



SPOLM 2007

ISSN 2175-6295

Rio de Janeiro- Brasil, 08 e 09 novembro de 2007.

TÉCNICAS ESTATÍSTICAS MULTIVARIADAS E DE ANÁLISE EXPLORATÓRIA DE DADOS ESPACIAIS APLICADAS AO ESTUDO DE VÍTIMAS FATAIS EM ACIDENTES DE TRÂNSITO NO MUNICÍPIO DE BELÉM NO ANO DE 2006

Valéria Sá de Paiva Pereira

Universidade Federal do Pará – UFPA
Rua Augusto Correa, 01, Guamá – 66075-110, Belém-PA
valeriasa87@gmail.com

Giselle Sanjad Ramos

Universidade Federal do Pará – UFPA
Rua Augusto Correa, 01, Guamá – 66075-110, Belém-PA
gisellesanjad@hotmail.com

Edson Marcos Leal Soares Ramos

Universidade Federal do Pará – UFPA
Rua Augusto Correa, 01, Guamá – 66075-110, Belém-PA
edson@ufpa.br

Silvia dos Santos de Almeida

Universidade Federal do Pará – UFPA
Rua Augusto Correa, 01, Guamá – 66075-110, Belém-PA
salmeida@ufpa.br

Adrilayne dos Reis Araújo

Universidade Federal do Pará – UFPA
Rua Augusto Correa, 01, Guamá – 66075-110, Belém-PA
adrilayne@ufpa.br

RESUMO

Este trabalho tem como objetivo utilizar técnicas estatísticas multivariadas e de análise exploratória de dados espaciais para estudar os casos de vítimas fatais, em acidentes de trânsito, ocorridos no município de Belém, no ano de 2006. Quatro técnicas estatísticas distintas foram utilizadas: Estatística Descritiva, Análise de Componentes Principais, Análise de Correspondência e Análise Exploratória dos Dados Espaciais. Dentre os principais resultados pode-se destacar que o *turno da manhã* é o turno que ocorre o maior número de acidentes de trânsito. O dia que mais ocorre acidentes é *sábado*; o principal autor é *homem*; a maior causa das vítimas fatais é por *atropelamento*; o local de maior ocorrência é em *via pública*; as maiores vítimas são os *homens*. Pode-se ver, ainda, que quanto maior a quantidade de *atropelamentos*, maiores serão as chances de vítimas fatais no trânsito no município de Belém. A autoria *ignorada* apresenta forte associação com os dias de *domingo* e *sexta-feira*, com nível de confiança de 82%. Isto é, acidentes de trânsito ocorridos nestes dias, no município de Belém, possuem alta probabilidade de o autor fugir do local do acidente. Finalmente, pode-se ver que 9 (nove) bairros, do município de Belém, apresentaram, no ano de 2006, um quantitativo dos acidentes de trânsito, com vítimas fatais, acima da faixa de valores esperados.

Palavras-Chaves: Análise de Componentes Principais; Análise de Correspondência; Análise Exploratória dos Dados Espaciais (AEDE).

ABSTRACT

The objective of this work is to use multivariate and Exploratory Spatial Data Analysis statistical techniques to study the occurrences of fatal victims involved in traffic accidents, in Belém, 2006. It was used four distinct statistic's techniques: Descriptive Statistics, Principal Components Analysis, Correspondence Analysis and Exploratory Spatial Data Analysis. Among the main results can be highlighted that the greater number of occurrences are *in the morning*. *Saturday* is the day with more accidents occurrences; the local of greater number of occurrences are the *public avenues*; *men* are the majority number of victims. It's also can be noted, as greater as the number of running overs greater will be the chances of fatal victims in transit's accidents in Belém. The *ignored* authorship shows a strong association with Sunday and Friday, on a reliable level of 82%. In other words, accidents with occurrences in these days, in Belém, get a high probability of the author runs away of the accident's local. Finally, can be concluded that 9(nine) quarters, in Belém, presented in 2006, a quantitative of traffic accidents with fatal victims, over the band of expected values

Keywords: Principal Components Analysis; Correspondence Analysis; Exploratory Spatial Data Analysis.

1. INTRODUÇÃO

Segundo o Código de Trânsito Brasileiro – CTB, acidente de trânsito tem por definição “um evento não intencional, envolvendo pelo menos um veículo motorizado ou não, que circula por uma via para trânsito de veículos”. E em caso de falecimento da vítima no local do acidente ou até 30 (trinta) dias após o mesmo, é considerado acidente de trânsito com vítima fatal.

Segundo Pires (2007), diretor do DENATRAN – Departamento Nacional de Trânsito, “A sociedade tem que entender que temos que ter um bom comportamento no trânsito. Se houver certeza da punição, nós vamos ter uma sociedade melhor e mais equilibrada”. A incerteza da punição gera a falta de conscientização de crianças, jovens e, principalmente, dos motoristas. A morte no trânsito não pode ser vista como uma fatalidade. Pois quase sempre é resultado da imprudência.

Diante disso, neste trabalho são utilizadas quatro técnicas estatísticas: Análise Descritiva, Análise de Componentes Principais, Análise de Correspondência e Análise Exploratória dos Dados Espaciais (AEDE), para estudar os casos de vítimas fatais em acidentes de trânsito no município de Belém, no ano de 2006, com a finalidade principal de gerar informações que subsidiem intervenções pelo por público, ou seja, a implantação de ações (ou medidas) que visem mudar a realidade dos acidentes de trânsito no município de Belém.

2. DESENVOLVIMENTO

2.1. DADOS

Os dados utilizados neste estudo são referentes à quantidade de vítimas fatais, em acidentes de trânsito, ocorridos no município de Belém, no ano de 2006, obtidos junto ao Sistema Integrado de Segurança Pública (SISP) do Estado do Pará, por meio da Sub-Gerência de Estatística, do Centro Estratégico Integrado (CEI), órgão pertencente à Secretaria de Segurança Pública do Estado do Pará (SEGUP).

2.2. ANÁLISE DESCRITIVA

A Tabela 1 apresenta a quantidade e o percentual por mês dos acidentes de trânsito, com vítimas fatais, ocorridos no município de Belém, no ano de 2006. Nela pode-se observar que os meses que mais ocorreram acidentes com vítimas fatais no trânsito, foram *junho* com 12,67%, seguido de *novembro* e *dezembro*, ambos com 11,33%.

Tabela 1: Quantidade e Percentual por Mês dos Acidentes de Trânsito, com Vítimas Fatais, Ocorridos no Município de Belém, no Ano de 2006.

Meses	Quantidade	Percentual
Janeiro	15	10,00
Fevereiro	9	6,00
Março	13	8,67
Abril	11	7,33
Mai	7	4,67
Junho	19	12,67
Julho	11	7,33
Agosto	7	4,67
Setembro	9	6,00
Outubro	15	10,00
Novembro	17	11,33
Dezembro	17	11,33
Total	150	100,00

Fonte: Centro Estratégico Integrado-CEI, Set/07

A Figura 1 apresenta a distribuição por turno dos acidentes de trânsito, com vítimas fatais, ocorridos no município de Belém, no ano de 2006. A partir dela pode-se ver que o *turno da manhã* com 28,67%, é o de maior ocorrência. A Figura 2 apresenta a distribuição por dia da semana dos acidentes de trânsito, com vítimas fatais, ocorridos no município de Belém, no ano de 2006. Nela vê-se que o dia que mais ocorre acidentes é *sábado* com 19,34%, seguido de *domingo* e *sexta-feira*, ambos com 18%.

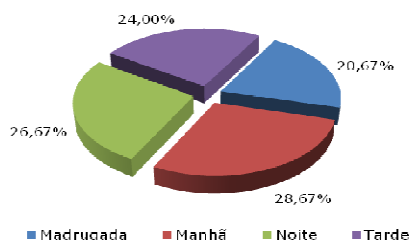


Figura 1: Distribuição por Turno dos Acidentes de Trânsito, com Vítimas Fatais, Ocorridos no Município de Belém, no Ano de 2006.

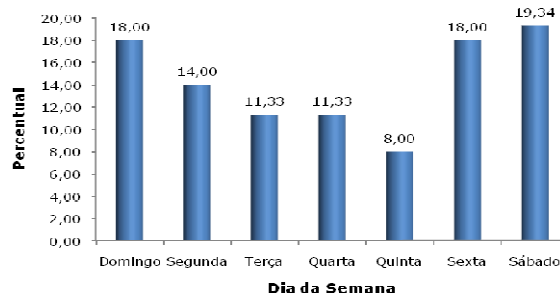


Figura 2: Distribuição por Dia da Semana dos Acidentes de Trânsito, com Vítimas Fatais, Ocorridos no Município de Belém, no Ano de 2006.

A Figura 3 apresenta a distribuição quanto à autoria dos acidentes de trânsito, com vítimas fatais, ocorridos no município de Belém, no ano de 2006. A partir dela pode-se ver que o principal autor é *homem*, com 64%, seguido de *ignorado* e *mulher*, com 33,33% e 2,67% das ocorrências, respectivamente. Vale ressaltar que autor *ignorado* é aquele que foge do local do crime e não é reconhecido. A Figura 4 apresenta a distribuição por motivo dos acidentes de trânsito, com vítimas fatais, ocorridos no município de Belém, no ano de 2006. Nela vê-se que a maior causa das vítimas fatais é por *atropelamento* com 74%, seguido de *colisão* com 17,33% das ocorrências.

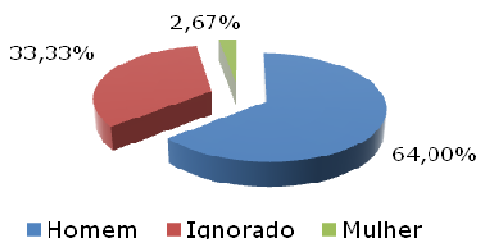


Figura 3: Distribuição por Autoria dos Acidentes de Trânsito, com Vítimas Fatais, Ocorridos no Município de Belém, no Ano de 2006.

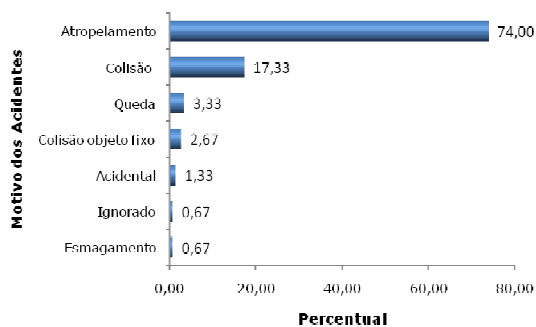


Figura 4: Distribuição por Motivo dos Acidentes de Trânsito, com Vítimas Fatais, Ocorridos no Município de Belém, no Ano de 2006.

A Figura 5 apresenta a distribuição quanto ao local dos acidentes de trânsito, com vítimas fatais, ocorridos no município de Belém, no ano de 2006. A partir dela pode-se ver que o maior percentual é em *via pública* com 92,67% dos casos. A Figura 6 apresenta a distribuição quanto ao meio dos acidentes de trânsito, com vítimas fatais, ocorridos no município de Belém, no ano de 2006. Nela observa-se que tem-se como a maioria das ocorrências, *veículo leve* com 47,34% dos casos, seguido de veículo pesado e ônibus, ambos com 18% das ocorrências.

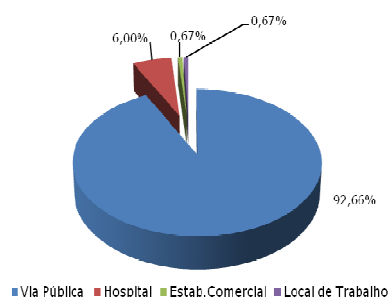


Figura 5: Distribuição por Local dos Acidentes de Trânsito, com Vítimas Fatais, Ocorridos no Município de Belém, no Ano de 2006.

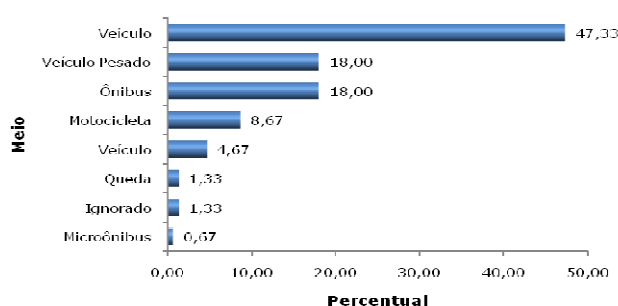


Figura 6: Distribuição por Meio dos Acidentes de Trânsito, com Vítimas Fatais, Ocorridos no Município de Belém, no Ano de 2006.

A Figura 7 apresenta a distribuição por Gênero da Vítima dos acidentes de trânsito, com vítimas fatais, ocorridos no município de Belém, no ano de 2006. Nela vê-se que as maiores vítimas são os *homens* com 72,67% das ocorrências. A Figura 8 apresenta a distribuição por faixa etária da vítima dos acidentes de trânsito, com vítimas fatais, ocorridos no município de Belém, no ano de 2006. Nela vê-se que as maiores vítimas fatais dos acidentes de trânsito estão na faixa etária de 20 a 49 anos com 57,33%.

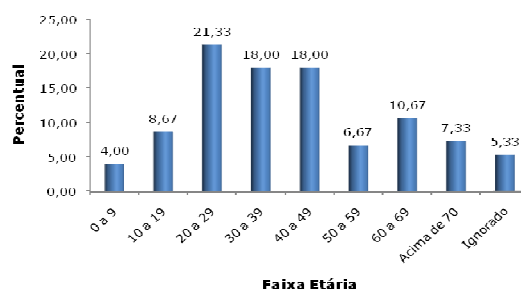
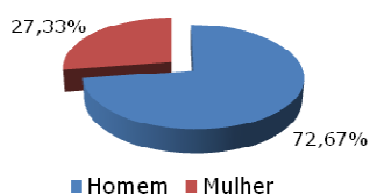


Figura 7: Distribuição por Gênero da Vítima dos Acidentes de Trânsito, com Vítimas Fatais, Ocorridos no Município de Belém, no Ano de 2006.

Figura 8: Distribuição por Faixa Etária da Vítima dos Acidentes de Trânsito, com Vítimas Fatais, Ocorridos no Município de Belém, no Ano de 2006.

2.3. ANÁLISE DE COMPONENTES PRINCIPAIS (ACP)

De acordo com Mingoti (2005), a técnica estatística denominada de Análise de Componentes Principais, popularmente chamada de ACP, foi introduzida por Karl Pearson em 1901 e está fundamentada em Hotelling (1933). Seu objetivo principal é o de explicar a estrutura de variância e covariância de um vetor aleatório, composto de p -variáveis aleatórias, a partir da construção de combinações lineares das variáveis originais. Estas combinações lineares são chamadas de componentes principais e são não correlacionadas entre si.

Quando se têm p -variáveis originais é possível obter-se p componentes principais. No entanto, em geral deseja-se obter "redução do número de variáveis a serem avaliadas e interpretação das combinações lineares construídas", ou seja, a informação contida nas p -variáveis originais é substituída pela informação contida em k ($k < p$) componentes principais não correlacionadas Mingoti (2005). Desta forma, o sistema de variabilidade do vetor aleatório composto das p -variáveis originais é aproximado pelo sistema de variabilidade do vetor aleatório que contém as k componentes principais. A qualidade da aproximação depende do número de componentes mantidas no sistema e pode ser medida a partir da avaliação da proporção de variância total explicada por essas.

O número de componentes principais a serem analisadas dependem do percentual de variância explicada pelos autovalores (λ_i). Um valor de percentual de variância explicada, comumente, aceito é maior ou igual a 70%. Logo, se o percentual de variância explicada da 1ª Componente Principal (Y_1) for maior ou igual a 70%, deve-se utilizar somente esta componente na análise. Porém, se o percentual de variância explicada da 1ª Componente Principal (Y_1) for menor que 70%, devem-se utilizar a 1ª e 2ª Componentes Principais, na análise, se suas somas de percentuais de variância explicada for maior que 70%. Caso contrário deve-se utilizar as 1ª; 2ª e 3ª Componentes Principais, na análise, se suas somas de % de variância explicada for maior que 70%. E assim por diante.

A Tabela 2 apresenta as estatísticas resultantes da aplicação da técnica estatística multivariada de Análise de Componentes Principais aos dados dos acidentes de trânsito, com vítimas fatais, ocorridos no município de Belém, no ano de 2006. Nesta análise foram utilizadas somente as variáveis: *Motivo*, *Vítima* e *Local*, devido terem variâncias significativas para modelar os acidentes de trânsito, com vítimas fatais, ocorridos no município de Belém, no ano de 2006. Assim, observa-se na Tabela 2, que são necessárias a 1ª (Y_1) e a 2ª componente (Y_2) para explicar os acidentes de trânsito, com vítimas fatais, ocorridos no município de Belém, no ano de 2006, pois juntas correspondem a 76,70% variância explicada.

Tabela 2: Estatísticas Resultantes da Aplicação da Análise de Componentes Principais aos Dados Acidentes de Trânsito, com Vítimas Fatais, Ocorridos no Município de Belém, no Ano de 2006.

Estatísticas	Componentes Principais			Soma
	Y_1	Y_2	Y_3	
Autovalores (λ_i)	0,076746	0,052511	0,039315	0,168572
Proporção	45,50	31,20	23,30	
% Acumulado	45,50	76,70	100,00	

As Componentes Principais que modelam os acidentes de trânsito, com vítimas fatais, registrados no município de Belém, no ano de 2006, são

$$Y_1 = 0,971118(\text{motivo}) - 0,13974(\text{vítima}) + 0,193393(\text{local}) \quad (1)$$

e

$$Y_2 = -0,14794(\text{motivo}) + 0,283293(\text{vítima}) + 0,947555(\text{local}) \quad (2)$$

A 1ª Componente Principal (Y_1) é considerada um Índice de Vítimas Fatais, em acidentes de trânsito ocorridos no município de Belém, no ano de 2006 e está fortemente correlacionada ($r = 0,98$) com a variável motivo, que é a variável com a maior variabilidade ($S^2 = 0,074904$). Vale lembrar que nesta variável, conforme Figura 4, a maior quantidade de registros é de *atropelamento*. Desta forma, quanto maior a quantidade de *atropelamentos*, maior serão as chances de vítimas fatais no trânsito no município de Belém.

A 2ª Componente Principal (Y_2), apresentada na Equação 2, pode ser considerada um segundo Índice de Vítimas Fatais em acidentes de trânsito ocorridos no município de Belém, no ano de 2006 e está fortemente correlacionada ($r = 0,95$) com a variável *local* que é a variável com a maior variabilidade (0,052564). Ressalta-se que nesta variável, conforme Figura 5, a maior quantidade de registros é de *via pública*. Isso implica dizer que quanto maior a quantidade de acidentes de trânsito em *via pública*, maior serão as chances de vítimas fatais no trânsito no município de Belém.

2.4. ANÁLISE DE CORRESPONDÊNCIA (AC)

Segundo Cunha Jr. (1997), profissionais de diversas áreas freqüentemente se vêem deparados com um problema crítico de análise e exploração de variáveis categóricas ou qualitativas. Essas questões são realmente muito comuns devido ao formato dos questionários de pesquisa e, em sua grande maioria não têm todas as suas variáveis utilizadas da forma mais proveitosa nas análises e produção de relatórios. A técnica de AC representa uma importante ferramenta para a exploração de dados categóricos.

Esta técnica tem como objetivo principal a representação ótima da estrutura dos dados observados que geralmente são introduzidos sem qualquer tratamento estatístico prévio, provando a flexibilidade e utilidade da técnica. Há dois tipos de análise de correspondência, análise simples (AC) e análise multivariada (ACM).

Para Moscarola (1991) e Lagarde (1995) o gráfico resultante da AC deve ser interpretado como um mapa geográfico, assumindo que as menores distâncias entre as categorias-linha e categorias-coluna representam as mais fortes associações entre elas, enquanto que as maiores distâncias representam dissociações entre as mesmas.

É recomendada a realização do teste do Critério β antes da aplicação da AC, utilizando-se

$$\beta = \frac{\chi^2 - (l-1)(c-1)}{\sqrt{(l-1)(c-1)}}, \quad (3)$$

onde χ^2 = valor do qui-quadrado; l = número de linhas e c = número de colunas. Se $\beta > 3$, as variáveis são ditas dependentes a um risco menor ou igual a 5%. Deve-se observar também a quantidade de informação restituída pelas variáveis deve ser $\geq 70\%$. Também é recomendado que o valor singular dos Autovalores seja maior que 0,20.

2.4.1. RESULTADO DA APLICAÇÃO DA ANÁLISE DE CORRESPONDÊNCIA ÀS VARIÁVEIS TURNO *VERSUS* IDADE

A Tabela 3 apresenta algumas das estatísticas (Autovalores, Total da Inércia, % de Inércia, % Cumulativo da Inércia, Qui-quadrado (χ^2), Graus de Liberdade ($g.l.$), Nível Descritivo (p) e Critério β) resultantes da aplicação da Análise de Correspondência às variáveis Turno e Idade. A partir do valor de $\beta = 3,36$, conclui-se que as variáveis Turno e Idade, são dependentes ao nível de significância menor que 5%, sendo então possível a aplicação da AC. Vê-se também que os percentuais de inércia para os Eixos 1 e 2, restituem

88,24% (62,36% + 25,88%) da informação, assim, a partir deste resultado a análise do gráfico perceptual pode se deter somente a estes dois eixos de componentes principais.

Tabela 3: Estatísticas Resultantes da Análise de Correspondência Aplicada às Variáveis Turno *versus* Idade, dos Acidentes de Trânsito, com Vítimas Fatais, Ocorridos no Município de Belém, no Ano de 2006.

Eixos	Autovalor	% Inércia	% Acumulado	χ^2	<i>g.l.</i>	<i>p</i>	β
1	0,185	62,36	62,36	27,71	24	0,00681	3,36
2	0,077	25,88	88,24	11,50			
3	0,035	11,76	100,00	5,23			

A Figura 9 apresenta gráfico resultante da aplicação da AC as variáveis Turno *versus* Idade, dos acidentes de trânsito, com vítimas fatais, ocorridos no município de Belém, no ano de 2006. A partir dela e da Tabela 4, que a apresenta os resíduos e seus respectivos níveis de confiança, verifica-se que o turno da *manhã* apresenta-se associado às vítimas fatais das faixas etárias: *ignoradas*, de 30 a 39 anos, de 60 a 69 anos e acima de 70 anos ao nível de confiança de 74,02%; 87,43%, 74,02% e 96,97%, respectivamente. O turno da *tarde* também está associado às faixas etárias de 60 a 69 anos e 40 a 49 anos, com nível de confiança de 72,96% e 83,32%, respectivamente. O turno da *noite* está associado à faixa etária *ignorada* e o turno da *madrugada* com a faixa etária de 20 a 29 anos, ao nível de confiança de 98,70%. A proximidade da faixa etária de 10 a 19 anos com o ponto de cruzamento dos Eixos 1 e 2, indica que os acidentes de trânsito, com vítimas fatais, no município de Belém, no ano de 2006, aconteceram em todos os turnos. Ainda, pode-se ver que as faixas etárias de 0 a 9 anos e 50 a 59 anos, não apresentam-se associadas a nenhum turno, isto é, acidentes de trânsito, com vítimas fatais, nessas faixas etárias, ocorrem de forma aleatória segundo o turno.

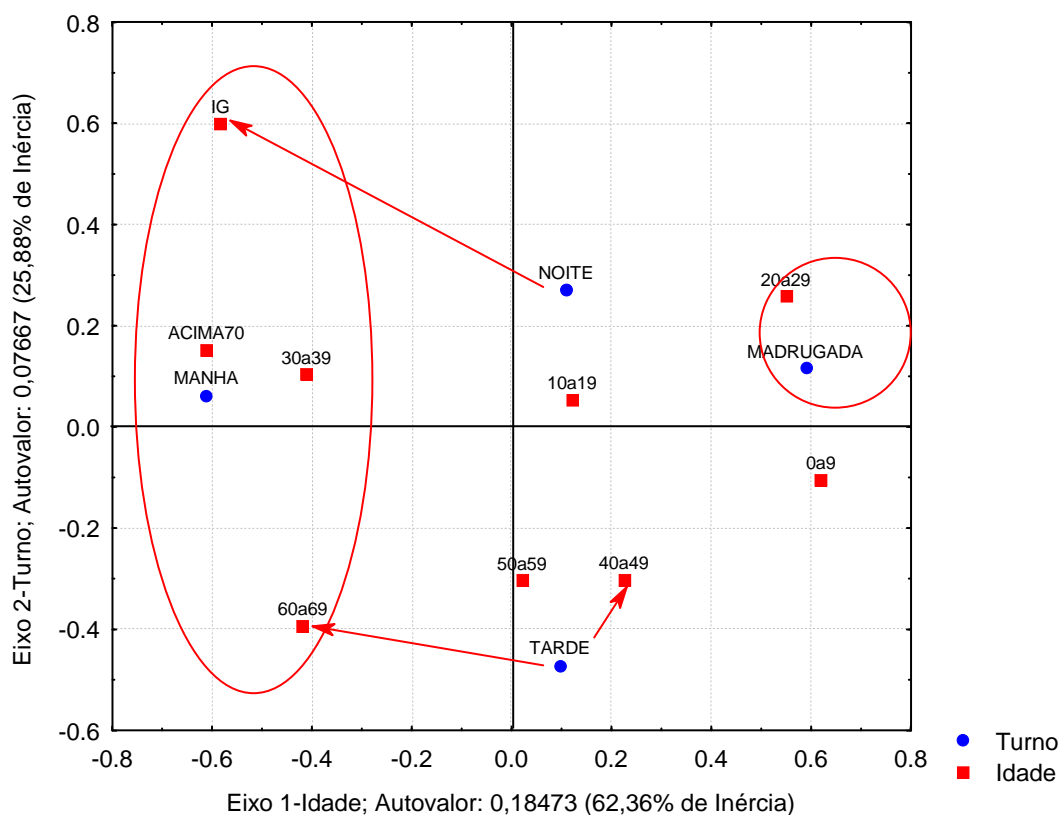


Figura 9: Gráfico Resultante da Aplicação da AC às Variáveis Turno *versus* Idade, dos Acidentes de Trânsito, com Vítimas Fatais, Ocorridos no Município de Belém, no ano de 2006.

Tabela 4. Resíduos e Seus Respectivos Níveis de Confiança Resultante da Aplicação da AC às Variáveis Turno *versus* Idade, dos Acidentes de Trânsito, com Vítimas Fatais, Ocorridos no Município de Belém, no ano de 2006.

Faixa Etária	Turno			
	Manhã	Tarde	Madrugada	Noite
0 a 9 anos	0,00	35,93	50,51	24,82
10 a 19 anos	0,00	0,00	15,16	22,55
20 a 29 anos	0,00	0,00	98,70	60,16
30 a 39 anos	87,43	0,00	0,00	23,44
40 a 49 anos	0,00	83,34	45,22	0,00
50 a 59 anos	0,00	69,83	0,00	16,17
60 a 69 anos	74,02	72,97	0,00	0,00
Acima de 70 anos	96,97	0,00	0,00	0,00
Ignorado	74,02	0,00	0,00	79,88

2.4.2. RESULTADO DA APLICAÇÃO DA ANÁLISE DE CORRESPONDÊNCIA ÀS VARIÁVEIS DIA DA SEMANA *VERSUS* AUTORIA

A Tabela 5 apresenta algumas das estatísticas (Autovalores, Total da Inércia, % de Inércia, % Cumulativo da Inércia, Qui-quadrado (χ^2), Graus de Liberdade (*g.l.*), Nível Descritivo (*p*) e Critério β) resultantes da aplicação da Análise de Correspondência às variáveis Dia da Semana e Autoria. A partir do valor de $\beta = 3,49$, conclui-se que as variáveis Dia da Semana e Autoria, são dependentes ao nível de significância menor que 5%, sendo então possível a aplicação da AC. Vê-se ainda, que os percentuais de inércia para os Eixos 1 e 2, restituem 100,00% (60,65% + 39,35%) da informação. Assim, a partir deste resultado a análise do gráfico perceptual pode se deter somente a estes dois eixos de componentes principais.

Tabela 5: Estatísticas Resultantes da Análise de Correspondência Aplicada às Variáveis Dia da Semana *versus* Autoria, dos Acidentes de Trânsito, com Vítimas Fatais, Ocorridos no Município de Belém, no Ano de 2006.

Eixos	Autovalor	% Inércia	% Acumulado	χ^2	<i>g.l.</i>	<i>p</i>	β
1	0,097	60,65	60,65	14,61	12	0,01983	3,49
2	0,063	39,35	100,00	9,48			

A Figura 10 apresenta gráfico resultante da aplicação da AC às variáveis Dia da Semana *versus* Autoria, dos acidentes de trânsito, com vítimas fatais, ocorridos no município de Belém, no ano de 2006. A partir dela e da Tabela 6, que a apresenta os resíduos e seus respectivos níveis de confiança, verifica-se que a autoria *ignorada* apresenta uma forte associação com os dias de *domingo* e *sexta-feira*, com nível de confiança de 81,76%. Isto é, acidentes de trânsito ocorridos nestes dias, no município de Belém, possuem alta probabilidade de o autor fugir do local do acidente. Autoria *homem* apresentou associação de 86,50% confiança com o *sábado*. E Autoria *mulher*, associou-se com *terça-feira* e *quinta-feira*, com nível de confiança 97,34% e 77,07%, respectivamente.

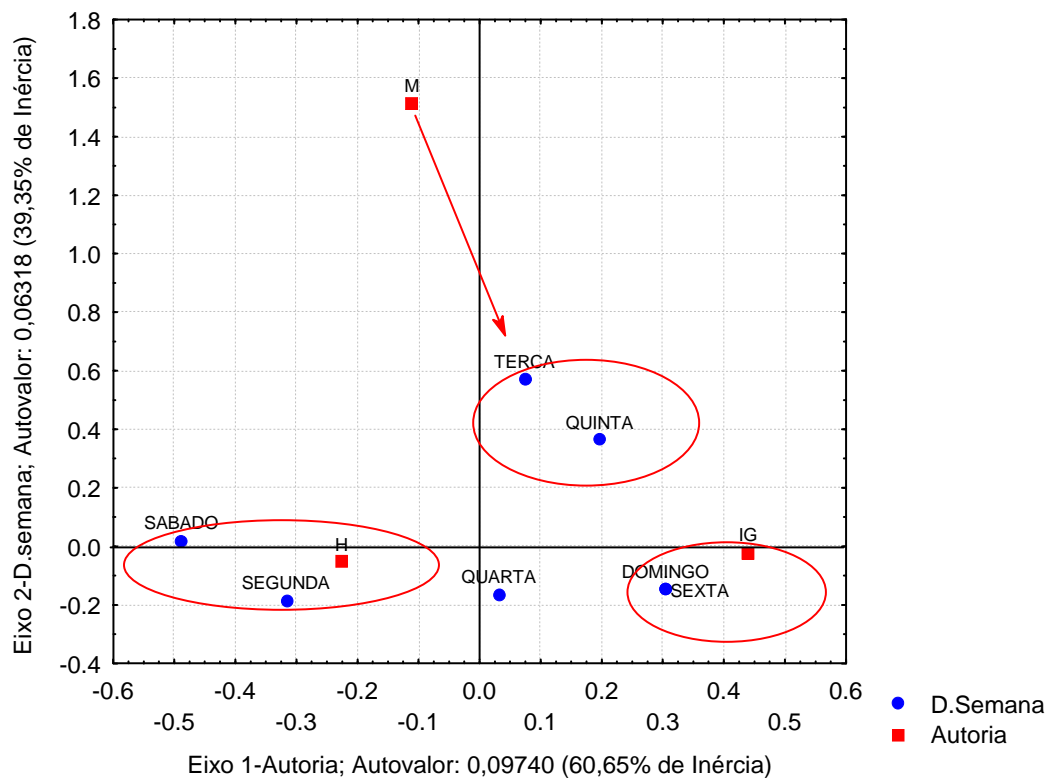


Figura 10 - Gráfico Resultante da Aplicação da AC às Variáveis Dia da Semana *versus* Autoria, dos Acidentes de Trânsito, com Vítimas Fatais, Registrados no Município de Belém, no ano de 2006.

Tabela 6. Resíduos e Seus Respectivos Níveis de Confiança Resultante da Aplicação da AC às Variáveis Dia da Semana *versus* Autoria, dos Acidentes de Trânsito, com Vítimas Fatais, Ocorridos no Município de Belém, no ano de 2006.

Dia da Semana	Autoria		
	Homem	Mulher	Ignorado
Domingo	0,00	0,00	81,76
Segundo	66,85	0,00	0,00
Terça	0,00	97,84	11,14
Quarta	2,90	0,00	11,14
Quinta	0,00	77,07	38,29
Sexta	0,00	0,00	81,76
Sábado	86,50	20,34	0,00

2.5. ANÁLISE EXPLORATÓRIA DE DADOS ESPACIAIS (AEDE)

Análise exploratória dos dados espaciais é um conjunto de técnicas utilizadas para descrever distribuições espaciais de variáveis, traçar padrões de correlação espacial ou apontar a ocorrência de clusters, ou mesmo apontar outliers (Anselin, 1998). Com o intuito, identificar padrões espaciais do comportamento, por exemplo, de acidentes com vítimas fatais de trânsito. E assim, produzir informações espaciais adequadas ao georeferenciamento, para auxiliar a tomada de decisão no combate e prevenção deste tipo de crime.

A análise exploratória de dados espaciais (AEDE) utiliza as ferramentas de visualização espaciais e a análise de autocorrelação espacial global e local para a interpretação das informações. A AEDE tem por princípio que os fenômenos espaciais tendem a estar

correlacionados com outros que se acham geograficamente próximos. Para capturar a ocorrência de tais associações, utiliza-se o indicador global de associação espacial.

A AEDE não supõe normalidade dos dados, pois na natureza nem tudo é normal. O estudo é feito principalmente para a avaliação da associação espacial das variáveis no espaço, ou seja, deseja-se verificar na distribuição espacial dos dados a presença ou não de um padrão aleatório. Neste trabalho, as ferramentas da AEDE implementadas utilizam o software GeoDa (Anselin, 2004).

2.5.1. CARTOGRAMA

O cartograma é um mapa onde o plano é o tamanho da área proporcional a uma determinada variável. A colocação é alinhada tão perto quanto possível ao local original das unidades. O tamanho (área) dos círculos é proporcional ao valor da variável selecionada. O cartograma realça em um mapa, informações adicionais sobre algum tipo de valor específico da variável, como negativos (cor preta), zeros (cor branca) e *outliers* superiores (cor vermelha) e *outliers* inferiores (cor azul), e verde é o *default* do programa para dados considerados dentro faixa de valores esperados. *Outliers* são valores que ocorrem fora da faixa de valores esperados. Assim, uma informação é classificada como um *outlier* quando medir um determinado múltiplo da amplitude interquartílica (a diferença entre os 75% e 25% das observações). Desse modo, os valores abaixo de 25% (limite inferior) ou acima de 75% (limite superior) são considerados *outliers*. Conforme Figura 11, verifica-se os múltiplos utilizados para a definição do limite superior e limite inferior são $\pm 1,5$ e $3,0$ vezes a amplitude interquartílica.

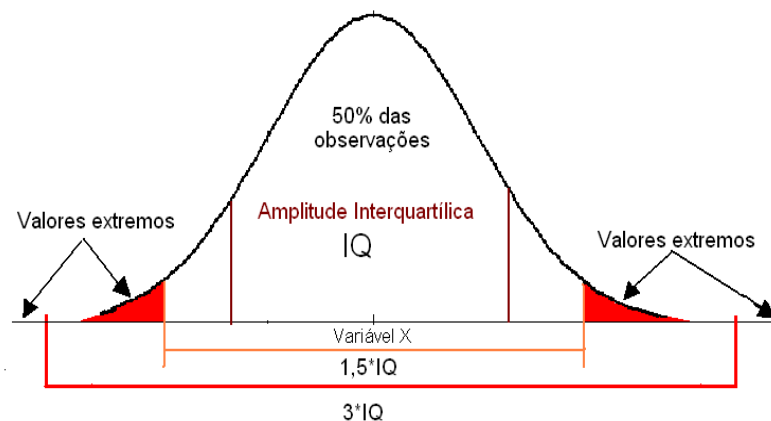


Figura 11: Identificação do limite inferior e limite superior para os valores extremos.

A Figura 12 apresenta o cartograma circular do quantitativo dos acidentes de trânsito, com vítimas fatais, ocorridos no município de Belém, no ano de 2006, por bairro, com limites de $\pm 1,5$ vezes a amplitude interquartílica. A partir dela, verifica-se a presença de *outliers* superiores de vítimas fatais de acidentes de trânsito no município de Belém, no ano de 2006. Ou seja, 9 (nove) bairros, do município de Belém, apresentaram, no ano de 2006, um quantitativo dos acidentes de trânsito, com vítimas fatais, acima da faixa de valores esperados. Estes bairros são: *Guamá, Campina, São Brás, Umarizal, Marco, Telégrafo, Pedreira, Sacramento e Marambaia*.

A Figura 13 apresenta o cartograma circular do quantitativo dos acidentes de trânsito, com vítimas fatais, ocorridos no município de Belém, no ano de 2006, por bairro, com limites de $\pm 3,0$ vezes a amplitude interquartílica. A partir dela, verifica-se a presença de *outliers* superiores de vítimas fatais de acidentes de trânsito no município de Belém, no ano de 2006. Nela, 4 (quatro) bairros, do município de Belém, apresentaram, no ano de 2006, um quantitativo dos acidentes de trânsito, com vítimas fatais, acima da faixa de valores esperados. Estes bairros são: *Marco, Pedreira, Sacramento e Marambaia*.

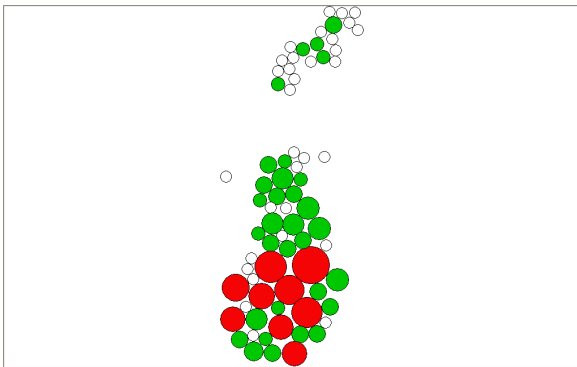


Figura 12: Cartograma Circular do Quantitativo dos Acidentes de Trânsito, com Vítimas Fatais, Ocorridos no Município de Belém, no ano de 2006, por Bairro, com Limites de $\pm 1,5$ vezes a Amplitude Interquartílica.

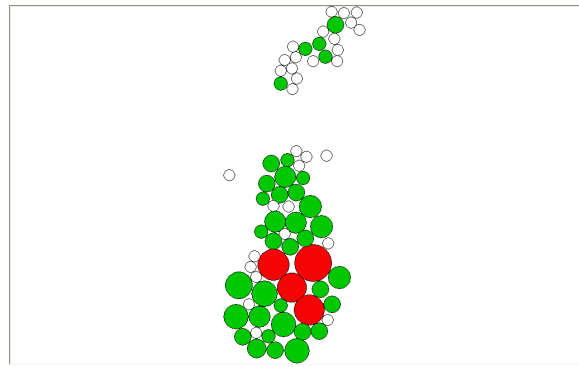


Figura 13: Cartograma Circular do Quantitativo dos Acidentes de Trânsito, com Vítimas Fatais, Ocorridos no Município de Belém, no ano de 2006, por Bairro, com Limites de $\pm 3,0$ vezes a Amplitude Interquartílica.

2.5.2. BOX MAP

O Box Map é uma versão mais completa de um mapa quartil, no qual os *outliers* no primeiro e quarto quartil são realçados separadamente. Além de mostrar os valores extremos, o mapa sugere um agrupamento espacial dos valores. No Box Map, cada objeto é classificado conforme sua posição em relação aos quadrantes do gráfico de Moran. Da mesma forma como no cartograma, a definição dos *outliers* é baseada na formulação de limites inferiores e superiores, que correspondem a múltiplos da amplitude interquartílica, ou seja, a 1,5 e 3,0 vezes a amplitude interquartílica (Tavares, 2007).

A Figura 14 apresenta o box map do quantitativo dos acidentes de trânsito, com vítimas fatais, ocorridos no município de Belém, no ano de 2006, por bairro, com limites de ± 1.5 vezes a amplitude interquartílica. A partir dela, verifica-se a presença de *outliers* superiores de vítimas fatais de acidentes de trânsito no município de Belém, no ano de 2006. Ou seja, 9 (nove) bairros, do município de Belém, apresentaram, no ano de 2006, um quantitativo dos acidentes de trânsito, com vítimas fatais, acima da faixa de valores esperados. Estes bairros são: *Guamá, Campina, São Brás, Umarizal, Marco, Telégrafo, Pedreira, Sacramento e Marambaia*.

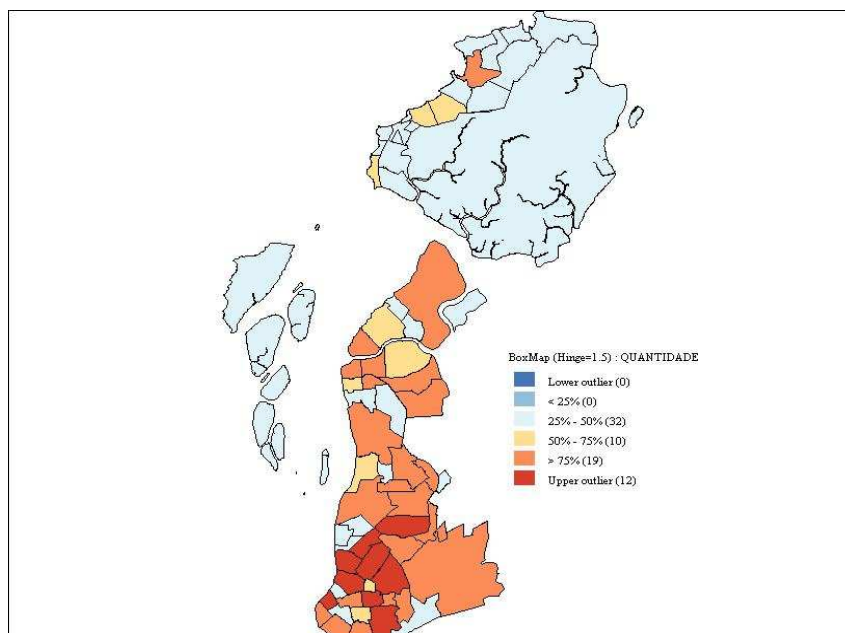


Figura 14: Box Map do Quantitativo dos Acidentes de Trânsito, com Vítimas Fatais, Ocorridos no Município de Belém, no Ano de 2006, por Bairro, com Limites de ± 1.5 vezes a Amplitude Interquartílica.

A Figura 15 apresenta o box map do quantitativo dos acidentes de trânsito, com vítimas fatais, ocorridos no município de Belém, no ano de 2006, por bairro, com limites de $\pm 3,0$ vezes a amplitude interquartílica. A partir dela, verifica-se a presença de *outliers* superiores de vítimas fatais de acidentes de trânsito no município de Belém, no ano de 2006. Ou seja, 4 (quatro) bairros, do município de Belém, apresentaram, no ano de 2006, um quantitativo dos acidentes de trânsito, com vítimas fatais, acima da faixa de valores esperados. Estes bairros são: *Marco, Pedreira, Sacramento e Marambaia*.

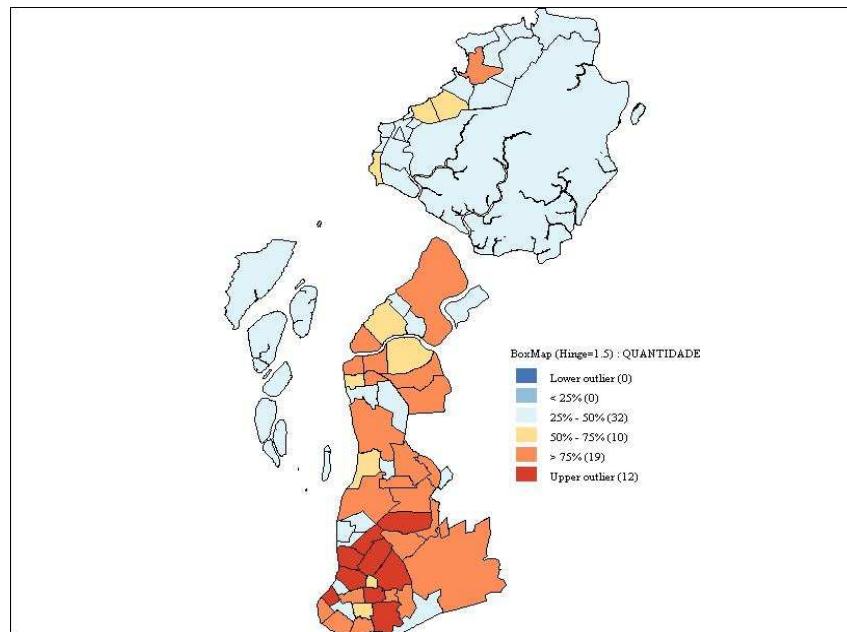


Figura 13: Box Map do Quantitativo dos Acidentes de Trânsito, com Vítimas Fatais, Ocorridos no Município de Belém, no Ano de 2006, por Bairro, com Limites de ± 3.0 vezes a Amplitude Interquartílica.

3. CONCLUSÃO

Este trabalho teve como objetivo utilizar técnicas estatísticas multivariadas e de análise exploratória de dados espaciais para estudar os casos de vítimas fatais, em acidentes de trânsito, ocorridos no município de Belém, no ano de 2006. A partir da Análise Descritiva dos dados pode-se ver que o turno da *manhã* com 28,67%, é o de maior ocorrência. O dia que mais ocorre acidentes é *sábado* com 19,34%, seguido de *domingo* e *sexta-feira*, ambos com 18,00%. O principal autor é *homem*, com 64,00%, seguido de *ignorado* e *mulher*, com 33,33% e 2,67% das ocorrências, respectivamente. A maior causa das vítimas fatais é por *atropelamento* com 74,00%, seguido de *colisão* com 17,33% das ocorrências. Quanto ao local de ocorrência dos acidentes de trânsito, pôde-se ver que o maior percentual é em *via pública* com 92,66% dos casos. As maiores vítimas são *homens* com 72,67% das ocorrências. E quanto ao meio, tem-se como a maioria, *veículo leve* com 47,33% dos casos.

A partir da técnica estatística multivariada de Componentes Principais pode-se obter dois modelos (componentes) capazes de explicar o comportamento acidentes de trânsito, com vítimas fatais, ocorridos no município de Belém, no ano de 2006. A 1ª Componente Principal (Y_1) é fortemente correlacionada ($r = 0,98$) com a variável motivo, que é a variável com a maior variabilidade ($S^2 = 0,074904$). Vale lembrar que nesta variável, conforme Figura 4, a maior quantidade de registros é de *atropelamento*. Desta forma, quanto maior a quantidade de

atropelamentos, maior serão as chances de vítimas fatais no trânsito no município de Belém. A 2ª Componente Principal (Y_2) é fortemente correlacionada ($r = 0,95$) com a variável local que é a variável com a maior variabilidade (0,052564). Ressalta-se que nesta variável, a maior quantidade de registros é de *via pública*. Isso implica dizer que quanto maior a quantidade de acidentes de trânsito em via pública, maior serão as chances de vítimas fatais no trânsito no município de Belém.

A partir da técnica estatística multivariada da Análise de Correspondência verificou-se que o turno da *manhã* apresenta-se associado às vítimas fatais das faixas etárias: *ignoradas*, de 30 a 39 anos e acima de 60 anos, ao nível de confiança de 74,02%; 87,43% e W%, respectivamente. O turno da *tarde* também está associado às faixas etárias de 60 a 69 anos e 40 a 49 anos, com nível de confiança de 72,96% e 83,32%, respectivamente. O turno da *noite* está associado à faixa etária *ignorada* e o turno da *madrugada* com a faixa etária de 20 a 29 anos, ao nível de confiança de 98,70%. As faixas etárias de 0 a 9 anos e 50 a 59 anos, não apresentam-se associadas a nenhum turno, isto é, acidentes de trânsito, com vítimas fatais, nessas faixas etárias, ocorrem de forma aleatória segundo o turno. A autoria *ignorada* apresenta uma forte associação com os dias de *domingo* e *sexta-feira*, com nível de confiança de 81,76%. Isto é, acidentes de trânsito ocorridos nestes dias, no município de Belém, possuem alta probabilidade do autor fugir do local do acidente. Autoria *homem* apresentou associação de 86,50% confiança com o *sábado* e autoria *mulher*, associou-se com *terça-feira* e *quinta-feira*, com nível de confiança 97,34% e 77,07%, respectivamente.

Finalmente, a partir da técnica estatística de Análise Exploratória de Dados Espaciais pode-se verificar no Cartograma Circular e no Box Map, com limites de ± 1.5 vezes a amplitude interquartílica, a presença de *outliers* superiores de vítimas fatais de acidentes de trânsito no município de Belém, no ano de 2006. Ou seja, 9 (nove) bairros, do município de Belém, apresentaram, no ano de 2006, um quantitativo dos acidentes de trânsito, com vítimas fatais, acima da faixa de valores esperados. Estes bairros são: *Guamá*, *Campina*, *São Brás*, *Umarizal*, *Marco*, *Telégrafo*, *Pedreira*, *Sacramenta* e *Marambaia*. E no Cartograma Circular e no Box Map, com limites de ± 3.0 vezes a amplitude interquartílica, verificou-se que 4 (quatro) bairros, do município de Belém, apresentaram, no ano de 2006, um quantitativo dos acidentes de trânsito, com vítimas fatais, acima da faixa de valores esperados. Estes bairros são: *Marco*, *Pedreira*, *Sacramenta* e *Marambaia*.

4. REFERENCIAS BIBLIOGRÁFICAS

- ANSELIN, Luc.. Exploratory Spatial Data Analysis in a Geocomputacional Environment. In: Longley P. A.; Brooks; S. M.; MCDONNELL, R.; MACMILLIAN; B. Geocomputation a primer. Chichester: John Willey & Sons Ltd, 1998, p.77-94.
- ANSELIN, Luc.. Geoda: geodata analysis software. Illinois, 2004.
- CUNHA Jr.; Marcus V. M. Análise Multidimensional de Dados Categóricos: A aplicação das análises de correspondência simples e múltipla em marketing e sua integração com técnicas de análise de dados quantitativos. Rio Grande do Sul: UFRGS, 1997.
- CRUZ, A. T. Análise Exploratória Espacial dos Crimes de Roubo a Transeunte no Distrito de Icoaraci no ano de 2006, Belém, 2007.
- LAGARDE; J. Initiation `a L'Analyse des Données. Paris: Dunod, 1995.
- LIMA, A. G. Um Estudo sobre Vítimas Fatais em Acidentes de Trânsito na Região Metropolitana de Belém no Ano de 2003, Belém, 2004.
- MINGOTI, S. A. Análise de Dados Através de Métodos de Estatística Multivariada: Uma Abordagem Aplicada, Belo Horizonte: UFMG, 2005.
- MOSCAROLA, J. Enquêtes et Analysis de Données. Vuibert. Paris, 1991.
- MENDONÇA, F. L. S. Análise de Componentes Principais Aplicadas à Identificação de Deficiências de Aprendizagem nos Alunos de Recuperação em Matemática da 5ª Série do Município de Uruará – Pará, Santarém, 2007.
- PIRES, Aílton Brasiliense. Departamento Nacional de Trânsito. Disponível em <<http://www.denatran.pa.gov.br>> Acesso em: Setembro de 2007.