

ANÁLISE FATORIAL EXPLORATÓRIA APLICADA NO QUESTIONÁRIO DO CADASTRO ÚNICO

Alexsandro Arruda Ferreira¹
Kuang Hongyu²

RESUMO: Este estudo tem por objetivo a extração e compreensão de fatores relevantes que caracterizam famílias de baixa renda por meio da técnica estatística multivariada Análise Fatorial Exploratória (AFE). O Cadastro Único é um importante instrumento de planejamento de políticas públicas, pois proporciona em uma só base, um retrato abrangente das condições de vida das famílias mais vulneráveis, incluindo informação de renda, características do domicílio, escolaridade, deficiências entre outras. Atualmente é utilizado em mais de 20 programas e políticas sociais, como o Programa Bolsa Verde, Tarifa Social de Energia Elétrica (TSEE) e Bolsa Família. Foi utilizada AFE, uma técnica multivariada que permite reduzir a complexidade do problema original, possibilitando a divisão do questionário em fatores e a identificação das questões mais representativas do instrumento. Das 13 variáveis originais selecionadas para análise fatorial, passaram a constituir cinco fatores relacionados a Serviços Públicos essenciais como Emprego e Renda, Deficiência e Aposentadoria, Divisão do Domicílio e Material Domicílio. Uma redução significativa dos dados com 69% de variância explicada.

Palavras-chave: Análise Multivariada; Análise Fatorial Exploratória; Cadastro Único.

FACTORIAL EXPLORATORY ANALYSIS APPLIED IN THE QUESTIONNAIRE OF THE “CADASTRO ÚNICO”

ABSTRACT: This study aims to extract and understand relevant factors that characterize low income families through the multivariate statistical technique called Exploratory Factor Analysis (EFA). The “Cadastro Único” is an important tool for planning public policies, as it provides a comprehensive picture of the living conditions of the most vulnerable families, including income information, characteristics of the home, schooling, disabilities, among others. It is currently used in more than 20 social programs and policies, such as the Bolsa Verde Program, Tarifa Social de Energia Elétrica (TSEE) and Bolsa Família Program. EFA was used, a multivariate technique that allows reducing the complexity of the original problem, allowing the division of the questionnaire into factors and the identification of the most representative questions of the instrument. Of the 13 original variables selected for factor analysis, five factors related to Essential Services, Employment and Income, Disability and Retirement, Household and Household Material Division were constituted. A significant reduction of the data with 69% explained variance.

Keywords: Multivariate Analysis; Exploratory Factor Analysis; Cadastro Único.

¹Graduando do Curso de Estatística. Universidade Federal de Mato Grosso – UFMT. Campus Cuiabá.
alex.ferreira.aaf@gmail.com

² 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

Desigualdade e pobreza são fenômenos complexos existentes ao longo da história do país, sendo assim não é adequado ter uma visão simplista sobre as políticas desenvolvidas que as combatem. A elaboração de políticas que efetivamente superem a desigualdade que existe no Brasil se tornou um enorme desafio para a gestão pública (BICHIR, 2010).

O Cadastro Único foi criado em Julho de 2001 como um instrumento de identificação e caracterização sócio econômicas das famílias brasileiras de baixa renda, com renda inferior a meio salário mínimo por pessoa ou renda total familiar de até 3 salários mínimos em todos os municípios brasileiros. Sua consolidação só ocorreu de fato com a implementação do Programa Bolsa Família (PBF), que unificou os programas de transferência de renda existentes até 2003 (TEIXEIRA et al., 2017).

Embora a utilização do Cadastro Único continue centrado na seleção dos beneficiários do Bolsa Família, existe um amplo leque de utilização que já vem sendo adotadas. No curto prazo, esses programas tem o objetivo de avaliar os problemas decorrentes da situação de pobreza, para longo prazo investir no capital humano, amenizando o ciclo de pobreza de geração em geração (BARROS et al., 2007)

Para a composição do questionário do Cadastro Único, são disponibilizadas 30 variáveis referentes à família nos anos de 2014 a 2016 e 34 variáveis relacionados a pessoa para todos os anos. (AFE) é uma técnica multivariada que permite reduzir a complexidade do problema original, possibilitando a divisão do questionário em fatores e a identificação das questões mais representativas do instrumento (KIRCH et al., 2017)

A construção do modelo da Análise Fatorial consiste em dois principais métodos para o cálculo dos fatores: método da máxima verossimilhança e o método dos componentes principais. Com o primeiro método, podemos testar hipóteses e cria intervalos de confiança para grandes amostras, entretanto, sugere uma normalidade multivariada dos dados, motivo pelo qual o método dos componentes principais é mais utilizado, pois não existe pressuposição das variáveis analisadas (BARROSO; ARTES, 2003).

AFE via Componentes Principais são técnicas multivariadas de modelagem da covariância por meio da transformação linear das variáveis originais em um número menor de componentes não correlacionados que explicam a maior parte da variabilidade dos dados (NEISSE; HONGYU, 2016; KIRCH et al., 2017).

Contudo ambas as técnicas estão relacionadas à redução de dados através da combinação de variáveis com variabilidades afins. A possibilidade de redução de dados

tornou tais técnicas populares em várias áreas do conhecimento como: agronomia, zootecnia, ecologia, florestal dentre outras (NEISSE; HONGYU, 2016).

O objetivo deste artigo é a extração e compreensão de fatores relevantes que caracterizam famílias de baixa renda por meio da técnica estatística multivariada da (AFE).

MATERIAL E MÉTODOS

Os dados foram obtidos por meio do site *aplicacoes.mds.gov.br*, referente ao Cadastro Único do Governo Federal, são aproximadamente 21 mil indivíduos com um questionário composto por 47 questões relacionados a família e pessoas. Foram selecionadas para esse estudo as famílias que tiveram as maiores quantidades de itens respondidos do questionário, as famílias que possuíam dados faltantes foram excluídas da análise juntamente com os itens de questionário que possuíam apenas uma resposta, ao final foram selecionados 3147 mil famílias e 13 itens de questionário que serão apresentados na Tabela 1 para a aplicação da (AFE).

Tabela 1. Itens selecionados do Cadastro Único que compõe o questionário.

local_domic_fam	O local onde está situado o seu domicílio tem, na maioria, características:
comodos_domic_fam	Quantos cômodos tem seu domicílio?
comodos_dormitorio_fam	Quantos cômodos estão servindo permanentemente de dormitório para os moradores do seu domicílio?
material_piso_fam	Qual é o material predominante no piso do seu domicílio?
material_domic_fam	Qual é o material predominante na construção das paredes externas do seu domicílio?
agua_canalizada_fam	O seu domicílio tem água canalizada para, pelo menos, um cômodo?
abaste_agua_domic_fam	Qual é a forma de abastecimento de água utilizada no seu domicílio?
escoa_sanitario_domic_fam	De que forma é feito o escoamento do banheiro ou sanitário?
destino_lixo_domic_fam	O lixo do seu domicílio
deficiencia_memb	Tem alguma deficiência permanente que limite as suas atividades habituais (como trabalhar, ir à escola, brincar, etc.)
remuner_emprego_memb	Qual foi a remuneração bruta de todos os trabalhos recebidos nesse período?
trabalho_12_meses_memb	Teve trabalho remunerado nos últimos 12 meses?
renda_aposent_memb	Valor recebido de aposentadoria

O primeiro passo durante a implementação da (AFE) é verificar se a aplicação da técnica tem validade para as variáveis escolhidas. Para isso, dois métodos de avaliação são comumente utilizados, o critério de Kaiser-Meyer-Olkin (KMO) e o teste de esfericidade de Bartlett (DZIUBAN; SHIRKEY, 1974; DAMÁSIO, 2012).

O índice de KMO comumente conhecido com o índice de adequação da amostra, é um teste estatístico que sugere a proporção de variância dos itens que podem estar sendo explicado por uma variável latente, sendo um indicador de quão adequada é a aplicação da AFE para o conjunto de dados (LORENZO-SEVA; TIMMERMAN; HAIR et al., 2005; DAMÁSIO, 2012). O KMO é obtido por meio dos quadrados das correlações totais dividido pelo quadrado das correlações parciais das variáveis analisadas, cuja expressão é (DZIUBAN; SHIRKEY, 1974).

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

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 o coeficiente de correlação linear entre resíduos. O KMO pode variar entre 0 e 1. Os valores próximos de 0 indicam que as somas das correlações parciais dos itens avaliados são bastante altas em relação a soma das correlações totais. Nesses casos, possivelmente a aplicação da Análise Fatorial será inadequada (PASQUALI, 1999; DAMÁSIO, 2012).

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 e valores maiores que 0,8 e 0,9 são considerados ótimos (HUTCHESON; SOFRONIOU, 1999). Contudo a medida de adequação da amostra (MAA) aceitável é maior que 0,5 (PEREIRA, 1999). O modelo de Análise de Fatores pode ser expresso como (GUTTMAN, 1954; JOHNSON; WICHERN, 2007):

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

Em que X_i é o i -ésimo escore após a padronização (média 0 e desvio padrão 1), $i = 1, 2, \dots, p$, sendo p o número de variáveis e $a_{i1}, a_{i2}, \dots, a_{im}$ são as cargas fatoriais 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 possui média 0.

Os p valores observados X_p são expressos em termos de $p+m$ variáveis aleatórias não observáveis ($F_1, F_2, \dots, F_m; e_1, e_2, \dots, e_p$), características que diferem a análise fatorial da regressão múltipla, com as variáveis independentes podendo ser observadas e suas posições são ocupadas por F no modelo fatorial (MINGOTI, 2005). Matricialmente temos (JOHNSON; WICHERN, 1998; NEISSE; HONGYU, 2016):

$$X_{(p \times 1)} = \Lambda_{(p \times m)} F_{m \times 1} + \varepsilon_{(p \times 1)}$$

Com o modelo de Análise Fatorial:

$$\text{Var}(X_j) = a_{j1}^2 \text{Var}(F_1) + a_{j2}^2 \text{Var}(F_2) + \dots + a_{jm}^2 \text{Var}(F_m) + \text{Var}(e_j) = a_{j1}^2 + a_{j2}^2 + \dots + a_{jm}^2 + \text{Var}(e_j)$$

Em que $a_{j1}^2 + a_{j2}^2 + \dots + a_{jm}^2$ são comumente denominadas de comunalidades X_j (a parte da sua variância que é explicada pelos fatores comuns). A comunalidade não pode ultrapassar o valor 1, ou seja, $-1 \leq a_{ij} \leq 1$. Podendo ser estabelecido que a correlação entre X_j e X_j seja:

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

Consequentemente, duas variáveis serão altamente correlacionadas se possuírem altas cargas no mesmo fator (NEISSE; HONGYU, 2016).

RESULTADOS E DISCUSSÃO

Os dados do Cadastro Único disponível no site contêm aproximadamente 21 mil famílias e 44 itens de questionário, porém existem muitas famílias com questões não respondidas e vários itens sem resposta. Cattell (1978) argumenta que o número mínimo da amostra deva ser de 250 sujeitos. Atualmente, é fortemente aceito que as utilizações de amostras grandes tendem a obter resultados mais precisos, diminuindo o efeito do erro amostral (MacCallum & Turcker, 1991). Foram selecionadas 3147 mil famílias que possuem o maior número de resposta para aplicação da análise fatorial. Após análise, os itens que apresentaram comunalidades abaixo de 0,40 foram eliminados, de acordo com Schawb (2007) usualmente o valor mínimo aceitável é de 0,50, contudo para nossos propósitos os itens acima de 0,40 foram mantidos.

Foi realizado o teste de KMO e teste de Bartlett, a fim de verificar a adequabilidade dos dados para aplicação da AFE. O valor do KMO foi de 0,61 um valor medíocre de acordo com (Hutchenso & Sofroniou, 1999), entretanto Hair et al (2006) sugere 0,50 como um valor aceitável, sendo assim o KMO obtido é adequado para aplicação da análise fatorial. O teste de Bartlett mostrou um valor de $\chi^2 = 279600$ com 12 graus de liberdade e um valor-p de $2,2 \times 10^{-16} < 0,001$, com esses valores apresentados pelo teste, rejeita-se a hipótese nula, indicando que a matriz de correlação da amostra não é uma matriz identidade e a aplicação da AFE é apropriada. A medida de adequação (MAA) foi considerada aceitável para cada item ($>0,5$) como mostra a Tabela 2, indicando o prosseguimento da AFE.

Tabela 2. Medida de Adequação para cada item do questionário.

Item	MAA
local_domic_fam (V1)	0,72
comodos_domic_fam (V2)	0,53
comodos_dormitorio_fam (V3)	0,53
material_piso_fam (V4)	0,81
material_domic_fam (V5)	0,71
agua_canalizada_fam (V6)	0,76
abaste_agua_domic_fam (V7)	0,75
escoa_sanitario_domic_fam (V8)	0,83
destino_lixo_domic_fam (V9)	0,73
deficiencia_memb (V10)	0,50
remuner_emprego_memb (V11)	0,50
trabalho_12_meses_memb (V12)	0,50
renda_aposent_memb (V13)	0,50

O *screeplot* da Figura 1 foi obtido por meio de uma AFE. Utilizando o método de componentes principais, aplicado em uma matriz com 13 variáveis, apesar de não existir um critério consensual para escolha de quantos fatores devem ser extraídos, foi utilizado o método de Kaiser-Guttman, em que sugere ser extraídos somente os fatores com autovalores maiores que 1, por meio da análise do *screeplot*, foi observado que a linha poligonal decresce rapidamente até o quinto fator, os quais explicam aproximadamente 69% da variância total. Tanto Hair et al quanto Schawb (2007) sugere um patamar de 60% como sendo aceitável, entretanto, Peterson(2000) realizou um estudo com o objetivo de avaliar os níveis de variância explicada no estudos que utilizaram análise fatorial, concluindo que em média a variância explicada é de 56,6%, 10% apresentaram variância explicada maior que 76% e 10% apresentaram variância explicada menor que 34%, dessa forma os valores obtidos nesse estudo são superiores ao valor médio.

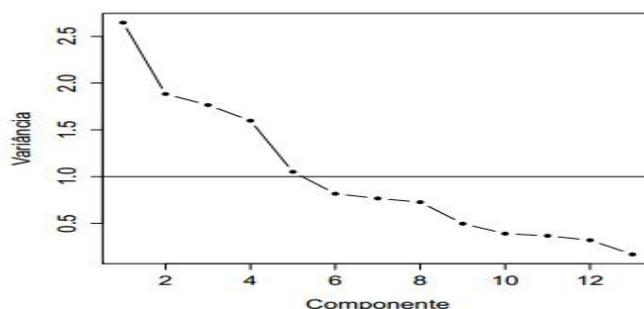


Figura 1. Scree plot dos autovalores obtidos na análise fatorial.

Para a análise fatorial, foram selecionados os autovalores e os percentuais de variância explicada por cada fator, como se pode observar na Tabela 3. O total dos fatores explicaram aproximadamente 69% da variação total dos itens do questionário. O primeiro fator explicou 18%, o segundo 15%, o terceiro 13%, o quarto 12% e o quinto 11%.

Tabela 3. Autovalores, variância e variância acumulada dos 5 fatores retidos.

Fator	Autovalor	Variância Explicada (%)	Variância Explicada Acumulada (%)
1	2,29	0,18	18
2	1,83	0,15	33
3	1,73	0,13	46
4	1,63	0,12	58
5	1,45	0,11	69

A Figura 2 mostra o gráfico de correspondência para as cargas prévias. Verifica-se que não é possível distinguir claramente as variáveis de cada fator. Foi utilizado o método de rotação varimax, um dos mais bem sucedidos e utilizados em pesquisas. Após aplicação da rotação varimax, que procura minimizar o número de variáveis que apresentam cargas altas em cada fator. Pode-se observar na Figura 3 a separação dos itens correspondente com cada fator.

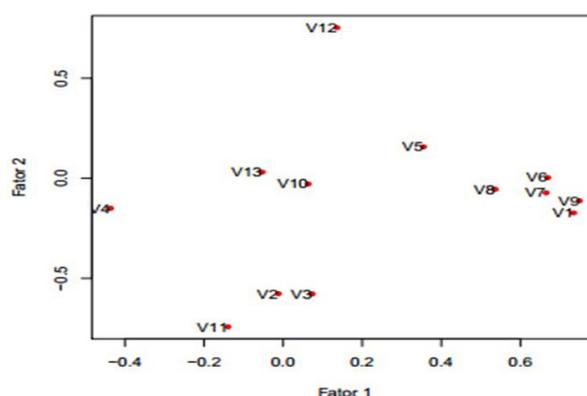


Figura 2. Gráfico de correspondência para cargas prévias antes da rotação varimax dos 13 itens selecionado do cadastro único.

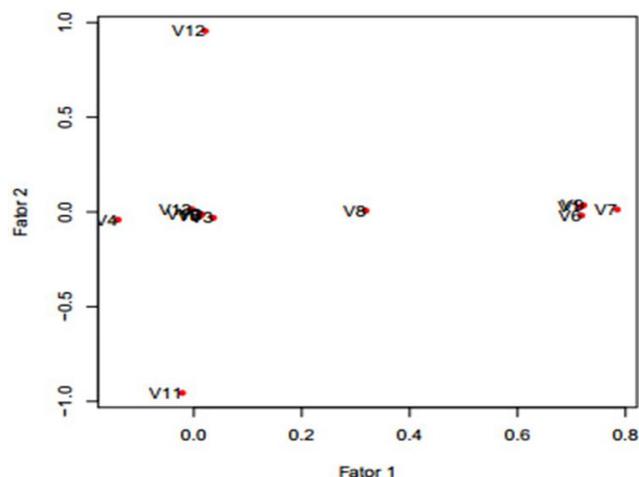


Figura 3. Gráfico de correspondência para cargas prévias após a rotação varimax dos 13 itens selecionado do cadastro único.

Tanto Hair et al (2006) quanto Schawb (2007) sugerem que a mesma variável não pode contribuir para a construção de fatores distintos. Adota-se 0,40 como limite aceitável da contribuição da variável na criação do fator com o objetivo de evitar o problema da indeterminação da relação entre variáveis e fatores. Seguindo a Tabela 4, nenhuma variável apresenta estrutura complexa, ou seja, cargas fatoriais acima de 0,40 em ambos os componentes, ouve um incremento generalizado no valor das cargas fatoriais associados aos fatores extraídos. Os valores apresentados em negritos são referentes as maiores cargas das variáveis nos fatores.

Tabela 4. Cargas fatoriais rotacionadas por método varimax e comunalidade de cada item do questionário.

Variáveis	Fator 1	Fator 2	Fator 3	Fator 4	Fator 5	Comum
local_domic_fam (V1)	0,717	0,028	0,201	-0,004	0,202	0,596
comodos_domic_fam (V2)	0,014	0,014	0,893	-0,006	0,143	0,817
comodos_dormitorio_fam (V3)	0,036	0,031	0,897	-0,005	-0,005	0,807
materia_piso_fam (V4)	-0,141	0,041	0,094	-0,086	0,666	0,481
material_domic_fam (V5)	0,012	0,022	-0,132	-0,015	-0,765	0,603
agua_canalizada_fam (V6)	0,718	0,019	-0,165	0,015	-0,108	0,555
abaste_agua_domic_fam (V7)	0,785	-0,013	-0,047	0,008	0,046	0,620
escoa_sanitario_domic_fam (V8)	0,319	-0,007	0,140	-0,039	-0,543	0,418
destino_lixo_domic_fam (V9)	0,722	-0,035	0,116	0,005	-0,231	0,590
deciencia_memb (V10)	0,011	0,009	-0,004	-0,901	-0,026	0,812
remuner_emprego_memb (V11)	-0,021	-0,956	0,014	0,008	0,015	0,914
trabalho_12_meses_memb (V12)	0,021	0,955	0,030	-0,016	-0,013	0,914
renda_aposent_memb (V13)	-0,004	-0,014	0,007	0,901	0,017	0,812

O primeiro fator (F1) está relacionado com as variáveis: V1, V6, V7 e V9, ou seja, itens que estão associa do são local da residência das famílias, em que moram essas famílias possui coleta de lixo, abastecimento de água, água canalizada e se moram na zona rural ou urbana. Pode-se observar pela Tabela 4 dos quatro itens que compõem o primeiro fator, todos tem

cargas fatoriais acima de 0,5, possuem valores muito próximos com destaque para os itens abastecimento de água (V7) e destino do lixo (V9) com as maiores cargas apresentadas 0,785 e 0,722. Com essas características apresentadas pelas variáveis, o F1 pode ser denominado de

Serviços Públicos Essenciais.

O segundo fator (F2) está associado às variáveis V11 e V12, esses itens tem o interesse em saber se algum membro da família esteve trabalhando de maneira remunerada nos últimos 12 meses e o valor dessa remuneração, como mostra a Tabela 4, a variável V12 possui uma carga alta positiva 0,955 e a variável V11 uma carga alta negativa -0,956, indicando assim que as famílias possuem emprego remunerado, porém com baixa remuneração, sendo assim esse fator pode ser denominado de **Emprego e Renda**.

Os itens cômodos por domicílio (V2) e cômodos dormitórios (V3) estão relacionados ao fator (F3), com cargas fatoriais altas e positivas 0,893 e 0,897, possuem comunalidades altas como mostra a Tabela 4, São itens relacionados com a quantidade de cômodos que tem a residência dessas famílias e quantos cômodos estão servindo permanentemente de dormitório, com essas características podemos denominar essa fator de **Divisão do Domicílio**.

Membros com deficiência (V10) e renda de membros aposentados (V13) constituem o Fator 4, V10 possui carga fatorial alta negativa -0,901 com uma alta comunalidade e V13 possui carga positivamente alta 0,901 com comunalidade de 0,812 uma das mais altas da análise. Famílias com a presença destes grupos tendem a ter despesas adicionais relacionadas a cuidados cotidianos como: segurança, saúde, alimentação e etc , sendo um fator indicador de

Deficiência e Aposentadoria.

As variáveis V4,V5 e V8 estão relacionados ao quinto fator (F5) com cargas 0,666, -0,765 e -0,543, respectivamente. V5 possui comunalidade acima de 0,5 sendo a que mais contribuiu nesse grupo, V4 e V8 possui comunalidade abaixo de 0,5 com valores de carga fatorial menores em relação a V8. Esse fator agrupou variáveis relacionadas com a estrutura do domicilio, o item V4 está associado material predominante utilizado na construção do piso das residências, V5 referente ao material usado para construção das paredes externas e V8 a forma do escoamento do banheiro. Com essas características o Fator 5 pode ser indicado como **Material do Domicílio**.

CONCLUSÃO

O uso de técnicas multivariadas como a análise fatorial exploratória pode auxiliar na redução da dimensão dos dados, sendo possível agrupar variáveis mais correlacionadas, fazendo com que a perda da informação seja a menor possível. Com os resultados apresentados nesse estudo foi possível identificar as variáveis mais representativas do cadastro único. Das 13 variáveis originais selecionadas para análise fatorial, passaram a constituir cinco fatores relacionados a Serviços públicos Essenciais, Emprego e Renda, Decidência e aposentadoria, Divisão do Domicílio e Material Domicílio. Uma redução significativa dos dados com 69% de variância explicada. O cadastro único pode funcionar como mecanismo de suporte ao planejamento, gerenciamento e controle de políticas sociais, sendo assim a aplicações de técnicas multivariadas auxiliam na interpretação dos dados, contribuindo para tomadas de decisões e diagnósticos que ajudam a identificar o público alvo a ser atingido.

REFERÊNCIAS BIBLIOGRÁFICAS

- BARROSO, L. P., ARTES, R. **Análise Multivariada**. Lavras: Região Brasileira da Sociedade Internacional de Biometria, volume (1), 150, 2003.
- BARROS, R.; CARVALHO, M.; DUTRA, R.; FRANCO, S.; HARTUNG, G. **O uso de preditores para a melhoria da qualidade das informações do Cadastro Único**. Ipea, 2007. Mimeografado.
- BICHIR, R. M. **O bolsa família na berlinda?** Revista Novos Estudos. São Paulo, n. 87, p. 115-129, 2010. BRASIL. Ministério do Desenvolvimento Social e Agrário. Programa Bolsa Família. Disponível em: < <http://www.mds.gov.br/bolsafamilia> >. Acesso em 12, dezembro, 2016b.
- CATTELL, R. B. **The Scientific Use of Factor Analysis in Behavioral and Life Sciences**. Nova York: Plenum, 1978. DAMASIO B.F. Uso da Análise Fatorial Exploratória Em Psicologia. Avaliação Psicológica, 2012, 11(2), pp. 213-228.
- DZIUBAN, C.D. SHIRKEY, E, S. **When is a correlation matrix appropriate for factor analysis? Some decision rules**. Psychological Bulletin, 81(6), 358-361, 1974.
- GUTTMAN, L. **Some necessary conditions for common factor analysis**. Psychometrika, 19, 149-162, 1954.
- HAIR, J. F., ANDERSON, R. E., TATHAM, R. L. BLACK, W. C. **Análise multivariada de dados**. A. S. Sant'Anna & A. C. Neto (Trad.). Porto Alegre: Bookman. 2005.
- HAIR, Jr; BLACK, W. C; BABIN, B. J; ANDERSON, R. E e TATHAM, R. L. **Multivariate Data Analysis**. 6ª edição. Upper Saddle River, NJ: Pearson Prentice Hall, 2006.
- HUTCHERSON, 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**. Madison: Prentice Hall International, 6ª edição, 2007.
- 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**. Engineering and Science, 2017.
- LORENZO-SEVA, U., TIMMERMAN, M. E., & KIERS, H. A. L. (2011) **The Hull Method for Selecting the Number of Common Factors**. Multivariate Behavioral Research, 46(2), 340-364. doi:10.1080/00273171.2011.564527.
- LORENZO-SEVA, U. FERRANDO, P. J. **FACTOR: A computer program to fit the exploratory factor analysis model**. Behavior Research Methods, 38(1), 88-91. 2006.
- MACCALLUM, R. C. & TUCKER, L. R. **Representing sources of error in the common factor model: Implications for theory and practice**. Psychological Bulletin, 109(3), 502-511, 1991.
- MINGOTI, S. **A Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada**. Editora UFMG, 2005.

- 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. 2016.
- PASQUALI, L. **Análise fatorial: um manual teórico-prático.** Brasília: Editora UnB. 1999.
- PEREIRA, J, C, R. **Análise de dados qualitativos: estratégias metodológicas para as ciências da saúde, humanas e sociais.** São Paulo: EDUSP. 1999.
- PETERSON, R. A. **A meta-analysis of variance accounted for and fator loadings in exploratory factor analysis.** . Marketing Letters, 11(3), 261-275, 2000.
- R CORE TEAM. **R: A language and environment for statistical computing.** Vienna, Austria. R Foundation for Statistical Computing, 2017.
- SCHAWB, A.J. **Eletronic Classroom,** 2007. [Online] Disponível em: <<http://www.utexas.edu/ssw/eclassroom/schwab.html>> Acesso em: [10 jan.2018].
- TEIXEIRA. A.O., JUNIOR. A.A.B., VIEIRA. M.T., KERN. A.P., FARIA. W.R. **Programa de transferência de renda, perfil socioeconômico e probabilidades de classificação: avaliando o critério de seleção dos beneficiários do Bolsa Família nas regiões Sul e Sudeste.** ANPEC, ECONOMIA SOCIAL E DEMOGRAFIA ECONÔMICA, I38, C38, 2017