

EXERCÍCIOS BOX PLOT

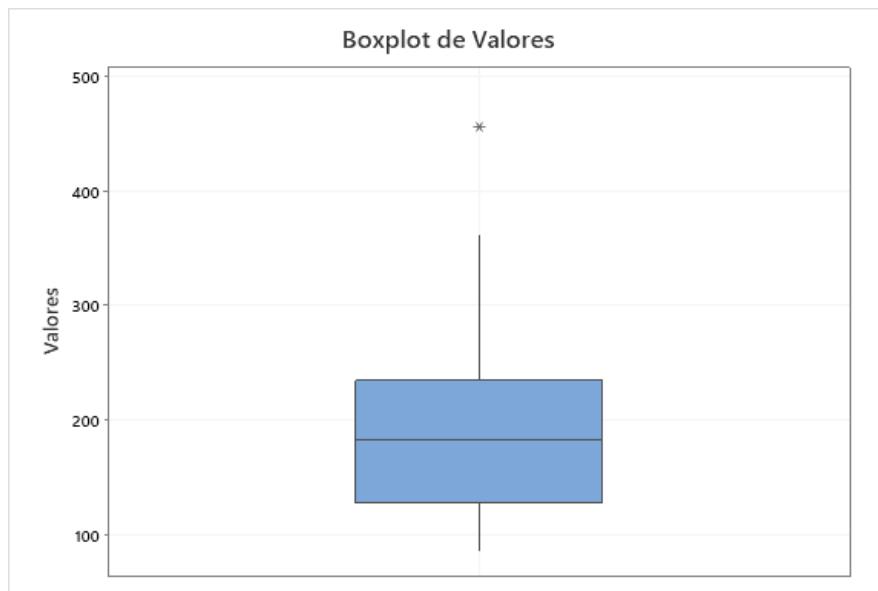
- 01.** A tabela de dados a seguir reporta o número de linhas telefônicas por mil habitantes em cada estado do Brasil em 2001.

Acre	183,8	Maranhão	86,1	Rio de Janeiro	347,5
Alagoas	125,4	M. Grosso	199,6	R. G. do Norte	150,1
Amapá	193,3	M. G. do Sul	235,3	R. G. do Sul	236,9
Amazonas	162	Minas Gerais	218,6	Rondônia	214,6
Bahia	142,3	Pará	128	Roraima	214,1
Ceará	140,6	Paraíba	125,4	Santa Catarina	257,3
D. Federal	456,8	Paraná	244,2	São Paulo	362,8
E. Santo	228,7	Pernambuco	147,8	Sergipe	140,7
Goiás	231,4	Piauí	118,2	Tocantins	113,8

- a) Determine os seis quartis;

$Q1 = 128$; Mediana = 183,8; $Q3 = 235,3$

- b) Construa o seu Box Plot;



- c) Há pontos discrepantes entre essas observações? Quais?

Sim, O ponto referente ao Distrito Federal

- 02.** A tabela a seguir contém o número de domicílios rústicos em alguns municípios do Estado do Rio de

Janeiro em 2000.

<i>Angra dos Reis</i>	572	<i>Miracema</i>	216
<i>Araruama</i>	117	<i>Niterói</i>	914
<i>Barra do Píraí</i>	741	<i>Nova Friburgo</i>	295
<i>Barra Mansa</i>	250	<i>Nova Iguaçu</i>	457
<i>Belford Roxo</i>	339	<i>Petrópolis</i>	1839
<i>Cabo Frio</i>	566	<i>Queimados</i>	81
<i>Campos</i>	1119	<i>Resende</i>	66
<i>Duque de Caxias</i>	556	<i>Rio das Ostras</i>	123
<i>Guapimirim</i>	51	<i>São Antônio de Pádua</i>	88
<i>Itaboraí</i>	132	<i>São Gonçalo</i>	394
<i>Itaguaí</i>	70	<i>São João de Meriti</i>	103
<i>Itaperuna</i>	74	<i>São Pedro da Aldeia</i>	77
<i>Japeri</i>	122	<i>Saquarema</i>	289
<i>Macaé</i>	143	<i>Seropédica</i>	159
<i>Magé</i>	567	<i>Teresópolis</i>	329
<i>Maricá</i>	64	<i>Valença</i>	229

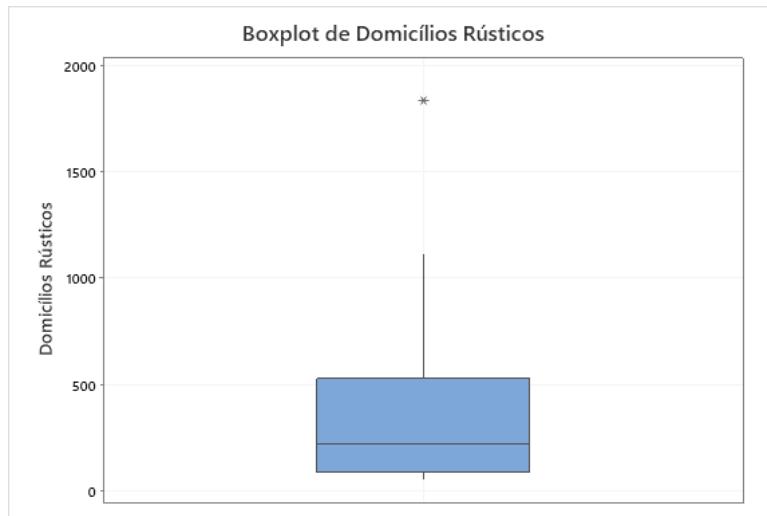
a) Determine os seis quartis;

$$Q1 = 91,75$$

$$\text{Mediana} = 222,5$$

$$Q3 = 531,25$$

b) Construa o seu Box Plot;



03. A tabela abaixo apresenta o número de gols sofridos pela seleção brasileira de futebol em cada uma das edições da copa do mundo.

04.

Ano	1930	1934	1938	1950	1954	1958	1962	1966	1970	1974
Gols sofridos	2	3	11	6	5	4	5	6	7	4
Ano	1978	1982	1986	1990	1994	1998	2002	2006	2010	
Gols sofridos	3	6	1	2	3	10	4	2	4	

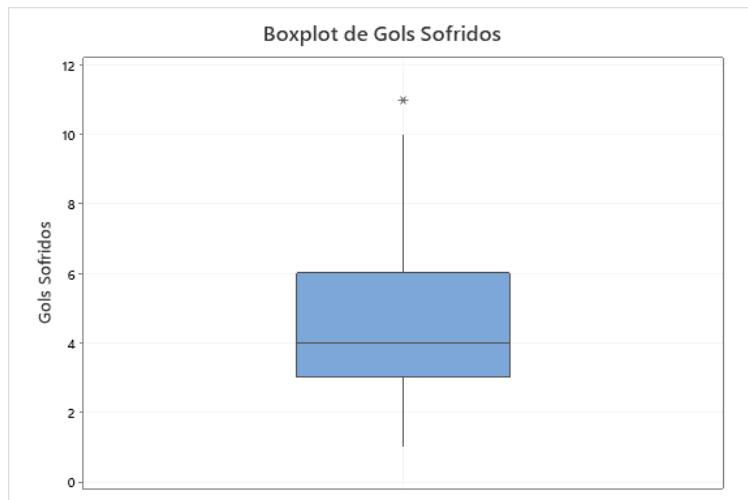
- a) Calcule os quartis e construa um box-plot. Existe algum valor que pode ser considerado discrepante?

$$Q1 = 3$$

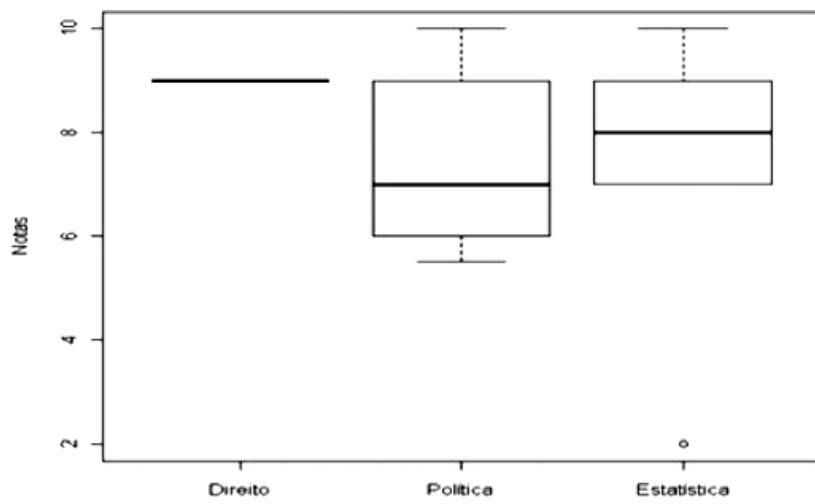
$$\text{Mediana} = 4$$

$$Q3 = 6$$

Pontos Discrepantes: Teve 1, sendo o ano de 1938

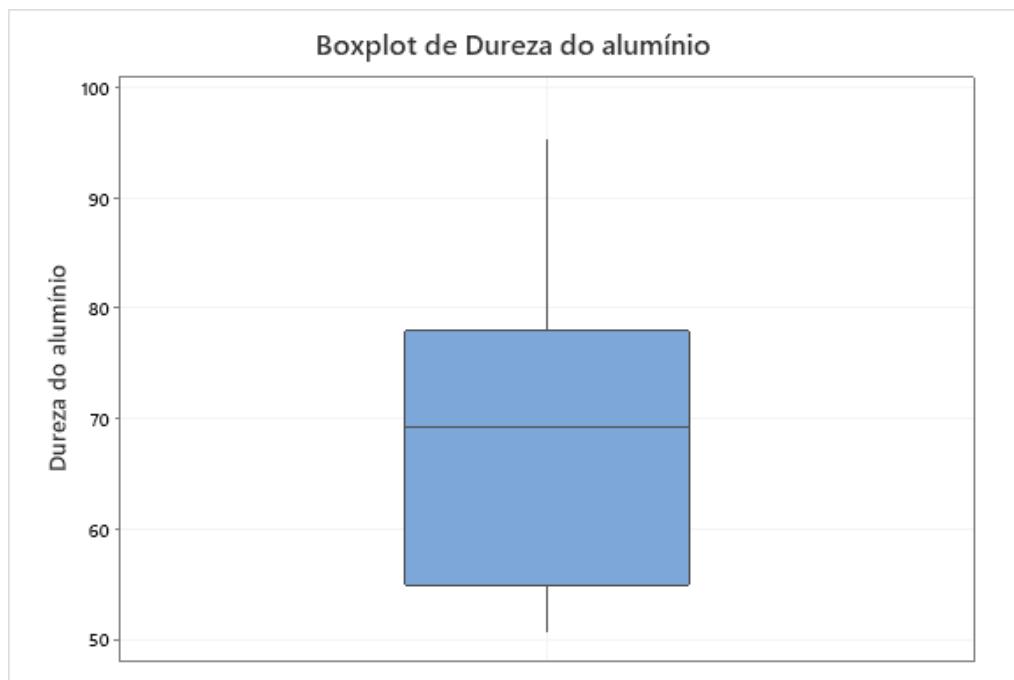


05. Compare e indique as diferenças existentes entre as distribuições das variáveis notas dos cursos de Direito, Política e Estatística, utilizando “boxplot” e as medidas calculadas em



- a) Por que em Direito o gráfico é uma reta? Pois a nota é constante (todos tiraram a mesma nota)
- b) Quem tem maior variabilidade e o porqual razão? Política possui maior variabilidade porque tem a maior caixa box-plot (tem maior desvio padrão)
- c) Porque a linha da mediana em Política não está centralizada? Pois os valores das notas no primeiro quartil estão mais próximo da mediana que os valores dos dados no terceiro quartil
- d) Existe algum dado discrepante? Sim, tem uma nota com valor 2 em estatística
06. Considere novamente os dados sobre a dureza do alumínio, e construa o boxplot dos dados.

53,0	70,2	84,3	69,5	77,8	87,5
53,4	82,5	67,3	54,1	70,5	71,4
95,4	51,1	74,4	55,7	63,5	85,8
53,5	64,3	82,7	78,5	55,7	69,1
72,3	59,5	55,3	73,0	52,4	50,7



- 07.** Cinco pessoas estão almoçando juntas. Os nomes dessas pessoas, ordenando-se em termos de idade (do mais novo pro mais velho), são: Paulo, Andrea, Jorge, Daniela e Vitor. Com base em suas idades, foram calculadas as seguintes medidas: Média = 21, Mediana = 25, diatânciam interquartil = 10, amplitude = 30 (todas elas expressas em anos) e variância = 146 anos.

OBS: Amplitude = (maior valor) – (menor valor).

- a) Quais são as idades de cada um deles?

Sejam $x_1 \leq x_2 \leq x_3 \leq x_4 \leq x_5$ as idades de Paulo, Andrea, Jorge, Daniela e Vitor, respectivamente. Então, temos:

$$\text{Mediana} = x_3 = 25, \quad (\text{I})$$

$$\text{Diatânciam Interquartil} = x_4 - x_2 = 10, \quad (\text{II})$$

$$\text{Amplitude} = x_5 - x_1 = 30, \quad (\text{III})$$

$$\text{Média} = 21 \rightarrow x_1 + x_2 + x_3 + x_4 + x_5 = 5 \times 21 = 105 \quad (\text{IV})$$

$$\text{Variância} = 146 \rightarrow x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 - 5 \times 21^2 = 4 \times 146 \quad (\text{V})$$

Substituindo (I), (II) e (III) em (IV), obtemos:

$$x_1 + x_2 + 25 + (x_2 + 10) + (x_1 + 30) = 105 \rightarrow 2x_1 + 2x_2 = 40 \rightarrow x_1 + x_2 = 20 \rightarrow x_2 = 20 - x_1 \quad (\text{VI})$$

$$\text{Além disso, de (II), } x_4 = x_2 + 10 = 20 - x_1 + 10 = 30 - x_1 \quad (\text{VII})$$

Substituindo (VI), (I), (VII) e (III) em (V), chegamos a:

$$x_1^2 + (20 - x_1)^2 + 25^2 + (30 - x_1)^2 + (30 + x_1)^2 = 4 \times 146 + 5 \times 441 = 2789.$$

Isto nos leva a equação do 2º grau: $x_1^2 - 10x_1 + 9 = 0$, cujas raízes são 9 e 1.

Experimentemos inicialmente a solução $x_1 = 9$.

Ela nos leva a: $x_2 = 20 - 9 = 11$, $x_3 = 25$, $x_4 = 30 - 9 = 21$ e $x_5 = 30 + 9 = 39$.

Essa solução não serve, porque aqui temos $x_3 = 25 > x_4 = 21$, contrariando a premissa de que os dados estão ordenados.

Experimentemos agora a solução $x_1 = 1$. Ela nos leva a:

$x_2 = 20 - 1 = 19$, $x_3 = 25$, $x_4 = 30 - 1 = 29$ e $x_5 = 30 + 1 = 31$. Solução correta!

As idades são, portanto, as seguintes: Paulo: 1, Andrea: 19, Jorge: 25, Daniela: 29 e Vitor: 31.

- b) Alguma dessas cinco observações pode ser considerada discrepante, pelo critério para identificação de outliers que se baseia nos quartis da variável? Qual?

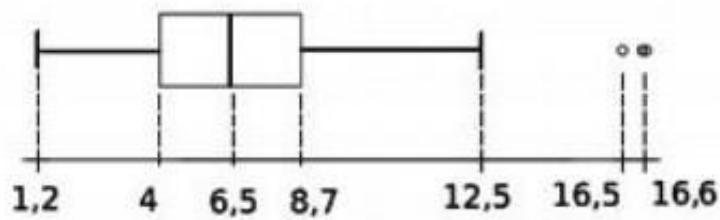
Os limites são:

$$\text{Lim. inferior} = x_2 - 1,5 \times \text{DIQ} = 19 - 1,5 \times 10 = 4$$

$$\text{Lim. superior} = x_4 + 1,5 \times \text{DIQ} = 29 + 1,5 \times 10 = 44.$$

Sendo assim, tem um ponto fora, que é Paulo com 1 ano.

- 08.** Considere uma amostra de 100 tempos de falha de um componente elétrico (meses) que tenha gerado o seguinte boxplot:



- a) Qual conjunto contém a maior quantidade de elementos da amostra referente aos tempos de falha: $(-\infty; 6,5)$ ou $(6,5; \infty)$? Justifique.

O conjunto $(6,5; \infty)$, pois esse conjunto se inicia a partir do valor mediana e a mediana está mais próxima do terceiro quartil, mostrando assim que maior parte dos dados está contido em Q3.