

JONAS ANTONIO VIEIRA JUNIOR

**ESTATÍSTICA MULTIVARIADA APLICADA A DADOS SOCIOECONÔMICOS  
PARA IDENTIFICAÇÃO DE TIPOLOGIAS ESPACIAIS: ESTUDO DE CASO DOS  
MUNICÍPIOS DE MINAS GERAIS - 2010**

BELO HORIZONTE  
2018

JONAS ANTONIO VIEIRA JUNIOR

**ESTATÍSTICA MULTIVARIADA APLICADA A DADOS SOCIOECONÔMICOS  
PARA IDENTIFICAÇÃO DE TIPOLOGIAS ESPACIAIS: ESTUDO DE CASO DOS  
MUNICÍPIOS DE MINAS GERAIS - 2010**

Monografia apresentada ao Curso de Especialização em Estatística Aplicada do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Especialista em Estatística Aplicada.

Orientadora: Profa. Dra. Sueli Aparecida Mingoti

BELO HORIZONTE  
2018

## RESUMO

A proposta desta monografia é propor e discutir uma contribuição ao estudo de tipologias espaciais dos municípios que compõem o Estado de Minas Gerais a partir de métodos de Estatística Multivariada, aplicados sobre a base de dados socioeconômicos disponibilizada pelo PNUD por meio do Atlas de Desenvolvimento Humano Brasileiro referente ao ano de 2010. O trabalho está estruturado em duas perspectivas: a primeira consiste em sintetizar as principais variáveis socioeconômicas disponíveis no banco de dados em índices estatisticamente representativos para a amostra; o segundo consiste no uso destes índices na formação de grupos de municípios com perfil socioeconômico semelhante. Após essa classificação são realizadas as análises de perfis dos grupos com uso de estatística exploratória. De modo complementar, utiliza-se Sistema de Informações Geográficas (SIG) para espacializar a estrutura formada, observando a distribuição espacial e localização dos grupos.

**Palavras-chave:** Estatística Multivariada; Geoprocessamento; Agrupamento; Cluster.

## SUMÁRIO

|          |  |    |
|----------|--|----|
| 1        | INTRODUÇÃO .....   | 8  |
| 2        | OBJETIVOS GERAL E ESPECÍFICOS .....  | 10 |
| 3        | REFERENCIAL TEÓRICO .....  | 10 |
| 3.1      | As Características da área objeto de Estudos: Minas Gerais .....           | 10 |
| 3.2      | Teoria dos Lugares Centrais: O modelo de Christaller.....                  | 13 |
| 3.3      | Classificação e Hierarquização Regional: o caso brasileiro .....           | 14 |
| 3.4      | Quantificação em Geografia e Estatística Multivariada .....                | 16 |
| 3.5      | Geoprocessamento, SIG e Análise Espacial .....                             | 18 |
| 4        | METODOLOGIA ESTATÍSTICA .....  | 21 |
| 4.1      | Estatística Multivariada: Conceitos e Definições .....                     | 22 |
| 4.2      | Análise Fatorial Ortogonal.....  | 23 |
| 4.2.1.   | Estimação do modelo de Análise Fatorial Ortogonal.....                     | 25 |
| 4.2.1.1. | Estimação do número de fatores.....  | 25 |
| 4.2.1.2. | Estimação das cargas fatoriais pelo método das componentes principais..... | 26 |
| 4.2.2.   | Avaliação da qualidade de ajuste do modelo .....                           | 26 |
| 4.2.3.   | Interpretação dos fatores .....  | 27 |
| 4.2.4.   | Rotação ortogonal dos fatores .....  | 27 |
| 4.2.5.   | Cálculo dos escores dos fatores.....                                       | 29 |
| 4.3      | Técnicas em Estatística Multivariada: Análise de Agrupamentos.....         | 30 |
| 5        | DESCRIÇÃO E ANÁLISE DOS RESULTADOS .....                                   | 33 |
| 5.1      | Construção do Banco de Dados e Seleção das Variáveis .....                 | 33 |
| 5.2      | Análise Descritiva dos Dados.....  | 35 |
| 5.3      | Análise Fatorial .....   | 37 |
| 5.3.1    | Estimação do número de fatores e cargas fatoriais.....                     | 37 |
| 5.3.2    | Interpretação dos fatores.....   | 39 |
| 5.4      | Análise de <i>Cluster</i> .....  | 50 |
| 5.4.1    | Estimação do número de grupos e respectivas composições .....              | 50 |
| 5.4.2    | Análise de perfil dos grupos de acordo com os escores dos fatores.....     | 54 |
| 6        | CONSIDERAÇÕES FINAIS .....   | 61 |
|          | <b>REFERÊNCIAS BIBLIOGRÁFICAS</b> .....                                    | 63 |
|          | <b>APÊNDICE</b> .....  | 66 |

## LISTA DE TABELAS E QUADROS

|  |    |
|--|----|
| QUADRO 1: Municípios cuja população ultrapassa 1 milhão de habitantes (IBGE, 2015).....  | 14 |
| QUADRO 2: Variáveis selecionadas para estudo .....   | 34 |
| QUADRO 3: Modelo fatorial para três fatores – Método de Componentes Principais .....   | 39 |
| QUADRO 4: Coeficientes do Modelo fatorial para três fatores (usar tamanho de fonte 12 para título de Quadros e tabelas) .....                    | 45 |
| QUADRO 5: Valores referentes ao município de Belo Horizonte – originais e padronizados.....  | 46 |
| QUADRO 6: Estatísticas descritivas dos escores em cada fator.....  | 46 |
| QUADRO 7: Modelo de Agrupamento baseado em escores resultados da análise fatorial ortogonal, distância euclidiana, método de ligação média ..... | 51 |
| QUADRO 8: Partição final – k=11 .....  | 52 |
| QUADRO 9: Estatísticas descritivas para o Fator 1 (Partição n=11) .....  | 55 |
| QUADRO 10: Estatísticas descritivas para o Fator2 (Partição n=11) .....  | 57 |
| QUADRO 11: Estatísticas descritivas para o Fator3 (Partição n=11) .....  | 58 |
| <br>   |    |
| TABELA 1: Estatísticas descritivas das variáveis originais .....   | 35 |
| TABELA 2: Aplicação dos critérios para determinação do número de grupos na partição ....   | 51 |

## LISTA DE GRÁFICOS E FIGURAS

|   |    |
|---|----|
| GRÁFICO 1: Autovalores da matriz de correlação por número de fatores. <i>Factor number</i> representa o número de fatores e <i>eigenvalue</i> representa os autovalores. .... | 38 |
| GRÁFICO 2: Cargas fatoriais das variáveis nos Fatores 1 e 2 .....   | 49 |
| GRÁFICO 3: Dendrograma – partição constituída de 11 grupos .....  | 53 |
| GRÁFICO 4: Boxplot dos escores no primeiro fator – 11 grupos .....  | 55 |
| GRÁFICO 5: Boxplot dos escores no segundo fator – 11 grupos .....   | 57 |
| GRÁFICO 6: Boxplot dos escores no terceiro fator – 11 grupos.....   | 58 |
| GRÁFICO 7: Gráfico de Dispersão dos escores por grupo – Fatores 1 e 2.....  | 59 |
| GRÁFICO 8: Gráfico de Dispersão dos escores por grupo – Fatores 1 e 3.....  | 59 |
| GRÁFICO 9: Gráfico de Dispersão dos escores – Fatores 2 e 3.....  | 60 |
| GRÁFICO 10: Gráfico Box-Plot do Índice de Desenvolvimento Humano em MG, 2010.....   | 60 |
|   |    |
| FIGURA 1: Mapa de Localização do Estado de Minas Gerais e seus municípios .....   | 12 |
| FIGURA 2: Mapa dos escores para Fator 1 .....   | 40 |
| FIGURA 3: Mapa dos escores para Fator 2 .....   | 41 |
| FIGURA 4: Mapa dos escores para Fator 3 .....   | 43 |
| FIGURA 5: Mapa da variável T_DES18M padronizada .....   | 43 |
| FIGURA 6: Histograma e boxplot dos escores por Fatores .....  | 48 |
| FIGURA 7: Histograma e boxplot dos valores padronizados da variável T_DES18M.....   | 49 |
| FIGURA 8: Espacialização da partição para 11 grupos – coleção de mapas.....   | 54 |

## LISTA DE ABREVIACÕES

|       |  |
|-------|--|
| ACP   | Análise de Componentes Principais                              |
| BDGEO | Bancos de Dados Geográficos                                    |
| GIS   | <i>Geographic Information System</i>                           |
| IBGE  | Instituto Brasileiro de Geografia e Estatística                |
| IDH   | Índice de Desenvolvimento Humano                               |
| IDHM  | Índice de Desenvolvimento Humano Municipal                     |
| IPEA  | Instituto de Pesquisa Econômica Aplicada                       |
| PCA   | <i>Principal Component Analysis</i>                            |
| SIG   | Sistemas de Informação Geográfica                              |
| SSB   | Soma de Quadrados Total Entre Grupos                           |
| SSTc  | Soma de Quadrados Total  |
| SSW   | Soma de Quadrados Total Dentro dos Grupos formados na partição |

## 1 INTRODUÇÃO

A busca pela compreensão do padrão de distribuição espacial da população e suas características tem sido objeto de estudo essencial à Geografia, pois compõem a dinâmica demográfica e possibilita sua gestão pelos pesquisadores interessados nos temas vinculados a hierarquia urbana, redes e sistemas regionais. Todavia, frente à dificuldade em definir parâmetros pré-determinados que sejam aplicáveis como modelos e a intensa dinâmica populacional, a caracterização destes sistemas é considerada atividade de grande complexidade, especialmente para o caso brasileiro (AMORIM FILHO e RIGOTTI, 2002).

Para viabilizar a classificação de cidades, autores como AMORIM FILHO e RIGOTTI (2002) destacaram que os limiares demográficos – apesar de extremamente importantes – não são suficientes para representar adequadamente a função e hierarquia urbana regional. Para suprir essa lacuna, destaca-se o uso de variáveis associadas à capacidade de modelar as diversas características socioeconômicas da população objeto de estudo, as quais diferem em termos estatísticos e locacionais.

Neste sentido, faz-se necessário o uso de métodos cuja capacidade de síntese permita abordar a gama de informações disponíveis, assegurando a representatividade das variáveis e proporcionando maior uso do potencial de análise. Conforme MINGOTI (2005), o conjunto de métodos estatísticos empregados em situações nas quais diversas variáveis são medidas simultaneamente para cada amostra compõe o que conhecemos como Estatística Multivariada.

Dentre as diversas aplicações, MINGOTI (2005) cita a construção de índices aplicados na hierarquização de elementos, onde estudos buscam classificar a qualidade e o desempenho das amostras integradas ao universo de dados. De modo semelhante, apresenta a segmentação do conjunto de dados em diferentes grupos. Assim, cada grupo é formado por elementos com características semelhantes nas variáveis observadas que, simultaneamente, permitem o delineamento de perfis representativos dentro do banco de dados em estudo.



Face à dificuldade em estabelecer critérios e, a partir destes, construir classificações entre as cidades, busca-se com esta monografia contribuir para o avanço dos estudos de regionalização geográfica fundamentados nos métodos de Estatística Multivariada, conhecidos por Análise Fatorial e Análise de Agrupamentos. Neste sentido, segmentamos o estudo em duas fases complementares. A primeira consiste na criação de índices sobre as variáveis sugeridas em estudos de regionalização territorial, aumentando a capacidade de análise por meio da síntese de variáveis em novos indicadores. A segunda é constituída pela formação de grupos de municípios semelhantes conforme índices construídos na etapa anterior.

Optamos por abordar os 853 municípios pertencentes ao Estado de Minas Gerais como universo amostral. As variáveis socioeconômicas foram extraídas do Atlas de Desenvolvimento Humano no Brasil referente ao ano de 2010 e consolidadas num banco de dados geográficos, onde cada registro corresponde a uma geometria (município) a qual, por sua vez, compreende uma linha na tabela de dados. A adoção destas premissas é justificada pela diversidade regional em Minas Gerais, perceptível até mesmo no padrão de variabilidade dos territórios de seus municípios.

Neste sentido, entendemos que estudos nessa linha podem contribuir para a melhor compreensão de sua dinâmica e direcionamento de políticas públicas regionais, contribuindo para a compreensão e intervenção na gestão territorial a partir de dados estatísticos. Adicionalmente, frente à intensa dinâmica populacional – principalmente migratória – e gama de dados disponíveis sobre a temática socioeconômica, este estudo mostra-se como potencial de relevante contribuição, pois favorece o poder e celeridade de análise a partir de sua síntese e agrupamento, preservando a representatividade e veracidade dos mesmos.

O restante deste trabalho está organizado da seguinte forma: na seção 2, apresenta-se os objetivos gerais e específicos; na seção 3, os fundamentos teóricos necessários para construção deste estudo; a seção 4 discorre acerca da metodologia e modelagem do banco de dados; apresentamos e discutimos os resultados na seção 5 e, finalmente, tecemos as considerações finais na seção 6.

## **2 OBJETIVOS GERAL E ESPECÍFICOS**

Este trabalho tem como principal objetivo apresentar uma proposta de agrupamento de unidades espaciais semelhantes para os 853 municípios de Minas Gerais a partir da aplicação de métodos estatísticos multivariados utilizando-se variáveis socioeconômicas.

Para tanto, torna-se necessário seguir pelos objetivos específicos elencados a seguir:

1. Construir índices a partir das variáveis disponíveis para a proposta de regionalização;
2. Apresentar os padrões de distribuição espacial dos índices encontrados por meio do uso de técnicas de geoprocessamento;
3. Apresentar a caracterização socioeconômica dos municípios agrupados.

## **3 REFERENCIAL TEÓRICO**

Nesta seção são apresentados os fundamentos teóricos que embasaram o desenvolvimento desta monografia. A caracterização da área objeto do estudo é apresentada na seção 3.1. Teorias de cunho geográfico e seus principais desafios são apresentados nas seções 3.2 e 3.3. Posteriormente, a seção 3.4 retrata conceitualmente os fundamentos estatísticos necessários a esta monografia. A seção 3.5 refere-se exclusivamente às técnicas de geoprocessamento aplicadas para o mapeamento e análise dos resultados.

### **3.1 As Características da área objeto de Estudos: Minas Gerais**

Situado na região sudeste do território brasileiro e assentado sobre 586.520,732 km<sup>2</sup>, Minas Gerais é o segundo estado mais populoso do país com 19.597.330 habitantes, abaixo apenas de São Paulo com 41.262.199, acima dos estados Rio de Janeiro e Bahia, com 15.989.929 e 14.016.906 habitantes, respectivamente. Em 2017, a população estimada ultrapassou os 20 milhões de habitantes IBGE (2010). Ocupa também a segunda posição no ranking nacional quando observadas as condições de domicílios particulares com situação inadequada de saneamento básico, abaixo apenas do estado da

Bahia. Mantém ainda a vice-liderança no número de matrículas de alunos nos níveis de educação básica, abaixo apenas de São Paulo.

Contudo, Minas Gerais ocupa 8ª posição quando observamos o percentual de habitantes residentes em área urbana com 85,3% de sua população. Esta é, em sua maioria, composta por adultos e idosos – situado entre as 10 primeiras posições para percentual de habitantes acima de 40 anos. Em contrapartida, não chega à vigésima posição quando o grupo etário está entre 0 e 24 anos (IBGE, 2017).

Em 2010, foi considerado o oitavo Estado com maior rendimento nominal mensal domiciliar per capita (R\$ 1.168) e 9º melhor Índice de Desenvolvimento Humano (IDH) nacional (0,731). O índice de Gini para o Estado, segundo IBGE (2010) é 0,46, refletindo a 10ª posição no quesito grau de concentração de renda em comparação à 3ª posição ocupada no volume de receitas orçamentárias realizadas.

Em termos geomorfológicos, Minas Gerais é formado por extensos alinhamentos de serras e característicos “mares de morros” e o Quadrilátero Ferrífero – região onde predomina a atividade minerária – alternando suas paisagens para o planalto central entalhado por uma das principais bacias hidrográficas do país: o rio São Francisco. Reflexo dessa formação pode ser observado na diversidade de formações vegetais ao longo do território. Apesar da predominância do bioma Cerrado, a Mata Atlântica prevalece na região da Zona da Mata (porção sudeste) e a Caatinga consolidada na região norte e nordeste do Estado. O conjunto destas informações confirma a heterogeneidade contida em seu território e até mesmo na variação da dimensão territorial entre seus 853 municípios. Notadamente, existem regiões onde estes índices demonstram comportamento extremamente diferente do apresentado. Itamarandiba e Juiz de Fora apresentam, respectivamente, o índice de Gini em 0,47 e 0,43, assemelhando aos estados Rondônia (0,48) e Santa Catarina (0,41), o que sugere que elevada disparidade regional em Minas Gerais pode ser comparada aos extremos do território nacional.

A FIGURA 1 permite identificar a localização no contexto nacional, ilustra o relevo, permite o levantamento de hipóteses sobre um possível padrão de variabilidade espacial dos tamanhos dos municípios ao longo de sua extensão.

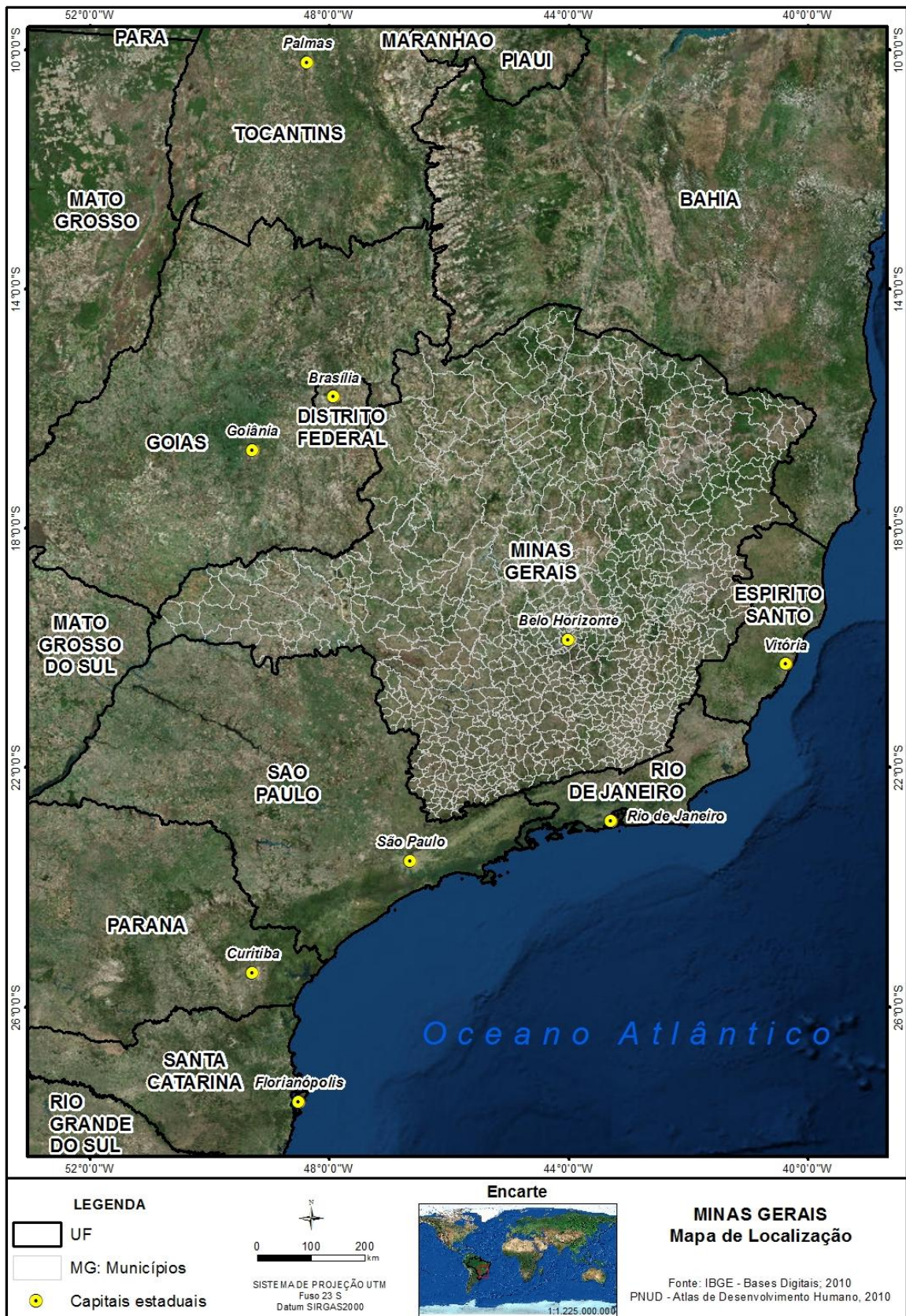


FIGURA 1: Mapa de Localização do Estado de Minas Gerais e seus municípios

### 3.2 Teoria dos Lugares Centrais: O modelo de Christaller

Um dos exemplos mais significativos e amplamente conhecidos pela abordagem científica em Geografia diz respeito ao estudo de povoamento, elaborado por Walter Christaller em 1933 na região sul da Alemanha. Segundo BRADFORD e KENT (1987), Christaller formulou uma teoria a partir de diversas premissas, que demonstraram a existência de padrões espaciais de povoamento. Estas premissas fundamentaram diversos conceitos, tais como centralidade, periferia e regiões.

Os modelos de análise regional apresentados até então refletiam a importância dos conceitos de sítio e situação das cidades nos estudos demográficos. Estes consideram as funções, origens e localização geográfica das cidades. Contudo, a abordagem de Christaller absorveu parte destes conceitos e expandiu sua interpretação para a posição relativa das cidades. Neste sentido, seria possível ponderar adequadamente a relevância e funções de cada cidade dentro de uma organização geral, composta por elementos hierárquicos e respectivas áreas de influência, corroborando para a compreensão da dinâmica regional (BRADFORD e KENT, 1987).

O principal objetivo da Teoria dos Lugares Centrais é “explicar a organização espacial das povoações e das áreas de influência, em particular a sua posição relativa e dimensão” (BRADFORD e KENT, 1987). Em linhas gerais, Christaller se baseou num modelo associado a dois conceitos fundamentais: o alcance ou raio de ação e o limiar mínimo. O alcance representa a distância que a pessoa está disposta a percorrer para usufruir de um bem. O limiar mínimo refere-se à demanda mínima necessária para que a oferta deste bem seja viável. Como resultado do trabalho de Christaller, foram sugeridas posições hierarquicamente associadas às cidades de determinada região:

- Centro de ordem superior;
- Centro de ordem média;
- Centro de ordem inferior;
- Centro de ordem mais baixa.

Apesar das diferenças características destes centros, existe relação intrínseca entre os mesmos. Esta relação é produto da interação entre os bens oferecidos – assim como seu alcance e limiar – e seus consumidores. Ademais, a dinâmica destas relações implica diretamente na construção do espaço geográfico e no surgimento de outros centros de ordem distinta, corroborando para a regionalização das cidades conforme propriedades específicas à estrutura hierárquica encontrada.

### 3.3 Classificação e Hierarquização Regional: o caso brasileiro

Os diversos fatos econômicos presenciados ao longo da história do Brasil tem causado efeito direto sobre a construção do espaço brasileiro. CARVALHO (2009) atenta para o número de cidades acima de um milhão de habitantes ter crescido de duas para cinco entre os anos 1960 e 1970 e para doze em 1991. Apresentamos no QUADRO 1 as informações divulgadas pelo Instituto Brasileiro de Geografia e Estatística (IBGE), onde este número alcançou 17 municípios no ano de 2015. É importante destacar que este aumento ocorreu de modo desigual, polarizado sobretudo na formação das principais metrópoles brasileiras e, conseqüentemente, resultou em mudanças nas características dos movimentos migratórios.

QUADRO 1: Municípios cuja população ultrapassa 1 milhão de habitantes (IBGE, 2015)

| <b>UF</b> | <b>MUNICÍPIO</b> | <b>POPULAÇÃO ESTIMADA</b> |
|-----------|------------------|---------------------------|
| SP        | São Paulo        | 11.967.825                |
| RJ        | Rio de Janeiro   | 6.476.631                 |
| BA        | Salvador         | 2.921.087                 |
| DF        | Brasília         | 2.914.830                 |
| CE        | Fortaleza        | 2.591.188                 |
| MG        | Belo Horizonte   | 2.502.557                 |
| AM        | Manaus           | 2.057.711                 |
| PR        | Curitiba         | 1.879.355                 |
| PE        | Recife           | 1.617.183                 |
| RS        | Porto Alegre     | 1.476.867                 |
| PA        | Belém            | 1.439.561                 |
| GO        | Goiânia          | 1.430.697                 |
| SP        | Guarulhos        | 1.324.781                 |
| SP        | Campinas         | 1.164.098                 |
| MA        | São Luís         | 1.073.893                 |
| RJ        | São Gonçalo      | 1.038.081                 |
| AL        | Maceió           | 1.013.773                 |



Em contrapartida ao aumento populacional heterogêneo, AMORIM FILHO e RIGOTTI (2002) apresentam o contexto teórico de surgimento de importantes conceitos regionalistas, dentre eles o polo de crescimento e as cidades médias, ambos associados às contribuições de teóricos como AMORIM FILHO (1976), AMORIM FILHO (1982) LAJUGIE (1971) e PERROUX (1961). Posteriormente, estes conceitos foram desmembrados em razão não somente do tamanho demográfico, mas também da posição geográfica, destacados por Christaller *apud* BRADFORD e KENT, 1987.

Tais conceitos refletem, direta ou indiretamente, os impactos causados pela dinâmica econômica sobre os aspectos como hierarquia, tipologia e funções dos centros urbanos na rede de cidades brasileiras (IPEA, 1999). No mesmo estudo, IPEA (1999) apresenta diversos critérios e notas metodológicas acerca das ponderações necessárias para construção da rede hierárquica urbana brasileira, associados diretamente aos indicadores selecionados sob os aspectos espaciais ou geográficos, sob o contexto da hierarquia do sistema urbano do Brasil, da tipologia de cidades e das caracterizações das funções urbanas.

Na busca por avanço na caracterização de um destes níveis denominado cidades médias, AMORIM FILHO (1976, *apud* AMORIM FILHO e RIGOTTI, 2002) aplicou os seguintes critérios funcionais:

- Interações constantes e duradouras tanto com seu espaço regional, quanto com aglomerações urbanas de hierarquia superior;
- Tamanho demográfico e funcional suficientes para que possam oferecer um leque bastante largo de bens e serviços ao espaço microrregional a elas ligado;
- Capacidade de receber e fixar os migrantes de cidades menores ou da zona rural, através do oferecimento de oportunidades de trabalho, funcionando, assim, como pontos de interrupção do movimento migratório na direção das grandes cidades, já saturadas;
- Condições necessárias ao estabelecimento de relações de dinamização com o espaço rural microrregional que as envolve;
- Diferenciação do espaço intra-urbano, com um centro funcional já bem individualizado e uma periferia dinâmica, evoluindo segundo um modelo bem parecido com o das grandes cidades, isto é, através da multiplicação de novos núcleos habitacionais periféricos;
- Aparecimento, embora evidentemente em menor escala, de certos problemas semelhantes aos das grandes cidades, como, por exemplo, a pobreza das populações de certos setores urbanos.

Entretanto, IPEA (1999) constata que “na realidade, as principais cidades brasileiras respondem, simultaneamente, por várias funções”, o que torna extremamente complexa a sua caracterização no sistema hierárquico urbano brasileiro.

AMORIM FILHO et al. (1982 *apud* AMORIM FLHO e RIGOTTI, 2002) aplicaram tais critérios e constataram, além da complexidade de caracterização das cidades médias, a existência de pelo menos quatro níveis hierárquicos, sendo:

- Grandes centros regionais: considerado um limite com o grupo das grandes cidades;
- Cidades médias de nível superior: abarcava cidades bem estruturadas com potencial de crescimento;
- Cidades médias propriamente ditas: caracterizado por cidades médias com funções bem definidas, principalmente pela intermediação aos polos regionais;
- Centros urbanos emergentes: definido como faixa de transição entre pequenas e as cidades médias.

Observa-se que o modelo de classificação regional prevê a descontinuidade espacial, também mencionada por BRUNET (1970 *apud* AMORIM FILHO e RIGOTTI, 2002) como descontinuidades em geografia, onde a variabilidade espacial dos diferentes grupos de cidades compõe a estrutura regional de modo coerente às funções e atributos municipais.

### 3.4 Quantificação em Geografia e Estatística Multivariada

O aumento do uso de métodos quantitativos associado aos estudos de cunho geográfico tem contribuído de modo expressivo para as análises espaciais, em concordância com a percepção de diversos autores, dentre eles GERARDI (1981), CARVALHO (2002) e CARVALHO e VIEIRA JUNIOR (2008).

SIMÃO (1999), SILVA (2002), BRAVO (2004) e SIMÕES (2005) apresentam exemplos de aplicação de métodos de Estatística Multivariada para fundamentar suas análises em diversas áreas, especialmente associados à análise regional. Dentre eles



podemos destacar o uso de análise fatorial na caracterização de grupos de cidades segundo seu potencial de crescimento (FAISSOL, 1978); a aplicação de técnicas multivariadas na descrição e redução dos domicílios paulistas segundo indicadores sociodemográficos (KAGEYAMA e LEONE, 1999); a dinâmica das cidades médias brasileiras (PEREIRA, 2003); a caracterização da rede urbana brasileira a partir da análise de agrupamentos (IPEA, 1999); a identificação de arranjos produtivos locais (CROCCO et al., 2003); a análise da inserção da Região Metropolitana de Belo Horizonte (RMBH) no contexto macrorregional brasileiro (SIMÕES et al., 2004) entre outros.

Denomina-se por Estatística Multivariada o conjunto de métodos estatísticos empregados em situações nas quais diversas variáveis são medidas simultaneamente para cada elemento amostral do banco de dados.

Entre as técnicas de Estatística Multivariada, a técnica de Análise Fatorial apresenta relevante potencial de aplicação, pois busca descrever a variabilidade original das variáveis em um número menor de variáveis aleatórias, denominadas fatores comuns que, adicionalmente, permitem “a rotação dos eixos (fatores) que sintetizam as informações contidas na matriz de dados” (HADDAD et al., 1989, *apud* SIMÕES, 2005). As cargas fatoriais que definem os fatores podem ser estimadas por meio de vários métodos, dentre eles a técnica de Análise de Componentes Principais (ACP), a qual objetiva explicar a estrutura de variância e covariância de um vetor aleatório por meio de combinações lineares (MINGOTI, 2005). Posteriormente, pode-se aplicar o recurso de rotação ortogonal sobre as cargas fatoriais que, apesar de não impactar na melhoria do ajuste do modelo, pode aperfeiçoar a visualização dos agrupamentos das variáveis nos respectivos fatores, acentuando as diferenças – caso existam – entre as cargas fatoriais.

Também conhecida como Análise de Conglomerados, Grupos ou *Cluster*, a Análise de Agrupamento consiste em formar grupos de acordo com a medida de similaridade entre os elementos amostrais contidos na base de dados. BAROUCHE e SAPORTA (1982 *apud* SIMÕES, 2005) esclarecem que:

O método de agrupamento (*cluster*) classifica seus elementos em grupos restritos homogêneos internamente, permitindo gerar estruturas agregadas significativas e desenvolver tipologias analíticas. Tais métodos podem ser hierárquicos – com sequências de partições em classes cada vez mais amplas – e não hierárquicos – com partições em números fixos de classes.

Aplicadas nesta monografia, as técnicas hierárquicas ainda são classificadas em aglomerativas e divisivas. Pelas técnicas aglomerativas, cada elemento amostral é considerado um único agrupamento – conhecido como *cluster* de tamanho 1 – e, em cada passo, estes elementos são agrupados até consolidar a partição final com um único grupo. Essa característica inerente ao processo hierárquico permite a construção do gráfico Dendrograma ou Dendograma, conforme TIMM (2002 *apud* MINGOTI, 2005). Apesar dos diferentes métodos disponíveis para agrupamentos hierárquicos, nesta monografia optamos por empregar o Método da Média das Distâncias (*Average Linkage*), o qual é baseado na média das distâncias entre todos os pares de elementos que podem ser formados entre os grupos comparados (SEIFODDINI, 1989).

Existem diversos outros métodos adotados como critério de distância entre os grupos com aplicações específicas em áreas distintas. Todavia, não é objetivo deste trabalho aprofundar no tema, mas fundamentar sua aplicação como forma de aporte aos métodos de regionalização geográfica. Para maiores detalhes e referências sobre o assunto, pode-se consultar MINGOTI (2005), JOHNSON e WICHERN (2002) e MANLY (1986).

Para concluir esta seção, faz-se importante ressaltar o uso do coeficiente de correlação linear de Pearson. Segundo Barbetta (2006), este coeficiente deve ser empregado para representar o grau de associação entre duas variáveis quantitativas, resultando num índice entre -1 e 1. Neste caso, quanto mais próximos dos extremos desse intervalo, maior a intensidade de associação (negativa ou positiva); quanto mais próxima de zero, menor a intensidade da relação linear entre as variáveis.

### 3.5 Geoprocessamento, SIG e Análise Espacial

Atualmente, a difusão das técnicas e aplicações das geotecnologias contribuiu para a consolidação dos conceitos em geoprocessamento. Segundo CÂMARA et al. (2001) “o termo Geoprocessamento denota a disciplina do conhecimento que utiliza técnicas matemáticas e computacionais para o tratamento da informação geográfica”. Neste

sentido, CÂMARA et al. (2001) sugerem que “Geoprocessamento é uma tecnologia interdisciplinar, que permite a convergência de diferentes disciplinas científicas para o estudo de fenômenos ambientais e urbanos”. Outros autores sugerem que o termo geoprocessamento evidencia o avanço da representação do espaço geográfico em ambiente computacional.

Em todos os casos, os autores abordam o uso da informação geográfica em sistemas informatizados. Carece destacar que se entende por “informação geográfica” toda e qualquer informação passível de localização, associada preferencialmente a um sistema de referência espacial conhecido onde os pares de coordenadas permitem, com precisão e acurácia, seu correto posicionamento (CÂMARA et al., 2001).

Considerado como um dos três principais pilares do geoprocessamento, podemos destacar os Sistemas de Informação Geográfica (SIG ou GIS<sup>1</sup>). Impulsionados pela revolução teórico-quantitativa, tais sistemas apareceram de modo muito embrionário nos anos 50, por meio de tentativas de automatização de dados com características espaciais. A década de 60 é considerada o surgimento dos primeiros SIG, com objetivo de auxiliar no diagnóstico de recursos ambientais em esfera governamental. Com o avanço nos recursos de *hardware* e desenvolvimento dos fundamentos matemáticos relacionados à geometria computacional ocorridos na década de 70, os SIG tornaram-se mais acessíveis e abriram caminho para novos sistemas. A década de 80 é considerada o ponto de crescimento acelerado das tecnologias SIG, o qual pode ser constatado até os dias atuais, associado principalmente pela difusão dos computadores pessoais e implantação de sistemas gerenciadores de bancos de dados (CÂMARA et al., 2001). As últimas décadas são marcadas pela intensa disseminação desta tecnologia e sua aplicação em diversas áreas, tais com agronomia, engenharia, avaliação de imóveis, planejamento urbano e regional, geologia, dentre outras.

Assim como o conceito de Geoprocessamento, diversos autores apresentaram uma definição para SIG. BURROUGH (1986) sugere que o SIG é “(...) conjunto poderoso de ferramentas para coletar, armazenar, recuperar, transformar e visualizar dados sobre o mundo real”. Em vista disso, COWEN (1995) define SIG como “(...) um sistema de

---

<sup>1</sup> Em inglês, *Geographic Information System*.

suporte à decisão que integra dados referenciados espacialmente num ambiente de respostas a problemas”. Outros autores sugerem que o SIG consiste num banco de dados indexados espacialmente e destacam sua capacidade de integrar a informação georreferenciada.

Em todas as definições, dois principais pontos são elencados: a capacidade do SIG no georreferenciamento, manuseio e gerenciamento de informações espaciais e sua relevante contribuição no suporte à tomada de decisões e aumento do potencial de análise de dados – considerando agora sua posição geográfica. Para isso, faz-se necessário atentar para a modelagem conceitual de dados e sua estruturação em algoritmos capazes de representar a diversidade de conceitos no espaço geográfico (CÂMARA e MONTEIRO, 2001).

As principais saídas de um SIG consistem em relatórios, tabelas, gráficos e, especialmente, mapas. “Um mapa pode ser interpretado como uma metáfora do mundo real” (CÂMARA et. al, 2001). Principal ferramenta de estudo na ciência geográfica, o mapa corrobora para a compreensão e comportamento dos dados no espaço geográfico, sendo este um grande desafio em diversas áreas do conhecimento.

A crescente importância associada à análise de dados e extração de informações estratégicas para suporte à tomada de decisão tem apresentado expressivo crescimento nos últimos anos. Neste sentido, a preocupação com o comportamento espacial de fenômenos estudados se mostra como um notável diferencial, pois busca compreender a existência de padrão na distribuição de dados georreferenciados, sua extensão e até mesmo correlação com outras variáveis. Estes pontos representam a essência da análise espacial, a qual mensura as propriedades e relacionamentos ponderados pela localização, ou seja, “a ideia central é incorporar o espaço à análise” (CÂMARA *et al*, 2004).

Um estudo clássico de análise espacial foi realizado por John Snow, no século XIX. Em 1854, ocorria um surto de cólera na cidade de Londres. Em suma, John Snow demarcou a localização das pessoas infectadas pela doença através de sua residência e as bombas de água que abasteciam a cidade. Na época, a elevada concentração de pessoas em torno de algumas bombas levou à suspensão de uso das mesmas, visto a contaminação da

água nestes locais. Esta hipótese foi confirmada posteriormente com estudos que associaram, também, o lançamento de rejeitos em cursos d'água.

Em análise espacial, a taxonomia mais utilizada para caracterizar os problemas consiste em três abordagens. A primeira considera a representação de elementos pontuais, expressos por ocorrências identificadas por pares de coordenadas geográficas. Como exemplos, temos a ocorrência de crimes, árvores em meio urbano e capitais federais. A segunda consiste em superfícies contínuas, estimadas a partir de um conjunto de amostras pontuais – independente de seu padrão de distribuição. Neste caso, os métodos de amostragem e interpolação assumem posição central na discussão, pois podem comprometer a correta representação da realidade. A terceira e última abordagem trata-se de dados agregados em unidades de análise (áreas, polígonos), nas quais se supõe que o valor de determinada variável é único, indivisível e homogêneo em toda sua extensão (CÂMARA et al., 2004). Observa-se que o uso do SIG para análise espacial ocasiona na construção de um banco de dados geográficos, no qual as áreas são armazenadas como geometrias (polígonos) e vinculados às tabelas, onde são dispostos os atributos que caracterizam os elementos de análise. Como exemplo, cada coluna desta tabela pode representar as diversas variáveis socioeconômicas dos municípios representados por suas divisas em formato de polígonos.

Assim como na estatística, a análise espacial parte da análise exploratória. Neste caso, faz-se o uso de mapas para compreender as medidas de tendência central, a existência de valores discrepantes (*outliers*) e a identificação de possíveis padrões de distribuição espacial. Carece destacar a incorporação de métodos estatísticos para aprimoramento da análise espacial, os quais permitiram o surgimento de conceitos como dependência espacial, autocorrelação espacial e até mesmo inferência estatística para dados espaciais (CÂMARA et al., 2004).

#### **4 METODOLOGIA ESTATÍSTICA**

Nesta seção apresenta-se os métodos estatísticos aplicados neste estudo. A primeira subseção traz conceitos e definições básicos para compreensão das subseções seguintes, as quais abordam Análise Fatorial e Análise de *Cluster*.

#### 4.1 Estatística Multivariada: Conceitos e Definições

Para melhor apresentarmos os conceitos teóricos que fundamentam os estudos de análise espacial apresentados na seção 3, faz-se pertinente o uso de técnicas de estatística multivariada. Associado a este cenário, a organização do banco de dados espaciais corroborou para a percepção do caráter multivariado dos dados em questão, pois existem diferentes medidas associadas a cada unidade amostral contida no Banco de Dados Geográficos (BDGEO), neste caso, os municípios. O conjunto das  $p$  variáveis aleatórias que representam tais medidas são representadas na forma de um vetor aleatório, tal que

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix},$$

onde cada variável aleatória,  $X_i = 1, \dots, p$ , que compõe este vetor possui uma determinada distribuição de probabilidade.

O valor esperado do vetor  $\mathbf{X}$  é dado por

$$\boldsymbol{\mu} = E(\mathbf{X}) = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix},$$

sendo  $\mu_i = E(X_i), i = 1, 2, \dots, p$ .

A variância de  $X_i$  é denotada por  $\sigma_{ii}$  e a covariância de  $X_i$  e  $X_j$ , para  $i \neq j$ , é denotada por  $\sigma_{ij}$ ,  $i=1, 2, \dots, p, j=1, 2, \dots, p$ , sendo  $\sigma_{ij} = E[(X_i - \mu_i)(X_j - \mu_j)]$ . A matriz de covariâncias de  $\mathbf{X}$  é denotada por  $\boldsymbol{\Sigma}_{p \times p}$ , sendo  $\sigma_{ij}$  o componente da  $i$ -ésima linha e  $j$ -ésima coluna dessa matriz. Por sua vez, a matriz de correlações de  $\mathbf{X}$  é denotada por  $\mathbf{P}_{p \times p}$ , onde o componente da  $i$ -ésima linha e  $j$ -ésima coluna representa a correlação linear entre as variáveis  $X_i$  e  $X_j$  dada por

$$\rho_{ij} = 1 \text{ se } i = j \text{ e } \rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}} \text{ se } i \neq j, \text{ sendo } -1 \leq \rho_{ij} \leq 1 \text{ para todo } i \text{ e } j.$$

## 4.2 Análise Fatorial Ortogonal

No intuito de construir índices que proporcionem a análise e agrupamento das unidades amostrais, a Análise Fatorial Ortogonal se mostra como técnica potencialmente favorável, principalmente porque a construção dos fatores é feita de tal forma que seja mantida, ao máximo, a estrutura de variabilidade original dos dados. Em resumo, as  $p$  variáveis originais padronizadas são agrupadas em um conjunto de  $\mathbf{m}$  novas variáveis latentes não correlacionadas denominadas fatores,  $1 \leq m \leq p$ . Se o valor de  $\mathbf{m}$  for igual a 1, temos apenas um grupo composto por todas as variáveis. De outro modo, se  $\mathbf{m}$  for igual à quantidade de variáveis, cada variável constituirá individualmente um único grupo. Deste modo, o objetivo é obter um valor de  $\mathbf{m}$  maior que 1 e menor que  $p$ .

Sejam  $\mathbf{X}$ ,  $\boldsymbol{\mu}$ ,  $\boldsymbol{\Sigma}_{p \times p}$  e  $\mathbf{P}_{p \times p}$  como definidos na seção 4.1. Denote por  $Z_i = (X_i - \mu_i)/\sqrt{\sigma_{ii}}$  a variável aleatória  $X_i$  padronizada,  $i=1, 2, \dots, p$ . Nesta monografia, assumimos a matriz de correlação das variáveis aleatórias observadas para construção do modelo de análise fatorial, a qual presume o relacionamento linear das variáveis padronizadas para identificar os fatores. Para este modelo, temos:

$$\begin{array}{rcllcl} Z_1 & = & l_{11}F_1 & + & l_{12}F_2 & + \dots & + & l_{1m}F_m & + & \varepsilon_1 \\ Z_2 & = & l_{21}F_1 & + & l_{22}F_2 & + \dots & + & l_{2m}F_m & + & \varepsilon_2 \\ \vdots & & \vdots & & \vdots & \vdots & \vdots & \vdots & & \\ Z_p & = & l_{p1}F_1 & + & l_{p2}F_2 & + \dots & + & l_{pm}F_m & + & \varepsilon_p \end{array}$$

onde  $Z_i$  é a variável  $X_i$  padronizada,  $l_{ij}$  é a carga fatorial da variável  $i$  no fator  $F_j$ ,  $\varepsilon_i$  é o erro aleatório,  $i=1, 2, \dots, p$ ,  $j=1, 2, \dots, \mathbf{m}$ , sendo  $\mathbf{m}$  o número de fatores no modelo,  $1 \leq \mathbf{m} \leq p$ .

Em sua forma matricial, o modelo de Análise Fatorial Ortogonal é representado por

$$\mathbf{D}(\mathbf{X}-\boldsymbol{\mu})=\mathbf{L} \mathbf{F} + \boldsymbol{\varepsilon}$$

onde

$$(\mathbf{X}-\boldsymbol{\mu})_{px1} = \begin{bmatrix} X_1-\mu_1 \\ X_2-\mu_2 \\ \vdots \\ X_p-\mu_p \end{bmatrix} \quad \boldsymbol{\varepsilon}_{px1} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_1 \\ \vdots \\ \varepsilon_p \end{bmatrix} \quad \mathbf{F}_{mx1} = \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_m \end{bmatrix} \quad \mathbf{L}_{pxm} = \begin{bmatrix} l_{11} & l_{12} & \dots & l_{1m} \\ l_{21} & l_{22} & \dots & l_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ l_{p1} & l_{p2} & \dots & l_{pm} \end{bmatrix},$$

$$\mathbf{D}_{pxp} = \begin{bmatrix} 1/\sqrt{\sigma_{11}} & 0 & 0 & \dots & 0 \\ 0 & 1/\sqrt{\sigma_{22}} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \dots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1/\sqrt{\sigma_{pp}} \end{bmatrix}.$$

Assim, temos os  $\mathbf{m}$  fatores aleatórios (variáveis latentes) armazenados no vetor  $\mathbf{F}_{mx1}$ . Os erros de medida e variação de  $Z_i$  não explicados pelos fatores comuns  $F_j, j=1, \dots, m$ , são armazenados no vetor aleatório  $\boldsymbol{\varepsilon}_{px1}$ . As cargas fatoriais são coeficientes armazenados na matriz  $\mathbf{L}_{pxm}$  e representam a correlação que os  $\mathbf{m}$  fatores tem com cada variável  $Z_i$ , isto é  $l_{ij} = \text{corr}(Z_i, F_j)$  e permitem a identificação das variáveis que estão mais correlacionadas com cada fator.

As suposições do modelo de Análise Fatorial Ortogonal são:

- Os fatores são variáveis aleatórias com média igual a zero e variância igual a 1; e são não correlacionados entre si (ortogonais);
- Os erros do modelo são tais que  $E[\boldsymbol{\varepsilon}_{px1}] = \mathbf{0}$ , ou seja,  $E[\varepsilon_i] = 0, i=1, \dots, p$ ;  $V[\varepsilon_i] = \psi_i$  e  $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0, \forall i \neq j$ , ou seja, os termos de erro são não-correlacionados e não precisam ter a mesma variância;
- Os fatores são não correlacionados com os erros do modelo.

A proporção da variabilidade total associada ao vetor aleatório  $\mathbf{Z} = (Z_1, \dots, Z_p)'$  que é explicada pelo fator  $F_j$  é dada por  $PFTE_{F_j} = \sum_{i=1}^p l_{ij}^2 / p$ , sendo que os fatores mais representativos para o modelo estão associados aos maiores valores desta quantidade.

Tanto o número de fatores,  $\mathbf{m}$ , quanto os fatores subjacentes ao modelo,  $\mathbf{F}_{mx1}$ , são desconhecidos *a priori* e, portanto, precisam ser estimados.



#### 4.2.1. Estimação do modelo de Análise Fatorial Ortogonal

Na prática, os parâmetros do modelo de análise fatorial são estimados a partir da matriz de correlação amostral. Essa matriz é dada por

$$\mathbf{R}_{p \times p} = \begin{bmatrix} 1 & r_{12} & r_{13} & \cdots & r_{1p} \\ r_{21} & 1 & r_{23} & \cdots & r_{2p} \\ r_{31} & r_{32} & 1 & \cdots & r_{3p} \\ \vdots & \vdots & \vdots & 1 & \vdots \\ r_{p1} & r_{p2} & r_{p3} & \cdots & 1 \end{bmatrix},$$

sendo  $r_{ij}$  a correlação linear amostral de Pearson entre as variáveis  $X_i$  e  $X_j$ ,  $i \neq j$ .

O primeiro passo no processo de estimação diz respeito à determinação do número de fatores,  $\mathbf{m}$ , e partir desse as cargas fatoriais  $L_{pxm}$  poderão ser estimadas por meio de métodos como o de componentes principais.

##### 4.2.1.1. Estimação do número de fatores

A construção do modelo de Análise Fatorial começa pela estimação do número de fatores a serem incluídos no modelo,  $\mathbf{m}$ . A determinação do valor de  $\mathbf{m}$  depende de alguns critérios que são baseados nos  $p$  autovalores extraídos da matriz de correlações amostrais, denotados por  $\hat{\lambda}_i$ ,  $i=1, \dots, p$ . Pode-se utilizar, por exemplo, o critério do percentual de variância total do vetor aleatório  $\mathbf{Z} = [Z_1, \dots, Z_p]'$  que é explicada pelos fatores incluídos no modelo ou o critério de Kaiser (KAISER, 1958).

No primeiro caso, o valor de  $\mathbf{m}$  é estimado como sendo aquele para o qual

$$\left( \sum_{i=1}^m \frac{\hat{\lambda}_i}{p} \right) \times 100 = \alpha,$$

sendo  $\alpha$  um valor pré-especificado pelo pesquisador,  $0 < \alpha < 100$ .

Com base no critério de Kaiser,  $\mathbf{m}$  é estimado como sendo o número de autovalores  $\hat{\lambda}_i$  maiores que 1. Os dois critérios, no entanto, podem ser utilizados em conjunto para uma melhor determinação do valor de  $\mathbf{m}$ .

#### 4.2.1.2. Estimação das cargas fatoriais pelo método das componentes principais

As estimação da matriz de cargas fatoriais pode ser realizada por vários métodos, dentre eles o de componentes principais o qual não exige informações ou suposições sobre a distribuição de probabilidades das variáveis  $\mathbf{Z} = (Z_1, \dots, Z_p)'$  (JOHNSON e WICHERN, 2002).

Para um dado valor de  $\mathbf{m}$ , a matriz de cargas fatoriais  $\mathbf{L}_{pxm}$  e a matriz de covariâncias dos erros do modelo  $\mathbf{\Psi}_{pxp} = \text{diag}(\psi_1, \dots, \psi_p)$  são estimadas por

$$\hat{\mathbf{L}}_{pxm} = \begin{bmatrix} \sqrt{\hat{\lambda}_1} \hat{\mathbf{e}}_1 & \sqrt{\hat{\lambda}_2} \hat{\mathbf{e}}_2 & \dots & \sqrt{\hat{\lambda}_m} \hat{\mathbf{e}}_m \end{bmatrix}$$

e

$$\hat{\mathbf{\Psi}}_{pxp} = \text{diag}(\mathbf{R}_{pxp} - \hat{\mathbf{L}}_{pxm} \hat{\mathbf{L}}_{m xp}),$$

onde  $\hat{\mathbf{e}}_i$  representa o autovetor normalizado correspondente ao  $i$ -ésimo autovalor ( $\hat{\lambda}_i$ ) da matriz de correlação amostral  $\mathbf{R}_{pxp}$  e  $\text{diag}()$  denota uma matriz diagonal. Os autovetores  $\hat{\mathbf{e}}_i, i = 1, \dots, p$  são tais que:  $\hat{\mathbf{e}}_i' \hat{\mathbf{e}}_i = 1, \forall i$ , e  $\hat{\mathbf{e}}_i' \hat{\mathbf{e}}_j = 0, \forall i \neq j$ .

#### 4.2.2. Avaliação da qualidade de ajuste do modelo

A construção do modelo fatorial ortogonal induz uma nova parametrização da matriz teórica de correlações  $\mathbf{P}_{pxp}$  associada às variáveis originais  $\mathbf{X} = [X_1, \dots, X_p]'$  de forma que  $\mathbf{P}_{pxp} = \mathbf{L}\mathbf{L}' + \mathbf{\Psi}$ . O objetivo da análise fatorial é, portanto, encontrar as matrizes  $\mathbf{L}_{pxm}$  e  $\mathbf{\Psi}_{pxp}$  que possam representar a matriz de correlação teórica  $\mathbf{P}_{pxp}$  para um número  $\mathbf{m}$  de fator inferior à quantidade de variáveis originais  $p$ . Como resultado, a variância de cada variável  $\mathbf{Z}_i, i=1, \dots, p$ , é decomposta em duas partes, denominadas *comunalidade* e *variância específica*.

A *comunalidade* é dada por  $h_i^2 = l_{i1}^2 + \dots + l_{im}^2$  e representa a parte da variabilidade de  $Z_i$  que é explicada pelos  $m$  fatores incluídos no modelo de análise fatorial. Em contrapartida, a parte da variabilidade que concerne ao erro aleatório corresponde à *variância específica*, representada por  $\psi_i$ . Como a variável  $Z_i$  tem variância igual a 1, já que representa a padronização da variável original  $X_i$ , temos  $h_i^2 + \psi_i = 1$ .

Precisamente, a comunalidade representa a parte da variância da variável  $Z_i$  que é explicada pelo modelo de análise fatorial ortogonal com  $m$  fatores. Neste sentido, quanto maiores os valores de comunalidade, maior a capacidade de explicação do modelo. Portanto, trata-se de uma quantidade útil para avaliar a qualidade do modelo ajustado.

#### 4.2.3. Interpretação dos fatores

Na Análise Fatorial almeja-se encontrar um número relativamente pequeno de fatores que possuam um alto grau de explicação da variabilidade original dos dados e que, além disso, tornem o conjunto de variáveis observadas mais facilmente interpretáveis na prática. Os valores e o padrão das cargas fatoriais são utilizados para determinar quais variáveis originais tem a maior influência em cada fator. Quanto maior a grandeza numérica da carga fatorial, maior a influência da variável.

Em cada problema prático, portanto, uma vez que as cargas fatoriais associadas a cada variável estejam determinadas, os fatores latentes podem ser utilizados para descrever os dados em um número menor de conceitos do que aqueles representados pelas próprias variáveis originais individualmente. A interpretação de um determinado fator vai depender daquelas variáveis que apresentam cargas fatoriais numericamente altas associadas a ele.

#### 4.2.4. Rotação ortogonal dos fatores

A solução encontrada para o modelo de análise fatorial nem sempre é facilmente interpretável, ou seja, a magnitude numérica das cargas fatoriais das variáveis ao longo dos fatores podem ter valores que não permitem a determinação do conceito prático

associado a cada um deles. Este cenário é caracterizado pela fragilidade na partição das variáveis originais em determinado número de grupos ou frente à possível incoerência na alocação destas, onde os coeficientes da matriz  $\mathbf{L}_{p \times m}$  apresentam valores próximos em termos de grandeza numérica.

Para ajudar na interpretação, torna-se necessário o emprego da transformação conhecida por rotação ortogonal dos fatores originais a fim de encontrar uma solução cujos fatores sejam mais facilmente interpretáveis, no sentido de serem altamente correlacionados com algumas variáveis e baixa correlação com as demais.

Interessante destacar que “a rotação ortogonal para um  $\mathbf{m}$  fixo preserva o valor da variância total explicada pelos  $\mathbf{m}$  fatores obtidos a partir do modelo de análise fatorial sem rotação” (MINGOTI, 2005). Igualmente, este procedimento não altera a aproximação da matriz de correlação amostral  $\mathbf{R}_{p \times p}$ , o que implica dizer as comunalidades e variâncias específicas são as mesmas do modelo de análise fatorial ortogonal não rotacionado.

Isso revela a possibilidade de outras soluções para o modelo fatorial ortogonal no que se refere aos valores das cargas fatoriais – preservando a modelagem construída sobre o relacionamento entre as variáveis originais no que se refere a qualidade de ajuste do modelo.

Dentre os diversos métodos de rotação ortogonal discutidos na literatura, nesta monografia é feito uso do método VARIMAX proposto por KAISER (1958 apud MINGOTI, 2005). Carece destacar que este método é amplamente aplicado e aceito em estudos de diversas áreas, os quais se beneficiaram do método de Análise Fatorial. Neste sentido, seu emprego neste estudo se baseou no fato que este método seleciona os coeficientes que maximizam a variação dos quadrados das cargas fatoriais originais das colunas da matriz  $\hat{\mathbf{L}}_{p \times m}$ . Importante ressaltar que a os métodos de rotação ortogonal tendem a uma concordância na solução. Isso porque foram construídos a partir da estrutura de variabilidade dos dados e, nos casos onde a estrutura se mostra presente na correlação destes, a partição final tende a convergir para uma solução semelhante – diferenciando apenas as cargas fatoriais.

#### 4.2.5. Cálculo dos escores dos fatores

Após a construção, validação e interpretação dos fatores faz-se necessário calcular os escores de cada elemento amostral, de modo a utilizá-los em análises posteriores de interesse. O escore da  $k$ -ésima unidade amostral no fator  $F_j$  é dada por

$$\hat{F}_{jk} = w_{j1} Z_{1k} + w_{j2} Z_{2k} + \dots + w_{jp} Z_{pk}$$

onde  $\mathbf{Z}_k = [Z_{1k}, \dots, Z_{pk}]$  é vetor de observações da  $k$ -ésima unidade amostral e os coeficientes  $w_{ji}, j = 1, \dots, m, i = 1, 2, \dots, p$ , são os pesos de ponderação de cada variável  $Z_i$  no fator  $F_j$ . Os escores  $\hat{F}_{jk}$  podem servir para uma diversidade de estudos como, por exemplo, a construção de gráficos ou modelos de regressão. Neste trabalho, os escores dos fatores serão aplicados em métodos de análise de agrupamentos (Cluster) com o intuito de caracterizar tipologias espaciais dos municípios do estado de Minas Gerais. Existem diferentes técnicas para determinação da matriz de ponderação  $\mathbf{W}_{m \times p}$ . Neste trabalho foi utilizado o método de regressão linear múltipla, no qual as cargas fatoriais fazem o papel das variáveis explicativas, as variáveis  $Z_i$  fazem o papel das variáveis dependentes (resposta) e os fatores fazem o papel dos parâmetros do modelo de regressão ajustado sem intercepto. A estimação dos parâmetros do modelo é realizada pelo método dos mínimos quadrados ponderado (JOHNSON e WICHERN, 2002).

O software estatístico Minitab for Windows (2011, versão 16), utilizado nessa monografia para desenvolvimento dos métodos estatísticos supracitados, calcula automaticamente os escores dos fatores determinados pelo usuário. Nesse cálculo, todas as variáveis padronizadas são incluídas, tanto aquelas mais correlacionadas com o respectivo fator quanto as menos correlacionadas. Entretanto, os valores dos coeficientes de ponderação para as variáveis de menor correlação são numericamente pequenos e, conseqüentemente, o efeito no valor final do escore do fator é de pouca importância. É possível manter na equação apenas as variáveis mais correlacionadas com o fator e, nesse caso, o usuário precisa implementar a equação correspondente.

### 4.3 Técnicas em Estatística Multivariada: Análise de Agrupamentos

Nesta monografia, os escores da análise dos fatores encontrados na análise fatorial assumiram papel fundamental para a proposta de classificação a partir da técnica de Estatística Multivariada de análise de agrupamento (*cluster*). Quando o modelo de repartição assume a técnica hierárquica aglomerativa, partiu-se da menor unidade amostral (cada município no estudo de interesse) até a construção de um único grupo geral (o estado de Minas Gerais no estudo de interesse). Neste modelo a variabilidade é menor no início do processo de agrupamento, visto a presença de poucos elementos em sua composição. Logicamente, a maior variabilidade é encontrada na construção de um único grupo, onde todos os elementos são incluídos. O método de agrupamento utilizado foi o da Média das Distâncias (*Average Linkage*), no qual a distância entre dois *Cluster* (conglomerados) é definida como a média das distâncias entre todos os pares de elementos que podem ser formados com elementos entre os dois grupos, neste caso, municípios (JOHNSON e WICHERN, 2002). Assim, a distância será definida por

$$d(C_1, C_2) = \sum_{l \in C_1} \sum_{k \in C_2} \left( \frac{1}{n_1 n_2} \right) d(X_l, X_k),$$

onde  $C_1$  e  $C_2$  são os conglomerados com  $n_1$  e  $n_2$  elementos, respectivamente, candidatos à união em determinado estágio do processo de agrupamento e  $d(X_l, X_k)$  é a distância entre os elementos amostrais  $l$  e  $k$ , calculada com base nos valores das variáveis aleatórias observadas e utilizadas para realizar o estudo. Em cada estágio do procedimento, a distância entre todos os pares de conglomerados passíveis de união é calculada, sendo unidos os dois grupos que resultam na menor valor. É permitida apenas a união de dois conglomerados e, portanto, a formação de um novo *cluster* em cada estágio. Nessa monografia a distância entre dois elementos amostrais utilizada foi a Euclidiana e as variáveis usadas para obter o agrupamento dos municípios foram os escores dos fatores provenientes da análise fatorial, pois esta distância permite avaliar a dissimilaridade entre os dados de modo geométrico, além de ser amplamente difundida pela literatura. A distância entre conglomerados pode ser transformada em uma medida de similaridade (DAVIS e BOULDIN, 1979). No *software* estatístico Minitab for Windows (2011, versão 16) está disponível a medida de similaridade definida como

$$s(C_1, C_2) = \left( 1 - \frac{d(C_1, C_2)}{M} \right) \cdot 100$$

sendo  $M = \max\{d_{jk}, j, k = 1, 2, \dots, n\}$ , ou seja M é o valor máximo da matriz de distâncias entre os elementos amostrais utilizada no primeiro passo do agrupamento. Eventualmente, essa medida definida no Minitab pode gerar um valor negativo nos últimos passos do agrupamento o que indica que as partições produzidas nesses passos são completamente inadequadas.

Para a definição do número de grupos foram praticados alguns critérios disponíveis na literatura. O primeiro deles refere-se à adoção da distância entre os conglomerados unidos em cada passo (também chamado de nível de fusão) como medida para delinear a faixa inicial de valores plausíveis para o número de grupos.

Enquanto o nível de similaridade decresce à medida que avançamos nas etapas do processo, o nível de fusão aumenta. Deste modo, buscou-se identificar pontos de ruptura, quebras bruscas entre os estágios do processo de agrupamento no que se refere ao nível de fusão (ou seja, pontos de aumentos expressivos do nível de fusão) ou equivalentemente pontos de decrescimento expressivos na medida de similaridades. Esses pontos são indicativos iniciais de interrupção do algoritmo de agrupamento.

O terceiro critério baseia-se no comportamento da variabilidade total da partição, formada pela soma das variabilidades entre grupos e dentro dos grupos visando buscar a partição com maior variabilidade entre os conglomerados e menor variabilidade dentro de cada grupo. Seu cálculo é baseado na Soma de Quadrados Total (SSTc), a qual consiste na soma das variâncias amostrais de todas as variáveis, utilizadas para realizar o agrupamento, multiplicadas por  $(n-1)$ , onde n é o número de observações da amostra – neste estudo, 853 municípios. A soma de quadrados dentro dos grupos pode ser calculada a partir das variâncias amostrais de cada variável por grupo. Tais variâncias devem ser somadas e multiplicadas por  $(n_i-1)$ , onde  $n_i$  corresponde ao número de observações alocadas no grupo i. Feito isto, basta somar todos os valores encontrados para se obter a Soma de Quadrados Total Dentro dos Grupos formados na partição

(SSW). De posse destes dados, a Soma de Quadrados Total Entre Grupos (SSB) é dada por:

$$SSB = SSTc - SSW$$

A partir das medidas de variabilidade apresentadas, pode-se encontrar a proporção da variação total dos dados que é explicada pela divisão desses em determinado número de grupos (*clusters*). Essa medida chamada de coeficiente de correlação intragrupos é definida por:

$$R^2 = \frac{SSB}{SSTc}$$

Quanto maior o valor de  $R^2$ , maior será a soma de quadrados entre grupos (SSB) e menor será a soma de quadrados residual (SSW). Contudo, este critério deve ser empregado após a aplicação dos demais supracitados, visto definir a “região crítica” – região de prováveis valores para o número de grupos. Tais valores podem ser utilizados na construção de gráficos para, assim como nos demais critérios, identificar pontos de forte ruptura, sugerindo o momento ideal de interromper o processo de agrupamento.

O quarto critério utilizado nessa monografia para contribuir na definição da partição ideal foi sugerido por CALINSKI e HARABASZ (1974 apud MINGOTI, 2005), denominado Estatística Pseudo F. O fundamento dessa estatística está baseado na análise de variância para a comparação dos vetores de médias dos grupos formados a cada passo. Busca-se o maior valor desta estatística, o qual representaria a menor probabilidade de significância e, assim, sugere a rejeição da hipótese de igualdade entre os vetores de médias populacionais, caso o teste estatístico fosse implementado, corroborando para a identificação da partição com maior heterogeneidade entre grupos. Essa estatística é definida por:

$$F = \frac{SSB/(g^* - 1)}{SSW/(n - g^*)} = \left( \frac{n - g^*}{g^* - 1} \right) \left( \frac{R^2}{1 - R^2} \right)$$

onde  $g^*$  é o número de grupos relacionados com a partição do respectivo estágio de agrupamento (MINGOTI, 2005).



Conforme CALINSKI e HARABASZ, se  $F$  é invariavelmente crescente com a quantidade de grupos, há indícios de inexistência de qualquer estrutura “natural” de partição dos dados. Em contrapartida, se  $F$  apresentar um valor máximo, há indícios de existência de uma “partição ideal” nos dados.

Existem vários outros métodos para estimação do número de grupos ou avaliação da qualidade da partição. No entanto, estudos mostram que os quatro critérios mencionados nessa monografia são bons critérios em termos práticos e de fato muito utilizados (MILLIGAN e COOPER, 1985).

## **5 DESCRIÇÃO E ANÁLISE DOS RESULTADOS**

Nesta seção são apresentados os resultados do estudo. A seção 5.1 aborda a construção do banco de dados georreferenciados. A seção 5.2 apresenta a análise das variáveis por meio das técnicas de estatística descritiva. A seção 5.3 evidencia os resultados obtidos através do modelo de Análise Fatorial. Por fim, na seção 5.4 são expostos os resultados do uso da técnica de Análise de Agrupamentos.

### **5.1 Construção do Banco de Dados e Seleção das Variáveis**

O Atlas de Desenvolvimento Humano no Brasil dispõe de 232 variáveis socioeconômicas segmentadas em diversos temas como Educação, Renda, Demografia e Saúde, todas referentes ao ano de 2010. Para alcançarmos o objetivo proposto neste estudo, foram selecionadas 10 variáveis de diferentes temas, apresentadas no Quadro 2.

QUADRO 2: Variáveis selecionadas para estudo

| Variável   | Descrição  | Sigla no BD       | Unidade de Medida        |
|--|--|-------------------|--------------------------|
| Esperança de Vida ao Nascer                          | Corresponde a média de anos que as pessoas deverão viver se mantidas as condições diagnosticadas no período da pesquisa.   | ESPVIDA           | Número médio de anos     |
| Mortalidade infantil                                 | Representa o número de crianças que não deverão sobreviver ao primeiro ano de vida em cada 1.000 crianças.   | MORT1             | Quantidade de indivíduos |
| Renda <i>per capita</i>                              | Caracteriza o somatório da renda total dos indivíduos residentes no município dividido pela quantidade de pessoas deste.   | RDPC              | Valores em Reais (R\$)   |
| Renda <i>per capita</i> do 1º quinto                 | Equivale ao valor máximo da renda domiciliar <i>per capita</i> dos 20% mais pobres do município.   | CORTE1            | Valores em Reais (R\$)   |
| Rendimento médio dos adultos ocupados                | Apresenta a renda média das pessoas empregadas acima de 18 anos de idade residentes no município amostrado.  | RENOCUP           | Valores em Reais (R\$)   |
| Expectativa de anos de estudo                        | Denota a média de anos de estudo que uma pessoa deverá completar ao atingir 18 anos de idade no município amostrado.   | E_ANOESTUDO       | Número médio de anos     |
| Percentual de adultos com ensino superior            | Apresenta a proporção entre as pessoas acima de 25 anos com ensino superior concluído e o total de pessoas nessa faixa etária do município.                              | T_SUPER25M        | Proporção de indivíduos  |
| Percentual infantil no ensino básico sem atraso      | Constitui a proporção de estudantes entre 6 e 17 anos matriculados regularmente em suas respectivas séries/ciclos e o total de pessoas nessa faixa etária por município. | T_ATRASO_0_BASICO | Proporção de indivíduos  |
| Percentual de adultos ocupados no setor agropecuário | Reflete a proporção de adultos empregados no setor agropecuário e o total desta faixa etária por município.  | P_AGRO            | Proporção de indivíduos  |
| Taxa de adultos desempregados                        | Percentual da população economicamente ativa (PEA) nessa faixa etária que estava desocupada por município na época da pesquisa.  | T_DES18M          | Proporção de indivíduos  |

Observamos que existe diferença entre a escala de mensuração das variáveis, dada à natureza e o método de apresentação disponibilizado pelos autores do dado. Para desenvolvimento do estudo foi necessário construir um banco de dados geográficos (BDGEO). Neste sentido, foi utilizado o Sistema de Informação Geográfica (SIG) ArcGIS 10.2, o qual permitiu associar os resultados das variáveis aleatórias aos municípios – colunas e linhas, respectivamente – garantindo a representação

alfanumérica na forma de tabela de atributos e espacial, reproduzida nas geometrias (polígonos) com coordenadas geográficas indexadas.

## 5.2 Análise Descritiva dos Dados

Numa abordagem exploratória, as variáveis foram analisadas por meio das técnicas de Estatística Descritiva cujos resultados estão apresentados na TABELA 1.

TABELA 1: Estatísticas descritivas das variáveis originais

|  | Média  | Mediana | Desvio padrão | Coef. de Variação | Mínimo | Máximo   |
|--|--------|---------|---------------|-------------------|--------|----------|
| Esperança de Vida ao Nascer                          | 74,42  | 74,41   | 1,79          | 2,41              | 68,37  | 78,15    |
| Mortalidade infantil                                 | 16,38  | 16,20   | 2,92          | 17,83             | 10,35  | 27,80    |
| Renda per capita                                     | 490,60 | 475,23  | 173,08        | 35,28             | 181,77 | 1.731,84 |
| Renda per capita do 1º quinto                        | 170,24 | 169,17  | 65,69         | 38,59             | 22,40  | 342,50   |
| Rendimento médio dos adultos ocupados                | 766,39 | 736,57  | 262,60        | 34,26             | 242,47 | 2.437,34 |
| Expectativa de anos de estudo                        | 9,09   | 9,15    | 0,76          | 8,36              | 6,03   | 11,04    |
| Percentual de adultos com ensino superior            | 5,78   | 5,28    | 2,59          | 44,87             | 1,31   | 22,93    |
| Percentual infantil no ensino básico sem atraso      | 60,17  | 60,53   | 7,27          | 12,08             | 34,37  | 80,71    |
| Percentual de adultos ocupados no setor agropecuário | 34,81  | 36,32   | 16,82         | 48,31             | 0,49   | 79,35    |
| Taxa de adultos desempregados                        | 5,81   | 5,34    | 2,93          | 50,45             | 0,59   | 20,77    |

Observa-se que as variáveis associadas à renda, educação e trabalho apresentam maior variabilidade, o que reflete a desigualdade destes temas. Em contrapartida, as demais variáveis não apresentam grandes diferenças entre as suas medidas de média, mediana e moda. Para discutir estas colocações, deve-se atentar para a unidade de medida e escala das variáveis utilizadas, destacadas anteriormente no QUADRO 2.

Para complementar a análise descritiva, elaboramos os gráficos histograma e box-plot das variáveis empregadas no estudo. Estes podem ser observados no APÊNDICE A. Tais gráficos sugerem que as variáveis que compõem o estudo desta monografia apresentam comportamento assimétrico, com valores predominantemente abaixo ou acima da média. Além disso, os gráficos box-plot sugerem que as medidas de tendência central apresentam diferenças em sua distribuição, em função da presença de valores discrepantes.

Ainda em caráter exploratório, buscou-se identificar a possibilidade de associação entre as variáveis. Para isso, utilizou-se o método de correlação de Pearson .

Como resultado, a variável Esperança de Vida ao Nascer apresentou correlações moderadas e fortes com as variáveis Mortalidade Infantil (-0,996), Renda *per capita* (0,745), Renda *per capita* do 1ºquinto (0,739), Rendimento médio dos adultos ocupados (0,713), e Percentual de adultos com ensino superior (0,553). Isso sugere que municípios com maior renda possuem maior desenvolvimento e, com isso, menores valores de mortalidade infantil e maior expectativa de vida.

A variável Percentual de adultos com ensino superior apresentou correlações moderadas e fortes com as variáveis Mortalidade infantil (-0,535), Renda *per capita* (0,757), Renda *per capita* do 1º quinto (0,606) e Rendimento médio dos adultos ocupados (0,724). Este cenário sugere a influência direta da educação superior na renda e redução da mortalidade infantil, representando também o grau de desenvolvimento municipal.

A variável Renda *per capita* apresentou fortes correlações com as variáveis Rendimento médio dos adultos ocupados (0,943), Renda *per capita* do 1º quinto (0,877) e moderada com a variável Percentual de adultos ocupados no setor agropecuário (-0,580). Assim, as variáveis renda apresentaram padrão inverso em relação as atividades formais desenvolvidas no setor agropecuário. Isso pode inferir nas relações de trabalho informais presentes no setor agropecuário, nas quais os benefícios e relações entre contratado e contratante não são claramente evidenciadas por meio de instrumentos formalizados, sejam contratos ou vínculos empregatícios.

A variável Mortalidade infantil apresentou correlações semelhantes com as variáveis Renda *per capita* (-0,730), Rendimento médio dos adultos ocupados (-0,700) e Renda *per capita* do 1º quinto (-0,730). Tal cenário é refletido nas análises anteriores.

A variável Renda *per capita* do 1º quinto apresentou correlações fortes e moderada com as variáveis Rendimento médio dos adultos ocupados (0,834) e Percentual de adultos ocupados no setor agropecuário (-0,574), caso semelhante à comparação da renda com os ocupados neste setor, discutido anteriormente.

Assim como a variável Percentual de adultos ocupados no setor agropecuário apresentou correlações moderadas com as variáveis Rendimento médio dos adultos ocupados (-0,636) e Percentual de adultos com ensino superior (-0,583).

Por fim, a variável Expectativa de anos de estudo apresentou correlação forte apenas com a variável Percentual infantil no ensino básico sem atraso (0,735), o que sugere forte associação entre as variáveis referentes a educação básica municipal.

As demais correlações apresentaram valores inferiores a 0,5 e, em função disso, foram indicadas apenas no APÊNDICE B, onde apresentamos o quadro completo de correlações.

### 5.3 Análise Fatorial

O modelo de análise fatorial foi ajustado para as variáveis padronizadas e os fatores foram extraídos a partir da matriz de correlação das variáveis, sendo utilizada a técnica de análise de componentes principais para estimação do número de fatores e cargas fatoriais (MINGOTI, 2005).

#### 5.3.1 Estimação do número de fatores e cargas fatoriais

Considerando os critérios para determinação do número de fatores descritos na seção 4.2, podemos constatar no GRÁFICO 1 dois momentos de forte ruptura entre os autovalores de ordem 1 e 3. Diante disso e em concordância com a regra de Kaiser (1958), assumimos o uso de três fatores no modelo, pois os outros autovalores

(*eigenvalues*) encontrados são inferiores a 1. No QUADRO 3 são apresentadas as cargas fatoriais estimadas de cada fator, respectivas comunalidades e percentual de variância total explicada, das soluções com  $m=3$  fatores com e sem rotação ortogonal VARIMAX (KAISER, 1958).

Optamos por trabalhar com a solução pelo método de rotação VARIMAX, pois reforçou a disparidade das cargas fatoriais das variáveis entre os fatores, contribuindo para sua interpretação. Ainda, o modelo encontrado abarca 82,7% da variabilidade total da estrutura de dados empregada para sua construção e, observadas as comunalidades encontradas, pode ser considerado válido e representativo para a proposta do estudo em questão. Adicionalmente, todas as comunalidades estão acima de 67,5%, sendo que a maioria está acima de 80%. As cargas fatoriais mais representativas em cada fator estão assinaladas em negrito.

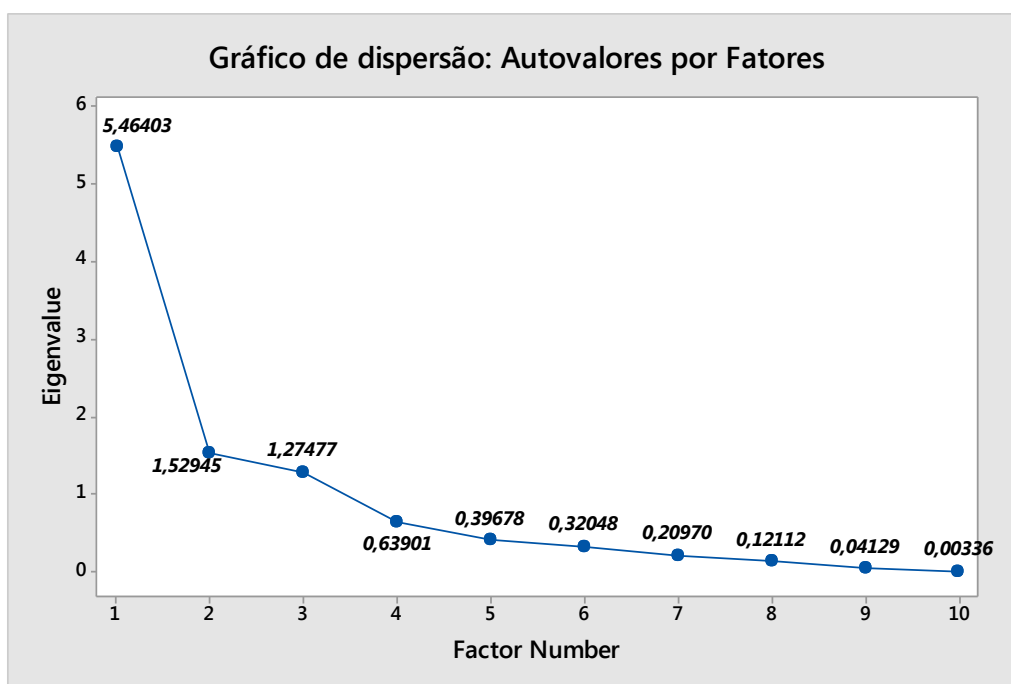


GRÁFICO 1: Autovalores da matriz de correlação por número de fatores. *Factor number* representa o número de fatores e *eigenvalue* representa os autovalores.

QUADRO 3: Modelo fatorial para três fatores – Método de Componentes Principais

| <b>Rotação</b>        | <b>Variáveis</b>      | <b>Fator 1</b> | <b>Fator 2</b> | <b>Fator 3</b> | <b>Comunalidades</b> |
|-----------------------|-----------------------|----------------|----------------|----------------|----------------------|
| Não aplicada          | ESPVIDA               | <b>0,871</b>   | -0,138         | 0,152          | 0,801                |
|                       | MORT1                 | <b>-0,861</b>  | 0,136          | -0,157         | 0,785                |
|                       | RDPC                  | <b>0,940</b>   | -0,105         | 0,019          | 0,894                |
|                       | CORTE1                | <b>0,908</b>   | -0,050         | 0,148          | 0,848                |
|                       | RENOCUP               | <b>0,913</b>   | -0,147         | -0,095         | 0,865                |
|                       | E_ANOESTUDO           | 0,341          | <b>0,872</b>   | -0,100         | 0,886                |
|                       | T_SUPER25M            | <b>0,774</b>   | -0,056         | -0,271         | 0,675                |
|                       | T_ATRASO_0_BASICO     | 0,431          | <b>0,823</b>   | 0,128          | 0,879                |
|                       | P_AGRO                | <b>-0,676</b>  | 0,113          | 0,550          | 0,772                |
|                       | T_DES18M              | -0,255         | 0,064          | <b>-0,890</b>  | 0,862                |
|                       | Autovalores           | 5,464          | 1,530          | 1,275          | -                    |
| Percentual explicação | 54,6                  | 15,3           | 12,7           | ,82,7          |                      |
| <b>Rotação</b>        | <b>Variáveis</b>      | <b>Fator 1</b> | <b>Fator 2</b> | <b>Fator 3</b> | <b>Comunalidades</b> |
| Método Varimax        | ESPVIDA               | <b>0,858</b>   | 0,101          | 0,233          | 0,801                |
|                       | MORT1                 | <b>-0,848</b>  | -0,101         | -0,236         | 0,785                |
|                       | RDPC                  | <b>0,928</b>   | 0,149          | 0,105          | 0,894                |
|                       | CORTE1                | <b>0,871</b>   | 0,196          | 0,228          | 0,848                |
|                       | RENOCUP               | <b>0,925</b>   | 0,100          | -0,010         | 0,865                |
|                       | E_ANOESTUDO           | 0,107          | <b>0,929</b>   | -0,106         | 0,886                |
|                       | T_SUPER25M            | <b>0,783</b>   | 0,147          | -0,201         | 0,675                |
|                       | T_ATRASO_0_BASICO     | 0,185          | <b>0,910</b>   | 0,130          | 0,879                |
|                       | P_AGRO                | <b>-0,731</b>  | -0,061         | 0,484          | 0,772                |
|                       | T_DES18M              | -0,178         | -0,021         | <b>-0,911</b>  | 0,862                |
|                       | Percentual explicação | 51,5           | 18,1           | 13,1           | 82,7                 |

### 5.3.2 Interpretação dos fatores

De acordo com a solução do QUADRO 3 e ressaltando que as cargas fatoriais estimadas representam as correlações dos fatores com cada uma das variáveis utilizadas na análise, o Fator 1 representa o índice de longevidade (ESPVIDA e MORT1) agregado com renda (RDPC, RENOCUP e CORTE1), população com ensino superior (T\_SUPER25M) e profissionais empregados no setor agropecuário (P\_AGRO). Pela estrutura apresentada este fator retém 51,5% da variabilidade dos dados, o qual pode ser interpretado como destaque nos municípios com melhores estruturas, condições e qualidade de vida. Além disso, quanto maiores os valores dos escores neste fator, melhor a posição do município nos quesitos elencados. As variáveis Percentual de adultos ocupados no setor agropecuário (P\_AGRO) e Mortalidade infantil (MORT1) assumiram sinal negativo neste fator. Isso confirma sua interpretação, pois os salários e

níveis de escolaridade exigidos para indivíduos empregados nos outros setores são, em geral, superiores ao do setor agropecuário. Deste modo, quanto menor o percentual de profissionais ocupados no setor agropecuário do município, maior a renda média formal dos indivíduos. De modo semelhante, quanto menor o número de crianças que não deverão sobreviver ao primeiro ano de vida em cada 1.000 crianças, maior o escore deste município. Em ambos os casos, tais variáveis refletem indiretamente as condições de estrutura municipal, renda e saúde no município. Para ilustrar este potencial, podemos observar a FIGURA 2 que apresenta a distribuição espacial dos escores do Fator 1 dos municípios de Gerais.

Como foi mencionado na seção 4.2.5 (p.29), os escores utilizados na análise estatística foram obtidos através do software Minitab for Windows, versão 16, e são compostos por todas as variáveis padronizadas, tanto as mais correlacionadas com o fator quanto as de menor correlação, com a ressalva que os coeficientes de ponderação dessas últimas são de menor importância numérica e, como consequência, não impactam significativamente no valor final do escore do Fator.

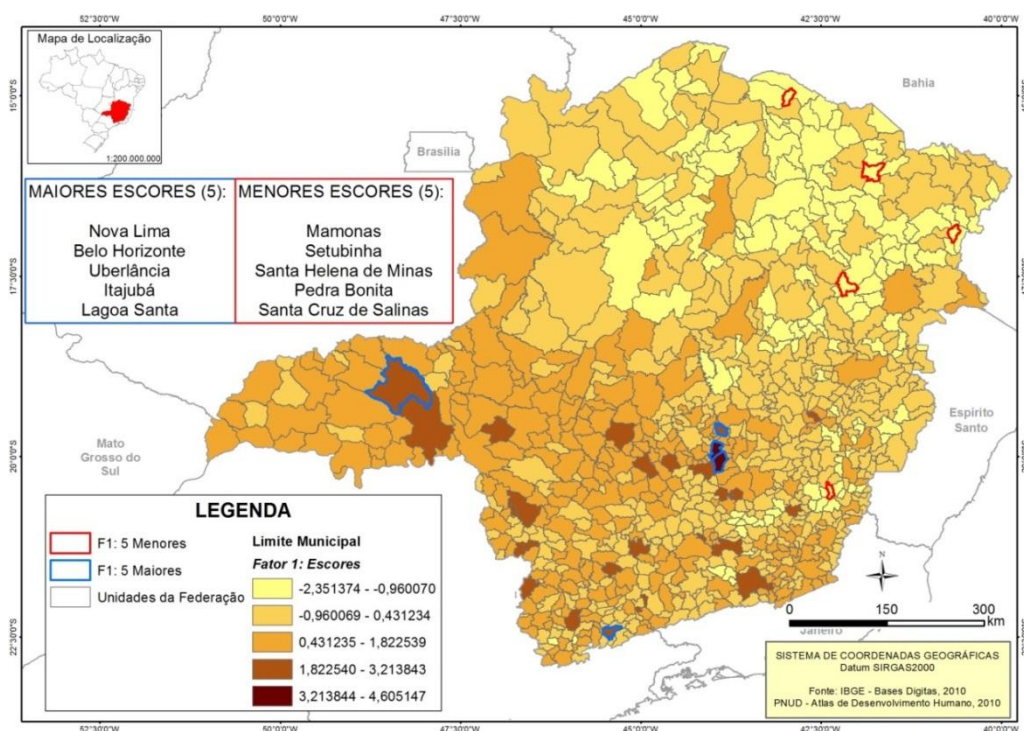


FIGURA 2: Mapa dos escores para Fator 1

Dos cinco maiores municípios de acordo com os escores do Fator 1, três estão inseridos na Região Metropolitana de Belo Horizonte – a própria capital mineira, Nova Lima e



Lagoa Santa. Os outros dois indicam a possível influência regional de São Paulo – Uberlândia e Itajubá. Tal constatação sugere indícios das teorias regionais na configuração espacial polarizada. Os cinco municípios com menores escores neste fator possuem população abaixo de 15.000 habitantes e PIB abaixo de R\$ 20.000 (IBGE, 2010), situados nas porções leste e nordeste do Estado.

O segundo fator representa um índice educacional vinculado à educação básica, composto pela expectativa de anos de estudo que um habitante deverá completar até atingir a fase adulta (E\_ANOSESTUDO) e pelo percentual de jovens regularmente matriculados na educação básica (T\_ATRASO\_0\_BASICO). Semelhante ao primeiro fator, maiores valores refletem melhor situação no quesito educacional. Este corresponde a 18,1% da variabilidade total dos dados e pode ser interpretado como um índice exclusivo para a educação básica municipal, semelhante a um indicador global para o desempenho dos estudantes inscritos até o ensino médio.

A FIGURA 3 ilustra a distribuição espacial dos escores no Fator 2 em Minas Gerais. Aparentemente, não é possível assegurar a existência de um padrão espacial, visto a dispersão dos escores ao longo do território.

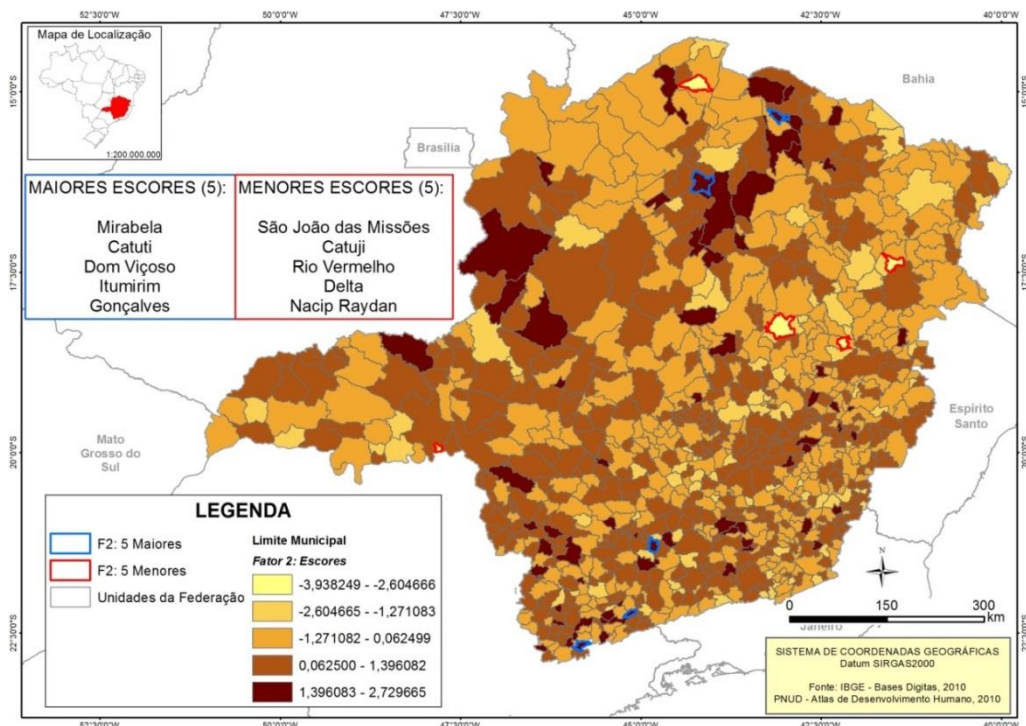


FIGURA 3: Mapa dos escores para Fator 2

Os cinco primeiros colocados no Fator 2 (Mirabela, Catuti, Dom Viçoso, Itumirim, Gonçalves) apresentam taxa de escolarização de 6 a 14 anos de idade acima de 98%, contra 97,6% em Belo Horizonte (IBGE, 2010). Os cinco últimos colocados neste fator possuem percentuais abaixo de 95% na mesma taxa, confirmando a interpretação dada.

O Fator 3 representa 13,1% da variabilidade explicada pelo modelo fatorial e, com maior peso, o percentual de desempregados inseridos na PEA (T\_DES18M) na data de coleta dos dados. Há, no entanto, uma correlação positiva moderada do fator com a variável P\_AGRO. Como o valor encontrado da carga fatorial para essa variável é bem maior no Fator 1, há maior evidência para que seja alocada no Fator 1 ao invés do Fator 3. Assim, para o Fator 3 predomina a interpretação relativa ao percentual de pessoas que fazem parte da População Economicamente Ativa e, entretanto, está desempregada. Deste modo, quanto maior os valores dos escores do Fator 3, menor o desemprego na população municipal. O mesmo pode ser útil para direcionar políticas públicas para incentivo à geração de empregos e estudos de desenvolvimento regional. Optamos por manter este fator visto a contribuição da variável no estudo e manutenção da estrutura encontrada – frente aos demais testes realizados.

A FIGURA 4 mostra a distribuição espacial dos escores no Fator 3 ao longo municípios de Minas Gerais. Neste caso, podem-se notar indícios de um padrão espacial, onde os maiores escores estão situados na porção sudoeste do Estado de Minas Gerais, principalmente na divisa com o Estado de São Paulo. Naturalmente, pode-se sugerir que estes municípios sofrem influência das cidades de nível superior da região, até mesmo da própria capital paulista.

De mesmo modo, a predominância das atividades de setores específicos pode contribuir para a formação de “ilhas” ou identificação de grupos de municípios semelhantes, como retratado na FIGURA 4, especificamente na região leste de Minas Gerais, onde o polo regional conhecido como Vale do Aço é claramente destacado. Observamos que os municípios com valores elevados no Fator 3 representam menor percentual de pessoas ocupadas formalmente. Isso pode ter ocorrido em função da influência da variável P\_AGRO na construção do índice.

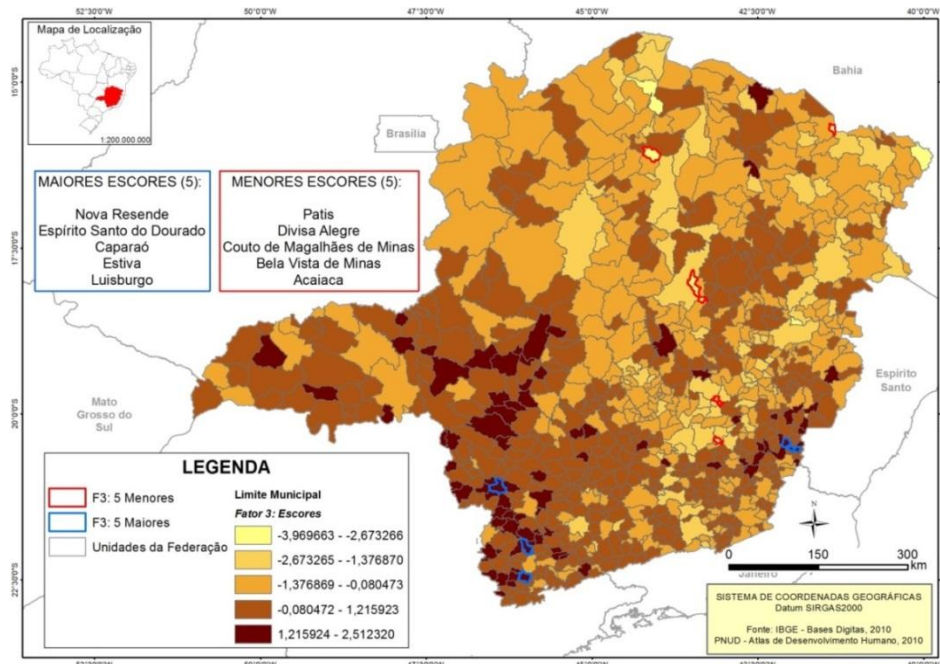


FIGURA 4: Mapa dos escores para Fator 3

No intuito de evitar possíveis confusões de interpretação sobre os valores apresentados para as variáveis padronizadas, deve-se considerar que estes foram obtidos por meio da subtração dos valores originais em relação à média da variável. Deste modo, municípios com valores menores que a média recebem valores negativos para  $Z_i$  e aqueles com valores maiores geram valores positivos. Considerando a predominância da variável T\_DES18M na composição do Fator 3, elaboramos a FIGURA 5 para representar exclusivamente sua distribuição espacial.

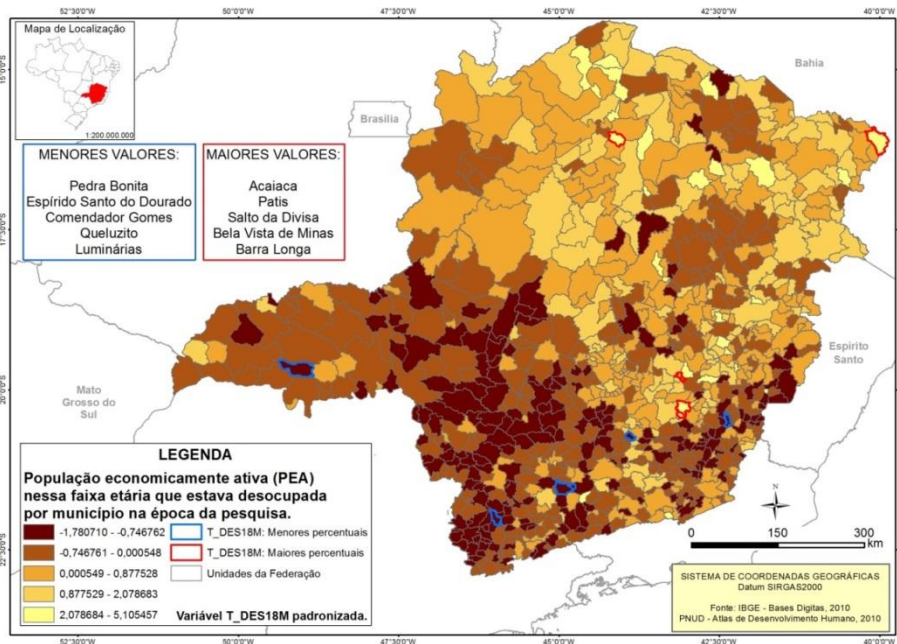


FIGURA 5: Mapa da variável T\_DES18M padronizada

Assim como na elaboração do modelo de análise fatorial, utilizamos a variável T\_DES18M padronizada para elaboração da FIGURA 5. Observamos que, apesar de o padrão de distribuição desta variável ser semelhante ao apresentado na distribuição espacial dos escores relativos ao Fator 3 (FIGURA 4), não são totalmente concordantes.

Alguns dos municípios apontados como maiores e menores escores do Fator 3 foram concordantes com os resultados obtidos pela variável padronizada, tais como Espírito Santo do Dourado, Patis, Bela Vista de Minas e Acaiaca. Porém, os municípios Divisa Alegre, Couto de Magalhães de Minas, Nova Resende, Caparaó, Estiva e Luisburgo não se foram indicados quando analisamos somente a variável T\_DES18M padronizada.

Isso se dá em função do próprio conceito e concepção do modelo de análise fatorial ortogonal aplicado neste estudo, o qual considerou os coeficientes de todas as variáveis no cálculo dos escores. Neste sentido, a predominância da variável T\_DES18M no Fator 3 não inibe a influência das demais que, apesar da baixa grandeza numérica representada pelas cargas fatoriais, contribuíram para os resultados do modelo.

Em outros testes, geramos novos modelos sem esta variável. Contudo, os resultados encontrados não foram superiores ao apresentado, reduzindo consideravelmente o percentual de representação do modelo e aumento da complexidade de análise dos demais fatores. Deste modo, optamos pela manutenção a estrutura apresentada – principalmente por sua representatividade e importância desta variável para caracterização regional.

Os coeficientes utilizados para cálculo dos escores de cada fator são apresentados no QUADRO 4. Tais coeficientes representam o peso relativo de cada variável na combinação linear das variáveis do respectivo fator e foram estimados a partir do método de regressão linear (JOHNSON e WICHERN, 2002).

QUADRO 4: Coeficientes do Modelo fatorial para três fatores

| VARIÁVEIS   | FATOR 1       | FATOR 2      | FATOR 3       |
|---|---------------|--------------|---------------|
| Esperança de Vida ao Nascer (ESPVIDA)                               | <b>0,166</b>  | -0,043       | 0,137         |
| Mortalidade infantil (MORT1)  | <b>-0,163</b> | 0,041        | -0,140        |
| Renda <i>per capita</i> (RDPC)                                      | <b>0,182</b>  | -0,020       | 0,033         |
| Renda <i>per capita</i> do 1º quinto (CORTE1)                       | <b>0,157</b>  | 0,0150       | 0,131         |
| Rendimento médio dos adultos ocupados (RENOCUP)                     | <b>0,193</b>  | -0,049       | -0,056        |
| Expectativa de anos de estudo (E_ANOESTUDO)                         | -0,083        | <b>0,565</b> | -0,096        |
| Percentual de adultos com ensino superior (T_SUPER25M)              | <b>0,166</b>  | -0,001       | -0,198        |
| Percentual infantil no ensino básico sem atraso (T_ATRASO_0_BASIC0) | -0,075        | <b>0,541</b> | 0,084         |
| Percentual de adultos ocupados no setor agropecuário (P_AGRO)       | <b>-0,179</b> | 0,046        | 0,416         |
| Taxa de adultos desempregados (T_DES18M)                            | 0,010         | 0,016        | <b>-0,701</b> |

Como ilustração, apresentamos a seguir o cálculo dos escores dos fatores para o município de Belo Horizonte no Fator 1. O asterisco na equação representa o uso de variáveis padronizadas, ou seja, o valor original da variável subtraído da média amostral e dividido pelo desvio-padrão amostral da variável.

$$\begin{aligned} \text{FATOR 1(escore)} = & (0,166 \cdot \text{ESPVIDA}^*) + (-0,163 \cdot \text{MORT1}^*) + (0,182 \cdot \text{RDPC}^*) + (0,157 \cdot \text{CORTE1}^*) + \\ & (0,193 \cdot \text{RENOCUP}^*) + (-0,083 \cdot \text{E\_ANOESTUDO}^*) + (0,166 \cdot \text{T\_SUPER25M}^*) + \\ & (-0,075 \cdot \text{T\_ATRASO\_0\_BASICO}^*) + (-0,179 \cdot \text{P\_AGRO}^*) + (0,010 \cdot \text{T\_DES18M}^*) \end{aligned}$$

$$\begin{aligned} \text{FATOR 1(escore)} = & (0,166 \cdot 1,087) + (-0,163 \cdot (-1,174)) + (0,182 \cdot 5,816) + \\ & (0,157 \cdot 2,470) + (0,193 \cdot 4,831) + (-0,083 \cdot 1,023) + \\ & (0,166 \cdot 6,619) + (-0,075 \cdot 0,702) + (-0,179 \cdot (-2,041)) + (0,010 \cdot 0,089) \\ & = \mathbf{4,073} \end{aligned}$$

O QUADRO 5 apresenta os dados originais e padronizados das variáveis utilizadas e relativas ao município de Belo Horizonte, utilizado como exemplo no cálculo dos escores para o Fator 1. Deste modo, é possível replicar a metodologia aplicada neste trabalho para outros casos.

QUADRO 5: Valores referentes ao município de Belo Horizonte – originais e padronizados

| Variáveis  | Valores originais | Valores padronizados | Média  | Desvio Padrão |
|--|-------------------|----------------------|--------|---------------|
| Esperança de Vida ao Nascer                          | 76,37             | 1,087                | 74,42  | 1,79          |
| Mortalidade infantil                                 | 12,95%            | -1,174               | 16,38  | 2,92          |
| Renda per capita                                     | R\$ 1.497,29      | 5,816                | 490,60 | 173,08        |
| Renda per capita do 1º quinto                        | R\$ 332,50        | 2,470                | 170,24 | 65,69         |
| Rendimento médio dos adultos ocupados                | R\$ 2.034,94      | 4,831                | 766,39 | 262,60        |
| Expectativa de anos de estudo                        | 9,87              | 1,023                | 9,09   | 0,76          |
| Percentual de adultos com ensino superior            | 22,93%            | 6,619                | 5,78   | 2,59          |
| Percentual infantil no ensino básico sem atraso      | 65,27%            | 0,702                | 60,17  | 7,27          |
| Percentual de adultos ocupados no setor agropecuário | 0,49%             | -2,041               | 34,81  | 16,82         |
| Taxa de adultos desempregados                        | 6,07%             | 0,089                | 5,81   | 2,93          |

Os valores das variáveis padronizadas para o município de Belo Horizonte sugerem desempenho superior em relação aos demais municípios. Com exceção das variáveis associadas ao percentual de empregados no setor agropecuário e mortalidade infantil, Belo Horizonte apresenta valores superiores às médias de todas as demais variáveis. Adicionalmente, o comportamento retratado pelos dados corrobora para sua classificação política, dada como capital estadual.

O QUADRO 6 apresenta as estatísticas descritivas dos escores por fator. As medidas descritivas Média e Desvio-padrão não são apresentados já que, pela metodologia utilizada na construção dos fatores, essas medidas são iguais a 0 e 1, respectivamente.

QUADRO 6: Estatísticas descritivas dos escores em cada fator

| Escores | Mínimo  | Mediana | Máximo | Assimetria | P-valor<br>(Teste de Normalidade) | Amplitude |
|---------|---------|---------|--------|------------|-----------------------------------|-----------|
| Fator 1 | -2,3514 | -0,0641 | 4,6051 | 0,41       | <0,010                            | 6,9565    |
| Fator 2 | -3,9382 | 0,0137  | 2,7297 | -0,18      | 0,072                             | 6,6679    |
| Fator 3 | -3,9697 | 0,0620  | 2,5123 | -0,37      | <0,010                            | 6,4820    |

Ao contrário das medianas dos escores dos Fatores 2 e 3, os escores do Fator 1 apresentaram mediana negativa (-0,0641). Os escores do Fator 1 também apresentam maiores valores e amplitude, seguidos pelos escores do Fator 2 e 3, nesta ordem. Para avaliar se os escores seguem ou não uma distribuição normal, foi realizado o teste de

normalidade de Ryan-Joiner – semelhante ao teste de Shapiro-Wilk (SHAPIRO e WILK, 1965). Neste teste a hipótese nula é a de que os dados são provenientes da distribuição normal enquanto que a hipótese alternativa é a de que eles não são provenientes dessa distribuição.

Considerando o nível de significância de 0,05 – ou seja, assumindo o risco de 5% de afirmar que os dados não seguem a distribuição normal quando eles, de fato, seguem – observamos que o teste de normalidade para a distribuição dos escores dos Fatores 1 e 3 apresentam p-valor inferior a 0,010. Nestes casos, deve-se rejeitar a hipótese nula e assumir que os dados não seguem uma distribuição normal. O Fator 2 apresentou p-valor 0,072, o que nos impede de rejeitar a hipótese nula e concluindo a 5% de que os dados são aderentes a uma distribuição normal.

Os valores de assimetria extremos nos Fatores 1 e 3: 0,41 e -0,37, respectivamente, corroboram para os resultados referentes ao teste de normalidade, representados pelo p-valor inferior a 0,010. Para os escores do Fator 1 espera-se um histograma com cauda maior para o lado direito, retratando a assimetria positiva. O oposto deve-se esperar para o histograma do Fator 3, com cauda maior para o lado esquerdo em função da assimetria negativa. O Fator 2 apresenta leve assimetria negativa – originada possivelmente pela presença de algum dado discrepante. Tais considerações podem ser observadas na FIGURA 5, onde apresentamos os escores dos três fatores por meio de histogramas e box-plots.

Apesar dos resultados do teste de normalidade de Ryan-Joiner (Shapiro-Wilk) sugerirem que os escores dos Fatores 1 e 3 não seguem uma distribuição normal, os histogramas apresentados na FIGURA 6 com a curva de distribuição normal refletem a sensibilidade deste teste em relação a presença de dados discrepantes. Nos três Fatores, os escores se concentram em torno do valor zero e possuem poucos registros distantes da massa de dados, conforme gráficos boxplot.

Neste ponto, tais informações acrescentariam para novos estudos associados à aplicação dos escores, onde as suposições de distribuição normal seriam exigidas – modelos de regressão, por exemplo

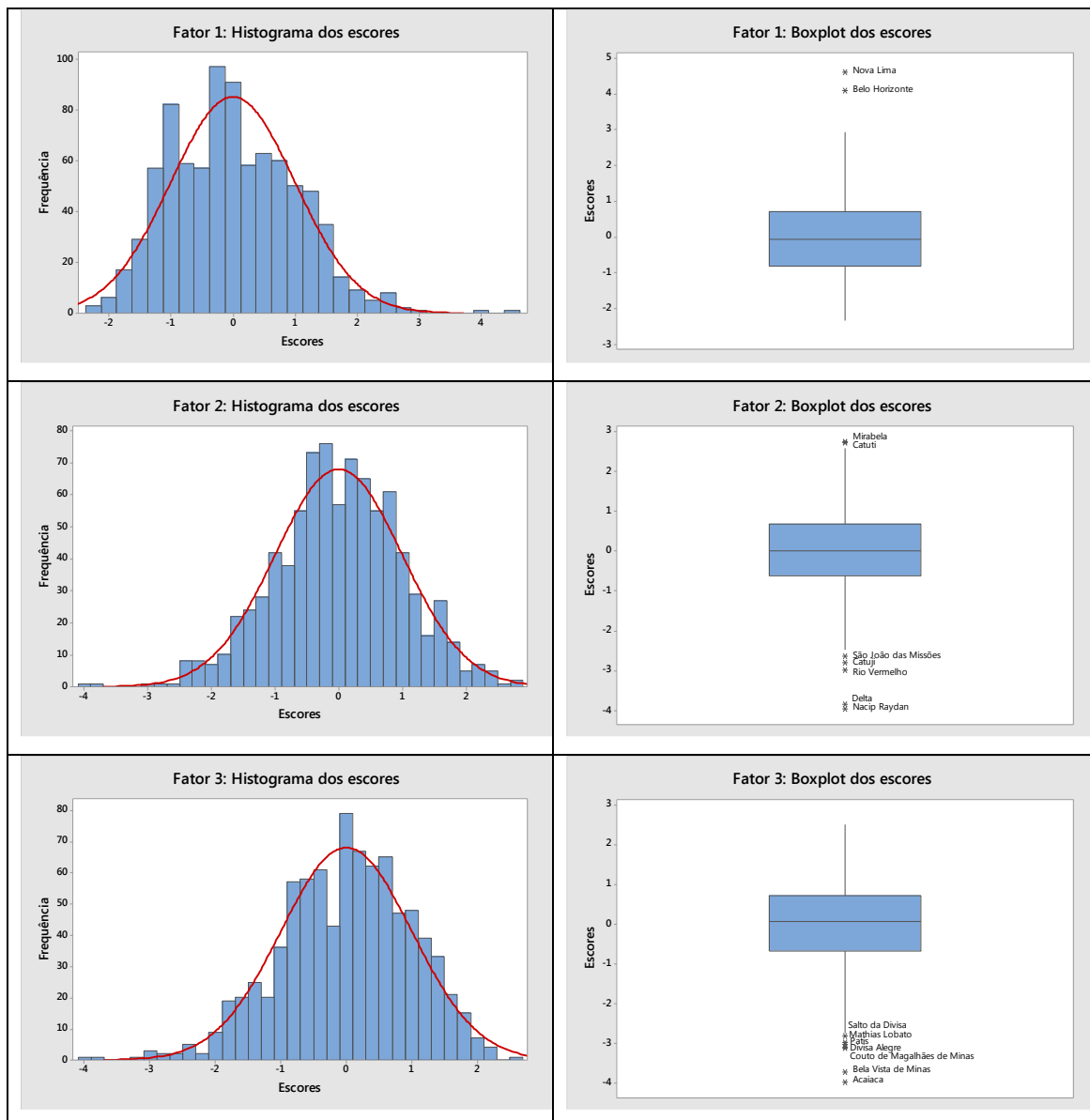


FIGURA 6: Histograma e boxplot dos escores por Fatores

Em razão da predominância da variável T\_DES18M na construção do Fator 3, apresentamos na FIGURA 7 os gráficos histograma e boxplot dos seus dados padronizados, utilizados na construção do modelo de análise fatorial.



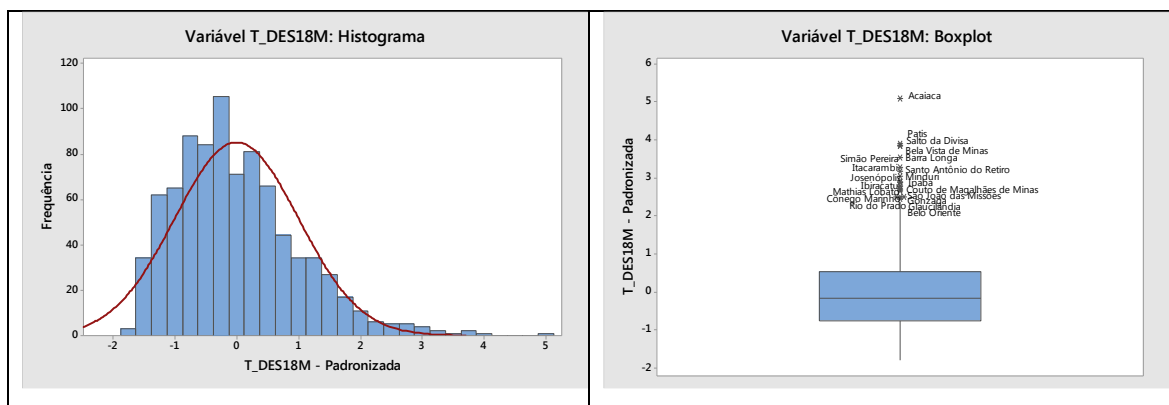


FIGURA 7: Histograma e boxplot dos valores padronizados da variável T\_DES18M

A FIGURA 7 retrata de modo semelhante o cenário do Fator 3, apresentado graficamente na FIGURA 6. Contudo, a alteração de sentido da assimetria refere-se à troca de sinal presente na construção do modelo de análise fatorial. O comportamento em ambos histogramas é praticamente o mesmo, visto a predominância desta variável. O mesmo ocorre para os gráficos boxplot.

O GRÁFICO 2 retrata as cargas fatoriais das variáveis no primeiro e segundo fator. O resultado encontrado reforça a interpretação das variáveis para os referidos fatores. O primeiro fator pondera melhores condições da população cujo score é maior. Valores baixos indicam piores condições. O mesmo pode ser confirmado para o segundo fator, porém relacionado às variáveis características da educação básica municipal.

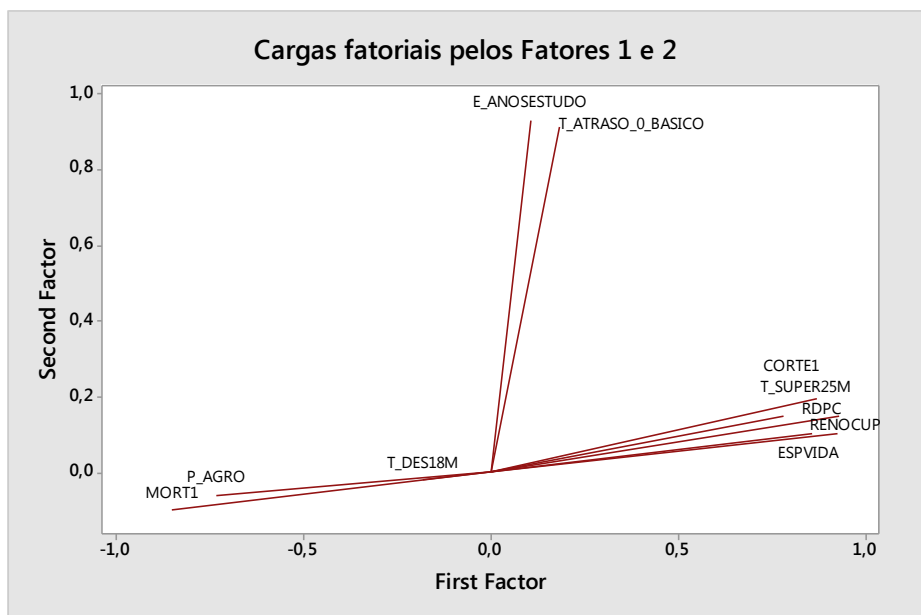


GRÁFICO 2: Cargas fatoriais das variáveis nos Fatores 1 e 2

## 5.4 Análise de *Cluster*

Para aplicar a Técnica de Análise de *Cluster* nos escores dos 3 fatores obtidos a partir da Análise Fatorial ortogonal, optou-se pelo método hierárquico aglomerativo baseado na média das distâncias euclidianas (*Average Linkage*) entre os elementos que compõem os grupos a cada etapa do processo de aglomeração dos municípios.

### 5.4.1 Estimação do número de grupos e respectivas composições

Considerando todos os 853 municípios em um único grupo o valor da Soma dos Quadrados Total (SST) encontrado foi 2.556, e as distâncias média e máxima do centróide (vetor de médias) foram iguais a 1,5971 e 4,7284, respectivamente. O valor 2.556 é adotado como referência para subsidiar o melhor nível para interromper o algoritmo de agrupamento e estimar o número de grupo da partição dos municípios.

O QUADRO 7 apresenta três principais cenários de ruptura para o nível de fusão. O primeiro ocorre no passo 3, onde o nível de fusão é 0,05357 e são formados 850 grupos na partição. Neste caso, a elevada quantidade de grupos impede que a proposta de agrupamento seja consolidada e devidamente analisada por meio das variáveis dispostas no BDGEO, o que nos levou descartar seu uso. O segundo ocorre no outro extremo da etapa de agrupamento, no passo 850, onde teríamos apenas 3 grandes grupos na partição. Esta alternativa também foi desconsiderada, pois o elevado nível de segregação oculta as características que poderiam contribuir para a proposta de caracterização das tipologias espaciais, sobrescrevendo assim as disparidades regionais de cada município. Finalmente, o cenário de ruptura entre os passos 840 e 843 se mostra promissor para a constituição da partição ideal. Isso porque estes passos apresentam elevado nível de similaridade (ainda acima de 70%) para uma quantidade de grupos mais próxima à proposta de agrupamento em nível estadual.

QUADRO 7: Modelo de Agrupamento baseado em escores resultados da análise fatorial ortogonal, distância euclidiana, método de ligação média

| Passo | Quantidade de Grupos | Nível de Similaridade | Nível de Fusão | Grupos unificados |     | Novo Grupo | Número de elementos no novo grupo |
|-------|----------------------|-----------------------|----------------|-------------------|-----|------------|-----------------------------------|
| 1     | 852                  | 99,6639               | 0,02488        | 317               | 600 | 317        | 2                                 |
| 2     | 851                  | 99,4278               | 0,04235        | 68                | 797 | 68         | 2                                 |
| 3     | 850                  | 99,2762               | 0,05357        | 583               | 587 | 583        | 2                                 |
| ...   | ...                  | ...                   | ...            | ...               | ... | ...        | ...                               |
| 840   | 13                   | 75,752                | 1,79472        | 1                 | 2   | 1          | 348                               |
| 841   | 12                   | 74,8535               | 1,86122        | 4                 | 67  | 4          | 21                                |
| 842   | 11                   | 74,7274               | 1,87055        | 35                | 108 | 35         | 44                                |
| 843   | 10                   | 73,6776               | 1,94825        | 8                 | 35  | 8          | 125                               |
| ...   | ...                  | ...                   | ...            | ...               | ... | ...        | ...                               |
| 850   | 3                    | 60,531                | 2,92129        | 1                 | 40  | 1          | 772                               |
| 851   | 2                    | 60,1131               | 2,95223        | 1                 | 4   | 1          | 851                               |
| 852   | 1                    | 42,3497               | 4,26698        | 1                 | 236 | 1          | 853                               |

Assim, passamos para o processo de determinação do número de grupos entre os passos 840 a 843. Para isso, aplicamos os critérios referentes à soma de quadrados entre grupos e residual. Feito isso, calculamos os valores do coeficiente de correlação intra classe ( $R^2$ ) e da Estatística Pseudo F, conforme TABELA 2.

TABELA 2: Aplicação dos critérios para determinação do número de grupos na partição

| Parâmetros                            | Número de grupos: 10 | Número de grupos: 11 | Número de grupos: 12 | Número de grupos: 13 |
|---------------------------------------|----------------------|----------------------|----------------------|----------------------|
| Variabilidade dentro dos grupos (SSW) | 1.072,11             | 1.004,84             | 993,78               | 926,72               |
| Variabilidade entre dos grupos (SSB)  | 1.483,89             | 1.551,16             | 1.562,22             | 1.629,28             |
| $R^2$ (%)                             | 58,06                | 60,69                | 61,12                | 63,74                |
| Estatística Pseudo F                  | 129,64               | 129,98               | 120,19               | 123,07               |

Considerando os resultados obtidos na TABELA 2, entendemos ser interessante paralisar o processo no passo em que se tem 11 grupos, pois apresenta poder de explicação representativo (60,69%) e maior valor da Estatística Pseudo F para a região analisada.

É importante salientar que o valor do coeficiente de correlação intra-classe  $R^2$  é crescente em relação ao número de conglomerados. Sendo assim, o valor do Pseudo-F

para soluções com valores de  $R^2$  semelhantes deve ser priorizado na decisão final sobre o número de *clusters* estimado. Pela Tabela 2 é possível verificar que as soluções para partições com 12 e 13 clusters apresentam valores de  $R^2$  próximos da solução com 11 clusters. Contudo, a solução com 11 *clusters* resulta num maior valor de Pseudo-F, o que fortalece sua concepção de partição ideal para o universo de dados observados.

A solução encontrada para a partição constituída de 11 grupos pode ser observada no QUADRO 8.

QUADRO 8: Partição final – k=11

| <b>Grupos</b>    | <b>Quantidade de elementos</b> | <b>Somas dos Quadrados (SSW)</b> | <b>Distância média do centróide</b> | <b>Distância máxima do centróide</b> |
|------------------|--------------------------------|----------------------------------|-------------------------------------|--------------------------------------|
| <i>Cluster1</i>  | 348                            | 460,908                          | 1,08214                             | 2,22157                              |
| <i>Cluster2</i>  | 21                             | 26,469                           | 1,06401                             | 1,88005                              |
| <i>Cluster3</i>  | 217                            | 266,394                          | 1,01844                             | 2,69172                              |
| <i>Cluster4</i>  | 64                             | 73,608                           | 0,9958                              | 2,11892                              |
| <i>Cluster5</i>  | 81                             | 72,65                            | 0,89408                             | 1,79543                              |
| <i>Cluster6</i>  | 56                             | 43,95                            | 0,82251                             | 1,4541                               |
| <i>Cluster7</i>  | 16                             | 13,932                           | 0,87269                             | 1,43843                              |
| <i>Cluster8</i>  | 44                             | 45,443                           | 0,9139                              | 1,87331                              |
| <i>Cluster9</i>  | 2                              | 0,355                            | 0,4216                              | 0,4216                               |
| <i>Cluster10</i> | 2                              | 0,565                            | 0,53147                             | 0,53147                              |
| <i>Cluster11</i> | 2                              | 1,435                            | 0,84705                             | 0,84705                              |

O centróide do grupo representa o vetor de médias do grupo.

A distância média do centróide retrata a média das distâncias das observações do grupo em relação ao vetor de médias do grupo.

No QUADRO 8, os grupos 9, 10 e 11 são formados por Araponga e Pedra Bonita, Belo Horizonte e Nova Lima, Nacip Raydan e Delta, respectivamente. Valores extremos nos escores destes municípios em diferentes fatores proporcionaram sua distinção específica dos demais grupos. O remanescente da base de dados foi distribuído de modo a formar três contextos: no primeiro temos grupos entre 16 e 81 elementos – *Clusters* 2, 4, 5, 6, 7 e 8; no segundo, observamos a existência de dois grandes grupos – *Clusters* 1 e 3 – com 348 e 217 elementos, respectivamente. A composição dos 11 grupos em termos dos municípios correspondentes pode ser encontrada no APÊNDICE C.

A organização destes dados num dendrograma pode ser observada no GRÁFICO 3. Neste, observamos a possibilidade de separar ainda mais a partição para próximos

grupos. Contudo, como destacado anteriormente, o ganho de explicação não seria relevante frente à complexidade do ganho analítico frente aos objetivos desta monografia.

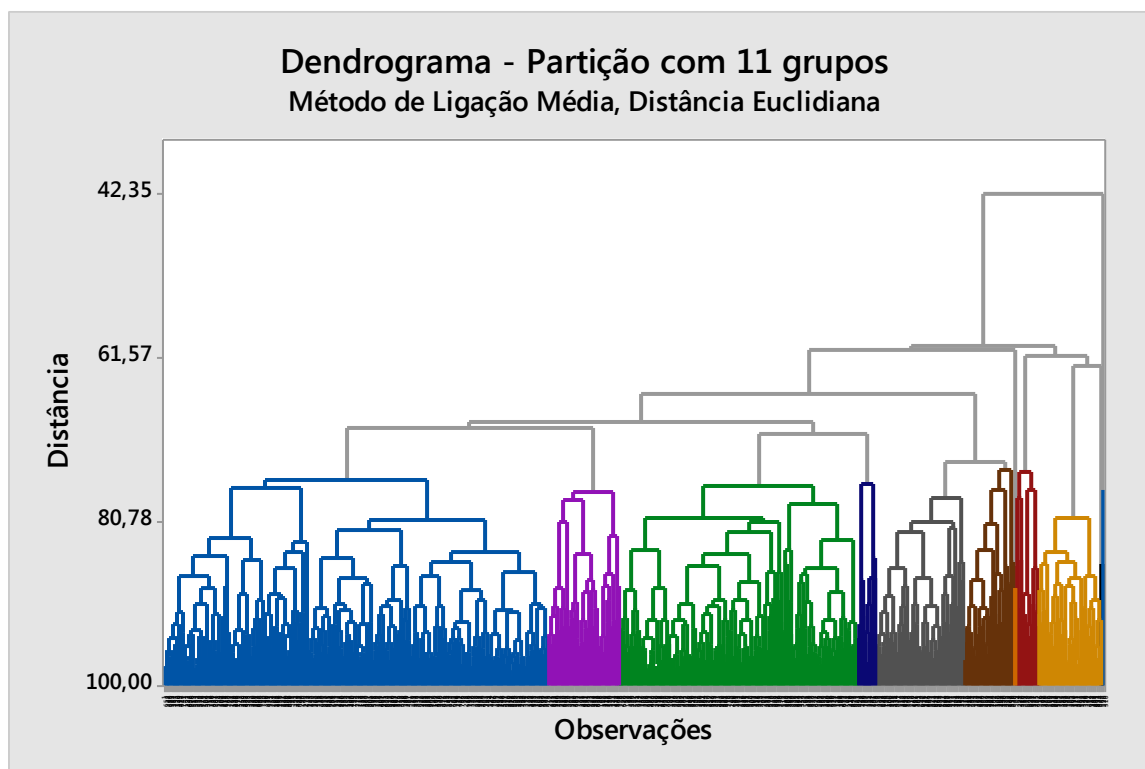


GRÁFICO 3: Dendrograma – partição constituída de 11 grupos

Para facilitar esta análise, representamos espacialmente na FIGURA 8 a partição apresentada no dendrograma do GRÁFICO 3 correspondente ao QUADRO 8. A técnica de coleção de mapas permite verificar com maior clareza a distribuição espacial dos elementos em cada grupo. Primeiramente, observa-se a descontinuidade espacial constatada entre as porções norte-nordeste e sul-sudoeste do Estado, representadas pelos Grupos 1 e 3. Este cenário corrobora para a hipótese de influência regional do Estado de São Paulo, associados principalmente a presença de infraestrutura rodoviária – a qual favorece a intensidade dos fluxos demográficos. Posteriormente, a variabilidade espacial dos Grupos 4 e 6 reforça as descontinuidades territoriais corroboradas pela polarização dos elementos destes grupos.

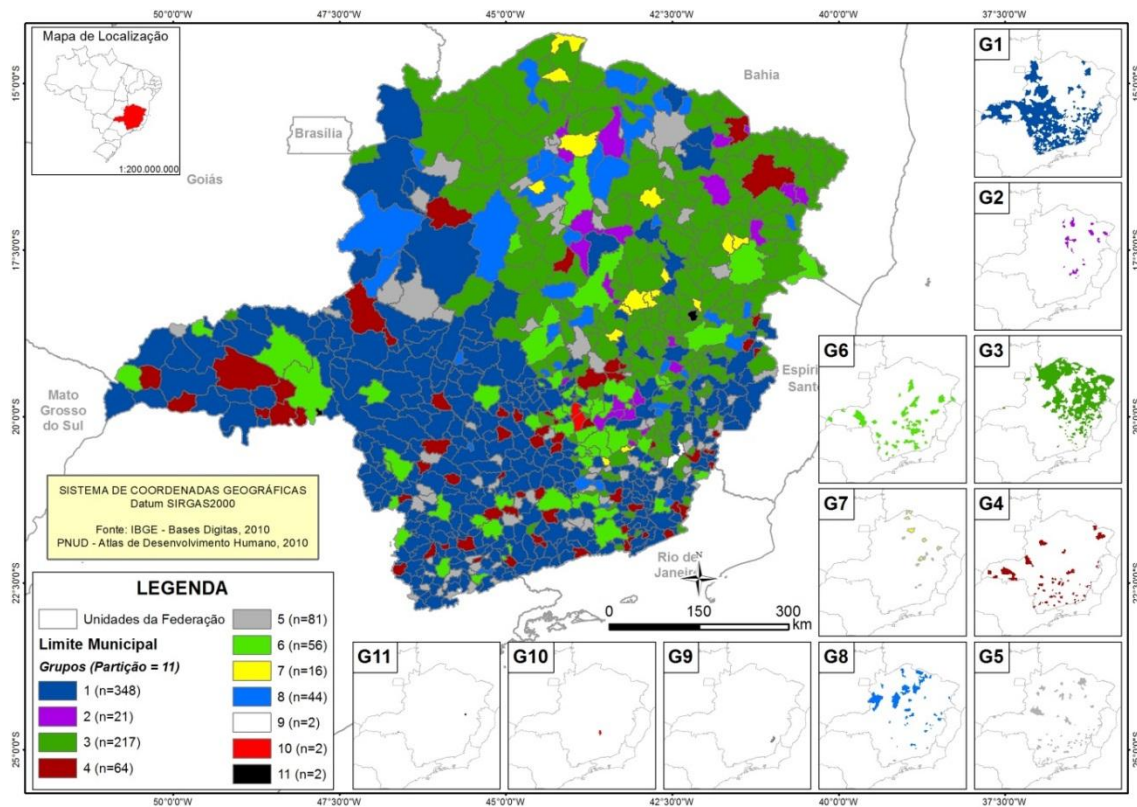


FIGURA 8: Espacialização da partição para 11 grupos – coleção de mapas

#### 5.4.2 Análise de perfil dos grupos de acordo com os escores dos fatores

Embora a representação gráfica da partição em formato de mapa contribua para a validação de hipóteses destacadas preliminarmente, é insuficiente para confirmação de tais cenários. Diante disso, optamos por comparar as estatísticas descritivas dos escores dos Fatores 1, 2 e 3 apresentadas nos QUADROS 9, 10 e 11 e GRÁFICOS 4, 5 e 6, respectivamente. Tais resultados podem ser encontrados no APÊNDICE C, o qual especifica em qual grupo cada município foi classificado.

QUADRO 9: Estatísticas descritivas para o Fator 1 (Partição n=11)

| Fator   | Grupo | Quantidade de elementos | Média   | Desvio Padrão | Mínimo  | Mediana | Máximo  | Amplitude |
|---------|-------|-------------------------|---------|---------------|---------|---------|---------|-----------|
| Fator 1 | 1     | 348                     | 0,4086  | 0,6855        | -1,4617 | 0,4024  | 2,1467  | 3,6084    |
|         | 2     | 21                      | -0,607  | 0,69          | -1,673  | -0,299  | 0,356   | 2,029     |
|         | 3     | 217                     | -0,7235 | 0,6339        | -2,3514 | -0,8064 | 1,0431  | 3,3945    |
|         | 4     | 64                      | 0,55    | 0,6702        | -0,8892 | 0,5478  | 1,8566  | 2,7459    |
|         | 5     | 81                      | -0,6937 | 0,5632        | -1,8665 | -0,7492 | 0,5987  | 2,4651    |
|         | 6     | 56                      | 1,7064  | 0,5991        | 0,7907  | 1,6327  | 2,9329  | 2,1422    |
|         | 7     | 16                      | -1,1378 | 0,3019        | -1,8761 | -1,0688 | -0,7423 | 1,1338    |
|         | 8     | 44                      | -0,789  | 0,729         | -2,079  | -0,889  | 0,919   | 2,998     |
|         | 9     | 2                       | -1,902  | 0,47          | -2,234  | -1,902  | -1,569  | 0,665     |
|         | 10    | 2                       | 4,339   | 0,376         | 4,073   | 4,339   | 4,605   | 0,532     |
|         | 11    | 2                       | 0,514   | 0,492         | 0,166   | 0,514   | 0,861   | 0,696     |

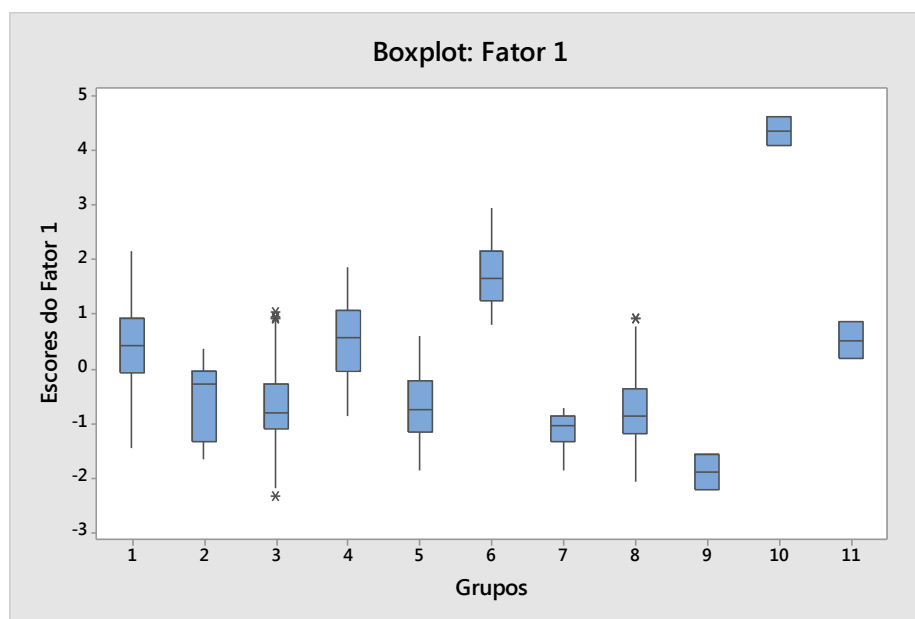


GRÁFICO 4: Boxplot dos escores no primeiro fator – 11 grupos

Considerando que o Fator 1 representa o índice de longevidade agregado com renda, população com ensino superior e profissionais empregados no setor agropecuário, observamos que os municípios que compõem o grupo 10 são considerados os de melhor desenvolvimento no Estado, cuja média neste fator é 4,339 – Belo Horizonte e Nova Lima.

Seguido pelo grupo 6 de média 1,7064, o qual pode ser indicado como mais próximo aos polos de desenvolvimento regional – semelhante aos conceitos de Cidades Médias. Este é composto por 56 cidades, dentre elas Uberlândia, Uberaba, Juiz de Fora, Poços de Caldas, Pouso Alegre, Araxá, Curvelo, Diamantina, Montes Claros, Pirapora, Governador Valadares, Teófilo Otoni, Itabira, Ouro Branco, Contagem, Betim e outros semelhantes. Sua função característica de Cidade Média pode fortalecer também ao se observar o padrão de distribuição espacial destes municípios (FIGURA 8).

Os Grupos 2, 3, 5, 7, 8 e 9, possuem as menores médias neste fator, representadas pelos sinais negativos em seus escores, o que caracteriza municípios identificados em níveis diferenciados, associadas principalmente ao desempenho nos demais fatores.

Os Grupos 1 e 3 possuem maior quantidade de elementos e, conforme esperado, maior amplitude (3,6084 e 3,3945, respectivamente).

Visto como índice de desempenho dos estudantes matriculados na educação básica em 2010, o Fator 2 mostra, através do QUADRO 10 e GRÁFICO 5, o melhor desempenho dos municípios contidos no grupo 5, 6 e 8 com médias 0,3916, 1,324 e 1,5754, respectivamente.

Os grupos 5 e 8 se apresentam-se como centros emergentes cujo desempenho está acima dos demais grupos formados. Uma das prováveis causas vinculadas a este desempenho pode estar associada ao volume populacional.

Considerando que a educação básica é de obrigação legal específica do poder municipal, o baixo volume populacional associado à atenção municipal para esta faixa pode contribuir para este resultado. Sua proximidade às cidades de nível superior pode estar vinculada a esta configuração, pois tais cidades concentram maior quantidade de instituições de ensino e potencializam seu poder de atração. Contudo, faz-se necessário detalhar estes casos para confirmar a existência destes cenários.



QUADRO 10: Estatísticas descritivas para o Fator 2 (Partição n=11)

| Fator   | Grupo | Quantidade de elementos | Média   | Desvio Padrão | Mínimo  | Mediana | Máximo  | Amplitude |
|---------|-------|-------------------------|---------|---------------|---------|---------|---------|-----------|
| Fator 2 | 1     | 348                     | 0,1505  | 0,5832        | -1,1982 | 0,1498  | 2,0721  | 3,2703    |
|         | 2     | 21                      | 0,385   | 0,604         | -1,05   | 0,513   | 1,655   | 2,704     |
|         | 3     | 217                     | -0,518  | 0,5745        | -1,6673 | -0,5235 | 0,8422  | 2,5095    |
|         | 4     | 64                      | -1,579  | 0,3993        | -2,4677 | -1,5748 | -0,6706 | 1,797     |
|         | 5     | 81                      | 1,324   | 0,528         | 0,5662  | 1,2531  | 2,5854  | 2,0192    |
|         | 6     | 56                      | 0,3916  | 0,443         | -0,5677 | 0,3739  | 1,6267  | 2,1943    |
|         | 7     | 16                      | -2,223  | 0,3888        | -2,9741 | -2,1672 | -1,4267 | 1,5474    |
|         | 8     | 44                      | 1,5754  | 0,4566        | 0,5567  | 1,5684  | 2,7297  | 2,1729    |
|         | 9     | 2                       | -1,198  | 0,291         | -1,404  | -1,198  | -0,993  | 0,411     |
|         | 10    | 2                       | 0,118   | 0,459         | -0,207  | 0,118   | 0,442   | 0,649     |
|         | 11    | 2                       | -3,8831 | 0,0779        | -3,9382 | -3,8831 | -3,828  | 0,1102    |

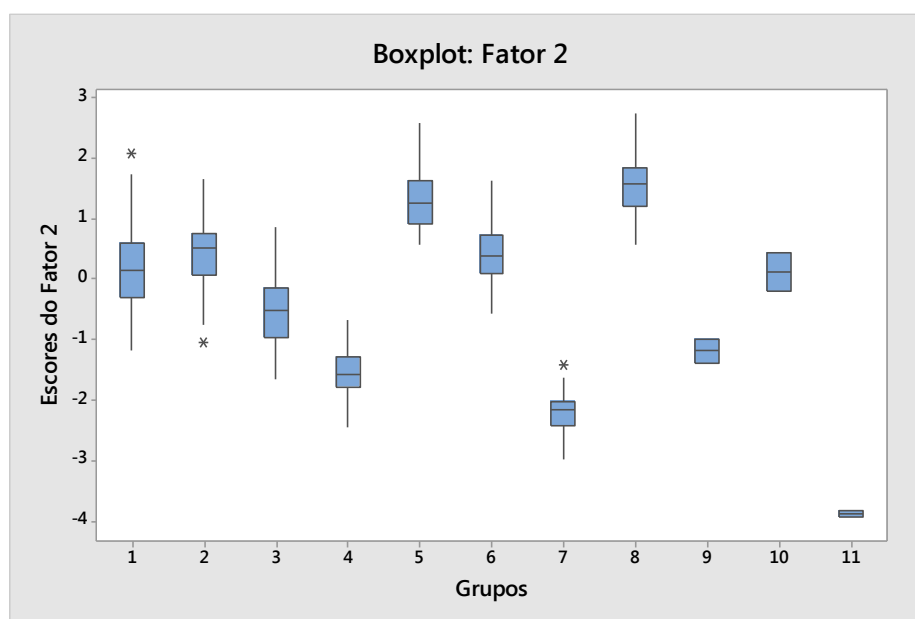


GRÁFICO 5: Boxplot dos escores no segundo fator – 11 grupos

No que se refere ao Fator 3 o QUADRO 11 e o GRÁFICO 6 permitem visualizar a elevada variabilidade na maioria dos grupos quando comparados ao comportamento dos mesmos nos outros fatores. Neste caso, os municípios com menores escores representam maior taxa de desempregados presentes na população economicamente ativa (PEA).

Assim, os municípios presentes no grupo 2 (média -2,337) possuem maiores taxas e podem ser objeto de políticas públicas específicas. Adicionalmente, deve-se destacar que todos os municípios presentes neste grupo estão geograficamente próximos – até mesmo cercados – por municípios do grupo 3, o qual predominam as regiões norte e nordeste de Minas Gerais (FIGURA 8). Neste fator, os Grupos 4, 5 e 9 apresentam maiores médias da partição (0,5507, 0,665 e 1,508, respectivamente), o que representa menores taxas de desemprego nos municípios agregados.

QUADRO 11: Estatísticas descritivas para o Fator 3 (Partição n=11)

| Fator   | Grupo | Quantidade de elementos | Média   | Desvio Padrão | Mínimo  | Mediana | Máximo | Amplitude |
|---------|-------|-------------------------|---------|---------------|---------|---------|--------|-----------|
| Fator 3 | 1     | 348                     | 0,533   | 0,7199        | -1,0841 | 0,5614  | 2,5123 | 3,5964    |
|         | 2     | 21                      | -2,337  | 0,694         | -3,97   | -2,042  | -1,459 | 2,51      |
|         | 3     | 217                     | -0,5261 | 0,7081        | -2,9656 | -0,5306 | 1,1744 | 4,14      |
|         | 4     | 64                      | 0,5507  | 0,7482        | -1,209  | 0,5835  | 2,2186 | 3,4276    |
|         | 5     | 81                      | 0,665   | 0,5587        | -0,3733 | 0,5525  | 2,1877 | 2,561     |
|         | 6     | 56                      | -1,2293 | 0,4939        | -2,4409 | -1,2569 | 0,0307 | 2,4716    |
|         | 7     | 16                      | -0,394  | 0,829         | -1,735  | -0,104  | 0,53   | 2,265     |
|         | 8     | 44                      | -0,8372 | 0,5634        | -1,9904 | -0,7062 | 0,3192 | 2,3095    |
|         | 9     | 2                       | 1,508   | 0,224         | 1,35    | 1,508   | 1,666  | 0,316     |
|         | 10    | 2                       | -1,379  | 0,461         | -1,705  | -1,379  | -1,052 | 0,652     |
|         | 11    | 2                       | 0,187   | 1,089         | -0,583  | 0,187   | 0,958  | 1,541     |

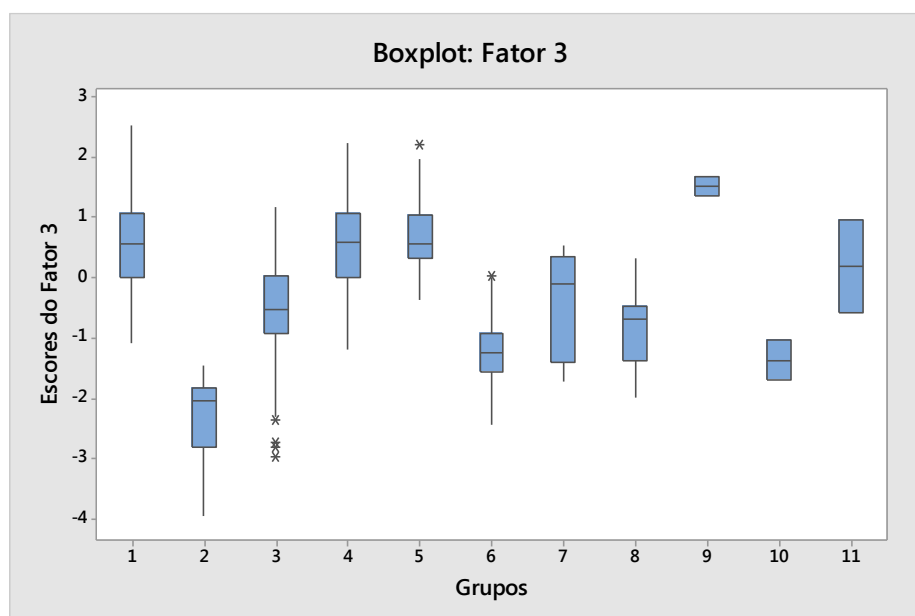


GRÁFICO 6: Boxplot dos escores no terceiro fator – 11 grupos

Os GRÁFICOS 7, 8 e 9 apresentam os diagramas dos pares de fatores com indicação dos grupos ao qual pertencem os municípios. A classificação dos escores por grupo reforça as tipologias espaciais e sua faixa de sobreposição, a qual poderia ser interpretada como algo semelhante à zona de transição dos municípios pelos fatores. Poucos municípios poderiam ser interpretados como observações atípicas, principalmente Belo Horizonte e Nova Lima – reunidos no Grupo 10 – com maiores escores no Fator 1.

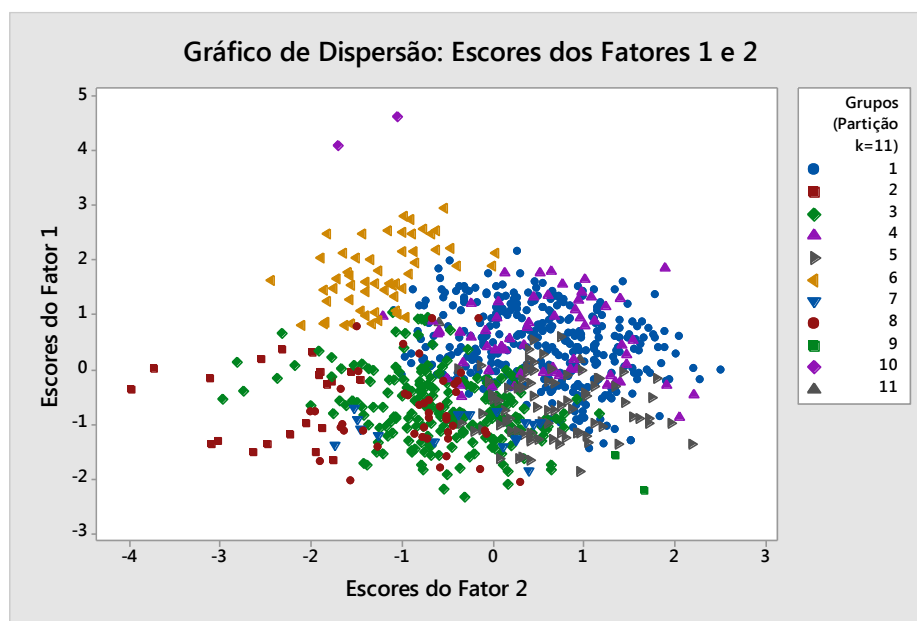


GRÁFICO 7: Gráfico de Dispersão dos escores por grupo – Fatores 1 e 2

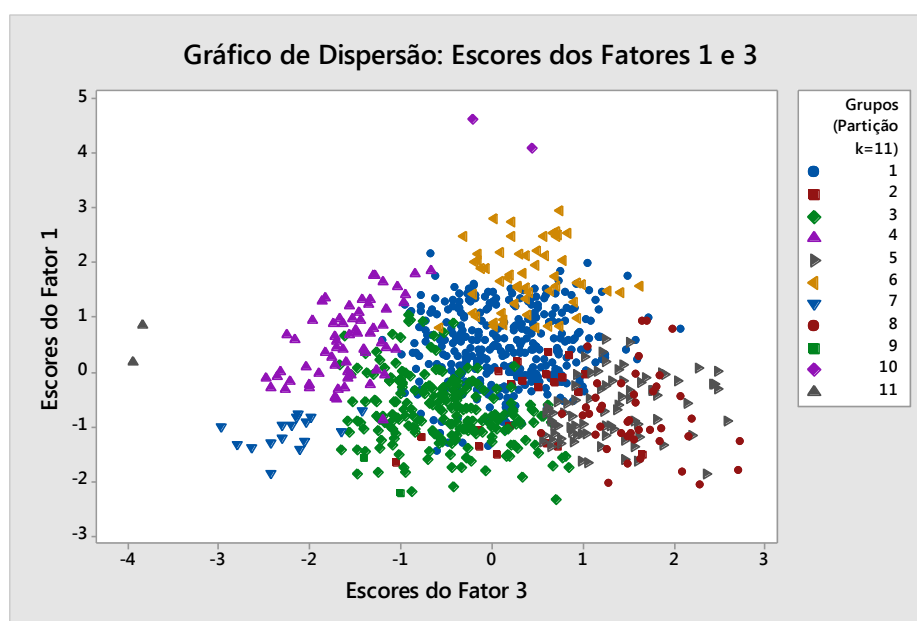


GRÁFICO 8: Gráfico de Dispersão dos escores por grupo – Fatores 1 e 3

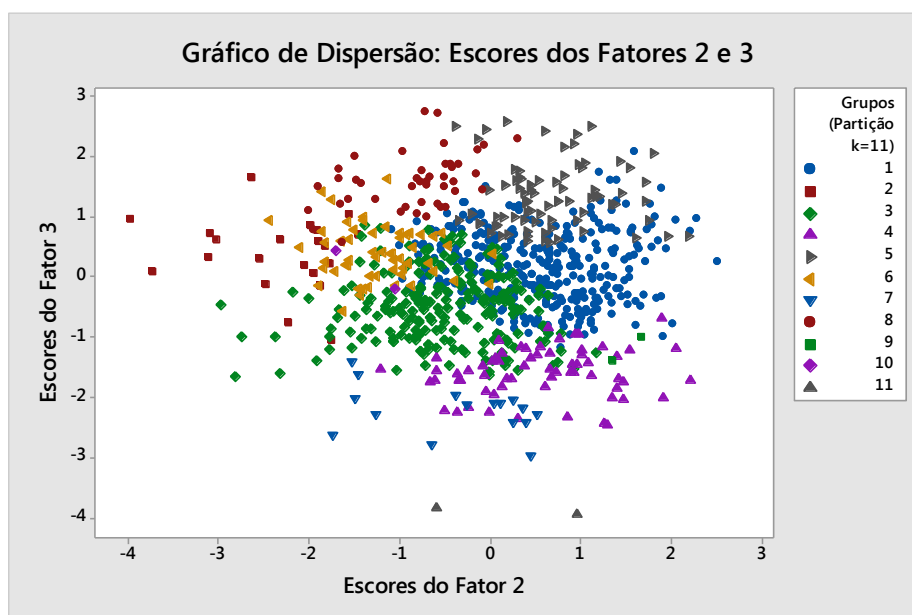


GRÁFICO 9: Gráfico de Dispersão dos escores – Fatores 2 e 3

Finalmente, apresentamos no GRÁFICO 10 o Box-Plot do Índice de Desenvolvimento Humano Municipal (IDHM) de 2010 por grupo. Apesar de não inserida no modelo, tal variável retrata o cenário apresentado nos resultados do modelo de análise fatorial e posterior agrupamento dos escores, corroborando para a proposta de caracterização das tipologias espaciais apresentada na seção 5.3, pois se assemelha ao padrão de distribuição dos municípios, baseado em seus escores para o Fator 1.

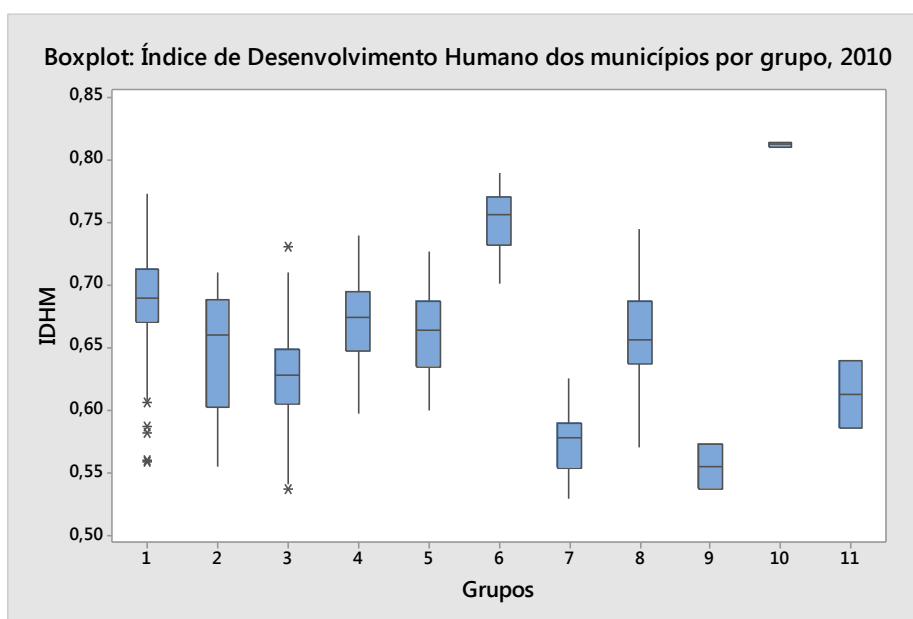


GRÁFICO 10: Gráfico Box-Plot do Índice de Desenvolvimento Humano em MG, 2010

## 6 CONSIDERAÇÕES FINAIS

A aplicação de técnicas de estatística multivariada se mostrou de grande contribuição para alcançar a proposta de identificação das tipologias espaciais formada pelos municípios em Minas Gerais a partir de variáveis socioeconômicas. A capacidade de sintetizar a informação de diversas variáveis em índices representativos e explicáveis incrementa o poder de análise e sua interpretação sobre o universo de dados. De modo complementar, as informações do modelo de análise fatorial permitem ao gestor público direcionar recursos, incentivos ou ações de caráter pragmático em áreas (variáveis) com maior poder de impacto no respectivo fator. Assim, estas técnicas favorecem a tomada de decisão frente aos cenários de desigualdade em Minas Gerais.

Os escores obtidos por meio da análise fatorial se mostraram úteis e aplicáveis em estudos de diversas finalidades. Dentre elas, o emprego destes dados para construção de grupos se mostrou de grande relevância, pois viabilizou a formação dos grupos e, até mesmo, a sugestão de hierarquia entre estes. Como resultado, vislumbramos a possibilidade de direcionar políticas públicas regionais para a gestão territorial. A oportunidade de analisar além de variáveis únicas ou mero somatório destas constitui benefício direto para o delineamento de perfis municipais. Isso porque a partição encontrada neste estudo permite, ainda, a interpretação por meio dos fatores construídos. Logo, pode-se observar o padrão característico dos municípios em índices, o que esclarece a leitura e interpretação da realidade.

Adicionalmente, o uso de geoprocessamento na espacialização destes dados proporciona ao pesquisador visualizar o comportamento espacial destes dados. Para o gestor, a simples lista dos maiores, melhores e seus extremos assume papel único e exclusivamente informativo. Contudo, observar o padrão de distribuição dos grupos associados ao desempenho nos respectivos índices, incentiva a reflexão sobre as causas e medidas para amenizar tais discrepâncias. A comparação com seus vizinhos, o diagnóstico da função do município em sua região e a possibilidade de estabelecer hierarquia instiga o direcionamento de estudos e ações direcionadas, as quais sugerem melhor resultado e emprego de recursos.

Ademais, a partir do estudo realizado, atentamos para a possibilidade de avançar com os estudos em menores níveis de agrupamento. Avançar nas etapas de análise de *Cluster* pode responder questões específicas aos elementos discrepantes e explicar as faixas de sobreposição encontradas nos gráficos de dispersão. Contudo, não foi objetivo deste estudo chegar a este nível. A apresentação das discrepâncias regionais, os elementos extremos e o reforço do papel das Cidades Médias na hierarquia regional servem de incentivo para continuidade de estudos mais detalhados, em nível e contexto adequado.

Finalmente, reforçamos a contribuição de estudos interdisciplinares para as temáticas envolvidas. O emprego de técnicas de geoprocessamento, teorias inerentes a Geografia e métodos específicos da Estatística propiciou a visão identificada neste estudo. Aplicados em conjunto, tais ciências ampliaram a visão sobre a realidade diagnóstica – o que, quando realizado separadamente, pode demonstrar apenas uma faceta desta realidade.

## REFERÊNCIAS BIBLIOGRÁFICAS

AMORIM FILHO, O. B. **Um esquema metodológico para o estudo das cidades médias** In: ENCONTRO NACIONAL DE GEÓGRAFOS, 2., 1976, Belo Horizonte IGC/UFMG, 1976, 600 p., p. 6-15.

AMORIM FILHO, O.; RIGOTTI, J. I. R. Os Limiares Demográficos na Caracterização das Cidades Médias. In: **XIII Encontro da Associação Brasileira de Estudos Populacionais**. Ouro Preto: PUC Minas, 2002.

AMORIM FILHO, O.B.; BUENO, M. E. T.; ABREU, J. F. **Cidades de porte médio e o programa de ações sócio-educativo-culturais para as populações carentes do meio urbano em Minas Gerais**. Boletim de Geografia Teórica. Rio Claro, 12 (23-24); 33-46, 1982.

BARBETTA, P. A. **Estatística aplicada às ciências sociais**. 6. ed. rev. Florianópolis: Ed. UFSC, 2006. 315p.

BAROUCHE, J.M.; SAPORTA, G. **Análise de dados**, Rio de Janeiro: Zahar, 1982.

BRADFORD, M. G; KENT, W. A. (William Ashley). **Geografia humana: teorias e suas aplicações**. Lisboa (Portugal): Gradiva, 1987. 264p. (Trajectos,7)

BRAVO, D. P. **Análise multivariada de assentamentos urbanos subnormais na Vila Alto Vera Cruz: um estudo exploratório**. Dissertação de mestrado. Pontifícia Universidade Católica de Minas Gerais, Belo Horizonte, 2004.

BURROUGH, P. A. **Principles of geographical information systems for land resources assessment**. 1986: 54-54.

CÂMARA, G; DAVIS, C; MONTEIRO, A.M. **Introdução à Ciência da Geoinformação**. São José dos Campos: INPE (Instituto Nacional de Pesquisas Espaciais), 2001.

CÂMARA, G; MONTEIRO, A. M.; FUCKS, S. D.; CARVALHO, M. S (eds). In: **Análise Espacial de Dados Geográficos. Análise Espacial e Geoprocessamento**. Brasília: EMBRAPA, 2004.

CARVALHO, P. F. B. **Subsistemas e tipologias da migração populacional no Brasil a partir de dados censitários: uma proposta metodológica**. 2009. 168f. Tese (Doutorado em Tratamento da Informação Espacial) – Instituto de Ciências Humanas, Pontifícia Universidade Católica de Minas Gerais, Belo Horizonte, 2009.

CARVALHO, P. F. B. **Uma proposta para o ensino introdutório da quantificação em Geografia, com uso do software MATLAB**. Dissertação (Mestrado em Geografia) – Pontifícia Universidade Católica de Minas Gerais. 2002.

CARVALHO, P. F. B.; VIEIRA JUNIOR, J. A. In: Resumos: 16º Seminário de Iniciação Científica. **O Uso de Métodos Quantitativos na Geografia**. Belo Horizonte:

Ed. PUC Minas, 2008. p.158.

COWEN, D. J.; JENSEN, J. R.; BRESNAHAN, P. J.; EHLER, G. B.; GRAVES, D.; HUANG, X.; WIESNER, C.; MACKEY, H. E. **The design and implementation of an integrated geographic information system for environmental applications.** Photogrammetric Engineering and Remote Sensing 61.11 (1995): 1393-1404.

CROCCO, M. A., GALINARI, R., SANTOS, F., LEMOS, M. B., SIMÕES, R. **Metodologia de identificação de arranjos produtivos locais potenciais: uma nota técnica.** Belo Horizonte: UFMG/CEDEPLAR, 2003. (Texto para Discussão: 191).

ESRI. **ArcGIS** for Desktop, versão 10.2. Copyright 1999-2014 Esri Inc. All rights Reserved.

FAISSOL, S. (org.) **Tendências atuais na geografia urbano/regional.** Rio de Janeiro: FIBGE, 1978.

GERARDI, L. H. O.; SILVA, B.-C. N. **Quantificação em Geografia.** São Paulo: Difel, 1981.

HADDAD, P. R; BANCO DO NORDESTE DO BRASIL. **Economia regional: teorias e métodos de análise.** Fortaleza: BNB, 1989. 694p. (Estudos econômicos e sociais (Banco do Nordeste do Brasil. Escritório Técnico de Estudos Econômicos do Nordeste).

IBGE. **Estimativas populacionais para os municípios e para as Unidades da Federação brasileiros em 01.07.2015.** Acesso: [http://www.ibge.gov.br/home/estatistica/populacao/estimativa2015/estimativa\\_dou.shtm](http://www.ibge.gov.br/home/estatistica/populacao/estimativa2015/estimativa_dou.shtm) em Junho de 2017.

IPEA; UNIVERSIDADE ESTADUAL DE CAMPINAS; BRASIL. **Caracterização e tendências da rede urbana do Brasil.** Campinas: UNICAMP: IPEA, 1999. 2v (Coleção pesquisa;. n.3)

JOHNSON, R. A.; WICHERN, D. W. **Applied Multivariate Statistical Analysis** . 5a. Ed. New Jersey: Prentice Hall, Inc, 2002

KAGEYAMA, A.; LEONE, E.T. **Uma tipologia de municípios paulistas com base em indicadores sociodemográficos.** Campinas: IE/UNICAMP, 1999. (Textos para discussão: nº 66)

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

LAJUGIE, J. **Os sistemas econômicos.** 3. ed. São Paulo: Difusão Europeia do Livro, 1971. 130p.

MANLY, B. F. J. **Multivariate statistical methods: a primer.** Chapman and Hall, London, 1986.

MILLIGAN, G. W.; COOPER, M. C. **An examination of procedures for determining**



**the number of cluster in a data set.** *Psychometrika*, v. 50, n.2, p. 159-179, 1985.

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

MINGOTI, S. A.; SILVA, A. F. **Um Exemplo de Aplicação de Técnicas de Estatística Multivariada na Construção de Índices de Preços.** *Nova Economia* (Belo Horizonte), v.7, n.2, p. 203-212, 1997.

**MINITAB.** versão 16.1. Copyright 1991-2004 LEAD Technologies, Inc. Allrightsreserved.

PEREIRA, F. M. **Cidades médias brasileiras: uma tipologia a partir de suas (des)economias de aglomeração.** Dissertação (Mestrado) – UFMG: CEDEPLAR. Belo Horizonte: 2003

PERROUX, F. **Economia e sociedade: coacção-troca-dom.** São Paulo: Duas Cidades, 1961. 215p.

PROGRAMA DAS NAÇÕES UNIDAS PARA O DESENVOLVIMENTO (PNUD). **Atlas do desenvolvimento humano no Brasil.** Disponível em:  
< <http://atlasbrasil.org.br/2013/pt/> >. Acesso em: 03 de setembro de 2017.

SEIFODDINI, H. K. Single linkage versus average linkage clustering in machine cells formation applications. **Computers & Industrial Engineering.** v. 16, n. 3, p. 419-426, 1989.

SHAPIRO, S.; WILK, M. B. **An analysis of variance test for normality.** *Biometrika*, 52, p. 591-611, 1965.

SILVA, L. V. D. **Tipologia e hierarquização no Sul de Minas utilizando métodos e técnicas de estatística multivariada.** Dissertação de mestrado. Pontifícia Universidade Católica de Minas Gerais, Belo Horizonte, 2002.

SIMÃO, M. L. R. **Caracterização espacial da produção cafeeira de Minas Gerais: um estudo exploratório utilizando técnicas de Análise Espacial de Estatística Multivariada.** Dissertação de mestrado. Pontifícia Universidade Católica de Minas Gerais, Belo Horizonte, 1999.

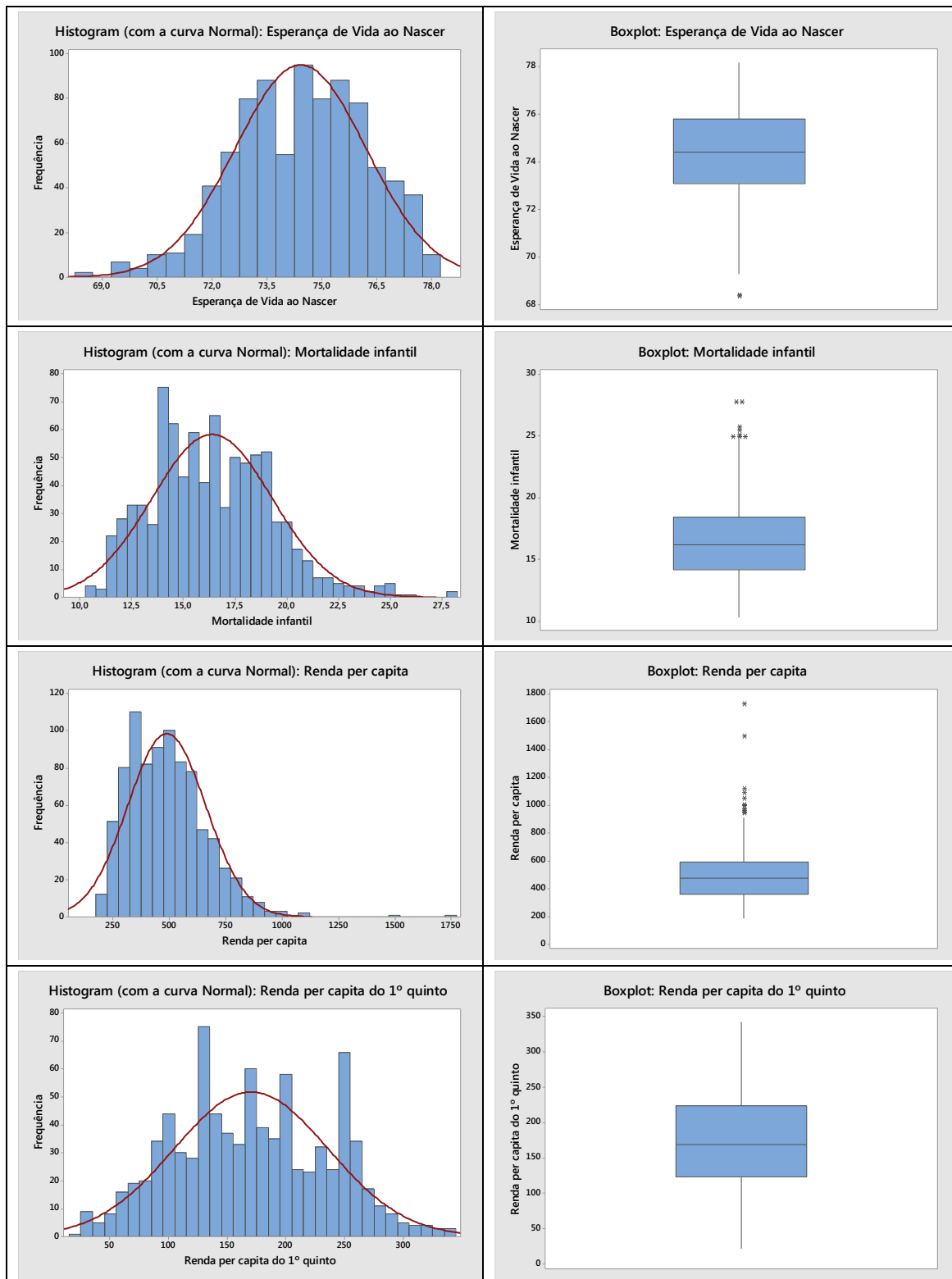
SIMÕES, R.; AMARAL, P.; GUIMARÃES, C. **Rede urbana na oferta de serviços de saúde em Minas Gerais: uma análise multivariada para 2002.** In: Anais... XIV ENCONTRO NACIONAL DE ESTUDOS POPULACIONAIS, Caxambu, ABEP, 2004.

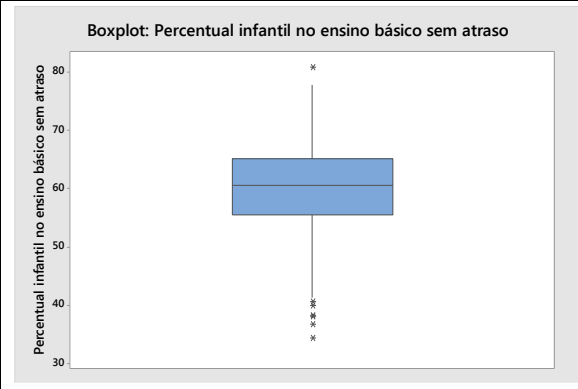
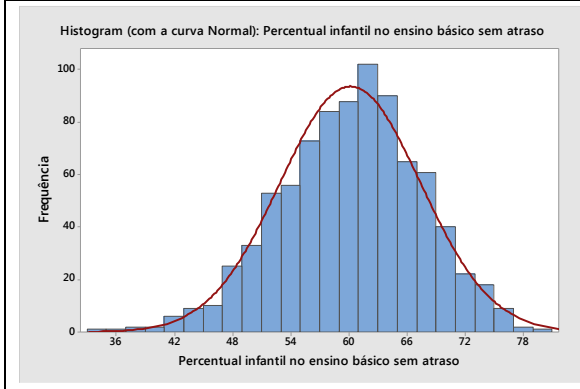
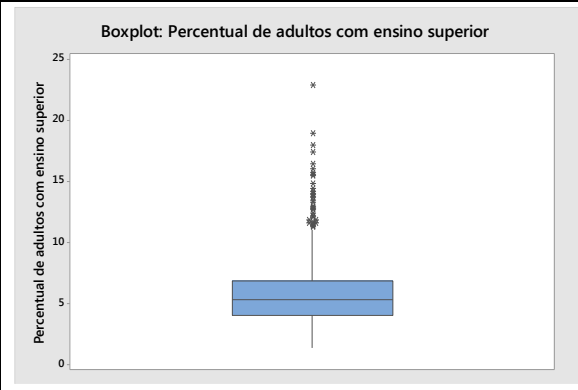
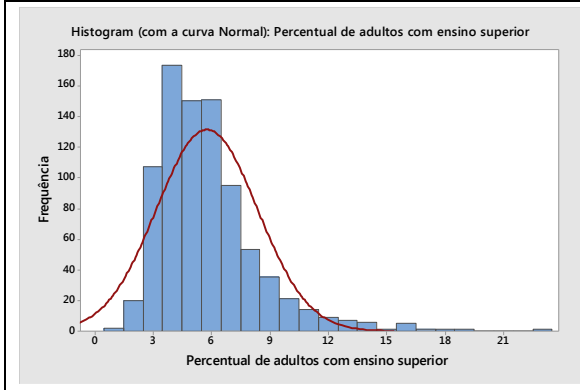
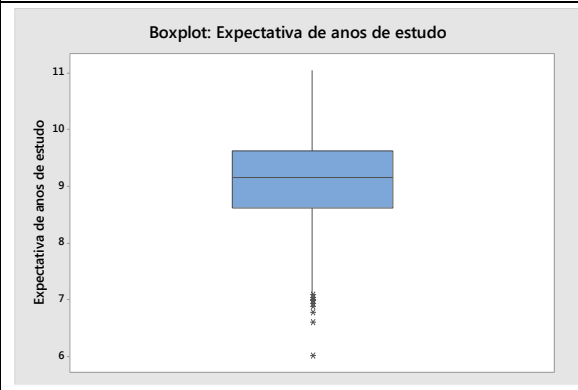
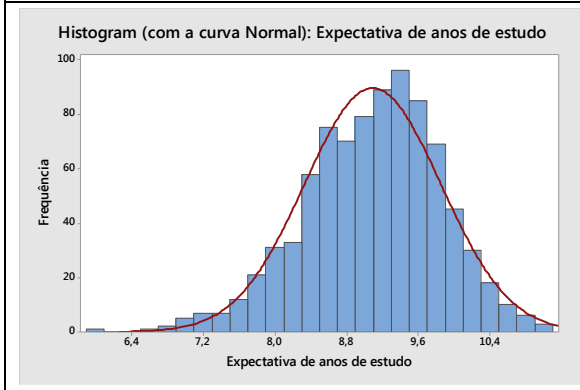
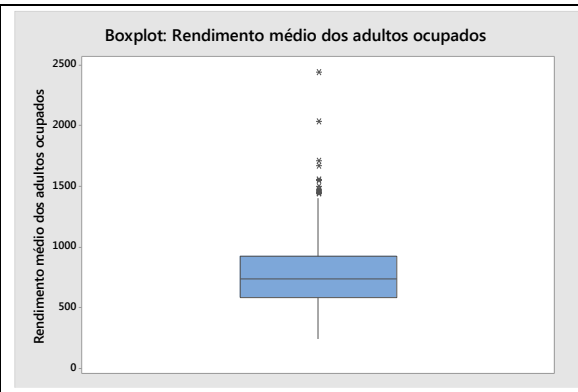
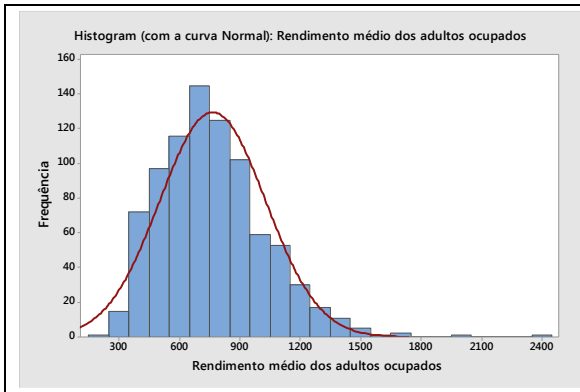
SIMÕES, R. F. **Métodos de análise regional e urbana: diagnóstico aplicado ao planejamento.** Belo Horizonte: UFMG – CEDEPLAR, 2005. (Textos para discussão: 259).

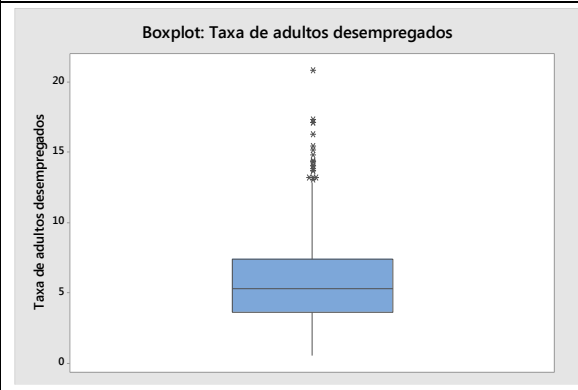
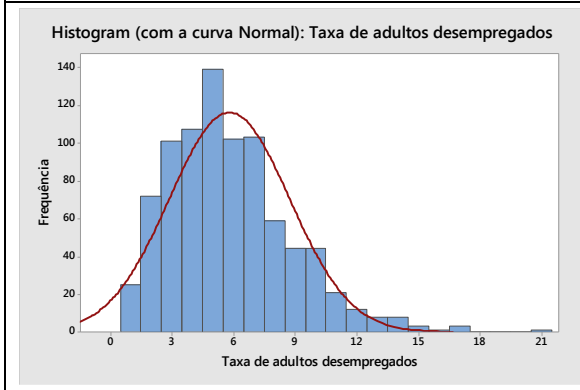
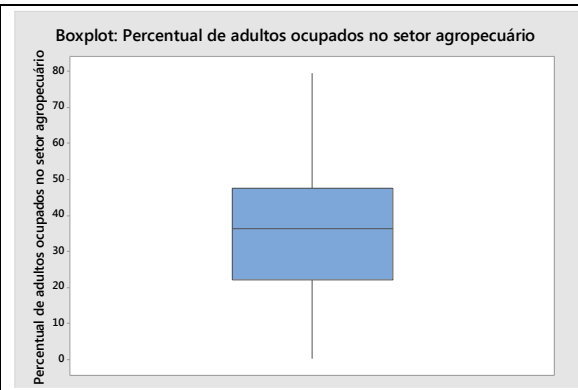
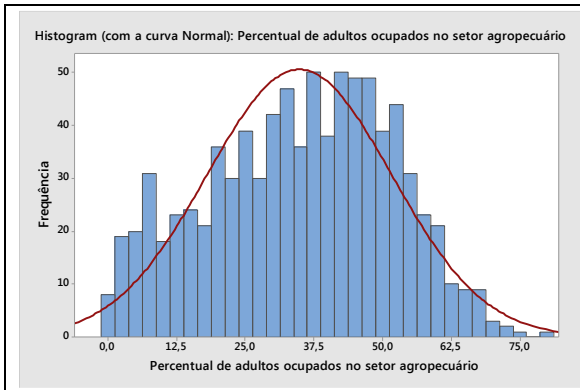
# APÊNDICE

## APÊNDICE A – Gráficos das Estatísticas Descritivas das variáveis utilizadas no estudo

Figura AA1







APÊNDICE B – Valores das correlações e respectivas siglas das variáveis utilizadas no estudo

Quadro AB1: Código das variáveis usadas no Quadro AB2

| <b>CÓDIGO</b> | <b>SIGLA</b>      | <b>DENOMINAÇÃO</b>                                   |
|---------------|-------------------|--|
| A             | ESPVIDA           | Esperança de Vida ao Nascer                          |
| B             | MORT1             | Mortalidade infantil                                 |
| C             | RDPC              | Renda per capita                                     |
| D             | CORTE1            | Renda per capita do 1º quinto                        |
| E             | RENOCUP           | Rendimento médio dos adultos ocupados                |
| F             | E_ANOESTUDO       | Expectativa de anos de estudo                        |
| G             | T_SUPER25M        | Percentual de adultos com ensino superior            |
| H             | T_ATRASO_0_BASICO | Percentual infantil no ensino básico sem atraso      |
| I             | P_AGRO            | Percentual de adultos ocupados no setor agropecuário |
| J             | T_DES18M          | Taxa de adultos desempregados                        |

Quadro AB2: Valores das correlações entre as variáveis empregadas no estudo

|   | A      | B      | C      | D      | E      | F      | G      | H      | I      | J |
|---|--------|--------|--------|--------|--------|--------|--------|--------|--------|---|
| A | -      |        |        |        |        |        |        |        |        |   |
| B | -0,996 | -      |        |        |        |        |        |        |        |   |
| C | 0,745  | -0,730 | -      |        |        |        |        |        |        |   |
| D | 0,739  | -0,730 | 0,877  | -      |        |        |        |        |        |   |
| E | 0,713  | -0,700 | 0,943  | 0,834  | -      |        |        |        |        |   |
| F | 0,181  | -0,181 | 0,221  | 0,231  | 0,174  | -      |        |        |        |   |
| G | 0,553  | -0,535 | 0,757  | 0,606  | 0,724  | 0,252  | -      |        |        |   |
| H | 0,277  | -0,275 | 0,318  | 0,377  | 0,277  | 0,735  | 0,255  | -      |        |   |
| I | -0,492 | 0,486  | -0,580 | -0,574 | -0,636 | -0,189 | -0,583 | -0,136 | -      |   |
| J | -0,262 | 0,260  | -0,273 | -0,372 | -0,160 | 0,030  | -0,044 | -0,139 | -0,200 | - |

APÊNDICE C – Organização dos municípios de Minas Gerais por grupo, conforme resultado da partição k=11

Quadro AC1: Composição dos Grupos para partição k=11

| GRUPO | MUNICÍPIOS  |
|-------|---|
| 1     | <p>Abadia dos Dourados, Abaeté, Abre Campo, Açucena, Aimorés, Aiuruoca, Alagoa, Albertina, Além Paraíba, Alpinópolis, Alto Caparaó, Alto Rio Doce, Alvinópolis, Andradas, Andrelândia, Antônio Carlos, Antônio Dias, Araguari, Arantina, Arapuá, Araújo, Arceburgo, Arcos, Areado, Argirita, Astolfo Dutra, Baependi, Bambuí, Bandeira do Sul, Belmiro Braga, Belo Vale, Bicas, Biquinhas, Boa Esperança, Bocaina de Minas, Bom Jardim de Minas, Bom Jesus da Penha, Bom Jesus do Amparo, Bom Repouso, Bom Sucesso, Bonfinópolis de Minas, Borda da Mata, Botelhos, Brasópolis, Bueno Brandão, Buritis, Cabo Verde, Cachoeira Dourada, Cajuri, Caldas, Camanducaia, Cambuí, Cambuquira, Campanha, Campestre, Campina Verde, Campo Belo, Campo do Meio, Campo Florido, Campos Altos, Campos Gerais, Canápolis, Candeias, Capela Nova, Capelinha, Capetinga, Capim Branco, Carangola, Careagu, Carmo da Mata, Carmo de Minas, Carmo do Cajuru, Carmo do Paranaíba, Carmo do Rio Claro, Carmópolis de Minas, Carneirinho, Carrancas, Carvalhópolis, Carvalhos, Casa Grande, Cascalho Rico, Cássia, Catas Altas, Central de Minas, Centralina, Chalé, Chiador, Claraval, Cláudio, Comendador Gomes, Conceição de Ipanema, Conceição do Pará, Conceição do Rio Verde, Conceição dos Ouros, Confins, Congonhal, Conquista, Conselheiro Pena, Coqueiral, Cordislândia, Coronel Pacheco, Coronel Xavier Chaves, Córrego Novo, Cristais, Cristina, Crucilândia, Cruzeiro da Fortaleza, Cruzília, Delfim Moreira, Delfinópolis, Descoberto, Desterro de Entre Rios, Divinésia, Divinópolis, Divisa Nova, Dom Joaquim, Dona Eusébia, Dolores de Campos, Dolores de Guanhães, Dolores do Indaiá, Dolores do Turvo, Doloresópolis, Elói Mendes, Entre Rios de Minas, Ervália, Esmeraldas, Espera Feliz, Espírito Santo do Dourado, Estiva, Estrela Dalva, Estrela do Indaiá, Estrela do Sul, Eugénópolis, Ewbank da Câmara, Extrema, Fama, Faria Lemos, Florestal, Formiga, Fortaleza de Minas, Fortuna de Minas, Fronteira, Frutal, Goianá, Grupiara, Guapé, Guaraciaba, Guaranésia, Guarda-Mor, Guidoal, Gurinhatã, Heliadora, Ibertioga, Ibiá, Ibiraci, Ibitiré, Ibitiúra de Minas, Igaratinga, Iguatama, Ijaci, Inconfidentes, Indianópolis, Inhapim, Inhaúma, Ipanema, Iraí de Minas, Itamarati de Minas, Itamogi, Itamonte, Itanhandu, Itanhomi, Itapagipe, Itapeçerica, Itapeva, Itaverava, Ituiutaba, Iturama, Itutinga, Jacuí, Jeceaba, Jequitibá, Jesuânia, João Pinheiro, Lagoa da Prata, Lagoa Dourada, Lagoa Formosa, Lagoa Grande, Lajinha, Lambari, Leandro Ferreira, Leopoldina, Liberdade, Lima Duarte, Luz, Machado, Manhuaçu, Manhumirim, Mantena, Maravilhas, Mar de Espanha, Maripá de Minas, Martinho Campos, Matias Barbosa, Matozinhos, Matutina, Medeiros, Mirai, Moema, Monsenhor Paulo, Monte Alegre de Minas, Monte Carmelo, Monte Santo de Minas, Monte Sião, Montezuma, Morada Nova de Minas, Munhoz, Muriaé, Mutum, Muzambinho, Natalândia, Nepomuceno, Nova Ponte, Nova Resende, Nova Serrana, Olaria, Olhos-d'Água, Oliveira, Onça de Pitangui, Ouro Fino, Paineiras, Pains, Papagaios, Pará de Minas, Paraguaçu, Paraisópolis, Passa Quatro, Passa Tempo, Passa-Vinte, Patos de Minas, Patrocínio, Patrocínio do Muriaé, Paula Cândido, Pedra do Indaiá, Pedrinópolis, Pedro Teixeira, Pequi, Perdizes, Perdões, Piau, Piedade do Rio Grande, Pimenta, Piracema, Piranguinho, Piraúba, Pitangui, Planura, Poço Fundo, Pocrane, Pompéu, Prados, Pratápolis, Pratinha, Presidente Bernardes, Quartel Geral, Queluzito, Raul Soares, Recreio, Resende Costa, Ribeirão das Neves, Rio Paranaíba, Rio Pomba, Ritápolis, Rochedo de Minas, Rodeiro, Romaria, Rosário da Limeira, Sacramento, Salinas, Santa Bárbara do Monte Verde, Santa Juliana, Santa Luzia, Santana da Vargem, Santa Rita de Caldas, Santa Rita de Jacutinga, Santa Rita do Itueto, Santa Rita do Sapucaí, Santa Rosa da Serra, Santa Vitória, Santo Antônio do Amparo, Santo Antônio do Aventureiro, Santo Antônio do Monte, São Bento Abade, São Domingos do Prata, São Francisco de Paula, São Geraldo, São Gotardo, São João Batista do Glória, São João do Manhuaçu, São João Nepomuceno, São José da Lapa, São José da Varginha, São Miguel do Anta, São Pedro da União, São Roque de Minas, São Sebastião da Bela Vista, São Sebastião da Vargem Alegre, São Sebastião do Maranhão, São Sebastião do Paraíso, São Sebastião do Rio Verde, São Tiago, São Tomás de Aquino, São Thomé das Letras, São Vicente de Minas, Sapucaí-Mirim, Sarzedo, Senador Firmino, Senhora de Oliveira, Sericita, Serra da Saudade, Serra do Salitre, Serrania,</p> |

|   |  |
|---|--|
|   | Silveirânia, Silvianópolis, Simonésia, Soledade de Minas, Tabuleiro, Taiobeiras, Tapira, Tapiraí, Teixeiras, Tiradentes, Tiros, Tocantins, Três Marias, Três Pontas, Tupaciguara, Turmalina, Turvolândia, Ubá, Unai, Vargem Bonita, Vermelho Novo, Virgínia, Visconde do Rio Branco, Wenceslau Braz.   |
| 2 | Acaiaca, Bela Vista de Minas, Belo Oriente, Bocaiúva, Couto de Magalhães de Minas, Cural de Dentro, Divisa Alegre, Felisburgo, Gonzaga, Ibiracatu, Itinga, Janaúba, Nova Era, Novo Oriente de Minas, Patis, Prudente de Moraes, Rio Doce, Rio do Prado, Rio Piracicaba, Santa Bárbara, São Gonçalo do Rio Abaixo.  |
| 3 | Água Boa, Águas Formosas, Alfredo Vasconcelos, Almenara, Alpercata, Alvarenga, Amparo do Serra, Cachoeira de Pajeú, Angelândia, Aracitaba, Araçuaí, Araporã, Arinos, Ataléia, Baldim, Bandeira, Barão de Monte Alto, Barra Longa, Berilo, Bertópolis, Berizal, Bom Jesus do Galho, Bonito de Minas, Botumirim, Braúnas, Buenópolis, Cabeceira Grande, Campanário, Cantagalo, Capitão Andrade, Carai, Carbonita, Carlos Chagas, Carmésia, Chapada do Norte, Chapada Gaúcha, Cipotânea, Comercinho, Conceição do Mato Dentro, Congonhas do Norte, Cordisburgo, Corinto, Coroaci, Coronel Murta, Crisólita, Cristiano Otoni, Cuparaque, Datas, Desterro do Melo, Divino das Laranjeiras, Divinolândia de Minas, Divisópolis, Engenheiro Caldas, Espinosa, Felício dos Santos, São Gonçalo do Rio Preto, Felixlândia, Fernandes Tourinho, Ferros, Fervedouro, Formoso, Francisco Dumont, Franciscópolis, Frei Gaspar, Fronteira dos Vales, Fruta de Leite, Funilândia, Glaucilândia, Guanhões, Guarará, Iapu, Ibiaí, Icaraí de Minas, Igarapé, Imbé de Minas, Indaiabira, Ipaba, Itacambira, Itacarambi, Itamarandiba, Itambacuri, Itaobim, Jacinto, Jaíba, Jampruca, Januária, Jenipapo de Minas, Jequeri, Jequitaiá, Joaíma, Joanésia, Jordânia, José Gonçalves de Minas, José Raydan, Josenópolis, Nova União, Ladainha, Lagoa dos Patos, Lassance, Malacacheta, Manga, Marilac, Mário Campos, Marliéria, Mata Verde, Materlândia, Mateus Leme, Matias Cardoso, Matipó, Medina, Mendes Pimentel, Minas Novas, Minduri, Miradouro, Miravânia, Moeda, Monjolos, Montalvânia, Monte Azul, Monte Formoso, Naque, Ninheira, Nova Belém, Nova Módica, Novo Cruzeiro, Orizânia, Ouro Verde de Minas, Padre Paraíso, Pai Pedro, Palmópolis, Paraopeba, Paulistas, Pavão, Peçanha, Pedra Azul, Pedra do Anta, Pedra Dourada, Pedras de Maria da Cruz, Periquito, Pescador, Piedade de Ponte Nova, Pingo-d'Água, Pintópolis, Piranga, Pirapetinga, Ponto Chique, Ponto dos Volantes, Presidente Juscelino, Presidente Kubitschek, Raposos, Resplendor, Riacho dos Machados, Rio Casca, Rio Espera, Rio Novo, Rubelita, Rubim, Sabinópolis, Salto da Divisa, Santa Bárbara do Leste, Santa Cruz de Salinas, Santa Cruz do Escalvado, Santa Efigênia de Minas, Santa Fé de Minas, Santa Helena de Minas, Santa Maria de Itabira, Santa Maria do Salto, Santa Maria do Suaçuí, Santana do Manhuaçu, Santana do Paraíso, Santa Rita de Minas, Santo Antônio do Gramma, Santo Antônio do Jacinto, Santo Hipólito, São Brás do Suaçuí, São Félix de Minas, São Francisco, São Geraldo da Piedade, São Gonçalo do Abaeté, São João do Oriente, São João do Paraíso, São João Evangelista, São José do Jacuri, São José do Mantimento, São Pedro dos Ferros, São Pedro do Suaçuí, São Romão, São Sebastião do Anta, São Sebastião do Rio Preto, Sardoá, Setubinha, Senador Modestino Gonçalves, Senhora do Porto, Serra dos Aimorés, Serro, Simão Pereira, Sobrália, Taquaraçu de Minas, Tarumirim, Tumiritinga, Ubaí, Ubaporanga, Umburatiba, Urucânia, Urucuia, Vargem Alegre, Várzea da Palma, Varzelândia, Verdelândia, Veredinha, Vespasiano, Mathias Lobato, Virgem da Lapa, Virginópolis, Virgolândia. |
| 4 | Água Comprida, Águas Vermelhas, Bonfim, Brasilândia de Minas, Caiana, Caparaó, Carmo da Cachoeira, Conceição da Barra de Minas, Chácara, Conceição das Pedras, Conceição das Alagoas, Coromandel, Córrego Danta, Divino, Goiabeira, Guarani, Guimarães, Ibituruna, Ilicínea, Ingaí, Ipuíuna, Itabirinha, Itaguara, Itambé do Mato Dentro, Itatiaiuçu, Jaboticatubas, Jacutinga, Jequitinhonha, Joaquim Felício, Juatuba, Juruáia, Laranjal, Madre de Deus de Minas, Martins Soares, Mercês, Morro do Pilar, Oratórios, Palma, Pequeri, Perdigão, Piedade dos Gerais, Pirajuba, Piumhi, Pouso Alto, Prata, Rio Acima, Rio Preto, Santana do Deserto, Santana do Jacaré, Santo Antônio do Rio Abaixo, São Francisco de Sales, São Geraldo do Baixio, São Gonçalo do Pará, São Gonçalo do Sapucaí, São João da Mata, São Joaquim de Bicas, São José da Barra, São Sebastião do Oeste, Senador Cortes, Seritinga, Serranos, União de Minas, Veríssimo, Volta Grande.   |
| 5 | Aguanil, Alterosa, Antônio Prado de Minas, Bias Fortes, Brás Pires, Cachoeira de Minas, Camacho, Campo Azul, Canaã, Cana Verde, Capitólio, Caputira, Caranaíba, Claro dos Poções, Coimbra, Conceição da Aparecida, Consolação, Córrego do Bom Jesus, Dom Bosco, Dom Silvério, Dom Viçoso, Douradoquara, Durandé, Francisco Badaró, Gonçalves, Guaraciama, Guiricema, Ipiacu, Itueta, Itumirim, Japaraíba, Japonvar, Juramento, Lagamar, Lamim, Lontra,   |

|    |   |
|----|---|
|    | Luisburgo, Luminárias, Marmelópolis, Monte Belo, Morro da Garça, Natércia, , Nazareno, Nova Porteirinha, Novorizonte, Olímpio Noronha, Padre Carvalho, Paiva, Passabém, Pedralva, Piedade de Caratinga, Piranguçu, Porto Firme, Poté, Presidente Olegário, Alto Jequitibá, Reduto, Ressaquinha, Riachinho, Rio Manso, Rio Pardo de Minas, Santa Bárbara do Tugúrio, Santa Margarida, Santana de Pirapama, Santana do Garambéu, Santana do Riacho, Santana dos Montes, São Domingos das Dores, São Francisco do Glória, São João da Lagoa, São José do Alegre, Sem-Peixe, Senador Amaral, Senador José Bento, Serranópolis de Minas, Taparuba, Tocos do Moji, Toledo, Varjão de Minas, Vieiras.                |
| 6  | Alfenas, Araxá, Barão de Cocais, Barbacena, Barroso, Betim, Bom Despacho, Brumadinho, Cachoeira da Prata, Caeté, Capinópolis, Cataguases, Caxambu, Congonhas, Conselheiro Lafaiete, Contagem, Coronel Fabriciano, Curvelo, Diamantina, Governador Valadares, Guaxupé, Ipatinga, Itabira, Itabirito, Itajubá, Itaú de Minas, Itaúna, João Monlevade, Juiz de Fora, Lagoa Santa, Lavras, Limeira do Oeste, Mariana, Montes Claros, Nanuque, Ouro Branco, Ouro Preto, Passos, Pedro Leopoldo, Pirapora, Poços de Caldas, Ponte Nova, , , Pouso Alegre, Sabará, Santos Dumont, São João del Rei, São Lourenço, Sete Lagoas, Teófilo Otoni, Timóteo, Tombos, Três Corações, Uberaba, Uberlândia, Varginha, Viçosa. |
| 7  | Alvorada de Minas, Aricanduva, Catas Altas da Noruega, Catuji, Coluna, Cristália, Diogo de Vasconcelos, Frei Lagonegro, Itaipé, Juvenília, Rio Vermelho, São João da Ponte, São João das Missões, São João do Pacuí, São José da Safira, Serra Azul de Minas.   |
| 8  | Araçaí, Augusto de Lima, Brasília de Minas, Bugre, Buritizeiro, Caetanópolis, Capitão Enéas, Carandaí, Catuti, Cedro do Abaeté, Cônego Marinho, Coração de Jesus, Córrego Fundo, Dionísio, Dom Cavati, Engenheiro Navarro, Entre Folhas, Francisco Sá, Frei Inocência, Galiléia, Gameleiras, Gouveia, Inimutaba, Leme do Prado, Luislândia, Machacalis, Mamonas, Maria da Fé, Mato Verde, Mesquita, Mirabela, Oliveira Fortes, Paracatu, Porteirinha, Ribeirão Vermelho, Santa Cruz de Minas, Santana de Cataguases, Santo Antônio do Retiro, São José do Divino, São José do Goiabal, Senhora dos Remédios, Uruana de Minas.   |
| 9  | Araponga, Pedra Bonita.   |
| 10 | Belo Horizonte, Nova Lima.  |
| 11 | Delta, NacipRaydan.   |