TRANSPORTES
www.revistatransportes.org.br
anpet
associação nacional de pesquisa e ensino em transportes
www.anpet.org.br

Factors related to highway crash severity in Brazil through a multinomial logistic regression model

Fatores relacionados à severidade de acidentes em rodovias no Brasil através de um modelo de regressão logística multinomial

Lucas Franceschi[1], Luciano Kaesemodel[2], Vera do Carmo Comparsi de Vargas[3], Andréa Cristina Konrath[4], Luiz Ricardo Nakamura[5], Thiago Gentil Ramires[6], Camila Belleza Maciel Barreto[7], Amir Mattar Valente[8]

[1]Federal University of Santa Catarina, Santa Catarina – Brazil, franceschi.lucas@gmail.com
[2]Federal University of Santa Catarina, Santa Catarina – Brazil, luciano.kaesemodel@gmail.com
[3]Federal University of Santa Catarina, Santa Catarina – Brazil, veradocarmo@gmail.com
[4]Federal University of Santa Catarina, Santa Catarina – Brazil, andreack@gmail.com
[5]Federal University of Santa Catarina, Santa Catarina – Brazil, luiz.rn@gmail.com
[6]Federal Technological University of Paraná, Paraná – Brazil, thiago.gentil@gmail.com
[7]Federal University of Santa Catarina, Santa Catarina – Brazil, camilabmaciel@gmail.com
[8]Federal University of Santa Catarina, Santa Catarina – Brazil, amir.labtrans@gmail.com

Recebido:
1 de março de 2021
Aceito para publicação:
14 de julho de 2021
Publicado:
7 de abril de 2022
Editor de área:
Sara Ferreira

Keywords:
Road transportation.
Injury severity.
Statistical learning.
Highway crashes.
Traffic safety.

Palavras-chave:
Transporte rodoviário.
Severidade de acidentes.
Aprendizagem estatística.
Acidentes em rodovias.
Segurança viária.

DOI:10.14295/transportes.v30i1.2756

ABSTRACT
Reducing the number of deaths by road crashes is an important priority for many countries around the world. Although focusing on the occurrence of crashes can provide safety policies that help reduce its numbers, studying their severity can provide different measures that may help further reduce the number of deaths by focusing on the most severe problems first. In this paper, a multinomial logistic regression model is fitted to nationwide highway crash data in Brazil from 2017 to 2019 to identify and estimate the associated factors to crash severity. Severity is classified as *without injury*, *with injured victims* or *with fatal victims*. Amongst other observations, results indicate that pedestrian involvement in highway crashes increase dramatically their severity. Also, conditions that favor greater speeds (clear weather, times when there are fewer vehicles on the road) are also related to an increase in crash severity, pointing to a proportional relation with traffic fluidity. Moreover, some observed limitations on the data may indicate that Brazil would benefit greatly from national crash records standards and unified databases, especially crossmatching crash reports with health data.

RESUMO
Reduzir o número de mortes por acidentes de trânsito é uma prioridade importante ao redor do mundo. O estudo da severidade dos acidentes pode melhorar as políticas públicas de segurança viária, concentrando esforços nas situações associadas a acidentes de maior severidade. Neste artigo, um modelo de regressão logística multinomial é ajustado a dados de acidentes em rodovias federais no Brasil de 2017 a 2019 para estimar os fatores associados à severidade dos acidentes. A severidade é classificada como sem lesão, com vítimas feridas ou com vítimas fatais. O envolvimento de pedestres é o principal fator identificado para aumento da severidade. Além disso, condições que favorecem maiores velocidades (como tempo limpo ou horários com menos tráfego) também estão associadas com maiores severidades. Em relação ao mês, as chances de maior severidade são menores no início do ano e maiores em agosto e em novembro. As limitações observadas indicam que o Brasil carece da adoção de padrões nacionais de registro de acidentes e de bancos de dados unificados, especialmente comparando registros de acidentes rodoviários com bancos de dados de saúde.

1. INTRODUCTION

Roadway crashes are a serious and important problem for all countries around the world. Recent reports show that 1.35 million road traffic deaths occur each year around the globe, amounting for the eighth leading cause of death for people of all ages, and the number one cause of death for children and young adults (World Health Organization, 2018). The number of injuries caused by traffic crashes can be upwards of 50 million per year worldwide. Therefore, great academic effort has been dedicated into understanding the main associated factors that result in such dramatic numbers to produce better safety policies, with measures that might reduce this impact. While much research is done in terms of analyzing the number of crashes, fostering policies for its reduction, the severity of crashes is also an important dimension. In this context, the approach of a Safe System, originated in Sweden and the Netherlands in the 1980s and 1990s, considers the failings of humans, accepting the validity of a simple ethical imperative that no human being should be killed or seriously injured as the result of a road crash (International Transport Forum, 2016). Thus, if the human factor of the safe system equation fails, the other elements must be activated, such as safe vehicles or safe roads.

Road transportation in Brazil accounts for more than 60% of the national freight transportation and almost 50% of all passenger traffic. The country has more than 1.7 million kilometers of highways, of which 213,208 are paved, and 75,744 kilometers (65,370 paved and 10,374 unpaved) are under federal jurisdiction (Brasil, 2019). On federal highways, crash data is made available via a crash reports database by the Federal Highway Police Department (in Portuguese: *Polícia Rodoviária Federal* - PRF), which includes important attributes such as crash cause and type, number of injured and fatal victims, weather conditions, road type and surface, amongst other attributes.

The PRF database provides a per crash number of victims for each injury severity classification (uninjured, injured, or dead). In this study, injury severity outcome is classified as either *without injuries*, *with victims* or *with fatal victims*. Savolainen *et al.* (2011) provide a review of methodological alternatives for crash-injury severity analysis. The authors indicate that related studies usually consider discrete outcome models due to the nature of crash severity classifications in most applications. This is the case for this paper, which seeks to identify which are the associated factors to road crash severity in Brazilian federal highways through a multinomial logistic regression model applied to the PRF database, with data from 2017 to 2019.

The goal of the present work is to provide information regarding the factors that are associated with crash severity in Brazil, so that this information can be used by public and private entities to prioritize safety policies and reduce accident severity. It is also the goal of this paper to present this information to the international road safety community, so that global reviews and comparisons can be made using these results. It is not the goal of this paper to present novel relationships between circumstantial variables and crash severity, since this is a subject that has been studied extensively over the decades (Evans, 2004), but rather to provide updated numbers regarding this reality in Brazil.

1.1. Literature Review

The factors that lead to road crash frequency and severity have been a topic of study for a long time. Many studies have applied statistical methods to obtain insights into this problem and to inform policy makers on the most critical factors that need to be addressed. Nonetheless, the growth in user accessible computing power and software availability have widened the possibilities for such analyses, and, even if the same model is applied, looking into different time periods is useful to measure the tendencies over time. In this section, a few papers with a related topic of study are reviewed.

The logistic regression model has been applied to analyze crash data on many occasions. For instance, Souza *et al.* (2016) identified which factors were related to the occurrence of crashes involving cyclists. Zhang *et al.* (2016) studied the effect of fatigue on the frequency and severity of traffic crashes in Guangdong, China, and Shakya and Marsani (2017) measured the influence of characteristics such as age, gender, vehicle type, hour of crash and type of collision over the severity of the crash, in a binary outcome response (with or without deaths). Although the logistic regression model has been widely used to study traffic crashes and severity data, one of its many assumptions is that the output (e.g., severity classification) must be binary. Therefore, severity classification is usually adopted as with or without fatalities. When further response levels are needed, a natural choice is to use the multinomial logistic regression model (Hosmer and Lemeshow, 2000).

One of the first studies to apply such technique to analyze crash data was presented in the paper by Shankar and Mannering (1996). The paper focused on single-vehicle motorcycle crashes using five years of data from the state of Washington, USA. The authors classified the motorcycle-rider injury severity outcome into five classes: property damage only, possible injury, evident injury, disabling injury and fatality. Amongst relevant conclusions regarding the relationships of the variables employed, the authors also concluded that the multinomial model was a promising methodology to evaluate the determinants of motorcycle crash severity.

Savolainen and Mannering (2007) also applied a multinomial regression model to analyze injury severity on motorcycle crashes using data from 2003 to 2005 in the state of Indiana, USA. In this case, the severity outcome was divided into four categories: no injury, non-incapacitating injury, incapacitating injury, and fatality. The main findings showed that the increase in rider age is related to more severe injuries and that other variables such as collision type, road characteristics, alcohol consumption, helmet use and unsafe speed are significantly related to the injury outcome of crashes.

Other studies that applied the multinomial logistic regression model to analyze crash severity data were those of Tay *et al.* (2011), Çelik and Oktay (2014), Wu *et al.* (2016) and Chen and Fan (2019). In Brazil, Girotto *et al.* (2016) studied the relationship between time working as a truck driver and the report of involvement in traffic crashes or near-misses. The authors concluded that there was an evident relationship between longer professional experience and a reduction in reported involvement in these occurrences.

Other methodologies have also been carried out, such as in Carrasco *et al.* (2012), who studied fatal motorcycle crashes from 2001 to 2009 in the city of Campinas, Brazil, using an Injury Severity Score (ISS) calculation. In this case, severity was represented by a continuous score, describing the most potentially fatal injuries. This study showed that alcohol consumption was a significant factor, and that head trauma was the most frequent and severe injury. It also showed that half of the victims died before receiving adequate medical care.

Almeida *et al.* (2013) analyzed risk factors associated with traffic crash severity in the city of Fortaleza, Brazil, using a non-concurrent cohort study of data between 2004 and 2008.

Deterministic and probabilistic relationship techniques were used to integrate different crash and health databases, and generalized linear models were used to investigate risk factors for death by traffic crash. The authors concluded that prevention activities should focus primarily on crashes involving two-wheeled vehicles that most often involve a single person, unskilled, male, at nighttime, during weekends and on roads that allow higher speeds. De Andrade *et al.* (2019) analyzed the trend in the number of fatalities, severe injuries, and minor injuries from traffic crashes on federal highways in Brazil, using a dataset from the PRF from year 2007 to 2017. The goal of this paper was to compare trends before and after the start of the Decade of Action for Road Safety using an Interrupted Time Series (ITS) study.

Cunto and Ferreira (2017) analyzed the injury severity of motorcycle crashes in Fortaleza, Brazil using mixed ordered response models. The study used data from urban crashes involving motorcycles, and levels of injury were classified as no apparent injury, slight injury, serious injury and fatal injury. Several variables were adopted as risk factors for this study, and the results have suggested that using a helmet amounted for a 9% reduced chance of suffering severe or fatal injuries. Crashes on daylight as well as on weekdays presented lower risk of resulting in fatal injuries. As shown, discrete outcome statistical models are widely used to analyze crash severity data and multinomial logistic regression model is proven a well-used method for this task. Furthermore, various possible severity classifications can be adopted. Crash severity analysis is observed to be most common for studies focusing on motorcycles. As for studies in Brazil, many studies about crash severity used datasets from single cities, whereas the PRF dataset for federal highways has been used primarily for crash frequency trends.

Therefore, Brazilian data on crash severity is found mainly in regional scopes, whereas nationwide data is more frequently used for crash frequency studies. One of the contributions of this paper (besides being informative of the factors affecting crash severity in Brazil) is to fill this gap, by performing a crash severity analysis using nationwide (rather than regional) crash data.

2. METHODS

As previously discussed, a common option to analyze crash severity is the multinomial logistic regression model. It is also a natural choice to model problems with more than two possible outcomes, that can be used as a classification tool, modelling the odds of the response (or target) variable Y as a function of a set of explanatory variables $(X_1, \dots, X_r)$, generalizing the logistic regression model (Hosmer and Lemeshow, 2000).

Mathematically, let us consider K possible outcome levels, in which one is chosen as the reference or baseline level and the other $K-1$ outcomes are separately modelled against this reference. As in Righetto et al. (2019), if $K = 3$, i.e., the response presents three different levels, as is the case for the severity outcome in the present study, and letting *no victims* be the referent outcome, we shall write

$$\frac{P(Y=1)}{P(Y=3)} = \mu \text{ and } \frac{P(Y=2)}{P(Y=3)} = \sigma \tag{1}$$

where $P(Y=m)$ represents the probability of $Y=m$ and μ and σ are unknown parameters representing the odds ratio between levels $Y=1$ versus $Y=3$ (i.e., injured victim versus uninjured victims) and $Y=2$ versus $Y=3$ (i.e., fatal victim versus uninjured victims), respectively.

To properly estimate and model parameters μ and σ as functions of a set of explanatory variables, i.e., to select which are the associated factors that mainly contribute for a road crash to outcome a fatal and/or injured victim, we may use appropriate link functions as follows.,

$$P(Y=1) = \frac{\exp(X_1\beta_1)}{1+\exp(X_1\beta_1)+\exp(X_2\beta_2)} \quad ,$$

$$P(Y=2) = \frac{\exp(X_2\beta_2)}{1+\exp(X_1\beta_1)+\exp(X_2\beta_2)} \quad \text{and} \qquad (2)$$

$$P(Y=3) = \frac{1}{1+\exp(X_1\beta_1)+\exp(X_2\beta_2)} \quad .$$

Model (1) is implemented in the *gamlss* package (Stasinopoulos and Rigby, 2007) in R software (The R Foundation, 2021) and can be accessed through the MN3() function. The papers of Stasinopoulos *et al.* (2017) and Rigby et al. (2019) contain further details regarding the estimation process, statistical modelling and distribution information used in this case. The associated factors (features) are selected using a stepwise-based procedure, called Strategy A, and can be accessed through the stepGAICAll.A() function in R. The procedure uses the Akaike information criterion (AIC) (Akaike, 1974) to perform the selection and is fully described by Nakamura *et al.* (2017) and Stasinopoulos *et al*. (2017).

2.1. Data

The dataset used in this study contains police-reported highway crash records on federal highways in Brazil from January 1st, 2017, up to December 31st, 2019. It is publicly available from PRF (Brasil, 2020), though it has been treated to filter some records that would not be able to be considered in the statistical analysis, e.g., there was no information in a specific characteristic. Furthermore, crash coordinates were crossmatched with the reported locations. That is, the coordinate of the crash was verified against infrastructure geographical databases to check that the coordinate was in fact located in the same highway as registered, in a similar kilometer mark as reported, and near the same location. Records found to be inconsistent were discarded. As a result, 172,778 valid records were considered, which is 76.7% of the total number (225,304 crashes) of records in the dataset. Geographical data was added in the database only in 2017, which is why previous data was not used in this study

The crash reports take note on the circumstances of each occurrence such as the type of the crash, its cause (as understood by the police officer), the weather, the time of day and the roadway geometry at the crash site. The crash cause variable is known to have a certain degree of uncertainty, since determining the exact cause of a crash requires that a more detailed study be performed. Therefore, the original 24 categories for this variable are grouped into three main categories: road-environmental, human and vehicle related causes. However, some uncertainty on the crash causes remains, which must be considered when interpreting the results.

All variables considered, including the severity outcome, had categorized values. To apply the multinomial logistic regression model, for each variable, a reference category was chosen to represent the baseline case. Some of the original variables and levels were reclassified to reduce the number of categories and enable the fitting of a regression model with a proper inference, due to the large amount of data and a small number of observations in a few cases. Table 1 shows how the variables and categories were grouped.

Table 1 – Original and grouped categories

Variable: Crash cause	
Grouped category	**Original categories**
Road-environmental	Road defect, animal on road, static object on road, low visibility, slippery surface, natural phenomena, insufficient/inadequate road signs or markings
Human	Driver lack of attention, pedestrian lack of attention, incompatible speed, external aggression, disobedience of traffic rules by the driver, sleeping driver, alcohol intake, undue overtaking, safety distance was not kept, disobedience of traffic rules by pedestrian, alcohol or drug ingestion by pedestrian, sudden illness, intake of psychoactive substances
Vehicle	Tyre damage or excessive wear, failure on the vehicle's lighting/signaling system, mechanical defect in the vehicle, excessive or poorly conditioned cargo
Variable: Crash type	
Grouped category	**Original categories**
Run-over	pedestrian run-over, animal run-over
rollover	rollover
tipping	tipping
vehicle-object collision	collision with static object, collision with moving object
vehicle-vehicle collision	frontal collision, side-to-side collision, rear-end collision, sideswipe
pileup	pileup
others	run-off-road, occupant ejection, cargo shedding, eventual damages, car fire
Variable: Weather	
Grouped category	**Original categories**
Rainy	Rain, drizzle, fog, snow, hail
Sunny	Clear skies, sunny
Cloudy	Cloudy
Others	High winds, ignored

All the variables and category names were also freely translated into English from the original (Portuguese), as presented in Table 2. The reference category is highlighted in bold text for each variable.

Table 2 – List of variables and categories, with reference categories highlighted in bold

Variable	Categories (reference in bold)
Business day	No, **yes**
Year	2017, **2018**, 2019
Holiday	**No**, yes
Region	Midwest, northeast, north, **southeast**, south
Month	**January**, February, March, April, May, June, July, August, September, October, November, December
Day of the week	Sunday, Monday, Tuesday, **Wednesday**, Thursday, Friday, Saturday
Time of day	Dawn, **day**, dusk, night
Pedestrians involved	Yes, **no**
Crash cause	**Human**, road-environmental, vehicle
Crash type	Run-over, rollover, **tipping**, vehicle-object collision, vehicle-vehicle collision, pileup, others
Weather	Sunny, foggy, **rainy**, others
Location	**Rural**, urban
Road type	Single-lane, **two-lane**, multiple lanes
Severity outcome (*output variable*)	**No victims**, injured victims, fatal victims

For the crash type, the difference between tipping and rollover crashes is that the police officer classifies a crash as rollover when the vehicle's roof touches the ground, whereas in tipping crashes the vehicle rolls from its upright position to laying on its side, without rolling further. Tipping crashes commonly involve large trucks and might happen primarily due to the lateral oscillation of loads, whereas rollovers usually involve small cars.

Table 2 presents all the variables that were considered from the PRF database. These variables include characteristics of the crash and other external factors that are believed to have an impact on the occurrence. When creating the crash registry forms (and therefore listing the data that would be recorded for each crash), PRF already hypothesized which factors are relevant for the crash occurrence and severity. Location-related variables (region and location) are relevant because the nature of the traffic and the road infrastructure in different places differ. Time-related ones (business day, year, holiday, month, day of the week) are relevant because the traffic flows change in nature and in volume over time, such as weekend versus weekday traffic, for instance, which can also affect crash probabilities. Environment related variables (time of day, weather, road type) directly affect the driving conditions such as low visibility or wet pavement, and crash circumstances (pedestrian involved, crash type, crash cause) directly affect the outcome of the crashes. Therefore, all variables listed in Table 2 were initially included in the statistical analysis.

The number of crashes per variable category in the considered database can be seen in Table 3, classified by severity outcome. The severity outcome distribution for each category is also presented in Figure 1.

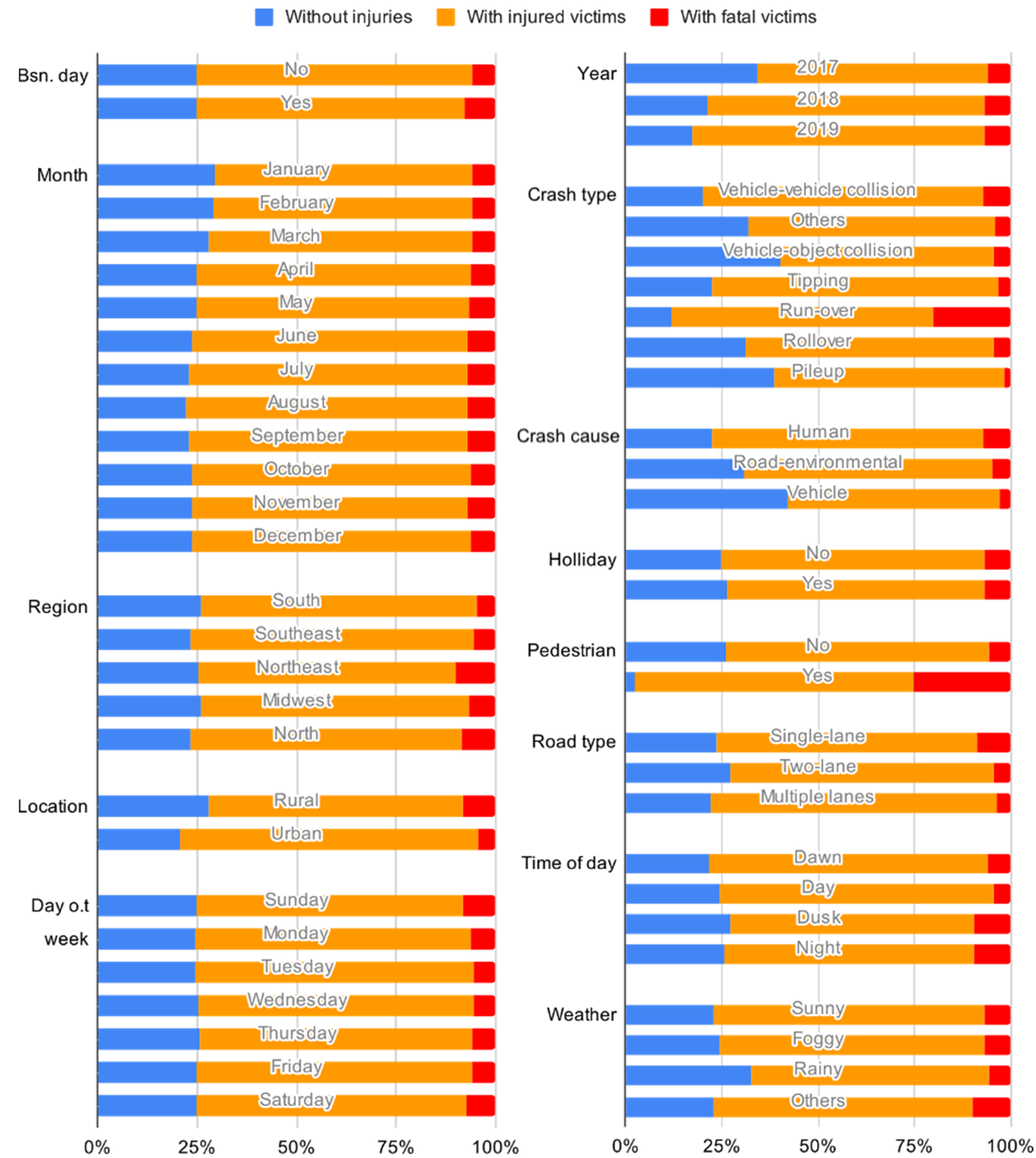

Figure 1. Injury severity distribution for each variable category

Table 3 – Total number of crashes in the database by variable category and by injury severity classification

Variable	Category	Number of crashes (2017 - 2019)							
		No injuries		W/ injured victims		W/ fatal victims		Total	
	Total	43,032	24.9%	118,487	68.6%	11,259	6.5%	172,778	100.0%
Business day	No	29,113	24.9%	80,731	69.1%	6,921	5.9%	116,765	67.6%
	Yes	13,919	24.8%	37,756	67.4%	4,338	7.7%	56,013	32.4%
Month	January	4,164	29.3%	9,185	64.7%	848	6.0%	14,197	8.2%
	February	3,804	29.0%	8,530	65.0%	788	6.0%	13,122	7.6%
	March	4,098	27.8%	9,774	66.4%	853	5.8%	14,725	8.5%
	April	3,550	24.7%	9,926	69.1%	881	6.1%	14,357	8.3%
	May	3,453	25.0%	9,402	68.2%	935	6.8%	13,790	8.0%
	June	3,428	23.8%	9,949	69.2%	1,006	7.0%	14,383	8.3%
	July	3,322	23.1%	10,046	69.8%	1,018	7.1%	14,386	8.3%
	August	3,166	22.4%	10,022	70.8%	976	6.9%	14,164	8.2%
	September	3,309	23.0%	10,053	70.0%	999	7.0%	14,361	8.3%
	October	3,457	23.6%	10,226	69.9%	941	6.4%	14,624	8.5%
	November	3,374	23.6%	9,915	69.5%	980	6.9%	14,269	8.3%
	December	3,907	23.8%	11,459	69.9%	1,034	6.3%	16,400	9.5%
Region	South	14,122	26.2%	37,311	69.1%	2,548	4.7%	53,981	31.2%
	Southeast	12,617	23.4%	38,318	71.1%	2,956	5.5%	53,891	31.2%
	Northeast	9,256	25.2%	23,801	64.7%	3,724	10.1%	36,781	21.3%
	Midwest	4,888	25.9%	12,750	67.6%	1,234	6.5%	18,872	10.9%
	North	2,149	23.2%	6,307	68.2%	797	8.6%	9,253	5.4%
Day of the week	Sunday	7,114	24.8%	19,327	67.2%	2,302	8.0%	28,743	16.6%
	Monday	5,742	24.5%	16,207	69.2%	1,458	6.2%	23,407	13.5%
	Tuesday	5,193	24.5%	14,762	69.8%	1,198	5.7%	21,153	12.2%
	Wednesday	5,454	25.4%	14,804	68.9%	1,220	5.7%	21,478	12.4%
	Thursday	5,797	25.5%	15,592	68.6%	1,333	5.9%	22,722	13.2%
	Friday	6,577	24.8%	18,320	69.1%	1,606	6.1%	26,503	15.3%
	Saturday	7,155	24.9%	19,475	67.7%	2,142	7.4%	28,772	16.7%
Year	2017	22,382	34.2%	39,105	59.8%	3,930	6.0%	65,417	37.9%
	2018	11,085	21.4%	37,121	71.8%	3,484	6.7%	51,690	29.9%
	2019	9,565	17.2%	42,261	75.9%	3,845	6.9%	55,671	32.2%
Crash type	Vehicle-vehicle collision	16,914	20.2%	61,004	72.7%	5,981	7.1%	83,899	48.6%
	Others	12,947	32.0%	25,952	64.2%	1,544	3.8%	40,443	23.4%
	Vehicle-object collision	5,534	40.0%	7,681	55.5%	623	4.5%	13,838	8.0%
	Tipping	2,837	22.4%	9,388	74.3%	416	3.3%	12,641	7.3%
	Run-over	1,359	11.9%	7,756	67.9%	2,316	20.3%	11,431	6.6%
	Rollover	2,534	31.0%	5,303	64.8%	343	4.2%	8,180	4.7%
	Pileup	907	38.7%	1,403	59.8%	36	1.5%	2,346	1.4%
Crash cause	Human	31,737	22.5%	99,259	70.4%	10,005	7.1%	141,001	81.6%
	Road-environmental	5,676	30.9%	11,829	64.4%	874	4.8%	18,379	10.6%
	Vehicle	5,619	41.9%	7,399	55.2%	380	2.8%	13,398	7.8%
Holliday	No	41,344	24.8%	114,236	68.6%	10,840	6.5%	166,420	96.3%
	Yes	1,688	26.5%	4,251	66.9%	419	6.6%	6,358	3.7%
Location	Rural	28,249	28.0%	64,657	64.0%	8,160	8.1%	101,066	58.5%
	Urban	14,783	20.6%	53,830	75.1%	3,099	4.3%	71,712	41.5%
Pedestrians involved	No	42,830	26.1%	112,059	68.4%	9,019	5.5%	163,908	94.9%
	Yes	202	2.3%	6,428	72.5%	2,240	25.3%	8,870	5.1%
Road type	Single-lane	20,991	23.5%	60,531	67.7%	7,864	8.8%	89,386	51.7%
	Two-lane	19,061	27.3%	47,847	68.5%	2,918	4.2%	69,826	40.4%
	Multiple lanes	2,980	22.0%	10,109	74.5%	477	3.5%	13,566	7.9%
Time of day	Day	23,278	24.5%	67,578	71.0%	4,286	4.5%	95,142	55.1%
	Night	15,367	25.8%	38,646	64.9%	5,563	9.3%	59,576	34.5%
	Dusk	2,027	21.6%	6,800	72.3%	574	6.1%	9,401	5.4%
	Dawn	2,360	27.3%	5,463	63.1%	836	9.7%	8,659	5.0%
Weather	Sunny	24,953	23.0%	76,337	70.3%	7,320	6.7%	108,610	62.9%
	Foggy	7,587	24.3%	21,537	69.1%	2,037	6.5%	31,161	18.0%
	Rainy	9,832	32.7%	18,648	62.0%	1,614	5.4%	30,094	17.4%
	Others	660	22.7%	1,965	67.5%	288	9.9%	2,913	1.7%

3. RESULTS

Table 4 displays the estimates of the final fitted multinomial logistic regression model obtained through the stepwise-based selection method. Coefficients with positive signs in parameter μ imply a greater probability (consequently, a greater chance) that a crash presents an injured victim (compared to crashes without victims), whilst a negative sign indicates a lesser chance. Similarly, for parameter σ, positive and negative coefficients indicate greater and lesser probabilities that a crash result in a fatality when compared to crashes without victims, respectively.

Table 4 – Estimates, standard errors (SE; in parenthesis) and odds ratio of the fitted multinomial logistic regression model

Feature		μ (injured victims vs. no victims)		σ (fatal victims vs. no victims)	
		Estimate (SE)	Odds Ratio	Estimate (SE)	Odds Ratio
Intercept		**0.861** (0.037)**		**-2.870 (0.082)**	
Region (Ref.: Southeast)	Midwest	-0.171** (0.020)	0.843	0.079** (0.040)	1.082
	Northeast	-0.290** (0.017)	0.748	0.342** (0.030)	1.408
	North	-0.219** (0.028)	0.803	0.181** (0.049)	1.198
	South	-0.223** (0.015)	0.800	-0.355** (0.031)	0.701
Month (Ref.: January)	February	0.023 (0.028)	1.023	0.007 (0.057)	1.007
	March	0.082** (0.027)	1.085	-0.001 (0.057)	0.999
	April	0.220** (0.028)	1.246	0.099* (0.056)	1.104
	May	0.185** (0.028)	1.203	0.154** (0.055)	1.166
	June	0.229** (0.028)	1.257	0.196** (0.055)	1.217
	July	0.257** (0.028)	1.293	0.251** (0.055)	1.285
	August	0.319** (0.028)	1.376	0.294** (0.055)	1.342
	September	0.288** (0.028)	1.334	0.268** (0.055)	1.307
	October	0.276** (0.028)	1.318	0.245** (0.055)	1.278
	November	0.282** (0.028)	1.326	0.339** (0.055)	1.404
	December	0.277** (0.027)	1.319	0.207** (0.054)	1.230
Day of the week (Ref.: Wednesday)	Sunday	0.075** (0.022)	1.078	0.033** (0.042)	1.034
	Monday	0.069** (0.023)	1.071	0.181** (0.045)	1.198
	Tuesday	0.054** (0.023)	1.055	0.048 (0.047)	1.049
	Thursday	0.001 (0.023)	1.001	0.021 (0.046)	1.021
	Friday	0.035 (0.022)	1.036	0.064 (0.044)	1.067
	Saturday	0.040* (0.022)	1.041	0.254** (0.042)	1.289
Time of day (Ref.: Day)	Dawn	-0.186** (0.027)	0.830	0.679 (0.046)	1.972
	Dusk	0.089** (0.027)	1.093	0.252 (0.052)	1.287
	Night	-0.175** (0.013)	0.839	0.505 (0.025)	1.657
Pedestrians involved (Ref.: No)	Yes	2.629** (0.078)	13.86	4.069** (0.094)	58.498
Crash cause (Ref.: Human)	Road-Environmental	-0.028 (0.021)	0.972	-0.505** (0.045)	0.604
	Vehicle	-0.711** (0.020)	0.491	-1.051** (0.057)	0.350
Crash type (Ref.: Tipping)	Run-over	-0.680** (0.046)	0.507	0.228** (0.084)	1.256
	Rollover	-0.440** (0.033)	0.644	-0.154* (0.079)	0.857
	Vehicle-object collision	-0.929** (0.029)	0.395	-0.150** (0.069)	0.861
	Vehicle-vehicle collision	-0.122** (0.024)	0.885	0.823** (0.056)	2.278
	Pileup	-0.963** (0.049)	0.382	-1.114** (0.179)	0.328
	Others	-0.459** (0.024)	0.632	-0.230** (0.060)	0.795
Weather (Ref.: Rainy)	Sunny	0.334** (0.016)	1.397	0.284** (0.033)	1.328
	Foggy	0.282** (0.019)	1.326	0.211** (0.039)	1.235
	Others	0.392** (0.048)	1.480	0.375** (0.081)	1.455
Location (Ref.: Rural)	Urban	0.314** (0.013)	1.369	-0.668** (0.026)	0.513
Road type (Ref.: Two-lane)	Multiple	0.098** (0.024)	1.103	-0.013 (0.056)	0.987
	Single-lane	0.175** (0.013)	1.191	0.778** (0.026)	2.177

Note: ** Statistical difference at 5% level. * Statistical difference at 10% level.

Using the individual odds ratio presented in Table 4, it is possible to measure how each of the factors affects the chance of an outcome. For instance, considering all other explanatory variables as fixed (i.e., under the same circumstances), a crash that occurs during dusk is $exp(0.089) = 1.093$ times more likely, (it has 9.3% more chance) to present an injured victim (compared to crashes with uninjured victims) than one occurred in daytime. Moreover, under the same circumstances, a pileup crash outcome is $exp(-1.114) = 0.328$ times as likely (it has 67.2% less chance) to be a fatal crash (compared to crashes with uninjured victims) than a tipping crash. It is noteworthy that a coefficient that is not statistically significant (Table 4) indicates that there is no statistical difference between the reference and the tested level of the factor. For example, there are no statistical evidence that crashes during February have a higher (or lesser) chance to present injured or fatal victims (when compared to uninjured victims) than the ones occurred in January.

Different combinations of factors may be observed, resulting in different probabilities for each possible outcome. An interactive dashboard is provided at https://ramires.shinyapps.io/acidentes/ so that the exact probability of the crash severity outcome as a function of the observed attributes can be observed.

To assess the adequacy of the fitted model presented in Table 4, we analyze the normalized (randomized) quantile residuals (Dunn and Smyth, 1996). Stasinopoulos *et al.* (2017) states that the advantage of using these residuals is that the true residuals always have a standard distribution when the assumed model is correct. Within the *gamlss* framework, we may obtain the normalized randomized quantile residuals (Dunn and Smyth, 1996) given by $\hat{r}_i = \Phi^{-1}(\hat{u}_i)$, where Φ^{-1} is the inverse cumulative function of the standard normal distribution, $\hat{u}_i = F(y|\widehat{\boldsymbol{\theta}})$, $F(y|\widehat{\boldsymbol{\theta}})$ is a step function with jumps at the integers $j \in R_y$ and $\widehat{\boldsymbol{\theta}} = (\mu, \sigma)$. (Stasinopoulos *et al.*, 2017).

The randomized quantile residuals are usually displayed in worm plots (van Buuren and Fredriks, 2001) (i.e., detrended qqplots), where it is possible to visualize whether the residuals follow a normal distribution or not. If a vertical shift, slope, quadratic, or cubic shape is observed, then there is a problem with the location, dispersion, skewness and/or kurtosis parameters of the residuals, respectively (and, consequently, with the response variable distribution). Nonetheless, as can be seen in Figure 2, the residuals follow a Gaussian distribution. Therefore, we can say that the model provides a good fit to the data. Further information regarding this diagnostic tool may be seen in (Stasinopoulos et al., 2017).

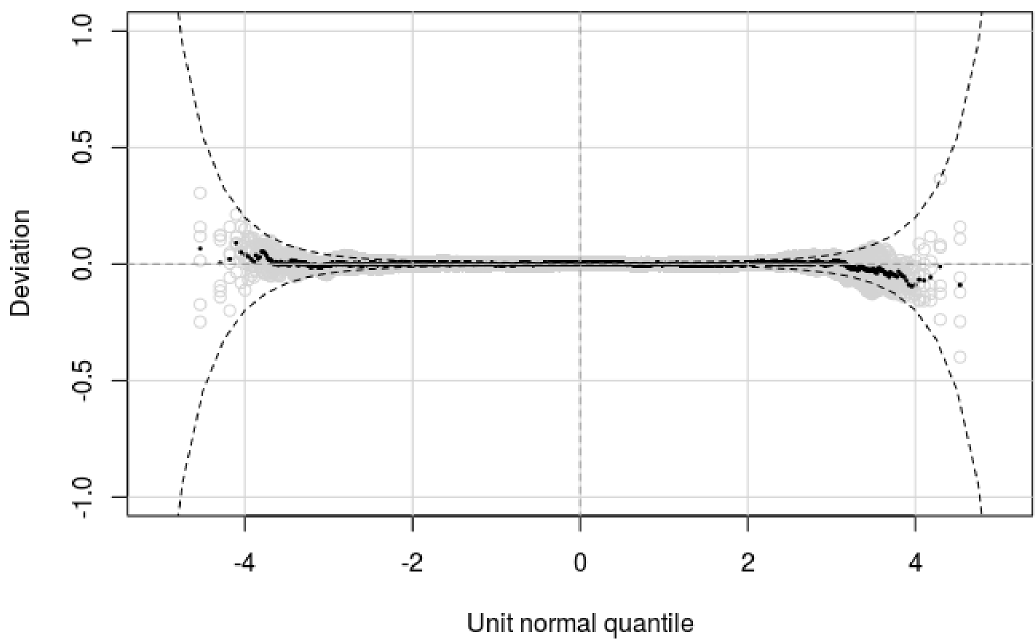

Figure 2. Worm plot of the residuals for the fitted multinomial logistic model

4. DISCUSSION

The most dramatic factor to produce higher severities is pedestrian involvement. If a pedestrian is involved, the crash has more than 58 times the chance of resulting in death than other crashes, and more than 13 times the chance of resulting in injuries. The dataset that was used in this study includes data from federal highways, i.e., crashes that occurred in long-distance, high-speed highways, where normally there is not much infrastructure for pedestrian traffic.

However, the highways cross urban perimeters, and in such places, there is normally a high urban concentration on the road's margins. Even in rural locations, people tend to build and live near the highways, which oftentimes mean that the highway serves as a local street for small villages, without proper infrastructure for pedestrian traffic.

In this dataset, of 8,870 crashes that registered pedestrian involvement (5.1% of the total crashes on the dataset), 5,563 (62.7%) occurred in urban locations. Also, 7,933 (89.4%) of such records were run-over occurrences. Moreover, out of a total of 11,259 crashes on the dataset with fatal outcomes, 2,240 (19.9%) involved a pedestrian. That is, even though they account for a little over 5% of the crashes, occurrences involving pedestrians represent almost 20% of all fatal crashes in federal highways.

In a policy-making perspective, this means that, even though pedestrians' involvement may not be very frequent in federal highways, they account for an important number of fatal crashes. This is primarily explained by the increased speed of the vehicles, since pedestrians have little chance of surviving crashes at highway speeds (Wang and Cicchino, 2020). Highway planning and maintenance entities must account for the fact that the highways are also used by pedestrians and invest in proper infrastructure in places where such traffic is likely.

Another important variable for this model is the crash type. Crashes where vehicles collide with other vehicles are much more likely to result in a fatal outcome than other types. For instance, they have 2.278 times the chance of resulting in death than tipping crashes. Run-over crashes have a 25.6% increased chance of fatality. Pileups have the lowest probability of resulting in deaths, with 67.2% less chance of fatality than tipping crashes. Looking into injuries, however, tipping crashes are the most likely to result in injuries than all other types, whereas pileups also have the lowest chance of injuries. When looking into the number of records in the database, vehicle on vehicle collisions account for 48.5% of all records, and 53.1% of all crashes with fatality.

It is noteworthy to mention that, of all the vehicle-to-vehicle collision with fatal outcomes, more than 50% are frontal collisions, that is, when a vehicle collides directly with an oncoming vehicle. Another study (Wang *et al.*, 2019) has found a different result, where the highest fatality rates were observed on read-end collisions.

The time of day is also related to the severity outcome. Crashes occurring at dawn have almost double the chance (97.2% more) of resulting in fatality than crashes occurring during daytime, even though crashes on daytime and dusk have a slightly increased chance of resulting in injuries. Crashes during the night also have a 65.7% increased chance of fatality than during the day. The fatality increase during the night and dawn could be associated with periods where traffic is still starting to appear on the highway, when it is easier for vehicles to speed up, but it is also more likely that other vehicles would be on the highway, increasing the odds for high-speed crashes, which in turn would be more severe. This result is consistent with those of Cunto and Ferreira (2017), where the authors also concluded that crashes during the daylight have a lower risk of being fatal. Almeida *et al.* (2013) also concluded that crashes in the early morning hours were associated with increased severity.

A similar effect can also be observed on the weather variable, where crashes in sunny weather have a 39.7% greater chance of injuries and 32.8% increased chance of fatality than crashes in the rain. The similarity being that in sunny weather, as well as during the night and dawn, the drivers could be more likely to speed, whereas under rainy weather, vehicles naturally slow down due to safety concerns. These results agree with what was found by Hordofa *et al.* (2018) where dry conditions were also associated with a higher fatality rate.

As for the location of the crash, results show that crashes in rural locations are 94.9% more likely to result in death than crashes in urban locations. However, crashes in cities are 36.9% more likely to result in injured victims. This also points to the effect of increased speeds in crash severity, since vehicles are more likely to be able to speed in open, rural areas than urban locations. Other studies that investigated this question have also come to similar conclusions. Darma, Karim and Abdullah (2017) showed that, in Malaysia, 66% of the traffic deaths occur in rural regions. Almeida *et al.* (2013) also concluded that crashes on roads that allow greater speeds are more severe. Rakauskas, Ward and Gerberich (2009) estimate that the increased severity in rural regions could be explained, amongst other factors, by the increased distance from medical care facilities, which in turn increase the time it takes for seriously injured patients to receive medical attention.

Results for crash cause indicate that crashes that are understood as caused by human errors, rather than vehicle defects or road-environmental factors, are the most likely to result in fatality. These results are in accordance with other studies, such as that by Mohanty and Gupta (2015), which found that personal or human behavioral factors were the main associated causes to road crashes, and by Iqbal *et al.* (2020), that found that 66.8% of collisions were caused by human factors. Crashes caused by vehicle malfunctions have the lowest probability of resulting in injury or deaths. The approach of a safe system (International Transport Forum, 2016), and the design of forgiving roads could very positively impact this scenario, creating an opportunity for drivers to safely stop or return to the road after a human error. As discussed previously, however, the PRF database presents crash causes only as understood by the police officer, and not from a robust analysis. Therefore, there is some uncertainty regarding the exact determination of the crash causes.

The geoeconomic region where the crash happened also influences crash severity chances. The behavior is distinct for injured and fatal victims. For the first model (injured victims), the Southeastern region, which was chosen as the reference case for being the most populated region in Brazil, has the greatest chances of crashes resulting in injuries. Regarding fatal crashes, the Northeastern region has a 40.8% greater chance of fatality than the Southeastern region. For comparison, Morais Neto *et al*. (2016) studied the regional disparities in road traffic injuries in Brazil and found that the number of lesions per accident is lowest in the Southeastern region, moderate in the South and Northeast, and highest in the Midwest and North. The paper by Barroso Jr. *et al.* (2019) also shows that the chances for an accident to be lethal when compared to those in the Southeast region are 78% higher in the Northeast, 58% higher in the North, 44% higher in the Midwest and 10% higher in the South.

Finally, another way to look at the results of the fitted model is by looking into the combination of factors that lead to the greater chances of fatality, injuries and of no injury. The crashes that have the greatest chance for fatal outcomes are vehicle on vehicle collision crashes on single-lane highways in the Northeast, during November, on Saturdays, at dawn, with pedestrian involvement, caused by human error and in rural regions. This combination results in a 69% chance of fatality. Crashes with the greatest chance of injuries are tipping crashes on single-lane highways, in the Southeast, during August, on Sundays, at dusk, with pedestrian involvement, caused by human error and in urban regions. These crashes have a 90% chance of non-fatal injury outcomes. Also, according to the fitted model, the least severe crashes are pileups on two-lane highways in the Northeast, during January, on Wednesdays, at dawn or night (both levels have similar effects on the model), without pedestrian involvement, caused by vehicle problems and in rural regions.

5. CONCLUSION

In this paper, a dataset from crashes in federal highways in Brazil was analyzed through a multinomial logistic regression model to observe the influence of a few different factors on crash severity. Also, an interactive online dashboard was provided where the fitted model's results can be observed by providing a set of input variables. By fitting the model and analyzing its resulting coefficients, a few conclusions regarding crash severity in federal highways in Brazil can be drawn, which are summarized as follows.

- Crashes with pedestrian involvement are not the most frequent on highways, but they have a dramatically increased chance to result in death. Therefore, it is necessary to account for pedestrian presence on federal highways, especially near populated areas where pedestrian traffic is more likely.
- Vehicle on vehicle collisions are the most likely to result in deaths, and crashes on single-lane roads also present double the risk of death than on two-lane highways. This emphasizes the safety benefits of duplicating single-lane roads, or otherwise isolating opposite traffic directions.
- Although crashes during the night and dawn are less frequent, they have a significantly higher chance of resulting in death than crashes during daylight, so that safety measures can be put in place specifically addressing the conditions in these hours of the day.
- Conditions that favor greater speeds (clear weather and times or places where there are fewer vehicles on the road) are associated with increased crash severity. This emphasizes the importance of speed control measures, such as enforcement or road safety education programs. It also means that conditions which are usually considered safer for traffic can be dangerous conditions as well, since drivers might feel safer and increase speeds, and incidents are likely to be more severe. This information could be used, for instance, to support education campaigns emphasizing the dangers in favorable driving conditions.
- Some results suggest that the nature of the traffic could be associated to crash severity, such as for the month variable or day of the week. To better investigate this, however, more in-depth regional studies must be performed considering economic activities and events throughout the year, such as harvest periods for specific crops or increased tourism seasons for each region.

Moreover, the presented results also provided a measure of each of the considered variable's influence on crash severity, so that the results in Table 4 may be used for further investigation on crash severity in federal highways in Brazil and to improve future road safety planning efforts.

5.1. Limitations and future research

There are a few limitations to the present work that must also be stated. Firstly, the registry done by the PRF officers consider only deaths that occur on the site of the crash, and victims that leave the crash site alive are considered as injured victims, even if they later succumb to their injuries in the hospital. In the US, by contrast, crashes are considered to have a fatal outcome if a death occurs, consequently, up to 30 days after the crash (National Highway Traffic Safety Administration, 2019). National health databases in Brazil do in fact consider road fatalities as those that occur because of a road crash, not only at the site. The problem is that those databases are not integrated with the PRF database, so it is not possible to study the crash factors that resulted in the fatality. To solve this problem, an integration between these databases should be established, which is not a trivial task. Therefore, since this study focused on the circumstances of the crash to study severity, it was limited to the information recorded by the police officer, which includes only the immediate result of the crash.

Another factor that affects the present work is the absence of a national standard on how to report and register traffic crashes. Crash records in Brazil would benefit greatly from improved policies that standardize the way that any entity would classify and record road crashes. A federal standard could even allow studies like this one to be performed with various datasets from different entities, which could be more easily integrated. A national standard was once created in 2016, by the National Traffic Council (in Portuguese: *Conselho Nacional de Trânsito* - CONTRAN*)*, called RENAEST (in Portuguese: *Registro Nacional de Acidentes e Estatísticas de Trânsito*). However, practice shows that only a few entities did follow the national standard, and to the present date there is no national unified database for traffic crashes.

This study also seeks to understand crash severity as a function of nationwide (rather than regional) data. Because of this, some underlying factors might hinder the results by distorting the data for different regions, such as there being different patterns for data registry in each region, different entities responsible for data collection, or even differences on the nature of the traffic flows across the country. The adoption of standard national practices for crash data collection would benefit this aspect as well. While such a standard is not adopted, the results for severity by region of the country should be considered cautiously.

Finally, it is noteworthy that data from the first months of year 2020 were already available at the time of the drafting of this paper. However, the COVID-19 global pandemic has affected human life on many areas, including the nature of transportation and economic activities. As such, the authors feel that including these data would require specific consideration to be given to the particularities of the transportation patterns during the pandemic and, therefore, chose not to include it in the study.

REFERENCES

Akaike, H. (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, v. 19, n.6, p. 716–723. DOI:10.1109/TAC.1974.1100705.

Almeida, R. L. F. de; Bezerra Filho, J. G.; Braga, J. U.; Magalhães, F. B.; Macedo, M. C. M. e Silva, K. A.. (2013) Man, road and vehicle: risk factors associated with the severity of traffic accidents. *Revista de Saúde Pública*, v.47, n.4, p. 718–731. DOI:10.1590/S0034-8910.2013047003657.

Andrade, F. R. e Antunes, J. L. F. (2019) Trends in the number of traffic accident victims on Brazil's federal highways before and after the start of the Decade of Action for Road Safety. *Cadernos de Saude Publica*, v.35, n.8, p. 1–11. DOI:10.1590/0102-311X00250218.

Barroso Jr., G. T.; Bertho, A. C. S. e Veiga, A. de C. (2019) A letalidade dos acidentes de trânsito nas rodovias federais brasileiras. *Revista Brasileira de Estudos de População*, v.36, p. 1–22. DOI:10.20947/S0102-3098a0074.

Brasil (2019) *Boletim Estatístico*. Brasília, DF: Confederação Nacional dos Transportes. Disponível em: <https://www.cnt.org.br/boletins> (acesso em 16/07/2021).

Brasil (2020) *Acidentes*. Brasília, DF: Polícia Rodoviária Federal. Disponível em: <https://www.gov.br/prf/pt-br/acesso-a-informacao/dados-abertos/dados-abertos-acidentes> (acesso em 16/07/2021).

van Buuren, S. e Fredriks, M. (2001) Worm plot: A simple diagnostic device for modelling growth reference curves. *Statistics in Medicine*, v.20, n.8, p. 1259–1277. DOI:10.1002/sim.746.

Carrasco, C. E.; Godinho, M.; Barros, M. B. A; Rizoli, S. e Fraga, G. P. (2012) Fatal motorcycle crashes: A serious public health problem in Brazil. *World Journal of Emergency Surgery*, v.7, n.S5, p. 7–10. DOI:10.1186/1749-7922-7-S1-S5.

Çelik, A. K. e Oktay, E. (2014) A multinomial logit analysis of risk factors influencing road traffic injury severities in the Erzurum and Kars Provinces of Turkey. *Accident Analysis & Prevention*, v.72, p. 66–77. DOI:10.1016/j.aap.2014.06.010.

Chen, Z. e Fan, W. (D.) (2019) A multinomial logit model of pedestrian-vehicle crash severity in North Carolina. *International Journal of Transportation Science and Technology*, v.8, n.1, p. 43–52. DOI:10.1016/j.ijtst.2018.10.001.

Cunto, F. J. C. e Ferreira, S. (2017) An analysis of the injury severity of motorcycle crashes in Brazil using mixed ordered response models. *Journal of Transportation Safety and Security*, v.9, p. 33–46. DOI:10.1080/19439962.2016.1162891.

Darma, Y., Karim, M. R. e Abdullah, S. (2017) An analysis of Malaysia road traffic death distribution by road environment. *Sādhanā*, v.42, n.9, p. 1605–1615. DOI:10.1007/s12046-017-0694-9.

Dunn, P. K. e Smyth, G. K. (1996) Randomized Quantile Residuals. *Journal of Computational and Graphical Statistics*, v.5, n.3, p. 236. DOI:10.2307/1390802.

Evans, L. (2004) *Traffic Safety*. Blooming Hills: Science Serving Society.

Girotto, E. *et al.* (2016) Professional experience and traffic accidents/near-miss accidents among truck drivers. *Accident Analysis and Prevention*, v.95, p. 299–304. DOI:10.1016/j.aap.2016.07.004.

Hordofa, G. G. ; Assehid, S. ; Girma, A. e Weldemarium, T. D. (2018) Prevalence of fatality and associated factors of road traffic accidents among victims reported to Burayu town police stations, between 2010 and 2015, Ethiopia. *Journal of Transport & Health*, v.10, p. 186–193. DOI:10.1016/j.jth.2018.06.007.

Hosmer, D. W. e Lemeshow, S. (2000) *Applied Logistic Regression* (2ª ed.) Hoboken: John Wiley & Sons, Inc.

International Transport Forum (2016) *Zero Road Deaths and Serious Injuries*. Paris: OECD Publishing. DOI:10.1787/9789282108055-en.

Iqbal, A.; ur Rehman, Z.; Ali, S.; Ullah, K. e Ghani, U. (2020) Road traffic accident analysis and identification of black spot locations on highway. *Civil Engineering Journal*, v.6, n.12, p. 2448–2456. DOI:10.28991/cej-2020-03091629.

Mohanty, M. e Gupta, A. (2015) Factors affecting road crash modeling. *Journal of Transport Literature*, v.9, n.2, p. 15–19. DOI:10.1590/2238-1031.jtl.v9n2a3.

Morais Neto, O. L. M. ; Andrade, A. L.; Guimarães, R. A.; Mandacarú, P. M. P. e Tobias, G. C. (2016) Regional disparities in road traffic injuries and their determinants in Brazil, 2013. *International Journal for Equity in Health*, v.15, n.1, p. 142. DOI:10.1186/s12939-016-0433-6.

Nakamura, L. R.; Rigby, R. A.; Stasinopoulos, D. M.; Leandro, R. A.; Villegas, C. e Pescim, R. R. (2017) Modelling location, scale and shape parameters of the Birnbaum-Saunders generalized t distribution. *Journal of Data Science*, v.15, n.2, p. 221–238. DOI: 10.6339/JDS.201704_15(2).0003

National Highway Traffic Safety Administration (2019) *Fatality Analysis Reporting System (FARS) Analytical User 's Manual, NHTSA*. Washington. Disponível em: < https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812827> Data de acesso: 19/07/2021.

Rakauskas, M. E., Ward, N. J. e Gerberich, S. G. (2009) Identification of differences between rural and urban safety cultures. *Accident Analysis & Prevention*, v.41, n.5, p. 931–937. DOI:10.1016/j.aap.2009.05.008.

Rigby, R. A.; Stasinopoulos, M. D.; Heller, G. Z.; Bastiani, F. D. (2019) *Distributions for Modeling Location, Scale, and Shape* (1ª ed.) Nova York, NY: Chapman and Hall/CRC. DOI:10.1201/9780429298547.

Righetto, A. J.; Ramires, T. G.; Nakamura, L. R.; Castanho, P. L. D. B.; Faes, C. e Savian, T. V. (2019) Predicting weed invasion in a sugarcane cultivar using multispectral image. *Journal of Applied Statistics*, v.46, n.1, p. 1–12. DOI:10.1080/02664763.2018.1450362.

Savolainen, P. e Mannering, F. L. (2007) Probabilistic models of motorcyclists' injury severities in single- and multi-vehicle crashes. *Accident Analysis and Prevention*, v.39, n.5, p. 955–963. DOI:10.1016/j.aap.2006.12.016.

Savolainen, P. T.; Mannering, F. L.; Lord, D. e Quddus, M. A;(2011) The statistical analysis of highway crash-injury severities: A review and assessment of methodological alternatives. *Accident Analysis and Prevention*, v.43, n.5, p. 1666–1676. DOI:10.1016/j.aap.2011.03.025.

Shakya, R. e Marsani, A. (2017) Using logistic regression to estimate the influence of accident factors on accident severity in Kathmandy Valley. *Proceedings of IOE Graduate Conference, 2017*, Lalitpur, Nepal: Institute of Engineering, Tribhuvan University, p. 311–324.

Shankar, V. e Mannering, F. (1996) An exploratory multinomial logit analysis of single-vehicle motorcycle accident severity. *Journal of Safety Research*, v.27, n.3, p. 183–194. DOI:10.1016/0022-4375(96)00010-2.

Souza, C. A. M.; Bahia, C. A. e Constantino, P. (2016) Analysis of factors associated with traffic accidents of cyclists attended in Brazilian state capitals. *Ciencia e Saude Coletiva*, v.21, n.12, p. 3683–3690. DOI:10.1590/1413-812320152112.24152016.

Stasinopoulos, D. M. e Rigby, R. A. (2007) Generalized additive models for location scale and shape (GAMLSS) in R. *Journal of Statistical Software*, v.23, n.7, p. 1–46. DOI:10.18637/jss.v023.i07.

Stasinopoulos, M. D.; Rigby, R. A; Heller, G. Z.; Voudouris, V e Bastiani, F. de (2017) *Flexible Regression and Smoothing: Using GAMLSS in R* (1ª ed.) Boca Raton: CRC Press.

Tay, R.; Choi, J., Kattan, L e Khan, A. (2011) A multinomial logit model of pedestrian-vehicle crash severity. *International Journal of Sustainable Transportation*, v.5, n.4, p. 233–249. DOI:10.1080/15568318.2010.497547.

The R Foundation (2021) *R: The R project for statistical computing*. Disponível em: https://www.r-project.org/. Data de acesso: 19/07/2021.

Wang, D., Liu, Q; Ma, L.; Zhang, Y. e Cong, H.(2019) Road traffic accident severity analysis: A census-based study in China. *Journal of Safety Research*, v.70, p. 135–147. DOI:10.1016/j.jsr.2019.06.002.

Wang, J. e Cicchino, J. B. (2020) Fatal pedestrian crashes on interstates and other freeways in the United States. *Journal of Safety Research*, v. 74, p. 1–7. DOI:10.1016/j.jsr.2020.04.009.

World Health Organization (2018) *Global Status Report on Road Safety*. Geneva: WHO.

Wu, Q., Zhang, G.; Ci, Y.; Wu, L.; Tarefder, R. A. e Alcántara, A. "D." (2016) Exploratory multinomial logit model–based driver injury severity analyses for teenage and adult drivers in intersection-related crashes. *Traffic Injury Prevention*, v.17, n.4, p. 413–422. DOI:10.1080/15389588.2015.1100722.

Zhang, G.; Yau, K. K. W; Zhang, X. e Li, Y.; (2016) Traffic accidents involving fatigue driving and their extent of casualties. *Accident Analysis and Prevention*. v.87, p. 34–42. DOI:10.1016/j.aap.2015.10.033.