

ASPECTOS CONCEITUAIS SOBRE O DESENVOLVIMENTO DE MODELOS DE PREVISÃO DE ACIDENTES DE TRÂNSITO

Flávio José Craveiro Cunto

Universidade Federal do Ceará
Departamento de Engenharia de Transportes

Christine Tessele Nodari

Universidade Federal do Rio Grande do Sul
LASTRAN – Laboratório de Sistemas de Transporte/UFRGS

RESUMO

Esse artigo tem por objetivo discutir a evolução dos modelos de previsão de acidentes de trânsito (MPA) e assim, esclarecer aspectos relevantes quanto aos principais conceitos e pressupostos envolvidos nessa temática. Ênfase é dada para a relevância dos modelos lineares generalizados (MLG) na evolução da estimação dos acidentes. Também são feitas considerações sobre o estado da arte no desenvolvimento dos MPA, onde são abordadas questões referentes à seleção das variáveis preditoras, dimensionamento da amostra, definição e estimação de sua expressão matemática e avaliação da qualidade do modelo. Como principal resultado do artigo espera-se ter contribuído para a melhor compreensão do papel dos MPA no estudo da segurança viária e dos conceitos envolvidos no seu desenvolvimento.

ABSTRACT

This paper aims to discuss the evolution of accident prediction models (APM) and thus to clarify important aspects regarding the main concepts and assumptions involved in this topic. Emphasis is given to the relevance of generalized linear models (GLM) on accident prediction. The state of the art in the development of APM are also discuss with emphasis on issues related to the selection of prediction variables, sample size, definition and estimation of its mathematical expression and evaluating of the goodness of fit related to the model. It is expected that this paper has contributed to a better understanding of the role of APM in road safety studies as well as the concepts and premises underlying their development.

1. INTRODUÇÃO

A maioria das análises de segurança em engenharia de transportes procura investigar duas questões fundamentais: 1) Quais são os locais com os maiores riscos de colisões os quais justifiquem intervenções de engenharia de segurança viária?; 2) Quais são as intervenções que produzirão a maior redução no risco de acidentes em relação ao investimento necessário à sua implementação?

Hauer (2002) observa que, de maneira geral, engenheiros e pesquisadores não podem alterar deliberadamente especificações geométricas e operacionais de componentes do sistema viário com o objetivo de investigar mudanças no nível de segurança de tais componentes, por motivos práticos e éticos. Desta forma, análises de segurança viária podem ser vistas como estudos observacionais (não experimentais) onde se procura a compreensão das relações de causa e efeito de fatores (variáveis) em um dado fenômeno (segurança viária), a partir da análise de eventos (acidentes de trânsito) que não foram formalmente desenhados para esse fim. (Rosenbaum, 2002; Cunto, 2008).

No contexto dos estudos observacionais, pesquisadores têm procurado representar a segurança viária como uma propriedade inerente às interseções e segmentos viários. Essa propriedade que é expressa pelo número esperado de acidentes κ em um dado intervalo de tempo, resulta da interação de um conjunto de fatores viários, humanos e ambientais presentes em determinada entidade viária (Hauer *et al.*, 1988; Lord, 2000).

Duas utilidades práticas de κ seriam então: 1) a seleção de pontos críticos, ou seja, entidades onde se observou uma quantidade de acidentes de trânsito K maior que o número esperado de

acidentes para entidades similares e 2) a comparação de cenários com diferentes atributos que permita a escolha de alternativas com base em critérios objetivos para expressar o desempenho da segurança viária. Esses usos atendem, respectivamente, as questões 1 e 2 formuladas no início dessa seção.

Ao longo das últimas duas décadas, diversas pesquisas procuraram contribuir para a estimação do número esperado de acidentes κ utilizando modelos estatísticos de regressão que relacionam o número observado de acidentes de trânsito com atributos geométricos e de operação da via, conhecidos como modelos de previsão de acidentes (MPA) ou funções de desempenho de segurança viária (Hauer *et al.*, 1988, Bélanger, 1994). Esses modelos são desenvolvidos através de técnicas estatísticas como a modelagem linear generalizada e lidam, com relativo sucesso, com a maioria dos problemas oriundos da natureza estocástica dos acidentes de trânsito (Hauer, 2002; Cunto, 2008). A primeira edição do *Highway Safety Manual* (AASHTO, 2010) dedica parte considerável de seu conteúdo à justificação, preceitos e desenvolvimento dos MPA em todas as etapas do planejamento estratégico, tático e operacional dos sistemas de transportes.

Para elucidar a importância dos modelos de previsão de acidentes (MPA) na pesquisa da segurança viária cabe a discussão sobre a diferença existente entre número (ou frequência) observado, previsto e esperado de acidentes. De forma sucinta, os MPA permitem a estimação do número previsto de acidentes em um determinado local. Tais modelos são desenvolvidos por meio de um conjunto de dados observados de vários locais similares. Já o número esperado de acidentes para um determinado local (por exemplo, para uma determinada interseção) é obtido pela combinação do número previsto pelo modelo combinado com o número observado neste local específico. Essa combinação de informações torna mais robusta a estimativa do número de acidentes que realmente deveria estar acontecendo neste local.

É importante considerar que a ocorrência de acidentes em um determinado local é considerada um fenômeno aleatório raro. Sendo assim, sabe-se que os acidentes ocorridos em um determinado local resultam de diversos fatores inter-relacionados. Alguns desses fatores são determinísticos e outros são estocásticos (aleatórios e imprevisíveis) (AASHTO, 2010). Essa fração estocástica indica que parte dos acidentes ocorre devido ao acaso e, portanto, não sofreu influência das características físicas e operacionais do local. Por esse motivo, o número observado de acidentes em um determinado local não é um bom indicador das reais condições de segurança deste local. A estimativa do número previsto de acidentes e a sua combinação com os dados locais para a obtenção do número esperado de acidentes é uma alternativa para se obter números mais confiáveis sobre as condições de segurança dos locais estudados.

Este trabalho tem por objetivo discutir a evolução dos modelos de previsão de acidentes de trânsito, ressaltando aspectos relacionados à sua base conceitual, escolha das variáveis preditoras, dimensionamento da amostra, definição e estimação de sua expressão matemática e avaliação da qualidade do modelo.

2. MODELOS LINEARES GENERALIZADOS NA SEGURANÇA VIÁRIA

Os estudos observacionais mais frequentemente aplicados em segurança viária são os do tipo “antes” e “depois” com ou sem grupo de comparação, estudos transversais (*cross-sectional studies*) e a meta-análise (Hauer, 2002; Persaud e Lyon, 2006; AASHTO, 2010). Uma característica comum aos estudos mencionados que pode afetar as conclusões obtidas é a

presença dos fatores de confusão ou covariáveis como fluxo veicular e raio de curvatura, velocidade operacional e largura de faixas de tráfego, etc. De acordo com Elvik e Vaa (2004) a modelagem estatística multivariada torna os estudos observacionais mais confiáveis por possibilitar um controle rigoroso dos fatores de confusão além de lidar com a presença de tendências e o fenômeno de regressão à média.

A elevada dispersão observada na frequência dos acidentes de trânsito quase sempre leva ao desenvolvimento de modelos com estrutura do erro assumindo a distribuição de Poisson ou binomial negativa (Poisson-Gama). Nestes casos, as técnicas de modelagem linear generalizada são consideradas plataformas adequadas para o desenvolvimento de modelos mais robustos, pois é possível o relaxamento da suposição a respeito da normalidade dos erros que existe, por exemplo, nos modelos de regressão linear (Cunto, 2008; Cardoso, 2006).

A estrutura geral dos modelos lineares generalizados (MLG), assim como nos modelos de regressão linear, pode ser dividida em um componente aleatório, um componente sistemático e uma função de ligação entre esses componentes (McCullagh e Nelder, 1989). O componente aleatório corresponde à variável que representa o fenômeno em estudo e a suposição em relação à sua distribuição de probabilidade e, no caso dos MLG, deve pertencer às distribuições da família exponencial.

Supondo Y_{it} a variável aleatória para a frequência de acidentes em uma entidade i em um dado intervalo de tempo t , uma das suposições mais frequentes é que Y_{it} segue a distribuição de Poisson com média λ_{it} . Nesse caso a probabilidade da ocorrência de y acidentes na entidade i é calculada pela expressão:

$$P(Y_{it} = y_{it}) = \frac{\lambda_{it}^{y_{it}} \cdot \exp(-\lambda_{it})}{y_{it}!} \quad (1)$$

Em que:

λ_{it} = valor esperado de acidentes de trânsito na entidade i no período t (igual à variância);

Essa estrutura de componente aleatório implica em termos práticos que se duas interseções, por exemplo, tiverem as mesmas características físicas e operacionais observadas, terão o mesmo valor esperado de acidentes. Entretanto ao utilizar os dados históricos de acidentes de trânsito, não é possível controlar todas as variáveis que influenciam no fenômeno e, além disso, as variáveis controladas não são coletadas com a mesma precisão. Essa linha de argumentação presente em Hauer (2001) aparentemente justifica os resultados empíricos obtidos por Bonneson e McCoy (1993) indicando que os acidentes apresentam dispersão incompatível com a estrutura do erro da distribuição de Poisson.

Uma alternativa para contornar o problema é, ao invés de assumir um valor fixo λ_{it} para todas as entidades com as mesmas características, considerar então (Hauer, 2004; Lord, 2006):

$$\lambda_{it} = \nu_{it} \theta_i \quad (2)$$

Em que ν_{it} é a média de todos os λ_{it} para uma população de entidades com as mesmas características da entidade i no período t ; $\theta_i \sim Gama(1, 1/\phi)$ é o multiplicador que representa o quanto o valor esperado de uma entidade difere da média de todos os valores esperados das

entidades com característica idênticas e $1/\phi_i$ é a variância, também denominado de parâmetro de “superdispersão”.

Pode-se demonstrar que essa conjugação de distribuições Poisson-Gama equivale a assumir desta feita $Y_{it} \sim BN(\lambda_{it}, \lambda_{it} + \lambda_{it}^2/\phi)$ (Hauer, 2004). Observe-se que o termo λ_{it}^2/ϕ na variância representa a dispersão adicional à distribuição de Poisson presente nos dados. A estrutura da função densidade de probabilidade Poisson-Gama (Binomial Negativa) para a ocorrência de y_{it} acidentes é calculada pela expressão:

$$P(Y_{it} = y_{it}; \phi, \lambda_{it}) = \frac{\Gamma(y_{it} + \phi)}{\Gamma(\phi)y_{it}!} \left(\frac{\phi}{\lambda_{it} + \phi}\right)^\phi \left(\frac{\lambda_{it}}{\lambda_{it} + \phi}\right)^{y_{it}} \quad (3)$$

Na estrutura tradicional dos MLG, o componente sistemático é formado pelas variáveis que “explicam” parte da variabilidade encontrada na distribuição de Y_{it} através de uma função linear ou preditor linear. A expressão matemática para o preditor linear pode ser descrita como:

$$\eta = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p \quad (4)$$

Em que β_i são os coeficiente do modelo e x_i as variáveis preditoras (fluxo veicular, número faixas, largura de acostamento, etc.). O último componente estrutural dos MLG denominado de função de ligação permite a união entre os componentes aleatório $E\{Y_{it}\} = \lambda_{it}$ e o preditor linear da seguinte forma:

$$\eta = g(\lambda_{it}) \quad (5)$$

Para o desenvolvimento dos MLG a função $g(\cdot)$ deve ser do tipo monotônica e diferenciável como ocorre nos casos das funções identidade e logarítmica para as distribuições de Poisson e Binomial (McCullagh e Nelder, 1989).

3. DESENVOLVIMENTO DOS MODELOS DE PREVISÃO DE ACIDENTES

A construção de modelos de regressão serve ao propósito fundamental de contribuir para identificar a natureza sistemática ou aleatória da influência de fatores relacionados a um determinado fenômeno (McCullagh e Nelder, 1989). Consequentemente, faz-se necessário que o modelador defina de forma satisfatória o fenômeno de interesse para então iniciar o processo de seleção dos possíveis fatores de influência, o desenho do experimento, a estimação da expressão matemática e finalmente a escolha do modelo mais eficiente. Uma adaptação desse processo construtivo para o caso da segurança viária está proposta na Figura 01 e será discutida de forma mais detalhada nas próximas seções.

3.1. Objetivo do estudo de segurança viária

A definição do objetivo do estudo de segurança viária estabelece a diretriz geral para a escolha das possíveis variáveis explicativas ou preditoras do modelo, além de auxiliar no desenho do experimento. Hauer (2004) define duas categorias básicas de estudos, sendo a primeira relacionada com a previsão da frequência de acidentes esperada em uma entidade viária em função de seus atributos e, a segunda categoria que procura estimar a mudança na frequência dos acidentes de trânsito ocasionada pela alteração de atributos específicos das entidades. Desta forma observa-se a tentativa de obtenção de relações causais nas questões

pertinentes à segunda categoria. Hauer (2004) afirma ainda que apesar da semelhança computacional, os modelos preditivos (categoria 1) são consideravelmente mais corriqueiros e confiáveis do que os modelos “causais” (categoria 2).

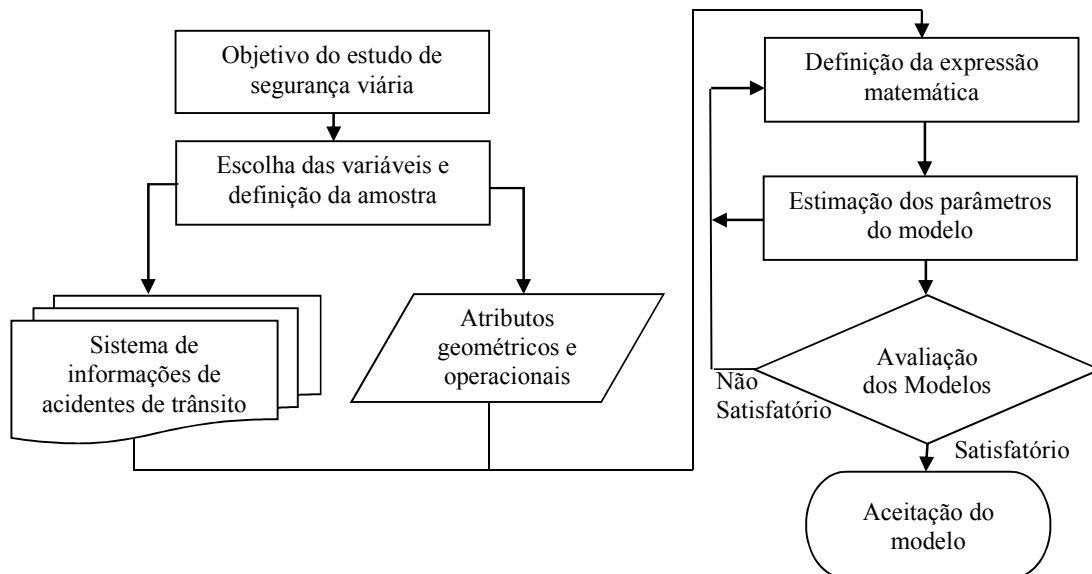


Figura 01: Metodologia proposta para a calibração de MPA

3.2. Escolha das variáveis e definição da amostra

O passo subsequente à definição do objetivo do estudo é a escolha das variáveis preditoras a serem consideradas na construção do MPA. De maneira geral essas variáveis podem ser divididas em fatores que refletem a exposição dos usuários a eventos que possam gerar acidentes como o período de análise, fluxo veicular e a extensão do segmento, e fatores ambientais que influenciem a probabilidade do envolvimento em acidentes de trânsito (Bonneson e McCoy, 1993; Cardoso e Goldner, 2007).

Observa-se que, de maneira geral, os modeladores escolhem as variáveis preditoras utilizando critérios subjetivos com base na experiência sobre a influência da variável nos acidentes de trânsito ou ainda considerando aspectos do custo de coleta de certos atributos. A Tabela 01 apresenta exemplos de variáveis preditoras utilizadas em modelos propostos para interseções e segmentos viários.

O dimensionamento da amostra nos estudos observacionais de segurança viária é um aspecto ainda sem critérios objetivos. Em virtude da impossibilidade de encontrar entidades perfeitamente iguais, o controle do “experimento” é feito separando-se grupos com características similares como interseções semaforizadas, segmentos de pista simples, etc.

A elevada dispersão nos acidentes de trânsito provoca normalmente a necessidade de um número elevado de observações para a obtenção de valores estatisticamente significativos dos parâmetros. O aumento da amostra pode ser obtido tanto pelo acréscimo de entidades como pelo aumento do tempo de estudo (Bonneson e McCoy, 1993).

Tabela 1: Variáveis preditoras exploradas nos MPA

Entidade ^(*)	Variáveis de Exposição	Variáveis Ambientais
1	VDMA _{total} , VDMA _{via principal} , VDMA _{via secundária}	# de faixas, número de aproximações e configuração geométrica (“cruz” ou “tê”), existência de canalização de tráfego e separadores centrais, ciclofaixas e ciclovias
2	Comprimento do segmento, VDMA, fluxo horário de pedestres, fluxo horário de motocicletas, fluxo horário de bicicletas	Largura e quantidade de faixa de tráfego, largura total da pista, uso do solo, densidade de acessos, separadores centrais, limite de velocidade, velocidade de fluxo livre, existência de paradas de ônibus

(*) 1-Interseções; 2-Segmentos viários

Fontes: Bonneson e McCoy (1993), Persaud e Mucsi (1995), Hakkert *et al.* (1996), Greibe (2003), Cardoso e Goldner (2007).

Lord (2006) investigou a influência do tamanho da amostra e da média dos acidentes no parâmetro de dispersão (ϕ) da distribuição binomial negativa usando simulação Monte Carlo. Os resultados sugeridos por Lord indicam que quanto menor a média dos acidentes maior deve ser o tamanho da amostra e como, limite inferior, desaconselhou o desenvolvimento de MPA com a estrutura BN para amostras menores do que 100 observações.

3.3. Coleta dos dados

Para o desenvolvimento de MPA é necessário a coleta de três tipos de dados: (i) dados de registro dos acidentes; (ii) dados das características físicas dos locais onde ocorreram os acidentes e (iii) dados de fluxo de veículos e pessoas nestes locais. Muitas vezes essas informações estão disponíveis de forma dispersa e necessitam tratamento adicional para serem úteis no processo de estimação dos modelos.

O registro do acidente de trânsito é o dado primário utilizado no desenvolvimento dos MPA e, conseqüentemente, os tipos de modelos possíveis irão depender da quantidade e nível de agregação da informação disponibilizada pelos sistemas de informações de acidentes de trânsito (SIATs). Dentre as informações desejáveis podem-se citar: data, hora, local da ocorrência (de preferência georeferenciado), tipo do acidente, severidade, idade, sexo e categoria das vítimas e tipo dos veículos envolvidos.

Com relação à qualidade da informação, os procedimentos de coleta, tabulação e análise de consistência dos SIATs devem procurar minimizar as principais fontes de erros que podem afetar a construção de MPA, dentre elas podem-se citar (Zegeer *et al.*, 1987; Hauer e Hakkert, 1989; Farmer, 2003; Queiroz, 2003):

- Sub-registro de acidentes com menor severidade;
- Diferenças no nível de detalhamento da informação com a severidade;
- Duplicidade de registros pela existência de várias fontes de coleta;
- Imprecisão na localização do acidente;
- Alterações na metodologia de coleta com o tempo.

As variáveis normalmente adotadas para representar a exposição de veículos e pedestres são o fluxo e o comprimento do trecho para o caso específico dos segmentos viários. No caso do fluxo veicular, a melhor medida é o Volume Médio Diário Anual (VDMA), pois está baseado em múltiplas contagens ao longo do ano. A obtenção do VDMA pode ser feita por meio de contagens 24 horas registradas automaticamente ou pela estimativa baseada em contagens amostrais expandidas para todo o período. Em um segmento viário, o VDMA é o volume de tráfego médio diário observado nas 24 horas do dia, nos dois sentidos da rodovia, ao longo de um ano. No caso de interseções, são considerados dois valores de VDMA, o da via principal e da secundária. Deve-se observar que, no caso dos valores de VDMA das duas aproximações da via principal serem diferentes, adota-se o maior valor para aquela via. O mesmo deve ser observado para a via secundária (AASHTO, 2010; NCHRP, 2008).

Os dados relativos às características físicas dos locais onde ocorreram acidentes tipicamente incluídos nos MPA são, no caso de interseções: o número de aproximações, o número de faixas nas aproximações, o tipo de controle, a configuração operacional e a existência de canteiros ou separadores centrais e a existência de iluminação artificial.

Para modelagem de segmentos de vias e/ou rodovias as características mais frequentemente adotadas são, além do comprimento do segmento, a classe da rodovia, o número de faixas, a existência de controladores de velocidade, a presença de estacionamento na via, a presença de acostamento e a presença de canteiros ou separadores centrais (AASHTO, 2010). Esses dados podem ser obtidos em inventários mantidos pelos órgãos gestores, em levantamentos de campo e, em alguns casos, através de pesquisa em imagens de satélite atualizadas e disponíveis na *internet*.

3.4. Expressão Matemática e Estimação dos Parâmetros dos MPA

Sawalaha e Sayed (2006) sugerem que os MPA devem, inicialmente, satisfazer critérios lógicos como não permitir a obtenção de números negativos de acidentes de trânsito para combinações das variáveis preditoras. Também não são permitidos valores positivos, quando as variáveis preditoras relacionadas à exposição como VDMA e comprimento do trecho são nulas.

A grande maioria das pesquisas disponível na literatura aponta para a utilização de modelos que podem incluir componentes multiplicativos e aditivos com a seguinte formulação geral (Hakkert *et al.*, 1996; Sawalha e Sayed, 2006; AASHTO, 2010):

$$Y = \alpha \left\{ \left[\prod_i (A_i)^{\beta_i} \right] \cdot e^{\sum_j (\gamma_j B_j)} + \left(\sum_k \delta_k C_k \right) \right\} \quad (6)$$

em que:

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

A, B, C = vetores de variáveis preditoras;

$\alpha, \beta, \gamma, \delta$ = parâmetros do modelo.

Inicialmente, as variáveis preditoras do vetor A são aquelas correspondentes à exposição como VDMA, comprimento do trecho, fluxo de pedestres, etc. No vetor de variáveis preditoras do tipo B observa-se principalmente a presença de atributos físicos e operacionais que alteram a probabilidade de ocorrência de acidentes ao longo de uma porção significativa da via como número e largura de faixas de tráfego, presença de separador central,

acostamentos, dentre outros. Finalmente, as variáveis da parte aditiva são aquelas que influenciam a ocorrência de acidentes de trânsito em pontos específicos de segmentos viários, como uma ponte estreita, entradas/saídas de lotes, retorno em nível, etc. (Hauer, 2004).

O coeficiente α acima, também chamado de parâmetro de escala, pode ser interpretado como um fator que reflete alterações anuais no padrão dos acidentes em função de mudanças em aspectos globais como clima, qualidade dos dados, crescimento urbano, alterações na fiscalização e legislação vigente. (Hauer, 2004; Lord e Persaud, 2000). Ainda, de acordo com Lord e Persaud (2000), para uma mesma amostra com vários anos de dados disponíveis, os MPA anuais devem ser estimados com os mesmos coeficientes β alterando-se somente o parâmetro de escala α .

A estimação dos parâmetros do modelo é feita na maioria das vezes pelo método da máxima verossimilhança o qual pode ser obtida com algoritmos de otimização como o de Newton-Rapson (Hauer, 2002; Lord, 2006). Atualmente existe um número considerável de aplicativos computacionais como o SAS, SPSS e R que possuem rotinas pré-definidas para a estimação dos parâmetros em função da estrutura assumida para o erro (Poisson ou binomial negativa).

Um aspecto importante sobre a definição da expressão matemática é o processo de inserção das variáveis no modelo. De acordo com Hauer (2004) deve-se iniciar com um modelo base contendo apenas a variável de exposição, acrescentando-se as demais variáveis de forma seqüencial. O logaritmo da máxima verossimilhança (ℓ) do modelo base é comparado com os valores de ℓ dos modelos com uma segunda variável. A variável que, ao ser acrescentada resulte no maior aumento nesse parâmetro, será a próxima a ser incluída. O processo é repetido até que todas as variáveis significativas estejam contempladas.

Um indicativo da correlação entre as variáveis é obtido analisando-se a alteração nos parâmetros estimados (β , γ , δ) ao inserir uma nova variável. Pequenas variações nesses parâmetros indicam baixa correlação entre as variáveis. Nestes casos, acredita-se que os modelos desenvolvidos possam ser usados em interpretações causais, do contrário os modelos devem ser usados somente como preditivos (Hauer, 2004).

3.5. Avaliação dos Modelos Propostos

A qualidade do modelo proposto deve ser avaliada com o objetivo de comprovar sua capacidade de predição em todas as faixas de valores esperados para as variáveis independentes. Para os modelos lineares generalizados que adotam a não normalidade dos erros, os indicadores tradicionais como o coeficiente de determinação R^2 e o R^2 ajustado não se mostram eficientes (Miaou, 1996, Hauer, 2004).

Após a verificação inicial da significância de cada coeficiente do modelo (valor p) um número considerável de indicadores matemáticos e gráficos têm sido aplicados aos modelos lineares generalizados, dentre eles ressaltam-se os a estatística generalizada de Pearson χ_p^2 , o parâmetro de dispersão σ_d (*overdispersion parameter*), o desvio escalonado S_p (*scaled deviance*), o critério de informação de Akaike (AIC) e o gráfico de resíduos acumulados (*CURE plot*) (McCullagh e Nelder, 1989, Bonneson e McCoy, 1993; Lord, 2000; Hauer, 2004).

A estatística generalizada de Pearson é uma media da discrepância de ajuste de um modelo ao conjunto de dados cuja expressão é descrita por:

$$\chi_p^2 = \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)} \quad (7)$$

Em que:

$\hat{\mu}_i$ = valores da variável resposta estimados pelo modelo

$V(\hat{\mu}_i)$ = função de variância estimada pelo modelo

A estatística χ_p^2 pode ser utilizada para confirmar a significância do modelo. Se $\chi_p^2 > \chi_{\alpha, n-p-1}^2$ em que α é o nível de significância, n é o número de observações e p o número de parâmetros, significa que a variância explicada pelo modelo difere da variância observada nos dados. De acordo com Bonneson e McCoy (1993), essa medida não apresenta bom desempenho para amostras pequenas (resultados assintóticos para distribuições não normais), sendo recomendada sua utilização em conjunto com outros indicadores.

O parâmetro de dispersão σ_d é obtido pela razão entre a estatística generalizada de Pearson e a diferença entre o número de observações e o número de parâmetros do modelo. Neste caso a obtenção de valores próximos a 1 indica que a variância assumida na estrutura do modelo é similar àquela observada nos dados. Valores de σ_d superiores a 1 indicam dispersão dos dados acima do modelo e vice-versa. Observa-se a utilização frequente de σ_d para avaliar a distribuição estatística mais adequada para os erros entre Poisson e binomial negativa (Bonneson e McCoy, 1993; Lord, 2000).

O desvio escalonado (S_p) compara a máxima verossimilhança do modelo completo (saturado) e do modelo proposto com a expressão dada por:

$$S_p = 2(\hat{\ell}_n - \hat{\ell}_p) \quad (8)$$

Em que:

$\hat{\ell}_n$ = valor do logaritmo da máxima verossimilhança – modelo saturado;

$\hat{\ell}_p$ = valor do logaritmo da máxima verossimilhança – modelo proposto;

De acordo com esse critério, um bom modelo deve utilizar poucos parâmetros e ainda assim apresentar pequeno desvio em comparação com o modelo saturado (número de parâmetros igual ao número de observações). Em geral pode-se considerar que $S_p \sim \chi_{n-p}^2$, permitindo, desta forma, a execução de um teste estatístico para confirmar a significância do desvio escalonado.

O critério de informação de Akaike (AIC) pode ser aplicado na comparação entre possíveis modelos para um mesmo conjunto de dados. Este indicador expressa a diferença entre a variância e o viés do modelo proposto de acordo com a expressão (Hu, 2007):

$$AIC = 2k - 2\hat{\ell}_p \quad (9)$$

Em que:

k = número de parâmetros do modelo (variância);

$\hat{\ell}_p$ = valor do logaritmo da máxima verossimilhança – modelo proposto (viés);

Para modelos com elevado número de parâmetros a variância será baixa, entretanto o modelo será viesado, ou seja, válido somente para aqueles dados. Por outro lado, com poucos parâmetros o modelo terá um pequeno viés, porém com elevada variância. Procura-se então equilibrar viés/variância escolhendo o modelo com o menor AIC.

A escolha do modelo mais adequado pode ser ainda auxiliada pela análise de gráficos que confrontem valores observados e previstos, ou então que mostrem os valores residuais do modelo ao longo dos valores possíveis das variáveis preditoras. Hauer e Bamfo (1997) mostram que os gráficos de resíduos tradicionais nem sempre são informativos e propõem a utilização nos MPA de gráficos de resíduos acumulados ou *CURE plot*. Um exemplo desse tipo de representação gráfica é apresentado na Figura 2.

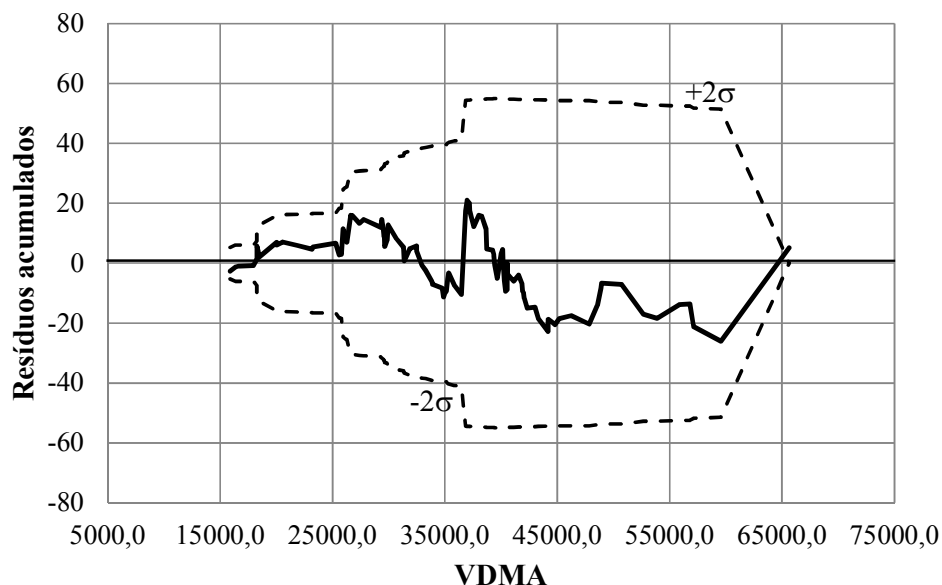


Figura 2: Exemplo de gráfico de resíduos acumulados (Fonte: Cunto *et al.*, 2011)

Para modelos satisfatórios, o gráfico de resíduos acumulados deve apresentar uma oscilação moderada em torno de zero. Recomenda-se ainda a construção de uma envoltória usando o dobro do desvio padrão como critério objetivo para confirmar o desempenho do modelo. Sempre que o gráfico ultrapassar os limites ($\pm 2\sigma$) delimita-se uma área em que o modelo é inadequado. Trechos ascendentes no gráfico delimitam áreas nas quais o modelo subestima a frequência observada e vice-versa. É possível ainda a obtenção de um indicativo de valores extremos (*outliers*) da amostra, os quais se apresentam em “saltos” bruscos no gráfico. Hauer (2004) apresenta um procedimento detalhado para a confecção de gráficos de resíduos acumulados.

No exemplo da Figura 02, observa-se que o modelo proposto pode ser utilizado no intervalo de valores observados de VDMA (15.000 a 65.000 veículos) e que duas observações (aproximadamente aos 37.000 e aos 65.000vph) merecem atenção especial.

A validação de qualquer modelo é uma etapa importante onde o objetivo é avaliar o desempenho do modelo proposto em um novo conjunto de dados. Acredita-se, portanto que os

MPA precisam ter seus desempenhos devidamente validados, muito embora este aspecto não tenha sido abordado com frequência na literatura.

A ausência de critérios mais objetivos de validação dos MPA pode ser atribuída, pelo menos em parte, ao processo tradicional de validação onde se aplica o modelo a um novo conjunto de dados. Nos estudos observacionais de segurança viária nem sempre é possível utilizar dados de outras regiões e, dentro de uma mesma jurisdição normalmente não é possível a obtenção de duas amostras. Contudo, Kutner *et al.* (2005) afirmam que a validação de um modelo também pode ser conduzida por meio da verificação da consistência do modelo com os fundamentos teóricos acerca do fenômeno e com os resultados empíricos obtidos na literatura.

4. CONSIDERAÇÕES FINAIS

O desenvolvimento e a utilização de MPA podem ser considerados tópicos relativamente recentes no estudo da segurança viária. Em decorrência disso, existem muitas questões sobre o tema ainda pouco discutidas e não suficientemente esclarecidas, mesmo no meio acadêmico, limitando o desenvolvimento das pesquisas nessa área. Nesse contexto, o presente artigo pretendeu uniformizar e esclarecer alguns termos e procedimentos atualmente vigentes na área da modelagem de acidente viários.

Entre outros aspectos relevantes, o artigo buscou-se esclarecer a adequação do uso das técnicas de MLG no desenvolvimento de MPA. Adicionalmente, foram discutidas as razões para a adoção da distribuição de Poisson ou da binomial negativa (Poisson-gama) na representação da estrutura do erro.

Na trajetória do desenvolvimento dos MPA muitos desafios foram enfrentados e solucionados e outros continuam exigindo investigação. Como desafio, permanece a necessidade de se investigar as possibilidades e as limitações da transferência, tanto espacial como temporal, dos modelos estimados, bem como suas aplicações no contexto decisório dos diversos níveis do planejamento de transportes. Provavelmente, uma das principais questões que carecem de reposta diz respeito à capacidade dos modelos em prever, não apenas a frequência de acidentes, mas também a relação causal entre as variáveis estudadas e a ocorrência dos mesmos. Espera-se que as reflexões aqui propostas possam contribuir para uma melhor compreensão por parte da comunidade técnico-científica brasileira a respeito da modelagem estatística de acidentes.

REFERÊNCIAS BIBLIOGRÁFICAS

- AASHTO (2010) *Highway Safety Manual*. American Association of State Highway and Transportation Officials. 1st. Edition, Washington, D.C.
- Bélanger, C. (1994) Estimation of safety of four-legged unsignalized intersections. *TRR: Journal of the Transportation Research Board*, No. 1467, TRB, National Research Council, Washington, D.C., p. 23-29.
- Bonneson, J.A., McCoy, P.T. (1993) Estimation of safety at two-way stop-controlled intersections on rural highways. *TRR: Journal of the Transportation Research Board*, No. 1401, TRB, National Research Council, Washington, D.C., p. 83-99.
- Cardoso, G. (2006) Modelos para previsão de acidentes em vias arteriais urbanas. Tese de Doutorado, Programa de Pós-Graduação em Eng. de Produção, Universidade Federal do Rio Grande do Sul–UFRGS.
- Cardoso, G. Goldner, L. (2007) Desenvolvimento e aplicação de modelos para previsão de acidentes de trânsito. *Transportes*, v.XV, p.43-51.
- Cunto, F.J.C. (2008) *Assessing Safety Performance of Transportation Systems using Microscopic Simulation* Tese de Doutorado, Department of Civil and Env. Engineering, Universidade de Waterloo, Ontário.

- Cunto, F.J.C., Castro Neto, M.M., Barreira, D.S. (2011) Modelos de Previsão de Acidentes em Interseções Semaforizadas de Fortaleza. *Artigo submetido ao XXI Congresso de Pesquisa e Ensino em Transportes*, Associação Nacional de Pesquisa e Ensino em Transportes-ANPET, Belo Horizonte, MG.
- Elvik, R., Vaa T. (2004) *The Handbook of Road Safety Measures*. (1^a ed.), Elsevier.
- Farmer, C.M. (2003) Reliability of Police-Reported Information for Determining Crash Injury and Severity. *Traffic Injury Prevention*, v.4, p.38-44
- Greibe, P.(2003) Accident prediction models for urban roads. *Accident Analysis and Prevention*, v.35,p.273-285.
- Hakkert, A.S., Hocherman, I., Mensah, A. (1996) Levels of safety on interurban roads. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1553, TRB, National Research Council, Washington, D.C., p. 95-102.
- Hauer, E. (2001). Overdispersion in modelling accidents on road sections and in empirical Bayes estimation. *Accident Analysis and Prevention*, v.33, p.799-808.
- Hauer, E. (2002) *Observational Before-after Studies in Road Safety*. (1^a ed.), Pergamon.
- Hauer, E. (2004) Statistical Road Safety Modeling. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1897, TRB, National Research Council, Washington, D.C., p. 81-87.
- Hauer, E. e Hakkert, A.S. (1989) The extent and implications of incomplete accident reporting. *TRR: Journal of the Transportation Research Board*, No. 1185, TRB, National Research Council, Washington, D.C., p.1-10.
- Hauer, E., Bamfo, J. (1997) Two Tools for Finding What Function Links the Dependent Variable to the Explanatory Variables. ICTCT Conference Proceedings. Lund, Suécia, 1997.
- Hauer, E., Ng Jerry, C.N., Lovell, J. (1988) Estimation of safety at signalized intersections. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1185, TRB, National Research Council, Washington, D.C., p. 48-61.
- Hu, S. (2007) Akaike Information Criteria. Notas de Aula. Center for Research in Scientific Computation. North Carolina State University, Carolina do Norte.
- Kutner, M.H., Nachtsheim, C.J., Neter, J., Li, W. (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 de Doutorado, Universidade de Toroto, 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 parameter. *Accident Analysis and Prevention*, v. 33, p. 751-766.
- Lord, D., Persaud, B.N. (2000) Application of the generalized estimating equations procedure. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1717, TRB, National Research Council, Washington, D.C., p. 102-108.
- 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.
- McCullagh P., Nelder, J.A. (1989) *Generalized Linear Models – 2nd Edition*. Chapman & Hall/CRC.
- NCHRP. (2008) Research Results Digest 329: Highway Safety Manual Data Needs Guide. National Cooperative Highway Research Program, TRB. National Research Council, Washington, DC.
- Persaud, B., Lyon, C. (2006) Safety Evaluation of Offset Improvements for Left-Turn Lanes. FHWA Publication No. FHWA-HRT-09-036. Federal Highway Administration, Washington, D.C.
- Persaud, B., Mucsi, K. (1995) Microscopic accident potential models for two-lane rural roads. *TRR: Journal of the Transportation Research Board*, No. 1485, TRB, National Research Council, Washington, D.C., p. 134-139.
- Queiroz, M.P. (2003). Análise Espacial dos Acidentes de Trânsito do Município de Fortaleza. Dissertação de Mestrado. Programa de Pós-graduação em Engenharia de Transportes – PETRAN. Universidade Federal do Ceará, Fortaleza
- Rosenbaum, P.R. (2002) *Observational studies*. Editora Springley-Verlag, Nova Iorque.
- Sawalha, Z., Sayed, T. (2006) Traffic accident modeling: some statistical issues. *Canadian Journal of Civil Engineering*, v.33, p. 1115-1124.
- Zegeer, C.V., Hummer, J., Reinfurt, D., Herf, L., Hunter, W. (1987) Safety Effects of Cross-section Design for Two-lane Roads. FHWA Publication No. FHWA-RD-87/008. Federal Highway Administration, Washington, D.C.

Flávio José Craveiro Cunto (flaviocunto@det.ufc.br). Departamento de Engenharia de Transportes, Universidade Federal do Ceará. Campus do PICI, s/n – Bloco 703 – CEP. 60455-760 – Fortaleza, CE, Brasil.

Christine Tessele Nodari (piti@producao.ufrgs.br) Departamento de Engenharia de Produção e Transportes, Universidade Federal do Rio Grande do Sul. Av.Osvaldo Aranha, 99 – sala 513 - CEP. 90035-001 – Porto Alegre, RS, Brasil.