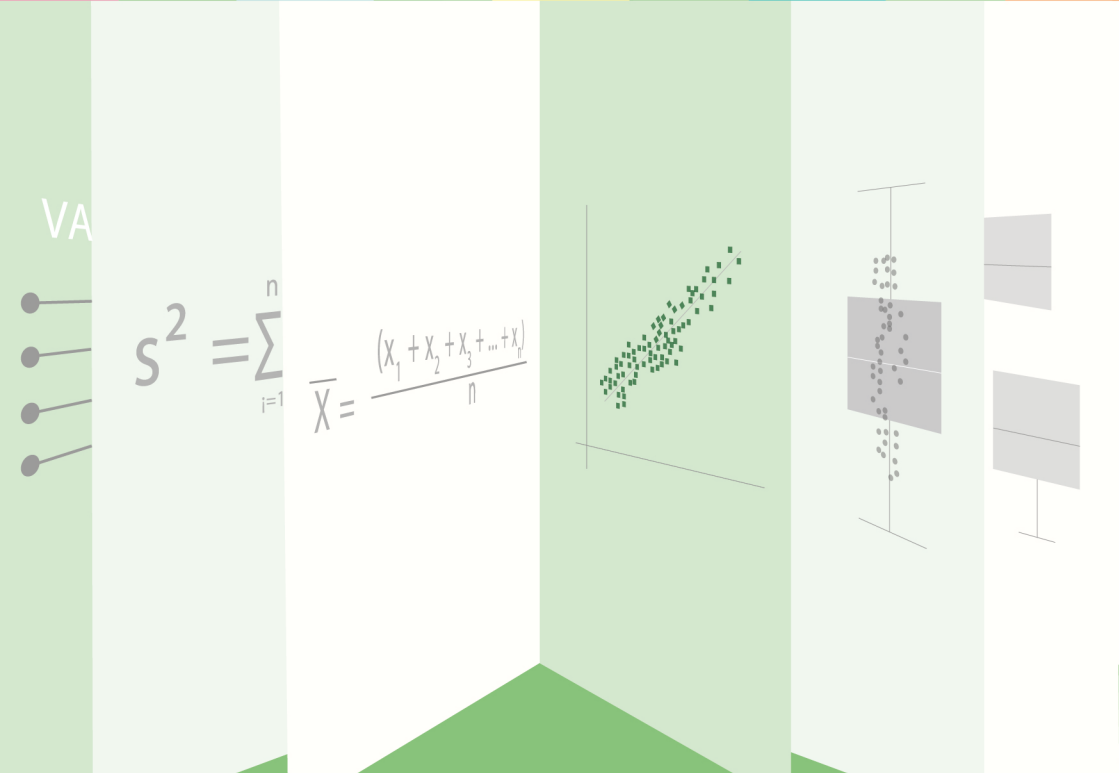




# MEDIDAS DE POSIÇÃO E DISPERSÃO



**Roteiro, Adaptação e Revisão:**

Maria Rita Marques de Oliveira

Karina Rubia Nunes

Guilherme Cardoso Contini

Vitor Marchi Moreno Dias

**Desenvolvimento do Conteúdo:**

Bethina da Rocha Camargo

Rogério Antonio de Oliveira

**Ilustrações:**

Giulia Marques Ranzini

**Design Gráfico e Programação Visual:**

Milton Nakata Studio

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.  
DIVISÃO DE BIBLIOTECA E DOCUMENTAÇÃO - CAMPUS DE BOTUCATU - UNESP  
BIBLIOTECÁRIA RESPONSÁVEL: **ROSEMEIRE APARECIDA VICENTE - CRB 8/5651**

EBAPOPOP : medidas de posição e dispersão / Roteiro, Adaptação e  
Revisão Maria Rita Marques de Oliveira ... [et al]. - Botucatu :  
UNESP/INTERSSAN, 2020  
ePub

Inclui bibliografia

Disponível em: <http://www.redesans.com.br>

ISBN: 978-65-86433-14-2

1. Estatística básica. 2. Matemática. 3. Bioestatística. 4. Medidas de  
posição. 5. Medidas de dispersão. 6. Políticas públicas. 7. Indicadores  
sociais. 8. Exercícios. I. Título. II. Oliveira, Maria Rita Marques de.  
III. Estatística Básica para Políticas Públicas. IV. Universidade Estadual  
Paulista "Júlio de Mesquita Filho". V. Centro de Ciência, Tecnologia e  
Inovação para Soberania e Segurança Alimentar e Nutricional da  
UNESP.

CDD 519.5

# Sumário

Lista de Símbolos .....	4
Lista de Abreviaturas .....	5
Assim os dados são transformados.....	6
O que você precisa saber para compreender melhor este e-book .....	7
<b>Capítulo 1</b> .....	8
Símbolos matemáticos .....	10
Medidas de tendência central .....	11
Exercício .....	18
<b>Capítulo 2</b> .....	19
Medidas de dispersão .....	21
Variância amostral .....	24
Desvio Padrão amostral (s) .....	26
Coeficiente de variação .....	27
Exercício .....	28
<b>Capítulo 3</b> .....	29
Medidas de posição.....	32
Medidas de dispersão .....	35
Exercício .....	39
<b>Capítulo 4</b> .....	40
Diagrama de caixa (Box-plot).....	42
Exercício .....	46
<b>Gabarito Exercícios</b> .....	47
<b>Para saber mais</b> ... ..	51

# Lista de Símbolos

$x_i$  : i-ésimo valor da observação

$\Sigma$  : Letra do alfabeto Grego, sigma. Lê-se: “somatório de”

$\bar{x}$  : Média

$Q_1$  : Primeiro quartil

$Q_2$  : Segundo quartil

$Q_3$  : Terceiro quartil

$s^2$  : Variância amostral

$s$  : Desvio padrão amostral



# Lista de Abreviaturas

**TAC** : Taxa de Atualização Cadastral

**TAFE** : Taxa de Acompanhamento da Frequência Escolar

**TAAS** : Taxa de Acompanhamento da Agenda de Saúde

**CV** : Coeficiente de variação

# Assim os dados são transformados...

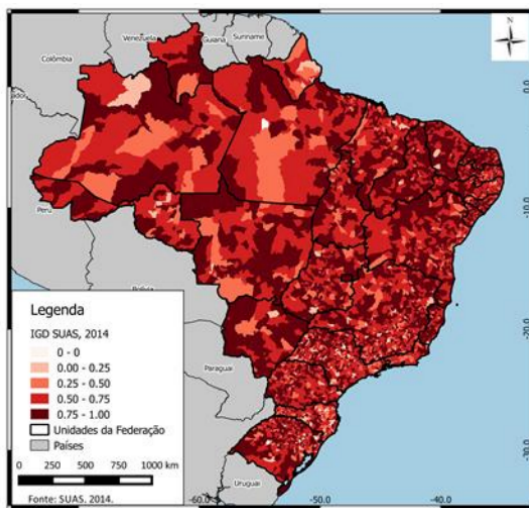


Figura 1: Distribuição espacial do IGD-Suas:Brasil - por municípios, 2014.

[Fonte: reprodução dos autores SATYRO; CUNHA; CAMPOS, 2016.](#)

Veja esse mapa, nele estão georreferenciados todos os municípios com suas respectivas taxas de execução de IGD-Suas. Acompanhando a cor da legenda é possível identificar os municípios conforme a Taxa de execução do IGD-Suas. No entanto, o banco de dados que gerou esse mapa pode ser transformado para se obter muitas outras informações, tais como: média geral ou regional, dispersão dos dados, a distribuição do IGD-Suas em quartis, entre outras. Os conhecimentos obtidos nesse e-book poderá auxiliar sua criatividade em olhar esses dados sob outros ângulos de análise.

Nesse e-book você irá aprender a interpretar, calcular e aplicar as diversas medidas de posição e dispersão, para um banco de dados ou uma amostra específica. Ao final de cada capítulo, você terá aprendido: *Capítulo 1* - Interpretar e calcular média, mediana e moda para uma amostra; *Capítulo 2* - Interpretar e calcular amplitude, quartil, coeficiente de variação; distância interquartílica, variância e desvio padrão dos dados; *Capítulo 3* - Aprimorar suas habilidades em aplicar as medidas de posição e dispersão em planilha eletrônica; *Capítulo 4* - construir Diagramas de Caixa.

# O que você precisa saber para compreender melhor este e-book

## Distribuição simétrica e assimétrica

A distribuição de um conjunto de dados pode ser:

**Simétrica:** se a há uma repartição exata do conjunto de dados em torno do ponto central, dessa forma, as medidas centrais: a média, mediana e moda são coincidentes.

**Assimétrica:** se não há simetria entre o conjunto de dados, os dados apresentam um deslocamento que pode ser à direita ou à esquerda.

**Figura:** Distribuição simetria e assimetria à esquerda ou à direita para um conjunto de dados.

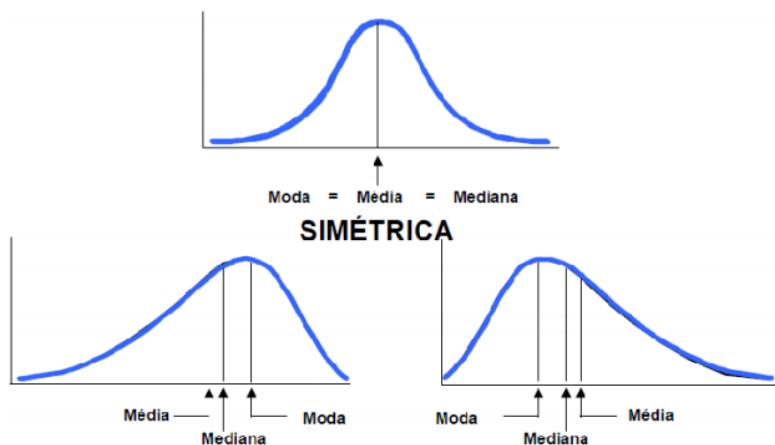


Figura: Distribuição simetria e assimetria à esquerda ou à direita para um conjunto de dados.

## Aproximação decimal

Considere um resultado numérico com muitas casas decimais, neste caso, podemos realizar um arredondamento, ou seja, uma aproximação decimal. Geralmente utilizamos duas ou três casas decimais para reportar valores numéricos. Por exemplo, considere o interesse em arredondar o número de 3 casas decimais 6,185 para apenas duas casas. Note que último número da casa decimal é igual a 5, logo temos que “arredondar para cima” toda vez que o número for maior ou igual a 5, ou seja, teremos 6,19. Agora se tivermos 6,183, então a 3ª. casa decimal é igual a 3 (menor que 5), logo podemos “arredondar para baixo”, ou seja, mantemos a 2ª casa decimal igual e obtemos o valor 6, 18.

# Capítulo 1

Analise as figuras seguintes: elas retratam o indicador de “Renda per Capita em U\$ por dia”. Temos todos os países com informações disponíveis, o tamanho das faixas representa a quantidade da população e os continentes possuem cores distintas.

Na década de 70 (Figura 2), aproximadamente 51% das pessoas no mundo viviam em extrema pobreza, ou seja, com menos de 2 dólares por dia, sendo a maioria residente da Ásia e na África.

Quando comparamos esse mesmo indicador, após 40 anos (Figura 3), percebemos o deslocamento da maioria da população que vivia com apenas 1 dólar por dia para a porção central, próxima de 10 dólares por dia.

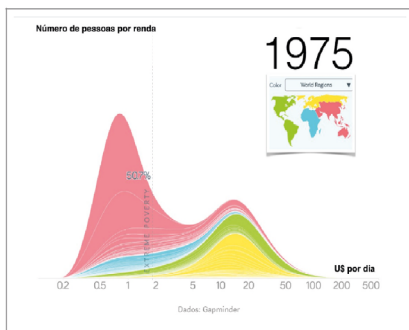


Figura 2: Renda per Capita em U\$ por dia, 1975

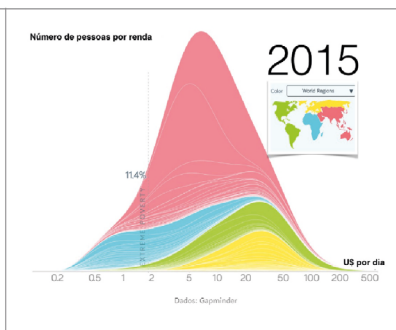


Figura 3: Renda per Capita em U\$ por dia, 2015.

Essas imagens evidenciam o deslocamento de uma população ao longo do tempo bem como a importância em se compreender medida de tendência central, quando trabalhamos com um conjunto grande de dados e queremos compreender como eles se comportam em diferentes momentos do tempo.

Nesse capítulo você irá conhecer as diversas medidas de posição, denominada também por medida de tendência central e será capaz de Interpretar as principais medidas de posição: média, mediana e moda.



# Símbolos matemáticos

Primeiramente, considere os valores coletados de uma amostra de tamanho  $n$ , representada por:

$$X_1, X_2, X_3, ..., X_i, ..., X_n$$

O subscrito indica a posição do valor na sequência das coletas. Portanto  $x_1$  representa o primeiro valor observado,  $x_2$  o segundo valor e assim por diante, o valor  $x_i$  é o  $i$ -ésimo valor no conjunto de  $x_n$  valores.

## Exemplo1:

Na Tabela 1 é apresentado a ordem dos valores coletados e, na última coluna, estão representandos os elementos que  $x_i$  pode assumir em relação às outras tabelas.

Tabela 1: Informações do Cadastro Único, Município Cachoeira do Norte, 2020.

Ordem	Nome	NIS	Idade do pai	Posição do valor
1	José	3975481380	18	$x_1$
2	Maria	1234567890	21	$x_2$
3	Matilde	7894561238	26	$x_3$
4	Larissa	2458779325	27	$x_4$
...	...	...	...	...
200	João	1543679852	65	$x_n$

Para indicar a soma dos  $n$  valores observados na amostra , usamos a letra grega  $\Sigma$  (sigma, em versão maiúscula), que deve serlida como“somatório de” e escrevemos:

$$\sum_{i=1}^n x_i = x_1+x_2 + ... + x_n$$

Obs: podemos ler o somatório acima como: soma dos valores observados  $x_i$ , variando do primeiro valor observado  $x_1$  até o último valor representado por  $x_n$

# Medidas de tendência central

Uma medida de tendência central é um ponto de equilíbrio dos dados, “ponto do meio”, sendo uma medida resumo do centro dos dados em torno do qual estes se distribuem. As medidas de tendência central dos dados mais comuns são: *média*, *mediana* e *moda*.

## Média

A média aritmética (média simples ou  $\bar{x}$ ) é a medida de tendência central mais conhecida e utilizada para resumir a informação contida em um conjunto de dados. Pode ser obtida por meio da soma de todos os dados e dividindo o resultado pela quantidade de valores observados.

A fórmula para o cálculo da média é dada por:

$$\text{Média} = \frac{\text{Soma de todos os dados}}{\text{Número de dados}} = \bar{x} = \frac{\sum_{i=1}^n x_i}{n}.$$

A média possui algumas características, como:

- É afetada quando os dados apresentam valores discrepantes, ou seja, valores muito altos, puxam o valor da média para cima, assim como valores discrepantes baixos podem puxar o valor da média para baixo;
- Quando possuímos distribuição simétrica, a média é bastante utilizada;
- Para variáveis categóricas, é não utilizado o cálculo da média, devido a natureza da variável estar associada as características observadas serem expressas por meio de palavras;
- Pode ser utilizada para algumas variáveis discretas, admitindo resultados finais com casas decimais.



### Exemplo 2:

A idade de seis crianças que estão na fila para entrar no espaço amigo, são: 7, 10, 9, 8, 10 e 12 anos. Calcule a média da idade dessas crianças.

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{7 + 10 + 9 + 8 + 10 + 12}{6} = \frac{56}{6} = 9,33.$$

A idade média das crianças na fila de espera para uma vaga do espaço amigo é de 9,33 anos. Embora a idade possa ser considerada uma variável inteira, é admitida a representação numérica com casas decimais.

Agora vamos fazer um exercício utilizando um indicador que você utiliza no seu dia a dia de trabalho na assistência social, o Índice de Gestão Descentralizada Municipal. Se tive dúvidas acesso o manual do IGD-SUAS ([http://www.mds.gov.br/webarquivos/publicacao/bolsa\\_familia/Guias\\_Manuais/ManualIGD.pdf](http://www.mds.gov.br/webarquivos/publicacao/bolsa_familia/Guias_Manuais/ManualIGD.pdf)).

### Fórmula do Índice de Gestão Descentralizada – Municipal:

$\text{IGD - M} = \text{Fator I} \times \text{Fator II} \times \text{Fator III} \times \text{Fator IV}$
---

Para relembrar as variáveis e também para guiar a construção do exercício, utilize a [tabela 2](#) (para o cálculo do Fator I) e a [tabela 3](#) (cálculo dos fatores II, III, IV), com as informações do município de Cachoeira do Norte.

Tabela 2: Informações para cálculo do Fator I, IGD-M Cachoeira do Norte, EBAPOP, 2020.

<b>Fator I</b> : média aritmética simples das seguintes variáveis (a+b/2).			
<b>a)</b>	<b>TAC</b> – Taxa de Atualização Cadastral	Total de cadastros válidos de famílias com renda per capita, até meio salário mínimo atualizados nos últimos dois anos no Cadastro Único do município	9800
		Cadastros de famílias com renda per capita, até meio salário mínimo no Cadastro Único no município	12500
<b>b)</b>	<b>TAFE</b> – Taxa de Acompanhamento da Frequência Escolar	Número de crianças e adolescentes pertencentes às famílias beneficiárias do PBF com perfil educação no município e com informações de frequência escolar	5200
<b>TAFE + TAAS/2</b>		Número total de crianças e adolescentes pertencentes a famílias beneficiárias do PBF com perfil educação no município	7850
	<b>TASS</b> – Taxa de Acompanhamento da Agenda de Saúde	Público com perfil saúde no município e com informações de acompanhamento de condicionalidade de saúde	5630
		Total do público com perfil saúde no município	7050

Tabela 3: Informações para cálculo dos Fatores II,III,IV IGD-M Cachoeira do Norte, EBAPOP, 2020.

<b>Fatores II, III e IV:</b> são atribuídos os seguintes valores.		
	<b>VALOR = 0</b>	<b>VALOR = 1</b>
<b>Fator II</b>	O município não tiver aderido ao SUAS	O município tiver aderido ao SUAS
<b>Fator III</b>	O município não tiver informado no SuasWeb, até as datas estipuladas pela Portaria GM/MDS nº 103, de 30 de setembro de 20137, a apresentação da comprovação de gastos dos recursos do IGD-M ao respectivo CMAS	O município tiver informado, no SuasWeb, a apresentação da comprovação de gastos dos recursos do IGD-M ao respectivo CMAS
<b>Fator IV</b>	O CMAS <sup>1</sup> não tiver informado no SuasWeb <sup>2</sup> , até as datas estipuladas pela Portaria GM/MDS nº 103, de 30 de setembro de 20138, a aprovação total da comprovação de gastos dos recursos transferidos	O CMAS tiver informado, no SuasWeb, a aprovação total da comprovação de gastos dos recursos transferidos

1

CMAS (Conselho Municipal de Assistência Social) e 2SuasWeb (Um dos aplicativos da Rede SUAS. Ele possibilita a fiscalização das ações e a obrigatoriedade do acompanhamento tanto do planejamento estadual e municipal da área, como da execução física e financeira da prestação de contas.

Vamos então começar o exercício:

Para o cálculo do Fator I - Fator de operação utilizaremos a seguinte fórmula:

$$\frac{\text{TAC taxa de atualização cadastral} + \left( \frac{\text{TAFE Taxa de Acompanhamento da Frequência Escolar} + \text{TAAS Taxa de Acompanhamento da Agenda de Saúde}}{2} \right)}{2}$$

Vamos escrever a fórmula de cálculo para o município Cachoeira do Norte utilizando as informações da tabela:

TAC - Taxa de Atualização Cadastral	Total de cadastros válidos de famílias com renda per capita, até meio salário mínimo atualizados nos últimos dois anos no Cadastro Único no município	9800	=0,78
	Total de cadastros de famílias com renda per capita, até meio salário mínimo no Cadastro Único no município	12500	
TAFE - Taxa de Acompanhamento da Frequência Escolar	Número de crianças e adolescentes pertencentes às famílias beneficiárias do PBF com perfil educação no município e com informações de frequência escolar	5200	=0,66
	Número total de crianças e adolescentes pertencentes a famílias beneficiárias do PBF com perfil educação no município	7850	
TAAS - Taxa de Acompanhamento da Agenda de Saúde	Público com perfil saúde no município e com informações de acompanhamento de condicionalidade de saúde	5630	=0,79
	Total do público com perfil saúde no município	7050	

Calculando o Fator I - IGD-M:

$$\text{Fator I} = \frac{\text{TAC taxa de atualização cadastral} + \left( \frac{\text{TAFE Taxa de Acompanhamento da Frequência Escolar} + \text{TAAS Taxa de Acompanhamento da Agenda de Saúde}}{2} \right)}{2}$$

$$\text{Fator I} = \frac{0,78 + \left( \frac{0,66 + 0,79}{2} \right)}{2}$$

$$\text{Fator I} = 0,75$$

Agora que temos o cálculo do Fator 1 vamos calcular o indicador. Considere que o município aderiu ao SUAS, alimentou o SuasWeb com a comprovação dos gastos e o CMAS informou SuasWeb com a comprovação de gastos de recursos transferidos. Portanto possui valor = 1 para os fatores II,III,IV

$$\text{IGD - M} = \text{Fator I} \times \text{Fator II} \times \text{Fator III} \times \text{Fator IV}$$

$$\begin{aligned}\text{IGD - M} &= 0,75 \times 1 \times 1 \times 1 \\ \text{IGD - M} &= 0,75\end{aligned}$$

## Mediana

A mediana é o valor que ocupa a posição central de um conjunto de dados ordenados, separando-os em duas metades.

A mediana possui algumas características, como:

- Para variáveis categóricas não é utilizável;
- Valores discrepantes não afetam a mediana;
- Mais utilizado e indicado para dados com distribuições assimétricas.

Para o caso em que o conjunto de dados analisado é composto por um número par de elementos, há dois valores na posição central e assim, é necessário fazer a média desses dois valores para encontrar a mediana. No caso em que a quantidade de elementos de um conjunto de dados é ímpar, encontra-se apenas um valor na posição central, sendo o valor mediano dos dados. Vejamos esses dois casos no exemplo 3 abaixo.

### Exemplo 3:

Se o número de dados é ímpar, existe um único valor na posição central.

$$3, \mathbf{5} \text{ e } 9 - \text{Mediana} = 5$$

Se o número de dados é par, existem dois valores na posição central. Então, a mediana é a média desses dois valores.

$$3, \mathbf{5, 7} \text{ e } 9 - \text{Mediana} = \frac{5+7}{2} = \frac{12}{2} = 6$$

## Moda

A moda é o valor que ocorre com maior frequência em um conjunto de dados, ou seja, é o valor que mais se repete ou que mais aparece no conjunto de dados.

A moda possui algumas características, como:

- O conjunto de dados pode ser plurimodal, ou seja, ter mais de uma moda;
- Para as variáveis numéricas e categóricas a moda pode ser calculada;
- O conjunto de dados pode ser amodal, ou seja, não possuir nenhuma moda.

### Exemplo 4:

Para os conjuntos de dados abaixo, encontre a moda:

0, 2, 4, 6, 8, 10 - não tem moda, pois nenhum valor se repete;

1, **2, 2**, 3, **4, 4**, 5, 6, 7 - em duas modas representadas pelos números: 2 e 4.

Exemplo 5:

A Tabela 4 apresenta as idades dos responsáveis pela família na cidade de Cachoeira do Norte no ano de 2020 para uma amostra composta por 6 famílias. Para estudar melhor a distribuição dos dados, podem ser calculadas as medidas de tendência central: *média*, *mediana* e *moda*:

Tabela 4: Idade do responsável pela família, Cachoeira do Norte, 2020.

Ordem	Idade
1	90
2	70
3	65
4	90
5	90
6	30

Média  $\frac{90 + 70 + 65 + 90 + 90 + 30}{6} = 72,50;$

Ordenando os dados: 30, 65, **70, 90**, 90 e 90

Mediana (30, 65, **70, 90**, 90, 90)  $\frac{70 + 90}{2} = \frac{160}{2} = 80$

Moda: 90 (apareceu três vezes = 30, 65, 70, **90, 90** e **90**).

Como você pode perceber, *média*, *mediana* e *moda* podem ter valores bem diferentes uns dos outros. A média é importante quando queremos encontrar o ponto médio do valor total da amostra (a soma), mas não é ideal quando queremos encontrar o valor central entre os valores que compõem uma amostra, aquele valor que divide a amostra em duas partes iguais. Para isso, o recomendado é usar a mediana. Já o valor mais frequente numa amostra não é indicado pela média nem pela mediana, mas pela moda. Por exemplo, se queremos a idade média da amostra, o valor é 72,50. Se queremos saber qual o valor que separa a metade mais velha da metade mais nova dos elementos da amostra, esse valor é 80. E se queremos saber qual valor que mais se repete, o número é 90.

**Exercícios** [\(ver gabarito no final do e-book\)](#)

1 - A Tabela 5 apresenta a renda familiar de 14 famílias cadastradas no Cadastro Único, no município Cachoeira do Norte. Organize os dados com as medidas de posição que conheceu aqui e escreva suas observações sobre o comportamento dos dados.

Tabela 5: Renda Familiar de 14 cadastros do Cadastro único, município, Cachoeira do Norte, 2020.

Renda Familiar	
R\$ 320,00	R\$ 250,00
R\$ 480,00	R\$ 480,00
R\$ 640,00	R\$ 510,00
R\$ 710,00	R\$ 250,00
R\$ 250,00	R\$ 480,00
R\$ 320,00	R\$ 640,00
R\$ 800,00	R\$ 250,00



# Capítulo 2

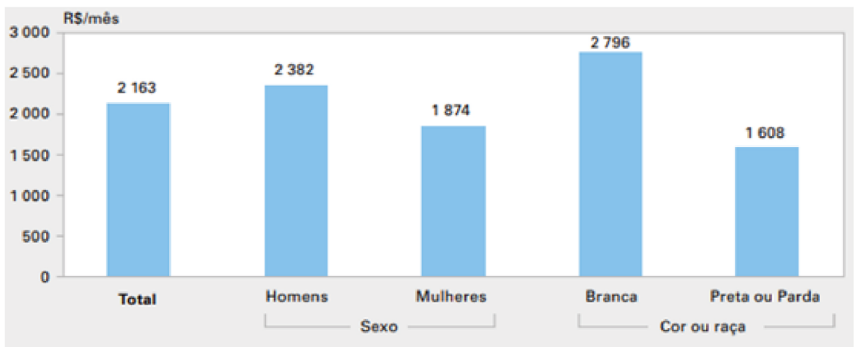


Figura 4: Rendimento médio real do trabalho principal das pessoas ocupadas, segundo o sexo e a cor ou raça - Brasil - 2018

Fonte: IBGE - Pesquisa Nacional por Amostragem de Domicílio, 2018.

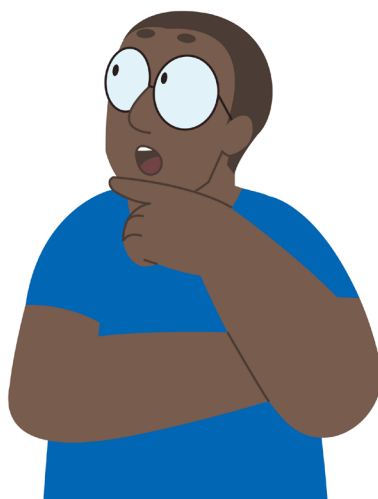
O gráfico acima é da Pesquisa Nacional de Domicílio - PNAD, realizada pelo Instituto Brasileiro de Geografia e Estatística (IBGE), e ilustra a média do rendimento real do trabalho principal das pessoas ocupadas no Brasil. Você parou para pensar quantas pessoas foram entrevistadas e o tamanho da planilha de dados que foi utilizada para a construção desse gráfico? Em 2018, nessa pesquisa, o IBGE visitou 211.344 domicílios a cada trimestre, o que nos faz inferir que a planilha para construção desse gráfico tem mais de um milhão de dados

Será que as médias falam tudo sobre esses dados? Qual a distância entre o que ganha mais e o que ganha menos? Qual o valor mínimo e o valor máximo? Será que se analisarmos a metade que ganha mais e a metade que ganha menos vamos encontrar um mesmo perfil de distribuição dos dados?

O recorte de análise desse gráfico procurou evidenciar as iniquidades do trabalho no Brasil, onde mulheres, negros e pardos ganham menos, mas poderíamos aprofundar a análise e construir classes de renda dentro de cada um desses grupos. Por exemplo, poderíamos distribuir o rendimento das mulheres em quatro classes, das que ganham menos às que ganham mais. Estudando esse capítulo você aprenderá como elaborar análise dessa natureza.



No Capítulo 2, você aprenderá como interpretar e calcular as medidas de amplitude, quartil, distância interquartilica, variância, desvio padrão e coeficiente de variação dos dados, também conhecidas como medidas de dispersão.



# ***Medidas de dispersão***

Para descrever o conjunto de dados, além das medidas de tendência central, pode-se utilizar medidas de variabilidade, também chamadas de medidas de dispersão.

As medidas de dispersão são utilizadas para avaliar o quanto cada observação do conjunto de dados varia em relação a sua média, ou seja, as medidas de dispersão calculam a distância de uma determinada observação em relação à média desse conjunto de dados. Dessa forma, podemos saber se a distribuição dos dados apresenta uma variabilidade baixa, moderada ou alta, ajudando a entender o comportamento dos dados.

## **Amplitude**

A amplitude é a medida de variação mais simples, é definida pela diferença entre o valor máximo e o valor mínimo de um conjunto de dados.

Primeiramente, para calcular a amplitude de um conjunto de dados é necessário entender o conceito de valor máximo e valor mínimo.

- Valor mínimo:é o menor valor de todas as observações do conjunto de dados;

- Valor máximo: é o maior valor de todas as observações do conjunto de dados.

Então a amplitude é calculada pela fórmula:

$$Amplitude = X_{maior} - X_{menor}$$

Exemplo 6:

Na Tabela 6, encontre o valor máximo, mínimo e calcule a amplitude.

Tabela 6: Idade do responsável pela família, Cachoeira do Norte, 2020.

Ordem	Idade
1	18
2	22
3	34
4	42
5	48
6	50

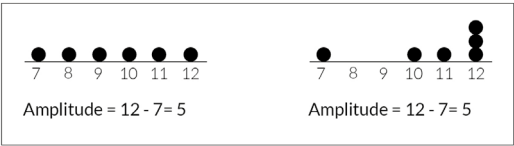
Valor máximo: 50

Valor mínimo: 18

Amplitude =  $X_{maior} - X_{menor} = 50 - 18 = 32$

A amplitude possui algumas desvantagens:

- Ignorar a forma como os dados estão distribuídos. Dois conjuntos de dados com variabilidade muito diferentes podem ter a mesma amplitude. Como mostrado abaixo:



- Sensível a valores extremos, pois para calculá-los apenas se usa os dois valores extremos, como apresentado a seguir:

**1,1,1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,3,3,3,3,4,5**  
Amplitude = 5 - 1 = 4

**1,1,1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,3,3,3,3,4,120**  
Amplitude = 120 - 1 = 119

## Quartil

Para encontrar a medida quartil é necessário: encontrar a mediana que irá dividir o conjunto de dados ordenados em dois subconjuntos com a mesma quantidade de dados, assim:

- O que antecede a mediana (dados iguais ou menores do que a mediana), seria um conjunto de dados.
- O que sucede a mediana (dados iguais ou maiores do que a Mediana), outro conjunto de dados.

O conjunto de dados ordenados poderá ser dividido não apenas em duas metades, mas em quatro quartos. A mediana divide o conjunto de dados em duas partes e os Quartis (como o nome sugere) divide o conjunto de dados em quatro partes.

Para obter quartis:

1. Organize os dados em ordem crescente, isto é, do menor para o maior. Encontre a mediana (que é também o segundo quartil -  $Q_2$  ).
2. Encontre o primeiro quartil, da seguinte forma: tome o conjunto de dados a esquerda da mediana; o primeiro quartil ( $Q_1$ ) é a mediana do novo conjunto de dados.
3. Encontre o terceiro quartil, da seguinte forma: tome o conjunto de dados a direita da mediana; o terceiro quartil ( $Q_3$ ) é a mediana do novo conjunto de dados.

### Exemplo 7:

Considere o conjunto de dados abaixo e encontre os quartis.

1, 2, 3, 4, **5**, 5, 7, 9, 10; Mediana ( $Q_2$ ) = 5.

Para obter o primeiro quartil ( $Q_1$ ), separamos os dados a esquerda da mediana.

$$1, \mathbf{2, 3}, 4; \quad Q_1 = \frac{2+3}{2} = \frac{5}{2} = 2,5$$

Para obter o terceiro quartil ( $Q_3$ ), separamos os dados a direita da mediana.

$$5, \mathbf{7, 9}, 10; \quad Q_3 = \frac{7+9}{2} = \frac{16}{2} = 8$$

### Distância interquartílica ( $Q_3 - Q_1$ )

A distância interquartílica é a distância entre o primeiro e o terceiro quartil.

$$\text{Distância interquartílica} = (Q_3 - Q_1)$$

### Exemplo 8:

Para os dados do exemplo 7, calcule a distância interquartílica.

$$1, 2, 3, 4, 5, 5, \mathbf{7, 9}, 10; \quad Q_1 = 2,5 \text{ e } Q_3 = 8$$

$$\text{Distância interquartílica} = 8 - 2,5 = 5,5.$$

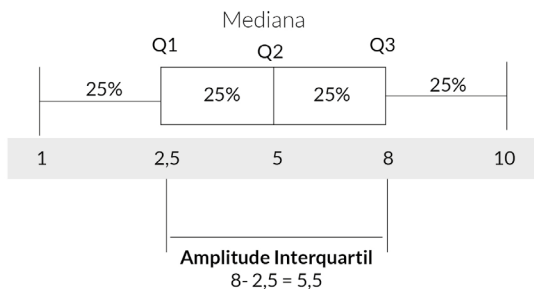


Figura 5: ilustração dos exemplos 7 e 8.

Obs: Para conjunto de dados com valores discrepantes, recomenda-se calcular a mediana, a amplitude e a distância interquartílica juntos. Já que a amplitude é muito sensível aos valores discrepantes.

## ***Variância amostral***

A variância amostral é mais recomendada quando a média é utilizada como medida de tendência central, ou seja, quando a média indica o centro da distribuição dos dados. Pode-se calcular a variância a partir da somatória das diferenças de cada observação, subtraindo-se o valor da média.

$$\text{Diferença} = \text{observação} - \text{média}$$

$$\text{Diferença} = x_i - \bar{x}$$

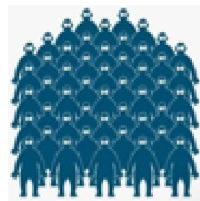
Se a diferença em relação à média é pequena, podemos concluir que a observação ( $x_i$ ) é próxima da média, caso contrário, a observação está afastada da média.

A variância amostral é a soma dos quadrados dos desvios de cada observação em relação à média, dividida por  $(n-1)$ , conforme a fórmula:

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

### **Exemplo 9:**

As idades das crianças da sala de atividades bolinha, são: 2, 4, 7, 8, 9, 10 e 8. Calcule a variância.



Primeiramente, precisamos calcular a média:

$$\text{Média} = \frac{2 + 4 + 7 + 8 + 9 + 10 + 8}{7} = \frac{48}{7} = 6,857; \bar{x} = 6,86$$

Calculando a variância:

$$s^2 = \frac{(2-6,8)^2 + (4-6,8)^2 + (7-6,8)^2 + (8-6,8)^2 + (9-6,8)^2 + (10-6,8)^2 + (8-6,8)^2}{7-1}$$

$$s^2 = \frac{23,04 + 7,84 + 0,04 + 1,44 + 4,84 + 10,24 + 1,44}{6}$$

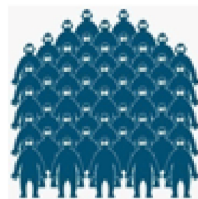
$$s^2 = \frac{48,88}{6} = 8,1466$$

$$s^2 = 8,15$$

Para a sala de atividades bolinha, pode-se observar que as crianças possuem várias idades e por isso a variância deu um valor mais alto, ou seja, as idades estão afastadas da média.

### Exemplo 10:

As idades das crianças da sala de atividades estrelinha, são: 3,3,3,3,2,3 e 2. Calcule a variância.



Primeiramente, precisamos calcular a média:

$$\text{Média} = \frac{3 + 3 + 3 + 3 + 2 + 3 + 2}{7} = \frac{19}{7} = 2,7142; \bar{x} = 2,71$$

Calculando a variância:

$$s^2 = \frac{(3-2,7)^2 + (3-2,7)^2 + (3-2,7)^2 + (3-2,7)^2 + (2-2,7)^2 + (3-2,7)^2 + (2-2,7)^2}{7-1}$$

$$s^2 = \frac{0,09 + 0,09 + 0,09 + 0,09 + 0,49 + 0,09 + 0,49}{6}$$

$$s^2 = \frac{1,43}{6} = 0,2383$$



Para a sala de atividades estrelinha as crianças possuem idades próximas, por isso a variância foi baixa, ou seja, as idades estão próximas à média.

## ***Desvio padrão amostral (s)***

O desvio padrão é uma medida de variabilidade muito usada para representar a dispersão dos dados. Em muitos artigos, é usada a simplificação DP para representar o desvio padrão. O Desvio padrão é a raiz quadrada da variância, com sinal positivo, dado pela fórmula:

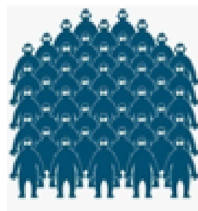
$$s = \sqrt{\text{variância}} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}.$$

### **Exemplo 11:**

Para os exemplos 10 e 9 acima, calcular o desvio padrão

**Sala de atividades bolinha:**  $s = \sqrt{8,15} = 2,85$

**Sala de atividades estrelinha:**  $s = \sqrt{0,24} = 0,49$



Uma maneira comum de expressar o desvio padrão (s) é juntamente com a média:

**Sala de atividades bolinha:**  $\bar{x}$  (média)  $\pm s = 6,7 \pm 2,85$  anos  
(lê-se: média de 6,7 e  $\pm 2,85$  anos de desvio).

**Sala de atividades estrelinha:**  $\bar{x}$  (média)  $\pm s = 2,7 \pm 0,49$  anos.

## Coeficiente de variação

O coeficiente de variação é usado para analisar a dispersão em relação ao seu valor médio (média), quando temos dois ou mais conjuntos de dados. Portanto, o CV é uma forma de expressar a variabilidade dos dados e costuma ser dado em porcentagem (%).

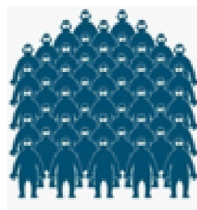
$$CV = \frac{\text{desvio padrão}}{\text{média}} = \frac{s}{\bar{x}} * 100\%$$

### Exemplo 12:

Para os exemplos 10 e 9, calcule o CV.

Sala de atividades bolinha:

$$CV = \frac{2,85}{6,85} = 0,41605 * 100\% = 41,61 \%$$



Sala de atividades estrelinha:

$$CV = \frac{0,48}{2,71} = 0,17712 * 100\% = 17,71 \%$$

Normalmente, o coeficiente de variação é classificado segundo a dispersão dos dados da seguinte forma:

- $CV < 15\%$  = indica baixa dispersão dos dados em torno da média (baixa variabilidade)
- $15\% \leq CV < 30\%$  = indica dispersão moderada (variabilidade moderada).
- $CV \geq 30\%$  = indica alta dispersão dos dados em torno da média (alta variabilidade).

Para o exemplo 12, a sala de atividades bolinha teve  $CV = 41,61\%$ , indicando alta variabilidade dos dados. Já a sala de atividades estrelinha, com  $CV = 17,71\%$ , indicou que seus dados apresentam uma variabilidade moderada.

## Exercício [\(ver gabarito no final do e-book\)](#)

1) Continuando o exercício resolvido do [Capítulo 1, Tabela 5 \(página 18\)](#), calcule as medidas de dispersão.

Tabela 5: Renda Familiar de 14 cadastros do Cadastro único, município, Cachoeira do Norte, 2020.

Renda Familiar	
R\$ 320,00	R\$ 250,00
R\$ 480,00	R\$ 480,00
R\$ 640,00	R\$ 510,00
R\$ 710,00	R\$ 250,00
R\$ 250,00	R\$ 480,00
R\$ 320,00	R\$ 640,00
R\$ 800,00	R\$ 250,00

# Capítulo 3

## Evolução no processamento das informações




Valor	1	10	100	1.000	10.000	100.000	1 milhão, ou mais
Hieróglifo	I	∩	9	✦	✦	 OU 	
Descrição	Corda simples ou Basão Calcanhar Corda enrolada Flor de De Lótus Dedo Indicador Girino ou Sapo						Homem com as mãos enfiadas.

Figura 6: [Assim surgiram os números arábicos. Egito, por volta de 3.000 a.C.](#)

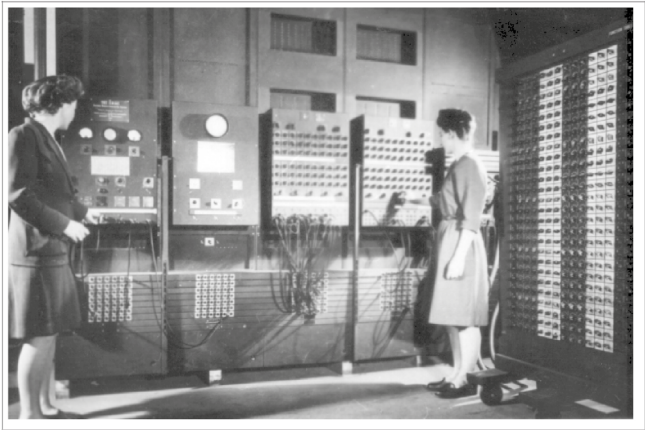


Figura 7: [USA - Forças armadas, ENIAC - Electrical Numerical Integrator and Computer, criado em 1946.](#)



Figura 8 : Smartphones com acesso às redes sociais e muitas outras funções. 2020.

O homem está na terra por volta de 200 milhões de anos. Lançando o olhar para a vida do homem na terra, somente há cerca de 3 mil anos antes de Cristo, a humanidade começou a criar os primeiros sistemas de numeração (Figura 6). Já o espaço de tempo entre a criação do número e a criação do computador foi de aproximadamente cinco mil anos (Figura 7). Setenta anos depois, o computador que ocupava uma sala se transformou em um aparelho que cabe na palma da mão e pode processar uma infinita quantidade de informações (Figura 8).

A tecnologia pode ser utilizada a seu favor, muito daquilo que se fazia manualmente com elevada exigência de raciocínio e atenção pode ser feito por meio do computador, de forma quase automática. Mas não se engane, o raciocínio humano continua sendo fundamental para que não nos tornemos máquinas que reproduzem operações sem saber o que estão fazendo e saibamos aplicar corretamente os resultados obtidos.

*Se você precisar retomar o significado de cada variável acesso o documento Perguntas frequentes do Cadastro Único:*

[https://aplicacoes.mds.gov.br/sagirmips/noticias/arquivos/files/perguntas\\_frequentes\\_dados\\_cadastro\\_unico\\_programas\\_sociais.pdf](https://aplicacoes.mds.gov.br/sagirmips/noticias/arquivos/files/perguntas_frequentes_dados_cadastro_unico_programas_sociais.pdf)

Nesse capítulo, você irá aprender como trabalhar com grandes volumes de dados utilizando um programa de computador. Iremos aplicar medidas de posição e de dispersão em planilhas eletrônicas. Neste caso, optou-se por utilizar o software Microsoft Excel, mas também há outras planilhas eletrônicas e outras versão do Excel disponíveis que realizam os mesmos cálculos, por exemplo, a planilha eletrônica CalcDo LibreOffice(software gratuito) .

Nesse capítulo vamos utilizar uma tabela de Microdados Original do Cadastro Único. Utilizaremos a base de dados do município de Encontro das Águas (Município fictício, identificado na base de Dados do IBGE com o Código 1100015).



Exemplo 13:

Base de dados do Cadastro Único com marcação do Bolsa Família - Micro-dados Dez/2018 - Cadastro Único e PBF (atualizado em 12/11/2019, 14:12).

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	cd_ibge	estrato	classf	id_fam	dat_cadastrament	dat_alteracao	vlr_ren	dat_atualizacao_f	cod_lo	cod_es	qtd_co	qtd_co	cod_mi
72290	1100015	1	3	87891	16/10/2007	30/09/2018	190	12/04/2018	2	1	4	2	4
72293	1100015	1	3	87894	18/08/2013	01/10/2018	83	27/07/2018	1	1	6	3	2
72326	1100015	1	3	87929	12/08/2010	30/09/2018	100	15/01/2018	2	1	5	2	2
72366	1100015	1	3	87975	28/08/2018	02/10/2018	416	28/08/2018	2	1	4	1	5
72383	1100015	1	3	87996	12/01/2018	02/10/2018	533	25/01/2018	2	1	6	2	2
72452	1100015	1	3	88075	31/07/2017	02/08/2017	600	31/07/2017	2	1	3	1	2
72485	1100015	1	3	88115	29/08/2014	25/05/2017	95	23/05/2017	2	1	5	2	2
72589	1100015	1	3	88227	16/05/2007	30/09/2018	41	13/08/2018	2	1	3	2	1
72596	1100015	1	3	88236	27/08/2018	29/08/2018	50	27/08/2018	1	1	3	2	5
72609	1100015	1	3	88251	16/12/2011	23/11/2018	954	23/11/2018	1	1	6	2	5
72629	1100015	1	3	88273	02/08/2018	02/10/2018	954	02/08/2018	1	1	5	2	2
72665	1100015	1	3	88314	11/03/2010	12/12/2018	83	12/12/2018	2	1	4	2	5
72764	1100015	1	3	88424	23/08/2018	02/10/2018	500	23/08/2018	2	1	5	2	2
72805	1100015	1	3	88473	02/10/2018	02/10/2018	954	02/10/2018	2	1	5	4	2
72821	1100015	1	3	88492	13/12/2011	01/10/2018	1431	25/07/2018	1	1	5	2	5
72984	1100015	1	3	88667	03/07/2018	02/10/2018	400	03/07/2018	1	1	3	1	5
72988	1100015	1	3	88671	25/10/2006	30/09/2018	50	28/03/2018	1	1	3	1	5
73000	1100015	1	3	88684	22/10/2018	22/10/2018	233	22/10/2018	2	1	4	2	2
73007	1100015	1	3	88691	20/02/2013	08/10/2018	66	08/10/2018	2	1	4	1	5
73059	1100015	1	3	88752	14/06/2018	02/10/2018	954	14/06/2018	1	1	7	3	5
73139	1100015	1	3	88840	06/10/2017	01/10/2018	318	22/08/2018	1	1	4	2	2

base amostra familia\_201812

361 de 1040575 registros localizados.

Figura 9: Base de dados do Cadastro Único com marcação do Bolsa Família, atualizado em 12/11/2019, 14:12.

Na Figura 9, é possível visualizar que a base de dados é bem extensa, com muitos dados. Dessa forma, realizar as contas manualmente como fizemos nos capítulos anteriores seria praticamente inviável.

Medidas de posição

Para realizar os cálculos das medidas de posição e dispersão da base de dados do Cadastro único, do município Encontro das Águas, é necessário definir e selecionar a variável que será utilizada. A variável utilizada neste caso é vlr\_renda\_media\_fam, composta pelo valor da renda que as famílias recebem, como mostra a Figura 10.

1	cd_ibge	vlr_renda_media_fam	
72290	1100015	190	
72293	1100015	83	
72326	1100015	100	
72366	1100015	416	
72383	1100015	533	
72452	1100015	600	
72485	1100015	95	
72589	1100015	41	
72596	1100015	50	
72609	1100015	954	
72629	1100015	954	
72665	1100015	83	
72764	1100015	500	
72805	1100015	954	
72821	1100015	1431	
72984	1100015	400	
72988	1100015	50	
73000	1100015	233	
73007	1100015	66	
73059	1100015	954	
73139	1100015	318	

Figura 10: Seleção da variável.

Para calcular as medidas de posição e dispersão, é necessário clicar em “Inserir função”, escolher qual cálculo deseja fazer (Fig 11), clicar em ok e selecionar a coluna ou as células que deseja avaliar. Como mostra a Figura 11, a opção “Inserir função” contém todos os tipos de funções que o Excel disponibiliza, não somente matemáticas e estatísticas.

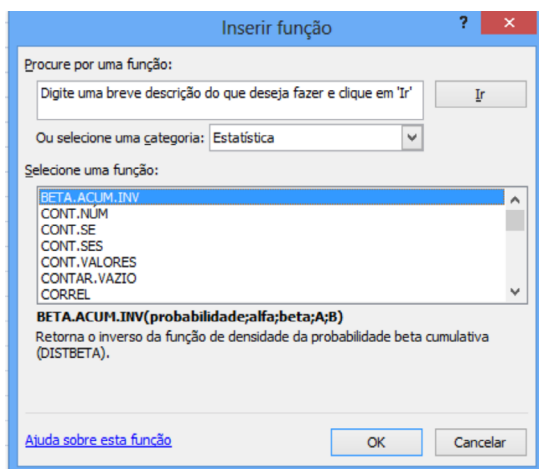


Figura 11: Inserir função desejada.



Outra maneira, para quem já está mais habituado com planilhas eletrônicas e conhece as funções que irá utilizar, é colocar o símbolo de igual (=) em uma célula limpa, escrever a nome da função que deseja, clicar duas vezes na função e depois selecionar o conjunto de dados desejado, como mostra a Figura 12.

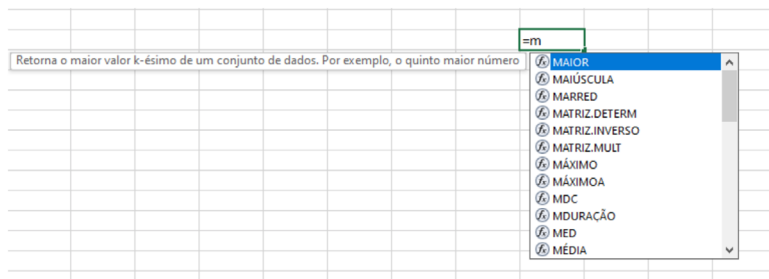


Figura 12: Inserir função com o símbolo igual (=).

## Média

Para calcular a média da variável `vlr_renda_media_fam`, é necessário ir em “Inserir função” (circulado em vermelho na Figura 13) e selecionar “MÉDIA”, ou colocar em uma célula limpa: =MÉDIA (selecionar dados).

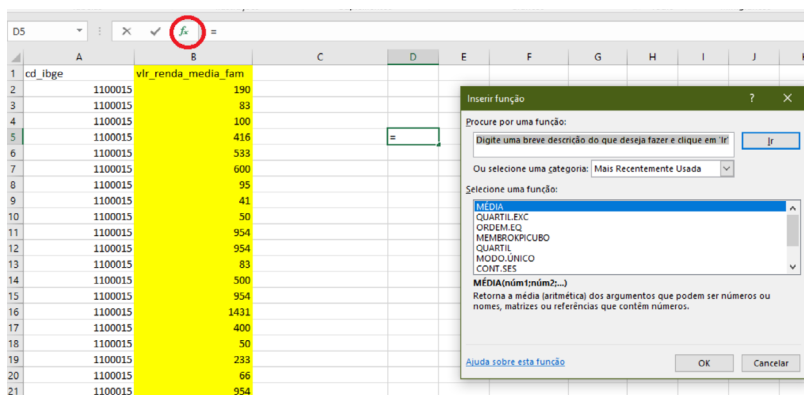


Figura 13: função média.

Para calcular a mediana e a moda, o processo é o mesmo que da média, só mudam as funções, que são:

- Mediana: =MED(selecionar dados);
- Moda: =MODO(selecionar dados), utilizados para dados numéricos.
- Moda: =CONT.SE(intervalo;"critérios"), utilizados para variáveis categóricas.

## Organização dos dados:

Para classificar os dados do menor valor até o maior valor, ou vice-versa, é necessário selecionar o conjunto de dados que deseja ordenar e ir na opção “classificar”, como mostra a Figura 14.

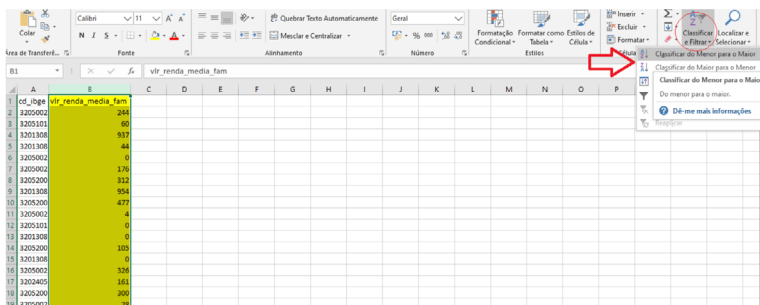


Figura 14: classificação dos dados.

## Medidas de dispersão

### Amplitude

Para encontrar a amplitude, depois de ordenar os dados, é necessário colocar em uma célula vazia o sinal de igual (=), depois selecionar a última célula com o maior valor, inserir o símbolo de – (menos), selecionar a célula de menor valor, por fim, apertar enter.

base_amostra_familia_201812						
	A	B	C	D	E	F
1	cd_ibge	vlr_renda_media_fam		Linha	cd_ibge	vlr_renda_media_fam
2	1100015	0		L2	1100015	0
3	1100015	0			...	...
4	1100015	0		L362	1100015	1908
5	1100015	0				
6	1100015	0				
7	1100015	0				
8	1100015	0				
9	1100015	0				
10	1100015	5				
11	1100015	13				
12	1100015	18				
13	1100015	25				
14	1100015	25				
15	1100015	25				
16	1100015	26				
17	1100015	26				
18	1100015	27				
19	1100015	33				
20	1100015	33				
21	1100015	33				
22	1100015	34				

Figura 15: cálculo da amplitude.

O desvio padrão e a variância podem ser calculados como a média, só altera a nome da função que se utiliza:

- Desvio Padrão (DV): =DESVPAD (em relação a amostra).
- Variância: =VAR.A(dados).
- Coeficiente de variação (CV).

A partir da média e do desvio padrão é possível construir o CV(Fig 16). Primeiramente, coloca-se o símbolo de igual (=). Depois, seleciona-se o valor do DV dos dados, coloca-se a barra (/) para fazer a divisão, seleciona-se o valor da média e multiplica tudo por 100, como ilustra a Figura 9.

cd_ibge	vlr_renda_media_fam				
1100015	190				
1100015	83				
1100015	100	Média	394,4472	CV	= (E4/E3)*100
1100015	416	DV	440,2899		
1100015	533				
1100015	600				
1100015	95				
1100015	41				
1100015	50				
1100015	954				
1100015	954				
1100015	83				
1100015	500				
1100015	954				
1100015	1431				
1100015	400				

Figura 16: Coeficiente de variação.

# Quartil

Para calcular o quartil é necessário ir em inserir função e optar por QUARTIL, selecionar o conjunto de dados (matriz) e inserir qual quartil deseja (1, 2, ou 3), como ilustra a Figura 17.

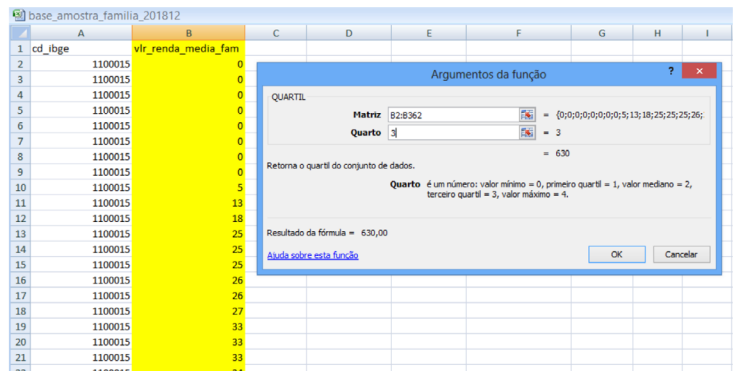


Figura 17: quartil

## Construção de tabelas

A partir dessas informações é possível construir tabelas, como a Tabela 7, construída a partir dos dados da Figura 18, dados que foram obtidos usando as funções acima para encontrar as medidas de posição e dispersão para os dados da renda familiar.

Tabela 7: Medidas resumo para os valores de renda familiar média do Cadastro único, Encontro das Águas, 2018.

Média	Mediana	Moda	Amplitude	$Q_1$	$Q_3$	Distância Interquartílica	Desvio Padrão	CV%
393,56	175,00	50,00	1908,00	83,00	630,00	547,00	440,00	111%

base_amostra_familia_201812			
	A	B	C
356	1100015	1874	
357	1100015	1874	
358	1100015	1874	
359	1100015	1900	
360	1100015	1900	
361	1100015	1900	
362	1100015	1908	
363			
364	Média	393,56	
365	Mediana	175,00	
366	Moda	50,00	
367	Desvio Padrão	440,00	
368	Coefficiente de Variação	111%	
369	Quartil 1	83,00	
370	Quartil 3	630	
371	Distância Interquartilica	547,00	
372			

Figura 18: dados para construção de Tabela 5.

É possível construir tabelas para mais variáveis também, como para os dados dos bairros de Encontro das Águas e Cachoeira da Norte. Como exemplo, observe a Tabela 8, que foi construída a partir dos dados da Figura 19.

Fonte: [Base do Cadastro Único com marcação do Bolsa Família - Microdados Dez/2018 - Cadastro Único e PBF \(atualizado em 12/11/2019, 14:12\). Município Código IBGE 1100015 e 1100023.](#)

Tabela 8: Medidas resumo para os valores de renda familiar média do Cadastro Único. Encontro das Águas e Cachoeira do Norte, 2018.

Município	Média	Mediana	Moda	Amplitude	$Q_1$	$Q_3$	Distância Interquartilica	Desvio padrão
Encontro das águas	394,48	175	954	1908	83	631,5	548,5	440,28
Cachoeira da Norte	356,9	200	100	2054	100	500	400	355,1

1100015	100	1100023	20
1100015	833	1100023	150
1100015	85	1100023	39
1100015	50	1100023	274
1100015	116	1100023	150
1100015	137	1100023	152
1100015	572	1100023	133
1100015	1380	1100023	116
1100015	83	1100023	954
1100015	216	1100023	937
Média	394,4472222	Média	356,9074658
Mediana	175	Mediana	200
Moda	954	Moda	100
Desvio Padrão	440,2899431	Desvio Padrão	355,0979433
Coeficiente de Vari	1,11	Coeficiente de	100,5095841
Quartil 1	83	Quartil 1	100
Quartil 3	631,5	Quartil 3	500
Distância Interquat	548,5	Distância Inter	400
Valor mínimo	0	Valor mínimo	0
Valor máximo	1908	Valor máximo	2054

Figura 19: Dados para a construção da tabela.

## Exercício [\(ver gabarito no final do e-book\)](#)

1) Com os dados obtidos nos Exercício resolvidos ao final dos Capítulos 1 e 2, para a Tabela 5: Renda de 14 famílias cadastradas no programa assistencial, construa uma tabela resumo com as medidas de posição e dispersão.

# Capítulo 4

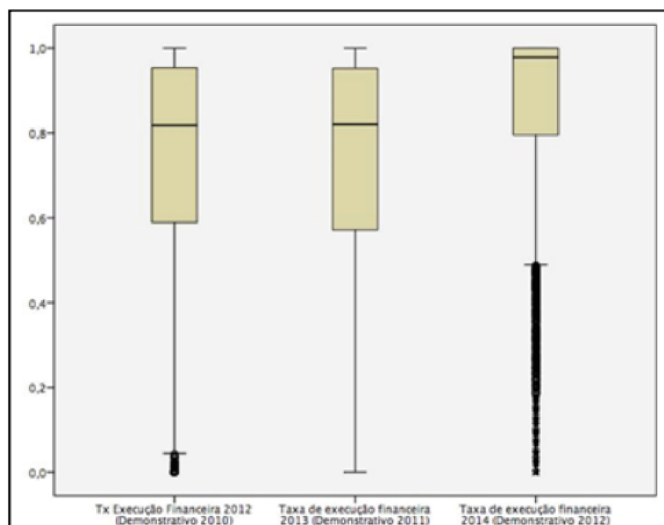


Figura 20: Taxa de Execução financeira do IGD-Suas para todos os municípios, por ano - 2015

Fonte: [Elaborado pelos autores SATYRO; CUNHA; CAMPOS, 2016.](#)

Olhando este gráfico, algum desavisado poderia pensar que se trata de uma obra de arte ou coisa do tipo, mas, com um pouco de conhecimento de estatística, podemos localizar nessa imagem as principais medidas de posição e dispersão aprendidas anteriormente, como a mediana e os quartis, que dividem os dados em 4 partes, contendo 25% em cada. Além disso, analisando o gráfico, podemos perceber a presença de observações discrepantes ao esperado para a grande maioria dos dados coletados. O gráfico mostra que há uma diferença nos valores das medianas e dos quartis ao longo do tempo, indicando uma mudança na distribuição dos dados ao longo dos anos.

Essa figura encontra-se no trabalho "Análise espacial da burocracia da assistência social nos municípios brasileiros: notas para uma reflexão sobre a capacidade de implementação dessa política" (Autores: Natália Guimarães Duarte; Sátyro Eleonora Schettini; Martins Cunha Jarvis Campos, 2016. Disponível em: <https://www.scielo.br/pdf/op/v22n2/1807-0191-op-22-2-0286.pdf>).

O objetivo do trabalho foi realizar uma análise da capacidade de implementação da política de assistência social e, entre os dados analisados, foi estudada a Taxa de Execução Financeira do Índice de Gestão Descentralizada (IGD-Suas).

Os dados foram representados na figura que você está vendo. A interpretação desse gráfico informa que houve significativa melhora da média de execução do IGD-Suas no ano de 2012, mas existe grande variação nos dados dos municípios que não conseguem executar.





## Diagrama de caixa (Box-plot)

O box-plot é uma forma gráfica de visualizar a distribuição dos dados que representa os quartis por meio de um retângulo, permitindo observar se a distribuição dos dados é simétrica ou assimétrica em relação ao ponto central, representado pela mediana ( $Q_1$ ).

A Figura 21 abaixo ilustra e especifica o box-plot, mostrando como ele é composto. A expressão “outlier” é utilizada para descrever observações discrepantes, ou seja, observações que estão fora dos limites calculados, considerando a distribuição observada de todos os dados.

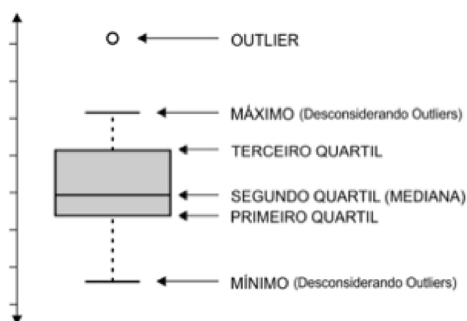


Figura 21: Representação e explicação do box-plot.

Para construir o gráfico é importante:

1. Calcular o valor mínimo e máximo, primeiro e terceiro quartil e a mediana.
2. Criar um segmento de reta em posição vertical, para representar a amplitude dos dados.
3. Marcar, nesse segmento, o primeiro, o segundo (mediana) e o terceiro quartis.
4. Criar uma caixa retangular (box) de maneira que o lado superior e o inferior passem exatamente sobre os pontos que marcam o primeiro e o terceiro quartis.
5. Fazer um ponto para representar a mediana (obedecendo a escala).

### Exemplo 14:

Criar um diagrama de caixa para os dados:

1, 2, 3, 4, 5, 6, 7, 8, 9, 10

Primeiramente é necessário calcular os quartis e os encontrar o valor máximo e mínimo:

Mínimo: 1

Primeiro quartil:  $\frac{2+3}{2} = 2,53$

Mediana:  $\frac{5+6}{2} = 5,5$

Terceiro quartil:  $\frac{8+9}{2} = 8,5$

Máximo: 10

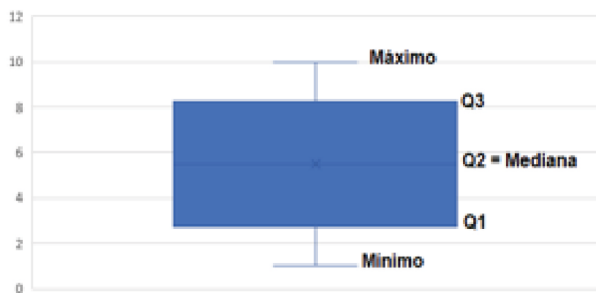


Figura 22: box-plot para o exemplo 14.

Para calcular o limite superior e inferior do boxplot usamos as fórmulas:

$$\text{Limite Inferior} = Q1 - 1,5 * (Q3 - Q1)$$

$$\text{Limite Superior} = Q3 - 1,5 * (Q3 - Q1)$$

$$L_{\text{inf}} = Q1 - 1,5 * (Q3 - Q1) = 2,5 - 1,5 * 6 = -6,5$$

$$L_{\text{sup}} = Q3 + 1,5 * (Q3 - Q1) = 8,5 + 1,5 * 6 = 17,5$$

Quando os dados não atingem os limites, a barra é indicada para o valor de máximo ou mínimo do conjunto de dados, que é o caso deste exemplo.

Exemplo 15:

Faça o boxplot para os dados da renda familiar média do Cadastro único, Encontro das águas, 2018.

Tabela 9: Medidas resumo para os valores de renda familiar média do Cadastro único, Encontro das Águas, 2018.

Média	Mediana	Moda	Amplitude	$Q_1$	$Q_3$	Distância Interquartílica	Desvio Padrão	CV%
393,56	175,00	50,00	1908,00	83,00	630,00	547,00	440,00	111%

Com os dados da Tabela 9, é possível construir manualmente o box-plot para os dados de renda familiar média do Cadastro Único do Município de Encontro das águas em 2018.

Entretanto, como são muitos dados, fica inviável fazer a construção do diagrama de caixa manualmente. Dessa forma, no Excel, é necessário selecionar os dados da renda, ir na aba inserir, depois em gráficos ou gráficos recomendados, escolher a opção todos os gráficos e, por fim, selecionar caixa, como ilustra a Figura 23.

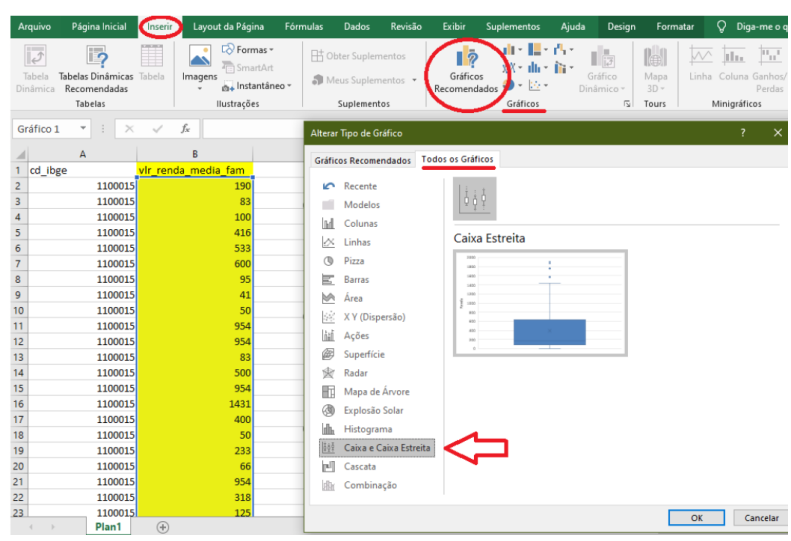


Figura 23: Esquema para a construção do diagrama de caixa (Box-plot).

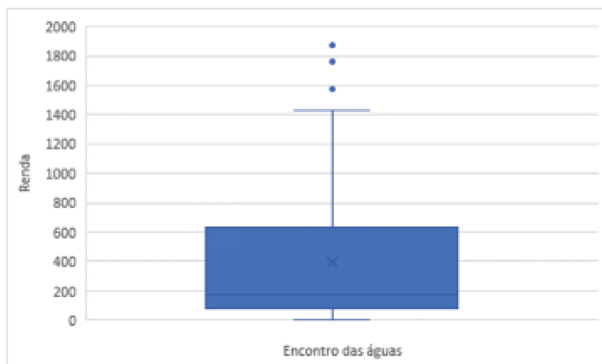


Figura 24: Box-plot para o exemplo 15.

Os valores observados fora dos limites inferior e superior são representados por pontos ou asteriscos na posição que ocorrem, pois representam as observações discrepantes (outliers), com comportamento diferente da maioria dos dados. Pode-se observar na Figura 24 a presença dos outliers.

## Exemplo 16:

Construa o box-plot para a Tabela 10.

Tabela 10: Medidas resumo para os valores de renda familiar média do Cadastro Único, Encontro das Águas e Cachoeira do Norte, 2018.

Município	Média	Mediana	Moda	Amplitude	$Q_1$	$Q_3$	Distância Interquartilica	Desvio padrão
Encontro das águas	394,48	175	954	1908	83	631,5	548,5	440,28
Cachoeira da Norte	356,9	200	100	2054	100	500	400	355,1

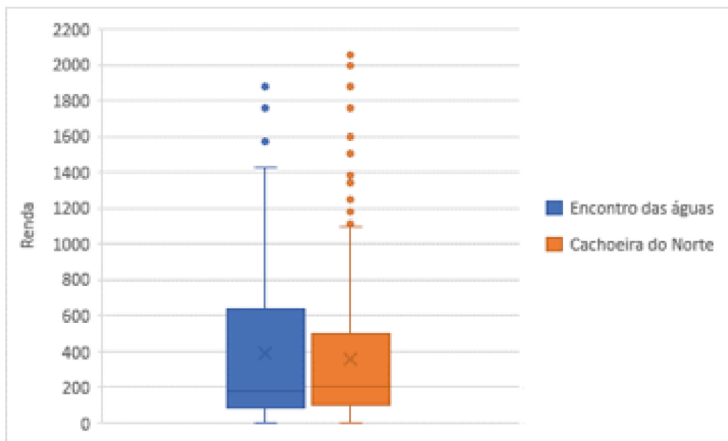


Figura 25: Box-plot para o exemplo 16.

Na Figura 25, pode-se observar que os dados do município de Cachoeira do Norte estão mais concentrados (caixa menor) e possuem mais valores discrepantes. O município de Encontro das águas possui dados com mais variação em relação ao de Cachoeira do Norte e menos observações discrepantes.

## Exercício [\(ver gabarito no final do e-book\)](#)

Vamos retomar a Tabela 4: Renda de 14 famílias cadastradas no programa assistencial (Capítulo 1)

Tabela 5: Renda Familiar de 14 cadastros do Cadastro único, município, Cachoeira do Norte, 2020.

Renda Familiar	
R\$ 320,00	R\$ 250,00
R\$ 480,00	R\$ 480,00
R\$ 640,00	R\$ 510,00
R\$ 710,00	R\$ 250,00
R\$ 250,00	R\$ 480,00
R\$ 320,00	R\$ 640,00
R\$ 800,00	R\$ 250,00

# Gabarito Exercícios

## Capítulo 1

### [Exercício 1 \(página 18\)](#)

Resposta: Primeiramente, é necessário construir a tabela de distribuição de frequência para facilitar o cálculo.

Tabela 5: Frequência das rendas das 14 famílias

Valor	Frequência
R\$250,00	4
R\$320,00	2
R\$480,00	3
R\$510,00	1
R\$640,00	2
R\$710,00	1
R\$800,00	1
<b>Total</b>	<b>14</b>

### Média

$$\frac{(2504) + (320 \times 2) + (480 \times 3) + (510 \times 1) + (640 \times 2) + (710 \times 1) + (800 \times 1)}{14} = \frac{6380}{14} = \text{R\$} : 455,71$$

Vale ressaltar que podemos calcular a média utilizando a fórmula acima, que multiplica o valor pela frequência que apareceu nos dados, ou calcular por meio do somatório de todos os valores apresentados na Tabela 3, pois chegaremos ao mesmo resultado.

Ordenando os dados:

R\$ 250,00   R\$ 250,00   R\$ 250,00   R\$ 250,00   R\$ 320,00   R\$ 320,00   **R\$ 480,00**  
**R\$ 480,00**   R\$ 480,00   R\$ 510,00   R\$ 640,00   R\$ 640,00   R\$ 710,00   R\$ 800,00

Como o conjunto de dados é par, é necessário calcular a média dos dois valores centrais.

Mediana:  $\frac{480 + 480}{2} = \frac{960}{2} = \text{R\$} : 480,00$

Moda: R\$250,00 (apareceu 4 vezes na Tabela 4).

As 14 famílias possuem em média de salário, o valor mediano é de R\$480,00 e a moda é R\$250,00. Note que a média é menor que a mediana, neste caso, isso ocorre porque temos 4 famílias recebendo apenas R\$250,00, que por coincidência corresponde a moda dos dados.

## Capítulo 2

### Exercício 1 (página 28)

**Amplitude:** Para calcular a amplitude, primeiramente precisamos ordenar os dados:

R\$ 250,00 R\$ 250,00 R\$ 250,00 R\$ 250,00 R\$ 320,00 R\$ 320,00 R\$ 480,00  
R\$ 480,00 R\$ 480,00 R\$ 510,00 R\$ 640,00 R\$ 640,00 R\$ 710,00 R\$ 800,00

Maior valor: 800,00

Menor valor: 250,00

Amplitude =  $800 - 250 = \text{R\$}550,00$ .

**Quartil:** Primeiramente, o que é a mediana, já foi calculado no capítulo anterior:

$$\text{Mediana: } \frac{480 + 480}{2} = \frac{960}{2} = \text{R\$} 480,00 = Q_2$$

O conjunto de dados que antecede o  $Q_2$  é:

R\$ 250,00 R\$ 250,00 **R\$ 250,00 R\$ 250,00** R\$ 320,00 R\$ 320,00

$$\text{A mediana desse novo conjunto é: } \frac{250+250}{2} = \frac{500}{2} = \text{R\$} 250,00 = Q_1$$

O conjunto de dados que sucede :

R\$ 480,00 R\$ 510,00 **R\$ 640,00 R\$ 640,00** R\$ 710,00 R\$ 800,00

$$\text{A mediana desse novo conjunto é: } \frac{640+640}{2} = \frac{1280}{2} = \text{R\$} 640,00 = Q_3$$

### Distancia interquartilica

$$\text{Distância interquartilica} = Q_3 - Q_1 = 640 - 250 = \text{R\$ } 390,00.$$

### Variância

Para calcular a variância podemos utilizar a Tabela 5 construída no exercício do capítulo 1, lembrando que Média = R\$:455,71 .

Tabela 5: Frequência das rendas das 14 famílias

Valor	Frequência
R\$250,00	4
R\$320,00	2
R\$480,00	3
R\$510,00	1
R\$640,00	2
R\$710,00	1
R\$800,00	1
<b>Total</b>	<b>14</b>

$$s^2 = \frac{(250-455,71)^2 * 4 + (320-455,71)^2 * 2 + (480-455,71)^2 * 3 + (510-455,71)^2 + (640-455,71)^2 * 2 + (710-455,71)^2 + (800-455,71)^2}{14-1}$$

$$s^2 = \frac{169266 + 36834 + 1770 + 2947 + 67926 + 64663 + 11853}{13}$$

$$s^2 = \frac{461942}{13} = \text{R\$ } 35,534$$

### Desvio padrão:

$$s: \sqrt{35,534} = \text{R\$ } 188,50$$

### Coefficiente de variação

$$\text{CV: } \frac{188,50}{455,71} * 100\% = 41,36$$



Capítulo 3

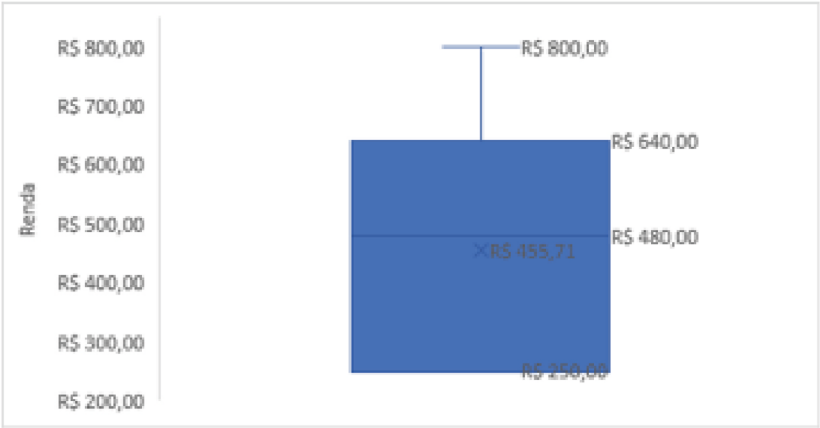
[Exercício 1 \(página 39\)](#)

Tabela 7: Medidas resumo para os valores de renda de 14 famílias cadastradas no programa assistencial, inventado

Média	Mediana	Moda	Amplitude	$Q_1$	$Q_3$	Distância Interquartílica	Desvio Padrão
455,71	480	250	550,00	250	607,5	357,5	188,50

Capítulo 4

[Exercício 1 \(página 46\)](#)



# Para saber mais...

ARNOT, Antônio. **Estatística Fácil**. São Paulo: Saraiva, 2002.

IBGE. INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Síntese de indicadores sociais : uma análise das condições de vida da população brasileira**: 2019 / IBGE, Coordenação de População e Indicadores Sociais. - Rio de Janeiro: IBGE, 2019.

MORETTIN, Pedro Alberto; BUSSAB, Wilton de Oliveira. **Estatística básica**. rev. e atual. São Paulo: Saraiva, 2010.

PADOVANI, Carlos Roberto. **Bioestatística**. São Paulo: Cultura Acadêmica: Universidade Estadual Paulista, Pró-Reitoria de Graduação, 2012.

PAGANO, M.; KIMBERLEE, G. **Princípios de Bioestatística**. 2a edição. Thomson. ed., São Paulo, 2004.

SATYRO, Natália Guimarães Duarte; CUNHA, Eleonora Schettini Martins;

CAMPOS, Jarvis. **Análise espacial da burocracia da assistência social nos municípios brasileiros: notas para uma reflexão sobre a capacidade de implementação dessa política**. Opin. Publica, Campinas , v. 22, n. 2, p. 286-317, Aug. 2016 . Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_art-text&pid=S0104-62762016000200286&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_art-text&pid=S0104-62762016000200286&lng=en&nrm=iso)>. Acessado em: 25 jun 20.

VIEIRA, Sonia. **Estatística básica**. São Paulo: Cengage Learning, v. 9, 2012.

**Realização**  
Ministério da Cidadania

**Execução**  
Centro de Ciência, Tecnologia e Inovação para Soberania e  
Segurança Alimentar e Nutricional da UNESP - INTERSSAN

