



Engenharia Civil/Mecatrônica

Estatística e Probabilidade

Aula 2

Prof. Guilherme Sada Ramos

Instituto Federal de Santa Catarina/ Câmpus Criciúma

25/26 de março de 2021



Dados agrupados

Quando a amostra coletada tem muitos dados, estes podem ser descritos, para efeito de análise, em intervalos.

Um critério para se estabelecer esse agrupamento é utilizando a Fórmula de Stürges:

$$k = 1 + 3,3 \log n$$

- k : quantidade de intervalos
- n : quantidade de dados



Exemplo: Medindo as alturas de 40 pessoas, obtivemos os resultados (em cm):

| | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 169 | 187 | 152 | 153 | 198 | 185 | 186 | 188 | 169 | 175 |
| 176 | 170 | 163 | 182 | 183 | 192 | 193 | 196 | 184 | 179 |
| 178 | 175 | 159 | 185 | 158 | 176 | 158 | 175 | 163 | 160 |
| 190 | 180 | 175 | 169 | 196 | 158 | 175 | 162 | 155 | 185 |

Pela fórmula, o número de intervalos é $1 + 3,3 \log 40 = 6,29 \approx 6$.
Para saber a amplitude (tamanho) dos intervalos, diminuimos a menor da maior altura ($198 - 152 = 46$) e dividimos pelo número de intervalos.

$$\frac{46}{6} = 7,67 \approx 8$$



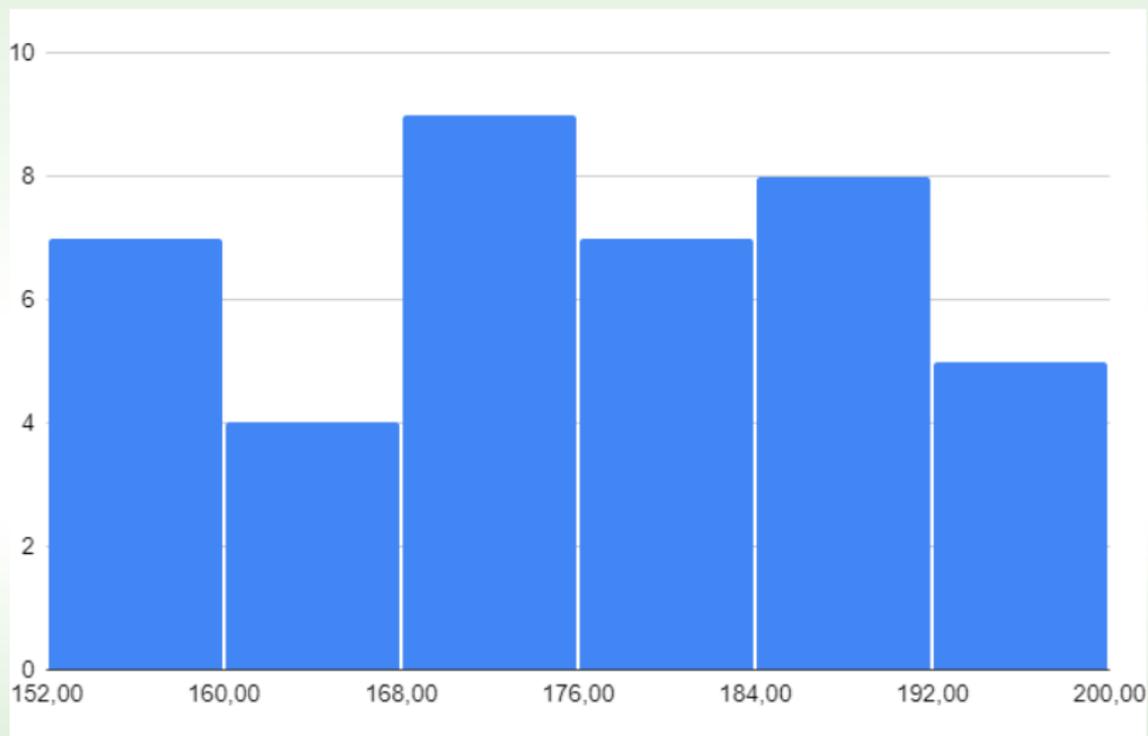
Os intervalos ficam assim:

- $152 \mapsto 160$ (7 pessoas)
- $160 \mapsto 168$ (4 pessoas)
- $168 \mapsto 176$ (9 pessoas)
- $176 \mapsto 184$ (7 pessoas)
- $184 \mapsto 192$ (8 pessoas)
- $192 \mapsto 200$ (5 pessoas)

$$a \mapsto b = [a, b[= \{x \in \mathbb{R} : a \leq x < b\}$$



INSTITUTO FEDERAL
Santa Catarina





Medidas estatísticas

Medidas de centralidade

Média aritmética simples

Se uma amostra tem n dados, designados por $x_1, x_2, x_3, \dots, x_n$, a média aritmética destes dados é:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$$

Exemplo: calcular a média dos dados

51 61 53 48 68 49 59 61 62 60 55

$$\bar{x} = \frac{\sum_{i=1}^{11} x_i}{11} = \frac{51 + 61 + \cdots + 60 + 55}{11} = \frac{627}{11} = 57$$



Exemplo: calcular a média dos dados

51 61 53 48 68 49 59 61 62 60 55

$$\bar{x} = \frac{\sum_{i=1}^{11} x_i}{11} = \frac{51 + 61 + \dots + 60 + 55}{11} = \frac{627}{11} = 57$$

Propriedades da média aritmética:

- Adicionando a cada um dos dados um número $p > 0$, a média aritmética fica aumentada em p unidades.
- Multiplicando a cada um dos dados por um número $q > 0$, a média aritmética fica multiplicada por q .
- Retirando um valor menor, ou acrescentando um valor maior que a média ao conjunto, esta média *aumenta*.
- Retirando um valor maior, ou acrescentando um valor menor que a média ao conjunto, esta média *diminui*.



Média aritmética ponderada

Em um conjunto de dados, algum deles pode aparecer mais de uma vez. Neste caso, atribuímos a ele um *peso*.

Se, em um conjunto de k dados x_1, x_2, \dots, x_k , considerarmos suas frequências absolutas (aqui chamadas pesos) n_1, n_2, \dots, n_k , respectivamente, a média aritmética ponderada é:

$$\bar{x} = \frac{\sum_{i=1}^k x_i n_i}{\sum_{i=1}^k n_i} = \frac{x_1 n_1 + x_2 n_2 + \dots + x_k n_k}{n_1 + n_2 + \dots + n_k}$$



Exemplo: Considere que, em uma determinada empresa, a renda dos trabalhadores seja:

- R\$ 2.500,00 para 12 deles;
- R\$ 4.000,00 para 5 deles;
- R\$ 5.900,00 para 2 deles;
- R\$ 9.600,00 para 1 deles.

Neste caso, qual é a renda média dos trabalhadores desta empresa?

$$\bar{x} = \frac{12 \times 2500 + 5 \times 4000 + 2 \times 5900 + 1 \times 9600}{12 + 5 + 2 + 1} = \frac{71400}{20}$$

$$\bar{x} = 3570$$



Uma outra maneira de se calcular é utilizando as frequências relativas:

$$\bar{x} = \sum_{i=1}^k x_i f_i = x_1 f_1 + x_2 f_2 + \dots + x_k f_k$$

No exemplo anterior, temos:

- $\frac{12}{20} = 0,60$ para R\$ 2.500,00
- $\frac{5}{20} = 0,25$ para R\$ 4.000,00
- $\frac{2}{20} = 0,10$ para R\$ 5.900,00
- $\frac{1}{20} = 0,05$ para R\$ 9.600,00

$$\bar{x} = \sum_{i=1}^4 x_i f_i = 2500 \times 0,60 + 4000 \times 0,25 + 5900 \times 0,10 + 9600 \times 0,05 = 3570$$

Mediana

Dados em rol: são dados escritos em ordem crescente, ou decrescente!

Exemplo: o conjunto

51 61 53 48 68 49 59 61 62 60 55

com os dados colocados em ordem crescente, fica

48 49 51 53 55 59 60 61 61 62 68

Chamamos de mediana o valor central desta distribuição de dados em rol. No caso, o valor mediano é 59.



Se o número de dados for par, tomamos a média aritmética dos dois valores centrais como valor mediano da amostra.

Exemplo: Acrescentando o dado 57 ao conjunto do slide anterior, o rol fica:

48 49 51 53 55 57 59 60 61 61 62 68

Os valores centrais são 57 e 59. A mediana no caso é $\frac{57+59}{2} = 58$.



De um modo geral, para dados x_1, x_2, \dots, x_n , com $x_1 \leq x_2 \leq \dots \leq x_n$, definimos mediana como sendo:

$$Me = \begin{cases} x_{\frac{n+1}{2}}, & \text{se } n \text{ é ímpar} \\ \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}, & \text{se } n \text{ é par.} \end{cases}$$

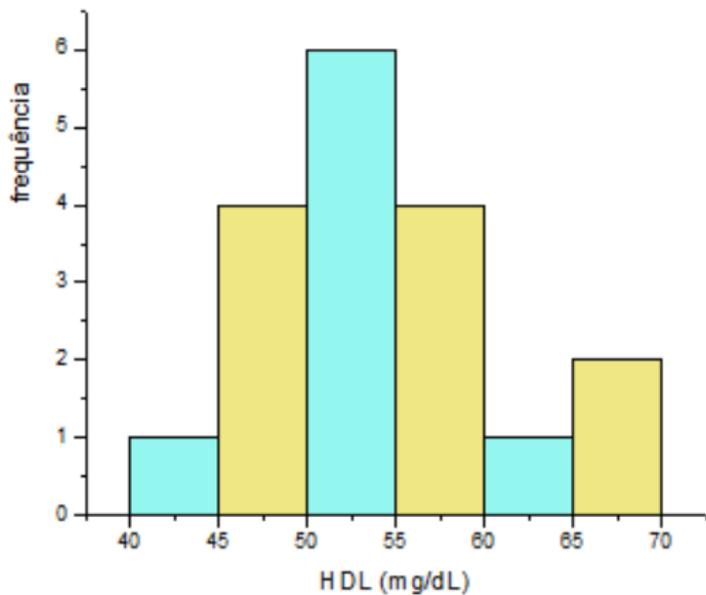
Percentis

O “k-ésimo percentil”, p_k , de uma amostra, é um valor maior ou igual a $k\%$ dos dados da amostra e menor ou igual a $(100 - k)\%$ dos dados do mesmo conjunto.

Exemplo: p_{20} é o 20º percentil: 20% dos dados são menores ou iguais a ele, e $(100 - 20 = 80)\%$ estão acima ou iguais.



INSTITUTO FEDERAL
Santa Catarina





Moda

Em um conjunto de dados, a moda é aquele dado com maior frequência absoluta. É o que dado que mais aparece no conjunto.

No conjunto de dados

3 4 5 4 4 2 1 9 7 5

o elemento que aparece com maior frequência é 4. Dizemos, portanto, que 4 é a moda deste conjunto de dados.



Para efeito de comparação, em

48 49 51 53 55 57 59 60 61 61 62 68

temos:

- média 57;
- mediana 58;
- moda 61.