

**UFSC- UNIVERSIDADE FEDERAL DE SANTA CATARINA**

# **Medidas Separatrizes**

**Prof<sup>a</sup> Andréa**

# Medidas Separatrizes

São números que dividem a sequência ordenada de dados em partes que contêm a mesma quantidade de elementos da série.

Desta forma, a mediana que divide a sequência ordenada em dois grupos, cada um deles contendo 50% dos valores da sequência, é também uma medida separatriz.

Além da mediana, as outras medidas separatrizes que destacaremos são: **quartis, decis, percentis e quintis**.

# Medidas Separatrizes

**Quartis** – Se dividirmos a série ordenada em quatro partes, cada uma ficará com seus 25% de seus elementos. Os elementos que separam estes grupos são chamados de quartis.

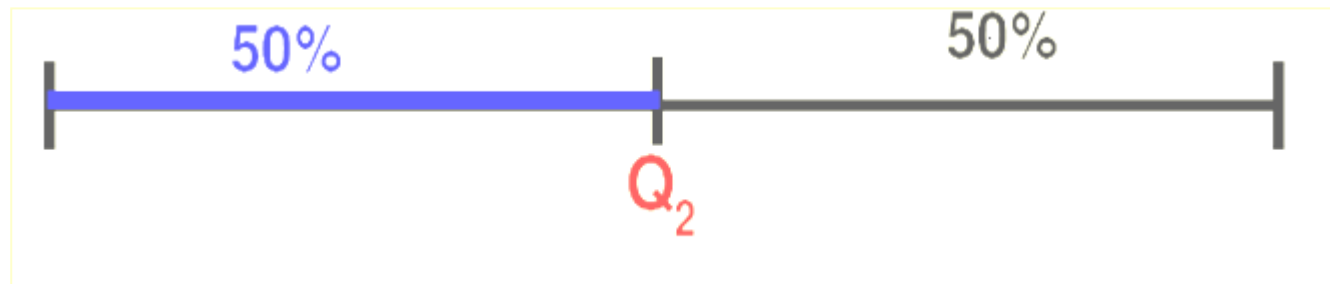
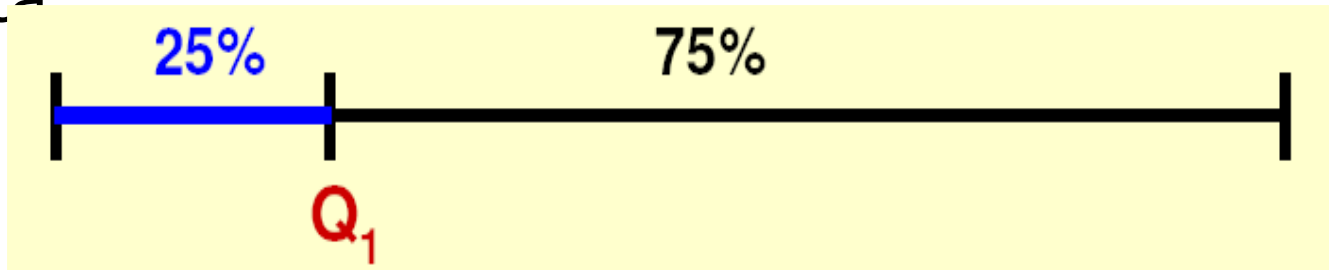
Assim, o primeiro quartil, que indicaremos por Q1, separa a seqüência ordenada deixando 25% de seus valores à esquerda e 75% de seus valores à direita.

O segundo quartil, que indicaremos por Q2, separa a seqüência ordenada deixando 50% de seus valores à esquerda e 50% de seus valores à direita. Note que o Q2 é a Mediana da série.

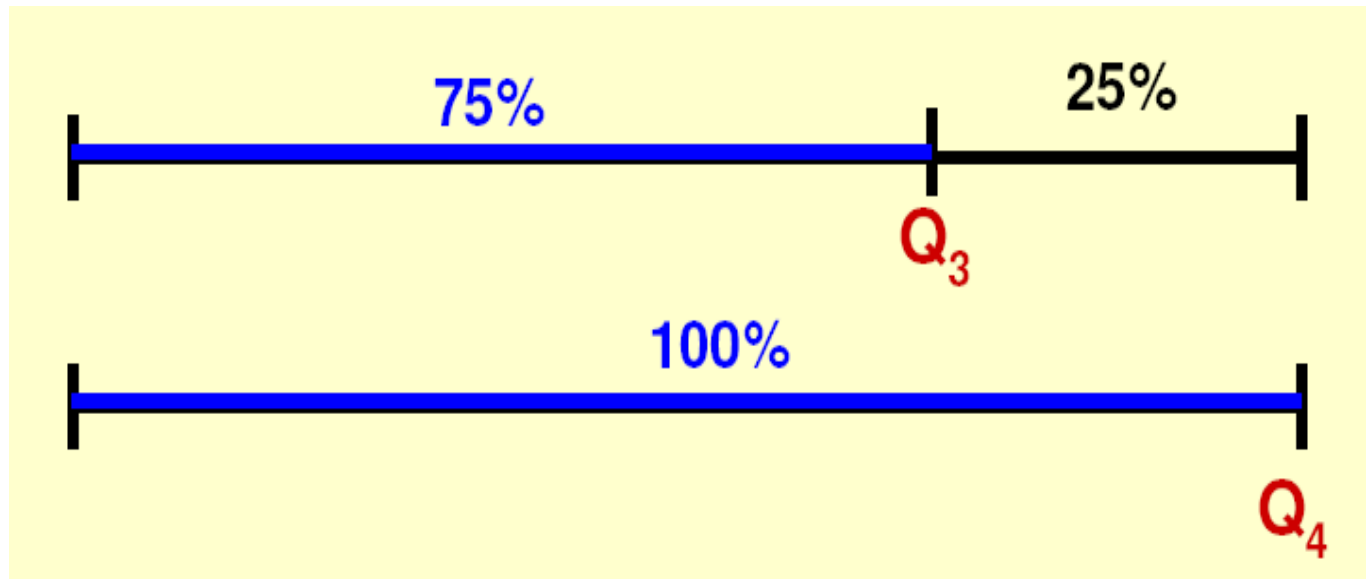
O terceiro quartil Q3 obedece a mesma regra dos anteriores.

# Medidas Separatrizes

O primeiro quartil ( $Q_1$ ) corresponde à primeira quarta parte (25%) da série, ou seja, separa 25% dos valores à esquerda dele e 75% à direita



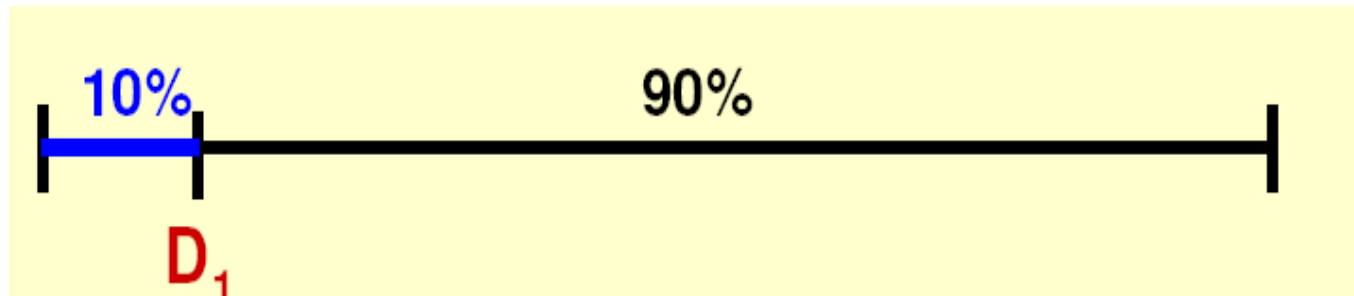
# Medidas Separatrizes



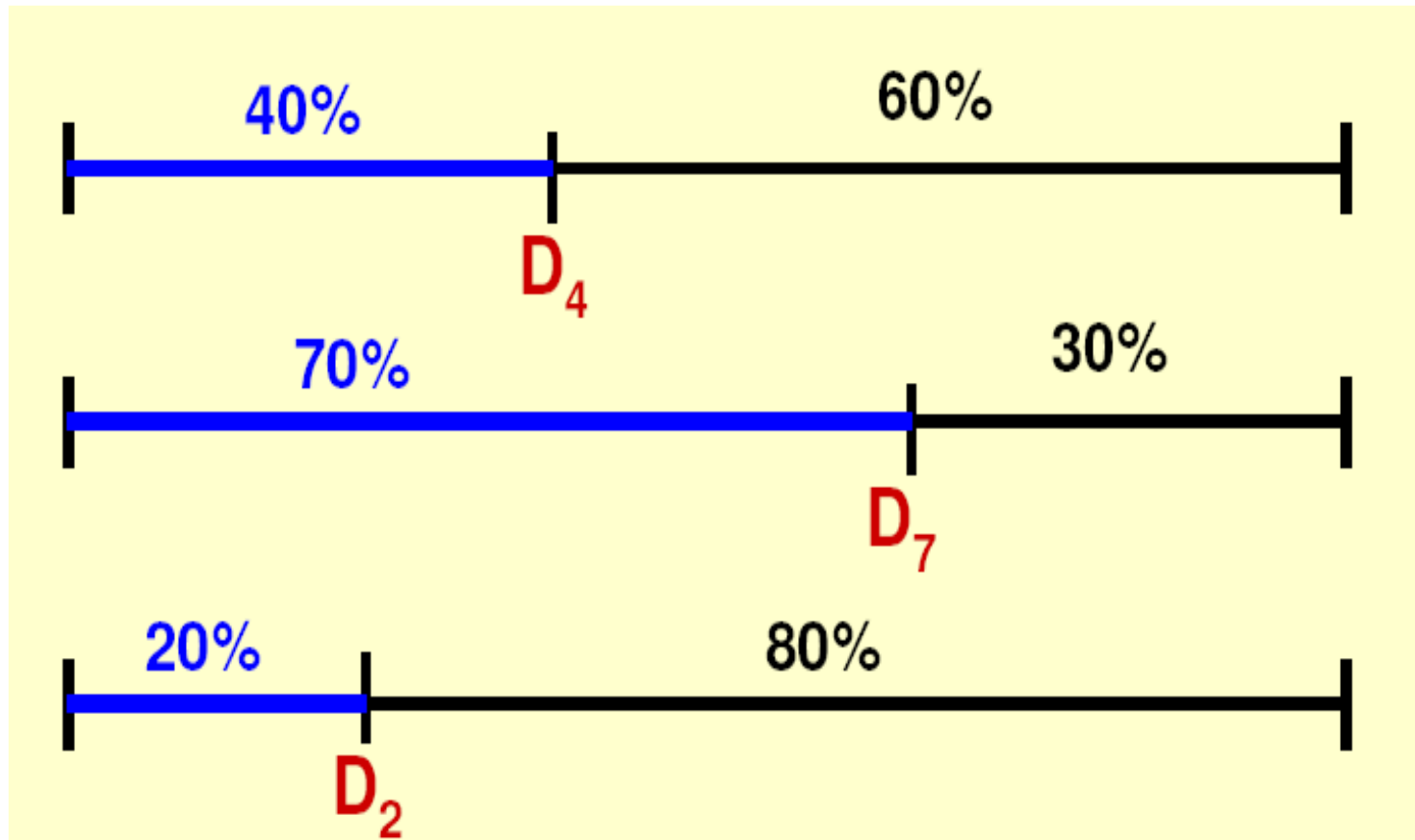
O terceiro quartil ( $Q_3$ ) está situado de tal modo que as três quartas partes (75%) dos termos são menores ou iguais a ele e uma quarta parte (25%) é maior ou igual. O quarto quartil corresponde à última quarta parte (100%) da série.

# Medidas Separatrizes

**Decis** - os decis separam uma série em 10 partes iguais, ou seja, deixam à esquerda da seqüência 10% de seus valores e 90% à direita. Eles são indicados por  $D_1$ ,  $D_2$ ,  $D_3$ , ...,  $D_{10}$ .



# Medidas Separatrizes



# Medidas Separatrizes

**Percentis** - Assim como os quartis dividem uma série de valores em quatro partes iguais, os percentis separam uma série em 100 partes iguais, em que cada parte fica com 1% de seus valores, assim, 1% fica à sua esquerda e 99% à direita. Eles são indicados por P1, P2, P3, ..., P99.



# Medidas Separatrizes

**Quintis** – Se dividirmos a série ordenada em cinco partes, cada uma ficará com seus 20% de seus elementos.

Os elementos que separam estes grupos são chamados de quintis.

Assim, o primeiro quintil, que indicaremos por  $K_1$ , separa a seqüência ordenada deixando 20% de seus valores à esquerda e 80% de seus valores à direita.

De modo análogo são definidos os outros quintis.

# Medidas Separatrizes

Para calcularmos um percentil precisamos definir qual o percentil desejado ( $i$ ), calcular sua posição por meio da Equação abaixo e verificar qual o valor de elemento que ocupa aquela posição.

$$Posição_{Pi} = \frac{n.i}{100}$$

# Medidas Separatrizes

Suponha que desejamos conhecer a idade limite que separa os 75% mais novos dos 25% mais velhos da empresa X.

**Primeiro passo: colocar em ordem crescente. Já está –próximo slide**

## Idade Func. Empresa X

Ordem	Idade
1ª	18
2ª	21
3ª	23
4ª	23
5ª	23
6ª	24
7ª	25
8ª	25
9ª	26
10ª	26

Ordem	Idade
11ª	27
12ª	27
13ª	28
14ª	28
15ª	29
16ª	32
17ª	32
18ª	33
19ª	35
20ª	36

Ordem	Idade
21ª	37
22ª	37
23ª	37
24ª	37
25ª	37
26ª	38
27ª	38
28ª	39
29ª	39
30ª	40

Ordem	Idade
31ª	40
32ª	42
33ª	42
34ª	42
35ª	43

Ordem	Idade
36ª	45
37ª	45
38ª	46
39ª	48
40ª	49

ROL

**Segundo passo: decidir qual percentil deverá ser calculado. (P75)**

**Terceiro passo: calcular a posição do percentil desejado.**

$$Posição_{Pi} = \frac{n.i}{100} = \frac{40.75}{100} = 30$$

**Quarto passo: ver na escala ordenada, qual o valor correspondente à posição calculada.**

## Idade Func. Empresa X

Ordem	Idade
1ª	18
2ª	21
3ª	23
4ª	23
5ª	23
6ª	24
7ª	25
8ª	25
9ª	26
10ª	26

Ordem	Idade
11ª	27
12ª	27
13ª	28
14ª	28
15ª	29
16ª	32
17ª	32
18ª	33
19ª	35
20ª	36

Ordem	Idade
21ª	37
22ª	37
23ª	37
24ª	37
25ª	37
26ª	38
27ª	38
28ª	39
29ª	39
30ª	40

Ordem	Idade
31ª	40
32ª	42
33ª	42
34ª	42
35ª	43

Ordem	Idade
36ª	45
37ª	45
38ª	46
39ª	48
40ª	49

ROL

# Medidas Separatrizes

Podemos afirmar que 75% dos funcionários possuem idade menor ou igual a 40 anos. De outro modo, dentre os 75% mais novos, o mais velho deles possui 40 anos.

Analisando os 25% restantes, podemos dizer que 25% possuem idade maior ou igual a 40 anos. De outro modo, dentre os 25% mais velhos, o mais novo deles possui 40 anos.

Note que o valor 40 se repete.  
Ele ocupa a 30ª e a 31ª posição no ROL.

# Medidas Separatrizes

Observe a distribuição a seguir. Ela mostra a nota de 50 alunos numa prova de recuperação .(Suponha que Desejamos calcular: a) As 15% piores notas;

Notas $X_i$	$f_i$
<b>4</b>	<b>2</b>
<b>5</b>	<b>4</b>
<b>6</b>	<b>3</b>
<b>7</b>	<b>6</b>
<b>8</b>	<b>8</b>
<b>9</b>	<b>15</b>
<b>10</b>	<b>12</b>
Total	50

Sabendo que  $i = 15$  e  $n = 50$  temos que

$$Posição_{Pi} = \frac{n.i}{100} = \frac{50.15}{100} = 7,5$$



# Medidas Separatrizes

Notas $X_i$	$f_i$	fac	$f_{acri}$	Observações
4	2	2	4%	O primeiro e o segundo tiraram 4
5	4	6	12%	3º , 4º , 5º e 6º tiraram 5
6	3	9	18%	7º , 8º e 9º tiraram 6
7	6	15	30%	Do 10º e 15º tiraram 7
8	8	23	46%	Do 16º e 23º tiraram 8
9	15	38	76%	Do 24º e 38º tiraram 9
10	12	50	100%	Do 39º e 50º tiraram 10
Total	50			

# Medidas Separatrizes

$$Posição_{Pi} = \frac{n.i}{100} = \frac{50.15}{100} = 7,5$$

P15 será a média entre o valor que ocupar a sétima e o que ocupar a oitava posição na série ordenada. Observando a frequência acumulada, vemos que ambos valem 6, então,  $P15 = (6 + 6) : 2 = 6$

Podemos afirmar que 15% tiraram nota menor ou igual a 6, e 85% nota maior ou igual a 6.

# Medidas Separatrizes

Identificamos a medida que queremos obter com o percentil correspondente,  $P_i$ .

Calculamos  $i\%$  de  $n(\sum f_i)$  para localizar a posição do percentil  $i$  no Rol, ou seja:

$$\frac{i \times \sum f_i}{100}$$

**Exemplo:** Calcule o  $D_4$  para a série

$x_i$	$f_i$	$F_{ac}$
2	3	3
4	5	8
5	8	16
7	6	22
10	2	24
	$\Sigma f_i = 24$	

# Medidas Separatrizes

$x_i$	$f_i$	fac
2	3	3
4	5	8
5	8	16
7	6	22
10	2	24
	$\Sigma f_i = 24$	

Solução:  $D_4 = P_{40}$ .

Calculamos 40% de 24 que é o número de elementos da série obtendo:

$$\frac{40 \times 24}{100} = 9,6$$

Este valor indica a posição do  $P_{40}$  é um valor compreendido entre o nono e o décimo elemento da série.

Observamos que o nono e o décimo elementos é 5.

Interpretação: 40% dos valores desta seqüência são valores menores ou iguais que 5 e 60% dos valores desta seqüência são valores maiores ou iguais que 5.

# Box-Plot ou Diagrama em Caixas

O Diagrama em Caixas, também chamado de Box-plot ou Box & Whisker plot é um gráfico que permite avaliar facilmente os valores típicos, a assimetria, a dispersão e os dados discrepantes de uma distribuição de dados de uma variável QUANTITATIVA. É indicado para grandes conjuntos de dados.

A construção do Diagrama em Caixas exige que sejam calculados previamente os valores da Mediana, primeiro quartil (Q1) e terceiro quartil (Q3) do conjunto de dados, bem como a identificação dos extremos superior (maior valor) e inferior (menor valor).

# Box-Plot ou Diagrama em Caixas

Traçam-se dois retângulos (duas caixas): um representa a “distância” entre o primeiro quartil(Q1), e a Mediana e o outro a distância entre a Mediana e o terceiro quartil(Q3) .

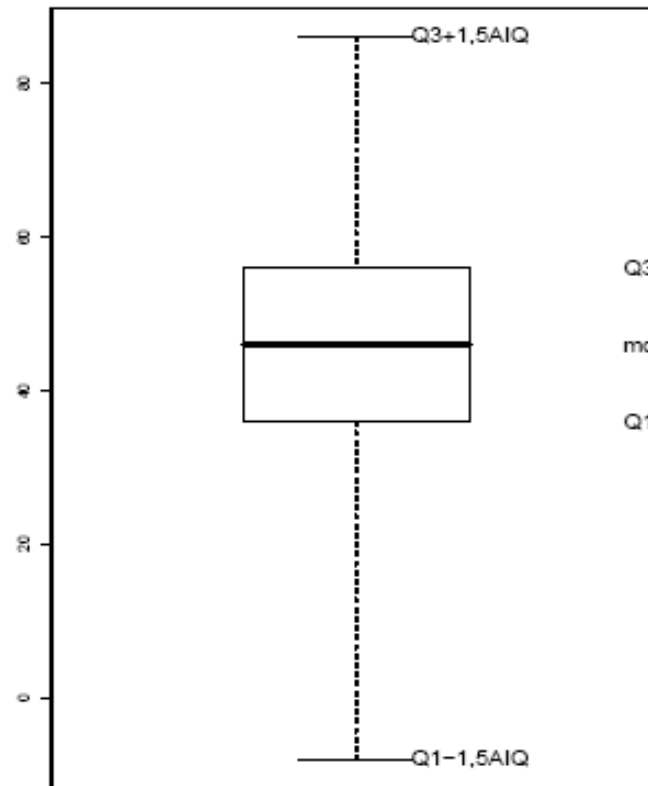
A partir dos Quartis são desenhadas linhas verticais até os últimos valores não discrepantes tanto abaixo quanto acima. Valores discrepantes (ou “outliers”) são aqueles que têm valores:

- maiores do que a expressão  $Q3 + 1,5 \times (Q3 - Q1)$  ou
- menores do que a expressão  $Q1 - 1,5 \times (Q3 - Q1)$

Todos os valores discrepantes são marcados para posterior estudo individual.

# BoxPlot ou Diagrama em Caixas

O Diagrama em Caixas “típico” seria:



# BoxPlot ou Diagrama em Caixas

Se as duas caixas tiverem “alturas” semelhantes ( $Q3 - Md \cong Md - Q1$ ) a distribuição é dita simétrica. Quanto maiores as “alturas” das caixas [maiores ( $Q3 - Md$ ) e ( $Md - Q1$ )] maior a dispersão do conjunto. O valor “típico” do conjunto será a Mediana ( $Md$ ), cujas características foram vistas anteriormente.

A seguir as etapas para construção do Diagrama em Caixas.

- 1) Ordenar os dados.
- 2) Calcular Mediana, Primeiro Quartil e Terceiro Quartil
- 3) Identificar Extremos.
- 4) Construir os retângulos ( $Q3 - Md$ ,  $Md - Q1$ ).
- 5) A partir dos retângulos, para cima e para baixo, seguem linhas até o último valor não discrepante.
- 6) Marcar as observações discrepantes.



# BoxPlot ou Diagrama em Caixas- Exemplo

Foram medidas as alturas de 35 homens adultos, cujos resultados estão abaixo.

Sejam as alturas de homens adultos a seguir:

181 174 145 150 168 173 163 184 178 165 173 165 166 205  
167 168 169 170 174 170 172 198 177 173 179 180 169 181  
169 183 142 183 163 204 165

Construa o diagrama em caixas para as alturas, avaliando valor típico, assimetria, dispersão e valores discrepantes.

# BoxPlot ou Diagrama em Caixas- Exemplo

## 1) Ordenar os dados crescentemente:

142 145 150 163 163 165 165 165 166 167 168 168 169 169 169 170  
170 172 173 173 173 174 174 177 178 179 180 181 181 183 183 184  
198 204 205

## 2) Calcular : mediana, primeiro quartil e terceiro quartil

Há 35 medidas:  $n = 35$

Posição da mediana =  $(n + 1) / 2 = 36 / 2 = 18^a \Rightarrow$  valor que está na 18ª posição. **Md = 172**

Posição do primeiro quartil =  $(n \times i) / 100 = (35 \times 25) / 100 = 8,75^a \Rightarrow$  valor que está na 8ª e 9ª posição. **Q1 = 165,5**

Posição do terceiro quartil =  $(n \times i) / 100 = (35 \times 75) / 100 = 26,25^a \Rightarrow$  valor que está na 26ª e 27ª posição. **Q3 = 179,5**

## 3) Identificar extremos

O maior valor do conjunto (extremo superior)  $Es = 205$

O menor valor do conjunto (extremo inferior)  $Ei = 142$

# Box-Plot ou Diagrama em Caixas- Exemplo

## 4) “Retângulos”

$$Q3 - Md = 179,5 - 172 = 7,5$$

$$Md - Q1 = 172 - 165,5 = 6,5$$

## 5) Identificação dos valores discrepantes

$$Q1 - 1,5 \times (Q3 - Q1) = 165,5 - 1,5(179,5 - 165,5) = 144,5$$

Valores menores do que 144,5 cm de altura serão discrepantes: só há um valor abaixo de 144,5 (142), então há apenas um valor discrepante inferior. Assim a linha vertical inferior irá até o último valor não discrepante, que vale 145,5 cm.

$$Q3 + 1,5 \times (Q3 - Q1) = 179,5 + 21 = 200,5$$

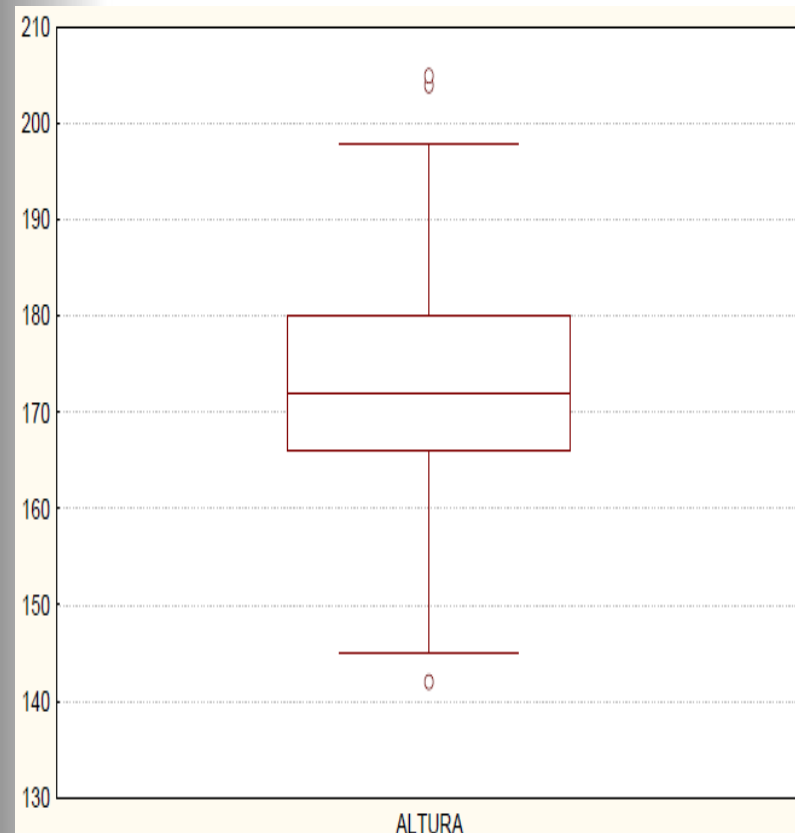
Valores maiores do que 200,5 cm de altura serão discrepantes: há dois valores acima que são (204 e 205), então há dois valores discrepantes superiores. A linha vertical superior irá até o último valor não discrepante, no caso 198.

# Box-Plot ou Diagrama em Caixas

O valor típico do conjunto é a mediana que vale 172 cm. Esse valor pode ser alto ou não, dependendo do objetivo (para selecionar jogadores de vôlei e basquete pode ser baixo, para jôqueis pode ser alto), exigindo conhecimentos mais aprofundados para ser interpretado.

As duas caixas têm “alturas” semelhantes, indicando simetria ou "leve assimetria". Quanto à dispersão não há muito o que se comentar pois não há um padrão para comparação.

Há apenas um valor discrepante inferior, e dois superiores. Estes valores talvez merecessem um estudo individual: primeiramente verificar se não houve erro de medição, se constatada a correção da medida identificar os indivíduos, estudar seu histórico médico, etc.



# Diagrama em Caixas Múltiplo

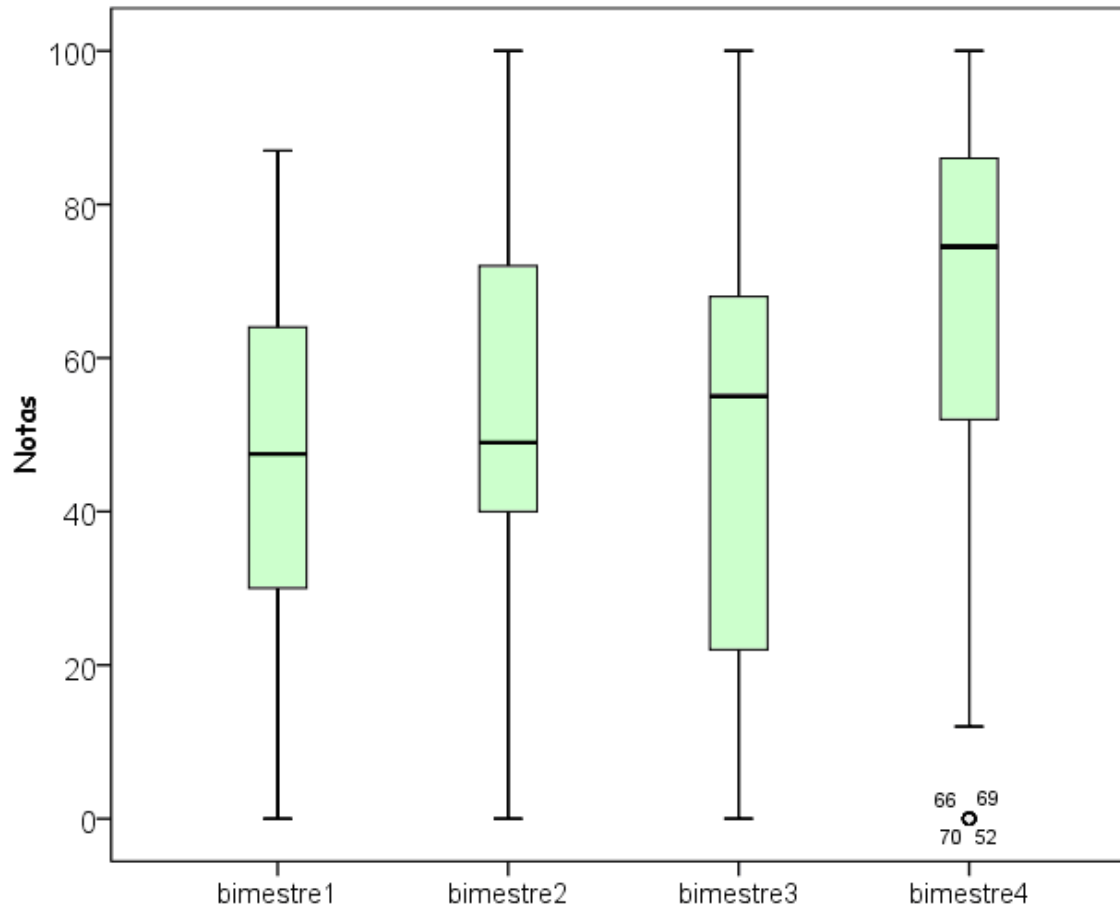
É bastante comum querer comparar vários conjuntos de dados, para avaliar seus valores típicos, dispersão, assimetria, e valores discrepantes.

Por exemplo, no Exemplo anterior, poderíamos ter interesse em comparar vários conjuntos de alturas, provenientes de diferentes grupos.

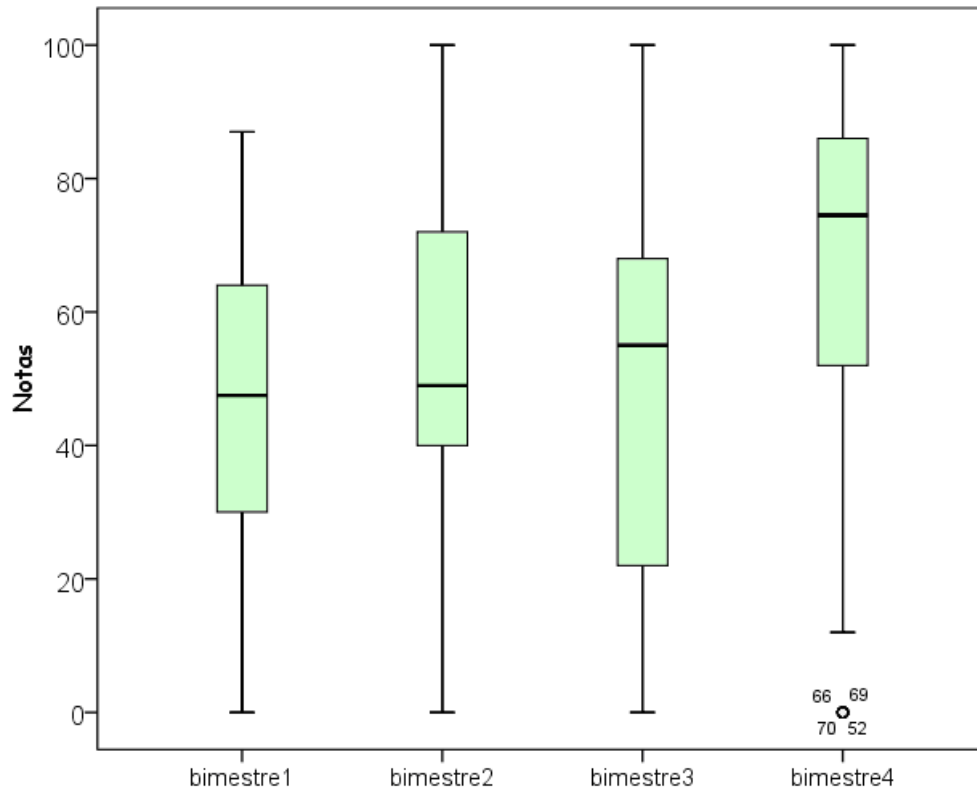
Para tanto precisamos construir um diagrama múltiplo, em que todos tenham a mesma escala, para possibilitar a comparação (diversos programas estatísticos permitem fazer isso).

# Diagrama em Caixas Múltiplo

Exemplo - O diagrama em caixas múltiplo abaixo apresenta as notas obtidas por alunos em quatro bimestres de uma disciplina anual de estatística. Faça a análise dos diagramas: valor típico, dispersão, assimetria, valores discrepantes.



# Diagrama em Caixas Múltiplo



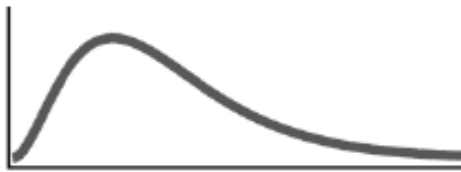
Em relação à variabilidade, observa-se que ela foi maior no segundo e terceiro bimestres, sendo neste último maior a variabilidade dos 50% dos valores centrais.

Os resultados no primeiro bimestre, foram praticamente simétricos, enquanto que no segundo bimestre, tem-se uma assimetria positiva;

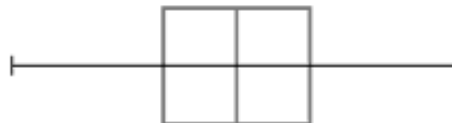
No terceiro e quarto bimestres, tem-se uma assimetria negativa. Sendo que no último bimestre, ocorreram alguns possíveis *outliers*

# BoxPlot ou Diagrama em Caixas

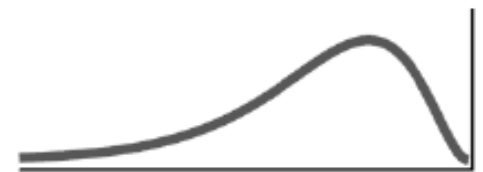
Assimetria positiva



Simetria



Assimetria negativa





# LISTA 8