Material Teórico - Módulo de ESTATÍSTICA BÁSICA I

Medidas de Posição

Primeiro Ano do Ensino Médio

Autor: Prof. Francisco Bruno Holanda
Revisor: Prof. Antonio Caminha M. Neto

1 Introdução

Após coletar os dados, organizá-los em tabelas de frequência e criar gráficos, a Estatística Descritiva também tem como objetivo criar medidas que, de certa forma, representem de maneira resumida os dados coletados.

Nesta aula abordaremos as principais **medidas estatísticas de posição** destacando suas propriedades e também os cuidados que devemos ter ao analisar estes dados "representativos".

2 Média

A média $\overline{x}$ de uma lista de n observações $(x_1, x_2, ..., x_n)$ é a média aritmética simples dos termos da lista, definida por

$$\overline{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}.$$

Em palavras, a média é a soma de todos os dados observados, dividida pela quantidade de observações realizadas. Por exemplo, considere a tabela de frequência a seguir, que reúne as quantidades de faltas registradas dos alunos do primeiro ano de uma determinada escola:

Tabela 1: Faltas por Quantidade de Alunos

Faltas	Quant. de Alunos
0	2
1	3
2	5
3	0
4	0
5	2
6	10
7	9
8	3
9	2
10	0
11	0
12	1

Veja que temos um experimento com um total de

$$2 + 3 + 5 + 2 + 10 + 9 + 3 + 2 + 1 = 37$$

observações, uma vez que há 37 alunos e estamos observando seus números de faltas. Por outro lado, a soma de todos os dados observados (i.e., das faltas dos 37 alunos) é

$$1 + 1 + 1 + 2 + 2 + 2 + 2 + 2 + 5 + 5 + \cdots + 9 + 9 + 12,$$

em que a quantidade de vezes que cada termo aparece é dada por sua frequência absoluta. Assim, a soma dos dados observados é

$$1 \cdot 3 + 2 \cdot 5 + 5 \cdot 2 + 6 \cdot 10 + 7 \cdot 9 + 8 \cdot 3 + 9 \cdot 2 + 12 = 193,$$

de forma que sua média $\overline{x}$ é dada por

$$\overline{x} = \frac{193}{37} \cong 5,21.$$

Isso significa que se ditribuíssemos a quantidade total de faltas (193) igualmente entre os 37 alunos da sala, cada um receberia aproximadamente $5,21$ faltas.

Agora, suponha que a mesma pesquisa foi realizada na turma de segundo ano dessa mesma escola, e verificou-se que a média de faltas dos alunos dessa turma foi de 5,32. Dessa forma, pergunta-se: *"será que podemos considerar que os alunos do segundo ano são* ***em geral*** *mais faltosos do que os alunos do primeiro ano?"*

A reposta é NÃO. Lembre-se de que a média é uma medida estatística que resume o conjuto de dados. Dessa forma, não podemos usar unicamente a média para tirar conclusões gerais sobre uma população. No caso particular do número de faltas dos alunos do segundo ano, considere sua respectiva tabela de frequência, dada a seguir:

Tabela 2: Faltas por Quantidade de Alunos

Faltas	Quant. de Alunos
0	12
1	3
2	5
3	1
4	0
5	0
6	10
7	2
8	0
9	0
10	0
$11 \vdash 90$	0
91	1

Na tabela de frequência acima, podemos verificar que existe um dos alunos que faltou 91 vezes. Esta observação. que está muito "distante" das demais, é capaz de *distorcer* a média. De fato, ao calcularmos a média da turma do segundo ano sem levar em consideração essa última observação, a média fica:

$$\overline{x} = \frac{90}{33} \cong 2,73.$$

Veja que o número $2,73$ é bem inferior tanto a $5,32$, que é a média de faltas da turma, quanto a $5,21$, que é a média de faltas da turma do primeiro ano.

Dessa forma, tome bastante cuidado ao analisar pesquisas focadas apenas em comparar duas médias de duas populações distintas. Tirar conclusões gerais utilizando apenas uma medida "resumitiva" dos dados podem gerar conclusões falaciosas.

Exercício 1 (ENEM 2011). *A participação dos estudantes na Olimpíada Brasileira de Matemática das Escolas Públicas (OBMEP) aumenta a cada ano. O quadro indica o percentual de medalhistas de ouro, por região, nas edições da OBMEP de 2005 a 2009:*

Tabela 3: Medalhas de Ouro na OBMEP

Região	**2005**	**2006**	**2007**	**2008**	**2009**
Norte	2%	2%	1%	2%	1%
Nordeste	18%	19%	21%	15%	19%
Centro-Oeste	5%	6%	7%	8%	9%
Sudeste	55%	61%	58%	66%	60%
Sul	20%	12%	13%	9%	11%

Em relação às edições de 2005 a 2009 da OBMEP, qual o percentual médio de medalhistas de ouro da região Nordeste?

Solução. Para descobrir o percentual médio de medalhistas de ouro de 2005 a 2009 da OBMEP referente à região Nordeste, é necessário somar os percentuais dessa região em cada ano e dividir pelos 5 anos, que nesse caso é o total de observações. Esses dados encontram-se na terceira linha da tabela:

Nordeste: 18%, 19%, 21%, 15% e 19%. Calculando a média, obtemos

$$\frac{18\% + 19\% + 21\% + 15\% + 19\%}{5} = 18,4\%.$$

□

Exercício 2. *Considere um estudo realizado sobre os salários dos funcionários de uma empresa durante os anos de* 2015 *e* 2016*, cujos dados coletados encontram-se resumidos nas tabelas a seguir:*

Tabela 4: Salários em 2015 (medidos em 1000 reais)

1,1	2	1,3	3	1,2
1,3	4,1	2	12,3	2
1,2	1,1	1,2	3	1,5
1,3	2	2,2	1,4	1,2

Tabela 5: Salários em 2016 (medidos em 1000 reais)

1	2	1,1	3	1,2
1,3	3,1	2	32,7	2
1,2	1,1	1,2	3	1,3
1,1	2	2	1	1,2

a) Calcule as médias dos salários dos funcionários nos anos de 2015 *e* 2016*.*

b) Examinando somente as médias calculadas no item (a), podemos garantir que os funcionários em geral melhoraram de situação após um ano?

Solução. A média dos salários em 2015 é obtida dividindo-se sua soma $(46,4)$ pelo total de funcionários (20), de sorte que vale $2,32$. Calculando de forma análoga a média dos salários em 2016, obtemos $3,22$.

Olhando apenas para as médias, podemos concluir erroneamente que, , de maneira geral, os funcionários melhoraram seus rendimentos de um ano para o outro. Porém, se examinarmos os dados com mais cuidado, veremos que apenas um funcionário (aquele que já tinha o salário mais alto em 2015) teve seu rendimento aumentado em 2016; todos os outros ou permaneceram com o mesmo salário de 2015 ou tiveram uma redução de um ano para o outro. □

O exemplo anterior ilustra um fato geral sobre a média aritmética de um conjunto de dados, o qual merece ser reforçado: a média é muito influenciada por *valores discrepantes*, isto é, valores muito grandes ou muito pequenos, que sejam marcadamente distantes da maior parte dos demais dados observados. Nesses casos, é necessário utilizar uma outra medida estatística de posição para representar o conjunto. Uma medida possível, que veremos na próxima seção, é a *mediana*.

3 Mediana

O primeiro passo para determinar a mediana de qualquer conjunto de dados é colocá-los em ordem crescente. Dessa forma, se $(x_1, x_2, ..., x_n)$ é uma lista de dados coletados, sejam $x_{(1)}$ o menor valor dessa coleção, $x_{(2)}$ o segundo menor e assim por diante, até chegarmos ao maior valor, definido por $x_{(n)}$. Então, temos

$$x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)},$$

e dizemos que

$$(x_{(1)}, x_{(2)}, \ldots, x_{(n)})$$

é a *sequência ordenada de dados*[1].

Caso o número de observações n seja ímpar, a mediana, denotada pela letra M é *definida* como o termo que ocupa a *posição central* na sequência ordenada de dados. Por outro lado, caso n seja par, a mediana é definida como a média aritmética dos dois termos centrais da sequência ordenada de dados. Em resumo:

[1] O leitor não deve confundir-se com o fato de que toda sequência é, em si, uma lista ordenada. Aqui, quando dizemos *sequência ordenada de dados*, estamos impondo uma *ordem* aos termos da sequência que, em princípio, é distinta da ordem original. Nesse sentido, os exemplos e exercícios discutidos a seguir esclarecerão as eventuais dúvidas que surjam.

$$M = x_{\left(\frac{n+1}{2}\right)},\ \text{se } n \text{ for ímpar}$$

e

$$M = \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2},\ \text{se } n \text{ for par.}$$

Por exemplo, considere as observações:

$$5, 2, 4, 3, 3, 4, 1, 8, 10$$

Colocando-as em ordem crescente, obtemos a sequência ordenada de dados:

$$1, 2, 3, 3, 4, 4, 5, 8, 10.$$

Como temos um número ímpar $n = 9$ de dados, existe um dado central, que no caso é o dado que ocupa a posição $\frac{n+1}{2} = 5$. Uma vez que tal dado é igual a 4, concluímos que a mediana M do conjunto de dados em questão é $M = 4$.

Por outro lado, tomemos como exemplo a seguinte coleção de dados:

$$2, 3, 6, 8, 3, 1, 7, 6, 3, 6, 3, 5,$$

que em ordem crescente fica

$$1, 2, 3, 3, 3, 3, 5, 6, 6, 6, 7, 8.$$

Desta feita, temos um número par $n = 12$ de dados, de modo que a mediana é igual à média dos dois termos centrais. Como $\frac{n}{2} = 6$, concluímos que tais termos centrais ocupam as posições 6 e 7, valendo, portanto, 3 e 5. Assim, $M = \frac{3+5}{2} = 4$.

Exercício 3. *Calcule as medianas dos dados apresentados nas tabelas 1 e 2. Explique o motivo pelo qual a diferença entre a média e a mediana no segundo grupo de dados é bem maior que a diferença correspondente no primeiro grupo.*

Solução. Uma maneira eficiente para descobrirmos a mediana é construindo as tabelas de frequências acumuladas correspondentes às tabelas 1 e 2. Assim procedendo, obtemos as tabelas 6 e 7 a seguir.

Na tabela 6 temos 37 oservações, que é uma quantidade ímpar. Portanto, a mediana será o termo de ordem $\frac{37+1}{2} = 19$ que, segundo a tabela, vale 6. Nesse caso, a diferença entre a média e a mediana será

$$5,21 - 6 = -0,79.$$

Na tabela 7, temos 34 observações, que é uma quantidade par. Portanto, a mediana será a média entre os dois termos centrais, os quais ocupam as posições $\frac{34}{2} = 17$ e 18. Segundo a tabela, os valores desses termos são ambos iguais a 2. Logo, a mediana também é igual a 2. Aqui, a diferença entre a média e a mediana será

$$5,32 - 2 = 3,32.$$

Tabela 6: Faltas Acumuladas Primeiro Ano

Faltas	Quant. de Alunos	Acumuladas
0	2	2
1	3	5
2	5	10
3	0	10
4	0	0
5	2	12
6	10	22
7	9	31
8	3	34
9	2	36
10	0	36
11	0	36
12	1	37

Tabela 7: Faltas Acumuladas Segundo Ano

Faltas	Quant. de Alunos	Acumuladas
0	12	12
1	3	15
2	5	20
3	1	21
4	0	21
5	0	21
6	10	31
7	2	33
8	0	33
9	0	33
10	0	33
$11 \vdash 90$	0	33
91	1	34

Nesse segundo caso, podemos ver como a presença de valores muito discrepantes faz com que a média e a mediana fiquem bem distantes uma da outra, o que não ocorre tão marcadamente no primeiro caso. $\square$

A seguinte pergunta cabe, agora, de maneira natural: se a mediana representa melhor a situação geral dos dados (desconsiderando os valores discrepantes), por que ela não é mais utilizada que a média?

A resposta é simples: imagine que os dados estejam desorganizados; a média é fácil de ser computada: basta somar tudo e dividir pela quantidade de observações. Por outrolado, para obter a mediana, devemos colocar os dados inicialmente em ordem crescente, um processo que é computacionalmente bem mais custoso.

À luz dos exemplos discutidos até aqui, o leitor pode pensar que esse argumento não justifica, uma vez que o trabalho que tivemos nesses casos para ordenar os dados foi bem pequeno. Entretanto, cumpre observar que em situações reais, como no senso de uma população, por exemplo, o total de dados coletados encontra-se na casa das dezenas de milhões, o que pode tornar sua ordenação bem mais trabalhosa.

Na próxima seção discutiremos uma terceira opção de medida de posição.

4 Moda

A **moda** é o valor que ocorre com maior frequência em um conjunto de dados, e será representada por x^*. (Recorde que já citamos rapidamente essa medida na aula anterior deste mesmo módulo.)

Caso os dados ainda não se encontrem agrupados em tabelas de frequência por classes, a moda pode ser facilmente identificada após organizarmos os dados em uma tal tabela de frequências. Por exemplo, considere o conjunto de observações a seguir:

5	3	7	6	3	3
4	5	1	7	5	6
5	6	5	6	6	2
1	1	4	5	1	7
5	2	4	3	5	6
1	4	5	2	1	3

Construindo a tabela de frequências correspondente, obtemos:

Tabela 8: Tabela de frequências

Valor	Frequência
1	6
2	3
3	5
4	4
5	9
6	6
7	3

A partir da tabela 8, percebemos facilmente que o valor que ocorre com maior frequência é 5 (o qual ocorre 9 vezes). Portanto, neste exemplo a moda é $x^* = 5$.

Observações

- Quando a série de dados coletados possuir dois valores com a mesma frequência máxima, cada um deles é uma moda, e o conjunto diz-se **bimodal**.
- Se mais de dois valores ocorrerem com a mesma frequência máxima, diremos que o conjunto de dados é **multimodal**.
- Quando nenhum valor é repetido, convencionamos dizer que o conjunto *não tem moda*.

Apesar da moda ser relativamente fácil de ser calculada, sua principal desvantagem é que não podemos assegurar sua existência em qualquer conjunto de dados. De fato, quando todas as observações são distintas, não há nenhum valor que represente a moda.

Finalizamos a aula resolvendo alguns exercícios para fixar as definições aprendidas até aqui.

Exercício 4 (ENEM 2010). *O quadro seguinte mostra o desempenho de um time de futebol no último campeonato. A coluna da esquerda mostra o número de gols marcados e a coluna da direita informa em quantos jogos o time marcou aquele número de gols.*

Tabela 9: Gols no campeonato

Gols Marcados	Quantidade de Partidas
0	5
1	3
2	4
3	3
4	2
5	2
7	1

Se X, Y e Z são, respectivamente, a média, a mediana e a moda dessa distribuição, é correto afirmar que:

a) $X = Y < Z$.

b) $Z < X = Y$.

c) $Y < Z < X$.

d) $Z < X < Y$.

e) $Z < Y < X$.

Solução. O primeiro passo é calcular a média de gols por partida. Como os dados estão agrupados em classes, a média é dada pela razão entre o somatório dos produtos dos valores pelas frequências com que esses valores aparecem e o somatório das frequências:

$$\begin{aligned} X &= \frac{0\cdot 5 + 1\cdot 3 + 2\cdot 4 + 3\cdot 3 + 4\cdot 2 + 5\cdot 2 + 7\cdot 1}{5+3+4+3+2+2+1} \\ &= \frac{45}{20} = 2,25. \end{aligned}$$

Para o cálculo da mediana, note inicialmente que, como temos um total de 20 observações, ela será a média aritmética dos dois valores centrais (o décimo e o décimo primeiro termos), quando a amostra está ordenada. Em ordem crescente, os onze primeiros termos são

$$0, 0, 0, 0, 0, 1, 1, 1, 2, 2, 2;$$

portanto,

$$Y = \frac{2+2}{2} = 2.$$

Por fim, pela tabela de frequências podemos verificar que a moda é $Z = 0$.

Logo, $Z < Y < X$, e a resposta correta é a letra E. $\square$

Exercício 5. *A tabela que segue é demonstrativa do levantamento realizado por determinado curso de línguas, no que se refere às idades dos alunos que são fluentes em mais de duas línguas:*

Tabela 10: Alunos Poliglotas

Idade	Número de Alunos
25	12
28	15
30	25
33	15
35	10
40	8

Calcule a moda, a média e a mediana dessa distribuição de frequências.

Solução. A moda é a observação de maior frequência; nesse caso, $x^* = 30$.

Para a mediana, veja que temos um total de

$$12 + 15 + 25 + 15 + 10 + 8 = 85$$

observações. Portanto, a mediana corresponderá à observação de ordem $\frac{85+1}{2} = 43$. Examinando a coluna da direita, notamos que $12 + 15 < 43 < 12 + 15 + 25$; logo, $M = 30$.

Por fim, para calcular a média, temos que somar todas as idades e dividir pela quantidade de alunos:

$$\begin{aligned}\overline{x} &= \frac{25\cdot 12 + 28\cdot 15 + 30\cdot 25 + 33\cdot 15 + 35\cdot 10 + 40\cdot 8}{12+15+25+15+10+8}\\ &= \frac{300+420+750+495+350+320}{85}\\ &= \frac{2635}{85} = 31.\end{aligned}$$

$\square$

5 Sugestões ao professor

Separe três encontros de 50 minutos cada para desenvolver o conteúdo desta aula. No primeiro encontro, apresente as definições e exemplos sobre a média. Na segunda apresente as definições de mediana e moda. Na terceira, verifique se os alunos aprenderam os conceitos através de exercícios propostos.

Apresente à turma cada uma das medidas de posição em sequência, passando de uma seção para a outra apenas quando ficar claro que todos compreenderam bem o conteúdo estudado e, principalmente, quando os alunos já tiverem percebido os pontos positivos e negativos de cada uma das medidas.