

TTI109 – Estatística

Aula 06 – Estatística Descritiva 05

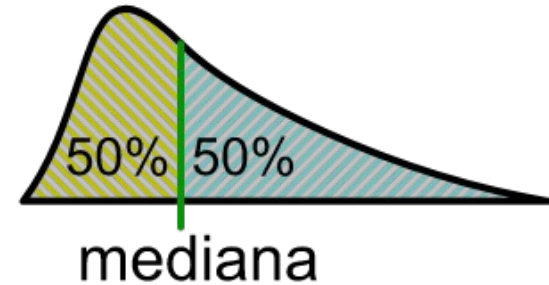


Medidas de posição

↯ Vamos investigar as *separatrizes* para especificar a posição de um elemento dentro de um conjunto de dados.

↯ *Separatrizes* são números que partilham, ou dividem, um conjunto de dados ordenado em partes iguais (com o mesmo número de elementos).

A mediana é uma separatriz porque divide um conjunto de dados ordenado em duas partes com quantidades iguais de elementos.



Medidas de posição

Quartis

Definição

Os três **quartis**, Q_1 , Q_2 e Q_3 , dividem um conjunto de dados ordenado em quatro partes iguais. Aproximadamente 1/4 dos dados recai sobre ou abaixo do **primeiro quartil** Q_1 . Aproximadamente metade dos dados recai sobre ou abaixo do **segundo quartil** Q_2 (o segundo quartil é o mesmo que a mediana do conjunto de dados). Aproximadamente 3/4 dos dados recaem sobre ou abaixo do **terceiro quartil** Q_3 .

Encontrando os quartis de um conjunto de dados

O número de usinas nucleares nos 15 maiores produtores de energia nuclear no mundo está listado a seguir. Encontre o primeiro, o segundo e o terceiro quartis do conjunto de dados. O que você observa? (*Fonte: International Atomic Energy Agency.*)

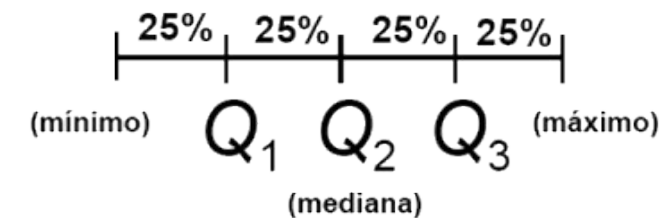
7 20 16 6 58 9 20 50 23 33 8 10 15 16 104



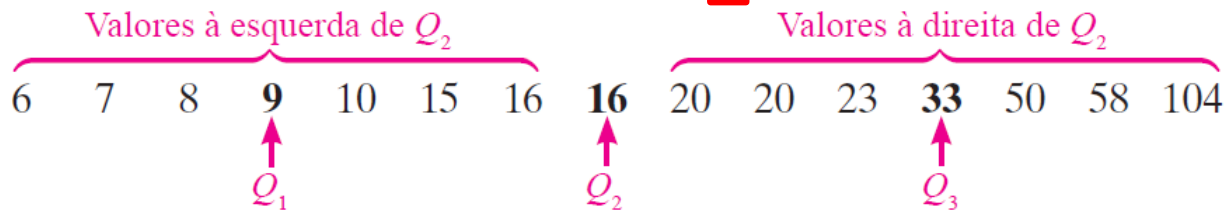
Como fazer?

Medidas de posição

- 1 Ordene o conjunto de dados e encontre a mediana
- 2 O primeiro quartil, Q_1 , é a mediana dos valores à esquerda de
- 3 O terceiro quartil, Q_3 , é a mediana dos valores à direita de



Aproximadamente 1/4 dos países tem 9 usinas nucleares ou menos; aproximadamente metade tem 16 ou menos; e cerca de 3/4 têm 33 ou menos.



Medidas de posição

↯ A **mediana** (o segundo quartil) é uma medida de **tendência central** baseada na **posição**. A medida de **variação** que é baseada na **posição** é a **amplitude interquartil**, que indica a dispersão dos 50% centrais.

Definição

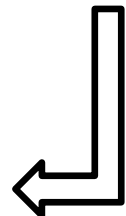
A **amplitude interquartil (AIQ)** de um conjunto de dados é uma medida de variação que fornece a amplitude da porção central (aproximadamente metade) dos dados. A AIQ é a diferença entre o terceiro e o primeiro quartis.

$$AIQ = Q_3 - Q_1$$



Lembre-se

Um *outlier* está muito distante dos demais valores do conjunto de dados.
Uma forma de identificar *outliers* é usar a amplitude interquartil.



Medidas de posição

Instruções

Usando a amplitude interquartil para identificar *outliers*

1. Encontre o primeiro (Q_1) e o terceiro (Q_3) quartis do conjunto de dados.
2. Encontre a amplitude interquartil: $AIQ = Q_3 - Q_1$.
3. Multiplique a AIQ por 1,5: 1,5 (AIQ).
4. Subtraia 1,5 (AIQ) de Q_1 . Qualquer valor menor que $Q_1 - 1,5$ (AIQ) é um *outlier*.
5. Adicione 1,5 (AIQ) à Q_3 . Qualquer valor maior que $Q_3 + 1,5$ (AIQ) é um *outlier*.

O número de usinas nucleares nos 15 maiores produtores de energia nuclear no mundo está listado a seguir. Encontre o primeiro, o segundo e o terceiro quartis do conjunto de dados. O que você observa? (*Fonte: International Atomic Energy Agency.*)

7 20 16 6 58 9 20 50 23 33 8 10 15 16 104



Existem
outliers nesse
conjunto de
dados?

Medidas de posição

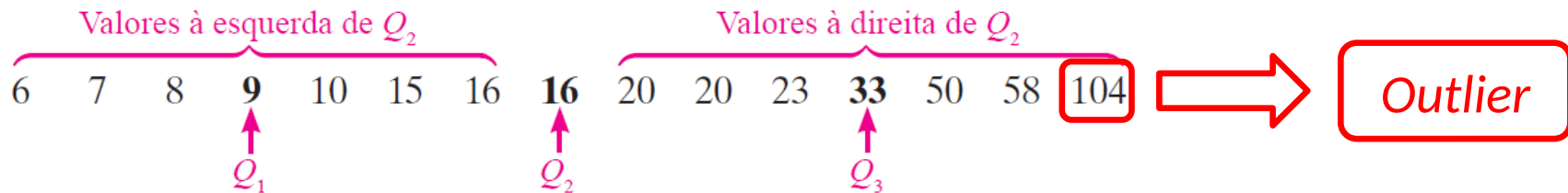


$$Q_1 = 9 \text{ e } Q_3 = 33$$

↗ Amplitude interquartil:

↗ Para se identificar outlier:

(não se aplica!)



↗ **Conclusão:** O número de usinas nucleares na porção central do conjunto de dados varia por, no máximo, 24. O outlier, 104, não afeta a .

Medidas de posição

↗ **Aplicação dos quartis:** representar conjuntos de dados usando diagramas de caixa-e-bigode ou **boxplots**.

↗ Um diagrama **boxplot** é uma ferramenta de análise exploratória que destaca características importantes de um conjunto de dados.

1. O valor mínimo
2. O primeiro quartil Q_1
3. A mediana Q_2
4. O terceiro quartil Q_3
5. O valor máximo

Instruções

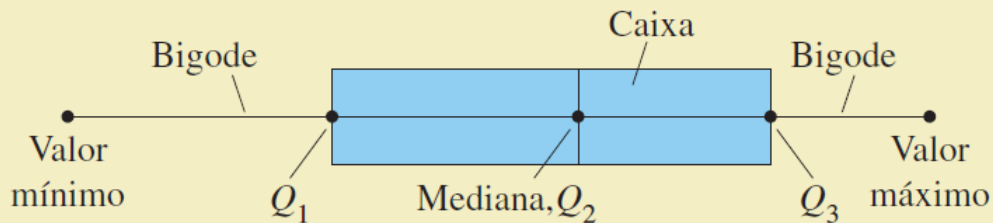
Desenhando um **boxplot**

1. Determine o resumo dos cinco números do conjunto de dados.
2. Construa uma escala horizontal que se estenda sobre a amplitude dos dados.
3. Represente os cinco números sobre a escala horizontal.
4. Desenhe uma caixa em cima da escala horizontal de Q_1 a Q_3 , e desenhe uma linha vertical na caixa, em Q_2 .
5. Desenhe os bigodes da caixa para os valores mínimo e máximo.

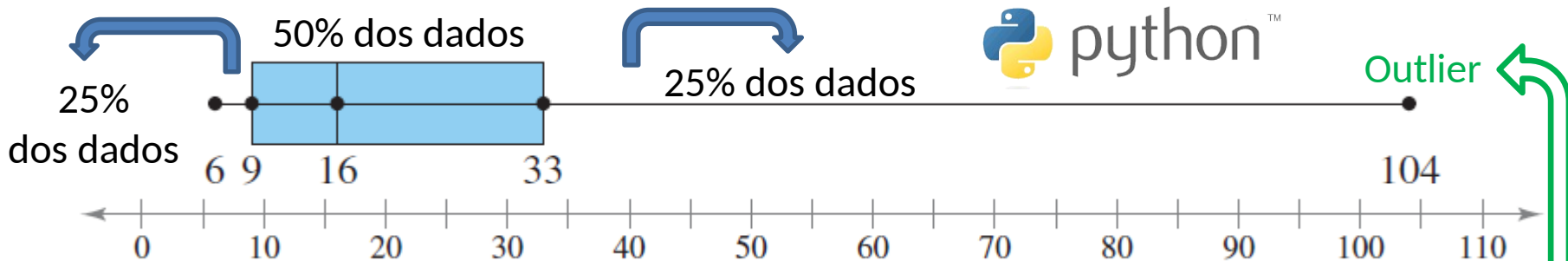


Medidas de posição

Boxplot com seus cinco elementos principais com a caixa e linhas (bigodes).



Exemplo das usinas nucleares:



O comprimento do bigode direito é muito maior do que o esquerdo...

Medidas de posição

Percentis e outras separatrizes

Separatriz	Resumo	Símbolos
Quartis	Divide um conjunto de dados em 4 partes iguais	Q_1, Q_2, Q_3
Decis	Divide um conjunto de dados em 10 partes iguais	$D_1, D_2, D_3, \dots, D_9$
Percentis	Divide um conjunto de dados em 100 partes iguais	$P_1, P_2, P_3, \dots, P_{99}$

Definição

Para encontrar o **percentil que corresponde a um valor específico x** , use a fórmula:

$$\text{Percentil de } x = \frac{\text{número de elementos menores que } x}{\text{número total de elementos}} \cdot 100$$

e então arredonde o resultado para o valor inteiro mais próximo.

Medidas de posição

Conjunto de dados
ordenado:

Custos com ensino
superior (em
milhares de dólares)

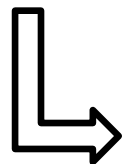
19	22	23	23	26	26	27	27	27	30	31	32	32
33	34	34	38	38	40	41	42	44	45	45	46	

Qual o percentil correspondente a US\$
30.000?

9 valores são menores do que US\$ 30.000

25 amostras no total

$$\text{Percentil de 30} = \frac{\text{número de elementos menores que 30}}{\text{número total de elementos}} = \frac{9}{25} \cdot 100 = 36$$



Interpretação: O custo com ensino em 36% das faculdade pesquisadas é inferior ou igual a US\$ 30.000.


Medidas de posição

Escore padrão

Definição

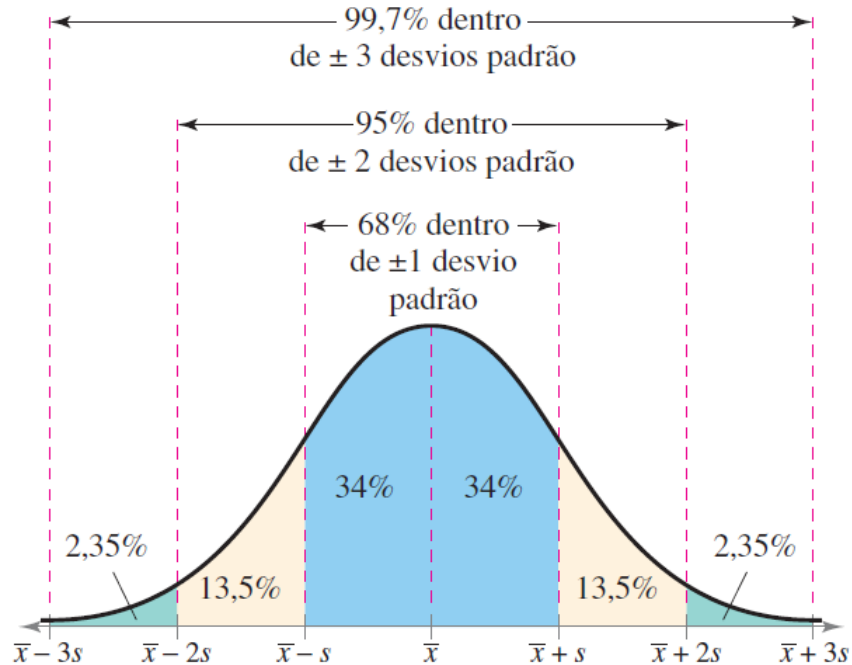
O **escore padrão** ou **escore-z** representa o número de desvios padrão em que um valor x encontra-se a partir da média μ . Para calcular o escore-z para um valor, use a fórmula a seguir:


$$z = \frac{\text{valor} - \text{média}}{\text{desvio padrão}} = \frac{x - \mu}{\sigma}.$$

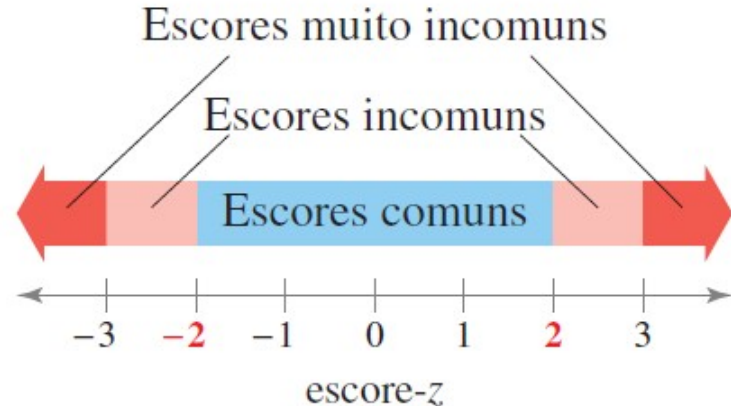


Um escore-z pode ser usado para identificar valores incomuns de um conjunto de dados que seja aproximadamente em formato de sino

Medidas de posição




$$z = \frac{x - \mu}{\sigma}$$



Medidas de posição

Médias e desvios padrão de populações de homens e mulheres



Altura dos homens	Altura das mulheres

Compare os escores-z de um homem e uma mulher com 1,83 m de altura. Admita que as distribuições das alturas sejam aproximadamente em formato de sino (chamadas de normais ou gaussianas).

$$z_H = \frac{x_H - \mu_H}{\sigma_H} = \frac{183 - 177,6}{7,6} = 0,71 \quad z_M = \frac{x_M - \mu_M}{\sigma_M} = \frac{183 - 163,3}{6,6} = 2,98$$

Medidas de posição

$$z_H = 0,71$$



O escore-z para o homem com 1,83 m de altura está dentro de *1 desvio padrão da média*, designando uma *altura típica* para um homem dessa população.

$$z_M = 2,98$$



O escore-z para a mulher com 1,83 m de altura está a quase *3 desvios padrão da média*, designando uma *altura incomum* para uma mulher dessa população.

TTI109 – Estatística

Aula 06 – Estatística Descritiva 05

