

FLG0243 - Cartografia Temática

Mapas e Estatística

Professores

Ligia Vizeu Barrozo

Reinaldo Paul Pérez Machado

Formas de abordagem sobre o universo

Abordagens	Características básicas			
Ciência	Provisório	Explicativo	Lógico	Empírico
Filosofia	Provisório	Explicativo	Lógico	
Religião	Definitivo	Explicativo	Lógico	
Arte	Estético			

Base empírica

- ▶ Ao contrário do conceito popular de empírico, o qual se aproxima mais de algo feito de forma casual, por *ensaio e erro*, na Filosofia da Ciência, significa as impressões que chegam ao nosso sentido, aquilo que é sentido. Refere-se a fatos objetivos presentes no mundo real.

Base empírica

- ▶ A Ciência explica os mistérios naturais por meio de um discurso *lógico* pautado em evidências empíricas (concretas, observáveis).
- ▶ Embora um resultado possa ser concreto, duradouro, ou mesmo eterno, as conclusões científicas serão sempre *provisórias*.

CIÊNCIA

avanço do conhecimento

- ✓ importância da pesquisa bibliográfica

CIÊNCIA

- ▶ **Conjunto organizado de conhecimentos relativos a um determinado objeto, especialmente os obtidos mediante a observação, a experiência dos fatos e um método próprio**

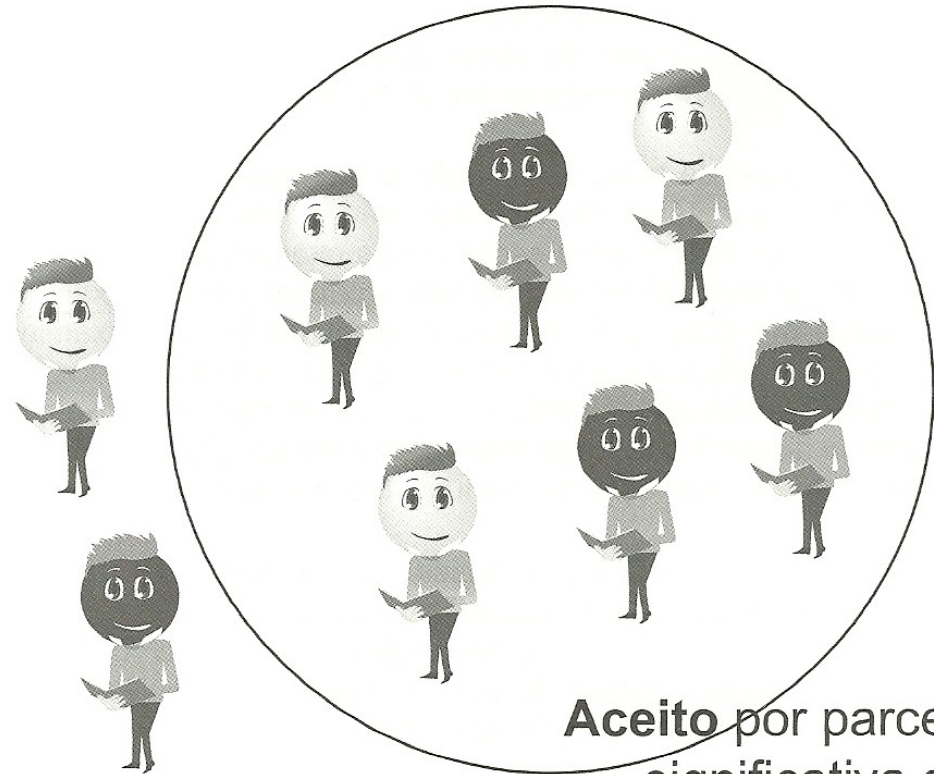
CIÊNCIA

método de estudo

- ▶ A possibilidade de uma ciência existir é determinada não pelo *o que* mas pelo *como*, não pelo assunto mas pelo *método*. Todos os assuntos podem ser objeto de investigação

(Cole & King, 1968)

O que é conhecimento científico?



Aceito por parcela
significativa da
comunidade
científica

**Requisitos Metodológicos
(viés científico)**

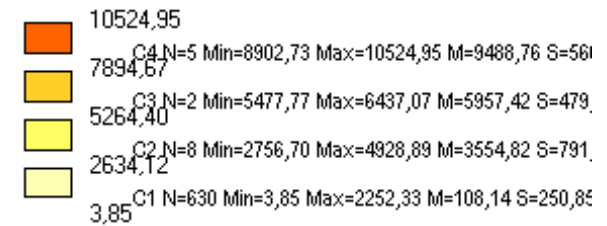
Conhecimento científico

- ▶ O conhecimento científico, num dado momento e área, será sempre aquele aceito nesse momento, pelos cientistas dessa área. Isso não significa que haja sempre concordância.
- ▶ O desafio do cientista não é apenas o de conseguir conclusões originais e interessantes... Ele deve convencer sua comunidade científica sobre a validade dessas conclusões. E, para isso, não bastam os dados. O emprego de técnicas e procedimentos aceitos é parte importante para se conseguir tal aceitação. Além disso, deve usar métodos de análise de dados válidos para a comunidade científica.

Quatro classes de mesmo intervalo (Valor máximo – valor mínimo)/4

Cartografia

Densidade Demográfica 1996 hab/km²

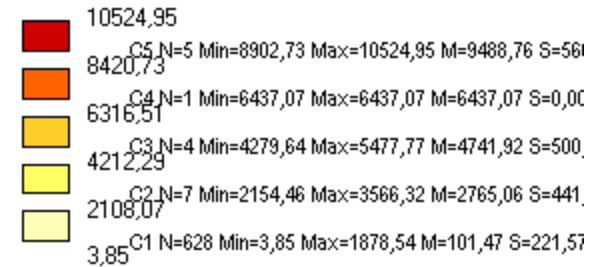


Densidade demográfica no Estado de São Paulo, 1996 (5 classes de mesmo intervalo)

Cartografia



Densidade Demográfica 1996 hab/km2



Densidade demográfica no Estado de São Paulo, 1996 (6 classes de mesmo intervalo)

Cartografia

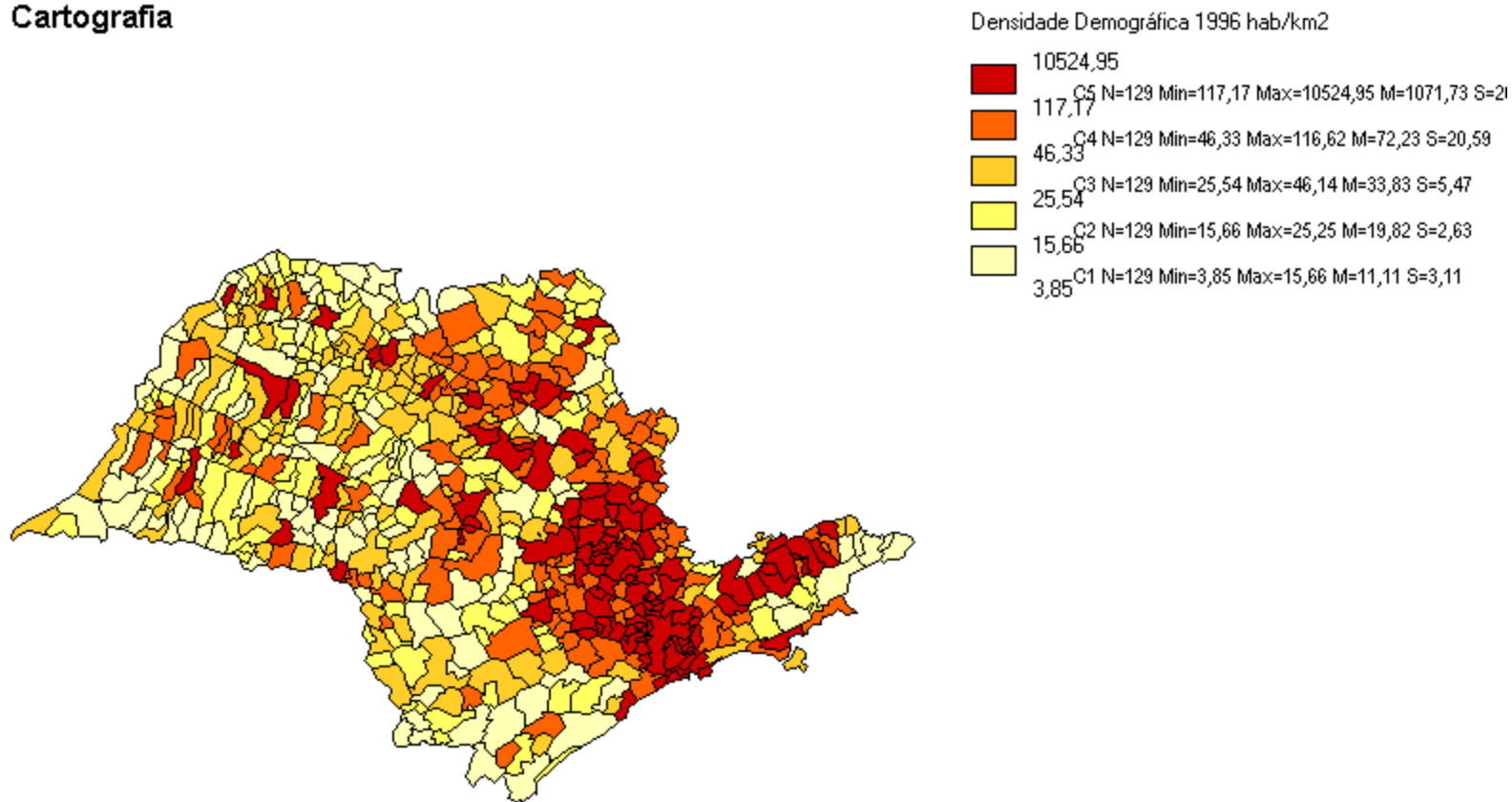


Densidade Demográfica 1996 hab/km2



Densidade demográfica no Estado de São Paulo, 1996 (5 classes definidas pelo número de municípios)

Cartografia



Porcentagem de moradias expostas a alto nível de ruído na Holanda

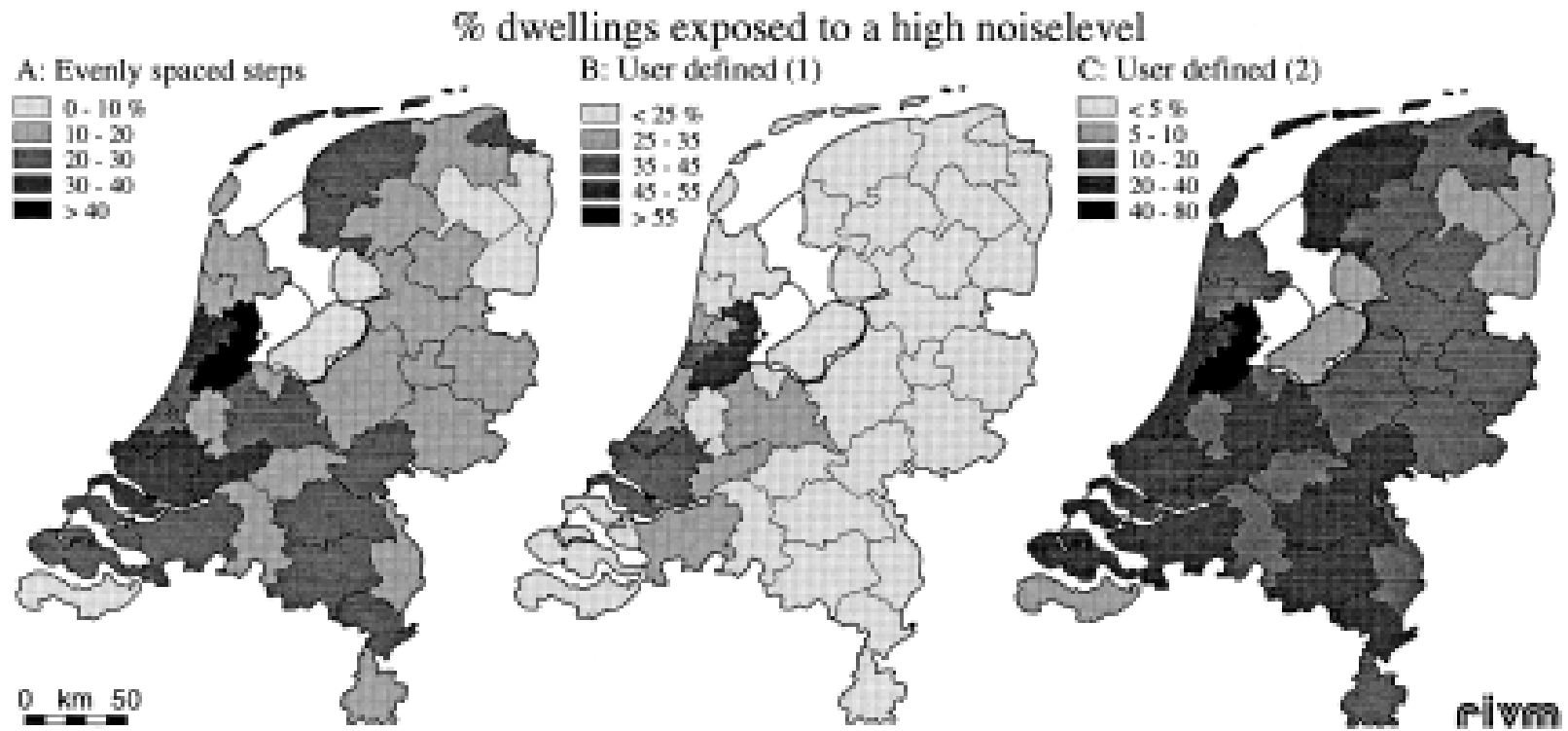
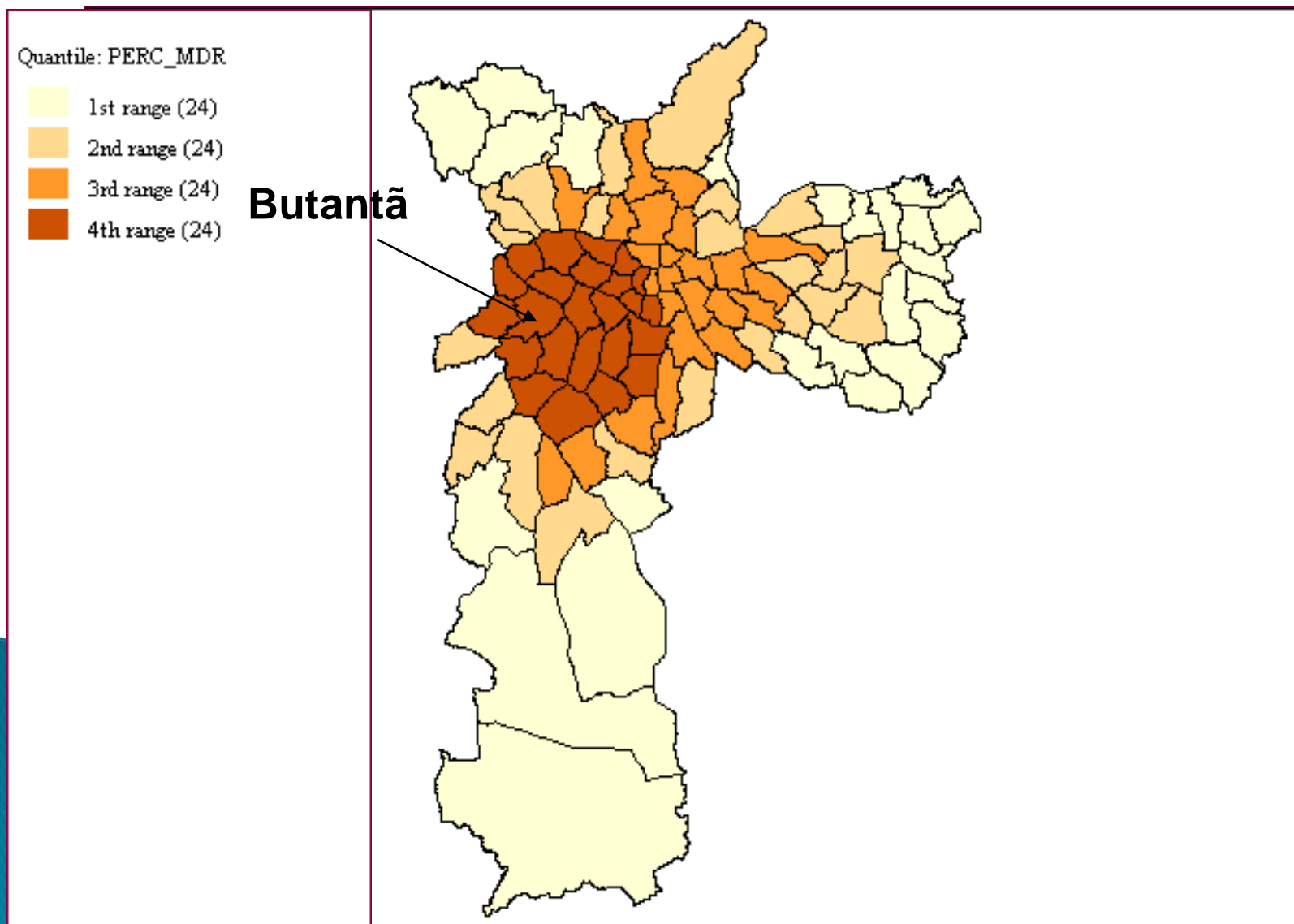
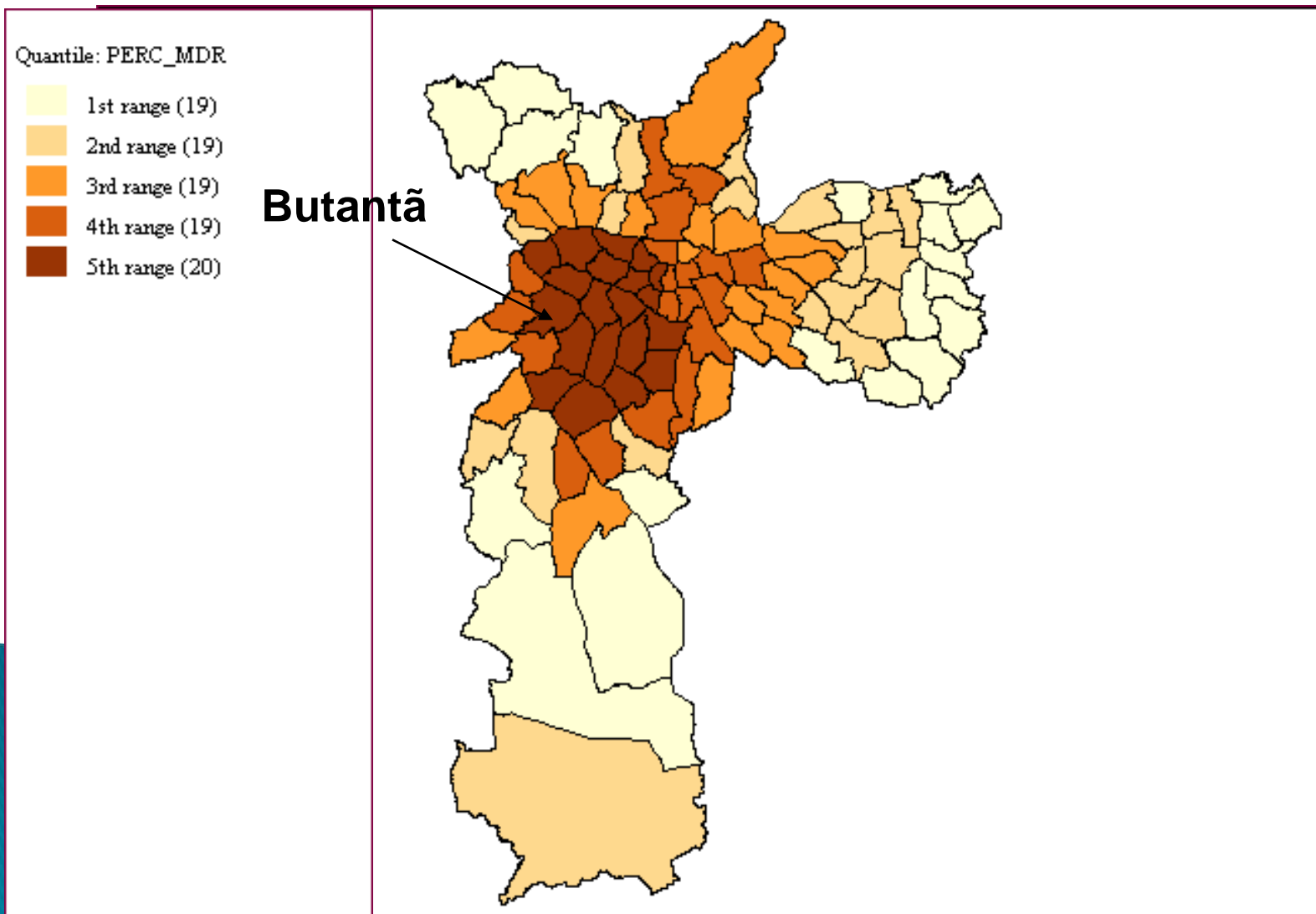


Fig. 3. Comparison of maps using different ranging methods.

Distribuição espacial do percentual de pós-graduados no município de SP

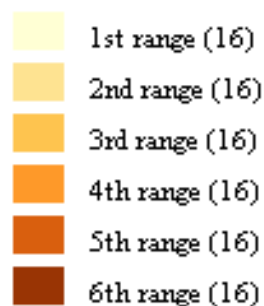


Distribuição espacial do percentual de pós-graduados no município de SP

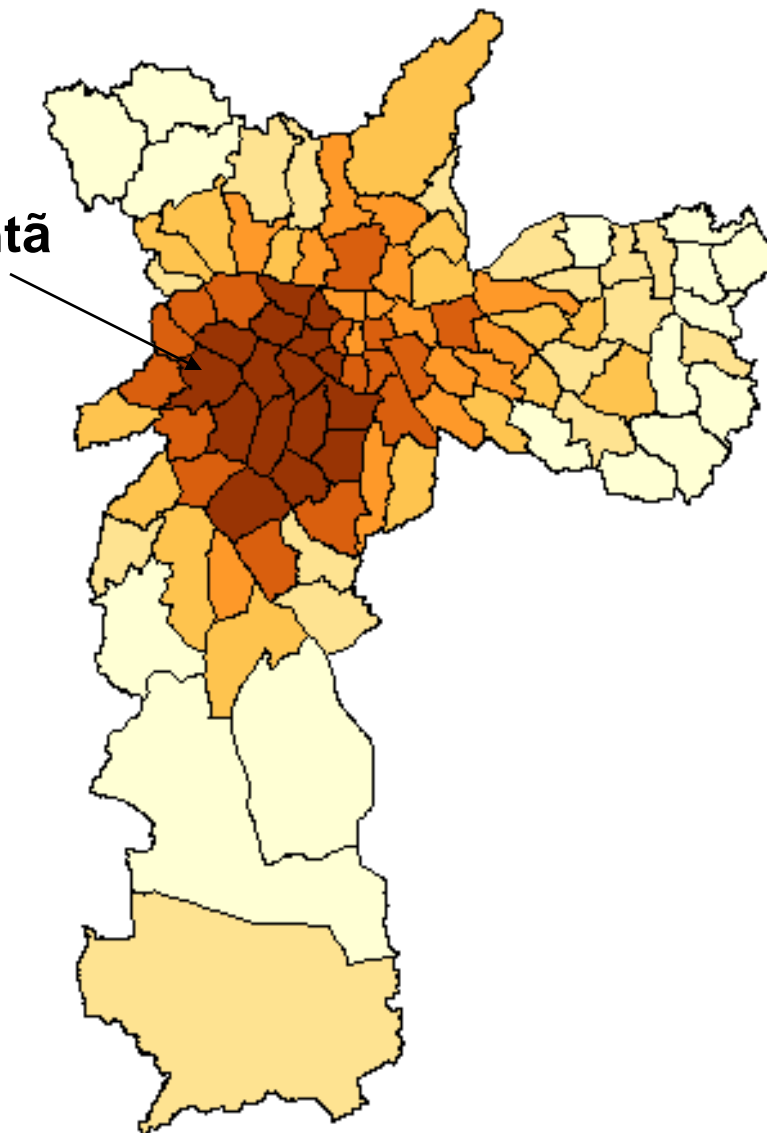


Distribuição espacial do percentual de pós-graduados no município de SP

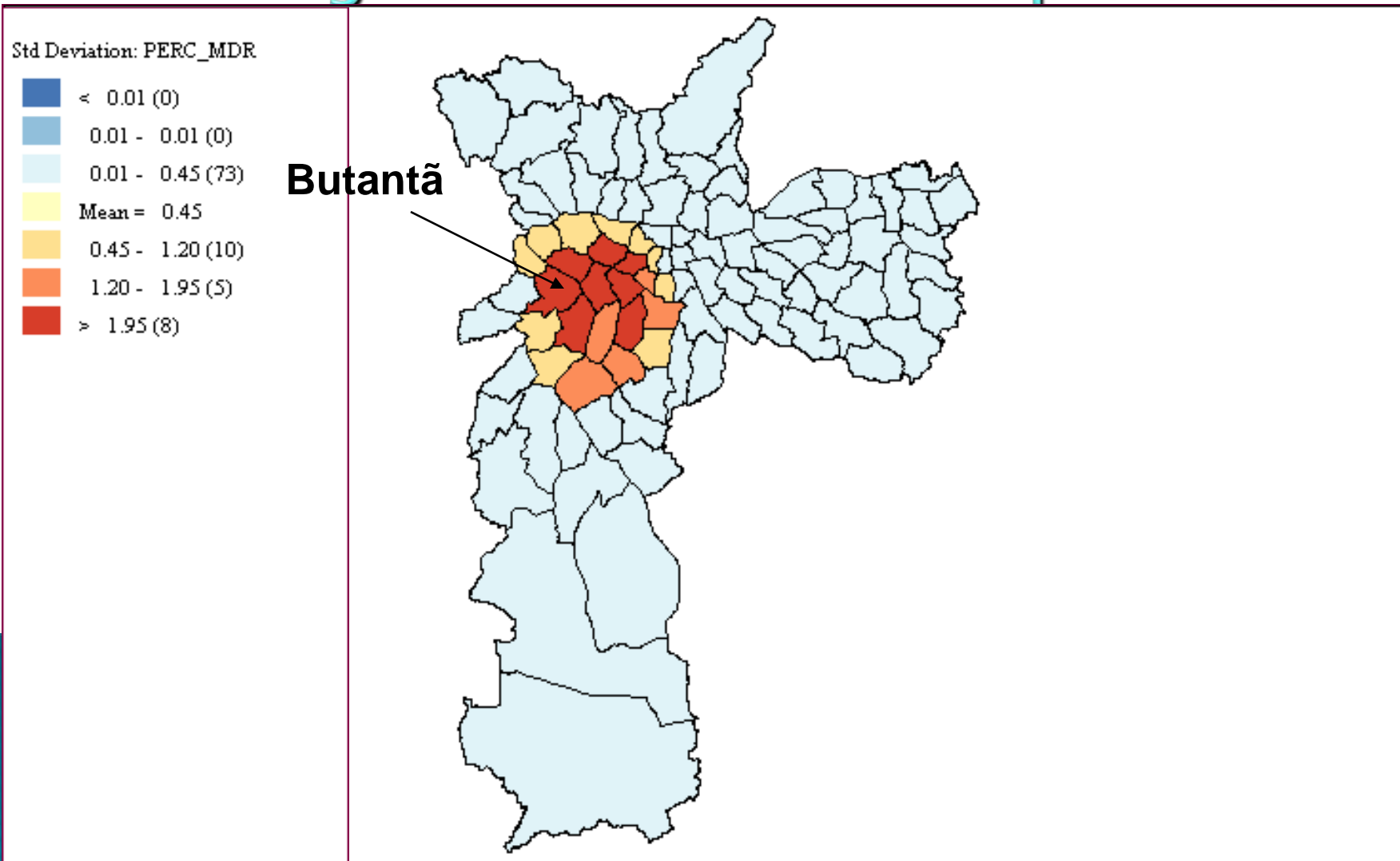
Quantile: PERC_MDR



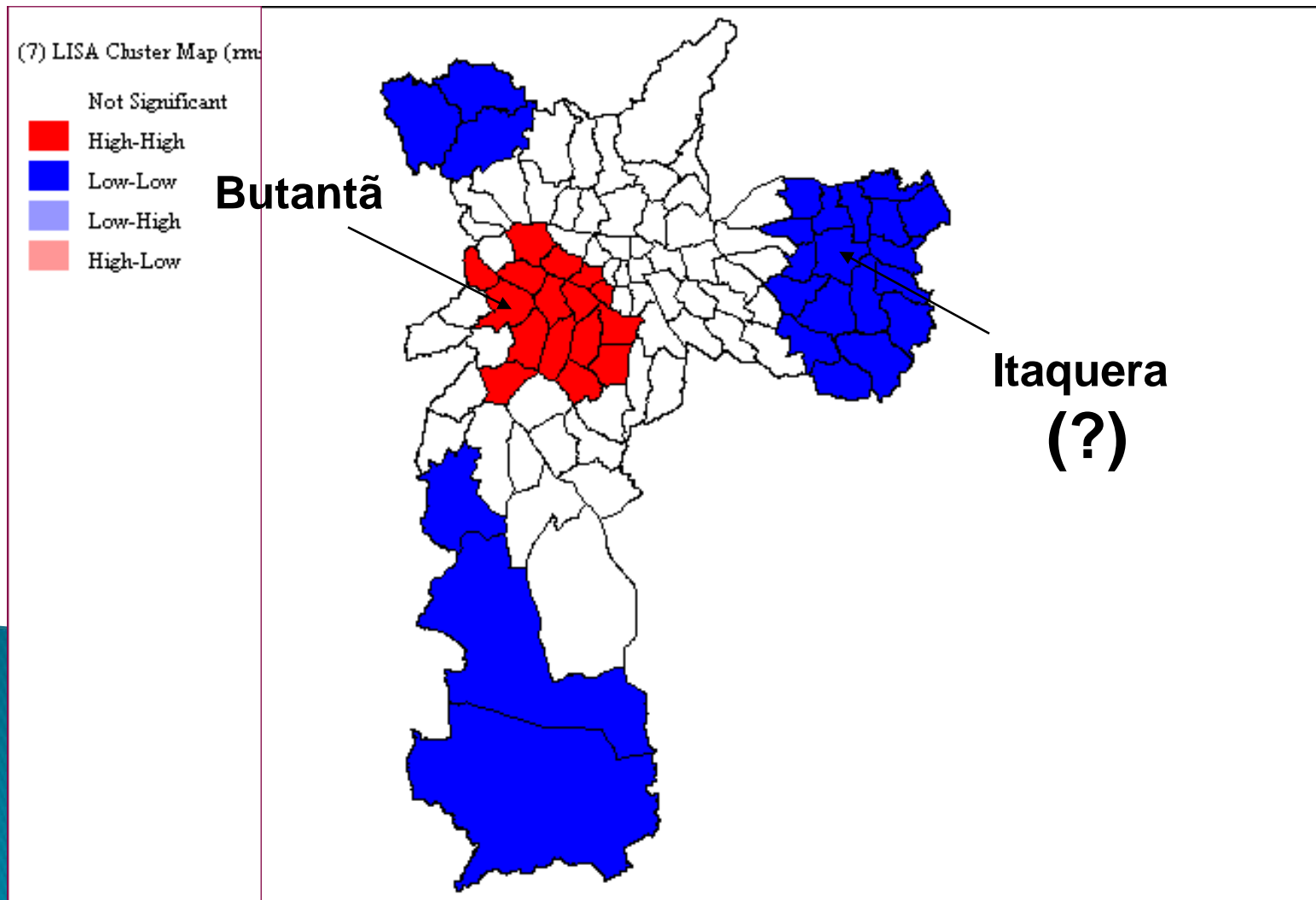
Butantã



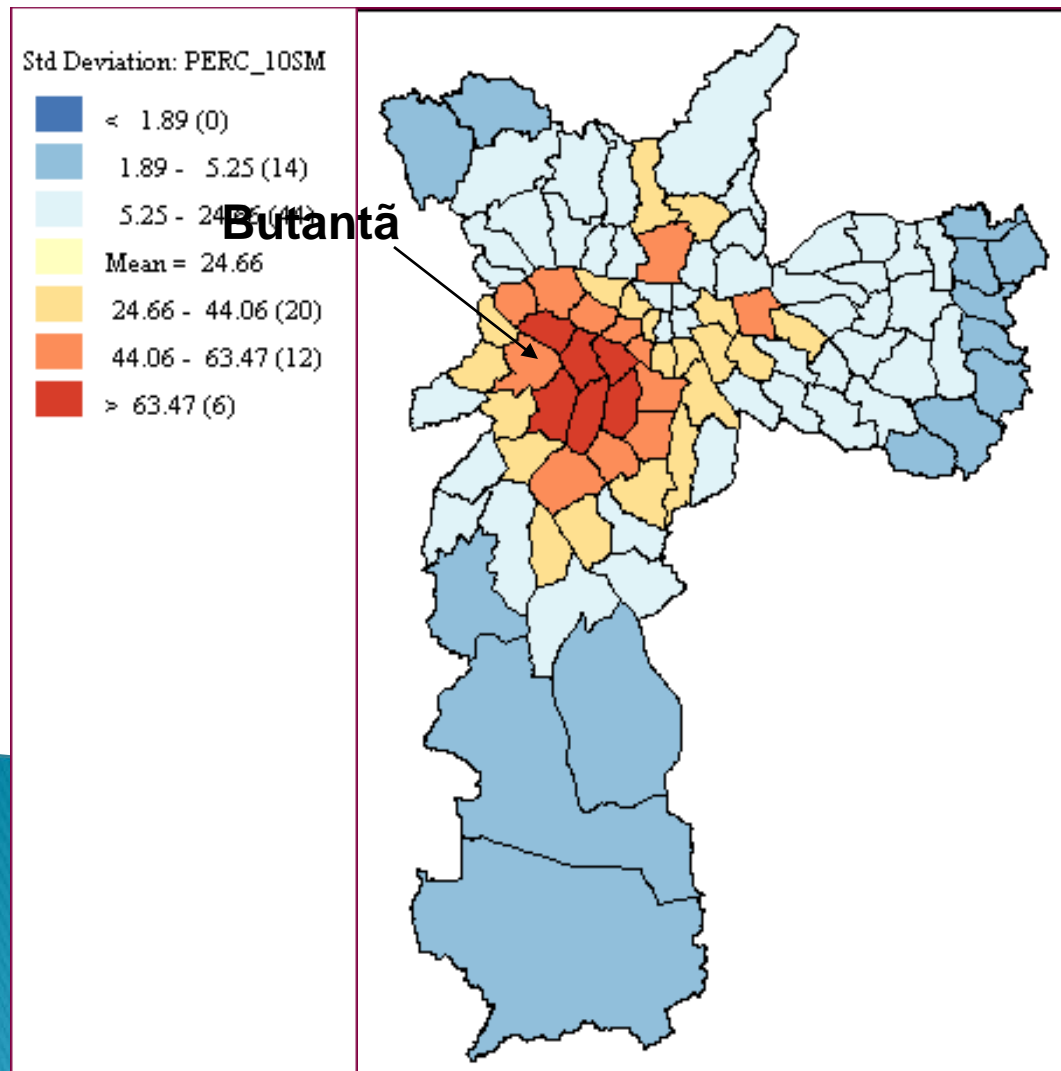
Desvio padrão do percentual de pós-graduados no município de SP



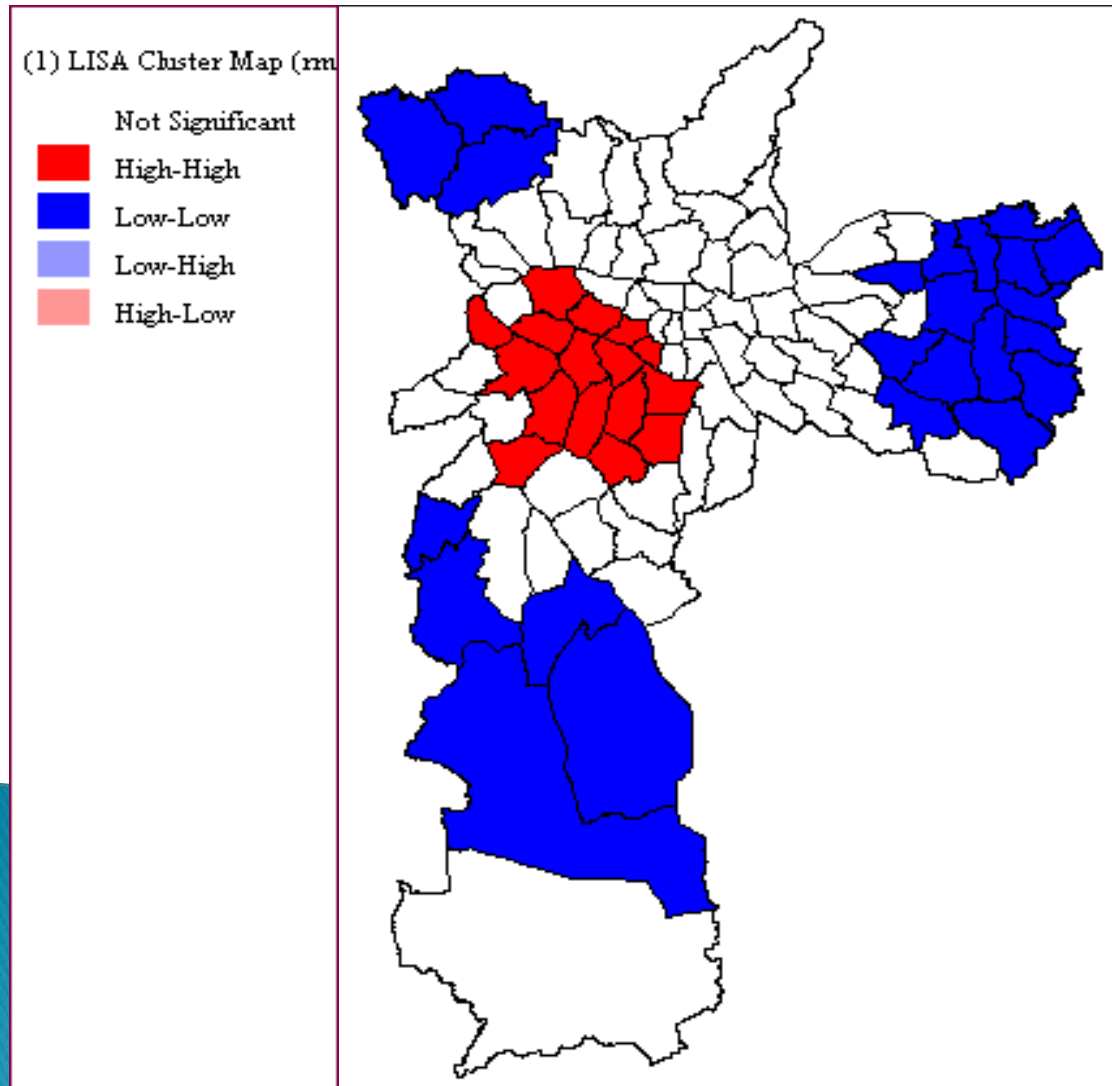
Agrupamentos (*clusters*) do percentual pós-graduados no município de SP



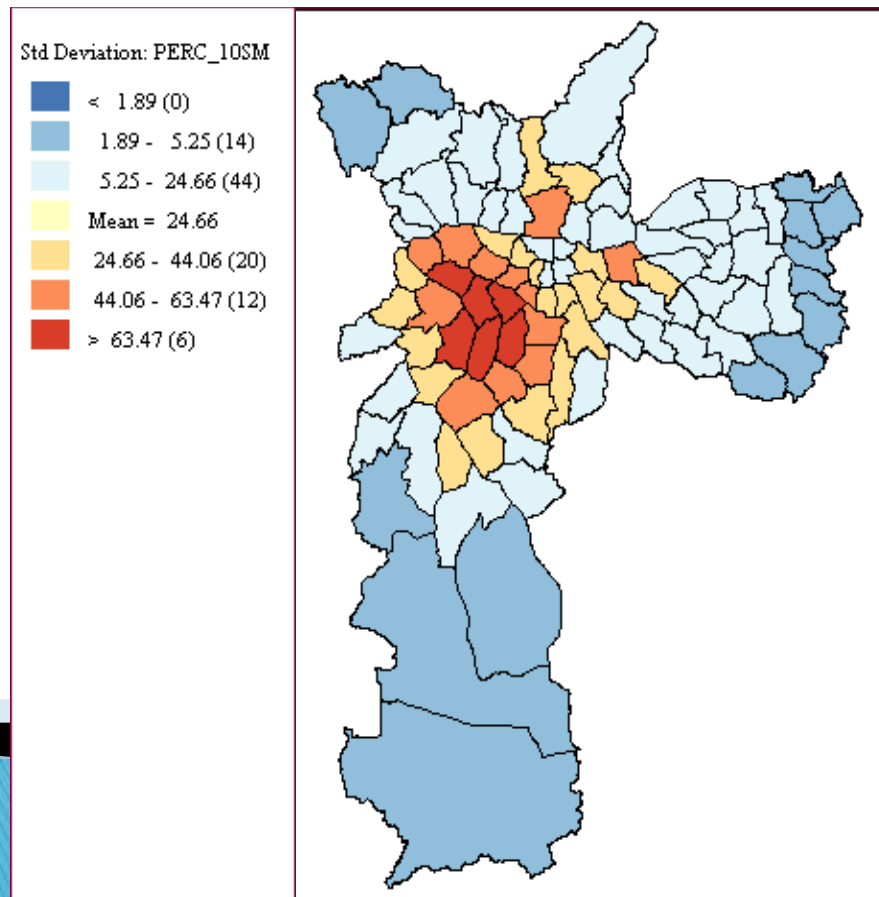
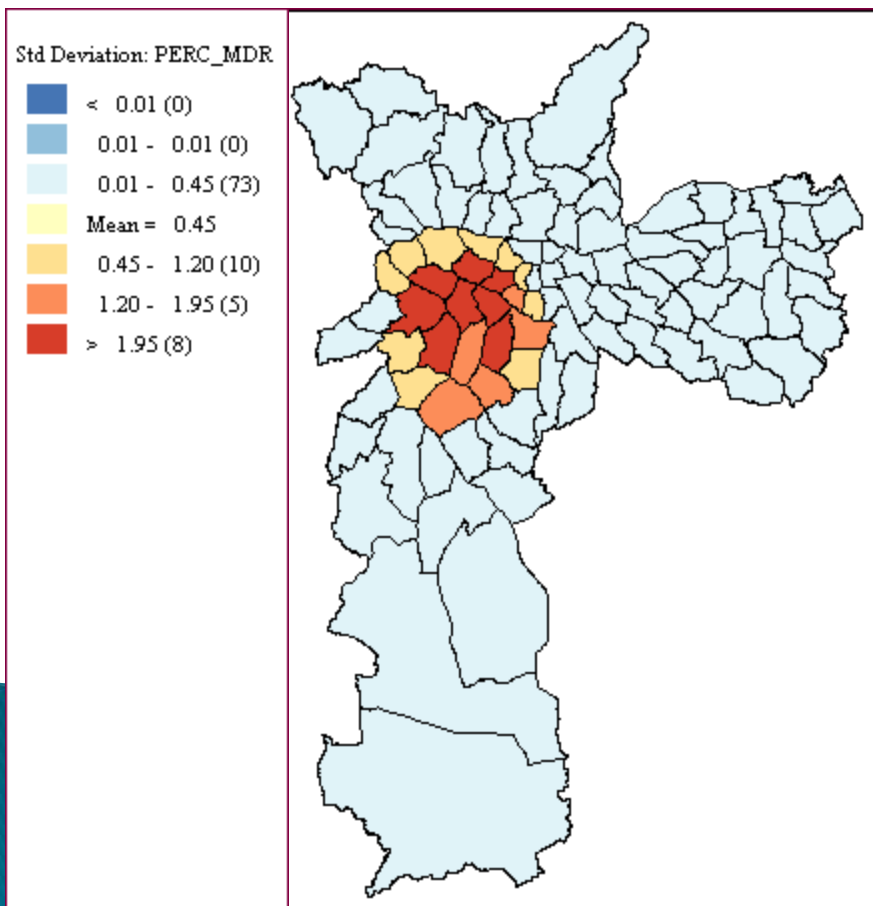
Desvio padrão % de responsáveis com rendimento nominal mensal acima de 10 SM



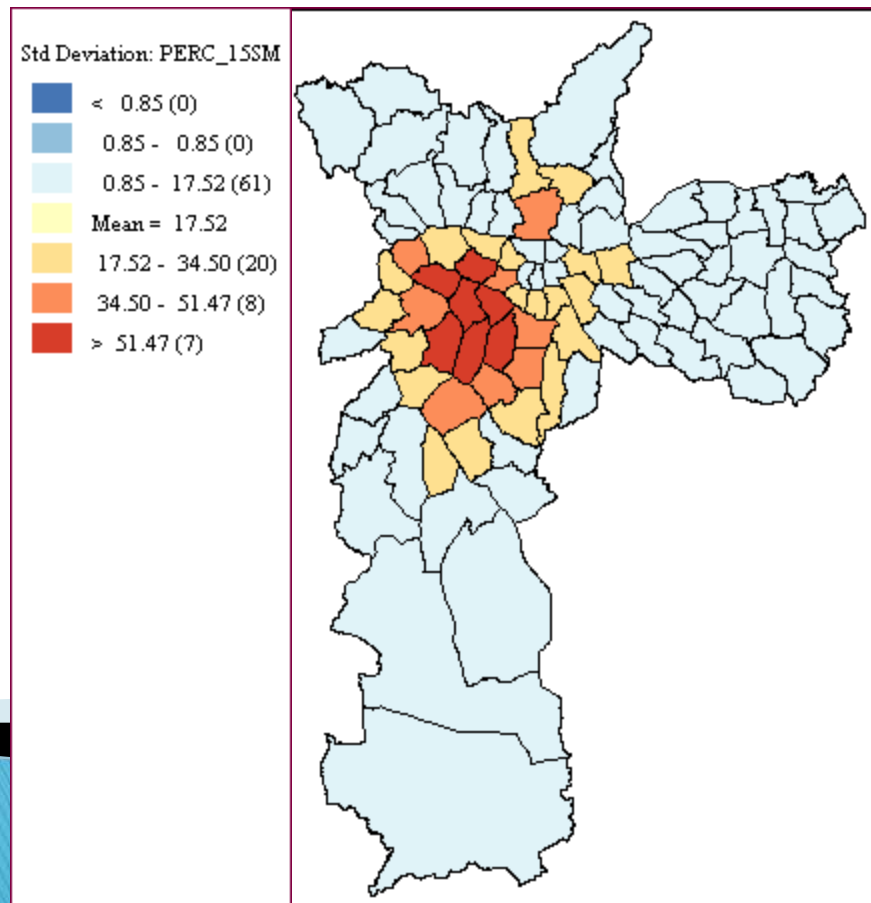
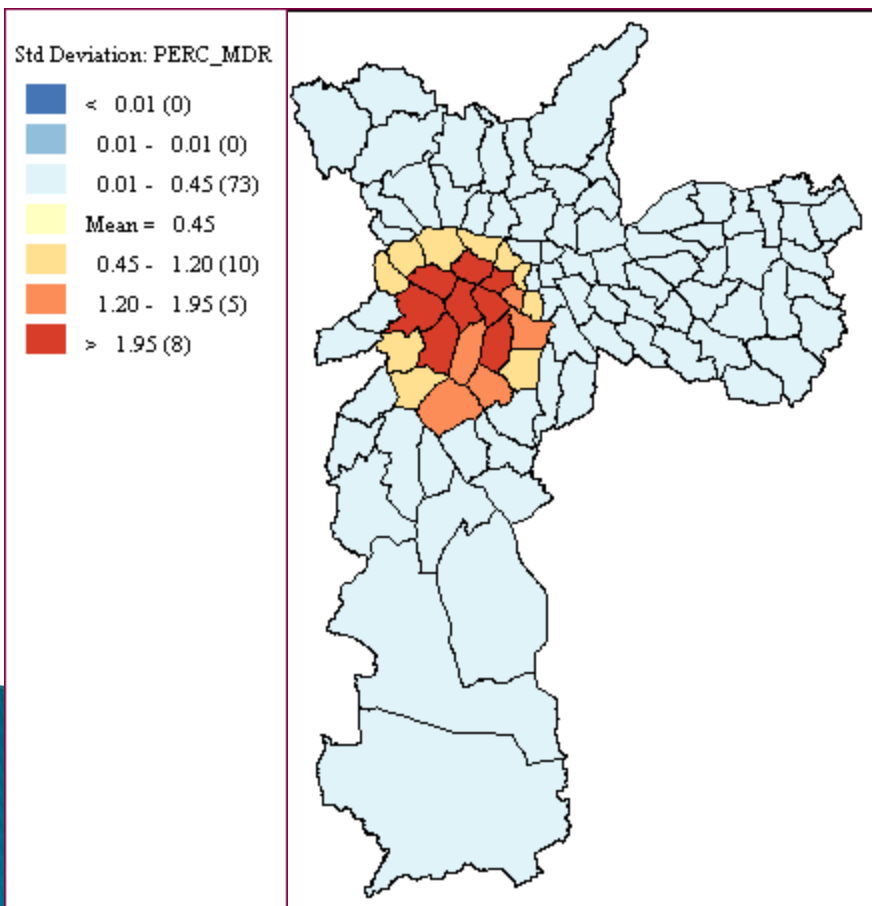
Agrupamentos de % de resp. com renda acima de 10 SM



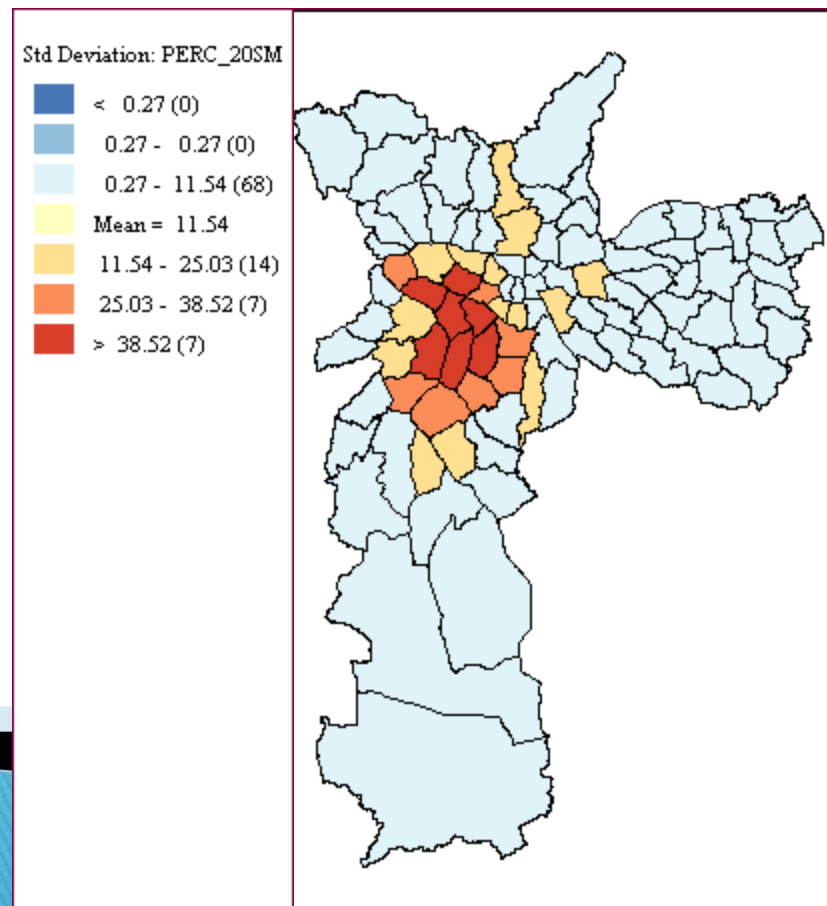
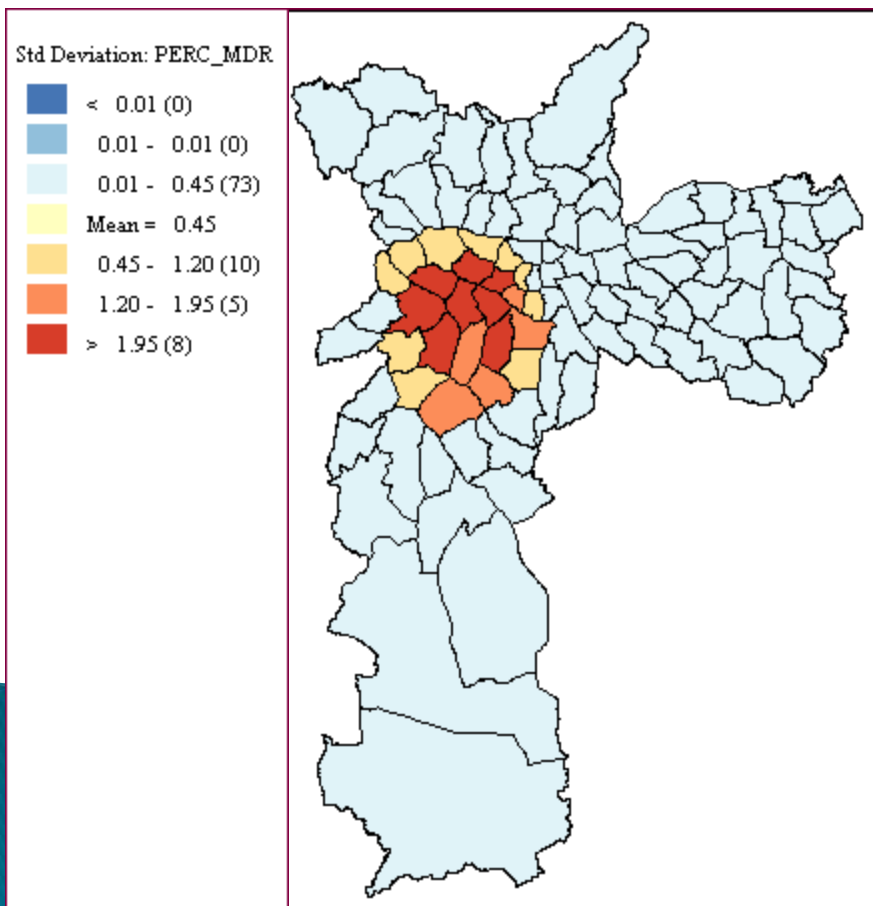
Desvio padrão de % de pós-graduados x % resp. com renda acima de 10 SM



Desvio padrão de % de pós-graduados x % resp. com renda acima de 15 SM



Desvio padrão de % de pós-graduados x % resp. com renda acima de 20 SM



Qual faixa de renda melhor explica a distribuição espacial do % de pós-graduados em SP?

- Visualmente podemos verificar que existe um padrão de distribuição espacial para a variável de interesse, mas não conseguimos afirmar com certeza qual faixa de renda melhor a explica
- **Uma análise estatística (espacial) pode MEDIR como essas variáveis se associam (correlação; regressão)**

Qual faixa de renda melhor explica a distribuição espacial do % de pós-graduados em SP?

➤ Resultados de análise de regressão (espacial):

10SM: $r^2 = 0,889$ $p < 0,0001$;

15SM: $r^2 = 0,901$ $p < 0,0001$;

20SM: $r^2 = 0,906$ $p < 0,0001$.

CIÊNCIA

avanzo do conhecimento

O estudo é necessário?

✓ importância da pesquisa bibliográfica

ANÁLISE QUANTITATIVA DOS DADOS

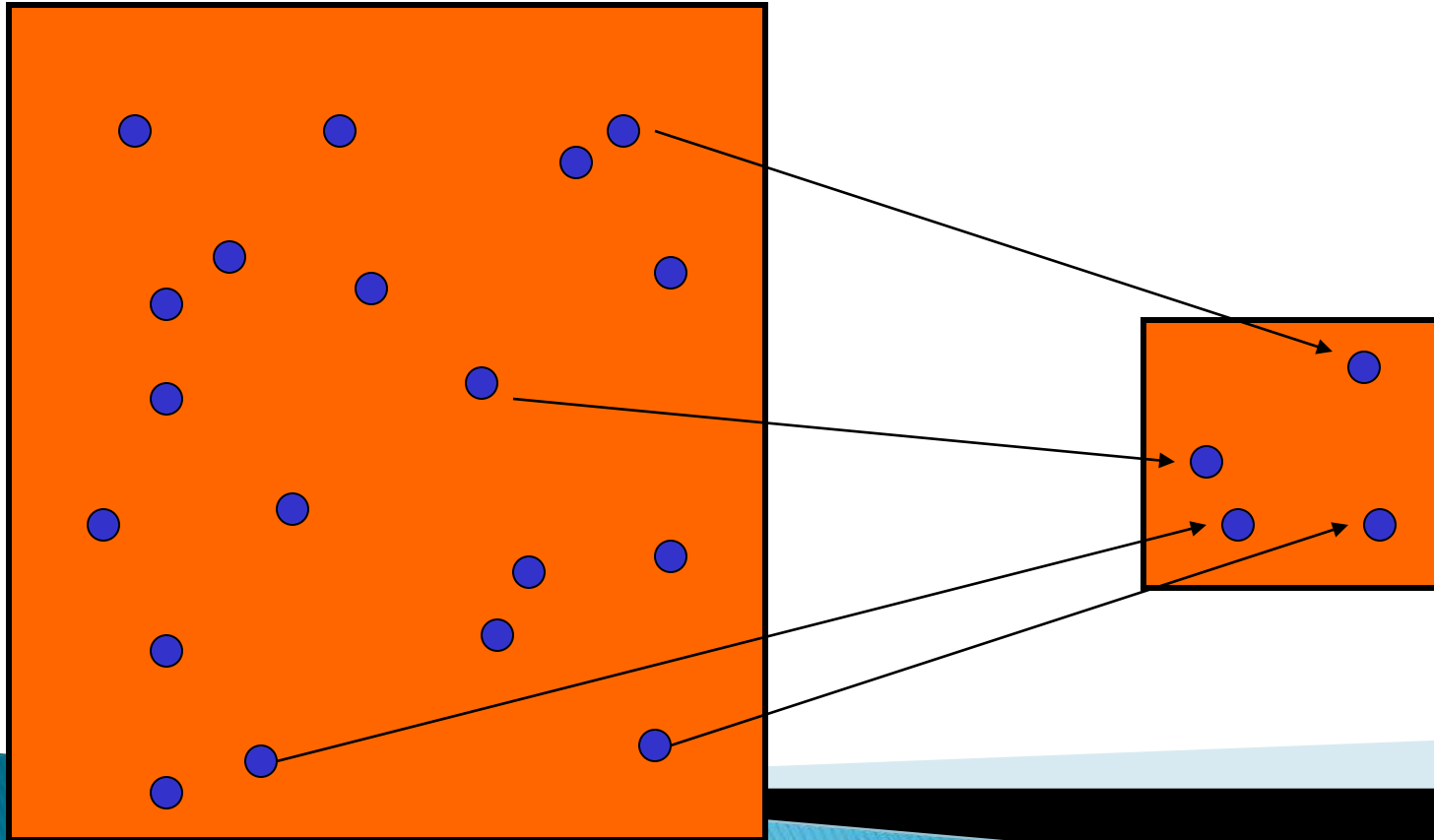
possibilita maior objetividade e precisão na análise;

evita-se generalizações com base em evidências insuficientemente analisadas;

linguagem científica interdisciplinar e universal

**População (indivíduos,
itens ou dados)**

amostra



Estatística descritiva

Estatística inferencial

População

Coleção de todos os indivíduos, itens ou dados sob consideração em um estudo estatístico

Amostra

Parte da população da qual a informação é coletada

Estadística descritiva

Compreende métodos para organizar e resumir a informação

Estadística inferencial

Compreende métodos para fazer predições ou generalizações sobre uma população com base nas características de uma amostra dessa população

Dados

Informação que descreve as características de diferentes objetos e eventos

Variável

Uma única característica de um conjunto de dados.

Ex.: **conjunto de dados**: municípios do Estado de São Paulo

Variáveis: população; renda da população; índice de desenvolvimento do município

Alguns símbolos

\bar{x} = x barra, média da amostra

μ = mi, média da população

Σ = sigma, somatória

Medidas de tendência central

Média aritmética

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

Soma de todos os valores (x) de um conjunto de dados dividida pelo número de valores (n).

Medidas de tendência central

Média

(SOMATÓRIA)

$$\sum_{i=1}^n x_i = \sum_{i=1}^n x_i = x_1 + x_2 + x_3 + x_4 + x_5$$

Medidas de tendência central - MÉDIA

Estados da Região Norte	Rendimento médio mensal da pop. ocupada
Rondônia	714,6
Acre	780,3
Amazonas	659,9
Roraima	643,5
Amapá	594,1
Pará	772,3
Tocantins	512,9

Fonte: IBGE, 2002

Média – exemplo: rendimento médio da população ocupada na Região Norte do Brasil

$$\sum_{i=1}^n x_i = \sum_{i=1}^n x_i = 714,6 + 780,3 + 659,9 + 643,5 + 594,1 + 772,3 + 512,9$$

$$\sum_{i=1}^n x_i = 4677,6$$

$$\mu = \frac{\sum_{i=1}^n x_i}{n} \quad \mu = \frac{4677,6}{7} \quad \mu = 668,23$$

Medidas de tendência central

MEDIANA (md)

É o valor que ocupa a posição central de um conjunto de dados ordenados.

Número de dados ímpar: a mediana é o valor que ocupa a posição central dos dados ordenados

Medidas de tendência central

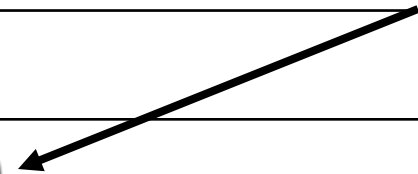
MEDIANA DE CONJUNTO ÍMPAR DE DADOS

valor que ocupa a posição central dos dados ordenados

MEDIANA DE CONJUNTO ÍMPAR DE DADOS

Estados da Região Norte	Rendimento médio mensal da pop. ocupada
Tocantins	512,9
Amapá	594,1
Roraima	643,5
Amazonas	659,9
Rondônia	714,6
Pará	772,3
Acre	780,3

MEDIANA (md)



MEDIANA DE CONJUNTO PAR DE DADOS

Estados da Região Sudeste	Rendimento médio mensal da pop. ocupada
Minas Gerais	636,70
Espírito Santo	719,70
Rio de Janeiro	895,40
São Paulo	1028,70

$$\textit{mediana} = \frac{719,70 + 895,40}{2}$$

$$\text{md} = 807,55$$

OBSERVAÇÕES SOBRE A MEDIANA

1
5
300
800
1000

290
295
300
301
302

Conjuntos bem diferentes podem ter a mesma MEDIANA

Média = 297,6

Média = 421,2

OBSERVAÇÕES SOBRE A MEDIANA

Em alguns casos a mediana descreve a tendência central dos dados melhor do que a média. É o caso de conjunto de dados com valores discrepantes.

Medidas de tendência central

MODA (mo)

É o valor que ocorre com maior frequência em um conjunto de dados.

Medidas de tendência central

MODA

7
8
5
7
7
9
8

MODA
(unimodal)

The diagram shows a vertical list of numbers: 7, 8, 5, 7, 7, 9, 8. Three arrows originate from the text 'MODA (unimodal)' and point to the three instances of the number 7 in the list, indicating that 7 is the mode.

Medidas de tendência central

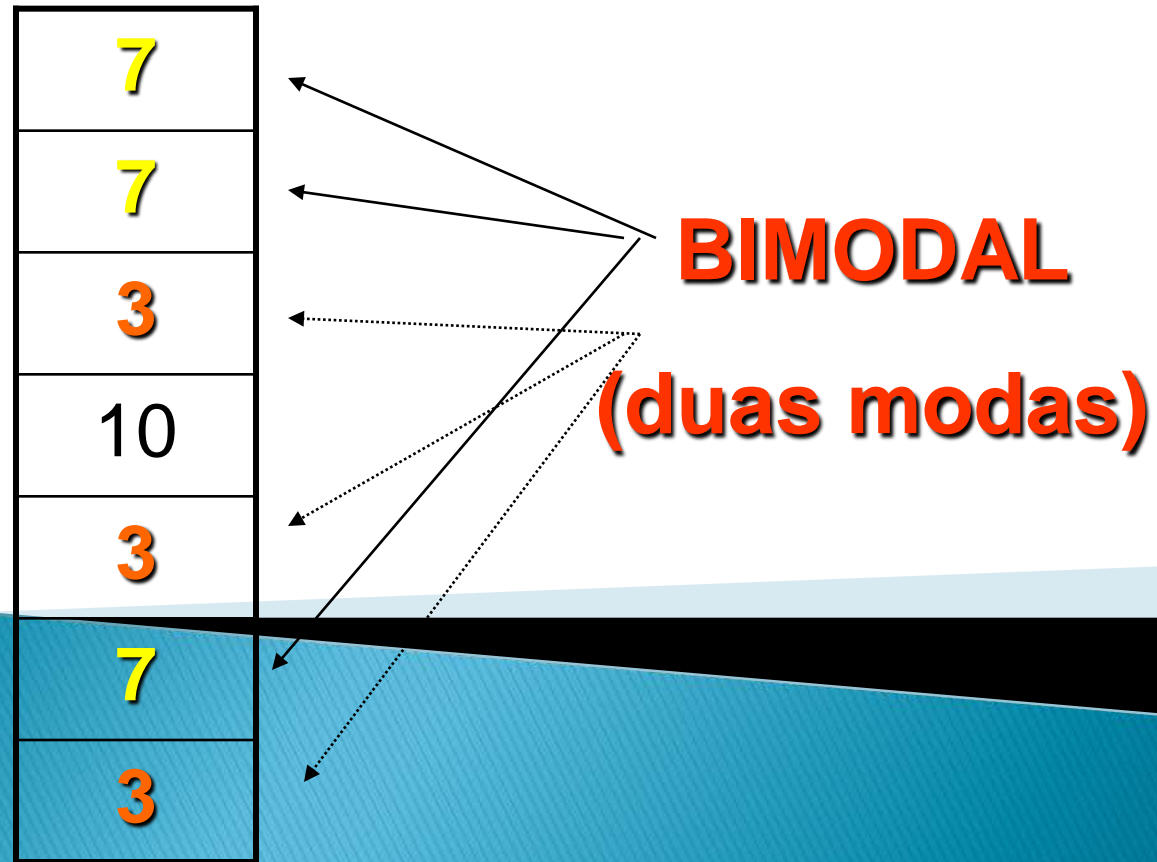
MODA

7
8
5
10
3
9
6

SEM MODA

Medidas de tendência central

MODA



OBSERVAÇÕES SOBRE A MODA

A moda mostra a tendência central dos dados identificando a área em que os dados estão mais concentrados. No entanto, na maior parte das vezes, a média e a mediana fornecem melhor descrição da tendência central dos dados

OBSERVAÇÕES SOBRE A MODA

- ✓ se o conjunto de dados é relativamente pequeno (20 ou 30 observações), você pode obter a moda, mas em geral, ela não terá qualquer sentido prático;
- ✓ a moda é bastante informativa quando o conjunto de dados é grande;
- ✓ não tem qualidades matemáticas;

Escolha da medida de tendência central

Depende:

Conjunto de dados;

Objetivo da pesquisa

HISTOGRAMA: representação gráfica que considera a frequência dos valores da série por classes de intervalos experimentais pequenos e iguais.

Tabela de dados

TABELA
BRASIL: DENSIDADE DEMOGRÁFICA SEGUNDO
AS UNIDADES DA FEDERAÇÃO — 1991

Unidades da federação	Densidade demográfica (hab./km ²)
Acre	2,71
Alagoas	88,34
Amazonas	1,34
Amapá	2,03
Bahia	20,91
Ceará	43,67
Distrito Federal	275,86
Espírito Santo	56,82
Goiás	11,80
Maranhão	14,96
Mato Grosso	2,24
Mato Grosso do Sul	4,98
Minas Gerais	26,82
Pará	4,16
Paraíba	59,32
Paraná	42,36
Pernambuco	70,50
Piauí	10,27
Rio de Janeiro	292,85
Rio Grande do Norte	45,41
Rio Grande do Sul	32,55
Rondônia	4,74
Roraima	0,96
Santa Catarina	47,61
São Paulo	127,07
Sergipe	68,24
Tocantins	3,32

Fonte: IBGE (1993).

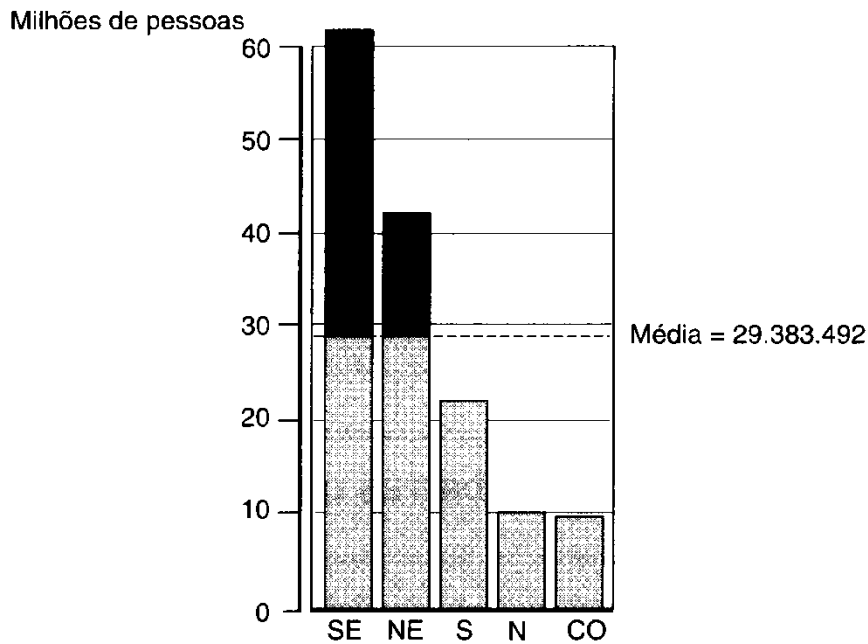
Tabela de frequência

APURAÇÃO

Classes de intervalos = 5	Frequência
0 ┆ 5	9
5 ┆ 10	—
10 ┆ 15	3
15 ┆ 20	—
20 ┆ 25	1
25 ┆ 30	1
30 ┆ 35	1
35 ┆ 40	—
40 ┆ 45	2
45 ┆ 50	2
50 ┆ 55	—
55 ┆ 60	2
60 ┆ 65	—
65 ┆ 70	1
70 ┆ 75	1
75 ┆ 80	—
80 ┆ 85	—
85 ┆ 90	1
// //	
120 ┆ 125	1
// //	
275 ┆ 280	1
280 ┆ 285	—
285 ┆ 290	—
290 ┆ 295	1

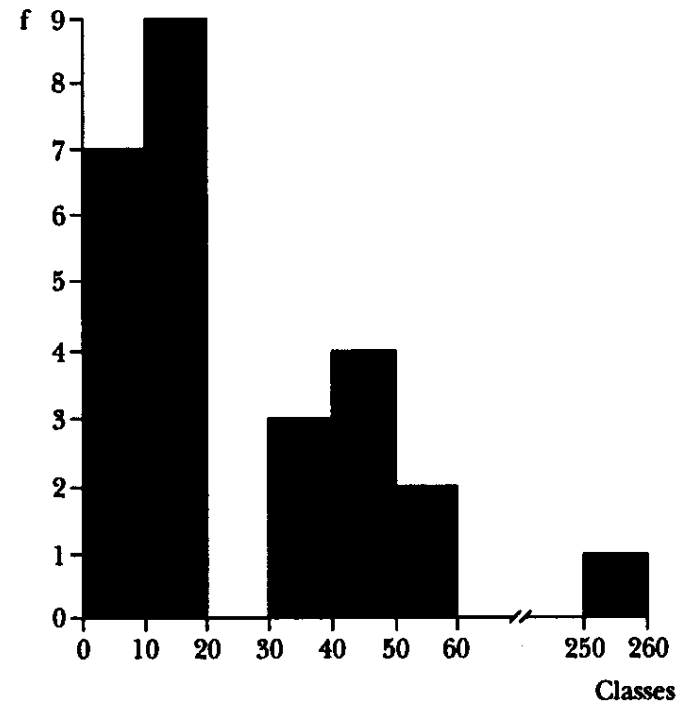
Gráfico de colunas x histograma

BRASIL: POPULAÇÃO RESIDENTE SEGUNDO AS GRANDES REGIÕES — 1991



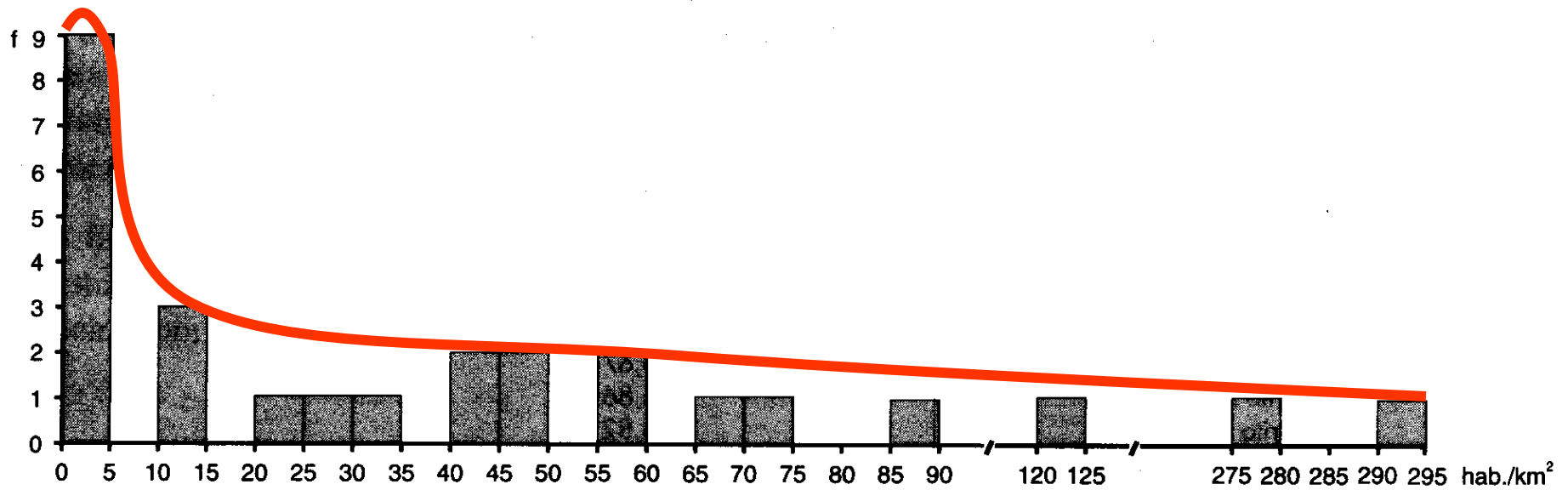
Fonte: IBGE (1993).

Gráfico 2.1 Histograma



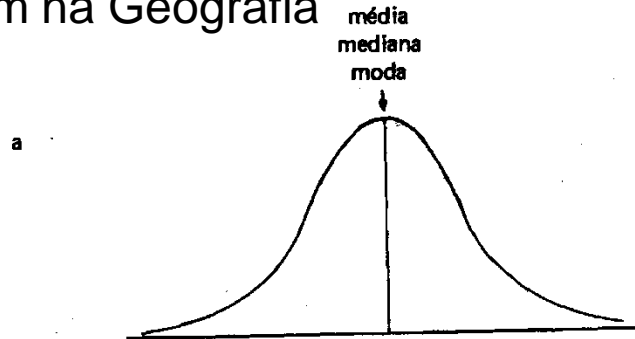
127

HISTOGRAMA

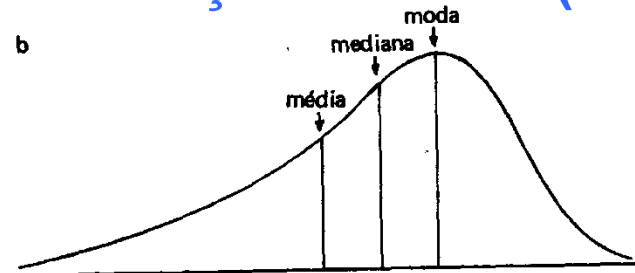


Tipos de curvas de frequência

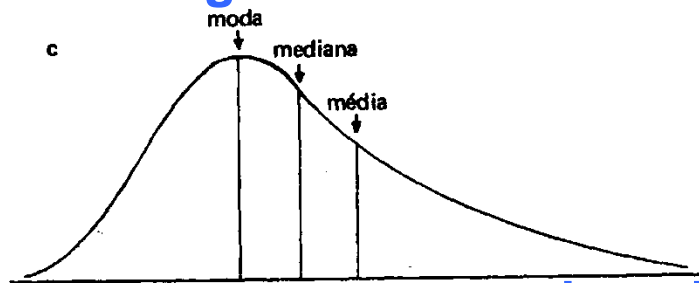
comum na Geografia
Física



distribuição simétrica (normal)



assimetria negativa



assimetria positiva

comum na Geografia Humana

Quanto maior
for a
assimetria,
menos
representativa
é a média
aritmética

Medidas de variabilidade ou dispersão

- ▶ Dois conjuntos de dados podem ter a mesma média mas a distribuição dos dados pode ser bem diferente entre os dois conjuntos
- ▶ Para compreender o grau com que os dados tendem a dispersar-se em torno de um valor central, deve-se calcular as medidas de dispersão

Medidas de variabilidade ou dispersão – AMPLITUDE TOTAL

▶ ATV – amplitude total de variação

Ex. 1) 1, 4, 7, 10, 13

Ex. 2) 4, 5, 7, 8, 11

Média = 7

No exemplo 1, a dispersão vai de 1 a 13
(amplitude total = 12)

No exemplo 2, a dispersão é de 4 a 11
(amplitude total = 7)

Medidas de variabilidade ou dispersão - AMPLITUDE TOTAL

$$ATV = x_{(\max)} - x_{(\min)}$$

$x_{(\max)}$ e $x_{(\min)}$ são os valores maiores e menores dos n-valores

Medidas de variabilidade ou dispersão – AMPLITUDE TOTAL

- Dá uma rápida informação sobre a dispersão dos dados mas é uma medida imprecisa, pois o cálculo só envolve dois valores observados
- Não importa o número de dados
- Não se tem informação sobre a distribuição dos dados dentro do intervalo ou sobre o número de valores que estão perto da média
- É uma medida relativamente boa de variação para um pequeno número de dados

Medidas de variabilidade ou dispersão – Desvio médio

$$DM = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n}$$

\bar{x} = média aritmética;

$|x - \bar{x}|$ = valor absoluto do desvio de x_i em relação à média

Medidas de variabilidade ou dispersão – variância

Na prática é mais usado o desvio médio quadrado – VARIÂNCIA

Média dos quadrados dos desvios em relação à média do conjunto

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

Medidas de variabilidade ou dispersão - variância

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

Como os desvios são elevados ao quadrado, a variância é expressa em unidades quadradas e assim, muito difícil de ser interpretada

Medidas de variabilidade ou dispersão - DESVIO PADRÃO

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

É a raiz da média dos quadrados dos desvios em relação à média do conjunto. Se os valores estão próximos uns dos outros, o desvio padrão é pequeno. Maior a dispersão dos dados, maior o desvio padrão.

VARIÂNCIA E DESVIO PADRÃO para população

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

RELAÇÃO ENTRE DESVIO MÉDIO E DESVIO PADRÃO

**Para distribuições normais o
desvio padrão é 1,25 vezes o
desvio médio**

Exemplo de cálculo de desvio padrão

TABELA 13: Precipitação anual em Salvador (1956-1975). Cálculo do desvio médio, da variância e do desvio padrão

Precipitações anuais (em mm)		
x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
946,1	- 1129,9	1276674,0
1389,8	- 686,2	470870,4
1527,8	- 548,2	300523,2
1559,9	- 516,1	266359,2
1752,6	- 323,4	104587,6
1821,9	- 254,1	64566,8
1838,0	- 238,0	56644,0
1842,7	- 233,3	54428,9
1936,6	- 139,4	19432,4
1938,7	- 137,3	18851,3
1958,3	- 117,7	13853,3
2150,5	74,5	5550,3
2208,0	132,0	17424,0
2409,7	333,7	111355,7
2427,6	351,6	123622,6
2459,7	383,7	147225,7
2479,7	403,7	162973,7
2647,1	571,1	326155,2
2807,0	731,0	534361,0
3417,7	1341,7	1800158,9
41519,4	8646,6	5875618,2

$$\bar{x} = \frac{41519,4}{20} = 2075,97 \approx 2076,0 \text{ mm}$$

$$\text{variância} = \frac{5875618,2}{20} = 293780,9$$

$$\text{desvio médio} = \frac{8646,6}{20} = 432,3 \text{ mm}$$

$$\text{desvio padrão} = \sqrt{293780,9} = 542,0 \text{ mm}$$

Fonte: Ministério da Agricultura, Departamento Nacional de Meteorologia.

FÓRMULA SIMPLIFICADA PARA CÁLCULOS

$$s^2 = \frac{\sum x^2}{n} - \bar{x}^2$$

$$s = \sqrt{\frac{\sum x^2 - (\sum x)^2 / n}{n}}$$

Exemplo de cálculo simplificado de desvio padrão

TABELA 14: Cálculo simplificado da variância e do desvio padrão para a precipitação anual em Salvador (1956-1975).

Precipitações anuais (em mm)	
x_i	x_i^2
946,1	895105,2
1389,8	1931544,0
1527,8	2334172,8
1559,9	2433288,0
1752,6	3071606,8
1821,9	3319319,6
1838,0	3378244,0
1842,7	3395543,3
1936,6	3750419,6
1938,7	3758557,7
1958,3	3834938,9
2150,5	4624650,3
2208,0	4875264,0
2409,7	5806654,1
2427,6	5893241,8
2459,7	6050124,1
2479,7	6148912,1
2647,1	7007138,4
2807,0	7879249,0
3417,7	11680673,3
41519,4	92068647,0

$$s^2 = \frac{\Sigma x^2 - (\Sigma x)^2 / n}{n}$$

$$s^2 = \frac{92068647,0 - \frac{1723860576,4}{20}}{20}$$

$$s^2 = 293780,9$$

$$s = 542,0 \text{ mm}$$

Medidas de dispersão relativa

O coeficiente de variação permite comparar a variabilidade entre diversos conjuntos de dados com médias bem diferentes. Indica a variação relativa.

$$V = \frac{s}{\bar{x}} * 100$$

Coeficiente de variação

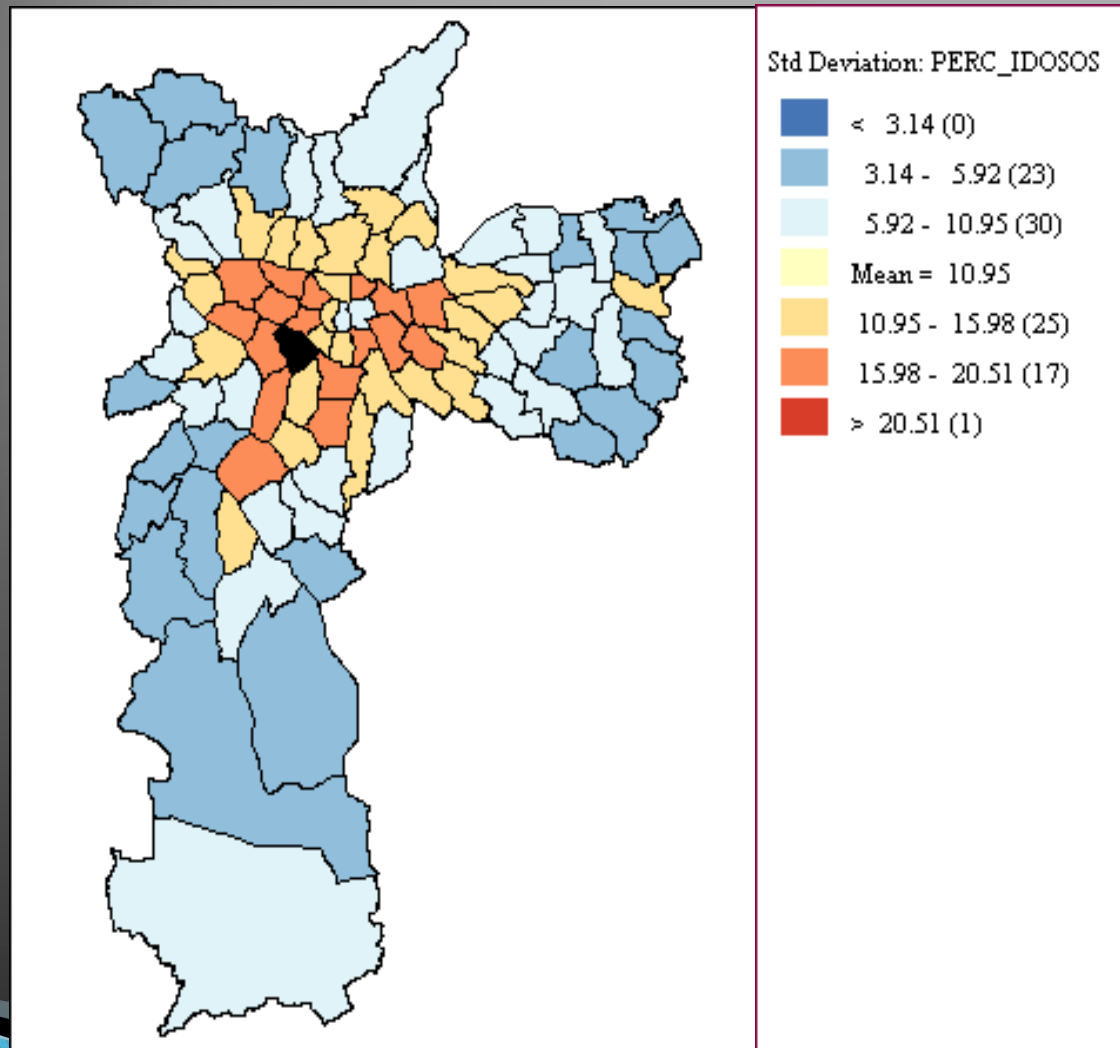
Como desvio padrão e média são dados na mesma unidade, V independe de unidade de medida. V é geralmente expresso em porcentagem.

Estação	Precipitação (mm)	S	V
Pirituba/BA	712,3	319,7	44,8%
Sto Antonio de Jesus/BA	1446,3	319,3	22,08%

Coeficiente de variação

Desvantagem: não é utilizável se a média estiver próxima de zero.

Exemplo: percentual de idosos por distrito no município de São Paulo, 2002



CONCLUINDO

- ▶ Quando os dados apresentam distribuição normal, os intervalos de classe nos mapas com base no desvio padrão permitem uma classificação não só objetiva mas também muito favorável para a interpretação geográfica, porque destaca facilmente as regiões com altos ou baixos valores e as regiões com valores em torno da média facilitando a interpretação e comparação entre regiões.