Medias de Posição e Dispersão: Percentis, Quartis e Boxplot

Gilberto Pereira Sassi

Universidade Federal Fluminense
Instituto de Matemática e Estatística
Departamento de Estatística

24 de maio de 2016

Medidas Descritivas – Variável Quantitativa

Medidas de Posição Valor representativo e/ou valor típico

Tabela 1: Medidas de Posição.

Média	Percentil
Moda	1° Quartil
Mediana	3° Quartil

Medidas de Dispersão Avalia como os dados se distribuem.

Tabela 2: Medidas de Dispersão.

Variância	Amplitude
Desvio Padrão	Intervalo Interquartil
Desvio Médio	

Medidas de Posição

Média $\bar{x} = \dfrac{x_1 + x_2 + \cdots + x_n}{n}$

Moda Valor mais frequente de uma variável quantitativa discreta

Mediana Posição central dos dados ordenados $x_{(1)} \leq \cdots \leq x_{(n)}$:

$$md(x) = \begin{cases} x_{(\frac{n}{n})}, & \text{se } n \text{ é ímpar} \\ \dfrac{x_{(\frac{n}{n})} + x_{(\frac{n}{n}+1)}}{2}, & \text{se } n \text{ é par} \end{cases}.$$

Medidas de Dispersão

Variância $s^2 = \frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+\cdots+(x_n-\bar{x})^2}{n-1} = \frac{\sum_{i=1}^{n} x_i^2 - n\cdot\bar{x}^2}{n-1}$

Desvio Padrão $s = \sqrt{s^2}$

Desvio Médio $dm = \frac{|x_1-\bar{x}|+\cdots|x_n-\bar{x}|}{n}$

Amplitude $A = \max - \min$, em que max é o maior valor e min é o menor valor presente na amostra

Exemplo

Considere as notas de uma turma de estatística com 10 aluno.

4.47	5.81	9.69	6.91	7.11
9.78	8.08	3.95	9.42	7.92

Média $\bar{x} = \frac{4,47+5,81+9,69+7,11+9,78+8,08+3,95+9,42+7,92}{10} = 7,314$

Mediana Dados Ordenados:

$x_{(1)}$	$x_{(2)}$	$x_{(3)}$	$x_{(4)}$	$x_{(5)}$	$x_{(6)}$	$x_{(7)}$	$x_{(8)}$	$x_{(9)}$	$x_{(10)}$
4.47	5.81	9.69	6.91	7.11	9.78	8.08	3.95	9.42	7.92

Então, $md(x) = \frac{x_{(\frac{n}{2})}+x_{(\frac{n}{2}+1)}}{2} = \frac{7,11+7,92}{2} = 7,515$

Variância $\sum_{i=1}^{10} x_i^2 = 573,93$, $\bar{x}^2 = 53,49$ e $s^2 = \frac{573,93-10\cdot 53,49}{9} = 4,33$

Desvio Padrão $s = \sqrt{s^2} = \sqrt{4,33} = 2,08$

Desvio Médio $dm = \frac{|4,47-7,314|+\cdots+|7,92-7,314|}{10} = 1,664$

Percentil

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

Notação: $q(p)$ é o percentil de ordem $p \cdot 100$.

- $q(0,25)$: 1° Quartil com notação Q_1
- $q(0,5)$: 2° Quartil ou Mediana com notação Q_2 ou $md(x)$
- $q(0,75)$: 3° Quartil com notação Q_3

Exemplo

Considere as notas de estatística para três turmas na Tabela 3.

Tabela 3: Notas para as três turmas de Estatística.

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

Na Tabela 4 ordenamos os dados presentes na Tabela 3.

Tabela 4: Notas em ordem crescente para as três turmas de Estatística.

Turma	Notas							
A	$x_{(1)}$ 4	$x_{(2)}$ 5	$x_{(3)}$ 5	$x_{(4)}$ 6	$x_{(5)}$ 6	$x_{(6)}$ 7	$x_{(7)}$ 7	$x_{(8)}$ 8
B	$x_{(1)}$ 1	$x_{(2)}$ 2	$x_{(3)}$ 4	$x_{(4)}$ 6	$x_{(5)}$ 6	$x_{(6)}$ 9	$x_{(7)}$ 10	$x_{(8)}$ 10
C	$x_{(1)}$ 0	$x_{(2)}$ 6	$x_{(3)}$ 7	$x_{(4)}$ 7	$x_{(5)}$ 7	$x_{(6)}$ 7	$x_{(7)}$ 8	

Exemplo

Turma A 1° **Quartil** posição $0,25 \cdot (8+1) = 2,25$ e $Q_1 = \frac{x_{(2)}+x_{(3)}}{2} = \frac{5+5}{2} = 5$

2° **Quartil** posição $0,5 \cdot (8+1) = 4,5$ e $Q_2 = \frac{x_{(4)}+x_{(5)}}{2} = \frac{6+6}{2} = 6$

3° **Quartil** posição $0,75 \cdot (8+1) = 6,75$ e $Q_3 = \frac{x_{(6)}+x_{(7)}}{2} = \frac{7+7}{2} = 6$

Turma B 1° **Quartil** posição $0,25 \cdot (8+1) = 2,25$ e $Q_1 = \frac{x_{(2)}+x_{(3)}}{2} = \frac{2+4}{2} = 3$

2° **Quartil** posição $0,5 \cdot (8+1) = 4,5$ e $Q_2 = \frac{x_{(4)}+x_{(5)}}{2} = \frac{6+6}{2} = 6$

3° **Quartil** posição $0,75 \cdot (8+1) = 6,75$ e
$Q_3 = \frac{x_{(6)}+x_{(7)}}{2} = \frac{9+10}{2} = 9,5$

Turma C 1° **Quartil** posição $0,25 \cdot (7+1) = 2$ e $Q_1 = x_{(2)} = 6$

2° **Quartil** posição $0,5 \cdot (7+1) = 4$ e $Q_2 = x_{(4)} = 7$

3° **Quartil** posição $0,75 \cdot (7+1) = 6$ e $Q_3 = x_{(6)} = 7$

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

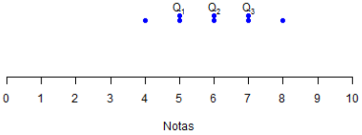


Figura 2: Turma B

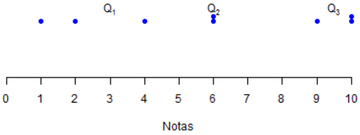
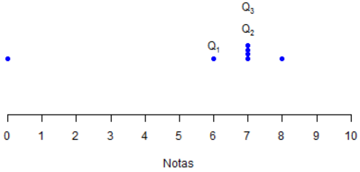


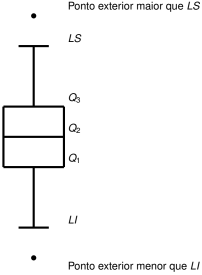
Figura 3: Turma C



Boxplot

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

Figura 4: Boxplot



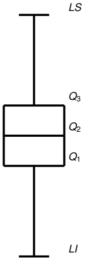
em que $LS = Q_3 + 1,5 \cdot (Q_3 - Q_1)$, $LI = Q_1 - 1,5 \cdot (Q_3 - Q_1)$.
Se um valor está entre *LI* e *LS*, este é chamado de valor Adjacente. Por outro, se um valor é maior que *LS* ou menor *LI*, este é chamado de ponto exterior e temo indícios de que o valor é atípico.

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.