Tópico 1- Revisão de medidas de resumo e variabilidade

Medidas a serem revistas

- Média
- Mediana
- Variância
- Desvio-padrão
- Coeficiente de variação de Pearson
- Quartil
- Percentil

Notação:

X $\rightarrow$ variável (X:idade)

N $\rightarrow$ tamanho da população

n $\rightarrow$ tamanho da amostra

μ $\rightarrow$Média populacional (parâmetro, geralmente desconhecido)

$\overline{X}$ $\rightarrow$ Estatística (fórmula)

$\overline{x}$ $\rightarrow$ Média amostral (estimativa; valor calculado na amostra)

Média aritmética (medidas de tendência central)

Considerar

X: Número de unidades básicas de saúde (UBS) em cinco bairros de uma cidade

3; 2; 5; 6; 4

Para calcular a média, soma-se os valores e divide-se pelo número de valores

$$\text{média aritmética} = \bar{x} = \frac{3+2+5+6+4}{5} = 4\,\text{UBS}$$

Por definição, a média aritmética é o valor que ocupa o centro de equilíbrio da distribuição de valores. Portanto, a soma dos desvios em torno da média (soma das diferenças entre cada valor e a média) é igual a zero.

2-4=	-2
3-4=	-1
4-4=	0
5-4=	1
6-4=	2
Soma=	0

Apresentação em fórmula

Em uma amostra aleatória simples de tamanho n, composta pelas observações x_1, x_2, ..., x_n, a média aritmética ($\bar{x}$) é igual a:

$$\bar{x} = \frac{x_1 + x_2 + ... + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

No exemplo, n=5 e x_1=3; x_2=2, x_3=5, x_4=6, x_5=4

Portanto, $\bar{x} = \frac{3+2+5+6+4}{5} = \frac{20}{5} = 4$ UBS

Observações
A média aritmética:

- só existe para variáveis quantitativas e seu valor é único;
- é da mesma natureza da variável considerada;
- sofre influência dos valores aberrantes (outlier).

Ex: (n=5) x_1=3; x_2=2, x_3=5,x_4=6,x_5=24;. Portanto, $\bar{x} = \frac{3+2+5+6+24}{5} = \frac{40}{5} = 8$ UBS

Exemplo:

Considerar os valores de número de doenças crônicas para idosos do sexo masculino

Masculino 3; 0; 1; 3; 2; 1; 3; 0; 2; 1; 0; 6; 0; 0; 1; 2
(n=16)

Calcular o número médio ($\bar{x}$) de doenças crônicas

n=16

Soma dos valores =25

$\bar{x} = \frac{25}{16} = 1{,}6$ doenças

<u>Mediana</u>

É o valor que ocupa a posição central de uma série de n observações, quando estas estão ordenadas de forma crescente ou decrescente.

Se número de observações (n) for **ímpar**:

a mediana é o valor da variável que ocupa o posto $\frac{n+1}{2}$

Exemplo
X: Número de unidades básicas de saúde (UBS) em cinco bairros de uma cidade (n=5, um número ímpar)

3; 2; 5; 6; 4

Ordenando-se os valores

x:	2;	3;	4;	5;	6
posto	1	2	3	4	5

Como n é impar, mediana se encontra no posto $\frac{n+1}{2} = \frac{5+1}{2} = 3$

Portanto o número mediano é 4 UBS

Se o número de observações (n) for **par**:

a mediana é a média aritmética dos valores da variável que ocupam os postos $\frac{n}{2}$ e $\frac{n+2}{2}$

<u>Exemplo</u>

Considerar os valores de número de doenças crônicas para idosos do sexo masculino

Masculino 3; 0; 1; 3; 2; 1; 3; 0; 2; 1; 0; 6; 0; 0; 1; 2

Ordenando-se os valores

Valores	0	0	0	0	0	1	1	1	1	2	2	2	3	3	3	6
Postos	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16

Postos centrais

$\boldsymbol{\frac{n}{2} = \frac{16}{2} = 8}$ **e** $\boldsymbol{\frac{n+2}{2} = \frac{18}{2} = 9}$

Valores	0	0	0	0	0	1	1	1	1	2	2	2	3	3	3	6
Postos	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16

Mediana= média dos valores da variável que ocupam os postos centrais $=\bar{x} = \frac{1+1}{2} = 1$
Mediana = 1 doença crônica

Observação:

A mediana
- existe para variável quantitativa e em alguns casos também para qualitativa ordinal;
- é da mesma natureza da variável considerada;

- torna-se inadequada quando há muitos valores repetidos;
- não sofre influência de valores aberrantes.

Ex: (n=5) x_1=3; x_2=2, x_3=5,x_4=6,x_5=24;.

Média aritmética = $\bar{x} = \frac{3+2+5+6+24}{5} = \frac{40}{5} = 8$ UBS

Ordenando-se os valores: 2; 3;5;6; 24 Mediana ocupa o posto $\frac{n+1}{2} = \frac{6}{2} = 3$; portanto,

mediana = 5 UBS

Medidas de dispersão (variância, desvio-padrão, coeficiente de variação e percentis)

Constituem medidas de dispersão

- Valores mínimo e máximo
- Amplitude de variação
- Variância
- Desvio padrão
- Coeficiente de variação de Pearson

Considerar os valores que representam o número de unidades básicas (UBS) em bairros da cidade X.

3; 2; 5; 6; 4

Valores mínimo e máximo: valores extremos da distribuição.

Valor mínimo = 2 UBS; valor máximo = 6 UBS

Amplitude de variação: é a diferença entre os 2 valores extremos da distribuição.

Valor máximo - valor mínimo = 6 - 2= 4 UBS

Variância

É uma medida de dispersão que fornece a distância média ao quadrado das observações em relação á media. As distâncias de cada observação em relação à média são denominadas **desvios** em relação à média. Se forem elevados ao quadrado, serão denominados **desvios quadráticos.** Então a variância também pode ser entendida como a média dos desvios quadráticos de cada observação em relação à média aritmética.
Considerar os valores

3; 2; 5; 6; 4

$\bar{x} = 4$ UBS

Valor	(valor-média)	(valor-média)^2
3	3-4= -1 UBS	1 UBS^2
2	2-4= -2 UBS	4 UBS^2
5	5-4= 1 UBS	1 UBS^2
6	6-4= 2 UBS	4 UBS^2
4	4-4= 0 UBS	0 UBS^2
Soma =	0 UBS	10 UBS^2

Variância = $\frac{10}{5} = 2$ UBS^2

Desvio padrão

É uma medida de dispersão calculada a partir da variância sedo a raiz quadrada desta. Indica o quanto em média "erramos em média" ao representarmos um conjunto de dados pela média. É, portanto, o desvio médio dos valores em relação à média.

Desvio padrão= $\sqrt{2} = 1{,}4$ UBS
O erro médio que se comete ao resumir os dados pela média é de 1,4 UBS.

Apresentando as fórmulas:

Na população a variância é representada pelo **parâmetro** σ^2 que pode ser estimado por dois **estimadores**:

Se os dados forem referentes a toda a **população**, o **estimador** é $S^2_{(N)} = \frac{\sum_{i=1}^{N}(X_i - \bar{X})^2}{N}$

É a soma dos desvios quadráticos dos valores em relação à média dividida por N, onde N é o número de observações.

Se os dados forem referentes a uma **amostra**, o **estimador** é $S^2_{(N-1)} = \frac{\sum_{i=1}^{N}(X_i - \bar{X})^2}{N-1}$

É a soma dos desvios quadráticos dos valores em relação à média dividida por N-1, onde N é o número de observações.

Desvio padrão

Na população, o desvio padrão é um parâmetro com notação σ sendo igual à a raiz quadrada da variância, ou seja $\sigma = \sqrt{\sigma^2}$.

O estimador do desvio padrão é representado por $S = \sqrt{S^2}$

Coeficiente de Variação de Pearson (CV):

É uma medida de dispersão que relaciona a média e o desvio padrão. É representado em porcentagem. Será próximo de zero quando a dispersão for pequena, próxima a zero. Pode ser maior do que 100%. Isto ocorrerá quando a dispersão for maior que a média.

$$\mathrm{CV} = \frac{\mathrm{S}}{\overline{\mathrm{X}}} x100$$, onde S é o desvio padrão e $\overline{X}$, a média.

Notação, resumo:

Estatística	População Parâmetro	Estimador	Estimativa (com dados da amostra)
Média	μ	$\overline{X} = \frac{\sum X_i}{N}$	$\overline{x} = \frac{\sum x_i}{n}$
Variância	σ^2	$S^2_{(N)} = \frac{\sum_{i=1}^{N}(X_i - \overline{X})^2}{N}$	$s^2_{(N)} = \frac{\sum_{i=1}^{N}(x_i - \overline{X})^2}{N}$
Desvio padrão	σ	$S^2_{(N-1)} = \frac{\sum_{i=1}^{N}(X_i - \overline{X})^2}{N-1}$ $S = \sqrt{S^2}$	$s^2_{(n-1)} = \frac{\sum_{i=1}^{n}(x_i - \overline{X})^2}{n-1}$ $s = \sqrt{s^2}$

Calcule as medidas de dispersão da variável “número de doenças crônicas” para idosos do sexo masculino

Masculino	3	0	1	3	2	1	3	0	2	1	0	6	0	0	1	2	

Masculino (X)	$(x-\overline{x})$	$(x-\overline{x})^2$
3		
0		
1		
3		
2		
1		
3		
0		
2		
1		
0		
6		
0		
0		
1		
2		

Valor mínimo = Valor máximo =

Variância (n-1) = Desvio padrão (n-1)=

Coeficiente de variação de Pearson= CV(%)=

Exercício
Considere os dados da Tabela 1 do artigo Tassara MP; Guilarde AO; Rocha BAM; Féres VCR; Martelli CMT . Neurological manifestations of dengue in Central Brazil Rev Soc Bras Med Trop 50(3):379-382, May-June, 2017 doi: 10.1590/0037-8682-0444-2016

TABLE 1

Neurological manifestations of dengue in a case series of patients recruited in Central Brazil, 2005-2006.

Patient/ number	Sex	Age (years)	Fever/ days	2009 WHO classification	Neurological/ manifestation	MAC-ELISA	Virus isolation	Molecular test	Outcome
1	F	52	4	Dengue	Paresthesia/hands	Pos	Neg	DENV-3	Cured
2	F	3	5	Dengue	Paresthesia/face	Pos	Neg	---	Cured
3	F	35	6	DWS	Paresthesia/UL+LL	Pos	Neg	---	Cured
4	M	36	3	Severe	Paresthesia/hands	Pos	DENV-3	DENV-3	Cured
5	F	17	5	DWS	Paresthesia/lips	Pos	Neg	---	Cured
6	F	56	8	Severe	Encephalo	Pos	Neg	---	Cured
7	F	24	5	Severe	Paresthesia/UL+LL	Undeter.	DENV-3	DENV-3	Cured
8	F	20	6	DWS	Paresthesia/LL	Pos	Neg	---	Cured
9	F	31	9	DWS	Paresthesia/LL	Pos	Neg	---	Cured
10	F	36	7	DWS	Paresthesia/hands	Pos	Neg	---	Cured
11	F	29	7	Dengue	Paresthesia/hands	Pos	Neg	---	Cured
12	F	47	6	Dengue	Paresthesia/hands	Pos	Neg	---	Cured
13	F	41	10	Severe	Encephalo /Seizures	Pos	Neg	DENV-3	Death
14	F	45	6	Dengue	Paresthesia/LL	Pos	Neg	---	Cured
15	F	25	7	Severe	Paresthesia/LL	Pos	Neg	---	Cured
16	F	17	8	Severe	Meningoenceph	Pos	Neg	---	Cured
17	F	19	1	DWS	Paresthesia/feet	Pos	Neg	---	Cured
18	F	54	7	Dengue	Paresthesia+Paresis	Pos	Neg	---	Cured
19	F	27	6	DWS	Paresthesia	Pos	Neg	---	Cured
20	F	34	7	DWS	Paresthesia/LL	Pos	Neg	---	Cured
21	M	68	10	Severe	Encephalo	Pos	Neg	---	Cured
22	F	39	3	DWS	Paresthesia/UL+LL	Pos	Neg	---	Cured
23	M	14	5	Severe	Encephalo	Pos	Neg	---	Cured
24	M	71	15	Severe	Encephalitis/Paresis	Pos	Neg	---	Cured
25	F	38	7	DWS	Paresthesia/hands	Pos	Neg	---	Cured
26	F	15	16	Severe	Meningoenceph	Pos	Neg	DENV-3	Death
27	M	24	1	Severe	Encephalo/Seizures	Pos	Neg	---	Cured
28	F	18	3	Severe	Seizures	Pos	Neg	---	Cured

WHO: World Health Organization; **MAC-ELISA:** immunoglobulin M (IgM) antibody-capture enzyme-linked immunosorbent assay; **F:** female; **M:** male; **DENV:** dengue virus; **DWS:** dengue warning sign;; **LL:** lower limbs; **UL:** upper limbs;; **Severe:** severe dengue; **Encephalo:** encephalopathy; **Meningoenceph:** meningoencephalitis.; **Pos**: positive; **Neg**: negative

Resumir os dados dos pacientes quanto à idade e dias com febre. Calcular as medidas: média, mediana, valores mínimo e máximo, variância, desvio padrão, coeficiente de variação de Pearson.

Resposta:

Medidas	Idade (anos)	Idade (anos)	dias com febre
n	28	27	28
Média ($\bar{x}$)	33,4	34,5	6,5
Mediana (med)	32,5	34	6
valores mínimo e máximo (min; max)	3; 71	14; 71	1; 16
Variância (n) (s_n^2)	266,2	240,6	11,2
Variância (n-1) (s_{n-1}^2)	276,0	249,8	11,6
Desvio padrão (n) (s_n)	16,3	15,5	3,3
desvio padrão (n-1) (s_{n-1})	16,6	15,8	3,4
Coeficiente de Variação de Pearson CV(%) (n)	50,2%	45,6%	55,7%
Coeficiente de Variação de Pearson CV(%) (n-1)	49,8%	45,8%	52,1%

No artigo, é mencionado que o menor valor da idade é 14, entretanto na tabela 1, aparece uma idade 3. Esse valor deve estar errado porque ao se calcular a idade média incluindo-se esse valor, a idade média é 33,4 anos. No artigo, a média=34 anos (dp=15,5).

Utilizando-se a definição da média aritmética:

Por definição, a média aritmética é o valor que ocupa o centro de equilíbrio da distribuição de valores. Portanto, a soma dos desvios em torno da média (soma das diferenças entre cada valor e a média) é igual a zero.

É possível calcular o valor que está errado.

id	idade	$(x - \overline{x})$
1	52	18
2		
3	35	1
4	36	2
5	17	-17
6	56	22
7	24	-10
8	20	-14
9	31	-3
10	36	2
11	29	-5
12	47	13
13	41	7
14	45	11
15	25	-9
16	17	-17
17	19	-15
18	54	20
19	27	-7
20	34	0
21	68	34
22	39	5
23	14	-20
24	71	37
25	38	4
26	15	-19
27	24	-10
28	18	-16
n	27	
	Soma dos desvios=	14

A soma dos desvios em torno da média ($\bar{x} = 34$) devia ser zero. Isto quer dizer que o valor que está faltando, menos a média tem que ser =-14

Chamando-se de x o valor faltante; $x - 34 = -14$; portanto x=20 anos

	idades	$(x - \overline{x})$
1	52	18
2	20	-14
3	35	1
4	36	2
5	17	-17
6	56	22
7	24	-10
8	20	-14
9	31	-3
10	36	2
11	29	-5
12	47	13
13	41	7
14	45	11
15	25	-9
16	17	-17
17	19	-15
18	54	20
19	27	-7
20	34	0
21	68	34
22	39	5
23	14	-20
24	71	37
25	38	4
26	15	-19
27	24	-10
28	18	-16
n	28	0

Calculando-se a média considerando-se o valor =20 tem-se ($\bar{x} = 34\ anos\ ; dp(n) = 15{,}5\ anos$)

Quartil

Valores da variável que dividem a distribuição em quatro partes iguais.

	¼	½	¾
25%	25%	25%	25%

Q1: deixa abaixo 25% das observações

25%	75%

Q2: deixa abaixo 50% das observações

50%	50%

Q3: deixa abaixo 75% das observações

75%	25%

$$Q_1 = x_{(\frac{1}{4}(n+1))} \quad \text{e} \quad Q_3 = x_{(\frac{3}{4}(n+1))}$$

onde x é o valor da variável e $(\frac{1}{4}(n+1))$ e $(\frac{3}{4}(n+1))$ são índices que representam as posições ocupadas por x.

Os dados abaixo são referentes ao peso ao nascer de 50 recém-nascidos que tiveram síndrome de desconforto respiratório idiopático grave. 23 crianças sobreviveram e 27 foram a óbito (*).

1.050*	2.500*	1.890*	1.760	2.830
1.175*	1.030*	1.940*	1.930	1.410
1.230*	1.100*	2.200*	2.015	1.715
1.310*	1.185*	2.270*	2.090	1.720
1.500*	1.225*	2.440*	2.600	2.040
1.600*	1.262*	2.560*	2.700	2.200
1.720*	1.295*	2.730*	2.950	2.400
1.750*	1.300*	1.130	2.550	3.160
1.770*	1.550*	1.575	2.570	3.400
2.275*	1.820*	1.680	3.005	3.640

Ordenando-se os dados, em cada grupo, obtém-se:

1.030*	1.310*	2.200*	1.680	2.550
1.050*	1.500*	2.270*	1.715	2.570
1.100*	1.550*	2.275*	1.720	2.600
1.175*	1.600*	2.440*	1.760	2.700
1.185*	1.720*	2.500*	1.930	2.830
1.225*	1.750*	2.560*	2.015	2.950
1.230*	1.770*	2.730*	2.040	3.005
1.262*	1.820*	1.130	2.090	3.160
1.295*	1.890*	1.410	2.200	3.400
1.300*	1.940*	1.575	2.400	3.640

Fonte: van Vliet PK; Gupta JM. Sodium bicarbonate in idiopatic respiratory distress syndrome. *Arch. Diseases in Childhood*,1973:48, 249-255.

<u>Entre os recém-nascidos que sobreviveram:</u>

$$Q_1 = x_{(\frac{1}{4}(23+1))} = x_6 = 1720g \; ; \quad Q_3 = x_{(\frac{3}{4}(23+1))} = x_{18} = 2830g$$

$$Q_2 = x_{(\frac{1}{2}(23+1))} = x_{12} = 2200g$$

Entre os recém-nascidos que foram a óbito

$$Q_1 = x_{(\frac{1}{4}(27+1))} = x_7 = 1230g \; ; \; Q_3 = x_{(\frac{3}{4}(27+1))} = x_{21} = 2200g$$

$$Q_2 = x_{(\frac{1}{2}(27+1))} = x_{14} = 1600g$$

Se o resultado for um valor fracionário:

Por exemplo, para n=22

$$Q_1 = x_{(\frac{1}{4}(22+1))} = x_{(\frac{23}{4})} = x_{(5\frac{3}{4})}$$

que é ¾ do caminho entre x_5=1715 e x_6=1720

$$Q_1 = 1715 + \frac{3}{4}(1720 - 1715) = 1718{,}8g$$

$$Q_3 = x_{(\frac{3}{4}(22+1))} = x_{(17\frac{1}{4})}$$

que é ¼ do caminho entre x_{17}=2700 e x_{18}=2830

$$Q_3 = 2700 + \frac{1}{4}(2830 - 2700) = 2732{,}5g$$

Decil
Valores da variável que dividem a distribuição em dez partes iguais.

Percentil
Valores da variável que dividem a distribuição em cem partes iguais.

Entre os recém-nascidos que sobreviveram

Percentil 5:

$$P_5 = x_{(\frac{5}{100}(23+1))} = x_{(\frac{120}{100})} = x_{(1\frac{1}{5})}$$

que é 1/5 do caminho entre x_1=1130 e x_2=1410

$$P_5 = 1130 + \frac{1}{5}(1410 - 1130) = 1186g$$

Percentil 10:

$$P_{10} = x_{(\frac{10}{100}(23+1))} = x_{(\frac{240}{100})} = x_{(2\frac{2}{5})} \; ; \; P_{10} = 1410 + \frac{2}{5}(1575 - 1410) = 1476g$$

Percentil 50:

$$P_{50} = x_{(\frac{50}{100}(23+1))} = x_{(\frac{1200}{100})} = x_{(12)} \; ; \; P_{50} = 2200g$$

Percentil 75:

$$P_{75} = x_{(\frac{75}{100}(23+1))} = x_{(\frac{1800}{100})} = x_{(18)} \; ; \; P_{75} = 2830g$$

Percentil 90:

$$P_{90} = x_{(\frac{90}{100}(23+1))} = x_{(\frac{2160}{100})} = x_{(21\frac{3}{5})} \; ; \; P_{90} = 3160 + \frac{3}{5}(3400 - 3160) = 3304g$$

Percentil 95:

$$P_{95} = x_{(\frac{95}{100}(23+1))} = x_{(\frac{2280}{100})} = x_{(22\frac{4}{5})} \; ; \; P_{95} = 3400 + \frac{4}{5}(3640 - 3400) = 3592g$$

Box plot e identificação de valores aberrantes (*outliers*)

O Box plot representa graficamente dados de forma resumida em um retângulo onde as linhas da base e do topo são o primeiro e o terceiro quartis, respectivamente. A linha entre estas é a mediana. Linhas verticais que iniciam no meio da base e do topo do retângulo, terminam em valores denominados adjacentes inferior e superior (Chambers *et al.*, 1983, pag 60).

O valor adjacente superior é o maior valor das observações que é menor ou igual a Q3+1,5(Q3-Q1).

O valor adjacente inferior é definido como o menor valor que é maior ou igual a Q1-1,5(Q3-Q1), sendo a diferença Q3-Q1 denominada intervalo inter-quartil (IIQ).

Valores *outliers* (discrepantes ou aberrantes) são valores que "fogem" da distribuição dos dados. O box plot além de apresentar a dispersão dos dados torna-se útil também para identificar a ocorrência destes valores como sendo os que caem fora dos limites estabelecidos pelos valores adjacentes superior e inferior.

O box plot permite também investigar a dispersão e simetria dos dados.

Comentários sobre o gráfico:

Utilizando-se os dados de imc tem-se quartil 1 = 24; quartil 2 = 27 e quartil 3 = 29
Intervalo Inter quartil = 29-24= 5

VAI: Menor valor dos dados que é maior ou igual a 16,5 (24-(1,5x5)). **VAI =16,5**
VAS: Maior valor dos dados que é menor ou igual a 36,5 (24-(1,5x5)). **VAS = 35**

Não existem valores abaixo do VAI, mas existem valores acima do VAS indicando existência de dois outliers.

Dados originais

id	imc	id	imc
1	26	26	24
2	31	27	34
3	24	28	25
4	22	29	20
5	27	30	27
6	27	31	45
7	26	32	35
8	27	33	24
9	28	34	22
10	26	35	31
11	24	36	27
12	27	37	23
13	23	38	20
14	29	39	29
15	24	40	29
16	35	41	27
17	29	42	30
18	37	43	34
19	19	44	25
20	23	45	34
21	19	46	29
22	28	47	27
23	28	48	23
24	26	49	29
25	28	50	27

Ordenando-se os dados

id	imc	posto
19	19	1
21	19	2
29	20	3
38	20	4
4	22	5
34	22	6
13	23	7
20	23	8
37	23	9
48	23	10
3	24	11
11	24	12
15	24	13
26	24	14
33	24	15
28	25	16
44	25	17
1	26	18
7	26	19
10	26	20
24	26	21
5	27	22

6	27	23
8	27	24
12	27	25
30	27	26
36	27	27
41	27	28
47	27	29
50	27	30
9	28	31
22	28	32
23	28	33
25	28	34
14	29	35
17	29	36
39	29	37
40	29	38
46	29	39
49	29	40
42	30	41
2	31	42
35	31	43
27	34	44
43	34	45
45	34	46
16	35	47
32	35	48
18	37	49
31	45	50

Calcular Q1, Q2, IIQ e VAI e VAS e construir o box plot.

Estatística	Fórmula	Aplicação	Valor
Q1	$x_{\left(\frac{1}{4}\right)*(n+1)}$	$x_{\left(\frac{1}{4}\right)*(50+1)} = x_{\left(12\frac{3}{4}\right)}$	$24 + \left(0{,}75 * (24 - 24)\right) = 24$
Q2	$x_{\left(\frac{1}{2}\right)*(n+1)}$	$x_{\left(\frac{1}{2}\right)*(50+1)} = x_{\left(25\frac{1}{2}\right)}$	$27 + \left(0{,}5 * (27 - 27)\right) = 27$
Q3	$x_{\left(\frac{3}{4}\right)*(n+1)}$	$x_{\left(\frac{3}{4}\right)*(50+1)} = x_{\left(38\frac{1}{4}\right)}$	$29 + \left(0{,}25 * (29 - 29)\right) = 29$
IIQ	Q3-Q1		29-24=5
VAI	Q1-(1,5*(Q3-Q1))	24-(1,5*5)=16,5	Menor valor dos dados que é maior ou igual a 16,5 = 19
VAS	Q3+(1,5*(Q3-Q1))	29+(1,5*5)=36,5	Maior valor dos dados que é menor ou igual a 36,5 = 35
Outlier			37 e 45

Exercício

Fazer o gráfico box plot para triglicérides. Existem valores outlier?

id	triglic	id	triglic
1	128	26	89
2	166	27	92
3	79	28	181
4	166	29	91
5	61	30	171
6		31	176
7	211	32	165
8	157	33	38
9	124	34	46
10	111	35	
11	80	36	153
12	73	37	
13	205	38	99
14	101	39	66
15		40	130
16	170	41	72
17	126	42	87
18	193	43	219
19	92	44	
20	47	45	125
21	221	46	233
22	86	47	118
23	119	48	56
24	75	49	80
25	145	50	104

Valores ordenados

id	triglic	posto
33	38	1
34	46	2
20	47	3
48	56	4
5	61	5
39	66	6
41	72	7
12	73	8
24	75	9
3	79	10
11	80	11
49	80	12
22	86	13
42	87	14
26	89	15
29	91	16
19	92	17
27	92	18
38	99	19
14	101	20

50	104	21
10	111	22
47	118	23
23	119	24
9	124	25
45	125	26
17	126	27
1	128	28
40	130	29
25	145	30
36	153	31
8	157	32
32	165	33
2	166	34
4	166	35
16	170	36
30	171	37
31	176	38
28	181	39
18	193	40
13	205	41
7	211	42
43	219	43
21	221	44
46	233	45
6		
15		
35		
37		
44		

Calcular Q1, Q2, IIQ e VAI e VAS e construir o box plot.

Resposta

Estatística	Fórmula	Aplicação	Valor
Q1	$x_{\left(\frac{1}{4}\right)*(n+1)}$	$x_{\left(\frac{1}{4}\right)*(\ \ +1)} = x_{(\ \ )}$	$___ + (___ * (___ - ___)) =$
Q2	$x_{\left(\frac{1}{2}\right)*(n+1)}$	$x_{\left(\frac{1}{2}\right)*(\ \ +1)} = x_{(\ \ )}$	$___ + (___ * (___ - ___)) =$
Q3	$x_{\left(\frac{3}{4}\right)*(n+1)}$	$x_{\left(\frac{3}{4}\right)*(\ \ +1)} = x_{(\ \ )}$	$___ + (___ * (___ - ___)) =$
IIQ	Q3-Q1		____ - ____= ____
VAI	Q1-(1,5*(Q3-Q1))	___-(1,5*___)=____	Menor valor dos dados que é maior ou igual a ___ = ____
VAS	Q3+(1,5*(Q3-Q1))	___+(1,5*___)=___	Maior valor dos dados que é menor ou igual a ___ = ____
Outlier			