

Universidade Federal do Paraná
Setor de Ciências Exatas Departamento de Estatística

**RELACÃO DO CRIME EM CIDADES PEQUENAS COM CUSTO, IDADE E ANOS
DE ESTUDO DE ESTADOS DA AMERICA DO NORTE**

CE225 - Modelos Lineares Generalizados

Alunos: Andréa Alves GRR 20096668

Everton Souza GRR 20124692

Curitiba, 25 de setembro de 2017

Sumário

| | |
|---|---|
| 1. Introdução | 2 |
| 2. Metodologia | 2 |
| 3. Modelagem | 2 |
| 3.1 Análise descritiva e exploratória..... | 3 |
| 3.2 Modelo com todos os efeitos aditivos..... | 4 |
| 3.3 Modelo proposto..... | 8 |
| 4 .Conclusões | 9 |

1. Introdução

O presente estudo tem por objetivo explicar e quantificar, a relação do crime em pequenas cidades, as informações contidas no arquivo de dados sobre o crime, educação e o custo policial para as pequenas cidades americanas nos 10 estados do sudeste e leste da América. O conjunto de dados consiste em um total de 100 observações que representam as 50 pequenas cidades da América dentro desses estados. As variáveis são: Total Crimes, Crimes violentos, as idades que mais tem frequência e qual a relação que a quantidade de anos de Estudo influencia na ocorrência de crimes. A Base de Dados que será realizada a análise é do “Life in America’s Small Cities”, por G.S. Thomas.

2. Metodologia

As variáveis serão analisadas utilizando Análise de Regressão Linear, com o objetivo de identificar fatores que afetam o crime em cidades menores e encontrar um Modelo de Regressão que melhor estime os dados.

3. Modelagem

Faremos as análises dos dados, com a descrição e modelo proposto. Abaixo uma descrição das variáveis que serão trabalhadas nessa análise.

Total: Total Crimes reportados por 1 milhão de residentes

Violento: Crimes violentos reportados por 100.000 residentes

Custo: Custo Anual por policial em \$/residentes

Idade1: % de pessoas com 25 anos + 4 anos de escolaridade

Idade2: % de pessoas de 16 a 19 anos sem escolaridade

Idade3: % de pessoas de 18 a 24 anos na escola

Idade4: % de pessoas de 25 anos com menos de 4 anos de escolaridade

3.1. Análise Descritiva e Exploratória

Neste tópico iremos explorar as medidas resumo das variáveis

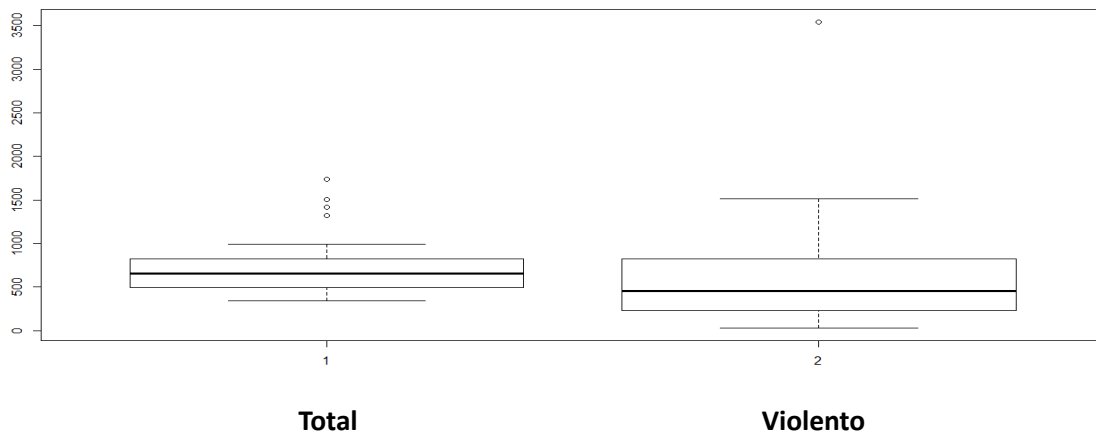
A primeira é a tabela com as medidas resumo das variáveis.

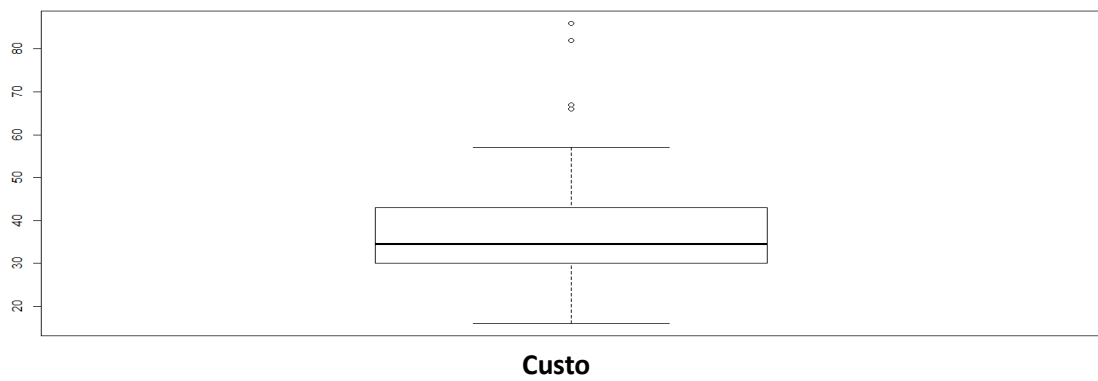
| | Total | Violento | Custo | Idade1 | Idade2 | Idade3 | Idade4 |
|---------|---------|----------|-------|--------|--------|--------|--------|
| Min. | 341,0 | 29,0 | 16,0 | 42,0 | 4,0 | 7,0 | 8,0 |
| 1st Qu. | 497,0 | 230,8 | 30,0 | 49,0 | 11,0 | 21,3 | 11,0 |
| Median | 654,5 | 454,0 | 34,5 | 59,0 | 14,0 | 25,0 | 12,0 |
| Mean | 718,0 | 616,2 | 37,8 | 58,8 | 15,4 | 29,9 | 13,8 |
| 3rd Qu | 820,5 | 822,5 | 42,3 | 67,0 | 19,0 | 34,3 | 15,8 |
| Max | 1.740,0 | 3.545,0 | 86,0 | 81,0 | 34,0 | 81,0 | 36,0 |

Podemos verificar pela tabela que quanto maior a quantidade de Crimes, maior é o custo e nas duas Bases de Dados Idade 1 (até 25 anos com +4 de estudo) e Idade 3 (18 a 24 anos na escola) são os maiores % nessa faixa maior de crimes. Apenas com essas informações não temos como chegar numa conclusão que o estudo interfere nos índices de criminalidade.

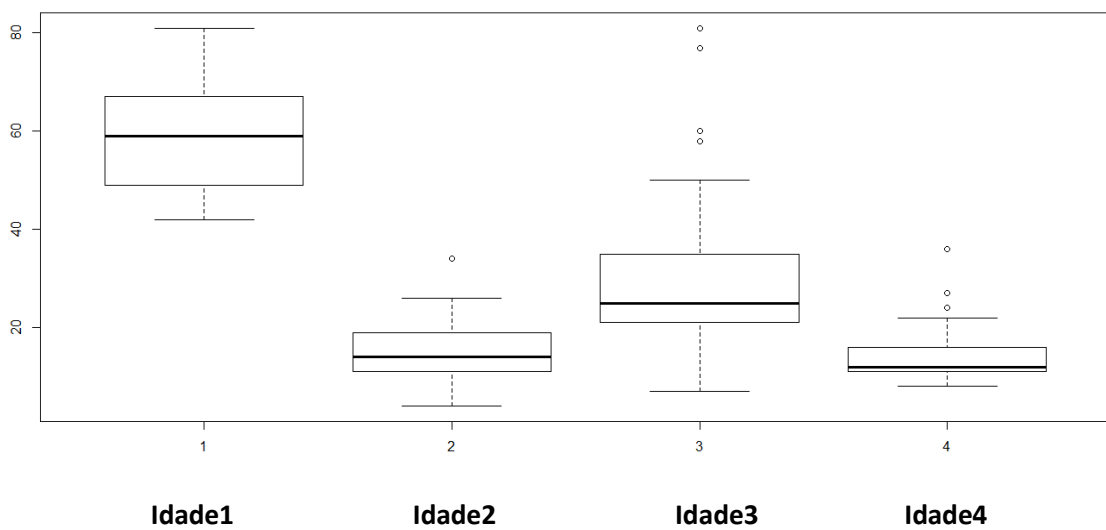
Mas o crime, o financiamento da polícia e a educação têm um relacionamento e um impacto na taxa de criminalidade nas pequenas cidades americanas?

Analisando os Blox Plots dos dados de Total Crimes vs Crimes Violentos, podemos verificar que quantidade de crimes violentos é maior que total crimes, ainda mais que a Proporção utilizada é Total Crimes por 1MM de Habitantes vs Crimes Violentos 100 mil habitantes.

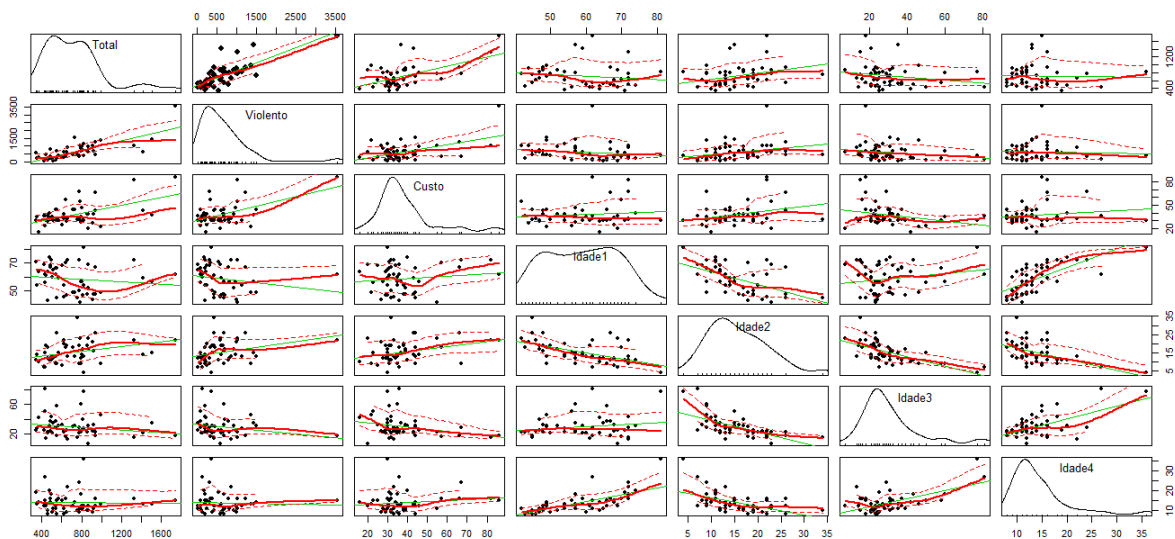




O Gráfico de Custo podemos verificar acima dos 35, é a base maior (35 a 42), ou seja quanto maior a quantidade de crimes, maior é o custo com policiamento.



Quanto ao Gráfico acima, é o boxplot de % de crimes com base na quantidade de crimes.



A figura acima apresenta uma matriz de gráficos, na diagonal temos o Gráfico da Normal a assimetria para o lado esquerdo. Os gráficos de dispersão fornecem um indicativo da tendência dessas informações.

3.2. Modelo com os efeitos aditivos

Fazendo um ajuste das variáveis para definir um modelo, primeiro utilizamos todas as variáveis Crimes Total, Crimes Violento, Custo, Idade1, Idade2, Idade3 e Idade4.

| | Estimador | Erro | T | P valor |
|------------------|-----------|--------|------|---------|
| Intercept | 100,39 | 370,70 | 0,27 | 0,79 |
| Violento | 0,33 | 0,06 | 5,57 | 0,00 |
| Custo | 3,99 | 2,68 | 1,50 | 0,14 |
| Idade1 | 1,85 | 5,24 | 0,36 | 0,72 |
| Idade2 | 7,83 | 7,75 | 1,01 | 0,32 |
| Idade3 | 2,55 | 3,42 | 0,70 | 0,45 |
| Idade4 | 3,23 | 10,72 | 0,30 | 0,76 |

F-statistic: 11,36

Os valores de Idade1 e Idade4 são altos com 0,72 e 0,76 são bastante elevados podemos excluir do modelo pois não são fatores significativos para afetar a condição de crime. Com base disso faremos um novo modelo removendo esses dois fatores.

É demonstrado que o nível de educação superior não é um fator significativo para afetar a condição de crime. Sabendo disso, eu removo as duas variáveis X3 e X6 para continuar fazendo a análise de regressão.

| | Estimador | Erro | T | P valor |
|------------------|-----------|--------|------|----------|
| Intercept | 211,30 | 164,70 | 1,28 | 0,21 |
| Violento | 0,32 | 0,05 | 5,83 | 5,55E-07 |
| Custo | 4,07 | 2,35 | 1,73 | 0,09 |
| Idade2 | 6,45 | 5,98 | 1,08 | 0,29 |
| Idade3 | 1,74 | 2,38 | 0,73 | 0,47 |

F-statistic: 17,74

Considerando as 4 variáveis o resultado da regressão não é satisfeita já que a F statistics é 17,74. Analisando os dados Idade 2 e Idade3 podemos verificar que os dados não influenciam diretamente no modelo, o que nos leva a entender que escolaridade nível de escolaridade não influencia no modelo proposto. Neste caso faremos uma nova regressão excluindo essas duas variáveis.

Fazendo uma nova regressão utilizando agora as três variáveis: Total, Violento e Custo

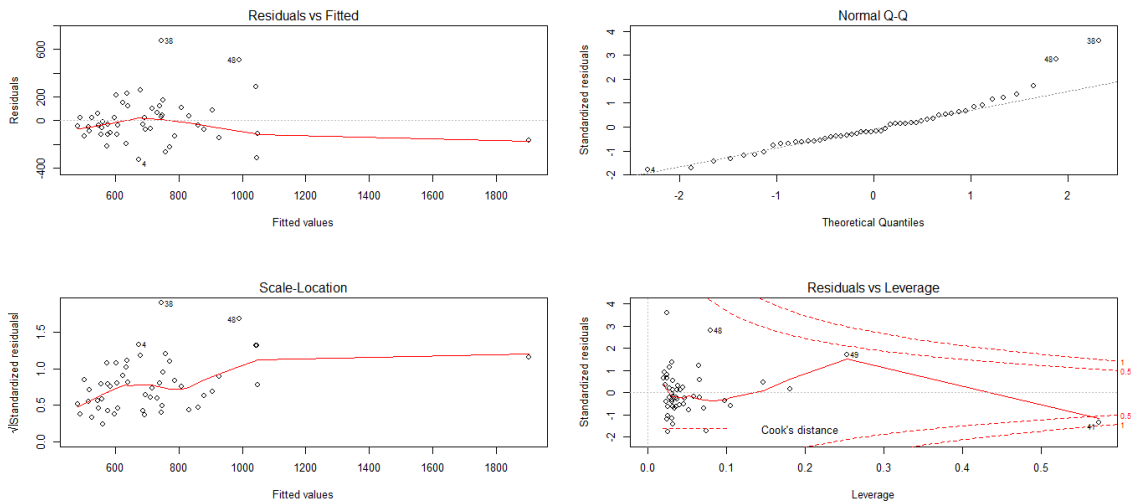
| | Estimador | Erro | T | P valor |
|------------------|-----------|-------|------|----------|
| Intercept | 350,88 | 79,25 | 4,43 | 5,65E-05 |
| Violento | 0,33 | 0,05 | 6,12 | 1,76E-07 |
| Custo | 4,24 | 2,27 | 1,87 | 0,07 |

F-statistic: 16,36

Verificamos, a partir do resultado da análise, que as estatísticas F são melhoradas para 35.5 e R ainda é 0.77, o que mostra que este modelo se encaixa melhor nas observações. Além disso, o valor P de Violento indica a necessidade da variável. Enquanto a variável Custo (financiamento anual da polícia em \$ / residente) não mostra o significado com base no nível de confiança de 95%.

Faremos um novo modelo removendo a variável Custo, assim a variável apenas contém Violento (taxa de criminalidade violenta relatada por 100.000 habitantes).

| | Estimador | Erro | T | P valor |
|------------------|-----------|-------|-------|----------|
| Intercept | 479,14 | 40,52 | 11,80 | 7,99E-16 |
| Violento | 0,39 | 0,05 | 8,01 | 2,10E-10 |



Na figura os residuais estão entre -400 e 200, os resíduos seguem uma normal com alguns outliers no final da curva, nesse caso não rejeitamos a hipótese dos resíduos se distribuírem normalmente. O último gráfico indica os outliers, os pontos fora das bandas em vermelhos podemos classificar como observações neste caso não influentes. Com isso podemos seguir com as análises do novo Modelo Proposto.

3.3. Modelo Proposto

Novamente especificaremos o modelo aditivo, mas agora considerando somente as variáveis selecionadas na subsecção anterior.

O modelo adotado é: $X \sim \text{Normal}$

$$Y = 479,14 + 0.39 \text{ crime } X1$$

Coefficients:

```
Estimate Std. Error t value Pr(>|t|)
(Intercept) 479.14487 40.52693 11.823 7.99e-16 ***
Violento 0.38757 0.04836 8.014 2.10e-10 ***
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 194.2 on 48 degrees of freedom
Multiple R-squared: 0.6723, Adjusted R-squared: 0.6634
F-statistic: 64.23 on 1 and 48 DF, p-value: 2.096e-10

Com base nesse último modelo com apenas duas variáveis temos o teste F com Aqui 65 e um ajuste R com 66%, os demais modelos ficaram abaixo desses níveis. Com relação a informação de Violento vs Crime total tem relação direta em casos das cidades pequenas.

4. Conclusão

O modelo utilizando uma variável, usando a taxa de criminalidade violenta relatada por 100.000 habitantes, é o melhor preditor da taxa de criminalidade total relatada global por 1 milhão de residentes. A taxa de criminalidade total por 1 milhão de residentes tem uma forte relação linear com a um conjunto de amostras menores nesse caso 100.000 habitantes. Podemos verificar que o objeto de estudo que a quantidade de estudos pode interferir, no modelo não teve relação direta.