

## Aplicação de Componentes Principais e Análise Fatorial a Dados Criminais de 26 Estados dos Eua

### Application of Principal Components and Factor Analysis to Crime Data From 26 Us States

<sup>1</sup>Anderson Cristiano Neisse, <sup>2</sup>Kuang Hongyu

<sup>1</sup>Graduando no Departamento de Estatística, Instituto de Ciência Exatas e da Terra - ICET da Universidade Federal de Mato Grosso – UFMT (a.neisse@gmail.com)

<sup>2</sup>Doutor em Estatística – USP, Professor do Departamento de Estatística, Universidade Federal de Mato Grosso – UFMT (prof.kuang@gmail.com)

---

**RESUMO:** A análise multivariada é um conjunto de técnicas estatísticas que permite a análise e interpretação de conjuntos de dados de natureza quantitativa com grande número de variáveis de forma simplificada. Dentre o rol de técnicas encontram-se a análise de componentes principais (ACP) e a análise fatorial (AF) que visam reduzir a quantidade de variáveis agrupando-as de acordo com sua variabilidade. Este artigo visa demonstrar a aplicabilidade das técnicas de ACP e AF na interpretação de dados de crime de 26 estados dos EUA.

**Palavras-chave:** Análise multivariada; análise de componentes principais; análise fatorial; crimes.

**ABSTRACT:** Multivariate analysis is a set of statistical techniques that allows the analysis and interpretation of quantitative data sets with large number of variables in a simplified form. Among the list of techniques there are the principal component analysis (PCA) and factor analysis (FA), both aims are to reduce the number of variables by grouping them according to their variability. This article aims to demonstrate the applicability of the PCA and FA techniques in the interpretation of the 26 US states crime data.

**Keywords:** multivariate analysis; principal component analysis; factorial analysis; crimes.

---

## INTRODUÇÃO

Crime é um dos problemas que continuamente perturbam a existência da humanidade, sendo ainda mais perturbador no complexo mundo moderno. Tendo começado nos primórdios da civilização como algo simples e desorganizado, atualmente é um problema complexo e organizado (GULUMBE, *et al*). A taxa de crimes é um fator importante para julgar a qualidade de vida em uma determinada área, seja ela um bairro, cidade ou estado, por constituir problemas de nível: econômico ao impor restrições no potencial de desenvolvimento; social por afetar a qualidade de vida das pessoas envolvidas e eventualmente toda ou parte de uma sociedade e político por tornar necessária a elaboração de planos estratégicos e alocação de recursos para combate ao crime (FARIAS, *et al*; 2008). Apesar de o crime ser inevitável em uma sociedade várias medidas preventivas podem e são aplicadas de forma a reduzir a ameaça. A introdução de métodos científicos e técnicos modernos

na prevenção de crimes tem se mostrado eficiente. Os métodos multivariados da estatística têm se provado eficientes tendo feito contribuições em muitas explicações criminológicas (GULUMBE, *et al*).

A análise multivariada é um conjunto de técnicas estatísticas que permite a análise e interpretação de conjuntos de dados de natureza quantitativa com grande número de variáveis de forma simplificada. O advento das técnicas multivariadas se deu pelo avanço tecnológico e o uso de computadores para análise estatística, sem os quais não seria possível analisar grande número de variáveis de forma conjunta. Por conta da facilidade de análise e interpretação de grande número de variáveis, as técnicas multivariadas têm se tornado muito populares em várias áreas do conhecimento apesar de terem sido desenvolvidas para problemas específicos, dentre as áreas de grande aplicação atualmente estão: agronomia, zootecnia, ecologia, florestal, etc. Dentre as várias possíveis aplicações das técnicas multivariadas estão as taxas de

ocorrência de crimes que são classificados em várias categorias como: Assassinato, Roubo, Estupro, Assalto, Arrombamento, Pequenos Furtos, Furtos Veiculares, etc.

A Análise de Componentes Principais (ACP) e a Análise Fatorial (AF) são técnicas multivariadas de modelagem da covariância por meio da transformação linear das variáveis originais em um conjunto significativamente menor de componentes não correlacionados que explicam a maior parte da variabilidade dos dados.

Portanto, ambas as técnicas estão relacionadas à redução de massa de dados através da combinação de variáveis de variabilidades afins. Apesar de as técnicas multivariadas terem sido desenvolvidas para soluções de problemas muito específicos, a possibilidade da redução de massa de dados tornou tais técnicas populares em várias áreas do conhecimento.

O presente estudo apresenta as técnicas multivariadas de ACP e AF como instrumentos de redução de massa de dados que possibilitam a análise de grandes bancos de dados através da redução da quantidade de variáveis sem perda significativa de informação. As duas análises são aplicadas a dados de diferentes categorias de crimes com frequências observadas em 26 estados nos Estados Unidos (EUA).

## MATERIAL E MÉTODOS

### *Banco de Dados*

O conjunto de dados utilizado neste trabalho foi retirado do *software SAS/STAT(R) 9.22 User's Guide (2008)* sobre um censo que forneceu informações sobre os índices de criminalidade por 100.000 pessoas em sete categorias para cada um dos 26 estados dos Estados Unidos (EUA) em 1997: *Missouri, Montana, Nebraska, Nevada, New Hampshire, New Jersey, New Mexico, New York, North Carolina, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, Rhode Island, South*

*Carolina, South Dakota, Tennessee, Texas, Utah, Vermont, Virginia, Washington, West Virginia, Wisconsin e Wyoming.*

O conjunto contém sete variáveis que representam categorias de crime: X1: Assassinato; X2: Estupro; X3: Roubo; X4: Assalto; X5: Arrombamento; X6: Pequenos furtos e X7: Roubo de veículos. Estas variáveis foram utilizadas para demonstrar as técnicas de ACP e AF. Toda a análise foi realizada por meio do software R 3.0.1 (R Development Core Team, 2014).

### *Análise de Componentes Principais*

A ACP possui várias aplicações nas mais variadas áreas científicas, o seu uso tem se popularizado após os avanços computacionais e o desenvolvimento de programações que possibilitem cálculos matriciais que diagonalizem matrizes simétricas positivas semi-definidas. Tem sido muito utilizada por pesquisadores por remover a multicolinearidade entre variáveis transformando-as em um conjunto de variáveis não correlacionadas, chamadas de componentes principais. Também reduz a massa de dados em poucos componentes ortogonais que explicam a variação dos dados em proporções de ordem decrescente. Entretanto, a ACP não é recomendada quando se têm muitos zeros na matriz de dados, ou muitos dados ausentes, pois é muito sensível a *outliers* (HONGYU, 2015). Também é recomendável sempre trabalhar com amostras maiores do que o número de variáveis para que não haja perda de informação da variabilidade original. Em casos de variáveis com baixa correlação o resultado da redução de variáveis é praticamente igual aos dados originais, ou seja, não há efeito (HONGYU, 2015; 2012; REGAZZI, 2000).

Seja  $X_{n \times p}$  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).

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

Com base na matriz de covariância são encontrados os pares de autovalores e autovetores  $(\lambda_1, e_1), (\lambda_2, e_2), \dots, (\lambda_p, e_p)$ , em que  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ , os componentes principais  $Z_i$  são obtidos através da combinação linear do  $i$ -ésimo autovetor com as  $p$  variáveis:

$$Z_i = e_{i1}X_1 + e_{i2}X_2 + \dots + e_{ip}X_p$$

Os componentes são, desta forma, variáveis não mensuradas a partir das informações levantadas no estudo (variáveis latentes). A variação contida nos componentes principais é igual à variação das variáveis originais, desta forma a contribuição de cada componente para a variação é expressa em porcentagem através da proporção da variação total que pertence a cada componente (HONGYU, 2016; JOHNSON; WICHERN, 1998).

É com base na porcentagem de contribuição da variabilidade total de cada componente que é realizada a escolha do modelo de  $k$  componentes, geralmente escolhendo de forma a obter 80% ou mais da variabilidade total. Um critério muito utilizado na retenção de fatores é o de Kaiser (KAISER, 1958) que afirma que os componentes com  $\lambda_i > 1$  representam parcela suficiente da variação total dos dados. Outro método que pode ser utilizado para reforçar a decisão com base no método de Kaiser é o *scree plot* ou teste de Cattell (CATTELL, 1966) que consiste na observação do gráfico dos autovalores pelo número de dimensões aonde o objetivo é localizar o ponto aonde os autovalores apresentam tendência linear decrescente (DAMÁSIO, 2012). Analisado de forma isolada, o *scree plot* pode ser inconclusivo em casos de fatores não muito bem definidos, então deve ser usado sempre em conjunto com outros critérios de forma a reafirmar a decisão.

## Análise Fatorial

A Análise Fatorial (AF) engloba técnicas multivariadas que buscam evidenciar variáveis latentes (fatores) que influenciam na variabilidade de um grupo ou de todas as variáveis originais sendo que as variáveis que partilham de variâncias comuns (são influenciadas pelas mesmas subjacências) são atribuídas a um mesmo fator (BROWN, 2006). O fator é uma variável não mensurada no estudo original que tem influência em mais de uma das variáveis originais (DAMÁSIO, 2012). Então a AF tem o objetivo de agrupar as variáveis de variabilidades afins em fatores de forma a reduzir a massa de dados com a mínima perda de informação possível.

Tanto a ACP quanto a AF assumem que a variabilidade dos dados é composta por três características: a variância específica, que diz respeito à variação que não é compartilhada com as demais variáveis; variância comum é a variância que é compartilhada entre variáveis e a variância do erro, que trata da variação não explicada pelo componente ou fator (DAMÁSIO, 2012).

Ao comparar os resultados de uma ACP e uma AF percebe-se que as cargas fatoriais e comunalidades da ACP são mais elevadas e taxas de variância explicada elevadas em relação a AF, pois a ACP considera a variância específica dos itens e não somente a variância comum, isso gera imprecisão quando o objetivo é entender as variáveis latentes que influenciam no grupo de variáveis (COSTELLO & OSBORNE, 2005).

O primeiro estágio para a aplicação da AF é verificar a possibilidade de fatoração da matriz de dados, permitindo assim a aplicação da análise fatorial aos dados. O teste mais aplicado é o critério de Kaiser-Meyer-Olkin (KMO), conhecido também por teste de adequação da amostra. Este teste indica a o nível de adequação da AF para o conjunto de dados por meio da estimação da porção da variância que pode estar sendo explicada por variáveis

subjacentes. O teste á baseado nos coeficientes de correlação linear  $r_{jm}$  entre as variáveis  $X_j$  e  $X_m$  e nos coeficientes de

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

O valor do KMO pode variar de zero a um, aonde valores próximos ou iguais a zero indicam uma soma de correlações parciais das variáveis alta em relação à soma das correlações totais. Nesses casos a possibilidade de a análise fatorial ser inapropriada é alta (PASQUALI, 1999). A

correlação entre os resíduos  $r_{pm}$ , é expresso da forma:

Tabela 1 traz a regra para interpretação dos índices de KMO e pode-se perceber que a adequação amostral com índice menor do que 0,5 é inaceitável para a aplicação de Análise Fatorial (PASQUALI, 1999; KAISER, 1958):

**Tabela 1:** Adequação Amostral de acordo com a medida KMO.

KMO	Adequação
> 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

A mais importante decisão de uma AF é a retenção dos fatores uma vez que se a quantidade de fatores retidos for equivocada os resultados não serão gerados apropriadamente. Um dos métodos muito utilizados, assim como na ACP, é o método de retenção com base nos autovalores de Kaiser (KAISER, 1958). O método do *scree plot* também pode ser utilizado na AF assim como na ACP.

Como as cargas fatoriais muitas vezes apresentam valores elevados em mais de um fator, isso pode levar a divisões de fatores inconclusivas. O método de rotação de fatores tem o objetivo de tornar a interpretação mais fácil sem alterar os resultados finais significativamente (COSTELLO & OSBORNE, 2005). O método de rotação *varimax* (KAISER, 1958) baseia-se nos pressupostos de Thurstone (1947) de que uma estrutura de fatores parcimoniosa é aquela construída de forma que cada variável se manifeste de forma significativa apenas em um fator (DAMÁSIO, 2012). Esse tipo de rotação permite a interpretação dos fatores

separando as variáveis em grupos independentes.

Neste trabalho os dados foram analisados por meio de ACP e em seguida, com base na quantidade de componentes definidos pelos métodos de Kaiser e o *scree plot* foi aplicada AF para definição dos fatores latentes às variáveis de crimes. Ambas as análises foram interpretadas a fim de evidenciar os estados que se destacam na questão de índice de crimes.

## RESULTADOS E DISCUSSÕES

Foi realizado teste de KMO na amostra das categorias de crimes medidos nos estados dos EUA de forma a verificar o nível de adequação dos dados à análise Fatorial. O resultado do teste foi de **0,7487**, este resultado é considerado de adequação aceitável e intermediária da amostra segundo os critérios apresentados na Tabela 1, desta forma é plausível a aplicação de AF aos dados de crimes.

Como é visível na Tabela 2, que contém os autovalores e porcentagens da

variação explicada por cada componente, somente os componentes 1 e 2 já explicam aproximadamente **79,18%** de toda a variação dos dados, sendo **64,23%** do CP1 e os demais **14,94%** pertencendo ao CP2.

A quantidade de dois componentes é considerada suficiente pelo fato de os dois primeiros CPs possuírem autovalores  $>1$  ( $\lambda_i > 1$ ) (KAISER, 1958) responsáveis por

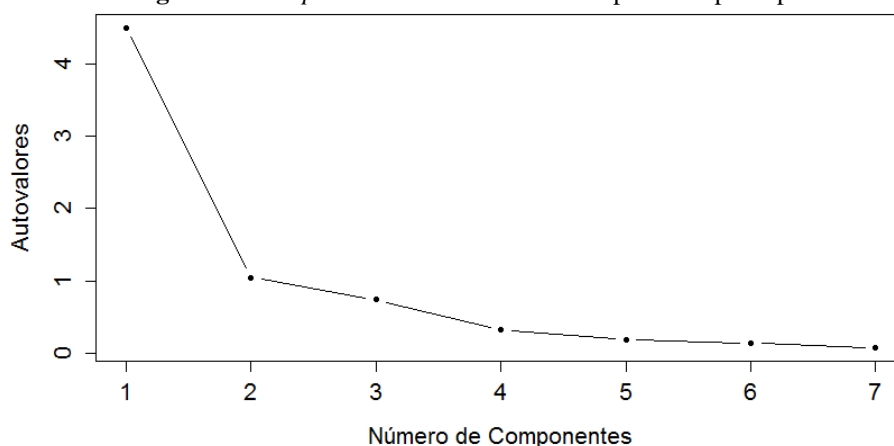
79,18% da variação dos dados. A decisão por dois CPs é reforçada pelo *screeplot* (Figura 1) que apresenta estabilização dos autovalores após o segundo ponto.

Pode-se afirmar com base na análise dos autovalores que os Componentes Principais 1 e 2 resumem a variação dos dados de maneira eficiente.

**Tabela 2:**Componentes principais, autovalores e porcentagem da variância explicada e proporção acumulada pelos componentes.

Componente Principal	Autovalores	Proporção (%)	Proporção Acumulada (%)
PC1	4,50	64,24	64,23
PC2	1,05	14,94	79,18
PC3	0,74	10,54	89,72
PC4	0,32	4,58	94,29
PC5	0,19	2,72	97,01
PC6	0,14	1,94	98,96
PC7	0,07	1,04	100,00

**Figura 1:**Scree plot dos autovalores dos componentes principais.



A redução de 7 variáveis para 2 componentes representando 79,18% da variação total é satisfatória. Por estemotivo

serão montadas as equações 1 e 2 utilizando somente os dois primeiros componentes.

$$CP1 = 0.37X_1 + 0.40X_2 + 0.38X_3 + 0.38X_4 + 0.44X_5 + 0.34X_6 + 0.31X_7 \quad (1)$$

$$CP2 = -0.43X_1 - 0.29X_2 + 0.22X_3 - 0.39X_4 + 0.09X_5 + 0.27X_6 + 0.66X_7 \quad (2)$$

O CP1 pode ser chamado de componente de crimes de estupro e arrombamento uma vez que as variáveis  $X_2$  (Estupro) e  $X_5$  (Arrombamento) se destacam neste componente de acordo com os coeficientes de ponderação e também de correlação contidos na Tabela 2. O CP2, pelo contraste presente entre as variáveis  $X_1$

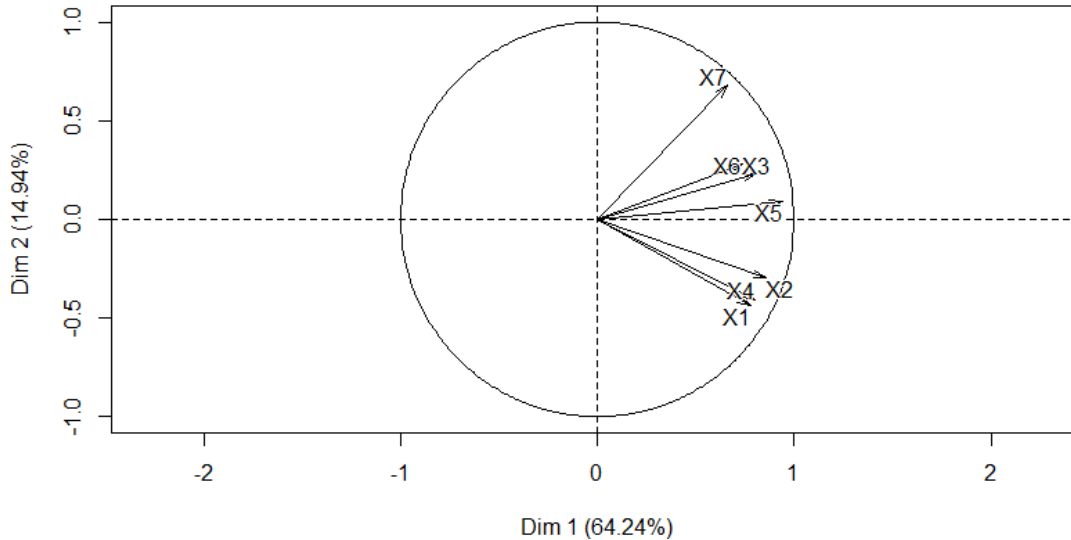
(Assassinato) e  $X_4$  (Assalto) com  $X_7$  (Roubo de Veículo) presente na equação e na Tabela 2, pode ser chamado de componente de contraste de assassinato e assalto com roubo de veículos.

As variáveis  $X_2$  (Estupro) e  $X_5$  (Arrombamento) possuem as maiores contribuições para o CP1, isso se observa

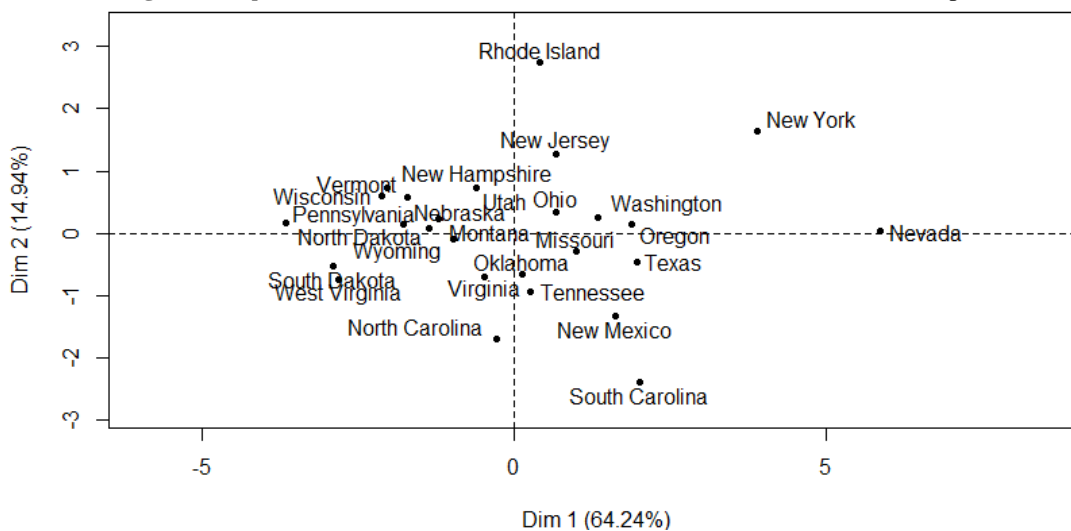
por serem os maiores vetores e também os mais próximos do eixo do CP1 na Figura 2, apesar de serem de tamanho aproximado  $X_2$  dista mais do eixo de CP1 do que  $X_5$ ,

portanto sua contribuição não é tão similar (Figura 2).

**Figura 2:** Biplot CP1 x CP2 sobre as variáveis de criminalidade dos estados dos EUA pela ACP.



**Figura 3:** Biplot CP1 x CP2 com os indivíduos (estados) sobre criminalidade pela ACP.



De acordo com Rencher (2002), o percentual da variância total que deve ser explicada pelos dois primeiros componentes principais é de no mínimo 70%, desta forma os CPs 1 e 2 se mostram suficientes para representação da variação dos dados originais.

De acordo com a Figura 3, pode-se observar que, dos estados dos EUA analisados, *Nevada* possui evidentemente os maiores índices de estupro e arrombamento do país, seguido de *New*

*York*. Os menores índices pertencem aos estados de *Wisconsin*, *Nebraska* e *South Dakota*.

O estado de *Rhode Island* possui os maiores índices de assassinato e assalto em contraste com os menores índices de Roubo a Veículos, a segunda colocação nestes índices fica com *New York* seguido pelo estado de *New Jersey* (Figura 3).

A análise de componentes principais foi efetiva quanto à redução da massa de dados, tendo reduzido a quantidade de

variáveis em dois componentes, permitindo a análise dos dados. Entretanto, a ACP não diferencia a variância específica da variância comum das variáveis, enquanto a Análise Fatorial (AF) considera somente a variância comum (a variância compartilhada entre as variáveis).

Quando se comparam os resultados de uma ACP e uma AF, é perceptível que as cargas fatoriais da ACP são geralmente mais elevadas por considerar a variância específica de cada item (COSTELLO &

OSBORNE, 2005). Desta forma, foi feita Análise Fatorial com a intenção de identificar os fatores latentes que explicam os índices de criminalidade nos 26 estados dos EUA.

Na Tabela 3 são apresentadas as communalidades e variâncias específicas de todas as variáveis. Os valores de communalidade permitem verificar que a Análise Fatorial é uma técnica adequada a ser aplicada nos dados visto que nenhuma dos valores é menor do que 0,60.

**Tabela 3:** Comunalidades e Variância Específica das sete categorias de crime em 26 estados dos EUA.

Variável	Comunalidade	Variância Específica
X1	0,795	0,205
X2	0,827	0,173
X3	0,708	0,292
X4	0,807	0,193
X5	0,891	0,109
X6	0,616	0,384
X7	0,898	0,102

Foi optado por seguir o critério de autovalores maiores que 1 de Guttman-Kaiser (GUTTMAN, 1954; KAISER, 1960) para seleção da quantidade de fatores,

resultando assim em dois fatores conforme as informações analisadas na ACP. Desta forma, foi feita a estimação das cargas para os Fatores 1 e 2, apresentadas na Tabela 4.

**Tabela 4:** Cargas Fatoriais de 2 fatores para as sete categorias de crime em 26 estados dos EUA.

Variável	Fator 1	Fator 2
X1	-0,776	0,439
X2	-0,859	0,299
X3	-0,810	-0,228
X4	-0,802	0,405
X5	-0,940	-0,089
X6	-0,733	-0,280
X7	-0,661	-0,679

Pode-se observar que as cargas fatoriais não surtiram o efeito esperado de separar os fatores nitidamente, uma vez que com exceção de X<sub>7</sub>, todas as demais

variáveis possuem maior carga no Fator 1. Para facilitar a interpretação foi aplicada uma rotação *Varimax* nas cargas e os resultados são apresentados na Tabela 5.

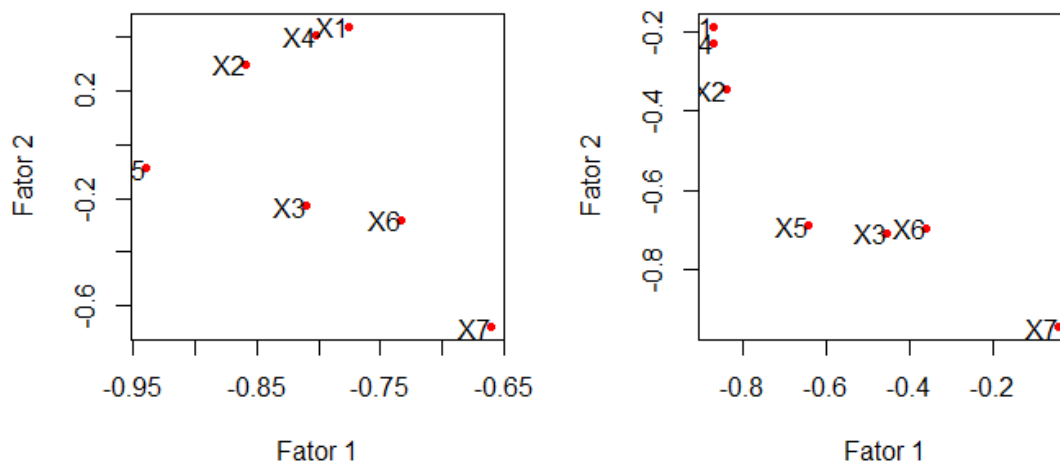
**Tabela 5:** Cargas Fatoriais Rotacionadas pelo método *Varimax* de 2 fatores para as sete categorias de crime em 26 estados dos EUA.

Variável	Fator 1	Fator 2
X1	<b>-0,872</b>	-0,185
X2	<b>-0,841</b>	-0,345
X3	-0,456	<b>-0,708</b>
X4	<b>-0,869</b>	-0,227
X5	-0,645	<b>-0,689</b>
X6	-0,363	<b>-0,696</b>
X7		<b>-0,946</b>

A Tabela 5 contém as cargas rotacionadas, foram marcadas em negrito as maiores dentre os dois fatores para cada variável, o que indica maior explicação da variável pelo Fator. Com a Rotação *Varimax* é possível definir claramente as variáveis de cada fator, com exceção da

variável X<sub>6</sub> que teve a maior carga considerada apesar da diferença ser pequena. Na Figura 4 fica clara a relação entre as variáveis e os fatores antes e depois da rotação, no segundo gráfico as variáveis se dividem entre os eixos enquanto que antes da rotação não havia separação clara.

**Figura 4:** Gráficos de Correspondência para cargas prévias e após a rotação *Varimax* das sete categorias de crime de 26 estados dos EUA.



O Fator 1, representando 64,24% da variância total, é composto pelas variáveis X<sub>1</sub> (Assassinato), X<sub>2</sub> (Estupro) e X<sub>4</sub> (Assalto) e pode ser chamado de Crimes de Ameaça à Vida. As variáveis de violência possuem cargas altas e mesmo Assalto que pode eventualmente envolver roubo é caracterizado pela agressão. Já o Fator 2 representa 14,94% da variação total e é composto pelas variáveis X<sub>3</sub> (Roubo), X<sub>5</sub>

(Arrombamento), X<sub>6</sub>(Pequenos Roubos) e X<sub>7</sub> (Roubo de Veículos). Este fator, por agrupar somente variáveis que caracterizam roubo em geral, pode ser chamado de Fator de Crimes de Roubos (Figura 4 e Tabela 5).

Finalmente, na Tabela 6, são apresentados os scores estimados para os 26 estados por meio do método de regressão tanto para o Fator 1 quanto o Fator 2.



**Tabela 6:** Scores dos Fatores 1 e 2 estimados por método de regressão para 26 estados dos EUA.

Estado	Fator 1	Fator 2
Missouri	0,4595	0,2722
Montana	-0,5529	-0,2337
Nebraska	-0,8161	-0,1334
Nevada	2,7179	-0,029
New Hampshire	-0,9389	-0,7002
New Jersey	0,3186	-1,2263
New Mexico	0,7559	1,2713
New York	1,8039	-1,5795
North Carolina	-0,1277	1,6139
North Dakota	-1,6907	-0,1743
Ohio	0,3151	-0,3324
Oklahoma	0,0643	0,6254
Oregon	0,871	-0,1424
Pennsylvania	-0,6277	-0,0827
Rhode Island	0,1978	-2,6366
South Carolina	0,9296	2,2857
South Dakota	-1,335	0,499
Tennessee	0,1254	0,8824
Texas	0,9103	0,433
Utah	-0,2808	-0,7115
Vermont	-0,7876	-0,5692
Virginia	-0,2189	0,6588
Washington	0,6258	-0,248
West Virginia	-1,2974	0,7184
Wisconsin	-0,9732	-0,5767
Wyoming	-0,4497	0,0815

Os valores de score para o Fator 1, que é o mais importante por conta de explicar 64,24% da variação dos dados, indicam que o estado de *Nevada* lidera os 26 estados no índice de crimes de ameaça à vida seguida pelo estado de *New York* em segundo lugar, o estado de *South Carolina* é o terceiro colocado. Dentre os estados com menor índice de crimes de ameaça à vida o primeiro colocado é o estado de *North Dakota* seguido por *South Dakota* que fica em segundo lugar e em terceiro lugar o estado de *West Virginia* (Tabela 6).

Os valores de score para o Fator 2, que explica 14,94% da variação dos dados, mostram um alto índice de crimes de roubo em *South Carolina*, sendo o primeiro dos 26 estados. O segundo estado com maior índice de crimes de roubo é *North Carolina*, *New Mexico* fica em terceiro lugar dentre os maiores índices de crimes de

roubo. Dentre os estados com menores índices de crimes de roubo *Rhode Island* se encontra em primeiro lugar, *New York* é o estado com segundo menor índice de roubos dentre os 26 analisados e em terceiro lugar fica o estado de *New Jersey*.

Todos os resultados mostrados nos scores de fatores podem ser confirmados analisando-se a os gráficos de *biplo* de variáveis (Figura 2) em conjunto com o de indivíduos (Figura 3).

## CONCLUSÃO

De acordo com os resultados, a Análise de Componentes Principais aplicada foi efetiva pelo fato de ter reduzido a quantidade de variáveis de 7 para dois componentes que explicam parte satisfatória dos dados originais. A redução se dá pelo fato de cinco variáveis possuírem

correlação entre si ou pouca variabilidade, desta forma atribuindo redundância aos dados analisados.

A Análise Fatorial permitiu agrupar os fatores com base no comportamento (Crimes de Violência e de Roubo) pode-se afirmar que este método foi eficaz e cumpriu os objetivos propostos que são os de Resumo e Redução dos dados além de permitir a interpretação dos mesmos.

O principal objetivo de ambas as análises foi cumprido: Redução do número de variáveis sem perda significativa de informação sobre a variabilidade do conjunto de variáveis originais permitindo a interpretação dos dados de forma simplificada.

## REFERÊNCIAS

- BROWN, Timothy A. Confirmatory factor analysis for applied research. 2006. Confirmatory factor analysis for applied research. xiii, 2006.
- CATTELL, Raymond B. The scree test for the number of factors. *Multivariate behavioral research*, 1966, 1.2: 245-276.
- DAMÁSIO, Bruno Figueiredo. Uso da análise fatorial exploratória em psicologia. *Avaliação psicológica*, 2012, 11.2: 213-228.
- FARIAS, Christiano Alves; FIGUEIREDO, Adelson Martins; LIMA, João Eustáquio de. Dependência espacial e análise de agrupamento de municípios para diferentes tipos de crime em Minas Gerais. *REUNA*, 2010, 13.3.
- GULUMBE, Shehu U.; DIKKO, H. G.; BELLO, Yusuf. Analysis of Crime Data using Principal Component Analysis: A case study of Katsina State. *CBN Journal of Applied Statistics*, 2013, 3.2: 39-49.
- GUTTMAN, Louis. Some necessary conditions for common-factor analysis. *Psychometrika*, 1954, 19.2: 149-161.
- HONGYU, Kuang. Distribuição empírica dos autovalores associados à matriz de interação dos modelos AMMI pelo método bootstrap não-paramétrico. 2012. PhD Thesis. Escola Superior de Agricultura “Luiz de Queiroz.
- HONGYU, Kuang. 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, Kuang; SANDANIELO, Vera Lúcia Martins; DE OLIVEIRA JUNIOR, Gilmar Jorge. Análise de Componentes Principais: Resumo Teórico, Aplicação e Interpretação. *E&S Engineering and Science*, 2016, 5.1: 83-90.
- JOHNSON, Richard Arnold, et al. *Applied multivariate statistical analysis*. Upper Saddle River, NJ: Prentice hall, 2002.
- KAISER, Henry F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 1958, 23.3: 187-200.

OSBORNE, Jason W.; COSTELLO, Anna B. Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Pan-Pacific Management Review*, 2009, 12.2: 131-146.

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

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, notas de aula INF 766, Departamento de Informática da Universidade Federal de Viçosa, v.2, 2000.

RENCHEER, Alvin C.; CHRISTENSEN, W. F. *Methods of Multivariate Analysis* ; JOHN WILEY&SONS. INC.: Toronto, ON, Canada, 2002.

SAS Institute Inc. 2008. SAS/STAT® 9.2 User's Guide. Cary, NC: SAS Institute Inc. 1st electronic book.

THURSTONE, Louis Leon. *Multiple factor analysis*. 1947.