

**Variáveis: descrição numérica
dos dados**

Medidas de tendência central

mediana e média

Onde está o centro da distribuição

Se você tiver que sugerir um único número para resumir uma variável numérica, qual você sugeriria?

Mediana

- Ordene as observações:
- Se o n é ímpar, a mediana é a observação central:
78, 91, 94, 98, 99, 101, 103, 105, 114.
mediana = 99.
- Se o n é par, divida as duas observações centrais:
78, 91, 94, 98, 99, 101, 103, 105, 114, 121.
mediana = 100.

Média

$$\bar{y} = \frac{\textit{Total}}{n} = \frac{\sum y}{n}$$

- 78, 91, 94, 98, 99, 101, 103, 105, 114.
média = 98

Mediana ou média?

Se acrescentamos 1200 (um valor extremo) na distribuição:

- 78, 91, 94, 98, 99, 101, 103, 105, 114, 1200.

mediana = 100.

- 78, 91, 94, 98, 99, 101, 103, 105, 114, 1200.

média = 208

Mediana ou média?

- A média é sensível a valores extremos.
- Como uma distribuição assimétrica, a mediana é melhor opção.
- Com uma distribuição simétrica, a média representa bem os dados.

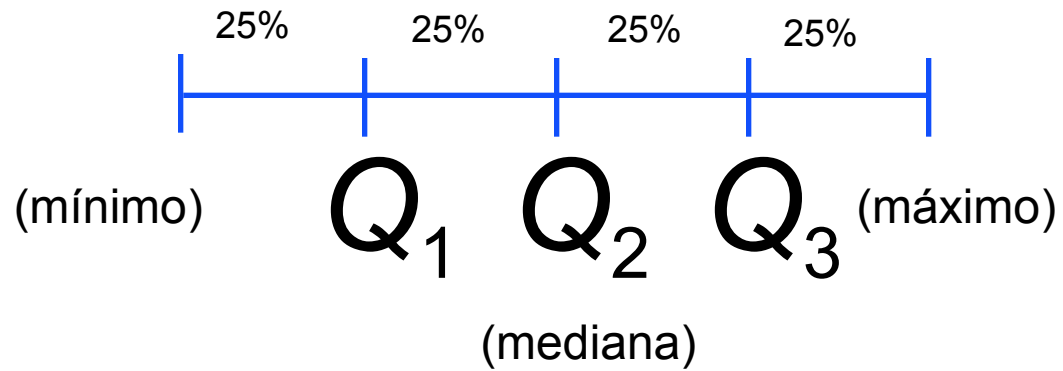
Medida de posição relativa

quartis e percentis

Quartis

- Q_1 (primeiro quartil): separe 25% dos valores ordenados dos 75% superiores.
- Q_2 (segundo quartil): separa os 50% dos valores abaixo, 50% dos valores acima
- Q_3 (terceiro quartil): separa os 75% dos valores ordenados dos 25% inferiores.

Quartis



Percentil

- Assim como existem três quartis que dividem um conjunto de dados em quatro partes, há também 99 percentis representados por P_1 , P_2 , P_3 ... P_{99} , que dividem os dados em cem partes com cerca de 1% dos valores em cada um.
- Em uma distribuição um indivíduo que tirou nota 8 em uma prova e está no 99 percentil, significa que 99% dos alunos receberam nota abaixo de 8.

Percentil

$$\text{Percentil valor de } x = \frac{\text{número de valores } < \text{ que } x}{\text{número total de valores}} \cdot 100$$

Exercício: percentil

19	19	20	20	20	20	22	22	22	22
23	23	23	23	23	23	23	24	24	24
24	24	25	25	25	25	25	25	25	26
26	26	26	26	26	27	27	28	28	30

- Numa distribuição de idades de um conjunto de alunos, encontre o percentil de um aluno que tem 23 anos.

Exercício: percentil

19	19	20	20	20	20	22	22	22	22
23	23	23	23	23	23	23	24	24	24
24	24	25	25	25	25	25	25	25	26
26	26	26	26	26	27	27	28	28	30

- Numa distribuição de idades de um conjunto de alunos, encontre o percentil de um aluno que tem 23 anos.
- Percentil de 23 = $10/40 = 25$
- 25% dos alunos têm menos do que 23 anos; 75% tem pelo menos 23 anos.

Exemplo: quartis e percentis: gastos de campanha 2010

Dinheiro_arrecadado

<hr/>				
	Percentiles	Smallest		
1%	180	8.25		
5%	606	18.07		
10%	1130	18.07	Obs	3763
25%	3325	20	Sum of Wgt.	3763
50%	14725		Mean	233124.3
		Largest	Std. Dev.	560945.7
75%	141828.8	4743205		
90%	726229	4767776	Variance	3.15e+11
95%	1334800	4953740	Skewness	3.950688
99%	2918540	5690642	Kurtosis	22.12244

Medidas de dispersão

desvio padrão e intervalo interquartil (IQR)

Desvio padrão

- O desvio padrão de um conjunto de dados, expresso pela letra s , é uma medida que expressa quanto os valores desviam da média.
- Fórmula:

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

Como calcular o desvio padrão

Salário de quatro indivíduos:

1792 1666 1362 1614 1460 1867 1439

$$\begin{aligned}\bar{X} &= \frac{1792 + 1666 + 1362 + 1614 + 1460 + 1867 + 1439}{7} \\ &= \frac{11,200}{7} \\ &= 1600\end{aligned}$$

Como calcular o desvio padrão

observações x_i	Desvios $x_i - \bar{x}$	Desvio ao Quadrado $(x_i - \bar{x})^2$
1792	1792-1600 = 192	(192) ² = 36,864
1666	1666 -1600 = 66	(66) ² = 4,356
1362	1362 -1600 = -238	(-238) ² = 56,644
1614	1614 -1600 = 14	(14) ² = 196
1460	1460 -1600 = -140	(-140) ² = 19,600
1867	1867 -1600 = 267	(267) ² = 71,289
1439	1439 -1600 = -161	(-161) ² = 25,921
soma = 0		soma = 214,870

$$s^2 = \frac{214,870}{7-1} = 35,811.67$$

$$s = \sqrt{35,811.67} = 189.24$$

Desvio padrão: propriedades

- O valor do desvio padrão é sempre positivo (nunca negativo).
- O valor do desvio padrão aumenta dramaticamente com a inclusão de um ou mais valores extremos (outliers).
- A unidade do desvio padrão é a mesma unidade dos dados originais.

O intervalo interquartil (IQR)

- Os quartis mais baixos e mais altos são os 25th e 75th percentis dos dados.
- O intervalo interquartil = $Q_3 - Q_1$
- O IQR indica qual “território” a metade central dos dados cobrem.

Exemplo: intervalo interquartil (IQR)

- O intervalo interquartil da distribuição:

Max	9.0
Q3	7.6
Median	7.0
Q1	6.6
Min	3.7

- $Q_3 - Q_1 = 1.0$
- Observamos que 50% dos dados estão entre 7.6 e 6.6; portanto próximos da mediana.

“Pares” de medidas

- Média + desvio padrão
- Mediana + intervalo interquartil.