

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ENGENHARIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

Gabriela Holz Boffo

**FORMATOS E TÉCNICAS DE MODELOS DE
PREVISÃO DE ACIDENTES DE TRÂNSITO**

Porto Alegre

2011

Gabriela Holz Boffo

Formatos e Técnicas de Modelos de Previsão de Acidentes de Trânsito

Dissertação submetida ao Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal do Rio Grande do Sul como requisito parcial à obtenção do título de Mestre em Engenharia de Produção, modalidade Acadêmica Profissional, na área de concentração em Sistemas de Transportes.

Orientadora: Christine Tessele Nodari, Dra.

Porto Alegre

2011

Gabriela Holz Boffo

Formatos e Técnicas de Modelos de Previsão de Acidentes de Trânsito

Esta dissertação foi julgada adequada para a obtenção do título de Mestre em Engenharia de Produção na modalidade Profissional e aprovada em sua forma final pelo Orientador e pela Banca Examinadora designada pelo Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal do Rio Grande do Sul.

Profa. Christine Tessele Nodari, Dra.
Orientador PPGE/UFGRS

Profa. Carla S. ten Caten, Dra.
Coordenador PPGE/UFGRS

Banca Examinadora:

Professor Flávio José Craveiro Cunto, *Ph.D.* (DET/UFC)

Professora Helena Beatriz Bettella Cybis, *Ph.D.* (DEPROT/UFGRS)

Professor João Fortini Albano, Dr. (DEPROT/UFGRS)

Dedicatória

À minha família e principalmente à minha mãe
que tanto se esforçou e me incentivou para que
eu chegasse até aqui.

AGRADECIMENTOS

Esta Dissertação de mestrado contou com o apoio e participação de várias pessoas em momentos distintos. Assim, gostaria de citar as pessoas e entidades que de certa forma contribuíram para a conclusão da mesma. A todos meu sincero agradecimento.

A minha orientadora, Prof^a Christine Tessele Nodari, Dr.^a, pela paciência, atenção e confiança que depositou em mim.

Aos membros da banca, Prof. Flávio José Craveiro Cunto, Ph.D., Prof. Helena Beatriz Bettella Cybis, Ph.D. e Prof. João Fortini Albano, Dr., por terem aceito o convite para participar desta defesa e plas valiosas contribuições para a conclusão do trabalho.

Aos Professores do PPGEPI-LASTRAN pela transmissão dos conhecimentos e vivências profissionais durante o mestrado.

Aos amigos e colegas do PPGEPI-LASTRAN pelo convívio e aprendizado durante esses anos.

Aos amigos e colegas do Consórcio Univias que sempre me incentivaram e disponibilizaram os dados necessários para a realização da Dissertação. E, especialmente ao Felipe Sousa, Carlo Framarim e Michele Marcondes que compreenderam os meus momentos de ausência.

Ao companheiro e amigo Gregório Bastian que esteve do meu lado durante todo esse período e sempre que possível me ajudou quando precisei.

À minha família pelo apoio, confiança e dedicação durante todos esses anos.

Por fim, os agradecimentos aos demais colegas e amigos que de alguma forma colaboraram para a conclusão dessa etapa.

BOFFO, G. H. *Formatos e técnicas de modelos de previsão de acidentes*. 2011. Dissertação (Mestrado em Engenharia) – Universidade Federal do Rio Grande do Sul, Brasil.

RESUMO

A ampliação acelerada da demanda por transporte, mais especificamente pelo transporte rodoviário, tem provocado um aumento expressivo no número de acidentes de trânsito nesse ambiente. Conseqüentemente, a redução dos acidentes de trânsito tem sido um grande desafio para os pesquisadores e gestores da área rodoviária. Porém, os acidentes de trânsito são eventos complexos se considerados os diversos fatores que podem influenciá-los. Dentro desse contexto esta dissertação apresenta um estudo de modelos de previsão de acidentes, que podem ser utilizados para a avaliação do potencial de segurança em determinados locais, identificação e classificação de localidades perigosas ou com propensão a acidentes e avaliação da eficácia de medidas de melhoria da segurança. Nessa dissertação é apresentado um levantamento teórico e metodológico dos modelos de previsão de acidentes, identificando as principais variáveis adotadas bem como as técnicas utilizadas. Para cada modelo revisado foram verificadas as principais diferenças e limitações, e ainda, a análise das variáveis mais influentes presentes nesses modelos. Após, é feita uma comparação de duas abordagens distintas para estimar modelos de previsão de acidentes. A primeira consiste em estimar a ocorrência de acidentes em segmentos da via com as mudanças de características dos elementos de infraestrutura. O segundo relaciona a frequência de acidentes para um único elemento de infraestrutura da via, chamado na literatura internacional de entidade (ex: interseção, curva, tangente, etc.), com base apenas na variável relacionada ao volume de tráfego. O estudo baseado na comparação dessas duas abordagens para a previsão de acidentes revelou que a utilização do volume de tráfego como única variável independente apresenta resultados semelhantes ou até melhores que os modelos baseados em diversos elementos de infraestrutura da rodovia.

Palavras-chave: Segurança Viária, Modelos de Previsão de Acidentes de Trânsito, Modelos Lineares Generalizados.

ABSTRACT

The enlargement and the accelerated development of transportation systems, more specifically the land system, have caused the number of road accidents to increase significantly. Therefore, the reduction of road accidents has been a great challenge for researchers and managers in the field of land transportation. However, considering the various factors that may influence them, road accidents are complex events. In this context, this paper presents a study of accident prediction models that can be used to assess the safety potential in certain locations, identify and rank dangerous locations or areas prone to accidents and evaluate the effectiveness of safety improvement measures. Initially, a theoretical and methodological review of accident prediction models is presented, and both the main variables adopted and the methodologies employed are identified. The main differences between all models reviewed and their limitations are presented, and the most influential variables are analyzed. In a second moment, a comparison of two different accident prediction methods is performed. The first method consists in estimating the occurrence of accidents in road sections with changes in the characteristics of infrastructure elements. The second one relates the frequency of accidents based on a single infrastructure element (intersection, curve, tangent, etc.) based on traffic volume only. The study based on the comparison of these two methods found that the use of traffic volume as the only independent variable yields similar or even better results than the models based on various road infrastructure elements.

Key words: Road Safety, Accident Prediction Models, Generalized Linear Models.

LISTA DE FIGURAS

1. INTRODUÇÃO

Figura 1

Estrutura do método de trabalho.....16

2. PRIMEIRO ARTIGO

Método para Previsão de Acidentes: Uma Revisão da Literatura.....19

Figura 1

Evolução dos modelos de previsão de acidentes de trânsito.....28

3. SEGUNDO ARTIGO

Modelos de Previsão de Acidentes de Trânsito: Quanto mais Variáveis Melhor?.....37

Figura 1

Estrutura da metodologia.....46

LISTA DE TABELAS

3. SEGUNDO ARTIGO

Modelos de Previsão de Acidentes de Trânsito: Quanto mais Variáveis Melhor?.....37

Tabela 1	
Estatísticas Descritivas.....	48
Tabela 2	
Estimativa dos coeficientes para o Modelo 1.....	53
Tabela 3	
Estimativa dos coeficientes para o Modelo 2.....	53
Tabela 4	
Estimativa dos coeficientes para o Modelo 3 – modelo inicial.....	54
Tabela 5	
Estimativa dos coeficientes para o Modelo 3 – 1ª etapa.....	55
Tabela 6	
Estimativa dos coeficientes para o Modelo 3 – 2ª etapa.....	56
Tabela 7	
Estimativa dos coeficientes para o Modelo 3 – 3ª etapa.....	57
Tabela 8	
Estimativa dos coeficientes para o Modelo 3 – 4ª etapa.....	58
Tabela 9	
Estimativa dos coeficientes para o Modelo 3 – 5ª etapa.....	58
Tabela 10	
Estimativa dos coeficientes para o Modelo 3.....	59
Tabela 11	
Estimativa dos coeficientes para o Modelo 5.....	59

LISTA DE QUADROS

3. PRIMEIRO ARTIGO

Método para Previsão de Acidentes: Uma Revisão da Literatura.....19

Quadro 1

Descrição dos principais modelos analisados.....30

Quadro 2

Descrição das variáveis dos principais modelos analisados.....32

4. SEGUNDO ARTIGO

Modelos de Previsão de Acidentes de Trânsito: Quanto mais Variáveis Melhor?.....37

Quadro 1

Resumo dos modelos estimados.....50

Quadro 2

Modelos calibrados.....60

SUMÁRIO

1. INTRODUÇÃO	13
1.1. <i>Comentários Iniciais</i>	13
1.2. <i>Tema e Objetivos.....</i>	14
1.3. <i>Justificativa do Tema e Objetivos.....</i>	14
1.4. <i>Método</i>	15
1.5. <i>Delimitação do Trabalho.....</i>	16
1.6. <i>Estrutura do Trabalho.....</i>	17
2. PRIMEIRO ARTIGO	19
<i>Métodos para Previsão de Acidentes: uma revisão da literatura</i>	19
1. <i>INTRODUÇÃO.....</i>	20
2. <i>MODELOS DE PREVISÃO DE ACIDENTES DE TRÂNSITO</i>	21
2.1. <i>Evolução dos modelos de previsão de acidentes de trânsito.....</i>	22
2.2. <i>Comprimento dos Segmentos</i>	26
3. <i>SÍNTESE E DISCUSSÃO DOS ESTUDOS REVISADOS</i>	27
4. <i>CONCLUSÕES.....</i>	33
<i>REFERÊNCIAS</i>	34
3. SEGUNDO ARTIGO.....	37
<i>MODELOS DE PREVISÃO DE ACIDENTES DE TRÂNSITO: quanto mais variáveis melhor?</i>	37
1. <i>INTRODUÇÃO.....</i>	38
2. <i>MODELOS DE PREVISÃO DE ACIDENTES</i>	39
2.1. <i>Estudos envolvendo modelos de previsão de acidentes</i>	39
3. <i>MÉTODO</i>	45
3.1. <i>Rodovia de estudo</i>	46
3.2. <i>Descrição das variáveis.....</i>	47

3.3. Calibração dos modelos	49
3.4. Comparação dos modelos.....	51
4. DESENVOLVIMENTO E RESULTADOS DOS MODELOS.....	53
4.1. Desenvolvimento dos modelos.....	53
5. CONCLUSÕES.....	61
REFERÊNCIAS	62
4. CONCLUSÕES.....	64
REFERÊNCIAS	68

1. INTRODUÇÃO

1.1. Comentários Iniciais

A Organização Mundial de Saúde (WHO) apresentou um estudo com dados de 2007 mostrando que o Brasil tem o quinto maior número de mortes no trânsito de todo o mundo (WHO, 2009). De acordo com o mesmo estudo, em termos absolutos, a quantidade de óbitos no Brasil (35,1 mil) só é inferior ao de quatro países: Índia (105,7 mil), China (96,6 mil), Estados Unidos (42,6 mil) e Rússia (35,9 mil). Relativizando esses números com a população, o Brasil ocupa uma posição intermediária, com 18 óbitos para cada 100 mil habitantes. Sendo, nesse caso, superior aos Estados Unidos (13 óbitos para cada 100 mil habitantes) e inferior a Rússia (25 óbitos para cada 100 mil habitantes). Esse estudo indicou que os maiores índices se encontram no Leste do Mediterrâneo e nos países africanos. E as menores taxas estão na Holanda, Suécia e Reino Unido. A conclusão da pesquisa é que os acidentes rodoviários já são a décima maior causa de mortes no mundo, totalizando 1,2 milhão de óbitos por ano.

Frente ao cenário atual de mortes por acidente de trânsito, diversas pesquisas têm sido conduzidas no sentido de desenvolver o conhecimento sobre o tema segurança viária (GAO, 2003; Miranda e Braga, 2004; Nodari, 2003). De forma geral, é consenso que os acidentes de trânsito são resultados da combinação de diversos fatores casuais que contribuem para sua ocorrência. Dessa forma, pesquisadores do mundo inteiro têm se dedicado a estudar os fatores que são mais relevantes na ocorrência de acidentes de trânsito a fim de desenvolver

medidas que, bloqueando ou reduzindo a ação desses fatores, reduza também a ocorrência dos acidentes.

Entender como esses fatores contribuem para a ocorrência de acidentes de trânsito e medir o efeito de medidas corretivas concebidas para reduzir a influência dos mesmos é fundamental. Nesse contexto, os modelos de previsão de acidentes surgem como ferramentas estatísticas capazes de relacionar os fatores contribuintes de acidentes de trânsito à quantidade de acidentes de trânsito permitindo a análise do potencial de geração de acidentes de diferentes projetos/configurações viárias.

1.2. Tema e Objetivos

Esta dissertação trata do tema relacionado à ocorrência de acidentes rodoviários que está incluído na área de segurança viária. Fazem parte desse estudo os fatores que estão relacionados e contribuem para a ocorrência de acidentes de trânsito, bem como os modelos de previsão da influência de tais fatores na ocorrência dos acidentes.

O objetivo geral desse estudo é identificar as principais tendências no processo de estimativa de modelos de previsão de acidentes rodoviários. Entre os objetivos específicos relacionados ao estudo estão: (i) a identificação das principais técnicas de modelagem; (ii) a identificação das principais variáveis relacionadas à acidentes de trânsito e (iii) a comparação das diferentes abordagens adotadas atualmente na modelagem da ocorrência de acidentes de trânsito em ambiente rodoviários.

1.3. Justificativa do Tema e Objetivos

A redução do número de acidentes de trânsito é uma meta constante do governo e órgãos rodoviários. A ocorrência de acidentes de trânsito é influenciada por diversos fatores. Entender como esses fatores influenciam a ocorrência de acidentes de trânsito é um dos desafios da pesquisa em segurança viária.

Estudos visando investigar a relação entre variáveis de projeto geométrico, bem como de outras variáveis, e a ocorrência de acidentes de trânsito tem sido realizados

utilizando modelos matemáticos para prever a frequência de acidentes de trânsito. Tais modelos relacionam a ocorrência de acidentes de trânsito em várias localidades da rodovia, ao tráfego e às características geométricas de cada localidade. Por meio dos resultados dessa modelagem, é possível obter avanços na avaliação do potencial de segurança de locais da via, na identificação e classificação de localidades perigosas ou com propensão a acidentes, na avaliação da eficácia de medidas de melhoria da segurança e o planejamento da segurança.

É possível apontar como uma das principais justificativas para a modelagem da ocorrência de acidentes de trânsito, a possibilidade de se atuar preventivamente na promoção da segurança viária. Tais modelos podem servir de suporte para avaliação das condições de segurança de projetos viários ainda na sua fase de concepção. Por meio dos modelos calibrados, é possível estimar o número esperado de acidentes de diferentes alternativas de projeto, ainda antes da sua implantação. Essa informação permite subsidiar a tomada de decisão e privilegiar opção por projetos mais seguros.

1.4. Método

Esta pesquisa se caracteriza por ser de natureza aplicada, uma vez que objetiva gerar conhecimentos, de aplicação prática, dirigidos à solução de problemas específicos. É ainda caracterizada por ser predominantemente quantitativa e por adotar procedimento experimental, onde são selecionadas variáveis que seriam capazes de influenciar o objeto de estudo.

Este trabalho iniciou por meio de uma revisão teórica sobre a evolução dos modelos de previsão de acidentes de trânsito. Nessa revisão buscou-se identificar as principais técnicas de modelagem adotadas e as variáveis utilizadas nesse tipo de estudo. Esse levantamento subsidiou a etapa seguinte, de escolha do método estatístico a ser utilizado no estudo. Por fim, foi possível proceder a coleta de dados e a modelagem da ocorrência de acidentes de trânsito por meio das abordagens revisadas. De posse dos resultados da modelagem foram comparadas as abordagens adotadas no estudo. Resumidamente, o método adotado segue as etapas apresentadas na figura 1.

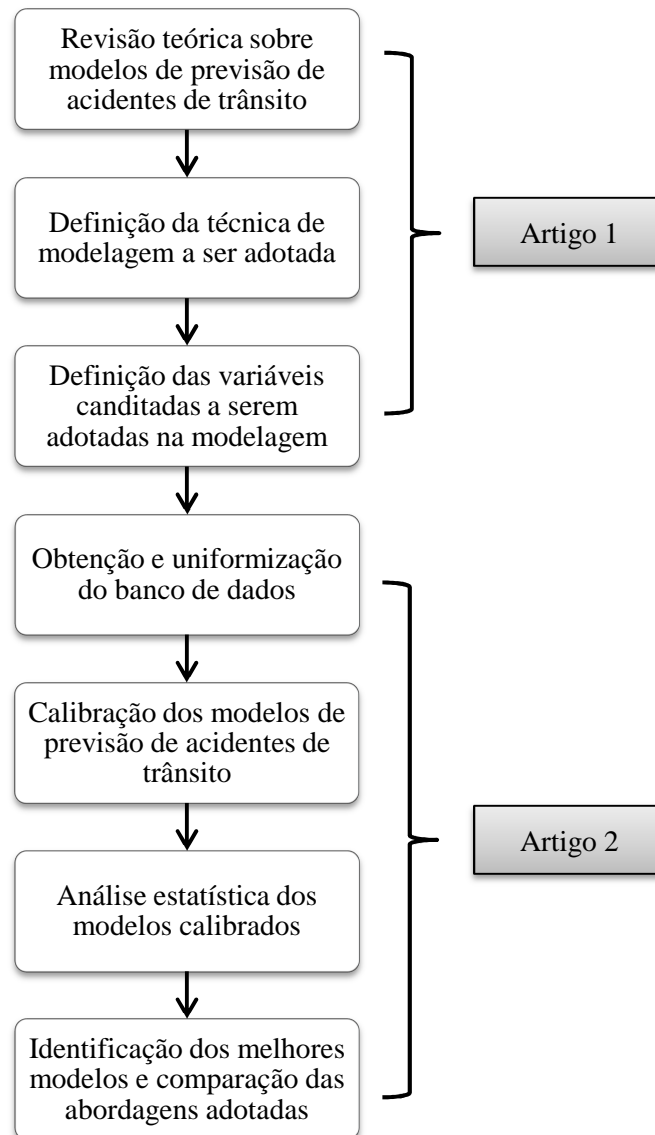


Figura 1: Estrutura do método de trabalho

1.5. Delimitação do Trabalho

O referencial teórico do estudo está baseado, principalmente, em trabalhos realizados com dados internacionais. Dessa forma é baseado em informações que não correspondem a dados coletados no Brasil. Grande parte das referências analisadas apresentam resultados de estudos realizados com dados dos Estados Unidos (EUA), os demais são, principalmente, estudos realizados com dados da Europa e Ásia. Tendo em vista as diferenças existentes no ambiente viário de cada país, os resultados obtidos nesses estudos podem ser diferentes dos resultados detectados no presente trabalho.

Com relação aos dados utilizados, o estudo é limitado à análise de um trecho de 214 quilômetros da rodovia BR-290 localizada no Estado do Rio Grande do Sul e administrada pelo Consórcio Univias. Essa rodovia é de pista simples com a existência de terceira faixa em alguns trechos. A análise é realizada com dados referente aos anos de 2008, 2009 e 2010. Como existe uma dificuldade para obter dados detalhados referente aos acidentes de trânsito, o estudo não considerou fatores relacionados aos condutores envolvidos nos acidentes de trânsito nem relacionados às condições climáticas no momento do acidente.

As variáveis usadas no estudo referem-se aos dados de infraestrutura e de tráfego da rodovia. Esses dados foram disponibilizados pela empresa que administra a rodovia, sendo uma parte referente ao levantamento técnico de engenharia da mesma e outra parte referente aos dados operacionais de volume e composição de tráfego.

1.6. Estrutura do Trabalho

Este trabalho está dividido em quatro capítulos. O primeiro capítulo apresenta uma visão geral do tema proposto destacando a importância desse estudo. Esse primeiro capítulo também destaca os objetivos da pesquisa, justificativa do tema e objetivos, o método utilizado e as limitações do trabalho.

O segundo capítulo é composto pelo artigo 1 que contém um levantamento dos estudos realizados sobre os modelos de previsão de acidentes de trânsito. Com esse levantamento percebe-se como e porque aconteceram as evoluções nos métodos de modelagem utilizadas, salientando as limitações de cada tipo de modelagem. Também são apresentados os objetivos de cada estudo, identificando as variáveis usadas e principalmente as que mostraram maior influência nos resultados.

O terceiro capítulo é composto pelo artigo 2, e apresenta uma descrição mais detalhada dos métodos estatísticos utilizados em estudos anteriores para prever a frequência de acidentes de trânsito. Essa descrição justifica a escolha do tipo de modelagem proposta para a aplicação nos dados levantados. A aplicação é feita com o intuito de comparar duas abordagens diferentes de estimar a frequência de acidentes de trânsito. Em uma delas o modelo é calibrado usando diversas variáveis de infraestrutura da rodovia (largura de pista,

existência de rampa, volume de tráfego, etc.), enquanto na outra somente a variável volume de tráfego é utilizada.

Por fim, o quarto capítulo retoma os conceitos do estudo, contextualizando os principais resultados alcançados e os comparando com outros estudos semelhantes. Assim, são estabelecidas as conclusões, visando melhorar as estimativas realizadas para a frequência de acidentes de trânsito em rodovias. Neste capítulo também são propostas sugestões para trabalhos futuros, discutindo as diferentes aplicações de modelos de previsão de acidentes de trânsito e como aplicá-los vinculados aos objetivos das pesquisas.

2. PRIMEIRO ARTIGO

MÉTODOS PARA PREVISÃO DE ACIDENTES: UMA REVISÃO DA LITERATURA

RESUMO

A redução dos acidentes rodoviários é um grande desafio aos pesquisadores e gestores de transportes. Neste artigo busca-se, através de um levantamento teórico e metodológico de modelos de previsão de acidentes de trânsito, identificar as principais variáveis adotadas bem como as técnicas utilizadas. Os modelos de previsão de acidentes de trânsito relacionam a ocorrência de acidentes em várias localidades da rodovia, ao tráfego e às características geométricas de cada localidade. Tais modelos têm diversas aplicações, tais como, a avaliação do potencial de segurança de locais da via, identificação e classificação de localidades perigosas ou com propensão a acidentes, avaliação da eficácia de medidas de melhoria da segurança e planejamento da segurança. A identificação das possibilidades de melhorias é feita através da análise variáveis mais influentes presentes nos modelos revisados. Para cada modelo revisado foram ainda identificadas suas principais diferenças e limitações.

Palavras-chave: Segurança Viária, Modelos de Previsão de Acidentes de Trânsito, Modelos Lineares Generalizados.

ABSTRACT

The reduction of road accidents is a major challenge for transportation authorities. This article aims at, by means of a theoretical and methodological review, identifying the key variables and the methodology adopted in accident prediction models. Accident prediction models relate the occurrence of accidents in various segments of motorways to both the traffic and geometric characteristics of each segment. Such models have several applications, such as the assessment of the safety potential of road segments, the identification and classification of dangerous locations or areas prone to accidents, the evaluation of the effectiveness of safety improvement measures and security planning. The identification of possible improvements is done by analyzing the most influential variables present in the models reviewed. The main differences between all models reviewed are presented, as well as their limitations.

Key- words: Road Safety, Accident Prediction Models, Generalized Linear Models

1. INTRODUÇÃO

A ocorrência de acidentes viários é influenciada por diversas variáveis. Entre as variáveis mais usadas estão o fluxo de tráfego, as características geométricas, as condições de superfície do pavimento, e as condições climáticas. A modelagem estatística é utilizada para desenvolver os modelos de previsão de acidentes em função de tais variáveis. Modelos que relacionam a ocorrência de acidentes em várias localidades da rodovia, ao tráfego e às características geométricas de cada localidade possuem diversas aplicações tais como a avaliação do potencial de segurança de locais da via, a identificação e classificação de localidades perigosas ou com propensão a acidentes, a avaliação da eficácia de medidas de melhoria da segurança e o planejamento da segurança.

Inicialmente, os modelos de previsão de acidentes viários foram baseados em modelos convencionais de Regressão Múltipla que partem dos pressupostos de erros distribuídos normalmente e de homocedasticidade, que significam variância do erro igual para quaisquer valores da variável preditora. Entretanto, diversos pesquisadores demonstraram que as suposições dos modelos convencionais de Regressão Linear Múltipla são violadas por dados de acidentes de trânsito (Jovanis e Chang, 1986; Joshua e Garber, 1990). Assim, os

pesquisadores da área de segurança viária passaram a adotar a utilização da Regressão de Poisson ou Regressão Binomial Negativa, através da técnica de Modelos Lineares Generalizados, para desenvolvimento de modelos de previsão de acidentes.

Entretanto os modelos que utilizam a distribuição de Poisson apresentam determinadas limitações. Uma delas é a consideração de que o número esperado de acidentes por unidade de tempo é igual à variância. Porém, em muitas análises de dados de acidente de trânsito, a variância mostra-se superior à média. Assim, a fim evitar a violação do pressuposto do modelo de distribuição de Poisson, a distribuição Binomial Negativa tem sido usada com frequência na modelagem da previsão de acidentes.

Na distribuição Binomial Negativa a média e a variância podem ser tratadas separadamente. Assim, a principal diferença entre os modelos que assumem a distribuição Binomial Negativa em comparação a distribuição de Poisson está na possibilidade de analisar casos onde os acidentes apresentam um padrão bastante disperso, que é incompatível com a distribuição de Poisson que assume igualdade entre média e variância.

O objetivo principal deste artigo é apresentar a evolução dos estudos realizados sobre modelos de previsão de acidentes, identificando as principais variáveis adotadas bem como as técnicas de modelagem utilizadas. Este artigo está organizado em 4 seções. As seções 2 e 3 apresentam, respectivamente, o referencial teórico sobre modelos de previsão de acidentes e a compilação dos fatores relacionados aos acidentes de trânsito. Por fim, a seção 4 contém as conclusões obtidas a partir da comparação dos estudos revisados.

2. MODELOS DE PREVISÃO DE ACIDENTES DE TRÂNSITO

Com o objetivo de entender a relação entre a frequência de acidentes de trânsito e as características geométricas e operacionais das rodovias, diferentes métodos de modelagem vem sendo adotadas pelos pesquisadores. Entre os mais adotados estão: Modelos de Regressão Linear e Modelos Lineares Generalizados. O Modelo de Regressão Linear utilizado é a Regressão Linear Múltipla, já os Modelos Lineares Generalizados são a Regressão de Poisson e Regressão Binomial Negativa. Pesquisas recentes mostram que na Regressão Linear Múltipla três pressupostos não são satisfeitos quando aplicada na análise de

acidentes, sendo esses a homocedasticidade, a existência de valores negativos da variável dependente e a ocorrência de distribuição normal no erro (Jovanis e Chang, 1986). Com o intuito de solucionar os problemas associados aos modelos de Regressão Linear Múltipla, esses autores propuseram a Regressão de Poisson. Eles concluíram que a Regressão de Poisson é uma alternativa superior a Regressão Linear quando aplicada na segurança viária.

2.1. Evolução dos modelos de previsão de acidentes de trânsito

Joshua e Garber (1990) pesquisaram a relação entre fatores geométricos da via e acidentes com veículos de carga, utilizando modelos de Regressão Linear e Regressão de Poisson. Novamente concluíram que as técnicas de Regressão Linear utilizadas em sua pesquisa não descreveram adequadamente o relacionamento entre acidentes com caminhão e as variáveis independentes. Já o modelo de Regressão de Poisson apresentou melhor desempenho na representação dessa relação.

O modelo de Regressão de Poisson também foi utilizado por Miaou *et al.* (1992) para estabelecer a relação entre acidentes com caminhões e fatores geométricos de estradas rurais. O modelo estimado sugeriu que o tráfego diário médio anual por pista, a curvatura horizontal e a inclinação das rampas eram correlacionados significativamente com a probabilidade de ocorrer um acidente com caminhões. Porém, no decorrer desse trabalho, uma violação de um pressuposto do modelo de Poisson foi encontrada. Para usar o modelo de Poisson é necessário que a média e a variância da variável dependente sejam iguais. No entanto, na maioria dos dados de acidente, a variância da frequência de acidente não é igual à sua média.

Em 1993 um estudo complementar (Miaou e Lum, 1993) teve como principal finalidade avaliar as propriedades estatísticas de dois modelos de Regressão Linear convencionais e de dois modelos de Regressão de Poisson. Os modelos estudados eram comparáveis àqueles desenvolvidos em estudos precedentes explorando o relacionamento entre acidentes e fatores geométricos de estradas. Os quatro tipos de modelos considerados foram: um modelo de Regressão Linear Aditivo, um modelo de Regressão Linear Multiplicativo, uma Regressão de Poisson Multiplicativa com função exponencial e uma Regressão de Poisson Multiplicativa com função não exponencial. Nesse estudo os autores

confirmaram que os modelos de Regressão de Poisson são superiores aos modelos de Regressão Linear na previsão da ocorrência de acidentes. Além disso, constatou-se que o modelo de Regressão de Poisson com a função exponencial teve o melhor desempenho entre os quatro modelos avaliados. No mesmo estudo, os autores ainda tentaram avaliar a existência da grande variabilidade na frequência de acidentes de trânsito. Assim, concluíram que, quando existe sobredispersão (variância maior que a média) nos dados e o modelo de Poisson é usado, os coeficientes estimados do modelo tendem a ser subestimados (apresentar valor inferior ao real). Por causa disso, os autores sugeriram o uso de uma distribuição de probabilidade mais geral como a Binomial Negativa.

As distribuições de Poisson e Binomial Negativa foram utilizadas em estudos conduzidos por Shankar *et al.* (1995) para relacionar acidentes de trânsito com fatores geométricos e ambientais. Esse estudo teve como objetivo avaliar os efeitos desses fatores na frequência de acidente em uma rodovia do Estado de Washington. Além da quantidade total de acidentes em seções da rodovia, os autores modelaram também a quantidade de alguns tipos específicos de acidentes. No final do trabalho os autores concluíram que os modelos de regressão separados por tipo específico de acidentes apresentaram melhores ajustes, sendo isto estatisticamente confirmado.

Poch e Mannering (1996) utilizaram o modelo de regressão Binomial Negativa para modelar a frequência de acidentes em cruzamentos de Bellevue em Washington. O estudo forneceu um passo importante para desenvolver uma abordagem sistemática e estatisticamente defensável para identificar os impactos das melhorias possíveis em interseções. Concluindo que a Regressão Binomial Negativa é uma ferramenta com boa capacidade de previsão, sendo eficiente sua utilização na previsão de frequência de acidentes.

Uma abordagem diferente da tradicionalmente adotada foi apresentada por Persaud *et al.* (2000). Ao invés de estimar os acidentes para trechos viários, esse estudo analisou separadamente a quantidade de acidentes em entidades específicas da rodovia como curvas e retas utilizando regressão através dos modelos lineares generalizados. As variáveis independentes utilizadas no estudo foram fluxo de tráfego e geometria da estrada. Uma variável do tipo “dummy” foi utilizada para avaliar “terreno plano” e “terreno montanhoso”. Os resultados mostraram que em curvas as variáveis volume diário médio anual, comprimento

da seção e grau de curvatura aumentam a quantidade de acidentes de trânsito. Já para as retas somente as variáveis volume diário médio anual e comprimento da seção aumentam a quantidade de acidentes de trânsito. Além disso, existiu um número maior de acidentes de trânsito em “terreno montanhoso”.

No mesmo ano Abdel-Aty e Essam Radwan (2000) utilizaram a distribuição Binomial Negativa para prever a quantidade de acidentes de trânsito em função das variáveis volume diário médio anual, grau de curvatura horizontal, comprimento da seção, pista e larguras, em vias urbanas e rurais. O diferencial deste estudo é que os modelos também foram desenvolvidos para explicar a influência das características do motorista, incluindo as variáveis sexo (feminino e masculino) e idade (separada em três classes). Os resultados mostraram que a probabilidade de ocorrência de acidentes de trânsito aumenta com o aumento do tráfego diário médio anual, grau de curvatura horizontal e comprimento da seção. Com relação aos condutores concluíram que condutores do sexo masculino tem tendência a se envolver em acidentes de trânsito relacionados ao excesso de velocidade. E ainda, o aumento do tráfego diário médio anual por faixa aumenta a probabilidade de ocorrência de acidentes em todas as idades. Porém, o efeito relacionado a esta variável é maior para os condutores jovens e idosos do que para os condutores de meia idade. Portanto, os condutores mais jovens e idosos sofrem mais problemas do que condutores de meia idade quando existe grande volume de tráfego.

Contrariando a tendência do uso de modelos de regressão lineares generalizados baseados na distribuição Binomial Negativa Greibe (2003) utilizou modelos baseados na distribuição de Poisson para prever a taxa de acidentes de trânsito em 1036 interseções e 142 km de segmentos viários da Dinamarca. Neste estudo a distribuição de Poisson foi escolhida simplesmente por ser mais fácil de usar e, como o estudo foi planejado e iniciado no ano de 1990, a distribuição de Poisson era a única solução viável com o software estatístico disponível no momento. Como resultado constatou-se que a variável volume de tráfego tem maior poder de explicação principalmente para os modelos de interseções. Já para os modelos de segmentos rodoviários, outras variáveis como largura da via, número de acessos e uso do solo também se mostraram variáveis significativas e importantes para prever a número de acidentes.

Ainda, no ano de 2004 Hauer (2004a) desenvolveu um modelo de previsão de acidentes usando a distribuição Binomial Negativa. A variável dependente utilizada foi o número de acidente por ano, e as independentes foram características geométricas da via e fluxo de tráfego. O autor sugeriu maneiras para atribuir a forma funcional a cada uma das variáveis do modelo, e observou que a equação modelo deve ter componentes multiplicativos e aditivos. O componente multiplicativo é utilizado para explicar a influência das variáveis existente ao longo de uma rodovia (tal como a largura da pista), enquanto que o componente aditivo explica a presença de pontos perigosos (tal como rodovias com pontes estreitas).

Em um estudo complementar Hauer (2004b) aplicou o modelo estatístico mencionado anteriormente à frequência de acidentes de trânsito em quatro estradas urbanas. Os modelos propostos avaliaram o número de acidentes por ano em função das variáveis independentes volume diário médio anual, percentual de caminhões, grau de curvatura e comprimento de curvas horizontais, classe das tangentes e do comprimento de curvas verticais, largura da pista, limite de velocidade, pontos de acesso, a presença e a natureza de estacionamento. Os resultados apontaram como variáveis significativas o volume diário médio anual, número de entradas comerciais de automóveis e limite de velocidade.

Em um estudo desenvolvido em Porto Alegre, capital do Rio Grande do Sul, apresentou modelos para previsão de acidentes de trânsito em vias arteriais urbanas. O objetivo do trabalho foi gerar modelos de previsão de acidentes de trânsito a partir de variáveis relacionadas à exposição e a características viárias. Foi utilizada a técnica de Modelos Lineares Generalizados com distribuição de Poisson para relacionar a quantidade de acidentes de trânsito com as características viárias e de tráfego. Os modelos de previsão de acidentes de trânsito calibrados explicaram mais de 60% da variabilidade dos dados, comprovando a relação entre os acidentes de trânsito e as variáveis estudadas (Cardoso, 2006).

No mesmo ano, utilizando Regressão Binomial Negativa, um estudo realizado em interseções rurais da Geórgia (Jutaek *et al.*, 2006) revelou que a variável volume de tráfego afeta significativamente o desempenho global da segurança em interseções, independente do tipo de cruzamento. Já a significância das variáveis relacionadas às características

geométricas das interseções varia em função do tipo de cruzamento assim como o tipo de acidente.

Em um estudo mais recente Kim *et al.* (2007) propuseram um modelo para relacionar os acidentes ocorridos em interseções de rodovias rurais do Estado da Geórgia (EUA). As características geométricas e ambientais utilizadas para medir a correlação foram: tipo de pavimento, tipo de sinalização, largura da plataforma da via, quantidade de movimentos conflitantes, tipo e ângulo de colisão, condições atmosféricas, nível de tráfego, entre outras. O método utilizado foi baseado na distribuição logística. Os resultados indicaram que existe uma relação entre as características ambientais e geométricas das interseções e os acidentes ocorridos nestes locais.

Com a utilização de modelos lineares generalizados Kwak *et al.* (2010) desenvolveram modelos de previsão de acidentes para vias expressas da Coreia. Os resultados mostraram que os acidentes de trânsito nessas vias podem ocorrer com mais frequência em segmentos da estrada onde o volume de tráfego aumenta de forma descontínua, como aqueles que possuem pedágio. Esses resultados também podem ser utilizados como ferramenta de avaliação de segurança a fim de escolher áreas para a implantação de contramedidas de acidentes.

Em resumo os pesquisadores mostraram que a Regressão Linear Múltipla não é um método apropriado para modelar a relação entre a ocorrência de acidentes, e os fatores geométricos e de tráfego da via. Assim, a Regressão de Poisson e a Regressão Binomial Negativa são mais apropriadas para modelar essa relação. Ressaltando a Regressão Binomial Negativa é a que apresenta pressupostos mais fáceis de serem atendidos.

2.2. Comprimento dos Segmentos

Outro aspecto importante discutido em artigos revisados refere-se à determinação do comprimento do segmento rodoviário a ser utilizado na modelagem. A definição do comprimento e dos critérios de determinação dos segmentos viários na utilização de modelos de previsão de acidentes é fundamental já que pode influenciar nos resultados dos mesmos de forma significativa (Nodari, 2003). As alternativas mais recomendadas para determinação do

comprimento do segmento são o uso de segmentos de tamanho fixo ou as seções homogêneas, ou seja, segmentos com características geométricas homogêneas (Miaou *et al.* 1991).

No que diz respeito aos segmentos homogêneos levantam-se diversos problemas importantes. Um destes problemas é que as rodovias com número elevado de curvas tendem a produzir segmentos menores que um quilômetro de comprimento para assegurar a homogeneidade na geometria. Isto pode conduzir a problemas com relação ao erro de localização dos acidentes, pois os mesmos, na maioria das vezes, são relatados com relação ao quilometro mais próximo. Outro problema que pode ser citado está relacionado ao fato do erro de previsão do modelo possuir um componente aleatório em cada amostra (Miaou e Lum, 1993). Assim, com a existência de segmentos curtos, existirá um aumento deste efeito em virtude de se ter maior quantidade de observações na amostra. Entretanto, se forem adotadas segmentos mais longos o poder de explicação do modelo será diminuído em consequência da redução de quantidades de amostras. Neste caso há o risco de existirem trechos não homogêneos por conta da grande extensão. O modelo proposto por Miaou e Lum (1993) adota segmentos longos e variáveis.

As desvantagens de usar segmentos com comprimento fixo são menos severas, comparando-se ao uso de segmentos homogêneos. A maioria das desvantagens podem ser superadas explicando a não homogeneidade dos segmentos incluindo medidas detalhadas da variabilidade entre os segmentos na especificação do modelo (por exemplo número das curvas). Se tais dados estão disponíveis, não há necessidade de restringir a análise baseando-se em segmentos homogêneos (Shankar *et al.*, 1995).

3. SÍNTESE E DISCUSSÃO DOS ESTUDOS REVISADOS

A utilização de modelos de previsão de acidentes normalmente tem como objetivo determinar o número esperado de acidentes de trânsito em um local determinado. Isto é realizado através da utilização de modelos estatísticos que resultam na estimativa da frequência com que os acidentes ocorrem considerando determinadas condições, tais como, de tráfego, de clima e de geometria da via. Assim pode-se concluir que não se tem como prever quando e onde irá ocorrer um acidente, mas tem-se como determinar um valor esperado de

acidentes de trânsito a partir de características com efeitos significativos sobre a ocorrência dos mesmos.

Para desenvolver um modelo de previsão de acidentes, inicialmente é necessário determinar quais as variáveis disponíveis que podem explicar a ocorrência de acidentes. Assim, após escolher o tipo de modelo que será utilizado devem-se buscar as variáveis significativas para serem utilizadas no modelo. A seguir mostram-se os tipos de modelos existentes e variáveis utilizadas a partir do referencial teórico analisado.

A Figura 1 apresenta uma linha do tempo mostrando a evolução dos tipos de modelos de previsão de acidentes utilizados em diversos tipos de estudos revisados. Analisando a Figura 1 tem-se que os primeiros estudos realizados para prever a frequência de acidentes de trânsito utilizavam modelos convencionais de Regressão Linear Múltipla. Porém, alguns pressupostos desse tipo de modelo eram violados quando aplicados a dados de acidentes de trânsito (Jovanis e Chang, 1986; Joshua e Garber, 1990). Assim, os modelos lineares generalizados passaram a ser utilizados para prever a frequência de acidentes de trânsito.

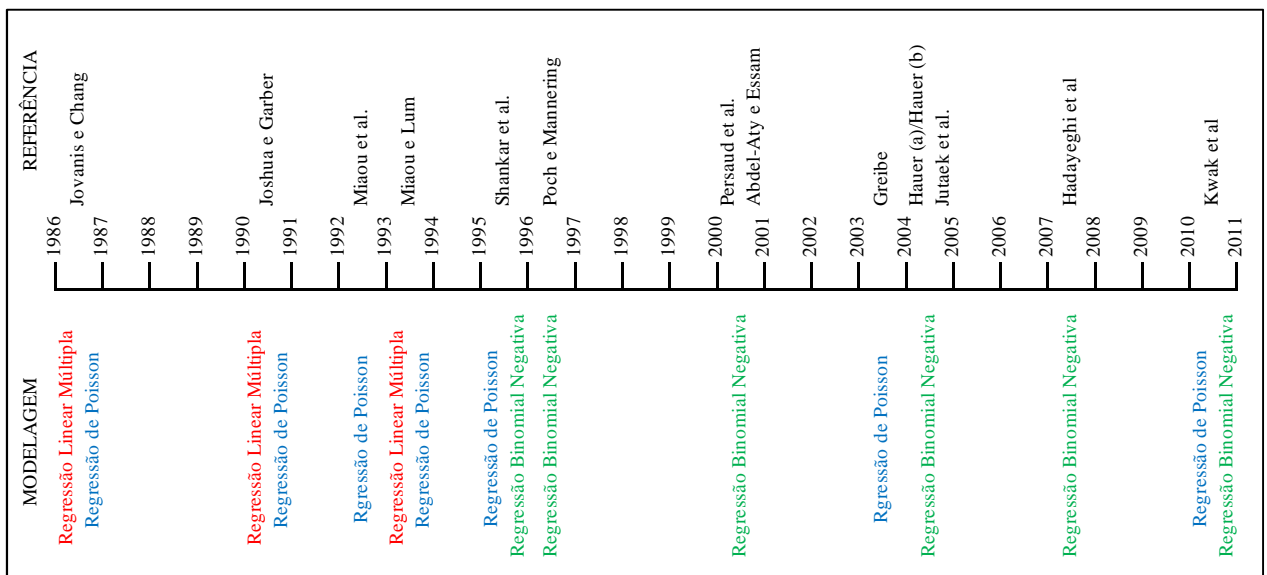


Figura 1: Evolução dos modelos de previsão de acidentes de trânsito.

As primeiras modelagens baseadas na utilização de modelos lineares generalizados utilizaram a distribuição de Poisson (Miaou *et al.*, 1992; Miaou e Lum, 1993 e Shankar *et al.*,

1995). Entretanto a distribuição de Poisson apresenta como definição média igual a variância, o que normalmente não ocorre em dados de acidentes de trânsito, como já foi dito anteriormente. Assim, os estudos passaram a utilizar modelos lineares generalizados baseados na distribuição Binomial Negativa (Persaud *et al.*, 2000; Abdel-Aty e Essam Radwan, 2000; Juttaek *et al.*, 2006 e Kwak *et al.*, 2010).

O Quadro 1 apresenta algumas características dos principais estudos analisados. Dentre essas características estão autor(es), ano, local, objetivo e técnica de análise do estudo. Através da análise desta tabela tem-se a evolução dos modelos de previsão de acidentes de trânsito.

Observando o Quadro 1 percebe-se que grande parte das referências existentes apresentam resultados de estudos realizados nos Estados Unidos (EUA), os demais são principalmente estudos realizados com dados da Europa e Ásia. Com relação aos objetivos dos estudos pode-se dizer que eles estão divididos em dois grupos, um que desenvolve modelos de previsão de acidentes ao longo de vias e outro em interseções.

Nos modelos de previsão de acidentes ao longo de vias as principais variáveis utilizadas para prever a frequência de acidentes estão relacionadas às características geométricas e de tráfego da via (Greibe 2003; Kwak *et al.*, 2010). Algumas outras variáveis utilizadas estão relacionadas a efeitos climáticos (Shankar *et al.*, 1995) ou até mesmo incluindo dados do condutor, como sexo e faixa etária (Abdel-Aty e Essam Radwan, 2000).

Já com relação aos modelos utilizados para prever acidentes em interseções, além da utilização da variável volume de tráfego, são normalmente utilizadas as variáveis relacionadas com as características geométricas da via (Poch e Mannering, 1996; Juttaek *et al.*, 2006). Para esse tipo de estudos pode-se também analisar os tipos de acidentes que ocorrem com maior frequência (Kim *et al.*, 2007). Além disso, pode ser de interesse comparar a frequência de acidentes em interseções semaforizadas e não semaforizadas (Poch e Mannering, 1996; Kim *et al.*, 2007).

Quadro 1: Descrição dos principais modelos analisados

REFERÊNCIA	LOCAL DO ESTUDO	OBJETIVO DO ESTUDO	TÉCNICA DE ANÁLISE
Shankar <i>et al.</i> (1995)	Rodovias de Washington	Explorar a frequência de acidentes em rodovias com base em uma análise multivariada utilizando variáveis geométricas da via, climáticas e alguns efeitos sazonais.	Modelo de Regressão Binomial Negativa e Poisson
Poch e Mannering (1996)	63 cruzamentos de Bellevue em Washington	Descobrir interações importantes entre elementos geométricos e relacionados ao tráfego e à frequência de acidentes em cruzamentos.	Modelo de Regressão Binomial Negativa
Abdel-Aty e Essam Radwan (2000)	Vias arteriais da Florida Central	Desenvolver um modelo matemático que explique a relação entre frequência de acidentes e características geométricas e de tráfego da via. Outros objetivos incluem o desenvolvimento de modelos por sexo e faixa etária do condutor.	Modelo de Regressão Binomial Negativa
Greibe (2003)	1036 interseções e 142 km de segmentos viários urbanos da Dinamarca	Estabelecer modelos para prever o número esperado de acidentes em interseções e segmentos viários em áreas urbanas.	Modelo de Regressão de Poisson
Jutaek <i>et al.</i> (2006)	Georgia, Califórnia e Michigan	Estabelecer relações entre a ocorrência de acidentes de trânsito e características geométricas de interseções semaforizadas.	Modelo de Regressão Binomial Negativa
Hadayeghi <i>et al.</i> (2007)	Toronto (Canadá)	Desenvolver modelos de previsão de acidentes zonais comumente utilizados para o planejamento do transporte urbano.	Modelo de Regressão Binomial Negativa
Kim <i>et al.</i> (2007)	91 interseções na Georgia	Identificar os fatores que afetam a probabilidade de que certos tipos de acidentes ocorram em interseções.	Modelo de Regressão Binomial Logística
Kwak <i>et al.</i> (2010)	Vias Expressas da Coreia	Desenvolver modelos de previsão de acidentes para vias expressas da Coreia com a utilização de variáveis geométricas e ambientais analisando os fatores influentes.	Modelo de Regressão Binomial Negativa, Poisson, Binomial Negativa com muitos zeros e Poisson com muitos zeros

O Quadro 2 apresenta a lista das variáveis utilizadas nos modelos dos principais estudos analisados. Observando o Quadro 2 percebe-se, em quase todos os estudos, a presença de alguma variável relacionada ao volume de veículos que trafegam no segmento estudado. No estudo de Abdel-Aty e Essam Radwan (2000) o volume de tráfego apareceu como o fator que mais influencia para o aumento da frequência de acidentes. O mesmo ocorreu em Greibe (2003) constatando-se que a variável volume de tráfego é a mais poderosa principalmente para os modelos de interseções visto que para os modelos de segmentos rodoviários, outras

variáveis como largura da via, número de acessos e uso do solo também se mostraram variáveis significativas e importantes para prever a frequência de acidentes. Ao averiguar que variáveis relacionadas a volume de tráfego são significativas para a previsão de acidentes de trânsito indica-se a necessidade de se obter valores confiáveis para essa variável (Jutaek et al., 2006).

O estudo específico com cruzamentos (Poch e Mannering, 1996) apresentou que o aumento do volume de tráfego que converge para a esquerda aumenta a frequência de acidentes. Verificando também que o aumento de 1% no volume de tráfego que converge para a esquerda aumenta em 2,28% a frequência de acidentes. O aumento de volume de tráfego que converge para a direita também aumenta a frequência de acidentes. Para um aumento de 1% no volume de tráfego que converge para a direita tem-se um aumento de apenas 0,92% na frequência de acidentes. No mesmo estudo constatou-se que quando as interseções são sinalizadas existe uma redução significativa na frequência de acidentes e, acessos que foram classificados como acesso local apresentam menor frequência de acidentes. Por definição, as ruas locais apresentam menor quantidade de volume de tráfego quando comparada as outras classificações. Assim, esta pequena quantidade de volume de tráfego faz com que exista uma menor possibilidade de acidente.

Com relação às variáveis relacionadas a geometria da via destaca-se que para reduzir a probabilidade de acidentes em áreas que sofrem problemas climáticos adversos, o projeto geométrico não deve ser focado apenas em pavimento molhado, ou seja, quando chove. Devendo ser realizado também um grande esforço para evitar curvas horizontais com velocidade de projeto alta em áreas que apresentam condições meteorológicas adversas como neve, por exemplo (Shankar et al., 1995).

A partir de modelos que utilizaram variáveis relacionadas aos condutores conclui-se que motoristas do sexo feminino têm maior probabilidade de se envolver em acidentes do que os condutores do sexo masculino quando existir maior volume de tráfego e a largura da faixa for reduzida. Além disso, condutores do sexo masculino tendem a se envolver mais em acidentes relacionados ao excesso de velocidade. Relacionado a idade tem-se que os condutores jovens e os idosos tem maior possibilidade de se envolver em acidente de trânsito

do que os condutores de meia idade quando o volume de tráfego for maior (Abdel-Aty e Essam Radwan, 2000).

Quadro 2: Descrição das variáveis dos principais modelos analisados

REFERÊNCIA	VARIÁVEIS UTILIZADAS	
Shankar <i>et al.</i> (1995)	<ul style="list-style-type: none"> - Número de curvas horizontais; - Raio horizontal máximo da curva; - Raio horizontal mínimo da curva; - Precipitação média mensal; - Máximo de precipitação diária do mês; 	<ul style="list-style-type: none"> - Mínimo de precipitação pluviométrica diária do mês; - Média mensal de neve; - Máxima queda de neve diária no mês; - Número de dias de neve no mês.
Poch e Mannering (1996)	<ul style="list-style-type: none"> - Volume diário médio; - Sinalização (1 semaforizada - 0 não semaforizada); - Tipo de sinalização; 	<ul style="list-style-type: none"> - Tipo de acesso; - Limites de velocidade; - Visibilidade.
Abdel-Aty e Essam Radwan (2000)	<ul style="list-style-type: none"> - Tráfego diário médio anual; - Grau de curvatura horizontal; - Classificação urbana/rural; - Limite de velocidade estabelecido; - Número de faixas; - Tipo de acostamento; 	<ul style="list-style-type: none"> - Pavimento no acostamento; - Superfície da faixa; - Largura divisória central (m); - Largura do acostamento (m); - Largura da faixa (m); - Comprimento da seção (m).
Greibe (2003)	<ul style="list-style-type: none"> - Fluxo de tráfego; - Velocidade; - Número de acessos; - Número de faixas; - Classe da via; - Ciclovias/ciclofaixas; 	<ul style="list-style-type: none"> - Divisor físico central; - Estacionamento; - Parada de ônibus; - Uso do solo; - Fluxo de pedestre.
Jutaek <i>et al.</i> (2006)	<ul style="list-style-type: none"> - Volume diário médio; - Ângulo entre as estradas principais e secundárias; - Grau da curva horizontal na estrada principal; - Grau da curva horizontal na estrada secundária; - Iluminação na interseção (0 não - 1 sim); - Largura média da estrada principal; 	<ul style="list-style-type: none"> - Porcentagem que converte à esquerda; - Percentual de caminhão passando na interseção; - Velocidade indicada na estrada principal nas proximidades da interseção; - Velocidade indicada na estrada secundária nas imediações da interseção.
Hadayeghi <i>et al.</i> (2007)	<ul style="list-style-type: none"> - Intensidade de tráfego; - Rede viária; - Uso do Solo; 	<ul style="list-style-type: none"> - Características socioeconômicas; - Características demográficas.
Kim <i>et al.</i> (2007)	<ul style="list-style-type: none"> - Condições do tempo (1 boa - 0 ruim); - Superfície (1 molhada - 0 seca); - Dia (1 dia - 0 noite); - Curva (1 curva horizontal - 0 c.c.); 	<ul style="list-style-type: none"> - Acostamento (1 com - 0 sem); - Sinalizada (1 semaforizada - 0 não semaforizada); - Ângulo da interseção (1 90° - 0 c.c.)
Kwak <i>et al.</i> (2010)	<ul style="list-style-type: none"> - Exposição - milhões de veículos por km; - Raio da curva horizontal (metros); - Grau de curva vertical; - Subida - (0 não - 1 sim); - Descida - (0 não - 1 sim); - Pista de acesso - (0 não - 1 sim); - Pista de saída - (0 não - 1 sim); 	<ul style="list-style-type: none"> - Número de vias - (0 não - 1 sim); - Cancela de Pedágio - (0 não - 1 sim); - Ponte - (0 não - 1 sim); - Túnel - (0 não - 1 sim); - Radares - (0 não - 1 sim); - Área de serviço - (0 não - 1 sim); - % de Caminhões.

No trabalho de Hadayeghi et al. (2007) a variável chamada de exposição, definida como quilometragem média percorrida pelos veículos, apareceu como a principal variável do estudo. Essa variável contribuiu significativamente para o aumento da frequência de acidentes.

Outras variáveis que também contribuem para esse o aumento são: velocidade média estabelecida, total de vias arteriais (km), total de vias expressas (km), total de vias coletoras (km), total de vias (km), número de cruzamentos sinalizados, número de escolas, total de área comercial (m²); total de área do governo e institucional (m²) e total de área residencial (m²). Já as variáveis que diminuem a frequência de acidentes são: total de ferrovias (km), total de vias locais (km) e total de área aberta (m²).

Utilizando Regressão Binomial Logística com o objetivo de identificar os fatores que afetam a probabilidade de que certos tipos de acidentes ocorram em interseções foi averiguado que colisões laterais apresentam maior chance de ocorrer em condições de tempo bom, durante o dia, em curvas horizontais e interseções não semaforizadas. Já as colisões traseiras apresentam maior chance de ocorrer durante o dia e em interseções semaforizadas. Os abalroamentos com veículos trafegando no mesmo sentido apresentam maior chance de ocorrer fora de curvas horizontais, vias que não possuem acostamento e com ângulo que não seja de 90°. Para os abalroamentos com veículos trafegando em sentido contrário existe maior chance de ocorrer em condições ruins de tempo, com superfície seca, durante a noite e fora de curvas horizontais. Por fim, as colisões frontais não apresentaram nenhuma variável significativa (Kim et al., 2007).

4. CONCLUSÕES

Este estudo apresenta a evolução dos tipos de modelagens utilizadas na previsão de acidentes bem como as variáveis utilizadas. O objetivo é apresentar alguns estudos realizados sobre previsão de acidentes, identificando as metodologias e variáveis utilizadas.

Inicialmente identifica-se que os modelos de Regressão Linear Múltipla eram sugeridos. Porém, após perceber que alguns pressupostos estavam sendo violados, outra técnica de modelagem foi sugerida, utilizando modelos lineares generalizados. Entre esses modelos, primeiramente, foram sugeridos os que são baseados na distribuição de Poisson. Como a distribuição de Poisson assume média igual a variância e nos dados de ocorrência de acidentes, normalmente, a variância é superior a média, a utilização desses modelos não é eficiente. Então os modelos lineares generalizados que utilizam a distribuição Binomial Negativa passaram a ser utilizados.

Com relação às variáveis utilizadas nos modelos apresentados nos estudo analisados, existe uma lista ampla. Porém, as que aparecem com mais frequência são as variáveis volume de tráfego, grau de curvatura, existência de rampa, largura de faixa, comprimento da seção e quantidade de cruzamentos. Dessas as que mais influenciam na ocorrência de acidentes de trânsito são as variáveis referente a cruzamentos, comprimento dos segmentos e principalmente volume de tráfego.

REFERÊNCIAS

Abdel-Aty, M.A., Essam Radwan, E.A., 2000. Modeling traffic accident occurrence and involvement. *Accident Analysis and Prevention* 32, 633–642.

Cardoso, G., 2006. Modelos de previsão de acidentes de trânsito em vias arteriais urbanas. Tese de Doutorado, Programa de Pós-graduação em Engenharia de Produção, Universidade federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brasil.

Greibe, P., 2003. Accident prediction models for urban roads. *Accident Analysis and Prevention* 35, 273–285.

Hadayeghi, A., Shalaby, A.S., Persaud, B.N., 2007. Safety Prediction Models: A Proactive Tool for Safety Evaluation in Urban Transportation Planning Applications. *Transportation Research Board of the National Academies*, January 225 – 236.

Hauer, E., 2004a. Statistical road safety modeling. In: *Proceedings of the 83rd TRB Annual Meeting*, Washington, DC, USA, January 11–15.

Hauer, E., 2004b. Safety models for urban four-lane undivided road segments. In: *Proceedings of the 83rd TRB Annual Meeting*, Washington, DC, USA, January 11–15.

Joshua, S., Garber, N., 1990. Estimating truck accident rate and involvement using linear and Poisson regression models. *Transportation Planning and Technology* 15, 41–58.

Jovanis, P.P., and Chang, H.L. 1986. Modeling the relationship of accidents to miles traveled. *Transportation Research Record* 1068, pp. 42–51.

Jutaek, O; Simon, W; Doohee, 2006, N. Accident Prediction Model for Railway - Highway Interfaces. *Accident Analysis and Prevention* 38, 346–356.

Kim, D.G.; Lee, Y.; Washington, S.; Choi, K. 2007. Modeling crash outcome probabilities at rural intersections: Application of hierarchical binomial logistic models. *Accident Analysis and Prevention*, v39, p.125-134.

Kwak, H.; Kim, D. K.; Park, S. H.; Kho, S. Y. 2010. Development of a Safety Performance Function for Korean Expressways. 12th WCTR, July 11-15, 2010 – Lisbon, Portugal.

Miaou, S. P.; Hu, P. S.; Wright, T.; Davis, S. C.; Rathi, A. K. 1991. Development of relationships between truck accidents and highway geometric design: Phase I. Technical memorandum prepared by the Oak Ridge National Laboratory. Washington, DC: Federal Highway Administration.

Miaou, S., Hu, P., Wright, T., Rathi, A., Davis, S., 1992. Relationship between truck accidents and highway geometric design: a Poisson regression approach. *Transportation Research Record* 1376, 10–18.

Miaou, S., and Lum, H. 1993. Modeling vehicle accident and highway geometric design relationships. *Accident Analysis and Prevention*, 25(6): 689–709.

Nodari C. T. 2003. Método de avaliação da segurança potencial de segmentos rodoviários rurais de pista simples. 215 p. Tese (Doutorado em Engenharia de Produção). Universidade Federal de Rio Grande do Sul.

Persaud, B., Retting, R.A., Lyon, C., 2000. Guidelines for the identification of Hazardous Highway Curves. *Transportation Research Record* 1717, 14–18.

Poch, M., Mannering, F., 1996. Negative binomial analysis of intersection accident frequencies, *Journal of Transportation Engineering* 122(2).

Shankar, V., Mannering, F., Barfield, W., 1995. Effect of roadway geometrics and environmental factors on rural freeway accident frequencies. *Accident Analysis and Prevention*. 27 (3), 542–555.

3. SEGUNDO ARTIGO

MODELOS DE PREVISÃO DE ACIDENTES DE TRÂNSITO: QUANTO MAIS VARIÁVEIS MELHOR?

RESUMO

Os modelos de previsão de acidentes são utilizados para diversas finalidades, tais como, avaliar a segurança de locais da rodovia, classificar localidades que apresentam maior propensão a ocorrência de acidentes além de avaliar a eficácia de medidas de melhorias de segurança realizadas e de contribuir no planejamento da segurança. Neste artigo busca-se comparar duas abordagens distintas para previsão de acidentes. A primeira consiste em estimar a ocorrência de acidentes em segmentos da via decorrentes de mudanças nas variáveis que se referem ao volume de tráfego e elementos de infraestrutura. A segunda relaciona a frequência de acidentes em um determinado elemento de infraestrutura da via (de interseção, curva, tangente, etc.), chamada entidade, com base apenas na variável referente ao volume de tráfego. Entre os resultados, destaca-se que a modelagem por entidade apresenta melhores resultados.

Palavras-chave: Segurança Viária, Modelos de Previsão de Acidentes, Modelos Lineares Generalizados.

ABSTRACT

Accident prediction models are employed for various purposes, such as to assess the safety of road segments and identify and classify dangerous locations that are more prone to accidents, besides evaluating the effectiveness of safety improvement measures carried out and security planning. This article seeks to compare two different accident prediction methods. The first method consists in estimating the occurrence of accidents in road sections with changes in the characteristics of infrastructure elements. The second one relates the frequency of accidents based on a single infrastructure element (intersection, curve, tangent, etc.) based on traffic volume only. The results indicated that the entity models produces better results.

Key- words: Road Safety, Accident Prediction Models, Generalized Linear Models

1. INTRODUÇÃO

Os modelos de previsão de acidentes são, normalmente, modelos matemáticos que buscam estimar o valor esperado para a quantidade de ocorrência de acidentes em função de variáveis capazes de prever esse fenômeno. As variáveis que influenciam na ocorrência dos acidentes estão, normalmente, associadas às características operacionais e físico-ambientais da rodovia. Entre as variáveis que se referem a características operacionais estão o volume e/ou composição de tráfego. Já em relação às características físico-ambientais, destacam-se as curvas horizontais e verticais, a presença de 3ª faixa e a largura da faixa e as condições meteorológicas no momento do acidente. Ainda, são candidatas a variáveis explicativas, as características do condutor.

Os modelos de previsão de acidentes podem ser concebidos basicamente de duas formas: inicialmente é possível utilizar modelos de regressão múltipla onde as rodovias são divididas em segmentos e, a esses segmentos, são associadas características operacionais, físicas ou ambientais para estimativa do número de acidentes no segmento. Nesse caso, um determinado segmento da rodovia terá associado a ele os valores de diversas variáveis, entre elas, a presença de interseção e/ou curva horizontal, bem como as tradicionais variáveis de volume de tráfego, largura de faixa, etc. Alternativamente, é possível optar pela estimativa de modelos de regressão simples específico por entidade. Nesse caso, será estimado um modelo

específico para a entidade curvas e outro modelo específico para a entidade interseções, entre outros. Nos modelos por entidades costuma-se usar com variável preditiva apenas o volume de tráfego.

Estimar modelos de previsão de acidentes separadamente para interseções e segmentos viários é interessante, uma vez que a tipologia dos acidentes pode ser diferente nessas entidades. Por exemplo, colisões laterais podem ocorrer mais frequentemente em cruzamentos. Já abalroamentos e colisões frontais podem ser predominantes em segmentos viários.

Nesse estudo são comparadas essas duas abordagens na estimativa de modelos de previsão de ocorrência de acidentes. A técnica de modelagem adotada é a de Modelos Lineares Generalizados. O objetivo do estudo é avaliar se os modelos por entidade apresentam resultados semelhantes ou não aos modelos de múltiplas variáveis.

Este artigo está organizado em 5 seções. A seção 2 apresenta o referencial teórico da evolução dos modelos de previsão de acidentes. Na seção 3 é descrito o método de análise utilizado na criação dos modelos. A seção 4 contém as discussões e resultados obtidos a partir dos modelos criados. Por fim, a seção 5 apresenta as conclusões do estudo.

2. MODELOS DE PREVISÃO DE ACIDENTES

Nessa seção são apresentados estudos envolvendo a modelagem da ocorrência de acidentes viários. É dado destaque as técnicas adotadas pelos seus autores e as principais contribuições desses estudos. Adicionalmente, são revisadas as principais técnicas estatísticas usadas nos modelos de estimativas de ocorrência de acidentes viários e discutidas as premissas básicas que fundamentam a sua adoção.

2.1. Estudos envolvendo modelos de previsão de acidentes

Entre as experiências revisadas sobre desenvolvimento de modelos de previsão de acidentes destaca-se o estudo desenvolvido por Hauer *et al.* (1988) onde foram estimados modelos de previsão de acidentes para interseções semaforizadas em vias urbanas. Nesse

estudo, diferentes modelos foram estimados para cada período do dia (pico da manhã, pico da tarde e diário) e para os diferentes padrões de acidentes registrados. Já os pesquisadores Bonneson e McCoy (1993) estimaram um modelo de previsão para acidentes de trânsito de todos os tipos em cruzamentos de estradas rurais. Ambos os estudos utilizaram o volume de tráfego como única variável independente.

Em outro estudo Greibe (2003), que teve por objetivo desenvolver modelos de previsão de acidentes para interseções e segmentos de via utilizando diversas variáveis relacionadas ao sistema viário. Como resultado desse estudo tem-se que, no modelo de previsão de acidentes em interseções, a variável volume de tráfego apareceu como responsável por mais de 90% da variância explicada. Segundo o autor, este resultado demonstra que no caso particular de interseções, métodos que utilizam apenas o volume de tráfego, como variável explicativa, pode ser tão bom quanto os modelos envolvendo diversas variáveis do sistema viário. Entretanto, no mesmo estudo, nos modelos para os segmentos de via o volume de tráfego explicou aproximadamente 30% da variância do modelo. Mostrando a necessidade de inclusão de outras variáveis para aumentar a capacidade de previsão dos modelos estimados para segmentos de via.

Por outro lado, Hauer (2004) relacionou a quantidade de acidentes por ano em função das variáveis independentes volume diário médio anual, percentual de caminhões, grau de curvatura e comprimento de curvas horizontais, classe das tangentes e do comprimento de curvas verticais, largura da faixa, limite de velocidade, pontos de acesso, a presença e a natureza de estacionamento. Os resultados apontaram como variáveis significativas o volume diário médio anual, número de entradas comerciais de automóveis e limite de velocidade. Em outro estudo Kim *et al.* (2007) propuseram um modelo para relacionar os acidentes ocorridos em interseções de rodovias rurais do Estado da Geórgia (EUA). As características geométricas e ambientais utilizadas para medir a relação foram: tipo de pavimento, tipo de sinalização, largura da plataforma da via, quantidade de movimentos conflitantes, tipo e ângulo de colisão, condições atmosféricas, nível de tráfego, entre outras. Os resultados indicaram que existe uma relação entre as características ambientais e geométricas das interseções e os acidentes ocorridos nestes locais.

Com o objetivo de desenvolver modelos de previsão de acidentes de trânsito utilizando variáveis preditoras que expressam fatores de exposição e de risco e considerando tipos distintos de acidentes em vias arteriais urbanas, Cardoso (2006) utilizou Modelos Lineares Generalizados com distribuição de Poisson. O modelo para acidentes de trânsito do tipo colisão apresentou como variáveis significativas as variáveis volume veicular médio diário, percentual de motocicletas no fluxo de tráfego, velocidade no 85º percentil, comprimento total do segmento, largura total do segmento, condições do pavimento, densidade de vegetação lateral, presença de parada de ônibus, volume de veículos que acessam o trecho da via arterial, volume de veículos em cruzamentos no trecho analisado, obtendo 68% de variabilidade explicada. Já o modelo para atropelamentos apresentou como variáveis significativas o volume veicular médio diário, percentual de travessia de pedestre, fluxo horário de travessia de pedestres ao longo de cada segmento, comprimento total do trecho, condições do pavimento e densidade de vegetação lateral, e obteve 61% da variabilidade explicada.

Grande parte dos estudos revisados apresentam uma abordagem ampla e quantitativa, com o objetivo de relacionar a ocorrência de acidentes com as especificações de diversas variáveis. No entanto, entre as variáveis usadas para prever a ocorrência de acidentes, o volume de tráfego, normalmente, é a variável preditora que apresenta maior influência (Abdel-Aty e Essam Radwan, 2000; Greibe, 2003; Jutak *et al*, 2004 e Hadayeghi *et al*, 2007).

2.2. Técnicas adotadas na modelagem da ocorrência de acidentes viários

Outra questão relevante, além da seleção das variáveis a serem incluídas na modelagem da ocorrência de acidentes de trânsito, é a escolha do tipo do modelo a ser usado. Inicialmente os modelos eram estimados através da técnica de Regressão Linear Múltipla, que é uma técnica estatística utilizada para estudar a relação entre uma variável dependente e diversas variáveis independentes.

O modelo genérico da Regressão Linear Múltipla é dado pela expressão:

$$y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (1)$$

onde:

y_i = estimativa da quantidade de acidentes de trânsito;

$\beta_{0,1,2,\dots,k}$ = parâmetros obtido através da calibração do modelo;

$X_{1i,2i,\dots,ki}$ = características que possivelmente influenciam na ocorrência de acidentes (geométrica da via, volume de tráfego, condições climáticas, etc.);

ε_i = erro aleatório em Y, para a observação i , $i = 1, 2, \dots, n$.

Na aplicação do modelo de Regressão Linear Múltipla existe o pressuposto de que $\varepsilon_i \sim N(0, \sigma^2)$, ou seja, os erros possuem distribuição Normal com média zero e variância constante (homoscedasticidade). No entanto, algumas das suposições dos modelos de Regressão Linear Múltipla não são satisfeitas quando relacionam a quantidade de acidentes com as características geométricas e de tráfego de vias (Jovanis e Chang, 1986; Joshua e Garber, 1990).

Para solucionar os problemas com as suposições da Regressão Linear Múltipla, pesquisas adicionais mostraram que melhores resultados podem ser obtidos quando se utiliza Modelos Lineares Generalizados. Os Modelos Lineares Generalizados podem ser considerados uma extensão dos Modelos Lineares convencionais por dois motivos. Primeiro pela distribuição considerada não precisar ser normal, podendo ser qualquer distribuição da família exponencial. Segundo porque mesmo mantendo a estrutura de linearidade, a função que relaciona o valor esperado e o vetor de covariáveis pode ser qualquer função diferenciável (Miaou *et al.*, 1992; Miaou e Lum, 1993 e Shankar *et al.*, 1995).

Inicialmente os Modelos Lineares Generalizados foram usados assumindo a distribuição de probabilidade de Poisson para relacionar a ocorrência de acidentes às variáveis explicativas (Miaou *et al.*, 1992; Miaou e Lum, 1993 e Shankar *et al.*, 1995). Assim, se Y é uma variável aleatória que representa a quantidade de acidentes de trânsito observados, a probabilidade de ocorrer y acidentes no local i em um dado intervalo de tempo, assumindo a distribuição de Poisson, é representado por:

$$P(Y = y_i) = \frac{\mu_i^{y_i} \exp(-\mu_i)}{y_i!} \quad (2)$$

onde μ_i é o número esperado de acidentes para o local i no mesmo intervalo de tempo. Esse valor esperado de acidentes seguindo a distribuição Poisson pode ser expresso em sua forma exponencial como:

$$\mu_i = \exp(\beta_0 + \sum_{k=1}^n X_i \beta_k) \quad (3)$$

onde:

X_i = características que possivelmente influenciam na ocorrência de acidentes (geométrica da via, volume de tráfego, condições climáticas, etc.);

β_k = parâmetros obtido através da calibração do modelo.

A Equação 3 mostra que o número esperado de acidentes é considerado como uma função exponencial linear de vários atributos presentes no trecho analisado. Para determinar os coeficientes β_k do modelo utiliza-se o método da máxima verossimilhança que é normalmente:

$$L(\beta) = \prod_{i=1}^n f_i(Y_i) = \prod \frac{[\mu(X_i, \beta)]^{y_i} \exp[-\mu(X_i, \beta)]}{y_i!}. \quad (4)$$

Algumas pesquisas mostraram que em determinados casos os acidentes de trânsito não apresentam a dispersão existente na suposição da distribuição de Poisson, média igual à variância. Portanto, os modelos mais recentes são desenvolvidos assumindo distribuição Binomial Negativa dos erros (Miaou *et al.*, 1992; Hauer, 2001; Persaud *et al.*, 2002).

Segundo Jonhson e Kotz (1969), suponha que se tem uma mistura de distribuições Poisson, tal que o valor esperado dessas distribuições varie de acordo com uma distribuição Gamma com função de densidade de probabilidade dada por

$$P(\lambda) = [\beta^\alpha \Gamma(\alpha)]^{-1} \lambda^{\alpha-1} \exp\left[-\frac{\lambda}{\beta}\right] \quad (5)$$

onde, $\lambda > 0$, $\alpha > 0$ e $\beta > 0$.

De forma que,

$$\begin{aligned} P(Y = y_i) &= [\beta^\alpha \Gamma(\alpha)]^{-1} \int_0^\infty \lambda^{\alpha-1} \exp\left[-\frac{\lambda}{\beta}\right] \left[\lambda^{y_i} \exp\left[-\frac{\lambda}{\beta}\right]\right] d\lambda \\ &= [\beta^\alpha \Gamma(\alpha)]^{-1} \int_0^\infty \lambda^{\alpha+y_i-1} \exp[-\lambda(\beta^{-1} + 1)] d\lambda \end{aligned}$$

$$= \binom{\alpha + y_i - 1}{\alpha - 1} \left(\frac{\beta}{\beta+1}\right)^{-1} \left(\frac{1}{\beta+1}\right)^\alpha \quad (6)$$

Então, Y tem uma distribuição Binomial Negativa com parâmetros (α, β) . Assim, Y é uma variável aleatória que representa a quantidade de acidentes de trânsito observados assumindo distribuição Binomial Negativa. E, a probabilidade de ocorrer y acidentes no local i em um dado intervalo de tempo é representado por:

$$= \binom{\alpha + y_i - 1}{\alpha - 1} \left(\frac{\beta}{\beta+1}\right)^{-1} \left(\frac{1}{\beta+1}\right)^\alpha \quad (7)$$

onde: $y = \alpha, \alpha + 1, \dots$ para $0 < \beta \leq 1$.

Na distribuição Binomial Negativa,

$$E(y) = \frac{\alpha}{\beta} \quad (8)$$

e

$$V(y) = \frac{\alpha(1-\beta)}{\beta^2}. \quad (9)$$

A média dos acidentes de trânsito pode ser representada em sua forma logarítmica como sendo uma função linear de uma série de variáveis explicativas do tipo:

$$\ln(\mu) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} \quad (10)$$

As principais diferenças entre os modelos que assumem a distribuição Binomial Negativa dos que assumem a distribuição Poisson está na possibilidade de trabalhar com dados de acidentes que apresentam dispersão elevada visto que a distribuição de Poisson assume média igual à variância (Miaou *et al.*, 1992; Miaou e Lum, 1993 e Shankar *et al.*, 1995).

Um conjunto de dados pode apresentar uma variabilidade maior do que a esperada pelos modelos probabilísticos descritos anteriormente. Esse fenômeno é chamado de superdispersão, podendo ser ocasionada por diversos fatores. Um deles pode ser devido ao excesso de zeros nos dados (Borgatto, 2004). Nestes casos, pode-se propor uma modelagem feita a partir dos chamados modelos inflacionados de zeros, nos quais os dados apresentam um ajuste em duas partes: uma para as contagens nulas e outra para as não-nulas (Lambert,

1992). Esses modelos são conhecidos como regressão de Poisson Inflacionada de Zeros (ZIP) e regressão Binomial Negativa Inflacionada de Zeros (ZINB) e também estão sendo utilizados em estudos de tráfego para prever acidentes de trânsito (Lee e Mannering, 2002; Qin *et al*, 2004 e Kumara e Chin, 2003).

3. MÉTODO

Nesta seção são descritas as etapas e os procedimentos aplicados para o desenvolvimento do estudo. Tanto a definição das etapas do estudo quanto os procedimentos usados para o levantamento dos dados e estimação dos modelos de previsão de acidentes de trânsito foram fundamentadas nas experiências existentes na literatura nacional e internacional revisadas nesse artigo.

O estudo foi realizado em quatro etapas. Na primeira etapa foram definidas a área de estudo e os trechos rodoviários para o levantamentos de dados. A segunda etapa apresenta a escolha das variáveis coletadas e o método para obtenção de cada uma delas. Na terceira etapa foram definidos os modelos a serem comparados, bem como as variáveis que compõem cada um deles e o método de estimação. Por fim, a quarta etapa apresenta como foram comparados os diferentes modelos estimados. A figura 1 apresenta as etapas do trabalho, as quais são descritas nos itens seguintes.

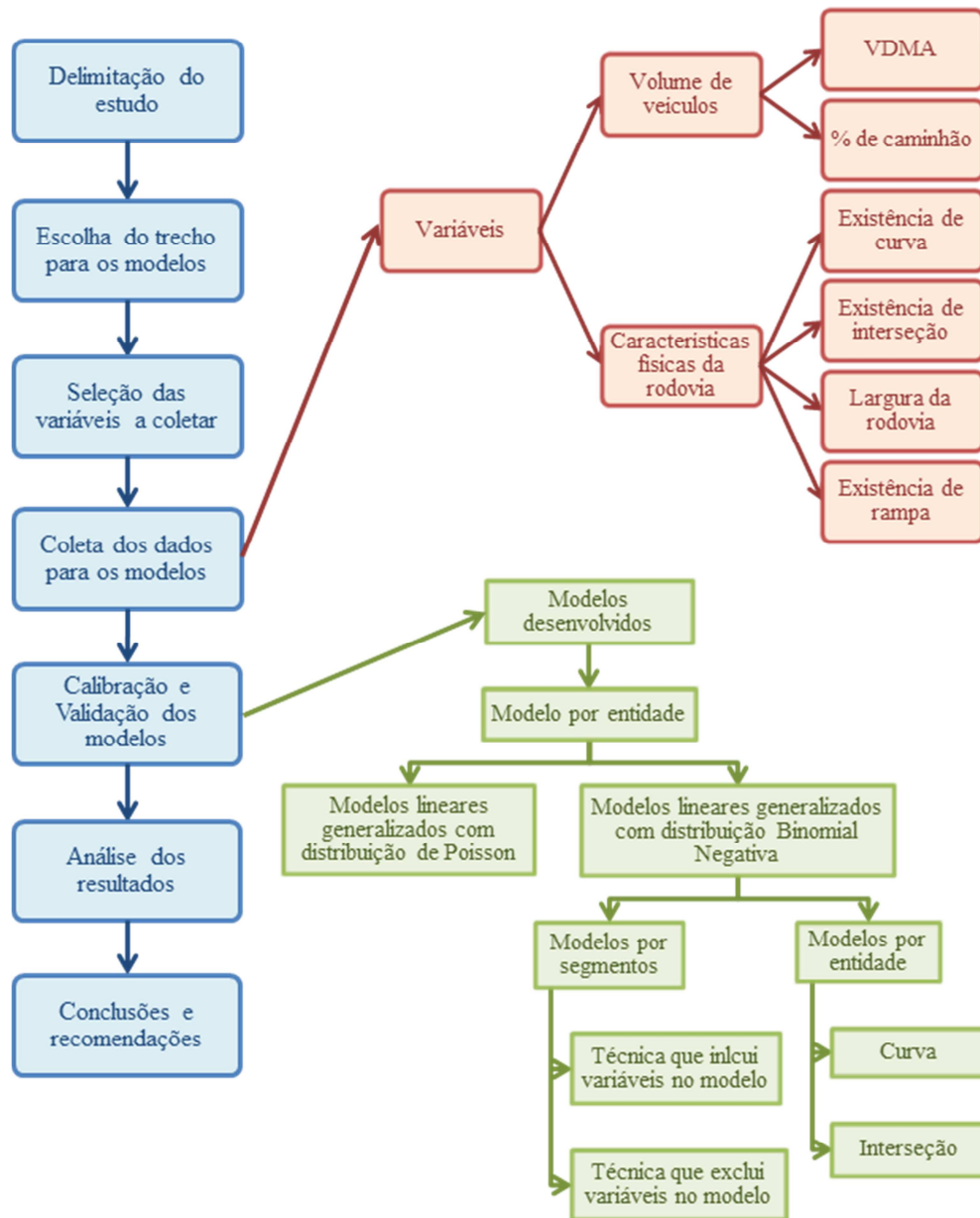


Figura 1: Estrutura da metodologia

3.1. Rodovia de estudo

A primeira etapa do estudo foi a delimitação do mesmo. A premissa adotada foi a seleção de uma rodovia, dentre as rodovias disponíveis, do Estado do Rio Grande do Sul que apresentasse um número expressivo de curvas e interseções. Esses locais costumam apresentar uma maior quantidade de acidentes, quando comparado às segmentos em tangente, e por isso são focos de interesse no estudo da segurança viária. Então, dentre as rodovias

disponíveis para o estudo, optou-se em utilizar a BR-290 na extensão do km 113 ao km 317, totalizando 214 quilômetros que foram divididos em 74 segmentos e contemplam 13 interseções e 27 curvas. Esse trecho está localizado entre as cidades de Eldorado do Sul, que fica próximo da capital (Porto Alegre), e Caçapava do Sul, cidade localizada na região da campanha.

3.2. Descrição das variáveis

A escolha das variáveis a serem utilizadas na calibração dos modelos foi feita através da consideração de dois aspectos: (i) relevância da variável como preditora da ocorrência de acidentes de trânsito e (ii) viabilidade de obtenção dos dados relativos a estas variáveis. Analisando esses aspectos foram selecionadas as seguintes variáveis:

- VDMA: volume diário médio anual;
- Segmento: comprimento do segmento da rodovia (m);
- Curva: assume valor 1 quando o segmento é uma curva e valor 0 caso contrário;
- Interseção: assume valor 1 quando o segmento é uma interseção e valor 0 caso contrário;
- Largura: largura da rodovia (m);
- Curva Vertical: apresenta valor 1 para segmentos em rampa e valor 0 caso contrário;
- % Caminhão: percentual de tráfego de caminhões sobre o tráfego total.

A Tabela 1 a seguir apresenta algumas estatísticas descritivas para os dados coletados.

O volume de tráfego estava disponível na base de dados da empresa que administra a rodovia nos segmentos onde existe praça de pedágio. Para os demais segmentos, foram realizadas contagens no mês de janeiro de 2011. A partir das contagens realizadas, e das contagens das praças de pedágio foram feitas extrapolações para obtenção do VDMA de cada segmento. Por meio da análise dos dados de volume de tráfego em cada segmento analisado foram obtidos também dados para outra variável, o percentual de caminhões na composição do volume de tráfego.

Tabela 1: Estatísticas Descritivas

Base de Dados	Parâmetros	Variáveis				
		Acidentes	VDMA	Largura	Segmento	% Caminhão
Segmentos	Mínimo	0	2.632	7	1	16
	Média	4	5.464	7	3	29
	Máximo	32	10.610	11	11	39
Interseções	Mínimo	0	2.657	7		16
	Média	3	5.464	7		28
	Máximo	10	10.610	7		39
Curvas	Mínimo	0	2.632	7		16
	Média	2	5.164	7		2
	Máximo	11	9.305	11		11

A variável segmento representa o comprimento, em metros, de cada parte da rodovia. Para essa segmentação utilizou-se segmentos do tipo homogêneos e de comprimento variável, onde cada segmento da rodovia é homogêneo em relação às características utilizadas na modelagem.

A variável curva é uma variável do tipo *dummy*, assumindo valor um quando existe curva horizontal e valor zero caso contrário. Da mesma forma, as variáveis interseção e curva vertical assumem valor um quando existe interseção e rampa, respectivamente, e zero caso contrário. Por essas variáveis serem *dummy* não é coerente calcular estatísticas descritivas sobre elas, sendo assim elas não constam na Tabela 1.

Por fim, a variável largura representa a largura total da pista, excluindo o acostamento. Como a rodovia escolhida é de pista simples, quando existe o aumento na largura da mesma é porque o segmento possui terceira faixa já não foram consideradas as superlarguras da rodovia.

Nesse estudo a variável de resposta, ou seja, a que se deseja estimar é a quantidade anual de acidentes. Não é de interesse desse estudo, estimar a quantidade de acidentes de trânsito de algum tipo específico. Assim, essa variável contempla todos os tipos de acidentes que ocorreram. Essa variável foi obtida através do sistema de registro interno da empresa que administra a rodovia. Para todo o tipo de acidente que ocorre no trecho concedido a empresa é acionada e um protocolo de acompanhamento é gerado.

3.3. Calibração dos modelos

A etapa de calibração dos modelos se desenvolveu em dois momentos. Primeiramente foram estimados Modelos Lineares Generalizados para a entidade curva usando distribuição de Poisson (Modelo 1) e usando a distribuição Binomial Negativa (Modelo 2). Esses dois modelos foram estimados para comprovar a tendência de uso da distribuição Binomial Negativa, verificando se esta última resulta em melhores ajustes dos modelos. A decisão de usar Modelos Lineares Generalizados por entidade se deu com base na tendência de seu uso, observada na revisão teórica. A entidade curvas foi usada neste momento por ser no banco de dados disponível com maior número de observações. A partir da confirmação do melhor ajuste do Modelo 2, foram estimados Modelos Linear Generalizado com distribuição Binomial Negativa considerando todo o conjunto de variáveis levantadas nesse estudo.

Em um segundo momento, com a finalidade de comparar duas maneiras distintas de calibração de modelos de previsão de acidentes criou-se dois modelos, utilizando Modelos Lineares Generalizados com distribuição Binomial Negativa e todas as variáveis independentes levantadas no estudo. Um dos modelos (Modelo 3) utiliza o método que em cada etapa inclui-se variáveis no modelo. Esse método consiste em começar o modelo com apenas uma a variável, ou seja, para cada variável preditora ajusta-se um modelo. Após são analisadas as estatísticas de significância para os coeficientes de cada modelo, teste t. Assim, entre os modelos calibrados escolhe-se a variável mais significativa, ou seja, com o menor *p-value*, probabilidade associada ao teste. Essa variável é a primeira a entrar no modelo, e a partir dela são criados modelos com duas variáveis onde as que apresentarem o menor *p-value* continuam no modelo. O processo termina quando não existe mais nenhuma variável significativa para ser colocado no modelo. É importante ressaltar que os coeficientes estimados não podem sofrer grandes alterações entre as etapas, pois se isso ocorrer é sinal de que essas variáveis são correlacionadas e somente uma deve permanecer no modelo (Lord e Persaud, 2000). O outro modelo (Modelo 4) utiliza o método que em cada etapa exclui-se uma variável do modelo. Assim, esse método começa com a inclusão de todas as variáveis independentes de interesse (para os casos de regressão múltipla). Após são analisadas as estatísticas de significância para os coeficientes de cada variável, teste t. Assim, entre as

variáveis não significativas retira-se do modelo a variável que apresentar maior *p-value*, probabilidade associada ao teste. Esse procedimento é realizado até que só restem variáveis significativas no modelo.

Por fim, o último modelo (Modelo 5) utiliza Modelos Lineares Generalizados com distribuição Binomial Negativa para estimar a frequência de acidentes de trânsito em interseções. Assim como no Modelo 1, para previsão de acidentes de trânsito em curvas, o Modelo 5 utiliza apenas a variável VDMA como variável preditora.

Partindo do objetivo do estudo serão calibrados cinco modelos de previsão de acidentes de trânsito, conforme apresentado no Quadro 1. O primeiro modelo, Modelo 1, será comparado com o Modelo 2, a fim de verificar qual o método de modelagem mais adequado para os dados coletados. Os Modelos 3 e 4 serão comparados a fim de identificar com o melhor método de calibração para os dados. Por fim os Modelos 2 e 5 serão comparados com o Modelo 3 para verificar se os modelos específicos para as entidades interseções e curvas apresentam melhores estimativas para a ocorrência e acidentes de trânsito.

Quadro 1. Resumo dos modelos estimados

Modelo	Abordagem	Variáveis consideradas	Distribuição considerada
1	Por entidade - CURVA	VDMA	Poisson
2	Por entidade - CURVA	VDMA	Binomial negativa
3	Para segmentos homogêneos – partindo do modelo estimado apenas com o VDMA foram incluídas, uma a uma, as demais variáveis.	VDMA Segmento Curva	Binomial negativa
4	Para segmentos homogêneos – partindo do modelo estimado com todas as variáveis consideradas foram retiradas as não significativas uma a uma.	Interseção Largura Curva Vertical % Caminhão	Binomial negativa
5	Por entidade - INTERSEÇÃO	VDMA	Binomial negativa

Para a calibração dos modelos utilizou-se o *software* R. A escolha desse software deve-se ao fato de ser um *software* livre e por ter a opção de realizar a modelagem linear generalizada, permitindo a escolha da distribuições estatísticas de Poisson ou Binomial Negativa.

3.4. Comparação dos modelos

Com o objetivo de verificar qual o modelo melhor se ajusta para estimar a frequência de acidentes, muitas medidas escalares têm sido utilizadas nos estudos revisados. Uma medida escalar pode ser útil para comparar modelos e em última instância selecionar um modelo final. Entre essas medidas pode-se citar as estatísticas *Akaike Information Criterion* (AIC), *Bayesian Information Criterion* (BIC) e *Deviance*, descritas a seguir.

Akaike Information Criterion (AIC)

O critério de informação de Akaike (1973) conhecido como AIC é uma estatística frequentemente utilizada para a escolha da melhor especificação de uma equação de regressão. O AIC é definido da seguinte forma:

$$AIC = 2 \times [k - \ln(L)] \quad (11)$$

onde:

k é o número de coeficientes estimados (incluindo a constante);

L é a função de máxima verossimilhança.

O modelo que apresentar o menor valor para AIC é considerado o modelo de melhor ajuste. Cabe ressaltar que esse critério permite ordenar modelos, mesmo que o ajuste não seja razoável.

Bayesian Information Criterion (BIC)

O critério de informação Bayesiano (BIC) foi proposto por Schwarz (1978) e é dado pela expressão:

$$BIC = -2 \ln(L) + k \ln(N) \quad (12)$$

onde:

k é o número de coeficientes estimados (incluindo a constante);

L é a função de máxima verossimilhança;

N é o número total de observações.

O modelo com menor BIC é considerado o melhor modelo ajustado.

Deviance

Para um conjunto de n observações pode ser ajustado um modelo contendo até n parâmetros. O modelo mais simples a ser ajustado é o modelo nulo, que tem um único parâmetro representado por um valor comum a todos os parâmetros. Esse modelo atribui toda a variação entre as respostas ao componente aleatório. No outro extremo, pode-se construir um modelo saturado ou completo no qual tem-se n parâmetros, um para cada observação. Ele atribui toda a variação ao componente sistemático, reproduzindo os próprios dados.

Os modelos nulo e saturado representam os dois extremos. O que se deseja é um modelo intermediário, que possua o menor número possível de parâmetros, mas explique bem a resposta. Dessa forma, busca-se um modelo com p parâmetros linearmente independentes, o qual denomina-se modelo corrente ou modelo sob pesquisa.

Nelder e Wedderburn (1972) propõe como medida de discrepância a chamada *deviance*, que é dada por:

$$S_p = 2(\hat{l}_n - \hat{l}_p), \quad (13)$$

onde \hat{l}_n e \hat{l}_p são os máximos do logaritmo da função de verossimilhança para os modelos saturado e corrente, com n parâmetros para o modelo saturado e p parâmetros para o modelo corrente.

A *deviance* cresce ou decresce de acordo com a entrada das covariáveis no modelo. Quanto maior for o número de covariáveis, menor o valor da *deviance*, mas a complexidade da interpretação dos dados aumenta. Portanto, na prática procura-se uma *deviance* moderada e um modelo que contemple os parâmetros necessários para ajustar os dados.

Para se testar a qualidade do ajuste de um modelo linear generalizado, o valor da *deviance* deve ser comparado a um valor de uma distribuição de probabilidade conhecida. A distribuição da *deviance* é desconhecida, mas, no caso da distribuição de Poisson, pode-se mostrar que ela tem uma distribuição assintótica qui-quadrado com $(n - p)$ graus de liberdade (Nelder e Wedderburn, 1972).

Para comparar os modelos de previsão de acidentes de trânsito calibrados, utilizou-se a estatística AIC. A escolha dessa estatística se deve ao fato de ser a mais utilizada nos estudos revisados e também por penalizar modelos com muitas variáveis. Pois, quanto mais variáveis tiver no modelo, menor é a chance de ele apresentar bons resultados quando aplicado em dados distintos dos utilizados na calibração.

4. DESENVOLVIMENTO E RESULTADOS DOS MODELOS

Nesta seção são apresentados os resultados da estimação dos modelos de previsão de acidentes. São também discutidos os valores obtidos para os coeficientes do modelo e seu impacto na estimativa da ocorrência de acidentes.

4.1. Desenvolvimento dos modelos

O primeiro modelo é calibrado através de Modelos Lineares Generalizados com distribuição de Poisson e estima a frequência de acidentes de trânsito em curvas da rodovia utilizando apenas a variável VDMA como preditora. O Modelo 2 é semelhante ao Modelo 1, porém é calibrado utilizando Modelos Lineares Generalizados com distribuição Binomial Negativa. Assim, as Tabelas 2 e 3 apresentam as estimativas para os coeficientes de cada variável, o erro padrão de cada coeficiente, o valor da estatística z para testar se os coeficientes são significativamente diferente de zero e o valor da probabilidade associado a estatística z, para os Modelos 1 e 2 respectivamente.

Tabela 2: Estimativa dos coeficientes para o Modelo 1

	Estimativa	Erro Padrão	Valor Z	Valor P
Intercepto	-0,7706	225,20	-3,42	0,00
VDMA	0,0001	0,04	3,48	0,00

Tabela 3: Estimativa dos coeficientes para o Modelo 2

	Estimativa	Erro Padrão	Valor Z	Valor P
Intercepto	-0,7898	0,26	-3,09	0,00
VDMA	0,0001	0,00	3,06	0,00

Observando os resultados apresentados nas Tabelas 2 e 3 percebe-se que os dois modelos são válidos, pois os coeficientes estimados são significativos ao nível de 5% de significância, uma vez que o Valor $P < 0,05$. Isso significa que o coeficiente dessas variáveis não é nulo, ou seja, não é igual a zero. Para comparar as duas modelagens utilizou-se o valor da estatística AIC. Então os valores de AIC para os Modelos 1 e 2 foram 361,11 e 357,03 respectivamente. Comparando os valores de AIC conclui-se que a modelagem utilizando Modelos Lineares Generalizados com distribuição Binomial Negativa apresenta o melhor ajuste, sendo assim essa modelagem foi adotada na calibração dos modelos seguintes.

O Modelo 3 estima a frequência de acidentes de trânsito em segmentos da rodovia. Para a calibração do Modelo 3 utilizou-se a técnica que começa com um modelo mais simples e vai se incluindo variáveis baseado na significância do coeficiente da variável. O modelo inicial contém as variáveis preditoras VDMA e Segmento. A variável VDMA foi incluída no modelo inicial por ter apresentado uma importância expressiva nos estudos revisados, mostrando ser essencial. Já a variável Segmento foi incluída no modelo porque a segmentação utilizada apresenta diferentes comprimentos, ou seja, são segmentos homogêneos e não de tamanho fixo. A Tabela 4 apresenta as estimativas dos coeficientes para esse modelo inicial.

Tabela 4: Estimativa dos coeficientes para o Modelo 3 – modelo inicial

	Estimativa	Erro Padrão	Valor Z	Valor P
Intercepto	-0,3035	0,19	-1,62	0,10
VDMA	0,0001	0,00	4,07	0,00
Segmento	0,2795	0,02	12,52	0,00

Observando a Tabela 4 tem-se que os coeficientes estimados para as variáveis VDMA e Segmento são significativos a 5% de significância. A partir desse modelo inicia-se as etapas de inclusão de variáveis. Na primeira etapa são calibrados 5 modelos a partir do modelo inicial. Em cada modelo é incluída mais uma das variáveis preditoras (Curva, Interseção, Largura, Curva Vertical e Caminhão), e a variável que fica no modelo é aquela que apresenta maior significância, ou seja, menor Valor P associado ao a estatística Z. A Tabela 5 apresenta as estimativas dos coeficientes para esses cinco modelos.

Tabela 5: Estimativa dos coeficientes para o Modelo 3 – 1ª etapa

	Estimativa	Erro Padrão	Valor Z	Valor P
Intercepto	0,0744	0,19	0,38	0,70
VDMA	0,0001	0,00	3,88	0,00
Curva	-0,7262	0,14	-5,02	0,00
Segmento	0,2371	0,02	10,81	0,00
Intercepto	-0,6329	0,18	-3,45	0,00
VDMA	0,0001	0,00	4,77	0,00
Interseção	1,1930	0,17	6,87	0,00
Segmento	0,3112	0,02	15,18	0,00
Intercepto	-1,3840	0,59	-2,35	0,02
VDMA	0,0001	0,00	3,67	0,00
Largura	0,1569	0,08	1,93	0,05
Segmento	0,2810	0,02	12,71	0,00
Intercepto	0,5096	0,19	2,67	0,01
VDMA	0,0001	0,00	4,53	0,00
Vertical	-0,9892	0,12	-8,36	0,00
Segmento	0,2520	0,02	13,52	0,00
Intercepto	-0,8677	0,26	-3,34	0,00
VDMA	0,0001	0,00	1,94	0,05
Caminhão	0,0289	0,01	2,96	0,00
Segmento	0,2658	0,02	12,05	0,00

Observando a Tabela 5 tem-se que a variável Curva Vertical foi a que apresentou maior nível de significância, ou seja, menor Valor P ou maior Valor Z em módulo. Assim a variável Vertical é incluída no modelo para a segunda etapa. Na segunda etapa são calibrados quatro modelos incluindo mais uma das variáveis preditoras (Curva, Interseção, Largura e Caminhão). A Tabela 6 apresenta as estimativas dos coeficientes para esses quatro modelos calibrados.

Tabela 6: Estimativa dos coeficientes para o Modelo 3 – 2ª etapa

	Estimativa	Erro Padrão	Valor Z	Valor P
Intercepto	0,7450	0,19	3,86	0,00
VDMA	0,0001	0,00	4,15	0,00
Curva	-0,9023	0,11	-7,94	0,00
Vertical	-0,5489	0,13	-4,21	0,00
Segmento	0,2252	0,02	12,11	0,00
Intercepto	0,1252	0,19	0,65	0,51
VDMA	0,0001	0,00	5,23	0,00
Vertical	-0,8275	0,11	-7,53	0,00
Interseção	0,9218	0,15	6,09	0,00
Segmento	0,2792	0,02	15,83	0,00
Intercepto	-1,1830	0,47	-2,54	0,01
VDMA	0,0001	0,00	3,79	0,00
Vertical	-1,0660	0,11	-9,29	0,00
Largura	0,2554	0,06	3,97	0,00
Segmento	0,2536	0,02	14,16	0,00
Intercepto	0,3459	0,28	1,23	0,22
VDMA	0,0001	0,00	3,52	0,00
Vertical	-0,9549	0,12	-7,75	0,00
Caminhão	0,0070	0,01	0,81	0,42
Segmento	0,2497	0,02	13,44	0,00

Observando a Tabela 6 tem-se que a variável Interseção foi a que apresentou maior nível de significância, ou seja, menor Valor P ou maior Valor Z em módulo. Assim a variável Interseção é incluída no modelo para a terceira etapa. Na terceira etapa são calibrados mais três modelos incluindo mais uma das variáveis preditoras (Curva, Largura e Caminhão). A Tabela 7 apresenta as estimativas dos coeficientes para os três modelos calibrados.

Tabela 7: Estimativa dos coeficientes para o Modelo 3 – 3ª etapa

	Estimativa	Erro Padrão	Valor Z	Valor P
Intercepto	0,3413	0,20	1,70	0,09
VDMA	0,0001	0,00	4,84	0,00
Vertical	-0,7944	0,11	-7,40	0,00
Interseção	0,7846	0,15	5,11	0,00
Curva	-0,3684	0,13	-2,83	0,00
Segmento	0,2574	0,02	14,00	0,00
Intercepto	-1,8040	0,43	-4,18	0,00
VDMA	0,0001	0,00	4,49	0,00
Vertical	-0,9125	0,11	-8,64	0,00
Interseção	0,9767	0,14	6,77	0,00
Largura	0,2862	0,06	4,94	0,00
Segmento	0,2838	0,02	16,72	0,00
Intercepto	-0,0825	0,28	-0,30	0,76
VDMA	0,0001	0,00	4,22	0,00
Vertical	-0,7906	0,11	-7,01	0,00
Interseção	0,9210	0,15	6,16	0,00
Caminhão	0,0085	0,01	1,08	0,28
Segmento	0,2767	0,02	15,84	0,00

Analisando a Tabela 7 tem-se que a variável Largura foi a que apresentou maior nível de significância, ou seja, menor Valor P ou maior Valor Z em módulo. Assim a variável Largura é incluída no modelo para a quarta etapa. Nessa etapa são calibrados mais dois modelos incluindo das variáveis predictoras (Curva e Caminhão). A Tabela 8 apresenta as estimativas dos coeficientes para os dois modelos calibrados.

Tabela 8: Estimativa dos coeficientes para o Modelo 3 – 4ª etapa

	Estimativa	Erro Padrão	Valor Z	Valor P
Intercepto	-1,8130	0,42	-4,34	0,00
VDMA	0,0001	0,00	3,79	0,00
Vertical	-0,8768	0,10	-8,58	0,00
Interseção	0,8050	0,14	5,56	0,00
Largura	0,3292	0,06	5,78	0,00
Curva	-0,4938	0,13	-3,85	0,00
Segmento	0,2570	0,02	14,61	0,00
Intercepto	-1,9040	0,45	-4,21	0,00
VDMA	0,0001	0,00	3,87	0,00
Vertical	-0,8886	0,11	-8,10	0,00
Interseção	0,9754	0,14	6,80	0,00
Largura	0,2808	0,06	4,85	0,00
Caminhão	0,0054	0,01	0,71	0,48
Segmento	0,2822	0,02	16,72	0,00

Observando a Tabela 8 tem-se que a variável Curva apresentou maior nível de significância, ou seja, menor Valor P ou maior Valor Z em módulo. Assim essa variável é incluída no modelo para a quinta e última etapa. Nessa etapa é calibrado mais um modelo incluindo a variável preditora Caminhão. A Tabela 9 apresenta as estimativas dos coeficientes para o modelo calibrado.

Tabela 9: Estimativa dos coeficientes para o Modelo 3 – 5ª etapa

	Estimativa	Erro Padrão	Valor Z	Valor P
Intercepto	-1,9020	0,44	-4,35	0,00
VDMA	0,0001	0,00	3,27	0,00
Vertical	-0,4899	0,13	-3,82	0,00
Interseção	0,8053	0,14	5,58	0,00
Largura	0,3235	0,06	5,65	0,00
Curva	-0,8549	0,11	-8,06	0,00
Caminhão	0,0050	0,01	0,68	0,50
Segmento	0,2560	0,02	14,56	0,00

Analisando os resultados apresentados na Tabela 9 percebe-se que a variável Caminhão, que representa o % de tráfego de caminhão, não é significativa ao nível de 5% de

significância, pois o Valor $P > 0,05$. Isso significa que o coeficiente dessa variável pode ser igual a zero. Então se sugere o modelo da etapa anterior como modelo final. Assim, as estimativas dos coeficientes para o Modelo 3 estão na Tabela 10 a seguir.

Tabela10: Estimativa dos coeficientes para o Modelo 3

	Estimativa	Erro Padrão	Valor Z	Valor P
Intercepto	-1,8130	0,42	-4,34	0,00
VDMA	0,0001	0,00	3,79	0,00
Vertical	-0,8768	0,10	-8,58	0,00
Interseção	0,8050	0,14	5,56	0,00
Largura	0,3292	0,06	5,78	0,00
Curva	-0,4938	0,13	-3,85	0,00
Segmento	0,2570	0,02	14,61	0,00

Para calibrar o Modelo 4, que estima a frequência de acidentes de trânsito em segmentos da rodovia, utilizou-se a técnica que inicia com um modelo completo e vai se excluindo variáveis com base na significância do coeficiente da variável. Assim o Modelo 4 inicia com todas as variáveis independentes de interesse, após são analisadas as estatísticas de significância para os coeficientes de cada variável. Assim, entre as variáveis não significativas retira-se do modelo a variável que apresentar maior *p-value*, probabilidade associada ao teste t. Esse procedimento é realizado até que só restem variáveis significativas no modelo. Após esse procedimento o Modelo 4 apresentou o mesmo formato que o Modelo 3, ou seja, nesse caso os dois procedimentos diferentes resultaram em uma mesma estimativa. O valor de AIC para estes modelos foi de 850,95.

O último modelo calibrado, Modelo 5, estima a frequência de acidentes de trânsito em interseções utilizando somente o VDMA como variável independente. A Tabela 11 apresenta a estimativa dos coeficientes para esse modelo.

Tabela11: Estimativa dos coeficientes para o Modelo 5

	Estimativa	Erro Padrão	Valor Z	Valor P
Intercepto	0,7167	0,26	2,76	0,01
VDMA	0,0001	0,00	2,08	0,04

Analisando os coeficientes do Modelo 5, apresentados na Tabela 11 tem-se que o coeficiente estimado para a variável VDMA é significativo ao nível de significância de 5%. Assim, o modelo é válido e seu AIC é de 175,15.

Para comparar o ajuste dos modelos foi utilizada a estatística AIC, descrita anteriormente. O Quadro 2 apresenta, de forma sintética, a forma funcional de cada modelo estimado e os valores de AIC para cada um deles.

Quadro 2: Modelos Calibrados

Modelo	Fórmula	AIC
1	$Acidentes = e^{(-0,7706+0,0001*VDMA)}$	361,11
2	$Acidentes = e^{(-0,7898+0,0001*VDMA)}$	357,03
3	$Acidentes = e^{(-1,8130+0,0001*VDMA-0,8768*Vertical+0,8050*Interseção+0,3292*Largura-0,4938*Curva+0,2570*Segmento)}$	850,95
4	$Acidentes = e^{(-1,8130+0,0001*VDMA-0,8768*Vertical+0,8050*Interseção+0,3292*Largura-0,4938*Curva+0,2570*Segmento)}$	850,95
5	$Acidentes = e^{(0,7167+0,0001*VDMA)}$	175,15

Comparando os valores de AIC calculados para cada modelo pode-se concluir que o modelo gerado através de Modelos Lineares Generalizados com distribuição de Binomial Negativa (Modelo 2) apresenta resultado superior ao modelo gerado através de Modelos Lineares Generalizados com distribuição de Poisson. Já comparando os valores de AIC entre os modelos por entidade, Modelo 2 e 5, e o modelo para segmentos, Modelo 4, conclui-se que os modelos por entidades apresentam melhores resultados.

Analisando os coeficientes estimados para a variável VDMA nos Modelos 2 e 5 conclui-se que para o aumento de 1000 veículos no VDMA tem-se um aumento de 14,4% na chance de ocorrer acidentes de trânsito em curvas e 9,5% de aumento na chance de ocorrer acidentes de trânsito em interseções. Para os Modelos 3 e 4 existe uma incoerência nos valores dos coeficientes, pois para a existência de curva tem-se uma redução de 39% na chance de ocorrer acidentes de trânsito e, para a existência de rampa tem-se uma redução de 58,4% na chance de ocorrer acidentes de trânsito. E ainda, para o aumento de um metro na largura da pista tem-se um aumento de 39% na chance de ocorrer acidentes de trânsito.

Assim, os resultados desses coeficientes não estão coerentes com a literatura nem com os resultados esperados. Uma das possíveis causas dessa discordância pode ser o fato da base de dados ser insuficiente para uma modelagem precisa da ocorrência de acidentes de trânsito. Nos mesmos modelos para o aumento de 1000 veículos no VDMA tem-se o aumento de 7,9% na chance de ocorrer acidentes de trânsito. Por fim, para a existência de interseções tem-se o aumento de 123,7% na chance de ocorrer acidentes de trânsito.

5. CONCLUSÕES

Este artigo apresenta as técnicas de modelagens utilizadas na previsão de acidentes e as variáveis mais comumente utilizadas. Também são apresentadas duas abordagens de modelagem distintas. Uma que modela os acidentes de trânsito em um segmento rodoviários dividindo-o em vários trechos e associando esses trechos a variáveis explicativas. E a outra que estima modelos específicos por entidades (interseção, curva, tangente) associando a ocorrência de acidentes apenas a variável relativa ao volume de tráfego. O objetivo do artigo é avaliar se os modelos por entidade utilizando apenas a variável volume de tráfego apresentam resultados semelhantes ou não aos modelos de múltiplas variáveis. Complementando o estudo, são calibrados mais dois modelos por entidade considerando como variáveis independentes as variáveis de geometria da via e o volume de tráfego.

A modelagem dos dados dos dados de um segmento de 214 quilômetros da rodovia BR-290 foi realizada por meio de Modelos Lineares Generalizados usando a distribuição Binomial Negativa. Como resultado do estudo foram estimados 5 modelos, que, ao serem comparados entre si, permitiram concluir que os modelos específicos para entidades utilizando como variável explicativa apenas o volume de tráfego apresentaram melhores ajustes que os modelos para segmentos baseados em múltiplas variáveis. Essa conclusão se deu com base na análise do critério de informação de Akaike.

Adicionalmente foram testadas duas formas de construção dos modelos, a primeira onde o modelo inicia com uma forma mais simples e vai se incluindo variáveis baseado na significância do coeficiente da mesma. Essa significância é medida através do *p-value* que é a probabilidade associada ao teste t. O procedimento de inclusão das variáveis termina quando não existem mais variáveis significativas para serem incluídas no modelo. A segunda forma

inicia com um modelo completo e vai se excluindo variáveis com base na significância do coeficiente da variável. Assim o modelo inicia com todas as variáveis independentes de interesse, após são analisadas as estatísticas de significância para os coeficientes de cada variável. Assim, entre as variáveis não significativas retira-se do modelo a variável que apresentar maior *p-value*, probabilidade associada ao teste t. Esse procedimento é realizado até que só restem variáveis significativas no modelo. Após esses dois procedimentos de modelagem os resultados para os modelos foram iguais, ou seja, para os dados coletados e as variáveis selecionadas a maneira como será calibrado o modelo é indiferente.

Com relação aos coeficientes, verificou-se que para os Modelos 3 e 4 existe uma incoerência nos valores dos coeficientes, pois para a existência de curva e rampa tem-se uma redução de 39% e 58,4% respectivamente na chance de ocorrer acidentes de trânsito. Assim, os resultados desses coeficientes não estão coerentes com a literatura nem com os resultados esperados. Uma das possíveis causas dessa discordância pode ser o fato da base de dados ser insuficiente para uma modelagem precisa da ocorrência de acidentes de trânsito.

REFERÊNCIAS

- Abdel-Aty, M.A., Essam Radwan, E.A., 2000. Modeling traffic accident occurrence and involvement. *Accident Analysis and Prevention* 32, 633–642.
- Bonneson, J. and McCoy, P., 1993. Estimation of safety at two-way STOP-controlled intersections on rural highways. *Transportation Research Record* 1401:83-89. Washington, DC: Transportation Research Board.
- Borgatto, A. F. Modelos para proporções com superdispersão e excesso de zeros - Um procedimento Bayesiano. Piracicaba, 2004. 90p. Tese (Doutorado) – Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo.
- Cardoso, G., 2006. Modelos de previsão de acidentes de trânsito em vias arteriais urbanas. Tese de Doutorado, Programa de Pós-graduação em Engenharia de Produção, Universidade federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brasil.

Greibe, P., 2003. Accident prediction models for urban roads. *Accident Analysis and Prevention* 35, 273–285.

Hadayeghi, A., Shalaby, A.S., Persaud, B.N., 2007. Safety Prediction Models: A Proactive Tool for Safety Evaluation in Urban Transportation Planning Applications. Transportation Research Board of the National Academies, January 225 – 236.

Hauer, E., Nolang, J. C. N., Lovell, J., 1988. Estimation of Signalized Intersections. In *Transportation Research Record TRB*, National Research Council, Washington, D.C., 1988, pp. 48-61.

Hauer, E., 2001. Overdispersion in modeling accidents on road sections and in Empirical Bayes estimation. *Accident Analysis and Prevention*, 33, 799-808.

Hauer, E., 2004. Statistical road safety modeling. In: *Proceedings of the 83rd TRB Annual Meeting*, Washington, DC, USA, January 11–15.

Joshua, S., Garber, N., 1990. Estimating truck accident rate and involvement using linear and Poisson regression models. *Transportation Planning and Technology* 15, 41–58.

Jovanis, P.P., and Chang, H.L. 1986. Modeling the relationship of accidents to miles traveled. *Transportation Research Record* 1068, pp. 42–51.

Jutaek, O., Washington, S., Choi, K., 2004. Development of accident prediction models for rural highway intersections. *Transportation Research Record*, 1897(2004), pp. 18-27.

Kim, D.G.; Lee, Y.; Washington, S.; Choi, K. 2007. Modeling crash outcome probabilities at rural intersections: Application of hierarchical binomial logistic models. *Accident Analysis and Prevention*, v39, p.125-134.

Kumara, S.S.P., Chin, H.C., 2003. Modeling accident occurrence at signalized tee intersections with special emphasis on excess zeros. *Traffic Injury Prev.* 3 (4), 53–57.

Kwak, H.; Kim, D. K.; Park, S. H.; Kho, S. Y. 2010. Development of a Safety Performance Function for Korean Expressways. 12th WCTR, July 11-15, 2010 – Lisbon, Portugal.

Lambert, D. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, v.34, n.1, p.1–14, 1992.

Lee, J., Mannering, F.L., 2002. Impact of roadside features on the frequency and severity of run-off-road accidents: an empirical analysis. *Accid. Anal. Prev.* 34 (2), 349–361.

Lord, D., Persaud, B. N., 2000. Accident prediction models with and without trend: Application of the generalized estimating equations procedure. *Transportation Research Record* 1717, 102–108.

Miaou, S., Hu, P., Wright, T., Rathi, A., Davis, S., 1992. Relationship between truck accidents and highway geometric design: a Poisson regression approach. *Transportation Research Record* 1376, 10–18.

Miaou, S., and Lum, H. 1993. Modeling vehicle accident and highway geometric design relationships. *Accident Analysis and Prevention*, 25(6): 689–709.

Persaud, B., Retting, R.A., Lyon, C., 2000. Guidelines for the identification of Hazardous Highway Curves. *Transportation Research Record* 1717, 14–18.

Persaud, B.N., D. Lord, and J. Palminaso (2002) Issues of Calibration and Transferability in Developing Accident Prediction Models for Urban Intersections. *Transportation Research Record* 1784, pp. 57-64.

Qin, X., Ivan, J.N., Ravishankar, N., 2004. Selecting exposure measures in crash rate prediction for two-lane highway segments. *Accid. Anal. Prev.* 36 (2), 183–191.

Shankar, V., Mannering, F., Barfield, W., 1995. Effect of roadway geometrics and environmental factors on rural freeway accident frequencies. *Accident Analysis and Prevention*. 27 (3), 542–555.

4. CONCLUSÕES

Este trabalho trata da ocorrência de acidentes rodoviários. Nesse contexto foi desenvolvido um modelo de previsão de acidentes para um trecho da rodovia BR-290 que está localizado no estado do Rio Grande do Sul. Faz parte desse estudo a investigação dos fatores que estão relacionados à ocorrência de acidentes de trânsito e dos modelos de previsão de acidentes de trânsito mais utilizados. Assim, pode-se dizer que a revisão teórica está dividida em duas partes, uma sobre os tipos de modelagens utilizadas e outra sobre as variáveis mais frequentemente utilizadas e com maior poder de explicação.

Por meio da revisão teórica constatou-se que existem diversos estudos desenvolvidos para investigar a relação entre o projeto geométrico de rodovias e a ocorrência de acidentes de trânsito. Geralmente, nesses estudos, são utilizados modelos matemáticos para prever a ocorrência de acidentes de trânsito. Esses modelos relacionam a quantidade de acidentes de trânsito em várias localidades da rodovia, ao volume de tráfego e às características geométricas e ambientais de cada localidade. Assim, os modelos de previsão de acidentes de trânsito possuem diversas aplicações tais como a avaliação do potencial de segurança de locais da via, a identificação e classificação de localidades perigosas ou com propensão a acidentes, a avaliação da eficácia de medidas de melhoria da segurança e o planejamento da segurança.

A revisão teórica também evidenciou que os modelos de Regressão Linear Múltipla foram inicialmente adotados na estimativa da ocorrência de acidentes. Porém, após perceber

que alguns pressupostos dessa técnica de modelagem estavam sendo violados, outro tipo de modelagem foi sugerida. Essa modelagem sugerida utiliza Modelos Lineares Generalizados que primeiramente, eram aplicados baseados na distribuição de Poisson. Porém, a distribuição de Poisson assume média igual a variância e nos dados de ocorrência de acidentes, normalmente, a variância é superior a média, sendo a utilização desses modelos não eficiente. Então foi sugerida a utilização de Modelos Lineares Generalizados que utilizam a distribuição Binomial Negativa. Dessa forma, ficou atendido o primeiro objetivo específico que previa a identificação das principais técnicas de modelagem.

Em resposta ao objetivo específico da dissertação relacionado a identificação das principais variáveis relacionadas a acidentes de trânsito, foi identificada uma ampla lista de possíveis variáveis utilizadas nos modelos de previsão de acidentes de trânsito revisados. Porém, as variáveis que aparecem com mais frequência são o volume de tráfego, o grau de curvatura, a existência de rampa, a largura de faixa, o comprimento da seção e a quantidade de cruzamentos. Dessas as que mais influenciam na ocorrência de acidentes de trânsito, com base nos estudos revisados, são as variáveis referentes a existência de cruzamentos, comprimento das seções e, principalmente, o volume de tráfego.

Identificadas as variáveis mais usadas na modelagem e a técnica de modelagem mais adequada a previsão de acidentes viários, partiu-se para o levantamento de dados da rodovia de estudo e a modelagem dos mesmos. Nessa etapa, foram adotadas e comparadas duas abordagens. Na primeira, a rodovia estudada foi dividida em trechos e baseado nas características desses trechos (variáveis preditoras) foi estimado o modelo de previsão de acidentes para a rodovia como um todo. Já na segunda abordagem adotada foram gerados modelos específicos para entidades da rodovia (curva e interseção).

Os resultados dessa etapa indicaram que os modelos estimados para entidades específicas e utilizando como variável preditora apenas o volume de tráfego, apresentaram melhor ajuste do que os demais modelos estimados. Dessa forma, foi atendido o terceiro e último objetivo específico que previa a comparação das diferentes abordagens adotadas na modelagem da ocorrência de acidentes em ambiente rodoviários.

Como conclusão geral do estudo obteve-se que a tendência atual na modelagem da ocorrência de acidentes de trânsito é a adoção da técnica de Modelos Lineares Generalizados que utilizam a distribuição Binomial Negativa. Adicionalmente, os modelos específicos para entidades conduzem a melhores resultados do que os modelos estimados para trechos rodoviários.

Para trabalhos futuros sugere-se a discussão das possíveis diferenças nas possibilidades de uso dos modelos gerados por entidades e por trechos rodoviários. O melhor entendimento dessas diferenças permitirá escolher a adequadamente a abordagem a ser utilizada. Sugere-se também ampliar a amostra para verificar a coerência da magnitude e sinal dos coeficientes estimados. Por fim, a utilização de modelos mais avançados como ZIP e ZINB também é sugerida para comparar com os resultados obtidos nesse trabalho.

REFERÊNCIAS

GAO. GOVERNMENT ACCOUNTABILITY OFFICE. Highway Research Continues on a Variety of Factors that Contribute to Motor Vehicle Crash. Washington: United Report to Congressional Requesters, n°. GAO-03-436, 2003.

Miranda, V. e M. Braga (2004) Segurança de trânsito no Brasil: Propostas para as empresas; XIII Congresso Panamericano de Engenharia de Trânsito e Transporte; Anais eletrônicos (CD); Nova York.

Nodari C. T. 2003. Método de avaliação da segurança potencial de segmentos rodoviários rurais de pista simples. 215 p. Teses (Doutorado em Engenharia de Produção). Universidade Federal de Rio Grande do Sul.

World Health Organization (WHO). Global status report on road safety: time for action. Geneva, World Health Organization, 2009. Disponível em: <www.who.int/violence_injury_prevention/road_safety_status/2009> Acesso em: 5 de março de 2011.