



INDICADORES BÁSICOS NOSSA SÃO PAULO
UM ESTUDO DE SUAS RELAÇÕES

Fabiola Dapuzzo Vinhas e Nathalie Perret

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO
FEA - Faculdade de Economia e Administração
Programa de Estudos Pós-Graduados em Administração

SUMÁRIO

RESUMO	3
1. INTRODUÇÃO.....	3
2. ENTENDENDO OS DADOS	4
1.1 A importância dos Indicadores.....	4
1.2 A Classificação dos Indicadores	4
1.3 Os Indivíduos	6
1.4 As Variáveis	6
3. ANÁLISE INDIVIDUAL DAS VARIÁVEIS.....	8
3.1 Variáveis categóricas	8
3.2 Variáveis Quantitativas	9
4. RELAÇÕES ENTRE VARIÁVEIS	15
4.1 Dendogramas.....	15
4.2 Gráfico de Regressões.....	15
4.3 Correlações.....	16
5. ANÁLISE DE CLUSTERS.....	23
5.1 Cluster de Variáveis	23
5.2 Cluster de Observações	24
6. ANÁLISE DE COMPONENTES PRINCIPAIS	25
7. CONSIDERAÇÕES FINAIS	28

RESUMO

Indicadores são percentuais, índices, informações qualificadas que servem como instrumentos fundamentais para avaliar e analisar determinadas realidades. Com eles é possível monitorar a qualidade de vida na cidade, as políticas sociais e a gestão pública municipal, visando dotar a sociedade civil de importantes recursos para exercer uma real influência nas metas das políticas públicas, de modo que a prioridade seja a construção de uma cidade justa e sustentável. A Lei N 14.173, publicada em 26 de junho de 2006, compromete a Prefeitura de São Paulo com o estabelecimento de indicadores de desempenho relativos à qualidade dos serviços públicos no Município de São Paulo. Nesse sentido, o presente trabalho visa efetuar uma análise exploratória de alguns dos indicadores selecionados para termos uma visão global de como se encontra a situação de qualidade de vida da população de São Paulo.

Palavras-chaves: indicadores, Movimento Nossa São Paulo

1. INTRODUÇÃO

O Movimento Nossa São Paulo foi lançado em maio de 2007 a partir da percepção de que a atividade política no Brasil, as instituições públicas e a democracia estão com a credibilidade abalada perante a população. Constatamos que é necessário promover iniciativas que possam recuperar para a sociedade os valores do desenvolvimento sustentável, da ética e da democracia participativa.

O Movimento pretende construir uma força política, social e econômica capaz de comprometer a sociedade e sucessivos governos com uma agenda e um conjunto de metas a fim de oferecer melhor qualidade de vida para todos os habitantes da cidade.

Atualmente, 586 organizações da sociedade civil integram o movimento, que é apartidário e inter-religioso, não tem presidente nem diretoria, se constituiu e se expande na forma de rede.

O Movimento espera contar com a participação de toda a sociedade para reunir idéias e propor ações que possam contribuir para o desenvolvimento justo e sustentável da cidade em áreas essenciais como Educação, Meio Ambiente, Segurança, Lazer e Cultura, Trabalho, Transporte, Moradia, Saúde e Serviços.

A missão do NOSSA SÃO PAULO é mobilizar diversos segmentos da sociedade para, em parceria com instituições públicas e privadas, construir e se comprometer com uma agenda e um conjunto de metas, articular e promover ações, visando a uma cidade de São Paulo justa e sustentável.

Para isso, criou o Observatório Cidadão Nossa São Paulo que disponibiliza um conjunto de indicadores sociais, ambientais, econômicos, políticos e culturais sobre a cidade de São Paulo e cada uma de suas 31 subprefeituras, que serão constantemente atualizados, avaliados e divulgados para toda a sociedade.

A maioria dos indicadores foi selecionada por meio de um trabalho coletivo que reuniu dezenas de pessoas e organizações da sociedade civil em 14 Grupos de Trabalho do Movimento, aos quais agregamos alguns outros importantes e os previstos em leis.

Para o trabalho de monitoramento que o Movimento Nossa São Paulo passa a exercer é fundamental a prioridade que deve ser dada à produção e divulgação de informações para

cada área da administração pública municipal e para cada subprefeitura, com transparência metodológica e periodicidade regular.

O presente trabalho tem por objetivo efetuar uma análise exploratória de alguns dos indicadores selecionados dentre as 13 áreas temáticas: cultura, educação ensino médio, educação ensino fundamental, educação infantil, esporte, habitação, meio ambiente, orçamento, saúde, trabalho e renda, transporte: acidentes de trânsito, transporte: mobilidade urbana e violência, para termos uma visão global de como se encontra a situação de qualidade de vida da população de São Paulo.

O software estatístico utilizado neste trabalho é o MINITAB 14.

2. ENTENDENDO OS DADOS

1.1 A importância dos Indicadores

Os indicadores são percentuais, índices, informações qualificadas que servem como instrumentos fundamentais para avaliar e analisar determinadas realidades. Com eles é possível monitorar a qualidade de vida na cidade, as políticas sociais e a gestão pública municipal, visando dotar a sociedade civil de importantes recursos para exercer uma real influência nas metas das políticas públicas, de modo que a prioridade seja a construção de uma cidade justa e sustentável.

Diante da complexidade da cidade de São Paulo a Nossa São Paulo optou por não trabalhar apenas com as médias da cidade como um todo, pois ela é extremamente desigual e as médias escondem onde estão as prioridades e as necessidades de maiores recursos e investimentos sociais. Por isso a decisão pela segmentação regional, estabelecendo indicadores por subprefeituras. São 31 subprefeituras, sendo elas: Aricanduva; Butantã; Campo Limpo; Capela do Socorro; Casa Verde/Cachoeirinha; Cidade Ademar; Cidade Tiradentes; Ermelino Matarazzo; Freguesia/Brasilândia; Guaianases; Ipiranga; Itaim Paulista; Itaquera; Jabaquara; Jaçanã / Tremembé; Lapa; M'Boi Mirim; Moóca; Parelheiros; Penha; Perus; Pinheiros; Pirituba; Santana/Tucuruvi; Santo Amaro; São Mateus; São Miguel; Sé; Vila Maria/Vila Guilherme; Vila Mariana; Vila Prudente/Sapopemba.

1.2 A Classificação dos Indicadores

A Lei N 14.173, publicada em 26 de junho de 2006, compromete a Prefeitura de São Paulo com o estabelecimento de indicadores de desempenho relativos à qualidade dos serviços públicos no Município de São Paulo. A classificação é a seguinte:

Saúde pública

- Nível de exames preventivos de saúde (adulto e infantil);
- Tempo médio de atendimento para consultas (adulto e infantil);
- Tempo médio de atendimento para análises clínicas (adulto e infantil);
- Tempo médio de atendimento para outros procedimentos (adulto e infantil);
- Tempo médio para a realização de procedimentos de alta complexidade;
- Número de crianças vacinadas.

Educação básica

- Nível de universalização da educação infantil;
- Nível de universalização do ensino fundamental;
- Nível de universalização do ensino médio;
- Nível de evasão escolar;
- Nível de alfabetização na faixa etária;
- Nível de repetência dos alunos;
- Nível de formação/graduação dos professores;
- Nível de adequação série/idade;
- Nível de compatibilidade bairro/escola;
- Desempenho apurado no Sistema de Avaliação de Aproveitamento Escolar dos Alunos da Rede Municipal de Ensino de São Paulo, estabelecido pela Lei n 14.063, de 14 de outubro de 2005.

Segurança no trânsito

- Número proporcional de acidentes fatais ocorridos no trânsito em cem mil habitantes, no conjunto e no período considerado;
- Número proporcional de acidentes no trânsito com lesões em cem mil habitantes, ocorridos no conjunto e no período considerados;
- Média aritmética mensal dos congestionamentos;
- Área verde por habitante por metro quadrado;
- Área de lazer por habitante por metro quadrado;
- Qualidade dos índices de qualidade do ar;
- Qualidade da água do sistema fluvial.

Limpeza pública

- População atendida por coleta de lixo;
- População atendida por coleta de lixo seletiva;
- Proporção de lixo seletivo coletado;
- Destinação final do lixo;
- Varrição de logradouros públicos.

Transportes públicos

- Tempo médio de espera nos terminais de transferência utilizados para o embarque de passageiros para o transporte urbano;
- Tempo médio de espera nas paradas intermediárias entre o terminal de transferência de saída e o de chegada;
- Tempo médio para o deslocamento dos trabalhadores de seus domicílios aos locais de trabalho;
- Velocidade média do deslocamento do ônibus em horário normal e em horário de pico;
- Nível médio de pontualidade por empresa;
- Nível de limpeza da área de circulação dos terminais de transferência;
- Nível de limpeza dos banheiros públicos dos terminais de transferência;

- Nível de limpeza, conservação e manutenção da frota.

1.3 Os Indivíduos

Os indivíduos desta pesquisa são as 31 subprefeituras de São Paulo: Aricanduva; Butantã; Campo Limpo; Capela do Socorro; Casa Verde/Cachoeirinha; Cidade Ademar; Cidade Tiradentes; Ermelino Matarazzo; Freguesia/Brasilândia; Guaianases; Ipiranga; Itaim Paulista; Itaquera; Jabaquara; Jaçanã / Tremembé; Lapa; M'Boi Mirim; Moóca; Parelheiros; Penha; Perus; Pinheiros; Pirituba; Santana/Tucuruvi; Santo Amaro; São Mateus; São Miguel; Sé; Vila Maria/Vila Guilherme; Vila Mariana; Vila Prudente/Sapopemba.

1.4 As Variáveis

São 13 as variáveis desta pesquisa, incluindo o nome das subprefeituras. As mesmas são mais bem explicadas na Tabela 1. Já a Tabela 2 traz a amostra total com o detalhamento da população e os indicadores utilizados.

Tabela 1. As Variáveis.

Variável	Significado	Tipo	Unidade de Medida
Subprefeitura	São os nomes das subprefeituras	Variável Categórica	N/A
Cinema	Porcentagem das salas de cinema em cada subprefeitura sobre o total de salas da cidade.	Variável Quantitativa	Percentual
Abandono no ensino médio total	Porcentagem de alunos que abandonaram o ensino médio nas redes pública e privada.	Variável Quantitativa	Percentual
Abandono no ensino fundamental total	Porcentagem de alunos que abandonaram o ensino fundamental nas redes pública e privada.	Variável Quantitativa	Percentual
Demanda de creches	Porcentagem de matrículas efetuadas sobre o total de procura por vagas.	Variável Quantitativa	Percentual
Unidades Esportivas	Porcentagem das unidade públicas de esporte em cada subprefeitura, sobre o total do município.	Variável Quantitativa	Percentual
Favelas	Porcentagem de domicílios em favelas sobre o total de domicílios da subprefeitura..	Variável Quantitativa	Percentual
Consumo de água per capita	Consumo de água (Residencial, Comercial, Público, Industrial e Misto) estimado, em metros cúbicos, por habitante.	Variável Quantitativa	Metros cúbicos
Orçamento por	Orçamento destinado a cada subprefeitura dividido	Variável	N/A

subprefeitura	peelo número de seus habitantes.	Quantitativa	
Leitos Hospitalares	Número de leitos hospitalares públicos e privados disponíveis por mil habitantes.	Variável Quantitativa	N/A
Desemprego de jovens de 16 a 29 anos	Taxa média de desemprego de jovens de 16 a 29 anos.	Variável Quantitativa	N/A
Mortes no trânsito	Número de mortes em acidentes de trânsito por cem mil habitantes, por local de moradia da vítima.	Variável Quantitativa	N/A
Acessibilidade	Percentual da frota de ônibus com acessibilidade para pessoas com deficiência.	Variável Quantitativa	Percentual
Agressão a mulheres	Número de internações de mulheres de 20 a 59 anos por causas relacionadas a possíveis agressões, por cem mil mulheres nessa faixa etária.	Variável Quantitativa	N/A

Tabela 2. Amostra Total

Subprefeitura	Cinemas	Abandono no ensino médio	Abandono no ensino fundamental	Demanda de creche	Unidades Esportivas	Favelas	Consumo de água
Aricanduva	8,78	3,94	0,86	53,18	4,53	2,07	6,58
Butantã	4,58	4,99	0,99	55,19	4,23	20,82	7,54
Campo Limpo	4,20	7,79	1,27	34,97	0,6	40,41	4,46
Capela do Socorro	1,53	6,07	1,34	41,96	0	14,59	5,30
Casa Verde/Cachoeirinha	0,00	7,04	1,03	40,51	1,51	10,75	4,71
Cidade Ademar	0,00	5,37	1,35	34,03	0	19,82	4,39
Cidade Tiradentes	0,00	11,79	1,55	49,48	2,42	6,64	3,00
Ermelino Matarazzo	0,00	6,87	1,27	46,48	0	7,27	5,55
Freguesia/Brasilândia	0,00	6,64	1,31	48,08	5,74	20,27	5,06
Guaianases	0,00	9,45	2,17	52,35	0	9,10	4,53
Ipiranga	0,38	5,20	1,05	45,42	4,53	21,27	5,23
Itaim Paulista	0,76	6,97	1,49	44,49	2,42	10,44	4,87
Itaquera	1,15	7,62	1,18	49,85	6,34	8,55	4,96
Jabaquara	0,00	7,93	2,30	46,39	3,93	26,84	6,08
Jaçanã / Tremembé	0,00	6,87	1,27	40,17	0,91	10,10	5,14
Lapa	4,20	3,58	1,00	66,85	7,85	8,16	6,16
M'Boi Mirim	0,00	9,29	1,88	37,13	0	27,89	4,23
Moóca	8,02	2,95	0,85	67,88	10,57	1,07	6,34
Parelheiros	0,00	7,36	1,47	36,05	0	14,11	3,83
Penha	3,05	5,45	1,08	61,07	3,02	6,60	5,39
Perus	0,00	9,37	0,85	41,02	0	35,67	5,36
Pinheiros	13,36	3,23	0,24	55,08	0	0,91	8,39
Pirituba	0,00	5,67	0,55	56,67	6,65	13,66	4,86
Santana/Tucuruvi	0,38	2,62	0,48	51,76	5,44	1,28	5,34
Santo Amaro	13,74	2,74	0,64	54,71	5,74	4,68	7,81
São Mateus	0,00	6,51	1,31	52,47	1,21	14,15	5,57
São Miguel	0,00	8,15	1,50	52,56	0,91	12,48	5,37
Sé	20,61	4,82	1,99	48,86	5,14	0,31	4,48
Vila Maria/Vila Guilherme	4,20	5,55	1,46	54,72	3,63	6,03	5,76
Vila Mariana	7,25	2,53	0,46	56,89	6,65	0,73	5,55
Vila Prudente/Sapopemba	3,82	7,82	1,20	59,46	6,04	15,29	5,19

Continuação Tabela 2

Subprefeitura	Orçamento por subprefeitura per capita	Leitos Hospitalares	Desemprego de jovens de 16 a 29 anos	Mortes no trânsito	Acessibilidade	Agressão a mulheres
Aricanduva	118,97	0,50	16,05	9,27	0,00	102,26
Butantã	113,18	3,45	13,86	11,15	0,00	144,34
Campo Limpo	70,89	0,52	20,62	10,55	0,00	93,48
Capela do Socorro	60,06	0,65	20,62	13,55	0,00	95,91
Casa Verde/Cachoeirinha	81,67	1,36	18,19	15,80	0,00	164,76
Cidade Ademar	83,81	0,61	20,62	11,39	0,00	126,71
Cidade Tiradentes	82,77	0,86	21,48	7,39	0,00	144,32
Ermelino Matarazzo	128,35	4,29	21,48	16,47	0,00	138,69
Freguesia/Brasilândia	81,75	0,48	18,19	12,98	0,00	197,97
Guaianases	107,43	1,21	21,48	9,42	0,00	234,48
Ipiranga	84,18	2,95	14,79	13,12	0,00	102,53
Itaim Paulista	81,07	0,85	21,48	9,72	0,00	136,72
Itaquera	82,46	1,96	21,48	10,19	0,00	145,00
Jabaquara	129,42	3,59	14,79	16,83	0,00	75,78
Jaçanã / Tremembé	105,10	3,57	18,35	17,48	0,00	147,06
Lapa	132,77	3,37	13,86	11,23	0,00	106,86
M'Boi Mirim	73,88	0,71	20,62	15,78	0,00	104,97
Moóca	122,88	9,86	16,05	12,22	0,00	106,28
Parelheiros	136,81	0,00	20,62	9,03	0,00	42,99
Penha	93,52	1,04	16,05	11,58	0,00	118,60
Perus	142,44	0,00	18,19	9,43	0,00	122,36
Pinheiros	170,80	14,73	13,86	11,55	0,00	60,64
Pirituba	78,81	1,85	18,19	14,04	0,00	121,55
Santana/Tucuruvi	103,81	4,05	18,35	7,36	0,00	86,35
Santo Amaro	170,35	5,59	14,79	18,66	0,00	62,97
São Mateus	97,39	0,68	21,48	11,69	0,00	150,18
São Miguel	88,43	0,86	21,48	14,82	0,00	147,85
Sé	238,96	24,29	12,78	10,45	0,00	105,19
Vila Maria/Vila Guilherme	120,81	1,79	18,35	13,72	0,00	145,55
Vila Mariana	112,89	17,19	14,79	12,03	0,00	61,26
Vila Prudente/Sapopemba	81,10	0,87	16,05	11,24	0,00	159,16

3. ANÁLISE INDIVIDUAL DAS VARIÁVEIS

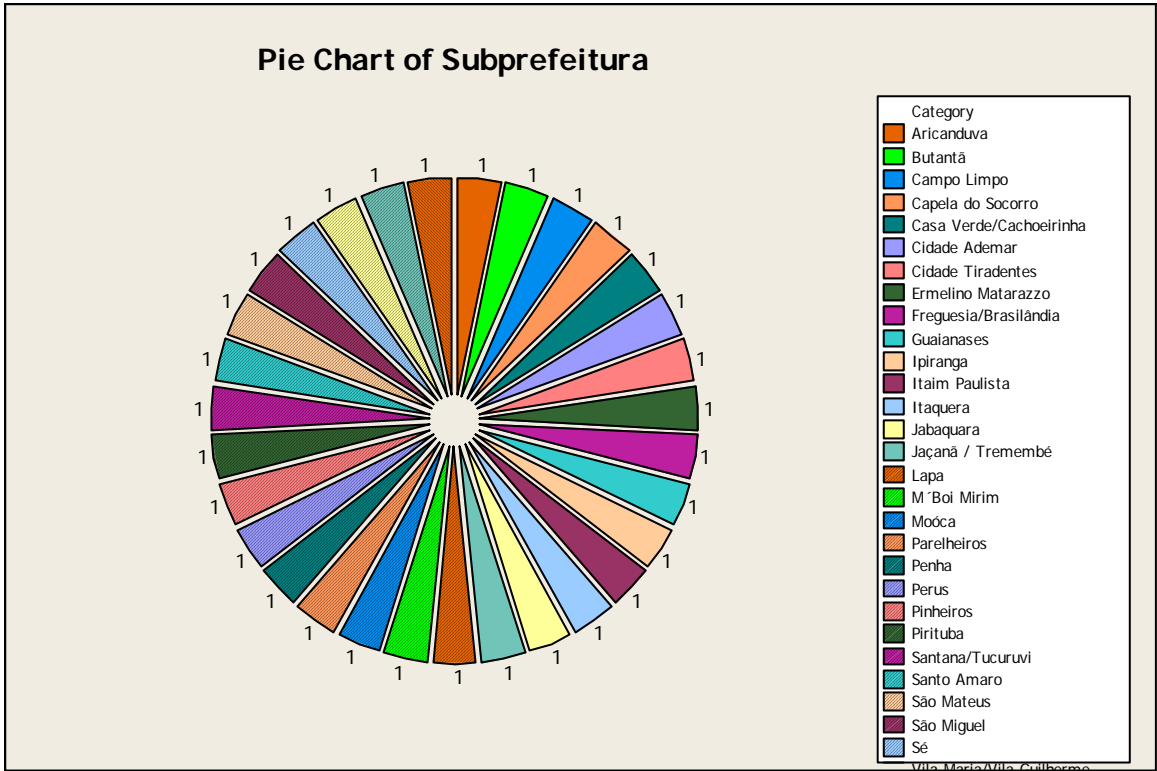
Para que se possa iniciar a análise crítica sobre a base de dados apresentada no tópico anterior, iniciam-se as análises individuais das variáveis.

3.1 Variáveis categóricas

Este tipo de variável indica que o foco de concentração deve ser a análise de gráficos do tipo pie chart e barras.

2.1.1. Variável: “Subprefeitura”

Nossa amostra totaliza 31 subprefeituras. Por se tratar do nome da subprefeitura não há como realizar uma análise aprofundada da variável, desse modo, apenas faremos o pie-chart.



3.2 Variáveis Quantitativas

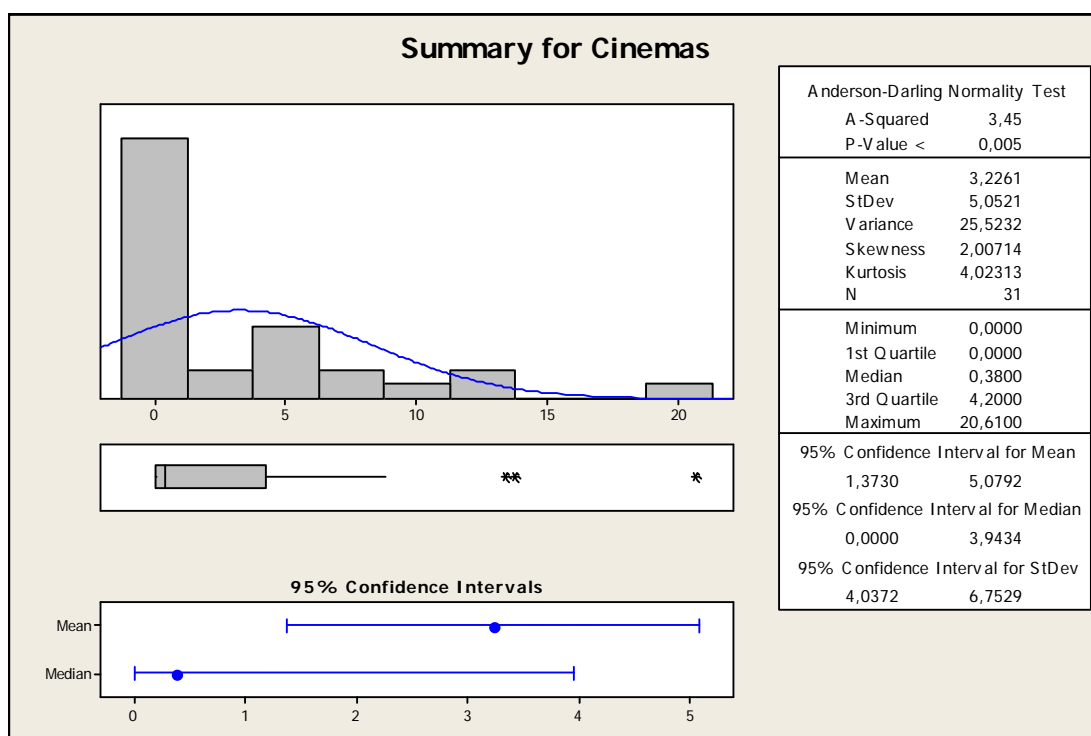
Na análise das variáveis quantitativas podemos utilizar um grande número de ferramentas de análise tais como histogramas, curvas de densidade, gráfico de ramos, bolhas, boxplot, dotplot. Nesta parte, utilizaremos histogramas e gráfico ramo de folhas para entender melhor o comportamento de cada variável.

3.2.1 Variável: “Cinema”

Os dados são de 2006 e representam a porcentagem das salas sobre o total de salas na cidade, que no ano era de 262 salas.

```
Stem-and-leaf of Cinemas  N = 31
Leaf Unit = 1,0

(19)  0  00000000000000000011
12    0  33
10    0  4444
6     0  7
5     0  88
3     1
3     1  33
2     1  1
1     1
1     2  0
```

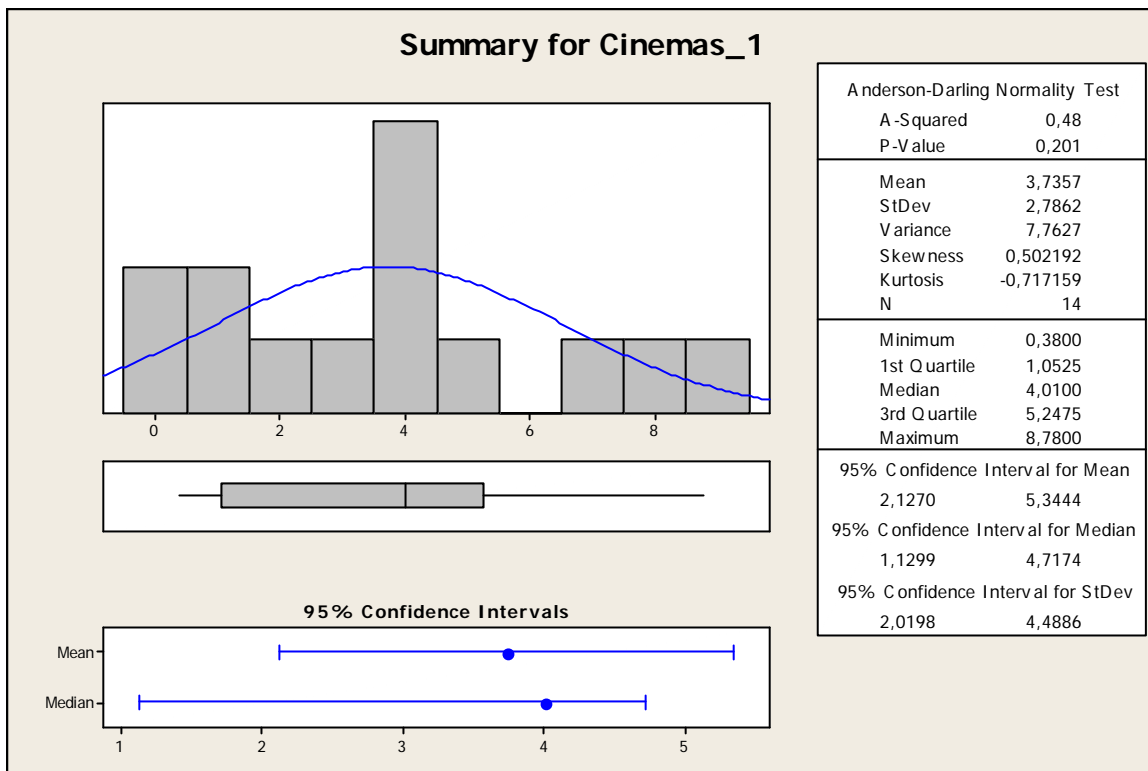


Observa-se que este indicador apresenta uma distribuição assimétrica à esquerda, mas tal assimetria é devida à existência de um grande número de subprefeituras não possuem nenhum cinema - observa-se que 14 subprefeituras, o que equivale há 45,2% do total, que não possuem nenhuma sala em sua região. A curva de densidade indica uma distribuição relativamente anormal, o que é corroborado pelo p-value de $< 0,05\%$ no teste de normalidade de Anderson-Darling. O boxplot nos permite confirmar a afirmação precedente, uma vez que a média se localiza mais à direita da linha de distribuição.

A média dos valores é 0,38, variância 25,5232, desvio padrão 5,05 e o fator de desigualdade das subprefeituras, valor representado por quantas vezes a subprefeitura com classificação mais baixa está pior daquela em classificação mais alta, é de 54,2 ($20,61 \text{ maximum} / 0,38 \text{ 1r quartile}$);

No gráfico podemos observar que a maioria das subprefeituras não possuem nenhum ou possuem muito poucas salas de cinema, pelo gráfico de ramo de folhas há 17 subprefeituras que estão abaixo de 1%, e apenas 2 entre 1 a 2%. Há também alguns outliers: Pinheiros, Sé e Santo Amaro.

Por possuir muitos dados zerados e alguns outliers se faz necessária a transformação logarítmica:



O novo gráfico apresenta uma assimetria a esquerda, observamos que no primeiro quartil ainda predomina subprefeituras com poucos cinemas. O boxplot permite corroborar a afirmação precedente, uma vez que a média se localiza mais à esquerda da linha de distribuição.

Com 95% de confiança, pode-se afirmar que a média para salas de cinema se localiza entre 2,12 e 5,344.

3.2.2 Variável: “Abandono no ensino médio total”

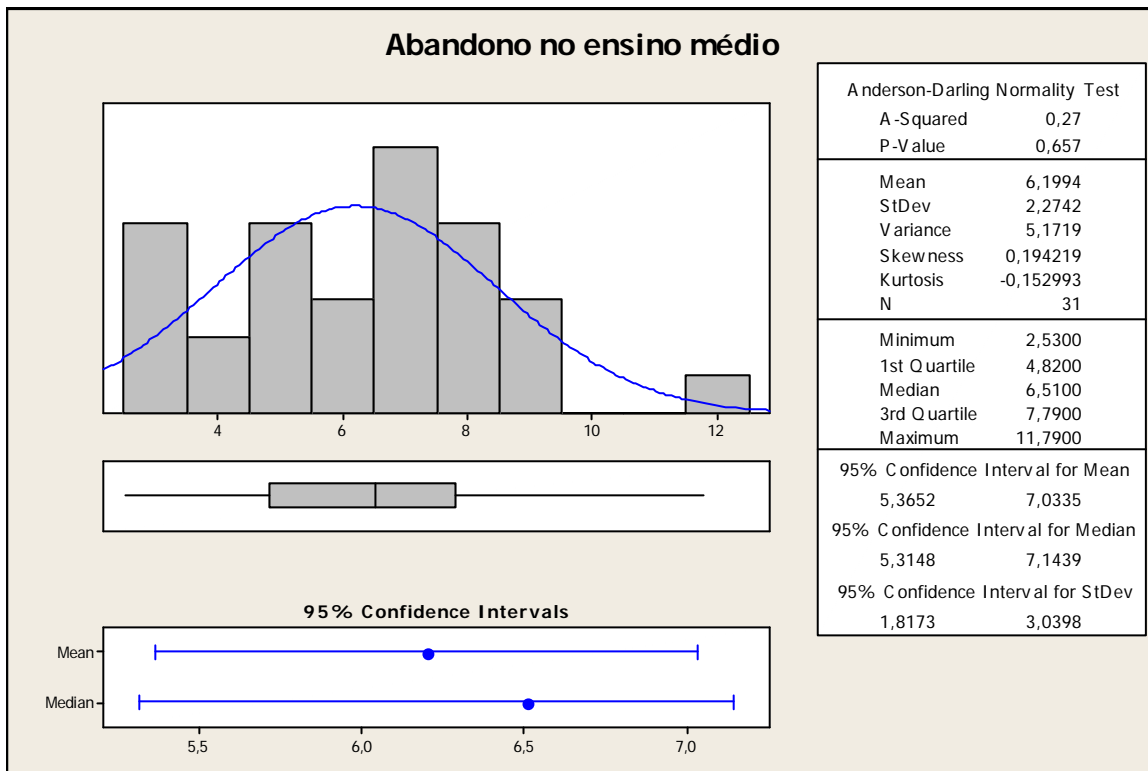
Os dados são de 2005 e representam porcentagem de alunos que abandonaram o ensino médio nas redes pública e privada, em seu valor absoluto representavam 31.769 abandonos em toda São Paulo.

```

Stem-and-leaf of Abandono no ensino médio  N  = 31
Leaf Unit = 0,10

 4  2  5679
 7  3  259
 9  4  89
14  5  23456
(6) 6  056889
11  7  036789
 5  8  1
 4  9  234
 1 10
 1 11 7

```



Com base nos gráficos e números da tabela acima, pode-se notar que se trata de uma distribuição ligeiramente assimétrica para a esquerda. Existe um pico à esquerda do histograma, que representam subprefeituras com baixíssimo abandono no ensino médio. De fato, a média apresentada é de 6,51, variância 5,1719, desvio padrão 2,2742 e o fator de desigualdade das subprefeituras é de 4,7 (11,79 maximum/2,53 minimum). Não encontrados valores atípicos nesta amostra, dado este reforçado pelo box-plot apresentado.

Observa-se que o valor maximum é o de 11,79, representado, de acordo com o gráfico de ramos e folhas, por apenas 1 (uma) subprefeitura sendo ela a de Cidade Tiradentes e o valor minimum é de 2,53 representado também por 1 subprefeitura – Vila Mariana.

3.2.3 Variável: “Abandono no ensino fundamental”

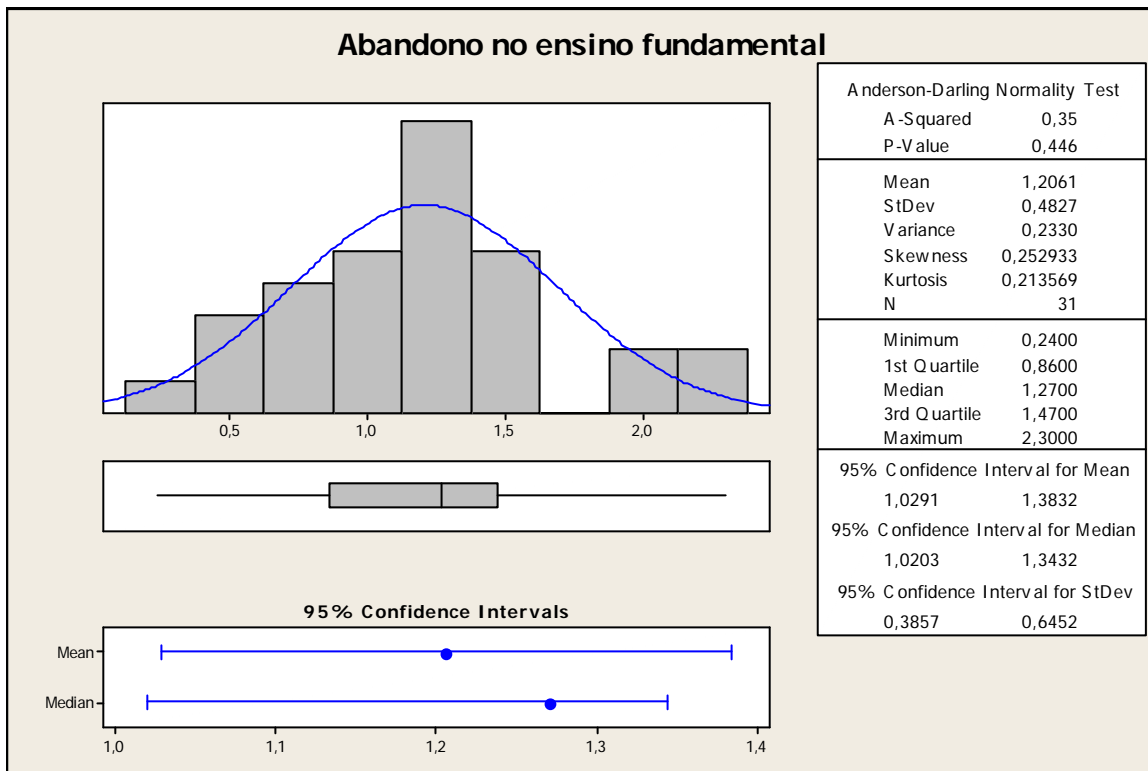
Os dados são de 2005 e representam porcentagem de alunos que abandonaram o ensino fundamental nas redes pública e privada, em seu valor absoluto representavam 19.853 abandonos em toda São Paulo.

```

Stem-and-leaf of Abandono no ensino fundamental  N  = 31
Leaf Unit = 0,10

 1  0  2
 4  0  445
 5  0  6
 9  0  8889
14  1  00001
(8) 1  22223333
 9  1  44455
 4  1
 4  1  89
 2  2  1
 1  2  3

```



Trata-se de uma distribuição simétrica, o que é evidenciado tanto pela curva de densidade, como pelo boxplot e pelo p-value. Não há valores atípicos, sendo o valor mínimo apresentado é da subprefeitura de Pinheiros (0,24) e o valor máximo apresentado se refere a Jabaquara (2,3).

A mediana indica que aproximadamente metade dos países tem o índice de 0,60 e metade está classificada em valores superiores àquele. O valor médio encontrado é de 1,27, com desvio-padrão de 0,48. O fator de desigualdade das subprefeituras é de 9,6 (2,3 maximum/0,24 minimum).

3.2.4 Variável: “Demanda de creches”

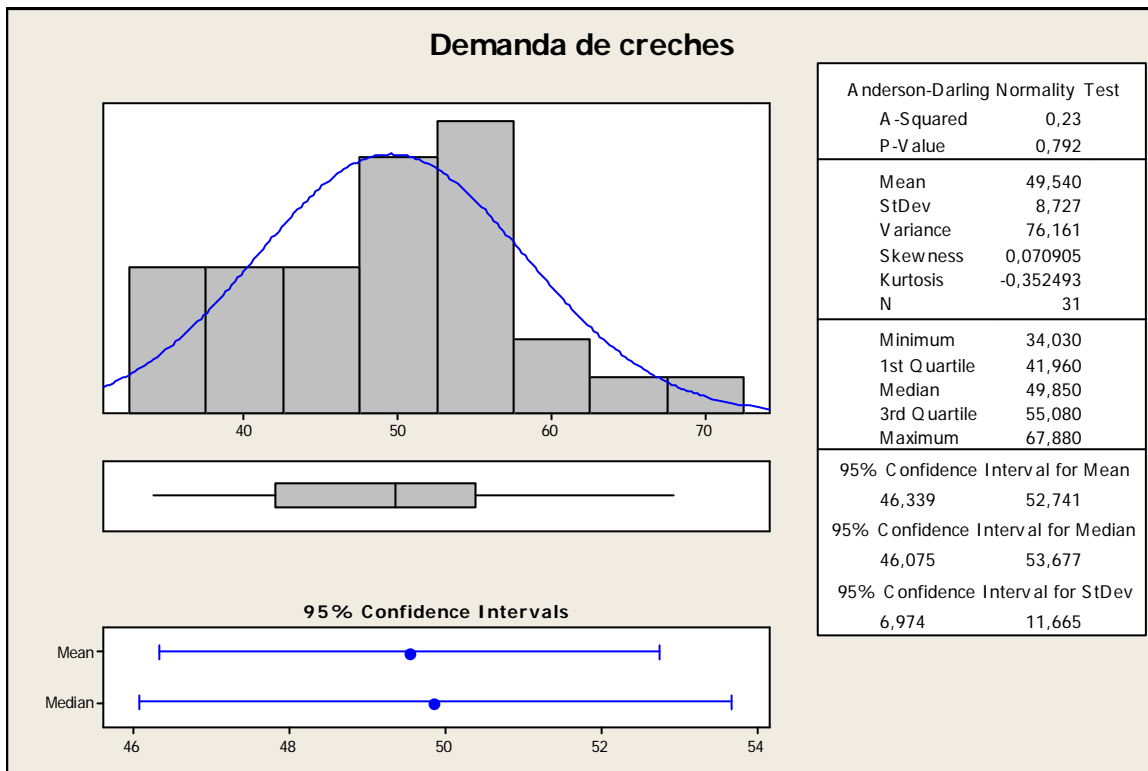
Os dados são de 2008 e representam porcentagem de matrículas efetuadas sobre o total de procura por vagas, o valor absoluto representa o número de matrículas e é de 101.966.

```

Stem-and-leaf of Demanda de creches  N  = 31
Leaf Unit = 1,0

 2  3  44
 4  3  67
 9  4  00114
(7) 4  5668899
15  5  1222344
 8  5  55669
 3  6  1
 2  6  67

```



Trata-se também de uma distribuição simétrica. Não há valores atípicos, sendo o valor mínimo apresentado é da subprefeitura de Pinheiros (0,24) e o valor máximo apresentado se refere a Jabaquara (2,3).

A mediana indica que aproximadamente metade dos países tem um escore menor do que 49,85 e metade desses, um índice maior. O índice médio apurado é de 49,85, mas o desvio-padrão é de 8,7, demonstrando a dispersão nesta amostra.

3.2.5 Variável: “Unidades Esportivas”

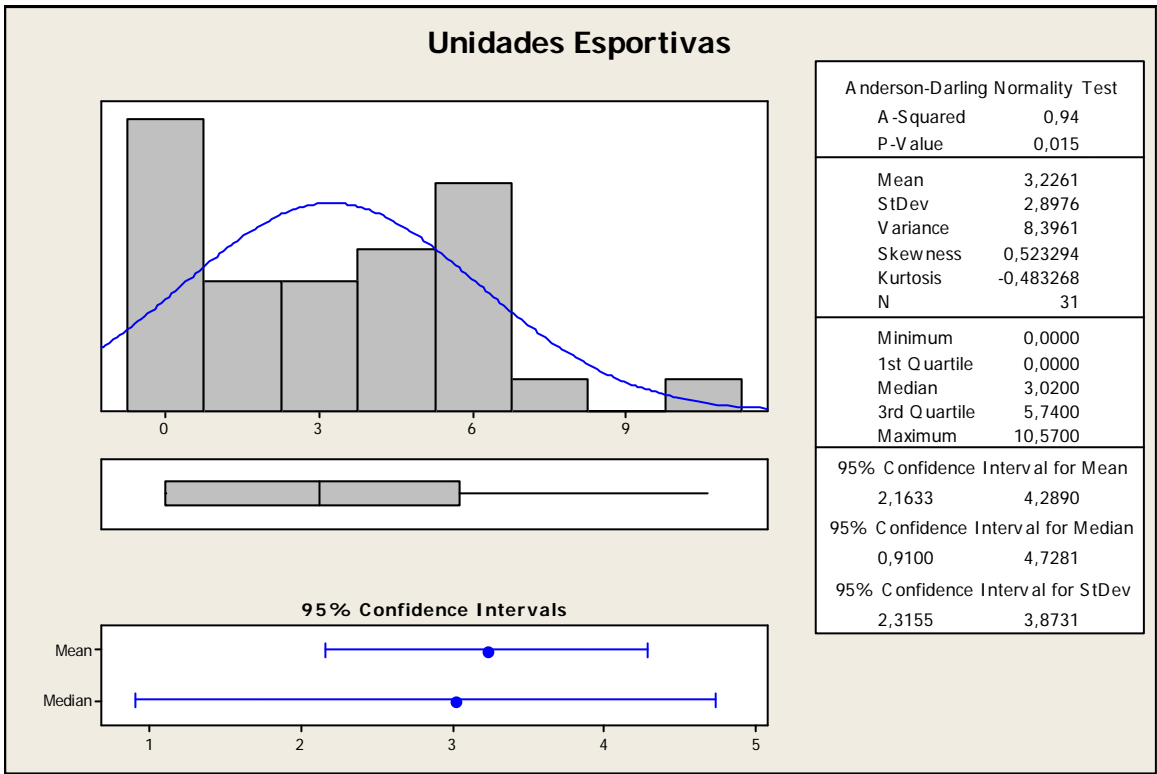
Os dados são de 2006 e porcentagem das unidade públicas de esporte em cada subprefeitura, sobre o total do município. O valor absoluto das unidades esportistas existentes em São Paulo é de 331.

Stem-and-leaf of Unidades Esportivas N = 31
Leaf Unit = 0,10

```

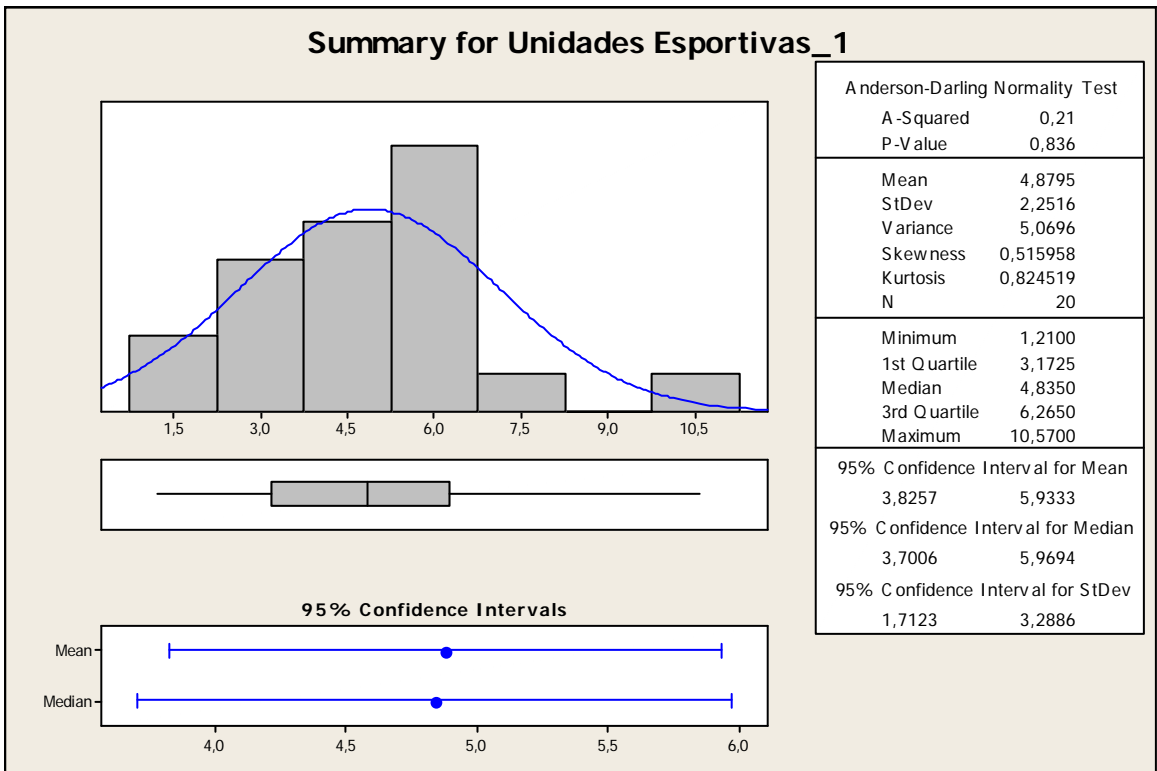
11 0 0000000699
13 1 25
15 2 44
(3) 3 069
13 4 255
10 5 1477
6 6 0366
2 7 8
1 8
1 9
1 10 5

```



Observa-se que este indicador apresenta uma distribuição assimétrica à esquerda, devido ao fato de que há oito subprefeituras com 0 unidades e três abaixo de 1%. Assim, a média apresentada é de 3,02, variância 8,3961 e desvio padrão 2,89 e o fator de desigualdade das subprefeituras é de 17,6 (10,57 maximum/3,02 median)

Por possuir muitos dados zerados e alguns abaixo de 1% se faz necessária a transformação logarítmica:



Com o gráfico simétrico podemos observar que a mediana está próxima da média, demonstrando certa simetria da distribuição. Com 95% de confiança, podemos afirmar que a média para este indicador se encontra em uma faixa de 3,82 e 5,933.

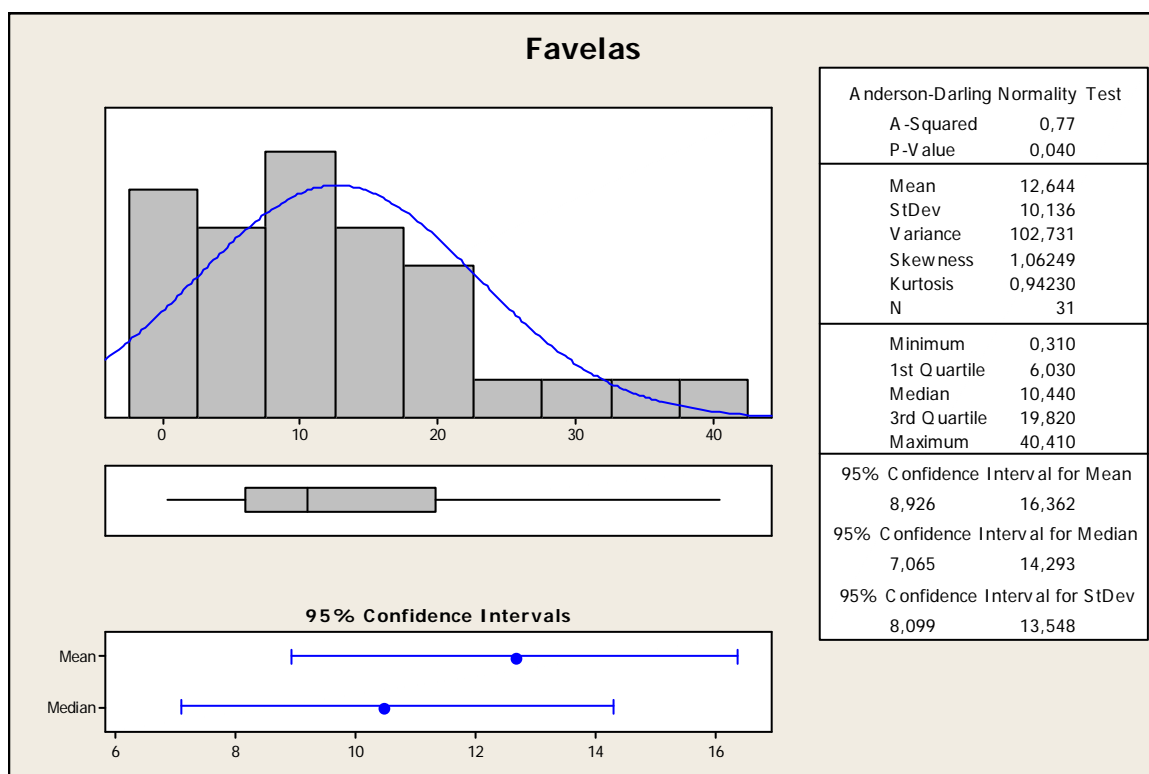
3.2.6 Variável: “Favelas”

Os dados são de 2008 e representam a porcentagem de domicílios em favelas sobre o total de domicílios da subprefeitura, o total de domicílios em favelas no município é de 385.081.

```

Stem-and-leaf of Favelas  N = 31
Leaf Unit = 1,0

 7  0  0001124
 14 0  6667889
(8) 1  00023444
 9  1  59
 7  2  001
 4  2  67
 2  3
 2  3  5
 1  4  0
    
```



Pode-se observar que se trata de uma distribuição assimétrica à esquerda. Com base na análise dos gráficos acima podemos tecer alguns comentários em relação ao comportamento desta variável:

A média apresentada é de 10,44, variância 102,731 e desvio padrão 10,136 e o fator de desigualdade das subprefeituras é de 130,4 (40,41 maximum/0,31 minimum), a mais alta de todos os indicadores.

Há três subprefeituras com porcentagem menor que 1% e duas abaixo de 2%, sendo elas, respectivamente, Pinheiros, Sé, Vila Mariana, Moóca e Santana/Tucuruvi.

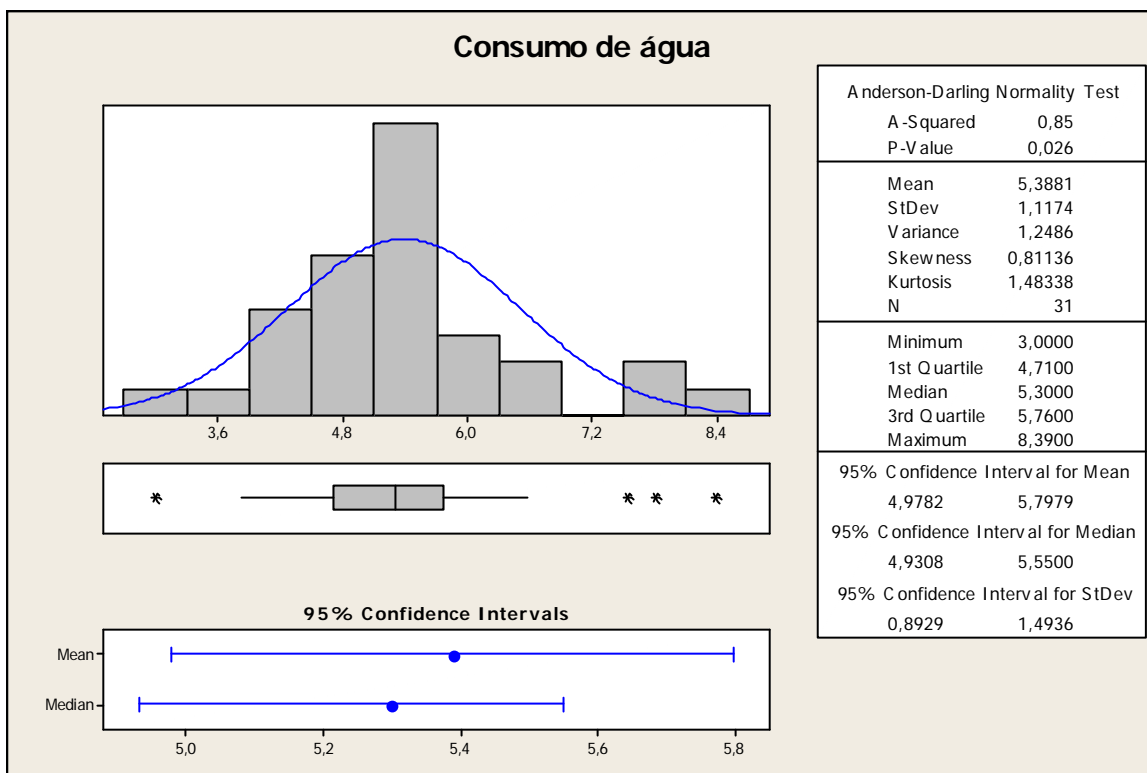
3.2.7 Variável: “Consumo de água”

Os dados são de 2007 e representa o consumo de água (Residencial, Comercial, Público, Industrial e Misto) estimado, em metros cúbicos, por habitante, o total absoluto em metros cúbicos da cidade é de 57.199.

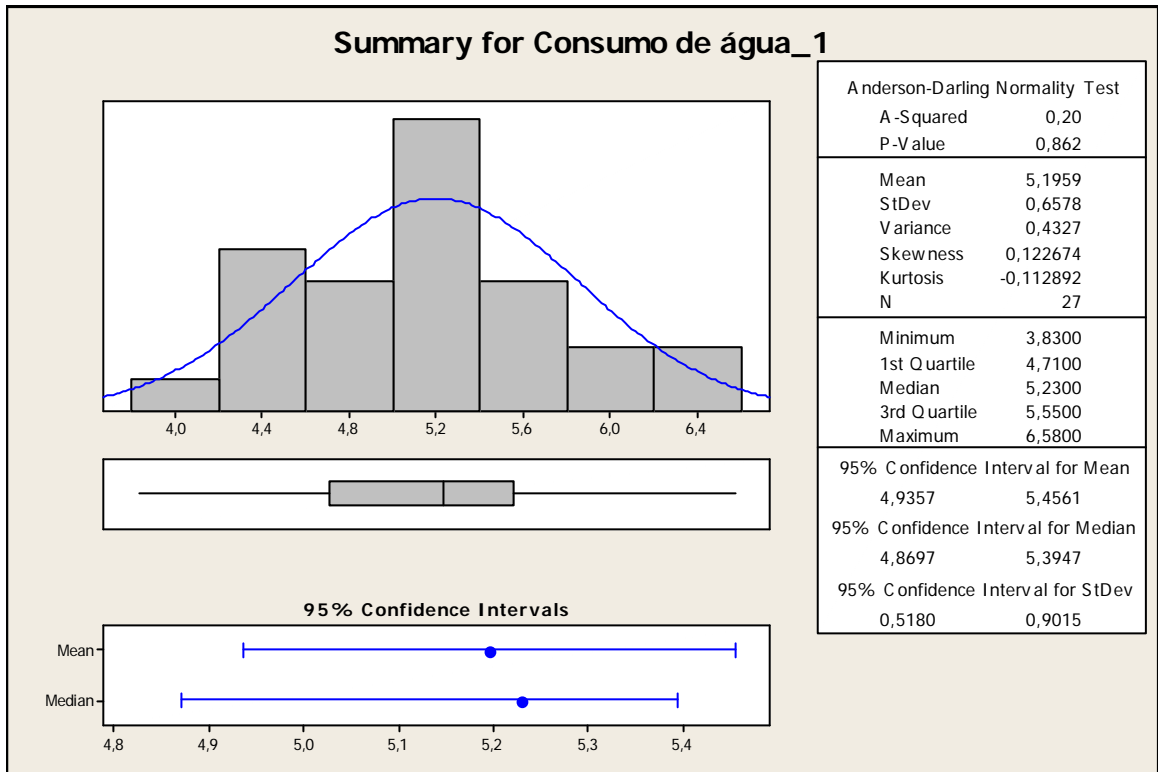
```

Stem-and-leaf of Consumo de água  N  = 31
Leaf Unit = 0,10

 1  3  0
 2  3  8
 6  4  2344
11  4  57889
(9) 5  011233333
11  5  5557
 7  6  013
 4  6  5
 3  7
 3  7  58
 1  8  3
    
```



Observa-se uma distribuição simétrica, porém constata-se a presença de valores atípicos que representam a mais baixa subprefeitura em relação ao consumo é o da Cidade Tiradentes com o consumo de apenas 3% ou 747 metros cúbicos por habitante, os outros pontos destacados são das subprefeituras Butanta (7,54%), Santo Amaro (7,81%) e Pinheiros (8,39%). Assim, resolvemos refazer o gráfico eliminando os outliers:



Trata-se de uma distribuição simétrica, o que é evidenciado tanto pela curva de densidade, como pelo boxplot e pelo p-value do teste de normalidade de Anderson-Darling.

A mediana está razoavelmente próxima da média, demonstrando certa simetria da distribuição. Com 95% de confiança, podemos afirmar que a média populacional para este indicador se encontra em uma faixa de 4,94 e 5,46.

3.2.8 Variável: “Orçamento por subprefeitura”

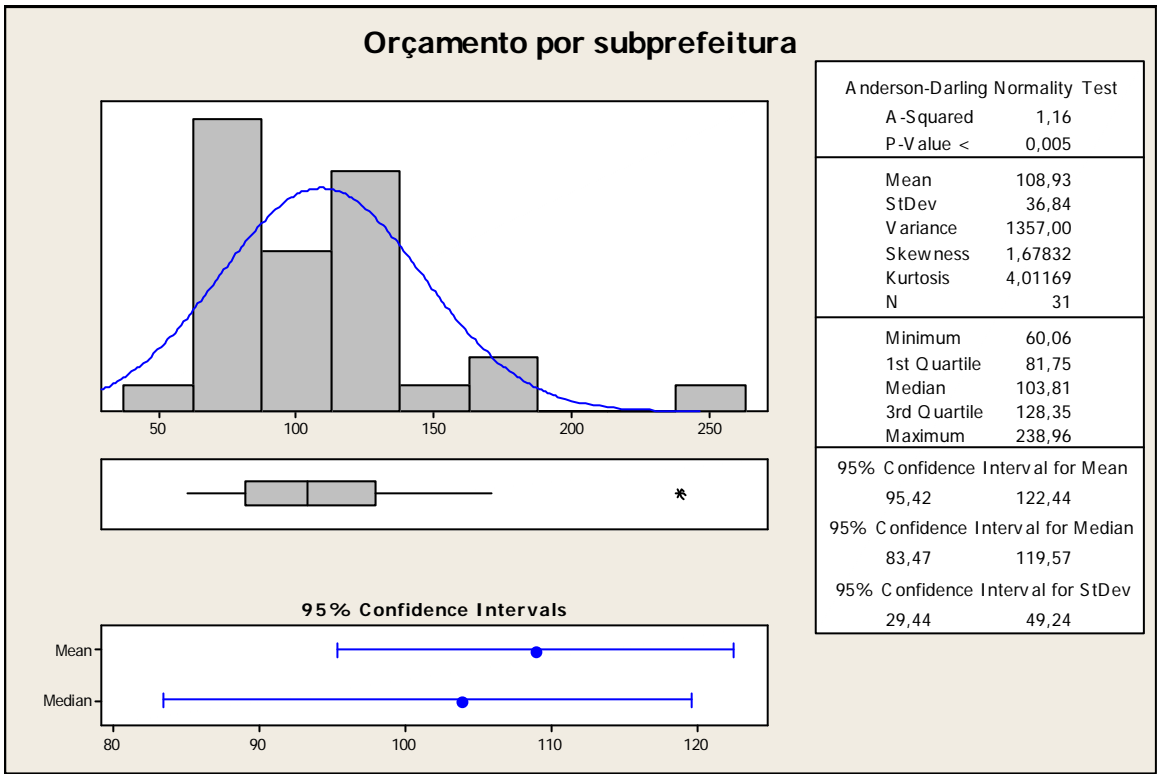
Os dados são de 2008 e representam o orçamento destinado a cada subprefeitura dividido pelo número de seus habitantes.

```

Stem-and-leaf of Orçamento por subprefeitura  N = 31
Leaf Unit = 10

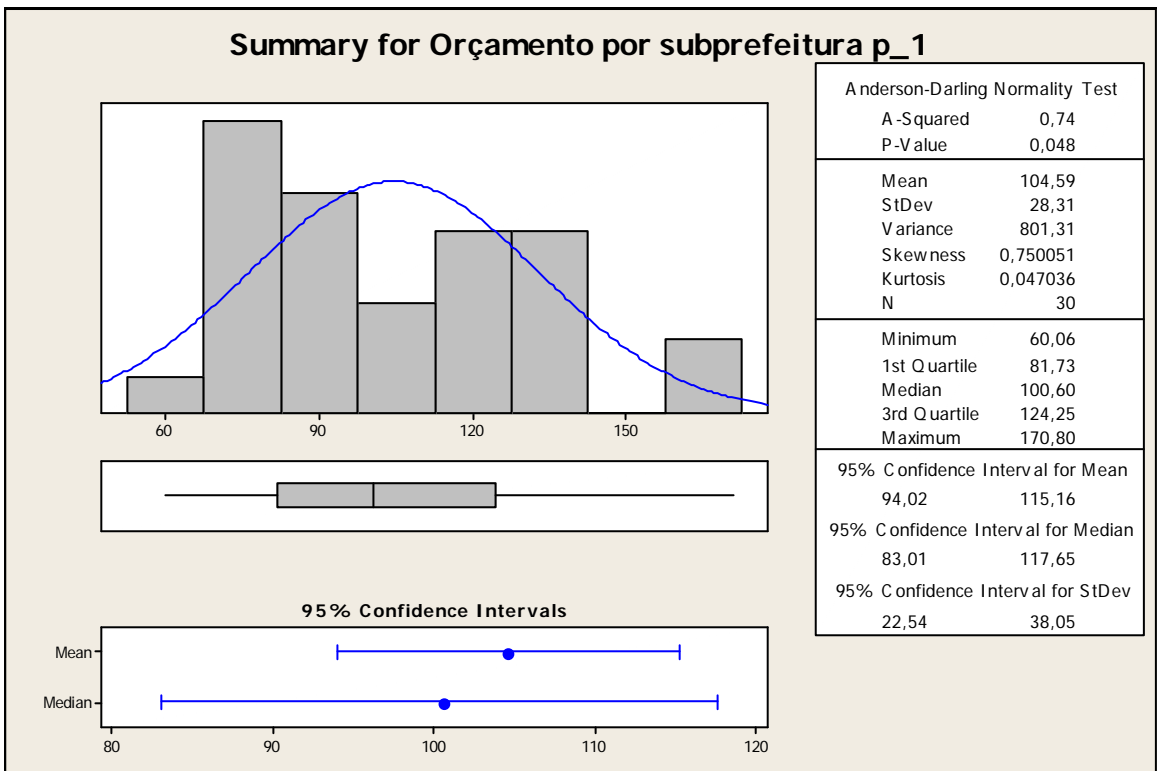
 4  0  6777
15  0  88888888899
(6) 1  000111
10  1  222233
 4  1  4
 3  1  77
 1  1
 1  2
 1  2  3

```



Este histograma apresenta uma distribuição assimétrica à direita. A curva de densidade indica uma distribuição anormal, o que é corroborado pelo p-value de $< 0,005$ no teste de normalidade de Anderson-Darling. O boxplot permite corroborar a afirmação precedente, uma vez que a média se localiza mais à direita da linha de distribuição. A média apresentada é de 103,81, variância 1357,00 e desvio padrão 36,84 e o fator de desigualdade das subprefeituras é de 4,0 (238,96 maximum/60,06 minimum);

Há apenas um outlier que representa a subprefeitura da Sé, por ser um número muito desproporcional aos outros, 238,96%, retiramos o valor e refizemos o histograma:



O gráfico agora apresenta simetria, observa-se que o A-squared está 0,74. Também a mediana está relativamente próxima da média, o que indica uma relativa simetria na distribuição.

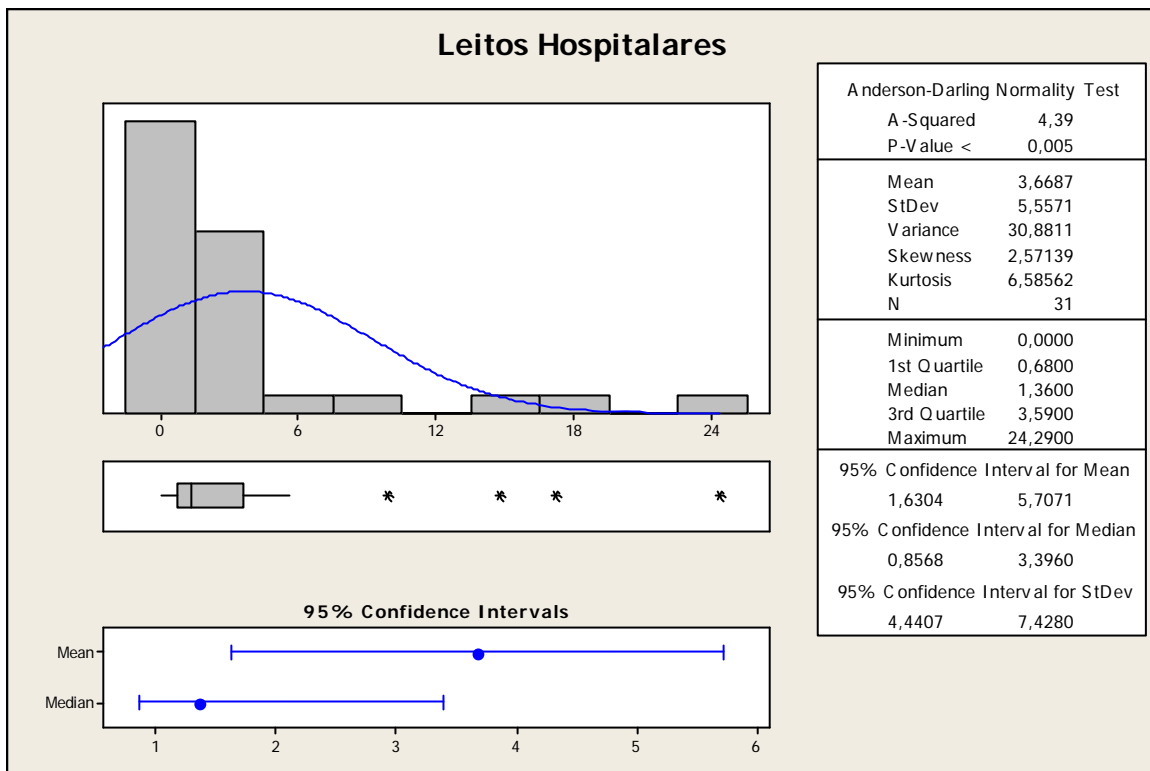
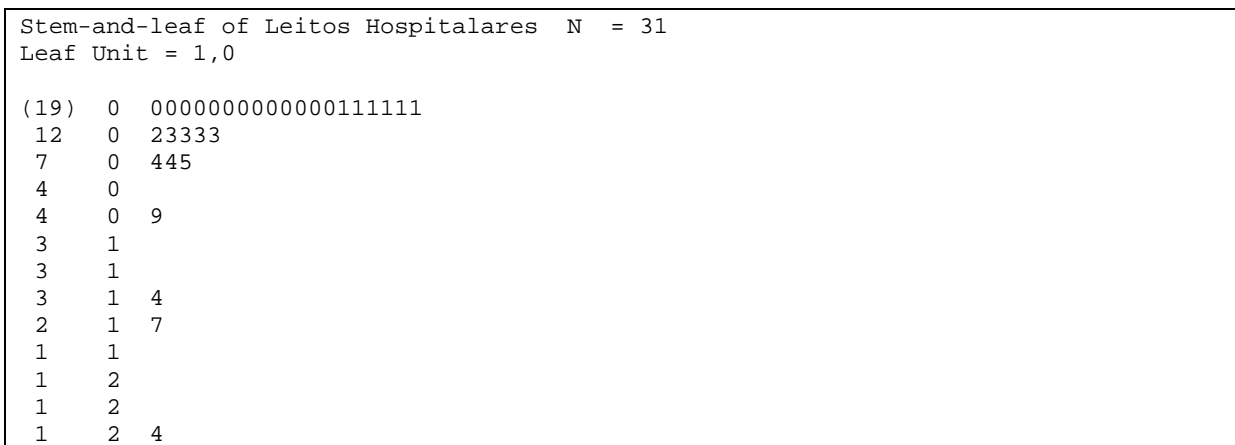
A mediana nos indica que aproximadamente metade das subprefeituras tem um escore menor do que 100 e metade desses, um índice maior. O índice médio apurado é de 104, mas o desvio-padrão (medida de dispersão) é de 28,31 - o índice mínimo é de 60,06, e o máximo de 170,80, o que corrobora a grande dispersão dos dados.

O fator de desigualdade das subprefeituras, valor representado por quantas vezes a subprefeitura com classificação mais baixa está pior daquela em classificação mais alta, é de 1,7 (21,48 maximum/12,78 minimum);

Com 95% de confiança, podemos afirmar que a média encontra-se entre os valores de 94 e 115,16.

3.2.9 Variável: “Leitos Hospitalares”

Os dados são de 2008 e representam o número de leitos hospitalares públicos e privados disponíveis por mil habitantes. Em São Paulo há 34.404 leitos hospitalares.



O histograma apresentado mostra uma distribuição assimétrica para a esquerda, o que pode ser comprovado pelo teste de normalidade de Anderson-Darling que apresenta p-value < 0,005. Há duas subprefeituras sem nenhum leito hospitalar, onze abaixo de 1%.

A média apresentada é de 1,36, variância 30,88 e desvio padrão 5,55 e o fator de desigualdade das subprefeituras é de 50,6 (24,29 maximum/0,68 1r quartil). Há outliers representados pelas subprefeituras da Móoca (9,86%), Pinheiros (14,73%), Vila Mariana (17,19%) e Sé (24,29%).

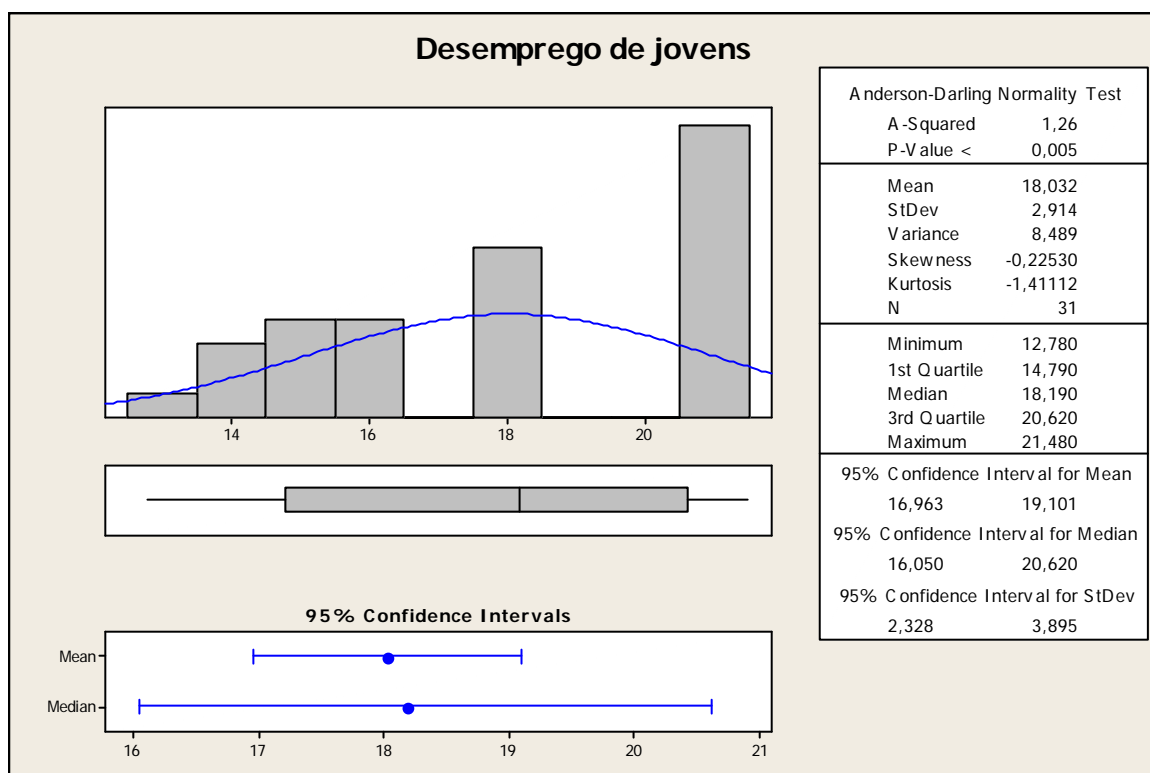
3.2.10 Variável: “Desemprego de jovens de 16 a 29 anos”

Os dados são de 2006 e representam a taxa média de desemprego de jovens de 16 a 29 anos.

```

Stem-and-leaf of Desemprego de jovens  N  = 31
Leaf Unit = 0,10

 1  12  7
 4  13  888
 8  14  7777
 8  15
12  16  0000
12  17
(7) 18  1111333
12  19
12  20  66666
 7  21  4444444
    
```



O histograma apresenta uma distribuição ligeiramente assimétrica para a direita. A curva de densidade indica uma distribuição relativamente anormal, o que é corroborado pelo p-value de < 0,05% no teste de normalidade de Anderson-Darling. O boxplot nos permite confirmar a afirmação precedente, uma vez que a média se localiza mais à direita da linha de distribuição.

Quanto à valores atípicos, esses não encontrados, dado este reforçado pelo box-plot apresentado.

A mediana nos indica que aproximadamente metade das subprefeituras tem um escore menor do que 18 e metade desses, um índice maior. O índice médio apurado é de 18,79, mas o desvio-padrão (medida de dispersão) é de 2,91. O fator de desigualdade das subprefeituras, valor representado por quantas vezes a subprefeitura com classificação mais baixa está pior daquela em classificação mais alta, é de 1,7 (21,48 maximum/12,78 minimum);

O índice mínimo é de 12,7, e o máximo de 21,4, o que corrobora a grande dispersão dos dados. A mediana é de 18,19, índice este relativamente próximo da média de 18,79, o que indica uma relativa simetria na distribuição.

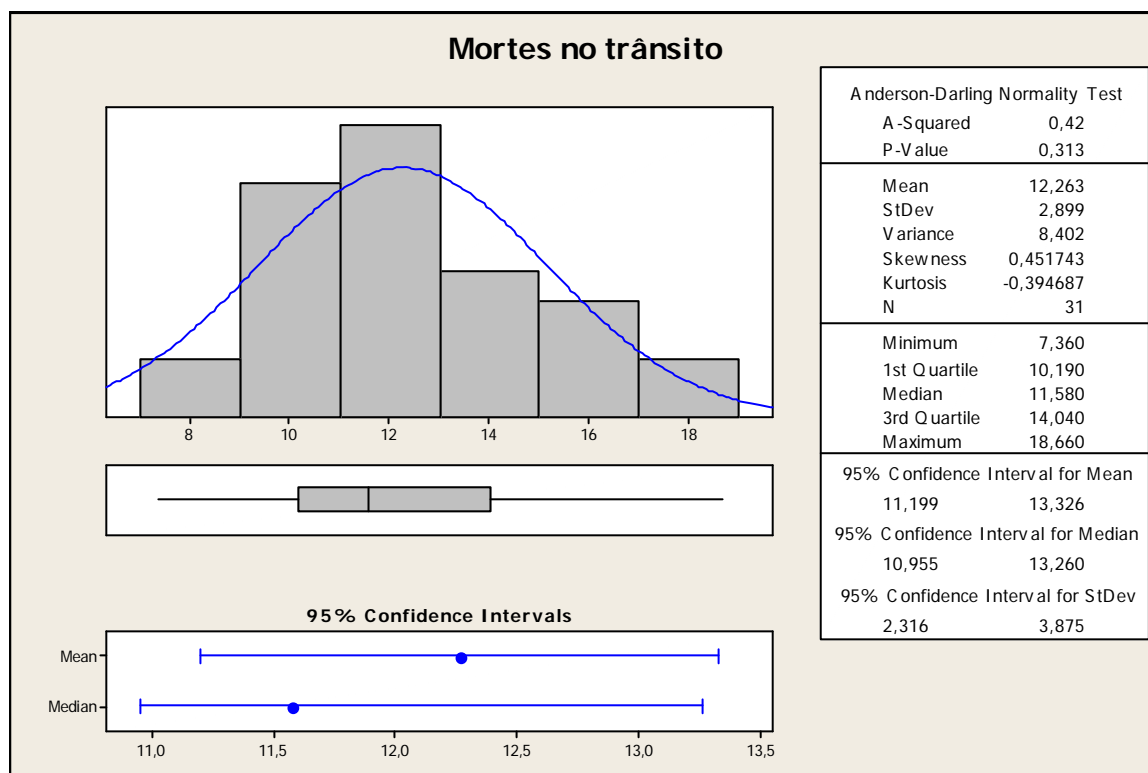
Com 95% de confiança, podemos afirmar que a média encontra-se entre os valores de 16,9 e 19,1.

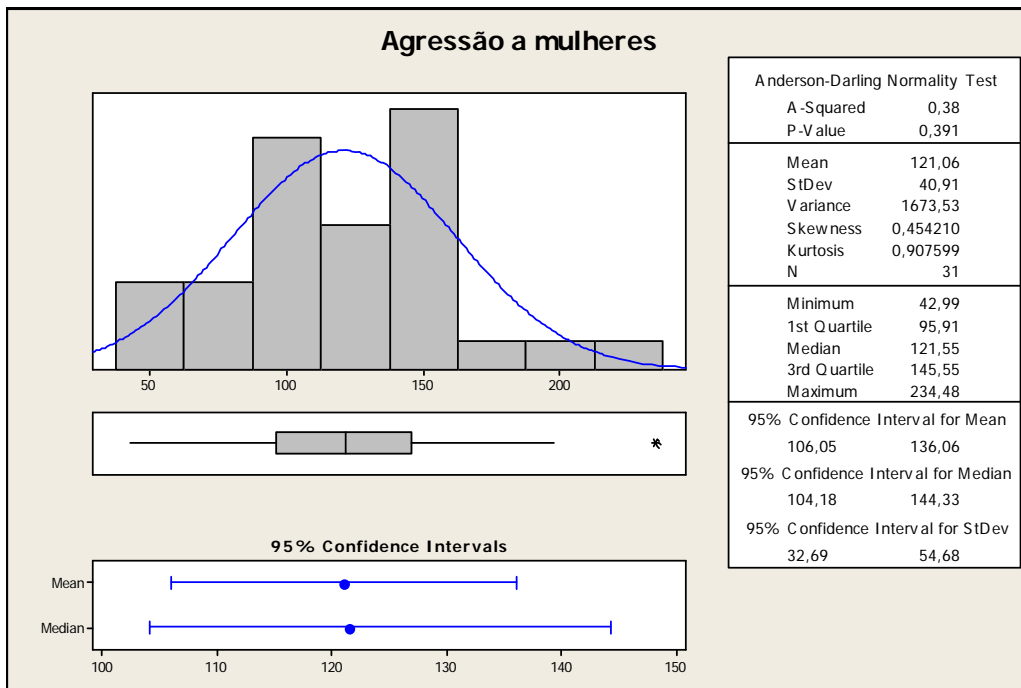
3.2.11 Variável: “Mortes no trânsito”

Os dados são de 2007 e representam o número de mortes em acidentes de trânsito por cem mil habitantes, por local de moradia da vítima.

```
Stem-and-leaf of Mortes no trânsito  N = 31
Leaf Unit = 0,10

 2  7  33
 2  8
 7  9  02447
10 10 145
(7) 11 1223556
14 12 029
11 13 157
 8 14 08
 6 15 78
 4 16 48
 2 17 4
 1 18 6
```

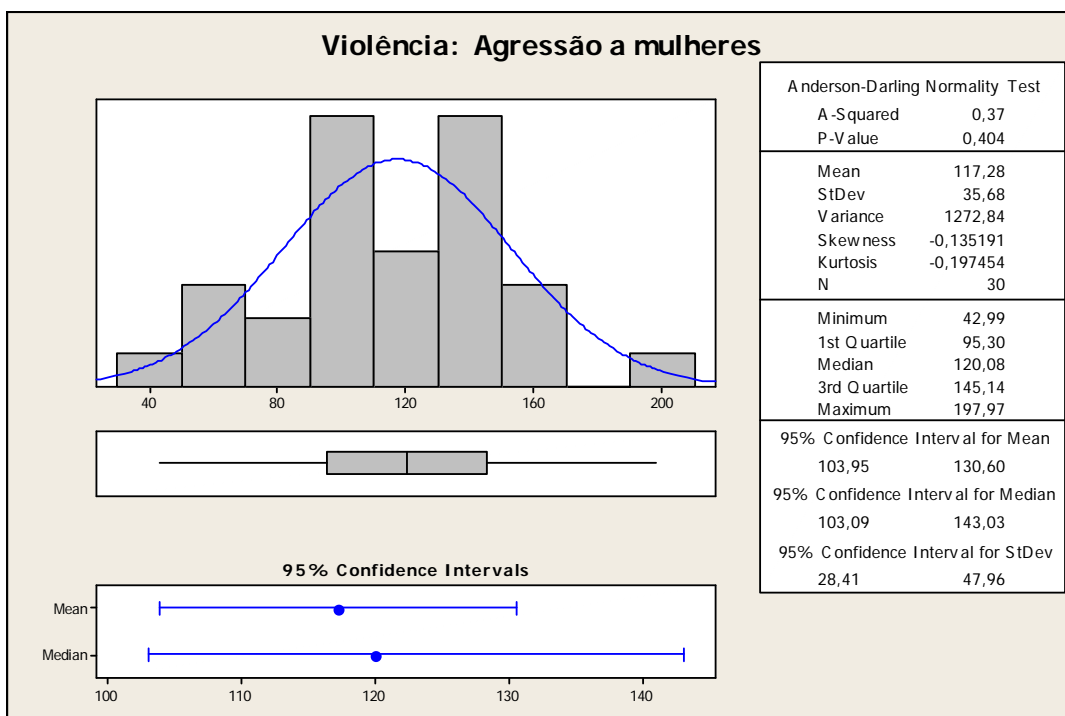




Trata-se de uma distribuição praticamente simétrica, o boxplot permite corroborar a afirmação precedente, e também observa-se que a mediana está relativamente próxima da média, demonstrando a simetria dos valores.

A mediana nos indica que aproximadamente metade das subprefeituras tem um índice de agressão contra mulheres menor do que 121,55 e metade dos países atingem um índice maior. O menor índice é de Parelheiros e o maior de Guainases – como podemos observar no gráfico, este outlier encontra-se longe da média.

Retirando o outlier o gráfico permanece em simetria (ver abaixo). Com 95% de confiança, pode-se afirmar que a média para este indicador está em uma faixa entre 103,95 e 130,60..

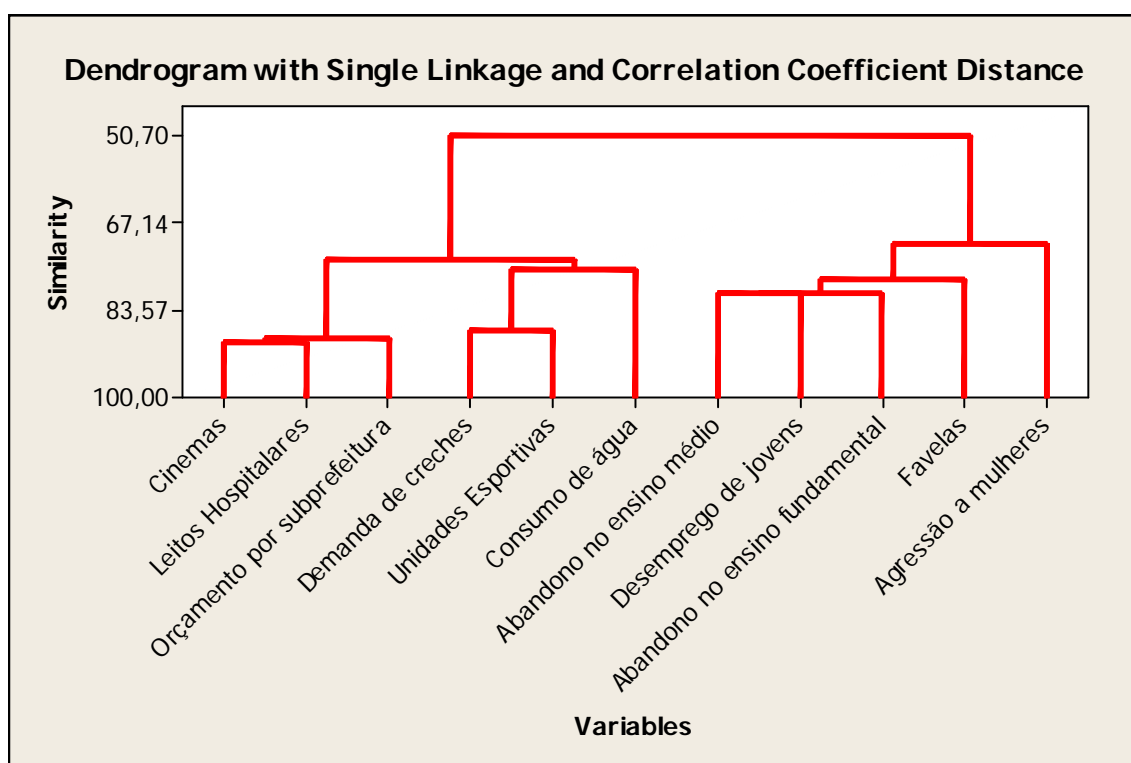


4. RELAÇÕES ENTRE VARIÁVEIS

4.1 Dendogramas

A análise de dendogramas nos permite verificar quais variáveis estão mais próximas uma das outras, quando os “grupo” são inicialmente desconhecidos. O objetivo é que se possa agrupar variáveis, diminuindo o número de fatores do modelo. Obviamente deve-se verificar quais os agrupamentos que fazem sentido para os dados analisados.

Segue abaixo o dendograma para a relação entre as variáveis, considerando os dados acima, ou seja, com o modelo contendo todas as variáveis:



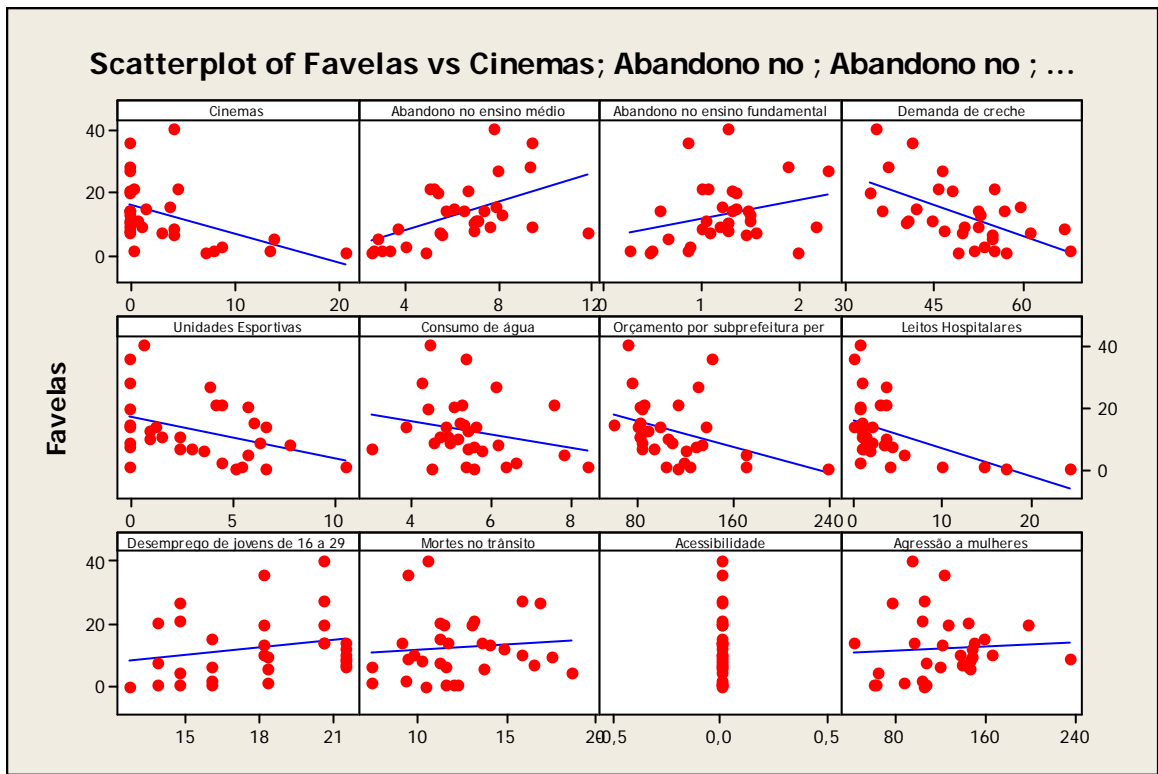
A partir do dendograma escolhemos como variável básica *Favelas* a fim de realizar as próximas análises.

4.2 Gráfico de Regressões

A relação entre duas variáveis quantitativas é apresentada através de um gráfico, e o mais adequado para esse fim é o gráfico de dispersão. Visando buscar prováveis relações entre as variáveis analisadas, procura-se interpretar os gráficos de dispersão, quanto a seu padrão geral e seus desvios relativos ao padrão.

Como dito anteriormente, compararemos as demais variáveis com a variável básica escolhida *Favelas*.

Primeiramente realizamos o gráfico scatterplot que nos mostra as correlações entre as variáveis:



Da análise das correlações acima, percebe-se que indicador Favelas possui associação relativamente fraca, porém positiva com as variáveis Abandono no ensino médio, abandono no ensino fundamental, desemprego de jovens, mortes no trânsito e agressão a mulheres. Assim, o movimento do SWB é acompanhado pelo movimento das demais variáveis citadas. Constata-se a correlação fraca e negativa entre Favelas e cinemas, demanda de creches, unidades esportivas, consumo de água, leitos hospitalares e orçamento.

Os indicadores Favelas e Demanda de creches e abandono no ensino médio são as variáveis que demonstram uma intensidade mais forte na correlação. As outras variáveis a intensidade das correlações é fraca, representada por uma grande dispersão nos grupos em análise.

4.3 Correlações

Para as variáveis em análise, apresenta-se a matriz de correlação incluindo o teste de significância p-value. Para a correlação, utilizou-se do índice de Pearson. A correlação é sempre um número entre zero e um e mede a intensidade de relações lineares.

Correlations: Favelas; Cinemas; Abandono no ; Abandono no ; Demanda de c; ...

	Favelas	Cinemas	Abandono no	Abandono no
Cinemas	-0,454 0,010			
Abandono no	0,520 0,003	-0,570 0,001		
Abandono no	0,294 0,109	-0,228 0,216	0,644 0,000	
Demanda de c	-0,576 0,001	0,363 0,045	-0,477 0,007	-0,365 0,043

Unidades Esp	-0,380 0,035	0,342 0,060	-0,532 0,002	-0,336 0,065
Consumo de á	-0,247 0,181	0,489 0,005	-0,653 0,000	-0,513 0,003
Orçamento po	-0,370 0,040	0,770 0,000	-0,379 0,035	-0,033 0,860
Leitos Hospi	-0,491 0,005	0,793 0,000	-0,515 0,003	-0,170 0,360
Desemprego d	0,235 0,204	-0,666 0,000	0,633 0,000	0,345 0,058
Mortes no tr	0,090 0,631	-0,017 0,926	-0,077 0,682	0,075 0,687
Acessibilida	* *	* *	* *	* *
Agressão a m	0,069 0,714	-0,403 0,025	0,486 0,006	0,366 0,043
	Demanda de c	Unidades Esp	Consumo de á	Orçamento po
Unidades Esp	0,717 0,000			
Consumo de á	0,503 0,004	0,269 0,144		
Orçamento po	0,221 0,232	0,144 0,441	0,400 0,026	
Leitos Hospi	0,291 0,112	0,321 0,078	0,281 0,126	0,721 0,000
Desemprego d	-0,501 0,004	-0,553 0,001	-0,611 0,000	-0,578 0,001
Mortes no tr	-0,086 0,646	-0,052 0,780	0,283 0,123	0,008 0,966
Acessibilida	* *	* *	* *	* *
Agressão a m	0,043 0,819	-0,108 0,565	-0,294 0,109	-0,365 0,044
	Leitos Hospi	Desemprego d	Mortes no tr	Acessibilida
Desemprego d	-0,597 0,000			
Mortes no tr	0,006 0,973	-0,116 0,533		
Acessibilida	* *	* *	* *	
Agressão a m	-0,373 0,039	0,392 0,029	-0,041 0,827	* *
Cell Contents: Pearson correlation				
P-Value				
* NOTE * All values in column are identical.				

Os coeficientes de correlação apresentados indicam a intensidade e o sentido da relação linear entre as variáveis quantitativas, ou seja, quanto maior a correlação, maior esta relação, que pode ser positiva ou negativa.

A confiabilidade da correlação deve ser analisada pelo p-value, quanto mais próximo de 0 mais confiável. Pode-se dizer que a maioria dos coeficientes são confiáveis, mesmo nenhum sendo igual a 0.

No caso do indicador Favelas, podemos observar que ele se correlaciona com as variáveis Demanda de Creches e Leitos Hospitalares, mesmo apresentando relação negativa com ambas, e a variáveis abandono no ensino médio que apresenta relação positiva.

As duas exceções são com as variáveis Mortes no trânsito que apresenta o p-value de 0,63 e a variável Agressão a mulheres com o p-value de 0,714, significando assim dizer que as correlações apresentadas não são absolutamente confiáveis.

4.4 Análise de Regressão

A regressão é a expressão numérica da reta de tendência visualizada nos sub-itens anteriormente analisados, ou seja, ela descreve a relação entre uma variável explicativa e uma variável de resposta.

Regression Analysis: Favelas versus Cinemas; Abandono no ensi; ...

* Acessibilidade has all values = 0
 * Acessibilidade has been removed from the equation.

The regression equation is
 Favelas = 58,2 - 0,256 Cinemas + 2,64 Abandono no ensino médio
 + 1,52 Abandono no ensino fundamental - 0,829 Demanda de creche
 + 0,820 Unidades Esportivas + 3,67 Consumo de água
 - 0,0691 Orçamento por subprefeitura per - 0,334 Leitos Hospitalares
 - 1,58 Desemprego de jovens de 16 a 29 - 0,315 Mortes no trânsito
 - 0,0268 Agressão a mulheres

Predictor	Coef	SE Coef	T	P
Constant	58,24	26,37	2,21	0,040
Cinemas	-0,2557	0,5508	-0,46	0,648
Abandono no ensino médio	2,637	1,153	2,29	0,034
Abandono no ensino fundamental	1,522	4,173	0,36	0,719
Demanda de creche	-0,8293	0,2806	-2,96	0,008
Unidades Esportivas	0,8195	0,8492	0,97	0,347
Consumo de água	3,669	2,425	1,51	0,147
Orçamento por subprefeitura per	-0,06907	0,06660	-1,04	0,313
Leitos Hospitalares	-0,3345	0,4666	-0,72	0,482
Desemprego de jovens de 16 a 29	-1,5764	0,8257	-1,91	0,071
Mortes no trânsito	-0,3149	0,5419	-0,58	0,568
Agressão a mulheres	-0,02685	0,04306	-0,62	0,540

S = 7,31051 **R-Sq = 67,1%** R-Sq(adj) = 48,0%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	11	2066,49	187,86	3,52	0,008
Residual Error	19	1015,43	53,44		
Total	30	3081,92			

Source	DF	Seq SS
Cinemas	1	635,82
Abandono no ensino médio	1	310,74
Abandono no ensino fundamental	1	0,40
Demanda de creche	1	390,14
Unidades Esportivas	1	62,61
Consumo de água	1	370,35
Orçamento por subprefeitura per	1	30,45
Leitos Hospitalares	1	13,25
Desemprego de jovens de 16 a 29	1	213,50
Mortes no trânsito	1	18,46
Agressão a mulheres	1	20,78

Unusual Observations

Obs	Cinemas	Favelas	Fit	SE Fit	Residual	St Resid
3	4,2	40,41	24,09	4,27	16,32	2,75R
7	0,0	6,64	17,58	5,06	-10,94	-2,08R
21	0,0	35,67	25,13	5,48	10,54	2,18R

R denotes an observation with a large standardized residual.

Pela análise de regressão as variáveis explicam 67,1% (r-quadrado) a variável que foi escolhida, sendo que pelo número P observamos que o menor número é o da variável *Demanda de creche*, podemos assim inferir que esta é a que mais explica a variável violência. Comprovaremos isto com a análise do stepwise.

Com o objetivo de procurarmos uma melhora nas informações estatísticas, ou seja, uma melhora no poder explicativo da regressão, opta-se por utilizar a metodologia Stepwise. Este método estabelece a inclusão no modelo, de uma variável de cada vez. Desta forma, são deixadas no modelo, as combinações de variáveis que otimizam seu poder de explicação:

Stepwise Regression: Favelas versus Cinemas; Abandono no ensi; ...

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

Response is Favelas on 10 predictors, with N = 31

Step	1	2	3	4
Constant	45,79	42,24	77,39	71,26
Demanda de creche	-0,67	-0,55	-0,72	-0,63
T-Value	-3,80	-3,22	-4,07	-3,67
P-Value	0,001	0,003	0,000	0,001
Leitos Hospitalares		-0,64	-1,01	-0,86
T-Value		-2,41	-3,38	-3,00
P-Value		0,023	0,002	0,006
Desemprego de jovens de 16 a 29			-1,41	-1,92
T-Value			-2,24	-3,02
P-Value			0,034	0,006
Abandono no ensino médio				1,63
T-Value				2,18
P-Value				0,039
S	8,43	7,81	7,30	6,84
R-Sq	33,19	44,64	53,31	60,51
R-Sq(adj)	30,89	40,68	48,12	54,43
Mallows C-p	12,7	7,9	4,8	2,5

O modelo gerado manteve apenas as variáveis Demanda de creche, leitos hospitalares e desemprego de jovens, incluídos na equação com sinal negativo, e Abandono no ensino médio incluído na equação com sinal positivo.

Como podemos observar a variável Demanda de Creche explica 33,19% da variável Favelas. As quatro variáveis acima representam juntas 60,51%.

4.5 Box-Plots e Two-sample Test

Realizaremos neste item a comparação com Box-plots e o two-sample das 4 variáveis que de acordo com a análise de regressão se correlacionam melhor com a variável Favela.

a. Favelas x Demanda de Creche

Two-Sample T-Test and CI: Favelas; Demanda de creche

Two-sample T for Favelas vs Demanda de creche

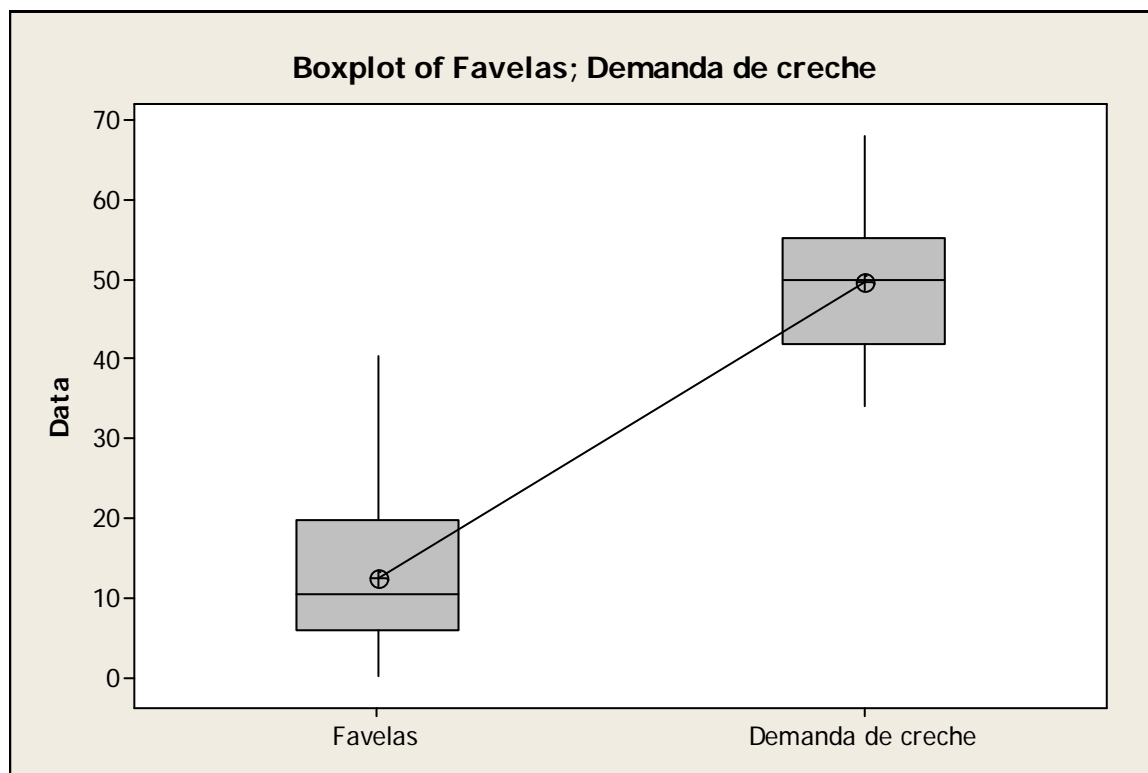
	N	Mean	StDev	SE Mean
Favelas	31	12,6	10,1	1,8
Demanda de crech	31	49,54	8,73	1,6

Difference = mu (Favelas) - mu (Demanda de creche)

Estimate for difference: -36,8958

95% CI for difference: (-41,7044; -32,0872)

T-Test of difference = 0 (vs not =): T-Value = -15,36 P-Value = 0,000 DF = 58



b. Favelas x Leitos Hospitalares

Two-Sample T-Test and CI: Favelas; Leitos Hospitalares

Two-sample T for Favelas vs Leitos Hospitalares

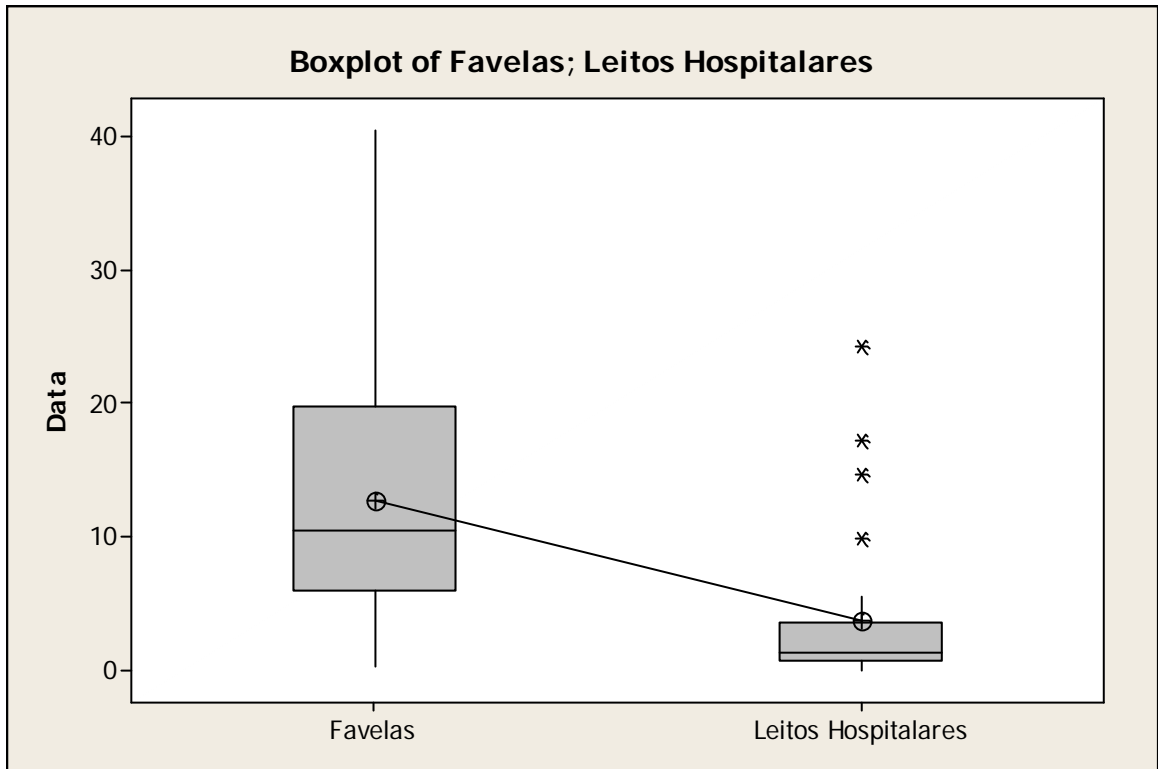
	N	Mean	StDev	SE Mean
Favelas	31	12,6	10,1	1,8
Leitos Hospitalares	31	3,67	5,56	1,0

Difference = μ (Favelas) - μ (Leitos Hospitalares)

Estimate for difference: 8,97516

95% CI for difference: (4,79625; 13,15407)

T-Test of difference = 0 (vs not =): T-Value = 4,32 P-Value = 0,000 DF = 46



c. Favelas x Desemprego de jovens

Two-Sample T-Test and CI: Favelas; Desemprego de jovens de 16 a 29

Two-sample T for Favelas vs Desemprego de jovens de 16 a 29

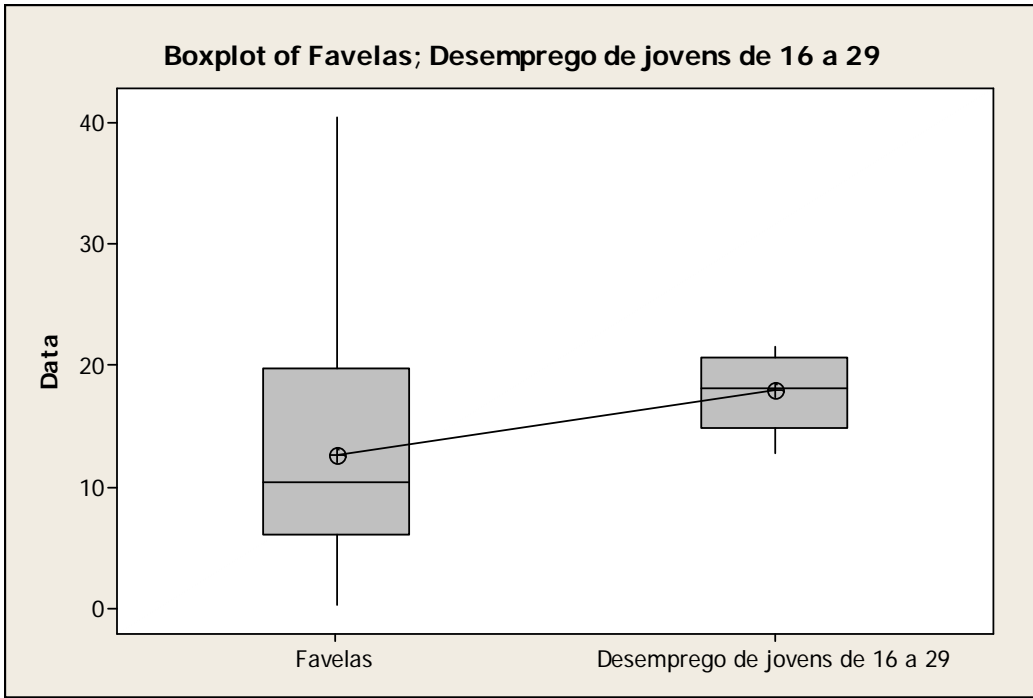
	N	Mean	StDev	SE Mean
Favelas	31	12,6	10,1	1,8
Desemprego de jo	31	18,03	2,91	0,52

Difference = μ (Favelas) - μ (Desemprego de jovens de 16 a 29)

Estimate for difference: -5,38806

95% CI for difference: (-9,23740; -1,53873)

T-Test of difference = 0 (vs not =): T-Value = -2,84 P-Value = 0,007 DF = 34



d. Favelas x Abandono no ensino médio

Two-Sample T-Test and CI: Favelas; Abandono no ensino médio

Two-sample T for Favelas vs Abandono no ensino médio

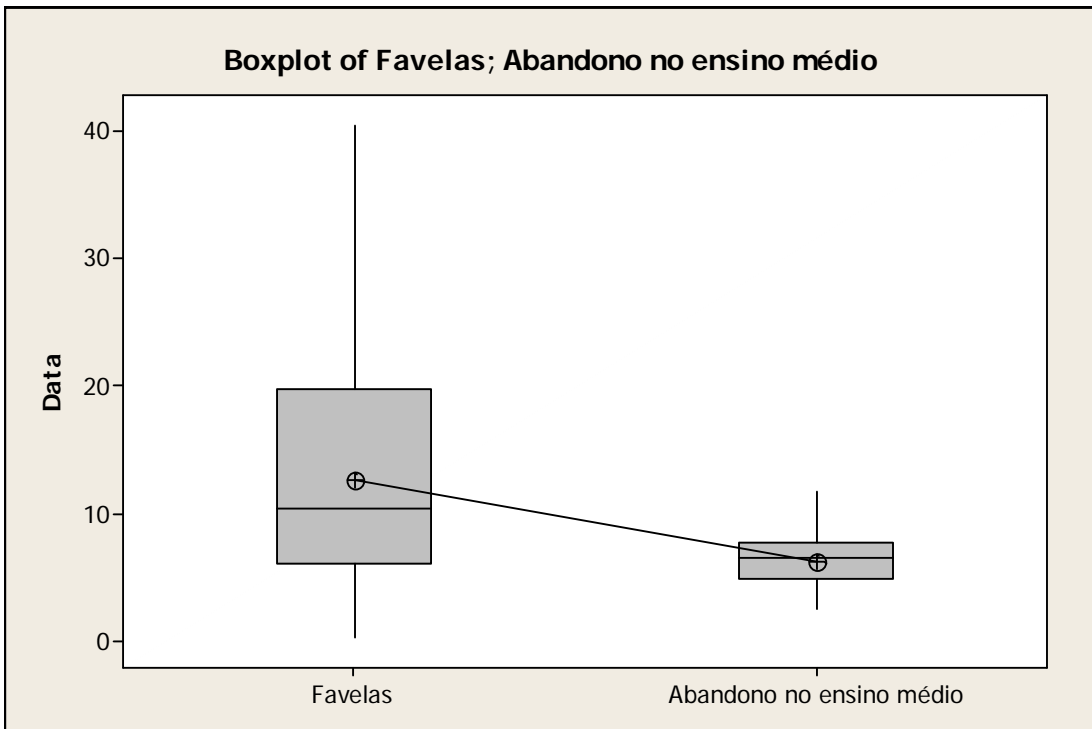
	N	Mean	StDev	SE Mean
Favelas	31	12,6	10,1	1,8
Abandono no ensi	31	6,20	2,27	0,41

Difference = mu (Favelas) - mu (Abandono no ensino médio)

Estimate for difference: 6,44452

95% CI for difference: (2,64878; 10,24025)

T-Test of difference = 0 (vs not =): T-Value = 3,45 P-Value = 0,002 DF = 33



5. ANÁLISE DE CLUSTERS

Neste tópico, o objetivo básico da análise é descobrir agrupamentos naturais entre as variáveis, sendo o agrupamento feito com base nas similaridades e dissimilaridades, caracterizadas pelas distâncias entre variáveis. O objetivo de agrupar observações é tentar unir em grupos distintos, países com características semelhantes em relação a todas as variáveis analisadas.

5.1 Cluster de Variáveis

Para as variáveis analisadas, temos a seguinte representação dos níveis de similaridade e distância entre estas:

Cluster Analysis of Variables: Cinemas; Abandono no ; Abandono no ; ...							
Correlation Coefficient Distance, Single Linkage Amalgamation Steps							
Step	Number of clusters	Similarity level	Distance level	Clusters joined	New cluster	Number of obs. in new cluster	
1	11	89,6627	0,206745	1	9	2	
2	10	88,4913	0,230175	1	8	3	
3	9	85,8421	0,283157	4	5	2	
4	8	82,1752	0,356495	2	3	2	
5	7	81,6464	0,367071	2	10	3	
6	6	75,9921	0,480159	2	6	4	
7	5	75,1726	0,496549	4	7	3	
8	4	74,4615	0,510771	1	4	6	
9	3	74,3195	0,513611	2	12	5	
10	2	64,1514	0,716972	1	11	7	
11	1	54,4887	0,910227	1	2	12	

Final Partition

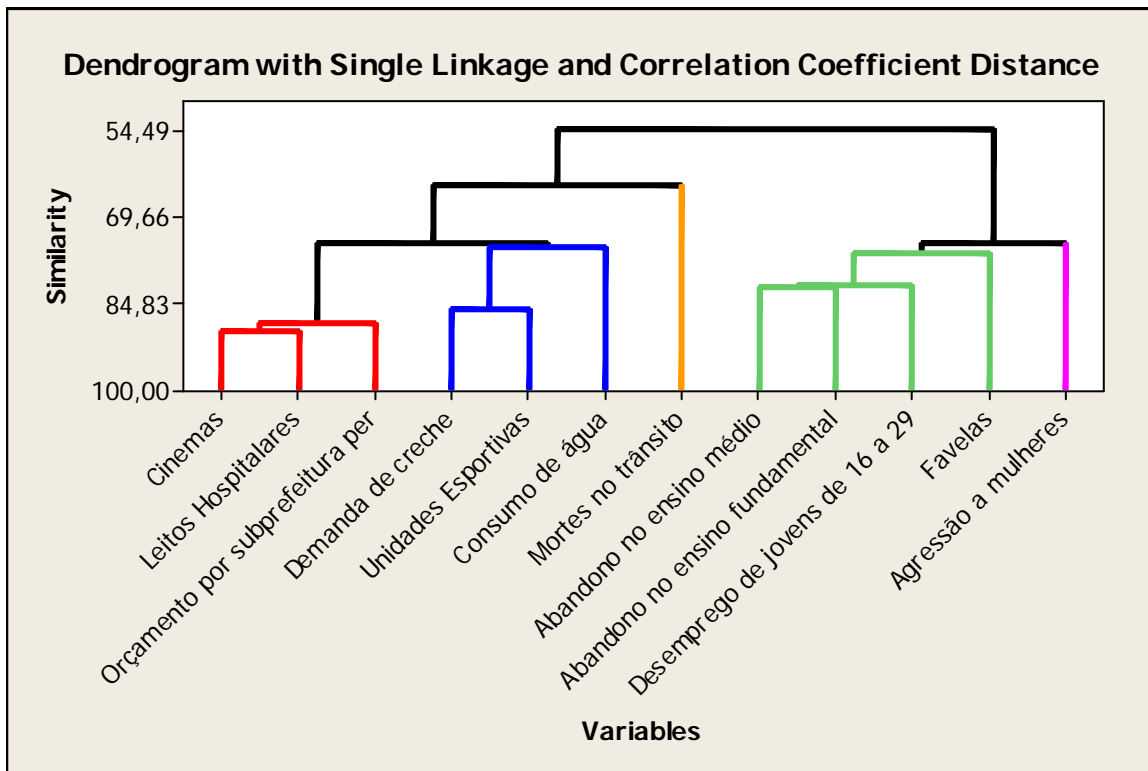
Cluster 1
Cinemas Demanda de creche Unidades Esportivas Consumo de água Orçamento
por subprefeitura per Leitos Hospitalares

Cluster 2
Abandono no ensino médio Abandono no ensino fundamental Favelas Desemprego
de jovens de 16 a 29

Cluster 3
Mortes no trânsito

Cluster 4
Agressão a mulheres

A representação dos clusters é feita via dendograma, que é um tipo específico de diagrama ou representação icônica que organiza determinados fatores e variáveis, que permite a visualização rápida das relações entre as variáveis.



Observamos representados no Dendrograma 4 grupos:

- Primeiro cluster composto por Cinemas, Demanda de creche, Unidades Esportivas, Consumo de água, Orçamento por subprefeitura e Leitos Hospitalares
- Segundo cluster composto por Abandono no ensino médio, Abandono no ensino fundamental, Favelas e Desemprego de jovens de 16 a 29
- Terceiro cluster composto por Mortes no trânsito
- Quarto cluster composto por Agressão a mulheres

5.2 Cluster de Observações

Para a verificação da possibilidade de agrupamentos por observações, temos:

Cluster Analysis of Observations: Cinemas; Abandono no ; Abandono no ; ...

Standardized Variables, 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	30	86,2763	1,24027	5 15	5	2
2	29	84,4517	1,40516	26 27	26	2
3	28	82,4803	1,58332	12 26	12	3
4	27	82,2733	1,60203	5 8	5	3
5	26	81,2957	1,69039	12 13	12	4
6	25	79,7549	1,82963	5 12	5	7
7	24	79,3585	1,86545	4 6	4	2
8	23	79,2123	1,87867	20 29	20	2
9	22	77,4046	2,04203	20 31	20	3
10	21	77,1534	2,06474	4 5	4	9
11	20	77,1263	2,06719	16 18	16	2
12	19	76,5219	2,12181	4 20	4	12
13	18	76,1033	2,15964	4 9	4	13

14	17	73,9112	2,35775	4	23	4	14
15	16	73,7235	2,37471	1	4	1	15
16	15	73,5105	2,39396	1	11	1	16
17	14	71,9108	2,53853	1	16	1	18
18	13	70,9976	2,62106	1	24	1	19
19	12	70,9275	2,62740	1	3	1	20
20	11	70,5678	2,65990	1	17	1	21
21	10	69,5938	2,74793	1	2	1	22
22	9	69,0116	2,80054	1	30	1	23
23	8	67,9632	2,89529	1	19	1	24
24	7	67,8415	2,90629	1	21	1	25
25	6	67,7758	2,91223	1	7	1	26
26	5	63,2127	3,32462	1	10	1	27
27	4	60,6031	3,56046	1	14	1	28
28	3	58,7466	3,72823	22	25	22	2
29	2	55,2826	4,04129	1	22	1	30
30	1	34,1990	5,94670	1	28	1	31

Final Partition
Number of clusters: 1

	Number of observations	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster1	31	*	*	0

6. ANÁLISE DE COMPONENTES PRINCIPAIS

A análise de componentes principais (AC) é um dos métodos estatísticos mais utilizados quando se pretende analisar dados multivariados. A análise de Componentes Principais permite transformar um conjunto de variáveis originais intercorrelacionadas, em um novo conjunto de variáveis não correlacionadas, as componentes principais, cujo objetivo principal é verificar se existe um pequeno número das primeiras componentes principais que seja responsável por explicar uma porção elevada da variação total associada ao conjunto original. Assim, procura-se encontrar uma transformação mais representativa e geralmente mais compacta das observações.

Principal Component Analysis: Favelas; Cinemas; Abandono no ; Abandono no ; Dem

Eigenanalysis of the Correlation Matrix

Eigenvalue	5,2941	1,6579	1,3822	1,0815	0,7556	0,6187	0,3595	0,2388
Proportion	0,441	0,138	0,115	0,090	0,063	0,052	0,030	0,020
Cumulative	0,441	0,579	0,695	0,785	0,848	0,899	0,929	0,949

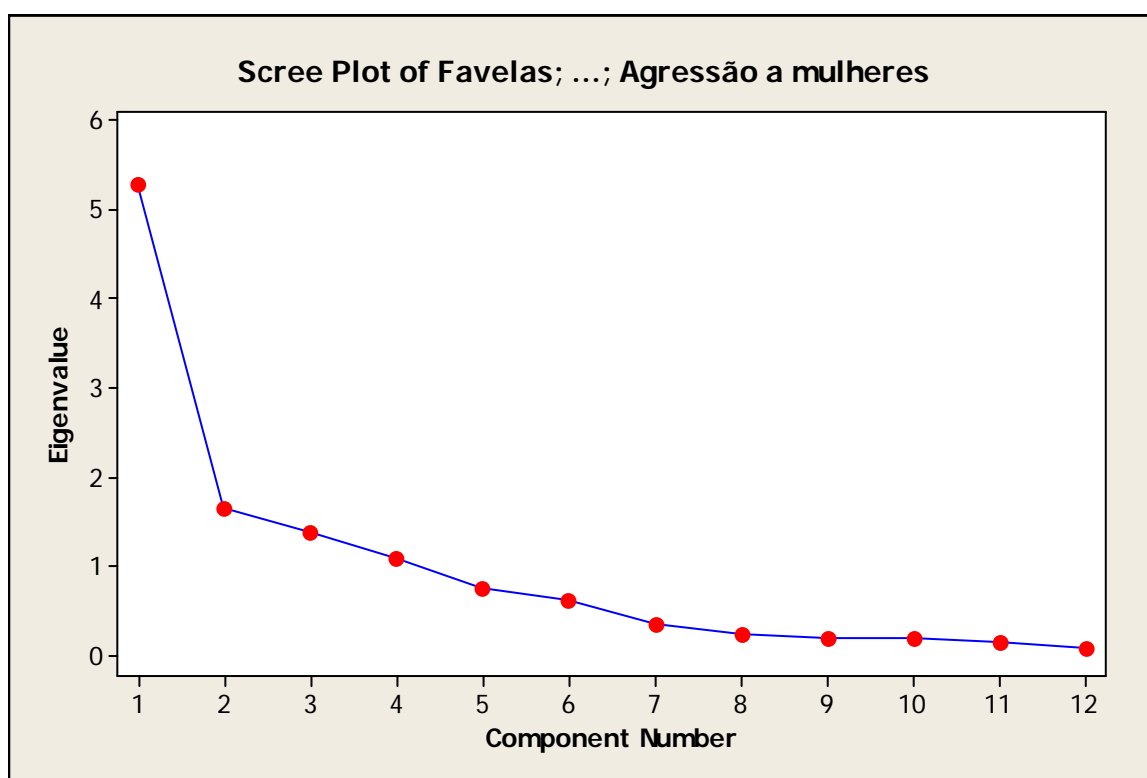
Eigenvalue	0,2002	0,1862	0,1467	0,0786
Proportion	0,017	0,016	0,012	0,007
Cumulative	0,966	0,981	0,993	1,000

Variable	PC1	PC2	PC3	PC4	PC5	PC6
Favelas	0,263	0,171	0,364	0,008	-0,655	-0,215
Cinemas	-0,358	0,291	-0,161	0,012	-0,010	-0,118
Abandono no ensino médio	0,368	0,127	-0,207	0,144	-0,145	-0,149
Abandono no ensino fundamental	0,235	0,327	-0,353	0,413	-0,197	0,115
Demanda de creche	-0,284	-0,448	-0,212	0,220	-0,069	-0,139
Unidades Esportivas	-0,268	-0,393	-0,150	0,108	-0,461	0,421
Consumo de água	-0,306	-0,081	0,383	0,211	0,110	-0,541
Orçamento por subprefeitura per	-0,294	0,441	-0,199	0,081	0,022	-0,249
Leitos Hospitalares	-0,326	0,322	-0,248	-0,006	0,055	0,217
Desemprego de jovens de 16 a 29	0,359	-0,067	-0,093	-0,152	0,455	0,088

Mortes no trânsito	-0,019	0,131	0,474	0,678	0,239	0,411
Agressão a mulheres	0,211	-0,288	-0,355	0,463	0,118	-0,360
Variable	PC7	PC8	PC9	PC10	PC11	PC12
Favelas	-0,215	0,043	0,208	-0,073	-0,325	0,303
Cinemas	-0,148	-0,084	0,668	-0,156	0,474	0,153
Abandono no ensino médio	0,118	0,695	0,123	0,025	0,218	-0,426
Abandono no ensino fundamental	0,434	-0,482	0,075	0,233	-0,094	-0,041
Demanda de creche	0,351	0,344	0,136	0,254	-0,154	0,504
Unidades Esportivas	-0,004	-0,051	0,110	-0,436	-0,113	-0,364
Consumo de água	0,189	-0,150	0,186	0,083	-0,230	-0,501
Orçamento por subprefeitura per	0,186	0,159	-0,400	-0,574	-0,206	0,138
Leitos Hospitalares	-0,427	0,204	0,085	0,410	-0,504	-0,158
Desemprego de jovens de 16 a 29	0,138	0,019	0,496	-0,359	-0,472	0,050
Mortes no trânsito	-0,057	0,192	0,022	-0,100	0,067	0,129
Agressão a mulheres	-0,576	-0,169	-0,100	-0,132	-0,008	0,035

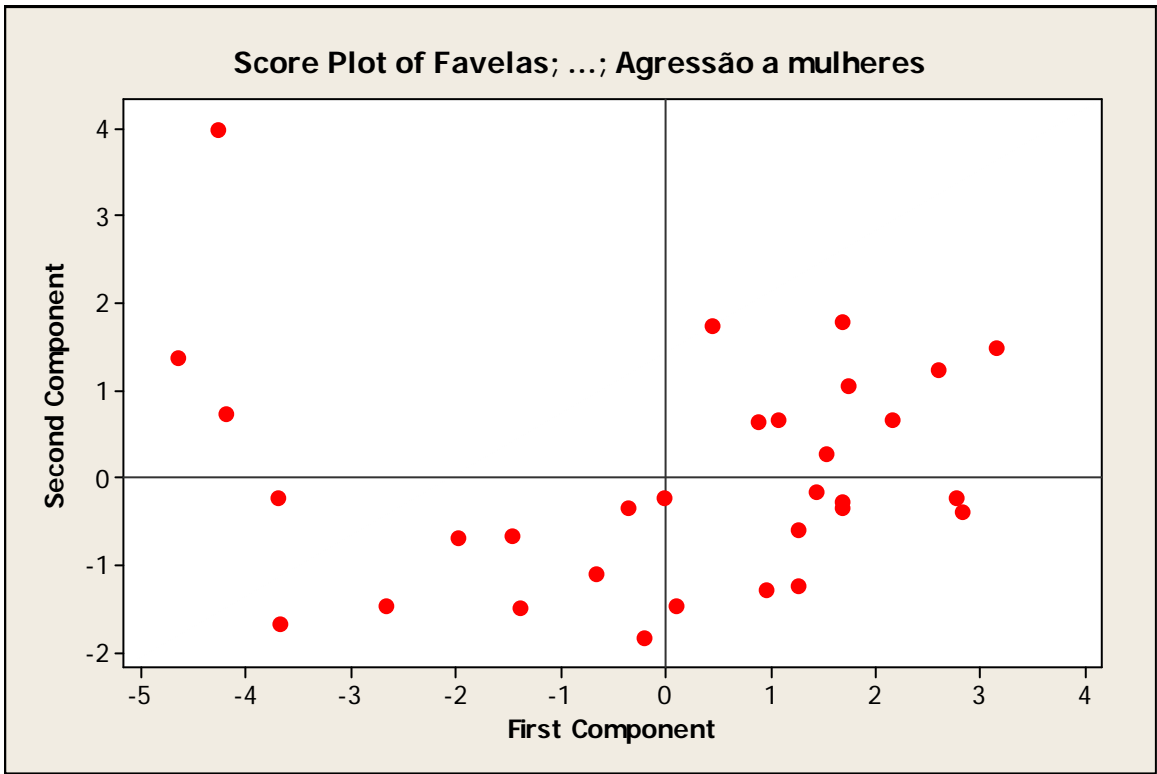
Com a análise dos resultados apurados aprendemos que o índice PC1 é capaz de explicar 44,1% do comportamento das variáveis totais analisadas, um índice baixo. Já os índices PC1 e PC2 somados, são capazes de explicar 57,3%. Ao se agregar o PC3, temos um valor de 69,5%. Incluindo o PC4 temos um poder explicativo de 78,5%, e assim sucessivamente. Só que à medida que vão sendo agregados novos índices, a contribuição marginal diminui sensivelmente.

Sendo assim, ao utilizarmos do PC1, PC2, PC3 e PC4 atingimos um poder explicativo relativamente satisfatório para representar o grupo de variáveis totais.

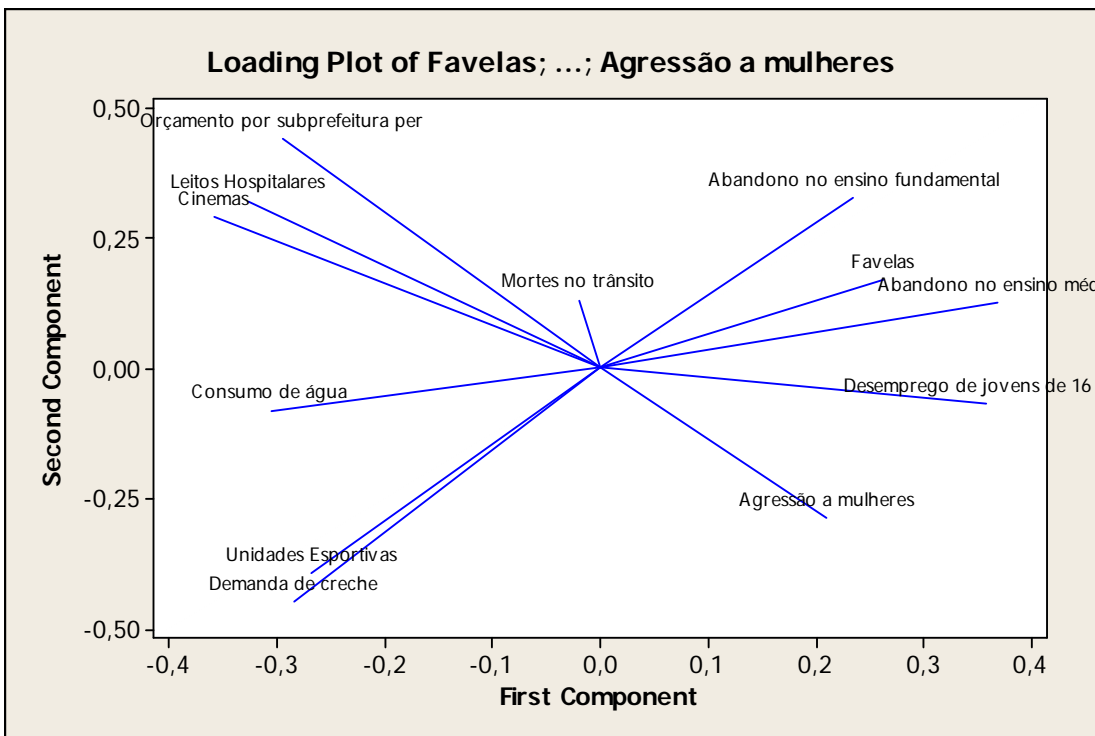


Como as variáveis não são fortemente correlacionadas, houve a necessidade de se ir além do primeiro eixo, chegando-se até o terceiro eixo, para que se atingisse um poder explicativo razoável.

Para o primeiro e segundo componente temos a seguinte representação gráfica através do score plot, que ilustra a importância de cada variável para o valor da variância.



Pelo loadingplot pode-se comprovar a complementariedade das variáveis, que no nosso caso não há correlações fortes:



7. CONSIDERAÇÕES FINAIS

O objetivo deste trabalho foi analisar algumas variáveis do Movimento Nossa São Paulo para verificarmos o atual cenário de qualidade de vida da cidade de São Paulo. Para a realização das análises de comparações escolhemos, aleatoriamente, a variável Favelas. Dessa maneira conseguimos utilizar grande dos gráficos estudados na disciplina Métodos Quantitativos e realizamos um amplo estudo.

Através dos gráficos pudemos observar a grande discrepância entre as subprefeituras e as variáveis, assim como a fraca correlação entre as 13 variáveis.

Vale ressaltar que a cidade ainda carece de inúmeros indicadores importantes, principalmente os referenciados por subprefeituras. O Nossa São Paulo tem como objetivo também de exigir dos poderes públicos que produzam, aprimorem e referenciem geograficamente as principais informações sobre a cidade em suas 31 subprefeituras

