Medidas Descritivas

- Medidas de localização, posição ou tendência central
- Medidas separatrizes
- Medidas de variação ou dispersão
- Medidas de formato

Medidas Descritivas

Objetivo → reduzir um conjunto de dados numéricos a um pequeno grupo de valores que deve fornecer toda a informação relevante a respeito desses dados

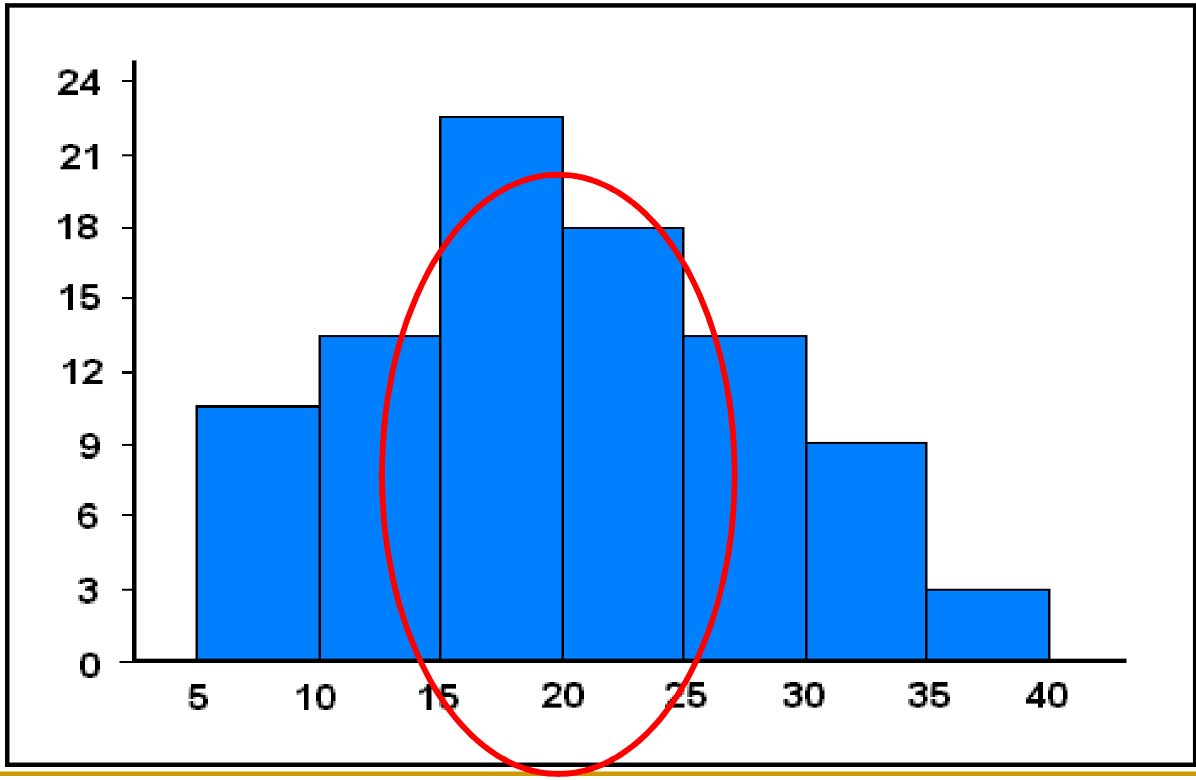
⇨ Existe uma grande variedade de medidas descritivas
Como escolher a mais adequada?

- **Com que objetivo a medida está sendo obtida?**
- **Existem valores atípicos que podem afetá-la exageradamente?**
- **O propósito da análise é meramente descritivo ou planeja-se fazer inferências?**

Medidas de localização

- Também denominadas medidas de tendência central ou medidas de posição.

- Indicam um ponto central onde, geralmente, está localizada a maioria das observações.

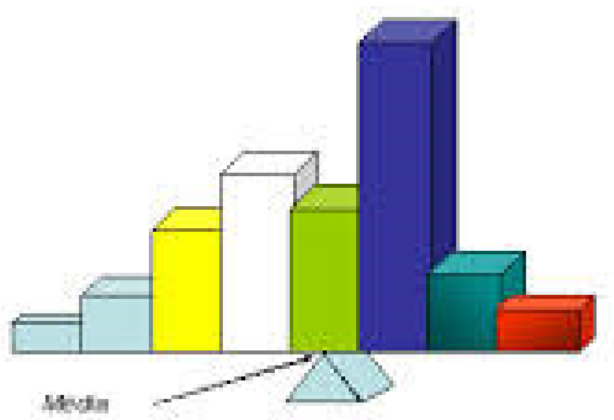


Medidas de localização

Objetivo → representar o ponto de equilíbrio ou o centro ou o ponto de concentração de uma distribuição

Medidas de localização mais utilizadas:

- Média aritmética
- Mediana
- Moda



Média aritmética

Medida mais conhecida e utilizada:

- ✓ facilidade de cálculo e de compreensão
- ✓ propriedades matemáticas e estatísticas

Média aritmética
- Simples → todos os valores participam do cálculo com o mesmo peso
- Ponderada → pelo menos um dos valores participa com peso diferente

Média aritmética simples

Para um conjunto de n valores:

Quando falarmos simplesmente **média** – é dela que estamos falando

$x_i = x_1, x_2, \ldots, x_n$

$$\bar{x} = \frac{\sum x_i}{n}$$

soma de todos os valores

total de valores somados

Exemplo:

X = tempo para realizar uma tarefa (min)

$x_i = 4, 5, 7, 9, 10$

$$\bar{x} = \frac{4+5+7+9+10}{5} = \frac{35}{5} = 7 \text{ min}$$

Média aritmética ponderada

Temos um conjunto de valores e um conjunto de pesos:

$x_i = x_1, x_2, \ldots, x_n$

$p_i = p_1, p_2, \ldots, p_n$

$$\bar{x}_p = \frac{\sum x_i\, p_i}{\sum p_i}$$

soma de produtos de valores e pesos

soma dos pesos

Exemplo:

X = nota

$x_i = 7,\ 8,\ 6,\ 10$

$p_i = 2, 2, 1, 1$

$$\bar{x}_p = \frac{7\times2+8\times2+6\times1+10\times1}{2+2+1+1}$$

$$\bar{x}_p = \frac{46}{6} = 7{,}67$$

Propriedades da média aritmética

1ª propriedade: A média de um conjunto de dados que não varia, ou seja, cujos valores são uma constante, é a própria constante.

Verificação numérica:

$$x_i = 7, 7, 7, 7, 7$$

$$\overline{x} = 7$$

2ª propriedade: Ao somar uma constante c por todos os valores de um conjunto de dados, sua média também é somada por esta constante.

Verificação numérica:

$x_i = 4, 5, 7, 9, 10 \qquad \bar{x} = 7$

Somar c=2

$x_i + 2 = 6, 7, 9, 11, 12$

$$\bar{x}_{x+2} = \frac{\sum (x_i + 2)}{n} = \frac{6+7+9+11+12}{5} = \frac{45}{5} = 9$$

$$\boxed{\begin{aligned} \bar{x}_{x+2} &= 7 + 2 \\ \bar{x}_{x+c} &= \bar{x} + c \end{aligned}}$$

3ª propriedade: Ao multiplicar uma constante c por todos os valores de um conjunto de dados, sua média também é multiplicada por esta constante.

Verificação numérica:

x_i = 4, 5, 7, 9, 10 $\qquad \bar{x} = 7$

Multiplicar por c=2

$2x_i$ = 8, 10, 14, 18, 20

$$\bar{x}_{2x} = \frac{\sum 2x_i}{n} = \frac{8+10+14+18+20}{5} = \frac{70}{5} = 14$$

$$\boxed{\begin{aligned} \bar{x}_{2x} &= 2 \times 7 \\ \bar{x}_{cx} &= c\bar{x} \end{aligned}}$$

4ª propriedade: A soma de todos os desvios em relação à média de um conjunto de valores é nula.

$$\sum \underbrace{(x_i - \bar{x})}_{\text{desvio}} = 0$$

diferença entre a observação e a média aritmética

Verificação numérica:

x_i = 4, 5, 7, 9, 10

$\bar{x} = 7$

$$(x_i - \bar{x}) \begin{cases} 4 - 7 = -3 \\ 5 - 7 = -2 \\ 7 - 7 = 0 \\ 9 - 7 = 2 \\ 10 - 7 = 3 \end{cases} \text{SOMA} = 0$$

5ª propriedade: A soma dos quadrados dos desvios em relação a uma constante c é mínima quando c = $\bar{x}$.

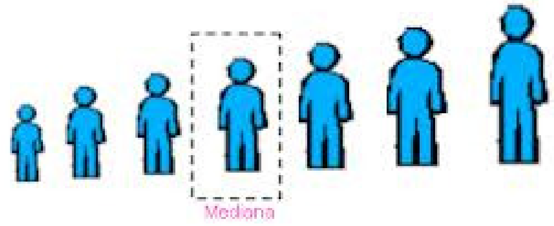
$$\sum (x_i - c)^2 \leftarrow \boxed{\text{é mínima quando } c = \bar{x}}$$

Verificação numérica:

$$\bar{x} = 7$$

i	x_i	$(x_i - \bar{x})^2$	$(x_i - 5)^2$	$(x_i - 10)^2$
1	4	9	1	36
2	5	4	0	25
3	7	0	4	9
4	9	4	16	1
5	10	9	25	0
Σ	35	26	46	71

Mediana (Md)



É a medida que divide um conjunto de dados **ordenado** em duas partes iguais: 50% dos valores ficam abaixo e 50% ficam acima da mediana.

Md

$x_{(1)}$ 50% 50% $x_{(n)}$

Para obter a mediana:

1. Ordenar os dados

2. Determinar a posição (p) da mediana

2. Determinar a posição (p) da mediana

Dois casos:

n ímpar $\rightarrow$ um valor central

posição $\rightarrow$ $p = \dfrac{n+1}{2}$ $Md = x_{(p)}$

n par $\rightarrow$ dois valores centrais

posições $\left\{ p = \dfrac{n+1}{2} \begin{array}{l} \nearrow p_1 \\ \searrow p_2 \end{array} \right.$ $Md = \dfrac{x_{(p_1)} + x_{(p_2)}}{2}$

Exemplo:

X = Tempo para realizar uma tarefa (min)

x_i = 5, 9, 7, 4, 12, 10

1. Ordenar os dados

$x_{(i)} = 4,\ 5,\ \boxed{7,\ 9},\ 10,\ 12$

2. Determinar a posição (p) da mediana

n = 6 (par)

$$p = \frac{n+1}{2} = \frac{6+1}{2} = 3{,}5$$

$p_1 = 3$

$p_2 = 4$

$$Md = \frac{x_{(3)} + x_{(4)}}{2} = \frac{7+9}{2} = 8$$

$$\boxed{Md = 8\ \text{min}}$$

Moda (Mo)

⇨ É o valor de maior ocorrência num conjunto de dados.

⇨ É a única medida que pode não existir e, existindo, pode não ser única.

Exemplos: X = Tempo para realizar uma tarefa (min)

x_i = 12, 8, 7, 5, 7, 4, 8, 8, 9 Mo = 8 min

x_i = 5, 7, 3, 7, 9, 5, 9, 3 não existe Mo (conjunto amodal)

x_i = 9, 5, 4, 5, 7, 1, 2, 2 Mo = 2 e 5 min (conjunto bimodal)

Exercício proposto:

Os valores que seguem são os tempos (em segundos) de reação a um alarme de incêndio, após a liberação de fumaça de uma fonte fixa:

12 9 11 7 9 14 6 10

Para este conjunto de valores, calcule as medidas de posição (média, mediana e moda).

$$\bar{X} = 9{,}75$$

$$Md = 9{,}5$$

$$Mo = 9$$

Média aritmética - características

- No cálculo da média participam todos os valores observados.

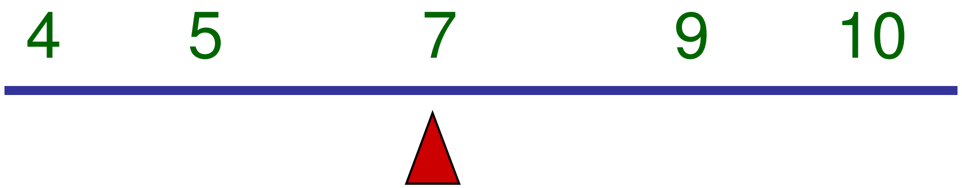
- É uma medida de fácil interpretação e presta-se muito bem a tratamentos estatísticos adicionais.

- É uma medida que sempre existe e é única.

- É o ponto de equilíbrio de uma distribuição, sendo tão mais eficiente quanto mais simétrica for a distribuição dos valores ao seu redor.

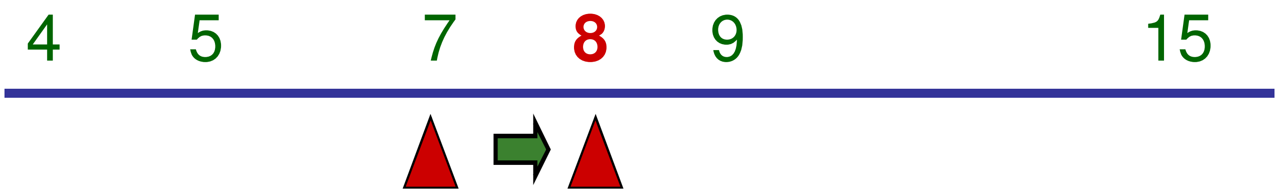
- **Desvantagem:**

  - É uma medida altamente influenciada por valores atípicos ou discrepantes (não resistente).



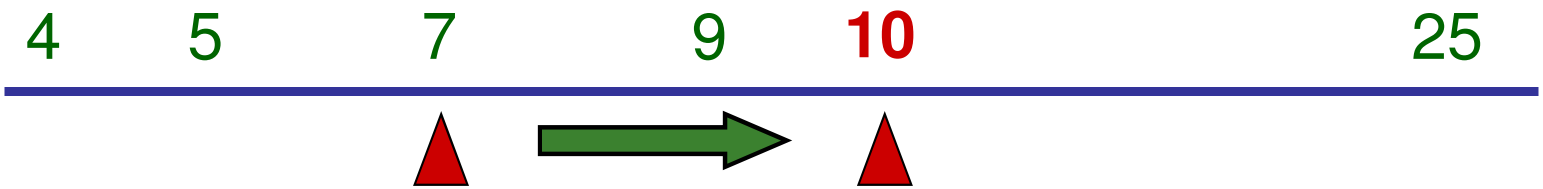
$Md = 7$

$\bar{x} = 7$



$Md = 7$

$\bar{x} = 8$



$Md = 7$

$\bar{x} = 10$

Mediana - características

➢ Define exatamente o centro de uma distribuição, mesmo quando os valores se distribuem assimetricamente em torno da média.

➢ Pode ser determinada mesmo quando não se conhece todos os valores do conjunto de dados.

➢ É uma medida que sempre existe e é única.

➢ É uma medida resistente, ou seja, não sofre influência de valores discrepantes.

Desvantagem:

- ➢ É uma medida que não se presta a cálculos matemáticos.

Moda - características

- É uma medida que têm existência real dentro do conjunto de dados e em grande número de vezes.

- Não exige cálculo, apenas uma contagem.

- Pode ser determinada também para variáveis categóricas.

- **Desvantagem**:
  - É uma medida que não se presta a cálculos matemáticos.
  - Deixa sem representação todos os valores do conjunto de dados que não forem iguais a ela.

Medidas separatrizes

⇨ Indicam limites para proporções de observações em um conjunto

Mediana → divide o conjunto ordenado em duas partes

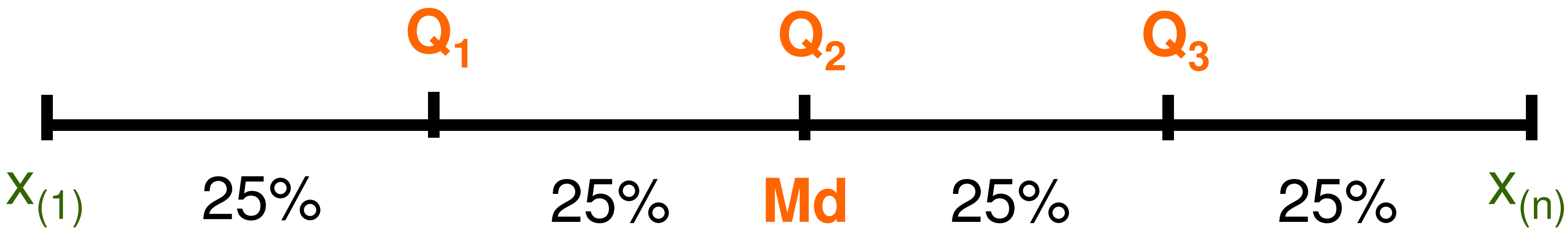
Quartis → dividem o conjunto ordenado em quatro partes

Decis → dividem o conjunto ordenado em dez partes

Percentis → dividem o conjunto ordenado em cem partes

Quartis (Q_i)

⇨ São três medidas que dividem um conjunto de dados ordenado em quatro partes iguais.



Primeiro quartil (Q_1): 25% dos valores abaixo e 75% acima dele

Segundo quartil (Q_2): 50% dos valores abaixo e 50% acima dele

Terceiro quartil (Q_3): 75% dos valores abaixo e 25% acima dele

Para obter os quartis:

1. Ordenar os dados

2. Determinar a posição (p) de cada quartil

Dois casos:

n ímpar

Posição do $Q_1 \rightarrow p_1 = \frac{n+1}{4}$

Posição do $Q_2 \rightarrow p_2 = \frac{2(n+1)}{4}$

Posição do $Q_3 \rightarrow p_3 = \frac{3(n+1)}{4}$

n par

Posição do $Q_1 \rightarrow p_1 = \frac{n+2}{4}$

Posição do $Q_2 \rightarrow p_2 = \frac{2n+2}{4}$

Posição do $Q_3 \rightarrow p_3 = \frac{3n+2}{4}$

$$Q_i = X_{(p_i)}$$

Se p não for inteiro, tomamos os dois inteiros mais próximos.

Exemplo: $p_i = 7,5$ → 7, 8

$$Q_i = \frac{X_{(7)} + X_{(8)}}{2}$$

O quartil será a média aritmética dos dois valores que ocupam essas duas posições.

Exercício proposto:

Foram registrados os tempos de frenagem para 21 motoristas que dirigiam a 30 milhas por hora. Os valores obtidos foram:

69	58	70	80	46	61	65	74	75	55	67
56	70	72	61	66	58	68	70	68	58	

Para o conjunto de valores, calcule os quartis e interprete esses valores.

46	55	56	58	58	58	61	61	65	66	67
68	68	69	70	70	70	72	74	75	80	

$Q_1 = 58$ $Q_2 = 67$ $Q_3 = 70$

Medidas de variação ou dispersão

i	x_i	y_i	z_i
1	4	2	1
2	4	5	8
3	4	4	5
4	4	6	4
5	4	3	2
$\sum$	20	20	20
Média	4	4	4

Medidas de variação ou dispersão

Objetivo → indicar quanto os valores diferem entre si ou quanto eles se afastam da média

⇨ Complementam as medidas de tendência central

Medidas de variação mais utilizadas:

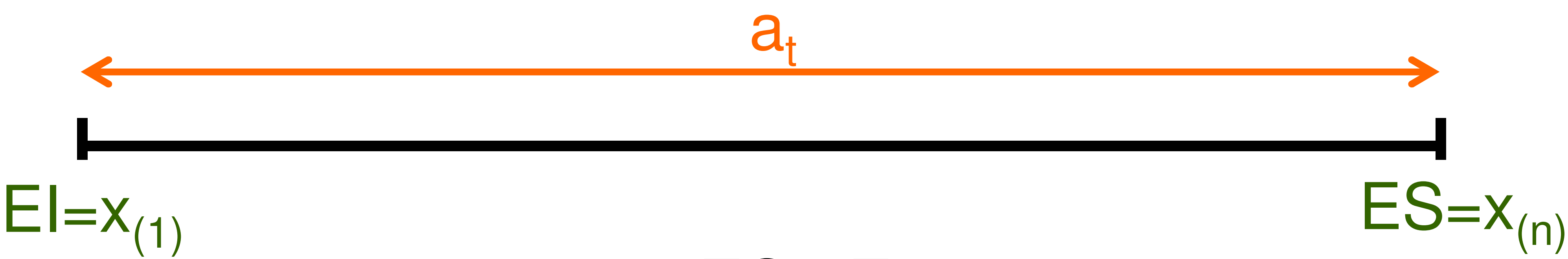
- Amplitude total
- Variância
- Desvio padrão
- Coeficiente de variação

Tendo a cabeça a arder e os pés enterrados no gelo, na média está tudo bem!

Amplitude total (a_t)

⇨ Fornece uma ideia inicial de variação

⇨ É obtida pela diferença entre o maior valor e o menor valor de um conjunto de dados



$$a_t = ES - EI$$

ES: extremo superior do conjunto de dados ordenado
EI: extremo inferior do conjunto de dados ordenado

$$a_t = x_{(n)} - x_{(1)}$$

Exemplo:

X = Tempo para realizar uma tarefa (min)

x_i = 3, 3, 4, 6, 7, 9, 9, 11, 12

$$a_t = ES - EI = 12 - 3 = 9\,\text{min}$$

EI 100% ES

Significado: todos os valores do conjunto de dados diferem, no máximo, em 9 minutos

Desvantagens

- pouco precisa
- extremamente influenciada por valores discrepantes

i	x_i	y_i	z_i
1	4	2	1
2	4	5	8
3	4	4	5
4	4	6	4
5	4	3	2
Σ	20	20	20
Média	4	4	4
a_t	0	4	7

Variância (s^2)

⇨ Medida de variação mais utilizada:

- facilidade de compreensão
- propriedades estatísticas importantes para a inferência

⇨ Leva em conta todos os valores do conjunto de dados

⇨ Considera o desvio da média como unidade básica da variação:

Desvio: $(x_i - \overline{x})$ ← mede quanto cada valor varia em relação à média

Exemplo:

$x_i = 4, 5, 7, 9, 10$

$\bar{x} = 7$

$$(x_i - \bar{x}) \begin{cases} 4 - 7 = -3 \\ 5 - 7 = -2 \\ 7 - 7 = 0 \\ 9 - 7 = 2 \\ 10 - 7 = 3 \end{cases}$$

variação do x_i em relação à média

Média dos desvios → variação média do conjunto de valores

soma de todos os desvios →

número de desvios somados →

$$\frac{\sum (x_i - \bar{x})}{n} = 0$$

4ª propriedade da média → $\sum (x_i - \bar{x}) = 0$

Solução: elevar os desvios ao quadrado $\rightarrow$ desvios negativos ficam positivos e podem ser somados

soma dos quadrados dos desvios

número de desvios somados

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

Média dos quadrados dos desvios

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

número de graus de liberdade ou desvios independentes

Por que utilizar n-1 como denominador?

Porque este denominador confere à variância melhores propriedades estatísticas (importante na inferência estatística).

⇨ Quando o objetivo for apenas descrever a variação de um conjunto de valores, podemos usar o denominador n.

$$s_n^2 = \frac{\sum (x_i - \overline{x})^2}{n}$$

⇨ Quando o objetivo for estimar a variação de uma população por meio da variação de um conjunto de valores (amostra), devemos usar o denominador n-1.

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

Exemplo:

X = tempo para realizar uma tarefa (min)

x_i = 4, 5, 7, 9, 10 $\longrightarrow$ $\bar{x} = 7$ min

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

$$= \frac{(4-7)^2 + (5-7)^2 + (7-7)^2 + (9-7)^2 + (10-7)^2}{5-1}$$

$$= \frac{9+4+0+4+9}{4} = \frac{26}{4} = 6{,}5$$

$$s^2 = 6{,}5 \text{ min}^2$$

unidade de medida fica elevada ao quadrado

Propriedades da variância

1ª propriedade: A variância de um conjunto de dados que não varia, ou seja, cujos valores são uma constante, é zero.

Verificação numérica:

$$x_i = 7, 7, 7, 7, 7 \longrightarrow \bar{x} = 7$$

$$s^2 = \frac{(7-7)^2 + (7-7)^2 + (7-7)^2 + (7-7)^2 + (7-7)^2}{5-1} = 0$$

$$s_c^2 = \frac{\sum (c-c)^2}{n-1} = 0$$

2ª propriedade: Ao somar uma constante **c** a todos os valores de um conjunto de dados, a variância destes dados **não se altera**.

Verificação numérica:

x_i = 4, 5, 7, 9, 10 → $\bar{x}=7$; $s^2 = 6{,}5$

Somar c=2

x_i+2 = 6, 7, 9, 11, 12 → $\bar{x}_{x+2} = 9 \rightarrow \boxed{\bar{x}_{x+c} = \bar{x} + c}$; $s^2_{x+2} = 6{,}5 \rightarrow \boxed{s^2_{x+c} = s^2}$

$$s^2_{x+2} = \frac{(6-9)^2 + (7-9)^2 + (9-9)^2 + (11-9)^2 + (12-9)^2}{5-1}$$

$$= \frac{9+4+0+4+9}{4} = \frac{26}{4} = 6{,}5\,\text{min}^2$$

3ª propriedade: Ao multiplicar todos os valores de um conjunto de dados por uma constante **c**, a variância destes dados fica multiplicada pelo **quadrado** desta constante.

Verificação numérica:

$x_i = 4, 5, 7, 9, 10$ → $\bar{x} = 7$; $s^2 = 6{,}5$

Multiplicar por c=2

$2x_i = 8, 10, 14, 18, 20$ → $\bar{x}_{2x} = 14 \rightarrow \boxed{\bar{x}_{cx} = c\bar{x}}$; $s^2_{2x} = 26 \rightarrow \boxed{s^2_{xc} = c^2 s^2}$

$$s^2_{2x} = \frac{(8-14)^2 + (10-14)^2 + (14-14)^2 + (18-14)^2 + (20-14)^2}{5-1}$$

$$= \frac{36+16+0+16+36}{4} = \frac{104}{4} = 26 \text{ min}^2 = 2^2 \times 6{,}5$$

Trabalhando com a expressão da variância, é possível encontrar uma fórmula mais prática de cálculo:

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

$$= \sum x_i^2 - \sum 2x_i\bar{x} + \sum \bar{x}^2$$

$$= \sum x_i^2 - 2\frac{\sum x_i}{n}\sum x_i + n\left(\frac{\sum x_i}{n}\right)^2$$

$$\boxed{s^2 = \frac{\sum x_i^2 - n\,\bar{x}^2}{n-1}}$$

$$= \sum x_i^2 - 2\frac{\left(\sum x_i\right)^2}{n} + n\frac{\left(\sum x_i\right)^2}{n^2}$$

$$= \sum x_i^2 - 2\frac{\left(\sum x_i\right)^2}{n} + \frac{\left(\sum x_i\right)^2}{n}$$

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

$$s^2 = \frac{\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}}{n-1}$$

Desvantagens da variância:

1. Como a variância é calculada a partir da média, é uma medida pouco resistente, ou seja, muito influenciada por valores atípicos.

2. Como a unidade de medida fica elevada ao quadrado, a interpretação da variância se torna mais difícil.

Para solucionar o problema de interpretação da variância surge outra medida: o desvio padrão.

Desvio padrão (s)

- É definido como a raiz quadrada positiva da variância

$$s = \sqrt{s^2}$$

Exemplo:

X = tempo para realizar uma tarefa (min)

x_i = 4, 5, 7, 9, 10

$$\bar{x} = 7\,\text{min}$$

$$s^2 = 6{,}5\,\text{min}^2$$

$$s = \sqrt{s^2}$$

$$s = \sqrt{6{,}5\,\text{min}^2}$$

$$s = 2{,}55\,\text{min}$$

Apresentação do desvio padrão:

$$\bar{x} \pm s$$

$$7 \pm 2{,}55$$

Tempo médio para realizar a tarefa de 7 minutos com uma variação média de 2,55 minutos acima e abaixo da média.

Significado: variação média em torno da média aritmética

ATENÇÃO: ver propriedades da variância e quais as adaptações necessárias para o desvio padrão!

Interpretação do Desvio Padrão

O desvio padrão é uma forma de medir distância entre as observações de um conjunto de dados em relação a sua média.

Exemplo:

Podemos comparar 85 em um exame de inglês com 80 em um exame de alemão?

Que nota é realmente mais alta?

Um raciocínio rápido mostra que isso depende do desempenho dos outros estudantes em cada turma, ou seja, da variabilidade das notas – do desvio padrão.

Esse mesmo raciocínio de comparação de escores em provas é utilizado nos concursos vestibulares.

Coeficiente de Variação (CV)

- O coeficiente de variação é definido como a proporção (ou percentual) da média representada pelo desvio padrão.

$$CV = \frac{s}{\bar{x}} \; 100\%$$

Exemplo:

X = tempo para realizar uma tarefa (min)

$$x_i = 4,\ 5,\ 7,\ 9,\ 10 \quad \begin{cases} \bar{x} = 7 \text{ min} \\ s = 2{,}55 \text{ min} \end{cases}$$

$$CV = \frac{s}{\bar{x}} \; 100\% = \frac{2{,}55 \text{ min}}{7 \text{ min}} \; 100\% = 36{,}4\%$$

- O CV é a medida mais utilizada para comparar variabilidades de diferentes conjuntos de dados

- Esta comparação não deve ser feita através de qualquer medida de variação em duas situações:
  - quando as médias dos conjuntos comparados são muito desiguais
  - quando as unidades de medida são diferentes

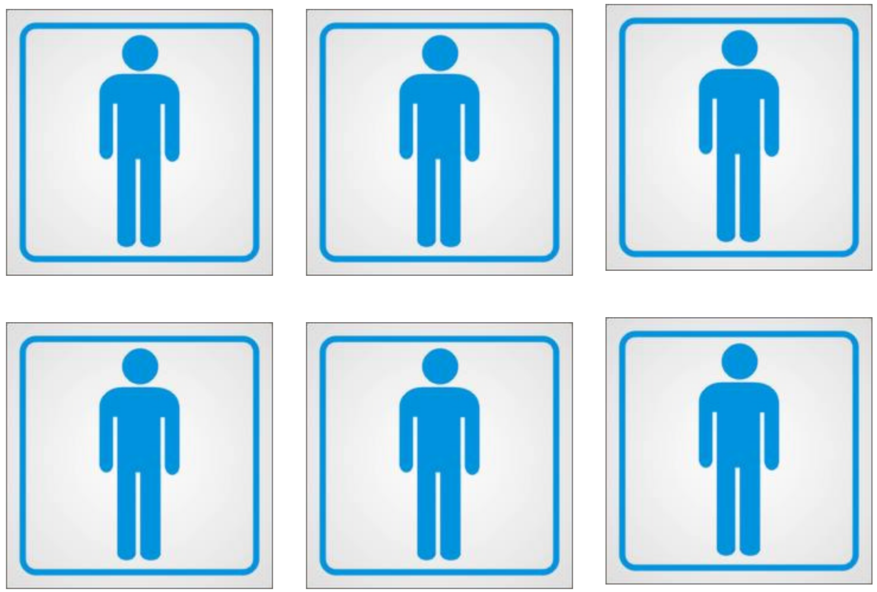
  Nessas situações devemos usar o CV.

Vantagens:

- O CV é desprovido de unidade de medida (expresso em percentagem)
- O CV é uma medida relativa, pois relaciona o desvio padrão com a sua respectiva média aritmética

Exemplo:

Consideremos que x_{1i} e x_{2i} são conjuntos de valores referentes aos salários (R$) de mulheres e homens , para os quais foram obtidas as seguintes medidas:



Mulheres (X_1): $\bar{x}_1 = 1.300$

$s_1 = 340$

Homens (X_2): $\bar{x}_2 = 2.500$

$s_2 = 420$

Qual grupo varia mais em relação aos salários?

$\bar{x}_1 = 1300$

$s_1 = 340$

$\boxed{CV_1 = 26{,}2\%}$

$\bar{x}_2 = 2500$

$s_2 = 420$

$\boxed{CV_2 = 16{,}8\%}$

O maior desvio padrão, quando comparado à sua média, representou menor variação.

Quando as médias são diferentes, devemos usar o CV.

Exercício proposto:

Contou-se o número de vendas de determinado produto durante os sete dias de uma semana, com os seguintes resultados:

14	20	20	20	15	16	18

a) Determine a média, a mediana e a moda.

b) Calcule a variância, o desvio padrão e o coeficiente de variação.

média = 17,6

mediana = 18

moda = 20

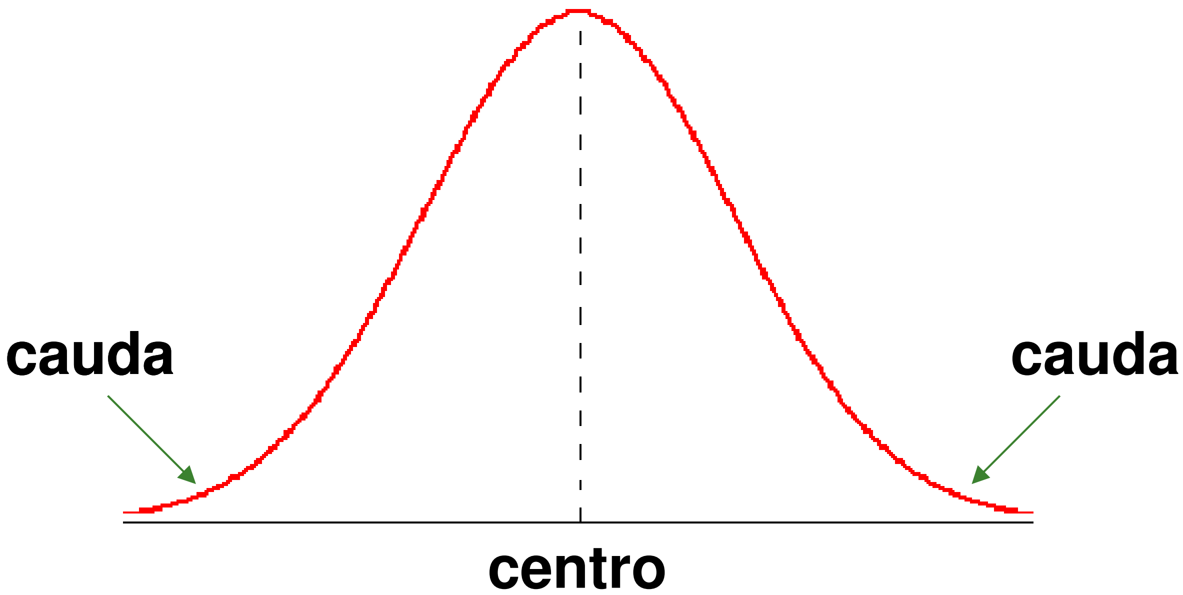
variância = 6,62

desvio padrão = 2,57

CV = 14,64%

Medidas de Formato

- O formato é um aspecto importante de uma distribuição. Está relacionado com as ideias de **simetria** e **curtose**.
- A **simetria** em torno de um eixo indica que o formato da distribuição à esquerda e à direita desse eixo é o mesmo.
- A **curtose** está relacionada com o grau de concentração das observações no centro e nas caudas da distribuição.



Coeficiente de Assimetria

⇨ Informa se a maioria dos valores se localiza à esquerda, ou à direita, ou se estão distribuídos uniformemente em torno da média aritmética.

⇨ O coeficiente de assimetria é calculado a partir do segundo e do terceiro momentos centrados na média:

$$a_3 = \frac{m_3}{m_2\sqrt{m_2}}$$

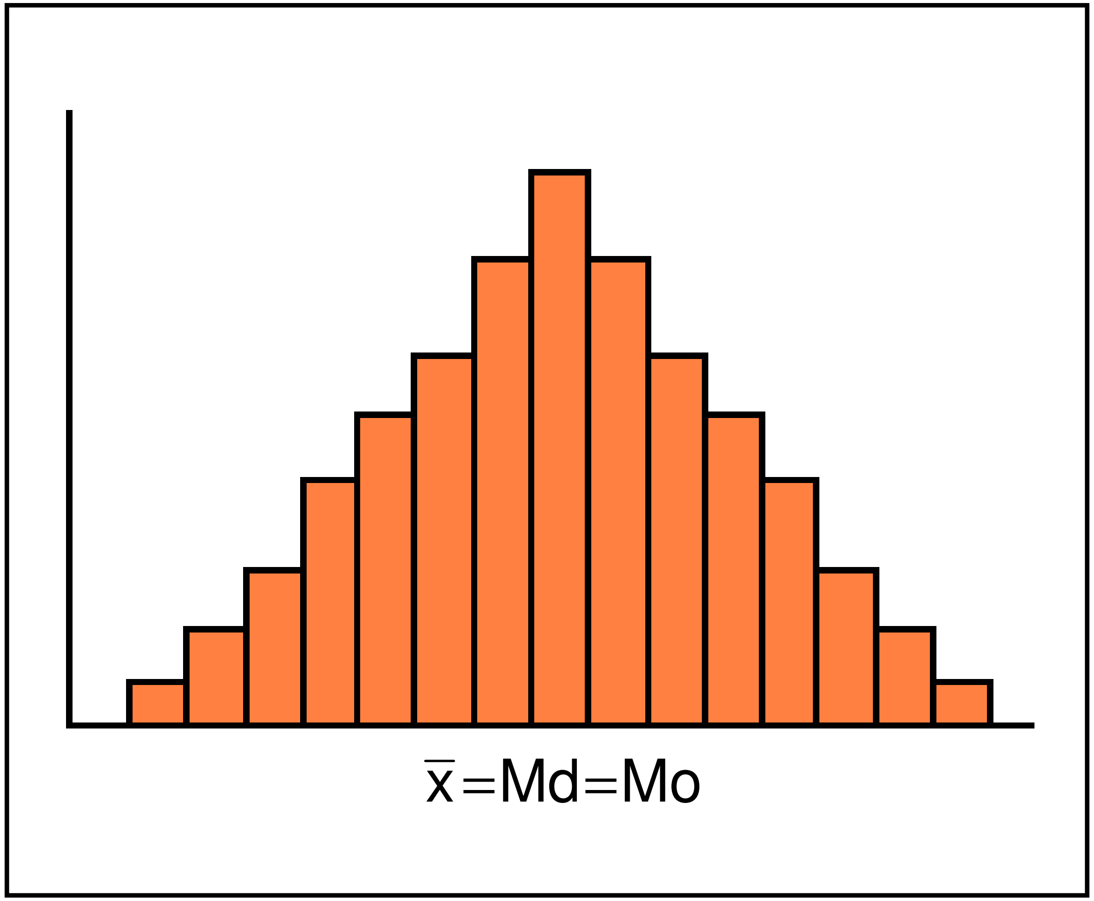
$$m_3 = \frac{\sum (x_i - \overline{x})^3}{n}$$

$$m_2 = \frac{\sum (x_i - \overline{x})^2}{n}$$

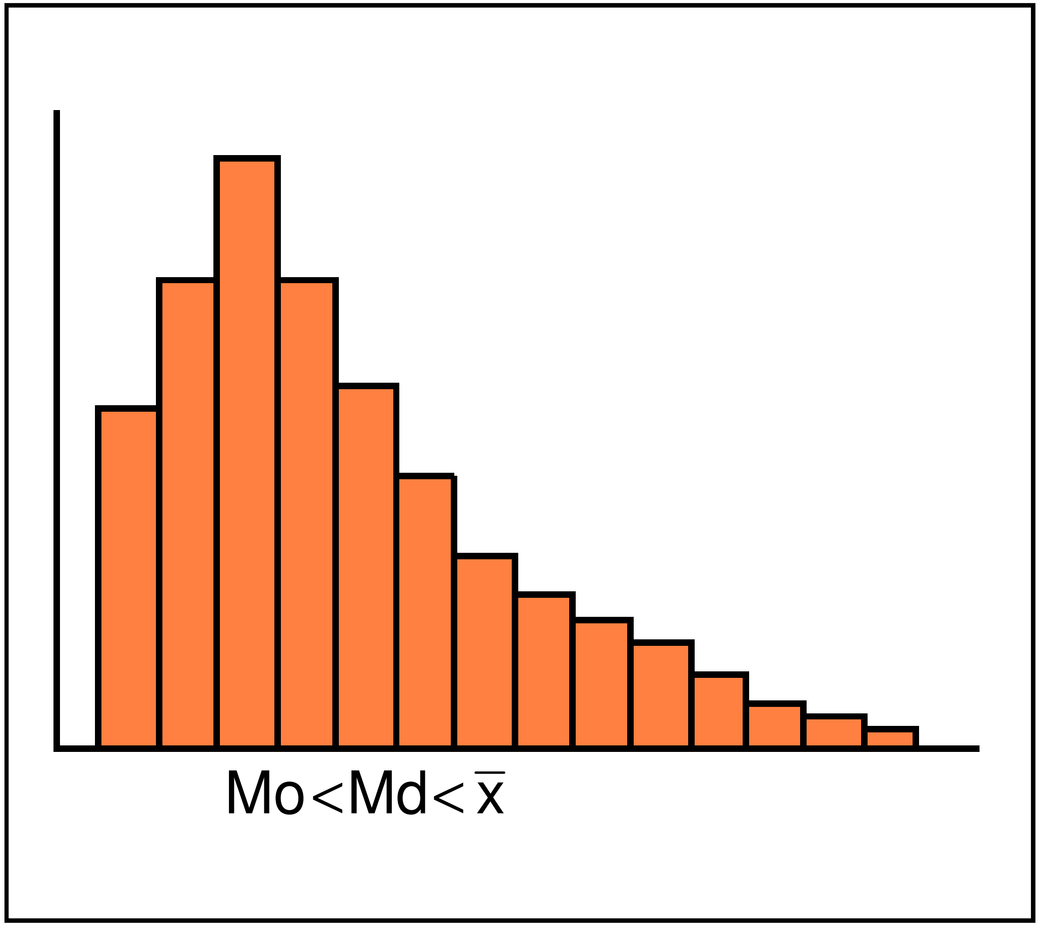
⇨ Indica o grau e o sentido do afastamento da simetria.

Classificação quanto à simetria

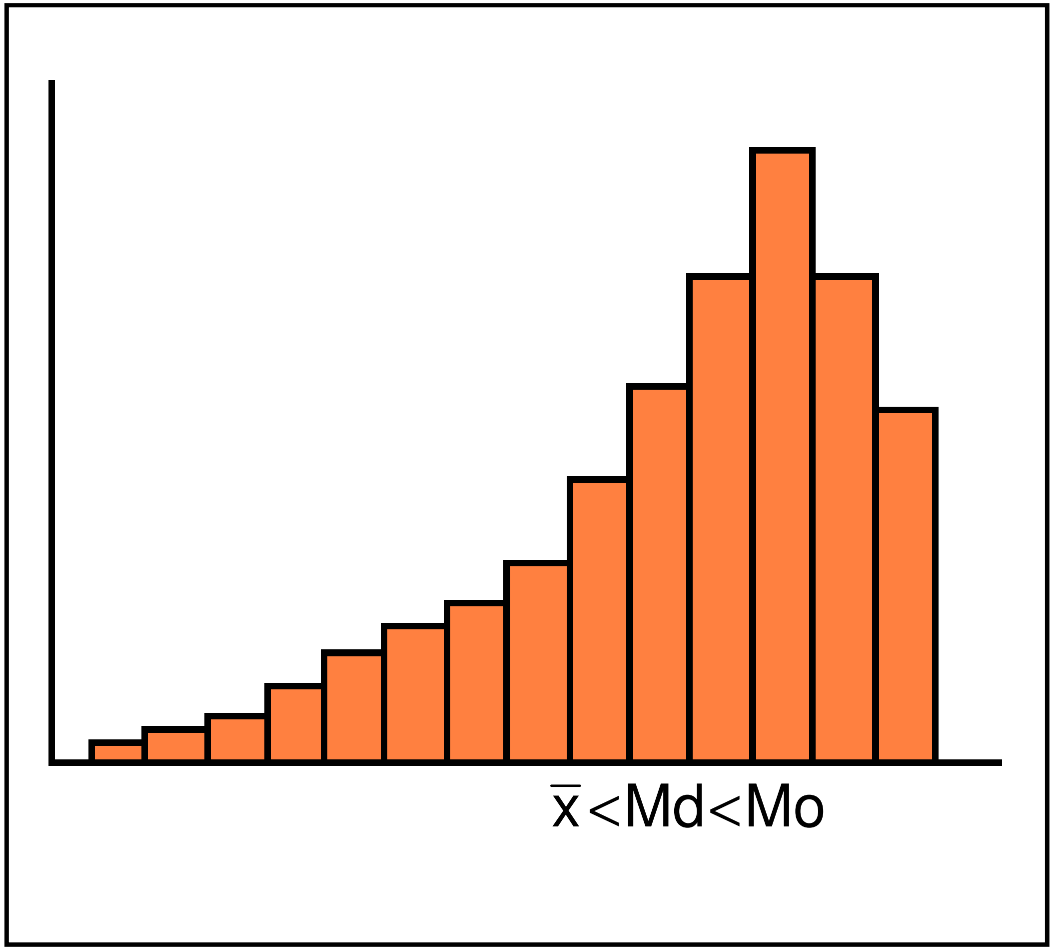
Se a_3=0, a distribuição é classificada como simétrica, indicando que os valores estão uniformemente distribuídos em torno da média.



Se $a_3>0$, a distribuição é classificada como assimétrica positiva, indicando que a maioria dos valores são menores ou se localizam à esquerda da média (cauda para direita).



Se $a_3<0$, a distribuição é classificada como assimétrica negativa, indicando que a maioria dos valores são maiores ou se localizam à direita da média (cauda para esquerda).



Coeficiente de Curtose

⇨ Indica o grau de achatamento de uma distribuição.

⇨ O coeficiente de curtose é calculado a partir do segundo e do quarto momentos centrados na média.

$$a_4 = \frac{m_4}{m_2^2}$$

$$m_4 = \frac{\sum (x_i - \bar{x})^4}{n}$$

$$m_2 = \frac{\sum (x_i - \bar{x})^2}{n}$$

⇨ A classificação é feita tendo por base a curtose que ocorre na distribuição normal.

Classificação quanto à curtose

$a_4 < 3$ ⇒ ***platicúrtica*** ⇒ baixa concentração de valores no centro, tornando a distribuição mais achatada que a distribuição normal.

$a_4 = 3$ ⇒ ***mesocúrtica*** ⇒ concentração das observações ocorre de forma semelhante à da distribuição normal.

$a_4 > 3$ ⇒ ***leptocúrtica*** ⇒ alta concentração de valores no centro, o que provoca um pico maior que o da distribuição normal.

Platicúrtica Mesocúrtica Leptocúrtica

Medidas Descritivas

- Dados não agrupados (medidas exatas)
- Dados agrupados em classe (medidas aproximadas)

3,11	8,88	9,26	10,81	12,69
13,78	15,23	15,62	17,00	17,39
18,36	18,43	19,27	19,50	19,54
20,16	20,59	22,22	23,04	24,47
24,58	25,13	26,24	26,26	27,65
28,06	28,08	28,38	32,03	36,37
38,98	38,64	39,16	41,02	42,97
44,08	44,67	45,40	46,69	48,65
50,39	52,75	54,80	59,07	61,22
70,32	82,70	85,76	86,37	93,34

Dados não agrupados

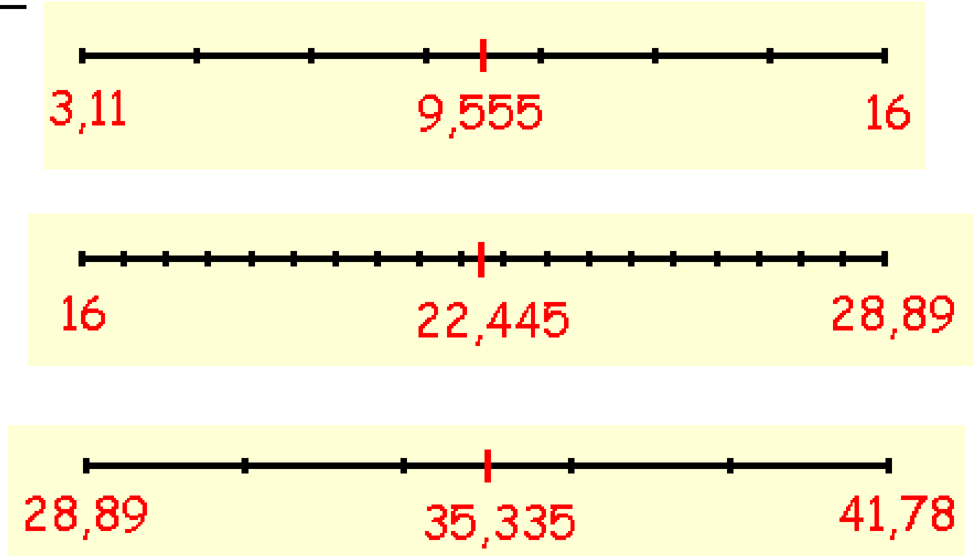
j	Classes	F_j
1	3,11 ├── 16,00	8
2	16,00 ├── 28,89	20
3	28,89 ├── 41,78	6
4	41,78 ├── 54,67	8
5	54,67 ├── 67,56	3
6	67,56 ├── 80,45	1
7	80,45 ├──┤ 93,34	4
	Σ	50

Dados agrupados em classe

Medidas para dados agrupados em classe

j	Classes	c_j	F_j
1	3,11 |— 16,00	9,555	8
2	16,00 |— 28,89	22,445	20
3	28,89 |— 41,78	35,335	6
4	41,78 |— 54,67	48,225	8
5	54,67 |— 67,56	61,115	3
6	67,56 |— 80,45	74,005	1
7	80,45 |—| 93,34	86,895	4
	Σ	–	50

Pressuposição



Distribuição simétrica dentro dos intervalos

$$\bar{x} = \frac{\sum x_i}{n}$$ ← não estão disponíveis

Média

j	Classes	c_j	F_j	$c_j\ F_j$
1	3,11 ├── 16,00	9,555	8	9,555 x 8 = 76,44
2	16,00 ├── 28,89	22,445	20	22,445 x 20 = 448,90
3	28,89 ├── 41,78	35,335	6	35,335 x 6 = 212,01
4	41,78 ├── 54,67	48,225	8	48,225 x 8 = 385,80
5	54,67 ├── 67,56	61,115	3	61,115 x 3 = 183,345
6	67,56 ├── 80,45	74,005	1	74,005 x 1 = 74,005
7	80,45 ├──┤ 93,34	86,895	4	86,895 x 4 = 347,58
	Σ	–	50	$\sum c_j F_j$ **= 1.728,08**

Média aritmética: $\bar{x} = \dfrac{\sum c_j F_j}{n} = \dfrac{1728,08}{50} = 34,56$ reais

gasto no supermercado

Mediana e Moda

Classe mediana → 2
Classe modal → 2

j	Classes	c_j	F_j	F'_j	$c_j\ F_j$
1	3,11 ├── 16,00	9,555	8	8	76,44
2	16,00 ├── 28,89	22,445	20	28	448,90
3	28,89 ├── 41,78	35,335	6	34	212,01
4	41,78 ├── 54,67	48,225	8	42	385,80
5	54,67 ├── 67,56	61,115	3	45	183,35
6	67,56 ├── 80,45	74,005	1	46	74,01
7	80,45 ├──┤ 93,34	86,895	4	50	347,58
	Σ	–	50	–	1.728,08

Posição da Mediana → $p = \dfrac{n+1}{2} = 25,5$ → 25, 26

Classe mediana: classe que compreende a mediana

Classe modal: classe com a maior frequência absoluta

Variância

$$\bar{x} = 34{,}56 \text{ reais}$$

j	Classes	c_j	F_j	$F_j(c_j - \bar{x})^2$
1	3,11 |— 16,00	9,555	8	$8\,(9{,}555 - 34{,}56)^2 = 5.002{,}00$
2	16,00 |— 28,89	22,445	20	$20\,(22{,}445 - 34{,}56)^2 = 2.935{,}46$
3	28,89 |— 41,78	35,335	6	$6\,(35{,}335 - 34{,}56)^2 = 3{,}60$
4	41,78 |— 54,67	48,225	8	
5	54,67 |— 67,56	61,115	3	
6	67,56 |— 80,45	74,005	1	
7	80,45 |—| 93,34	86,895	4	
	Σ	–	50	$\sum F_j(c_j - \bar{x})^2$

24.062,15

Variância: $$s^2 = \frac{\sum F_j(c_j - \bar{x})^2}{n-1}$$

$$s^2 = \frac{\sum F_j \left(c_j - \bar{x}\right)^2}{n-1} = \frac{24062{,}15}{49} = 491{,}06 \ \text{reais}^2$$

$$s = \sqrt{s^2} = \sqrt{491{,}06} = 22{,}16 \ \text{reais}$$

$$\bar{x} \pm s$$

$$34{,}56 \pm 22{,}16 \, \text{reais}$$

$$CV = \frac{s}{\bar{x}} 100\% = \frac{22{,}16}{34{,}56} 100\% = 64{,}12\%$$

Dados não agrupados	Dados agrupados em classe
$\overline{X}$ = 34,78 reais	$\overline{X}$ = 34,56 reais
Md = 27,31 reais	Classe mediana: [16,00 ; 28,89)
Mo não existe	Classe modal: [16,00 ; 28,89)
s^2 = 471,32 reais^2	s^2 = 491,06 reais^2
s = 21,71 reais	s = 22,16 reais
CV = 62,42%	CV = 64,12%

Exercício proposto:

Calcule as medidas descritivas para o conjunto de dados referente ao número de pães não vendidos em uma certa padaria até a hora do encerramento do expediente.

j	Classes	F_j
1	0	20
2	1	7
3	2	7
4	3	3
5	4	2
6	5	1
	$\sum$	40

Solução:

j	Classes	F_j	F_j'	c_jF_j	$F_j(c_j-\bar{x})^2$
1	0	20	20	0	23,11
2	1	7	27	7	0,04
3	2	7	34	14	5,99
4	3	3	37	9	11,12
5	4	2	39	8	17,11
6	5	1	40	5	15,41
	$\sum$	40	-	43	72,78

média = 1,075
moda = 0
mediana = 0,5

Assimétrica positiva

Variância = 1,87
Desvio padrão = 1,37
CV = 127,07