

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA
DEPARTAMENTO DE ESTATÍSTICA**

**PROPOSTA DE UM ÍNDICE DE AVALIAÇÃO DO POTENCIAL
DOS DISTRITOS OPERACIONAIS DO RIO GRANDE DO SUL
ATRAVÉS DE ANÁLISE FATORIAL**

Eduardo Ramos

Porto Alegre, 2004

Este Trabalho de Conclusão foi analisado e julgado adequado para a obtenção do título de Bacharel em Estatística e aprovada em sua forma final pelo orientador e pelo Departamento de Estatística, Universidade Federal do Rio Grande do Sul.

Prof.^a Márcia Echeveste

Orientador

Departamento de Estatística

Universidade Federal do Rio Grande do Sul

BANCA EXAMINADORA

Prof.^a Luciana Neves Nunes

Departamento de Estatística - UFRGS

*"Há homens que lutam um dia e são bons.
Há outros que lutam um ano e são melhores.
Há os que lutam muitos anos e são muito bons.
Porém, há os que lutam toda a vida.
Esses são os imprescindíveis."*

Bertolt Brecht

AGRADECIMENTOS

Gostaria de agradecer à minha orientadora Márcia E. Echeveste e aos demais professores do Departamento de Estatística da UFRGS.

Agradeço aos meus colegas e amigos da UFRGS que estiveram comigo durante estes longos anos em que estive no curso.

Agradeço a minha namorada Vivian de B. Formiga por ter sempre estado do meu lado e me apoiado durante todo este período.

E, acima de tudo, agradeço aos meus pais, Adroaldo R. Ramos e Maria E. Ramos, que sempre me incentivaram a estudar e nunca me deixaram desistir.

ÍNDICE

1	COMENTÁRIOS INICIAIS.....	11
1.1	INTRODUÇÃO.....	11
1.2	TEMA.....	13
1.3	OBJETIVOS.....	13
1.4	JUSTIFICATIVA DO TEMA.....	13
1.5	MÉTODO DE TRABALHO.....	14
1.6	ESTRUTURA.....	15
1.7	DELIMITAÇÕES DO TRABALHO.....	16
2	REVISÃO BIBLIOGRÁFICA.....	17
2.1	INTRODUÇÃO.....	17
2.2	DEFINIÇÃO DO OBJETIVO DO ESTUDO.....	17
2.3	ESCOLHA DAS VARIÁVEIS.....	18
2.3.1	<i>Padronização das variáveis.....</i>	18
2.4	ANÁLISE FATORIAL.....	19
2.4.1	<i>Esquema de uma Análise Fatorial.....</i>	21
2.4.2	<i>Suposições para uso da Análise Fatorial.....</i>	22
2.4.3	<i>CrITÉRIOS para o número de fatores a serem extraídos.....</i>	24
2.4.3.1	CrITÉrio da Raiz Latente.....	24
2.4.3.2	CrITÉrio a Priori.....	24
2.4.3.3	CrITÉrio do teste Scree.....	25
2.4.4	<i>Estimação dos Escores Fatoriais.....</i>	26
2.4.5	<i>Rotação dos Fatores.....</i>	27
2.4.5.1	O conceito de estrutura simples.....	29
2.4.5.2	Rotação Quartimax.....	30
2.4.5.3	Rotação Varimax.....	30
2.4.5.4	Rotação Equimax.....	30
2.4.5.5	Rotação Oblimin.....	30
2.4.5.6	Rotação Promax.....	31
2.4.6	<i>Validação da Análise Fatorial.....</i>	31
2.5	CONSTRUÇÃO DOS ÍNDICES.....	31
2.5.1	<i>Índices de Desenvolvimento.....</i>	32
2.5.2	<i>Índices de Hierarquização.....</i>	34
2.5.3	<i>Propriedades dos índices.....</i>	35
2.6	DISCUSSÃO DO CAPÍTULO.....	35
3	UM EXEMPLO DE APLICAÇÃO.....	36

3.1	APRESENTAÇÃO DO PROBLEMA.....	36
3.2	OBJETIVOS DO ESTUDO.....	37
3.3	ESCOLHA DAS VARIÁVEIS.....	37
3.3.1	<i>Padronização das variáveis.....</i>	40
3.3.2	<i>Descrição das variáveis.....</i>	40
3.4	ANÁLISE FATORIAL.....	41
3.4.1	<i>Suposições para uso da análise fatorial.....</i>	41
3.4.2	<i>Número de fatores a serem extraídos.....</i>	43
3.4.3	<i>Estimação dos Escores Fatoriais.....</i>	44
3.4.4	<i>Rotação dos Fatores.....</i>	49
3.4.5	<i>Validação da análise fatorial.....</i>	51
3.5	CONSTRUÇÃO DOS ÍNDICES.....	52
3.5.1	<i>Índices de Desenvolvimento.....</i>	52
3.5.1.1	<i>Utilizando variáveis por Ranks.....</i>	53
3.5.1.2	<i>Utilizando variáveis padronizadas pela Normal.....</i>	55
3.5.2	<i>Índices de Hierarquização.....</i>	57
3.5.2.1	<i>Utilizando variáveis por Ranks.....</i>	58
3.5.2.2	<i>Utilizando variáveis padronizadas pela Normal.....</i>	59
3.5.3	<i>Propriedades dos índices.....</i>	60
3.5.3.1	<i>Modelo de Desenvolvimento.....</i>	60
3.5.3.2	<i>Modelo de Hierarquização.....</i>	60
3.5.4	<i>Diferenças entre os índices propostos.....</i>	61
3.5.5	<i>Correlações entre os índices e as variáveis.....</i>	62
3.6	DISCUSSÃO DOS RESULTADOS.....	65
4	COMENTÁRIOS FINAIS.....	68
4.1	CONCLUSÕES.....	68
4.2	SUGESTÕES PARA TRABALHOS FUTUROS.....	69
	REFERÊNCIAS BIBLIOGRÁFICAS.....	70

LISTA DE FIGURAS

Figura 1 Gráfico dos autovalores para o critério do Teste Scree.....	25
Figura 2 Rotação Ortogonal de fatores.....	28
Figura 3 Rotação Oblíqua de fatores.....	28
Figura 4 Gráfico dos autovalores na utilização da Análise Fatorial para a elaboração do índice.....	43

LISTA DE TABELAS

Tabela 1 Exemplo de pesos no índice de desenvolvimento	33
Tabela 2 Descrição das variáveis e fonte de dados por município	38
Tabela 3 Descrição das variáveis e fonte de dados por DOP	39
Tabela 4 Descrição dos DOPs atuais	39
Tabela 5 Medidas-resumo das variáveis consideradas	40
Tabela 6 Teste KMO e Teste de Bartlett para dados organizados por ranks	42
Tabela 7 Medida KMO e Teste de Bartlett para dados padronizados pela Normal	42
Tabela 8 Escores gerados pelos métodos de Regressão e Bartlett para dados “rankeados”	45
Tabela 9 Escores gerados pelos métodos de Regressão e Bartlett para dados padronizados pela Normal	47
Tabela 10 Fatores gerados a partir das rotações ortogonais	50
Tabela 11 Pesos no índice para variáveis rankeadas, nas rotações Varimax e Equimax	53
Tabela 12 Pesos no índice para variáveis rankeadas, na rotação Quartimax	54
Tabela 13 Índice de Desenvolvimento para variáveis organizadas por Ranks	55
Tabela 14 Pesos no índice para variáveis padronizadas pela Normal, nas três rotações	56
Tabela 15 Índice de Desenvolvimento para variáveis padronizadas pela Normal	57
Tabela 16 Índice de Hierarquização para variáveis organizadas por Ranks	58
Tabela 17 Índice de Hierarquização para variáveis padronizadas pela Normal	59
Tabela 18 Descritivas dos índices gerados através do modelo de desenvolvimento	61
Tabela 19 Descritivas dos índices gerados através do modelo de hierarquização	62
Tabela 20 Correlação entre as variáveis e os índices do modelo de desenvolvimento	63
Tabela 21 Correlação entre as variáveis e os índices do modelo de hierarquização	64
Tabela 22 Classificação das variáveis de acordo com especialistas	65
Tabela 23 Propostas para um índice de potencialidade dos DOPs do RS	66

RESUMO

Esta monografia aborda a construção de índices através da análise fatorial. Análise Fatorial é uma técnica de análise que avalia a relação de várias variáveis, cujo propósito é condensar a informação contida nas variáveis originais, reduzindo a dimensionalidade e formando fatores comuns com uma perda mínima de informação. Para a realização de uma análise fatorial, é necessário escolher o método de estimação dos escores fatoriais e o tipo de rotação dos fatores. Da mesma forma, para a elaboração de alguns índices é necessária a utilização de várias variáveis, portanto, a análise fatorial mostra-se uma técnica capaz de fornecer este índice capturando a informação das variáveis que irão compor o índice. Essa monografia apresenta alternativas de índices resultantes de diferentes combinações entre os métodos de estimação dos escores fatoriais e dos métodos de rotação dos fatores na elaboração de um índice potencial. Para ilustrar estas diferenças, é apresentado um exemplo de aplicação que envolve a criação de um índice que mede a potencialidade para os distritos operacionais do Rio Grande do Sul utilizando alguns métodos de estimação dos escores fatoriais e métodos de rotação dos fatores.

ABSTRACT

This monograph approaches the construction of indexes through the factor analysis. Factor Analysis is an analysis technique that evaluates the relationship of several variables, whose purpose is to condense the information contained in the original variables, reducing the dimensionality and forming factors common with a minimum loss of information. For the accomplishment of a factor analysis, it is necessary to choose the method of estimate of the factorial scores and the type of rotation of the factors. In the same way, for the elaboration of some indexes are necessary the use of several variables, therefore, the factor analysis is shown a technique capable to supply this index capturing the information of the variables that will compose the index. This monograph presents alternatives of resulting indexes of different combinations among the methods of estimate of the factorial scores and of the methods of rotation of the factors in the elaboration of a potential index. To illustrate these differences, it is presented an application example that involves the creation of an index that measures the potentiality for the operational districts of Rio Grande do Sul using some methods of estimate of the factorial scores and methods of rotation of the factors.

1 COMENTÁRIOS INICIAIS

1.1 Introdução

O estado do Rio Grande do Sul possui 397 municípios, e segundo classificação do Departamento Autônomo de Estradas de Rodagem (DAER), esses municípios podem ser classificados dentro de 17 Distritos Operacionais (DOPs), de acordo com seus limites geográficos.

Estes Distritos Operacionais apresentam características diferenciadas, pois enquanto um DOP pode ser considerado extremamente industrializado, algum outro pode não possuir indústrias, por exemplo.

O estudo aqui contextualizado trata-se da avaliação do uso das faixas de domínio dos DOPs. Entende-se por faixas de domínio a extensão de terra entre a estrada e o município que a abriga.

As faixas de domínio são de gerência, atualmente, dos DOPs. O seu uso pode ser, em linhas gerais: uso comercial (quiosques, *outdoors*, etc.) ou uso industrial (passagem de fibra ótica, gasoduto, telefone, etc.).

O uso comercial ou industrial identificados em cada faixa de domínio é determinado pelo potencial (valor de exploração) de cada faixa pertencente a um determinado DOP.

No entanto, sabe-se que os DOPs não tem o mesmo valor comercial. Por isso, deseja-se criar um índice único que avalie o potencial de cada região.

Esse índice, futuramente, auxiliará a estabelecer um preço para o uso das faixas de domínio que pertencem ao DOP.

Dado o exposto acima, constatou-se uma oportunidade de elaborar um índice que represente a potencialidade de cada Distrito Operacional, e uma das utilizações deste índice poderia ser de estabelecer diferenças na comercialização de um produto ou na implantação de qualquer outro tipo de serviço.

A denominação de índice pode ter vários significados. Em Economia, há índices de preços, quantidades e valor dos bens, de custo de vida, de (des)emprego, de bolsas de valores, de concentração dos mercados, de monopólio de empresas, de importação e exportação. Em Administração, índices de produção, de liquidez, velocidade de vendas; em Administração Pública, diversos índices permitem avaliar a qualidade de vida, a permanência ou evasão escolar, o nível de criminalidade e o padrão de saúde das populações. Na Engenharia há índices de desempenho, de motivação; na área da educação existem índices que medem o quociente de inteligência, coeficiente de aprovação, etc.

No Brasil existem diversos números índices conhecidos, como Índice Nacional de Preços ao Consumidor (INPC), Índice Geral de Preços (IGP), Índice da Bolsa de Valores de São Paulo (Ibovespa), entre vários outros, e suas formas de obtenção são muito variadas, e uma destas formas de obtenção é através do uso de técnicas estatísticas multivariadas, entre elas a Análise Fatorial.

Segundo Mulaik (1972), a Análise Fatorial é uma técnica que trata das relações internas de um conjunto de variáveis, substituindo um conjunto inicial de variáveis correlacionadas por um conjunto menor de fatores (ou variáveis hipotéticas) que podem ser não correlacionadas ou correlacionadas, e que explicam a maior parte da variância do conjunto original.

Spearman (1904) publicou o primeiro trabalho sobre análise fatorial, limitado a busca de um único fator. Sua extensão para vários fatores se deve a Thurstone (1941) que publicou um artigo intitulado "*Factorial Studies on Intelligence*".

Na realização de uma análise fatorial é necessário realizar a estimação dos escores fatoriais, que são o valor de cada fator para cada observação, e também se deve realizar uma rotação na matriz de cargas fatoriais, que é simplesmente multiplicá-la por uma matriz ortogonal, para proporcionar uma melhor interpretação dos fatores comuns.

Análise fatorial é uma ferramenta muito poderosa que auxilia a resolução de problemas

em muitas áreas do conhecimento, como por exemplo, em Administração, Engenharia de Produção, Economia. Ela pode ser utilizada em diversos tipos de pesquisas, como pesquisa de marketing, de satisfação, de importância.

1.2 Tema

Esse trabalho tem como tema a aplicação da análise fatorial para construir um índice que represente a potencialidade de cada DOP do Rio Grande do Sul. Esse estudo abrange um levantamento de referências teóricas sobre os assuntos mencionados, assim como a aplicação em dados reais.

O tema desenvolvido nesse trabalho torna-se relevante pela crescente necessidade de desenvolvimento de índices e valores que condensem informação.

1.3 Objetivos

O objetivo principal desse trabalho é estudar e comparar as formas de aplicação dos diferentes métodos de estimação dos escores fatoriais e dos métodos de rotação dos fatores na análise fatorial.

Como objetivo secundário desse trabalho, tem-se a aplicação das técnicas abordadas nessa monografia em um exemplo de aplicação na construção de um índice que demonstre a potencialidade de cada DOP do Rio Grande do Sul, através da análise fatorial.

1.4 Justificativa do tema

De acordo com Endo (1988), índices têm sua importância pelo fato de que é muito mais agradável monitorar diversas variáveis através de uma única medida, e não através de vários valores.

A criação destes índices, conforme Fachel (1976), pode ser realizada através da utilização de análise fatorial quando há um grande número de variáveis necessário para obtê-los, onde todas as variáveis são importantes.

Swisher, Beckstead e Bebeau (2004) afirmam que a análise fatorial pode ser usada para aumentar e clarear o discernimento do pesquisador na realização de seus projetos de pesquisa quando há o envolvimento de diversas variáveis.

Por essa razão, acredita-se que esse trabalho auxiliará no uso da análise fatorial a profissionais da área de estatística e de outras áreas interessadas neste tipo de aplicação.

1.5 Método de trabalho

Essa seção descreve a metodologia desenvolvida neste trabalho.

Esta é uma pesquisa de natureza aplicada, pois tem como objetivo gerar conhecimentos para aplicação prática dirigidos à solução de problemas específicos, envolvendo verdades e interesses locais.

Quanto à abordagem, esta pesquisa pode ser considerada quantitativa, pois considera que tudo pode ser quantificável, o que significa traduzir em números opiniões e informações para classificá-los e analisá-los.

Do ponto de vista dos procedimentos técnicos, esta pesquisa pode ser classificada como bibliográfica, pois foi elaborada a partir de material já publicado, constituído principalmente de livros e material disponibilizado na internet.

Esta pesquisa será aplicada em um banco de dados organizado a partir de variáveis de interesse para o estudo.

Os passos percorridos para a elaboração desta monografia foram:

- Pesquisa bibliográfica sobre análise fatorial e construção de índices;
- Pesquisa sobre o uso de diferentes abordagens de estimação e seus reflexos na análise fatorial;
- Coleta de dados;
- Construção de alternativas de índices;

- Construção de mecanismos de comparação entre os índices;
- Estabelecimento das conclusões da pesquisa.

1.6 Estrutura

O trabalho está dividido em quatro capítulos principais.

O capítulo 1 trata dos comentários iniciais, subdivididos em sete seções que compreendem os tópicos: introdução, tema, objetivos, justificativa do tema e objetivos, método de trabalho, estrutura, e limitações.

O capítulo 2 consiste de uma revisão bibliográfica, contendo um resumo das técnicas estatísticas utilizadas para a resolução do problema em questão e dos resultados obtidos de estudos de diversos autores. Essa é a base teórica do assunto que está sendo estudado neste trabalho. O item 2.1 apresenta uma introdução ao assunto do uso de análise fatorial para a criação de números-índices. No item 2.2 estão os objetivos do estudo. O item 2.3 mostra como se realiza a escolha das variáveis e também apresenta dois tipos de padronização das mesmas. O item 2.4 apresenta o esquema de uma análise fatorial, com seus objetivos, suposições e alguns critérios para o seu funcionamento. Na seção 2.5 são apresentados os números-índices, mostrando suas propriedades e uma ligeira explicação sobre cada um, sendo os mais importantes para a aplicação da análise fatorial o índice de desenvolvimento e o índice de hierarquização. E na seção 2.6 é realizada uma discussão em cima do que foi visto em todo o capítulo.

O capítulo 3 é composto por um exemplo de aplicação que consiste em averiguar a potencialidade de cada distrito operacional do estado do Rio Grande do Sul, elaborando para esse fim, através da análise fatorial, um índice. Nesta parte são aplicadas as técnicas revisadas no capítulo 2 deste trabalho.

O capítulo 4 trata das considerações finais, conclusões, recomendações e sugestões para trabalhos futuros.

1.7 Delimitações do Trabalho

Essa monografia restringe-se ao estudo da aplicação de análise fatorial para a criação de um índice que avalie a potencialidade das regiões do Rio Grande do Sul. Outras técnicas poderiam ser úteis para criar índices; esta monografia trata da comparação dos itens provenientes da criação de índices através da análise fatorial. Não se discute índices de desempenho estimados através de outras abordagens.

Não faz parte do objetivo deste trabalho aprofundar os conhecimentos em aspectos da análise fatorial, e sim o suficiente para solucionar o problema em questão.

2 REVISÃO BIBLIOGRÁFICA

2.1 Introdução

Em diversas áreas do conhecimento, o pesquisador quase sempre se depara com estudos que envolvem várias variáveis. Se o número de variáveis for relativamente grande, o pesquisador pode estar interessado em encontrar uma maneira de condensar a informação contida nas variáveis, reduzindo a dimensionalidade e formando fatores com uma perda mínima de informação. Este é o objetivo da técnica estatística multivariada conhecida como Análise Fatorial.

Para a formação de índices que avaliem uma população com várias variáveis, é possível a utilização de algumas técnicas estatísticas multivariadas, e uma destas técnicas a serem utilizadas pode vir a ser a análise fatorial. Posteriormente, serão identificados propriedades e critérios na avaliação dos itens criados pela análise fatorial.

2.2 Definição do Objetivo do estudo

O primeiro passo para o estudo é a definição clara dos objetivos de pesquisa. A definição dos objetivos deve estar alinhada ao problema de pesquisa. É neste momento que podem ser definidas as variáveis que serão necessárias para atingir o objetivo e elucidar o problema.

O entendimento do objetivo é fundamental para a solução do problema de pesquisa e

deve servir de guia para o estudo.

2.3 Escolha das variáveis

Para a realização de uma análise fatorial, devem-se escolher variáveis que possuem medidas métricas. A análise fatorial só será possível se todas as variáveis forem quantitativas, não sendo possível realizá-la quando as variáveis forem qualitativas.

Na criação de um índice, todas as variáveis que forem escolhidas devem possuir um vínculo com o que este índice irá representar. Por exemplo, se um índice mede grau de urbanização, deve possuir variáveis relacionadas a este índice, tais como área urbanizada, extensão de rodovias com pavimentação, número de indústrias, entre outras, e não deve ser incluída nenhuma variável que nada represente para ele, como altura ou peso das pessoas.

2.3.1 Padronização das variáveis

Quando as variáveis possuem grandezas diferentes, é necessário padronizá-las para a realização da análise fatorial, gerando assim resultados, na análise, para variáveis na mesma escala de medida.

Um método de padronização de variáveis é através dos *ranks*, o qual consiste em organizar as variáveis originais de acordo com os seus tamanhos e atribuir um número de 1 a N, sendo N o tamanho da amostra.

Outro método bastante divulgado de padronização das variáveis é através da distribuição normal, que é feito pegando-se o valor de um caso, diminuindo-se da média da variável e dividindo-se pelo seu desvio-padrão, conforme é mostrado na equação 1.

$$Z = \frac{X - \mu}{\sigma} \quad (1)$$

Onde:

μ = média da variável

σ = desvio-padrão da variável

2.4 **Análise Fatorial**

Análise fatorial é uma técnica de interdependência na qual todas as variáveis são simultaneamente consideradas. Conforme Possoli (1992), esta técnica analisa a inter-relação de um grande número de variáveis e então explica estas variáveis em termos das suas dimensões comuns (fatores), que podem ser não correlacionados (fatores ortogonais) ou correlacionados (fatores oblíquos).

Esta técnica tem como objetivo principal a redução da dimensionalidade das variáveis e sumarização dos dados. A análise fatorial converte um conjunto de variáveis inter-relacionadas em grupos menores, chamados de fatores os quais são interdependentes (independentes entre si).

De acordo com Bravo (1994), a análise fatorial pode identificar a estrutura de relações entre as variáveis examinando as correlações entre as variáveis.

Além disso, pode identificar variáveis representativas de um arranjo muito grande de variáveis para usar em análises multivariadas subseqüentes ou criar um arranjo completamente novo de variáveis, muito menor, para parcialmente ou completamente substituir o arranjo inicial de variáveis para inclusão em técnicas subseqüentes. Em ambas as instâncias o objetivo é manter a natureza e características das variáveis originais, mas reduzir seu número para simplificar as subseqüentes análises multivariadas.

Segundo Kendall (1950), as técnicas de análise multivariada podem ser divididas em dois tipos: quando o objetivo é estudar a dependência de uma ou mais variáveis em relação às outras, e quando há interesse em estudar as relações de um conjunto de variáveis entre si. Ao primeiro tipo dá-se o nome de análise de dependência, e ao segundo tipo dá-se o nome de análise de interdependência, onde está contida a análise fatorial.

A técnica de análise fatorial apareceu pela primeira vez em 1904 através de um artigo publicado por Spearman intitulado “*General Intelligence Objectively Determined and*

Measured?. O modelo proposto por Spearman pode ser considerado uma primeira aproximação, pois apresentava apenas um fator comum. Pelo fato de possuir apenas um fator comum, Bartlett (1962) considerou este modelo inadequado, sugerindo que mais fatores comuns fossem introduzidos ao modelo, surgindo assim o conceito de fatores múltiplos.

Matematicamente, a análise fatorial é semelhante à análise de regressão múltipla, onde cada variável pode ser expressa como uma combinação linear dos fatores. (JOHNSON, 1988).

Dado um conjunto X_1, X_2, \dots, X_i variáveis aleatórias. Cada variável no modelo de análise fatorial pode ser descrita conforme a equação 2:

$$X_i = A_{i1}F_1 + A_{i2}F_2 + A_{i3}F_3 + \dots + A_{im}F_m + e_i \quad (2)$$

Onde:

A_{ij} = carga fatorial da i -ésima variável no j -ésimo fator comum. Reflete a importância do j -ésimo fator na composição da i -ésima variável.

F_m = Fator comum (pois está presente em todas as variáveis X_i)

e_i = Fatores específicos que descrevem a variação residual específica da i -ésima variável (resíduo que afeta somente X_i). É a parte da variável X_i que não é explicada pelos fatores comuns F.

Os fatores comuns podem ser não correlacionados entre si (ortogonais), ou seja $Cov(F_m, F_{m'}) = 0$, ou correlacionados entre si (oblíquos), onde $Cov(F_m, F_{m'}) \neq 0$. Estes fatores comuns podem ser expressos como combinações lineares de variáveis observáveis (HAIR, 1995).

$$F_i = W_{i1}X_1 + W_{i2}X_2 + W_{i3}X_3 + \dots + W_{ik}X_k \quad (3)$$

Onde:

F = estimativa do i -ésimo fator

W_i = peso ou coeficiente do escore fatorial

K = número de fatores

Supõe-se que os fatores específicos e_i são não correlacionados entre si e não correlacionados com os fatores comuns. A variância destes fatores específicos é denominada variância residual (ou especificidade), e é denotada por ω_i .

Sob a suposição de fatores não correlacionados, a variância das variáveis observadas pode ser descrita pela equação 4:

$$\sigma_i^2 = \sum_{j=1}^n \lambda_{ij}^2 + \omega_i \quad (4)$$

A parte da variância da variável X_i que é explicada só pelos fatores comuns é chamada de comunalidade da variável X_i , e é dada pela equação 5:

$$\sum_{j=1}^n \lambda_{ij}^2 = \sigma_i^2 - \omega_i \quad (5)$$

Os parâmetros deste modelo a serem estimados na prática são as cargas fatoriais e as especificidades, mas se os fatores comuns são supostos correlacionados, os coeficientes de correlação também devem ser estimados.

E finalmente para uma melhor interpretação dos fatores comuns, o pesquisador deve realizar a rotação dos fatores após obter as cargas fatoriais.

2.4.1 Esquema de uma Análise Fatorial

O cálculo de uma análise fatorial envolve três decisões básicas: cálculo da matriz de correlação para conhecer os objetivos especificados por agrupar as variáveis, o projeto de estudo em termos de número de variáveis e propriedades de medição das variáveis, e por

último o tamanho de amostra necessário (HAIR, 1995).

Variáveis que compõem uma análise fatorial possuem normalmente medidas métricas. Em alguns casos, podem ser usadas variáveis *dummy* (codificadas com 0 ou 1), embora consideradas não métricas.

O pesquisador geralmente não deve realizar uma análise fatorial com uma amostra menor que 50 observações, e preferencialmente o tamanho da amostra deve ser de 100 observações ou mais (MULAİK, 1972).

Como uma regra geral, deve-se realizar a análise fatorial se o número de observações for no mínimo cinco vezes o número de variáveis a serem analisadas (CATTEL, 1977). Kim (1978) e Latif (1994) também propõem um mínimo de 20 observações para cada variável.

2.4.2 *Suposições para uso da Análise Fatorial*

Em uma análise fatorial é suposto que os fatores específicos são não-correlacionados entre si e também não-correlacionados com os fatores comuns, lembrando que os fatores comuns podem ser correlacionados entre si (fatores oblíquos) ou não-correlacionados entre si (fatores ortogonais).

Assume-se também que os fatores comuns e os fatores específicos possuem distribuição Normal Multivariada, o que implica dizer que os vetores de variáveis originais tenham distribuição Normal Multivariada. Para a verificação desta suposição, é necessário realizar um teste de aderência.

As hipóteses para a verificação são as seguintes:

H_0 : Os vetores das variáveis originais possuem distribuição Normal Multivariada

H_1 : Os vetores das variáveis não seguem uma distribuição Normal Multivariada

A estatística de teste utilizada encontra-se na equação 6:

$$D^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \quad (6)$$

Onde:

O_i : valor observado

E_i : valor esperado

De acordo com Meyer (1983), se n for suficientemente grande, a distribuição de D^2 tem, aproximadamente, uma distribuição qui-quadrado, com $(k-1)$ graus de liberdade.

Rejeita-se H_0 quando $D^2 > C$, onde C é obtido das tábuas de distribuição qui-quadrado, isto é, $C = \chi_{k-1}^2$.

Assim, deve-se obter cada um dos parâmetros da distribuição Normal Multivariada $(\mu_1, \mu_2, \dots, \mu_n, \sigma_1, \sigma_2, \dots, \sigma_n)$, encontrar a estatística de teste e compará-la com a distribuição qui-quadrado para então rejeitar ou aceitar a hipótese nula.

Existem ainda duas maneiras de verificar qual o grau de ajuste dos dados a análise fatorial, isto é, qual é o nível de confiança que se pode esperar dos dados quando do seu tratamento pelo método multivariado de análise fatorial.

A primeira é obter a Medida Kayser-Meyer-Olkin (KMO), que apresenta valores normalizados (entre 0 e 1) e mostra qual é a proporção de variância que as variáveis apresentam em comum ou a proporção desta que são devidas a fatores comuns. Para a interpretação do resultado obtido, valores próximos de 1 indicam que o método de análise fatorial é perfeitamente adequado para o tratamento dos dados. Por outro lado, valores menores que 0,5 indicam a inadequação do método (PEREIRA, 2001).

A segunda maneira é realizar o Teste de Esfericidade de Bartlett, que é baseado na distribuição estatística qui-quadrado e testa a hipótese nula H_0 de que a matriz de correlação é

uma matriz identidade (cuja diagonal é 1 e todas as outras iguais a zero), isto é, que não há correlação entre as variáveis (PEREIRA, 2001).

2.4.3 Critérios para o número de fatores a serem extraídos

Como decidir o número de fatores a serem extraídos? Na decisão do número de fatores extraídos, o pesquisador normalmente utiliza algum critério pré-determinado, como o critério da raiz latente ou o critério a priori, para alcançar um número específico de fatores a serem extraídos.

2.4.3.1 Critério da Raiz Latente

A técnica mais utilizada para determinar o número de fatores a serem extraídos é o critério da raiz latente, também conhecido como critério do autovalor. Esta técnica é simples de aplicar. Cada variável contribui com um valor 1 para o autovalor total. Por conseguinte, somente os fatores que possuem autovalor maior que 1 são considerados significantes; todos os fatores com um autovalor menor que 1 são considerados insignificantes e são descartados.

Digamos que A é uma matriz quadrada $k \times k$ e I é uma matriz identidade $k \times k$. Então os escalares $\lambda_1, \lambda_2, \dots, \lambda_k$ que satisfazem a equação 7 são chamados de autovalores da matriz A . A equação 7 é chamada de equação característica.

$$|A - \lambda I| = 0 \quad (7)$$

2.4.3.2 Critério a Priori

Antes de ser realizada a análise fatorial o pesquisador já sabe quantos fatores deve extrair. O pesquisador simplesmente instrui o computador para encerrar a análise quando o número desejável de fatores for alcançado.

2.4.3.3 Critério do teste Scree

Todos os fatores possuem alguma variância única, e a proporção de variância única é substancialmente maior nos últimos do que nos primeiros fatores (KIM e MUELLER, 1978). O teste Scree é utilizado para identificar o número ótimo de fatores que podem ser extraídos antes que o total de variância única comece a dominar a estrutura da variância comum. Este teste é realizado colocando os autovalores em um gráfico contra o número de fatores na sua ordem de extração, e o formato desta curva é utilizado para observar o ponto de corte. A figura 1 mostra um gráfico de autovalores com os primeiros 20 fatores extraídos de um estudo hipotético. Começando pelo primeiro fator, a curva começa a decrescer e depois lentamente começa a se transformar em uma linha horizontal. O ponto onde a curva começa a virar uma linha horizontal indica o número máximo de fatores a serem extraídos. Nesta figura, o número máximo de fatores a serem extraídos é dez.

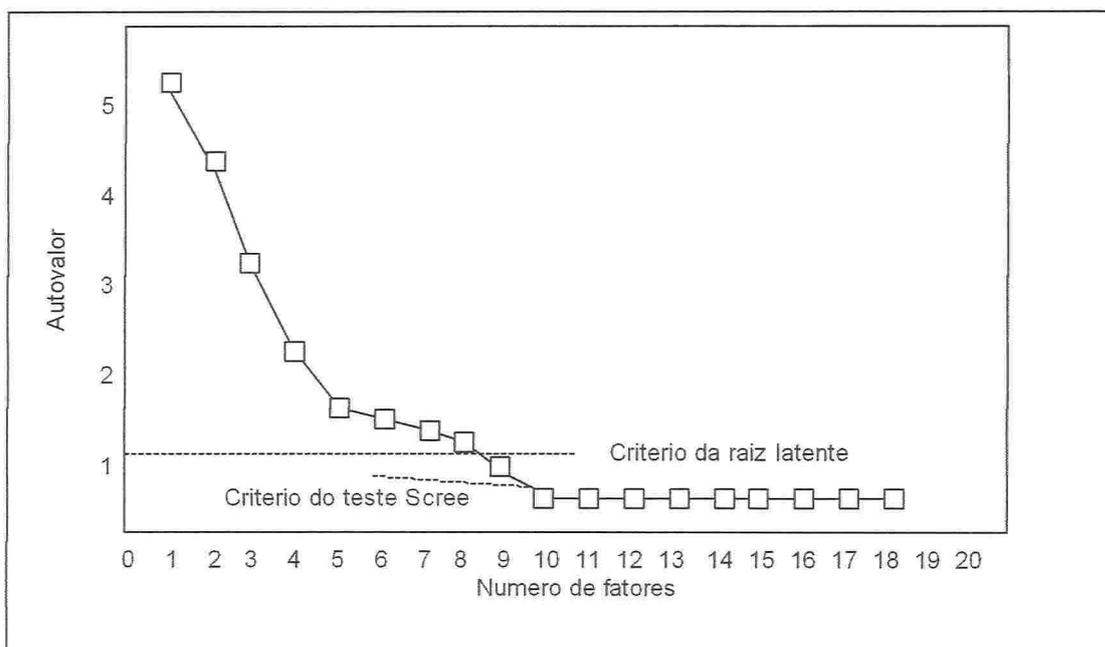


Figura 1 Gráfico dos autovalores para o critério do Teste Scree

No gráfico acima se pode notar também que pelo critério da raiz latente, o número de fatores a serem extraídos é igual a oito.

2.4.4 Estimação dos Escores Fatoriais

Em algumas aplicações, como por exemplo, quando a análise fatorial é preliminar a algum outro tipo de análise multivariada ou quando é utilizada para a construção de índices, é necessário tentar descrever os fatores em termos das variáveis observadas. Feito isto, estima-se o valor de cada fator para cada observação. Estes valores são chamados de Escores Fatoriais.

Existem dois principais métodos para a obtenção das estimativas dos escores fatoriais, método de regressão e o método de Bartlett.

O método de regressão foi desenvolvido por Thompsom (1961), e é baseado na regressão dos vetores dos fatores comuns sobre os vetores das variáveis originais. Este método é equivalente a encontrar para cada j , $j = 1, 2, \dots, m$, uma função linear das observações que dará um bom preditor de F_j dado pela equação 8:

$$\hat{F}_j = \underline{X}' \underline{b}_j \quad (8)$$

Onde:

\underline{X}' : vetor das variáveis originais transposto

\underline{b}_j : vetor dos coeficientes dos escores fatoriais

O vetor dos escores fatoriais é escolhido de tal forma que $(\hat{F}_j - F_j)$ seja mínimo.

No método de Bartlett os coeficientes dos escores fatoriais são obtidos de tal forma que a soma dos resíduos padronizados seja mínima em relação aos elementos do vetor dos fatores comuns. A expressão para obter-se a matriz dos coeficientes dos escores fatoriais tanto para o caso de fatores ortogonais como para o caso de fatores oblíquos é dada pela equação 9:

$$\hat{F}_j = \hat{B} \underline{X} \quad (9)$$

Onde $\hat{\underline{B}}$ é o vetor dos coeficientes dos escores fatoriais.

Lawley e Maxwell (1971) demonstraram que os estimadores obtidos pelo método de regressão são viciados, enquanto que os obtidos pelo método de Bartlett são não viciados. Entretanto, os estimadores pelo método de regressão possuem menor variância do que os estimadores do método de Bartlett.

2.4.5 Rotação dos Fatores

Para proporcionar uma melhor interpretação dos fatores comuns, faz-se uma rotação na matriz de cargas fatoriais, fazendo-a mais identificável com a natureza das variáveis originais. De acordo com Hair (1995), rotação de uma matriz significa multiplicá-la por uma matriz ortogonal, sendo que esta rotação não é única, dado que a matriz ortogonal pode ser qualquer matriz que não apresente correlação entre os fatores.

O caso mais simples de rotação é a rotação ortogonal, onde os eixos são mantidos em 90 graus. Também é possível rotacionar os eixos e não manter um ângulo de 90 graus entre os eixos de referência. Quando não é ortogonal, o procedimento de rotação é chamado de rotação oblíqua.

Na prática, o objetivo de todos os métodos de rotação é simplificar as linhas e colunas da matriz fatorial para facilitar a interpretação. Em uma matriz fatorial, colunas representam fatores, com cada linha correspondendo a uma variável. Entende-se por “simplificar as linhas”, fabricar tantos valores em cada linha perto de zero quanto possível. Por “simplificar as colunas” entende-se fabricar tantos valores em cada coluna perto de zero quanto possível.

Rotações oblíquas são semelhantes a rotações ortogonais, exceto que rotações oblíquas permitem fatores correlacionados em vez de manter independência entre os fatores rotacionados. A rotação dos fatores Ortogonal e Oblíqua estão demonstradas nas figuras 2 e 3, respectivamente.

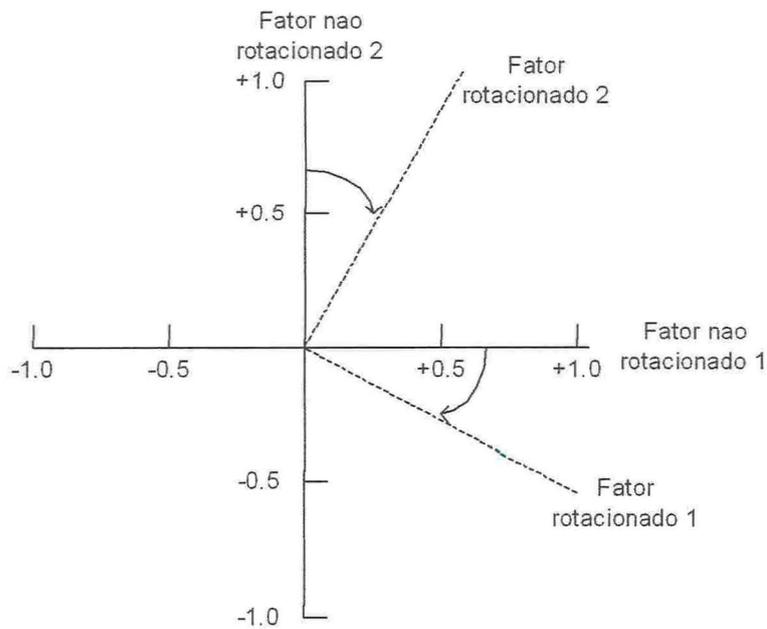


Figura 2 Rotação Ortogonal de fatores

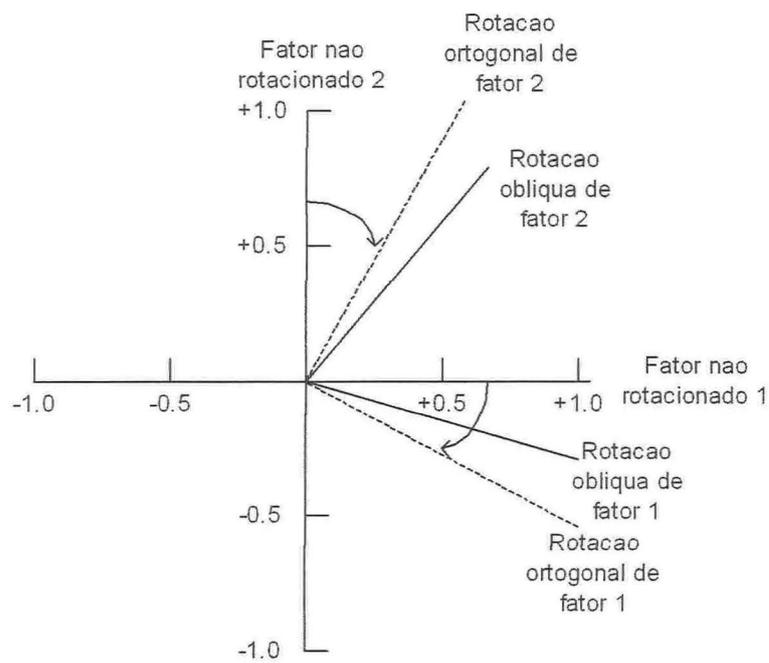


Figura 3 Rotação Oblíqua de fatores

2.4.5.1 *O conceito de estrutura simples*

O conceito de estrutura simples foi formulado por Thurstone (1945), que diz que muitas variáveis não dependeriam de todos os fatores, por isso as cargas fatoriais destes fatores deveriam ser nulas.

Dado, por exemplo, x fatores comuns e y variáveis aleatórias, a estrutura simples será obtida quando em cada linha houver somente um elemento distinto de zero e entre qualquer par de colunas os elementos distintos de zero não coincidirem.

Os princípios da estrutura simples foram empregados inicialmente nos métodos gráficos de transformação dos fatores, e por último têm sido utilizados em procedimentos analíticos. Estes métodos analíticos simplificam a estrutura das cargas fatoriais e fornecem soluções únicas. Pode-se aplicá-los tanto em rotações ortogonais quanto em rotações oblíquas dos fatores.

Os métodos mais conhecidos de rotação ortogonal são o Quartimax, Varimax, e Equimax, e embora a rotação Quartimax seja analiticamente mais simples que a rotação Varimax, esta apresenta uma separação mais clara dos fatores, e como a rotação Equimax não possui muita aceitação e é muito pouco usada, a rotação Varimax é a mais recomendada (HAIR, 1995). Os métodos mais difundidos de rotação oblíqua são o Oblimin e o Promax, sendo que normalmente se utiliza o método Oblimin, exceto os casos onde os bancos de dados são muito grandes, sendo recomendado então a utilização do método de rotação Promax.

Conforme Press (1972), a escolha de uma rotação ortogonal ou oblíqua deve ser feita com base nas necessidades particulares de um determinado problema de pesquisa. Se a meta da pesquisa é reduzir o número de variáveis originais, a solução apropriada seria uma rotação ortogonal. Ainda, se o investigador quer reduzir um número maior de variáveis para um número menor de variáveis não-correlacionadas para uso subsequente em regressão ou outras técnicas de predição, uma solução ortogonal é a melhor. Porém, se o objetivo final da análise fatorial é obter alguns fatores teoricamente significantes ou construtos, uma solução oblíqua é apropriada. Chega-se a esta conclusão porque, realisticamente, poucos fatores são não-correlacionados, como em uma rotação ortogonal.

2.4.5.2 Rotação *Quartimax*

Este método é equivalente à maximização da soma da quarta potência das cargas fatoriais. Por isso o nome *Quartimax*. Este critério foi desenvolvido inicialmente por Nenhaus e Wrigley (1954).

No método *Quartimax*, os grandes valores das cargas fatoriais tendem a se tornar maiores e os pequenos valores das cargas fatoriais tendem a ficar menores em valor absoluto, comparando-os com os tamanhos originais. Este método diz respeito à simplificação da descrição de cada linha (que corresponde as variáveis) da matriz de cargas fatoriais.

2.4.5.3 Rotação *Varimax*

Este método foi desenvolvido por Kaiser (1958), em contraposição ao método *Quartimax*, dando mais ênfase à simplificação das colunas (que corresponde aos fatores) da matriz de cargas fatoriais, tentando satisfazer os requisitos para a estrutura simples, se aproximando bastante dela.

No método *Varimax*, a variância atinge um máximo, e quando isso ocorre o fator tem maior simplicidade, no sentido de que as cargas deste fator tendem a zero.

2.4.5.4 Rotação *Equimax*

Este critério considera simultaneamente a simplificação das linhas e das colunas da matriz de cargas fatoriais, tentando minimizar a complexidade dos fatores e das variáveis simultaneamente.

2.4.5.5 Rotação *Oblimin*

É um método de rotação não ortogonal, onde a covariância entre os fatores é minimizada. Quando o seu parâmetro é igual a zero, as soluções são mais oblíquas. Quanto mais seu parâmetro vai ficando negativo, mais os fatores ficam menos oblíquos.

2.4.5.6 Rotação Promax

É um método de rotação oblíqua, que permite que os fatores estejam correlacionados. Ele pode ser calculado mais rapidamente que a rotação Oblimin, podendo ser usado em bancos de dados muito grandes.

2.4.6 Validação da Análise Fatorial

Um dos aspectos para validar o modelo de análise fatorial é verificar a estabilidade dos resultados. A estabilidade deste modelo é inicialmente dependente do tamanho da amostra utilizado e do número de casos por variável. O pesquisador deve sempre obter a maior amostra possível e desenvolver um modelo parcimonioso. Se o tamanho de amostra permitir, o pesquisador deve dividir aleatoriamente a amostra em duas e comparar os resultados das análises fatoriais nas duas novas amostras.

Outro ponto importante na validação da análise fatorial é a detecção de observações discrepantes. O pesquisador deve estimar o modelo de análise fatorial com e sem observações consideradas discrepantes, se existirem, para visualizar seus impactos nos resultados. Se a perda de uma observação considerada discrepante não for importante, deve-se retirá-la da amostra e realizar uma nova análise fatorial.

2.5 Construção dos índices

Conforme Fonseca (1985), número-índice é um quociente que expressa uma dada quantidade em comparação a uma quantidade base.

Denomina-se número-índice o quociente da variável em estudo em situações ao longo do tempo ou espaço. Nesse quociente, o numerador é chamado valor considerado e o denominador é conhecido como valor base. Algebricamente, o número-índice entre b (base) e c (considerado) é dado por (10):

$$I_{h,c} = \frac{\text{valor considerado}}{\text{valor base}} \quad (10)$$

Por convenção, índice é adimensional, na forma percentual, com dois dígitos e sem sinal indicativo desta condição. O que lida com grandezas simples é dito elementar; o que trata de grandezas complexas (resume muitos valores) é conhecido como agregativo.

Em razão da sua definição, o número-índice também é conhecido como valor relativo ou simplesmente relativo (MILONE, 2004).

2.5.1 Índices de Desenvolvimento

O Índice de desenvolvimento é um índice sintético que abrange um conjunto amplo de indicadores com o objetivo de mensurar o grau de desenvolvimento dos objetos de estudo do pesquisador. O índice de desenvolvimento é resultado da agregação, com a mesma ponderação, de N blocos de indicadores. Cada um dos N blocos, por sua vez, resulta da agregação de diferentes variáveis.

Para cada uma das variáveis componentes dos blocos é calculado um índice, entre 0 (nenhum desenvolvimento) e 1 (desenvolvimento total), que indica a posição relativa dos objetos de estudo. Com esse fim, são fixados, a partir de parâmetros internacionais, os valores de referência máximo (1) e mínimo (0) de cada variável. Os índices dos blocos são obtidos através de uma média ponderada dos índices de cada uma das variáveis componentes do bloco.

Os objetos de estudo do pesquisador podem ser classificados pelo Índice de desenvolvimento em três grupos: baixo desenvolvimento (índices até 0,499), médio desenvolvimento (entre 0,500 e 0,799) e alto desenvolvimento (maiores que 0,800).

A tabela 1 apresenta um exemplo de pesos no índice de desenvolvimento, sendo que a soma dos pesos dentro de cada bloco deve ser igual a 1, e soma de todos os pesos no índice de desenvolvimento também deve ser igual a 1.

Tabela 1 Exemplo de pesos no índice de desenvolvimento

Blocos	Índices	Peso no Bloco	Peso no Índice de desenv.	Limite Inferior	Limite Superior
Bloco 1	índice 1	0,20	0,20/N	mínimo	máximo
	índice 2	0,10	0,10/N	mínimo	máximo
	⋮	⋮	⋮	⋮	⋮
	índice x	0,05	0,05/N	mínimo	máximo
Bloco 2	índice 1	0,08	0,08/N	mínimo	máximo
	índice 2	0,22	0,22/N	mínimo	máximo
	⋮	⋮	⋮	⋮	⋮
	índice y	0,15	0,15/N	mínimo	máximo
⋮	⋮	⋮	⋮	⋮	⋮
Bloco N	índice 1	0,30	0,30/N	mínimo	máximo
	índice 2	0,02	0,02/N	mínimo	máximo
	⋮	⋮	⋮	⋮	⋮
	índice z	0,18	0,18/N	mínimo	máximo

A fórmula de cálculo para operar a transformação das variáveis e dos indicadores no índice de desenvolvimento é apresentada na equação 11:

$$I_d = \frac{(X - L_i)}{(L_s - L_i)} \quad (11)$$

Onde:

I_d = índice de desenvolvimento

X = indicador do objeto de estudo

L_i = limite inferior do indicador

L_s = limite superior do indicador

2.5.2 Índices de Hierarquização

Este índice foi obtido de um estudo encontrado na tese de mestrado em Estatística intitulada Análise Fatorial de autoria de Fachel (1976).

Neste índice, a principal técnica utilizada é a análise fatorial, devido ao grande número de variáveis necessário para obtê-lo, sendo todas as variáveis importantes para o modelo.

A metodologia utilizada para a construção deste índice tem como base o fato de que os fatores resultantes da análise fatorial são ortogonais e, logo, linearmente independentes.

Seja f_{ij} , $i = 1, 2, \dots, N$, $j = 1, 2, \dots, m$ o valor do escore fatorial para a i -ésima observação e o j -ésimo fator e seja P_j o percentual de variância comum explicada pelo fator j . Um índice preliminar é obtido através da seguinte fórmula em (12).

$$I = \sqrt{\sum_{j=1}^m P_j f_{ij}^2} \quad (12)$$

Onde:

N = número de observações

m = número de fatores

Após o cálculo destes índices, faz-se a normalização deles, diminuindo seus valores pelo menor e dividindo-o pela diferença entre o maior e o menor índice, obtendo assim o índice de hierarquização, conforme a fórmula descrita em (13).

$$I_h = \frac{I - \min[I]}{\max[I] - \min[I]} \quad (13)$$

2.5.3 *Propriedades dos índices*

Pode-se medir o desvio de uma variável qualquer de diferentes modos. Apesar de muitos estudiosos terem proposto diversas formas de cálculo, não se obteve ainda um índice que se possa considerar perfeito nem uma fórmula que permita a quantificação exata da mudança conjunta de múltiplas variáveis (cada forma em uso tem uma vantagem, defeito ou restrição). Um dos melhores estudos sobre esta questão foi realizado por Fisher e está registrado em “*The making of index number*”, onde são propostas as seguintes propriedades esperadas para um número-índice ideal:

- *Determinabilidade*: medida finita e determinável;
- *Homogeneidade*: quantia não alterável por mudanças nas unidades de medida;
- *Invariância*: não é afetado pelo acréscimo ou decréscimo de itens cujos relativos sejam iguais;
- *Proporcionalidade*: termo proporcional aos relativos dos itens considerados;
- *Identidade*: para situações coincidentes, o índice é igual a 1;

2.6 *Discussão do Capítulo*

Esse capítulo mostrou um resumo do funcionamento de uma análise fatorial, abrangendo seus objetivos, suposições e vários outros aspectos. Também mostrou algumas técnicas de obtenção de números-índices. A ênfase maior é dada aos números-índices e aos métodos de estimação dos escores fatoriais e métodos de rotação dos fatores, já que para a obtenção dos índices utilizam-se os escores fatoriais obtidos na análise e também se pode variar o tipo de método de rotação dos fatores.

3 UM EXEMPLO DE APLICAÇÃO

3.1 Apresentação do problema

O problema que será utilizado para ilustrar esse trabalho consiste em elaborar um índice de potencialidade para os Distritos Operacionais do RS. Propõe-se a construção deste índice através da utilização de uma análise fatorial. Serão realizadas algumas alternativas de análises, variando seus métodos de estimação dos escores fatoriais e seus métodos de rotação dos fatores. Além disso, constatou-se que as variáveis deste estudo possuem escalas com grandezas muito diferenciadas, e por esta razão as análises fatoriais serão feitas padronizando as variáveis originais de duas maneiras: a primeira maneira é através dos ranks, que consiste em organizar as variáveis originais de acordo com os seus tamanhos; a segunda maneira é através da padronização pela distribuição Normal.

O objetivo da elaboração deste índice é verificar, através dos resultados obtidos visualizados em tabelas, as diferenças que existem nas variações dos métodos de estimação dos escores fatoriais e nos métodos de rotação dos fatores na criação deste índice. Como serão utilizados dois tipos de variáveis (“*rankeadas*” e padronizadas pela Normal), dois tipos de estimação dos escores fatoriais (Regressão e Bartlett), três tipos de rotação ortogonal dos fatores (Varimax, Quartimax, Equimax) e dois métodos de criação dos índices (Desenvolvimento e Hierarquização), serão obtidos 24 indicadores, podendo assim compará-los.

Nesta monografia não serão testadas todas as combinações possíveis para a realização da análise fatorial, o que daria $2 \times 2 \times 5 = 20$ possíveis resultados. E considerando os dois tipos

de índices, as possibilidades aumentam para 40. Então se optou por utilizar apenas os métodos de rotação ortogonal de fatores para diminuir este número.

Os dados foram obtidos em fontes de dados secundárias que disponibilizavam as variáveis mais atualizadas para a realização do estudo. Como algumas destas variáveis não são obtidas anualmente, coletaram-se variáveis desde o ano 2000 até o ano 2003. As fontes de dados das variáveis podem ser visualizadas nas tabelas 2 e 3.

Os dados foram analisados nos softwares *Excel* e *SPSS*, e os resultados são apresentados nos próximos itens desse capítulo.

3.2 Objetivos do estudo

O objetivo do estudo é criar um índice que avalie o potencial de cada Distrito Operacional. Do ponto de vista da técnica estatística, pretende-se verificar quais as principais implicações na elaboração deste índice quando se variam alguns aspectos na realização da análise fatorial, como o método de estimação dos escores fatoriais e o método de rotação dos fatores.

Após a obtenção destes indicadores, será possível averiguar qual deles produziu o resultado mais satisfatório para o estudo no que tange a escolha do método de estimação dos escores fatoriais, o método de rotação dos fatores e o tipo de padronização das variáveis na elaboração deste índice.

3.3 Escolha das variáveis

A escolha das variáveis foi realizada segundo os objetivos do estudo, a caracterização do desenvolvimento populacional, do desenvolvimento industrial e do desenvolvimento econômico de cada região.

As variáveis foram pesquisadas de acordo com a disponibilidade nas fontes de dados secundários. As variáveis constantes na Tabela 2 referem-se aos municípios do estado do Rio

Grande do Sul (397 municípios de acordo com o censo 2000) e apresentam a descrição de cada variável, a fonte de dados de onde se extraiu esta variável e o ano em que as variáveis foram obtidas, enquanto que as variáveis constantes da Tabela 3 referem-se a cada Distrito Operacional do RS, e apresentam as mesmas informações contidas na Tabela 2.

Tabela 2 Descrição das variáveis e fonte de dados por município

Variável	Descrição	Fonte	Ano Base	Endereço eletrônico
Área total	Área em km de extensão do município	FEE	2003	www.fee.tche.br
População	Número de habitantes	FEE	2001	www.fee.tche.br
IDH	Índice de Desenvolvimento Urbano	ONU	2000	www.onu-brasil.org.br
PIB	Produto Interno Bruto	FEE	2001	www.fee.tche.br
Veículos	Número de veículos cadastrados	DETRAN	2003	www.detran.rs.gov.br
Nº. indústrias	Número de estabelecimentos industriais	RAIS	2002	www.rais.gov.br
Nº. comércio	Número de estabelecimentos comerciais	RAIS	2002	www.rais.gov.br
Nº. construção civil	Número de estabelecimentos de C. civil	RAIS	2002	www.rais.gov.br
Nº. serviços	Número de estabelec. prestadores de serviço	RAIS	2002	www.rais.gov.br
Nº. agropecuária	Número de estabelecimentos agropecuários	RAIS	2002	www.rais.gov.br

Tabela 3 Descrição das variáveis e fonte de dados por DOP

Variável	Descrição	Fonte	Ano Base	Endereço eletrônico
Extensão pavimentada		DAER	2002	www.daer.rs.gov.br
Extensão não pavimentada		DAER	2002	www.daer.rs.gov.br
VDM	Número de veículos passantes	DAER	2002	www.daer.rs.gov.br

A tabela 4 apresenta uma lista com os 17 Distritos Operacionais vigentes na elaboração desta monografia.

Tabela 4 Descrição dos DOPs atuais

DOP	Regiões	DOP	Regiões
1º DOP	Esteio	11º DOP	Lajeado
2º DOP	Bento Gonçalves	12º DOP	Santiago
3º DOP	Santa Cruz	13º DOP	Erechim
4º DOP	Santa Maria	14º DOP	Santa Rosa
5º DOP	Cruz Alta	15º DOP	São Francisco de Paula
6º DOP	Passo Fundo	16º DOP	Osório
7º DOP	Pelotas	17º DOP	Palmeira das Missões
8º DOP	Bagé		
9º DOP	Alegrete		
10º DOP	Cachoeira do Sul		

3.3.1 Padronização das variáveis

As variáveis deste estudo possuem escalas com grandezas muito diferenciadas, e por esta razão as análises fatoriais serão feitas padronizando as variáveis.

Serão realizados dois tipos de padronização nas variáveis: padronização por *ranks* e padronização pela distribuição Normal.

3.3.2 Descrição das variáveis

A Tabela 5 apresenta a descrição das variáveis utilizadas e as medidas-resumo utilizadas para representar cada DOP.

Tabela 5 Medidas-resumo das variáveis consideradas

Variável	Medida-resumo
Área total	total
População	total
IDH	mediana
PIB	mediana
Veículos	total
Nº. indústrias	total
Nº. comércio	total
Nº. construção civil	total
Nº. serviços	total
Nº. agropecuário	total
Extensão pavimentada	Valor por DOP
Extensão não pavimentada	Valor por DOP
VDM	Valor por DOP

O banco de dados foi montado considerando todos os municípios do estado do Rio Grande do Sul, com base no censo 2000, e logo após agrupou-se esses municípios dentro dos seus respectivos DOPs.

Na tabela 5, o total refere-se à soma dos valores de determinada variável para cada município que compõe o DOP. Para as variáveis extensão pavimentada, extensão não pavimentada e VDM foi utilizado o valor disponibilizado pelo DAER para o respectivo DOP.

A mediana foi a medida resumo escolhida para representar os valores dos municípios dentro de um mesmo DOP para as variáveis IDH e PIB. A média e o desvio-padrão também foram calculados, no entanto a mediana se mostrou mais adequada para as análises devido a grande variabilidade existente entre os valores de diferentes municípios em um mesmo DOP.

3.4 Análise Fatorial

Como as variáveis originais possuem grandezas diferentes, foi necessário criar novas variáveis, padronizando as variáveis originais, para realizar a análise fatorial. Utilizaram-se dois novos bancos de dados originados a partir do banco de dados original, sendo um deles com as mesmas variáveis do banco original organizadas por *ranks*, e no outro se utilizou as mesmas variáveis do banco de dados original padronizadas pela distribuição Normal.

A análise foi realizada utilizando-se dois métodos de estimação dos escores fatoriais (Regressão e Bartlett), três métodos de rotação ortogonal dos fatores (Varimax, Quartimax e Equimax), e dois tipos de padronização das variáveis originais (*ranks* e padronização pela Normal) para a utilização em dois modelos de índices (Desenvolvimento e Hierarquização).

3.4.1 Suposições para uso da análise fatorial

Uma das suposições para a utilização da análise fatorial é que os vetores das variáveis originais tenham distribuição Normal Multivariada. Para a verificação desta suposição, é preciso que se realize um teste de aderência para o caso onde a distribuição é completamente especificada, pois no banco de dados estão contidas informações de todo o estado do Rio Grande do Sul, podendo assim se obter as médias e variâncias para cada variável. A realização deste teste envolve uma equação muito complexa e o autor desta monografia resolveu não apresentá-la aqui, pois como cada variável possui dois parâmetros, média e variância, e como o banco de dados possui 13 variáveis, isso dá um total de 26 parâmetros. Após a obtenção de

cada parâmetro, deve-se encontrar a estatística de teste para compará-la com a distribuição qui-quadrado para então rejeitar ou aceitar a hipótese nula que diz que os vetores das variáveis originais possuem distribuição Normal Multivariada. Assim, para satisfazer esta condição para a utilização da análise fatorial, supõe-se que os vetores das variáveis originais possuem distribuição Normal Multivariada.

Outras suposições para o uso da análise fatorial indicam qual o grau de ajuste dos dados a análise fatorial, e a verificação destas suposições se dão através da Medida KMO de Adequação da Amostra e do Teste de Esfericidade de Bartlett. Os resultados obtidos na realização destes dois testes podem ser observados nas tabelas 6 e 7.

Tabela 6 Teste KMO e Teste de Bartlett para dados organizados por ranks

KMO e Teste de Bartlett		
Medida KMO de Adequação da Amostra		0,6417
Teste de Esfericidade de Bartlett	Aprox. Qui-Quadrado	231,3581
	gl	78
	Sig.	0,000

Tabela 7 Medida KMO e Teste de Bartlett para dados padronizados pela Normal

KMO e Teste de Bartlett		
Medida KMO de Adequação da Amostra		0,6508
Teste de Esfericidade de Bartlett	Aprox. Qui-Quadrado	297,5390
	gl	78
	Sig.	0,000

De acordo com Chatfield e Collins (1980), os dados possuem um razoável grau de

ajuste à análise fatorial se o valor resultante da Medida KMO estiver entre 0,6 e 0,8. Observando então os valores obtidos pela Medida KMO para os dois tipos de dados nas tabelas 6 e 7, nota-se que eles possuem um razoável grau de ajuste à análise fatorial. E os valores obtidos pelo Teste de Esfericidade de Bartlett nos permitem também confirmar a possibilidade de adequação do método de análise fatorial para o tratamento dos dados.

3.4.2 Número de fatores a serem extraídos

O número de fatores a serem extraídos nesta análise fatorial seguiu o critério da raiz latente, onde somente os fatores que possuem um autovalor maior que 1 são considerados, e todos os fatores com autovalores menor que 1 são descartados.

Optou-se pelo critério da raiz latente (ou autovalor), pois ele mostrou-se mais apropriado, fornecendo uma menor quantidade de fatores para se trabalhar, facilitando, assim, os cálculos para a obtenção dos futuros índices.

Independentemente do tipo de índice que será criado, e dos métodos que serão utilizados na análise fatorial, o critério da raiz latente apontou três fatores a serem extraídos, conforme mostra a figura 4.

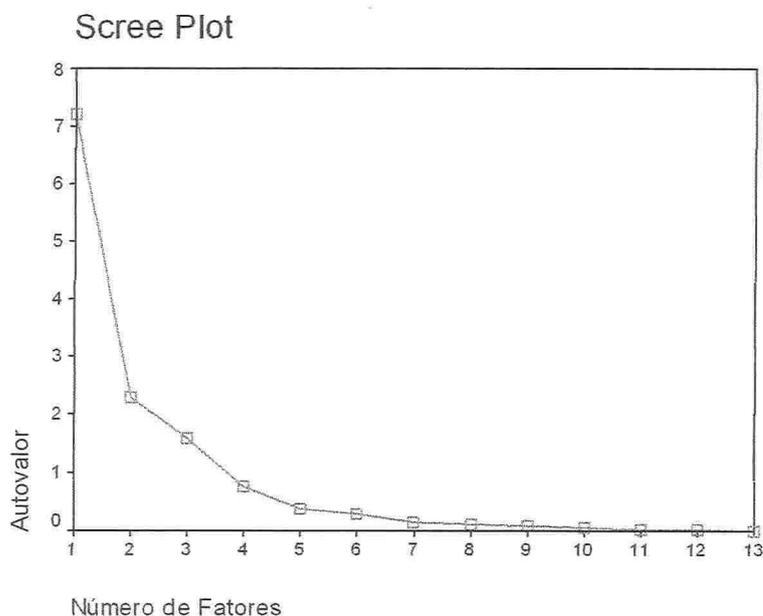


Figura 4 Gráfico dos autovalores na utilização da Análise Fatorial para a elaboração do índice

3.4.3 Estimação dos Escores Fatoriais

Os escores fatoriais foram obtidos para os dois tipos de dados (por *ranks* e padronizados pela Normal) e para os três tipos de rotação ortogonal dos fatores (Varimax, Quartimax e Equimax).

Os resultados obtidos nos dois tipos de estimação dos escores estão nas tabelas 8 e 9.

Tabela 8 Escores gerados pelos métodos de Regressão e Bartlett para dados “rankeados”

DOP	Rotação Varimax						Rotação Quartimax					
	Regressão			Bartlett			Regressão			Bartlett		
	Fator 1	Fator 2	Fator 3	Fator 1	Fator 2	Fator 3	Fator 1	Fator 2	Fator 3	Fator 1	Fator 2	Fator 3
Esteio	1,311	0,045	1,402	1,311	0,045	1,402	1,429	0,476	1,191	1,429	0,476	1,191
Bento Gonçalves	1,458	-0,059	1,449	1,458	-0,059	1,449	1,535	0,626	1,217	1,535	0,626	1,217
Santa Cruz	0,706	0,411	-1,225	0,706	0,411	-1,225	0,624	-0,194	-1,320	0,624	-0,194	-1,320
Santa Maria	0,615	-1,117	-0,246	0,615	-1,117	-0,246	0,143	1,249	-0,325	0,143	1,249	-0,325
Cruz Alta	-0,893	-0,410	1,518	-0,893	-0,410	1,518	-0,754	0,141	1,637	-0,754	0,141	1,637
Passo Fundo	0,816	-0,128	1,248	0,816	-0,128	1,248	0,888	0,458	1,114	0,888	0,458	1,114
Pelotas	1,173	-1,254	-1,688	1,173	-1,254	-1,688	0,407	1,508	-1,833	0,407	1,508	-1,833
Bagé	-1,240	-1,429	0,135	-1,240	-1,429	0,135	-1,630	0,913	0,330	-1,630	0,913	0,330
Alegrete	-0,208	-1,705	-0,418	-0,208	-1,705	-0,418	-0,850	1,506	-0,368	-0,850	1,506	-0,368
Cachoeira do Sul	-1,082	-0,103	-0,469	-1,082	-0,103	-0,469	-1,104	-0,300	-0,302	-1,104	-0,300	-0,302
Lajeado	0,262	1,659	0,523	0,262	1,659	0,523	0,898	-1,440	0,464	0,898	-1,440	0,464
Santiago	-1,602	-0,599	0,355	-1,602	-0,599	0,355	-1,643	0,020	0,594	-1,643	0,020	0,594
Erechim	-0,745	0,911	-0,171	-0,745	0,911	-0,171	-0,395	-1,120	-0,067	-0,395	-1,120	-0,067
Santa Rosa	0,526	0,489	-0,387	0,526	0,489	-0,387	0,603	-0,292	-0,465	0,603	-0,292	-0,465
São Francisco de Paula	-1,422	1,316	-0,081	-1,422	1,316	-0,081	-0,867	-1,730	0,119	-0,867	-1,730	0,119
Osório	-0,107	1,037	-0,350	-0,107	1,037	-0,350	0,215	-1,024	-0,340	0,215	-1,024	-0,340
Palmeira das Missões	0,431	0,936	-1,594	0,431	0,936	-1,594	0,501	-0,796	-1,648	0,501	-0,796	-1,648

DOP	Rotação Equimax					
	Regressão			Bartlett		
	Fator 1	Fator 2	Fator 3	Fator 1	Fator 2	Fator 3
Esteio	1,247	0,140	1,453	1,247	0,140	1,453
Bento Gonçalves	1,400	0,049	1,505	1,400	0,049	1,505
Santa Cruz	0,713	0,490	-1,191	0,713	0,490	-1,191
Santa Maria	0,721	-1,054	-0,233	0,721	-1,054	-0,233
Cruz Alta	-0,911	-0,510	1,477	-0,911	-0,510	1,477
Passo Fundo	0,776	-0,073	1,278	0,776	-0,073	1,278
Pelotas	1,345	-1,121	-1,654	1,345	-1,121	-1,654
Bagé	-1,113	-1,535	0,070	-1,113	-1,535	0,070
Alegrete	-0,040	-1,710	-0,445	-0,040	-1,710	-0,445
Cachoeira do Sul	-1,049	-0,191	-0,512	-1,049	-0,191	-0,512
Lajeado	0,093	1,668	0,552	0,093	1,668	0,552
Santiago	-1,555	-0,744	0,284	-1,555	-0,744	0,284
Erechim	-0,815	0,844	-0,191	-0,815	0,844	-0,191
Santa Rosa	0,495	0,539	-0,361	0,495	0,539	-0,361
São Francisco de Paula	-1,529	1,186	-0,123	-1,529	1,186	-0,123
Osório	-0,185	1,029	-0,343	-0,185	1,029	-0,343
Palmeira das Missões	0,408	0,993	-1,565	0,408	0,993	-1,565

Tabela 9 Escores gerados pelos métodos de Regressão e Bartlett para dados padronizados pela Normal

DOP	Rotação Varimax						Rotação Quartimax					
	Regressão			Bartlett			Regressão			Bartlett		
	Fator 1	Fator 2	Fator 3	Fator 1	Fator 2	Fator 3	Fator 1	Fator 2	Fator 3	Fator 1	Fator 2	Fator 3
Esteio	3,721	-0,073	0,046	3,721	-0,073	0,046	3,695	-0,228	-0,384	3,695	-0,228	-0,384
Bento Gonçalves	0,231	0,568	2,483	0,231	0,568	2,483	0,545	0,699	2,400	0,545	0,699	2,400
Santa Cruz	0,117	0,500	-1,107	0,117	0,500	-1,107	0,002	0,430	-1,142	0,002	0,430	-1,142
Santa Maria	0,178	-0,736	-0,120	0,178	-0,736	-0,120	0,137	-0,749	-0,094	0,137	-0,749	-0,094
Cruz Alta	-0,555	-0,355	1,199	-0,555	-0,355	1,199	-0,421	-0,262	1,275	-0,421	-0,262	1,275
Passo Fundo	-0,301	0,434	1,513	-0,301	0,434	1,513	-0,104	0,532	1,508	-0,104	0,532	1,508
Pelotas	0,226	-1,087	-0,968	0,226	-1,087	-0,968	0,070	-1,149	-0,919	0,070	-1,149	-0,919
Bagé	-0,516	-1,772	0,195	-0,516	-1,772	0,195	-0,552	-1,735	0,363	-0,552	-1,735	0,363
Alegrete	-0,141	-1,799	-0,213	-0,141	-1,799	-0,213	-0,229	-1,800	-0,084	-0,229	-1,800	-0,084
Cachoeira do Sul	-0,319	-0,226	-0,871	-0,319	-0,226	-0,871	-0,428	-0,262	-0,812	-0,428	-0,262	-0,812
Lajeado	-0,226	1,528	0,530	-0,226	1,528	0,530	-0,107	1,564	0,457	-0,107	1,564	0,457
Santiago	-0,707	-0,994	0,443	-0,707	-0,994	0,443	-0,684	-0,936	0,583	-0,684	-0,936	0,583
Erechim	-0,505	0,583	-0,341	-0,505	0,583	-0,341	-0,521	0,583	-0,315	-0,521	0,583	-0,315
Santa Rosa	-0,304	0,726	-0,105	-0,304	0,726	-0,105	-0,288	0,731	-0,113	-0,288	0,731	-0,113
São Francisco de Paula	-0,479	0,441	-0,730	-0,479	0,441	-0,730	-0,546	0,419	-0,695	-0,546	0,419	-0,695
Osório	-0,177	1,047	-0,836	-0,177	1,047	-0,836	-0,238	1,004	-0,873	-0,238	1,004	-0,873
Palmeira das Missões	-0,241	1,215	-1,117	-0,241	1,215	-1,117	-0,330	1,158	-1,154	-0,330	1,158	-1,154

DOP	Rotação Equimax					
	Regressão			Bartlett		
	Fator 1	Fator 2	Fator 3	Fator 1	Fator 2	Fator 3
Esteio	3,715	-0,045	0,231	3,715	-0,045	0,231
Bento Gonçalves	0,101	0,492	2,508	0,101	0,492	2,508
Santa Cruz	0,167	0,535	-1,084	0,167	0,535	-1,084
Santa Maria	0,191	-0,731	-0,133	0,191	-0,731	-0,133
Cruz Alta	-0,611	-0,396	1,158	-0,611	-0,396	1,158
Passo Fundo	-0,381	0,384	1,508	-0,381	0,384	1,508
Pelotas	0,284	-1,054	-0,989	0,284	-1,054	-0,989
Bagé	-0,509	-1,782	0,114	-0,509	-1,782	0,114
Alegrete	-0,113	-1,792	-0,275	-0,113	-1,792	-0,275
Cachoeira do Sul	-0,273	-0,201	-0,893	-0,273	-0,201	-0,893
Lajeado	-0,267	1,509	0,565	-0,267	1,509	0,565
Santiago	-0,719	-1,013	0,376	-0,719	-1,013	0,376
Erechim	-0,493	0,589	-0,348	-0,493	0,589	-0,348
Santa Rosa	-0,305	0,727	-0,098	-0,305	0,727	-0,098
São Francisco de Paula	-0,446	0,460	-0,739	-0,446	0,460	-0,739
Osório	-0,145	1,071	-0,811	-0,145	1,071	-0,811
Palmeira das Missões	-0,197	1,247	-1,090	-0,197	1,247	-1,090

Observando as tabelas 8 e 9, nota-se que os escores fatoriais não variam dentro de cada método de rotação dos fatores, independente do método de estimação dos escores utilizado e do tipo de padronização utilizado nas variáveis.

Como os escores fatoriais não variam em nenhum dos dois tipos de dados, pode-se utilizar qualquer um dos métodos estudados de estimação dos escores fatoriais para realizar a análise fatorial que ajudará na construção do índice potencial para os distritos operacionais do Rio Grande do Sul. Assim, optou-se pelo método de Regressão, pois é o mais comumente utilizado na realização das análises fatoriais.

3.4.4 Rotação dos Fatores

Após a escolha do método de regressão para estimar os escores fatoriais, a análise fatorial foi realizada utilizando os três tipos de rotação ortogonal para os dois tipos de dados, gerando assim seis tabelas com três fatores cada. Os fatores foram denominados preliminarmente como: expansão populacional, expansão industrial, expansão econômica para todos os casos, com exceção para a rotação Quartimax para dados classificados por ranks, pois a configuração dos seus fatores sugere que o primeiro fator seja denominado como expansão populacional/industrial, o segundo seja denominado como expansão agrícola e o terceiro continua como expansão econômica.

Esses fatores foram gerados a partir de todas as variáveis consideradas no estudo. Os fatores gerados a partir de todas variáveis podem ser visualizados na tabela 10.

Tabela 10 Fatores gerados a partir das rotações ortogonais

Dados organizados por Ranks								
Rotação Varimax			Rotação Quartimax			Rotação Equimax		
Fator 1	Fator 2	Fator 3	Fator 1	Fator 2	Fator 3	Fator 1	Fator 2	Fator 3
População	Agropec.	PIB	VDM	Agropec.	PIB	População	Agropec.	PIB
Veículos	Área	IDH	Comércio	Área	IDH	Veículos	Área	IDH
Comércio	Indústria	Área não Pavim.	Veículos			Comércio	Indústria	Área não Pavim.
VDM	Área Pavim.		População			VDM	Serviços	
Const. Civil	Serviços		Const. Civil			Const. Civil	Área Pavim.	
			Serviços					
			Área Pavim.					
			Indústria					
			Área não Pavim.					

Dados Padronizados pela distribuição Normal								
Rotação Varimax			Rotação Quartimax			Rotação Equimax		
Fator 1	Fator 2	Fator 3	Fator 1	Fator 2	Fator 3	Fator 1	Fator 2	Fator 3
População	Serviços	PIB	Veículos	Área Pavim.	PIB	População	Serviços	PIB
Comércio	Área Pavim.	IDH	Comércio	Serviços	IDH	Comércio	Const. Civil	IDH
Veículos	Const. Civil		População	Área		Veículos	Área Pavim.	
VDM	Área		VDM	Const. Civil		VDM	Área	
Indústria	Agropec.		Indústria	Agropec.		Indústria	Agropec.	
Área não Pavim.			Área não Pavim.			Área não Pavim.		

Nos dados organizados por *ranks*, a rotação Varimax e a rotação Equimax geraram os mesmos fatores; nos dados padronizados pela Normal, as três rotações geraram os mesmos blocos.

Percebe-se certa tendência na formação dos fatores, a exceção do caso anteriormente mencionado, e por isso os índices serão calculados para todas as combinações possíveis (utilizando o método de regressão para estimar os escores fatoriais), inclusive no caso diferente dos demais, entre os dois tipos de dados (por *ranks* e padronizados pela Normal), e os três tipos de rotação ortogonal (Varimax, Quartimax e Equimax).

A diferença entre os fatores criados para os dados organizados por *ranks* e os dados padronizados pela Normal se deve ao diferente método de padronização utilizado para estas variáveis. As variáveis padronizadas pela distribuição Normal geraram os mesmos fatores, independentemente do método de rotação dos fatores. Nas variáveis organizadas por *ranks*, a rotação Quartimax gerou fatores diferentes dos outros dois tipos de rotação, e isso pode ter ocorrido pelo fato de que a rotação Quartimax diz respeito à simplificação da descrição de cada linha (que corresponde as variáveis) da matriz de cargas fatoriais, enquanto que as outras rotações não se utilizam da simplificação das linhas da matriz das cargas fatoriais.

3.4.5 Validação da análise fatorial

Para validar este modelo é necessário verificar sua estabilidade e fazer a detecção de observações discrepantes.

Como a estabilidade é dependente do tamanho da amostra, neste caso o tamanho é o maior possível, pois existem 17 Distritos Operacionais no RS e todos eles são contemplados no banco de dados, o que constitui toda a população.

No caso da detecção de observações discrepantes, notou-se que o 1º DOP possuía informações bem diferenciadas das demais regiões, pois este DOP compreende o município de Porto Alegre. Como esta informação é muito importante para a obtenção dos resultados, não é possível retirá-la para a realização de uma nova análise fatorial.

Assim, com a estabilidade do modelo verificada e com a observação de que o único fator considerado discrepante não pode ser retirado do banco de dados, os resultados obtidos na realização desta análise fatorial podem ser considerados válidos.

3.5 Construção dos índices

Para a elaboração de um índice de potencialidade dos DOPs do RS será realizada a análise fatorial para os dois tipos de dados (organizados por *ranks* e padronizados pela Normal) com o método de regressão para estimar os escores fatoriais e com os três métodos de rotação ortogonal dos fatores.

A elaboração deste índice se realizará através da utilização dos dois modelos de índice visualizados anteriormente, e após será verificada as propriedades dos índices construídos segundo a seção 2.5 desta monografia.

3.5.1 Índices de Desenvolvimento

O primeiro modelo de índice usou como referência o índice de desenvolvimento mencionado na revisão bibliográfica, item 2.5.1.

Este modelo irá determinar o grau de desenvolvimento dos Distritos Operacionais do Rio Grande do Sul através da análise das variáveis contidas no banco de dados do estudo, podendo caracterizar a potencialidade de cada região.

Os aspectos da análise fatorial que o modelo de desenvolvimento leva em consideração são os fatores (blocos) formados na análise e também o máximo e o mínimo de cada variável do estudo.

Por exemplo, para a criação do índice de desenvolvimento para o DOP de Esteio, utilizando a rotação Varimax para os dados organizados por *ranks*, a fórmula utilizada pode ser visualizada em 14.

$$I_d = 0,1333 \cdot \frac{17-1}{17-1} + 0,05 \cdot \frac{17-1}{17-1} + \dots + 0,05 \cdot \frac{2-1}{17-1} = 0,8063 \quad (14)$$

Esta fórmula utiliza valores encontrados no banco de dados original e na tabela 11.

3.5.1.1 Utilizando variáveis por Ranks

Neste modelo de índice utilizando as variáveis organizadas por *ranks*, os blocos possuem pesos de acordo com a importância de cada variável nos seus respectivos blocos, e os limites inferior e superior são sempre os mesmos para todas as variáveis, conforme se pode observar nas tabelas 11 e 12.

Tabela 11 Pesos no índice para variáveis rankeadas, nas rotações Varimax e Equimax

		Peso no Bloco	Peso no Índice	LI	LS
Expansão Populacional	população	0,40	0,1333	1	17
	veículos	0,15	0,0500	1	17
	comércio	0,20	0,0667	1	17
	VDM	0,10	0,0333	1	17
	construção civil	0,15	0,0500	1	17
Expansão Industrial	agropecuária	0,05	0,0167	1	17
	área	0,15	0,0500	1	17
	indústria	0,40	0,1333	1	17
	área pavimentada	0,20	0,0667	1	17
	serviços	0,20	0,0667	1	17
Expansão Econômica	PIB	0,50	0,1667	1	17
	IDH	0,35	0,1167	1	17
	área não pavimentada	0,15	0,0500	1	17

Tabela 12 Pesos no índice para variáveis rankeadas, na rotação Quartimax

		Peso no Bloco	Peso no Índice	LI	LS
Expansão Populacional / Industrial	população	0,25	0,0833	1	17
	veículos	0,10	0,0333	1	17
	comércio	0,10	0,0333	1	17
	VDM	0,05	0,0167	1	17
	construção civil	0,05	0,0167	1	17
	indústria	0,25	0,0833	1	17
	serviços	0,05	0,0167	1	17
	área pavimentada	0,10	0,0333	1	17
	área não pavimentada	0,05	0,0167	1	17
Expansão Agrícola	agropecuária	0,60	0,2000	1	17
	área	0,40	0,1333	1	17
Expansão Econômica	PIB	0,60	0,2000	1	17
	IDH	0,40	0,1333	1	17

Conforme visto nas tabelas 11 e 12, os fatores formados para as variáveis organizadas por *ranks* são iguais nas rotações Varimax e Equimax, gerando assim os mesmos índices.

Tabela 13 Índice de Desenvolvimento para variáveis organizadas por Ranks

	Varimax e Equimax	Quartimax
DOP	ÍNDICE	ÍNDICE
Esteio	0,8063	0,7219
Bento Gonçalves	0,8927	0,8552
Santa Cruz do Sul	0,5161	0,4250
Santa Maria	0,4281	0,5198
Cruz Alta	0,4656	0,5490
Passo Fundo	0,7146	0,6885
Pelotas	0,4594	0,5281
Bagé	0,2927	0,5156
Alegrete	0,2990	0,4979
Cachoeira do Sul	0,2276	0,2875
Lajeado	0,7083	0,5104
Santiago	0,2979	0,4354
Erechim	0,4792	0,4000
Santa Rosa	0,6333	0,5635
São Francisco de Paula	0,3724	0,3010
Osório	0,4042	0,2885
Palmeira das Missões	0,5026	0,4125

3.5.1.2 Utilizando variáveis padronizadas pela Normal

Neste mesmo modelo de índice, utilizou-se as variáveis padronizadas pela distribuição Normal. Neste caso, os blocos também possuem pesos de acordo com a importância de cada variável nos seus respectivos blocos, e os limites inferior e superior dependem da variável, conforme a tabela 14.

Tabela 14 Pesos no índice para variáveis padronizadas pela Normal, nas três rotações

		Peso no Bloco	Peso no Índice	LI	LS
Expansão Populacional	população	0,40	0,1333	-0,53	3,78
	veículos	0,10	0,0333	-0,47	3,75
	comércio	0,15	0,0500	-0,69	3,68
	VDM	0,10	0,0333	-0,75	3,36
	indústria	0,15	0,0500	-1,00	3,20
	área não pavimentada	0,10	0,0333	-1,96	1,66
Expansão Industrial	agropecuária	0,05	0,0167	-1,52	1,55
	área	0,20	0,0667	-1,27	2,05
	construção civil	0,15	0,0500	-1,53	2,08
	área pavimentada	0,40	0,1333	-1,93	1,66
	serviços	0,20	0,0667	-1,75	1,99
Expansão Econômica	PIB	0,60	0,2000	-1,44	2,11
	IDH	0,40	0,1333	-1,30	2,10

Como os fatores formados para as variáveis padronizadas pela distribuição Normal são iguais nas três rotações, eles acabaram gerando o mesmo índice, pois se utilizaram dos mesmos pesos e limites inferiores e superiores.

Tabela 15 Índice de Desenvolvimento para variáveis padronizadas pela Normal

Varimax, Quartimax e Equimax	
DOP	ÍNDICE
Esteio	0,2726
Bento Gonçalves	0,0131
Santa Cruz do Sul	0,1530
Santa Maria	0,0843
Cruz Alta	-0,1430
Passo Fundo	0,0155
Pelotas	0,2252
Bagé	-0,0419
Alegrete	0,1442
Cachoeira do Sul	-0,0132
Lajeado	0,0654
Santiago	-0,1340
Erechim	0,0821
Santa Rosa	0,1603
São Francisco de Paula	-0,0184
Osório	0,2673
Palmeira das Missões	0,1889

3.5.2 Índices de Hierarquização

Este outro modelo de índice usou como referência o índice de hierarquização mencionado na revisão bibliográfica, no item 2.5.2.

Este modelo irá classificar os Distritos Operacionais do Rio Grande do Sul através de uma hierarquia das regiões, que levam em consideração a análise das variáveis contidas no banco de dados do estudo.

Os aspectos da análise fatorial que o modelo de hierarquização utiliza são os escores

fatoriais e a percentagem da variância comum explicada pelos fatores.

3.5.2.1 Utilizando variáveis por Ranks

Neste modelo de índice utilizando as variáveis organizadas por *ranks*, o valor do escore fatorial é sempre o mesmo para um determinado tipo de rotação, independente do método de estimação dos escores fatoriais, e a percentagem da variância comum explicada pelos fatores é igual para os três tipos de rotação.

Tabela 16 Índice de Hierarquização para variáveis organizadas por Ranks

	Varimax	Quartimax	Equimax
DOP	ÍNDICE	ÍNDICE	ÍNDICE
Esteio	0,8259	0,8526	0,7632
Bento Gonçalves	0,9636	0,9616	0,8984
Santa Cruz do Sul	0,3123	0,2297	0,3196
Santa Maria	0,2634	0,0918	0,3142
Cruz Alta	0,5321	0,4097	0,5364
Passo Fundo	0,3781	0,3938	0,3508
Pelotas	0,9395	0,5940	1,0000
Bagé	0,8401	1,0000	0,7525
Alegrete	0,3773	0,5355	0,3599
Cachoeira do Sul	0,4733	0,4521	0,4444
Lajeado	0,3761	0,5507	0,3526
Santiago	1,0000	0,9535	0,9435
Erechim	0,2826	0,1010	0,3207
Santa Rosa	0,0022	0,0223	0,0000
São Francisco de Paula	0,9604	0,6221	0,9939
Osório	0,0000	0,0000	0,0246
Palmeira das Missões	0,3837	0,3596	0,3784

3.5.2.2 Utilizando variáveis padronizadas pela Normal

Neste outro modelo de índice, agora utilizando as variáveis padronizadas pela Normal, o valor do escore fatorial também é sempre o mesmo para um determinado tipo de rotação, independente do método de estimação dos escores fatoriais, e a percentagem da variância comum explicada pelos fatores também é igual para os três tipos de rotação.

Tabela 17 Índice de Hierarquização para variáveis padronizadas pela Normal

	Varimax	Quartimax	Equimax
DOP	ÍNDICE	ÍNDICE	ÍNDICE
Esteio	1,0000	1,0000	1,0000
Bento Gonçalves	0,2120	0,2440	0,2064
Santa Cruz do Sul	0,0296	0,0255	0,0354
Santa Maria	0,0000	0,0000	0,0016
Cruz Alta	0,0933	0,0719	0,1054
Passo Fundo	0,0886	0,0819	0,0964
Pelotas	0,1057	0,1066	0,1085
Bagé	0,2283	0,2319	0,2295
Alegrete	0,2014	0,2080	0,2010
Cachoeira do Sul	0,0064	0,0280	0,0000
Lajeado	0,1610	0,1624	0,1627
Santiago	0,1444	0,1392	0,1488
Erechim	0,0442	0,0497	0,0431
Santa Rosa	0,0159	0,0165	0,0171
São Francisco de Paula	0,0438	0,0583	0,0396
Osório	0,0864	0,0894	0,0878
Palmeira das Missões	0,1380	0,1429	0,1391

3.5.3 *Propriedades dos índices*

Nos modelos propostos acima, foram desenvolvidos ao total 9 índices, sendo que alguns não apresentam as propriedades necessárias para serem configurados como tal. Para cada modelo será feita uma verificação destas propriedades, com o intuito de poder classificá-los como índices.

3.5.3.1 *Modelo de Desenvolvimento*

Foram desenvolvidos 3 índices no modelo de desenvolvimento, e todos os índices deste modelo possuem a propriedade de determinabilidade, pois suas medidas são finitas e determináveis.

Quanto à homogeneidade, os três índices também apresentam esta propriedade, pois alterando as unidades de medidas das variáveis originais não irá alterar a padronização das mesmas, e assim os índices também não são alterados.

Caso se acrescente ou se diminua os itens cujos relativos sejam iguais, os índices não são afetados (pelo mesmo motivo citado acima), e por isso respeitam a propriedade da invariância.

Como não há situações coincidentes, não é possível verificar a propriedade da identidade.

3.5.3.2 *Modelo de Hierarquização*

Em todos os 6 índices desenvolvidos no modelo de hierarquização, as propriedades de determinabilidade, homogeneidade e invariância são atendidas plenamente, pelos motivos citados nestas propriedades para o modelo de desenvolvimento.

Como não há situações coincidentes, não é possível verificar a propriedade da identidade.

3.5.4 Diferenças entre os índices propostos

Durante o processo de elaboração dos modelos de índices, verificou-se que estes modelos apresentaram algumas diferenças entre si, as quais se destacam:

- Os dois modelos de criação de índices geraram seis índices, mas como no modelo de desenvolvimento se utiliza fatores para a sua criação, e alguns métodos de rotação geraram os mesmos fatores, este modelo acabou criando índices iguais independente do tipo de rotação utilizada, podendo assim considerar apenas três índices neste modelo, enquanto que no de hierarquização pode-se considerar os seis índices;
- As amplitudes dos índices desenvolvidos através do modelo de desenvolvimento variam entre 0,416 e 0,665 e a amplitude para os índices desenvolvidos através do modelo de hierarquização é sempre igual a 1, o que mostra que os índices de desenvolvimento possuem uma menor variabilidade;
- Conforme se pode observar nas tabelas 18 e 19, as médias dos índices desenvolvidos para variáveis padronizadas são sempre as menores, e também pode-se notar que os desvios-padrão dos índices de desenvolvimento são menores que os desvios-padrão dos índices de hierarquização, mostrando novamente que os índices de desenvolvimento possuem menor variabilidade;
- Os coeficientes de variação para os índices gerados para as variáveis padronizadas pela Normal são todos maiores que 1, o que indica que eles são menos precisos que os índices gerados para as variáveis organizadas por *ranks*.

Tabela 18 Descritivas dos índices gerados através do modelo de desenvolvimento

ROTAÇÃO	RANKS		PADRONIZADO
	Varimax e Equimax	Quartimax	Varimax, Quartimax e Equimax
AMPLITUDE	0,665	0,568	0,416
MÉDIA	0,500	0,500	0,078
DESVIO PADRÃO	0,192	0,153	0,127
COEF. VARIAÇÃO	0,383	0,307	1,635

Tabela 19 Descritivas dos índices gerados através do modelo de hierarquização

ROTAÇÃO	RANKS			PADRONIZADO		
	Varimax	Quartimax	Equimax	Varimax	Quartimax	Equimax
AMPLITUDE	1,000	1,000	1,000	1,000	1,000	1,000
MÉDIA	0,524	0,478	0,515	0,153	0,156	0,154
DESVIO PADRÃO	0,333	0,329	0,320	0,230	0,230	0,230
COEF. VARIAÇÃO	0,636	0,688	0,621	1,504	1,472	1,489

3.5.5 Correlações entre os índices e as variáveis

A correlação tem como objetivo básico medir o grau de associação linear entre duas variáveis. Ela é medida através do coeficiente de correlação (r), que fica entre os limites -1 e $+1$, isto é, $-1 \leq r \leq +1$.

O coeficiente de correlação linear de Pearson pode ser calculado a partir da equação 15:

$$r = \frac{n \sum X_i Y_i - (\sum X_i)(\sum Y_i)}{\sqrt{[n \sum X_i^2 - (\sum X_i)^2][n \sum Y_i^2 - (\sum Y_i)^2]}} \quad (15)$$

Para este caso, n é o tamanho da amostra, X_i é o valor da variável e Y_i é o valor do índice.

O autor deste trabalho propõe que, como houve uma condensação de várias variáveis em um único número, é necessário averiguar o grau de associação entre os índices propostos e as variáveis que compõem o banco de dados mais importantes (importância das variáveis é subjetiva, normalmente obtida através de especialistas), e para isso devem-se obter as correlações entre os índices e as variáveis.

Mediu-se então a correlação entre os índices propostos e as variáveis que compõem o banco de dados, e os resultados estão nas tabelas 20 e 21.

Tabela 20 Correlação entre as variáveis e os índices do modelo de desenvolvimento

	RANKS		PADRONIZADO
	Varimax e Equimax	Quartimax	Varimax, Quartimax e Equimax
Área	-0,488	-0,033	-0,022
População	0,766 *	0,695 *	0,463
IDH	0,713 *	0,679 *	-0,072
PIB	0,529	0,617 *	-0,576
Veículos	0,844 *	0,756 *	0,416
Indústria	0,786 *	0,325	0,403
Comércio	0,801 *	0,637 *	0,510
Construção Civil	0,813 *	0,401	0,467
Serviços	0,854 *	0,409	0,296
Agropecuária	-0,004	0,582	-0,031
VDM	0,842 *	0,496	0,436
Área Pavimentada	0,829 *	0,405	0,276
Área não Pavimentada	-0,462	-0,412	-0,173

* Correlação é significante a 1%

Tabela 21 Correlação entre as variáveis e os índices do modelo de hierarquização

	RANKS			PADRONIZADO		
	Varimax	Quartimax	Equimax	Varimax	Quartimax	Equimax
Área	0,147	0,120	0,161	-0,216	-0,218	-0,215
População	-0,081	-0,069	-0,082	0,937 *	0,936 *	0,936 *
IDH	-0,022	0,161	-0,076	0,466	0,466	0,469
PIB	0,419	0,516	0,360	0,286	0,287	0,289
Veículos	-0,142	-0,061	-0,157	0,938 *	0,940 *	0,937 *
Indústria	-0,074	-0,087	-0,077	0,774 *	0,782 *	0,771 *
Comércio	-0,054	-0,081	-0,039	0,900 *	0,898 *	0,901 *
Construção Civil	-0,347	-0,282	-0,357	0,113	0,117	0,114
Serviços	-0,163	-0,118	-0,176	0,197	0,210	0,195
Agropecuária	0,359	0,436	0,332	0,315	0,315	0,316
VDM	-0,104	-0,107	-0,100	0,849 *	0,854 *	0,848 *
Área Pavimentada	-0,279	-0,234	-0,289	0,009	0,017	0,009
Área não Pavimentada	0,164	0,114	0,180	-0,321	-0,319	-0,323

* Correlação é significativa a 1%

Observando as tabelas 20 e 21, quando as variáveis são organizadas por *ranks*, nenhuma correlação é significativa à 1% entre as variáveis observadas e os índices gerados pelo modelo de hierarquização. E quando as variáveis estão padronizadas pela distribuição Normal, nenhuma correlação é significativa à 1% entre as variáveis observadas e os índices gerados pelo modelo de desenvolvimento.

O índice que possui o maior número de correlações significativas a 1% é aquele que foi gerado através do modelo de desenvolvimento, para variáveis organizadas por *ranks* e cuja análise fatorial foi realizada utilizando-se o método de rotação dos fatores Varimax ou Equimax, onde de 13 variáveis no total possui correlação significativa à 1% com 9 destas variáveis.

3.6 *Discussão dos Resultados*

Os resultados apresentados neste capítulo mostram que a criação de índices através do uso de análise fatorial confrontando com as propriedades dos números-índices são atendidas para quase todos os índices propostos.

A escolha do melhor índice pode ser considerada subjetiva, pois depende dos objetivos do estudo que se está realizando e da opinião de especialistas. Neste exemplo de aplicação deseja-se um índice que represente a potencialidade de cada Distrito Operacional do estado do Rio Grande do Sul, podendo este índice ser utilizado como um fator multiplicador de um valor pré-fixado para a comercialização de um produto ou implantação de um serviço.

De acordo com especialistas que foram consultados, a classificação quanto a importância de cada variável pode ser visualizada na tabela 22.

Tabela 22 Classificação das variáveis de acordo com especialistas

VARIÁVEIS	Média
ÁREA	3,73
POPULAÇÃO	7,73
IDH	8,16
PIB	6,86
VEÍCULOS	7,18
INDÚSTRIA	7,67
COMÉRCIO	6,90
CONSTRUÇÃO CIVIL	4,30
SERVIÇO	5,42
AGROPECUÁRIA	4,06
VDM	13,15
ÁREA PAVIM.	16,19
ÁREA NÃO PAVIM.	8,66
TOTAL	100

Nesta tabela, pode-se observar que as variáveis que possuem as maiores notas são, respectivamente, área pavimentada, VDM e PIB. Observando agora as tabelas 20 e 21, nota-se que o índice gerado através do modelo de desenvolvimento, para variáveis organizadas por *ranks* e em cuja análise fatorial utilizou-se o método de rotação dos fatores Varimax ou Equimax, possui correlação significativa à 1% com as duas primeiras variáveis de maiores notas. Já o mesmo modelo de índice, só que para rotação de fatores Quartimax, possui correlação significativa à 1% com a terceira variável com maior nota.

Tabela 23 Propostas para um índice de potencialidade dos DOPs do RS

DOP	1ª PROPOSTA DE ÍNDICE DE POTENCIALIDADE	2ª PROPOSTA DE ÍNDICE DE POTENCIALIDADE
1º Esteio	0,8063	0,7219
2º Bento Gonçalves	0,8927	0,8552
3º Santa Cruz do Sul	0,5161	0,4250
4º Santa Maria	0,4281	0,5198
5º Cruz Alta	0,4656	0,5490
6º Passo Fundo	0,7146	0,6885
7º Pelotas	0,4594	0,5281
8º Bagé	0,2927	0,5156
9º Alegrete	0,2990	0,4979
10º Cachoeira do Sul	0,2276	0,2875
11º Lajeado	0,7083	0,5104
12º Santiago	0,2979	0,4354
13º Erechim	0,4792	0,4000
14º Santa Rosa	0,6333	0,5635
15º São Francisco de Paula	0,3724	0,3010
16º Osório	0,4042	0,2885
17º Palmeira das Missões	0,5026	0,4125

Assim, o índice para representar a potencialidade de cada Distrito Operacional do RS

pode ser um dos dois acima citados, sendo que a escolha final depende da análise dos especialistas que poderão considerar outros aspectos subjetivos, mas tomando como base a avaliação sugerida na análise fatorial. A Tabela 23 apresenta os dois índices que podem ser considerados como sendo o Índice de Potencialidade.

4 COMENTÁRIOS FINAIS

4.1 Conclusões

Índices têm sido amplamente utilizados em diversas áreas para monitorar aspectos variados. Uma das propostas para criação de índices é através da utilização de uma análise fatorial.

Essa monografia abordou alguns métodos de criação de índices para dois tipos de variáveis (por *ranks* e padronizadas pela Normal) através do uso de análise fatorial. Na realização da análise fatorial é necessário escolher o método de estimação dos escores fatoriais e o método de rotação dos fatores, que influenciam na formação dos fatores. Observou-se que, independentemente de qual método de extração dos fatores for escolhido, os métodos de estimação dos escores fatoriais geraram os mesmos resultados, deixando, assim, a escolha livre para qualquer um deles para este exemplo de aplicação. Utilizou-se três tipos de rotação ortogonal de fatores para gerar os índices, para os dois tipos de variáveis, e, observando as correlações entre as variáveis e os índices e também a opinião de especialistas quanto a importância de cada variável, apresentou-se duas propostas para um índice de potencialidade.

Existem várias situações na elaboração de números-índices onde são perfeitamente aplicáveis os conceitos de análise fatorial, o que constitui numa contribuição deste trabalho.

Para ilustrar a pesquisa, a monografia apresentou um exemplo de aplicação onde é avaliada a potencialidade de cada DOP do Rio Grande do Sul através da elaboração de um índice criado a partir das técnicas vistas na revisão bibliográfica. Após realizar algumas considerações a respeito dos índices criados, das correlações obtidas e da opinião de

especialistas, escolheu-se duas propostas de índices para exemplificar um índice potencial, um deles gerado através do modelo de desenvolvimento, realizando uma rotação Varimax ou Equimax para variáveis organizadas por *ranks*, e o outro é obtido através do mesmo modelo e do mesmo tipo de variáveis, alterando apenas o método de rotação dos fatores para Quartimax.

4.2 Sugestões para trabalhos futuros

O assunto abordado nessa monografia dispõe de uma ampla quantidade de técnicas, pois existem diversos métodos de extração de rotação dos fatores e também um leque muito grande de índices que podem ser criados a partir do uso de análise fatorial. Entre essas técnicas, muitas são recentes e surgiram nos últimos anos.

Uma grande quantidade de estudos pode ser realizada, onde utilização de análise fatorial para criação de índices não são muito comuns. Como sugestões para trabalhos futuros, destacam-se:

- Criação de índices que trabalhem com variáveis observadas ao longo do tempo, utilizando a análise fatorial;
- Utilização de métodos de rotação oblíqua de fatores para a obtenção de um índice potencial;
- O desenvolvimento de novas técnicas e procedimentos ou aprofundamento nas técnicas existentes;
- Um comparativo entre as novas técnicas com as abordadas nesta monografia.

REFERÊNCIAS BIBLIOGRÁFICAS

- BRAVO, R. S. *Análisis Estadístico Multivariable*. Madrid, España. Editorial Paraninfo, 1994.
- CATELL, R. B. *Factor Analysis: An Introduction and Manual for the Psychologist and Social Scientist*. Connecticut, USA. Greenwood Press, 1977.
- CHATFIELD, C., COLLINS, A. J. *Introduction to Multivariate Analysis*. New York, USA. Chapman and Hall, 1980.
- ENDO, S. K. *Métodos Quantitativos - Números Índices*. São Paulo, Brasil. Atual Editora Ltda., 1988.
- FACHEL, J. M. G. *Análise Fatorial*. São Paulo, Dissertação de Mestrado – Universidade de São Paulo, 1976.
- FONSECA, J. *Estatística Aplicada*. São Paulo, Brasil. Atlas, 1985.
- HAIR, J. F., ANDERSON, R. E., TATHAM, R. L., BLACK, W. C. *Multivariate Data Analysis*. New Jersey, USA. Prentice Hall, 1995.
- JOHNSON, R. A., WICHERN, D. W. *Applied Multivariate Statistical Analysis*. Washington, USA. Prentice Hall, 1988.
- KENDALL, M. G. Factor *Analysis as a Statistical Technique*. J.R. Statist. Soc., B, 12, 60-73.
- KIM, J. O., MUELLER, C. W. *Factor Analysis: Statistical Methods and Practical Issues*. Iowa, USA. Sage Publications, 1978.
- LATIF, S. A. *A Análise Fatorial Auxiliando a Resolução de um Problema Real de Pesquisa de Marketing*. Caderno de Pesquisas em Administração, v. 00, n. 0, p. 1-10. 1994.

MEYER, P. L. *Probabilidade: Aplicações à Estatística*. Washington, USA. LTC Editora, 1983.

MILONE, G. *Estatística Geral e Aplicada*. São Paulo, Brasil. Pioneira Thomson Learning, 2004.

MULAİK, S. A. *The Foundations of Factor Analysis*. Georgia, USA. McGraw-Hill, 1972.

PEREIRA, J. C. R. *Análise de Dados Qualitativos: Estratégias Metodológicas para as Ciências da Saúde, Humanas e Sociais*. São Paulo, Brasil. EDUSP, 2001.

POSSOLI, S. *Análise Multivariada*. Porto Alegre. Cadernos de Matemática e Estatística da UFRGS, 1992.

PRESS, J. *Applied Multivariate Analysis*. New York, USA. Holt, Rinehart and Winston, 1972.

SWISHER, L. L.; BECKSTEAD, J. W.; BEBEAU, M. J. *Factor Analysis as a Tool for Survey Analysis Using a Professional Role Orientation Inventory as an Example*. Journal of Marketing Research, v. 84, n. 9, p. 784-799, 2004.