

Muitos textos didáticos de Matemática para a escola fundamental introduzem o conhecimento estatístico por meio da construção e observação de gráficos e tabelas. Como medidas para resumir os dados (medidas-resumo) são normalmente apresentadas as chamadas medidas de locação, a saber: **média, mediana e moda**. Isso se repete na escola média, mas dessa vez há o acréscimo, vez ou outra, das medidas de variabilidade como variância e desvio padrão, medidas sofisticadas para uma primeira abordagem do tema – isso talvez explique a apresentação mecânica dos conceitos, com ênfase nos cálculos, sem significado explícito para os alunos.

Nas orientações aos professores, muitas vezes se encontra que o *tratamento da informação* (substituto do termo *estatística* nos documentos definidores dos currículos) permite interpretar dados e fazer inferências (isso é colocado tanto no fundamental como no médio). O termo inferência na área de estatística tem um significado próprio, e deve ser usado na perspectiva de ampliar para uma população o resultado encontrado na amostra, ampliação essa que será mediada pela probabilidade (isso não será discutido aqui). Apesar de o raciocínio inferencial não ser objeto de estudo do ensino básico, a preparação para ele não pode prescindir da introdução de medidas de variabilidade desde a primeira abordagem do tratamento de dados.

Ao utilizar somente as medidas média, mediana e moda, mesmo que no início da alfabetização estatística, mostra-se ao aluno um lado do problema – seria como uma visão estática da situação. A sugestão é introduzir, juntamente com as medidas de locação mencionadas, uma medida de variabilidade muito simples, a **amplitude**, que permitirá ao aluno uma visão mais abrangente de um conjunto de dados. O exemplo a seguir mostra a importância dessa abordagem.

Será mencionado aqui um problema ligado à área de controle de qualidade, em uma atividade desenvolvida no chamado “chão” de fábrica. Imagine que se está diante de uma máquina de cortar tampas de uma caixa de CD, segundo um quadrado de 12 cm de lado. Para fazer o controle diário da produção desde o início do processo, a instrução que se dá ao operário é do seguinte tipo: retire as 5 primeiras tampas cortadas em cada dia e avalie o comprimento de um dos lados, para saber se o processo está sob controle ou se ajustes têm que ser feitos na máquina. Sem entrar nos detalhes do controle de qualidade propriamente dito, podem ser analisadas algumas situações de observação das cinco primeiras peças em três dias diferentes:

Dia 1 (5 peças) 12 cm 10 cm 14 cm 12 cm 12 cm

É imediato verificar que **média** = 12 cm. Nesse caso, pode-se observar que também a mediana é 12 cm.

Dia 2 (5 peças) 11,5 cm 12 cm 11,5 cm 12,5 cm 12,5 cm

É também imediato verificar que **média** = 12 cm e também nesse caso pode-se observar que a mediana é 12 cm.

Dia 3 (5 peças) 11,95 cm 12 cm 12,05 cm 12,01 cm 11,99 cm

Novamente nessa situação temos que **média** = **mediana** = 12 cm.

Em todos os dias, somente com o conhecimento de média (ou mediana), o operário estaria inclinado a aceitar o processo, pois a média é o próprio valor-alvo. No entanto, observando os resultados, é possível ver que as produções diferem bastante entre si quanto à *variabilidade* e que as medidas calculadas até então não conseguem quantificar esse fato. Ou seja, somente com a informação da média (ou da mediana) não é possível ter uma visão completa do comportamento de um conjunto de dados. Isso sugere o uso de outras medidas-resumo para complementar a análise. A mais imediata medida-resumo que traduz variabilidade, de fácil entendimento, e que poderia ser explorada nesse contexto tanto no nível fundamental como no médio é a **amplitude**, que é uma medida de variabilidade obtida da diferença entre o maior valor (máximo) e o menor valor (mínimo), ou seja:

$$\text{amplitude} = \text{máximo} - \text{mínimo}.$$

Para os três dias analisados, tem-se:

$$\text{Dia 1} \quad \text{média} = 12 \text{ cm} \quad \text{amplitude} = 14 \text{ cm} - 10 \text{ cm} = 4 \text{ cm}$$

$$\text{Dia 2} \quad \text{média} = 12 \text{ cm} \quad \text{amplitude} = 12,5 \text{ cm} - 11,5 \text{ cm} = 1 \text{ cm}$$

$$\text{Dia 3} \quad \text{média} = 12 \text{ cm} \quad \text{amplitude} = 12,05 \text{ cm} - 11,95 \text{ cm} = 0,1 \text{ cm}$$

O operário poderia escrever em seu relatório que em todos os dias as máquinas tiveram um funcionamento inicial, **em média**, de acordo com o esperado, mas que a variabilidade foi muito diferente. Isso mostra a diferença na qualidade e precisão do produto, e sugere a revisão da máquina. Quanto menor for a variabilidade na produção, maior será a precisão do produto, indicando então que no dia 3 a máquina apresenta resultados mais precisos do que nos outros dias. É interessante notar que mesmo o dia 1, em que a moda é 12 (que é o alvo), valor similar ao das outras medidas de locação, apresenta valores descrepantes, o que resulta na maior dispersão encontrada, caracterizada pela amplitude. Se um possível requisito fosse não se afastar do valor-alvo (12 cm) por mais do que 0,05 cm, obter uma amplitude igual a 0,1 cm (dobro da variabilidade permitida para cada lado) seria adequado em princípio, fazendo com que o dia 3 tivesse a produção inicial dentro dos requisitos de especificação da produção. Assim, com uma medida de posição, a **média**, e uma medida de variabilidade, a **amplitude**, o usuário estaria mais preparado para tomar decisões. Também é importante lembrar que só a medida de variabilidade seria insuficiente, pois caso a amplitude fosse 0 cm (situação aparentemente desejável), como o que seria obtido se as medidas fossem, por exemplo, 11 cm 11 cm 11 cm 11 cm 11 cm, isso inviabilizaria a produção, pois pela média (= 11 cm) seria detectado que o processo estaria fora do controle, uma vez que, como visto, o alvo é 12 cm.

Isso posto, devemos sempre associar a uma medida de locação uma outra de variabilidade. Evidentemente, calculando o desvio padrão, chegariamos à mesma conclusão sobre a precisão das medidas para os três dias (os valores de desvio padrão de cada amostra para os dias 1, 2 e 3 são, respectivamente: 1,41 cm, 0,50 cm e 0,036 cm). Apesar de o desvio padrão ser uma medida mais precisa de variabilidade do que a amplitude, o uso da amplitude nas primeiras análises de um conjunto de dados permite ao aluno quantificar a variação presente no conjunto de dados. De fato, isso é usado em uma etapa inicial de controle de qualidade, pois é mais fácil que um operário seja treinado inicialmente para calcular amplitude do que para calcular o desvio padrão.

Com medidas de posição e de variabilidade tem-se mais conhecimento do comportamento do processo, o que facilita a tomada de decisão. Nesse processo, em que cinco peças são retiradas, o intuito é saber se a máquina está pronta para a produção em série das peças ou se ela necessita ser calibrada. Devemos lembrar que a produção é em série e em larga escala e a precisão é vital para a redução de desperdício. Com esse exemplo, vê-se que a introdução da noção de variabilidade através da simplicidade de cálculo da amplitude favorece a aprendizagem da estatística e de seus propósitos, sendo que, quando oportuna, a inclusão de discussão do desvio padrão ficará mais natural, já que essa é uma medida mais completa de variabilidade, pois inclui todos os valores em seu cálculo e não somente os valores extremos.

Provavelmente, a não apresentação da simples medida de variabilidade (a amplitude) no contexto escolar se deve a sua ausência dos Parâmetros Curriculares Nacionais (PCN) 1997 (fundamental I) e 1998 (fundamental II), que, apesar de estabelecerem a necessidade de sintetizar informações e elaborar inferências, não mencionam a existência de medidas de variabilidade, colocando somente média-mediana-moda como medidas necessárias. Somente nos PCN do ensino médio de 1998 são mencionadas as medidas variância e desvio padrão. Mas ainda aí a amplitude não é mencionada.

Garfield & Ben-Zvi (2008), pesquisadores renomados internacionalmente em Educação Estatística, discutindo o lugar de variabilidade no currículo da escola básica, colocam

“A ideia de dispersão, ou variabilidade, deveria permear todo o currículo”.

Associar medidas de variabilidade ao ensino de medidas de locação desde o início da aprendizagem em estatística permitirá ao aluno desenvolver ferramentas para interpretar corretamente situações relacionadas ao tratamento estatístico de dados.

Referências

Parâmetros Curriculares Nacionais, MEC (1997, 1998).
Garfield, J.B. & Ben-Zvi, D. *Developing students' statistical reasoning*. Springer, 2008.