

Análise Fatorial por Meio da Matriz de Correlação de Pearson e Policórica no Campo das Cisternas

Factorial Analysis by means of the Pearson and Polychoric Correlation in the Field of Cisterns

¹Lidiane Mendes Kruschewsky Lordelo, ²Kuang Hongyu, ³Patrícia Campos Borja, ⁴Milton José Porsani

¹ Professora Doutora, Centro de Ciências Exatas e Tecnológicas - Universidade Federal do Recôncavo da Bahia (lidiane@ufbr.edu.br)

² Professor Doutor, Departamento de Estatística - Universidade Federal de Mato Grosso, (prof.kuang@gmail.com)

³ Professora Doutora, Departamento de Engenharia Ambiental - Universidade Federal da Bahia (borja@ufba.br)

⁴ Professor Doutor, Instituto de Geociências da UFBA e do Centro Interdisciplinar de Energia e Ambiente/CIENAM - Universidade Federal da Bahia (Milton.porsani@gmail.com)

RESUMO: A partir da pesquisa realizada em cinco municípios do semiárido baiano sobre o uso e funcionamento do sistema de captação da água de chuva segundo as dimensões ambientais, técnicas, sanitárias, institucionais e sociais, o presente trabalho buscou identificar o método da Análise Fatorial Exploratória (AFE) que melhor se ajusta a variabilidade dos dados e cujos fatores gerados são capazes de explicar o conjunto de variáveis. Foram aplicados 455 questionários nos e estudadas 36 variáveis. Foram analisados dois métodos de Análise Fatorial Exploratória, sendo eles: (i) Matrizes de Correlação de Pearson e (ii) Policórica, ambas utilizando rotação *Varimax*. Os dois métodos permitiram a geração de 6 Fatores. Os dois métodos da análise fatorial com a matriz de correlação de Pearson e Policórica foram adequados para este conjunto de dados, logo, o método pela matriz de policórica foi o mais eficaz na identificação de grupos das variáveis, portanto foi o melhor para este conjunto de dados. O Fator 1 foi considerado o fator mais influente do estudo, pois explicou a maior variância dos dados (11%), agrupando variáveis relacionadas as dimensões técnica e sanitária que interferem diretamente na qualidade da água armazenada na cisterna.

Palavras chave: P1MC, cisternas, análise fatorial, rotação *Varimax*.

ABSTRACT: From the research carried out in five municipalities of the Bahian semi-arid region on the use and operation of the rainwater harvesting system, according to the environmental, technical, health, institutional and social dimensions, the present work sought to identify the method of Exploratory Factor Analysis (AFE) that best adjusts to the variability of the data and whose generated factors are able to explain the set of variables. 455 questionnaires were applied and studied 36 variables. Two methods of Exploratory Factor Analysis were analyzed, being: (i) Pearson Correlation Matrices and (ii) Polychoric, both using *Varimax* rotation. Both methods allowed the generation of 6 Factors. The two methods of the factorial analysis with the Pearson and Polychoric correlation matrix were adequate for this data set, but the polychoric matrix method was the most effective in identifying groups of variables, so it was the best for this set of data. The Factor 1 was considered the most influential factor in the study, because it explained the greater variance of the data (11%), grouping variables related to the technical and sanitary dimensions that directly interfere in the quality of the water stored in the cisterna.

Keywords: P1MC, cisterns, factorial analysis, *Varimax* rotation.

1. INTRODUÇÃO

A cisterna é uma tecnologia que vem sendo difundida cada vez mais em locais da zona rural como uma alternativa de convivência com a seca. A partir do ano de 2003, o Programa Um Milhão de Cisternas (P1MC) objetivou implantar de um milhão de cisternas na zona rural, região do Semiárido Brasileiro (ASA, 2017). No Brasil, apesar do uso da cisterna já está consolidado, principalmente nas regiões com dificuldades de abastecimento de água para consumo humano, foi a partir do P1MC, que uma parte significativa da população do semiárido passou uma alternativa para acesso à água.

O PIMC foi incorporado pelo Governo Federal no âmbito de uma política pública do então Ministério do Desenvolvimento Social e Combate à Fome (MDSCF). Segundo a Articulação Semiárido Brasileiro (ASA) (ASA, 2016), até outubro de 2016, foram construídas 588.935 cisternas (59% da meta), sendo que esse número corresponde às ações executadas unicamente pela ASA, as quais são financiadas pelo Governo Federal e também por financiamento da Fundação Banco do Brasil, da Fundação Avina, da Federação Brasileira de Bancos (Febran), do Instituto Brasileiro de Desenvolvimento e Sustentabilidade (Iabs), da Petrobrás, do Banco Nacional do Desenvolvimento Econômico e Social (BNDES), dentre outras.

Em face dos objetivos do Programa e da necessidade de se garantir o acesso à água no semiárido, estudos vêm sendo desenvolvidos para avaliar o desempenho do Programa e da tecnologia (BRITO; GNADLINGER, 2006; BRASIL, 2006; LORDELO et al., 2016). Muitos estudos foram desenvolvidos possibilitando avaliar questões de ordem técnica, institucional, sanitária, ambiental, social e de saúde. Resultados importantes apontam para o impacto positivo do PIMC, que além de proporcionar acesso à água com maior facilidade, também tem contribuído para alterações nas relações de gênero. O impacto na vida das famílias beneficiadas vai desde a redução da incidência de doenças relacionadas com a água até a diminuição da sobrecarga de trabalho das mulheres e crianças, que geralmente são as que exercem a tarefa de coletar água. Com a implantação das cisternas, elas deixam de percorrer longas distâncias em busca de água, em fontes muitas vezes inseguras (ASA, 2002). Segundo Silva *et al.* (2014), a liberação do tempo consumido para obtenção de água, que agora pode ser destinado a outras atividades, abre oportunidades para os moradores desenvolverem outras tarefas, como trabalho remunerado. Além disso, contribui para aumento da frequência escolar de crianças e adolescentes. As mulheres, por questões de ordem socioculturais, exercem papel fundamental em relação ao uso e funcionamento das cisternas, uma vez que estão à frente do manejo da água dentro do domicílio e colaboram com a saúde e higiene no ambiente familiar. Entretanto, sua atuação pode ir mais além que o manejo doméstico das águas. Moraes e Rocha (2013) efetuaram um estudo de caso qualitativo sobre a participação das mulheres no PIMC e constataram que algumas delas assumiram papéis significantes como construtoras de cisternas e membros de comissões locais de água, que até então, eram funções exercidas apenas por homens. Porém, estudos também já sinalizaram diversas questões que precisam ser melhor debatidas afim de garantir para a população beneficiada a água em qualidade e quantidade para o consumo humano.

Questões relativas ao projeto executivo e ao uso e funcionamento já foram evidenciados. Segundo Andrade Neto (2013), buscando atender as especificidades das famílias (quantidade de pessoas por domicílio e área da captação) e índices pluviométricos da região, o projeto executivo deveria contemplar mais de um volume de reservação. De acordo com Lordelo *et al.* (2016), o projeto técnico não contempla todas as instalações necessárias para que a cisterna atenda a população no que tange à qualidade de água. Nesse mesmo estudo, também foi detectado a falta de continuidade de ações de educação sanitária voltadas para orientações para o uso e manuseio da cisterna por parte da população. Silva *et al.* (2006) ao efetuarem uma pesquisa participante no Assentamento de Paus Brancos, município de Campina Grande, e nas comunidades rurais Curral do Meio e Poços das Pedras no município de São João do Cariri, todas no estado da Paraíba, revelaram que embora as famílias conhecessem os procedimentos corretos de higiene e uso, elas não utilizavam de forma correta a água das cisternas e poucos eram os cuidados para evitar a contaminação das águas. Por exemplo, os autores constataram que em relação à conservação do telhado das residências, em Paus Brancos, predominaram tetos sujos (45,17%), nos quais havia folhas secas, pedaços de madeira, latas e sapatos velhos, jogados nos tetos pela população, além de poeira e fezes de animais, sem que houvesse uma preocupação com a qualidade da água da

cisterna. Outro fator que mereceu atenção foi o uso de cordas e baldes para retirar a água da cisterna. A observação direta feita pelos autores nas comunidades rurais de São João de Cariri revelou que a bomba manual não era utilizada devido esforço e demora para retirar a água. Outra constatação foi que o balde usado para a coleta não apresentava condições adequadas de higiene na maioria dos domicílios, sendo usado também para outros fins.

A partir da pesquisa realizada, com financiamento do CNPq, em cinco municípios baianos sobre o uso e funcionamento do sistema de captação da água de chuva segundo as dimensões ambientais, técnicas, sanitárias, institucionais e sociais, o presente trabalho visa identificar, por meio da Análise Fatorial Exploratória (AFE), o método que melhor se ajusta a variabilidade dos dados e cujos fatores gerados são capazes de explicar o conjunto de variáveis avaliadas nesse estudo, apresentadas no Quadro 1.

A AFE é um conjunto de técnicas multivariadas que tem como objetivo encontrar a estrutura subjacente em uma matriz de dados e determinar o número e a natureza das variáveis latentes (fatores) que melhor representam um conjunto de variáveis observadas (BROWN, 2006). Ao analisar a estrutura das inter-relações de um determinado número de variáveis observadas, a AFE define os fatores que melhor explicam a sua covariância (HAIR *et al.*, 2005).

Para identificar qual é o método que melhor se ajusta na formação dos fatores foram utilizados os métodos: análise fatorial pela matriz de Correlação de Pearson e análise fatorial Policórica, por meio da rotação *Varimax*. Nesse sentido, o presente trabalho visa contribuir com esses estudos identificando o melhor método da AFE que apresente o conjunto de variáveis que expliquem o uso e funcionamento do sistema de captação da água de chuva.

2. MATERIAIS E MÉTODOS

A pesquisa foi realizada nos municípios de Abaré, Chorrochó, Macururé, Glória e Santa Brígida localizados no sertão do Estado da Bahia, selecionados por apresentar maior vulnerabilidade pelos critérios de: situação de escassez hídrica (precipitação média anual menor que 600mm) e população do município (menor que 20.000 habitantes).

Para a definição da amostra do estudo procedeu-se ao cálculo de uma amostra aleatória estratificada com alocação proporcional e amostragem aleatória simples sem reposição. As estimativas de tamanho da amostra foram feitas segundo as variáveis de estudo: qualidade bacteriológica da água, considerando-se a proporção de amostras fora do padrão bacteriológico da Portaria do Ministério da Saúde n^o 2.914/2011; consumo médio per capita de água; e as variáveis sobre o uso funcionamento das cisternas. Considerando essas proposições, a amostra de estudo foi constituída de 455 cisternas.

As variáveis estudadas foram estabelecidas a partir de aspectos contemplados no projeto executivo do P1MC e da revisão de literatura, em especial os estudos de Ruskin e Callender (1988), abordando as temáticas: ambiental, técnica, sanitária, social e institucional. O questionário para o levantamento de dados foi composto por 59 perguntas (variáveis), sendo que destas foram selecionadas 36 (Quadro 1), a partir de discussões realizadas em um Grupo Focal (MINAYO, 2005).

A coleta de dados foi realizada no período de dezembro de 2013 a março de 2015. As análises se respaldaram na técnica da Análise Fatorial Exploratória pelo método de rotação *Varimax*. Essa técnica visou reduzir o conjunto de variáveis em fatores mais influentes capazes de explicar a variabilidade dos dados e o uso e funcionamento das cisternas. Os resultados foram gerados por meio do software livre R.

Quadro 1 – Variáveis de estudo sobre o uso e funcionamento das cisternas do PIMC

| DIMENSÃO | VARIÁVEIS |
|---------------|---|
| Técnica | p42 - Estado da estrutura da cisterna p43 - Estado da tampa da cisterna p47 - Estado da bomba p50 - Manutenção da cisterna (consertos e reparos) p34 - Existência da calha de captação da água de chuva na cisterna p36 - Existência da canalização condutora da água de chuva captada para a cisterna p39 - Existência de sangrador com tamponamento na cisterna p51 - Tempo de construção em anos p38 - Existência de dispositivo para o desvio das primeiras águas da chuva p40 - Distância da cisterna a instalações de esgoto ou vala a céu aberto p10 - Área do telhado p63 - Consumo da água da família por semana em litros depois da cisterna |
| Sanitária | p48 - Caso a água seja retirada da cisterna com balde: Estado de conservação e limpeza do balde p49 - Caso a água seja retirada da cisterna com balde: O balde é usado para outros usos? p52 - Recipiente usado para armazenar a água retirada da cisterna (tipo) p53 - Estado de limpeza dos recipientes p11 - Material do telhado p46 - Forma de retirar a água da cisterna p44 - Limpeza da cisterna p45 - Forma de limpeza |
| Institucional | p57 - Educação sanitária e capacitação para uso da cisterna |
| Social | p1 - Quantas pessoas vivem na casa p2 - Bens que existem na casa p3 -Anos de estudo do chefe da família p6 - Área da casa p7 - Material das paredes da moradia |
| Ambiental | Média de precipitação mensal |

Fonte: autores.

2.1 Matriz de correlação de Pearson e Policórica

A matriz de correlação é utilizada para iniciar a análise estatística dos dados históricos dos diversos universos estudados, identificando visualmente as variáveis envolvidas no estudo que se relacionam entre si. A determinação do grau de relação entre duas variáveis é dada pelo coeficiente de Pearson, também chamado de coeficiente de correlação, ou simplesmente correlação para os pares de variáveis. Esse coeficiente é um número entre -1 e 1 que expressa o grau de dependência linear entre duas variáveis quantitativas, que quando negativa indica que uma variável diminui com o aumento da outra, e positiva quando uma variável aumenta com o aumento da outra (JOHNSON; WICHERN, 2007).

Para duas variáveis quaisquer (x, y) sendo testadas, calcula-se o coeficiente de correlação de Pearson, como segue (HAIR et al., 2005; JOHNSON; WICHERN, 2007):

$$r = \frac{\sum_i(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i(x_i - \bar{x})^2 \sum_i(y_i - \bar{y})^2}} \quad (1)$$

em que \bar{x} e \bar{y} são as médias amostrais dos conjuntos de dados.

A correlação policórica é uma medida de associação bivariada utilizada quando as duas variáveis são ordinais com 3 ou mais categorias. Suponha duas variáveis categóricas, A e B, que são relacionadas com as variáveis contínuas latentes X e Y por:

$$\begin{aligned} A &= c_i, \text{ se } \gamma_{i-1} \leq X \leq \gamma_i && \text{com } i = 1, 2, \dots, r \\ B &= d_j, \text{ se } \tau_{j-1} \leq Y \leq \tau_j && \text{com } j = 1, 2, \dots, s \end{aligned}$$

em que γ_i e τ_j são os limiares, $\gamma_0 = \tau_0 = -\infty$ e $\gamma_r = \tau_s = +\infty$.

A distribuição conjunta das variáveis latentes X e Y é assumida ser normal padrão bivariada com coeficiente de correlação ρ . O coeficiente de correlação ρ é o coeficiente de correlação policórica e é definido como a correlação produto-momento entre as variáveis normais latentes X e Y (BISTAFFA, 2010).

2.2 Modelo teórico da Análise Fatorial Exploratória - AFE

O primeiro passo durante a implementação de AFE (pressupostos) é verificar se a aplicação da análise fatorial tem validade para as variáveis escolhidas, sendo justificado pela pouca quantidade de participantes da pesquisa. Para isso, dois métodos de avaliação que são mais utilizados: o critério de *Kaiser-Meyer-Olkin* (KMO) e o Teste de Esfericidade de *Bartlett*.

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. Esse índice indica quão adequada é a aplicação da AFE para o conjunto de dados (LORENZO-SEVA, 2006; TIMMERMAN; KIERS, 2011; KIRCH *et al.*, 2017).

O modelo da análise de fatores pode ser escrito como (JOHNSON; WICHERN, 2007; NEISSE; HONGYU, 2016):

$$X_i = a_{i1}F_1 + a_{i2}F_2 + \dots + a_{im}F_m + e_i \quad (2)$$

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 tem-se (JOHNSON; WICHERN, 2007):

$$\mathbf{X}_{(p \times 1)} = \mathbf{\Lambda}_{(p \times m)} \mathbf{F}_{(m \times 1)} + \mathbf{\varepsilon}_{(p \times 1)} \quad (3)$$

com o modelo AF:

$$\begin{aligned} \text{Var}(X_j) &= a_{j1}^2 \text{Var}(F_1) + a_{j2}^2 \text{Var}(F_2) + \dots + a_{jm}^2 \text{Var}(F_m) + \text{Var}(e_j) \\ &= a_{j1}^2 + a_{j2}^2 + \dots + a_{jm}^2 + \text{Var}(e_j) \end{aligned} \quad (4)$$

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 qual não pode exceder 1, é preciso que $-1 \leq a_{ij} \leq +1$.

Pode também ser estabelecido que a correlação entre X_j e $X_{j'}$:

$$r_{jj'} = a_{j1}a_{j'1} + a_{j2}a_{j'2} + \dots + a_{jm}a_{j'm} \quad (5)$$

consequentemente, duas variáveis somente serão altamente correlacionadas se elas tiverem altas cargas no mesmo fator (NEISSE; HONGYU, 2016; KIRCH *et al.*, 2017).

3. RESULTADOS E DISCUSSÕES

Realizado o levantamento de dados e organizando-o em um banco de dados no programa Excel foi aplicado teste de Kaiser-Meyer-Olkin (KMO). Esse teste foi o critério adotado para identificar se o modelo de análise fatorial proposto é adequadamente ajustado aos dados, testando a sua consistência geral. O resultado encontrado para o KMO foi igual a 0,59. Quando esse valor é igual ou superior a 0,50 indica que a análise fatorial é apropriada para o determinado estudo.

Após a verificação dos resultados e confirmando a possibilidade da realização da análise fatorial nos dados estudados, foi aplicada a análise fatorial utilizando a matriz de correlação de Pearson e correlação policórica por método de rotação *Varimax*. Os resultados das análises realizadas pelos dois métodos possibilitaram a identificação da associação entre as variáveis, podendo assim apresentar a tendência geral dos dados.

O Quadro 1 apresenta os resultados das variâncias explicadas calculadas para os métodos de AFE (de Pearson e policórica) estudados sendo gerados seis fatores (MR1, MR2, MR3, MR4, MR5, MR6). Percebe-se que os dois primeiros fatores, independente do método utilizado, representam 11% e 9%, respectivamente, das variâncias explicadas pelos fatores. Avaliando os resultados obtidos na AFE por matriz de correlação de Pearson, observa-se que os fatores explicaram 51% das variâncias dos dados e para a matriz de correlação policórica 54%.

Quadro 1 – Variâncias explicadas para cada fator para os métodos estudados

| Métodos | Proporção explicada por cada fator | | | | | | Proporção acumulada |
|---------------------------------|------------------------------------|------|------|------|------|------|---------------------|
| | MR1 | MR2 | MR3 | MR5 | MR4 | MR6 | |
| Matriz de correlação de Pearson | 0,11 | 0,09 | 0,08 | 0,08 | 0,08 | 0,07 | 0,51 |
| Matriz de correlação policórica | MR1 | MR3 | MR5 | MR2 | MR6 | MR4 | 0,54 |

Fonte: Própria (2017).

Apesar de haver uma ampla literatura sobre os principais aspectos das AFE, há certa lacuna sobre o tópico da variância explicada (PETERSON, 2000). Tal limitação deixa, por vezes, os pesquisadores no dilema de tentar compreender o valor percentual obtido em suas análises, questionando-se sobre a adequação ou inadequação da estrutura fatorial, bem como a aceitabilidade dos índices obtidos (DAMÁSIO, 2012).

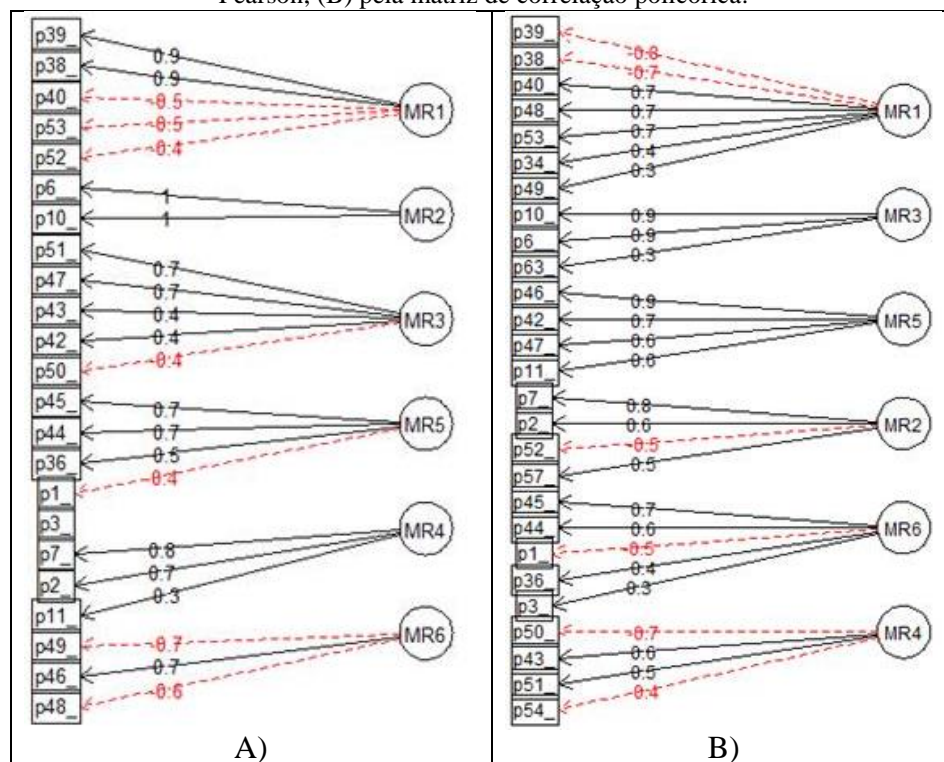
De acordo com Tinsley e Tinsley (1987), soluções fatoriais que expliquem apenas entre 30% a 40% da variância comum entre os itens sugerem uma ampla porcentagem de variância não-explicada (resíduos). Por meio de um estudo de revisão da literatura, os autores concluíram que, em geral, as soluções fatoriais nas pesquisas explicavam menos de 50% da variância total (DAMÁSIO, 2012). Peterson (2000) realizou um estudo meta-analítico com o objetivo de avaliar os níveis de variância explicada nos estudos que utilizaram AFs, bem como suas relações com aspectos referentes ao delineamento metodológico utilizado em tais estudos. De acordo com os critérios de inclusão/exclusão adotados pelo autor, foram avaliadas 803 análises fatoriais, reportadas em 568 artigos, publicados entre 1964 e 1999. Do total das AFs executadas, 67% utilizaram o método dos componentes principais e 82% utilizaram o método de rotação *Varimax*. A média da variância explicada foi de 56,6%. Do total das AFs avaliadas, 10% apresentaram variância explicada maior que 76%, e outros 10% apresentaram variância explicada menor que 34%. Uma vez que os resultados das variâncias do presente estudo foram de 51% e 54% respectivamente, constata-se uma boa representatividade da variabilidade dos dados.

As Tabelas 2 e 3 apresentam os fatores, as cargas fatoriais correspondentes e a comunalidade pela matriz de Correlação de Pearson e Policórica, respectivamente.

Na AFE, com base na matriz de correlação de Pearson (Tabela 2), é possível constatar, a partir dos valores das comunalidades, que a variável p6 - área da casa é melhor explicada pelos fatores (com 97%) e a pior é a p11 - material do telhado (com 14%). Já na AFE pela matriz de correlação policórica (Tabela 3), a partir dos valores das comunalidades, é possível constatar que a variável que é melhor explicada pelos fatores é a p46 - Forma de retirar água da cisterna (com 90%) e a pior é a p63 - per capita (com 15%).

A Figura 1 apresenta os resultados encontrados da análise fatorial com os dois métodos utilizados. Pela matriz de correlação de Pearson via *Varimax*, algumas variáveis apresentaram baixas cargas fatoriais, implicando na sua exclusão pelo próprio procedimento. Pela matriz de correlação policórica todas as variáveis apresentaram cargas fatoriais que justificaram o agrupamento em pelo menos um dos fatores. Silva *et al.* (2006) utilizou análise fatorial pela matriz de correlação de Pearson via *Varimax* para analisar Índice de sustentabilidade do uso da água de cisternas no Semiárido brasileiro, permitiu formação de 4 fatores. De acordo com Moraes (2016), utilizou análise fatorial com componentes principais para estudo de Cisternas domiciliares para qualidade da água, foi formado 2 fatores e algumas variáveis apresentaram baixas cargas fatoriais.

Figura 1 – Diagrama da análise fatorial exploratória no campo das cisternas (A) pela matriz de correlação de Pearson; (B) pela matriz de correlação policórica.



Fonte: Própria (2017).

Tabela 2 – Cargas fatoriais das variáveis estudadas e comunalidade pela matriz de correlação de Pearson.

| | MR1 | MR2 | MR3 | MR5 | MR4 | MR6 | CUM |
|-----|-------|-------|-------|-------|-------|-------|-------------|
| p1 | 0,04 | -0,04 | 0,13 | -0,39 | 0,27 | 0,08 | 0,25 |
| p2 | 0,08 | 0,03 | -0,03 | 0,12 | 0,74 | 0,13 | 0,59 |
| p3 | 0,00 | 0,06 | 0,06 | 0,29 | 0,19 | 0,28 | 0,20 |
| p6 | 0,02 | 0,98 | -0,03 | -0,07 | 0,00 | 0,02 | 0,97 |
| p7 | -0,01 | 0,08 | -0,09 | -0,19 | 0,78 | -0,05 | 0,66 |
| p10 | 0,02 | 0,98 | -0,03 | -0,07 | 0,00 | 0,02 | 0,97 |
| p11 | -0,10 | -0,08 | 0,07 | 0,13 | 0,32 | -0,04 | 0,14 |
| p36 | 0,24 | -0,17 | 0,11 | 0,49 | -0,06 | 0,25 | 0,40 |
| p38 | 0,89 | 0,08 | 0,03 | 0,14 | -0,03 | 0,04 | 0,82 |
| p39 | 0,89 | 0,10 | 0,02 | 0,13 | -0,03 | 0,06 | 0,83 |
| p40 | -0,53 | 0,00 | -0,29 | 0,02 | -0,15 | 0,04 | 0,39 |
| p42 | 0,02 | -0,01 | 0,36 | -0,20 | 0,02 | 0,14 | 0,19 |
| p43 | -0,01 | -0,03 | 0,45 | -0,06 | 0,10 | -0,06 | 0,22 |
| p44 | 0,12 | -0,03 | 0,10 | 0,69 | 0,20 | -0,13 | 0,56 |
| p45 | -0,09 | -0,08 | -0,17 | 0,74 | 0,03 | -0,08 | 0,60 |
| p46 | -0,10 | 0,14 | 0,23 | 0,07 | -0,11 | 0,70 | 0,58 |
| p48 | -0,36 | 0,07 | 0,19 | 0,20 | 0,10 | -0,56 | 0,53 |
| p47 | 0,00 | 0,04 | 0,74 | 0,16 | 0,10 | 0,20 | 0,63 |
| p49 | -0,05 | 0,06 | -0,01 | 0,13 | -0,15 | -0,72 | 0,57 |
| p50 | 0,05 | 0,07 | -0,36 | 0,15 | 0,28 | 0,19 | 0,27 |
| p51 | 0,14 | 0,00 | 0,75 | 0,12 | -0,21 | 0,02 | 0,64 |
| p52 | -0,37 | 0,07 | -0,12 | 0,15 | -0,33 | 0,24 | 0,35 |
| p53 | -0,52 | 0,15 | 0,26 | 0,09 | 0,13 | -0,12 | 0,40 |

Fonte: Própria (2017).

Tabela 3 – Cargas fatoriais das variáveis estudadas e comunalidade calculada pela matriz de correlação de policórica.

| | MR1 | MR3 | MR5 | MR2 | MR6 | MR4 | CUM |
|-----|-------|-------|-------|-------|-------|-------|-------------|
| p1 | -0,01 | -0,05 | 0,18 | 0,25 | -0,47 | 0,21 | 0,37 |
| p2 | -0,20 | -0,11 | -0,03 | 0,62 | 0,06 | -0,10 | 0,46 |
| p3 | -0,05 | 0,26 | 0,20 | 0,11 | 0,34 | -0,11 | 0,25 |
| p6 | -0,06 | 0,91 | -0,13 | -0,04 | -0,09 | 0,04 | 0,85 |
| p7 | 0,22 | 0,10 | 0,03 | 0,80 | -0,01 | -0,12 | 0,71 |
| p10 | -0,06 | 0,91 | -0,13 | -0,04 | -0,09 | 0,04 | 0,85 |
| p11 | 0,33 | 0,04 | 0,57 | 0,40 | 0,29 | 0,01 | 0,68 |
| p34 | 0,37 | -0,20 | 0,22 | 0,05 | -0,11 | 0,15 | 0,26 |
| p36 | -0,24 | 0,23 | 0,08 | -0,12 | 0,44 | 0,44 | 0,53 |
| p38 | -0,72 | 0,06 | 0,16 | 0,15 | 0,19 | 0,16 | 0,63 |
| p39 | -0,76 | -0,01 | 0,04 | 0,03 | 0,15 | 0,16 | 0,62 |
| p40 | 0,68 | 0,26 | 0,23 | -0,34 | 0,17 | 0,00 | 0,72 |
| p42 | 0,04 | -0,01 | 0,67 | -0,05 | -0,24 | 0,11 | 0,52 |
| p43 | 0,11 | 0,08 | 0,07 | 0,15 | -0,02 | 0,57 | 0,38 |
| p44 | -0,18 | -0,07 | 0,04 | 0,39 | 0,62 | 0,13 | 0,60 |
| p45 | 0,10 | -0,21 | 0,06 | 0,01 | 0,70 | -0,04 | 0,55 |
| p46 | -0,09 | -0,08 | 0,86 | 0,00 | 0,04 | -0,38 | 0,90 |
| p48 | 0,66 | -0,08 | -0,18 | 0,07 | 0,50 | 0,08 | 0,74 |
| p47 | -0,01 | 0,07 | 0,58 | -0,04 | 0,22 | 0,37 | 0,53 |
| p49 | 0,33 | 0,25 | -0,08 | 0,20 | 0,07 | 0,12 | 0,23 |
| p50 | -0,07 | 0,14 | 0,19 | 0,27 | 0,26 | -0,72 | 0,72 |
| p51 | -0,19 | -0,04 | 0,41 | -0,03 | 0,12 | 0,51 | 0,48 |
| p52 | 0,17 | -0,23 | 0,14 | -0,55 | 0,12 | -0,37 | 0,55 |
| p53 | 0,66 | -0,24 | 0,11 | 0,14 | 0,07 | 0,11 | 0,54 |
| p54 | 0,13 | 0,34 | 0,11 | 0,34 | 0,10 | -0,35 | 0,39 |
| p57 | 0,10 | 0,43 | 0,22 | 0,46 | 0,10 | 0,13 | 0,48 |
| p63 | -0,08 | 0,31 | 0,17 | 0,14 | 0,00 | -0,03 | 0,15 |

Fonte: Própria (2017).

A partir dos resultados obtidos da AFE pelo método da correlação de Pearson, podem-se organizar as variáveis nos fatores da seguinte forma:

– MR1 - observou-se que nesse fator foram agrupadas variáveis das dimensões técnica (“Existência de sangrador com tamponamento na Cisterna”, “Desvio das primeiras águas” e “Distância da cisterna das instalações de esgotos”) e as variáveis da dimensão sanitária (“Estado de limpeza dos recipientes” e “Recipiente usado para armazenar a água retirada da cisterna”);

– MR2 - relacionaram-se nesse fator duas variáveis sendo uma da dimensão social (“Área da casa (m²)” e uma da dimensão técnica (“Área do telhado - m²”),

– MR3 - todas as variáveis presentes nesse fator fazem parte da dimensão técnica, sendo elas: “Tempo de construção da cisterna”, “Estado da bomba”, “Estado da tampa da cisterna”, “Estado da estrutura da cisterna” e “Manutenção”;

– MR5 - a maioria das variáveis que fazem parte desse fator são da dimensão sanitária, sendo elas: “Forma de limpeza” e “Limpeza da cisterna”. Mas também encontra-se nesse fator a variável “Existência da canalização condutora da água de chuva captada para a cisterna” que faz parte da dimensão técnica e a variável “Quantidade de pessoas no domicílio” da dimensão social;

– MR4 - houve uma predominância de variáveis da dimensão social, encontrando nesse fator as variáveis “Material das paredes” e “Bens que existem na casa”. Somou-se a essas variáveis o “Material do telhado”; e,

– MR6 - nesse fator todas as variáveis existentes são da dimensão sanitária, sendo elas: “Caso a água seja retirada da cisterna com balde: O balde é usado para outros usos? ”, “Forma de retirar a água da cisterna” e “Caso a água seja retirada da cisterna com balde: Estado de conservação e limpeza do balde”.

Diante da análise feita sobre os grupos de fatores pode-se perceber que foi necessário eliminar as variáveis “Existência de calha de captação de água de chuva”; “Reserva da água de beber”; “Educação sanitária e capacitação para uso da cisterna”; “per capita” do grupo de variáveis para que o método já que essas variáveis apresentaram carga fatorial muito baixa. Essas variáveis não apresentaram variação nas respostas, logo, suas variâncias foram nulas, causando problema na singularidade da matriz de correlação de Pearson.

A variável “Anos de estudo do chefe da família” não teve representatividade na AFE pela matriz de correlação Pearson, e em decorrência disso, não foi agrupada em nenhum fator.

No caso do método para a aplicação da matriz da correlação policórica foi possível agrupar todas as variáveis estudadas nos fatores. Porém, conforme já foi exposto, o primeiro fator, devido ao seu valor da variância explicada, tem a capacidade de explicar as demais, e, portanto, é o fator de maior significância do estudo. As variáveis foram agrupadas nos fatores na AFE pela matriz de correlação policórica da seguinte forma:

– MR1 - Predominou a presença de variáveis da dimensão técnica, sendo elas: “Existência de sangrador com tamponamento na Cisterna”, “Desvio das primeiras águas”, “Distância da cisterna das instalações de esgotos”, Existência da calha de captação da água de chuva na Cisterna”. Também participa desse fator variáveis da dimensão sanitária, sendo: “Caso a água seja retirada da cisterna com balde: Estado de conservação e limpeza do balde (observar) ”, “Estado de limpeza dos recipientes” e “Caso a água seja retirada da cisterna com balde: O balde é usado para outros usos? ”. Independente da dimensão, as variáveis desse fator têm relação direta com a qualidade da água;

– MR2 - a maioria das variáveis presentes nesse fator são da dimensão social (“Material das paredes”, “Bens que existem na casa”). Porém, encontrou-se também nesse fator as variáveis: “Recipiente usado para armazenar a água retirada da cisterna” e “Educação sanitária e capacitação para uso da cisterna”. Essa última, apesar de estar em um fator com variância

explicada de 9%, tem grande importância no estudo, pois as ações de educação sanitária reuniram informações capazes de influenciar diretamente na qualidade e quantidade de água consumida;

– MR3 - Esse fator reuniu variáveis que caracterizam a quantidade de água a ser armazenada e consumida pela família, sendo elas: “Área da casa (m²)”, “Área do telhado (m²)” e “Per capita”. – MR4 - as variáveis desse fator impactam diretamente na qualidade da água armazenada na cisterna sendo elas: “Manutenção”, “Estado da tampa da cisterna”, “Tempo de construção da cisterna” e “Reserva da água de beber”.

– MR5 - as variáveis agrupadas nesse fator fazem parte das dimensões sanitária (“Forma de retirar a água da cisterna” e “Material do telhado”) e técnica (“Estado da estrutura da cisterna”, e “Estado da bomba”). Todas elas impactam na qualidade da água armazenada na cisterna;

– MR6 - Neste fator agrupou-se as variáveis “Forma de limpeza”, “Limpeza da cisterna”, “Quantidade de pessoas no domicílio”, “Existência da canalização condutora da água de chuva captada para a cisterna” e “Anos de estudo do chefe da família”.

O primeiro fator gerado pela correlação de Pearson e policórica reuniram variáveis com características das dimensões técnicas e sanitárias. Todas elas apresentam relação direta com a qualidade da água armazenada na cisterna. Porém, as variáveis da dimensão sanitária agrupadas pela matriz de correlação policórica, melhor explicam a interferência na qualidade da água.

A qualidade da água se agrava quando o balde utilizado para retirar a água da cisterna não tem conservação, não existe local adequado para seu armazenamento, e ainda recebe outros usos além de retirar água da cisterna. A forma como o balde é guardado pode permitir o contato direto com animais, impurezas e facilidade para usos indevidos. Segundo Blackburn *et al.* (2003), as fontes de contaminação das águas de cisternas com coliformes estão relacionadas ao mau acondicionamento dos baldes; à presença de animais nas instalações e estruturas de captação e no manuseio inadequado do balde quando da retirada da água nas cisternas.

Os dois métodos estudados mostraram que existe uma preponderância entre os fatores, quando esses apresentam variação nas variâncias explicadas (o primeiro fator tem o valor da variância explicada superior aos demais e sucessivamente). A pouca variação, no entanto, entre os valores das variâncias explicadas foi devido à grande quantidade de variáveis estudadas. A presença de tantas variáveis com temáticas diversificadas (dimensões técnica, sanitária, Institucional, social e ambiental) fez-se necessária devido à quantidade de informações importantes a serem analisadas quando se estuda o uso e funcionamento das cisternas para consumo humano. Mesmo não sendo de grande variação os valores das variâncias explicadas, elas atenderam as considerações de parcimônia da AFE.

Diante dos resultados, considerando valores do KMO, variância explicada, percentual de variáveis presentes no estudo e características dos fatores gerados, pode-se afirmar que a AFE para o estudo do uso e funcionamento da cisterna pela matriz de correlação policórica foi a que melhor explicou os resultados.

4. CONCLUSÃO

O trabalho teve como premissa aplicar métodos de Análise Fatorial para identificar qual deles melhor relaciona as variáveis estudadas, a partir da formação de grupos de variáveis que se traduzam em respostas para estudar as variáveis relacionadas ao uso e funcionamento do sistema de captação de água de chuva implantado pelo PIMC. Os métodos estudados foram matriz de correlação policórica e de Pearson pela rotação *Varimax*.

A área de estudo envolveu os municípios de Abaré, Chorrochó, Macururé, Glória e Santa Brígida no semiárido Baiano, com uma amostra de 455 cisternas. O questionário proposto

foi formado por 36 perguntas objetivas, capturou informações relacionadas com a cisterna no campo técnico, sanitário, ambiental, social e institucional.

O teste de Kaiser-Meyer-Olkin (KMO) realizado para avaliar se a análise fatorial é apropriado (0,59), indicando bons resultados para as correlações. Os resultados evidenciaram que a variabilidade dos dados de 36 variáveis nos dois métodos pode ser explicada por seis fatores. Esses fatores foram reunidos a partir das suas cargas fatoriais. As variáveis dos dois primeiros fatores (Fator 1 - MR1 e Fator 2 - MR3), gerados nos métodos estudados explicaram a grande variância dos dados, sendo os fatores mais importantes desde estudo.

O Fator 1, mais representativo, para os dois métodos estudados, relacionou as variáveis das dimensões técnica e sanitária, relacionadas o projeto executivo e à eficiência das ações de educação sanitária. Muitas das variáveis são influenciadas por essas ações, já que elas estão voltadas para a formação dos usuários no uso e funcionamento das cisternas. Essa formação, possibilita ao usuário ter informações referentes à forma de utilizar a cisterna, forma de limpeza, manutenção, dentre outros temas abordados. A cisterna é uma tecnologia que apesar de necessitar do órgão público para a assistência técnica, fiscalização e manutenção tem forte influência das práticas quanto ao manuseio e cuidados das famílias, especialmente quanto à qualidade de água.

No Fator 1 da correlação policórica ainda pôde-se perceber a influência da variável “uso de balde” para retirar água da cisterna. A forma apropriada de retirar a água da cisterna é por meio da bomba, já acoplada na cisterna. O uso da bomba, além de evitar o contato de objetos do meio externo com a água, também evita a constante abertura da tampa, evitando entrada de impurezas.

Ainda pôde-se identificar nos resultados encontrados para os dois métodos em estudo que a variância explicada atendeu as considerações de parcimônia, sendo primeiro fator o que mais e melhor explica as demais variáveis, seguido dos demais fatores, com valores das variâncias decrescendo.

Os fatores gerados na correlação policórica permitiram agrupar todas as variáveis em questão, garantindo que houvesse uma identificação de quais delas melhor representam as demais.

Assim, o estudo pôde indicar que a AFE para o uso e funcionamento da cisterna pela matriz de correlação policórica foi a que melhor explicou os resultados.

5. REFERÊNCIAS

ANDRADE NETO, C. O. Aproveitamento imediato da água de chuva. *Revista Eletrônica de Gestão e Tecnologias Ambientais*, v.1, p. 67-80, ISSN: 2317-563X, 2013.

_____. Aproveitamento imediato da água de chuva. **Revista Eletrônica de Gestão e Tecnologias Ambientais**, v.1, p. 67-80, 2013.

ARTICULAÇÃO NO SEMIÁRIDO BRASILEIRO (ASA). **AÇÕES – P1MC**. Disponível em: <<http://www.asabrasil.org.br/acoes/p1mc>>. Acesso em: 09 mar. 2017.

MINAYO, M. C. S. Pesquisa avaliativa por triangulação de métodos. **In: Avaliação por triangulação de métodos: abordagens de programas sociais**. Rio de Janeiro: Fiocruz, 2005.

BISTAFFA, B.C. **Incorporação de indicadores categóricos ordinais em modelos de equações estruturais**. 142p. Dissertação (Mestrado em Estatística). Universidade de São Paulo. 2010.

BLACKBURN, D.M.; RETAMAL, Y. P. B.; LIMA, M.; JALFIM, F.; VIANA, A. A.; FARIAS JÚNIOR. M. Avaliação da contaminação microbiológica de água para consumo doméstico na região de atuação da Diaconia no Semi-Árido nordestino. **In: SIMPÓSIO BRASILEIRO DE CAPTAÇÃO E MANEJO DE ÁGUA DE CHUVA**, 4, 2003, Juazeiro, BA. Anais... Juazeiro, BA: ABCMAC; Petrolina, PE: Embrapa Semi-Árido, 2003. 1 CD-ROM.

BRASIL. Ministério da Saúde. Portaria nº 2914, de 12/2011. Diário Oficial da República Federativa do Brasil, Poder Executivo, Brasília, DF, 2011.

BRASIL. Tribunal de Contas da União. **Relatório de avaliação de programa: Ação Construção de Cisternas para Armazenamento de Água**. Tribunal de Contas da União Relator Ministro Guilherme Palmeira. – Brasília: TCU, Secretaria de Fiscalização e Avaliação de Programas de Governo, 2006. 129 p.

BRITO, L. T. L.; GNADLINGER, J. **Relatório sobre a oficina: avanços nos estudos sobre cisternas: qualidade de água e cisterna de alambrado**. Petrolina: ABCMAC, 2006.

BROWN, Timothy A. Confirmatory factor analysis for applied research. **Confirmatory factor analysis for applied research**. 2006.

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

HAIR, J. F.; ANDERSON, R. E.; TATHAM, R. L.; BLACK, W. C. **Análise multivariada de dados**. A. Porto Alegre: Bookman. 2005.

HELMREICH, H.; HORN, H. Opportunities in rainwater harvesting. **Desalination**, v. 248, p. 118-124, 2009.

JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**, 6ª edição, Madison: Prentice Hall International. 2007. 816p.

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, 4-13. 2017.

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.

LORDELO, L M K. **Cisterna de captação de água de chuva para abastecimento de humano: discutindo a efetividade enquanto tecnologia social no semiárido baiano**. Relatório de pesquisas realizadas para fins de exame de qualificação. Programa de pós Graduação de Ciência e Energia. Universidade Federal da Bahia, 2016.

LORDELO, L M K. BORJA P. C. MORAES, L. R. S. PORSANI M. J. Contribuição de parâmetros para construção do índice de efetividade da cisterna – P1MC. **In: 29º Congresso Brasileiro de Engenharia Sanitária e Ambiental**, 2017a.

LORDELO, L M K.; BORJA P. C.; PORSANI M. J.; MORAES, L. R. S. Avaliação do uso e funcionamento das cisternas do PIMC – um estudo no semiárido baiano. **Revista Eletrônica Gestão e Tecnologias Ambientais**, v. 5, n. 2, p. 107-121, 2017b.

MORAES, A. F. J.; ROCHA, C. Gendered waters: the participation of women in the ‘One Million Cisterns’ rainwater harvesting program in the Brazilian Semi-arid region. **Journal of Cleaner Production**, n. 60, p. 163-169, 2013.

MORAIS, G. F. O. **Cisternas domiciliares: qualidade da água para consumo humano em comunidades rurais do semiárido Sergipano**. Dissertação (mestrado). Programa Regional de Pós-Graduação em Desenvolvimento e Meio Ambiente. Universidade Federal da Paraíba, 2016.

NEISSE, A.; 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, 105-115. 2016.

Peterson, R. A. A meta-analysis of variance accounted for and factor loadings in exploratory factor analysis. *Marketing Letters*, 11(3), 261-275, 2000.

RUSKIN, R. H.; CALLENDER, P. S. **Maintenance of cistern water quality in the Virgin Islands**. Technical Report N. 30, Caribbean Research Institute, University of the Virgin Islands, St. Thomas, U. S. Virgin Islands, 1988.

SILVA, A. S.; BRITO, L. T. L.; SILVA, C. M. M. S.; FILHO, P. P. S. **Índice de sustentabilidade do uso da água de cisternas no Semi Árido brasileiro**. Empresa Brasileira de Pesquisa Agropecuária. Ministério da Agricultura, Pecuária e Abastecimento, 2006.

SILVA, N. M. D.; PERELO, L. W.; MORAES, L. R. S. Fatores intervenientes da qualidade microbiológica das águas de chuva armazenadas em cisternas da área rural do município de Inhambupe no semiárido baiano. **In: XII Simpósio Ítalo Brasileiro De Engenharia Sanitária E Ambiental**, 12, 2014, Natal. Anais. Natal: ABES, 2014.

TINSLEY, H. E. A. & Tinsley, D. J. Uses of factor analysis in counseling psychology research. **Journal of Counseling Psychology**, 34(4), 414-424. 1987.