

Revista Brasileira de Cartografia (2016), N^o 68/2 Edição Especial Aplicações dos SIG: 263-273
Sociedade Brasileira de Cartografia, Geodésia, Fotogrametria e Sensoriamento Remoto
ISSN: 1808-0936

ANÁLISE DA RELAÇÃO ENTRE AS CARACTERÍSTICAS DAS INTERSEÇÕES, O USO DO SOLO E A OCORRÊNCIA DE ACIDENTES DE TRÂNSITO NA ZONA SUL DO RIO DE JANEIRO

*Analysis of the Link between the Characteristics of the Intersection, the Land Use
and Traffic Accidents Frequency in South Zone of Rio de Janeiro*

Camila Marangoni¹, Marcos Rocha¹, Paulo Maranhão² & Carlos David Nassi³

¹Instituto Militar de Engenharia - IME

Seção de Engenharia Cartográfica

Praça General Tibúrcio, 80. CEP 22290-270 - Rio de Janeiro – RJ - Brasil
eng.camilamarangoni@gmail.com, menesesrocha@yahoo.com.br

²Instituto Militar de Engenharia – IME

Seção de Engenharia Básico

Praça General Tibúrcio, 80. CEP 22290-270 - Rio de Janeiro – RJ - Brasil
maranhao@ime.eb.br

³Universidade Federal do Rio de Janeiro – UFRJ

Programa de Engenharia de Transportes- COPPE

Av. Horácio Macedo 2030. Bloco H, sala 106, Cidade Universitária. CEP 21941-972 - Rio de Janeiro – RJ - Brasil
nassi@pet.coppe.ufrj.br

Recebido em 17 de Novembro, 2014/ Aceito em 31 de Maio, 2015

Received on November 17, 2014/ Accepted on May 31, 2015

RESUMO

Este artigo busca verificar a relação entre as características das interseções, o uso do solo e o número de acidentes na zona sul do Rio de Janeiro. Para tal, utilizaram-se os coeficientes de Pearson para obter a correlação das variáveis independentes entre si e com o número de acidentes. Para avaliar o quão as mesmas explicariam a distribuição dos acidentes empregaram-se dois tipos de modelos estatísticos: modelos de regressão múltipla e modelos lineares generalizados. O número de interseções semaforizadas e a extensão das vias com áreas de lazer e de uso misto apresentaram os maiores valores de correlação com o número de acidentes. No entanto, a variável extensão das vias com uso misto não foi empregada na modelagem por ser correlacionada com o número de vias semaforizadas. O modelo de regressão múltipla apresentou resultado superior ao obtido pelo modelo linear generalizado com distribuição binomial negativa.

Palavras-chave: Modelagem de Acidentes, Pearson, Modelo de Regressão Múltipla, Binomial Negativa.

ABSTRACT

This paper aims to analyse the link between the intersection characteristics, land use and the number of accidents in the south zone of Rio de Janeiro. For this purpose, Pearson's correlation was utilized to correlate the independent variables with each other and with the number of accidents. To assess how these variables explain the distribution of the accidents, two types of statistical models were employed: multiple regression models and generalized linear models.

The number of signalized intersections and the road length with recreation areas and mixed use showed the highest correlation with the number of accidents. However, the road length with mixed use variable has not been utilized in modeling due to be correlated with number of signalized intersections. The multiple regression model showed best results than generalized liner model with negative binomial distribution.

Keywords: Accident Modeling, Pearson Correlation, Multiple Regression Model, Negative Binomial.

1. INTRODUÇÃO

A morte por acidentes de trânsito é atualmente a maior causa de morte na população entre 15 e 29 anos e será a 5ª maior causa de morte no mundo considerando todas as idades em 2030 (WHO, 2009). O Brasil, por sua vez, ocupa o 5º lugar entre os países com maior número de acidentes de trânsito. Tal fato fez com que a ONU declarasse o período de 2011-2020 como sendo a Década de Ação pelo Trânsito Seguro, na qual os governos se comprometem a tomar medidas visando diminuir o número de acidentes.

O emprego de modelos estatísticos pode ser um poderoso aliado na compreensão da distribuição espacial e dos fatores que possam estar relacionados com os acidentes, sendo comumente empregado nos países desenvolvidos, principalmente nos países da Europa Ocidental, Estados Unidos e Austrália. No Brasil, por sua vez, poucos estudos sobre modelagem de acidentes vêm sendo desenvolvidos (CUNTO, 2011). Grande parte dos mesmos detém-se mais na análise exploratória do que na modelagem propriamente dita.

Este artigo emprega os modelos de regressão linear multivariada e os lineares generalizados na modelagem do número de acidentes em um dado local em função das características das interseções das vias e do uso do solo. Busca-se, desta forma, oferecer subsídios para uma melhor compreensão da distribuição dos acidentes no Brasil e verificar se as variáveis explicativas empregadas nos países desenvolvidos se aplicam à realidade brasileira.

2. FUNDAMENTAÇÃO TEÓRICA

Os modelos estatísticos de previsão de acidentes podem ser classificados como aqueles que levam em consideração somente o número total de acidentes em um dado local, conhecido como modelo de frequência de acidentes e aqueles que consideram não somente o total, mas os dividem por severidade, que vai desde os acidentes com somente

danos materiais até aqueles com mortes fatais (ANASTASOPOULOS & MANNERING, 2011).

Os modelos utilizados neste estudo são os de frequência, tendo em vista que não se tem as informações sobre a severidade dos acidentes. Tal tipo de modelagem tem como característica a de não empregar como variáveis explicativas as características dos acidentes, tais como aquelas do condutor, da via e dos veículos envolvidos nos mesmos, o que o torna adequado para o caso do Brasil, onde ocorrem comumente preenchimentos incorretos ou incompletos dos Boletins de Ocorrência. Empregam-se no lugar outras variáveis, tais como aquelas relacionadas com as características socioeconômicas da população do local onde ocorreu o acidente, uso do solo e características geométricas e de conectividade das vias (SAVOLAINEN et al, 2011). Lord e Mannering (2010) fizeram uma extensa revisão sobre os diversos erros que podem ocorrer na análise de dados de frequência de acidentes, propondo diversos tipos de modelos tais como: modelos lineares generalizados, aditivos generalizados, modelos multinível, dentre outros.

De acordo com Cunto *et al* (2011), um dos esforços pioneiros para a modelagem de acidentes de acidentes no Brasil foi a tese de doutorado de Cardoso (2006), na qual se modelaram os acidentes de trânsito em vias artérias urbanas na cidade de Porto Alegre-RS. Mânica (2007) em sua dissertação de mestrado modelou a distribuição de acidentes nas rodovias também no Rio Grande do Sul com a aplicação de modelos de regressão. Cunto *et al* (2011) modelaram a distribuição de acidentes de trânsito em interseções semaforizadas na cidade de Fortaleza. Boffo (2011) fez uma revisão bibliográfica dos principais modelos de previsão de acidentes. Holz (2011) aplicou a modelagem da distribuição de acidentes de motocicleta na cidade de Porto Alegre com emprego de regressão linear multivariada. Rocha et al (2012a) fizeram uma modelagem estatística

da distribuição espacial dos acidentes na zona sul do Rio de Janeiro. Barbosa et al (2014) aplicaram modelos de previsão de acidentes nas interseções das vias das cidades de Fortaleza, Belo Horizonte e Brasília.

2.1 Variáveis Explicativas

Os locais onde se tem uma maior quantidade de pedestres e de volume de tráfego, bem como maiores valores de velocidade nas vias costumam apresentar maior exposição ao risco de acidentes de trânsito. Miranda-Moreno et al (2011) mostraram que uma redução de 30% no volume de veículos nas interseções de sua área de estudo, diminuía em 50% o risco médio dos pedestres na região e em 35% os acidentes com danos aos pedestres.

O volume de veículos e de pedestres é tradicionalmente medido por meio de contadores instalados em locais estratégicos das vias ou por meio de Instrumentos Autônomos de Medição de Velocidade, conhecidos como radares, pardais ou lombadas eletrônicas. No caso do Brasil, tais medições estão limitadas a poucas vias pelo seu alto custo, sendo muitas vezes realizadas somente quando se planeja algum evento.

Verificou-se que a relação entre o número de acidentes ocorridos nas interseções e o total de acidentes é considerável, sendo que cerca de 43% dos acidentes nos Estados Unidos e 40% dos acidentes da Noruega ocorrem nas interseções (LORD et al, 2005 *apud* TAYANDRIFAAT, 2007). Segundo Pulugurtha et al (2011), cerca de 21% das fatalidades em acidentes de trânsito nos Estados Unidos ocorreram em interseções, o que comprova a importância de se trabalhar com as mesmas. As interseções semaforizadas possuem um destaque especial por serem o encontro de vias com certo fluxo de veículos ou de pedestres que justificam a colocação dos semáforos.

Como variáveis explicativas, diversos autores vêm utilizando aquelas associadas à geometria das vias, tais como largura e comprimento, bem como à conectividade como, por exemplo, densidade de interseções (ZHANG et al, 2013; UKKUSURI et al, 2012; HA et al, 2011). Rocha e Nassi (2012) encontraram na mesma região de estudo uma maior quantidade de acidentes nas zonas de tráfego com maior densidade de vias e maior densidade de vias do tipo arterial primária, sendo esta hierarquia um

forte indicador do fluxo de veículos.

Diversos estudos empregam variáveis socioeconômicas na modelagem de acidentes como densidade demográfica, idade, renda e número de empregos (NOLANDANDQUDDUS, 2004; PULUGURTHA & SAMBHARA, 2011; HA et al, 2011; UKKUSURI et al, 2012, KIM et al, 2010). Alguns deles vêm associando os acidentes a índices de desenvolvimento de regiões da cidade (NOLAND & QUDDUS, 2004; LICAJ, 2011; COTRILL & THAKURIAH, 2010). A acessibilidade aos transportes públicos, tais como número de pontos de ônibus, de estações de metrô ou de trem, também vêm sendo utilizados como indicadores do fluxo de pedestres em uma região, o qual teria impacto no número de acidentes (THAKURIAH et al, 2012; COTRILL & THAKURIAH, 2010; UKKUSUKI, 2012, KIM et al, 2010). O uso do solo, seja para fins comerciais, industriais ou residenciais (PULUGURTHA et al, 2013; UKKUSUKI, 2012), bem como número de escolas, de estabelecimentos que vendam bebidas alcoólicas, dentre outros, também vem sendo apontados em alguns estudos como relacionados à quantidade de acidentes em uma região (CLIFTON & KREAMER-FULTS, 2007; UKKUSUKI, 2012). Diversos estudos envolvendo a presença de escolas, costumam associar o risco de acidentes com a capacidade de se caminhar em torno das mesmas, o que está relacionado com a qualidade das calçadas, a geometria das vias e até mesmo com o fluxo de veículos Giles-Corti et al (2011).

3. ÁREA DE TRABALHO E DADOS EMPREGADOS

A área de trabalho utilizada foi a zona sul do Rio de Janeiro, região esta com grande importância econômica e turística para a cidade, possuindo cerca de 10% da população da cidade e 13% dos acidentes da mesma. Apresenta grande quantidade de áreas de lazer, tais como cinemas, teatros, museus e *shopping centers*, destacando-se as áreas de lazer ao ar livre, tais como as praias situadas às margens do oceano Atlântico e da baía da Guanabara, bem como a região da Lagoa Rodrigo de Freitas e da Floresta da Tijuca. Neste estudo dividiu-se a região em 32 (trinta e duas) zonas de tráfego (Figura 1), definidas no Plano Diretor de Transporte Urbano- PDTU em 2003/2005.

Segundo Martínez et al (2009), pouca referência pode ser encontrada sobre estes critérios adotados para a criação de zonas de tráfego, mas sabe-se que são o nível de agregação espacial comumente empregado nos estudos de transporte, onde se busca obter áreas mutuamente excludentes, baseando-se na similaridade com respeito a alguns critérios dentro de cada uma das regiões.



Fig. 1 - Zonas de tráfego empregadas na pesquisa.

Os dados de acidentes foram obtidos para os anos de 2008 a 2011 pela Polícia Militar do Rio de Janeiro no local do acidente com o emprego de rastreadores de satélites *Global Positioning System* (GPS) e fornecidos na forma de coordenadas em uma planilha eletrônica. Outra parte dos dados de acidentes ocorridos em vias denominadas de vias especiais foi fornecida pela Companhia de Engenharia de Tráfego da cidade do Rio de Janeiro (CET-Rio) em formato de planilha eletrônica e agregado por ano e por via. Vias especiais são vias de grande importância por concentrarem grande fluxo de veículos e ligarem importantes regiões da cidade.

A verificação da característica das interseções das vias e o uso do solo foram obtidos por meio da visualização das imagens panorâmicas (Figura 2) disponibilizadas pelo aplicativo *Google Street View*.



Fig. 2 - Exemplo de interseção.
Fonte: *Google Street View*

A Figura 3 mostra uma representação temática das condições de trânsito às 18h próximo ao acesso ao túnel Rebouças, importante via de acesso e escoamento da zona Sul, empregando a opção de trânsito típico do *Google Maps*. Considerando que as vias com trânsito rápido estão representadas com a cor verde e as de trânsito lento com a cor vermelha, é possível verificar que as principais vias encontram-se na cor vermelha, o que mostra as difíceis condições de trânsito na zona sul, principalmente nos horários de pico.

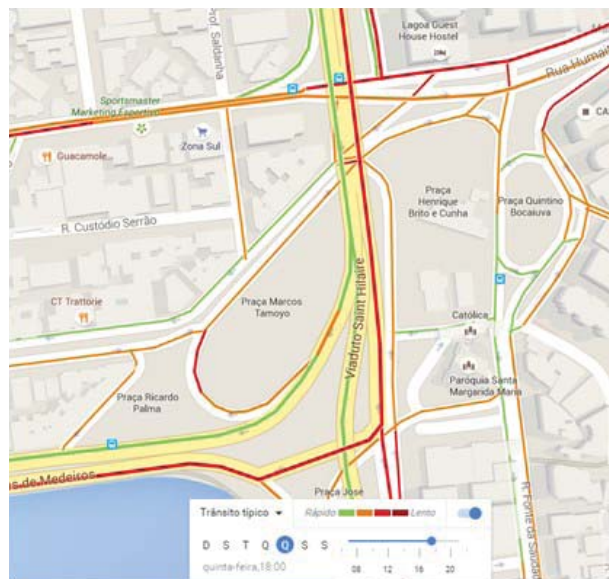


Fig. 3 - Mapa com a representação das condições em trânsito na região de acesso ao túnel Rebouças. Fonte: *Google Maps*.

As informações do número de escolas foram obtidas a partir da Secretaria de Educação da Cidade do Rio de Janeiro e do Estado do Rio de Janeiro. Foram consideradas no estudo somente as creches e unidades públicas de ensino fundamental e médio. Não foram consideradas, portanto, as escolas particulares e as públicas de ensino superior.

Quanto ao uso do solo, embora a bibliografia costume mensurar o mesmo em termos de área, neste estudo classificou-se o uso do solo como sendo aquele predominante em cada trecho de via, obedecendo à seguinte escala de prioridade: lazer, comercial, misto, residencial e outros. Como exemplo, quando se tem um trecho de via de casas (uso residencial) em frente a uma praça, este trecho de via passa a ser classificado como sendo lazer. A descrição de cada um deles

será dada a seguir:

-Lazer: caracterizado por locais públicos de lazer públicos a céu aberto tais como praças, praias, parques, ao redor de espelhos d'água, etc. Aqui não estão considerados locais fechados ou privados, tais como *shopping-centers*, cinemas, clubes, etc.

- Residencial: área caracterizada predominantemente pela presença de casas e/ou prédios residenciais.

- Comercial: área com presença predominante de comércio.

- Misto: áreas que apresentam construções tanto residenciais, como comerciais ou que apresentem em um mesmo prédio ambos os usos.

- Outros: caracterizado por locais que não se encaixam em nenhuma das descrições acima, como por exemplo, hospitais, túneis, entroncamentos, áreas militares entre outros.

4. COLETA E PREPARAÇÃO DOS DADOS

Antes da execução da modelagem propriamente dita, fez-se a coleta e o processamento dos dados nos aplicativos de SIG. Inicialmente visualizaram-se as interseções e o uso do solo ao longo das vias pelo aplicativo *Google Street View* e associaram-se os atributos das mesmas às feições pontuais, no caso das interseções e lineares no caso dos trechos de vias, respectivamente, com o emprego do aplicativo *Quantum GIS 1.8.0*.

As informações da localização das escolas, obtida por meio de planilhas eletrônicas, foram

fornecidas por meio do endereço completo das mesmas. Tendo em vista que a base de ruas possui os nomes das ruas, a locação das mesmas foi feita diretamente no *Quantum GIS*.

A agregação dos dados coletados no *Quantum GIS* nas zonas de tráfego foi feita no aplicativo *ArcGIS 9.3*. Entende-se por agregação a obtenção nas zonas de tráfego do somatório do número de pontos representativos dos locais das interseções e escolas, bem como extensão das vias. Exportou-se, por fim, a tabela de atributos das zonas de tráfego, as quais contêm as informações do número de acidentes e das variáveis explicativas por zona de tráfego, para o programa estatístico *R*, onde se processou a modelagem.

A descrição das variáveis utilizadas encontra-se na Tabela 1 e o sumário estatístico encontra-se na Tabela 2.

5. ANÁLISE DA CORRELAÇÃO DAS VARIÁVEIS

Antes da modelagem, calcularam-se as correlações de Pearson das variáveis dependentes envolvidas na modelagem com o emprego do programa *SPSS 15*, utilizando-se o método *Stepwise*. Selecionaram-se as variáveis explicativas com maior correlação com a variável resposta e que, ao mesmo tempo, apresentassem baixa correlação entre si. Nesta pesquisa foram consideradas correlacionadas aquelas variáveis cujo coeficiente de correlação de Pearson fosse menor que -0,3 ou maior que +0,3.

Tabela 1: Descrição das variáveis

Categorias	Variáveis	Nomes das variáveis	Descrição
Acidentes	ZONEACI	Número de acidentes	Número de acidentes por zona de tráfego
Uso do Solo	ESCOLAS	Número de escolas	Número de escolas públicas por zona de tráfego, incluindo escolas estaduais municipais e creches
	COMP_LAZER	Extensão das vias de lazer	Extensão das vias ao longo de áreas de lazer a céu aberto
	COMP_MISTO	Extensão das vias mistas	Extensão das vias predominantemente com uso misto
	COMP_RESID	Extensão das vias residenciais	Extensão das vias predominantemente residenciais, com pouco ou nenhum comércio
	COMP_COMER	Extensão das vias comerciais	Extensão das vias predominantemente comerciais, com pouca ou nenhuma residência
	COMP_OUTRO	Extensão de outros tipos de vias	Extensão das vias que incluem túneis, entroncamentos, entre outros
Característica da Interseção	NUM_SEMAFO	Número de Semáforos	Número de semáforos por zonas de tráfego

Tabela 2: Estatística descritiva das variáveis utilizadas na modelagem

Variáveis	Média	Dev. Pad.	Min	Max	N
ZONEACI	1222,84	622,17	368	2676	1222,84
ESCOLAS	2,18	1,84	0	8	70
COMP_LAZER	1636,39	1776,02	0	6778	52364,70
COMP_MISTO	1586,69	1683,35	0	7661,06	50774,17
COMP_RESID	6612,12	3411,92	1475,85	15571,3	211588,13
COMP_COMER	107,64	557,96	0	3205,62	3444,52
COMP_OUTRO	825,53	1155,76	0	5051,43	26417,02
NUM_SEMAFO	14,96	10,63	0	40	479

6. MODELAGEM ESTATÍSTICA

Os modelos estatísticos empregados foram o modelo de regressão linear múltipla e os modelos lineares generalizados, do inglês *Generalized Linear Models* (GLM) com distribuição normal e binomial negativa.

6.1 Regressão Múltipla

A análise de regressão é um método estatístico muito empregado quando se deseja prever o valor de uma variável dependente (Y) em função de outras variáveis denominadas explicativas (X). Os principais objetivos de um modelo de regressão são, portanto, identificar as variáveis explicativas mais significativas assim como o vetor dos coeficientes o qual será empregado, juntamente com as variáveis explicativas, na previsão dos valores Y em um segundo momento.

A equação da regressão múltipla pode ser representada por:

$$Y = \beta X + \varepsilon \quad (1)$$

onde β é o vetor de coeficientes e ε é o vetor de erros aleatórios. Esse modelo deve satisfazer aos pressupostos de que os erros são normalmente distribuídos com média zero e variância constante e não serem correlacionados. A matriz X de variáveis explicativas deve ser formada por um conjunto fixo de valores, ou seja, não deve ser estocástica e não pode haver alta correlação ou multicolinearidade entre essas variáveis.

Para atender o pressuposto da normalidade dos resíduos, é usual nos modelos de regressão aplicar uma transformação na variável dependente. Uma das mais conhecidas é a transformação de Box e Cox (1964) a qual

emprega a seguinte transformação na variável resposta z

$$z = \begin{cases} \frac{y^\lambda - 1}{\lambda} & \text{se } \lambda \neq 0 \\ \log(y) & \text{se } \lambda = 0 \end{cases} \quad (2)$$

onde λ é uma constante desconhecida. A ideia dessa transformação é encontrar um valor de λ de tal forma que a nova variável se torne aproximadamente normal e, além disso, consiga produzir constância na variância e a linearidade entre a esperança da nova variável z e as covariáveis selecionadas para compor o modelo. Embora tal transformação seja muito aplicada nos modelos de regressão múltipla, pode ser utilizada também nos modelos lineares generalizados.

6.2 Modelos Lineares Generalizados

Os modelos lineares generalizados foram propostos por Nelder e Weddeburn em 1972, como sendo uma extensão dos modelos lineares clássicos onde as suposições de normalidade, homogeneidade de variâncias e relacionamento linear entre os efeitos das covariáveis e a média eram relaxados, produzindo uma abordagem unificada para análise de uma ampla classe de dados contínuos e discretos.

Segundo Cordeiro e Demétrio (2008), um modelo linear generalizado é definido por uma distribuição de probabilidade para a variável resposta pertencente à família exponencial, formada de uma componente aleatória, a qual especifica a distribuição de probabilidade da variável resposta, uma componente sistemática, a qual especifica uma função linear das variáveis explicativas e uma função de ligação que

relaciona uma combinação linear das variáveis explicativas com o valor médio da variável resposta.

A componente aleatória de um GLM considera que se dispõe de um vetor de observações $y = (y_1, \dots, y_n)^T$ como realização das variáveis aleatórias $Y = (Y_1, \dots, Y_n)^T$ (independentes ou pelo menos não correlacionadas, cada uma com distribuição pertencente à família exponencial que tem as seguintes propriedades:

a) A distribuição de cada Y_i é da forma canônica e depende de um único parâmetro, digamos θ_i , onde os θ_i 's não tem que ser os mesmos. Assim,

$$f(y_i; \theta_i) = \exp[y_i b(\theta_i) + c(\theta_i) + d(y_i)] \quad (3)$$

onde $b(\cdot)$, $c(\cdot)$ e $d(\cdot)$ são funções conhecidas.

b) A distribuição de todos os Y_i 's é da mesma forma, ou seja, a função densidade conjunta de Y_1, Y_2, \dots, Y_n pode ser escrita da seguinte forma:

$$f(y_1, \dots, y_n; \theta_1, \dots, \theta_n) = \exp \left[\sum_{i=1}^n y_i b_i(\theta_i) + \sum_{i=1}^n c_i(\theta_i) + \sum_{i=1}^n d_i(y_i) \right] \quad (4)$$

Distribuições conhecidas como a binomial, Poisson, gaussiana, gaussiana inversa e Gamma, só para citar alguns exemplos, são membros da família exponencial. Além disso, para a escolha de uma distribuição adequada para o erro, deve-se examinar o tipo de dado, especialmente quanto aos aspectos de assimetria, natureza contínua ou descontínua e intervalo de variação. Essa componente é representada pelo erro aleatório ε no modelo linear clássico. Os modelos com distribuição de Poisson e Binomial Negativa vêm sendo utilizados na modelagem de acidentes por serem adequados aos dados de contagem. No entanto somente o modelo com distribuição Binomial Negativa contempla o efeito da superdispersão, comumente encontrado nos dados de acidentes.

A componente sistemática também chamada de preditor linear é definida como uma função linear dos parâmetros desconhecidos $\beta = (\beta_1, \dots, \beta_k)^T$ representado no modelo linear por βX , onde X é a matriz de variáveis explicativas do modelo, de dimensões $n \times k$ do modelo. A escolha do preditor linear adequado leva em consideração técnicas

de seleção de covariáveis.

A função de ligação faz a ligação entre a média da variável resposta e a estrutura linear do modelo. Assim, a escolha dessa função deve ser compatível com a distribuição proposta para o erro, de forma a facilitar a interpretação do modelo.

A estimação do vetor de parâmetros desconhecidos é feita utilizando-se o método da máxima verossimilhança, quando o sistema de equações é linear.

A qualidade do ajuste de um MGL é medida por meio da função desvio (*deviance*) dada por:

$$S_p = 2(\hat{l}_n - \hat{l}_p) \quad (5)$$

onde S_p é o desvio do modelo e \hat{l}_n e \hat{l}_p são respectivamente, os máximos da log-verossimilhança para o modelo saturado (número de observações n igual ao número de parâmetros p dos coeficientes β 's) e para a investigação ($p < n$).

O desvio é uma medida de distância entre os valores ajustados e observados e segundo Cordeiro (1986) embora pouco seja conhecido sobre a distribuição do desvio, na prática compara-se S_p com o valor crítico $\chi_{n-p;\alpha}^2$ da distribuição qui-quadrado a um nível de significância α com $n-p$ graus de liberdade.

6.3 Comparação entre os modelos GLM

A comparação entre os modelos pode ser feita, entre outros métodos, por meio do critério de informação de Akaike (1974), o qual propôs um método para comparação entre os modelos o qual se baseia na estimação da máxima verossimilhança, sendo representada por:

$$AIC = -2L + 2q \quad (6)$$

onde L representa o log da verossimilhança, q é o número de parâmetros do modelo. Ao se comparar dois modelos, o melhor modelo será aquele que tiver o menor valor do AIC.

De modo a comparar os resultados da regressão linear múltipla com o GLM com distribuição binomial negativa, pode-se utilizar um GLM com distribuição gaussiana que é semelhante ao modelo de regressão múltipla.

7. RESULTADOS

Observando-se a correlação o coeficiente de Pearson na Tabela 3, é possível verificar que as variáveis com maior correlação com o número de acidentes são o número de interseções semaforizadas, o comprimento de vias com o uso do solo misto e o uso destinado ao lazer. No entanto, o comprimento de vias com o uso do solo misto apresenta correlação com o número de vias semaforizadas acima de 0,6, o que fez com que considerasse somente o uso do solo destinado ao lazer.

A correlação do número de interseções semaforizadas com o número de acidentes faz sentido na medida em que se costuma colocar semáforos nas interseções com maior fluxo de veículos.

Quanto ao uso do solo, o coeficiente de correlação de Pearson mostrou que na zona sul as vias com maior número de acidentes são aquelas de uso misto e lazer, o que se verifica na prática devido ao maior fluxo de veículos e de pessoas, que na sua maioria ocorrerem nas vias que apresentam

estabelecimentos comerciais caracterizados predominantemente por prédios com comércio no térreo e residência em cima.

No caso das vias que se encontram ao longo do mar ou da Lagoa Rodrigo de Freitas, a grande quantidade de acidentes pode ser explicada não somente pelo grande número de pessoas que utilizam tais áreas de lazer, mas também por serem vias largas e com menor quantidade de interseções, o que faz com que a velocidade dos veículos também aumente. Estas vias também apresentam grande fluxo de veículos sendo, pela sua importância, classificadas em grande parte como estruturais ou arteriais.

Verificou-se, por meio dos testes de normalidade de Anderson-Darling e Shapiro Wilks, que a variável resposta não apresentava uma distribuição normal. Para tal, aplicou-se a transformação de Box-Cox sobre a esta variável e comprovou-se que os pressupostos dos modelos de regressão múltipla passaram a ser atendidos. A seguir são apresentados na Tabela 4 o resumo dos modelos estatísticos ajustados e nas Figuras 2 a 6, a análise gráfica dos resíduos.

Tabela 3: Resultados do coeficiente de Pearson

	ZONE ACI	NUM_ SEMAFO	COMP_ LAZER	COMP_ RESID	COMP_ MISTO	COMP_ COMER	COMP_ OUTRO	ESCO LAS
ZONEACI	1,000	0,779	0,422	0,256	0,621	0,148	-0,016	0,107
NUM_SEMAFO	0,779	1,000	0,162	0,245	0,637	-0,013	-0,029	0,141
COMP_LAZER	0,422	0,162	1,000	-0,247	-0,116	-0,011	-0,164	-0,215
COMP_RESID	0,256	0,245	-0,247	1,000	0,297	-0,164	0,413	0,458
COMP_MISTO	0,621	0,637	-0,116	0,297	1,000	0,032	-0,084	0,261
COMP_COMER	0,148	-0,013	-0,011	-0,164	0,032	1,000	0,086	-0,095
COMP_OUTRO	-0,016	-0,029	-0,164	0,413	-0,084	0,086	1,000	0,326
ESCOLAS	0,107	0,141	-0,215	0,458	0,261	-0,095	0,326	1,000

Tabela 4: Variáveis mais significativas e resultados dos modelos

Categoria de variáveis	Variáveis	Regressão linear ¹		GLM Gaussiana		GLM Binomial Negativa	
		Coef	e.p	Coef	e.p	Coef	e.p
Interseções	SEMAFO	0,03445***	0,0000001	0,03445***	0,0000001	0,03672***	0,00515
Uso do solo	LAZER	0,000079*	0,0117	0,000079*	0,0117	0,000084**	0,000031
R ²		0,67	-	-	-	-	-
Função desvio				2,476		32,511	
AIC		-		16,916		473,07	-

(e.p) erro padrão.

*Nível de significância de 0,05.

**Nível de significância de 0,01.

***Nível de significância de 0,001.

¹ Número de acidentes na zona de tráfego obtido a partir dos coeficientes obtidos da modelagem - $ACI = 0,03445 \times SEMAFO + 0,000079 \times LAZER$

A análise gráfica dos resíduos, embora seja uma técnica informal, é uma das formas mais usuais e simples de verificar se as suposições dos modelos estatísticos testados foram satisfeitas. Dessa forma, serão apresentados a seguir os resultados da referida análise para o modelo considerado mais adequado nesse estudo – modelos de regressão.

Assim, analisando o gráfico da probabilidade normal na Figura 4, pode-se verificar que os valores dos resíduos da regressão estão próximos a uma distribuição normal.

Na Figura 5 observa-se que a hipótese de homocedasticidade, ou seja, variância constante é obedecida. Ainda, verifica-se por meio da Figura 6 que não há evidências de tendência nos resíduos e portanto, pode-se dizer que o pressuposto de independência dos resíduos é atendida.

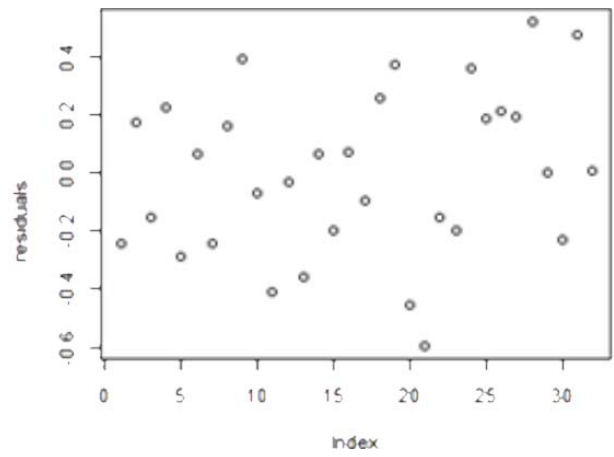


Fig. 6 - Gráfico dos Resíduos com a Ordem de Coleta dos Dados.

Já as Figuras 7 e 8 mostram a inexistência de outliers ou valores aberrantes na variável explicada Y e nas variáveis explicativas X, respectivamente.

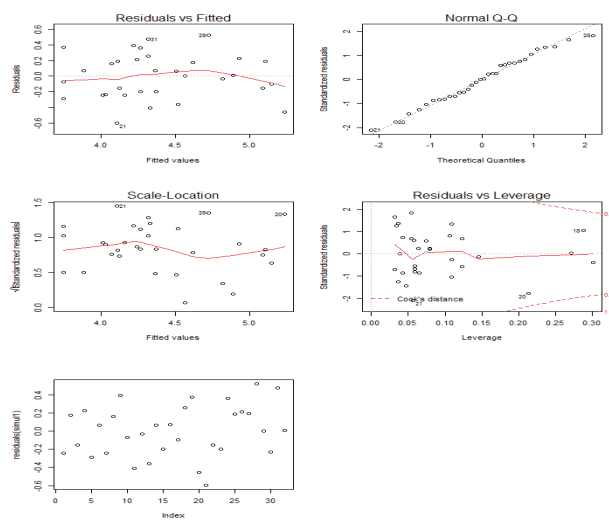


Fig. 4 - Gráfico da Probabilidade Normal (Q-Q Plot).

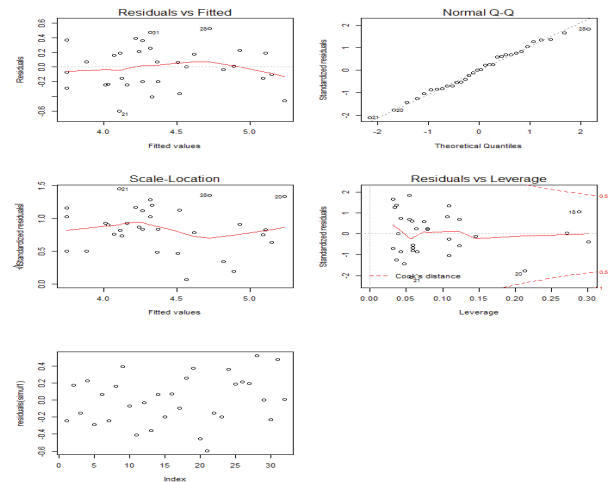


Fig. 7 - Gráfico dos Resíduos Studentizados com Valores Ajustados.

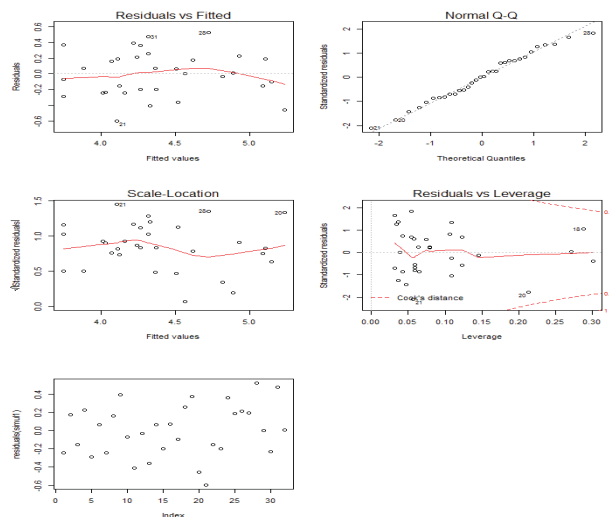


Fig. 5 - Gráfico dos Resíduos com Valores Ajustados.

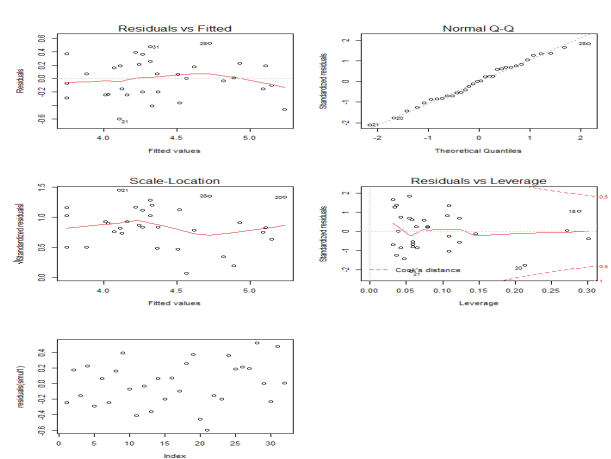


Fig. 8 - Gráfico Resíduos Studentizados com Lavarage.

Observando-se a Tabela 4, verifica-se que o valor do R^2 da regressão linear múltipla ficou em 0,67. Quando da utilização dos GLM, aplicou-se a mesma transformação de Box-Cox na variável dependente no caso da distribuição gaussiana para que tivesse um resultado igual ao da regressão linear múltipla. Comparando-se os valores do critério de informação de Akaike e o valor da função desvio dos modelos lineares generalizados com distribuição gaussiana e binomial negativa, é possível verificar que a gaussiana apresentou menores valores do AIC e da função desvio, o que torna o mesmo melhor que aquele com a distribuição binomial negativa. A dependência espacial dos resíduos foi determinada a partir do índice de Moran, o qual teve o valor de -0,16, sendo considerado um valor pequeno e portanto não descartou a hipótese de independência dos resíduos.

CONSIDERAÇÕES FINAIS

Buscou-se no estudo empregar variáveis explicativas comumente utilizadas na bibliografia. No entanto, a forma como o uso do solo foi mensurado em unidade de comprimento foi diferente do encontrado na bibliografia que utiliza a unidade de área.

O modelo de regressão múltipla após aplicação da transformação Box-Cox apresentou resultado superior ao do GLM com distribuição binomial negativa, o que mostra a possibilidade de se utilizar regressões lineares múltiplas devidamente transformadas na modelagem de acidentes de trânsito.

Tais resultados não são conclusivos tendo em vista a pequena quantidade de variáveis explicativas e o reduzido número de áreas de agregação. No entanto podem servir como indicadores de variáveis pertinentes na medida em que elas fornecem uma contribuição considerável ao modelo de previsão de acidentes.

REFERÊNCIAS BIBLIOGRÁFICAS

AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**. Texas, v. 19, n. 6, p. 716-723, dec. 1974.

ANASTASOPOULOS, P. C. & MANNERING, F. L. An empirical assessment of fixed and random parameter logit models using crash and non crash specific injury data. **Accident Analysis and Prevention**. V. 43, issue 3, p. 1140-1147, may.

2011.

BOFFO, G. H. **Formatos e técnicas de modelos de previsão de acidentes de trânsito**. Dissertação. (Mestrado em Engenharia de Produção), Pós-Graduação em Engenharia de Produção, Escola de Engenharia da UFRGS, Universidade Federal do Rio Grande do Sul, 2011.

BOX, G. E. P. & COX, D. R. An analysis of transformations. **Journal of the Royal Statistical Society**. V. 26, n. 2, p.211-252, 1964

CARDOSO, G. **Modelos para previsão de acidentes de trânsito em vias arteriais urbanas**, 2006. 289f. Tese de doutorado. (Doutorado em Engenharia de Produção), Pós-Graduação em Engenharia de Produção, Escola de Engenharia da UFRGS, Universidade Federal do Rio Grande do Sul, 2006.

CLIFTON, K. J. & KREAMER-FULTS, K. An examination of the environmental attributes associated with pedestrian_vehicular crashes near public schools. **Accident Analysis and Prevention**. V. 39, issue 4, p. 708-715, 2007.

CORDEIRO, G. M. Modelos Lineares Generalizados. In: VII SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA. 1986. **Anais**. Campinas, São Paulo, Brasil. p. 1-23.

CORDEIRO, G. M., Demétrio C.G.B. **Modelos Lineares Generalizados**. Minicurso para o 12º SEAGRO e a 52ª Reunião Anual da RBRAS, Universidade Federal de Santa Maria. Santa Maria, Brasil, 2008. p.1-165.

COTRILL, C. D. & THAKURIAH, P. V. Evaluating pedestrian crashes in areas with high low-income or minority populations. **Accident Analysis and Prevention**. V 42, issue 6, p. 1718-1728, nov. 2010.

CUNTO, F. J. C.; CASTRO NETO, M. M. & BARREIRA, D. S. Modelos de previsão de acidentes de trânsito em interseções semaforizadas de Fortaleza. XXV CONGRESSO DE PESQUISA E ENSINO EM TRANSPORTES. Fortaleza, Brasil, 2011.

GILES-CORTI, B.; WOOD, G.; PIKORA, T.; LEARNIHAN, V.; BULSARA, M.; NIEL, K.V.; TIMPERIO A.; MCCORMACK, G. & VILLANUEVA, K. School site and the potential to walk to school: The impact of street

- connectivity and traffic exposure in school neighborhoods. **Health & Place**. V. 17, issue 2, p. 545-550, mar. 2011.
- HA, H. H. & THILL, J. C. Analysis of traffic hazard intensity: a spatial epidemiology case study of urban pedestrians. **Computers, Environment and Urban Systems**. V. 35, issue 3, p. 230-240, may. 2011.
- HOLZ, R. F.; KORZENOWSKI, A.; NODARI, C. T.; TEN CARTEN, C. S. & LINDAU, L.A., Modelagem de acidentes envolvendo motociclistas em Porto Alegre, In: *XXV Congresso de Pesquisa e Ensino em Transportes*. **Anais...** Belo Horizonte-MG, ANPET(Ed.), 2011.
- KIM, K.; PANT, P. & YAMASHITA, E. Accidents and accessibility: measuring the influences of demographic and land use variables in Honolulu, Hawaii. **Journal of the Transportation Research Board**. n. 2147 p. 9-17. 2010.
- LORD, D.; SCHALKWYK, I.; STAPLIN, L. & CHRYSLER, S. Reducing older driver injuries at intersection using more accommodating design practices. College Station: **Texas Transportation Institute**. 2005.
- LICAJ, I.; HADDACK, M.; POCHET, P. & CHIRON, M. Contextual deprivation, daily travel and road traffic injuries among the young in the Rhône Département (France). **Accident Analysis and Prevention**. V. 43 (5), p. 1617-1623, set. 2011.
- MÂNICA, A. G. **Modelo de previsão de acidentes rodoviários envolvendo motocicletas**, 2007. 177f. Tese de doutorado. (Doutorado em Engenharia de Produção), Pós-Graduação em Engenharia de Produção, Escola de Engenharia da UFRGS, Universidade Federal do Rio Grande do Sul, 2007.
- MIRANDA-MORENO, L.F., P. MORENCY, EL-GENEIDY A. M. The link between built environment, pedestrian activity and pedestrian - vehicle collision occurrence at signalized intersections. **Accident Analysis and Prevention**. V. 43 (5), p.1624-1634, set.2011.
- NELDER, J. A. & WEDDERBURN, R. W. M. Generalized linear models. **Journal of the Royal Statistical Society**. A 135, p. 370, 1972
- NOLAND, R. B. & QUDDUS, M.A. A spatially disaggregate analysis of road casualties in England. **Accident Analysis and Prevention**. V 36, issue 6, p. 973-984, nov. 2004.
- PULUGURTHA, S. S. & SAMBHARA, V. R. Pedestrian crash estimation models for signalized intersections. **Accident Analysis and Prevention**. V. 43, issue 1, p. 439-446, jan. 2011.
- PULUGURTHA, S.S.; DUDDU, V. & KOTAGIRI, Y. Traffic analysis zone level crash estimation models based on land use characteristics. **Accident Analysis and Prevention**. V. 50, 678-687, jan. 2013.
- ROCHA, M. M. & NASSI, C, D. Modelagem estatística dos acidentes de trânsito na cidade do Rio de Janeiro com emprego de sistema de informações geográficas. Anais do 17º CONGRESSO PANAMERICANO DE TRANSPORTES (PANAM), Santiago, Chile, 2012.
- SAVOLAINEN, P. T.; MANNERING, F. L.; LORD, D. & QUDDUS, M. A. The statistical analysis of highway crash-injury severities: a review and assessment of methodological alternatives. **Accident Analysis and Prevention**. V. 43, issue 5, 1666-1676, set.2011.
- THAKURIAH, P.V.; METAXATOS, P.; LIN, J. & JENSEN, E. An examination of factors affecting propensities to use bicycle and pedestrian facilities in suburban locations. **Transportation Research Part D, Transport and Environment**. V. 17, issue 4, p. 341-348, jun.2012.
- UKKUSURI, S.; MIRANDA-MORENO, L. F.; RAMADURAI G. & ISA-TAVAREZ, J. The role of built environment on pedestrian crash frequency. **Safety Science**. V. 50, issue 4, 1141-1151, abr.2012.
- WHO. Global Status Report on road safety: time for action, **World Health Organization**, Genebre. 2009.
- ZHANG, Y.; BIGHAM, J.; LI, Z.; RAGLAND, D. & CHEN, X. Associations between road network structure and pedestrian-bicyclist accidents, In: **Proceedings of the 92th ANNUAL MEETING OF THE TRANSPORTATION RESEARCH BOARD**, Washington, DC, 2013.