Medidas de tendência central (média e mediana)

Medidas de tendência central

<u>Média aritmética</u>
Notação:

X $\rightarrow$ variável

N $\rightarrow$ tamanho da população

n $\rightarrow$ tamanho da amostra

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

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

$\bar{x}$ $\rightarrow$ Média amostral (estimativa, valor calculado na amostra)

Média aritmética

Considerar

X: Número de ovos de Aedes aegypti

3	2	5	6	4

Para calcular a média soma-se os valores de uma variável e divide-se a soma pelo número de valores.

Média aritmética = $\frac{3+2+5+6+4}{5} = 4$ ovos

2	3	4	5	6

média

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

Média aritmética é o valor que ocupa o centro de equilíbrio de uma distribuição de frequências de uma variável quantitativa. Portanto, a soma das diferenças entre cada valor e a média é igual a zero.

Apresentação da estatística (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:

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

No exemplo, $x_1=3$; $x_2=2$, $x_3=5$,$x_4=6$,$x_5=4$; n=5. Portanto, $\bar{x} = \frac{3+2+5+6+4}{5} = \frac{20}{5} = 4$ ovos

OBS: 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: $x_1=3$; $x_2=2$, $x_3=5$,$x_4=6$,$x_5=20$; n=5. Portanto, $\bar{x} = \frac{3+2+5+6+24}{5} = \frac{40}{5} = 8$ ovos

Exemplo:

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

Masculino	3	0	1	3	2	1	3	0	2	1	0	6	0	0	1	2	
(n=16)																	
Feminino	1	4	4	0	2	1	2	3	2	1	3	1	2	3	3	2	3
(n=33)	1	3	3	1	3	2	3	1	3	1	0	2	2	1	2	4	

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

<u>Homens n=16</u>

	Masculino (X)
	3
	0
	1
	3
	2
	1
	3
	0
	2
	1
	0
	6
	0
	0
	1
	2
Total	

$\bar{x} = \frac{}{16} =$ doenças

Mulheres

	Feminino (X)
	1
	1
	4
	3
	4
	3
	0
	1
	2
	3
	1
	2
	2
	3
	3
	1
	2
	3
	1
	1
	3
	0
	1
	2
	2
	2
	3
	1
	3
	2
	2
	4
	3
Total	69

Exercício

Os dados a seguir são provenientes de um estudo que avaliou o tempo médio de vida em dias de 22 machos e 31 fêmeas de *Triatoma sordida*, nos estágios de ninfa e adulto, em condições de laboratório. Utilizou-se neste exemplo apenas os dados de tempo de vida em estágio de ninfa. [Fonte: Souza JMP de, 1978. *Triatoma sórdida* – Considerações sobre o tempo de vida das formas adultas e sobre a oviposição das fêmeas. Revista de Saúde Pública. São Paulo, 12:291-6.]

Calcule o número médio de dias no estágio de ninfa para machos e fêmeas:

Machos

136	157	154	129	247	164	133	126	247	139
139	148	221	248	131	139	135	143	249	173
241	241								

$\bar{x}_{Machos} =$

Fêmeas

126	126	127	130	129	128	131	126	132	136
146	128	150	136	158	134	126	128	128	139
203	208	242	241	250	244	259	241	253	234
250									

$\bar{x}_{Fêmeas} =$

Média geométrica

É a raiz n-ésima do produto de n observações $\bar{X}_G = \sqrt[n]{X_1 X_2 X_3 .... X_n} = \sqrt[n]{\prod_{i=1}^{n} X_i}$

A média geométrica também pode ser calculada como o anti logaritmo da média aritmética dos logaritmos dos valores, onde o logaritmo pode estar em qualquer base.

$$\bar{X}_G = anti\log\left(\frac{\log X_1 + \log X_2 + ... + \log X_n}{n}\right) = \frac{\sum_{i=1}^{n} \log X_i}{n}$$

É apropriada somente para valores positivos. Se os valores forem todos iguais, a média aritmética e a geométrica serão idênticas, caso contrário, $\bar{X}_G < \bar{X}$ **. É útil para razões onde se deseja dar pesos iguais a razões e para o cálculo de mudanças relativas percentuais.**

Lembrando sobre logaritmo

$\log_a b = x$ então $a^x = b$; em que $a > 0;\ a \neq 1;\ b > 0$

$\log_e a = \ln a$ é o logaritmo natural ou neperiano, em que e é um número irracional e vale aproximadamente 2,71.

mudança de base : $\log_e a = \dfrac{\log_{10} a}{\log_{10} e}$

mas $\log_{10} e \cong 0{,}43$, então $\log_e a = \dfrac{1}{0{,}43}\log_{10} a$; assim, $\log_e a = 2{,}3\log_{10} a$

Exemplo

Valor (a)	$\log_{10} a$	$\ln_e a$
3	0,477	1,099
2	0,301	0,693
5	0,699	1,609
6	0,778	1,792
4	0,602	1,386
Soma	2,857	6,579

Antilogaritmo

Se $\log_a b = x$ então $anti\log_a x = b$

Se $anti\log_a x = b$, então $a^x = b$

<u>Exemplo:</u>

$anti\log_3 2 = b$

$3^2 = b$ portanto $b = 9$

$\log_3 9 = x$; $3^x = 9$; portanto $x = 2$

<u>Exemplo:</u>

X: Número de ovos de Aedes aegypti

3	2	5	6	4

$\bar{x}_G = \sqrt[5]{3x2x5x6x4} = \sqrt[5]{\prod_{i=1}^{5} X_i} = \sqrt[5]{720} = 3{,}7$ ovos

ou

$$\bar{x}_G = anti\log\left(\frac{\log 3 + \log 2 + ... + \log 6}{5}\right) = \frac{\sum_{i=1}^{5}\log X_i}{5}$$

$\bar{x}_G = anti\ \log\left(\dfrac{2{,}857}{5}\right) = anti\ \log(0{,}5714) = 3{,}73$ ovos

$\bar{x}_G = anti\ \ln\left(\dfrac{6{,}579}{5}\right) = anti\ \ln(1{,}3158) = 3{,}73$ ovos

$\bar{x} = 4{,}0$ ovos

Exercício

Considere as observações; calcule e compare as medidas de resumo

3	2	5	6	47

Média aritmética

$\bar{x} = 12{,}6$ ovos

Média geométrica

$\bar{x}_G = 6{,}1$ ovos

Mediana

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

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

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

Quando 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 $\frac{n+2}{2}$

OBS:
- existe para variável quantitativa e qualitativa ordinal;
- é da mesma natureza da variável considerada;
- torna-se inadequada quando há muitos valores repetidos;
- não sofre influência de valores aberrantes.

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

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

n=16

Ordenando-se os valores

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

Postos centrais

Mediana= $\frac{(1+1)}{2} = 1$

Feminino	1	4	4	0	2	1	2	3	2	1	3	1	2	3	3	2	3
	1	3	3	1	3	2	3	1	3	1	0	2	2	1	2	4	

n=33

Ordenando-se os valores

Valor	Posto
0	1
0	2
1	3
1	4
1	5
1	6
1	7
1	8
1	9
1	10
1	11
2	12
2	13
2	14
2	15
2	16
2	17
2	18
2	19
2	20
3	21
3	22
3	23
3	24
3	25
3	26
3	27
3	28
3	29
3	30
4	31
4	32
4	33

Posto central $\frac{n+1}{2} = \frac{33+1}{2} = 17$

Portanto, mediana=2 doenças crônicas não infecciosas

Exemplo:

Os dados a seguir são relativos à quantidade mensal de larvas de *Aedes albopictus* coletadas em dois ambientes do Parque Ecológico do Tietê, Guarulhos, SP, no período de abril de 2001 a março de 2003. Os dados foram extraídos de Urbinatti PR. "Observações ecológicas de *Aedes albopictus* (Diptera:Culicidae) em áreas de proteção ambiental e urbana da periferia na Grande São Paulo. São Paulo, 2004". [Tese de Doutorado, Faculdade de Saúde Pública da Universidade de São Paulo]. (Adaptado).

Ambiente A (n=20)

111	117	170	113	163	212	173	155	114	167
109	158	220	118	129	112	130	128	135	119

Ordenando-se os valores:

109	112	114	118	128	130	155	163	170	212
111	113	117	119	129	135	158	167	173	220

Número mediano de larvas de *Ae.albopictus* =

Valor mediano: (129+130)/2= 129,5 larvas

Ambiente B (n=20)

118	105	159	113	149	72	76	83	92	104
137	84	87	158	138	137	130	112	122	142

Ordenando-se os valores:

72	76	83	84	87	92	104	105	112	113
118	122	130	137	137	138	142	149	158	159

n=20

Número mediano de larvas de Ae.albopictus =

Exercício
Os dados a seguir são provenientes de um estudo que avaliou o tempo médio de vida em dias de 22 machos e 31 fêmeas de *Triatoma sordida*, nos estágios de ninfa e adulto, em condições de laboratório (Souza JMP de, 1978. *Triatoma sórdida* – Considerações sobre o tempo de vida das formas adultas e sobre a oviposição das fêmeas. Revista de Saúde Pública. São Paulo, 12:291-6).

Utilizou-se neste exemplo apenas os dados de tempo de vida em estágio de ninfa.

Calcule o número mediano de dias no estágio de ninfa para machos:

Machos

136	157	154	129	247	164	133	126	247	139
139	148	221	248	131	139	135	143	249	173
241	241								

Ordenando-se os valores

126	133	139	143	157	21	247	249
129	135	139	148	164	241	247	
131	136	139	154	173	241	248	

Valor mediano=

Número mediano de dias no estágio de ninfa para fêmeas:

Fêmeas

126	126	127	130	129	128	131	126	132	136
146	128	150	136	158	134	126	128	128	139
203	208	242	241	250	244	259	241	253	234
250									

Valor mediano=

Ordenando-se os valores

126	127	128	132	139	203	241	250
126	128	129	134	146	208	242	253
126	128	130	136	150	234	244	259
126	128	131	136	158	241	250	

n=31

Valor mediano=

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

Valores mínimo e máximo: valores extremos da distribuição.

Ambiente A

109	112	114	118	128	130	155	163	170	212
111	113	117	119	129	135	158	167	173	220

Valor mínimo = 109 larvas; valor máximo = 220 larvas

Ambiente B

118	105	159	113	149	72	76	83	92	104
137	84	87	158	138	137	130	112	122	142

Ordenando-se os valores:

72	76	83	84	87	92	104	105	112	113
118	122	130	137	137	138	142	149	158	159

Valor mínimo = 72 larvas; valor máximo = 159 larvas

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

Ambiente A
valor máximo - valor mínimo = 220 - 109= 111 larvas
Ambiente B
valor máximo - mínimo = 159 - 72 = 87 larvas

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 media são denominados desvios em relação à média. Se forem elevados ao quadrado, são denominados desvios quadráticos. Então a variância também pode ser entendida como a média dos dos desvios quadraticos de cada observação em relação à média aritmética.
Considerar os valores

3	2	5	6	4

$\bar{x} = 4$ ovos

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

2	3	4	5	6

Variância = $\frac{10}{5} = 2$ ovos^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 media

Desvio padrão= $\sqrt{2} = 1{,}4$ ovos

O erro médio que se comete ao resumir os dados pela média é de 1,4 ovos.

Apresentando as fórmulas:

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

Se os dados forem referentes á 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 $S = \sqrt{S^2}$

Notação, resumo:

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

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.

Exercício

Calcule as medidas de dispersão da variável "número de doenças crônicas" para cada um dos sexos

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)

Variância (n-1)

Desvio padrão (n)

Desvio padrão (n-1)

Coeficiente de variação de Pearson

Sexo

Feminino	1	4	4	0	2	1	2	3	2	1	3	1	2	3	3	2	3
	1	3	3	1	3	2	3	1	3	1	0	2	2	1	2	4	

Feminino (X)	$(x-\bar{x})$	$(x-\bar{x})^2$
1		
4		
4		
0		
2		
1		
2		
3		
2		
1		
3		
1		
2		
3		
3		
2		
3		
1		
3		
3		
1		
3		
2		
3		
1		
3		
1		
0		
2		
2		
1		
2		
4		

Valor mínimo

Valor máximo

Variância (n)

Variância (n-1)

Desvio padrão (n)

Desvio padrão (n-1)

Coeficiente de variação de Pearson

Apresentação das medidas-resumo

A tabela abaixo foi extraída do artigo: Diagnóstico de sobrepeso em adolescentes: estudo do desempenho de diferentes critérios para o Índice de Massa Corporal de MONTEIRO POA *et al*. (*Rev. Saúde Pública* , 2000;.34(5):506 - 13).

Discuta os resultados obtidos ignorando a coluna do valor de p.

Tabela 1 – Estatística descritiva da população em estudo, por sexo (n=493). Pelotas, RS, Brasil. 1998.

Variável	Meninos (n=242) Média	DP	Meninas (n=251) Média	DP	p - valor
Idade (anos)	16,1	0,2	16,1	0,2	0,6
Peso (kg)	65,2	12,3	57,5	10,5	<0,001
Altura (cm)	170,6	6,6	159,8	6,2	<0,001
IMC (kg/m)	22,1	3,7	22,1	3,5	0,8
Dobra subescapular (mm)*	19,9	7,5	23,7	6,3	<0,001
Dobra tricipital (mm)*	19,6	6,3	26,3	5,4	<0,001

*As dobras cutâneas foram medidas apenas nos 92 meninos e 96 meninas cujo IMC foi igual ou superior ao percentil 85 para idade e sexo conforme Nhanes I (OMS).[8]

Tabela 3 – Medidas de tendência central, de dispersão e intervalos de confiança do consumo alimentar dos escolares estimados pelos DA. Escola de Aplicação da USP, São Paulo, 2009.

Estatística	Energia (Kcal)	Carboidrato (g)	Proteína (g)	Lipídios (g)
Média	1730,7	238,6	64,1	59,1
Mediana	1702,0	236,8	61,1	56,4
Desvio padrão	493,2	71,0	21,1	20,8
Valor mínimo	480,0	97,8	12,7	4,6
Valor máximo	3711,3	465,8	157,2	139,4
Q1; Q3	1408,6; 1947,3	179,0; 271,9	51,9; 72,5	46,7; 67,4
IC 95%	(1624,3 -1837,1)	(223,3 - 253,9)	(59,6 – 68,7)	(54,6 – 63,6)

(n=85)

HINNIG PF. Construção de um Questionário de Frequência Alimentar Quantitativo para crianças de 7 a 10 anos [dissertação de mestrado]. São Paulo: Faculdade de Saúde Pública da USP; 2010.

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.

Entre os recém-nascidos que sobreviveram:

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

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

Entre os recém-nascidos que foram a óbito

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

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

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:

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

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.

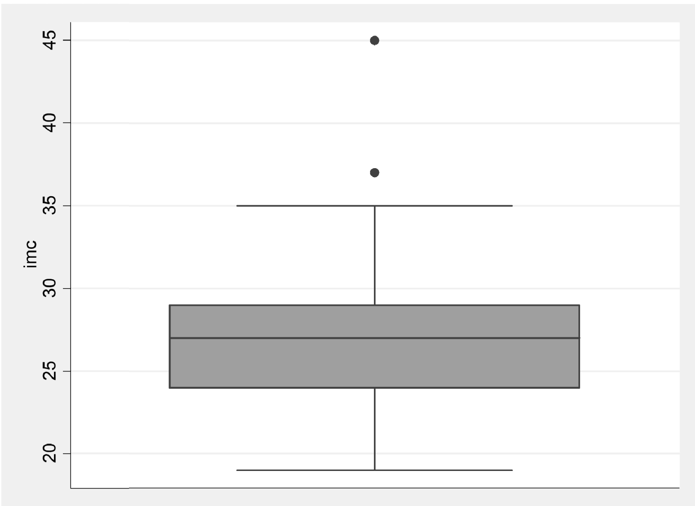
O box plot permite também investigar a dispersão e simetria dos dados.

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.



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	
19	19	
21	19	
29	20	
38	20	
4	22	
34	22	
13	23	
20	23	
37	23	
48	23	
3	24	
11	24	
15	24	
26	24	
33	24	
28	25	
44	25	
1	26	
7	26	
10	26	
24	26	
5	27	

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

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

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	
33	38	
34	46	
20	47	
48	56	
5	61	
39	66	
41	72	
12	73	
24	75	
3	79	
11	80	
49	80	
22	86	
42	87	
26	89	
29	91	
19	92	
27	92	
38	99	

14	101	
50	104	
10	111	
47	118	
23	119	
9	124	
45	125	
17	126	
1	128	
40	130	
25	145	
36	153	
8	157	
32	165	
2	166	
4	166	
16	170	
30	171	
31	176	
28	181	
18	193	
13	205	
7	211	
43	219	
21	221	
46	233	
6		
15		
35		
37		
44		

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

Exercício

Os dados a seguir são adaptados de artigo publicado por Honório NA & Lourenço-de-Oliveira R. 2001, cujo estudo avaliou a frequência mensal de larvas e pupas de *Aedes aegytpi* e *Aedes albopictus* coletadas em pneus, no período de novembro de 1997 a outubro de 1998, em Nova Iguaçu, Rio de Janeiro.

Número de larvas – *Ae.albopictus*

123	243	215	142	153	118	164	194	160	120
128	122	151	155	137	129	216	157	145	182

Número de larvas – Ae.aegypti

90	104	140	72	67	78	60	61	101	81
117	89	111	83	98	101	74	88	132	66

a) Calcule a o número médio de larvas em cada grupo utilizando a média aritmética
b) Calcule a o número médio de larvas em cada grupo utilizando a média geométrica
c) Calcule o número mediano de larvas em cada grupo.
d) Desenhe o *box plot* do número de larvas representando os dois grupos em um só gráfico.
e) Comente o gráfico *box plot* quanto a dispersão dos dados, existência de valores aberrantes e simetria dos dados.

Atenção: é necessário ordenar os valores para fazer os itens c e d.

Exercício 11

Os dados a seguir são adaptados de estudo, publicado por Devicari et al. 2013, que avaliou o tamanho das asas em (mm) de *Aedes scapularis* para machos e fêmeas da espécie, capturados no Município de São Paulo:

Machos

1,78	1,91	2,02	2,11	2,13	2,21	2,30	2,41	1,87	2,01
2,03	2,11	2,15	2,21	2,31	3,50	1,90	2,01	2,10	2,11
2,15	2,21	2,32							

Fêmeas

1,01	1,62	2,30	2,40	2,56	2,61	2,71	2,80	1,52	1,89
2,31	2,45	2,60	2,65	2,75	2,89	1,58	1,97	2,34	2,45
2,60	2,70	2,78							

a) Calcule o tamanho médio da asa (mm) para cada sexo. Utilize a média aritmética;
b) Calcule o tamanho mediano da asa (mm) para cada sexo;
c) Calcule a variância, o desvio-padrão e o coeficiente de variação de Pearson do tamanho da asa em (mm) para cada sexo.
d) Machos e fêmeas são parecidos quanto ao tamanho da asa (mm)?
e) E quanto à variabilidade?
f) Apresente o box plot do tamanho da asa (mm) para os sexos e interprete o gráfico.

Exercício

A tabela abaixo foi extraída do artigo: Influência da Altitude, Latitude e Estação de Coleta (Regra de Bergmann) na dimensão de *Lutzomyia intermedia* (Lutz & Neiva, 1912) (Diptera:Psychodidae, Phlebotominae). Marcondes CB et al. (Memórias do Instituto Oswaldo Cruz, 1999;. vol94(5):693-700).

Discuta os resultados obtidos.

Dimensions (in µm) of females and some of their respective ratios for *Lutzomyia intermedia* from Viana, a low altitude and lower latitude locality in the State of Espírito Santo (ES), and from low altitude and higher latitude localities in the states of Rio de Janeiro and São Paulo

	Viana (ES)				Rio de Janeiro and São Paulo			
Structures and ratios	Mean	s	N	C. V.	Mean	s	N	C. V.
Width of head[b]	342.6	12.2	17	3.6	374.1	21.3	61	5.7
Length of eye[b]	200.4	11	17	5.5	222.1	14.2	56	6.4
Width of eye[b]	110.6	8.8	17	7.9	126.7	8.3	55	6.6
Length of palpomere 3[a]	163.5	6.6	15	4	171.6	7.1	60	7.1
Length of palpomere 5[b]	135.3	11.5	14	8.5	145.9	12.3	58	8.5
Total length of palpus[b]	552	21.5	14	3.9	580	29.3	58	5
Maximum width of wing[b]	576	29.9	13	5.2	620	36.4	64	5.9
Length wing/maximum width of wing[a]	3.53	0.201	13	5.7	3.35	0.225	61	6.7
Length of R_2[a]	547	30.4	13	5.5	585	58.9	65	9.7
δ[a]	278	34.2	12	12.3	317	58.6	65	18.5
Length of R_3 [a]	683	30.1	13	4.41	723	60.2	65	8.33
Length of anterior femur[a]	692	33.8	11	4.9	722.5	45.5	50	6.3
Maximum width of spermathecal head[a]	10.49	1.8	18	17.5	11.87	2.13	59	17.9

s: standard deviation; N: number of observations; C.V.: coefficient of variation; *a*: significant at 5%; *b*: significant at 1%; δ: distance between the distal extremity of R_1 and the fork of R_{2+3}.

Exercício

A tabela abaixo foi extraída do artigo: Influência da Altitude, Latitude e Estação de Coleta (Regra de Bergmann) na dimensão de *Lutzomyia intermedia* (Lutz & Neiva, 1912) (Diptera:Psychodidae, Phlebotominae). Marcondes CB et al. (Memórias do Instituto Oswaldo Cruz, 1999;. vol94(5):693-700).

Discuta os resultados obtidos.

Tabela - Distribuição das médias, desvios padrão e variâncias das absorbâncias de amostras de **Lutzomyia longipalpis** ***coletadas no campo, alimentadas em laboratório e de*** **Lutzomyia almerioi** ***procedentes de campo no período de 2002 a 2004.***

		Lutzomyia longipalpis			*Lutzomyia almerioi*		
Amostras/Ano	Local	X	S	V	X	S	V
2002	campo	0,518	0,247	0,061	0,781	0,167	0,028
	laboratório	0,815	0,030	0,001	-	-	-
2003	campo	0,652	0,148	0,022	0,764	0,031	0,001
2004	campo	0,668	0,197	0,039	-	-	-

X = média; S = desvio-padrão; V = variância