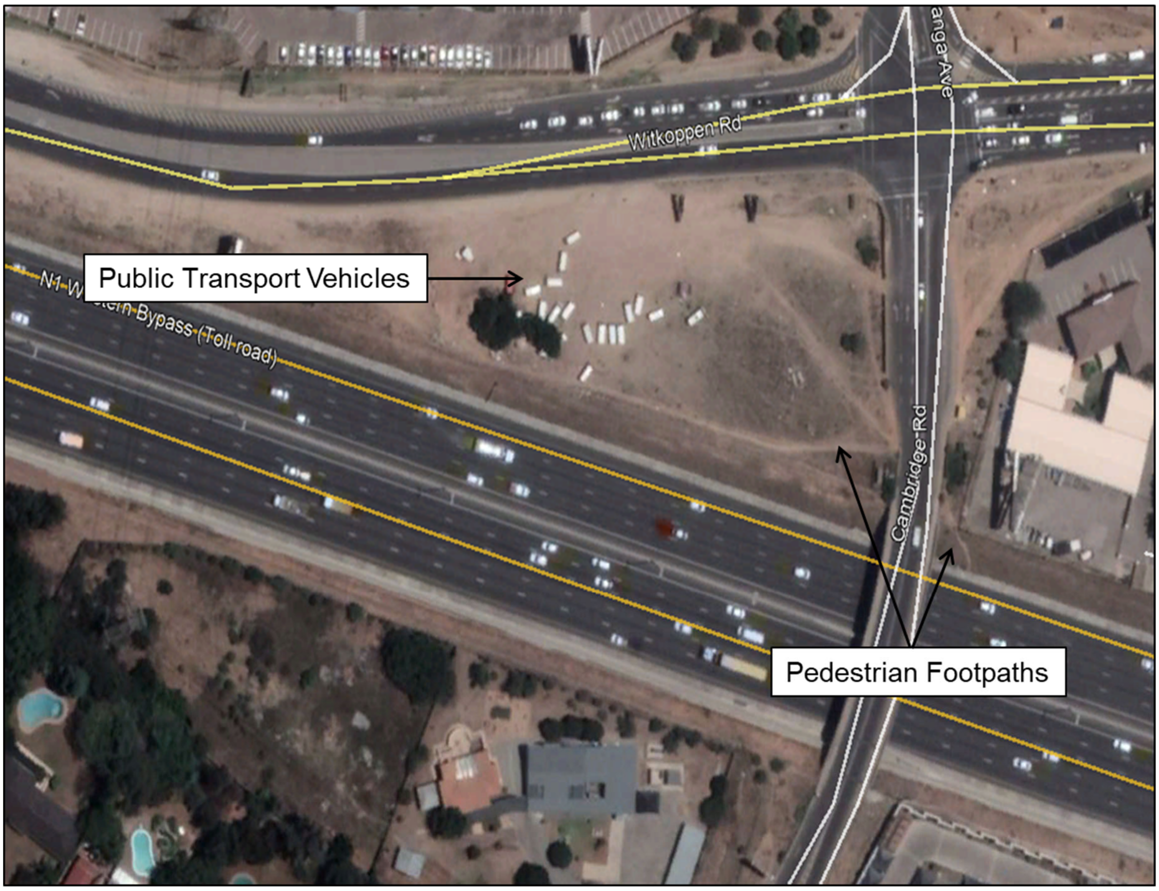



*Figure 9: Pedestrian footpaths and public transport vehicles at the Witkoppen Interchange*

Figure 9 above is a photo of the Witkoppen interchange. It is evident from Figure 9 that there are major pedestrian and public transport activity in the area, with the majority of the pedestrian footpaths leading to the ramp of the interchange. It is evident that there is an informal mini-bus taxi holding facility at this interchange. There is a pedestrian footpath that leads to the freeway; however, there is no evidence that the pedestrians cross the freeway in a straight line since the footpath is only located on the one side of the road. The pedestrians might, therefore, walk on the shoulder of the road to a different location before they cross the road. Another possibility is that the pedestrians are picked up and dropped off by public transport vehicles at this point, from which they can take another mini-bus taxi at the informal holding facility.

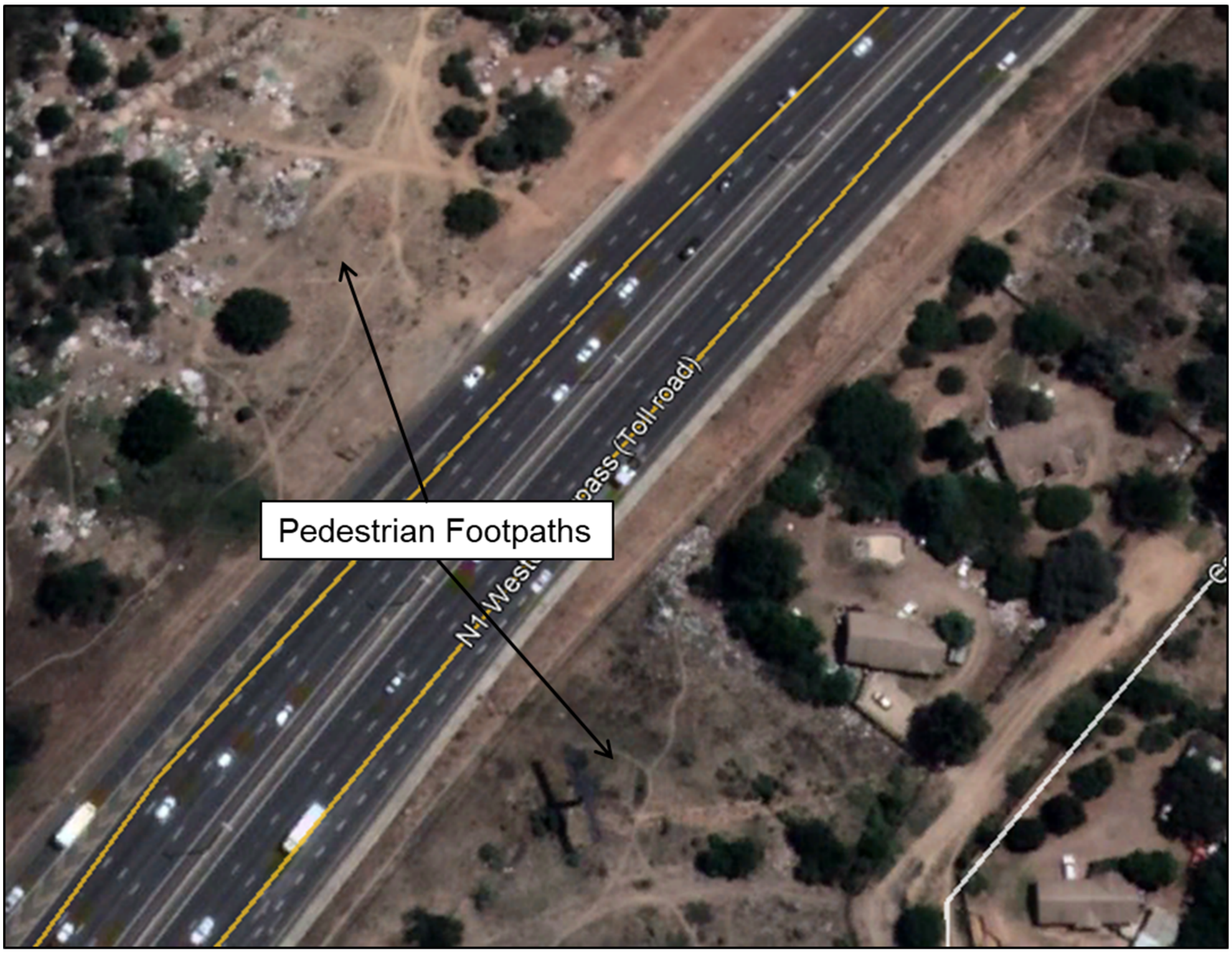



*Figure 10: Pedestrian footpaths leading to the N1 Western Bypass*

A high number of pedestrian footpaths are visible in Figure 10. The majority of the footpaths are located on the western side of the N1 Western Bypass. A reason for this is due to the informal settlement on the western side of the road, resulting in a high number of pedestrians that walk to the freeway from different sides of the settlement. Pedestrians might also walk on the shoulder of the eastern side of the road before they divert away from the freeway to their destinations.


*Figure 11: Pedestrian footpaths leading to and from the ramps of the Malibongwe Interchange*

Figure 11 gives a representation of the footpaths at the Malibongwe Interchange. It is clear from the figure that the pedestrians cross the freeway at the ramps as well as between the on- and off-ramps. It can, however, be assumed that the majority of the pedestrians cross the freeway at the ramps since the majority of the footpaths leads to and from the ramps of the freeway. These types of footpaths confirm the observation made in the pedestrian count data.

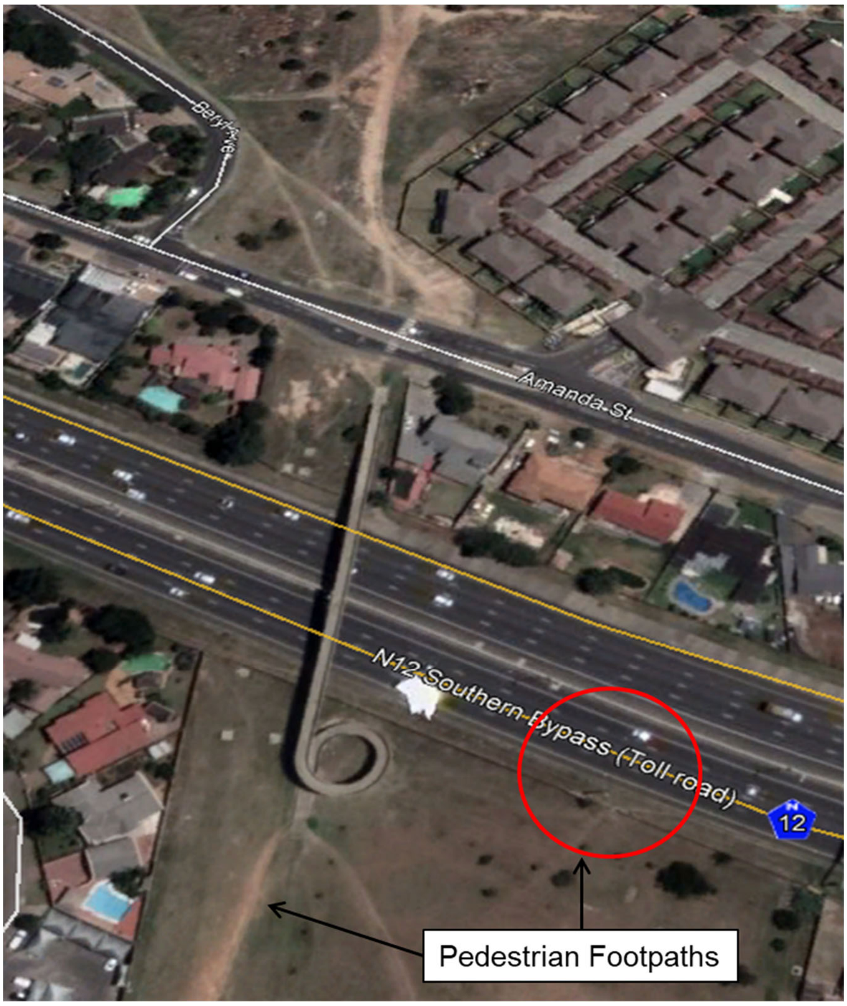



*Figure 12: Pedestrian bridge and pedestrian footpath at N12 Southern Bypass*

It is clear from Figure 12 that pedestrians make use of the existing pedestrian bridge, located to the east of the Klip Rivier interchange. There is, however, also evidence of pedestrians who do not make use of the pedestrian bridge but instead cross the freeway midblock.

### 3.4.2 Average Traffic Volume

The traffic data of three different types of cross-sections were used in this study. Figure 13 gives a graphical representation of the different type of volume sections.

i. Traffic volume between two adjacent interchanges

   This is the section of road located between two adjacent interchanges, i.e., from the on-ramp at interchange 1 to the off-ramp at interchange 2. For this section of road, the ramp volume was added to the traffic that was already travelling on the freeway. This is referred to as the total traffic volume.

ii. Traffic volume between ramps of an interchange

This is the section of freeway between the off- and on-ramps of the same interchange. This traffic volume is the traffic volume where the traffic that left the freeway via the off-ramp was subtracted from the total traffic volume, but the traffic that was travelling on the on-ramp to enter the freeway was not yet added to the freeway traffic.

iii. Ramp traffic volume

This is the traffic that was travelling on the ramps of the interchanges. The traffic volume travelling on the freeway was not added to this volume.

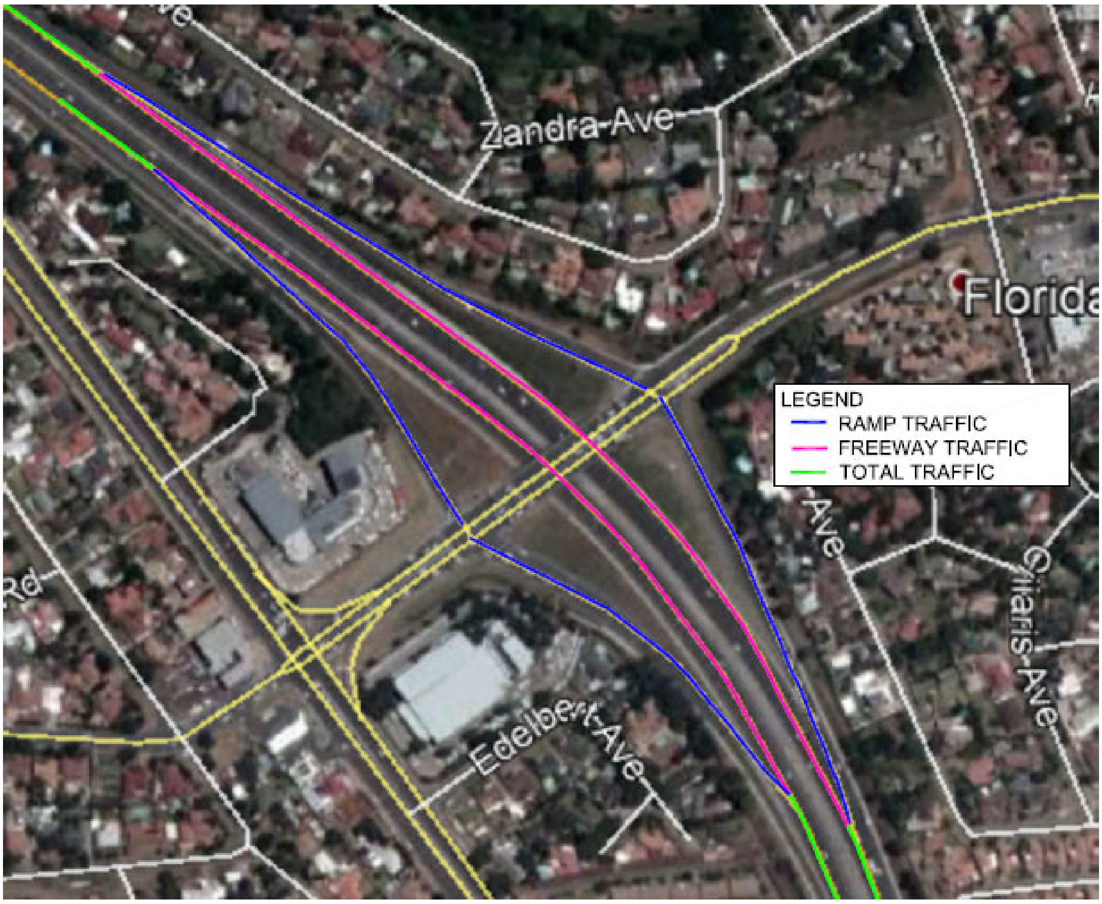



*Figure 13: Different sections of an interchange which was used for traffic volume assignment*

The traffic volume will not vary along with the sub-section, since traffic can only get on and off of the freeway at the interchanges. The same traffic volume was therefore assigned to the crashes that occurred on each sub-section of the freeway, i.e., the same traffic volume was assigned to all the crashes that happened in the AM peak on sub-section 1 in 2013.

***Average traffic volume at the time of the incident***

It is known that traffic volumes differ for all hours of a day. Research, however, indicates that the traffic flow patterns are similar for every day of the week, e.g., every Monday follows the same traffic flow pattern, the same traffic flow pattern is observed for every Tuesday, and so on (Yeon, Hernandez and Elefteriadou, 2009). The exact traffic volume that was travelling on the road, at the exact time and location where the crash occurred would be the most accurate volume to use, however, the time at which the crash was reported and logged might be slightly different from the time at which the crash happened. The traffic volume per minute of the day was also not available, and the traffic volume would be estimated to the nearest hour if the exact traffic volume was to be used. Also, some of the CTO stations did not record 100% of the time during the year, resulting in some of the crashes for which there will be no data to be used. Since it has been determined through the literature review that the average traffic on every Monday, Tuesday, etc. is similar, it was assumed that the average traffic per time of day would be sufficiently accurate for the purpose of this study. The average traffic flow per peak hour for all days of the week was, therefore, assigned to the pedestrian crashes. This data was also obtained from the CTO stations. Assumptions of the traffic volume at the locations where no data was available for specific days or years, was made. The typical daily traffic volumes for the year 2013 are available in the 2013 SANRAL Yearbook. The typical volume per hour of the day for every day of the week is presented on graphs. The average traffic that was expected per time of day for every day of the week for the year 2013 was therefore determined by reading off the typical traffic per hour of the day for all seven days of the week. The average growth rate for the traffic volumes can be determined using the 2014 to 2018 data. The 2013 traffic volumes could also be determined using this method; however, it was preferred to use the recorded data from the SANRAL Yearbook, since the graphs represent the actual recorded data, while calculating the traffic volumes would mean that estimated growth rates would have to be applied. The methods used to determine the unavailable traffic volume is described below.

***Methods to determine the unavailable / missing traffic volume data***

1. The same CTO station

This method was used to determine the missing traffic volumes for the stations where the data of one year was missing. The growth rate over the available four years was determined, and this was applied to calculate the volumes of the year for which the data was not available. The calculated traffic volumes were thereafter compared with the data at the adjacent sub-sections of the same year using the adjacent CTO stations. This was done to ensure that there were no discrepancies between the recorded and calculated traffic volumes. The trend, i.e. difference in volumes between adjacent interchanges per time of day and per year, for the available years were determined and compared with the trend for the calculated data to determine whether the estimated traffic volumes were in the correct category or not.

2. Adjacent CTO stations

For some of the CTO stations, the data for only one year was available, e.g. only for the year 2018. Traffic growth patterns cannot be determined from the traffic volume data of only one year. The data for the adjacent stations were therefore used to determine a growth rate. The two separate growth rates, i.e. the growth rate for both adjacent CTO stations were compared with one another. If the growth rate for both subsections were similar, it was applied to the CTO where the data was outstanding to determine the traffic volumes for the missing years. Only the total (freeway + ramp) volume could be determined using this method since the volumes that were travelling on the on- and off-ramps for the different interchanges are not the same and would not necessarily follow the same growth patterns. The problem with this method is that the ramp volume cannot be estimated accurately. There were, however, no crashes that happened on the ramps during the years where there were no data are available, and this was therefore assumed to not be a problem for this study.

3. SANRAL Yearbook information

The SANRAL yearbook information for the historic years was used to determine the typical traffic flow for CTO stations where data was missing. This was done for all the stations to determine the typical traffic for every day of the week for the year 2013. This data was compared to the traffic volumes recorded for the 2014 – 2018 years to determine whether the estimated traffic volumes were in the same range than the data for the rest of the analysis years.

4. AADT information

The Average Annual Daily Traffic (AADT) for each station can be found in the Google Earth SANRAL yearbook data. The information for the years 2013, 2015, 2016, and 2017 was available. This data was used to determine whether the estimated data for years where no data was available, fall within the same range of the AADT of the station. The 2013 AADT was compared with the AADT of the other years to determine whether the estimated 2013 data, which was derived using the typical volume graphs in the 2013 SANRAL yearbook, was estimated within a reasonable margin of accuracy. Growth rates were also calculated using the AADT data for the different years. This growth rate was compared with the growth rate determined from the hourly CTO data to ensure that an accurate growth was used to determine outstanding data.

It is expected that there is a difference in the estimated data and the data that probably travelled on the freeway during the year for which the data was not available (Table 4), however, all the data used for the model was average data rather than exact data. It is therefore assumed that the small discrepancy between the calculated data and the estimated data will not have a significant impact on the model.

*Table 4: CTO stations with incomplete traffic volume data*

| CTO station | Data missing | Solved |
|---|---|---|
| 1939 | 2018 | Determined average traffic volume using average growth rates of the available years |
| 1938 | 2015 | Determined average traffic volume using average growth rates of the available years |
| 1904 | 2014-2016 | Determined average traffic per peak hour using adjacent CTO station data |
| 1895 | 2014-2017 | Determined average traffic per peak hour using adjacent CTO station data |
| 1946 | 2014-2017 | Determined average traffic per peak hour using adjacent CTO station data |

The traffic volume in the direction of the crash, and not in both directions of the crash, was used as part of the model. The reason for this is because the pedestrian who crossed the road was forced to cross in two stages, resulting in the opportunity to wait on the median or inside shoulder of the road for another gap in the traffic stream before the pedestrian crossed the second half of the road. It was, therefore, assumed that the traffic flow that was present on the road in the direction of the crash was the only volume that played a role in the crash.

### 3.4.3 Average Travel Speed

Previous studies have determined that the impact speed in a pedestrian crash has a significant impact on the injuries sustained by the pedestrian involved in the crash (Kröyer, 2015; Martin and Wu, 2018). Speed also increases a pedestrian’s risk of being involved in a pedestrian crash (Forbes, 2012).

The speed travelled on the freeway is dependent on factors such as the cross-section (geometry) of the road as well as on the traffic flow, i.e., if there are high traffic volumes, the travel speed will approach congested, or breakdown, speed. The travel speed will be close to free-flow speed on sections where there are low traffic volumes. It is, therefore, unlikely that there would be a significant deviation in the travel speed on every sub-section. It was, hence, assumed that the entire sub-section of the freeway in between the interchanges would have the same travel speed and traffic volume.

Two different speed types are recorded at the CTO stations, namely the $85^{th}$ percentile speed and the average speed travelled. The speed travelled per lane and per vehicle type (light vehicles and heavy vehicles) are recorded. The average speed travelled for both vehicle types were used as part of this study. This was done because the speed that the pedestrian had to take into account in choosing the crossing gaps before crossing the freeway includes a combination of the light and heavy vehicle speed. The speed that was used as part of this study was therefore not the impact speed (which is not, in fact, recorded), but rather the average speed of all the vehicles on the road since it is assumed that this speed will play a role in the decision to cross or to wait for another gap.

The exact location, namely the lane in which the crash occurred, was not known. The speed of the specific lane in which the crash occurred could, therefore, not have been used as part of the study.

The $85^{th}$ percentile speed is the speed limit at or below which 85% of the vehicles on the road travel. This means that 15% of the drivers exceed this speed limit. The reason why the $85^{th}$ percentile speed is not used as for the analyses is because there is no indication as to the percentage of people who actually travel at the $85^{th}$ percentile speed. The speed that most of the people travel is also unknown since it is only known that most people drive at or below this speed, but not the actual travel speed. The average speed, on the other hand, takes into account the speed travelled by all the vehicles over a certain distance and time. This gives an actual indication as to how fast the average person drives, while the $85^{th}$ percentile speed does not. The $85^{th}$

percentile speed might give a wrong impression of the actual speed that is driven on the road. If 50% of the drivers keep to a 60km/h speed limit, 10% at a 65km/h and 5% at an 80km/h speed limit, a total of 85% of the drivers drive at or below 80km/h. This speed, however, does not give an indication of the speed that most people on the road travel at.

As mentioned before, the exact impact speed in pedestrian crashes was not recorded, and the exact impact speed can therefore not be used as part of the model. Average speed data can, however, be extracted from comprehensive traffic observation (CTO) stations. The average travel speed was used since this gives a better representation of the actual speed that the majority of the vehicles travel at. The speed limit of a residential road gives a good indication of the speed that is travelled on the road. The speed limit on the South African freeways is; however, not a good representation of the speed travelled on the freeway, since the speed limit is not always adhered to, especially at night or during free-flow conditions. It is normally not possible to drive at the speed limit during the times when the freeway is congested. The most probable travel speed would then be the congested (breakdown) speed. Research has also indicated that average speed has an impact on the risk and injuries sustained in pedestrian crashes (Golob, Recker and Alvarez, 2004, Tanishita and van Wee, 2017). It was therefore assumed that the average travel speed would be the most suitable speed to use for this study.

Speed data

The hourly average speed travelled per sub-section and direction of travel (e.g., Northbound) on the road was extracted from the speed data obtained from the CTO stations for the years 2014 to 2018. The average speed per time of day, i.e. AM peak, PM peak and so on, was calculated for every sub-section of the freeway by taking the average of the hourly average speed travelled per direction for the hours that are included in every peak period. The average travel speed used for crashes that happened at an interchange between the on- and off-ramps were assumed to be the same as for the crashes that happened on the sections between interchanges. It was assumed that the number of vehicles that enter and leave the freeway at the ramps would not make a significant difference in the average travel speed on the freeway since the ramp volumes accelerate to the freeway speed and not the other way around. At locations where there is major congestion, resulting in vehicle queues on the ramps or freeways, the speed of the freeway and ramps will be similar, since the vehicles will queue back onto the ramps. The conditions on the freeway are therefore the factor that influences the speed travelled on both the ramps and the freeway sections.

The speed data was assigned to all the pedestrian crashes, namely the 2014 AM peak data for sub-section 1 was assigned to the crashes that happened in the 2014 AM peak in sub-section 1, the 2018 PM data for sub-section 15 was assigned to the crashes that happened in the 2018 PM peak in sub-section 15, and so on.

There are CTO stations where no data for certain years were available. Table 5 gives a summary of the CTO stations and the corresponding sub-sections where there are gaps in the data.

*Table 5: CTO stations with incomplete travel speed data*

| CTO station | Data missing | Solved |
|---|---|---|
| 1939 | 2018 | Determined average traffic speed using average growth rates of the available years* |
| 1938 | 2015 | Determined average traffic speed using average growth rates of the available years* |
| 1904 | 2014-2016 | Determined average traffic speed per peak hour using adjacent CTO station data |
| 1895 | 2014-2017 | Determined average traffic speed per peak hour using adjacent CTO station data |
| 1946 | 2014-2017 | Determined average traffic speed per peak hour using adjacent CTO station data |

*The average speed data that was available was compared with one another, and it was determined that the travel speeds over the analysis period did not change significantly. The average speed travelled on the sub-section of the previous year was therefore used for the sub-sections where no speed data was available.

### 3.4.4 Land Use Along the Freeway

Research has indicated that there is a relationship between pedestrian crashes and land uses in the area surrounding the crash site (Priyantha Wedagama, Bird and Metcalfe, 2006, Osama and Sayed, 2017). The land use surrounding the Gauteng FMS Network consists mainly of residential, business, industrial and retail land use. These four categories were, therefore, used as part of this study. The definitions of each of these land use zonings, according to the City of Tshwane Town Planning Scheme are given below.

*Residential: "Means a hotel, block of flats, tenements, boarding house, and hostel together with such outbuildings as are ordinarily used therewith."* (City Planning and Development Division, 2008)

*Retail (Retail Industry): "Means, inter alia, catering, a confectionary, dress-making, and tailoring, engraving, instant printing and copying, jewellery manufacturing, photographic processing, picture framing, and screen printing; as well as the servicing and repair of air conditioners, audio equipment, basket ware and cane furniture, canvass goods and tents, bicycles, electronic equipment, domestic equipment, leather-works and shoes, office equipment, television and video equipment, upholstery, watches, weighing machines and window blinds, but does not include a Light Industry and the wholesale selling of goods."* (City Planning and Development Division, 2008)

*Industrial: "Means land and buildings where a product or part of a product is manufactured, mounted, processed, repaired, rebuilt or packed, including a power station and incinerator plant and may include a cafeteria and a caretaker's flat and any other activities connected to or incidental to the activities mentioned herein, excluding noxious industries, light industries and retail industries."* (City Planning and Development Division, 2008)

*Light Industry: "Means land and buildings used for, inter alia, a bakery, a builder's yard, a car wash, a contractor's yard, dry-cleaners, carpet cleaners, joinery workshop, launderette, laundry, lawnmower workshop, painter's workshop, plumber's workshop, printing workshop, transport depot, panel-beater, motor workshops, a ready-mix plant and any other such industries, workshops or yards which in the opinion of the Municipality do not cause a nuisance to the environment, may be used for similar purposes and may include the retail sale of products ancillary and subservient to the main use on the same property."* (City Planning and Development Division, 2008)

*Business: "Means land and buildings used as an office, financial institution, fitness centre, hairdresser, receiving depot for dry-cleaning and shoe repairs, dental workshop, medical and dental consulting rooms, optometrist or for other business purposes such as inter alia beauty salon, pet salon, but does not include any building mentioned whether by way of inclusion or exclusion in the definition of Institution nor a building designed or used as a Place of Instruction, Place of Amusement, Shop, Public Garage, Parking Garage, Industry, Noxious Industry, Warehouse, Vehicle Sales Mart or a factory or workshop."* (City Planning and Development Division, 2008)

The land use data was incorporated into the model to determine whether a relationship between the land use in close proximity of the Gauteng FMS Network and the pedestrian crashes exists. It was assumed that the land use from 2013 to 2018 did not change considerably, and the existing 2018 land use patterns were therefore used to determine the land use in proximity of every crash.

The land use in the surrounding area of the freeway is such that "mixed-use" is the appropriate description of the land use to the majority of the crashes. The impact of individual land uses is; however, of importance since this variable aims to determine whether specific land uses have a more significant impact on pedestrian crashes or not. The selection of the land use that was assigned to the crash is discussed below.

The land use in the area was observed, and it was not only the land use adjacent to the freeway that was recorded but also the land use in the area surrounding the freeway that was taken into consideration. This was done using Google Earth aerial and street view, as well as using the GCRO GIS Viewer. These land use zonings were linked with the pedestrian crashes that occurred along the freeway. It should be noted that the informal settlements were classified as "residential" on the GCRO GIS Viewer, and these settlements were, therefore, classified as residential land use as part of this study. The historical aerial photos on Google Earth were used to ensure that the land use did not change between 2013 and 2018.

It was assumed that the majority of the pedestrians that were involved in pedestrian crashes were most likely to be residents of the informal township settlements located along the freeway (this assumption was confirmed by Sinclair & Zuidgeest (2016), in their study near Khayelitsha, Western Cape). It was not expected that the residential houses along the freeway would be the major attraction of the pedestrian trips, but rather that the pedestrians would walk to the retail centres to wait for job opportunities. Other pedestrians were expected to work at the retail centres, business developments as well as at the industrial zones. For the zones where only residential units were present along the freeway, the land use that was linked to the crash was assumed to be residential. If there was, however, a mix between low-density housing and retail or industrial land uses, the retail and industrial land use was assumed to be the attraction of the trips. The day on which the crash occurred was thereafter taken into account in the decision to link the specific land use with the crash. If retail and business land uses were in close proximity of the crash that occurred during the weekend, the retail land use was expected to be the attraction land use. Specific land uses are known for generating higher traffic volumes during certain hours of the day than other land uses, e.g., the critical peak wherein traffic are generated for retail zonings are Friday PM and Saturday peak periods. Business areas, on the other hand, generate traffic in the

AM and PM peak periods of typical weekdays. It was therefore not expected that crashes that happened on Saturday nights were due to the business land use, but would rather be influenced by retail land use.

### 3.4.5 Vehicle Type

Research indicated that the type of vehicle that was involved in a low speed crash had an impact on the injuries sustained (Martin and Wu, 2018, Ballesteros, Dischinger and Langenberg, 2004). The vehicle type that was involved in the crash was obtained from the SANRAL crash data. The vehicle type involved in the pedestrian crash was not recorded for 11% of the crashes. The percentage split between the different vehicle categories was calculated and was distributed in the same proportion to the crashes where no vehicle type was logged. Three different vehicle categories were taken into account as part of the study, namely:

1. Passenger vehicles
2. Light Delivery Vehicles (LDV)
3. Heavy vehicles (HV)

The crashes where the vehicle involved in the crash was classified as a “bakkie” or “SUV” was regarded to be part of the passenger vehicle category. The reason for this is because the majority of the bakkies and SUV that are present on the road are used for private family use and can, therefore, be classified as passenger vehicles.

### 3.4.6 Day of the Week

The purpose of trips made during the week usually is different from the trip purpose during the weekend. The trips conducted on a weekday are typically regarded as work-based trips, while the trips made during weekends are classified as leisure trips. This variable was included in determining whether the day of the week on which a pedestrian crash occurred had an impact on the crash rate and injuries sustained. The Cambridge definition of a weekday states that a weekday is any day between Sunday and Saturday (Cambridge Dictionary), and the definition of the weekend is the time between Friday evening and Sunday night. Fridays, therefore, fall in both categories. The AM and off peak hours of a Friday was therefore assumed to be part of the weekday hours, but the PM and night time crashes were assumed to be part of the weekend crashes.

The day of the week at which the pedestrian crashes occurred, are logged in the SANRAL crash data. The data that was used as part of the model was simplified by only using “weekday” and “weekend” instead of all seven different days of the week. It was not expected that this would have a negative influence on the crash data since it is known that the number of crashes that happen during weekdays does not vary much (RTMC, 2017)

### 3.4.7 Time of the Day

Research has indicated that the majority of the fatal pedestrian crashes happen during night time (Amoh-Gyimah *et al.*, 2017; Regev, Rolison and Moutari, 2018). In South Africa, the majority of the fatal crashes occur during the late afternoons and early evenings (RTMC, 2017). The incident time variable was used to determine whether there was a relationship between the time at which a pedestrian crash occurred and the crash rate as well as injuries sustained in the crash. The exact time of the incidents is not always recorded accurately, as mentioned previously. It was, therefore, decided not to use the exact time of the crash, but rather the peak period in which the crash falls. It was assumed that the incident time was recorded accurately enough to fall in the same peak period as in which it occurred. The time at which the crash occurred were grouped into four different peaks, namely:

- AM Peak (06:00-08:59)
- Off Peak (09:00-14:59)
- PM Peak (15:00-18:59)
- Night Peak (19:00-05:59)

The peak period at which every crash occurred was determined and assigned to the crash data. This data was used to determine which traffic flow volume and travel speed have to be assigned to the individual crashes.

### 3.4.8 Number of Lanes to Cross

The distance to cross the freeway is dependent on the number of lanes that the pedestrian has to cross. Research indicated that the number of lanes to be crossed have a negative impact on the pedestrian crashes (Zhang, Chen, and Wei, 2019). The higher the number of lanes to cross, the more conflict points between the pedestrian and the vehicles occur. A longer gap is required to cross the road, with speed differentials to consider between every lane. In the Gauteng province, the right-hand lane is usually associated with the highest speed, and the left-hand lane is associated with the lowest speed. Pedestrians do not always consider this when waiting for a gap in the traffic stream to cross the road. When they start to cross the road from the slow lane, the pedestrians do not always realise that they not only have to take a long enough gap to be able to cross but that they also have to take the higher travel speeds of the vehicles in the fast lanes into account.

The exact distance to cross the freeway was not taken into account. This was done since it was assumed that vehicles do not travel in the shoulders of the freeway. The number of lanes was, therefore, the attribute that was considered for the analysis purposes in this study. Another reason why the number of lanes to cross was considered, was due to the reason that the number of conflict points increases as the number of lanes increase and not necessarily as the distance to cross increases.

The number of lanes that the pedestrian had to cross at the pedestrian crash location was determined using Google Earth. The number of lanes for each crash location as well as for the different analysis years was confirmed using the historical imagery on Google Earth. It was determined that there were no lanes added to any of the crash locations between the years 2013 and 2018.

### 3.4.9 Injuries Sustained

The injuries sustained in the pedestrian crashes were determined using the SANRAL crash data. The injuries were classified as none, slight, severe, and fatal. It is expected that there might be underreporting in the crashes where no injuries were recorded, since, if there was no damage to the vehicle and the pedestrian was able to walk away without medical treatment, the crash might not be attended to and would therefore not be logged in the system.

It was expected that there might be underreporting of fatal injuries. The injury classification of a pedestrian who sustained severe injuries in the crash, but who passed away later at the hospital was not necessarily changed from severe to fatal injuries. A person sustained fatal injuries when he passed away on the scene of the crash or within 30 days at the hospital. This threshold is not always able to be applied, since the information that the patient has passed away in the hospital is not always communicated to SANRAL, and the change in the injury type is therefore not logged into the system.

### 3.4.10 Crash Rate

The definition of a crash rate according to the Federal Highway Association is given below.

“Crash rates describe the number of crashes in a given period as compared to the traffic volume (or exposure) to crashes” (Federal Highway Association).

In traffic engineering, the daily crash rate is usually calculated. The average annual daily traffic (AADT) is used for this calculation. The risk of being in a pedestrian crash for different times of the day was one of the attributes used in the model. The daily crash rate was therefore not used, but the crash rates for certain times of the day were determined so that it can be related to the time of day attribute.

The crash rate was determined per sub-section of the road. The average annual traffic volume per time of day was used, instead of the average annual daily traffic volume. The calculated crash rate is therefore not a daily crash rate but is broken down into a crash rate for the AM peak, PM peak, off peak and night peak. One crash rate for the six years was determined, using the CTO station data and 2013 SANRAL Yearbook data.

It is noted that the crash rates that are typically calculated should preferably be calculated using a constant length of road. In the case of this study, this will not be feasible, since the interchange spacing between the interchanges on the Gauteng FMS Network is not the same for all the interchanges. The traffic volume used in the crash rate calculation should be constant for every sub-section. If the study area were divided into sections of equal length, it would not be possible to use an accurate traffic volume, since the traffic volume on the section of road will change if the section includes more than one interchange. It was therefore assumed that the most accurate calculation for the crash rate would be when the sections of road in between the interchanges are used, to ensure that the traffic volume that is used stay constant.

The formula to calculate the crash rate on a road segment is given below:

$$Crash\ Rate = \frac{Number\ of\ crashes\ per\ segment \times 1\ 000\ 000}{N \times AAPPT \times 365}$$

$$N = number\ of\ years$$

$$AAPPT = Average\ annual\ peak\ period\ traffic$$

The crash rate per sub-section and per time of day was calculated and assigned to every crash. If a crash occurred in the night peak of sub-section 7, the crash rate that was calculated for the night peak on sub-section 7, was assigned to the crash. This was done for all the crashes.

## 3.5 MODELLING

### 3.5.1 Introduction to the Multinomial Logit (MNL) Model

MNL models predict the categorical placement of a dependent variable based on a list of independent variables. These variables are assumed to be unordered and as the statement above already indicated, independent of each other. Dummy variables are applied to these types of models, with the dependent variable, i.e., the y-coefficient, that takes on the dummy variable. The dependent variable can have more than one category. If the dependent variable has K variables, K-1 dummy variables will be applied, and K-1 models will be developed to explain the behaviour of the independent variables on the dependent variable (Liao, 2011, Starkweather and Key Moske, 2002).

An MNL model is typically used to predict the probable outcome of a dependent variable by investigating the relationship between the dependent variable with a list of independent variables into account, (Statistics Solutions, 2019). The data used to develop an MNL model has to apply to a few assumptions and requirements,

- **Multicollinearity:** The variable included in the data set should be independent of each other. When the variables are dependent on each other, one of the variables should be removed from the data set;
- **Normality:** the distribution of the errors are assumed to follow a normal distribution;
- The **variance** of all the variables should be homogeneous;
- One or more of the independent variables should be **continuous**; and
- The dependent variable should be **binary** or **non-continuous**.

Multiple logistic regressions will be applied to the set of data to validate the data set. The coefficients of the independent variables of the model are also determined through the regression analyses.

The following tests can be used to determine which variables are to be included in the model and to rule out any variable that may be dependent on another variable as well as to determine which variables are statistically non-significant.

- ***R-square test:***
  An R-square value between 0.3 and 0.6 is regarded as a good behaviour model. The better the R-square value, the closer the model is to the perfect model.
- ***RMSE Test Value:***
  This value represents the root mean square error. This variable is regarded as a better value to use than an R-square value since the RMSE value is an absolute measure of fit, whereas the R-square value represents the relative measure of fit. An RMSE value as close as possible to 0 is aimed for.

- ***P-test:***
  For this test, a p-stats value has to be calculated. This value indicates whether the variables included in the model is statistically significant. The p-stats value is related to the confidence interval. The p-stats value can take on any value between 0 and 1. If the p-stats value is higher than the significant level (alpha-value), the null hypothesis cannot be rejected, while p-stats values smaller than the significant level indicates that the null hypothesis can be rejected.
- ***T-Test:***
  A **null hypothesis** is formed ($H_0$: b=0), where it is assumed that all the coefficients, except for $\beta_0$, are equal to zero. If the Z-value of a coefficient is greater than 1.96, the null hypothesis can be rejected with a 95% certainty
- ***F-test*:**
  $H_0$=a=b: This test is used to test the significance of all the variables together; however, this is not a very strong test, and will not be used as part of this study.
- ***Sign of the coefficients:***
  Does the sign of the coefficients make logical sense?

These tests are to be applied after every multiple linear regression, to ensure that only the significant variables are included in the model. The model is calibrated when all the variables and the R-square or RMSE variable comply with the abovementioned tests. The coefficients can now be applied to the general model given below.

$$\log{(y/y_0)} = \beta_{Lu}X + \beta_D X + \beta_{Ph}X + \beta_S X + \beta_{TVol}X + \beta_{VehT}X + \beta_L X + C$$

### 3.5.2 Variables Included in the Data Set

Two multinomial logistic regression models were developed as part of this study to determine the relationship between the dependent and the independent variables. The models will be referred to as shown in Table 6.

*Table 6: Multinomial Logistic Models to be developed*

| Model | Dependent Variable | Independent Variables |
|---|---|---|
| Model 1 | Injuries sustained (fatal vs. non-fatal) | Land use in the area<br>Day on which crash occurred<br>The peak in which crash occurred<br>Average travel speed on the road<br>Average traffic volume on the road<br>Vehicle type involved in the crash<br>Number of lanes to cross |
| Model 2 | Crash Rate | Land use in the area<br>Day on which crash occurred<br>The peak in which crash occurred<br>Average travel speed on the road<br>Average traffic volume on the road<br>Vehicle type involved in the crash<br>Number of lanes to cross |

The variables are not ordinal, and an ordinal regression model will therefore not be suitable to model the relationship between the variables. The attributes included in the study are also not hierarchical; thus, the hierarchical MNL model will not be a suitable model to use as part of this study. A multinomial logistic regression model was; however, expected to be suitable since this model does not assume ordinal or hierarchical attributes. It was expected that some of the attributes are dependent on each other. These factors were excluded from the model when the regression was applied to the data.

As already mentioned, the variables of an MNL model should be binary or non-continuous for the dependent variable and continuous, binary, or represented as a ratio for the independent variables. Some of the independent variables included in the models did not comply with the criteria listed above. Variables such as the peak hour in which the crash occurred, or the day on which the crash occurred did not comply with the criteria. These variables were converted to numerical values. Variables with only two outcomes were converted to binary variables, i.e. variables that only take on values equal to 1 or 0, as indicated in Table 7.

*Table 7: Variables included in the Multinomial Logistic Model*

| Variable | Category and Assigned Value |
|---|---|
| **Independent Variables** | |
| Day of crash ($\beta_D$) | Weekday = 1<br>Weekend = 0 |
| Time of day ($\beta_{Ph}$) | AM Peak = 1<br>Off Peak = 2<br>PM Peak = 3<br>Night Peak = 4 |
| Land use surrounding the crash ($\beta_{LU}$) | Residential = 1<br>Industrial = 2<br>Business = 3<br>Retail = 4 |
| Traffic volume ($\beta_{TVol}$) | Average traffic volume per peak period as determined from the CTO data |
| Travel speed ($\beta_s$) | Average speed per peak period as determined from the CTO data |
| Vehicle Type involved ($\beta_{VehT}$) | Passenger vehicle = 1<br>Light delivery truck = 2<br>Heavy vehicle = 3 |
| Cross-section $\beta_L$) | Number of lanes to cross |
| **Dependent Variables** | |
| Injury sustained | Fatal = 0<br>Non-fatal = 1 |
| Crash rate | Crash rates were divided into three categories since this value is continuous:<br>Low crash rate: 0.00 – 0.33<br>Medium crash rate: 0.34 – 0.66<br>High crash rate: 0.67 – 1.25 |

It can also be seen from Table 7 that the crash rate was classified into three different groups. This was done because one of the assumptions of an MNL model is that the dependent variable is not continuous. Crash rates are, in theory, continuous variables; however, it can be converted into a non-continuous variable by dividing the rates into three different groups, as specified above.

### 3.5.3 Significant Testing Methods

#### a) P-test

P testing is widely used in research to determine whether a model is significant or not. The null hypothesis is accepted or rejected based on the p-value. The majority of researchers use this approach as part of their significance testing. Gigerenzer, Krauss, and Vitouch, (2004), however, does not agree with this approach. According to Gigerenzer, Krauss and Vitouch (2004), other tests, such as the power test, can provide more information on the significance of a model than the null hypothesis testing. He also argues that the null hypothesis testing should instead only be done for experiments for which no or almost no information is available. One of the most well-known statisticians, R.A. Fisher, argues that the fact that the generally accepted application of a $p<0.05$ indicates a lack of statistical thinking.

Gigerenzer, Krauss and Vitouch (2004) supports Fisher's thinking by noting that if researchers always use the same significance testing methods, they result in only focusing on a few specific attributes, tend to use the same standard values so that they do not break the rules and rather tend to apply wishful thinking instead of critical thinking. These statements are rather harsh to make; however, it does seem as if researchers blindly apply the same significance tests and values, without checking whether it will suit the type of model and experiment that they are using.

The p-value is a statistical value to test the value of a model. The p-value tests all the assumptions made in a model and not only the hypothetical assumptions in a model. It should be noted that the p-value does not indicate whether certain assumptions are correct or incorrect. The p-value indicates whether the null hypothesis can be rejected or not. The significance of a model is determined using the p-values. P-values are used to determine whether a model (or coefficients in the model) is significant or not, i.e., whether the null hypothesis can be rejected or not.

A large p-value indicates that the variable might be statistically insignificant. The probability of the null hypothesis to be true increases as the p-value increases. Care should, however, be taken with regards to small and large p-values. The fact that a small or a large p-value was obtained does not necessarily mean that the null hypothesis should be accepted or rejected. Other, erroneous assumptions included in the model might cause p-values to be small or large.

When keeping the above in mind, it is generally assumed that a small p-value, say $p=0.05$, only indicates that the null hypothesis can be rejected for 95% of the time. There is, however, room for other p-values, and even large p-values can indicate significant results (Kim, 2015).

In statistics, a p-value of 0.05 or lower is generally accepted as a good p-value. It should be noted that this is only an indication and not a requirement that the p-value should be smaller than 0.05. This is, however, only a guideline and other values might be accepted based on the data set and the significance level (alpha value). The p-value is directly linked to the alpha value. An alpha value is a value that is chosen beforehand, i.e., before the model is developed and calibrated, while a p-value has to be calculated afterwards (Kim, 2015).

The alpha level chosen for an experiment relates to two different types of errors, namely, Type I and Type II errors. A type I error is an error where it is assumed where the null hypothesis is true, while a Type II error assumes that the alternative hypothesis is true (Kim, 2015, Stephanie Glen, 2012).

The p-value is directly related to the confidence interval chosen for the study. The confidence interval is equal to one minus the alpha value. The p-value should be smaller than the alpha value for the confidence interval to be maintained or approved. If a large alpha value is chosen, it means that the confidence interval of the study will be small.

In some cases when a model is forced to comply with a small p-value, the model can become meaningless, since some of the variables, which have an influence on the study, are either forced to fit a specific pattern or are left out since the confidence level is not high enough. There is therefore not a specific value that the p-value should comply to and the fact that some models comply with a higher p-value does not mean that the model is inaccurate (Greenland *et al.*, 2016, Greenland, 2011). The following reasons to use a larger p-value were found:

- Small sample sizes
- Samples with a low statistical power
- Studies with one-sided hypotheses
- Studies attempting to confirm a null hypothesis

An MNL model requires a certain number of data points per independent variable to be statistically acceptable. For models with a few data points, a larger p-value would probably be accepted. However, if the data set has many data points, a large p-value would most probably not be accepted. It should also be noted that p-tests are not the only statistical significant tests that indicate whether a model is acceptable or not. T-tests, RMSE, and $R^2$ values can also be used to indicate the statistical significance of a model (Greenland *et al.*, 2016).

### b) T-test

T-tests are used to compare mean-values obtained in a statistical model. The t-test can be based on a one-sample or a two-sample t-test. As with the p-values, a t-value is calculated after the model is developed. This test is also used to determine whether the null hypothesis can be rejected or not (Veazie, 2015; Ogee *et al.* 2016).

A t-value is a value that compares the means obtained with the null hypothesis. Large t-values are typically regarded as good indicators of statistical significance. If the t-value is equal to 0, it means that it is also equal to the null hypothesis. T-values, as well as t-distribution tables, are used to determine the confidence interval of the data (Ogee *et al.* 2016).

To determine the significance of a study using the t-test, the degrees of freedom, as well as the t-value, should be available. The degrees of freedom are determined by subtracting one from the number of data points. For example, if the data set contains 9 data points, the degree of freedom is equal to 8. The t-value obtained in the study is looked up in the table in the line corresponding to the degrees of freedom. The majority of the t-distribution tables already provide the researcher with the confidence interval for both one- and two-tailed hypotheses. The confidence level can, therefore, be read directly off the table if the t-value and the degrees of freedom are known, as indicated in Figure 14.

**t-distribution**

| | Confidence Level | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 60% | 70% | 80% | 85% | 90% | 95% | 98% | 99% | 99.8% | 99.9% |
| | Level of Significance | | | | | | | | | |
| 2 Tailed | 0.40 | 0.30 | 0.20 | 0.15 | 0.10 | 0.05 | 0.02 | 0.01 | 0.002 | 0.001 |
| 1 Tailed | 0.20 | 0.15 | 0.10 | 0.075 | 0.05 | 0.025 | 0.01 | 0.005 | 0.001 | 0.0005 |
| df | | | | | | | | | | |
| 1 | 1.376 | 1.963 | 3.133 | 4.195 | 6.320 | 12.69 | 31.81 | 63.67 | — | — |
| 2 | 1.060 | 1.385 | 1.883 | 2.278 | 2.912 | 4.271 | 6.816 | 9.520 | 19.65 | 26.30 |
| 3 | 0.978 | 1.250 | 1.637 | 1.924 | 2.352 | 3.179 | 4.525 | 5.797 | 9.937 | 12.39 |
| 4 | 0.941 | 1.190 | 1.533 | 1.778 | 2.132 | 2.776 | 3.744 | 4.596 | 7.115 | 8.499 |
| 5 | 0.919 | 1.156 | 1.476 | 1.699 | 2.015 | 2.570 | 3.365 | 4.030 | 5.876 | 6.835 |
| 6 | 0.906 | 1.134 | 1.440 | 1.650 | 1.943 | 2.447 | 3.143 | 3.707 | 5.201 | 5.946 |
| 7 | 0.896 | 1.119 | 1.415 | 1.617 | 1.895 | 2.365 | 2.999 | 3.500 | 4.783 | 5.403 |
| 8 | 0.889 | 1.108 | 1.397 | 1.592 | 1.860 | 2.306 | 2.897 | 3.356 | 4.500 | 5.039 |
| 9 | 0.883 | 1.100 | 1.383 | 1.574 | 1.833 | 2.262 | 2.822 | 3.250 | 4.297 | 4.780 |
| 10 | 0.879 | 1.093 | 1.372 | 1.559 | 1.813 | 2.228 | 2.764 | 3.170 | 4.144 | 4.586 |

*Figure 14: Example of calculating the confidence interval using the t-test*

In the example indicated in Figure 14, a data set with 9 data points resulted in 8 degrees of freedom. If it is assumed that a t-value of approximately 1.397 was obtained for a two-tailed hypothesis test. It is clear from the figure that this t-value corresponds to an 80% confidence level. If the confidence level for this example was set at 90%, the null hypothesis could not be rejected, but if the confidence level for this example was set at 80%, the null hypotheses could be rejected. Some t-distribution tables provide the significance level instead of the confidence level associated with the t-value. If the significance level is provided, the confidence level can be calculated as follows:

$$CI = 1 - significance\ level$$

**c) Odds ratio and probabilities**

In statistics, the probability, as well as the odds ratio, is used to predict the outcome of an experiment or model. These factors are very similar; however, there are a few differences between them:

- The odds ratio indicates the constant effect of an attribute on the dependent variable. The probability is not based on the constant effect and can therefore not present the results in one number, but will change as the variables changes.
- Probabilities usually are more intuitive and easier to understand. Probabilities are generally used to explain the effect of a specific attribute on the dependent variable when all the other attributes are kept constant.
- Probabilities are also useful when the independent variable is categorical. The effect of this variable is simple to explain through probabilities.
- When the independent variable is continuous, it is easier to use the odds ratio, since this factor will stay constant while the probability will not stay constant as the independent variable changes.

The odds ratio explains how the outcome will change with a change in the independent variable; for example, the odds of sustaining fatal injuries is two times higher for high speeds than for low speeds. If this ratio is explained in probability terms, it can only be determined that the probability of sustaining fatal injuries at high travel speeds is higher than for low travel speeds (Karen Grace-Martin, 2019).

### d) Root Mean Square Error and $R^2$ value

This value represents the difference between the predicted and the actual data. One of the advantages that this value has above the $R^2$ value is the fact that the RMSE value is the absolute difference, while the $R^2$ value predicts the relative difference between the actual and predicted data. Also, one of the downfalls of the $R^2$ value is that it can only be made more accurate by adding or removing variables. The RMSE value has the same units as the dependent variable and is also calculated using the standard deviation, which makes this value even more attractive to use as the goodness of fit measure, (Karen Grace-Martin, 2019).

The interpretation of the significance of the variables included in the model does not change with a high or a low R squared value (Jim Frost, 2018).

### e) Sample size

The sample size of an experiment plays a significant role in the accuracy of the outcome of the model. The required sample size is dependent on several attributes, such as the number of independent variables, the standard deviation, the $R^2$ value, and the power of the model. There are different methods to determine the sample size. Two of the sampling methods are described below.

**Power test**

The power test of a model relates to the Type II error. This test is applied to determine the sample size that is required to obtain a particular result. The power test can also be used to determine the power of the test after the experiment, and the model has been done. In the case of this study, the power test was not applied to calculate the power of the model, but to calculate the minimum sample size that had to be used to obtain results of a certain confidence interval and power.

One of the inputs required for the power test is the effect size. The effect size can be calculated using the population size and the sample size. In the case of this study, an effect size was assumed, since the data required to determine the effect size is not known. If the total number of pedestrians that walk along and cross the freeway on a daily basis was known, the effect size, i.e., the number of people involved in a pedestrian crash compared to the number of pedestrians that walk along the road, could have been calculated. It is, however, assumed that a small proportion of the pedestrians that cross the freeway on a daily basis were involved in a pedestrian crash and a small effect size was therefore assumed. Ned and Pierre, (2018) determined that a value of 10% can be regarded as a small effect size and an effect size of 10% was therefore used for further calculations as part of this study. The calculations to determine the minimum sample size was done using the G*Power calculation tool (Faul, F. et al., 2007).

**Ten Times rule**

A rule of thumb method that can be used to determine the minimum sample size required to conduct a study is the ‘10 times rule’. This rule states that for every group, a minimum of ten data points for every independent variable should be available to obtain reasonable results. This method has proven to result in inaccurate results since it is only dependent on the number of variables that are included in the study and not on any other statistical value (Ned and Pierre, 2018, Starkweather and Key Moske, 2002, Hair Joseph F et al., 2017).

## 3.6 LIMITATIONS OF THE DATA

This study was not without limitations. The limitations that were found in the data collection process are discussed below.

The exact location of the crash

The crash location that was recorded in the SANRAL data is not necessarily accurate. Pedestrians that were struck at high speeds tend could be displaced meters from the initial point of impact. This is due to the momentum and force of the crash. The force and momentum result in two impact points, one being the impact between the pedestrian and the vehicle and the second impact point the location where the pedestrian hit the ground. Only one impact point was recorded in the data, namely the location where the pedestrian was found after the second impact point (i.e., hitting the ground). This again has to do with the speed at which the vehicle was travelling when it hit the pedestrian. Also, the recording of the crash is done to a certain degree of accuracy. The location which is logged is usually a close estimate of where the crash occurred. This limitation has been dealt with by applying the average traffic of the sub-section as well as by taking the land use in the surrounding area into consideration.

Impact speed

The exact speed at which a pedestrian was struck was not known, and an estimated travel speed at which the pedestrian was struck is not logged in the SANRAL data. The average travel speed was therefore used for the analyses, which was expected to give a reasonably accurate relationship between the impact that speed had on pedestrian injuries and crash rates. The inclusion of the average speed in the model decreased the accuracy of the relationship between the injuries sustained and the travel speed. It was, however, expected that the average speed gave a fairly accurate representation of the speed that the pedestrian had to anticipate when they waited on the side of the road for a gap to cross the road.

The lane in which the crash occurred or in which the pedestrian was found was also not logged in the SANRAL data. The lane in which the pedestrian ended up might be different from the lane in which he/she was struck. Due to this, the speed travelled per lane was not used, since the lane to be used was not known. The speed differential was, therefore, also not taken into account.

<u>Traffic volume</u>

The traffic volume that was used as part of the study was average traffic volumes over twelve months. The CTO stations did not always record 100% of the available recording time per day, which might lead to an inaccurately calculated average. Also, 2013 data was not available, and this data was estimated using the typical data from the 2013 yearbook graphs. Since the data was read of a graph, this could lead to parallax mistakes in the reading of the data. The locations where other data was also not available, e.g. where no data for 2018 was available, the data also had to be estimated. The estimated volumes were compared with the available data, and the estimated values are expected to be in an acceptable range of accuracy.

<u>Trip purpose and land use</u>

The trip purpose and the location where the pedestrians were walking to were unknown. Educated guesses were made to determine the trip purpose by comparing the day and time of the crash with the surrounding land use. This can result in inaccurate land use data that was assigned to the pedestrian crashes since the origin and the destination of the trip were not taken into account. This resulted in the origin, and destination land uses that were used interchangeably. Some of the pedestrian crashes that occurred were therefore assumed to be at the destination location, while it might have happened at the pedestrian's origin. This, therefore, resulted in the wrong land use that was assigned to the pedestrian crash. Another possibility is that the pedestrian was hit somewhere in the middle of his trip, resulting in unrelated land use that was assigned to the crash.

<u>Vehicle Type</u>

The vehicle type that was involved in the pedestrian crash was recorded for 89% of the crashes. The vehicle type for the remaining 11% of the crashes was unknown. The proportional weight between the different vehicle types was calculated and distributed amongst the unknown crashes. There is, therefore, a chance that the vehicle type that was involved in the crash was different from the vehicle type that was assigned to the crash. The portion of the crashes to which a wrong vehicle type could have been assigned to is small, and it was assumed that the impact that this will have on the model, is negligibly small.

# 4. ANALYSIS

Two MNL models were developed to determine what the relationship between the various independent variables is on:

1. The injuries sustained in a pedestrian crash; and
2. The pedestrian crash rate during different times of the day.

For the first model, namely the model explaining the relationship between the injuries sustained and the independent variables, four different injury types were identified, namely:

- No injuries: 42 crashes
- Slight injuries: 19 crashes
- Severe injuries: 38 crashes
- Fatal injuries: 89 crashes

This, therefore, resulted in three models. The "fatal" category was assumed to be the reference category. The three models that were developed are given below:

- *Model A: log(probability of no injury/probability of fatal injury)*
- *Model B: log(probability of slight injury/probability of fatal injury)*
- *Model C: log(probability of severe injury/probability of fatal injury)*

A total number of 188 pedestrian crashes were used to develop these models. It can be seen that the numbers of slight and severe injuries are very low. A reasonable number of data points per category are required to develop a suitable MNL model. Research indicates that the absolute minimum data points for a model should be ten data points per independent variable, known as the ten times rule (Ned and Pierre, 2018, Starkweather and Key Moske, 2002). In this study, a total of seven independent variables are investigated, resulting in a minimum of 70 data points that are required per model. When this is compared to the dataset that is used as part of this study, it can be concluded that the minimum number of data points are not met for Models B and C.

It was therefore decided to combine all the non-fatal injuries in one category in order to develop a binomial (since the dependent variable has only two levels) logistic regression model that will predict the probability of sustaining fatal injuries compared to non-fatal injuries.

The same approach was followed for the crash rate model. The crash rate was first divided into five groups, namely:

- Very low: 0 – 0.25 crashes per million vehicle kilometres travelled, with **140 crashes** associated with very low crash rates;
- Low: 0.25 – 0.50 crashes per million vehicle kilometres travelled, with **34 crashes** associated with low crash rates;
- Medium: 0.50 – 0.75 crashes per million vehicle kilometres travelled, with **8 crashes** associated with medium crash rates;
- High: 0.75 – 1.00 crashes per million vehicle kilometres travelled, with **10 crashes** associated with high crash rates; and
- Very high: 1.00 – 1.25 crashes per million vehicle kilometres travelled, with **14 crashes** associated with very high crash rates.

As already mentioned for the data set used for Model 1, the absolute minimum data points that should be used for the seven independent variables are 70 data points per group. When looking at the number of data points available for the five crash data categories, it is clear that these categories do not comply with the absolute minimum standard. The crash rate was, therefore, divided into three categories to be able to get more data points per group, namely:

- Low: 0.0 - 0.33 crashes per million vehicle kilometres travelled, with 102 crashes associated with low crash rates
- Medium: 0.34 – 0.66 crashes per million vehicle kilometres travelled, with 54 crashes associated with medium crash rates
- High: 0.67 - 1.25 crashes per million vehicle kilometres travelled, with 32 crashes associated with high crash rates

It is noted that the high category covers a bigger range of crash rates; however, this was not expected to be a problem, since crash rates of 0.66 and higher are typically regarded as high crash rates, resulting in the high crash rate category that contains a bigger range of crash rates.

The G*Power analysis tool was used to determine the minimum sample size that is required to determine a statistical power of 90% with a confidence interval of 95% with seven independent variables. The same calculation was run, only with a 50% confidence interval, and the following sample sizes were calculated, as indicated in Table 8.

*Table 8: Required sample size for different confidence intervals*

| No | Power | Effect size | Confidence Interval | Number of variables | Required sample size |
|---|---|---|---|---|---|
| 1 | 90% | 10% | 95% | 7 | 201 |
| 2 | 90% | 10% | 50% | 7 | 102 |

It can, therefore, be concluded that for a confidence interval of 50% a sufficient data set is available, however, if a confidence interval of 95% is aimed for, the data set obtained from the SANRAL crash data have to consist of more data points to improve the confidence of the study.

## 4.1 MODEL 1: INJURIES SUSTAINED

### 4.1.1 Developing of Model 1

The simplified model was developed to predict the probability of sustaining fatal injuries versus non-fatal injuries. This resulted in more data points that could have been used for the regression analysis to determine the relationship between the different variables and injuries sustained. The p-stats and t-stats values before and after omitting the statistically insignificant variables are summarised in Table 9.

*Table 9: P-stats and T-stats values for the Injury Model (Model 1)*

| Independent Variable | P stats value (before omitting data) | P stats value (after omitting data) | T stats value (before omitting data) | T stats value (after omitting data) |
|---|---|---|---|---|
| Constant | 0.2201 | 0.2000 | 1.2262 | 1.2816 |
| Land use | 0.7469 | n/a | 0.3228 | n/a |
| Day of crash | 0.1142 | 0.1292 | -1.5796 | -1.5173 |
| Peak period | 0.6626 | n/a | -0.4363 | n/a |
| Speed | 0.2477 | 0.2684 | -1.1560 | -1.1067 |
| Volume | 0.0330 | 0.0026 | 2.1318 | 3.0064 |
| Vehicle type | 0.3479 | 0.3788 | -0.9388 | -0.8801 |
| Lanes to cross | 0.6199 | 0.4570 | -0.4960 | -0.7439 |

The coefficients of the independent variables that were determined before and after omitting the statistical dependent or insignificant values are summarised in Table 10.

*Table 10: Coefficient values before and after omitting independent variables*

| Independent Variable | Coefficient Value (before omitting data) | Coefficient Value (after omitting data) |
|---|---|---|
| Constant | 2.1888 | 2.0236 |
| Land use | 0.0512 | n/a |
| Day of crash | -0.5548 | -0.5216 |
| Peak period | -0.0842 | n/a |
| Speed | -0.0160 | -0.0150 |
| Volume | 0.0003 | 0.0003 |
| Vehicle type | -0.2095 | -0.1944 |
| Lanes to cross | -0.1125 | -0.1531 |

The model that was developed to explain the relationship between the fatal and non-fatal injuries can, therefore, be expressed as follows:

$$\log\left(\frac{non-fatal}{fatal}\right) = -0.5216X_D - 0.0150X_s + 0.0003X_{TVol} - 0.1944X_{VehT} - 0.1531X_l + 2.0236$$

### 4.1.2 Significance testing of Model 1

The significance level that was chosen for this model is 0.50. This is a rather large value, resulting in a low confidence interval:

$$CI = 1 - \alpha = 1 - 0.5 = 50\%$$
$$with\ CI\ the\ confidence\ interval$$

A 50% confidence interval was assumed to be applicable because of the relatively small number of data points as well as the variability of the travel speed and traffic volume variables which are quite high. The p-stats values that were greater than 0.50 were therefore omitted from the model since these variables are statistically insignificant for this model.

***P-test***

The significance of the attributes included in the model was investigated using the p-stats values obtained after developing the model. The p-stats values are given in Table 9 above. It is clear that if all seven independent variables are included in the study, the p-stats values of some of the variables are large and exceed the significance level that was decided on beforehand. This leads to one of two conclusions:

- The variable is not statistically significant for this model and therefore has to be excluded from the data set; or
- The variable is dependent on another variable also included in the data set and therefore has to be excluded from the data set to prevent the two variables having an influence on one another in the model.

The variables with p-stats values that are larger than the significance level were excluded one by one from the model to eliminate all the variables that are either dependent on other variables or insignificant. The variables that were left with are indicated in Table 9.

As discussed in Section 3.5.3, the p-stats values obtained after developing a model might be large due to another assumption that was made to develop the model. It was therefore decided to use other statistical tests to determine whether the null hypothesis can be rejected or not.

***T-test***

The t-test was performed to determine what the significance of Model 1 is. This was done by determining the t-stats values after developing the model, refer to Table 9 for these values. These values, together with the 187 degrees of freedom were used to find the t-distribution values for a two-tailed hypothesis in the t-distribution table, provided in Appendix A. After obtaining these values, the significance level was determined, as shown in Figure 15.

| | Area in right tail = 0.25 | Area in right tail = 0.20 | Area in right tail = 0.15 | Area in right tail = 0.10 | Area in right tail = 0.05 | Area in right tail = 0.025 | Area in right tail = 0.02 | Area in right tail = 0.01 | Area in right tail = 0.005 | Area in right tail = 0.0025 | Area in right tail = 0.001 | Area in right tail = 0.0005 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| DF | t-score | t-score | t-score | t-score | t-score | t-score | t-score | t-score | t-score | t-score | t-score | t-score |
| 173 | 0.676 | 0.844 | 1.040 | 1.286 | 1.654 | 1.974 | 2.069 | 2.348 | 2.605 | 2.843 | 3.138 | 3.348 |
| 174 | 0.676 | 0.844 | 1.040 | 1.286 | 1.654 | 1.974 | 2.069 | 2.348 | 2.604 | 2.843 | 3.138 | 3.347 |
| 175 | 0.676 | 0.844 | 1.040 | 1.286 | 1.654 | 1.974 | 2.069 | 2.348 | 2.604 | 2.843 | 3.137 | 3.347 |
| 176 | 0.676 | 0.844 | 1.039 | 1.286 | 1.654 | 1.974 | 2.069 | 2.348 | 2.604 | 2.843 | 3.137 | 3.347 |
| 177 | 0.676 | 0.844 | 1.039 | 1.286 | 1.654 | 1.973 | 2.069 | 2.348 | 2.604 | 2.843 | 3.137 | 3.346 |
| 178 | 0.676 | 0.844 | 1.039 | 1.286 | 1.653 | 1.973 | 2.069 | 2.347 | 2.604 | 2.842 | 3.137 | 3.346 |
| 179 | 0.676 | 0.844 | 1.039 | 1.286 | 1.653 | 1.973 | 2.069 | 2.347 | 2.604 | 2.842 | 3.136 | 3.346 |
| 180 | 0.676 | 0.844 | 1.039 | 1.286 | 1.653 | 1.973 | 2.069 | 2.347 | 2.603 | 2.842 | 3.136 | 3.345 |
| 181 | 0.676 | 0.844 | 1.039 | 1.286 | 1.653 | 1.973 | 2.069 | 2.347 | 2.603 | 2.842 | 3.136 | 3.345 |
| 182 | 0.676 | 0.844 | 1.039 | 1.286 | 1.653 | 1.973 | 2.069 | 2.347 | 2.603 | 2.842 | 3.136 | 3.345 |
| 183 | 0.676 | 0.844 | 1.039 | 1.286 | 1.653 | 1.973 | 2.068 | 2.347 | 2.603 | 2.841 | 3.135 | 3.344 |
| 184 | 0.676 | 0.844 | 1.039 | 1.286 | 1.653 | 1.973 | 2.068 | 2.347 | 2.603 | 2.841 | 3.135 | 3.344 |
| 185 | 0.676 | 0.844 | 1.039 | 1.286 | 1.653 | 1.973 | 2.068 | 2.347 | 2.603 | 2.841 | 3.135 | 3.344 |
| 186 | 0.676 | 0.844 | 1.039 | 1.286 | 1.653 | 1.973 | 2.068 | 2.347 | 2.603 | 2.841 | 3.135 | 3.344 |
| 187 | 0.676 | 0.844 | 1.039 | 1.286 | 1.653 | 1.973 | 2.068 | 2.346 | 2.602 | 2.841 | 3.134 | 3.343 |
| 188 | 0.676 | 0.844 | 1.039 | 1.286 | 1.653 | 1.973 | 2.068 | 2.346 | 2.602 | 2.841 | 3.134 | 3.343 |

*Figure 15: T-distribution values for 187 degrees of freedom*

Figure 15 is an extract of the t-distribution table that was used as part of this study. Since this study consists of 188 data points, the degrees of freedom are calculated to be 187. The t-stats value determined for the average traffic volume is 3.0064, which lies between the t-values 2.841

and 3.134 in the t-distribution table. The calculated t-stats value lies approximately in the middle of the two values, resulting in a significance level of approximately 0.0035, as calculated below.

$$\frac{0.001 + 0.0025}{2} = 0.00175$$

$$for\ a\ two - tailed\ hypothesis, the\ t - distribution\ value\ is:$$

$$0.00175 \times 2 = 0.0035$$

This result again in a confidence level of

$$(1 - 0.0035) = 0.996\ \approx\ 100\% - 95\%$$

Similar confidence intervals were determined for the t-test and the p-test, as indicated in Table 11.

*Table 11: Confidence Intervals determined from the p- and t-tests*

| Independent Variable | Confidence Interval of p-test | Confidence Interval of t-test |
|---|---|---|
| Constant | n/a | n/a |
| Land use | n/a | n/a |
| Day of crash | 87.08% | 90% - 80% |
| Peak period | n/a | n/a |
| Speed | 73.16% | 80% - 70% |
| Volume | 99.74% | 100% - 95% |
| Vehicle type | 62.12% | 70% - 60% |
| Lanes to cross | 54.3% | 60% - 50% |

These values indicate that there is not a random error or an assumption in the model that will produce a skew p-stats value since the same confidence level was determined using the t- and p-tests.

***RMSE value***

The RMSE value that was obtained for this model is 0.41. This is regarded as a good value, and the model is therefore expected to have a good fit since this value is less than 1, which was aimed for.

***Odds ratio***

The odds ratio for the variables included in the model was determined as follows:

$$Odds\ ratio\ speed =\ e^{\beta_s}$$

With β the coefficient of the independent variable.

*Table 12: Odds ratios for fatal vs. non-fatal injuries*

| Independent Variable | Odds ratio calculation | Odds ratio |
|---|---|---|
| Day of crash | $OR = e^{\beta_D}$ | 0.59357 |
| Average speed travelled | $OR = e^{\beta_s}$ | 0.985112 |
| Average traffic volume | $OR = e^{\beta_{TVol}}$ | 1.0003 |
| Vehicle type | $OR = e^{\beta_{VehT}}$ | 0.823329 |
| Number of lanes to cross | $OR = e^{\beta_l}$ | 0.858044 |

It is evident from Table 12 that the odds of sustaining non-fatal injuries versus fatal injuries relative to the average traffic volume increases 1.0003 times for every unit increase in X. It is further evident from Table 12 that the odds of being in a non-fatal crash for every increase in X are the smallest for the day on which the crash occurs and the highest for the average traffic volume present on the road at the time of the crash.

***Signs of coefficients***

The equation predicting the probability of sustaining fatal injuries can be seen below.

$$\log\left(\frac{non - fatal}{fatal}\right) = -0.5216X_D - 0.0150X_s + 0.0003X_{TVol} - 0.1944X_{VehT} - 0.1531X_l + 2.0236$$

It is evident when looking at the equation that all the variables have a negative impact on the injuries sustained, i.e., increasing the probability of sustaining fatal injuries when:

i) The crash occurs during a weekend;
ii) The average travel speed increases;
iii) With larger vehicles; and
iv) When the number of lanes to cross increases.

The probability of sustaining fatal injuries decreases as the traffic volume on the road increases.

It is concluded that, even though the p- and the t-tests indicated that the predicted model has a relatively large probability of accepting the null hypothesis, the model can still be deemed accurate for this study. The model is accepted due to the following reasons:

1. The sample size is small and will, therefore, result in lower confidence intervals
2. The results obtained for the confidence interval and the sample size were as predicted by the G*Power analysis tool.

## 4.2 MODEL 2: CRASH RATE

### 4.2.1 Developing of Model 2

The second model that was developed is the model that predicts the probability of being involved in a pedestrian crash based on the seven independent variables already listed in Section 3.5.2.

The calculated crash rates are continuous variables, while one of the requirements of a dependent variable should be that it is not continuous. The crash rate used for the Crash Rate Model (Model 2) was divided into three categories in order to be able to develop more accurate MNL models, as already discussed. This will result in two sub-models predicting the probability to obtain high, medium, or low crash rates on the different subsections of the road. It was assumed that if the crash rate were divided into two categories, the model would not be suitable since this will force the model to apply only two different crash rates, while the variation in the crash rates is, in fact, large on the different sub-sections and during the different times of the day.

The p-values and t-values before and after omitting the statistically insignificant data can be seen in Table 13 for the first sub-model, namely the model predicting the probability to obtain low crash rates in relation to high crash rates. Table 14 gives a summary of the p-values and t-values for the second sub-model, namely the model predicting the probability to obtain medium crash rates in relation to high crash rates.

*Table 13: P-stats and T-stats values determined for the low versus high crash rate model*

| Independent Variable | P-stats value (before omitting data) | P-stats value (after omitting data) | T-stats value (before omitting data) | T-stats value (after omitting data) |
|---|---|---|---|---|
| Constant | 0.4045 | 0.4011 | 0.8335 | 0.8397 |
| Land use | 0.1107 | 0.1143 | 1.5950 | 1.5793 |
| Day of crash | 0.2317 | 0.2101 | -1.1959 | -1.2533 |
| Peak period | 0.1710 | 0.1542 | 1.3690 | 1.4249 |
| Speed | 0.1226 | 0.0985 | -1.5439 | -1.6523 |
| Volume | 0.0091 | 0.0064 | 2.6066 | 2.7284 |
| Vehicle type | 0.8935 | n/a | -0.1339 | n/a |
| Lanes to cross | 0.8207 | n/a | -0.2266 | n/a |

*Table 14: P-stats and T-stats values determined for the medium versus high crash rate model*

| Independent Variable | P-stats value (before omitting data) | P-stats value (after omitting data) | T-stats value (before omitting data) | T-stats value (after omitting data) |
|---|---|---|---|---|
| Constant | 0.1560 | 0.1770 | 1.4188 | 1.3500 |
| Land use | 0.4858 | 0.5070 | 0.6969 | 0.6635 |
| Day of crash | 0.3441 | 0.2837 | -0.9460 | -1.0720 |
| Peak period | 0.1525 | 0.1502 | 1.4307 | 1.4387 |
| Speed | 0.0372 | 0.0224 | -2.0840 | -2.2833 |
| Volume | 0.0280 | 0.0260 | 2.1966 | 2.2256 |
| Vehicle type | 0.6312 | n/a | -0.4800 | n/a |
| Lanes to cross | 0.6239 | n/a | -0.4904 | n/a |

The coefficients that were determined for the two crash rate models predicting the probability of obtaining low and medium crash rates in relation to the high crash rates are provided in Table 15 and Table 16, respectively.

*Table 15: Coefficients determined for the low versus high crash rate model*

| Independent Variable | Coefficient Value (before omitting data) | Coefficient Value (after omitting data) |
|---|---|---|
| Constant | 2.4711 | 2.3620 |
| Land use | 0.3750 | 0.3695 |
| Day of crash | -0.5568 | -0.5768 |
| Peak period | 0.3694 | 0.3387 |
| Speed | -0.0370 | -0.0379 |
| Volume | 0.0005 | 0.0005 |
| Vehicle type | -0.0420 | n/a |
| Lanes to cross | -0.0758 | n/a |

*Table 16: Coefficients determined for the medium versus high crash rate model*

| Independent Variable | Coefficient Value (before omitting data) | Coefficient Value (after omitting data) |
|---|---|---|
| Constant | 4.3700 | 3.9459 |
| Land use | 0.1787 | 0.1696 |
| Day of crash | -0.4876 | -0.5448 |
| Peak period | 0.4304 | 0.3787 |
| Speed | -0.0520 | -0.0542 |
| Volume | 0.0005 | 0.0004 |
| Vehicle type | -0.1679 | n/a |
| Lanes to cross | -0.1753 | n/a |

The models predicting the crash rates (CR) based on the independent variables listed in Table 15 and Table 16 can be seen below.

$$\log\left(\frac{low\ CR}{high\ CR}\right) = 0.3695X_{LU} - 0.5768X_D + 0.3387X_{Ph} - 0.0379X_S + 0.0005X_{TVol} + 2.3620$$

and

$$\log\left(\frac{medium\ CR}{high\ CR}\right) = 0.1696X_{LU} - 0.5448X_D + 0.3787X_{Ph} - 0.0542X_S + 0.0004X_{TVol} + 3.9459$$

### 4.2.2 Significance test of Model 2

The same significant testing methods that were used to test the significance of Model 1 are used to test the significance level of Model 2. The level of significance of this model was also set to be 0.5, resulting in a confidence interval of 50%. This was done due to the same reasons as for Model 1. The limited data points result in difficulty to obtain a model predicting results with high confidence levels.

***P-test***

The significance of the attributes included in the model was investigated using the p-stats values obtained after developing the model. The p-values are given in Table 13 and Table 14 above. It is clear that if all seven independent variables are included in the study, the p-stats values of some of the variables are very large and exceed the significance level that was decided on beforehand. This leads to one of two conclusions:

- The variable is not significant for this model and therefore should be excluded from the data set; or
- The variable is dependent on another variable also included in the data set and, therefore, should be excluded from the data set to prevent the two variables having an influence on one another in the model.

The variables with p-stats values that are larger than the significance level were excluded one by one from the model to eliminate all the variables that are either dependent on other variables or insignificant. The variables that were left with are indicated in Table 13 and Table 14. From the p-stats values that were obtained before any data was omitted, it is clear that the same independent variables were insignificant for both models, which leads to the conclusion that the same independent variables have an influence on the all three different crash rate categories, which was expected.

As discussed in Section 3.5.3, the p-stats values obtained after developing a model might be large due to another assumption that was made to develop the model. It is therefore decided to use other statistical tests to determine whether the null hypothesis can be rejected or not.

***T-test***

The t-test was performed to determine what the significance of Model 1 is. This was done by determining the t-stats values after developing the model, refer to Table 13 and Table 14 for these values. These values, together with the degrees of freedom were used to find the t-distribution values for a two-tailed hypothesis in the t-distribution table, provided in Appendix A. Similar confidence levels were determined in the t-test and the p-test, as indicated in Table 17 and Table 18.

*Table 17: Confidence Intervals determined from the p- and t-tests for the low vs high crash rate model*

| Independent Variable | Confidence Interval of p-test | Confidence Interval of t-test |
|---|---|---|
| Constant | n/a | n/a |
| Land use | 89% | 80%-90% |
| Day of crash | 79& | 70%-80% |
| Peak period | 85% | 80%-90% |
| Speed | 90% | 90% |
| Volume | 99% | 99%-99.5% |
| Vehicle type | n/a | n/a |
| Lanes to cross | n/a | n/a |

*Table 18: Confidence Intervals determined from the p- and t-tests for the medium vs high crash rate model*

| Independent Variable | Confidence Interval of p-test | Confidence Interval of t-test |
|---|---|---|
| Constant | n/a | n/a |
| Land use | 49% | 50%-60% |
| Day of crash | 72% | 70%-80% |
| Peak period | 85% | 80%-90% |
| Speed | 98% | 96%-98% |
| Volume | 97% | 96%-98% |
| Vehicle type | n/a | n/a |
| Lanes to cross | n/a | n/a |

The values summarised in Table 17 and Table 18 indicates that there is not a random error or an assumption in the model that will produce a skew p-stats value since similar confidence levels for the independent variables were determined using the t- and p-tests.

### *RMSE value*

The RMSE value obtained for this model is 0.71, which is regarded as a reasonable fit. The fit of this model is not as good as Model 1. This was; however expected, since the dependent variable of Model 2 consists of three variables and not two variables as in Model 1, resulting in fewer data points per sub-model from which information can be drawn.

### *Odds ratio*

The odds ratios for the variables included in the two sub-models were determined as follows:

$$Odds\ ratio\ speed = e^{\beta_s}$$

With β the coefficient of the independent variable.

*Table 19: Odds ratios for the independent variables for the low versus high crash rate model*

| Independent Variable | Odds ratio calculation | Odds ratio |
|---|---|---|
| Land use | $OR = e^{\beta_{LU}}$ | 1.447011 |
| Day of crash | $OR = e^{\beta_D}$ | 0.561693 |
| The peak in which crash occurred | $OR = e^{\beta_{Ph}}$ | 1.403122 |
| Average speed travelled | $OR = e^{\beta_S}$ | 0.962809 |
| Average traffic volume | $OR = e^{\beta_{TVol}}$ | 1.0005 |

*Table 20: Odds ratios for the independent variables for the medium versus high crash rate model*

| Independent Variable | Odds ratio calculation | Odds ratio |
|---|---|---|
| Land use | $OR = e^{\beta_{LU}}$ | 1.184831 |
| Day of crash | $OR = e^{\beta_D}$ | 0.579958 |
| The peak in which crash occurred | $OR = e^{\beta_{Ph}}$ | 1.460385 |
| Average speed travelled | $OR = e^{\beta_S}$ | 0.947243 |
| Average traffic volume | $OR = e^{\beta_{TVol}}$ | 1.0004 |

It is evident from Table 19 and Table 20 that the land use and the peak period in which the crash occurred have the highest impact on the odds of being in a pedestrian crash or not. The day on which the crash occurred has, just as in the Injury Model, the smallest impact on the risk of being involved in a pedestrian crash.

***Signs of coefficients***

The models predicting the probability of obtaining different crash rates on the sub-sections of the FMS Network is given below.

$$\log\left(\frac{low\ CR}{high\ CR}\right) = 0.3695X_{LU} - 0.5768X_D + 0.3387X_{Ph} - 0.0379X_S + 0.0005X_{TVol} + 2.3620$$

and

$$\log\left(\frac{medium\ CR}{high\ CR}\right) = 0.1696X_{LU} - 0.5448X_D + 0.3787X_{Ph} - 0.0542X_S + 0.0004X_{TVol} + 3.9459$$

It is evident from the two equations above that the land use surrounding the crash, the peak in which the pedestrian crash occurred as well as the average traffic volume on the section of road at the time of the crash had a positive impact on the low and medium crash rates.

The speed travelled as well as the day on which the crash occurred had a negative influence on the crash rate, resulting in a higher risk of being in a pedestrian crash.

## 4.3 REMARKS ON THE SIGNIFICANCE TESTS OF MODEL 1 AND MODEL 2

After developing the two models, it is concluded that the model that predicts the injuries sustained in the pedestrian crash, i.e., Model 1, is the model that fit the data best. The model that predicts the crash rates on the freeway using the seven independent variables did result in an acceptably accurate model. A low significance level, which resulted in high p-stats values, were used for both models. As determined using the power test, a higher confidence level would require a more extensive data set, which is currently unavailable. The results and discussion of the results can be found in the next chapter.

# 5. RESULTS AND DISCUSSION

The results obtained after developing the Injury Model (Model 1) and the Crash Rate Model (Model 2) are shown and discussed in this chapter. The results are discussed and represented in different ways to clarify the importance and impact of every attribute according to the predicted as well as actual data.

## 5.1 COMPARISON BETWEEN MODEL 1 AND MODEL 2

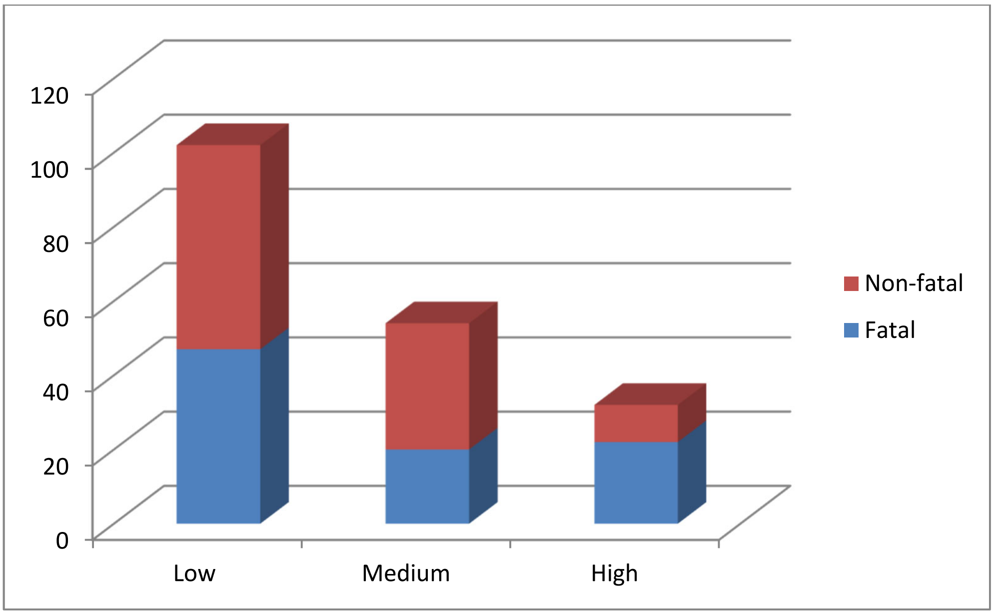
When the two models, i.e., the Injury Model and the Crash Rate Model, are compared with each other, it can be determined that the independent variables that have an impact on the outcome of the models differ. Also, it is clear that Model 1 has a better fit than Model 2, due to the following reasons:

i) The p-values obtained for the injury model is smaller, indicating a larger confidence interval;

ii) The t-values obtained for the injury model is larger, indicating a smaller significance level, which results in a larger confidence interval; and

iii) The root mean square error for the injury model is smaller, which indicate a better fit.

Although, even with a different combination of independent variables, a relationship between the crash rate and the injuries sustained might be made.




*Graph 1: Relationship between fatal and non-fatal injuries for the different crash rate categories*

It is evident from Graph 1 that the number of non-fatal and fatal injuries for the low crash rate sections is approximately the same. The majority of the pedestrian crashes that occurred in the medium crash rate category are non-fatal injuries with the fatal injuries comprising of the majority of the high crash rate category. It seems as if there is an increased risk of sustaining fatal injuries on a high crash rate segment, with a decreased risk of sustaining fatal injuries on the medium and low crash rates segments. This relationship and other similarities are aimed to be determined and confirmed in the remainder of this chapter.

The comparison between the two models is discussed in more detail in the sections below. These sections cover the relationship between the two models to indicate whether the two models can be applied together or whether the models should be used in isolation. The discussion on the relationship between the observed data and the modelled data can be seen in Section 5.2 for the Injury Model and in Section 5.3 for the Crash Rate Model.

### 5.1.1 Land use

The coefficients and the odds ratios related to the land use attribute for the Injury Model (Model 1) and Crash Rate Model (Model 2) can be seen in Table 21.

*Table 21: Coefficients and Odds Ratios associated with the land use attribute*

| Model Description | Land Use | |
|---|---|---|
| | Coefficient | Odds Ratio |
| Model 1 (Injuries sustained) | 0 | n/a |
| Model 2.1 (Low crash rate vs high crash rate) | 0.3695 | 1.447011 |
| Model 2.2 (Medium crash rate vs high crash rate) | 0.1696 | 1.184831 |

It is clear from Table 21 that the land use in the vicinity of the crash does not have a statistically significant impact on the injuries sustained. The land use surrounding the crash, however, influenced the crash rates calculated for the different road segments. The land use had a positive effect on the crash rates, i.e., the odds of being involved in a crash increase with every one unit increase in X. It is evident that the land use has a more significant impact on the low crash rate sections.

### 5.1.2 Day of the week

The coefficients and the odds ratios related to the day of the week on which the pedestrian crash occurred for the Injury Model (Model 1) and Crash Rate Model (Model 2) can be seen in Table 22.

*Table 22: Coefficients and Odds Ratios associated with the day of the week attribute*

| Model Description | Day of the week | |
|---|---|---|
| | Coefficient | Odds Ratio |
| Model 1 (Injuries sustained) | -0.5216 | 0.593570 |
| Model 2.1 (Low crash rate vs high crash rate) | -0.5768 | 0.561693 |
| Model 2.2 (Medium crash rate vs high crash rate) | -0.5448 | 0.579958 |

It is evident from Table 22 that the coefficients associated with the day on which the pedestrian crash happened, are almost the same for the three models. The relationship that the day on which the crash occurred has on the risk of fatal injuries as well as on the risk of being involved in a pedestrian crash is therefore also similar. The relationship was negative in all three cases, i.e., the risk of being involved in a fatal pedestrian crash increases during the weekend. The crash rates also indicate that there is a decrease in the risk of pedestrian crashes that occur during weekdays. When the odds ratio of the Injury Model is taken into account, it can be seen that the odds of sustaining fatal injuries is 0.594 for every one unit increase of X.

The odds of being involved in a pedestrian crash during the weekend increase with approximately 57% with every one unit increase in X.

### 5.1.3 Peak Period

The coefficients and the odds ratios related to the peak period in which the pedestrian crash occurred for the Injury Model (Model 1) and Crash Rate Model (Model 2) can be seen in Table 23.

*Table 23: Coefficients and Odds Ratios associated with the peak period attribute*

| Model Description | Peak period | |
|---|---|---|
| | Coefficient | Odds Ratio |
| Model 1 (Injuries sustained) | 0 | 0 |
| Model 2.1 (Low crash rate vs high crash rate) | 0.3387 | 1.403122 |
| Model 2.2 (Medium crash rate vs high crash rate) | 0.3787 | 1.460385 |

The land use, as well as the peak in which the crash happened, did not have an influence on the injuries sustained. It is expected that the peak hour variable is already taken into account in the traffic volume on the road at different times of the day, as will be further elaborated on in Section 5.2.4. It is, however, evident from Table 23 that the peak in which the crash happened had a positive influence on the crash rates, which was expected.

When looking at the odds ratio for the peak period, it is clear that as the day progress, the odds of being involved in a pedestrian crash, increases. The AM peak hour is, therefore, the peak in which the probability is the lowest to be involved in a pedestrian crash. This is confirmed in the literature (Amoh-Gyimah *et al.*, 2017).

### 5.1.4 Average Travel Speed

The coefficients and the odds ratios related to the average travel speed attribute for the Injury Model (Model 1) and the Crash Rate Model (Model 2) can be seen in Table 24.

*Table 24: Coefficients and Odds Ratios associated with the average travel speed attribute*

| Model Description | Average travel speed | |
|---|---|---|
| | Coefficient | Odds Ratio |
| Model 1 (Injuries sustained) | -0.0150 | 0.985112 |
| Model 2.1 (Low crash rate vs high crash rate) | -0.0379 | 0.962809 |
| Model 2.2 (Medium crash rate vs high crash rate) | -0.0542 | 0.947243 |

When the coefficients of the injuries sustained model, and the crash rate models are compared, it is evident that the average speed travelled had the most significant impact on the injuries sustained model; however, the odds ratio for the models are similar. The odds of sustaining fatal injuries is 0.985 with every unit increase in X. It is therefore evident that the average speed travelled increase the probability of being in a pedestrian crash, as well as to sustain fatal injuries in the crash. The sign of the coefficient for the average speed travelled attribute is negative in the models, which confirm the odds ration that the probability of sustaining non-fatal injuries decreases as the speed increase and that the probability of obtaining low or medium crash rates also decrease.

### 5.1.5 Average Traffic Volume

The coefficients and the odds ratios related to the average traffic volume attribute for the Injury Model (Model 1) and Crash Rate Model (Model 2) can be seen in Table 25.

*Table 25: Coefficients and Odds Ratios associated with the average traffic volume attribute*

| Model Description | Average traffic volume | |
|---|---|---|
| | Coefficient | Odds Ratio |
| Model 1 (Injuries sustained) | 0.0003 | 1.0003 |
| Model 2.1 (Low crash rate vs high crash rate) | 0.0005 | 1.0005 |
| Model 2.2 (Medium crash rate vs high crash rate) | 0.0004 | 1.0004 |

The traffic volume on the road has almost the same effect in the crash rates as on the injuries sustained in pedestrian crashes. The probability of sustaining non-fatal injuries in a pedestrian crash increases as the traffic volume increases. The risk of being involved in a pedestrian crash also increases as the traffic volume increases. In the same way, the odds to be involved in a pedestrian crash and to sustain non-fatal injuries increase as the traffic volume increases.

### 5.1.6 Vehicle Type

The coefficients and the odds ratios related to the type of vehicle involved in the pedestrian crash for the Injury Model (Model 1) and Crash Rate Model (Model 2) can be seen in Table 26.

*Table 26: Coefficients and Odds Ratios associated with the vehicle type attribute*

| Model Description | Vehicle type | |
|---|---|---|
| | Coefficient | Odds Ratio |
| Model 1 (injuries sustained) | -0.1944 | 0.823329 |
| Model 2.1 (Low crash rate vs high crash rate) | 0 | 0 |
| Model 2.2 (Medium crash rate vs high crash rate) | 0 | 0 |

It is evident from Table 26 that the vehicle type that was involved in the pedestrian crash, and, therefore, also the vehicle composition on the road, does not have an impact on the crash rates sustained on the FMS Network.

The vehicle type does have an impact on the severity of the injuries sustained in a pedestrian crash. The coefficient associated with the vehicle type is negative, which means that the probability to sustain non-fatal injuries decrease for the crashes where the vehicle involved is larger than a passenger vehicle (Tefft, 2013). This is in correlation with what was found in the literature review. When the odds ratio is investigated, it is clear that the odds of sustaining non-fatal injuries are 0.82 for every one unit increase X.

### 5.1.7 Lanes to Cross

The coefficients and the odds ratios related to the number of lanes to cross attribute for the Injury Model (Model 1) and the Crash Rate Model (Model 2) can be seen in Table 27.

*Table 27: Coefficients and Odds Ratios associated with the number of lanes to cross attribute*

| Model Description | Lanes to cross | |
|---|---|---|
| | Coefficient | Odds Ratio |
| Model 1 (Injuries sustained) | -0.1531 | 0.858044 |
| Model 2.1 (Low crash rate vs high crash rate) | 0 | 0 |
| Model 2.2 (Medium crash rate vs high crash rate) | 0 | 0 |

As with the vehicle type attribute, the number of lanes to cross does not have an impact on the crash rates of the different sub-sections.

The number of lanes to cross, however, did have an impact on the injuries sustained in the pedestrian crash. The coefficient associated with the number of lanes to cross is again negative.

This indicates that the probability of sustaining fatal injuries increases as the number of lanes to cross increases.

### 5.1.8 Remarks on the Comparison between Model 1 and Model 2

It was determined that the odds ratios of the average traffic volume and average travel speed attributes are similar. This result in the odds of being involved in a fatal pedestrian crash as well as the odds of being involved in a pedestrian crash being approximately the same. These two attributes were the only similarities that were found when the two models were compared. Since a relationship between the increased risk of sustaining fatal injuries in a pedestrian crash that happened on a high crash rate segment could not be confirmed, as suggested in Graph 1, it is concluded that the Injury and Crash Rate Models cannot be used interchangeably.

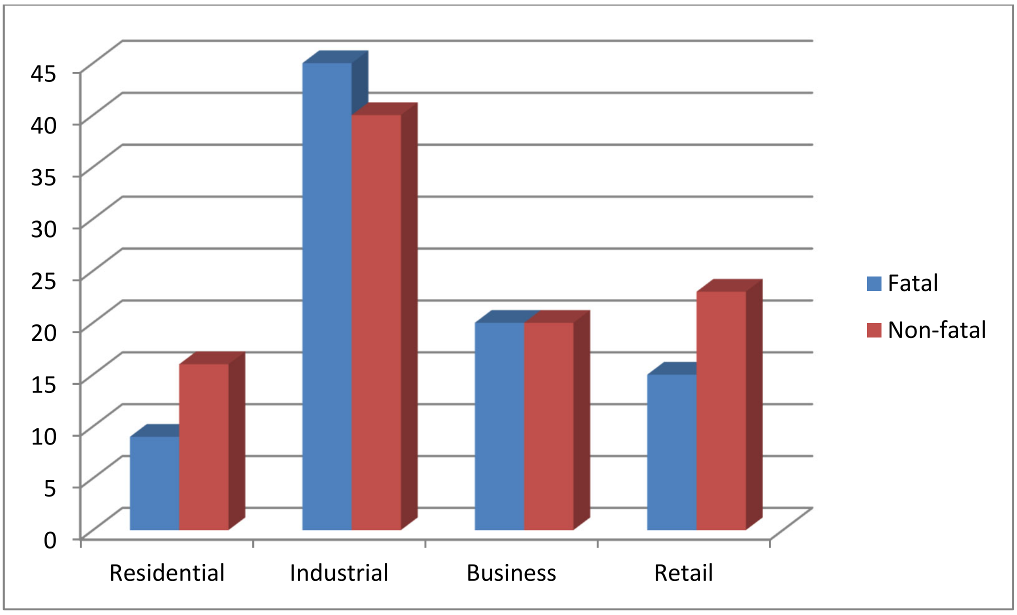
When the above is taken into account, it can be concluded that Model 2 should preferably be simplified by only taking into account whether there is an increased risk of being involved in a pedestrian crash or not. This approach will be followed in the remaining part of the study.

## 5.2 THE INJURY MODEL VERSUS THE ACTUAL DATA

The observed data (i.e., the SANRAL crash data) that was used to develop the Injury Model were analysed separately from the binomial logistic regression model. Graphs were plotted to see what the distribution between the fatal and non-fatal injuries are when it is compared to the Injury Model. The results of the comparison are summarised below.

### 5.2.1 Land use

The number of fatal and non-fatal injuries sustained in the pedestrian crashes for different land uses are summarised in Graph 2.




*Graph 2: Number of crashes for the different land use zonings in the study area*

This variable was found to be statistically insignificant in the Injury Model, as already mentioned in Section 5.1.1. It is clear from Graph 2 that the land use that poses the highest risk to sustain fatal injuries is industrial land use. Contradictory results on the relationship between land use and pedestrian crashes were found in the literature. Some literature indicated that the land use, especially commercial and industrial land use, has an impact on the injuries sustained in pedestrian crashes Lin *et al.*, (2019), Osama and Sayed, (2017), while other research indicated that there is not a clear relationship between the type of land use and the pedestrian injuries (Clifton, Burnier and Akar, 2009). The results of this study suggest that there is not a significant difference between the retail, business, and residential land uses, however, it can be concluded that the industrial land use has a negative effect on the pedestrian crashes.

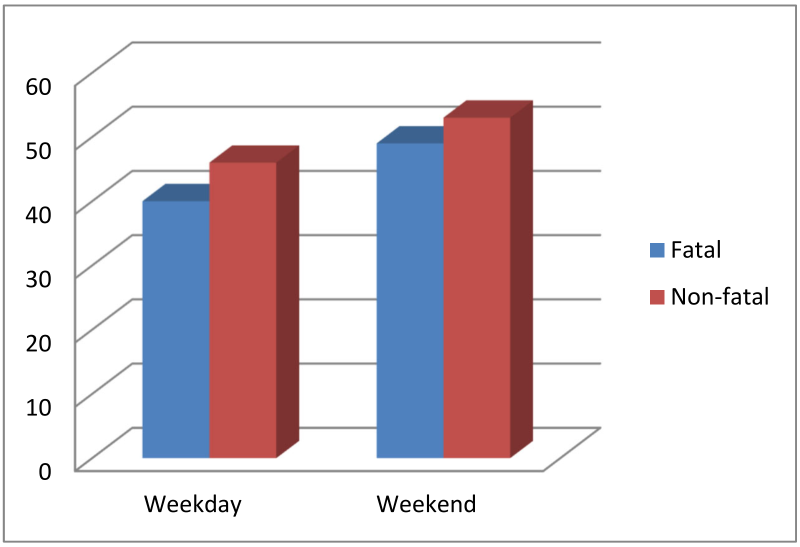
Some of the industrial developments along the freeway operate on a continuous basis. Shifts are introduced at these developments to ensure that there are staff members at these developments during all hours of the day and during all days of the week. These shifts were, however, not considered as part of this study. This was done due to the fact that a low correlation that was found between the land use and the pedestrian crashes.

A reason why the land use does not have a significant impact on the injuries sustained in pedestrian crashes might be because the origin and destination patterns were not taken into account, leading to inaccurate land uses being assigned to crashes. Another inaccurate assumption that could have been made was that the land use that was assigned to the crash was just the land use that the pedestrian had to pass to get to the destination where he wanted to be. The land use that was assigned to the crash was therefore assumed to be the origin or destination land use, while it only might have been a type of land use that had to be passed for the pedestrian to get to his destination.

The majority of the study area is surrounded by industrial land use. It was therefore expected that the majority of the crashes would also have happened at these locations. This, therefore, only indicates that industrial land use has an influence on the number of crashes but does not have any relationship with the injuries sustained. This assumption might still be inaccurate, since the ratio between the industrial land use and the other land uses are out of proportion and might, therefore, result in a biased result.

### 5.2.2 Day of the week

A summary of the pedestrian crashes that occurred during weekdays and weekends can be seen in Graph 3.



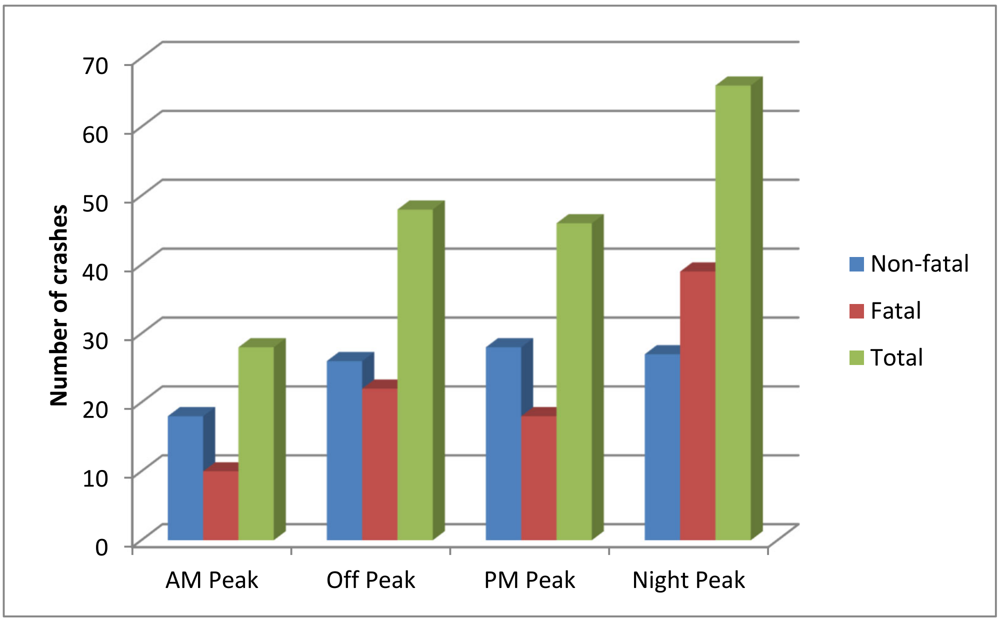

*Graph 3: Number of crashes per day of the week*

It is clear from Graph 3 that the majority of the crashes (both fatal and non-fatal) occur during weekends, as also determined in the Injury Model. There is, however, not a significant difference between the number of crashes that occur during the weekday versus the crashes that occur during the weekend. It should be taken into account that the number of crashes that occur on weekdays are distributed amongst five days (excluding Friday evening and night) while the crashes that occur on weekends are distributed amongst just more than two days (Saturday, Sunday and Friday evening and night). When this is taken into account, it is clear that the crash rate for weekends is higher than for weekdays. A total of 86 (47%) of the fatal crashes occurred during the weekday, with 102 (53%) that occurred during the weekends, resulting in an average of 19.11 crashes per weekday and an average of 40.8 crashes per day during the weekend.

The observations made when comparing the actual data with the Injury model is in line with the literature review, where it was determined that the risk of sustaining fatal injuries in a pedestrian crash increases during the weekend, (Amoh-Gyimah *et al.*, 2017).

### 5.2.3 Peak Period

The distribution of the pedestrian crashes during the peak periods of the day can be seen in Graph 4.




*Graph 4: Number of crashes per peak period*

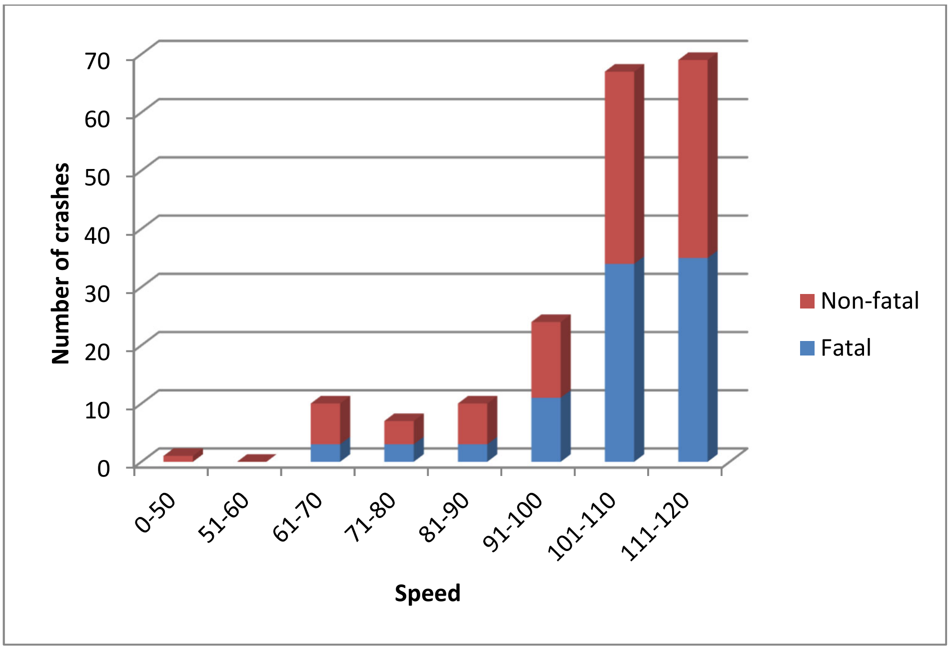
It was determined in the Injury Model that the peak in which the pedestrian crashes occur, does not have a significant impact on the injuries sustained in pedestrian crashes. It is evident from Graph 4 that the majority of the pedestrian crashes occur during the night peak period. A total of 66 pedestrian crashes occurred during the night peak. This amounts to 35% of all the pedestrian crashes. This is in conjunction with the research, where it is proven that the risk of being in a fatal pedestrian crash is higher in the night peak than in the other peaks. The pattern indicated in Graph 4 is consistent with the study conducted by Amoh-Gyimah *et al.*, (2017), who determined that there is a high number of fatal pedestrian crashes during the night and off peak periods, and a lower number of fatal pedestrian crashes during the AM and PM peak.

The number of pedestrian crashes is significantly lower in the AM peak than in the other peaks. The number of crashes that happened in the PM peak period is similar to the number of crashes that occurred in the off peak period; however, the severity of the injuries sustained in the crashes that occurred in the PM peak period is typically non-fatal injuries. This is in line with the research, where it was determined that the majority of the fatal injuries occur outside the peak periods. Also, the increase in the number of crashes in the PM peak periods can be explained due to the fact that the drivers tend to be more stressed after work. Fatigue and loss of concentration are also rather experienced in the PM peak than in the AM peak, (Kononov *et al.*, 2013).

According to Huang, Sun, and Zhang (2018), the risk of sustaining fatal injuries are higher in free-flow conditions than in congested conditions. Higher travel speeds and lower traffic volumes are associated with free-flow conditions. The off and night peak periods are typically the time of day in which free-flow conditions are expected. The latter variables were found to have a significant impact on the injuries sustained in the pedestrian crashes, and it was therefore concluded that the peak period is collinear with these variables and that it should, therefore, be excluded from the model.

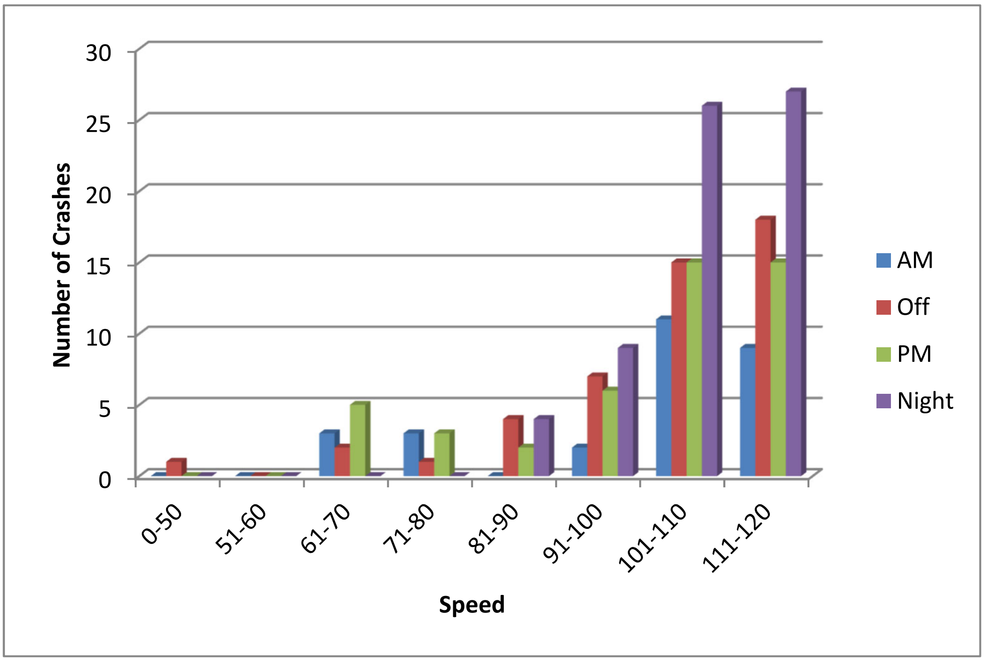
### 5.2.4 Average Travel Speed

The average travel speed that was recorded on the road was determined to have an influence on the type of injuries sustained in pedestrian crashes. The higher the speed travelled, the higher the probability that the pedestrian will sustain fatal injuries. This trend is also confirmed in the literature (Kim *et al.*, 2010; Moore *et al.*, 2011). The injury types sustained at different speed categories are indicated in Graph 5.




*Graph 5: Number of crashes for different speed categories*

The number of pedestrian crashes that occurred during the different times of day for the different speed categories can be seen in Graph 6.




*Graph 6: Number of crashes per peak period for different speed categories*

From Graph 5 and Graph 6, it is evident that the majority of the crashes occur at higher travel speeds. It is also clear from these graphs that the majority of the high-speed crashes occur during the night and off peak periods with the majority of the AM and PM crashes occurring during low speed (or congested speed). This proves that the assumption made that the peak period is collinear with the average travel speed is correct and that the peak period in which the pedestrian crash happened is indirectly included in the average travel speed variable.

The speed at which a vehicle travels has an impact on the sight distance required to come to a complete stop. The higher the travel speed, the longer distance is required to come to a complete stop. This relationship is explained by the equations of motion.

*Table 28: Equations of movement*

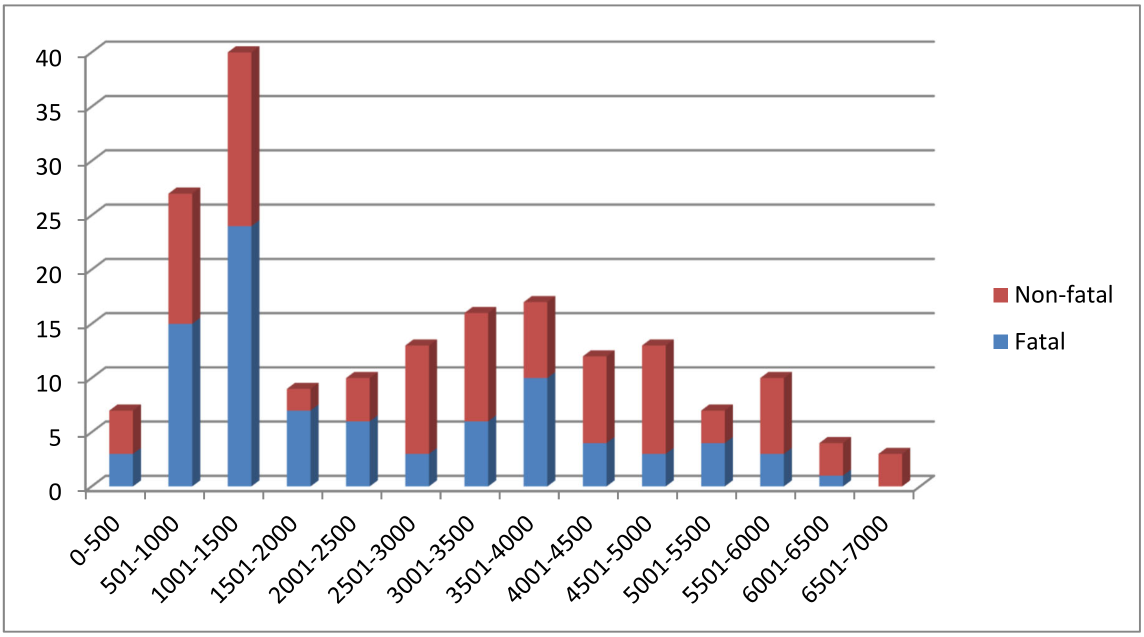
| Equation Name | Equation |
|---|---|
| Speed (v) | $v = v_0 + at$ |
| Distance (s) | $s = s_0 + v_0 t + \frac{1}{2}at^2$ |
| Acceleration (a) | $a = \frac{\Delta v}{\Delta t}$ |

The research found that the typical deceleration rate ($a$) for a vehicle is between 3m/s$^2$ and 4.5m/s$^2$ (Van As et al., 2002). If a pedestrian suddenly enters the roadway, the driver of the vehicle has to decelerate to avoid a crash with the pedestrian; however, since a vehicle has a limited deceleration rate, more time would be required to come to a complete stop for higher travel speeds than for low travel speeds. When this is taken into account, it is clear that the risk of being involved in a fatal pedestrian crash increases as the speed increases. Also, high speeds, or free-flow speeds, are often associated with night time driving, since the road is at free-flow condition. During night time or low light conditions, the sight distance is limited, resulting in a shorter available distance to stop, which equate to less reaction time. This also indicates that if high speeds are travelled in low light conditions, the probability to be involved in a fatal pedestrian crash increase.

A similar number of fatal and non-fatal crashes occurred at high speeds. This was not expected, and this observation is not confirmed by the literature. A reason why this trend has occurred might be because of the street lights that are provided on the entire FMS Network provided sufficient sight distance. This, therefore, results in sufficient reaction time for the vehicle that was involved in the pedestrian crash, to decelerate enough to hit the pedestrian at a low speed, resulting in non-fatal injuries during the night time crashes (which are associated with high-speed crashes). The literature indicated that street lighting, which improves sight distance, resulted in fewer fatal pedestrian injuries (Bianco, 2017; Griffith, 1994).

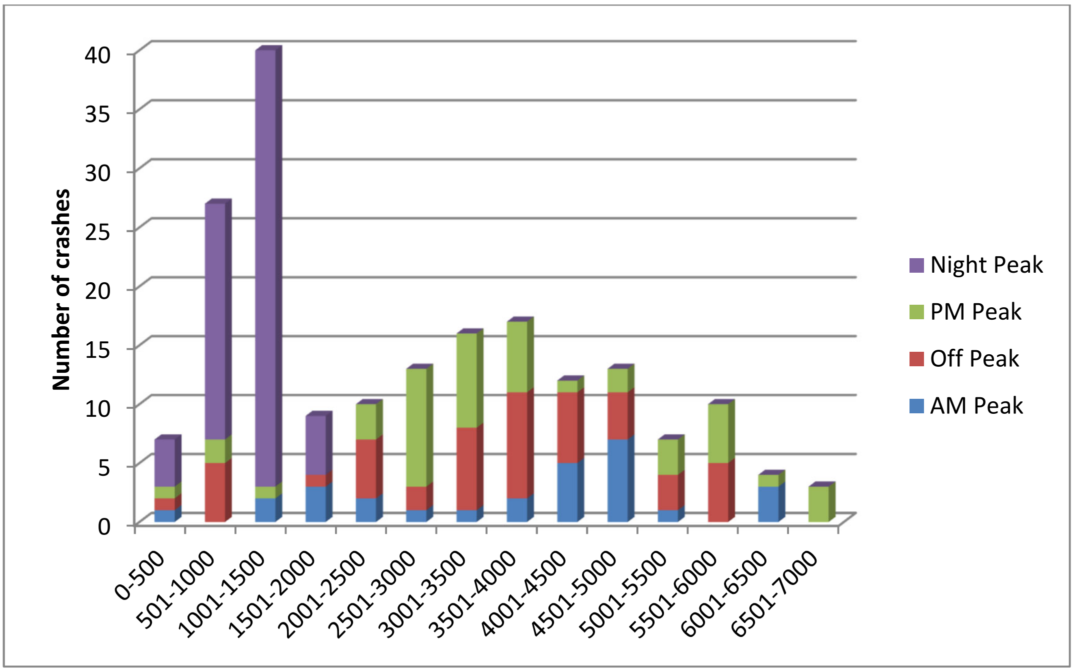
### 5.2.5 Average Traffic Volume

The average traffic volume is discussed in this section. It was determined in the Injury model that this attribute has a positive impact on the injuries sustained, resulting in a higher probability to sustain non-fatal injuries as the traffic volume on the road increases. The number of fatal and non-fatal injuries for the different traffic volume categories is indicated in Graph 7.




*Graph 7: Number of crashes for different traffic volume categories*

It is evident from Graph 7 that the majority of the pedestrian crashes occurred during the times of the day when the traffic volume on the road was low, which will typically be during the night peak periods. The non-fatal crashes are distributed more evenly across the different traffic volume categories. It is interesting to note that only a few fatal crashes occurred during the time of day when high traffic volumes were recorded on the road. These periods are typically the AM and PM peak periods. The reason for this is that the speed at which the vehicles travelled at the high volumes are low, resulting in sufficient time to react when a pedestrian cross the road. The lower speeds also result in lower impact force, which leads to less severe injuries. The peak hours in which the different traffic volumes were recorded can be seen in Graph 8.




*Graph 8: Number of crashes per peak period for different traffic volume categories*

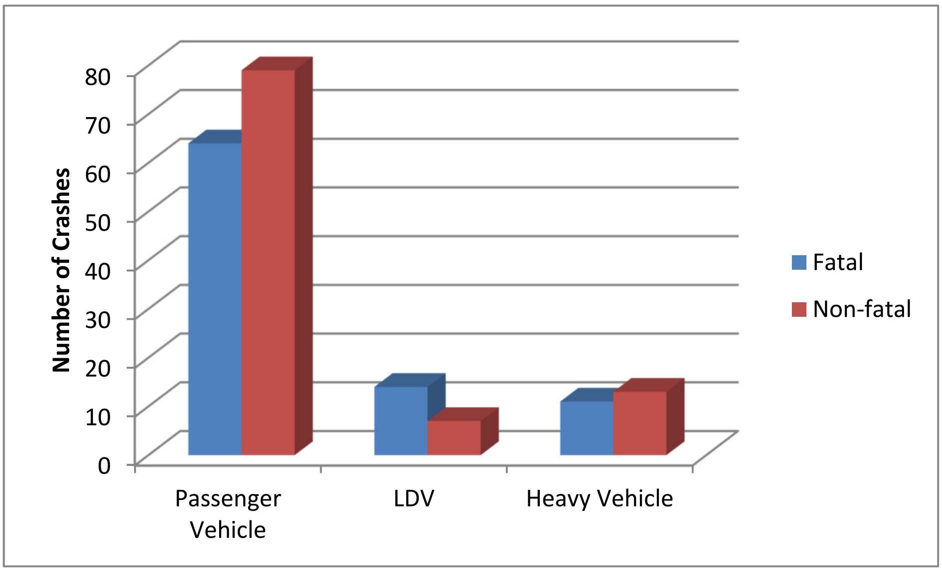
It is evident that the lowest traffic volumes were recorded during the night time peak, and it is therefore concluded from Graph 7 and Graph 8 that the majority of the pedestrian crashes occur during the night peak. The PM peak hour is associated with traffic volumes between 2500 and 4000 vehicles per hour. When this is compared with Graph 7, it can be assumed that the majority of the pedestrian crashes that happen during the PM peak are non-fatal crashes. This trend can also be seen in Graph 4.

It is interesting to note that the variation in the traffic volume in the AM peak period is significantly lower than in the other peak periods. It is clear from Graph 8 that traffic volumes between 4000 and 5000 vehicles per hour (vph) are associated with the AM peak hour. From Graph 7, it is evident that the majority of the pedestrian crashes that occur when the traffic volume on the road is between 4000vph and 5000vph is non-fatal crashes. This concurs with the literature where it is determined that the risk of sustaining fatal injuries at high traffic volumes decreases, but increases for free-flow conditions, (Parmet, Lynm and Glass, 2002b). When the traffic volume is considered, it is clear that the peak periods in which the crashes occurred are indirectly also included in the Injury Model. This is, therefore, consistent with the assumption that the peak period is collinear with the traffic volume.

It evident from Graph 7 that the number of pedestrian crashes, as well as the risk of sustaining fatal injuries, decreases rapidly for the very high traffic volumes (>6000vph). The average travel speed on the road for these types of crashes are typically approaching breakdown speed. The pedestrians can walk through the traffic stream when these traffic volumes and travel speeds are observed on the road. This, therefore, results in a decrease in pedestrian crashes, since the vehicles can stop in time when a pedestrian starts to cross the road. In the case where the pedestrian is hit, the impact of the crash will be low due to the low travel speeds, resulting in non-fatal injuries. This assumption is confirmed in Graph 6, where it is clear that the number of crashes, as well as the injuries sustained at low speeds, decreases. This, therefore, results in the conclusion that the attributes that are included in the Injury Model cannot necessarily be observed in isolation. A dynamic relationship exists between the attributes, where the combination of the attributes has a different outcome than what would be expected if the attributes are investigated in isolation.

### 5.2.6 Vehicle Type

The number of fatal and non-fatal crashes that occurred between the three different types of vehicles included in this study can be seen below. In the Injury Model, it was determined that the vehicle type had an impact on the type of injuries sustained, with the crashes involving heavy vehicles contributing to the highest probability to sustain fatal injuries.




*Graph 9: Number of crashes per vehicle type*

It can be concluded from Graph 9 that the majority of the pedestrian crashes that happened on the Gauteng FMS Network was between pedestrians and passenger vehicles. The number of fatal injuries associated with passenger vehicles is lower than for non-fatal injuries, however, if these vehicles are travelling at high speeds, the impact sustained in the crash might be the same as between a heavy vehicle that was travelling at a low speed. This phenomenon is explained by Newton's Second Law of Motion as well as by the law of momentum and is a reason why the number of fatal injuries in the pedestrian crashes is high.

The speed at which a vehicle travelled has an influence on in the distance required to stop. When looking at the movement equations, it is evident that the mass of the vehicles does not have an influence on the time or distance required to come to a complete stop. It is, however, a well-known fact that heavy vehicles require more time and distance to come to a complete stop than passenger vehicles. The friction that the different vehicles have to sustain is also negligibly small to make a significant impact on the acceleration or deceleration of a vehicle. There are, however, other laws of physics which explains why heavy vehicles require a longer distance to stop, or in other words, a longer reaction time, than passenger vehicles. These laws are explained by Newton's Second Law of Motion as well as the momentum law. Newton argued that the force that any object experience is equal to:

$$F = ma \ \ (Newton\ II)$$
$$with\ m\ the\ mass\ of\ the\ object$$
$$and\ a\ the\ acceleration\ of\ the\ object$$

It is clear from the equation above that the force that an object experience and the force at which a vehicle hit a pedestrian is directly related to the product of its mass and acceleration. The momentum of a vehicle that travels on the road can be calculated using the equation below:

$$p = mv$$
$$with\ m\ the\ mass\ of\ the\ object$$
$$and\ v\ the\ speed\ of\ the\ object$$

It is therefore evident that heavy vehicles have more momentum and force and require more time to come to a complete stop, refer to Table 28. The reaction time available when pedestrians suddenly enter the roadway is not long enough to enable the driver of the heavy vehicle to come to a complete stop. Due to the weight of the vehicle, the impact force is also much higher, which increase the probability to sustain fatal injuries, as determined in Model 1.

The distribution between the fatal and non-fatal injuries between the three vehicle classes are provided in Table 29.

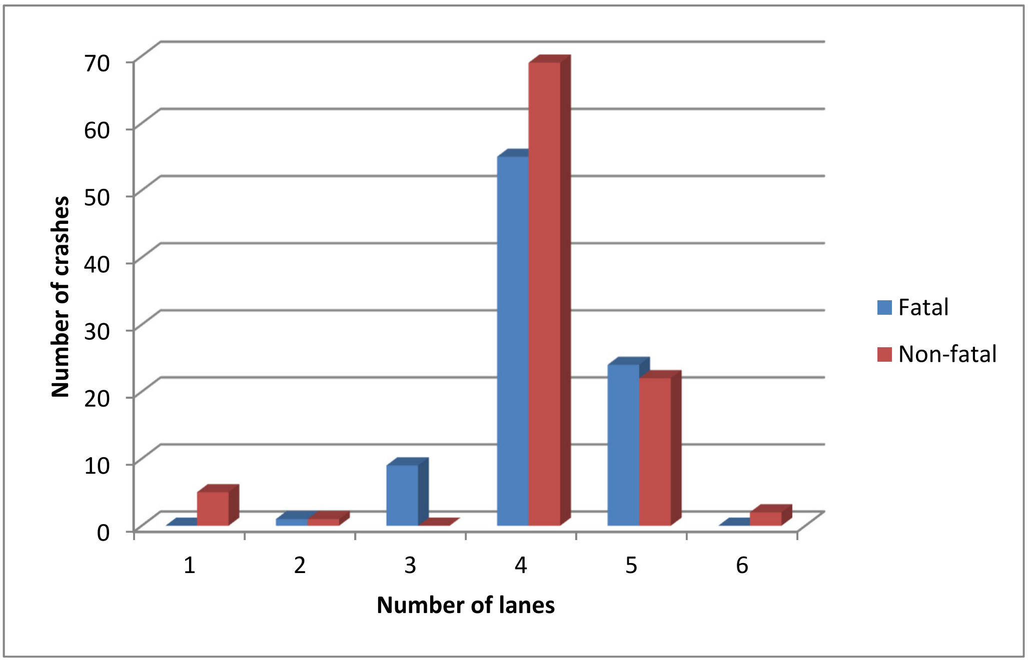
*Table 29: Proportion split between fatal and non-fatal crashes*

| Vehicle Type | % Fatal | % Non-fatal |
|---|---|---|
| Passenger Vehicle | 45% | 55% |
| LDV | 67% | 33% |
| Heavy Vehicle | 46% | 54% |

It is interesting to note that the majority of the pedestrians that are hit by light delivery vehicles sustained fatal injuries, while this split was not determined for the heavy vehicles. In the literature, it was found that the probability of sustaining fatal injuries increases as the vehicle size and weight increases (Kashani and Besharati, 2017; Tefft, 2013). This phenomenon was also found in the Injury Model, where the coefficient indicated that the probability of sustaining fatal injuries increases as the vehicle size and weight increases.

### 5.2.7 Lanes to cross

The number of fatal and non-fatal pedestrian crashes that happened on the different cross-sections of the FMS Network included in the study can be seen in Graph 10.




*Graph 10: Relationship between the number of crashes and the number of lanes per direction*

According to the Injury Model, the number of lanes to cross had a negative impact on the injuries sustained, where the higher number of lanes resulted in a higher probability to sustain fatal injuries. Graph 10 indicates that the majority of the pedestrian crashes occur on the sections of road with four lanes.

It should be noted that the majority of the freeway that was included in the study consisted of four lanes per direction. This result was, therefore, expected. The crashes that happened at the section of road with one or two lanes were all crashes that occurred on the ramps of the interchanges. It can be seen from Graph 10 that the number of fatal injuries increased as the number of lanes increased. This agrees with the Injury Model, where the probability of sustaining fatal injuries increases as the number of lanes increase. When the number of lanes on a road increases, the pedestrians need a longer gap to be able to cross the road. The sight distance might also decrease for the pedestrian when the number of lanes on a road increases, since the vehicles in the far lane might be in the line of sight of other vehicles in the lanes closer to the pedestrian. This, therefore, makes it difficult to assess the gap that is available between the vehicles in the different lanes. The ability to assess the speed at which vehicles are travelling in the far lane also diminishes when the number of lanes increases, which makes it challenging to know when a gap is large enough to be able to cross the road or not. This is in line with what was found in the literature (Zhang, Chen, and Wei, 2019).

### 5.2.8 Remarks on the Injury Model

From the Injury Model, it is clear that the land use surrounding the crash site as well as the peak hour in which the crash occurred, were not statistically significant for the Injury Model. Also, when comparing the coefficients of the model, it is clear that some of the variables have a more significant impact on the risk of being fatally injured than others.

*Table 30: Comparison of the weight of each variable for the Injury Model*

| Independent Variable | Coefficient Value | Odds Ratio |
|---|---|---|
| Constant | 2.0236 | n/a |
| Day of crash | -0.5216 | 0.59357 |
| Speed | -0.0150 | 0.985112 |
| Volume | 0.0003 | 1.0003 |
| Vehicle type | -0.1944 | 0.823329 |
| Lanes to cross | -0.1531 | 0.858044 |

It is evident from Table 30 that the average travel speed, as well as the average traffic volume on the road, has the most significant impact on the probability to sustain fatal injuries in a pedestrian crash. It can be seen that the coefficient representing the number of lanes to cross is relatively small. This variable, therefore, does not have a major impact on the probability to sustain fatal or non-fatal injuries for the cases where the number of lanes to cross is low. As the number of lanes starts to increase, this coefficient, however, start to make a more significant impact on the injuries sustained. The same was observed for the vehicle type involved in the crash attribute. This is proof that the odds ratio will, in some cases be easier to interpret, since the odds ratio give the constant increase, while probabilities change as the attribute changes. It is therefore clear that the odds of sustaining fatal injuries as the number of lanes increase, is equal to 0.858 for every increase in X (assuming the other variables stay constant). The day on which the pedestrian crash happened has a significant influence on the probability to sustain fatal injuries. When the odds ratio of this attribute is taken into account, it is clear that the odds to sustain fatal injuries when the crash happens during the weekend are 59%, if all the other variables are kept constant.

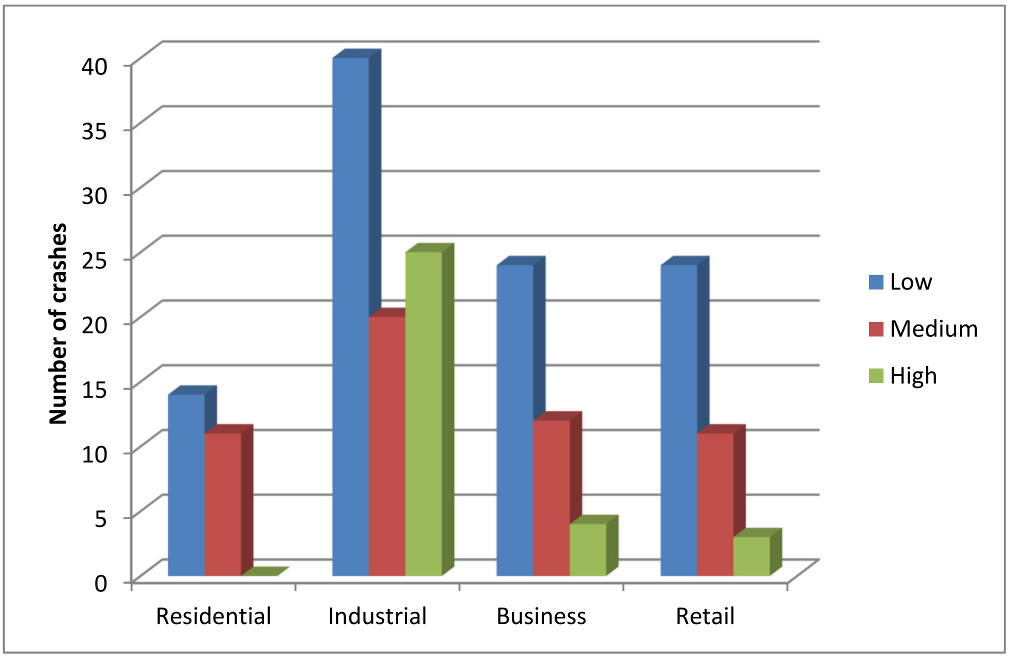
The variables included in the Injury Model cannot be viewed in isolation. The variables change constantly in the field, resulting in an interactive model where the probabilities of sustaining fatal or non-fatal injuries are determined using a combination of factors. It is, however, concluded that the risk of fatal injuries is the highest for high travel speeds and low vehicle volumes on the road.

## 5.3 THE CRASH RATE MODEL VERSUS THE ACTUAL DATA

The MNL models that predict the probability to be involved in a pedestrian crash are discussed in this section of the study. These models were developed to predict the probability of being involved in a pedestrian crash for different crash rate categories. The model was determined to have a relatively low fit; however, this model can still be used to predict the risk of being in a pedestrian crash or not. As with the Injury Model, the actual data is again compared with the predicted data.

### 5.3.1 Land use

The number of pedestrian crashes categorised into the different crash rates versus the land use in the vicinity of the pedestrian crash is indicated in Graph 11. This attribute was found to be statistically significant in the Crash Rate Model.

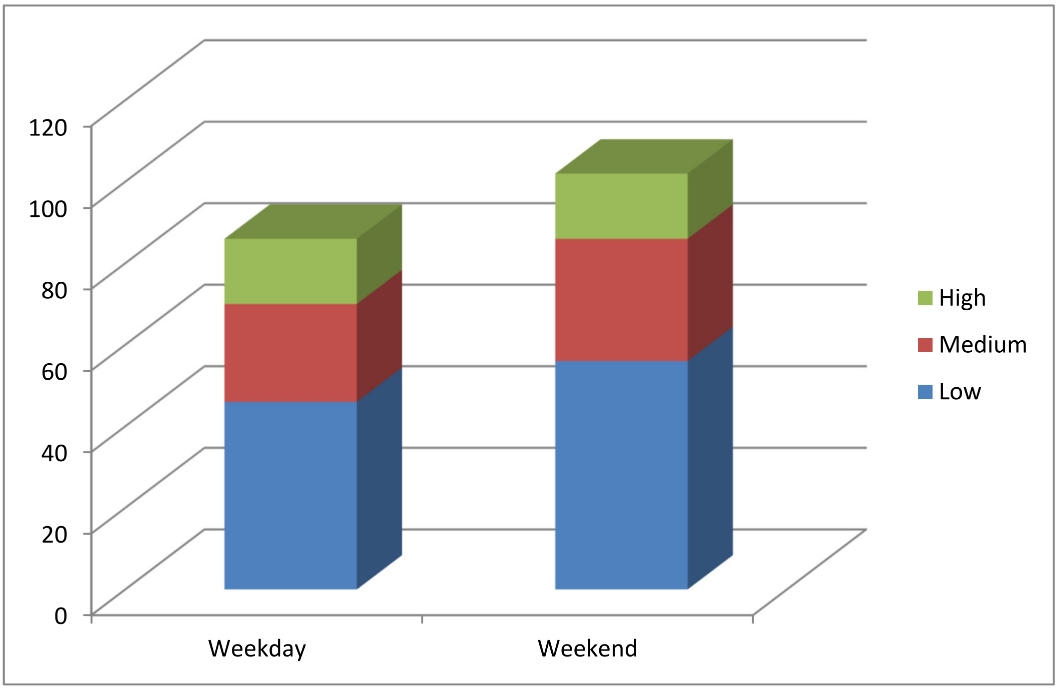



*Graph 11: Number of pedestrian crashes per land use type*

It can be seen from Graph 11 that the majority of the pedestrian crashes occurred at the non-residential land uses. There is a significant increase in the number of pedestrian crashes at the locations where the industrial land use was nearby the pedestrian crash. A total of 85 out of the 188 pedestrian crashes occurred close to industrial land uses. The majority of the crashes that occurred on the high crash rate sections were located at the industrial land uses. It is, therefore, evident that the risk of being involved in a pedestrian crash increase for industrial land uses. This conclusion was confirmed in the Crash Rate Model, where it was determined that the risk to be involved in a pedestrian crash increase for industrial and commercial land uses. This trend was also confirmed in the literature review (Graham, Glaister and Anderson, 2005).

### 5.3.2 Day of the Week

The number of pedestrian crashes that occurred on weekdays and during weekends for the different crash rate categories can be seen in Graph 12.

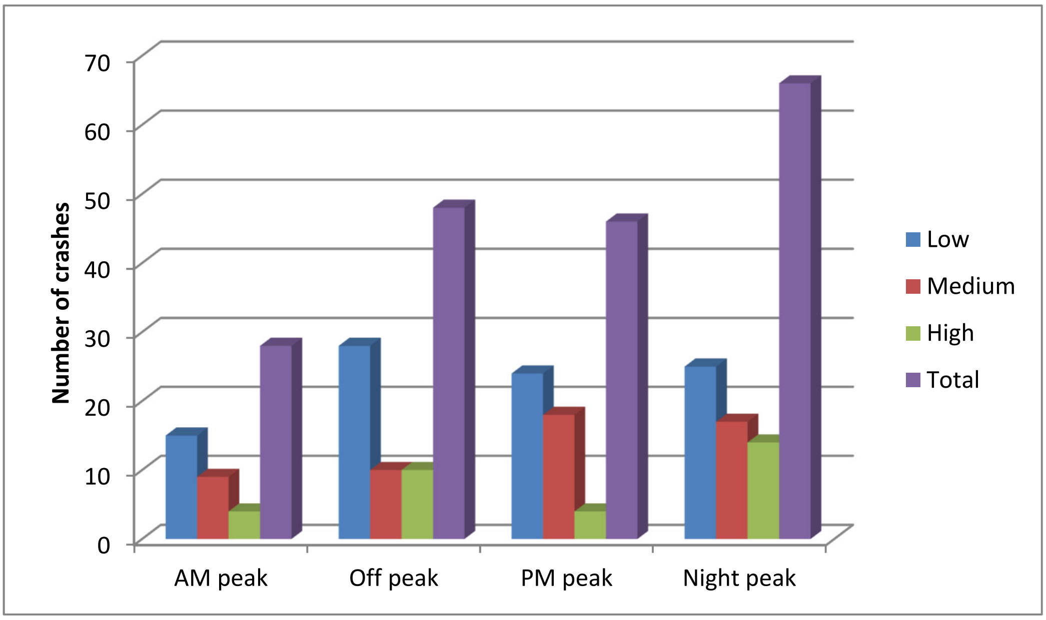



*Graph 12: Number of pedestrian crashes per crash rate per day of the week*

It is clear from Graph 12 that the majority of the crashes occurred during the weekend. This trend can also be observed from the Crash Rate Model, in which it was determined that the probability of being in a pedestrian crash increases during the weekend. This also confirms what was determined in the literature review, where it was determined that the use of alcohol and the risk of being involved in a pedestrian crash due to alcohol increased during the weekends, (Amoh-Gyimah *et al.*, 2017, RTMC, 2017).

### 5.3.3 Peak Period

The number of pedestrian crashes that occurred per time of day for the three different crash rate categories can be seen in Graph 13.

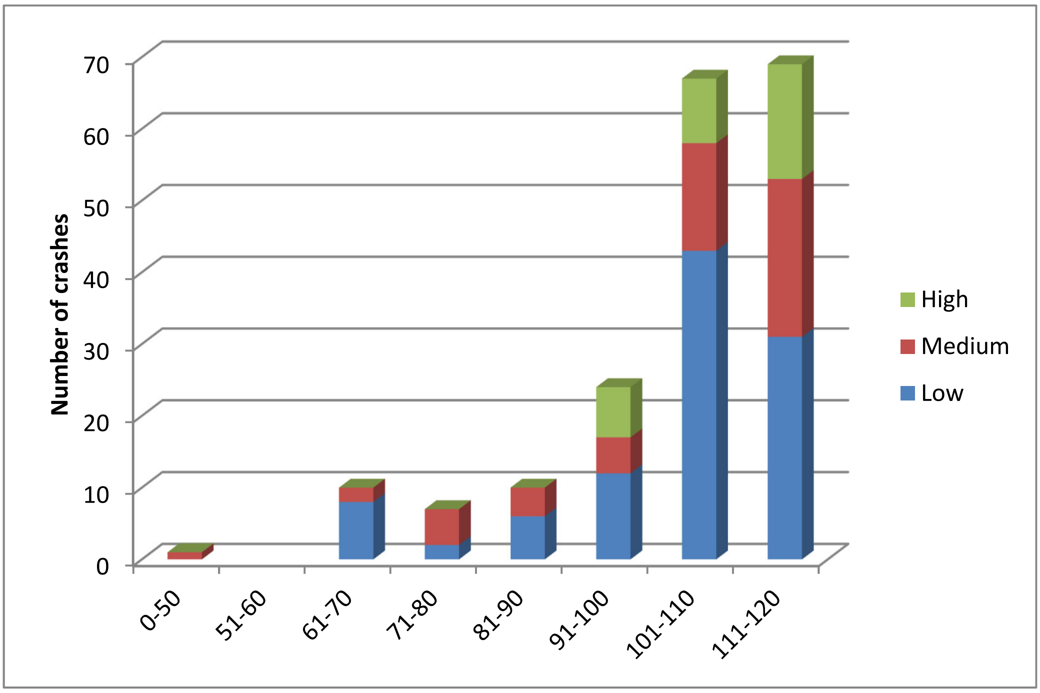



*Graph 13: Number of pedestrian crashes per time of day and crash rate category*

It can be seen from Graph 13 that the risk of being involved in a pedestrian crash increases during the night peak. The AM peak period is the time of day in which the risk of being involved in a pedestrian crash is the lowest, with the night peak period posing the highest risk of being involved in a pedestrian crash. This phenomenon was also observed in the literature review (Bianco, 2017). This variable was determined to be significant in the Crash Rate Model. It is interesting to note that this variable was not determined to be collinear to the average travel speed and traffic volume, as was determined in the Injury Model. The reason for this is since the crash rates used for the Crash Rate Model were determined per time of day, resulting in specific crash rates that were applied in the AM, off, PM and night peak periods. It is therefore expected that the Crash Rate Model will be dependent on the peak period.

### 5.3.4 Average Travel Speed

The number of pedestrian crashes that occurred for different travel speed categories can be seen in Graph 14.

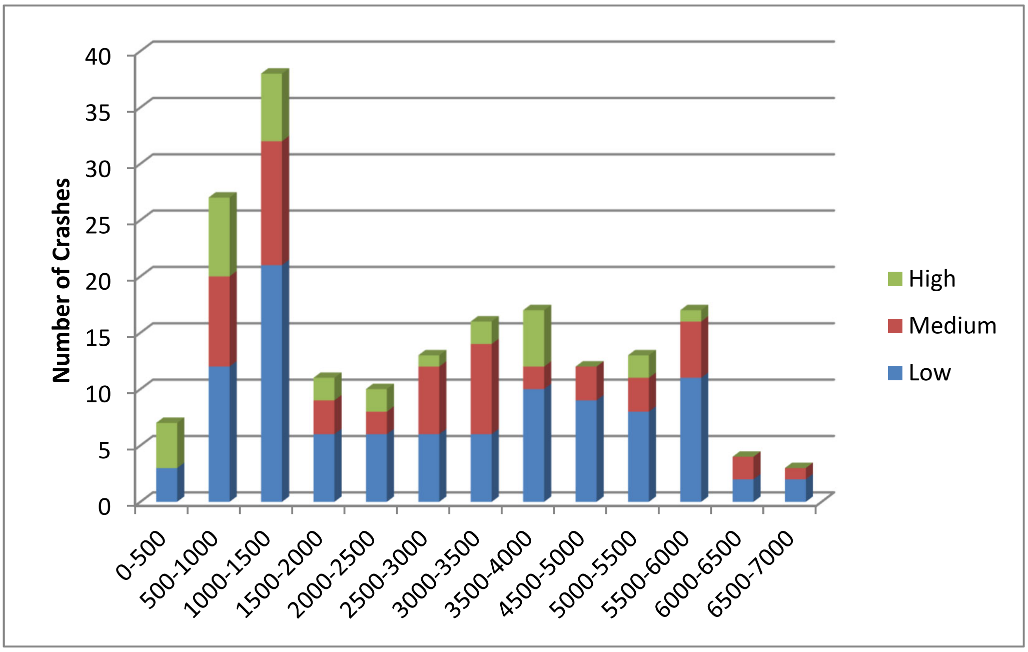



*Graph 14: Pedestrian crashes for different travel speed categories*

Graph 14 indicates that the majority of the pedestrian crashes occur during the times when the average travel speed is approaching or at free-flow speed (120 km/h). This trend was also determined in the Crash Rate Model, where it was found that the risk of being involved in a pedestrian crash increases as the travel speed increase. A reason for this is due to the equations of motion where it can be seen that the time and distance required to come to a complete stop, increases as the travel speed increases. Refer to Section 5.1.4 for more detail on the equations of motion. In the cases where speeds that are close or equal to break down speed (congested speed) are travelled, the risk of being involved in a pedestrian crash decreases rapidly. The decrease in the risk is because the vehicles move at low enough speeds that pedestrians can walk through the traffic stream without being hit since the vehicles travel slow enough to stop for the pedestrians on the road.

### 5.3.5 Average Traffic Volume

The number of pedestrian crashes for different traffic volume categories, and the three different crash rate categories can be seen in Graph 15.




*Graph 15: Number of pedestrian crashes for different traffic volume categories*

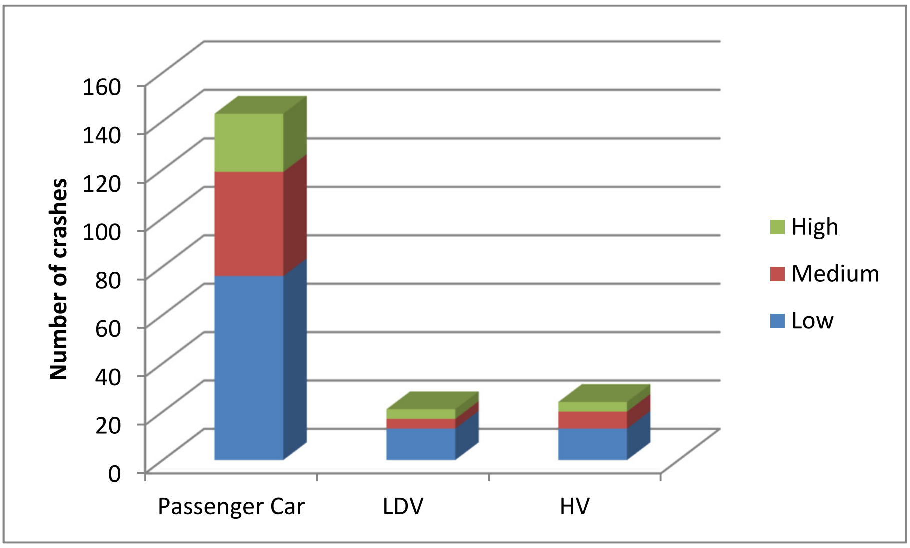
It can be concluded from Graph 15 that the majority of the pedestrian crashes occurred when low traffic flow conditions were recorded on the road. This observation was also made in the Crash Rate Model, where it was determined that the risk of being involved in a pedestrian crash on a section of the road associated with a low or medium crash rate increases when the traffic volumes decrease. The risk of being in a pedestrian crash is the lowest when the highest traffic volumes are recorded on the road. There is, however, a clear trend of an increase in the risk of being involved in a pedestrian crash as the traffic volume increases from 1500 vehicles per hour to 6000 vehicles per hour with a rapid decrease in the crash risk for traffic volumes that exceed 6000 vehicles per hour. The increased crash risk for the medium to high traffic volume can be explained by the increased number of conflict points on the road, while the increased risk of being involved in a pedestrian crash for low traffic volume conditions has to be viewed together with the average travel speed (see the section above). This attribute is, therefore, also not able to be viewed in isolation since it is clear that the traffic volume has a complex relationship with the other variables that are included in the model. This does not mean that the average traffic volume on the road has a collinear relationship with the other variables - it only means that the combination of these factors has a dynamic relationship where the attributes together have an impact on the pedestrian crash risk, which cannot necessarily be explained when looking at the attributes individually.

For example, the reason why the crash risk decreases rapidly for the very high traffic volumes (>6000vph), is due to the decrease in the average travel speed. Although the number of conflict points increases as the traffic volume increases, the average speed travelled decrease rapidly to breakdown speed, resulting in a traffic stream that moves so slow that the pedestrians can walk through the traffic stream on the road without the risk of being hit by a vehicle.

This, therefore, results in the conclusion that the odds of being in a pedestrian crash on a low crash rate segment decreases for very high traffic volumes. This is an interesting conclusion since literature indicated that the number of conflict points and therefore also the number of crashes increase as the traffic volume increase. A reason why this observation is made might be since the traffic volume on the Gauteng FMS Network increases drastically during the AM and PM peak hours of the day. The number of pedestrian crashes does not increase with such a steep gradient. Also, the injuries sustained in these crashes are much lower than during low traffic conditions. The pedestrian crashes where the pedestrian was, therefore, able to walk away might have not been reported, which will result in skew results.

### 5.3.6 Vehicle Type

The number of pedestrian crashes per vehicle type can be seen in Graph 16. This variable was found to be statistically insignificant in the Crash Rate Model.




*Graph 16: Number of pedestrian crashes per vehicle type and crash rate category*

Graph 16 provides a summary of the number of crashes that occur per crash rate category and vehicle type. It is evident from the graph that passenger vehicles were involved in the majority of the crashes. No specific trend can be determined in terms of the crash rate when this graph is studied. The percentage split between the low, medium and high crash rates per vehicle type is approximately the same, as indicated in Table 31.

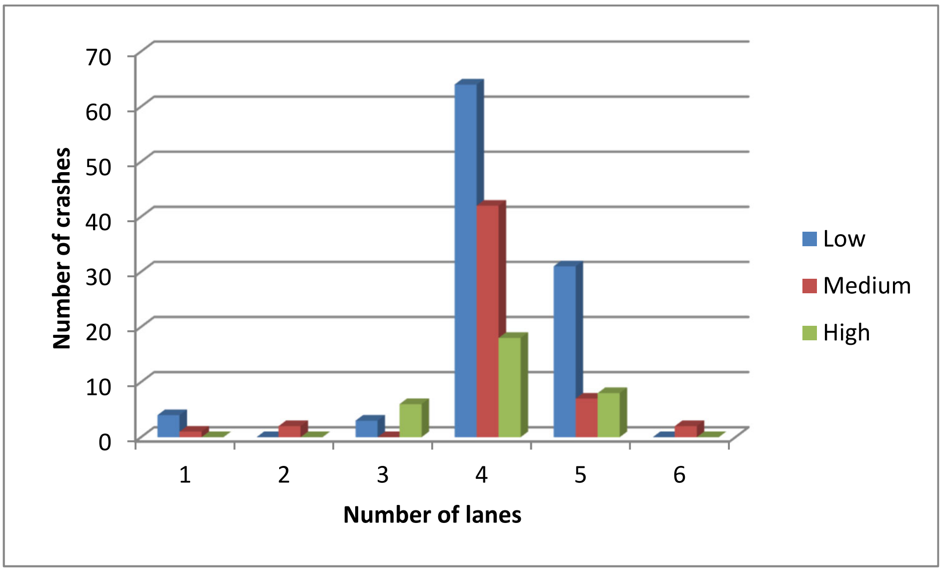
*Table 31: Percentage split between the vehicle types and crash rates*

| The proportion of crashes involving | Passenger vehicles | LDV's | Heavy vehicles |
|---|---|---|---|
| Low Crash Rate | 74% | 13% | 13% |
| Medium Crash Rate | 80% | 7% | 13% |
| High Crash Rate | 74% | 13% | 13% |

The low crash category is also the category in which the majority of the crashes occurred, which can also be concluded from the other attributes. There are, however, no clear relationship between the number of pedestrian crashes involving passenger vehicles and the crash rates.

### 5.3.7 Number of Lanes to Cross

The number of pedestrian crashes for the different cross-sections can be seen in Graph 17.




*Graph 17: Number of pedestrian crashes per cross-section type and crash rate category*

From Graph 17 it can be determined that the majority of the crashes occur on the sections of road which have four lanes. This trend is observed for all three crash rate categories. As the number of lanes on a road increases, the capacity of the road increases. This can either result in a longer period where free-flow conditions are experienced (when the traffic volume stays constant), which result in higher travel speeds, or it can result in an increase in the number of vehicles on the road since the additional lanes increase the capacity. In this scenario, the speed will stay constant, and the number of conflict point will increase due to the increase in the traffic volume. The two scenarios described above have a direct influence on the crash rate and is already incorporated in the model. It is therefore concluded that the cross-section does not influence the crash rates since it is already accounted for in the average traffic volume and average travel speed attributes.

The literature indicated that an increase in the number of lanes also increases the number of conflict points on the road. This would, therefore, increase the risk of being involved in a pedestrian crash. It is clear from the Crash Rate Model that this observation cannot be made as part of this study (Zhang, Chen, and Wei, 2019). It is, however, assumed that the increased risk of being involved in a pedestrian crash due to an increase in conflict points is included in the traffic volume attribute.

### 5.3.8 Remarks on the Crash Rate Model

From the Crash Rate Model, it is clear that the number of lanes to cross as well as the vehicle type involved in the pedestrian crash were not statistically significant for the Crash Rate Model. Also, when comparing the coefficients of the model, it is clear that some of the variables have a more significant impact on the risk of being fatally injured than others.

*Table 32: Comparison of the weight of each variable for the Crash Rate Model*

| Independent Variable | Coefficient Value | Odds Ratio |
|---|---|---|
| Constant | 2.3620 | n/a |
| Land use | 0.3695 | 1.447011 |
| Day of crash | -0.5768 | 0.561693 |
| Peak period | 0.3387 | 1.403122 |
| Speed | -0.0379 | 0.962809 |
| Volume | 0.0005 | 1.0005 |

From Table 32, it is evident that the land use surrounding the crash has a significant impact on the crash risk. There is a high risk of being involved in a pedestrian crash for non-residential land uses; however, it is expected that this variable might present skew results since the majority of the study area is surrounded by non-residential land uses. It is, however, concluded that this model might not be applicable in other provinces or on other roads since it is expected that the results obtained when applying the Crash Rate Model might be skew if the land use attribute is included in the model.

The day on which the pedestrian crash occurs has the smallest impact on the risk of being involved in a pedestrian crash. It can be seen that the odds of being in a pedestrian crash increase with 56% during the weekend, however, the other variables that are included in the model have a more significant impact on the pedestrian crash risk than the day on which the crash occurred. The peak period and average traffic volume on the road was used to calculate the specific crash rates for the 26 sub-sections included in the study area. It was therefore expected that these two attributes would have a significant impact on the Crash Rate Model. The speed travelled on the road has a direct link with the traffic volume on the road. The speed will decrease in congested conditions and will increase under free-flow conditions. This variable was, therefore, also expected to have a significant impact on the Crash Rate Model.

The variables included in the Crash Rate Model cannot be viewed in isolation. The variables change constantly in the field, resulting in an interactive model where the probabilities to be involved in a pedestrian crash are determined using a combination of factors. It can be concluded that the peak period in which the crash occurred as well as the average speed travelled at the time of the crash have the most significant impact on the probability to be involved in a pedestrian crash.

## 5.4 APPLICATION OF THE INJURY MODEL IN OTHER LOCATIONS

It was determined from the comparison between the actual data and the Injury Model that the independent variables included in the model cannot be viewed in isolation when determining the probability of sustaining fatal injuries in a pedestrian crash. This is due to the dynamic relationship between the independent variables that result in a combined probability.

The purpose of this section is to give an outline and the conditions under which the Injury Model can be applied in other provinces in South Africa and on different roads than the roads included in the study area. Pedestrians walking along and crossing the freeway are not allowed in most countries Sinclair and Zuidgeest (2016), and since this model is specifically designed for freeways, it can only be applied in locations where pedestrians are observed on the freeways.

The Injury Model was developed in a specific environment. The variables that were included in the model are based on pedestrian crash data collected for the years 2013 – 2018. The entire study area is median divided, resulting in pedestrians that cross the road using the rolling-gap crossing method. The entire study area is also provided with street lighting, which improves the sight distance during the dusk and dark periods of the day. In order to use the Injury Model to predict the probability of sustaining fatal injuries in pedestrian crashes, the study area should consist of a dual carriageway. Street lighting should also be provided along the section of road under investigation. When looking at the data recorded at the CTO stations, it is clear that further research has to be done to derive an equation which can predict the average speed travelled for the different traffic volumes that are recorded throughout the day. A few observations can, however, be made. The travel speed of the vehicles decreases when high traffic volumes are present on the road, as shown in Graph 6 and Graph 8. It is only at very high traffic volumes that there is a significant decrease in the travel speed, which confirms the observations made in the Highway Capacity Manual 2010, (HCM 2010). The travel speed and traffic volume profiles for weekdays as well as weekends were investigated and were plotted on one graph to determine the relationship between the two attributes. It was determined that there is an indirect relationship between the average traffic volume and the average travel speed during all four peaks of the day. This relationship can be explained by two third-order polynomial equations. The relationship between the travel speed and traffic volume on every subsection is different; however, all the segments complied with the indirect third order polynomial relationship. These relationships were observed on weekdays as well as during the weekends. A selection of graphs indicating this relationship for a random selection of CTO stations and days can be seen in Appendix B.

It is clear from the graphs that the traffic volume profile on the FMS Network included in the study does not follow the same pattern for every CTO station. Some of the sub-sections or CTO stations are associated with a high AM peak, while other CTO stations indicate that the PM peak has the highest peak. The reason for this is due to the directional split that is observed on the freeway. Also, the average traffic volumes that were recorded on the road during the off peak period are within the same range as either the AM or PM peak hour (depending on the directional split). This is because the AM and PM peak periods are typically associated with a sudden increase in the traffic volume in a short period of time to the highest peak volume. This, therefore, results in a low traffic volume that increases rapidly to the highest peak before descending again, refer to Figure 16.

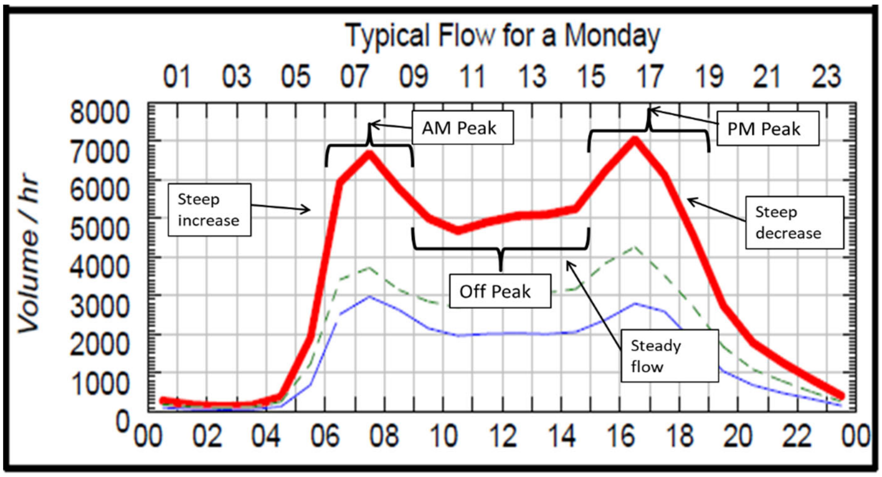



*Figure 16: Typical flow for a Monday indicating a steep increase and decreases for the AM and PM Peak*

It is evident from Figure 16 that the average traffic volume calculated for the AM and PM peak hour includes very low to very high volumes, while the off peak volumes are relatively steady throughout the peak period. This, therefore, results in the average AM or PM peak period volume being in the same range as the average off peak period volume. It does not mean that the congestion levels experienced during the off peak are higher than during the AM and PM peaks. The congestion is related to the jam density, which is reached during the AM and PM peaks and not during the off peaks. The jam density results in blocking back of vehicles, which therefore result in lower vehicle flows and lower travel speeds (breakdown speed).

It should also be noted that the model that was developed was based on a low confidence interval as well as a small data set. It is therefore concluded that, even though this model can be used at other locations with the same road environment as listed above, more research should be done to develop a more accurate model. Also, the relationship between the traffic volume and the speed travelled for the different traffic volumes should also be further investigated to obtain a clear understanding of how the traffic and speed patterns in other places would fit into this model. Adjustment factors have to be determined to make up for other traffic pattern observations.

It is concluded that the Injury Model can be applied in other locations if the section of road on which the investigation is done, is similar to the section of road that was investigated as part of this study. The section of road should comply with the following factors:

i) Dual carriageway provided with street lights;
ii) High order road; and
iii) The traffic volume and travel speed profiles during the day should have an inverse, third order polynomial relationship.

# 6. CONCLUSION

The MNL model predicting the injuries sustained, i.e., the probability of sustaining fatal injuries versus non-fatal injuries is the model which predicted the most accurate results. It can be concluded that the attributes that increase the probability of sustaining fatal injuries are:

- i) Pedestrian crashes that occur during a weekend;
- ii) Increase in travel speed;
- iii) Pedestrian crashes with larger vehicles; and
- iv) Increase in the number of lanes to cross.

The probability of sustaining fatal injuries in a pedestrian crash decreases as the traffic volume increases. The land use surrounding the crash site as well as the peak hour in which the pedestrian crash occurred was determined to be statistically insignificant for this model. The literature review indicated that the peak period in which the pedestrian crash occurs influences the injuries sustained in the pedestrian crash. This relationship was, however, not found in this study. A reason for this is the fact that the peak period in which the pedestrian crash occurred, is already incorporated in the traffic volume and the travel speed variables. It can, therefore, be concluded that the peak period does have an influence on the injuries sustained in a pedestrian crash, but that this attribute is collinear with the travel speed and traffic volume attributes.

It was determined that this model could be applied in other locations in South Africa or at locations where pedestrians are observed along with high speed and high-volume roads. It was concluded that if the surrounding area and the variables included in the model comply with the following factors, it can be applied:

- i) The section of road on which the model is applied should be median divided, forcing the pedestrian to cross the road using the rolling-gap crossing method;
- ii) The section of road should be a high speed road (higher order road);
- iii) The relationship between the speed and traffic volumes travelled on the road should have an indirect third order polynomial relationship; and
- iv) The section of road should be provided with street lighting, to improve the sight distance during the dusk and dark conditions of a day.

It was concluded that the Crash Rate Model does not have such a good fit as the Injury Model, due to larger p-stats values as well as smaller t-stats values. The RMSE value for this model was also larger than for the Injury Model, resulting in more proof that the Crash Rate Model has a less good fit.

It was determined that the risk of being involved in a pedestrian crash increases at the locations where non-residential land use is located nearby the pedestrian crash. The crash rate also increased during the times of the day associated with high speeds, i.e., during the off and night peak periods. The traffic volume recorded on the road had an impact on the pedestrian crash rates. The risk of being involved in a pedestrian crash for free-flow conditions increased dramatically. This result is consistent with the increased risk of being involved in a pedestrian crash during the times when high travel speeds are recorded on the road. The high travel speeds are typically recorded during the periods of the day when the road is at free-flow condition. The risk of being involved and injured in a pedestrian crash decreases rapidly for very high traffic volumes, which is also associated with low travel speeds.

This study was not without limitations. A limited number of data points were available for this study, resulting in MNL models which have a relatively low confidence level. It can also be assumed that the recording of the data might include some errors, due to late detection times, erroneous logging of information as well as due to pedestrians that pass away in the hospital and which are not included in the data set. Some of the speed and traffic volume data was also missing, resulting in some pedestrian crashes which were linked with estimated traffic volumes and travel speeds. Assumptions on the vehicle type that was involved in the pedestrian crash had to be made. This could have had an impact on the analysis results, since the wrong traffic speed, traffic volume, or vehicle type could have been assumed for some of the pedestrian crashes.

It can, however, be concluded that even with these limitations, it is only the land use variable that might be erroneous and which may have resulted in inaccurate results. This is due to the way in which the land use was assigned to the pedestrian crash without knowing whether the crash occurred at the origin, destination or only at a pass-by location. Another reason why it is expected that the land use attribute might result in skewed results is that the majority of the land use surrounding the study area was industrial land use.

More research is needed to determine a more accurate analysis and understanding of the models and the interaction between the variables, especially for the relationship between the traffic volume and the travel speed during the different times of a day. The number of data points used in this study limited the confidence interval and therefore, also the accuracy of the results. This was considered as one of the major limitations of the study.

The variables that have an impact on the injuries sustained in pedestrian crashes as well as on the crash rate cannot be viewed individually. This is because the independent variables included in the study have a dynamic relationship resulting in a combined probability which will not be able to be determined when the variables are investigated in isolation. Even when all the limitations are taken into account, it can be concluded that the Injury Model has a good fit and that this model can be applied in other locations in South Africa to determine the probability that a pedestrian will sustain fatal injuries in a pedestrian crash.

# 7. REFERENCES

Al-Ghamdi, A. S. (2002) 'Pedestrian-vehicle crashes and analytical techniques for stratified contingency tables', *Accident Analysis and Prevention*, 34(2), pp. 205–214. doi: 10.1016/S0001-4575(01)00015-X.

Amoh-Gyimah, R. *et al.* (2017) 'The effect of natural and built environmental characteristics on pedestrian-vehicle crash severity in Ghana', *International Journal of Injury Control and Safety Promotion*. Taylor & Francis, 24(4), pp. 459–468. doi: 10.1080/17457300.2016.1232274.

Askin, R. G. *et al.* (1991) 'ContentServer (1).pdf', *Physical Therapy*, pp. 1081–1100. doi: 10.1177/017084068800900203.

Ballesteros, M. F., Dischinger, P. C. and Langenberg, P. (2004) 'Pedestrian injuries and vehicle type in Maryland, 1995-1999', *Accident Analysis and Prevention*, 36(1), pp. 73–81. doi: 10.1016/S0001-4575(02)00129-X.

Bianco, Z. (2017) 'Analysis of Pedestrian Crash Characteristics and Contributing Causes in Central Florida'.

Brodsky, H. and Hakkert, A. S. (1988) 'Risk of a road accident in rainy weather', *Accident Analysis and Prevention*, 20(3), pp. 161–176. doi: 10.1016/0001-4575(88)90001-2.

Cameron, M. H. and Elvik, R. (2010) 'Nilsson's Power Model connecting speed and road trauma: Applicability by road type and alternative models for urban roads', *Accident Analysis and Prevention*. Elsevier Ltd, 42(6), pp. 1908–1915. doi: 10.1016/j.aap.2010.05.012.

Campbell, B. J. *et al.* (2004) 'A Review of Pedestrian Safety Research in the United States and Abroad', (January), p. 150.

Chen, H. *et al.* (2016) 'Analysis of risk factors affecting driver injury and crash injury with drivers under the influence of alcohol (DUI) and non-DUI', *Traffic Injury Prevention*. Taylor & Francis, 17(8), pp. 796–802. doi: 10.1080/15389588.2016.1168924.

Cinnamon, J., Schuurman, N. and Hameed, S. M. (2011) 'Pedestrian injury and human behaviour: Observing road-rule violations at high-incident intersections', *PLoS ONE*, 6(6). doi: 10.1371/journal.pone.0021063.

City Planning and Development Division (2008) 'Tshwane town-planning scheme, 2008', 2008(497).

Classen, S., Awadzi, K. D. and Mkanta, W. W. (2008) 'Person-vehicle-environment interactions predicting crash-related injury among older drivers', *American Journal of Occupational Therapy*, 62(5), pp. 580–587. doi: 10.5014/ajot.62.5.580.

Clifton, K. J., Burnier, C. V. and Akar, G. (2009) 'Severity of injury resulting from pedestrian-vehicle crashes: What can we learn from examining the built environment?', *Transportation Research Part D: Transport and Environment*. Elsevier Ltd, 14(6), pp. 425–436. doi: 10.1016/j.trd.2009.01.001.

Constant, A. and Lagarde, E. (2010) 'Protecting vulnerable road users from injury', *PLoS Medicine*, 7(3), pp. 1–4. doi: 10.1371/journal.pmed.1000228.

Crocetta, G. *et al.* (2015) 'The influence of vehicle front-end design on pedestrian ground impact', *Accident Analysis and Prevention*. Elsevier Ltd, 79, pp. 56–69. doi: 10.1016/j.aap.2015.03.009.

Dada, M., Zuidgeest, M. and Hess, S. (2019) 'Modelling pedestrian crossing choice on Cape Town's freeways: Caught between a rock and a hard place?', *Transportation Research Part F: Traffic Psychology and Behaviour*. Elsevier Ltd, 60, pp. 245–261. doi: 10.1016/j.trf.2018.10.005.

David, S. S. J. *et al.* (1986) 'Peripheral vision and the aetiology of child pedestrian accidents', *British Journal of Psychology*, pp. 117–135. doi: 10.1111/j.2044-8295.1986.tb01987.x.

DiMaggio, C., Durkin, M. and Richardson, L. D. (2006) 'The association of light trucks and vans with paediatric pedestrian deaths.', *International journal of injury control and safety promotion*, 13(2), pp. 95–99. doi: 10.1080/17457300500310038.

Forbes, G. (2012) 'Human Factors, Safety and the Impacts of Speed', *Conference and exhibition of the transportation association of Canada- Transportation: Innovations and Opportunities*, pp. 1–14. Available at: http://conf.tac-atc.ca/english/annualconference/tac2012/docs/session17/forbes2.pdf.

Gigerenzer, G., Krauss, S. and Vitouch, O. (2004) 'The null ritual: What you always wanted to know about null hypothesis testing but were afraid to ask', *Handbook on Quantitative Methods in the Social Sciences*, pp. 391–408.

Golob, T. F., Recker, W. W. and Alvarez, V. M. (2004) 'Freeway safety as a function of traffic flow', *Accident Analysis and Prevention*, 36(6), pp. 933–946. doi: 10.1016/j.aap.2003.09.006.

Graham, D., Glaister, S. and Anderson, R. (2005) 'The effects of area deprivation on the incidence of child and adult pedestrian casualties in England', *Accident Analysis and Prevention*, 37(1), pp. 125–135. doi: 10.1016/j.aap.2004.07.002.

Greenland, S. (2011) 'Null misinterpretation in statistical testing and its impact on health risk assessment', *Preventive Medicine*. Elsevier B.V., 53(4–5), pp. 225–228. doi: 10.1016/j.ypmed.2011.08.010.

Greenland, S. *et al.* (2016) 'Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations', *European Journal of Epidemiology*, 31(4), pp. 337–350. doi: 10.1007/s10654-016-0149-3.

Griffith, M. (1994) 'Comparison of the safety of lighting options on urban freeways', *Public Roads*, 58(2), pp. 8–15. Available at: http://trid.trb.org/view.aspx?id=413910.

Hair Joseph F (Kennesaw State University) Hult G Tomas M (Michigan State University) Ringle Christian M (Hamburg University of Technology, Germany and The University of Newcastle, Australia) Sarstedt Marko (Otto-von-Guericke University , Madeburg, Germany, A. (2017) *3b.HairBook2017PLS2ndEd. (1).pdf*.

doi: 10.1007/s10995-012-1023-x [doi].

Hine, J. and Russell, J. (1993) 'Traffic barriers and pedestrian crossing behaviour', *Journal of Transport Geography*, 1(4), pp. 230–239. doi: 10.1016/0966-6923(93)90047-4.

Huang, Y., Sun, D. (Jian) and Zhang, L. H. (2018) 'Effects of congestion on drivers' speed choice: Assessing the mediating role of state aggressiveness based on taxi floating car data', *Accident Analysis and Prevention*. Elsevier, 117(800), pp. 318–327. doi: 10.1016/j.aap.2018.04.030.

Jang, K. *et al.* (2010) 'Influential Factors on Level of Injury in Pedestrian Crashes: Applications of Ordered Probit Model with Robust Standard Errors', p. 17p. Available at: https://trid.trb.org/view/910750.

Kashani, A. T. and Besharati, M. M. (2017) 'Fatality rate of pedestrians and fatal crash involvement rate of drivers in pedestrian crashes: a case study of Iran', *International Journal of Injury Control and Safety Promotion*. Taylor & Francis, 24(2), pp. 222–231. doi: 10.1080/17457300.2016.1166139.

Kim, J. H. (2015) 'How to Choose the Level of Significance: A Pedagogical Note', *SSRN Electronic Journal*, (66373). doi: 10.2139/ssrn.2652773.

Kim, J. K. *et al.* (2010) 'A note on modeling pedestrian-injury severity in motor-vehicle crashes with the mixed logit model', *Accident Analysis and Prevention*. Elsevier Ltd, 42(6), pp. 1751–1758. doi: 10.1016/j.aap.2010.04.016.

Kononov, J. *et al.* (2013) 'Relationship between Traffic Density, Speed, and Safety and Its Implications for Setting Variable Speed Limits on Freeways', *Transportation Research Record: Journal of the Transportation Research Board*, 2280(1), pp. 1–9. doi: 10.3141/2280-01.

Kröyer, H. R. G. (2015) 'Is 30km/h a "safe" speed? Injury severity of pedestrians struck by a vehicle and the relation to travel speed and age', *IATSS Research*. International Association of Traffic and Safety Sciences, 39(1), pp. 42–50. doi: 10.1016/j.iatssr.2014.08.001.

Kröyer, H. R. G., Jonsson, T. and Várhelyi, A. (2014) 'Relative fatality risk curve to describe the effect of change in the impact speed on fatality risk of pedestrians struck by a motor vehicle', *Accident Analysis and Prevention*. Elsevier Ltd, 62, pp. 143–152. doi: 10.1016/j.aap.2013.09.007.

Lee, G. *et al.* (2016) 'Association between intersection characteristics and perceived crash risk among school-aged children', *Accident Analysis and Prevention*. Elsevier Ltd, 97, pp. 111–121. doi: 10.1016/j.aap.2016.09.001.

Li, D. *et al.* (2017) 'Analyzing pedestrian crash injury severity under different weather conditions', *Traffic Injury Prevention*. Taylor & Francis, 18(4), pp. 427–430. doi: 10.1080/15389588.2016.1207762.

Li, Y. and Fernie, G. (2010) 'Pedestrian behavior and safety on a two-stage crossing with a center refuge island and the effect of winter weather on pedestrian compliance rate', *Accident Analysis and Prevention*. Elsevier Ltd, 42(4), pp. 1156–1163. doi: 10.1016/j.aap.2010.01.004.

Liao, T. (2011) 'Multinomial Logit Models', *Interpreting Probability Models*, pp. 48–59. doi: 10.4135/9781412984577.n6.

Lin, P. S. *et al.* (2019) 'Development of countermeasures to effectively improve pedestrian safety in low-income areas', *Journal of Traffic and Transportation Engineering (English Edition)*. Elsevier Ltd, 6(2), pp. 162–174. doi: 10.1016/j.jtte.2019.02.001.

Luoma, J. and Peltola, H. (2013) 'Does facing traffic improve pedestrian safety?', *Accident Analysis and Prevention*. Elsevier Ltd, 50, pp. 1207–1210. doi: 10.1016/j.aap.2012.09.023.

Martin, J. (2002) 'Relationship between crash rate and hourly traffic flow on.pdf', 34, pp. 619–629.

Martin, J. L. and Wu, D. (2018) 'Pedestrian fatality and impact speed squared: Cloglog modeling from French national data', *Traffic Injury Prevention*. Taylor & Francis, 19(1), pp. 94–101. doi: 10.1080/15389588.2017.1332408.

Mayr, J. M. *et al.* (2018) 'Causes and consequences of pedestrian injuries in children', *European Journal of Pediatrics*, 162(3), pp. 184–190. doi: 10.1007/s00431-002-1099-1.

Miao, M., Yang, Y. and Liang, Y. (2016) 'Pedestrian crash risk assessment and intervention', *Advances in Mechanical Engineering*, 8(7), pp. 1–10. doi: 10.1177/1687814016653296.

Miranda-Moreno, L. F., Morency, P. and El-Geneidy, A. M. (2011) 'The link between built environment, pedestrian activity and pedestrian-vehicle collision occurrence at signalized intersections.', *Accident; analysis and prevention*. Elsevier Ltd, 43(5), pp. 1624–34. doi: 10.1016/j.aap.2011.02.005.

Moore, D. N. *et al.* (2011) 'Mixed logit analysis of bicyclist injury severity resulting from motor vehicle crashes at intersection and non-intersection locations', *Accident Analysis and Prevention*. Elsevier Ltd, 43(3), pp. 621–630. doi: 10.1016/j.aap.2010.09.015.

Ned, K. and Pierre, H. (2018) 'Minimum sample size estimation in PLS-SEM: The inverse square root and gamma-exponential methods', *Information Systems Journal*, 28(1), pp. 227–261. doi: 10.1111/isj.12131.

Noland, R. B., Klein, N. J. and Tulach, N. K. (2013) 'Do lower income areas have more pedestrian casualties?', *Accident Analysis and Prevention*. Elsevier Ltd, 59, pp. 337–345. doi: 10.1016/j.aap.2013.06.009.

Osama, A. and Sayed, T. (2017) 'Evaluating the Impact of Socioeconomics, Land Use, Built Environment, and Road Facility on Cyclist Safety', *Transportation Research Record: Journal of the Transportation Research Board*, 2659(1), pp. 33–42. doi: 10.3141/2659-04.

Ossiander, E. M. and Cummings, P. (2002) 'Freeway speed limits and traffic fatalities in Washington State', *Accident Analysis and Prevention*, 34(1), pp. 13–18. doi: 10.1016/S0001-4575(00)00098-1.

Oxley, J. *et al.* (2018) 'Falling while walking: A hidden contributor to pedestrian injury', *Accident Analysis and Prevention*. Elsevier Ltd, 114, pp. 77–82. doi: 10.1016/j.aap.2017.01.010.

Özkan, T., Porter, B. E. and Lajunen, T. (2016) 'Field studies in traffic safety behavior', *Transportation Research Part F: Traffic Psychology and Behaviour*, 42, pp. 423–424. doi: 10.1016/j.trf.2016.09.030.

Papadimitriou, E., Lassarre, S. and Yannis, G. (2016) 'Introducing human factors in pedestrian crossing behaviour models', *Transportation Research Part F: Traffic Psychology and Behaviour*. Elsevier Ltd, 36, pp. 69–82. doi: 10.1016/j.trf.2015.11.003.

Parmet, S., Lynm, C. and Glass, R. M. (2002a) 'Pedestrian safety', *Journal of the American Medical Association*, 288(17), p. 2212. doi: 10.1001/jama.288.17.2212.

Parmet, S., Lynm, C. and Glass, R. M. (2002b) *Pedestrian safety*, *Journal of the American Medical Association*. doi: 10.1001/jama.288.17.2212.

Peden, M. M. *et al.* (1996) 'Injured pedestrians in Cape Town - The role of alcohol', *South African Medical Journal*, 86(9 II), pp. 1103–1105.

Poó, F. M., Ledesma, R. D. and Trujillo, R. (2018) 'Pedestrian crossing behavior, an observational study in the city of Ushuaia, Argentina', *Traffic Injury Prevention*, 19(3), pp. 305–310. doi: 10.1080/15389588.2017.1391380.

Priyantha Wedagama, D. M., Bird, R. N. and Metcalfe, A. V. (2006) 'The influence of urban land-use on non-motorised transport casualties', *Accident Analysis and Prevention*, 38(6), pp. 1049–1057. doi: 10.1016/j.aap.2006.01.006.

Qin, X., Ivan, J. N. and Ravishanker, N. (2004) 'Selecting exposure measures in crash rate prediction for two-lane highway segments', *Accident Analysis and Prevention*, 36(2), pp. 183–191. doi: 10.1016/S0001-4575(02)00148-3.

Regev, S., Rolison, J. J. and Moutari, S. (2018) 'Crash risk by driver age, gender, and time of day using a new exposure methodology', *Journal of Safety Research*. The Authors, 66, pp. 131–140. doi: 10.1016/j.jsr.2018.07.002.

Rosén, E. and Sander, U. (2009) 'Pedestrian fatality risk as a function of car impact speed', *Accident Analysis and Prevention*, 41(3), pp. 536–542. doi: 10.1016/j.aap.2009.02.002.

Schmucker, U. *et al.* (2010) 'Real-world car-to-pedestrian-crash data from an urban centre', *Journal of Trauma Management & Outcomes*, 4(1), pp. 2–7. doi: 10.1186/1752-2897-4-2.

Sinclair, M. and Zuidgeest, M. (2016) 'Investigations into pedestrian crossing choices on Cape Town freeways', *Transportation Research Part F: Traffic Psychology and Behaviour*. Elsevier Ltd, 42, pp. 479–494. doi: 10.1016/j.trf.2015.07.006.

South Africa Committee of Transport Officials (2013) 'TMH 17 South African Trip Data Manual', *Coto*, 1(September).

Starkweather, J. and Key Moske, A. (2002) 'Multinomial logistic regression', *Nursing research*, 44(1), pp.

197–200. Available at: http://journals.lww.com/nursingresearchonline/Abstract/2002/11000/Multinomial_Logistic_Regression.9.aspx.

Tanishita, M. and van Wee, B. (2017) 'Impact of vehicle speeds and changes in mean speeds on per vehicle-kilometer traffic accident rates in Japan', *IATSS Research*. International Association of Traffic and Safety Sciences, 41(3), pp. 107–112. doi: 10.1016/j.iatssr.2016.09.003.

Tefft, B. C. (2013) 'Impact speed and a pedestrian's risk of severe injury or death', *Accident Analysis and Prevention*. Elsevier Ltd, 50, pp. 871–878. doi: 10.1016/j.aap.2012.07.022.

Thompson, L. L. *et al.* (2013) 'Impact of social and technological distraction on pedestrian crossing behaviour: An observational study', *Injury Prevention*, 19(4), pp. 232–237. doi: 10.1136/injuryprev-2012-040601.

Veazie, P. J. (2015) 'Understanding statistical testing', *SAGE Open*, 5(1). doi: 10.1177/2158244014567685.

Verster, T. and Fourie, E. (2018) 'The good, the bad and the ugly of South African fatal road accidents', *South African Journal of Science*, 114(7/8), pp. 63–70. doi: 10.17159/sajs.2018/20170427.

Xiaoyun, Z. *et al.* (2011) 'The first collision point position identification method in vehicle-pedestrian impact accident', *International Journal of Crashworthiness*, 16(2), pp. 181–194. doi: 10.1080/13588265.2010.548135.

Yeo, H. *et al.* (2013) 'Impact of traffic states on freeway crash involvement rates', *Accident Analysis and Prevention*. Elsevier Ltd, 50, pp. 713–723. doi: 10.1016/j.aap.2012.06.023.

Yeon, J., Hernandez, S. and Elefteriadou, L. (2009) 'Differences in Freeway Capacity by Day of the Week, Time of Day, and Segment Type', *Journal of Transportation Engineering*, 135(7), pp. 416–426. doi: 10.1061/(asce)0733-947x(2009)135:7(416).

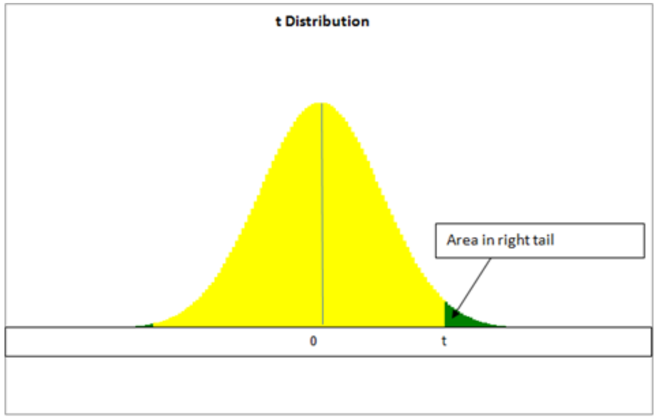
Zegeer, C. V. and Bushell, M. (2012) 'Pedestrian crash trends and potential countermeasures from around the world', *Accident Analysis and Prevention*. Elsevier Ltd, 44(1), pp. 3–11. doi: 10.1016/j.aap.2010.12.007.

Zhang, C., Chen, F. and Wei, Y. (2019) 'Evaluation of pedestrian crossing behavior and safety at uncontrolled mid-block crosswalks with different numbers of lanes in China', *Accident Analysis and Prevention*. Elsevier, 123(January 2018), pp. 263–273. doi: 10.1016/j.aap.2018.12.002.

Zheng, Z., Ahn, S. and Monsere, C. M. (2010) 'Impact of traffic oscillations on freeway crash occurrences', *Accident Analysis and Prevention*, 42(2), pp. 626–636. doi: 10.1016/j.aap.2009.10.009.

# APPENDIX A

*T-Distribution Table*




# *t* Distribution

| | Area in right tail = 0.25 | Area in right tail = 0.20 | Area in right tail = 0.15 | Area in right tail = 0.10 | Area in right tail = 0.05 | Area in right tail = 0.025 | Area in right tail = 0.02 | Area in right tail = 0.01 | Area in right tail = 0.005 | Area in right tail = 0.0025 | Area in right tail = 0.001 | Area in right tail = 0.0005 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| DF | t-score | t-score | t-score | t-score | t-score | t-score | t-score | t-score | t-score | t-score | t-score | t-score |
| 1 | 1.000 | 1.376 | 1.963 | 3.078 | 6.314 | 12.706 | 15.895 | 31.821 | 63.657 | 127.321 | 318.309 | 636.619 |
| 2 | 0.816 | 1.061 | 1.386 | 1.886 | 2.920 | 4.303 | 4.849 | 6.965 | 9.925 | 14.089 | 22.327 | 31.599 |
| 3 | 0.765 | 0.978 | 1.250 | 1.638 | 2.353 | 3.182 | 3.482 | 4.541 | 5.841 | 7.453 | 10.215 | 12.924 |
| 4 | 0.741 | 0.941 | 1.190 | 1.533 | 2.132 | 2.776 | 2.999 | 3.747 | 4.604 | 5.598 | 7.173 | 8.610 |
| 5 | 0.727 | 0.920 | 1.156 | 1.476 | 2.015 | 2.571 | 2.757 | 3.365 | 4.032 | 4.773 | 5.893 | 6.869 |
| 6 | 0.718 | 0.906 | 1.134 | 1.440 | 1.943 | 2.447 | 2.612 | 3.143 | 3.707 | 4.317 | 5.208 | 5.959 |
| 7 | 0.711 | 0.896 | 1.119 | 1.415 | 1.895 | 2.365 | 2.517 | 2.998 | 3.499 | 4.029 | 4.785 | 5.408 |
| 8 | 0.706 | 0.889 | 1.108 | 1.397 | 1.860 | 2.306 | 2.449 | 2.896 | 3.355 | 3.833 | 4.501 | 5.041 |
| 9 | 0.703 | 0.883 | 1.100 | 1.383 | 1.833 | 2.262 | 2.398 | 2.821 | 3.250 | 3.690 | 4.297 | 4.781 |
| 10 | 0.700 | 0.879 | 1.093 | 1.372 | 1.812 | 2.228 | 2.359 | 2.764 | 3.169 | 3.581 | 4.144 | 4.587 |
| 11 | 0.697 | 0.876 | 1.088 | 1.363 | 1.796 | 2.201 | 2.328 | 2.718 | 3.106 | 3.497 | 4.025 | 4.437 |
| 12 | 0.695 | 0.873 | 1.083 | 1.356 | 1.782 | 2.179 | 2.303 | 2.681 | 3.055 | 3.428 | 3.930 | 4.318 |
| 13 | 0.694 | 0.870 | 1.079 | 1.350 | 1.771 | 2.160 | 2.282 | 2.650 | 3.012 | 3.372 | 3.852 | 4.221 |

| | Area in right tail = 0.25 | Area in right tail = 0.20 | Area in right tail = 0.15 | Area in right tail = 0.10 | Area in right tail = 0.05 | Area in right tail = 0.025 | Area in right tail = 0.02 | Area in right tail = 0.01 | Area in right tail = 0.005 | Area in right tail = 0.0025 | Area in right tail = 0.001 | Area in right tail = 0.0005 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| DF | t-score | t-score | t-score | t-score | t-score | t-score | t-score | t-score | t-score | t-score | t-score | t-score |
| 15 | 0.691 | 0.866 | 1.074 | 1.341 | 1.753 | 2.131 | 2.249 | 2.602 | 2.947 | 3.286 | 3.733 | 4.073 |
| 16 | 0.690 | 0.865 | 1.071 | 1.337 | 1.746 | 2.120 | 2.235 | 2.583 | 2.921 | 3.252 | 3.686 | 4.015 |
| 17 | 0.689 | 0.863 | 1.069 | 1.333 | 1.740 | 2.110 | 2.224 | 2.567 | 2.898 | 3.222 | 3.646 | 3.965 |
| 18 | 0.688 | 0.862 | 1.067 | 1.330 | 1.734 | 2.101 | 2.214 | 2.552 | 2.878 | 3.197 | 3.610 | 3.922 |
| 19 | 0.688 | 0.861 | 1.066 | 1.328 | 1.729 | 2.093 | 2.205 | 2.539 | 2.861 | 3.174 | 3.579 | 3.883 |
| 20 | 0.687 | 0.860 | 1.064 | 1.325 | 1.725 | 2.086 | 2.197 | 2.528 | 2.845 | 3.153 | 3.552 | 3.850 |
| 21 | 0.686 | 0.859 | 1.063 | 1.323 | 1.721 | 2.080 | 2.189 | 2.518 | 2.831 | 3.135 | 3.527 | 3.819 |
| 22 | 0.686 | 0.858 | 1.061 | 1.321 | 1.717 | 2.074 | 2.183 | 2.508 | 2.819 | 3.119 | 3.505 | 3.792 |
| 23 | 0.685 | 0.858 | 1.060 | 1.319 | 1.714 | 2.069 | 2.177 | 2.500 | 2.807 | 3.104 | 3.485 | 3.768 |
| 24 | 0.685 | 0.857 | 1.059 | 1.318 | 1.711 | 2.064 | 2.172 | 2.492 | 2.797 | 3.091 | 3.467 | 3.745 |
| 25 | 0.684 | 0.856 | 1.058 | 1.316 | 1.708 | 2.060 | 2.167 | 2.485 | 2.787 | 3.078 | 3.450 | 3.725 |
| 26 | 0.684 | 0.856 | 1.058 | 1.315 | 1.706 | 2.056 | 2.162 | 2.479 | 2.779 | 3.067 | 3.435 | 3.707 |
| 27 | 0.684 | 0.855 | 1.057 | 1.314 | 1.703 | 2.052 | 2.158 | 2.473 | 2.771 | 3.057 | 3.421 | 3.690 |
| 28 | 0.683 | 0.855 | 1.056 | 1.313 | 1.701 | 2.048 | 2.154 | 2.467 | 2.763 | 3.047 | 3.408 | 3.674 |
| 29 | 0.683 | 0.854 | 1.055 | 1.311 | 1.699 | 2.045 | 2.150 | 2.462 | 2.756 | 3.038 | 3.396 | 3.659 |
| 30 | 0.683 | 0.854 | 1.055 | 1.310 | 1.697 | 2.042 | 2.147 | 2.457 | 2.750 | 3.030 | 3.385 | 3.646 |
| 31 | 0.682 | 0.853 | 1.054 | 1.309 | 1.696 | 2.040 | 2.144 | 2.453 | 2.744 | 3.022 | 3.375 | 3.633 |
| 32 | 0.682 | 0.853 | 1.054 | 1.309 | 1.694 | 2.037 | 2.141 | 2.449 | 2.738 | 3.015 | 3.365 | 3.622 |
| 33 | 0.682 | 0.853 | 1.053 | 1.308 | 1.692 | 2.035 | 2.138 | 2.445 | 2.733 | 3.008 | 3.356 | 3.611 |
| 34 | 0.682 | 0.852 | 1.052 | 1.307 | 1.691 | 2.032 | 2.136 | 2.441 | 2.728 | 3.002 | 3.348 | 3.601 |
| 35 | 0.682 | 0.852 | 1.052 | 1.306 | 1.690 | 2.030 | 2.133 | 2.438 | 2.724 | 2.996 | 3.340 | 3.591 |
| 36 | 0.681 | 0.852 | 1.052 | 1.306 | 1.688 | 2.028 | 2.131 | 2.434 | 2.719 | 2.990 | 3.333 | 3.582 |
| 37 | 0.681 | 0.851 | 1.051 | 1.305 | 1.687 | 2.026 | 2.129 | 2.431 | 2.715 | 2.985 | 3.326 | 3.574 |
| 38 | 0.681 | 0.851 | 1.051 | 1.304 | 1.686 | 2.024 | 2.127 | 2.429 | 2.712 | 2.980 | 3.319 | 3.566 |
| 39 | 0.681 | 0.851 | 1.050 | 1.304 | 1.685 | 2.023 | 2.125 | 2.426 | 2.708 | 2.976 | 3.313 | 3.558 |
| 40 | 0.681 | 0.851 | 1.050 | 1.303 | 1.684 | 2.021 | 2.123 | 2.423 | 2.704 | 2.971 | 3.307 | 3.551 |

| | Area in right tail = 0.25 | Area in right tail = 0.20 | Area in right tail = 0.15 | Area in right tail = 0.10 | Area in right tail = 0.05 | Area in right tail = 0.025 | Area in right tail = 0.02 | Area in right tail = 0.01 | Area in right tail = 0.005 | Area in right tail = 0.0025 | Area in right tail = 0.001 | Area in right tail = 0.0005 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| DF | t-score | t-score | t-score | t-score | t-score | t-score | t-score | t-score | t-score | t-score | t-score | t-score |
| 42 | 0.680 | 0.850 | 1.049 | 1.302 | 1.682 | 2.018 | 2.120 | 2.418 | 2.698 | 2.963 | 3.296 | 3.538 |
| 43 | 0.680 | 0.850 | 1.049 | 1.302 | 1.681 | 2.017 | 2.118 | 2.416 | 2.695 | 2.959 | 3.291 | 3.532 |
| 44 | 0.680 | 0.850 | 1.049 | 1.301 | 1.680 | 2.015 | 2.116 | 2.414 | 2.692 | 2.956 | 3.286 | 3.526 |
| 45 | 0.680 | 0.850 | 1.049 | 1.301 | 1.679 | 2.014 | 2.115 | 2.412 | 2.690 | 2.952 | 3.281 | 3.520 |
| 46 | 0.680 | 0.850 | 1.048 | 1.300 | 1.679 | 2.013 | 2.114 | 2.410 | 2.687 | 2.949 | 3.277 | 3.515 |
| 47 | 0.680 | 0.849 | 1.048 | 1.300 | 1.678 | 2.012 | 2.112 | 2.408 | 2.685 | 2.946 | 3.273 | 3.510 |
| 48 | 0.680 | 0.849 | 1.048 | 1.299 | 1.677 | 2.011 | 2.111 | 2.407 | 2.682 | 2.943 | 3.269 | 3.505 |
| 49 | 0.680 | 0.849 | 1.048 | 1.299 | 1.677 | 2.010 | 2.110 | 2.405 | 2.680 | 2.940 | 3.265 | 3.500 |
| 50 | 0.679 | 0.849 | 1.047 | 1.299 | 1.676 | 2.009 | 2.109 | 2.403 | 2.678 | 2.937 | 3.261 | 3.496 |
| 51 | 0.679 | 0.849 | 1.047 | 1.298 | 1.675 | 2.008 | 2.108 | 2.402 | 2.676 | 2.934 | 3.258 | 3.492 |
| 52 | 0.679 | 0.849 | 1.047 | 1.298 | 1.675 | 2.007 | 2.107 | 2.400 | 2.674 | 2.932 | 3.255 | 3.488 |
| 53 | 0.679 | 0.848 | 1.047 | 1.298 | 1.674 | 2.006 | 2.106 | 2.399 | 2.672 | 2.929 | 3.251 | 3.484 |
| 54 | 0.679 | 0.848 | 1.046 | 1.297 | 1.674 | 2.005 | 2.105 | 2.397 | 2.670 | 2.927 | 3.248 | 3.480 |
| 55 | 0.679 | 0.848 | 1.046 | 1.297 | 1.673 | 2.004 | 2.104 | 2.396 | 2.668 | 2.925 | 3.245 | 3.476 |
| 56 | 0.679 | 0.848 | 1.046 | 1.297 | 1.673 | 2.003 | 2.103 | 2.395 | 2.667 | 2.923 | 3.242 | 3.473 |
| 57 | 0.679 | 0.848 | 1.046 | 1.297 | 1.672 | 2.002 | 2.102 | 2.394 | 2.665 | 2.920 | 3.239 | 3.470 |
| 58 | 0.679 | 0.848 | 1.046 | 1.296 | 1.672 | 2.002 | 2.101 | 2.392 | 2.663 | 2.918 | 3.237 | 3.466 |
| 59 | 0.679 | 0.848 | 1.046 | 1.296 | 1.671 | 2.001 | 2.100 | 2.391 | 2.662 | 2.916 | 3.234 | 3.463 |
| 60 | 0.679 | 0.848 | 1.045 | 1.296 | 1.671 | 2.000 | 2.099 | 2.390 | 2.660 | 2.915 | 3.232 | 3.460 |
| 61 | 0.679 | 0.848 | 1.045 | 1.296 | 1.670 | 2.000 | 2.099 | 2.389 | 2.659 | 2.913 | 3.229 | 3.457 |
| 62 | 0.678 | 0.847 | 1.045 | 1.295 | 1.670 | 1.999 | 2.098 | 2.388 | 2.657 | 2.911 | 3.227 | 3.454 |
| 63 | 0.678 | 0.847 | 1.045 | 1.295 | 1.669 | 1.998 | 2.097 | 2.387 | 2.656 | 2.909 | 3.225 | 3.452 |
| 64 | 0.678 | 0.847 | 1.045 | 1.295 | 1.669 | 1.998 | 2.096 | 2.386 | 2.655 | 2.908 | 3.223 | 3.449 |
| 65 | 0.678 | 0.847 | 1.045 | 1.295 | 1.669 | 1.997 | 2.096 | 2.385 | 2.654 | 2.906 | 3.220 | 3.447 |
| 66 | 0.678 | 0.847 | 1.045 | 1.295 | 1.668 | 1.997 | 2.095 | 2.384 | 2.652 | 2.904 | 3.218 | 3.444 |
| 67 | 0.678 | 0.847 | 1.045 | 1.294 | 1.668 | 1.996 | 2.095 | 2.383 | 2.651 | 2.903 | 3.216 | 3.442 |

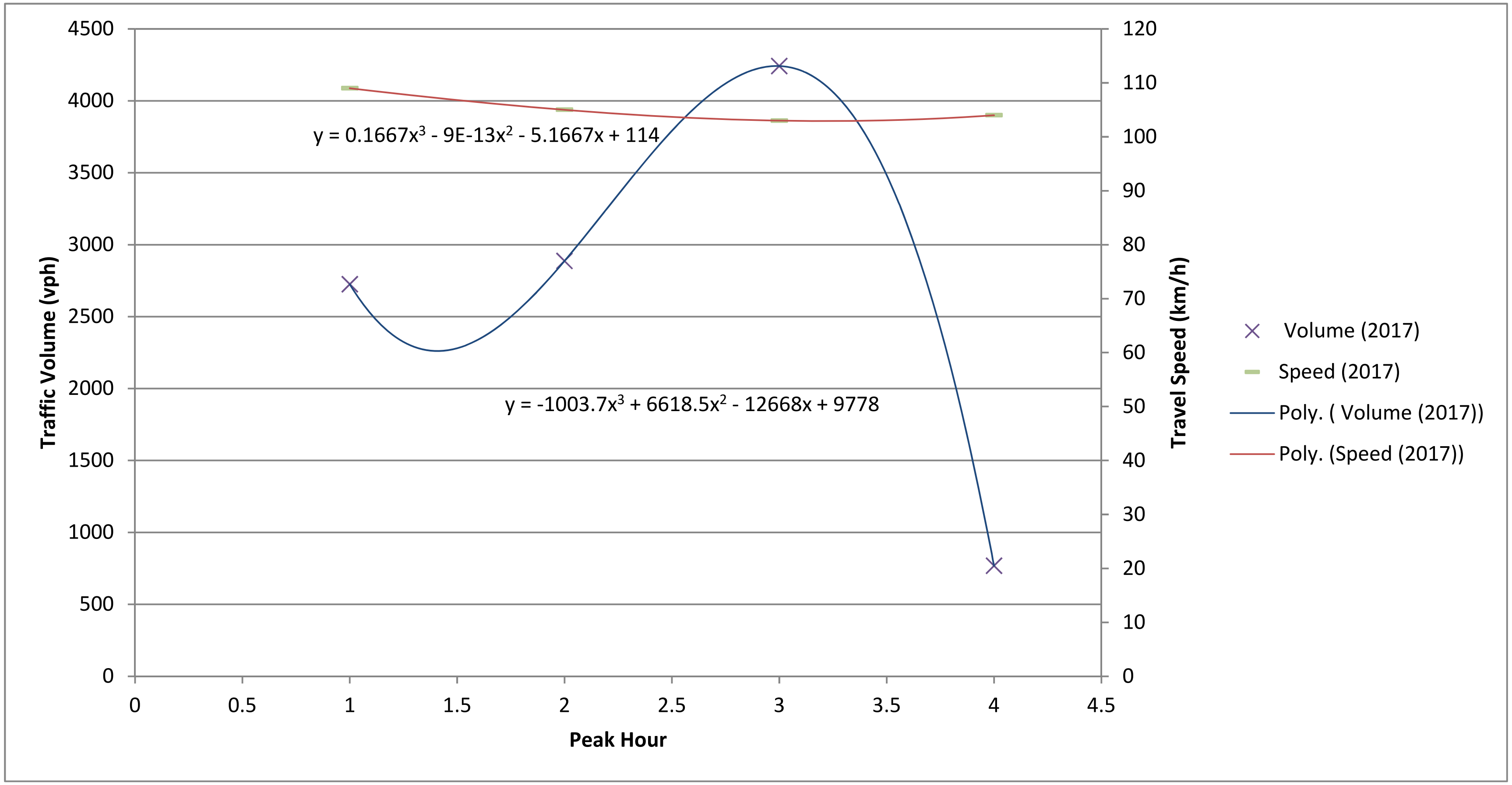
| | Area in right tail = 0.25 | Area in right tail = 0.20 | Area in right tail = 0.15 | Area in right tail = 0.10 | Area in right tail = 0.05 | Area in right tail = 0.025 | Area in right tail = 0.02 | Area in right tail = 0.01 | Area in right tail = 0.005 | Area in right tail = 0.0025 | Area in right tail = 0.001 | Area in right tail = 0.0005 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| DF | t-score | t-score | t-score | t-score | t-score | t-score | t-score | t-score | t-score | t-score | t-score | t-score |
| 69 | 0.678 | 0.847 | 1.044 | 1.294 | 1.667 | 1.995 | 2.093 | 2.382 | 2.649 | 2.900 | 3.213 | 3.437 |
| 70 | 0.678 | 0.847 | 1.044 | 1.294 | 1.667 | 1.994 | 2.093 | 2.381 | 2.648 | 2.899 | 3.211 | 3.435 |
| 71 | 0.678 | 0.847 | 1.044 | 1.294 | 1.667 | 1.994 | 2.092 | 2.380 | 2.647 | 2.897 | 3.209 | 3.433 |
| 72 | 0.678 | 0.847 | 1.044 | 1.293 | 1.666 | 1.993 | 2.092 | 2.379 | 2.646 | 2.896 | 3.207 | 3.431 |
| 73 | 0.678 | 0.847 | 1.044 | 1.293 | 1.666 | 1.993 | 2.091 | 2.379 | 2.645 | 2.895 | 3.206 | 3.429 |
| 74 | 0.678 | 0.847 | 1.044 | 1.293 | 1.666 | 1.993 | 2.091 | 2.378 | 2.644 | 2.894 | 3.204 | 3.427 |
| 75 | 0.678 | 0.846 | 1.044 | 1.293 | 1.665 | 1.992 | 2.090 | 2.377 | 2.643 | 2.892 | 3.202 | 3.425 |
| 76 | 0.678 | 0.846 | 1.044 | 1.293 | 1.665 | 1.992 | 2.090 | 2.376 | 2.642 | 2.891 | 3.201 | 3.423 |
| 77 | 0.678 | 0.846 | 1.043 | 1.293 | 1.665 | 1.991 | 2.089 | 2.376 | 2.641 | 2.890 | 3.199 | 3.421 |
| 78 | 0.678 | 0.846 | 1.043 | 1.292 | 1.665 | 1.991 | 2.089 | 2.375 | 2.640 | 2.889 | 3.198 | 3.420 |
| 79 | 0.678 | 0.846 | 1.043 | 1.292 | 1.664 | 1.990 | 2.088 | 2.374 | 2.640 | 2.888 | 3.197 | 3.418 |
| 80 | 0.678 | 0.846 | 1.043 | 1.292 | 1.664 | 1.990 | 2.088 | 2.374 | 2.639 | 2.887 | 3.195 | 3.416 |
| 81 | 0.678 | 0.846 | 1.043 | 1.292 | 1.664 | 1.990 | 2.087 | 2.373 | 2.638 | 2.886 | 3.194 | 3.415 |
| 82 | 0.677 | 0.846 | 1.043 | 1.292 | 1.664 | 1.989 | 2.087 | 2.373 | 2.637 | 2.885 | 3.193 | 3.413 |
| 83 | 0.677 | 0.846 | 1.043 | 1.292 | 1.663 | 1.989 | 2.087 | 2.372 | 2.636 | 2.884 | 3.191 | 3.412 |
| 84 | 0.677 | 0.846 | 1.043 | 1.292 | 1.663 | 1.989 | 2.086 | 2.372 | 2.636 | 2.883 | 3.190 | 3.410 |
| 85 | 0.677 | 0.846 | 1.043 | 1.292 | 1.663 | 1.988 | 2.086 | 2.371 | 2.635 | 2.882 | 3.189 | 3.409 |
| 86 | 0.677 | 0.846 | 1.043 | 1.291 | 1.663 | 1.988 | 2.085 | 2.370 | 2.634 | 2.881 | 3.188 | 3.407 |
| 87 | 0.677 | 0.846 | 1.043 | 1.291 | 1.663 | 1.988 | 2.085 | 2.370 | 2.634 | 2.880 | 3.187 | 3.406 |
| 88 | 0.677 | 0.846 | 1.043 | 1.291 | 1.662 | 1.987 | 2.085 | 2.369 | 2.633 | 2.880 | 3.185 | 3.405 |
| 89 | 0.677 | 0.846 | 1.043 | 1.291 | 1.662 | 1.987 | 2.084 | 2.369 | 2.632 | 2.879 | 3.184 | 3.403 |
| 90 | 0.677 | 0.846 | 1.042 | 1.291 | 1.662 | 1.987 | 2.084 | 2.368 | 2.632 | 2.878 | 3.183 | 3.402 |
| 91 | 0.677 | 0.846 | 1.042 | 1.291 | 1.662 | 1.986 | 2.084 | 2.368 | 2.631 | 2.877 | 3.182 | 3.401 |
| 92 | 0.677 | 0.846 | 1.042 | 1.291 | 1.662 | 1.986 | 2.083 | 2.368 | 2.630 | 2.876 | 3.181 | 3.399 |
| 93 | 0.677 | 0.846 | 1.042 | 1.291 | 1.661 | 1.986 | 2.083 | 2.367 | 2.630 | 2.876 | 3.180 | 3.398 |
| 94 | 0.677 | 0.845 | 1.042 | 1.291 | 1.661 | 1.986 | 2.083 | 2.367 | 2.629 | 2.875 | 3.179 | 3.397 |

| | Area in right tail = 0.25 | Area in right tail = 0.20 | Area in right tail = 0.15 | Area in right tail = 0.10 | Area in right tail = 0.05 | Area in right tail = 0.025 | Area in right tail = 0.02 | Area in right tail = 0.01 | Area in right tail = 0.005 | Area in right tail = 0.0025 | Area in right tail = 0.001 | Area in right tail = 0.0005 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| DF | t-score | t-score | t-score | t-score | t-score | t-score | t-score | t-score | t-score | t-score | t-score | t-score |
| 96 | 0.677 | 0.845 | 1.042 | 1.290 | 1.661 | 1.985 | 2.082 | 2.366 | 2.628 | 2.873 | 3.177 | 3.395 |
| 97 | 0.677 | 0.845 | 1.042 | 1.290 | 1.661 | 1.985 | 2.082 | 2.365 | 2.627 | 2.873 | 3.176 | 3.394 |
| 98 | 0.677 | 0.845 | 1.042 | 1.290 | 1.661 | 1.984 | 2.081 | 2.365 | 2.627 | 2.872 | 3.175 | 3.393 |
| 99 | 0.677 | 0.845 | 1.042 | 1.290 | 1.660 | 1.984 | 2.081 | 2.365 | 2.626 | 2.871 | 3.175 | 3.392 |
| 100 | 0.677 | 0.845 | 1.042 | 1.290 | 1.660 | 1.984 | 2.081 | 2.364 | 2.626 | 2.871 | 3.174 | 3.390 |
| 101 | 0.677 | 0.845 | 1.042 | 1.290 | 1.660 | 1.984 | 2.081 | 2.364 | 2.625 | 2.870 | 3.173 | 3.389 |
| 102 | 0.677 | 0.845 | 1.042 | 1.290 | 1.660 | 1.983 | 2.080 | 2.363 | 2.625 | 2.869 | 3.172 | 3.388 |
| 103 | 0.677 | 0.845 | 1.042 | 1.290 | 1.660 | 1.983 | 2.080 | 2.363 | 2.624 | 2.869 | 3.171 | 3.388 |
| 104 | 0.677 | 0.845 | 1.042 | 1.290 | 1.660 | 1.983 | 2.080 | 2.363 | 2.624 | 2.868 | 3.170 | 3.387 |
| 105 | 0.677 | 0.845 | 1.042 | 1.290 | 1.659 | 1.983 | 2.080 | 2.362 | 2.623 | 2.868 | 3.170 | 3.386 |
| 106 | 0.677 | 0.845 | 1.042 | 1.290 | 1.659 | 1.983 | 2.079 | 2.362 | 2.623 | 2.867 | 3.169 | 3.385 |
| 107 | 0.677 | 0.845 | 1.041 | 1.290 | 1.659 | 1.982 | 2.079 | 2.362 | 2.623 | 2.866 | 3.168 | 3.384 |
| 108 | 0.677 | 0.845 | 1.041 | 1.289 | 1.659 | 1.982 | 2.079 | 2.361 | 2.622 | 2.866 | 3.167 | 3.383 |
| 109 | 0.677 | 0.845 | 1.041 | 1.289 | 1.659 | 1.982 | 2.079 | 2.361 | 2.622 | 2.865 | 3.167 | 3.382 |
| 110 | 0.677 | 0.845 | 1.041 | 1.289 | 1.659 | 1.982 | 2.078 | 2.361 | 2.621 | 2.865 | 3.166 | 3.381 |

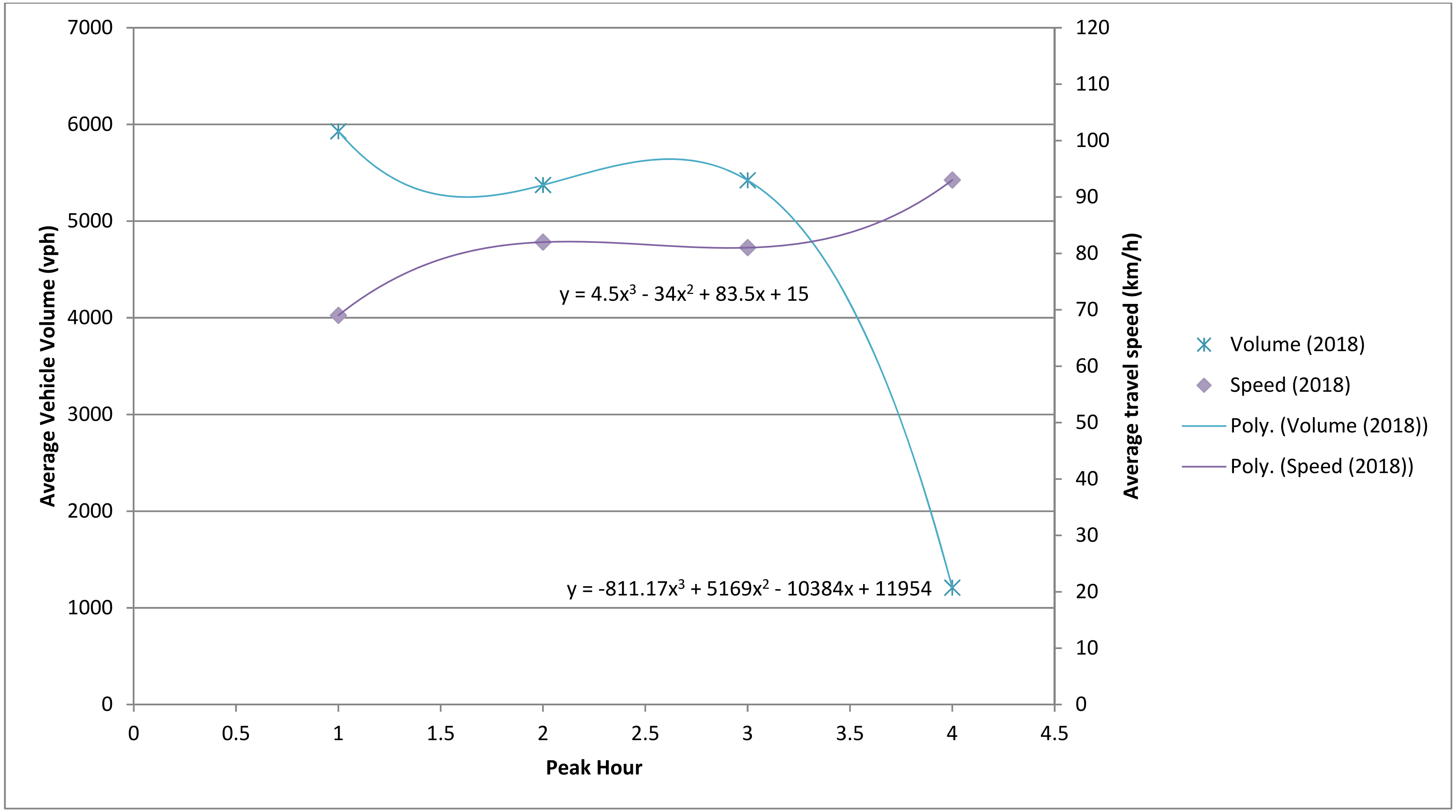
# APPENDIX B

*Speed versus Volume Graphs*

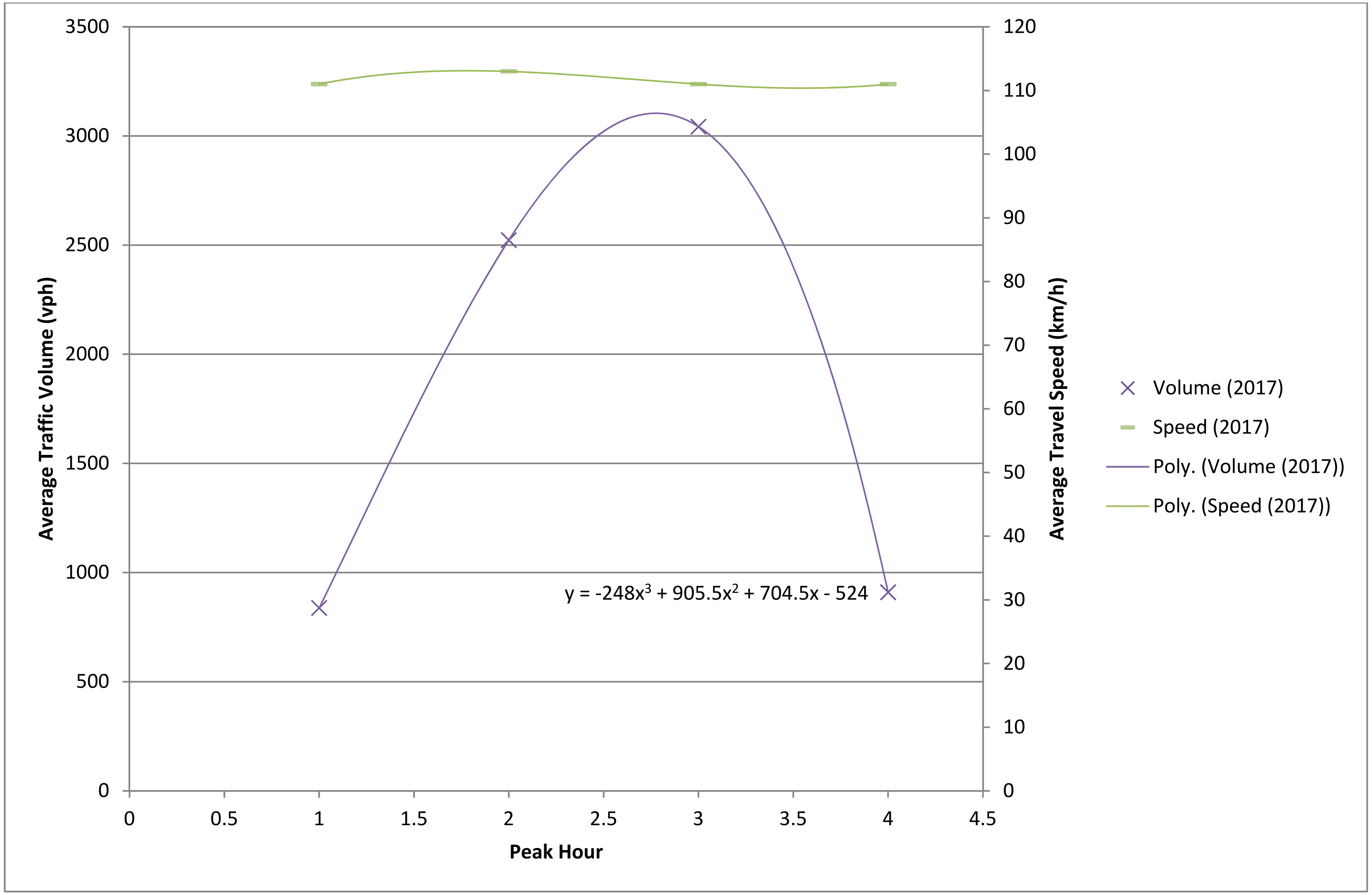
**List Of Graphs**



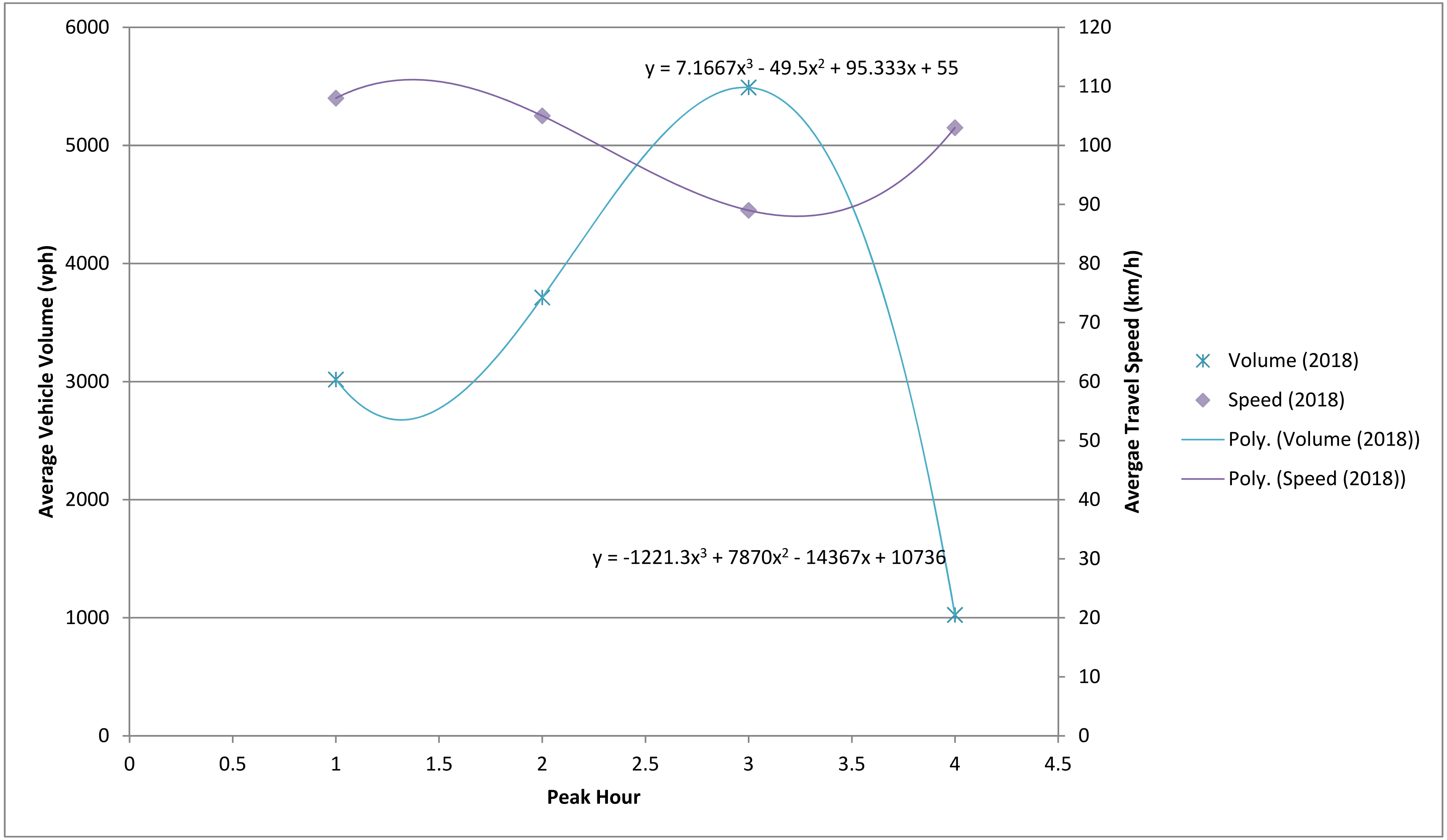

*Graph B 1: Relationship between the average traffic volume and average travel speed on a typical Thursday for CTO station 1945*



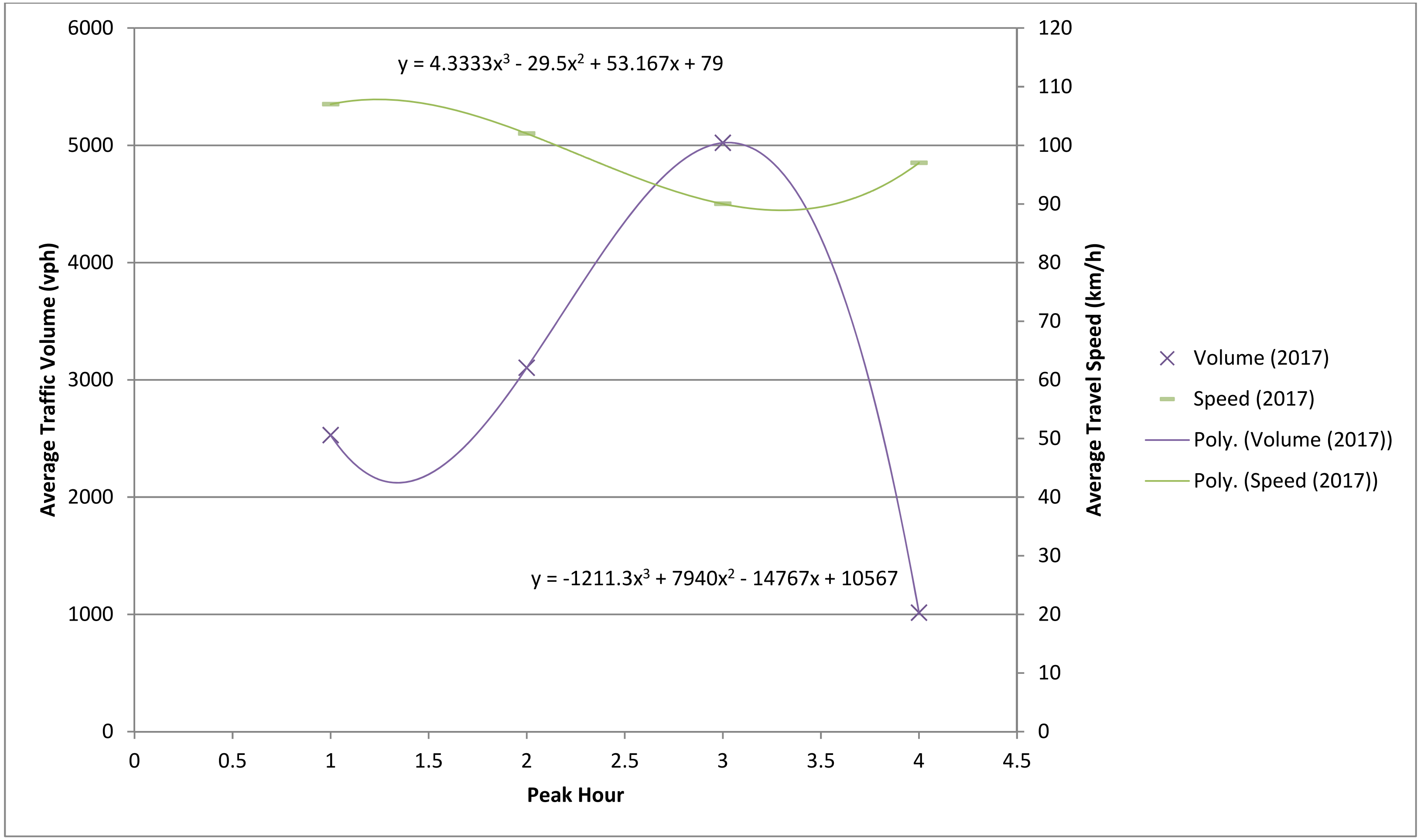

*Graph B 2: Relationship between the average traffic volume and average travel speed on a typical Monday for CTO station 1905*




*Graph B 3: Relationship between the average traffic volume and average travel speed on a Sunday for CTO station 1925*




*Graph B 4: Relationship between the average traffic volume and average travel speed on a typical Tuesday for CTO station 1925*




*Graph B 5: Relationship between the average traffic volume and average travel speed on a typical Wednesday for CTO station 1868*