

CET797
Elementos de Estatística
Universidade Estadual de Santa Cruz

Medidas Estatísticas

Kaique Brandão Macêdo da Silva

Adaptação dos slides anteriores (Nilo Fernandes Varela e Ellison
W. M. Guimarães)

Ilhéus – BA

2019

Tendência Central



- Medidas que orientam quanto aos valores centrais.
- Representam os fenômenos pelos seus valores médios, em torno dos quais tendem a se concentrar os dados.
- Também chamados de **centro da distribuição**.

1. **Média**
2. **Moda**
3. **Mediana**

Média

Aritmética, Geral, Geométrica e Harmônica

Média

Média Aritmética

Dados Não Agrupados

Seja $Y = (y_1, y_2, \dots, y_n)$ e (N, n) = Quantidade de variáveis em Y , temos que:

$$\begin{array}{l} \text{Parâmetro: } \bar{Y} \text{ ou } \mu = \frac{\sum y}{N} \\ \text{Estimativa: } \bar{y} \text{ ou } m = \frac{\sum y}{n} \end{array}$$

Média Populacional (μ)

$$\mu = \frac{\sum Y}{N}$$

N sendo o
tamanho
da população

Média Amostral (m)

$$m = \frac{\sum y}{n}$$

n sendo o
tamanho
da amostra

Média Aritmética

Dados Não Agrupados

Ex:

As notas dos alunos da disciplina Elementos de Estatística estão representadas no seguinte quadro:

Aluno	A	B	C	D	E	F	G
Nota	6,5	10	8	9,4	8	6,4	7

$$y = (6.5, 10, 8, 9.4, 8, 6.4, 7), n = 7$$

$$m = \frac{\sum y}{n} = \frac{6,5 + 10 + 8 + 9,4 + 8 + 6,4 + 7}{7} = \frac{55.3}{7} = 7,89 \checkmark$$



Média Aritmética com R

Dados Não Agrupados

A função `mean()` calcula a média ao passar um vetor como argumento.

```
> y <- c(6.5, 10, 8, 9.4, 8, 6.4, 7)
> mean(y)
[1] 7.9
>
```

Média Aritmética

Dados Agrupados

Quando a amostra for um valor relativamente grande, agrupamos os dados em uma tabela para facilitar a compreensão.

$$Y = (y_1, y_2, \dots, y_n)$$

$$F = (F_1, F_2, \dots, F_n)$$

Y sendo as médias de cada intervalo.

F sendo a frequência de cada intervalo

O cálculo é dado por:

$$m = \frac{\sum y \cdot F}{\sum F}$$

Média Aritmética

Dados Agrupados (SEM INTERVALO)

Exemplo:

Dadas as seguintes idades de alunos do 5º semestre do curso de Agronomia, determine a média amostral:

15, 15, 16, 16, 16, 16, 16, 16, 16, 16, 16, 17, 17, 17, 17, 17, 17,
17, 18, 18, 18, 18, 18, 18, 22, 22, 22, 22, 22, 26, 26, 26, 26, 26,
26

OBS: Para facilitar a compreensão, iremos organizar os dados numa tabela por frequência...

Média Aritmética

Dados Agrupados (SEM INTERVALO)

Y	Fi	Y*Fi
15	2	30
16	9	144
17	7	119
18	6	108
22	5	110
26	6	156
Σ	35	667

$$m = \frac{\Sigma y.F}{\Sigma F} = \frac{(15 \times 2) + (16 \times 9) + (17 \times 7) + (18 \times 6) + (22 \times 5) + (26 \times 6)}{2 + 9 + 7 + 6 + 5 + 6} = \frac{667}{35} = 19,06 \quad \checkmark$$



Média Aritmética com R: Ex 1

Dados Agrupados (SEM INTERVALO)

A função `table()` fornece uma matriz. Sendo que a primeira linha contém os elementos sem repetição, e na segunda linha contém a frequência dos respectivos valores acima.

```
>
> #Declarando o vetor
> y <- c(15, 15, 16, 16, 16, 16, 16, 16, 16, 16, 16, 16, 17, 17, 17, 17, 17, 17, 17, 18, 18, 18, 18, 18, 18,
, 22, 22, 22, 22, 22, 26, 26, 26, 26, 26, 26)
> #Obtendo a tabela
> table(y)
y
15 16 17 18 22 26
 2  9  7  6  5  6
> mean(y)
[1] 19.05714
>
```



Média Aritmética com R: Ex 2

Dados Agrupados (SEM INTERVALO)

A função `sum()` soma todos os elementos de um vetor.
A multiplicação de vetores acontece com o operador de multiplicação.

```
>
> #Declarando um vetor com as idades
> y <- c(15, 16, 17, 18, 22, 26)
>
> #Declarando um vetor com a frequência de cada idade
> f <- c(2, 9, 7, 6, 5, 6)
>
> #Calculo:
> sum(y*f)/sum(f)
[1] 19.05714
>
> y*f
[1] 30 144 119 108 110 156
> |
```

Média Aritmética

Dados Agrupados (COM INTERVALO)

Exemplo:

Dadas as seguintes idades de pessoas que foram em um determinado consultório médico, determine a média amostral:

Idade	Fi	Y	Y*Fi
0 † 10	70	5	350
10 † 20	20	15	300
20 † 30	15	25	375
30 † 40	15	35	525
40 † 50	12	45	540
50 † 60	18	55	990
60 † 70	50	65	3250
Σ	200		6330

$$\begin{aligned}
 m &= \frac{\Sigma y.F}{\Sigma F} = \frac{(5 \times 70) + (15 \times 20) + (25 \times 15) + (35 \times 15) + (45 \times 12) + (55 \times 18) + (65 \times 50)}{70 + 20 + 15 + 15 + 12 + 18 + 50} \\
 &= \frac{6330}{200} = 31,65 \quad \checkmark
 \end{aligned}$$



Média Aritmética com R

Dados Agrupados (COM INTERVALO)

A função `mean()` suporta uma tabela de distribuição de frequência (FDT) como argumento:

```
> #Agrupando classes - Reconstituindo um FDT
> tbl <- make.fdt(f=c(70, 20, 15, 15, 12, 18, 50), start=0, end=70)
> tbl
  Class limits  f   rf rf(%)  cf cf(%)
  [0,10)  70 0.35  35.0   70  35.0
  [10,20)  20 0.10  10.0   90  45.0
  [20,30)  15 0.08   7.5  105  52.5
  [30,40)  15 0.08   7.5  120  60.0
  [40,50)  12 0.06   6.0  132  66.0
  [50,60)  18 0.09   9.0  150  75.0
  [60,70)  50 0.25  25.0  200 100.0
> mean(tbl)
[1] 31.65
> |
```

Média Geral

$$K = (y_{\mathbf{1}}, y_{\mathbf{2}}, \dots, y_{\mathbf{k}})$$

$$N_{\mathbf{K}} = (n_{\mathbf{1}}, n_{\mathbf{2}}, \dots, n_{\mathbf{k}})$$

K sendo as estimativas das médias aritméticas de k series.

N sendo o número de termos de cada estimativa de uma série K

O cálculo é dado por:

$$m = \frac{\sum n \cdot y}{\sum n}$$

$$\bar{y} \text{ ou } m = \frac{n_1 \bar{y}_1 + n_2 \bar{y}_2 + \dots + n_k \bar{y}_k}{n_1 + n_2 + \dots + n_k}$$

Média Geral

1) (4, 5, 6, 7, 8)

$$n1 = 5 \quad y1 = 6$$

2) (1, 2, 3)

$$n2 = 3 \quad y2 = 2$$

3) (9, 10, 11, 12, 13)

$$n3 = 5 \quad y3 = 11$$

$$m = \frac{\sum n \cdot y}{\sum n} = \frac{(5 \times 6) + (3 \times 2) + (5 \times 11)}{5 + 3 + 5} = \frac{91}{13} = 7$$

$$\bar{Y} = \frac{n_1 \bar{y}_1 + n_2 \bar{y}_2 + \dots + n_k \bar{y}_k}{n_1 + n_2 + \dots + n_k} = \frac{5 \cdot 6 + 3 \cdot 2 + 5 \cdot 11}{5 + 3 + 5} = 7$$



Média Geral com R

```
> y1 <- 4:8
> y2 <- 1:3
> y3 <- 9:13
> y1
[1] 4 5 6 7 8
> y2
[1] 1 2 3
> y3
[1] 9 10 11 12 13
> medGera1<-function(y1,y2,y3){
+   (length(y1)*mean(y1) + length(y2)*mean(y2) + length(y3)*mean(y3)) /
+   (length(y1) + length(y2) + length(y3))
+ }
> medGera1(y1,y2,y3)
[1] 7
```

Média Geométrica

- Usada para medidas proporcionais de crescimento quando uma medida subsequente depende de medidas prévias.
- A média geométrica pode ser entendida em termos da geometria. A média geométrica de dois números, **a** e **b**, é o tamanho do lado de um quadrado cuja área é igual à área de um retângulo com lados de tamanho **a** e **b**.

Uma amostra:

$$Y = (y_1, y_2, \dots, y_n)$$

O cálculo é dado por:

$$mg = \sqrt[n]{y_1 * y_2 * \dots * y_n}$$

Média Geométrica

Ex:

Se um investimento rende 10% no primeiro ano e 20% no segundo ano, qual será o rendimento médio desse investimento?

Sendo C o capital inicial, após dois anos o montante M será:

$$M = C \times 1,10 \times 1,20 = 1,32 \times C$$

Se aplicarmos a média aritmética das porcentagens rendidas a cada ano, teríamos 15% como média.

$$M = C \times 1,15 \times 1,15 = 1,3225 \times C$$

Média Geométrica

Se um investimento rende 10% no primeiro ano e 20% no segundo ano, qual será o rendimento médio desse investimento?

Sendo C o capital inicial, após dois anos o montante M será:

$$M = C \times 1,10 \times 1,20 = 1,32 \times C$$

Se aplicarmos a média aritmética das porcentagens de 10% e 20% de cada ano, teríamos 15% como média:

$$M = C \times 1,15 \times 1,15 = 1,3225 \times C$$


Média Geométrica

Se um investimento rende 10% no primeiro ano e 20% no segundo ano, qual será o rendimento médio desse investimento?

$$M = C \times 1,10 \times 1,20 = 1,32 \times C$$

Se aplicarmos a média geométrica das porcentagens rendidas a cada ano:

$$mg = \sqrt[2]{1,10 \times 1,20} \cong 1.1489 \text{ ou } 14,89\%$$



Média Geométrica com R

Com pacote `psych` você consegue calcular a média geométrica através da `geometric.mean()`, passando um vetor ou um data frame como argumento. Também, através da função `prod()` e `length()`, aplicando a fórmula da mg, se obtém a média de forma fácil.

```
> library(psych)

> y <- c(1.10, 1.20)

> geometric.mean(y)
[1] 1.148913

> (prod(y)^(1/length(y)))
[1] 1.148913
```

Média Harmônica

- Usada para médias de crescimento e proporções de velocidade.
- A média harmônica está relacionada ao cálculo matemático das situações envolvendo as grandezas inversamente proporcionais.

$$Y = (y_1, y_2, \dots, y_n)$$

$$F = (F_1, F_2, \dots, F_n)$$

Y sendo os elementos

F sendo a frequência de cada elemento

O cálculo é dado por:

$$mh = \frac{n}{\frac{F_1}{y_1} + \frac{F_2}{y_2} + \dots + \frac{F_n}{y_n}}$$

Média Harmônica

Problema: Cláudio vai de casa para o trabalho com uma velocidade de 30km/h. Já na volta, devido ao trânsito, faz o mesmo caminho com velocidade média de 20km/h. Qual foi sua velocidade média considerando ida e volta?



IDA:

$$t_1 = \frac{d}{30}$$

VOLTA:

$$t_2 = \frac{d}{20}$$

$$V_m = \frac{\Delta s}{\Delta t} = \frac{2d}{\frac{d}{30} + \frac{d}{20}} = \frac{2}{\frac{1}{30} + \frac{1}{20}} = \frac{2}{\frac{5}{60}} = \frac{2 \times 60}{5} = \frac{120}{5} = 24 \text{ km/h}$$

Média Harmônica

Média Harmônica com R

Com pacote `psych` você consegue calcular a média harmônica através da `harmonic.mean()`, passando um vetor ou um data frame como argumento. Também, através da função `sum()` e `length()`, aplicando a fórmula da mh, se obtém a média de forma fácil.

```
>  
> library(psych)  
> y <- c(30, 20)  
>  
> harmonic.mean(y)  
[1] 24  
>  
> length(y)/(sum(1/y))  
[1] 24  
> |
```

Média

Vantagens

- Fácil de compreender e calcular;
- Utiliza todos os valores da série;
- É um valor único;
- É fácil de ser incluída em expressões matemáticas;
- Pode ser determinada nas escalas: intervalar e proporcional.

Desvantagens

- Muito afetada por valores extremos;
- Necessário conhecer todos os valores da série.

Moda



Moda

- A moda de um conjunto de dados trata do valor que ocorre com maior frequência ou o valor mais comum em um conjunto de dados.
- Notação adotada:
 - **MO**: para o parâmetro.
 - **mo**: para a amostra.
- Uma série sem moda é uma série **amodal**.
- Uma série com mais de uma moda é denominada de série **multimodal**.

Moda

Sem Agrupamento de Classes

Ex:

Dadas as seguintes idades de alunos de uma classe do segundo ano noturno de uma escola, determine a moda:

15, 15, 16, 16, 16, 16, 16, 16, 16, 16, 16, 17, 17, 17, 17, 17, 17, 17, 18, 18, 18, 18, 18, 22, 22, 22, 22, 22, 22, 26, 26, 26, 26, 26, 26

Para facilitar, iremos organizar os dados numa tabela por frequência:

y_i	15	16	17	18	22	26
F_i	2	9	7	6	5	6

Analisando, a idade mais frequente na turma é 16, com frequência igual a 9. Ou seja, **mo = 16**.



Moda com o R

Sem Agrupamento de Classes

A função `mfv()`, pertencente ao pacote `fdth`, calcula a moda.

```
>
> y <- c(15, 15, 16, 16, 16, 16, 16, 16, 16, 16, 16, 16, 17, 17, 17, 17, 17, 17,
17, 18, 18, 18, 18, 18, 18, 18, 22, 22, 22, 22, 22, 22, 26, 26, 26, 26, 26, 26)
>
> table(y)
y
15 16 17 18 22 26
 2  9  7  6  5  6
>
> mfv(y)
[1] 16
>
|
```

Moda

Com Agrupamento de Classes

- Passo 1: Identificar a classe modal (De maior frequência)
- Passo 2: Utilizar a fórmula de Czuber

$$mo = l + \frac{D1}{D1+D2} \times (L - l)$$

L e l: Limite superior e inferior da classe modal

D1: Diferença entre a frequência da classe modal e a imediatamente anterior

D2: Diferença entre a frequência da classe modal e a imediatamente posterior

Moda

Com Agrupamento de Classes

Dada uma tabela que contém a quantidade de salários mínimos e a quantidade de pessoas que recebem as respectivas quantidades:

<i>Salário Mínimo</i>	4 + 8	8 + 12	12 + 16	16 + 20	20 + 24	Σ
F_i	5	7	4	3	1	20

- Passo 1: Identificar a classe modal (De maior frequência)
- Passo 2: Utilizar a fórmula de Czuber

$$mo = 8 + \frac{2}{2+3} \times (12 - 8) = 8 + \frac{2}{5} \times 4 = 9,6 \text{ sm} \checkmark$$



Moda com o R

Com Agrupamento de Classes

A função `mfv()`, pertencente ao pacote `fdth`, calcula a moda.

```
> #Distribuição com agrupamento de classes
> y <- make.fdt(f=c(5, 7, 4, 3, 1), start=4, end=24)
> y
  class limits f   rf rf(%) cf cf(%)
    [4,8)  5 0.25   25  5   25
    [8,12)  7 0.35   35 12   60
   [12,16)  4 0.20   20 16   80
   [16,20)  3 0.15   15 19   95
   [20,24)  1 0.05    5 20  100
>
> mfv(y)
[1] 9.6
>
```

Moda

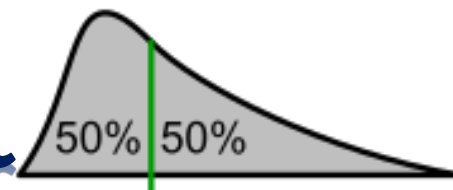
Vantagens

- Fácil de compreender e calcular
- Não é afetada por valores extremos
- Pode ser aplicada em todas as escalas: nominal, ordinal, intervalar e proporcional.

Desvantagens

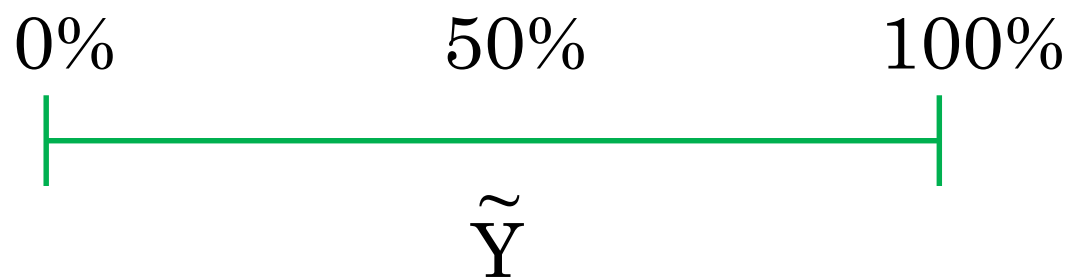
- Pode estar afastada do centro dos valores
- É difícil de ser incluída em expressões matemáticas
- Não usa todos os valores da série
- Os dados podem ter mais de uma moda (bimodal ou multimodal).
- Alguns dados não possuem moda.

Mediana



Mediana

Medida de tendência muito usada quando o interesse é a determinação do valor que separa a série de dados em duas partes iguais, 50% situados acima e 50% situados abaixo da medida.



Notação:

\tilde{Y} ou MD para o parâmetro.

\tilde{y} ou md para a estimativa.

Mediana

Com Variáveis Discretas

- Se o número de elementos n for ímpar, a mediana é o elemento central:

$$\tilde{Y} = \frac{n + 1}{2}$$

- Se o número de elementos n for par, a mediana é a média dos dois elementos centrais:

$$\tilde{Y} = \textit{Media} \left[\frac{n}{2}, \frac{n + 1}{2} \right]$$

Mediana

Com Variáveis Discretas

Ex:

Dadas as seguintes idades de alunos de uma classe do segundo ano noturno de uma escola, determine a mediana:

15, 15, 16, 16, 16, 16, 16, 16, 16, 16, 16, 17, 17, 17, 17, 17, 17,
17, 18, 18, 18, 18, 18, 18, 22, 22, 22, 22, 22, 26, 26, 26, 26, 26, 26

Para facilitar a compreensão, iremos organizar os dados numa tabela por frequência:

Yi	Fi	Fac
15	2	2
16	9	11
17	7	18
18	6	24
22	5	29
26	6	35
Σ	35	119

Podemos observar que $n = 35$, ou seja, é **ímpar**, logo:

$$\tilde{Y} = \frac{n + 1}{2} = \frac{35 + 1}{2} = 18$$

Ou seja, a mediana está na posição 18, logo, pela tabela de frequência acumulada observamos que a mediana é **17**.



Mediana com R: Exemplo 1

Com Variáveis Discretas

A função `median()`, retorna a mediana de um vetor.

Obs: Ela ordena o vetor passado como parâmetro

```
>
> y <- c(15, 15, 16, 16, 16, 16, 16, 16, 16, 16, 16, 16, 17, 17, 17, 17, 17, 17,
17, 18, 18, 18, 18, 18, 18, 18, 22, 22, 22, 22, 22, 22, 26, 26, 26, 26, 26, 26)
>
> table(y)
y
15 16 17 18 22 26
 2  9  7  6  5  6
>
> median(y)
[1] 17
> |
```

Mediana

Com Variáveis Discretas

Ex 2:

Dadas as seguintes idades de alunos de uma classe do segundo ano noturno de uma escola, determine a mediana:

15, 15, 16, 16, 16, 16, 16, 16, 16, 16, 16, 17, 17, 17, 17, 17, 17,
17, 18, 18, 18, 18, 18, 18, 22, 22, 22, 22, 22, 26, 26, 26, 26, 26,
26, 27.

Para facilitar a compreensão, iremos organizar os dados numa tabela por frequência:

Yi	Fi	Fac
15	2	2
16	9	11
17	7	18
18	6	24
22	5	29
26	6	35
27	1	36
Σ	36	155

Podemos observar que $n = 36$, ou seja, é **par**. logo:

$$\tilde{Y} = Media \left[\frac{n}{2}, \frac{n+1}{2} \right] \quad \frac{n}{2} = \frac{36}{2} = 18$$

$$\frac{n+1}{2} = \frac{36+1}{2} = 18,5$$

Observamos que em $n/2$ obtemos 18, que pela Fa, a idade é 17. E com $n+1/2$ obtemos 18,5, que pela tabela de Fa fica entre o 18 e o 24, logo a idade é 18. E para obter a mediana fazemos a média de 17 e 18 que resulta em **17,5**.



Mediana com R: Exemplo 2

Com Variáveis Discretas

```
>
> y <- c(15, 15, 16, 16, 16, 16, 16, 16, 16, 16, 16, 16, 17, 17, 17, 17, 17, 17,
17, 18, 18, 18, 18, 18, 18, 18, 22, 22, 22, 22, 22, 26, 26, 26, 26, 26, 26, 27)
>
> table(y)
y
15 16 17 18 22 26 27
 2  9  7  6  5  6  1
>
> median(y)
[1] 17.5
> |
```

Mediana

Com Variáveis Contínuas

Para variáveis contínuas, aplicamos a fórmula de interpolação linear e identificamos pela FAC a classe que contém a mediana:

$$\tilde{Y} = l + \frac{\frac{n}{2} - \Sigma f_{ant}}{F_{md}} \times (L - l)$$

L e l: Limite superior e inferior da classe md

F_{md}: Frequência da classe md.

Σf_{ant} : Soma das frequências anteriores à classe md

n: Tamanho da série

Mediana

Com Variáveis Contínuas

Dadas as seguintes idades de pessoas que foram em um determinado consultório médico, determine a mediana:

Idades	Fi	Fac
0 + 10	70	70
10 + 20	20	90
20 + 30	15	105
30 + 40	15	120
40 + 50	12	132
50 + 60	18	150
60 + 70	50	200
Σ	200	867

Primeiro calcula-se a ordem, aplicando: $\frac{n}{2} = \frac{200}{2} = 100$; a classe da Mediana é verificada pela Fac, nesse caso, a classe onde $Fac = 105$.

20 + 30
15
105

Mediana

Com Variáveis Contínuas

Idades	Fi	Fac
0 + 10	70	70
10 + 20	20	90
20 + 30	15	105
30 + 40	15	120
40 + 50	12	132
50 + 60	18	150
60 + 70	50	200
Σ	200	867

Com a fórmula da interpolação linear:

$$\tilde{Y} = l + \frac{\frac{n}{2} - \Sigma f_{ant}}{F_{md}} \times (L - l) =$$

$$20 + \frac{\frac{200}{2} - 90}{15} \times (30 - 20) =$$

$$\tilde{Y} = 20 + \frac{10}{15} \times 10 = 20 + \frac{100}{15} = 26,66$$

l e L : 20 e 30

F_{md} : 15

Σf_{ant} : 90

n : 200



Mediana com R

Com Variáveis Contínuas

A função `median()`, também aceita objetos do tipo fdt:

```
>
> tb1 <- make.fdt(f=c(70, 20, 15, 15, 12, 18, 50), start=0, end=70)
>
> tb1
  Class limits  f   rf rf(%)  cf cf(%)
  [0,10)  70 0.35  35.0   70  35.0
  [10,20)  20 0.10  10.0   90  45.0
  [20,30)  15 0.08   7.5  105  52.5
  [30,40)  15 0.08   7.5  120  60.0
  [40,50)  12 0.06   6.0  132  66.0
  [50,60)  18 0.09   9.0  150  75.0
  [60,70)  50 0.25  25.0  200 100.0
>
> median(tb1)
[1] 26.66667
> |
```

Mediana

Vantagens

- Fácil de compreender e aplicar;
- Não é afetada por valores extremos;
- É um valor único;
- Pode ser determinada nas escalas: ordinal, intervalar e proporcional.

Desvantagens

- É difícil de ser incluída em expressões matemáticas
- Não usa todos os valores da série.

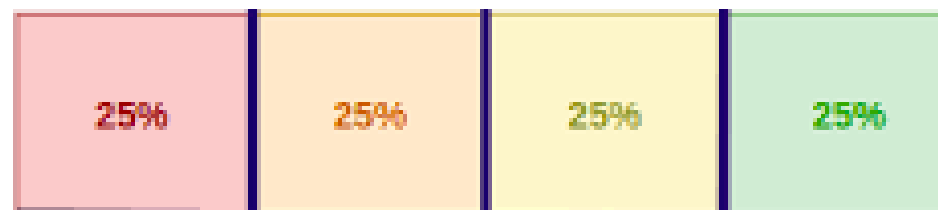
Posição ou Separatrizes

Quartis, Decis e Percentis

Posição ou Separatrizes

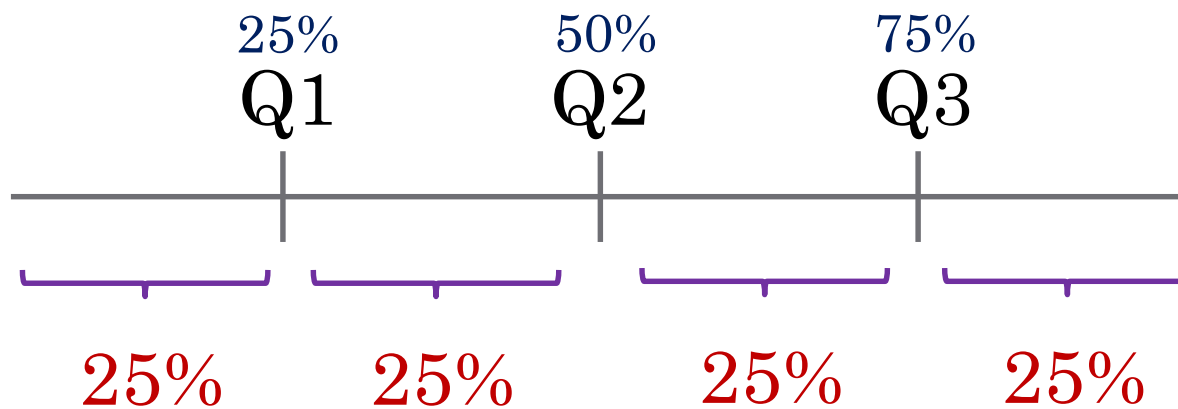
- Permitem determinar valores que particionam a série de **n** observações em partes iguais.
- Genericamente denominadas de **quantis**.

Quartis



Quartis

Quartis são medidas separatrizes que dividem os dados de um rol em 4 grupos que contém 25% do número de dados cada.



Notação adotada: (Q) para o parâmetro e (q) para a estimativa.

Quartis

Com Dados Brutos

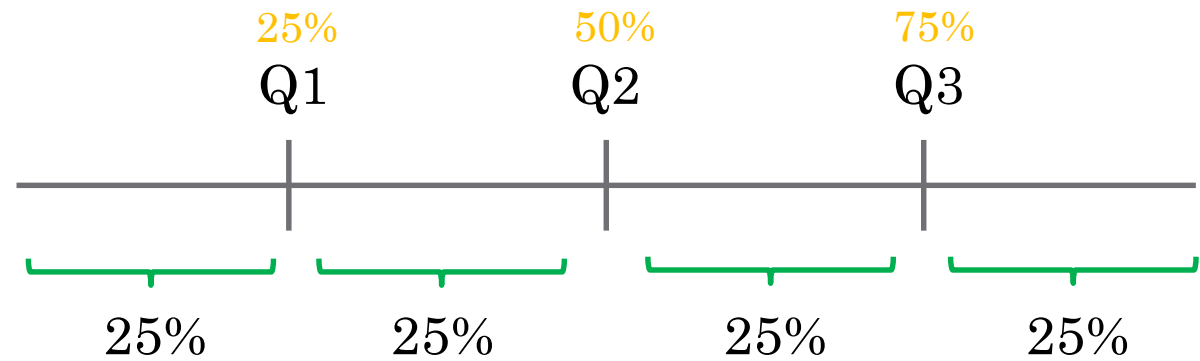
Para achar a posição do quartil desejado:

$$POS_{Qi} = \frac{i \times n}{4}$$

i: Quartil
n: Tamanho da série

Quartis: Exemplo 1

Com Dados Brutos



- Encontre os quartis da série:

$$Q1 = 6$$

{ 5, 6, 6, 9, 10, 13, 15 }

$$Q2 = 9$$

{ 5, 6, 6, 9, 10, 13, 15 }

$$Q3 = 13$$

{ 5, 6, 6, 9, 10, 13, 15 }

$$POS_{Q1} = \frac{1 \times 7}{4} = 1,75$$

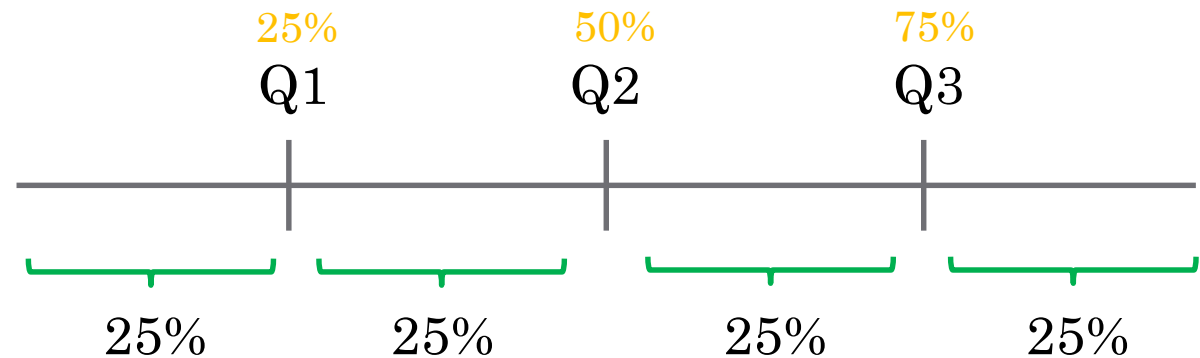
$$POS_{Q2} = \frac{2 \times 7}{4} = 3,5$$

$$POS_{Q3} = \frac{3 \times 7}{4} = 5,25$$

Quartis: Exemplo 2

Com Dados Brutos

- Encontre os quartis da série:



$$Q1 = \frac{2+3}{2} = \frac{5}{2} = 2,5 \quad \{ 1, 1, \textcircled{2, 3}, 5, 5, 6, 7, 9, 9, 10, 13 \} \quad POS_{Q1} = \frac{1 \times 12}{4} = 3$$

$$Q2 = \frac{5+6}{2} = \frac{11}{2} = 5,5 \quad \{ 1, 1, 2, 3, 5, \textcircled{5, 6}, 7, 9, 9, 10, 13 \} \quad POS_{Q2} = \frac{2 \times 12}{4} = 6$$

$$Q3 = \frac{9+9}{2} = \frac{18}{2} = 9 \quad \{ 1, 1, 2, 3, 5, 5, 6, 7, \textcircled{9, 9}, 10, 13 \} \quad POS_{Q3} = \frac{3 \times 12}{4} = 9$$



Quartis com R: Exemplo 2

Com Dados Brutos

A função `quantile()`, retorna os quartis referente ao vetor passado por argumento.

```
>  
> k <- c(1, 1, 2, 3, 5, 5, 6, 7, 9, 9, 10, 13)  
> quantile(k)[2:4]  
 25%  50%  75%  
2.75 5.50 9.00  
>
```

Quartis

Com Variáveis Contínuas

$$Q_i = lqi + \frac{\left(\frac{i \times n}{4} - \sum f\right) \times h}{Fqi}$$

lqi : Limite inferior da classe Q_i

i : Quartil (1, 2, 3)

$\sum f$: Soma das frequências anteriores à classe Q_i

n : Tamanho da série

h : Amplitude da classe Q_i

Fqi : Frequência da classe Q_i

Quartis: Exemplo

Com Variáveis Contínuas

- Encontre os quartis Q1 e Q2:

$$q1 \in \frac{1 \times 40}{4} = 10$$

lqi : Limite inferior da classe Q_i

i : Quartil (1, 2, 3)

$\sum f$: Soma das frequências anteriores à classe Q_i

n : Tamanho da série

h : Amplitude da classe Q_i

Fqi : Frequência da classe Q_i

Idades	50 ┤ 54	54 ┤ 58	58 ┤ 62	62 ┤ 66	66 ┤ 70	70 ┤ 74
f_i	4	9	11	8	5	3
f_a	4	13	24	32	37	40

$$Q_i = lqi + \frac{\left(\frac{i \times n}{4} - \sum f\right) \times h}{Fqi} = 54 + \frac{(10 - 4) \times 4}{9} = 54 + \frac{24}{9} = 56,66$$

Quartis:

Com Variáveis Contínuas

- Encontre os quartis Q1 e Q2:

$$q2 \in \frac{2 \times 40}{4} = 20$$

lqi : Limite inferior da classe Q_i

i : Quartil (1, 2, 3)

$\sum f$: Soma das frequências anteriores à classe Q_i

n : Tamanho da série

h : Amplitude da classe Q_i

Fqi : Frequência da classe Q_i

Idades	50 ┤ 54	54 ┤ 58	58 ┤ 62	62 ┤ 66	66 ┤ 70	70 ┤ 74
fi	4	9	11	8	5	3
fa	4	13	24	32	37	40

$$Q2 = lqi + \frac{\left(\frac{i \times n}{4} - \sum f\right) \times h}{Fqi} = 58 + \frac{(20 - 13) \times 4}{11} = 58 + \frac{28}{11} = 60,54$$



Quartis com R: Exemplos

Com Variáveis Contínuas

```
> k <- make.fdt(f=c(4, 9, 11, 8, 5, 3), start=50, end=74)
> k
```

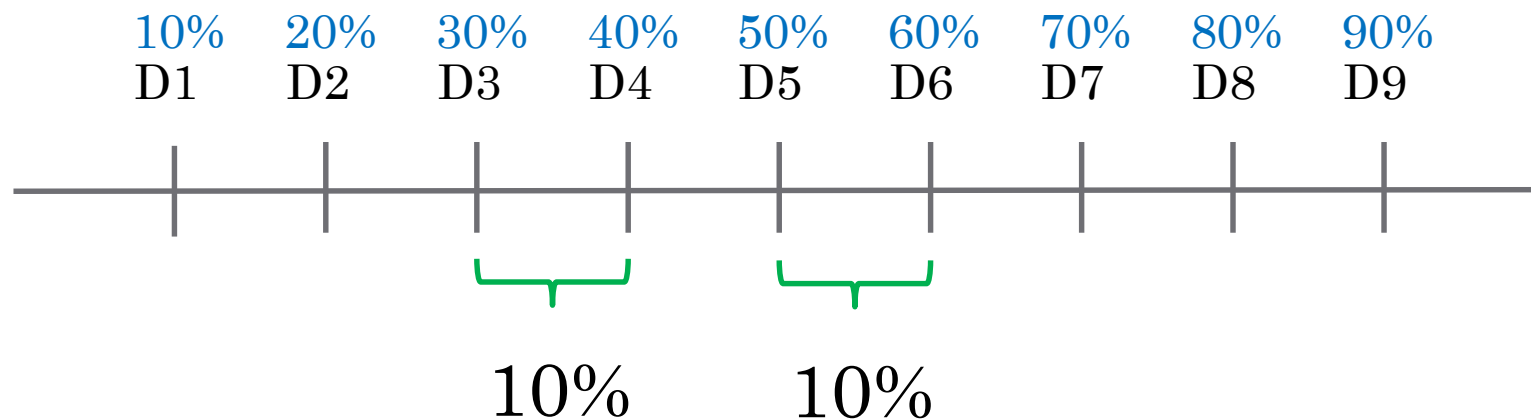
Class	limits	f	rf	rf(%)	cf	cf(%)
[50,54)		4	0.10	10.0	4	10.0
[54,58)		9	0.22	22.5	13	32.5
[58,62)		11	0.28	27.5	24	60.0
[62,66)		8	0.20	20.0	32	80.0
[66,70)		5	0.12	12.5	37	92.5
[70,74)		3	0.08	7.5	40	100.0

```
>
> quantile(k, i=1)
[1] 56.66667
>
> quantile(k, i=2)
[1] 60.54545
>
```

Decis

Decis

- Decis são medidas separatrizes que dividem os dados de um rol em 10 grupos que contém 10% do número de dados cada.
- Notação adotada: (D) para o parâmetro e (d) para a estimativa.



Decis

Com Dados Brutos

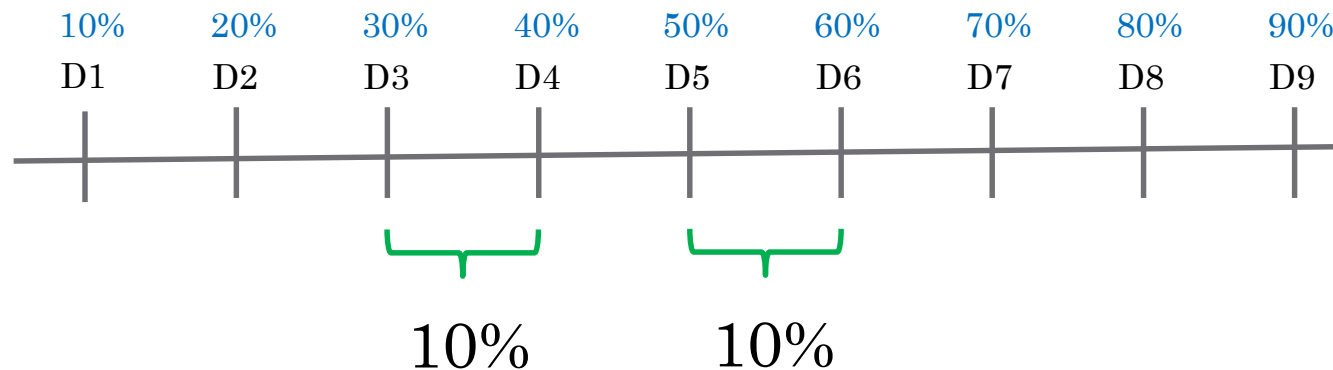
- Para achar a posição do decil desejado:

$$POS_{Di} = \frac{i \times n}{10}$$

i : Decis
 n : Tamanho da série

Decis: Exemplo 1

Com Dados Brutos



- Encontre os decis da série:

{ 1, 3, 5, 6, 6, 9, 10, 13, 15, 20, 22, 24, 25, 29, 31, 31, 39, 40, 55, 90, 91}

{ 1, 3, 5, 6, 6, 9, 10, 13, 15, 20}

{24, 25, 29, 31, 31, 39, 40, 55, 90, 91}

$$D2 = 6$$

$$POS_{D2} = \frac{2 \times 21}{10} = 4,2$$

$$D5 = 22$$

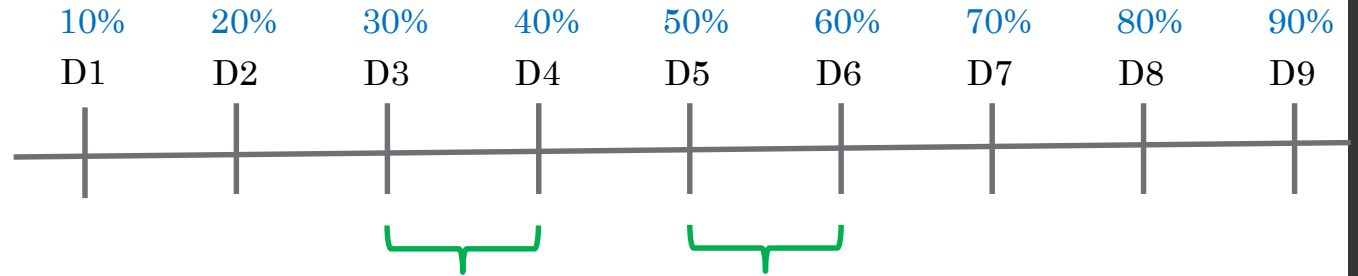
$$POS_{D5} = \frac{5 \times 21}{10} = 10,5$$

$$D7 = 31$$

$$POS_{D7} = \frac{7 \times 21}{10} = 14,7$$

Decis: Exemplo 2

Com Dados Brutos



- Encontre os decis da série:

{ 1, 3, 5, 6, 6, 9, 10, 13, 15, 20, 22, 24, 25, 29, 31, 31, 39, 40, 55, 90 }

$$D5 = \frac{22+20}{2} = \frac{42}{2} = 21$$

$$D2 = \frac{6+6}{2} = \frac{12}{2} = 6$$

$$D7 = \frac{29+31}{2} = \frac{60}{2} = 30$$

$$POS_{D2} = \frac{2 \times 20}{10} = 4$$

$$POS_{D5} = \frac{5 \times 20}{10} = 10$$

$$POS_{D7} = \frac{7 \times 20}{10} = 14$$



Decis com R: Exemplo 1

Com Dados Brutos

A função `quantile()`, retorna os decis referente ao vetor passado por argumento.

```
>  
> k <- c(1, 3, 5, 6, 6, 9, 10, 13, 15, 20, 22, 24, 25, 29, 31, 31, 39, 40, 55, 90, 91)  
>  
> length(k)  
[1] 21  
>  
> quantile(k, p=seq(0, 1, .1))[2:10]  
10% 20% 30% 40% 50% 60% 70% 80% 90%  
   5   6  10  15  22  25  31  39  55  
>
```

Decis

Com Variáveis Contínuas

$$Di = ldi + \frac{(\frac{i \times n}{10} - \Sigma f) \times h}{Fdi}$$

ldi: Limite inferior da classe Di

i: Decis (1, 2, ... , 9)

Σf : Soma das frequências anteriores à classe Di

n: Tamanho da série

h: Amplitude da classe Di

Fdi: Frequência da classe Di

Decis: Exemplo

Com Variáveis Contínuas

$$d3 \in \frac{3 \times 40}{10} = 12$$

- Encontre os decis D3:

Idades	50 ┆ 54	54 ┆ 58	58 ┆ 62	62 ┆ 66	66 ┆ 70	70 ┆ 74
fi	4	9	11	8	5	3
fa	4	13	24	32	37	40

$$Di = ldi + \frac{\left(\frac{i \times n}{10} - \sum f\right) \times h}{Fdi} = 54 + \frac{(12 - 4) \times 4}{9} = 54 + \frac{32}{9} = 57,55$$



Decis com R: Exemplo

Com Variáveis Contínuas

```
> tb = make.fdt(f=c(4, 9, 11, 8, 5, 3), start=50, end=74)
> tb
```

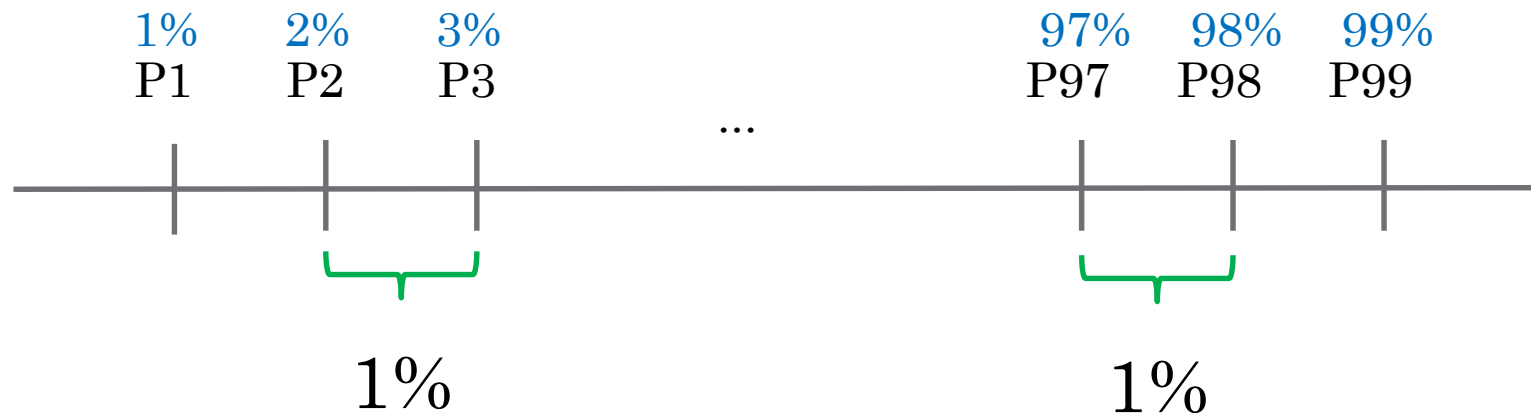
Class	limits	f	rf	rf(%)	cf	cf(%)
	[50,54)	4	0.10	10.0	4	10.0
	[54,58)	9	0.22	22.5	13	32.5
	[58,62)	11	0.28	27.5	24	60.0
	[62,66)	8	0.20	20.0	32	80.0
	[66,70)	5	0.12	12.5	37	92.5
	[70,74)	3	0.08	7.5	40	100.0

```
>
> quantile(tb, i=3, probs=seq(0, 1, 0.1))
[1] 57.55556
>
> |
```

Percentis

Percentis

- Percentis são medidas separatrizes que dividem os dados de um rol em 100 grupos que contém 1% do número de dados cada.
- Notação adotada: (P) para o parâmetro e (p) para a estimativa.



Percentis

Com Dados Brutos

- Para achar a posição do percentis desejado:

$$POS_{Pi} = \frac{i \times n}{100}$$

i: Percentis
n: Tamanho da série

Percentis: Exemplo

Com Dados Brutos

Em uma seleção para determinada vaga de emprego a empresa pretende levar para próxima etapa, candidatos com pontuações acima de 70º (30% melhores) percentil, de acordo com as notas abaixo:

31 31 37 40 48 50 51 51 60 62 64 65 65 65 66 74 74 88 91 92

$$POS_{P70} = \frac{70 \times 20}{100} = 14$$

$$P70 = \frac{65+66}{2} = \frac{131}{2} = 65,5$$

$$P70 = 65,5$$



Percentis com R: Exemplo

Com Dados Brutos

```
> k <- c(31, 31, 37, 40, 48, 50, 51, 51, 60, 62, 64, 65, 65, 65, 66, 74, 74,  
88, 91, 92)
```

```
> quantile(k, prob=seq(0, 1, .01))[71]
```

```
70%
```

```
65.3
```

Percentis

Com Variáveis Contínuas

$$P_i = l_{pi} + \frac{(\frac{i \times n}{100} - \sum f) \times h}{F_{pi}}$$

l_{pi}: Limite inferior da classe P_i

i: Percentis (1, 2, ... , 98, 99)

$\sum f$: Soma das frequências anteriores à classe P_i

n: Tamanho da série

h: Amplitude da classe P_i

F_{pi}: Frequência da classe P_i

Percentis

Com Variáveis Contínuas

Calcular P_{75} , na seguinte distribuição dos salários (em salários mínimos) para uma amostra de 20 funcionários.

Salários	4 † 8	8 † 12	12 † 16	16 † 20	20 † 24	Σ
f _i	5	7	4	2	2	20
f _a	5	12	16	18	20	

$$\frac{75 \times 20}{100} = 15$$

1

$$POS_{P75} = \frac{75 \times 20}{100} = 15$$

2

$$Pi = lpi + \frac{(\frac{i \times n}{100} - \Sigma f) \times h}{Fpi}$$

$$P75 = 12 + \frac{(15 - 12) \times 4}{4} = 12 + \frac{12}{4} = 12 + 3 = 15$$



Percentis com R: Exemplo

Com Variáveis Contínuas

```
> library(fdth)

> tb <- make.fdt(f=c(5, 7, 4, 2, 2), start=4, end=24)

> tb
Class limits f   rf rf(%) cf cf(%)
  [4,8) 5 0.25   25  5   25
  [8,12) 7 0.35   35 12   60
 [12,16) 4 0.20   20 16   80
 [16,20) 2 0.10   10 18   90
 [20,24) 2 0.10   10 20  100

> for(i in 75)
+ print(quantile(tb, i=i, probs=seq(0, 1, .01)))
[1] 15
```

Muito Obrigado!



Referências

- Canal de Estatística, professora Zuleica Pires Lage, disponível em: <https://www.youtube.com/user/Zuzinhah/>. Acesso em: 20 set. 2018.
- FARIA, JOSÉ CLÁUDIO, Notas de aulas expandidas: CET 018 – Elementos de estatística, UESC. Disponível em: http://nbcgib.uesc.br/lec/download/faria/apostilas/CET018_10ed_1pf.pdf, Acesso em: 19 set. 2018.
- Média Aritmética, Wikipédia, disponível em: https://pt.wikipedia.org/wiki/M%C3%A9dia_aritm%C3%A9tica, Acesso em: 21 set. 2018.
- Média Geométrica, Wikipédia, disponível em: https://pt.wikipedia.org/wiki/M%C3%A9dia_geom%C3%A9trica, Acesso em: 21 set. 2018.
- Média Harmônica, Wikipédia, disponível em: https://pt.wikipedia.org/wiki/M%C3%A9dia_harm%C3%B4nica, Acesso em: 21 set. 2018.