

# MAT02018 - Estatística Descritiva

## Medidas de tendência central (continuação)

Rodrigo Citton P. dos Reis  
citton.padilha@ufrgs.br

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA  
DEPARTAMENTO DE ESTATÍSTICA

Porto Alegre, 2022

# Mediana

# Mediana

- ▶ A **mediana**, geralmente representada por ***md***, é o valor de ***x***, em uma **série ordenada de dados**, que divide a série em dois subgrupos de igual tamanho.
- ▶ Considere que observamos as notas de cinco alunos:
  - ▶ André: nota 5,0;
  - ▶ Carla: nota 5,5;
  - ▶ Eliana: nota 8,5;
  - ▶ Júlio: nota 7,0;
  - ▶ Pedro: nota 8,0.

# Mediana

- ▶ Para obter a mediana desde conjunto de dados, primeiro ordenamos de maneira crescente as notas observadas:

5, 0; 5, 5; 7, 0; 8, 0; 8, 5

- ▶ A mediana é o valor que está no centro, ou seja, 7, 0<sup>1</sup>.
- ▶ Como interpretação, temos que metade das notas da turma estão abaixo de 7, e metade estão acima de 7.

---

<sup>1</sup>A mediana é também definida por alguns autores como o **valor que ocupa a posição central** de um **conjunto de dados ordenados**.

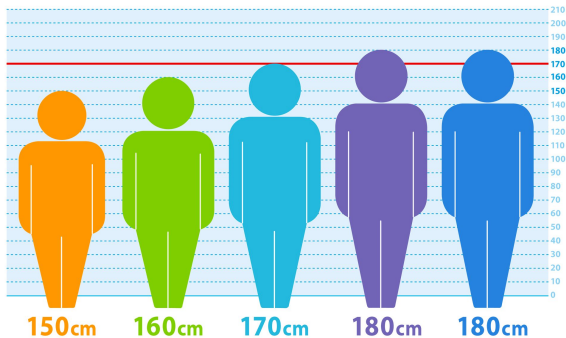
# Mediana

- ▶ Uma característica importante da mediana é a de que ela **não é afetada** pelos extremos da série.
- ▶ No exemplo anterior, se observássemos a nota **3,5** para o aluno André e **9,0** para a aluna Eliana, a mediana continuaria sendo o valor **7,0**:

**3,5; 5,5; 7,0; 8,0; 9,0**

# Mediana

- Para calcular a mediana, **ordenamos os dados** para que se possa identificar em que posição ela se localiza.



# Mediana

- ▶ Em grandes conjuntos de dados, a **posição da mediana**, ou seja, a posição central na versão ordenada deste conjunto, é encontrada facilmente por intermédio do seguinte cálculo:

$$\frac{n + 1}{2}.$$

- ▶ **Exemplo:** em uma amostra de 35 medidas de estatura ( $n = 35$ ), a mediana é o o valor que encontra-se no  $(35+1)/2 = 18$  da **série dos dados ordenados**<sup>2</sup>.

---

<sup>2</sup>Lembre: para encontrarmos a mediana de uma distribuição é necessário ordenarmos os dados do conjunto observado.

# Mediana

- ▶ Quando o conjunto contiver um **número par** de elementos, a **mediana é a média dos dois valores centrais** (do conjunto ordenado).
  - ▶ Assim, se observamos os seguintes valores para uma certa variável:

1, 3, 7, 98

- ▶ A mediana está na posição 2,5<sup>3</sup>, e portanto, a mediana é a média dos valores centrais  $md = (3 + 7)/2 = 5^4$ .

---

<sup>3</sup>De maneira mais formal, a mediana está entre as posições 2 e 3 da série de dados ordenados.

<sup>4</sup>**Sua vez!** Calcule a mediana de altura da sua casa.



## Mediana em tabelas de frequências com agrupamento simples

- Considere mais uma vez o exemplo do **número de carburadores** em um conjunto de 32 modelos de automóveis.

Número de carburadores ( $x_i$ )	$n_i$
1	7
2	10
3	3
4	10
6	1
8	1
Total	32

# Mediana em tabelas de frequências com agrupamento simples

- ▶ Note que os valores já estão ordenados.
- ▶ Como o conjunto possui **32** elementos, e portanto  **$n$  é par**, a fórmula  $(n + 1)/2 = (32 + 1)/2$  nos diz que a mediana está entre as posições **16** e **17**.
  - ▶ A mediana será a média dos valores destas duas posições.
- ▶ A frequência acumulada ( $n_{ac}$ ) pode nos ajudar.

## Mediana em tabelas de frequências com agrupamento simples

Número de carburadores ( $x_i$ )	$n_i$	$n_{ac}$
1	7	7
2	10	17
3	3	20
4	10	30
6	1	31
8	1	32
Total	32	-

## Mediana em tabelas de frequências com agrupamento simples

- Percebemos que ambas as posições assumem o valor 2, e assim,  $md = (2 + 2)/2 = 2$ .

## Mediana em tabelas de frequências com agrupamento simples

- Uma forma alternativa para se obter a mediana quando os dados estão agrupados é usar a frequência acumulada relativa ( $f_{ac}$ ).
- O valor de  $x$  para o qual  $f_{ac} = 0,5$  é a mediana, pois metade dos valores é igual ou menor  $f_{ac}^{-1}(x)$ .

Número de carburadores ( $x_i$ )	$n_i$	$n_{ac}$	$f_i$	$f_{ac}$
1	7	7	0,22	0,22
2	10	17	0,31	0,53
3	3	20	0,09	0,62
4	10	30	0,31	0,94
6	1	31	0,03	0,97
8	1	32	0,03	1,00
Total	32	-	1,00	-

## Mediana em tabelas de frequências com agrupamento simples

- ▶ Na tabela acima, 22% são iguais ou menores que 1 e 53% são iguais ou menores do que 2; logo, a mediana é 2.

## Mediana em tabelas de frequências com dados agrupados por intervalos

Quando os dados estiverem organizados em intervalos de classe, os valores individuais não podem ser identificados.

Nesse caso, pode-se **estimar** a mediana usando a seguinte expressão:

$$md = LIR_{md} + h \left( \frac{n/2 - n_{ac}^{(ant)}}{n_{md}} \right),$$

em que

- ▶  $LIR_{md}$ : limite inferior real do intervalo que contém a mediana;
- ▶  $h$ : amplitude do intervalo;
- ▶  $n$ : tamanho da amostra;
- ▶  $n_{ac}^{(ant)}$ : frequência absoluta acumulada no intervalo anterior ao que contém a mediana;
- ▶  $n_{md}$ : frequência absoluta simples no intervalo que contém a mediana.

## Mediana em tabelas de frequências com dados agrupados por intervalos

- ▶ Assim, como nos casos anteriores, o primeiro passo é encontrar a posição do valor central do conjunto de dados ordenados pela fórmula  $(n + 1)/2$ , e em seguida, encontrar a classe em que a mediana se encontra utilizando a frequência acumulada.
- ▶ Como exemplo, utilizaremos os dados de idades de 30 crianças de uma escola.

Idade (anos)	$n_i$	$n_{ac}$
5,5 └ 6,5	1	1
6,5 └ 7,5	20	21
7,5 └ 8,5	7	28
8,5 └ 9,5	2	30
Total	30	-



## Mediana em tabelas de frequências com dados agrupados por intervalos

- ▶ Neste exemplo, a mediana está entre as posições **15** e **16**, pois  $(30 + 1)/2 = 15,5$ .
- ▶ Esse valor encontra-se no intervalo **6,5** ─ **7,5**, porque ali estão desde o 2º até o 20º valor deste conjunto de dados (ordenados).
- ▶ Então, estima-se a idade mediana como:

$$md = 6,5 + 1,0 \times \left( \frac{30/2 - 1}{20} \right) = 7,2.$$

- ▶ Este resultado nos informa que metade das crianças desta escola, que coletamos informação, tem idade superior a 7,2 anos.
- ▶ **Sua vez!** Calcule a mediana do **Teste de Desempenho Escolar** (TDE; exercício da aula passada).

# Moda

# Moda

- ▶ A **moda** é o **valor mais frequente** (que apresenta a maior frequência) de uma série de valores<sup>5</sup>.
- ▶ **Exemplo:** considere que em uma certa quadra da cidade observamos as cores das faixadas das casas:

*amarela, laranja, laranja, amarela, verde, amarela, laranja, verde, laranja, laranja*

- ▶ A cor laranja é a **moda** nesta quadra, pois apresenta a maior frequência (5).

---

<sup>5</sup>Note que esta medida de tendência central se aplica a dados qualitativos, pois se refere às frequências da distribuição. Já a média se aplica apenas para dados quantitativos. Pense por um minuto na média da variável *cor dos olhos*. Faz algum sentido?

# Moda

- ▶ **Exemplo:** considere novamente o exemplo do número de carburadores.
  - ▶ Temos que 10 carros apresentam dois carburadores e 10 carros apresentam quatro carburadores.
  - ▶ Assim, temos **duas modas**, ou dizemos que a distribuições do **número de carburadores** é **bimodal**, com uma moda em **2** e outra em **4**.

# Moda

- ▶ Quando os dados estão apresentados em intervalos de classe, costuma-se indicar um **intervalo modal**.
- ▶ No exemplo de **milhas por galão** a classe  $14,3 \vdash 18,2$  é o intervalo modal, pois apresenta maior frequência (10).
- ▶ **Sua vez!** Encontre o intervalo modal do exemplo do **Teste de Desempenho Escolar** (TDE; exercício da aula passada).

# ComplementaR

# ComplementaR



Esta seção é complementar. São apresentadas algumas poucas funções em R relacionadas a discussão da aula. Para tal, vamos utilizar o exemplo original de (BUSSAB; MORETTIN, 2017) sobre os dados dos empregados da seção de orçamentos da Companhia MB. A planilha eletrônica correspondente encontra-se no arquivo `companhia_mb.xlsx`. Vamos começar carregando os dados para o R. Existem várias formas de se carregar **arquivos de dados** em diferentes no R. Como arquivo de interesse encontra-se no formato do Excel (xlsx), vamos utilizar a função `read_excel` do pacote `readxl`<sup>6</sup>.

---

<sup>6</sup>Caso você não tenha o pacote, instale-o: `install.packages("readxl")`.

# ComplementaR

```
# install.packages("readxl")  
library(readxl)  
  
dados <- read_excel(path = "companhia_mb.xlsx")
```

```
class(dados) # classe do objeto dados
```

```
## [1] "tbl_df"      "tbl"        "data.frame"
```

```
dim(dados) # dimensão do objeto dados
```

```
## [1] 36  7
```

► Note que o objeto dados é uma tabela de dados bruto.

```
head(dados) # apresenta as primeiras linhas do objeto dados
```



## ComplementaR

```
## # A tibble: 6 x 7
##       N `Estado Civil` `Grau de Instrução` `N de Filhos` `Salario (x Sa~` Id
##   <dbl> <chr>          <chr>                <dbl>          <dbl> <d
## 1     1   1 solteiro      ensino fundamental          NA           4
## 2     2   2 casado       ensino fundamental           1          4.56
## 3     3   3 casado       ensino fundamental           2          5.25
## 4     4   4 solteiro      ensino médio                NA          5.73
## 5     5   5 solteiro      ensino fundamental          NA          6.26
## 6     6   6 casado       ensino fundamental           0          6.66
```

- A função `mean` calcula a média aritmética de um conjunto de dados.

```
mean(dados$Idade)
```

```
## [1] 34.58333
```

```
mean(dados$`Salario (x Sal Min)`)
```

```
## [1] 11.12222
```

## ComplementaR

- ▶ No caso em que a variável possui **dados ausentes**, como é o caso da variável *número de filhos*, precisamos utilizar o argumento `na.rm = TRUE` para remover os dados ausentes do conjunto antes de calcular a média.

```
mean(dados$`N de Filhos`, na.rm = TRUE)
```

```
## [1] 1.65
```

- ▶ A mediana de um conjunto de dados pode ser obtida utilizando a função `median`.

```
median(dados$Idade)
```

```
## [1] 34.5
```

```
median(dados$`Salario (x Sal Min)`)
```

```
## [1] 10.165
```

```
median(dados$`N de Filhos`, na.rm = TRUE)
```

```
## [1] 2
```

## ComplementaR

- ▶ A moda pode ser obtida observando-se a maior frequência em uma tabela de frequências.

```
table(dados$`Estado Civil`)
```

```
##  
##   casado solteiro  
##      20       16
```

- ▶ Casado é a moda na seção de orçamentos da Companhia MB. Outra possibilidade é recuperar a maior frequência do vetor de frequências utilizando a função `which.max`.

```
which.max(table(dados$`Região de Procedência`))
```

```
## outra  
##      3
```

```
which.max(table(dados$`N de Filhos`))
```

```
## 2  
## 3
```

## Para casa

1. Resolver os exercícios 1 a 3 do Capítulo 8.5 do livro **Fundamentos de Estatística**<sup>7</sup> (disponível no Sabi+).
2. Para o seu levantamento estatístico, calcule médias, modas e medianas, de acordo com a classificação das variáveis. Compartilhe no Fórum Geral do Moodle.
3. Ler os slides sobre “Médias ponderada, geométrica e harmônica”.

---

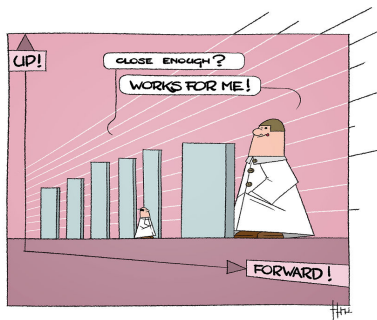
<sup>7</sup>Vieira, S. **Fundamentos de Estatística**, Atlas, 2019, p. 135-136.

# Próxima aula

- ▶ Medidas de variabilidade.

# Por hoje é só!

Bons estudos!



BUSSAB, W. de O.; MORETTIN, P. A. **Estatística básica**. 9. ed. São Paulo: Saraiva, 2017.