
Medidas de tendência central: onde a maior parte dos indivíduos se encontra?

Measures of central tendency: where are the bulk of people?

RODRIGO PEREIRA DUQUIA¹
JOÃO LUIZ DORNELLES BASTOS²

1 MEDIDAS DE TENDÊNCIA CENTRAL

Conforme mencionado na primeira edição das Notas de Epidemiologia e Estatística desta revista, existem várias formas de se classificar as informações coletadas em um estudo. Inicialmente, podemos dividir as variáveis em dois grandes grupos: o das variáveis quantitativas e o das qualitativas.¹⁻³ Quando trabalhamos com variáveis quantitativas, muitas vezes o número de observações é grande e necessitamos de parâmetros para descrever de forma sucinta o comportamento desse conjunto de informações. Utilizamos as medidas de tendência central (MTC) para expressar, através de um único número, em torno de que valor tende a se concentrar um conjunto de dados numéricos.¹ Por exemplo, um pesquisador interessado em estudar a creatinina, substância dosada pelo sangue que avalia a função renal de um indivíduo, em um grupo de idosos de um asilo, afere a creatinina de 20 idosos. Posteriormente, ele registra em uma tabela (Tabela 1) a creatinina de cada indivíduo, por ordem de entrevista, a fim de estudar o valor desse metabólito nesse grupo de pessoas. Da forma como está apresentado na Tabela 1, podemos observar especificamente a creatinina de cada idoso. Entretanto, qual o valor numérico que po-

deria representar esses 20 indivíduos? Dito de outra forma, próximo de que valor encontra-se a maioria das medidas de creatinina dos idosos investigados?

TABELA 1 - Descrição da creatinina sérica de 20 indivíduos idosos.

Número do professor	Peso (kg)
1	0,6
2	0,8
3	1,0
4	0,8
5	0,9
6	1,0
7	0,7
8	0,5
9	1,1
10	0,5
11	0,8
12	1,1
13	0,8
14	0,9
15	0,8
16	0,9
17	0,7
18	0,6
19	0,8
20	0,7

¹ Dermatologista. Mestre em Epidemiologia pela Universidade Federal de Pelotas.

² Odontólogo. Mestre em Epidemiologia pela Universidade Federal de Pelotas.

Algumas pessoas responderiam a esta questão citando a média da creatinina nesses 20 idosos. No entanto, conforme veremos posteriormente, nem sempre a média é a medida mais adequada. Para responder a este questionamento, freqüentemente utilizamos as MTC, que têm como objetivo descrever de forma sucinta um conjunto de dados. Existem três principais medidas de tendência central. São elas: média, mediana e moda.

1.1 Média

Esta é a MTC mais utilizada e melhor compreendida. O cálculo da média (\bar{X}) [leia-se X barra] é feito pelo somatório [representado aqui pela letra grega sigma (Σ)] dos valores de todas as observações (indivíduos), dividido pelo número de observações (n), conforme a fórmula abaixo.¹⁻⁴

$$\bar{X} = \frac{\sum X}{n}$$

Utilizando os dados da Tabela 1 para calcular a média de creatinina do grupo de 20 idosos, verificamos que seu valor equivale a 0,8.

$$\bar{x} = \frac{0,6+0,8+1,0+0,8+0,9+1,0+0,7+0,5+1,1+0,5+0,8+1,1+0,8+0,9+0,8+0,9+0,7+0,6+0,8+0,7}{20} = 0,8$$

Ao representarmos de forma gráfica (histograma) um conjunto de dados quantitativos contínuos ou numéricos, nós podemos ter duas for-

mas de distribuição: simétrica (também chamada Normal ou Gaussiana) ou assimétrica (Não-Gaussiana) (Figura 1).¹⁻³ Em uma distribuição perfeitamente simétrica, a maioria dos dados se encontra próximo de um determinado valor, que pode ser expresso pela média, ou mediana, uma vez que nesse tipo de distribuição ambas as MTC são idênticas.¹⁻³ Os demais dados se distribuem igualmente afastando-se dos valores centrais, conforme demonstrado no gráfico A da Figura 1.

Nesta figura, representamos de forma gráfica o peso de mil sacos de arroz de um determinado depósito. A média e a mediana de peso desses sacos é de 3.200g. Como podemos observar, a maioria apresenta um peso igual ou muito próximo de 3.200g e os sacos mais pesados ou mais leves que 3.200g se distribuem igualmente para ambos os lados da média, fazendo com que a representação gráfica desse conjunto de dados seja simétrica e lembre o formato de um sino.

No gráfico B da Figura 1 adicionamos 15 sacos de arroz de 10 kg aos mil sacos já existentes. Como resultado, a forma gráfica altera-se e não mais apresenta uma distribuição perfeitamente simétrica (normal) como aquela do gráfico A. No gráfico B, a média de peso dos sacos passa a ser 3.550g, enquanto que a mediana permanece sendo 3.200g. Dessa forma, verificamos que em distribuições assimétricas não podemos utilizar o valor da média. Devemos, nestas situações, utilizar a mediana.

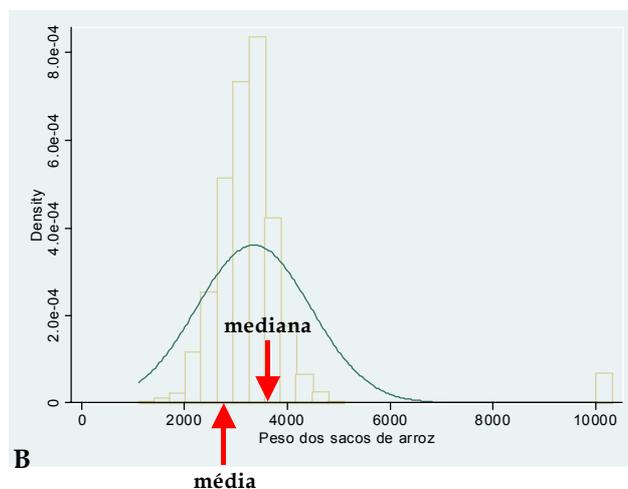
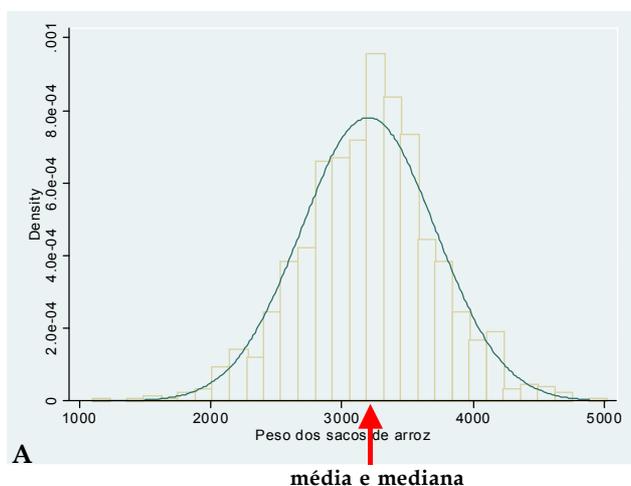


Figura 1 - À esquerda (gráfico A) temos a representação uma distribuição normal e à direita (gráfico B) a mesma representação após a inclusão de 15 sacos de arroz com 10 kg.

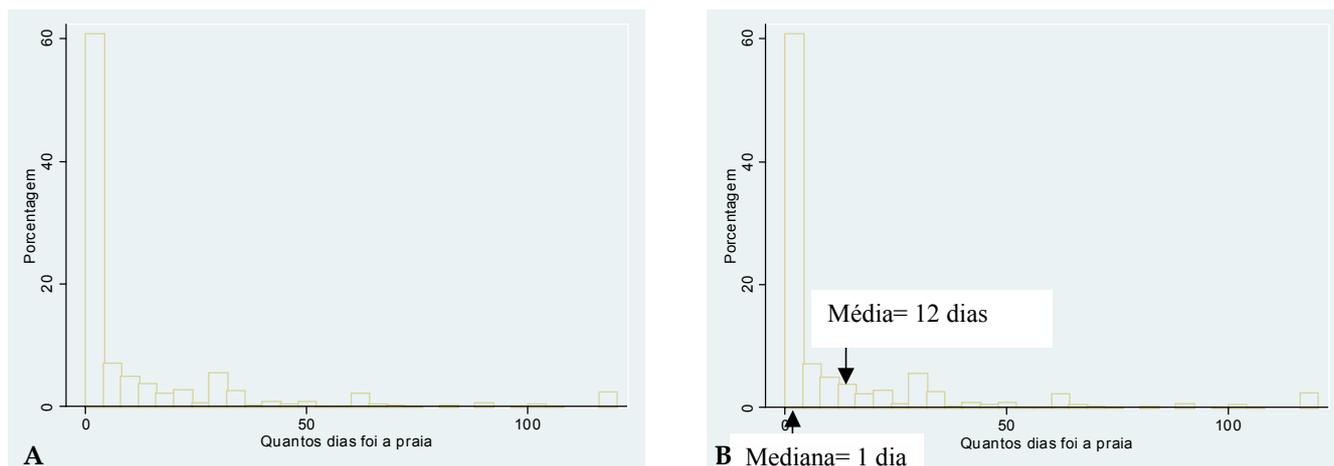


Figura 2 - (A) Representação do número de dias de exposição solar na praia. (B) Valores da média e mediana do número de dias freqüentados na praia.

No gráfico B, apresentamos o mesmo gráfico com os respectivos valores da média e da mediana do número de dias freqüentados na praia. Como podemos observar, a distribuição dos dados é assimétrica e, portanto, os valores da média e da mediana se distanciam. Neste caso, a média de dias de freqüência à praia foi 12, mas esse valor não representa o ponto onde se encontra a maioria dos valores. O ponto que representa mais adequadamente o número de dias freqüentados na praia pela maioria dos indivíduos é o valor 1. Este é o valor que divide esse conjunto de dados em duas partes iguais, ou seja, é a mediana.

Dessa forma, nos casos em que a distribuição dos dados for assimétrica, devemos utilizar como medida de tendência central preferencialmente a mediana, pois seu valor não será influenciado por valores extremos de distribuições assimétricas.¹⁻³ Sempre que utilizarmos a média como MTC para distribuições assimétricas, seu valor será influenciado pelos valores extremos do conjunto de dados, ou seja, os valores extremos de uma distribuição assimétrica “puxarão” o valor da média para perto deles.

1.3 Moda

Não foi ao acaso que deixamos essa medida por último lugar. Sua utilização é pouco freqüente e serve apenas para demonstrar qual o valor é o mais freqüente (que mais se repete) em um conjunto de dados.^{1,2} Por exemplo, na Tabela 2 demonstramos um conjunto de valores com a ida-

de em anos de alunos de uma classe de aula. Como podemos notar, existem mais alunos com 7 anos de idade nessa turma. Desta forma, dizemos que a moda desse conjunto de dados é 7, pois a maioria das crianças apresenta 7 anos de idade. Devemos lembrar que sempre que uma distribuição for perfeitamente simétrica, o valor da moda também será igual ao da média e ao da mediana. Outra observação a ser feita é que, algumas vezes, a distribuição de um conjunto de dados pode não ter moda. Isto ocorre caso as observações sejam todas diferentes entre si nesse conjunto de dados.

TABELA 2 - Lista de alunos com suas respectivas idades.

Nome dos alunos	Idade em anos
1. João	7
2. Samuel	9
3. Rodrigo	7
4. Luciano	7
5. Gustavo	10
6. Luis Artur	11
7. Luiz Henrique	7
8. Gustavo	8
9. Paulo	9
10. Gerson	7
11. Julio	7
12. Lucio	12
13. Lucas	7
14. Marcelo	9
15. Mateus	7

2 CONSIDERAÇÕES FINAIS

O conhecimento das propriedades das MTC é fundamental para descrição, interpretação e análise de dados em pesquisa.

Frente a um artigo, devemos sempre ter o cuidado de avaliar se as MTC foram bem aplicadas, pois, caso contrário, todas as conclusões dos autores e, conseqüentemente, as nossas poderão estar distorcidas. O mesmo se aplica para o tipo de teste utilizado para realizar análises estatísticas, já que o pressuposto de muitos deles é que a distribuição da variável de interesse (desfecho) tenha distribuição simétrica (normal). Dessa forma, muitas vezes, observando a forma da distribuição dos dados, sabemos se um determinado teste em um artigo foi corretamente utilizado e se seus resultados são válidos.

REFERÊNCIAS

- 1 Altman DG. Practical statistics for medical research. London: Chapman & Hall; 1997.
- 2 Kirkwood BR, Sterne JAC. Essential medical statistics. Oxford: Blackwell Science; 2003.
- 3 Massad E, Menezes RX, Silveira PSP, Ortega NRS. Métodos quantitativos em medicina. São Paulo: Manole; 2004.
- 4 Pereira MG. Epidemiologia: teoria e prática. Rio de Janeiro: Guanabara Koogan; 1995.
- 5 Duquia RP, Menezes AMB, Reichert FF, Almeida HL. Prevalence and associated factors with sunscreen use in Southern Brazil: A population-based study. J Am Acad Dermatol. 2007: in press.

Endereço para correspondência:
JOÃO LUÍZ DORNELLES BASTOS
Avenida do Antão, 353 - Morro da Cruz
CEP 88025-150, Florianópolis, SC, Brasil
Telefone: (48) 3028-1345
E-mail: joao@pilotis.com.br