

Modelos de previsão de acidentes de trânsito em interseções semaforizadas de Fortaleza

Flávio José Craveiro Cunto¹, Manoel Mendonça de Castro Neto² e Davi Sales Barreira³

Resumo: As interseções viárias urbanas concentram grande parte dos acidentes de trânsito em virtude do elevado grau de interação entre usuários, veículos e via. Os modelos de previsão de acidentes são modelos de regressão que relacionam a frequência dos acidentes de trânsito com atributos geométricos e operacionais da via e lidam, com relativo sucesso, com o elevado grau de aleatoriedade desse fenômeno. O objetivo deste trabalho é desenvolver modelos de previsão de acidentes de trânsito para interseções semaforizadas da cidade de Fortaleza, com ênfase na metodologia para a construção de modelos simples e confiáveis. Os modelos foram estimados com uma amostra de 101 interseções em função do volume diário médio anual, número de faixas, número de aproximações e tipo de separador central. O modelo contendo o fluxo e número de faixas apresentou desempenho satisfatório para a predição do número total de acidentes de trânsito nas interseções semaforizadas avaliadas. DOI:10.4237/transportes.v20i2.558.

Palavras-chave: modelos de previsão de acidentes, acidentes de trânsito, modelos lineares generalizados, modelagem de segurança viária.

Abstract: Urban road intersections concentrate a significant portion of traffic accidents due to the high degree of interaction between users, vehicles and road. Accident prediction models are regression models that establish a relationship between the frequency of traffic accidents and geometric and operational attributes of the road, dealing, with relative success, with the high degree of randomness of this phenomenon. The objective of this work is to develop traffic accident prediction models for signalized intersections located in the city of Fortaleza, with emphasis on the methodology for the development of simple and reliable models. The models were estimated based on a sample of 101 intersections, having as prediction variables annual average daily traffic (AADT), number of lanes, number of approaches, and type of central median. The model containing AADT and number of lanes showed satisfactory performance for the estimation of the total number of traffic accidents at the evaluated intersections.

Keywords: safety prediction models, safety performance functions, generalized linear models, road safety modeling.

1. INTRODUÇÃO

Por ano, mais de um milhão de pessoas são mortas e entre 20 e 50 milhões são gravemente feridas (Dinn-Zarr, 2008). Acidentes de trânsito são a principal causa de mortes em jovens entre 15 e 29 anos. Do ponto de vista econômico, os custos dos acidentes de trânsito chegam a mais de 30 bilhões de reais por ano (valor atualizado para junho de 2008), o que corresponde a aproximadamente 1,2% do PIB brasileiro (Ferraz, 2008; IPEA, 2003). Deste montante, aproximadamente 7 bilhões referem-se a acidentes em vias urbanas (IPEA, 2006). Com relação ao impacto causado na fluidez, colisões de tráfego são responsáveis por 25% do congestionamento nos EUA e na Inglaterra (Alkaabi *et al.*, 2011). Um estudo feito em Riyadh, capital da Arábia Saudita mostrou que cerca de 50% dos acidentes ocorreram em interseções, e destas, mais da metade foram consideradas severas (Al-Ghamadi, 2003).

O processo de tomada de decisão no gerenciamento da segurança viária depende de indicadores que possam expressar de forma objetiva o nível de segurança dos componentes de uma determinada rede de transportes

(AASHTO, 2010). Nesse sentido, pesquisadores têm procurado relacionar séries históricas de acidentes de trânsito com atributos geométricos e operacionais da via com o uso de modelos estatísticos de regressão denominados modelos de previsão de acidentes (MPA) ou funções de desempenho de segurança viária. Um dos principais objetivos dos MPA é melhorar as análises de segurança viária através de um controle mais rigoroso das características inerentes ao processo estocástico dos acidentes de trânsito, como o fenômeno de regressão à média e a dispersão excessiva dos dados (Hauer, 2002).

Se por um lado os MPA têm sido explorados há mais de duas décadas em países como Canadá, Estados Unidos, Inglaterra e Suécia, por outro lado, no Brasil seu estágio de desenvolvimento ainda é incipiente. Os primeiros esforços de modelagem para o ambiente brasileiro podem ser atribuídos à pesquisa de doutoramento de Cardoso (2006) e sintetizados no trabalho posterior de Cardoso e Goldner (2007), nos quais foram analisados segmentos arteriais urbanos de Porto Alegre. Acredita-se que essa discrepância possa ser atribuída em parte a problemas relacionados com a disponibilidade e qualidade das informações sobre acidentes de trânsito associados à relativa escassez de procedimentos para a calibração e validação de modelos dessa natureza no âmbito nacional.

Este trabalho tem por objetivo desenvolver modelos de previsão de acidentes de trânsito para interseções semaforizadas da cidade de Fortaleza com ênfase nos aspectos necessários para a construção de modelos simples e confiáveis.

¹ Flávio José Craveiro Cunto, Programa de Pós-Graduação em Engenharia de Transportes, Departamento de Engenharia de Transportes, Universidade Federal do Ceará, Fortaleza, CE, Brasil. (e-mail: flaviocunto@det.ufc.br).

² Manoel Mendonça de Castro Neto, Programa de Pós-Graduação em Engenharia de Transportes, Departamento de Engenharia de Transportes, Universidade Federal do Ceará, Fortaleza, CE, Brasil. (e-mail: manoiel@det.ufc.br).

³ Davi Sales Barreira, Departamento de Engenharia de Transportes, Universidade Federal do Ceará, Fortaleza, CE, Brasil. (e-mail: davibar1@gmail.com).

2. MODELOS DE PREVISÃO DE ACIDENTES

A modelagem estatística da segurança viária baseia-se na premissa de que a partir da frequência de acidentes em vários locais é possível, utilizando ferramentas estatísticas, determinar empiricamente como estratégias relacionadas ao projeto e operação do ambiente viário podem alterar as chances de ocorrência de acidentes de trânsito (Davis, 2004). Nesse sentido os MPA têm sido amplamente considerados como uma das ferramentas mais promissoras na modelagem da natureza aleatória rara associada aos acidentes de trânsito.

Inicialmente os MPA foram desenvolvidos a partir de modelos de regressão linear assumindo a normalidade dos erros e homocedasticidade. Essas suposições foram consideradas inconsistentes com a natureza e frequência dos acidentes de trânsito (Hauer *et al.*, 1988). Pesquisas posteriores confirmaram resultados promissores com a aplicação de modelos assumindo a distribuição do número observado de acidentes do tipo Poisson, ou a distribuição binomial negativa em casos onde os acidentes apresentam um padrão consideravelmente disperso (variância superior à média) entre entidades similares (Bonneson e McCoy, 1993; Persaud e Mucsi, 1995). Uma alternativa possível para lidar com esse padrão excessivo de dispersão dos dados, incompatível com as suposições da regressão linear tradicional (distribuição normal) é o emprego dos modelos lineares generalizados (Cunto, 2008; Cardoso, 2006).

A formulação geral mais comum para os MPA em interseções encontrada na literatura pode ser expressa por (Hakkert *et al.*, 1996; Sawalha e Sayed, 2006; AASHTO, 2010):

$$Y = \alpha \left[\prod_i (A_i)^{\beta_i} \right] \cdot e^{\sum_j (\gamma_j B_j)} \quad (1)$$

em que,

Y: número esperado de acidentes em um intervalo de tempo (ano);

A e *B*: vetores de variáveis preditoras; e

α , β , γ : vetores de coeficientes do modelo.

A estimação dos coeficientes do modelo é quase sempre feita com a utilização de rotinas e aplicativos estatísticos como o GENSTAT, GENMOD no SAS e GLM no R os quais aplicam o método da máxima verossimilhança e o algoritmo de otimização Newton-Raphson (Hauer, 2002;

Lord, 2006). A Tabela 1 apresenta as principais características de MPA propostos para interseções.

3. METODOLOGIA PARA ESTIMAÇÃO DOS MPA

A metodologia adotada para o desenvolvimento dos MPA nesse trabalho seguiu as seguintes etapas: definição dos objetivos da modelagem, escolha das variáveis e definição da amostra, coleta de dados, estimação dos parâmetros dos modelos investigados, avaliação e seleção do modelo mais adequado. O objetivo da modelagem é estimar o número total de acidentes de trânsito e o número de acidentes de trânsito com vítimas feridas e fatais em interseções semaforizadas a partir de seus atributos físicos e operacionais.

3.1. Escolha das variáveis e definição da amostra

O processo de escolha das variáveis para os MPA não possui critérios objetivos e, na maioria das vezes depende, além dos objetivos do estudo, de fatores relacionados à experiência do modelador e à disponibilidade das informações sobre o ambiente viário da entidade. Tomando por base os critérios descritos e considerando a frequência com que certas variáveis aparecem na literatura, as quatro variáveis preditoras consideradas neste estudo foram volume diário médio anual (VDMA), número total de faixas da interseção, número total de aproximações e presença ou não de canteiro central. A seguir serão apresentados aspectos adicionais que justificaram a inclusão das variáveis acima:

- **VDMA:** variável de exposição que usualmente explica a maior parte da variação dos acidentes de trânsito em interseções;
- **Número total de faixas:** acredita-se que essa variável aumente a frequência de avanços de semáforo (colisões transversais) e frequência de mudanças de faixas (colisões laterais);
- **Número de aproximações:** Uma via com mais de um sentido pode gerar maior interação entre os veículos, inclusive com conversões proibidas além de exigir um grau maior de atenção dos condutores; e
- **Presença de canteiro central:** É possível que sua presença possa de alguma forma reduzir a quantidade de acidentes entre veículos em sentidos opostos.

Tabela 1. Modelos de previsão de acidentes de trânsito para interseções

Autores	Local	#Anos	Entidade		Variáveis preditoras testadas
			#Int.	Tipo de controle	
Bélangier (1994)	Quebec/CA	–	149	Não semaforizada	VDMA total e por aproximação, distância de visibilidade, presença de faixas de conversão, limite de velocidade e indicação luminosa da interseção
Bonneson e McCoy (1993)	Nebraska/EUA	3	125	Não semaforizada	VDMA da via principal e secundária
Hauer <i>et al.</i> (1988)	Toronto/CA	3	145	Semaforizada	Fluxo veicular por hora e período do dia
Persaud e Nguyen (1998)	Ontário/CA	6	500	Semaforizada	VDMA total, da principal e secundária
Sayed e Rodrigues (1999)	Vancouver/CA	3	419	Não semaforizada	VDMA da via principal e secundária
Oh <i>et al.</i> (2004)	Georgia, Califórnia, Michigan/EUA	6	460	Sem./Não sem.	53 variáveis preditoras

Assim como na escolha das variáveis, um método de definição da amostra para o desenvolvimento de MPA ainda não está satisfatoriamente estabelecido. De maneira geral procura-se controlar as variáveis com objetivo de criar grupos de entidades homogêneas para reduzir a influência de fatores desconhecidos (ou não controlados) que reduzam o poder preditor do modelo.

Bonneson e McCoy (1993) e Lord (2006) observam, entretanto, que o modelador deve evitar a criação de grupos excessivamente homogêneos uma vez que isso pode culminar com amostras reduzidas comprometendo a obtenção de valores estatisticamente significativos dos parâmetros. Uma indicação mais objetiva do tamanho da amostra é apresentada por Lord (2006) o qual recomenda, em função da média dos acidentes de trânsito e do parâmetro de dispersão da distribuição binominal negativa, a utilização de amostras com mais de 100 observações.

A amostra inicialmente planejada para este estudo tinha um total de 288 interseções semaforizadas da rede viária com controle centralizado do CTAFOR-Controle de Tráfego em Área de Fortaleza. Essa amostra foi reduzida para 119 interseções semaforizadas em virtude de problemas no armazenamento dos dados de fluxo veicular junto ao CTAFOR.

Uma análise utilizando imagens aéreas do aplicativo *Google Earth* foi realizada para remover interseções com características consideradas atípicas e que poderiam influenciar negativamente os resultados. A seleção e remoção dessas interseções obedeceram aos seguintes critérios:

- Interseções com mais de quatro ramos;
- Interseções com ramos desalinhados;
- Ângulo entre vias menor que 70 graus;
- Espaçamento menor que 50 m entre interseções; e
- Interseções com estágio para conversões à esquerda.

Um exemplo de interseções retiradas da amostra (espaçamento menor que 50 metros) está ilustrado na Figura 1. Como resultado desse processo de seleção, 18 interseções foram eliminadas e, desta forma, a amostra utilizada no processo de modelagem foi composta por 101 interseções semaforizadas.

3.2. Coleta de dados e análise descritiva da amostra das interseções semaforizadas

Os dados brutos de acidentes foram coletados pelo Sistema de Informações de Acidentes de Trânsito de Fortaleza (SIAT-FOR). O SIAT-FOR é um banco de dados informatizado e georeferenciado que compila os acidentes de trânsito registrado por 10 órgãos responsáveis pelo atendimento às ocorrências de trânsito do município de Fortaleza desde o ano de 2000. Dentre as informações disponíveis no SIAT-FOR destacam-se: data, hora, local da ocorrência (nome da via e número/referência), tipo do acidente, veículos envolvidos, severidade do acidente (sem vítimas, com vítimas feridas, e com vítimas fatais), além de informações sobre as vítimas como gênero, idade, tipo (condutor, passageiro, pedestre, ciclista, etc).

Os dados utilizados neste estudo referem-se ao ano de 2009 e têm como variáveis de interesse o número de acidentes por aproximação categorizados de acordo com a severidade em dois tipos: número total de acidentes e número total de acidentes com feridos e fatais.

Um aspecto importante sobre a coleta dos acidentes diz respeito à sua localização. O processo de coleta das informações do órgão gestor de Fortaleza conta com 10 fontes diferentes de informações e, apesar da existência de rotinas de consistência de dados e do georeferenciamento, não existe um procedimento padronizado para a alocação de um determinado acidente como sendo na interseção ou ao longo de um segmento. Uma análise preliminar realizada em boletins de ocorrência do órgão gestor de trânsito (AMC) sugere que, na maioria das vezes, os agentes atribuem a localização do acidente como sendo na interseção quando o impacto inicial se deu no quadrilátero formado pelo prolongamento das faixas de retenção das aproximações (interseções semaforizadas). Desta forma, as observações (acidentes) utilizadas nesse trabalho, foram caracterizadas como sendo em interseções de acordo com os mesmos critérios utilizados pelos agentes públicos.

O VDMA foi estimado com dados de fluxo veicular armazenados no banco de dados do CTAFOR, cujo controle semaforizado é feito pelo aplicativo inglês SCOOT (*Split Cycle Offset Optimization Technique*). Tal sistema estima e armazena fluxos horários em períodos de 15 minutos em

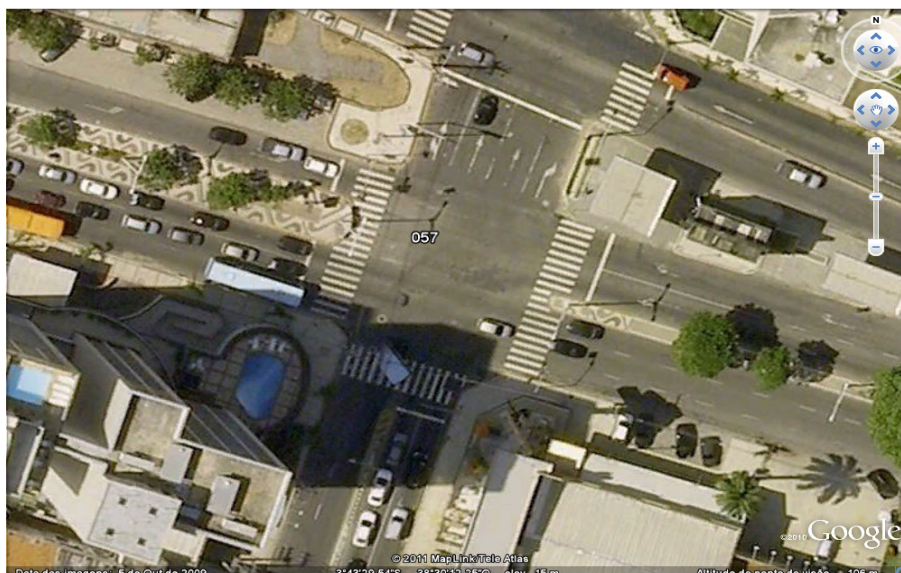


Figura 1. Exemplo de interseção atípica removida da amostra

Tabela 2. Descrição das variáveis

<i>Variável</i>	<i>Descrição</i>	<i>Média</i> (*)	<i>Desv. pad.</i>	<i>Mínimo</i>	<i>Máximo</i>
<i>v1_vdma09</i>	VDMA de 2009- via principal	23.583	7.850	8.047	44.312
<i>v2_vdma09</i>	VDMA de 2009 - via secundária	11.735	5.267	641	28.563
<i>vdma09</i>	VDMA de 2009 - interseção	35.319	10.438	15.887	65.618
<i>ta09</i>	Total de acidentes - interseção	7,5	7,7	0	48
<i>taff09</i>	Total de acidentes com feridos e fatais	1,5	1,6	0	7
<i>nfx</i>	Número total de faixas - interseção	5,8	1,4	4	12
<i>nap</i>	Número total de aproximações	2,7	0,6	2	4
<i>cc</i>	Canteiro central; 0=ausência, 1=em uma via, 2=em duas vias	0,7	0,6	0	2

(*) Amostra = 101 observações

todos os 650 *links* das 288 interseções semaforizadas da região controlada.

Inicialmente uma amostra dos volumes de todos os *links* das interseções disponíveis (165) contendo os 31 dias do mês de maio de 2009 foi coletada para a estimação do volume diário médio mensal (VDMM). Realizou-se então uma verificação para retirar da amostra os dias com possíveis falhas nos dados com a utilização de ferramenta *box plot*.

Para dispensar a necessidade da aplicação de fatores de expansão do volume de um dia para a média semanal, considerou-se que o VDMM deveria ser estimado pela média de pelo menos uma semana completa de observações. Desta forma ao final do processo, os *links* que não tivessem pelo menos uma semana de dados válidos eram excluídos da amostra. Vale ressaltar que a retirada de um *link* inválido provocou também a exclusão de sua respectiva interseção. Ao todo foram removidas da amostra 46 interseções.

Finalmente o VDMM de maio foi expandido para o VDMA com a aplicação do fator de expansão médio de 0,991. Esse fator foi obtido com a utilização dos resultados de uma análise do padrão da variação espaço-temporal do volume veicular em Fortaleza apresentado no trabalho de Oliveira (2004).

No tocante as variáveis preditoras relacionadas com a geometria das interseções, sua coleta foi realizada com a utilização de fotografias aéreas do aplicativo *Google Earth*, e também através de pesquisas em plantas com projetos geométricos das vias fornecidas pelo órgão gestor do trânsito de Fortaleza. A Tabela 2 apresenta uma análise descritiva das variáveis utilizadas nesse esforço de modelagem.

3.3. Estimação dos parâmetros dos modelos investigados

Com base nos objetivos da modelagem e nas variáveis coletadas foram investigadas duas categorias de modelos: a categoria 1 utilizando como variável resposta o número total de acidentes de trânsito (*ta09*), e a categoria 2 com o número total de acidentes de trânsito com vítimas fatais e feridas (*taff09*) como variável dependente.

Para as duas categorias adotou-se como modelo de partida o que contém somente a variável *vdma09* na parte multiplicativa da Equação 01 e a partir daí foram acrescentadas sequencialmente as outras variáveis preditoras no componente exponencial dessa equação. Desta forma, para se considerar todas as combinações das variáveis envolvidas, um total de oito modelos foi estimado para cada categoria de MPA.

Um teste inicial foi realizado com os modelos de partida para se verificar, entre Poisson e binomial negativa, qual a distribuição mais adequada para os dados de acidentes de trânsito coletados. Os parâmetros de dispersão σ_d para os modelos da categoria 1 foram respectivamente iguais a 3,92 para a distribuição de Poisson e 1,04 para a binomial negativa. Nos modelos da categoria 2 estes foram de 1,78 e 1,13, respectivamente. Os resultados desses testes sugerem que a dispersão observada nos dados coletados é mais compatível com a premissa da distribuição binomial negativa ($\sigma_d \cong 1$) para as duas categorias de modelos.

Os parâmetros de cada modelo foram então estimados com a utilização da função *glm.nb* (binomial negativa) do aplicativo estatístico R, versão 2.13.0 adotando a função logarítmica como função de ligação (exemplo da sintaxe do R: `<glm.nb(formula = ta09 ~ log(vdma09), init.theta = 2.566679449, link = log)>`).

3.3.1. Modelos da categoria 01 – número total de acidentes de trânsito

Os coeficientes de cada modelo estimado com os respectivos valores do erro-padrão estão apresentados na Tabela 3. Essa tabela mostra ainda os parâmetros de “superdispersão” (ϕ) da distribuição binomial negativa e o valor do logaritmo da verossimilhança (-2ℓ) obtidos para cada modelo testado.

Conforme mencionado anteriormente, a construção dos possíveis modelos seguiu um processo sequencial de inserção de variáveis a partir do modelo com apenas a variável VDMA (modelo de partida). Esse processo baseia-se na comparação entre os valores do logaritmo da máxima verossimilhança (-2ℓ) do modelo mais simples com o modelo imediatamente mais complexo (com uma variável adicional). A variável que, ao ser acrescentada resulte no maior aumento no parâmetro -2ℓ será a próxima a ser incluída. O processo é repetido até que todas as variáveis significativas sejam incluídas no modelo.

Observando-se os valores da Tabela 3, por exemplo, pode-se comparar o modelo 1 (*vdma09*) com os possíveis modelos de duas variáveis (modelos 2, 3 e 4). Observa-se que a inclusão da variável *nfx* acarreta o maior acréscimo no valor de -2ℓ (27), e portanto, esta variável deve ser incluída imediatamente. Seguindo esse raciocínio foram selecionados os modelos 1, 2, 6 e 8 para uma avaliação mais minuciosa com o emprego de indicadores estatísticos e ferramentas gráficas.

Tabela 3. Parâmetros dos modelos investigados – Categoria 01

Variável	Modelos (*)								
	01	02	03	04	05	06	07	08	
α	ln(α)	-12,09	-5,13	-7,72	-5,65	-5,27	-7,49	-7,12	-7,23
	Coef.	5,61E-06	5,91E-03	4,44E-04	3,52E-03	5,16E-03	5,61E-04	8,09E-04	7,22E-04
	$\hat{\sigma}_\beta^{(**)}$	2,59	2,52	3,06	2,85	2,65	2,68	2,89	2,61
$vdma_{09}$	Coef.	1,34	0,52	0,90	0,59	0,53	0,70	0,60	0,55
	$\hat{\sigma}_\beta$	0,23	0,26	0,30	0,14	0,28	0,26	0,29	0,26
nfx	Coef.	–	0,28	–	–	0,52	0,41	–	0,34
	$\hat{\sigma}_\beta$	–	–	–	–	–	–	–	–
cc	Coef.	–	–	0,31	–	–	-0,39	-0,66	-0,90
	$\hat{\sigma}_\beta$	–	–	–	–	–	–	–	–
nap	Coef.	–	–	0,14	–	–	0,27	0,31	0,28
	$\hat{\sigma}_\beta$	–	–	–	–	–	–	–	–
ϕ	$\hat{\sigma}_\beta$	–	–	–	0,14	0,20	–	0,33	0,31
-2ℓ		2,92	4,47	3,16	3,58	4,47	4,81	3,90	5,39
		-575	-548	-570	-561	-548	-543	-556	-537

(*) parâmetros estatisticamente significantes ($\alpha=0,05$)

3.3.2. Modelos da categoria 02 – número total de acidentes de trânsito com vítimas feridas e fatais

Os parâmetros das variáveis testadas em todos os possíveis modelos para o número de acidentes de trânsito com vítimas feridas e fatais se mostraram estatisticamente não significantes ($\alpha = 0,05$). Uma análise visual da Figura 2 confirma que os acidentes da amostra apresentaram um comportamento disperso e sem um padrão aparente de variação com o fluxo. Essa mesma análise foi verificada para as outras variáveis.

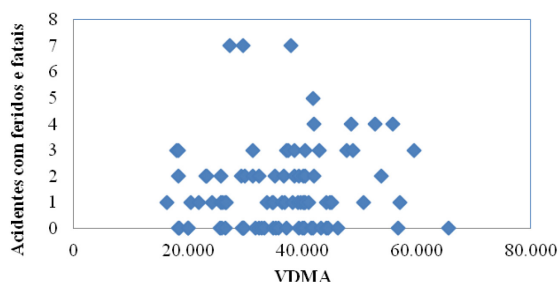


Figura 2. Gráfico de acidentes com feridos e vítimas fatais e exposição

Em termos práticos, os resultados obtidos confirmam que, no ambiente das interseções onde predominam as colisões traseiras, a existência de uma maior exposição acarreta em maior probabilidade da ocorrência de acidentes de trânsito, não implicando, contudo, que os mesmos serão mais graves. A severidade dos acidentes nas interseções parece estar mais associada às colisões transversais que, para o ambiente semaforizado, tem forte ligação com o avanço semaforizado (voluntário ou involuntário). Nesse contexto a consideração de outras variáveis como hora do dia, local da via, existência de fiscalização eletrônica pode representar melhor esta categoria de acidentes.

3.4. Avaliação dos modelos da categoria 01

Após a verificação inicial da significância de cada coeficiente do modelo (*valor-p*) um número considerável de indicadores matemáticos e gráficos tem sido aplicado ao processo de escolha de modelos lineares generalizados.

Busca-se o equilíbrio entre a simplicidade (número de parâmetros) e a aderência aos dados. Para modelos com elevado número de parâmetros, a variância dos erros será baixa, entretanto o modelo será viesado para a amostra utilizada (*overfitting*); ou seja, o modelo se ajustará tão firmemente aos dados existentes que perderá a capacidade de previsão de acidentes em novas interseções. Por outro lado, modelos com menos parâmetros terão um menor viés, porém possuirão menor aderência, pois variáveis importantes para explicar o número de acidentes estarão ausentes.

Este trabalho balizou a escolha do modelo mais adequado com a utilização dos seguintes indicadores: estatística generalizada Pearson χ_p^2 , parâmetro de dispersão σ_d (*overdispersion parameter*), desvio escalonado S_p (*scaled deviance*), critério de informação de Akaike (AIC) e gráfico de resíduos acumulados (*CURE plot*) (McCullagh e Nelder, 1989; Bonneson e McCoy, 1993; Lord, 2000; Hauer, 2004). A Tabela 4 mostra os valores obtidos desses indicadores para cada um dos modelos selecionados.

O indicador χ_p^2 é utilizado para avaliar o quanto a variância explicada pelo modelo difere da variância observada nos dados, ou seja, se o modelo proposto se ajusta aos dados. Para um determinado nível de significância α o modelo é considerado adequado se $\chi_p^2 < \chi_{p,critico,\alpha}^2$. Os valores apresentados na Tabela 4 indicam que, de acordo com esse critério ($\alpha = 0,05$), os

Tabela 4. Indicadores dos modelos selecionados – Categoria 01

Indicador	Modelos			
	1	2	6	8
-2ℓ	-575,46	-547,92	-542,85	-536,80
$gI^{(*)}$	99	98	97	96
$\chi_{p,critico,0,05}^2$	123	122	121	120
χ_p^2	135	129	119	108
S_p	103	102	101	101
AIC	581	556	559	549
σ_d	1,04	1,04	1,04	1,05

(*) grau de liberdade

modelos 6 e 8 são os que melhor se ajustam aos dados. O modelo 2 entretanto teve um desempenho muito próximo ao aceitável não devendo ser totalmente descartado. Vale ressaltar que esse indicador tem seu desempenho questionado para amostras pequenas (Bonneson e McCoy, 1993), ou seja, aumentando-se a amostra o modelo 2 poderia ser considerado aceitável.

Para o desvio escalonado (S_p), que é um outro indicador da falta de aderência do modelo proposto, conclui-se que os quatro modelos estão aproximadamente a uma mesma distância (abaixo) dos seus valores críticos (Tabela 4). Ou seja, quando comparado com modelos saturados (uma variável para cada observação), os modelos parecem aderir satisfatoriamente aos dados.

O critério AIC compara diferentes modelos com base no equilíbrio entre o viés e a variância explicada pelos mesmos. Por sua formulação matemática o modelo mais equilibrado é aquele que apresenta o menor AIC. Segundo Simonoff (2003), uma diferença de AIC acima de 10 já pode ser considerada expressiva. O modelo 1 é o mais simples, porém seu valor de AIC é bem superior ao dos demais modelos. Já o modelo 2 oferece uma melhor aderência, com a redução de 25 pontos com relação ao modelo 1. A diferença do AIC do modelo 2 para os demais ficou abaixo de 10 pontos, favorecendo portanto a escolha do modelo 2 como a mais plausível alternativa.

Para modelos satisfatórios, o gráfico de resíduos acumulados (*CURE plot*) deve apresentar uma oscilação moderada e aleatória em torno de zero (“*random walk*”) além de permanecer confinado na envoltória de dois desvios padrões (Hauer, 2004). A Figura 3 apresenta o *CURE plot* para o modelo 2 com as respectivas envoltórias de $\pm 2\sigma_d$. Os mesmos gráficos foram produzidos para os outros modelos os quais tiveram comportamento similar. Em todos os casos os modelos estiveram entre os limites da envoltória e com oscilação em torno de zero. Um aspecto interessante do *CURE plot* é o indicativo de observações extremas (*outliers*). Esses pontos podem ser identificados no gráfico

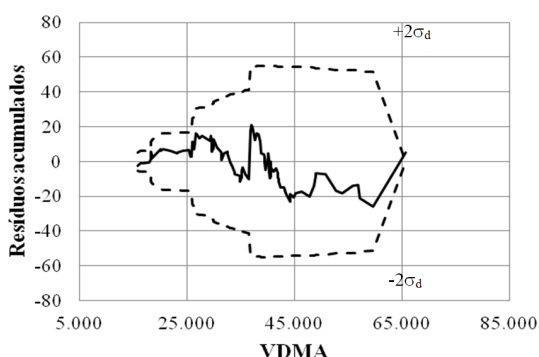


Figura 3. Gráfico de resíduos acumulados do modelo 2

por “saltos” verticais. No caso da Figura 3, pode-se identificar dois possíveis *outliers* um aos 37.000vph e outro no extremo do modelo (aprox. 65.000 vph). Nestes casos observaram-se 35 e 48 acidentes enquanto que o modelo previu 7 e 17 acidentes respectivamente. Esses pontos podem ser investigados de forma mais detalhada.

Outro importante aspecto a se considerar é a multicolineariedade existente entre as variáveis preditoras. A Tabela 5 apresenta a matriz de correlação para as variáveis preditoras utilizadas nessa investigação. Observa-se um alto grau de correlação entre as variáveis *nfx* e *cc* (0,86), entre *nfx* e *nap* (0,86) e entre *cc* e *nap* (0,96). Portanto, os modelos 6 e 8 possuem alta multicolineariedade, podendo até possuir boa capacidade preditora, porém seus coeficientes não permitem qualquer interpretação sobre relações de causa e efeito das variáveis canteiro central e número de aproximações nos acidentes de trânsito. Essa multicolineariedade é também confirmada pela mudança nos coeficientes das variáveis à medida que suas correlatas entram no modelo (Tabela 3).

Pela análise global dos indicadores, conclui-se que, apesar de não haver uma preferência definitiva por um modelo específico, a expressão com as variáveis *vdma* e *nfx* (modelo 2) apresenta características mais adequadas dentre os analisados. Considerando a formulação geral mais comum para os MPA em interseções e os coeficientes estimados, o modelo 2 tem a seguinte expressão matemática:

$$Y = 0,00591 \cdot (vdma_09^{0,52}) \cdot e^{0,28 \cdot nfx} \quad (2)$$

em que,

Y: número esperado de acidentes no ano de 2009 (interseções semaforizadas).

Cabe ressaltar que em virtude da elevada dependência entre o fluxo e o número de faixas, o modelo deve ser encarado predominantemente como preditor. Relações de causa e efeito em virtude de alterações no número de faixas devem ser analisadas com cuidado. A Figura 4 representa graficamente o modelo proposto.

Modelos de regressão, assim como qualquer outro tipo de modelo, precisam passar pelo processo de validação. Segundo Kutner *et al.* (2005), na ausência de uma amostra para se testar o desempenho do modelo, um esforço de validação pode ser feito por meio da verificação da consistência do modelo com os fundamentos teóricos acerca do fenômeno. A Figura 5 permite uma comparação visual entre o modelo proposto e o indicado no *Highway Safety Manual* (AASHTO, 2010) para interseções arteriais semaforizadas. Observa-se que o número esperado de acidentes cresce monotonicamente nos dois modelos e que as curvas possuem inclinações semelhantes. A análise

Tabela 5. Matriz de correlação das variáveis preditoras

	<i>nfx</i>	<i>nap</i>	<i>cc</i>	<i>v1_vdma09</i>	<i>v2_vdma09</i>	<i>vdma_09</i>
<i>nfx</i>	1					
<i>nap</i>	0,86	1				
<i>cc</i>	0,86	0,94	1			
<i>v1_vdma09</i>	0,55	0,55	0,55	1		
<i>v2_vdma09</i>	0,16	0,18	0,19	-0,28	1	
<i>vdma_09</i>	0,59	0,61	0,62	0,60	0,59	1

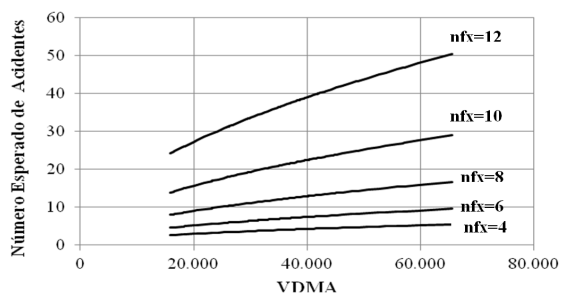


Figura 4. Representação gráfica do modelo 02

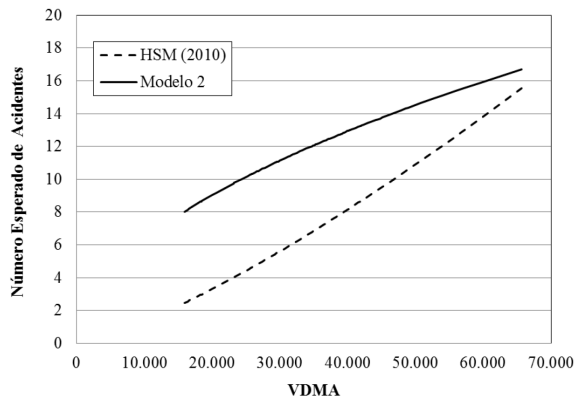


Figura 5. Representação gráfica do modelo 02 (8 faixas) e do HSM (AASHTO, 2010)

sugere, entretanto, que modelos desenvolvidos em outras jurisdições devem ser cuidadosamente avaliados tendo em vista uma possível recalibração para refletir as condições locais.

4. CONCLUSÕES

Este trabalho apresentou um esforço de desenvolvimento de modelos de previsão de acidentes para interseções semaforizadas em Fortaleza. Mais especificamente, foram discutidos aspectos relativos à formulação geral do modelo, estrutura mais adequada para a representação dos erros do modelo, definição das variáveis e coleta dos dados, estimação dos parâmetros do modelo e escolha dos modelos mais adequados. A amostra utilizada no desenvolvimento do modelo foi composta por 101 interseções semaforizadas de onde foram coletados os dados sobre a frequência total de acidentes para o ano de 2009. Como variáveis preditoras foram investigadas o volume diário médio anual, o número total de faixas da interseção, a presença de canteiro central e o número de aproximações.

Os resultados da modelagem indicaram que, para a estimação do número total de acidentes de trânsito, o modelo com o VDMA e número de faixas apresentou desempenho mais adequado dentre aqueles analisados. Pela elevada dependência entre as variáveis explicativas o modelo serve mais ao propósito de predição do que ao propósito de estimar relações de causa e efeito.

Para a estimação do número de acidentes de trânsito com vítimas feridas e fatais, entretanto, nenhuma das variáveis investigadas foi significativa. Acredita-se que a severidade dos acidentes nas interseções possa estar ligada ao avanço

semafórico que, por sua vez, sofre influência de variáveis como hora do dia, local da via e existência de fiscalização eletrônica, não contempladas nesse trabalho.

A consolidação do uso dessa ferramenta estatística (MPA) no ambiente brasileiro carece entretanto de outros trabalhos no sentido de aprimorar os bancos de dados de acidentes de trânsito, integrar esses bancos de dados às outras possíveis variáveis utilizadas na modelagem, investigar a transferência espacial e temporal dos modelos e apresentar metodologias de validação mais robustas.

REFERÊNCIAS BIBLIOGRÁFICAS

- AASHTO (2010) *Highway Safety Manual*. American Association of State Highway and Transportation Officials. 1st. Edition, Washington, D.C.
- Al-Ghamadi, A. S. (2003) Analysis of Traffic Accidents at Urban Intersections in Riyadh. *Accident Analysis and Prevention*, v. 35, n. 5, p. 717-24. DOI:10.1016/S0001-4575(02)00050-7.
- Alkaabi, A., D. Dissanayake, e R. Bird (2011) Analyzing Clearance Time of Urban Traffic Accidents in Abu Dhabi using Hazard-based Duration Modelling Method. In *Transportation Research Record*, TRB, Washington, D.C., n. 2229, p. 46-54. DOI:10.3141/2229-06.
- Bonneson, J. A., P. T. McCoy (1993) Estimation of safety at two-way stop-controlled intersections on rural highways. *Transportation Research Record*, n. 1401, TRB, Washington, D.C., p. 83-99.
- Cardoso, G. (2006) *Modelos para previsão de acidentes em vias arteriais urbanas*. Tese (Doutorado). Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal do Rio Grande do Sul-UFRGS.
- Cardoso, G. e L. Goldner (2007) Desenvolvimento e aplicação de modelos para previsão de acidentes de trânsito. *Transportes*, v. 15, n. 2, p. 43-51.
- Cunto, F. J. C. (2008) *Assessing Safety Performance of Transportation Systems using Microscopic Simulation*. Tese (Doutorado). Department of Civil and Environmental Engineering, University of Waterloo, Ontário. Disponível em: <http://hdl.handle.net/10012/4111>.
- Davis, G. A. (2004) Possible Aggregation Biases in Road Safety Research and a Mechanism Approach to Accident Modeling. *Accident Analysis and Prevention*, v. 36, p. 1119-1127. DOI:10.1016/j.aap.2004.04.002.
- Dinh-Zarr, T. B., 2008. United Nations Debates Road Safety this Spring. *Journal of the Institute of Transportation Engineers*, April 2008, p. 42-44.
- Ferraz, C., A. Raia, e B. Bezerra (2008) *Segurança no Trânsito*. Grupo São Francisco.
- Hakkert, A. S., L. Hoeherman e A. Mensah (1996) Levels of safety on interurban roads. *Transportation Research Record*, TRB, National Research Council, Washington, D.C., n. 1553, p. 95-102.
- Hauer, E. (2002) *Observational Before-after Studies in Road Safety*. (1ª ed.). Pergamon.
- Hauer, E. (2004) Statistical Road Safety Modeling. *Transportation Research Record*, No. 1897, TRB, Washington, D.C., p. 81-87. DOI:10.3141/1897-11.
- Hauer, E. e A. S. Hakkert (1989) The extent and implications of incomplete accident reporting. In *Transportation Research Record*, TRB, Washington, D.C., n. 1185, p. 1-10.
- Hauer, E., C. N. Ng Jerry e J. Lovell (1988) Estimation of safety at signalized intersections. *Transportation Research Record*, TRB, Washington, D.C., n. 1185, p. 48-61.
- IPEA (2003). Impactos sociais e econômicos dos acidentes de trânsito nas aglomerações urbanas. Instituto de Pesquisas Econômicas Aplicadas. Brasília.
- Kutner, M., C. Nachtsheim, J. Neter e W. Li (2005) *Applied Linear Statistical Models*. (5a. ed.). McGraw-Hill/Irwin
- Lord, D. (2000) *The prediction of accidents on digital networks: characteristics and issues related to the application of accident prediction models*. Tese (Doutorado), Universidade de Toronto, Canadá.
- Lord, D. (2006) Modeling motor vehicle crashes using poisson-gamma models: Examining the effects of low sample mean values and small sample size on the estimation of the fixed dispersion pa-

- parameter. *Accident Analysis and Prevention*, v. 33, p. 751–766. DOI:10.1016/j.aap.2006.02.001.
- McCullagh P. e J. A. Nelder (1989) *Generalized Linear Models – 2nd Edition*. Chapman & Hall/CRC.
- Miaou, S. P. (1996) Measuring the Goodness-of-Fit of Accident Prediction Models. FHWA Publication No. FHWA-RD-96/040. Federal Highway Administration, Washington, D.C.
- Oliveira, M. V. (2004). *A Natureza dos Padrões de Variação Espaço-temporal do Volume Veicular em Ambiente Urbano: Estudo de Caso em Fortaleza*. Dissertação (Mestrado). Programa de Pós-graduação em Engenharia de Transportes. Universidade Federal do Ceará, Fortaleza.
- Persaud, B., e T. Nguyen (1998) Disaggregate Safety Performance Models for Signalized Intersections on Ontario Provincial Roads. *Transportation Research Record*, TRB, Washington, D.C., n. 1635, p. 113–120. DOI:10.3141/1635-15.
- Persaud, B. e K. Mucsi (1995) Microscopic accident potential models for two-lane rural roads. *Transportation Research Record: Journal of the Transportation Research Board*, n. 1485, p. 134–139.
- Simonoff J. S. (2003). *Analyzing Categorical Data*. (1a. ed), Springer.
- Sawalha, Z. e T. Sayed (2006) Traffic accident modeling: some statistical issues. *Canadian Journal of Civil Engineering*. v. 33, p. 1115–1124. DOI:10.1139/L06-056.