

# Análise da severidade dos acidentes com motocicletas utilizando modelos probit e logit ordenados

Carlos Henrique Loiola Coutinho<sup>1</sup>, Flávio José Craveiro Cunto<sup>2</sup> e Sara Maria Pinho Ferreira<sup>3</sup>

**Resumo:** O ambiente de circulação rodoviário brasileiro tem apresentado um aumento desproporcional da utilização de motocicletas ao longo dos últimos 15 anos. A mesma tendência acompanha o número de acidentes de trânsito da categoria em parte por sua relativa vulnerabilidade. O objetivo deste trabalho é identificar fatores que influenciam a severidade dos acidentes com motocicletas nas vias urbanas de Fortaleza com a utilização dos modelos ordenados do tipo probit e logit. Foram desenvolvidos modelos ordenados categóricos utilizando uma amostra com 3.232 observações de acidentes de trânsito de 2004 a 2011. A severidade resultante dos acidentes de trânsito foi classificada em quatro categorias. Os resultados da calibração dos modelos indicaram que motociclistas que utilizam capacete e pilotam durante o dia têm um menor risco de sofrer lesões mais graves. De outra forma, motociclistas mais velhos e que sofreram acidente ocorrido em finais de semana apresentaram um maior risco de lesões mais graves.

*Palavras-chave:* modelos ordenados, severidade dos acidentes com motocicletas, modelos categóricos probit e logit.

**Abstract:** The Brazilian roadway traffic environmental has experienced a disproportionate development in the use of motorcycles over the last 15 years. The same trend has been observed for the number of traffic accidents in the category in part by their relative vulnerability. The objective of this study is to identify factors that influence the severity of motorcycle accidents in urban streets of Fortaleza with the use of ordered probit and logit models. Ordered categorical models using a sample of 3,232 observations of traffic accidents from 2004 to 2011 were developed. The resulting severity of traffic accidents was classified into four categories. The results of the model calibration indicated that riders who use helmets and traveled during the day have a lower risk of suffering severe injuries. Otherwise, older riders and those involved in an accident on the weekend had a higher risk of suffering more severe injuries.

*Keywords:* ordered models, severity of motorcycle crashes, probit and logit categorical models.

## 1. INTRODUÇÃO

O ambiente de circulação brasileiro tem apresentado uma evolução desproporcional da utilização de motocicletas ao longo dos últimos 15 anos. Em termos de representação na frota de veículos, as motocicletas passaram de 4,5 milhões em 2001 para 19,9 milhões em 2012 (aumento de 4,4 vezes) enquanto a frota total de veículos nesse mesmo período evoluiu de 34,9 milhões para 76,1 milhões (aumento de 2,2 vezes), passando assim de 14,2% do total de veículos em 2001 para 26,2 % em 2012 (INCT, 2013).

Alguns dos fatores incentivadores desse crescimento listados por Holz et al. (2011) estão ligados às características físicas e operacionais dessa categoria veicular que permitem otimizar aspectos da circulação em áreas urbanas, como por exemplo menor tempo de viagem em comparação ao modo privado e ao transporte público, menor custo de aquisição e operação e facilidade para estacionamentos. Sob essa ótica, é possível afirmar que o aumento relativo da frota de motocicletas deve perdurar nos próximos anos, principalmente caso seja mantido o processo histórico de urbanização e adensamento das cidades brasileiras e o cenário socioeconômico recente marcado pelo aumento da renda e a facilidade de crédito.

O crescimento da utilização das motocicletas e a inerente vulnerabilidade de seus usuários aliados à utilização desse modo em atividades que estimulam o comportamento de risco por parte dos condutores, além de aumentar a exposição às condições de tráfego (quilômetros percorridos por dia), compõem um cenário potencial para a ocorrência de acidentes de trânsito dessa categoria. No Brasil, os acidentes com vítimas fatais envolvendo motocicletas correspondem a aproximadamente 34% do total, enquanto que o de automóveis em torno de 29%. A comparação relativa à frota das duas categorias fornece indícios objetivos do risco de morte bastante superior para o caso das motocicletas (Waiselfisz, 2013).

Sob a premissa de que a análise metódica das informações disponíveis nos bancos de dados de acidentes de trânsito pode aumentar a compreensão dos principais fatores contribuintes para a ocorrência de acidentes e de suas consequências, diversos pesquisadores vêm desenvolvendo esforços de modelagem estatística voltados à estimação do grau de severidade como variável de resposta qualitativa (Ferreira, 2010; Kockelman e Kweon, 2002). Esses modelos também denominados modelos categóricos ou modelos de escolha discreta têm sido utilizados com relativo sucesso na modelagem da demanda por transportes em sua etapa de divisão modal, passaram a ser explorados na modelagem da segurança viária em suas estruturas mais convencionais: binário/multinomial logit ou probit (O'Donnell e Connor, 1996; Ferreira, 2010; Kononen et al, 2011; Ye e Lord, 2014).

Para o caso específico dos acidentes de trânsito em que as categorias de severidade seguem um grau sequencial ou ordenado, os modelos categóricos ordenados (ou ordinais) representam melhor esse tipo de estrutura quando comparados aos modelos categóricos tradicionais (Bajra-

<sup>1</sup> Carlos Henrique Loiola Coutinho, Universidade Federal do Ceará Departamento de Engenharia de Transportes. (c.henrique.loiola@gmail.com)

<sup>2</sup> Flávio José Craveiro Cunto, Universidade Federal do Ceará Departamento de Engenharia de Transportes. (flaviocunto@det.ufc.br)

<sup>3</sup> Sara Maria Pinho Ferreira, Universidade do Porto Faculdade de Engenharia. (sara@fe.up.pt)

charya, 2013; Mannering e Bhat, 2014). Diante da contextualização da problemática apresentada este trabalho tem como objetivo identificar fatores que influenciam a severidade dos acidentes com motocicletas nas vias urbanas de Fortaleza com a utilização dos modelos ordenados do tipo probit e logit.

## 2. MODELOS CATEGÓRICOS APLICADOS À SEGURANÇA VIÁRIA

Os modelos categóricos são muito utilizados em diversas áreas de estudo como: social, educacional e biomédica. As variáveis categóricas podem ser nominais, quando as categorias não têm uma ordenação natural, ou ordinais quando existe uma ordenação natural. Quando a variável resposta categórica possui mais de 2 categorias, utiliza-se o modelo de múltiplas categorias. Nos modelos com variáveis nominais, costuma-se modelar a probabilidade da variável resposta estar em uma determinada categoria. Já nos modelos com variáveis ordenadas, costuma-se modelar a probabilidade acumulada da variável resposta estar numa determinada categoria ou abaixo dela, ou seja, usa-se o modelo de ligação acumulado (Veloza, 2009).

No estudo da severidade dos acidentes, os modelos utilizados para a análise da gravidade dos acidentes são os modelos ordenados, pois as categorias dos acidentes possuem uma ordenação tais como: “danos materiais”, “ferido leve”, “ferido grave”, e “morto”. O modelo ordenado pode ser utilizado com uma estrutura logit, quando a distribuição considerada é a logística, ou probit, quando a distribuição é normal (Ferreira, 2010). Devido à rigidez do modelo ordenado, por considerar as mesmas variáveis explicativas para todos os níveis de gravidade, alguns autores analisaram também a aplicação do modelo multinomial logit ou probit. No entanto, estes modelos ignoram a natureza ordenada da variável dependente (Abdel-Aty, 2003).

O’Donell e Connor (1996) apresentaram uma das primeiras aplicações dos modelos probit e logit ordenados para estimar a gravidade dos acidentes de trânsito. Dados de aproximadamente 18 mil acidentes ocorridos em 1991 em vias públicas de um condado na Austrália foram utilizados para estimar os dois modelos. Os resultados mostraram que aumento na idade da vítima e na velocidade estimada do veículo no momento da colisão aumentam a probabilidade de lesões graves e morte. Outros fatores que tiveram grande relevância nas probabilidades foram o lugar dos ocupantes do veículo, nível de álcool no sangue, tipo de veículo e tipo de colisão.

Abdel-Aty (2003) aplicou modelos probit ordenados para analisar a severidade de acidentes em múltiplos locais. Foram desenvolvidos modelos para seções de rodovias, interseções sinalizadas e praças de pedágio no estado da Flórida. Todos os modelos mostraram significância na idade do motorista, gênero, uso do cinto de segurança, ponto do impacto, velocidade e tipo de veículo no nível de severidade do acidente.

Holz et al. (2011) realizaram uma modelagem dos acidentes envolvendo motociclistas em Porto Alegre. Foram utilizadas duas técnicas de análise: Regressão logística, para identificação das variáveis influentes; e Regressão Linear Múltipla, para caracterização dos efeitos das variáveis

significativas na taxa de accidentalidade. Os resultados indicaram que os riscos de acidentes com motociclistas aumentam com o número de veículos, a presença de semáforos aumenta significativamente o risco dos motociclistas e a condição de tempo nublado e chuvoso contribui para redução da taxa de acidentes de motociclistas.

Chung, Song e Yoon (2013) fizeram um estudo da severidade de acidentes em “motoboys” na área metropolitana de Seul aplicando modelos probit ordenados. A maioria dos resultados foi consistente com os casos gerais dos acidentes de motocicletas: acidentes entre motos de entrega e com veículos de carga especiais tiveram o efeito mais visível na gravidade dos acidentes, seguido por infrações envolvendo motoristas alcoolizados e manobras perigosas. A pressão para fazer uma entrega rápida também foi um fator que pode levar a aumentar a probabilidade de acidentes graves.

Os modelos ordenados logit e probit são empregados para estimar valores de variáveis latentes ( $y_i^*$ ) através de equações lineares. As variáveis latentes não são observadas diretamente nos dados disponíveis sendo segregadas por intervalos numéricos ordenados de acordo com categorias da variável observada ( $y_i$ ). A estrutura geral do modelo pode ser descrita por:

$$y_i^* = \beta' x_i + \varepsilon_i, i = 1, \dots, n \quad (1)$$

em que o vetor  $x_i$  é o conjunto de  $k$  covariáveis supostamente independentes do erro  $\varepsilon_i$  e  $\beta$  é o vetor de  $k$  parâmetros a serem estimados. Para o caso específico desse trabalho,  $y_i^*$  é a variável latente contínua observada em sua forma discreta em quatro categorias ligadas ao grau de lesão dos motociclistas envolvidos em acidentes, classificadas de acordo com o seguinte critério:

$$\begin{aligned} y_i &= 0 \text{ se } -\infty < y_i^* \leq \mu_0 \text{ (Ileso),} \\ y_i &= 1 \text{ se } \mu_0 < y_i^* \leq \mu_1 \text{ (Ferido leve),} \\ y_i &= 2 \text{ se } \mu_1 < y_i^* \leq \mu_2 \text{ (Ferido grave),} \\ y_i &= 3 \text{ se } \mu_2 < y_i^* \leq +\infty \text{ (Vítima fatal),} \end{aligned} \quad (2)$$

em que  $\mu_i$  são os limites de cada categoria proposta a serem determinados durante o processo de calibração do modelo. Assim, a probabilidade de uma vítima do acidente  $i$  sofrer uma lesão de gravidade  $j$  pode ser representada por:

$$\begin{aligned} P(y_i = 0 | x_i) &= F(\mu_0 - \beta' x_i) \\ P(y_i = j | x_i) &= F(\mu_j - \beta' x_i) - F(\mu_{j-1} - \beta' x_i), \quad j = 1, 2 \\ P(y_i = 3 | x_i) &= 1 - F(\mu_2 - \beta' x_i) \end{aligned} \quad (3)$$

em que  $F(\cdot)$  é a função de distribuição cumulativa do termo de erro aleatório  $\varepsilon_i$  estimado em  $x$ . Essas probabilidades serão positivas se os limites satisfizerem a restrição  $\mu_0 < \mu_1 < \mu_2$ . O modelo ordenado logit é quando a função  $F(\cdot)$  utilizada é a distribuição logística e o modelo ordenado probit quando  $F(\cdot)$  é a cumulativa normal (Greene e Hensher, 2010).

Os modelos são calibrados utilizando o método da máxima verossimilhança que estima os valores dos diferentes parâmetros do modelo estatístico de modo a maximizar

a probabilidade dos dados observados. Note-se, no entanto, que na análise de modelos probabilísticos não-lineares como é o caso do modelo ordenado, os parâmetros estimados não representam o efeito marginal. Por esse fato, na análise dos valores calibrados deve-se considerar adicionalmente o cálculo dos efeitos marginais e/ou das probabilidades associadas a cada categoria.

Os parâmetros calibrados são analisados quanto à sua significância estatística com base nos níveis de confiança através do teste *t-student*. A validade dos modelos pode ser avaliada através de um teste de significância e do teste de linhas paralelas. O primeiro teste compara a verossimilhança do 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. O teste de linhas paralelas tem como objetivo avaliar a necessidade de se flexibilizar os parâmetros limites  $\mu_i$  através da análise do ajuste da função de ligação comum às várias categorias da variável dependente.

### 3. PROCEDIMENTO METODOLÓGICO

Nesta seção são descritas a metodologia de coleta e análise dos dados, bem como a codificação e resumo das variáveis. Em seguida são explicadas as etapas de desenvolvimento e avaliação da qualidade dos modelos propostos.

#### 3.1. Coleta e Análise das Informações de Acidentes de Trânsito

Os dados foram coletados do banco de dados do Sistema de Informações de Acidentes de Trânsito de Fortaleza (SIATFOR) para os anos de 2004 a 2011, sendo registrados um total de 62.338 acidentes de trânsito. A filtragem inicial dessa amostra foi realizada para considerar somente os acidentes envolvendo pelo menos uma motocicleta restando um total de 4.899 observações.

O SIATFOR disponibiliza um conjunto de informações sobre o acidente de trânsito além das informações básicas como dia, hora, local e tipo do acidente que engloba características das vias, veículos, meio ambiente e os condutores envolvidos. Em sua rotina operacional o SIATFOR agrega informações advindas de várias fontes diferentes, além dos boletins de ocorrência coletados por seus agentes de trânsito.

Se por um lado, a presença de outras fontes de informação aumenta a abrangência dos bancos de dados, por outro lado, observa-se uma heterogeneidade no preenchimento de alguns campos presentes no sistema. Desta forma, considerando a qualidade das informações presentes e o escopo deste estudo, as variáveis relacionadas ao uso do capacete, hora da ocorrência, dia da semana, condições do pavimento da hora do acidente, uso da motocicleta, idade e sexo dos condutores foram utilizadas

De forma complementar à filtragem inicial, foram eliminadas as observações com campos sem informações referentes às variáveis explicativas mencionadas. Após esse processo, restaram 3.238 observações para o estudo. Do total de acidentes, 17,3% foram classificados como somente com danos materiais, ( $y_i = 0$ ); 68,2% foram classificados como tendo ao menos um ferido leve ( $y_i = 1$ ); 13,1% foram

classificados tendo ao menos um ferido grave ( $y_i = 2$ ), e 1,4% como tendo ao menos uma vítima fatal ( $y_i = 3$ ).

#### 3.2. Codificação e Resumo das Variáveis Explicativas

Conforme descrição anterior, em virtude da disponibilidade e confiabilidade dos dados, foram escolhidas as seguintes variáveis explicativas para a modelagem: capacete, luz do dia, final de semana, idade, sexo, superfície e categoria. Para o processo de calibração do modelo as variáveis foram codificadas em variáveis categóricas.

A variável *capacete* foi codificada em: (1 = com capacete) e (0 = sem capacete). A variável *luz do dia* é referente ao horário entre 06:00 e 18:00, a qual foi codificada como (1) e (0) o horário entre 18:00 e 06:00, com pouca luminosidade. A *idade* dos condutores foi dividida em quatro categorias, sendo de 0 a 20 anos (0), de 21 a 40 anos (1), de 41 a 60 anos (2) e acima de 60 (3).

A variável final de semana (*fds*) foi dividida em dois grupos e codificada como 0 para os acidentes ocorridos de segunda a sexta-feira e 1 para os acidentes ocorridos aos finais de semana. Em relação ao gênero dos condutores, a variável *sexo* foi codificada como (Masculino = 1) e (Feminino = 0). A variável *categoria* diz respeito ao uso da motocicleta e foi codificada como: (Aluguel = 1) e (Particular = 0). Superfície foi codificada como: (Molhado = 1) e (Seco = 0). A gravidade é a variável dependente e foi classificada em 4 categorias: ileso (0), leve (1), grave (2) e fatal (3). A Tabela 1 apresenta a estatística descritiva das variáveis envolvidas no estudo.

#### 3.3. Desenvolvimento dos Modelos Categóricos

Baseado nos objetivos do estudo foram estimados os modelos ordenados do tipo logit e probit com o emprego da função *<regressão ordinal>* do aplicativo SPSS©. A calibração inicial com todas as variáveis explicativas disponíveis (Modelo 1) apresentou as variáveis *superfície*, *sexo* e *categoria*, como não significativas ( $\alpha = 0,05\%$ ). Após a estimação do Modelo 1 foram retiradas as variáveis não significativas, sendo estimados novos coeficientes para o modelo com as variáveis *capacete*, *luz do dia*, *idade* e *fds*, denominado Modelo 2. Os principais parâmetros dos modelos calibrados estão resumidos na Tabela 2.

Observou-se que a retirada das variáveis não significativas quase não alterou o valor dos parâmetros. Houve apenas uma pequena alteração no valor do erro padrão das variáveis, porém as que eram significativas permaneceram dessa forma. O mesmo aconteceu com as não significativas. Apesar de existirem métodos de seleção das variáveis para escolher o melhor modelo, esse não é o objetivo do estudo em questão, mas sim verificar a influência das variáveis na severidade dos acidentes.

Para analisar a validade dos modelos foram feitos dois testes estatísticos: o teste de linhas paralelas e o teste de significância do modelo. O teste de linhas paralelas avalia se os coeficientes da equação linear que define a variável latente modelo são os mesmos através das categorias de resposta, ou seja, se as linhas para cada categoria de resposta são paralelas. O teste é baseado na comparação da hipótese nula, na qual assume-se que as linhas são paralelas, com o modelo geral, no qual as linhas não são paralelas.

**Tabela 1.** Estatística descritiva das variáveis

Variável	Média	Desvio		Descrição
		Padrão		
capacete	0,952	0,214		= 1 se a vítima estava de capacete; se não = 0
luz do dia	0,738	0,440		= 1 se o acidente ocorreu durante o dia; se não = 0
idade	1,100	0,470		= 0 se a vítima tinha menos que 21 anos; = 1 entre 21 e 40 anos; = 2 entre 41 e 60 anos; = 3 acima de 60 anos
fds	0,217	0,412		= 1 se ocorreu no sábado ou domingo; se não = 0
sexo	0,970	0,171		= 1 se masculino; = 0 se feminino
categoria	0,048	0,214		= 1 se o veículo é de aluguel; = 0 se é particular
superfície	0,071	0,257		= 1 se a pista estava molhada; = 0 se estava seca

**Tabela 2.** Estimativa dos parâmetros

Variável	Modelo 1		Modelo 2					
	Probit		Logit		Probit*		Logit*	
	$\beta$	E.P.	$\beta$	E.P.	$\beta$	E.P.	$\beta$	E.P.
capacete	-0,283	0,095	-0,515	0,173	-0,279	0,094	-0,509	0,172
luz do dia	-0,275	0,047	-0,522	0,086	-0,277	0,047	-0,526	0,086
idade	0,151	0,044	0,276	0,080	0,154	0,043	0,282	0,079
fds	0,185	0,050	0,320	0,091	0,185	0,050	0,321	0,091
sexo**	-0,032	0,119	-0,044	0,217	-	-	-	-
categoria**	0,041	0,096	0,103	0,175	-	-	-	-
superfície**	-0,079	0,079	-0,113	0,144	-	-	-	-
$\mu_0$	-1,259	0,160	-2,153	0,292	-1,218	0,111	-2,097	0,204
$\mu_1$	0,773	0,159	1,258	0,289	0,814	0,110	1,313	0,201
$\mu_2$	1,942	0,167	3,769	0,321	1,982	0,122	3,824	0,245

\* Modelo apenas com as variáveis significativas.

\*\* Variável não significativa.

O teste de linhas paralelas analisa o valor qui-quadrado ajuste tão bom quanto o modelo com as variáveis explicati-

**Tabela 3.** Teste de linhas paralelas

Modelo		Verossimilhança (-2log)		Qui-quadrado		Graus de liberdade		Significância	
		Probit	Logit	Probit	Logit	Probit	Logit	Probit	Logit
(1)	Hipótese nula	507,8	508,4						
	Geral	491,0	491,1	16,8	17,3	14	14	0,265	0,240
(2)	Hipótese nula	272,3	272,6						
	Geral	259,7	259,6	12,6	13,0	8	8	0,125	0,113

drado que é a diferença entre os valores da verossimilhança dos dois modelos. Se as linhas são paralelas, a significância observada deve ser grande, ou seja, o modelo geral não aprima

o ajuste. Com a significância maior que o grau de confiança adotado, a hipótese nula é mantida. Os modelos calibrados nesse estudo apresentaram significâncias maiores que 0,05. Conclui-se que os modelos são bem ajustados com a função de ligação escolhida. Caso a hipótese nula fosse rejeitada, deve-se considerar a utilização da regressão multinomial, a qual estima coeficientes separados para cada categoria. A Tabela 3 resume as informações do teste de linhas paralelas para os modelos.

O teste de significância do modelo 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 mudança da verossimilhança dos dois modelos segue uma distribuição qui-quadrado e o teste estabelece como hipótese nula que o modelo apenas com o intercepto tem um

ajuste tão bom quanto o modelo com as variáveis explicativas. Os dois modelos analisados obtiveram significâncias aproximadamente iguais a zero como mostrado na Tabela 4, ou seja, as variáveis explicativas melhoram o ajuste do modelo.

Na regressão linear, os valores de  $R^2$  resumem a proporção da variação da variável dependente explicada pelas variáveis independentes. Para os modelos cujo método de estimação é o da máxima verossimilhança, tal como é o caso do modelo ordenado, utilizam-se estatísticas alternativas conhecidas como pseudo- $R^2$  para avaliar o ajuste do modelo. As três métricas usadas neste estudo foram as seguintes (Greene e Hensher, 2010):

$$\text{McFadden: } R_M^2 = 1 - \left( \frac{\hat{L}(B)}{L(B^{(0)})} \right) \quad (4)$$

$$\text{Cox \& Snell: } R_{CS}^2 = 1 - \left( \frac{\hat{L}(B^{(0)})}{\hat{L}(B)} \right)^{2/n} \quad (5)$$

$$\text{Nagelkerke: } R_N^2 = \left( R_{CS}^2 / (1 - L(B^{(0)}))^{2/n} \right) \quad (6)$$

Tabela 4. Informações de ajuste do modelo

Modelo		Verossimilhança (-2log)		Qui-quadrado		Graus de liberdade		Significância	
		Probit	Logit	Probit	Logit	Probit	Logit	Probit	Logit
(1)	Intercepto	583,1	583,1						
	Final	507,8	508,4	75,3	74,7	7	7	0,000	0,000
(2)	Intercepto	346,4	346,4						
	Final	272,3	272,7	74,1	73,7	4	4	0,000	0,000

em que  $L(B^{(0)})$  é o logaritmo da verossimilhança para o modelo apenas com o intercepto;  $L(\hat{B})$  é o logaritmo da verossimilhança para o modelo com os parâmetros estimados; e  $n$  é o número de observações da amostra.

Para todos os modelos, os valores do pseudo- $R^2$  mantiveram-se constantes:  $R^2_M$  igual a 0,013,  $R^2_{CS}$  igual a 0,023 e  $R^2_N$  igual a 0,027. Tal como se pode concluir pelas Equações (4), (5) e (6), os valores do pseudo- $R^2$  são normalmente utilizados para avaliar o desempenho do modelo proposto através de uma relação com o modelo reduzido (só com o intercepto). Neste trabalho, os valores próximos dos pseudo- $R^2$  para os dois modelos considerados serviram para confirmar que não houve superioridade entre as duas estruturas propostas para o erro dos modelos (probit e logit).

#### 4. INTERPRETAÇÃO DOS COEFICIENTES E ANÁLISE DE SENSIBILIDADE

A influência das variáveis independentes na severidade dos acidentes pode ser estimada pelo efeito marginal, o qual é o efeito que a mudança de uma unidade das variáveis independentes causa na probabilidade estimada para as categorias de severidade previstas no modelo. Greene e Hensher (2010) mostram que, ao contrário da regressão linear tradicional, nem o sinal ou a magnitude dos coeficientes são informativos a respeito dos efeitos parciais (ou marginais) das variáveis explicativas dos modelos ordenados.

Greene e Hensher (2010) ressaltam que os coeficientes estão relacionados diretamente com os valores da variável latente  $y^*$  e não com os valores da variável categórica observada ( $y$ ). Desta forma, é possível mostrar que o efeito da mudança em uma variável no modelo, depende de todos os outros parâmetros, dos dados observados e da categoria de interesse.

Em relação aos sinais dos coeficientes pode-se, entretanto, associar positivamente o sentido de crescimento da variável de interesse com a probabilidade da última categoria e negativamente esse mesmo sentido para a primeira categoria, ou seja, coeficientes positivos indicam crescimento na probabilidade da última categoria e decréscimo na probabilidade da primeira categoria (Greene e Hensher, 2010).

A análise dos sinais dos coeficientes para os Modelos 2 (Tabela 2) indica que quanto maior a idade da vítima, maiores as chances de haver acidente fatal. A variável final de semana também apresentou um coeficiente positivo, revelando que nos sábados e domingos há uma maior probabilidade de acidentes mais graves. Esses resultados podem estar relacionados com uma maior velocidade média desenvolvida pelas motocicletas devido ao menor fluxo nas vias e ao uso de álcool ou outras drogas durante esse período.

A variável associada ao uso do capacete apresentou coeficiente negativo corroborando estudos que apontam a

redução dos riscos de morte e lesões graves em acidentes com motocicletas. Da mesma forma, a variável *luz do dia* também apresentou coeficiente negativo, sugerindo que durante o dia (6:00 às 18:00hrs) há um menor risco de acidentes com vítimas fatais. Uma hipótese levantada para explicar esse fenômeno é que durante o dia há um maior fluxo e uma menor velocidade média desenvolvida pelos veículos e motocicletas, além da iluminação natural facilitar a visibilidade de outros veículos, pedestres e obstáculos, permitindo assim um maior tempo de reação para evitar o acidente.

Os parâmetros estimados permitem analisar o efeito das variáveis, mas não a magnitude das mesmas. Para tal deve ser considerado o cálculo dos efeitos marginais quando as variáveis explicativas são contínuas. No caso do presente estudo, as variáveis explicativas são todas categóricas (binárias ou multcategóricas no caso da variável Idade). Neste caso a magnitude é analisada por comparação das probabilidades correspondentes a cada categoria da variável explicativa (por exemplo, 0 ou 1), mantendo-se as restantes variáveis constantes.

Consequentemente, e de forma a complementar a análise da sensibilidade do modelo em relação às variáveis, considerou-se uma vítima de acidente de referência e em seguida verificou-se a variação das probabilidades com a mudança das características. A vítima de referência pode ter, por exemplo, todas as variáveis com valores iguais a zero, ou seja:

- i) Não utilizava capacete,
- ii) Envolveu-se em acidente entre 18:00 e 06:00 horas,
- iii) Tinha idade entre 0 e 20 anos,
- iv) Envolveu-se em acidente ocorrido entre segunda-feira e sexta-feira.

As estimativas das probabilidades de acidentes para a vítima de referência estão representadas na Tabela 5. Para ambos os modelos Logit e Probit, as probabilidades permaneceram as mesmas para os diferentes níveis de severidade. As linhas restantes da Tabela 5 apresentam as probabilidades quando as variáveis estão com valor igual a um (ou dois e três no caso da Idade), mantendo-se as restantes iguais à "referência".

#### 5. CONCLUSÕES

Esse trabalho apresentou uma análise do impacto de variáveis relacionadas ao momento do acidente e a características dos condutores na severidade dos acidentes envolvendo motocicletas em Fortaleza.

Foram desenvolvidos modelos ordenados categóricos do tipo logit e probit utilizando uma amostra com 3.232 observações de acidentes de trânsito coletadas do Sistema de Informações de Acidentes de Trânsito de Fortaleza

**Tabela 5.** Probabilidades estimadas

	<i>Ileso</i>	<i>Leve</i>	<i>Grave</i>	<i>Fatal</i>
<b>Referência</b>	<b>0,109</b>	<b>0,679</b>	<b>0,191</b>	<b>0,021</b>
Capacete = 1	0,170	0,691	0,126	0,013
Luz do Dia = 1	0,172	0,691	0,124	0,013
Idade = 1	0,085	0,652	0,235	0,028
Idade = 2	0,065	0,613	0,284	0,038
Idade = 3	0,045	0,574	0,333	0,048
Final de semana = 1	0,082	0,647	0,241	0,030

(SIATFOR) para os anos de 2004 a 2011. A severidade resultante dos acidentes de trânsito foi classificada em quatro categorias, a saber: somente danos materiais, ( $y_i = 0$ ), ferido leve ( $y_i = 1$ ), ferido grave ( $y_i = 2$ ), e vítima fatal ( $y_i = 3$ ).

As variáveis explicativas do estudo foram limitadas à qualidade e disponibilidade no sistema, tendo sido utilizadas as variáveis relacionadas ao uso do capacete, hora da ocorrência, dia da semana, condições do pavimento da hora do acidente, uso da motocicleta, idade e sexo dos condutores.

Os resultados do processo de calibração indicaram que as variáveis relacionadas ao sexo dos condutores, condição do pavimento (seco ou molhado) e uso da motocicleta (particular ou aluguel) não influenciaram significativamente na severidade dos acidentes.

Com os resultados da calibração dos modelos, percebe-se que motociclistas que utilizam capacete e pilotam durante o dia têm um menor risco de sofrer lesões mais graves. De outra forma, motociclistas mais velhos e que sofreram acidente ocorrido em finais de semana apresentaram um maior risco de lesões mais graves.

Testes de validação foram realizados para mostrar a qualidade dos modelos. Notou-se que todos os modelos apresentaram coeficientes iguais entre as categorias de gravidade e que os modelos com as variáveis explicativas tornam os modelos melhores do que aqueles apenas com o intercepto. Também foi possível fazer uma comparação entre os modelos probit e logit, observando-se que ambos produzem resultados similares.

Em trabalhos futuros recomenda-se explorar técnicas estatísticas com o objetivo de considerar, por exemplo, a heterogeneidade entre observações ou a flexibilidade dos parâmetros estimados para os valores limites que separam as categorias da variável dependente (Yasmin e Eluru, 2013). Estas técnicas permitem por um lado, um melhor ajuste do modelo à base de dados, e por outro lado, aprofundar a análise das variáveis explicativas na gravidade das vítimas.

## AGRADECIMENTOS

Os autores agradecem o apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) para o desenvolvimento dessa pesquisa, além das Bolsas de Iniciação de Pesquisa e Produtividade em Pesquisa concedidas. Os autores agradecem ainda a Autarquia Municipal de Trânsito, Serviços Públicos e Cidadania de Fortaleza -

AMC pelas informações disponibilizadas sobre os acidentes de trânsito neste trabalho.

## REFERÊNCIAS

- Abdel-Aty, M. (2003) Analysis of Driver Injury Severity Levels at Multiple Locations Using Ordered Probit Models. *Journal of Safety Research*, v. 34, n. 5, p. 597-603. DOI:10.1016/j.jsr.2003.05.009.
- AMC (2012) *Anuário estatístico de acidentes de trânsito de Fortaleza – 2004 à 2011*. Autarquia Municipal de Trânsito, Serviços Públicos e Cidadania, Prefeitura Municipal de Fortaleza, Fortaleza, CE.
- Bajracharya, S. (2013) Intersection Accident Analysis Using Ordered Probit Model. *Proceedings of Eastern Asia Society for Transportation Studies*, v. 9, p. 379. ISSN:1881-1132
- Chung, Y.; Song, T. E B. Yoon (2014) Injury Severity in Delivery-motorcycle to Vehicle Crashes in the Seoul Metropolitan Area. *Accident Analysis and Prevention*, v. 62, p. 79-86. DOI: 10.1016/j.aap.2013.08.024.
- Couto, A. E S. Ferreira (2012) Método Probabilístico para Identificação de Zonas de Acumulação de Acidentes. *Anais do XXVI Congresso de Pesquisa e Ensino em Transportes*, ANPET, Joinville, v. 21, n. 3, p. 48-55. DOI:10.4237/transportes.v21i3.683.
- Ferreira, S. (2010) *A Segurança Rodoviária no Processo de Planeamento de Redes de Transportes*. Ph.D. dissertation, University of Porto.
- Greene, W.H. E D.A Hensher (2010) *Modeling Ordered Choices: A Primer and Recent Developments*, Cambridge University Press, Cambridge.
- Holz, R. F. (2013) *Motocicletas: Uma Realidade Inconveniente*. Qualificação de Doutorado, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS.
- Holz, R. F.; Korzenowski, A.; Nodari, C. T.; Lindau, L. A. E C. S. T Caten (2011) Modelagem dos Acidentes Envolvendo Motociclistas em Porto Alegre. *Anais do XXV Congresso de Pesquisa e Ensino em Transportes*, ANPET, Belo Horizonte. Panorama Nacional da Pesquisa em Transportes, v. 1, p. 1553-1564.

INCT (2013) *Evolução da Frota de Automóveis no Brasil 2001 – 2012*. Observatório das Metrôpoles, Instituto Nacional de Ciência e Tecnologia, Ministério de Ciência e Tecnologia, Brasília, DF.

Kockelman, K. M. E Y. J. Kweon (2002) Driver Injury Severity: an Application of Ordered Probit Models. *Accident Analysis and Prevention*, v. 34, p. 313-321. DOI: 10.1016/S0001-4575(01)00028-8.

Kononen, D. W.; Flannagan, C. E S. Wong (2011) Identification and Validation of a Logistic Regression Model for Predicting Serious Injuries Associated with Motor Vehicle Crashes. *Accident Analysis and Prevention*, v. 43, n. 1, p. 112-122. DOI: 10.1016/j.aap.2010.07.018.

Mannering, F. L. E C.R. Bhat (2014) Analytic Methods in Accident Research: Methodological Frontier and Future Directions. *Analytic Methods in Accident Research*, v. 1, p. 1-22. DOI: 10.1016/j.amar.2013.09.001.

O'donnell, C. J. E D.H. Connor (1996) Predicting the Severity of Motor Vehicle Accident Injuries Using Models of Ordered Multiple Choice. *Accident Analysis and Prevention*, v. 28, n. 6, p. 739-753. DOI: 10.1016/S0001-4575(96)00050-4.

Obeng, K. (2011) Gender Differences in Injury Severity Risks in Crashes at Signalized Intersections. *Accident Analysis and Prevention*, v. 43, n. 4, p. 1521-1531. DOI: 10.1016/j.aap.2011.03.004.

Veloza, P. L. C. (2009) *Modelos para Dados Categóricos com Estrutura Temporal*. Dissertação de Mestrado, Instituto de Matemática, Departamento de Métodos Estatísticos, UFRJ, Rio de Janeiro.

Waiselfisz, J.J. (2013) *Mapa da violência 2013: Acidentes de Trânsito e Motocicletas*. Rio de Janeiro.

Yasmin, S. E N. Eluru (2013) Evaluating Alternate Discrete Outcome Frameworks for Modeling Crash Injury Severity. *Accident Analysis & Prevention*, v. 59, p. 506-521. DOI: 10.1016/j.aap.2013.06.040.

Ye, F. E D. Lord (2014) Comparing Three Commonly Used Crash Severity Models on Sample Size Requirements: Multinomial Logit, Ordered Probit, and Mixed Logit. *Analytic Methods in Accident Research*, v. 1, p. 72-85. DOI: 10.1016/j.amar.2013.03.001.