**PONTIFÍCIA UNIVERSIDADE CATÓLICA DE CAMPINAS**

**CENTRO DE ECONOMIA E ADMINISTRAÇÃO FACULDADE DE CIÊNCIAS ECONÔMICAS**

**JOÃO PEDRO TOLEDO TRICOLI DE LUCAS**

**ECONOMETRIA ESPACIAL PARA ANÁLISE DO CRIME NO SUDESDE BRASILEIRO**

**CAMPINAS**

**2023**

**JOÃO PEDRO TOLEDO TRICOLI DE LUCAS**

**ECONOMETRIA ESPACIAL PARA ANÁLISE DO CRIME NO ESTADO DE SÃO PAULO**

Projeto de Pesquisa apresentado como requisito parcial para aprovação na disciplina de Monografia I.

Orientadora: Prof.ª. Dra. Paulo R S Oliveira

**PUC – CAMPINAS**

**2023**

**RESUMO**

LUCAS, João Pedro Toledo Tricoli. TÍTULO. 2023. 99 f. Trabalho de Conclusão de Curso. Faculdade de Ciências Econômicas, Centro de Economia e Administração, Pontifícia Universidade Católica de Campinas, Campinas, 2023.

O Resumo apresenta em **um parágrafo** uma visão geral sobre o conteúdo da monografia, entre 15 e 20 linhas, para informar ao leitor o seguinte: I) os objetivos = tema analisado na monografia; II) a metodologia = fundamento teórico, métodos e técnicas de abordagem e de análise; III) os resultados = destaca ideias principais, fatos novos, descobertas significativas, relações e implicações verificadas; e IV) as conclusões = consequências dos resultados e suas relações com os objetivos.

Os itens III e IV só devem ser confeccionados quando a monografia estiver pronta.

(Obs: A redação deve ser na forma dissertativa (frases concisas, objetivas e diretas em parágrafos com introdução, desenvolvimento e conclusão), devendo-se evitar enumeração de tópicos, fórmulas, equações, tabelas, quadros e gráficos.)

**Palavras-chave:** Crime, Econometria Espacial, Analise Espacial, Crime São Paulo, Economia do Crime.

**LISTA DE GRÁFICOS**

[**Gráfico 1 -** A 15](#_Toc55247348)

**LISTA DE QUADROS**

[**Quadro 1 -** A 13](#_Toc55247196)

**LISTA DE TABELAS**

[**Tabela 1 - Matriz com Pesos Binários** 20](#_Toc137714358)

[**Tabela 2 - Matriz de Vizinhança com Pesos Padronizados** 21](#_Toc137714359)

**LISTA DE FIGURAS**

[**Figura 1 - Conceitos das Matrizes de Vizinhança** 18](#_Toc137714435)

[**Figura 2 - Matriz de Vizinhança de Segunda Ordem** 19](#_Toc137714436)

[**Figura 3 - Comparação de Modelos de Estimação** 26](#_Toc137714437)

**LISTA DE SIGLAS**

IBGE – Instituto Brasileiro de Geografia e Estatística

IPEA – Instituto de Pesquisa Econômica e Aplicada

**SUMÁRIO**

[1. INTRODUÇÃO 8](#_Toc150266711)

[1.1 TEÓRIA ECONÔMICA SOBRE A CRIMINALIDADE 10](#_Toc150266712)

[1.1.1 DANOS 11](#_Toc150266713)

[1.1.2 O CUSTO DE APREENSÃO E CONVICÇÃO 12](#_Toc150266714)

[1.1.3 OFENÇAS 13](#_Toc150266715)

[1.1.4 PUNIÇÃO 14](#_Toc150266716)

[1.3 REVISAO DA LITERATURA 16](#_Toc150266717)

[1.3.1 DETERMINANTES DO CRIME 16](#_Toc150266718)

[2. METODOLOGIA 22](#_Toc150266719)

[2.1 REGRESSÃO ESPACIAL 23](#_Toc150266720)

[2.1.1 AUTOCORRELACAO ESPACIAL 24](#_Toc150266721)

[2.1.2 MATRIZ DE VIZINHANCA 25](#_Toc150266722)

[2.1.3 MATRIZ DE VIZINHANCA COM PESOS 27](#_Toc150266723)

[2.1.4 ÍNDICE DE MORAN 29](#_Toc150266724)

[2.1.5 MÉTODO DOS MINÍMOS QUADRADOS ORDINÁRIOS (MQO) 30](#_Toc150266725)

[2.1.6 MODELO SPATIAL LAG 32](#_Toc150266726)

[2.1.7 MODELO SPATIAL ERROR 34](#_Toc150266727)

[2.1.8 ESCOLHA DO MODELO DE REGRESSÃO ESPACIAL 35](#_Toc150266728)

[2.2.1 CRIAÇÃO DO MODELO MQO PARA ESTIMAÇÃO 36](#_Toc150266729)

[3 REFERÊNCIAS BIBLIOGRÁFICAS 39](#_Toc150266730)

# 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 arbitrarias do estado e a escola positivista, com o intuito de aplicar métodos científicos ao estudo do crime. (JEFFERY, [1959-1960])

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 ([1959-1960]) como a tentativa de responder por meio de estudos científicos a criminalidade.

The main characteristic of positivism is its attempt to answer the riddle of criminality by means of scientific studies of the individual offender. The use of scientific method is one of the major characteristics of positivism. (JEFFERY, [1959-1960], 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, [1959-1960])

Neste sentindo 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 individuo como racional, custo benéfico e teorização sistemática de problemas econômicos.

Desta forma, a base cientifica do positivismo promoveu um ambiente propicio há 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 entorno de análises econométrica espacial 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 ([1959-1960]) 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 cientificas 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, 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, diversas contribuições foram feitas, 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, dentre outros, 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 tem como objetivo geral 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 socioeconômicos com a criminalidade na região Sudeste do Brasil.

Os objetivos específicos são: a) Analisar os pilares dos determinantes do crime, sendo estes, danos, custo de apreensão e convicção, ofensas e punição. Seguindo o referencial teórico de Becker (1968); b) Revisão da literatura, pautada na conclusão de resultados obtidos em estudos similares, com diferentes níveis territoriais, e a decisão das escolhas de variáveis; d) Especificar os princípios da econometria espacial e Análise Exploratória de Dados Espaciais (AEDE), incluindo os principais modelos: OLS, Spatial Lag, Spatial Erro além dos critérios de escolha por Anselin (ref) e) Construção do modelo e processo de escolha das variáveis e) Interpretação do modelo, resultados encontrados e comparação com estudos similares.

A metodologia utilizada nessa pesquisa, considerando objetivos 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 as variáveis socioeconômicas o ATLAS Brasil, que sumariza dados por municípios, pautado no Instituto Brasileiro de Geografia e Estatística (IBGE) será a principal fonte utilizada.

A Monografia será dívida em três capítulos. O primeiro apresentara o referencial teórico da economia do crime, seguido da revisão de literatura sobre os principais estudos relacionados à teoria, focando nos critérios das principais variáveis socioeconômicas escolhidas e resultados obtidos.

O segundo capítulo irá trazer a metodologia, conceitos da econometria espacial e método de estimação do modelo, assim como o processo de escolha das variáveis que proporcione o melhor ajuste.

O terceiro capítulo trará a interpretação dos resultados e uma comparação com as conclusões de estudos similares, explicitadas na revisão da literatura.

## 1.1 TEÓRIA ECONÔMICA SOBRE A CRIMINALIDADE

O referencial teórico da presente pesquisa procurou compreender a economia do crime, baseado no estudo, Crime and Punishment: An Economic Approach de 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 um crime.

Becker (1968) em seu estudo sobre a economia do crime, desenvolve com bases econômicas, uma análise de políticas públicas e privadas com o objetivo de 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 a fórmula abaixo. Viapiana (2006) explicita o modelo de Becker, com o intuito de entender a determinante da atividade criminosa.

|  |  |  |
| --- | --- | --- |
|  |  | (1) |

Onde:

: benefício gerado ao criminoso por praticar o delito;

: custos da atividade criminosa;

: probabilidade de apreensão

Desta forma o indivíduo é incentivado a escolher a atividade criminosa quando o benefício () excede o custo da atividade criminosa () multiplicado pela chance de ser apreendido ().

Neste sentido a teoria é dividia em quatro (4) principais áreas de atuação, danos, custo de apreensão e convicção, o fornecedor de 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 pautar a teoria.

### 1.1.1 DANOS

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

|  |  |  |
| --- | --- | --- |
|  |  | (2) |

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

|  |  |  |
| --- | --- | --- |
|  |  | (3) |

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

|  |  |  |
| --- | --- | --- |
|  |  | (4) |

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

### 1.1.2 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 que são alocados em esforços referente ao combate à criminalidade.

Uma aproximação para medir o nível de atividade criminal seria o número de ofensas inocentadas por condenação.

|  |  |  |
| --- | --- | --- |
|  |  | (4) |

Onde:

: É a relação entre os delitos apurados pelas condenações e todos os delitos.

Em outras palavras 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 esta relação.

|  |  |  |
| --- | --- | --- |
|  |  | (4) |

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.

### 1.1.3 OFENÇAS

A literatura traz diversos pontos na consideração dos determinantes da atividade criminosa, sendo estes, especificados na revisão da literatura. Contudo, a maioria concorda no ponto, que com tudo ou mais constante, temos uma tendencia a redução da atividade criminosa quando se aumenta a probabilidade de apreensão. Becker evidencia o método principal da escolha racional do indivíduo baseado na expectativa de utilidade.

The approach taken here follows the economists' usual analysis of choice and assumes that a person commits an offense if the expected utility to him exceeds the utility he could get by using his time and other resources at other activities. Some persons become" criminals," therefore, not because their basic motivation differs from that of other persons, but because their benefits and costs differ. (BECKER, 1968, v. 76, p. 176)

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 para ele em atividades legais e outras atividades ilegais, a frequência de prisões e vontade de cometer um ato ilegal. (BECKER, 1968)

|  |  |  |
| --- | --- | --- |
|  |  | (4) |

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

O aumento de ou 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)

### 1.1.4 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, dentre elas se encontram, redução da maioridade penal, maior rigor nas condenações, dentre outros.

De modo geral a punição deve se encaixar com o crime, logo, para infrações mais simples temos multas, que são uma punição financeira para quem realiza a ofensa, por outro lado em casos mais sérios temos 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 temos 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.

|  |  |  |
| --- | --- | --- |
|  |  | (4) |

Onde é o custo social e é o coeficiente que transforma em . O tamanho de difere para cada tipo de punição. Para temos crimes menores, com multas, enquanto seria prisão, e outros tipos mais sérios de punição.

Quando falamos sobre as punições, como método de redução na atividade criminosa, é importante manter um certo nível de interpretação, pois caímos em 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 consequentemente a maior violência em crimes simples.

Em suma, Becker (1968), concentra a ideia de determinação de um modelo que permita a melhor alocação de recursos para políticas públicas, com o objetivo de 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) e o ofensor seja preso e condenado.

O peso da punição pode ser visto como juntamente com o tipo, 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 acerca do que é considerado atividade ilegal.

To some, any wage rates set by competitive labor markets are permissible, while to others, rates below a certain minimum are violations of basic rights; to some, gambling, prostitusion, and even abortion should be freely available to anyone willing to pay the market price, while to others, gambling is sinful and abortion is murder. These differences are basic to the development and implementation of public policy but have been excluded from my inquiry. (BECKER, 1968, v. 76, p. 209)

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.

## 1.3 REVISAO DA LITERATURA

Esta monografia utilizou como revisão literária, estudos, em diversos níveis territoriais, nacionais e internacionais, que utilizam da teoria do crime como base para escolha de variáveis na construção do modelo econométrico para análise espacial. Os trabalhos buscam, de modo geral, responder à questão principal, *“Quais as variáveis socioeconômicas que impactam as atividades criminais?”.*

As escolhas das variáveis para análise no modelo de regressão levam em conta as preposições da teoria de Becker (1968). Um exemplo seria a relação da educação, que poderia influenciar positivamente na redução de atividade criminosa, ou um maior aumento na renda das atividades legais, reduzindo a recompensa na atividade criminosa, dentre outros fatores.

Com este conceito, é analisado em um modelo econométrico um conjunto de variáveis socioeconômicas, que segundo a teoria, impactariam sobre determinado tipo de crime. Neste sentido, também é incluindo o conceito do espaço e vizinhança, que teria uma relevância no estudo do crime.

A revisão terá como objetivo explorar as decisões nas escolhas de variáveis para estimação do modelo econométrico, explicitando as diferenças e objetivos nos estudos e por fim concluindo seus principais resultados.

Desta forma será possível ter uma base que suporte as escolhas de variáveis para o modelo econométrico deste trabalho, assim como uma comparação final para os resultados encontrados.

### 1.3.1 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.

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 significante, FAJNZYLBER; LEDERMAN; LOAYZA, (1998) indica que maior desigualdade está associada com maior homicídio intencional, contudo, o nível de renda percapita não é significante.

O problema em relação a análise única da renda ou renda percapita 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 cometer algum crime. 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 percapita, a recompensa pela atividade criminosa também aumenta. Logo, em regiões com maior nível de renda, teria 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. Sendo detectado a relação positiva da desigualdade com crimes financeiros, com uma elasticidade de 1,46, segundo os dados do censo demográfico do ano 2000.

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, 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. Ehrich (1968), em seu estudo participation in illegitimate activities, 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, MADDAH, M (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 estudos, contudo, no caso brasileiro se mostrou significante 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?” GLAESER; SACERDOTE, (1999), traz 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 são mais efetivas em cidades menores. (GLAESER; SACERDOTE, 1999)

Também é enfatizado um dos grandes problemas dos dados criminais, sendo este a subnotificação, segundo GLAESER; SACERDOTE, (1999), indivíduos em grandes cidades estão menos inclinados e 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, MADDAH, M (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 binaria para regiões metropolitanas.

Neste sentido, as referências tendem a ir de encontro, localidades mais densas demograficamente, tendem a estar relacionadas de maneira positiva com maiores taxas de crimes, incluindo financeiros (Scorzafave e Soares, 2009) e homicídios (MADDAH, M 2013) e (FAJNZYLBER; LEDERMAN; LOAYZA, 1998).

Até o momento foi revisado variáveis de teor mais específico e objetivo, renda, desigualdade, desemprego e densidade demográfica. Contudo, outros fatores podem afetar a decisão do indivíduo no momento da escolha em realizar uma atividade ilícita.

Com isso, 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. Neste tópico é evidenciado outra relação, que em grandes cidades existe a tendencia de maior concentração de famílias monoparentais.

O estudo “The impacto f income and Family Structure on Delinquency”, 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 primeiro a renda familiar.

De acordo com a análise empírica, os incentivos financeiros governamentais à 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”, de LOBOS, J. (2018), que entrevista mais de 105 pessoas, de idades de 15 a 31 anos, concluiu que quase metade (44,8%) viviam em familiar nucleares, mais de um quarto (26,7%) pertenciam a famílias desestruturadas.

Os resultados de LOBOS, J. (2018) relacionam também 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.

Por outro lado, 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.

O conceito de instabilidade familiar é amplo, 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.

O conceito de instabilidade familiar, como visto, enfatiza um efeito mais forte em uma faixa etária mais jovem, 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.

O estudo avalia que uma das principais razões para a maior predisposição dos jovens é a pressão dos pares, 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 influência.

A importância da faixa etária também é vista em FRANCISCO; FAJNZYLBER, (2001), onde ele indica que na faixa etária mais jovem as taxas de homicídios ficam bem maiores do que as de faixa 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, e mais intuitiva, 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, temos 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 em seu estudo relataram que a escola teve menos influencia 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.

Por fim, temos a variável relacionada ao risco para cometer a atividade ilegal, neste caso temos algumas opções, dependendo muito da disponibilidade de dados disponível para 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. Podemos 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 de proposito em reduzir o crime, como investimentos em treinamento policial, equipamentos como 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.

Com isso, temos a definição das principais justificativas e comparação dos resultados obtidos em diversos estudos similares, incluindo a escolha de variáveis e suas especificidades.

# **METODOLOGIA**

O presente trabalho teve como objetivo aplicar, sobre a econometria espacial e Análise Exploratória de Dados Espaciais, os pressupostos apresentados na teoria de Becker (1968). Foi levado em consideração as contribuições de estudos similares, que serviram como base para a escolha de variáveis para a construção do modelo.

A metodologia foi distribuída em dois tópicos centrais. O primeiro, consistem em uma análise fundamental dos conceitos que orbitam as regressões espaciais.

Para a Econometria Espacial e Análise Exploratória de Dados Espaciais (AEDE), as referências bases foram os livros Advanced Statistics in Criminology and Criminal Justice de Weisburd et al. (2022) e Exploring spatial data with GeoDaTM Anselin, L. (2005).

Neste tópico inclui definições sobre Autocorrelação Espacial, Matrizes de Vizinhança, Indicador de Moran I e os Modelos de Regressão, sendo eles, Método dos Mínimos Quadrados Ordinários (MQO), Modelo Spatial Lag e Modelo Spatial Error.

O segundo tópico consiste no processo de construção do modelo, incluindo a escolha das variáveis, assim como a expectativa dos resultados, pautadas na revisão da literatura.

Por fim, será realizado o processo metodológico para a escolha do melhor ajuste de variáveis, juntamente com o a suavização para as taxas de homicídio.

Desta forma, podendo concluir com a especificação do processo de escolha do melhor modelo de regressão (MQO, Spatial Lag ou Spatial Error), pautado em Anselin, L (2005).

Todo o processo realizado na construção metodológica e manipulação de dados foi feito no Software de Estatística e Programação R.

Os dados socioeconômicos utilizados nas variáveis independentes foram retirados do Atlas Brasil e Instituto Brasileiro de Geografia e Estatística (IBGE), em nível territorial municipal para os estados pertencentes ao Sudeste Brasileiro, enquanto para a variável dependente taxa de homicídios, o Atlas da Violência do IPEA foi utilizado.

O método de análise foi de corte, utilizando dados para todas as variáveis no ano de 2010.

## 2.1 REGRESSÃO ESPACIAL

Os modelos de regressões tradicionais, analisam dados de séries 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 para um número de pessoas na região , assumindo que a escolha de método e densidade do bairro na região de não teria uma influência no tempo de transporte da região .

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, de certa, forma 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.

### 2.1.1 AUTOCORRELACAO 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 componente do espaço apresenta grande peso.

Neste estudo, o homicídio, possui por origem um componente espacial forte, relacionados com si próprio, em outras palavras, regiões mais violentas tendem a ter em sua proximidade, outras regiões violentas.

Também temos o relacionamento com outras variáveis, como visto no estudo de Becker (1968), como que 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.

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

As dependências espaciais estão relacionadas as interações no espaço, causando agrupamentos de eventos ao longo do espaço, 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 próximas maiores níveis de violência. A intensidade de violência em uma região afeta a intensidade da violência nas localidades próximas. Neste sentido, 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 eventos ao longo do espaço. (WEISBURD et al., 2022).

Um exemplo seria o nível de violência aumentar durante grandes eventos, como nas Olímpiadas do Rio de Janeiro.

Estes pontos são importantes de ter em mente, 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, uma vez que a intenção do estudo, com o componente espacial, é detectar o grau de correlação entre as variáveis considerando o componente de vizinhança, permitindo assim, continuação nas especificações dos modelos.

### 2.1.2 MATRIZ DE VIZINHANCA

A Matriz de Vizinhança, é necessária para detectar e corrigir Correlação Espacial. Ela é um modo numérico de representar relações de vizinhança entre pontos, par de coordenadas e , representando latitude e longitude ou polígonos, representando a combinação de 3 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.

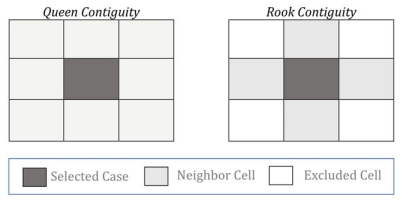
A matriz de contiguidade espacial com pesos se preocupa com o conceito dos adjacentes. Utilizado comumente para polígonos, a contiguidade verifica se um polígono está tocando no outro, assim como o modo em que ele está tocando. (WEISBURD et al., 2022).

Os dois principais métodos de contiguidade são os 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.

A imagem abaixo traz exemplificação de ambas as contiguidades.

**Figura 1 - Conceitos das Matrizes de Vizinhança**

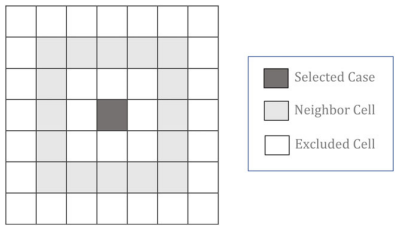


Fonte: (WEISBURD et al.). Advanced Statistics in Criminology and Criminal Justice. 5. ed. Switzerland: Springer, 2022. p. 552, fig. ()

Ambos os métodos permitem adaptação no nível de células ou polígonos, que estão sendo considerados como vizinhos, neste caso da Figura 1, temos ordem um 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. A Figura 2 traz o exemplo do funcionamento de ordens na matriz de vizinhança.

**Figura 2 - Matriz de Vizinhança de Segunda Ordem**



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

Outra opção, como mencionado, é a matriz de vizinhança com peso baseada na distância. Normalmente calculada para a escolha de polígonos 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 centroides dentro deste limite vão ser considerados vizinhos.

Por outro lado, existe a opção de utilizar a matriz por número de vizinhos, chamada de k-nearest neighbors. Onde indica o número de vizinhos que é incuido na análise. Este método inclui o vizinho mais próximo primeiro e segue para o próximo, até chegar no limite determinado. (WEISBURD et al., 2022).

O método k-nearest neighbors não exige contiguidade e a distância entre polígonos também não é um problema. Ele é útil para resolver problemas 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.

### 2.1.3 MATRIZ DE VIZINHANCA COM PESOS

Até o momento consideramos 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 considerar que vizinhos mais próximos teriam um peso maior na nossa análise.

Desta forma, nesta sessão, 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 de unidades espaciais, podendo ser polígonos, são consideradas como vizinhoss, temos , se não . Isto se aproxima da matriz de correlação, onde a diagonal reflete uma unidade espacial com si mesmo.

De modo geral temos duas opções, a matriz com pesos binários e a com pesos padronizados em linha.

Na primeira, consideramos como sendo vizinho ou não sendo vizinho, a soma da linha da matriz pode ter valores diversos, dependendo do número de vizinhos.

Por outro lado, quando temos linhas de peso padronizadas o resultado da soma das linhas da matriz resulta no máximo em 1. A imagem abaixo traz a exemplificação da matriz com pesos binarios.

Tabela 1 - Matriz com Pesos Binários

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Região | | | | | | |
| Região |  | 1 | 2 | 3 | 4 | Sum |
| 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. 552, fig. ()

Neste caso, estamos 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. Podemos 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 a 1.

Tabela 2 - Matriz de Vizinhança com Pesos Padronizados

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Região | | | | | | |
| Região |  | 1 | 2 | 3 | 4 | Sum |
| 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. 552, fig. ()

Podemos observar, que apesar da padronização, ainda temos que os pesos assinados a cada vizinhança são os mesmos, sendo a divisão do total pelo número de vizinhos, portanto, proporcional.

Para conseguirmos atribuir peso maior para vizinhanças mais próximas e decair com o nível de distância a utilização da matriz de vizinhança com pesos inversa é necessária.

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 tradados e como eles vão se relacionar para as variáveis escolhidas, impactando o resultado obtido e as interpretações.

### 2.1.4 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 explicitaria 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.

|  |  |  |
| --- | --- | --- |
|  |  | (4) |

Onde:

: Número de unidades espaciais;

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

: Valor de outra variável em uma outra localidade ;

: A média da variável;

: É a soma de todos os pesos espaciais

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

|  |  |  |
| --- | --- | --- |
|  |  | (4) |

O indicador de Moran I é interpretado de maneira similar ao coeficiente de correlação. Caso computemos o indicador para taxa de homicídio, o indicador irá retornar um valor que indica o quão similar ou não similar 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 tipos, com o objetivo de identificar agrupamentos baseado nas Autocorrelação Espacial. A tabela X resumi as categorias.

|  |  |
| --- | --- |
| 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. |

### 2.1.5 MÉTODO DOS MINÍMOS QUADRADOS ORDINÁRIOS (MQO)

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

|  |  |  |
| --- | --- | --- |
|  |  | (4) |

Sendo a variável dependente e a independente, é coeficiente de regressão, sendo a inclinação a magnitude da relação estudada, por fim temos , que representa o erro aleatório do modelo.

O MQO traça a melhor regressão ajustada, minimizando a Soma dos Erros Quadrados. 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 , temos primeiro a necessidade da transformação dos Erros Quadrados e depois a busca pelo menor erro possível no somatório.

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.

O fator de proximidade geográfica também seria um significante, uma vez que países da união europeia teriam legislações similares, dado a características históricas de desenvolvimento.

O fator principal, é que a heterocedasticidade 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 viés devido a violação do princípio da homoscedasticidade dos erros. 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).

### 2.1.6 MODELO SPATIAL LAG

Segundo Weisburd et al. (2022) o modelo de Spatial Lag 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 espacial incorporando a defasagem da variável dependente.

|  |  |  |
| --- | --- | --- |
|  |  | (4) |

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

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

Rho indica a forca da autocorrelação espacial presente na variável dependente.

Temos como teste de Hipótese:

|  |  |  |
| --- | --- | --- |
|  | 𝐻𝐴: 𝐴𝑢𝑡𝑜𝑐𝑜𝑟𝑟𝑒𝑙𝑎𝑐𝑎𝑜 𝑒𝑠𝑝𝑎𝑐𝑖𝑎𝑙 | (4) |

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

Neste caso, a estimativa de depende de e , sendo assim, a interpretação de não é a mesma que no MQO, pois consideramos o efeito espacial para a variável 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 3 traz uma sintetização da relação entre as variáveis e o que o modelo de Spatial Lag (SAR) consegue analisar, comparando com o já discutido MQO e o Spatial Error, que será analisado em seguida.

Diagrama, Gráfico de caixa estreita

Descrição gerada automaticamente**Figura 3 - Comparação de Modelos de Estimação**

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

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

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

### 2.1.7 MODELO SPATIAL ERROR

O modelo Spatial Error 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).

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 , juntamente com o termo de erro usual. (WEISBURD et al., 2022).

|  |  |  |
| --- | --- | --- |
|  |  | (4) |

Onde:

|  |  |  |
| --- | --- | --- |
|  |  | (4) |

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

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

As especificações dos modelos de regressão espaciais são importantes, isto porque elas fazem do processo de seleção do melhor modelo para determinada estrutura de dados e objeto de análise.

O próximo tópico mostra a metodologia de seleção de modelo de regressão espacial, seguido da escolha das variáveis com o melhor ajuste.

### 2.1.8 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.

Neste sentido Anselin (2005) especifica um processo de decisão que seleciona o modelo de regressão mais apropriado para a relação estudada. A imagem X traz os passos realizados para a escolha final do modelo.

[imagem]

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

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

Anselin (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, temos qual deles será 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, ele serve como o último divisor na decisão de escolha do modelo.

De modo geral, 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 imagem X.

### 2.2.1 CRIAÇÃO DO MODELO MQO PARA ESTIMAÇÃO

A variável dependente do modelo utilizada neste estudo foi a taxa de homicídio por 100 mil habitantes para todos os municípios do sudeste brasileiro.

Esta taxa passou pelo processo bayesiano 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.

“plotting observed rates can have serious drawbacks when sample sizes vary by area, since very high (and low) observed rates are found disproportionately in poorly-sampled areas” (Gelman and Price 1999, 3221).

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

Por fim, é criado uma taxa local para o estimado suavizado, o objetivo é não apenas reduzir o risco global criado para cada município, podendo mascarar heterogeneidades não observadas.

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.

O mapa X traz a comparação da taxa de homicídio por 100 mil habitantes e a taxa suavizada.

[mapa]

Em relação as variáveis dependentes, uma divisão de categorias foi feita previamente, considerando o referencial teórico de Becker (1082) e a revisão da literatura, podendo ser visualizado na tabela X.

|  |  |
| --- | --- |
| Categoria | Variáveis |
| Desigualdade | Indicador de Gini  Indicador de Theill  % de População Extremamente Pobre  % de População Pobre  Indice de Desenvolvimento Humano (IDHM)  Indice de Desenvolvimento Humano da Renda (IDHM Renda) |
| Predisposição do Homem Jovem | % População Homem de 15 a 29 Anos |
| Desemprego | % da Desocupação de 18 Anos ou mais |
| 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 |

Todas as variáveis estão em nível municipal e são do ano de 2010, assim como a taxa de homicídio suavizada. Portanto a construção inicial do modelo MQO foi pautada em dados de corte.

Para escolha do melhor ajuste um modelo de regressão stepwise foi feito. 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.

O principal ponto para realizar este processo foi considerando as diversas variáveis escolhidas em diversos estudos, e como estas, podem ter efeitos distintos pautados em características regionais.

Desta forma, considerando um agrupamento de variáveis gerais, vistas na tabela X, assim como algumas transformações delas, pautadas na revisão de literatura, o processo de seleção stepwise traz um modelo com um melhor ajuste com os dados em questão.

O modelo final escolhido pode ser visto na tabela X, diferente do resultado do processo de seleção, não foi considerado as variáveis IDHM e IDHM Renda, removidas posteriormente devido ao problema de multicolineariedade.

|  |  |
| --- | --- |
| Variáveis | Expectativa de Impacto |
| log(Taxa Homicídio Suavizada) |  |
| ln(% População Homem de 15 a 29 Anos) | + |
| ln(% de Mulheres de 15 a 17 Anos com ao menos um filho) | + |
| log(% de População Extremamente Pobre) | + |
| ln(Indicador de Gini Normalizado) | + |
| ln(% Desocupação de 18 Anos ou mais) | + |
| ln(% População Urbana) | + |
| ln(% de Crianças de 6 a 14 Anos Fora da Escola) | + |
| Binária para Região Metropolitana | + |

A utilização do log para porcentagem da população extremamente pobre foi pautada em Ye, X., & Wu, L. (2011, April 1). A seleção da log da taxa de homicídio suavizada foi feita pelo método de seleção, provendo um melhor ajuste ao modelo de modo geral.

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

Por fim, temos a expectativa dos resultados, levando em consideração qual seria o impacto das variáveis independentes na taxa de homicídio suavizada. Neste caso, o impacto esperado é positivo para todas. O resultado do modelo MQO pode ser visto na tabela X.

[MQO]

De modo geral, o modelo contempla todas as categorias evidenciadas, com a inclusão de ao menos uma variável para cada categoria além de que os impactos seguem o esperado.

O modelo apresenta significância estatística para todas as variáveis, assim como um R2 de XX, sem apresentar multicolineariedade, com todas as variáveis com VIF < 2.X.

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

3 REFERÊNCIAS BIBLIOGRÁFICAS

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: ATheoretical 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 Sócio-econômico 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.

JEFFERY, Clarence Ray. The Historical Development of Criminology. Journal of Criminal Law and Criminology, [s. l.], v. 50, p. 3-19, [1959-1960].

LOBOS, J. Factors Affecting Juvenile Delinquency. Faculty Research JournalBulacan State University - Meneses Campus, v. 7, p. 1–6, 2018.

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

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

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, [2014-2015].

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.

WEISBURD, David et al. Advanced Statistics in Criminology and Criminal Justice. 5. ed. Switzerland: Springer, 2022. 552 p.