

UNIVERSIDADE DE SÃO PAULO  
ESCOLA DE ENGENHARIA DE SÃO CARLOS

PEDRO HENRIQUE CALDEIRA CALIARI

Análise de efeitos espaciais para um problema de escolha modal: uma  
abordagem com ênfase na variabilidade espacial

São Carlos

2024



PEDRO HENRIQUE CALDEIRA CALIARI

Análise de efeitos espaciais para um problema de escolha modal: uma  
abordagem com ênfase na variabilidade espacial

**Versão corrigida**

Dissertação apresentada à Escola de  
Engenharia de São Carlos da Universidade de  
São Paulo, como parte dos requisitos para a  
obtenção do Título de Mestre em Ciências.

Orientadora: Profa. Dra. Cira Souza Pitombo

São Carlos

2024

## FICHA CATALOGRÁFICA

AUTORIZO A REPRODUÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica elaborada pela Biblioteca Prof. Dr. Sérgio Rodrigues Fontes da EESC/USP com os dados inseridos pelo(a) autor(a).

C153a Caliari, Pedro Henrique Caldeira  
Análise de efeitos espaciais para um problema de escolha modal: uma abordagem com ênfase na variabilidade espacial / Pedro Henrique Caldeira Caliari; orientadora Cira Souza Pitombo. São Carlos, 2024.

Dissertação (Mestrado) - Programa de Pós-Graduação em Engenharia de Transportes e Área de Concentração em Planejamento e Operação de Sistemas de Transporte -- Escola de Engenharia de São Carlos da Universidade de São Paulo, 2024.

1. Escolha modal. 2. Dependência espacial. 3. Variabilidade espacial. 4. Regressão Logística Geograficamente Ponderada. 5. Teste de variabilidade espacial. I. Título.

## FOLHA DE APROVAÇÃO

Candidato: Engenheiro Pedro Henrique Caldeira Caliari

Título da Dissertação: “Análise de efeitos espaciais para um problema de escolha modal: uma abordagem com ênfase na variabilidade espacial”.

Data da defesa: 06/02/2024.

### Comissão julgadora

### Resultado

Prof<sup>ª</sup> Associada Dra. **Cira Souza Pitombo**

(**Orientadora**)

(Escola de Engenharia de São Carlos – EESC/USP)

Aprovado

Prof. Dr. **Cassiano Augusto Isler**

(Escola Politécnica – EP/USP)

Aprovado

Prof. Dr. **Gustavo Garcia Manzato**

(Universidade Estadual Paulista – UNESP)

Aprovado

Coordenadora do Programa de Pós-Graduação em Engenharia de Transportes:

Prof<sup>ª</sup> Associada **Dra. Ana Paula Camargo Larocca**

Presidente da Comissão de Pós-Graduação:

Prof. Titular **Carlos De Marqui Junior**



## AGRADECIMENTOS

Aos meus pais, Vanessa Medeiros Caldeira Caliari e Durion José Caliari Voss, pelo amor e apoio incondicionais, desde a concepção da ideia de cursar o mestrado, do estudo para o ingresso no curso, o período de aulas na pandemia até problemas de saúde inesperados.

À minha orientadora, professora Cira Souza Pitombo, pela orientação nesta dissertação e nos produtos derivados, pela eventual paciência, momentos de convivência e ensinamentos que deseje levar para o resto da minha vida acadêmica.

Aos orientados da professora Cira, Jorge Ubirajara Pedreira Junior e Samuel de França Marques, que me ajudaram em especial no começo da minha pesquisa, com valiosas dicas e sugestões.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 617463/2021-00.





## EPÍGRAFE

“Instead of using this approach I invoke the first law of geography: everything is related to everything else, but near things are more related than distant things.”

Tobler (1970)



## RESUMO

CALIARI, P. H. C. **Análise de efeitos espaciais para um problema de escolha modal: uma abordagem com ênfase na variabilidade espacial**. 2024. Dissertação (Mestrado) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2024.

O presente trabalho de mestrado tem como objetivo principal analisar os efeitos espaciais na modelagem da escolha modal, com foco na heterogeneidade espacial. A escolha modal estudada é binomial, com os modos automóvel privado e transporte público. A principal lacuna está associada à análise simultânea de dependência e variabilidade espacial, por meio da aplicação de modelos espaciais locais. Também é proposto um teste de variabilidade espacial, associado a resultados agregados e desagregados. As etapas metodológicas gerais propostas seguidas foram: (1) Modelagem logit não espacial; (2) Análise exploratória de dados espaciais; (3) Regressão logística geograficamente ponderada (*GWLR*) e teste de variabilidade espacial de referência; (4) Proposta de teste de variabilidade espacial. Foram comparados modelos logit não espacial e *GWLR* considerando métricas de qualidade de ajuste, matriz de confusão e coerência e capacidade explicativa dos parâmetros estimados. Foi utilizado um banco de dados sintético, criado a partir de interpolação de dados reais (Pesquisa Origem Destino de São Carlos (SP), realizada em 2007 e 2008) onde a dependência espacial foi comprovada. Os resultados do modelo *GWLR* apresentaram melhoria de ajuste frente ao modelo não espacial e tiveram maior ganho explicativo pelos coeficientes estimados locais. Isto corrobora a presença e importância da dependência e variabilidade espacial na escolha modal. Os resultados do teste proposto, agregados e desagregados, corroboram a variabilidade espacial das variáveis adotadas, em consonância com o teste de referência adotado. Os procedimentos seguidos podem ser aplicados futuramente a outras bases de dados com efeitos espaciais. Além disso, o teste de variabilidade espacial proposto traz contribuições na análise, considerando o caráter desagregado de seus resultados e visualizações espaciais de variabilidade.

Palavras-chave: Escolha modal. Dependência espacial. Variabilidade espacial. Regressão Logística Geograficamente Ponderada. Teste de variabilidade espacial.



## ABSTRACT

CALIARI, P. H. C. **Analysis of spatial effects for a modal choice problem: An approach focusing on spatial variability.** 2024. Dissertação (Mestrado) – São Carlos School of Engineering, University of São Paulo, 2024.

The main objective of this study is to analyze the spatial effects on modal choice, focusing on the spatial heterogeneity. A binomial mode choice between private car and public transport is adopted. The main research gap is related to the simultaneous analysis of spatial dependence and variability, with local spatial models. A spatial variability test is also proposed, with both aggregated and disaggregated results. The general methodological steps are: (1) Non-spatial logit modeling; (2) Exploratory spatial data analysis; (3) Geographically weighted logistic regression (*GWLR*) and spatial variability test (reference); (4) Proposal of a spatial variability test. Non-spatial logit and *GWLR* models were compared using measures of goodness of fit, hit rate and coherence and explanatory power of the estimated parameters. We used a synthetic database created by interpolating real data whose spatial dependence was confirmed in previous studies (São Carlos Origin Destination Survey, São Paulo, Brazil, from 2007 to 2008). The *GWLR* results present an improvement over the non-spatial model, while also presenting more explanatory value due to the non-stationary coefficients. The results of the proposed test, aggregated and disaggregated, corroborate the spatial variability of the concerned variables, in accordance with the adopted reference test. The presented procedures can be applied in future studies to other databases with the presence of spatial effects. Also, the proposed spatial variability test brings contributions to the analysis, considering the local nature of its results and the visualization of spatial variability.

Key words: Travel Mode choice. Spatial dependence. Spatial variability. Geographically weighted logistic regression. Spatial variability test.



## LISTA DE ILUSTRAÇÕES

Figura 1 – Lacunas associadas à modelagem da escolha modal .....	33
Figura 2 – Banco de dados – Santa Felícia (São Carlos, São Paulo) .....	37
Figura 3 – Banco de dados adaptado, variável escolha modal .....	39
Figura 4 – Fluxograma do método proposto .....	47
Figura 5 – Fluxograma do teste proposto .....	50
Figura 6 – Mapas do I de Moran local, variável Viagens.....	57
Figura 7 – Mapas do I de Moran local, variável Idade.....	57
Figura 8 – Mapas do I de Moran local, variável Quantidade de automóveis .....	58
Figura 9 – Mapa do <i>LJC</i> na variável escolha opção automóvel.....	58
Figura 10 – Mapa do <i>I</i> de Moran local da probabilidade estimada de escolha do automóvel no modelo não espacial.....	59
Figura 11 – Mapas dos coeficientes estimados .....	62
Figura 12 – Mapa das probabilidades de escolha modal automóvel pelo modelo <i>GWLR</i> . .....	63
Figura 13 – Mapa de coeficientes significativos em todas as variáveis .....	63
Figura 14 – Exemplo de distribuição espacial da variável idade para dados originais (A) e uma das bases de dados permutados (B, C, D).....	65
Figura 15 – Análise local da variabilidade espacial .....	66
Figura 16 – Teste <i>Local Join Count</i> na comparação dos coeficientes dentro do IC .....	67





## LISTA DE TABELAS

Tabela 1 – Sumário dos trabalhos consultados e lacunas de pesquisa .....	36
Tabela 2 – Medidas descritivas das variáveis quantitativas (87 observações).....	38
Tabela 3 – Frequências das categorias das variáveis qualitativas (87 observações).....	38
Tabela 4 – Medidas descritivas básicas do banco de dados das variáveis quantitativas (601 observações) .....	39
Tabela 5 – Percentuais de observações das categorias das variáveis binárias.....	39
Tabela 6 – Teste de Wilcoxon-Mann-Whitney para comparações entre as variáveis numéricas que compõem a amostra original e sintética.....	40
Tabela 7 – Teste Qui-quadrado para comparações entre as variáveis binárias que compõem a amostra original e sintética .....	40
Tabela 8 – Etapas metodológicas e pacotes computacionais .....	54
Tabela 9 – Resultados gerais do modelo logit não espacial .....	55
Tabela 10 – Síntese dos parâmetros estimados do modelo logit não espacial .....	55
Tabela 11 – Resultados globais do I de Moran .....	56
Tabela 12 – Resultados gerais do modelo <i>GWLR</i> .....	60
Tabela 13 – Síntese dos parâmetros estimados do modelo <i>GWLR</i> .....	60
Tabela 14 – Resultados de multicolinearidade do modelo espacial local .....	61
Tabela 15 – Resultado do teste Monte Carlo de variabilidade espacial .....	61
Tabela 16 – Métricas de ajuste dos modelos logit não espacial e <i>GWLR</i> .....	64
Tabela 17 – Matrizes de confusão dos modelos .....	64
Tabela 18 – Resultados do teste proposto .....	65



## LISTA DE ABREVIATURAS E SIGLAS

AEDE	Análise Exploratória de Dados Espaciais
<i>AICc</i>	<i>Corrected Akaike Information Criterion</i>
CV	Coefficiente de Variação
<i>GWLR</i>	<i>Geographically Weighted Logistic Regression</i>
<i>GWR</i>	<i>Geographically Weighted Regression</i>
IC	Intervalo de confiança
<i>KNN</i>	<i>K Nearest Neighbors</i>
<i>LJC</i>	<i>Local Join Count</i>
<i>MGWR</i>	<i>Multiscale Geographically Weighted Regression</i>
<i>SGWR</i>	<i>Semi-parametric Geographically Weighted Regression</i>



## SUMÁRIO

CAPÍTULO 1: INTRODUÇÃO .....	23
1.1 Considerações iniciais .....	23
1.2 Objetivos.....	25
1.3 Hipóteses .....	25
1.4 Contribuições.....	25
1.5 Definições adotadas neste trabalho.....	26
1.6 Produtos da dissertação .....	26
1.7 Estrutura do texto .....	27
CAPÍTULO 2: REVISÃO BIBLIOGRÁFICA .....	28
2.1 Contexto inicial de escolha modal e modelos não espaciais .....	28
2.2. Abordagem espacial para análise da escolha modal.....	29
2.3. Testes de variabilidade espacial .....	34
2.4 Lacunas de pesquisa .....	35
CAPÍTULO 3: MATERIAIS E MÉTODO .....	37
3.1 Dados.....	37
3.2 Modelo logit não espacial.....	41
3.3 Análises Exploratórias de Dados Espaciais (AEDE) .....	42
3.4 Modelo espacial local e teste de variabilidade espacial de Monte Carlo .....	44
3.5 Procedimento metodológico .....	47
3.6 Teste de variabilidade espacial proposto .....	50
3.6.1. Modelagem não espacial e obtenção dos intervalos de confiança dos parâmetros	51
3.6.2. Testes univariados de dependência espacial.....	51
3.6.3. Modelo GWLR no banco de dados sem permutação .....	51
3.6.4. Comparação dos coeficientes estimados com o intervalo de confiança do modelo não espacial.....	52
3.6.5. Criação de bancos de dados permutados e processamento dos modelos <i>GWLR</i> permutados .....	52
3.6.6. Aplicação do teste proposto.....	52
3.6.7. Validação do teste proposto.....	53
3.6.8. Testes locais adicionais para variabilidade espacial.....	53
3.7 Recursos computacionais .....	54
CAPÍTULO 4: RESULTADOS .....	55
4.1 Modelo logit não espacial e obtenção dos intervalos de confiança dos parâmetros.....	55
4.2 Análises exploratórias espaciais .....	56
4.3 Modelo <i>GWLR</i> e teste de variabilidade espacial Monte Carlo.....	59
4.4 Comparações entre modelos logit não espacial e espacial local .....	64
4.5 Teste de variabilidade espacial local proposto neste trabalho .....	65

CAPÍTULO 5: CONCLUSÕES E SUGESTÕES PARA TRABALHOS FUTUROS .....	68
5.1 Síntese do trabalho .....	68
5.2 Contribuições e conclusões .....	69
5.3 Limitações e sugestões para trabalhos futuros .....	70
REFERÊNCIAS .....	72

# CAPÍTULO 1: INTRODUÇÃO

*Este capítulo introdutório contextualiza o presente trabalho, descrevendo, em seguida, objetivos, hipóteses, contribuições e principais produtos da presente dissertação de mestrado.*

## 1.1 Considerações iniciais

Tradicionalmente, problemas de escolha modal são analisados com uso de modelos de escolha discreta, baseados em funções utilidade aleatórias (McFadden, 1974; Ben-Akiva, 1973; Ben-Akiva e Lerman, 1985). Entretanto, normalmente tais modelos não são capazes de lidar com a dimensão espacial; ora por não incorporarem esses dados explicitamente, ora porque efeitos espaciais costumam violar hipóteses nos modelos não espaciais (Anselin, 1988; Lesage, 1999). Sabe-se, ainda, que as variáveis associadas ao comportamento de viagens possuem efeitos espaciais intrínsecos (Mohammadian e Kanaroglou, 2003; Páez, 2006; Erdogan, 2009; Zhang *et al.*, 2012; Lindner *et al.*, 2021; Marques e Pitombo, 2023), assim como a escolha modal (Rajamani *et al.*, 2003; Lindner e Pitombo, 2018; Nkeki e Asikhia, 2019; Tao *et al.*, 2019; Assirati e Pitombo, 2021).

Os efeitos espaciais são abordados pela econometria espacial, que os define como a dependência e heterogeneidade espaciais, que podem estar presentes simultaneamente (Anselin, 1988). A dependência espacial é a interação multidirecional entre as observações, de acordo com a primeira lei da geografia de Tobler: “tudo está relacionado entre si, mas as coisas mais próximas têm mais impacto do que coisas distantes” (Tobler, 1970). Em outras palavras, na presença de dependência espacial uma variável (dependente ou independente) em dada observação também está em função dos valores das observações vizinhas. Já a heterogeneidade espacial pode ser compreendida como uma instabilidade estrutural de resposta das variáveis (Anselin, 1988). Ou seja, a presença da heterogeneidade espacial implica que o parâmetro real de uma variável é uma superfície de valores que variam regionalmente. Apesar da maior complexidade de obtenção, resultados locais possibilitam uma análise mais diversa e verossímil que não é disponível com estimadores únicos (globais ou estacionários). Neste trabalho, o termo variabilidade espacial é utilizado como sinônimo de heterogeneidade espacial.

Existem três formas mais comuns para incorporação dos efeitos espaciais para variáveis de interesse. A primeira e mais simples é a Análise Exploratória de Dados Espaciais (AEDE), como o *Local Moran's I* (Anselin, 1995) e o *Local Join Count (LJC)* (Anselin e Li, 2019). Suas principais limitações no escopo deste trabalho são: dificuldade de fazer uma análise multivariada (Anselin e Li, 2019, 2020) e que analisam apenas a dependência espacial. A

segunda opção são modelos de regressão espaciais globais, que incorporam dados e estrutura espaciais, porém cujos coeficientes estimados são estacionários, não atendem hipótese de variabilidade espacial. A terceira opção, e a adotada neste estudo como etapa final, são os modelos espaciais locais, no caso o modelo *Geographically Weighted Logistic Regression* (*GWLR*, Brunson *et al.*, 1996; Fotheringham *et al.*, 2002), em português Regressão Logística Geograficamente Ponderada. A sua principal vantagem frente aos modelos espaciais globais é a possibilidade de estimar um coeficiente para cada observação das variáveis selecionadas, logo com maior fidedignidade caso a variabilidade espacial esteja presente.

Considerando que a escolha modal é um fenômeno com presença intrínseca dos efeitos espaciais (Rajamani *et al.*, 2003; Lindner e Pitombo, 2018; Nkeki e Asikhia, 2019; Tao *et al.*, 2019; Assirati e Pitombo, 2021), neste trabalho são explorados modelos espaciais locais e testes de dependência espacial. Anselin (1988) e Fotheringham (1997) destacam três motivos que podem indicar não estacionariedade dos parâmetros estimados: (1) erros de coleta de dados, (2) natureza do fenômeno estudado e (3) má especificação dos modelos. Assim, há um interesse em incorporar simultaneamente a dependência e a variabilidade espacial nos estudos de escolha modal. Isso se dá pelo maior potencial explicativo dos resultados locais e expectativa de sua validade, bem como para evitar a violação de hipóteses dos modelos não espaciais na presença dos efeitos espaciais (Anselin, 1988; LeSage, 1999).

Porém, mesmo com a expectativa da variabilidade espacial, ela não é necessariamente presente em todas as variáveis (Fotheringham *et al.*, 2002; Nakaya *et al.*, 2014), mesmo com a presença de dependência espacial. Logo, testar a hipótese de variabilidade espacial de uma variável é uma importante etapa que deve ser atendida para plena análise dos resultados de modelos locais. O teste de variabilidade espacial relaciona-se a duas perguntas chaves: (1) os valores dos parâmetros estimados têm variação estatística na região de estudo? (2) A variação dos parâmetros locais estimados possui um padrão espacial?

Os testes de variabilidade espacial encontrados na literatura estão associados a resultados agregados, como indicadores, relacionados a pseudo p-valor (Leung *et al.*, 2000; Fotheringham *et al.*, 2002; Nakaya *et al.*, 2014; Nakaya, 2015). Nestes testes foram encontradas limitações como comparações arbitrárias de métricas de ajuste (Nakaya *et al.*, 2014; Nakaya, 2015), impossibilidade para aplicação em variáveis binárias ou de contagem (Leung *et al.*, 2000). Também, comum a todos os três testes consultados, a não consideração de resultados de modelos estacionários ou não espaciais. Dos três, o considerado mais adequado, e utilizado neste estudo como referência para validação do teste proposto, é o de Fotheringham *et al.* (2002).



## 1.2 Objetivos

O objetivo principal deste trabalho de mestrado é analisar efeitos espaciais para um problema de escolha modal, com ênfase na variabilidade espacial.

De forma específica, tem-se o objetivo de propor um teste de variabilidade espacial, com resultados que podem ser analisados posteriormente de forma desagregada.

## 1.3 Hipóteses

Neste trabalho a seguinte hipótese é avaliada:

- I. A incorporação de dependência e variabilidade espacial traz melhorias às estimativas relacionadas à escolha modal.

## 1.4 Contribuições

O presente trabalho tem duas contribuições, a primeira é a elaboração de um procedimento metodológico coeso, com procedimentos bem fundamentados na literatura; que permita a análise simultânea dos efeitos espaciais para o fenômeno da escolha modal.

O procedimento metodológico foi elaborado para que as etapas seguintes atendam ao menos uma lacuna importante das anteriores. Isto garante a coesão teórica e prática das ferramentas e a maturação dos resultados entre cada etapa, tanto na capacidade explicativa quanto no ganho de ajuste dos modelos. Para corroborar a hipótese adotada e a validade do procedimento, este está sendo aplicado em um banco de dados sintético de escolha modal, proveniente de um estudo de caso real (dados da Pesquisa Origem/Destino de 2007/2008, realizada em São Carlos – SP). Outra qualidade do procedimento é a facilidade de acesso às ferramentas computacionais, pois todas elas são gratuitas e de acesso livre, com disponibilidade de manuais e de bancos de dados de treino.

Com o procedimento completo, espera-se que o conjunto de resultados sejam valiosos para que pesquisadores elaborem diretrizes mais eficientes e efetivas para situações de escolha modal. Na realidade, nenhuma etapa proposta é limitada ao escopo de escolha modal ou comportamento de viagens. Logo, o procedimento pode ser replicado para quaisquer estudos de variáveis dependentes binárias em que se espera a presença de dependência espacial dos dados e variabilidade espacial dos coeficientes estimados.

A segunda contribuição deste trabalho é a elaboração, aplicação e validação de um teste de variabilidade espacial. Mais detalhes das etapas do teste são apresentados no Capítulo 3. Existem três contribuições relativas ao teste proposto: (1) a consideração explícita do resultado

do modelo não espacial; (2) não há restrição sobre o tipo ou distribuição da variável dependente do modelo; (3) resultados do teste podem ser analisados de forma desagregada, avaliando a aleatoriedade espacial.

### 1.5 Definições adotadas neste trabalho

Para normalização dos termos deste trabalho são adotadas as seguintes definições:

- **Dependência espacial:** efeito espacial que representa a interação multidirecional dos indivíduos/observações. Um valor de uma variável em dada localização fica em função do valor da mesma variável em localizações vizinhas (Anselin, 1988; LeSage, 1999).
- **Variabilidade espacial:** instabilidade estrutural de resposta das variáveis sob um padrão espacial não aleatório (Anselin, 1988; Fotheringham *et al.*, 2002). Neste trabalho, o termo variabilidade espacial é sinônimo de heterogeneidade espacial e de não-estacionariedade espacial.
- **Efeitos espaciais:** dependência espacial e heterogeneidade espacial (Anselin, 1988), usualmente imbricadas (Anselin, 1988; LeSage, 1999; Fotheringham *et al.*, 2002).
- **Estacionariedade espacial:** o oposto de variabilidade espacial: os coeficientes estimados são os mesmos em todas as observações analisadas. É uma hipótese adotada em modelos não espaciais e modelos espaciais globais (SAR, SEM, SLX, Durbin, GSM, *etc*).

### 1.6 Produtos da dissertação

Esta dissertação está associada a três produtos principais, listados a seguir:

- **Aplicação de regressão logística multinomial geograficamente ponderada para um problema de escolha modal.** Pedro Henrique Caldeira Caliari e Cira Souza Pitombo, relatório de teses e dissertações em andamento, apresentado e publicado no 36º Congresso de Pesquisa e Ensino em Transportes, (ANPET 2022) (<https://proceedings.science/anpet-2022/trabalhos/aplicacao-de-regressao-logistica-multinomial-geograficamente-ponderada-para-um-p?lang=pt-br>).
- **Proposta de um teste para verificar a variabilidade espacial em modelos espaciais locais: uma abordagem com ênfase na escolha modal,** Pedro Henrique Caldeira Caliari e Cira Souza Pitombo, artigo científico (congresso), apresentado e publicado nos anais do 37º Congresso de Pesquisa e Ensino em Transportes (ANPET 2023)

(<https://proceedings.science/anpet-2023/trabalhos/proposta-de-um-teste-para-verificar-a-variabilidade-espacial-em-modelos-espaciai?lang=pt-br>).

- **Analysis of spatial effects for a modal choice problem: An approach focusing on spatial variability**, Pedro Henrique Caldeira Caliari e Cira Souza Pitombo, artigo científico submetido em periódico internacional (em análise).

### 1.7 Estrutura do texto

O texto está dividido em cinco capítulos, incluso o atual. O Capítulo 2 apresenta um breve contexto da literatura de escolha modal não espacial. Também discorre da evolução dos trabalhos do tema no contexto da inclusão dos efeitos espaciais, com estudos de caso que utilizaram os programas computacionais escolhidos ou similares. Nesse capítulo também está presente a revisão da literatura feita sobre os testes de variabilidade espaciais disponíveis.

No Capítulo 3 serão apresentados os materiais utilizados, as adaptações realizadas, um detalhamento de cada ferramenta aplicada no procedimento, o procedimento metodológico proposto, um fluxograma síntese e o modelo aplicado como validação do procedimento. Também é detalhado o procedimento do teste de variabilidade espacial desenvolvido e apresentada a listagem dos recursos computacionais utilizados em cada etapa do trabalho.

O Capítulo 4 trata dos resultados de cada seção do Capítulo 3, mais a comparação entre o modelo logit não espacial e o modelo *GWLR*. Por fim, o Capítulo 5 traz as conclusões do trabalho, com implicações dos resultados obtidos, limitações e sugestões para trabalhos futuros.

## CAPÍTULO 2: REVISÃO BIBLIOGRÁFICA

*Este capítulo traz os principais trabalhos consultados ligados ao tema escolha modal e análise espacial. Descreve modelagens utilizadas em problemas de escolha modal, com foco na econometria espacial. Além disso, são debatidos testes propostos para análise de variabilidade espacial, bem como as principais lacunas de pesquisa associadas à presente dissertação.*

### 2.1 Contexto inicial de escolha modal e modelos não espaciais

O problema de escolha modal envolve uma série de fatores, tais como necessidades e características individuais, atividades que serão realizadas, a acessibilidade urbana e infraestrutura de transportes, os modos de transporte disponíveis, fatores relacionados à viagem (Ben-Akiva e Lerman, 1985), além da distribuição espacial das atividades no meio urbano (Páez, 2006; Sheng e Sharp, 2019).

Tradicionalmente, problemas de escolha modal são analisados com auxílio de modelos de escolha discreta, baseados em funções utilidade aleatórias, ou *RUM* (*Random Utility Models*, Ben-Akiva e Lerman, 1985). Nos modelos *RUM*, cada modo de transporte que pode ser escolhido é representado por uma equação que descreve quantitativamente sua utilidade, simbolizando seus benefícios e atratividade.

Uma boa representação dos fatores mais influentes da escolha modal implica em um ganho explicativo e de acurácia (Ben-Akiva e Lerman, 1985). Com as utilidades calculadas, obtêm-se as probabilidades de escolha de cada modo para cada indivíduo, e derivam-se métricas da qualidade global do modelo.

O tipo de modelo aqui mencionado faz parte da econometria tradicional. Explicando os termos, um modelo é uma representação formal de teorias, que visam se aproximar da realidade (Bollen, 1989). A econometria é a aplicação de ferramentas matemáticas, estatísticas e de economia para investigar relações quantitativas (Tintner, 1953).

Contudo, o campo tradicional da econometria não tem ferramentas para lidar com a dimensão espacial e suas consequências nos procedimentos (Anselin, 1988; LeSage, 1999, Almeida, 2012). Ainda, variáveis que representam a escolha modal usualmente não são aleatórias no espaço (Rajamani *et al.*, 2003; Goetzke, 2008; Nkegbe *et al.*, 2012; Mendiola *et al.*, 2014; Lindner e Pitombo, 2018; Nkeki e Asikhia, 2019; Assirati e Pitombo, 2021; Mondal e Bhat, 2022). Portanto, a incorporação do espaço na estrutura dos modelos torna-se necessária para a compreensão adequada do fenômeno, além do escopo da econometria tradicional.

Uma tentativa de mitigar a falta das informações espaciais pode ser vista, por exemplo, em Rajamani *et al.* (2003) e Nkegbe *et al.* (2012), com a introdução de variáveis que tentam explicitar informações espaciais, associadas a parâmetros estimados estatisticamente significativos e importantes na interpretação dos resultados apresentados. Porém, tal estratégia é limitada dado que os efeitos espaciais costumam violar as hipóteses dos modelos de regressão não espaciais (Anselin, 1988; LeSage, 1999; Almeida, 2012). Novamente, um esforço mais adequado para lidar com os efeitos espaciais intrínsecos à escolha modal é a adoção da econometria espacial, introduzida na próxima seção.

## **2.2. Abordagem espacial para análise da escolha modal**

O campo da econometria espacial foi cunhado por Jean HP Paelinck (1979), sendo consolidada por diversos estudos de referência nas décadas de 1980 a 2000, são exemplos: Anselin (1988), Getis e Ord (1992), Anselin (1995), Brunsdon *et al.* (1996), LeSage (1999) e Fotheringham *et al.* (2002). A econometria espacial pode ser definida como uma área derivada da econometria, que lida com interações e estruturas espaciais em modelos de regressão (Anselin, 1988). Por extensão, os modelos de econometria espacial podem amenizar ou evitar vieses decorrentes de quebras de suposições presentes nos procedimentos não espaciais (Anselin, 1988; LeSage, 1999; LeSage e Pace, 2009; Almeida, 2012). Uma visão mais prática da econometria espacial para o atual trabalho pode ser o foco no estudo da dependência e heterogeneidade espaciais e suas diversas implicações na calibração e interpretação de testes e modelos. Além das definições apresentadas na Seção 1.5, os próximos parágrafos definem e detalham melhor estes dois termos, bem como a importância para a escolha modal.

Remetendo à primeira lei da geografia de Tobler: “tudo está relacionado entre si, mas as coisas mais próximas têm mais impacto do que coisas distantes” (Tobler, 1970). A dependência espacial é equivalente à lei de Tobler, sendo o impacto desse efeito inversamente proporcional à distância. A dependência espacial pode decorrer da natureza intrínseca ao tema de estudo, erros de medida (como agrupamentos artificiais), ou má especificação do modelo (Anselin, 1988; LeSage, 1999; Almeida, 2012).

Os erros de medida ocorrem por imprecisão das coletas e/ou da agregação espacial. A primeira causa é pouco provável nos dias atuais considerando o avanço da precisão dos aparelhos de medida. Contudo, podem ocorrer tais erros por falhas do processo de coleta e questionários, por associação errônea das coordenadas com um dos eventos/observações ou uso de um sistema de coordenadas diferente do original. Em outros casos, como dados censitários

agregados, há uma inerente agregação de informações que pode introduzir a dependência espacial, ou distorcer a dependência espacial já existente.

Os erros de especificação merecem um estudo cauteloso, mesmo que na ausência de efeitos espaciais. São causas desse tipo de erro: (1) má escolha do modelo, (2) adoção de hipóteses incorretas e (3) a omissão de variáveis pertinentes (Ramsey, 1969; Anselin, 1988; LeSage, 1999). Visto a complexidade e diversidade de informações pertinentes ao comportamento de viagens e à escolha modal (Ben-Akiva e Lerman, 1985; Pitombo *et al.*, 2011; Zhang *et al.*, 2012; Nkeki e Asikhia, 2019), a omissão de variáveis pertinentes é frequentemente uma limitação dos dados disponíveis no estudo, podendo ser uma fonte adicional da dependência espacial.

No tópico da heterogeneidade espacial, ela é definida pela presença de diferenças estruturais de resposta de uma variável em diferentes localidades (Anselin, 1988). Ou seja, o coeficiente estimado de uma variável não é mais um valor único, mas sim um vetor de coeficientes que variam não aleatoriamente na região. Suas causas são as mesmas da dependência espacial: (1) erros de medidas, (2) erros de especificação e a (3) natureza espacial dos dados. Isto reforça a natureza íntima entre a dependência espacial e a heterogeneidade espacial, sendo que ambas costumam ocorrer simultaneamente (Anselin, 1988; LeSage, 1999; Almeida, 2012), problematizando uma modelagem apropriada. A heterogeneidade espacial também é frequentemente encontrada na literatura com sinônimos de variabilidade espacial, variabilidade local, não-estacionariedade espacial ou instabilidade espacial (Anselin, 1988; Fotheringham, 1997; Fotheringham e Brunson, 1999; Almeida, 2012).

Além dos parâmetros variáveis no espaço, a heterogeneidade espacial também causa heterocedasticidade, onde o resíduo da regressão apresenta dependência espacial (Anselin, 1988; Anselin e Griffith, 1988; LeSage, 1999; Almeida, 2012). Neste trabalho, a análise será restrita à variabilidade espacial dos coeficientes estimados.

Analogamente à econometria tradicional, um passo inicial indispensável em estudos de econometria espacial com modelos de regressão espacial é a aplicação de análises exploratórias, tais como as Análises Exploratórias de Dados Espaciais (AEDE). Essas análises permitem entender melhor os padrões espaciais, ou a falta deles, em cada variável considerada, ou em grupos de variáveis. Os resultados podem ser interpretados de forma global, considerando todas as observações ao mesmo tempo em métricas sintetizadas, ou local, que permite a identificação de aglomerados e observações limítrofes (*outliers*) (Anselin, 1995).

Uma AEDE local usual na literatura de transportes é o *Univariate Local Moran's I* (Anselin, 1995), pela sua facilidade de interpretação e riqueza de resultados. AEDE

multivariadas também estão disponíveis, como ramificações multivariadas do índice I de Moran (Anselin, 2017) e o Geary Multivariado Local (Anselin, 2019). É indispensável a identificação de se, e como, os efeitos espaciais ocorrem nas variáveis e observações, ou grupos de observações, uma vez que essas informações podem impactar e restringir os tipos de modelo de regressão passíveis de uso (LeSage e Pace, 2009; Almeida, 2012).

Por conta de seu poder exploratório, as AEDE possibilitam identificações de padrões muitas vezes elusivos sem a dimensão espacial (Ramos e Da Silva, 2007; Erdogan, 2009; Aljoufie *et al.*, 2013; Sánchez-Díaz *et al.*, 2016).

Para variáveis binárias, caso da variável escolha modal, a maioria das AEDE disponíveis, inclusive o *Univariate Local Moran's I*, não são adequadas (Boots, 2003; Anselin e Li, 2019). Uma alternativa prática e direta é o *Local Join Count (LJC)* (Boots, 2003; Anselin e Li, 2019). Como exemplo, em Rybarczyk e Shaker (2021) o *LJC* foi utilizado como etapa exploratória para a variável de preferência declarada de uso de bicicletas em transporte público para viagens de última milha. Seu uso corroborou a existência da dependência espacial e ajudou a entender os padrões espaciais da variável binária em estudo.

Contudo, as AEDE multivariadas (Cook *et al.*, 1994; Dykes, 1998; Anselin, 2017; Anselin, 2019) são limitadas para quantificação e confirmação de relações entre três ou mais variáveis (Anselin e Li, 2019, 2020). Através da análise espacial confirmatória, por modelos de regressão espaciais globais e locais, é possível preencher tal lacuna e melhor compreender o fenômeno da escolha dos modos de transporte.

Assim, propõe-se o uso dos modelos de regressão da econometria espacial. Os modelos econométricos espaciais podem ser classificados em globais (ou estacionários), em que o vetor de parâmetros estimados é único para todas as observações ou locais, onde os coeficientes, *t*-valores e demais métricas dos estimadores são específicos para cada observação (Brunsdon *et al.*, 1996).

Modelos espaciais globais, ou modelos de autocorrelação espacial, possuem diversos tipos de estruturas, podendo considerar a dependência espacial na variável dependente, variáveis independentes e resíduos ou uma combinação dessas opções. São exemplos: SAR, SEM, SLX, Durbin e GSM. Para mais detalhes sugere-se Anselin (1988), Lesage (1999) e Almeida (2012). Neste tipo de modelo selecionou-se um recorte da literatura (Mohammadian e Kanaroglou, 2003; Miyamoto *et al.*, 2004; Goetzke, 2008; Sheng e Sharp, 2019; Assirati e Pitombo, 2021). Neste conjunto, as justificativas da aplicação dos modelos frente às alternativas não-espaciais são: o interesse na dependência espacial intrínseca ao tema e obtenção de estimadores não viesados (ou menos viesados). Contudo, uma limitação de todos os modelos

espaciais globais é que os coeficientes estimados de uma variável são únicos para toda a região do estudo, o que costuma não ser verossímil dada a presença de efeitos espaciais (Fotheringham, 1997; Fotheringham *et al.*, 2002).

Para modelos espaciais locais foram consultados os estudos de Páez (2006), Albuquerque *et al.* (2017), Nkeki e Asikhia (2019) e Rybarczyk e Shaker (2021). Dado o foco no procedimento, trabalhos fora do escopo de escolha modal ou engenharia de transportes também foram considerados. A maior riqueza dos resultados obtidos, por conta da variabilidade espacial, foi uma justificativa importante dos trabalhos acima e um ganho frente aos modelos globais. Porém, considera-se falha a falta de aplicação de testes para variabilidade espacial nos trabalhos citados.

Com partida os trabalhos de outras áreas, Páez (2006) aplicou um modelo probit em regressão geograficamente ponderada (RGP, ou *GWR* em inglês) para analisar o impacto da infraestrutura de transportes e vacância de lotes vizinhos na mudança do uso do solo. Os resultados obtidos pelo autor corroboraram as hipóteses de significância dos parâmetros estimados associados a variáveis, ainda que só com duas variáveis, e de variabilidade espacial.

Em Albuquerque *et al.* (2017), lançou-se mão de um modelo *GWLR* no programa computacional *GWR4*, para estudo de *credit scoring* no Distrito Federal, agregado por regiões. Com foco no objetivo de comparar o ganho explicativo dos modelos *GWLR* frente aos logit não espacial. Para a comparação, foram utilizados valores de *AICc*, porcentagens de acertos e falso positivos e métricas dos escores de crédito. Apesar da pouca diferença encontrada entre os modelos, a análise de coeficientes espacialmente variáveis agrega um maior potencial explicativo do fenômeno que não seria possível com um modelo logit não espacial.

Rybarczyk e Shaker (2021) estudaram a adoção de integração de bicicletas no transporte público para viagens de primeira ou última milha, com variável dependente categórica e de preferência declarada. Os modelos utilizaram variáveis dos indivíduos e informações da região em uso do solo e infraestrutura disponível para os modos ônibus, bicicleta e a pé. Foram elaborados mapas de calor e *LJC* na etapa exploratória, e em seguida, foram comparados três modelos de regressão, sendo um deles o *GWLR*. Os resultados apontam uma pequena melhoria do ajuste do modelo *GWLR* frente aos modelos não espaciais e, segundo os autores, a variabilidade espacial dos estimadores forneceu relações não observadas na literatura consultada.

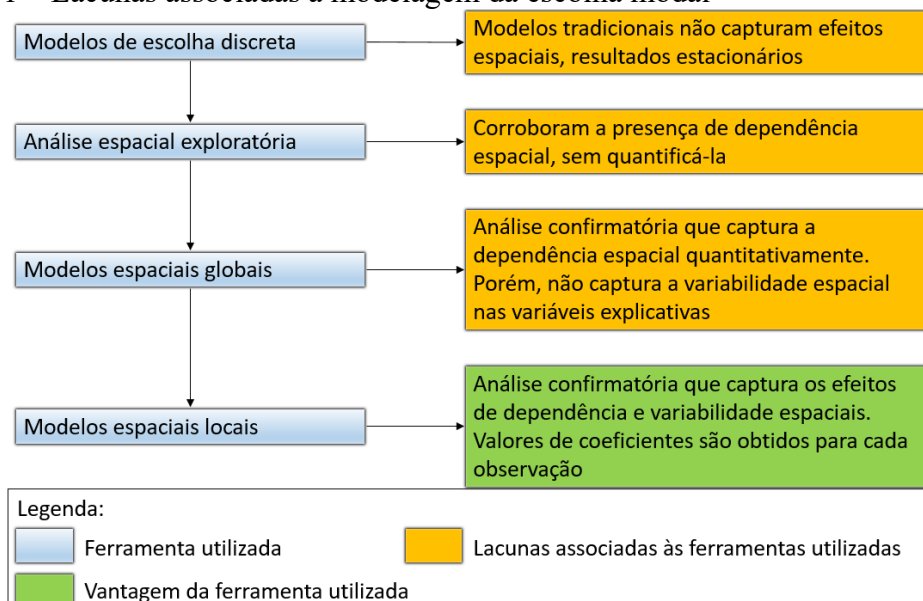
Já em Nkeki e Asikhia (2019), o modelo *GWLR* é aplicado a um contexto de escolha modal, sendo até o momento o único estudo encontrado no âmbito da revisão da literatura aqui realizada. Os autores avaliam diversas variáveis explanatórias (combinadas em fatores por



análise dos componentes principais), com a variável de escolha modal, onde vários modos foram agregados em duas categorias: transportes públicos e transportes privados. Apesar desta agregação dificultar a análise dos resultados, por não atribuí-los a um único modo em cada categoria, esta é uma limitação compreensível, dado que um modelo *GWR* para logit multinomial ainda não está disponível em termos computacionais. Contudo, o estudo citado possui duas limitações importantes: a primeira é a aplicação de um teste de variabilidade espacial que não é visto como adequado. A segunda é que o resultado deste teste dá indícios que nem todas as variáveis possuem variabilidade espacial, sendo mais adequado um modelo *GWLR* semi-paramétrico (Fotheringham *et al.*, 2002; Nakaya *et al.*, 2014). Porém, não é feito um novo modelo para diferenciar as variáveis espacialmente estacionárias das não estacionárias. A importância das limitações do estudo citado será melhor detalhada na seção seguinte.

Ainda que esperada a rejeição da estacionariedade espacial dada a presença de dependência espacial (Anselin 1988; Fotheringham, 1997), é prudente realizar um teste dessa hipótese (Fotheringham e Brunson, 1999; Fotheringham *et al.*, 2002; Nakaya *et al.*, 2014; Fotheringham e Oshan, 2016). Caso rejeitada, os modelos espaciais globais não são adequados e devem ser usados modelos espaciais locais para acomodar a variabilidade espacial dos estimadores, provendo maior verossimilhança e potencial explicativo aos resultados (Brunson *et al.*, 1996; Fotheringham e Brunson, 1999; Bhat e Zhao, 2002; Nakaya, 2015). A Figura 1, resume e ilustra as lacunas de cada opção de modelagem considerada para escolha modal e a vantagem da opção escolhida.

Figura 1 – Lacunas associadas à modelagem da escolha modal



Fonte: elaborado pelo autor.

Ressalta-se que os testes de variabilidade espacial são específicos ao modelo elaborado e feitos para cada variável. Para situações em que o conjunto de variáveis não seja unanimemente estacionário ou não-estacionário, sugere-se o uso de modelos semi-paramétricos de regressão geograficamente ponderada (*SGWR*), com variáveis ao nível global e local, vide Nakaya *et al.* (2014). Uma melhor opção ao *SGWR* é o *Multiscale GWR* (Fotheringham *et al.*, 2017), adequado para conjuntos em que todas, ou parte das variáveis possuem variabilidade espacial corroborada.

### 2.3. Testes de variabilidade espacial

Anselin (1988) e Fotheringham (1997) destacam três indícios de não estacionariedade espacial dos parâmetros estimados: erros de coleta de dados, natureza do fenômeno estudado e má especificação dos modelos. Porém, mesmo com a expectativa da variabilidade espacial, ela não é necessariamente presente em todas as variáveis, ainda que haja dependência espacial e mesmo com um certo grau de variação regional do parâmetro estimado no modelo *GWR*.

Uma pergunta chave ao considerar a variabilidade espacial é se os parâmetros estimados em uma variável são estatisticamente diferentes entre si na região (Leung *et al.*, 2000). O que implica na necessidade de testar a hipótese de variabilidade espacial para justificar o modelo e para possibilitar uma melhor análise dos resultados locais. Alguns autores propuseram testes diferentes ao longo da literatura relacionada à análise espacial de variáveis (Brunsdon *et al.*, 1996; Leung *et al.*, 2000; Nakaya *et al.*, 2014; Oshan *et al.*, 2019).

Nakaya *et al.* (2014) sugeriram testes de comparações entre as métricas dos modelos *GWR* e os modelos não espaciais. São eles a diferença do AICc ser de -2 ou menor, da diferença do pseudo R<sup>2</sup> (Nakaya *et al.*, 2014), a diferença de graus de liberdade e diferenças de graus de liberdade e desvio das variáveis. Porém, novamente, a interpretação dessas diferenças ou é arbitrária, ou é questionável por não considerar a escala das métricas. Enquanto isso, o teste proposto por Leung *et al.* (2000), mesmo com as melhorias sugeridas por Fotheringham *et al.* (2002), apresenta três complicações: a formulação do teste assume uma variável com distribuição normal; falta de vantagens frente ao teste Monte Carlo proposto por Fotheringham *et al.* (2002), com as mesmas desvantagens; o teste de Leung *et al.* (2000) não está prontamente disponível nas ferramentas computacionais consultadas.

Outro teste da literatura foi proposto por Brunsdon *et al.* (1996) e Fotheringham *et al.* (2002) e implementado no programa MGWR (Oshan *et al.*, 2019). Nele, após a execução do modelo *GWR* original, são feitas permutações nas coordenadas, criando bancos de dados (por padrão mil) espacialmente aleatórios. Neste contexto, permutar as coordenadas é trocar as

coordenadas das observações, de forma aleatória e sem repetição. Por exemplo: a observação 30 assume os valores das variáveis da observação 71, a observação 71 assume os valores das variáveis que antes eram da observação 256, assim em diante. Feito isso, o modelo *GWR* é aplicado em cada uma das simulações. Além da permutação das coordenadas, também é selecionada uma nova largura de banda, dada a expectativa de aleatoriedade espacial (Fotheringham *et al.*, 2022). Depois, o desvio padrão do modelo de referência é comparado aos desvios padrão dos modelos simulados. O pseudo p-valor do teste é a posição do desvio padrão do modelo original neste *rank*. Por exemplo, se uma variável do modelo tem o terceiro menor desvio padrão dentre mil simulações, seu pseudo p-valor é 0.003. Este teste foi selecionado como referência para validação do teste proposto.

Uma ferramenta menos comum, mas ainda válida, para teste da variabilidade espacial é a comparação do intervalo de confiança dos parâmetros do modelo não espacial com os parâmetros do modelo local (Fotheringham *et al.*, 2002; Propastin, 2009), que será utilizada neste trabalho. São usados um modelo de coeficientes estacionários (modelo *logit* não espacial, neste trabalho), modelos espaciais locais permutados e um modelo espacial local não permutado (modelo de referência). Como etapa adicional do teste proposto, são analisados mapas temáticos e testes locais disponíveis utilizados para complementar o teste proposto, visto que resultados agregados, ainda que dentro da econometria espacial, podem não representar bem o contexto local real (Anselin, 1988; Getis e Ord, 1992).

## 2.4 Lacunas de pesquisa

O presente trabalho apresenta uma análise simultânea de efeitos espaciais para um caso de escolha modal, motivada pela literatura consultada. No âmbito da revisão da literatura realizada pelos autores, na maioria dos trabalhos encontrados que abordam efeitos espaciais em escolha modal, os coeficientes estimados são espacialmente estacionários (Lindner e Pitombo, 2018; Assirati e Pitombo, 2021; Mondal e Bhat, 2022). Ou seja, nestes casos a variabilidade espacial não é analisada, exceto no estudo de Nkeki e Asikhia (2019). Neste último estudo, nem todas as variáveis do modelo tiveram sua variabilidade espacial corroborada, porém, como mencionado anteriormente, não foi feito um novo modelo GWLR semi-paramétrico. Considerar tais variáveis espacialmente estacionárias como globais poderia trazer resultados ainda mais realistas (Fotheringham *et al.*, 2002; Nakaya *et al.*, 2014). Assim, é possível verificar esta lacuna importante, relacionada à análise mais detalhada da variabilidade espacial em problemas de escolha modal.

Para abordar esta lacuna, há uma sequência de etapas a serem realizadas para atender certos requisitos, como será melhor discutido no Capítulo 3 (Seções 3.3 a 3.5). Nelas o foco é a contribuição relacionada a um procedimento metodológico que incorpore a análise da dependência e variabilidade espaciais simultaneamente, com enfoque na escolha modal.

A segunda contribuição, também detalhada no Capítulo 3 (Seção 3.6), é a proposta de um teste de variabilidade espacial. O teste proposto, assim como os demais testes encontrados na literatura (Leung *et al.*, 2000; Fotheringham *et al.*, 2002; Nakaya *et al.*, 2014), fornece um p-valor agregado para cada variável. Também tem resultado final associado a um pseudo p-valor em cada variável, como em Fotheringham *et al.* (2002). No entanto, neste trabalho também são apresentados resultados que podem ser utilizados para análises locais da variabilidade espacial.

A Tabela 1 apresenta um sumário dos principais artigos consultados, o tópico principal deles e suas contribuições e lacunas pertinentes, bem como as lacunas relacionadas ao presente trabalho.

Tabela 1 – Sumário dos trabalhos consultados e lacunas de pesquisa

Trabalho	Tópico principal	Contribuições	Lacunas	Lacuna abordada neste trabalho
Ben-Akiva e Lerman (1985)	Modelagem da escolha modal por funções utilidade aleatórias.	Trabalho pioneiro para modelagem da escolha modal.	Não considera o aspecto geográfico na estrutura do modelo.	Análise simultânea de dependência e variabilidade espacial para escolha modal
Assirati e Pitombo (2021)	Modelo com dependência espacial para escolha modal.	Incorporação da dependência espacial por AEDE e geoestatística e modelo logit espacial global.	Modelo de regressão utilizado não incorpora a variabilidade espacial.	
Nkeki e Asikhia (2019)	Modelo <i>GWLR</i> para escolha modal	Incorporação de ambos os efeitos espaciais por um modelo <i>GWLR</i> .	Nem todas as variáveis têm a variabilidade espacial corroborada, mas o modelo <i>SGWLR</i> não é adotado.	
Leung <i>et al.</i> (2000)	Elaboração de testes estatísticos para modelos <i>GWR</i> .	Testes para ganho do <i>GWR</i> frente aos modelos não espaciais.	Não pode ser aplicado em estudos de variável dependente binária ou de contagem. Não possibilita elaborar testes desagregados.	Proposta de teste de variabilidade espacial, com resultados desagregados
Nakaya <i>et al.</i> (2014)	Ferramenta computacional para modelos <i>GWR</i> semi-paramétricos.	Teste simples e, fácil, não paramétrico.	Comparações arbitrarias de métricas de ajuste do modelo: faltam valores de corte claros. Não possibilita testes desagregados.	
Fotheringham <i>et al.</i> (2002)	Fundamentos dos modelos <i>GWR</i> .	Teste não paramétrico, não é arbitrário, fácil compreensão.	Não possibilita elaborar testes desagregados.	

Fonte: elaborado pelo autor.

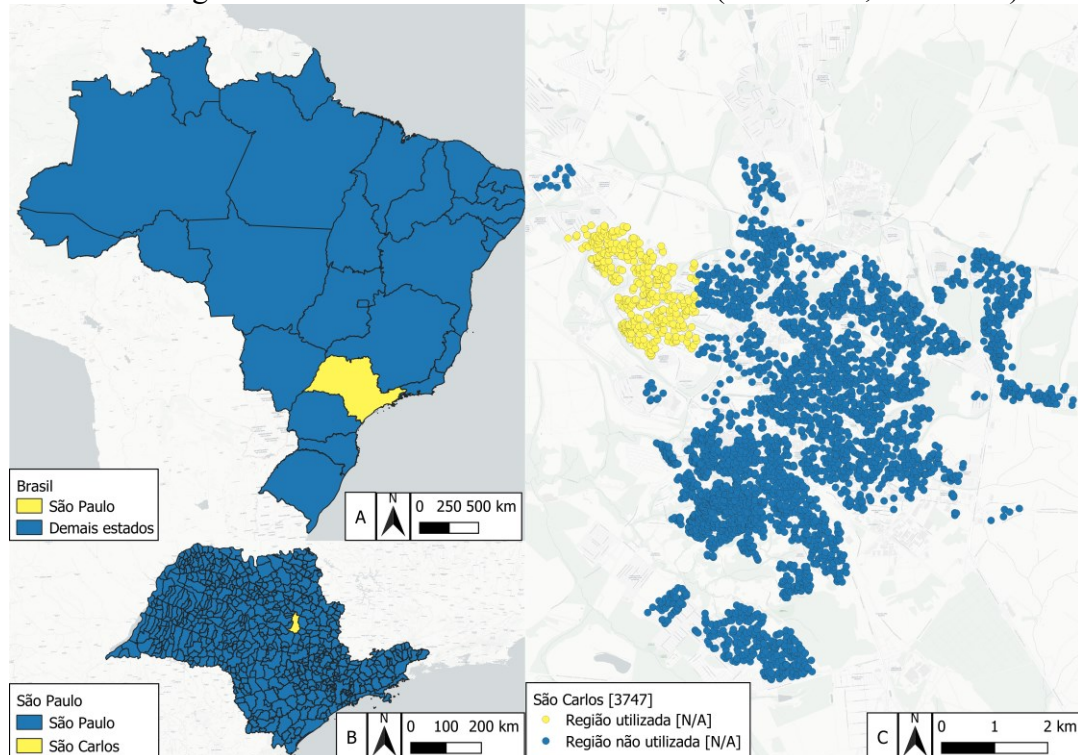
## CAPÍTULO 3: MATERIAIS E MÉTODO

*Este capítulo descreve os principais materiais deste estudo (dados e pacotes computacionais). O fluxograma metodológico é apresentado e descrito, bem como o procedimento associado à proposta do teste de variabilidade espacial é detalhado nas seções subsequentes.*

### 3.1 Dados

Neste trabalho foi utilizado um banco de dados sintético, com observações extrapoladas dos dados originais, provenientes da Pesquisa Origem Destino, realizada em São Carlos (SP) em 2007/2008 (Rodrigues da Silva, 2008). O banco de dados, corresponde à região de Santa Felícia, na cidade de São Carlos, São Paulo (Figura 2) e possui observações de 87 pontos georreferenciados e duas escolhas modais (Ônibus ou Automóvel). Nota-se que estes 87 pontos são uma sub-amostra dos pontos em amarelo da Figura 2C. Este estudo de caso real foi escolhido por ter dependência espacial corroborada em trabalhos anteriores (Pitombo *et al.*, 2015; Gomes *et al.*, 2016). A Tabela 2 e Tabela 3 mostram as medidas descritivas das variáveis quantitativas e frequência de ocorrência das variáveis qualitativas do banco de dados original, respectivamente.

Figura 2 – Banco de dados – Santa Felícia (São Carlos, São Paulo)



As regiões em amarelo mostram, progressivamente, a região do banco de dados utilizado.

Fonte: Rodrigues da Silva (2008).

Tabela 2 – Medidas descritivas das variáveis quantitativas (87 observações)

Variável	Média	Mediana	Desvio Padrão	Mínimo	Máximo
Viagens <sup>1</sup>	2.16	2	0.61	1	4
Idade	41.97	40	12.09	20	78
Automóveis no domicílio <sup>2</sup>	0.92	1	0.67	0	3
Motos no domicílio <sup>2</sup>	0.23	0	0.47	0	2

<sup>1</sup>: Quantidade de viagens diárias no domicílio. <sup>2</sup>: Quantidade de carros e quantidade de motos no domicílio, respectivamente.

Tabela 3 – Frequências das categorias das variáveis qualitativas (87 observações)

Variável	Frequência categoria 0	Frequência categoria 1
Escolha modal <sup>1</sup>	0.66	0.34
CNH <sup>2</sup>	0.81	0.19
Estuda <sup>3</sup>	0.93	0.07
Trabalha <sup>4</sup>	0.77	0.23
Gênero <sup>5</sup>	0.79	0.21
Grau de instrução <sup>6</sup>	0.49	0.51

<sup>1</sup>: Escolha modal entre transporte público (0), e carro (1); <sup>2</sup>: CNH (Posse de Carteira Nacional de Habilitação): Não possui (0), Possui (1); <sup>3</sup>: Estuda: Não (0), Sim (1); <sup>4</sup>: Trabalha: Não (0), Sim (1). <sup>5</sup>: Gênero: Masculino (0), Feminino (1); <sup>6</sup>: Grau de instrução: ensino médio incompleto ou inferior (0), ensino médio completo, ensino superior e pós-graduação (1).

O banco de dados precisou ser alterado para incluir mais observações para o modelo espacial local, pois a quantidade original era insuficiente para uma boa calibração do modelo (Fotheringham *et al.*, 2002; Fotheringham e Oshan, 2016). As coordenadas das observações acrescentadas foram escolhidas aleatoriamente na proximidade dos dados observados, com uma densidade aproximadamente homogênea.

Para o procedimento proposto é necessário que a base de dados: possua um número satisfatório de observações; que todas as observações estejam georreferenciadas; a variável de escolha modal seja binária (em números 0 ou 1) e presença de todos os valores das variáveis em cada observação. Sugere-se o uso de coordenadas UTM para melhor performance computacional.

Os valores das variáveis nas novas observações foram obtidos por uma interpolação inversamente ponderada à distância simples (*Inverse Distance Weighting, IDW*, Shepard, 1968), com potência de decaimento dois. Os valores das variáveis nas novas observações foram obtidos de formas diferentes, a depender do tipo da variável.

Para variáveis quantitativas inteiras ou bem discretizadas (como a idade), o valor adotado foi o *IDW* arredondado. Para variáveis binárias, salvo a escolha modal, a fórmula utilizada manteve a frequência relativa igual ao banco de dados original. Na variável de escolha modal, foi dado o valor 1 (uso do carro), caso o valor do *IDW* da observação estivesse acima da mediana de todas as novas observações. Isto foi adotado para amenizar a discrepância da escolha entre os dois modos.

O banco de dados adaptado do original (Santa Felícia, São Carlos – SP) possui 601 observações e 9 variáveis. A Tabela 4 apresenta cada variável quantitativa e medidas descritivas básicas. Já a Tabela 5 apresenta o percentual de observações das categorias das variáveis qualitativas, com todas tendo apenas duas categorias. A Figura 3 mostra a distribuição espacial do banco de dados adaptado na variável escolha modal, com o valor 0 para transporte público e valor 1 para transporte por automóvel particular.

Tabela 4 – Medidas descritivas básicas do banco de dados das variáveis quantitativas (601 observações)

Variável	Média	Mediana	Desvio Padrão	Mínimo	Máximo
Viagens	2.05	2	0.317	1	4
Idade	40.6	41	6.70	20	78
Auto <sup>1</sup>	0.943	1	0.387	0	3
Moto <sup>1</sup>	0.115	0	0.329	0	2

<sup>1</sup>: Quantidade de automóveis no domicílio, Quantidade de motocicletas no domicílio, respectivamente.

Fonte: elaborada pelo autor.

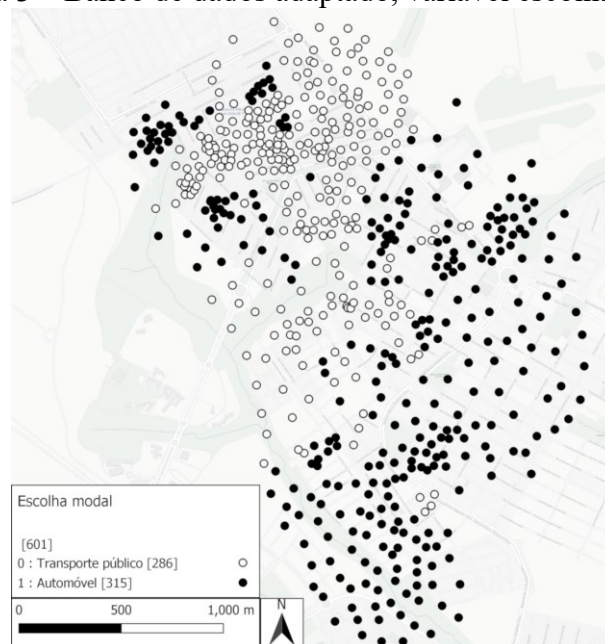
Tabela 5 – Percentuais de observações das categorias das variáveis binárias

Variável	Frequências categoria 0	Frequências categoria 1
Escolha <sup>1</sup>	0.476	0.524
CNH <sup>2</sup>	0.198	0.802
Estuda <sup>3</sup>	0.070	0.930
Trabalha <sup>4</sup>	0.084	0.906
Gênero <sup>5</sup>	0.210	0.790
GI <sup>6</sup>	0.494	0.506

1: Escolha modal entre transporte público (0), e carro (1); 2: CNH (Posse de Carteira Nacional de Habilitação): Não possui (0), Possui (1); 3: Estuda: Não (0), Sim (1); 4: Trabalha: Não (0), Sim (1). 5: Gênero: Masculino (0), Feminino (1); 6: Grau de instrução: ensino médio incompleto ou inferior (0), ensino médio completo, ensino superior e pós-graduação (1).

Fonte: elaborada pelo autor.

Figura 3 – Banco de dados adaptado, variável escolha modal



Fonte: Adaptado de Rodrigues da Silva (2008).

Com o objetivo de comparar a amostra original (87 observações) com a base de dados sintética, proveniente da original (601 observações), foram realizados testes estatísticos. Na Tabela 6 e Tabela 7 são apresentados os testes comparativos de cada variável disponível e os respectivos p-valores. Para as variáveis numéricas, foi aplicado o teste de Wilcoxon-Mann-Whitney (Hollander *et al.*, 1999) e para variáveis binárias o teste Chi-quadrado (Pearson, 1900). Pela Tabela 6, apenas a variável “posse de motocicletas” teve uma média estatisticamente diferente. No entanto, esta variável não foi usada na modelagem posterior. Para as variáveis binárias (Tabela 7), observou-se similaridade entre todos os pares de variáveis que compõem a amostra original (87 observações) e a sintética (601 observações). Assim, é verificado que a base de dados sintética conserva características básicas da amostra original.

Tabela 6 – Teste de Wilcoxon-Mann-Whitney para comparações entre as variáveis numéricas que compõem a amostra original e sintética

Variável	P-valor
Viagens	0.186
Idade	0.814
Auto	0.271
Moto	0.009

Fonte: elaborada pelo autor.

Tabela 7 – Teste Qui-quadrado para comparações entre as variáveis binárias que compõem a amostra original e sintética

Variável	P-valor
Escolha modal	0.626
CNH	0.228
Estuda	0.397
Trabalha	0.730
Gênero	0.731
Grau de instrução	0.667

Fonte: elaborada pelo autor.

As variáveis adotadas foram: viagens, idade e quantidade de carros no domicílio (Auto) para calibração dos modelos. Os critérios para escolha foram: pertinência da variável para a escolha modal, ajuste da variável no modelo não espacial, presença de efeitos espaciais pelas AEDE e ajuste da variável no modelo espacial local. As variáveis independentes selecionadas, todas quantitativas, foram normalizadas, para possibilitar que os resultados dos modelos fossem analisados na mesma escala. É recomendado que essa ou outras transformações sejam feitas diretamente no banco de dados e não pelas opções das ferramentas disponíveis no *software* utilizado para calibração dos modelos locais, pois algumas podem transformar também a variável dependente (caso do MGWR versão 2.2.1), invalidando o processo para modelos logit ou poisson.



Nota-se que podem existir diferenças entre os ajustes do modelo não espacial e espacial local e ausência de efeitos espaciais nas variáveis iniciais. Assim, o ponto de partida pode ser pelo modelo não espacial, com seleção das variáveis de acordo com a literatura ou particularidades do caso de estudo, ou pelas AEDE, caso o pesquisador queira encurtar este processo iterativo e focar apenas nas variáveis com dependência espacial.

As variáveis também podem ser selecionadas de acordo com a variabilidade espacial, porém isto não é recomendado por dois motivos. Primeiro, que a hipótese variabilidade espacial pode ser rejeitada. Este é um resultado válido e importante pois, por mais que eventualmente contraintuitivo ou em desacordo com a literatura, pode ser verossímil e útil para explicar particularidades do caso de estudo. O outro motivo é que existem modelos semiparamétricos de regressão logística geograficamente ponderada (*SGWLR*) que podem trabalhar com variáveis estacionárias e não estacionárias simultaneamente (Nakaya *et al.*, 2014, Nakaya, 2015). Logo, parte das variáveis escolhidas ser estacionária não impossibilita a calibração de um modelo *GWLR*. Portanto, o procedimento propõe que a seleção do conjunto de variáveis do(s) modelo(s) seja feita na etapa não espacial e alterada na etapa de AEDE, caso necessário.

### **3.2 Modelo logit não espacial**

O modelo logit não espacial, ou modelo logit tradicional, é um modelo de regressão para variáveis dependentes qualitativas, como o caso de escolha modal. A probabilidade de cada escolha é calculada pela função utilidade aleatória, que quantifica a atratividade de cada opção para cada indivíduo (Ben-Akiva e Lerman, 1985).

O critério de utilidade é um tipo de regra de escolha compensatória, onde um baixo desempenho em uma característica pode ser compensado por um bom desempenho em outros aspectos, mesmo que sejam de diferentes naturezas (Ben-Akiva e Lerman, 1985).

O processo de escolha modal é representado através de diversas características em diferentes níveis: dos usuários, da viagem, do espaço construído e dos modos disponíveis. Portanto, trata-se de um assunto multidisciplinar, englobando teorias microeconômicas e da psicologia comportamental (Ben-Akiva e Lerman., 1985). A Equação 1 descreve a utilidade do modo  $m$  e a Equação 2 a probabilidade de escolha do modo de transporte  $m$  dentre todos os  $M$  modos disponíveis (Ben-Akiva, 1973; Ben-Akiva e Lerman, 1985). Visto que não é o objetivo deste documento discutir ou melhorar o processo de calibração dos modelos apresentados, sugere-se McFadden (1974), Manski (1977), Ben-Akiva (1973) Ben-Akiva e Lerman (1985) para detalhamentos conceituais.

$$U_{m,i} = \sum_{j=1}^{k_m} x_i * \beta_j + e_{m,i} = V_{m,i} + e_{m,i} \quad (1)$$

Onde:

$U_{m,i}$ : função de utilidade percebida na observação  $i$ , para o modo de transporte  $m$ .

$\beta_j$ : coeficientes do subconjunto  $k_m$  de variáveis selecionadas para caracterizar a escolha do modo de transporte  $m$ .

$V_{m,i}$ : função de utilidade verdadeira do indivíduo  $i$  para o modo  $m$

$e_{m,i}$ : termo de erro do indivíduo  $i$  para o modo  $m$ .

$$P_{m,i} = \frac{e^{V_{m,i}}}{\sum_{n=1}^M e^{V_{n,i}}} \quad (2)$$

Onde:

$P_{m,i}$ : probabilidade de escolha do modo  $m$ , para o indivíduo  $i$ , no conjunto de  $M$  possíveis modos.

Dado que esta etapa não contempla os efeitos espaciais, pode ocorrer que ela seja cíclica com as etapas de AEDE e modelagem espacial. Ou seja, um modelo logit não espacial pode ser adequado por si só. No entanto, havendo indicações de dependência e/ou variabilidade espacial, isto pode demandar novas análises e implementos nas estimativas.

No processo de modelagem, é importante a aplicação de modelos não espaciais. Primeiro que a comparação entre resultados espaciais e não espaciais é importante para o estudo de caso. E, visto que a literatura ainda não está maturada com modelos espaciais locais para este tipo de aplicação, é importante justificar os esforços adicionais para aplicação de um modelo como o *GWLR*.

Neste trabalho, todas as variáveis foram declaradas na função utilidade da opção carro (opção binária 1). Isto foi feito para ter uma conformidade com o recurso computacional MGWR, pois nele não é possível declarar ambas as funções utilidade, conforme a Equação 4 (Seção 3.4).

### 3.3 Análises Exploratórias de Dados Espaciais (AEDE)

A etapa de AEDE, como discutido anteriormente, é o primeiro passo das análises espaciais, servindo como arcabouço para os modelos espaciais. Sua importância no procedimento proposto é analisar a presença de efeitos espaciais para cada variável e, se existem, como são os padrões locais. Por isso, foram selecionadas ferramentas com resultados locais e globais.

A aplicação de AEDE depende do tipo de variável. Para variáveis quantitativas, o *Univariate Local Moran's I* (Anselin, 1995) é adequado e amplamente adotado na literatura de diversas áreas, e fornece resultados locais e globais.

O índice I de Moran local apresenta resultados associados à tendência de dependência espacial global, índices I de Moran locais, quadrante com o valor global e valores locais do índice e mapa de agrupamentos de acordo com valor do I de Moran local (Anselin *et al.*, 2010). Os grupos dos agrupamentos podem ser alto-alto (valor alto da observação  $i$  e vizinhos), alto-baixo, baixo-baixo e baixo-alto. Quanto mais o módulo do índice I se aproxima de um, maior a dependência espacial estimada, e quanto mais próxima de zero menor a dependência espacial (Anselin, 1995).

Para variáveis binárias, são necessárias AEDE específicas. Neste estudo, selecionou-se o *Local Join Count* (Anselin e Li, 2019) para a variável de escolha modal. O *LJC* calcula, para cada observação  $i$  de valor binário 1, a quantidade de vizinhos com o mesmo valor de  $i$ , e compara esta contagem com a quantidade esperada sob a hipótese de aleatoriedade espacial (proporção global do valor binário 1). Caso a diferença entre essas duas contagens seja estatisticamente significativa, rejeita-se a hipótese nula de aleatoriedade espacial, corroborando a dependência espacial da variável (Boots, 2003; Anselin e Li, 2019). Seu resultado é um pseudo  $p$ -valor local, sem índice global pré-determinado.

É importante notar que todas as AEDE dependem da definição de uma matriz de pesos espaciais. A largura de banda, o critério que define os vizinhos de cada observação, usualmente é definida por uma distância, ou por um número de vizinhos mais próximos (*k nearest neighbours*, *KNN*) e no caso das matrizes de peso espaciais também podem ser usadas regras de contiguidade. Contudo, para as AEDE, não há um código computacional que indique a largura de banda ótima a ser utilizada (Almeida, 2012). Portanto, recomenda-se a adoção de uma largura de banda nem pequena demais, para não desconsiderar vizinhos próximos importantes, nem grande demais, para não acrescentar ruído (Fotheringham *et al.*, 2022). É interessante que após a definição da largura de banda do modelo *GWLR*, sejam refeitas as AEDE com a nova largura de banda e, se possível, comparação com os novos resultados das primeiras AEDE.

Neste estudo foi utilizada, na etapa de AEDE, uma largura de banda inicial de 20 vizinhos mais próximos, número relativamente pequeno frente à quantidade de vizinhos disponíveis, porém grande o suficiente para representar uma região de vizinhos. Ressalta-se que, mesmo neste banco de dados sintético, a relação da dependência espacial (se existir) é desconhecida. Esta limitação é um fator comum da econometria espacial, e requer cautela e

pensamento crítico do modelador (Goodchild e Janelle, 2010;). Assim, não se sabe *a priori* qual critério de vizinhança representa ou se aproxima melhor da estrutura de vizinhança real, tampouco o valor da largura de banda mais adequado. Para uma discussão mais aprofundada, que foge do escopo desta dissertação, sugere-se Kopczewska e Kubara (2020).

### 3.4 Modelo espacial local e teste de variabilidade espacial de Monte Carlo

Para a etapa de modelagem espacial local, foi adotado o modelo *GWR*, elaborado por Brunson *et al.* (1996) e Fotheringham *et al.* (2002), com um modelo logístico (*GWLR*). O modelo *GWR* funciona como um agregado de regressões, ou regressão de janela móvel, em que cada observação tem sua regressão, considerando apenas as observações vizinhas, inversamente ponderadas pela distância. O critério de corte de vizinhança é dado pela largura de banda, observações além dessa distância (geográfica ou de vizinhos mais próximos) não são consideradas como vizinhas (Brunson, *et al.*, 1996; Fotheringham *et al.*, 2002). Há também a função *kernel*, adotada nos *GWR*, que prevê a diminuição da influência de observações vizinhas mais distantes (Fotheringham *et al.*, 2002). Existem diferentes tipos de função *kernel* disponíveis, cada uma com um comportamento de decaimento diferente (Oshan *et al.*, 2019). Neste trabalho, foi adotada a função *kernel* bi-quadrada, vide Equação 3 (Fotheringham *et al.*, 2002), e a largura de banda por *KNN*.

$$w_{ij} = \left[ 1 - \left( \frac{d_{ij}}{b} \right)^2 \right]^2 \text{ para } d_{ij} < b, w_{ij} = 0 \text{ caso contrário} \quad (3)$$

Onde:

$w_{ij}$ : peso kernel na observação  $i$  para a observação vizinha  $j$ .

$d_{ij}$ : distância entre  $i$  e  $j$ .

$b$ : distância de  $i$  até o último vizinho mais próximo considerado.

Com as funções *kernel*, atende-se à dependência espacial, pois apenas as observações vizinhas influenciam os resultados, com as mais próximas influenciando mais a regressão local (Fotheringham *et al.*, 2002; Oshan *et al.*, 2019). E dado que cada observação possui sua própria regressão, os coeficientes estimados são locais, respeitando a premissa de variabilidade espacial (Fotheringham *et al.*, 2002).

Para processar o modelo, é necessário declarar as coordenadas, o conjunto de variáveis, o tipo de variável dependente, a largura de banda e o tipo de largura de banda (distância fixa ou variável por vizinhos mais próximos). Também é necessário, para este procedimento, a ativação das opções do teste de Monte Carlo e testes de multicolinearidade. No caso do modelo *GWLR*, as variáveis estimadas estão de acordo com a Equação 4 (Li *et al.*, 2019; Oshan *et al.*, 2019).

Por esta equação, percebe-se que a variável dependente estimada é a probabilidade da ocorrência do valor 1 da variável depende observada. Ainda, as variáveis explanatórias não estão em uma função utilidade propriamente dita, como no modelo logit não espacial (vide Equação 1) mas é semelhante a ela para a categoria de valor 1.

$$\hat{Y}_i \sim \text{Bernoulli}[p_i]$$

$$\text{logit}[p_i] = \sum_k \hat{\beta}_{k,i} * x_{k,i} \quad (4)$$

Onde:

$\hat{Y}_i$ : probabilidade estimada da escolha modal de valor 1 (automóvel) do indivíduo  $i$

$\hat{\beta}_{k,i}$ : coeficiente estimado da variável  $k$  na observação  $i$

$x_{k,i}$ : variável  $k$  na observação  $i$

Recomenda-se que a largura de banda seja por um dos seguintes critérios: a ótima calculada pelo programa (usualmente por critério de minimização do AICc ou BIC do modelo); a mínima que possibilite a execução do modelo simultaneamente com o teste de variabilidade espacial ou uma suficientemente grande para conseguir uma quantidade desejada de estimadores estatisticamente significativos, sem comprometer muito a qualidade do ajuste do modelo.

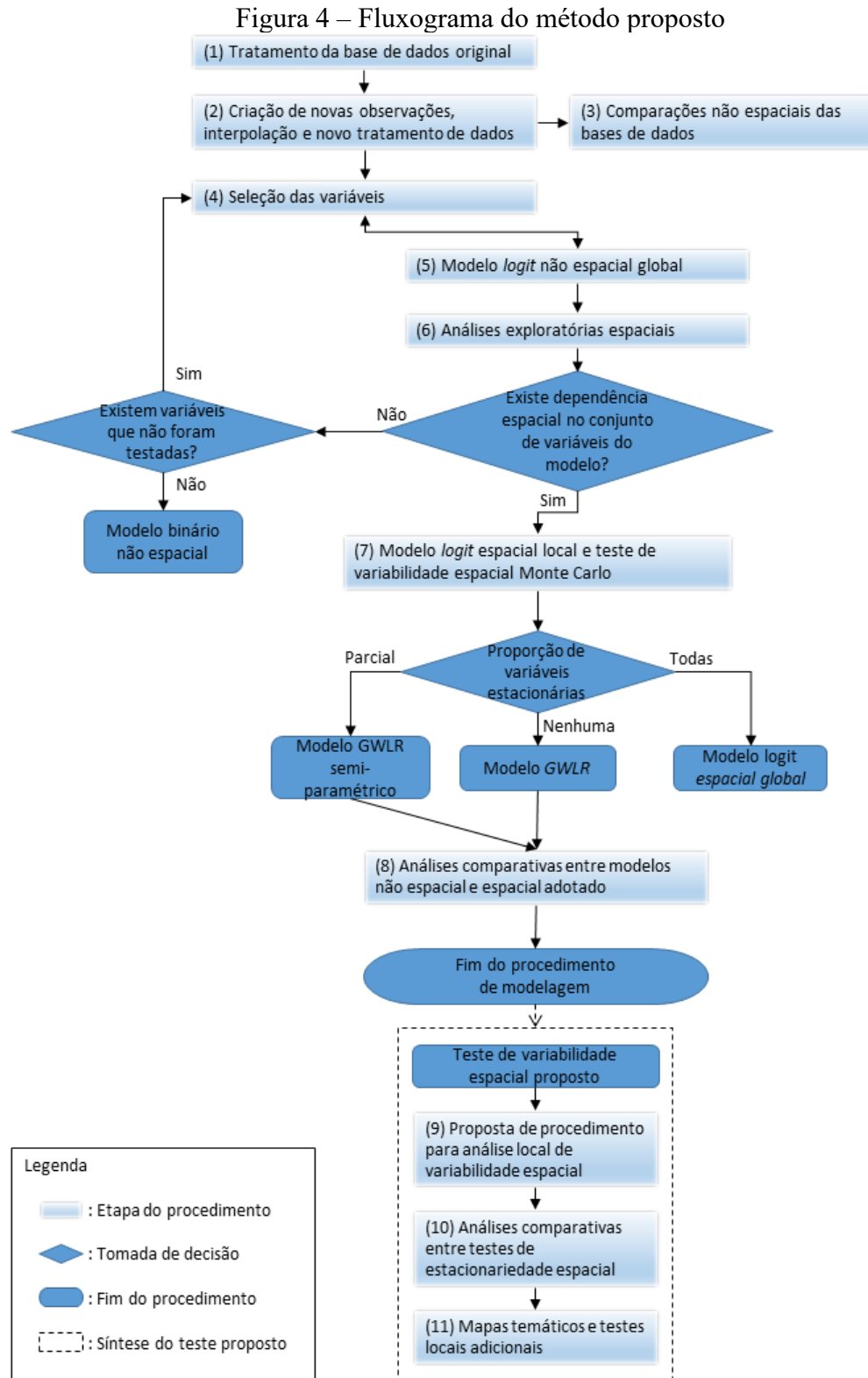
O uso da largura de banda mínima permitida pode ser interessante pois, quanto menor ela é, maior a influência esperada dos efeitos espaciais (Fotheringham *et al.*, 2022). Sob uma ótica puramente computacional dos efeitos espaciais e dos modelos *GWR*, um bom conjunto de variáveis tende a minimizar a largura de banda escolhida frente a outras opções de conjuntos, pois requer menos informações para executar o código e fornece estimadores com menor variância (Fotheringham *et al.*, 2022). Fotheringham *et al.* (2022) também comentam que esta abordagem pode resultar em maior viés dos estimadores e menor quantidade de observações com coeficientes estimados estatisticamente significativos. Uma compensação semelhante ocorre com as AEDE nos índices de dependência espacial, conforme resultados da Seção 4.2 (Capítulo 4).

A execução de um teste de variabilidade espacial é necessária para o teste de hipótese de estacionariedade espacial para os estimadores, premissa dos modelos *GWR* (Fotheringham *et al.*, 2002). Também é requerida, neste procedimento, para determinar o tipo de modelo correto a ser adotado, a depender de quantas variáveis têm a hipótese de estacionariedade espacial rejeitada. Um modelo mais adequado ao resultado do teste de variabilidade implica em um modelo mais verossímil e usualmente com melhor ajuste (Fotheringham *et al.*, 2002; Nakaya *et al.*, 2014; Nakaya, 2015).

O teste de Monte Carlo para variabilidade espacial, proposto por Brunson *et al.* (1996) e Fotheringham *et al.* (2002), opera pelo rearranjo das coordenadas das observações em um número grande de iterações. O teste compara se a diferença do desvio padrão dos estimadores entre os diferentes modelos das iterações e o modelo sem permutação podem ter sido ao acaso ou não. Ele tem como hipótese nula a estacionariedade espacial, e como hipótese alternativa a variabilidade espacial do parâmetro associado à variável (Fotheringham *et al.*, 2002). Conforme Fotheringham *et al.* (2002) e Oshan *et al.* (2019), uma desvantagem é a exigência computacional do teste, que pode tornar o código inviável em modelos com muitas observações e/ou variáveis.

### 3.5 Procedimento metodológico

A Figura 4 sintetiza todo o procedimento metodológico, ilustrando a sequência de etapas e as tomadas de decisão envolvidas. As etapas estão detalhadamente descritas em seguida.



Fonte: elaborado pelo autor.

A primeira etapa a ser realizada é o tratamento dos dados, para que estejam conforme os requisitos mencionados na Seção 3.1, bem como facilidade de uso e formatação adequada para a execução das ferramentas computacionais.

Com o banco de dados tratado, aplica-se o modelo logit não espacial tradicional. É um passo necessário para conhecer as relações entre as variáveis independentes e a dependente, e se os resultados estão de acordo com características da área de estudo e relações observadas na literatura. Os resultados também são usados para comparação com o modelo espacial local.

A etapa de AEDE é a primeira chave para saber se os efeitos espaciais estão presentes nas variáveis independentes e na dependente. Caso a variável dependente e pelo menos uma variável independente não tenham a hipótese de dependência espacial corroborada, o procedimento acaba nesta etapa, com a escolha do modelo logit não espacial como mais adequado. Vale ressaltar que, nesta etapa, também são utilizadas as probabilidades de escolha modal, obtidas no modelo logit não espacial, para cálculo do índice de Moran, o qual é originalmente aplicado para variáveis numéricas contínuas.

Na presença dos efeitos espaciais, é possível entender melhor as distribuições de valores de cada variável no banco de dados e, conseqüentemente, agrupamentos espaciais e localização de valores limítrofes (*outliers*). Os resultados podem indicar mudanças nas variáveis selecionadas do modelo, ou alternativas para serem testadas em paralelo com o conjunto original. Porém, como comentado anteriormente, as AEDE são normalmente limitadas para análises uni ou bivariadas.

Prosseguindo com a corroboração da dependência espacial, a quarta etapa é a aplicação de um modelo *GWLR*. Após a execução do modelo, conforme explicado na Seção 3.4, avalia-se a quantidade de variáveis independentes estacionárias, conforme teste de Monte Carlo ou equivalente, sendo possível três cenários.

Para o cenário em que todas as variáveis são estacionárias, deve ser utilizado um modelo espacial global, visto que os efeitos espaciais foram corroborados na etapa de AEDE e que os modelos espaciais globais trabalham sob hipótese de estacionariedade espacial. Dada a variedade de modelos disponíveis (LeSage, 1999; Almeida, 2012) e que o mais recomendado varia de acordo com o caso de estudo, a escolha do modelo logit espacial global não é abordada neste trabalho.

Nos casos em que parte das variáveis é estacionária e parte é espacialmente variável, adota-se os modelos de regressão logística geograficamente ponderada semiparamétrica (*SGWLR*). A diferença deste tipo de modelo para o *GWLR* é que são informados dois grupos de



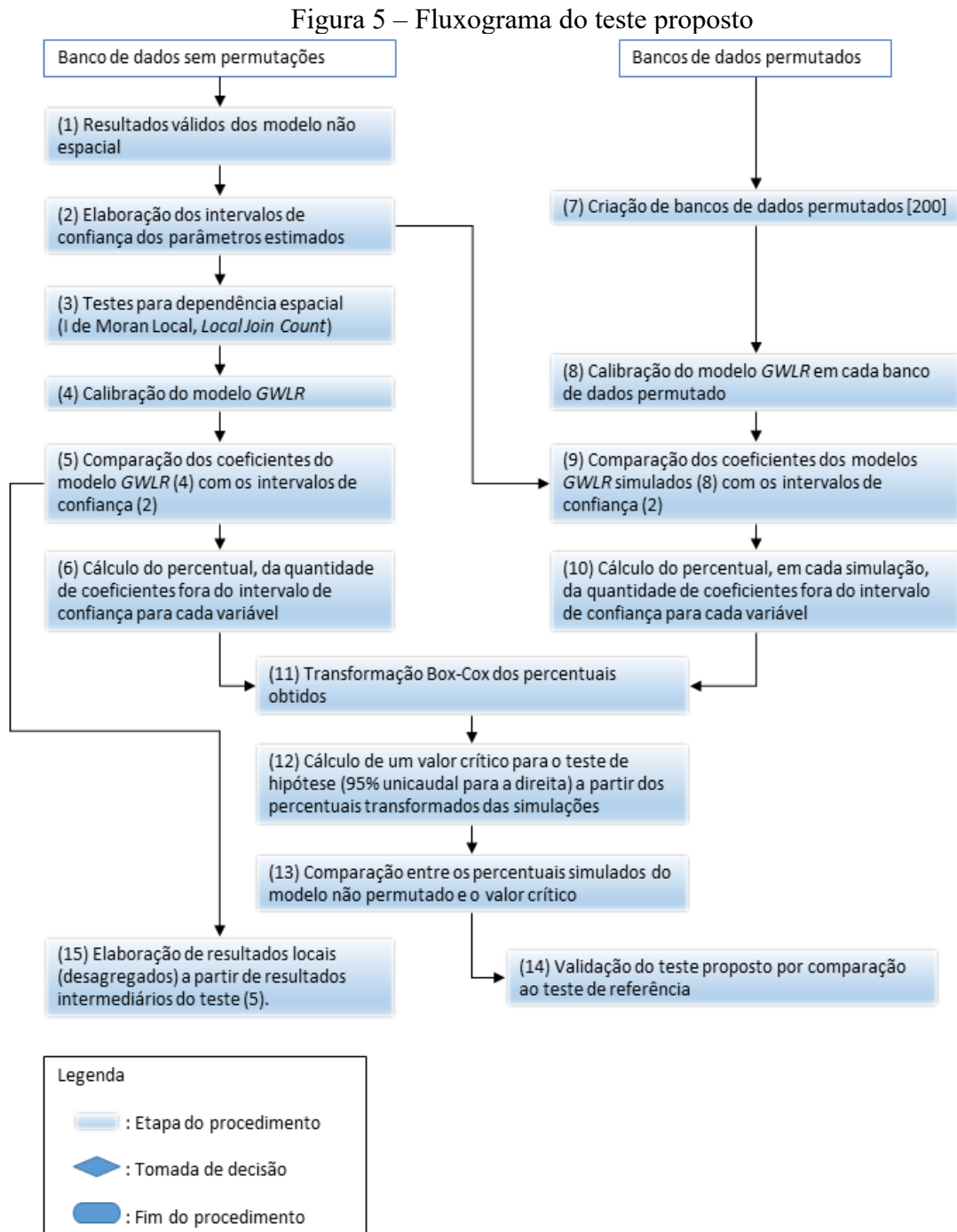
variáveis, um local e um global, respeitando o resultado do teste de variabilidade espacial. As cautelas descritas na Seção 3.4 devem ser observadas nas variáveis locais.

O terceiro cenário ocorre quando nenhuma das variáveis é espacialmente estacionária, com aplicação do modelo *GWLR*, conforme Seção 3.4. Nota-se que a variabilidade espacial de uma variável é um resultado derivado do modelo utilizado, podendo mudar com a inclusão e exclusão de outras variáveis e alteração da largura de banda.

A última etapa do procedimento principal é a comparação dos resultados obtidos da etapa anterior com os resultados do modelo logit não espacial, por métricas de ajuste global e matrizes de confusão. Os ganhos explanatórios dos resultados locais também devem ser considerados para estas comparações. As etapas 9 a 11 da Figura 4, que tratam do teste proposto, são melhor detalhadas na próxima seção.

### 3.6 Teste de variabilidade espacial proposto

Para o teste proposto, foram criadas 200 permutações, sem repetições, do conjunto de variáveis a partir do banco de dados original para calibração de modelos *GWLR* sob a hipótese de aleatoriedade espacial. O procedimento do teste proposto é ilustrado na Figura 5, que corresponde às etapas 9 a 11 do fluxograma da Figura 4.



Fonte: elaborado pelo autor.

Este teste tem os seguintes requerimentos: (i) os parâmetros estimados no modelo não espacial são significativos; (ii) presença de dependência espacial corroborada por AEDE para, pelo menos, a variável dependente e uma variável independente; (iii) os coeficientes obtidos pelo modelo espacial local não possuem ordens de grandeza muito discrepantes entre si (*outliers* são permitidos); (iv) não há multicolinearidade considerável entre variáveis independentes e parâmetros estimados.

A hipótese nula do teste é a aleatoriedade espacial dos coeficientes estimados do modelo local, com a hipótese alternativa de não aleatoriedade espacial dos coeficientes estimados. Assim, a hipótese alternativa corrobora a variabilidade espacial. As subseções subsequentes descrevem cada etapa.

### **3.6.1. Modelagem não espacial e obtenção dos intervalos de confiança dos parâmetros**

Após a calibração do modelo não espacial, verifica-se se os parâmetros estimados das variáveis são significativos (1), atendendo ao primeiro requerimento. Também é realizado teste de multicolinearidade das variáveis independentes neste caso. Em seguida, são obtidos os intervalos de confiança dos parâmetros de cada variável (2), considerando valores de erro padrão e suposição de distribuição normal dos parâmetros estimados.

### **3.6.2. Testes univariados de dependência espacial**

Para o segundo requerimento, são realizadas AEDE para teste de dependência espacial (3). Para variáveis quantitativas contínuas aplica-se o I de Moran Local (Anselin, 1995), e para variáveis binárias o *Local Join Count* (Anselin e Li, 2019). Caso a variável dependente e pelo menos uma variável independente possuam dependência espacial, o segundo requerimento do teste é atendido. Senão, o mais apropriado é adotar um modelo não espacial, sem cabimento para a aplicação do teste.

### **3.6.3. Modelo GWLR no banco de dados sem permutação**

Em seguida, aplica-se o modelo *GWLR* em uma largura de banda adequada para o caso de estudo (4). Recomenda-se a menor largura de banda capaz de calibrar o modelo, e caso a densidade das observações seja heterogênea o uso do critério de *k* vizinhos mais próximos (*KNN*). Se coeficientes de uma mesma variável forem de ordens de grandeza muito diferentes, o indicador do teste no modelo espacial local ou o valor crítico obtido podem não ser confiáveis. Visto que as variáveis são testadas separadamente pelos seus parâmetros estimados, é

necessária a avaliação da multicolinearidade nos coeficientes dos modelos (salvo os simulados). Caso não existam indícios de multicolinearidade significativa o último pré-requisito para aplicação do teste é atendido. Neste trabalho a métrica *VIF* (*Variance Inflation Factor*) foi adotada para avaliar a multicolinearidade, com critério de ter a maioria dos *VIF* até 5, e nenhum ou poucos coeficientes com *VIF* entre 5 e 10 (Wheeler, 2007; Oshan *et al.*, 2019).

#### **3.6.4. Comparação dos coeficientes estimados com o intervalo de confiança do modelo não espacial**

Após os resultados do modelo GWLR, os coeficientes estimados são comparados com os intervalos de confiança do modelo não espacial (5). Também é feita a sumarização dos valores binários de dentro e fora do intervalo dos coeficientes estimados em percentuais (6), um para cada variável do modelo.

#### **3.6.5. Criação de bancos de dados permutados e processamento dos modelos *GWLR* permutados**

São simulados duzentos novos bancos de dados com os pares de coordenadas permutados (sem repetição) a partir do banco de dados original (7). O modelo espacial local é calibrado para cada novo banco de dados (8), obtendo modelos em que se espera a ausência de efeitos espaciais. A única diferença do processamento do modelo *GWLR* no banco de dados original e o modelo *GWLR* nos bancos de dados permutados é que a largura de banda foi escolhida pela maximização do *AICc*, deixando-a variável entre a largura de banda do modelo de referência e o máximo de vizinhos mais próximos (N-1). Isso é necessário pois, em bancos de dados com aleatoriedade espacial, espera-se que a largura de banda otimizada pelo *AICc* em modelos *GWR* seja muito grande (Fotheringham *et al.*, 2022). Posteriormente, são comparados os coeficientes locais estimados de todas as variáveis e todas as simulações com o intervalo de confiança dos modelos não espaciais (9), criando um percentual de coeficientes estimados fora de intervalo de confiança (10).

#### **3.6.6. Aplicação do teste proposto**

Com a aplicação da décima etapa, obtém-se uma distribuição de percentuais em cada variável, cada uma com 200 elementos. Aplica-se a transformação para normalidade Box Cox (11) (Box e Cox, 1964). Em seguida, obtém-se o valor de corte para o nível de significância de 5% à direita (12). É um teste unicaudal visto que, pelo intervalo de confiança ser elaborado por

coeficientes estimados de um modelo não espacial, uma porcentagem pequena de parâmetros estimados fora do intervalo de confiança não corrobora a não aleatoriedade espacial.

Assim, o valor de coeficientes fora do IC obtido pelo modelo de referência é comparado ao valor de corte obtido pelas simulações (13). Caso o valor de referência do modelo seja maior do que o de corte, rejeita-se a hipótese nula do teste, por consequência corroborando a variabilidade espacial da variável.

### **3.6.7. Validação do teste proposto**

Para a validação do teste proposto, aplicou-se o teste de variabilidade espacial disponível no programa MGWR (versão 2.2.1, Oshan et al., 2019) e os resultados dos testes foram comparados (14). Esse teste realiza mil simulações (por padrão) por permutação nas coordenadas. Cada banco de dados permutado é modelado com a seleção de uma nova largura de banda, armazenando o erro padrão dos coeficientes das variáveis. Esses erros padrão são postos em ordem crescente, e o p-valor do teste é a posição do modelo de referência nesse *rank*. Assim, um menor erro padrão de uma variável do modelo frente aos modelos permutados corrobora a hipótese de variabilidade espacial.

Nota-se que não é necessário, nem no teste proposto nem no de referência, que a variável dependente seja binária, tampouco que o modelo utilizado seja o *GWLR*, é possível adotar qualquer modelo local. A variabilidade espacial é específica para o conjunto de variáveis independentes escolhido do modelo de referência, e seu resultado não pode ser extrapolado caso este conjunto seja alterado.

### **3.6.8. Testes locais adicionais para variabilidade espacial**

Adicionalmente, sabe-se que os efeitos espaciais podem não ser fidedignos quando representados por métricas agregadas (Getis e Ord, 1992). O que justifica a investigação local da variabilidade, por exemplo, com AEDE para análise da não-aleatoriedade espacial dos estimadores (15) por resultados intermediários do teste proposto, ou de fácil obtenção a partir deles. O mapa temático dos coeficientes locais do modelo de referência dentro e fora do IC do modelo não espacial e o teste *Local Join Count* sobre esse mapa são aplicados para avaliação local da variabilidade espacial.

### 3.7 Recursos computacionais

A Tabela 8 apresenta a etapa metodológica associada ao recurso computacional utilizado.

Tabela 8 – Etapas metodológicas e pacotes computacionais

Etapa metodológica	Software/biblioteca ou ferramenta
Interpolação simples de dados	ArcGIS (ESRI, 2019)
Mapas temáticos	QGIS (QGIS.org, 2022)
Modelo logit não espacial; <i>VIF</i>	<i>R</i> (R Core Team, 2003); pacote CAR (Fox e Weisberg, 2018)
AEDE para variáveis quantitativas	<i>GeoDa – Local Moran's I</i> (Anselin <i>et al.</i> , 2010; Anselin, 1995)
AEDE para variáveis qualitativas	<i>GeoDa – Local Join Count</i> (Anselin <i>et al.</i> , 2010; Anselin e Li, 2019)
Modelo <i>logit</i> espacial local ( <i>GWLR</i> )	<i>MGWR – GWLR</i> (Oshan <i>et al.</i> , 2019; Li <i>et al.</i> , 2019)

Fonte: elaborada pelo autor.

## CAPÍTULO 4: RESULTADOS

*Este capítulo apresenta os resultados da aplicação do método, detalhado no capítulo anterior. Os resultados são descritos na ordem do procedimento proposto para realização do objetivo principal e, posteriormente, do objetivo específico.*

### 4.1 Modelo logit não espacial e obtenção dos intervalos de confiança dos parâmetros

O modelo logit não espacial foi calibrado utilizando-se o *software* R (R Core Team, 2023). A Tabela 9 e Tabela 10 apresentam os resultados globais de ajuste do modelo e os resultados dos parâmetros estimados, respectivamente. A Tabela 10 também apresenta os intervalos de confiança obtidos por parâmetros estimados, considerando suposição de distribuição gaussiano para os parâmetros.

Tabela 9 – Resultados gerais do modelo logit não espacial

Métrica	Valor
AICc	743.27
Log-Likelihood	-367.60
Pseudo R <sup>2</sup> ajustado	0.1161

Fonte: elaborada pelo autor.

Tabela 10 – Síntese dos parâmetros estimados do modelo logit não espacial

Variável	Valor	Erro Padrão	Intervalo de confiança	t-valor	p-valor (95%)	VIF
Intercepto	-5.177	0.828	-6.80, -3.55	-6.254	0	- <sup>1</sup>
Viagens <sup>2</sup>	5.735	1.774	2.26, 9.21	3.233	0.001	1.008
Idade <sup>3</sup>	3.163	0.870	1.46, 4.87	3.635	0	1.056
Auto <sup>4</sup>	6.889	0.591	4.70, 9.08	6.177	0	1.064

<sup>1</sup> O VIF não se aplica ao intercepto. <sup>2</sup> Quantidade de viagens diárias realizadas pelo domicílio. <sup>3</sup> Idade do entrevistado do domicílio. <sup>4</sup> Quantidade de automóveis por domicílio.

Fonte: elaborada pelo autor.

Todos os valores dos coeficientes obtidos estão de acordo com o esperado, ou seja, os resultados dos parâmetros estimados são condizentes com a literatura, comprovando uma relação positiva entre posse de automóveis, idade e número de viagens com a escolha do automóvel (Ibrahim, 2003; Palma e Rochat, 2000; Ben-Akiva et al., 1993). Ainda, como os VIF são menores que 5, não há indícios de multicolinearidade (Wheeler, 2007). No entanto, caso a dependência espacial seja corroborada para a maior parte das variáveis adotadas, é plausível que o modelo logit não espacial não seja o mais adequado, uma vez que as observações podem não ser espacialmente independentes entre si.

## 4.2 Análises exploratórias espaciais

Para as variáveis quantitativas do modelo, a Tabela 11 traz os resultados globais do I de Moran. Nas Figuras Figura 6, Figura 7 e Figura 8 é possível visualizar os mapas dos resultados locais. Foram utilizadas duas larguras de banda por *KNN*: 20, um valor inicial para teste da dependência espacial antes da aplicação do *GWLR* e 232, por ser a mesma adotada no *GWLR*. O motivo de também adotar a mesma largura de banda do modelo *GWLR* é representar a dependência espacial na mesma escala da variabilidade espacial. Todos os comandos foram feitos com o máximo de permutações permitida no GeoDa, para maior precisão dos resultados, sob nível de significância de 5%. Além dos resultados globais do I de Moran, a Tabela 11 apresenta a quantidade de observações cujo índice I de Moran local é estatisticamente significativo. O resultado do *LJC* na variável escolha modal é apresentado na Figura 9.

Tabela 11 – Resultados globais do I de Moran

Variável	Número de vizinhos mais próximos	I de Moran	Pseudo p-valor	Resultados significativos no teste I de Moran local (%) <sup>2</sup>
Probabilidade <sup>1</sup>	20	0.249	0	52.41
Viagens	20	0.161	0	22.12
Idade	20	0.442	0	52.91
Auto	20	0.135	0	33.78
Probabilidade <sup>1</sup>	232	0.050	0	76.36
Viagens	232	0.022	0	63.72
Idade	232	0.105	0	92.34
Auto	232	0.009	0.005	35.61

<sup>1</sup> Probabilidade da escolha do automóvel - resultado do modelo logit não espacial.

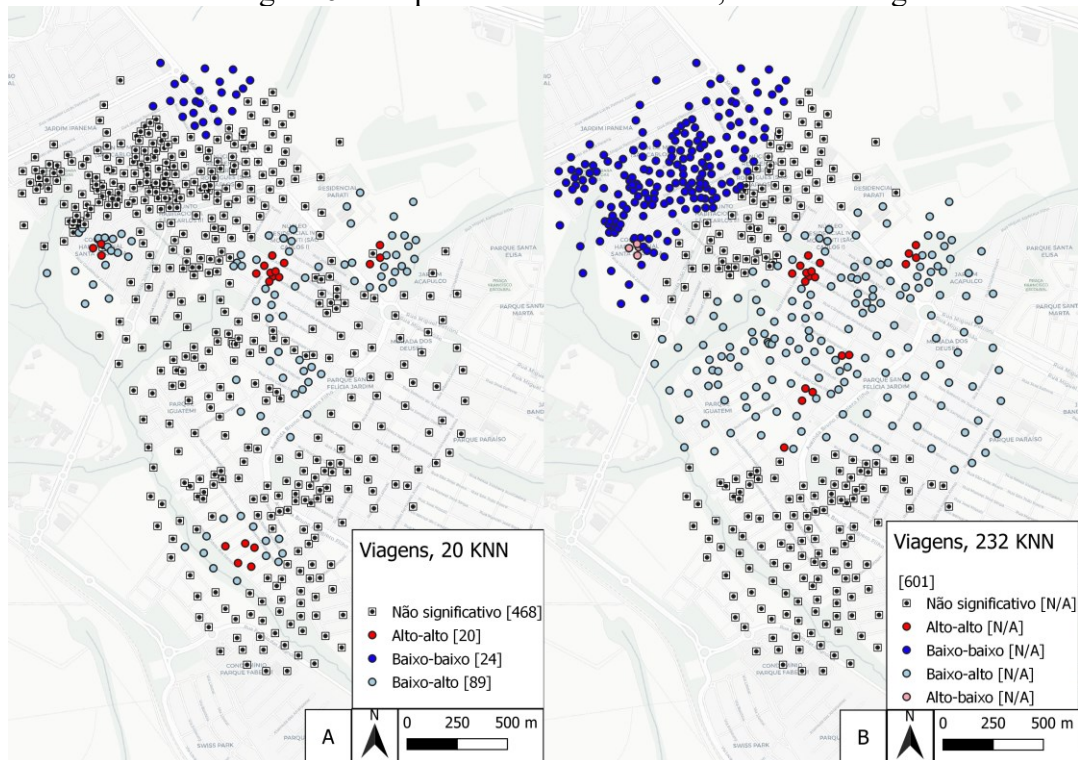
<sup>2</sup> Quantidade percentual de observações significativas no teste de hipótese do I de Moran Local.

Fonte: elaborada pelo autor.

Sob a hipótese nula de aleatoriedade espacial, todas as variáveis independentes do modelo têm a presença de dependência espacial corroborada. A presença da dependência espacial também sugere um ajuste limitado do modelo logit não espacial. Um resultado interessante é que uma maior largura de banda diminui a dependência espacial observada pelo I de Moran, porém aumenta a quantidade de observações em que o índice é estatisticamente significativo. Estas diferenças estão de acordo com o encontrado em Fotheringham *et al.* (2022), mesmo que fora do contexto de modelos de regressão, e reforça a importância da escolha da largura de banda para as AEDE.

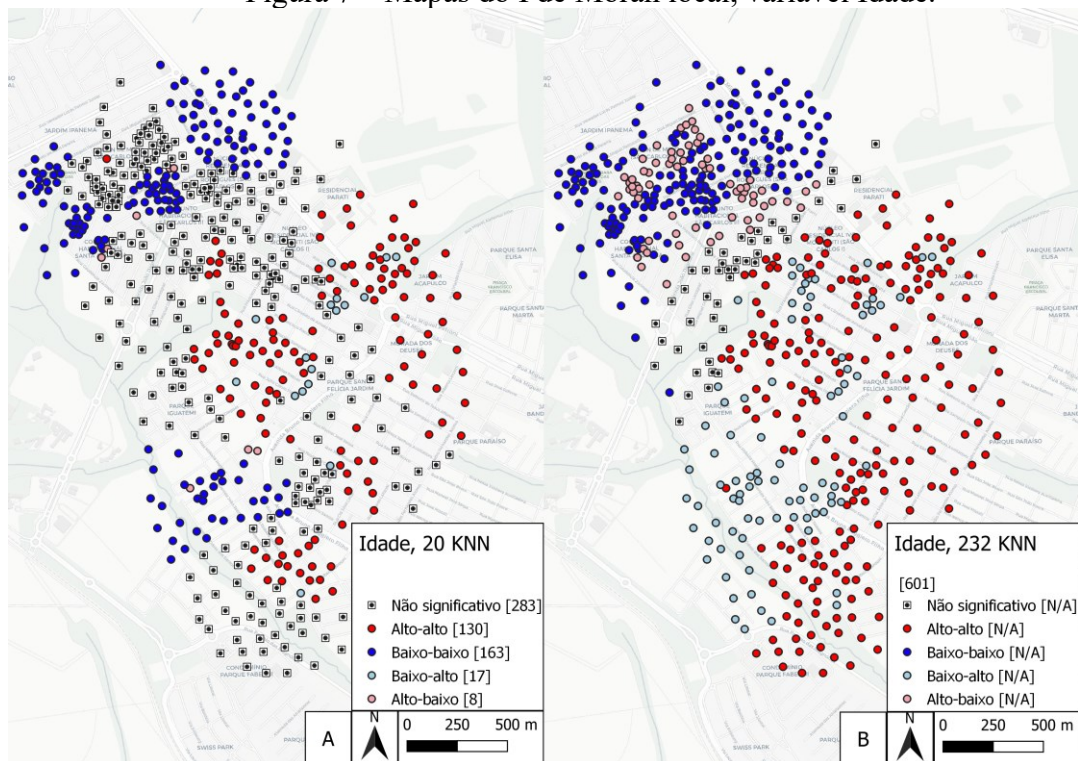


Figura 6 – Mapas do I de Moran local, variável Viagens



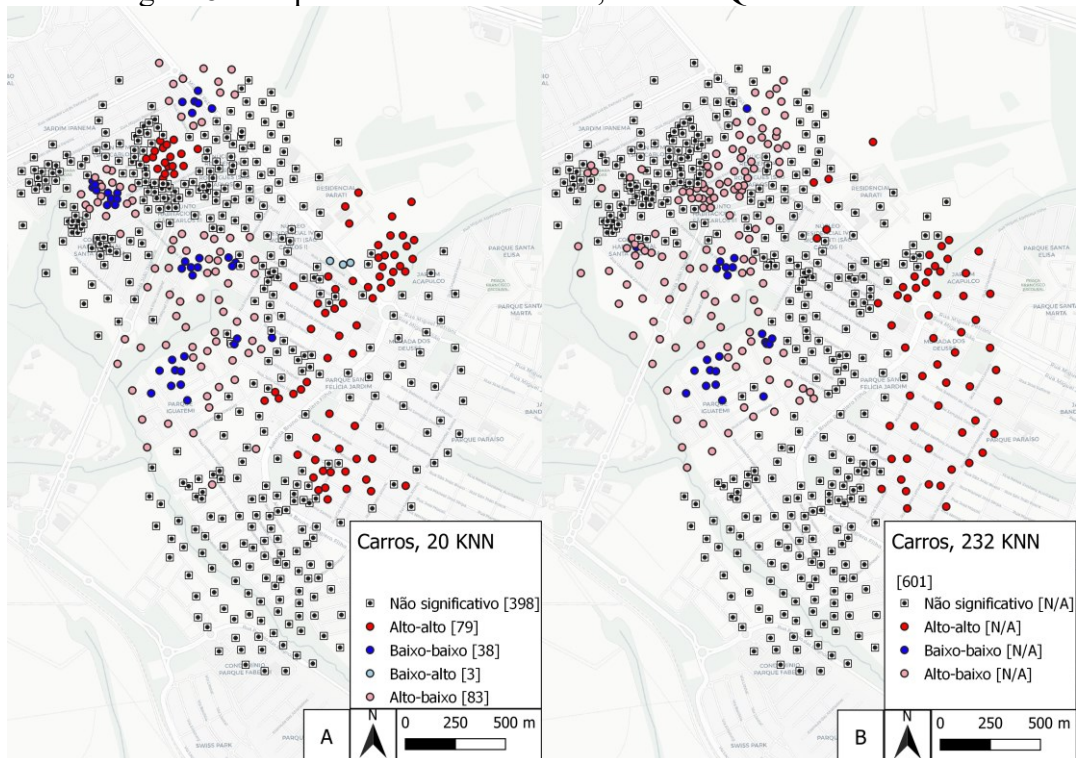
Fonte: elaborada pelo autor.

Figura 7 – Mapas do I de Moran local, variável Idade.



Fonte: elaborado pelo autor.

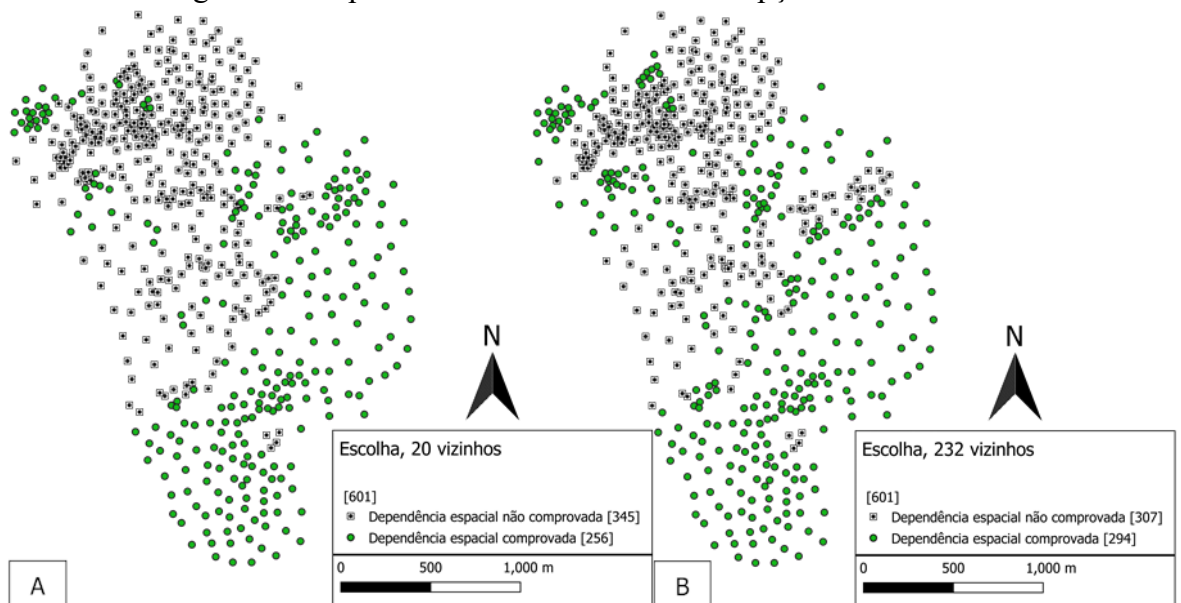
Figura 8 – Mapas do I de Moran local, variável Quantidade de automóveis



Fonte: elaborado pelo autor.

Para o *LJC*, o resultado é um mapa de pseudo p-valores, indicando as observações em que a hipótese de aleatoriedade espacial é rejeitada, conforme descrito em detalhes na Seção 3.3. A Figura 9 ilustra resultado do *LJC* na variável escolha para a opção automóvel.

Figura 9 – Mapa do *LJC* na variável escolha opção automóvel

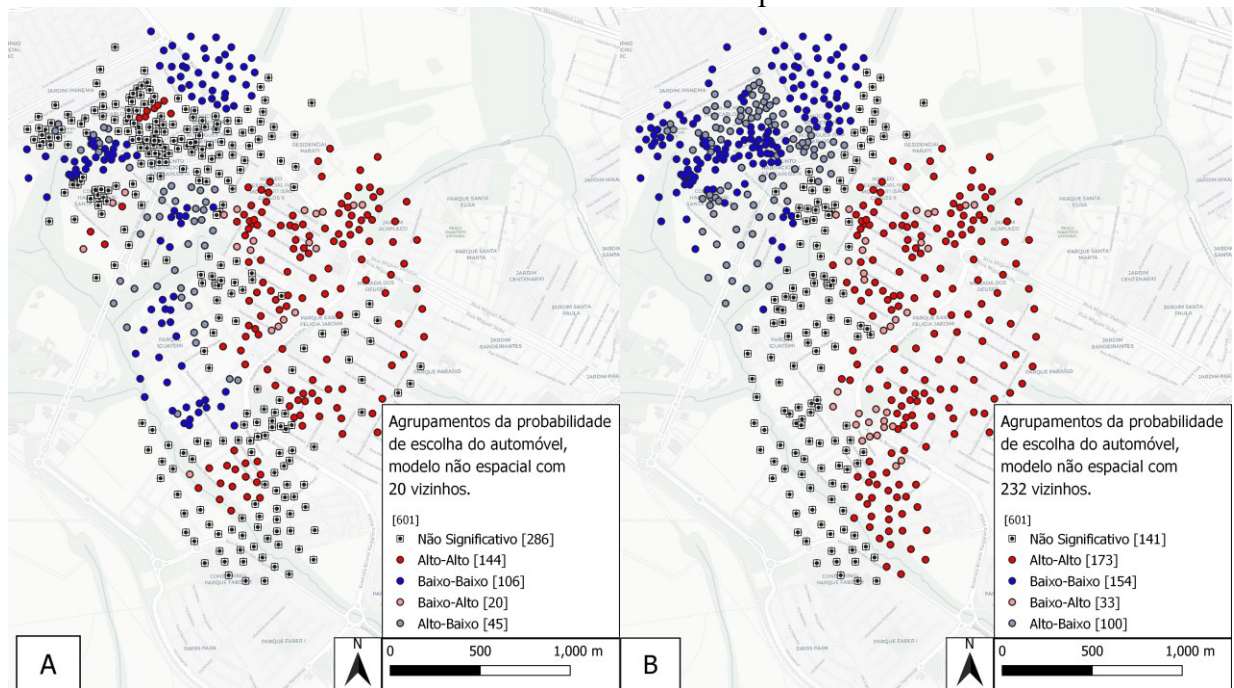


Fonte: elaborado pelo autor.

Para 20 vizinhos, 42.6% das observações tiveram a hipótese nula de aleatoriedade espacial refutada para o teste do *LJC*, e para 232 vizinhos 48.92%. Ainda, percebe-se que o padrão das observações com dependência espacial significativa pelo teste *LJC* não é aleatório, sendo mais um indício da presença de dependência espacial na variável escolha modal.

A Figura 10 é o resultado gráfico do *I* de Moran local para as probabilidades das escolhas do modo de transporte carro, obtidos por meio do modelo logit não espacial.

Figura 10 – Mapa do *I* de Moran local da probabilidade estimada de escolha do automóvel no modelo não espacial



Fonte: elaborado pelo autor.

### 4.3 Modelo *GWLR* e teste de variabilidade espacial Monte Carlo

Os resultados principais do *GWLR* são: os parâmetros estimados ( $\beta$ ) associados a cada variável em cada observação, t-valores e p-valores dos parâmetros estimados, desvio padrão ou erro padrão e métricas locais de multicolinearidade. No caso do *software* MGWR, também é fornecida a distância de *Cooks* para auxiliar a análise de influência das observações (Fotheringham *et al.*, 2002). Os valores obtidos não indicam a presença de *outliers*.

A Tabela 12 fornece os resultados gerais do modelo *GWLR*, e a Tabela 13 sumariza os parâmetros estimados das variáveis. Nota-se que apenas as observações com coeficientes estatisticamente significativos em cada variável foram utilizadas para cálculo dos valores, para um t-valor de 1.96.

Tabela 12 – Resultados gerais do modelo *GWLR*

Métrica	Valor
AICc	425.2
BIC	513.3
Log-Likelihood	-191.49
Pseudo R <sup>2</sup> ajustado	0.523
Largura de banda ( <i>KNN</i> ) <sup>1</sup>	232
Distância de <i>Cooks</i> máxima	0.021

<sup>1</sup>: Menor largura de banda que executou o modelo e o teste de variabilidade espacial.

Fonte: elaborada pelo autor.

Tabela 13 – Síntese dos parâmetros estimados do modelo *GWLR*

Variável	Média	Desvio padrão	Mínimo	Mediana	Máximo	Coefficiente de variação (CV) <sup>1</sup>	Coefficientes significativos (%)
Intercepto	-12.18	8.93	-27.74	-6.761	-2.808	0.733	90 (14.98)
Viagens	7.360	2.35	3.689	6.899	12.17	0.319	56 (9.32)
Idade	5.914	11.07	-12.19	7.274	28.56	1.872	277 (46.09)
Auto	9.554	3.12	2.734	9.729	18.65	0.306	560 (93.18)

<sup>1</sup>: Razão entre o desvio padrão e a média, em módulo.

Fonte: elaborada pelo autor.

A largura de banda adotada foi escolhida de forma iterativa, sendo observado o mesmo padrão da etapa de AEDE: maiores larguras de banda fornecem mais observações com coeficientes significativos, porém com menor impacto de efeitos espaciais. No caso do *GWLR*, também há uma piora na performance do modelo. Este resultado também está de acordo com Fotheringham *et al.* (2022), e destaca a escolha do modelador de acordo com o tipo de resultado necessário para o caso de estudo. Visto que o presente trabalho visa comparar os ganhos esperados entre o modelo não tradicional e o *GWLR*; e não uma predição da maior quantidade possível de observações, foi mantido o critério de menor largura de banda que execute a regressão e o teste de variabilidade espacial.

Os parâmetros estimados, na Tabela 10 e Tabela 13, se mantiveram com os mesmos sinais para as variáveis, e com pesos relativos que se mantêm proporcionais. Em ambos, o intercepto é bastante influente, o que era esperado pela pouca disponibilidade de variáveis em um contexto de escolha modal, que depende de diversos fatores. Em todas as observações, o intercepto foi negativo e com forte variação. No modelo tradicional, a posse de automóveis foi a variável mais influente, seguida de viagens e idade. Esta tendência se manteve no modelo local, com algumas diferenças em módulo.

A Tabela 14 sumariza os resultados de multicolinearidade das variáveis no modelo, para as observações associadas aos coeficientes estatisticamente significativos.

Tabela 14 – Resultados de multicolinearidade do modelo espacial local

Variável	VIF até 2 (% <sup>2</sup> )	VIF entre 5 e 10 (% <sup>2</sup> )	VIF acima de 10 (% <sup>2</sup> )
Intercepto	- <sup>1</sup>	- <sup>1</sup>	- <sup>1</sup>
Viagens	56 (100)	0 (0)	0 (0)
Idade	277 (100)	0 (0)	0 (0)
Auto	556 (99.28)	4 (0.72)	0 (0)

<sup>1</sup> O *VIF* não se aplica ao intercepto. <sup>2</sup> Porcentagem do número de observações com coeficientes significativos da variável.

Fonte: elaborada pelo autor.

Os resultados acima indicam pouca ou nenhuma multicolinearidade pelo *VIF*, que foi o critério adotado neste estudo para análise de multicolinearidade. Foi adotada como referência o valor de *VIF* abaixo de 5 como ótimo (Oshan *et al.*, 2019).

Para o teste de Monte Carlo no recurso computacional MGWR, são fornecidos os p-valores do teste, para nível de significância de 5%, de cada variável fornecida no modelo. Para p-valores abaixo de 0.05, rejeita-se a hipótese nula de estacionariedade espacial em favor da hipótese alternativa de variabilidade espacial. A Tabela 15 apresenta os resultados do teste para o modelo computado.

Tabela 15 – Resultado do teste Monte Carlo de variabilidade espacial

Variável	p-valor	Hipótese corroborada
Intercepto	0.001	Variabilidade espacial
Viagens	0.000	Variabilidade espacial
Idade	0.000	Variabilidade espacial
Auto	0.019	Variabilidade espacial

Fonte: elaborada pelo autor.

A hipótese de estacionariedade espacial foi rejeitada para todos os coeficientes das variáveis independentes, permitindo a análise da variabilidade dos estimadores. Percebe-se, pela Tabela 13, que o intercepto e a idade possuem uma grande variabilidade em seus coeficientes. Dessa forma, além de atestar a variabilidade espacial das variáveis, a amplitude de valores é importante e não deve ser ignorada. Na Tabela 13, é apresentado o coeficiente de variação (CV), definido como a razão entre o desvio padrão e a média. Nota-se que este coeficiente foi elaborado apenas para auxiliar a análise dos resultados, e não é, necessariamente, um indício que a diferença dos estimadores locais é estatisticamente significativa. Inclusive, o CV só pode ser analisado após a corroboração da variabilidade espacial.

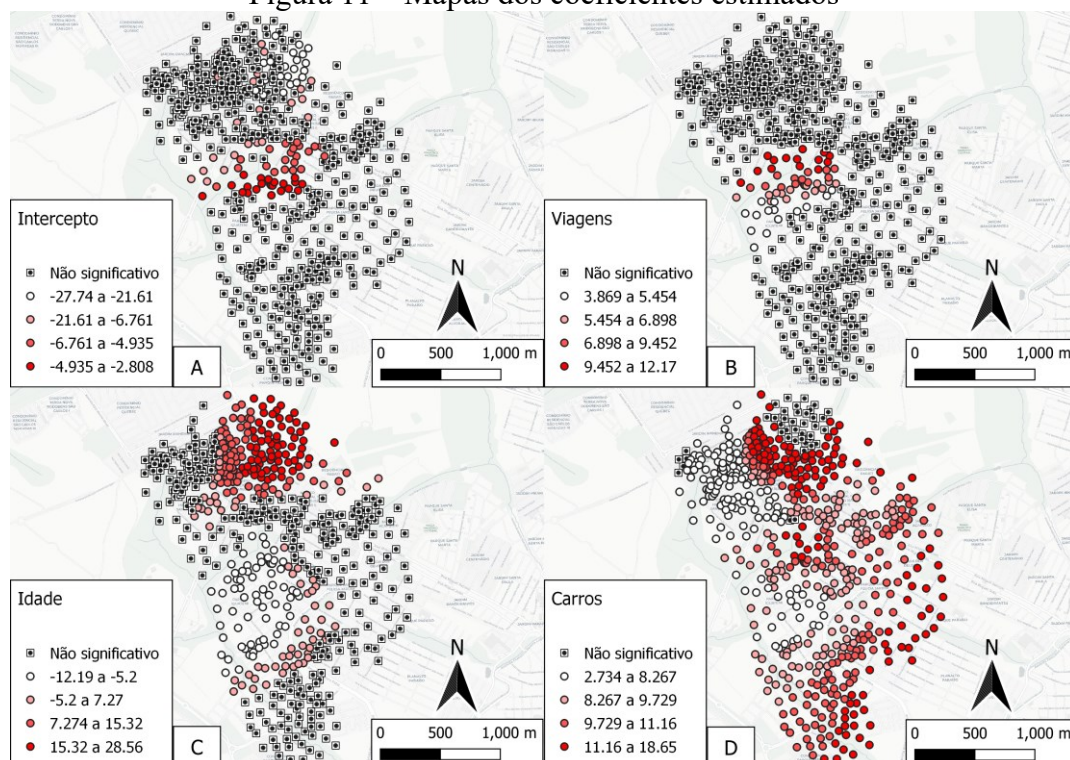
O coeficiente da idade possui a média menos influente, porém é a com maior CV. Além disso, pelos mínimos e máximos, o modelo sugere que, para alguns indivíduos, a idade pode ser mais explicativa do que a posse de automóveis ou a quantidade de viagens na escolha modal. Em outras palavras, a média pode ser uma métrica insuficiente em modelos *GWR*. A idade também é a única variável selecionada que os coeficientes podem assumir valores negativos. Estas particularidades da variável idade podem ocorrer pois diferentes faixas etárias apresentam

comportamentos de viagens diferentes (Golob e Hensher, 1998; Páez *et al.*, 2007; Zheng *et al.*, 2012; Tyrinopoulos e Antoniou, 2013; Basarić *et al.*, 2016; Liu *et al.*, 2020). A fase idosa, por exemplo, pode implicar que a idade tenha uma influência esperada não linear no estimador (Zheng *et al.*, 2012). A idade também pode representar outras variáveis sociais importantes que afetam a escolha modal (Tyrinopoulos e Antoniou, 2013; Liu *et al.*, 2020), mas que não estão presentes neste estudo. Apesar do maior potencial explicativo dos coeficientes locais, é justamente pelas limitações do banco de dados utilizado que esta análise por faixas etárias não foi aprofundada.

Os CV dos coeficientes estimados de automóveis e viagens são relativamente baixos quando comparado ao da idade e intercepto. Um motivo razoável dessas diferenças é, ao contrário da idade, a quantidade de viagens e de automóveis é diretamente proporcional com a escolha do modo automóvel, ainda que não necessariamente linear. A variação encontrada no intercepto, por sua vez, pode ocorrer pela omissão de variáveis pertinentes no modelo, sobretudo aquelas que caracterizariam as alternativas (Anselin, 1988; LeSage, 1999).

Os mapas dos parâmetros estimados são apresentados na Figura 11, com os coeficientes estatisticamente não significativos e a classificação dos significativos em quebras naturais de *Jenks*. Os intervalos de quebra foram calculados sem considerar os coeficientes não significativos.

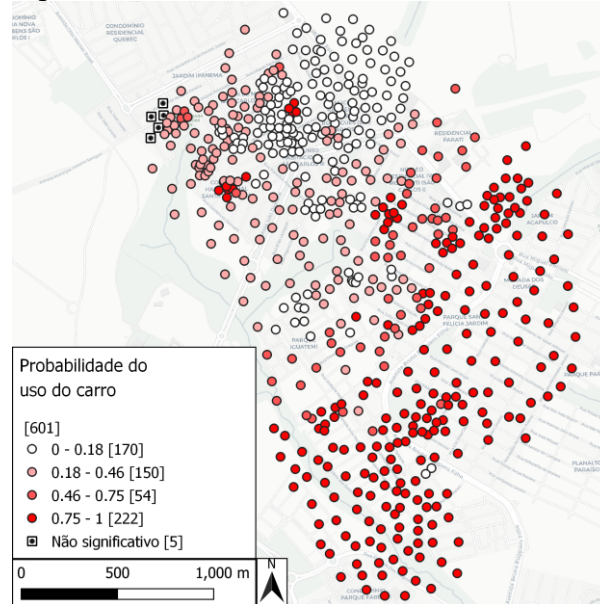
Figura 11 – Mapas dos coeficientes estimados



Fonte: elaborado pelo autor.

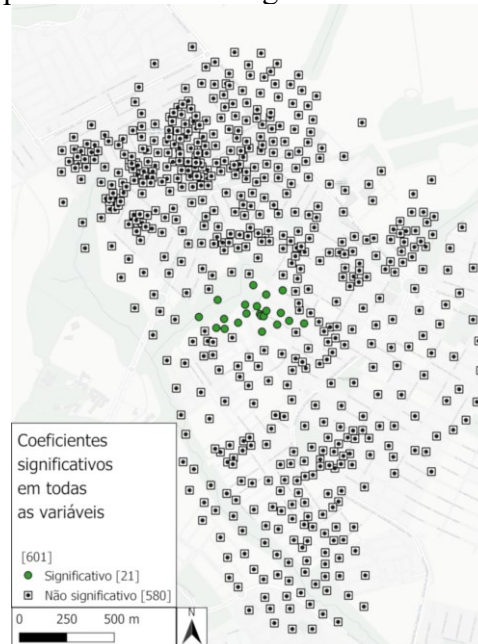
Na Figura 12, é apresentado o mapa da probabilidade estimada de uso do automóvel, sendo que as observações em que nenhuma das variáveis teve ao menos um coeficiente significativo foram desconsideradas. Outro resultado que pode ser obtido é a quantidade e localização das observações em que todas as variáveis têm coeficientes significativos simultaneamente, na Figura 13.

Figura 12 – Mapa das probabilidades de escolha modal automóvel pelo modelo *GWLR*.



Fonte: elaborado pelo autor.

Figura 13 – Mapa de coeficientes significativos em todas as variáveis



Fonte: elaborado pelo autor.

#### 4.4 Comparações entre modelos logit não espacial e espacial local

Na comparação entre o modelo logit não espacial e o modelo *GWLR*, o último teve melhor ajuste em todas as métricas consideradas (*AICc*, *log-Likelihood*, pseudo  $R^2$  ajustado e taxa de acerto global). A Tabela 16 descreve as métricas de ajuste comparadas, enquanto a Tabela 17 mostra a matriz de confusão de cada modelo.

Tabela 16 – Métricas de ajuste dos modelos logit não espacial e *GWLR*

Métrica	Modelo logit não espacial	Modelo <i>GWLR</i>
<i>AICc</i>	743.262	425.206
Log-Likelihood	-367.598	-191.491
Pseudo $R^2$ ajustado	0.112	0.523
Taxa de acerto global	64.22 %	85.19 %

Fonte: elaborado pelo autor.

Tabela 17 – Matrizes de confusão dos modelos

Valor Observado	Modelo logit não espacial		Modelo <i>GWLR</i>	
	0	1	0	1
0	21.96 %	25.62 %	44.09 %	3.49 %
1	10.15 %	42.26 %	11.31 %	41.1 %

Fonte: elaborado pelo autor.

Pela Tabela 16, o modelo *GWLR* teve melhor ajuste em todas as métricas consideradas. O *VIF* das variáveis (Tabela 10 e Tabela 14) não apontam aumento significativo da multicolinearidade dos parâmetros estimados do *GWLR*, que é uma consideração importante ao considerar este tipo modelo (Fotheringham e Oshan, 2016). Ainda, pelas matrizes de confusão, o *GWLR* tem uma melhoria considerável no acerto no modo transporte público com um aumento aproximado de 22%. O modo automóvel privado teve uma leve piora no acerto, de 42.26% no logit não espacial para 41.1% no *GWLR* (-1.16%).

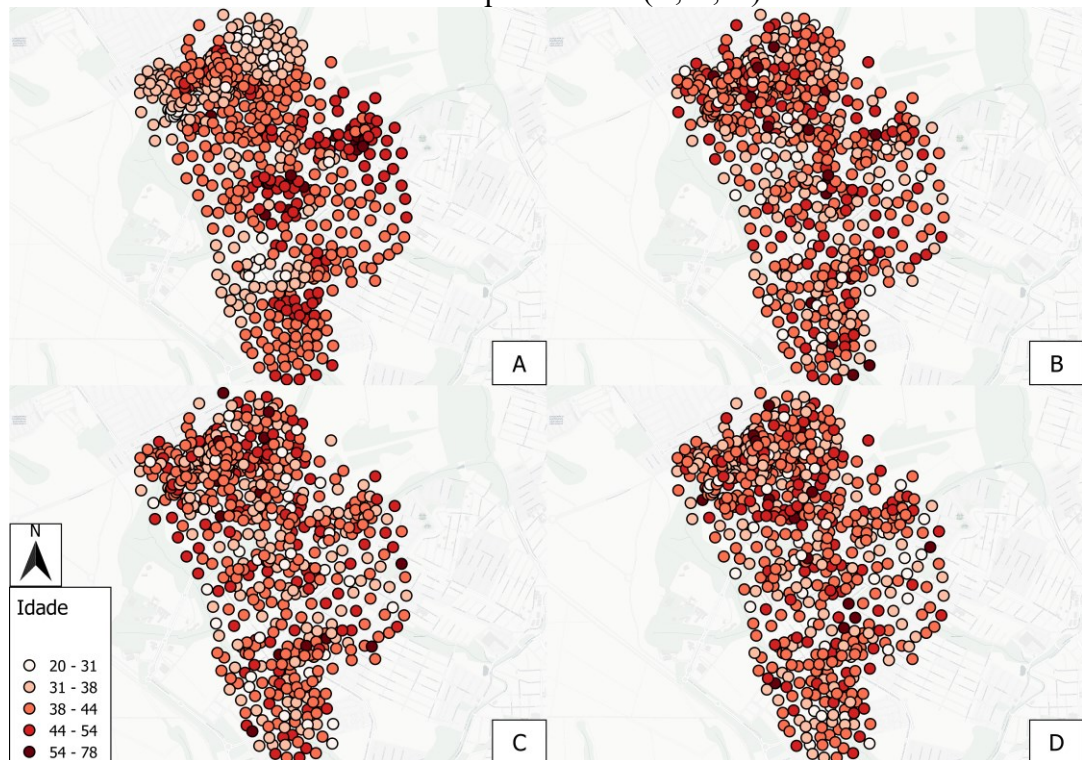
Assim, a única desvantagem que pode ser atribuída aos resultados do modelo *GWLR* deste estudo é o baixo percentual de coeficientes estimados significativos, salvo na variável posse de automóveis (vide Tabela 13). Isto pode ser amenizado aumentando a largura de banda da calibração do modelo, ao custo de pior ajuste e piores coeficientes estimados e perda da captura dos efeitos espaciais (Fotheringham *et al.*, 2022). Dado que um dos objetivos deste estudo é a análise dos efeitos espaciais na escolha modal, manteve-se a largura de banda de 232 vizinhos, ciente da menor quantidade de coeficientes locais estimados significativos.



#### 4.5 Teste de variabilidade espacial local proposto neste trabalho

Conforme descrito na Seção 3.6, as coordenadas geográficas foram aleatoriamente permutadas 200 vezes, obtendo-se 200 bases de dados sorteadas onde a aleatoriedade espacial é esperada. A Figura 14 apresenta o exemplo de valores da variável idade do banco de dados sem permutação (A) e três bancos de dados com coordenadas geográficas permutadas (B, C, D). A Tabela 18 sumariza os valores críticos do teste e os do modelo local de referência.

Figura 14 – Exemplo de distribuição espacial da variável idade para dados originais (A) e uma das bases de dados permutados (B, C, D)



Fonte: elaborado pelo autor.

Tabela 18 – Resultados do teste proposto

Variável	Valor do modelo sem permutação <sup>1</sup>	Valor de corte do teste <sup>1</sup>	Resultado do teste proposto	Resultado do teste de referência
Intercepto	-0.126	-0.809	Variabilidade espacial	Variabilidade espacial
Viagens <sup>2</sup>	-0.130	-1.064	Variabilidade espacial	Variabilidade espacial
Idade <sup>3</sup>	-0.142	-3.088	Variabilidade espacial	Variabilidade espacial
Auto <sup>4</sup>	-0.310	-2.359	Variabilidade espacial	Variabilidade espacial

<sup>1</sup>: Percentuais de coeficientes locais fora do intervalo de confiança dos coeficientes do modelo não espacial; transformados por Box-Cox (Box e Cox, 1964). <sup>2</sup> Quantidade de viagens diárias realizadas pelo domicílio. <sup>3</sup> Idade do entrevistado do domicílio. <sup>4</sup> Quantidade de automóveis por domicílio

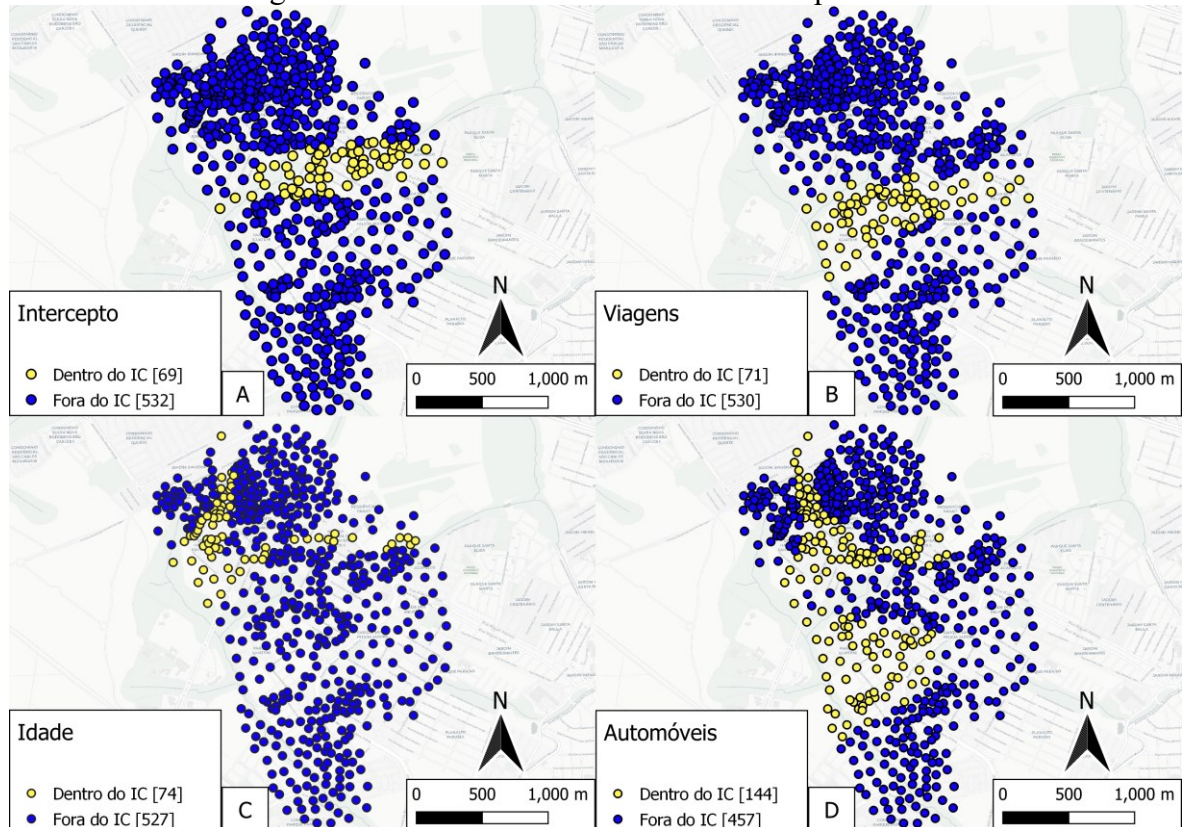
Fonte: elaborado pelo autor.

Em todas as variáveis, a hipótese de variabilidade espacial é validada pelo teste proposto, em consonância com o teste de referência utilizado (vide p-valores na Tabela 15).

Ainda, as Figuras Figura 15 e

Figura 16 mostram os resultados desagregados do teste. São, respectivamente, a comparação dos coeficientes estimados do modelo *GWLR* sem permutação e o teste LJC sobre estas categorias.

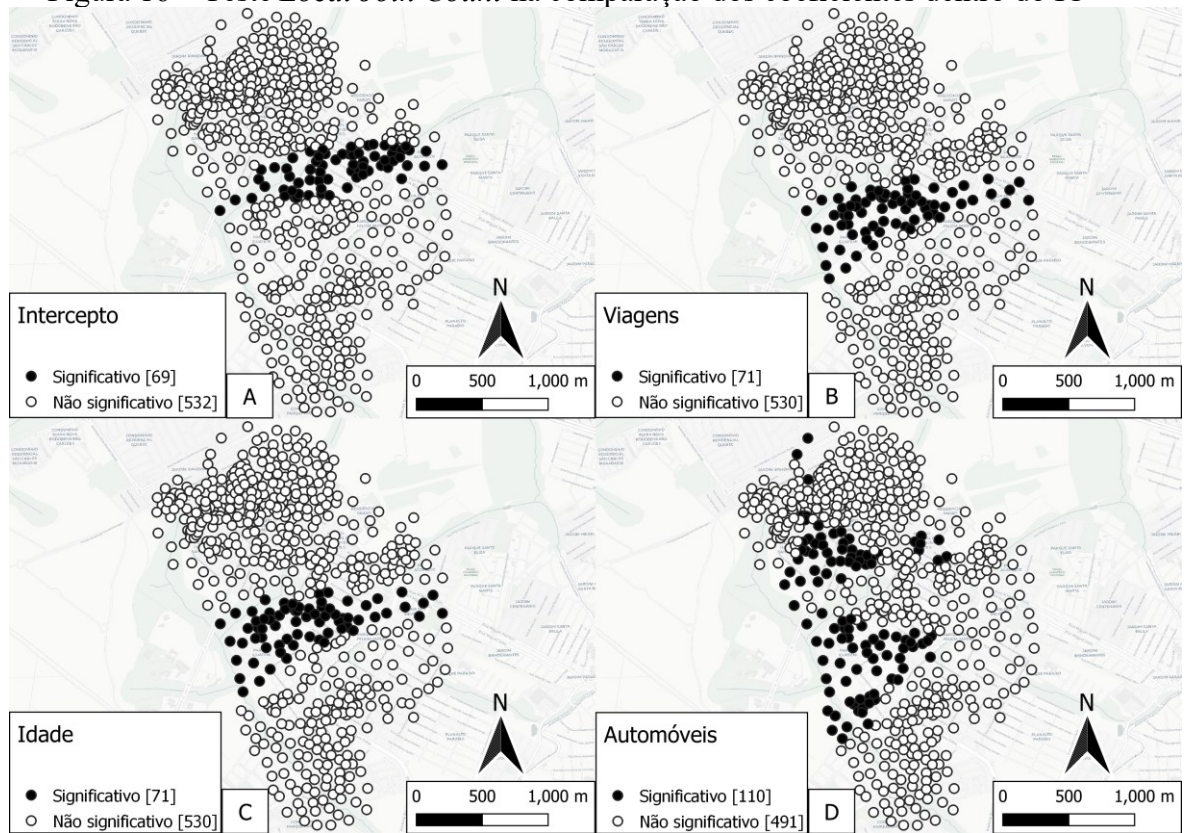
Figura 15 – Análise local da variabilidade espacial



Fonte: elaborado pelo autor.

Conforme sugerido por Anselin e Li (2019), a categoria de valor 1 é a de menor proporção global, ou seja, o teste avaliou a dependência espacial das observações dentro do intervalo de confiança, em todas as variáveis. Assim, as regiões formadas pelos pontos pretos são de estacionariedade espacial, enquanto as regiões com pontos brancos têm indício de presença da variabilidade espacial. Ambas as figuras corroboram a não-aleatoriedade espacial da diferença estatística entre os coeficientes do modelo logit não espacial e do *GWLR*, e sugerem regiões em que a variabilidade espacial pode não estar presente.

Figura 16 – Teste *Local Join Count* na comparação dos coeficientes dentro do IC



Fonte: elaborado pelo autor.

## CAPÍTULO 5: CONCLUSÕES E SUGESTÕES PARA TRABALHOS FUTUROS

### 5.1 Síntese do trabalho

O principal objetivo deste trabalho foi analisar os efeitos espaciais para um caso de estudo de escolha modal, com foco na análise de variabilidade espacial. De forma específica, foi proposto um teste de variabilidade espacial, utilizando resultados de ambos os modelos locais e não espacial, por meio de simulações e comparações com intervalos de confiança.

Foi proposta uma sequência metodológica a partir de calibração de um modelo logit não espacial, seguindo por análises exploratórias espaciais e calibração posterior de um modelo *GWLR*. A ordem destas etapas foi elaborada para atender os pressupostos necessários do modelo *GWR*, destacar o melhor ajuste dos modelos e ganhos qualitativos de resultados locais, além de guiar os pesquisadores a entender melhor o estudo de caso, dada a carência de trabalhos com enfoque nos modelos *GWR* na literatura de escolha modal.

Os coeficientes estimados do modelo logit não espacial estavam dentro da expectativa da literatura, com a predominância da posse de automóveis. Os resultados das AEDE mostraram que todas as variáveis consideradas possuem dependência espacial e agrupamentos espaciais distintos, dando prosseguimento ao procedimento proposto. O modelo *GWLR* foi melhor em todas as métricas de ajuste e nas taxas de acertos, comparado ao modelo logit não espacial. Dentre as variáveis consideradas (idade, quantidade de viagens e quantidade de automóveis no domicílio), a quantidade de automóveis foi a variável com maior predominância na escolha modal no modelo logit não espacial e na maioria das observações do modelo *GWLR*. A quantidade de viagens tem o segundo maior coeficiente em ambos os modelos, mas os coeficientes estimados locais da idade têm os maiores valores mínimos e máximos. Os resultados da variável idade estão parcialmente em acordo com a literatura consultada, mas isto pode ter ocorrido pois a faixa etária idosa não foi tratada como uma sub-amostra e estudada a fundo, por estar além do escopo deste trabalho.

O teste de variabilidade espacial proposto compara os intervalos de confiança dos coeficientes de um modelo estacionário, neste estudo o modelo logit não espacial, com os coeficientes de um modelo espacial local (inclusive simulações), neste estudo o *GWLR*. Estas comparações criam valores binários que são sumarizados em percentagens de coeficientes locais dentro e fora do intervalo, que por sua vez são transformados por Box-Cox (Box e Cox, 1964). Com os percentuais transformados das simulações, são feitas distribuições e obtidos

valores críticos, para teste de hipótese com o percentual transformado do modelo sem permutações, obtendo um resultado para cada parâmetro estimado.

## 5.2 Contribuições e conclusões

Os resultados apresentados corroboram a viabilidade da aplicação de um modelo *GWLR* para o estudo da escolha modal, e fornecem indícios que os resultados locais disponíveis ampliam a interpretação dos resultados além do que seria possível com coeficientes estacionários. Com estes resultados, a hipótese central do estudo: “A incorporação de dependência e variabilidade espacial traz melhorias às estimativas relacionadas à escolha modal” é corroborada. Além da melhoria observada nas métricas globais de ajuste, como *AICc*, pseudo  $R^2$ , *log-likelihood* e matriz de confusão, há a melhoria explicativa decorrente da disponibilidade de coeficientes estimados locais.

Visto que os resultados do teste de variabilidade espacial proposto foram similares ao teste de referência adotado em todas as variáveis, considera-se que o teste foi bem-sucedido para o caso de estudo adotado. Destaca-se que o teste pode ser aplicado também com um modelo espacial global ao invés de um não espacial para a elaboração do intervalo de confiança das variáveis. Os modelos espaciais globais podem incorporar a dependência espacial, mas seus parâmetros estimados são estacionários. Assim, há maior garantia que a diferença estatística entre os estimadores locais e o intervalo de confiança adotado seja devido apenas à variabilidade espacial. Neste caso sugere-se o teste de hipótese bilateral, dado que o parâmetro estimado pelo modelo espacial global pode ser próximo da média dos parâmetros espaciais locais estimados; enquanto os coeficientes dos modelos dos bancos de dados permutados se afastam do intervalo de confiança, por serem espacialmente aleatórios.

Conforme Getis e Ord (1992) e Fotheringham *et al.* (2002), métricas globais, mesmo que intrinsecamente espaciais, podem esconder tendências locais importantes. Portanto, uma etapa adicional proposta foi utilizar resultados locais intermediários do teste para corroborar o teste proposto e avaliar o ganho de potencial exploratório. Outro aspecto interessante é avaliar se a diferença estatística, caso corroborada no teste proposto, também é acompanhada por uma não-aleatoriedade espacial. Foram utilizados mapas de comparação direta dos parâmetros locais com o intervalo de confiança do modelo não espacial e um teste *Local Join Count* sobre os valores utilizados. O teste *LJC* se mostrou viável para análise local da variabilidade espacial, adicionando um novo ponto de vista para exploração e compreensão dos fenômenos locais. Por esses mapas foi possível avaliar a segregação regional entre grupos e suas coesões internas, esta não-aleatoriedade espacial está de acordo com os resultados anteriores.

### 5.3 Limitações e sugestões para trabalhos futuros

Uma das principais restrições deste trabalho é o banco de dados utilizado. Trata-se de uma base de dados adaptada de uma base real, tendo parte de suas observações interpoladas, incorporando uma dependência espacial artificial. Logo, os resultados apresentados não devem ser utilizados fora de seu contexto ou para representar a cidade de São Carlos/SP. Também houve falta de disponibilidade das variáveis das alternativas, uma vez que a pesquisa Origem/Destino é uma pesquisa de Preferência Revelada, havendo a caracterização apenas da alternativa realmente escolhida. A não consideração da distância em rede como opção para definição de vizinhos e calibração do modelo *GWLR* também é considerada uma limitação deste estudo. Ainda que o modelo *GWR* foi considerado a ferramenta computacional final neste caso de estudo, existem modelos mais recentes que podem trazer ainda mais riqueza aos resultados, como o *Multiscale GWR* (Fotheringham *et al.*, 2017; Oshan *et al.*, 2019), que possibilita o tratamento das variáveis em escalas diferentes, o que, além da esperada melhoria dos índices de ajuste, também tem potencial explicativo pela importância da escala em fenômenos com efeitos espaciais (Tobler, 1970; Anselin, 1988; Getis e Ord, 1992).

Também devem ser ressaltadas limitações importantes do teste proposto: implementação em mais de uma ferramenta computacional, dificultando sua reprodução; a falta de um modelo espacial global para cálculo dos intervalos de confiança; a ausência de variáveis espacialmente estacionárias no estudo de caso (pelo teste de referência), sendo desconhecida a eficácia do teste proposto na rejeição da variabilidade espacial.

Por isso, sugerem-se trabalhos futuros para resolver as limitações supracitadas do teste proposto, bem como as do caso de estudo do objetivo principal. Destacam-se as seguintes sugestões:

- Utilização de um banco de dados real;
- Uso de mais variáveis esperadas pertinentes pela literatura de escolha modal, bem como devida caracterização das alternativas;
- Uso de variáveis agregadas relacionadas às zonas de tráfego das residências, bem como variáveis de uso do solo e inerentes ao sistema de transportes;
- Avaliação do potencial de uso dos coeficientes locais estimados e/ou análises exploratórias espaciais para formação de diferentes perfis comportamentais de indivíduos;
- Aplicação do teste de variabilidade espacial proposto em outras bases de dados relativas à escolha modal, com efeitos espaciais esperados;

- Para o teste proposto, as demais sugestões importantes para trabalhos futuros coincidem com as três limitações citadas no parágrafo anterior.

## REFERÊNCIAS

- ALBUQUERQUE, P. H. M.; MEDINA, F. A. S.; SILVA, A. R. Regressão logística geograficamente ponderada aplicada a modelos de credit scoring. **Revista Contabilidade & Finanças**, v. 28, p. 93-112, 2017.
- ALJOUFIE, M.; BRUSSEL, M.; ZUIDGEEST, M.; VAN MAARSEVEEN, M. Urban growth and transport infrastructure interaction in Jeddah between 1980 and 2007. **International Journal of Applied Earth Observation and Geoinformation**, v. 21, p. 493-505, 2013.
- ALMEIDA, E. **Econometria espacial aplicada**. Campinas–SP. Alínea, 2012.
- ANSELIN, L. **Spatial econometrics: methods and models**. Springer Science & Business Media, 1988.
- ANSELIN, L. Local indicators of spatial association—LISA. **Geographical analysis**, v. 27, n. 2, p. 93-115, 1995.
- ANSELIN, L. **Cluster Analysis (3): Spatially Constrained Clustering Methods**. 2017. Disponível em: [https://geodacenter.github.io/workbook/8\\_spatial\\_clusters/lab8.html](https://geodacenter.github.io/workbook/8_spatial_clusters/lab8.html). Acesso em 10 dez. 2021.
- ANSELIN, L. A local indicator of multivariate spatial association: extending Geary's C. **Geographical Analysis**, v. 51, n. 2, p. 133-150, 2019.
- ANSELIN, L.; GALLO, J. L.; JAYET, H. Spatial panel econometrics. **The econometrics of panel data: Fundamentals and recent developments in theory and practice**, p. 625-660, 2008.
- ANSELIN, L.; GRIFFITH, D. A. Do spatial effects really matter in regression analysis? **Papers in Regional Science**, v. 65, n. 1, p. 11-34, 1988.
- ANSELIN, L.; LI, X. Operational local join count statistics for cluster detection. **Journal of Geographical Systems**, v. 21, n. 2, p. 189-210, 2019.
- ANSELIN, L.; LI, X. Tobler's law in a multivariate world. **Geographical Analysis**, v. 52, n. 4, p. 494-510, 2020.



ANSELIN, L.; REY, S. J. Introduction to the special issue on spatial econometrics. **International Regional Science Review**, v. 20, n. 1-2, p. 1-7, 1997.

ANSELIN, L.; SYABRI, I.; KHO, Y. GeoDa: an introduction to spatial data analysis. **Handbook of applied spatial analysis**. Springer, Berlin, Heidelberg, 2010. p. 73-89.

ASSIRATI, L.; PITOMBO, C. S. Modeling spatial effect on travel mode choice using a synthetic spatially correlated data set. **Boletim de Ciências Geodésicas**, v. 27, 2021.

BASARIĆ, V.; VUJIČIĆ, A.; SIMIĆ, J. M.; BOGDANOVIĆ, V.; SAULIĆ, N. Gender and age differences in the travel behavior—a Novi Sad case study. **Transportation research procedia**, 14, 4324-4333, 2016.

BEN-AKIVA, E. M. **Structure of passenger travel demand models**. 1973. Tese de Doutorado. Massachusetts Institute of Technology.

BEN-AKIVA, E. M.; LERMAN, S. R. **Discrete choice analysis: theory and application to travel demand**. MIT press, 1985.

BHAT, C.; ZHAO, H. The spatial analysis of activity stop generation. **Transportation Research Part B: Methodological**, v. 36, n. 6, p. 557-575, 2002.

BOLLEN, K. A. **Structural equations with latent variables**. John Wiley & Sons, 1989.

BOOTS, B. Developing local measures of spatial association for categorical data. **Journal of Geographical Systems**, v. 5, n. 2, p. 139-160, 2003.

BOX, G. E. P.; COX, D. R. An analysis of transformations. **Journal of the Royal Statistical Society Series B: Statistical Methodology**, v. 26, n. 2, p. 211-243, 1964.

BRUNSDON, C.; FOTHERINGHAM, A. S.; CHARLTON, M. E. Geographically weighted regression: a method for exploring spatial nonstationarity. **Geographical analysis**, v. 28, n. 4, p. 281-298, 1996.

COOK, D.; CRESSIE, N.; MAJURE, J.; SYMANZIK, J. Some dynamic graphics for spatial data (with multiple attributes) in a GIS. In: **Compstat: Proceedings in Computational Statistics 11th Symposium held in Vienna**, Austria, 1994. Physica-Verlag HD, 1994. p. 105-119.

CONOVER, W. J. **Practical nonparametric statistics**. John Wiley & Sons, p. 309-314, 1999.

DYKES, J. Cartographic visualization. **Journal of the Royal Statistical Society: Series D (The Statistician)**, v. 47, n. 3, p. 485-497, 1998.

ERDOGAN, S. Explorative spatial analysis of traffic accident statistics and road mortality among the provinces of Turkey. **Journal of safety research**, v. 40, n. 5, p. 341-351, 2009.

ESRI. ArcGIS Desktop 10.8. **ESRI INC**, 2019.

FOTHERINGHAM, A. S. Trends in quantitative methods I: stressing the local. **Progress in human geography**, 21(1), 88-96, 1997. <https://doi.org/10.1191/030913297676693207>.

FOTHERINGHAM, A. S. Context-dependent spatial analysis: A role for GIS?. **Journal of Geographical Systems**, v. 2, n. 1, p. 71-76, 2000.

FOTHERINGHAM, A. S.; BRUNSDON, C. Local forms of spatial analysis. **Geographical analysis**, v. 31, n. 4, p. 340-358, 1999.

FOTHERINGHAM, A. S.; BRUNSDON, C.; CHARLTON, M. **Geographically weighted regression: the analysis of spatially varying relationships**. John Wiley & Sons, 2002.

FOTHERINGHAM S. A.; CHARLTON, M.; BRUNSDON, C. The geography of parameter space: an investigation of spatial non-stationarity. **International Journal of Geographical Information Systems**, v. 10, n. 5, p. 605-627, 1996.

FOTHERINGHAM, A. S.; OSHAN, T. M. Geographically weighted regression and multicollinearity: dispelling the myth. **Journal of Geographical Systems**, v. 18, n. 4, p. 303-329, 2016.

FOTHERINGHAM, A. S., YANG, W., KANG, W. Multiscale geographically weighted regression (MGWR). **Annals of the American Association of Geographers**, 107(6), 1247-1265, 2017. <https://doi.org/10.1080/24694452.2017.1352480>.

FOTHERINGHAM, A. S.; YU, H.; WOLF, L. J.; OSHAN, T. M.; LI, Z. On the notion of 'bandwidth' in geographically weighted regression models of spatially varying processes. **International Journal of Geographical Information Science**, p. 1-18, 2022.

FOX, J.; WEISBERG, S. **An R companion to applied regression**. Sage publications, 2018.

GETIS, A.; ORD, J. K. The analysis of spatial association by use of distance statistics. **Geographical analysis**, v. 24, n. 3, p. 189-206, 1992.

GOETZKE, F. Network effects in public transit use: evidence from a spatially autoregressive mode choice model for New York. **Urban Studies**, v. 45, n. 2, p. 407-417, 2008.

GOLOB, T. F.; HENSHER, D. A. Greenhouse gas emissions and Australian commuters' attitudes and behavior concerning abatement policies and personal involvement. **Transportation Research Part D: Transport and Environment**, v. 3, n. 1, p. 1-18, 1998.

GOODCHILD, M. F.; JANELLE, D. G. Toward critical spatial thinking in the social sciences and humanities. **GeoJournal**, v. 75, p. 3-13, 2010.

GOMES, V. A.; PITOMBO, C. S.; ROCHA, S. S.; & SALGUEIRO, A. R. Kriging geostatistical methods for travel mode choice: a spatial data analysis to travel demand forecasting. **Open Journal of Statistics**, v. 6, n. 3, p. 514-527, 2016.

HOLLANDER, M.; WOLFE, D. A.; CHICKEN, E. **Nonparametric statistical methods**. John Wiley & Sons, 1999.

KOPCZEWSKA, K.; KUBARA, M. Spatial weights matrix, distance measurement, tessellation, spatial statistics. **In: Applied spatial statistics and econometrics**. Routledge, 2020. p. 151-211.

LESAGE, J. P. The theory and practice of spatial econometrics. **University of Toledo. Toledo, Ohio**, v. 28, n. 11, 1999.

LESAGE, J. P.; PACE, R. K. **Introduction to spatial econometrics**. Chapman and Hall/CRC, 2009.

LEUNG, Y.; MEI, C.; ZHANG, W. Statistical tests for spatial nonstationarity based on the geographically weighted regression model. **Environment and Planning A**, v. 32, n. 1, p. 9-32, 2000.

LI, Z.; OSHAN, T; FOTHERINGHAM, A. S.; WEI, K.; LEVI, W.; YU, H.; SACHDEVA, M.; BARDIN, S. **MGWR 2.2 User Manual**. Spatial Analysis Research Center (SPARC) Arizona State University, Tempe, USA, Tempe, p. 38, 2019.

LINDNER, A.; PITOMBO, C. S. A conjoint approach of spatial statistics and a traditional method for travel mode choice issues. **Journal of Geovisualization and Spatial Analysis**, 2, 1-13, 2018. <https://doi.org/10.1007/s41651-017-0008-0>.

LINDNER, A.; PITOMBO, C. S.; ASSIRATI, L.; PEDREIRA JUNIOR, J. U.; SALGUEIRO, A. R. Estimation of Travel Mode Choice Using Geostatistics: a Brazilian Case Study. **Revista Brasileira de Cartografia**, v. 73, n. 1, p. 182-197, 2021.

LIU, X.; GAO, L.; NI, A.; YE, N. Understanding better the influential factors of commuters' multi-day travel behavior: Evidence from Shanghai, China. **Sustainability**, 12(1), 376, 2020. <https://doi.org/10.3390/su12010376>.

MANSKI, C. F. The structure of random utility models. **Theory and decision**, v. 8, n. 3, p. 229, 1977.

MCFADDEN, D. The measurement of urban travel demand. **Journal of public economics**, v. 3, n. 4, p. 303-328, 1974.

MENDIOLA, L.; GONZÁLEZ, P.; CEBOLLADA, À. The link between urban development and the modal split in commuting: The case of Biscay. **Journal of Transport Geography**, v. 37, p. 1-9, 2014.

MIYAMOTO, K.; VICHIANAN, V.; SHIMOMURA, N.; PÁEZ, A. Discrete choice model with structuralized spatial effects for location analysis. **Transportation Research Record**, v. 1898, n. 1, p. 183-190, 2004.

MOHAMMADIAN, A.; KANAROGLOU, P. S. Applications of spatial multinomial logit model to transportation planning. In: **Proceedings of the 10th international conference on travel behaviour research**. 2003.

MONDAL, A.; BHAT, C. R. A Spatial Rank-Ordered Probit Model with an Application to Travel Mode Choice. **Transportation Research Part B: Methodological**, v. 155, p. 374-393, 2022.

NAKAYA, T. Geographically weighted generalised linear modelling. **Geocomputation: a practical primer**, v. 217, n. 20, p. 201-220, 2015.

NAKAYA, T.; CHARLTON, M.; LEWIS, P.; BRUNSDON, C.; YAO, J.; FOTHERINGHAM, S. **GWR4 user manual**. Windows Application for Geographically Weighted Regression Modelling, 2014.

NKEGBE, P. K.; KUUNIBE, N.; ABDUL-MUMIN, Y. Choice of transport mode by non-resident university students in Ghana. **Centre for Promoting Ideas**, v. 3, n. 20 2012.

NKEKI, F. N.; ASIKHIA, M. O. Geographically weighted logistic regression approach to explore the spatial variability in travel behaviour and built environment interactions: Accounting simultaneously for demographic and socioeconomic characteristics. **Applied geography**, 108, 47-63, 2019. <https://doi.org/10.1016/j.apgeog.2019.05.008>.

OSHAN, T. M.; LI, Z.; KANG, W.; WOLF, L. J.; FOTHERINGHAM, A. S. mgwr: A Python implementation of multiscale geographically weighted regression for investigating process spatial heterogeneity and scale. **ISPRS International Journal of Geo-Information**, v. 8, n. 6, p. 269, 2019.

PAELINCK, J. H. **Spatial econometrics** (Vol. 1). Saxon House. 1979.

PÁEZ, A. Exploring contextual variations in land use and transport analysis using a probit model with geographical weights. **Journal of Transport Geography**, v. 14, n. 3, p. 167-176, 2006.

PÁEZ, A.; SCOTT, D.; POTOGLOU, D.; KANAROGLOU, P.; NEWBOLD, K. B. Elderly mobility: demographic and spatial analysis of trip making in the Hamilton CMA, Canada. **Urban Studies**, v. 44, n. 1, p. 123-146, 2007.

PEARSON, K. X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. **The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science**, v. 50, n. 302, p. 157-175, 1900.

PITOMBO, C. S.; KAWAMOTO, E.; SOUSA, A. J. An exploratory analysis of relationships between socioeconomic, land use, activity participation variables and travel patterns. **Transport Policy**, 18(2), 347-357, 2011. <https://doi.org/10.1016/j.tranpol.2010.10.010>.

PITOMBO, C. S.; SALGUEIRO, A. R.; DA COSTA, A. S. G.; ISLER, C. A. A two-step method for mode choice estimation with socioeconomic and spatial information. **Spatial Statistics**, v. 11, p. 45-64, 2015.

QGIS.org. QGIS Geographic Information System. **QGIS Association**, 2022. <http://www.qgis.org>

RAJAMANI, J.; BHAT, C. R.; HANDY, S.; KNAAP, G.; SONG, Y. Assessing impact of urban form measures on nonwork trip mode choice after controlling for demographic and level-of-service effects. **Transportation research record**, v. 1831, n. 1, p. 158-165, 2003.

RAMOS, R. A. R.; DA SILVA, A. N. R. A spatial analysis approach for the definition of metropolitan regions—the case of Portugal. **Environment and Planning B: Planning and Design**, v. 34, n. 1, p. 171-185, 2007.

RAMSEY, J. B. Tests for specification errors in classical linear least-squares regression analysis. **Journal of the Royal Statistical Society: Series B (Methodological)**, v. 31, n. 2, p. 350-371, 1969.

RODRIGUES DA SILVA, A. N. Elaboração de um Banco de Dados de Viagem para Auxílio ao Desenvolvimento de Pesquisas na Área de Planejamento dos Transportes. **Universidade de São Paulo, Escola de Engenharia de São Carlos**, Relatório FAPESP, Processo, n. 04/15843, p. 4, 2008.

RYBARCZYK, G.; SHAKER, R. R. Predicting bicycle-on-board transit choice in a university environment. **Sustainability**, v. 13, n. 2, p. 512, 2021.

SÁNCHEZ-DÍAZ, I. HOLGUÍN-VERAS, J.; WANG, X. An exploratory analysis of spatial effects on freight trip attraction. **Transportation**, v. 43, n. 1, p. 177-196, 2016.

SHEPARD, D. A two-dimensional interpolation function for irregularly-spaced data. In: **Proceedings of the 1968 23rd ACM national conference**. 1968. p. 517-524.

TINTNER, G. The definition of econometrics. **Econometrica: Journal of the Econometric Society**, p. 31-40, 1953.

TOBLER, W. R. A computer movie simulating urban growth in the Detroit region. **Economic geography**, v. 46, n. sup1, p. 234-240, 1970.

TYRINOPOULOS, Y.; ANTONIOU, C. Factors affecting modal choice in urban mobility. **European Transport Research Review**, v. 5, n. 1, p. 27-39, 2013.

ZHANG, L.; HONG, J.; NASRI, A.; SHEN, Q. How built environment affects travel behavior: A comparative analysis of the connections between land use and vehicle miles traveled in US cities. **Journal of transport and land use**, 5(3), 40-52, 2012. <https://doi.org/10.5198/jtlu.v5i3.266>.