Usando Distribuição de Frequências Percentis e Boxplot

Gilberto Pereira Sassi

Universidade Federal Fluminense
Instituto de Matemática e Estatística
Departamento de Estatística

25 de maio de 2016

Medidas de Posição

Valor Representativo ou valor típico

Média	Percentil
Moda	1 Quartil
Mediana	3 Quartil

Medidas de Dispersão

Avalia como os dados estão dispersos.

Variância	Amplitude
Desvio Padrão	Intervalo Interquartil
Desvio Médio	

Exemplo

Considere as notas finais de três turmas de estatística na Tabela 1.

Tabela 1: Notas das turmas de estatística.

Turma	Notas							
A	4	5	5	6	6	7	7	8
B	1	2	4	6	6	9	10	10
C	0	6	7	7	7	7	8	

Tabelas de distribuição para as três turmas.

Tabela 2: Turma A

Notas	Frequência	Proporção	Porcentagem
4	1	0,125	12,5
5	2	0,25	25
6	2	0,25	25
7	2	0,25	25
8	1	0,125	12,5
total	8	1	100

Tabela 3: Turma B

Notas	Frequência	Proporção	Porcentagem
1	1	0,125	12,5
2	1	0,125	12,5
3	0	0	0
4	1	0,125	12,5
5	0	0	0
6	2	0,25	25
7	0	0	0
8	0	0	0
9	1	0,125	12,5
10	2	025	25
total	8	1	100

Tabela 4: Turma C

Notas	Frequência	Proporção	Porcentagem
0	1	0,1429	14,29
1	0	0,0000	0,00
2	0	0,0000	0,00
3	0	0,0000	0,00
4	0	0,0000	0,00
5	0	0,0000	0,00
6	1	0,1429	14,29
7	4	0,5714	57,14
8	1	0,1429	14,29
total	7	1	100

Média e Variância usando distribuição de frequência

Turma A

$$\bar{x} = \frac{4 + 2 \cdot 5 + 2 \cdot 6 + 2 \cdot 7 + 8}{8} = 6$$
$$s^2 = \frac{(4-6)^2 + 2 \cdot (5-6)^2 + 2 \cdot (6-6)^2 + 2 \cdot (7-6)^2 + (8-6)^2}{8-1} \simeq 1,71$$

Turma B

$$\bar{x} = \frac{1 + 2 + 4 + 2 \cdot 6 + 9 + 2 \cdot 10}{8} = 6$$
$$s^2 = \frac{(6-1)^2 + (2-6)^2 + (4-6)^2 + 2 \cdot (6-6)^2 + (9-6)^2 + 2 \cdot (10-6)^2}{8-1}$$
$$= 12,29$$

Turma C

$$\bar{x} = \frac{0 + 6 + 4 \cdot 7 + 8}{7} = 6$$
$$s^2 = \frac{(0-6)^2 + (6-6)^2 + 4 \cdot (7-6)^2 + (8-6)^2}{6} = 7,33$$

Exemplo

Considere os dados de taxa de mortalidade infantil em 34 cidades do oeste catarinense na Tabela 5.

Tabela 5: Taxa de mortalidade.

9,9	10,3	11,9	13,1	13,9	15,7	17	18
18,3	18,4	19,6	20	20,3	21,7	22	22,6
22,7	23,5	23,7	23,8	25,4	27,2	27,3	28,9
29,7	29,9	32,3	32,9	33	36,3	36,4	38,3
39,2	62,2						

Distribuição de Frequência

Tabela 6: Distribuição de Frequência para a taxa de mortalidade infantil nas cidades do oeste catarinense.

Taxa de Mortalidade Infantil	Frequência	Proporção	Porcentagem
0 ⊢ 10	1	0.0294	2.94
10 ⊢ 20	10	0.2941	29.41
20 ⊢ 30	15	0.4412	44.12
30 ⊢ 40	7	0.2059	20.59
40 ⊢ 50	0	0.0000	0.00
50 ⊢ 60	0	0.0000	0.00
60 ⊢ 70	1	0.0294	2.94
Total	34	1.0000	100.00

Exemplo

Ao usarmos a tabela de distribuição de frequência, não sabemos o valor exato de cada elemento da amostra. A solução é atribuir o ponto do médio da classe que elemento conforme Tabela 7.

Tabela 7: Ponto médio de cada classe com a respectiva frequência.

Taxa de Mortalidade Infantil	**Frequência**	**ponto médio**
0 ⊢ 10	0	5
10 ⊢ 20	1	15
20 ⊢ 30	0	25
30 ⊢ 40	0	35
40 ⊢ 50	0	45
50 ⊢ 60	0	55
60 ⊢ 70	0	65
Total	1	–

Exemplo

Observe que ao usarmos a distribuição estamos calculando uma aproximação para a média, variância e desvio médio.

Aproximação Usando Distribuição de Frequência

$$\bar{x} = \frac{5 \cdot 1 + 15 \cdot 10 + 25 \cdot 15 + 35 \cdot 7 + 55}{34} = 24.71$$

$$s^2 = \frac{(5-24,71)^2 + 10 \cdot (15-24,71)^2 + 15 \cdot (25-24,71)^2 + 7 \cdot (35-24,71)^2 + (55-24,71)^2}{33} = 112,03$$

$$dm = \frac{|5-24,71| + 10 \cdot |15-24,71| + 15 \cdot |25-24,71| + 7 \cdot |35-24,71| + |55-24,71|}{34} = 6,87$$

Aproximação Usando Distribuição de Frequência

$$\bar{x} = 24,86$$
$$s^2 = 107,63$$
$$dm = 7,70$$

Percentil, Quartil e Boxplot

O percentil de ordem $p \cdot 100$ $(0 < p < 1)$ é o valor que ocupa a posição $p \cdot (n+1)$ dos dados ordenados.

Notação:

- $q(p)$ é o percentil de ordem $p \cdot 100$;
- se $p = 0,25$, chamamos o percentil de primeiro quartil e usamos a notação matemática Q_1;
- se $p = 0,25$, chamamos o percentil de segundo quartil ou mediana e usamos a notação matemática Q_2 ou md
- se $p = 0,75$, chamamos o percentil de terceiro quartil e usamos a notação matemática Q_3

Exemplo

Notas finais em três turmas de estatística.

Tabela 8: Notas finais para as turmas A, B e C.

Turma A	Turma B	Turma C
6	6	0
8	10	7
5	2	7
4	1	8
6	6	7
5	4	7
7	9	6
7	10	

Exemplo – Turma A

Primeiro ordenamos os dados:

$x_{(1)}$	$x_{(2)}$	$x_{(3)}$	$x_{(4)}$	$x_{(5)}$	$x_{(6)}$	$x_{(7)}$	$x_{(8)}$
4	5	5	6	6	7	7	8

1° Quartil Posição: $0,25 \cdot (8+1) = 2,25$ em que 8 é tamanho da amostra. Como $2,25$ não é um número inteiro, usamos o valor na posição 2 e 3 e tiramos uma média:

$$Q_1 = \frac{x_{(2)} + x_{(3)}}{2} = \frac{5+5}{2} = 5$$

2° Quartil Posição: $0,5 \cdot (8+1) = 4,5$ em que 8 é tamanho da amostra. Como $4,5$ não é um número inteiro, usamos o valor na posição 4 e 5 tirando uma média:

$$Q_2 = \frac{x_{(4)} + x_{(5)}}{2} = \frac{6+6}{2} = 6$$

3° Quartil Posição: $0,75 \cdot (8+1) = 6,75$ em que 8 é tamanho da amostra. Como $6,75$ não é um número inteiro, usamos o valor na posição 6 e 7 tirando uma média:

$$Q_2 = \frac{x_{(6)} + x_{(7)}}{2} = \frac{7+7}{2} = 7$$

Exemplo – Turma B

Primeiro ordenamos os dados:

$x_{(1)}$	$x_{(2)}$	$x_{(3)}$	$x_{(4)}$	$x_{(5)}$	$x_{(6)}$	$x_{(7)}$	$x_{(8)}$
1	2	4	6	6	9	10	10

1° Quartil Posição: $0,25 \cdot (8+1) = 2,25$ em que 8 é tamanho da amostra. Como $2,25$ não é um número inteiro, usamos o valor na posição 2 e 3 e tiramos uma média:

$$Q_1 = \frac{x_{(2)} + x_{(3)}}{2} = \frac{2+4}{2} = 3$$

2° Quartil Posição: $0,5 \cdot (8+1) = 4,5$ em que 8 é tamanho da amostra. Como $4,5$ não é um número inteiro, usamos o valor na posição 4 e 5 tirando uma média:

$$Q_2 = \frac{x_{(4)} + x_{(5)}}{2} = \frac{6+6}{2} = 6$$

3° Quartil Posição: $0,75 \cdot (8+1) = 6,75$ em que 8 é tamanho da amostra. Como $6,75$ não é um número inteiro, usamos o valor na posição 6 e 7 tirando uma média:

$$Q_3 = \frac{x_{(6)} + x_{(7)}}{2} = \frac{9+10}{2} = 9,5$$

Exemplo – Turma C

Primeiro ordenamos os dados:

$x_{(1)}$	$x_{(2)}$	$x_{(3)}$	$x_{(4)}$	$x_{(5)}$	$x_{(6)}$	$x_{(7)}$
0	6	7	7	7	7	8

1° Quartil Posição: $0,25 \cdot (7+1) = 2$ em que 7 é tamanho da amostra. O primeiro quartil é o valor na posição 2:

$$Q_1 = x_{(2)} = 6$$

2° Quartil Posição: $0,5 \cdot (7+1) = 4$ em que 7 é tamanho da amostra. O primeiro quartil é o valor na posição 4:

$$Q_2 = x_{(4)} = 7$$

3° Quartil Posição: $0,75 \cdot (7+1) = 6$ em que 7 é tamanho da amostra. O primeiro quartil é o valor na posição 6:

$$Q_3 = x_{(6)} = 7$$

Exemplo

Observe que o primeiro, segundo o terceiro quartis podem ser usados para analisar a variabilidade e assimetria dos dados. A diferença $Q_3 - Q_1$ avalia como os dados se distribuem e é denominado de *intervalo interquartil*.

Figura 1: Turma A

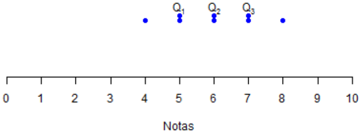


Figura 2: Turma B

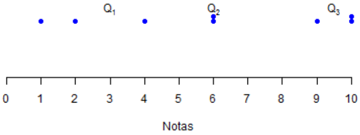
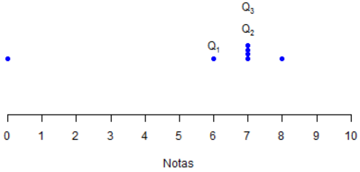


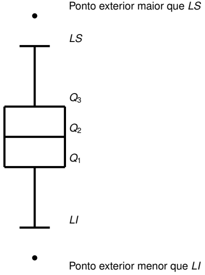
Figura 3: Turma C



Boxplot

O primeiro, o segundo e o terceiro quartil podem ser apresentados visualmente usando o gráfico um diagrama chamado *boxplot*.

Figura 4: Boxplot



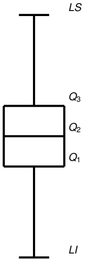
em que $LS = Q_3 + 1,5 \cdot (Q_3 - Q_1)$, $LI = Q_1 - 1,5 \cdot (Q_3 - Q_1)$.
Se um valor está entre LI e LS, este é chamado de valor Adjacente. Por outro, se um valor é maior que LS ou menor LI, este é chamado de ponto exterior e temo indícios de que o valor é atípico.

Exemplo

Vamos fazer o boxplot para as notas finais da A de estatística apresentado anteriormente.

- $Q_1 = 5$, $Q_2 = 6$ e $Q_3 = 7$;
- $Q_3 - Q_1 = 2$;
- $LS = Q_3 + 1,5 \cdot (Q_3 - Q_1) = 10$;
- $LI = Q_1 - 1,5 \cdot (Q_3 - Q_1) = 2$;

Figura 5: Boxplot – Turma A.



Exemplo

Considere as notas finais já ordenadas de uma turma de estatística com 12 alunos:

$x_{(1)}$	$x_{(2)}$	$x_{(3)}$	$x_{(4)}$	$x_{(5)}$	$x_{(6)}$	$x_{(7)}$	$x_{(8)}$	$x_{(9)}$	$x_{(10)}$	$x_{(11)}$	$x_{(12)}$
0	3.35	3.41	3.83	3.98	4.08	4.19	4.55	5	5.55	5.84	10

1° Quartil Posição: $0,25 \cdot (12+1) = 3,25$ em que 12 é tamanho da amostra. Como $3,25$ não é um número inteiro, usamos o valor na posição 3 e 4 e tiramos uma média:

$$Q_1 = \frac{x_{(3)} + x_{(4)}}{2} = \frac{3,41 + 3,83}{2} = 3,62$$

2° Quartil Posição: $0,5 \cdot (12+1) = 6,5$ em que 12 é tamanho da amostra. Como $6,5$ não é um número inteiro, usamos o valor na posição 6 e 7 tirando uma média:

$$Q_2 = \frac{x_{(6)} + x_{(7)}}{2} = \frac{4,08 + 4,19}{2} = 4,135$$

3° Quartil Posição: $0,75 \cdot (12+1) = 9,75$ em que 12 é tamanho da amostra. Como $9,75$ não é um número inteiro, usamos o valor na posição 9 e 10 tirando uma média:

$$Q_3 = \frac{x_{(9)} + x_{(10)}}{2} = \frac{5 + 5,55}{2} = 5,275$$

Exemplo – continuação

- $Q_1 = 3,62$, $Q_2 = 4,135$ e $Q_3 = 5,275$;
- $LS = Q_3 + 1,5 \cdot (Q_3 - Q_1) = 7,7575$;
- $LI = Q_1 - 1,5 \cdot (Q_3 - Q_1) = 1,1375$.

Figura 6: Boxplot de notas finais.