

CAMILA RAFAELA GOMES DIAS

**DESEMPENHO DO MÉTODO DE ESTIMAÇÃO POR COMPONENTES  
PRINCIPAIS E DA ROTAÇÃO VARIMAX NA ANÁLISE DE FATORES**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Estatística Aplicada e Biometria, para obtenção do título de *Magister Scientiae*.

VIÇOSA  
MINAS GERAIS – BRASIL  
2018

**Ficha catalográfica preparada pela Biblioteca Central da Universidade  
Federal de Viçosa - Câmpus Viçosa**

T

D541d  
2018  
Dias, Camila Rafaela Gomes, 1991-  
Desempenho do método de estimação por componentes  
principais e da rotação varimax na análise de fatores / Camila  
Rafaela Gomes Dias. – Viçosa, MG, 2018.  
ix, 44f. : il. ; 29 cm.

Orientador: José Ivo Ribeiro Júnior.

Dissertação (mestrado) - Universidade Federal de Viçosa.

Referências bibliográficas: f.42-44.

1. Análise multivariada. 2. Correlação (Estatística).

I. Universidade Federal de Viçosa. Departamento de Estatística.  
Programa de Pós-graduação em Estatística Aplicada e Biometria.

II. Título.

CDD 22 ed. 519.535

CAMILA RAFAELA GOMES DIAS

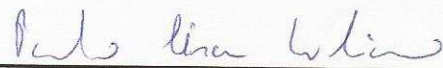
**DESEMPENHO DO MÉTODO DE ESTIMAÇÃO POR COMPONENTES  
PRINCIPAIS E DA ROTAÇÃO VARIMAX NA ANÁLISE DE FATORES**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Estatística Aplicada e Biometria, para obtenção do título de *Magister Scientiae*.

APROVADA: 26 de fevereiro de 2018.



Alexandre Navarro da Silva



Paulo César Emiliano



José Ivo Ribeiro Junior  
(Orientador)

*Aos meus pais Creuza e André e ao meu irmão João Paulo.*

## AGRADECIMENTOS

Agradeço primeiramente a Deus que me deu essa oportunidade, me sustentou, me deu forças e sabedoria para chegar ao passo de concluí-la.

Em segundo, ao meu professor, pai e orientador José Ivo, pois sem ele com certeza não conseguiria ter chegado até aqui. Agradeço não apenas pelo seu papel de professor, que aliás, exerceu com maestria, como também pela sua amizade, por ser sempre alguém com quem pude contar e buscar conselhos. A você minha eterna gratidão.

Agradeço também aos meus pais e ao meu irmão, por contribuírem para que eu alcançasse esse objetivo, por suportarem minhas ausências, meus nervosismos e minhas crises de ansiedade. Por estarem sempre ao meu lado nessa conquista. A vocês todo meu amor.

Aos meus amigos que por infinitas vezes me fortaleceram nessa caminhada, sempre me incentivando e torcendo por mim, em especial a minha amiga Juliana, meu presente de Deus, minha companheira de todas as horas, dentro e fora da UFV.

Aos meus colegas de trabalho, em especial, à Professora Daniela, Diretora da Editora UFV e minha chefe, pelo fundamental apoio e compreensão durante esses dois anos.

À toda minha família, tios e tias, primos e primas, por todo o carinho e orações que recebi, contribuindo sempre para meu fortalecimento e crescimento pessoal.

À FAPEMIG pelo apoio financeiro.

Por fim, agradeço a todas as pessoas que torceram por mim e que, de maneira direta ou indireta, fizeram parte dessa jornada, que eu espero não terminar por aqui, com fé em Deus.

## **BIOGRAFIA**

CAMILA RAFAELA GOMES DIAS, filha de Creuza Gomes Pereira Dias e de André Luiz de Faria Dias, nasceu em Conselheiro Lafaiete, Minas Gerais, em 17 de maio de 1991.

Em março de 2010, ingressou no curso de Licenciatura em Matemática na Universidade Federal de Viçosa, Viçosa-MG, graduando-se em janeiro de 2016.

Em março de 2016, iniciou o curso de Mestrado no Programa de Pós-graduação em Estatística Aplicada e Biometria na Universidade Federal de Viçosa, submetendo-se à defesa de dissertação em 26 de fevereiro de 2018.

## SUMÁRIO

	Página
LISTA DE TABELAS .....	vi
RESUMO .....	viii
ABSTRACT .....	ix
1. INTRODUÇÃO GERAL .....	1
2. OBJETIVOS .....	3
2.1. Geral.....	3
2.2. Específicos .....	3
3. REVISÃO DE LITERATURA .....	4
3.1. Considerações Gerais sobre a Análise de Fatores.....	4
3.2. Matriz de Correlações .....	5
3.3. Método de Estimação dos Componentes Principais .....	7
3.4. Método de Rotação Varimax .....	9
3.5. Viabilidade da Análise de Fatores .....	9
4. MATERIAL E MÉTODOS .....	13
4.1. Modelo Fatorial.....	13
4.2. Fatores Teóricos .....	14
4.3. Matrizes de Correlações.....	18
4.4. Método dos Componentes Principais.....	21
4.4.1. Sem Rotação .....	21
4.4.2. Com Rotação .....	23
4.5. Medidas de Avaliação.....	24
5. RESULTADOS E DISCUSSÃO .....	26
5.1. Matrizes de Correlações.....	26
5.2 Método dos Componentes Principais.....	30
5.2.1. Comunalidade .....	30
5.2.2. Cargas Fatoriais .....	33
5.2.2.1. Sem Rotação .....	33
5.2.2.2. Com Rotação.....	36
5.2.3. Autovalores.....	37
6. CONCLUSÕES .....	41
7. REFERÊNCIAS BIBLIOGRÁFICAS .....	42

## LISTA DE TABELAS

Tabela 3.1 - Classificação da análise de fatores segundo o KMO.....	12
Tabela 4.1 - Matrizes teóricas $\Gamma$ e $\Psi$ para $h_w^2 = 0,49$ ( $w = 1, 2, 3$ e $4$ ) .....	15
Tabela 4.2 - Matrizes teóricas $\Gamma$ e $\Psi$ para $h_w^2 = 0,56$ ( $w = 1, 2, 3$ e $4$ ).....	15
Tabela 4.3 - Matrizes teóricas $\Gamma$ e $\Psi$ para $h_w^2 = 0,64$ ( $w = 1, 2, 3$ e $4$ ) .....	16
Tabela 4.4 - Matrizes teóricas $\Gamma$ e $\Psi$ para $h_w^2 = 0,72$ ( $w = 1, 2, 3$ e $4$ ).....	16
Tabela 4.5 - Matrizes teóricas $\Gamma$ e $\Psi$ para $h_w^2 = 0,81$ ( $w = 1, 2, 3$ e $4$ ) .....	16
Tabela 4.6 - Matrizes teóricas $\Gamma$ e $\Psi$ para $h_w^2 = 0,90$ ( $w = 1, 2, 3$ e $4$ ).....	16
Tabela 4.7 - Autovalores ( $\lambda_j$ ) obtidos para os fatores $F_1$ e $F_2$ , independentemente de $d_w$ , para $h_w^2$ ( $w = 1, 2, 3$ e $4$ ) .....	17
Tabela 4.8 - Matrizes de correlações ( $\rho$ ) representadas pelos elementos da diagonal e acima dela para $h_w^2 = 0,49$ ( $w = 1, 2, 3$ e $4$ ).....	18
Tabela 4.9 - Matrizes de correlações ( $\rho$ ) representadas pelos elementos da diagonal e acima dela para $h_w^2 = 0,56$ ( $w = 1, 2, 3$ e $4$ ).....	18
Tabela 4.10 - Matrizes de correlações ( $\rho$ ) representadas pelos elementos da diagonal e acima dela para $h_w^2 = 0,64$ ( $w = 1, 2, 3$ e $4$ ).....	18
Tabela 4.11 - Matrizes de correlações ( $\rho$ ) representadas pelos elementos da diagonal e acima dela para $h_w^2 = 0,72$ ( $w = 1, 2, 3$ e $4$ ).....	19
Tabela 4.12 - Matrizes de correlações ( $\rho$ ) representadas pelos elementos da diagonal e acima dela para $h_w^2 = 0,81$ ( $w = 1, 2, 3$ e $4$ ) .....	19
Tabela 4.13 - Matrizes de correlações ( $\rho$ ) representadas pelos elementos da diagonal e acima dela para $h_w^2 = 0,90$ ( $w = 1, 2, 3$ e $4$ ).....	19
Tabela 5.1 - Valores calculados e entre parênteses os $p$ – valores do teste de esfericidade de Bartlett e do KMO para $h_w^2 = 0,49$ .....	26
Tabela 5.2 - Valores calculados e entre parênteses os $p$ – valores do teste de esfericidade de Bartlett e do KMO para $h_w^2 = 0,56$ .....	26
Tabela 5.3 - Valores calculados e entre parênteses os $p$ – valores do teste de esfericidade de Bartlett e do KMO para $h_w^2 = 0,64$ .....	26
Tabela 5.4 - Valores calculados e entre parênteses os $p$ – valores do teste de esfericidade de Bartlett e do KMO para $h_w^2 = 0,72$ .....	26

Tabela 5.5 - Valores calculados e entre parênteses os $p$ – valores do teste de esfericidade de Bartlett e do KMO para $h_w^2 = 0,81$ .....	27
Tabela 5.6 - Valores calculados e entre parênteses os $p$ – valores do teste de esfericidade de Bartlett e do KMO para $h_w^2 = 0,90$ .....	27
Tabela 5.7 - Equações de regressão ajustadas para <i>valor calculado</i> , $p$ – valor e KMO .....	28
Tabela 5.8 - Comunalidades ( $h_{wCP}^2$ ) obtidas de acordo com as seis matrizes de correlações e os percentuais de variação em relação às comunalidades teóricas ( $h_w^2$ ), independentemente de $d_w$ .....	30
Tabela 5.9 - Equação de regressão ajustada para a variável $\Delta_{h^2}$ .....	30
Tabela 5.10 - Somas de quadrados dos elementos da matriz de resíduos obtidas de acordo com as seis matrizes de correlações, independentemente de $d_w$ ....	31
Tabela 5.11 - Equação de regressão ajustada para $sqr_{res_{CP}}$ .....	31
Tabela 5.12 - Matrizes $\Gamma_{CP}$ e $\Psi_{CP}$ obtidas para $h_w^2 = 0,49$ ( $w = 1, 2, 3$ e $4$ ).....	33
Tabela 5.13 - Matrizes $\Gamma_{CP}$ e $\Psi_{CP}$ obtidas para $h_w^2 = 0,56$ ( $w = 1, 2, 3$ e $4$ ) .....	33
Tabela 5.14 - Matrizes $\Gamma_{CP}$ e $\Psi_{CP}$ obtidas para $h_w^2 = 0,64$ ( $w = 1, 2, 3$ e $4$ ).....	34
Tabela 5.15 - Matrizes $\Gamma_{CP}$ e $\Psi_{CP}$ obtidas para $h_w^2 = 0,72$ ( $w = 1, 2, 3$ e $4$ ).....	34
Tabela 5.16 - Matrizes $\Gamma_{CP}$ e $\Psi_{CP}$ obtidas para $h_w^2 = 0,81$ ( $w = 1, 2, 3$ e $4$ ).....	34
Tabela 5.17 - Matrizes $\Gamma_{CP}$ e $\Psi_{CP}$ obtidas para $h_w^2 = 0,90$ ( $w = 1, 2, 3$ e $4$ ).....	34
Tabela 5.18 - Equação de regressão ajustada para $\Delta_{\Gamma}$ .....	35
Tabela 5.19 - Matrizes $\Gamma_{CPR}$ obtidas para as seis comunalidades $h_w^2$ ( $w = 1, 2, 3$ e $4$ )...36	36
Tabela 5.20 - Autovalores ( $\lambda_{j_{CP}}$ ) obtidos para as seis comunalidades $h_w^2$ ( $w = 1, 2, 3$ e $4$ )38	38
Tabela 5.21 - Autovalores ( $\lambda_{j_{CP}}$ ) obtidos após a rotação para as seis comunalidades $h_w^2$ ( $w = 1, 2, 3$ e $4$ ).....	39
Tabela 5.22 - Equações de regressão ajustadas a partir dos autovalores estimados para as variáveis $\Delta_{\lambda}$ , $\Delta_{\lambda_1}$ e $\Delta_{\lambda_2}$ .....	40

## RESUMO

Dias, Camila Rafaela Gomes, M.Sc., Universidade Federal de Viçosa, fevereiro de 2018. **Desempenho do método de estimação por componentes principais e da rotação varimax na análise de fatores.** Orientador: José Ivo Ribeiro Júnior. Coorientadores: Ana Carolina Campana Nascimento e Moysés Nascimento.

Com o objetivo de avaliar o desempenho do método dos componentes principais utilizando a análise de fatores ortogonais, com e sem rotação varimax, a partir das estimativas dos seus parâmetros, foram estabelecidas 18 matrizes de correlações que configuraram diferentes graus de relações entre quatro variáveis aleatórias normalmente distribuídas. Inicialmente, avaliou-se a adequação das matrizes de correlações para a aplicação da análise de fatores pelo teste de esfericidade de Bartlett e pelo critério de KMO. Definiram-se como variáveis explicativas os autovalores, a matriz de cargas fatoriais, a comunalidade e a soma de quadrados da matriz de resíduos. Em seguida, foram calculados os erros relativos médios dos desvios entre os valores teóricos e estimados dessas variáveis, definidos, respectivamente, por:  $\Delta_\lambda$ ,  $\Delta_{\lambda_1}$ ,  $\Delta_{\lambda_2}$ ,  $\Delta_\Gamma$  e  $\Delta_{h^2}$ . Além disso, calculou-se também a soma de quadrados da matriz de resíduos para avaliar a qualidade de ajuste de cada modelo fatorial ortogonal. Ao total, foram efetuadas 36 análises de fatores, sem e com rotação varimax. Posteriormente, para cada variável avaliada foi realizada uma análise de superfície de resposta. Concluiu-se que para os dados que seguem distribuição normal p-variada o método de estimação da análise de fatores ortogonais por componentes principais não foi adequado e a rotação varimax não melhorou os resultados quando utilizado com esse método. As análises estatísticas foram realizadas no software R.

## ABSTRACT

Dias, Camila Rafaela Gomes, M.Sc., Universidade Federal de Viçosa, February, 2018. **Performance of the principal components estimation method and the varimax rotation in factor analysis.** Adviser: José Ivo Ribeiro Júnior. Co-advisers: Ana Carolina Campana Nascimento and Moysés Nascimento.

To evaluate the performance of the main components method using orthogonal factors analysis, with and without varimax rotation, 18 matrices of correlations were established from the estimates of their parameters, which set up different degrees of relations between four random normally distributed variables. Firstly, the adequacy of correlation matrices for the application of factor analysis by the Bartlett's sphericity test and the KMO criterion was evaluated. The eigenvalues, the matrix of factorial loads, the commonality and the sum of squares of the residue matrix were defined as explanatory variables. Then, the mean relative errors of the deviations between the theoretical and estimated values of these variables, defined respectively as:  $\Delta_{\lambda}$ ,  $\Delta_{\lambda_1}$ ,  $\Delta_{\lambda_2}$ ,  $\Delta_{\Gamma}$  e  $\Delta_{h^2}$  were calculated. In addition, the sum of squares of the residue matrix was also calculated to evaluate the adjustment quality of each orthogonal factorial model. In total, 36 factor analyzes were carried out, with and without varimax rotation. Subsequently, a response surface analysis was performed for each evaluated variable. It was concluded that for the data following normal p-varied distribution, the method of estimation of orthogonal factors analysis by principal components was not adequate and the varimax rotation did not improve the results when used with this method. Statistical analyzes were carried out in software R.

## 1. INTRODUÇÃO GERAL

A Estatística Multivariada baseia-se em um conjunto de métodos que analisam situações nas quais várias variáveis são medidas simultaneamente em cada elemento amostral e pode ser separada em dois grupos. O primeiro é composto por métodos exploratórios de sintetização (ou simplificação) da estrutura de variabilidade dos dados e, o segundo, por métodos de inferência estatística (MINGOTI, 2005).

Entre os métodos utilizados para a redução da estrutura de variabilidade dos dados destaca-se a análise de fatores, que basicamente consiste em encontrar um modo de sintetizar a informação contida em um grupo de variáveis originais em outro grupo menor de novas variáveis latentes, chamadas de fatores, e com perda mínima de informação (HAIR et al., 2014).

Pode-se dizer que, de modo geral, o objetivo da análise de fatores é identificar as relações de covariâncias entre as variáveis originais e descrever essas relações usando alguns poucos fatores (CRUZ et al., 2014).

No modelo fatorial, cada variável analisada é definida como sendo uma combinação linear de fatores, de modo que esta combinação represente a decomposição da variabilidade das variáveis. Considera-se que cada variável pode ser explicada por um conjunto de fatores comuns e por um fator específico (CRUZ et al., 2014), de tal modo que a cada fator corresponde um conjunto de variáveis altamente correlacionadas.

O resultado de uma análise de fatores depende da matriz de covariâncias utilizada e não é invariante à mudança de escala. Além disso, a solução também não é única, ou seja, existem diferentes soluções capazes de reproduzir a mesma matriz, e isto depende do número de fatores comuns do modelo, bem como do método de estimação e de rotação utilizados na análise.

Segundo Ferreira (2011), a não unicidade para os parâmetros do modelo fatorial pode ser utilizada de forma favorável para melhorar a interpretação dos fatores. Por outro lado, ela pode gerar dúvidas por possibilitar diferentes resultados para um mesmo conjunto de dados.

A análise de fatores surgiu no início do século XX, quando pesquisadores tentaram definir e mensurar a inteligência humana. Nomes importantes como Karl Pearson e Thurstone surgem nesse contexto, mas foi Charles Spearman considerado o “pai” dessa teoria (HARMAN, 1976; JOHNSON; WICHERN, 2007).

Mingoti (2005) cita alguns exemplos de trabalhos de outros autores sobre o uso da análise de fatores, como no posicionamento de produtos e serviços perante os concorrentes de mercado, para identificar o perfil de consumidores, na descoberta dos fatores relacionados à satisfação do indivíduo com sua vida pessoal, atividade profissional ou com produtos de um modo geral.

Já na Ciências Agrárias alguns trabalhos foram desenvolvidos para determinar a capacidade de combinação de linhagens de matrizes de corte (ABREU et al., 1999), para auxiliar na seleção genômica de suínos (TEIXEIRA et al., 2015) e para avaliar a dependência espacial dos atributos químicos do solo (ALMEIDA; GUIMARÃES, 2016).

Nesse campo do conhecimento, a análise de fatores é de ampla aplicação entre os melhoristas e a sua utilização pelo modelo fatorial ortogonal, ou seja, fatores não correlacionados, viabiliza uma solução para o problema de multicolinearidade (FERREIRA et al., 2005).

Como a análise de fatores nem sempre proporciona um ajuste satisfatório, bem como vários tipos de ajustes, dúvidas surgem quando se pretende realizar esta análise, tendo em vista que esse método dispõe de vários critérios de estimação dos parâmetros do modelo e também conta com alguns artifícios, como rotações, que podem ser realizados na tentativa de se alcançar uma estrutura entre os fatores mais simples de ser interpretada.

Ademais, grande parte dos trabalhos publicados em que se realiza a análise de fatores, não relata o método de estimação utilizado, mas somente a realização, ou não, da rotação e o método de rotação aplicado.

Isso sugere, portanto, que nesses trabalhos, utilizou-se o método de estimação dos componentes principais, dado que para a análise de fatores ele é o mais simples (FERREIRA, 2011). Portanto, é de suma importância testar sua eficiência em prover estimativas adequadas na análise de fatores.

Sendo a análise de fatores um método de análise de dados amplamente conhecido e aplicado, surge a necessidade de compreender, entre as várias possibilidades de determinações dos parâmetros do modelo, o método de estimação dos componentes principais, verificando a sua eficácia em função de diferentes situações teóricas definidas *a priori*. Além disso, buscou-se verificar também se a rotação auxilia na melhoria da estimativa e da interpretação, sob as mesmas situações.

## **2. OBJETIVOS**

### **2.1. Geral**

Avaliar o desempenho do método dos componentes principais utilizado na análise de fatores, sem e com rotação varimax, a partir das estimativas dos seus parâmetros em diferentes configurações de graus de relações entre variáveis aleatórias normalmente distribuídas.

### **2.2. Específicos**

- Verificar as eficiências do teste de esfericidade de Bartlett e do critério de Kaiser-Meyer-Olkin (KMO), sobre a adequação ou não de diferentes matrizes de correlações para a análise de fatores.
- Verificar o efeito da comunalidade sobre a adequação do uso da análise de fatores.
- Verificar o efeito do grau de relação entre as variáveis na qualidade da análise de fatores.
- Verificar os efeitos da comunalidade teórica e das diferenças entre as correlações das variáveis nas estimativas dos autovalores, das cargas fatoriais, da matriz residual e da própria comunalidade.
- Verificar o efeito da rotação varimax sobre a interpretação da análise de fatores.

### 3. REVISÃO DE LITERATURA

#### 3.1. Considerações Gerais sobre a Análise de Fatores

Existem dois tipos de análise de fatores: a exploratória e a confirmatória. Conforme Hair et al. (2014), a análise de fatores exploratória é útil na busca da estrutura de um conjunto de dados ou na redução da dimensão das variáveis em estudo quando, *a priori*, não se estabelece restrições sobre a estimação de fatores e nem do número de fatores a serem extraídos. Já na análise de fatores confirmatória, tem-se uma ideia preconcebida da real estrutura dos dados, ou seja, tem-se um modelo fatorial pré-especificado e, pretende-se, portanto, testar hipóteses para avaliar o quanto os dados se aproximam desse modelo.

A análise de fatores exploratória ou aqui simplesmente denominada análise de fatores é um método estatístico multivariado utilizado para identificar uma possível estrutura latente dos dados e que fornece as ferramentas necessárias para estudar a estrutura de interdependência de um grande número de variáveis com o propósito de reduzir, sem perda significativa de informação, a dimensão dos dados.

Os fatores são dimensões latentes que separaram variância comum e compartilhada entre as variáveis de qualquer variação de erro única ou de medida dentro das variáveis. Assim, acredita-se que um fator representa a dimensão verdadeira subjacente em um conjunto de variáveis, depois de remover a porção de variância nas variáveis que não é comum às outras, isto é, única ou devido ao erro (HARLOW, 2005).

Supõe-se que na análise de fatores as correlações ou covariâncias entre um conjunto de variáveis surgem das relações dessas variáveis com um pequeno número de variáveis implícitas, não observáveis, chamadas de variáveis latentes e usualmente conhecidas como fatores ou fatores comuns. Os fatores são variáveis latentes que geram as variáveis originais (RENCHE; CHRISTENSEN, 2012). A cada fator corresponde um subconjunto das variáveis originais altamente correlacionadas entre si, mas que tenham baixa correlação com as demais variáveis de um outro fator.

Para Hair et al. (2014), a análise de fatores comum considera apenas a variância comum entre as variáveis, assumindo que a variância específica do erro não é de interesse na definição da estrutura dessas variáveis. Portanto, os fatores comuns que são estimados por essa análise são baseados apenas na variância comum ou compartilhada pelos dados.

De acordo com Ferreira (2011), os fatores comuns são responsáveis por explicar as variâncias e as covariâncias das variáveis em estudo, enquanto que as variáveis específicas abrangem apenas as variâncias das variáveis consideradas no modelo.

Cada fator representa um conjunto de variáveis altamente correlacionadas e eles devem ser estimados de modo que incorporem o caráter e a natureza das variáveis de seu grupo a fim de reduzir a redundância entre essas variáveis. O objetivo da simplificação do banco de dados é tornar mais fácil e precisa a interpretação das variáveis e das relações entre elas.

Segundo Ferreira (2011) e Rencher e Christensen (2012), cada variável é expressa como uma combinação linear de fatores. Entretanto, é possível olhar cada fator como sendo uma função do conjunto inteiro de variáveis observadas (HAIR et al., 2014).

### 3.2. Matriz de Correlações

Para iniciar uma análise de fatores, deve-se verificar a estrutura de interdependência do conjunto de dados, sendo que ela pode ser representada pela matriz de covariâncias das variáveis originais ( $\Sigma_Y$ ) ou padronizadas ( $\Sigma_Z$ ). Vale ressaltar que quando a padronização das variáveis originais é feita com base nas suas respectivas médias e desvios-padrão, obtém-se novas variáveis centradas em zero e com variâncias iguais a um. Nesse caso, a matriz de covariâncias das variáveis padronizadas ( $\Sigma_Z$ ) coincide com a matriz de correlações das variáveis originais ( $\rho_Y$ ).

Usualmente, opta-se por analisar as variáveis em escala padronizada, ou seja, trabalhar com a matriz de correlações das variáveis originais ( $\rho_Y = \rho$ ), com o objetivo de minimizar o efeito da diferença entre as unidades de medidas ou escalas das variáveis originais.

Dessa forma, dada uma matriz  $Y$ , de ordem  $n \times p$ , representando  $p$  variáveis  $Y_1, Y_2, \dots, Y_p$ , com  $n$  elementos amostrais, considere  $\mu' = [\mu_1 \mu_2 \dots \mu_p]'$  o vetor de médias das variáveis  $Y_s$  e  $\rho$  a matriz de correlações dos dados originais. Desse modo, o modelo da análise de fatores é definido como segue:

$$\begin{aligned} Y_1 - \mu_1 &= \gamma_{11}F_1 + \gamma_{12}F_2 + \dots + \gamma_{1k}F_k + \epsilon_1; \\ Y_2 - \mu_2 &= \gamma_{21}F_1 + \gamma_{22}F_2 + \dots + \gamma_{2k}F_k + \epsilon_2; \\ &\vdots \\ Y_p - \mu_p &= \gamma_{p1}F_1 + \gamma_{p2}F_2 + \dots + \gamma_{pk}F_k + \epsilon_p. \end{aligned}$$

Matricialmente, essa definição pode ser expressa por:

$$\mathbf{D}^{-1}(\mathbf{Y} - \boldsymbol{\mu}) = \boldsymbol{\Gamma}\mathbf{F} + \boldsymbol{\epsilon},$$

em que  $\boldsymbol{\Gamma} = [\gamma_{wj}]$  é uma matriz de ordem  $p \times k$ , cujos coeficientes  $\gamma_{wj}$  são chamados *loadings* ou cargas fatoriais e servem como pesos, mostrando o grau de relacionamento linear entre as variáveis e os fatores ( $w = 1, 2, \dots, p$  e  $j = 1, 2, \dots, k$ );  $\mathbf{F}$  é um vetor de ordem  $k \times 1$  dos fatores comuns não observáveis, ou seja, *a priori*, não são mensuráveis e de algum modo deverão ser identificados; e  $\boldsymbol{\epsilon}$  é um vetor de ordem  $p \times 1$  de erros aleatórios que corresponde aos erros de medição e às variações dos  $Y_s$  que não são explicadas pelos fatores comuns.

O vetor  $\boldsymbol{\epsilon}$  é composto pelos fatores específicos  $\epsilon_1, \epsilon_2, \dots, \epsilon_p$  e  $\mathbf{D}$  é a matriz diagonal  $p \times p$ , cujos elementos da diagonal principal correspondem aos respectivos desvios-padrão das variáveis  $Y_1, Y_2, \dots, Y_p$ .

Analogamente, escreve-se:

$$\begin{bmatrix} 1/\sigma_1 & 0 & \dots & 0 \\ 0 & 1/\sigma_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1/\sigma_p \end{bmatrix} \begin{bmatrix} Y_1 - \mu_1 \\ Y_2 - \mu_2 \\ \vdots \\ Y_p - \mu_p \end{bmatrix} = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \dots & \gamma_{1k} \\ \gamma_{21} & \gamma_{22} & \dots & \gamma_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{p1} & \gamma_{p2} & \dots & \gamma_{pk} \end{bmatrix} \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_k \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_p \end{bmatrix}.$$

Para proceder à estimação do modelo fatorial são necessárias algumas suposições.

São elas:

1.  $E(\mathbf{F}) = E(\boldsymbol{\epsilon}) = \mathbf{0}$ , o que garante que todos os possíveis fatores comuns e específicos possuem médias iguais a 0, em que  $\mathbf{0}$  é o vetor nulo de ordem  $p \times 1$ ;

2.  $Cov(\mathbf{F}) = E(\mathbf{F}\mathbf{F}') = \mathbf{I}_k$ , o que implica que todos os fatores possuem variâncias iguais a 1 e sejam independentes entre si;

3.  $Cov(\boldsymbol{\epsilon}) = E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}') = \boldsymbol{\Psi}_p$  sendo  $\boldsymbol{\Psi}_p = \begin{bmatrix} \psi_1 & 0 & \dots & 0 \\ 0 & \psi_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \psi_p \end{bmatrix}$ , com  $\psi_w > 0$ ,

$\forall w = 1, 2, \dots, p$ , o que indica que os erros são não correlacionados entre si e não necessariamente possuem a mesma variância;

4.  $Cov(\mathbf{F}, \boldsymbol{\epsilon}) = E(\boldsymbol{\epsilon}\mathbf{F}') = \mathbf{0}_{p \times k}$  o que implica a independência entre os fatores comuns e específicos.

Se tais propriedades são verificadas no modelo, ele é denominado de modelo fatorial ortogonal e neste caso, a matriz  $\boldsymbol{\rho}$  pode ser decomposta, como segue:

$$\boldsymbol{\rho} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}' + \boldsymbol{\Psi}.$$

Segundo Ferreira (2011), tem-se:

$$\begin{aligned} \text{Var}(Z_w) &= 1 = \gamma_{1w}^2 + \gamma_{2w}^2 + \dots + \gamma_{kw}^2 + \psi_w = h_w^2 + \psi_w; \\ \text{Cov}(Z_w, Z_{w'}) &= \rho_{ww'} = \gamma_{w1}\gamma_{w'1} + \gamma_{w2}\gamma_{w'2} + \dots + \gamma_{wk}\gamma_{w'k}, \end{aligned}$$

em que:

$$Z_w = \frac{Y_w - \mu_w}{\sigma_w}.$$

A variância de  $Z_w$  pode ser decomposta em duas partes: a primeira,  $h_w^2 = \sum_{j=1}^k \gamma_{wj}^2$ , denomina-se comunalidade e representa a variância da variável  $Z_w$  explicada pelos  $k$  fatores comuns incluídos no modelo; a segunda,  $\psi_w$ , é a parte da variabilidade de  $Z_w$  associada ao erro aleatório e, ou, à variância específica da variável. Portanto, na análise de fatores, a variância de cada variável é explicada pelos fatores comuns e pelo fator específico, enquanto que as correlações são totalmente explicadas pelos fatores comuns.

A matriz  $\Gamma\Gamma'$  tem em sua diagonal principal as comunalidades das  $p$  variáveis e, fora da diagonal principal, as correlações que serão, aproximadamente, os produtos das cargas fatoriais para duas variáveis diferentes ( $\gamma_{wj} \cdot \gamma_{w'j}$ ).

O propósito da análise de fatores é estimar as matrizes  $\Gamma$  e  $\Psi$  que possam representar a matriz  $\rho$ , usando um número  $k$  de fatores, sendo  $k < p$  (MINGOTI, 2005).

Tendo em vista que os parâmetros do modelo fatorial são desconhecidos, o próximo passo será escolher o método estatístico para estimar os fatores.

### 3.3. Método de Estimação dos Componentes Principais

Considere a matriz  $Y$ , de ordem  $n \times p$ , representando  $p$  variáveis  $Y_1, Y_2, \dots, Y_p$  com  $n$  elementos amostrais,  $\mu' = [\mu_1 \mu_2 \dots \mu_p]'$  o vetor de médias das variáveis  $Y_s$  e a matriz de correlações  $\rho$ .

Este método baseia-se na decomposição espectral da matriz  $\rho$ , que garante que toda matriz simétrica pode ser decomposta da seguinte forma:

$$P\Lambda P' = \underbrace{P\Lambda^{1/2}}_{\Gamma} \underbrace{\Lambda^{1/2}P'}_{\Gamma'} = \Gamma\Gamma'.$$

Assim, existem matrizes  $P$  e  $\Lambda$ , tais que:

$$\rho = P\Lambda P' = \underbrace{P\Lambda^{1/2}}_{\Gamma} \underbrace{\Lambda^{1/2}P'}_{\Gamma'} = \Gamma\Gamma',$$

sendo  $\Lambda$  uma matriz diagonal, cujos elementos da diagonal principal são os autovalores de  $\rho$ ,  $P = [e_1 e_2 \dots e_p]$  é a matriz dos autovetores normalizados, em suas colunas, associados a cada autovalor ( $\lambda$ ) e  $\Gamma = P\Lambda^{1/2}$  é a matriz de cargas fatoriais.

Note que  $\mathbf{\Gamma} = \mathbf{P}\mathbf{\Lambda}^{1/2} = [\sqrt{\lambda_1}\mathbf{e}_1\sqrt{\lambda_2}\mathbf{e}_2 \dots \sqrt{\lambda_p}\mathbf{e}_p]$ , sendo que na análise de fatores, os autovalores ( $\lambda$ ) são definidos pela seguinte equação:

$$\lambda_{wj} = \sum_{w=1}^k \gamma_{wj}^2, \text{ para } j = 1, 2, \dots, k,$$

sendo  $\gamma_{wj}$  a carga fatorial da  $w$ -ésima variável padronizada  $Z_w$  e o  $j$ -ésimo fator  $F_j$ , para  $w = 1, 2, \dots, p$  e  $j = 1, 2, \dots, k$ .

Estimam-se as matrizes  $\mathbf{\Gamma}$  e  $\mathbf{\Psi}$  da seguinte forma:

$$\hat{\mathbf{\Gamma}}_{pkk} = \left[ \sqrt{\hat{\lambda}_1}\hat{e}_1 \sqrt{\hat{\lambda}_2}\hat{e}_2 \dots \sqrt{\hat{\lambda}_k}\hat{e}_k \right];$$

$$\hat{\mathbf{\Psi}}_{ppp} = \text{diag}[\hat{\rho} - \hat{\mathbf{\Gamma}}\hat{\mathbf{\Gamma}}'],$$

em que:

$$\hat{\rho} = \mathbf{R}(\text{matriz amostral de correlações}).$$

A matriz  $\hat{\mathbf{\Psi}}_{ppp}$  também pode ser estimada como sendo:

$$\hat{\mathbf{\Psi}}_{ppp} = \mathbf{R} - \hat{\mathbf{\Gamma}}\hat{\mathbf{\Gamma}}',$$

o que não retorna necessariamente uma matriz diagonal.

Obtém-se, portanto, uma aproximação da matriz de correlações teórica  $\rho$ , de tal modo que:  $\mathbf{R} \cong \hat{\mathbf{\Gamma}}\hat{\mathbf{\Gamma}}' + \hat{\mathbf{\Psi}}$ . Isso ocorre pois é desconsiderada a variação aleatória não comum aos fatores e a contribuição dos últimos  $p - k$  autovetores e autovalores de  $\mathbf{R}$ .

Um dos métodos para avaliar a qualidade de ajuste do modelo fatorial é calcular a matriz amostral residual dada por:

$$\mathbf{MRes} = \mathbf{R} - (\hat{\mathbf{\Gamma}}\hat{\mathbf{\Gamma}}' + \hat{\mathbf{\Psi}}).$$

Quanto mais próximo de zero forem os seus valores, mais o modelo  $\hat{\mathbf{\Gamma}}\hat{\mathbf{\Gamma}}' + \hat{\mathbf{\Psi}}$  aproxima-se da matriz  $\mathbf{R}$ , indicando uma boa representatividade (MINGOTI, 2005).

Sabe-se que a variância total pode ser calculada pelo traço da matriz amostral de correlação estimada  $\mathbf{R}$ . Define-se como sendo a proporção da variação total explicada pelo  $k$ -ésimo fator comum como sendo:

$$\frac{\sum_{w=1}^p \hat{\gamma}_{wk}^2}{\text{tr}(\mathbf{R})} = \frac{\sum_{w=1}^p \hat{\gamma}_{wk}^2}{p}.$$

Este resultado pode auxiliar na interpretação e identificação dos fatores mais representativos no modelo.

O método de estimação por componentes principais (CP) recebeu este nome por basear-se na utilização dos autovalores e dos autovetores da matriz de covariâncias para estimar as cargas fatoriais.

Esse método não é invariante às escalas, ou seja, resultados diferentes podem ser obtidos quando a análise de fatores for realizada pelas matrizes  $\Sigma$  ou  $\rho$ .

### 3.4. Método de Rotação Varimax

Para facilitar a interpretação dos fatores e buscar uma estrutura mais simples para a matriz de cargas fatoriais é possível fazer uma rotação ortogonal dos fatores, de modo a se preservar a orientação original entre eles.

As rotações fatoriais têm o objetivo de facilitar a interpretação dos fatores, visto que muitas vezes as variáveis analisadas apresentam cargas fatoriais elevadas em mais de um fator (DAMASIO, 2012)

Existem vários tipos de transformações, como por exemplos as transformações oblíquas e não ortogonais. Entretanto, essas normalmente não produzem soluções independentes, isto é, obtêm-se fatores correlacionados (FERREIRA, 2011).

O objetivo da rotação fatorial é simplificar as linhas e as colunas da matriz de cargas fatoriais, a fim de tornar máximo apenas um fator em cada linha, tentando reduzir ou aproximar o quanto for possível os demais fatores do valor zero.

Segundo Johnson e Wichern (2007), todas as cargas de fator obtidas a partir de uma transformação ortogonal das cargas iniciais têm a mesma capacidade de reproduzir a matriz de covariâncias. A partir da álgebra de matrizes, sabe-se que uma transformação ortogonal corresponde a uma rotação rígida (ou reflexão) dos eixos de coordenadas.

Já Mingoti (2005) diz que a rotação ortogonal não altera o ajuste do modelo fatorial determinado a partir das matrizes  $\Gamma$  e  $\Psi$ , uma vez que a matriz residual, as comunalidades e as variâncias específicas permanecem inalteradas.

Uma matriz quadrada  $T_m$  é dita ser uma matriz ortogonal se, e somente se,  $TT' = T'T = I_{mxm}$ .

A transformação ortogonal se dá a partir da multiplicação de uma matriz ortogonal  $T$  qualquer pela matriz das cargas fatoriais  $\Gamma$ . Assim,

$$\hat{\Gamma}^*_{pxm} = \hat{\Gamma}_{pxm} \hat{T}_{mxm}$$

é a matriz das cargas fatoriais originais transformadas pela rotação ortogonal.

O critério varimax seleciona cargas fatoriais  $\tilde{\gamma}_{wj}$  que maximizam a seguinte expressão:

$$v^* = \frac{1}{p^2} \sum_{j=1}^k \left[ \sum_{w=1}^p \tilde{\gamma}_{wj}^4 - \frac{1}{p} \left( \sum_{w=1}^p \tilde{\gamma}_{wj}^2 \right)^2 \right],$$

em que  $\tilde{\gamma}_{wj} = \frac{\hat{\gamma}_{wj}}{\hat{h}_w}$  sendo  $\hat{\gamma}_{wj}$  a carga fatorial estimada para a  $w$ -ésima variável no  $j$ -ésimo fator e  $\hat{h}_w$  a raiz quadrada da comunalidade associada a variável  $Z_w$ .

A rotação varimax é um dos métodos de rotações ortogonais que produzem fatores que são não correlacionados entre si.

Pallant (2007) afirma que dentre os vários tipos de rotação ortogonal, o método varimax é o mais comumente utilizado, pois procura minimizar o número de variáveis que apresentam altas cargas em cada fator.

Segundo Johnson e Wichern (2007), escalar os coeficientes girados  $\tilde{\gamma}_{wj}$  têm o efeito de dar às variáveis com pequenas comunalidades mais peso na determinação da estrutura simples. Após a transformação  $T'$  ser determinada, as cargas  $\tilde{\gamma}_{wj}$  são multiplicadas por  $\hat{h}_w$ , para que as comunalidades originais sejam preservadas.

Na prática, o critério varimax é o mais utilizado e normalmente produz soluções mais simplificadas que os outros métodos (MINGOTI, 2005).

### 3.5. Viabilidade da Análise de Fatores

O ajuste de um modelo fatorial pressupõe que as variáveis-resposta sejam correlacionadas entre si (MINGOTI, 2005).

A partir do nível de correlação entre as variáveis pode-se definir a viabilidade, ou não, da aplicação da análise de fatores.

Se for concluído que todas as correlações são baixas ou que todas elas são iguais (denotando que nenhuma estrutura existe para agrupar variáveis), então deve-se questionar a utilização da análise de fatores (HAIR et al., 2014).

Se a correlação entre os dados não for significativa, a análise de fatores não será apropriada. Apenas se as variáveis forem suficientemente correlacionadas umas com as outras, será possível identificar fatores representativos.

Para Hair et al. (2014), deve existir um número substancial de correlações entre as variáveis maiores que 0,30 (em módulo) para justificar a aplicação desse método.

Tabachnick e Fidell (2007 *apud* Willians et al., 2010) recomendam que a matriz de correlações seja inspecionada para quantificar coeficientes de correlação acima de 0,30.

Hair et al. (2014) classificam os valores absolutos das correlações como sendo 0,30 mínimo, 0,40 importante e 0,50 significativo. Se nenhuma correlação ultrapassar 0,30, deve-se questionar se a análise de fatores é o método apropriado para ser usado.

Ressalta-se ainda que o modelo fatorial só faz sentido quando a matriz de correlações ( $\rho$ ) difere da matriz identidade. Se os dados possuem distribuição de probabilidades normal p-variada, pode-se realizar um teste para avaliar as correlações dos dados e verificar se a matriz de correlações ( $\rho$ ) se aproxima ou não da matriz identidade. Este teste é chamado de teste de esfericidade de Bartlett, cuja hipótese testada é dada por:

$$H_0: \rho_{p \times p} = I_{p \times p}.$$

Para que o modelo da análise de fatores possa ser bem ajustado, espera-se que a hipótese do teste de esfericidade de Bartlett seja rejeitada.

Outro modo de inspeção, é o critério de Kaiser-Meyer-Olkin (KMO), que é uma medida de adequação da amostragem que compara as correlações simples com as correlações parciais.

Este coeficiente é definido por:

$$KMO = \frac{\sum_{w \neq w'} \rho_{ww'}^2}{\sum_{w \neq w'} \rho_{ww'}^2 + \sum_{w \neq w'} \rho_{ww'|x}^2},$$

em que  $\rho_{ww'}$  é a correlação entre as variáveis aleatórias  $Z_w$  e  $Z_{w'}$ , e  $\rho_{ww'|x}$  é correlação parcial entre as variáveis  $Z_w$  e  $Z_{w'}$  ajustada para as demais variáveis que são representadas pela letra  $x$ .

Os valores do índice KMO variam de 0 a 1, sendo que valores próximos de 0 indicam que a análise de fatores pode não ser adequada, pois entende-se que as correlações entre as variáveis é fraca, enquanto que valores próximos de 1 sugerem que a matriz  $\rho^{-1}$  está próxima de uma matriz diagonal, considerando-se, portanto, mais adequada a utilização dessa técnica.

Uma discussão sobre os valores recomendados para essa medida é apresentada em Pereira (2001), apontando que para valores até 0,50, a adequação dos dados é considerada imprópria para o uso da análise de fatores. Mingoti (2005) também afirma que valores do coeficiente KMO próximos a 0,5 seriam péssimos.

Kaiser e Rice (1974, *apud* Rencher e Christensen, 2012) recomendam que o KMO deve exceder 0,8 para resultados satisfatórios. Por outro lado, Pallant (2007) sugere valores

superiores à 0,6, enquanto que para Hair et al. (2014), valores do KMO acima de 0,50 são aceitáveis.

Já Sharma (1996) aponta que, embora não haja testes estatísticos para o KMO, a análise de fatores pode ser classificada conforme se apresenta na Tabela 3.1

Tabela 3.1 - Classificação da análise de fatores segundo o KMO

KMO	Classificação
$\geq 0,90$	Ótima
0,80 à 0,89	Boa
0,70 à 0,79	Média
0,60 à 0,69	Razoável
0,50 à 0,59	Ruim
$\leq 0,50$	Inaceitável

Fonte: Sharma (1996) e Manso e Simões (2009).

## 4. MATERIAL E MÉTODOS

### 4.1. Modelo Fatorial

O primeiro passo para conduzir uma análise fatorial é estimar a matriz de correlações dos dados (MINGOTI, 2005). Tendo em vista que para simular variáveis aleatórias, a matriz de correlações deve ser informada, optou-se em estabelecer um critério para definir matrizes de correlações de ordem 4 x 4 distintas que caracterizassem diferentes graus de relações entre variáveis.

Para tanto, optou-se por realizar um estudo exploratório supondo-se quatro variáveis aleatórias ( $Y_1, Y_2, Y_3$  e  $Y_4$ ) que seguem distribuição de probabilidades normal multivariada:

$$Y_1, Y_2, Y_3 \text{ e } Y_4 \sim N_4(\boldsymbol{\mu}_Y; \boldsymbol{\Sigma}_Y),$$

em que:  $\boldsymbol{\mu}_Y$  é o vetor aleatório de médias de ordem 4 x 1 e  $\boldsymbol{\Sigma}_Y$  é a matriz de covariâncias de ordem 4 x 4.

Para evitar problemas de escalas, o estudo das variáveis aleatórias  $Y_1, Y_2, Y_3$  e  $Y_4$  foi realizado por meio das variáveis aleatórias padronizadas  $Z_1, Z_2, Z_3$  e  $Z_4$  sendo:

$$Z_w = \frac{Y_w - \mu_w}{\sigma_w}, \text{ para } w = 1, 2, 3 \text{ e } 4.$$

Assim definidas, as quatro variáveis aleatórias padronizadas seguiram distribuição de probabilidades normal multivariada, como segue:

$$Z_1, Z_2, Z_3 \text{ e } Z_4 \sim N_4(\boldsymbol{\mu}_Z = 0; \boldsymbol{\Sigma}_Z),$$

em que:  $\boldsymbol{\mu}_Z$  é o vetor nulo de ordem 4 x 1 e  $\boldsymbol{\Sigma}_Z$  é a matriz de covariâncias entre as variáveis padronizadas ( $Z_s$ ) ou de correlações entre as variáveis originais ( $Y_s$ ) de ordem 4 x 4.

Neste caso, obtém-se que:

$$\boldsymbol{\Sigma}_Z = \boldsymbol{\rho}_Y = \boldsymbol{\rho}.$$

É importante ressaltar que para o modelo fatorial ter utilidade prática, deve-se considerar um número de fatores menor que o número de variáveis. Assim, a fim de se obter um modelo fatorial parcimonioso, construíram-se dois fatores ( $F_1$  e  $F_2$ ), definindo-se, portanto, o modelo fatorial ortogonal, baseado nas variáveis aleatórias padronizadas, como sendo:

$$Z_1 = \gamma_{11}F_1 + \gamma_{21}F_2 + \epsilon_1;$$

$$Z_2 = \gamma_{21}F_1 + \gamma_{22}F_2 + \epsilon_2;$$

$$Z_3 = \gamma_{31}F_1 + \gamma_{32}F_2 + \epsilon_3;$$

$$Z_4 = \gamma_{41}F_1 + \gamma_{42}F_2 + \epsilon_4;$$

em que:  $\gamma_{wj}$  são as cargas fatoriais da  $w$ -ésima variável padronizada  $Z_w$  no  $j$ -ésimo fator  $F_j$ , representando o grau de correlação entre  $Z_w$  e  $F_j$ , e  $\epsilon = [\epsilon_1 \ \epsilon_2 \ \epsilon_3 \ \epsilon_4]'$ , o vetor de erros aleatórios devido à variação de  $Z_w$  que não é explicada pelos fatores comuns  $F_j$ .

De forma matricial, tem-se:

$$\mathbf{Z} = \mathbf{\Gamma}\mathbf{F} + \boldsymbol{\epsilon};$$

$$\begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \end{bmatrix} = \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \\ \gamma_{31} & \gamma_{32} \\ \gamma_{41} & \gamma_{42} \end{bmatrix} \begin{bmatrix} F_1 \\ F_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \end{bmatrix}.$$

## 4.2. Fatores Teóricos

Para as construções dos dois fatores ortogonais, estabeleceram-se seis valores de comunalidades comuns às quatro variáveis (0,49; 0,56; 0,64; 0,72; 0,81 e 0,90), que se referem as correlações de 0,7; 0,75; 0,8; 0,85 e 0,90 entre as variáveis e o fator, e três distâncias proporcionais (0,81; 0,45 e 0,20) entre as cargas fatoriais dos dois fatores, representando distâncias grande, média e pequena, respectivamente, e também comum às quatro variáveis, originando 18 combinações que permitiram definir 18 matrizes teóricas de cargas fatoriais ( $\mathbf{\Gamma}_{4 \times 2}$ ) e de variâncias específicas ( $\mathbf{\Psi}_{4 \times 4}$ ), de modo que:

$$\mathbf{\Gamma}_{4 \times 2} = \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \\ \gamma_{31} & \gamma_{32} \\ \gamma_{41} & \gamma_{42} \end{bmatrix};$$

$$\mathbf{\Psi}_{4 \times 4} = \text{diag} (\mathbf{I}_{4 \times 4} - \mathbf{\Gamma}\mathbf{\Gamma}'_{4 \times 4}),$$

sendo:

$$\mathbf{\Gamma}\mathbf{\Gamma}'_{4 \times 4} = \begin{bmatrix} \gamma_{11}^2 + \gamma_{22}^2 & \gamma_{11}\gamma_{21} + \gamma_{12}\gamma_{22} & \gamma_{11}\gamma_{31} + \gamma_{12}\gamma_{32} & \gamma_{11}\gamma_{41} + \gamma_{12}\gamma_{42} \\ \gamma_{21}\gamma_{11} + \gamma_{22}\gamma_{12} & \gamma_{21}^2 + \gamma_{22}^2 & \gamma_{21}\gamma_{31} + \gamma_{22}\gamma_{32} & \gamma_{21}\gamma_{41} + \gamma_{22}\gamma_{42} \\ \gamma_{31}\gamma_{11} + \gamma_{32}\gamma_{12} & \gamma_{31}\gamma_{21} + \gamma_{32}\gamma_{22} & \gamma_{31}^2 + \gamma_{32}^2 & \gamma_{31}\gamma_{41} + \gamma_{32}\gamma_{42} \\ \gamma_{41}\gamma_{11} + \gamma_{42}\gamma_{12} & \gamma_{41}\gamma_{21} + \gamma_{42}\gamma_{22} & \gamma_{41}\gamma_{31} + \gamma_{42}\gamma_{32} & \gamma_{41}^2 + \gamma_{42}^2 \end{bmatrix}.$$

A distância proporcional entre as cargas fatoriais dos fatores  $F_1$  e  $F_2$ , comum às quatro variáveis em cada uma das 18 combinações, foi definida com o objetivo de estabelecer diferentes graus de separação entre as quatro variáveis nos dois fatores.

Além disso, foi possível definir, pelas magnitudes das cargas fatoriais, a representação das variáveis  $Z_1$  e  $Z_2$  por  $F_1$ , e  $Z_3$  e  $Z_4$  por  $F_2$ , fixando-se:

$$\gamma_{11} = \gamma_{21} = \gamma_{32} = \gamma_{42};$$

$$\gamma_{12} = \gamma_{22} = \gamma_{31} = \gamma_{41},$$

e permitindo-se analisar, *a posteriori*, o quão próximos foram as estimativas dos fatores teóricos assim determinados.

Para que as combinações fossem satisfeitas, algumas relações foram estabelecidas:

$$\gamma_{w1}^2 + \gamma_{w2}^2 = h_w^2, \text{ para } w = 1, 2, 3 \text{ e } 4;$$

$$d_w = \frac{|\gamma_{w1} - \gamma_{w2}|}{\sqrt{h_w^2}}, \text{ para } w = 1, 2, 3 \text{ e } 4,$$

sendo:  $\gamma_{wj}$  a carga fatorial da variável  $Z_w$  no fator  $F_j$ , para  $j = 1$  e  $2$ ,  $h_w^2$  a comunalidade da variável  $Z_w$  e  $d_w$  a distância proporcional entre as cargas fatoriais  $\gamma_{wj}$  e  $\gamma_{w1j}$ .

Assim sendo, obtiveram-se as 18 matrizes de cargas fatoriais e as 18 matrizes de variâncias específicas conforme apresentado nas Tabelas 4.1, 4.2, 4.3, 4.4, 4.5 e 4.6.

Tabela 4.1 - Matrizes teóricas  $\Gamma$  e  $\Psi$  para  $h_w^2 = 0,49$  ( $w = 1, 2, 3$  e  $4$ )

Matriz	$d_w = 0,81$	$d_w = 0,45$	$d_w = 0,20$
$\Gamma$	$\begin{bmatrix} 0,69 & 0,12 \\ 0,69 & 0,12 \\ 0,12 & 0,69 \\ 0,12 & 0,69 \end{bmatrix}$	$\begin{bmatrix} 0,628 & 0,31 \\ 0,628 & 0,31 \\ 0,31 & 0,628 \\ 0,31 & 0,628 \end{bmatrix}$	$\begin{bmatrix} 0,56 & 0,42 \\ 0,56 & 0,42 \\ 0,42 & 0,56 \\ 0,42 & 0,56 \end{bmatrix}$
$\Psi$	$\begin{bmatrix} 0,51 & 0 & 0 & 0 \\ 0 & 0,51 & 0 & 0 \\ 0 & 0 & 0,51 & 0 \\ 0 & 0 & 0 & 0,51 \end{bmatrix}$	$\begin{bmatrix} 0,51 & 0 & 0 & 0 \\ 0 & 0,51 & 0 & 0 \\ 0 & 0 & 0,51 & 0 \\ 0 & 0 & 0 & 0,51 \end{bmatrix}$	$\begin{bmatrix} 0,51 & 0 & 0 & 0 \\ 0 & 0,51 & 0 & 0 \\ 0 & 0 & 0,51 & 0 \\ 0 & 0 & 0 & 0,51 \end{bmatrix}$

Tabela 4.2 - Matrizes teóricas  $\Gamma$  e  $\Psi$  para  $h_w^2 = 0,56$  ( $w = 1, 2, 3$  e  $4$ )

Matriz	$d_w = 0,81$	$d_w = 0,45$	$d_w = 0,20$
$\Gamma$	$\begin{bmatrix} 0,738 & 0,134 \\ 0,738 & 0,134 \\ 0,134 & 0,738 \\ 0,134 & 0,738 \end{bmatrix}$	$\begin{bmatrix} 0,672 & 0,333 \\ 0,672 & 0,333 \\ 0,333 & 0,672 \\ 0,333 & 0,672 \end{bmatrix}$	$\begin{bmatrix} 0,6 & 0,45 \\ 0,6 & 0,45 \\ 0,45 & 0,6 \\ 0,45 & 0,6 \end{bmatrix}$
$\Psi$	$\begin{bmatrix} 0,44 & 0 & 0 & 0 \\ 0 & 0,44 & 0 & 0 \\ 0 & 0 & 0,44 & 0 \\ 0 & 0 & 0 & 0,44 \end{bmatrix}$	$\begin{bmatrix} 0,44 & 0 & 0 & 0 \\ 0 & 0,44 & 0 & 0 \\ 0 & 0 & 0,44 & 0 \\ 0 & 0 & 0 & 0,44 \end{bmatrix}$	$\begin{bmatrix} 0,44 & 0 & 0 & 0 \\ 0 & 0,44 & 0 & 0 \\ 0 & 0 & 0,44 & 0 \\ 0 & 0 & 0 & 0,44 \end{bmatrix}$

Tabela 4.3 - Matrizes teóricas  $\Gamma$  e  $\Psi$  para  $h_w^2 = 0,64$  ( $w = 1, 2, 3$  e  $4$ )

Matriz	$d_w = 0,81$	$d_w = 0,45$	$d_w = 0,20$
$\Gamma$	$\begin{bmatrix} 0,788 & 0,138 \\ 0,788 & 0,138 \\ 0,138 & 0,788 \\ 0,138 & 0,788 \end{bmatrix}$	$\begin{bmatrix} 0,715 & 0,359 \\ 0,715 & 0,359 \\ 0,359 & 0,715 \\ 0,359 & 0,715 \end{bmatrix}$	$\begin{bmatrix} 0,64 & 0,48 \\ 0,64 & 0,48 \\ 0,48 & 0,64 \\ 0,48 & 0,64 \end{bmatrix}$
$\Psi$	$\begin{bmatrix} 0,36 & 0 & 0 & 0 \\ 0 & 0,36 & 0 & 0 \\ 0 & 0 & 0,36 & 0 \\ 0 & 0 & 0 & 0,36 \end{bmatrix}$	$\begin{bmatrix} 0,36 & 0 & 0 & 0 \\ 0 & 0,36 & 0 & 0 \\ 0 & 0 & 0,36 & 0 \\ 0 & 0 & 0 & 0,36 \end{bmatrix}$	$\begin{bmatrix} 0,36 & 0 & 0 & 0 \\ 0 & 0,36 & 0 & 0 \\ 0 & 0 & 0,36 & 0 \\ 0 & 0 & 0 & 0,36 \end{bmatrix}$

Tabela 4.4 - Matrizes teóricas  $\Gamma$  e  $\Psi$  para  $h_w^2 = 0,72$  ( $w = 1, 2, 3$  e  $4$ )

Matriz	$d_w = 0,81$	$d_w = 0,45$	$d_w = 0,20$
$\Gamma$	$\begin{bmatrix} 0,837 & 0,148 \\ 0,837 & 0,148 \\ 0,148 & 0,837 \\ 0,148 & 0,837 \end{bmatrix}$	$\begin{bmatrix} 0,76 & 0,38 \\ 0,76 & 0,38 \\ 0,38 & 0,76 \\ 0,38 & 0,76 \end{bmatrix}$	$\begin{bmatrix} 0,68 & 0,51 \\ 0,68 & 0,51 \\ 0,51 & 0,68 \\ 0,51 & 0,68 \end{bmatrix}$
$\Psi$	$\begin{bmatrix} 0,28 & 0 & 0 & 0 \\ 0 & 0,28 & 0 & 0 \\ 0 & 0 & 0,28 & 0 \\ 0 & 0 & 0 & 0,28 \end{bmatrix}$	$\begin{bmatrix} 0,28 & 0 & 0 & 0 \\ 0 & 0,28 & 0 & 0 \\ 0 & 0 & 0,28 & 0 \\ 0 & 0 & 0 & 0,28 \end{bmatrix}$	$\begin{bmatrix} 0,28 & 0 & 0 & 0 \\ 0 & 0,28 & 0 & 0 \\ 0 & 0 & 0,28 & 0 \\ 0 & 0 & 0 & 0,28 \end{bmatrix}$

Tabela 4.5 - Matrizes teóricas  $\Gamma$  e  $\Psi$  para  $h_w^2 = 0,81$  ( $w = 1, 2, 3$  e  $4$ )

Matriz	$d_w = 0,81$	$d_w = 0,45$	$d_w = 0,20$
$\Gamma$	$\begin{bmatrix} 0,886 & 0,158 \\ 0,886 & 0,158 \\ 0,158 & 0,886 \\ 0,158 & 0,886 \end{bmatrix}$	$\begin{bmatrix} 0,805 & 0,402 \\ 0,805 & 0,402 \\ 0,402 & 0,805 \\ 0,402 & 0,805 \end{bmatrix}$	$\begin{bmatrix} 0,72 & 0,54 \\ 0,72 & 0,54 \\ 0,54 & 0,72 \\ 0,54 & 0,72 \end{bmatrix}$
$\Psi$	$\begin{bmatrix} 0,19 & 0 & 0 & 0 \\ 0 & 0,19 & 0 & 0 \\ 0 & 0 & 0,19 & 0 \\ 0 & 0 & 0 & 0,19 \end{bmatrix}$	$\begin{bmatrix} 0,19 & 0 & 0 & 0 \\ 0 & 0,19 & 0 & 0 \\ 0 & 0 & 0,19 & 0 \\ 0 & 0 & 0 & 0,19 \end{bmatrix}$	$\begin{bmatrix} 0,19 & 0 & 0 & 0 \\ 0 & 0,19 & 0 & 0 \\ 0 & 0 & 0,19 & 0 \\ 0 & 0 & 0 & 0,19 \end{bmatrix}$

Tabela 4.6 - Matrizes teóricas  $\Gamma$  e  $\Psi$  para  $h^2 = 0,90$  ( $w = 1, 2, 3$  e  $4$ )

Matriz	$d_w = 0,81$	$d_w = 0,45$	$d_w = 0,20$
$\Gamma$	$\begin{bmatrix} 0,935 & 0,168 \\ 0,935 & 0,168 \\ 0,168 & 0,935 \\ 0,168 & 0,935 \end{bmatrix}$	$\begin{bmatrix} 0,852 & 0,42 \\ 0,852 & 0,42 \\ 0,42 & 0,852 \\ 0,42 & 0,852 \end{bmatrix}$	$\begin{bmatrix} 0,76 & 0,57 \\ 0,76 & 0,57 \\ 0,57 & 0,76 \\ 0,57 & 0,76 \end{bmatrix}$
$\Psi$	$\begin{bmatrix} 0,1 & 0 & 0 & 0 \\ 0 & 0,1 & 0 & 0 \\ 0 & 0 & 0,1 & 0 \\ 0 & 0 & 0 & 0,1 \end{bmatrix}$	$\begin{bmatrix} 0,1 & 0 & 0 & 0 \\ 0 & 0,1 & 0 & 0 \\ 0 & 0 & 0,1 & 0 \\ 0 & 0 & 0 & 0,1 \end{bmatrix}$	$\begin{bmatrix} 0,1 & 0 & 0 & 0 \\ 0 & 0,1 & 0 & 0 \\ 0 & 0 & 0,1 & 0 \\ 0 & 0 & 0 & 0,1 \end{bmatrix}$

Foram calculados os autovalores teóricos de acordo com a seguinte fórmula:  $\lambda_j = \sum_{w=1}^4 \gamma_{wj}^2$ , para  $j = 1$  e  $2$ , conforme apresentado na Tabela 4.7.

Tabela 4.7 - Autovalores ( $\lambda_j$ ) obtidos para os fatores  $F_1$  e  $F_2$ , independentemente de  $d_w$ , para  $h_w^2$  ( $w = 1, 2, 3$  e  $4$ )

$h_w^2$	$d_w = 0,81; 0,45; \text{ e } 0,20$	
	$F_1$	$F_2$
0,49	0,98	0,98
0,56	1,13	1,13
0,64	1,28	1,28
0,72	1,45	1,45
0,81	1,62	1,62
0,90	1,81	1,81

O critério de Kaiser para os autovalores, é baseado na comparação do valor numérico de cada autovalor ( $\lambda_j$ ) com o valor 1, e determina que a quantidade de fatores a ser utilizada na análise de fatores deve ser igual ao número de autovalores maiores que 1.

Esse critério substancia-se no fato de que cada fator tem um autovalor, que representa a quantidade de variância explicada por ele, de modo que se um autovalor é menor do que 1, significa que o fator representa menos variância do que uma única variável. Sendo um dos objetivos da análise de fatores reduzir o conjunto de variáveis, fatores com autovalores menores que 1 não parecem servir a esse propósito. Assim, apenas os fatores com autovalores maiores devem ser retidos pela análise (FLOYD; WIDAMAN, 1995).

Nota-se que, com exceções dos autovalores obtidos para as matrizes com comunalidades iguais a 0,49, todos os outros foram maiores que 1. Porém, como neste caso, eles são ligeiramente abaixo do valor estabelecido pelo critério de Kaiser, optou-se por mantê-los.

Como a variância total das variáveis aleatórias padronizadas  $Z_1, Z_2, Z_3$  e  $Z_4$  é igual a quatro e com base no critério de Kaiser, concluiu-se que os fatores assim estabelecidos foram representativos e proporcionaram a parcimônia necessária para a análise, bem como descreveram com qualidade a estrutura de variabilidade dos  $Z_s$ .

### 4.3. Matrizes de Correlações

Definindo-se a padronização das variáveis aleatórias  $Y_s$  centradas em zero e com variâncias iguais a 1, é sabido que a matriz de covariâncias ( $\Sigma_z$ ) das variáveis aleatórias padronizadas  $Z_s$  corresponde à matriz de correlações ( $\rho$ ) dos  $Y_s$ .

Pela teoria e baseado no modelo fatorial ortogonal com dois fatores, a matriz de correlações ( $\rho$ ) pode ser decomposta da seguinte forma (MINGOTI, 2005):

$$\Gamma\Gamma'_{4 \times 4} + \Psi_{4 \times 4} = \rho_{4 \times 4},$$

sendo  $\Gamma_{4 \times 2}$  a matriz de cargas fatoriais obtida a partir da decomposição de  $\rho$ , e  $\Psi_{4 \times 4}$  a matriz de variâncias específicas das variáveis aleatórias padronizadas  $Z_s$ .

A partir dos resultados obtidos para as matrizes  $\Gamma\Gamma'_{4 \times 4}$  e  $\Psi_{4 \times 4}$  de maneira teórica, em que foram utilizadas 18 combinações entre as seis comunalidades ( $h_w^2$ ) e as três distâncias das cargas fatoriais dos dois fatores inerentes a cada variável aleatória padronizada, foram obtidas as matrizes de correlações teóricas ( $\rho$ ), conforme apresentados nas Tabelas 4.8, 4.9, 4.10, 4.11, 4.12 e 4.13.

Tabela 4.8 - Matrizes de correlações ( $\rho$ ) representadas pelos elementos da diagonal e acima dela para  $h_w^2 = 0,49$  ( $w = 1, 2, 3$  e  $4$ )

$d_w = 0,81$				$d_w = 0,45$				$d_w = 0,20$			
1	0,49	0,17	0,17	1	0,49	0,39	0,39	1	0,49	0,47	0,47
	1	0,17	0,17		1	0,39	0,39		1	0,47	0,47
		1	0,49			1	0,49			1	0,49
			1				1				1

Tabela 4.9 - Matrizes de correlações ( $\rho$ ) representadas pelos elementos da diagonal e acima dela para  $h_w^2 = 0,56$  ( $w = 1, 2, 3$  e  $4$ )

$d_w = 0,81$				$d_w = 0,45$				$d_w = 0,20$			
1	0,56	0,20	0,20	1	0,56	0,45	0,45	1	0,56	0,54	0,54
	1	0,20	0,20		1	0,45	0,45		1	0,54	0,54
		1	0,56			1	0,56			1	0,56
			1				1				1

Tabela 4.10 - Matrizes de correlações ( $\rho$ ) representadas pelos elementos da diagonal e acima dela para  $h_w^2 = 0,64$  ( $w = 1, 2, 3$  e  $4$ )

$d_w = 0,81$				$d_w = 0,45$				$d_w = 0,20$			
1	0,64	0,22	0,22	1	0,64	0,51	0,51	1	0,64	0,61	0,61
	1	0,22	0,22		1	0,51	0,51		1	0,61	0,61
		1	0,64			1	0,64			1	0,64
			1				1				1

Tabela 4.11 - Matrizes de correlações ( $\rho$ ) representadas pelos elementos da diagonal e acima dela para  $h_w^2 = 0,72$  ( $w = 1, 2, 3$  e  $4$ )

$d_w = 0,81$				$d_w = 0,45$				$d_w = 0,20$			
1	0,72	0,25	0,25	1	0,72	0,58	0,58	1	0,72	0,70	0,70
	1	0,25	0,25		1	0,58	0,58		1	0,70	0,70
		1	0,72			1	0,72			1	0,72
			1				1				1

Tabela 4.12 - Matrizes de correlações ( $\rho$ ) representadas pelos elementos da diagonal e acima dela para  $h_w^2 = 0,81$  ( $w = 1, 2, 3$  e  $4$ )

$d_w = 0,81$				$d_w = 0,45$				$d_w = 0,20$			
1	0,81	0,28	0,28	1	0,81	0,65	0,65	1	0,81	0,78	0,78
	1	0,28	0,28		1	0,65	0,65		1	0,78	0,78
		1	0,81			1	0,81			1	0,81
			1				1				1

Tabela 4.13 - Matrizes de correlações ( $\rho$ ) representadas pelos elementos da diagonal e acima dela para  $h_w^2 = 0,90$  ( $w = 1, 2, 3$  e  $4$ )

$d_w = 0,81$				$d_w = 0,45$				$d_w = 0,20$			
1	0,90	0,31	0,31	1	0,90	0,72	0,72	1	0,90	0,87	0,87
	1	0,31	0,31		1	0,72	0,72		1	0,87	0,87
		1	0,90			1	0,90			1	0,90
			1				1				1

Em uma análise quantitativa, notou-se que apenas quando a distância proporcional entre cargas fatoriais é 0,81, existem um número substancial (50%) de correlações entre as variáveis com valores absolutos próximos à 0,30, o que segundo Hair et al. (2014) indica uma não adequação para a aplicação da análise de fatores.

Com base na suposição de normalidade das variáveis aleatórias estudadas, pode-se realizar um teste para verificar se as 18 matrizes de correlações ( $\rho$ ) acima relacionadas são próximas ou não da matriz identidade. Ele é chamado de teste de esfericidade de Bartlett, cuja hipótese testada nesse estudo foi dada por:

$$H_0: \rho_{4 \times 4} = I_{4 \times 4}.$$

O nível de significância escolhido para o teste foi de 5%, de tal modo que se  $p$  – valor  $< 0,05$ , a hipótese nula foi rejeitada e concluiu-se que a matriz  $\rho$  não é uma matriz identidade. Como as variáveis aleatórias seguem distribuição normal, pode-se concluir também que as correlações não são iguais a zero e, conseqüentemente, elas não são mutuamente independentes entre si.

Quando a hipótese  $H_0$  do teste de esfericidade de Bartlett é rejeitada, infere-se que a matriz de correlações é adequada e justifica a aplicação da análise de fatores.

Outro critério utilizado para verificar a viabilidade do uso da análise de fatores foi o de Kaiser-Meyer-Olkin (KMO), cujo coeficiente é dado por:

$$KMO = \frac{\sum_{w \neq w'} \rho_{ww'}^2}{\sum_{w \neq w'} \rho_{ww'}^2 + \sum_{w \neq w'} \rho_{ww'}^2 x'}$$

em que  $\rho_{ww'}$  é a correlação entre as variáveis aleatórias  $Z_w$  e  $Z_{w'}$ , e  $\rho_{ww'} x'$  é correlação parcial entre as mesmas variáveis  $Z_w$  e  $Z_{w'}$  ajustada para as demais.

Os valores da estatística KMO podem variar de 0 a 1, sendo que quanto mais próximos de 1, maior será a proximidade da matriz  $\rho^{-1}$  de uma matriz diagonal, indicando mais adequada a utilização da análise de fatores.

De modo geral, os resultados desses dois critérios coincidem com relação a recomendação ou não da aplicação da análise de fatores (DZIUBAN; SHIRKEY, 1974 *apud* DAMÁSIO, 2012).

Assim, iniciou-se o presente estudo analisando quantitativamente as matrizes e avaliando a adequabilidade dos dados, a partir do teste de esfericidade de Bartlett e do critério de KMO em cada uma das 18 matrizes de correlações.

Tais análises foram viabilizadas pelo pacote *psych* do software R, versão 3.3.1, por meio das seguintes funções:

*KMO(cor)*

*cortest.bartlett(cor)*

sendo *cor* a matriz de correlações ( $\rho$ ) das variáveis  $Y_g$ .

Em seguida, para verificar o efeito da comunalidade e da distância proporcional entre as cargas fatoriais sobre o valor calculado e o *p - valor* do teste de esfericidade de Bartlett e para o KMO, foi estimada a superfície de resposta para cada uma dessas variáveis, conforme o modelo:

$$y = \beta_0 + \beta_1 h + \beta_2 d + \beta_3 hd + \varepsilon,$$

em que:

*h* = comunalidade (0,49; 0,56; 0,64; 0,72; 0,81; e 0,90);

*d* = diferença entre as cargas fatoriais em proporção (0,81; 0,45; e 0,20);

$\varepsilon \sim N(0, \sigma^2)$ .

## 4.4. Método dos Componentes Principais

### 4.4.1. Sem Rotação

De acordo com modelo fatorial ortogonal tem-se que:

$$\boldsymbol{\rho} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}' + \boldsymbol{\Psi},$$

em que  $\boldsymbol{\Gamma}$  e  $\boldsymbol{\Psi}$  são as matrizes de cargas fatoriais e das variâncias específicas do modelo fatorial, respectivamente.

Posto isso, aplicou-se o método dos componentes principais para realizar a análise de fatores, fixando-se a estimação de apenas dois fatores comuns em cada matriz de correlações, com o objetivo de avaliar a qualidade de estimação das matrizes  $\boldsymbol{\Gamma}_{CP}$  e  $\boldsymbol{\Psi}_{CP}$  obtidas por esse método baseado na matriz  $\boldsymbol{\rho}$ , em que:

$$\boldsymbol{\rho} \cong \boldsymbol{\Gamma}_{CP}\boldsymbol{\Gamma}'_{CP} + \boldsymbol{\Psi}_{CP}.$$

Da teoria do método dos componentes principais, a matriz  $\boldsymbol{\Gamma}_{CP}$  é obtida por meio da decomposição espectral da matriz de correlações ( $\boldsymbol{\rho}$ ) que é especificada em termos dos pares de autovalores e autovetores  $(\lambda_{1CP}, \mathbf{e}_{1CP})$ ,  $(\lambda_{2CP}, \mathbf{e}_{2CP})$ , sendo  $\lambda_{1CP} \geq \lambda_{2CP}$ , no qual obteve-se as matrizes  $\boldsymbol{\Gamma}_{CP}$  e  $\boldsymbol{\Psi}_{CP}$  por:

$$\begin{aligned}\boldsymbol{\Gamma}_{4 \times 2_{CP}} &= \mathbf{P}_{4 \times 2} \boldsymbol{\Lambda}_{2 \times 2}^{1/2}; \\ \boldsymbol{\Psi}_{4 \times 4_{CP}} &= \text{diag}(\mathbf{I}_{4 \times 4} - \boldsymbol{\Gamma}_{CP}\boldsymbol{\Gamma}'_{CP}).\end{aligned}$$

De outro modo,  $\boldsymbol{\Gamma}_{CP}$  e  $\boldsymbol{\Psi}_{CP}$  também podem ser definidas como segue:

$$\begin{aligned}\boldsymbol{\Gamma}_{CP} &= \left[ \sqrt{\lambda_{1CP}} \mathbf{e}_{1CP} \quad \sqrt{\lambda_{2CP}} \mathbf{e}_{2CP} \right]; \\ \boldsymbol{\Psi}_{CP} &= \begin{bmatrix} \psi_{1CP} & 0 & 0 & 0 \\ 0 & \psi_{2CP} & 0 & 0 \\ 0 & 0 & \psi_{3CP} & 0 \\ 0 & 0 & 0 & \psi_{4CP} \end{bmatrix},\end{aligned}$$

em que:

$\lambda_{1CP}$  = primeiro autovalor da matriz  $\boldsymbol{\rho}$  obtido pelo método dos componentes principais;

$\lambda_{2CP}$  = segundo autovalor da matriz  $\boldsymbol{\rho}$  obtido pelo método dos componentes principais;

$\mathbf{e}_{1CP}$  = autovetor normalizado associado ao autovalor  $\lambda_{CP1}$ ;

$\mathbf{e}_{2CP}$  = autovetor normalizado associado ao autovalor  $\lambda_{CP2}$ ;

$\psi_{jCP} = 1 - \sum_{j=1}^2 \gamma_{CPj}^2$ ;

$\sum_{j=1}^2 \gamma_{wjCP}^2 = \gamma_{w1CP}^2 + \gamma_{w2CP}^2 = h_{wCP}^2$ , para  $w = 1, 2, 3$  e  $4$   $j = 1$  e  $2$ .

Logo, obteve-se o modelo fatorial ortogonal obtido pelo método dos componentes principais definido por:

$$\begin{aligned} Z_1 &= \gamma_{11_{CP}}F_{1_{CP}} + \gamma_{21_{CP}}F_{2_{CP}} + \epsilon_{1_{CP}}; \\ Z_2 &= \gamma_{21_{CP}}F_{1_{CP}} + \gamma_{22_{CP}}F_{2_{CP}} + \epsilon_{2_{CP}}; \\ Z_3 &= \gamma_{31_{CP}}F_{1_{CP}} + \gamma_{32_{CP}}F_{2_{CP}} + \epsilon_{3_{CP}}; \\ Z_4 &= \gamma_{41_{CP}}F_{1_{CP}} + \gamma_{42_{CP}}F_{2_{CP}} + \epsilon_{4_{CP}}, \end{aligned}$$

em que:  $\gamma_{wj_{CP}}$  são as cargas fatoriais obtidas pelo método dos componentes principais sem rotação da  $w$ -ésima variável padronizada  $Z_w$  no  $j$ -ésimo fator  $F_{j_{CP}}$  e  $\epsilon_{CP} = [\epsilon_{1_{CP}} \ \epsilon_{2_{CP}} \ \epsilon_{3_{CP}} \ \epsilon_{4_{CP}}]'$  é o vetor de erros aleatórios devido à variação de  $Z_w$  que não é explicada pelos fatores comuns  $F_{j_{CP}}$ .

De forma matricial, tem-se:

$$\mathbf{Z} = \mathbf{\Gamma}_{CP}\mathbf{F}_{CP} + \epsilon_{CP};$$

$$\begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \end{bmatrix} = \begin{bmatrix} \gamma_{11_{CP}} & \gamma_{12_{CP}} \\ \gamma_{21_{CP}} & \gamma_{22_{CP}} \\ \gamma_{31_{CP}} & \gamma_{32_{CP}} \\ \gamma_{41_{CP}} & \gamma_{42_{CP}} \end{bmatrix} \begin{bmatrix} F_{1_{CP}} \\ F_{2_{CP}} \end{bmatrix} + \begin{bmatrix} \epsilon_{1_{CP}} \\ \epsilon_{2_{CP}} \\ \epsilon_{3_{CP}} \\ \epsilon_{4_{CP}} \end{bmatrix}.$$

Foram obtidas as matrizes de cargas fatoriais ( $\mathbf{\Gamma}_{CP}$ ) e das variâncias específicas ( $\mathbf{\Psi}_{CP}$ ) para cada uma das 18 matrizes de correlações estabelecidas de acordo com as combinações entre as seis comunalidades (0,49; 0,56; 0,64; 0,72; 0,81 e 0,90) e as três distâncias proporcionais (0,81; 0,45 e 0,20) das cargas fatoriais.

Foram estimadas as comunalidades ( $h_{w_{CP}}^2$ ) e os autovalores  $\lambda_{j_{CP}}$  em cada uma das 18 matrizes de correlações, a fim de verificar quão próximas são as estimativas dos parâmetros.

Segundo Johnson e Wichern (2007), os elementos da diagonal da matriz  $\mathbf{\rho}_{CP}$  são iguais aos elementos da diagonal da matriz  $\mathbf{\Gamma}_{CP}\mathbf{\Gamma}'_{CP} + \mathbf{\Psi}_{CP}$ . Porém, os elementos fora da diagonal da matriz  $\mathbf{\rho}$  não são reproduzidos por  $\mathbf{\Gamma}_{CP}\mathbf{\Gamma}'_{CP} + \mathbf{\Psi}_{CP}$ , indicando uma fragilidade do método dos componentes principais.

Isso significa que os parâmetros do modelo fatorial ortogonal, caso não sejam iguais, após a aplicação do método dos componentes principais aplicado à matriz de correlações  $\mathbf{\rho}$ , será devido único e exclusivamente ao próprio método.

Portanto, adotou-se a nomenclatura com o índice  $CP$  aos parâmetros do modelo obtido pelo método dos componentes principais, dado que ele foi aplicado a uma matriz paramétrica e não a uma matriz estimada  $\mathbf{R}$ .

Considerando-se a matriz residual  $MRes_{CP} = \rho - (\Gamma_{CP}\Gamma'_{CP} + \Psi_{CP})$ , como sendo o resultado da aproximação da matriz  $\rho$  pela solução do método de estimação dos componentes principais ( $\Gamma_{CP}\Gamma'_{CP} + \Psi_{CP}$ ), em cada uma das 18 análises, foi estimada a soma de quadrados dos elementos da matriz residual  $MRes_{CP}$ , definida por:

$$sqres_{CP} = tr(MRes_{CP}^2).$$

Ao total, foram realizadas 18 análises de fatores pelo método dos componentes principais, sem rotação, no software R versão 3.3.1 relacionadas às 18 matrizes  $\rho$ , cujos resultados apresentados, utilizando-se a estimação pelo método dos componentes principais, foram obtidos de acordo com o seguinte *script*:

```
library(openxlsx)
dados = read.xlsx("cor teorica.xlsx")
attach(dados)
cor = cbind(Z1,Z2,Z3,Z4)
library(psych)
analise.cp = principal(cor,nfactors=2,rotate="none")
comu = analise.cp$communality
a = analise.cp$loadings
at=t(a)
gg=a%*%at
varexp = diag(1-comu)
corest = gg+varexp
autovalor = analise.cp$values
e = cor-gg-varexp
e2 = e%*%e
sqres = tr(e2)
sqres
```

#### 4.4.2. Com Rotação

Para cada uma das 18 matrizes de correlações, foi realizada a análise de fatores utilizando-se os métodos de estimação dos componentes principais e de rotação varimax, a fim de avaliar qual o efeito da rotação na estimação das cargas fatoriais e qual a melhoria gerada para a interpretação dos fatores, ou seja, verificar se a rotação favoreceu o agrupamento das variáveis nos fatores.

Dessa forma, obteve-se novamente as matrizes das cargas fatoriais ( $\Gamma_{CPR}$ ), das variâncias específicas ( $\Psi_{CPR}$ ), os autovalores ( $\lambda_{CPR}$ ), da comunalidade ( $h_{CPR}^2$ ), dos resíduos ( $MRes_{CPR}$ ) e a soma de quadrados dos elementos da matriz de resíduos ( $sqres$ ).

De acordo com Pereira (2001), a rotação da matriz de cargas fatoriais não afeta a comunalidade das variáveis, nem a proporção da variância explicada pelos fatores.

Como esperado, observou-se que para as estimativas de comunalidade, conseqüentemente, da matriz de variâncias específicas, da matriz de resíduos e da soma de quadrados dos elementos da matriz de resíduos não houve alteração com relação às análises de fatores procedidas com e sem a rotação varimax, sendo iguais em ambas as análises para todas as 18 matrizes de correlações.

Portanto, segue que  $\gamma_{wj_{CP}} \neq \gamma_{wj_{CPR}}$ , porém  $h_{w_{CP}}^2 = h_{w_{CPR}}^2$ ,  $\Psi_{CP} = \Psi_{CPR}$ ,  $\lambda_{j_{CP}} = \lambda_{j_{CPR}}$  e  $MRes_{CP} = MRes_{CPR}$ .

No total, foram realizadas 36 análises de fatores (com e sem rotação), sendo cada análise com o método de estimação dos componentes principais e com a rotação varimax viabilizada pelo seguinte *script* no software R versão 3.3.1:

```
library(openxlsx)
dados = read.xlsx("cor teorica.xlsx")
attach(dados)
cor = bind(Z1,Z2,Z3,Z4)
library(psych)
analise = principal(cor,nfactors=2,rotate="varimax")
comu = analise$communality
a = analise$loadings
at = t(a)
gg = a%*%at
varexp = diag(1-comu)
corest = gg+varexp
autovalor = analise$values
e = cor-gg-varexp
e2 = e%*%e
sqres = tr(e2)
sqres
```

#### 4.5. Medidas de Avaliação

Definiram-se como variáveis explicativas os autovalores ( $\lambda_j$ ,  $j = 1$  e  $2$ ), a matriz de cargas fatoriais ( $\Gamma$ ), a comunalidade ( $h_w^2$ ), a diferença entre as cargas dos fatores  $F_1$  e  $F_2$  para cada variável  $Z_w$ ,  $w = 1, 2, 3$  e  $4$  e a soma de quadrados dos resíduos (*sqres*).

Para cada uma das 36 análises realizadas, foram estimados os erros relativos médios dos desvios entre os parâmetros e as suas respectivas estimativas, como segue:

$$\Delta_\lambda = \frac{1}{2} \left( \sum_{j=1}^2 \frac{|\lambda_j - \lambda_{j_{CP}}|}{\lambda_j} \right);$$

$$\Delta_{\lambda_1} = \frac{|\lambda_1 - \lambda_{1CP}|}{\lambda_1} \text{ e } \Delta_{\lambda_2} = \frac{|\lambda_2 - \lambda_{2CP}|}{\lambda_2};$$

$$\Delta_{\Gamma} = \frac{1}{8} \left( \sum_{w=1}^4 \sum_{j=1}^2 \frac{|\gamma_{wj} - \gamma_{wjCP}|}{\gamma_{wj}} \right);$$

$$\Delta_{h^2} = \frac{1}{4} \left( \sum_{w=1}^4 \frac{|h_w^2 - h_{wCP}^2|}{h_w^2} \right);$$

$$\Delta_F = |\gamma_{wj} - \gamma_{wjCP}|;$$

$$sqres_{CP} = tr(\mathbf{MRes}_{CP}^2), \text{ em que } \mathbf{MRes}_{CP} = \boldsymbol{\rho} - (\boldsymbol{\Gamma}_{CP}\boldsymbol{\Gamma}'_{CP} + \boldsymbol{\Psi}_{CP}).$$

Com base nas variáveis delta ( $\Delta$ ) assim definidas, foi verificado o desempenho de cada análise, partindo-se do princípio de que quanto mais próximos de 0 (zero) forem os valores das variáveis delta, mais próximos são os valores teóricos e os estimados pela análise de fatores, permitindo-se, portanto, avaliar a precisão do método de estimação.

Para avaliar a qualidade de ajuste do modelo fatorial foi utilizado o critério de KMO, que segundo Sharma (1996) classifica como média a adequação dos dados para o uso da análise de fatores quando o  $KMO \geq 0,70$ . Além disso, foi aplicado o teste de esfericidade de Bartlett a 5% de significância.

A fim de avaliar os efeitos da comunalidade, da distância proporcional entre as cargas e da rotação varimax sob cada variável avaliada, foi realizada uma análise de superfície de resposta de primeira ordem conforme o seguinte modelo:

$$y = \beta_0 + \beta_1 h + \beta_2 d + \beta_3 r + \beta_4 hr + \beta_5 hd + \beta_6 dr + \varepsilon,$$

em que:

$h$  = comunalidade (0,49; 0,56; 0,64; 0,72; 0,81; e 0,90);

$d$  = diferença entre as cargas fatoriais em proporção (0,81; 0,45; e 0,20);

$r$  = rotação (sem=0 e com=1);

$\varepsilon \sim N(0, \sigma^2)$ , sendo o resíduo

Os coeficientes não significativos foram retirados do modelo, um de cada vez e começando pelas interações, de acordo com o teste de t de Student a 5% de significância. O modelo foi composto somente por efeitos significativos, com exceções dos lineares que apresentaram interações significativas.

As análises estatísticas foram realizadas no software R, versão 3.3.1 e no Microsoft Excel 2016.

## 5. RESULTADOS E DISCUSSÃO

### 5.1. Matrizes de Correlações

Um pré-requisito básico para aplicar a análise de fatores é certificar-se de que as variáveis aleatórias não sejam independentes (RENCHEER; CHRISTENSEN, 2012), ou seja, garantir que a matriz de correlações seja apropriada para justificar a sua aplicação.

Para avaliar este requisito, foram utilizados o teste de esfericidade de Bartlett e a medida de adequação da amostragem, proposta por Kaiser-Meyer-Olkin (KMO), cujos resultados são apresentados nas Tabelas 5.1, 5.2, 5.3, 5.4, 5.5 e 5.6.

Tabela 5.1 - Valores calculados e entre parênteses os  $p$  – valores do teste de esfericidade de Bartlett e do KMO para  $h_w^2 = 0,49$

Estatística	$d_w = 0,81$	$d_w = 0,45$	$d_w = 0,20$
Bartlett	58,20* (0,000104)	84,16* ( $4,92 \times 10^{-10}$ )	102,43* ( $7,82 \times 10^{-14}$ )
KMO	0,56	0,75	0,79

\*significativo pelo teste de Bartlett ( $P < 0,05$ ).

Tabela 5.2 - Valores calculados e entre parênteses os  $p$  – valores do teste de esfericidade de Bartlett e do KMO para  $h_w^2 = 0,56$

Estatística	$d_w = 0,81$	$d_w = 0,45$	$d_w = 0,20$
Bartlett	80,11* ( $3,38 \times 10^{-9}$ )	112,18* ( $7,13 \times 10^{-16}$ )	136,57* ( $5,29 \times 10^{-21}$ )
KMO	0,58	0,76	0,81

\*significativo pelo teste de Bartlett ( $P < 0,05$ ).

Tabela 5.3 - Valores calculados e entre parênteses os  $p$  – valores do teste de esfericidade de Bartlett e do KMO para  $h_w^2 = 0,64$

Estatística	$d_w = 0,81$	$d_w = 0,45$	$d_w = 0,20$
Bartlett	109,11* ( $3,13 \times 10^{-15}$ )	150,27* ( $6,80 \times 10^{-24}$ )	181,86* ( $1,37 \times 10^{-30}$ )
KMO	0,57	0,77	0,83

\*significativo pelo teste de Bartlett ( $P < 0,05$ ).

Tabela 5.4 - Valores calculados e entre parênteses os  $p$  – valores do teste de esfericidade de Bartlett e do KMO para  $h_w^2 = 0,72$

Estatística	$d_w = 0,81$	$d_w = 0,45$	$d_w = 0,20$
Bartlett	151,31* ( $4,1 \times 10^{-24}$ )	200,56* ( $1,44 \times 10^{-34}$ )	244,22* ( $7,04 \times 10^{-44}$ )
KMO	0,57	0,78	0,85

\*significativo pelo teste de Bartlett ( $P < 0,05$ ).

Tabela 5.5 - Valores calculados e entre parênteses os  $p$  – valores do teste de esfericidade de Bartlett e do KMO para  $h_w^2 = 0,81$

Estatística	$d_w = 0,81$	$d_w = 0,45$	$d_w = 0,20$
Bartlett	216,43* (6,01 x10 <sup>-38</sup> )	275,79* (1,25 x10 <sup>-50</sup> )	336,52* (1,21 x10 <sup>-63</sup> )
KMO	0,56	0,77	0,86

\*significativo pelo teste de Bartlett ( $P < 0,05$ ).

Tabela 5.6 - Valores calculados e entre parênteses os  $p$  – valores do teste de esfericidade de Bartlett e do KMO para  $h_w^2 = 0,90$

Estatística	$d_w = 0,81$	$d_w = 0,45$	$d_w = 0,20$
Bartlett	337,37* (7,95 x10 <sup>-64</sup> )	406,77* (9,79 x10 <sup>-79</sup> )	497,61* (2,75 x10 <sup>-98</sup> )
KMO	0,56	0,74	0,86

\*significativo pelo teste de Bartlett ( $P < 0,05$ ).

Foi possível verificar que para as 18 matrizes de correlações, rejeitou-se ( $P < 0,05$ ) pelo teste de esfericidade de Bartlett, a hipótese de nulidade,  $\rho = I$ , indicando que as correlações existentes entre as variáveis aleatórias padronizadas  $Z_1, Z_2, Z_3$  e  $Z_4$  foram suficientes para aplicar a análise de fatores.

Para o KMO, foram obtidos valores estimados que variaram entre 0,56 e 0,86 (Tabelas 5.1, 5.2, 5.3, 5.4, 5.5 e 5.6.).

De acordo com Sharma (1996), valores do KMO maiores que 0,70 caracterizam uma adequação média dos dados para a aplicação da análise de fatores. Portanto, de acordo com o critério de KMO ( $\geq 0,70$ ) apenas para os casos em que as distâncias proporcionais entre as cargas fatoriais são iguais a 0,45 e 0,20, independentemente da comunalidade entre as variáveis aleatórias padronizadas  $Z_1, Z_2, Z_3$  e  $Z_4$ , as matrizes de correlações apresentaram-se adequadas para a utilização da análise de fatores.

Para a distância proporcional entre cargas fatoriais de 0,81, existem um número substancial (50%) de correlações entre as variáveis com valores próximos à 0,30, que para Tabachnick e Fidell (2007 *apud* Willians et al., 2010) e Hair et al. (2014), indicam uma não adequação para a aplicação da análise de fatores.

Por outro lado, o teste de esfericidade de Bartlett apresentou significância ( $P < 0,05$ ) em todos os casos estudados, indicando que há adequabilidade das 18 matrizes de correlações para a aplicação da análise de fatores, conforme os resultados apresentados nas Tabelas 5.1 a 5.6.

Em geral, os resultados do critério de KMO e do teste de esfericidade de Bartlett tendem a ser uniformes em indicar a adequabilidade dos dados para o uso da análise de fatores (DZIUBAN; SHIRKEY, 1974 *apud* DAMÁSIO, 2012), fato que não foi constatado no presente estudo. Todavia, os dois métodos de avaliação coincidiram em suas recomendações apenas para todos os casos cujas distâncias proporcionais entre as cargas fatoriais foram de 0,45 e 0,20. Nessas situações as matrizes de correlações se caracterizaram por apresentarem todos os seus valores maiores que 0,30 e mais próximos entre si.

Porém, na prática, a proximidade entre as correlações das variáveis dificulta o agrupamento delas em fatores distintos.

Pelo critério de KMO foi sinalizado que quanto menor a distância entre as cargas fatoriais, mais adequado foi o uso da análise de fatores, independentemente da comunalidade entre as variáveis.

De acordo com a superfície de resposta de primeira ordem estimada para os valores calculados e o *p* – *valor* do teste de esfericidade de Bartlett e para o KMO, obtiveram-se as seguintes equações de regressão ajustadas (Tabela 5.7).

Tabela 5.7 - Equações de regressão ajustadas para *valor calculado*, *p* – *valor* e KMO

Variável	Equação de regressão ajustada	R <sup>2</sup>
<i>valor calculado</i> (Bartlett)	$- 256,84 + 775,46*h - 148,08*d$	0,93
<i>p</i> – <i>valor</i> (Bartlett)	$5,79 \times 10^{-6}$	-----
KMO	$0,93771 - 0,44736*d$	0,95

\*significativo pelo teste t de Student ( $P < 0,05$ ).

O aumento da comunalidade proporcionou aumento ( $P > 0,05$ ) no valor calculado do teste de esfericidade de Bartlett, porém não ( $P < 0,05$ ) em sua significância. Por outro lado, a diminuição da distância proporcional entre as cargas fatoriais dos dois fatores aumentou ( $P < 0,05$ ) o valor calculado do referido teste. Esse resultado demonstrou que o teste de esfericidade de Bartlett está relacionado com a estrutura da matriz de correlações. Quanto mais ela se distancia da matriz identidade, como pode-se verificar nos casos de maior comunalidade e menor distância, esse teste aumenta o seu valor calculado. No entanto, todos esses valores calculados estiveram associados com altas significâncias.

A relação positiva com a comunalidade confere maior viabilidade para a aplicação da análise de fatores. Por outro lado, e contrária à expectativa, a relação negativa com a

distância não proporciona a mesma viabilidade, dado que as correlações ficam mais próximas entre si, o que pode dificultar o agrupamento das variáveis em fatores independentes.

Diante do exposto, a confiabilidade do teste de esfericidade de Bartlett em avaliar a qualidade de ajuste dos dados para a aplicação da análise de fatores é baixa, tendo em vista que quanto menor for distância entre as cargas fatoriais, mais difícil será a identificação de fatores comuns e, conseqüentemente, o agrupamento das variáveis nesses fatores.

Além disso, percebeu-se que quanto menor foi a distância entre as cargas fatoriais, maiores foram as correlações, e conseqüentemente, as correlações parciais entre as variáveis, elevando-se assim os valores do índice KMO (Tabela 5.7).

As matrizes de correlações com distância proporcional de 0,81 entre as cargas fatoriais (Tabelas 4.8 a 4.13) apresentaram um número substancial (> 50%) de correlações entre variáveis menor que 0,3, exceto para a matriz com comunalidade de 0,9 (Tabela 4.13) e, nesses casos, de acordo com Hair et al. (2014), a aplicação da análise de fatores pode ser considerada inapropriada. Tal fato coincide com a interpretação do índice KMO, que apresentou, para essas matrizes, valores de 0,56 ou 0,57. Porém, por sofrer somente influência negativa da distância, de modo contrária à expectativa, na prática, o uso de apenas esse critério pode dificultar e, ou comprometer os resultados da análise de fatores.

No entanto, enquanto o teste de esfericidade de Bartlett foi significativo para as 18 matrizes de correlações, apresentando grande sensibilidade em rejeitar  $H_0: \rho = I$ , independentemente das correlações entre as matrizes de correlações estudadas (Tabelas 4.8 a 4.13), o KMO conseguiu diferenciar as matrizes em termos de viabilidade para a aplicação da análise de fatores em termos das proximidades entre as cargas fatoriais, mas notou-se que são obtidos altos ou baixos valores de KMO, independentemente se as comunalidades são altas ou baixas.

Porém, como a comunalidade está relacionada com a qualidade da análise de fatores, os métodos que avaliam tal particularidade deveriam se relacionar a essa grandeza. Assim, com base no modelo fatorial ortogonal e como apenas o teste de esfericidade de Bartlett apresentou essa relação, concluiu-se que ele foi mais adequado em avaliar a viabilidade para a utilização da análise de fatores. Já o KMO não é um bom critério para essa avaliação.

Apesar de o teste de Bartlett variar em função da variação da comunalidade, ele foi significativo ( $p - valor < 0,05$ ) para  $h_w^2 \geq 0,49$ . Isso indicou que este teste foi sensível à rejeição de  $H_0$  e que, conseqüentemente, poderá validar a análise de fatores mesmo em situações não muito adequadas. Portanto, recomenda-se como a melhor referência para

utilizar a análise de fatores de maneira mais representativa da realidade das variáveis estudadas, a estimativa da comunalidade.

## 5.2 Método dos Componentes Principais

### 5.2.1. Comunalidade

Segundo Cruz (2006), a comunalidade é a medida da eficiência da representação da variável padronizada  $Z_w$  por uma parte comum, também envolvida nas demais variáveis.

Para cada uma das 18 matrizes de correlações, foram obtidas as comunalidades ( $h_{wCP}^2$ ) e calculados a variação percentual a fim de descrever a relação entre a comunalidade teórica e a estimada pelo método dos componentes principais, cujos resultados são apresentados na Tabela 5.8.

Tabela 5.8 - Comunalidades ( $h_{wCP}^2$ ) obtidas de acordo com as seis matrizes de correlações e os percentuais de variação em relação às comunalidades teóricas ( $h_w^2$ ), independentemente de  $d_w$

$h_w^2$	$h_{wCP}^2$	% de variação
0,49	0,75	53,1
0,56	0,78	39,3
0,64	0,82	28,1
0,72	0,86	19,4
0,81	0,91	12,4
0,90	0,95	5,6

Em seguida, estimaram-se os erros relativos médios dos desvios entre as comunalidades teórica e estimada ( $\Delta_{h^2}$ ) para cada uma das 18 matrizes, cuja equação de regressão estimada está apresentada na Tabela 5.9.

Tabela 5.9 - Equação de regressão ajustada para a variável  $\Delta_{h^2}$

Variável	Equação de regressão ajustada	$R^2$
$\Delta_{h^2}$	$1,02331 - 1,11124 * h$	0,97

\*significativo pelo teste t de Student ( $P < 0,05$ ).

Notou-se que  $\Delta_{h^2}$  relacionou-se ( $P < 0,05$ ) de maneira inversa com a comunalidade teórica, de tal modo que quanto menor for essa comunalidade, maior será seu erro relativo médio de obtenção pelo método dos componentes principais. Segue, portanto, que quanto maior for a comunalidade real dos dados, melhor será sua estimativa e, conseqüentemente, as interpretações baseadas na análise de fatores.

Como o método dos componentes principais superestima a comunalidade, avaliou-se a  $sqres$ , a fim de recomendar melhor, o menor valor  $h^2$  que viabiliza a análise de fatores.

Ferreira (2011) afirma que a soma de quadrados da matriz de resíduos serve de parâmetro para avaliar a qualidade de ajuste do modelo fatorial. Quanto maior for a comunalidade das variáveis, menor será a variância específica delas, e conseqüentemente, menor será a diferença  $MRes_{CP} = \rho - (\Gamma_{CP}\Gamma'_{CP} + \Psi_{CP})$ , uma vez que a diagonal da matriz  $\Psi_{CP}$  equivale a  $1 - h^2_{CP}$ .

Assim, obteve-se a  $sqres_{CP}$  para as 18 matrizes de correlações e seus resultados são apresentados na Tabela 5.10.

Tabela 5.10 - Somas de quadrados dos elementos da matriz de resíduos obtidas de acordo com as seis matrizes de correlações, independentemente de  $d_w$

$h^2_w$	$sqres_{CP}$
0,49	0,26
0,56	0,19
0,64	0,13
0,72	0,08
0,81	0,04
0,90	0,01

Na seqüência, foi feita também a análise de regressão da soma de quadrados dos resíduos  $sqres_{CP}$  resultantes das 18 matrizes de correlações, cuja equação de regressão ajustada é apresentada na Tabela 5.11.

Tabela 5.11 - Equação de regressão ajustada para  $sqres_{CP}$

Variável	Equações de regressão ajustada	R <sup>2</sup>
$sqres_{CP}$	$0,53537 - 0,60782 * h$	0,96

\*significativo pelo teste t de Student ( $P < 0,05$ ).

A variável  $sqres$  diminuiu ( $P < 0,05$ ) à medida que a comunalidade ( $h_w^2$ ) aumentou. Portanto, quanto maior for a comunalidade real das variáveis em estudo, melhor será o ajuste do modelo fatorial.

Verificou-se que para todas as 18 matrizes de correlações estudadas os valores da soma de quadrados dos resíduos ( $sqres_{CP}$ ) foram menores que 0,26. Porém, segundo Mingoti (2005), idealmente esses valores devem ser o mais próximo possível de zero, e apenas para as matrizes com comunalidades  $h_w^2 \geq 0,64$ , tal grandeza tem valores consideravelmente pequenos ( $sqres_{CP} \leq 0,13$ ).

Segundo Cruz et al. (2014), valores de comunalidades superiores a 0,64 têm sido aceitos como razoáveis, pois são equivalentes a uma correlação próxima de 0,80 entre a variável padronizada e a parte comum ( $\gamma_{wj}$ ), que explica esta variável. Isto pode ser constatado nos fatores teóricos (Tabelas 4.3 a 4.6) em que  $h_w^2 \geq 0,64$  e as distâncias proporcionais entre as cargas fatoriais foram de 0,81 e 0,45, diferentemente de quando a distância proporcional de 0,20 entre as cargas, conforme esperado.

Constatou-se que o aumento da comunalidade melhorou a qualidade do ajuste do modelo fatorial, por provocar menos erro na estimativa da própria comunalidade e também na soma de quadrados de resíduos, bem como aumentou o valor calculado do teste de esfericidade de Bartlett.

Souza (1988) (*apud* Cruz (2006)) relata serem aceitos como razoáveis valores de comunalidades superior a 0,64. Já Figueiredo e Silva (2010) afirmam que o valor mínimo aceitável para a comunalidade é de 0,50 e para Hair et al. (2014), variáveis com comunalidades inferiores a 0,50 não apresentam explicações suficientes pela solução fatorial.

Entretanto, a fim de se obter uma soma de quadrados do resíduo e um erro relativo médio do desvio de comunalidade suficientemente pequenos, e uma estimativa mais adequada para análise de fatores pelo teste de Bartlett, recomenda-se comunalidades  $h^2 \geq 0,64$ .

Tal conclusão se justifica porque a medida que à comunalidade teórica aumenta, seu erro de estimativa diminui.

## 5.2.2. Cargas Fatoriais

### 5.2.2.1. Sem Rotação

Segundo Figueiredo e Silva (2010), o pesquisador deve observar as cargas fatoriais de cada variável, procurando identificar variáveis que apresentam elevada carga em um fator, no caso, acima de 0,40, uma vez que a mesma variável não pode contribuir para as construções de fatores distintos.

Hair et al. (2014) também recomendam que se adote 0,40 como limite aceitável de contribuição de uma variável na criação do fator, para evitar problemas da indeterminação da relação entre variáveis e fatores.

São apresentadas nas Tabelas 5.12, 5.13, 5.14, 5.15, 5.16 e 5.17 as matrizes de cargas fatoriais e de variâncias específicas obtidas de acordo com o método dos componentes principais para cada uma das 18 matrizes de correlações estudadas.

Tabela 5.12 - Matrizes  $\Gamma_{CP}$  e  $\Psi_{CP}$  obtidas para  $h_w^2 = 0,49$  ( $w = 1, 2, 3$  e  $4$ )

Matriz	$d_w = 0,81$	$d_w = 0,45$	$d_w = 0,20$
$\Gamma_{CP}$	$\begin{bmatrix} 0,68 & 0,54 \\ 0,68 & 0,54 \\ 0,68 & -0,54 \\ 0,68 & -0,54 \end{bmatrix}$	$\begin{bmatrix} 0,75 & -0,42 \\ 0,75 & -0,42 \\ 0,75 & 0,42 \\ 0,75 & 0,42 \end{bmatrix}$	$\begin{bmatrix} 0,78 & 0,37 \\ 0,78 & 0,37 \\ 0,78 & -0,37 \\ 0,78 & -0,37 \end{bmatrix}$
$\Psi_{CP}$	$\begin{bmatrix} 0,26 & 0 & 0 & 0 \\ 0 & 0,26 & 0 & 0 \\ 0 & 0 & 0,26 & 0 \\ 0 & 0 & 0 & 0,26 \end{bmatrix}$	$\begin{bmatrix} 0,26 & 0 & 0 & 0 \\ 0 & 0,26 & 0 & 0 \\ 0 & 0 & 0,26 & 0 \\ 0 & 0 & 0 & 0,26 \end{bmatrix}$	$\begin{bmatrix} 0,26 & 0 & 0 & 0 \\ 0 & 0,26 & 0 & 0 \\ 0 & 0 & 0,26 & 0 \\ 0 & 0 & 0 & 0,26 \end{bmatrix}$

Tabela 5.13 - Matrizes  $\Gamma_{CP}$  e  $\Psi_{CP}$  obtidas para  $h_w^2 = 0,56$  ( $w = 1, 2, 3$  e  $4$ )

Matriz	$d_w = 0,81$	$d_w = 0,45$	$d_w = 0,20$
$\Gamma_{CP}$	$\begin{bmatrix} 0,70 & -0,54 \\ 0,70 & -0,54 \\ 0,70 & 0,54 \\ 0,70 & 0,54 \end{bmatrix}$	$\begin{bmatrix} 0,78 & -0,41 \\ 0,78 & -0,41 \\ 0,78 & 0,41 \\ 0,78 & 0,41 \end{bmatrix}$	$\begin{bmatrix} 0,81 & 0,35 \\ 0,81 & 0,35 \\ 0,81 & -0,35 \\ 0,81 & -0,35 \end{bmatrix}$
$\Psi_{CP}$	$\begin{bmatrix} 0,22 & 0 & 0 & 0 \\ 0 & 0,22 & 0 & 0 \\ 0 & 0 & 0,22 & 0 \\ 0 & 0 & 0 & 0,22 \end{bmatrix}$	$\begin{bmatrix} 0,22 & 0 & 0 & 0 \\ 0 & 0,22 & 0 & 0 \\ 0 & 0 & 0,22 & 0 \\ 0 & 0 & 0 & 0,22 \end{bmatrix}$	$\begin{bmatrix} 0,22 & 0 & 0 & 0 \\ 0 & 0,22 & 0 & 0 \\ 0 & 0 & 0,22 & 0 \\ 0 & 0 & 0 & 0,22 \end{bmatrix}$

Tabela 5.14 - Matrizes  $\Gamma_{CP}$  e  $\Psi_{CP}$  obtidas para  $h_w^2 = 0,64$  ( $w = 1, 2, 3$  e  $4$ )

Matriz	$d_w = 0,81$	$d_w = 0,45$	$d_w = 0,20$
$\Gamma_{CP}$	$\begin{bmatrix} 0,72 & -0,55 \\ 0,72 & -0,55 \\ 0,72 & 0,55 \\ 0,72 & 0,55 \end{bmatrix}$	$\begin{bmatrix} 0,82 & -0,39 \\ 0,82 & -0,39 \\ 0,82 & 0,39 \\ 0,82 & 0,39 \end{bmatrix}$	$\begin{bmatrix} 0,85 & -0,32 \\ 0,85 & -0,32 \\ 0,85 & 0,32 \\ 0,85 & 0,32 \end{bmatrix}$
$\Psi_{CP}$	$\begin{bmatrix} 0,18 & 0 & 0 & 0 \\ 0 & 0,18 & 0 & 0 \\ 0 & 0 & 0,18 & 0 \\ 0 & 0 & 0 & 0,18 \end{bmatrix}$	$\begin{bmatrix} 0,18 & 0 & 0 & 0 \\ 0 & 0,18 & 0 & 0 \\ 0 & 0 & 0,18 & 0 \\ 0 & 0 & 0 & 0,18 \end{bmatrix}$	$\begin{bmatrix} 0,18 & 0 & 0 & 0 \\ 0 & 0,18 & 0 & 0 \\ 0 & 0 & 0,18 & 0 \\ 0 & 0 & 0 & 0,18 \end{bmatrix}$

Tabela 5.15 - Matrizes  $\Gamma_{CP}$  e  $\Psi_{CP}$  obtidas para  $h_w^2 = 0,72$  ( $w = 1, 2, 3$  e  $4$ )

Matriz	$d_w = 0,81$	$d_w = 0,45$	$d_w = 0,20$
$\Gamma_{CP}$	$\begin{bmatrix} 0,75 & -0,55 \\ 0,75 & -0,55 \\ 0,75 & 0,55 \\ 0,75 & 0,55 \end{bmatrix}$	$\begin{bmatrix} 0,85 & -0,38 \\ 0,85 & -0,38 \\ 0,85 & 0,38 \\ 0,85 & 0,38 \end{bmatrix}$	$\begin{bmatrix} 0,88 & -0,29 \\ 0,88 & -0,29 \\ 0,88 & 0,29 \\ 0,88 & 0,29 \end{bmatrix}$
$\Psi_{CP}$	$\begin{bmatrix} 0,14 & 0 & 0 & 0 \\ 0 & 0,14 & 0 & 0 \\ 0 & 0 & 0,14 & 0 \\ 0 & 0 & 0 & 0,14 \end{bmatrix}$	$\begin{bmatrix} 0,14 & 0 & 0 & 0 \\ 0 & 0,14 & 0 & 0 \\ 0 & 0 & 0,14 & 0 \\ 0 & 0 & 0 & 0,14 \end{bmatrix}$	$\begin{bmatrix} 0,14 & 0 & 0 & 0 \\ 0 & 0,14 & 0 & 0 \\ 0 & 0 & 0,14 & 0 \\ 0 & 0 & 0 & 0,14 \end{bmatrix}$

Tabela 5.16 - Matrizes  $\Gamma_{CP}$  e  $\Psi_{CP}$  obtidas para  $h_w^2 = 0,81$  ( $w = 1, 2, 3$  e  $4$ )

Matriz	$d_w = 0,81$	$d_w = 0,45$	$d_w = 0,20$
$\Gamma_{CP}$	$\begin{bmatrix} 0,77 & -0,56 \\ 0,77 & -0,56 \\ 0,77 & 0,56 \\ 0,77 & 0,56 \end{bmatrix}$	$\begin{bmatrix} 0,88 & -0,36 \\ 0,88 & -0,36 \\ 0,88 & 0,36 \\ 0,88 & 0,36 \end{bmatrix}$	$\begin{bmatrix} 0,92 & 0,25 \\ 0,92 & 0,25 \\ 0,92 & -0,25 \\ 0,92 & -0,25 \end{bmatrix}$
$\Psi_{CP}$	$\begin{bmatrix} 0,10 & 0 & 0 & 0 \\ 0 & 0,10 & 0 & 0 \\ 0 & 0 & 0,10 & 0 \\ 0 & 0 & 0 & 0,10 \end{bmatrix}$	$\begin{bmatrix} 0,10 & 0 & 0 & 0 \\ 0 & 0,10 & 0 & 0 \\ 0 & 0 & 0,10 & 0 \\ 0 & 0 & 0 & 0,10 \end{bmatrix}$	$\begin{bmatrix} 0,10 & 0 & 0 & 0 \\ 0 & 0,10 & 0 & 0 \\ 0 & 0 & 0,10 & 0 \\ 0 & 0 & 0 & 0,10 \end{bmatrix}$

Tabela 5.17 - Matrizes  $\Gamma_{CP}$  e  $\Psi_{CP}$  obtidas para  $h_w^2 = 0,90$  ( $w = 1, 2, 3$  e  $4$ )

Matriz	$d_w = 0,81$	$d_w = 0,45$	$d_w = 0,20$
$\Gamma_{CP}$	$\begin{bmatrix} 0,80 & -0,56 \\ 0,80 & -0,56 \\ 0,80 & 0,56 \\ 0,80 & 0,56 \end{bmatrix}$	$\begin{bmatrix} 0,91 & -0,34 \\ 0,91 & -0,34 \\ 0,91 & 0,34 \\ 0,91 & 0,34 \end{bmatrix}$	$\begin{bmatrix} 0,95 & 0,21 \\ 0,95 & 0,21 \\ 0,95 & -0,21 \\ 0,95 & -0,21 \end{bmatrix}$
$\Psi_{CP}$	$\begin{bmatrix} 0,05 & 0 & 0 & 0 \\ 0 & 0,05 & 0 & 0 \\ 0 & 0 & 0,05 & 0 \\ 0 & 0 & 0 & 0,05 \end{bmatrix}$	$\begin{bmatrix} 0,05 & 0 & 0 & 0 \\ 0 & 0,05 & 0 & 0 \\ 0 & 0 & 0,05 & 0 \\ 0 & 0 & 0 & 0,05 \end{bmatrix}$	$\begin{bmatrix} 0,05 & 0 & 0 & 0 \\ 0 & 0,05 & 0 & 0 \\ 0 & 0 & 0,05 & 0 \\ 0 & 0 & 0 & 0,05 \end{bmatrix}$

Com base nas cargas fatoriais obtidas, as variáveis  $Z_1, Z_2, Z_3$  e  $Z_4$  foram relacionadas ao fator  $F_1$ , o que implicaria conclusões erradas para todas as 18 matrizes de

correlações, dado que, *a priori*,  $Z_1$  e  $Z_2$  associam-se com o fator  $F_1$  e  $Z_3$  e  $Z_4$  com  $F_2$ . Tal conclusão se baseou na diferença entre as matrizes de cargas fatoriais.

De acordo com as diferenças entre as cargas fatoriais de cada variável em cada fator, foi realizada a análise de regressão baseada no erro relativo médio dos desvios das cargas fatoriais cuja equação de regressão ajustada é apresentada na Tabela 5.18.

Tabela 5.18 - Equação de regressão ajustada para  $\Delta_r$

Variável	Equação de regressão ajustada	R <sup>2</sup>
$\Delta_r$	$0,0535 + 2,4886*d$	0,82

\*significativo pelo teste t de Student ( $P < 0,05$ ).

Concluiu-se, portanto, que a qualidade da estimativa das cargas fatoriais aumentou com a diminuição da distância, ao contrário da expectativa, uma vez que quanto menor for o distanciamento entre os valores das cargas fatoriais, mais próximos entre si serão os valores das correlações, dificultando, todavia, o agrupamento das variáveis nos fatores.

Esperava-se que o  $\Delta_r$  diminuísse em função do aumento da comunalidade e da diminuição da distância, cujos cenários caracterizam maior relação entre os fatores e as suas respectivas variáveis agrupadas. Nesses casos, o modelo fatorial está bastante associado à realidade do fator. Logo, a falta desse resultado levou à conclusão de que o método de estimação dos fatores por componentes principais não se mostrou adequado, sem rotação.

Pode-se constatar também um distanciamento entre os valores obtidos e teóricos das cargas fatoriais, bem como a disposição dessas correlações dentro de cada fator (Tabelas 4.1 a 4.6 e Tabelas 5.12 a 5.17).

Em todas as matrizes de cargas fatoriais obtidas sem rotação, independentemente da distância proporcional entre as cargas fatoriais, o método dos componentes principais igualou, em módulo, os valores das cargas em cada fator, de maneira oposta à que foi determinada, podendo induzir uma interpretação errada de que as quatro variáveis são altamente correlacionadas entre si, e representadas por um único fator comum, mesmo partindo-se do pressuposto de que a  $h_w^2 \geq 0,64$  (Tabelas 5.12 a 5.17).

### 5.2.2.2. Com Rotação

Na expectativa de melhorar a interpretação dos fatores, para cada uma das 18 matrizes de correlações, foi realizada a análise de fatores utilizando-se, após o método de estimação dos componentes principais, a rotação varimax.

As matrizes  $\Gamma_{CPR}$  obtidas a partir das análises de fatores realizadas em cada uma das 18 matrizes de correlações, utilizando-se a rotação varimax são descritas na Tabela 5.19.

Tabela 5.19 - Matrizes  $\Gamma_{CPR}$  obtidas para as seis comunalidades  $h_w^2$  ( $w = 1, 2, 3$  e  $4$ )

$h_w^2$	$d_w = 0,81$	$d_w = 0,45$	$d_w = 0,20$
0,49	$\begin{bmatrix} \mathbf{0,80} & -0,32 \\ \mathbf{0,80} & -0,32 \\ 0,49 & \mathbf{0,71} \\ 0,49 & \mathbf{0,71} \end{bmatrix}$	$\begin{bmatrix} 0,26 & \mathbf{0,82} \\ 0,26 & \mathbf{0,82} \\ \mathbf{0,84} & 0,21 \\ \mathbf{0,84} & 0,21 \end{bmatrix}$	$\begin{bmatrix} 0,58 & \mathbf{0,64} \\ 0,58 & \mathbf{0,64} \\ \mathbf{0,86} & -0,05 \\ \mathbf{0,86} & -0,05 \end{bmatrix}$
0,56	$\begin{bmatrix} \mathbf{0,71} & -0,53 \\ \mathbf{0,71} & -0,53 \\ \mathbf{0,69} & 0,55 \\ \mathbf{0,69} & 0,55 \end{bmatrix}$	$\begin{bmatrix} \mathbf{0,88} & -0,02 \\ \mathbf{0,88} & -0,02 \\ 0,53 & \mathbf{0,71} \\ 0,53 & \mathbf{0,71} \end{bmatrix}$	$\begin{bmatrix} \mathbf{0,81} & 0,35 \\ \mathbf{0,81} & 0,35 \\ \mathbf{0,82} & -0,34 \\ \mathbf{0,82} & -0,34 \end{bmatrix}$
0,64	$\begin{bmatrix} 0,62 & \mathbf{0,67} \\ 0,62 & \mathbf{0,67} \\ \mathbf{0,80} & -0,42 \\ \mathbf{0,80} & -0,42 \end{bmatrix}$	$\begin{bmatrix} 0,58 & \mathbf{0,70} \\ 0,58 & \mathbf{0,70} \\ \mathbf{0,91} & -0,02 \\ \mathbf{0,91} & -0,02 \end{bmatrix}$	$\begin{bmatrix} 0,43 & \mathbf{0,80} \\ 0,43 & \mathbf{0,80} \\ \mathbf{0,85} & 0,32 \\ \mathbf{0,85} & 0,32 \end{bmatrix}$
0,72	$\begin{bmatrix} \mathbf{0,78} & -0,51 \\ \mathbf{0,78} & -0,51 \\ \mathbf{0,71} & 0,60 \\ \mathbf{0,71} & 0,60 \end{bmatrix}$	$\begin{bmatrix} \mathbf{0,92} & -0,13 \\ \mathbf{0,92} & -0,13 \\ \mathbf{0,71} & 0,59 \\ \mathbf{0,71} & 0,59 \end{bmatrix}$	$\begin{bmatrix} \mathbf{0,91} & -0,20 \\ \mathbf{0,91} & -0,20 \\ \mathbf{0,85} & 0,38 \\ \mathbf{0,85} & 0,38 \end{bmatrix}$
0,81	$\begin{bmatrix} \mathbf{0,91} & -0,29 \\ \mathbf{0,91} & -0,29 \\ 0,55 & \mathbf{0,77} \\ 0,55 & \mathbf{0,77} \end{bmatrix}$	$\begin{bmatrix} \mathbf{0,89} & -0,33 \\ \mathbf{0,89} & -0,33 \\ \mathbf{0,87} & 0,39 \\ \mathbf{0,87} & 0,39 \end{bmatrix}$	$\begin{bmatrix} \mathbf{0,95} & 0 \\ \mathbf{0,95} & 0 \\ \mathbf{0,82} & 0,49 \\ \mathbf{0,82} & 0,49 \end{bmatrix}$
0,90	$\begin{bmatrix} \mathbf{0,83} & -0,51 \\ \mathbf{0,83} & -0,51 \\ \mathbf{0,76} & 0,62 \\ \mathbf{0,76} & 0,62 \end{bmatrix}$	$\begin{bmatrix} \mathbf{0,93} & -0,31 \\ \mathbf{0,93} & -0,31 \\ \mathbf{0,90} & 0,38 \\ \mathbf{0,90} & 0,38 \end{bmatrix}$	$\begin{bmatrix} \mathbf{0,96} & -0,20 \\ \mathbf{0,96} & -0,20 \\ \mathbf{0,95} & 0,22 \\ \mathbf{0,95} & 0,22 \end{bmatrix}$

Para Damásio (2012) as rotações fatoriais têm o objetivo de facilitar a interpretação dos fatores, tendo em vista que muitas vezes as variáveis analisadas apresentam cargas fatoriais elevadas em mais de um fator.

Idealmente, deseja-se que cada variável tenha uma carga fatorial alta, em módulo, em um único fator, devendo, portanto, apresentar cargas absolutas pequenas ou moderadas nos demais fatores (JOHNSON; WICHERN, 2007).

Conforme já mencionado, definiu-se, *a priori*, que representar as variáveis  $Z_1$  e  $Z_2$  pelo fator  $F_1$  e as variáveis  $Z_3$  e  $Z_4$  pelo fator  $F_2$ . De acordo com as matrizes de cargas fatoriais obtidas a partir do método dos componentes principais com rotação varimax, verificou-se que, na maioria dos casos (55,6%) não houve favorecimento da separação das variáveis nos dois fatores.

Mesmo nos casos em que a rotação varimax possibilitou uma correta separação das variáveis  $Z_1$  e  $Z_2$  em  $F_1$  e as variáveis  $Z_3$  e  $Z_4$  em  $F_2$ , tal resultado não apresentou um padrão de explicação. Isso implica em não recomendar essa rotação combinada com o método de estimação dos componentes principais.

Contudo, nem a rotação dos fatores para as comunalidades  $h_w^2 \geq 0,64$ , com exceção do cenário com comunalidade  $h_w^2 = 0,81$  e  $d_w = 0,81$ , foi capaz de possibilitar interpretações corretas.

### 5.2.3. Autovalores

É sabido que cada fator possui um autovalor associado que representa a quantidade total de variância explicada por ele.

De acordo com Garson (2009, *apud* FIGUEIREDO; SILVA, 2010), o autovalor mede a variância em todas as variáveis que é devida ao fator. Se um fator tem um autovalor baixo ele contribui pouco para a explicação das variâncias das variáveis originais e pode ser ignorado como redundante em relação a fatores mais importantes.

Os autovalores dos dois primeiros fatores obtidos para as 18 matrizes de correlações pelo método dos componentes principais sem rotação varimax são apresentados na Tabela 5.20.

Tabela 5.20 - Autovalores ( $\lambda_{j_{CP}}$ ) obtidos para as seis comunalidades  $h_w^2$  ( $w = 1, 2, 3$  e  $4$ )

$h_w^2$	$d_w = 0,81$		$d_w = 0,45$		$d_w = 0,20$	
	$F_1$	$F_2$	$F_1$	$F_2$	$F_1$	$F_2$
0,49	1,82	1,16	2,27	0,71	2,43	0,55
0,56	1,96	1,17	2,46	0,67	2,64	0,48
0,64	2,07	1,21	2,67	0,62	2,87	0,41
0,72	2,22	1,23	2,88	0,57	3,11	0,34
0,81	2,37	1,25	3,11	0,52	3,36	0,25
0,90	2,53	1,27	3,33	0,47	3,63	0,17

Ao utilizar o método dos componentes principais para realizar a análise de fatores, notou-se que os autovalores do primeiro fator foram superestimados, uma vez que mantendo o uso do critério de Kaiser ( $\lambda_j \geq 1$ ) para a escolha do número de fatores, apenas quando a distância proporcional entre as cargas fatoriais foi de 0,81, conservar-se-ia a utilização dos dois fatores na análise, ou seja, para todas as matrizes com cargas distantes em 0,45 e 0,20, um único fator poderia ser mantido no modelo fatorial, diferente do determinado de maneira teórica.

Esse resultado indicou que a grande distância (0,81) entre as variáveis  $Z_1$  e  $Z_2$  com as variáveis  $Z_3$  e  $Z_4$  possibilitou as obtenções dos fatores  $F_1$  e  $F_2$ , facilitando o agrupamento correto das variáveis. No entanto, para as distâncias de 0,45 e 0,20 que retratam mais proximidades entre as relações das variáveis com os dois fatores, o método dos componentes principais não possibilitou tal obtenção, por acarretar somente a estimativa de  $\lambda_1 \geq 1$  e consequentemente agrupar todas elas em um único fator.

Percebeu-se, desse modo, que a determinação do uso de um ou dois fatores obtida com base na análise dos autovalores contrariou a interpretação obtida com base nas cargas fatoriais estimadas a partir da análise de fatores realizada através do método dos componentes principais sem rotação varimax que, conforme Tabelas 5.11 a 5.16, verificou-se a necessidade de apenas um fator para a representação das quatro variáveis para todas as 18 matrizes de correlações estudadas.

Isso deveu-se ao fato de que as cargas do primeiro fator, isto é, as correlações entre as variáveis e o primeiro fator, também foram superestimadas, tendo em vista que o autovalor é calculado por:

$$\lambda_j = \sum_{w=1}^4 \gamma_{wj}^2, \text{ para } j = 1 \text{ e } 2,$$

em que  $\gamma_{wj}^2$  representa o quadrado da carga fatorial associado a  $w$ -ésima variável no  $j$ -ésimo fator.

Logo, o método dos componentes principais parece ter uma característica de superestimar o máximo da variabilidade presente na matriz de correlações no primeiro autovalor. Por isso, quanto maior for a comunalidade e a distância entre as correlações das variáveis, menor será o erro de estimativa desse autovalor.

Qualquer variação por menor que seja da maior comunalidade e da menor distância provoca um erro acentuado.

Se ocorre um erro muito grande na estimativa do primeiro autovalor, a estimativa do segundo estará totalmente comprometida, o que inviabiliza a interpretação da equação  $\Delta_{\lambda_2}$ .

Ao realizar a análise de fatores com a rotação varimax, a matriz de cargas fatoriais ( $\Gamma$ ), obviamente tem seus valores modificados, mas de tal modo que:

$$\Gamma_{CP}\Gamma'_{CP} + \Psi_{CP} = \Gamma_{CP}^* \Gamma_{CP}^{*'} + \Psi_{CP},$$

sendo  $\Gamma_{CP}^* = \Gamma_{CP}T$ , em que  $TT' = T'T = I$ .

Porém, a partir da definição anterior para o cálculo dos autovalores, suas estimativas não se mantiveram as mesmas após a rotação, como mostra a Tabela 5.21.

Tabela 5.21 - Autovalores ( $\lambda_{jCP}$ ) obtidos após a rotação para as seis comunalidades  $h_w^2$  ( $w = 1, 2, 3$  e  $4$ )

$h_w^2$	$d_w = 0,81$		$d_w = 0,45$		$d_w = 0,20$	
	$F_1$	$F_2$	$F_1$	$F_2$	$F_1$	$F_2$
0,49	1,77	1,22	1,53	1,45	2,16	0,82
0,56	1,96	1,17	2,12	1,01	2,64	0,48
0,64	2,05	1,23	2,31	0,97	1,80	1,48
0,72	2,22	1,23	2,71	0,74	3,08	0,36
0,81	2,26	1,36	3,10	0,52	3,14	0,47
0,90	2,52	1,28	3,33	0,48	3,63	0,17

Do mesmo modo que na análise de fatores sem rotação, neste caso também notou-se que na maioria dos cenários utilizaria apenas um fator, conforme o critério de Kaiser, também de modo contrário ao determinado teoricamente.

Foram calculados os erros relativos médios dos desvios entre os autovalores teórico e estimado, que subsidiaram análises de regressão a partir dos resultados com e sem rotação varimax.

As equações de regressão ajustadas para as variáveis  $\Delta_{\lambda}$ ,  $\Delta_{\lambda_1}$  e  $\Delta_{\lambda_2}$  são apresentadas na Tabela 5.22, para resultados a partir da estimação sem e com rotação varimax.

Tabela 5.22 - Equações de regressão ajustadas a partir dos autovalores estimados para as variáveis  $\Delta_\lambda$ ,  $\Delta_{\lambda_1}$  e  $\Delta_{\lambda_2}$

Variáveis	Equações de regressão ajustadas	R <sup>2</sup>
$\Delta_\lambda$	$1,11477 - 0,82181*d - 0,12339*r$	0,74
$\Delta_{\lambda_1}$	$1,88082 - 0,72695*h - 0,86473*d - 0,16009*r$	0,71
$\Delta_{\lambda_2}$	$-0,4207 + 1,7789*h + 0,3836d - 1,6882*hd$	0,80

\*significativo pelo teste t de Student (  $P < 0,05$  ).

De modo geral, a estimativa do autovalor se aproximou mais do valor real à medida que a comunalidade e a distância entre as cargas aumentam.

A rotação varimax proporcionou uma melhora na estimativa do autovalor.

De modo geral, apesar da rotação varimax ter melhorado a interpretação dos fatores e das estimativas dos autovalores, mesmo assim ela não foi suficiente para viabilizar a utilização do método dos componentes principais.

Em suma, se a análise de fatores for realizada pelo método dos componentes principais, recomenda-se seguir às seguintes etapas:

- determinar o número de fatores a serem estimados com base no critério de Kaiser;
- obter as estimativas das comunalidades das variáveis analisadas; para aquelas com comunalidades maiores que 0,64, deve-se incluí-las no modelo fatorial; caso contrário, não;
- verificar a coerência das interpretações dos autovalores e das cargas fatoriais com relação ao número de fatores; se forem coerentes, entende-se que a análise de fatores está bem aplicada; caso contrário, deve-se desconsiderar sua aplicação;
- não há necessidade de realizar a rotação varimax.

## 6. CONCLUSÕES

O teste de esfericidade de Bartlett é muito sensível em validar a adequação da análise de fatores. Já o critério de KMO está relacionado com as magnitudes das correlações da matriz utilizada, o que pode não ser adequado para a análise de fatores quando essas correlações são altas e não há a possibilidade de discriminar as variáveis estudadas em diferentes fatores.

O erro de estimativa da comunalidade diminui com o aumento da comunalidade teórica das variáveis estudadas e o seu aumento melhora a qualidade de ajuste do modelo fatorial, uma vez que as variâncias residuais diminuem. Contudo, recomenda-se a estimativa mínima de 0,64 para a comunalidade de cada variável, quando a análise de fatores for obtida pelo método dos componentes principais.

As estimativas das cargas fatoriais aproximam-se dos seus respectivos parâmetros à medida que as distâncias proporcionais entre as cargas fatoriais diminuam.

As estimativas dos autovalores não são adequadas nem antes e nem após a rotação.

Para a análise de fatores realizada com o método dos componentes principais não se recomenda a rotação varimax.

O método dos componentes principais utilizado na análise de fatores quando as variáveis aleatórias seguem distribuição de probabilidades normal p-variada e, principalmente, com comunalidades teóricas abaixo que 0,64, não proporciona boas estimativas.

A análise de fatores por meio do método dos componentes principais com rotação varimax nem sempre fornecerá interpretações estatísticas próximas das que realmente são. Por isso, as interpretações técnicas e as informações de estudos especificamente relacionados, serão de grande importância para ajudar na decisão de aceitar, ou não, os resultados dessa análise.

## 7. REFERÊNCIAS BIBLIOGRÁFICAS

ABREU, V. M. N.; CRUZ, C. D.; SILVA, M. A.; FIGUEIREDO, E. A. P.; ABREU, P. G. **Estudo da capacidade de combinação de linhagens de matrizes de frangos de corte, por meio da análise de fatores.** Produção de ovos. R. Bras. Zootec., Viçosa, v. 28, n. 2, p. 258-263, Apr. 1999.

ALMEIDA, L. S.; GUIMARÃES, E. C. Geoestatística e análise fatorial exploratória para representação espacial de atributos químicos do solo, na cafeicultura. **Coffee Science**, v. 11, n. 2, p. 195 - 203, abr./jun. 2016.

CRUZ, C. D.; **Programa GENES: análise multivariada e simulação.** Viçosa: Editora UFV, 2006.

CRUZ, C. D.; CARNEIRO, P. C. S.; REGAZZI, A. J. **Modelos biométricos aplicados ao melhoramento genético: volume 2.** 3ª ed. Viçosa: Editora UFV, 2014.

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

FABRIGAR, L. R., WEGENER, D. T., MACCALLUM, R. C., STRAHAN, E. J. Evaluating the use of exploratory factor analysis in psychological research. **Psychological Methods**, v. 4, n. 3, p. 272-299, 1999.

FERREIRA, A.; CECON, P. R.; CRUZ, C. D.; FERRÃO, R. G.; SILVA, M. F.; FONSECA, A. F. A.; FERRÃO, M. A. G. Seleção simultânea de Coffea canephora por meio da combinação de análise de fatores e índices de seleção. **Pesquisa Agropecuária Brasileira**, Brasília, v. 40, n. 12, p. 1189-1195, 2005.

FERREIRA, D. F. **Estatística multivariada.** 2ª ed. Lavras: Editora UFLA, 2011.

FIGUEIREDO FILHO, D. B.; JÚNIOR, J. A. S. Visão além do alcance: uma introdução à análise fatorial. **Opinião Pública**, Campinas, v. 16, n. 1, p. 160-185, 2010.

FLOYD, F. J. e WIDAMAN, K. F. Factor analysis in the development and refinement of clinical assessment instruments. **Psychological Assessment, Psychological Assessment**, v. 7, n. 3, p. 286-299, 1995.

HAIR, J. F.; BLACK, W. C.; BABIN, B. J.; ANDERSON, R. E.; **Multivariate Data Analysis**. 7ª ed. United States of America: Pearson, 2014.

HARLOW, L. L. **The essence of multivariate thinking: basic themes and methods**. New Jersey: Lawrence Erlbaum Associates, Publishers, 2005.

HARMAN, H. H. **Modern factor analysis**. 3ª ed. Chicago: The University of Chicago Press, 1976.

JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. 6ª ed. New Jersey: Pearson, 2007.

MANSO, J. R. P.; SIMÕES, N.M. **Indicador sintético de desenvolvimento económico e social ou de bem-estar dos municípios do continente português**, Universidade da Beira Interior, Observatório para o Desenvolvimento Económico e Social, 2009. Acessado em 13/11/2017. Disponível em: [https://www.researchgate.net/profile/Jose\\_Manso/publication/267546116\\_INDICADOR\\_SINTETICO\\_DE\\_DESENVOLVIMENTO\\_ECONOMICO\\_E\\_SOCIAL\\_OU\\_DE\\_BEM-ESTAR\\_DOS\\_MUNICIPIOS\\_DO\\_CONTINENTE\\_PORTUGUES/links/5469cba00cf2397f782d9d0e/INDICADOR-SINTETICO-DE-DESENVOLVIMENTO-ECONOMICO-E-SOCIAL-OU-DE-BEM-ESTAR-DOS-MUNICIPIOS-DO-CONTINENTE-PORTUGUES.pdf](https://www.researchgate.net/profile/Jose_Manso/publication/267546116_INDICADOR_SINTETICO_DE_DESENVOLVIMENTO_ECONOMICO_E_SOCIAL_OU_DE_BEM-ESTAR_DOS_MUNICIPIOS_DO_CONTINENTE_PORTUGUES/links/5469cba00cf2397f782d9d0e/INDICADOR-SINTETICO-DE-DESENVOLVIMENTO-ECONOMICO-E-SOCIAL-OU-DE-BEM-ESTAR-DOS-MUNICIPIOS-DO-CONTINENTE-PORTUGUES.pdf)

MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada: uma abordagem prática**. Belo Horizonte: Editora UFMG, 2005.

PALLANT, J. **SPSS Survival Manual: a step-by-step guide to data analysis using SPSS para Windows**. 3ª ed. Open University Press, England, 2007.

PEREIRA, J. C. R. **Análise de dados qualitativos: estratégias metodológicas para as ciências da saúde, humanas e sociais.** São Paulo: Edusp, 2001.

RENCHEER, A. C.; CHRISTENSEN, W. F. **Methods of multivariate analysis.** 3<sup>a</sup> ed. Department of Statistics, Brigham Young University, Provo, UT., 2012.

SHARMA, S. **Applied Multivariate Techniques.** John Wiley & Sons, Inc., United States of American, 1996.

TEIXEIRA, F. R. F.; NASCIMENTO, M.; NASCIMENTO, A. C. C.; PAIXÃO, D. M.; AZEVEDO, C. F.; SILVA, F. F.; CRUZ, C. D.; LOPES, P. S.; GUIMARÃES, S. E. F. Determinação de fatores em características de suínos, **Revista Brasileira de Biometria**, São Paulo, v. 33, n. 2, p. 130-138, 2015.

WILLIAMS, B., BROWN, T., ONSMAN, A. Exploratory factor analysis: A five-step guide for novices. **Australasian Journal of Paramedicine**, v. 8, n. 3, p. 1-13, 2010. (Disponível em: <https://ajp.paramedics.org/index.php/ajp/article/viewFile/93/90>).