

Cálculo e uso de mediana

Rodinei Antonio Pachani
Uninove. São Paulo – SP [Brasil]
pachani@uninove.br

O cálculo e o uso da mediana aparecem como dúvida em exercícios e avaliações realizados por alunos de vários cursos em que a disciplina de Estatística é ministrada. Neste trabalho, é apresentada uma comparação entre os resultados do cálculo da mediana, por meio de dois métodos específicos (A e B), para séries estatísticas que apresentem seus dados agrupados em intervalos de classe. Também é realizada uma breve análise do uso da mediana comparado ao da média. Foram utilizados oito exemplos hipotéticos para determinar a mediana e apresentados os principais resultados.

Palavras-chave: Agrupamento em intervalos de classe.
Mediana. Séries estatísticas.



1 Introdução

O cálculo da mediana constitui o erro mais freqüente cometido pelos alunos quando se trabalha com a estatística descritiva, em qualquer curso. Tal cálculo é simples para séries com dados agrupados e não-agrupados; no entanto, em se tratando de dados agrupados em intervalos, apresenta-se um tanto complexo. A maioria dos autores pesquisados utiliza a mesma forma para calcular a mediana.

Para a verificação do método utilizado no cálculo da mediana, especificamente para séries com intervalos de classe, foram pesquisados três livros (COSTA NETO, 1977; TIBONI, 2002; NAZARETH, 2003), nos quais se observaram as expressões que cada um utiliza. Verificou-se que elas apresentam diferenças, o que, contudo, não significa uma outra maneira de calcular a mediana, como será constatado. Por essa razão, nos exemplos que seguem, utilizamos apenas dois métodos para seu cálculo: um com base nos autores pesquisados e outro mais simplificado.

Nossa análise é iniciada pela definição que cada autor atribui à mediana. Em seguida, são expostas as fórmulas adotadas para o cálculo da mediana, em que se observam semelhanças. É proposto, então, o segundo método e, finalmente, por intermédio de oito exemplos, determina-se a mediana.

1.1 Desenvolvimento

Costa Neto (1977) define a mediana como uma quantidade que, a exemplo da média, também procura caracterizar o centro da distribuição de freqüências, porém conforme um critério diferente. A mediana é calculada com base na ordem dos valores que formam o conjunto de dados.

Tiboni (2002) trata-a como uma medida de posição, uma separatriz, pois divide o conjunto em duas partes iguais, com o mesmo número de

elementos. Para casos em que a série é aberta nos extremos, a mediana caracteriza o ponto médio mais confiável.

Na concepção de Nazareth (2003), a mediana é o valor que divide a distribuição ao meio, de tal modo que 50% dos dados estejam acima desse valor, e os outros 50%, abaixo. Verifica-se que as definições apresentadas não diferem muito umas das outras, daí o motivo de serem parecidas as expressões matemáticas adotadas para o cálculo da mediana.

Para o uso das expressões a seguir, todos os autores identificam antes em que classe se encontra a mediana, pois essa informação trará um dado necessário à utilização das expressões: o limite inferior da classe à qual pertence.

A expressão utilizada para determinar a classe mediana é:

$$C_m = \frac{\sum f_i}{2} \tag{1}$$

Em que:

f_i = total de elementos da série.

Depois de determinar a classe mediana e, conseqüentemente, seu limite inferior, são mostradas as expressões utilizadas para seu cálculo. Confira suas semelhanças:

Segundo Costa Neto (1977):

$$Md = L_i + \frac{(\frac{n}{2}) - Fa}{f_{md}} h_{md} \tag{2}$$

Em que:

L_i = limite inferior da classe que contém a mediana;

n = número de elementos do conjunto de dados;

Fa = soma das freqüências das classes anteriores à que contém a mediana;

f_{md} = freqüência da classe que contém a mediana;

h_{md} = amplitude da classe que contém a mediana.

Conforme Tiboni (2002):

$$Md = l_{Md} + \left[\frac{\left(\frac{\sum fi}{2} \right) - Fa}{f_{md}} \right] h_{Md} \quad (3)$$

Em que:

l_{Md} = limite inferior da classe da mediana;

F_{ant} = freqüência acumulada da classe anterior à classe mediana;

f_{Md} = freqüência da classe mediana;

h_{Md} = amplitude do intervalo de classe mediana.

Consoante Nazareth (2003):

$$Md = Li + \frac{d \cdot h}{F} \quad (4)$$

$$p = \frac{n + 1}{2} \quad (5)$$

$$d = p - Fa \quad (6)$$

Em que:

L_i = limite inferior da classe da mediana;

h = amplitude do intervalo de classe mediana;

F = freqüência da classe à qual pertence a mediana;

n = total de elementos existentes na série;

F_a = Freqüência do intervalo de classe imediatamente inferior a P.

Como se pode observar, as expressões são muito semelhantes. Em todas, são empregados os mesmos pontos referenciais: limite inferior, freqüência acumulada da classe mediana, amplitude do intervalo etc. A diferença está apenas na forma

como eles são escritos. Em suma, a expressão matemática utilizada é a mesma.

Diante disso, propomos um método de cálculo que utiliza como referência, o ponto médio da classe mediana. Para determinar a classe correspondente à mediana, faz-se uso de uma expressão parecida com a dos autores consultados (1), porém com um fator de ajuste, isto é:

$$C_m = \frac{\sum fi + 1}{2} \quad (7)$$

Nos próximos exemplos, calculamos a mediana por meio dos dois métodos: o dos autores (A) e o nosso (B). As situações e os dados das “pesquisas” são fictícios.

Intervalo (reais)	fi	xi	fac
0 a 20	15	10	15
+ de 20 a 40	5	30	20
+ de 40 a 60	7	50	27
+ de 60 a 80	3	70	30
Total	30		

Quadro 1: Quanto se gasta em um restaurante (30 entrevistados)

Fonte: O autor.

Método A

Identificação da classe mediana: $30/2 = 15$.

Portanto, 1ª classe.

Cálculo:

$$md = 0 + \left[\frac{\left(\frac{30}{2} \right) - 0}{15} \right] \cdot 20 = 20$$

Método B

Identificação da classe da mediana:

$$\frac{\sum fi + 1}{2} = \frac{31}{2} = 15,5$$

Portanto, entre a 1ª e a 2ª classe.

Cálculo:

$$md = \frac{10 + 30}{2} = 20$$



Intervalo (reais)	fi	xi	fac
0 a 100	37	50	37
+ de 100 a 200	28	150	65
+ de 200 a 300	12	250	77
+ de 300 a 400	27	350	104
+ de 400 a 500	48	450	152
Total	152		

Quadro 2: Quanto se gasta em uma casa noturna (152 entrevistados)

Fonte: O autor.

Método A

Identificação da classe mediana: $152/2 = 76$.

Portanto, 3ª classe.

Cálculo:

$$md = 200 + \left[\frac{\left(\frac{152}{2}\right) - 65}{12} \right] \cdot 100 = 291,67$$

Método B

Identificação da classe da mediana:

$$\frac{\sum fi + 1}{2} = \frac{152 + 1}{2} = 76,5$$

Portanto, 3ª classe.

Cálculo:

$$md = 250$$

Intervalo (reais)	fi	xi	fac
0 a 5	8	2,5	8
+ de 5 a 10	7	7,5	15
+ de 10 a 15	2	12,5	17
+ de 15 a 20	15	17,5	32
+ de 20 a 25	9	22,5	41
+ de 25 a 30	10	27,5	51
Total	51		

Quadro 3: Quanto se gasta* em transporte (51 entrevistados)

Obs.: *diariamente.

Fonte: O autor.

Método A

Identificação da classe mediana: $51/2 = 25,5$.

Portanto, 4ª classe.

Cálculo:

$$md = 15 + \left[\frac{\left(\frac{51}{2}\right) - 17}{15} \right] \cdot 5 = 17,83$$

Método B

Identificação da classe da mediana:

$$\frac{\sum fi + 1}{2} = \frac{52 + 1}{2} = 26$$

Portanto, 4ª classe.

Cálculo:

$$md = 17,5$$

Intervalo (reais)	fi	xi	fac
50 a 60	58	55	58
+ de 60 a 70	37	65	95
+ de 70 a 80	62	75	157
+ de 80 a 90	99	85	256
Total	256		

Quadro 4: Quanto se gasta em uma casa de repouso (256 entrevistados)

Fonte: O autor.

Método A

Identificação da classe mediana: $256/2 = 128$.

Portanto, 3ª classe.

Cálculo:

$$md = 70 + \left[\frac{\left(\frac{256}{2}\right) - 95}{62} \right] \cdot 10 = 75,32$$

Método B

Identificação da classe da mediana:

$$\frac{\sum fi + 1}{2} = \frac{257}{2} = 128,5$$

Portanto, 3ª classe.

Cálculo:

$$md = 75$$

Intervalo (reais)	fi	xi	fac
30 a 34	8	32	8
+ de 34 a 38	7	36	15
+ de 38 a 42	2	40	17
+ de 42 a 46	9	44	26
Total	26		

Quadro 5: Quando se gasta em uma viagem (26 entrevistados)

Fonte: O autor.

Método A

Identificação da classe mediana: $26/2 = 13$.

Portanto, 2ª classe.

Cálculo:

$$md = 34 + \left[\frac{\left(\frac{26}{2}\right) - 8}{7} \right] \cdot 4 = 36,86$$

Método B

Identificação da classe da mediana:

$$\frac{\sum fi + 1}{2} = \frac{27}{2} = 13,5$$

Portanto, 2ª classe.

Cálculo:

$$md = 36$$

Intervalo X (ano)	fi	xi	fac
1972-1974	123	1973,5	123
1975-1977	275	1976,5	398
1978-1980	256	1979,5	654
1981-1983	805	1982,5	1.459
1984-1986	145	1985,5	1.604
1987-1989	548	1988,5	2.152
Total	2.152		

Quadro 6: Ano de fabricação de seu automóvel (2.152 entrevistados)

Fonte: O autor.

Método A

Identificação da classe mediana: $2152/2 = 1076$. Portanto, 4ª classe.

Cálculo:

$$md = 1981 + \left[\frac{\left(\frac{2152}{2}\right) - 654}{805} \right] \cdot 3 = 1982,57$$

Método B

Identificação da classe da mediana:

$$\frac{\sum fi + 1}{2} = \frac{2153}{2} = 1076,5$$

Portanto, 4ª classe.

Cálculo:

$$md = 1982,5$$

Intervalo (reais)	fi	xi	fac
0 a 100	37	50	37
+ de 100 a 200	28	150	65
+ de 200 a 300	12	250	77
+ de 300 a 400	27	350	104
+ de 400 a 500	358	450	462
Total	462		

Quadro 7: Quanto se gasta em convênio médico empresarial (462 entrevistados)

Fonte: O autor.

Método A

Identificação da classe mediana: $462/2 = 231$.

Portanto, 5ª classe.

Cálculo:

$$md = 400 + \left[\frac{\left(\frac{462}{2}\right) - 104}{358} \right] \cdot 100 = 435,47$$

Método B

Identificação da classe da mediana:

$$\frac{\sum fi + 1}{2} = \frac{462+1}{2} = 231,5$$



Portanto, 5ª classe.

Cálculo:

$$md = 450$$

Intervalo X (unidades)	fi	xi	fac
22 a 30	12	26	12
31 a 39	27	35	39
40 a 48	25	44	64
49 a 57	80	53	144
58 a 66	90	62	234
67 a 75	54	71	288
Total	288		

Quadro 8: Quantidade de itens comprados em hipermercados (288 entrevistados)

Fonte: O autor.

Método A

Identificação da classe mediana: $288/2 = 144$.

Portanto, 4ª classe.

Cálculo:

$$md = 49 + \left[\frac{\left(\frac{288}{2}\right) - 64}{80} \right] \cdot 9 = 58$$

Método B

Identificação da classe da mediana:

$$\frac{\sum fi + 1}{2} = \frac{289}{2} = 144,5$$

Portanto, entre a 4ª e a 5ª classes.

Cálculo:

$$md = \frac{53 + 62}{2} = 57,5$$

2 Resultados e discussão

Os resultados obtidos foram os seguintes:

Exemplo	Método A	Método B
1	20,00	20,00
2	291,67	250,00
3	17,83	17,50
4	75,32	75,00
5	36,86	36,00
6	1982,57	1982,50
7	435,47	450,00
8	58,00	57,5

Quadro 9: Mediana calculada por meio dos dois métodos

Fonte: O autor.

- Nota-se que nos exemplos 1 e 8 os resultados foram exatamente iguais, tanto pelo método A quanto pelo B. Outros exemplos realizados em séries semelhantes, que não constam neste estudo, mostram que essa igualdade é constante, pois a mediana é encontrada entre duas classes, sendo, portanto, determinada pela média do ponto médio de cada classe;
- Nos exemplos de 3 a 6, observam-se resultados muito próximos. Analisando a dispersão dos dados, vê-se que eles não estão polarizados, isto é, a frequência com que os dados se repetem aparece distribuída pela série de forma homogênea, apresentando-se simetricamente. Isso leva à dedução de que, quando os dados estão dispersos em frequências não muito diferentes (como é o caso), os resultados da mediana, pelos dois métodos, se aproximam muito;
- Nos exemplos 2 e 7 encontra-se a maior discrepância, porque a dispersão está polarizada, ou seja, os dados se apresentam assimetricamente. Aqui, ocorreu o inverso do que acontece nos exemplos de 3 a 6.

Infelizmente, não é possível ter um método que seja confiável apenas para alguns tipos de séries. O método proposto, apesar de seu cálculo tentador, simples de realizar, não apresenta com

rigor os mesmos resultados obtidos pelos autores consultados.

O valor da mediana calculado pelo método A apresenta-se mais consistente e se mostra invariável nas situações em que o método B é deficiente, como se verifica nos resultados dos exemplos 2 e 7. É importante lembrar que a função da mediana é informar a posição central ou “o ponto médio” dos valores analisados, bem como a própria média, porém cada uma tem sua aplicação definida. No cálculo da média, por exemplo, todos os valores da amostra são levados em conta, ao passo que, no caso da mediana, isso não ocorre. Por essa razão, valores muito grandes ou muito pequenos, comparados aos demais valores da amostra, causam grandes variações na média, o que, em geral, não se observa em relação à mediana, ao se utilizar o método A, sendo, portanto, resistente até a valores atípicos.

3 Considerações finais

Neste trabalho, é apresentada uma comparação entre os resultados do cálculo da mediana, por meio de dois métodos específicos (A [dos três autores estudados] e B [nosso]), para séries estatísticas que apresentem seus dados agrupados em intervalos de classe. A partir dos resultados obtidos, podemos inferir que a média, ao contrário da mediana, é uma medida pouco resistente, sendo influenciada por valores “muito grandes” ou “muito pequenos”, mesmo que surjam em pequeno número na amostra. Esses valores são os responsáveis pela má utilização da média em muitas situações em que seria mais coerente adotar a mediana. Ademais, não se pode dizer generalizadamente qual dessas medidas de localização é a mais correta, pois isso depende do contexto em que se encontram. No entanto, identificamos situações em que a mediana é a mais correta:

- Quando desejamos obter o ponto que divide a distribuição em duas partes iguais;
- Quando há valores extremos que afetam, de maneira acentuada, a média aritmética;
- Quando a variável em estudo é salário, em que as distâncias entre os que ganham pouco e os que ganham muito são enormes, ou ainda, quando aqueles que ganham acima de um determinado valor estiverem na última classe. Nesses casos, a renda mediana não seria afetada por essa falta de informação, e a média não poderia ser calculada.

The calculation and the median use

The calculation and the median use appear as doubt in exercises and evaluations carried out by the students in several courses where the statistic is taught. A comparison between the median calculation results is presented by means of two methods (A and B), for statistical series that presents their data grouped in class intervals. It's also presented a short analysis of its use comparatively with the average. It has been used eight hypothetic situations to determine the median and the main results were presented.

Key words: Grouping in class intervals. Median. Statistical series.

Referências

- COSTA NETO, P. L. O. *Estatística*. 1. ed. São Paulo: Edgard Blücher, 1977.
- NAZARETH, H. *Curso básico de estatística*. 12. ed. São Paulo: Ática, 2003.
- TIBONI, C. G. R. *Estatística básica para o curso de turismo*. 1. ed. São Paulo: Atlas, 2002.

Recebido em 26 set. 2006 / aprovado em 24 nov. 2006

Para referenciar este texto

PACHANI, R. A. Cálculo e uso de mediana. *Exacta*, São Paulo, v. 4, n. 2, p. 417-423, jul./dez. 2006.

