

Análise de relação entre variáveis de ocorrências de crimes da cidade do Recife

Estudo de caso por meio da avaliação de dados da Secretaria de Defesa Social

Carolina Lima Gomes de Melo¹  orcid.org/0000-0001-7826-6392

Rodrigo Lins Rodrigues²  orcid.org/0000-0002-3598-5204

¹ Escola Politécnica de Pernambuco, Universidade de Pernambuco, Recife, Brasil,

² Departamento de Educação, Universidade Federal Rural de Pernambuco, Recife, Brasil

E-mail do autor principal: clgm@ecom.poli.br

Resumo

O aumento da criminalidade nos últimos anos na cidade do Recife tem deixado em alerta a população e a polícia. A sociedade tem modificado seus hábitos com receio da violência, enquanto a polícia tenta prevenir e combater novos crimes de acordo com as ocorrências relatadas às delegacias da cidade. Este estudo tem por objetivo encontrar e analisar possíveis relações existentes entre variáveis representativas de ocorrências de crimes das delegacias da cidade do Recife, para auxiliar na análise e na criação de estratégias para melhor prevenção de crimes. Inicialmente foi realizado um pré-processamento das bases por meio das ferramentas Pentaho Kettle e R, em seguida foram realizadas estatísticas descritivas e por fim foram aplicados os testes de hipótese de Qui-quadrado e o de Fisher.

Palavras-Chave: *Ocorrências de crimes em Recife; Qui-quadrado; Fisher; R; Kettle;*

Abstract

The increase in crime in recent years in the city of Recife has left the population and the police alert. Society has modified its habits for fear of violence, while police try to prevent and fight new crimes according to the incidents reported to city police stations. This study aims to find and analyze possible relationships between representative variables of crime occurrences in the city of Recife, to assist in the analysis and creation of strategies for better crime prevention. Initially, a pre-processing of the bases was carried out using the Pentaho Kettle and R tools, followed by descriptive statistics and finally the Chi-square and Fisher hypothesis tests were applied.

Key-words: *Occurrences of crimes in Recife Chi-square; Fisher; R; Kettle;*

1 Introdução

Segundo George Kelling [1], um dos autores do artigo "Teoria da Janela Quebrada" e do livro "Consertando janelas quebradas", se um prédio tem uma janela quebrada, é melhor consertá-la logo, antes que um vândalo quebre todas as outras. Nesta perspectiva acredita-se que crimes pequenos podem atrair crimes mais graves. Essa ideia tem orientado programas nos Estados Unidos, onde o controle das pequenas infrações tem ajudado a prevenir crimes mais graves[1].

O crescente aumento da violência no estado de Pernambuco aponta para uma urgência na criação de soluções estratégicas, desafiadoras para a gestão da segurança pública. Algumas cidades pelo mundo têm adotado medidas proativas de controle e prevenção a crimes com a ajuda da gestão de informações registradas em sistema de informação, apresentando significativo impacto na redução da taxa de criminalidade. Um exemplo disso é o CompStat, utilizado pela prefeitura de Nova Iorque, que pode ser definido como uma técnica de gestão de processos orientada por metas, baseada na utilização de tecnologia computacional, estratégia operacional e responsabilidade gerencial para estruturar o modo como o departamento de polícia fornece serviços voltados para o controle da criminalidade [2]. Logo, o Compstat não é um método de mapeamento do crime por si só, mas um instrumento de gestão baseado em informação gerada pela tecnologia da informação. Sua utilização teve impactos significativos na experiência de Nova Iorque nos anos 90 e foi expandida para outros departamentos de polícia, dentro e fora dos Estados Unidos. O sistema Compstat caracteriza-se pela utilização de dados criminais com o mapeamento geográfico das áreas a serem policiadas. Esse instrumento está associado a técnicas como o mapeamento de "zonas quentes" (*hotspot mapping*), que são definidas a partir da utilização das estatísticas criminais [3].

Exemplos como este nos encorajou a desenvolver este trabalho, onde buscamos analisar a relação existente entre as variáveis de ocorrência de crimes fomentando, assim, possíveis orientações à Segurança Pública no estado de Pernambuco, através da promoção de estratégias adequadas para prevenção e combate a crimes.

2 Trabalhos relacionados

Vários trabalhos vêm sendo desenvolvidos nesta temática nos últimos anos, um estudo para identificar o *modus operandi* do crime de roubo a transeuntes em Belém do Pará foi realizado em 2015 [4] e, nele, a partir de dados coletados nos registros dos boletins de ocorrência, foi possível realizar uma análise descritiva, identificando o modo como os assaltantes agem e o meio empregado para a locomoção no espaço geográfico por ocasião do cometimento do crime. Além disso, pôde-se analisar a variável temporal (horário e dia da semana) e o número de autores envolvidos.

No estudo [4], foi feito um levantamento dos dados de 2011, 2012 e 2013 para observar se houve alguma alteração ao longo dos anos no comportamento do cometimento do crime de roubo, em termos da faixa de hora predominante. Em 2011, o delito ocorria sobretudo das 12h00 às 17h59; já em 2012 e 2013, a tônica se dava entre 18h00 e 23h59; e o ano de 2013 caracterizou-se pelo distanciamento entre as duas faixas de hora, fugindo do padrão dos anos anteriores. Este conjunto de informações revela a relação entre a luz do dia e a escuridão na preferência do delinquente no cometimento do crime de roubo.

O mesmo estudo também mostrou uma nítida distribuição da ocorrência do roubo ao longo dos dias da semana: em 2011 foi o sábado; em 2012, a sexta-feira; e em 2013, a quarta-feira [4].

Um outro estudo semelhante [5], realizado com variáveis de criminalidade em Teresina, Piauí, apontou os locais e horários onde há o maior índice de criminalidade, quais os principais crimes cometidos, causas e instrumentos utilizados. Também traça um perfil das vítimas e criminosos. Nesse estudo foi possível verificar a diferença entre os crimes cometidos contra homens e mulheres. Enquanto os homens são as principais vítimas de homicídios dolosos, as mulheres são as principais atingidas por lesões dolosas [5].

3 Referencial Teórico

Os métodos utilizados para analisar dados em busca de relações existentes entre as variáveis diferem de acordo com o tipo de dado coletado. Neste trabalho, as variáveis analisadas são

categóricas, e, para descobrir possíveis dependências entre elas, foram utilizados dois tipos de testes de hipótese, o teste do *qui-quadrado* e o teste exato de Fisher.

Um teste de hipótese é um procedimento estatístico que permite tomar uma decisão (aceitar ou rejeitar a hipótese nula H_0) entre duas ou mais hipóteses (hipótese nula H_0 ou hipótese alternativa H_1), utilizando os dados observados de um determinado experimento [6]. Uma hipótese nula geralmente afirma que não existe relação entre dois fenômenos medidos, enquanto que a hipótese alternativa é contraditória à primeira.

3.1 Teste de Qui quadrado

O teste do *qui quadrado* pode não ser tão preciso quando os valores das amostras são muito pequenos, sendo, nesse caso, recomendado utilizar testes exatos. Por esse motivo, duas abordagens foram utilizadas neste trabalho, o teste de aproximação de *quiquadrado* e o teste exato de Fisher.

O teste do *qui quadrado* compara as frequências observadas com as frequências esperadas da amostra. As frequências observadas são obtidas diretamente dos dados das amostras, enquanto que as frequências esperadas são calculadas matematicamente a partir destas.

É preciso obter as duas estatísticas: o χ^2 calculado e o χ^2 tabelado. O χ^2 calculado é obtido através da equação (1). Já o χ^2 tabelado depende do número de graus de liberdade e do nível de significância adotado.

O teste é feito sobre uma tabela de contingência formada por (m) linhas e (n) colunas e o grau de liberdade é dado pelo produto de (m-1) x (n-1). O nível de significância representado por α é definido previamente, mas costuma-se adotar, por convenção, o valor de $\alpha = 0,05$. O nível de significância representa a máxima probabilidade de erro (valor-p) que se tem ao rejeitar uma hipótese. Em outras palavras, isso significa que os resultados experimentais que atingem esse nível de significância têm, no máximo, 5% de chance de ser resultado do mero acaso.

O teste calcula a relação: quadrado da diferença entre as frequências obtida (O_i) e esperada (O_e) em cada casa da tabela de contingência, dividido pela frequência esperada, e soma esses quadrados. O

resultado dessa soma é o valor do χ^2 encontrado (1).

$$\chi^2 = \sum_{i=1}^n \frac{(o_i - e_i)^2}{e_i} \quad (1)$$

Considerando o grau de liberdade calculado e o nível de significância pré-definido, consultando-se a tabela da distribuição de χ^2 , encontramos o valor crítico de χ^2 e o valor p. Uma vez que o valor de χ^2 encontrado é maior ou igual ao valor crítico, e, conseqüentemente, o valor-p é menor do que o nível de significância adotado, rejeita-se a hipótese nula. Ou seja, quando o valor obtido no teste estatístico excede o correspondente do valor crítico estabelecido, a hipótese nula deve ser rejeitada e aceita-se a associação entre as variáveis. Caso contrário, a hipótese nula não pode ser descartada.

3.2 Teste Exato de Fisher

O teste exato de Fisher é mais preciso do que o teste do *qui quadrado* quando os números esperados são pequenos. John H. McDonald recomenda que seja utilizado o teste exato de Fisher quando o tamanho total da amostra for inferior a 1000, e o *qui quadrado* seja utilizado para tamanhos de amostra maiores [7].

O Teste Exato de Fisher se caracteriza por fornecer diretamente o valor-p, sem o uso de uma estatística intermediária.

O teste também pode ser realizado no caso de amostras grandes, porém, como envolve cálculos maiores e mais complexos que utilizam fatoriais, o que pode conduzir a números excessivamente elevados, necessita de maiores recursos do computador.

O teste exato de Fisher é baseado na distribuição hipergeométrica (2):

$$P = \frac{(A + B)! (C + D)! (A + C)! (B + D)!}{N! A! B! C! D!} \quad (2)$$

A probabilidade calculada será igual ao produto dos fatoriais dos totais marginais pelo fatorial do total geral multiplicado pelo inverso do produto dos fatoriais dos valores observados em cada classe.

4 Metodologia

A análise realizada neste artigo utilizou cinco bases de dados da Secretaria de Defesa Social do Estado de Pernambuco¹. Essas bases contêm ocorrências que foram registradas nas delegacias do estado no período de 01 de janeiro de 2017 a 30 de setembro de 2017, isto é, durante um período de nove meses, resultando em um total de 52.873 instâncias. São elas: CVLI (Crimes Violentos Letais Intencionais), CVP (Crimes Violentos contra o Patrimônio), Furto, Furto de Veículo e Roubo de Veículo.

O processo de análise envolveu:

- Verificar se existe relação significativa entre o dia da semana ser dia útil ou final de semana e a natureza do crime.
- Verificar se existe relação significativa entre o período do dia e a natureza do crime.
- Verificar se existe relação entre o trimestre e a natureza do crime.
- Verificar se existe relação significativa entre a área onde ocorreu o crime e a natureza do mesmo.
- Verificar se existe relação significativa entre a faixa etária da vítima e a natureza do crime para os casos de crimes violentos letais intencionais.
- Verificar se existe relação significativa entre o sexo da vítima e a natureza do crime para os casos de crimes violentos letais intencionais.

4.1 Pré-processamento

Para dar início à pesquisa, foi necessário realizar um trabalho de extração, transformação e carga (ETL – *Extract, Transform and Load*) nas bases para que a integração entre elas e suas análises fossem possíveis.

O processo de ETL envolveu a extração dos dados das bases obtidas, a transformação dos mesmos, incluindo higienização, padronização e categorização de variáveis, e, por fim, a integração das bases em uma base única. Para esta última etapa do ETL,

foram selecionadas as variáveis mais significativas e em comum entre as bases.

As bases foram extraídas dentro da ferramenta Kettle, também conhecida como PDI (*Pentaho Data Integration*²). Nela, cada base passou individualmente por uma transformação. Uma variável representando um contador para os registros foi criada, além de mais outras duas variáveis categóricas, ambas derivadas da variável data. Uma delas representando o mês, e a outra, o dia da semana. As mesmas transformações foram reproduzidas para cada base.

Em seguida, as saídas das bases tratadas no Kettle continuaram o pré-processamento com mais algumas categorias sendo criadas. Dessa vez, utilizando o R³, um ambiente de software livre para computação estatística e gráficos. As categorias trimestre e semana – indicando quando se trata de dia útil ou final de semana, mais especificamente – foram criadas, derivadas das variáveis recém-criadas mês e dia da semana, respectivamente. Também foi criada – apenas para a base CVLI – a categoria faixa etária, derivada da variável idade. A base CVLI era a única que continha as variáveis idade e gênero.

O pré-processamento no R gerou novas saídas, que foram utilizadas novamente no Kettle para realizar a integração.

A integração envolveu a junção de todas as cinco bases utilizando apenas as suas variáveis em comum, resultando em uma base única, conforme Figura 1.

¹<http://www.sds.pe.gov.br/>

² <http://www.pentaho.com/product/data-integration>

³ <https://www.r-project.org/>

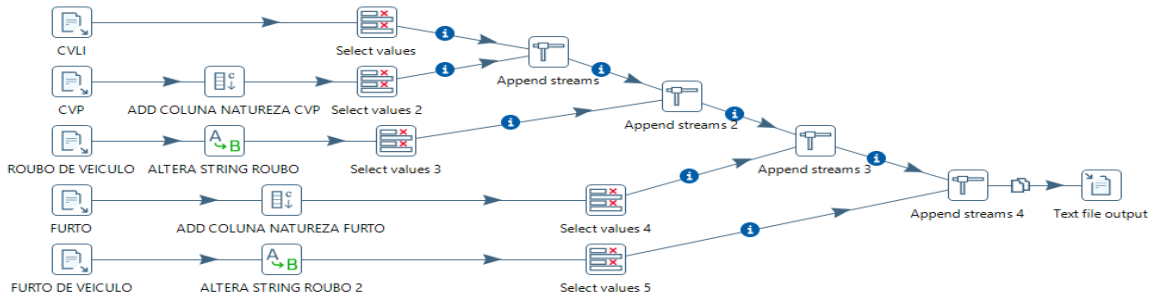


Figura 1 – Pré-processamento e integração das bases no PDI

4.2 Dicionário de dados

A base integrada, resultante da fase de ETL, contém nove variáveis, como mostra o Quadro 1:

Quadro 1 – Dicionário de Dados das bases integradas

AREA	Área onde ocorreu o crime, representada pela AIS – Área Integrada de Segurança, divisão territorial feita em Pernambuco para acompanhamento de ações e resultados. Existem, ao todo, 26 AIS em Pernambuco, mas as bases só envolvem as da cidade do Recife: AIS 1 – Santo Amaro, Boa Vista, Ilha Joana Bezerra e São José. AIS 2 – Espinheiro, Iputinga, Cordeiro, Madalena, Água Fria e Campo Grande. AIS 3 – Boa Viagem, Pina, Imbiribeira, Ibura, Brasília Teimosa AIS 4 – Várzea, Curado, Jardim São Paulo, Torrões e Afogados. AIS 5 – Apipucos, Guabiraba, Brejo da Guabiraba, Passarinho, Dois Unidos, Vasco da Gama e Alto do Mandú.
DATA	Informa a data e a hora em que ocorreu o crime.
DIA_SEMANA	Variável categórica derivada da variável DATA. Informa o dia da semana quando ocorreu o crime.
DIA_ÚTIL	Variável categórica derivada da variável DIA_SEMANA. Informa se o dia em que ocorreu o crime era dia útil ou final de semana.
MÊS	Variável categórica derivada da variável DATA. Informa o mês em que ocorreu o crime.
NATUREZA	Informa a natureza do crime (homicídio, roubo, latrocínio, etc.).
PERÍODO	Variável categórica derivada da variável DATA. Informa o período do dia em que o crime ocorreu (manhã, tarde, noite ou madrugada).
TRIMESTRE	Variável categórica derivada da variável MÊS. Informa o trimestre em que o crime ocorreu.
TOTAL	Contador de registros.

A base CVLI, além das variáveis mostradas na tabela acima, também contém três variáveis exclusivas, conforme o Quadro 2:

Quadro 2 – Dicionário de Dados da base CVLI

SEXO	Variável categórica que representa o sexo da vítima de crimes violentos letais intencionais.
IDADE	Informa a idade da vítima.
FAIXA_ETARIA	Variável categórica derivada da variável IDADE que representa a faixa etária da vítima, onde "Adolescente" representa a faixa de 12 a 18 anos, "Jovem Adulto" representa a faixa de 19 a 30 anos, "Adulto" representa a faixa de 31 a 49 anos, e "Idoso", a partir de 60 anos.

4.3 Análises Descritivas

4.3.1 Tabelas de Frequência

Inicialmente, utilizando a ferramenta R, foram construídas individualmente tabelas de frequência para todas as nove variáveis da base integrada e para as duas variáveis categóricas exclusivas da base CVLI.

A tabela de frequência auxilia na orientação da pesquisa, mostrando a distribuição de frequência das variáveis. A frequência de uma variável é o número de ocorrências ou repetições dessa variável.

Utilizando as funções *table()* e *prop.table()*, uma tabela de frequência foi construída para cada variável, assim como uma tabela de frequência relativa.

A tabela de frequência da variável que representa o sexo da vítima de crimes violentos letais intencionais, por exemplo, apontou para um

percentual maior de vítimas do sexo masculino, com 92,19% de incidência.

Ainda na base de crimes violentos letais intencionais, a tabela de frequência construída em cima da variável que representa a faixa etária da vítima mostrou que as maiores vítimas são jovens adultos, com uma incidência de 55,41%, seguidos de adultos e adolescentes, com 30,45% e 13,48% respectivamente.

Todas as distribuições citadas até aqui foram construídas em cima da base CVLI. A seguir, serão apresentadas as tabelas de distribuição de frequências em cima da integração de todas as cinco bases (CVLI, CVP, ROUBO DE VEÍCULO, FURTO E FURTO DE VEÍCULO). A Tabela1, referente à natureza do crime, mostra que há um índice maior de casos de crimes violentos ao patrimônio, e o segundo lugar ficando com crimes de furto.

Tabela 1 - Distribuição de frequência relativa referente à natureza do crime

NATUREZA	FREQUÊNCIA
CVP	56,05%
FURTO	34,72%
ROUBO DE VEÍCULO	5,65%
FURTO DE VEÍCULO	2,45%
HOMICÍDIO	1,11%
LATROCÍNIO	0,03%
LESÕES CORPORAIS SEGUIDA DE MORTE	0,00%

Como o período da base vai de janeiro a setembro do mesmo ano, a variável trimestre, sozinha, não indica algo muito significativo, já a tabela de frequência relativa da variável referente ao mês da ocorrência apresentou uma distribuição praticamente uniforme para todos os meses, de forma que, sozinha, na base integrada, ela não tem grande relevância.

Tabela 2 - Distribuição de Frequência relativa referente ao dia da semana

DIA DA SEMANA	FREQUÊNCIA
SEXTA-FEIRA	15,65%
QUARTA-FEIRA	15,05%
TERÇA-FEIRA	14,74%
QUINTA-FEIRA	14,70%
SEGUNDA-FEIRA	14,47%
SÁBADO	14,04%
DOMINGO	11,34%

Pode-se observar na Tabela2 que ocorreram mais crimes nas sextas-feiras e nas quartas-feiras, com o

domingo apresentando o menor percentual de ocorrências, coincidindo com o estudo realizado em Belém do Pará [4], em que as sextas-feiras e as quartas-feiras apresentaram mais incidentes de crimes de roubo.

Consequentemente, a tabela de frequências da variável que categorizou o dia da semana em dia útil ou final de semana ratificou o resultado acima, revelando que os crimes ocorreram com mais frequência nos dias úteis do que nos finais de semana.

O estudo também identificou que, assim como em Teresina [5], crimes de violência contra o patrimônio e homicídio são mais recorrentes no período da noite. Neste estudo, encontramos um resultado de 40,45% de incidência de crimes de CVP durante a noite e 8,81% no período da madrugada. Para homicídio, houve uma incidência de 42,56% no período da noite e 17,42% durante a madrugada, enquanto que o estudo realizado em Teresina apresentou uma incidência de 49% para crimes de furto e 55% para crimes de roubo - ambos são categorizados como crimes de violência ao patrimônio (CVP) - e uma incidência de 76% para crimes de homicídio, ambos no período de 18h às 03h (noite e madrugada).

Tabela 3 - Distribuição de frequência relativa referente ao período do dia.

PERÍODO	FREQUÊNCIA
NOITE	34,26%
TARDE	28,12%
MANHÃ	23,81%
MADRUGADA	13,81%

A distribuição de frequências relativas por período, conforme apresentado na Tabela 5, mostra que houve uma maior incidência de crimes gerais durante a noite.

Tabela 4 - Distribuição de frequência relativa referente à área da ocorrência.

ÁREA	FREQUÊNCIA
1	24,29
3	23,25
4	20,23
2	19,05
5	12,20
99	0,99

Por fim, a distribuição por área (Tabela 4) indicou que a AIS 1 e a AIS3 apresentaram um percentual

maior de crimes. A AIS 1 é composta dos seguintes bairros: Santo Amaro, Boa Vista, Ilha Joana Bezerra e São José, e a AIS 3 é formada pelos bairros Boa Viagem, Pina, Imbiribeira, Ibura e Brasília Teimosa. A área 99 corresponde aos valores ausentes.

Este resultado pode incentivar um outro estudo que envolva o levantamento das características urbanas dos bairros do estado para detectar se esses bairros mais violentos possuem características que indiquem degradação urbana, negligência ou desmazelo. As características socioeconômicas dos bairros também devem ser levadas em consideração, para identificar se existe um alto índice de desemprego, pobreza, baixa escolaridade e outras características que podem ser reveladas potenciais para desencadear a violência.

5 Testes de relação

Analisando o Gráfico 1, é possível ratificar que existe uma incidência maior de crimes de violência ao patrimônio durante o período da noite.

Aplicando o teste de *qui quadrado* na tabela de contingência construída a partir dessas variáveis, "natureza" e "período", utilizando a função *chisq.test()* do R, obtivemos o valor do $X^2 = 3348.9$, com o grau de liberdade = 12 e o valor $p < 0,0000000000000022$, isto é, muito abaixo do nível de significância 0,05. Portanto, há um forte indício de que podemos rejeitar a hipótese nula e concluir que existe uma relação entre a natureza do crime e o período do dia em que ele ocorre.

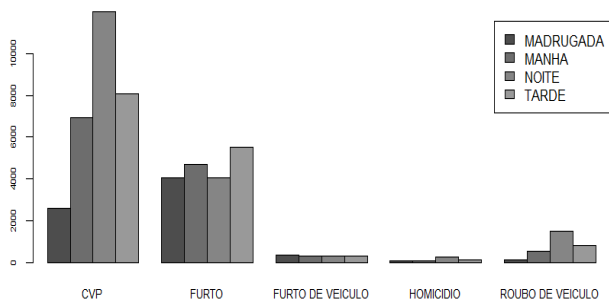


Gráfico 1 – Incidência de crime por natureza e por período

Para identificar se a possível relação que parece existir entre as variáveis "natureza" e "dia_util" no Gráfico 2, também aplicamos o teste de *qui-*

quadrado na tabela de contingência das duas variáveis, obtendo um resultado de 233.51 para o X^2 , com grau de liberdade = 4 e valor $p \approx 0,000$, indicando que podemos rejeitar a hipótese nula e constatar que existe relação entre essas duas variáveis.

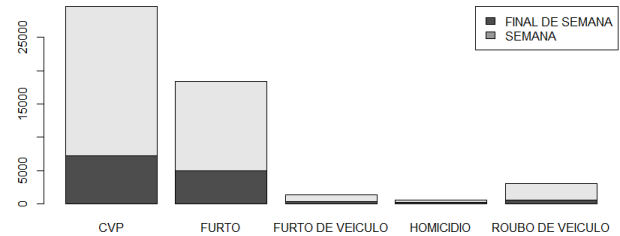


Gráfico 2 – Incidência de crime por natureza e por dia de semana (dia útil ou final de semana)

Ainda utilizando o teste de *qui quadrado*, vimos que a hipótese nula, que afirma não existir relação entre as variáveis que representam a natureza do crime e sexo da vítima não pode ser descartada, visto que o resultado do X^2 foi de 0.23795, com um valor $p = 0.9935$, acima do nível de significância estabelecido, de 0,05. Portanto, não podemos afirmar que essas variáveis são dependentes.

Por meio do teste exato de *Fisher*, pôde-se concluir que a distribuição do número de homicídios difere ($p = 0,01212$) entre as faixas etárias das vítimas e o sexo, sendo maior para jovens adultos e adultos do sexo masculino, com 53% e 26,36% respectivamente. O Gráfico 3 ilustra bem essa diferença. Este resultado converge com o observado no estudo realizado em Teresina, Piauí.

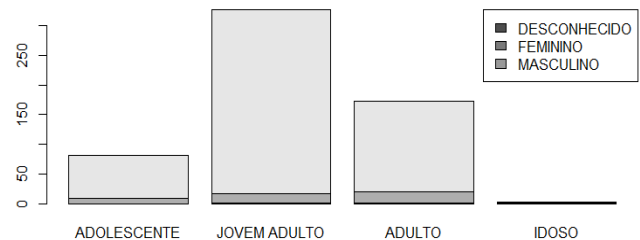


Gráfico 3 – Incidência de homicídios por faixa etária e por sexo

Analisando também a relação entre o período do dia e o dia da semana das ocorrências de crimes de homicídio isoladamente, vimos que, assim como o estudo em Teresina também apontou, eles seguem

um padrão temporal, com uma maior incidência à noite e nos finais de semana, com o sábado à noite apresentando o maior número de ocorrências, 13,67%. Aplicando o teste de *qui quadrado*, pudemos constatar que existe uma relação entre essas duas variáveis (período do dia e dia da semana). O resultado do teste apresentou um $X^2 = 39.876$ e um valor $p = 0.00217$.

O Gráfico 4 mostra uma distribuição dos crimes por área e natureza, indicando que crimes contra o patrimônio (roubos e furtos) ocorrem mais na AIS 1, enquanto que crimes relacionados a veículos (roubos e furtos) ocorrem mais na AIS 4, formada pelos bairros Várzea, Curado, Jardim São Paulo, Torrões e Afogados, e AIS 3, formada pelos bairros Boa Viagem, Pina, Imbiribeira, Ibura e Brasília Teimosa. Isso foi ratificado através do teste do *qui-quadrado*, que resultou em um p valor menor do que 0,05.

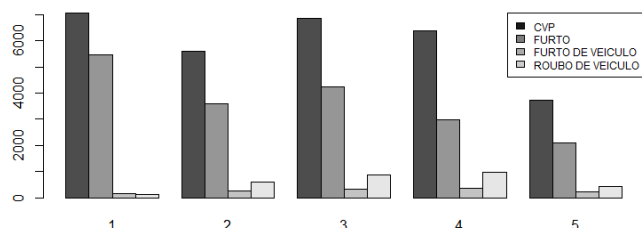


Gráfico 4 – Incidência de crimes de CVP, furto ou roubo de veículo por AIS.

Tabela 5 – Distribuição de Frequência Natureza X AIS

	CVP	FURTO	FURTO DE VEICULO	ROUBO DE VEICULO
1	13,35%	10,35%	0,28%	0,19%
2	10,62%	6,77%	0,48%	1,14%
3	12,99%	8,01%	0,61%	1,64%
4	12,05%	5,64%	0,68%	1,85%
5	7,03%	3,94%	0,40%	0,82%

A Tabela 6 destrincha os crimes relacionados apenas a veículos (roubos e furtos) por período do dia e dia da semana. Nela pode-se observar que existe maior incidência nas noites de sexta e quarta-feira. Aplicando o teste de *qui quadrado*, obtivemos o resultado de 0.4866 para o valor p.

Tabela 6 – Distribuição de Frequência de roubos e furtos de veículos por período e dia da semana

	Madrugada	Manhã	Noite	Tarde
DOM	1,40%	1,87%	3,60%	2,27%
SEG	1,78%	2,69%	5,65%	3,79%
TER	1,82%	3,18%	6,61%	4,30%
QUA	1,89%	2,80%	7,31%	4,58%
QUI	1,73%	3,39%	6,82%	3,74%
SEX	1,85%	3,41%	7,34%	4,18%
SAB	1,61%	2,29%	4,77%	3,34%

A Tabela 7 representa uma síntese de todas as variáveis e suas respectivas significâncias encontradas através dos testes de relação.

Tabela 7 – Resultado do valor p nos testes de relação entre as variáveis

	natureza	periodo	dia_semana	area	sexo	faixa etaria	dia da semana (veículo)
natureza		2,2E-15		<2,2e-16			
periodo	<2,2e-16		0,00217				
dia_semana	0,9935	0,00217					
area	<2,2e-16						
sexo						0,01212	
faixa etaria					0,01212		
Período (veículo)							0,4866

6 Análise e Discussões

Vimos nos resultados das análises que o período de maior ocorrência de crimes em geral é o da noite. Isso acontece possivelmente pela visibilidade ser menor com a pouca iluminação que existe à noite e por ter menos movimento nas ruas. Esse tipo de informação merece ser identificado para que haja uma melhor e mais estratégica distribuição do policiamento por horário. Outra possível solução também seria a readequação da iluminação da cidade de acordo com o tipo de movimentação de cada bairro.

Vimos também nos resultados das análises que alguns bairros são mais suscetíveis a crimes de roubos e furtos de veículos. Com poucos dados suficientes para inferir a causa dessa estatística, podemos apenas conjecturar que esses números podem estar relacionados ao fato de se tratarem de bairros residenciais compostos, em sua maioria, de casas, implicando em muitos carros estacionados nas ruas e moradores que precisam descer dos carros para abrir os portões de suas residências, ficando mais vulnerável e se tornando um alvo fácil para os criminosos.

A preferência por alvos de determinada faixa etária por parte dos criminosos poderia ter sua causa descoberta, caso existissem algumas outras variáveis, tais como condição social da vítima, a localização exata do crime e o sexo do criminoso. Uma base de dados mais completa e mais integrada poderia auxiliar nessa descoberta.

Vimos também que o tipo de ocorrência de crimes varia de acordo com a área. A base utilizada nesse estudo tem limitações de quantidade de variáveis a um período de tempo pequeno. Porém, as relações encontradas aqui podem ser utilizadas para alimentar estudos mais detalhados e mais aprofundados, e também para um planejamento estratégico da segurança segmentado por AIS, onde cada AIS teria um policiamento adequado.

Ainda em relação à distribuição de ocorrências por AIS, este estudo nos possibilitou observar que a AIS 1 possui maior ocorrência em todos os tipos de crime, e talvez isso comprove a teoria das janelas quebradas, sugerindo que crimes menores atraíam crimes mais graves. A partir da prevenção de pequenos crimes pode ser possível evitar que a área vá se transformando em um espaço cada vez mais violento. Para isso é necessário identificar os pontos

e quais os crimes ali praticados, como o mostrado nesse estudo.

7 Considerações Finais

Com esses levantamentos, é possível elaborar um diagnóstico melhor do problema da violência no estado e começar a ter uma visão estratégica para solucioná-lo. A mudança do tipo de policiamento em determinadas áreas, ou o aumento do efetivo policial em determinados dias e horários, de acordo com os resultados dessas estatísticas, pode ser uma das soluções, juntamente com a instalação de câmeras de segurança em pontos estratégicos.

8 Contribuições Futuras

Cabe fazer algumas considerações relevantes para que haja uma melhor gestão das informações sobre os dados das ocorrências de crimes e delitos no Estado de Pernambuco.

Este estudo pode fomentar outros estudos que devem ser realizados envolvendo mais variáveis além das mostradas aqui, mas que não foram fornecidas pela Secretaria de Defesa Social por razões de sigilo da informação, como variáveis geográficas e variáveis relacionadas a dados urbanos, buscando possíveis relações entre tipos de crimes e características urbanas do local de ocorrência.

Como trabalhos futuros serão utilizadas técnicas de mineração de dados, com o objetivo de encontrar padrões nos crimes relatados, além da criação de uma ferramenta web para melhor visualização e análise dos resultados.

Este estudo e possíveis extensões do mesmo podem orientar a Segurança Pública na promoção de estratégias adequadas para a redução da criminalidade no Estado. Também podem fornecer subsídios para serem feitos posteriormente monitoramentos para identificar se houve real redução na criminalidade após a implementação de novos controles.

Agradecimentos

À Secretaria de Defesa Social de Pernambuco, pelo fornecimento das bases de dados utilizadas neste trabalho.

Referências

[1] WILSON, James Q.; KELLING, George L. BROKEN WINDOWS: The police and neighborhood safety. **The Atlantic**, Mar. 1982. Disponível em: <<https://www.theatlantic.com/magazine/archive/1982/03/broken-windows/304465/>>. Acesso em: 30 set. 2017.

[2] WALSH, William F. Compstat: An analysis of an emerging police managerial paradigm. Policing: An International. **Journal of Police Strategies & Management**, v. 24, n.3, p. 347, 2001.

[3] RATCLIFFE, J.erry H. Crime mapping and the training needs of law enforcement. **European Journal on Criminal Policy and Research**, v. 10, n.1, p.10-65, 2004. Disponível em: <<https://doi.org/10.1023/B:CRIM.0000037550.40559.1c>>. Acesso em: 1 fev. 2018.

[4] CAVALCANTE, Lucidéia S.; ALMEIDA, Silvia dos S.; ARAÚJO, Adrilayne dos R. O *Modus operandi* do crime de roubo a transeuntes em Belém. **Planejamento e Políticas Públicas**, n. 47, jul/dez. 2016. Disponível em: <<http://www.ipea.gov.br/ppp/index.php/PPP/article/view/614>>. Acesso em: 30 jan. 2018

[5] SANTOS, Laura Castro de Carvalho dos. Violência e criminalidade: Um estudo dos dados existentes em Teresina - PI. **Âmbito Jurídico**, Rio Grande, v. 15, n. 99, abr 2012. Disponível em: <http://www.ambito-juridico.com.br/site/?n_link=revista_artigos_leitura&artigo_id=11448>. Acesso em: mar 2018.

[6] DÁVILA, Víctor Hugo L. **Teste de Hipóteses**. Instituto de Matemática, Estatística e Computação, UNICAMP. Disponível em: <https://www.ime.unicamp.br/~hlauchos/Inferencia_Hipo1.pdf> p. 3. Acesso em: 15 mar. 2018.

[7] MCDONALD, John H. **Handbook of Biological Statistics**. 3rd ed. Baltimore: Sparky House Publishing, 2014. Disponível em: <<http://www.biostathandbook.com/HandbookBioStatThird.pdf>>. Acesso em: mar. 2018.