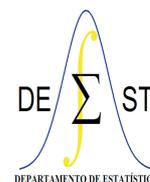




UNIVERSIDADE FEDERAL DE OURO PRETO
INSTITUTO DE CIÊNCIAS EXATAS E BIOLÓGICAS
DEPARTAMENTO DE ESTATÍSTICA



Análise da caracterização de desigualdade social e demográfica dos setores censitários de Mariana-MG

Willian de Paula da Silva

Ouro Preto-MG

Maio de 2021

Willian de Paula da Silva

Análise da caracterização de desigualdade social e demográfica dos setores censitários de Mariana-MG

Monografia de Graduação apresentada ao Departamento de Estatística do Instituto de Ciências Exatas e Biológicas da Universidade Federal de Ouro Preto como requisito parcial para a obtenção do grau de bacharel em Estatística.

Orientador

Prof. Dr. Ricardo Tavares

UNIVERSIDADE FEDERAL DE OURO PRETO – UFOP
DEPARTAMENTO DE ESTATÍSTICA – DEEST

Ouro Preto-MG

Maio de 2021

SISBIN - SISTEMA DE BIBLIOTECAS E INFORMAÇÃO

S586a Silva, Willian de Paula da .
Análise da caracterização de desigualdade social e demográfica dos setores censitários de Mariana-MG. [manuscrito] / Willian de Paula da Silva. - 2021.
63 f.: il.: gráf., tab., mapa.

Orientador: Prof. Dr. Ricardo Tavares.
Monografia (Bacharelado). Universidade Federal de Ouro Preto.
Instituto de Ciências Exatas e Biológicas. Graduação em Estatística .

1. Mariana (MG). 2. Censo. 3. IBGE. I. Tavares, Ricardo. II. Universidade Federal de Ouro Preto. III. Título.

CDU 316.34(815.1)

Bibliotecário(a) Responsável: CELINA BRASIL LUIZ - CRB6-1589



FOLHA DE APROVAÇÃO

Willian de Paula da Silva

Análise da caracterização de desigualdade social e demográfica dos setores censitários de Mariana-MG

Monografia apresentada ao Curso de Estatística da Universidade Federal de Ouro Preto como requisito parcial para obtenção do título de Bacharel em Estatística

Aprovada em 20 de maio de 2021

Membros da banca

Dr. Ricardo Tavares - Orientador - Universidade Federal de Ouro Preto
Dr. Tiago Martins Pereira - Universidade Federal de Ouro Preto
Dr. Flávio dos Reis Moura - Universidade Federal de Ouro Preto

Professor Dr. Ricardo Tavares, orientador do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 20/05/2021



Documento assinado eletronicamente por **Ricardo Tavares, PROFESSOR DE MAGISTERIO SUPERIOR**, em 26/07/2021, às 17:56, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Flavio dos Reis Moura, PROFESSOR DE MAGISTERIO SUPERIOR**, em 26/07/2021, às 19:16, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Tiago Martins Pereira, COORDENADOR(A) DE CURSO DE ESTATÍSTICA**, em 26/07/2021, às 21:00, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_organizacao_acesso_externo=0, informando o código verificador **0197617** e o código CRC **4DC1B4F0**.

Agradeço aos meus pais por todo apoio nessa minha jornada acadêmica, meus maiores e melhores orientadores na vida. Este trabalho é inteiramente dedicado a eles. Gratidão eterna.

Agradecimentos

Primeiramente agradeço a Deus pelo dom da vida e por todos os momentos em que cuidou de mim e sustentou meus passos, mesmo em momentos onde eu não tinha mais esperanças.

Agradeço aos meus pais Ana Maria e José Gonçalves por me proporcionarem chegar até aqui, sob os seus cuidados me educaram, protegeram e ensinaram os melhores e maiores valores da vida. Aqui qualquer palavra escrita não mensura a gratidão, o amor, o orgulho, o respeito e a saudade que me acompanham. Hoje dói lembrar do "é nosso sonho ver você formando", pois não os tenho para comemorar e abraçar, mas a certeza de dever cumprido me conforta.

Agradeço aos meus irmãos Ismael e Gilberto que sempre me apoiaram e não me deixaram desanimar nos momentos mais difíceis de nossas vidas. Agradeço a Deus pela renovação de vida de nossas famílias, meus sobrinhos Thales, Sarah e Carlos Eduardo que nos proporcionam tanta alegria e amor, e que mesmo de forma indireta alimentam nossas perspectivas para o futuro. Também agradeço às minhas cunhadas Fabíola e Letícia por me apoiarem em todos os momentos.

Agradeço imensamente à minha namorada Patrícia por todo apoio e carinho desde o início, sempre ao meu lado nos bons e maus momentos, que em diversas situações suas palavras me confortavam e demonstravam que eu poderia chegar até aqui. Muito obrigado!

Aos meus amigos de vida que sempre me apoiaram e tiveram participação importante nesse processo, Eduardo, Gustavo, Marina e Júnia, muito obrigado! Não menos importantes mas que a Universidade me deu o prazer de conhecer e conviver, agradeço os momentos vividos, as conversas e o compartilhamento de conhecimento, meus amigos Marcus, Pedro, Denis, Gabriella, Leylaine e Raphaella, amigos que levarei pra vida toda.

Ao meu orientador, Ricardo Tavares, agradeço a atenção, solicitude e os ensinamentos em todos os momentos ao longo do curso e deste trabalho.

Por fim agradeço os professores do departamento pelos conhecimentos compartilhados, pelos bons momentos vividos e pela dedicação à missão que lhes foi confiada.

Mas, buscai primeiro o reino de Deus, e a sua justiça, e todas estas coisas vos serão acrescentadas.

Mateus 6:33

Análise da caracterização de desigualdade social e demográfica dos setores censitários de Mariana-MG

Autor: Willian de Paula da Silva

Orientador(a): Prof. Dr. Ricardo Tavares

RESUMO

Nos últimos anos tem sido cada vez mais consensual que as questões econômicas e sociais não devam ignorar a garantia de mais cidadania para as sociedades. A cidadania não apenas como direito de votar, mas como um direito de ter educação de qualidade, saúde, informação, participação, igualdade de oportunidades etc. O Brasil, embora seja o maior e mais populoso país da América Latina, se encontra entre os dez países mais desiguais do mundo segundo o IBGE (Instituto Brasileiro de Geografia e Estatística). Nos últimos anos as desigualdades sociais e econômicas foram objetos de estudos sociais e demográficos das cidades brasileiras, tornando-se importantes para a construção e aprimoramento de políticas que estimulem a redução destas diferenças. Com base nos microdados do censo demográfico do IBGE de 2010 e a preocupação em identificar geograficamente tais diferenças, este trabalho propôs a construção de um índice de caracterização de desigualdade social e demográfica (ICDSD) para os setores censitários de Mariana. Um modelo fatorial foi ajustado a partir das variáveis econômicas e sociais, e o escore do primeiro fator foi definido como o ICDSD. Uma análise de associação espacial foi realizada nesse ICDSD e identificou-se associações entre algumas regiões em que a maioria dos setores tiveram ICDSD baixo, e seus vizinhos também. Uma região da cidade apresentou, de forma significativa, alto ICDSD, bem como seus vizinhos. O *Software R* (R Core Team (2020)) foi utilizado neste trabalho.

Palavras-chave: Caracterização, Mariana, setores censitários, IBGE.

Analysis of the characterization of social and demographic inequality in the census sectors of Mariana-MG

Author: Willian de Paula da Silva

Advisor: Prof. Dr. Ricardo Tavares

ABSTRACT

In recent years it has been increasingly agreed that economic and social issues should not ignore the guarantee of more citizenship for societies. Citizenship is not just a right to vote, but a right to quality education, health, information, participation, equal opportunities, etc. Although Brazil is the largest and most populous country in Latin America, it is among the ten most unequal countries in the world according to IBGE (Brazilian Institute of Geography and Statistics). In recent years, social and economic inequalities have been the object of social and demographic studies in Brazilian cities, becoming important for the construction and improvement of policies that encourage the reduction of these differences. Based on the microdata from the 2010 IBGE demographic census and the concern to identify such differences geographically, this work proposed the construction of an index of characterization of social and demographic inequality (ICDSD) for the census sectors of Mariana. A factorial model was adjusted based on economic and social variables, and the score for the first factor was defined as the ICDSD. An analysis of spatial association was performed in this ICDSD and associations were identified between some regions in which most sectors had low ICDSD, and their neighbors as well. A region of the city showed, significantly, high ICDSD, as well as its neighbors. Software R (R Core Team (2020)) was used in this work.

Keywords: Characterization, Mariana, census sectors, IBGE.

Lista de figuras

1	Exemplo simples de mapa para vizinhança de primeira ordem	p. 34
2	Diagrama de espalhamento de Moran	p. 36
3	Exemplo de mapa dividido por 7 áreas (polígonos)	p. 37
4	Correlograma contendo as 12 variáveis originais	p. 41
5	Distribuição dos 12 indicadores na escala logarítmica na base 10	p. 42
6	Distribuição dos 10 distritos na escala logarítmica na base 10	p. 50
7	Distribuição espacial do ICDS dos setores do Município	p. 51
8	Distribuição espacial da renda média por morador dos setores do Município	p. 52
9	Distribuição espacial do ICDS dos setores na sede	p. 53
10	Distribuição espacial da renda média por morador dos setores na sede .	p. 54
11	Distribuição espacial do ICDS dos setores na área urbana	p. 55
12	Distribuição espacial do ICDS da renda média por morador dos setores na área urbana	p. 56
13	Diagrama de espalhamento de Moran	p. 57
14	Distribuição espacial de Moran local	p. 58
15	Mapa de Moran significativo do ICDS	p. 59

Lista de tabelas

1	Base de dados (transposta) para a análise fatorial	p. 18
2	Interpretação do KMO	p. 31
3	Cargas Fatoriais baseada no tamanho da amostra	p. 32
4	Medidas descritivas das variáveis estudadas	p. 41
5	Valores de MSA para cada variável	p. 43
6	Matriz de cargas fatoriais dos dados originais	p. 44
7	Variância explicada dos dados originais	p. 44
8	Comunalidades	p. 44
9	Especificidades	p. 45
10	Valores de MSA após exclusão da variável X2	p. 45
11	Matriz de cargas fatoriais após exclusão da variável X2	p. 46
12	Variância explicada pelos fatores após exclusão da variável X2	p. 46
13	Comunalidades após exclusão da variável X2	p. 46
14	Matriz de cargas fatoriais rotacionada após a exclusão da variável X2	p. 47
15	Variância explicada pelos fatores rotacionada após exclusão da variável X2	p. 47
16	Cargas iniciais e de rotação varimax e variância explicada para os fatores retidos	p. 48

Sumário

1	Introdução	p. 12
2	Materiais e Métodos	p. 16
2.1	Análise Fatorial	p. 18
2.1.1	Estimação de Cargas Fatoriais	p. 23
2.1.2	Rotação dos Fatores	p. 26
2.1.3	Escores Fatoriais	p. 28
2.2	Análise Espacial	p. 32
2.2.1	Análise Exploratória dos dados	p. 33
2.2.2	Autocorrelação Espacial	p. 33
2.2.3	Diagrama de Espalhamento de Moran	p. 36
2.2.4	Indicadores de Autocorrelação Espacial	p. 37
3	Resultados	p. 40
3.1	Análise Descritiva	p. 40
3.2	Análise Fatorial	p. 43
3.3	Análise Espacial	p. 49
4	Considerações Finais	p. 60
	Referências	p. 61
	Apêndice A – 1	p. 63

1 Introdução

Na discussão infindável sobre a realidade das questões econômicas, sociais, não se deve ignorar a garantia de mais cidadania para as sociedades. A cidadania não apenas como direito de votar, mas que envolvam o direito de educação de qualidade, informação, saúde, igualdade de oportunidades, participação, entre outros aspectos para promoção da qualidade de vida dos cidadãos.

O Brasil se apresenta como o maior e mais populoso país da América Latina e segundo dados do IBGE (2020), se encontra entre os dez países mais desiguais do mundo. As pesquisas apontam que as desigualdades sociais e econômicas do brasileiro tem se intensificado nos últimos anos, principalmente em alguns aspectos que envolvem baixa escolaridade, mercado de trabalho informal e desocupação. O IBGE destaca diferenças maiores entre população definida como branca versus as pardas ou pretas. Esse fato, acaba impactando diretamente o comportamento demográfico de uma população, uma vez que ele está diretamente relacionado com às desigualdades sociais e econômicas da população, como citado por Wood e Carvalho (1994).

Perante o exposto, os estudos demográficos e sociais das cidades brasileiras se tornam importantes para a construção e aprimoramento de políticas que estimulem a redução destas diferenças. No caso da primeira capital de Minas Gerais, Deus (2019) destaca que em 2017, Mariana possuía a sexta maior arrecadação tributária advinda da atividade extrativa mineral. No entanto, ainda apresenta indicadores sociais e financeiros que não imprimem uma dinâmica econômica mais favorável ao município e aos seus habitantes.

Schneider e Waquil (2001) apresentaram uma caracterização dos municípios gaúchos a partir dos indicadores socioeconômicos baseados no censo agropecuário de 1995/96. Os autores utilizaram uma análise fatorial e uma análise de agrupamentos para construir classes ou grupos de municípios homogêneos a partir dos escores fatoriais. Os resultados destacaram cinco grupos de municípios com diferentes níveis de pobreza rural e de degradação dos recursos naturais. Os autores ainda rejeitaram uma possível polarização entre

o sul e o norte do estado, ou seja, a pobreza rural está tanto nos municípios do norte quanto nos do sul.

Perobelli et al. (2013) propuseram identificar os potenciais de desenvolvimento dos municípios localizados em torno do município de Juiz de Fora levando em consideração a importância do processo de planejamento regional. Foi utilizado pelos autores a aplicação da análise fatorial e Análise de agrupamentos com o objetivo de verificar a capacidade das cidades de potencializar seu desenvolvimento ou absorver novos investimentos, apresentando cinco categorias a partir dos escores fatoriais normalizados. Os resultados obtidos apontaram uma heterogeneidade muito grande dos municípios da região, como esperado Juiz de Fora é o grande centro de desenvolvimento da região e 25 cidades apresentaram como opções atraentes a absorção de investimentos.

Xerxenevsky e Fochezatto (2015) propuseram construir um índice de desenvolvimento relativo utilizando o banco de dados FEEDADOS da Fundação de Economia e Estatística (FEE) e dados do Instituto Nacional de Estudos e Pesquisa Educacionais (INEP) com a finalidade de hierarquizar os municípios do Litoral Norte do Rio Grande do Sul de acordo com os seus níveis de desenvolvimento. Para o estudo os autores escolheram a análise fatorial baseada no método de Componentes Principais afim de identificar os fatores causais de desenvolvimento dos municípios. Por decorrência apresentou-se quatro fatores distintos, destacando baixos níveis de desenvolvimento para a maioria dos municípios. Também ressaltou que os indicadores obtidos de municípios acima da média não indicam uma situação ideal, e sim que esses apresentaram uma melhor situação que os demais.

No levantamento de indicadores socioeconômicos Abreu et al. (2011) apresentaram a espacialização setorializada do IDH (Índice de Desenvolvimento Humano) da área urbana do município de Viçosa- MG a partir do Censo Demográfico de 2000. Como base para esse estudo os autores realizaram uma análise espacial por setores censitários a partir da média simples entre três subíndices (renda, longevidade e educação). A análise mostrou IDH alto na região central da área urbana, valores medianos de IDH para a periferia e um número pequeno de setores considerados como valores baixos. Os resultados obtidos pela utilização da krigagem ordinária destacou três áreas quentes declaradas de desenvolvimento e três áreas frias declaradas de subdesenvolvimento na região urbana do município.

Investigando como as transformações dos serviços de saúde se expressam para fins da gestão intergovernamental do sistema único de saúde, Albuquerque et al. (2016) buscaram identificar as diferenças e sua distribuição espacial nas 438 regiões de saúde existentes

no território brasileiro. Os autores utilizaram agrupamentos de perfis socioeconômicos a partir da análise fatorial e da análise de Componentes Principais para identificar e explicar situações observadas em 2016 com aquelas encontradas em dados do ano 2000 a partir do Censo Demográfico - IBGE, Datasus e Sistema de Contas Regionais - IBGE. Os resultados obtidos indicaram melhora dos indicadores nas regiões, persistindo a diferença entre eles. Para os autores a explicação desse cenário tem como hipótese a distribuição da população acompanhado das políticas sociais e econômicas herdadas.

Barbosa, Goncalves e Santana (2019) avaliaram o índice de Vulnerabilidade Social do Município de Natal afim de verificar a realidade humana, suas fragilidades sociais e ambientais em nível de setor censitário. Com base no recorte geográfico pelo IBGE utilizando os 895 setores censitários do município de Natal com dados obtidos do Censo Demográfico 2010, foi aplicado a técnica de análise de Componentes Principais para identificação dos fatores de acordo com suas características individuais e a estimação de cada setor foi obtida a partir do score Z. Com isso a classificação de cada setor foi utilizada para produção do mapa. Concluiu-se que existe diferenças significativas nas condições de vulnerabilidade social na periferia dos distritos norte e oeste do município, atrelados nas precárias condições de educação, poder econômico e riscos ambientais, comparado com os demais.

Farias, Figueiredo e Lima (2008) em um estudo sobre criminalidade sugeriram analisar a existência de clusters utilizando análise espacial com relação a diferentes crimes violentos em municípios de Minas Gerais em 2002, a partir de dados primários da Polícia Militar de Minas Gerais obtidos junto à Fundação João Pinheiro. Para o desenvolvimento desse trabalho foram utilizadas as técnicas de análise fatorial, análise de agrupamentos e associação espacial com o objetivo de verificar os aspectos espaciais da criminalidade. Os autores evidenciaram dois tipos de crimes de forma distinta, fatores relacionados a violência contra o patrimônio e violência contra a pessoa. Os resultados obtidos pelo autores consideraram uma política heterogênea de atuação para o combate à violência contra a pessoa e a violência contra o patrimônio tendo em vista a dependência espacial entre os municípios.

Moreira e Fochezatto (2017) utilizaram a técnica estatística do I de Moran para elaborar o Índice de Criminalidade dos Municípios Baianos identificando os fatores determinantes para explicar o progresso e o padrão da violência nesses municípios com dados disponibilizados pela Secretaria de Segurança Pública do Estado da Bahia em 2014. Aplicou-se a teoria da Econometria Espacial e a análise exploratória de dados espaciais

para compreender em que ponto uma região pode influenciar as condições das demais regiões vizinhas. A partir do estudo os autores apontaram que a atividade criminosa atinge outros municípios vizinhos causando um efeito de transbordamento da criminalidade. Por outro lado pode-se constatar que as variáveis socioeconômicas densidade demográfica, percentual de alunos matriculados por município e PIB per capita apresentaram estatísticas significativas exceto a variável relacionada com indivíduos empregados.

A aplicação de uma análise socioeconômica de uma determinada região pode ser obtida através de vários indicadores que possibilitam a identificação de agentes determinantes para o desenvolvimento econômico e social. A análise multivariada consiste em um conjunto de métodos estatísticos que estudam várias variáveis de forma simultânea com o propósito de simplificar a interpretação dessas informações, contribuindo para a otimização do desenvolvimento democrático das regiões em estudo.

Para pesquisas dessa origem o IBGE (Instituto Brasileiro de Geografia e Estatística) disponibiliza dados coletados em âmbito Nacional a partir de Censos demográficos, que corresponde ao conhecimento das condições de vida da população. É com essa perspectiva que esse estudo teve como objetivo propor e avaliar alguns indicadores por setor censitário do município mineiro de Mariana para análise do censo demográfico 2010 do IBGE.

O próximo capítulo apresenta os materiais e métodos, capítulo 3 mostra os resultados da aplicação a dados reais e por último o capítulo 4 trás as considerações finais deste trabalho.

2 Materiais e Métodos

Para o desenvolvimento deste trabalho, o banco de dados utilizado foi obtido no site do IBGE no dia 09 de Novembro de 2019 às 16 horas e 30 minutos a partir do Censo Demográfico de 2010. As informações presentes neste trabalho são constituídos por setores censitários sendo a menor unidade territorial formada por área contínua, contida em área urbana ou rural. A seguir é apresentado os passos utilizados para obtenção dos dados:

<https://www.ibge.gov.br/> > Estatísticas > Downloads > Censos > Censo Demográfico 2010 > Resultados do Universo > Agregados por Setores Censitários > Download (MG-20171016.zip e Documentação Agregados dos Setores_20180416.zip)

Para realizar a análise espacial no presente trabalho, foi utilizado um arquivo Shapefile proveniente de dados referenciados geograficamente a partir do site do IBGE do Censo de 2010. Abaixo é apresentado o passo a passo utilizado para obtenção das malhas territoriais:

<https://www.ibge.gov.br/> > Geociências > Downloads > Organização do Território > Malhas Territoriais > Malhas de Setores censitários divisões intramunicipais > Censo 2010 > Setores censitários.shp > mg > Download (mg distritos.zip; mg subdistritos.zip; mg setores censitários.zip; mg municípios.zip)

Para a análise dos dados do presente trabalho foi utilizado a linguagem de programação e ambiente para análise estatística R, (R Core Team (2020)), *Software* gratuito que possui técnicas robustas para análise de dados .

Como o Censo disponibiliza dados característicos dos domicílios com totalidade da população do país, tornou-se necessário a extração e manipulação de informações. Para formulação das variáveis deste estudo, foram utilizados ao total 8 planilhas do Censo demográfico. Os arquivos utilizados com seus respectivos códigos e nomes encontram-se no Apêndice A-1, no final deste trabalho. É de suma importância destacar que a descrição das planilhas utilizadas se encontram no arquivo PDF da Base de informações do Censo Demográfico 2010: Resultados do Universo por setor censitário, contido no arquivo zip dos dados. O procedimento para tratamento das variáveis que compõem o banco de dados

do presente estudo são dados a seguir:

- **X1:** Densidade domiciliar:

$$\text{Arquivo Basico (V003)} \times 100$$

- **X2 :** Proporção de mulheres:

$$\frac{\text{Arquivo Pessoa12 (V002)}}{\text{Arquivo Basico(V002)}} \times 100$$

- **X3:** Proporção de pessoas de 15 a 24 anos de idade:

$$\frac{\text{Arquivo Pessoa13 (V049 + \dots + V058)}}{\text{Arquivo Basico(V002)}} \times 100$$

- **X4:** Proporção de pessoas pretas ou pardas:

$$\frac{\text{Arquivo Pessoa03 : Pretas(V003) + Parda(V005)}}{\text{Arquivo Basico (V002)}} \times 100$$

- **X5:** Proporção de pessoas alfabetizadas:

$$\frac{\text{Arquivo Pessoa01 (V001)}}{\text{Arquivo Basico (V002)}} \times 100$$

- **X6:** Renda média dos responsáveis por cada morador:

$$\frac{\text{Arquivo Basico(V005)}}{\text{Arquivo Basico (V002)}} \times 100$$

- **X7:** Razão de dependência:

$$\frac{\text{Arquivo Pessoa13 V022 + (V035 + \dots + V048) + (V100 + \dots + V134)}}{\text{Arquivo Pessoa13 (V049 + \dots + V099)}} \times 100$$

- **X8:** Proporção de domicílios sem energia elétrica:

$$\frac{\text{Arquivo Domicilio01 (V046)}}{\text{Arquivo Basico (V001)}} \times 100$$

- **X9:** Proporção de domicílios sem coleta de lixo:

$$\frac{\text{Arquivo Basico (V001) - Arquivo Domicilio01 (V036)}}{\text{Arquivo Basico (V001)}} \times 100$$

- **X10:** Proporção de domicílios com até um banheiro:

$$\frac{\text{Arquivo Domicilio01 (V023 + V025)}}{\text{Arquivo Basico (V001)}} \times 100$$

- **X11:** Proporção de responsáveis com rendimento mensal de até meio salário mínimo:

$$\frac{\text{Arquivo Responsavel Renda (V001 + V010)}}{\text{Arquivo Basico (V001)}} \times 100$$

- **X12:** Proporção de pessoas responsáveis por domicílio com idade acima de 65 anos:

$$\frac{\text{Arquivo Responsavel02 (V057 + \dots + V092)}}{\text{Arquivo Basico (V001)}} \times 100$$

A seguir são apresentados na Tabela 1 os doze indicadores utilizados no estudo.

Tabela 1: Base de dados (transposta) para a análise fatorial

Variáveis	Setores Censitários			
	1	2	...	95
X1	330.69	319.94	...	346.74
X2	53.08	49.16	...	47.29
X3	18.98	18.89	...	20.11
X4	65.07	80.81	...	73.04
X5	88.01	83.48	...	80.11
X6	102.03	64.47	...	86.60
X7	37.99	41.60	...	48.12
X8	0.00	21.20	...	0.38
X9	2.07	27.22	...	9.20
X10	67.93	86.39	...	89.27
X11	28.62	13.61	...	0.00
X12	11.72	6.33	...	17.62

Para a metodologia aplicada no estudo, foi realizado o tratamento de imputação de média simples para dados faltantes, considerando a estrutura geográfica dos vizinhos de primeira ordem entre os dados para métodos que necessitam de casos completos para serem utilizados na análise. Para descrever as metodologias empregadas neste trabalho, a seção 2.1 apresenta os conceitos e pressuposições da Análise Fatorial, em seguida na seção 2.2 são apresentados as metodologias da Análise Espacial e suas ferramentas de visualização.

2.1 Análise Fatorial

Segundo Rencher (2002) a análise fatorial é um método estatístico cujo principal interesse é verificar a estrutura peculiar existente entre as variáveis de estudo em uma matriz

de dados. Em um estudo com grande número de variáveis altamente correlacionadas, tem-se como objetivo condensar a informação dessas variáveis transformando-a em um conjunto menor com uma perda mínima de informação, para este fim duas ferramentas estatísticas utilizadas são a análise fatorial e análise de componentes principais.

Para Rencher (2002) o objetivo da análise de componentes principais é explicar o máximo possível a variância total nas variáveis reduzindo os dados a um número menor de componentes. Já a análise fatorial trata-se de combinações lineares dos fatores, para explicar as covariâncias ou correlações entre as variáveis. Opcionalmente pode ser realizado as duas ferramentas nos mesmos dados.

Em uma combinação de números que contém n linhas e p colunas, dispõe-se um arranjo chamado de matriz de dimensões $n \times p$. Do ponto de vista geométrico considerando como pontos no espaço p -dimensional as coordenadas são dadas por $(x_1, x_2, \dots, x_p)^t$ denominadas por vetores. Constantemente as matrizes para análises multivariadas se resumem a n indivíduos e p variáveis.

Denominado por linhas e colunas a matriz pode ser representada da seguinte forma:

$$\mathbf{X}_{n \times p} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

$$\mathbf{X}_{n \times p} = [x_{ij}] \quad i = 1, 2, \dots, n \quad j = 1, 2, \dots, p$$

Segundo Ferreira (2008) em uma análise fatorial a verificação das relações entre as variáveis permitem identificar subconjuntos de variáveis correlacionadas entre si a partir da matriz de correlação. Aplicada para a padronização dos dados com diferentes magnitudes das variáveis a matriz de correlação é dada por:

$$\mathbf{P} = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{21} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \cdots & 1 \end{bmatrix}$$

em que ρ_{ij} é o coeficiente de correlação ij .

$$\rho_{ij} = \frac{\text{cov}(X_i, X_j)}{\sqrt{\text{var}(X_i)}\sqrt{\text{var}(X_j)}} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{jj}}} \quad (2.1)$$

para todo $i \neq j$, em que σ_{ij} é a covariância entre as variáveis X_i e X_j , já σ_{ii} é a variância da variável X_i .

Modelo Fatorial Ortogonal

Para explicação do comportamento das variáveis, segundo Ferreira (2008) é pressuposto no modelo fatorial que poucos fatores comuns são suficientes para determinar as variáveis observáveis, $m < p$, em que são estabelecidas algumas relações. Dado um vetor aleatório $\mathbf{x} = [X_1, X_2, \dots, X_p]^t$, sejam $E(\mathbf{x}) = \boldsymbol{\mu} = [\mu_1, \mu_2, \dots, \mu_p]^t$ o vetor de médias e $\boldsymbol{\Sigma}_{p \times p}$ a matriz de variâncias-covariâncias do vetor \mathbf{x} , onde:

$$\begin{aligned} X_1 - \mu_1 &= l_{11}F_1 + l_{12}F_2 + \dots + l_{1m}F_m + \epsilon_1 \\ X_2 - \mu_2 &= l_{21}F_1 + l_{22}F_2 + \dots + l_{2m}F_m + \epsilon_2 \\ &\vdots \\ X_p - \mu_p &= l_{p1}F_1 + l_{p2}F_2 + \dots + l_{pm}F_m + \epsilon_p \end{aligned}$$

na forma matricial temos,

$$\mathbf{x} - \boldsymbol{\mu} = \mathbf{L}\mathbf{F} + \boldsymbol{\epsilon} \quad (2.2)$$

em que $\mathbf{F} = [F_1, F_2, \dots, F_m]^t$ é o vetor contendo os fatores comuns não observáveis ou latentes ou constructos que buscam maximizar o poder de explicação, \mathbf{L} é igual a uma matriz $p \times m$ em que os coeficientes l_{ij} , $i = 1, 2, \dots, p$, $j = 1, 2, \dots, m$ são as cargas fatoriais ou *loadings* e os fatores específicos relacionados a cada variável são $\boldsymbol{\epsilon} = [\epsilon_1, \epsilon_2, \dots, \epsilon_p]^t$.

Para explicar o padrão de respostas de um indivíduo através do valor que ela tem a partir dos fatores comuns, as cargas fatoriais apontam a importância que cada constructo tem na definição do valor de cada variável, logo os fatores específicos informam quanto de cada variável não foi explicada pelo fatores comuns.

De acordo com Ferreira (2008) as suposições necessárias para formação do modelo fatorial sobre os vetores \mathbf{F} e $\boldsymbol{\epsilon}$, são que \mathbf{F} assume distribuição com média igual a 0 e os elementos de \mathbf{F} são independentes possuindo covariância \mathbf{I} , de natureza igual $\boldsymbol{\epsilon}$ possui média igual a 0, ou seja, $\text{Cov}(\boldsymbol{\epsilon}) = \boldsymbol{\Psi}$, logo se estabelece:

$$E(\mathbf{F}) = 0 \quad e \quad \text{Cov}(\mathbf{F}) = \mathbf{I} \quad (2.3)$$

em que os fatores tem média igual a 0, variâncias iguais a 1 e não correlacionadas.

$$E(\boldsymbol{\epsilon}) = \mathbf{0} \quad e \quad Cov(\boldsymbol{\epsilon}) = \boldsymbol{\Psi} \quad (2.4)$$

em que a média dos erros é igual a 0, sem obrigatoriedade de conter variâncias iguais e não são correlacionadas.

$$Cov(\mathbf{F}, \boldsymbol{\epsilon}) = E(\boldsymbol{\epsilon}\mathbf{F}^t) = \mathbf{0} \quad (2.5)$$

em que \mathbf{F} e $\boldsymbol{\epsilon}$ são independentes.

A partir das pressuposições definidas acima juntamente com o modelo 2.2, a estrutura de covariância de \mathbf{x} é:

$$Cov(\mathbf{x}) = \boldsymbol{\Sigma} = E(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^t \quad (2.6)$$

em que $\boldsymbol{\Sigma}$ é a matriz de variância-covariância dos dados originais, logo substitui-se pela definição do modelo (2.2), portanto

$$\begin{aligned} \boldsymbol{\Sigma} &= E \left[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^t \right] = \\ &= E \left[(\mathbf{L}\mathbf{F} + \boldsymbol{\epsilon})(\mathbf{L}\mathbf{F} + \boldsymbol{\epsilon})^t \right] = \\ &= E \left[(\mathbf{L}\mathbf{F} + \boldsymbol{\epsilon}) \left((\mathbf{L}\mathbf{F})^t + \boldsymbol{\epsilon}^t \right) \right] = \\ &= E \left[\mathbf{L}\mathbf{F}(\mathbf{L}\mathbf{F})^t + \boldsymbol{\epsilon}(\mathbf{L}\mathbf{F})^t + \mathbf{L}\mathbf{F}\boldsymbol{\epsilon}^t + \boldsymbol{\epsilon}\boldsymbol{\epsilon}^t \right] = \\ &= \mathbf{L}E(\mathbf{F}\mathbf{F}^t)\mathbf{L}^t + E(\boldsymbol{\epsilon}\mathbf{F}^t)\mathbf{L}^t + \mathbf{L}E(\mathbf{F}\boldsymbol{\epsilon}^t) + E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^t) \end{aligned} \quad (2.7)$$

Determinada as premissas destacadas em (2.3), (2.4), (2.5), é dado que:

$$\begin{aligned} \boldsymbol{\Sigma} &= cov(\mathbf{L}\mathbf{F}) + cov(\boldsymbol{\epsilon}) = \\ &= \mathbf{L}cov(\mathbf{F})\mathbf{L}^t + \boldsymbol{\Psi} = \\ &= \mathbf{L}\mathbf{I}\mathbf{L}^t + \boldsymbol{\Psi} = \\ &= \mathbf{L}\mathbf{L}^t + \boldsymbol{\Psi} \end{aligned} \quad (2.8)$$

Deste modo, a partir das suposições apresentadas anteriormente, as covariâncias entre

os elementos de \mathbf{x} e \mathbf{F} podem ser observadas,

$$\begin{aligned}
 Cov(\mathbf{x}, \mathbf{F}) &= E\left[(\mathbf{x} - \boldsymbol{\mu})\mathbf{F}^t\right] = E\left[(\mathbf{L}\mathbf{F} + \boldsymbol{\epsilon})\mathbf{F}^t\right] = \\
 &= E(\mathbf{L}\mathbf{F}\mathbf{F}^t + \boldsymbol{\epsilon}\mathbf{F}^t) = E(\mathbf{L}\mathbf{F}\mathbf{F}^t) + E(\boldsymbol{\epsilon}\mathbf{F}^t) = \\
 &= \mathbf{L}E(\mathbf{F}\mathbf{F}^t) + E(\boldsymbol{\epsilon}\mathbf{F}^t) = \\
 &= \mathbf{L}\mathbf{I} + 0 = \mathbf{L}
 \end{aligned} \tag{2.9}$$

portanto,

$$Cov(X_i, F_j) = \ell_{ij} \tag{2.10}$$

de modo que os termos ℓ_{ij} descrevem a covariância entre a variável X_i e o fator F_j .

De acordo com Ferreira (2008) a decomposição da matriz de variâncias-covariâncias $\boldsymbol{\Sigma}$ em uma soma de duas matrizes é dada pela matriz simétrica $\mathbf{L}\mathbf{L}^t$ de posto $m < p$ que contém a parte comum ao grupo de variáveis. A segunda parte dada por $\boldsymbol{\Psi}$ é uma matriz diagonal que contém a parte específica de cada variável ou especificidade.

Logo as variâncias das variáveis podem ser descritas na forma:

$$\sigma_{ii} = \ell_{i1}^2 + \ell_{i2}^2 + \cdots + \ell_{im}^2 + \psi_i = \sum_{j=1}^m \ell_{ij}^2 + \psi_i, \quad i = 1, \cdots, p \tag{2.11}$$

em que a soma do primeiro termo é denominado como a soma dos fatores comuns ou comunalidade, portanto

$$h_i^2 = \sum_{j=1}^m \ell_{ij}^2 \tag{2.12}$$

O segundo termo ψ_i é denominado como os fatores específicos ou especificidade, em que

$$\sigma_{ii} = h_i^2 + \psi_i, \quad i = 1, \cdots, p \tag{2.13}$$

A finalidade da análise de fatores é a definição dos elementos da matriz \mathbf{L} e dos elementos de $\boldsymbol{\Psi}$. Uma propriedade é demonstrada a partir de uma matriz ortogonal $\mathbf{T}_{m \times m}$ em que $\mathbf{T}\mathbf{T}^t = \mathbf{T}^t\mathbf{T} = \mathbf{I}$. Sendo assim a expressão (2.2) é reproduzida da seguinte forma:

$$\mathbf{x} - \boldsymbol{\mu} = \mathbf{L}\mathbf{F} + \boldsymbol{\epsilon} = \mathbf{L}\mathbf{T}\mathbf{T}^t\mathbf{F} + \boldsymbol{\epsilon} = \mathbf{L}^*\mathbf{F}^* + \boldsymbol{\epsilon} \quad (2.14)$$

em que: $\mathbf{L}^* = \mathbf{L}\mathbf{T}$ e $\mathbf{F}^* = \mathbf{T}^t\mathbf{F}$.

Firmado por Ferreira (2008) em que $E(\mathbf{F}^*) = \mathbf{T}^t(\mathbf{F}) = \mathbf{T}^t \mathbf{0} = \mathbf{0}$ e $\text{Cov}(\mathbf{F}^*) = \mathbf{T}^t \text{Cov}(\mathbf{F})\mathbf{T} = \mathbf{T}^t\mathbf{T} = \mathbf{I}$, é impossível diferenciar as cargas de \mathbf{L} das de \mathbf{L}^* , ou seja, \mathbf{F} e $\mathbf{F}^* = \mathbf{T}^t\mathbf{F}$ possuem as mesmas propriedades, uma vez que geram a matriz de variâncias-covariâncias $\boldsymbol{\Sigma}$ iguais, mesmo que as cargas fatoriais \mathbf{L} e \mathbf{L}^* sejam de forma geral diferentes.

Estimação do valor de m

Segundo Rencher (2002) vários critérios tem sido propostos para escolher o número de fatores (m), referindo sempre às matrizes de covariâncias ou correlações populacionais consideraremos quatro critérios para a escolha do número de fatores a serem retidos.

1 - Escolher m igual ao número de fatores necessários para a variância acumulada em uma percentagem predeterminada, por exemplo 80% da variância total.

2 - Escolher m igual ao número de autovalores maior que o autovalor médio. Para a matriz de correlação \mathbf{R} a média é 1, para a matriz de covariância \mathbf{S} é $\sum_{j=1}^p \frac{\lambda_j}{p}$.

3 - O screeplot é útil para revelar o número ótimo de fatores que podem ser extraídos antes que a quantidade de variância específica comece a dominar a estrutura de variância comum.

4 - Testar a hipótese de que m é o número correto de fatores, $H_0: \boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}^t + \boldsymbol{\Psi}$ em que \mathbf{L} é a matriz de cargas fatoriais $p \times m$.

2.1.1 Estimação de Cargas Fatoriais

Um dos passos mais importantes para a análise fatorial é a determinação do número de fatores para melhor entendimento e interpretação das informações contidas nos dados. Para a utilização da análise de fatores, $\boldsymbol{\Sigma}$ deve-se diferenciar de uma matriz diagonal, ou a matriz \mathbf{P} de correlações se diferenciar da identidade. Vale enfatizar que parâmetros do modelo de fatores em ocorrências reais são anônimas e que sua estimação deve ocorrer a partir das observações amostrais.

Dada uma amostra X_1, X_2, \dots, X_n de tamanho n em p variáveis correlacionadas, \mathbf{R} é dado como estimador de \mathbf{P} , já a matriz \mathbf{S} passa a ser um estimador de $\boldsymbol{\Sigma}$ em que se

torna possível a realização do teste de hipótese, que consiste em verificar a igualdade de Σ a uma matriz diagonal. Segundo Ferreira (2008) caso a hipótese não for rejeitada os fatores específicos apresentam papel predominante, nisso a utilização da análise de fatores não terá utilidade, caso contrário o modelo de fatores será útil.

Seguidamente será abordado métodos de estimação para os parâmetros do modelo.

Método de Componentes Principais

Conforme Ferreira (2008) na abordagem de componentes principais para fatorar Σ é utilizado a decomposição espectral, em que a matriz contém autovetores e autovalores com $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$, logo:

$$\Sigma = \mathbf{P}\mathbf{A}\mathbf{P}^t = \mathbf{P}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{P}^t = \mathbf{L}\mathbf{L}^t \quad (2.15)$$

em que $\mathbf{L} = \mathbf{P}\mathbf{\Lambda}^{\frac{1}{2}}$.

Tendo-se em vista que a matriz acima (2.15) não é útil dada a utilização do mesmo número de fatores quanto de variáveis fazendo com que não haja sequer variação alguma para os fatores específicos, é tido como a solução desse problema a utilização de um número menor m de fatores comuns do que o de p variáveis, nisso os autovetores já não são mais considerados juntamente com os $p - m$ autovalores. A partir disso a contribuição de $\lambda_{m+1}\epsilon_{m+1}\epsilon_{m+1}^t + \lambda_{m+2}\epsilon_{m+2}\epsilon_{m+2}^t + \dots + \lambda_p\epsilon_p\epsilon_p^t$ para Σ é desprezada, logo a aproximação é dada por,

$$\Sigma \cong \begin{bmatrix} \sqrt{\lambda_1}\epsilon_1 & & & \\ & \sqrt{\lambda_2}\epsilon_2 & & \\ & & \ddots & \\ & & & \sqrt{\lambda_m}\epsilon_m \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1}\epsilon_1 \\ \sqrt{\lambda_2}\epsilon_2 \\ \vdots \\ \sqrt{\lambda_m}\epsilon_m \end{bmatrix} = \hat{\mathbf{L}}\hat{\mathbf{L}}^t \quad (2.16)$$

em que não há parcela dos fatores específicos. Entretanto estimando a diagonal de $\Sigma - \mathbf{L}\mathbf{L}^t$ como definido em (2.16) contendo a parcela dos fatores específicos, a aproximação de Σ é dada por $\Sigma \cong \mathbf{L}\mathbf{L}^t + \Psi$ em que

$$\Psi = \text{Diag}(\Sigma - \mathbf{L}\mathbf{L}^t) \quad (2.17)$$

tem o intuito de evitar que uma variável com elevada variação influencie de forma errônea na determinação das cargas fatoriais. A matriz (2.17) apresentada acima, quando Σ ou \mathbf{P} são substituídos pelos seus respectivos estimadores \mathbf{S} ou \mathbf{R} é dada segundo Ferreira

(2008) como a solução dos componentes principais, onde \mathbf{S} é estabelecida em função dos pares de autovalores e autovetores. Assim, abaixo é apresentado a matriz diagonal dos estimadores das variâncias específicas, portanto:

$$\hat{\Psi} = \text{Diag}(\mathbf{S} - \hat{\mathbf{L}}\hat{\mathbf{L}}^t) \quad (2.18)$$

Estabelecida em função dos autovalores e autovetores a análise de fatores por componentes principais é obtida a partir da covariância amostral \mathbf{S} , nisso disposto do número de fatores $m < p$ segundo Ferreira (2008), a matriz de cargas fatoriais estimadas é dada por:

$$\hat{\mathbf{L}} = \left[\sqrt{\hat{\lambda}_1} \hat{\epsilon}_1 \sqrt{\hat{\lambda}_2} \hat{\epsilon}_2 \cdots \sqrt{\hat{\lambda}_m} \hat{\epsilon}_m \right] = \hat{\mathbf{P}}_1 \hat{\Lambda}_1^{1/2} \quad (2.19)$$

No modelo de fatores $\hat{\mathbf{P}} = \mathbf{L}\mathbf{L}^t + \Psi$, os m fatores comuns restabelecerão os elementos que estão fora da diagonal principal de ρ , definida como uma aproximação modificada na solução fatorial principal, sendo válido tanto para \mathbf{R} quanto para \mathbf{S} . Para definição da matriz (\mathbf{R}_r) denominada como matriz de correlação amostral reduzida, é suposto que os fatores específicos não façam parte de ρ e que as estimativas iniciais da variância específica tenham sido encontradas. Logo pode ser definida assim:

$$\mathbf{R}_r = \mathbf{R} - \Psi^* \quad (2.20)$$

Segundo Ferreira (2008) os estimadores aplicados no método fatorial principal são os pares de autovalor-autovetor obtidos pela matriz reduzida \mathbf{R}_r , logo as comunalidades são dadas por:

$$\hat{\mathbf{h}}_i^{*2} = \sum_{j=1}^m \ell_{ij}^{*2} \quad (2.21)$$

Método da Máxima Verossimilhança

Para utilização desse método é necessário que os fatores comuns \mathbf{F} e os fatores específicos ϵ possuam distribuição normal. Dado vetor aleatório $\mathbf{x} \sim \mathbf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, em que $\mathbf{x} - \boldsymbol{\mu} = \mathbf{L}\mathbf{F} + \epsilon$ é igualmente normalmente distribuído, deste modo a função de veros-

similhança é dada por:

$$\begin{aligned} \ell(\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \prod_{i=1}^n f_i(x_i) = \frac{1}{(2\pi)^{\frac{np}{2}} |\boldsymbol{\Sigma}|^{\frac{n}{2}}} \exp \left\{ -\frac{1}{2} \sum_{j=1}^n (x_j - \mu_j)^t \boldsymbol{\Sigma}^{-1} (x_j - \mu_j) \right\} \\ &= \frac{1}{(2\pi)^{\frac{np}{2}} |\mathbf{L}\mathbf{L}^t + \boldsymbol{\Psi}|^{\frac{n}{2}}} \exp \left\{ -\frac{1}{2} \sum_{j=1}^n (x_j - \mu_j)^t (\mathbf{L}\mathbf{L}^t + \boldsymbol{\Psi})^{-1} (x_j - \mu_j) \right\} \end{aligned} \quad (2.22)$$

De acordo com Ferreira (2008) a função de verossimilhança é dependente dos parâmetros \mathbf{L} e $\boldsymbol{\Psi}$ mediante a matriz de covariâncias $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}^t + \boldsymbol{\Psi}$. Para que o modelo esteja bem estabelecido, deseja-se aplicar uma condicionante em que $\boldsymbol{\Delta}$ é uma matriz diagonal dada por:

$$\boldsymbol{\Delta} = \mathbf{L}^t \boldsymbol{\Psi}^{-1} \mathbf{L} \quad (2.23)$$

A maximização da função de verossimilhança (2.22) é estabelecida a partir dos estimadores das matrizes \mathbf{L} e $\boldsymbol{\Psi}$, em que sua aplicação é dada com possibilidade somente em métodos numéricos. Mais detalhes técnicos podem ser observados em (FERREIRA, 2008).

2.1.2 Rotação dos Fatores

No processo de extração, costumeiramente os fatores gerados por vezes não se encontram em condições espontaneamente interpretáveis. Segundo Rencher (2002) na rotação das cargas fatoriais que podem ser representadas como ponto entre eixos, a capacidade essencial para girar as cargas para um novo quadro de referência sem afetar quaisquer propriedades ou suposições é muito útil na interpretação dos fatores. A matriz de cargas fatoriais \mathbf{L} através da multiplicação por uma matriz ortogonal \mathbf{T} conduz a uma fatoração válida. A rotação tem como objetivo posicionar os eixos mais próximos possível dos pontos, associando os grupos de variáveis a um fator com interpretação mais objetiva.

São considerados dois tipos básicos de rotação, sendo elas ortogonal e oblíqua. Na rotação ortogonal os eixos são preservados e mantêm-se perpendiculares, onde suas comunalidades não são alteradas. A rotação oblíqua difere na forma em que a estrutura dos eixos não permanecem perpendiculares, ou seja, estão dispostos a se aproximarem do grupo de pontos com a finalidade de gerar interpretações mais claras. A rotação ortogonal é comumente utilizado, nesse caso, o material abaixo será voltado à métodos utilizados propriamente na rotação ortogonal. Mais detalhes técnicos podem ser encontrados em (RENCHE, 2002).

Dada uma matriz $p \times m$ de cargas fatoriais em que é sujeita a uma rotação pela matriz \mathbf{T} em que $\mathbf{L}^* = \mathbf{LT}$. A ortogonalidade de \mathbf{T} estabelece que as comunalidades não sejam alteradas, logo:

$$\sum_{j=1}^m \ell_{ij}^2 \Leftrightarrow h_i^2 \quad (2.24)$$

Neuhaus e Wrigley (1954) sugeriram maximizar a variância do quadrado das $p \times m$ cargas fatoriais em que a variância do quadrado das cargas fatoriais é dada por:

$$V = \sum_{i=1}^p \sum_{j=1}^m \ell_{ij}^4 - \frac{1}{pm} \left(\sum_{i=1}^p \sum_{j=1}^m \ell_{ij}^2 \right)^2 \quad (2.25)$$

Logo a soma das cargas fatoriais de quarta potência corresponde à maximização de V , critério esse denominado como quartimax.

O método varimax proposto por Kaiser (1958) representa a matriz dos pesos fatoriais rotacionada, de uma estrutura simples de soma das variâncias das cargas fatoriais quadráticas dentro de cada coluna da matriz \mathbf{L} . Como se considera que os pesos sejam ele de respostas com grandes e pequenas comunalidades, o critério alternativo de melhor acurácia é expresso por:

$$v = \frac{1}{p^2} \sum_{j=1}^m \left[p \sum_{i=1}^p x_{ij}^4 - \left(\sum_{i=1}^p x_{ij}^2 \right)^2 \right] \quad (2.26)$$

em que a j -ésima carga fatorial da i -ésima variável resposta dividida pela raiz quadrada de sua comunalidade é dada por:

$$x_{ij} = \frac{\ell_{ij}}{\sqrt{\sum_{j=1}^m \ell_{ij}^2}} \quad (2.27)$$

Dado par de fatores r e s com cargas normalizadas x_{ir} e x_{is} que difere da expressão 2.26 onde esses fatores apresentam o ângulo ϕ em que deve satisfazer a seguinte relação:

$$tg(\phi) = \frac{2 \left[2p \sum_{i=1}^p (x_{ir}^2 - x_{is}^2) x_{ir} x_{is} - \sum_{i=1}^p (x_{ir}^2 - x_{is}^2) \left(2 \sum_{i=1}^p x_{ir} x_{is} \right) \right]}{p \sum_{i=1}^p [(x_{ir}^2 - x_{is}^2) - (2x_{ir} x_{is})^2] - \left\{ \left[\sum_{i=1}^p (x_{ir}^2 - x_{is}^2) \right]^2 - 2 \left(p \sum_{i=1}^p x_{ir} x_{is} \right)^2 \right\}} \quad (2.28)$$

De acordo com Ferreira (2008) alguns processos para a solução iterativa da rotação são essenciais, em que os ângulos sejam menores que um critério de convergência especificado ϵ no meio de um ciclo. Logo esses processos são dados pela rotação do primeiro e segundo fator como ângulo ϕ , nisso o primeiro fator é rotacionado juntamente com o terceiro fator original, e nesse sentido deve se prosseguir até que os pares de rotações

$$\frac{m(m-1)}{2} \quad (2.29)$$

tenham ocorrido.

Na elaboração de intervalos de confiança e teste de hipóteses em grandes amostras, proveito oriundo de se utilizar a inferência estatística garantindo sua consistência e normalidade assintótica, a avaliação do número de fatores tem-se pelas seguintes hipóteses:

$$\begin{aligned} H_0 : \Sigma &= LL^t + \Psi \\ H_a : \Sigma &\neq LL^t + \Psi \end{aligned}$$

Sob a hipótese nula H_0 , a estatística TRV é dada por:

$$TRV = -2 \ln \left(\frac{|\hat{\Sigma}|}{|\hat{S}_n|} \right)$$

A estatística TRV acima segue uma distribuição qui-quadrado onde $v = \frac{1}{2}\{(p-m)^2 - p - m\}$ graus de liberdade. Dado que o resultado do teste traria a mesma solução de fatores da matriz de correlação \mathbf{R} , para o teste da falta de ajuste é preciso que os graus de liberdade sejam positivos, logo:

$$m < \frac{1}{2} \left(2p + 1 - \sqrt{8p + 1} \right) \quad (2.30)$$

No teste de razão de verossimilhança H_0 é geralmente rejeitada quando m for pequeno em comparação a p , em que o modelo apresenta um número maior de fatores comuns. Caso contrário em que m seja grande em comparação a p , a hipótese não é rejeitada, nesse caso acontece devido a Σ estar próximo de \mathbf{S}_n .

2.1.3 Escores Fatoriais

Segundo Ferreira (2008) os escores fatoriais são valores estimados dos fatores comuns em que suas quantidades correspondentes de cada fator para cada indivíduo podem ser

utilizadas para estruturação de índices ou diagnósticos futuros em análises posteriores. Definido como valores atribuídos às variáveis hipotéticas os escores fatoriais não podem ser estimados costumeiramente no sentido estatístico.

A seguir será apresentado dois métodos para previsão de escores fatoriais.

Método dos Mínimos Quadrados Ponderados

Suponhamos que os parâmetros do modelo fatorial $\mathbf{x} - \boldsymbol{\mu} = \mathbf{LF} + \boldsymbol{\epsilon}$ sejam conhecidos. A partir da $\text{Var}(\epsilon_i) = \psi_i$, a soma dos quadrados de resíduos é dada por:

$$\sum_{i=1}^p \frac{\epsilon_i^2}{\psi_i} = \boldsymbol{\epsilon}^t \boldsymbol{\Psi}^{-1} \boldsymbol{\epsilon} = (\mathbf{x} - \boldsymbol{\mu} - \mathbf{LF})^t \boldsymbol{\Psi}^{-1} (\mathbf{x} - \boldsymbol{\mu} - \mathbf{LF}) \quad (2.31)$$

Para minimizar f da equação 2.31, o estimador $\hat{\mathbf{f}}$ é dado por:

$$\hat{\mathbf{f}} = (\mathbf{L}^t \boldsymbol{\Psi}^{-1} \mathbf{L})^{-1} \mathbf{L}^t \boldsymbol{\Psi}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \quad (2.32)$$

Logo, os estimadores $\hat{\mathbf{L}}^t$, $\hat{\boldsymbol{\Psi}}$ e $\hat{\boldsymbol{\mu}}$ para o j -ésimo fator é:

$$\hat{\mathbf{f}}_j = (\mathbf{L}^t \boldsymbol{\Psi}^{-1} \mathbf{L})^{-1} \mathbf{L}^t \boldsymbol{\Psi}^{-1} (\mathbf{x}_j - \hat{\mathbf{x}}) \quad (2.33)$$

\mathbf{x}_j corresponde ao vetor das p observações equivalentes ao j -ésimo indivíduo, $j = 1, 2, \dots, n$.

Método de Regressão

Segundo Ferreira (2008) dado o modelo fatorial original 2.2, suponhamos que os vetores \mathbf{X} , \mathbf{F} e $\boldsymbol{\epsilon}$ possuam distribuição normal multivariada. Nisso a combinação linear $\mathbf{x} - \boldsymbol{\mu} = \mathbf{LF} + \boldsymbol{\epsilon}$ tem distribuição $N_p(\mathbf{0}, \mathbf{LL}^t + \boldsymbol{\Psi})$, distribuição conjunta $\mathbf{x} - \boldsymbol{\mu}$ e \mathbf{F} é normal $(p + m)$ -variada, com média zero e matriz de variâncias-covariâncias dada por:

$$\boldsymbol{\Sigma}^* = \begin{bmatrix} \mathbf{LL}^t + \boldsymbol{\Psi} & \mathbf{L} \\ \mathbf{L}^t & \mathbf{I} \end{bmatrix}$$

As equações a seguir referem-se à distribuição condicional de $\mathbf{F}|\mathbf{x}$, normal multivariada com média e variância dada por:

$$\begin{aligned} E(\mathbf{F}|\mathbf{x}) &= \mathbf{L}^t \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) = \mathbf{L}^t (\mathbf{LL}^t + \boldsymbol{\Psi})^{-1} (\mathbf{x} - \boldsymbol{\mu}) \\ \text{Cov}(\mathbf{F}|\mathbf{x}) &= \mathbf{I} - \mathbf{L}^t \boldsymbol{\Sigma}^{-1} \mathbf{L} = \mathbf{I} - \mathbf{L}^t (\mathbf{LL}^t + \boldsymbol{\Psi})^{-1} \mathbf{L} \end{aligned} \quad (2.34)$$

Os coeficientes de uma regressão multivariada com as variáveis originais, de acordo com Ferreira (2008) são dadas por $\mathbf{L}^t(\mathbf{L}\mathbf{L}^t + \mathbf{\Psi})^{-1}$ que produzem os escores fatoriais. Deste modo, o vetor \mathbf{x}_j , tomando os estimadores \mathbf{L} e $\mathbf{\Psi}$ pelo método da máxima verossimilhança são:

$$\hat{\mathbf{f}}_j = \hat{\mathbf{L}}^t \hat{\mathbf{\Sigma}}^{-1}(\mathbf{x}_j - \bar{\mathbf{x}}) = \hat{\mathbf{L}}^t(\hat{\mathbf{L}}\hat{\mathbf{L}}^t + \hat{\mathbf{\Psi}})^{-1}(\mathbf{x}_j - \bar{\mathbf{x}}) \quad (2.35)$$

Portanto, quando a matriz diagonal $\hat{\mathbf{\Sigma}}^{-1}$ estimada pela máxima verossimilhança ter seu valor próximo de zero, conclui-se que os estimadores anteriores produzirão os mesmos escores.

Ambas as abordagens de escore fatorial tem dois elementos em comum.

1 -Eles tratam as cargas fatoriais estimadas $\hat{\ell}_{ij}$ e as variâncias específicas $\hat{\Psi}_i$ como se fossem os valores verdadeiros.

2 - Não se substitue os ℓ 's forem rotacionados ou não-rotacionados, embora a gente tenha a achar que o rotacionado tenha sido mais rigoroso

Passos para a realização de uma análise fatorial

Para a definição do método de extração dos fatores é aplicado o teste de normalidade multivariado, que consiste na distribuição mais adequada segundo Ferreira (2008) para os dados e resíduos dos modelos. O teste baseia-se na estatística W de Shapiro-Wilk para dados multivariados.

As hipóteses de uma distribuição normal multivariada são dadas a seguir:

H_0 : A amostra provem de uma distribuição normal multivariada;

H_a : A amostra não provem de uma distribuição normal multivariada;

em que a rejeição da hipótese nula implica que só poderemos estimar o modelo fatorial utilizando o método de componentes principais ou o método dos fatores principais, uma vez que o método da máxima verossimilhança supõe normalidade multivariada dos dados.

O próximo passo segundo Fávero et al. (2009) é verificar a adequabilidade da amostra pelo critério de Kaiser-Meyer-Olkin (KMO), que consiste em avaliar o grau de adequação da amostra para aplicação da análise fatorial. Variando entre 0 e 1, quanto mais próximo de 1 o valor do KMO estiver, mais adequada é a amostra. De forma análoga ao KMO, o MSA é o cálculo de cada variável separadamente, tendo como objetivo analisar se uma variável pode ser explicada pelas demais. Abaixo é dada a Tabela 2 para interpretação do

valor do KMO:

Tabela 2: Interpretação do KMO

KMO	Interpretação
0,90 - 1,00	Excelente
0,80 - 0,90	Ótimo
0,70 - 0,80	Bom
0,60 - 0,70	Regular
0,50 - 0,60	Ruim
0,00 - 0,50	Inadequado

Para validar a aplicação da análise fatorial em um estudo, o teste de esfericidade de Bartlett requer que as variáveis X_1, \dots, X_p sejam correlacionadas. Caso as variáveis não sejam correlacionadas duas a duas, então a matriz de correlação é a matriz identidade. Logo as hipóteses do teste são:

$$H_0 : R = I \quad \text{vs} \quad H_a : R \neq I$$

em que a rejeição de H_0 indica que existe relação suficiente entre os indicadores para aplicação da análise fatorial.

Para se obter o número de fatores o critério de retenção de fatores de Kaiser-Guttman segundo Laros (2012) considera a capacidade de um fator explicar pelo menos a quantidade de variância que é explicada por uma única variável, onde o autovalor é igual a 1. Portanto, a soma dos autovalores correspondem ao número de variáveis analisadas, em que apenas autovalores > 1 são retidos.

Para a extração dos fatores segundo Fávero et al. (2009), a análise fatorial busca maximizar a variância total explicada dos dados pelo fator formado. Pela complexidade de interpretação muita das vezes na análise multivariada pelos fatores não-rotacionados, a utilização do método de rotação pode auxiliar como outra opção para interpretação desses fatores.

Para o critério de significância das cargas fatoriais segundo Hair et al. (2009), a significância das cargas fatoriais é baseada no tamanho da amostra de acordo com a Tabela 3 a seguir:

Tabela 3: Cargas Fatoriais baseada no tamanho da amostra

Carga Fatorial	Tamanho da Amostra
0,30	350
0,35	250
0,40	200
0,45	150
0,50	120
0,55	100
0,60	85
0,65	70
0,70	60
0,75	50

Segundo Ferreira (2008) as comunalidades são a proporção da variabilidade comum de cada variável que é explicada pelos fatores variando entre 0 e 1, as variáveis que contém comunalidades mais próximas de 1 terão maior poder de explicação pelo fator, de forma análoga, variáveis com comunalidades próximas de 0 não serão significativas para comporem determinado fator. Da mesma forma a porcentagem de variabilidade não explicada pelos fatores é dada pela especificidade que consiste na parcela da variância não relacionada aos fatores comuns.

Proposto por Kaiser (1958) a rotação varimax é um método de rotação ortogonal que maximiza a soma de variâncias de cargas exigidas da matriz fatorial. Portanto tenta agregar um número menor de variáveis sobre cada fator para gerar fatores mais interpretáveis.

2.2 Análise Espacial

Nessa seção descreveremos alguns dos métodos de análise de dados em que a localização geográfica é usada explicitamente na análise, que formam a estatística espacial.

Os tipos de dados espaciais podem ser classificados em:

- Superfície aleatória: Quando os dados são observados apenas em alguns locais, e a partir de interpolação a informação para toda a superfície é estimada.
- Processos pontuais: Quando os dados são a própria posição exata dos pontos ou eventos que estão sendo avaliados.

- **Área:** Quando a localização está associada a áreas delimitadas por polígonos e os eventos são agregados por municípios, bairros ou setores censitários.
- **Interação espacial:** Quando os dados são informações envolvendo origem e destino, e não apenas um local específico.

Este trabalho ficou restrito aos dados de área, em que não se dispõe da localização exata dos eventos, mas de um valor que representa a área. Esse tipo de estudo é também conhecido como estudos ecológicos, cuja principal limitação é que a informação ou o valor não pode ser considerada uniforme em toda esta área.

2.2.1 Análise Exploratória dos dados

Segundo Gilberto et al. (2004) na análise exploratória uma visualização mais palpável ao fenômeno em estudo é dada a partir de valores extremos nos mapas. Os SIGs dispõem de quatro métodos de corte de variável sendo eles em quebras naturais, intervalos iguais, percentis e desvios padrão. Para as quebras naturais: Os valores de categorização da variável não seguem uma regra predefinida, e sim o julgamento próprio de quem está propondo esta visualização; intervalos iguais: valores máximo e mínimo são divididos pelo número de classes, decorrendo de uma distribuição concentrada tendo de um lado o resultado de alocação a uma ou duas cores; para os percentis tem-se por definição de classes a necessidade de atribuição dos polígonos em quantidades iguais pelas cores, diante dessa hipótese diferenças significativas em valores extremos podem dificultar o reconhecimento de áreas críticas. Logo os desvios padrão com suposição de distribuição normal tem a apresentação de cores diferentes na distribuição da variável para valores acima e abaixo da média.

2.2.2 Autocorrelação Espacial

A autocorrelação espacial é caracterizada por uma correlação em uma variável entre locais próximos no espaço.

Matriz de Proximidade Espacial

A estimação da variabilidade espacial de dados de área é desenvolvida pela matriz de proximidade espacial que é conhecida como a matriz de vizinhança. Na utilização em cálculos de indicadores é de grande proveito a normalização de suas linhas em que a soma dos pesos de cada linha seja igual a 1.

Dado um determinado conjunto de n áreas A_1, \dots, A_n constrói-se a matriz $W^{(1)}$ ($n \times n$), onde cada um dos elementos W_{ij} representa uma medida de proximidade entre A_i e A_j , calculada a seguir pelos seguintes critérios:

- $W_{ij} = 1$, se o centróide de A_i está a uma determinada distância de A_j ; caso contrário $W_{ij} = 0$.
- $W_{ij} = 1$, se A_i compartilha um lado comum com A_j , caso contrário $W_{ij} = 0$.
- $W_{ij} = \frac{l_{ij}}{l_i}$, onde l_{ij} é o comprimento da fronteira entre A_i e A_j e l_i é o perímetro de A_i .

A matriz de proximidade espacial difunde com vizinhos de maior ordem podendo construir matrizes $W^{(2)}, \dots, W^{(n)}$. A Figura 1 apresenta de forma simples um exemplo de mapa dividido por polígonos com critério de vizinhança de primeira ordem.

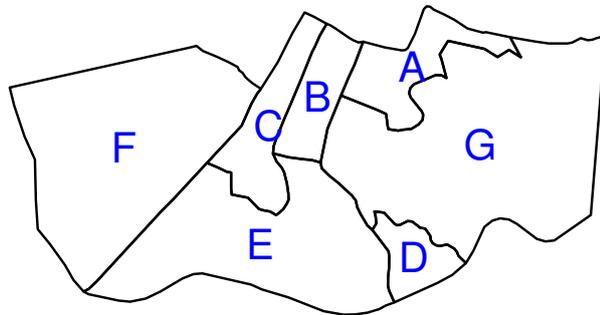


Figura 1: Exemplo simples de mapa para vizinhança de primeira ordem

A matriz de vizinhança tem como apresentação dos dados as subdivisões territoriais, em que as relações de adjacência são representadas pelo valor 1 onde as regiões delimitadas pelas fronteiras são vizinhas de primeira ordem, e 0 para as regiões não consideradas como vizinhas de primeira ordem.

Matriz de proximidade de primeira ordem

$$W^{(1)} = \begin{matrix} & \begin{matrix} A & B & C & D & E & F & G \end{matrix} \\ \begin{matrix} A \\ B \\ C \\ D \\ E \\ F \\ G \end{matrix} & \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 0 \end{pmatrix} \end{matrix}$$

A matriz abaixo apresenta um modelo claro para o critério de proximidade espacial em que a soma dos pesos espaciais de cada linha é igual um.

Matriz de pesos espaciais de primeira ordem

$$W^{(1)} = \begin{matrix} & \begin{matrix} A & B & C & D & E & F & G \end{matrix} \\ \begin{matrix} A \\ B \\ C \\ D \\ E \\ F \\ G \end{matrix} & \begin{pmatrix} 0 & 0,5 & 0 & 0 & 0 & 0 & 0,5 \\ 0,25 & 0 & 0,25 & 0 & 0,25 & 0 & 0,25 \\ 0 & 0,33 & 0 & 0 & 0,33 & 0,33 & 0 \\ 0 & 0 & 0 & 0 & 0,5 & 0 & 0,5 \\ 0 & 0,2 & 0,2 & 0,2 & 0 & 0,2 & 0,2 \\ 0 & 0 & 0,5 & 0 & 0,5 & 0 & 0 \\ 0,25 & 0,25 & 0 & 0,25 & 0,25 & 0 & 0 \end{pmatrix} \end{matrix}$$

Sob o critério de parcimônia nomeia-se os coeficientes da matriz de primeira ordem por W_{ij} , já as matrizes de ordem K por $W_{ij}^{(k)}$. Se tratando de vizinhos de ordem maior para o critério de relacionamento de vizinhos de ordem $W^{(2)}$, adiciona-se os polígonos vizinhos de segunda ordem aos de primeira ordem, ressaltando a permanência de grandeza binária. De acordo com O'Sullivan e Unwin (2010) algumas ligações de proximidade contendo o valor 1 podem apresentar relações diferentes, umas mais fortes que outras variando de 0 para interação fraca e 1 para interação forte, possibilidade presente mediante a ponderação de pesos entre dois polígonos.

Média Móvel

Compreendido por Gilberto et al. (2004) para explorar a variação da tendência espacial dos dados, a média móvel espacial na área i é a média dos valores da variável (em estudo) nos vizinhos da área i . Denominada como $\hat{\mu}_i$, é associada ao atributo Z_i relativo à i -ésima área, fundamentada em cálculo nos elementos W_{ij} da matriz de proximidade espacial $W^{(1)}$

obtendo a média dos vizinhos.

$$\hat{\mu}_i = \sum_{j=1}^n W_{ij} Z_j \quad (2.36)$$

2.2.3 Diagrama de Espalhamento de Moran

Para a visualização da autocorrelação existente é tomado o diagrama de espalhamento de Moran, ferramenta que proporciona a possibilidade de visualização do comportamento da variabilidade espacial relacionando os valores normalizados do atributo com a média da vizinhança. O diagrama é construído a partir dos valores normalizados da variável que consiste na subtração dos valores originais ao valor da média, divididos pelo desvio padrão.

O diagrama de espalhamento de Moran relaciona e divide em quatro quadrantes z (valores normalizados) e wz (média dos vizinhos) de:

- Q1 (valores positivos, médias positivas) e Q2 (valores negativos, médias negativas):
Pontos de associação positiva com vizinhos semelhantes.
- Q3 (valores positivos, médias negativas) e Q4 (valores negativos, médias positivas):
Pontos de associação negativa, com vizinhos desiguais.

A Figura 2 apresenta o gráfico de dispersão com base no diagrama de espalhamento de Moran:

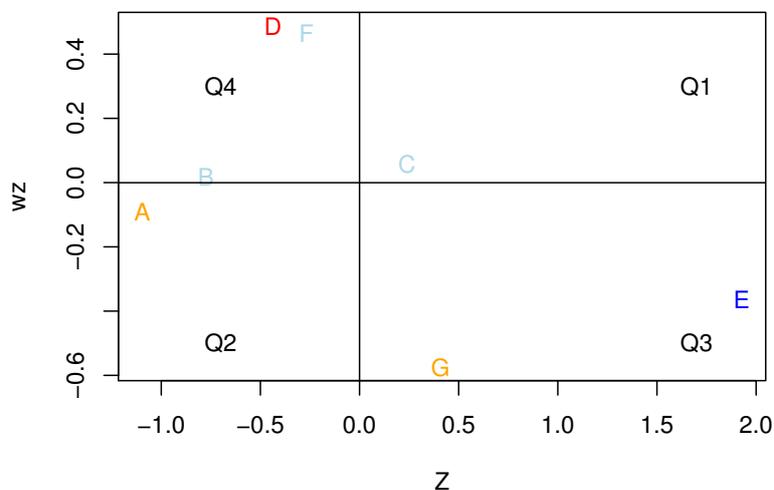


Figura 2: Diagrama de espalhamento de Moran

Cada ponto do diagrama apresentado é uma área conforme destacado na Figura 3, em que pode-se observar a distribuição espacial dos dados.

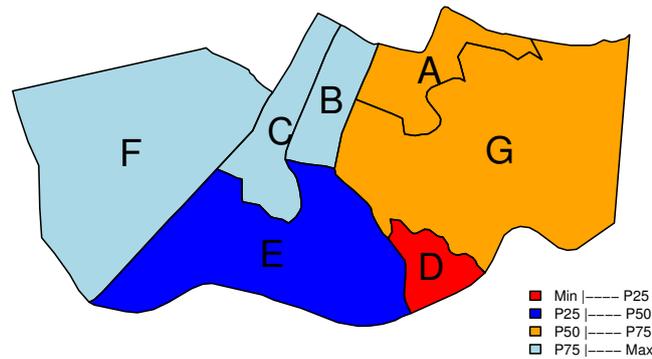


Figura 3: Exemplo de mapa dividido por 7 áreas (polígonos)

Dentre as sete áreas delimitadas pelos polígonos acima, o diagrama de espalhamento de Moran aponta duas regiões presentes nos quadrantes 1 e 2 respectivamente, que apresentam associação espacial em relação aos homicídios. Pode-se observar também que a maioria das áreas presentes nos quadrantes 3 e 4 não apresentam o mesmo processo de associação espacial. E na relação é avaliada pelo indicadores de Moran.

2.2.4 Indicadores de Autocorrelação Espacial

Indicadores Globais

Baseado nos conceitos aplicados por Gilberto et al. (2004) a caracterização da dependência espacial é oriunda de observações de uma determinada região em que se pondera os valores correlacionados no espaço. Nessas condições denomina-se a autocorrelação espacial como funções aplicadas para estimar quanto determinada região é dependente dos valores desta mesma variável nas localizações vizinhas.

Nesse contexto, é dada abaixo a expressão da autocorrelação do índice global de Moran I , considerando apenas o primeiro vizinho:

$$I = \frac{\sum_{i=1}^n \sum_{j=1}^n W_{ij} (Z_i - \bar{Z})(Z_j - \bar{Z})}{\sum_{i=1}^n (Z_i - \bar{Z})^2} \quad (2.37)$$

em que n é o número de áreas, Z_i o valor do atributo considerado na área i , \bar{z} é o valor médio do atributo na região e W_{ij} os elementos da matriz normalizada de proximidade espacial. Nesse contexto observa-se que o cálculo será avaliado apenas para vizinhos de primeira ordem. Posteriormente para cálculo de matrizes de proximidade de maior ordem é utilizado a função de autocorrelação expressa abaixo:

$$I^{(k)} = \frac{n \sum_{i=1}^n \sum_{j=1}^n W_{ij}^{(k)} (Z_i - \bar{Z})(Z_j - \bar{Z})}{\sum_{i=1}^N (Z_i - \bar{Z})^2} \quad (2.38)$$

O Índice de Moran dispõe-se de um teste em que a hipótese nula é de independência espacial com valor igual a zero, para valores entre 0 e +1 é considerado correlação positiva e valores entre 0 e -1 correlação negativa(inversa). Nesse âmbito para validação dos resultados é necessário verificar a correlação espacial, em que sua juricidade é associada a uma distribuição estatística, sendo mais comum relacionar a estatística de teste à distribuição normal. Outra alternativa em relação à distribuição é o utilizar um teste de pseudo-significância, em que são elaboradas diferentes permutações de atributos agregados às regiões, produzindo um novo ajuste espacial.

Para o cálculo do índice de Moran a estacionariedade dos dados de primeira e segunda ordem é ponto importante a se observar, logo o índice perde eficácia para dados não-estacionários onde a função de autocorrelação segue em redução após transpassar o espaço onde há interferências locais.

Indicadores Locais

Aplicado para tratar de grande número de áreas, o índice local de Moran desenvolvido por Anselin (1994) tem como objetivo testar a autocorrelação local e detectar objetos espaciais com influência no indicador de global de Moran.

A interdependência espacial entre todos os polígonos em estudo avaliada pelo Moran global difere do Moran local, em que o Moran local informa a interdependência entre um determinado polígono e uma certa vizinhança estabelecida.

Os indicadores locais analisam diferentes processos de autocorrelação espacial em regiões onde é mais externada determinadas associações criando possibilidades para examinar padrões com maiores detalhes. O índice local de Moran é expresso por:

$$I_i = \frac{Z_i \sum_{j=1}^n W_{ij} Z_j}{\sum_{j=1}^n Z_j^2} \quad (2.39)$$

em que W_{ij} é o valor n da matriz de vizinhança para a região i com a região j , e Z_i e Z_j são os desvios em relação à média.

Na identificação de desvios de uma distribuição espacial aleatória, o teste Ipop tem sensibilidade ao alto número de eventos em uma pequena população de uma única localidade. Para o teste Ipop a hipótese nula assume a variação geográfica do número de casos iguais à variação geográfica do tamanho da população, tendo em vista a particularidade de áreas não homogêneas da região em estudo.

Com base para este cálculo Gilberto et al. (2004) apresenta outra aplicação que difere o teste I de Moran, o teste C de Geary, que consiste na utilização da diferença entre os pares. De forma semelhante ao teste I de Moran a estatística C de Geary apresenta a autocorrelação espacial ou não, com magnitudes diferentes para transmitir o quanto de interação espacial há nas regiões. Maiores detalhes técnicos podem ser encontrados em (ALMEIDA, 2012).

3 Resultados

Este capítulo apresenta os resultados deste trabalho, sendo a seção 3.1 com a descrição estatística das variáveis, a seção 3.2 com os resultados da análise fatorial e a seção 3.3 com as análises espaciais.

3.1 Análise Descritiva

A Tabela 4 apresenta as principais medidas descritivas para as variáveis deste estudo. Para não ser repetitivo, a interpretação das medidas será realizada apenas para a variável rendamedia. Observa-se que a renda média dos responsáveis por morador foi de R\$ 651,40, podendo variar, em média, R\$ 1952,00 para mais ou para menos. Essa dispersão foi aproximadamente de 300% do valor da média. Em relação as separatrizes podemos afirmar que o menor valor de renda média foi de R\$ 45,70 e o maior de R\$ 12.000,00; sendo que 25% dos setores censitários tem rendimento médio mensal por morador de até R\$ 103,30; 50% dos setores censitários tem rendimento médio mensal por morador de até R\$ 168,10; 75% dos setores censitários tem rendimento médio mensal por morador de até R\$ 282,60.

Tabela 4: Medidas descritivas das variáveis estudadas

Variáveis	média	dp	cv	min	Q1	Q2	Q3	max
densidade	345.90	45.80	13.20	200.00	321.50	344.70	368.20	514.30
propmulher	51.00	2.90	5.60	40.90	49.60	50.70	52.60	60.50
prop15a24	17.80	5.60	31.70	0.00	16.70	19.20	21.00	26.20
propracapp	63.10	19.80	31.40	0.00	57.10	66.80	76.10	97.60
propalfa	85.90	7.10	8.30	60.00	80.80	87.20	91.20	103.40
rendamedia	651.40	1952.00	299.70	45.70	103.30	168.10	282.60	12000.00
razdep	43.80	12.00	27.30	20.90	34.80	44.00	51.70	90.50
propsenergia	3.90	17.60	452.60	0.00	0.00	0.00	0.90	150.00
propsclixo	100.10	391.50	391.10	0.00	0.70	4.10	39.60	3283.30
prop1banh	66.40	24.10	36.30	0.00	56.70	75.40	83.50	98.00
propmeiosal	54.10	211.90	391.60	0.00	6.80	13.80	30.60	1533.30
propresp65	72.90	290.80	398.70	3.00	8.40	14.70	22.10	2200.00

A Figura 4 a seguir procura descrever as correlações bivariadas das variáveis. Podemos observar que o conjunto de dados aponta aleatoriedade entre as variáveis com correlações positivas e negativas, onde algumas são altamente correlacionadas, por outro lado, vale destacar que as variáveis X2 e X7 não apresentaram correlação com a maioria das variáveis.

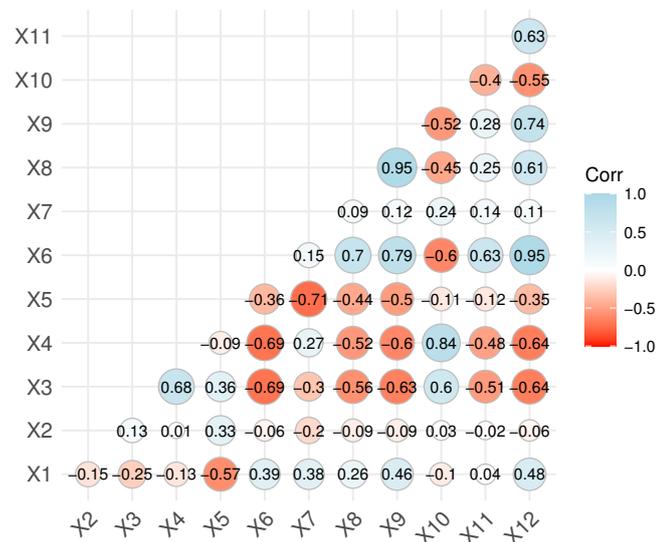


Figura 4: Correlograma contendo as 12 variáveis originais

A Figura 5 apresenta a distribuição das variáveis deste estudo. A Tabela 4 já deu uma idéia da descrição das variáveis, portanto, para complementar a tabela os boxplots ficarão restritos a mencionar somente a forma, para termos uma noção dos níveis de assimetria das variáveis. Analisando o comportamento das variáveis, observa-se que apenas a variável X1 apresentou distribuição aproximadamente simétrica. As variáveis X2, X6, X8, X11 e X12 apresentaram assimetria à direita, com níveis levemente assimétricos de X2 e X11. Já X3, X4, X5, X7, X9 e X10 apresentaram assimetria à esquerda, com níveis levemente assimétricos de X4 e X9.

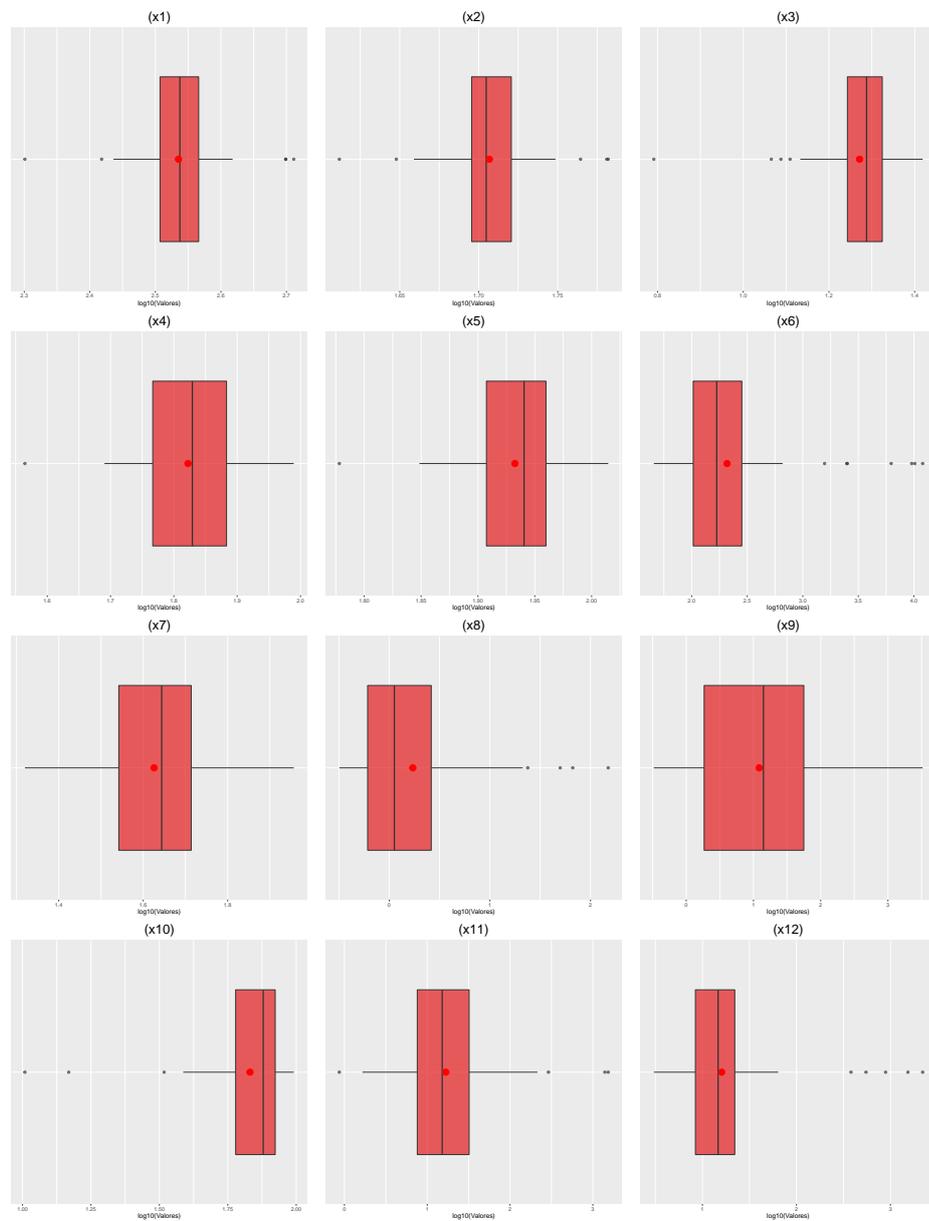


Figura 5: Distribuição dos 12 indicadores na escala logarítmica na base 10

3.2 Análise Fatorial

Aqui na análise fatorial o objetivo foi obter um modelo fatorial dos 95 (tamanho da amostra) setores censitários, cujo fator 1 tivesse a maior percentagem de variância explicada. Os escores fatoriais desse fator 1 serão utilizados como indicador de caracterização da desigualdade social e demográfica dos setores censitários de Mariana.

Para definir o método de extração dos fatores, o teste de normalidade multivariado avaliou a hipótese nula da amostra ser proveniente de uma distribuição normal multivariada, logo o resultado obtido não foi significativo (valor $p < 0.0001$), portanto, rejeitamos a possibilidade de normalidade multivariada dos dados. Consequentemente o método de extração dos fatores será o de componentes principais.

Ao avaliar o grau de correlação parcial, foi obtido uma boa adequabilidade dos dados para a realização de uma análise fatorial exploratória com valor de $KMO = 0,72$. De forma análoga, as medidas de MSA fornecem o KMO para cada variável. A Tabela 5 inicialmente nos dá indícios de que as variáveis $X2 = 0,44$ e $X7 = 0,49$ podem ser retiradas da análise, por apresentar capacidade de explicação insuficiente ($MSA < 0,50$).

Tabela 5: Valores de MSA para cada variável

MSA											
X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
0.59	0.44	0.80	0.74	0.76	0.78	0.49	0.68	0.71	0.91	0.73	0.71

Para-se verificar a existência de correlação suficiente entre as variáveis, o teste de esfericidade de Bartlett avaliou a hipótese nula de igualdade da matriz de correlação com a matriz identidade. O resultado obtido foi significativo (valor $p < 0,0001$), portanto, rejeitamos a hipótese nula, logo existe correlação suficiente entre as variáveis para aplicação da técnica multivariada.

Na análise de retenção dos fatores pelo critério do tamanho da amostra com $n = 95$, temos que as cargas fatoriais serão significativas acima de 0,60, portanto, foram retidos 3 fatores. Na Tabela 6 podemos observar as cargas fatoriais não-rotacionadas, onde as variáveis X1 e X2 não se apresentaram significativas para nenhum fator. Foi obtido no primeiro fator cargas significativas para a maioria das variáveis (X3, X4, X6, X8, X9, X10, X12), em que são observadas correlações positivas e negativas. O segundo fator apresentou correlações apenas para as variáveis X5 e X7, já o terceiro fator com a variável X11.

Posteriormente, a Tabela 7 apresentou a parcela de variabilidade explicada dos dados, onde o primeiro fator é igual a 5,679, o segundo fator uma parcela menor igual a 2,356 e o terceiro fator 1,033. A variabilidade total dos dados pelo solução dos 3 fatores é considerável, sendo aproximadamente de 76%, em que o primeiro fator explica 47,3% da variabilidade total dos dados, o segundo fator 19,6% e o terceiro 8,6%.

Tabela 6: Matriz de cargas fatoriais dos dados originais

	PC1	PC2	PC3
X1	0.46	0.53	0.22
X2	-0.14	-0.37	0.10
X3	-0.82	0.01	0.21
X4	-0.76	0.53	-0.01
X5	-0.46	-0.82	-0.03
X6	0.94	-0.05	-0.08
X7	0.19	0.81	-0.35
X8	0.80	0.06	0.41
X9	0.89	0.09	0.40
X10	-0.69	0.53	-0.02
X11	0.59	-0.18	-0.68
X12	0.90	-0.03	-0.10

Tabela 7: Variação explicada dos dados originais

	PC1	PC2	PC3
Variância Contab.	5.679	2.356	1.033
Proporção da Variância total	0.473	0.196	0.086
Proporção Acumulada	0.473	0.670	0.756

A Tabela 8 apresenta a proporção que cada variável é explicada pelo fator através da análise das comunalidades. Foi observado que algumas variáveis apresentaram alto poder de explicação pelos fatores, já a variável X2 apresentou a menor porcentagem, 17,0% somente, logo é identificada como uma variável passível de eliminação do modelo.

Tabela 8: Comunalidades

X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
0.55	0.17	0.71	0.87	0.89	0.89	0.82	0.81	0.95	0.75	0.85	0.82

Com base na Tabela 9 da variabilidade específica, pode-se observar que a maioria

das variáveis possuem uma pequena parcela não explicada pelos fatores. Já a variável X2 apresentou um valor alto, cerca de 83,0% de variabilidade não explicada pelos fatores.

Tabela 9: Especificidades

1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
0.45	0.83	0.29	0.13	0.11	0.11	0.18	0.19	0.05	0.25	0.15	0.18

Vimos inicialmente que os resultados obtidos utilizando todas as variáveis não apresentaram uma solução satisfatória para os objetivos do estudo, mesmo com uma proporção interessante de explicação. Inicialmente, observado que as variáveis X2 e X7 apresentaram $MSA < 0,50$, iremos reespecificar o modelo fatorial excluindo apenas a variável X2 por apresentar $MSA = 0,44$ e parcela de variabilidade explicada pelo fator de apenas 17%, com intuito de se obter um modelo com maior poder de explicação. Já a variável X7 será mantida pela proximidade do valor mínimo aceitável de MSA e pela alta explicação da variável pelo fator, acima de 80%.

Realizou-se novamente a estatística KMO após a exclusão da variável X2, onde foi observado uma leve melhora no $KMO = 0,74$, indicando boa adequabilidade para aplicação da análise fatorial. Na Tabela 10 a maioria das variáveis apresentaram capacidade de explicação suficiente ($MSA > 0,50$), somente a variável X7 manteve o valor abaixo igual a 0,48.

Tabela 10: Valores de MSA após exclusão da variável X2

X1	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
0.57	0.80	0.75	0.81	0.79	0.48	0.69	0.72	0.91	0.75	0.72

O teste de esfericidade de Bartlett novamente avaliou a hipótese nula de igualdade da matriz de correlação com a matriz identidade. O resultado obtido foi significativo (valor $p < 0,0001$), portanto, após a exclusão da variável X2 podemos prosseguir com a análise, pois os dados apresentaram correlação suficiente.

Observando a Tabela 11 das cargas fatoriais, novamente 3 fatores foram retidos. Não houve alteração dos fatores após a exclusão da variável x2, demonstrando que a variável não era representativa para o modelo. Na Tabela 12 com as 11 variáveis, houve um aumento da variabilidade total explicada para 81,5%. Percebemos que o primeiro fator

aumentou o poder de explicação por constituir a maioria das variáveis, aproximadamente 52% da variabilidade total dos dados. O segundo fator 20,7% e o terceiro 9,4%.

Tabela 11: Matriz de cargas fatoriais após exclusão da variável X2

	PC1	PC2	PC3
X1	0.46	0.55	-0.21
X3	-0.82	0.01	-0.21
X4	-0.77	0.54	0.03
X5	-0.44	-0.82	0.03
X6	0.94	-0.03	0.09
X7	0.18	0.82	0.36
X8	0.80	0.07	-0.41
X9	0.89	0.12	-0.39
X10	-0.69	0.54	0.04
X11	0.59	-0.17	0.69
X12	0.90	-0.01	0.11

Tabela 12: Variância explicada pelos fatores após exclusão da variável X2

	PC1	PC2	PC3
Variância Contab.	5.653	2.280	1.038
Proporção da variância total	0.514	0.207	0.094
Proporção Acumulada	0.514	0.721	0.815

Podemos observar novamente na Tabela 13, das comunalidades, que todas as variáveis apresentaram bom relacionamento com os fatores.

Tabela 13: Comunalidades após exclusão da variável X2

X1	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
0.55	0.71	0.88	0.88	0.89	0.85	0.81	0.95	0.77	0.86	0.83

Logo percebemos que a solução encontrada para o modelo reespecificado apresenta um grau de explicação até o momento útil. Mas analisando as cargas fatoriais disponíveis na Tabela 11, percebemos que algumas variáveis apresentaram cargas interessantes, próximas de 0,60, onde iremos aplicar a método da rotação varimax para investigar a melhor alocação para composição de cada fator.

Logo, após a rotação das cargas fatoriais apresentadas na Tabela 14, vimos que o resultado obtido não foi positivo para nosso estudo. As variáveis X1 e X10 não são sig-

nificativas para nenhum dos fatores e cargas cruzadas para três variáveis. O valor da variabilidade total apresentada na Tabela 15 não se altera, logo permanece aproximadamente 81,5%. onde 33,5% é explicada pelo primeiro fator, 21,6% pelo segundo e 26,5% pelo terceiro fator. Portanto pode-se perceber que apesar das mudanças observadas, o modelo rotacionado não se mostrou interessante.

Tabela 14: Matriz de cargas fatoriais rotacionada após a exclusão da variável X2

	PC1	PC2	PC3
X1	0.46	0.59	-0.01
X3	-0.48	-0.19	-0.67
X4	-0.62	0.37	-0.60
X5	-0.32	-0.87	-0.04
X6	0.65	0.18	0.66
X7	-0.13	0.89	0.16
X8	0.87	0.16	0.18
X9	0.92	0.23	0.24
X10	-0.57	0.39	-0.54
X11	0.00	0.07	0.92
X12	0.61	0.19	0.64

Tabela 15: Variância explicada pelos fatores rotacionada após exclusão da variável X2

	PC1	PC2	PC3
Variância Contab.	3.684	2.376	2.910
Proporção da variância total	0.335	0.216	0.265
Proporção Acumulada	0.335	0.551	0.815

Considerando os resultados obtidos até o momento, vimos que o método utilizado da rotação varimax não foi satisfatório. Diante dos dados apresentados para o estudo, temos que a melhor solução encontrada pelo método de componentes principais é o fator 1 das cargas não-rotacionadas, onde foi obtido aproximadamente 52% de explicação conjunta da variabilidade total das 11 variáveis. A Tabela 16 a seguir, apresenta a comparação dos resultados obtidos da melhor solução encontrada para prosseguir com o estudo.

Tabela 16: Cargas iniciais e de rotação varimax e variância explicada para os fatores retidos

Variáveis	Cargas Fatoriais Antes da Rotação			Cargas Fatoriais Após a Rotação Varimax			Comu. \hat{h}_i^2
	f1	f2	f3	f1	f2	f3	
X1	0.46	0.55	-0.21	0.46	0.59	-0.01	0.55
X3	-0.82	0.01	-0.21	-0.48	-0.19	-0.67	0.71
X4	-0.77	0.54	0.03	-0.62	0.37	-0.60	0.88
X5	-0.44	-0.82	0.03	-0.32	-0.87	-0.04	0.88
X6	0.94	-0.03	0.09	0.65	0.18	0.66	0.89
X7	0.18	0.82	0.36	-0.13	0.89	0.16	0.85
X8	0.80	0.07	-0.41	0.87	0.16	0.18	0.81
X9	0.89	0.12	-0.39	0.92	0.23	0.24	0.95
X10	-0.69	0.54	0.04	-0.57	0.39	-0.54	0.77
X11	0.59	-0.17	0.69	0.00	0.07	0.92	0.86
X12	0.90	-0.01	0.11	0.61	0.19	0.64	0.83
Variância Contab.	5.65	2.28	1.04	3.68	2.37	2.91	8.98
Proporção da Variância total	0.51	0.21	0.09	0.33	0.22	0.26	0.81
Proporção Acumulada	0.51	.72	0.81	0.33	0.55	0.81	0.81

Portanto, o primeiro fator do modelo fatorial ajustado pelo método dos mínimos quadrados ponderados (Bartlett) com 11 variáveis é representado pela seguinte equação:

$$\begin{aligned} \hat{f}_{1k} = & 0,01664682Z_{1k} - 0,07043409Z_{2k} - 0,14306007Z_{3k} - \\ & 0,07017190Z_{4k} + 0,20555396Z_{5k} + 0,04587964Z_{6k} + 0,06955680Z_{7k} + \\ & 0,32454629Z_{8k} - 0,06684540Z_{9k} + 0,15092047Z_{10k} + 0,12491210Z_{11k} \end{aligned} \quad (3.1)$$

em que $k = 1, 2, 3, \dots, 95$ representa o setor censitário. Observamos as variáveis que aumentam o escore deste fator, assim como, variáveis que diminuem. Logo o conjunto de variáveis que proporcionam o aumento do valor do escore são: (X1) Densidade domiciliar; (X3) Proporção de pessoas de 15 a 24 anos de idade; (X6) Renda média dos responsáveis por cada morador; (X7) Razão de dependência; (X8) Proporção de domicílios sem energia elétrica; (X9) Proporção de domicílios sem coleta de lixo; (X11) Proporção de responsáveis com rendimento mensal de até meio salário mínimo; (X12) Proporção de responsáveis de pessoas por domicílio com idade acima de 65 anos. As variáveis que diminuem o valor desses escores são: (X4) Proporção de pessoas pretas ou pardas; (X5) Proporção de

alfabetizados; (X10) Proporção de domicílios com até um banheiro.

Podemos destacar que as variáveis (X9) proporção de domicílios sem coleta de lixo e a variável (X6) Renda média dos responsáveis por cada morador, possuem peso positivo, e portanto, proporcionam aumento no escore do fator. Por outro lado as variáveis (X4) Proporção de pessoas pretas ou pardas e a variável (X5) Proporção de pessoas alfabetizadas diminuem o escore, já que seus pesos são negativos.

Portanto, o fator 1 foi composto por 11 variáveis referentes a caracterização de desigualdade social e demográfica, sendo 8 numa direção de mais desigualdade e 3 de menos desigualdade.

Para facilitar a interpretação, o escore do fator 1 foi padronizado para a escala 0 a 100 conforme a expressão abaixo:

$$\hat{f}_{1k}^{pad} = \frac{\hat{f}_{1k} - \min(\hat{f}_{1k})}{\max(\hat{f}_{1k}) - \min(\hat{f}_{1k})} \times 100 \quad (3.2)$$

A partir de agora, o escore do fator 1 padronizado será considerado o índice de caracterização de desigualdade social e demográfica (ICDSD) dos setores censitários de Mariana.

3.3 Análise Espacial

Como a unidade de análise deste trabalho é setor censitário vimos a necessidade de apresentar espacialmente os escores nos respectivos setores censitários de Mariana, bem como o seu nível de dependência espacial.

A Figura 6 apresenta a distribuição do índice da caracterização de desigualdade social e demográfica (ICDSD) dos distritos de Mariana na escala logarítmica na base 10. O distrito da sede de Mariana apresentou assimetria à direita com presença de outliers, demonstrando uma variabilidade significativa do ICDSD. Os distritos de Cachoeira do Brumado, Monsenhor Horta, Cláudio Manuel e Furquim apresentaram assimetria à esquerda e baixa variabilidade no ICDSD. Os distritos de Padre Viegas e Camargos apresentaram aparentemente simétricos, com pouca variabilidade do ICDSD no distrito, enquanto Bandeirantes apresentou simetria do ICDSD, mas com uma certa variabilidade. Os distritos Santa Rita Durão e Passagem de Mariana apresentaram maior variabilidade entre os demais distritos do município, com assimetria à direita.

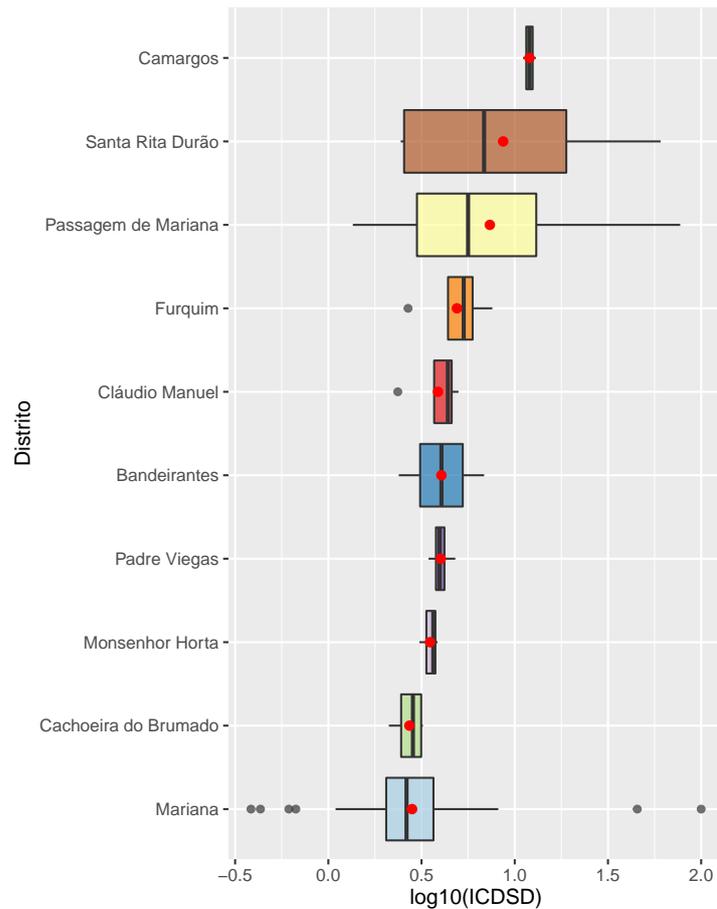


Figura 6: Distribuição dos 10 distritos na escala logarítmica na base 10

A Figura 7 apresenta a distribuição espacial do índice de moran local para os valores do índice da caracterização de desigualdades social e demográfica (ICDSD) dos setores censitários de Mariana. A distribuição do ICDSD foi categorizada em cinco faixas de 20% cada uma, denominadas quintis. Os distritos de Camargos, Santa Rita Durão, Passagem de Mariana, Bandeirantes e Furquim com áreas hachuradas em vermelho representam os 20% setores censitários com maiores valores de ICDSD, enquanto Padre Viegas, Monsenhor Horta e Cláudio Manuel com áreas hachuradas em vermelho claro representam os próximos 20% setores censitários com maiores valores de ICDSD. O distrito de Cachoeira do Brumado apresentou estar entre os 20% setores censitários medianos em relação aos valores de ICDSD, e para a sede o mapa não traz uma fotografia clara, onde se terá uma melhor visualização nas Figuras 9 e 11.

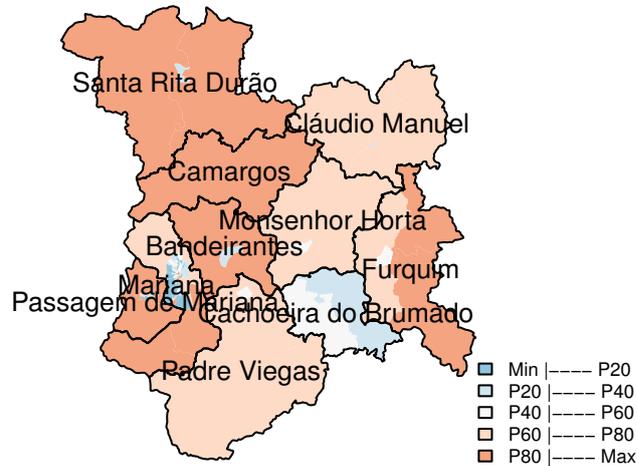


Figura 7: Distribuição espacial do ICDSD dos setores do Município

A Figura 8 apresenta a distribuição espacial do índice de moran local para os valores da renda dos responsáveis por morador dos setores censitários de Mariana. Os distritos de Camargos, Bandeirantes, Santa Rita Durão, Cláudio Manuel e Passagem de Mariana apresentaram setores com áreas hachuradas em azul que representam os 20% setores censitários com maiores rendas de responsáveis por morador, enquanto as áreas hachuradas em azul claro representam os próximos 20% setores censitários com maiores valores de renda de responsáveis por morador. Os distritos de Monsenhor Horta, Padre Viegas, Cachoeira do Brumado e Furquim apresentaram áreas hachuradas em vermelho que representam os 20% setores censitários com menores rendas de responsáveis por morador, onde pode-se observar que alguns setores desses distritos apresentam renda de responsáveis por morador entre os 40% menores.

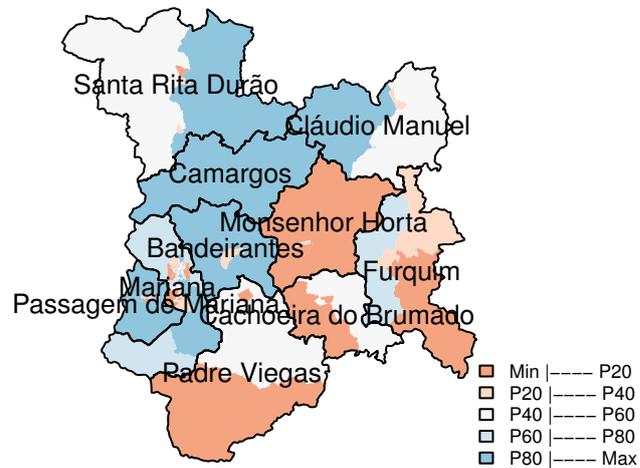


Figura 8: Distribuição espacial da renda média por morador dos setores do Município

A Figura 9 apresenta a distribuição espacial do índice de moran local para os valores do ICDS para os setores censitários do distrito da sede do município. Os setores localizados nas zonas rurais ao redor do município apresentam áreas hachuradas em vermelho que representam os 20% setores censitários com maiores valores de ICDS e vermelho claro que representam os próximos 20% setores censitários com maiores valores de ICDS. Os setores localizados na área urbana serão apresentados com maiores detalhes pela Figura 11, pois o mapa não traz uma fotografia clara.

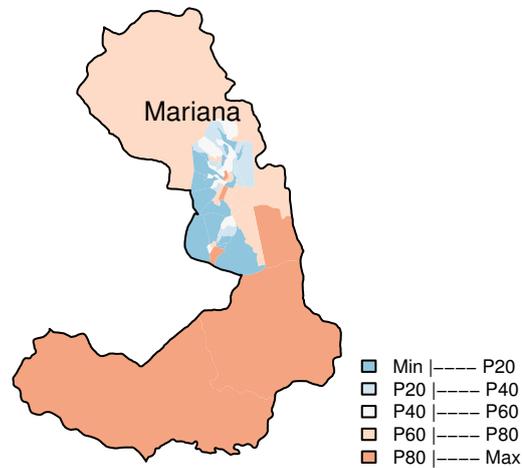


Figura 9: Distribuição espacial do ICDSD dos setores na sede

A Figura 10 apresenta a distribuição espacial do índice de moran local para os valores da renda dos responsáveis por morador dos setores censitários na sede do município. Os setores censitários localizados nas zonas rurais ao redor do município apresentaram setores com áreas hachuradas em azul e azul claro, que representam, respectivamente, os 20% setores censitários com maiores rendas de responsáveis por morador e os próximos 20% setores censitários com maiores rendas de responsáveis por morador. Os setores localizados na área urbana serão apresentados com maiores detalhes na Figura 12.

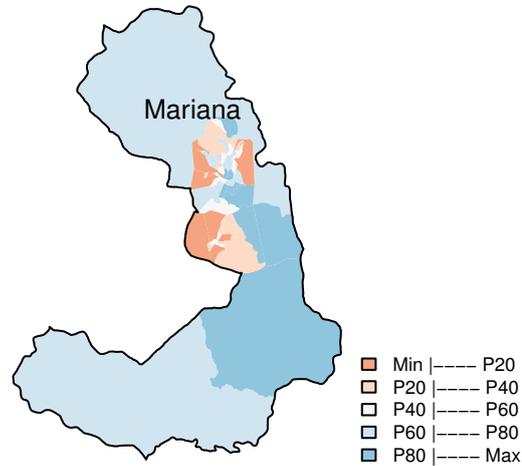


Figura 10: Distribuição espacial da renda média por morador dos setores na sede

A Figura 11 apresenta a distribuição espacial do índice de moran local para os valores do índice da caracterização de desigualdades social e demográfica (ICDSD) dos setores censitários da área urbana de Mariana. Os setores com áreas hachuradas em azul representam os 20% setores censitários com menores valores de ICDSD, localizados em algumas das regiões extremas da área urbana, enquanto os setores com áreas hachuradas em azul claro representam os próximos 20% setores censitários com menores valores de ICDSD, também localizados em algumas das regiões extremas e próximas da região central. Os setores que apresentam áreas hachuradas em vermelho, representam os 20% setores censitários com maiores valores de ICDSD, onde estão mais localizados na região central do mapa, assim como, os setores com áreas hachuradas em vermelho claro que representam os próximos 20% setores com maiores valores de ICDSD.

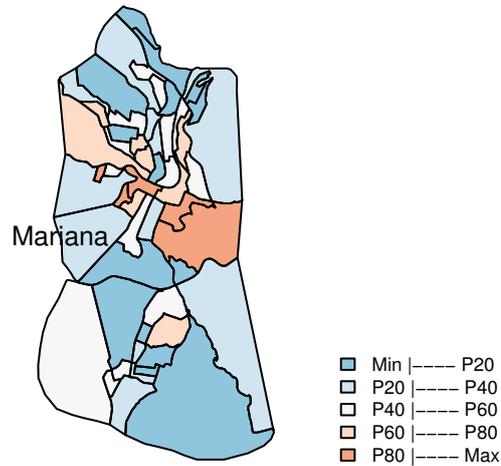


Figura 11: Distribuição espacial do ICDS dos setores na área urbana

A Figura 12 apresenta a distribuição espacial do índice de moran local para renda média dos responsáveis por morador dos setores censitários da área urbana de Mariana. Os setores com áreas hachuradas em azul representam os 20% setores censitários com maiores valores de renda média dos responsáveis por morador, localizados nas regiões centrais do mapa, enquanto os setores com áreas hachuradas em azul claro representam os próximos 20% setores censitários com maiores valores de renda média dos responsáveis por morador, distribuídos espacialmente pelos setores da área urbana. Os setores que apresentam áreas hachuradas em vermelho, representam os 20% setores censitários com menores valores de renda média dos responsáveis por morador, onde estão distribuídos espacialmente, assim como, os setores com áreas hachuradas em vermelho claro que representam os próximos 20% setores com menores renda média de responsáveis por morador.

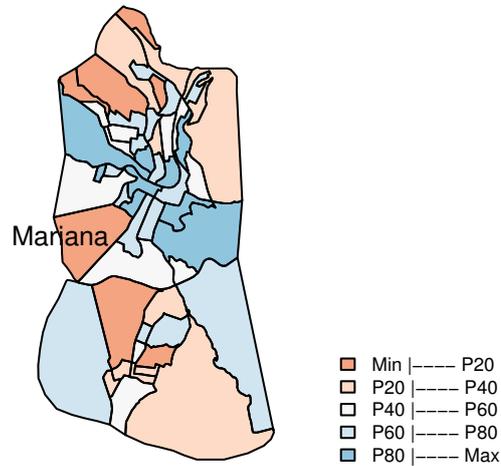


Figura 12: Distribuição espacial do ICDSD da renda média por morador dos setores na área urbana

Dividido em quatro quadrantes, a Figura 13 apresenta o diagrama de espalhamento de Moran, que permite analisar o comportamento da variabilidade espacial para os setores censitários de Mariana. Os quadrantes 1 e 2 indicam dependência espacial entre os setores, enquanto os quadrantes 3 e 4 não apresentam associação espacial. Podemos observar que a grande maioria dos setores estão localizados nos quadrantes 1 e 2, em que, o segundo quadrante apresenta a maior quantidade de setores e o primeiro quadrante uma quantidade menor, ambas indicando associação espacial. Os setores presentes nos quadrantes 3 e 4 não apresentam associação espacial, onde o quadrante 3 possui somente uma variável e o quadrante 4 possui uma quantidade menor, em relação ao quadrante com maior quantidade.

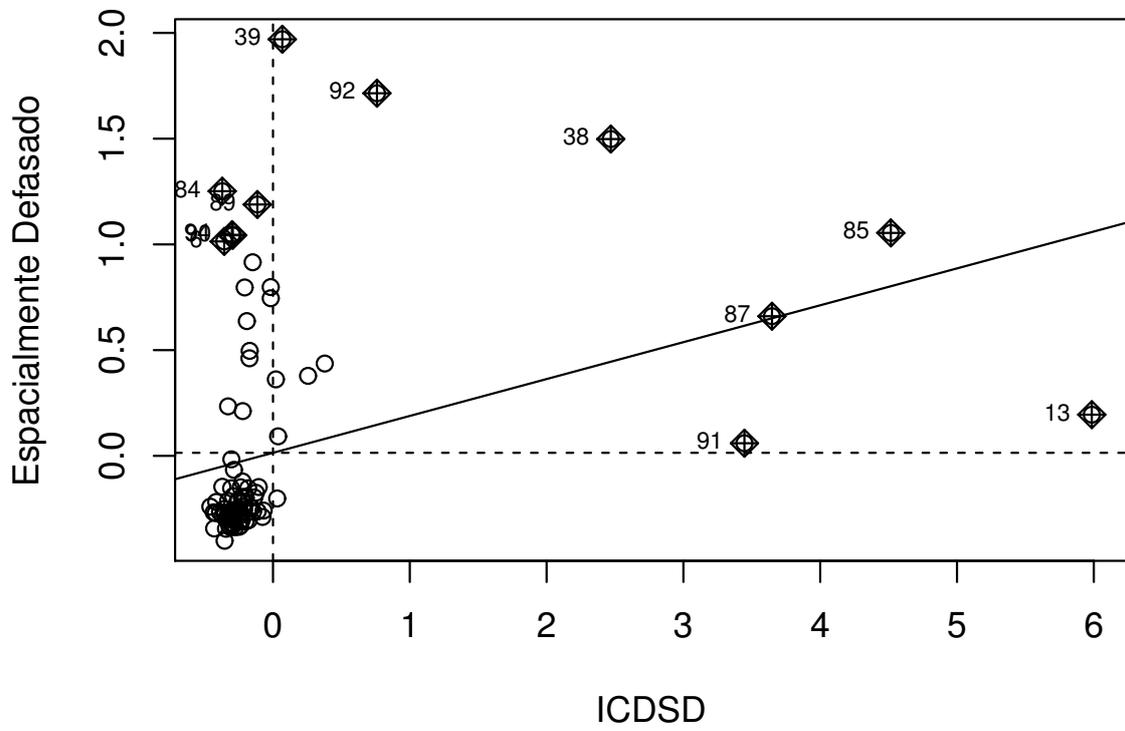


Figura 13: Diagrama de espalhamento de Moran

A Figura 14 apresenta o mapa de Moran do ICDS D segundo os setores censitários de Mariana em 2010. As regiões com áreas hachuradas com a cor azul escuro representam 72,60% dos setores censitários com valor do ICDS D baixo na área e valor médio do ICDS D para os setores vizinhos também baixo. As áreas hachuradas com a cor azul claro representam 14,70% dos setores censitários em que o valor do ICDS D é baixo na área e o valor médio do ICDS D para os setores vizinhos é alto. A cor vermelho claro representa 1,05% dos setores censitários em que o valor do ICDS D na área é alto e o valor médio do ICDS D para a vizinhança é baixo. A cor vermelho escuro representa 11,65% dos setores censitários em que o valor de ICDS D na área é alto e o valor médio de ICDS D para os setores vizinhos também é alto.

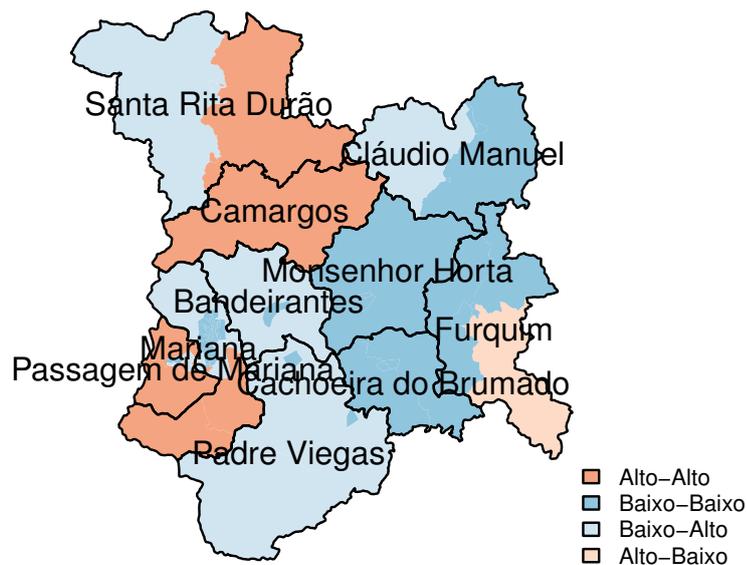


Figura 14: Distribuição espacial de Moran local

A Figura 15 representa o mapa de Moran significativo do ICDSD segundo os setores censitários de Mariana. Observa-se que as áreas hachuradas em vermelho destacam as regiões cuja associação de Moran foi significativa, representando os setores com alto índice da caracterização de desigualdade social e demográfica e que são vizinhos de setores com alto ICDSD.

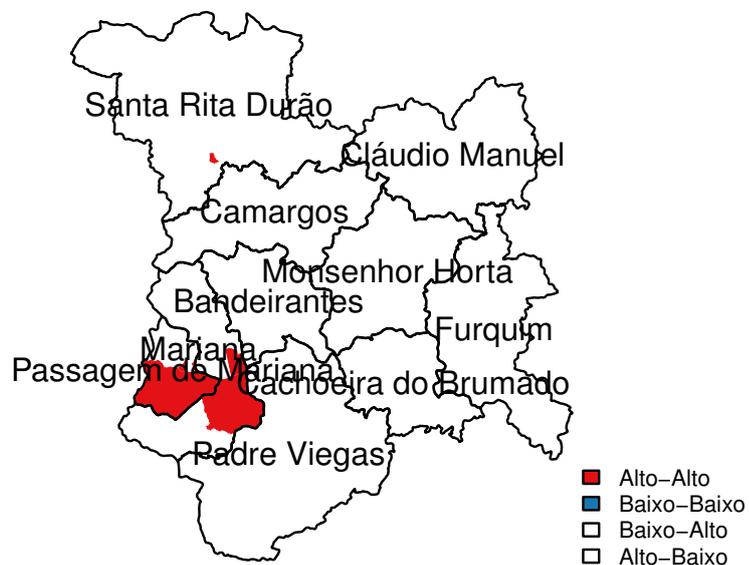


Figura 15: Mapa de Moran significativo do ICDSD

4 Considerações Finais

Este trabalho propôs construir e avaliar um índice de caracterização de desigualdade social e demográfica (ICDSD) para os setores censitários de Mariana com dados do censo de 2010. Nesse sentido, um modelo fatorial foi ajustado a partir de variáveis econômicas e sociais. O escore do primeiro fator foi definido como sendo esse índice, e realizou-se uma análise de associação espacial para verificar possíveis conjuntos de setores censitários com índice alto ou baixo.

O modelo fatorial obtido reteve 3 fatores cujo o autovalor foi maior que 1, sendo que o fator 1 que gerou o índice proposto neste trabalho teve a capacidade de explicar 51,4% da variabilidade dos dados originais. Este índice foi composto por 11 variáveis sendo que, das quais 8 estavam fortemente correlacionadas com este primeiro fator. O escore desse índice foi analisado a partir do índice de Moran, e concluiu-se que 72,60% dos setores apresentaram ICDSD baixo na área e também nos seus vizinhos. Em uma região da cidade encontrou-se alto ICDSD nos setores e nos vizinhos desses setores, com significância pelo índice local de Moran. Esta área portanto deveria receber mais atenção/estudos por parte dos gestores públicos.

Uma proposta de trabalho futuro seria comparar o ICDSD deste censo de 2010 com os censos anteriores e posteriores, por exemplo, 2000 e 2022 que será coletado ainda. Este trabalho poderia ser estendido por aplicar modelos de regressão espacial. Uma outra possível extensão seria dividir a cidade em grupos a partir de uma análise de clusters e assim auxiliar na formulação de políticas públicas focalizadas.

Referências

- ABREU, M. V. S. et al. Proposta metodológica para o cálculo e análise espacial do idh intraurbano de viçosa-mg. *Revista Brasileira de Estudos da População*, v. 28, n. 1, p. 169–186, 2011.
- ALBUQUERQUE, M. V. d. et al. *Desigualdades regionais na saúde: Mudanças observadas no Brasil de 2000 a 2016*. Nov 2016. Cien Saude Colet. [Citado em 13/02/2021]. Está disponível em: <http://www.cienciaesaudecoletiva.com.br/artigos/desigualdades-regionais-na-saude-mudancas-observadas-no-brasil-de-2000-a-2016/15903?id=15903>.
- ALMEIDA, E. *Econometria Espacial Aplicada: Análise exploratória de dados espaciais*. Campinas - SP: atomoealinea, 2012. 102–147 p. (1207510, v. 1).
- ANSELIN, L. *Geographical Analysis: Local indicators of spatial association-lisa*. Rio de Janeiro - RJ: The Ohio State University, 1994. 93–115 p.
- BARBOSA, I. R.; GONCALVES, R. C. B.; SANTANA, R. L. *Social vulnerability map for the municipality of Natal (Northeast Brazil) at a census sector level*. Dec 2019. J Hum Growth Dev; 29(1): 48-56. <http://dx.doi.org/10.7322/jhgd.157749>.
- DEUS, J. L. d. *Inclusão financeira na perspectiva de gênero, renda e situação ocupacional: Determinantes e entraves para o município de Mariana (Minas Gerais)*. Mariana, MG, Brasil: [s.n.], jul. 2019.
- FARIAS, C. A.; FIGUEIREDO, A. M.; LIMA, J. E. d. Dependência espacial e análise de agrupamentos de municípios para diferentes tipos de crime em minas gerais. *Revista Reuna*, v. 13, n. 3, p. 67–83, 2008.
- FERREIRA, D. F. *Estatística Multivariada*. 1th. ed. Lavras: UFLA, 2008.
- FÁVERO, L. P. et al. *Análise de Dados, MODELAGEM MULTIVARIADA PARA TOMADA DE DECISÕES: Análise fatorial*. Rio de Janeiro: Elsevier, 2009. 235–266 p.
- GILBERTO, C. et al. *Análise Espacial de Dados Geográficos*. Brasília: Livraria Virtual da EMBRAPA, 2004. 157–206 p.
- HAIR, J. F. J. et al. *Análise Multivariada de Dados: Análise fatorial*. São Paulo - SP: Pearson Education, 2009. 100–145 p.
- KAISER, H. F. *O critério varimax para rotação analpitica em análise fatorial*. Sep 1958. <https://doi.org/10.1007/BF02289233>.
- LAROS, J. A. *O Uso da Análise Fatorial: Algumas Diretrizes para Pesquisadores*. Brasília - DF: [s.n.], 2012. 141-160 p. ISBN 978-85-65091-00-8.

- MOREIRA, R. d. C.; FOCHEZATTO, A. Análise espacial da criminalidade no estado da bahia. *RDE - Revista de Desenvolvimento Econômico*, v. 3, n. 38, p. 52–80, 2017.
- NEUHAUS, J. O.; WRIGLEY, C. *Mathematical and Statistical Psychology*. United States: February, May and November, 1954. 81–91 p. (2, v. 7th).
- O'SULLIVAN, D.; UNWIN, D. *Geographic information analysis*. 2th. ed. New Jersey: John Wiley & Sons, 2010.
- PEROBELLI, F. S. et al. *Planejamento regional e potenciais de desenvolvimento dos municípios de Minas Gerais na região em torno de Juiz de Fora: uma aplicação de análise fatorial*, v. 9, n. 1. Dec 2013. Dec., 2013. Disponível em: <<https://revistas.face.ufmg.br/index.php/novaeconomia/article/view/2194>>. Acesso em: 13 fev. 2021.>.
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2020. Disponível em: <<https://www.R-project.org/>>.
- RENCHER, C. A. *Methods of Multivariate Analysis*. 2th. ed. United States of America: A JOHN WILEY & SONS, INC. PUBLICATION, 2002.
- SCHNEIDER, S.; WAQUIL, P. D. Caracterização socioeconômica dos municípios gaúchos e desigualdades regionais. *Revista de Economia e Sociologia Rural (RESR)*, v. 39, n. 3, p. 117–142, 2001.
- WOOD, H. C.; CARVALHO, J. A. M. d. *A demografia da desigualdade no Brasil*. Rio de Janeiro: IPEA, 1994.
- XERXENEVSKY, L. L.; FOCHEZATTO, A. Índice relativo de desenvolvimento socioeconômico dos municípios do litoral norte do rio grande do sul: Uma aplicação da análise fatorial. *Boletim Geográfico do Rio Grande do Sul*, v. 45, n. 25, p. 31–55, 2015.

APÊNDICE A - 1

Arquivo	Código	Descrição
Básico	V001	Domicílios particulares permanentes ou pessoas responsáveis por domicílios particulares permanentes
	V002	Moradores em domicílios particulares permanentes ou população residente em domicílios particulares permanentes
	V003	Média do número de moradores em domicílios particulares permanentes (obtida pela divisão de Var2 por Var1)
	V005	Valor do rendimento nominal médio mensal das pessoas responsáveis por domicílios particulares permanentes (com e sem rendimento)
Pessoa01	V001	Pessoas alfabetizadas com 5 ou mais anos de idade
Pessoa03	V003	Pessoas Residentes e cor ou raça - preta
	V005	Pessoas Residentes e cor ou raça - parda
Pessoa12	V002	Mulheres em domicílios particulares e domicílios coletivos
Pessoa13	V022	Pessoas com menos de 1 ano de idade
	V035	Pessoas de 1 ano de idade
	V036	Pessoas com 2 anos de idade
	⋮	⋮
	V133	Pessoas com 99 anos de idade
	V134	Pessoas com 100 anos ou mais de idade

Domicílio01	V023	Domicílios particulares permanentes sem banheiro de uso exclusivo dos moradores e nem sanitário
	V025	Domicílios particulares permanentes com 1 banheiro de uso exclusivo dos moradores
	V036	Domicílios particulares permanentes com lixo coletado por serviço de limpeza
	V046	Domicílios particulares permanentes sem energia elétrica
Responsável02	V057	Pessoas Responsáveis com 65 anos de idade
	V058	Pessoas Responsáveis com 66 anos de idade
	⋮	⋮
	V092	Pessoas Responsáveis com 100 ou mais anos de idade
ResponsávelRenda	V001	Pessoas responsáveis com rendimento nominal mensal de até 1/2 salário mínimo
	V010	Pessoas responsáveis sem rendimento nominal mensal