

## BoxPlot para variáveis quantitativas

Prof. Gabriel Miranda

### Resumo

---

#### Definição

Também denominado **diagrama de caixa**, **diagrama de extremos e quartis** o boxplot ou box plot é uma ferramenta gráfica para representar a variação de dados observados de uma variável numérica por meio de quartis. O objetivo dessa representação é fornecer informações sobre a variabilidade dos dados e detectar a presença de pontos fora dos limites calculados, frequentemente denominados valores discrepantes ou outliers. A Figura 1 mostra um esquema boxplot, com seus respectivos elementos.

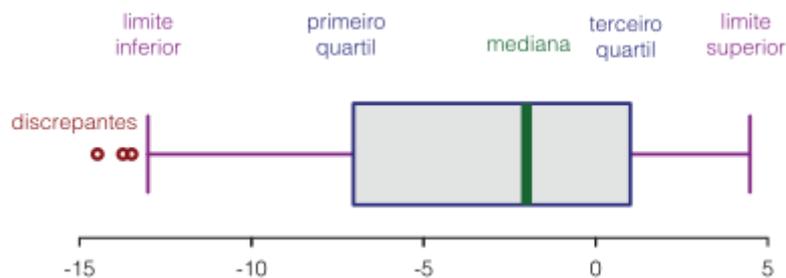


Figura 1: Boxplot e seus elementos (wikipedia)

Um boxplot possui como elementos a mediana do conjunto de dados que este representa, além dos limites inferior e superior e primeiro e terceiro quartil. Os limites inferior e superior são os valores em que respectivamente 5% e 95% das medidas são menores. Da mesma forma, o primeiro quartil representa o valor para qual 25% das medidas são menores, enquanto que o terceiro quartil são os valores em que 75% das medidas são menores.

#### Procedimento para elaboração de um BoxPlot

Usaremos como exemplo o conjunto de medidas dado por  $C = \{18,18,19,20,20,20,20,20,20,21,21,22,23,24,25,25,25,26,29,30,25,37\}$ , com 22 elementos. Colocaremos os dados em ordem crescente, atribuindo a cada medida um número de posição. como mostrado na tabela da Figura 2. A menor medida está na posição 1, o na próxima na posição 2 e assim em diante.

$i$	$x_i$
1	18
2	18
3	19
4	20
5	20
6	20
7	20
8	20
9	20
10	21
11	21
12	22
13	23
14	24
15	25
16	25
17	25
18	26
19	29
20	30
21	35
22	37

Figura 2: Valores das medidas colocados em ordem crescente.  $i$  é o índice. Por exemplo, a medida na posição  $i = 1$  é  $x_1 = 18$ , que é a menor medida. Valores repetidos são computados cada um com um índice. Assim as 6 medidas iguais a 20, cada uma tem um índice.

O procedimento segue para elaboração do bloxplot segue as seguintes etapas:

### 1. Calcular a mediana.

Colocaremos os dados em ordem crescente, associando uma ordenação a cada um deles, como mostrado na tabela da Figura 2. Como temos 22 elementos, a mediana é a média aritmética entre o décimo e o décimo primeiro valores.

$$M_d = \frac{21 + 22}{2} = 21,5$$

### 2. Calcular o primeiro e o terceiro quartis

$n = 22$ . Número de elementos par.

**Primeiro quartil (25%):**

$$k = \frac{n + 2}{4} = \frac{24}{4} = 6$$

Assim:

$$Q_1 = x_k = x_6 = 20$$

**Terceiro Quartil (75%):**

$$k = \frac{3n + 2}{4} = 17$$

assim :

$$Q_3 = x_k = x_{17} = 25$$

### 3. Calcular os limites inferior e superior

Utilizando as seguintes relações para calcular os valores limite.

Limite inferior:

$$L_i = Q_1 - 1,5.(Q_3 - Q_1) = 20 - 1,5.(25 - 20) = 12,5$$

Que é menor que a menor medida do conjunto. Quando isso ocorrer, este valor é desprezado e o limite inferior adotado é a menor medida do conjunto. No nosso exemplo, esse valor é 18.

Limite superior:

$$L_s = Q_3 + 1,5.(Q_3 - Q_1) = 25 + 1,5.(25 - 20) = 32,5$$

### 4. Traçar o gráfico

De posse dos limites, quartis e mediana, é traçado o gráfico da mesma maneira que a Figura 1. A Figura 3 mostra o boxplot do exemplo. Perceba que os vértices do retângulo horizontal ( boxplots também podem aparecer sob a forma de retângulos verticais) estão alinhados com o primeiro e terceiro quartis. As linhas são estendidas para o limite inferior, determinado como a maior medida do conjunto e o limite superior calculado. A mediana também está marcada. O gráfico permite portanto visualizar a distribuição dos valores do conjunto de dados.

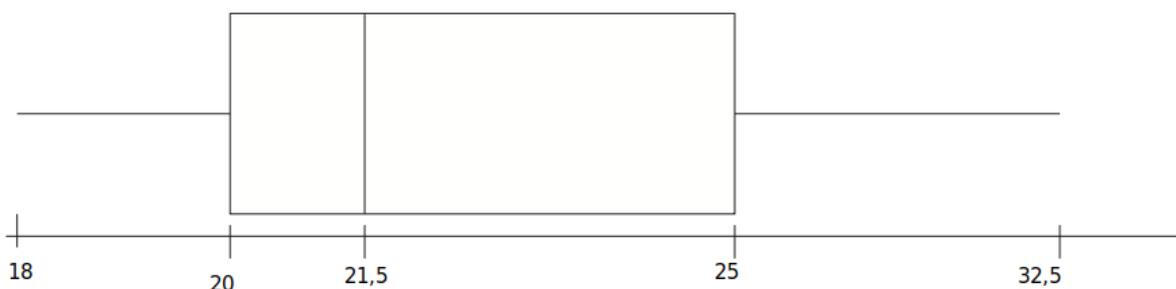


Figura 3: Boxplot resultante do exemplo