

Análise de Crimes em Curitiba

Equipe

Nome: CSI Curitibanos

Repositório GitLab da equipe: <https://gitlab.com/thiagorbernardo/IntroCD>

Membros:

- | | | |
|--|-----------------|---------------------------|
| • João Felipe Sarggin Machado,
joaomachado@alunos.utfpr.edu.br , | 1905023,
EC, | jfsmachado,
UTFPR |
| • João Lucas Mizuguchi Ferreira,
joaolucasferreira@alunos.utfpr.edu.br , | 2135965,
EC, | jlmizug,
UTFPR |
| • Pedro Sodré dos Santos,
psantos.2000@alunos.utfpr.edu.br , | 2126745,
EC, | pedro_sodre,
UTFPR |
| • Thiago Ramos Bernardo,
thiagobernardo@alunos.utfpr.edu.br , | 2126753,
EC, | thiagorbernardo,
UTFPR |

Introdução

Sabe-se que medidas de segurança pública são muito importantes para termos uma vida mais tranquila e pacífica, portanto esta pesquisa se propõe a verificar padrões nos crimes cometidos em Curitiba, a cidade mais populosa do Paraná. Estes padrões foram focados nos crimes nos bairros, a semelhança entre os horários de determinados tipos de crimes cometidos e também a época do ano/dia da semana em que ocorrem mais/menos crimes.

Para isso, foram utilizados os parâmetros dos bairros (renda, taxa de analfabetismo) e também a distância em relação ao centro de cada bairro, porque foram fatores considerados relevantes e que influenciam na quantidade de crimes em cada bairro. Consideramos outros fatores também relevantes, porém que não haviam dados facilmente disponíveis.

Outra coisa também realizada, foi identificar padrões nos horários e tipos de crimes que ocorriam naqueles determinados períodos do dia, procurando picos semelhantes em tipos de crimes distintos, porém relativos entre si.

Além disso, também foi verificado o período do ano em que ocorria a maior quantidade de crimes, além de épocas como fim de semana e feriados, e se isso influencia na quantidade de crimes cometidos ao longo do ano.

Em resumo, procuramos responder 3 perguntas de pesquisa:

- A variável renda, a taxa de analfabetização e distância em relação ao centro influenciam na quantidade de crimes em determinados bairros de Curitiba?
- Como o horário do crime influencia no tipo de crime cometido em Curitiba?

- Como o período do ano/ dia da semana influencia na quantidade total de crimes na cidade de Curitiba?

Em relação à primeira pergunta, apesar de ser encontrada alguma influência desses fatores em relação a quantidade de crimes, ele não foi tão elevada quanto se esperava ($R^2 = 0.121$, com análise com o centro; e $R^2 = 0.196$, com análise sem o centro). Na segunda questão, foram encontrados crimes que possuem muita semelhança em relação ao seus horários cometidos, justamente pela naturalidade que estes tipos de crimes ocorrem. Já na última questão, pode-se constatar que alguns tipos de crime sobem sua média nos finais de semana e outros diminuem. O mesmo acontece para os meses do ano, onde alguns crimes diminuem nos períodos de férias e outros aumentam.

Processamento de Dados

Primeiramente, o foco seria a análise da relação entre os parâmetros/índices dos bairros de Curitiba, com os crimes. Para isso, foram pegos estes dados através do site <http://infocuritiba.ippuc.org.br/>, porém, existem muitos dados que estão indisponíveis, então métricas e índices como quantidade de empregos, PIB, PIB per capita, IDH, entre outros apareceram como a string: “...” (vazio) no dataset, além do dataset estar com os parâmetros nas linhas ao invés das colunas, então foi feita uma transposição, as colunas com valores nulos (“...”) foram excluídas, e também, a coluna “ANO”, que, como só havia dados de 2010 disponíveis, era uma coluna com único valor, e por isso foi excluída também, e vamos induzir que não houve grandes mudanças nestes parâmetros na última década. Dessa forma o dataset sofreu uma transformação como se pode ver na imagem abaixo:

	Variável	Ano	Abranches	Água Verde	Alib	Alto Boqueirão	Alto da Glória	Alto da Rua XV	Alubia	Augusta	...	Sítio Cercado	Taboão	Tarumã	Tatuquara	Tingui	Uberaba	Umará	Vila Izabel	Vista Alegre	Xaxim
0	População Censitária - Total (Habitante)	2010.0	13.189	51.425	11.506	53.671	5.548	8.531	15.935	6.598	...	115.525	3.396	8.072	52.780	12.319	72.056	18.730	11.610	11.199	57.182
1	Classe de Rendimento - Pessoas de 10 Anos ou Mais	2010.0	3.533	12.044	2.920	14.447	1.318	1.936	4.097	1.798	...	31.281	763	2.155	14.962	3.065	19.815	5.421	2.876	2.947	15.940
2	Rendimento Nominal Mensal Médio das Pessoas de...	2010.0	1.179,72	3.170,07	3.054,92	867,36	3.054,18	2.854,97	1.203,26	731,64	...	726,28	1.381,97	2.100,23	590,06	1.370,39	1.101,82	762,65	2.467,22	2.036,11	1.040,56
3	Rendimento Nominal Mensal Mediano das Pessoas de...	2010.0	650,00	1.511,00	1.500,00	650,00	2.000,00	1.500,00	700,00	520,00	...	600,00	700,00	1.000,00	510,00	800,00	600,00	600,00	1.500,00	810,00	700,00
4	Empregos - Total (Habitante)	2010.0
5	PIB (R\$1.000,00) (Real)	2010.0
6	PIB - Per Capita (R\$1,00) (Real)	2010.0
7	Alfabetização - Taxa de Analfabetismo - Total ...	2010.0	3,71	0,82	0,87	3,36	0,84	0,99	3,50	4,85	...	4,44	4,10	2,24	6,11	1,97	3,51	4,37	1,15	1,87	3,07
8	IDHM - Geral (Unidade)	2010.0
9	IDHM - Renda (Unidade)	2010.0
10	IDHM - Longevidade (Unidade)	2010.0
11	IDHM - Educação (Unidade)	2010.0
12	IFDM - Geral (Unidade)	2010.0
13	IFDM - Educação (Unidade)	2010.0
14	IFDM - Saúde (Unidade)	2010.0

Figura 1. Raw Dataset (antes da transformação)

	População Censitária - Total (Habitante)	Classe de Rendimento - Pessoas de 10 Anos ou Mais de 10 Anos - Total (Habitante)	Rendimento Nominal Mensal Médio das Pessoas de 10 Anos ou Mais de 10 Anos - Total (Real)	Rendimento Nominal Mensal Mediano das Pessoas de 10 Anos ou Mais de 10 Anos - Total (Real)	Alfabetização - Taxa de Analfabetismo - Total (Porcentual)
ABRANCHES	13.189	3.533	1.179,72	650,00	3,71
ÁGUA VERDE	51.425	12.044	3.170,07	1.511,00	0,82
ALIB	11.506	2.920	3.054,92	1.500,00	0,87
ALTO BOQUEIRÃO	53.671	14.447	867,36	650,00	3,36
ALTO DA GLÓRIA	5.548	1.318	3.054,18	2.000,00	0,84
...
UBERABA	72.056	19.815	1.101,82	600,00	3,51
UMARÁ	18.730	5.421	762,65	600,00	4,37
VILA IZABEL	11.610	2.876	2.467,22	1.500,00	1,15
VISTA ALEGRE	11.199	2.947	2.036,11	810,00	1,87
XAXIM	57.182	15.940	1.040,56	700,00	3,07

Figura 2. Dataset limpo

Também foi feita a mudança dos nomes dos bairros para *uppercase*, para posteriormente fazer a junção das duas tabelas.

Outra tabela também adquirida e utilizada, foi a de ocorrências em Curitiba, disponibilizados pela própria Guarda Municipal da cidade, disponibilizada no site <https://www.curitiba.pr.gov.br/dadosabertos/busca/?termo=guarda>, também foi utilizada uma tabela semelhante com a geolocalização (Latitude e Longitude) dos crimes, de um dataset disponibilizado pelo professor.

Através da geolocalização dos bairros, também foi construído um dataset com a distância de cada bairro em relação ao centro, para verificar o parâmetro da primeira hipótese.

Visualizando as coisas básicas do dataset, foi visto que existiam muitas (185) naturezas de crime (Roubo, Furto, Desacato, Invasão, etc.), então, foi considerado as naturezas mais relevantes e comuns, reduzindo o número de naturezas de 185 para 41, para serem analisadas.

Para entender melhor a relação dos horários, e termos gráficos mais relevantes foi feito uma seleção das naturezas mais relevantes e comuns, como roubos, agressões, furtos, violência doméstica, depois um agrupamento dos tipos de ocorrências, juntando roubo e furto, crimes de pudor etc, assim reduzindo a quantidade de tipos de ocorrência, como é possível ver na tabela abaixo:

	NATUREZA	TOTAL_POR_CRIME
7	ROUBO	26839
0	AGRESSAO	15361
6	RISCO_VIDA	4382
1	CRIME_MULHER	3681
5	PUDOR	2805
8	SEQUESTRO	415
4	MORTES	141
2	ESTUPRO	108
3	EXPLORACAO	17

Figura 3. Tabela de crimes mais cometidos

Após a filtragem dos crimes, esses crimes foram agrupados por bairro, verificando a quantidade de crimes cometidos em cada bairro. Para normalizar o dataset, a quantidade de crimes foi dividida pelo número de habitantes em cada bairro sob a variável nomeada de “Criminalidade”, ou seja, quantidade de crimes por habitante. Esse dataset foi juntado com o dataset de cada parâmetro do bairro, assim como a distância também calculada para realizar a regressão múltipla presente nos resultados.

Para entender melhor como essas ocorrências acontecem ao longo do dia, foi separado pela natureza e plotado a quantidade de cada um, assim podendo ter uma ideia inicial de como cada tipo de crime se comporta ao longo do dia, visualizado na Figura 9.

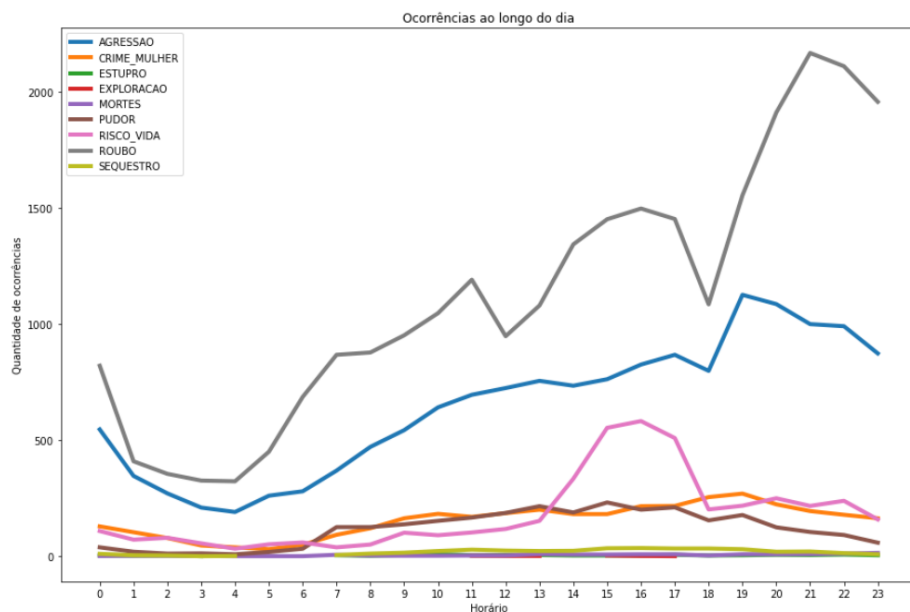


Figura 4. Gráfico de linha mostrando a ocorrência de cada crime em cada hora

Com essa visualização tínhamos uma ideia do que esperar para o modelo, porém para podermos relacionar a natureza e os horários foi preciso fazer uma normalização de cada tipo de crime durante o dia, então todos os crimes ficaram no intervalo de 0 a 1. Assim facilitando a análise por distância euclidiana, como apresentado na Figura 10.

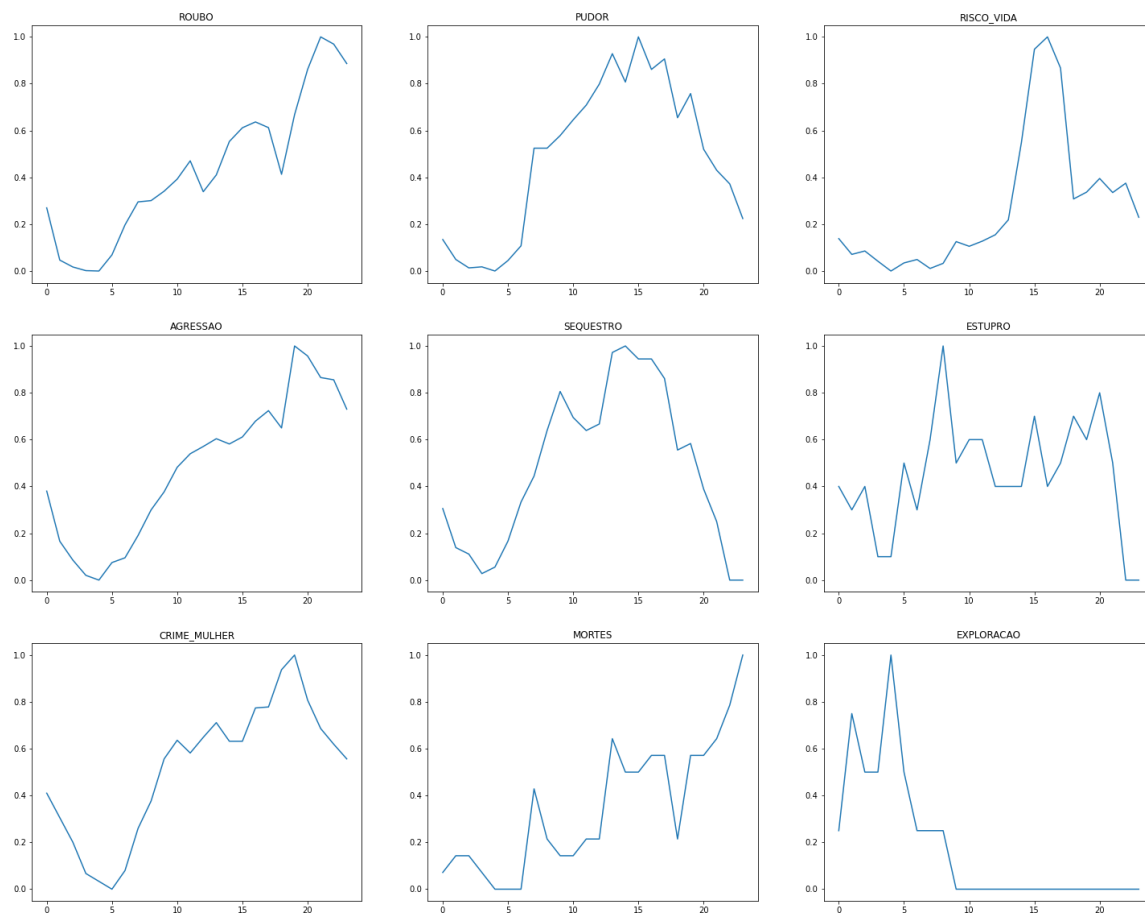


Figura 5. Crimes normalizados

Por fim, foi analisado como se dá a distribuição das diferentes naturezas de crime para cada dia da semana e mês do ano, como apresentado na figura 11.

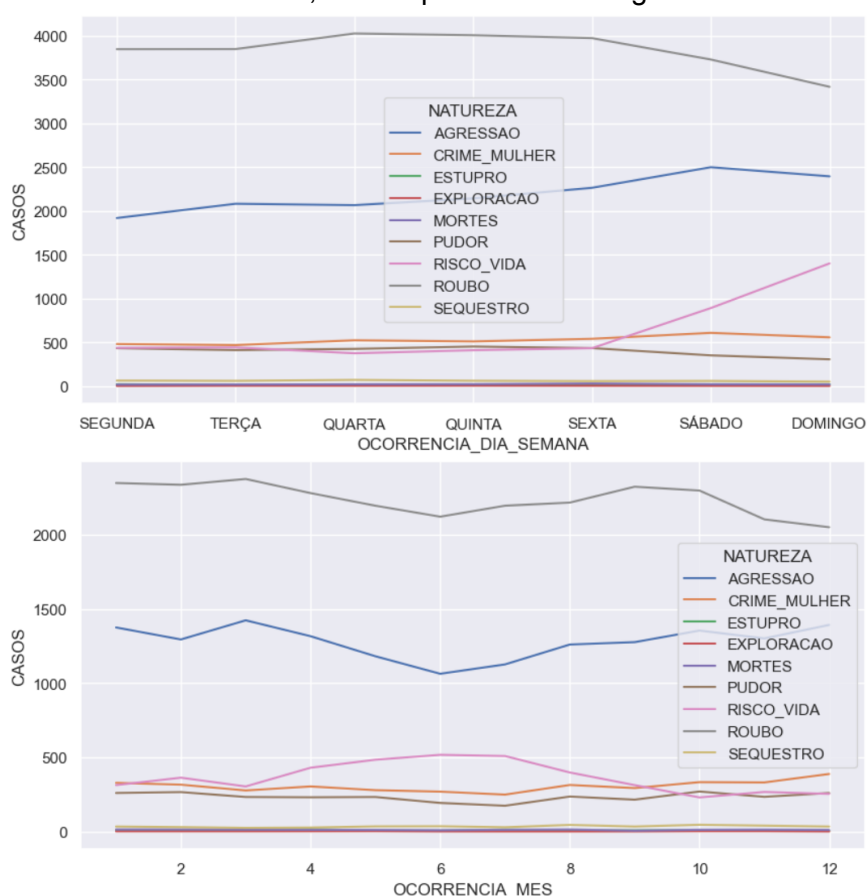


Figura 6. Gráfico de linha mostrando a ocorrência de cada crime em cada dia da semana e depois em cada mês

Resultados

Para responder a primeira pergunta foi feito regressão múltipla a partir da variável dependente (criminalidade) e das variáveis independentes (taxa de analfabetismo, distância em relação ao centro e rendimento mensal médio), com isso chegamos no seguinte resultado demonstrado na tabela abaixo:

OLS Regression Results			
Dep. Variable:	CRIMINALIDADE	R-squared:	0.196
Model:	OLS	Adj. R-squared:	0.135
Method:	Least Squares	F-statistic:	3.244
Date:	Wed, 07 Dec 2022	Prob (F-statistic):	0.0318
Time:	23:07:31	Log-Likelihood:	124.66
No. Observations:	44	AIC:	-241.3
Df Residuals:	40	BIC:	-234.2
Df Model:	3		
Covariance Type:	nonrobust		

Figura 7. Métricas de resultado da regressão múltipla (1º hipótese)

Com esse resultado e os resultados das outras tabelas descritas no item anterior conseguimos chegar na conclusão que por nosso estudo se basear na cidade de Curitiba, as variáveis socioeconômicas não são tão impactantes quanto o esperado, talvez por a escala ser muito pequena, além disso para obter resultados mais concretos seria necessário adquirir mais informações, como mais variáveis relevantes, entre elas o policiamento, fluxo de pessoas, iluminação, entre outras, e também seria necessário mais índices socioeconômicos de cada bairro como IDH, grau de escolaridade, taxa de desemprego, etc.

Na segunda pergunta foi feito clusterização hierárquica com as séries temporais de cada crime e com isso foi gerado o seguinte dendrograma abaixo:

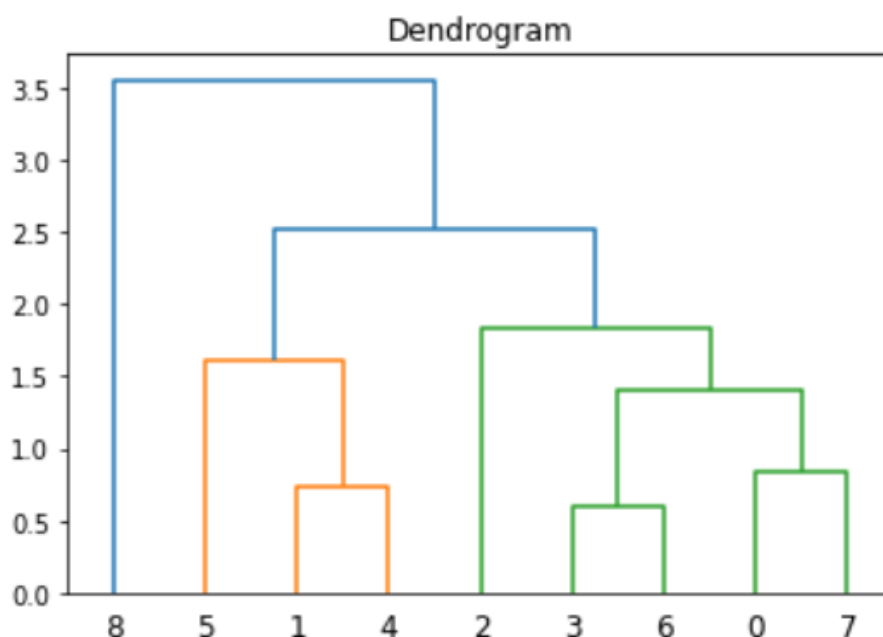


Figura 8. Dendrograma dos tipos de crimes montado através da clusterização

Com esse dendrograma e as séries temporais citadas no item anterior, conseguimos chegar na conclusão que o horário tem uma influência nos tipos de crimes, pois segundo o dendrograma, os crimes de agressão (3) e crime contra a mulher (6) são muito semelhantes, ou seja eles ocorrem frequentemente nos mesmos horários do dia, o mesmo acontece com crime contra pudor (1) e sequestro (4) e roubos (0) e morte (7), além disso, temos crimes que são menos semelhantes, como o estupro (5) com o primeiro grupo e o crime de risco a vida (2) com o segundo grupo e por fim podemos concluir que o crime de exploração (8) é o outlier do dendrograma e não possui quase nenhuma semelhança com os outros crimes.

Após a análise do dendrograma chegamos em algumas suposições para explicar alguns grupos formados no dendrograma.

Primeira suposição é sobre o primeiro grupo, que supõe que os crimes contra a mulher também envolvam denúncias de violência ou agressões, segundo o seguinte trecho retirado da [Veja abril](#), “Entre as agressões relatadas nos últimos doze meses, a maioria ocorreu em casa (42%), enquanto 29% na rua e o restante em ambientes como o trabalho, um bar/ balada ou a internet – na lista dos locais mais comuns.” podemos supor que as

agressões que ocorrem dentro de casa são consideradas crime contra a mulher por conta da Lei Maria da Penha e as que ocorrem fora são tratadas como agressões, e grande parte das vítimas de crime contra a mulher não reportam o abuso segundo o trecho retirado da [Veja abril](#) *“Das que buscaram ajuda, 23,8% disseram ter procurado algum órgão especializado (em instituições como delegacias da mulher ou delegacias comuns, além de ligações para o 190). Já 15% compartilharam a situação com alguém da família ou pessoas próximas.”.*

Segunda suposição é sobre o crime contra o pudor (1) e o estupro (5), que se baseia na teoria de que ocorre a semelhança, pois dentro da categoria de estupros também está incluso os atentados violentos ao pudor, essa teoria tem como base a seguinte afirmação retirada do [ConJur](#) *“O crime de atentado violento ao pudor foi inserido dentro do tipo penal do crime de ESTUPRO a partir do ano de 2009 e, dessa forma, a conduta do autor que atuara sob o revogado artigo 214 do Código Penal passou a ser classificada como crime de ESTUPRO (art. 213) e não de estupro de vulnerável (artigo 217-A).”.*

A última suposição é sobre o terceiro grupo, que tem a teoria de que ocorre essa semelhança, pois esse crimes ocorrem no fim da noite e uma parte dos roubos são assaltos à mão armada, o que aumenta as chances de morte da vítima, um trecho retirado do [G1](#) fala um pouco acerca desse assunto *“Somente na capital, foram 79 casos de roubos ou tentativa de roubos com tiros, que resultaram em 28 mortos e 52 feridos — 80 baleados no total. Em toda a Região Metropolitana, somente no mês de junho houve 20 casos de tiroteios durante tentativas de assalto, com oito mortos e 12 feridos.”.*

Já para a terceira pergunta, os dados foram agrupados por natureza, mês e dia da semana, e através de uma regressão múltipla, foi analisado se essas variáveis possuem influência na quantidade de crimes, os resultados do modelo se apresentam abaixo:

Dep. Variable:	CASOS_NORMALIZADOS	R-squared:	0.972
Model:	OLS	Adj. R-squared:	0.971
Method:	Least Squares	F-statistic:	1709.
Date:	Sun, 11 Dec 2022	Prob (F-statistic):	0.00
Time:	19:17:27	Log-Likelihood:	1272.1
No. Observations:	756	AIC:	-2512.
Df Residuals:	740	BIC:	-2438.
Df Model:	15		
Covariance Type:	nonrobust		

Figura 9. Métricas de resultado do modelo para a hipótese 3

Com esse resultado pode-se constatar que os dias da semana e o meses do ano conseguem explicar bem a variância da quantidade de ocorrências por natureza de crime.

Limitações e Trabalhos Futuros

Um dos problemas encontrados foi a falta de diversas informações em relação aos bairros, como o índice de desemprego de cada bairro, que é considerado como um fator determinante para a quantidade de crimes cometidos nos bairros. Além disso, o fato de não

termos encontrado influência na regressão múltipla talvez tenha sido devido a analisar todos os crimes, ao invés de analisar cada tipo de crime separadamente, e poderíamos ver em quais bairros seriam os crimes mais comuns, e como as nossas variáveis influenciam nesses tipos de crimes.

Outra limitação na segunda pergunta é que foi necessário agrupar os crimes, por possuir muitos tipos de ocorrências (185) e que foi necessário fazer uma redução e com isso talvez tenha-se perdido dados relevantes, e às vezes que possa ter causado uma interpretação errada. Outro ponto importante foi que não tínhamos a informação do sexo da pessoa afetada no crime, e com isso não tínhamos como separar as análises por sexo e assim entender melhor como ambos os sexos são afetados.