

Determinantes dos Crimes nos Municípios do Rio de Janeiro

Lucas Santos Leal

12/10/2020

Introdução

A redução dos índices de criminalidade é uma das principais tarefas do poder público em um tópico de grande interesse para a sociedade. O tema é ainda mais importante em países como o Brasil, que possuem características como desigualdade, altas taxas de desemprego e baixa escolaridade, que favorecem a ocorrência de crimes. A relevância do tópico exige a elaboração de pesquisas para sustentar a elaboração e implementação de políticas públicas. Nesse sentido, este trabalho avalia brevemente a situação da criminalidade nos municípios do Rio de Janeiro.

Em 2018, ocorreram aproximadamente 58 mil homicídios, no Brasil (27,8 mortes por 100 mil habitantes) (IPEA 2020). Esse número representou uma queda de 12% em comparação com 2017 (ano em que ocorreram 65,6 mil assassinatos), mas ainda reflete o elevado patamar em que o país se encontra em comparação com o restante do mundo. Em 2016, por exemplo, o Brasil ocupava a oitava posição entre os países com as maiores taxas de assassinato (UNODC, 2019).

Nos estados brasileiros, a violência é maior no Norte e Nordeste, regiões que apresentaram, em 2018, leve queda na tendência de crescimento dos últimos 10 anos da série, mas uma queda menos expressiva que as outras regiões do país, diferença parcialmente explicada pela guerra entre facções no sistema penitenciário (IPEA 2020). Naquele ano, os três estados mais violentos (e suas taxas por 100 mil habitantes) foram Roraima (71,8), Pará (53,2) e Ceará (54). Já os menos violentos foram São Paulo (8,2), Santa Catarina (11,9) e Minas Gerais (16).

O estado do Rio de Janeiro, de interesse desse trabalho, não se encontra entre os mais

violentos quando comparadas as taxas de homicídios, mas é um caso relevante para o país por conta da presença marcante do crime organizado, mais precisamente no tráfico de drogas. As taxas de homicídio no Rio apresentaram uma tendência favorável entre 2007 e 2012, chegando ao nível mais baixo dos últimos dez anos (29,4 homicídios por 100 mil) neste último ano. No entanto, após 2012, as taxas de violência tornaram a crescer, chegando a 38,4 em 2017, colocando-o em primeiro lugar na região Sudeste e em 13º no país (IPEA 2020). No estado, as maiores taxas de homicídios em 2018 ocorreram na região da Baixada Fluminense (56,2 por 100 mil). Na região da Grande Niterói foi observada a segunda maior taxa: 44 por 100 mil, seguida pelo interior com 37, e pela capital com 30,2.

A literatura é extensa no tratamento do assunto da criminalidade. São diversos os fenômenos que estão relacionados com a ocorrência de assassinatos, roubos, sequestros, entre outros crimes. Desde a existência de uma rede de crime organizado na região, desigualdades econômicas locais e baixa escolaridade da população, a ausência de policiamento e sistemas que falham em punir transgressores, diversos fatores influenciam as ações individuais dos criminosos e das vítimas, e também a ação estrutural e institucional.

Este trabalho seguirá o exposto por (???; Dos Anjos Júnior, Lombardi Filho, and Maia do Amaral 2018), que analisam fatores postos pela literatura como determinantes do crime, replicando suas variáveis explicativas (com algumas alterações operacionais devido à disponibilidade de dados). O objetivo do trabalho é identificar de que forma variáveis socioeconômicas e criminais estão associadas aos crimes nos municípios do estado do Rio de Janeiro, mais especificamente homicídios e roubos. A identificação de padrões municipais pode ser útil na elaboração de políticas públicas localizadas.

Marco Teórico

Diversos são os estudos que buscam compreender os determinantes de crimes (Cerqueira and Lobão 2004; Wolpin 1978; Soares 2004). Dada a existência de um grande número de tipos criminais, os motivos que os causam também são vários. Ainda, há um grande número de disciplinas interessadas nas causas de crimes, assim como nos padrões de distribuição dos eventos. Podem ser avaliados fatores psicológicos, individuais, ou mesmo fatores sociais econômicos que atuam de forma estrutural. Os diferentes pontos de vista muitas vezes utilizam as mesmas variáveis, algumas delas são educação, renda, desigualdade e presença de um aparato institucional consolidado. Esta seção faz um breve apanhado das principais vertentes teóricas existentes.

(Cerqueira and Lobão 2004) revisaram a literatura teórica e algumas explicações se destacam.

A primeira é a teoria da desorganização social. Para essa vertente, as comunidades têm papel fundamental na determinação dos crimes, pois são parte essencial da socialização dos indivíduos. Seguindo a teoria de Shaw e McKay, Sampson e Groves (1989) identificaram que nas comunidades inglesas caracterizadas por redes dispersas de relacionamento, grupos de adolescentes não vigiados e pouca participação organizacional, as taxas de criminalidade eram mais elevadas. Além disso, que essas características estavam associadas a níveis socioeconômicos mais baixos, heterogeneidade étnica e estruturas familiares frágeis. Outros trabalhos acrescentam mobilidade residencial como fator explicativo, associando-a a maiores índices de criminalidade (Boggess e Hipp, 2010; Dugan, 1999). Além da desorganização social, os autores expõem a teoria do aprendizado social em que os indivíduos moldam seu comportamento de acordo com as suas experiências em situações de conflito; a teoria do controle social, que se baseia na ideia de que quanto maior a ligação entre o indivíduo e a sociedade, traduzida em suas crenças nos valores culturais e sociais, menor a probabilidade de que cometa crimes. Para a teoria, o agente crê nos valores sociais que condenam atitudes criminosas. Os transgressores são, portanto, pessoas sem amarras sociais (Cerqueira and Lobão 2004).

Uma das teorias de maior repercussão é a da escolha racional. Iniciada por Gary Becker, em 1968, a abordagem passou a utilizar modelos formais de avaliação de custos e benefícios econômicos da atitude criminosa (Cerqueira and Lobão 2004). Com base nos pressupostos econômicos de racionalidade, a escolha de agir fora da lei teria como base fundamentos puramente pecuniários (Becker and Landes 1974). O que se seguiu foi uma série de estudos com base nos determinantes econômicos do crime, inspirados na ideia de que a criminalidade funciona da mesma forma que o mercado, fornecendo informações que são avaliadas pelos indivíduos, ainda que inconscientemente, que formulam suas ações a partir disto, numa ótica de oferta e demanda.

(???) um modelo na tentativa de explicar taxas nacionais de crimes, partindo dos pressupostos da escolha racional, mas de forma agregada. As variáveis explicativas do modelo utilizadas foram: (1) educação, que segundo o modelo, deve possuir efeito negativo na propensão a cometer crimes, devido às maiores expectativas de rendimento no mercado legal; (2) o passado criminal do agente, com efeito esperado negativo, considerando que o indivíduo com histórico criminal fica estigmatizado na população; (3) o nível e o crescimento da atividade econômica, que tem como efeito esperado a redução na taxa de criminalidade devido a maiores oportunidades no mercado de trabalho legal; (4) a existência de atividades criminais lucrativas como tráfico de drogas e de armas, com efeito esperado positivo; por fim (5) a força do sistema de polícia e de justiça, que reduz a expectativa de cometer crimes impunemente, assim como

a possibilidade de penas sérias. Os resultados de seus modelos econométricos mostram que crescimento econômico (PIB), expectativa de ser preso, e severidade do sistema judicial têm impacto negativo nas taxas de criminalidade. Do outro lado, desigualdade econômica, taxas de criminalidade dos anos anteriores, existência de um mercado de drogas consolidado possuem efeitos positivos nas taxas. No Brasil, os trabalhos também possuem origens explicativas distintas. Os estudos brasileiros costumam utilizar variáveis macroeconômicas para explicar violência (Cerqueira and Lobão 2004). Os resultados apontam para relações positivas entre crimes e taxa de urbanização, densidade demográfica, e desigualdade de renda. Soares and Naritomi (2010) identificam que três fatores relevantes estão associados com a alta incidência de crimes na América Latina: desigualdade de renda, presença policial e baixas taxas de encarceramento.

Outros estudos, notadamente (Buonanno, n.d.; Lochner and Moretti 2004) identificam a educação como fator importante no nível de criminalidade. É de se esperar que indivíduos que possuem maior acesso à educação acabam por ter maiores oportunidades além do crime, ou seja, o custo de oportunidade da ação criminosa aumenta (Lochner and Moretti 2004). Ainda na ótica da escolha racional, o indivíduo, ao considerar o tempo perdido na prisão, percebe que não é racional agir de forma a acarretar tal punição, reduzindo a propensão ao crime.

Alguns estudos que analisam as características espaciais de fenômenos criminais. Anselin (1988) afirma que, de acordo com as teorias ecológicas, a distribuição de crimes é determinada pela interseção espacial entre agressores motivados, vítimas potenciais e a ausência de opressores de crimes. Essa situação pode ser potencializada ou atenuada de acordo com determinadas características ecológicas em dado local. Para os autores, dois fatores são fundamentais para facilitar ou inibir a ocorrência de crimes. Em primeiro lugar, características físicas de determinada localidade como prédios, casas abandonadas, espaços abertos, têm a capacidade de ampliar ou diminuir o controle social. Em segundo lugar, o espaço pode ser influenciado por atividades rotineiras, ou seja, áreas comerciais, estacionamentos, locais com grande quantidade de consumo de álcool, podem potencializar a ocorrência de delitos e violência (Anselin et. al. 2000, Cohen, Gorr, e Olligschlaeger, 1993). Locais onde há grandes vias de acesso também estão relacionados com uma presença maior de atividade criminosa (Ribeiro e Diniz, 2005)

(Dos Anjos Júnior, Lombardi Filho, and Maia do Amaral 2018) incrementam a análise do crime adicionando a dimensão espacial aos seus modelos. Utilizando dados da região sudeste, os autores avaliam o impacto da densidade populacional, do PIB per capita, da taxa de desemprego, do índice de gini, da taxa de urbanização, da proporção de homens, da proporção

de jovens e da taxa de analfabetismo. Os resultados encontrados foram que apenas o gini, a proporção de homens e a taxa de analfabetismo não tiveram valores significativos. Todas as variáveis significativas tiveram impacto positivo na taxa de homicídios por 100 mil habitantes, com exceção da urbanização, que teve impacto significativo e negativo, de encontro com o esperado pela teoria.

Tomando como base o que foi exposto, este trabalho busca avaliar os determinantes da criminalidade. Em primeiro lugar, a educação aparece como fator explicativo relevante. Indivíduos que estudam acabam por ampliar suas oportunidades de empregabilidade e até mesmo de socialização, podendo portanto causar variações na criminalidade. Em segundo lugar, a renda aparece como variável explicativa. Em locais com maior acesso a renda e emprego, indivíduos têm menos motivos para entrar na atividade criminosa. Da ótica da escolha racional, caso o payoff de uma atividade criminosa seja menor que o de um trabalho, o indivíduo busca o trabalho, caso contrário, busca o crime. Outro fator utilizado para explicar a presença de criminalidade é a existência de um mercado do crime em determinado local que permita a uma pessoa iniciar a vida no crime mais rapidamente, novamente reduzindo os custos da atividade criminal. A densidade populacional também é um fator relevante, pois em locais com muitas pessoas, é seria mais difícil identificar criminosos do que em locais com baixa densidade, alta densidade também favorece a existência de locais aglomerados onde alguns crimes são mais fáceis de serem realizados. Por fim, a desigualdade econômica deve ser considerada, onde há maior desigualdade de renda, os indivíduos mais pobres têm menor acesso a bens do que os mais ricos do que em locais com menor desigualdade, o que pode favorecer a atitude criminosa.

Hipóteses

A partir da teoria as seguintes hipóteses foram formuladas:

1. Quanto maior o nível de educação de um município menor será a criminalidade;
2. Quanto maior a densidade populacional de um município, maior será a criminalidade;
3. Quanto maior a renda média dos trabalhadores, menor a criminalidade;
4. Quanto maior o mercado de drogas em um município maiores serão as taxas de homicídio e de roubo;
5. Quanto maior a desigualdade de renda maior a criminalidade;
6. Quanto maior o desenvolvimento do município maior será o nível de criminalidade.

As tabelas a seguir relacionam as variáveis e os sentidos causais esperados.

Variáveis independentes	Fonte	Ano
Taxa de homicídios	ISP	2019
taxa de roubos	ISP	2019
acima da média de homicídios	ISP	2019
acima da média de roubos	ISP	2019

Variável	Sentido esperado	Fonte
Densidade Populacional	+	IBGE
Renda média trabalhadores formais	-	IBGE
Índice de Gini	+	IBGE
IDHM	-	IBGE
Nota IDEB	-	IBGE
Prisões por tráfico	+	ISP

Operacionalização e descrição das variáveis

As quatro variáveis independentes não estão nos bancos de dados originais. Portanto será necessário adicioná-las. Para calcular a taxa de homicídios e de roubos basta dividir o número de interesse pela população e multiplicar por 100 mil. Por fim, para computar se um estado está ou não acima da média, basta criar uma variável que assume os valores 0 quando o valor de interesse estiver abaixo da média e 1 quando estiver acima (categoria de referência). A variável independente de tráfico também será elaborada posteriormente considerando-se a taxa de prisões por tráfico por 100 mil habitantes.

A densidade demográfica é a razão entre população e área territorial do município. A renda média dos trabalhadores formais computa o salário médio dos trabalhadores de carteira assinada no município. Essa variável não é a mais adequada para verificar a renda dos trabalhadores do município, afinal grande parcela da população possui empregos informais. No entanto, é possível utilizá-la como proxy das diferenças entre as rendas municipais. Para verificar a desigualdade de renda, o índice de gini varia de 0 a 1, com valores mais próximos de zero correspondendo a situações de maior distribuição. O índice não considera os diferentes níveis de renda entre os municípios, ou seja, é possível que um município pobre possua uma distribuição igualitária. Para verificar o nível de desenvolvimento de um município será utilizado o índice de desenvolvimento humano municipal. Para verificar a relação entre crime e educação utilizarei as notas do IDEB municipal. É necessário observar que estes dados se

referem às escolas públicas naquele município, o que exclui uma parcela da população. No entanto, considerando as hipóteses de que a renda está associada com a criminalidade, é plausível considerar a variável relevante, já que a população mais pobre está mais representada nas escolas públicas. Por fim, para verificar a presença do tráfico de drogas será utilizada a variável taxa de prisões por tráfico por 100 mil habitantes.

Algumas observações finais merecem ser feitas. Uma das variáveis relevantes abordadas pela teoria não pode ser verificada neste estudo: a presença de um aparato policial forte. Considerando que a responsabilidade pela segurança pública local é em sua maior parte do nível estadual, não é comum encontrar dados como efetivo policial e gastos policiais por município, portanto este trabalho não irá testar hipóteses relacionadas à força policial.

Os dados utilizados neste trabalho são relativos aos 90 municípios do estado do Rio de Janeiro para que todas as variáveis estavam disponíveis. No total serão utilizadas 10 variáveis, das quais quatro são dependentes. A primeira variável independente será a taxa de homicídios por 100 mil habitantes, forma padrão nos estudos sobre violência. A segunda será taxa de roubos por 100 mil habitantes. O objetivo de utilizar taxas e não valores brutos é para controlar pela população, afinal, um local com muito mais habitantes não tem mais homicídios ou roubos que outro necessariamente, mas também por possuir mais pessoas. A utilização de taxas por 100 mil, no entanto, não é suficiente para controlar todos os problemas existentes. Um problema que surge é o dos municípios com população muito pequena. Como homicídios são eventos raros, é comum que em muitas cidades o número de pessoas assassinadas em um determinado ano seja zero, ou um número pequeno. Mas até mesmo pequenas variações no número absoluto, podem causar variações extremas na taxa. No caso de um ano com muitos homicídios, a alteração na taxa pode ser grande, dependendo da população, mas o evento pode ter ocorrido por causas aleatórias. As formas de evitar esse problema são a avaliação das tendências temporais, para verificar qual em que direção o crime está indo ao longo do tempo para cada cidade, e a verificação das tendências espaciais, quando a taxa de homicídios em um lugar é ponderada pelas taxas das localidades vizinhas. Nenhum desses métodos será utilizado neste trabalho, pois fogem dos objetivos.

Esse problema ocorre também com roubos, mas em menor grau, considerando-se que não são eventos tão raros quando homicídios, o que permite uma variação maior, e melhor interpretação das taxas, ainda que quando baixas. Seguindo a teoria e as hipóteses elaboradas, as tabelas a seguir destacam as estatísticas descritivas das variáveis.

##

##

Mean

Std.Dev

Min

Max

##	-----	-----	-----	-----	-----
##	Dens_dem_km_2_2010	700.91	1892.76	12.67	13024.56
##	Ideb_rede_publica	4.23	0.48	3.40	5.30
##	IDHM_2010	0.71	0.04	0.61	0.84
##	media_hom	0.21	0.41	0.00	1.00
##	media_roubos	0.11	0.32	0.00	1.00
##	media_salarial_salarios_minimos	2.24	0.70	1.50	6.40
##	PIB_percap_2017	32897.13	25550.23	13505.21	151721.49
##	taxa_hom	20.96	14.50	0.00	56.00
##	taxa_roubos	345.04	489.99	0.00	2043.11
##	trafico	127.81	85.02	13.88	437.76

É relevante compreender a distribuição das variáveis utilizadas. Para tanto, a figura a seguir exemplifica as distribuições de frequência das variáveis independentes, assim como suas correlações entre si. A figura abaixo evidencia tais características. Alguns fatores merecem atenção. A correlação entre o IDHM e o log densidade demográfica foi de 0.52, significativa a nível de 0.5, valor considerado moderado. A relação era esperada, devido ao fato de as cidades com maior índice de desenvolvimento estarem situadas na região metropolitana do Rio de Janeiro, as quais possuem maior urbanização e população. Outra correlação que merece atenção ocorre entre as variáveis salário médio dos trabalhadores formais e o PIB per capita, que teve valor 0.6. No entanto, deve-se estar atento à possibilidade de multicolinearidade quando forem efetuadas as regressões, para isso, serão realizados os testes indicados, na seção metodologia. As outras correlações foram fracas ou próximas de zero.

Modelos de regressão

Para testar as hipóteses elaboradas, será utilizado um modelo de regressão linear e um modelo de regressão logística. No primeiro caso, as variáveis dependentes serão a taxa de homicídios por cem mil habitantes e a taxa de roubos por cem mil habitantes. No segundo caso, as variáveis independentes serão o fato de um município específico estar ou não acima da taxa média de homicídios e de roubos. Ou seja, quatro modelos serão testados no total.

A regressão linear é utilizada para estimar uma relação linear entre as variáveis. A estrutura é a seguinte:

$$Y = \alpha + \beta_1 X + e$$

em que α representa o termo constante do modelo, ou seja, o valor médio da variável dependente quando as variáveis independentes têm valor zero, β indica o coeficiente linear

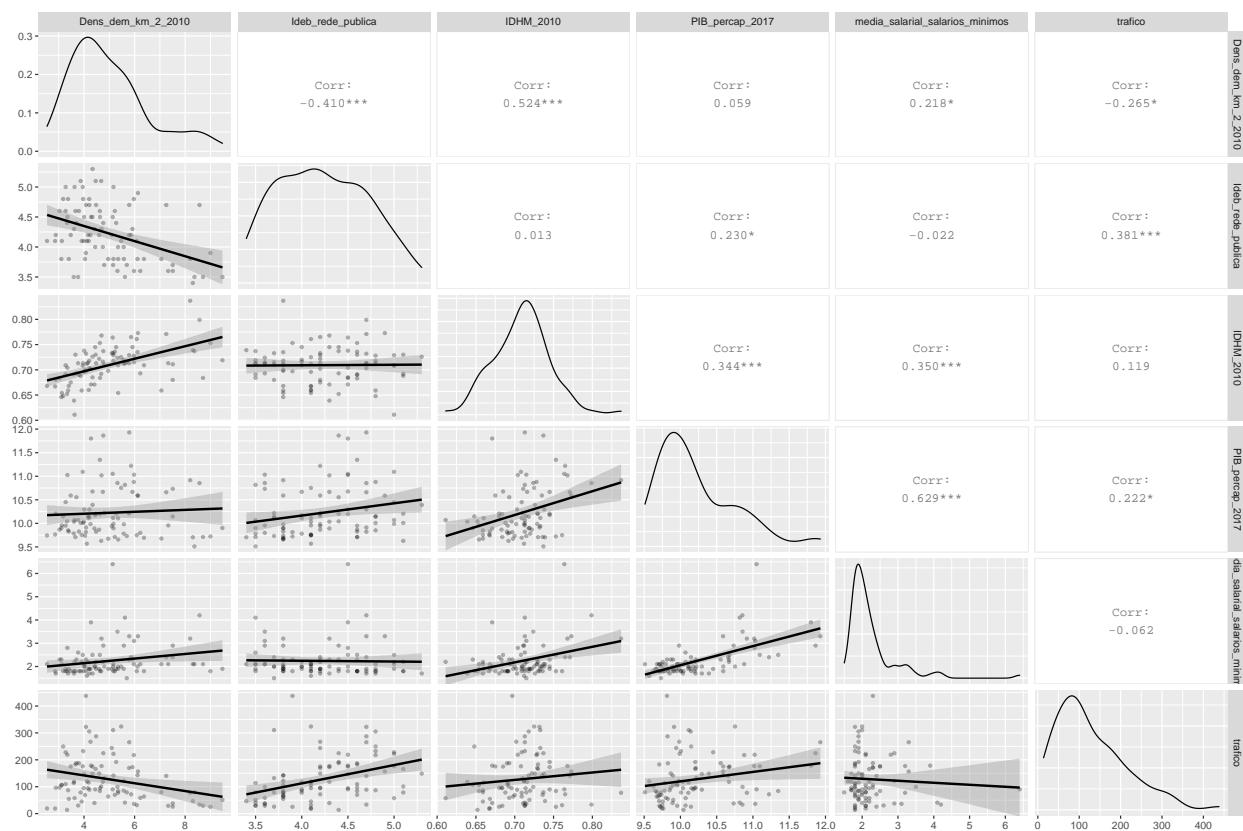


Figure 1: Tabela de correlação das variáveis independentes

do modelo, ou seja, a variação esperada em Y para cada variação em X (que representa a variável independente) e por fim o termo estocástico e que absorve toda a variação que não for considerada pelas variáveis do modelo. Observe que o coeficiente de regressão é constante para todos os valores da variável independente, daí a natureza linear do modelo.

A utilização de um modelo de regressão linear requer a observação de uma série de pressupostos (Gujarati et al. 2011; James et al. 2013; Wooldridge 2016). Os mais relevantes para esse trabalho são:

1. linearidade dos parâmetros;
2. valor esperado do erro igual a zero ($E[e_i|X_i] = 0$);
3. homoscedasticidade, ou seja, variância constante do erro ($\text{Var}(e_i|X_i) = \sigma^2$);
4. ausência de autocorrelação entre os erros ($\text{cov}(e_i, e_j|X_i, X_j) = 0$);
5. ausência de colinearidade perfeita.

O pressuposto da linearidade dos parâmetros não diz sobre a relação entre Y e X que pode ser não linear. Trata fundamentalmente da importância da interpretação dos coeficientes (Wooldridge 2016) que são constantes e refletem mudanças diretas em Y baseadas em mudanças de X . Tal relação é constante ao longo dos valores de X , algo que não é verdadeiro para coeficientes não lineares.

Já o segundo pressuposto, da esperança do erro ser igual a zero se refere à característica estocástica do erro em um modelo bem especificado. O erro representa tudo aquilo que não está explicitado no modelo e deve eliminar, de preferência, quaisquer elementos sistemáticos que possam explicar a variação de Y . Se a média condicional dos erros é zero, para determinados valores de X , a diferença entre o valor estimado e o valor real de Y é, em média, zero.

A homoscedasticidade está relacionada com a variância do erro, que deve ser constante ao para determinados valores das variáveis independentes. A heteroscedasticidade ocorre justamente quando a variância dos erros não é constante. A heteroscedasticidade não causa viés na estimativa, mas aumenta a incerteza no modelo, à medida que afeta os erros-padrão e as estatísticas de teste (???). O teste de Breusch-Pagan é o mais utilizado para identificação da presença de heteroscedasticidade em um modelo linear. A ausência de autocorrelação, pressuposto 4, está associada à existência de fatores sistemáticos além de X que influenciam a variável dependente. Quando os erros entre observações estão relacionados pode-se considerar que há fatores não aleatórios não considerados que estão associados a mudanças unitárias das variáveis.

Por fim, a ausência de colinearidade perfeita, ou seja, a existência de multicolinearidade. O que se espera é que nenhuma variável explicativa seja uma combinação linear das outras,

ou que não estejam fortemente associadas. O que a ruptura desse pressuposto implica é que a mesma parcela de variação de Y está sendo explicada por dois fatores distintos (nominalmente). Esta redundância pode interferir no erro-padrão e também nas estimativas dos coeficientes. Os remédios são vários e podem depender da causa do problema.

O método de estimação da reta de regressão mais utilizado é o de mínimos quadrados ordinários. Este método busca minimizar a soma dos quadrados dos resíduos, ou

$$\operatorname{argmin} \sum \hat{e}_i^2 = \operatorname{argmin} \sum (Y_i - \hat{Y}_i)$$

. A estimativa por MQO é baseada nos pressupostos elucidados acima e caso sejam verdadeiras, o estimador $\hat{\beta}$ é o melhor estimador linear não viesado (Gujarati et al. 2011). A interpretação do output de um modelo de regressão será explicada na avaliação dos resultados deste trabalho.

O modelo de regressão logística tem como objetivo identificar a probabilidade de pertencimento a uma categoria (James et al. 2013). Para isso é necessário que a variável dependente seja binária, com uma categoria de referência. É possível estimar uma regressão linear com a variável dependente binária, mas alguns problemas são comuns. O principal deles é o fato de que as estimativas das probabilidades podem estar além do intervalo (0,1), fato que é eliminado com o modelo logístico. A especificação do modelo é: para dados valores de X , a probabilidade de Y assumir o valor 1 (categoria de referência) é:

$$p(Y = 1|X) = \frac{e^z}{1 + e^z}$$

para

$$z = \beta_0 + \beta_1 X$$

A função logística é o link que assume valores entre zero e um. Através do método de máxima verossimilhança o modelo é estimado. A relação linear se encontra após algumas transformações que resultam no log da razão de probabilidades (*odds ratio*):

$$\log\left(\frac{P(X)}{1 - P(X)}\right) = \log(e^z) = z = \beta_0 + \beta_1 X$$

Alguns pressupostos devem ser observados(Stoltzfus 2011):

1. Independência dos erros;
2. linearidade entre a razão de probabilidades e as variáveis independentes;
3. Ausência de multicolinearidade;

4. Ausência de outliers influentes.

A interpretação dos resultados de uma regressão logística também é diferente. Como o coeficiente da reta tangente a um determinado ponto da reta de regressão, o output do modelo não pode ser avaliado como a variação na probabilidade para uma mudança na variável independente. A interpretação deve considerar a razão de probabilidades (logits) para cada aumento unitário das variáveis independentes.

Metodologia

Neste trabalho serão utilizadas os dois tipos de regressão acima expostos para testar as hipóteses. Primeiramente serão realizados dois modelos de regressão linear múltipla, um para cada variável dependente. Depois será realizada uma regressão logística para verificar a probabilidade de estar acima ou abaixo da média das variáveis dependentes. Por fim serão realizados os testes necessários para verificar os pressupostos. Os dois primeiros modelos assumirão a forma:

$$taxahomicidio = \alpha + \beta_1.densidade + \beta_2.educação + \beta_3.IDHM + \beta_4.Gini + \beta_5.tráfico + \beta_6.renda + e \quad (1)$$

$$taxaroubos = \alpha + \beta_1.densidade + \beta_2.educação + \beta_3.IDHM + \beta_4.Gini + \beta_5.tráfico + \beta_6.renda + e \quad (2)$$

O último:

$$\log\left(\frac{P(X)}{1 - P(X)}\right) = mediahomi = \alpha + \beta_1.densidade + \beta_2.educação + \beta_3.IDHM + \beta_4.Gini + \beta_5.tráfico + \beta_6.renda + e \quad (3)$$

Os resultados estão exibidos abaixo. O primeiro modelo testou a variável dependente taxa de homicídios por 100 mil habitantes. Os resultados mostram que apenas a nota do IDEB e o log da densidade demográfica tiveram coeficientes estatisticamente diferentes de zero. Cada mudança percentual na densidade está associada a uma mudança de 2.5% na taxa de homicídios por 100 mil, conforme esperado. A relação com o IDEB também foi a esperada, que para cada variação unitária na nota do IDEB, a taxa de homicídios varia negativamente em 7.5 unidades. Embora não seus coeficientes não tenham sido significativos, as outras

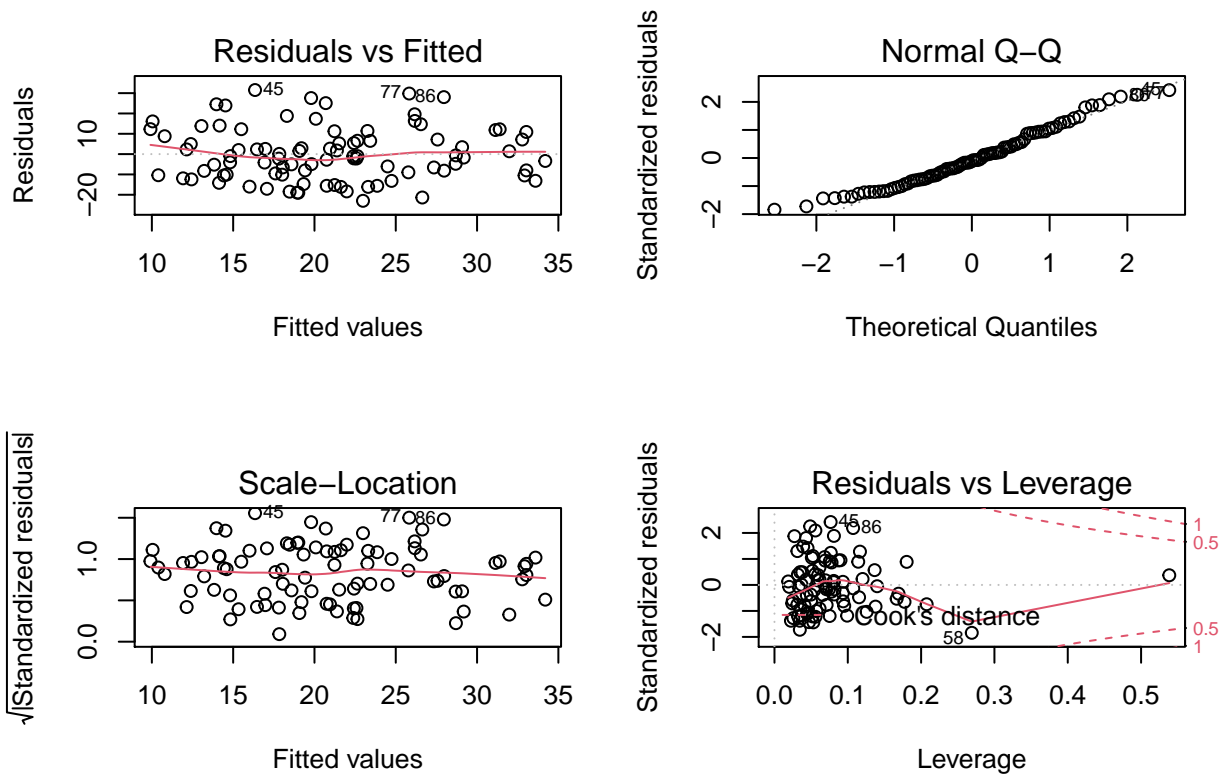
variáveis do modelo tiveram os sinais esperados, com exceção da renda média. O R2 ajustado do modelo foi de 12%, indicando que o modelo ajusta este valor da variação da taxa de homicídios. Considerando que a média da variável dependente está em torno de 20, o erro padrão residual foi alto, 13.56.

```
##
## Call:
## lm(formula = taxa_hom ~ log(Dens_dem_km_2_2010) + as.numeric(Ideb_rede_publica) +
##     IDHM_2010 + log(PIB_percap_2017) + trafico + media_salarial_salarios_minimos,
##     data = banco_final)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -22.985 -10.108  -1.392   8.399  31.529
##
## Coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      35.74502    42.10823   0.849   0.3984
## log(Dens_dem_km_2_2010)   2.51041    1.28972   1.946   0.0550 .
## as.numeric(Ideb_rede_publica)  -7.49066    3.63506  -2.061   0.0425 *
## IDHM_2010        -75.83538   52.21768  -1.452   0.1502
## log(PIB_percap_2017)    4.86794    3.77485   1.290   0.2008
## trafico           0.02067    0.01958   1.055   0.2943
## media_salarial_salarios_minimos  2.56762    2.81104   0.913   0.3637
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.56 on 83 degrees of freedom
## Multiple R-squared:  0.1847, Adjusted R-squared:  0.1257
## F-statistic: 3.133 on 6 and 83 DF,  p-value: 0.008104
```

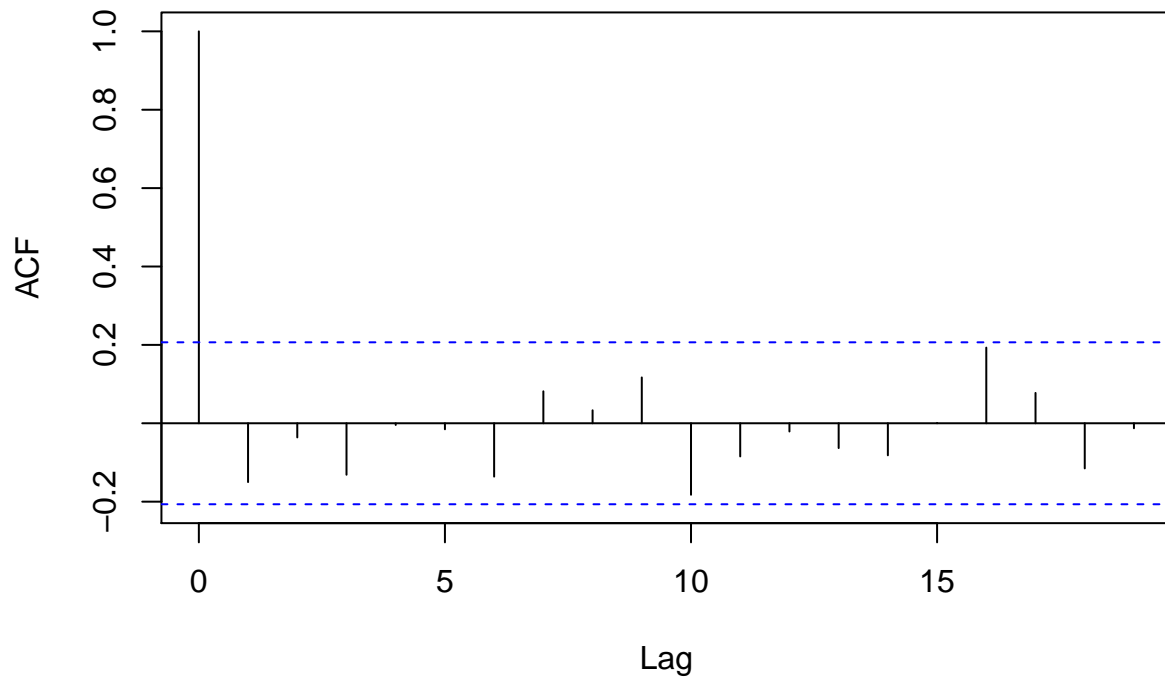
A seguir foi realizado o teste de variância constante dos erros. Os gráficos mostram que não há indícios de heteroscedasticidade forte, apresentando uma distribuição aparentemente aleatória dos resíduos. O gráfico da distância de Cook também indica que não houve outliers muito influentes. Pra confirmar as expectativas, foi realizado um teste de Breusch-Pagan. O resultado mostra um p-valor elevado, indicando a impossibilidade de rejeição da hipótese nula de que a variância dos erros é constante. Ainda, o teste de autocorrelação indica que as

observações são independentes. Por fim, o teste de VIF indica que não há multicolinearidade no modelo.

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 0.1189018, Df = 1, p = 0.73023
```



Series modelo_taxa_hom\$residuals



```
##          log(Dens_dem_km_2_2010)  as.numeric(Ideb_rede_publica)
##                               1.932277                      1.450872
##          IDHM_2010                log(PIB_percap_2017)
##                               1.781522                      1.996711
##          trafico media_salarial_salarios_minimos
##                               1.341414                      1.889256
```

O próximo passo é interpretar o modelo com a variável dependente taxa de roubos por 100 mil habitantes. A princípio, o modelo possui três variáveis com coeficientes diferentes de zero: densidade demográfica, IDHM, e tráfico de drogas. Os sentidos são esperados para densidade e IDHM, quando maior a densidade, maior a taxa de roubos, quando maior o IDHM, menor é a taxa. Para tráfico, o coeficiente está no sentido contrário do esperado, quando mais prisões por tráfico, menos crimes foram registrados. O R² é de 79% e o erro padrão 230.

```
##
## Call:
## lm(formula = taxa_roubos ~ log(Dens_dem_km_2_2010) + as.numeric(Ideb_rede_publica) +
##     IDHM_2010 + log(PIB_percap_2017) + media_salarial_salarios_minimos +
```

```

##      trafico, data = banco_final)
##
## Residuals:
##      Min        1Q    Median        3Q        Max
## -561.05 -166.54   -1.26   149.53   623.14
##
## Coefficients:
##                                Estimate Std. Error t value
## (Intercept)                   755.2503    716.2958   1.054
## log(Dens_dem_km_2_2010)       260.2202     21.9392  11.861
## as.numeric(Ideb_rede_publica) -94.4229     61.8354  -1.527
## IDHM_2010                     -1528.7431    888.2657  -1.721
## log(PIB_percap_2017)          -27.1364     64.2133  -0.423
## media_salarial_salarios_minimos  69.6525     47.8182   1.457
## trafico                       -0.7880      0.3331  -2.365
##                                Pr(>|t|)
## (Intercept)                   0.2948
## log(Dens_dem_km_2_2010)       <0.0000000000000002 ***
## as.numeric(Ideb_rede_publica)  0.1306
## IDHM_2010                     0.0890 .
## log(PIB_percap_2017)          0.6737
## media_salarial_salarios_minimos 0.1490
## trafico                       0.0203 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 230.7 on 83 degrees of freedom
## Multiple R-squared:  0.7933, Adjusted R-squared:  0.7783
## F-statistic: 53.08 on 6 and 83 DF,  p-value: < 0.00000000000000022

```

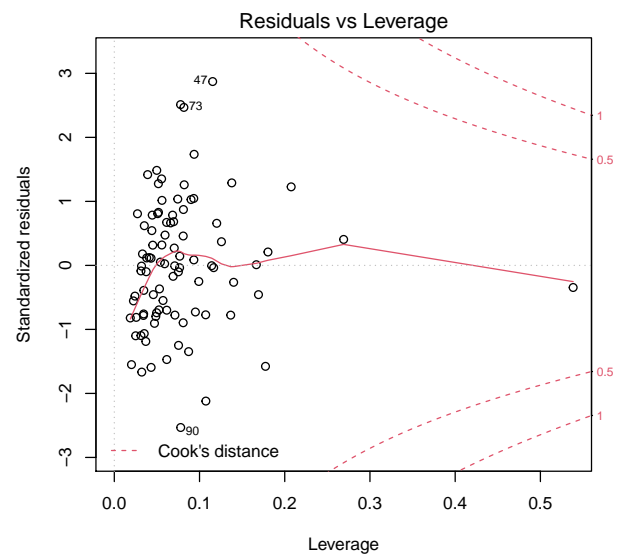
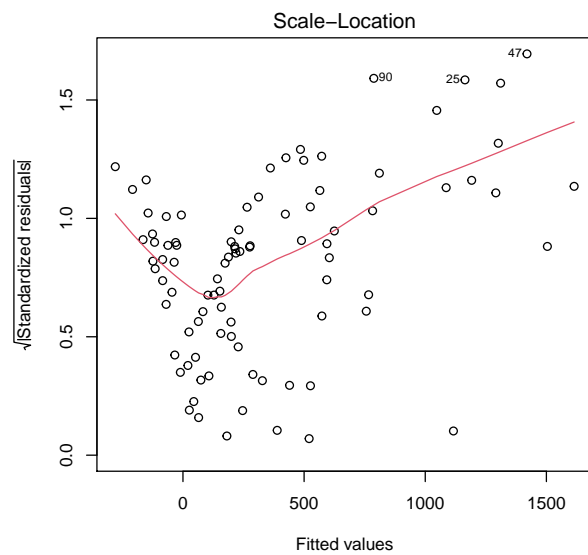
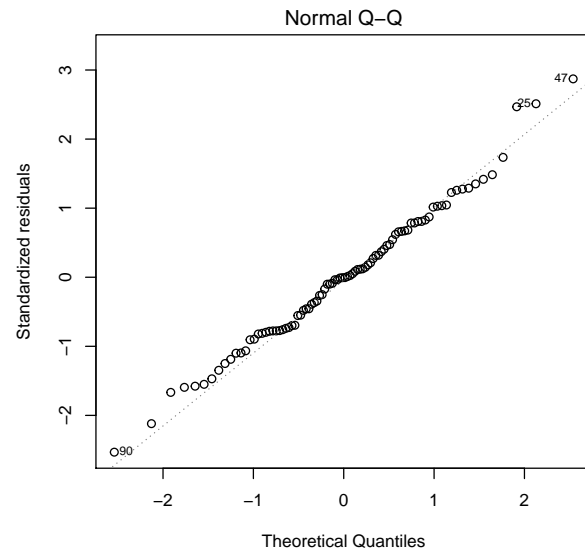
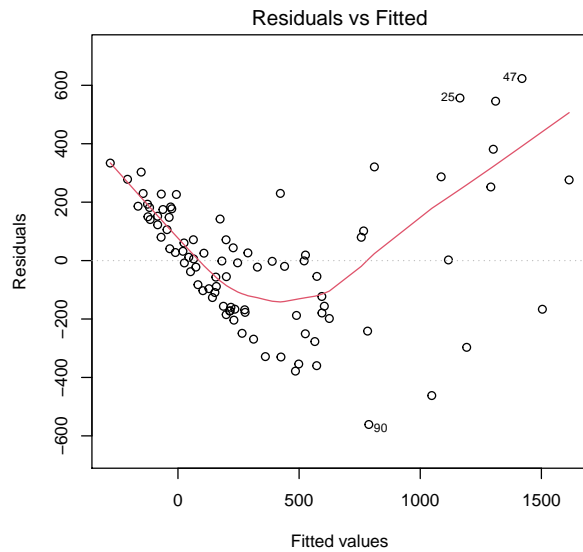
Mas é necessário identificar se o modelo está de acordo com os pressupostos. Os gráficos de diagnóstico indicam a possibilidade de heteroscedasticidade, suspeita que é confirmada com o teste de variância constante, que tem p-valor significativo. Os outros testes, de multicolinearidade e de autocorrelação indicam que o modelo está de acordo com os pressupostos.

```

## Non-constant Variance Score Test
## Variance formula: ~ fitted.values

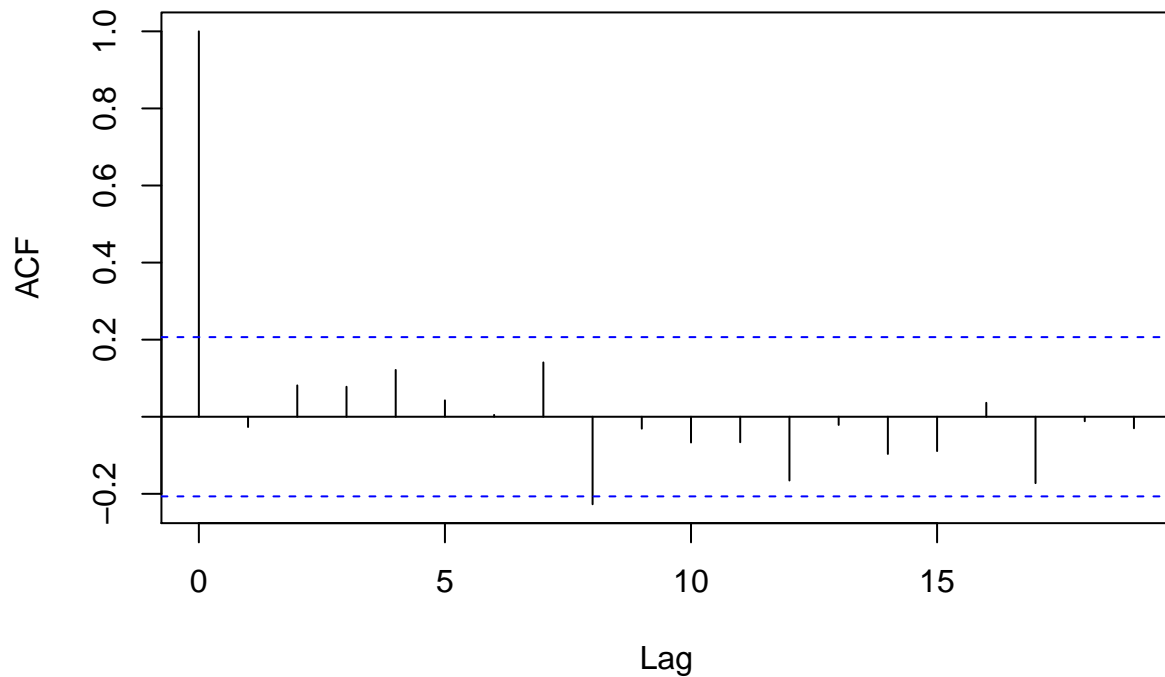
```


Chisquare = 28.05388, Df = 1, p = 0.00000011798



##	log(Dens_dem_km_2_2010)	as.numeric(Ideb_rede_publica)
##	1.932277	1.450872
##	IDHM_2010	log(PIB_percap_2017)
##	1.781522	1.996711
##	media_salarial_salarios_minimos	trafico
##	1.889256	1.341414

Series modelo_taxa_roubos\$residuals



Por fim vamos avaliar o último modelo deste trabalho. Decidi não utilizar um modelo logístico para a variável taxa de roubos por conta do pequeno número de casos que seria atribuído a uma das categorias, dificultando a interpretação do modelo. Para a variável “acima da média de homicídios” PIB per capta, densidade e IDEB foram as variáveis significativas do modelo. A interpretação inicial dos resultados implica que mudanças de uma unidade da nota do IDEB geram mudanças de -3.15 no log das probabilidades. Já o PIB per capta está associado a uma mudança de 1.68% no log das probabilidades, para cada aumento percentual. A densidade demográfica está associada a um aumento de 0.95% no log das chances, para cada variação percentual.

```
##
## Call:
## glm(formula = media_hom ~ log(Dens_dem_km_2_2010) + as.numeric(Ideb_rede_publica) +
##     IDHM_2010 + log(PIB_percap_2017) + trafico + media_salarial_salarios_minimos,
##     family = binomial(link = "logit"), data = banco_final)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
```

```

## -2.10399 -0.36397 -0.13576 -0.03559 1.86216
##
## Coefficients:
##
## Estimate Std. Error z value Pr(>|z|)
## (Intercept) -20.062256 11.215261 -1.789 0.07364 .
## log(Dens_dem_km_2_2010) 0.963576 0.347599 2.772 0.00557 **
## as.numeric(Ideb_rede_publica) -3.152090 1.400644 -2.250 0.02442 *
## IDHM_2010 10.767332 14.719216 0.732 0.46446
## log(PIB_percap_2017) 1.682258 0.905014 1.859 0.06305 .
## trafico -0.001263 0.006208 -0.203 0.83882
## media_salarial_salarios_minimos 0.476826 0.521942 0.914 0.36095
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 92.777 on 89 degrees of freedom
## Residual deviance: 43.457 on 83 degrees of freedom
## AIC: 57.457
##
## Number of Fisher Scoring iterations: 7
##
## Average marginal effects
## glm(formula = media_hom ~ log(Dens_dem_km_2_2010) + as.numeric(Ideb_rede_publica) +
## Dens_dem_km_2_2010 Ideb_rede_publica IDHM_2010 PIB_percap_2017 trafico
## 0.0003859 -0.2309 0.7889 0.000004301 -0.00009252
## media_salarial_salarios_minimos
## 0.03494
##
## Confusion Matrix and Statistics
##
## Reference
## Prediction 0 1
## 0 68 3
## 1 6 13
##

```

```

##                Accuracy : 0.9
##                95% CI : (0.8186, 0.9532)
##      No Information Rate : 0.8222
##      P-Value [Acc > NIR] : 0.03029
##
##                Kappa : 0.6814
##
##  McNemar's Test P-Value : 0.50499
##
##      Sensitivity : 0.9189
##      Specificity : 0.8125
##      Pos Pred Value : 0.9577
##      Neg Pred Value : 0.6842
##      Prevalence : 0.8222
##      Detection Rate : 0.7556
##      Detection Prevalence : 0.7889
##      Balanced Accuracy : 0.8657
##
##      'Positive' Class : 0
##

```

A matriz de confusão fornece informações sobre a capacidade de previsão do modelo. A precisão do modelo foi de 90%, 7.2% maior do que um modelo que considere apenas os valores da variável dependente. Portanto as variáveis independentes ajudaram a gerar previsões mais precisas.

Considerações Finais

Este trabalho buscou avaliar que fatores determinam os homicídios e roubos nos municípios do Rio de Janeiro. Os resultados encontrados estão de acordo com a literatura abordada. Quando avaliada a taxa de homicídios, foi identificado que educação, densidade demográfica têm impactos significativos. Para roubos, educação, IDHM e tráfico tiveram valores significantes, embora o resultado para tráfico tenha sido no sentido contrário ao esperado. Por fim, o modelo logístico utilizando as variáveis resultou numa capacidade de classificação entre municípios acima da média ou abaixo da média maior do que num modelo sem variáveis. Os resultados são interessantes, mas algumas ressalvas são importantes. Em primeiro lugar é necessário destacar que crimes são um evento naturalmente sazonais. Este efeito não foi avaliado neste

trabalho, pois não fazia parte de sua proposta, mas certamente interfere na capacidade de interpretação dos resultados. Um modelo de dados em painel poderia considerar esses efeitos. Em segundo lugar, crimes apresentam padrões espaciais, e as taxas em uma localidade são por muitas vezes afetadas pelas localidades vizinhas, estes efeitos também não foram levados em consideração. Todas essas questões são relevantes e podem ser conferidas em trabalhos futuros, afinal há muito se conhecem os determinantes da criminalidade, no entanto, as políticas públicas por muitas vezes falham em conseguir atingir seus objetivos de redução nos crimes.

Referências

- Anselin, Luc. 1988. *Spatial Econometrics: Methods and Models*. Vol. 4. Studies in Operational Regional Science. Dordrecht: Springer Netherlands. <https://doi.org/10.1007/978-94-015-7799-1>.
- Becker, Gary S., and William M. Landes. 1974. *Essays in the Economics of Crime and Punishment*. Human Behavior and Social Institutions 3. New York: National Bureau of Economic Research : distributed by Columbia University Press.
- Buonanno, Paolo. n.d. “The Socioeconomic Determinants of Crime. A Review of the Literature,” 34.
- Cerqueira, Daniel, and Waldir Lobão. 2004. “Determinantes Da Criminalidade: Arcabouços Teóricos E Resultados Empíricos.” *Dados* 47 (2): 233–69. <https://doi.org/10.1590/S0011-52582004000200002>.
- Dos Anjos Júnior, Otoniel Rodrigues, Stélio Coêlho Lombardi Filho, and Pedro Vasconcelos Maia do Amaral. 2018. “Determinantes Da Criminalidade Na Região Sudeste Do Brasil: Uma Aplicação de Painel Espacial.” *Economía Sociedad Y Territorio*, no. 57 (May). <https://doi.org/10.22136/est20181176>.
- Gujarati, Damodar N., Dawn C. Porter, Denise Durante, Mônica Rosemberg, Maria Lúcia G. L. Rosa, Claudio D. Shikida, Ari Francisco de Araújo Júnior, and Márcio Antônio Salvato. 2011. *Econometria Básica*. 5ª Edição. AMGH.
- IPEA. 2020. “© Instituto de Pesquisa Econômica Aplicada – Ipea 2020,” 96.
- James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. *An Introduction to Statistical Learning*. Vol. 103. Springer Texts in Statistics. New York, NY: Springer New York. <https://doi.org/10.1007/978-1-4614-7138-7>.

Lochner, Lance, and Enrico Moretti. 2004. "The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports." *THE AMERICAN ECONOMIC REVIEW* 94 (1): 65.

Soares, Rodrigo, and Joana Naritomi. 2010. "Understanding High Crime Rates in Latin America: The Role of Social and Policy Factors." In *The Economics of Crime: Lessons for and from Latin America*, 472. A National Bureau of Economic Research Conference Report. Chicago: University of Chicago Press.

Soares, Rodrigo R. 2004. "Development, Crime and Punishment: Accounting for the International Differences in Crime Rates." *Journal of Development Economics*, 30.

Stoltzfus, Jill C. 2011. "Logistic Regression: A Brief Primer: LOGISTIC REGRESSION: A BRIEF PRIMER." *Academic Emergency Medicine* 18 (10): 1099–1104. <https://doi.org/10.1111/j.1553-2712.2011.01185.x>.

Wolpin, Kenneth I. 1978. "An Economic Analysis of Crime and Punishment in England and Wales, 1894-1967." *Journal of Political Economy* 86 (5): 815–40. <http://www.jstor.org/stable/1828411>.

Wooldridge, Jeffrey M. 2016. *Introductory Econometrics: A Modern Approach*. Sixth edition. Boston, MA: Cengage Learning.