Identification of the Determinants of Severity for Pedestrian Injuries on Brazilian Federal Highways between 2017 and 2019: A Multinomial Logistic Regression Analysis

English translation of the article published in Scientia Plena, vol. 17, no. 4, 2021

https://doi.org/10.14808/sci.plena.2021.049901

Received: 27 August 2020 | Accepted: 20 April 2021

R. Miranda[1]; W. P. da Silva[1]; S. Dutt-Ross[2]

[1] Universidade Federal Fluminense, 24220-900, Niterói, Rio de Janeiro, Brazil

[2] Universidade Federal do Estado do Rio de Janeiro, 22290-240, Rio de Janeiro, Brazil

*rian_miranda@id.uff.br

ABSTRACT

Walking is one of the main forms of mobility among the Brazilian population, representing 41% of the trips made in the country in 2016. However, the sharing of space between pedestrians and motor vehicles in locations lacking adequate infrastructure has created conflict situations, increasing the risk of injuries and fatalities among pedestrians. This study analysed the determinants of injury severity among pedestrians on Brazilian federal highways between 2017 and 2019. Official data from the Brazilian Federal Highway Police were used to create a multinomial logistic regression model that classified the injury severity of pedestrians involved in highway crashes based on a set of explanatory variables. The results indicate that pedestrians struck by large vehicles, such as trucks, buses, and utility vehicles, have higher odds of death; periods of the day with lower natural light, such as night, dawn, and dusk, are associated with a higher risk of severe injuries and fatalities; urban road segments are less prone to fatalities than those located in rural areas; male pedestrians have higher odds of severe injury or death; and multi-carriageway roads and weekends are associated with higher odds of pedestrian death or severe injury. Finally, an annual upward trend was identified in the odds of severe injury (9.9% per year) and death (3.3% per year) among Brazilian pedestrians, alongside a higher risk of pedestrian injuries in the North, Northeast, and Central-West regions.

Keywords: pedestrian crashes; public safety; multinomial logistic regression.

1 INTRODUCTION

Walking is one of the most popular transportation modes among the Brazilian population, accounting for 41% of the trips made in Brazil in 2016 [1]. Studies show that walking provides important health benefits, such as reducing the risk of cardiovascular and cognitive diseases, as well as reducing cholesterol levels, body fat, and depression [2-6]. For this reason, research groups and public health organisations have encouraged the development of urban planning policies that provide safety and quality for pedestrian movement [7-10]. However, the worldwide growth in trips made by motor vehicles, combined with urban infrastructure that is insufficient for walking trips, has considerably increased the risk of traffic crashes between vehicles and pedestrians [11].

At the global level, road crashes remain a complex problem that public health and safety authorities have actively addressed. According to data from the World Health Organization (WHO), traffic crashes cause approximately 1.35 million fatalities annually, which corresponds to a rate of 18.2 deaths per 100,000 inhabitants [12]. The WHO also estimates that the financial costs associated with emergency and health care for crash victims are around 3% of the GDP in most countries. Studies conducted by the Pan American Health Organization (PAHO) indicate that this rate is lower among countries in the Region of the Americas, with an average of 15.6 deaths per 100,000 inhabitants in 2016 [13]. However, PAHO emphasises that 13 of the 30 surveyed countries registered mortality rates above the regional average, indicating that the different realities observed in each country may have led to discrepancies in their individual rates.

Brazil is currently among the countries with mortality rates above the regional average, with 19.7 deaths per 100,000 inhabitants. According to PAHO (2019) [13], pedestrians are the third road-user group most affected by traffic mortality, accounting for approximately 18.1% of these fatalities. Furthermore, Ladeira et al. (2017) [14] indicate that traffic crashes involving pedestrians are the second main factor responsible for both reducing the lifespan of Brazilians and increasing the number of years lived with physical disabilities. In 2015, pedestrian crashes resulted in a disability-adjusted life-year (DALY) rate of 430.1 years per 100,000 inhabitants [14]. The high number of pedestrians affected by traffic crashes in the country can be explained by the vulnerability of these users in road infrastructures that prioritise motorised transport modes, thereby creating unsafe conditions for walking trips.

It is important to note that vehicle-pedestrian crashes tend to cause more severe injuries on highways, due to their higher speed limits compared with local streets. The WHO estimates that the probability of pedestrian death following a collision with a vehicle travelling at 50 km/h is approximately 80% [15]. On the Brazilian federal highways, where the maximum speed limit for cars is 110 km/h, pedestrian crashes have resulted in substantial numbers of injuries and deaths between 2007 and 2019. During this period, 54,324 collisions between vehicles and pedestrians were recorded by the Brazilian Federal Highway Police (PRF for its acronym in Portuguese) [16]. Out of 54,407 pedestrians involved in these crashes, 71.9% sustained severe or minor injuries, 26.9% died, and only 1.2% were not injured. In addition, 70.6% of these crashes were recorded on urban sections, where the higher traffic and pedestrian flow increase the risk of pedestrian crashes [11].

Milton et al. (2008) [17] emphasise that transportation agencies have sought to understand the relationships between crash-related variables and the physical condition of victims, with the ultimate goal of developing policies for crash prevention. Accordingly, researchers worldwide have adopted logistic regression techniques to model traffic crashes based on their injury outcome [18-21]. This method has produced satisfactory results when applied to road crash data, enabling researchers to build mathematical models using both numeric and categorical variables and the logit link function. By using these models, they can estimate the odds of a given crash severity level (e.g., minor injuries, severe injuries, fatalities, etc) based on the set of explanatory variables associated with that crash event. Traditionally, simple logistic regression is adopted when the response variable is dichotomous, that is, when it has two possible categories. The multinomial logistic model should be used when the response variable has more than two categories [22].

In this context, several studies have used logistic regression to identify statistically significant factors associated with pedestrian crashes and to analyse the contribution of these factors to pedestrian injury severity. For example, Tay et al. (2011) [23] applied multinomial logistic regression to analyse factors associated with pedestrian crashes in South Korea in 2006, considering three severity levels: minor injuries, severe injuries, and fatalities. The study concluded that fatalities and severe injuries were more likely when drivers were male, over 65 years old, or under the influence of alcohol. Furthermore, crash severity increased when trips occurred during the afternoon or night, or when heavy vehicles were involved. Chen and Fan (2019) [19] analysed pedestrian crashes recorded in North Carolina, USA, between 2005 and 2012. In their study, vehicle-pedestrian crashes were classified into five levels: no injury, possible injury, evident injury, disabling injury, and fatality. The authors identified that disabling injuries and fatalities were more likely to occur in crashes involving motorcycles or buses, during dawn or dusk, on rain-wet pavement, and when pedestrians were aged 26 years or older. Finally, Park and Ko (2020) [20] used binary logistic regression to model the physical condition of pedestrians involved in crashes in South Korea between 2015 and 2017. Their study concluded that higher-severity crashes tend to involve large vehicles, take place near schools, and occur at night.

The objective of this study is therefore to analyse the determinants of injury severity among pedestrians on Brazilian federal highways from 2017 to 2019. This analysis is carried out in three stages as follows: 1) generation of a multinomial logistic regression model relating injury severity to variables in the database, such as period of day, vehicle type, gender, and others; 2) verification of the statistical significance of each variable as a predictor of injury severity; and 3) analysis of the odds ratios associated with the various levels of injury severity based on the explanatory variables.

This study is expected to contribute to road safety management activities carried out by highway concessionaires and public safety agencies. The results presented here aim to support competent authorities in adopting preventive measures to reduce both the number and severity of pedestrian crashes on federal highways, such as social-educational campaigns and physical interventions in road infrastructure. Such initiatives are essential for promoting the safety and well-being of drivers and pedestrians, as well as for reducing the potential costs borne by road administrators through compensation paid to victims.

2 MATERIALS AND METHODS

The flowchart in Figure 1 presents the procedures adopted in this study, from the characterisation of the study area to the identification of factors associated with pedestrian crashes on federal highways. All analyses were developed in the R programming language [24].

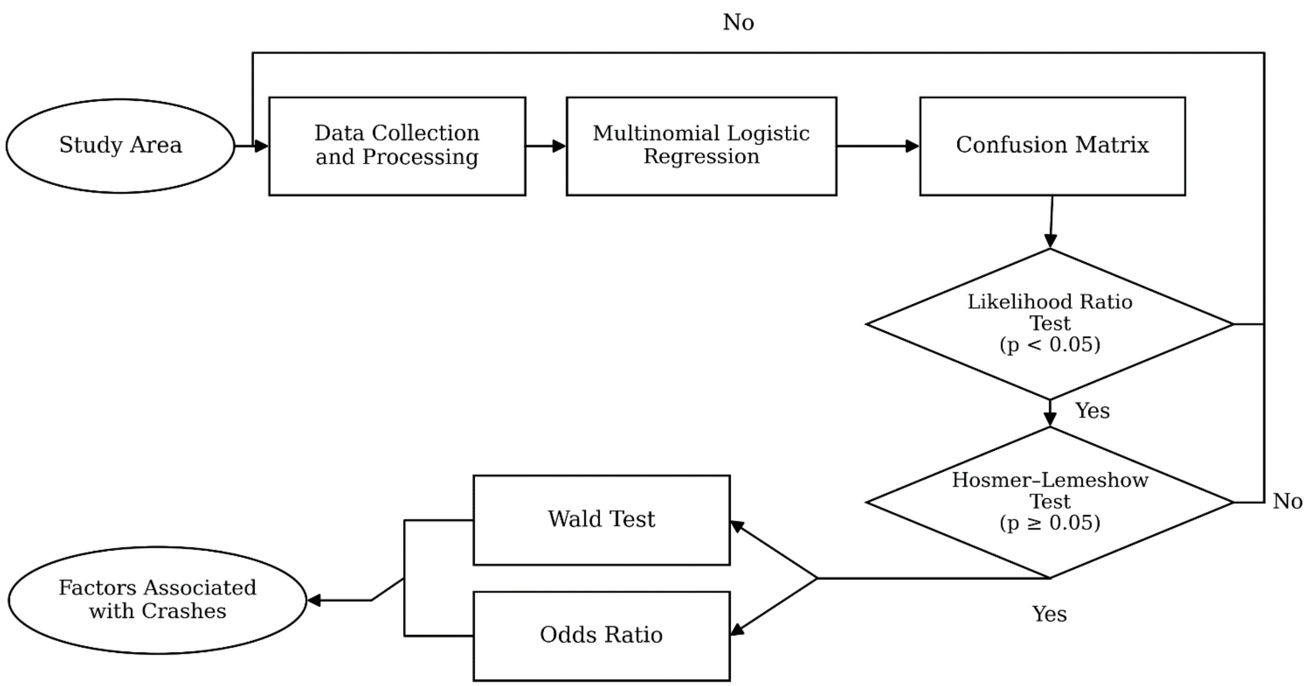

Figure 1 - Study flowchart

2.1 Study area

The study area considered in this work covers the Brazilian territory and its entire federal highway network. According to surveys conducted by the National Confederation of Transport (CNT for its acronym in Portuguese) [25], Brazil has 65,370 km of paved federal highways, of which 6,932 km (10.6%) are dual-carriageway roads, 57,275 km (87.6%) are single-carriageway roads, and 1,163 km (1.8%) are undergoing duplication works. Operating on this infrastructure, the total nationwide vehicle fleet comprised 100,746,553 vehicles in 2018 [26].

Data presented by the Brazilian National Traffic Department (DENATRAN for its acronym in Portuguese) [27] indicate that Brazil had 74,268,555 licensed drivers as of May 2020. The spatial distribution of drivers shows that the Southeast region concentrates the largest number of licensed drivers, with 37,792,667 registered drivers, which is equivalent to 50.9% of the national total. São Paulo leads the state-level statistics, with 23,071,445 drivers (31.1%), followed by Minas Gerais with 7,388,662 (9.9%) and Rio de Janeiro with 5,809,312 (7.8%).

2.2 Data collection and processing

The PRF is officially responsible for collecting traffic crash data whenever a collision occurs on the federal highway network. This collection is conducted through two different instruments: (1) the Electronic Traffic Crash Declaration (e-DAT for its acronym in Portuguese), completed by crash victims in cases where PRF officers identify that no drivers were under the influence of psychoactive substances and the crash did not result in injured victims or damage to society. The e-DAT is completed online through the official PRF website; and (2) Traffic Crash Reports (BAT for its acronym in Portuguese), completed by PRF officers in cases involving injured victims, drivers under the influence of psychoactive substances, or damage to society [28]. These documents record variables associated with each crash, such as date of occurrence, weather condition, period of day, land-use context (rural or urban), road type, number of people involved, their gender and age, their post-crash physical condition, their road-use classification (driver, passenger, or pedestrian), and the types of vehicles involved, among others. The classification of victims' physical condition (i.e., injury severity) is carried out by PRF officers at the crash scene, as shown in Table 1.

Table 1 - Official classification of the physical condition of pedestrian victims on Brazilian federal highways

Physical condition	Description
Not injured	The victim shows no signs or symptoms of injuries resulting from the crash, regardless of whether they were sent to hospital care.
Minor injuries	The victim presents one of the following injuries: (1) pain in a body area without vital organs; (2) bruises, abrasions, and minor lacerations; (3) first-degree burns affecting no more than 10% of body surface; (4) dental fractures; (5) pain in the chest wall, seat-belt abrasion, or abrasion caused by other parts of the vehicle; (6) minor external bleeding; (7) minor sprains and dislocations; (8) headache or dizziness, without loss of consciousness; (9) neck pain; and (10) eye bruising or abrasion.
Severe injuries	The victim does not show signs of death, but their injuries cannot be classified as minor.
Fatality	The victim shows evident signs of death identified at the scene, such as decapitation, crushing of the body, carbonisation, and the presence of *livor mortis* or *rigor mortis*. If the victim dies during transport to medical care, their physical condition is classified as severe injury.

Source: Adapted from PRF (2015) [28].

This information feeds PRF's electronic databases, which are updated annually and then made available for public access on the agency's official website [16]. The databases covering the period between 2017 and 2019 were therefore obtained via open access. In these databases, each record represents an individual involved in a reported crash. Overall, PRF identified 104,526 people involved in traffic crashes during this period, considering both pedestrians and vehicle drivers.

Data processing was then conducted to prepare the records for the construction of the logistic model. This stage began with sampling, in which only records relating to pedestrians involved in crashes were retained in the database. This process resulted in 10,518 pedestrian records, corresponding to 10.1% of the total records. Following this, data completion was assessed to identify variables that had been filled with invalid categories, such as “not informed”, “not identified”, “ignored”, and “invalid”. In these cases, invalid entries were replaced with missing values (represented by NA), which R treats as missing information. The same procedure was adopted for records where pedestrians were reported as being over 100 years old, as suggested by Barroso et al. (2019) [18].

Fourteen potential explanatory variables were then selected for inclusion in the multinomial logistic regression model: day of week (Sunday to Saturday), month (January to December), year (a discrete quantitative variable with values between 2017 and 2019), period of day (dawn, daytime, dusk, and night), road type (single carriageway, dual carriageway, or multiple carriageway), road alignment (straight section or curved section), land use (urban or rural), vehicle type, physical condition (not injured, minor injury, severe injury, or fatality), state, weather condition, vehicle year of manufacture (a discrete quantitative variable), age (a discrete quantitative variable), and gender (female or male).

The construction of logistic regression models requires all adopted variables to be completely filled, with no missing information. For this reason, Milton et al. (2008) [17] recommend excluding all records containing variables with missing data, as a way to enable the analyses to proceed. However, given the high occurrence of invalid entries in the database, it was first necessary to verify the percentage of missing information among the selected variables in order to avoid the exclusion of an excessive amount of records. Consequently, variables with more than 2% missing values (Figure 2), such as weather condition, road alignment, vehicle year of manufacture, and age, were excluded from the analysis. This approach was adopted to limit the number of excluded records and thus ensure a robust amount of data for model development. After these exclusions, 10,348 cases remained in the database, corresponding to 98.4% of the initial sample.

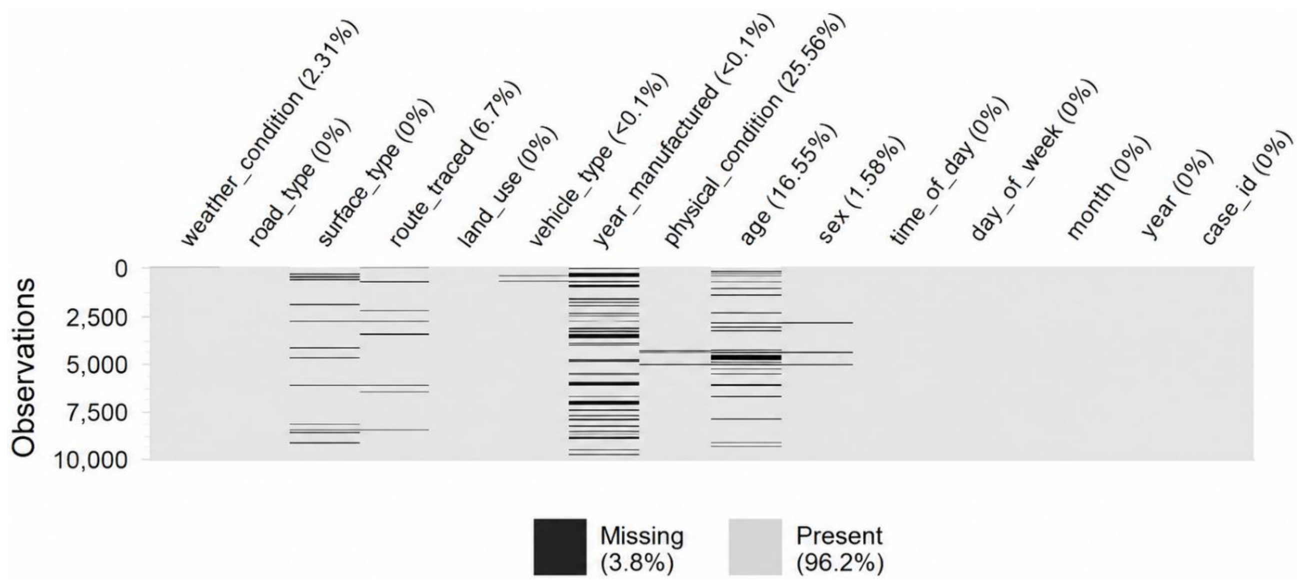

Figure 2 - Pattern of missing data in the records of pedestrians involved in road crashes.

Finally, simplifications were applied to the variables “state”, “month”, “day of week”, “vehicle type”, and “physical condition” in order to reduce the number of categories they contained. The variable “state”, which had 27 categories corresponding to Brazilian states and the Federal District, was converted into the variable “region”, with five categories corresponding to Brazil's macro regions (North, Northeast, South, Southeast, and Central-West). The variable “month”, which had 12 categories, was converted into the variable “season”, with four categories (spring, summer, autumn, and winter). The variable “day of week”, represented by seven categories, was transformed into the variable “day type”, with two categories (weekday and weekend). The variable “vehicle type”, which initially had 17 categories, was reduced to six main categories: car; utility vehicle (covering the utility vehicle, pickup truck, and light truck categories); motorcycle (covering motorcycle and scooter categories); truck (covering the truck, tractor units, and semi-trailer categories); bus (covering the bus and minibus categories); and other (covering the "other", bicycle, moped, trailer, wheeled tractor, and mixed tractor categories). No simplification was applied to the discrete quantitative variable "year", which therefore had three possible values (2017, 2018, and 2019).

For the variable "physical condition", which initially had four categories, the relative frequency of the "not injured" category (1.4%) was considerably lower than those observed for the other categories (minor injuries: 35.7%; severe injuries: 37.0%; fatality: 25.9%). According to Elvik and Mysen (1999) [29] and Hakkert and Hauer (1988) [30], crashes of lower severity tend to have higher levels of underreporting, which explains the low frequency of not injured pedestrians observed in the sample. This result suggests that the "not injured" category can be classified as a rare event, making the logistic regression model more likely to underestimate its probability of occurrence [31]. In addition, King and Zeng (2001) [31] indicate that the presence of rare events in the sample tends to produce biased estimates of logistic regression coefficients. To address these problems, records of not injured pedestrians were excluded from the analysis, resulting in a final sample of 10,205 cases. Finally, "physical condition" was renamed into "injury severity".

2.3 Multinomial logistic regression

Multinomial logistic regression was applied to fit a mathematical model describing pedestrian injury severity based on the explanatory variables in the database. The response variable therefore had three categories: (1) minor injuries; (2) severe injuries; and (3) fatality. Among these, the "minor injuries" category was adopted as the reference category. The multinomial logistic regression model is expressed by the logit function shown in Equation 1 [22].

$$g_k(x) = Ln\left[\frac{\Pr(k|x)}{\Pr(0|x)}\right] = \beta_{k0} + \beta_{k1}x_1 + \ldots + \beta_{kp}x_p \qquad (1)$$

$$\forall (k,i)/k \in \{0,1,2,\ldots,K\} \cap i \in \{0,1,2,\ldots,p\}$$

Where: k: injury severity level sustained by a given pedestrian;
K: number of categories of the "injury severity" variable;
p: total number of degrees of freedom of the explanatory variables;
x: vector of the explanatory variables;
Pr(k|x): conditional probability of injury severity k given the vector x;
Pr(0|x): conditional probability of the reference injury severity given the vector x;
β_{ki}: model parameter for the injury severity level k associated with the explanatory variable xi.

Initially, the database was divided into training and test samples following a 3:1 ratio. Thus, 75% of the data was used to develop the multinomial logistic regression model, while 25% was reserved to test its accuracy. Then, the model was built and evaluated across six procedures: (1) stepwise method; (2) confusion matrix; (3) likelihood ratio test; (4) Hosmer-Lemeshow test; (5) Wald test; and (6) odds ratio analysis. A 5% significance level was adopted for the statistical tests.

Following this workflow, the independent variables were first selected using the stepwise method. This method consists of adding variables to the model in consecutive steps. At each step, the model fit was evaluated using the Akaike Information Criterion (AIC). According to Alvarenga (2015) [32], logistic regression models present better fit when their AIC values are lower. Consequently, the stopping criterion of the stepwise method was satisfied when the selected model presented the minimum AIC value in comparison with models estimated in previous steps [33, 34]. In this context, adding a new variable would increase the AIC of the resulting model, thereby reducing goodness of fit. Equation 2 presents the AIC calculation formula:

$$AIC = -2[LL(\beta_U) - p] \tag{2}$$

Where: $LL(\beta_U)$: log-likelihood of the estimated model;
p: number of model parameters.

After variable selection and model construction, the test sample was used to generate a 3 x 3 confusion matrix, which compared the observed injury severity values in each record with those predicted by the regression model. Accuracy was then measured as the ratio between the number of times the model correctly predicted pedestrian injury severity and the total number of predictions made, as shown by Equation 3 [35].

$$Accuracy = \frac{a_{11} + a_{22} + a_{33}}{\sum_{i=1}^{k=3} \sum_{j=1}^{k=3} a_{ij}} \tag{3}$$

Where: a_{11}: number of correct predictions in the "minor injuries" category;
a_{22}: number of correct predictions in the "severe injuries" category;
a_{33}: number of correct predictions in the "fatality" category;
a_{ij}: number of predictions in row i and column j.

Next, the likelihood ratio and Hosmer-Lemeshow tests were adopted to assess model quality. The likelihood ratio test assessed the statistical significance of the previously selected set of variables by comparing the log-likelihood of the model proposed by the stepwise method with the log-likelihood of the null model. The null hypothesis of this test states that the selected variables do not significantly improve the model fit; therefore, these variables can only be accepted if the test rejects the null-hypothesis (p-value < 0.05), which indicates a significant improvement of the model when they are present [22, 36]. In turn, the Hosmer-Lemeshow test was used to assess the goodness of fit of the model in relation to the sample data. The null hypothesis of this test states that the model adequately fits the sample data [22, 37]. Therefore, when the test presents a p-value greater than or equal to 0.05, it is assumed that the model presents satisfactory fit. The test statistics adopted in the likelihood ratio and Hosmer-Lemeshow tests are shown in Equations 4 and 5, respectively.

$$G = -2 \times [LL(\beta_R) - LL(\beta_U)] \tag{4}$$

Where: $LL(\beta_R)$: log-likelihood of the null model;

$LL(\beta_U)$: log-likelihood of the model composed of the selected variables.

$$\chi^2 = \sum_{j=1}^{J} \sum_{k=0}^{K-1} \frac{(O_{jk} - E_{jk})^2}{E_{jk}} \tag{5}$$

Where: J: number of groups into which the sample is divided;
O_{jk}: observed frequency for severity k in group j;
E_{jk}: expected frequency for severity k in group j.

The Wald test was then performed to assess the statistical significance of the estimated β_{ki} parameters associated with the model variables. The null hypothesis of the test states that parameter β_{ki} is equal to zero. Therefore, a variable is considered statistically significant when the p-value is lower than 0.05. The Wald test statistic is presented by Equation 6.

$$W_{ki} = \frac{\hat{\beta}_{ki}}{SE(\hat{\beta}_{ki})} \tag{6}$$

Where: $\hat{\beta}_{ki}$: estimated parameter for injury severity level k and explanatory variable xi;
$SE(\hat{\beta}_{ki})$: standard error of β_{ki}.

Finally, an odds ratio analysis was conducted to identify the relationship between the values of the explanatory variables in the vector x and the odds of a pedestrian sustaining a given injury severity level k. This procedure was applied individually to all categories of explanatory variables that presented statistical significance, in order to identify their influence on the response variable [22]. The odds ratio and its 95% confidence interval were calculated using Equations 7 and 8, respectively.

$$OR_k(i,0) = \frac{\Pr(k \mid x = i) \,/\, \Pr(0 \mid x = i)}{\Pr(k \mid x = 0) \,/\, \Pr(0 \mid x = 0)} = \exp(\hat{\beta}_{ki}) \tag{7}$$

$$IC95\% = OR_k(i,0) \pm 1{,}96\, SE(\hat{\beta}_{ki}) \tag{8}$$

Where: $OR_k(i,0)$: odds ratio of injury severity k given the occurrence of category i of the explanatory variable;
k: pedestrian injury severity;
x = i: category analysed for the explanatory variable;
x = 0: reference category of the explanatory variable.

3 RESULTS

The stepwise method produced a multinomial logistic regression model composed of the variables “vehicle type”, “period of day”, “land use”, “gender”, “road type”, “day type”, “region”, and “year”, thereby excluding the variable “season”. The steps performed and their respective AIC values are presented in Table 2. The confusion matrix, shown in Table 3, indicated an accuracy of 48% for the model predictions, with 52.8% correct classifications for minor-injury cases, 37.7% for severe-injury cases, and 56.7% for fatality cases. The model was then validated using the likelihood ratio test (p-value < 0.001) and the Hosmer-Lemeshow test (p-value = 0.252). These results indicate not only that the selected set of variables is suitable for describing pedestrian injury severity, but also that the model resulting from the stepwise method provides satisfactory goodness of fit to the sample data.

Table 2 - Steps of the stepwise method for variable selection using the training sample

Steps	Model	AIC
1	injury_severity ~ 1	16654.92
2	injury_severity ~ vehicle_type	15817.56
3	injury_severity ~ vehicle_type + period_of_day	15492.77
4	injury_severity ~ vehicle_type + period_of_day + land_use	15364.12
5	injury_severity ~ vehicle_type + period_of_day + land_use + region	15310.47
6	injury_severity ~ vehicle_type + period_of_day + land_use + region + gender	15275.15
7	injury_severity ~ vehicle_type + period_of_day + land_use + region + gender + road_type	15255.65
8	injury_severity ~ vehicle_type + period_of_day + land_use + region + gender + road_type + day_type	15249.79
9	injury_severity ~ vehicle_type + period_of_day + land_use + region + gender + road_type + day_type + year	15245.23
10	injury_severity ~ vehicle_type + period_of_day + land_use + region + gender + road_type + day_type + year + season	15253.23

Table 3 - Confusion matrix for assessing model accuracy

Observed Values (Test Sample)	**Predicted Values (Model)**		
	Minor injuries	**Severe injuries**	**Fatality**
Minor injuries	492	283	156
Severe injuries	387	366	219
Fatality	86	195	368

Subsequently, the Wald test was performed. The results revealed that severe injuries are not associated with factors such as utility vehicles (p-value = 0.608), single-carriageway roads (p-value = 0.902), the North region (p-value = 0.740), and the South region (p-value = 0.232). For pedestrian crashes featuring trucks, the odds of pedestrians sustaining severe injuries were 72.9% higher than in cases involving cars. Vehicles classified as "other" had 37.9% higher odds of causing severe pedestrian injuries. In contrast, crashes involving buses reduced the odds of severe injury by 6.9%, while crashes involving motorcycles reduced them by 13.7%. Dawn, dusk, and night presented higher odds of severe injury than daytime, with increases of 13.0%, 38.6%, and 29.8%, respectively. The odds of severe injuries on urban sections were 10.8% lower than those observed in rural areas. Male pedestrians had 34.1% higher odds of severe injury than female pedestrians. The regional analysis indicated higher odds of severe injury in the Northeast region (32.0% higher) compared with the Southeast region. Pedestrians in the Central-West region, however, had lower odds of severe injury than pedestrians in the Southeast region (18.9% lower). In addition, pedestrians were found to have 18.5% higher odds of severe injury on multi-carriageway road sections than on dual-carriageway road sections, as well as 15.1% higher odds of severe injury on weekends than on weekdays. Finally, an annual increase of 9.9% in the odds of severe injury was observed. Table 4 presents the parameter estimates of each variable in the model, as well as the results of the Wald test and odds ratio analysis of severe pedestrian injuries.

Table 4 - Multinomial logistic regression model for the response category "Severe injury"

Explanatory variable	**β**	**Odds ratio (OR)**	**95% CI (OR)**	**P-value**
(Intercept)	-190.969			<0.001
Vehicle type (ref: Car)				
Truck	0.548	1.729	(1.603; 1.865)	<0.001
Motorcycle	-0.147	0.863	(0.776; 0.960)	0.007
Bus	-0.071	0.931	(0.883; 0.982)	0.009
Other	0.321	1.379	(1.252; 1.519)	<0.001
Utility vehicle	0.025	1.025	(0.932; 1.129)	0.608
Period of day (ref: Daytime)				
Dawn	0.122	1.130	(1.112; 1.149)	<0.001
Dusk	0.326	1.386	(1.353; 1.419)	<0.001
Night	0.261	1.298	(1.171; 1.439)	<0.001
Land use (ref: Rural)				
Urban	-0.114	0.892	(0.805; 0.990)	0.031
Gender (ref: Female)				
Male	0.293	1.341	(1.213; 1.483)	<0.001
Region (ref: Southeast)				
South	0.060	1.062	(0.962; 1.171)	0.232
Northeast	0.278	1.320	(1.200; 1.451)	<0.001
North	0.013	1.013	(0.938; 1.095)	0.740
Central-West	-0.209	0.811	(0.733; 0.898)	<0.001
Road type (ref: Dual carriageway)				
Multiple carriageway	0.170	1.185	(1.050; 1.339)	0.006
Single carriageway	0.006	1.006	(0.909; 1.115)	0.902
Day type (ref: Weekday)				
Weekend	0.141	1.151	(1.026; 1.291)	0.017
Year	0.094	1.099	(1.099; 1.099)	<0.001

The Wald test further revealed that, among all variables in the model, only the South region was not associated with pedestrian deaths (p-value = 0.066). Table 5 shows that pedestrians struck by buses had 52.0% higher odds of death than those involved in collisions with cars. The odds of death were even higher in cases comprising utility vehicles, with 54.3% higher odds of death compared with crashes featuring cars. However, the highest odds of death were recorded in pedestrian crashes caused by trucks and by vehicles classified as "other", with 264.1% and 163.2% higher odds of death, respectively. Crashes with motorcycles, by contrast, had 79.6% lower odds of resulting in pedestrian death. Furthermore, dusk increased the odds of pedestrian death by 87.2% compared with daytime. This increase was even greater at dawn (240.5%) and at night (227.0%). Urban sections represented lower risk for pedestrians, with 53.4% lower odds of death than rural areas. The regional analysis indicated that the odds of pedestrian death in the North region were 23.7% higher than in the Southeast region, while in the Central-West and Northeast regions the odds were 26.8% and 89.0% higher, respectively. Crashes occurring on single-carriageway roads reduced the odds of death by 24.7% compared with cases recorded on dual carriageways, while crashes on multi-carriageway roads increased the odds of death by 14.9%. Male pedestrians had 51.0% higher odds of death than female pedestrians, while weekends increased the odds of pedestrian death by 22.4% relative to weekdays. The results also indicated an annual increase of 3.3% in the odds of pedestrian death.

Table 5 - Multinomial logistic regression model for the response category "Fatality"

Explanatory variable	**β**	**Odds ratio (OR)**	**95% CI (OR)**	**P-value**
(Intercept)	-66.809			<0.001
Vehicle type (ref: Car)				
Truck	1.292	3.641	(3.364; 3.940)	<0.001
Motorcycle	-1.588	0.204	(0.196; 0.213)	<0.001
Bus	0.419	1.520	(1.446; 1.597)	<0.001
Other	0.968	2.632	(2.366; 2.927)	<0.001
Utility vehicle	0.434	1.543	(1.413; 1.686)	<0.001
Period of day (ref: Daytime)				
Dawn	1.225	3.405	(3.311; 3.501)	<0.001
Dusk	0.627	1.872	(1.832; 1.914)	<0.001
Night	1.185	3.270	(2.896; 3.692)	<0.001
Land use (ref: Rural)				
Urban	-0.765	0.466	(0.418; 0.519)	<0.001
Gender (ref: Female)				
Male	0.412	1.510	(1.361; 1.676)	<0.001
Region (ref: Southeast)				
South	0.095	1.100	(0.994; 1.218)	0.066
Northeast	0.637	1.890	(1.721; 2.076)	<0.001
North	0.213	1.237	(1.180; 1.297)	<0.001
Central-West	0.237	1.268	(1.174; 1.368)	<0.001
Road type (ref: Single carriageway)				
Multiple carriageway	0.139	1.149	(1.059; 1.247)	0.001
Single carriageway	-0.284	0.753	(0.672; 0.844)	<0.001
Day type (ref: Weekday)				
Weekend	0.202	1.224	(1.073; 1.397)	<0.001
Year	0.033	1.033	(1.033; 1.033)	<0.001

4 DISCUSSION AND CONCLUSION

This study analysed the determinants of injury severity among pedestrians struck by vehicles on Brazilian federal highways between 2017 and 2019. The PRF traffic crash database indicates that 10,518 pedestrians were victims of road crashes during this period, of which 10,205 composed the dataset used in the analyses. Using the training sample (75% of the data), a multinomial logistic regression model was developed, revealing that pedestrian injury severity is directly related to factors such as the type of vehicle involved in the crash, period of day, land use, region where the crash was recorded, pedestrian gender, road type, day type (weekday or weekend), and year of occurrence. According to the confusion matrix derived from the test sample (25% of the data), the model achieved an accuracy of 48%.

The results show a significant annual increase in the risk of severe injuries and deaths during the period between 2017 and 2019. The odds of pedestrian death in the North, Northeast, and Central-West regions were higher than in the Southeast region, while the Northeast region also presented greater propensity for recording severely injured pedestrians. This may be explained by the large number of pedestrians and drivers sharing the road in these regions. For example, the Secretariat of Infrastructure of the State of Bahia [38] identified pedestrian crashes as the fourth leading crash type recorded in the metropolitan region of Salvador, which includes sections of the BR-324 and BR-420 highways. The study also indicated that walking accounted for the largest share of trips among people in Bahia (35.3%), while buses and cars were the second and third most used transport modes, respectively (45% in total). This context favours risky situations for pedestrians on federal highways because of the high traveling speeds and the scarcity of crossing points that correspond to pedestrian desire lines.

The results also revealed that large vehicles, such as trucks, buses, and utility vehicles, tend to cause pedestrian fatalities. In addition, pedestrians hit by trucks are more likely to sustain severe injuries than those involved in crashes with cars. These findings are consistent with the study conducted by Roudsari et al. (2004) [39], which identifies trucks as a risk factor for pedestrian physical integrity, increasing the odds of severe injury by up to 210% and death by up to 240%. Vehicles classified as “other” also presented elevated odds of severe crashes and fatalities, which may be explained by the presence of heavy vehicles in this category, such as wheeled tractors, mixed tractors, and trailers. Collisions with motorcycles, by contrast, represented lower risk for pedestrians, with odds ratios below 1 for severe injury and death. This result indicates that motorcycles tend to cause fewer severe injuries or fatalities to pedestrians than cars, as previously observed by Tay et al. (2011) [23] and Wang et al. (2013) [40].

The odds ratio analysis for the period of day indicated that severe injuries and fatalities tend to occur at dawn, dusk, and night. Similar results were obtained by Tay et al. (2011) [23] and Wang et al. (2013) [40], who identified periods of lower natural light as aggravating factors for pedestrian injury severity. This situation can be explained by the fact that, during the daytime, drivers have better visibility of crossing points, making it easier both to reduce speed and to perform safety manoeuvres to avoid collisions with pedestrians. Thus, even when a crash is unavoidable, timely driver action may reduce impact severity. By contrast, other periods of the day have reduced natural lighting, making it more difficult for drivers to perceive pedestrians on the road, thereby increasing the odds of severe or fatal crashes .

Regarding land use, fatal crashes and severe injuries had higher odds of occurring in rural areas. These results are consistent with the findings of Tay et al. (2011) [23] and Chen and Fan (2019) [19], who also observed higher odds of severe crashes in rural areas. Urban sections usually have more traffic control devices, such as pedestrian crossings and pedestrian-crossing signs, as well as speed-reduction devices, such as raised crossings, speed humps, and speed cameras. These features improve pedestrian crossing safety, thereby reducing the odds of severe injuries and deaths. Rural areas, by contrast, favour higher travelling speeds, increasing fatality risk for pedestrians, especially in locations without footbridges. Consequently, urban sections present 53.4% lower odds of pedestrian death than rural areas.

Additionally, male pedestrians had higher odds of severe and fatal injuries than female pedestrians. As indicated by Ferenchak (2016) [41], the behavioural pattern observed among male pedestrians involves greater risk than that observed among female pedestrians. In general, men have lower tolerance for waiting time before crossing and use pedestrian crossings less often than women [41]. This behaviour not only increases crash risk among men in urban sections, but may also result in more frequent severe injuries and deaths.

The odds of severe injuries and fatalities were also higher on multi-carriageway roads, while single-carriageway sections presented lower pedestrian fatality risk. The presence of central medians on multi and dual carriageways, combined with the wider cross-sections of these road types, are possible explanations for this phenomenon. As shown in previous studies, wider roads not only impose longer crossing distances on pedestrians, but also increase drivers' sense of safety, making them more likely to travel at higher speeds [21, 23]. This situation is further aggravated for pedestrians who naturally have lower walking speeds, such as older adults, children, and those with physical disabilities. In addition, Tulu et al. (2013) [42] indicate that, in developing countries, pedestrians tend to climb over central barriers on multi and dual-carriageway roads due to the scarcity of footbridges at their desired crossing locations. As a result, pedestrians become even more susceptible to fatalities or severe injuries resulting from collisions with high-speed vehicles.

Finally, the results indicate that weekend crashes were more likely to cause pedestrian deaths and severe injuries. These results are consistent with the study by Alvarenga (2015) [32], which identified greater risk of severe crashes on Saturdays and Sundays. Adanu et al. (2018) [43] suggest that the risk of severe injury among drivers during weekends is associated with human factors, such as driving fatigue and susceptibility to distraction. Although those authors did not focus on pedestrian injury severity, the behavioural tendency presented by drivers during weekends is a possible explanation for the increase in the odds of pedestrian deaths on Saturdays and Sundays.

It can therefore be concluded that the factors representing the greatest risk to pedestrian physical integrity on Brazilian federal highways are large vehicles, such as trucks, buses, and utility vehicles; periods of the day with lower natural light, such as dawn, dusk, and night; rural sections; male gender; multi-carriageway road sections; weekends; and the North, Northeast, and Central-West regions. Moreover, the study revealed that the odds of pedestrians being diagnosed with severe injuries on federal highways have been increasing by 9.9% per year. The odds of fatality have been increasing by 3.3% per year, indicating that the high mortality rates observed by PAHO tend to increase if measures to prevent pedestrian crashes are not adopted [13].

These results suggest the need for infrastructural and operational road-planning interventions to protect pedestrians from possible collisions with motor vehicles. The interventions with the greatest potential to reduce the number of cases involving severe injuries or fatalities are the construction of elevated crossings in accordance with pedestrian desire lines in rural areas; the installation of road lighting on sections with high risk of pedestrian crashes between dusk and dawn; and the use of signage and speed-reduction devices in areas with high pedestrian movement. There is also a need for social-educational campaigns directed at drivers of large vehicles, male pedestrians, and weekend travellers. These campaigns should alert road users to the risks of shared use of federal highways by pedestrians and motor vehicles; to pedestrian vulnerability in relation to heavy vehicles; to the higher likelihood of severe pedestrian crash outcomes during periods of lower natural light; to the importance of crossing at safe locations; and to the need for drivers to pay greater attention during weekends and nighttime periods. These measures should be adopted primarily in the North, Northeast, and Central-West regions, given the higher odds of pedestrian fatality in these areas.

Finally, future research should prioritise the identification of pedestrian crash hotspots in these regions, in order to support decision-making by road administrators regarding the exact locations where infrastructure interventions should be implemented. These analyses can be carried out using kernel density maps or through the methodologies presented in the PARE Program of the Ministry of Transportation [44]. In addition, future studies should test other logistic regression techniques, such as ordinal, binomial, and mixed logistic regression models, with the aim of comparing them with the multinomial logistic regression used in this study. Differences between techniques may be expressed in terms of model accuracy, significant variables, coefficient estimates, and odds ratios. Finally, it is recommended that future studies adopt data imputation techniques to handle the incomplete or invalid entries observed in the PRF database. This procedure could allow more variables to be included in the model, such as pedestrian age and weather condition, enabling new conclusions about the influence of these factors on pedestrian injury severity.

5 ACKNOWLEDGEMENTS

This study was supported by the Coordination for the Improvement of Higher Education Personnel - Brazil (CAPES) - Finance Code 001.

6 REFERENCES

1. National Association of Public Transport (ANTP). Urban Mobility Information System of the National Association of Public Transport - Simob/ANTP: 2016 general report. São Paulo (SP): ANTP; 2016. Available at: http://files.antp.org.br/simob/simob-2016-v6.pdf

2. Abbott RD, White LR, Ross GW, Masaki KH, Curb JD, Petrovitch H. Walking and dementia in physically capable elderly men. JAMA. 2004 Sep;292(12):1447-1453. doi: 10.1001/jama.292.12.1447

3. Hamer M, Chida Y. Active commuting and cardiovascular risk: a meta-analytic review. Preventive Medicine. 2008 Jan;46(1):9-13. doi: 10.1016/j.ypmed.2007.03.006

4. Hanson S, Jones A, Petrovitch H. Is there evidence that walking groups have health benefits? A systematic review and meta-analysis. British Journal of Sports Medicine. 2015 Jan;49(11):710-715. doi: 10.1136/bjsports-2014-094157

5. Tainio M, Nazelle AJ, Götschi T, Kahlmeier S, Rojas-Rueda D, Nieuwenhuijsen MJ, Sá TH, Kelly P, Woodcock J. Can air pollution negate the health benefits of cycling and walking? Preventive Medicine. 2016 Jun;87(1):233-236. doi: 10.1016/j.ypmed.2016.02.002

6. Shih YH, Pai MC, Lin HS, Sung PS, Wang JJ. Effects of walking on sundown syndrome in community-dwelling people with Alzheimer's disease. International Journal of Older People Nursing. 2020 Jun;87(1):e12292. doi: 10.1111/opn.12292

7. Sallis JF, Cervero RB, Ascher W, Henderson KA, Kraft MK, Kerr J. An ecological approach to creating active living communities. Annual Review of Public Health. 2006;27(1):297-322. doi: 10.1146/annurev.publhealth.27.021405.102100

8. World Health Organization (WHO). Global recommendations on physical activity for health. Geneva (CH): WHO; 2011. Available at: https://apps.who.int/iris/bitstream/handle/10665/44399/9789241599979_eng.pdf?sequence=1

9. Panter J, Heinen E, Mackett R, Ogilvie D. Impact of new transport infrastructure on walking, cycling, and physical activity. American Journal of Preventive Medicine. 2016 Feb;50(2):e45-e53. doi: 10.1016/j.amepre.2015.09.021

10. Moura F, Cambra P, Gonçalves AB. Measuring walkability for distinct pedestrian groups with a participatory assessment method: A case study in Lisbon. Landscape and Urban Planning. 2017 Jan;157(1):282-296. doi: 10.1016/j.landurbplan.2016.07.002

11. Zegeer CV, Bushell M. Pedestrian crash trends and potential countermeasures from around the world. Accident Analysis & Prevention. 2012 Jan;44(1):3-11. doi: 10.1016/j.aap.2010.12.007

12. World Health Organization (WHO). Global status report on road safety 2018. Geneva (CH): WHO; 2018. Available at: https://www.who.int/publications/i/item/global-status-report-on-road-safety-2018

13. Pan American Health Organization (PAHO). Status of Road Safety in the Region of the Americas. Washington (DC): PAHO; 2019. Available at: https://iris.paho.org/handle/10665.2/51088

14. Ladeira RM, Malta DC, Morais Neto OLD, Montenegro MDMS, Soares Filho AM, Vasconcelos CH, Mooney M, Naghavi M. Road traffic accidents: Global Burden of Disease study, Brazil and federated units, 1990 and 2015. Brazilian Journal of Epidemiology. 2017 May;20(1):157-170. doi: 10.1590/1980-5497201700050013

15. World Health Organization (WHO). Speed management: a road safety manual. Geneva (CH): WHO; 2017. Available at: https://montevideo.gub.uy/sites/default/files/control_de_velocidad_1.pdf

16. Federal Highway Police (PRF). Open data: crashes. 2019. Available at: https://www.gov.br/prf/pt-br/acesso-a-informacao/dados-abertos/dados-abertos-da-prf

17. Milton JC, Shankar VN, Mannering FL. Highway accident severities and the mixed logit model: an exploratory empirical analysis. Accident Analysis & Prevention. 2008 Jan;40(1):260-266. doi: 10.1016/j.aap.2007.06.006

18. Barroso Junior GT, Bertho ACS, Veiga ADC. Lethality of traffic accidents on Brazilian federal highways in 2016. Brazilian Journal of Population Studies. 2019;36(1):1-22. doi: 10.20947/S0102-3098a0074

19. Chen Z, Fan WD. A multinomial logit model of pedestrian-vehicle crash severity in North Carolina. International Journal of Transportation Science and Technology. 2019 Mar;8(1):43-52. doi: 10.1016/j.ijtst.2018.10.001

20. Park S, Ko D. A multilevel model approach for investigating individual accident characteristics and neighborhood environment characteristics affecting pedestrian-vehicle crashes. International Journal of Environmental Research and Public Health. 2020 Apr;17(1):3107. doi: 10.3390/ijerph17093107

21. Sanz NC, Guirao B, Attard M. Analysis of the risk factors affecting the severity of traffic accidents on Spanish crosstown roads: The driver's perspective. Sustainability. 2020 Mar;12(6):2237. doi: 10.3390/su12062237

22. Hosmer DW, Lemeshow S, Sturdivant RX. Applied logistic regression. 3rd ed. Hoboken (NJ): Wiley & Sons, Inc.; 2013.

23. Tay R, Choi J, Kattan L, Khan A. A multinomial logit model of pedestrian-vehicle crash severity. International Journal of Sustainable Transportation. 2011 Mar;5(4):233-249. doi: 10.1080/15568318.2010.497547

24. R Core Team. R: A language and environment for statistical computing. Vienna (AT): R Foundation for Statistical Computing; 2020. Available at: https://www.R-project.org/

25. National Confederation of Transport (CNT). CNT highway survey 2019. Brasília (DF): CNT: SEST SENAT; 2019. Available at: https://pesquisarodovias.cnt.org.br/downloads/ultimaversao/gerencial.pdf

26. Brazilian Institute of Geography and Statistics (IBGE). Vehicle fleet 2018. 2018. Available at: https://cidades.ibge.gov.br/brasil/pesquisa/22/28120?ano=2018

27. Brazilian National Traffic Department (DENATRAN). Number of licensed drivers - DENATRAN: May 2020. 2020. Available at: https://www.gov.br/infraestrutura/pt-br/assuntos/transito/conteudo-denatran/estatisticas-quantidade-de-habilitados-denatran1790

28. Federal Highway Police (PRF). MP-015 Crash response. 2nd ed. Brasília (DF): PRF; 2015.

29. Elvik R, Mysen AB. Incomplete accident reporting meta-analysis of studies made in 13 countries. Transportation Research Record. 1999 Jan;1665(1):133-140. doi: 10.3141/1665-18

30. Hauer E, Hakkert AS. The extent and some implications of incomplete accident reporting. Transportation Research Record. 1988 Jan;1185(1):1-10. Available at: https://www.researchgate.net/publication/243764350_The_Extent_and_Some_Implications_of_Incomplete_Accident_Reporting

31. King G, Zeng L. Logistic regression in rare events data. Political Analysis. 2001;9(2):137-163. doi: 10.1093/oxfordjournals.pan.a004868

32. Alvarenga AMT. Generalised linear models: application to road crash data [dissertation]. Lisbon (PT): University of Lisbon; 2015.

33. Burnham KP, Anderson DR. Multimodel inference: Understanding AIC and BIC in model selection. Sociological Methods & Research. 2004 Nov;33(2):261-304. doi: 10.1177/0049124104268644

34. Hastie T, Tibshirane R, Friedman J. The elements of statistical learning. 2nd ed. Stanford (CA): Springer; 2017.

35. Visa S, Ramsay B, Ralescu A, van der Knaap E. Confusion matrix-based feature selection. In: Visa S, Inoue A, Ralescu A, editors. Proceedings of the Twenty-second Midwest Artificial Intelligence and Cognitive Science Conference; 2011 Apr 16-17; Cincinnati (OH). Madison (WI): Omnipress; c2011. p. 120-127. Available at: http://ceur-ws.org/Vol-710/MAICS2011Proceedings.pdf

36. Washington SP, Karlaftis MG, Mannering FL. Methods for transportation analysis. 2nd ed. London: Chapman and Hall/CRC; 2003.

37. Fagerland MW, Hosmer DW, Bofin AM. Multinomial goodness-of-fit tests for logistic regression models. Statistics in Medicine. 2008 Jan;27(21):4238-4253. doi: 10.1002/sim.3202

38. Secretaria de Infraestrutura da Bahia (SEINFRA). Mobility survey - Salvador metropolitan region. Salvador (BA): SEINFRA; 2012.

39. Roudsari BS, Mock CN, Kaufman R, Grossman D, Henary BY, Crandall J. Pedestrian crashes: higher injury severity and mortality rate for light truck vehicles compared with passenger vehicles. Injury Prevention. 2004 Jun;10(3):154-158. doi: 10.1136/ip.2003.003814

40. Wang WW, Haque M, Chin HC, Yun JGJ. Injury severity of pedestrian crashes in Singapore. In: Australasian Transport Research Forum 2013 Proceedings (ATRF). 2013 Oct 1-15; Brisbane (AU). Available at: https://eprints.qut.edu.au/63892/

41. Ferenchak NN. Pedestrian age and gender in relation to crossing behavior at midblock crossings in India. Journal of Traffic and Transportation Engineering. 2016 Aug;3(4):345-351. doi: 10.1016/j.jtte.2015.12.001

42. Tulu GS, Washington S, King MJ, Haque M. Why are pedestrian crashes so different in developing countries? A review of relevant factors in relation to their impact in Ethiopia. Australasian Transport Research Forum 2013 Proceedings; 2013 Oct 2-4; Brisbane (AU). p. 1-18. Available at: https://www.australasiantransportresearchforum.org.au/sites/default/files/2013_tulu_washington_king_haque.pdf

43. Adanu EK, Hainen A, Jones S. Latent class analysis of factors that influence weekday and weekend single vehicle crash severities. Accident Analysis & Prevention. 2018 Apr;113(1):187-192. doi: 10.1016/j.aap.2018.01.035

44. Brazilian Ministry of Transportation. Manual for the treatment of critical traffic crash locations. PARE Program. Brasília (DF): Ministry of Transportion; 2002.