

# Estatística descritiva

Profa. Dra. Rúbia Gomes Morato

## ***Tipos de dados de acordo com a medição (Rogerson, 2012) ou tipos de escalas estatísticas***

- **Qualitativos**
  - Nominal: atributo sem ordenamento (tipo de solo, vegetação)
  - Ordinal: observações hierarquizadas (tamanhos de cidades)
- **Quantitativos**
  - Intervalar: não possui zero natural (temperatura em graus Celsius ou Fahrenheit)
  - Razão: apresenta a ideia do zero (temperatura em Kelvin)

# Medidas de tendência central

- Média aritmética
- Média Geométrica
- Moda
- Mediana
- Valor máximo e mínimo
- Amplitude

# Medidas de dispersão

- Desvio em relação à média
- Variância da amostra
- Desvio padrão
- Erro padrão
- Coeficiente de variação
- Assimetria
- Curtose

- Correlação
- Regressão

# Medidas de tendência central

# Média aritmética

- Somatório de todos os elementos da série divididos pelo número de elementos.
- Exemplo: 5, 3, 6, 8, 4, 5, 7, 5, 9
- $M_A = (5 + 3 + 6 + 8 + 4 + 5 + 7 + 5 + 9) / 9$
- $M_A = 52/9$
- **A média aritmética é 5,77**

# Média geométrica

- A média geométrica é definida, para números positivos, como a raiz  $n$ -ésima do produto de  $n$  elementos de um conjunto de dados.

$$M_G = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdot \dots \cdot x_n}$$

- Exemplo: 5, 3, 6, 8, 4, 5, 7, 5, 9
- $M_G = \sqrt[9]{5 * 3 * 6 * 8 * 4 * 5 * 7 * 5 * 9}$
- $M_G = \sqrt[9]{4536000} = 5,49$



# Média geométrica

- No Excel: = 4536000^(1/9)
- =num^(1/n)
- "num" é o número cuja raiz se busca encontrar e "n" é a raiz (no exemplo, a nona).
- A média geométrica é usada em alguns casos, como em aplicações financeiras, por exemplo. Entretanto, em certas situações, a geométrica não faz sentido, como por exemplo, quando algum dos valores for zero.

# Moda

- A moda é o valor que ocorre mais vezes ou com maior frequência.
- Exemplo: **5**, 3, 6, 8, 4, **5**, 7, **5**, 9
- O valor mais frequente é **5** (ocorre três vezes), portanto a moda é **5**.

# Mediana

- A mediana é determinada ordenando-se os dados de forma crescente ou decrescente e determinando o valor central da série.
- Exemplo: 3, 4, 5, 5, **5**, 6, 7, 8, 9
- Ou: 9, 8, 7, 6, **5**, 5, 5, 4, 3
- **A mediana é 5**
- Metade dos dados estão à esquerda da mediana e a outra metade à direita da mediana.
- Quando os dados são muito discrepantes, a mediana pode ser afetada por dados extremos, e a moda pode ser mais representativa.

# Valor mínimo e máximo

- O menor e o maior valor da série
- Exemplo: 5, 3, 6, 8, 4, 5, 7, 5, 9
- Ordenando: 3, 4, 5, 5, 5, 6, 7, 8, 9
- **O valor mínimo é 3**
- **O valor máximo é 9**

# Amplitude

- Diferença entre o valor máximo e mínimo
- Exemplo: **3**, 4, 5, 5, 5, 6, 7, 8, **9**
- Amplitude =  $9 - 3$
- **A amplitude é 6**

# Separatrizes/Quantis

- Qualquer separatriz que divide o intervalo de frequência de uma população, ou de uma amostra, em partes iguais:
  - Tercil: cada parte tem 33,3% dos dados
  - Quartil: cada parte tem 25% dos dados
  - Quintil: cada parte tem 20% dos dados
  - Decil: cada parte tem 10% dos dados
  - Duodecil: cada parte tem 8,33% dos dados
  - Percentil: cada parte tem 1% dos dados

# Quartil

- O primeiro quartil corresponde aos primeiros 25% dos dados (começa no menor valor até o primeiro quarto dos dados)
- O segundo quartil corresponde ao intervalo entre 25 e 50% (a mediana)
- O terceiro quartil corresponde ao intervalo entre 50 e 75%
- O quarto quartil corresponde ao intervalos entre 75 e 100% (ou o valor máximo)

# Quartis de uma amostra

- Exemplo: 5, 3, 6, 8, 4, 5, 7, 5, 9
- Ordenando: 3, 4, 5, 5, 5, 6, 7, 8, 9
- **O valor mínimo é 3, o máximo é 9 e a mediana 5**
- A identificação do quartil é determinado por:  
***Número de observações (ordem do quantil/  
quantil)***



# Cálculo de quartis

- **Cálculo:**

***Número de observações (ordem do quantil/quantil)***

***Para quartis (1/4 ou 0,25 ou 25%):***

*Primeiro quartil: número de observações \* 1/4 (ou 0,25)*

*Segundo quartil: número de observações \* 2/4 (ou 0,5)*

*Terceiro quartil: número de observações \* 3/4 (ou 0,75)*

*Quarto quartil: número de observações \* 4/4 (máx)*

## ***Amostra ordenada: 3, 4, 5, 5, 5, 6, 7, 8, 9***

- O primeiro quartil é determinado por:
- $9 \cdot (1/4) = 2,25$  (que pode ser arredondado para 2), correspondendo ao segundo valor, que é 4.
  
- O segundo quartil é determinado por:
- $9 \cdot (2/4) = 4,5$  (que pode ser arredondado para 5), correspondendo ao quinto valor, que é 5.
  
- O terceiro quartil é determinado por:
- $9 \cdot (3/4) = 6,75$  (que pode ser arredondado para 7), correspondendo ao sétimo valor, que é 7.

## ***Amostra ordenada: 3, 4, 5, 5, 5, 6, 7, 8, 9***

- Assim, temos:
- Primeiro quartil: 3 e 4
- Segundo quartil: 5, 5 e 5
- Terceiro quartil: 6 e 7
- Quarto quartil: 8 e 9

# Amplitude interquartílica

## intervalo interquartil

- O intervalo interquartil é utilizado para avaliar o grau de espalhamento de dados (dispersão) em torno da medida de centralidade (mediana).
- Corresponde a diferença entre o primeiro e o terceiro quartil e concentra os 50% dos dados.

## ***Amostra ordenada: 3, 4, 5, 5, 5, 6, 7, 8, 9***

- No exemplo, temos:
- Primeiro quartil: 3 e 4
- ***Segundo quartil: 5, 5 e 5***
- ***Terceiro quartil: 6 e 7***
- Quarto quartil: 8 e 9
  
- O intervalo interquartil corresponde aos valores entre 5 e 7, que concentram 50% dos dados centralizados na mediana.

# Quartil

- O primeiro quartil corresponde aos primeiros 20% dos dados (começa no menor valor até o primeiro quinto dos dados)
- O segundo quartil corresponde ao intervalo entre 20 (segundo decil) e 40% (ou quarto decil)
- O terceiro quartil corresponde ao intervalo entre 40 (quarto decil) e 60% (ou sexto decil)
- O quarto quartil corresponde ao intervalos entre 60 (sexto decil) e 80% (ou oitavo decil)
- O quinto e último quartil corresponde ao intervalos entre 80 (oitavo decil) e 100% dos dados

# Cálculo de quintis

- **Cálculo:**

***Número de observações (ordem do quantil/  
quantil)***

***Para quintis (1/5 ou 0,2 ou 20%):***

*Primeiro quintil : número de observações \* 1/5*

*Segundo quintil : número de observações \* 2/5*

*Terceiro quintil : número de observações \* 3/5*

*Quarto quintil : número de observações \* 4/5*

*Quinto quintil: número de observações \* 5/5(máx)*

# Vantagens/desvantagens dos quantis

- Definição dos intervalos para mapa coropléticos de modo equilibrado (cada classe tem aproximadamente a mesma quantidade de unidades)
- Desvantagens: pode separar unidades semelhantes e resultar em classes heterogêneas, agrupando unidades diferentes e separando unidades semelhantes



# Medidas de dispersão

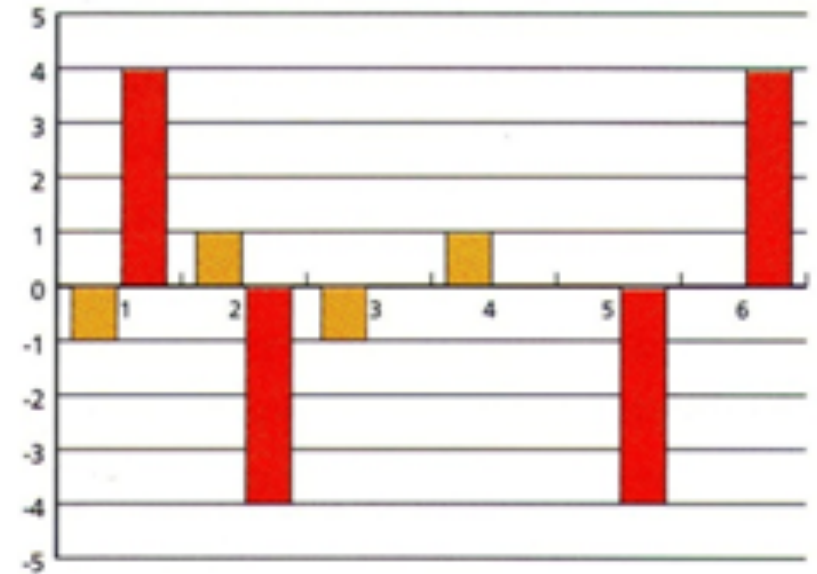
# Desvio em relação à média

- Diferença entre o valor observado e a média
- Fornece uma ideia da variabilidade dos dados em torno da média.

$$DM = x_i - \bar{x}$$

# Desvio em relação à média (Galvani, 2011)

A	B	DM "A"	DM "B"
4	9	-1	4
6	1	1	-4
4	5	-1	0
6	5	1	0
5	1	0	-4
5	9	0	4
$\bar{x}=5$	$\bar{x}=5$	$\Sigma=0$	$\Sigma=0$



Mesma média  
Desvios diferentes!

# Variância da amostra

- Somatória do quadrado do desvio em relação à média, dividida pela quantidade de elementos da série menos 1.

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}$$

# Variância da amostra (Galvani, 2011)

$x$	$x - \bar{x}$	$(x - \bar{x})^2$
4	-1	1
6	1	1
4	-1	1
6	1	1
5	0	0
5	0	0
$\bar{x} = 5$		$\sum(x - \bar{x})^2 = 4$

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}$$

$$S^2 = \frac{4}{6-1}$$

$$S^2 = 0,8$$

# Encontrar a variância da amostra

- **Amostra: 5, 10, 15, 5, 25**
- Média ( $\bar{x}$ ):  $60/5 = 12$
- $n = 5$

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}$$

- **Variância:**
- $[(-7)^2 + (-2)^2 + (3)^2 + (-7)^2 + (13)^2] / (5-1)$
- $[49 + 4 + 9 + 49 + 169] / 4$
- $280 / 4 = 70$



# Desvio padrão

- Raiz da variância

$$S = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

- Medida do grau de dispersão em relação à média.

# Cálculo do desvio-padrão

<i>Série</i>	<i>Desvio da média</i>	<i>Quadrado do desvio da média</i>
9	4	16
1	-4	16
1	-4	16
2	-3	9
8	3	9
9	4	16
<i>Média</i>	$(9+1+1+2+8+9)/6 \Rightarrow$	$30/6 = 5$
<i>Soma do quadrado dos desvios (16+16+16+9+9+16)</i>		82
<i>Variância (Soma do quadrado dos desvios/n-1) = 82/5</i>		16,4
<i>Desvio-padrão (Raiz da variância)</i>		4,049691



# Erro padrão

- O erro padrão é uma medida de variação de uma média amostral em relação à média da população.
- A aplicada para verificar a confiabilidade da média amostral calculada.
- É obtida pela divisão do desvio padrão pela raiz quadrada do tamanho amostral.
- Quanto menor o erro padrão, menor a dispersão e mais provável que qualquer média de amostra esteja próxima à média da população.

# Erro padrão

- $S_x$ : é o erro padrão
- $s$ : é o desvio padrão
- $n$ : é o tamanho da amostra

$$S_x = \frac{s}{\sqrt{n}}$$

- Exemplo: 9, 1, 1, 2, 8, 9
- Média: 5
- Desvio padrão: 4,049691
- $N = 6$

- $S_x = 4,049691 / \sqrt{6}$
- $S_x = 4,049691 / 2,4494$
- $S_x = \mathbf{1,6533}$

# Coeficiente de variação

- Expresso em porcentagem, permite comparar variáveis diferentes.

$$CV = \frac{100 \cdot S}{\bar{x}}$$

- Multiplica-se o desvio padrão por 100 e divide-se pela média.

# Coeficiente de variação (Galvani, 2011)

A	B	C
4	9	9
6	1	1
4	5	1
6	5	2
5	1	8
5	9	9

$$CV = \frac{100 \cdot S}{\bar{x}}$$

Desvio-padrão de A = 0,9

Desvio-padrão de B = 3,6

Desvio-padrão de C = 4,0

## **Série C**

Média = 5

Desvio-padrão = 4,0

$$CV = (100 * 4) / 5$$

$$CV = 400/5 = \mathbf{80\%}$$

$$CV_A = 18,0\%$$

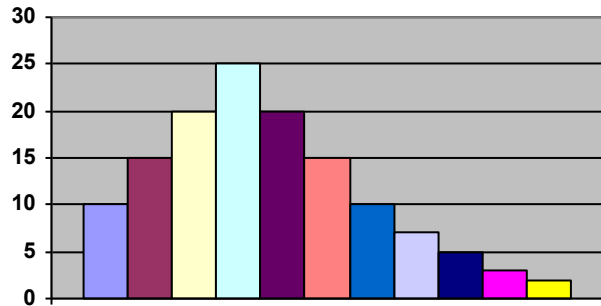
$$CV_B = 72,0\%$$

$$CV_C = 80,0\%$$

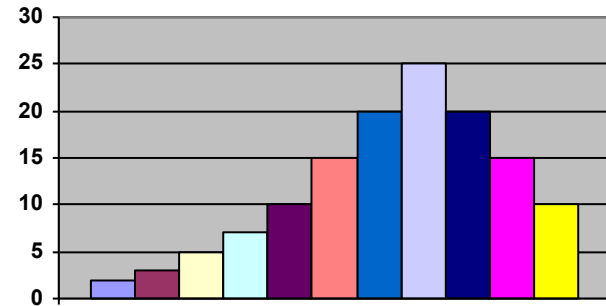
# Assimetria

- Mede o grau de assimetria exibido pelos dados e pelo histograma.
- Quando há mais observações abaixo da média a assimetria é positiva.
- Quando há mais observações acima da média a assimetria é negativa.

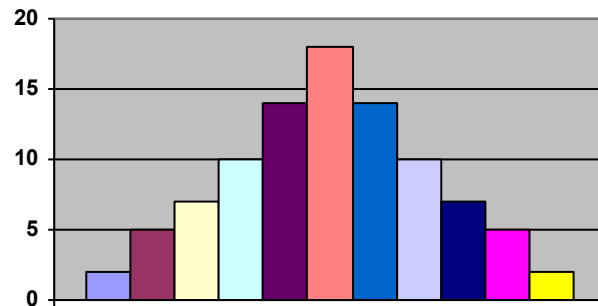
# Assimetria (GUIMARÃES, 1997)



assimetria positiva



assimetria negativa



Distribuição simétrica

# Assimetria

- A assimetria é calculada primeiro somando os cubos dos desvios da média, e então, dividindo desvio padrão.

$$\frac{1}{n} \sum \left[ \frac{X_i - \bar{X}}{s} \right]^3$$

# Curtose

- Mede o alongamento/achatamento do histograma. A fórmula é semelhante à da assimetria, com a ressalva de que a quarta potência é usada em vez da terceira:

$$\frac{1}{n} \sum \left[ \frac{x_i - \bar{x}}{s} \right]^4$$

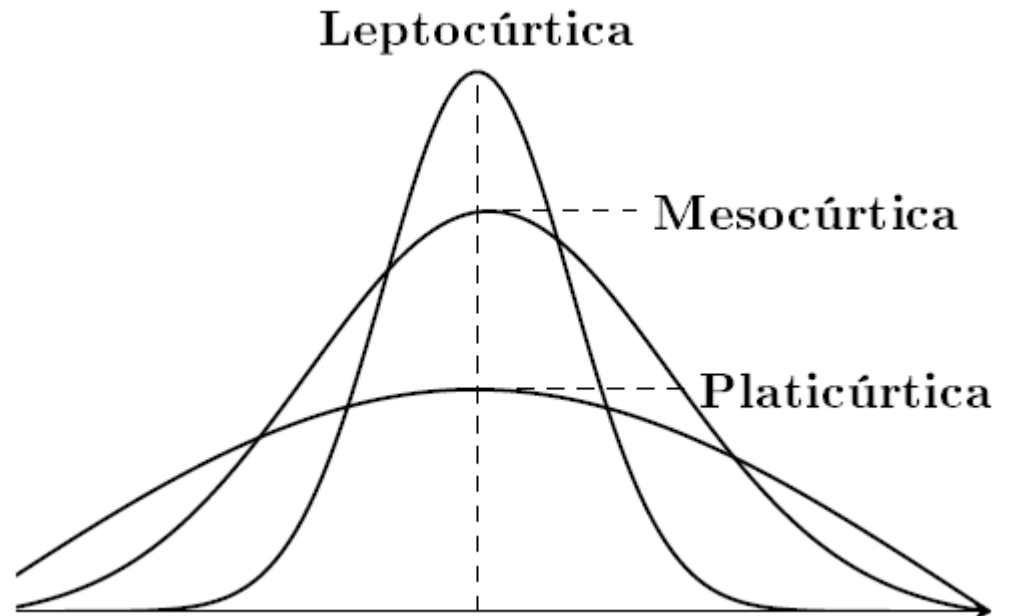
- Soma-se as quartas potências dos desvios da média, e então, divide-se pelo desvio padrão.



# Curtose (Previdelli, 2018)

- De acordo com esta medida temos a seguinte classificação:

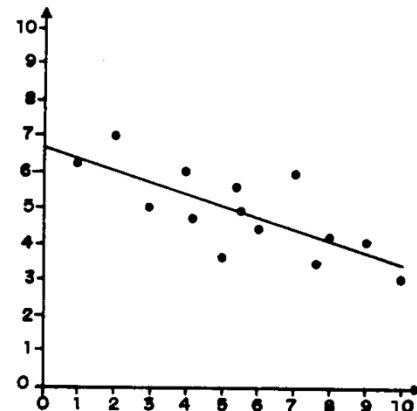
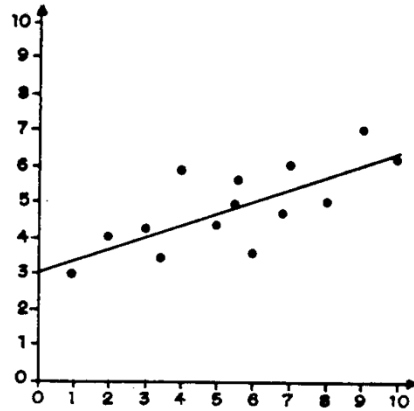
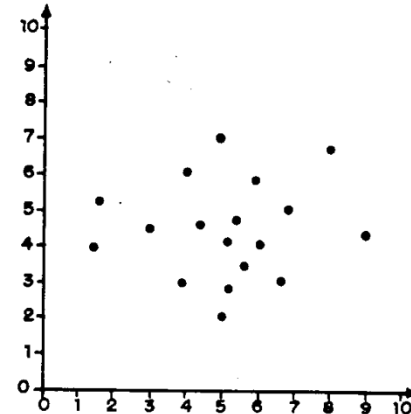
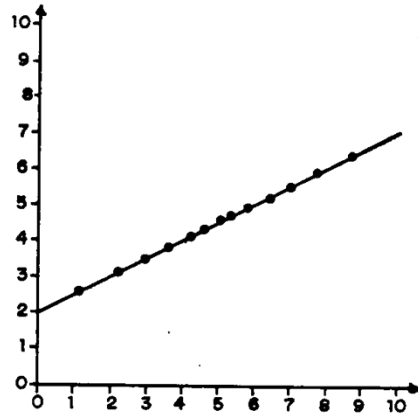
- $k < 0$ , Platicúrtica
- $k = 0$ , Mesocúrtica
- $k > 0$ , Leptocúrtica



# Correlação e Regressão (LUCHIARI, 2010)

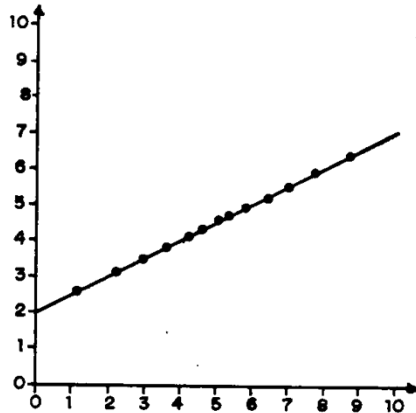
- Verificação se há correlação entre variáveis?
- Quando existe é possível efetuar estimativas?
- A regressão linear permite estimar uma variável em relação a outra.

# Método gráfico

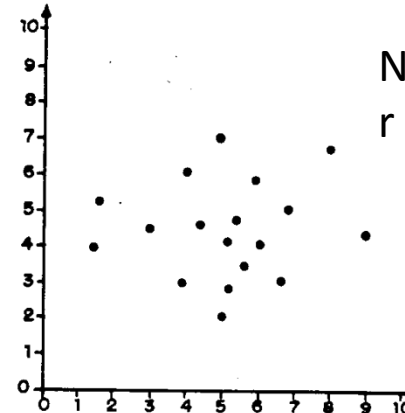


# Correlação e Regressão

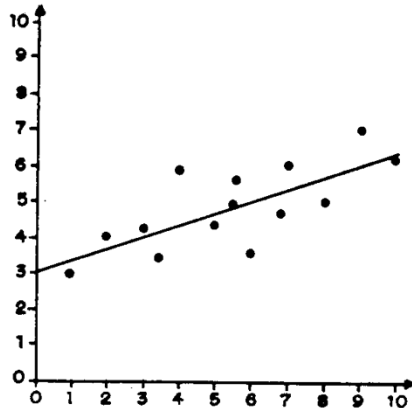
Correlação perfeita  
 $r = 1$



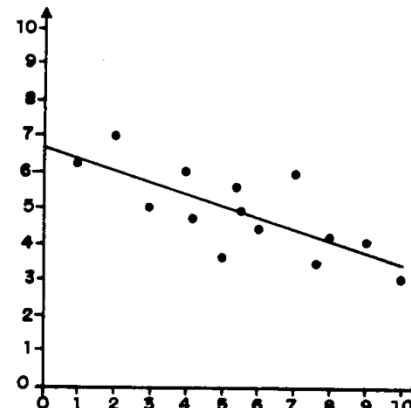
Nenhuma Correlação  
 $r = 0$



Correlação linear  
positiva  $r = 0,74$



Correlação linear  
negativa  $r = -0,74$



## Correlação – Como calcular (LUCIARI, 2010)

$$r = \frac{N \sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{[N \sum x_i^2 - (\sum x_i)^2][N \sum y_i^2 - (\sum y_i)^2]}}$$

<b>Município</b>	<b>Área = X</b>	<b>Estabel.</b>	<b>cabeças = Y</b>
Andradina	48 137	885	87 430
Alto Alegre	8 703	392	22 683
Araçatuba	59 120	499	81 699
Avanhandava	10 642	118	14 848
Barbosa	5 324	83	8 886
Bento de Abreu	6 822	65	15 588
Bilac	7 725	259	16 720
Birigui	21 226	516	46 864
Braúna	5 627	208	12 682
Brejo Alegre	1 528	65	4 853
Buritama	14 535	362	32 318
Castilho	38 319	590	61 824
Clementina	7 025	174	14 091
Coroados	9 029	198	15 240
Gabriel Monteiro	9 159	178	16 183
Glicério	7 802	163	11 604
Guaraçai	26 999	415	53 291

<b>Município</b>	<b>Área = X</b>	<b>Estabel.</b>	<b>cabeças = Y</b>	<b>X<sub>i</sub><sup>2</sup></b>
Andradina	48 137	885	87 430	2 317 178 471
Alto Alegre	8 703	392	22 683	75 740 817
Araçatuba	59 120	499	81 699	3 495 175 582
Avanhandava	10 642	118	14 848	113 253 441
Barbosa	5 324	83	8 886	28 343 698
Bento de Abreu	6 822	65	15 588	46 543 641
Bilac	7 725	259	16 720	59 679 024
Birigui	21 226	516	46 864	450 543 501
Braúna	5 627	208	12 682	31 658 628
Brejo Alegre	1 528	65	4 853	2 334 448
Buritama	14 535	362	32 318	211 277 562
Castilho	38 319	590	61 824	1 468 313 573
Clementina	7 025	174	14 091	49 346 129
Coroados	9 029	198	15 240	81 529 703
Gabriel Monteiro	9 159	178	16 183	83 883 434
Glicério	7 802	163	11 604	60 876 197
Guaraçai	26 999	415	53 291	728 924 402

<b>Município</b>	<b>Área = X</b>	<b>Estabel.</b>	<b>cabeças = Y</b>	<b>X<sub>i</sub><sup>2</sup></b>	<b>y<sub>i</sub><sup>2</sup></b>
Andradina	48 137	885	87 430	2 317 178 471	7 644 004 900
Alto Alegre	8 703	392	22 683	75 740 817	514 518 489
Araçatuba	59 120	499	81 699	3 495 175 582	6 674 726 601
Avanhandava	10 642	118	14 848	113 253 441	220 463 104
Barbosa	5 324	83	8 886	28 343 698	78 960 996
Bento de Abreu	6 822	65	15 588	46 543 641	242 985 744
Bilac	7 725	259	16 720	59 679 024	279 558 400
Birigui	21 226	516	46 864	450 543 501	2 196 234 496
Braúna	5 627	208	12 682	31 658 628	160 833 124
Brejo Alegre	1 528	65	4 853	2 334 448	23 551 609
Buritama	14 535	362	32 318	211 277 562	1 044 453 124
Castilho	38 319	590	61 824	1 468 313 573	3 822 206 976
Clementina	7 025	174	14 091	49 346 129	198 556 281
Coroados	9 029	198	15 240	81 529 703	232 257 600
Gabriel Monteiro	9 159	178	16 183	83 883 434	261 889 489
Glicério	7 802	163	11 604	60 876 197	134 652 816
Guaraçai	26 999	415	53 291	728 924 402	2 839 930 681



<b>Município</b>	<b>Área = X</b>	<b>Estabel.</b>	<b>cabeças = Y</b>	<b>X<sub>i</sub><sup>2</sup></b>	<b>Y<sub>i</sub><sup>2</sup></b>	<b>X<sub>i</sub>Y<sub>i</sub></b>
Andradina	48 137	885	87 430	2 317 178 471	7 644 004 900	4 208 624 904
Alto Alegre	8 703	392	22 683	75 740 817	514 518 489	197 408 334
Araçatuba	59 120	499	81 699	3 495 175 582	6 674 726 601	4 830 045 697
Avanhandava	10 642	118	14 848	113 253 441	220 463 104	158 013 307
Barbosa	5 324	83	8 886	28 343 698	78 960 996	47 307 998
Bento de Abreu	6 822	65	15 588	46 543 641	242 985 744	106 345 857
Bilac	7 725	259	16 720	59 679 024	279 558 400	129 165 678
Birigui	21 226	516	46 864	450 543 501	2 196 234 496	994 735 733
Braúna	5 627	208	12 682	31 658 628	160 833 124	71 356 541
Brejo Alegre	1 528	65	4 853	2 334 448	23 551 609	7 414 850
Buritama	14 535	362	32 318	211 277 562	1 044 453 124	469 754 734
Castilho	38 319	590	61 824	1 468 313 573	3 822 206 976	2 369 007 890
Clementina	7 025	174	14 091	49 346 129	198 556 281	98 984 766
Coroados	9 029	198	15 240	81 529 703	232 257 600	137 607 751
Gabriel						
Monteiro	9 159	178	16 183	83 883 434	261 889 489	148 216 699
Glicério	7 802	163	11 604	60 876 197	134 652 816	90 538 121
Guaraçai	26 999	415	53 291	728 924 402	2 839 930 681	1 438 782 393

Município	Área = X	Estabel.	cabeças = Y	$X_i^2$	$Y_i^2$	$X_i Y_i$
Andradina	48 137	885	87 430	2 317 178 471	7 644 004 900	4 208 624 904
Alto Alegre	8 703	392	22 683	75 740 817	514 518 489	197 408 334
Araçatuba	59 120	499	81 699	3 495 175 582	6 674 726 601	4 830 045 697
Avanhandava	10 642	118	14 848	113 253 441	220 463 104	158 013 307
Barbosa	5 324	83	8 886	28 343 698	78 960 996	47 307 998
Bento de Abreu	6 822	65	15 588	46 543 641	242 985 744	106 345 857
Bilac	7 725	259	16 720	59 679 024	279 558 400	129 165 678
Birigui	21 226	516	46 864	450 543 501	2 196 234 496	994 735 733
Braúna	5 627	208	12 682	31 658 628	160 833 124	71 356 541
Brejo Alegre	1 528	65	4 853	2 334 448	23 551 609	7 414 850
Buritama	14 535	362	32 318	211 277 562	1 044 453 124	469 754 734
Castilho	38 319	590	61 824	1 468 313 573	3 822 206 976	2 369 007 890
Clementina	7 025	174	14 091	49 346 129	198 556 281	98 984 766
Coroados	9 029	198	15 240	81 529 703	232 257 600	137 607 751
Gabriel Monteiro	9 159	178	16 183	83 883 434	261 889 489	148 216 699
Glicério	7 802	163	11 604	60 876 197	134 652 816	90 538 121
Guaraçá	26 999	415	53 291	728 924 402	2 839 930 681	1 438 782 393
	<b>287.721,70</b>		<b>516.804,00</b>	<b>9 304 602 252</b>	<b>26 569 784 430</b>	<b>15 503 311 253</b>

$$r = \frac{N \sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{[N \sum x_i^2 - (\sum x_i)^2][N \sum y_i^2 - (\sum y_i)^2]}}$$

$$N = 17$$

$$\sum X_i Y_i = 15\,503\,311\,253$$

$$\sum X_i = 287\,722$$

$$\sum Y_i = 516\,804$$

$$\sum X_i^2 = 9\,304\,602\,252$$

$$\sum (X_i)^2 = 82\,783\,776\,651$$

$$\sum Y_i^2 = 26\,569\,784\,430$$

$$\sum (Y_i)^2 = 2,67086E+11$$

$$r = \frac{N \sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{[N \sum x_i^2 - (\sum x_i)^2][N \sum y_i^2 - (\sum y_i)^2]}}$$

$$\text{Numerador} = (17) * (15\,503\,311\,253) - (287\,722) * (516\,804)$$

$$\text{Numerador} = (2,63556\text{E}+11) - (1,48696\text{E}+11)$$

$$\text{Numerador} = 1,14861\text{E}+11$$

$$r = \frac{N \sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{[N \sum x_i^2 - (\sum x_i)^2][N \sum y_i^2 - (\sum y_i)^2]}}$$

$$\text{Denominador} = [(17) * (9\ 304\ 602\ 252) - (82\ 783\ 776\ 651)] * [(17) * (26\ 569\ 784\ 430) - (2,67086E+11)]^{1/2}$$

$$\text{Denominador} = [(1,58178E+11) - (82\ 783\ 776\ 651)] * [(4,51686E+11) - (2,67086E+11)]^{1/2}$$

$$\text{Denominador} = [75\ 394\ 461\ 625] * [1,846E+11]^{1/2}$$

$$\text{Denominador} = [1,39178E+22]^{1/2}$$

$$\text{Denominador} = 1,17974E+11$$

$$r = \frac{N \sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{[N \sum x_i^2 - (\sum x_i)^2][N \sum y_i^2 - (\sum y_i)^2]}}$$

$$r = 1,14861E+11 / 1,17974E+11$$

$$r = 0,973610899$$

$$r = 0,97$$

# Regressão Linear

- Ajustar uma reta à distribuição de pontos.
- Mínimos quadrados

# Regressão Linear – Como calcular (LUCHIARI, 2010)

$$\hat{y} = a + bx$$



Encontrar os dois termos (a e b)

$$b = \frac{\sum xy - (\sum x \sum y) / N}{\sum x^2 - (\sum x)^2 / N}$$

$$a = \mu_y - b \mu_x$$

<b>Município</b>	<b>Área = X</b>	<b>Estabel.</b>	<b>cabeças = Y</b>
Andradina	48 137	885	87 430
Alto Alegre	8 703	392	22 683
Araçatuba	59 120	499	81 699
Avanhandava	10 642	118	14 848
Barbosa	5 324	83	8 886
Bento de Abreu	6 822	65	15 588
Bilac	7 725	259	16 720
Birigui	21 226	516	46 864
Braúna	5 627	208	12 682
Brejo Alegre	1 528	65	4 853
Buritama	14 535	362	32 318
Castilho	38 319	590	61 824
Clementina	7 025	174	14 091
Coroados	9 029	198	15 240
Gabriel Monteiro	9 159	178	16 183
Glicério	7 802	163	11 604
Guaraçai	26 999	415	53 291

<b>Município</b>	<b>Área = X</b>	<b>Estabel.</b>	<b>cabeças = Y</b>	$X_i Y_i$	$X^2$
Andradina	48 137	885	87 430	4 208 624 904	2 317 178 471
Alto Alegre	8 703	392	22 683	197 408 334	75 740 817
Araçatuba	59 120	499	81 699	4 830 045 697	3 495 175 582
Avanhandava	10 642	118	14 848	158 013 307	113 253 441
Barbosa	5 324	83	8 886	47 307 998	28 343 698
Bento de Abreu	6 822	65	15 588	106 345 857	46 543 641
Bilac	7 725	259	16 720	129 165 678	59 679 024
Birigui	21 226	516	46 864	994 735 733	450 543 501
Braúna	5 627	208	12 682	71 356 541	31 658 628
Brejo Alegre	1 528	65	4 853	7 414 850	2 334 448
Buritama	14 535	362	32 318	469 754 734	211 277 562
Castilho	38 319	590	61 824	2 369 007 890	1 468 313 573
Clementina	7 025	174	14 091	98 984 766	49 346 129
Coroados	9 029	198	15 240	137 607 751	81 529 703
Gabriel Monteiro	9 159	178	16 183	148 216 699	83 883 434
Glicério	7 802	163	11 604	90 538 121	60 876 197
Guaraçai	26 999	415	53 291	1 438 782 393	728 924 402

Município	Área = X	Estabel.	cabeças = Y	$X_i Y_i$	$X^2$
Andradina	48 137	885	87 430	4 208 624 904	2 317 178 471
Alto Alegre	8 703	392	22 683	197 408 334	75 740 817
Araçatuba	59 120	499	81 699	4 830 045 697	3 495 175 582
Avanhandava	10 642	118	14 848	158 013 307	113 253 441
Barbosa	5 324	83	8 886	47 307 998	28 343 698
Bento de Abreu	6 822	65	15 588	106 345 857	46 543 641
Bilac	7 725	259	16 720	129 165 678	59 679 024
Birigui	21 226	516	46 864	994 735 733	450 543 501
Braúna	5 627	208	12 682	71 356 541	31 658 628
Brejo Alegre	1 528	65	4 853	7 414 850	2 334 448
Buritama	14 535	362	32 318	469 754 734	211 277 562
Castilho	38 319	590	61 824	2 369 007 890	1 468 313 573
Clementina	7 025	174	14 091	98 984 766	49 346 129
Coroados	9 029	198	15 240	137 607 751	81 529 703
Gabriel Monteiro	9 159	178	16 183	148 216 699	83 883 434
Glicério	7 802	163	11 604	90 538 121	60 876 197
Guaraçai	26 999	415	53 291	1 438 782 393	728 924 402
	<b>287 722</b>		<b>516 804</b>	<b>15 503 311 253</b>	<b>9 304 602 252</b>

$$b = \frac{\sum xy - (\sum x \sum y) / N}{\sum x^2 - (\sum x)^2 / N}$$

$$\text{Numerador} = 15\,503\,311\,253 - (287\,722 * 516\,804) / 17$$

$$\text{Numerador} = 15\,503\,311\,253 - (148\,695\,725\,447) / 17$$

$$\text{Numerador} = 15\,503\,311\,253 - (8746807379 )$$

$$\text{Numerador} = 6\,756\,503\,873$$

$$b = \frac{\sum xy - (\sum x \sum y) / N}{\sum x^2 - (\sum x)^2 / N}$$

**Denominador = 9 304 602 252 – (82783776651) / 17**

**Denominador = 9 304 602 252 – (4869633921)**

**Denominador = 4 434 968 331**

$$b = \frac{\sum xy - (\sum x \sum y) / N}{\sum x^2 - (\sum x)^2 / N}$$

$$b = 6\,756\,503\,873 / 4\,434\,968\,331$$

$$b = 1,523461583$$

$$a = \mu_y - b \mu_x$$

$$a = 30400,23529 - (1,523461583 * 16924,8059)$$

$$a = 30400,23529 - (25784,2916 )$$

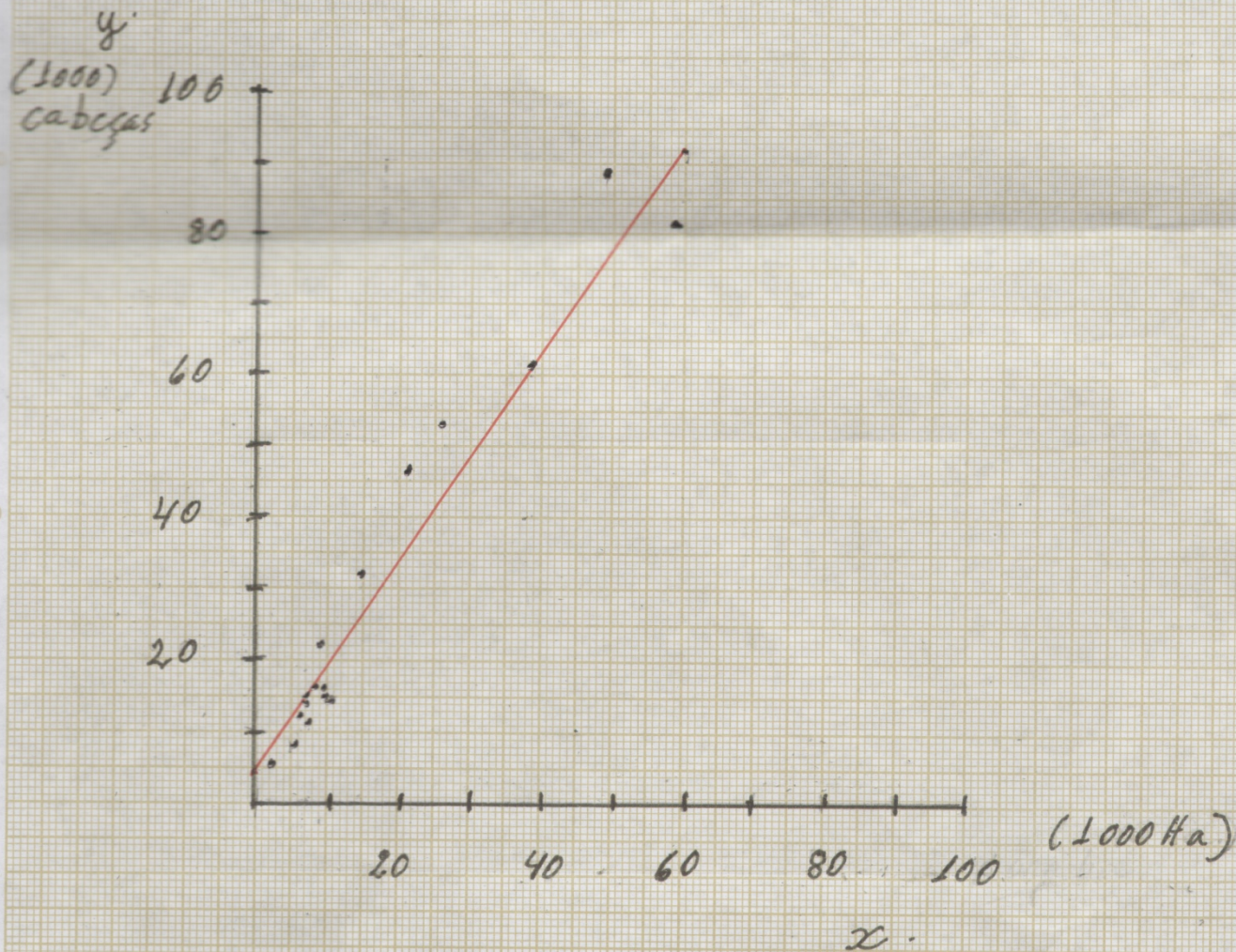
$$a = 4615,943734$$



$$\hat{y} = a + bx$$

$$Y_{\text{estim}} = 4615,943734 + 1,523461583 X$$

# Regressão (Área x Bovinos)



# Exercício 1

Estatística descritiva

Past



# Arquivo com dados dos setores censitários da Vila Andrade (formato do Excel)

The image shows a screenshot of an Excel spreadsheet titled "Dados\_Ex\_1\_V\_Andrade.xls". The spreadsheet displays data for 34 census sectors in Vila Andrade. The columns are labeled as follows: A (CD\_GEOCODI), B (TIPO), C (NM\_DISTRI), D (A\_AGU), E (C\_ESG), F (C\_LIXO), G (ENERGIA), H (RAMP/ARBOR), I (JUST\_AM), J (RENDA\_M), K (PESS), L (BRANC), M (NEGROS), N (AMAREL), O (PARDOS), P (INDIGENAS), and Q (INDIGENAS). The data is organized in a table with 34 rows, each representing a census sector. The first row is the header, and the subsequent rows contain numerical values for each indicator. The status bar at the bottom indicates "Normal View" and "Ready".

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	CD_GEOCODI	TIPO	NM_DISTRI	A_AGU	C_ESG	C_LIXO	ENERGIA	RAMP/ARBOR	JUST_AM	RENDA_M	PESS	BRANC	NEGROS	AMAREL	PARDOS	INDIGENAS				
2	355030883000073	URBANC	VILA ANDRADE	0,986	0,958	0,846	0,406	0,000	0,000	0,533	932,71	515	0,299	0,070	0,010	0,620	0,000			
3	355030883000074	URBANC	VILA ANDRADE	1,000	0,092	0,989	0,867	0,000	0,000	0,491	837,90	856	0,303	0,070	0,000	0,630	0,000			
4	355030883000075	URBANC	VILA ANDRADE	1,000	0,517	1,000	1,000	0,034	0,983	0,756	15470,64	202	0,965	0,000	0,000	0,040	0,000			
5	355030883000076	URBANC	VILA ANDRADE	0,996	0,996	1,000	0,996	0,000	0,000	0,665	1148,97	857	0,270	0,030	0,010	0,690	0,000			
6	355030883000077	URBANC	VILA ANDRADE	1,000	0,994	0,994	1,000	0,018	0,343	0,725	3449,11	520	0,706	0,010	0,010	0,260	0,000			
7	355030883000078	URBANC	VILA ANDRADE	1,000	1,000	1,000	1,000	0,008	0,000	0,668	6465,88	372	0,782	0,040	0,070	0,110	0,000			
8	355030883000079	URBANC	VILA ANDRADE	1,000	1,000	0,044	1,000	0,122	0,000	0,528	6432,57	198	0,848	0,010	0,030	0,110	0,000			
9	355030883000080	URBANC	VILA ANDRADE	0,989	1,000	0,011	1,000	0,033	0,000	0,505	6388,37	214	0,930	0,000	0,030	0,040	0,000			
10	355030883000081	URBANC	VILA ANDRADE	1,000	1,000	0,075	1,000	0,050	0,000	0,521	5978,75	214	0,930	0,010	0,010	0,060	0,000			
11	355030883000082	URBANC	VILA ANDRADE	1,000	1,000	1,000	1,000	0,020	1,000	0,837	20901,47	351	0,915	0,000	0,080	0,000	0,000			
12	355030883000083	URBANC	VILA ANDRADE	1,000	1,000	1,000	1,000	0,017	0,000	0,669	26711,02	188	1,000	0,000	0,000	0,000	0,000			
13	355030883000084	URBANC	VILA ANDRADE	1,000	0,970	1,000	0,990	0,020	0,945	0,821	2806,76	655	0,602	0,060	0,000	0,340	0,000			
14	355030883000085	URBANC	VILA ANDRADE	0,987	0,027	1,000	0,007	0,000	0,000	0,337	974,22	556	0,345	0,090	0,000	0,570	0,000			
15	355030883000086	URBANC	VILA ANDRADE	1,000	0,849	0,692	1,000	0,009	0,405	0,659	2381,69	1089	0,518	0,120	0,020	0,350	0,000			
16	355030883000185	URBANC	VILA ANDRADE	1,000	1,000	1,000	1,000	0,000	1,000	0,833	16067,28	312	0,795	0,030	0,120	0,060	0,000			
17	355030883000186	URBANC	VILA ANDRADE	1,000	0,955	0,970	1,000	0,045	1,000	0,828	19174,84	243	0,733	0,030	0,100	0,140	0,000			
18	355030883000187	URBANC	VILA ANDRADE	1,000	0,954	1,000	1,000	0,080	1,000	0,839	19196,30	313	0,850	0,010	0,050	0,090	0,000			
19	355030883000188	URBANC	VILA ANDRADE	1,000	1,000	0,960	1,000	0,027	1,000	0,831	20660,40	282	0,887	0,030	0,040	0,040	0,000			
20	355030883000189	URBANC	VILA ANDRADE	1,000	1,000	0,927	1,000	0,000	1,000	0,821	22837,56	143	0,867	0,000	0,070	0,060	0,000			
21	355030883000190	URBANC	VILA ANDRADE	1,000	0,955	1,000	1,000	0,000	1,000	0,826	18494,45	308	0,812	0,010	0,150	0,030	0,000			
22	355030883000191	URBANC	VILA ANDRADE	0,997	1,000	0,784	1,000	0,136	1,000	0,820	11584,89	288	0,531	0,050	0,200	0,220	0,000			
23	355030883000192	URBANC	VILA ANDRADE	1,000	1,000	0,725	0,978	0,088	1,000	0,799	14245,30	313	0,514	0,030	0,230	0,230	0,000			
24	355030883000193	URBANC	VILA ANDRADE	0,994	1,000	0,973	1,000	0,047	0,604	0,770	7879,00	870	0,800	0,030	0,010	0,160	0,000			
25	355030883000194	URBANC	VILA ANDRADE	1,000	0,980	0,569	1,000	0,039	1,000	0,765	14535,29	157	0,675	0,050	0,030	0,240	0,000			
26	355030883000195	URBANC	VILA ANDRADE	0,946	0,973	0,905	1,000	0,000	0,000	0,645	1304,21	1599	0,291	0,110	0,000	0,600	0,000			
27	355030883000196	URBANC	VILA ANDRADE	1,000	0,924	0,996	1,000	0,027	0,996	0,824	5316,24	655	0,692	0,060	0,020	0,230	0,000			
28	355030883000197	URBANC	VILA ANDRADE	1,000	0,938	0,992	0,992	0,031	1,000	0,817	5795,75	325	0,840	0,030	0,030	0,100	0,000			
29	355030883000198	URBANC	VILA ANDRADE	1,000	0,980	0,260	1,000	0,030	1,000	0,712	8823,95	546	0,811	0,010	0,030	0,160	0,000			
30	355030883000001	URBANC	VILA ANDRADE	1,000	0,836	1,000	0,995	0,817	1,000	0,941	8527,94	723	0,674	0,020	0,010	0,290	0,000			
31	355030883000002	URBANC	VILA ANDRADE	1,000	0,898	1,000	1,000	0,054	1,000	0,825	11207,31	515	0,893	0,020	0,020	0,070	0,000			
32	355030883000003	URBANC	VILA ANDRADE	1,000	0,978	0,987	0,991	0,142	0,987	0,848	10319,24	871	0,900	0,020	0,010	0,060	0,000			
33	355030883000004	URBANC	VILA ANDRADE	1,000	1,000	0,048	1,000	0,000	1,000	0,675	1196,11	444	0,295	0,140	0,000	0,560	0,000			
34	355030883000005	URBANC	VILA ANDRADE	1,000	1,000	1,000	1,000	0,000	1,000	0,833	16012,45	503	0,905	0,000	0,070	0,020	0,000			



# Estatística descritiva no Past

- Fazer o download do arquivo Dados\_Ex\_1\_V\_Andrade.xls e salvar numa pasta conhecida.
- Não é recomendado abrir o arquivo no Excel e Past ao mesmo tempo. É necessário abrir apenas pelo Past para fazer o exercício!
- Abrir o Past
- Seguir o caminho no menu do Past: Past > Open > indicar o diretório onde o arquivo Dados\_Ex\_1\_V\_Andrade.xls foi salvo, selecionar o arquivo e clicar em abrir.

# Estatística descritiva no Past

- Em import setting, manter selecionado Names, data em Rows contains
- Manter selecionado Names, data em Columns contains e clicar em ok.
- Para calcular as estatísticas descritivas, é necessário clicar na coluna e no menu clicar em Univariate > Summary statistics.
- É necessário clicar em cada coluna e repetir o caminho para cada variável da tabela.

# Estatística descritiva no Past

Não é necessário preencher todas as células da coluna, mas apenas as que apoiam a respostas das questões:

1. Quais as médias aritméticas e geométricas do abastecimento de água e do esgotamento sanitário? São semelhantes? Por quê?
2. Qual a mediana da renda? Qual a faixa de renda compreende o intervalo entre o mínimo e o primeiro quartil? Qual o intervalo entre o limite do terceiro quartil e o máximo (quarto quartil)?
3. Qual o desvio-padrão da arborização nas calçada? O que ele revela?
4. Compare as assimetrias das proporções de pessoas brancas e de pessoas pardas. O que elas representam?



# Referências

- BUGNI, R. P., JACOB, M. S. 2017. Índice de vulnerabilidade social: uma análise da cidade de São Paulo. Disponível em:  
[http://www.ipea.gov.br/agencia/images/stories/PDFs/livros/livros/170828\\_livro\\_territorios\\_numeros\\_insumos\\_politicas\\_publicas\\_2\\_cap04.pdf](http://www.ipea.gov.br/agencia/images/stories/PDFs/livros/livros/170828_livro_territorios_numeros_insumos_politicas_publicas_2_cap04.pdf)
- GALVANI, E. Estatística descritiva em sala de aula. In: VENTURI, L. A. B. Geografia: Práticas de campo, laboratório e sala de aula. São Paulo: Editora Sarandi, 2011.
- GUIMARÃES, I. A. Estatística I (Notas de aulas). 1997. Disponível em:  
[http://www.cin.ufpe.br/~rosf/public\\_html/Notas%20de%20Aula%20de%20Estat%EDstica%20I.doc](http://www.cin.ufpe.br/~rosf/public_html/Notas%20de%20Aula%20de%20Estat%EDstica%20I.doc)
- LUCHIARI, A. 2010. Apontamentos da aula de Cartografia Temática.
- MARTINELLI, M. Gráficos e Mapas: Construa-os Você Mesmo. São Paulo: Moderna, 1998. 120 p.
- PREVIDELLI, I. Bioestatística, 2018. Disponível em:  
<https://biostatistics-uem.github.io/Bio/descritiva.html>
- ROGERSON, P. A. Métodos estatísticos para Geografia: um guia para o estudante. Porto Alegre: Bookman, 2012.