

## MODELOS CATEGÓRICOS APLICADOS À ANÁLISE EXPLORATÓRIA DOS NÍVEIS DE AGREGAÇÃO DA SEVERIDADE DE ACIDENTES COM PEDESTRES EM ÁREAS URBANAS

**Mateus Nogueira Silva**

**Flávio José Craveiro Cunto**

Departamento de Engenharia de Transportes  
Universidade Federal do Ceará

**Marcos José Timbó Lima Gomes**

Universidade Federal do Cariri

### RESUMO

A promoção de um ambiente viário mais seguro para os pedestres requer a compreensão dos fatores de risco associados às lesões sofridas por esses usuários quando envolvidos em acidentes de trânsito. Esse estudo é dedicado a investigar diferentes configurações de classificação da severidade na modelagem categórica e identificar fatores que influenciam na gravidade dos acidentes com pedestres. Para isso foram estimados modelos logit multinomiais utilizando registros de acidentes da cidade de Fortaleza entre os anos de 2017 e 2019. Foram investigadas variáveis relacionadas ao pedestre, aspectos do acidente e características da estrutura urbana. Os resultados indicaram que o modelo com três níveis de severidade (leve/moderado, grave e fatal) apresentou o melhor desempenho. A partir desse modelo observou-se que fatores como idade avançada do pedestre, acidentes ocorridos a noite, com veículos pesados, em finais de semana e localizados em vias de maior classificação viária, estão associados a lesões mais severas.

### ABSTRACT

Promoting a safer road environment for pedestrians requires an understanding of the risk factors associated with the injuries suffered by these users when they are involved in crashes. This study is dedicated to investigating different severity classification configurations in the categorical modeling and identifying factors that influence the severity of crashes with pedestrians. For this purpose, multinomial logit models were estimated by using crashes records from the city of Fortaleza between the years 2017 and 2019. Variables related to pedestrians, aspects of the crash and characteristics of the urban structure were investigated. The results indicated that the model with three levels of severity (light/moderate, severe and fatal) showed the best performance. From this model, it was observed that factors such as pedestrian's advanced age, accidents at night, with the involvement of heavy vehicles, on weekends and located on roads with higher road classification, are associated with more severe injuries.

### 1. INTRODUÇÃO

O novo paradigma da segurança viária, fundamentado nas abordagens dos Sistemas Seguros e da Visão Zero, se baseia em um entendimento de que os seres humanos cometem erros e são vulneráveis a lesões em caso de um acidente de trânsito. No entanto, o sistema viário deve ser projetado de modo que o erro humano não leve a um desfecho grave ou fatal. Essas abordagens trazem uma perspectiva mais profunda das causas por trás das fatalidades e das lesões graves no trânsito e engloba, também, o conceito de responsabilidade compartilhada, em que tanto os governos, quanto o setor privado e a sociedade compartilham a responsabilidade de um sistema de transportes seguro (Belin *et al.*, 2012 ; Welle *et al.*, 2018).

Os sistemas seguros e a visão zero levam em consideração a suscetibilidade das pessoas às lesões graves no trânsito e dentro do ambiente viário os pedestres são considerados os usuários mais vulneráveis. Shinar (2017) destaca dois pontos que aumentam a probabilidade de lesões mais severas dos pedestres quando envolvidos em um acidente de trânsito: primeiro, o fato de estarem totalmente expostos sem nenhuma barreira de proteção e segundo por possuírem uma massa significativamente menor em relação aos veículos motorizados. Além disso, alguns

aspectos que distinguem os pedestres dos demais usuários, como um movimento no tráfego menos regulado e mais variável, tornam mais difícil sua proteção.

Bhat *et al.* (2017) destacam que os esforços para a priorização desses usuários precisam ser coordenados com estratégias que aumentem sua segurança. Isso, por sua vez, requer uma compreensão dos fatores de risco associados às lesões de pedestres nos acidentes de trânsito, para permitir a identificação de configurações inseguras do espaço urbano e ajudar a definir contramedidas adequadas para políticas de transporte. Na literatura esses fatores são comumente categorizados em grupos tais como características dos pedestres, dos motoristas, atributos dos veículos, características da via, condições ambientais, aspectos do acidente, entre outros.

Diversos pesquisadores vêm desenvolvendo esforços de modelagem voltados à estimação do grau de severidade como variável de resposta qualitativa (Kwigizile, *et al.*, 2011; Rothman *et al.*, 2012; Dong *et al.*, 2019). De acordo com Savolainen *et al.* (2011), as variáveis dependentes desses modelos são particularmente um resultado de resposta binária (lesão ou não lesão) ou um resultado de resposta múltipla (lesão fatal, grave, leve ou ileso). Esse tipo de dado é modelado principalmente usando modelos de regressão logística. Nestes casos, não existe uma agregação temporal dos acidentes de trânsito (anual, mensal, etc), como nas funções de desempenho da segurança viária ou modelos de previsão de acidentes, mas uma análise desagregada e de natureza mais explicativa entre os fatores contribuintes e a severidade do acidente, em função de sua ocorrência.

Embora diferentes metodologias de modelagem estejam disponíveis na literatura para examinar a gravidade de acidentes relacionada a vários fatores de influência, pouco se sabe sobre como a configuração de classificação da severidade adotada melhora o resultado da modelagem e o entendimento da influência desses fatores analisados. Como os dados de acidentes são extraídos de diferentes fontes como relatórios policiais ou registros de hospitais, é possível que haja divergências entre essas informações (Tsui *et al.*, 2009). Além disso, ainda há uma dificuldade na reportagem de acidentes de menor gravidade (sem vítimas). Dessa forma, é comum encontrar diversos estudos que utilizam diferentes configurações de agregação dos níveis de severidade, especialmente das categorias intermediárias entre ileso e fatal (Clifton *et al.*, 2009; Jang *et al.*, 2013).

Diante do exposto, esse trabalho tem como objetivo principal avaliar diferentes configurações de classificação da severidade para a modelagem da gravidade de acidentes com pedestres e identificar fatores que influenciam a severidade desses acidentes através de modelos multinomiais. O município de Fortaleza será utilizado como estudo de caso devido a possibilidade de acesso aos dados necessários para realização desse trabalho.

## 2. MODELAGEM DA SEVERIDADE DOS ATROPELAMENTOS DE PEDESTRES

O desenvolvimento e a implementação de medidas eficazes para a segurança de pedestres exigem uma exploração e análise abrangentes dos fatores que influenciam a probabilidade de ocorrência de acidentes e os níveis de gravidade das lesões. Muitos estudos tentaram determinar o impacto de fatores específicos como a idade e gênero do pedestre, velocidade do veículo, condições da via, etc., em atropelamentos. Devido à vulnerabilidade às lesões, a gravidade pode diferir de acordo com a idade, com lesões mais graves associadas principalmente a crianças e idosos (Abay, 2013; Eluru *et al.*, 2008). Já pedestres do sexo feminino estão associados a níveis mais altos de severidade do que pedestre do sexo masculino (Lee e Abdel-Aty, 2005).

Alguns estudos concentraram-se nos efeitos do tipo de controle de tráfego nos níveis de gravidade do acidente (Lee e Abdel-Aty, 2005; Eluru *et al.*, 2008). Esses pesquisadores descobriram que os níveis de lesões aumentam na ausência de controles de tráfego, como semáforos para pedestres. Além disso, vários estudos também examinaram o impacto do tipo de veículo e dos limites de velocidade das vias e observaram que veículos pesados, assim como limites de velocidade mais altos estão relacionados a lesões mais severas nos pedestres (Chen e Fan, 2019; Aziz *et al.*, 2013; Eluru *et al.*, 2008; Sze e Wong, 2007). Diversos estudos investigaram essa relação entre velocidade e severidade da lesão do pedestre e indicaram que lesões mais severas estão associadas a velocidades de impacto mais altas. O risco de ferimentos graves ou morte de pedestres aumenta exponencialmente com a velocidade (Rosén *et al.*, 2011; Anderson *et al.*, 1997; Garder, 2004; OPAS, 2012). Apesar da sua importância, trata-se de uma variável de extrema dificuldade de observação no momento da ocorrência e por isso variáveis *proxy* como a velocidade da via ou a classificação viária são comumente empregadas (Jang *et al.*, 2013; D. Li *et al.*, 2017).

Para Clifton *et al.* (2009), vias arteriais e rodovias, são projetadas para maiores velocidades e volumes de veículos e tendem a ser menos seguras para os pedestres. Já em áreas com alta densidade populacional e comercial deve-se esperar um grande número de acidentes, embora as lesões possam ser menos severas devido aos congestionamentos e às baixas velocidades dos veículos, proporcionados pelas características urbanas do ambiente. Outros aspectos do espaço urbano também podem ter influência significativa na gravidade dos atropelamentos, principalmente numa escala micro, como o número de faixas de rolamento que aumenta a distância de travessia e consequentemente a exposição do pedestre (Aziz *et al.*, 2013; Pour-Rouholamin e Zhou, 2016).

A severidade de lesões no trânsito em geral é analisada usando modelos categóricos. Devido à natureza ordenada da gravidade das lesões, modelos como logit/probit ordenados são comumente empregados (Jang *et al.*, 2013; Lee e Abdel-Aty, 2005). Contudo, essa abordagem assume que as variáveis tem o mesmo impacto (para o valor e o sinal do parâmetro) em todos os níveis de gravidade da lesão. Como consequência desse aspecto, existe uma restrição do modo como as variáveis explicativas afetam as probabilidades do resultado. Além disso, modelos ordenados são mais suscetíveis à subnotificação de dados de lesões, resultando em estimativas de parâmetros tendenciosas ou inconsistentes (Washington *et al.*, 2003; Savolainen *et al.*, 2011).

Modelos que não consideram a natureza ordenada das lesões, tal como o logit multinomial, são também frequentes na análise da severidade de acidentes de trânsito (Tay *et al.*, 2011; Chen e Fan, 2019). Apesar de não considerarem a ordem dos resultados da gravidade dos ferimentos, esses modelos não são afetados por essas restrições impostas pelos modelos ordenados. De modo geral, cada abordagem tem premissas e restrições que têm implicações sobre as inferências dos modelos, todavia, a superioridade de uma abordagem sobre outra pode depender muitos dos dados utilizados (Mannering e Bhat, 2014).

A escolha do tipo de modelo mais adequado, assim como a seleção de fatores específicos e exclusivos relacionados à gravidade da lesão do atropelamento, dependem muito das circunstâncias do local de estudo, do conjunto de dados usados para a análise e dos objetivos da pesquisa. Para Wang *et al.* (2013), mesmo em países ou cidades desenvolvidos, as diferenças

nas infraestruturas viárias, nas condições de tráfego, nos padrões comportamentais dos pedestres e motoristas podem resultar em um conjunto diferente de fatores significativos associados à gravidade dos ferimentos dos pedestres.

### 3. METODOLOGIA

Esta seção apresenta, inicialmente, informações sobre o banco de dados de acidentes de trânsito e as bases complementares de atributos da via/estrutura urbana usados nesse trabalho. Em seguida, são descritas as variáveis selecionadas para análise e as configurações de classificação da severidade que serão avaliadas. Por fim, é apresentada a estrutura de modelagem utilizada para essa análise.

#### 3.1 Dados

Os dados de atropelamento foram coletados do Sistema de Informações de Acidentes de Trânsito de Fortaleza (SIAT/FOR) para os anos de 2017 a 2019. Esses dados consistem em registros individuais com informações pessoais das vítimas, características dos veículos envolvidos e aspectos relacionados ao acidente. Para determinar as variáveis relacionadas a estrutura urbana utilizou-se a base georreferenciada de semáforos e equipamentos de fiscalização eletrônica fornecidos pelo órgão gestor de trânsito do município (AMC).

**Tabela 1:** Descrição das variáveis independentes

Variável	Descrição
<i>Pedestre</i>	
Sexo	1 - Homem (66%); 0 - Mulher (34%)
Ida_0_21	*Idade menor que 21 anos (16%);
Ida_22_40	1 - Idade entre 22 e 40 anos (31%); 0 - Outra
Ida_41_60	1 - Idade entre 41 e 60 anos (34%); 0 - Outra
Ida_60	1 - Idade maior que 60 anos (19%); 0 - Outra
<i>Acidente</i>	
Periodo_Ano	1 - Jan e Jul (16%); 0 - Fev a Jun e Ago a Dez (84%)
FDS	1 - Final de semana (30%); 0 - Dia útil (70%)
Noite	1 - 18 às 5hrs (43%); 0 - 5 às 18hrs (57%)
Hora_pico	1 - 7 às 9hrs e 17 às 19hrs (28%); 0 - Demais horas (72%)
M_Quadra	1 - Meio de quadra (84%); 0 - Interseção (16%)
Auto	*Carro de passeio (46%);
Veic_Pes	1 - Veículo pesado (8%); 0 - Outro
Moto	1 - Moto (46%); 0 - Outro
<i>Estrutura Urbana</i>	
V_Loc	*Via Local (49%);
V_Col	1 - Via coletora (4%); 0 - Outra
V_Art	1 - Via arterial (35%); 0 - Outra
V_Exp	1 - Via expressa (13%); 0 - Outra
Jur_M	*Jurisdição municipal (92%);
Jur_E	1 - Jurisdição estadual (3%); 0 - Outra
Jur_F	1 - Jurisdição Federal (5%); 0 - Outra
P_Equi_FE	1 - Presença de equipamento de fiscalização eletrônica em um raio de 100 metros do local do acidente (6%); 0 - Ausência (94%)
P_Semaf	1 - Presença de semáforos em um raio de 100 metros do local do acidente (30%); 0 - Ausência (70%)

\*Categoria de referência

As observações com campos sem informações referentes às variáveis explicativas investigadas foram eliminadas da amostra inicial (N = 4.658 observações), resultando em uma amostra de 2.771 observações de atropelamentos. A categoria pedestre ileso foi eliminada, pois apresentou somente um registro para esse período. Esse aspecto é bastante comum nos bancos de dados de acidentes, visto que acidentes sem vítimas feridas são menos prováveis de serem relatados, já que geralmente os órgãos responsáveis não são acionados para prestar atendimento. Além disso,

difícilmente um atropelamento resultará em vítima ileso, o que reduz também o número de registros desta categoria.

Algumas variáveis foram obtidas a partir da base georreferenciada. Para isso, criou-se um *buffer* com raio de 100 metros em torno do acidente. As variáveis foram divididas em três grupos: características do pedestre, aspectos relacionados ao acidente e características da estrutura urbana. A Tabela 1 apresenta as variáveis selecionadas, juntamente com as suas frequências relativas dentro da amostra utilizada.

### 3.2 Classificação da severidade

Um dos objetivos desse trabalho consiste na avaliação de diferentes configurações de classificação da gravidade das lesões sofridas pelos pedestres na modelagem da severidade. Para isso, três modelos categóricos foram estimados a partir da amostra de dados obtida: i) No Modelo 01, a variável dependente conta com quatro categorias, utilizando a classificação original do banco de dados: lesão leve, moderada, grave e fatal; ii) No Modelo 02 as lesões leves e moderadas foram agrupadas em uma única categoria devido a semelhança em termos de consequência, o que pode gerar divergências na classificação entre esses dois níveis pelos agentes responsáveis; iii) No Modelo 03 optou-se por unir as lesões graves e fatais em uma categoria, tomando como base os princípios dos Sistemas Seguros e Visão Zero, que almejam a redução de lesões graves e fatais no trânsito. A Tabela 02 apresenta a frequência dos níveis de severidade dos dados em cada configuração.

**Tabela 2:** Frequência dos níveis de severidade

<i>Modelo 01</i>	Leve (55%); Moderado (25%); Grave (12%); Fatal (8%)
<i>Modelo 02</i>	Leve/Moderado (80%); Grave (12%); Fatal (8%)
<i>Modelo 03</i>	Leve (55%); Moderado (25%); Grave/Fatal (20%)

Agresti (2006) recomenda verificar a quantidade de preditores possíveis para um modelo em caso de desbalanço nas categorias da variável resposta. Peduzzi *et al.* (1996) sugere que o modelo não deve conter mais do que  $n/10$  parâmetros, onde  $n$  representa o número de observações da categoria com a menor frequência, para evitar problemas de variâncias superestimadas e subestimadas e, portanto, cobertura pobre de intervalos de confiança baseados em testes de Wald. Desse modo, no Modelo 2, com o maior desbalanço na frequência das categorias, o número de preditores não deve ser superior a 22, o que de fato não ocorre como será observado nas próximas seções.

### 3.3 Desenvolvimento dos modelos categóricos e medidas de ajuste

Diferentes abordagens estatísticas foram aplicadas à análise da gravidade das lesões causadas por acidentes com pedestres (Savolainen *et al.*, 2011). Nesse trabalho, optou-se por utilizar um modelo não ordenado do tipo logit multinomial que permite efeitos variáveis mais flexíveis, pois não impõem nenhuma restrição aos parâmetros e efeitos marginais das variáveis (Abay, 2013). Além disso, esses modelos são menos suscetíveis à subnotificação. No Modelo Logit Multinomial (MLM) a estrutura geral usada para estimar a severidade inicia com a definição de uma função linear  $S$  que determina o resultado da lesão  $i$  para a observação  $n$  como:

$$S = \beta_i X_{in} + \varepsilon_{in} \quad (1)$$

onde  $\beta_i$  é um vetor de parâmetros estimados,  $X_{in}$  é um vetor de características observáveis que afetam a gravidade da lesão sustentada pela observação  $n$ , e  $\varepsilon_{in}$  é um termo de erro que é responsável por efeitos não observados (Washington *et al.*, 2011). A probabilidade para cada



nível de severidade é dada pela Equação 2, assumindo que  $\varepsilon_{in}$  é distribuído de forma idêntica e independente com uma distribuição de valor extremo do tipo 1. Quando um MLM é estimado, um nível de lesão é usado como um grupo de comparação e, portanto, seus coeficientes são definidos como zero. Neste estudo, a primeira categoria de cada modelo foi usada como categoria de referência.

$$P_n(i) = \frac{e^{\beta_i X_{in}}}{\sum_i e^{\beta_i X_{in}}} \quad (2)$$

Uma limitação do modelo MLM é que ele pode ser suscetível ao compartilhamento de efeitos não observados entre as categorias de severidade. Por exemplo, as categorias leve e moderado são próximas em termos de consequências, gerando muitas vezes até dificuldade do agente em classificar a lesão entre os dois níveis. Dessa forma, é possível que certos fatores não observados, como características do local, aspectos físicos do pedestre ou mesmo seu comportamento durante a travessia, que poderiam influenciar na severidade do acidente, tenham o mesmo efeito nas duas categorias. Matematicamente, isso corresponde à presença de correlação nos termos do erro entre duas ou mais categorias. Uma das premissas do modelo MLM é que a razão das probabilidades de quaisquer duas alternativas (categorias de respostas) para uma observação específica não é influenciada por outra alternativa, ou seja, assume-se que as categorias sejam independentes. Essa propriedade é denominada de Independência das Alternativas Irrelevantes (IAI) e se violada pode resultar em estimativas erradas dos coeficientes (Holdridge, *et al.*, 2005; Savolainen *et al.*, 2011).

Para avaliar a propriedade da IAI foi utilizado o teste de Hausman-McFadden. Esse teste baseia-se na comparação do modelo com todas as categorias da variável resposta (modelo completo) e outro modelo no qual se elimina os dados de uma determinada categoria (modelo reduzido). Se os parâmetros do modelo reduzido não mudarem significativamente em relação ao modelo completo, então não se pode rejeitar a hipótese nula de que a premissa do MNL se verifica (Hausman e McFadden, 1984).

Foram realizados ainda dois testes para avaliar a qualidade dos modelos: o teste de significância do modelo e o teste de Hosmer-Lemeshow. O primeiro compara o modelo proposto com o modelo contendo apenas o intercepto para verificar a hipótese de que as variáveis incluídas no modelo não melhoram significativamente seu poder preditivo quando comparado com o valor do intercepto (média). A estatística de teste corresponde à diferença entre a verossimilhança entre os dois modelos e a hipótese nula é que o modelo apenas com o intercepto tem um ajuste tão bom quanto o modelo com as variáveis explicativas.

O segundo teste é baseado na divisão da amostra segundo suas probabilidades ajustadas com base nos valores dos parâmetros estimados pela regressão. No caso do modelo multinomial, calcula-se a probabilidade estimada e atribui-se um score para cada observação. Em seguida, as observações são ordenadas com base nos scores obtidos e formam-se  $g$  grupos de tamanho igual, de modo que o 1º grupo contenha as observações com os menores scores e o último grupo as observações com os maiores scores. É recomendado pelo menos 10 grupos para o teste. Por fim uma tabela de contingência é criada com  $g$  linhas (número de grupos) e com  $c$  colunas (número de categorias da variável dependente). A estatística de teste é dada pela Equação 3, em que  $n_{vj}$  representa o número de observações pertencentes ao  $v$ -ésimo grupo e a  $j$ -ésima categoria e  $E_{vj}$  representa o número esperado de observações pertencentes ao  $v$ -ésimo grupo e a  $j$ -ésima categoria.  $C_g$  segue uma distribuição qui-quadrado com  $(g - 2)(c - 1) + (c - 2)$  graus

de liberdade. Um valor p não significativo indica que não há evidências de que as frequências observadas e esperadas sejam diferentes (ou seja, evidências de bom ajuste).

$$C_g = \sum_{v=1}^g \sum_{j=1}^c \frac{(n_{vj} - E_{vj})^2}{E_{vj}} \quad (3)$$

Além desses testes foram adotados ainda outros critérios de qualidade do ajuste dos modelos para comparar as três configurações de classificação da severidade: *Akaike Information Criterion* (AIC) e  $R^2$  de McFadden. Além de identificar um arranjo que melhore o ajuste do modelo, o interesse nessa comparação é avaliar se possíveis problemas relacionados a classificação utilizada no banco de dados pode alterar o entendimento dos resultados encontrados. A seleção das variáveis independentes incluídas nos modelos foi baseada na técnica *stepwise*, usando um nível de significância de 0,10, para reduzir a possibilidade de eliminar variáveis importantes. Os parâmetros foram estimados utilizando o método da máxima verossimilhança no programa estatístico R.

#### 4. RESULTADOS E DISCUSSÕES

Esta seção apresenta os resultados dos modelos estimados e discute a escolha da melhor configuração de classificação da severidade. Em seguida são apresentadas possíveis hipóteses sobre os efeitos dos fatores investigados na severidade dos acidentes de trânsito com pedestre.

##### 4.1 Classificação da severidade

A Tabela 3 apresenta os coeficientes estimados para os três modelos investigados e os indicadores de qualidade de ajuste. O primeiro teste realizado foi o teste de verossimilhança. Os três modelos analisados obtiveram significância aproximadamente igual a zero indicando que as variáveis explicativas melhoram o ajuste do modelo.

Pelo teste de Hosmer-Lemeshow, não há evidências para se rejeitar a hipótese nula de ajustamento nos três modelos estimados. O Modelo 02 apresentou uma significância maior que os demais, mas como mencionado, o teste baseia-se na comparação entre as frequências observadas e esperadas em grupos de observações, definidas pela probabilidade estimada. É possível que escolhas diferentes de categoria de referência, como no caso do Modelo 02, que tem como categoria base a junção dos níveis leve e moderado, resultem em conclusões diferentes para o teste. Fagerland *et al.*, (2008) destacam que a sensibilidade do teste à escolha da categoria de referência é geralmente pequena, mas grandes diferenças podem ocorrer em conjuntos de dados específicos. Dessa forma, não seria indicado associar ao Modelo 02 uma superioridade em relação aos outros pela maior significância do teste.

No modelo de regressão linear, o coeficiente de determinação ( $R^2$ ) resume a proporção de variância na variável dependente associada às variáveis independentes, com valores maiores de  $R^2$ , indicando que mais da variação é explicada pelo modelo. Já nos modelos de regressão com uma variável dependente categórica, os coeficientes de determinação  $R^2$  (denominado também de Pseudo -  $R^2$ ) representa a proximidade do modelo com o melhor ajuste (Greene e Hensher, 2010). O  $R^2$  de McFadden, por exemplo, é limitado entre 0 e 1, e mede alterações nas funções de probabilidade, mas não tem uma interpretação óbvia quando analisado isoladamente. Por outro lado, ele pode ser usado para comparar o desempenho entre modelos, com valores mais próximos de 1 associados a uma superioridade do modelo. É importante destacar que seus valores tendem a ser consideravelmente mais baixos do que aqueles esperados para o  $R^2$  da regressão linear. Valores entre 0,2 e 0,4 podem representar um ajuste excelente (Ortúzar e

Willumsen, 2011). No caso dos três modelos investigados ambos apresentaram valores relativamente baixos, porém o Modelo 02 apresentou o maior R<sup>2</sup>. Vale destacar ainda que o R<sup>2</sup> tende a aumentar com o número de variáveis preditoras e o Modelo 02 mesmo com o menor número de variáveis explicativas ainda demonstrou um desempenho melhor.

**Tabela 3:** Resultados da estimativa dos modelos logit multinomial

	Modelo 01			Modelo 02		Modelo 03	
	Moderado	Grave	Fatal	Grave	Fatal	Moderado	Grave/Fatal
	<i>coef</i>	<i>coef</i>	<i>coef</i>	<i>coef</i>	<i>coef</i>	<i>coef</i>	<i>coef</i>
Intercepto	-1,113	-2,254	-4,119	-2,562	-4,588	-1,111	-2,217
Periodo_Ano	0,276	-	-	-	-	0,275	-
FDS	-	0,291	0,348	0,282	0,359	-	0,314
Noite	0,180	0,551	0,595	0,493	0,548	0,180	0,566
Sexo	-	-	0,509	-	0,560	-	0,210
Ida_41_60	0,294	-	1,055	-	0,963	0,294	0,401
Ida_60	0,283	-	2,342	-	2,240	0,282	1,081
M_Quadra	0,310	-	-	-	-	0,310	-
Veic_Pes	-	0,796	1,407	0,879	1,402	-	1,050
Moto	-0,215	-	-	-	-	-0,215	-0,203
V_Art	-	-	-	-	0,379	-	-
V_Exp	-	0,460	-	0,509	0,546	-	0,388
Jur_E	-	0,883	1,235	1,070	1,256	-	1,015
Jur_F	-	1,162	1,311	1,189	1,313	-	1,214
<i>Teste de Verossimilhança</i>	$\chi^2 = 371,97$ ; valor p = 0,000			$\chi^2 = 339,7$ ; valor p = 0,000		$\chi^2 = 269,67$ ; valor p = 0,000	
<i>Teste de Hosmer-Lemeshow</i>	$\chi^2 = 26,24$ ; valor p = 0,341			$\chi^2 = 7,39$ ; valor p = 0,965		$\chi^2 = 9,93$ ; valor p = 0,870	
<i>R<sup>2</sup> McFadden</i>	0,06			0,10		0,05	
<i>AIC</i>	5998			3221,9		5323,8	

- Não significativo para um nível de significância de 0,10.

O AIC é uma medida de ajuste, baseada na verossimilhança, frequentemente utilizada para comparar modelos não aninhados quando não há critério ou regra óbvia para comparar ajustes. Valores menores do AIC são preferíveis, contudo, é importante destacar que a medida penaliza o modelo por ter um grande número de parâmetros (Greene e Hensher, 2010). O Modelo 01 tem uma variável explicativa a mais que o Modelo 03 e mesmo com a penalização apresentou um AIC bem próximo do terceiro modelo. Já o Modelo 02 com a diferença de apenas uma variável preditora em relação ao Modelo 03 e duas variáveis em relação ao Modelo 01 apresentou um AIC bem menor. Isso demonstra que essa diferença não se deve somente pela penalização imposta pela medida, mas também por um melhor ajuste do Modelo 02.

Como já mencionado, o modelo multinomial tem como pressuposto a IAI. Para verificar se os modelos estimados neste estudo atendem essa premissa foi realizado o teste de Hausman-McFadden. Os resultados do teste são apresentados na Tabela 4. Cada um dos três modelos testados foi comparado a um modelo estimado com um subconjunto do grupo de categorias da variável dependente. Nenhum dos casos atendeu ao pressuposto assintótico do teste ( $\chi^2$  negativo). Hausman e McFadden (1984) sugerem que um resultado negativo do  $\chi^2$  ainda pode ser tomado como suporte à hipótese nula por este ser um teste robusto. Entretanto, os resultados encontrados podem indicar problemas devido a uma dependência entre as categorias, ocasionando por exemplo, estimativas incorretas dos coeficientes e das probabilidades do modelo. Neste caso seria recomendada a utilização de outra abordagem estatística como modelos logit de parâmetros aleatórios.



**Tabela 4:** Teste de Hausman-McFadden

	Categoria Omitida	X <sup>2</sup>	gl	Evidência
Modelo 01	Fatal	-3,02	26	Não rejeitar a H0
	Grave	-14,3	26	Não rejeitar a H0
	Moderada	-1,09	26	Não rejeitar a H0
Modelo 02	Fatal	-0,21	11	Não rejeitar a H0
	Grave	-0,37	11	Não rejeitar a H0
Modelo 03	Grave/Fatal	-0,7	13	Não rejeitar a H0
	Moderada	-0,27	13	Não rejeitar a H0

Nota: se  $\chi^2 < 0$ , o modelo estimado não atende os pressupostos assintóticos do teste; valor p superior a um nível de significância de 0,10 em todos os testes.

Levando-se em consideração os testes realizados e as métricas consideradas para a comparação entre os três modelos, o Modelo 02 com três níveis de severidade (leve/moderado, grave e fatal) apresentou o melhor desempenho. Diante disso, a análise preliminar do efeito dos fatores de risco associados a gravidade das lesões com pedestres será feita a partir desse modelo.

#### 4.2 Avaliação dos fatores de risco

Os parâmetros estimados permitem analisar o efeito das variáveis entre os níveis de severidade, por exemplo, um sinal positivo indica um aumento da probabilidade do nível mais alto, enquanto um sinal negativo uma redução da probabilidade desse nível. Entretanto, não é possível identificar a magnitude desse efeito de forma direta. Como todas as variáveis explicativas são categóricas, a magnitude é analisada pela variação percentual na probabilidade quando uma variável é alterada de 0 para 1, mantendo-se as variáveis restantes constantes. Esse valor é denominado de pseudo-elasticidade da probabilidade. A Tabela 5 apresenta os resultados da pseudo-elasticidade para as variáveis investigadas.

**Tabela 5:** Pseudo-elasticidade das variáveis

	Leve/Moderado	Grave	Fatal
FDS	-3%	29%	39%
Noite	-5%	56%	64%
Veic_Pes	-11%	113%	260%
Sexo	-1%	*-1%	74%
Ida_41_60	-1%	*-1%	158%
Ida_60	-7%	*-7%	771%
V_Art	0%	*0%	45%
V_Exp	-5%	58%	64%
Jur_E	-14%	151%	203%
Jur_F	-16%	176%	213%

Nota: \*Parâmetro não significativo

Em relação aos aspectos ligados ao acidente os resultados sugerem que há chances maiores de vítimas graves ou fatais em acidentes ocorridos nos finais de semana. Isso pode estar associado com o menor fluxo de veículos motorizados durante o sábado e domingo, o que possibilita velocidades maiores dos condutores. Além disso, um acidente durante o período noturno tem 64% de chances a mais de ocasionar lesões fatais em relação a um acidente durante o dia. A suposição levantada para explicar esse fenômeno é que durante a noite, além das maiores velocidade dos veículos devido ao menor volume, a baixa iluminação pode dificultar a visibilidade dos condutores e pedestres. Resultados semelhantes foram verificados nos trabalhos de Chen e Fan (2019) e Jang *et al.* (2013). Quanto ao tipo de veículo, a variável *Veic\_Pes* (veículos pesados), que incluem caminhões, caminhonetes e ônibus, influencia significativamente a severidade das lesões com pedestres. Esse tipo de veículo tem cerca de

duas vezes mais de chances de ocasionar uma lesão grave e três vezes mais chances de gerar uma lesão fatal no pedestre. Essa relação pode ser encontrada em estudos como os de Wang *et al.*, (2013) e Aziz *et al.*, (2013);

Em relação às características do pedestre, a idade e o sexo foram significativos. Pedestres com idade mais avançada estão mais propensos a sofrerem lesões mais graves quando envolvidos em um acidente de trânsito, de acordo com a Tabela 5. Esse fato é também demonstrado nos trabalhos de Abay (2013) e Jang *et al.* (2013). Quando comparados a pedestres com idade inferior a 21 anos (crianças e jovens), vítimas com idade superior a 60 anos tem quase nove vezes mais chances de sofrerem uma lesão fatal. Já quanto ao sexo do pedestre o risco maior está associado aos homens. Uma hipótese para este resultado pode estar relacionada a uma tendência de um comportamento mais arriscado por parte dos homens, especialmente em relação às brechas disponíveis para travessia (Torres *et al.*, 2020). Esse resultado pode estar ainda associado a um possível consumo de bebidas alcólicas pelo pedestre do sexo masculino, que pode ser corroborada pelos resultados encontrados para os fatores fim de semana (*FDS*) e período noturno (*Noite*) os quais apresentaram também uma elevada relação com lesões mais graves.

A classificação da via onde ocorreu o acidente é geralmente utilizada como uma variável *proxy* para a velocidade. Apesar do Código de Trânsito Brasileiro estabelecer o limite de 80km/h para vias expressas urbanas, em Fortaleza adota-se o limite de 60km/h para essas vias, assim como para as vias arteriais. Já nas vias coletoras o limite é de 40km/h e em vias locais 30km/h. Vale destacar que a cidade iniciou em 2018 um processo de readequação dos limites de velocidade de algumas vias arteriais de 60 para 50km/h. De acordo com a Tabela 5, ambas as vias arteriais e expressas estão associadas a um risco maior de lesões graves e fatais nos pedestres, justamente por permitirem velocidades maiores.

Vias federais e estaduais são caracterizadas por um fluxo maior de veículos, especialmente de veículos pesados. Além disso, são caracterizadas por um padrão geométrico mais favorável, com faixas mais largas, presença de acostamento, entre outros aspectos, o que pode levar a excessos de velocidade, principalmente em trechos mais distantes de equipamentos de fiscalização eletrônica. Isso pode explicar as maiores chances de acidentes com lesões severas para os pedestres. De acordo com a Tabela 5, um acidente em uma via estadual ou federal tem cerca de três vezes mais chances de resultarem em uma lesão fatal do pedestre, em comparação com um acidente numa via municipal.

## 6. CONSIDERAÇÕES FINAIS

Esse trabalho apresentou uma análise dos fatores de risco associados a severidade de acidentes com pedestres. Foram investigadas variáveis relacionadas ao pedestre, ao momento do acidente, e aspectos da estrutura urbana. Além disso, foram avaliadas três diferentes configurações de classificação da severidade na modelagem desses fatores de risco. Para isso foram estimados modelos logit do tipo multinomial utilizando uma amostra com 2.771 observações de atropelamentos de pedestres coletadas do Sistema de Informações de Acidentes de Trânsito de Fortaleza (SIAT-FOR) para os anos de 2017 a 2019.

Os resultados da análise comparativa entre as três diferentes configurações da variável dependente (severidade das lesões) demonstraram que o modelo com níveis de severidade leve e moderado unidos em uma única categoria e mais os níveis grave e fatal em categorias

diferentes apresentou o melhor desempenho. Diante disso, esse modelo foi usado para avaliar os efeitos dos fatores de risco na severidade das lesões.

Os resultados do processo de calibração indicaram que as variáveis relacionadas ao entorno do local dos acidentes, como a presença de fiscalização eletrônica e a presença de semáforos não influenciaram significativamente na severidade dos acidentes. Por outro lado, percebeu-se que alguns fatores como idade mais avançada do pedestre, período noturno, finais de semana e vias que permitem velocidades maiores dos veículos estão associados a lesões mais severas do pedestre quando envolvido em um acidente de trânsito.

Entre as limitações encontradas nesse estudo pode-se destacar a redução no número de observações do banco de dados devido a problemas de georreferenciamento dos acidentes e informações faltantes de algumas observações. Além disso, alguns resultados encontrados poderiam facilmente serem atestados com a incorporação de mais informações no banco de dados, tais como um teste de alcoolemia nos envolvidos, especialmente naqueles acidentes com vítimas graves ou fatais, e pelo menos a velocidade média dos veículos no momento do acidente, entre outros aspectos.

Em relação ao tipo de modelagem utilizada, os modelos logit multinomiais flexibilizam algumas restrições dos modelos ordenados tradicionais, todavia não consideram a natureza ordenada. O teste realizado para verificar a premissa da independência das alternativas irrelevantes não demonstrou resultados completamente satisfatórios. Assim, em trabalhos futuros recomenda-se explorar técnicas estatísticas que considerem tanto a natureza ordenada da variável dependente quanto uma flexibilidade dos parâmetros estimados (Savolainen *et al.*, 2011). Essas técnicas possibilitam tanto um ajuste melhor do modelo aos dados quanto um aprofundamento na análise das variáveis explicativas.

#### Agradecimentos

O presente trabalho foi realizado com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001, do Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq e da Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico (Funcap).

#### REFERÊNCIAS BIBLIOGRÁFICAS

- Abay, K. A. (2013) Examining pedestrian-injury severity using alternative disaggregate models. *Research in Transportation Economics*, v. 43, n. 1, p. 123–136.
- Agresti, A. (2006) *An Introduction to Categorical Data Analysis*. 2.ed., New Jersey: Wiley.
- Amoh-Gyimah, R.; Saberi, M. e Sarvi, M. (2016) Macroscopic modeling of pedestrian and bicycle crashes: A cross-comparison of estimation methods. *Accident Analysis and Prevention*, v. 93, p. 147–159.
- Anderson, R.W.G.; Mclean, A.J; Farmer, M.J.B; Lee, B.H. e Brooks, C.G. (1997) Vehicle Travel Speeds and The Incidence of Fatal Pedestrian Collisions. *Accident Analysis and Prevention*, v. 29, n. 5, p. 667–674.
- Aziz, H; M. A.; Ukkusuri, S. V.; Hasan, S. (2013) Exploring the determinants of pedestrian-vehicle crash severity in New York City. *Accident Analysis and Prevention*, v. 50, p. 1298–1309, 2013.
- Bhat, C. R.; Astroza, S. e Lavieri, P. S.(2017) A new spatial and flexible multivariate random-coefficients model for the analysis of pedestrian injury counts by severity level. *Analytic Methods in Accident Research*, v.16, p. 1- 22.
- Belin, M. A.; Tillgren, P.; Vedung, E.. (2012) Vision Zero—a road safety policy innovation. *International journal of injury control and safety promotion*, v. 19, n. 2, p. 171-179.
- Chen, Z. e Fan, W. (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.
- Clifton, K. J.; Burnier, C. V. e Akar, G..(2009) Severity of injury resulting from pedestrian-vehicle crashes: What can we learn from examining the built environment? *Transportation Research Part D: Transport and Environment*, v. 14, n. 6, p. 425–436.
- Eluru, N.;Bhat, C. R. e Hensher, D. A.(2008) A mixed generalized ordered response model for examining

- pedestrian and bicyclist injury severity level in traffic crashes. *Accident Analysis and Prevention*, v. 40, n. 3, p. 1033–1054.
- Fagerland, M. W.; Hosmer, D. W. e Bofin, A. M. (2008) Multinomial goodness-of-fit tests for logistic regression models. *Statistics in Medicine*, v. 27, n. 21, p. 4238-4253
- Garder, P. E. (2004) The impact of speed and other variables on pedestrian safety in Maine. *Accident Analysis and Prevention*, v. 36, n. 4, p. 533–542.
- Gehl, Jan. (2010). *Cidades para Pessoas* (2ªed.). Editora Perspectiva, São Paulo.
- Greene, W. H., e Hensher, D. A. (2010). *Modeling ordered choices: A primer*. Cambridge University Press.
- Hausman, Jerry; McFadden, Daniel. (1984) Specification tests for the multinomial logit model. *Econometrica: Journal of the Econometric Society*, p. 1219-1240.
- Holdridge, J. M.; Shankar, V. N e Ulfarsson, G. F. (2005) The crash severity impacts of fixed roadside objects. *Journal of Safety Research*, v. 36, n. 2, p. 139-147.
- Jang, Kitae; Park, Shin; Kang, Sanghyeok; Song, Ki Kang, Seungmo e Chung, Sungbong. (2013) Evaluation of pedestrian safety. *Transportation Research Record*, n. 2393, p. 104–116.
- Kim, J. K.; Ulfarsson, G. F.; Shankar, V. N. e Mannering, F. L. (2010) A note on modeling pedestrian-injury severity in motor-vehicle crashes with the mixed logit model. *Accident Analysis and Prevention*, v. 42, n. 6, p. 1751–1758.
- Kwigizile, V.; Sando, T. e Chimba, D.. (2011) Inconsistencies of ordered and unordered probability models for pedestrian injury severity. *Transportation Research Record*, n. 2264, p. 110–118
- Lee, C. e Abdel-Aty, M.. (2005) Comprehensive analysis of vehicle-pedestrian crashes at intersections in Florida. *Accident Analysis and Prevention*, v. 37, n. 4, p. 775–786.
- Li, D; Ranjitkar, P.; Zhao, Y.; Yi, H. e Rashidi, S. (2016). Analyzing pedestrian crash injury severity under different weather conditions. *Traffic Injury Prevention*, v.18, n.4, p. 427–430.
- Mannering, F. L.; Bhat, C. R. Analytic methods in accident research: Methodological frontier and future directions. *Analytic methods in accident research*, v. 1, p. 1-22, 2014.
- Organização Pan-Americana de Saúde (OPAS). *Gestão da velocidade: um manual de segurança viária para gestores e profissionais da área*. 2012.
- Ortúzar, J. D.; Willumsen, L. G. (2011) *Modelling Transport*. 4 ed., London: Wiley, 2011.
- Peduzzi, P. N.; Concato, J.; Kemper, E.; Holford, T. R. e Feinstein, A. (1996). A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology*, v.99, p. 1373–1379.
- Pour-Rouholamin, M, e Zhou, H.. (2016) Investigating the risk factors associated with pedestrian injury severity in Illinois. *Journal of Safety Research*, v. 57, p. 9–17.
- Rosén, E.; Stigson, H. e Sander, U.. (2011) Literature review of pedestrian fatality risk as a function of car impact speed. *Accident Analysis and Prevention*, v. 43, n. 1, p. 25–33.
- Rothman, L.; Howard, A. W.; Camden, A.; Macarthur, C. (2012) Pedestrian crossing location influences injury severity in urban áreas. *Injury Prevention*, v. 18, n.6, p. 365-370.
- 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.
- Shinar, D. (2017). *Traffic safety and human behavior* (2ª ed). Emerald Group Publishing.
- Sze, N. N. e Wong, S. C. (2007) Diagnostic analysis of the logistic model for pedestrian injury severity in traffic crashes. *Accident Analysis and Prevention*, v. 39, n. 6, p. 1267–1278.
- 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.
- Torres, C.; Sobreira, L.; Castro-Neto, M.; Cunto, F.; Vecino-Ortiz, A.; Allen, K.; Hyder, A e Bachani, A. (2020) Evaluation of Pedestrian Behavior on Mid-block Crosswalks: A Case Study in Fortaleza—Brazil. *Frontiers in Sustainable Cities*, v.2, p. 1 – 6.
- Tsui, K. L.; So, F. L.; Sze, N. N.; Wong, S. C.; Leung, T. F. (2009) Misclassification of injury severity among road casualties in police reports. *Accident Analysis and Prevention*, v. 41, p. 84 – 89.
- Wang, Y. Y.; Haque, M. M.; Chin, H. C. e Yun, J. G. J. (2013) Injury severity of pedestrian crashes in Singapore. *Australasian Transport Research Forum*, ATRF 2013 – Proceedings.
- Washington, P. S.; Karlaftis G. M. e Mannering F. L. (2003) *Statistical and Econometric Methods for Transportation Data Analysis*. Chapman & Hall/CRC, Nova Iorque, 2003
- Welle, B. et al (2018). *Sustentável e seguro - Visão e Diretrizes para Zerar as Mortes no Trânsito*. World Resources Institute, Washington