



**PUC - SP**

**MÉTODOS QUANTITATIVOS  
TRABALHO FINAL  
ANALISE DA POPULAÇÃO E NÍVEL DE HOMICÍDIOS NA  
CIDADE DE SÃO PAULO.**

**MÁRCIO CARDOSO SANTOS  
MESTRADO EM ADMINISTRAÇÃO DE EMPRESAS  
MÉTODOS QUANTITATIVOS  
PROF. DR. ARNOLDO HOYOS**

**SÃO PAULO  
2011**

## SUMÁRIO

1. Introdução .....	3
2. Metodologia da Pesquisa .....	4
2.1 Os indivíduos .....	4
2.2 As variáveis .....	5
2.3 A tabela de dados .....	5
3. Análise das variáveis .....	11
4. Correlação entre variáveis .....	15
5. Dendograma .....	18
6. ANOVA .....	21
7. Principal Component Analysis .....	22
8. Análise de Regressão .....	23
9. Regressão Stepwise .....	24
10. Correlação .....	24
11. Análise dos Principais Componentes .....	25
12. Análise Discriminante .....	25
13. Regressão Logística Ordinal .....	26
14. Árvore de Decisão .....	30
15. Conclusão .....	30

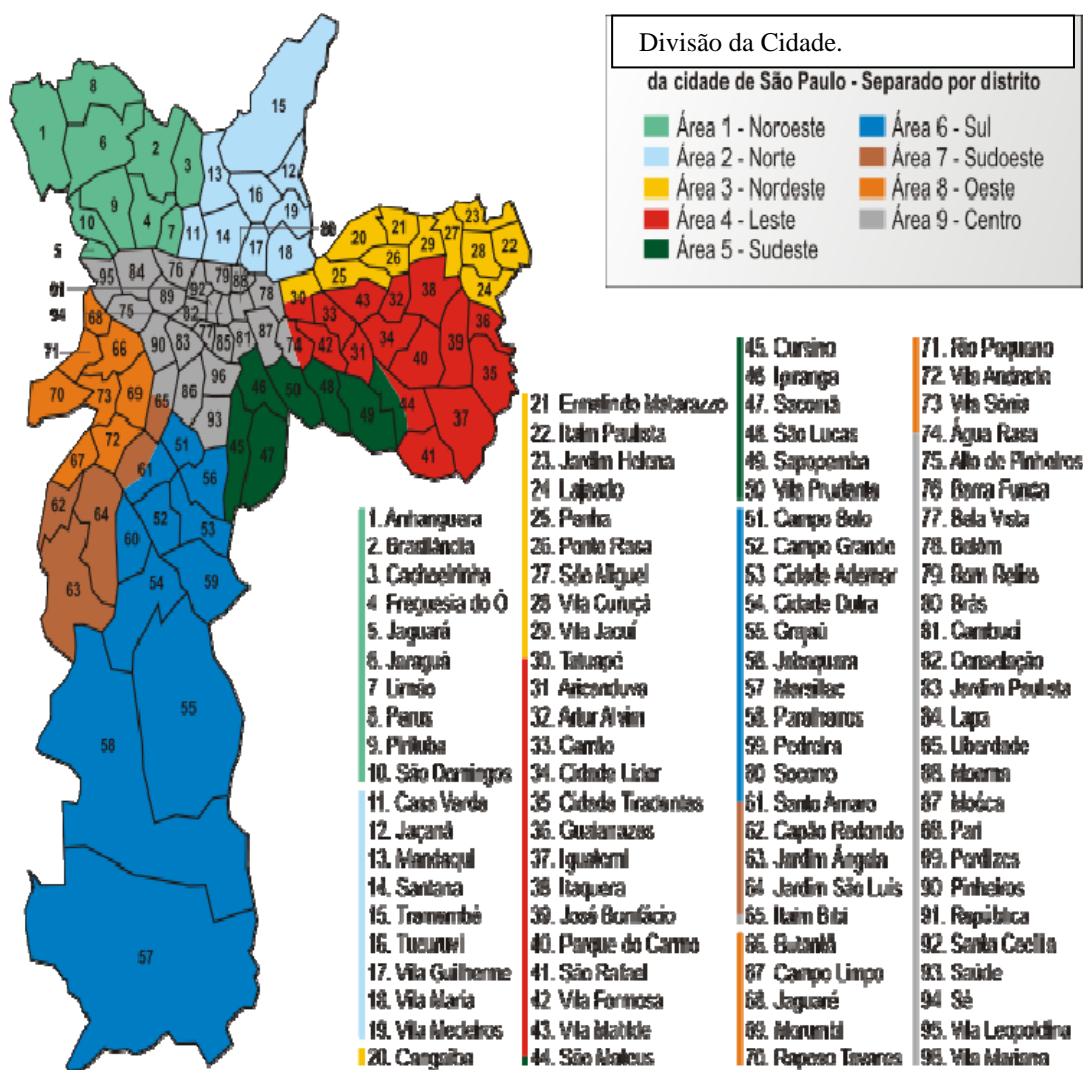
## 1. INTRODUÇÃO

O presente trabalho tem por objetivo efetuar uma análise estatística de dados populacionais e o nível de homicídios nos bairros da cidade de São Paulo, tendo em vista, como uma das maiores cidades do mundo, o fator de violência que preocupa toda a população paulistana. São Paulo tem hoje 96 distritos e comporta mais de 11 milhões de habitantes. Inicia-se o presente trabalho com o entendimento dos dados, incluindo as definições que se fizerem necessárias das variáveis, suas classificações em variáveis categóricas ou quantitativas, os significados e unidades de medida, além da apresentação da tabela de dados. Na seqüência, faz-se a análise de cada uma das variáveis separadamente quanto a sua forma de distribuição, os valores atípicos, medidas de centro e dispersão. Em conclusão, analisa-se comparativamente as análises efetuadas no desenvolvimento do trabalho. O software estatístico utilizado é o **MINITAB**.

## 2. METODOLOGIA DE PESQUISA

### 2.1 Os Indivíduos

Os indivíduos desta análise são 96 distritos da cidade de São Paulo, que estão distribuídos em 9 áreas e 5 zonas, conforme podemos observar no mapa a seguir do site da SPTransporte em junho de 2011. Levou-se em consideração fatores que podem colaborar para o maior ou menor nível de homicídios, segundo as características específicas de cada bairro estudado.



### 2.2 As Variáveis

Foram analisadas 5 variáveis: população total, ensino fundamental incompleto em percentual, ensino fundamental completo em percentual, ensino médio completo em percentual, homicídios em percentual.

As informações deste trabalho foram obtidas do portal da prefeitura da cidade de São Paulo: <http://infocidade.prefeitura.sp.gov.br/>, e têm o período de 2007 como referencia de todas das tabelas.

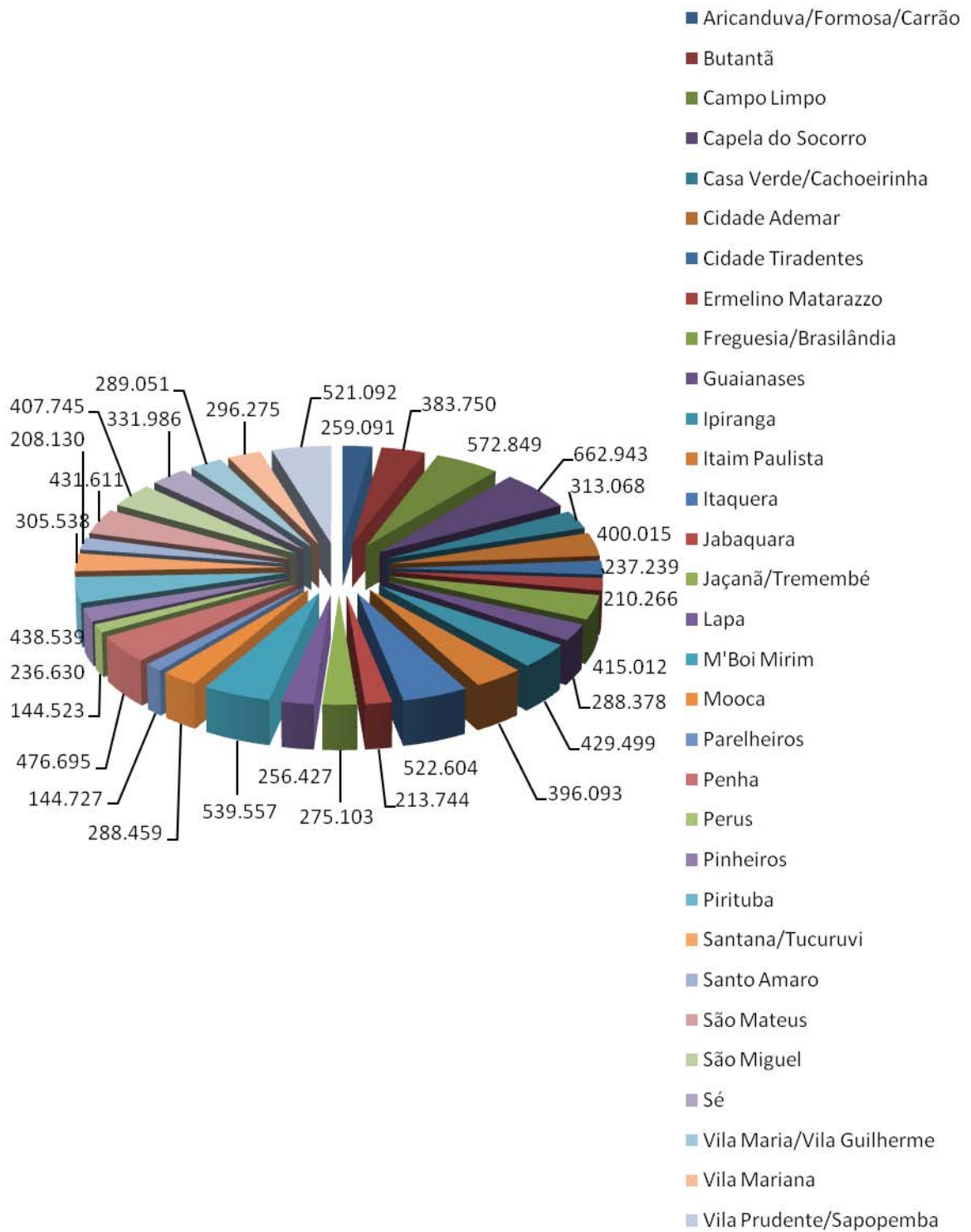
### 2.3 A Tabela de Dados

Com o objetivo de facilitar a interpretação dos dados pelo fato de que a tabela geral apresenta dados dos 96 distritos do município de São Paulo, são apresentadas as tabelas abaixo correspondendo a população e as variáveis para estudo do presente trabalho, com suas respectivas taxas percentuais.

Unidades Territoriais	Total
<b>Município de São Paulo</b>	<b>10.896.639</b>
<b>Aricanduva/Formosa/Carrão</b>	<b>259.091</b>
Aricanduva	94.287
Carrão	73.519
Vila Formosa	91.285
<b>Butantã</b>	<b>383.750</b>
Butantã	49.171
Morumbi	32.700
Raposo Tavares	96.216
Rio Pequeno	115.631
Vila Sônia	90.032
<b>Campo Limpo</b>	<b>572.849</b>
Campo Limpo	211.498
Capão Redondo	268.403
Vila Andrade	92.948
<b>Capela do Socorro</b>	<b>662.943</b>
Cidade Dutra	202.409
Grajaú	424.150
Socorro	36.384
<b>Casa Verde/Cachoeirinha</b>	<b>313.068</b>
Cachoeirinha	156.329
Casa Verde	78.129
Limão	78.610
<b>Cidade Ademar</b>	<b>400.015</b>
Cidade Ademar	248.407
Pedreira	151.608
<b>Cidade Tiradentes</b>	<b>237.239</b>
Cidade Tiradentes	237.239
<b>Ermelino Matarazzo</b>	<b>210.266</b>
Ermelino Matarazzo	114.837
Ponte Rasa	95.429

<b>Freguesia/Brasilândia</b>	<b>415.012</b>
Brasilândia	274.009
Freguesia do Ó	141.003
<b>Guaianases</b>	<b>288.378</b>
Guaianases	108.757
Lajeado	179.621
<b>Ipiranga</b>	<b>429.499</b>
Cursino	95.688
Ipiranga	96.625
Sacomã	237.186
<b>Itaim Paulista</b>	<b>396.093</b>
Itaim Paulista	235.868
Vila Curuçá	160.225
<b>Itaquera</b>	<b>522.604</b>
Cidade Líder	127.951
Itaquera	217.239
José Bonifácio	108.662
Parque do Carmo	68.752
<b>Jabaquara</b>	<b>213.744</b>
Jabaquara	213.744
<b>Jaçanã/Tremembé</b>	<b>275.103</b>
Jaçanã	93.548
Tremembé	181.555
<b>Lapa</b>	<b>256.427</b>
Barra Funda	10.866
Jaguara	23.664
Jaguare	41.864
Lapa	53.999
Perdizes	99.780
Vila Leopoldina	26.254
<b>M'Boi Mirim</b>	<b>539.557</b>
Jardim Ângela	281.178
Jardim São Luís	258.379
<b>Mooca</b>	<b>288.459</b>
Água Rasa	80.981
Belém	36.189
Brás	22.192
Mooca	58.777
Pari	13.104
Tatuapé	77.216
<b>Parelheiros</b>	<b>144.727</b>
Marsilac	9.659
Parelheiros	135.068
<b>Penha</b>	<b>476.695</b>
Artur Alvim	107.391
Cangaíba	149.193
Penha	120.184
Vila Matilde	99.927
<b>Perus</b>	<b>144.523</b>
Anhanguera	56.639
Perus	87.884
<b>Pinheiros</b>	<b>236.630</b>
Alto de Pinheiros	40.960

Itaim Bibi	69.942
Jardim Paulista	71.506
Pinheiros	54.222
<b>Pirituba</b>	<b>438.539</b>
Jaraguá	183.929
Pirituba	164.659
São Domingos	89.951
<b>Santana/Tucuruvi</b>	<b>305.538</b>
Mandaqui	102.936
Santana	110.835
Tucuruvi	91.767
<b>Santo Amaro</b>	<b>208.130</b>
Campo Belo	59.682
Campo Grande	96.354
Santo Amaro	52.094
<b>São Mateus</b>	<b>431.611</b>
Iguatemi	126.359
São Mateus	158.473
São Rafael	146.779
<b>São Miguel</b>	<b>407.745</b>
Jardim Helena	150.772
São Miguel	94.569
Vila Jacuí	162.404
<b>Sé</b>	<b>331.986</b>
Bela Vista	59.280
Bom Retiro	22.035
Cambuci	25.415
Consolação	47.563
Liberdade	54.245
República	43.080
Santa Cecília	63.531
Sé	16.837
<b>Vila Maria/Vila Guilherme</b>	<b>289.051</b>
Vila Guilherme	47.783
Vila Maria	109.257
Vila Medeiros	132.011
<b>Vila Mariana</b>	<b>296.275</b>
Moema	66.358
Saúde	113.484
Vila Mariana	116.433
<b>Vila Prudente/Sapopemba</b>	<b>521.092</b>
São Lucas	131.911
Sapopemba	295.084
Vila Prudente	94.097



Unid Territoriais	Total	EF inc tx	EF com tx	EM comp tx	Homicídios
Aricanduva	94287	21,4	19,1	31,5	45,35
Carrão	73519	21,8	15,6	34,7	33,26
Vila Formosa	91285	17,3	16,5	29,0	28,77
Butantã	49171	15,2	14,0	31,4	32,29
Morumbi	32700	13,4	11,9	21,0	54,93
Raposo Tavares	96216	21,0	17,0	31,7	44,95
Rio Pequeno	115631	24,5	18,2	26,0	17,90
Vila Sônia	90032	18,8	12,1	28,4	24,03
Campo Limpo	211498	25,7	16,7	26,6	52,21
Capão Redondo	268403	25,3	16,4	25,4	63,96
Vila Andrade	92948	19,2	16,4	22,0	29,87
Cidade Dutra	202409	19,0	16,2	33,9	77,85
Grajaú	424150	29,5	22,4	18,5	85,77
Socorro	36384	18,1	14,7	28,9	33,25
Cachoeirinha	156329	22,0	19,2	30,5	88,05
Casa Verde	78129	16,3	20,6	35,4	26,31
Limão	78610	21,6	14,6	33,4	51,19
Cidade Ademar	248407	19,8	20,0	28,8	68,21
Pedreira	151608	21,0	20,3	26,9	47,09
Cidade Tiradentes	237239	19,5	19,2	29,9	102,28
Ermelino Matarazzo	114837	25,9	19,0	27,9	68,33
Ponte Rasa	95429	17,0	10,3	30,7	31,60
Brasilândia	274009	27,1	16,0	26,6	87,33
Freguesia do Ó	141003	19,9	17,2	30,8	32,43
Guaianases	108757	24,5	17,5	27,8	102,49
Lajeado	179621	31,2	17,8	22,6	65,92
Cursino	95688	20,2	14,2	29,9	48,00
Ipiranga	96625	20,2	17,6	26,8	36,41
Sacomã	237186	19,8	15,9	34,9	50,81
Itaim Paulista	235868	24,4	20,1	26,7	58,76
Vila Curuçá	160225	21,2	18,3	27,2	79,87
Cidade Líder	127951	21,7	17,3	32,5	60,77
Itaquera	217239	25,0	17,0	29,0	69,47
José Bonifácio	108662	27,3	19,1	28,2	48,56
Parque do Carmo	68752	20,3	16,8	25,9	60,87
Jabaquara	213744	18,3	17,1	28,4	55,12
Jaçanã	93548	22,1	16,0	33,2	51,19
Tremembé	181555	23,6	17,8	26,2	44,57
Barra Funda	10866	8,6	10,0	29,6	30,85
Jaguara	23664	24,2	14,1	27,3	19,45
Jaguaré	41864	20,1	17,5	22,4	42,37
Lapa	53999	12,8	10,8	32,4	8,31
Perdizes	99780	10,1	8,7	23,8	5,86

Vila Leopoldina	26254	15,5	14,2	23,7	11,16
Jardim Ângela	281178	24,6	18,7	23,2	90,72
Jardim São Luís	258379	20,9	18,4	29,6	84,46
Água Rasa	80981	20,1	17,5	26,5	30,27
Belém	36189	25,0	16,9	27,7	37,86
Brás	22192	15,9	19,6	32,8	71,55
Mooca	58777	14,2	14,6	32,2	34,77
Pari	13104	24,3	18,0	26,0	33,73
Tatuapé	77216	17,4	14,5	30,2	20,16
Marsilac	9659	17,5	17,4	24,5	11,90
Parelheiros	135068	27,3	23,8	21,7	105,99
Artur Alvim	107391	26,7	13,0	32,3	44,96
Cangaíba	149193	22,2	15,4	29,0	32,74
Penha	120184	22,9	13,8	31,3	23,33
Vila Matilde	99927	23,7	12,6	30,3	29,14
Anhanguera	56639	25,9	17,6	27,2	23,42
Perus	87884	26,7	20,3	26,4	32,54
Alto de Pinheiros	40960	11,9	10,2	23,3	18,00
Itaim Bibi	69942	7,8	8,6	24,6	12,28
Jardim Paulista	71506	7,3	7,9	19,5	3,59
Pinheiros	54222	8,1	7,9	25,8	9,52
Jaraguá	183929	22,1	19,3	27,1	54,15
Pirituba	164659	28,8	13,9	24,6	41,41
São Domingos	89951	23,6	16,5	29,8	18,11
Mandaqui	102936	21,6	13,0	31,3	23,28
Santana	110835	16,1	13,0	31,1	22,46
Tucuruvi	91767	21,0	12,5	31,1	22,14
Campo Belo	59682	11,4	9,5	25,6	19,51
Campo Grande	96354	13,3	14,3	32,5	53,63
Santo Amaro	52094	12,1	9,3	31,6	26,43
Iguatemi	126359	19,3	17,5	23,0	75,65
São Mateus	158473	23,9	19,7	28,0	91,70
São Rafael	146779	24,5	16,1	28,3	81,54
Jardim Helena	150772	28,1	13,8	23,7	66,86
São Miguel	94569	19,4	15,1	33,1	56,48
Vila Jacuí	162404	21,1	16,8	29,0	52,13
Bela Vista	59280	13,2	14,2	31,2	28,49
Bom Retiro	22035	18,8	14,6	32,5	41,36
Cambuci	25415	18,4	16,8	29,8	31,34
Consolação	47563	12,1	11,3	26,7	14,67
Liberdade	54245	15,0	12,8	28,0	27,47
República	43080	20,9	14,5	29,3	56,58
Santa Cecília	63531	16,2	17,1	30,9	30,91
Sé	16837	19,6	18,1	36,0	99,43
Vila Guilherme	47783	24,1	15,8	24,8	26,01
Vila Maria	109257	23,0	18,8	28,4	39,53

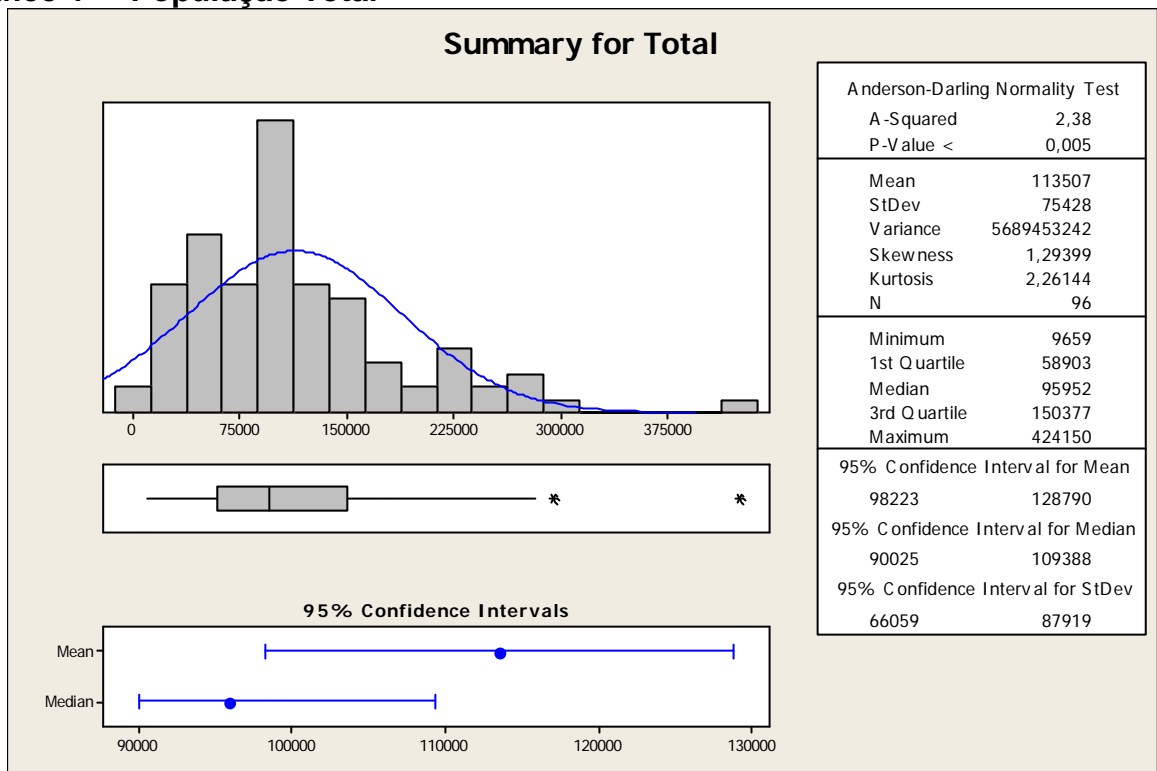
Vila Medeiros	132011	19,0	18,0	32,6	52,65
Moema	66358	8,0	8,3	22,3	7,01
Saúde	113484	12,6	8,3	25,9	13,55
Vila Mariana	116433	8,0	9,5	25,7	16,17
São Lucas	131911	22,0	16,2	33,7	40,19
Sapopemba	295084	24,6	20,1	28,6	72,99
Vila Prudente	94097	14,0	20,7	31,3	28,40

### 3. ANÁLISE DAS VARIÁVEIS

#### 3.1 Variáveis

#### Gráficos Summary

Gráfico 1 – População Total

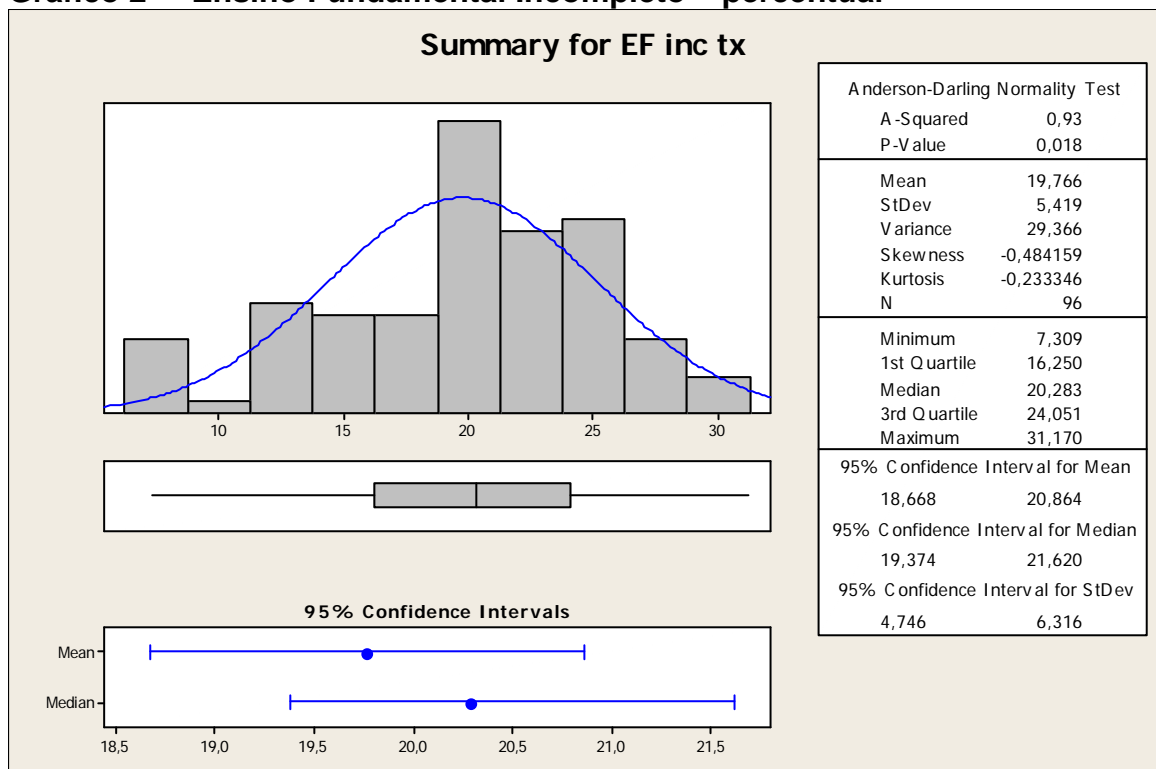


As principais observações que podemos fazer são:

- **Forma:** O Histograma nos permite verificar que se trata de uma distribuição assimétrica à direita. O intervalo de confiança dos retornos está entre 113507 e 95952. Através da análise do teste estatístico de Anderson-Darling medimos quão bem os dados seguem uma distribuição específica, considerando 5% como parâmetro no teste de hipótese, concluímos alto nível de significância para a variável que, portanto, segue uma distribuição específica.

- Valores Atípicos: Possui dois valores atípicos que são Grajaú com uma população de 424150 habitantes e Sapopemba com 295084 habitantes, constituindo os dois distritos mais populosos.
- Centro e Dispersão: É referente à distribuição dos dados que está espalhada em torno de sua média. O valor do desvio-padrão 75428, o que expressa a diferença entre os bairros mais populosos dos menos populosos.

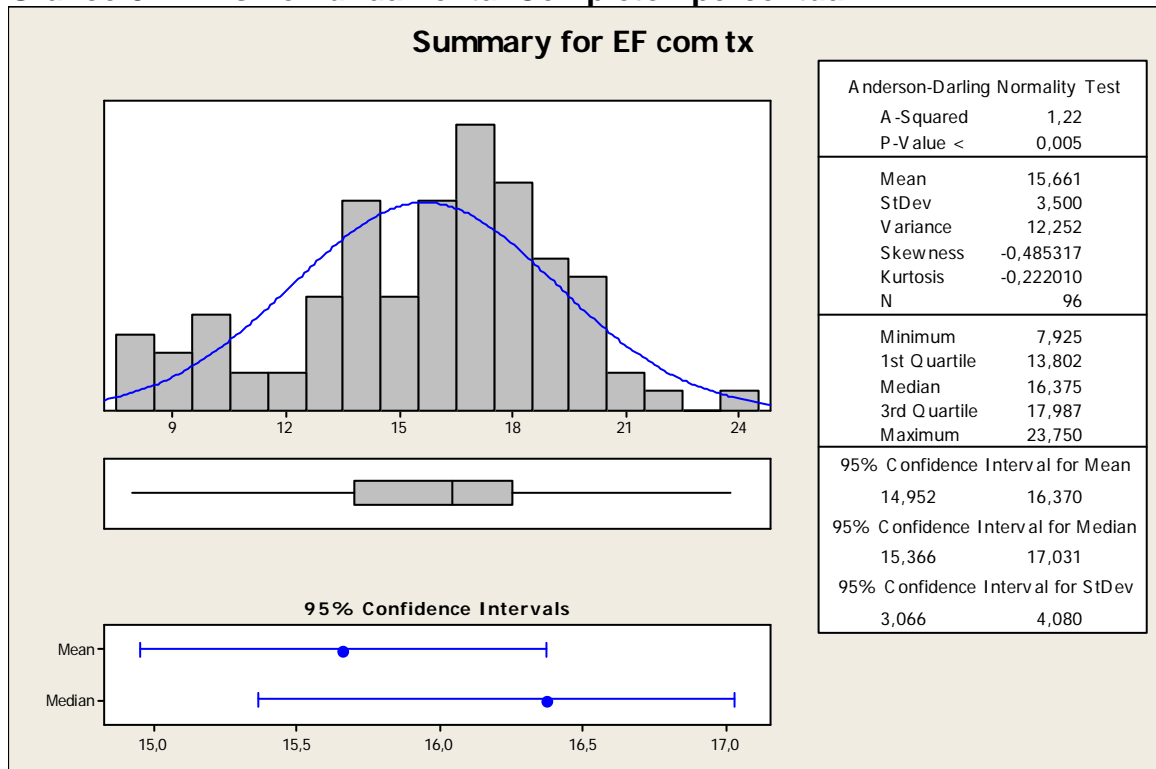
**Gráfico 2 – Ensino Fundamental Incompleto – percentual**



As principais observações que podemos fazer são:

- Forma: O Histograma nos permite verificar que se trata de uma distribuição simétrica. O intervalo de confiança dos retornos está entre 19,766 e 20,283. Através da análise do teste estatístico de Anderson-Darling medimos quão bem os dados seguem uma distribuição específica, considerando 1,8% como parâmetro no teste de hipótese, conclui-se alto nível de significância para a variável que, portanto, segue uma distribuição específica.
- Valores Atípicos: Não há valor atípico.
- Centro e Dispersão: É referente à distribuição dos dados que está espalhada em torno de sua média. O valor do desvio-padrão 5,419.

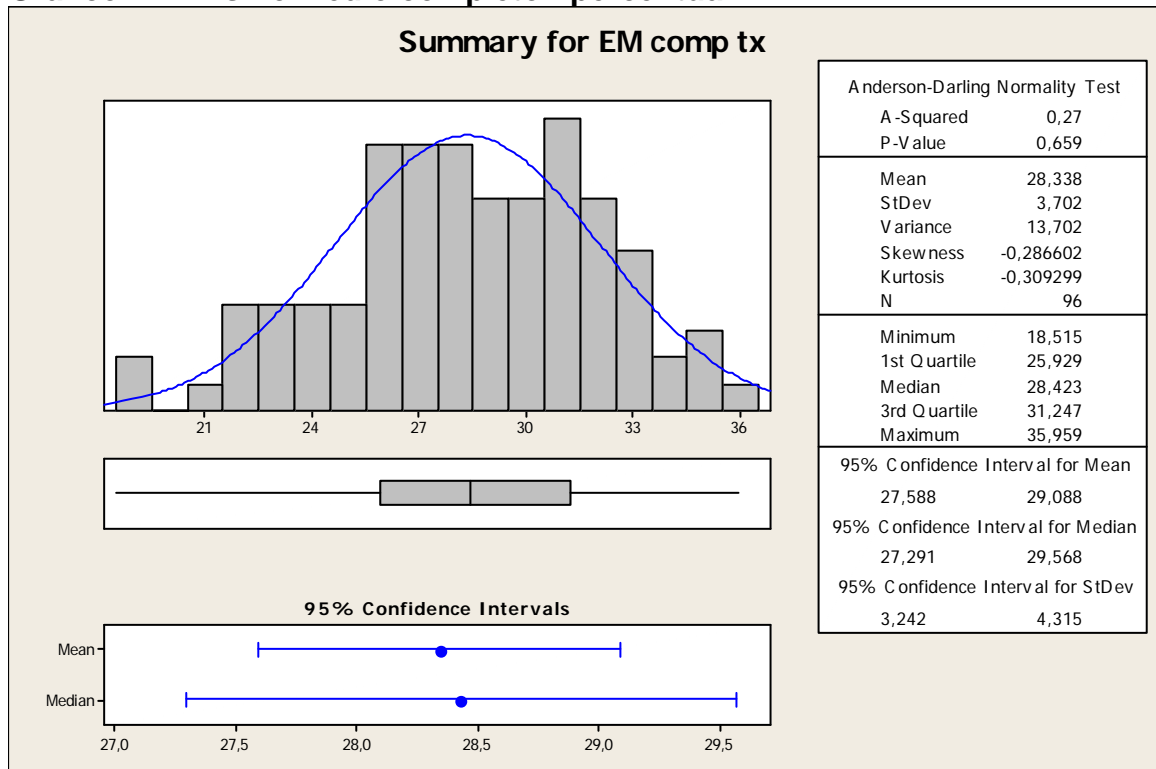
**Gráfico 3 – Ensino Fundamental Completo - percentual**



As principais observações que podemos fazer são:

- **Forma:** O Histograma nos permite verificar que se trata de uma distribuição assimétrica à direita. O intervalo de confiança dos retornos está entre 15,661 a 16,375. Através da análise do teste estatístico de Anderson-Darling medimos quão bem os dados seguem uma distribuição específica, considerando 5% como parâmetro no teste de hipótese, concluímos alto nível de significância para a variável que, portanto, segue uma distribuição específica.
- **Valores Atípicos:** Não há valores atípicos.
- **Centro e Dispersão:** É referente à distribuição dos dados que está espalhada em torno de sua média. O valor do desvio-padrão 3,50, o que pode significar que a população matriculada no ensino fundamental o conclui, nos distritos analisados.

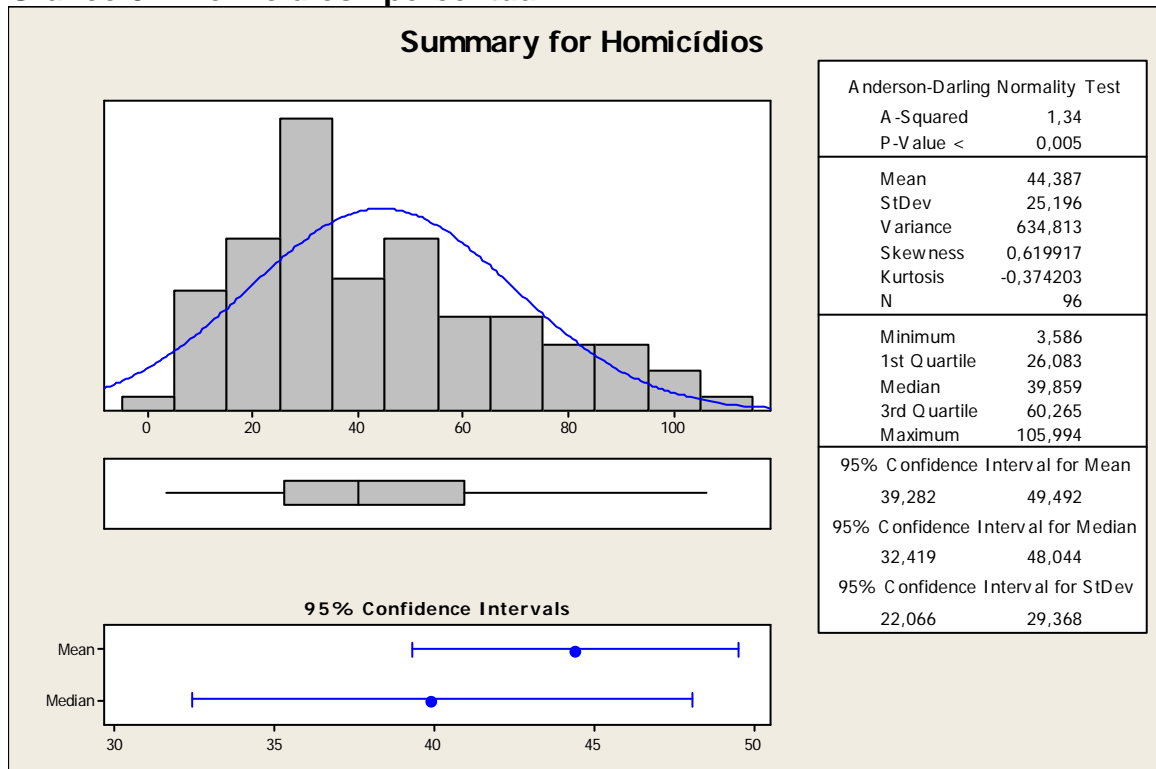
Gráfico 4 – Ensino Médio completo - percentual



As principais observações que podemos fazer são:

- **Forma:** O Histograma nos permite verificar que se trata de uma distribuição assimétrica à direita. O intervalo de confiança dos retornos está entre 28,338 e 28,423. Através da análise do teste estatístico de Anderson-Darling medimos quão bem os dados seguem uma distribuição específica, considerando 65,9%, o que se considera alto nível de significância para a variável que, portanto, segue uma distribuição específica.
- **Valores Atípicos:** Não há valores atípicos.
- **Centro e Dispersão:** É referente à distribuição dos dados que está espalhada em torno de sua média. O valor do desvio-padrão 3,702 demonstrando baixíssima variação entre os distritos estudados quanto ao percentual de indivíduos que concluem o ensino médio.

Gráfico 5 – Homicídios - percentual



As principais observações que podemos fazer são:

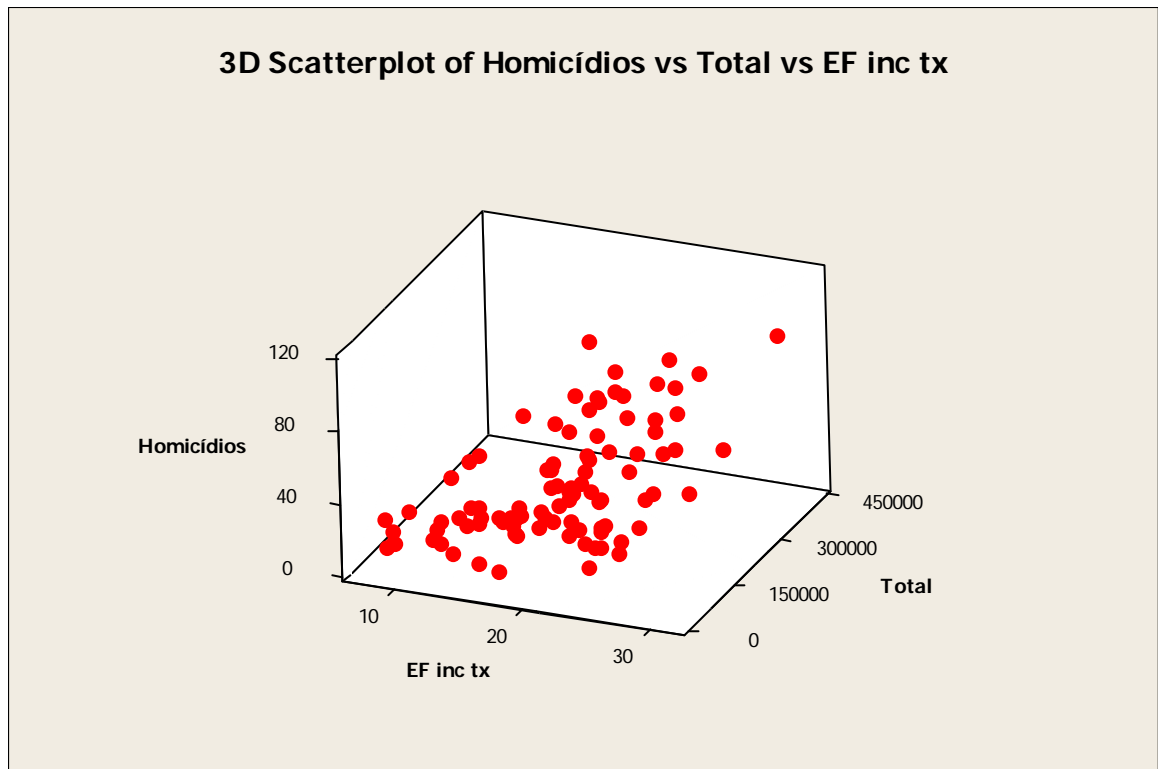
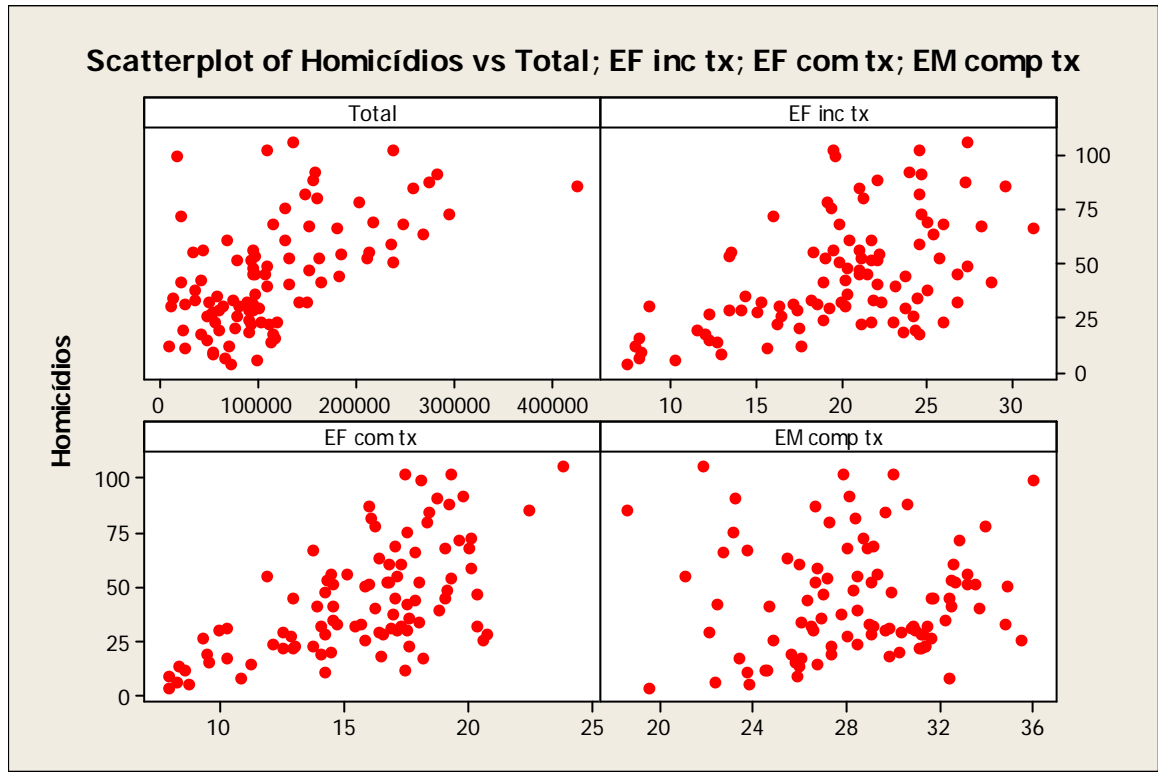
- **Forma:** O Histograma nos permite verificar que se trata de uma distribuição assimétrica à esquerda. O intervalo de confiança dos retornos está entre 44,387 e 39,859. Através da análise do teste estatístico de Anderson-Darling medimos quão bem os dados seguem uma distribuição específica, considerando 5% como parâmetro no teste de hipótese, concluímos alto nível de significância para a variável que, portanto, segue uma distribuição específica.
- **Valores Atípicos:** Não há valores atípicos.
- **Centro e Dispersão:** É referente à distribuição dos dados que está espalhada em torno de sua média. O valor do desvio-padrão 25,196 demonstrando relativa variação em número de homicídios nos distritos estudados.

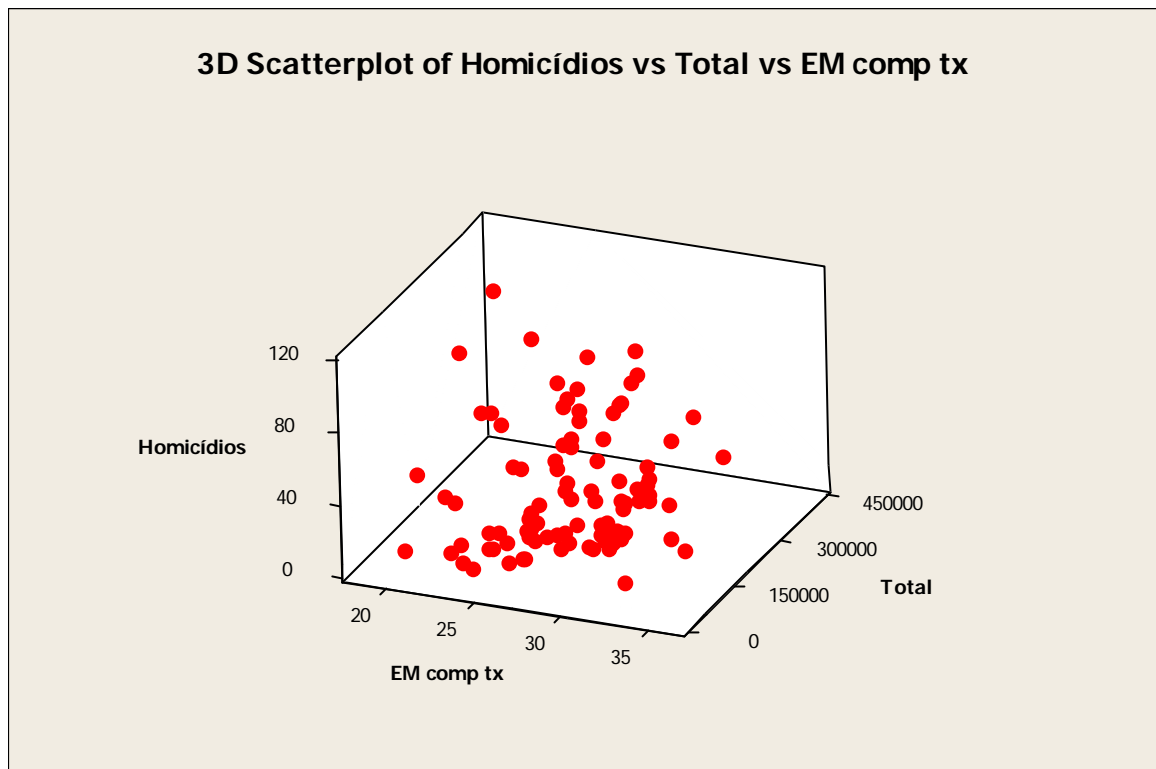
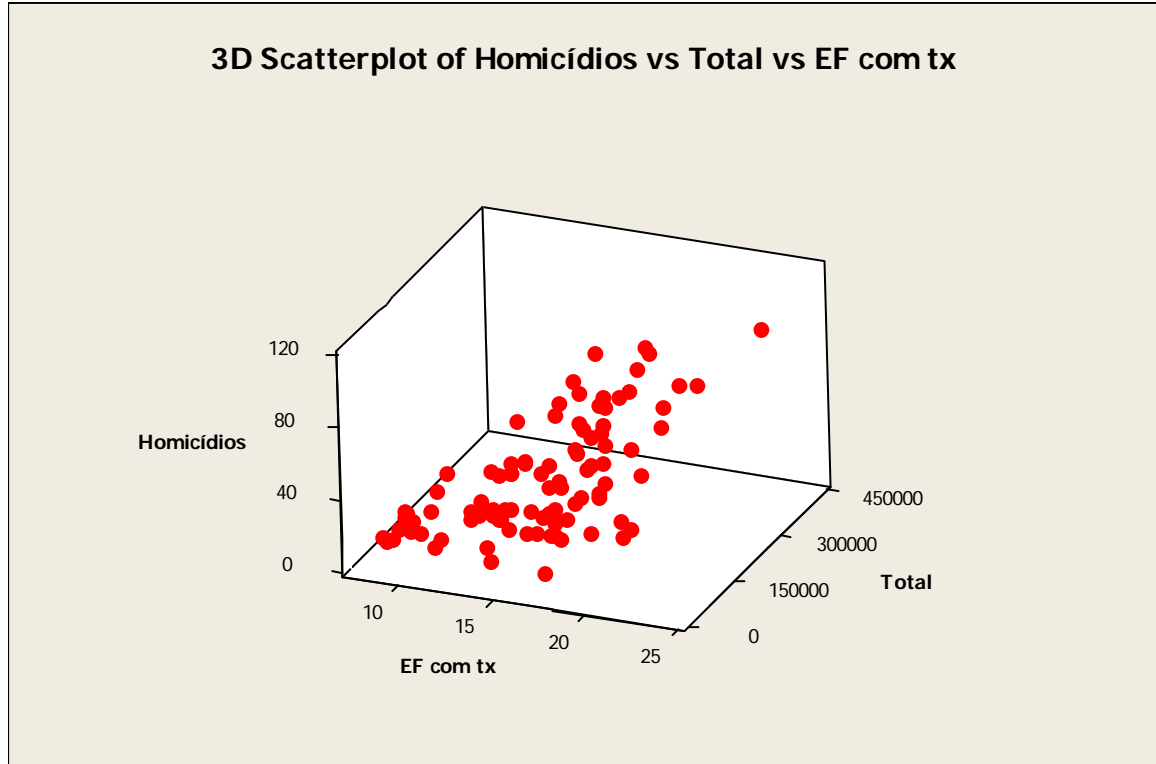
#### 4. CORRELAÇÃO ENTRE VARIÁVEIS

Através da correlação gerada através do software Minitab 14, é possível avaliar que existe uma forte correlação entre todas as variáveis e que todas apresentam alto grau de confiança.

É plenamente observável a relação existente entre o número de homicídios em relação as demais variáveis, expressando de forma significativa alguns dos fatores que podem conduzir a este típico específico de violência.

Os gráficos de dispersão que seguem demonstram essas correlações:





Segue abaixo uma breve análise dos gráficos acima:

Direção: Grande parte das dispersões acima nos permitem verificar um padrão linear na maioria dos gráficos, demonstrando uma maior acentuação na parte inferior. Nos

mostrando que há uma relação, ao menos de evidência visual. Em muitos casos o aumento de X não é acompanhado por um forte aumento de Y.

Intensidade: A intensidade de uma relação linear é mais facilmente observada mediante o cálculo de índices de correlação, que veremos mais adiante. Alguns gráficos acima parecem indicar a existência de relações lineares.

Forma: De maneira visual apenas, os gráficos acima parecem denotar forma linear crescente.

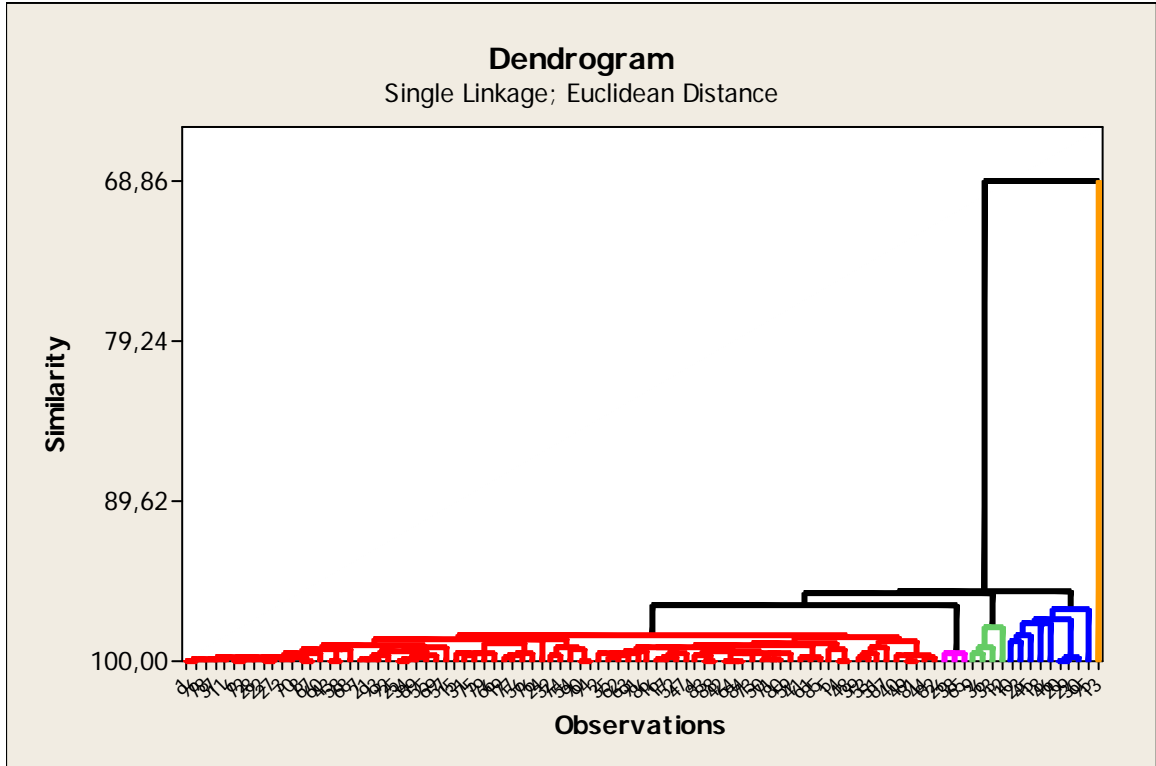
Valores Atípicos: Alguns gráficos indicam a existência de valores atípicos, ou seja, indivíduos ou estados, que estão localizados longe dos demais. Isto pode ser verificado também na análise exploratória que vimos no item anterior.

A matriz de correlação abaixo nos ajuda tirar conclusões mais precisas sobre a associação entre as variáveis.

**Parelheiros é o distrito que apresentou o maior índice de homicídios na cidade de São Paulo, no ano de 2007.**

## 5 - Dendograma

Dendogramas são estruturas gráficas em forma de árvore, utilizadas para representar as junções (métodos hierárquicos) ou divisões (métodos de partição) que ocorreram a partir de valores provenientes da matriz de distâncias (JOHNSON & WICHERN, 1988). De acordo com Bussab et al (1990), para construirmos um dendograma utilizando os valores da matriz de distâncias com o objetivo de ilustrar as junções, devemos colocar no eixo horizontal os elementos, em uma ordem conveniente de acordo com os grupos formados, de onde partirá de cada um desses elementos uma linha vertical até a altura correspondente ao nível (o valor da distância) em que ocorreu a junção (a um outro elemento ou grupo). Essa altura é marcada no eixo vertical.



**Cluster Analysis of Observations: Total; EF inc tx; EF com tx; EM comp tx; ...**

Euclidean Distance, Single Linkage  
Amalgamation Steps

Step	Number of clusters	Similarity level	Distance level	Clusters joined		New cluster	Number of obs. in new cluster
1	95	99,9927	30	64	84	64	2
2	94	99,9821	74	20	29	20	2
3	93	99,9803	81	8	67	8	2
4	92	99,9757	101	90	94	90	2
5	91	99,9736	109	25	34	25	2
6	90	99,9666	139	6	72	6	2
7	89	99,9639	150	43	58	43	2
8	88	99,9614	160	49	81	49	2
9	87	99,9539	191	1	96	1	2
10	86	99,9529	195	14	48	14	2
11	85	99,9468	221	83	88	83	2
12	84	99,9462	223	42	64	42	3
13	83	99,9374	260	22	27	22	2
14	82	99,9344	272	6	28	6	3
15	81	99,9319	282	1	78	1	3
16	80	99,9030	402	71	80	71	2
17	79	99,8838	482	16	17	16	2
18	78	99,8837	482	3	70	3	2
19	77	99,8786	503	50	71	50	3
20	76	99,8784	504	25	89	25	3
21	75	99,8726	528	6	22	6	5
22	74	99,8674	550	1	37	1	4
23	73	99,8551	600	1	11	1	5
24	72	99,8080	796	7	21	7	2
25	71	99,8065	802	7	93	7	3
26	70	99,7982	836	19	77	19	2
27	69	99,7975	839	44	82	44	2
28	68	99,7924	860	1	6	1	10
29	67	99,7818	904	41	61	41	2

30	66	99,7797	913	16	52	16	3
31	65	99,7151	1181	1	3	1	12
32	64	99,7126	1191	35	62	35	2
33	63	99,7087	1207	39	53	39	2
34	62	99,7066	1216	41	85	41	3
35	61	99,6977	1253	1	8	1	14
36	60	99,6934	1271	25	55	25	4
37	59	99,6820	1318	20	30	20	3
38	58	99,6733	1354	7	92	7	4
39	57	99,6651	1388	4	83	4	3
40	56	99,6446	1473	40	49	40	3
41	55	99,6227	1564	35	63	35	3
42	54	99,6193	1578	25	69	25	5
43	53	99,6189	1579	19	56	19	3
44	52	99,6159	1592	32	74	32	2
45	51	99,5775	1751	40	44	40	5
46	50	99,5773	1752	31	75	31	2
47	49	99,5404	1905	42	73	42	4
48	48	99,5334	1934	26	38	26	2
49	47	99,5143	2013	2	35	2	4
50	46	99,5013	2067	1	60	1	15
51	45	99,4842	2138	50	59	50	4
52	44	99,4827	2144	15	31	15	3
53	43	99,4743	2179	15	79	15	4
54	42	99,4600	2238	39	51	39	3
55	41	99,4581	2246	9	36	9	2
56	40	99,4560	2255	15	66	15	5
57	39	99,4280	2371	16	47	16	4
58	38	99,4272	2374	26	65	26	3
59	37	99,4224	2394	42	50	42	8
60	36	99,4223	2395	2	91	2	5
61	35	99,4175	2415	19	76	19	4
62	34	99,3609	2649	7	25	7	9
63	33	99,3179	2827	2	86	2	6
64	32	99,2948	2923	4	42	4	11
65	31	99,2741	3009	43	68	43	3
66	30	99,2623	3058	54	90	54	3
67	29	99,2388	3155	1	43	1	18
68	28	99,1582	3489	5	14	5	3
69	27	99,1568	3495	9	33	9	3
70	26	99,1081	3697	2	16	2	10
71	25	99,0992	3734	39	87	39	4
72	24	99,0950	3751	7	57	7	10
73	23	99,0714	3849	2	4	2	21
74	22	99,0446	3960	32	54	32	5
75	21	98,9252	4455	1	7	1	28
76	20	98,9184	4483	2	41	2	24
77	19	98,8960	4576	2	5	2	27
78	18	98,8610	4721	15	19	15	9
79	17	98,7459	5198	39	40	39	9
80	16	98,6475	5606	10	23	10	2
81	15	98,6064	5776	15	24	15	10
82	14	98,5680	5935	15	32	15	15
83	13	98,5102	6175	1	15	1	43
84	12	98,4448	6446	2	39	2	36
85	11	98,3346	6903	1	2	1	79
86	10	98,2704	7169	10	45	10	3
87	9	97,8072	9089	9	12	9	4
88	8	97,5942	9972	18	46	18	2
89	7	97,5816	10024	10	18	10	5
90	6	97,3056	11168	10	20	10	8
91	5	96,6450	13906	10	95	10	9
92	4	96,3903	14962	1	26	1	82
93	3	95,5415	18480	1	9	1	86
94	2	95,5056	18629	1	10	1	95
95	1	68,8616	129066	1	13	1	96

Final Partition  
Number of clusters: 5

	Number of observations	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster1	79	1,33721E+11	34209,2	79257,2
Cluster2	4	1,20310E+08	4406,8	8813,5
Cluster3	9	3,70218E+09	17902,6	35555,9
Cluster4	1	0,00000E+00	0,0	0,0
Cluster5	3	9,31182E+06	1485,0	2227,3

Cluster Centroids

Variable	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5	Grand centroid
Total	85401,8	211223	259528	424150	181702	113507
EF inc tx	19,0	22	23	30	26	20
EF com tx	15,1	17	18	22	18	16
EM comp tx	28,5	29	28	19	25	28
Homicídios	38,9	64	76	86	55	44

Distances Between Cluster Centroids

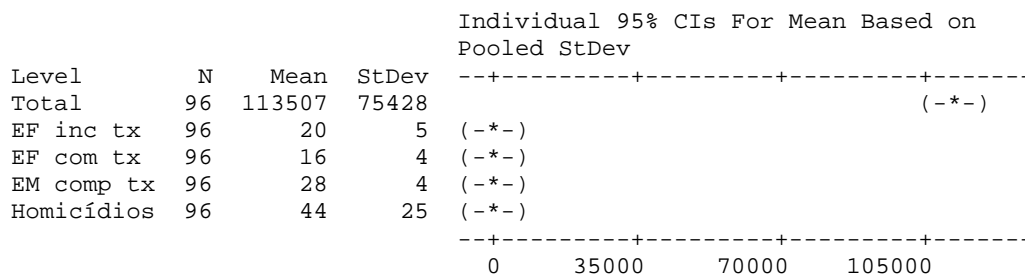
	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5
Cluster1	0	125821	174126	338748	96300
Cluster2	125821	0	48306	212928	29521
Cluster3	174126	48306	0	164622	77826
Cluster4	338748	212928	164622	0	242448
Cluster5	96300	29521	77826	242448	0

**6 – ANOVA**

**One-way ANOVA: Total; EF inc tx; EF com tx; EM comp tx; Homicídios**

Source	DF	SS	MS	F	P
Factor	4	9,89002E+11	2,47250E+11	217,29	0,000
Error	475	5,40498E+11	1137890786		
Total	479	1,52950E+12			

S = 33733 R-Sq = 64,66% R-Sq(adj) = 64,36%



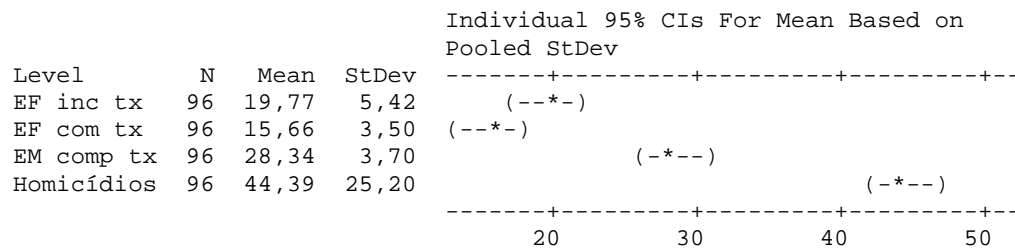
Pooled StDev = 33733

**One-way ANOVA: EF inc tx; EF com tx; EM comp tx; Homicídios**

Source	DF	SS	MS	F	P
Factor	3	46560	15520	89,95	0,000
Error	380	65563	173		

Total 383 112123

S = 13,14 R-Sq = 41,53% R-Sq(adj) = 41,06%



Pooled StDev = 13,14

## 7- Principal Component Analysis: Total; EF inc tx; EF com tx; EM comp tx; Homicídi

Eigenanalysis of the Correlation Matrix

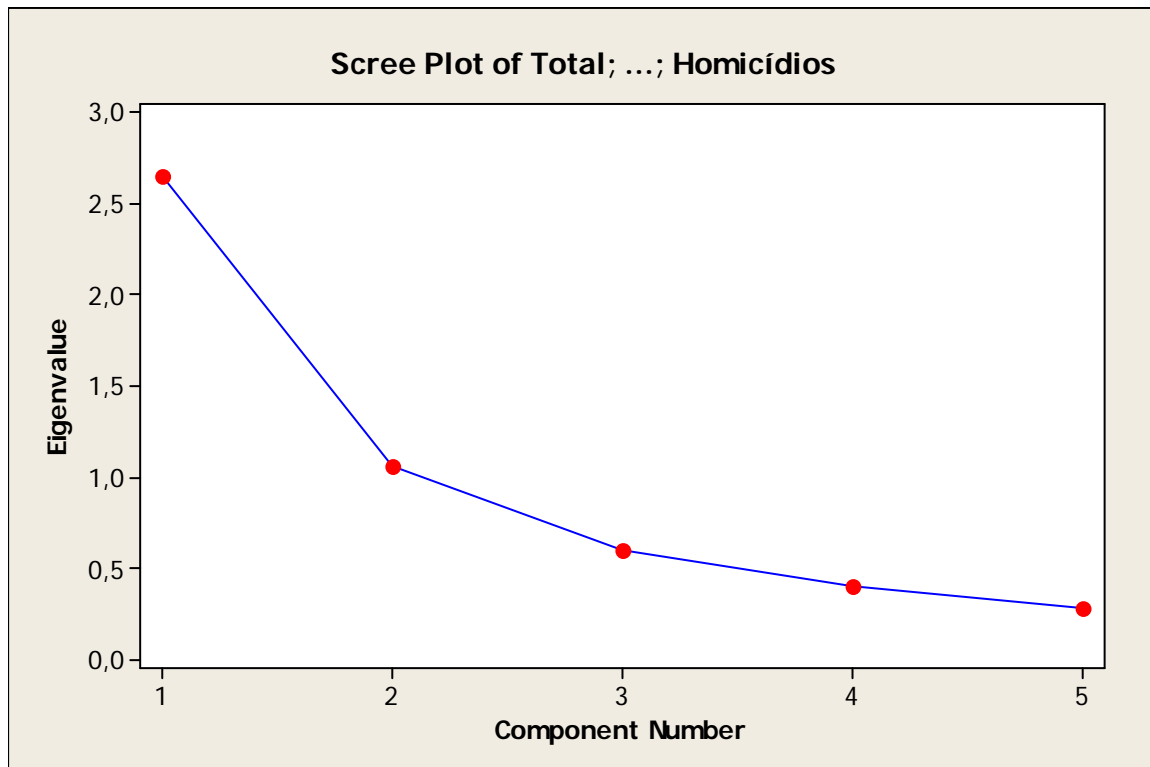
Eigenvalue	2,6576	1,0608	0,5988	0,3992	0,2836
Proportion	0,532	0,212	0,120	0,080	0,057
Cumulative	0,532	0,744	0,863	0,943	1,000

Variable	PC1	PC2	PC3	PC4	PC5
Total	0,457	-0,274	0,667	0,440	0,279
EF inc tx	0,506	0,011	-0,489	0,526	-0,477
EF com tx	0,517	0,177	-0,404	-0,251	0,690
EM comp tx	-0,010	0,943	0,253	0,218	-0,007
Homicídios	0,518	0,072	0,297	-0,648	-0,468

Dois fatores podem ser observados como fatores de análise: 1) em PC1 identifica-se a relação entre homicídios e o percentual ensino médio completo; 2) em PC2 verifica-se a relação existente entre população total e o percentual ensino médio completo; 3) Identifica-se a relação entre população total e percentual ensino fundamental incompleto; 4) em PC4 está estabelecida entre população total e homicídios; 5) em PC5 a relação está entre percentual ensino fundamental completo e homicídios.

Das relações estabelecidas o que se pode compreender é que, pelo sistema educacional adotado na cidade de São Paulo, a população tende a ter um maior nível de escolaridade, com baixa evasão e repetência. Além disso, os níveis de homicídios tende a ser menor quanto maior for o nível de escolaridade da população, enquanto que quanto mais populoso for o distrito maior a probabilidade de se ter homicídios, tendo em vista o baixo nível de escolaridade observado em PC4.

Scree Plot of Pop.total;...;Homicídios



Pela análise dos detalhes e do gráfico acima tem-se que PC1 apresenta índice de análise bastante expressivo. Isto é algo extremamente significativo, pois ao invés de se trabalhar com 5 variáveis poderia trabalhar com o índice PC1, que já explica grande parte das variáveis.

Pela análise dos valores / participação de cada variável no índice PC1, poderíamos denominá-lo “média”, uma vez que as participações são bastante equitativas e possuem o mesmo sinal.

## 8 - Análise de Regressão

### Regression Analysis: Homicídios versus Total; EF inc tx; ...

The regression equation is

$$\text{Homicídios} = -37,7 + 0,000124 \text{ Total} + 0,470 \text{ EF inc tx} + 2,84 \text{ EF com tx} + 0,504 \text{ EM comp tx}$$

Predictor	Coef	SE Coef	T	P
Constant	-37,70	16,18	-2,33	0,022
Total	0,00012414	0,00002839	4,37	0,000
EF inc tx	0,4701	0,4748	0,99	0,325
EF com tx	2,8364	0,7248	3,91	0,000
EM comp tx	0,5042	0,5079	0,99	0,323

S = 17,8665    R-Sq = 51,8%    **R-Sq(adj) = 49,7%**

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	4	31259,1	7814,8	24,48	0,000
Residual Error	91	29048,2	319,2		

Total 95 60307,3

Source	DF	Seq SS
Total	1	19991,5
EF inc tx	1	5509,3
EF com tx	1	5443,7
EM comp tx	1	314,6

#### Unusual Observations

Obs	Total	Homicídios	Fit	SE Fit	Residual	St Resid
5	32700	54,93	16,83	4,95	38,10	2,22R
13	424150	85,77	101,72	8,74	-15,94	-1,02 X
25	108757	102,49	50,89	2,65	51,60	2,92R
54	135068	105,99	70,21	6,25	35,79	2,14R
87	16837	99,43	42,94	4,97	56,48	3,29R

R denotes an observation with a large standardized residual.  
X denotes an observation whose X value gives it large leverage.

## 9 – Regressão Stepwise

### Stepwise Regression: Homicídios versus Total; EF inc tx; ...

Alpha-to-Enter: 0,15 Alpha-to-Remove: 0,15

Response is Homicídios on 4 predictors, with N = 96

Step	1	2
Constant	-26,12	-22,35
EF com tx	4,50	3,35
T-Value	7,77	5,79
P-Value	0,000	0,000
Total		0,00013
T-Value		4,71
P-Value		0,000
S	19,8	17,9
R-Sq	39,12	50,85
R-Sq(adj)	38,48	49,79
Mallows Cp	23,0	2,9

## 10 – Correlação

### Correlations: Total; EF inc tx; EF com tx; EM comp tx; Homicídios

	Total	EF inc tx	EF com tx	EM comp tx
EF inc tx	0,471 0,000			
EF com tx	0,425 0,000	0,670 0,000		
EM comp tx	-0,147 0,153	-0,029 0,776	0,079 0,444	
Homicídios	0,576 0,000	0,538 0,000	0,625 0,000	0,048 0,645

Cell Contents: Pearson correlation  
P-Value

## 11 – Análise de Principal Componentes

### Principal Component Analysis: Total; EF inc tx; EF com tx; EM comp tx; Homicídi

Eigenanalysis of the Correlation Matrix

Eigenvalue	2,6576	1,0608	0,5988	0,3992	0,2836
Proportion	0,532	0,212	0,120	0,080	0,057
Cumulative	0,532	0,744	0,863	0,943	1,000

Variable	PC1	PC2	PC3	PC4	PC5
Total	0,457	-0,274	0,667	0,440	0,279
EF inc tx	0,506	0,011	-0,489	0,526	-0,477
EF com tx	0,517	0,177	-0,404	-0,251	0,690
EM comp tx	-0,010	0,943	0,253	0,218	-0,007
Homicídios	0,518	0,072	0,297	-0,648	-0,468

## 12 – Análise Discriminante

### Discriminant Analysis: C14 versus Total; EF inc tx; ...

Linear Method for Response: C14

Predictors: Total; EF inc tx; EF com tx; EM comp tx; Homicídios

Group	1	2	3	5
Count	79	4	9	4

Summary of classification

	True Group			
Put into Group	1	2	3	5
1	75	0	0	0
2	4	3	1	1
3	0	0	7	0
5	0	1	1	3
Total N	79	4	9	4
N correct	75	3	7	3
Proportion	0,949	0,750	0,778	0,750

N = 96

N Correct = 88

Proportion Correct = 0,917

Squared Distance Between Groups

	1	2	3	5
1	0,0000	8,6724	16,3479	14,0662
2	8,6724	0,0000	1,5245	4,3969
3	16,3479	1,5245	0,0000	3,6364
5	14,0662	4,3969	3,6364	0,0000

Linear Discriminant Function for Groups

	1	2	3	5
Constant	-40,677	-52,641	-56,287	-48,256
Total	0,000	0,000	0,000	0,000
EF inc tx	0,355	0,237	0,116	0,351
EF com tx	0,912	0,848	1,004	1,171

EM comp tx	2,146	2,301	2,208	1,834
Homicídios	-0,109	-0,094	-0,087	-0,134

#### Summary of Misclassified Observations

Observation	True Group	Pred Group	Group	Squared Distance	Probability
9**	2	5	1	8,671	0,015
			2	2,043	0,416
			3	4,091	0,149
			5	2,029	0,419
15**	1	2	1	6,349	0,244
			2	4,494	0,617
			3	7,808	0,118
			5	11,173	0,022
29**	3	2	1	18,316	0,001
			2	4,142	0,638
			3	5,319	0,354
			5	13,195	0,007
30**	3	5	1	13,039	0,001
			2	3,011	0,210
			3	2,421	0,282
			5	1,244	0,507
31**	1	2	1	5,367	0,257
			2	3,959	0,520
			3	6,663	0,135
			5	7,526	0,087
65**	5	2	1	6,339	0,068
			2	2,767	0,408
			3	4,591	0,164
			5	3,021	0,359
75**	1	2	1	6,509	0,311
			2	5,551	0,502
			3	8,544	0,112
			5	9,370	0,074
79**	1	2	1	3,206	0,249
			2	1,502	0,584
			3	5,271	0,089
			5	5,530	0,078

## 13 – Regressão Logística Ordinal

### Ordinal Logistic Regression: Homicídios versus Total; EF inc tx; ...

Link Function: Logit

#### Response Information

Variable	Value	Count
Homicídios	3,586	1
	5,857	1
	7,015	1
	8,308	1
	9,524	1
	11,165	1
	11,899	1
	12,277	1
	13,550	1
	14,673	1
	16,170	1
	17,896	1
	17,996	1
	18,109	1
	19,445	1
	19,506	1

20,156	1
22,140	1
22,462	1
23,275	1
23,332	1
23,421	1
24,033	1
26,008	1
26,307	1
26,429	1
27,475	1
28,402	1
28,486	1
28,769	1
29,145	1
29,871	1
30,269	1
30,852	1
30,908	1
31,340	1
31,596	1
32,289	1
32,431	1
32,537	1
32,741	1
33,251	1
33,259	1
33,729	1
34,766	1
36,414	1
37,858	1
39,527	1
40,191	1
41,356	1
41,410	1
42,374	1
44,566	1
44,954	1
44,960	1
45,352	1
47,087	1
47,997	1
48,561	1
50,814	1
51,191	1
51,193	1
52,128	1
52,212	1
52,645	1
53,626	1
54,147	1
54,932	1
55,116	1
56,484	1
56,582	1
58,759	1
60,766	1
60,874	1
63,955	1
65,917	1
66,855	1
68,208	1
68,328	1
69,475	1
71,548	1
72,988	1
75,653	1
77,852	1
79,873	1
81,543	1

84,462	1
85,774	1
87,333	1
88,047	1
90,722	1
91,702	1
99,428	1
102,278	1
102,490	1
105,994	1
Total	96

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds	95% CI	
					Ratio	Lower	Upper
Const (1)	5,24894	1,90339	2,76	0,006			
Const (2)	6,09392	1,77603	3,43	0,001			
Const (3)	6,63651	1,74661	3,80	0,000			
Const (4)	7,04456	1,74223	4,04	0,000			
Const (5)	7,37776	1,74722	4,22	0,000			
Const (6)	7,66910	1,75675	4,37	0,000			
Const (7)	7,90336	1,76737	4,47	0,000			
Const (8)	8,11858	1,77913	4,56	0,000			
Const (9)	8,33309	1,79252	4,65	0,000			
Const (10)	8,53698	1,80651	4,73	0,000			
Const (11)	8,73587	1,82110	4,80	0,000			
Const (12)	8,91840	1,83507	4,86	0,000			
Const (13)	9,09176	1,84868	4,92	0,000			
Const (14)	9,25813	1,86187	4,97	0,000			
Const (15)	9,40460	1,87346	5,02	0,000			
Const (16)	9,54936	1,88481	5,07	0,000			
Const (17)	9,69240	1,89585	5,11	0,000			
Const (18)	9,82272	1,90570	5,15	0,000			
Const (19)	9,94323	1,91459	5,19	0,000			
Const (20)	10,0552	1,92265	5,23	0,000			
Const (21)	10,1585	1,92991	5,26	0,000			
Const (22)	10,2538	1,93645	5,30	0,000			
Const (23)	10,3448	1,94255	5,33	0,000			
Const (24)	10,4332	1,94834	5,35	0,000			
Const (25)	10,5166	1,95368	5,38	0,000			
Const (26)	10,5989	1,95884	5,41	0,000			
Const (27)	10,6844	1,96409	5,44	0,000			
Const (28)	10,7680	1,96911	5,47	0,000			
Const (29)	10,8488	1,97386	5,50	0,000			
Const (30)	10,9282	1,97844	5,52	0,000			
Const (31)	11,0051	1,98279	5,55	0,000			
Const (32)	11,0809	1,98700	5,58	0,000			
Const (33)	11,1554	1,99107	5,60	0,000			
Const (34)	11,2312	1,99514	5,63	0,000			
Const (35)	11,3086	1,99922	5,66	0,000			
Const (36)	11,3846	2,00317	5,68	0,000			
Const (37)	11,4619	2,00711	5,71	0,000			
Const (38)	11,5418	2,01112	5,74	0,000			
Const (39)	11,6206	2,01502	5,77	0,000			
Const (40)	11,6952	2,01864	5,79	0,000			
Const (41)	11,7666	2,02206	5,82	0,000			
Const (42)	11,8379	2,02542	5,84	0,000			
Const (43)	11,9093	2,02874	5,87	0,000			
Const (44)	11,9797	2,03196	5,90	0,000			
Const (45)	12,0513	2,03519	5,92	0,000			
Const (46)	12,1236	2,03840	5,95	0,000			
Const (47)	12,1957	2,04155	5,97	0,000			
Const (48)	12,2666	2,04459	6,00	0,000			
Const (49)	12,3350	2,04749	6,02	0,000			
Const (50)	12,4035	2,05034	6,05	0,000			
Const (51)	12,4731	2,05319	6,07	0,000			
Const (52)	12,5442	2,05605	6,10	0,000			
Const (53)	12,6157	2,05887	6,13	0,000			

Const (54)	12,6858	2,06159	6,15	0,000			
Const (55)	12,7559	2,06426	6,18	0,000			
Const (56)	12,8255	2,06686	6,21	0,000			
Const (57)	12,8932	2,06934	6,23	0,000			
Const (58)	12,9611	2,07179	6,26	0,000			
Const (59)	13,0294	2,07420	6,28	0,000			
Const (60)	13,0955	2,07649	6,31	0,000			
Const (61)	13,1613	2,07873	6,33	0,000			
Const (62)	13,2284	2,08096	6,36	0,000			
Const (63)	13,2960	2,08318	6,38	0,000			
Const (64)	13,3632	2,08533	6,41	0,000			
Const (65)	13,4299	2,08743	6,43	0,000			
Const (66)	13,4983	2,08955	6,46	0,000			
Const (67)	13,5680	2,09166	6,49	0,000			
Const (68)	13,6397	2,09379	6,51	0,000			
Const (69)	13,7144	2,09597	6,54	0,000			
Const (70)	13,7913	2,09818	6,57	0,000			
Const (71)	13,8728	2,10047	6,60	0,000			
Const (72)	13,9556	2,10276	6,64	0,000			
Const (73)	14,0376	2,10499	6,67	0,000			
Const (74)	14,1235	2,10729	6,70	0,000			
Const (75)	14,2128	2,10965	6,74	0,000			
Const (76)	14,3031	2,11201	6,77	0,000			
Const (77)	14,3980	2,11447	6,81	0,000			
Const (78)	14,4956	2,11699	6,85	0,000			
Const (79)	14,5941	2,11953	6,89	0,000			
Const (80)	14,6957	2,12216	6,92	0,000			
Const (81)	14,8030	2,12495	6,97	0,000			
Const (82)	14,9122	2,12783	7,01	0,000			
Const (83)	15,0242	2,13083	7,05	0,000			
Const (84)	15,1460	2,13419	7,10	0,000			
Const (85)	15,2782	2,13797	7,15	0,000			
Const (86)	15,4257	2,14237	7,20	0,000			
Const (87)	15,5871	2,14750	7,26	0,000			
Const (88)	15,7460	2,15291	7,31	0,000			
Const (89)	15,9055	2,15879	7,37	0,000			
Const (90)	16,0845	2,16603	7,43	0,000			
Const (91)	16,2894	2,17532	7,49	0,000			
Const (92)	16,5316	2,18801	7,56	0,000			
Const (93)	16,8407	2,20764	7,63	0,000			
Const (94)	17,2677	2,24348	7,70	0,000			
Const (95)	17,9821	2,33995	7,68	0,000			
Total	-0,0000103	0,0000030	-3,48	0,001	1,00	1,00	1,00
EF inc tx	-0,105180	0,0486788	-2,16	0,031	0,90	0,82	0,99
EF com tx	-0,309753	0,0755775	-4,10	0,000	0,73	0,63	0,85
EM comp tx	-0,139565	0,0522901	-2,67	0,008	0,87	0,79	0,96

Log-Likelihood = -397,991

Test that all slopes are zero: G = 80,372, DF = 4, P-Value = 0,000

#### Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	9917,59	9021	0,000
Deviance	795,98	9021	1,000

#### Measures of Association:

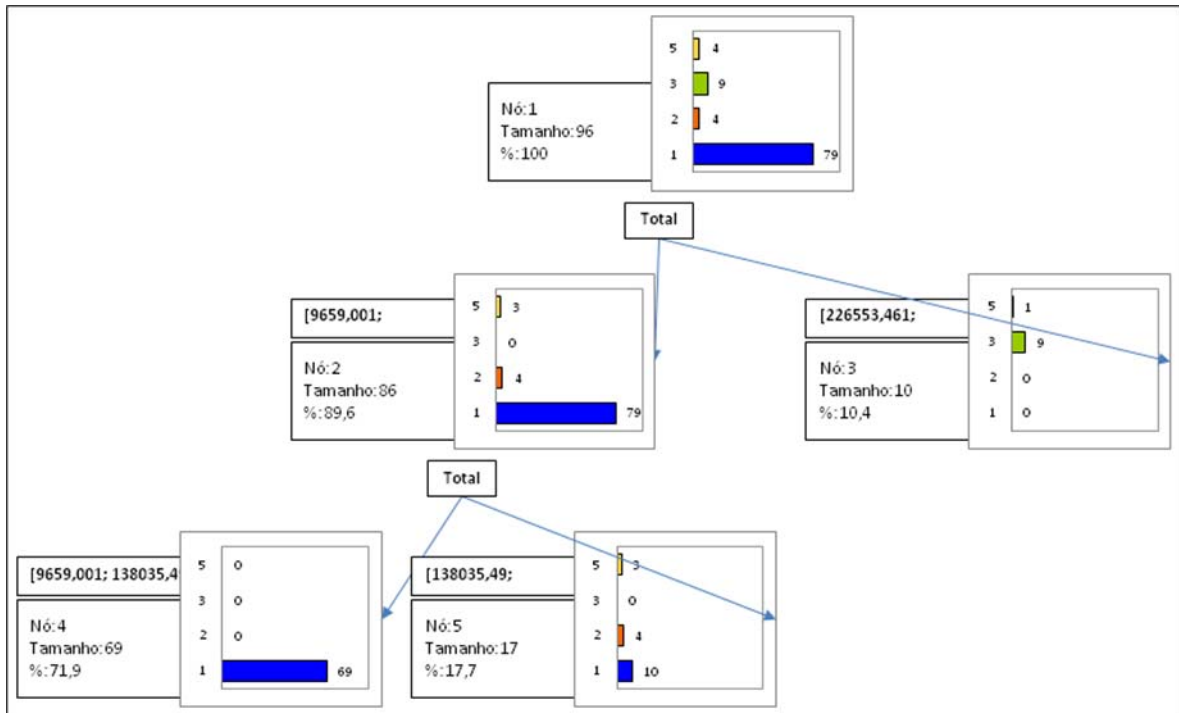
(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures
Concordant	3539	77,6	Somers' D 0,56
Discordant	1003	22,0	Goodman-Kruskal Gamma 0,56
Ties	18	0,4	Kendall's Tau-a 0,56
Total	4560	100,0	

#### Resultado das analises

REGRESSÃO LOGÍSTICA	77,6
ANALISE DE REGRESSÃO	49,7
ANALISE DISCRIMINANTE	91,7

**14 – Arvore de Decisão**



**15 – Conclusão**

Vimos neste trabalho a relação existente entre o número de habitantes de uma região e as características que implicam no número de homicídios na cidade de São Paulo. Os distritos de Parelheiros e Grajáú são os mais populosos, com um número de população com baixa escolaridade, em comparação aos demais distritos estudados.

Tem-se como fator de observação as implicações destas características no número de homicídios em Parelheiros, ocorridos no ano de 2007.

Esses fatores de análise levam a conclusão de que há uma forte relação entre as características de um bairro e o nível de homicídios, deduzindo-se que quanto menos populoso for um bairro, quanto maior o nível de escolaridade da população, menores serão os índices de criminalidade.

Compreende-se que há a necessidade de uma melhor distribuição territorial da população, diminuindo-se a concentração populacional e promovendo ações educacionais, capazes de promover o desenvolvimento sócio-educacional das populações de modo a diminuir ou extinguir todo o tipo de violência, em especial o número de homicídios na cidade de São Paulo, especificamente nas regiões mais periféricas do município paulista.

Ao mesmo tempo que o poder público deve assumir o seu papel de responsabilidade político, econômico, social, educacional, ambiental, cabe aos agentes comunitários estabelecerem participações de desenvolvimento cultural, esportivo, econômico, de modo a possibilitar a melhoria das comunidades inibindo pacífica e solidariamente a violência nas regiões mais carentes do município de São Paulo.