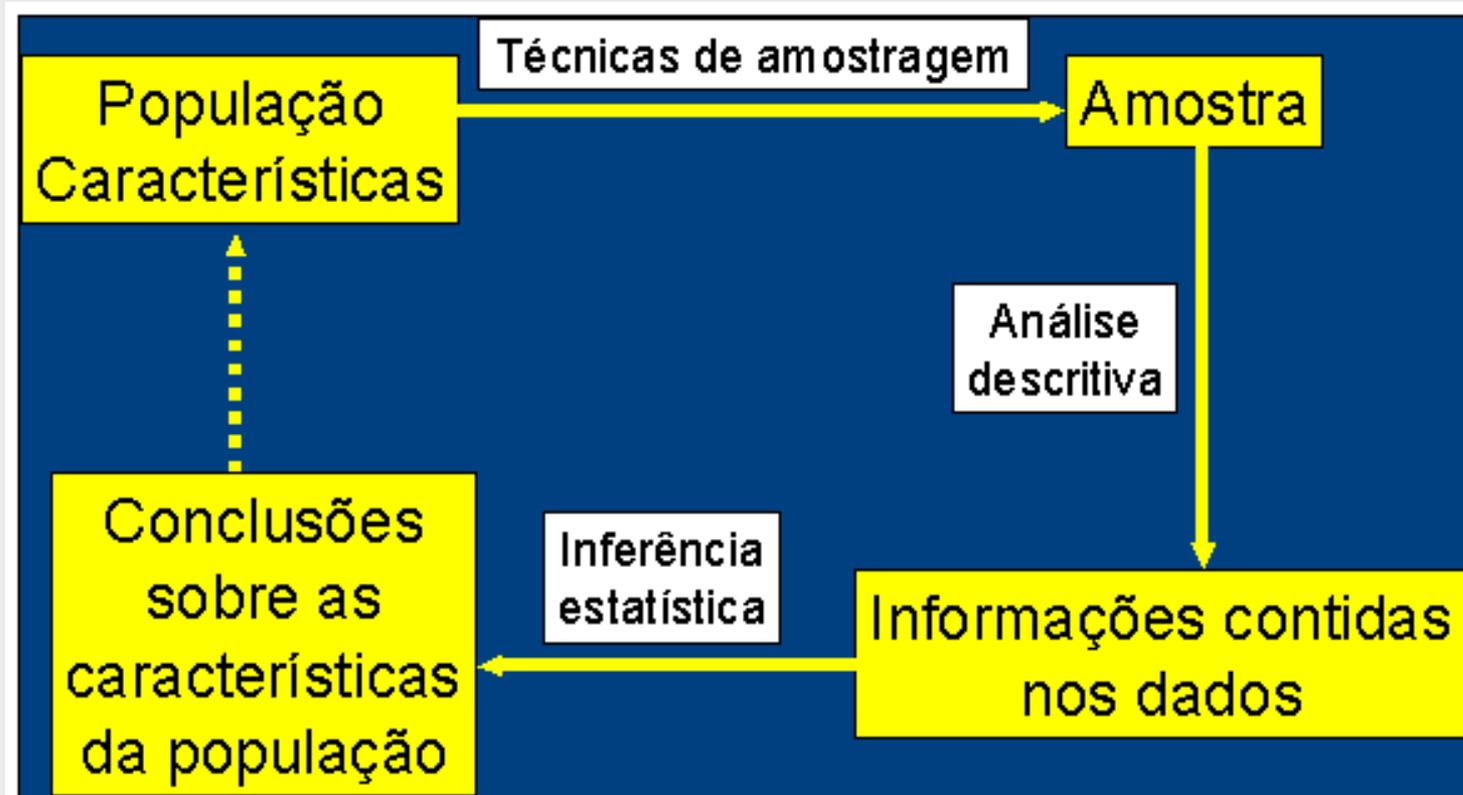


Medidas de Dispersão

Professora Ana Amélia Benedito Silva
aamelia@usp.br

Etapas da Análise Estatística



ANÁLISE DESCRITIVA

- tabelas
- gráficos
- medidas
 - média, mediana, moda
 - **desvio-padrão, coeficiente de variação**
 - percentis, quartis, decis

Medidas de Variabilidade

expressam a variabilidade ou a dispersão de um conjunto de dados

- amplitude total**
- variância, desvio-padrão**
- coeficiente de variação**

Variabilidade/Dispersão

temperatura máxima
em 3 dias na cidade A:

- dia 1: 28° C,
- dia 2: 29° C
- dia 3: 30° C

média = 29° C

temperatura máxima
em 3 dias na cidade B:

- dia 1: 23° C
- dia 2: 29° C
- dia 3: 35° C

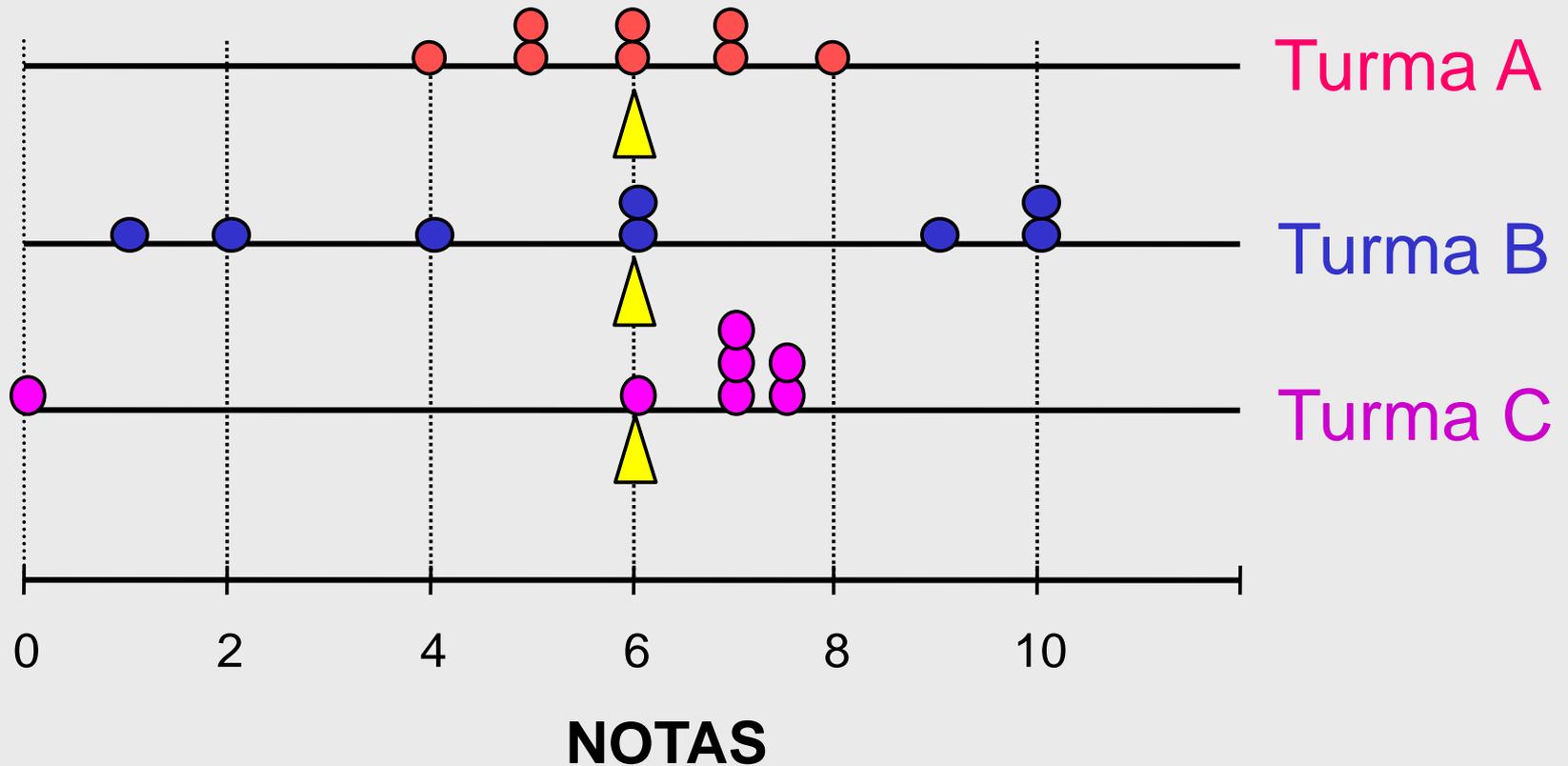
média = 29° C

Medida de variabilidade: permite distinguir o comportamento da temperatura máxima nas 2 cidades

Notas finais dos alunos de três turmas

turma	notas								\bar{x}
A	4	5	5	6	6	7	7	8	6
B	1	2	4	6	6	9	10	10	6
C	0	6	7	7	7	7,5	7,5		6

Diagrama de pontos das três turmas



 Média das notas

Amplitude total

$$A = \text{valor}_{\text{máximo}} - \text{valor}_{\text{mínimo}}$$

- determina a dispersão entre o maior e o menor valor de uma sequência **ordenada** de dados

Amplitude Total

turma	notas								\bar{x}
A	4	5	5	6	6	7	7	8	6
B	1	2	4	6	6	9	10	10	6
C	0	6	7	7	7	7,5	7,5		6

$$\text{Amplitude total (A)} = 8 - 4 = 4$$

$$\text{Amplitude total (B)} = 10 - 1 = 9$$

$$\text{Amplitude total (C)} = 7,5 - 0 = 7,5$$

Amplitude total

temperatura máxima
em 3 dias na cidade A:

- dia 1: 28° C,
- dia 2: 29° C
- dia 3: 30° C

média = 29° C

Amplitude total = 30-28=2° C

temperatura máxima
em 3 dias na cidade B:

- dia 1: 23° C
- dia 2: 29° C
- dia 3: 35° C

média = 29° C

Amplitude total = 35-23=12° C

Desvantagem da amplitude total: não considera todos os valores, apenas os valores extremos.

Desvio-padrão

- é a medida de dispersão **mais usada**
- determina a dispersão de **todos** os valores em relação à média
- quanto menor o desvio-padrão, menor a dispersão
- vem sempre junto com a média – **um não vive sem o outro!!!**

Qual das 2 cidades tem a menor dispersão em torno da média?

O desvio-padrão responde a esta questão.

temperatura máxima em 3 dias na cidade A:

- dia 1: 28° C,
- dia 2: 29° C
- dia 3: 30° C

média = 29° C

temperatura máxima em 3 dias na cidade B:

- dia 1: 23° C
- dia 2: 29° C
- dia 3: 35° C

média = 29° C

Desvio padrão

FUVEST

Prova matemática

nota aluno: 7

média: 5

desvio-padrão = 1

FUVEST

Prova física

nota aluno: 7

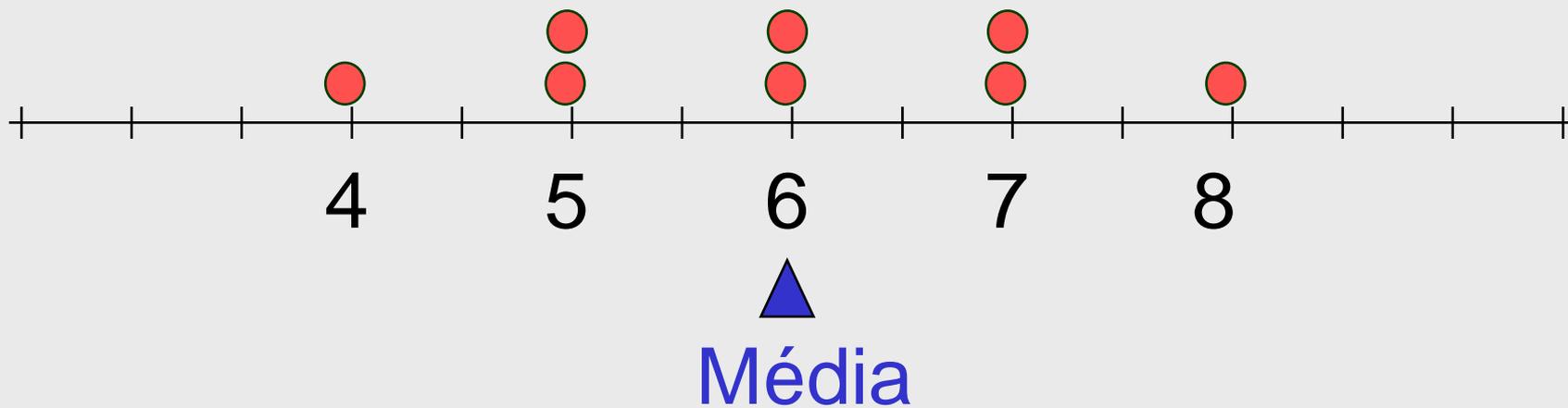
Média: 5

desvio-padrão = 2

Em qual das 2 provas ele foi melhor?

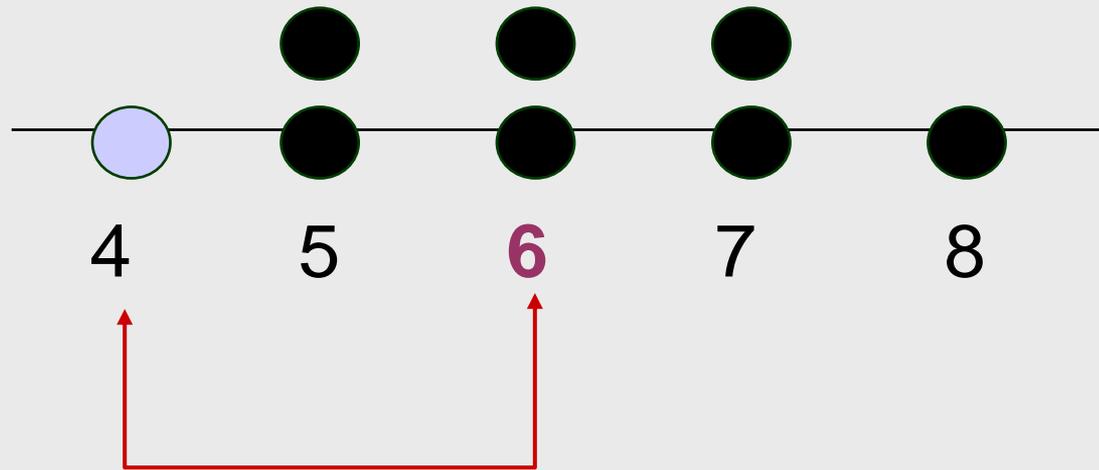
**Voltando ao
desvio-padrão**

A: 4 5 5 6 6 7 7 8



Como medir a dispersão?

Exemplo: Turma A : 4 5 5 6 6 7 7 8



Distância (desvio) de um valor em relação à média

Turma A : 4 5 5 6 6 7 7 8
Média = 6

**A soma dos desvios em
relação à média dá **sempre**
zero!!!!**

desvios em relação à média

$$4-6 = -2$$

$$5-6 = -1$$

$$5-6 = -1$$

$$6-6 = 0$$

$$6-6 = 0$$

$$7-6 = +1$$

$$7-6 = +1$$

$$8-6 = +2$$

$$\mathbf{Soma = 0}$$

Já que a soma dos desvios em relação à média dá sempre zero, o melhor é calcular os desvios ao quadrado!!!!

desvios ao quadrado em relação à média:

$$(4-6)^2 = (-2)^2 = +4$$

$$(5-6)^2 = (-1)^2 = +1$$

$$(5-6)^2 = (-1)^2 = +1$$

$$(6-6)^2 = (0)^2 = 0$$

$$(6-6)^2 = (0)^2 = 0$$

$$(7-6)^2 = (+1)^2 = +1$$

$$(7-6)^2 = (+1)^2 = +1$$

$$(8-6)^2 = (+2)^2 = +4$$

soma dos desvios ao quadrado = 12

Turma A : 4 5 5 6 6 7 7 8 Média = 6

Variância e Desvio-Padrão

Se $n \leq 30$ (amostra):

variância amostral = (soma dos desvios ao quadrado) / (n-1)

desvio-padrão amostral = raiz quadrada da variância amostral

Se $n > 30$ (população):

variância populacional = (soma dos desvios ao quadrado) / (n)

desvio-padrão populacional = raiz quadrada da variância populacional

n= número de elementos

Já que a soma dos desvios em relação à média dá sempre zero, o melhor é calcular os desvios ao quadrado!!!!

desvios ao quadrado em relação à média:

$$\begin{aligned}(4-6)^2 &= (-2)^2 = +4 \\(5-6)^2 &= (-1)^2 = +1 \\(5-6)^2 &= (-1)^2 = +1 \\(6-6)^2 &= (0)^2 = 0 \\(6-6)^2 &= (0)^2 = 0 \\(7-6)^2 &= (+1)^2 = +1 \\(7-6)^2 &= (+1)^2 = +1 \\(8-6)^2 &= (+2)^2 = +4\end{aligned}$$

soma dos desvios ao quadrado = 12

$$\text{variância} = 12 / (8-1) = 1,71$$

$$\text{desvio-padrão} = \sqrt{1,71} = 1,31$$

Comparação das três turmas pela média e desvio padrão

turma	notas								\bar{X}	S
A	4	5	5	6	6	7	7	8	6	1,31
B	1	2	4	6	6	9	10	10	6	3,51
C	0	6	7	7	7	7,5	7,5		6	2,69

Desvio-padrão

temperatura máxima
em 3 dias na cidade A:

- dia 1: 28° C,
- dia 2: 29° C
- dia 3: 30° C

média = 29° C

temperatura máxima
em 3 dias na cidade B:

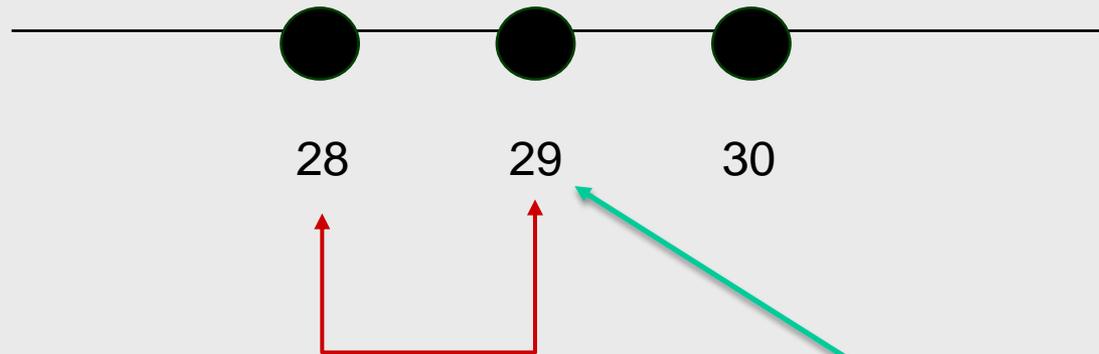
- dia 1: 23° C
- dia 2: 29° C
- dia 3: 35° C

média = 29° C

Desvio-padrão: permite distinguir o comportamento da temperatura máxima nas 2 cidades

Como medir a dispersão?

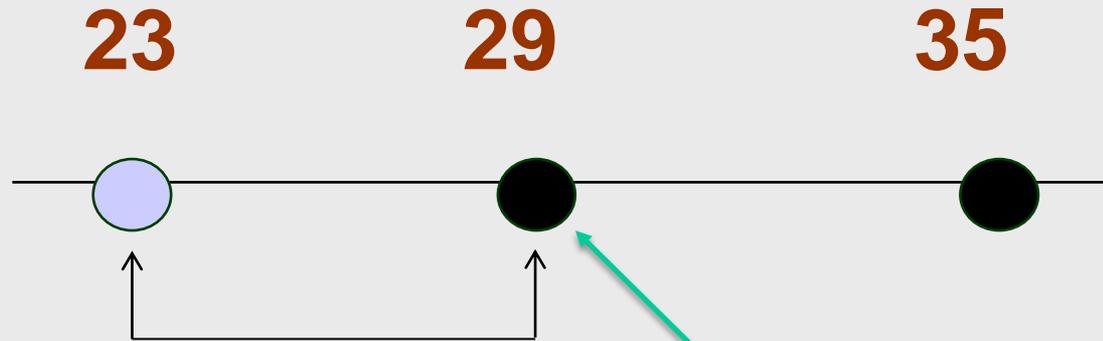
Cidade A : 28 29 30



Média = 29 ° C

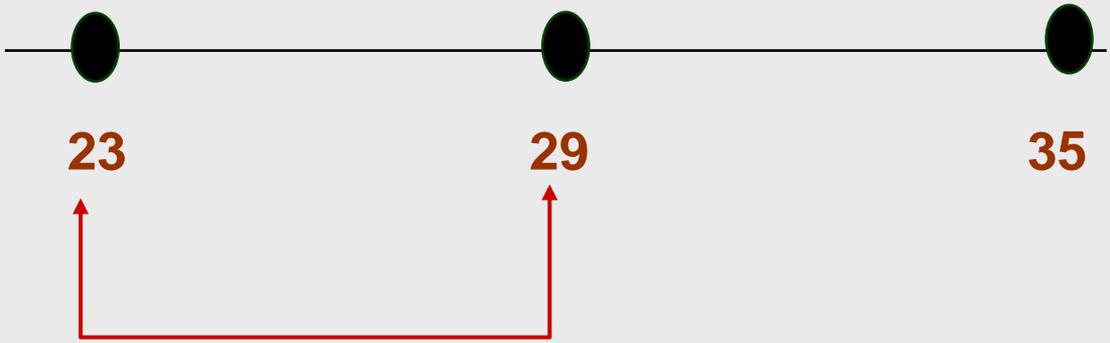
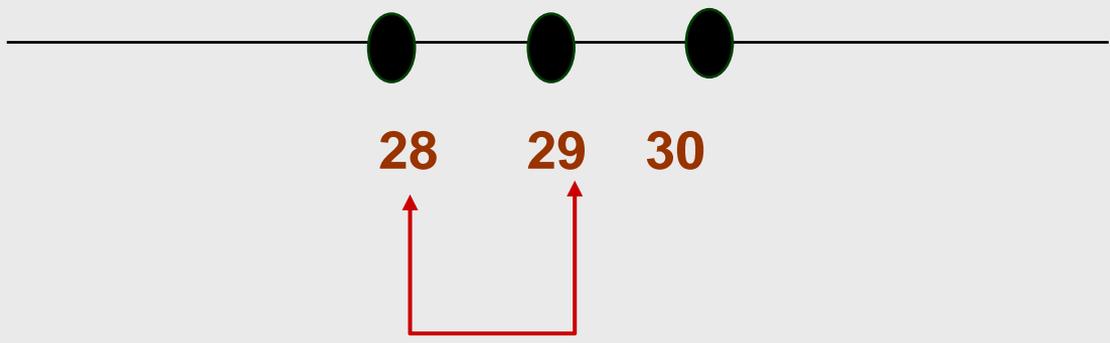
Como medir a dispersão?

Cidade B



Média = 29 ° C

Cidade A: 28° C 29° C 30° C



Cidade B: 23° C 29° C 35° C

Desvio-padrão

temperatura máxima na cidade A:

- dia 1: 28° C,
- dia 2: 29° C
- dia 3: 30° C

média = 29° C
desvio-padrão = 1° C

temperatura máxima na cidade B:

- dia 1: 23° C
- dia 2: 29° C
- dia 3: 35° C

média = 29° C
desvio-padrão = 6° C

Cálculo do desvio-padrão

	Cidade A	Cidade A	Cidade B	Cidade B
	28	$(28-29)^2 = 1$	23	$(23-29)^2 = 36$
	29	$(29-29)^2 = 0$	29	$(29-29)^2 = 0$
	30	$(30-29)^2 = 1$	35	$(35-29)^2 = 36$
	média = 29		média = 29	
soma		$1+0+1 = 2$		$36+0+36=72$
variância		$2/(3-1)=1$		$72/(3-1)=36$
desvio-padrão		raiz (1)=1		raiz (36) = 6

Coeficiente de Variação (CV)

- É uma medida de dispersão relativa
- Exprime o desvio-padrão em relação à média

$$CV = \frac{\text{Desvio-padrão}}{\text{média}} \times 100\%$$

altura e peso de um grupo de crianças

	MÉDIA	DESVIO-PADRÃO	COEFICIENTE DE VARIAÇÃO
ALTURA	100 cm	6 cm	6 %
PESO	40 Kg	6 Kg	15%

altura de crianças e de adolescentes num mesmo grupo

	MÉDIA	DESVIO-PADRÃO	COEFICIENTE DE VARIAÇÃO
Recém-nascidos	50 cm	5 cm	10 %
Adolescentes	160 cm	16 cm	10%

Exercício

O Procedimento de perda rápida de "peso" é comum entre os atletas dos esportes de combate. Para participar de um torneio, quatro atletas da categoria até 66 kg, Peso-Pena, foram submetidos a dietas balanceadas e atividades físicas. Realizaram três "pesagens" antes do início do torneio. Pelo regulamento do torneio, a primeira luta deverá ocorrer entre o atleta mais regular e o menos regular quanto aos "pesos". As informações com base nas pesagens dos atletas estão no quadro.

Atleta	1ª pesagem (kg)	2ª pesagem (kg)	3ª pesagem (kg)	Média	Mediana	Desvio padrão
I	78	72	66	72	72	4,90
II	83	65	65	71	65	8,49
III	75	70	65	70	70	4,08
IV	80	77	62	73	77	7,87

Após as três "pesagens", os organizadores do torneio informaram aos atletas quais deles se enfrentariam na primeira luta.

A primeira luta foi entre os atletas

- a) I e III.
- b) I e IV.
- c) II e III.
- d) II e IV.
- e) III e IV

Exercício

Um produtor de café irrigado em Minas Gerais recebeu um relatório de consultoria estatística, constando, entre outras informações, o desvio padrão das produções de uma safra dos talhões de sua propriedade. Os talhões têm a mesma área de $30\,000\text{ m}^2$ e o valor obtido para o desvio padrão foi de 90 kg/talhão . O produtor deve apresentar as informações sobre a produção e a variância dessas produções em sacas de 60 kg por hectare ($10\,000\text{ m}^2$). A variância das produções dos talhões expressa em $(\text{sacas/hectare})^2$ é:

- a) 20,25
- b) 4,50
- c) 0,71
- d) 0,50
- e) 0,25.

Exercício

Marco e Paulo foram classificados em um concurso. Para classificação no concurso o candidato deveria obter média aritmética na pontuação igual ou superior a 14. Em caso de empate na média, o desempate seria em favor da pontuação mais regular. No quadro a seguir são apresentados os pontos obtidos nas provas de Matemática, Português e Conhecimentos Gerais, a média, a mediana e o desvio padrão dos dois candidatos.

Dados dos candidatos no concurso

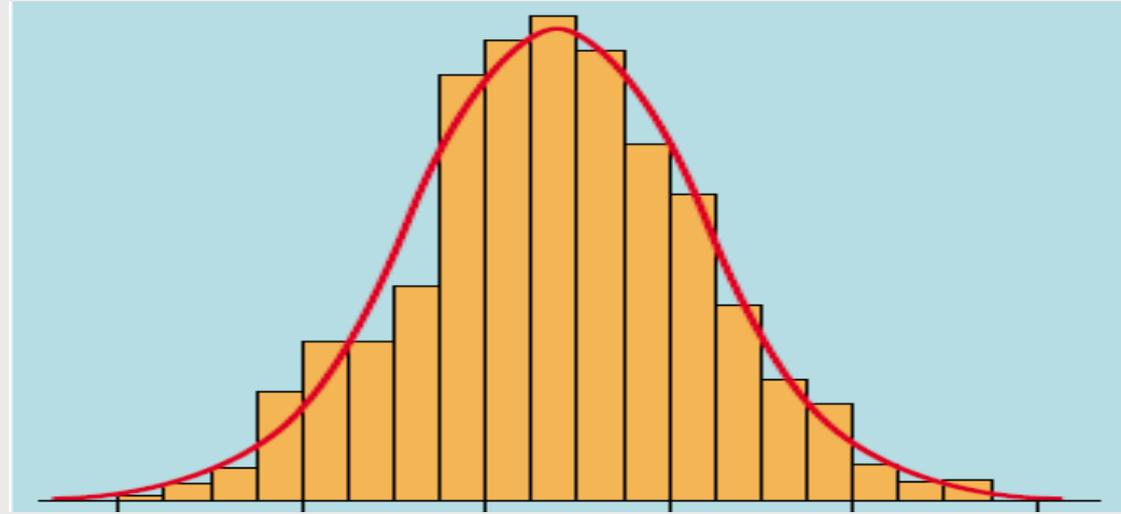
	Matemática	Português	Conhecimentos Gerais	Média	Mediana	Desvio Padrão
Marco	14	15	16	15	15	0,32
Paulo	8	19	18	15	18	4,97

O candidato com pontuação mais regular, portanto mais bem classificado no concurso, é

- a) Marco, pois a média e a mediana são iguais.
- b) Marco, pois obteve menor desvio padrão.
- c) Paulo, pois obteve a maior pontuação da tabela, 19 em Português.
- d) Paulo, pois obteve maior mediana.
- e) Paulo, pois obteve maior desvio padrão.

A DISTRIBUIÇÃO NORMAL

curva em forma de sino comum em muitos tipos de observações biológicas, psicológicas e sociais



$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad -\infty < x < \infty.$$

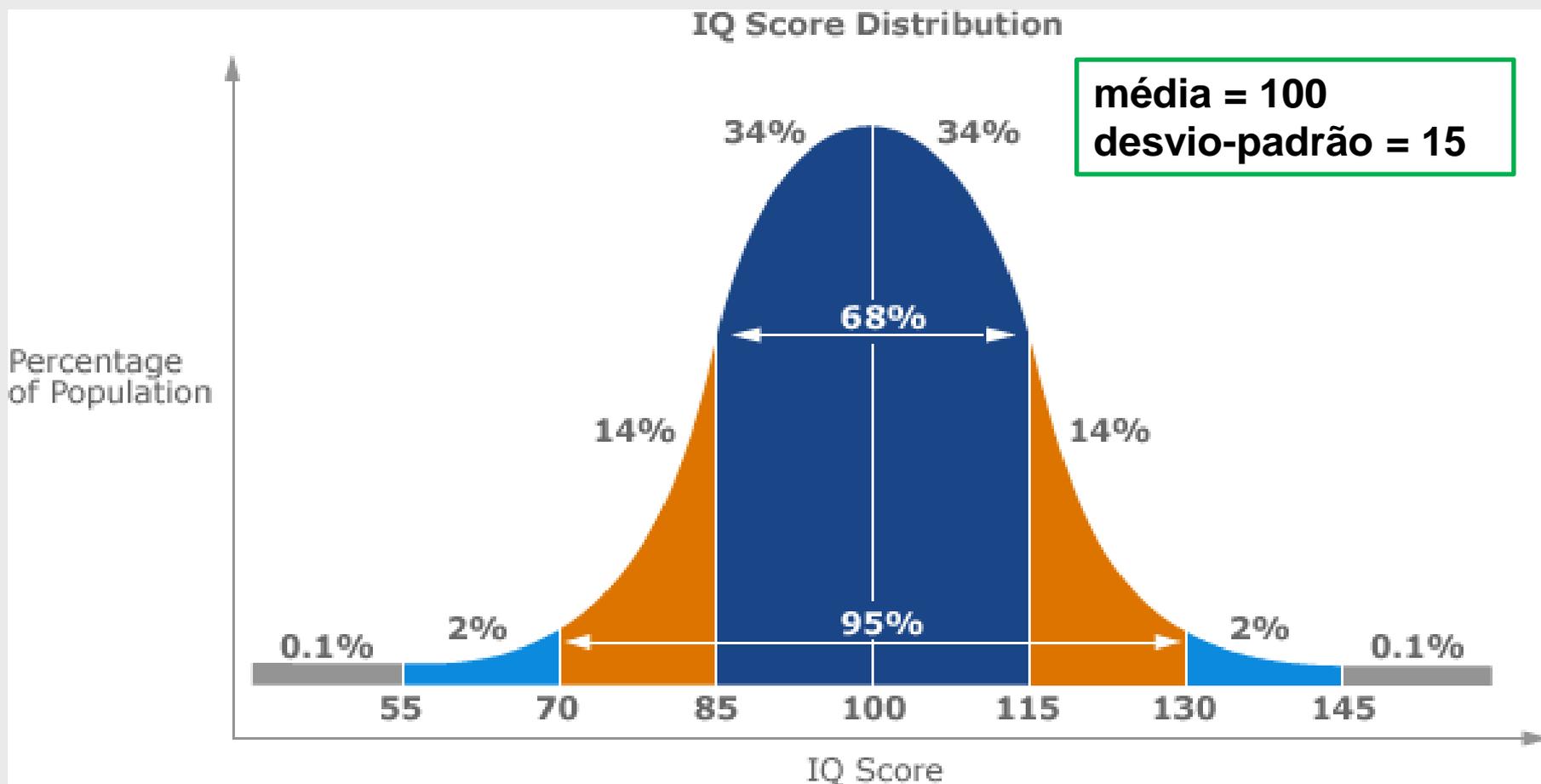
μ é o valor esperado (média) de X , com $-\infty < \mu < \infty$;

σ^2 é a variância de X , com $\sigma^2 > 0$.

Faixa de normalidade

- **média aritmética \pm desvio-padrão**
- **corresponde à aproximadamente 68% dos indivíduos da amostra**

Distribuição do QI numa população



	Nº de sujeitos	Idade (anos)	Peso (Kg)
homens	100	35±2	75±5
mulheres	100	28±4	60±4

Faixa de normalidade = média aritmética ± desvio-padrão

FN idade dos homens = 35 ± 2 anos : 33 a 37 anos

Faixa de normalidade (FN)

	Nº de sujeitos	Idade (anos)	Peso (Kg)
homens	100	35±2	75 ± 8
mulheres	100	28±4	60±4

Faixa de normalidade = média aritmética ± desvio-padrão

FN idade dos homens = 35 ± 2 anos (33 a 37)

FN peso dos homens = 75 ± 8 Kg (67 a 83)

NÃO inclui todos os sujeitos (só 68%)



100
sujeitos

média
35 anos

desvio-
padrão
2 anos

68% (**verde**) idade entre 33 e 37 anos

95% (**verde+azul**) idade entre 31 e 39 anos

99% (**verde+azul+laranja**) idade entre 29 e 41 anos

Exercício

Considere o conjunto de observações: 1 4 6 2 2

- a) Calcule a média e o desvio padrão dos dados.
- b) Some duas unidades a cada observação e repita o item (a). Compare com a resposta em (a) e comente.
- c) Multiplique cada observação original por 3 e repita (a).
- d) Compare com a resposta em (a) e comente,
- e) Se multiplicarmos cada observação original por uma constante b e somarmos a constante a , qual seria a média e o desvio padrão dos dados?

**Peso
recém-
nascidos**

Sobreviveram		Morreram	
1130	2550	1030	1600
1410	2570	1050	1720
1575	2600	1100	1750
1680	2700	1175	1770
1715	2830	1185	1820
1720	2950	1225	1890
1760	3005	1230	1940
1930	3160	1262	2200
2015	3400	1295	2270
2040	3640	1300	2275
2090		1310	2440
2200		1500	2500
2400		1550	2560
			2730

	Sobreviveram	Morreram
media	2307.4	1691.7
desvio-padrão	664.7	517.6
Coeficiente variação	28.8%	30.6%

OBRIGADA!!