

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE CAMPINAS
ESCOLA DE ECONOMIA E NEGÓCIOS
FACULDADE DE CIÊNCIAS ECONÔMICAS

JOÃO PEDRO TOLEDO TRICOLI DE LUCAS

DETERMINANTES DO CRIME NO SUDESTE BRASILEIRO: APLICAÇÃO DE
TÉCNICAS DE ECONOMETRIA ESPACIAL

CAMPINAS

2023

JOÃO PEDRO TOLEDO TRICOLI DE LUCAS

DETERMINANTES DO CRIME NO SUDESTE BRASILEIRO: APLICAÇÃO DE
TÉCNICAS DE ECONOMETRIA ESPACIAL

Monografia apresentada à Escola de Economia e
Negócios da Pontifícia Universidade Católica de
Campinas como exigência parcial para a obtenção do
título de Bacharel em Ciências Econômicas.

Orientador: Prof. Me. Pedro de Miranda Costa

CAMPINAS

2023

JOÃO PEDRO TOLEDO TRICOLI DE LUCAS

DETERMINANTES DO CRIME NO SUDESTE BRASILEIRO: APLICAÇÃO
DE TÉCNICAS DE ECONOMETRIA ESPACIAL

Monografia apresentada à Escola de Economia e
Negócios da Pontifícia Universidade Católica de
Campinas como exigência parcial para a obtenção do
título de Bacharel em Ciências Econômicas.

Orientador: Prof. Me. Pedro de Miranda Costa

Prof. Me. Pedro de Miranda Costa
Orientador e Presidente da comissão examinadora
Pontifícia Universidade Católica de Campinas

Prof. Dr. Paulo Ricardo da Silva Oliveira
Examinador
Pontifícia Universidade Católica de Campinas

Prof. Me. Roberto Brito de Carvalho
Examinador
Pontifícia Universidade Católica de Campinas

CAMPINAS

2023

Sistema de Bibliotecas e Informação - SBI
Gerador de fichas catalográficas da Universidade
PUC-Campinas Dados fornecidos pelo(a) autor(a).

L933d	<p>Lucas, João Pedro Toledo Tricoli de</p> <p>Determinantes do Crime no Sudeste Brasileiro : Aplicações de Técnicas de Econometria Espacial / João Pedro Toledo Tricoli de Lucas. - Campinas: PUC-Campinas, 2023.</p> <p>72 il.</p> <p>Orientador: Pedro de Miranda Costa Costa. Coorientador: Paulo Ricardo da Silva Oliveira</p> <p>TCC (Bacharelado em Ciências Econômicas) - Escola de Economia e Negócios , Escola de Economia e Negócios, Pontifícia Universidade Católica de Campinas, Campinas, 2023. Inclui bibliografia.</p> <p>1. Economia do Crime. 2. Econometria Espacial. 3. Análise Exploratória de Dados Espaciais. I. Costa, Pedro de Miranda Costa. II. Oliveira, Paulo Ricardo da Silva III. Pontifícia Universidade Católica de Campinas. Escola de Economia e Negócios. Escola de Economia e Negócios . IV. Título.</p>
-------	---

AGRADECIMENTOS

Agradeço especialmente à minha mãe, Lilia de Jesus Toledo Tricoli de Lucas, e à minha madrinha, Dirlene Aparecida de Oliveira, pelo apoio incondicional e dedicação ao longo da minha vida e trajetória acadêmica.

Ao estimado professor Paulo Ricardo da Silva Oliveira, responsável por grande parte do meu desenvolvimento acadêmico e profissional, agradeço a dedicação, paciência, apoio e incentivo. Suas perspectivas compartilhadas serão levadas por toda a minha vida.

Ao professor e orientador Pedro de Miranda Costa, expresse minha gratidão pela paciência e apoio durante a elaboração da monografia, bem como pela parceria nos projetos do Observatório.

Expresse minha gratidão a todos os professores do Observatório PUC-Campinas, cujo suporte foi fundamental em todos os projetos de extensão. Contribuindo para criar um ambiente de aprendizado agradável e promovendo meu crescimento acadêmico.

Ao professor Brito de Carvalho, agradeço por me incluir em diversos projetos que representaram a universidade, além de oferecer valiosas perspectivas econômicas.

A todos, muito obrigado.

"I am saying that the economic approach provides a valuable unified framework for understanding all human behaviour".

Gary Becker

RESUMO

LUCAS, João Pedro Toledo Tricoli. DETERMINANTES DO CRIME NO SUDESTE BRASILEIRO: APLICAÇÃO DE TÉCNICAS DE ECONOMETRIA ESPACIAL. 2023. Trabalho de Conclusão de Curso. Faculdade de Ciências Econômicas, Escola de Economia e Negócios, Pontifícia Universidade Católica de Campinas, Campinas, 2023.

Diversas áreas do conhecimento investigam os determinantes da criminalidade, buscando compreender o "como" e "porquê" as pessoas cometem crimes. Este estudo, analisa os determinantes do crime na região sudeste do Brasil, utilizando técnicas de econometria espacial. Embasado na economia do crime e na Análise Exploratória de Dados Espaciais (AEDE), o modelo econométrico Spatial Lag (SAR) foi aplicado aos dados de taxas de homicídio por cem mil habitantes de 2010, suavizado pelo método bayesiano, em logaritmo, no nível municipal juntamente a grupos de variáveis socioeconômicas que abrangem desigualdade, predisposição do homem jovem ao crime, instabilidade familiar, educação, urbanização e risco. A inclusão do componente espacial não valida a significância encontrada no modelo linear simples para as seguintes variáveis, desigualdade e predisposição do homem jovem ao crime. Por outro lado, destaca-se o impacto positivo da educação, estabilidade familiar, urbanização, taxa de desocupação e risco de apreensão em regiões metropolitanas. Vale ressaltar a presença de clusters de alta criminalidade no Rio de Janeiro e Espírito Santo e cluster de baixa criminalidade em regiões de Minas Gerais e São Paulo. Em resumo, políticas voltadas para a redução de desigualdades, promoção da educação, fortalecimento da estabilidade familiar, investimento em segurança e criação de oportunidades de trabalho têm o potencial de gerar impactos positivos na diminuição da taxa de homicídios.

Palavras-chave: Economia do Crime; Econometria Espacial; Análise Exploratória de Dados Espacial; Sudeste Brasileiro

ABSTRACT

LUCAS, João Pedro Toledo Tricoli. DETERMINANTS OF CRIME IN BRAZILIAN SOUTHEAST: APPLICATION OF SPATIAL ECONOMETRIC TECHNIQUES. 2023. Undergraduate Thesis. School of Economics, School of Economics and Business, Pontifical Catholic University of Campinas, Campinas, 2023.

Several fields of knowledge delve into the determinants of criminality, aiming to understand "how" and "why" individuals commit crimes. This study analyzes the determinants of crime in the southeastern region of Brazil, by means of spatial econometric techniques based on the theoretical framework of crime economics and Exploratory Spatial Data Analysis (ESDA). The Spatial Lag (SAR) model was applied to homicide rate (per 100,000 people) in 2010, smoothed by empirical Bayesian methodology, in logarithms, at the municipal level and across socio-economic variable groups encompassing inequality, predisposition of young men to commit crimes, family instability, education, urbanization, and risk. Including the spatial component does not validate the significance found in the simple linear model for variables such as inequality and young men's predisposition to crime. Conversely, it highlights the positive impact of education, family stability, urbanization, the unemployment rate, and apprehension risk in metropolitan regions. It is worth noting the presence of high crime clusters in Rio de Janeiro and Espírito Santo and low crime clusters in the regions of Minas Gerais and São Paulo. In summary, policies aimed at reducing inequalities, promoting education, strengthening family stability, investing in security, and creating job opportunities have the potential to have positive impacts on reducing homicide rates.

Keywords: Economics of Crime; Spatial Econometrics; Spatial Exploratory Data Analysis, Brazilian Southeast.

LISTA DE QUADROS

Quadro 1 - Matrizes de Vizinhança	36
Quadro 2 - Matriz de Vizinhança de Segunda Ordem	36
Quadro 3 - Matriz de Vizinhança com Pesos Binários	38
Quadro 4 - Matriz de Vizinhança com Pesos Padronizados	38
Quadro 5 - Categorias do Indicador de Associação Espacial Local.....	40
Quadro 6 - Interpretações de Impacto Modelo Spatial Lag (SAR).....	65

LISTA DE TABELAS

Tabela 1 - Variáveis Independentes Regressão Stepwise	51
Tabela 2 - Resultado Modelo MQO	53
Tabela 3 - Resultado Global Moran I	55
Tabela 4 - Resultado Multiplicador de Lagrange	57
Tabela 5 - Comparação dos Resultados do MQO e SAR	59
Tabela 6 - Impactos das Variáveis do Modelo (SAR).....	62

LISTA DE TABELAS

Figura 1 - Comparação de Modelos de Regressão Espacial	43
Figura 2 - Processo de Escolha Modelo de Regressão Espacial.....	46
Figura 3 - Taxa de Homicídio por Cem Mil Habitantes no Sudeste Brasileiro em 2010	49
Figura 4 - Taxa de Homicídio por Cem Mil Habitantes Suavizada no Sudeste Brasileiro em 2010	50
Figura 5 - Mapa dos Resíduos.....	54
Figura 6 - Correlograma de Moran I	55
Figura 7 - Indicador Local de Associação Espacial (LISA)	56

LISTA DE TABELAS

AEDE	Análise Exploratória de Dados Espaciais
GMM	Gaussian Mixture Model
IBGE	Instituto Brasileiro de Geografia e Estatística
IPEA	Instituto de Pesquisa Econômica e Aplicada
LISA	Local Indicator Spatial Assosiation
MQO	Mínimo Quadrados Ordinários
OLS	Ordinary Least Squares
SAR	Spatial Autoregressive Model
SEM	Spatial Error Model
SLM	Standart Linear Model
VIF	Variance Inflation Factor

SUMÁRIO

LISTA DE QUADROS.....	9
1 INTRODUÇÃO.....	14
2 FUNDAMENDOS DA ECONOMIA DO CRIME.....	19
2.1 Danos	20
2.2 O Custo de Apreensão e convicção.....	20
2.3 Ofensas.....	21
2.4 Punição.....	22
2.5 Revisão da Literatura	24
2.5.1 Determinantes do Crime.....	25
3 METODOLOGIA	33
3.1 Regressão Espacial.....	33
3.1.1 Autocorrelação Espacial	34
3.1.2 Matrizes de Vizinhança.....	35
3.1.4 Índice de Moran	39
3.1.5 Mínimos Quadrados Ordinários (MQO)	41
3.1.6 Modelo Spatial Lag (SAR)	42
3.1.7 Modelo Spatial Error (SEM)	44
3.1.8 Escolha do Modelo de Regressão Espacial	45
4. RESULTADOS.....	48
4.1 Modelo Mínimo Quadrados Ordinários (MQO)	48
4.2 Autocorrelação Espacial – Moran I e LISA	54
4.3 Seleção do Modelo de Regressão Espacial	57
4.4 Modelo Spatial Lag (SAR).....	58
4.4.1 Interpretação Estatística.....	60
5 ANÁLISE E DISCUSSÃO DOS RESULTADOS	63
6 CONSIDERAÇÕES FINAIS	65
REFERÊNCIAS	70

1 INTRODUÇÃO

Há procura de respostas sobre a criminalidade sempre foi algo estudado por diversas áreas acadêmicas, tendo como pioneirismo a sociologia, com estudos empíricos, baseado em observação. O desenvolvimento da criminologia pode ser sintetizado em duas classificações. A escola clássica desenvolvida no século XVIII, com a tentativa de reformar o sistema legal e proteger o acusado contra ações arbitrárias do estado e a escola positivista, com o intuito de aplicar métodos científicos ao estudo do crime. (JEFFERY, 1960)

A escola positivista ao contrário da clássica rejeitava a definição de crime como um termo legal, tendo como diferença, o foco de estudo, sendo o indivíduo criminoso, como elemento central, e não o crime em si. A caracterização do positivismo pode ser vista segundo Jeffery (1960) como a tentativa de responder por meio de estudos científicos a criminalidade.

A principal característica do positivismo é a tentativa de responder o enigma da criminalidade por meio de estudos científicos do indivíduo ofensor. O uso de métodos científicos é uma das principais características do positivismo. (JEFFERY, 1960, v.50, p.3)

Cesare Lombroso, criminologista italiano foi um dos pioneiros na visão positivista, mudando a atenção para o indivíduo no escopo dos estudos. A pergunta do “como e porquê?” as pessoas cometem crimes é a principal indagação do positivismo. (JEFFERY, 1960)

Neste sentido, diversas áreas acadêmicas contribuíram para esta pergunta, partindo da sociologia, psicologia, incluindo as ciências econômicas. A criação de modelos com fundo científico para explicações econômicas iniciou de forma mais constante a partir de Adam Smith, com considerações ao indivíduo como racional, custo benéfico e teorização sistemática de problemas econômicos.

Desta forma, a base científica do positivismo promoveu um ambiente propício à aplicação de modelos matemáticos, com fundo econômico, tendo como objetivo entender os fatores que estão relacionados com a decisão do indivíduo ao cometer um crime. A base destes estudos, se encontra em torno de análises econométricas espaciais das variáveis socioeconômicas, que podem se relacionar com a ação criminosa, em outras palavras, quais fatores podem influenciar na

decisão do indivíduo em cometer ou não um crime.

A literatura discute vários indicadores socioeconômicos que podem estar vinculado com a criminalidade. Segundo Jeffery (1960) a sociologia e psicologia enfatizam a importância da família na relação do crime, por outro lado, a economia busca relação com o desemprego, renda, desigualdade social, dentre outras.

Uma das teses científicas baseadas no pensamento econômico, que irá servir como base para este trabalho, é a de Becker (1968), explicitado em seu estudo, "*Crime and Punishment an Economic Approach*" (Crime e Punição: uma abordagem econômica, em tradução livre), onde um modelo econômico e matemático analisa a escolha de um indivíduo de cometer ou não um crime. Assumindo a racionalidade, uma pessoa cometeria um ato ilícito ao considerar que a utilidade de cometer tal delito supere a utilidade que ele poderia ter ao utilizar seu tempo em outra atividade legal.

A visão utilitarista traz um ideal sobre a escolha do indivíduo racional, ponderando os riscos e possíveis recompensas no momento de escolha da prática criminosa. A inclusão das variáveis socioeconômicas em modelagem matemática tem por objetivo identificar os possíveis determinantes que pesem na decisão do indivíduo, sendo uma das contribuições econômicas.

Com isso, diversos estudos foram feitos, em vários níveis territoriais, para as principais variáveis socioeconômicas e a relação com a atividade criminal. A literatura concorda que, por exemplo, um maior nível de educação, menor desigualdade, efetividade dos aparatos de segurança pública e sistema judiciário, levariam a uma redução no nível de criminalidade.

Contudo, no passar dos séculos vemos uma maior alfabetização, democratização, renda, controle inflacionário e diversos outros avanços socioeconômicos que contribuem para a melhora do bem-estar. Apesar disso, os indicadores de violência e criminalidade continuam a crescer de maneira constante. Filho e Reis (1999) explicitam este fenômeno como um paradoxo, onde apesar dos avanços socioeconômicos a criminalidade continua a aumentar, principalmente em grandes centros urbanos.

Portanto, este trabalho visa estudar a economia do crime, com referencial teórico pautado na teoria de Becker (1968), utilizando de modelagem a econometria

espacial e Análise Exploratória de Dados Espaciais (AEDE) para identificar a relação das principais variáveis socioeconômicas com a criminalidade na região Sudeste do Brasil.

Os objetivos específicos são: a) identificar as principais variáveis relacionadas com os determinantes do crime; b) atestar a relação espacial do modelo econométrico; c) analisar clusters de violência no sudeste brasileiro; d) definir modelo de regressão espacial apropriado; e) comparar os resultados encontrados a luz da literatura, esclarecendo os impactos das variáveis sobre o viés social e econômico.

A metodologia utilizada nessa pesquisa, considerando objetivo e finalidades, será feita por método empírico e pesquisa bibliográfica para método de comparação. Os dados coletados sobre criminalidade serão do Atlas da Violência do Instituto de Pesquisa Econômica Aplicada (IPEA). Em relação às variáveis socioeconômicas, o ATLAS Brasil, que sumariza dados por municípios, pautado no Instituto Brasileiro de Geografia e Estatística (IBGE) será a fonte utilizada.

A monografia será dividida em seis capítulos distintos. Inicialmente, introduzindo e contextualizando as abordagens científicas no quesito da pesquisa da economia do crime e o comportamento do indivíduo. Em seguida, será abordado o referencial teórico da economia do crime por meio de Becker (1968).

A revisão da literatura examinará estudos relevantes na área, focando os critérios de seleção das principais variáveis socioeconômicas e os resultados obtidos nessas pesquisas, sustentando as variáveis posteriormente escolhidas neste estudo.

O terceiro capítulo apresentará a metodologia, incorporando os princípios da econometria espacial e da Análise Exploratória de Dados Espaciais (AEDE). Em especial os modelos de regressão espacial e processo de seleção, enfatizando a diferença do modelo de regressão clássico.

O quarto capítulo discorrerá sobre os resultados da metodologia aos dados selecionados, evidenciando como a inclusão do componente espacial influenciou o modelo inicial clássico.

Finalmente, no capítulo cinco e seis será feita uma discussão e comparação dos resultados e conclusões de estudos semelhantes destacados na revisão da literatura. Essa análise visa observar as semelhanças e divergências dos produtos obtido neste estudo. Além disso, será fornecido esclarecimentos do ponto de vista econômico e social, acerca dos impactos das variáveis encontradas.

2 FUNDAMENDOS DA ECONOMIA DO CRIME

O referencial teórico da presente pesquisa procurou compreender a economia do crime, baseado no estudo, Becker (1968), que explicita a ação criminosa sendo pautada pela decisão do indivíduo racional, que analisa o custo benéfico, incluindo vantagens e riscos, pesando na decisão de cometer ou não um crime.

Em seu estudo sobre a economia do crime, Becker (1968) desenvolve, com bases econômicas, uma análise de políticas públicas e privadas visando minimizar o comportamento ilegal na sociedade. A consideração principal do modelo é pautada na decisão do indivíduo, considerando o custo-benefício em alocar seu tempo em uma atividade legal ou ilegal.

Esta decisão pode ser entendida segundo Viapiana (2006), dado a formulação abaixo, explicitando o modelo de Becker com o intuito de entender o determinante da atividade criminosa. (*apud* SALVIATO; MOURÃO, 2015, p. 356)

$$b > cp \quad (1)$$

Onde:

b: benefício gerado ao criminoso por praticar o delito;

c: custos da atividade criminosa;

p: probabilidade de apreensão

Desta forma, o indivíduo é incentivado a escolher a atividade criminosa quando o benefício (*b*) excede o custo da atividade criminosa (*c*) multiplicado pela chance de ser apreendido (*p*). (*apud* SALVIATO; MOURÃO, 2015, p. 356)

Neste sentido, a teoria é dividida em quatro (4) principais áreas de atuação, danos, custo de apreensão e convicção, ofensas e a punição. Nas próximas sessões, cada área de atuação terá um foco maior, explicitando a metodologia matemática de forma sintetizada como meio de sustentar a teoria.

2.1 Danos

Segundo Becker (1968), a quantidade de danos tende a aumentar com o maior nível de atividade criminal.

$$H'_i = \frac{dH_i}{dO_i} > 0 \quad (2)$$

Onde H_i é o dano da atividade i e O_i é o nível de atividade. Também temos que o valor social de ganho para o ofensor tende a aumentar com o número de ofensas.

$$\begin{aligned} G &= G(O) \\ G' &= \frac{dG}{dO} > 0 \end{aligned} \quad (3)$$

Logo, o custo ou danos a sociedade é a diferença entre dano e ganho.

$$D(O) = H(O) - G(O) \quad (4)$$

Estas métricas, apesar de serem importantes, não podem ser generalizadas, segundo Becker (1968), os custos a sociedade por atividade criminosa se diferem. O exemplo é o custo do assassinato, podendo ser medido simplesmente pela perda de renda da vítima, excluindo o valor dado a vida pela sociedade.

2.2 O Custo de Apreensão e convicção

Neste ponto, Becker (1968) traz a ideia dos gastos relativos a policiamento, corte, juizado, dentre outros fatores do aparato governamental público alocados em esforços referente ao combate à criminalidade.

Sendo postulado a relação entre os resultados provindos de policiamento, cortes e outros poderes, incluindo capital e materiais empregados para suprimir a criminalidade. (BECKER, 1968).

Estes resultados, provindos das diversas "Atividades" pode ser sumarizado na seguinte função, $A = f(m, r, c)$, onde f compila os esforços, dado os insumos. Logo, a "Atividade" seria mais dispendiosa, como sumariza a relação.

$$C = C(A)$$

$$C = \frac{dC}{dA} > 0 \quad (5)$$

Um aprimoramento para medir o nível de “Atividade”, seria o número de crimes inocentadas por condenação.

$$A \cong p0 \quad (6)$$

Onde p é a relação entre os delitos apurados, pelas condenações e todos os delitos. Em outras palavras, p seria a probabilidade geral de que um delito seja resolvido por condenação. A relação principal, é que aumentar a probabilidade de convicção ou o número de ofensas, aumentaria o custo total da ação criminosa. A formulação abaixo traz este pressuposto. (BECKER, 1968)

$$C_o = Cp > 0 \text{ se } p0 \neq 0 \quad (7)$$

Uma das argumentações na modelagem de regressão, é o fator dos investimentos e qualidade de policiamento, que envolvem esforços governamentais, com objetivo de monitorar e minimizar as atividades criminosas.

2.3 Ofensas

A literatura traz diversos pontos na consideração dos determinantes da atividade criminosa, sendo estes, especificados na revisão da literatura. Contudo, a maioria concorda, que com tudo ou mais constante, temos uma tendência a redução da atividade criminosa quando se aumenta a probabilidade de apreensão. Becker (1968) evidencia o método principal da escolha racional do indivíduo baseado na expectativa de utilidade.

A abordagem adotada aqui, segue a análise usual dos economistas sobre a escolha e assume que uma pessoa comete um delito se a utilidade esperada para ela exceder a utilidade que poderia obter usando seu tempo e outros recursos em outras atividades. Algumas pessoas se tornam “criminosas”, não porque a sua motivação básica se difere das outras pessoas, mas porque os benefícios e custos são diferentes. (BECKER, 1968, v. 76, p. 176, tradução livre)

Portanto, a metodologia se embasa na pressuposição que existe uma função, relacionando o número de ofensas de qualquer pessoa à sua probabilidade de condenação, à sua punição se condenado, e a outras variáveis, como a renda disponível em atividades legais além de outras atividades ilegais, a frequência de prisões e predisposição de cometer um ato ilegal. (BECKER, 1968)

$$O_i = O_i(p_i, f_i, u_i) \quad (8)$$

Onde O_i é o número de delitos que ele cometeria durante um determinado período, p_i sua probabilidade de condenação por delito, f_i sua punição por ofensa, e u_i uma variável que representa todas as outras Influências. (BECKER, 1968)

O aumento de p_i ou f_i reduziria a utilidade esperada por ofensa, em consequência, reduzindo o número de crimes, isto porque o risco por pagar pelo crime cometido iria aumentar, uma vez que a probabilidade de condenação ou a punição por ofensa aumentou. (BECKER, 1968)

2.4 Punição

A punição para a atividade criminosa é um fator discutido de forma polemica e com diversas vertentes. No âmbito político e de políticas públicas, um grande esforço é alocado em possíveis soluções para a redução da criminalidade. A punição acaba sendo um dos principais, pela facilidade e aceitação.

Alguns exemplos são a redução da maioria penal, maior rigor nas condenações, dentre outros.

De modo geral, a punição deve se encaixar ao crime, logo, para infrações mais simples se aplicaria multas, que são uma punição financeira para quem realiza a ofensa. Por outro lado, em casos mais sérios teria prisão, confisco de bens, dentre outros.

É importante ressaltar que o custo da punição não reside somente sobre o criminoso, e sim sobre toda sociedade. Por exemplo, na prisão, existe a exigência de pessoal, segurança física e tecnológica, manutenção, portanto, estes gastos acabam não sendo redirecionados a outros setores sociais, e são, em sua maioria, originários de tributos e impostos pagos por toda sociedade.

Sendo assim, Becker (1968) determina que o custo da punição é o custo para o indivíduo que cometeu a ofensa mais o custo ou menos o ganho de outros. A relação abaixo mostra a condição em termos do ofensor.

$$f' \equiv bf \quad (9)$$

Onde f' é o custo social e b é o coeficiente que transforma f em f' . O tamanho de b difere para cada tipo de punição. Para $b \cong 0$ temos crimes menores, com multas, enquanto $b > 1$ seria prisão, e outros tipos mais sérios de punição. (BECKER, 1968)

Quando se fala sobre as punições, como método de redução na atividade criminosa, é importante manter um certo nível de interpretação, pois existem ideologias distintas relacionadas a punição, a ideia como sempre é encontrar um equilíbrio. Em outras palavras, uma penalidade que se encaixa no crime.

Isto é evidenciado principalmente em crimes menores, uma maior punição não significa uma redução nas ofensas, podendo gerar o resultado contrário, uma vez que o risco pelo crime fica drasticamente maior, levando conseqüentemente a maior violência em crimes menores.

Em suma, Becker (1968), concentra a ideia de determinação de um modelo que permita a melhor alocação de recursos para políticas públicas, visando reduzir as ações criminosas, incluindo variáveis de gastos governamentais na segurança pública, policiais, cortes, juizados, dentre outros, que ajudariam a aumentar a probabilidade de um crime ser descoberto p , por conseqüência, aumentando a chance do ofensor ser preso e condenado.

O peso da punição pode ser visto como f , sendo o tipo de punição, prisão em regime fechado ou aberto, multa, tudo dependendo do nível de severidade do crime cometido.

Sendo assim, participar em atividades ilegais poderia ser explicado como uma escolha econômica, de um indivíduo racional, onde, alocar tempo em certa atividade ilegal geraria maior retorno, considerando os riscos, do que uma atividade legal.

É relevante destacar as diferenças de cada sujeito na consideração de danos causados e riscos, levando em conta que cada um pode ter uma visão em relação ao que é considerado atividade ilegal.

Para alguns, qualquer taxa salarial estabelecida em mercados de trabalhos competitivos é permitida, enquanto para outros, taxas abaixo de um certo mínimo são violações dos direitos básicos; para alguns, o jogo, a prostituição e até o aborto devem ser livremente disponíveis para qualquer pessoa disposta a pagar o preço de mercado, enquanto para outros, o jogo é pecado e o aborto é homicídio. Estas diferenças são fundamentais para o desenvolvimento e implementação de políticas públicas, mas foram excluídas da minha investigação (BECKER, 1968, v.76, p. 209, tradução livre)

Em linhas gerais, a teoria trabalha com o consenso de dano e benefício, alinhando com o custo-benefício e a decisão do indivíduo, contribuindo para uma melhor visão na alocação de recursos com bases econômicas.

As especificações de danos, custo de apreensão e convicção, ofensas e punição, permitem uma seleção de variáveis, adaptável a cada periodicidade de estudo e dados disponíveis para o nível territorial escolhido. O objetivo é contemplar os quatro eixos que impactam na decisão de cometer um crime.

2.5 Revisão da Literatura

Os trabalhos que relacionam a economia do crime com modelagem econométrica espacial, de modo geral, procuram responder à questão central: *“Quais as variáveis socioeconômicas que impactam as atividades criminais?”*.

Desta forma, esta monografia utilizou como revisão literária, estudos, em diversos níveis territoriais, no âmbito nacional e internacional. Evidenciando as justificativas para cada escolha de variável e seus respectivos resultados, esclarecendo se os impactos seguem o esperado pela lógica econômica.

Com isso, será possível ter uma base consolidada de categorias de variáveis comumente utilizadas, assim como as principais problemáticas encontradas, servindo como suporte para a construção do modelo de regressão deste trabalho.

Por fim, esta análise vai permitir uma comparação dos produtos obtidos neste estudo com os resultados já visitados na literatura, confirmando tendências já estabelecidas ou evidenciando divergências.

2.5.1 Determinantes do Crime

A ideia presente nos determinantes do crime está baseada no conceito teórico de Becker (1968), onde o indivíduo racional considera os riscos e benefícios no momento de decidir alocar seu tempo em uma atividade legal ou ilegal.

Os fatores envolvidos nos riscos e benefícios tem relação com o aparato do estado, no quesito de aumentar a probabilidade de apreensão e condenação, por meio de investimentos em segurança pública, judiciário e outros. Ademais, servindo propulsor do aumento das recompensas em atividades legais, ligado diretamente a melhor condição socioeconômica de maneira geral.

No tocante das severidades e penas aplicadas, que se relaciona com o custo-benefício, uma vez que, se punição por determinado crime não gera perdas o suficiente, quando comparado com os possíveis ganhos, a conduta criminosa do indivíduo se torna viável.

Segundo Silva e Riker (2019), este ponto é central no pensamento de Becker, além de ser um delimitador, uma vez que as punições estão ancoradas em códigos penais rígidos que já estabelecem as sentenças e possíveis regressões de regimes, podendo impactar diretamente no processo de escolha do indivíduo.

Outro ponto discutido são os gastos relacionados a prevenção, visto como delimitador. De acordo com Silva e Riker (2019) os gastos se iniciariam desde antes do indivíduo se tornar criminoso, atuando de forma repressiva e não razoável para o aparato Estatal.

A reflexão sobre a colaboração da população também é uma inconsistência evidenciada por Silva e Riker (2019), apesar de considerar o senso comum, não é regra que os indivíduos considerem o dano causado como significantes. Isto pode ser visível principalmente em regiões de alta violência, crimes menores podem não ser considerados significantes pela população.

De acordo com Becker (1968) as motivações dos indivíduos criminosos não se diferem das demais pessoas, o custo-benefício entra como determinante no processo de decisão. Do ponto de vista filosófico existem divergências em relação a este pensamento, Silva e Riker (2019) elucidam que Rousseau enfatiza que o homem nasce bom, e a sociedade o corrompe, já em Becker (1968), o homem é racional e a partindo do custo-benefício o homem se transforma em criminoso ou não.

De modo geral, as reflexões que envolvem o pensamento de Becker são diversas, e partem para outras áreas da ciência assim como as limitações do arcabouço governamental, judicial e outros. Contudo, a perspectiva econômica e matemática permite adaptações para estudos de cunho econométrico, elucidando variáveis significantes que impactam o processo de escolha do indivíduo.

Neste sentido, de início, o primeiro grupo de variáveis explanatórias que entram no escopo da análise giram em torno da renda e desigualdade. A renda por si só acaba não sendo significativa segundo FAJNZYLBER; LEDERMAN; LOAYZA, (1998), indicando que maior desigualdade está associada com maior homicídio intencional, contudo, o nível de renda per capita não é significativo.

O problema em relação a análise única da renda ou renda per capita é que ela possui dois vieses de interpretação, no modelo de Becker (1986) a decisão do indivíduo é pautada no custo-benefício em participar ou não de atividade ilegal, considerando riscos e recompensas.

O maior nível de renda, tende a ser interpretado como uma maior recompensa por realizar atividades legais, portanto, reduzindo a atratividade de alocar o tempo em atividades ilícitas. Isto pode levar a uma interpretação perigosa, que considera regiões pobres como mais propensas ao crime. Contudo, outro ponto de vista, é que, com maior renda per capita, a recompensa pela atividade criminosa também aumenta. Logo, em regiões com maior nível de renda, existiria uma maior oportunidade para grandes ganhos, considerando a atividade criminal.

Por este motivo, a literatura considera a renda, contudo, engloba o componente de desigualdade como o mais relevante, pois este sim, estaria positivamente relacionado com as atividades criminosas.

Segundo FRANCISCO; FAJNZYLBER (2001), ao analisarem as taxas de homicídio, abrangendo os estados da federação brasileira de 1981 até 1996, foi encontrado que um aumento de 1% na desigualdade, medida pela fração da renda total dos 20% mais pobres, levaria a uma queda de 3,1% na taxa de homicídio.

Scorzafave e Soares (2009) em um estudo para os municípios do estado de São Paulo, considerando o impacto da renda e outras variáveis sobre crimes com retornos financeiros, detectando uma elasticidade de 1,46, segundo os dados do censo demográfico do ano 2000.

Ambos os estudos indicam que o componente da desigualdade é relevante e positivamente relacionado com a criminalidade, tanto no âmbito dos homicídios como crimes com retornos financeiros.

A literatura traz diversas maneiras de tratar a desigualdade, Scorzafave e Soares (2009) utilizam a renda mediana, combinação de porcentagem da população mais pobre e indicadores de desigualdade, como Gini e Theill.

Outro ponto relevante apontado por FAJNZYLBER; LEDERMAN; LOAYZA, (1998), é que a desigualdade é uma variável “sensível à política”, portanto, ações públicas que se voltem para a redução da criminalidade devem estar atentas a desigualdade.

De modo geral, o consenso é que a desigualdade apresenta um peso relevante para a análise das ações criminosas, contudo é importante ter cuidado com a análise única da renda, sendo preferível realizar transformações para tratar mais do ângulo desigual.

Continuando na linha de variáveis de teor mais econômico, como renda e desigualdade, temos o desemprego. Ehrlich (1973), em seu estudo “*participation in illegitimate activities*” (Participação em atividades ilegais, em tradução livre), conclui que os resultados relacionados a taxa de desemprego, relacionado a taxa de crimes, são, em suas palavras, “desapontadores”.

A utilização da variável pode servir como um indicador complementar de oportunidade de renda, que estaria disponível no mercado. Por outro lado, MAJID MADDAH, (2013), traz uma análise empírica da relação da taxa de desemprego e taxa crimes relacionados a roubos para o Irã. Os resultados encontrados são significativos e positivos.

Os resultados também são positivos para o caso brasileiro, FRANCISCO; FAJNZYLBER, (2001) estimam que para um acréscimo de 1% no desemprego teria resultado de um aumento em 3,4% nas taxas de homicídio. O estudo de Scorzafave e Soares, (2009) também encontra significância para a taxa de desemprego no estado de São Paulo, com dados do censo de 2000.

Em linhas gerais, a taxa de desemprego apresenta divergência entre certos estudos, contudo, no caso brasileiro se mostrou significativa na análise. Em suma, as recomendações de políticas públicas acrescentam a redução na taxa de desemprego como positiva, uma vez que melhoram a estabilidade econômica e contribui para o aumento na recompensa em atividades legais.

Quanto se trata de dados com peso espacial, a densidade demográfica acaba sendo uma das variáveis que apresentam grande relação com as atividades criminais, principalmente pois está relacionada com as recompensas e os riscos.

Em seu estudo, *“Why is There More Crime in Cities?”* (Por que há mais crimes nas cidades, em tradução livre), GLAESER; SACERDOTE, (1999), trazem o conceito da urbanização e densidade demográfica como um fator importante na análise de variáveis que influenciam as ações criminosas.

É discutido principalmente o prêmio relacionado ao crime urbano, isto porque, grandes cidades podem tornar os seus moradores mais propensos ao crime, quando comparado com cidades pequenas. Os fatores podem ser relacionados a menor chance de ser detectado, apreendido e punido, isto porque, as ações policiais tendem a ser mais efetivas em cidades menores. (GLAESER; SACERDOTE, 1999)

Ademais, áreas com alta densidade apresentam uma maior diversidade de vítimas, que podem ser escolhidas com maior facilidade pelo indivíduo criminoso, ampliando as suas possíveis recompensas.

Também é enfatizado um dos grandes problemas dos dados criminais, sendo este a subnotificação, de acordo com GLAESER; SACERDOTE, (1999), indivíduos em grandes cidades estão menos inclinados a denunciar atividades criminosas, quando comparado com cidades menores.

Nesse caso, o número de prisões por crime relatado pode ser o mesmo em cidades grandes e pequenas, contudo o número de prisões por crime real pode ser muito menor nas grandes cidades. (GLAESER; SACERDOTE, 1999)

A literatura analisada também concorda com a importância da densidade demográfica, MAJID MADDAH, (2013) e FAJNZYLBER; LEDERMAN; LOAYZA, (1998), às utilizam em seus estudos, Scorzafave e Soares (2009) utilizam a porcentagem de população urbana em 2000, adicionando também uma variável binária para regiões metropolitanas.

Neste sentido, as referências tendem a ir ao encontro, localidades mais densas demograficamente tendem a estar relacionadas de maneira positiva com maiores taxas de criminalidade, incluindo financeiros e homicídios.

Até o momento foram revisadas variáveis de teor mais específico e objetivo, como renda, desigualdade, desemprego e densidade demográfica.

Contudo, outros fatores podem afetar a decisão do indivíduo no momento da escolha em realizar ou não uma atividade ilícita.

Com isto, é introduzido a variável de instabilidade da família, no sentido que, indivíduos em famílias mais estáveis, tenderiam a ter uma menor inclinação a realizar atividades criminosas.

Uma das maneiras de medir a instabilidade familiar, segundo GLAESER; SACERDOTE, (1999), é com a compreensão de famílias chefiadas por mulheres. Explicitando a ausência da figura paterna, tanto para o desenvolvimento pessoal como suporte econômico. Neste tocante, é evidenciado outra relação, onde em grandes cidades existe a tendência de maior concentração de famílias monoparentais.

Neste sentido, o estudo *“The impact of income and Family Structure on Delinquency”* (O impacto da renda e estrutura familiar na delinquência, em tradução livre), por COMANOR; PHILLIPS, (2002), considera a influência dos pais extremamente importante na economia familiar, evidenciando que o indicador da renda, por si próprio, é muito menos importante, criticando a análise da violência durante a adolescência, que observa em primeiro lugar a renda familiar.

De acordo com a análise empírica, os incentivos financeiros governamentais a famílias sem presença paterna, teria pouco efeito na delinquência adolescente. COMANOR; PHILLIPS, (2002) reforça que pais substitutos também teriam pouco efeito.

O estudo *“Factors Affecting Juvenile Delinquency”* (Fatores que afetam a delinquência juvenil, em tradução livre), de LOBOS, J. (2018), entrevista mais de 105 pessoas, de idades de 15 a 31 anos, concluindo que quase metade (44,8%) viviam em familiar nucleares e mais de um quarto (26,7%) pertenciam a famílias desestruturadas.

Os resultados de LOBOS, J. (2018) relacionam que para um terço dos entrevistados (35,2%) o lar não teve influência para que se tornassem indivíduos delinquentes, por outro lado, quase um quarto (22,9%) relataram que este fator teve influência.

FRANCISCO; FAJNZYLBER, (2001) utilizam a instabilidade familiar em seu estudo, como famílias chefiadas por mulheres, e concluem que para um aumento de 1% na porcentagem de famílias instáveis teria como impacto um aumento de 4,5% na taxa de homicídio, concordando com COMANOR; PHILLIPS,

(2002) na escolha de variável e confirmando sua relevância para o recorte brasileiro.

De modo similar, Scorzafave e Soares (2009) utilizam para instabilidade familiar a porcentagem de mães adolescentes entre 15 e 17 anos no período de 1991 para os municípios do estado de São Paulo, obtendo resultado significativo.

A concepção da fragilidade familiar é ampla, podendo ser tratado de diversas formas, principalmente dependendo da disponibilidade dos dados para o nível territorial desejado, contudo, se considera como uma variável relevante, que para o caso brasileiro, nos estudos revisados, estão presentes e apresentam significância.

Como visto, a instabilidade familiar enfatiza um efeito mais forte na faixa etária mais jovem, que acaba sendo a mais afetada nas famílias nucleares, desta forma, a variável de predisposição do jovem a atividades criminais também pode ser incluída no estudo.

Os jovens, de modo geral, estão mais predispostos a escolherem por atividades criminosas, LOBOS, J (2018) conclui que 78,1% dos respondentes em suas entrevistas para avaliar a delinquência juvenil eram entre 17 e 21 anos, enquanto 17,1% estavam entre 12 e 16 anos.

Além disso, conclui que uma das principais razões para a maior predisposição dos jovens é a pressão dos pares, sendo a persuasão exercidas nos indivíduos de um mesmo grupo social, tendo uma maior influência para tornar o indivíduo delinquente. LOBOS, J (2018) também aponta que os correspondentes perceberam a escola como um fator que menos impacta.

A importância da faixa etária também é vista em FRANCISCO; FAJNZYLBER, (2001), onde ele indica que para os mais jovem as taxas de homicídios ficam bem maiores do que as de faixa etária mais velhas. Também encontra que para a desorganização social o indicador teve representação significativa até 32 anos de idade.

Com isso, uma variável de predisposição para os jovens é consenso na maioria dos estudos, FRANCISCO; FAJNZYLBER (2001) concluem que o problema da criminalidade é mais acentuado nos jovens, principalmente os homens, contando com efeitos maiores de variáveis como renda, desemprego e chefia de família.

A linha de pensamento para a população mais jovem continua com a variável de nível educacional, como visto, este recorte de população apresenta maiores efeitos de variáveis socioeconômicas e impactos na criminalidade.

Contudo, a educação, similar a renda, apresenta uma dupla interpretação. A primeira, é mais intuitiva, onde um maior nível educacional em determinada região resultaria em melhores oportunidades no futuro, aumentando a recompensa pelo trabalho legal, assim como outros fatores de conscientização e socialização.

Por outro lado, existe um segundo viés, regiões com maior nível educacional, poderiam representar locais com maior nível de recompensa para o crime, uma vez que, sua população mais especializada, tendem a aumentar o prêmio e recompensa por atividades ilegais.

Neste sentido, FAJNZYLBER; LEDERMAN; LOAYZA, (1998) apresentam a relação de educação e crime como um quebra cabeça, um dos fatores para justificar a inconsistência pode vir pelo atraso no esforço educativo na redução na criminalidade, ou seja, o efeito da educação não se materializa quando os jovens estão sendo educados, mas quando se tornam adultos.

Por sua vez, LOBOS J, (2018) indica que 30,5% dos entrevistados relataram que a escola teve menos influência em sua delinquência, seguindo por mais de um quarto 27,6% que afirmaram não ter influência alguma.

O efeito da escolaridade é utilizado nos estudos revisados, e a justificativa se encontra principalmente na análise da população mais jovem, onde o consenso é que são uma faixa etária mais afetada pelas variáveis socioeconômicas, portando maiores níveis educacionais tendem a mitigar efeitos de menor renda, desemprego e outros no longo prazo.

Por fim, temos a variável relacionada ao risco para cometer a atividade ilegal, neste caso existe algumas opções, dependendo da disponibilidade de dados em determinado recorte.

O nível de investimento em segurança pública pode ser uma proxy para o risco em cometer atividade ilegais, pois em tese aumentaria as chances de ser descoberto e apreendido. Pode-se considerar também a taxa de encarceramento, indicando o nível de eficiência do âmbito judiciário em identificação, apreensão e condenação.

Outros fatores relacionados ao policiamento podem ser utilizados, como índice de investimento específico nos mecanismos policiais, com propósito de reduzir o crime, como investimentos em treinamento policial, equipamentos de armamentos, tecnologias, veículos.

A variável de regiões metropolitanas e densidade demográfica podem funcionar como *proxy*, relacionando o menor risco em ser apreendido nestes municípios, servindo como suporte, no caso de insuficiência de dados de segurança e/ou investimento policial ao nível territorial e periodicidade do estudo.

Em linhas gerais, as literaturas explicitadas trazem uma variedade de variáveis, assim como suas justificativas em termos de custo e benefício. As justificativas são baseadas na racionalidade do indivíduo, e interpretação de impactos socioeconômicos.

Portanto, é possível identificar cinco categorias principais de variáveis, sendo estas, desigualdade, predisposição do homem jovem, escolaridade, instabilidade da família, urbanização e risco. Esta divisão servirá como base para o processo metodológico de construção e estimação do modelo.

3 METODOLOGIA

O presente trabalho teve como objetivo aplicar, sobre a econometria espacial e Análise Exploratória de Dados Espaciais (AEDE¹), os pressupostos apresentados na teoria de Becker (1968). Foi levado em consideração as contribuições de estudos similares, visitados na revisão da literatura, servindo como base para a estruturação do modelo.

A metodologia tem como objetivo discutir as análises fundamentais dos conceitos que orbitam Econometria Espacial e Análise Exploratória de Dados Espaciais (AEDE).

Sendo assim, as referências bases foram: “*Advanced Statistics in Criminology and Criminal Justice*” (Estatística Avançada em Criminologia e Justiça Criminal, em tradução livre) de Weisburd et al. (2022) e “*Exploring spatial data with GeoDaTM*” (Explorando dados espaciais com GeoDATM, em tradução livre) por Anselin, L. (2005).

Neste tópico, inclui definições sobre Autocorrelação Espacial, Matrizes de Vizinhos, Indicador de Moran e os Modelos de Regressão, sendo eles, Método dos Mínimos Quadrados Ordinários (MQO), Modelo Spatial Lag (SAR) e Modelo Spatial Error (SEM), juntamente com o processo de escolha de regressão espacial por Anselin, L (2005).

3.1 Regressão Espacial

Os modelos de regressões tradicionais, analisam dados de séries, corte, temporais e em painéis, considerando que as observações e regiões são independentes entre elas. (LESAGE, J. P., 2008)

Diversos tipos de modelos podem ser analisados sem o componente regional, contudo, por racionalidade, podemos considerar que mesmo não sendo explicitado no modelo tradicional o conceito de espaço e região sejam importantes.

Um exemplo de LESAGE, J. P (2008) seria, com um modelo de regressão que relaciona o tempo de transporte para o trabalho de uma região i para um número de pessoas na região i , assumindo que a escolha de método e densidade do bairro na região j não teria uma influência no tempo de transporte da região i .

¹ AEDE é um combinado de técnicas que descrevem e visualizam dados espaciais, identificando padrões e anomalias. O método ajuda no entendimento e identificação de padrões em dados com forte componente espacial.

Logo, assumir que as linhas de transporte público, incluindo estrutura e densidade demográfica dos bairros não possuem relação entre si, e, portanto, seriam independentes é irrealista.

Neste sentido, o campo da Econometria Espacial utiliza de técnicas para incorporar o componente de dependência entre observações, incluindo as regiões do espaço que estão próximas geograficamente. (LESAGE, J. P., 2008).

Os tópicos a seguir contém os principais conceitos e interpretações, que são utilizadas no processo de construção de um modelo e análise da Regressão Espacial.

3.1.1 Autocorrelação Espacial

O exemplo do transporte público, já mencionado, e a relação espacial que eles compartilham entre bairros, estruturas e outros, é uma de muitas interações em que o espaço apresenta grande peso.

Os dados de criminalidade, tendem a ter uma forte Correlação Espacial, uma vez que regiões com maiores níveis de violência tendem a estar cercadas por outras regiões violentas, apesar de haver exceções. (WEISBURD et al., 2022)

Também existe a relação com outras variáveis, como visto no estudo de Becker (1968) e os determinantes do crime, a influência de desigualdade, educação, esforço policial, dentre outras, tendem a pesar na balança de risco e benefício, impactando a escolha do indivíduo em escolher ou não praticar uma atividade ilícita.

A Correlação Espacial se refere à força da associação entre casos e entre o espaço, explicitando o quão similar as observações são entre os vizinhos.

As dependências espaciais estão relacionadas as interações no espaço, causando agrupamentos de eventos, surgindo devido a interação das observações que estão próximas geograficamente. (WEISBURD et al., 2022).

Neste caso, poderíamos ter como exemplo, regiões extremamente violentas, que espalham para as regiões próximas maiores níveis de violência. A intensidade da violência em uma região afeta a intensidade da violência nas localidades próximas. Em suma, a Correlação Espacial traz à força entre variáveis no espaço.

Por outro lado, heterogeneidade espacial está relacionada a estrutura do espaço, causando uma distribuição irregular de evento. (WEISBURD et al., 2022). Um exemplo seria o nível de violência aumentar durante grandes eventos, como nas Olimpíadas do Rio de Janeiro.

Estes pontos são importantes, uma vez que afetam a escolha de modelagem para a regressão, dependendo da estrutura dos dados e o objetivo da pesquisa.

Em suma, a Autocorrelação Espacial é uma das primeiras análises feitas no modelo de regressão linear básico, pois a intenção do estudo é detectar a relação espacial exercida pelas variáveis, assim como o grau de correlação entre elas, considerando o componente de vizinhança.

3.1.2 Matrizes de Vizinhança

A Matriz de Vizinhança, é necessária para detectar e corrigir Correlação Espacial. Ela é um modo numérico de representar relações de vizinhança entre pontos, sendo estes, pares de coordenadas x e y , representando latitude e longitude ou polígonos, representando a combinação de três ou mais pontos de latitude e longitude. (WEISBURD et al., 2022).

Os principais métodos de identificação de vizinhança estão pautados em contiguidade, distância ou número de vizinhos. Cada método se ajusta de maneira diferente aos dados observados. Analisaremos com mais detalhe cada um dos princípios.

A matriz de contiguidade espacial com pesos se preocupa com o conceito dos adjacentes. Utilizado comumente para polígonos, a contiguidade verifica se um polígono está tocando no outro, assim como o modo em que ele está tocando. (WEISBURD et al., 2022)

Os dois principais métodos de contiguidade são o da rainha e torre. Exemplificado pelo xadrez, temos que no método de contiguidade da rainha, todos os polígonos adjacentes, incluindo os que se encontram nos vértices são considerados vizinhos.

Pelo método de contiguidade da torre, temos os polígonos dos vértices excluídos como vizinhos, mantendo apenas os que se encontram nas bordas,

similares aos movimentos das peças em um jogo de xadrez. O quadro 1 traz a exemplificação de ambas as contiguidades.

Quadro 1 - Matrizes de Vizinhança

Contiguidade da Rainha			Contiguidade do Rei		
	Cédula Selecionada				
	Cédula Vizinha				
	Cédula Excluída				

Fonte: Fabricação própria com base em Weisburd et al. (2022), *Advanced Statistics in Criminology and Criminal Justice*. 5. ed. Switzerland: Springer, 2022. p. 504, fig. (12.2)

Ambos os métodos permitem adaptação no nível de células ou polígonos, que estão sendo considerados como vizinhos, neste caso do Quadro 1, temos ordem um (1) de contiguidade, poderíamos aumentar para ordem dois, adicionando à matriz, vizinhos apenas de segunda ordem.

É importante mencionar que os vizinhos de primeira ordem não serão considerados na matriz de segunda ordem automaticamente, a não ser que seja especificado. O Quadro 2 traz o exemplo do funcionamento de ordens na matriz de vizinhança.

Quadro 2 - Matriz de Vizinhança de Segunda Ordem

	Cédula Selecionada			
	Cédula Vizinha			
	Cédula Excluída			

Fonte: Fabricação própria com base em Weisburd et al. (2022), *Advanced Statistics in Criminology and Criminal Justice*. 5. ed. Switzerland: Springer, 2022. p. 504, fig. (12.3)

Outra opção, como mencionado, é a matriz de vizinhança com peso baseada na distância. Normalmente calculada para a escolha de polígono vizinhos, e pode ser utilizada com o centroide de cada polígono, portanto, supondo que o limite de vizinhança seja um raio de 50 quilômetros, todos os polígonos que possuem centroides² dentro deste limite vão ser considerados vizinhos.

Por outro lado, existe a opção de utilizar a matriz por número de vizinhos, chamada de K-Vizinhos Mais Próximos (KNN) onde k indica o número de vizinhos que é incluído na análise. Este método adiciona o vizinho mais próximo primeiro e segue para o posterior, até chegar no limite determinado. (WEISBURD et al., 2022).

O método KNN não exige contiguidade, a distância entre polígonos também não é um problema. Ele é útil para resolver complicações que normalmente acontecem com regiões isoladas ou ilhas, pois estes não possuem relação de contiguidade. Contudo, assinamos de maneira fixa um número de vizinhos para todos os polígonos, em outras palavras, todos terão um número de vizinho, independente da localidade.

Até o momento considerou-se as diversas matrizes de vizinhança e as situações nas quais seriam feitas as suas utilizações, dependendo da característica geográfica ou das necessidades de análise dos dados. Contudo, durante o estudo, podemos concluir que vizinhos mais próximos teriam um peso maior na nossa análise.

Desta forma, será explicitado a formulação e utilização dos conceitos relacionados as matrizes, observando como são feitas as métricas e atribuição de pesos.

O método mais simples de quantificar a relação entre vizinhança é utilizando pesos binários, quando um par (i, j) de unidades espaciais, podendo ser polígonos, são consideradas como vizinhos tendo $w_{ij} = 1$, se não $w_{ij} = 0$. Isto se aproxima da matriz de correlação, onde a diagonal reflete uma unidade espacial com si mesmo. (WEISBURD et al., 2022)

De modo geral, temos duas opções, a matriz com pesos binários e a com pesos padronizados em linha. Na primeira, é considerado como sendo vizinho

$w_{ij} = 1$ ou não sendo vizinho $w_{ij} = 0$, a soma da linha da matriz pode ter valores diversos, dependendo do número de vizinhos. O Quadro 3 traz a

² Centroide ou centro geométrico, representa o centro do polígono, neste caso do município em questão.

exemplificação da matriz com pesos binário.

Quadro 3 - Matriz de Vizinhança com Pesos Binários

		Região				Soma
		1	2	3	4	
Região	1	0	1	1	0	2
	2	1	0	1	0	2
	3	1	1	0	1	3
	4	0	0	1	0	1

Fonte: Fabricação própria com base em Weisburd et al. (2022), *Advanced Statistics in Criminology and Criminal Justice*. 5. ed. Switzerland: Springer, 2022. p. 506, fig. (12.4)

Neste caso, é assinando para vizinhos o valor 1 e para não vizinhos o valor 0. Neste exemplo a região 4 é vizinha apenas da região 3, enquanto a região 3 é vizinha da 4, 2 e 1. Pode-se observar que a soma das linhas na matriz de contiguidade com pesos binários varia, dependendo da quantidade de regiões e respectivas vizinhanças.

Neste caso, a matriz com a padronização das linhas atribui proporcionalmente pesos para determinadas vizinhanças, reduzindo a soma de todas as linhas para 1.

Quadro 4 - Matriz de Vizinhança com Pesos Padronizados

		Região				Soma
		1	2	3	4	
Região	1	0	.50	.50	0	1
	2	.50	0	.50	0	1
	3	.33	.33	0	.33	1
	4	0	0	1	0	1

Fonte: Fabricação própria com base em Weisburd et al. (2022), *Advanced Statistics in Criminology and Criminal Justice*. 5. ed. Switzerland: Springer, 2022. p. 506, fig. (12.4)

Apesar da padronização, os pesos assinados a cada vizinhança são os mesmos, sendo a divisão do total pelo número de vizinhos, portanto, proporcional. Para atribuir peso maior nas vizinhanças mais próximas e decair com o nível de distância a utilização da matriz de vizinhança com pesos inversa é necessária. (WEISBURD et al., 2022)

De modo geral, a escolha de matriz de vizinhança é parte fundamental do processo de construção da análise, pois consiste no modo que o espaço e os vizinhos são tratados e como eles vão se relacionar com as variáveis escolhidas, impactando o resultado obtido juntamente com as interpretações.

3.1.4 Índice de Moran

A matriz de vizinhança com pesos é utilizada para calcular o indicador de Autocorrelação Espacial Moran I. Sendo o coeficiente que quantifica a força da Correlação Espacial, estando entre 1 e -1 . Desta forma, 1 indicaria existência de forte Autocorrelação Espacial positiva, e -1 negativa. (WEISBURD et al., 2022).

A definição do indicador de Moran I, segundo Weisburd et al. (2022) é dado pela equação abaixo.

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{w \sum (x - \bar{x})^2} \quad (10)$$

Onde:

x : Número de unidades espaciais;

x_i : Valor de um intervalo de uma variável em uma localidade i ;

x_j : Valor de outra variável em uma outra localidade j ;

\bar{x} : A média da variável;

w : É a soma de todos os pesos espaciais w_{ij}

Desta forma, caso não exista nenhuma Autocorrelação Espacial, o Moran I pode ser descrito de acordo com a fórmula abaixo.

$$E(I) = \frac{1}{n - 1} \quad (11)$$

O indicador de Moran I é interpretado de maneira similar ao coeficiente de correlação. Caso computemos o indicador para taxa de homicídio, o mesmo irá retornar um valor que indica o quão similar ou não em média a taxa de homicídio é comparado com os vizinhos próximos, pautado pela matriz de vizinhança escolhida.

Um exemplo de Weisburd et al. (2022) está relacionado com o teste global de Moran I para chamadas reportando crimes contra propriedade recebidas pela polícia, sendo um único valor que sumarizaria a Autocorrelação Espacial.

Considerando os registros pelo CEP, o resultado mostra uma Autocorrelação Espacial positiva, uma vez que taxas de crime contra propriedade são similares aos CEPs vizinhos.

Usualmente, juntamente ao teste de Moran I, é a utilização do Indicador de Associação Espacial Local (LISA) para visualização espacial. Ele ilustra em um nível mais detalhado e micro como a Correlação Espacial varia entre regiões.

A relação é padronizada para quatro classes, com o objetivo de identificar agrupamentos baseado nas Autocorrelação Espacial.

De modo geral, categorias Alta-Alta mostra clusters relacionados a altos níveis de criminalidade em um determinado município que por sua vez é acompanhado de municípios vizinhos com altos níveis de criminalidade.

O mesmo pressuposto segue para os cluster Baixo-Baixo, onde identifica-se regiões com agrupamentos de municípios com baixos níveis de criminalidade.

Quadro 5 - Categorias do Indicador de Associação Espacial Local

Categoria	Descrição
Alta-Alta	Indicador de similaridade, mostrando que áreas vizinhas tem alto valores. (i.e regiões que possuem altos níveis de violência, e por consequência seus vizinhos também possuem altos níveis de violência.)
Baixa-Baixa	Indicador de similaridade, mostrando que áreas vizinhas tem baixos valores. (i.e regiões que possuem baixos níveis de violência, e por consequência seus vizinhos também possuem baixos níveis de violência.)
Alta-Baixa	Indicador de não similaridade, mostrando que áreas vizinhas tem baixos valores comparados com a área central. (i.e regiões extremamente ricas cercadas de áreas extremamente pobres.)
Baixa-Alta	Indicador de não similaridade, mostrando que áreas vizinhas tem altos valores comparados com a área central.

Fonte: Fabricação própria, 2023

3.1.5 Mínimos Quadrados Ordinários (MQO)

O entendimento dos Mínimos Quadrados Ordinários (MQO) é importante para os outros modelos que serão apresentados a frente, apesar de não representar um bom ajuste para dados com relação espacial ele é o início da análise e estruturação da relação que é objetivo do estudo.

O MQO representa uma relação linear entre uma ou diversas variáveis independentes com uma variável dependente. Weisburd et al. (2022) traz a simplificação do modelo na formulação abaixo.

$$y = x\beta + u \quad (12)$$

Sendo y a variável dependente e x a independente, β é coeficiente de regressão, representando a magnitude da relação estudada. Por fim temos u , que retrata o erro aleatório do modelo.

O MQO traça a melhor regressão ajustada, minimizando a Soma dos Erros Quadrados (SQR). Considerando que a diferença entre os valores da predição do modelo e as observações reais variam de forma positiva e negativa, gerando assim os erros u , temos primeiro a necessidade da transformação dos Erros Quadrados para posteriormente buscar pelo menor erro possível no somatório.

$$\begin{aligned} SQR &= \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \\ &= \sum_{i=1}^n \hat{u}_i^2 \end{aligned} \quad (13)$$

Desta forma, o MQO utiliza como base de estimativa para os coeficientes os erros quadrados, contudo, a regressão assume que estes resíduos possuem variância constante (Homoscedasticidade). No caso de dados com Autocorrelação Espacial, os erros podem não ser estocásticos (aleatórios), existindo assim, padrões de associação entre regiões próximas.

Um exemplo seria tentar quantificar países que realizam esforços para combater corrupção dentro de suas legislações, realizando ativamente práticas de investigação, punição e outros, para reduzir práticas corruptas dentro do governo.

Ao realizar um modelo de regressão com MQO para prever este resultado em um mapa, com os erros de cada país, teríamos que a predição seria melhor em países com democracias mais fortes e legislações estáveis, pois estes seguem uma regulação mais estrita neste específico caso. (WEISBURD et al., 2022)

O fator de proximidade geográfica também seria significativa, uma vez que países da união europeia teriam legislações similares, dado a características históricas de desenvolvimento. (WEISBURD et al., 2022)

O fator principal, é que a heterocedasticidade (resíduos possuem variância desigual), viola o princípio da regressão pelo MQO, estando presente nos erros das relações espaciais e levantando a hipótese de que o componente espacial deve ser tratado de maneira especial.

Com isso, temos que a regressão pelo MQO, com dados que possuam relação espacial não é recomendado, uma vez que as estimativas podem conter vies devido a violação do princípio da homoscedasticidade dos erros.

Sendo assim, apesar de ser um bom início de exploração das relações entre variáveis, outras opções são aconselhadas, sendo as principais, Spatial Lag (SAR) e Spatial Error (SEM).

3.1.6 Modelo Spatial Lag (SAR)

Segundo Weisburd et al. (2022) o modelo de Spatial Lag (SAR) é utilizado para regressão quando a variável dependente é influenciada pelos seus vizinhos, sendo um dos modos mais comum para investigar dependência espacial. O modelo lida com a dependência incorporando a defasagem da variável dependente.

$$y = \rho W y + x\beta + u \quad (14)$$

O modelo é similar ao MQO, contudo adiciona a defasagem da variável espacial dependente. Temos para y um valor n por um vetor de observações na variável dependente ao longo do espaço.

Adicionado ao modelo de MQO, os valores dos vizinhos y que possuem o peso da matriz espacial W , multiplicado por ρ , representando o coeficiente autorregressivo para a variável dependente. (WEISBURD et al., 2022).

Rho (ρ) indica a força da autocorrelação espacial presente na variável dependente. Tendo como teste de Hipótese:

H0: Não autocorrelação espacial

HA: Autocorrelação espacial

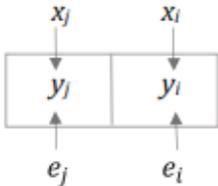
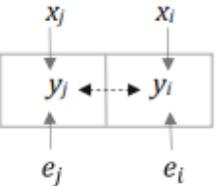
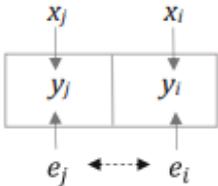
Sendo que, x é uma n por k da matriz de variáveis independentes, β é um k por um vetor de respectiva regressão de coeficiente, e u é uma n por um vetor de variáveis independentes e resíduos normalmente distribuídos. (WEISBURD et al., 2022).

Neste caso, a estimativa de y depende de $x\beta$ e ρW , sendo assim, a interpretação de $x\beta$ não é a mesma que no MQO, pois consideramos o efeito espacial para a variável y na matriz de vizinhança.

De acordo com Weisburd et al. (2022), no modelo MQO o efeito da variável independente na variável dependente é constante ao longo das observações. Isso não acontece quando temos incorporado o efeito espacial no modelo, pois ele varia os resultados para diferentes vizinhanças em cada observação do banco de dados.

Portanto, quando queremos descobrir um processo de interação espacial e os efeitos nos vizinhos na variável dependente podemos utilizar o modelo de Spatial Lag (SAR). A Figura 1 traz uma sintetização da relação entre os modelos de regressão espacial.

Figura 1 - Comparação de Modelos de Regressão Espacial

	OLS	SPATIAL LAG	SPATIAL ERROR
Estimation process	Least squares of errors	Maximum likelihood	Maximum likelihood
Location of spatial lag	N/A	Dependent variable as separate coefficient	Within error term
Equation	$y = x\beta + e$	$y = \rho W y + x\beta + e$	$y = x\beta + e$, where $e = \lambda W e + \xi$
Adjusts for	None	Spatial dependence	Spatial heterogeneity
Neighbor influence	 <p>No neighbor influence</p>	 <p>Dependent variable influenced by neighbors</p>	 <p>Residuals influenced by neighbors</p>

Fonte: (WEISBURD et al.). *Advanced Statistics in Criminology and Criminal Justice*. 5. ed. Switzerland: Springer, 2022. p. 552, tab. (12.3)

Podemos observar que diferente do modelo MQO, utiliza-se no Spatial Lag (SAR) a seleção de parâmetros com maior verossimilhança, deixando de lado os Mínimos Quadrados Ordinários, isto traz um β , ρ e λ mais ajustado para dados com relação espacial.

Para a influência dos vizinhos, como dito anteriormente, no modelo Spatial Lag (SAR), os valores de y em uma unidade i está diretamente relacionado aos valores de y encontrados em vizinhos j . (WEISBURD et al. 2022).

3.1.7 Modelo Spatial Error (SEM)

O modelo Spatial Error (SEM) é utilizado na regressão quando acreditamos que a autocorrelação espacial observada tem relação com o processo espacial cuja intensidade varia ao longo do espaço (Heterogeneidade espacial). (WEISBURD et al., 2022).

Este modelo lida com heterogeneidade espacial estimando o coeficiente espacial dentro do erro da regressão. O modelo inclui um termo de erro para os vizinhos, definido pela matriz de vizinhança com pesos W , juntamente com o termo de erro usual. (WEISBURD et al., 2022).

$$\text{Onde} \quad y = x\beta + u \quad (15)$$

$$\begin{aligned} u &= \lambda Wu + \varepsilon|\lambda| \\ u &= \lambda Wu + \xi \end{aligned} \quad (16)$$

Temos que u é um compilado de λ , um coeficiente espacial autorregressivo para os resíduos que captura a dependência espacial, Wu , sendo os resíduos das localizações dos vizinhos que estão na matriz espacial de vizinhança com pesos, e, ξ , um vetor independente e com erros distribuídos de forma idêntica, chamados de ruídos brancos. (WEISBURD et al., 2022).

Em outras palavras, de acordo com Darmofal (2015) a dependência espacial observada nos dados não reflete um verdadeiro processo espacial, mas meramente um agrupamento geográfico das fontes do comportamento de interesse.

Um exemplo seria o estudo de criminalidade em uma área que seria influenciado não só pelo seu fator, mas também pela criminalidade das regiões vizinhas, desta forma, o modelo capta a correlação espacial por meio dos termos de erros, como explicitado na formulação (15), logo os resíduos são influenciados pelos vizinhos.

A diferença do modelo Spatial Error (SEM), reside no fato que o coeficiente que indica o efeito dos vizinhos na variável dependente não é estimado separadamente igual no modelo Spatial Lag (SAR), isto porque, ele incorpora dentro do termo de erro como explicitado na formulação (16). (WEISBURD et al., 2022)

Portanto, o Spatial Error (SEM) considera a dependência espacial como possível variável omitida ou não observada, relacionada à correlação espacial, influenciando o termo de erro da regressão. O Spatial Lag (SAR), considera a dependência espacial ao incluir uma versão defasada da variável dependente como uma variável independente, assumindo que esta variável em uma localidade é influenciada não só pelas variáveis independentes, mas também pelos valores da variável dependente dos vizinhos.

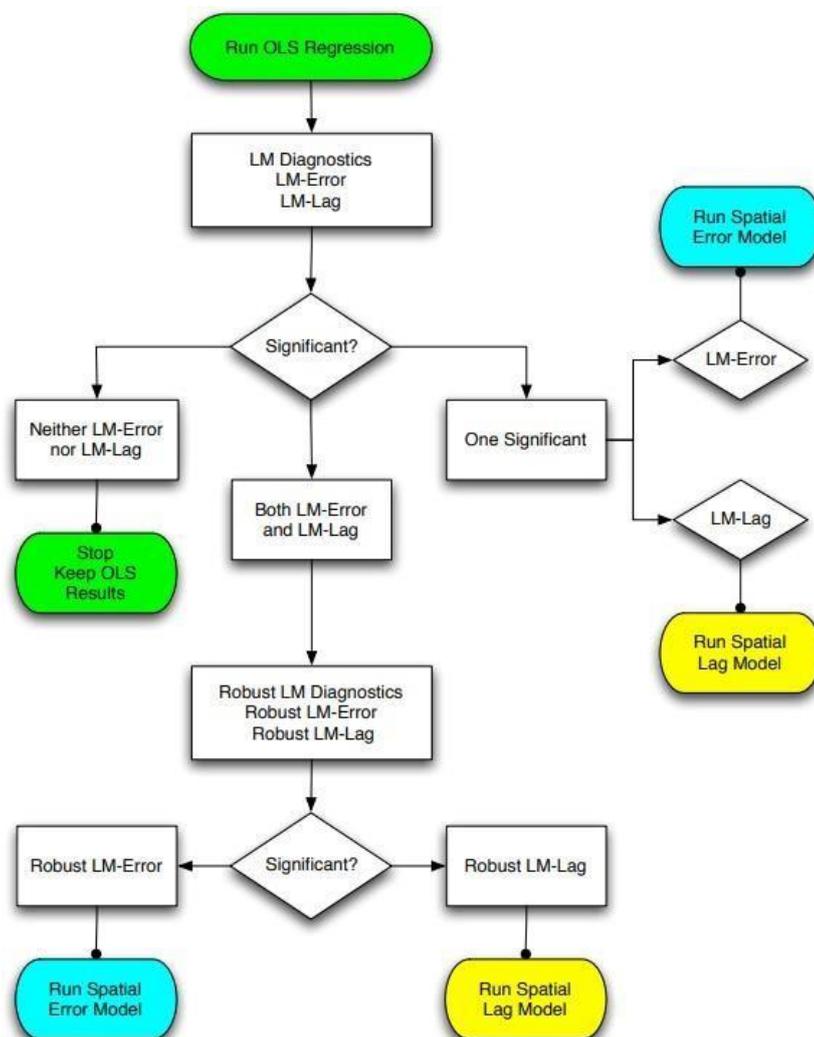
Com dois modelos de regressão espacial explicitados, evidenciando suas diferenças, a dúvida posta é qual seria a melhor escolha para os dados analisados. Neste sentido, o próximo tópico mostra a metodologia de seleção de modelo de regressão espacial baseado em Anselin. L (2005).

3.1.8 Escolha do Modelo de Regressão Espacial

O processo de seleção de modelo mais apropriado para os dados e a relação estudada é importante, pois não existe certeza se os erros observados na autocorrelação espacial provem da dependência espacial ou da heterogeneidade espacial.

Anselin. L (2005) especifica um processo de decisão que seleciona o modelo de regressão mais apropriado. A Figura 2 traz os passos realizados para a escolha final da regressão espacial.

Figura 2 - Processo de Escolha Modelo de Regressão Espacial



Fonte: (ANSELIN, 2005). Exploring Spatial Data with GeoDaTM: A Workbook. Illinois: Center For Spatially Integrated Social Science, 2005. p. 199, fig. (23.25)

O processo inicia com a construção de um modelo MQO (OSL), com objetivo de estudar a relação entre as variáveis e evidenciar a existência ou não de Autocorrelação Espacial nos resíduos.

Além disso, o modelo MQO traz o primeiro diagnóstico da relação entre as variáveis, podendo explicitar outliers e multicolinearidade extrema.

Anselin. L (2005) utiliza dois testes de multiplicador de Lagrange para acessar a dependência espacial, contudo é comum acessar os resíduos do modelo MQO e realizar o teste de Moran I.

Para realizar a identificação de Autocorrelação Espacial é necessário a criação de uma matriz de vizinhança. A escolha de qual tipo de matriz parte da

natureza da relação estudada.

Com ambos os testes de multiplicador de Lagrange, podemos identificar a Autocorrelação Espacial do modelo, e, de maneira mais específica, selecionar qual deles proveria um melhor ajuste, Spatial Lag (SAR) ou Spatial Error (SEM).

Onde Spatial Lag (SAR) testa por uma variável dependente espacialmente defasada, que estaria ausente e Spatial Error (SEM) testa a autocorrelação dos resíduos espaciais. (WEISBURD et al., 2022)

Adicionalmente, para ambos os testes, existe a versão robusta, que no caso de suficiência estatística para ambos os testes não robustos, serve como o último divisor na decisão de escolha do modelo.

Em síntese, o processo inicia com a criação do modelo MQO e análise da relação estudada, seguida pela identificação de autocorrelação espacial, provinda dos testes de Moran I e Lagrange, com versão robusta. Nesta etapa, dado a significância dos testes, a escolha é definida entre Spatial Lag (SAR) e Spatial Error (SEM), como mostra a Figura 2.

4. RESULTADOS

Este capítulo consiste nos resultados dos processos metodológicos. Incluem a construção e estimação modelo MQO, esclarecendo sobre a seleção de variáveis, pautado em Becker (1968) e na revisão da literatura. Além disso, é ordenado o processo de análise de Autocorrelação Espacial e escolha do modelo de regressão espacial, sistematizado por Anselin. L (2005).

Desta forma, os resultados obtidos no modelo de regressão espacial são interpretados e comparados com o MQO, levando em consideração as razões socioeconômicas, impactos esperados e a relevância do componente espacial para as variáveis analisadas.

Por fim, é feito um confronto dos resultados obtidos com o que foi visto na literatura, explicitando as principais semelhanças e divergências dos resultados obtidos e as possíveis justificativas para tal.

4.1 Modelo Mínimo Quadrados Ordinários (MQO)

Todo o processo realizado na construção metodológica e manipulação de dados foi feito no Software de Estatística e Programação R. O código para reprodução dos resultados podem ser visualizados no repositório GitHub³.

A variável dependente utilizada no modelo neste estudo foi a taxa de homicídio por cem mil habitantes para todos os municípios do sudeste brasileiro, extraída do Atlas da Violência do Instituto de Pesquisa Econômica Aplicada (IPEA).

Esta taxa passou pelo processo bayesiano⁴ de suavização empírica, isto está ancorado na ideia de que taxas observadas podem ter problemas, principalmente quando existe diferentes tamanhos de amostra que variam por área.

Neste caso, o estudo das taxas de homicídio por municípios pode trazer um desajuste quando o município em questão tem uma população menor, quando comparado com outros municípios.

(...) a plotagem das taxas observadas pode ter sérias desvantagens quando os tamanhos das amostras variam por área, uma vez que taxas observadas muito altas (e baixas) são encontradas desproporcionalmente e em áreas com baixa amostragem. (GELMA; PRICE 1999, 3221-34 apud MEDINA; SOLYMOSI, 2022, tradução livre).

³ [Repositório GitHub](#) – Contém código e bancos de dados utilizados na produção dos resultados.

⁴ A técnica aplica teorema de Bayes para estimar o padrão subjacente nos dados. O principal objetivo é reduzir ruídos e lidar com inconsistência de valores extremos em baixa quantidade em uma determinada região, distorcendo a interpretação e visualização do modelo.

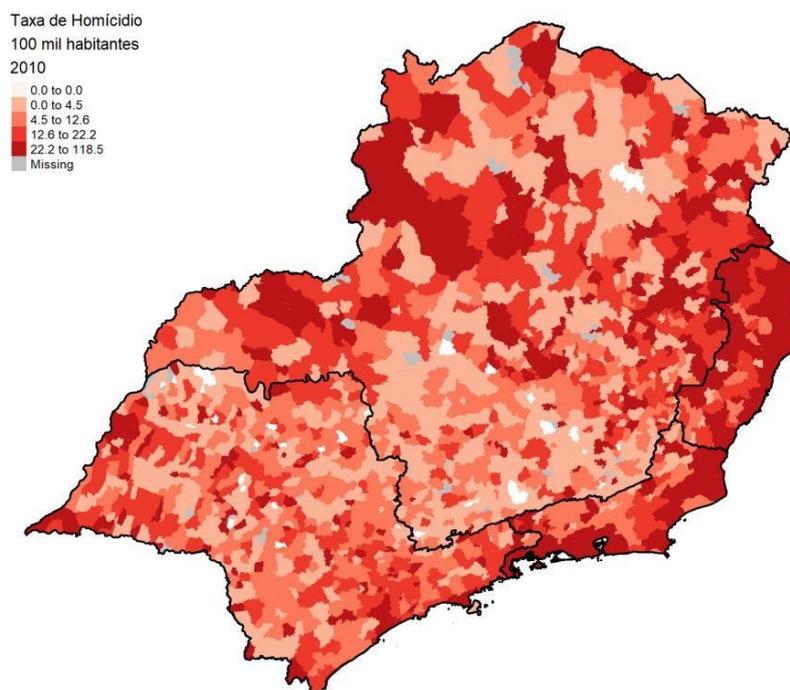
A criação da taxa foi feita no software de programação e estatística R, por meio do pacote *DCluster*, elaborado por epidemiologistas. A função utilizada foi *empbaysmooth()*, que cria um risco relativo suavizado de certos valores esperados e números observados de casos.

Com os resultados, é possível utilizar os parâmetros obtidos para criar o risco relativo bruto que por sua vez é utilizado para criar o risco relativo suavizado. (MEDINA; SOLYMOSI, 2022)

Por fim, é elaborada uma taxa local para o estimador suavizado, o objetivo é não apenas reduzir o risco global criado para cada município, algo que poderia mascarar heterogeneidades não observadas. (MEDINA; SOLYMOSI, 2022)

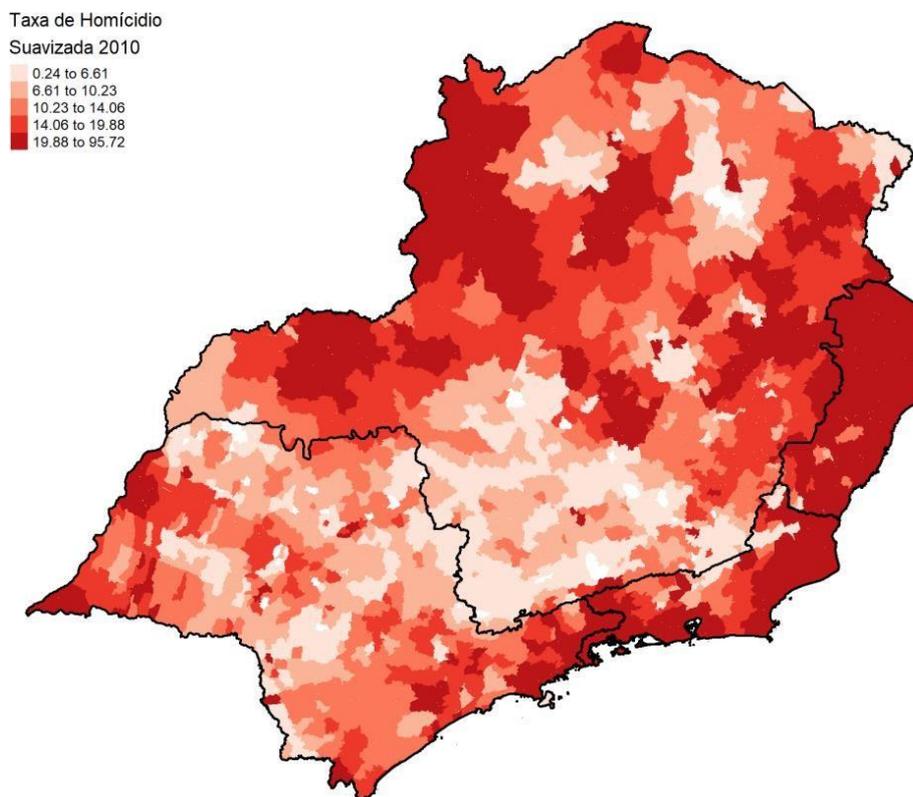
Para computar este estimador de suavização local é utilizado a matriz de vizinhança, neste caso a de contiguidade com pesos de ordem um. Desta forma, o risco bruto é reduzido, levando em consideração a taxa da vizinhança. As Figuras 3 e 4 trazem a comparação da taxa de homicídio por 100 mil habitantes e a taxa suavizada.

Figura 3 - Taxa de Homicídio por Cem Mil Habitantes no Sudeste Brasileiro em 2010



Fonte: Fabricação própria com base nos dados do Atlas da Violência e IBGE, 2023

Figura 4 - Taxa de Homicídio por Cem Mil Habitantes Suavizada no Sudeste Brasileiro em 2010



Fonte: Fabricação própria com base nos dados do Atlas da Violência e IBGE, 2023

A visualização evidência que a suavização melhora a qualidade da estimativa, tornando-a uma variável mais estável, corroborando para um melhor ajuste na estimação do modelo.

Em relação as variáveis independentes, uma divisão de categorias foi feita previamente, considerando o referencial teórico de Becker (1968) e a revisão da literatura.

É contemplado os principais tópicos utilizados para a análise socioeconômica e impactantes na atividade criminal, incluindo variáveis relacionadas a desigualdade, predisposição do homem jovem, escolaridade, instabilidade da família, urbanização e risco.

Todas as variáveis são de 2010 a nível municipal, assim como a taxa de homicídio por cem mil habitantes suavizada. Tal preferência segue a perspectiva de Ingram e Costa (2016), que justificam a escolha pela proximidade com a variável dependente de homicídio, além da disponibilidade em nível municipal.

Os dados socioeconômicos utilizados nas variáveis independentes foram retirados do Atlas Brasil e Instituto Brasileiro de Geografia e Estatística (IBGE). Portanto, a construção dos modelos deste estudo foi pautada em dados decorte.

Tabela 1 - Variáveis Independentes Regressão Stepwise

Categoria	Variáveis
Desigualdade	Indicador de Gini
	Indicador de Theill
	% de População Extremamente Pobre
	% de População Pobre
	% da Desocupação de 18 Anos ou mais
	Índice de Desenvolvimento Humano (IDHM)
	Índice de Desenvolvimento Humano da Renda (IDHM Renda)
Predisposição do Homem Jovem	% População Homem de 15 a 29 Anos
Escolaridade	% de Crianças de 6 a 14 Anos Fora da Escola
Instabilidade da Família	% de Mulheres de 15 a 17 Anos com ao menos um filho
Urbanização	% População Urbana
	População Total dividido por Mil
Risco	Binária para Região Metropolitana

Fonte: Fabricação própria, 2023

Para escolha do melhor ajuste, um modelo de regressão *stepwise* foi feito, similar a Silva et al. (2020). Em linhas gerais, a regressão *stepwise* é uma técnica que inicia com um modelo completo e em cada passo, elimina gradualmente variáveis independentes do modelo de regressão.

Desta forma, considerando um agrupamento de variáveis independentes gerais, pautadas na revisão de literatura e teoria de Becker (1968), o processo de seleção *stepwise*⁵ recomenda um ajuste com melhor ajuste considerando os dados em questão.

O modelo final pode ser visto no Quadro 6. O resultado provindo da regressão *stepwise* considerava IDHM e IDHM Renda, contudo, estas foram removidas devido ao problema de multicolinearidade.

⁵ O método *stepwise* seleciona o modelo com melhor ajuste ao adicionar ou remover variáveis independentes potenciais de maneira sucessiva, testando a significância estatística em cada iteração.

Quadro 6 - Variáveis Independentes Finais e Expectativa de Impacto

Variáveis	Expectativa de Impacto	Fonte
In(% População Homem de 15 a 29 Anos)	+	ATLAS BRASIL
In(% de Mulheres de 15 a 17 Anos com ao menos um filho)		
log(% de População Extremamente Pobre)		
In(Indicador de Gini Normalizado)		
In(% Desocupação de 18 Anos ou mais)		
In(% População Urbana)		
In(% de Crianças de 6 a 14 Anos Fora da Escola)		
Binária para Região Metropolitana		IBGE

Fonte: Fabricação própria, 2023

A utilização do log para porcentagem da população extremamente pobre foi pautada em Ye, X., & Wu, L. (2011, April 1). Já a seleção da log da taxa de homicídio por cem mil habitantes suavizada foi feita na realização de testes na regressão *stepwise*. A versão em log promoveu um resultado mais robusto do que a taxa de homicídio por cem mil habitantes suavizada.

A normalização do indicador de Gini foi feita com intuito de trazer para uma escala de maior comparabilidade, além de ser um ponto considerável para a escolha de modelo pelo processo *stepwise*, que permite alocar um maior número de variáveis independentes para escolha de um melhor ajuste.

De modo geral, é contemplado todas as categorias evidenciadas, com a inclusão de ao menos uma variável para cada.

Por fim, temos a expectativa dos resultados, levando em consideração qual seria o impacto das variáveis independentes na taxa de homicídio por cem mil habitantes suavizada em log. Neste caso, o impacto esperado é positivo para todas.

O resultado do modelo dos Mínimos Quadrados Ordinários (MQO) pode ser visto na tabela 2. Ele confirma os impactos esperados das variáveis selecionadas.

Tabela 2 - Resultado Modelo MQO

	Variável Dependente:	
	log(Taxa Homicídio Suavizada)	
	(MQO)	(VIF)
Intercepto	1.868*** (0.165)	
% Pop. Homem 15 a 29	1.882* (1.092)	1.06
% Pop. Mulher 15 a 17 Ao Menos um Filho	1.447*** (0.476)	1.04
log(% Pop Extremamente Pobre)	0.101*** (0.023)	2.59
Índice de Gini Normalizado	0.060*** (0.019)	1.18
% Desocupação 18 Anos ou Mais	3.537*** (0.802)	1.58
% Pop. Urbana	0.288*** (0.134)	2.21
% Crianças de 6 a 14 Anos Fora da Escola	4.835*** (1.168)	1.08
Binária Região Metropolitana	0.323*** (0.045)	1.15
Observações		1,631
R2		0.134
R2 Ajustado		0.130
Resíduos Erro Padrão		0.695 (df = 1622)
Estatística F		31.349*** (df = 8;1622)
Critério de Informação Akaike (AIC)		3454.9
Log-Verossimilhança		-1717.44 (df=10)
Nota:	*p<0.1; **p<0.05; ***p<0.01	
Erro Padrão em parênteses		
VIF: Fator de inflação de variância (Mede multicolinearidade)		

Fonte: Fabricação própria com base nos dados do Atlas Brasil, IBGE e Atlas da Violência, 2023

O modelo apresenta significância estatística para todas as variáveis, assim como um R^2 ajustado de 0.130, portanto, as variáveis independentes explicam em média, 13% da variação da taxa de homicídio por cem mil habitantes suavizada em log.

Em relação a multicolinearidade, a exclusão das variáveis de IHDM e IDHM Renda foram positivas, pois resultaram em um modelo com todas as variáveis restantes com $VIF < 2.60$.

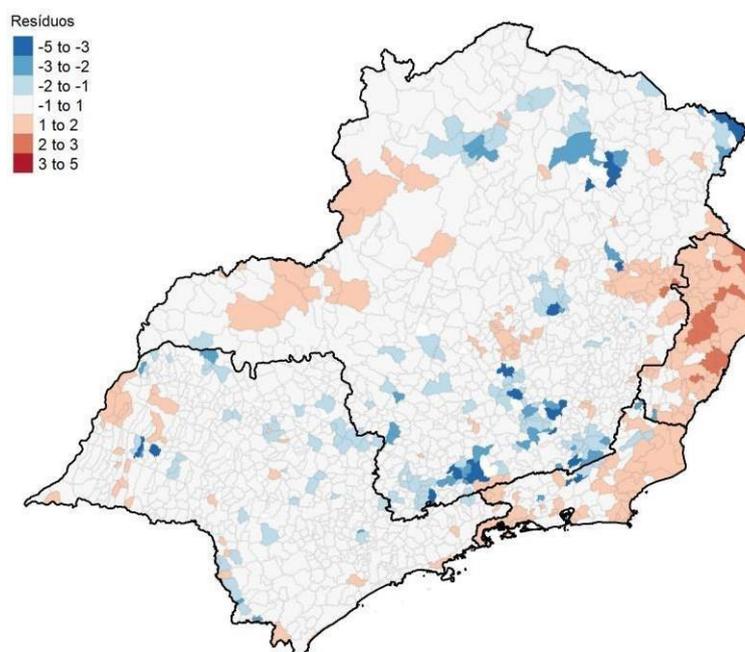
Após a análise inicial da relação estudada, assim como a confirmação da significância das variáveis e do modelo, é dado início a investigação sobre Autocorrelação Espacial, com o teste de Moran I e a visualização dos cluster com o mapa de LISA.

4.2 Autocorrelação Espacial – Moran I e LISA

Primeiramente, como início na análise da autocorrelação espacial um mapa dos resíduos do modelo linear foi feito. A ideia é olhar de modo geral o padrão da distribuição espacial dos resíduos.

Como método de visualização é utilizado o desvio padrão dos resíduos, ele indica quanto que os dados diferem da média, neste caso, cada desvio padrão se torna uma categoria no mapa *coropléticos* que pode ser visualizado Figura 5. (MEDINA; SOLYMOSI, 2022)

Figura 5 - Mapa dos Resíduos



Fonte: Fabricação própria com base nos dados do Atlas Brasil, IBGE e Atlas da Violência, 2023

O resultado mostra que certas áreas em vermelho estão superestimando e outras áreas em azul estão subestimando. Além disso, é possível visualizar clusters, principalmente no estado do Rio de Janeiro e Espírito Santo, indicando que possivelmente existe Autocorrelação Espacial.

Um dos possíveis passos para atestar a autocorrelação espacial é o teste de Moran I, nele, são fornecidos o modelo linear e a matriz de vizinhança, neste caso a de contiguidade da rainha com ordem um. O resultado pode ser visualizado na Tabela 3.

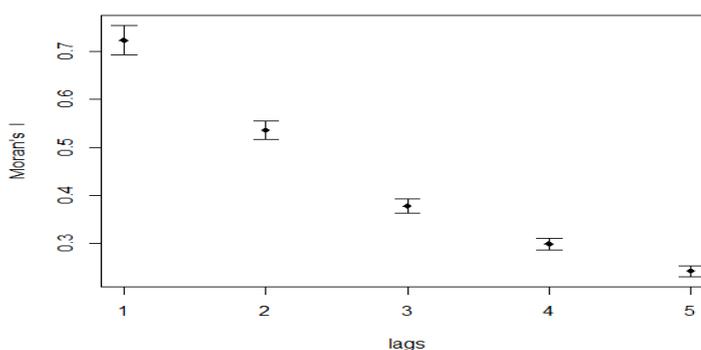
Tabela 3 - Resultado Global Moran I

Global Moran I for regression residuals	
Moran I Observado	0.538***
Expectativa	-0.002
Variância	0.0002
Estatística Moran I SD	35.597
Note:	*p<0.1; **p<0.05; ***p<0.01
Fonte: Fabricação própria com base nos dados do Atlas Brasil, IBGE e Atlas da Violência, 2023	

O teste apresenta significância estatística e o Moran I Observado indica uma autocorrelação moderada e positiva. Portanto, podemos concluir que para este modelo, os resíduos estão relacionados sistematicamente.

O correlograma do índice de Moran I evidencia a autocorrelação espacial, impactada pelo aumento na ordem dos vizinhos. Portanto, quanto maior o *lag*, ou o número da ordem de vizinhança, menor é a semelhança. A figura 6 mostra esta relação.

Figura 6 - Correlograma de Moran I



Fonte: Fabricação própria com base nos dados do Atlas Brasil, IBGE e Atlas da Violência, 2023

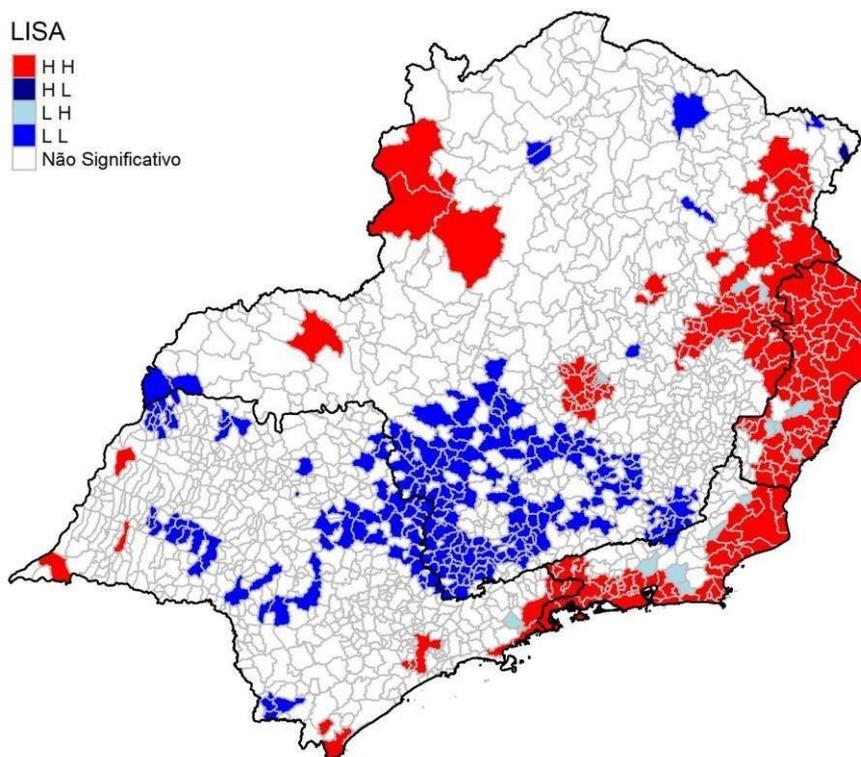
Em outras palavras, ao considerar um município, o seu vizinho de ordem cinco, por exemplo, teria uma semelhança menor do que seu vizinho de ordem um.

Como fim de análise de Autocorrelação Espacial, cabe a construção do Indicador Local de Associação Espacial (LISA). Ele especifica de modo micro, permitindo encontrar clusters de autocorrelação. Como explicitado, as categorias são Alto- Alto, Baixo-Baixo, Alto-Baixo e Baixo-Alto.

De modo geral, é possível identificar áreas onde a taxa de homicídio por cem mil habitantes suavizada em log apresenta altos níveis e são acompanhadas pelos seus municípios vizinhos. A mesma lógica interpretativa segue para as outras categorias.

A representação Figura 7 traz as categorias e especifica os clusters em regiões onde o teste de local de Moran obteve significância estatística à um p – value < 0.1 .

Figura 7 - Indicador Local de Associação Espacial (LISA)



Fonte: Fabricação própria com base nos dados do Atlas Brasil, IBGE e Atlas da Violência, 2023

De modo geral, é possível identificar que as regiões do Rio de Janeiro e Espírito Santo apresentam clusters Alto-Alto, enquanto partes de São Paulo e

Minas Gerais apresentam clusters Baixo-Baixo, com exceção de certos municípios, dos quais se encaixa São Paulo capital.

Com a análise de autocorrelação espacial, foi possível atestar que apenas um modelo linear não suporta a relação das variáveis, pois o componente espacial é relevante, logo, os modelos de regressão espacial poderiam gerar um melhor ajuste. (MEDINA; SOLYMOSI, 2022)

4.3 Seleção do Modelo de Regressão Espacial

Após atestar a autocorrelação espacial do modelo linear e a construção do teste de Moran I, com a matriz de vizinhança de contiguidade da rainha de ordemum, e a visualização dos clusters, é necessário concluir qual o melhor modelo de regressão espacial que se ajuste aos dados em questão.

O método de escolha é pautado em Anselin. L (2005), onde se utiliza doteste de multiplicador de Lagrange para atestar a significância do modelo para os testes de Spatial Lag (SAR) e Spatial Error (SEM) e devido os critérios de exclusão, escolher o melhor ajuste. O resultado do teste pode ser visto na Tabela 4.

Tabela 4 - Resultado Multiplicador de Lagrange

Multiplicador de Lagrange - Diagnostico de Dependência Espacial	
LMerr	< 0.01
LMlag	< 0.01
RLMerr	< 0.99
RLMlag	< 0.01

Fonte: Fabricação própria com base nos dados do Atlas Brasil, IBGE e Atlas da Violência, 2023

O resultado mostra que para os testes não robustos LMlag e LMerr há significância em < 0.01. Sendo assim, seguindo o processo de seleção os testes robustos para LMlag e LMerr são necessários.

O fim do processo de decisão de modelo de regressão espacial reside na escolha do Spatial Lag (SAR), o teste robusto RLMLag é estatisticamente significativo em < 0.01 enquanto o RLMerr não mostra significância.

Anselin. L (2005) resume o processo e as devidas escolhas pautadas no multiplicador de Lagrange.

Quando ambos os testes estatísticos do multiplicador de Lagrange rejeitarem a hipótese nula, proceda para a parte inferior e considere as formas robustas dos testes. Normalmente, apenas uma delas será significativa, ou uma terá uma ordem de significância superior a outra (e.g., $p < 0.000001$ comparado com $p < 0.03$). Nesse caso a decisão é simples: estime o modelo de regressão espacial correspondente à estatística robusta (mais) significativa. Na rara instância em que ambas forem altamente significativas, opte pelo modelo com maior valor para a estatística de teste. No entanto, nessa situação, é necessário algum cuidado, pois pode haver outras fontes de inadequação. Uma ação óbvia a ser tomada é considerar os resultados para diferentes pesos espaciais e/ou alterar as especificações básicas (ou seja, não a parte espacial) do modelo. Existem também casos raros em que nenhuma das estatísticas de teste de Lagrange robusto é significativa. Nestes casos, é provável que existam problemas mais sérios de especificações e esses devem ser abordados primeiro. (ANSELIN, 2005, p.198-200, tradução livre)

Por fim, podemos seguir estimando e interpretando o modelo Spatial Lag (SAR) para os dados em questão, verificando se a matriz de vizinhança escolhida provê um bom ajuste.

4.4 Modelo Spatial Lag (SAR)

A escolha do modelo Spatial Lag (SAR) vem após a realização de todas as etapas previstas na metodologia, incluindo a primeira análise com o modelo MQO, atestar a autocorrelação espacial com Moran I e posteriormente escolher, com base nos testes do multiplicador de Lagrange, incluindo os robustos, o modelo Spatial LAG (SAR) ou Spatial Error (SEM).

O modelo proposto (SAR), lida com a dependência espacial ao introduzir a defasagem da variável dependente como uma variável independente. Isso significa que o modelo leva em conta a influência espacial ou o impacto que as observações de uma determinada área têm sobre as observações das áreas próximas.

$$y = \rho W y + x\beta + u \quad (17)$$

O resultado pode ser visualizado na Tabela 5, ela acompanha os resultados do MQO como parâmetro de comparação.

Tabela 5 - Comparação dos Resultados do MQO e SAR

	Variável Dependente:	
	log(Taxa Homicídio Suavizada)	
	(MQO)	(SAR)
Constante	1.868*** (0.165)	0.128 (0.113)
% Pop. Homem 15 a 29	1.882* (1.092)	0.846 (0.712)
% Pop. Mulher 15 a 17 Ao Menos um Filho	1.447*** (0.476)	0.601* (0.310)
log(% Pop Extremamente Pobre)	0.101*** (0.023)	0.017 (0.015)
Índice de Gini Normalizado	0.060*** (0.019)	0.004 (0.012)
% Desocupação 18 Anos ou Mais	3.537*** (0.802)	0.931* (0.523)
% População Urbana	0.288*** (0.134)	0.130 (0.088)
% Crianças de 6 a 14 Anos Fora da Escola	4.835*** (1.168)	3.281*** (0.761)
Binaria Região Metropolitana	0.323*** (0.045)	0.079*** (0.030)
Observação	1,631	1,631
R2	0.134	
R2 Ajustado	0.130	
Resíduos Erro Padrão	0.695 (df = 1622)	
Estatística F	31.349*** (df = 8;1622)	
Critério de Informação Akaike (AIC)	3454.9	2,336.787
Nagelkerke Pseudo-R2		0.564
Rho		0.801***
Log-Verossimilhança	-1,717.44 (df=10)	-1,157.394
Sigma2		0.205
Teste de Wald		2,311.347*** (df = 1)
Teste de LR		1,120.093*** (df = 1)
Nota:		* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$
Erro Padrão em parênteses		
Fonte: Fabricação própria com base nos dados do Atlas Brasil, IBGE e Atlas da Violência, 2023		

Essencialmente, o modelo é semelhante ao Método dos Mínimos Quadrados Ordinários (MQO), mas incorpora a defasagem espacial. Aqui, y representa observações da variável dependente em todo o espaço, e ρ indica a intensidade da autocorrelação espacial.

Ao contrário do MQO convencional, o efeito da variável independente na dependente varia entre observações devido ao efeito espacial, influenciando diferentes vizinhanças em cada ponto de dados. (WEISBURD et al., 2022).

Resumindo, no modelo de regressão Spatial Lag (SAR), a variável dependente (como a taxa de homicídios por cem mil habitantes, suavizada em log) é influenciada por sua própria área e pelas variáveis dependentes de regiões vizinhas. Assim, a inclusão da defasagem da variável dependente como independente é uma característica do Spatial Lag (SAR).

4.4.1 Interpretação Estatística

Para interpretação estatística dos resultados, pela recomendação de Anselin. L (2005), é preciso ter cuidado ao analisar R^2 , pois, este não é comparável com o R^2 do MQO, sendo apenas um Pseudo – R^2 , não apropriado para regressão espacial. O Pseudo – R^2 assume observações independentes, não sendo condizente com os modelos espaciais, pois estes são interdependentes devido a autocorrelação.

Por outro lado, as medidas de Log-Verossimilhança e Critério de Informação Akaike (AIC) evidenciam a melhora, quando comparado com o modelo MQO. A Log-Verossimilhança aumenta de -1,717.44 para -1,157.394, enquanto AIC reduz de 3454.9 para 2,336.787.

Em síntese, AIC é uma medida estatística para seleção de modelo, realizando um balanço de ajuste e complexidade, valores menores deste indicador indicam um melhor ajuste. Por outro lado, Log- Verossimilhança é uma medida que representa o quão bem o modelo estatístico explica os dados observados, valores maiores indicam melhor ajuste dos dados.

O valor de $\rho = 0.801$ é significativo a $p < 0.01$, sendo consideravelmente alto. Isto indica que existe dependência espacial forte entre os

preditores significantes. Sendo assim, a mudança em um preditor de um município apresenta forte efeito em seus vizinhos.

Vale ressaltar que algumas variáveis perderam significância na modelagem do MQO, apesar do melhor ajuste do modelo. As variáveis são as seguintes: % Pop. Homem 15 a 29, log(Pop. Extremamente Pobre), Índice de Gini Normalizado, % Pop. Urbana, esta última apresenta significância em < 0.1 na análise dos impactos.

Desta forma, este resultado indica que não é possível afirmar que estas variáveis em questão estejam relacionadas com o aumento na taxa de homicídio por cem mil habitantes suavizada em log para o recorde de dados em analisados.

A matriz de vizinhança, falta de variável independente determinante no modelo, podem ser fatores que colaborem com a perda de significância das variáveis explicitadas. A matriz de contiguidade de ordem um pode não servir como o melhor ajuste para as variáveis de desigualdade e população de homens jovens. Um caminho possível é a utilização da matriz de pesos inversa, uma vez que atribui maior peso aos vizinhos próximos enquanto decai com o aumento da distância.

Neste tocante, Anselin. L (2005) explicita que a queda na magnitude dos coeficientes e a redução da significância observada no modelo MQO indica que, em certa medida, o poder explicativo das variáveis, que foi atribuído ao seu próprio município, estava na verdade relacionado às localizações vizinhas. Tal efeito é capturado pelo coeficiente da variável dependente defasada espacialmente na variável independente, portanto, reduz a magnitude vista previamente no MQO.

Além disso, é possível existir falta de variável determinante, relacionada com fatores de difícil mensuração, o exemplo seria atividades de facções, crime organizado, tráfico de drogas, dentre outros.

Para as variáveis significantes, temos o indicador que os grupos de instabilidade da família, desigualdade pela proxy desocupação, densidade demográfica, menor presença escolar e menor risco pela proxy binária para regiões metropolitanas, possuem impactos positivos sobre a taxa de homicídio suavizada.

O impacto das variáveis não pode ser analisado apenas nos indicadores dos preditores, isto porque, não é possível interpretar os coeficientes como efeitos marginais devido a dependência espacial.

Uma vez que o valor de y_i influenciará o nível de homicídio “em outros” municípios y_j e, por sua vez, estes y_j retroalimemente para y_i , precisamos levar em consideração os efeitos adicionais que o impacto momentâneo de x_i exerce sobre y_i por meio de seu impacto no nível de homicídio em outros municípios (WARD; GLEDITSCH, 2008 apud MEDINA; SOLYMOSI, 2022, tradução livre)

Em resumo, temos que a mudança de i^{th} municípios preditores podem afetar o resultado da região j^{th} . As duas situações são, o impacto direto de uma observação preditora em seu próprio resultado, e o impacto indireto de uma observação do preditor vizinho em seu resultado.

A tabela 6 traz o resultado dos efeitos indiretos, diretos e totais dos preditores na taxa de homicídio por cem mil habitantes suavizada em log.

Tabela 6 - Impactos das Variáveis do Modelo (SAR)

Variável	Direto	Indireto	Total
% Pop. Homem 15 a 29	1.050	3.354	4.404
% Pop. Mulher 15 a 17 Ao Menos um Filho	0.745**	2.380**	3.126**
log(% Pop Extremamente Pobre)	0.021	0.068	0.089
Índice de Gini Normalizado	0.005	0.017	0.023
% Desocupação 18 Anos ou Mais	1.155*	3.688*	4.843*
% População Urbana	0.160*	0.513*	0.674*
% Crianças de 6 a 14 Anos Fora da Escola	4.071***	13.003***	17.074***
Binaria Região Metropolitana	0.098***	0.314***	0.412***
Nota:	* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$		
Fonte:	Fabricação própria com base nos dados do Atlas Brasil, IBGE e Atlas da Violência, 2023		

5 ANÁLISE E DISCUSSÃO DOS RESULTADOS

Apesar das variáveis com não significância estatística, o resultado é o esperado considerando a revisão de literatura. O aumento da desigualdade, medida pelo Indicador de Gini e Percentual dos Extremamente Pobres impactariam positivamente na taxa de homicídio suavizada.

Para o Indicador de Gini FAJNZYLBER; LEDERMAN; LOAYZA, (1998) em modelo GMM de regressão de painel obtém significância e apresenta impacto positivo de 0.012 na log da taxa intencional de homicídio.

Para o mesmo indicador, Scorzafave e Soares (2009) obtém significância e elasticidade positiva no modelo Spatial Lag (SAR), relativa a 2.04, impactando crimes envolvendo ganhos financeiros.

Por sua vez, Ingram e Costa (2016), utiliza como variável dependente a log da taxa de homicídio suavizada em 2011, assim como variáveis independentes de 2010, similar a este estudo. O resultado para variável desigualdade contudo, é ambíguo, com impacto negativo no modelo SLM.

Para Silva et al. (2020), no estudo sobre taxas de homicídio em Pernambuco com dados em painel, o resultado para no modelo MQO sugere impacto positivo de 0.338 e significativo.

Em relação ao percentual dos extremamente pobres, Prates et al. (2023) no modelo MQO obtém impacto negativo enquanto Silva et al. (2020) positivo na taxa de homicídio, ambos com significância.

No quesito da propensão ao homem jovem cometer crimes, Prates et al. (2023) apresenta impacto negativo e significativo para proporção de homens de 15 a 29 anos no MQO e Ingram e Costa (2016) também impacto negativo e significativo no modelo SLM. Ambos resultados são contrários ao observado neste estudo.

Para instabilidade da família, Scorzafave e Soares (2009) apresentam impacto positivo e significativo de mães adolescentes em 1991 no modelo Spatial Lag (SAR).

Em relação ao desemprego, Prates et al. (2023) segue com resultado significativo e positivo para o MQO na taxa de homicídio. Scorzafave e Soares (2009) no SAR e MAJID MADDAH, (2013) no GMM também obtém resultado positivo para desemprego em crimes de roubos e com ganhos financeiros.

No quesito de densidade urbana, o resultado concorda com FAJNZYLBER; LEDERMAN; LOAYZA, (1998), Prates et al. (2023), Ingram e Costa

(2016), Scorzafave e Soares (2009) e Silva et al. (2020), onde está associado positivamente. Importante evidenciar que existe significância à 10% da variável % População Urbana nos impactos. Vale ressaltar, que MAJID MADDAH, (2013) encontra relação negativa no GMM⁶ para densidade populacional em alguns tipos de roubos.

Como complemento da variável binária, Scorzafave e Soares (2009), no modelo Spatial Lag (SAR), apresentam impacto negativo e significativo. Analisando pela ótica do risco, segundo GLAESER; SACERDOTE, (1999), o resultado positivo é mais condizente, indicando que regiões metropolitanas apresentam menor risco de apreensão.

Por fim, a variável relacionada a educação, no estudo de Scorzafave e Soares (2009), apresenta impacto negativo e significativo para taxa de frequência na escola. Prates et al. (2023) com taxa de pessoas sem completar quinta série e com mais de 15 anos encontra impacto positivo e não significativo.

Adicionalmente, FAJNZYLBER; LEDERMAN; LOAYZA, (1998), concluem que a educação, esta negativamente relacionada com taxa de homicídio, similar a Scorzafave e Soares (2009), observando a taxa frequência.

A ótica de interpretação para certos indicadores se altera dependendo da variável. Por exemplo, uma maior taxa de crianças fora da escola impacta positivamente a taxa de homicídio, como explicitado neste estudo. Contudo, uma maior taxa de frequência impacta negativamente.

Em resumo, os resultados observados nos modelos apresentados seguem o pressuposto da literatura, onde, maior desigualdade, maior desemprego, falta de escolaridade, maior instabilidade da família, maior grau de urbanização e regiões de menor risco teriam impacto positivo na taxa de homicídio.

⁶ Gaussian Mixture Model (GMM) é usualmente utilizado para determinar cluster e estimadores de densidade.

6 CONSIDERAÇÕES FINAIS

O presente trabalho teve como objetivo analisar os determinantes do crime para o sudeste brasileiro, utilizando-se do referencial teórico de Becker (1968) e revisão de estudos similares para o processo de escolhas de variáveis sobre os determinantes do crime.

Foi utilizado de modelagem econométrica e aplicação de ferramentas e metodologias de Análise Exploratória de Dados Espaciais (AEDE) para a criação e estimação dos modelos, observando o impacto das variáveis, considerando o componente espacial.

Ficou evidente, que para esta estrutura de análise, considerando dados de corte do ano de 2010 em nível municipal para o sudeste brasileiro em ambas as variáveis dependente e independentes, a adição do componente espacial não suporta totalmente a hipótese explicitada no modelo MQO. Os resultados interpretativos do Spatial Lag seguem no quadro 6.

Quadro 6 - Interpretações de Impacto Modelo Spatial Lag (SAR)

Categoria	Variável	Interpretação
Instabilidade familiar	% Pop. Mulher 15 a 17 Ao Menos um Filho	Aumento de uma unidade na variável independente em questão resulta em média no aumento de 3.125 na taxa de homicídio suavizada.
Desigualdade	% Desocupação 18 Anos ou Mais	Aumento de uma unidade na variável independente em questão resulta no aumento de 4.843 na taxa de homicídio suavizada.
Urbanização	% População Urbana	Aumento de uma unidade na variável independente em questão resulta em média no aumento de 0.674 na taxa de homicídio suavizada.
Educação	% Crianças de 6 a 14 Anos Fora da Escola	Aumento de uma unidade na variável independente em questão resulta em média no aumento de 17.074 na taxa de homicídio suavizada.
Risco	Binaria Região Metropolitana	Aumento em média na taxa de homicídio suavizada de 0.412 resultado da mudança de não região metropolitana para região metropolitana.

Fonte: Fabricação própria, 2023

Foi possível observar que os resultados convergem para a literatura, onde as variáveis-chaves dos grupos de desigualdade, instabilidade familiar, educação, urbanização e risco confluem para os pressupostos estabelecidos.

O modelo indica que a variável mais sensível foi a porcentagem de crianças de 6 a 14 anos fora da escola, indicando que o aumento de uma unidade na variável em questão impactaria em média 17.074 na taxa de homicídio por cem mil habitantes suavizada em log.

A ideia segue para as variáveis de porcentagem de desocupação de pessoas com 18 anos ou mais e porcentagem de população mulher de 15 a 17 anos com ao menos um filho. Logo, o aumento de uma unidade da variável independente em questão impacta em média 4.834 e 3.125 na taxa de homicídio suavizada em log, respectivamente.

Estes indicadores por si só trazem uma ideia conjuntural importante, pois conclui que existe uma alta sensibilidade da população jovem e variáveis socioeconômicas que afetam este grupo.

A educação é mais clara de observar, uma vez que esta diretamente ligada à população jovem, contudo, ela tem impacto direto e indireto no desenvolvimento, pois acarreta o desemprego futuro e serve como minimizador de instabilidades familiares, que nesta análise foi feita por meio de mães jovens.

Uma das problemáticas, já evidenciadas, é o processo de acompanhamento do indivíduo antes de se tornar criminoso, algo que seria inviável em termos de administração e gastos governamentais. Ademais, a educação tende a ter um efeito de longo prazo, portanto, investimentos tomam um maior tempo para se concretizarem, assim como os impactos em outros setores socioeconômicos.

Em consequência, uma das alternativas observadas no âmbito de políticas e segurança públicas é a tentativa da redução da maioria penal. O viés mais rápido de ação corrobora com a ideia de solução mais imediata, sendo assim mais atrativa, mas não necessariamente a mais efetiva.

Neste tocante, Lochner (2004) reconhece que efeitos na escolaridade tendem a ser mais positivos durante o período de *“high school”*, equivalente ao ensino médio, indicando os benefícios sociais da redução do crime por meio do aumento na formação do ensino médio.

Logo, é possível conectar novamente com outras variáveis o impacto educacional, no caso da formação do ensino médio a mais intuitiva é diretamente o desemprego ou desocupação.

Para o caso brasileiro, conclui-se que o desemprego impacta positivamente na criminalidade. O pressuposto converge para as teorias estudadas, onde a baixa remuneração por atividades legais torna as atividades ilegais mais recompensadoras.

Diferentemente da educação, a empregabilidade tende a ser mais sensível a políticas governamentais com retornos imediatos, contudo, cabe a reflexão de onde reside a raiz do problema.

Para instabilidade familiar, como mencionado, o foco deste estudo reside na população mulher de 15 a 17 anos com ao menos um filho, portanto, a relação com a educação é mais direta. Contudo, é possível observar a instabilidade por diversos ângulos.

O principal ponto, é que instabilidade familiar tende a ser uma causa menos objetiva no quesito de ação de políticas otimizadoras, usualmente, como mencionado, ela é impactada por outras variáveis fortemente relacionadas, como educação. Ainda assim, cabe a relevância do suporte direto a famílias instáveis, uma vez que se identifica a importância do jovem na situação da criminalidade.

Neste sentido, a tendência de famílias monoparentais em grande centro urbanos foi mencionada por GLAESER; SACERDOTE, (1999), relacionando com a instabilidade familiar. Atualmente, este conceito pode ser questionável, pois não necessariamente famílias com apenas um responsável são instáveis.

A relação posta traz um indicador de que certas variáveis tendem a funcionar de maneira diferente em centros urbanos, por isto que a variável de urbanização presente neste estudo segue os resultados da literatura. O aumento de uma unidade na variável em questão impacta em média 0.674 na taxa de homicídio por cem mil habitantes suavizada em log.

A reflexão sobre o resultado segue o pressuposto de que grandes centros urbanos possuem uma estrutura mais propícia para a criminalidade. É possível observar estas diferenças em relações a variáveis de impacto, principalmente nos quesitos de desigualdade e risco.

O problema da aglomeração urbana é dificilmente resolvido por meio de políticas governamentais, cabendo um foco maior nas diferenças estruturais que

levam grandes os centros urbanos a se tornarem um ambiente mais atrativo para a criminalidade.

Sendo assim, um dos pontos que estão relacionados com os centros urbanos, fazendo parte da estrutura que os torna mais propícios ao crime é o risco. O aumentar o risco de descoberta, apreensão e convicção tendem a reduzir a criminalidade, pois está diretamente relacionado com o processo de decisão do indivíduo.

A inclusão da variável binária para municípios pertencentes a regiões metropolitanas traz o pressuposto de que grandes centros tendem a ter menor risco, confirmado pelo resultado, indicando que a mudança de não região metropolitana para região metropolitana tende a impactar em média 0.412 na taxa de homicídio por cem mil habitantes suavizada em log.

O risco apresenta uma relevância importante, pois ele sozinho pode não impactar positivamente na redução da criminalidade, podendo gerar o efeito contrário. O exemplo já mencionado, seria aumentar drasticamente a pena de condenação para um crime menor, isto pode levar a maior nível de violência, uma vez que o risco aumentou drasticamente, contudo, a oportunidade em atividades legais ainda é extremamente baixa, mantendo a balança de decisão do indivíduo e aumentando os danos para sociedade.

Os produtos desta monografia convergem para os resultados dos estudos similares, sustentado pelo referencial teórico de Becker (1968). Apesar da desigualdade ser grande foco na análise dos determinantes do crime, conclui-se que a partir do recorde analisado a educação impacta de maneira relevante na taxa de homicídio.

Ademais, a reflexão também consiste nos efeitos de longo prazo da educação, assim como os impactos positivos em outros grupos de variáveis socioeconômicas, como desemprego e estabilidade familiar e desigualdade, nas quais ela está diretamente relacionada.

As conclusões sobre grandes centros urbanos seguem os pressupostos, estando vinculada a problemas estruturais que favorecem a atividade criminosa, dentre elas o risco, identificado como sendo menor em regiões metropolitanas.

Em linhas gerais, a redução do risco é relevante, contudo, deve ser feita de maneira cautelosa, pois aumentar o risco sem necessariamente aumentar a recompensa por atividades legais pode gerar ainda mais danos a sociedade.

Desta forma, políticas que tenham o intuito de reduzir a criminalidade podem tomar como norte os resultados dos estudos da economia do crime pela visão dos principais determinantes e traçar perspectivas que foquem na melhora dos índices socioeconômicos de maneira equilibrada, impactando a criminalidade.

Outro ponto importante é em relação a amplitude do território analisado. Este estudo utilizou todos os municípios do sudeste brasileiro, portanto, regiões com altos níveis de criminalidade podem mascarar o impacto da criminalidade em outras regiões. Sendo assim, recortes menores, como por exemplo estaduais, podem ser mais efetivos em determinar o impacto das variáveis, assim como as regiões mais afetadas.

De modo geral, a visão de Becker (1968) contribui para organizar, quantificar e mensurar o processo de decisão, permitindo traçar estratégias que mitiguem o aumento da criminalidade. Portanto, é plausível identificar variáveis de peso para o processo de decisão do indivíduo em relação a atividades criminosas, trazendo maior efetividade para ações governamentais e políticas de segurança pública.

REFERÊNCIAS

- ANSELIN, Luc. **Exploring Spatial Data with GeoDaTM: A Workbook**. Illinois: Center For Spatially Integrated Social Science, 2005. 224 p.
Disponível em:
https://edisciplinas.usp.br/pluginfile.php/4130006/mod_folder/content/0/geodaworkbook.pdf. Acesso em: 03 jun. 2023.
- ARAUJO, Ari Francisco; FAJNZYLBER, Pablo. **CRIME E ECONOMIA: UM ESTUDO DAS MICRORREGIÕES MINEIRAS**. IX Seminário sobre a Economia Mineira, [s. l.], p. 809-840, 2000.
- BECKER, Gary S. **Crime and Punishment: An Economic Approach**. Journal of Political Economy, Columbia, v. 76, ed. 01, p. 169-217, 1968.
- CERQUEIRA, DANIEL RICARDO DE CASTRO. **Causas e consequências do crime no Brasil**. Orientador: João Manoel Pinho de Mello Rodrigo Reis Soares. 2014. 188 p. Tese (Doutorado) - Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2014.
- COMANOR, W. S.; PHILLIPS, L. **The Impact of Income and Family Structure on Delinquency**. Journal of Applied Economics, v. 5, n. 2, p. 209–232, nov. 2002.
- EHRlich, Isaac. **Participation in Illegitimate Activities: A Theoretical and Empirical Investigation**. Journal of Political Economy, [s. l.], v. 83, ed. 3, p. 521-565, 1973.
- FAJNZYLBER, P.; LEDERMAN, D.; LOAYZA, N. **Determinants of crime rates in Latin America and the world**. [s.l.] The World Bank, 1998.
- FILHO, Claudio Chaves Beato; REIS, Ilka Afonso. **Desigualdade, Desenvolvimento Socioeconômico e Crime**. Atlas da Violência - IPEA, [s. l.], p. 1-14, 1999.
- FRANCISCO; FAJNZYLBER, P. **O que causa a criminalidade violenta no Brasil? Uma análise a partir do modelo econômico do crime: 1981 a 1996**. 1 jan. 2001.
- GLAESER, EDWARD L.; SACERDOTE, B. **Why is There More Crime in Cities?** Journal of Political Economy, v. 107, n. S6, p. S225–S258, dez. 1999.
- HEINEKE, John; BLOCK, Michael Kent. **The supply of legal and illegal activity: A choice theoretic analysis**. Santa Clara Business Review, [s. l.], p. 1-16, 1973.
In-Text Citation: (Ye & Wu, 2011)
- JEFFERY, Clarence Ray. **The Historical Development of Criminology**. Journal of Criminal Law and Criminology, [s. l.], v. 50, p. 3-19, [1960].
- LOBOS, J. **Factors Affecting Juvenile Delinquency**. Faculty Research Journal Bulacan State University - Meneses Campus, v. 7, p. 1–6, 2018.
- LOCHNER, Lance; MORETTI, Enrico. **The effect of education on crime:**

evidence from prison inmates, arrests, and self-reports. The American Economic Review, Nashville, v. 94, n. 1, 2004, p. 155-189.

MAJID MADDAH. **AN EMPIRICAL ANALYSIS OF THE RELATIONSHIP BETWEEN UNEMPLOYMENT AND THEFT CRIMES.** 1 jan. 2013.

MEDINA, Juanjo; SOLYMOSI, Reka. **Crime Mapping and Spatial Data Analysis using R.** [S. L.]: Github, 2022. Disponível em: https://maczokni.github.io/crime_mapping/. Acesso em: 06 jun. 2023.

NAKABASHI, Luciano et al. **Boletim de Segurança Pública - Ribeirão Preto.** Centrode Pesquisa em Economia Regional da FUNDACE (CEPER), [s. l.], v. 24, p. 1-10, 2020.

PRATES, Thierry Molnar et al. **HOMICIDES IN ALAGOAS – BRAZIL: SPATIAL ANALYSIS.** Estudios Económicos, Bahía Blanca, v., n. 80, p. 11-41, jan. 2023. Disponível em: <https://www.redalyc.org/journal/5723/572373856001/html/>. Acesso em: 14 set. 2023.

RESENDE, João Paulo; ANDRADE, Mônica Viegas. **Crime Social, Castigo Social: Desigualdade de Renda e Taxas de Criminalidade nos Grandes Municípios Brasileiros.** Estudos Econômicos, São Paulo, v. 41, ed. 1, p. 174-195, 2011.

SALVIATO, Rafael Buttini; MOURÃO, Gustavo Nunes. **ECONOMIA DO CRIME: UM ESTUDO SOBRE A TEORIA ECONÔMICA DO CRIME E AS SUAS VERIFICAÇÕES NO ESTADO DO PARANÁ.** Memorial TCC – Caderno da Graduação, [s. l.], p. 353-368, 2015.

SCORZAFAVE, Luiz Guilherme; SOARES, Milena Karla. **Income inequality and pecuniary crimes.** Economics Letters, [s. l.], v. 104, p. 40-42, 2009.

SHIKIDA, Pery Francisco Assis. **ECONOMIA DO CRIME: TEORIA E EVIDÊNCIAS EMPÍRICAS A PARTIR DE UM ESTUDO DE CASO NA PENITENCIÁRIA ESTADUAL DE PIRAQUARA (PR).** Revista de Economia e Administração, [s. l.], p. 1-20, 2005.

SILVA, Carlos et al. **Spatial Modeling for Homicide Rates Estimation in Pernambuco State-Brazil.** International Journal of Geo-Information. Switzerland, p. 1-19. dez. 2020.

SILVA, Michele Lins Aracaty e; RIKER, Daniel Garcia Jaña (org.). **ECONOMIA DO CRIME: UMA PERSPECTIVA ECONÔMICA DA TEORIA DE GARY BECKER COM FOCO NO CRIMINOSO RACIONAL.** In: COSTA, Alvaro Daniel (org.). Cultura, Cidadania e Políticas Públicas. 4. ed. [S. l.]: Atena, 2019. p. 1-278.

VIAPIANA, Luiz Tadeu. **Economia do crime: uma explicação para a formação do criminoso.** Editora AGE Ltda. 2006

Ward, M. D., & Gleditsch, K. S. (2008). **Spatial Regression Models.** SAGE Publications.

WEISBURD, David et al. **Advanced Statistics in Criminology and Criminal Justice**. 5. ed. Switzerland: Springer, 2022. 550 p.

Ye, X., & Wu, L. (2011, April 1). **Analyzing the dynamics of homicide patterns in Chicago: ESDA and spatial panel approaches**. Applied Geography, Elsevier BV. <https://doi.org/10.1016/j.apgeog.2010.08.006>