Material Teórico - Módulo Noções Básicas de Estatística

Fixação dos Conceitos

Sétimo Ano do Ensino Fundamental

Autor: Prof. Fabrício Siqueira Benevides
Revisor: Prof. Antonio Caminha M. Neto

27 de agosto de 2023

Portal da Matemática
OBMEP

1 Introdução

A maior parte desta aula é dedicada à resolução de exercícios sobre os conceitos estudados na aula anterior, seguindo os mesmos exemplos que constam nas vídeo-aulas. Ao final, introduziremos um novo parâmetro estatístico para variáveis quantitativas, a *amplitude* de um conjunto de números.

2 Fixação dos conceitos

Exemplo 1. *Em uma turma de sétimo ano, as notas obtidas pelos alunos na prova de Matemática (numa escala de 0 a 100) foram:*

$$13, 34, 45, 26, 19, 27, 48, 63, 81, 76, 52, 86, 92, 98.$$

Vamos calcular a média e a mediana das notas e discutir um pouco esses dados.

Solução. Antes de mais nada, observe que há 14 números na lista, o que indica que 14 alunos fizeram a prova.

Agora, para calcular a média dos números, basta somá-los e dividir o resultado por 14. Assim fazendo, temos:

$$\text{Média} = \frac{760}{14} \approx 54{,}28.$$

Observação: usamos o símbolo "$\approx$" com o significado de "aproximadamente".

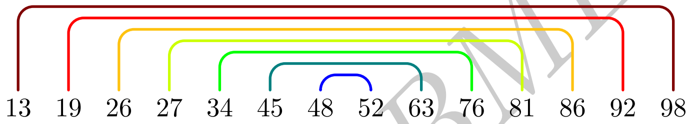
A interpretação da média é que se todos os alunos tivessem tirado a mesma nota, a fim de obter a mesma soma total de 760 pontos pelos 14 alunos, essa nota seria 54,28. Imaginando que a nota mínima para passar de ano fosse 50 pontos, se todos os alunos tivessem obtido nota igual à média, todos teriam sido aprovados (apesar de que teriam passado de ano com uma nota relativamente baixa).

Acontece que a média é uma informação simplificada da realidade. Observando os dados mais atentamente, vemos que a turma é bastante heterogênea e há muitos alunos com notas bem maiores e outros com notas bem menores que a média.

Para calcular a mediana, precisamos ordenar as notas. Listando-as da menor para a maior, obtemos:

$$13, 19, 26, 27, 34, 45, 48, 52, 63, 76, 81, 86, 92, 98$$

Como há 14 números (uma quantidade par deles), não há um único número que esteja no centro da lista; como $14/2 = 7$, os dois números mais ao centro são o sétimo e o oitavo números. Veja:



Assim, tais números são 48 e 52, de forma que a mediana é a média desses dois números:

$$\text{Mediana} = \frac{48 + 52}{2} = 50.$$

Uma vez que a mediana é igual a 50, a turma tem uma quantidade par de alunos e nenhum aluno tirou exatamente 50 pontos, concluímos que exatamente metade dos alunos ficaram com nota abaixo de 50 e metade acima de 50.

Quanto à moda, ela acaba não sendo uma informação relevante, já que cada uma das notas que aparece na lista aparece exatamente uma vez. Ou seja, não há valor que se destaque. De fato, temos um conjunto multimodal, no qual todos os números do conjunto fazem parte da moda. □

Exemplo 2. *O peso médio dos 100 alunos de uma academia de ginástica é igual a* 75 kg. *O peso médio dos homens é* 90 kg *e o peso médio das mulheres* 65 kg

(a) *Quantos homens frequentam a academia?*

(b) *Se não consideramos os 10 alunos mais pesados, o peso médio cai de* 75 kg *para* 72 kg*. Qual o peso médio desses 10 alunos?*

Solução.
(a) Seja H a quantidade de homens. Como o total de pessoas na academia é 100, temos que $100 - H$ é a quantidade de mulheres. Como o peso médio dos homens é 90 kg, temos que a soma dos pesos de todos os homens é $90H$. Da mesma forma, como o peso médio das mulheres é 65 kg, temos que a soma dos pesos de todas as mulheres é $65(100 - H)$. Assim, a soma dos pesos de todos os alunos é $90H + 65(100 - H)$.

Para terminar, lembre-se de que o peso médio de todas as 100 pessoas é 75 kg, logo

$$\frac{90H + 65(100 - H)}{100} = 75.$$

Resolvendo essa equação, obtemos

$$\begin{aligned} 90H + 6500 - 65H &= 7500 \\ \Longrightarrow\ 25H &= 1000 \\ \Longrightarrow\ H &= 40. \end{aligned}$$

Ou seja, há 40 homens na academia e, consequentemente, há 60 mulheres.

(b) Vamos chamar de P o peso médio das 10 pessoas mais pesadas da academia. Isso quer dizer que a soma dos pesos dessas pessoas é $10P$. Por outro lado, a soma dos pesos de todas as pessoas da academia é $75 \cdot 100 = 7500$. Assim, temos que a soma dos pesos das 90 pessoas mais leves é $7500 - 10P$.

O enunciado diz que a média dos pesos dessas 90 pessoas é 72 kg. Logo, obtemos a equação:

$$\frac{7500 - 10P}{90} = 72,$$

de sorte que

$$7500 - 10P = 90 \cdot 72 = 6480.$$

Portanto, $10P = 7500 - 6480 = 1020$, o que implica que $P = 102$. Assim, o peso médio das 10 pessoas mais pesadas é $102\,\text{kg}$. □

Exemplo 3. *As empresas A e B produzem um mesmo tipo de produto e oferecem condições de trabalho equivalentes, a menos de suas políticas de remuneração. A tabela seguinte traz os valores da média, mediana e moda dos salários das duas empresas.*

Empresa	*A*	*B*
Média	2500 *reais*	2000 *reais*
Mediana	1700 *reais*	1900 *reais*
Moda	1500 *reais*	1900 *reais*

Se você estivesse procurando emprego, qual das empresas escolheria?

Solução. Primeiramente, note que não há uma resposta definitiva para essa pergunta, uma vez que os dados apresentados são insuficientes. Por exemplo, não sabemos quantos funcionários cada empresa possui e também não é possível inferir qual o maior salário pago a um funcionário de cada empresa. De toda forma, podemos fazer algumas observações.

Na empresa A, a média salarial é de 2500 reais, mas a mediana é de 1700 e a moda apenas 1500. Para que isso aconteça, é **necessário** que os salários da empresa A variem bastante. Cerca de metade dos funcionário ganham abaixo de 1700 reais, enquanto o salário mais comumente pago é de apenas 1500 reais.

Apesar disso, a média salarial é 2500, o que garante que há pelo menos um funcionário que recebe muito acima de 2500. De fato, como existem funcionários que ganham 1500, ou seja, 1000 reais abaixo da média, para compensar é preciso que existam funcionários que ganhem 1000 reais acima da média. Ou seja, conclui-se que alguém da empresa A recebe salário de pelo menos 3500 reais.

Assim, se você tiver um currículo altamente qualificado e acreditar que será contratado para um cargo de alto escalão, a empresa A pode ser uma boa opção. Por outro lado, se seu currículo não for tão bom, há bastante chance de receber apenas 1500, que é o que a maioria das pessoas que trabalha na empresa A recebe.

Já na empresa B, tanto a média, quanto a moda e a mediana ficam na faixa de 1900 a 2000 reais. Se você for um empregado com currículo mediano, tudo indica que na empresa B você irá receber cerca de 1900 reais, que, além de ser o salário da maioria dos funcionários da empresa B, não está muito distante da média de salários da empresa.

Não podemos descartar completamente a possibilidade de que haja alguns salários muito acima dos 2000 ou muito abaixo dos 1900, mas, como empregado mediano, você provavelmente receberia um salário perto dos 1900 na empresa B, que é bem maior do que o salário mais comum da empresa A (1500). $\square$

Exemplo 4. *Discuta sobre qual medida central (média, moda ou mediana) é a mais adequada nos seguintes casos.*

(a) Estão disponíveis dados mensais sobre a incidência de envenenamento por picada de cobra. Deseja-se planejar a compra mensal de antídoto.

(b) Um fabricante de baterias deseja divulgar a durabilidade de seu produto e coleta informações sobre a duração de 100 *de suas baterias.*

Solução. Para o item (a), a medida mais adequada para o planejamento seria a quantidade média de picadas.

Realmente, de posse dessa quantidade, podemos estimar adequadamente a quantidade total de picadas no longo prazo, assim como calcular a quantidade total de antídoto necessária no período.

É possível que, num dado mês, haja mais picadas do que o usual e seja gasto mais antídoto. Porém, esse gasto precisa ser compensado por um mês onde haja menos picadas e, consequentemente, menor uso do antídoto. Assumindo que a sobra de um mês possa ser utilizada no mês seguinte, a média é a medida mais adequada para o planejamento.

Para o item (b), o fabricante deveria considerar a divulgação da moda do tempo de duração das baterias. Isto porque, ao fazê-lo, ele divulgará, para o cliente que comprar uma bateria nova, o tempo de vida mais provável da mesma. Além disso, os processos de fabricação provavelmente garantem que, na maioria dos casos, esse tempo de vida é respeitado.

É possível que algumas baterias venham com defeito de fábrica e durem pouquíssimo tempo, o que pode afetar negativamente a média.

Por exemplo, digamos que de cada 100 baterias 90 durem 5 anos, enquanto 10 deles durem 0 anos. Nesse caso, a média é $(95 \cdot 5 + 10 \cdot 0)/100 = 4{,}5$ anos (quatro anos e meio). Porém a moda é de 5 anos.

Num tal cenário, o fabricante está sendo honesto ao divulgar que a maioria de suas baterias dura 5 anos (e usando uma estratégia de marketing mais adequada do que se divulgasse uma durabilidade de apenas 4 anos e meio).

Além disso, os clientes que recebem baterias defeituosas devem ser compensados pelo fabricante, por meio de uma política de garantia do produto. Assim, a moda também acaba sendo a informação mais relevante para os clientes. □

Exemplo 5. *Dentre os recém formados de uma determinada área de estudos e com empregos considerados bons, foi sorteada uma amostra e analisada a quantidade de anos de estágio realizados no decorrer da faculdade (e anteriormente à formatura). Observando, na tabela abaixo, os dados obtidos, calcule a média, a moda e a mediana da quantidade de anos de estágio realizada.*

Anos de estágio	0	1	2	3	4	5	6	***Total***
Frequência	25	58	147	105	72	45	10	462

Solução. O número 462 indica que a amostra possui 462 estudantes. A primeira linha da tabela indica a quantidade de anos de estágio que cada um desses alunos realizou (os alunos que fizeram estágio por mais tempo fizeram 6 anos de estágio), ao passo que a segunda linha traz a frequência absoluta de cada quantidade de anos. Ou seja, a tabela nos diz que 25 alunos não realizaram estágio, 58 alunos realizaram um ano de estágio e assim sucessivamente.

Observe que a variável que estamos analisando é o "número de anos de estágio"; queremos saber a média do número de anos, a moda do número de anos e a mediana do número de anos.

A informação mais direta de se obter a partir da tabela é a moda: fica bastante claro que 2 anos é a quantidade de anos de estágio mais frequente, logo, a moda do número de anos de estágio é igual a 2. De fato, 147 alunos realizaram 2 anos de estágio e 147 é o maior número da segunda linha da tabela.

Para calcular a média, o número de anos precisa ser ponderado pelas frequências absolutas (conforme estudamos na aula anterior). Calculando dessa forma, vemos que a média do tempo de estágio realizado pelo 462 alunos é:

$$\frac{0\cdot 25+1\cdot 58+2\cdot 147+3\cdot 105+4\cdot 72+5\cdot 45+6\cdot 10}{462}=$$
$$=\frac{58+294+315+288+225+60}{462}=\frac{1240}{462}\approx 2{,}68.$$

Assim, os alunos ficaram, em média, 2,68 anos realizando estágio.

Por fim, para obter a mediana, também não podemos ignorar as frequências. Imagine uma lista, extensa, com 462 números, onde os 25 primeiros números são iguais a 0, os 58 números seguintes são iguais a 1, os 147 números seguintes são iguais a 2 e assim sucessivamente, até listarmos todos os 462 números.

$$\underbrace{0,\ldots,0}_{25},\underbrace{1,\ldots,1}_{58},\underbrace{2,\ldots,2}_{147},\underbrace{3,\ldots,3}_{105},\ldots$$

A mediana é a média dos números que estão no centro dessa lista. Como há 462 números, precisamos saber quais números ocupam as posições 231 e 232 da lista. Como $25 + 58 + 147 = 230$, o último número 2 ocupa a posição 230. Assim, tanto a posição 231 como a posição 232 são ocupadas pelo número 3. Logo, a mediana é igual a 3. □

Observação 6. *Caso você não tenha compreendido bem a informação contida na tabela do exemplo anterior, sugerimos que revise a seção sobre "frequências" da aula anterior. Em problemas de Estatística, é muito importante compreender a natureza dos dados. Não é produtivo tentar "chutar" uma resposta, uma vez que, pela lista de números da tabela, pode-se fazer vários tipos de cálculos errôneos. Por exemplo, a média dos números*

$$\frac{25 + 58 + 147 + 105 + 72 + 45 + 10}{7} = 62{,}14,$$

não é algo relevante para o problema. Da mesma forma, também não é relevante a média não ponderada da quantidade de anos

$$\frac{0 + 1 + 2 + 3 + 4 + 5 + 6}{7} = 3.$$

Para fixar os conceitos, vejamos outros exemplo bastante semelhante.

Exemplo 7. *Foi realizada uma pesquisa onde perguntou-se o número de computadores que cada entrevistado possui em sua residência. O resultado da pesquisa foi o seguinte:*

número de computadores	0	1	2	3	4	***Total***
Frequência	156	135	47	25	8	371

Calcule a média, a moda e a mediana da quantidade de computadores.

Solução. Fica claro que a moda é igual a 0, que é o número com maior frequência.

Por outro lado, para calcular a média, basta fazer:

$$\frac{0\cdot 156 + 1\cdot 135 + 2\cdot 47 + 3\cdot 25 + 4\cdot 8}{371} =$$
$$= \frac{135 + 94 + 75 + 32}{371} = \frac{336}{371} \approx 0{,}9.$$

Agora, para calcular a mediana, precisamos saber qual número está na posição central da lista de números

$$\underbrace{0,\ldots,0}_{156},\underbrace{1,\ldots,1}_{135},\underbrace{2,\ldots,2}_{47},\underbrace{3,\ldots,3}_{28},\underbrace{4,\ldots,4}_{8}.$$

Essa lista possui 371 elementos, que é uma quantidade ímpar. Como $371/2 = 185{,}5$, o número que está no centro da lista é o que ocupa a posição 186.

Ora, há 156 números iguais a zero na lista e, como $186 - 156 = 30$, queremos o trigésimo número listado após os zeros. Como o número 1 é listado 135 vezes (e $135 > 30$) fica claro que o número que está na posição 186 é o número 1. Assim, a mediana é igual a 1. □

3 Um novo parâmetro: a amplitude.

Uma máquina de cortar tampas de caixas de CD tem formato quadrado, de 12 cm de lado. Para fazer o controle diário da produção, a instrução dada ao operário é a seguinte: retire as 5 primeiras tampas cortadas em cada dia e avalie o comprimento de um dos lados. O objetivo é saber se o processo está sob controle, com o maquinário ajustado corretamente, no momento em que a produção está sendo iniciada, ou se há necessidade de ajustes.

A seguir, temos os valores que o operário anotou em cada dia.

Dia 1: 12 cm, 10 cm, 14 cm, 12 cm, 12 cm.

Dia 2: 11,5 cm, 12,0 cm, 11,5 cm, 12,5 cm, 12,5 cm.

Dia 3: 11,9 cm, 12,0 cm, 12,05 cm, 12,01 cm, 11,99 cm.

Imagine, ainda, que o funcionário foi orientado a anotar em seu relatório apenas a média e a mediana dos valores que ele aferiu.

Para o conjunto de dados acima, verifique que obtemos:

Dia	**Média**	**Mediana**
1	12	12
2	12	12
3	12	12

É curioso que, a média e a mediana não conseguem distinguir os três dias, escondendo potenciais defeitos apresentados pela máquina.

Para reduzir esse efeito, introduziremos um novo parâmetro, chamado de *amplitude*.

> A **amplitude** de uma conjunto de dados é uma medida de variabilidade obtida pela diferença entre o maior valor (máximo) e o menor valor (mínimo) do conjunto.

No dia 1, o maior tamanho que a máquina cortou foi de 14 cm e o menor foi de 10 cm; assim, a amplitude no dia 1 foi de $14 - 10 = 4\,\text{cm}$. Procedendo analogamente, vemos que, no dia 2, a amplitude foi dada por $12{,}5 - 11{,}5 = 1\,\text{cm}$; já no dia 3, a amplitude foi de $12{,}05 - 11{,}95 = 0{,}1\,\text{cm}$.

Vamos estender a tabela anterior, incluindo a informação sobre a amplitude das medidas de cada dia.

Dia	**Média**	**Mediana**	**Amplitude**
1	12	12	4,0 cm
2	12	12	1,0 cm
3	12	12	0,1 cm

Agora, fica claro que no primeiro dia a performance da máquina estava muito ruim. No segundo dia melhorou bastante, mas provavelmente ainda estava longe do esperado. A amplitude de 1,0 cm vem de desvios de pelo menos 0,5 cm em relação à média, o que ainda significa que estão sendo produzidas tampas que não irão encaixar bem nas caixas. Já no terceiro dia houve uma melhora considerável, e é possível que uma erro de 0,1 cm não seja algo significativo nesse tipo de processo de fabricação.

Veja que uma amplitude pequena garante que o desvio máximo em relação à média também foi pequeno. Assim, amplitudes pequenas nos garantem conjuntos de dados mais homogêneos.

Dicas para o Professor

Este material pode ser abordado em dois encontros de 50 minutos. Ele tem como pré-requisitos a familiaridade com números decimais, razões e proporções e a conversão da notação fracionária para a notação de porcentagem.