

# ME951 - Estatística e Probabilidade I

## Parte 3

Notas de aula de ME414 produzidas pelos professores **Samara Kiihl**, **Tatiana Benaglia** e **Benilton Carvalho** modificadas e alteradas pela Profa. **Larissa Avila Matos**

# Simetria e Assimetria da Distribuição

O formato da distribuição influencia se a média será maior ou menor do que a mediana.

# Simetria e Assimetria da Distribuição

O formato da distribuição influencia se a média será maior ou menor do que a mediana.

Um valor extremamente grande na cauda direita, puxa a média para a direita.

# Simetria e Assimetria da Distribuição

O formato da distribuição influencia se a média será maior ou menor do que a mediana.

Um valor extremamente grande na cauda direita, puxa a média para a direita.

Em geral, se o formato é:

# Simetria e Assimetria da Distribuição

O formato da distribuição influencia se a média será maior ou menor do que a mediana.

Um valor extremamente grande na cauda direita, puxa a média para a direita.

Em geral, se o formato é:

- perfeitamente simétrico:  $\text{média} = \text{mediana}$ .

# Simetria e Assimetria da Distribuição

O formato da distribuição influencia se a média será maior ou menor do que a mediana.

Um valor extremamente grande na cauda direita, puxa a média para a direita.

Em geral, se o formato é:

- perfeitamente simétrico:  $\text{média} = \text{mediana}$ .
- assimétrico à direita:  $\text{média} > \text{mediana}$ .

# Simetria e Assimetria da Distribuição

O formato da distribuição influencia se a média será maior ou menor do que a mediana.

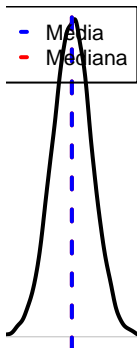
Um valor extremamente grande na cauda direita, puxa a média para a direita.

Em geral, se o formato é:

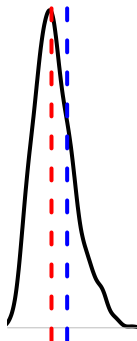
- perfeitamente simétrico:  $\text{média} = \text{mediana}$ .
- assimétrico à direita:  $\text{média} > \text{mediana}$ .
- assimétrico à esquerda:  $\text{média} < \text{mediana}$ .

# Simetria e Assimetria da Distribuição

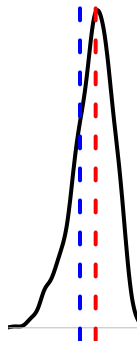
**Média=Mediana**



**Média > Mediana**



**Média < Mediana**





# Dispersão dos Dados

Considere dois conjuntos de dados:

$$A = \{10, 20, 30\}, \bar{x}_A = 20, s_A = 10.$$

$$B = \{10000, 10010, 10020\}, \bar{x}_B = 10010, s_B = 10.$$

Ambos têm o mesmo desvio padrão.

# Dispersão dos Dados

Considere dois conjuntos de dados:

$$A = \{10, 20, 30\}, \bar{x}_A = 20, s_A = 10.$$

$$B = \{10000, 10010, 10020\}, \bar{x}_B = 10010, s_B = 10.$$

Ambos têm o mesmo desvio padrão.

Se compararmos as escalas de cada conjunto de dados, poderíamos dizer que o segundo conjunto tem menor dispersão.

# Dispersão dos Dados

Considere dois conjuntos de dados:

$$A = \{10, 20, 30\}, \bar{x}_A = 20, s_A = 10.$$

$$B = \{10000, 10010, 10020\}, \bar{x}_B = 10010, s_B = 10.$$

Ambos têm o mesmo desvio padrão.

Se compararmos as escalas de cada conjunto de dados, poderíamos dizer que o segundo conjunto tem menor dispersão.

Por exemplo, a maior observação do conjunto  $B$ , 10020, é 0.2% maior do que a menor observação, 10000.

# Dispersão dos Dados

Considere dois conjuntos de dados:

$$A = \{10, 20, 30\}, \bar{x}_A = 20, s_A = 10.$$

$$B = \{10000, 10010, 10020\}, \bar{x}_B = 10010, s_B = 10.$$

Ambos têm o mesmo desvio padrão.

Se compararmos as escalas de cada conjunto de dados, poderíamos dizer que o segundo conjunto tem menor dispersão.

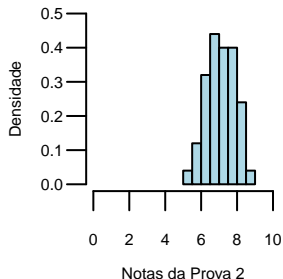
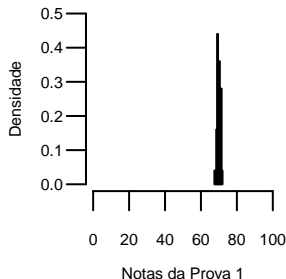
Por exemplo, a maior observação do conjunto  $B$ , 10020, é 0.2% maior do que a menor observação, 10000.

A maior observação do conjunto  $A$ , 30, é 3 vezes maior do que a menor observação, 10.

# Exemplo

Considere notas de 2 provas:

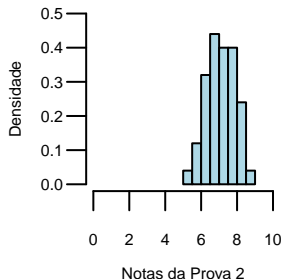
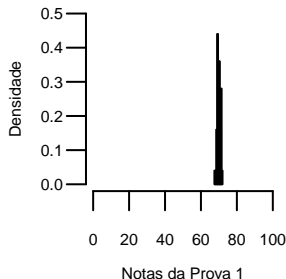
- Prova 1: 0 a 100. Média da turma: 70. Desvio padrão 1.
- Prova 2: 0 a 10. Média da turma: 7. Desvio padrão 1.



# Exemplo

Considere notas de 2 provas:

- Prova 1: 0 a 100. Média da turma: 70. Desvio padrão 1.
- Prova 2: 0 a 10. Média da turma: 7. Desvio padrão 1.



Neste caso, como as escalas são diferentes, não podemos tirar conclusões usando apenas o desvio padrão.

# Coeficiente de Variação

$$\text{Coeficiente de Variação (CV)} = \frac{s}{\bar{x}}$$

# Coeficiente de Variação

$$\text{Coeficiente de Variação (CV)} = \frac{s}{\bar{x}}$$

**Exemplo:**  $A = \{10, 20, 30\}$ ,  $\bar{x}_A = 20$ ,  $s_A = 10$ .

$B = \{10000, 10010, 10020\}$ ,  $\bar{x}_B = 10010$ ,  $s_B = 10$ .



# Coeficiente de Variação

$$\text{Coeficiente de Variação (CV)} = \frac{s}{\bar{x}}$$

**Exemplo:**  $A = \{10, 20, 30\}$ ,  $\bar{x}_A = 20$ ,  $s_A = 10$ .

$B = \{10000, 10010, 10020\}$ ,  $\bar{x}_B = 10010$ ,  $s_B = 10$ .

$CV_A = \frac{s_A}{\bar{x}_A} = 0.5$  e  $CV_B = \frac{s_B}{\bar{x}_B} \approx 0.0009$ .

# Coeficiente de Variação

$$\text{Coeficiente de Variação (CV)} = \frac{s}{\bar{x}}$$

**Exemplo:**  $A = \{10, 20, 30\}$ ,  $\bar{x}_A = 20$ ,  $s_A = 10$ .

$B = \{10000, 10010, 10020\}$ ,  $\bar{x}_B = 10010$ ,  $s_B = 10$ .

$CV_A = \frac{s_A}{\bar{x}_A} = 0.5$  e  $CV_B = \frac{s_B}{\bar{x}_B} \approx 0.0009$ .

**Exemplo:** Prova 1: 0 a 100. Média da turma: 70. Desvio padrão 1.

Prova 2: 0 a 10. Média da turma: 7. Desvio padrão 1.

# Coefficiente de Variação

$$\text{Coeficiente de Variação (CV)} = \frac{s}{\bar{x}}$$

**Exemplo:**  $A = \{10, 20, 30\}$ ,  $\bar{x}_A = 20$ ,  $s_A = 10$ .

$B = \{10000, 10010, 10020\}$ ,  $\bar{x}_B = 10010$ ,  $s_B = 10$ .

$$CV_A = \frac{s_A}{\bar{x}_A} = 0.5 \text{ e } CV_B = \frac{s_B}{\bar{x}_B} \approx 0.0009.$$

**Exemplo:** Prova 1: 0 a 100. Média da turma: 70. Desvio padrão 1.

Prova 2: 0 a 10. Média da turma: 7. Desvio padrão 1.

$$CV_1 = \frac{s_1}{\bar{x}_1} = 0.014 \text{ e } CV_2 = \frac{s_2}{\bar{x}_2} \approx 0.14.$$

# Usando medidas de posição para descrever dispersão

Média e mediana: medidas de posição **central**.

# Usando medidas de posição para descrever dispersão

Média e mediana: medidas de posição **central**.

Amplitude e desvio padrão: medidas de dispersão.

# Usando medidas de posição para descrever dispersão

Média e mediana: medidas de posição **central**.

Amplitude e desvio padrão: medidas de dispersão.

Há outros tipos de medida de posição para descrever a distribuição dos dados: **quartis** e **percentis**.

# Usando medidas de posição para descrever dispersão

Média e mediana: medidas de posição **central**.

Amplitude e desvio padrão: medidas de dispersão.

Há outros tipos de medida de posição para descrever a distribuição dos dados: **quartis** e **percentis**.

**Quartis** dividem os dados em 4 partes iguais: primeiro quartil ( $Q_1$ ), segundo quartil ( $Q_2$ ) e o terceiro quartil ( $Q_3$ ).

# Usando medidas de posição para descrever dispersão

Média e mediana: medidas de posição **central**.

Amplitude e desvio padrão: medidas de dispersão.

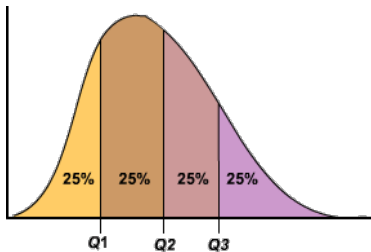
Há outros tipos de medida de posição para descrever a distribuição dos dados: **quartis** e **percentis**.

**Quartis** dividem os dados em 4 partes iguais: primeiro quartil ( $Q_1$ ), segundo quartil ( $Q_2$ ) e o terceiro quartil ( $Q_3$ ).

O **p-ésimo percentil** é o valor tal que uma porcentagem **p** dos dados ficam abaixo dele.

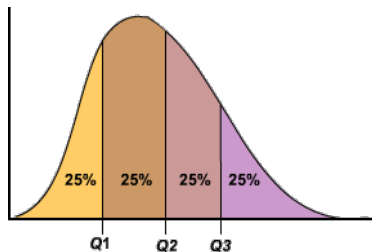


# Quartis



Para obter os quartis:

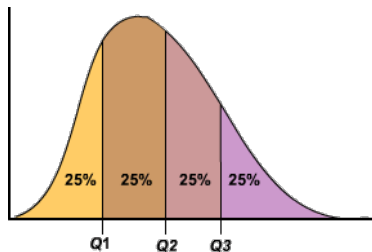
# Quartis



Para obter os quartis:

- Ordene os dados em ordem crescente.

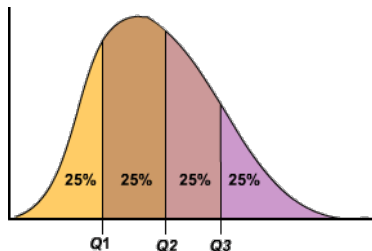
# Quartis



Para obter os quartis:

- Ordene os dados em ordem crescente.
- Encontre a mediana:  $Q_2 = \text{mediana}$ .

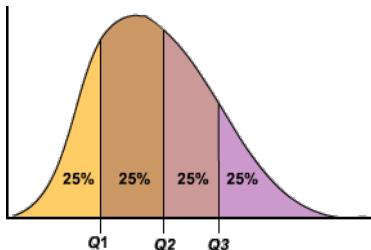
# Quartis



Para obter os quartis:

- Ordene os dados em ordem crescente.
- Encontre a mediana:  $Q_2 = \text{mediana}$ .
- Considere o subconjunto de dados abaixo da mediana.  $Q_1$  é a mediana deste subconjunto de dados.

# Quartis



Para obter os quartis:

- Ordene os dados em ordem crescente.
- Encontre a mediana:  $Q_2 = \text{mediana}$ .
- Considere o subconjunto de dados abaixo da mediana.  $Q_1$  é a mediana deste subconjunto de dados.
- Considere o subconjunto de dados acima da mediana.  $Q_3$  é a mediana deste subconjunto de dados.

## Exemplo: Quantidade de sódio (mg) em 20 cereais matinais

0, 70, 125, 125, 140, 150, 170, 170, 180, **200**

**200**, 210, 210, 220, 220, 230, 250, 260, 290, 290

Para calcular  $Q_1$ : calcula-se a mediana considerando apenas as 10 primeiras observações ordenadas:

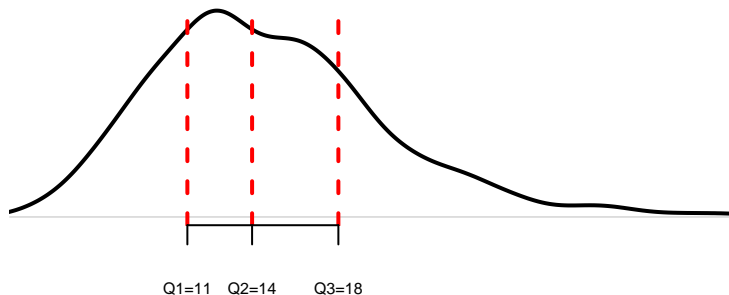
0, 70, 125, 125,  $\underbrace{140, 150}_{Q_1=145}$ , 170, 170, 180, **200**

Para calcular  $Q_3$ : calcula-se a mediana considerando apenas as 10 últimas observações ordenadas:

**200**, 210, 210, 220,  $\underbrace{220, 230}_{Q_3=225}$ , 250, 260, 290, 290

# Quartis e Assimetria

Os quartis também fornecem informação sobre o formato da distribuição.



A mediana  $Q_2$  é 14. A distância entre  $Q_1$  e  $Q_2$  é 3, enquanto que a distância entre  $Q_2$  e  $Q_3$  é 4, indicando que a distribuição é assimétrica à direita.

# Usando quartis para medir dispersão

A vantagem do uso de quartis sobre o desvio padrão ou a amplitude, é que os quartis são mais resistentes a dados extremos.



# Usando quartis para medir dispersão

A vantagem do uso de quartis sobre o desvio padrão ou a amplitude, é que os quartis são mais resistentes a dados extremos.

**Intervalo interquartílico (IQ)**  $= Q_3 - Q_1$ .

# Usando quartis para medir dispersão

A vantagem do uso de quartis sobre o desvio padrão ou a amplitude, é que os quartis são mais resistentes a dados extremos.

**Intervalo interquartílico (IQ)** =  $Q_3 - Q_1$ .

Representa 50% dos dados localizados na parte central da distribuição.

# Usando quartis para medir dispersão

A vantagem do uso de quartis sobre o desvio padrão ou a amplitude, é que os quartis são mais resistentes a dados extremos.

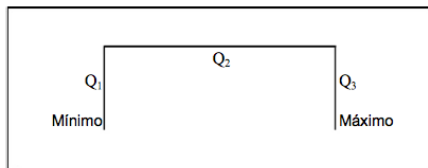
**Intervalo interquartílico (IQ)** =  $Q_3 - Q_1$ .

Representa 50% dos dados localizados na parte central da distribuição.



## Esquema dos 5 números

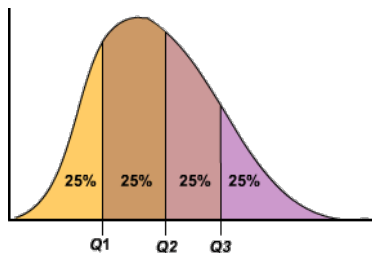
# Esquema dos 5 números



Notação:  $x_{(1)}$  = mínimo,  $x_{(n)}$  = máximo, onde  $x_{(k)}$  é a  $k$ -ésima observação depois de ordenar os dados.

$$Q_2 = \begin{cases} x_{(\frac{n+1}{2})}, & \text{se } n \text{ é ímpar} \\ \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2}, & \text{se } n \text{ é par} \end{cases}$$

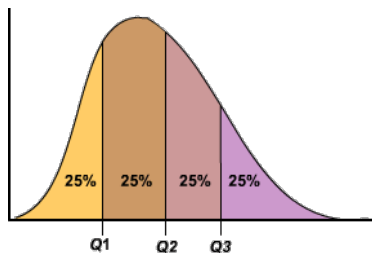
# Quartis e simetria da distribuição



Para uma distribuição simétrica ou aproximadamente simétrica:

■  $Q_2 - x_{(1)} \approx x_{(n)} - Q_2$

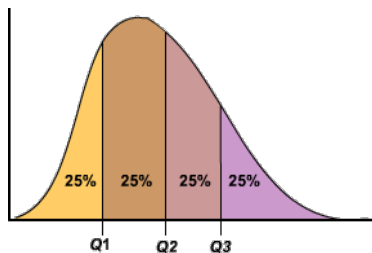
# Quartis e simetria da distribuição



Para uma distribuição simétrica ou aproximadamente simétrica:

- $Q_2 - x_{(1)} \approx x_{(n)} - Q_2$
- $Q_2 - Q_1 \approx Q_3 - Q_2$

# Quartis e simetria da distribuição

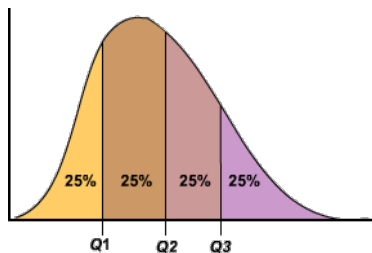


Para uma distribuição simétrica ou aproximadamente simétrica:

- $Q_2 - x_{(1)} \approx x_{(n)} - Q_2$
- $Q_2 - Q_1 \approx Q_3 - Q_2$
- $Q_1 - x_{(1)} \approx x_{(n)} - Q_3$



# Quartis e simetria da distribuição



Para uma distribuição simétrica ou aproximadamente simétrica:

- $Q_2 - x_{(1)} \approx x_{(n)} - Q_2$
- $Q_2 - Q_1 \approx Q_3 - Q_2$
- $Q_1 - x_{(1)} \approx x_{(n)} - Q_3$
- distâncias entre a mediana e  $Q_1$ ,  $Q_3$  menores do que as distâncias entre os extremos e  $Q_1$ ,  $Q_3$ .

## Dados discrepantes (*outliers*)

Importante: examinar os dados para verificar se há observações discrepantes.

# Dados discrepantes (*outliers*)

Importante: examinar os dados para verificar se há observações discrepantes.

- Média e desvio padrão são muito afetados por observações discrepantes.
- Após detectar a observação discrepante, verificar se não é um erro de digitação ou um caso especial da sua amostra.

# Dados discrepantes (*outliers*)

Importante: examinar os dados para verificar se há observações discrepantes.

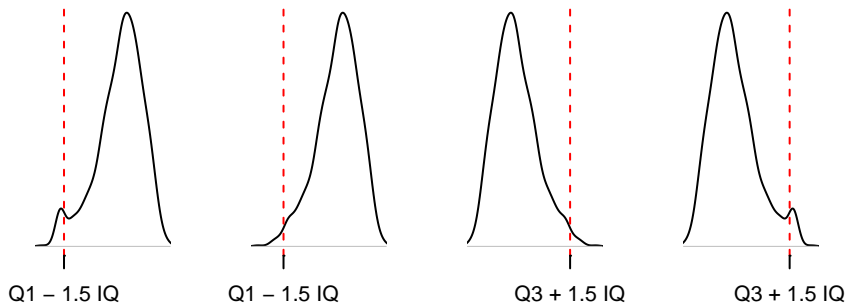
- Média e desvio padrão são muito afetados por observações discrepantes.
- Após detectar a observação discrepante, verificar se não é um erro de digitação ou um caso especial da sua amostra.
- Com poucos dados, podemos detectar um dados discrepante facilmente, apenas observando a sequência ordenada.
- Podemos usar o IQ como um critério mais geral de detecção de dados discrepantes.

## Dados discrepantes (*outliers*)

Uma observação é um potencial *outlier* se está abaixo de  $Q_1 - 1.5 \times IQ$  ou se está acima de  $Q_3 + 1.5 \times IQ$ .

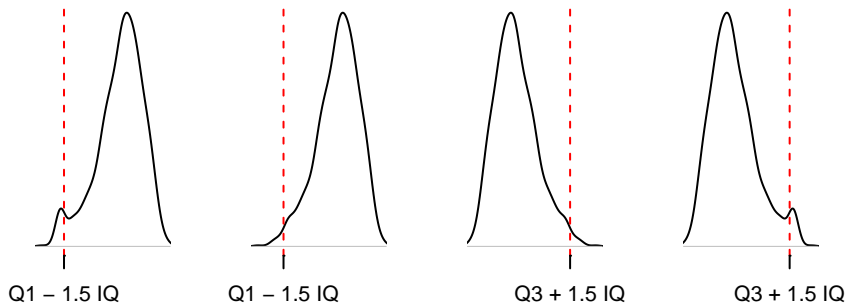
## Dados discrepantes (*outliers*)

Uma observação é um potencial *outlier* se está abaixo de  $Q_1 - 1.5 \times IQ$  ou se está acima de  $Q_3 + 1.5 \times IQ$ .



## Dados discrepantes (*outliers*)

Uma observação é um potencial *outlier* se está abaixo de  $Q_1 - 1.5 \times IQ$  ou se está acima de  $Q_3 + 1.5 \times IQ$ .



Dizemos *potencial outlier*, pois se a distribuição tem cauda longa (à direita ou à esquerda), algumas observações irão cair no critério, apesar de não serem outliers.

# Boxplot

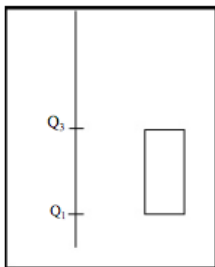
O esquema dos 5 números forma a base do gráfico denominado **boxplot**.



# Boxplot

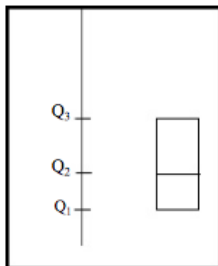
O esquema dos 5 números forma a base do gráfico denominado **boxplot**.

**Primeiro passo:** construir uma caixa que vai do primeiro ao terceiro quartil.



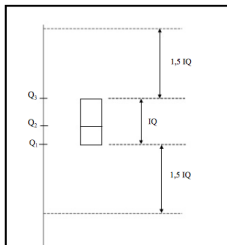
# Boxplot

**Segundo passo:** construir uma linha no meio da caixa, na altura da mediana ( $Q_2$ ).



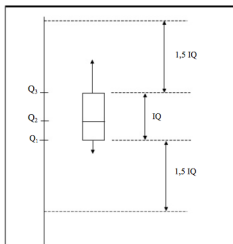
# Boxplot

**Terceiro passo:** definir os limites para que uma observação seja considerada *outlier*.



# Boxplot

**Quarto passo:** desenhar uma linha que saia da parte inferior da caixa e desça até o menor valor dos dados, mas que não ultrapasse os limites do critério de outliers. Desenhar uma linha que saia da parte superior da caixa e suba até o maior valor dos dados, mas que não ultrapasse os limites do critério de outliers. Outliers, quando existem, aparecem indicados separadamente no gráfico.



## Exemplo: População, em 1000 habitantes, dos estados brasileiros

RR	325	MS	2079	PB	3444	PR	9564
AP	478	MT	2505	GO	5004	RS	10188
AC	558	RN	2777	SC	5357	BA	13071
TO	1158	AM	2813	MA	5652	RJ	14392
RO	1380	AL	2823	PA	6193	MG	17892
SE	1785	PI	2844	CE	7431	SP	37033
DF	2052	ES	3098	PE	7919		

27 estados,  $n$  é ímpar, mediana é  $x_{(\frac{n+1}{2})} = x_{(\frac{27+1}{2})} = x_{(14)} = 3098$  (ES).

## Exemplo: População, em 1000 habitantes, dos estados brasileiros

RR	325	MS	2079	PB	3444	PR	9564
AP	478	MT	2505	GO	5004	RS	10188
AC	558	RN	2777	SC	5357	BA	13071
TO	1158	AM	2813	MA	5652	RJ	14392
RO	1380	AL	2823	PA	6193	MG	17892
SE	1785	PI	2844	CE	7431	SP	37033
DF	2052	ES	3098	PE	7919		

27 estados,  $n$  é ímpar, mediana é  $x_{(\frac{n+1}{2})} = x_{(\frac{27+1}{2})} = x_{(14)} = 3098$  (ES).

A metade inferior dos dados: 13 observações. A mediana deste subconjunto é  $Q_1 = x_{(7)} = 2052$  (DF).

## Exemplo: População, em 1000 habitantes, dos estados brasileiros

RR	325	MS	2079	PB	3444	PR	9564
AP	478	MT	2505	GO	5004	RS	10188
AC	558	RN	2777	SC	5357	BA	13071
TO	1158	AM	2813	MA	5652	RJ	14392
RO	1380	AL	2823	PA	6193	MG	17892
SE	1785	PI	2844	CE	7431	SP	37033
DF	2052	ES	3098	PE	7919		

27 estados,  $n$  é ímpar, mediana é  $x_{(\frac{n+1}{2})} = x_{(\frac{27+1}{2})} = x_{(14)} = 3098$  (ES).

A metade inferior dos dados: 13 observações. A mediana deste subconjunto é  $Q_1 = x_{(7)} = 2052$  (DF).

A metade superior dos dados: 13 observações. A mediana deste subconjunto é  $Q_3 = x_{(21)} = 7919$  (PE).

## Exemplo: População dos estados brasileiros

$$IQ = Q_3 - Q_1 = 7919 - 2052 = 5867$$



## Exemplo: População dos estados brasileiros

$$IQ = Q_3 - Q_1 = 7919 - 2052 = 5867$$

$$Q_1 - 1.5 \times IQ = -6748.5$$

## Exemplo: População dos estados brasileiros

$$IQ = Q_3 - Q_1 = 7919 - 2052 = 5867$$

$$Q_1 - 1.5 \times IQ = -6748.5$$

$$Q_3 + 1.5 \times IQ = 16720$$

## Exemplo: População dos estados brasileiros

$$IQ = Q_3 - Q_1 = 7919 - 2052 = 5867$$

$$Q_1 - 1.5 \times IQ = -6748.5$$

$$Q_3 + 1.5 \times IQ = 16720$$

Temos outliers?

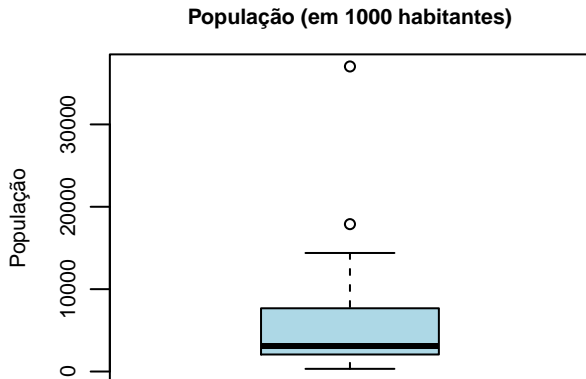
## Exemplo: População dos estados brasileiros

$$IQ = Q_3 - Q_1 = 7919 - 2052 = 5867$$

$$Q_1 - 1.5 \times IQ = -6748.5$$

$$Q_3 + 1.5 \times IQ = 16720$$

Temos outliers?



## Exemplo: Quantidade de sódio (mg) em 20 cereais matinais

$$\begin{array}{ccccccccccc} 0, & 70, & 125, & 125, & \underbrace{140, 150}, & 170, & 170, & 180, & \mathbf{200} \\ & & & & Q_1=145 & & & & \\ \mathbf{200}, & 210, & 210, & 220, & \underbrace{220, 230}, & 250, & 260, & 290, & 290 \\ & & & & Q_3=225 & & & & \end{array}$$

## Exemplo: Quantidade de sódio (mg) em 20 cereais matinais

$$\begin{array}{ccccccccccc} 0, & 70, & 125, & 125, & \underbrace{140, 150}, & 170, & 170, & 180, & \mathbf{200} \\ & & & & Q_1=145 & & & & \\ \mathbf{200}, & 210, & 210, & 220, & \underbrace{220, 230}, & 250, & 260, & 290, & 290 \\ & & & & Q_3=225 & & & & \end{array}$$

$$IQ = Q_3 - Q_1 = 225 - 145 = 80$$

## Exemplo: Quantidade de sódio (mg) em 20 cereais matinais

$$0, 70, 125, 125, \underbrace{140, 150}_{Q_1=145}, 170, 170, 180, \mathbf{200}$$
$$\mathbf{200}, 210, 210, 220, \underbrace{220, 230}_{Q_3=225}, 250, 260, 290, 290$$

$$IQ = Q_3 - Q_1 = 225 - 145 = 80$$

$$Q_1 - 1.5 \times IQ = 145 - 1.5 \times 80 = 25$$

## Exemplo: Quantidade de sódio (mg) em 20 cereais matinais

0, 70, 125, 125,  $\underbrace{140, 150}_{Q_1=145}$ , 170, 170, 180, **200**

**200**, 210, 210, 220,  $\underbrace{220, 230}_{Q_3=225}$ , 250, 260, 290, 290

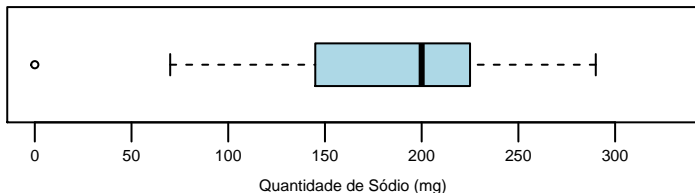
$$IQ = Q_3 - Q_1 = 225 - 145 = 80$$

$$Q_1 - 1.5 \times IQ = 145 - 1.5 \times 80 = 25$$

$$Q_3 + 1.5 \times IQ = 225 + 1.5 \times 80 = 345$$

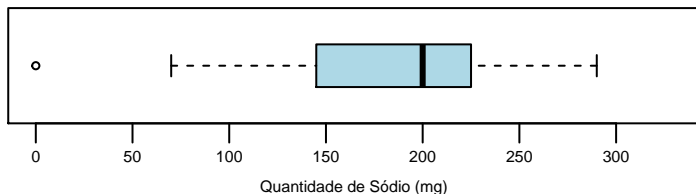


## Exemplo: Sódio em cereais matinais



Outlier: observação menor do que 25 ou maior do que 345.

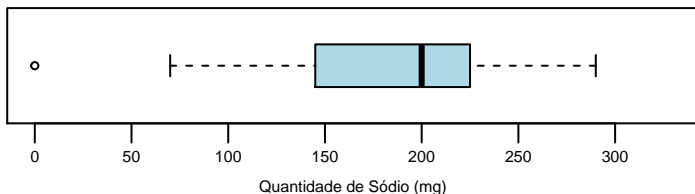
## Exemplo: Sódio em cereais matinais



Outlier: observação menor do que 25 ou maior do que 345.

As linhas pontilhadas denotam o mínimo/máximo dos dados que estão na região entre  $Q_1 - 1.5 \times IQ$  e  $Q_3 + 1.5 \times IQ$ .

## Exemplo: Sódio em cereais matinais

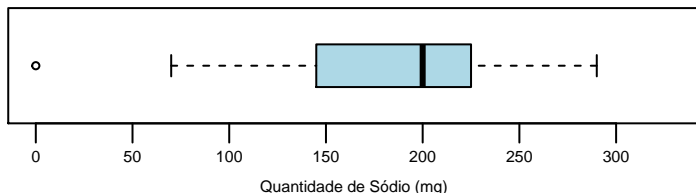


Outlier: observação menor do que 25 ou maior do que 345.

As linhas pontilhadas denotam o mínimo/máximo dos dados que estão na região entre  $Q_1 - 1.5 \times IQ$  e  $Q_3 + 1.5 \times IQ$ .

A observação máxima dos dados, 290, está no intervalo, então a linha do lado direito vai até 290.

## Exemplo: Sódio em cereais matinais



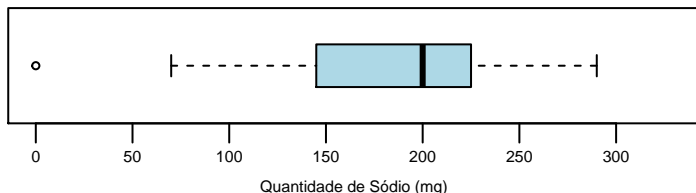
Outlier: observação menor do que 25 ou maior do que 345.

As linhas pontilhadas denotam o mínimo/máximo dos dados que estão na região entre  $Q_1 - 1.5 \times IQ$  e  $Q_3 + 1.5 \times IQ$ .

A observação máxima dos dados, 290, está no intervalo, então a linha do lado direito vai até 290.

A observação mínima dos dados, 0, está fora do intervalo (outlier=0).

## Exemplo: Sódio em cereais matinais



Outlier: observação menor do que 25 ou maior do que 345.

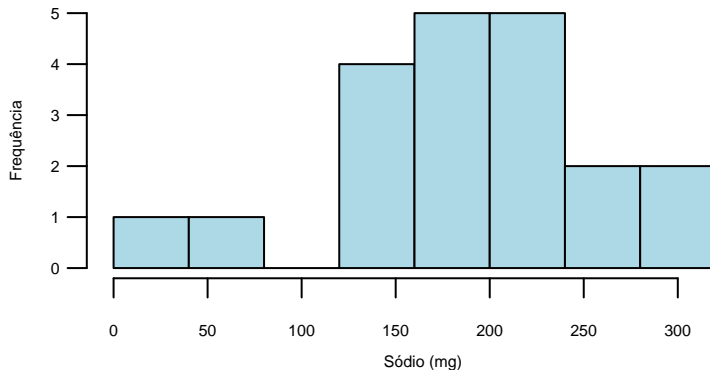
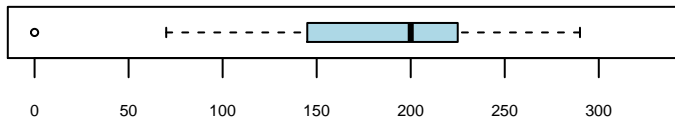
As linhas pontilhadas denotam o mínimo/máximo dos dados que estão na região entre  $Q_1 - 1.5 \times IQ$  e  $Q_3 + 1.5 \times IQ$ .

A observação máxima dos dados, 290, está no intervalo, então a linha do lado direito vai até 290.

A observação mínima dos dados, 0, está fora do intervalo (outlier=0).

Desconsiderando o outlier, o valor mínimo dos dados é 70, que está no intervalo. Portanto, a linha do lado direito vai até 70.

## Exemplo: Sódio em cereais matinais



# Boxplot x Histograma

Boxplot não substitui o histograma e vice-versa.

# Boxplot x Histograma

Boxplot não substitui o histograma e vice-versa.

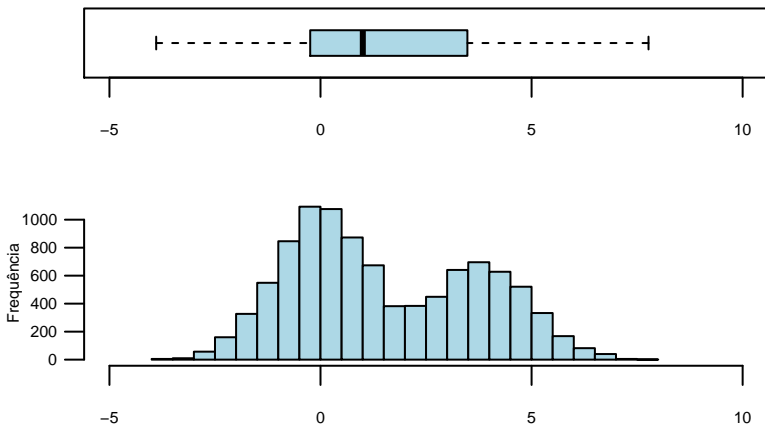
Por exemplo, se a distribuição é bimodal, não observamos isso pelo boxplot.



# Boxplot x Histograma

Boxplot não substitui o histograma e vice-versa.

Por exemplo, se a distribuição é bimodal, não observamos isso pelo boxplot.



Exemplo: Peso (em libras) de 64 alunas de educação física

$$\bar{x} = 133, Q_1 = 119, \text{ mediana} = 131.5, Q_3 = 144.$$

## Exemplo: Peso (em libras) de 64 alunas de educação física

$\bar{x} = 133$ ,  $Q_1 = 119$ , mediana=131.5,  $Q_3 = 144$ .

Como interpretar os quartis?

## Exemplo: Peso (em libras) de 64 alunas de educação física

$\bar{x} = 133$ ,  $Q_1 = 119$ , mediana=131.5,  $Q_3 = 144$ .

Como interpretar os quartis?

- 25% das alunas pesa até 119 libras.

## Exemplo: Peso (em libras) de 64 alunas de educação física

$\bar{x} = 133$ ,  $Q_1 = 119$ , mediana=131.5,  $Q_3 = 144$ .

Como interpretar os quartis?

- 25% das alunas pesa até 119 libras.
- 25% das alunas pesa mais do que 144 libras.

## Exemplo: Peso (em libras) de 64 alunas de educação física

$\bar{x} = 133$ ,  $Q_1 = 119$ , mediana=131.5,  $Q_3 = 144$ .

Como interpretar os quartis?

- 25% das alunas pesa até 119 libras.
- 25% das alunas pesa mais do que 144 libras.
- 75% das alunas pesa até 144 libras.

## Exemplo: Peso (em libras) de 64 alunas de educação física

$\bar{x} = 133$ ,  $Q_1 = 119$ , mediana=131.5,  $Q_3 = 144$ .

Como interpretar os quartis?

- 25% das alunas pesa até 119 libras.
- 25% das alunas pesa mais do que 144 libras.
- 75% das alunas pesa até 144 libras.

Você acredita que a distribuição seja simétrica?

## Exemplo: Peso (em libras) de 64 alunas de educação física

$\bar{x} = 133$ ,  $Q_1 = 119$ , mediana=131.5,  $Q_3 = 144$ .

Como interpretar os quartis?

- 25% das alunas pesa até 119 libras.
- 25% das alunas pesa mais do que 144 libras.
- 75% das alunas pesa até 144 libras.

Você acredita que a distribuição seja simétrica?

$$Q_2 - Q_1 \approx Q_3 - Q_2 \quad (?)$$



## Exemplo: Peso (em libras) de 64 alunas de educação física

$\bar{x} = 133$ ,  $Q_1 = 119$ , mediana=131.5,  $Q_3 = 144$ .

Como interpretar os quartis?

- 25% das alunas pesa até 119 libras.
- 25% das alunas pesa mais do que 144 libras.
- 75% das alunas pesa até 144 libras.

Você acredita que a distribuição seja simétrica?

$$Q_2 - Q_1 \approx Q_3 - Q_2 \quad (?)$$

$$\underbrace{Q_2 - Q_1}_{131.5 - 119 = 12.5} = \underbrace{Q_3 - Q_2}_{144 - 131.5 = 12.5}$$

## Exemplo: Taxa de desemprego na UE em 2003

	Taxa	Pais
1	8.3	Bélgica
2	6.0	Dinamarca
3	9.2	Alemanha
4	9.3	Grécia
5	11.2	Espanha
6	9.5	França
7	6.7	Portugal
8	4.4	Holanda
9	3.9	Luxemburgo
10	4.6	Irlanda
11	8.5	Itália
12	8.9	Finlândia
13	4.5	Áustria
14	6.0	Suécia
15	4.8	Reino Unido

Qual a amplitude?

## Exemplo: Taxa de desemprego na UE em 2003

	Taxa	País
1	8.3	Bélgica
2	6.0	Dinamarca
3	9.2	Alemanha
4	9.3	Grécia
5	11.2	Espanha
6	9.5	França
7	6.7	Portugal
8	4.4	Holanda
9	3.9	Luxemburgo
10	4.6	Irlanda
11	8.5	Itália
12	8.9	Finlândia
13	4.5	Áustria
14	6.0	Suécia
15	4.8	Reino Unido

Qual a amplitude?

$Q_1$ ,  $Q_3$ , mediana?

## Exemplo: Taxa de desemprego na UE em 2003

	Taxa	País
1	8.3	Bélgica
2	6.0	Dinamarca
3	9.2	Alemanha
4	9.3	Grécia
5	11.2	Espanha
6	9.5	França
7	6.7	Portugal
8	4.4	Holanda
9	3.9	Luxemburgo
10	4.6	Irlanda
11	8.5	Itália
12	8.9	Finlândia
13	4.5	Áustria
14	6.0	Suécia
15	4.8	Reino Unido

Qual a amplitude?

$Q_1$ ,  $Q_3$ , mediana?

Desenhe um boxplot.

## Exemplo: Taxa de desemprego na UE em 2003

Ordenando os dados:

3.9, 4.4, 4.5, **4.6**, 4.8, 6.0, 6.0, **6.7**, 8.3, 8.5, 8.9, **9.2**, 9.3, 9.5, 11.2

## Exemplo: Taxa de desemprego na UE em 2003

Ordenando os dados:

3.9, 4.4, 4.5, **4.6**, 4.8, 6.0, 6.0, **6.7**, 8.3, 8.5, 8.9, **9.2**, 9.3, 9.5, 11.2

Qual a amplitude?

## Exemplo: Taxa de desemprego na UE em 2003

Ordenando os dados:

3.9, 4.4, 4.5, **4.6**, 4.8, 6.0, 6.0, **6.7**, 8.3, 8.5, 8.9, **9.2**, 9.3, 9.5, 11.2

Qual a amplitude?

$$11.2 - 3.9 = 7.3$$

## Exemplo: Taxa de desemprego na UE em 2003

Ordenando os dados:

3.9, 4.4, 4.5, **4.6**, 4.8, 6.0, 6.0, **6.7**, 8.3, 8.5, 8.9, **9.2**, 9.3, 9.5, 11.2

Qual a amplitude?

$$11.2 - 3.9 = 7.3$$

$$Q_1 = 4.6, Q_3 = 9.2, \text{ mediana} = 6.7.$$



## Exemplo: Taxa de desemprego na UE em 2003

Ordenando os dados:

3.9, 4.4, 4.5, **4.6**, 4.8, 6.0, 6.0, **6.7**, 8.3, 8.5, 8.9, **9.2**, 9.3, 9.5, 11.2

Qual a amplitude?

$$11.2 - 3.9 = 7.3$$

$$Q_1 = 4.6, Q_3 = 9.2, \text{ mediana} = 6.7.$$

$$IQ = Q_3 - Q_1 = 4.6$$

## Exemplo: Taxa de desemprego na UE em 2003

Ordenando os dados:

3.9, 4.4, 4.5, **4.6**, 4.8, 6.0, 6.0, **6.7**, 8.3, 8.5, 8.9, **9.2**, 9.3, 9.5, 11.2

Qual a amplitude?

$$11.2 - 3.9 = 7.3$$

$$Q_1 = 4.6, Q_3 = 9.2, \text{ mediana} = 6.7.$$

$$IQ = Q_3 - Q_1 = 4.6$$

$$Q_1 - 1.5 \times IQ = -2.3$$

## Exemplo: Taxa de desemprego na UE em 2003

Ordenando os dados:

3.9, 4.4, 4.5, **4.6**, 4.8, 6.0, 6.0, **6.7**, 8.3, 8.5, 8.9, **9.2**, 9.3, 9.5, 11.2

Qual a amplitude?

$$11.2 - 3.9 = 7.3$$

$$Q_1 = 4.6, Q_3 = 9.2, \text{ mediana} = 6.7.$$

$$IQ = Q_3 - Q_1 = 4.6$$

$$Q_1 - 1.5 \times IQ = -2.3$$

$$Q_3 + 1.5 \times IQ = 16.1$$

## Exemplo: Taxa de desemprego na UE em 2003

Ordenando os dados:

3.9, 4.4, 4.5, **4.6**, 4.8, 6.0, 6.0, **6.7**, 8.3, 8.5, 8.9, **9.2**, 9.3, 9.5, 11.2

Qual a amplitude?

$$11.2 - 3.9 = 7.3$$

$$Q_1 = 4.6, Q_3 = 9.2, \text{ mediana} = 6.7.$$

$$IQ = Q_3 - Q_1 = 4.6$$

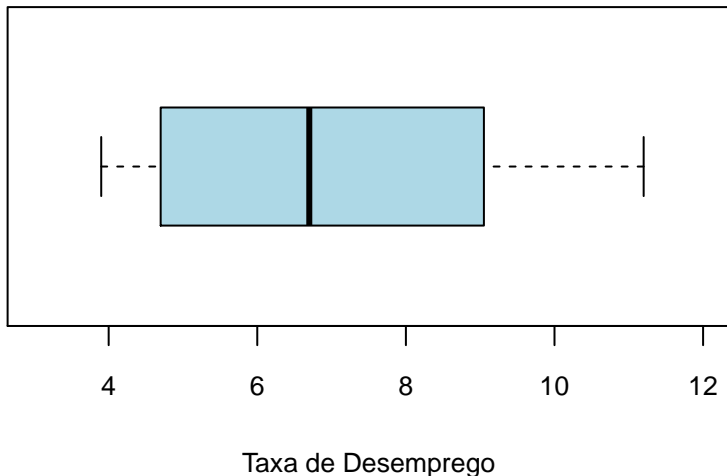
$$Q_1 - 1.5 \times IQ = -2.3$$

$$Q_3 + 1.5 \times IQ = 16.1$$

O mínimo e o máximo pertencem ao intervalo  $(-2.3, 16.1)$ , portanto as linhas pontilhadas terminam no máximo (11.2) e no mínimo (3.9).

## Exemplo: Taxa de desemprego na UE em 2003

$Q_1 = 4.6$ , mediana ( $Q_2$ )=6.7,  $Q_3 = 9.2$  e  $IQ = 4.6$



# Exercício

Deseja-se comparar 3 técnicas cirúrgicas para a extração de dente de siso. Cada uma das técnicas foi aplicada em 20 pacientes e os resultados são apresentados a seguir.

Responda:

- 1 Qual os valores aproximados para a mediana de cada técnica?
- 2 Qual é o intervalo interquartil? Conclusão?
- 3 Discuta a variabilidade do tempo de recuperação em cada técnica.
- 4 Se você é otimista, qual a técnica você escolheria?

