(ISSN: 3085-6124)
Cognitus Interdisciplinary Journal

STATISTICAL ANALYSIS OF ROAD ACCIDENTS IN BRAZIL

Análise Estatística De Acidentes De Trânsito No Brasil

ABSTRACT

Road traffic accidents rank among the leading causes of death for individuals aged 5 to 29, according to the World Health Organization. Studies indicate that men are approximately three times more likely than women to die in road collisions. This article examines whether there is a positive correlation between the number of active drivers (by age group and gender) and the demographic profile of accident victims (by age group and gender). Aggregated data on licensed drivers were used, organized into seven age brackets (18–30; 31–40; 41–50; 51–60; 61–70; 71–80; 80+), for both men and women, along with data on accident victims in the same age groups (grouped as 18–29; 30–39; 40–49; 50–59; 60–69; 70–79; 80+). Pearson's correlation coefficient was calculated separately for men and women, as well as the correlation between the proportion of male drivers and male victims. A strong positive correlation was found for both sexes ($r \approx 0.862$ for men; $r \approx 0.925$ for women), as well as a high correlation ($r \approx 0.961$) between male proportions. We conclude that age groups with more drivers also exhibit more victims, and that the proportion of male victims mirrors the proportion of male drivers in each age group. Preventive policies should prioritize young male drivers in prevention policies.

Eduardo Sansão Sozinho Maurício
Pós-graduando em Engenharia Mineral pela Escola Politécnica da USP.
Graduado em Engenharia Mecânica, com interesse em Ciências de Dados e Tomada de Decisão.
Nacionalidade: angolana | Residência: São Paulo, Brasil.
https://orcid.org/0009-0004-0340-8236

KEYWORDS: Traffic accidents; Active drivers; Victims; Age group; Gender.

(ISSN: 3085-6124)
Cognitus Interdisciplinary Journal

*Autor correspondente:
Eduardo Sansão Sozinho Maurício
eduardosansao052@gmail.com

Recebido em: [08-07-2025]
Publicado em: [18-07-2025]

RESUMO

Os acidentes de trânsito estão entre as principais causas de morte de pessoas com idades entre 5 e 29 anos, de acordo com a Organização Mundial da Saúde. Estudos indicam que os homens têm aproximadamente três vezes mais chances de morrer em colisões nas estradas do que as mulheres. Este artigo examina se há uma correlação positiva entre o número de motoristas ativos (por faixa etária e gênero) e o perfil demográfico das vítimas de acidentes (por faixa etária e gênero). Foram usados dados agregados sobre motoristas licenciados, organizados em sete faixas etárias (18-30; 31-40; 41-50; 51-60; 61-70; 71-80; 80+), tanto para homens quanto para mulheres, juntamente com dados sobre vítimas de acidentes nas mesmas faixas etárias (agrupadas como 18-29; 30-39; 40-49; 50-59; 60-69; 70-79; 80+). O coeficiente de correlação de Pearson foi calculado separadamente para homens e mulheres, bem como a correlação entre a proporção de motoristas do sexo masculino e vítimas do sexo masculino. Foi encontrada uma forte correlação positiva para ambos os sexos ($r \approx 0{,}862$ para homens; $r \approx 0{,}925$ para mulheres), bem como uma alta correlação ($r \approx 0{,}961$) entre as proporções de homens. Concluímos que as faixas etárias com mais motoristas também apresentam mais vítimas, e que a proporção de vítimas do sexo masculino reflete a proporção de motoristas do sexo masculino em cada faixa etária. As políticas preventivas devem priorizar os motoristas jovens do sexo masculino nas políticas de prevenção.

PALAVRAS-CHAVES: Acidentes de trânsito; Motoristas ativos; Vítimas; Faixa etária; Gênero.

INTRODUCTION

Currently, traffic injuries rank as the leading cause of death worldwide for individuals aged 5 to 29. In particular, World Health Organization data show that in 2020 approximately 1.3 million people died in road crashes, and estimates indicate that men are about three times more likely than women to die under these circumstances. This sex and age-specific risk disparity motivates an investigation of how the demographic profile of drivers *who represent the "at-risk" population* in crashes relates to the profile of registered crash victims. Building on these mortality and exposure insights, we turn now to directly compare driver and victim profiles [1].

Despite the extensive literature describing overall mortality and injury rates from collisions, few studies directly link the profile of active drivers (the number of licensed drivers by age group and sex) to the profile of victims (the number of men and women injured or killed by age group) [2]. Understanding this relationship is important for informing road-safety public policies, guiding targeted educational campaigns, and directing traffic enforcement to regions with higher concentrations of higher-risk drivers.

Objectives

This study aims to assess whether a positive correlation exists between the number of active drivers (by age group and sex) and the number of accident victims (by age group and sex).

- Describing the distribution of active drivers by age group and sex.

- Describing the demographic profile of accident victims by age group and sex.

- Calculating Pearson's "r" between the number of drivers and victims in each age group for both men and women.

- Evaluating whether the proportion of male victims in each age group mirrors the proportion of male drivers in the same group.

(ISSN: 3085-6124)
Cognitus Interdisciplinary Journal

Hypotheses

- **Hypothesis 1:** There is a strong positive correlation between the number of active drivers and the number of accident victims in each age group, for both men and women.

- **Hypothesis 2:** The proportion of male victims in each age group is higher in the age groups where the proportion of male drivers is higher.

MATERIALS AND METHODS

The numbers of active drivers were obtained from the DETRAN website, containing the counts of licensed men and women in the following twelve original age groups:

Active Driver

Table 1: Active Drivers by Age Group and Sex

Age group	Men	Women
0 - 10	64	0
11 - 17	2.905	273
18 - 24	100.322	20.159
25 - 29	76.089	20.959
30 - 34	71.262	21.761
35 - 39	71.336	22.762
40 - 44	66.801	20.807
45 - 49	54.064	15.303
50 - 54	44.464	11.444
55 -59	36.977	8.623
60 - 64	27.518	6.077
65 - 69	18.554	3.957
70 -74	11.400	2.353
75 - 79	6.263	1.280
≥ 18	4.161	593

Source: authors (2025)

Table 2. Profile of Involved Victims

Age group	Men	Woman
0 - 10	64	0
11 - 17	2.905	273
18 - 24	100.322	20.159
25 - 29	76.089	20.959
30 - 34	71.262	21.761
35 - 39	71.336	22.762
40 - 44	66.801	20.807
45 - 49	54.064	15.303
50 - 54	44.464	11.444
55 -59	36.977	8.623
60 - 64	27.518	6.077
65 - 69	18.554	3.957
70 -74	11.400	2.353
75 - 79	6.263	1.280
≥ 18	4.161	593

Source: authors (2025)

For analysis purposes, only victims aged ≥ 18 were included, and their ages were regrouped into 10-year bands (with the exception of the first band, which spans 12 years, and the last, open-ended).

Age-Group Grouping

To align the twelve original age groups for drivers and the fifteen original age groups for victims into comparable seven age groups, we regrouped as follows:

Table 3. Age-Group Grouping (Drivers)

Age Group No	Original Drivers Group	Grouped Age group
1	18-21 + 22-25 + 26-30	18-30
2	31-40	31-40
3	41-50	41-50
4	51-60	51-60
5	61-70	61-70
6	71-80	71-80

7	81-90 + 91-100 + 101-120 + 120+	+80

Source: authors (2025)

Table 4. Age-Group Grouping (Victims)

Age Group No	Original Victim Groups	Grouped Age group
1	18-24 + 25-29	18-29
2	30-34 + 35-39	30-39
3	40-44 + 45-49	40-49
4	50-54 + 55-59	50-59
5	60-64 + 65-69	60-69
6	70-74 + 75-79	70-79
7	+80	+80

Source: authors (2025)

This study employed secondary data sourced from the Brazilian Department of Traffic (DETRAN), which included the number of active licensed drivers and the number of road accident victims, both disaggregated by age group and sex. The analysis focused on individuals aged 18 years and older.

To ensure comparability between datasets, original age groups were regrouped into seven standardized intervals. For active drivers, the age brackets were consolidated into: 18–30, 31–40, 41–50, 51–60, 61–70, 71–80, and 80+. For accident victims, the regrouped intervals included: 18–29, 30–39, 40–49, 50–59, 60–69, 70–79, and 80+. While this approach aimed to maximize age alignment, minor discrepancies were noted—such as the inclusion of 30-year-olds in the driver group but not explicitly in the victim group. These differences were acknowledged as limitations, yet the strategy was considered appropriate given the exploratory nature of the study and the aggregated structure of the data.

Three types of variables were defined for the analysis. The exposure variables consisted of the absolute number of active drivers, disaggregated by age group and sex. The outcome variables referred to the absolute number of accident victims, also stratified by age and sex.

Additionally, proportional variables were computed for each age group, indicating the proportion of male drivers and male victims in the respective categories.

Descriptive analyses were conducted to generate frequency tables (absolute and relative) for both drivers and victims across all age groups. To test the research hypotheses, Pearson's correlation coefficient (r) was calculated in two scenarios: (1) the correlation between the number of active drivers and the number of accident victims, separately for males and females; and (2) the correlation between the proportion of male drivers and the proportion of male victims across age groups. A statistical significance level of $\alpha = 0.05$ was adopted.

Data were processed using Python, with the pandas and numpy libraries applied for manipulation and computation. Visualizations, including scatter plots to illustrate correlations, were created using the matplotlib library.

RESULTS

Table 5. Summary of Active Drivers by Age Group and Sex

Age Group	Mem	Women
18–30	2 485 327	1 746 600
31–40	3 099 192	2 423 540
41–50	3 225 908	2 480 860
51–60	2 417 121	1 641 479
61–70	1 769 049	1 093 496
71–80	949 448	512 994
80+	439 006	177 718

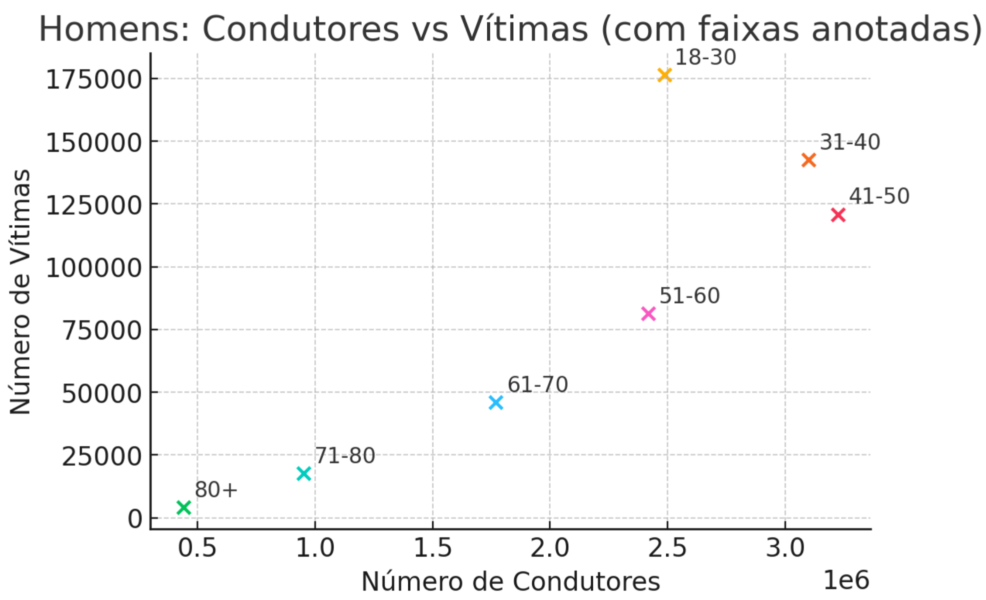
Source: authors (2025)

Among men, the age group with the highest number of accident victims was 18–29, totaling 176,411 individuals, followed by the 30–39 group with 142,598 victims. Among women, the highest concentration of victims occurred in the 30–39 age group, with 44,523, followed closely by the 18–29 group, with 41,118. In both sexes, a consistent decline in the number of victims was observed as age increased, particularly in the older age groups, with significantly lower figures recorded in the 70–79 and 80+ brackets.

To test the first hypothesis, the Pearson correlation coefficient was calculated between the number of active drivers and the number of accident victims across the seven age groups, analyzed separately for men and women. The results demonstrated a strong and statistically significant positive correlation: for men, r = 0.8618 ($p < 0.05$); and for women, r = 0.9246 ($p < 0.05$). These findings indicate that, for both sexes, age groups with a higher number of licensed drivers also tend to have a higher number of accident victims.

Scatter plots corresponding to these correlations (Figures 1 and 2) visually reinforce this pattern, showing a clear upward trend between the number of active drivers and the number of victims by age group for both men and women.

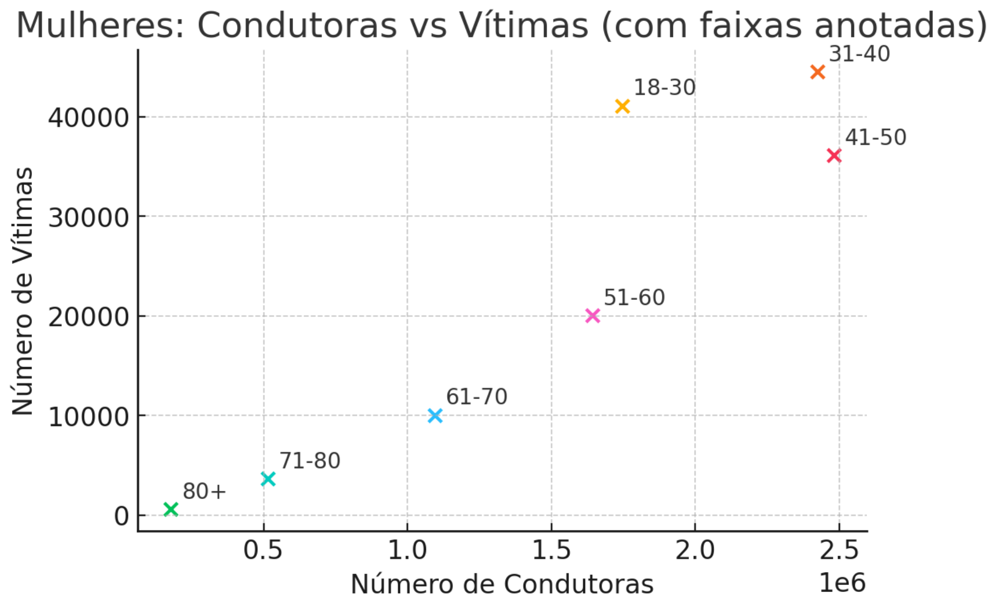
Figure 1: Active Drivers vs. Victims (Men)



Source: authors (2025)

(ISSN: 3085-6124)
Cognitus Interdisciplinary Journal

Figure 2: Active Drivers vs. Victims (Women)



Source: authors (2025)

In testing the second hypothesis, the analysis examined, for each age group, the proportion of men among all active drivers and the proportion of men among all accident victims. These proportions were calculated using the formula:

Proportion = (Number of males) / (Number of males + Number of females)

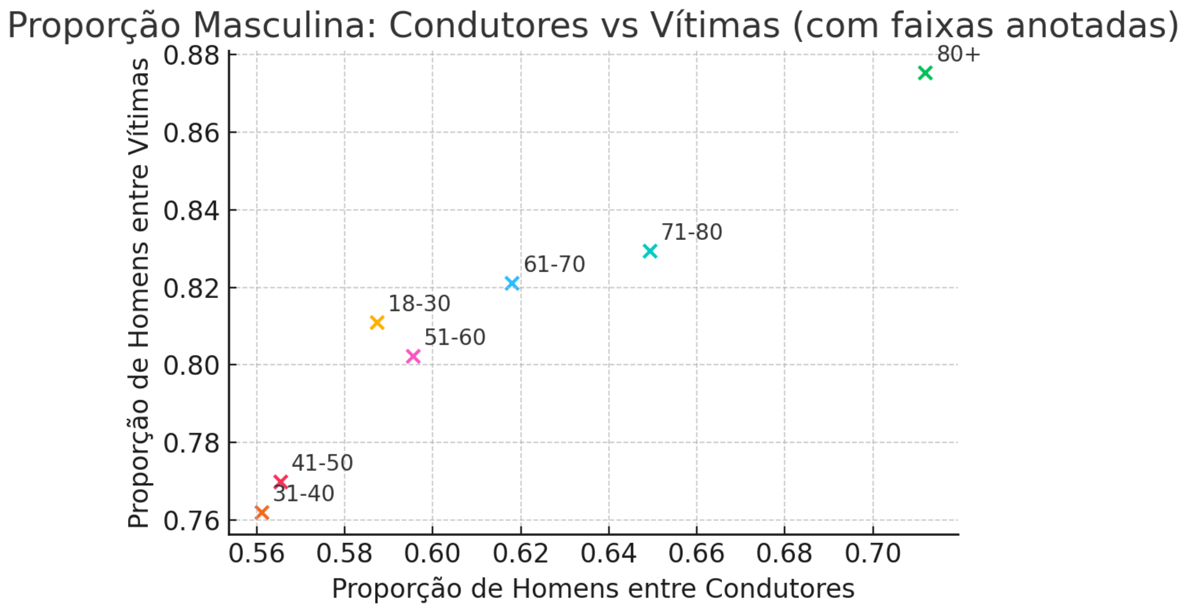
This calculation was applied separately to the driver and victim datasets across the seven defined age groups.

The Pearson correlation coefficient between the proportion of male drivers and the proportion of male victims across age groups was $r = 0.9610$ ($p < 0.05$), indicating a very strong and statistically significant positive association. This suggests that age groups with a higher percentage of male drivers also tend to have a higher percentage of male victims.

Figure 3 presents a scatter plot illustrating this relationship, where the alignment of data points along an upward trend line visually confirms the strength of the correlation.

(ISSN: 3085-6124)
Cognitus Interdisciplinary Journal

Figure 3. Proportion of Male Active Drivers vs. Proportion of Male Victims by Age Group



Source: authors (2025)

Table 7. Calculated Correlations

Sex / Measure	Pearson's r
Men (active drivers × victims)	0,8618
Women (active drivers × victims)	0,9246
Male proportion (active drivers vs. victims)	0,9610

Source: authors (2025)

DISCUSSION

The findings of this study provide empirical support for both proposed hypotheses. Regarding Hypothesis 1, the results revealed a strong positive correlation between the number of active drivers and the number of accident victims across all age groups, for both men (r = 0.8618) and women (r = 0.9246). This suggests that age groups with a greater number of licensed drivers are more frequently involved in traffic accidents, reinforcing the use of the number of drivers as a viable proxy for exposure to road risk.

(ISSN: 3085-6124)
Cognitus Interdisciplinary Journal

In relation to Hypothesis 2, the very high correlation between the proportion of male drivers and the proportion of male victims (r = 0.9610) indicates that age groups predominantly composed of male drivers are also those in which the majority of victims are male. This aligns with widely recognized evidence that men—especially younger men—are more prone to risk-taking behaviors in traffic environments, which increases their vulnerability to accidents.

These results are consistent with previous studies that identify young male drivers as a high-risk group. Lopes et al.[3] ,for example, reported that 60% of severe urban crashes in Brazil involve drivers under the age of 30, most of whom are male. In our analysis, the highest concentrations of both drivers and victims are found in the youngest age groups (18–30 for drivers and 18–29 for victims), which further supports the association between higher exposure and increased crash frequency.

In the Latin American context, Reyna-Castillo et al.[4] found a Pearson correlation coefficient of approximately 0.88 between the number of drivers and victims in Mexico, while Theofilatos[5] reported r ≈ 0.90 in a similar study conducted in Spain. Both are remarkably close to the values identified in this study. Additionally, data from the Pan American Health Organization indicate that up to 80% of young crash victims are male, which aligns closely with the gender-related correlations observed in our findings.

Traffic accidents in Brazil disproportionately affect men, young people, and residents of the Midwest region. Drivers, passengers, and motorcyclists are the groups most frequently involved, and although seat belt use in the front seat is relatively high, the same cannot be said for the rear seat. Almost one-third of victims reported abandoning t heir usual activities after the accident, reflecting significant impacts on quality of life[6].

This pattern of victimization becomes even more serious when considering the lethality on Brazilian federal highways. Men, pedestrians, and the elderly are the groups at greatest risk of death, especially in accidents occurring on Sundays, during the early hours of the morning, on curves, and in rural areas. The Northeast region has the highest proportion of fatalities, indicating that the victim's profile and the context of the accident are critical determinants in the severity of the outcomes [7].

In addition, trend analyses point to significant regional disparities. While capital cities maintained stable mortality rates from land transport accidents (LTA), municipalities in the interior of 14 states showed an increase, particularly in Piauí. In contrast, capitals such as

Curitiba recorded significant declines. Only São Paulo and Rio Grande do Sul showed a reduction in inland cities, highlighting regional inequality in addressing the problem [8].

Significant changes occurred after the implementation of the Decade of Action for Road Safety (DARS) in 2011. Until then, there had been a growing trend of deaths and injuries on federal highways, especially in the South and Midwest regions. Following this initiative, a reversal of the scenario was observed, with a significant decline in accident mortality and morbidity rates. The coordinated national action contributed to mitigating the continuous growth of these events, although each death still represents, on average, more than 12 injuries [9].

More specific studies, such as the one that analyzed Federal Highway 163 in Mato Grosso, reveal additional risk factors. Poor pavement, head-on collisions, and pedestrian accidents were associated with high risks of accidents with victims. Road maintenance problems and disobedience of traffic signs were also found to be relevant determinants. These findings reinforce the need for comprehensive strategies that integrate structural, behavioral, and regional contextual aspects [10].

Despite these contributions, some methodological limitations must be acknowledged. First, there is a mismatch between the age groups of drivers and victims, which can lead to distortions in correlation values. In addition, the use of the number of licensed drivers as a proxy for exposure does not consider actual driving activity, such as kilometers traveled, which can vary significantly between age groups. The use of aggregated data also prevents the analysis of crucial variables, such as vehicle type, road conditions, or time of accident. Finally, the absence of behavioral information, such as alcohol consumption, speeding, and cell phone use while driving, limits the depth of the analyses and makes more robust statistical models unfeasible.

From a public policy perspective, the findings strongly support the prioritization of interventions directed toward young male drivers, the group most exposed and most frequently victimized in traffic accidents. This includes educational campaigns tailored to risk reduction, increased enforcement through targeted speed controls and sobriety checkpoints, and infrastructure solutions such as reduced speed zones or segregated motorcycle lanes in high-risk areas.

As for future research, several paths are recommended. First, refined exposure metrics such as vehicle-miles traveled should be employed to improve accuracy in risk estimation. Second, longitudinal studies could provide insight into whether the correlations identified here remain consistent over time. Third, incorporating behavioral and environmental variables into multivariable models would allow for a deeper understanding of the mechanisms underlying traffic risk. Lastly, further analyses that distinguish between vehicle types—particularly cars and motorcycles—are essential, given the different risk profiles associated with each.

CONCLUSION

This study confirms that the number of road traffic victims is strongly associated with the number of active drivers across age and sex groups, particularly among young males. The high correlation between male driver proportions and male victimization highlights a critical demographic requiring focused intervention. These results underscore the importance of implementing targeted public policies, educational campaigns, and enforcement strategies aimed at reducing traffic injuries in high-risk populations. Future research should improve exposure metrics, integrate behavioral data, and account for vehicle types to enhance the precision and scope of traffic safety analyses.

REFERENCES

1. Bachani AM, Peden M, Gururaj G, Norton R, Hyder AA. Road Traffic Injuries. *Disease Control Priorities, Third Edition (Volume 7): Injury Prevention and Environmental Health*. Published online October 27, 2017:35-54. doi:10.1596/978-1-4648-0522-6_CH3
2. Braver ER, Trempel RE. Are older drivers actually at higher risk of involvement in collisions resulting in deaths or non-fatal injuries among their passengers and other road users? *Injury Prevention*. 2004;10(1):27-32. doi:10.1136/ip.2003.002923
3. Lopes LGF, Ribeiro SHP, Sousa SRGR, et al. Levantamento do perfil epidemiológico dos óbitos por acidentes de trânsito no estado de Pernambuco de 2015 a 2019. *Research, Society and Development*. 2022;11(8):e14511830681. doi:10.33448/rsd-v11i8.30681
4. Reyna-Castillo M, Santiago A, Martínez SI, Rocha JAC. Social Sustainability and Resilience in Supply Chains of Latin America on COVID-19 Times: Classification Using Evolutionary Fuzzy Knowledge. *Mathematics*. 2022;10(14):2371. doi:10.3390/math10142371
5. Theofilatos A. Incorporating real-time traffic and weather data to explore road accident likelihood and severity in urban arterials. *J Safety Res*. 2017;61:9-21. doi:10.1016/j.jsr.2017.02.003

6. Malta DC, Mascarenhas MDM, Bernal RTI, et al. Análise das ocorrências das lesões no trânsito e fatores relacionados segundo resultados da Pesquisa Nacional por Amostra de Domicílios (PNAD) Brasil, 2008. *Cien Saude Colet*. 2011;16(9):3679-3687. doi:10.1590/S1413-81232011001000005

7. Barroso Junior GT, Bertho ACS, Veiga A de C. A LETALIDADE DOS ACIDENTES DE TRÂNSITO NAS RODOVIAS FEDERAIS BRASILEIRAS. *Rev Bras Estud Popul*. 2019;36:1-22. doi:10.20947/S0102-3098a0074

8. Aquino ÉC de, Antunes JLF, Morais Neto OL de. Mortalidade por acidentes de trânsito no Brasil (2000–2016). *Rev Saude Publica*. 2020;54:122. doi:10.11606/s1518-8787.2020054001703

9. Andrade FR de, Antunes JLF. Tendência do número de vítimas em acidentes de trânsito nas rodovias federais brasileiras antes e depois da Década de Ação pela Segurança no Trânsito. *Cad Saude Publica*. 2019;35(8). doi:10.1590/0102-311x00250218

10. Almeida LV de C, Pignatti MG, Espinosa MM. Principais fatores associados à ocorrência de acidentes de trânsito na BR 163, Mato Grosso, Brasil, 2004. *Cad Saude Publica*. 2009;25(2):303-312. doi:10.1590/S0102-311X2009000200008