

# ANÁLISE FATORIAL E ESPACIAL EM DADOS DE CRIMES E ACIDENTES DE TRÂNSITO DO ESTADO DE MATO GROSSO

Damallys de Assis Oliveira<sup>1</sup>  
Gabriela Ferreira Rosa<sup>1</sup>  
Monique Melania Kempa<sup>1</sup>  
Névio Lotufo Neto<sup>1</sup>  
Kuang Hongyu<sup>2</sup>

**RESUMO:** Os acidentes de trânsito são um dos problemas que mais afetam a sociedade hoje em dia, a quantidade de feridos e mortos são números que só crescem no decorrer dos anos. O presente estudo teve como objetivo analisar dados de crimes e acidentes de trânsito em municípios do estado de Mato Grosso, por meio de técnicas estatísticas. Para tanto aplicou-se técnicas multivariadas, como a análise de componentes principais e análise fatorial, respectivamente, afim de redução, resumo e produção dos *escores*, gerando variáveis criminais mais significativas, combinadas com a estatística espacial, para ilustrar a situação espaço-temporal. As técnicas multivariadas foram eficientes, pois conseguiram reduzir as oito variáveis estudadas compostas por diferentes tipos de crimes para dois fatores e a abordagem espacial identificou regiões críticas em cada fator, porém não observou dependência espacial.

**Palavras-chave:** criminalidade, acidentes de trânsito, análise de componentes principais, análise fatorial, estatística espacial.

## FACTORIAL ANALYSIS AND SPATIAL IN CRIMES AND TRAFFIC ACCIDENTS DATA OF THE STATE OF MATO GROSSO

**ABSTRACT:** Traffic accidents are one of the problems that most affect society today, the number of wounded and dead are numbers that only grow over the years. The present study had as objective to analyze data of crimes and traffic accidents in municipalities of the state of Mato Grosso, through statistical techniques. For this purpose, multivariate techniques such as principal component analysis and factor analysis were used to reduce, summarize and produce scores, generating more significant criminal variables, combined with spatial statistics, to illustrate the space-temporal situation. The multivariate techniques were efficient because they were able to reduce the eight studied variables composed of different types of crimes for two factors and the spatial approach identified critical regions in each factor, but did not observe spatial dependence.

**Keywords:** crime, traffic accidents, principal component analysis, factorial analysis, spatial statistical analysis.

---

<sup>1</sup>Graduandos em Estatística da Universidade Federal de Mato Grosso, Campus Universitário de Cuiabá - damallys@gmail.com; gabrielarosa5@hotmail.com; moniquemkempa@gmail.com; neviolotufo@gmail.com.

<sup>2</sup>Prof. Dr. Adjunto do Departamento de Estatística. Universidade Federal de Mato Grosso, Av. Fernando Corrêa da Costa, nº 2367, Bairro Boa Esperança. CEP: 78060-900, Cuiabá, MT, Brasil - prof.kuang@gmail.com

## INTRODUÇÃO

O debate sobre a violência e a criminalidade mobilizou nos últimos anos muitos estudos e pesquisas no âmbito nacional e internacional em busca do entendimento deste fenômeno social, que instaura um campo de análise problemática e conflituosa (SOARES, 2014).

No Brasil, o crescimento das taxas de criminalidade tem gerado sentimento de insegurança à população nos últimos anos, sobretudo para os crimes de trânsito, em que o Brasil é campeão mundial. E nesse sentido, os custos do crime são significativamente altos, como prejuízos materiais, gastos públicos e privados em segurança, redução da qualidade de vida e a perda de capital humano (SANTOS e KASSOUF, 2007).

Embora o Brasil possua legislação específica com abrangência aos cinco fatores de risco, sua aplicação é ineficaz, pois não conseguiu reduzir os índices de acidentes, estando com uma taxa próxima a 20 mortes por 100 mil habitantes desde 2004. Segundo o Ministério da Saúde, anualmente 1,3 milhão de pessoas morrem no trânsito com até 50 milhões de sobreviventes feridas. Estes números são alarmantes e mostram a dimensão do problema e com base nisso, este artigo tem como objetivo utilizar análises estatísticas, como os métodos de análise multivariados e estatística espacial, que são ferramentas muito úteis para a avaliação desses dados criminológicos (GULUMBE *et al.*, 2013).

A análise multivariada é um conjunto de técnicas estatísticas, que permite a análise e interpretação de um conjunto de dados, que com o auxílio de programas computacionais possibilitam realizar análises com um grande número de variáveis de forma conjunta muito mais rápido e eficiente para a interpretação dos resultados. Por conta disso o uso dessa metodologia tem grande aplicação em várias áreas do conhecimento.

Dentre as técnicas multivariadas, será utilizada a Análise de Componentes Principais (ACP) e a Análise Fatorial (AF), que são técnicas multivariadas de modelagem da covariância e através da formação linear das variáveis originais as transformam em um conjunto significativamente menor de componentes não correlacionados, elucidando grande parte da variabilidade dos dados.

A Estatística Espacial de Área (EEA) é um método estatístico que utiliza a referência geográfica no modelo, com as coordenadas especiais no processo de coleta, descrição e análise dos dados. Sendo que, o enfoque está nos processos que acontecem no espaço, as técnicas estatísticas procuram então interpretar e analisar o comportamento desses processos (ARAÚJO *et al.*, 2014).

O presente estudo apresenta a combinação das técnicas multivariadas com a estatística espacial. A ACP e AF como instrumentos de resumo e sumarização da massa de dados, sem perda significativa de informações. E a estatística espacial cooperará na identificação das regiões mais alarmantes e verificação de municípios que possuam dependência espacial para os fatores gerados pela análise fatorial. Assim, o presente trabalho desenvolve uma metodologia que cooperará para identificar as regiões mais alarmantes quanto aos crimes e acidentes de trânsito no estado de Mato Grosso.

## MATERIAIS E MÉTODOS

### Banco de dados

O conjunto de dados foi obtido pelo site da Secretaria de Estado de Planejamento (SEPLAN) de Mato Grosso, provenientes de ocorrências registrados pela Polícia Militar do Estado de Mato Grosso e cadastrados no Sistema Estatístico de Ocorrência e Produtividade da Polícia Militar (SEOP/PM), referente ao ano de 2015.

O conjunto contém oito variáveis que representam categorias de crimes e acidentes de trânsito: X1: Abalroamento; X2: Atropelamento; X3: Capotamento; X4: Choque Mecânico; X5: Colisão; X6: Homicídio culposo na direção de veículo; X7: Lesão corporal culposa na direção de veículo e X8: Outros tipos de acidentes. Estas variáveis foram utilizadas para demonstrar as técnicas de análise de componentes principais (ACP) e análise fatorial (AF) juntamente com a estatística espacial (EE). O *shapefile* do estado de Mato Grosso para a aplicação da estatística espacial foi obtida no site do Instituto Brasileiro de Geografia e Estatística (IBGE). Todas as análises foram realizadas por meio do software R 3.3.0 (R Development Core Team, 2014).

### Análise de Componentes Principais (ACP)

A ACP é uma técnica multivariada que pode ser utilizada para gerar índices, agrupamento de indivíduos ou exclusão de variáveis. Sendo que agrupa os indivíduos de acordo com sua variação em que os indivíduos são agrupados segundo suas variâncias, ou seja, conforme seu comportamento dentro da população, representado pela variação do conjunto de características que define o indivíduo (HONGYU, 2015).

O objetivo da análise principal é encontrar combinações para  $n$  variáveis  $X_1, X_2, \dots, X_n$  gerando os índices  $Z_1, Z_2, \dots, Z_n$  que devem ser não correlacionados e que descreva a variação

dos dados. A ausência da correlação significa que os índices estão tendo "dimensões" diferentes dos dados, ou seja, o número de linhas ( $n$ ) é diferente do de colunas ( $p$ ), sendo que a ordem destes  $Var(Z_1) \geq Var(Z_2) \geq \dots \geq Var(Z_n)$ , em que  $Var(Z_1)$  expressa a variância de  $Z_1$  e os índices  $Z$  os componentes principais (ARAÚJO; COELHO, 2009).

Seja  $\mathbf{X}_{n \times p}$  matriz de  $p$  variáveis com médias  $\mu_p$  e  $\sigma_p^2$  variâncias com  $n$  observações não independentes entre si, ou seja, as variáveis possuem covariância, sendo que as covariâncias entre os pares de variáveis são representadas na matriz de covariância  $\Sigma_{p \times p}$  (HONGYU *et al.*, 2016).

$$\Sigma = \begin{bmatrix} \sigma^2_{11} & \cdots & \sigma^2_{1p} \\ \vdots & \ddots & \vdots \\ \sigma^2_{p1} & \cdots & \sigma^2_{pp} \end{bmatrix}$$

De acordo com REGAZZI (2000), para aplicações em diversas áreas do conhecimento o número de componentes utilizados tem sido aquele que acumula 70% ou mais de proporção da variância total (JOLLIFFE 1972, 1973; SAVEGNAGO *et al.*, 2011).

### Análise Fatorial (AF)

Com o propósito de estudar a criminalidade do estado do Mato Grosso (MT) a Análise Fatorial (AF) será aplicada como objetivo agrupar e construir grupos com as variáveis de variabilidades afins em fatores onde será obtido um resumo da massa de dados com a mínima perda de informação possível (NEISSE; HONGYU, 2016).

Para a implementação da AF primeiro é necessário verificar se a aplicação da análise fatorial tem validade para as variáveis escolhidas, para esse objetivo utilizou-se o teste de *Kaiser-Meyer-Olkin* (KMO):

$$KMO = \frac{\sum_{j=1}^p \sum_{m=1, m \neq j}^p r^2_{jm}}{\sum_{j=1}^p \sum_{m=1, m \neq j}^p r^2_{jm} + \sum_{j=1}^p \sum_{m=1, m \neq j}^p r^2_{pjm}}$$

Sendo que  $r^2_{jm}$  e  $r^2_{pjm}$  correspondem às correlações simples e parciais respectivamente.

A estatística KMO tem por finalidade verificar se há uma estrutura de dependência que seja bem definida entre as variáveis analisadas e que esta possa ser expressa pela matriz de correlações ou de covariância. Se caso tenha existência, a estrutura implica que a variável pode ser prevista pelas demais com certa margem de segurança. Portanto, o modelo somente

apresentará uma estrutura de dependência clara se a correlação parcial entre os pares de variáveis for baixa (FARIAS *et al.*, 2010).

**TABELA 1 - Adequação Amostral de acordo com a medida KMO.**

<b>KMO</b>	<b>Adequação</b>
> 0,9	Excelente
(0,8; 0,9]	Meritória
(0,7; 0,8]	Intermediária
(0,6; 0,7]	Medíocre
(0,5; 0,6]	Mísera
< 0,5	Inaceitável

**Fonte: NEISSE; HONGYU (2016).**

Para o teste KMO o resultado deve variar de 0 a 1, em que quanto menor for o valor, abaixo de 0,50 (KAISER,1958; PASQUALI, 1999), aumentam as chances de os dados não serem adequados. Uma das mais importantes decisões a ser tomada durante a execução de AFE se refere ao número de fatores a ser retido (GLORFELD, 1995). Uma extração inadequada impossibilita a interpretação dos resultados de maneira apropriada.

Muitas vezes as cargas fatoriais apresentam valores muito elevados em mais de um fator, o que acaba levando divisão de fatores inconclusivos. O método rotação dos fatores, tem como principal objetivo tornar o resultado empírico, ou seja, facilitar a interpretação conservando as suas propriedades estatísticas. De acordo com Pallant (2007), o tipo de rotação *Varimax* é o mais comumente utilizado pois esse método procura minimizar o número de variáveis que apresentam altas cargas em cada fator. O método de rotação *Varimax* tem como base os pressupostos do autor Thurstone (1947), em que uma estrutura de fatores parcimoniosa é construída com o objetivo que cada variável se manifeste de forma significativa apenas em um fator (DAMÁSIO, 2012). Por esse motivo, esse artigo utilizará esse tipo de rotação de fatores.

Depois da rotação e interpretação dos fatores, Malhotra (2004) e Hair *et al.* (1998) explicam que é preciso calcular as cargas fatoriais. Isso porque cada variável possui para cada fator encontrado um escore fatorial, sendo que as cargas fatoriais podem ser obtidas na matriz de coeficientes de cargas fatoriais. Para isso, o método utilizado nesse estudo será pela regressão, a fim de prever as mudanças na variável dependente como resposta as mudanças das variáveis independentes, por meio da regra estatística dos mínimos quadrados.

## Estatística Espacial (EE)

A estatística espacial é uma análise da localização, visualização e identificação da distribuição dos eventos que ocorrem no espaço. Dessa forma, consegue-se identificar padrões e modelar ocorrências no espaço (SANTOS; SOUZA, 2007).

São três os tipos de dados em análise espacial, Eventos ou Padrões Pontuais – que são ocorrências identificadas como pontos no espaço; Superfície Contínua – obtida a partir de amostras de campo, e normalmente utilizada em mapas topográficos, geológicos, ecológico, etc. Agora o último tipo de dado que são de Áreas – aqui são os dados de levantamentos populacionais e que se refere a ocorrências localizadas em cidades, bairros, etc. (DRUCK; CARVALHO; CÂMARA, 2004).

## Matriz Vizinhança

O primeiro passo é verificar o tipo da matriz vizinhança, também é conhecida como matriz de distância, matriz de conectividade ou matriz de proximidade. A análise de dados de área é a construção de uma matriz de vizinhança, esta matriz aponta a relação espacial de cada área do estudo, podendo ser ligada pela lista de vizinhos de cada município, ou pelas distâncias dos municípios ligados pelas estradas, e pela conectividade ponderada pelo comprimento da fronteira comum. É importante convencionar as formas de vizinhança nas matrizes de vizinhança que consideram a contiguidade sendo que existem três tipos, *Rook*, *Bishop* e *Queen*, nosso caso será a vizinhança do tipo “*Queen*” (ARAÚJO *et al.*, 2014, SEFFRIN *et al.*, 2018)

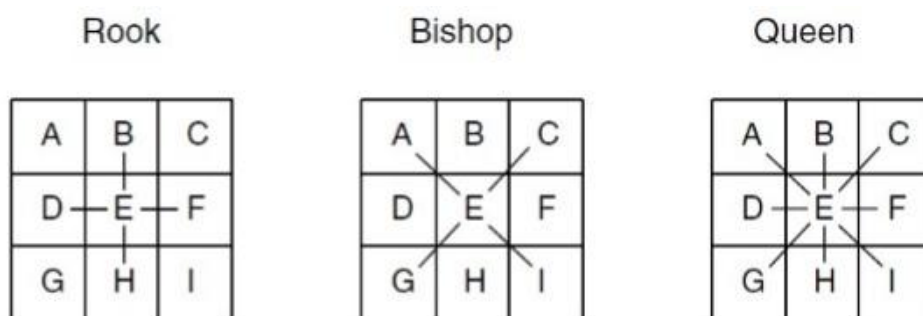


FIGURA 1 - Regras de *Rook*, *Bishop* e *Queen*.

Na Figura 1 mostrou as regras de *Rook*, *Bishop* e *Queen*, em que a regra *Rook*, que considera como vizinho somente as áreas com lado em comum da área **E**, a regra *Bishop*, que usa como critério de vizinhança as áreas que se das diagonais da área **E** e a regra de *Queen* aonde todas as áreas que têm intersecção não-nula com a área **E** serão vizinhas de **E**.

Para estimar a variabilidade espacial dos dados, usou-se a matriz de vizinhança espacial. Com um conjunto de  $n$  áreas (A) então  $\{A_1, \dots, A_n\}$ , então construí a matriz de primeira ordem  $W^{(1)}(n \times n)$ , cada um dos elementos  $w_{ij}$  representa uma medida de vizinhança entre  $A_i$  e  $A_j$ . Pode ser calculada a partir dos critérios (CÂMARA *et al.*, 2002):

- Se o centroide de  $A_i$  está a uma distância de  $A_j$ , então  $w_{ij}=1$ , caso contrário  $w_{ij}=0$ ;
- Se  $A_i$  tem um lado comum com  $A_j$ , então  $w_{ij}=1$ , caso contrário  $w_{ij}=0$ ;
- $w_{ij} = \frac{l_{ij}}{l_i}$  em que  $l_{ij}$  é o comprimento da fronteira entre  $A_i$  e  $A_j$  e  $l_i$  é o perímetro de  $A_i$ .

### Índice de Moran

Para a verificação da existência de dependência espacial utiliza-se o índice de Moran I, que foi proposto em 1948 sendo o mais usado em análise espacial. O método é usado para saber se existe ou não correlação espacial na variável estudada. A equação abaixo demonstra a fórmula para calcular a autocorrelação espacial global das variáveis (SARAIVA; CONCEIÇÃO; FRANCO, 2017).

$$I = \frac{S_0 \sum_{i=1}^n \sum_{j=1}^n Z_i Z_j W_{ij}}{\sum_{i=1}^n Z_i^2}$$

Em que o valor de  $S_0$ :

$$S_0 = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}}$$

Sabendo que  $Z_i = (x_i - \bar{x})$  e  $Z_j = (x_j - \bar{x})$ ;  $n$ : é o número de observações ou localidades;  $x_i, x_j$  são os valores das variáveis de estudo (interesse) nas localidades  $i$  e  $j$ ;  $\bar{x}$  é a média da variável  $x$ ;  $w_{ij}$  é o peso espacial das regiões de cada localidade  $i$  e  $j$ .

Assim, a estatística I de Moran é baseada nas somas de produtos cruzados de  $x_i$  para regiões vizinhas, segundo um critério de vizinhança dado pela matriz de pesos espaciais  $W$  (ALMEIDA, 2004).

De acordo com Khan (2010) e Seffrin, *et al.* (2018) os valores negativos próximos de zero indicam a ausência de autocorrelação espacial significativa para os valores da variável de estudo na região analisada. Os valores que são negativos e significativos apontam para uma autocorrelação negativa, indicando que a variável de interesse tende a ter valores diferentes para os locais mais próximos na região analisada. Quando os valores são positivos e significativos

indicam a autocorrelação espacial positiva, portanto, o valor da variável de interesse tende a ser semelhante nos vizinhos mais próximos na região observada. Ou seja:

- $I < 0$  correlação negativa então, a existência de regularidade.
- $I = 0$  ausência de correlação, portanto a distribuição espacial é aleatória.
- $I > 0$  correlação positiva, ou seja, têm agrupamento ou conglomerados.

Após calcular o Índice de Moran deve-se testar as hipóteses a fim de saber se existe ou não a presença de correlação:

$$\begin{cases} H_0: I = 0 \text{ ausência de correlação espacial} \\ H_1: I \neq 0 \text{ existência de correlação espacial} \end{cases}$$

Para obter a significância do índice de Moran é satisfeita duas suposições básicas a normalidade e a aleatoriedade. Em ambos os casos a significância do índice de Moran utiliza-se a estatística  $Z_{calculado}$ :

$$Z_{calculado} = \frac{I - \mu}{\sqrt{\sigma^2}}$$

Em que,  $\mu$  e  $\sigma^2$  são a média e variância esperados respectivamente, portanto, rejeita-se  $H_0$  se o valor se o valor-p  $\geq \alpha$  ou se,  $Z_{calculado} > Z_\alpha$ . Para  $Z_\alpha$  o valor é obtido na distribuição padrão.

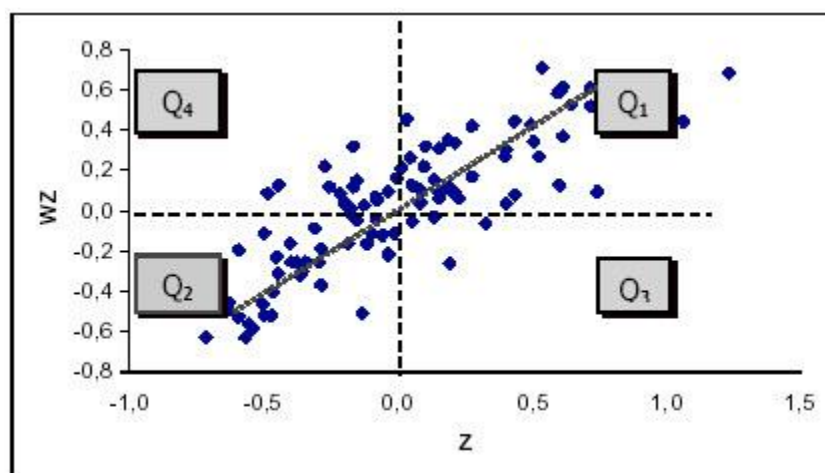
A hipótese implícita do cálculo do índice de Moran é a estacionariedade de primeira e segunda ordem, e o índice perde sua validade ao ser calculado para dados não estacionários. Quando existir não-estacionariedade de primeira ordem (tendência), os vizinhos tenderão a ter valores mais parecidos que áreas distantes, pois cada valor é comparado à média global, inflacionando o índice. Igualmente, se a variância não é constante, nos locais de maior variância o índice será mais baixo, e vice-versa. Quando o dado é não-estacionário, a função de autocorrelação continua decaindo mesmo após ultrapassar a distância onde há influências locais (CÂMARA, *et al.*, 2002).

### **Diagrama de Espalhamento de Moran**

Quando se aplica o Índice de Moran global é importante realizar uma análise gráfica da autocorrelação espacial por meio do diagrama de dispersão de Moran ou Diagrama de Espalhamento de Moran (Figura 2) que permite comparar o valor de cada município com seus vizinhos, identificando pontos com associação ou não espacial. Utiliza o indicador normalizado. Essa técnica permite visualizar graficamente, através do diagrama de espalhamento de Moran, o grau de similaridade entre vizinhos para definição de padrões espaciais. Colocando-se no eixo **X** o valor do indicador do município  $i$  normalizado e no eixo



Y o valor da estatística de Moran apenas entre os vizinhos de i, cada quadrante do gráfico que é interpretado como (PAIVA, 2007; NUNES, 2013):



Fonte: Paiva (2007)

FIGURA 2 - Diagrama de Espalhamento de Moran (*box map*).

Em que  $Z$  são os valores normalizados e  $WZ$  média dos vizinhos também normalizados gerando um mapa (*box map*). Portanto, para Anselin (1995), Araújo (2012) no eixo vertical é mostrada a defasagem espacial da variável de interesse, e já no eixo horizontal resulta no valor da variável de interesse. O diagrama está dividido em quatro partes: **Q1** = HH (*High-High*) (valores positivos, médias positivas) e **Q2** = LL (*Low-Low*) (valores negativos, médias negativas): que indicam os pontos de associação espacial positiva, no sentido que uma localização possui vizinhos com valores semelhantes; **Q3** = LH (*Low-High*) (valores positivos, médias negativas) e **Q4** = HL (*Hig-Low*) (valores negativos, médias positivas): indicam pontos de associação espacial negativa, no sentido que uma localização possui vizinhos com valores distintos.

Neves *et al.* (2000) e Anselin (1995) definem que no *box map*, os objetos são classificados em relação aos quadrantes do gráfico de espalhamento conforme sua posição, podendo receber uma cor correspondente no mapa gerado. Para gerar o LISA map, é avaliada a significância dos valores do índice de Moran Local obtido para cada objeto, em relação à hipótese  $H_0$  (ausência de correlação espacial). Na avaliação da significância é utilizada a abordagem de permutação dos atributos dos vizinhos. Os objetos são classificados em quatro grupos: não significantes; com significância entre 0,05 e 0,01; com significância entre 0,01 e 0,001; e maior que 0,001. Já no *Moran map*, semelhantemente ao LISA map, apenas os objetos em que os valores de LISA foram considerados significantes ( $p > 0,05$ ), são destacados, mas,

aparecem classificados em quatro grupos, conforme sua localização no quadrante do gráfico de espalhamento. Os restantes dos objetos ficam classificados como sem significância.

## RESULTADOS E DISCUSSÕES

O índice de KMO foi realizado na amostra das categorias de crimes medidos no estado de Mato Grosso sendo que o resultado do teste apontou um valor de 0,72 indicando que os dados são adequados ao método de Análise Fatorial, considerando os critérios apresentados na Tabela 1.

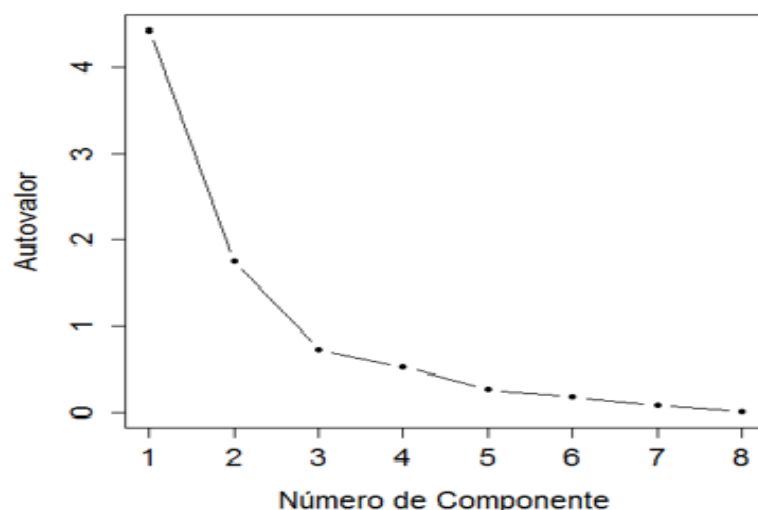
Na Tabela 2, são apresentados os resultados obtidos pela técnica da análise de componentes principais, os autovalores e a porcentagem da variância explicada por cada componente principal. Os dois primeiros componentes principais (CP1 e CP2) explicam aproximadamente 77,38% da variabilidade total, sendo que o CP1 explica cerca de 55,38% e 22% pertence ao CP2.

**TABELA 2 - Autovalores e percentual da variância explicada**

Componente Principal	Autovalores	Variância explicativa (%)	Variância acumulativa (%)
CP1	2,10	0,5538	55,38
CP2	1,33	0,2200	77,38
CP3	0,85	0,0912	86,50
CP4	0,73	0,0658	93,08
CP5	0,52	0,0340	96,48
CP6	0,43	0,0227	98,75
CP7	0,29	0,0109	99,84
CP8	0,11	0,0016	100,00

Observamos que a partir do CP3 o autovalor é  $<1$ , o que não satisfaz o critério de *Kaiser*. Portanto, os dois primeiros componentes principais são considerados eficientes, pois possuem autovalores  $> 1$  (KAISER, 1958) e conjuntamente representam 77,38% da variação dos dados (Tabela 2 e Figura 3).

Pelo gráfico *scree plot* (Figura 3) que foi obtido por meio da AF, em que é utilizado o método de extração dos componentes principais e o método de rotação *Varimax* em uma matriz de correlação composta por 8 variáveis, confirma que os dois componentes principais obedecem ao critério de *Kaiser*.



**FIGURA 3 - Scree plot dos autovalores obtidos da AF.**

Na Tabela 3 apresentada abaixo, é possível verificar as cargas rotacionadas pelo método *Varimax*, foram marcadas em negrito as maiores entre os dois fatores para cada variável, ou seja, aquelas que indicam maior explicação da variável pelo fator. Percebe-se que as variáveis “Abalroamento” (0,981), “Colisão” (0,953), “Choque Mecânico” (0,884), “Capotamento” (0,817) e “Outros tipos de acidentes” (0,806) são as que apresentam maiores valores de cargas fatoriais no Fator 1(MR1), e no Fator 2 (MR2) as maiores cargas pertencem as variáveis “Homicídio culposo na direção de veículo” (0,923) e “Lesão corporal culposa na direção de veículo” (0,874).

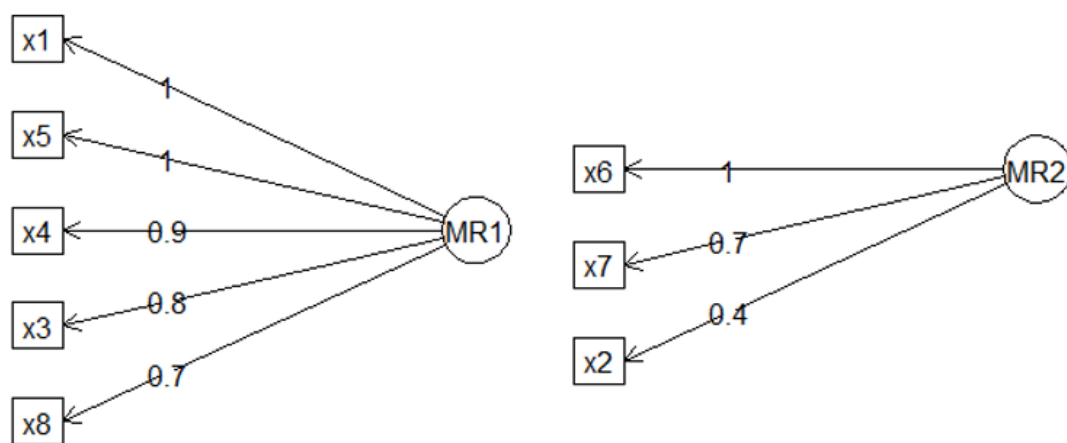
**TABELA 3 - Cargas fatoriais rotacionadas pelo método *Varimax*.**

Cód	Variável	Fator (MR1)	Fator (MR2)
x1	Abalroamento	<b>0,981</b>	
x2	Atropelamento	0,407	<b>0,516</b>
x3	Capotamento	<b>0,817</b>	0,262
x4	Choque Mecânico	<b>0,884</b>	0,309
x5	Colisão	<b>0,953</b>	
x6	Homicídio culposo na direção de veículo		<b>0,923</b>
x7	Lesão corporal culposa na direção de veículo		<b>0,874</b>
x8	Outros tipos de acidentes	<b>0,806</b>	

Na Figura 4, como gráfico auxiliar para corroborar as interpretações da Tabela 3, o Fator 1 (MR1) agrupou as variáveis relacionadas aos acidentes de trânsito e são eles “Abalroamento” (x1), “Colisão” (x5), “Choque Mecânico” (x4), “Capotamento” (x3) e “Outros tipos de acidentes” (x8). Observa-se na Figura 4, que todas as cinco variáveis que compõem o MR1

possuem cargas fatoriais maiores que 0,5. Sendo que os itens “Abalroamento” (x1) e “Colisão” (x5) são as que apresentam os maiores valores.

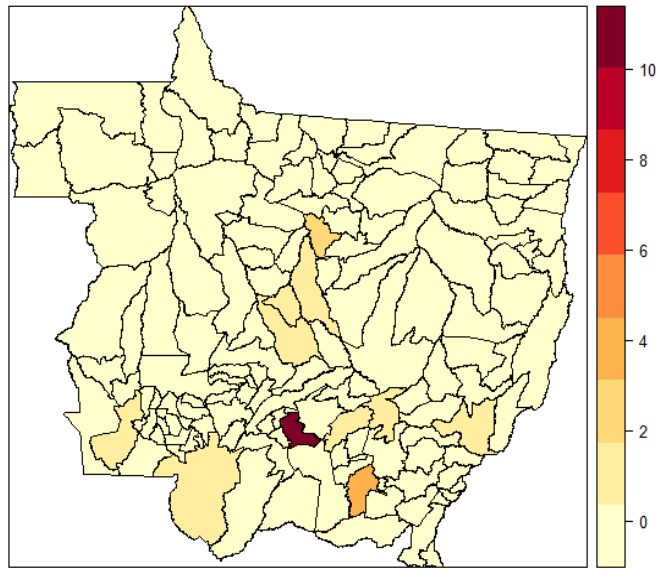
O Fator 2 (MR2) agrupou as variáveis relacionadas aos crimes de trânsito, são eles “Homicídio culposo na direção de veículo” (x6), “Lesão corporal culposa na direção de veículo” (x7) e “Atropelamento” (x2). O item que possui a maior carga fatorial é “Homicídio culposo na direção de veículo”, logo em seguida vem “Lesão corporal culposa na direção de veículo” e a que possui a menor carga fatorial é o item “Atropelamento” sendo igual a 0,4 (Figura 4).



**FIGURA 4 – Diagrama da AF com as variáveis que compõem os dois fatores e suas respectivas cargas fatoriais.**

Após definidas as oito variáveis que foram resumidas em apenas dois fatores, foi realizada a análise espacial que a seguir verificaremos se cooperou na identificação das regiões mais alarmantes e verificação de municípios que possuam dependência espacial para os fatores gerados pela análise fatorial.

Na Figura 5 representa o mapa temático da distribuição de ocorrências relacionadas ao Fator 1 (Acidentes de Trânsito) no mapa, onde nota-se que Cuiabá é o município com mais ocorrência, seguido por Rondonópolis, Sinop, Tapurah, Lucas do Rio Verde, Nova Mutum, Campo Verde, Primavera do Leste, Cáceres, Pontes e Lacerda e Barra do Garças.

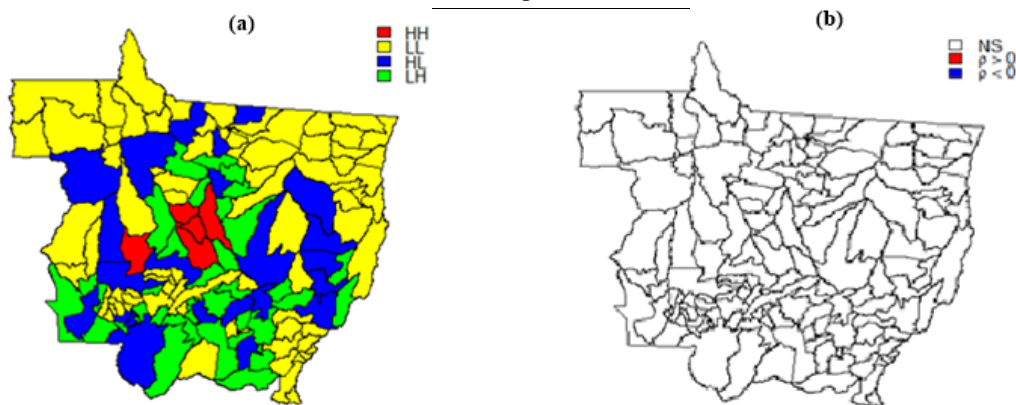


**FIGURA 5 - Mapa temático da distribuição espacial de acidentes de trânsito (Fator 1 - MR1) nos municípios de Mato Grosso no ano de 2015.**

O índice global I de Moran e seu teste de significância do Fator1 (MR1) associado à suposição aleatória estão representados pela Tabela 4. Diante do alto valor-p (0,6272) conclui-se que não há evidências para a rejeição da hipótese nula, ou seja, a hipótese de independência espacial não foi rejeitada.

**TABELA 4 - Índice global I de Moran para o fator 1 (acidentes de trânsito).**

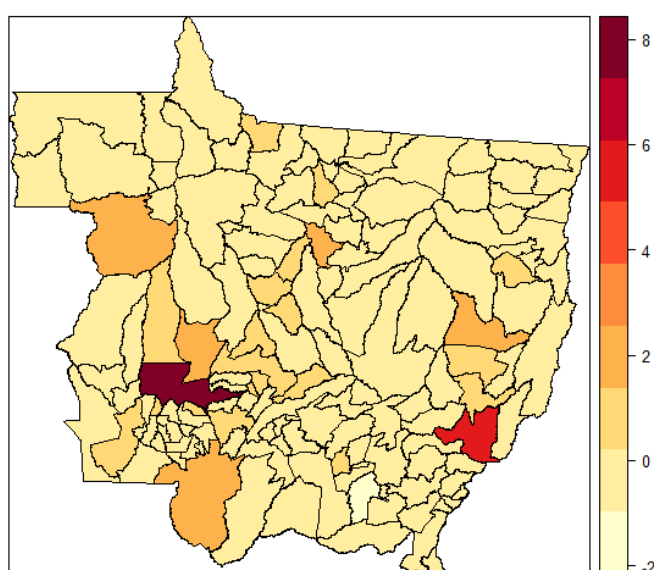
Moran	Fator MR1
<i>I</i>	-0,0238
valor-p	0,6272



**FIGURA 6 - Mapa da autocorrelação espacial *Box map* (a) e *Lisa Map* (b) que representa as regiões de significância obtidas através do indicador local de Moran dos Acidentes de Trânsito (Fator MR1) no Estado de Mato Grosso no ano de 2015.**

Como análise adicional para a visualização da falta de dependência espacial para o Fator 1 (MR1), foi realizado o diagrama de espalhamento ilustrado pela Figura 06 (a). Porém, o resultado representado pela Figura 6 (b) para o índice local  $I$  de Moran, não indica a formação de clusters locais, sendo assim, não possui nenhum município com ocorrências alarmantes considerando o Fator 1 (Acidentes de Trânsito).

O mapa temático do Fator 2 (Crimes de Trânsito) representado pela Figura 7 indicou a sua distribuição pelo estado de Mato Grosso. O município de maior destaque para o Fator 2 (MR2) é Tangará da Serra seguido por Barra do Garças, Sinop, Cáceres, Juína, Canarana, Campo Novo do Parecis.

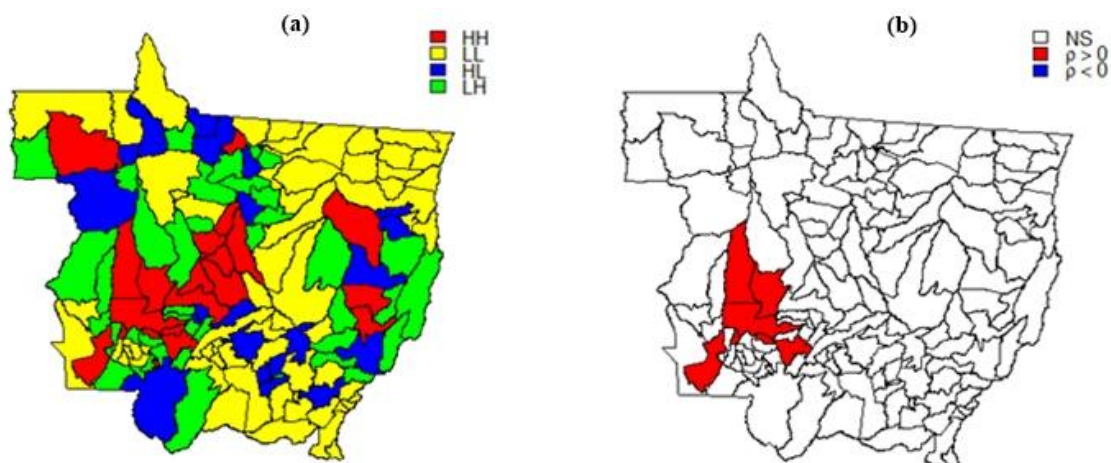


**FIGURA 7 - Mapa temático da distribuição espacial de Crimes de Trânsito (Fator MR2) nos municípios de Mato Grosso no ano de 2015.**

Na Tabela 5 apresentou o índice global  $I$  de Moran e seu teste de significância para o Fator 2 (MR2), também associado à suposição aleatória. O alto valor-p (0,1839) demonstra que não se pode obter evidências para a rejeição da hipótese nula, ou seja, a hipótese de independência espacial não é rejeitada.

**TABELA 5 - Índice global  $I$  de Moran para o fator 1 (Crimes de trânsito).**

Moran	Fator MR2
$I$	0,0393
valor-p	0,1839



**FIGURA 8 -** Mapa da autocorrelação espacial *Box map* (a) e *Lisa Map* (b) que representa as regiões de significância obtidas através do indicador local de Moran dos Crimes de Trânsito (Fator MR2) no Estado de Mato Grosso no ano de 2015.

O diagrama de espalhamento de Moran representado pelo *Box map* - Figura 8 (a) corrobora com a hipótese de independência espacial entre os dados, como demonstrado pelo índice global I de Moran. Apesar do índice global não indicar presença de dependência espacial, o índice local I de Moran – Figura 8 (b) - indica a existência de autocorrelação local, com a formação de clusters. São eles Pontes e Lacerda Tangará da Serra, Sapezal e Campo Novo do Parecis, que formaram uma região de alarde para o Fator MR2 (Crimes de Trânsito).

## CONCLUSÃO

O presente artigo teve o intuito de utilizar a união de técnicas estatísticas multivariadas juntamente com estatística espacial para o estudo dos crimes e acidentes de trânsito no Estado de Mato Grosso em 2015.

De acordo com os resultados obtidos, a análise fatorial permitiu agrupar as oito variáveis compostas por diferentes tipos de crimes e acidentes de trânsito em apenas dois fatores, o que facilita a interpretação e utilização de outras técnicas. Pode-se concluir então, que o método foi eficaz e cumpriu os objetivos propostos que são os de resumo, sumarização dos dados e geração de *scores* para utilização na estatística espacial.

Com a estatística espacial se obteve um mapeamento dos fatores no estado de Mato Grosso, mostrando a distribuição dos diferentes crimes e acidentes de trânsito, assim informando regiões críticas que precisam de intervenção do poder público para melhoria desses números, porém não conseguiu identificar dependência espacial dos municípios em relação aos fatores estudados.

## REFERÊNCIAS BIBLIOGRÁFICAS

ARAÚJO, E. C.; URIBE-OPAZO, M. A.; JOHANN, J. A. **Modelo de regressão espacial para estimativa da produtividade da soja associada a variáveis agrometeorologias na região oeste do estado do Paraná.** Eng. Agric., Jaboticabal, v. 34, n. 2, p. 286-299, abr. 2014.

ARAÚJO, W. O.; COELHO, C. J. **Análise de Componentes Principais (PCA)**, Relatório Mestrando em Sociedade, Tecnologia e Meio Ambiente – Centro Universitário de Anápolis UniEVANGÉLICA, 2009. Orientador: CLARIMAR JOSE COELHO.

BÔAS, E.F, SILVA, M.A. **Trânsito e segurança pública: Impactos e consequências.** Revista Humanidades, Bahia, v.4, n.2, 2015.

BRASIL. MINISTÉRIO DA SAÚDE (MS). SECRETARIA DE VIGILÂNCIA EM SAÚDE, FUNDAÇÃO OSWALDO CRUZ (FIOCRUZ). **Introdução a Estatística Espacial para Saúde Pública.** 2007.

DAMÁSIO, B. F. **Uso da análise fatorial exploratória em psicologia.** Avaliação psicológica, 2012, 11.2: 213-228.

DRUCK, S; CARVALHO, M. S; CÂMARA, G; MONTEIRO, A. M. V. **Análise Espacial de Dados Geográficos.** Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA). Brasília, 2004.

FARIAS, C.A; FIGUEIREDO, A.M; LIMA, J.E. **Dependência espacial e análise de agrupamento de municípios para diferentes tipos de crime em Minas Gerais.** REUNA, 13.3, 2010.

Glorfeld, L. W. *An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain.* Educational and Psychological Measurement, 55 (3), 377-393, 1995.

GULUMBE, S.U; DIKKO, H.G; BELLO, Y. *Analysis of Crime Data using Principal Component Analysis: A case study of Katsina State.* CBN Journal of Applied Statistics, 2013, 3.2: 39-40.

HAIR, J. F.; ANDERSON, R. E.; TATHAM, R. L.; BLACK, W. C. *Multivariate data analysis.* 4 ed. New Jersey: Prentice Hall, 1998. 730p.

HONGYU, K. **Comparação do GGE biplot-ponderado e AMMI-ponderado com outros modelos de interação genótipo × ambiente.** 2015. PhD Thesis. Escola Superior de Agricultura "Luiz de Queiroz.

HONGYU, K; SANDANIELO, V.L.M; JUNIOR, G.J.O. **Análise de Componentes Principais: Resumo Teórico, Aplicação e Interpretação.** E&S Engineering and Science, 2016, 5.1: 83-90.

JOLLIFFE, I.T. *Discarding Variables in a Principal Component Analysis.* I: Artificial Journal of the Royal Statistical Society. Series C (Applied Statistics), Vol. 21, No. 2 p. 160-173, 1972.



JOLLIFFE, I.T. *Discarding Variables in a Principal Component Analysis*. II: Real Data. Journal of the Royal Statistical Society, v. 22, n. 1, p. 21–31, 1973.

KAISER, H.F. *The varimax criterion for analytic rotation in factor analysis*. Psychometrika, 1958, 23.3: 187-200.

KHAN, A. S. Dependência espacial e emprego formal. O que é possível afirmar para indústria cearense. **In. Anais do VI Encontro de Economia do Ceará em Debate. Fortaleza - CE, novembro de 2010.**

MALHOTRA, N. K. *Marketing research*. 4 ed. New York: Prentice Hall, 2004. 864p.

NEISSE, A.C; HONGYU, K. **Aplicação de componentes principais e análise fatorial a dados criminais de 26 estados dos EUA**. E&S Engineering and Science, v.5, n.2, p.105-115, 2016.

NEVES, M. C; RAMOS, F. R; CAMARGO, E. C. G; CÂMARA, G; MONTEIRO, A. M. **Análise exploratória espacial de dados sócio-econômicos de São Paulo**. Salvador, 2000.

NUNES, F. G. **Análise exploratória espacial de indicadores de desenvolvimento socioambiental das regiões de planejamento do norte e nordeste goiano**. Ateliê Geográfico, Goiânia-GO, v.7, n.1, p.237-259. Abril/2013

OMS - ORGANIZAÇÃO MUNDIAL DE SAÚDE. **Relatório Global de Segurança no Trânsito 2013**. Disponível em:< <http://www.detran.rs.gov.br/decadars/?p=1655>>. Acesso em: setembro de 2018.

PAIVA, C. **Dependência Espacial: Setores censitários, Zonas OD, Distritos, Sub Prefeituras, etc**. Carlos Eduardo de Paiva Cardoso, comunicação pessoal, CET/SP e PUC/SP, 2007.

PALLANT, J. *SPSS Survival Manual*. Open University Press, 2007.

PASQUALI, L. **Análise fatorial: um manual teórico-prático**. Brasília: Editora da UnB, no prelo, 1998.

PEIXOTO, B.T; LIMA, R.S; DURANTE, M.O; **Metodologias e criminalidade violenta no Brasil. São Paulo em Perspectiva**. Vol.18 nº.1:13-21. São Paulo Jan./Mar. 2004.

R DEVELOPMENT CORE TEAM. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing, Vienna, 2014.

REGAZZI, A.J. **Análise multivariada**. Departamento de Informática da Universidade Federal de Viçosa, v.2, 2000.

RESENDE M.L; FERNANDES L.P.S; SILVA A.M.R. **Utilização da análise fatorial para determinar o potencial de crescimento econômico em uma região do sudeste do Brasil**. UFSM. Revista Economia e Desenvolvimento, n.19, 2017.

SANTOS, M. J; KASSOUF, A. L. **Uma investigação econômica da influência do mercado de drogas ilícitas sobre a criminalidade brasileira.** Revista Economia, v. 8, n. 2, p. 187-210, 2007.

SANTOS, S. M; SOUZA, W. V. **Introdução à Estatística Espacial para a Saúde Pública.** Brasília: Ministério da saúde, p 122. 2007.

SARAIVA, M. V; CONCEIÇÃO, O. C; FRANÇA, Marco Tulio Aniceto. **Os determinantes da criminalidade nos municípios gaúchos: evidências de um modelo econométrico espacial.**Ensaio FEE, v. 38, n. 3, p. 521-552, 2017.

SAVEGNAGO, R.P; CAETANO, S.L; RAMOS, S.B; NASCIMENTO, G.B; SCHMIDT, G.S; LEDUR, M.C. MUNARI, D.P. *Estimates of genetic parameters, and cluster and principal components analyses of breeding values related to egg production traits in a White Leghorn population,* Poultry Science, 90, p.2174-2188. 2011.

SEFFRIN, R; ARAÚJO, E.C; BAZZI, C.L. **Spatial analysis of area applied to soybean productivity in the west region of Paraná using software R.** Brazilian Journal of Geomatics, v. 6, n. 1, p. 23-43, 2018.

SOARES, T.C; ZABOT U.C; RIBEIRO G.M. **Índice geral de criminalidade: uma abordagem a partir da análise envoltória de dados para os municípios catarinenses.** Leituras de Economia Política, Campinas, (19): 89-109, dez.2011.

SOARES, A. M. **O acúmulo da violência e da criminalidade na sociedade brasileira e a corrosão dos direitos humanos.** Ridh, Bauru, v.2, n.2, p. 161-189, 2014.

THURSTONE, L. L. *Multiple-factor analysis.* University of Chicago Press. Frist Edition, December1947.