



ANÁLISE ESTATÍSTICA DOS PRINCIPAIS INDICADORES DO MUNICÍPIO DE SÃO PAULO 2004

Fátima Alexandre

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

Objetivo

O objetivo do trabalho é apresentar uma avaliação estatística sobre alguns dados disponíveis no Sumário de Dados do Município de São Paulo 2004.

Introdução

A Prefeitura do Município de São Paulo disponibiliza através de sua página na internet um sumário com os principais indicadores do município, organizados em capítulos temáticos de demografia, indicadores econômicos, saúde, educação, habitação e outros. As 31 sub-prefeituras são mostradas através de indicadores sociais e disponibilidade de equipamentos, serviços e programas municipais em cada uma delas. A Prefeitura coloca que este trabalho é o resultado de um esforço mais amplo na criação de um projeto de gestão da informação, elemento que julga indispensável no processo de modernização da administração pública. De modo geral, a disponibilização de informação sistematizada e compreensível é um pressuposto para o exercício da cidadania.

Neste trabalho são analisados alguns indicadores que podem ajudar na avaliação da Qualidade de Vida Urbana neste município. Foram escolhidos os seguintes dados referentes a cada uma das 31 subprefeituras mencionadas no Sumário:

- população
- renda mensal média do chefe de família
- número de empregos disponíveis
- taxa de analfabetismo
- população favelada
- IDH

O indicador de renda mensal é um valor aproximado, porque foi fornecido através de um gráfico esquemático com coordenadas que dão apenas ordem de grandeza.

O valor do IDH (Índice de Desenvolvimento Humano) apresentado no Sumário é referente a cada bairro componente da Subprefeitura, diferente dos demais indicadores que tem um único valor para a região, portanto foi adotado para este trabalho a média dos valores de IDH apresentados.

O Sumário de Dados do MSP 2004 utiliza como ano de referência e fonte para as tabelas de dados:

- o censo do IBGE de 2000, para quantificar a população da região, a renda mensal e a taxa de analfabetismo;
- a RAIS 2000 do Ministério do Trabalho e Emprego, para o número de empregos;
- a Base Cartográfica Digital das Favelas do MSP 2000 da Secretaria de Habitação e Desenvolvimento Urbano da PMSP / PRODAM, para a população favelada;
- o IDH 2000 fornecido pela Secretaria de Desenvolvimento, Trabalho e Solidariedade - PMSP.

A base de dados

Os dados a serem analisados estão apresentados na tabela abaixo :

subprefeitura	população	renda	n ^a empregos	taxa analfabetismo	pop favelada	IDH
Aricanduva/V.Formosa/Carrão	266838	1220	32704	3,65	7267	0,547
Butantã	377576	2050	58398	4,42	49948	0,658
Campo Limpo	505969	1010	25876	7,13	122738	0,498
Capela do socorro	563922	680	41853	6,81	132177	0,505
Casa Verde/Cachoeirinha	313323	980	39369	4,89	35001	0,505
Cidade Ademar	370797	760	17605	7,03	77510	0,448
Cidade Tiradentes	190657	480	2274	5,78	5873	0,446
Ermelino Matarazo	204951	800	12060	4,90	37818	0,477
Freguesia do O/Brasilândia	392251	780	15081	5,83	51882	0,479
Guaianazes	256319	460	13844	7,72	39742	0,428
Ipiranga	429235	1280	99373	4,15	61590	0,564
Itaim Paulista	359215	520	8262	7,05	28332	0,433
Itaquera	489502	740	30827	5,10	23345	0,475
Jabaquara	214095	1380	27658	4,58	24186	0,528
Lapa	270656	2600	156636	2,42	15130	0,606
Moóca	308161	1720	216354	2,26	6500	0,592
M'Boi Mirim	484966	580	32778	7,42	126559	0,422
Parelheiros	111240	440	4592	9,63	14793	0,315
Penha	475879	980	41285	3,76	28285	0,525
Perus	109116	540	2939	6,87	24608	0,485
Pinheiros	272574	4400	237758	0,94	552	0,824
Pirituba	390530	920	24708	4,63	50766	0,502
Santana/Tucuruvi	327135	1800	51035	2,26	3869	0,625
Santo Amaro	218558	2980	122171	1,92	6985	0,727
São Mateus	381718	600	26949	6,53	38770	0,426
São Miguel	378438	600	13817	7,34	45919	0,422
Sé	373914	2200	488539	2,14	2744	0,634
Tremembe/Jaçanã	255612	980	15866	5,70	21623	0,482
V.Maria/V.Guilherme	304393	940	65817	4,98	15014	0,502
Vila Mariana	313036	3600	137430	1,10	1956	0,774
V.Prudente/Sapopemba	523676	860	39124	5,09	59608	0,491

Apresentação dos resultados

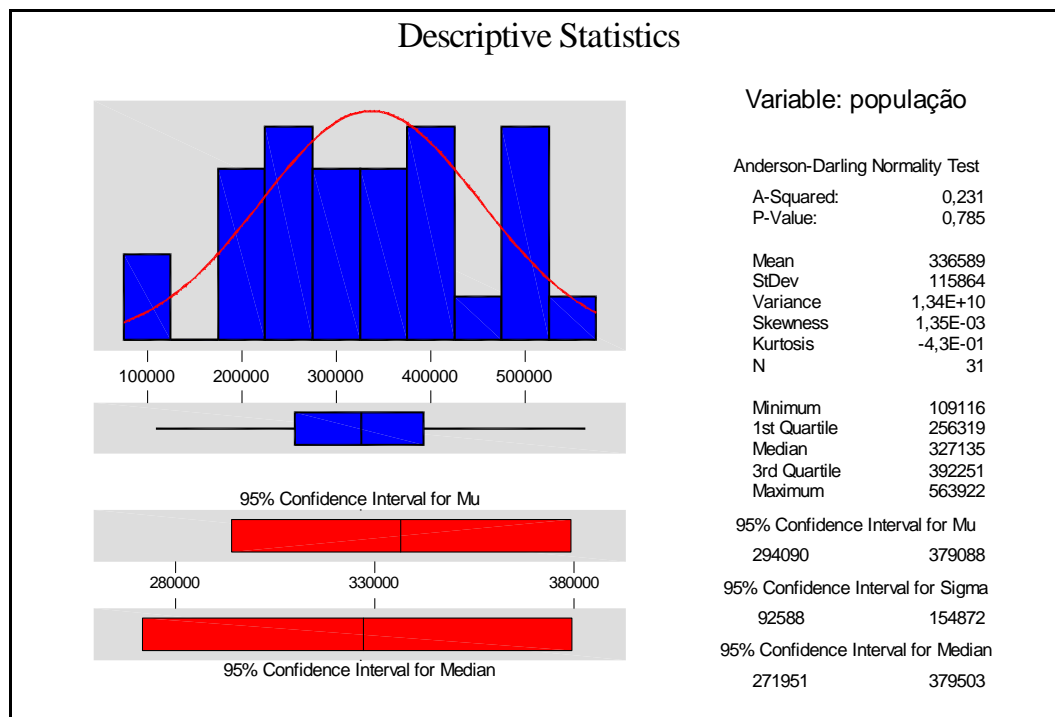
Análise Estatística Descritiva

Um dos componentes da análise estatística é a exploração de dados, comumente denominada de estatística descritiva, que fornece informações sobre a qualidade de seus dados e indica eventuais tendências. Com base nos dados tabelados acima, apresentamos os valores da média, mediana, desvio padrão, valores mínimos, máximos e quartis, utilizando o programa Mini Tab.

Descriptive Statistics: população; renda; nª empregos; taxa analfab; pop favela; IDH						
Variable	N	Mean	Median	TrMean	StDev	SE Mean
população	31	336589	327135	338011	115864	20810
renda	31	1286	940	1147	973	175
nª empre	31	67838	32704	50795	98613	17711
taxa ana	31	4,969	4,980	4,987	2,158	0,388
pop fave	31	37455	28285	33328	35922	6452
IDH	31	0,5273	0,5020	0,5189	0,1101	0,0198
Variable	Minimum	Maximum	Q1	Q3		
populaçã	109116	563922	256319	392251		
renda	440	4400	600	1720		
nª empre	2274	488539	15081	65817		
Taxa ana	0,940	9,630	3,650	6,870		
pop fave	552	132177	7267	50766		
IDH	0,3150	0,8240	0,4480	0,5920		

População

A representação gráfica pode ser feita através do histograma com o ajuste da curva normal, permitindo a visualização dos valores encontrados para a **população** das subprefeituras:



No gráfico acima encontramos abaixo do histograma outra forma de representação gráfica, a do Box Plot, ou seja, um desenho esquemático representando a mediana, Q1, Q3, máximo e mínimo. Para o HDI, podemos verificar a concentração da população, aproximadamente, entre os valores 256.000 e 392.000, sem outlier. Outlier é um valor discrepante que se encontra abaixo de $(Q1 - 1,5 * (Q3 - Q1))$ ou

acima de $(Q3+1,5*(Q3-Q1))$ A variável população apresenta média de 336.589 , média aparada de 338.011 e mediana de 327.135, muito próximas. O desvio padrão é de 115864, relativamente alto.

O Stem and Leaf da população é um esboço do histograma, e serve para dar uma idéia do comportamento dos dados:

Stem-and-Leaf Display: população

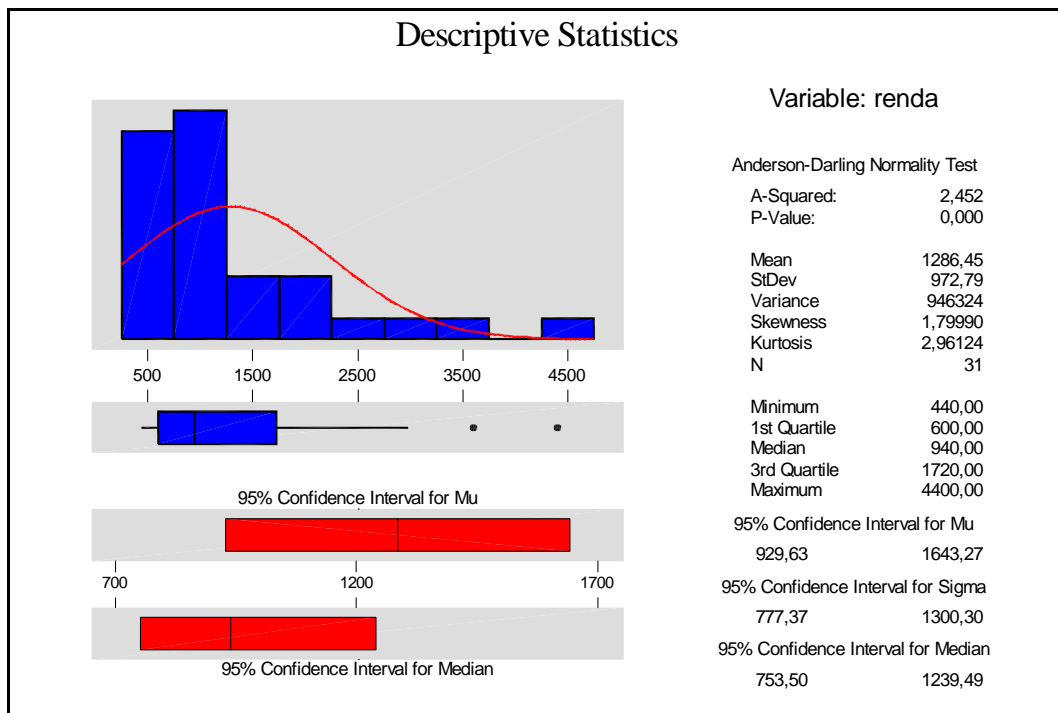
Stem-and-leaf of populaçã N = 31
Leaf Unit = 10000

```

 2    1 01
 3    1 9
 6    2 011
11    2 55677
(5)   3 00112
15    3 57777899
 7    4 2
 6    4 788
 3    5 02
 1    5 6
    
```

Renda mensal média do chefe de família

Abaixo se encontra o histograma com o ajuste da curva normal, permitindo a visualização dos valores encontrados para a variável **renda**:



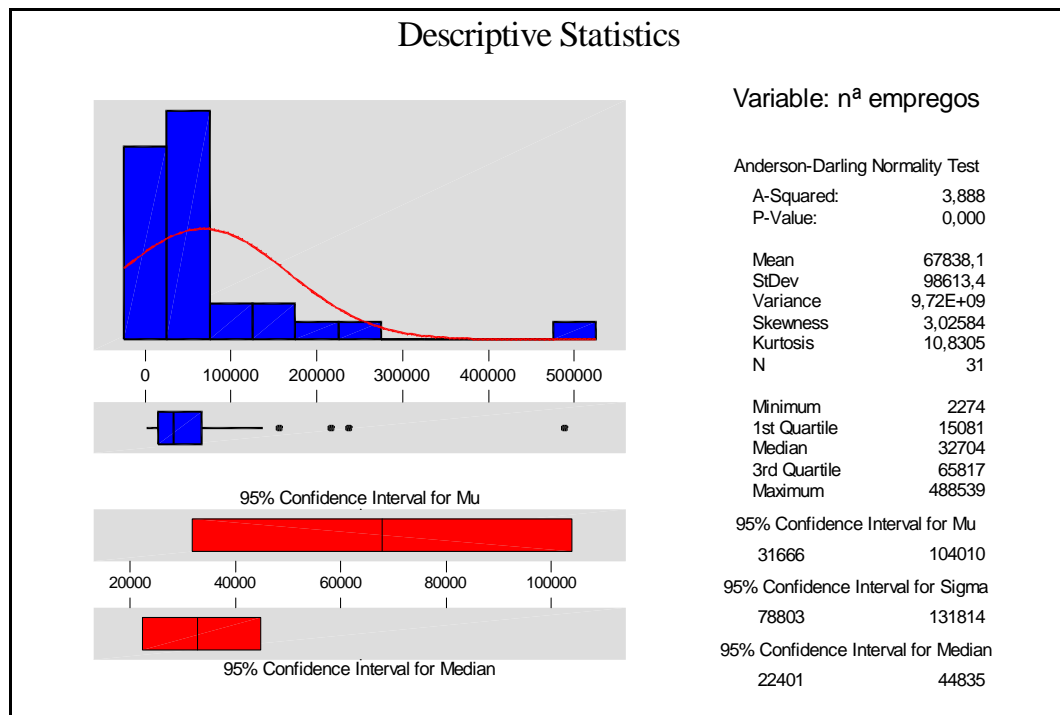
No Box Plot da renda, podemos verificar a posição dos dois Outliers (*) (Pinheiros e Vila Mariana), e a concentração de renda, aproximadamente, entre os valores R\$ 600,00 e R\$ 1720,00. Pinheiros e Vila Mariana apresentam uma renda mensal muito superior às demais subprefeituras e à média do Município de São Paulo (R\$ 1300,00), isso justifica o alto valor do IDH dessas regiões.

O Stem and Leaf dos valores da renda é o seguinte:

Stem-and-Leaf Display: renda	
Stem-and-leaf of renda N = 31	
Leaf Unit = 100	
3	0 444
(16)	0 5556667778899999
12	1 0223
8	1 78
6	2 02
4	2 69
2	3
2	3 6
1	4 4

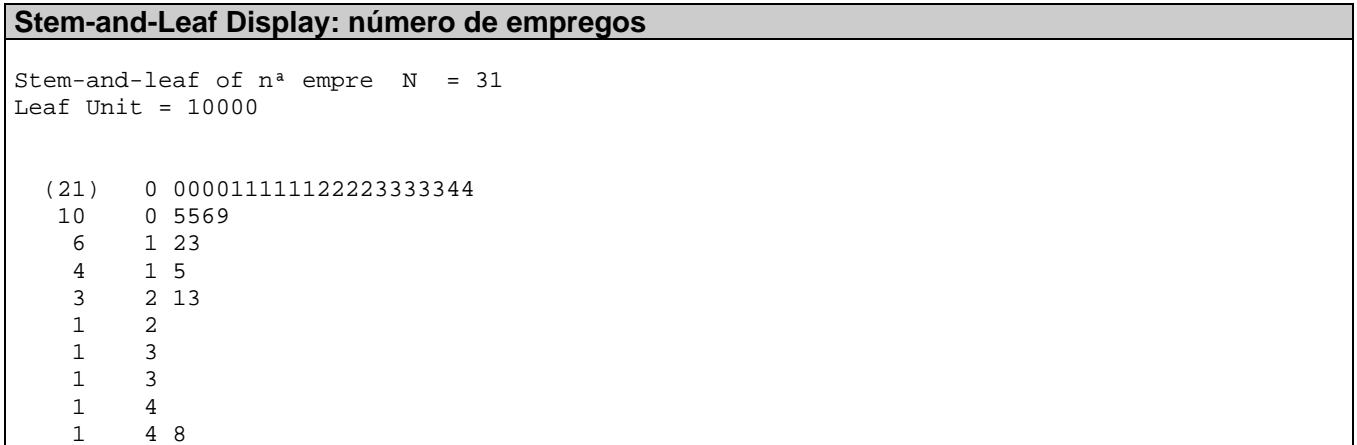
Número de empregos disponíveis

Para visualizar os valores encontrados do índice **número de empregos** existentes na região da subprefeitura, apresentamos o histograma com o ajuste da curva normal:



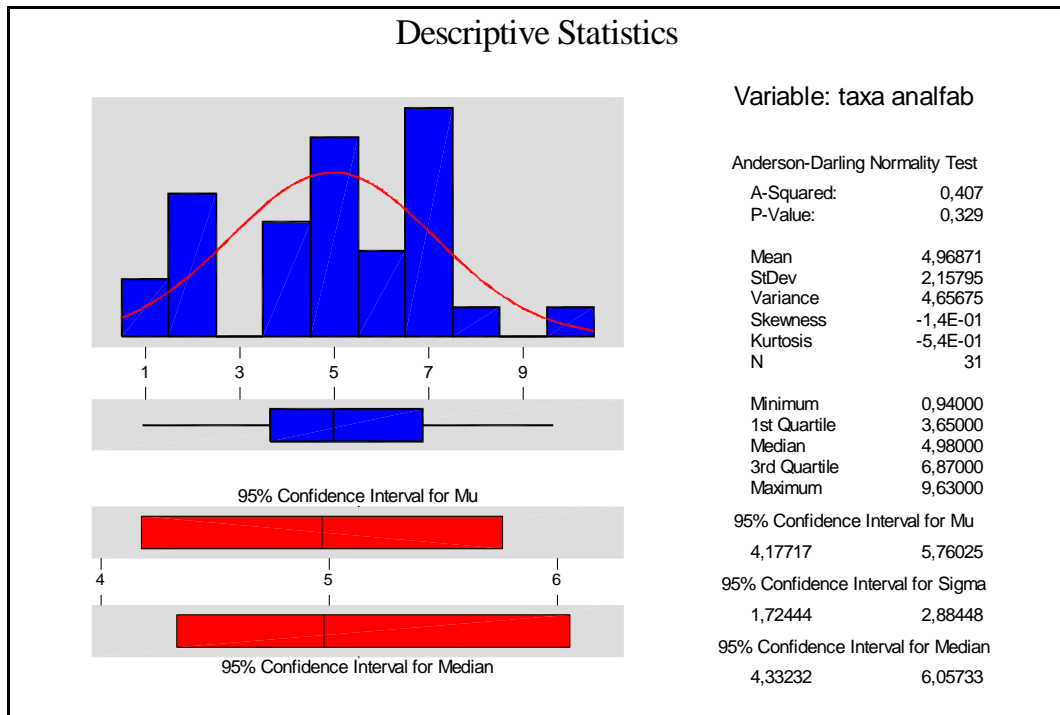
No Box Plot do número de empregos, podemos verificar a posição dos quatro Outliers(*) (Lapa, Mooca, Pinheiros e Sé), e a concentração da quantidade de empregos oferecida está, aproximadamente, entre 15.000 e 66.000.

O Stem and Leaf do número de empregos fica da seguinte forma:



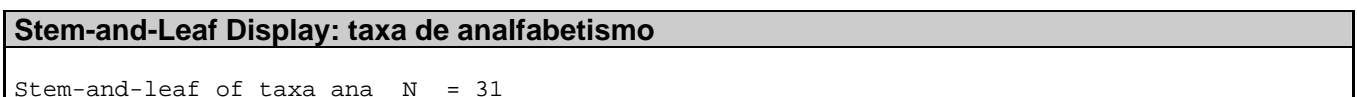
Taxa de analfabetismo

No histograma com o ajuste da curva normal, visualizamos os valores encontrados para o taxa de analfabetismo na região das subprefeituras:



No Box Plot da taxa de analfabetismo verifica-se que não existem Outliers(*), e a concentração da taxa está, aproximadamente, entre 3,65 e 6,87:

O Stem and Leaf para a taxa de analfabetismo fica assim:



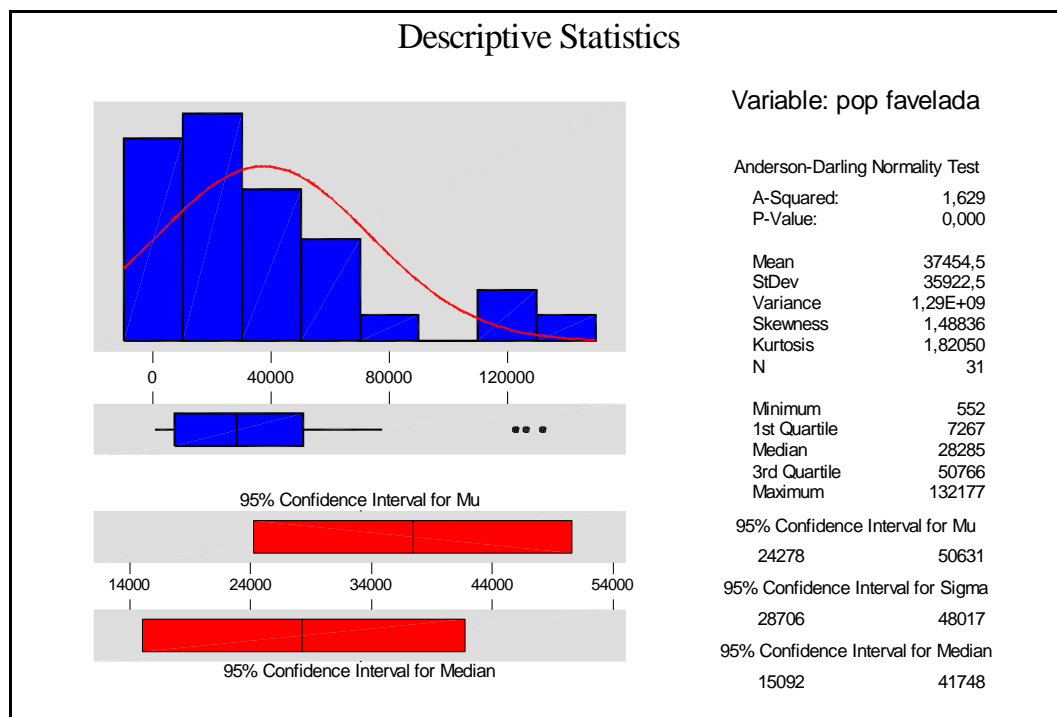
Leaf Unit = 0,10

```
1 0 9
3 1 19
7 2 1224
9 3 67
(7) 4 1456899
15 5 01778
10 6 588
7 7 001347
1 8
1 9 6
```

População Favelada

Na conceituação da Prefeitura, são considerados favela e assemelhados os agrupamentos de, no mínimo, duas unidades habitacionais precariamente construídas e dispostas de forma desordenada em um mesmo lote, cuja propriedade não é legalizada para aqueles que o ocupam.

No histograma com o ajuste da curva normal, visualizamos dos valores encontrados para a quantidade de pessoas que vivem em favelas na região das subprefeituras:



No Box Plot da população favelada verifica-se a presença de três Outliers(*) (Campo Limpo, M'Boi Mirim e Capela do Socorro), e a concentração de favelados está, aproximadamente, entre 7.200 e 50.800.

O Stem and Leaf para a população favelada fica assim:

Stem-and-Leaf Display: população favelada

Stem-and-leaf of pop fave N = 31
Leaf Unit = 1000

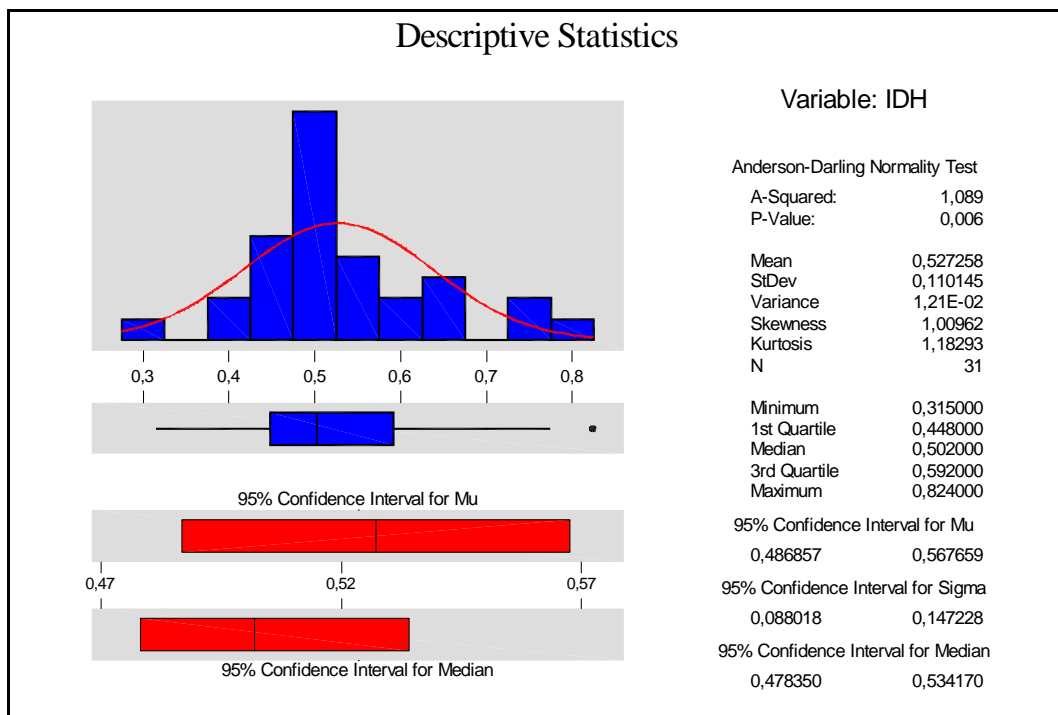
```
 8   0 01235667
11   1 455
(6)  2 134488
14   3 5789
10   4 59
 8   5 019
 5   6 1
 4   7 7
 3   8
 3   9
 3  10
 3  11
 3  12 26
 1  13 2
```

IDH

Na conceituação do documento de referencia, o IDH Índice de Desenvolvimento Humano é uma adaptação de índice criado pela ONU, realizada pela Secretaria de Desenvolvimento, Trabalho e Solidariedade / PMSP, com o objetivo de comparar o grau de desenvolvimento humano entre distritos do município de São Paulo. É composto pelos seguintes indicadores: taxa de mortalidade infantil, porcentagem da população alfabetizada e rendimento dos chefes de família.

Estes indicadores são transformados em índices, que somados formam o IDH. Ele varia de zero (piores condições de desenvolvimento) a um (melhores condições de desenvolvimento).

Para visualizar os valores encontrados do IDH na região da subprefeitura, apresento o histograma com o ajuste da curva normal:



No Box Plot do IDH, podemos verificar a posição de um Outlier(*) (Pinheiros), e a concentração, aproximadamente, entre 0,45 e 0,59.

O Stem and Leaf do IDH fica da seguinte forma:

Stem-and-Leaf Display: IDH	
Stem-and-leaf of IDH	N = 31
Leaf Unit = 0,010	
1	3 1
1	3
8	4 2222344
15	4 7778899
(7)	5 0000224
9	5 69
7	6 023
4	6 5
3	7 2
2	7 7
1	8 2

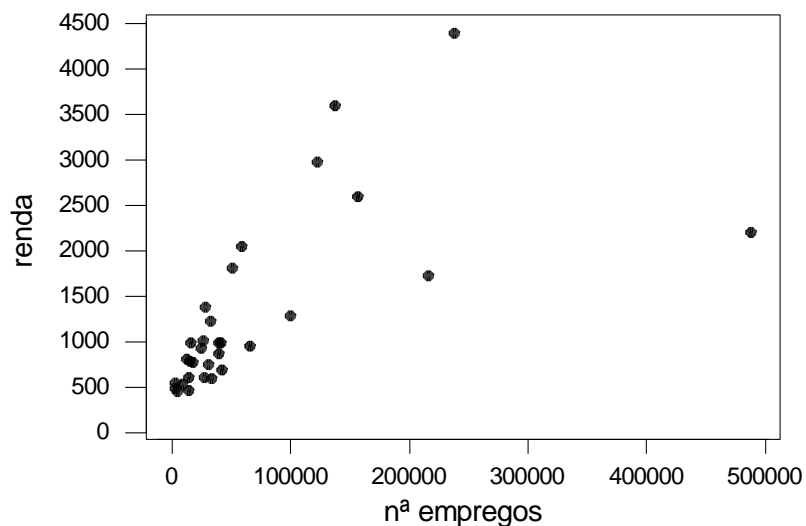
Diagramas de dispersão

O diagrama de dispersão é um gráfico de coordenadas cartesianas, onde em cada eixo temos uma variável, desta forma podemos verificar como se comporta uma variável quantitativa em função de outra também quantitativa.

Foram analisados os seguintes pares de variáveis:

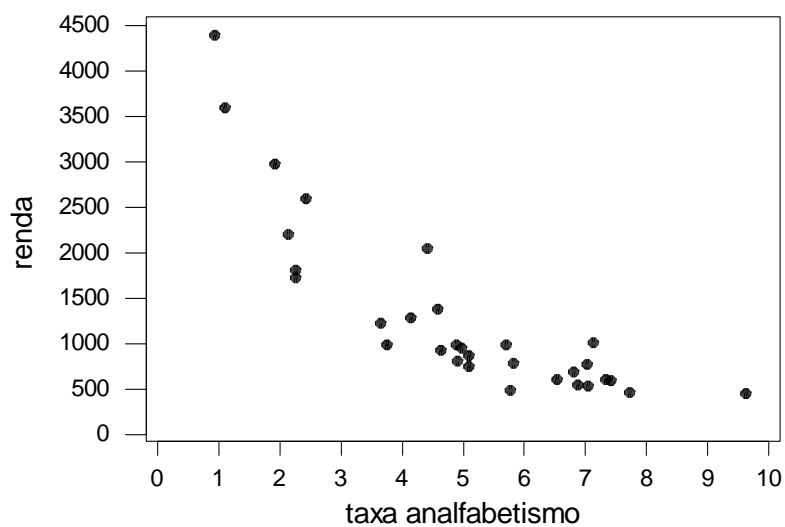
- renda x nº empregos
- renda x taxa de analfabetismo
- renda x população favelada
- renda x IDH
- nº de empregos x taxa de analfabetismo
- nº de empregos x população favelada
- nº de empregos x IDH
- taxa de analfabetismo x população favelada
- taxa de analfabetismo x IDH
- população favelada x IDH

Utilizando-se os valores da renda com o número de empregos, obtemos o seguinte diagrama:



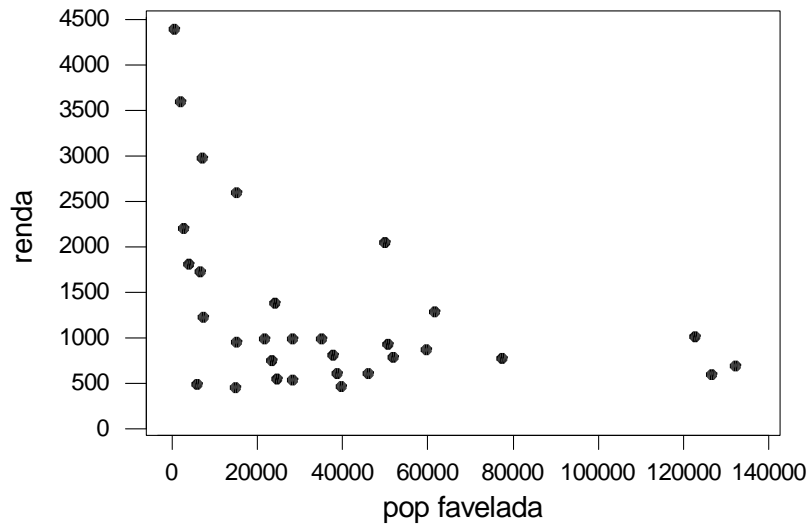
verificamos que o aumento do número de empregos não necessariamente significa um aumento de renda.

Utilizando-se os valores da renda com os de taxa de analfabetismo, obtemos o seguinte diagrama:



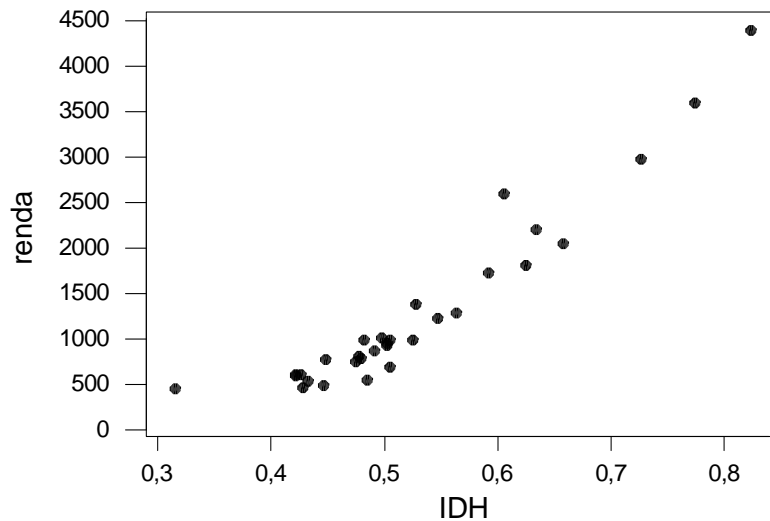
Verifica-se que existe uma tendência na diminuição da renda à medida que aumenta a taxa de analfabetismo.

Utilizando-se os valores de renda com o número de pessoas que vivem em favelas, obtemos :



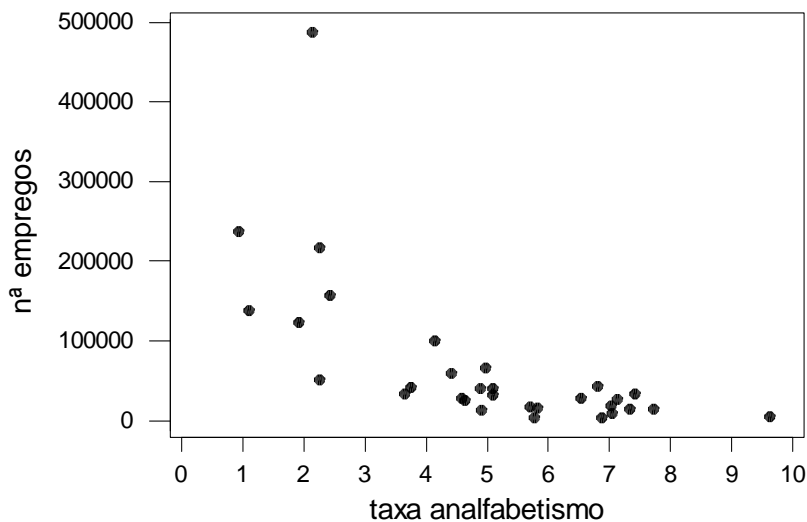
Verifica-se que nos extremos da renda temos os extremos de quantidade de pessoas que vivem em favela, e há uma concentração de renda na faixa de R\$ 500,00 a R\$1500,00, para uma população favelada aproximada entre 5.000 e 60.000 pessoas, mas não uma correlação clara entre as variáveis.

Utilizando-se os valores de renda com os do IDH, obtemos o seguinte diagrama:



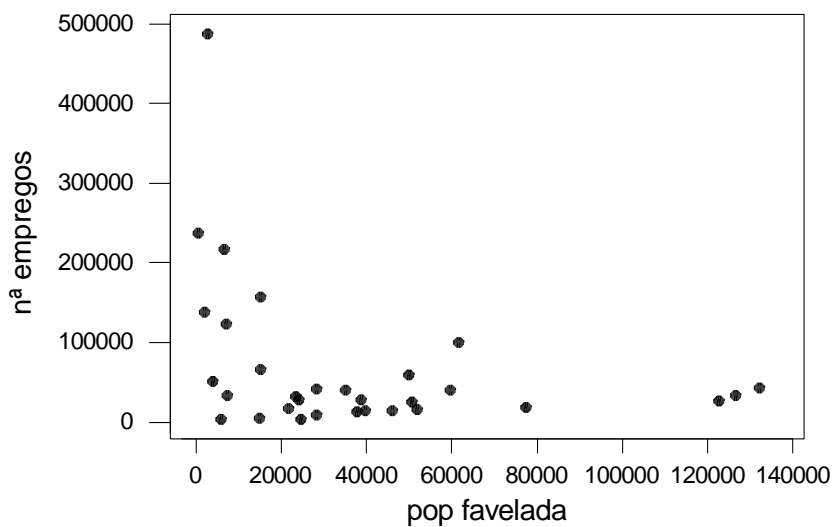
Verifica-se que há uma tendência no aumento do IDH à medida que ocorre um aumento de renda,

Utilizando-se os valores do número de empregos com a taxa de analfabetismo na região, temos :



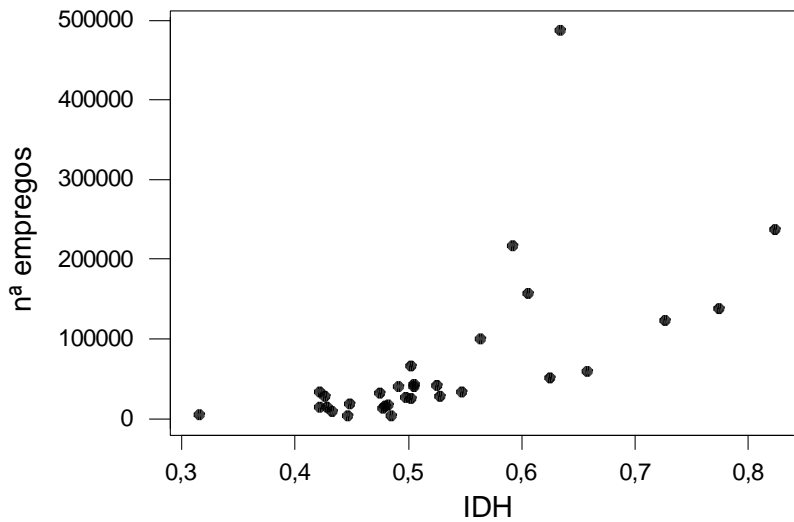
Verifica-se que há uma leve tendência na diminuição no número de empregos à medida que cresce a taxa de analfabetismo da região.

Utilizando-se os valores do número de empregos com a população favelada, obtemos o diagrama:



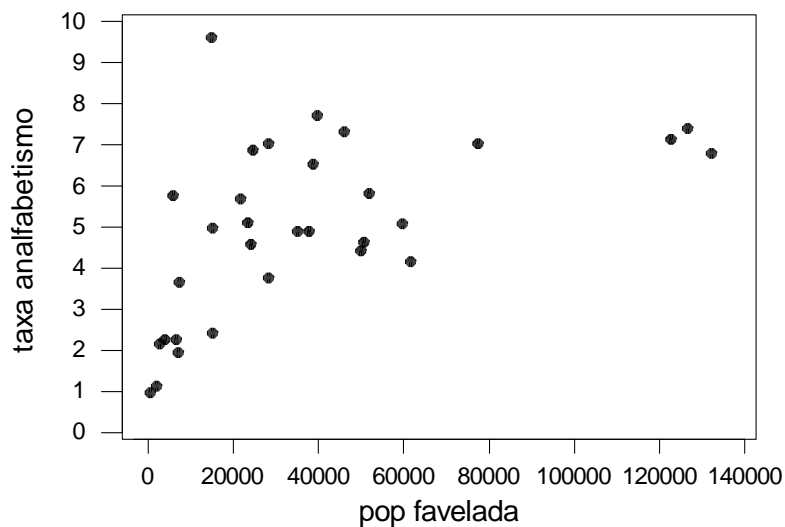
Não verificamos tendência nesta análise, somente uma concentração na quantidade do número de empregos (até 100.000), para diferentes faixas de população vivendo em favela.

Utilizando-se os valores do número de empregos com o IDH, obtemos o seguinte diagrama:



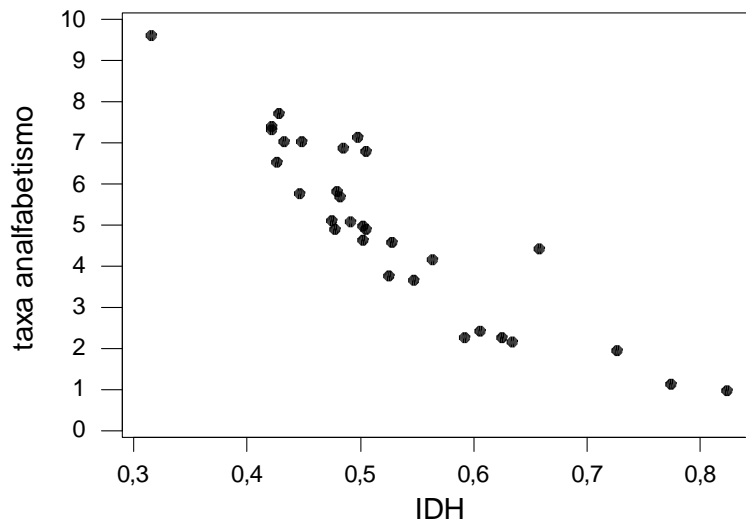
Observa-se que há pouca tendência no aumento de empregos à medida que o IDH aumenta.

Utilizando-se os valores do número de taxa de analfabetismo com a população favelada, temos :



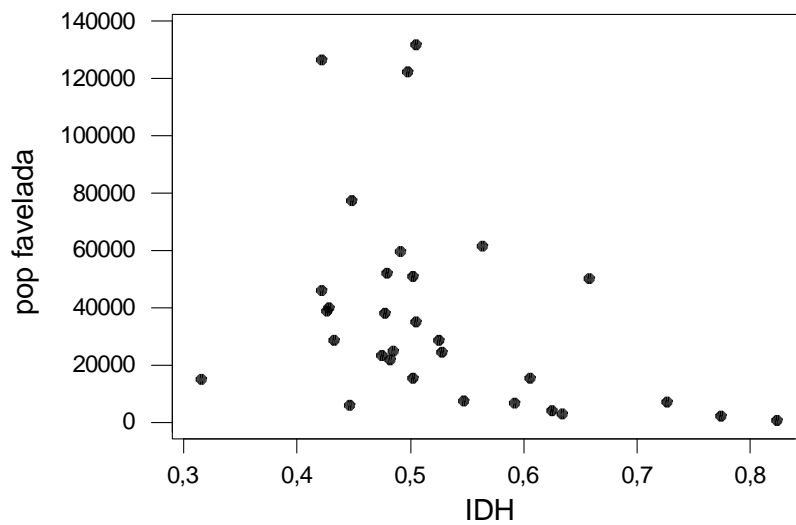
Não podemos afirmar que há tendência, porque a dispersão é grande, mas observa-se que para os extremos das variáveis existe uma certa correlação.

Utilizando-se os valores do número de taxa de analfabetismo com o IDH, obtemos o diagrama :



que demonstra uma tendência no aumento do IDH à medida que diminui a taxa de analfabetismo.

Utilizando-se os valores do número de população favelada com o IDH, obtemos o diagrama :



Observamos uma relação muito dispersa, talvez uma tendência no aumento significativo do IDH somente quando temos uma baixa população favelada.

CORRELAÇÃO

Para medir a relação entre duas variáveis quantitativas podemos usar o Coeficiente de Correlação Linear de Pearson, que quantifica quão linear é a relação. Para as relações estabelecidas no item anterior temos:

Correlations: renda; nª empregos; taxa analfabetismo; pop favelada; IDH				
	renda	nª empre	taxa ana	pop fave
nª empre	0,644 0,000			
taxa ana	-0,840 0,000	-0,651 0,000		
pop fave	-0,420 0,019	-0,339 0,062	0,537 0,002	
IDH	0,950 0,000	0,619 0,000	-0,903 0,000	-0,371 0,040
Cell Contents: Pearson correlation P-Value				

Verificamos que as observações feitas nos gráficos de dispersão se confirmam através da correlação de Pearson.

Quanto mais próximo de 1 (positivo ou negativo) melhor é a relação linear entre as variáveis, portanto para os valores que estamos estudando **a melhor correlação linear é do IDH com a renda.**

REGRESSÃO

A análise de regressão permite ajustar a relação entre duas variáveis quantitativas. Em função das análises feitas acima vamos fazer a regressão para as variáveis (IDH x renda):

Encontramos a seguinte equação:

Regression Analysis: IDH versus renda				
The regression equation is IDH = 0,389 +0,000108 renda				
Predictor	Coef	SE Coef	T	P
Constant	0,38890	0,01054	36,90	0,000
renda	0,00010755	0,00000657	16,36	0,000

S = 0,03502 R-Sq = 90,2% R-Sq(adj) = 89,9%

Analysis of Variance

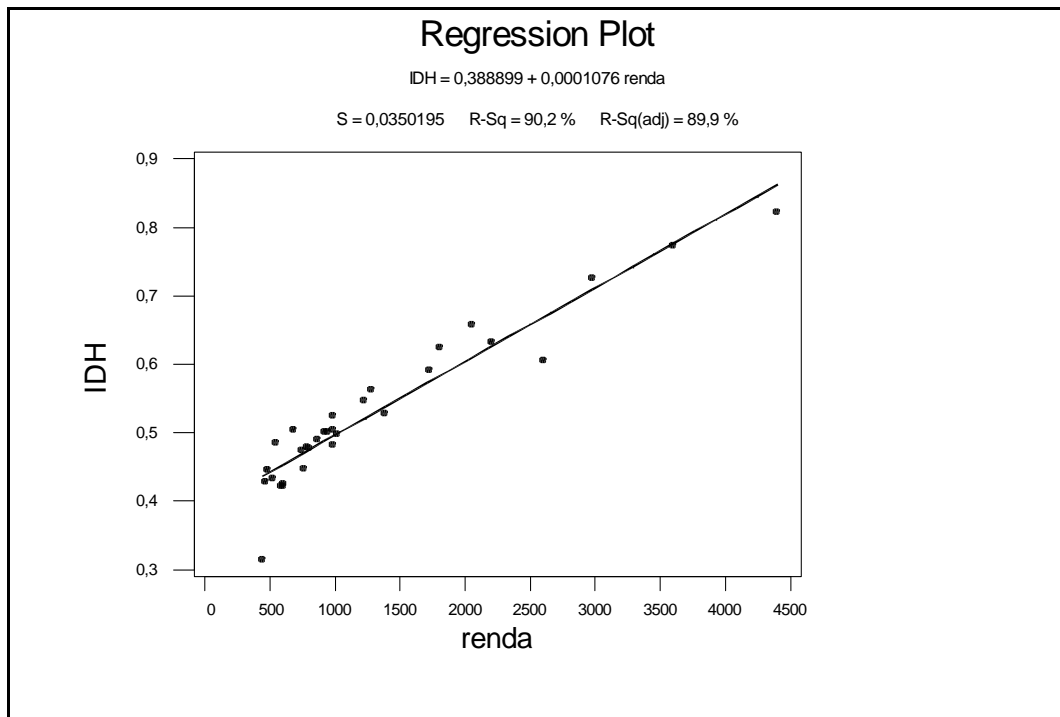
Source	DF	SS	MS	F	P
Regression	1	0,32839	0,32839	267,78	0,000
Residual Error	29	0,03556	0,00123		
Total	30	0,36396			

Unusual Observations

Obs	renda	IDH	Fit	SE Fit	Residual	St Resid
18	440	0,31500	0,43622	0,00840	-0,12122	-3,57R
21	4400	0,82400	0,86212	0,02141	-0,03812	-1,38 X
30	3600	0,77400	0,77608	0,01646	-0,00208	-0,07 X

R denotes an observation with a large standardized residual

X denotes an observation whose X value gives it large influence.

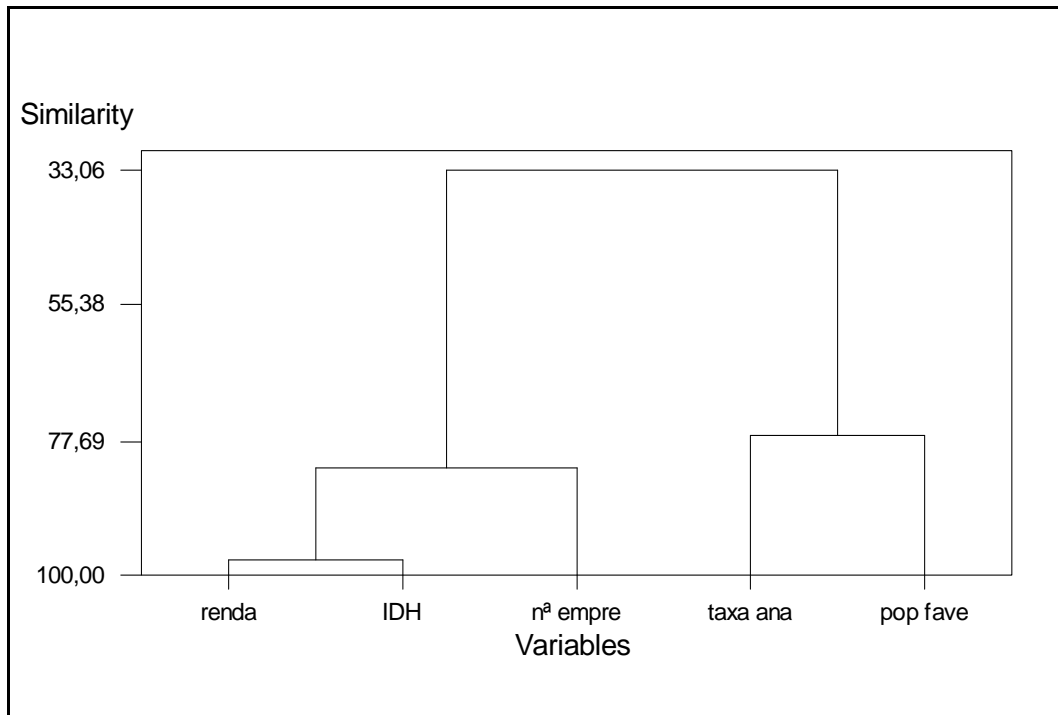


A reta de regressão de mínimos quadrados encontrada é: $IDH = 0,389 + 0,000108 \cdot \text{renda}$. Esta equação faz com a soma dos quadrados das distâncias dos pontos à reta seja a menor possível. O coeficiente angular da reta ($b=0,000108$) é a taxa média de variação do índice de desenvolvimento humano (HDI), à medida que o padrão de vida aumenta.

Verificamos que a relação linear entre o IDH e a renda explica 90,2% (r square) da variação do IDH.

ANÁLISE MULTIVARIADA

Neste tipo de análise vamos verificar o agrupamento das variáveis escolhidas, reunindo-as em clusters por similaridade:



Observamos que existe um agrupamento com maior similaridade entre o IDH e a renda, sendo que as outras variáveis tem uma correlação menor com anteriores.

CONCLUSÕES SOBRE AS ANÁLISES ESTATÍSTICAS EFETUADAS

De acordo com a análise feita, concluímos que:

- . o aumento do número de empregos não necessariamente significa um aumento de renda;
- . existe uma tendência na diminuição da renda à medida que aumenta a taxa de analfabetismo;
- . nos extremos da renda temos os extremos de quantidade de pessoas que vivem em favela;
- . há uma concentração de renda na faixa de R\$ 500,00 a R\$1500,00, para uma população favelada aproximada entre 5.000 e 60.000 pessoas
- . há uma tendência no aumento do IDH à medida que ocorre um aumento de renda;
- . há uma leve tendência na diminuição no número de empregos à medida que cresce a taxa de analfabetismo da região;
- . há pouca tendência no aumento de empregos à medida que o IDH da região aumenta;
- . não podemos afirmar que há tendência, porque a dispersão é grande, mas nos extremos das variáveis existe uma certa correlação entre analfabetismo e pop favelada;
- . há tendência de aumento do IDH à medida que diminui a taxa de analfabetismo;
- . há uma relação muito dispersa, talvez uma tendência no aumento significativo do IDH somente quando temos uma baixa população favelada.

As observações feitas nos gráficos de dispersão se confirmam através da correlação de Pearson. Observamos que existe um agrupamento com maior similaridade entre o IDH e a renda, sendo que as outras variáveis tem uma correlação menor com anteriores.

A análise dos outliers mostra bem o panorama de desigualdade do município:

Pinheiros, maior índice de renda (4.400 x 1.147 de média aparada) , tem população abaixo da média (272.574 x 338.011 trmean), alto nº de empregos (237.758 x 50.795 trmean), mínima taxa de analfabetismo (0,94 x 4,98 trmean), mínima população favelada (552 x 33.328 trmean) e o IDH mais alto do município (0,82 x 0,51 trmean).

No outro extremo temos M´Boi Mirim acima de Q3 em população (484.966 x 338.011 trmean), abaixo de Q1 em renda (580 x 1.147 trmean), nº de empregos na mediana , taxa de analfabetismo acima de Q3 (7,42 x 4,98 trmean); outlier em população favelada (126.559 x 33.328 trmean) e IDH abaixo de Q1 (0,42 x 0,51 trmean).

Resumindo o exemplo:

Pinheiros: população ↓ renda ↑ emprego ↑ analfabetismo ↓ favela ↓ IDH ↑
M´Boi Mirim : população ↑ renda ↓ emprego ↔ analfabetismo ↑ favela ↑ IDH ↓

Se tomarmos dados mundiais de IDH, Pinheiros se assemelha em desenvolvimento humano ao 51º país do mundo nesse índice, acima da posição do Brasil que é 63º, e M´BoiMirim se assemelha ao 163º país, num ranking de 177 países.

Considerando que o Brasil é o segundo maior índice GINI do continente americano, ou seja, campeão de desigualdade de distribuição (ao lado da Guatemala) podemos dizer que estes números confirmam e a cidade de São Paulo espelha a vergonhosa situação nacional que, no dizer de Bernardo Kliksberg, é como se houvessem dois Brasis: suas chances de ter um mínimo de qualidade de vida dependem se você nasce no Brasil pobre ou no Brasil rico.

REFERÊNCIAS

Botter, Denise et al, Noções de Estatística com apoio computacional, Instituto de Matemática e Estatística – USP, fevereiro de 1996.

Human Development Report 2005 – International cooperation at a crossroads: Aid, trade and security in an unequal world publicado pelo UNDP – United Nations Development Programme

Minitab Statistical Software Release 13.20

Sumário de Dados do Município de São Paulo 2004 – PMSP

http://www2.prefeitura.sp.gov.br/secretarias/governo/sumario_dados/0002