GET00170 Estatística Básica Aplicada às Ciências Humanas II
Lista de Exercícios 3 – 15/11/2017
Profa. Ana Maria Farias

---

1. Em um estudo sobre a pontualidade de empresas aéreas, coletaram-se dados relativos ao tempo total de atraso (em minutos) de duas empresas durante 40 dias, selecionados aleatoriamente, no primeiro semestre de 2007. (Obs.: Dados fictícios)

Empresa 1

105	102	108	111	105	108	99	126	111	102
117	117	108	102	129	130	131	114	114	93
114	102	114	99	126	117	111	108	105	114
120	111	111	129	93	96	111	114	111	105

$\sum_{i=1}^{40} x_i = 4443 \qquad \min = 93 \qquad \max = 131$

Empresa 2

93	105	93	120	108	96	99	117	135	96
123	87	105	120	147	87	93	123	123	129
105	168	90	120	120	90	126	108	102	126
114	96	156	99	111	105	102	96	96	99

$\sum_{i=1}^{40} x_i = 4428 \qquad \min = 87 \qquad \max = 168$

(a) Construa um diagrama de ramos-e-folhas lado a lado para comparar os tempos de atraso das duas empresas.

(b) Calcule a média e a amplitude de cada conjunto de dados, interpretando os resultados obtidos.

(c) Em uma mesma escala, construa os diagramas de caixa para os dados das 2 empresas. Comente.

Solução

(a) Considerando os dados na ordem dada:

Empresa 2 (folha=unidade)		**Empresa 1** (folha=unidade)
7 7	8	
9 6 6 9 6 0 0 3 6 9 6 3 3	9	9 3 9 3 3
2 5 2 8 5 5 8 5	10	5 2 8 5 8 2 8 2 2 8 5 5
1 4 7	11	1 1 7 7 4 4 4 4 7 1 4 1 1 1 4 4
6 6 0 0 9 3 3 0 3 0	12	6 9 6 0 9
5	13	0 1
7	14	
6	15	
8	16	

Ordenando os dados em cada ramo:

Empresa 2 (folha=unidade)		**Empresa 1** (folha=unidade)
7 7	8	
9 9 9 6 6 6 6 6 3 3 3 0 0	9	3 3 3 9 9
8 8 5 5 5 5 2 2	10	2 2 2 2 5 5 5 5 8 8 8 8
7 4 1	11	1 1 1 1 1 1 4 4 4 4 4 4 4 7 7 7
9 6 6 3 3 3 0 0 0 0	12	0 6 6 9 9
5	13	0 1
7	14	
6	15	
8	16	

Ambas empresas têm a grande parte dos tempos de atraso entre 90 e 129 minutos, mas a empresa 2 tem três tempos grandes, o que torna a amplitude dos seus dados maior do que a da empresa 1.

(b) Empresa 1: $\overline{x}_1 = \dfrac{4341}{40} = 111,075$ minutos $\qquad \Delta_1 = 131 - 93 = 38$ minutos

Empresa 2: $\overline{x}_2 = \dfrac{4428}{40} = 110,700$ minutos $\qquad \Delta_2 = 168 - 87 = 81$ minutos

O tempo médio de ambas é bem próximo, mas a amplitude para a Empresa 2 é bem maior, caracterizando uma maior variabilidade.

(c) Como cada conjunto de dados tem 40 observações, a mediana será a média da 20ª e 21ª observações. O primeiro quartil será a média da 10ª e da 11ª observações. O terceiro quartil será a média da 30ª e 31ª observações.

Empresa 1: $Q_1 = \frac{105+105}{2} = 105 \qquad Q_2 = \frac{111+111}{2} = 111 \qquad Q_3 = \frac{114+117}{2} = 115,5$

Empresa 2: $Q_1 = \frac{96+96}{2} = 96 \qquad Q_2 = \frac{105+105}{2} = 105 \qquad Q_3 = \frac{120+123}{2} = 121,5$

Verificação de valores discrepantes

Empresa 1:

$AIQ = Q_3 - Q_1 = 115,5 - 105 = 10,5$

$Q_1 - 1,5 \times AIQ = 105 - 1,5 \times 10,5 = 89,25 \qquad Q_3 + 1,5 \times AIQ = 115,5 + 1,5 \times 10,5 = 131,25$

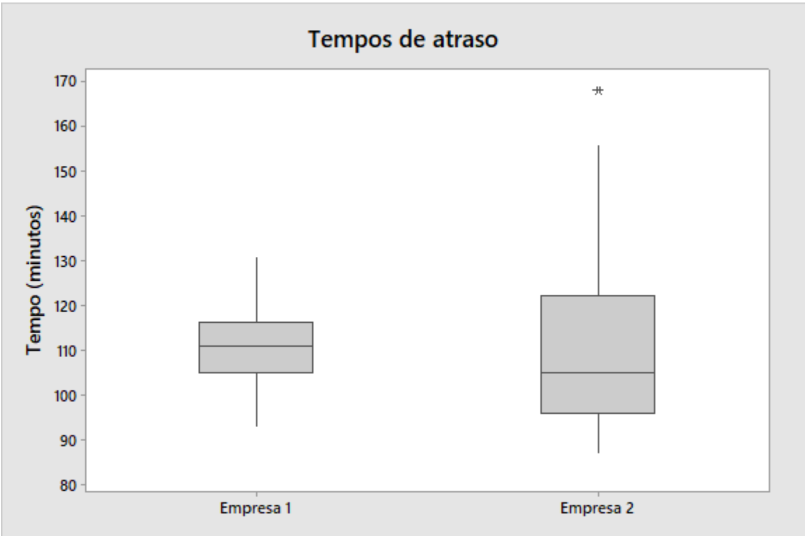
Não há valores discrepantes para a Empresa 1.

Empresa 2:

$AIQ = Q_3 - Q_1 = 121,5 - 96 = 25,5$

$Q_1 - 1,5 \times AIQ = 96 - 1,5 \times 25,5 = 57,75 \qquad Q_3 + 1,5 \times AIQ = 121,5 + 1,5 \times 25,5 = 159,75$

O valor 168 é um valor discrepante para a Empresa 2.



Vemos que na Empresa 2 há muito mais variabilidade; o tempo mediano da empresa 1 é maior, mas os 50% centrais estão mais concentrados.

2. Na tabela a seguir temos a distribuição das notas na prova de redação de um conjunto de candidatos a emprego no departamento de marketing de uma grande empresa.

Nota	Frequência Simples		Frequência Acumulada	
	Absoluta (n_i)	Relativa (f_i)%	Absoluta (N_i)	Relativa (F_i) %
$0 \vdash 2$	25	17,857	25	17,857
$2 \vdash 4$	32	22,857	57	40,714
$4 \vdash 6$	48	34,286	105	75,000
$6 \vdash 8$	27	19,286	132	94,286
$8 \vdash 10$	8	5,714	140	100,000
Total	140	100,000		

Calcule a nota média dos candidatos.

Solução

Os pontos médios das classes são

$x_1 = \frac{0+2}{2} = 1 \qquad x_2 = \frac{2+4}{2} = 3 \qquad x_2 = \frac{4+6}{2} = 5 \qquad x_2 = \frac{6+8}{2} = 7 \qquad x_5 = \frac{8+10}{2} = 9$

$$\begin{aligned}
\overline{x} &= \frac{25 \times 1 + 32 \times 3 + 48 \times 5 + 27 \times 7 + 8 \times 9}{140} \\
&= \frac{25}{140} \times 1 + \frac{32}{140} \times 3 + \frac{48}{140} \times 5 + \frac{27}{140} \times 7 + \frac{8}{140} \times 9 \\
&= 0,17857 \times 1 + 0,22857 \times 3 + 0,34286 \times 5 + 0,19286 \times 7 + 0,05714 \times 9
\end{aligned}$$

Trabalhar com as frequências simples é a maneira mais fácil e que evita erros de arredondamento; veja a primeira linha do cálculo acima. Para realizar esse cálculo de forma organizada, considere a seguinte tabela (note que ignoramos as informações irrelevantes):

Nota	Frequência absoluta n_i	Ponto médio x_i	$n_i \cdot x_i$
$0 \vdash 2$	25	1	$25 \cdot 1 = 25$
$2 \vdash 4$	32	3	$32 \cdot 3 = 96$
$4 \vdash 6$	48	5	$48 \cdot 5 = 240$
$6 \vdash 8$	27	7	$27 \cdot 7 = 189$
$8 \vdash 10$	8	9	$8 \cdot 9 = 72$
Total	**140**		$25 + 96 + 240 + 189 + 72 = \mathbf{622}$

Média: $\qquad \overline{x} = \frac{622}{140} = 4,443$

Essa forma de cálculo é mais simples, pois não trabalha com as frequências relativas, que, em geral, não são números inteiros.

3. A seguir temos informações sobre o número de latas de um refrigerante vendidoas no bar do Hotel Brisa Mar nas últimas 3 semanas. Construa um boxplot para esses dados.

45	45	46	46	46	47	48	48	48	48	48
49	49	49	50	50	51	52	53	54	54	

Solução

$n = 21$

Q_2 é a 11ª; Q_1 é a média da 5ª e da 6ª observações. Q_3 é a média da 16ª e da 17ª observações.

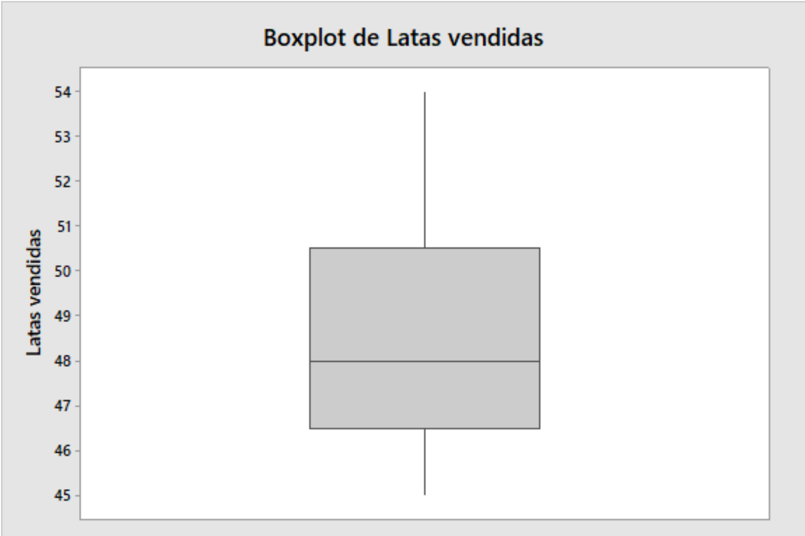
$Q_1 = \frac{46+47}{2} = 46,5 \qquad Q_2 = 48 \qquad Q_3 = \frac{50+51}{2} = 50,5$

Verificação de valores discrepantes

$AIQ = Q_3 - Q_1 = 50,5 - 46,5 = 4,0$

$Q_1 - 1,5AIQ = 46,5 - 1,5 \times 4 = 46,5 - 6 = 40,5 \qquad Q_3 + 1,5 \times AIQ = 50,5 + 1,5 \times 4 = 50,5 + 6 = 56,5$

Não há valores discrepantes.



4. A seguir temos dados que representam a diferença (em minutos) entre o horário efetivo de chegada e o horário previsto de voos de uma companhia aérea (valores negativos indicam que o voo chegou antes do horário previsto). Construa um boxplot para esses dados.

–18	–11	–2	6	7	11	15	16	17	19	19	20	22
22	23	24	25	25	26	28	33	45	48	50	51	

Solução

$n = 25$

Q_2 é a 13ª; Q_1 é a média da 6ª e da 7ª observações. Q_3 é a média da 19ª e da 20ª observações.

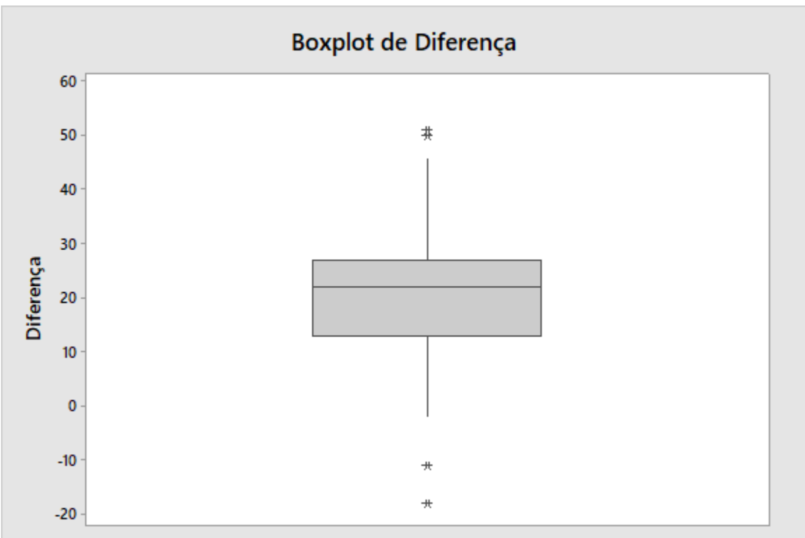
$Q_1 = \dfrac{46 + 4711 + 15}{2} = 13 \qquad Q_2 = 22 \qquad Q_3 = \dfrac{26 + 28}{2} = 27$

Verificação de valores discrepantes

$AIQ = Q_3 - Q_1 = 27 - 13 = 14$

$Q_1 - 1,5AIQ = 13 - 1,5 \times 14 = 13 - 21 = -8 \qquad Q_3 + 1,5 \times AIQ = 27 + 1,5 \times 14 = 27 + 21 = 48$

As duas observações inferiores e as duas observações superiores são valores discrepantes.



5. Calcule a variância dos seguintes dados:

10	12	15	19	21

Solução

A média dos dados é $\overline{x} = \frac{10+12+15+19+21}{5} = \frac{77}{5} = 15,4$

Vamos organizar os cálculos em uma tabela.

	Dado x_i	Desvio $= x_i - \overline{x}$	Desvio$^2 = (x_i - \overline{x})^2$
	10	$10 - 15,4 = -5,4$	$(-5,4)^2 = 29,16$
	12	$12 - 15,4 = -3,4$	$(-3,4)^2 = 11,56$
	15	$15 - 15,4 = -0,4$	$(-0,4)^2 = 0,16$
	19	$19 - 15,4 = 3,6$	$(3,6)^2 = 12,96$
	21	$21 - 15,4 = 5,6$	$(5,6)^2 = 31,36$
Soma	77	$-9,2 + 9,2 = 0$	$29,16 + 11,56 + 0,16 + 12,96 + 31,36 = 85,2$

A variância é a média dos desvios quadráticos:

$$\sigma^2 = \frac{85,2}{5} = 17,04$$

Uma forma mais fácil ainda de calcular a variância é usar o resultado de que a variância é a *média dos quadrados menos o quadrado da média*. Para os dados acima, a média dos quadrados é

$$\frac{10^2 + 12^2 + 15^2 + 19^2 + 21^2}{5} = \frac{100 + 144 + 225 + 361 + 441}{5} = \frac{1271}{5} = 254,2$$

e a variância é

$$\sigma^2 = 254,2 - (15,4)^2 = 17,04$$

6. Calcule a variância dos seguintes dados:

3	9	8	8	6	11	8	7

Solução

A média dos dados é

$$\overline{x} = \frac{3+9+8+8+6+11+8+7}{8} = \frac{60}{8} = 7,5$$

A média dos quadrados dos dados é

$$\overline{x} = \frac{3^2 + 9^2 + 8^2 + 8^2 + 6^2 + 11^2 + 8^2 + 7^2}{8} = \frac{9 + 81 + 64 + 64 + 36 + 121 + 64 + 49}{8} = \frac{488}{8} = 61$$

e a variância é

$$\sigma^2 = 61 - 7,5^2 = 4,75$$

7. A seguir temos a distribuição da receita (em unidades monetárias) de uma amostra de 50 hotéis do estado do Rio de Janeiro (dados fictícios). Calcule a média desses dados, interpretando o resultado.

Receita			Número de hotéis
1	⊢	5	12
5	⊢	10	23
10	⊢	20	8
20	⊢	50	5
50	⊢	100	2

Solução

Vamos completar a tabela:

Receita			Número de hotéis n_i	Ponto médio x_i	$n_i \cdot x_i$
1	⊢	5	12	3	$12 \cdot 3 = 36$
5	⊢	10	23	7,5	$23 \cdot 7,5 = 172,5$
10	⊢	20	8	15	$8 \cdot 15 = 120$
20	⊢	50	5	35	$5 \cdot 35 = 175$
50	⊢	100	2	75	$2 \cdot 75 = 150$
Soma			50		$653,50$

A média é

$$\overline{x} = \frac{653,50}{50} = 13,07$$

A receita média dos 50 hotéis é de 13,07 unidades monetárias.

8. A seguir temos a taxa de pulsação (em batimentos por minuto, bpm) de uma amostra de 40 mulheres.

46	50	52	54	56	56	58	58	60	60
60	60	62	62	64	64	64	66	66	66
68	68	68	68	68	70	70	70	72	74
74	74	76	78	80	80	84	86	88	93

(a) Construa um boxplot para esses dados.

(b) Sabendo que a soma dos dados é 2693, calcule a média dos dados, interpretando seu resultado.

(c) Construa uma tabela de frequência completa para esses dados, com 5 classes de mesmo tamanho e tomando 45 como limite inferior da primeira classe e 95 como o limite superior da última classe.

(d) Calcule a média dos dados a partir da tabela de frequências construída no item anterior. Compare com o resultado obtido anteriormente no item (8b). Explique o motivo da diferença. Quando você deve usar cada um desses resultados?

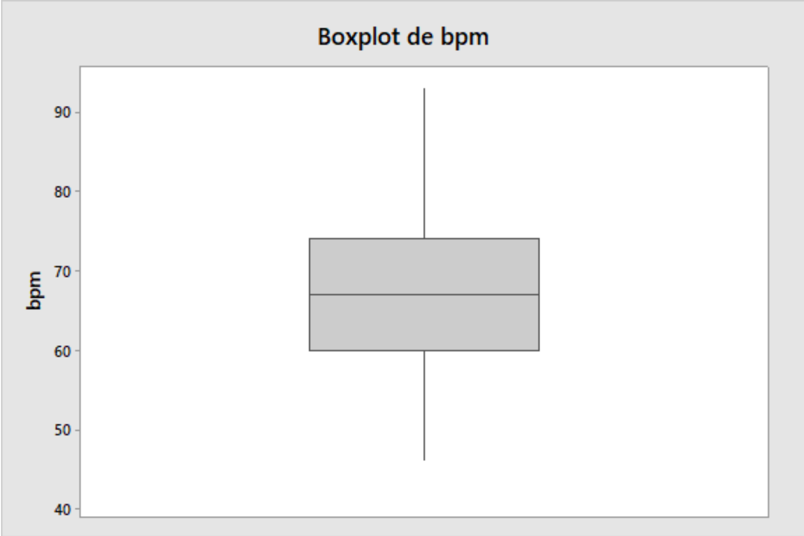
Solução

(a)

$$Q_1 = \frac{60+60}{2} = 60 \qquad Q_2 = \frac{66+68}{2} = 67 \qquad Q_3 = \frac{74+74}{2} = 74$$

$$AIQ = Q_3 - Q_1 = 74 - 60 = 14 \qquad Q_1 - 1,5AIQ = 60 - 1,5 \times 14 = 39 \qquad Q_3 + 1,5AIQ = 74 + 1,5 \times 14 = 95$$

Não há valores discrepantes.



(b) $\overline{x} = \dfrac{2693}{40} = 67,325$ bpm.
A taxa média de pulsação dessas 40 mulheres é de 67,325 bpm.

(c) Dados o limite inferior da primeira classe e o limite superior da última classe, temos que a amplitude dos dados na tabela de frequência é de $95 - 45 = 50$ e o comprimento de cada classe tem de ser $50/5 = 10$.

BPM	Frequência Simples		Frequência Acumulada	
	Absoluta (n_i)	Relativa (f_i)%	Absoluta (N_i)	Relativa (F_i) %
$45 \vdash 55$	4	10,0	4	10,0
$55 \vdash 65$	13	32,5	17	42,5
$65 \vdash 75$	15	37,5	32	80,0
$75 \vdash 85$	5	12,5	37	92,5
$85 \vdash 95$	3	7,5	40	100,0
Total	40	100,0		

(d) Os pontos médios das classes são 50, 60, 70, 80, 90 e a média é

$$\overline{x} = \frac{4 \times 50 + 13 \times 60 + 15 \times 70 + 5 \times 80 + 3 \times 90}{40} = \frac{200 + 780 + 1050 + 400 + 270}{40} = \frac{2700}{40} = 67,5$$

A diferença entre os valores (67,325 e 67,5) se deve ao fato de termos substituído cada valor pelo ponto médio da respectiva classe. Se os dados brutos estiverem disponíveis, sempre devemos calcular quaisquer medidas a partir deles. Medidas estatísticas a partir de dados agrupados só devem ser usadas quando os dados brutos não estiverem disponíveis e devem sempre ser usadas com bastante cautela.