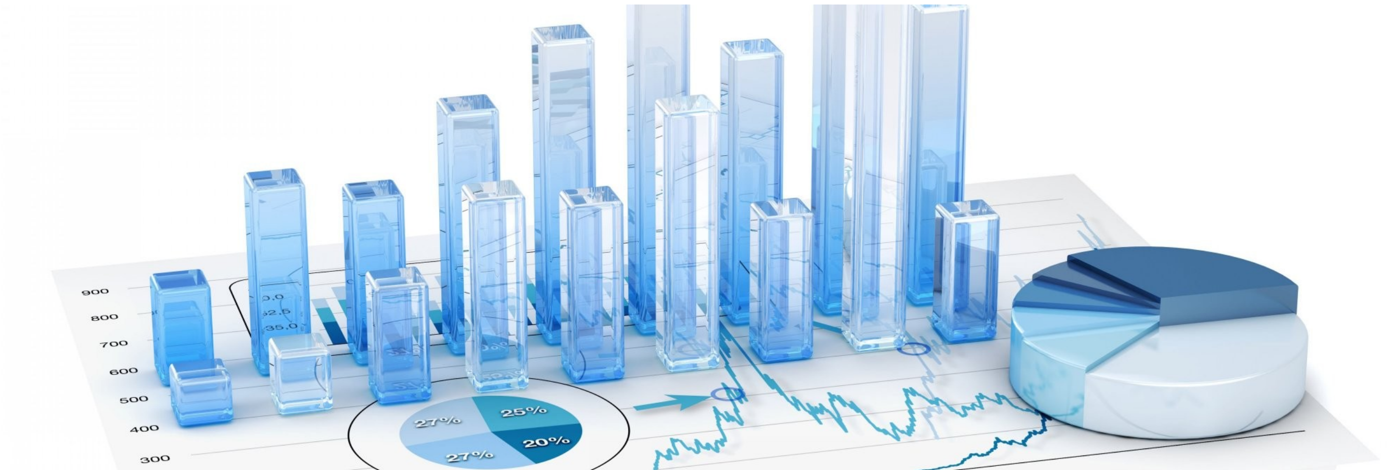
TTI109 – Estatística

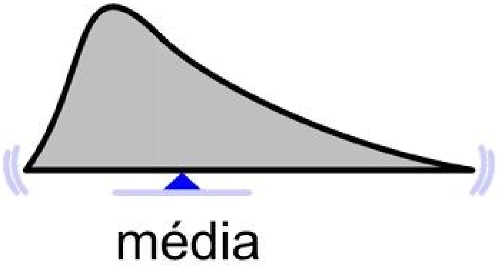
Aula 04 – Estatística Descritiva 03



Medidas de tendência central

- Uma *medida de tendência central* é um valor que representa uma *observação típica* ou central de um conjunto de dados.

Média



Definição

A **média** de um conjunto de dados é a soma dos valores dos dados dividida pelo número de observações. Para determinar a média de um conjunto de dados, use uma das fórmulas a seguir.

$$\text{Média populacional: } \mu = \frac{\Sigma x}{N} \qquad \text{Média amostral: } \bar{x} = \frac{\Sigma x}{n}$$

A letra grega minúscula μ (pronuncia-se mi) representa a média populacional e $\bar{x}$ (lê-se "x" barra) representa a média amostral. Note que N representa o número de observações em uma *população* e n representa o número de observações em uma *amostra*.

Medidas de tendência central

Encontrando a média amostral

Os pesos (em libras) de uma amostra de adultos antes de iniciarem um estudo sobre perda de peso estão listados. Qual é o peso médio dos adultos?

274	235	223	268	290	285	235

Solução

A soma dos pesos é:

$$\Sigma x = 274 + 235 + 223 + 268 + 290 + 285 + 235 = 1.810$$

Há 7 adultos na amostra, logo $n = 7$. Para encontrar o peso médio, divida a soma dos pesos pelo número de adultos na amostra.

$$\bar{x} = \frac{\Sigma x}{n} = \frac{1.810}{7} \approx 258{,}6.$$

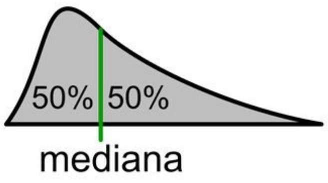
Resultado arredondado para uma casa decimal.

__Regra (típica) para arredondamento:__

Empregue uma casa decimal a mais do que o apresentado no conjunto original de dados.

Medidas de tendência central

Mediana



Definição

A **mediana** de um conjunto de dados é um valor que está no meio dos dados quando o conjunto está ordenado. A mediana indica o centro de um conjunto de dados ordenado, dividindo-o em duas partes com quantidades iguais de valores. Quando o conjunto de dados tem um número ímpar de observações, a mediana é o elemento do meio. Se o conjunto de dados tem um número par de observações, a mediana é a média dos dois elementos que ocupam as posições centrais.

Encontrando a mediana

Os pesos (em libras) de uma amostra de adultos antes de iniciarem um estudo sobre perda de peso estão listados.

274 235 223 268 290 285 235

Qual a mediana da amostra?

Medidas de tendência central

Solução

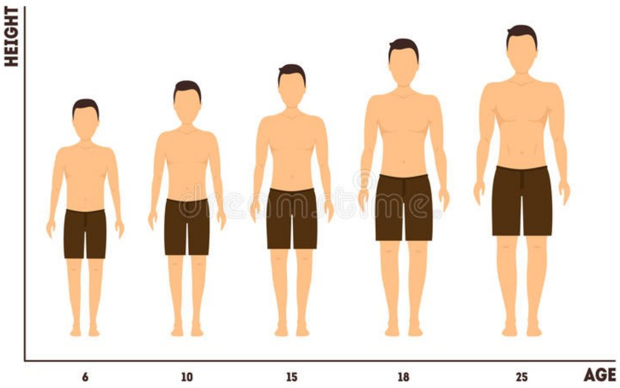
Para encontrar o peso mediano, primeiro ordene os dados.

223	235	235	**268**	274	285	290

Em razão de termos sete observações (um número ímpar), a mediana está no meio, é a quarta observação. Então, o peso mediano é 268 libras.

Dados na sequência original

274	235	223	268	290	285	235



Admita que o adulto pesando 285 libras decide não participar do estudo. Qual é o peso mediano dos adultos restantes?

Medidas de tendência central

Solução

Os pesos restantes em ordem são:

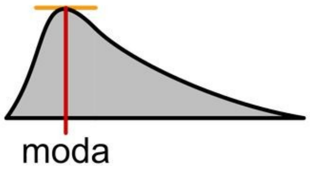
223 235 $\boxed{235 \quad 268}$ 274 290

Em razão de termos seis observações (um número par), a mediana é a média dos dois elementos do meio.

$$\text{Mediana} = \frac{235 + 268}{2} = \boxed{251{,}5}$$

Então, o peso mediano dos adultos restantes é 251,5 libras.

Moda



Definição

A **moda** de um conjunto de dados é o valor que ocorre com a maior frequência. Um conjunto de dados pode ter uma moda, mais de uma moda, ou não ter moda. Quando nenhum valor se repete, o conjunto de dados não tem moda. Quando dois valores ocorrem com a mesma maior frequência, cada um é uma moda e o conjunto é chamado de **bimodal**.

Medidas de tendência central

Os pesos (em libras) de uma amostra de adultos antes de iniciarem um estudo sobre perda de peso estão listados. Qual é o peso médio dos adultos?

274	235	223	268	290	285	235

Qual a moda da amostra?

Solução

Para encontrar a moda, primeiro ordene os dados.

223	235	235	268	274	285	290

A partir dos dados ordenados, podemos ver que o valor 235 ocorre duas vezes, enquanto os demais ocorrem somente uma vez. Então, a moda dos pesos é 235 libras.

Medidas de tendência central

Preferência por partido político em uma amostra de membros de um debate.

Partido político	Frequência, f
Democrata	46
Republicano	34
Independente	39
Outro/não sabe	5

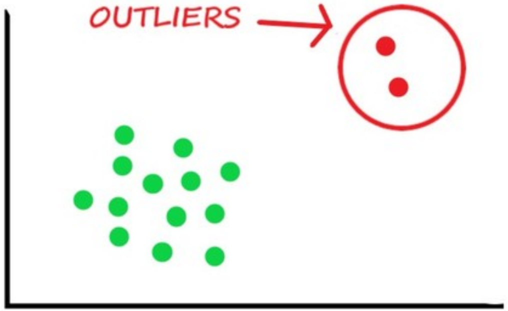
Qual a moda da amostra?

A média pode ser muito afetada quando o conjunto de dados contém *valores discrepantes* (*outliers*)

Definição

Um ***outlier*** é um valor que está muito afastado dos demais valores do conjunto de dados.

Medidas de tendência central



- Enquanto alguns *outliers* são dados válidos, outros podem ocorrer por causa de erros no registro dos dados.
- Um conjunto de dados pode ter um ou mais *outliers*, causando lacunas em uma distribuição.
- As conclusões que são tomadas com base em um conjunto de dados que contém *outliers* podem ser falhas.

INSTITUTO MAUÁ DE TECNOLOGIA

MAUÁ

Medidas de tendência central

Exemplo

Comparando a média, a mediana e a moda

Encontre a média, a mediana e a moda da amostra das idades dos alunos de uma turma mostradas na Tabela 2.15. Qual medida de tendência central melhor descreve um valor típico (representante) desse conjunto de dados? Há *outliers*?

A média considera todos os valores e é afetada pelo *outlier*. A mediana também leva em consideração todos os valores, e não é afetada pelo *outlier*.

Idades em uma turma.

20	20	20	20	20	20	21
21	21	21	22	22	22	23
23	23	23	24	24	65	

Outlier

Solução

Média: $\bar{x} = \frac{\Sigma x}{n} = \frac{475}{20} \approx 23{,}8$ anos.

Mediana: mediana $= \frac{21+22}{2} = 21{,}5$ anos.

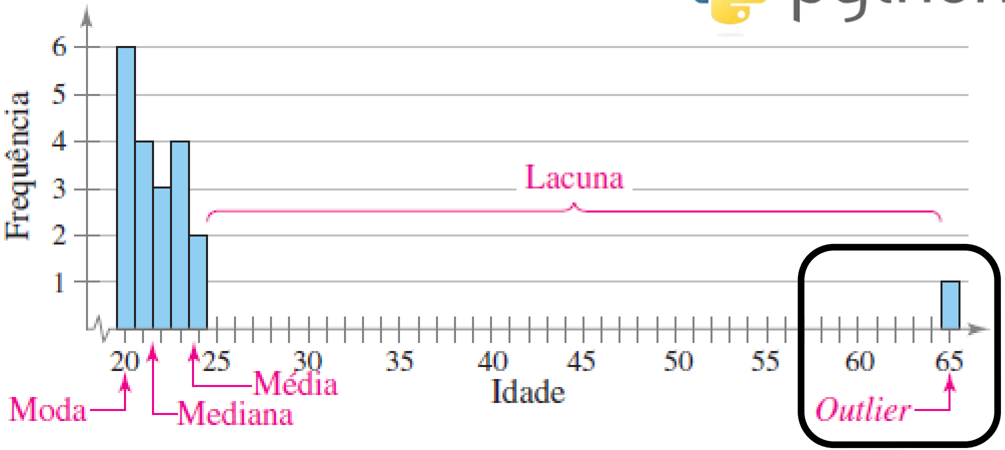
Moda: o valor que ocorre com maior frequência é 20 anos.

Medidas de tendência central

Idades em uma turma.

20	20	20	20	20	20	21
21	21	21	22	22	22	23
23	23	23	24	24	65	

Outlier



Interpretação

- A moda não representa um valor típico (região central).
- Nesse caso, a mediana é uma medida central mais adequada.

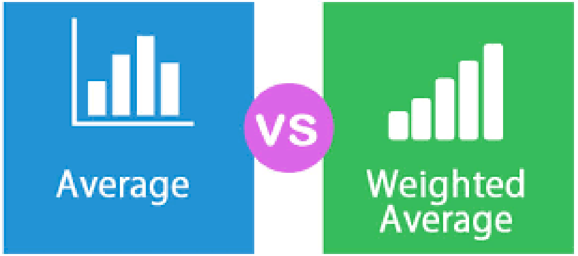
Medidas de tendência central

Definição

Uma **média ponderada** é a média de um conjunto de dados cujos valores têm pesos variados. A média ponderada é dada por:

$$\bar{x} = \frac{\Sigma(x \cdot w)}{\Sigma w}$$

em que *w* é o peso de cada valor de *x*.



Exemplo

Encontrando a média ponderada

Você está frequentando uma disciplina na qual sua nota é determinada com base em 5 fontes: 50% da média de seu teste, 15% de sua prova bimestral, 20% de sua prova final, 10% de seu trabalho no laboratório de informática e 5% de seus deveres de casa. Suas notas são: 86 (média do teste), 96 (prova bimestral), 82 (prova final), 98 (laboratório) e 100 (dever de casa). Qual é a média ponderada de suas notas? Se a média mínima para um conceito A é 90, você obteve um A?

INSTITUTO MAUÁ DE TECNOLOGIA

Medidas de tendência central

Solução Tipos de avaliação, notas e pesos para o cálculo da média ponderada.

Fonte	Nota, x	Peso, w	$x \cdot w$
Média do teste	86	0,50	43,0
Prova bimestral	96	0,15	14,4
Prova final	82	0,20	16,4
Informática	98	0,10	9,8
Dever de casa	100	0,05	5,0
		$\Sigma w = 1$	$\Sigma(x \cdot w) = 88{,}6$

$$\bar{x} = \frac{\Sigma\,(x \cdot w)}{\Sigma w} = \frac{88{,}6}{1} = 88{,}6$$

Não obteve conceito A

python™

- De início, vamos calcular a média ponderada de uma série.
- Depois, vamos estender a estratégia para painéis de dados.

Medidas de tendência central

Definição

A **média de uma distribuição de frequência** para uma amostra é aproximada por:

$$\bar{x} = \frac{\Sigma(x \cdot f)}{n} \qquad \text{Note que } n = \Sigma f$$

em que x e f são os pontos médios e as frequências de cada classe, respectivamente.

Instruções

Encontrando a média de uma distribuição de frequência

EM PALAVRAS	EM SÍMBOLOS
1. Determine o ponto médio de cada classe.	$x = \frac{(\text{limite inferior}) + (\text{limite superior})}{2}$
2. Calcule a soma dos produtos dos pontos médios pelas frequências.	$\Sigma(x \cdot f)$
3. Calcule a soma das frequências.	$n = \Sigma f$
4. Determine a média da distribuição de frequência.	$\bar{x} = \frac{\Sigma(x \cdot f)}{n}$

python™

Exemplo:
preços dos GPS

Formas das distribuições

Definição

Uma distribuição de frequência é **simétrica** quando uma linha vertical pode ser desenhada pelo meio do gráfico da distribuição e as metades resultantes são imagens espelhadas. Em termos práticos, um espelhamento aproximado pode caracterizar uma distribuição simétrica.

Uma distribuição de frequência é **uniforme** (ou **retangular**) quando todos os valores ou classes na distribuição têm frequências iguais ou aproximadamente iguais. Uma distribuição uniforme também é simétrica.

Uma distribuição de frequências é **assimétrica** quando a "cauda" do gráfico se alonga mais em um dos lados. Uma distribuição é **assimétrica à esquerda** (**assimetria negativa**) quando sua cauda se estende para a esquerda, e **assimétrica à direita** (**assimetria positiva**) quando sua cauda se estende para a direita.

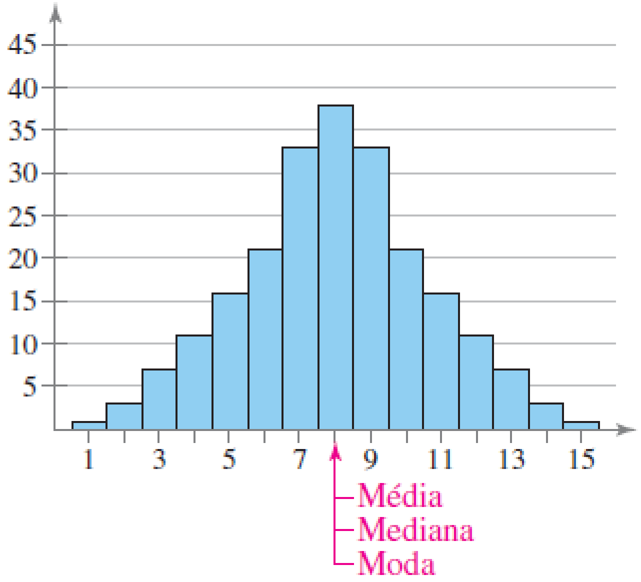
Simétrica e unimodal: média, mediana e moda são iguais

Assimétrica à esquerda: a média é menor que a mediana que é, geralmente, menor que a moda

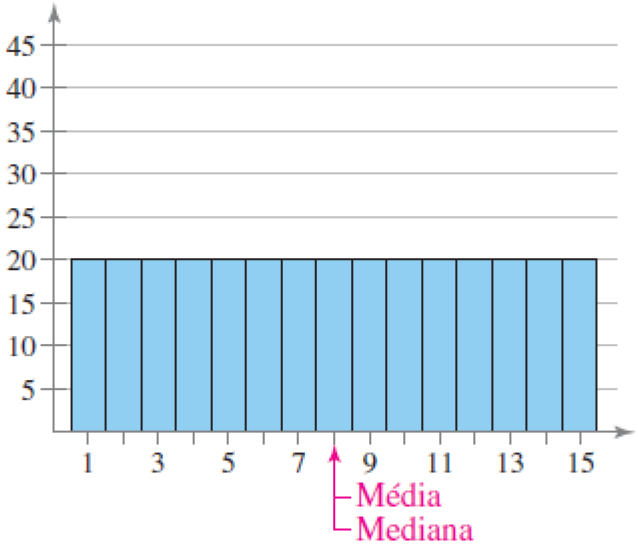
Assimétrica à direita: a média é maior que a mediana que é, geralmente, maior que a moda

Formas das distribuições

(a) Distribuição simétrica.

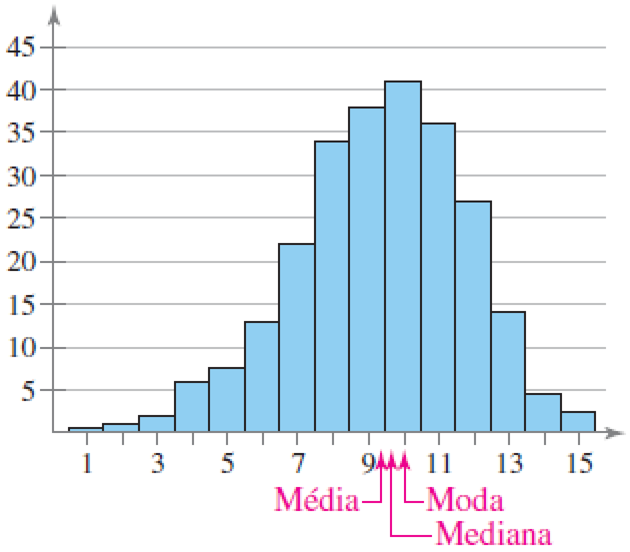


(b) Distribuição uniforme (simétrica).

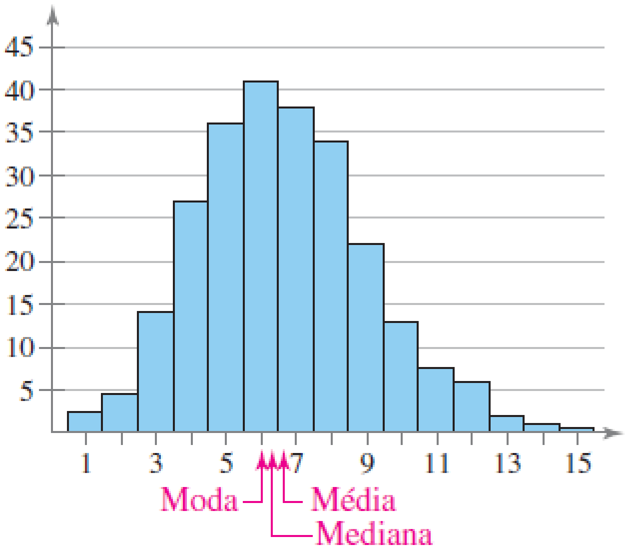


Formas das distribuições

(c) Distribuição assimétrica à esquerda.



(d) Distribuição assimétrica à direita.



TTI109 – Estatística

Aula 04 – Estatística Descritiva 03