

## Análise Fatorial Exploratória: resumo teórico, aplicação e interpretação

### Exploratory Factorial Analysis: theoretical summary, application and interpretation

<sup>1</sup>Kuang Hongyu

<sup>1</sup>Professor Adjunto do Departamento de Estatística/Instituto de Ciências Exatas e da Terra - Universidade Federal de Mato Grosso (prof.kuang@gmail.com)

---

**RESUMO:** A análise multivariada de uma forma geral refere-se aos métodos estatísticos que analisam simultaneamente múltiplas medidas em cada indivíduo ou objeto sob investigação. Nesse contexto, entre as técnicas de multivariadas, a análise fatorial exploratória (AFE) é umas das técnicas estatísticas mais utilizadas na análise de dados em diversas áreas do conhecimento, como agronomia, zootecnia, ecologia, floresta, medicina, entre outras. AFE é uma técnica dentro da análise fatorial cujo objetivo abrangente é identificar as relações subjacentes entre as variáveis medidas. Assim, o objetivo principal deste trabalho é apresentar os principais aspectos desta técnica, como resumo teórico, apresentando uma aplicação sobre os dados de crimes do Estado de Mato Grosso e nas interpretações dos resultados obtidos por diferentes métodos da AFE. Com este trabalho pretende auxiliar pesquisadores das diversas áreas a utilizarem AFE com maior discernimento teórico e metodológico.

**Palavras Chave:** Análise multivariada, análise componente principal, rotação *Varimax*.

**ABSTRACT:** Multivariate analysis is based on the statistical methods that simultaneously analyzing the multiple measures in each individual or subject which under investigation. Among all the multivariate techniques in this context, the exploratory factorial analysis (AFE) is one of the statistical techniques most frequent used in data analysis. It is used frequently in several areas of analysis, such as agronomy, zootechny, ecology, forest, medicine, among others. AFE's overarching goal is to identify the underlying relationships between measured variables. Thus, as an example of theoretical summary with an application about the crime data of the State of Mato Grosso, the AFE presents different methods to obtain the results of the interpretation. This work aims to help researchers from different areas to use AFE with greater theoretical and methodological insight.

**Keywords:** Multivariate analysis, principal component analysis, Varimax rotation.

---

## 1. INTRODUÇÃO

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 (NEISSE; HONGYU, 2016). 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 como agronomia, zootecnia, ecologia, florestal, entre outras. Dentre as técnicas multivariadas, este trabalho foca na análise fatorial (AF) ou "*Factor Analysis*", que é a principal e a mais antiga técnica de análise multivariada.

A origem da análise fatorial foi proposta por Spearman e Pearson no início do século 20, quando Spearman (1904) desenvolveu um método para a criação de um índice geral de inteligência (fator "g") com base nos resultados de vários testes (escalas), que supostamente refletiriam essa aptidão. Tratava-se de um primeiro método de AF, adequado para a estimação de um único fator.

A análise fatorial exploratória (AFE) ou "*exploratory factoranalysis*" é uma técnica dentro da análise fatorial cujo objetivo abrangente é identificar as relações subjacentes entre as variáveis medidas. A AFE é uma técnica estatística que estuda correlações entre um grande número de variáveis agrupando-as em fatores. Essa técnica permite a redução de dados,

identificando as variáveis mais representativas ou criando um novo conjunto de variáveis, bem menor que o original (HAIR *et al.*, 2009; KIRCH *et al.*, 2017).

Durante a realização de AFE, diversas decisões precisam ser tomadas a fim de se obter uma estrutura fatorial adequada (COSTELLO; OSBORNE, 2005). Uma vez que os resultados obtidos nas AFEs dependem, em grande medida, das decisões tomadas pelo pesquisador, a técnica possui um alto potencial de produzir resultados errôneos e/ou não confiáveis (PATIL *et al.*, 2008). Assim, todas as decisões tomadas durante a realização de uma AFE devem ser pautadas em critérios teóricos e metodológicos claros, buscando a obtenção de modelos fatoriais adequados (DAMÁSIO, 2012).

Uma situação comum em diversas áreas do conhecimento é aquela na qual, para cada elemento de uma amostra, observa-se um grande número de variáveis. Essas variáveis podem ser, por exemplo, um conjunto de itens de uma escala ou os escores obtidos por um indivíduo em diferentes escalas de avaliação. Diante de um quadro como esse, o pesquisador enfrenta dois problemas, que podem ser resolvidos por meio da AFE: a) a caracterização dos avaliados, levando-se em conta um conjunto eventualmente grande de variáveis, e b) a descrição da inter-relação dessas variáveis, eventualmente explicitando uma estrutura de interdependência subjacente aos dados (HAIR *et al.*, 2009).

O objetivo principal deste trabalho é apresentar os principais aspectos desta técnica, na forma de resumo teórico, fazendo-se posteriormente uma aplicação prática sobre os dados de crimes do Estado de Mato Grosso, na qual foca-se nas interpretações dos resultados obtidos por diferentes métodos da AFE. Com este trabalho, pretende auxiliar pesquisadores das diversas áreas a utilizarem AFE com maior discernimento teórico e metodológico.

## 2. MATERIAIS E MÉTODOS

O conjunto de dados utilizado neste (artigo) foram obtidos por meio do site da Secretaria de Estado de Planejamento (SEPLAN) de Mato Grosso, cujo levantamento foi realizado em 2015 pela Gerência de Estatística e Indicadores, por meio de dados de ocorrências e produtividades registrados pela Polícia Militar do Estado de Mato Grosso e cadastrados no Sistema Estatístico de Ocorrência e Produtividade da Polícia Militar (SEOP/PM) (<http://www.seplan.mt.gov.br/>).

Os dados contêm informações dos 56 municípios do estado de Mato Grosso, conforme a classificação do Instituto Brasileiro de Geografia e Estatística (IBGE). Para neste estudo foi utilizado a técnica de amostragem aleatória simples para obter os 56 municípios e as 8 variáveis. As variáveis trabalhadas foram: Ameaça (X1), Lesão Corporal (X2), Homicídio (X3), Roubo a pessoa (X4), Roubo de veículo (X5), Roubo em residência (X6), Roubo Estabelecimento Comercial (X7) e Estupro (X8).

Todas as análises deste artigo foram feitas por meio de rotinas computacionais implementadas no software R 3.5.1 (R Development Core Team, 2018) com os pacotes “*corrplot*”, “*vegan*”, “*factoextra*” e “*psych*”.

### 2.1 Pressupostos das Análises Fatoriais Exploratórias

Os pressupostos da AFE é verificar se a aplicação da análise fatorial tem validade para as variáveis escolhidas, sendo justificado pela pouca quantidade de respondentes da pesquisa. Para isso, dois métodos de avaliação são mais comumente utilizados, o critério de Kaiser-Meyer-Olkin (KMO) e o Teste de Esfericidade de Bartlett (DZIUBAN; SHIRKEY, 1974).

O índice de KMO, também conhecido como índice de adequação da amostra, é um teste estatístico que sugere a proporção de variância dos itens que pode estar sendo explicada por uma variável latente, tal índice indica o quão adequada é a aplicação da AFE para o conjunto de dados (HAIR *et al.*, 2009; LORENZO-SEVA; TIMMERMAN; KIERS, 2011). O

KMO é calculado por meio do quadrado das correlações totais dividido pelo quadrado das correlações parciais, das variáveis analisadas, cuja expressão é dada na Equação 1 (DZIUBAN; SHIRKEY, 1974):

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

em que  $r_{jm}^2$  é o coeficiente de correlação linear entre  $X_j$  e  $X_m$ ;  $r_{pjm}^2$  é o coeficiente de correlação parcial amostral entre  $X_j$  e  $X_m$ , definido como sendo o coeficiente de correlação linear entre os resíduos.

O KMO pode variar de zero a um. Valores iguais ou próximos a zero indicam que a soma das correlações parciais dos itens avaliados é bastante alta em relação à soma das correlações totais. Nesses casos, possivelmente a análise fatorial será inapropriada (PASQUALI, 1999). Como regra para interpretação dos índices de KMO, valores menores que 0,5 são considerados inaceitáveis, valores entre 0,5 e 0,7 são considerados medíocres; valores entre 0,7 e 0,8 são considerados bons; valores maiores que 0,8 e 0,9 são considerados ótimos e excelentes, respectivamente (HUTCHESON; SOFRONIOU, 1999; PEREIRA, 1999).

A medida de adequação da amostra (MAA) é considerável aceitável e meritória quando  $MAA > 0,5$ , sendo esta, avaliada por meio da Equação 2 para cada variável  $X_j$ :

$$MAA_j = \frac{\sum_{l=1, l \neq j}^p r_{jl}^2}{\sum_{l=1, l \neq j}^p r_{jl}^2 + \sum_{l=1, l \neq j}^p r_{pjl}^2}, \quad j = 1, \dots, p \quad (2)$$

O teste de esfericidade de Bartlett avalia em que medida a matriz de covariância é similar a uma matriz identidade, ou seja, não apresentam correlações entre si (FIELD, 2005). Esse teste avalia, também, a significância geral de todas as correlações em uma matriz de dados (HAIR *et al.*, 2009). Os valores do teste de esfericidade de Bartlett com níveis de significância ( $p < 0,05$ ) indicam que a matriz é fatorável, rejeitando a hipótese nula de que a matriz de dados é similar a uma matriz identidade (TABACHNICK; FIDELL, 2007). Em geral, os resultados dos testes de KMO e de esfericidade de Bartlett tendem a ser uniformes, aceitando ou negando a possibilidade de fatoração da matriz de dados (DZIUBAN; SHIRKEY, 1974).

## 2.2 Modelo teórico da análise fatorial

O modelo da análise de fatores pode ser escrito conforme Equação 3 (JOHNSON; WICHERN, 2008):

$$X_i = a_{i1}F_1 + a_{i2}F_2 + \dots + a_{im}F_m + e_i \quad (3)$$

em que  $X_i$  é o  $i$ -ésimo escore depois dele ter sido padronizado (média 0 e desvio-padrão 1);  $i = 1, \dots, p$ ;  $p$  é o número de variáveis;  $a_{i1}, a_{i2}, \dots, a_{im}$  são as cargas dos fatores para o  $i$ -ésimo teste;  $F_1, F_2, \dots, F_m$  são  $m$  fatores comuns não correlacionados, cada um com média 0 e variância 1 e  $e_i$  é um fator específico somente para o  $i$ -ésimo teste que é não correlacionado com qualquer dos fatores comuns e tem média zero.

Os  $p$  valores observados  $X_p$  são expressos em termos de  $p + m$  variáveis aleatórias não observáveis ( $F_1, F_2, \dots, F_m; \varepsilon_1, \varepsilon_2, \dots, \varepsilon_p$ ). Isso distingue o modelo fatorial do modelo de regressão múltipla, no qual as variáveis independentes podem ser observadas, e cujas posições são ocupadas por  $F$  no modelo fatorial. Matricialmente, o modelo é expresso pela Equação 4:

$$\mathbf{X}_{(p \times 1)} = \mathbf{\Lambda}_{(p \times m)} \mathbf{F}_{(m \times 1)} + \boldsymbol{\varepsilon}_{(p \times 1)} \quad (4)$$

Já o modelo AF é expresso na Equação 5:

$$\begin{aligned} \text{Var}(X_j) &= a_{j1}^2 \text{Var}(F_1) + a_{j2}^2 \text{Var}(F_2) + \dots + a_{jm}^2 \text{Var}(F_m) + \text{Var}(e_j) \\ &= a_{j1}^2 + a_{j2}^2 + \dots + a_{jm}^2 + \text{Var}(e_j) \end{aligned} \quad (5)$$

em que  $a_{j1}^2 + a_{j2}^2 + \dots + a_{jm}^2$  é chamada a comunalidade de  $X_j$  (a parte da sua variância que é explicada pelos fatores comuns). A comunalidade não pode exceder 1, sendo necessário que fique compreendida entre  $-1 \leq a_{ij} \leq +1$ . Pode também ser estabelecido que a correlação entre  $X_j$  e  $X_j$ , seja dada pela Equação 6:

$$r_{jj'} = a_{j1}a_{j'1} + a_{j2}a_{j'2} + \dots + a_{jm}a_{j'm} \quad (6)$$

Consequentemente, duas variáveis somente serão altamente correlacionadas se elas tiverem altas cargas no mesmo fator (JOHNSON; WICHERN, 2008; NEISSE; HONGYU, 2016).

### 2.3 Retenção de Fatores

Diversos procedimentos e critérios de retenção fatorial foram desenvolvidos (DAMÁSIO, 2012). Dentre eles, o critério de Kaiser-Guttman (GUTTMAN, 1954; KAISER, 1960) é um dos mais utilizados, também conhecido como critério do *eigenvalue* (autovalor) maior do que 1,0 ( $\lambda_i > 1$ ). Tal critério propõe uma avaliação rápida, objetiva do número de fatores a ser retido e cada fator retido apresenta um autovalor que se refere ao total de variância explicada por este fator (DAMÁSIO, 2012).

Existem outros critérios de retenção de Fatores na literatura, como o critério baseado no gráfico do *screeplot*, ou também conhecido como teste de Cattell (1966). O critério pelo *screeplot* é encontrar o ponto "cotovelo" no gráfico em que os autovalores apresentam uma tendência desce linear. Na análise fatorial com a utilização do critério de Kaiser–Guttman ou *screeplot*, geralmente, apresentam resultados bem semelhantes entre si.

O terceiro critério é conhecido como o método das análises paralelas (AP) (HORN, 1965), que vem sendo cada vez mais consolidado na literatura internacional, com implementação no *software R* no pacote “*psych*”. O método da AP foi desenvolvido para ser utilizado como critério de retenção de componentes. Entretanto, tem sido adaptado para o uso no contexto das AFE (Crawford *et al.*, 2010), e é considerado um procedimento adequado para determinar o número de fatores a serem retidos (LORENZO-SEVA; TIMMERMAN; KIERS, 2011). O método de extração da AP (HORN, 1965) consiste em comparar os valores dos *eigenvalues* da amostra piloto com os valores gerados a partir de uma amostra aleatória do mesmo tamanho. Apenas são retidos os *eigenvalues* com valor superior aos gerados na amostra aleatória. Por fim, o critério baseado no teste qui-quadrado, de Barlett (1950) e o critério de Velicer (1976), baseado nas correlações parciais.

### 2.4 Procedimentos gerais para a análise fatorial

Existem três estágios na análise de fatores. O problema inicial é a determinação das cargas dos fatores ( $a_{jk}$ ) com  $j = 1$  a  $p$  e  $k = 1$  a  $m$  ( $m < p$ ). Entretanto, a primeira etapa é definir os métodos de estimação das cargas dos fatores. Existem vários métodos de estimação referentes à análise fatorial na literatura, como Máxima Verossimilhança (*Maximum Likelihood*, ML); Eixos Principais fatoriais (*Principal Axis Factoring*, PAF); Componente Principal (CP); Mínimos Quadrados Generalizados; Mínimos Quadrados Não Ponderados e Fatoração Alfa.

Em geral, os métodos de componente principal e máxima verossimilhança fornecem os melhores resultados e mais recomendados para análise fatorial quando as amostras

apresentam distribuição normal e não-normal, respectivamente (COSTELLO; OSBORNE, 2005; JOHNSON; WICHERN, 2008).

O PAF é usualmente padronizado para ter uma média zero e uma variância igual à variância total considerada, o componente principal é o mesmo eixo principal, exceto que sua média não é padronizada para zero. O fator principal é normalizado para ter uma média zero e variância unitária (KIM, 1975). De acordo com HARMAN (1967), a solução da fatoraçoão com valores unitários nas diagonais da matriz de correlaçoão pode ser chamada de soluçoão de componente principal, e a soluçoão com comunalidades nas diagonais da matriz de correlaçoão é denominada de soluçoão do fator principal.

Definido o método usado para a determinaçoão das cargas dos fatores, é possível mostrar que tais cargas não são únicas. Se  $F_1, F_2, \dots, F_m$  são os fatores provisórios, então combinaçoões lineares desses fatores podem ser expressos na forma apresentada na Equaçoão 7 (JOHNSON; WICHERN, 2008):

$$\begin{aligned} F'_1 &= d_{11}F_1 + d_{12}F_2 + \dots + d_{1m}F_m \\ F'_2 &= d_{21}F_1 + d_{22}F_2 + \dots + d_{2m}F_m \\ &\dots \\ F'_m &= d_{m1}F_1 + d_{m2}F_2 + \dots + d_{mm}F_m \end{aligned} \quad (7)$$

Podem ser construídos de modo a serem não correlacionados e explicar os dados tão bem quanto os fatores provisórios. De fato, há uma infinidade de soluçoões alternativas para o modelo de análise de fatores. Isto leva ao segundo estágio na análise: rotaçoão de fator. Rotaçoão de fatores é consiste em encontrar os valores  $d_{ij}$ . Dessa forma, os fatores provisórios são transformados, com o objetivo de se encontrar novos fatores, mais fáceis de serem interpretados.

Depois de verificar a adequabilidade da base de dados e determinar a técnica de extraçoão e o número dos fatores, o pesquisador deve seguir para o próximo estágio: decidir o tipo de rotaçoão dos fatores. Existem dois principais tipos de rotaçoão: ortogonal e oblíqua. Por um lado, as rotaçoões ortogonais são mais fáceis de reportar e de interpretar. No entanto, o pesquisador deve assumir que os construtos são independentes (na prática esse pressuposto é mais difícil de ser respeitado).

Por outro lado, as rotaçoões oblíquas permitem que os fatores sejam correlacionados. Todavia, são mais difíceis de descrever e interpretar. De acordo com Hair *et al.* (2009), em geral, as duas formas de rotaçoão produzem resultados bastante semelhantes. Dentre esse tipo de rotaçoão, vários métodos são apresentados na literatura, tais como quartimax; equimax e *varimax*. Os métodos quartimax e equimax não foram bem sucedidos, sendo raramente utilizados na literatura. O *varimax* é o mais bem sucedido e o mais utilizado nas pesquisas entre os métodos ortogonais (DAMÁSIO, 2012).

O método *varimax* foi proposto por Kaiser (1958), quando a variância atinge o máximo, o fator tem maior interpretabilidade ou simplicidade, no sentido de que as cargas deste fator tendem ou à unidade, ou à zero. O critério de máxima simplicidade de uma matriz fatorial completa é definido como a maximizaçoão da soma destas simplicidades. Portanto, este método requer que as cargas dos fatores finais sejam tais que maximizem a função apresentada na Equaçoão 8:

$$V = \sum_{k=1}^m s_k^2 = \frac{1}{P} \sum_{k=1}^m \sum_{j=1}^p \left( \frac{a_{jk}^2}{h_j^2} \right)^2 - \frac{1}{P^2} \sum_{k=1}^m \left( \sum_{j=1}^p \frac{a_{jk}^2}{h_j^2} \right)^2 \quad (8)$$

em que  $h_j^2$  é a comunalidade da variável  $j$ . Ou, de uma maneira mais simples, após a multiplicaçoão da expressaçoão acima por  $p^2$ , já que a multiplicaçoão por uma constante não afeta o processo de maximizaçoão conforme Equaçoão 9:

$$V = p \sum_{k=1}^m \sum_{j=1}^p \left( a_{jk} / h_j \right)^4 - \sum_{k=1}^m \left( \sum_{j=1}^p a_{jk}^2 / h_j^2 \right)^2 \quad (9)$$

Esta expressão foi chamada por Kaiser como critério *varimax* normal ou simplesmente critério *varimax*.

O último estágio de análise fatorial envolve calcular os escores dos fatores. Estes são os valores dos fatores rotacionados  $F'_1, F'_2, \dots, F'_m$  para cada um dos  $n$  indivíduos para os quais os dados estão disponíveis. Os escores dos fatores são muitas vezes usados na construção de índices, para fazer diagnósticos ou como entrada em análises subsequentes. Existem dois métodos de estimação dos escores fatoriais: o método dos mínimos quadrados ponderados e o método de regressão (JOHNSON; WICHERN, 2008).

### 2.5 Análise Fatorial por Componentes Principais

Partindo-se da matriz  $\mathbf{X}_{n \times p}$ , realiza-se uma análise de componentes principais, obtendo-se  $p$  componentes principais (JOHNSON; WICHERN, 2008) conforme Equação 10:

$$Z_i = b_{i1}X_1 + b_{i2}X_2 + \dots + b_{ip}X_p \quad (10)$$

em que  $i = 1, \dots, p$  e  $b_{ij}$  são dados pelos autovetores da matriz de correlações. Por se uma transformação ortogonal, o relacionamento inverso é dado pela Equação 11:

$$X_i = b_{i1}Z_1 + b_{i2}Z_2 + \dots + b_{ip}Z_p \quad (11)$$

O modelo de  $m < p$  fatores fica então representado pela Equação 12:

$$X_i = b_{i1}Z_1 + b_{i2}Z_2 + \dots + b_{im}Z_m + e_i \quad (12)$$

em que  $e_i$  é uma combinação linear dos componentes de  $Z_{m+1}$  a  $Z_p$ .

Divide-se  $Z_i$  por seu desvio padrão  $\sqrt{\lambda_i}$  (obtendo-se  $F_i$ ), a raiz quadrada do correspondente autovalor na matriz de correlações. As equações são calculadas por meio da Equação 13:

$$X_i = \sqrt{\lambda_1}b_{i1}F_1 + \sqrt{\lambda_2}b_{i2}F_2 + \dots + \sqrt{\lambda_m}b_{im}F_m + e_i \quad (13)$$

Supondo que  $a_{ij} = \sqrt{\lambda_j}b_{ji}$ , é possível demonstrar que  $X_i$  é dado pela Equação 14:

$$X_i = a_{i1}F_1 + a_{i2}F_2 + \dots + a_{im}F_m + e_i \quad (14)$$

Procedendo a rotação *varimax* conforme Equação 14, chega-se que  $X_i$  é dado pela Equação 15:

$$X_i = g_{i1}F'_1 + g_{i2}F'_2 + \dots + g_{im}F'_m + e_i \quad (15)$$

em que  $F'_i$  representa o novo  $i$ -ésimo fator. Os escores fatoriais são calculados por mínimos quadrados ponderados para estimativas de máxima verossimilhança (Equação 16):

$$\mathbf{F}^* = (\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'\mathbf{X}' \quad (16)$$

em que  $\mathbf{F}^*$  é uma matriz  $n \times m$ , contendo os valores para os  $m$  fatores rotacionados em sua coluna;  $\mathbf{X}$  é a matriz  $n \times p$  de dados originais e padronizados;  $\mathbf{G}$  é a matriz  $p \times m$  das cargas fatoriais rotacionados (JOHNSON; WICHERN, 2008; HAIR *et al.*, 2009).

### 3. RESULTADOS/DISCUSSÕES

Pela Figura 1, apresenta-se os valores e cores das correlações, sendo que quando a correlação é próxima de 1, esta apresenta a cor azul escura e, quando o valor está próximo de -1, esta apresenta cor vermelha escura. As variáveis de criminalidades dos municípios do estado de Mato Grosso apresentaram correlações altas, principalmente entre as variáveis de ameaça e lesão corporal (X1 e X2) com correlação de 0,99; ameaça e roubo a pessoa (X1 e X4): 0,99; ameaça e estupro (X1 e X8): 0,98 e lesão corporal com roubo a pessoa (X2 e X4) correlação de 0,98. As menores correlações foram entre roubo estabelecimento comercial (X7) com estupro (X8) correlação de 0,6; X2 e X7 correlação de 0,61 e X3 e X7 com correlação de 0,63, portanto, os registros dos crimes dessas variáveis não foram muito correlacionados.

Caso uma variável seja estatisticamente independente das demais, o pesquisador pode cogitar excluí-la da análise, por exemplo, uma variável apresenta valores baixos de correlações com as demais. Em uma perspectiva mais conservadora, o pesquisador poderia excluí-la e estimar novamente a análise fatorial. Isso porque como a análise fatorial depende do padrão de correlação entre as variáveis observadas, espera-se que variáveis estatisticamente independentes não contribuam para a construção de um fator comum. Portanto, o teste final para a inclusão ou exclusão da variável será o nível de associação entre a variável e o fator extraído, sinalizado pelo valor da comunalidade (HAIR *et al.*, 2009).

Figura 1 - Gráfico de correlação das variáveis de criminalidade do estado de Mato Grosso



Em relação ao tamanho da amostra para AFE, quanto maior, mais adequada para esta se torna para a realização das análises. Hair *et al.* (2009) sugerem que a amostra deve ser superior a 50 observações, sendo aconselhável no mínimo 100 casos para assegurar resultados mais robustos. A razão entre o número de casos e a quantidade de variáveis deve exceder cinco para um ou mais. A medida global de adequação amostral do critério de KMO (Kaiser-Meyer-Olkin) varia entre 0 e 1, quanto mais perto de 1, tanto melhor. Pela Tabela 1, o KMO foi igual 0,7056, logo, isso indica que os fatores encontrados na AF conseguem descrever, satisfatoriamente, as variações dos dados originais. Hair *et al.* (2009) sugerem 0,50 como patamar aceitável.

Outro teste que pode ser avaliado é o Teste de Esfericidade de Bartlett (DZIUBAN; SHIRKEY, 1974) que indica se existe relação suficiente entre as variáveis para aplicação da

AF. Para que seja possível a aplicação da AF recomenda-se que o valor-p seja menor que 5%. Na Tabela 1 apresentou valor- $p < 0,001$ , logo, pelo teste de esfericidade constata-se que os dados são adequados para a aplicação da AF nas variáveis analisadas, indicando que a matriz é fatorável, rejeitando a hipótese nula de que a matriz de dados é similar a uma matriz-identidade.

Tabela 1 - A estatística KMO e teste de Bartlett

Critério de Kaiser-Mayer-Olkin		0,70563
Teste de Esfericidade de Bartlett	Qui-quadrado aprox.	1850,998
	gl	28
Valor-p		<0,001

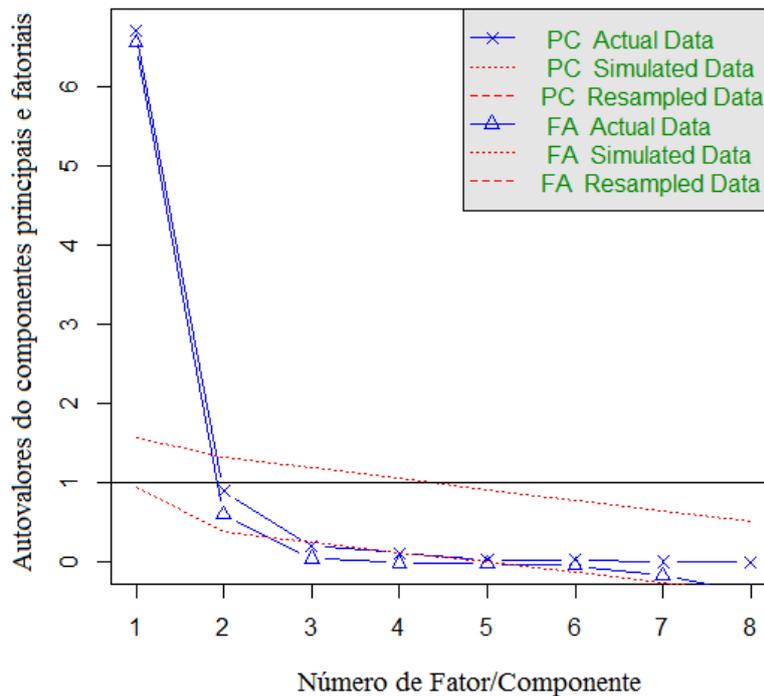
Em geral, os resultados dos testes de KMO e de esfericidade de Bartlett tendem a ser uniformes, aceitando ou negando a possibilidade de fatoração da matriz de dados (DZIUBAN; SHIRKEY, 1974; DAMÁSIO, 2012). As medidas de adequação amostral (MMA) foram feitas para cada uma das oito variáveis. Destaca-se que quando maior o valor da MMA mostra, mais adequado a variável é para essa análise. Todas as variáveis apresentaram MMA  $> 0,05$ , portanto, deve-se prosseguir com a AFE, principalmente as variáveis X8, X1 e X4 que apresentaram maiores valores, respectivamente (Tabela 2).

Tabela 2 - Medida de adequação amostral entre as 8 variáveis estudadas

Variáveis	Medida de adequação amostral (MMA)
X1	0,8136
X2	0,6748
X3	0,6295
X4	0,8082
X5	0,6657
X6	0,6114
X7	0,5468
X8	0,9758

O *Screeplot* apresentado na Figura 2 demonstra os autovalores dos componentes principais e da fatorial, sendo estes gerados pelo método das análises paralelas (AP) de Horn (HORN, 1965). De acordo com Laros (2004), a AP é um procedimento estatístico de simulação Monte-Carlo que consiste na construção aleatória de um conjunto hipotético de matrizes de correlação de variáveis, utilizando como base a mesma dimensionalidade (o mesmo número  $p$  de variáveis e o mesmo número  $n$  de sujeitos) do conjunto de dados reais. Pela análise da Figura 2, nota-se que o critério AP indicou que dois fatores foram ideais para ser extraídos para continuação da análise fatorial.

Figura 2 - Screeplot das análises de componentes principais e fatorial



Pela Tabela 3, apresenta-se os resultados de autovalores e variâncias explicadas e acumuladas pelos os três métodos de estimação de Componente Principal, Fator Principal e Máxima Verossimilhança. A inspeção nos autovalores ( $\lambda_i$ ) demonstra que o primeiro fator pelo método de Componente Principal consegue captar uma quantidade significativa da variação dos dados:  $\lambda_1 = 4,59$ . Como o total de variáveis do conjunto original é 8, isso significa que somente o primeiro autovalor explica  $4,59/30 = 57,38\%$  da variação. Como o segundo maior autovalor é  $\lambda_2 = 3,02$ , este explica  $37,75\%$  da variância dos dados originais.

Dessa forma, os três métodos apresentaram autovalores e variâncias e variâncias acumuladas próximos para os dois fatores, sendo que o método de Componente Principal apresentou a maior variância acumulada  $95,13\%$  para os dois fatores, enquanto os métodos Fator Principal e Máxima Verossimilhança apresentaram ( $94,26\%$  e  $94\%$ ) (Tabela 3).

Tabela 3 - Comunalidades e fatores iniciais, extraídos pelos Métodos do Componente Principal e Fator principal, com os seus respectivos autovalores e percentuais de variância explicada

Fator	Autovalor	Var. Expl. <sup>2</sup> (%)	Var. Acum. <sup>3</sup> (%)
Método de Componente Principal			
1	4,59	57,38	57,38
2	3,02	37,75	95,13
Método do Fator Principal			
1	4,57	57,13	57,13
2	2,97	37,13	94,26
Método da Máxima Verossimilhança			
1	4,56	57,0	57,0
2	2,96	37,0	94,0

<sup>1</sup>Variância Explicada; <sup>2</sup>Variância Acumulada.

Na Tabela 4, apresenta-se as cargas fatoriais e comunalidades do Fator 1 e 2 dos métodos de Componente Principal, Fator Principal e Máxima Verossimilhança após da rotação *Varimax*. 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. As comunalidades representam a proporção da variância para cada variável incluída na análise que é explicada pelos componentes extraídos (SCHAWB, 2007). Por exemplo, X1 e X4 apresentaram maiores comunalidades pelo Componente Principal, logo, os dois fatores extraídos explicam 99% da variância da X1 e X4. Pelo método Fator Principal, os dois fatores explicam 100% da variância da X4 e X7 e pela Máxima Verossimilhança, explicam 100% das variáveis X1 e X7 (Tabela 4).

Tabela 4 - Cargas iniciais dos dois primeiros fatores extraídos pelos dois métodos de estimação e pela rotação *Varimax*

Variáveis	Fator 1	Fator 2	Comunalidades ( $h^2$ )
Método de Componente Principal			
X1	<b>0,92</b>	0,37	0,99
X2	<b>0,91</b>	0,36	0,96
X3	<b>0,85</b>	0,43	0,90
X4	<b>0,89</b>	0,44	0,99
X5	0,52	<b>0,80</b>	0,91
X6	0,43	<b>0,85</b>	0,91
X7	0,29	<b>0,95</b>	0,98
X8	<b>0,93</b>	0,34	0,97
Método do Fator Principal			
X1	<b>0,92</b>	0,38	0,99
X2	<b>0,90</b>	0,37	0,95
X3	<b>0,82</b>	0,44	0,87
X4	<b>0,90</b>	0,44	1,00
X5	0,53	<b>0,75</b>	0,85
X6	0,46	<b>0,78</b>	0,82
X7	0,26	<b>1,00</b>	1,00
X8	<b>0,92</b>	0,35	0,97
Método da Máxima Verossimilhança			
X1	<b>0,92</b>	0,38	1,00
X2	<b>0,92</b>	0,36	0,97
X3	<b>0,83</b>	0,41	0,86
X4	<b>0,89</b>	0,44	0,98
X5	0,52	<b>0,79</b>	0,90
X6	0,43	<b>0,82</b>	0,85
X7	0,29	<b>0,95</b>	1,00
X8	<b>0,92</b>	0,34	0,96
Alfa de Cronbach	<b>0,86</b>	<b>0,89</b>	

Usualmente o valor mínimo aceitável é de 0,50 (SCHAWB, 2007). Logo, caso o pesquisador encontre alguma comunalidade abaixo desse patamar a variável deve ser removida da amostra e a análise fatorial deve ser realizada novamente. Além disso, baixa comunalidade entre um grupo de variáveis é um indício de que elas não estão linearmente

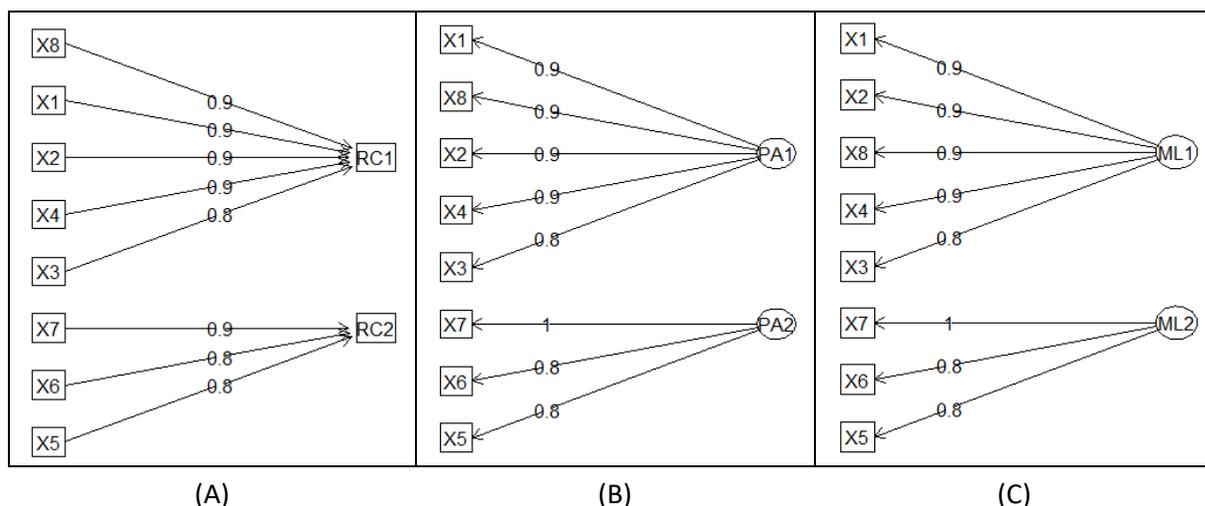
correlacionadas e, por isso, não devem ser incluídas na análise fatorial (FIGUEIREDO FILHO; SILVA JÚNIOR, 2010).

Pela análise da Tabela 4, é possível verificar as cargas fatoriais, com os coeficientes das colunas representando o relacionamento entre cada uma das variáveis e seus respectivos fatores. Em negrito estão as cargas fatoriais com maior valor para as variáveis. Os três métodos apresentaram bastantes semelhanças entre si em relação à formação dos fatores com as variáveis. O fator 1 representa (57,38%, 57,13% e 57%) da variância total pelos métodos de Componente Principal, Fator Principal e Máxima Verossimilhança, composto por 5 variáveis positivamente relacionadas, formados pelas as variáveis de ameaça (X1), lesão corporal (X2), homicídio (X3), roubo a pessoa (X4) e estupro (X8). Os crimes foram todos contra as pessoas, logo, o fator 1 será denominado de “crimes contra pessoa” (Tabela 4 e Quadro 1).

Pela análise da Tabela 4 e Quadro 1, nota-se que o fator 2 foi composto por 3 variáveis: roubo de veículo (X5), roubo em residência (X6) e roubo estabelecimento comercial (X7). Os crimes do fator 2 foram todos em relação ao roubo, logo, o nome mais adequado para fator 2 é “crimes de roubo”. Assim, visando facilitar a compreensão da análise, elaborou-se o diagrama apresentado no Quadro 1, sendo este uma forma de demonstrar a formação dos fatores, principalmente quando o estudo apresenta grande número de variáveis.

A avaliação da confiabilidade de uma estrutura fatorial foi verificada por meio do coeficiente de consistência interna chamado de Alpha de Cronbach (CRONBACH, 1951), que avalia o grau em que os itens de uma matriz de dados estão correlacionados entre si. Os itens que compõem uma escala devem apresentar um alto valor de alpha, e como esse coeficiente varia de 0 a 1, quanto mais próximo de 1, maior é a confiabilidade das escalas (HAIR *et al.*, 2009). De acordo com Hair *et al.* (2009), o limite inferior geralmente aceito para o alpha é de 0,70, apesar de se aceitar sua redução para 0,60 em pesquisas exploratórias. Nesse estudo, os coeficientes de alpha dos fatores foram iguais 0,86 e 0,89, apontando adequada consistência interna para estudos de natureza exploratória (Tabela 4).

Quadro 1 - Diagrama da análise fatorial exploratória: (A) pelo método de Componente Principal com RC1: fator 1 e RC2: fator 2; (B) pelo método do Eixo principal com PA1: fator1 e PA2: fator2; (C) pelo método da Máxima Verossimilhança com ML1: fator 1 e ML2: fator 2



Os modelos da análise fatorial podem ser construídos utilizando as cargas fatoriais pelo método de componente principal para as variáveis. Com as equações podem ser identificados as contribuições das variáveis para os fatores escolhidos pela AFE conforme indicado na sequência.

$$\begin{aligned}
 X_1 &= 0,92F_1 + 0,37F_2, \\
 X_2 &= 0,91F_1 + 0,36F_2, \\
 X_3 &= 0,85F_1 + 0,43F_2, \\
 X_4 &= 0,89F_1 + 0,44F_2, \\
 X_5 &= 0,52F_1 + 0,80F_2, \\
 X_6 &= 0,43F_1 + 0,85F_2, \\
 X_7 &= 0,29F_1 + 0,95F_2, \\
 X_8 &= 0,93F_1 + 0,34F_2.
 \end{aligned}$$

Após definição do modelo a ser utilizado, é usual a apresentar os resíduos padronizados (Tabela 5) com o intuito de verificar a qualidade do ajuste do modelo selecionado. Percebe-se que todos os resíduos foram baixos, possuindo valores absolutos menores que 2,5. Sendo assim, o bom comportamento residual indica que não há evidências suficientes para inclusão ou exclusão de parâmetros no modelo ajustado, implicando assim em um adequado ajuste do modelo.

Tabela 5 - Matriz de resíduos padronizados do modelo ajustado pelo critério de rotação *Varimax* entre as 8 variáveis aferidas referentes aos 46 municípios de Mato Grosso

Variável	X1	X2	X3	X4	X5	X6	X7	X8
X1	0,000							
X2	0,012	0,000						
X3	-0,022	-0,052	0,000					
X4	0,001	0,001	-0,002	0,000				
X5	0,005	0,005	-0,015	-0,008	0,000			
X6	-0,014	-0,010	0,033	-0,001	-0,081	0,000		
X7	0,009	0,011	-0,027	0,005	0,001	-0,022	0,000	
X8	-0,004	-0,004	-0,012	-0,008	-0,003	0,002	0,004	0,000

A última etapa da análise fatorial foi gerar as escores fatoriais do fator 1 e 2 obtidos pelo método de mínimos quadrados ponderados para os 46 municípios de Mato Grosso (Tabela 6). Embora o interesse principal da AF seja, usualmente, estimar parâmetros do modelo, os valores estimados dos fatores comuns, os escores também podem ser obtidos. Essas quantidades, que correspondem aos valores de cada fator para cada indivíduos, são muitas vezes usadas na construção de índices, para fazer diagnósticos ou como entrada em análises subsequentes. Pelas as escores neste estudo foi possível identificar os municípios que apresentaram maiores números de crimes em relação os dois fatores (crimes contra pessoa e roubo).

De acordo com a Tabela 6, percebe-se que os municípios de Cuiabá (5,633), Rondonópolis (3,606), Várzea Grande (1,777) e Sinop (1,256) apresentam escores elevados no fator 1, fato este que os classificam como os municípios que possuem os maiores registros nos “crimes contra pessoa” no ano de 2015. Mas, esse resultado é também devido ao fato de o município Cuiabá ser capital do estado de Mato Grosso, considerada uma pequena metrópole no centro da América do Sul. O município Rondonópolis tem o segundo maior produto interno bruto (PIB) do estado de Mato Grosso. Esses quatro municípios também são as maiores cidades do estado de Mato Grosso.

Tabela 6 - Escores fatoriais obtidos pelo método de mínimos quadrados ponderados para os municípios de Mato Grosso

Municípios	Fator 1	Fator 2	Municípios	Fator 1	Fator 2
Acorizal	0,428	0,015	Itanhangá	0,356	0,088
Água Boa	0,172	0,378	Itaúba	0,429	0,023
Alta Floresta	0,470	0,105	Itiquira	0,320	0,193
Apiacás	0,382	0,109	Luciara	0,440	0,012
Araputanga	0,327	0,034	Marcelândia	0,396	0,069
Arenápolis	0,383	0,010	Matupá	0,283	0,162
Aripuanã	0,333	0,074	Nobres	0,370	0,031
Barão de Melgaço	0,400	0,092	Nortelândia	0,395	0,037
Barra do Bugres	0,334	0,242	Paranaíta	0,329	0,194
Barra do Garças	0,696	0,809	Paranatinga	0,233	0,013
Brasnorte	0,319	0,206	Poconé	0,201	0,422
Cáceres	0,697	0,729	Poxoréo	0,352	0,136
Campinápolis	0,415	0,043	Primavera do Leste	0,716	0,488
Canarana	0,271	0,036	Querência	0,267	0,147
Carlinda	0,337	0,078	Ribeirãozinho	0,423	0,029
Chapada dos Guimarães	0,303	0,156	Rio Branco	0,436	0,005
Claudia	0,387	0,032	Rondolândia	0,428	0,037
Cocalinho	0,417	0,053	Rondonópolis	<b>3,606</b>	<b>5,800</b>
Colíder	<b>0,086</b>	0,648	Rosario Oeste	0,355	0,138
Colniza	0,325	0,131	Sapezal	<b>0,061</b>	0,165
Comodoro	0,332	0,052	Sinop	<b>1,256</b>	<b>2,179</b>
Confresa	0,330	0,122	Sorriso	0,195	0,153
Cuiabá	<b>5,633</b>	<b>3,680</b>	Tabaporã	0,389	0,047
Curvelândia	0,403	0,000	Tangará da Serra	0,251	0,544
Denise	0,317	0,031	Tapurah	0,307	0,239
Diamantino	<b>0,106</b>	0,282	Terra Nova do Norte	0,392	0,037
Guiratinga	0,416	0,036	Várzea Grande	<b>1,777</b>	<b>1,516</b>
Indiavaí	0,447	0,008	Vera	0,324	0,102

Por outro lado, os municípios Sapezal (0,061), Colíder (0,086) e Diamantino (0,106) apresentam escores baixos, classificando-os como municípios que apresentam menores registros nos “crimes contra pessoa” no ano de 2015. O fato de esses municípios possuírem baixos registros pode ser compreendido em função que possuir baixa extensão territorial, ou ainda por apresentarem os menores PIBs do estado de Mato Grosso.

No fator 2, os municípios Rondonópolis, Cuiabá, Sinop e Várzea Grande apresentaram os maiores escores: 5,8; 3,68; 2,179 e 1,516, respectivamente. Nota-se que o Município de Rondonópolis apresentou maiores números de registro nos “crimes de roubo”, superando inclusive os registrados na capital Cuiabá no ano 2015. Por outro lado, O município Curvelândia foi que menos contribuiu para o fator 2, apresentando escore 0,000, enquanto Rio Branco e Indiavaí apresentaram escores iguais a 0,005 e 0,008, respectivamente, o que indica baixa frequência de registros de “crimes de roubo” no ano de 2015.

#### 4. CONCLUSÃO

Neste trabalho teve objetivo de apresentar as teorias sobre os principais aspectos da AFE, aplicação com diversos métodos de extração e suas interpretações práticas. Por meio dela foi possível identificar dois fatores entre os três métodos pesquisados (Componente Principal, Fator Principal e Máxima Verossimilhança) que explicaram aproximadamente entre 94% a 95% da variância total encontrada. Os três métodos apresentaram resultados bastante semelhantes, sendo que o método do Componente principal foi o que apresentou melhor resultados no estudo de caso proposto.

A AFE é uma técnica relativamente complexa que exige dos pesquisadores uma série de decisões para que possa se obter uma estrutura fatorial adequada e com essa análise foi propiciou identificar os municípios com maiores crimes contra pessoa e contra roubo. Entretanto, espera-se que as informações aqui presentes possam auxiliar os pesquisadores na condução da técnica de AFE.

#### 5. REFERÊNCIAS

BARLETT, M. S. Tests of significance in factor analysis. *British Journal of Psychology*, v.3, p.77-85, 1950.

CATTELL, R. B. The scree test for the number of factors. *Multivariate Behavioral Research*, v.1, n.1, p.245-276, 1966.

COSTELLO, A. B.; OSBORNE, J. W. Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation*, v. 10, n.7, p.1-9, 2005.

CRAWFORD, A. V.; GREEN, S. B.; LEVY, R., LO, W.; SCOTT, L.; SVETINA, D. E COLS. Evaluation of parallel analysis methods for determining the number of factors. *Educational and Psychological Measurement*, v.70, n.6, p.885-901. 2010.

CRONBACH, L. J. Coefficient alpha and the internal structure of tests. *Psychometrika*, v.16, p.297-334, 1951.

DAMÁSIO, B. F. Uso da análise fatorial exploratória em psicologia. *Avaliação Psicológica*, v.11, n.2, p. 213-228, 2012.

DZIUBAN, C.D.SHIRKEY, E,S. When is a correlation matrix appropriate for factor analysis? Some decision rules, *Psychol, Bull*, v.81, p.358-361, 1974.

FIELD, A. **Discovering Statistics Using SPSS**. Ed. 2, London: Sage. 2005.

FIGUEIREDOFILHO, D. B.; SILVA JÚNIOR, J. A. Visão além do alcance: uma introdução à análise fatorial. *OPINIÃO PÚBLICA*, v.16, n.1, p. 160-185, 2010.

GUTTMAN, L. Some necessary conditions for common factor analysis, *Psychometrika*, v.19, p.149-162, 1954.

HAIR JR., J. F.; WILLIAM, B.; BABIN, B.; ANDERSON, R. E. **Análise multivariada de dados**. 6.ed. Porto Alegre: Bookman, 2009.

HARMAN, H.H. **Modern Factor Analysis**. 2ª Edição. Chicago: University of Chicago Press, 1967.

HORN, J. L. A rationale and technique for estimating the number of factors in factor analysis. **Psychometrika**, v.30, n.1, p.179-185. 1965.

HUTCHESON, G. D.; SOFRONIOU, N. **The multivariate social scientist: Introductory statistics using generalized linear models**. London: Sage Publications. 1999.

JOHNSON, R.A.; WICHERN, D.W. **Applied multivariate statistical analysis**. 6ª Edição. Madison: Prentice Hall International. 2008. 816p.

KAISER, H.F. The varimax criterion for analytic rotation in factor analysis. **Psychometrika**, v.23, p.187-200. 1958.

KAISER, H. F. The application of electronic computers to factor analysis. **Educational and Psychological Measurement**, v.20, p.141-151, 1960.

KIM, J.; MUELLER, C. W. **Factor analysis: Statistical methods and practical issues**. Beverly Hills, CA: Sage Publications. 1978.

KIRCH, J. L.; HONGYU, K.; SILVA, F. L.; DIAS, C. T. S. Análise Fatorial para Avaliação dos Questionários de Satisfação do Curso de Estatística de uma Instituição Federal. **E&S Engineering and Science**, v.6, n.1, 2017.

LAROS, J. A. O uso da análise fatorial: algumas diretrizes para pesquisadores. Em: L. Pasquali (Org.), **Análise fatorial para pesquisadores**. Petrópolis: Vozes. 2004.

LORENZO-SEVA, U.; TIMMERMAN, M. E.; KIERS, H.A. The hull method for selecting the number of common factors. **Multivariate Behavioral Research**, v.46, n.2, p.340-364, 2011.

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.6, n.2, 2016.

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

PASQUALI, L. **Análise fatorial: um manual teórico-prático**. Brasília: EditoraUnB, 1999.

PATIL, V. H.; SINGH, S. N.; MISHRA, S.; DONAVAN, D.T. Efficient theory development and factor retention criteria: Abandon the 'eigenvalue greater than one' criterion. **Journal of Business Research**, v.61, v.2, p.162-170, 2008.

PEREIRA, J, C, R. **Análise de dados qualitativos: estratégias metodológicas para as ciências da saúde, humanas e sociais**. São Paulo: EDUSP. 1999.

R DEVELOPMENT CORE TEAM. R: a language and environment for statistical computing. **R Foundation for Statistical Computing**, Vienna, Austria, 2018. Disponível em: <https://www.R-project.org/>. Acessado em: 01/11/18.

SCHWAB, A. J. **Data analysis and computers II**. 2007. Disponível em: [http://www.utexas.edu/courses/schwab/sw388r7\\_spring\\_2007/SW388R7\\_Syllabus\\_Spring\\_2007.pdf](http://www.utexas.edu/courses/schwab/sw388r7_spring_2007/SW388R7_Syllabus_Spring_2007.pdf). Acessado em: 01/11/18.

SPEARMAN, C. General intelligence, objectively determined and measured. **American Journal of Psychology**, v.15, p.201-293, 1904.

TABACHNICK, B.; FIDELL, L. **Using multivariate analysis**. Needham Heights: Allyn & Bacon, 2007.

VELICER, W. F. Determining the number of components from the matrix of partial correlations. **Psychometrika**, v.41, n.3, p.321-327, 1976.