



MATEUS SANTOS PEIXOTO

**MODELAGEM TOBIT EM DADOS DE ÁREA DOS CASOS
NOTIFICADOS POR TUBERCULOSE NOS MUNICÍPIOS DE
MINAS GERAIS**

**LAVRAS – MG
2023**

MATEUS SANTOS PEIXOTO

**MODELAGEM TOBIT EM DADOS DE ÁREA DOS CASOS
NOTIFICADOS POR TUBERCULOSE NOS MUNICÍPIOS DE
MINAS GERAIS**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária para a obtenção do título de Mestre.

Prof. Dr. João Domingos Scalon
Orientador

**LAVRAS – MG
2023**

**Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).**

Peixoto, Mateus Santos.

Modelagem tobit em dados de área dos casos notificados por
tuberculose nos municípios de Minas Gerais. / Mateus Santos
Peixoto. - 2023.

50 p. : il.

Orientador(a): João Domingos Scalon.

Dissertação (mestrado acadêmico) - Universidade Federal de
Lavras, 2023.

Bibliografia.

1. Estatística Espacial. 2. Dados de Área. 3. Modelo Tobit. I.
Scalon, João Domingos. II. Título.

MATEUS SANTOS PEIXOTO

**MODELAGEM TOBIT EM DADOS DE ÁREA DOS CASOS NOTIFICADOS POR
TUBERCULOSE NOS MUNICÍPIOS DE MINAS GERAIS**

**TOBIT MODELING ON CASE AREA DATA REPORTED FOR TUBERCULOSIS IN
THE MUNICIPALITIES OF MINAS GERAIS**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária para a obtenção do título de Mestre.

APROVADA em 28 de Novembro de 2023.
Prof. DSc. Ricardo Alves de Olinda UEPB
Prof. DSc. Lourenço Manuel UEM

Prof. Dr. João Domingos Scalon
Orientador

**LAVRAS – MG
2023**

A todos os meus familiares e amigos que me incentivaram durante todo o processo.
Dedico.

AGRADECIMENTOS

Primeiramente agradecer a Deus, pela minha vida, e por me permitir ultrapassar todos os obstáculos encontrados durante todos os meus anos de estudos.

Aos meus pais Ana Rita e Cristiano Messias por nunca terem medido esforços para me proporcionar um ensino de qualidade durante todo o meu período escolar e irmãos Thalita, Gabriel e Lais, que mesmo longe sempre estiveram torcendo por mim e me incentivando nos momentos difíceis e compreenderam a minha ausência enquanto eu me dedicava.

Aos meus tios Ana Patricia e Tiago Almeida que me acolheram na casa deles durante a graduação e me ajudaram durante a minha jornada pessoal e profissional ao ingressar no mestrado, pelos puxões de orelha e todo cuidado que tiveram comigo. A minha avó Carmem (in memoriam) que sempre me incentivava e fazia de tudo por mim.

Sou grato à minha família pelo apoio que sempre me deram durante toda a minha vida e pela compreensão de minha ausência nas reuniões familiares.

Aos amigos Gabriel, Mateus, Joabe, Valdinei, Alexsander, Guilherme, Viviane, Diana, Yhan, Victória, Arlley, Lucas, Wélson pela amizade e por todos os estudos e conhecimentos compartilhados.

As pessoas com quem convivi ao longo desses anos de curso, que me incentivaram e que certamente tiveram impacto na minha formação acadêmica.

O presente trabalho foi realizado com apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

Por último, mas não menos importante a todos os meus professores do programa de Estatística e experimentação agropecuária da Universidade Federal de Lavras e ao grupo de estudos GPS, pela excelência da qualidade técnica de cada um que me permitiu apresentar um melhor desempenho no meu processo de formação profissional, pelas correções e ensinamentos ao longo do curso, deixo um agradecimento especial ao meu orientador João Domingos Scalon pela ajuda e pela paciência com a qual guiou o meu aprendizado me orientando no meu trabalho que está sendo um divisor de águas para minha formação.

A estatística é a gramática da ciência.
(Karl Pearson)

Não existe um caminho para a felicidade. A felicidade é o caminho.
(Mahatma Gandhi)

O otimista é um tolo. O pessimista, um chato. Bom mesmo é ser um realista esperançoso.
(Ariano Suassuna)

RESUMO

A Tuberculose (TB) é uma das doenças de notificação compulsória mais negligenciadas do país, uma significativa questão de Saúde Pública, especialmente no Brasil, onde sua incidência persiste como um desafio. Para aprofundar a compreensão da distribuição geográfica da doença e identificar áreas com maior risco de transmissão, este estudo utiliza dados provenientes do DATASUS (Departamento de Informática do Sistema Único de Saúde) órgão do Ministério da Saúde, através da ferramenta de tabulação TabNet. A análise é conduzida considerando o arquivo de mapa "shape" do Instituto Brasileiro de Geografia e Estatística (IBGE) do estado de Minas Gerais.

O objetivo principal é mapear áreas propensas à propagação da tuberculose, explorando os impactos socioeconômicos e sociodemográficos. A abordagem adotada inclui a análise de dados de área, permitindo a investigação da relação entre incidência de tuberculose e características geográficas, socioeconômicas e demográficas.

Os modelos Tobit e Tobit espacial (SAR-TOBIT) são ferramentas essenciais nesse processo, uma vez que o modelo espacial leva em consideração a autocorrelação espacial presente nos dados, que frequentemente apresentam um padrão positivo.

A investigação das variáveis socioeconômicas destacou uma clara disparidade, apontando que as regiões Norte e Nordeste de Minas Gerais apresentam condições menos favoráveis em comparação com outras áreas. De maneira interessante, Itambé de Mato Dentro foi o único município localizado no quadrante 1, sugerindo uma possível influência no aumento dos casos de tuberculose para os vizinhos.

A dependência espacial revelada pelo tobit espacial é crucial para identificar áreas onde estratégias de controle podem não estar alcançando os resultados desejados. A implementação eficaz desse modelo é facilitada pelo uso do software R. Varias medidas podem ser recomendadas para abordar a tuberculose, isso inclui a melhoria do acesso aos serviços de saúde para diagnóstico e tratamento adequados, campanhas educativas para aumentar a conscientização, programas de rastreamento e testagem em massa, garantia de tratamento e acompanhamento adequados para pacientes diagnosticados, além de esforços para abordar os determinantes sociais da saúde. Dessa forma, esta pesquisa contribui substancialmente para o monitoramento e controle da tuberculose no Estado de Minas Gerais.

Palavras-chave: Tuberculose. Incidência. Fatores de risco. Análise espacial. Controle da tuberculose. Modelo Tobit.

ABSTRACT

Tuberculosis (TB) is one of the most neglected notifiable diseases in the country, a significant Public Health issue, especially in Brazil, where its incidence remains a challenge. To deepen the understanding of the geographic distribution of the disease and identify areas with a higher risk of transmission, this study uses data from DATASUS (Department of Informatics of the Unified Health System), an organization of the Ministry of Health, using the TabNet tabulation tool. The analysis is conducted considering the "shape" map file from the Brazilian Institute of Geography and Statistics (IBGE) for the state of Minas Gerais.

The main objective is to map areas prone to the spread of tuberculosis, exploring the socioeconomic and sociodemographic impacts. The approach adopted includes the analysis of area data, allowing the investigation of the relationship between tuberculosis incidence and geographic, socioeconomic and demographic characteristics.

The Tobit and spatial Tobit (SAR-TOBIT) models are essential tools in this process, since the spatial model takes into account the spatial autocorrelation present in the data, which often presents a positive pattern.

The investigation of socioeconomic variables highlighted a clear disparity, pointing out that the North and Northeast regions of Minas Gerais present less favorable conditions compared to other areas. Interestingly, Itambé de Mato Dentro was the only municipality located in quadrant 1, suggesting a possible influence on the increase in tuberculosis cases for neighbors.

The spatial dependence revealed by spatial tobit is crucial for identifying areas where control strategies may not be achieving desired results. Effective implementation of this model is facilitated by the use of R software. Various measures can be recommended to address tuberculosis, this includes improving access to health services for proper diagnosis and treatment, educational campaigns to increase awareness, screening programs and mass testing, ensuring adequate treatment and follow-up for diagnosed patients, as well as efforts to address social determinants of health. In this way, this research contributes substantially to the monitoring and control of tuberculosis in the State of Minas Gerais.

Keywords: Tuberculosis. Incidence. Risk factors. Spatial analysis. Tuberculosis control. Tobit model.

Impactos sociais, tecnológicos, econômicos e culturais

Este estudo aborda a distribuição geográfica da tuberculose (TB) em Minas Gerais, utilizando dados do DATASUS e do IBGE. Identificaram-se áreas com maior risco de transmissão de TB, permitindo medidas preventivas direcionadas, e destacou-se a disparidade socioeconômica entre as regiões Norte e Nordeste de Minas Gerais, evidenciando a necessidade de intervenções específicas. A análise dos dados socioeconômicos revela condições desfavoráveis em certas regiões, afetando a capacidade das comunidades de acessar serviços de saúde adequados. A implementação eficaz de estratégias de controle da TB pode reduzir os custos associados ao tratamento e o impacto socioeconômico negativo da doença. Foram utilizadas ferramentas como TabNet e modelos Tobit e Tobit espacial (SAR-TOBIT) para análise de dados e identificação de padrões geográficos e socioeconômicos. O uso do software R facilitou a implementação eficiente desses modelos, permitindo uma análise detalhada dos resultados. Reconheceu-se a influência dos determinantes sociais da saúde na propagação da TB, destacando a importância de abordagens culturais e educativas para aumentar a conscientização e promover a adesão ao tratamento.

Social, technological, economic and cultural impacts

This study addresses the geographic distribution of tuberculosis (TB) in Minas Gerais, using data from DATASUS and IBGE. Areas with a higher risk of TB transmission were identified, allowing targeted preventive measures, and the socioeconomic disparity between the North and Northeast regions of Minas Gerais was highlighted, highlighting the need for specific interventions. Analysis of socioeconomic data reveals unfavorable conditions in certain regions, affecting communities' ability to access adequate health services. Effective implementation of TB control strategies can reduce costs associated with treatment and the negative socioeconomic impact of the disease. Tools such as TabNet and spatial Tobit and Tobit models (SAR-TOBIT) were used to analyze data and identify geographic and socioeconomic patterns. The use of R software facilitated the efficient implementation of these models, allowing for a detailed analysis of the results. The influence of social determinants of health on the spread of TB was recognized, highlighting the importance of cultural and educational approaches to increase awareness and promote adherence to treatment.

LISTA DE FIGURAS

Figura 4.1 – Mapa das mesorregiões do Estado de Minas Gerais	18
Figura 4.2 – Matriz de proximidade espacial usando como critério a fronteira entre as áreas.	21
Figura 4.3 – Matriz de proximidade espacial usando como critério a fronteira entre as áreas.	21
Figura 4.4 – Diagrama de espalhamento	24
Figura 5.1 – Q-Q Plot	31
Figura 5.2 – Mapa dos quintis com relação a taxa de incidência de Tuberculose nos municípios do estado de Minas Gerais.	32
Figura 5.3 – Mapa dos quintis com relação a IDHM - Desenvolvimento dos municípios do estado de Minas Gerais.	33
Figura 5.4 – Mapa dos quintis com relação a Gini - Desigualdade de renda dos municípios do Estado de Minas Gerais.	34
Figura 5.5 – Mapa dos quintis com relação a proporção de pobres com renda menor a 140 reais mensais dos municípios do Estado de Minas Gerais.	35
Figura 5.6 – Mapa dos quintis com relação a proporção de pessoas que vive em domicílios com água encanada	36
Figura 5.7 – Mapa dos quintis com relação a proporção de pessoas que vive em domicílios com banheiro e água encanada	37
Figura 5.8 – Mapa dos quintis com relação a proporção de pessoas que vive em domicílios com energia elétrica	38
Figura 5.9 – Mapa dos quintis com relação a proporção de pessoas que vive em domicílios com água e esgotamento sanitário inadequados nos municípios do Estado de Minas Gerais.	39
Figura 5.10 – Mapa dos quartis em relação ao Lisa Map das Taxas de Tuberculose	40
Figura 5.11 – Mapa dos quintis com relação ao Moran Map das Taxas de Tuberculose . .	41

LISTA DE TABELAS

Tabela 5.1 – Taxa de Incidência da Tuberculose nos municípios do estado de Minas Gerais no ano de 2019 pra cada 100.000 habitantes.	30
Tabela 5.2 – Estatística Descritiva	30
Tabela 5.3 – Resultados do teste de normalidade Shapiro-Wilk	31
Tabela 5.4 – Tabela das estimativas do Índice de Moran e Estatística de Geary.	39
Tabela 5.5 – Observações censuradas do modelo Tobit.	42
Tabela 5.6 – Resultados da estimação do modelo Tobit.	42
Tabela 5.7 – Resultados do Teste de Normalidade Shapiro-Wilk para os Resíduos.	42
Tabela 5.8 – Resumo do modelo Sar-tobit.	43
Tabela 5.9 – Resultados da estimação do modelo SAR-Tobit.	43
Tabela 5.10 – Resultado do Teste de Normalidade de Shapiro-Wilk.	44
Tabela 5.11 – Comparação dos modelos	44

SUMÁRIO

1	INTRODUÇÃO	13
2	OBJETIVOS	15
2.1	Objetivos Específicos	15
3	REFERENCIAL TEÓRICO	16
3.1	Análise Espacial	16
3.2	Modelagem	17
4	MATERIAL E MÉTODOS	18
4.1	Dados e variáveis	19
4.2	Métodos	19
4.3	Matriz de Vizinhança (Matriz W)	20
4.4	Média Móvel	21
4.5	Autocorrelação Espacial (<i>I</i> de Moran)	22
4.6	Diagrama de Espalhamento	23
4.7	Box Map, Lisa Map e Moran Map	24
4.8	Modelo de Regressão Tobit	25
4.8.1	Censura e Truncamento	27
4.8.2	Métodos de estimação	27
4.8.2.1	Modelo Sar-tobit	27
4.8.3	Comparação dos modelos	28
4.9	Software	29
5	Resultados	30
5.1	Análise Espacial	39
5.2	Modelo de Regressão Tobit	42
5.3	Modelo de Regressão Tobit Espacial	43
6	Conclusão	45
	REFERÊNCIAS	46

1 INTRODUÇÃO

A Tuberculose é uma das doenças infecciosas mais antigas e persistentes da história da humanidade. Existem evidências de que a Tuberculose existe desde os tempos pré-históricos. A doença já foi encontrada em esqueletos de múmias do antigo Egito (3000 A.C). É causada pela bactéria *Mycobacterium tuberculosis*, que afeta principalmente os pulmões, mas pode se manifestar em outros órgãos do corpo e apesar de ser uma doença curável, ela ainda é responsável por um grande número de mortes em todo o mundo. No Brasil, a Tuberculose continua sendo um problema de saúde pública, com uma alta incidência em algumas regiões do país. A Organização Mundial da Saúde (OMS) (Petersen et al., 2022) estima que, em 2020, a TB foi responsável por 1,6 milhões de mortes, tornando-se uma das principais causas de morte por doenças infecciosas no mundo.

A Tuberculose se manifesta em diversas apresentações clínicas (formas) que podem variar de acordo com o local afetado no corpo e a gravidade da doença. As principais formas da Tuberculose incluem: Tuberculose pulmonar - a forma mais comum da doença e afeta os pulmões; Tuberculose extrapulmonar - a infecção ocorre fora dos pulmões e pode afetar vários órgãos e sistemas do corpo como ossos, rins, sistema nervoso central e sistema linfático; Tuberculose latente - a pessoa está infectada pela bactéria da Tuberculose, mas não apresenta sintomas e não é contagiosa; Tuberculose multirresistente (TB-MR) - é uma forma mais grave da doença em que as bactérias são resistentes aos medicamentos antiTuberculose mais comumente utilizados.

A transmissão ocorre principalmente por meio do ar, quando uma pessoa infectada com Tuberculose pulmonar ativa libera pequenas partículas chamadas aerossóis no ambiente ao tossir, espirrar, falar ou cantar. Essas partículas contêm as bactérias da Tuberculose e podem ser inaladas por outras pessoas próximas, especialmente em ambientes fechados e com pouca ventilação, também pode ser transmitida por meio do leite materno, mas é relativamente raro.

Em muitos casos, o sistema imunológico é capaz de controlar a infecção e impedir que a pessoa fique doente. Essa condição é conhecida como Tuberculose latente, no entanto, a Tuberculose latente pode se tornar ativa e causar sintomas no futuro. Fatores como a imunidade do indivíduo e a carga bacteriana do infectado desempenham um papel na probabilidade de infecção.

Os sintomas mais comuns da Tuberculose incluem tosse persistente, que pode durar mais de três semanas, produção de escarro (muco) com possível presença de sangue, dor no

peito, fraqueza, fadiga, perda de peso inexplicada, febre baixa (especialmente à tarde ou à noite) e suores noturnos. Em estágios avançados, a Tuberculose também pode causar falta de ar, falta de apetite e inchaço dos gânglios linfáticos (MOREIRA; KRITSKI; CARVALHO, 2020). E o diagnóstico é feito por meio de exames de imagem, como radiografias e tomografias dos pulmões, e exames laboratoriais, como análise de escarro ou amostras de outros fluidos corporais.

Para prevenir a transmissão da Tuberculose, medidas de controle devem ser adotadas. Isso inclui identificar e tratar precocemente as pessoas com Tuberculose ativa, garantir o acesso a medicamentos adequados, promover a adesão ao tratamento e melhorar as condições de ventilação e higiene em ambientes fechados (SOUZA; SILVA; MEIRELLES, 2010).

Neste sentido, a estatística espacial, por meio da análise de dados de área, desempenha um papel crucial na determinação de localidades com alto risco e incidência de doenças infectocontagiosas. Ademais, esse tipo de variável, em geral, possui estrutura que diverge do padrão de normalidade.

Ao lidar com modelos normais e não normais em dados de área, é comum encontrar fenômenos como autocorrelação espacial, heterogeneidade espacial e variações locais. A autocorrelação espacial refere-se à tendência de observações próximas umas das outras serem mais semelhantes do que observações distantes. A heterogeneidade espacial refere-se à variação na estrutura dos dados em diferentes regiões geográficas. Esses padrões espaciais podem violar a suposição de independência dos dados, fundamental em muitas análises estatísticas tradicionais.

Para lidar com essas questões, os modelos estatísticos espaciais são utilizados. Esses modelos consideram explicitamente a estrutura espacial dos dados, permitindo incorporar a dependência espacial e capturar padrões. Alguns exemplos de modelos estatísticos espaciais incluem o modelo de regressão espacial, modelos de mistura espacial, modelos de autoregressão espacial e krigagem espacial.

Neste sentido, modelos que podem ser aplicados para explorar a relação entre variáveis georreferenciadas e avaliar a influência de fatores espaciais na distribuição dos dados. Além disso, a estatística espacial também pode ser utilizada para mapear padrões espaciais, identificar áreas de alta ou baixa ocorrência de eventos.

Portanto, ao incluir dados de área em análises com modelos não normais, é possível obter *insights* mais precisos sobre a estrutura espacial dos dados, considerar a dependência espacial e melhorar a capacidade de previsão e tomada de decisão em contextos geográficos.

2 OBJETIVOS

O objetivo principal desse estudo é identificar áreas favoráveis à propagação da tuberculose, devido a fatores socioeconômicos.

2.1 Objetivos Específicos

- a) Analisar a estrutura de dependência espacial na incidência da tuberculose;
- b) Mapear a incidência da tuberculose na região de estudo com vista a identificar os locais de maior risco;
- c) Identificar os factores sócio-econômicos e demográficos que afetam a ocorrência da tuberculose

3 REFERENCIAL TEÓRICO

A tuberculose é uma das doenças infecciosas mais antigas e persistentes da história da humanidade. No Brasil, a tuberculose continua sendo um problema de saúde pública, com uma alta incidência em algumas regiões do país (ORGANIZATION et al., 2020). "Apesar de já existirem recursos tecnológicos capazes de promover o controle da tuberculose (TB), ainda não há perspectiva de obter sua erradicação em futuro próximo no Brasil."(VIEGAS et al., 2014).

O tratamento da tuberculose geralmente envolve uma combinação de medicamentos antibióticos específicos, tomados por um período prolongado. Além disso, a prevenção da tuberculose inclui a vacinação com a vacina BCG em países com alta incidência, bem como a adoção de medidas de controle da infecção em ambientes de saúde e comunitários (SOUZA; SILVA; MEIRELLES, 2010).

3.1 Análise Espacial

A Tuberculose quando em conjunto com a análise Espacial apresenta diversas vantagens. Primeiro, ela permite a identificação de áreas de alto risco, direcionando os esforços de prevenção e controle para as regiões mais afetadas (VYNNYCKY; FINE, 1997). Isso é especialmente útil em países com recursos limitados, corrupção e má gestão pública onde é necessário otimizar a utilização dos recursos disponíveis.

Nesse contexto, a análise espacial da tuberculose tem se tornado cada vez mais essencial como uma ferramenta para compreender a distribuição espacial da doença e, consequentemente, aprimorar as estratégias de prevenção e controle. Autores como (CRESSIE, 2015), (DIGGLE, 2003), (ANSELIN; SYABRI; KHO, 2006) e (CÂMARA et al., 2004) dentre outros, fizeram contribuições significativas para o desenvolvimento teórico e prático da estatística espacial. Suas pesquisas e publicações continuam a moldar o campo, exercendo uma influência duradoura.

Estudos recentes realizados em Minas Gerais têm destacado a importância da análise espacial para o controle da tuberculose na região. Um exemplo é o estudo de (PEREIRA; NOGUEIRA; CAMPOS, 2021), que analisou a distribuição espacial da doença em um município de médio porte no estado de Minas Gerais, relacionando-a com indicadores socioeconômicos.

Além disso, a análise espacial pode revelar padrões de disseminação da tuberculose, identificando *clusters* de casos e auxiliando na investigação de surtos (WOOD et al., 2011). Isso possibilita uma resposta rápida e direcionada, visando a interrupção da cadeia de trans-

missão. Em um estudo recente (MACIEL et al., 2019), por exemplo, foram utilizados dados de incidência de tuberculose em São Paulo para identificar aglomerados da doença e avaliar a influência dos fatores socioeconômicos na sua distribuição.

Outra contribuição importante da análise espacial é a identificação de fatores de risco associados à tuberculose. Ao examinar as características geográficas das áreas com maior incidência, é possível identificar correlações com fatores como densidade populacional, condições socioeconômicas, acesso a serviços de saúde, entre outros (LIN et al., 2008). Essas informações são fundamentais para o desenvolvimento de estratégias de prevenção e intervenção mais eficazes.

3.2 Modelagem

Os modelos Tobit e SAR-TOBIT merecem destaque. O modelo Tobit é eficaz em lidar com dados censurados, enquanto o SAR-TOBIT, um avanço incorporando componentes espaciais, se mostra particularmente relevante para análises em que a dependência espacial é um fator crítico. Ambos oferecem abordagens para entender a relação entre a incidência de tuberculose e fatores de risco, alinhando-se com a complexidade dos dados espaciais (ANSELIN; SYABRI; KHO, 2006).

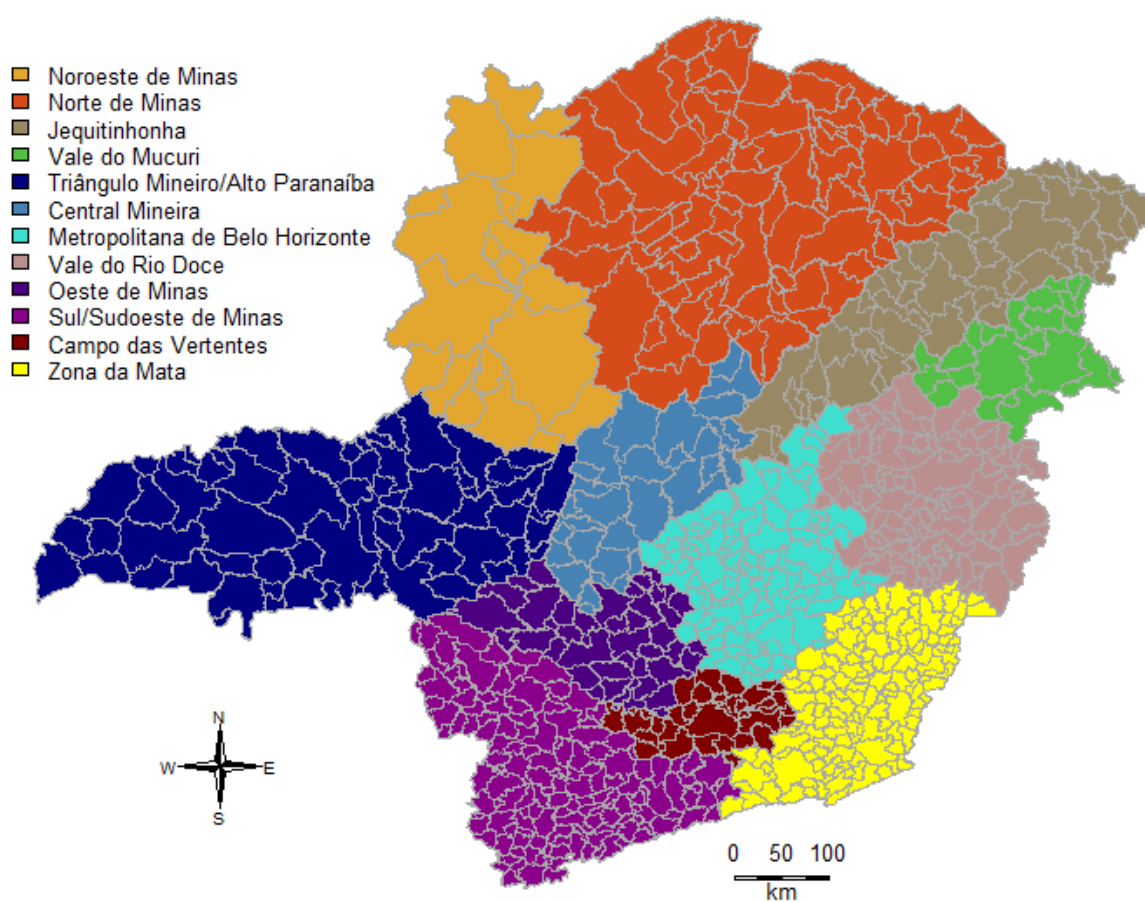
Portanto, a análise espacial da tuberculose aplicado a dados de notificação compulsória tem se mostrado uma ferramenta importante para aprimorar as ações de prevenção e controle da doença como a TB em Minas Gerais e no Brasil como um todo. Através da análise espacial e modelos estatísticos adequados, é possível identificar áreas com maior risco de transmissão da tb e compreender as características socioeconômicas e demográficas associadas à ocorrência da doença.

Em um estudo conduzido por (ARAÚJO; GARCIA, 2019) onde buscou analisar os determinantes da inovação, os padrões espaciais e a dependência espacial da inovação nas microrregiões brasileiras. Para analisar esses fatores, foi estimado um modelo empírico baseado na Função de Produção do Conhecimento (FPC) com dados de patente no Brasil por um Tobit Espacial Autoregressivo (SAR-Tobit). Os resultados indicaram que níveis mais elevados de Pesquisa e Desenvolvimento (P&D) Industrial Regional implicam em maior inovação e que maiores níveis de pesquisa universitária impactam positivamente a inovação na região. Além disso, regiões mais aglomeradas e diversificadas apresentam melhor desempenho inovador.

4 MATERIAL E MÉTODOS

Minas Gerais é uma das 27 unidades federativas do Brasil. Localizada na região Sudeste, faz divisa com o estado do Rio de Janeiro ao sudeste, São Paulo ao sul, Espírito Santo ao leste, Bahia a nordeste, Goiás a noroeste, Mato Grosso do Sul a oeste e Distrito Federal ao norte. Possui uma área total de 586.528,293 km^2 e 853 municípios. Segundo as estimativas do IBGE de 2021, a população do estado era de aproximadamente 21.411.923 habitantes.

Figura 4.1 – Mapa das mesorregiões do Estado de Minas Gerais



Fonte: Peixoto (2023).

4.1 Dados e variáveis

Os dados utilizados foram coletados no *DataSUS*, *site* do Ministério da Saúde disponível em (DATASUS, 2023) para o estado de Minas Gerais, os shapes (polígonos dos mapas) e as variáveis socioeconômicas e sociodemográficas foram retirados da plataforma do IBGE. Com base nos dados coletados, foi calculada a taxa de tuberculose, o resultado obtido fornece a taxa por 100.000 habitantes, utilizando a seguinte fórmula:

$$\frac{\text{casos_notificados}}{\text{pop_estimada_do_ano}} \times 100.000, \quad (4.1)$$

Isso nos permite comparar áreas com diferentes tamanhos populacionais de forma mais justa. Se usássemos apenas os números absolutos de casos, áreas mais populosas tenderiam a ter números maiores, independentemente da gravidade real da situação. As variáveis independentes utilizadas no estudo encontram-se descritas :

- a) IDHM - Índice de Desenvolvimento Humano Municipal;
- b) GINI - Índice de Gini;
- c) PMPOB - Proporção de pobres;
- d) T AGUA - Percentual da população que vive em domicílios com água encanada;
- e) T BANAGUA - Percentual da população que vive em domicílios com banheiro e água encanada;
- f) T LUZ - Percentual da população que vive em domicílios com energia elétrica;
- g) AGUA ESGOTO - % de pessoas em domicílios com abastecimento de água e esgotamento sanitário inadequados;

4.2 Métodos

Neste estudo, empregamos as técnicas da estatística espacial, que visam analisar, caracterizar e modelar variáveis aleatórias com estrutura espacial ou espaço-temporal. O objetivo principal é investigar a presença de um fenômeno específico em um determinado espaço geográfico. Essa abordagem metodológica é amplamente aplicada em diversas áreas do conhecimento, como agronomia, geologia, epidemiologia, entre outras.

Dentre os três tipos de dados georreferenciados frequentemente considerados - processos pontuais, superfície contínua, e dados de área, optamos por utilizar a metodologia de dados de área. É importante ressaltar que essa mesma metodologia foi empregada em estudos anteriores realizados por (COSTA, 2017) e (PEIXOTO et al., 2020).

Posteriormente foi avaliada a estrutura espacial das variáveis socioeconômicas e demográficas por meio de um modelo de regressão não normal espacial que será avaliado possíveis associações das demais variáveis com a dependente por área. As abreviações utilizadas das variáveis socioeconômicas e demográficas estão de acordo com as legendas do censo demográfico 2010 (IBGE, 2020). Os dados da TB foram coletados para o ano de 2019 pelo do site do Ministério da Saúde (Tabnet) (DATASUS, 2023), pois durante a pandemia da COVID-19, diversos países enfrentaram desafios relacionados ao controle de outras doenças de notificação compulsória, como é o caso da tuberculose. Com a sobrecarga dos sistemas de saúde e a necessidade de redirecionar recursos para o enfrentamento da pandemia, é possível que haja uma subnotificação de casos de tuberculose nos anos de 2020, 2021 e 2022. Por esse motivo será analisado apenas 2019. Além disso, medidas de distanciamento social podem ter dificultado o acesso aos serviços de saúde e a realização de diagnósticos e tratamentos.

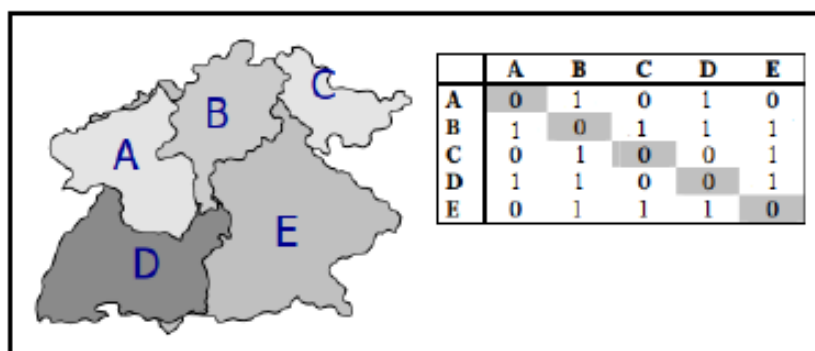
4.3 Matriz de Vizinhança (Matriz $\mathbf{W}$)

A associação de observações na vizinhança definido para cada localização pode ser expressa por meio de uma contiguidade ou pesos espaciais numa matriz $\mathbf{W}$ de ordem $n \times n$, também conhecida como matriz de distância ou matriz de proximidade. A matriz de vizinhança é definida como,

$$\mathbf{W} = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1n} \\ w_{21} & w_{22} & \dots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \dots & w_{nn} \end{bmatrix},$$

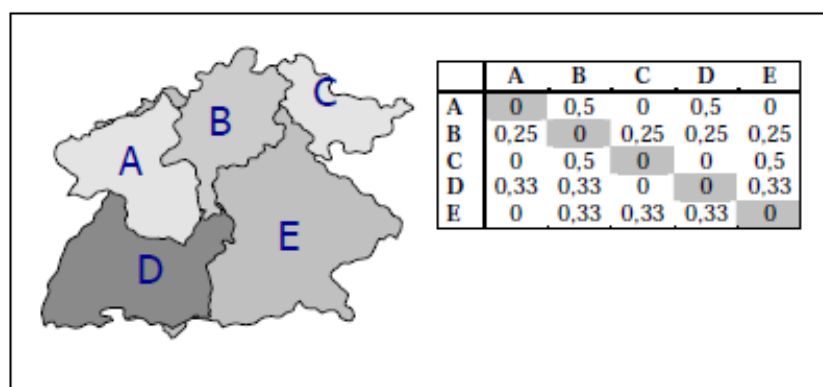
onde n representa o número de locais (observações). A entrada na i -ésima linha e j -ésima coluna, denotado como w_{ij} , corresponde ao par (i, j) de localizações. Os elementos da matriz w_{ij} assumem um valor diferente de zero quando as áreas (observações) i e j são consideradas vizinhas, e zero caso contrário (DUNCAN et al., 2014) conforme (Figura 4.1). Após o preenchimento da matriz faz-se necessário normalizar suas linhas de maneira que a soma dos pesos de cada linha seja igual a 1 (CÂMARA et al., 2004) conforme a (Figura 4.2).

Figura 4.2 – Matriz de proximidade espacial usando como critério a fronteira entre as áreas.



Fonte: CÂMARA et al. (2004).

Figura 4.3 – Matriz de proximidade espacial usando como critério a fronteira entre as áreas.



Fonte: CÂMARA et al. (2004).

4.4 Média Móvel

Para analisar a variação de tendência espacial da taxa de pessoas com tuberculose no estado de Minas Gerais, pode-se utilizar a média móvel local para cada município.

Uma forma simples de explorar a variação da tendência espacial dos dados é calcular a média dos valores dos vizinhos. Isto reduz a variabilidade espacial, pois a operação tende a produzir uma superfície com menor flutuação que os dados originais. A média móvel $\hat{\mu}_i$ associada ao atributo y_i , relativo a i -ésima área, pode ser calculada a partir dos elementos w_{ij} da matriz normalizada de proximidade espacial de primeira ordem, isso é a soma dos pesos de cada linha seja igual a 1 (CÂMARA et al., 2004). Isto simplifica muitos cálculos de índices de autocorrelação espacial, sendo definida como:

$$\hat{\mu}_i = \sum_{j=1}^n \sum_{i=1}^n w_{ij}(y_j - \bar{y}), \quad (4.2)$$

sendo $\hat{\mu}_i$ as médias ponderadas, w_{ij} elementos da matriz da matriz de proximidade espacial normalizada $\mathbf{w}$, y_j é o valor do atributo em cada área e n é o número de polígonos (áreas) e $\bar{y}$ é a média.

4.5 Autocorrelação Espacial (*I* de Moran)

A autocorrelação espacial (associação) é a correlação entre as observações de uma única variável estritamente atribuível à proximidade destas observações no espaço geográfico. As medidas de autocorrelação espacial lidam com a correlação entre as observações vizinhas de uma determinada variável, por exemplo a autocorrelação espacial do número de indivíduos com tuberculose nos municípios do estado de Minas Gerais

As medidas e testes de autocorrelação espacial podem ser diferenciadas pelo alcance ou escala de análise. Geralmente distingue-se entre as medidas globais e locais:

- a) Medidas Globais: Implica que todos os elementos na matriz $\mathbf{W}$ estão incluídas no cálculo da correlação espacial, produzindo um valor de autocorrelação espacial para qualquer matriz de pesos espaciais.
- b) Medidas Locais: Avaliam a autocorrelação espacial associada com uma ou mais unidades de áreas particulares. Ambas indicam o grau de associação espacial do conjunto de dados.

O índice *I de Moran* calcula a autocorrelação espacial como uma correlação a partir do produto dos desvios em relação à média. Esse índice indica a magnitude da associação espacial presente no conjunto de dados com n localizações, tendo como hipótese nula a independência espacial, dada pela expressão:

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{(\sum_{i=1}^n \sum_{j=1}^n w_{ij}) \sum_{i=1}^n (y_i - \bar{y})^2}, \quad (4.3)$$

em que irá variar no domínio dos reais mas, em geral, fica no intervalo de -1 e 1, em que -1 significa dissimilaridade, 0 representa um comportamento aleatório e 1 significa similaridade entre observações vizinhas.

Ainda que essa medida global seja muito útil para fornecer uma indicação de agrupamento global dos dados, esta medida precisa ser complementada por estatísticas locais. A fórmula para o cálculo do índice *I de Moran Local* para cada área A_i é dada por:

$$I_i = \frac{(y_i - \bar{y}) \sum_{j=1}^n \sum_{j=1}^n w_{ij} (y_j - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^n}, \quad (4.4)$$

onde I_i é o índice de Moran para a unidade espacial i , que mede a autocorrelação espacial da variável em estudo nessa unidade, y_i representa o valor da variável em estudo na unidade espacial i , $\bar{y}$ é a média dos valores da variável em estudo em todas as unidades espaciais, n representa o número total de unidades espaciais, j índice de iteração utilizado na fórmula, w_{ij} representa o peso espacial entre as unidades i e j , indicando a força da relação espacial entre elas, $(y_j - \bar{y})$ é a diferença entre o valor da variável na unidade j e a média dos valores da variável em todas as unidades espaciais, $(y_i - \bar{y})$ é a diferença entre o valor da variável na unidade i e a média dos valores da variável em todas as unidades espaciais, $(y_i - \bar{y})^n$ é o desvio padrão dos valores da variável em todas as unidades espaciais, elevado à potência de n onde n é o número total de unidades espaciais. E a estatística pode ser interpretada da seguinte forma: valores positivos de I_i significam que existem grupos espaciais com valores similares (alto ou baixo) da variável em estudo, valores negativos significam que existem grupos espaciais com valores dissimilares da variável entre as áreas e seus vizinhos.

4.6 Diagrama de Espalhamento

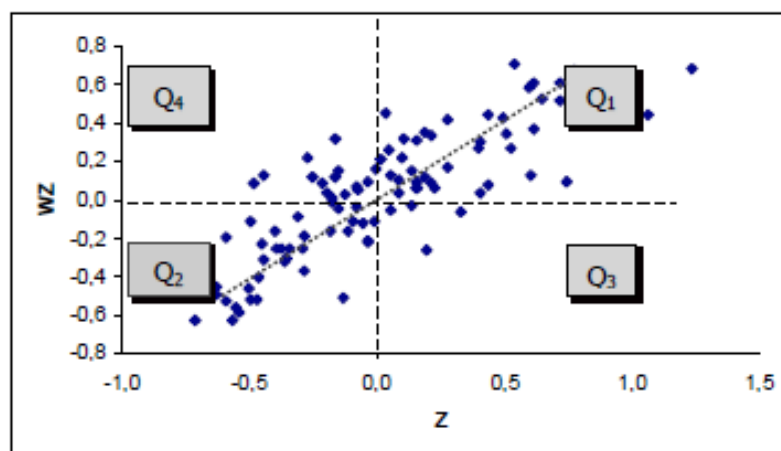
O diagrama de espalhamento de Moran é uma maneira adicional de visualizar a dependência espacial. Construído com base nos valores normalizados (valores de atributos subtraídos de sua média e divididos pelo desvio padrão), permite analisar comportamento da variabilidade espacial. Compara-se os valores normalizados do atributo numa área com a média dos seus vizinhos construindo um gráfico de Z (valores normalizados) por $\mathbf{WZ}$ (média dos vizinhos), que é dividido em quatro quadrantes (CÂMARA; MONTEIRO, 2001). Os quadrantes podem ser interpretados como:

- a) **Q1: Primeiro quadrante** (valores positivos, médias positivas) classificado como alto-alto (AA, + +);
- b) **Q2: Segundo quadrante** (valores negativos, médias negativas) classificado como baixo-baixo (BB, - -);
- c) **Q3: Terceiro quadrante** (valores positivos, médias negativas) classificado como alto-baixo (AB, + -);

d) **Q4: Quarto quadrante** (valores negativos, médias positivas) classificado como baixo alto (BA, - +).

Os municípios localizados nos quadrantes Q1 e Q2 indicam pontos de associação espacial positiva, no sentido que um município possui vizinhos com valores semelhantes. E municípios localizados nos quadrantes Q3 e Q4 indicam pontos de associação espacial negativa, no sentido que um município possui vizinhos com valores distintos como ilustra a Figura 4.4.

Figura 4.4 – Diagrama de espalhamento



Fonte: CÂMARA et al. (2004).

4.7 Box Map, Lisa Map e Moran Map

Os mapas são uma maneira simples de visualização e na estatística espacial podemos utilizar de três principais mapas (PEIXOTO et al., 2020), sendo eles:

- a) **Box Map:** É uma técnica que consiste em dividir a região de estudo em caixas de tamanho uniforme e plotar as médias dos valores de cada caixa em um mapa. Essa técnica é útil para visualizar padrões espaciais globais em dados de área. Por exemplo, se os valores dos dados estiverem agrupados em algumas caixas, enquanto outras caixas tiverem valores mais baixos, isso indicaria um padrão global de aglomeração;
- b) **LISA Map ou Local Indicators of Spatial Association Map:** É uma técnica que permite identificar padrões espaciais locais. Utiliza a estatística LISA para avaliar a associação espacial entre os valores de uma variável em diferentes locais. A técnica é útil para identificar agrupamentos locais de valores altos ou baixos em uma região de estudo, permitindo a identificação de áreas de maior ou menor incidência;

- c) Moran Map: É uma técnica que utiliza o coeficiente de correlação espacial de Moran para avaliar a autocorrelação espacial dos dados. O Moran Map é útil para identificar padrões globais de associação espacial entre as variáveis. A técnica avalia a similaridade dos valores de cada local com os valores de seus vizinhos, permitindo identificar áreas onde a variável tem valores altos ou baixos e, conseqüentemente, onde a doença tem maior ou menor incidência.

Cada uma dessas técnicas tem suas vantagens e desvantagens e a escolha da técnica mais adequada depende do objetivo da análise e das características dos dados. A análise estatística espacial é uma área em constante evolução e novas técnicas e métodos estão sendo desenvolvidos para lidar com os desafios da análise de dados espaciais e fornecer insights úteis para a saúde pública e outras áreas.

4.8 Modelo de Regressão Tobit

O Modelo Tobit é um modelo de regressão linear usado quando algumas observações da variável resposta encontram-se incompletas devido a algum tipo de censura. Porém existem duas causas para ocorrência de tal evento, quando ocorre truncamento e quando ocorre censura. Primeiramente iremos esclarecer a diferença entre esses dois casos.

Quando se trabalha com dados espaciais, é comum encontrar distribuições não normais, o que pode dificultar a aplicação de modelos estatísticos tradicionais. No entanto, existem modelos espaciais que podem ser utilizados para estes tipos de dados.

Segundo (LESAGE; PACE, 2009) o modelo Tobit espacial é um dos modelos que podem ser incorporados para essa situação, que é uma extensão do modelo Tobit clássico que leva em consideração a presença de dependência espacial nos dados e também a possibilidade de truncamento das variáveis contínuas. O modelo tobit, desenvolvido por (TOBIN, 1958), é comumente usado quando temos variáveis dependentes que são censuradas e truncadas, ou seja, quando elas não são completamente observadas, mas apenas parcialmente, embora o modelo Tobit seja baseado na suposição de normalidade dos erros, ele ainda pode ser aplicado a dados com distribuições não normais, desde que a censura ou truncamento ocorra de forma apropriada.

No modelo Tobit espacial, além da presença do truncamento nas variáveis dependentes, também levamos em conta a influência da dependência espacial entre as observações. Isso significa que a probabilidade de observar um determinado valor da variável dependente é influenciada não apenas pelas características individuais dessa observação, mas também pelas características das observações vizinhas.

A equação estrutural do modelo Tobit pode ser expressa da seguinte forma:

$$y^* = X'\beta + \varepsilon \quad (4.5)$$

sendo,

y^* representa a variável latente não observada;

X' é o transposto da matriz X , $K \times n$ de variáveis explicativas;

β é o vetor $K \times 1$ de parâmetros;

ε é o vetor $n \times 1$ de erros aleatórios normalmente distribuídos.

A partir do modelo genérico, a especificação do modelo Tobit pressupõe que a variável latente y^* seja observada somente quando um valor limite "c" é excedido. Dessa forma, a variável dependente é expressa da seguinte maneira:

$$y = \begin{cases} y_1^*, & \text{se } y^* > c \\ 0, & \text{caso contrario} \end{cases} \quad (4.6)$$

Fonte: (AMARAL; ANSELIN, 2013)

A variável y^* é então transformada em y , a variável observada. Se y^* for maior que c , então y assume os valores de y_1^* . Caso contrário, y assume o valor de 0. Isso representa o truncamento nos dados, onde valores abaixo de um certo limite são considerados como truncados.

E com o objetivo de capturar os efeitos espaciais a equação 4.7 representa a estrutura do modelo Tobit espacial, e segundo (AMARAL; ANSELIN, 2013) introduzimos a dependência espacial na variável latente dependente. Onde a dependência espacial é levada em consideração através da matriz de ponderação espacial S e o truncamento nos dados é tratado através da transformação da variável latente y^* em y . Nesse sentido, a variável latente dependente é agora representada da seguinte forma:

$$y^* = S^{-1}X'\beta + S^{-1}\varepsilon$$

sendo

$$(4.7)$$

$$S = I_n - \rho W$$

A variável dependente é denotada como y^* , que é uma variável latente não observada. Essa variável latente é relacionada com as variáveis independentes X através do vetor de coeficientes β . A expressão $S^{(-1)}X'\beta$ representa o efeito das variáveis independentes X sobre a variável

latente y^* e a matriz $S^{(-1)}$ é o inverso da matriz S , que é a matriz de ponderação espacial. sendo S definido como $\mathbf{I}n - \rho W$, e $\mathbf{I}n$ é uma matriz identidade de ordem n , e ρ é o parâmetro de autocorrelação espacial. W é uma matriz de pesos espaciais que descreve a relação espacial entre as observações. A expressão $S^{(-1)}\varepsilon$ representa o erro residual do modelo, sendo ε é um vetor de erros aleatórios. A combinação linear $S^{(-1)}\varepsilon$ é adicionada ao efeito das variáveis independentes para formar a variável latente y^* .

4.8.1 Censura e Truncamento

A censura ocorre quando, para determinadas observações da amostra, os dados relacionados à variável resposta não estão totalmente disponíveis devido a limitações, mas, ao contrário do truncamento, as informações sobre as variáveis independentes estão disponíveis. A censura representa essencialmente uma lacuna na coleta da amostra, seja devido à falta de informações ou à impossibilidade de medir as observações devido a restrições específicas. Nesse caso, os valores dentro dessa área de limitação são relatados com um mesmo valor.

O truncamento ocorre (principalmente) quando um estudo é conduzido com base em um conjunto de dados previamente utilizado, do qual são extraídas apenas as informações desejadas. Isso resulta na indisponibilidade de algumas observações tanto na variável resposta quanto nas variáveis independentes.

4.8.2 Métodos de estimação

A estimação dos parâmetros do modelo Tobit é frequentemente realizada utilizando o método da Máxima Verossimilhança. Esse método busca encontrar os valores dos parâmetros que maximizam a probabilidade de observar os dados truncados e não truncados dados o modelo estatístico adotado.

4.8.2.1 Modelo Sar-tobit

O procedimento é análogo ao modelo linear, mas como existem particularidades existentes no Modelo Tobit, alguns cálculos precisam ser adaptados ou feitos com ajuda computacional.

A seguir vemos o caso da obtenção da função de verossimilhança, para truncamento no valor c :

$$L(\theta) = \prod_{i=1}^n f(y_i|x_i, \theta) I_i F(c|x_i, \theta)^{1-I_i}, \quad (4.2)$$

onde $I_i = 1$ quando a observação está com o valor exato e $I_i = 0$ quando a observação está truncada à esquerda. θ representa o vetor de parâmetros β e σ^2 .

A função de log-verossimilhança para a regressão truncada à esquerda é dada por:

$$\ln L = \sum_{i:y_i > c} \left(-\frac{1}{2} \log(2\pi) + \ln \sigma^2 + \frac{(y_i - x_i^T \beta)^2}{\sigma^2} \right) + \sum_{i:y_i < c} \ln (F(x_i^T \beta / \sigma)), \quad (4.3)$$

Pode-se ver que a primeira parte da função é semelhante à verossimilhança para as observações que não estão presentes no truncamento, e a segunda se refere às probabilidades de truncamento das observações truncadas. Esse resultado se aplica apenas a truncamento à esquerda.

4.8.3 Comparação dos modelos

O Erro Quadrático Médio (EQM) e o Erro Absoluto Médio (EAM) são duas métricas comumente usadas para avaliar a precisão de modelos. O EQM é a média dos quadrados dos erros e o EAM é a média dos valores absolutos dos erros. Os "erros" neste caso são a diferença entre o valor real e o valor previsto. A fórmula para o EQM é:

$$EQM = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

em que:

n é o número total de observações.

y_i é o valor real da i -ésima observação.

$\hat{y}_i$ é o valor previsto para a i -ésima observação

E a fórmula para o EAM é:

$$EAM = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

em que:

n é o número total de observações.

y_i é o valor real da i -ésima observação.

$\hat{y}_i$ é o valor previsto para a i -ésima observação.

Ambas as métricas fornecem uma medida de quão perto as previsões do modelo estão dos valores reais.

4.9 Software

O software utilizado para realizar as análises neste trabalho é o (R Core Team, 2023), com as bibliotecas `spdeep`, `geobr`, `raster`, `maptools`, `spatialprobit`, `censreg`, entre outros. A versão específica utilizada foi a 4.3.1.

5 RESULTADOS

Inicialmente elaborou-se uma tabela com os dados dos municípios com as maiores taxas de incidência da TB, que leva em consideração os casos notificados e a base populacional de referência. Sendo assim, de acordo com a tabela 5.1, pode-se observar os municípios com maiores taxas de incidência por TB no ano de 2019. O fato de que a transmissão tem padrão aéreo por meio de gotículas, localidades com elevada densidade populacional são mais propícias ao desenvolvimento da infecção que pode justificar o elevado número de casos nos municípios que se encontram próximos da capital Belo horizonte, por ter um elevado número de habitantes.

Tabela 5.1 – Taxa de Incidência da Tuberculose nos municípios do estado de Minas Gerais no ano de 2019 pra cada 100.000 habitantes.

Municípios	Taxa de Incidência	População
Juramento	6811,359963	4.331
Consolação	3982,05272	1.783
Passabem	3032,140691	1.649
Turvolândia	1111,111111	5.040
São José do Mantimento	931,565747	2.791
Ubaporanga	825,9161254	12.471
Divinolândia de Minas	594,3732664	7.571
⋮	⋮	⋮

Fonte: Peixoto (2023).

Dos 853 municípios, 487 tiveram pelo menos 1 notificação, isso indica que os restantes 366 municípios não apresentaram nenhum caso notificado. Depois foram realizadas análises estatísticas descritivas dos dados coletados.

Tabela 5.2 – Estatística descritiva da taxa de incidência de TB

Estatísticas	Taxa
Nº. observações	853
Mínimo	0
Máximo	6811,36
Média	42,56
Mediana	6,91
Variância	89032,49
Desvio Padrão	298,38

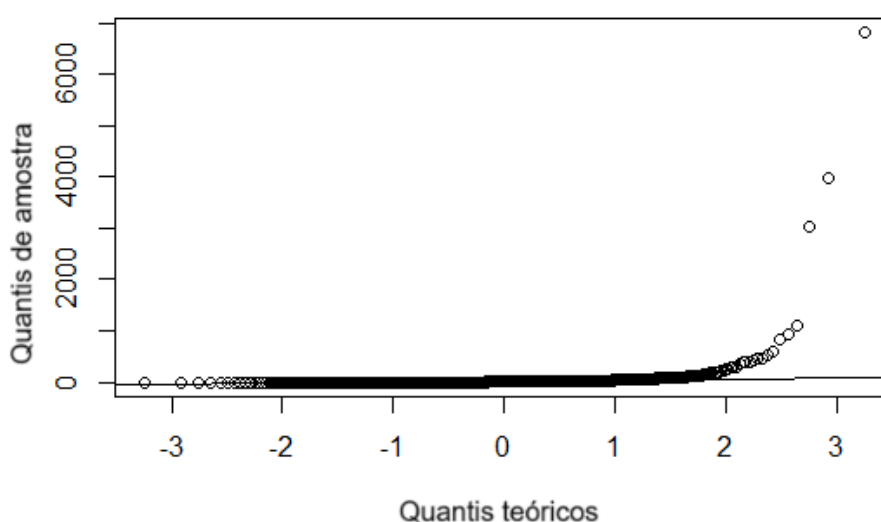
Fonte: Peixoto (2023).

Como pode-se observar na tabela 5.2, existe uma variabilidade considerável nos valores da taxa de incidência entre os municípios analisados. Essa variabilidade é evidenciada pelo

valor mínimo de 0, indicando a ausência de notificações em certas áreas, e pelo valor máximo de 6811.36, apontando para a existência de municípios com taxas de incidência excepcionalmente altas. Além disso, a média de 42.56 e a mediana de 6.91 mostram que a maioria dos municípios apresenta uma taxa de incidência relativamente baixa, embora existam casos com taxas mais altas. A ampla dispersão dos valores em relação à média, conforme indicado pela variância e pelo desvio padrão, ressalta a importância de considerar a variabilidade regional na análise da taxa de incidência.

No segundo momento, foi realizado teste para investigar a normalidade dos dados e de sua distribuição, sendo realizados por dois métodos: o QQ plot e o teste de Shapiro-Wilk. Figura 5.1, Tabela 5.3.

Figura 5.1 – Q-Q Plot



Fonte: Peixoto (2023).

Tabela 5.3 – Resultados do teste de normalidade Shapiro-Wilk

Estatística W	Valor - p
0,095624	< 0,05*

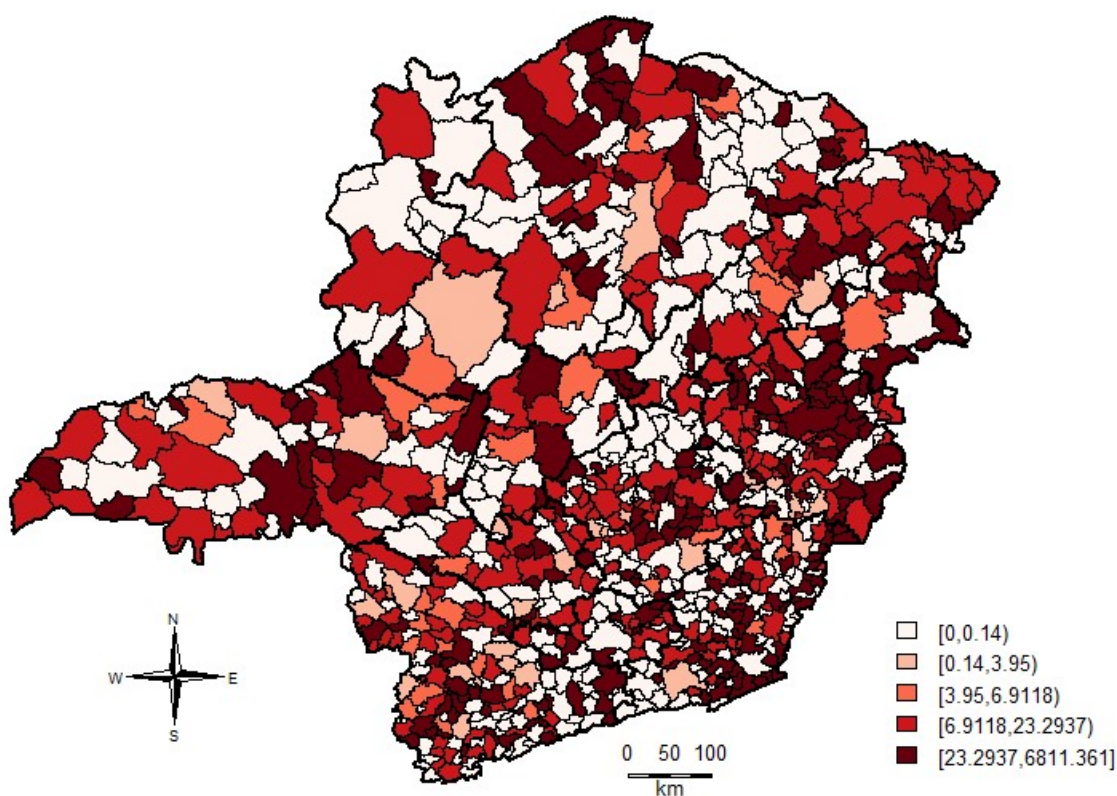
Fonte: Peixoto (2023).

Ao analisar o QQ plot, observa-se que os pontos se curvam para cima em relação à linha de referência, indicando caudas mais pesadas na distribuição dos dados do que o padrão normal. Esse indício motivou a realização do teste de Shapiro-Wilk para obter uma confirmação estatística mais precisa sobre a normalidade, assegurando, assim, uma análise confiável. Ao examinar o teste de Shapiro, percebe-se um valor-p menor que o nível de significância 0.05,

fornecendo evidências suficientes para rejeitar a hipótese nula de normalidade e indicando que os dados não seguem uma distribuição normal.

Por meio da inspeção visual da Figura 5.2, é possível identificar as áreas de maior incidência de Tuberculose (TB) em tons mais escuros e as áreas sem incidência em tons mais claros no primeiro quintil. De modo geral, o estado exibe aglomerados de municípios com altas e baixas taxas, sugerindo uma distribuição aparentemente aleatória. Contudo, é importante apontar que as maiores taxas estão concentradas nos municípios descritos pela tabela 5.1.

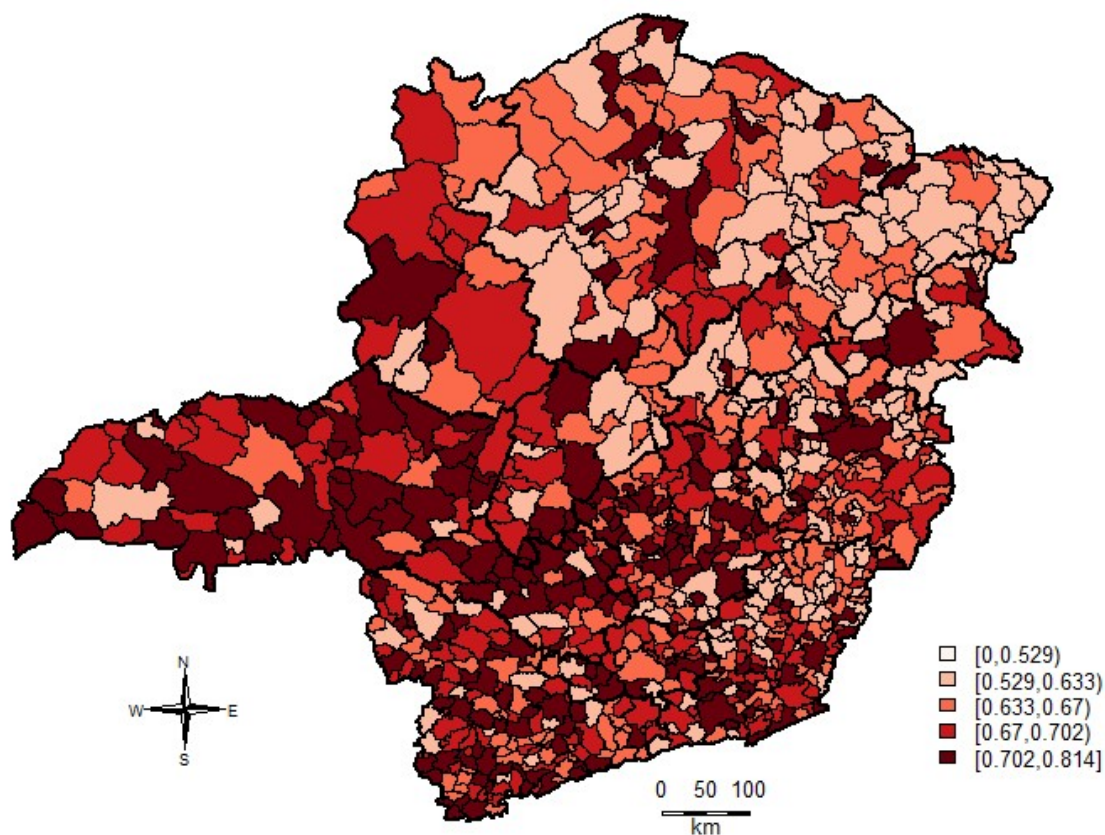
Figura 5.2 – Mapa dos quintis com relação a taxa de incidência de Tuberculose nos municípios do estado de Minas Gerais.



Fonte: Peixoto (2023).

Após analisar as características do mapa dos quintis da taxa de incidência de tuberculose, prosseguimos com a mesma análise para as variáveis explicativas.

Figura 5.3 – Mapa dos quintis com relação a IDHM - Desenvolvimento dos municípios do estado de Minas Gerais.



Fonte: Peixoto (2023).

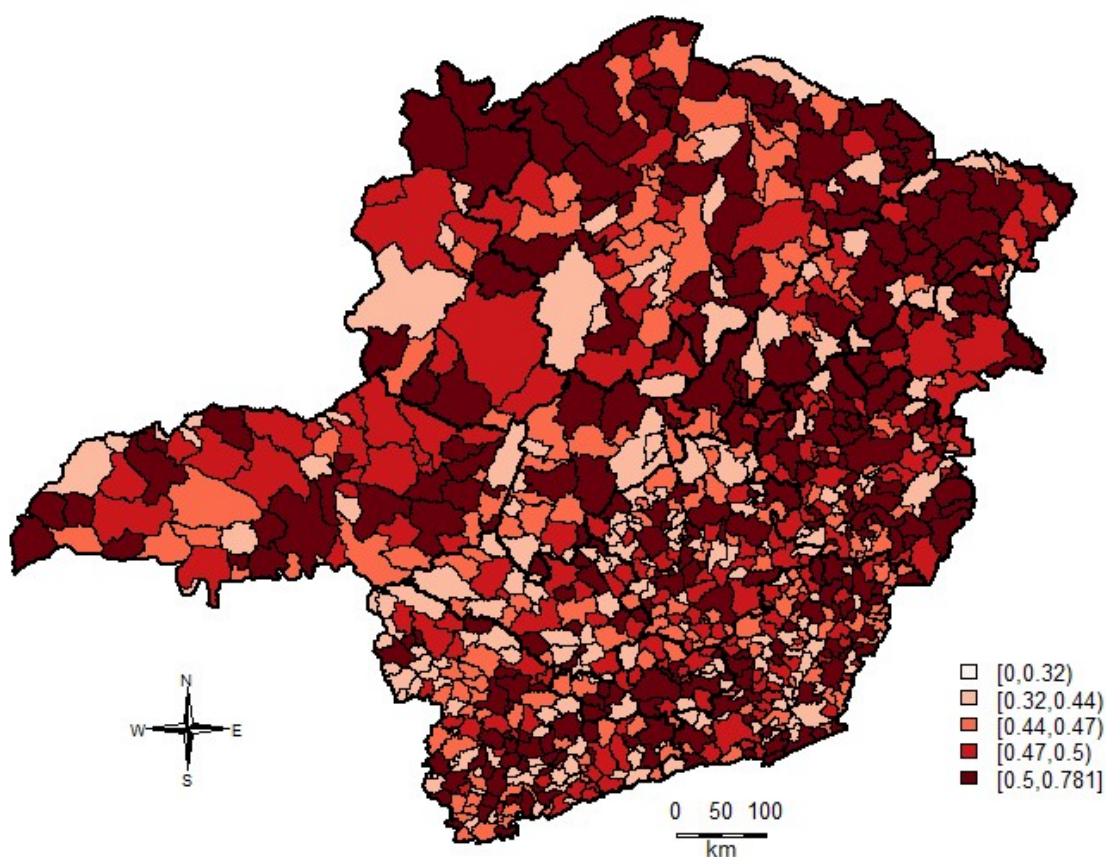
Com base na análise do mapa de Índice de Desenvolvimento Humano Municipal (IDHm), é possível observar que a região sul do estado de Minas Gerais apresenta uma maior concentração de municípios que são considerados mais desenvolvidos. O IDHm é um indicador que avalia o desenvolvimento humano com base em fatores como saúde, educação e renda. Isso sugere que essa região possui um maior acesso a serviços básicos, melhores condições socioeconômicas e maior nível educacional em comparação com outras áreas do estado.

Essa concentração pode ser atribuída a diversos fatores, como investimentos em infraestrutura, presença de indústrias e empresas, universidades, acesso a recursos naturais, entre outros. Esses fatores contribuem para um ambiente propício ao crescimento econômico e ao desenvolvimento humano.

De acordo com (LOPES, 2022) "Os resultados alcançados na análise exploratória de dados espacial (AEDE) do Índice de Desenvolvimento Humano Municipal (IDHM) e o Índice de Vulnerabilidade Social (IVS) e suas dimensões indicaram que, de modo geral, houve avanços significativos nos indicadores nos municípios, no entanto, esses avanços não se deram de forma

igual para todos. A estatística espacial do *I* de Moran Local corroborou a AEDE, mostrando disparidades no território mineiro. Nas regiões Norte de Minas, Vales do Jequitinhonha e do Mucuri e leste formaram clusters de municípios com valores baixos do IDHM e valores altos do IVS. Noutro extremo, o triângulo, Sul de Minas, parte Central Mineira e da Zona da Mata formaram clusters de situação opostas a esses indicadores."

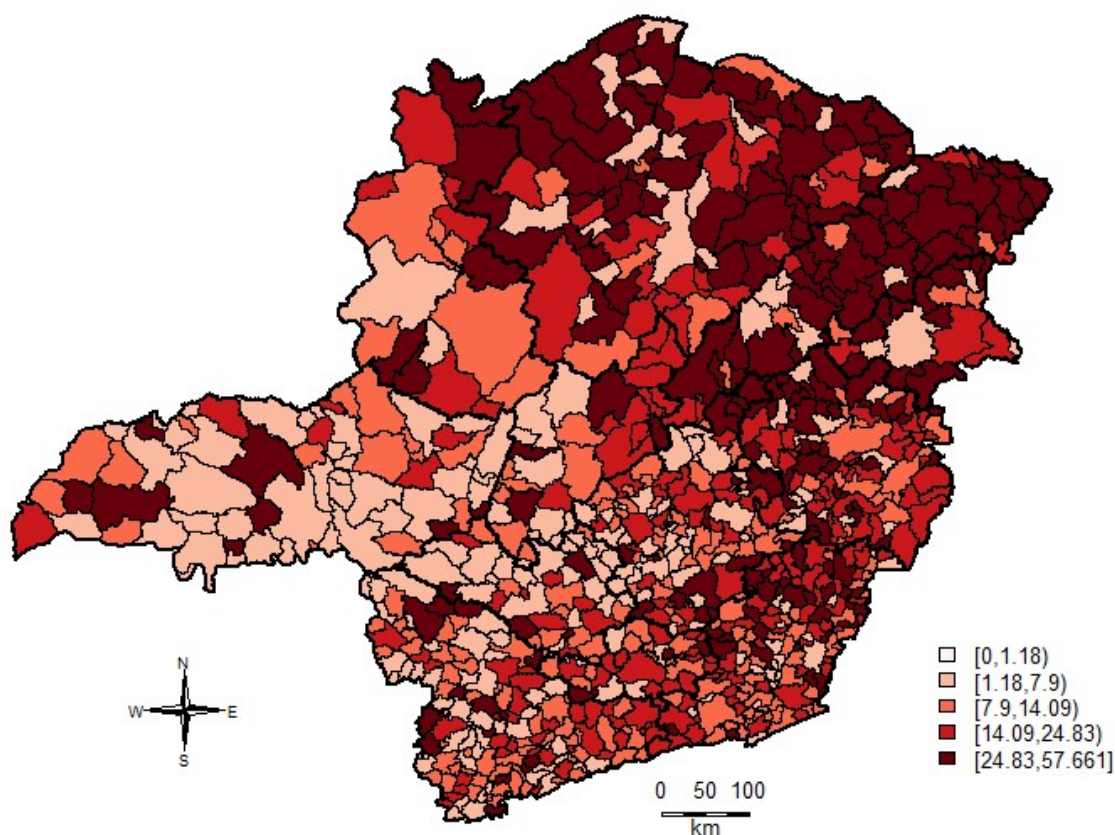
Figura 5.4 – Mapa dos quintis com relação a Gini - Desigualdade de renda dos municípios do Estado de Minas Gerais.



Fonte: Peixoto (2023).

Observa-se que em todo o estado de Minas Gerais, identificamos a presença de clusters de municípios com índice de Gini alto no entanto, é notável uma concentração mais significativa desses clusters na região norte. Essa concentração na região norte sugere a existência de desigualdades socioeconômicas mais acentuadas nessa área. Possíveis fatores que contribuem para essa situação podem incluir condições econômicas desfavoráveis, menor acesso a recursos e oportunidades, bem como desafios relacionados à infraestrutura e serviços básicos.

Figura 5.5 – Mapa dos quintis com relação a proporção de pobres com renda menor a 140 reais mensais dos municípios do Estado de Minas Gerais.



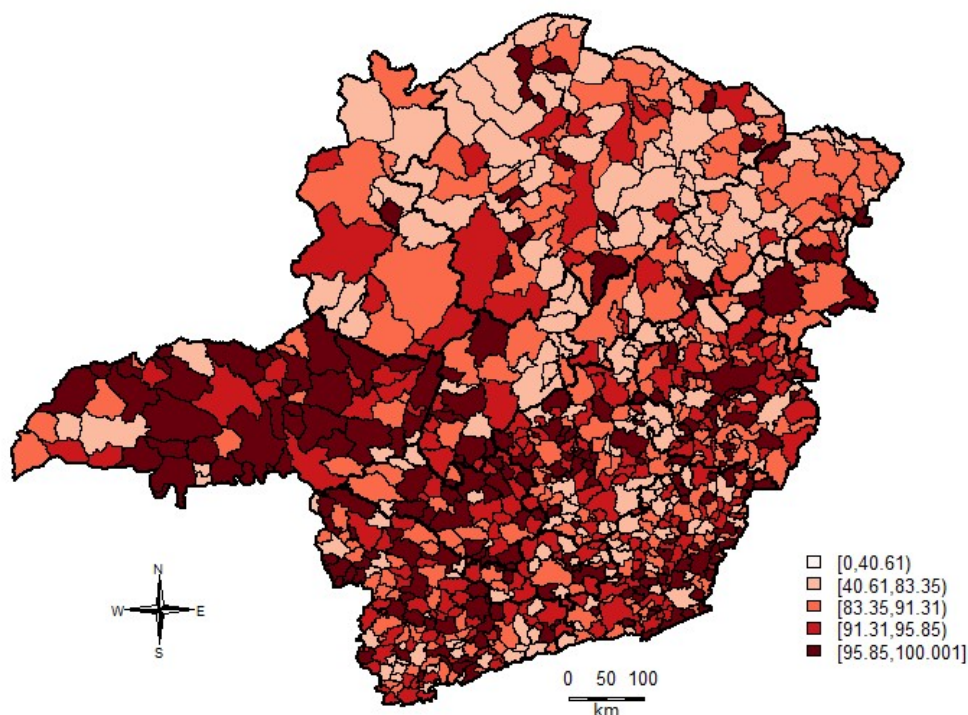
Fonte: Peixoto (2023).

Ao analisar a proporção de pobres da Figura 5.5, é notável uma maior concentração de pessoas em situação de extrema pobreza, definida como aquelas que recebem menos de 140 reais mensais, na região norte e nordeste.

Vários fatores podem contribuir para essa concentração de extrema pobreza na região. Entre eles, podemos citar a falta de acesso a oportunidades econômicas, como empregos formais e fontes de renda estáveis, além da carência de políticas de inclusão social e programas de combate à pobreza.

A ausência de oportunidades econômicas pode ser resultado da falta de investimentos na região, a escassez de indústrias e empreendimentos, além da dificuldade de acesso aos mercados e a serviços básicos de qualidade. A precariedade das condições de infraestrutura, como transporte, saneamento básico e educação, também contribui para a perpetuação da pobreza nessa região.

Figura 5.6 – Mapa dos quintis com relação a proporção de pessoas que vive em domicílios com água encanada.



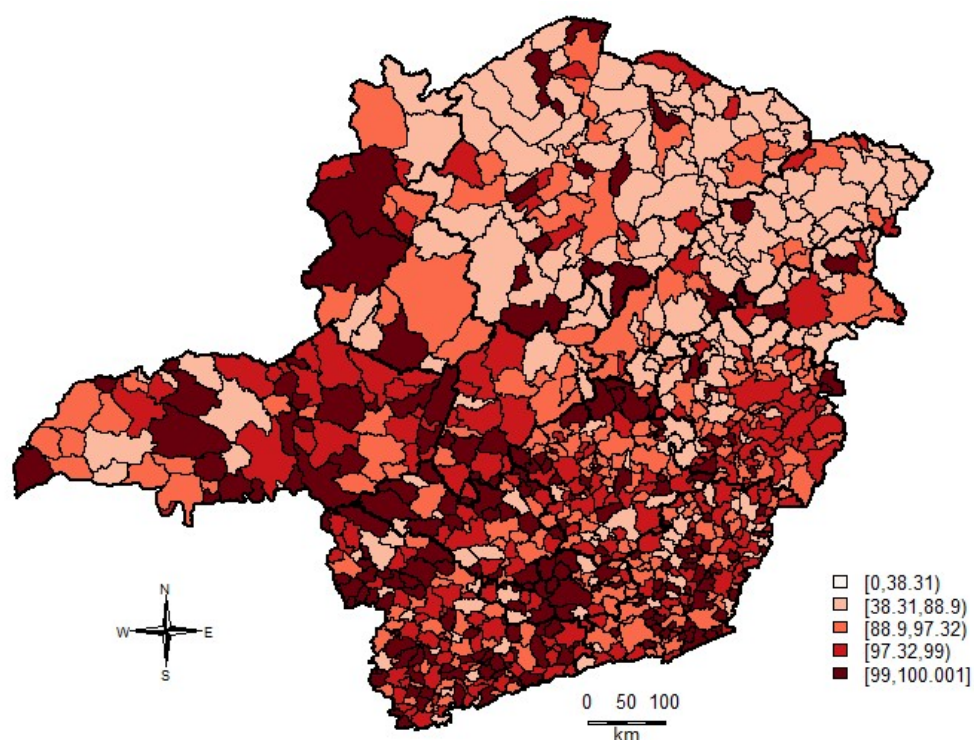
Fonte: Peixoto (2023).

Ao considerarmos as variáveis "proporção de pessoas que vivem em domicílios com água encanada" e "proporção de pessoas que vivem em domicílios com banheiro e água encanada", podemos chegar a uma interpretação semelhante àquela evidenciada nos mapas anteriores. De fato, é perceptível que a região sul do estado de Minas Gerais possui um maior acesso a esses benefícios em comparação com a região norte.

A análise dessas variáveis revela que, na região sul, uma proporção maior da população reside em domicílios com água encanada e com banheiro, indicando melhores condições de saneamento básico e infraestrutura. Isso sugere que os serviços de abastecimento de água e saneamento estão mais amplamente disponíveis e acessíveis na região sul, contribuindo para uma melhor qualidade de vida e bem-estar dos moradores.

Por outro lado, a região norte mostra uma proporção menor de pessoas vivendo em domicílios com água encanada e banheiro. Isso pode ser resultado da falta de investimentos em infraestrutura, limitações geográficas ou dificuldades socioeconômicas específicas dessa região. Essa disparidade na disponibilidade de água encanada e infraestrutura sanitária entre as regiões sul e norte de Minas Gerais evidencia a existência de desigualdades regionais em termos de

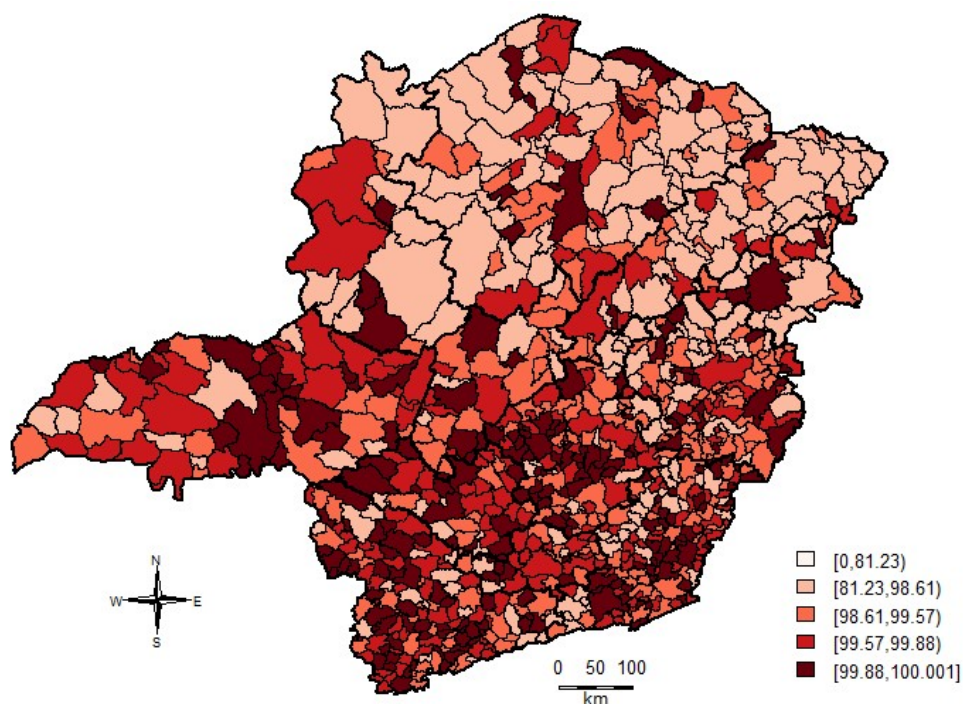
Figura 5.7 – Mapa dos quintis com relação a proporção de pessoas que vive em domicílios com banheiro e água encanada.



Fonte: Peixoto (2023).

acesso aos serviços básicos. É importante destacar que o acesso adequado a esses recursos é essencial para a promoção da saúde, higiene e qualidade de vida da população.

Figura 5.8 – Mapa dos quintis com relação a proporção de pessoas que vive em domicílios com energia elétrica.

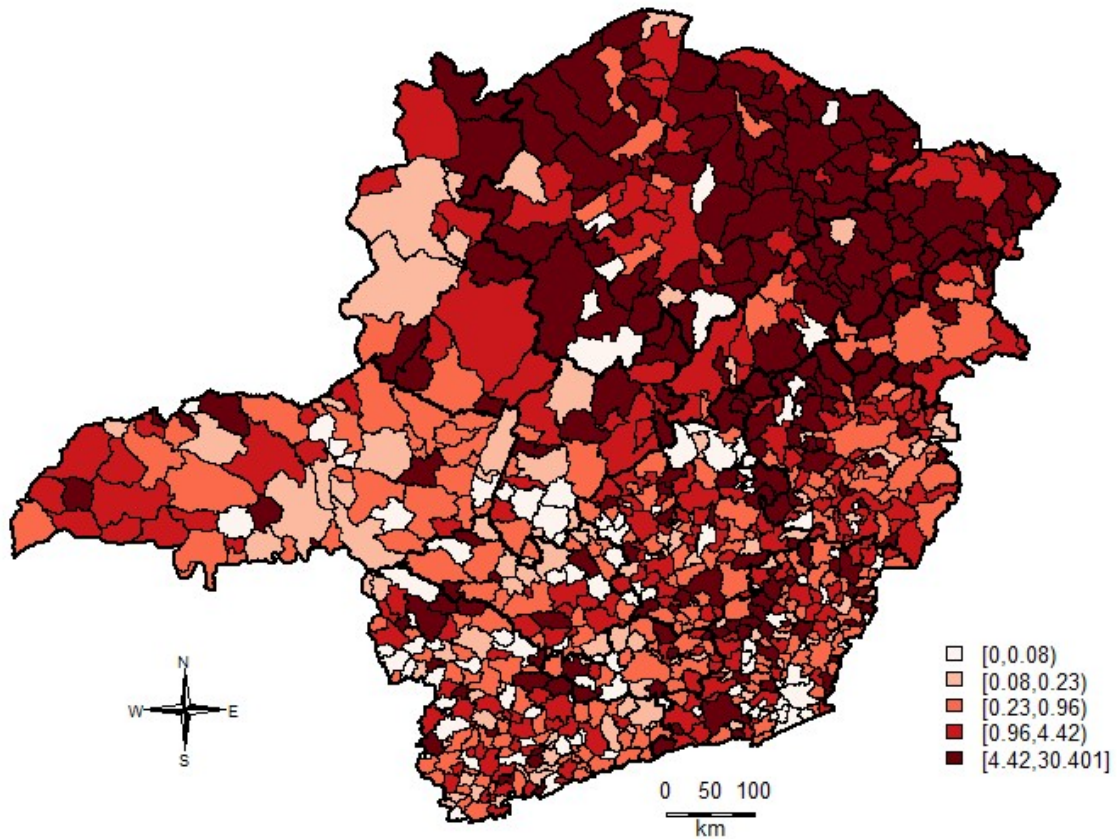


Fonte: Peixoto (2023).

Seguindo com a variável "energia elétrica"(Figura 5.8) observa-se que a grande maioria dos municípios da região sul está situada nos quintis superiores, indicando uma proporção mais elevada de acesso à energia elétrica.

Em contraste com a situação favorável em relação ao acesso à energia elétrica na região Sul de Minas Gerais, uma realidade oposta é verificada quando se trata da proporção de pessoas que residem em domicílios com condições inadequadas de água e esgotamento sanitário. A maior quantidade de casos desse cenário preocupante é identificada na região Norte e Nordeste de Minas Gerais.

Figura 5.9 – Mapa dos quintis com relação a proporção de pessoas que vive em domicílios com água e esgotamento sanitário inadequados nos municípios do Estado de Minas Gerais.



Fonte: Peixoto (2023).

5.1 Análise Espacial

Tabela 5.4 – Tabela com as estimativa e significâncias do Índice de Moran Global.

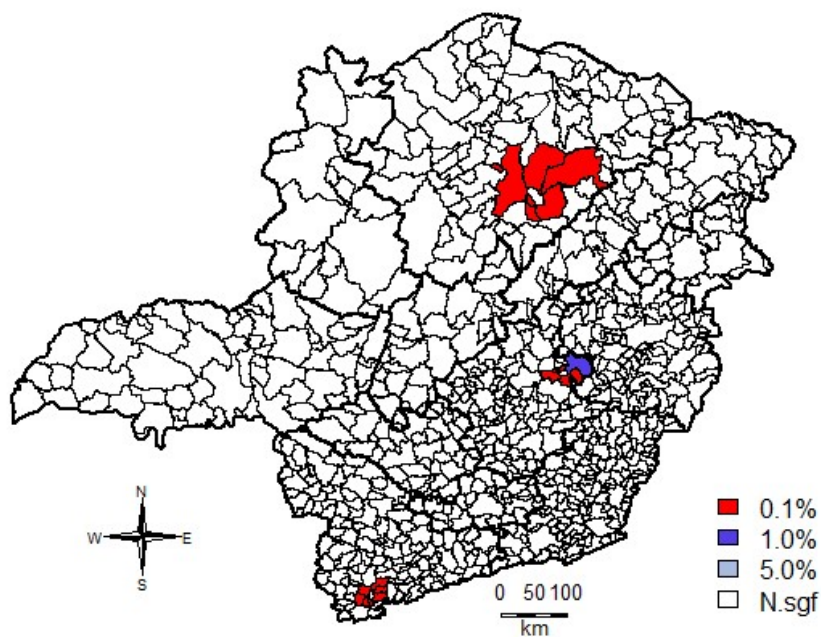
Variáveis	Estimativa	Valor - p
Taxa	-0,0090	0,6881
Idhm	0,1855	< 0,05*
Gini	0,0259	0,09915
Prop.Pobres	-0,0090	0,6881
Perc. Água Encanada	0,1649	< 0,05*
Perc. Banheiro e água	0,2367	< 0,05*
Perc. Energia Elétrica	0,1919	< 0,05*
Esgotamento sanitário	0,1422	< 0,05*

Fonte: Peixoto (2023).

Com base na Tabela 5.4, podemos analisar as estimativas e os valores-p das variáveis. Observa-se que apenas as variáveis taxa, Gini e proporção de pobres não apresentaram significância estatística, isso significa que não há evidências de autocorrelação espacial entre as áreas

consideradas. No entanto, para as demais variáveis com índices de Moran estatisticamente significantes, podemos afirmar que existe uma dependência espacial global. Porém, não descarta a possibilidade de existir dependência espacial em nível local e dado o grande número de áreas em consideração, é necessário investigar essa dependência em escala local. Para isso, utilizamos o mapa LISA.

Figura 5.10 – Mapa dos quartis em relação ao Lisa Map das Taxas de Tuberculose

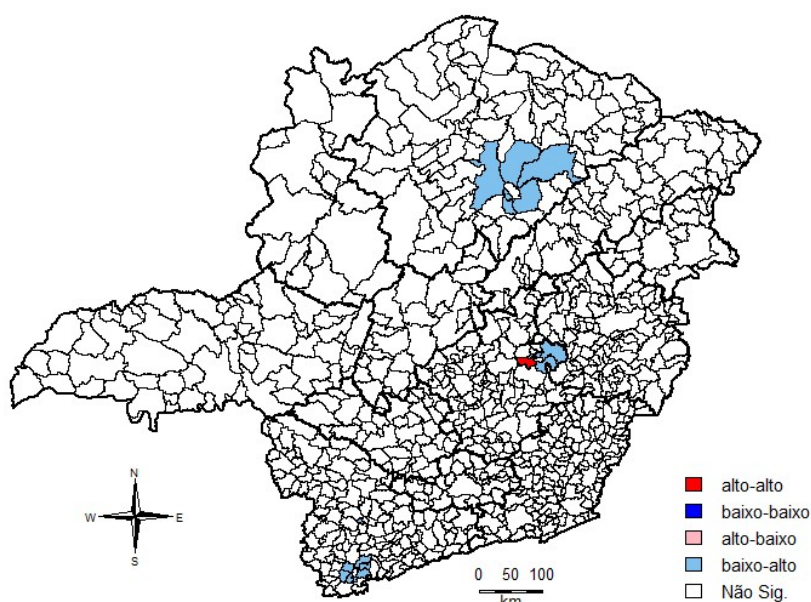


Fonte: Peixoto (2023).

No mapa dos quartis (Figura 5.10), em relação ao Lisa Map, foram identificados três agrupamentos significativos. Isso indica que, apesar da ausência de dependência espacial global para a taxa de Tuberculose, há áreas específicas onde existe uma correlação espacial significativa entre as taxas. Esses agrupamentos sugerem que a ocorrência da Tuberculose está concentrada em determinadas regiões, indicando possíveis padrões espaciais na distribuição da doença.

Ao identificar esses agrupamentos significativos no LISA map, podemos aprofundar nossa compreensão da dinâmica da Tuberculose em nível local.

Figura 5.11 – Mapa dos quintis com relação ao Moran Map das Taxas de Tuberculose



Fonte: Peixoto (2023).

Após a análise do LISA map, foi realizado o Moran map, revelando um padrão interessante de autocorrelação espacial nos dados da taxa de Tuberculose (Figura 5.11). O resultado mostrou que a maioria dos municípios está localizada no terceiro quadrante do Moran map, indicando uma tendência de baixas taxas de Tuberculose em relação aos seus vizinhos.

No entanto, um único município se destaca ao ser identificado no primeiro quadrante do Moran map, Itambé de mato dentro, apontando para uma alta taxa de Tuberculose em comparação com as áreas circundantes. Esse agrupamento espacial de altas taxas de incidência da doença é uma informação crucial para direcionar esforços de prevenção e controle.

5.2 Modelo de Regressão Tobit

Após a análise dos mapas, procedemos com a etapa subsequente: a realização dos ajustes dos modelos Tobit e Tobit Espacial.

Tabela 5.5 – Observações censuradas do modelo Tobit.

Observações	Total	Censurados a esquerda	Não censurados	Censurados a direita
	853	366	487	0

Fonte: Peixoto (2023).

Tabela 5.6 – Resultados da estimação do modelo Tobit.

Coefficientes	Estimativa	Desvio padrão	valor t	Valor - p
(Intercepto)	-1,590e+03	1,102e+03	-1,443	0,1491
IDHM	1,345e+03	6,972e+02	1,929	0,0537
GINI	-4,385e+02	3,630e+02	-1,208	0,2271
PMPOB	2,478e+00	3,626e+00	0,683	0,4945
T_AGUA	-3,023e+00	2,094e+00	-1,444	0,1487
T_BANAGUA	-3,029e+00	3,812e+00	-0,795	0,4268
T_LUZ	1,315e+01	1,103e+01	1,192	0,2333
AGUA_ESGOTO	-4,394e+00	6,791e+00	-0,647	0,5176
logSigma	6,008e+00	3,285e-02	182,907	< 0,05 ***

Fonte: Peixoto (2023).

Das 853 observações avaliadas, o software R censurou 366 observações à esquerda (o valores iguais a 0), sinalizando possíveis desafios na coleta abrangente de dados, especialmente em determinadas regiões. O coeficiente estimado para o Índice de Desenvolvimento Humano Municipal (IDHM) é de 1345, sugerindo que um aumento unitário nesse índice está associado a um acréscimo de 1345 na taxa de Tuberculose. No entanto, vale destacar a variável não é estatisticamente significativa a um nível de 5 %, apenas para 1 0%. É relevante observar que o coeficiente para logSigma não representa uma variável independente, mas sim o logaritmo do desvio padrão do erro.

Tabela 5.7 – Resultados do Teste de Normalidade Shapiro-Wilk para os Resíduos.

Teste de Normalidade	Estatística	Valor p
Shapiro-Wilk	0,74848	< 0,05*

Fonte: Peixoto (2023).

Devido à detecção de violações nos pressupostos, conforme evidenciado pela não normalidade dos resíduos por meio do teste de Shapiro-Wilk, a aplicação do modelo para a condução de inferências torna-se inviável.

5.3 Modelo de Regressão Tobit Espacial

Os resultados do modelo Sar-Tobit (Tabelas 5.8 e 5.9) indicam que a variável Índice de Desenvolvimento Humano Municipal (IDHM) apresenta significância estatística ($p < 0,05$), com uma estimativa positiva de 1408, sugerindo que um aumento no IDHM está associado a um aumento na variável resposta. Por outro lado, as variáveis Índice de Gini, Proporção de pobres, Percentual da população que vive em domicílios com água encanada, Percentual da população que vive em domicílios com banheiro e água encanada, Percentual da população que vive em domicílios com energia elétrica e Percentual de pessoas em domicílios com abastecimento de água e esgotamento sanitário inadequados, não demonstram significância estatística, com p-values superiores a 0,05.

Tabela 5.8 – Resumo do modelo Sar-tobit.

N observações	853
K covariáveis	8
# valores censurados	366
# valores observados	487
Min rho	-1.000
Max rho	1.000

Fonte: Peixoto (2023).

Tabela 5.9 – Resultados da estimação do modelo SAR-Tobit.

Variável	Estimativa	Desvio Padrão	Valor t	Valor p
(Intercept)	-1,724e+03	1,163e+03	-1,483	0,1385
IDHM	1,408e+03	7,085e+02	1,987	< 0,05*
GINI	-4,456e+02	3,752e+02	-1,188	0,2354
PMPOB	2,458e+00	3,759e+00	0,654	0,5135
T_AGUA	-2,847e+00	2,123e+00	-1,341	0,1803
T_BANAGUA	-3,434e+00	4,006e+00	-0,857	0,3916
T_LUZ	1,411e+01	1,155e+01	1,221	0,2224
AGUA_ESGOTO	-4,820e+00	6,945e+00	-0,694	0,4879
sige	1,696e+05	1,187e+04	14,295	< 0,05*
rho	-1,548e-01	9,801e-02	-1,579	0,1146

Fonte: Peixoto (2023).

A estimativa de "sige" indica que os resíduos têm uma variabilidade considerável em relação ao modelo, é significativa ($p < 0,05$), indicando uma forte associação estatística com a resposta, enquanto a variável rho, embora apresente uma estimativa negativa, não atinge significância estatística ($p = 0,1146$).

Tabela 5.10 – Resultado do Teste de Normalidade de Shapiro-Wilk.

Estatística de Teste (W)	0,17266
P-valor	< 0,05*

Fonte: Peixoto (2023).

Novamente foi detectado violações nos pressupostos, conforme evidenciado pela não normalidade dos resíduos por meio do teste de Shapiro-Wilk.

Realizando a comparação dos modelos temos os seguintes valores:

Tabela 5.11 – Comparação dos modelos

	MSE	MAE
Tobit	599560.8	415.0121
SAR-Tobit	116543	165.446

Fonte: Peixoto (2023).

O MSE do modelo SAR-Tobit (116543) é substancialmente menor do que o do modelo Tobit (599560.8). Isso sugere que o modelo SAR-Tobit apresenta um desempenho superior em termos de precisão na previsão dos dados, pois tem um menor valor de MSE.

Similarmente, o MAE do modelo SAR-Tobit (165.446) é consideravelmente inferior ao do modelo Tobit (415.0121). Um MAE menor indica uma melhor capacidade do modelo em prever os valores reais.

É notável que tanto o modelo Tobit quanto o modelo Sar-tobit apresentaram limitações significativas ao serem aplicados aos dados. As inadequações observadas sugerem a necessidade de explorar outras abordagens para melhor capturar a complexidade subjacente aos dados analisados.

Nesse contexto, uma alternativa promissora é a consideração de modelos alternativos, como o probit. O modelo probit, diferentemente do Tobit, é especialmente adequado para lidar com situações em que a variável dependente é binária, sendo útil quando a ocorrência ou não de um evento é o foco da investigação. No contexto desses dados, seria necessário a transformação dos dados para dados binários, onde o "0" seria os sem a doença e o "1" com a doença.

6 CONCLUSÃO

Com base nas análises realizadas, constatamos que não foi encontrada uma autocorrelação espacial global nas taxas de tuberculose nos municípios do Estado de Minas Gerais. No entanto, identificamos a presença de dependência espacial em nível local para alguns municípios, assim como em algumas variáveis explicativas.

Diante do comportamento das variáveis socioeconômicas, torna-se evidente a fragilidade da região norte quando em comparação com a região sul do estado de Minas. Ressaltando os baixos valores para as variáveis IDHM e T_BANAGUA que deixam mais explícito a condição de suas regiões.

Esses resultados têm implicações importantes para os órgãos responsáveis pela saúde no Estado de Minas Gerais. A detecção de dependência espacial local indica que existem áreas específicas com taxas de tuberculose mais elevadas, sugerindo a presença de agrupamentos ou padrões espaciais na ocorrência da doença.

Na etapa de modelagem, a variável IDHM destacou-se de maneira consistente ao longo de todo o processo de análise, revelando-se como um componente estatisticamente relevante no modelo SAR-TOBIT. Sua significância estatística indica que o aumento no Índice de Desenvolvimento Humano Municipal está associado a um aumento na variável de resposta, oferecendo uma perspectiva e confiável sobre o impacto positivo do desenvolvimento humano na incidência da tuberculose.

Com base nesses achados, é crucial que sejam elaborados programas de enfrentamento da tuberculose que visem reduzir as chances de transmissão da doença para a população. Esses programas devem considerar as áreas identificadas como significativas e direcionar recursos e esforços para implementar medidas preventivas, diagnóstico precoce e tratamento adequado nessas regiões.

Ao tomar medidas proativas e específicas com base nas informações fornecidas pela análise espacial, as autoridades de saúde podem desempenhar um papel fundamental na redução da incidência e no controle da tuberculose em Minas Gerais, protegendo a saúde da população e contribuindo para uma melhor qualidade de vida.

REFERÊNCIAS

- AMARAL, P. V.; ANSELIN, L. Finite sample properties of Moran's I test for spatial autocorrelation in tobit models. **Papers in Regional Science**, Wiley Online Library, v. 93, n. 4, p. 773–781, 2013.
- ANSELIN, L.; SYABRI, I.; KHO, Y. Geoda: An introduction to spatial data analysis. **Geographical analysis**, Wiley Online Library, v. 38, n. 1, p. 5–22, 2006.
- ARAÚJO, V. d. C.; GARCIA, R. Determinantes e dependência espacial da inovação nas regiões brasileiras: evidências a partir de um modelo tobit espacial. **Nova Economia**, SciELO Brasil, v. 29, p. 375–400, 2019.
- CÂMARA, G. et al. Análise espacial de áreas. **Análise espacial de dados geográficos. Brasília: EMBRAPA**, p. 157–82, 2004.
- CÂMARA, G.; MONTEIRO, A. M. V. Conceitos básicos em ciência da geoinformação. **Câmara G, Davis C, Monteiro AMV, organizadores. Introdução à ciência da geoinformação**. [acessado 2016 Fev 01]. Disponível em: [http://www.dpi.inpe.br/gilberto/livro/introd\[Links\]](http://www.dpi.inpe.br/gilberto/livro/introd[Links]), 2001.
- COSTA, A. O. Modelagem espacial no mapeamento de doenças transmitidas pelo aedes aegypti no estado da Paraíba. 2017.
- CRESSIE, N. **Statistics for spatial data**. [S.l.]: John Wiley & Sons, 2015.
- DATASUS, M. da S. **Casos de Tuberculose - Desde 2001 (SINAN)**. 2023. Disponível em: <<https://datasus.saude.gov.br/acesso-a-informacao/casos-de-tuberculose-desde-2001-sinan/>>.
- DIGGLE, P. J. **Statistical Analysis of Spatial Point Patterns**. [S.l.]: Academic Press, 2003.
- DUNCAN, D. T. et al. Demographic disparities in the tobacco retail environment in boston: A citywide spatial analysis. **Public Health Reports**, Association of Schools of Public Health, v. 129, n. 2, p. 209, 2014.
- IBGE. 2020. Disponível em: <<https://www.ibge.gov.br>>.
- LESAGE, J.; PACE, R. K. **Introduction to spatial econometrics**. [S.l.]: Chapman and Hall/CRC, 2009.
- LIN, H.-H. et al. Effects of smoking and solid-fuel use on copd, lung cancer, and tuberculosis in china: a time-based, multiple risk factor, modelling study. **The Lancet**, Elsevier, v. 372, n. 9648, p. 1473–1483, 2008.
- LOPES, P. C. B. Análise espacial dos índices de desenvolvimento humano municipal e de vulnerabilidade social em minas gerais. 2022.
- MACIEL, I. C. d. F. et al. Could the breed composition improve performance and change the enteric methane emissions from beef cattle in a tropical intensive production system? **PloS one**, Public Library of Science San Francisco, CA USA, v. 14, n. 7, p. e0220247, 2019.
- MOREIRA, A. d. S. R.; KRITSKI, A. L.; CARVALHO, A. C. C. Determinantes sociais da saúde e custos catastróficos associados ao diagnóstico e tratamento da tuberculose. **Jornal Brasileiro de Pneumologia**, SciELO Brasil, v. 46, 2020.

ORGANIZATION, W. H. et al. **Global hepatitis report 2020**. [S.l.]: World Health Organization, 2020.

PEIXOTO, M. S. et al. Modelagem espacial dos casos de tuberculose dos municípios do estado da Paraíba. **Sigmae**, v. 9, n. 1, p. 7–17, 2020.

PEREIRA, T. V.; NOGUEIRA, M. C.; CAMPOS, E. M. S. Análise espacial da tuberculose e sua relação com indicadores socioeconômicos em um município de médio porte em Minas Gerais. **Revista Brasileira de Epidemiologia**, SciELO Brasil, v. 24, 2021.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2023. Disponível em: <<https://www.R-project.org/>>.

SOUZA, S. d. S. d.; SILVA, D. M. G. V. d.; MEIRELLES, B. H. S. Representações sociais sobre a tuberculose. **Acta Paulista de Enfermagem**, SciELO Brasil, v. 23, p. 23–28, 2010.

TOBIN, J. Estimation of relationships for limited dependent variables. **Econometrica: journal of the Econometric Society**, JSTOR, p. 24–36, 1958.

VIEGAS, A. M. et al. Epidemiologia da tuberculose no município de Contagem, Minas Gerais, Brasil, entre 2002 e 2011. **Revista de Epidemiologia e Controle de Infecção**, Universidade de Santa Cruz do Sul, v. 4, n. 2, p. 146–153, 2014.

VYNNYCKY, E.; FINE, P. The natural history of tuberculosis: the implications of age-dependent risks of disease and the role of reinfection. **Epidemiology & Infection**, Cambridge University Press, v. 119, n. 2, p. 183–201, 1997.

WOOD, R. et al. Burden of new and recurrent tuberculosis in a major South African city stratified by age and HIV-status. **PloS one**, Public Library of Science San Francisco, USA, v. 6, n. 10, p. e25098, 2011.