O Campo Conceitual das Medidas de Dispersão: Discussões Iniciais

The conceptual field of dispersion measures: initial discussions

Irene Mauricio Cazorla[1]
Edicarlos Pereira de Sousa[2]
Carlos Eduardo Ferreira Monteiro[3]

Resumo: O ensino de Estatística aborda vários conceitos na Educação Básica, destacando-se as Medidas de Dispersão, essenciais para analisar variáveis estatísticas em diversas situações cotidianas. Este trabalho constitui-se num ensaio teórico que tem por objetivo apresentar uma aproximação do campo conceitual das Medidas de Dispersão, propondo situações vinculadas à natureza da variável e à organização dos dados. Para os aspectos metodológicos, mapeou-se cinco situações com base no arcabouço da Teoria dos Campos Conceituais. Nossas reflexões ratificam a importância de compreender a distribuição das variáveis, envolvendo um número razoável de dados e o uso de gráficos adequados. Sugere-se que estudantes experimentem diversas situações para ampliar seus repertórios e reconhecer qual é o algoritmo das Medidas de Dispersão mais adequado para cada caso.

Palavras-chave: Situações de Ensino. Medidas de Dispersão. Educação Estatística. Educação Básica. Teoria dos Campos Conceituais.

Abstract: The teaching of Statistics covers various concepts in Basic Education, highlighting the Measures of Dispersion, which are essential for analyzing statistical variables in different everyday situations. This work constitutes a theoretical essay aimed at presenting an approximation to the conceptual field of Measures of Dispersion, proposing situations related to the nature of the variable and the organization of the data. For methodological aspects, five situations were mapped based on the framework of the Theory of Conceptual Fields. Our reflections confirm the importance of understanding the distribution of variables, involving a reasonable number of data points and the use of appropriate graphs. It is suggested that students experiment with various situations to expand their repertoires and recognize which algorithm for Measures of Dispersion is most suitable for each case.

Keywords: Teaching Situations. Measures of Dispersion. Statistical Education. Basic Education. Theory of Conceptual Fields.

1 Introdução

O ensino de Estatística foi institucionalizado no âmbito da Educação Básica brasileira com a publicação dos Parâmetros Curriculares Nacionais – PCN (Brasil, 1997). A Base Nacional Comum Curricular – BNCC (Brasil, 2017) reiterou essa inserção dos conceitos estatísticos no currículo escolar ao vinculá-los à unidade temática Probabilidade e Estatística, uma das cinco associadas ao componente curricular Matemática.

Alguns estudos indicam avanços no campo da Educação Estatística, com uma quantidade crescente de trabalhos e de pesquisadores (Lima, Paula & Giordano, 2022; Schreiber & Porciúncula, 2019). A inserção do ensino de Estatística nos currículos da Educação

[1] Universidade Estadual de Santa Cruz • Itabuna, BA - Brasil • ✉ icazorla@uol.com.br • ORCID https://orcid.org/0000-0003-3028-5513
[2] Universidade Federal do Cariri • Brejo Santo, CE - Brasil • ✉ edicarlos.pereira@ufca.edu.br • ORCID https://orcid.org/0000-0003-4262-544X
[3] Universidade Federal de Pernambuco • Recife, PE - Brasil • ✉ carlos.fmonteiro@ufpe.br • ORCID https://orcid.org/0000-0003-4355-0793

CC BY NC SA
ISSN 2764-3158

Básica foi discutida por Cazorla e Utsumi (2010), que evidenciaram um rompimento com a cultura determinística nas aulas de matemática, alinhando-se a uma tendência internacional que ratifica o papel da Estatística na formação de cidadãos críticos. Apesar dos avanços, o ensino de conceitos estatísticos ainda se constitui um desafio (Samá & Silva, 2023).

No caso do Brasil, os documentos oficiais orientam que sejam trabalhados os conceitos estatísticos básicos, dentre eles as Medidas de Dispersão (MD): amplitude total, desvio médio, variância e desvio padrão. Apesar de sua importância, Barbosa, Santos e Lopes (2019) alertam que o ensino das MD não tem recebido a merecida atenção.

As MD são essenciais para analisar o comportamento de uma variável e estão intrinsecamente relacionadas com a compreensão conceitual da variabilidade, considerada um dos principais e mais complexos conceitos estatísticos (Cobb & Moore, 1997; Garfield & Ben-Zvi, 2005; Shaughnessy, 2007; Watson, Fitzallen, Wright & Kelly, 2023). Assim, percebe-se que articular a compreensão das MD com a variabilidade é primordial para a análise das variáveis estatísticas que fazem parte de diversas situações em nosso cotidiano. A abordagem da variabilidade no currículo se dá com maior ênfase no Ensino Médio, para o qual a BNCC defende abordagens de "variação e constância" e "certeza e incerteza", diretamente relacionadas com contextos em que prevalecem concepções de variabilidade.

Em geral, no ensino de variabilidade e das MD, estudantes e professores se limitam ao cálculo e leitura dos resultados, desconsiderando as implicações e a razoabilidade dos valores obtidos. Essas práticas dificultam a compreensão desses conceitos. Para Isoda, Chitmun e González (2018), compreender a variabilidade transcende o cálculo de medidas de dispersão, tais como amplitude (total), variância, desvio padrão ou coeficiente de variação. Os autores ratificam a importância de transpor o tratamento matemático nas aulas de Estatística, que conduzem os discentes a determinados cálculos, quase sempre, sem entender o que realmente representam. As Diretrizes para Avaliação e Instrução em Educação Estatística – GAISE, da Associação Estatística Americana, declaram que "é esse foco na variabilidade dos dados que diferencia a estatística da matemática" (Franklin *et al*., 2007, p. 6).

Os esforços para sistematizar situações de ensino, exemplificando possibilidades de apresentação das MD em sala de aula, tornam-se uma importante contribuição para a Educação Estatística. Neste trabalho, situações de ensino são entendidas conforme as situações da Teoria dos Campos Conceituais – TCC, com base na diversidade de possíveis situações que o professor vai se deparar ao ensinar as MD. Diante disso, este ensaio teórico concentra-se em discutir quais situações devem ser consideradas quando abordamos as MD na Educação Básica, apontando especificidades e chamando a atenção para que sejam feitas escolhas apropriadas na resolução de situações-problema em que a variabilidade estatística se faz presente.

Tomando como base o que já foi investigado sobre as MD e o que foi proposto para o campo conceitual das Medidas de Tendência Central - MTC (Cazorla, Utsumi & Magina, 2023), este artigo objetiva apresentar uma primeira aproximação do campo conceitual das MD, fundamentando-se na TCC, discutida mais adiante. Desta forma, propomos as situações de acordo com a natureza da variável e a organização dos dados (brutos ou agrupados). Também exemplificamos diversas possibilidades de apresentação das MD. Trata-se de uma reflexão teórica e de uma alternativa de sistematização do conhecimento produzido, até então, para o campo conceitual das MD na Educação Básica.

Este trabalho é composto por seis seções, contando com esta introdução. Na segunda seção, sintetizamos as principais ideias que estruturam a TCC. Na terceira seção, apresentamos a importância da compreensão conceitual das MD. Na quarta seção, definimos cada uma das MD contempladas nesta discussão teórica. Na quinta seção, apresentamos nossa proposta para

ISSN 2764-3158

a estrutura do campo conceitual das MD. Por fim, na sexta seção, tecemos nossas considerações finais em relação às discussões apresentadas.

2 A Teoria dos Campos Conceituais (TCC): uma breve revisão

A TCC é uma teoria de cunho cognitivista, desenvolvida para discutir e explicar o processo de formação de conceitos matemáticos no ambiente da sala de aula (Cazorla, Utsumi & Magina, 2021). A TCC foi proposta por Gérard Vergnaud (1933-2021), matemático, filósofo e psicólogo francês com notáveis contribuições para os campos da Psicologia Cognitiva e da Educação Matemática. No contexto da sala de aula, os conceitos de Esquema e de Zona de Desenvolvimento Proximal, respectivamente, são alguns dos aspectos que fazem da TCC uma teoria com fortes influências da Epistemologia Genética de Piaget (1966) e do socio construtivismo de Vygotsky (2001).

Os conceitos matemáticos são universais, porém as práticas pedagógicas diferem de acordo com fatores como, por exemplo, a formação docente e a realidade sociocultural das escolas. Sendo assim, a TCC considera que cada contexto ensinado por um determinado professor é único. Nessa perspectiva, Cazorla *et al.* (2023) ressaltam que o ensino de Matemática ocorre de diferentes formas e níveis e isso está relacionado com a cultura, com o avanço tecnológico da sociedade, com a concepção teórico-pedagógica da escola, e com a formação dos professores, em termos de conhecimentos específicos e pedagógicos. Segundo as autoras, todos esses fatores levam a objetivos e pontos de chegadas distintos.

Partindo da concepção epistemológica de que a aprendizagem do estudante é influenciada por fatores externos ao ambiente escolar, destacamos algumas inquietações de Vergnaud (1990), quando questiona sobre a natureza e a função de um novo conceito. Em sua obra, o autor interpela sobre como relacionar novas competências com concepções matemáticas e problemas práticos e teóricos, tornando-as valiosas e significativas na estruturação de novos conceitos.

Cedran e Kiouranis (2019, p. 65), sustentados teoricamente na TCC, destacam que um campo conceitual é constituído de "conceitos definidos socialmente e que são aplicados em diversos eventos por um indivíduo". Ao visitarem os principais fundamentos e perspectivas dessa teoria, as autoras definem que um campo conceitual pode ser pensado em termos de conhecimento científico e oriundo da produção intelectual de qualquer indivíduo, com base em suas experiências cotidianas.

Cazorla *et al.* (2023) ratificam que é de fundamental importância para o processo de ensino que o professor saiba escolher as situações que vão favorecer a formação de um determinado conceito por seus alunos e, ao mesmo tempo, que identifique os diversos conceitos que estão presentes em cada uma dessas situações-problema. Segundo as autoras, também é preciso observar as propriedades relevantes para a sua resolução (invariantes), pensando sobre as possíveis representações simbólicas, por meio das quais o aluno estabelece relações entre as situações e seus invariantes. Desse modo, as autoras sintetizam algumas ideias contidas em Vergnaud (1982, 1987, 1996), ao afirmarem que a aprendizagem de um conceito perpassa o seguinte tripé: um conjunto de situações (S), nas quais estão inseridas; um conjunto de invariantes (I) presentes nesse conjunto de situações; e o conjunto de representações simbólicas (R) que expressam os invariantes das situações. Essa tríade é geralmente descrita como: C = S, I, R.

Podemos observar que Vergnaud (1987) usa os termos "referente", "significado" e "significante" na perspectiva psicológica. Segundo esse autor, a Situação (S) está ligada ao "referente" do mundo real, das experiências das pessoas, ou seja, relaciona-se com a atividade

ISSN 2764-3158

e o problema no qual o objeto (conteúdo) matemático está presente. O "significado" tem relação com o conjunto de Invariantes (I), em outras palavras, são as propriedades que pertencem ao objeto matemático. Por fim, tem-se a Representação Simbólica (R), a qual, do ponto de vista epistemológico, é vista a partir da interação entre o conjunto de situações (S – referente) e o conjunto de invariantes (I – significado). As propriedades e os procedimentos podem estar implícitos para o estudante que esteja utilizando-os, constituindo-se o que Vergnaud (1996, 1998) denomina de "teorema-em-ação".

3 Qual a importância da compreensão conceitual das Medidas de Dispersão?

A compreensão da variabilidade tem amparado grandes avanços na indústria, agricultura, medicina e em outras áreas que lidam com a otimização de processos, em geral, controlando a variabilidade inerente a esses contextos. A compreensão da variabilidade a nível do ensino básico pode ser realizada com a observação e tratamento de variáveis próximas aos estudantes, como, por exemplo, o tamanho das camisetas do uniforme escolar que devem ser compradas. Esse exemplo mostrará aos estudantes que precisam considerar a distribuição de meninas e meninos, a idade, a altura, a massa corpórea, dentre outros fatores que interferem na constituição das pessoas. Nesse caso, se faz necessário levantar alguns dados para tomar decisões sobre a compra.

Garfield e Ben-Zvi (2005) destacam a importância das MD quando argumentam que os resumos estatísticos são mais que números propriamente ditos, pois revelam particularidades sobre os conjuntos de dados e, por isso, precisam ser compreendidos corretamente e satisfatoriamente. Para os autores, a amplitude indica a dispersão global dos dados, o desvio padrão dá indícios da concentração dos dados em torno da média e o intervalo interquartílico mostra o distanciamento da metade intermediária de uma distribuição.

É válido destacar também o papel da natureza das variáveis estatísticas para estudos que tratam da compreensão conceitual das MD, pois nem todas as medidas estatísticas e representações gráficas se aplicam a todos os tipos de variáveis. Essa observação coaduna Cazorla, Utsumi e Monteiro (2021), quando consideram a importância do reconhecimento da natureza da variável e dos dados gerados para refletirmos sobre as melhores decisões acerca da operacionalização e do tratamento estatístico dos dados.

A escolha do tipo de gráfico estatístico requer a compreensão de qual variável estatística dispomos. Sem esse conhecimento, podemos usar representações equivocadas e gerar dificuldades conceituais nos estudantes, sustentando conclusões superficiais ou inadequadas acerca dos dados. A representação gráfica pode permitir, por exemplo, a análise da distribuição dos dados, identificando como se concentram ou se "espalham" ao longo do intervalo dos valores da variável. É válido ressaltar que não há como compreender o comportamento da dispersão dos dados sem atentarmos para sua distribuição e as medidas de centro (Garfield & Ben-Zvi, 2005). Para esses autores, é preciso combinar resumos globais de dispersão e de centro para se ter uma visão mais abrangente, ao invés de focarmos a atenção em valores pontuais dos dados.

Os conceitos mobilizados para compreensão das MD são indissociáveis da noção de variabilidade ou variação dos dados. Nesse contexto, a literatura nem sempre apresenta consenso quanto ao uso desses dois termos. A "variabilidade" é vista como a forma substantiva do adjetivo "variável", significando que algo provavelmente irá mudar, enquanto a "variação" é tida como o substantivo que descreve o ato de variar (Reading & Shaughnessy, 2004). Neste trabalho, optamos por usá-los sem distinção, assim como Garfield *et al.* (2008) e Sousa Cazorla & Monteiro (2024).

ISSN 2764-3158

As práticas pedagógicas envolvendo conceitos estatísticos, quando desenvolvidas em contextos investigativos, auxiliam os estudantes a compreenderem como as variáveis são organizadas e como se comportam, motivando-os a perceber a existência de padrões e tornando-os sujeitos críticos e participativos nos diferentes espaços que habitam. No contexto investigativo, Rodríguez-Alveal e Maldonado-Fuentes (2023) recomendam a abordagem de problemas a partir de diferentes fontes exploratórias de variabilidade dos dados (real ou natural, induzida e amostral).

Adicionalmente, Franklin *et al*. (2007) indicam algumas das principais fontes de variabilidade em um conjunto de dados: a) *variabilidade de medição*: a medição realizada pelo mesmo indivíduo pode variar, seja porque o dispositivo de medição produz resultados duvidosos, como quando queremos medir uma grande distância usando uma régua pequena, ou porque os resultados obtidos pelo mesmo indivíduo podem variar em várias medições, ainda que o dispositivo seja preciso. Este é o caso da medida da pressão arterial, que difere de um momento para outro no mesmo indivíduo; b) *variabilidade natural*: é inerente à natureza da variável. As pessoas têm diferentes alturas e diferentes habilidades, opiniões e respostas emocionais; c) *variabilidade induzida*: se plantarmos uma saca de feijão em uma região e outra saca em outra região com clima diferente, as mudanças observadas na germinação das sementes e no crescimento das plantas de um local para outro podem ser explicadas pela variabilidade natural porque os locais não são os mesmos. No entanto, se forem usados adubos diferentes, as respostas de crescimento do feijão podem ser devido ao fertilizante ou, até mesmo, por causas desconhecidas. Os experimentos poderiam ajudar a determinar o motivo das diferenças. A comparação entre variabilidade natural e induzida é que tem permitido grandes avanços em diversas áreas, como na ciência médica, para se obterem conclusões sobre a eficácia de medicamentos; d) *variabilidade amostral*: em um processo político, é comum utilizarem-se resultados de uma amostra aleatória para estimar a proporção desconhecida de todos os eleitores que apoiam um candidato. No entanto, se uma segunda amostra for utilizada, mantendo o número de elementos, é quase certo não obtermos a mesma proporção. O valor da proporção amostral varia de amostra para amostra. Isso é chamado de variabilidade amostral.

As variáveis estatísticas são caracterizadas pela natureza não-determinística. Por isso, calcular e compreender conceitualmente a dispersão dos dados é importante para reconhecer que o conhecimento de Estatística pode auxiliar nas nossas escolhas diárias como, por exemplo, decidir abastecer no posto de combustível que mantém os preços mais estáveis, num certo período.

Os procedimentos matemáticos para o cálculo das MD são simples e demandam conhecimento apenas de operações básicas, tais como potenciação e radiciação. Porém, alguns autores salientam que a dispersão e as suas medidas têm apresentado dificuldades persistentes associadas aos processos de ensino e de aprendizagem (Moore, 1990; del Pino & Estepa, 2019). Assim, é necessário sugerir alternativas para apoiar a compreensão e apropriação desse campo conceitual da Estatística.

Os esforços contínuos para atualizar o ensino e o currículo da Estatística têm o potencial de promover novas aprendizagens a docentes, a partir da implementação de práticas pedagógicas inovadoras, bem como de tornar a aprendizagem mais estimulante para estudantes (Garfield, Le, Zieffler & Ben-Zvi, 2015). Para esses autores, esses esforços de atualização da Educação Estatística contribuem para uma nova geração de cidadãos, cujos conhecimentos são cruciais para uma sociedade marcada por um crescente fluxo de produção de dados. Em alguns casos, as propriedades matemáticas das MD permitem que sejam escolhidas como ferramental estatístico adequado para sintetizar o comportamento das informações produzidas,

BY NC SA
ISSN 2764-3158

principalmente quando ocorrem em contextos cuja característica dos dados assume a existência de variação.

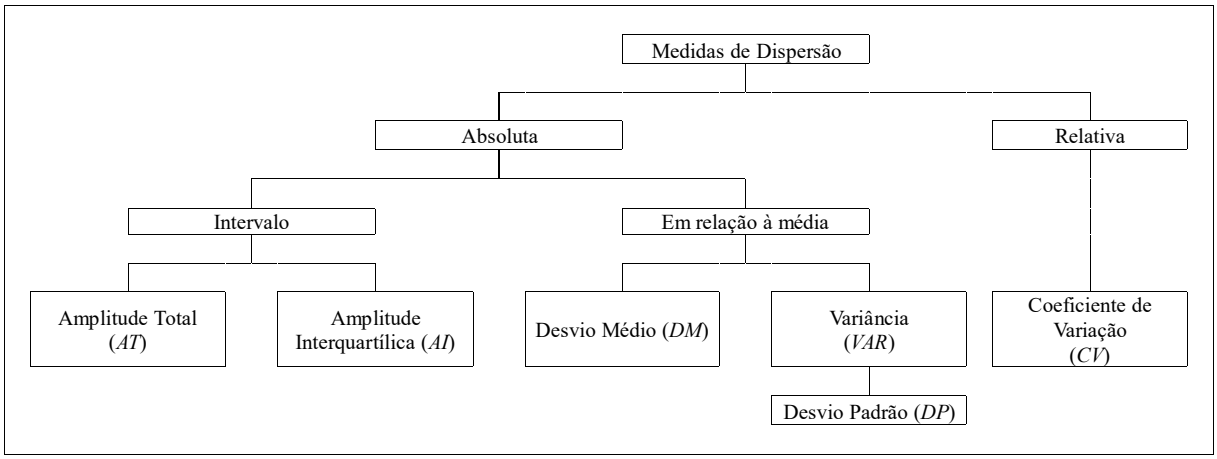
Na próxima seção, apresentamos brevemente a definição de cada uma das MD trabalhadas neste campo conceitual.

4 As Medidas de Dispersão – MD

As MD têm como objetivo mensurar a variabilidade de um conjunto de dados proveniente de uma variável estatística, indicando o grau de espalhamento ou concentração dos dados no intervalo dos possíveis valores que a variável pode tomar.

Nesta seção, apresentamos brevemente a definição das MD que são trabalhadas neste campo conceitual: Amplitude Total (*AT*), Amplitude Interquartílica (*AI*), Desvio Médio (*DM*), Variância (VAR), Desvio Padrão (*DP*) e, a medida relativa, o Coeficiente de Variação (*CV*), conforme Figura 1. Nesta seção, optou-se pelo uso da língua materna para descrição das MD, pois, de acordo com a TCC, há um conjunto de diferentes possíveis representações simbólicas (linguagem natural, gráficos e diagramas, sentenças formais, dentre outras) que podem ser usadas para indicar e representar os invariantes relacionados ao conceito (Vergnaud, 1982, 1987, 1996).

Figura 1: Medidas de Dispersão



Fonte: Construção dos autores

Amplitude (A) ou Amplitude Total (AT)

Uma variável quantitativa (X) toma valores em um intervalo, em geral descrito como: $[X_{min}; X_{max}]$, isto é, $X_{min} \leq X_i \leq X_{max}$, que contém 100% dos dados. Segundo Triola (2017), a *AT* de um conjunto de dados é a diferença entre os valores máximo (X_{max}) e mínimo (X_{min}). Trata-se de uma medida que fornece a primeira noção da variabilidade geral dos dados, mas não explica a distribuição interna dos dados.

Intervalo Interquartil (II) ou Amplitude Interquartílica (AI)

A medida *AI* é a distância entre os valores do terceiro quartil (Q_3) e do primeiro quartil (Q_1) de um conjunto de dados (Triola, 2017). Sendo assim, 50% dos dados centrais estão contidos no intervalo $[Q_1; Q_3]$.

As medidas *AT* e *AI* não informam a distribuição interna dos dados e não têm ponto de referência para medir a variabilidade dos dados. Por isso, é conveniente que sejam definidas medidas que sintetizem a variabilidade de um conjunto de observações, auxiliando, por

ISSN 2764-3158

exemplo, na comparação de conjuntos diferentes de valores, com base em algum critério estabelecido (Bussab & Morettin, 2017). Nesse sentido, a média tem sido utilizada como ponto de referência, pois é a única medida que está mais próxima de todos os valores da variável, a partir da qual se originam duas medidas: Desvio Médio e Variância.

Os desvios dos dados em relação à média nos dão uma ideia de como os valores da variável se afastam ou não em relação a essa medida. Contudo, não se pode construir nenhuma medida a partir deles, pois, para qualquer conjunto de dados, a soma de todos os desvios é zero (Bussab & Morettin, 2017), uma vez que os desvios positivos compensam os desvios negativos.

Na busca de contornar a soma zero dos desvios, algumas operações são propostas: encontrar o valor absoluto dos desvios, elevá-los ao quadrado e depois calcular a média.

Desvio Médio (DM)

O *DM* é obtido pela razão entre a soma da distância dos valores da variável em relação à média e a quantidade de dados (Triola, 2017). Logo, essa medida é não negativa e tomará o valor zero caso todos os valores coincidam, isto é, sejam constantes.

Variância (VAR) e Desvio Padrão (DP)

Outra forma de evitar que desvios positivos se anulem com desvios negativos é elevar ao quadrado todos esses desvios e depois dividir o resultado por n, encontrando a *VAR*. Dessa forma, definimos a *VAR como a média da soma dos quadrados dos desvios dos valores da variável em relação à média do conjunto de dados* (Bussab & Morettin, 2017). Na Educação Básica, utilizamos a variância apenas como uma forma de chegar ao *DP*. Além disso, a variância não tem referente no mundo real e não tem sentido tentar interpretá-la.

É válido ressaltar que, ao elevarmos os desvios ao quadrado, "inflacionamos" essas medidas, por isso revertemos o efeito dessa operação extraindo a raiz quadrada, encontrando o *DP*. Isto é, o Desvio Padrão é definido como *a raiz quadrada positiva da VAR* (Bussab & Morettin, 2017).

Todas as medidas, sejam de tendência central, dispersão ou os percentis, estão expressas nas unidades de medida da variável que as deu origem, porém a Variância é a única medida de dimensão igual ao quadrado da dimensão dos dados originais. Uma curiosidade que destacamos é que o *DM* é menor ou igual ao *DP*. Isso ocorre porque, ao elevarmos ao quadrado cada desvio e depois extrairmos a raiz quadrada, não conseguimos reverter essa "inflação".

O ensino do *DM* é mais pedagógico do que instrumental, pois o valor absoluto apresenta limitações matemáticas para a construção de outras medidas estatísticas, ao contrário da Variância. Todavia, sua interpretação tem um sentido intuitivo, pois é a distância média dos valores da variável em relação à média, já o Desvio Padrão é mais complexo de interpretar.

Coeficiente de variação (CV)

As medidas anteriormente apresentadas são denominadas de medidas de dispersão absoluta, pois levam em consideração apenas os valores dos dados. Essa é uma limitação muito grande, pois as variáveis podem ter grandezas diferentes como, por exemplo, a altura das pessoas (em metros ou centímetros) e a massa corpórea (em quilogramas); ou quando comparamos a distribuição de renda de países em unidades monetárias diferentes ou, ainda, quando analisamos uma única variável quantitativa pelas categorias de uma variável qualitativa como, por exemplo, a altura por sexo.

BY NC SA
ISSN 2764-3158

Nesses casos, o *CV*, que é uma medida de dispersão relativa, permite comparar a variação de diferentes conjuntos de dados, pois é definido pela razão entre o *DP* em relação à média (Bussab & Morettin, 2017), retirando o efeito das unidades de medida em que a variável foi coletada, sendo, portanto, adimensional. O *CV* é expresso em porcentagem e permite a comparação, o que auxilia na compreensão de quão homogênea/estável são as variáveis em estudo ou os grupos dentro de cada variável.

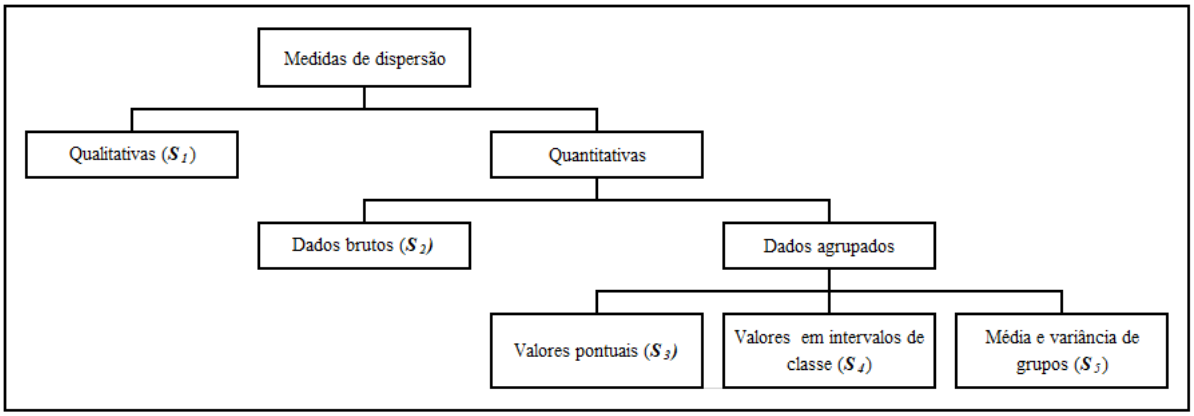
5 A Estrutura das Medidas de Dispersão: uma aproximação

Nesta seção, explicitamos como se estrutura a tríade (S, R, I) em termos de MD.

5.1 As Situações (S) presentes nas MD

O algoritmo para encontrar as MD depende da natureza da variável e de como os dados estão dispostos, se estão na sua forma original (dados brutos) ou agrupados, sejam em Tabelas de Distribuição de Frequência (TDF) ou em gráficos que ainda permitem o acesso aos valores da variável e à frequência absoluta ou relativa. Para Lautert e Cruz (2021), é necessário entender e vivenciar diversas situações para reconhecer e conseguir dar sentido aos conceitos. Assim, apresentamos as cinco situações para as MD, conforme Figura 2.

Figura 2: Situações para as Medidas de Dispersão



Fonte: Construção dos autores

S_1 (Variáveis qualitativas)

A única medida possível neste tipo de variável é a quantidade de categorias que a variável pode assumir. Nesse sentido, podemos ter variáveis dicotômicas, que apresentam apenas duas categorias, geralmente mutuamente excludentes; variáveis nominais ou ordinais previamente fixadas, que apresentam uma quantidade fixa de categorias como, por exemplo, cor favorita (verde, azul, amarelo e outra) ou classe social (baixa, média e alta, ou ainda, A, B, C, D e E); e as variáveis nominais abertas, cujas categorias são definidas à posteriori como, por exemplo, cor favorita. Neste caso, o respondente pode escolher uma ampla gama de cores que, a rigor, pode diferir dos outros respondentes.

A Situação ***S_1*** não se refere à distribuição dos dados de uma variável categórica, mas somente às categorias que pode assumir. Assim como quando estudamos as MTC e concluímos que a única medida possível é a Moda, como sendo a categoria que se repete com maior frequência, alguns autores defendem que existe a Mediana para variáveis ordinais (Cazorla *et al.*, 2021). Na nossa concepção, a única e possível MD seria a quantidade das categorias. O consenso sobre essa situação ainda não existe na Educação Estatística, pois os textos em Estatística excluem as MD para as variáveis qualitativas. Acreditamos que este é um ponto

ISSN 2764-3158

importante para ser debatido pelo GT12 no IX Sipem, para que a comunidade de educadores estatísticos possa se posicionar.

S_2 (Variáveis quantitativas com dados brutos)

No caso de variáveis quantitativas, cujos dados estão na sua forma bruta, a *AT* pode ser obtida pela diferença entre os valores máximo e mínimo. Para calcular a média, basta somar os valores da variável e dividir pela quantidade de dados. Para encontrar o Desvio Médio (*DM*) e a Variância (*VAR*), basta encontrar os desvios dos valores em relação à média, tomar o valor absoluto ou elevar ao quadrado cada um deles, em seguida somar e dividir pela quantidade de dados, respectivamente. Para encontrar o Desvio Padrão (*DP*), basta extrair a raiz quadrada da *VAR*.

S_3 (Variáveis quantitativas com dados agrupados em valores pontuais)

Este é o caso de variáveis discretas que tomam poucos valores agrupados em uma TDF (valores pontuais) ou em um gráfico que permita ainda a leitura das frequências (absolutas ou relativas) como, por exemplo, gráfico de hastes, de pontos ou de barras/colunas.

Neste caso, calculamos a média "ponderando" os valores pontuais (X_i) pela frequência absoluta (f_i) ou relativa (h_i) e dividindo o total obtido pela quantidade de dados (n). Em seguida, após calcular os desvios, os valores absolutos e os quadrados desses desvios devem ser ponderados pela frequência. A soma dos valores do produto gera o numerador do *DM* e da *VAR*, respectivamente. Para calculá-los, basta dividir esses valores por n.

S_4 (Variáveis quantitativas com dados agrupados em intervalos de classe)

No caso de variáveis discretas que tomam muitos valores ou variáveis contínuas, cujos dados se encontram agrupados em uma TDF em intervalos de classe, basta identificar a marca de classe (X_i), que é o ponto médio do intervalo de classe, e seguir os mesmos passos que na situação anterior (S_3).

Observamos que, neste caso, os valores das medidas (Média, *DM* e *DP*) serão por aproximação, pois não temos acesso aos dados brutos, e a marca de classe (ponto médio) representa todos os dados contidos nos intervalos.

Observamos que, quando apresentamos o *DM* e a *VAR* com dados agrupados, sejam com valores pontuais (S_3) ou em intervalos de classe (S_4), utilizamos a frequência absoluta como "fator de ponderação" e depois dividimos por n. Todavia, podemos utilizar a frequência relativa (em probabilidade) e, nesse caso, não precisaríamos dividir por n, uma vez que, $h_i = f_i/n$.

S_5 (Variáveis quantitativas com média e variância de grupos)

Em muitas situações, desejamos saber se o comportamento de uma variável quantitativa sofre a influência de uma variável qualitativa. No caso da média, para recompor a média geral a partir das médias parciais, basta ponderar as médias parciais pelo tamanho dos grupos, como apresentado por Cazorla *et al*. (2023). Todavia, isso não acontece com a variância, que se divide entre a variância dentro dos grupos e a variância entre grupos. Destaca-se que isso é muito complexo para ser ensinado na Educação Básica.

5.2 As Representações Simbólicas (R) presentes nas MD

As MD podem fazer uso da linguagem materna e das representações numérica e algébrica. Quanto à representação gráfica, observamos que as MD tomam um único valor, logo

BY NC SA
ISSN 2764-3158

é muito complexo tentar representá-las graficamente. No entanto, observamos a importância da representação gráfica da distribuição da variável, pois a sua configuração permite dar significado às MD.

A representação na língua materna, numérica e algébrica

Enfatizamos a importância do uso da linguagem adequada, conforme apresentada abaixo, uma vez que nem sempre é escolhida de forma correta.

Língua materna: a variância é a média dos quadrados dos desvios dos valores da variável em relação a sua média; o desvio padrão é a raiz quadrada da variância; o coeficiente de variação é a relação entre o desvio padrão e a média, é uma medida adimensional.

Representação numérica: M = 166,3 cm; VAR = 55,56 cm^2; DP = 7,45 cm.

Representação algébrica:

M: Média; $M = \frac{\sum_{i=1}^{n} x_i}{n}$;

DM: Desvio Médio; $DM = \frac{\sum_{i=1}^{n} |x_i - M|}{n}$; $DM = \frac{\sum_{i=1}^{k} f_i * |x_i - M|}{n}$; $DM = \sum_{i=1}^{k} h_i * |x_i - M|$;

VAR: Variância (S^2); $VAR = \frac{\sum_{i=1}^{n} (x_i - M)^2}{n}$; $VAR = \frac{\sum_{i=1}^{k} f_i * (x_i - M)^2}{n}$; $VAR = \sum_{i=1}^{k} h_i * (x_i - M)^2$;

DP: Desvio Padrão; $DP = \sqrt{VAR}$;

CV: Coeficiente de Variação; $CV = 100 * \frac{DP}{M}$.

A representação algébrica utiliza uma simbologia sofisticada para o nível da Educação Básica. Por essa razão, em muitos casos, tem-se evitado utilizá-la. Entretanto, neste trabalho se faz necessário explicitar a simbologia inerente às MD:

X: nome genérico da variável em estudo;

x_i: designa o i-ésimo valor/categoria da variável. No intervalo de classe, denota a marca de classe ou ponto médio da classe;

n: tamanho da amostra, quantidade de dados;

X_{min}: valor mínimo da variável quantitativa (no caso da variável qualitativa ordinal: categoria inferior);

X_{max}: valor máximo da variável quantitativa (no caso da variável qualitativa ordinal: categoria superior);

Intervalo de dados: $[x_{min}; x_{max}]$; AT: Amplitude Total; $AT = x_{max} - x_{min}$;

Q_1: Primeiro quartil; Q_3: Terceiro quartil;

AI: Amplitude Interquartílica; $AI = Q_3 - Q_1$;

k: quantidade de categorias (qualitativa), de valores pontuais (discreta que toma poucos valores), de intervalos de classes ou de grupos;

f_i: frequência absoluta, número de ocorrências da variável na i-ésima categoria, no i-ésimo valor ou no i-ésimo intervalo de classe, tal que $n = \sum_{i=1}^{k} f_i$;

n_i: tamanho do i-ésimo grupo, tal que $n = \sum_{i=1}^{k} n_i$;

h_i: frequência relativa, razão entre a frequência absoluta e a quantidade de dados ($h_i = f_i/n$), tal que $\sum_{i=1}^{k} h_i = 1$;

ISSN 2764-3158

A representação gráfica

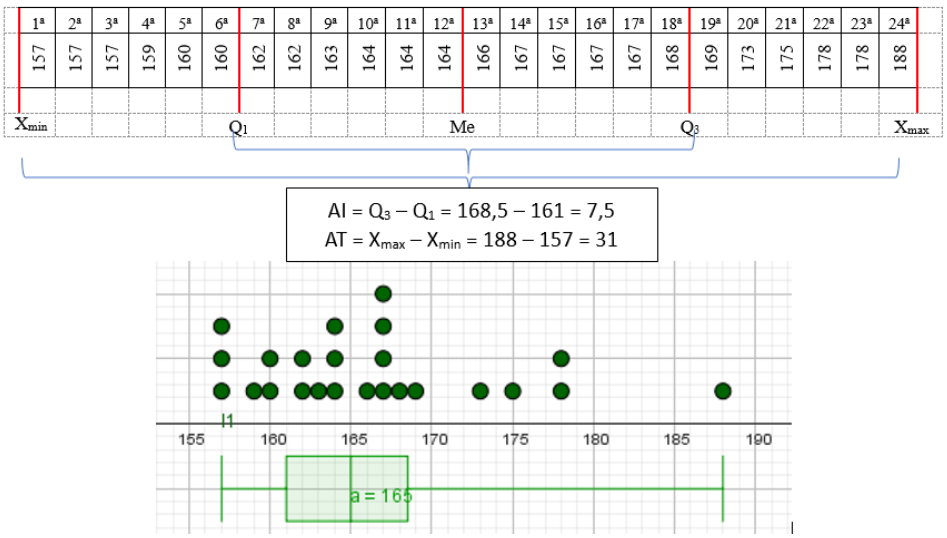
Como já salientamos, as MD tomam um único valor e, por isso, a sua representação gráfica pode não ter sentido. Porém, a representação da distribuição dos dados de forma adequada pode dar significado às MD. Isto é, o leitor deve compreender como os dados se distribuem para assim relacionar o comportamento da variabilidade dos dados com as medidas. Variáveis concentradas em torno da média também tendem a ter menor variabilidade do que aquelas que se espalham ao longo da amplitude da variável.

Os gráficos que podem ser utilizados para representar uma variável quantitativa, no contexto univariado, são: o gráfico de pontos, o histograma, o diagrama de ramo e folha, o de caixa (boxplot). Para variáveis discretas que tomam poucos valores, pode-se utilizar o gráfico de hastes e, eventualmente, o gráfico de barra/coluna. Quando a variável está indexada pelo tempo, é possível utilizar o gráfico de colunas e o de linhas. Neste caso, a variável está representada na ordenada. E, quando uma variável quantitativa é analisada pelas categorias de uma variável qualitativa, indica-se utilizar o diagrama de ramo e folha opostas, ou boxplot superpostos.

Esses gráficos apresentam a distribuição dos dados e suas principais características. A partir da configuração do formato da distribuição da variável, podemos intuir e verificar o grau de dispersão dos dados e sua aderência aos valores das medidas.

Na Figura 3, apresentamos a distribuição das alturas de 24 estudantes (Santana, 2020), cujos dados brutos são apresentados em ordem crescente, sinalizando os valores mínimo e máximo, o primeiro e o terceiro quartil e a mediana. Nessa representação, assinalamos a amplitude total e a amplitude interquartílica. Essa estratégia é útil para mostrar aos estudantes a posição e encontrar o valor dos quartis. Contudo, essa representação é enganosa, pois não leva em consideração os valores da variável, mas apenas a ordem. Nesse sentido, é importante o uso do diagrama de pontos que corrige essa falsa impressão.

Figura 3: Estatísticas e representações da altura (em cm) de 24 estudantes



$X_{min} = 157$
$X_{max} = 188$
AT = 31
M = 166,3
Me = 165
$Q_1 = 161$
$Q_3 = 168,5$
AI = 7,5
DM = 5,61
DP = 7,45
CV = 4,48%

Fonte: Construção dos autores a partir dos dados de Santana (2020)

5.3 Os invariantes (I) presentes nas MD

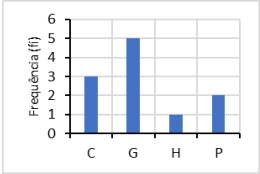
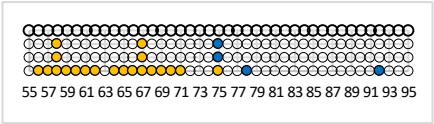
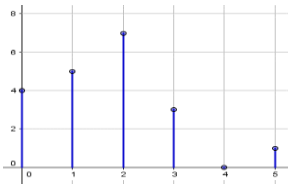
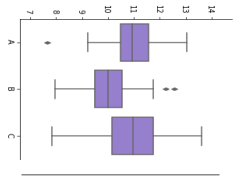
Os *Invariantes (I)* são formados pela rede de conceitos que ancoram o campo conceitual, bem como as propriedades das MD que unificam e distinguem as situações que lhe dão sustentação. Desse modo, a seguir apresentamos as propriedades das MD: a) todas as MD

ISSN 2764-3158

sempre assumem valores não negativos e serão maiores quanto maior for a variabilidade dos dados; b) as MD só tomarão valores iguais a zero quando não existe variabilidade, isto é, se os dados forem todos iguais; c) o *DP* será sempre maior ou igual ao *DM*; d) se se somar ou subtrair uma constante à variável, a *AT*, a *AI*, o *DP* e o *DM* não se alteram; e) se se multiplicar ou dividir a variável por uma constante, a *AT*, o *DM* e o *DP* ficam multiplicados ou divididos por essa constante; f) a *VAR* é a única medida que pode ser recomposta a partir das variâncias dos grupos e da variância entre grupos, princípio utilizado na construção dos testes de hipóteses para igualdade de médias, regressão linear etc.

Na Figura 4, apresentamos uma síntese da nossa proposta. Observamos que a *AT*, a *AI* e o *CV* não mudam ao longo das situações. Já o Q_1, Q_3, *DM* e o *DP* dependem da configuração dos dados, observando que apenas no caso de os dados estarem em intervalos de classe que iremos necessitar de processos de interpolação para calcular Q_1 e Q_3.

Figura 4: Situações para as MD

Tipo	Forma	Situações	Representação gráfica dos dados
Qualitativas	Dados brutos ou agrupados	***S₁*** (x_i: categorias) Quantidade de categorias da variável	
Quantitativas	Dados brutos	***S₂*** (x_i: valores) $DM = \frac{\sum_{i=1}^{n} \lvert x_i - M \rvert}{n}$ $DP = \sqrt{\frac{\sum_{i=1}^{n} (x_i - M)^2}{n}}$	Diagrama de pontos 
	Discreta que toma poucos valores, com dados agrupados	***S₃*** (x_i: valores pontuais) $DM = \frac{\sum_{i=1}^{k} f_i * \lvert x_i - M \rvert}{n}$ $DP = \sqrt{\frac{\sum_{i=1}^{n} f_i * (x_i - M)^2}{n}}$	Diagrama de hastes/bastão 
	Discreta que toma muitos valores; Contínuas com dados agrupados em intervalos de classe	***S₄*** (x_i: marca de classe) $DM = \frac{\sum_{i=1}^{k} f_i * \lvert x_i - M \rvert}{n}$ $DP = \sqrt{\frac{\sum_{i=1}^{n} f_i * (x_i - M)^2}{n}}$	Histograma
	Variância entre grupos	***S₅ (****média e variância de grupos****)*** *VAR = VAR-Entre +VAR-Dentro*	Boxplot (Análise de Variância) 

Fonte: Construção dos autores

6 Considerações finais

Este artigo dá continuidade à construção do campo conceitual das MTC e teve como objetivo apresentar uma primeira aproximação da estrutura do campo conceitual das MD. Sua

ISSN 2764-3158

originalidade reside na sistematização das situações que podem auxiliar estudantes e professores na compreensão dessas medidas.

Com base na TCC, elencamos um conjunto de cinco situações que podem ser abordadas com estudantes da Educação Básica. Os tipos de representações são definidos a partir de cada contexto, em que os estudantes devem escolher conforme o nível escolar, a quantidade, a forma de organização e a natureza dos dados. Além disso, as MD trazem diferentes explicações sobre a distribuição dos dados, desde aspectos mais gerais quanto padrões internos de variabilidade dos dados, portanto, conhecer as suas propriedades e o ferramental apropriado para cada caso ampliará a formação do conceito de dispersão, constituindo-se o tripé proposto neste trabalho.

Insistimos que, pela própria natureza da Estatística, é importante compreender como a variável se distribui, o que implica envolver um número razoável de dados e o uso de gráficos adequados, a fim de explorar todo o seu potencial. Por isso, é preciso que o estudante se defronte com as diversas situações, a fim de que amplie seus repertórios e saiba reconhecer qual é o algoritmo adequado para cada situação.

Apesar de reconhecermos que houve avanços do ensino de Estatística na Educação Básica brasileira, ainda há um caminho longo a ser percorrido para desenvolver situações de ensino do conceito de variabilidade e suas medidas. Um dos maiores desafios encontrados, em especial para abordar as MD, é a quantidade de dados que pode tornar a atividade maçante e improdutiva. Por essa razão, a estratégia de apresentar os conceitos com poucos dados para mostrar o algoritmo e as noções básicas parece adequada, embora não seja suficiente. Sugerimos que os estudantes experimentem a recolha de dados desde os anos iniciais do Ensino Fundamental, utilizando materiais concretos manipuláveis, reconhecendo a natureza das variáveis, organizando-as em tabelas e gráficos. Essas abordagens educacionais podem contribuir para a compreensão da variabilidade, a partir da representação dos dados coletados e, posteriormente, realizar os cálculos e conferir quão apurada é essa intuição.

Nossas reflexões, ao revisitarmos as situações das MTC, nos possibilitaram ter uma perspectiva mais ampla das situações das MD e sugerirmos que se faz necessária a construção do campo conceitual da Estatística Descritiva. Esse campo englobaria a natureza das variáveis, os gráficos e as tabelas estatísticas. Porém, isso será feito posteriormente.

O escopo deste trabalho foi estruturar as situações de ensino que, posteriormente, irão auxiliar professores no desenvolvimento de atividades em sala de aula. Afinal, trata-se de um artigo focado em uma reflexão teórica e uma sistematização do conhecimento produzido na área de Educação Estatística. Espera-se que este texto auxilie professores, estudantes e pesquisadores interessados na compreensão das MD, sob a ótica da Teoria dos Campos Conceituais, e que contribua para o desenvolvimento do pensamento e do letramento estatístico dos estudantes da Educação Básica.

Referências

Barbosa, G. C.; Santos, S. S. & Lopes, C. E. (2019). Um cenário das pesquisas em Educação Estatística no XII ENEM. *Revista de Ensino de Ciências e Matemática*, *10*(6), 319-339.

Brasil. Ministério da Educação e do Desporto. Secretaria de Educação Fundamental. (1997). *Parâmetros Curriculares Nacionais: Matemática*. Brasília, DF.

Brasil. Ministério da Educação. Secretaria de Educação Básica. (2017). *Base Nacional Comum Curricular: Educação Infantil e Ensino Fundamental*. Brasília, DF.

Bussab, W. O. & Morettin, P. A. (2017). *Estatística Básica*. (9. ed.). São Paulo, SP: Saraiva.

CC BY NC SA
ISSN 2764-3158

Cazorla, I. & Utsumi, M. (2010). Reflexões sobre o ensino de Estatística na Educação Básica. In: Cazorla, I. M. & Santana, E. (Orgs.). *Do tratamento da informação ao letramento estatístico* (pp. 113-144). Itabuna, BA: Via Litterarum.

Cazorla, I. M.; Utsumi, M. C. & Magina, S. M. (2021). Revisitando o conceito de mediana na perspectiva dos campos conceituais: uma aproximação teórica. In: *Anais do 8° Seminário Internacional de Pesquisa em Educação Matemática* (pp. 2556-2569). Uberlândia, MG.

Cazorla, I. M.; Utsumi, M. C. & Magina, S. M. (2023). The conceptual field of measures of central tendency: A first approximation. *International Electronic Journal of Mathematics Education*, *18*(4), 1-14.

Cazorla, I. M.; Utsumi, M. C. & Monteiro, C. E. F. (2021). Dos dados brutos à informação: o papel das técnicas transnumerativas no ensino de Estatística. *Educação Matemática Pesquisa*, *23*(4), 109-139.

Cedran, D. P. & Kiouranis, N. M. M. (2019). Teoria dos Campos Conceituais: visitando seus principais fundamentos e perspectivas para o ensino de ciências. *ACTIO*, *4*(1), p. 63-86.

Cobb, G. & Moore, D. (1997). Mathematics, statistics and teaching. *American Mathematical Monthly*, *104*(9), 801-823.

del Pino, J. & Estepa, A. (2019). Análisis de la enseñanza de las medidas de dispersión em libros de texto de educación secundaria. *Avances de Investigación en Educación Matemática*, *16*, 86-102.

Franklin, C.; Kader, G.; Mewborn, D.; Moreno, J.; Peck, R.; Perry, M. & Scheaffer, R. (2007). *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report: a Pre-K-12 Curriculum Framework*. Alexandria, VA: American Statistical Association.

Garfield, J. & Ben-Zvi, D. (2005). A framework for teaching and assessing reasoning about variability. *Statistics Education Research Journal*, *4*(1), 92-99.

Garfield, J.; Ben-Zvi, D.; Chance, B.; Medina, E.; Roseth, C. & Zieffler, A. (2008). *Developing students' statistical reasoning: connecting research and teaching practice*. Dordrecht, The Netherlands: Springer.

Garfield, J.; Le, L.; Zieffler, A. & Ben-Zvi, D. (2015). Developing students' reasoning about samples and sampling variability as a path to expert statistical thinking. *Educational Studies in Mathematics*, *88*(3), 327-342.

Isoda, M.; Chitmun, S. & González, O. (2018). Japanese and Thai senior high school mathematics teachers' knowledge of variability. *Statistics Education Research Journal*, *17*(2), 196-215.

Lautert, S. & Cruz, T. V. Q. F. (2021). Os conceitos estatísticos mobilizados por universitários em situações envolvendo medidas de tendência central e medidas de dispersão. In: *Anais do 8° Seminário Internacional de Pesquisa em Educação Matemática* (pp. 1909-1922). Uberlândia, MG.

Lima, R. F.; Paula, M. C. & Giordano, C. C. (2022). A Educação Estatística na Educação Infantil e nos Anos Iniciais do Ensino Fundamental: identificações em revistas brasileiras. *Revista de Educação Matemática*, *19*(01), 01-19.

Moore, D. S. (1990). Uncertainty. In: Steen, L. A. (Ed.). *On the Shoulders of Giants: New Approaches to Numeracy.* (pp. 95-137). Washington, DC: National Academy Press.

Piaget, J. (1966). *A epistemologia genética* [The genetic epistemology]. Martins Fontes.

ISSN 2764-3158

Reading, C. & Shaughnessy, J. M. (2004). Reasoning about variation. In: Ben-Zvi, D.; Garfield, J. (Eds.). *The challenge of developing statistical literacy, reasoning and thinking*. (pp. 201-226). Dordrecht, The Netherlands: Kluwer.

Rodríguez-Alveal, F. & Maldonado-Fuentes, A. C. (2023). Tipología de las preguntas sobre variabilidad en los textos escolares y su relación con la alfabetización y pensamiento estadístico. *Uniciencia*, *37*(1), 1-19.

Samá, S. & Silva, R. C. S. (2023). Rede de conversação universidade e escola: discussões e reflexões sobre o ensino de Estatística. *Revista de Educação PUC-Campinas*, *28*(e238489), 1-16.

Santana, C. V. (2020). *Relações entre variáveis estatísticas na contextualização e apropriação da função afim*. 2020. 163f. Dissertação (Mestrado em Educação Matemática). Universidade Estadual de Santa Cruz. Ilhéus, BA.

Schreiber, K. P. & Porciúncula, M. (2019). Mapeamento das pesquisas sobre Educação Estatística na Biblioteca Digital Brasileira de Teses e Dissertações: um olhar para a formação do professor de matemática. *Revista Eletrônica de Educação Matemática*, *14*(Edição Especial Educação Estatística), 1-17.

Shaughnessy, J. M. (2007). Research on statistics learning and reasoning. In: Lester Jr., F. K. (Ed.). *Second handbook of research on mathematics teaching and learning* (v. 2, pp. 957-1009). Charlotte, NC: Information Age Publishers.

Sousa, E. P.; Cazorla, I. M. & Monteiro, C. E. F. (2024). Abordagens conceituais sobre variabilidade estatística em pesquisas brasileiras: uma revisão sistemática de literatura. *Ensino Em Re-Vista*, *31*(e2024 – 13), 1-24.

Triola, M. F. (2017). *Introdução à Estatística*. (12. ed.). Rio de Janeiro, RJ: LTC.

Vergnaud, G. (1982). A classification of cognitive tasks and operations of thought involved in addition and subtraction problems. In: Carpenter, T. & Moser, J. (Eds.). *Addition and subtraction: A cognitive perspective*. (pp. 39-59). Lawrense Erlbaun.

Vergnaud, G. (1987). Conclusion. In: Janvier, C. (Ed.). *Problem of representation in the teaching and learning of mathematics*. (pp. 227-232). LEA Publishers.

Vergnaud, G. (1990). Epistemology and psychology of mathematics education. In: Nesher, P. & Kilpatrick, J. (Eds.). *Cognition and practice*. (pp. 14-30). Cambridge University Press.

Vergnaud, G. (1996). A teoria dos campos conceituais. In: Brun, J. (Ed.). *Didáctica das matemáticas* (pp. 155-191). Instituto Piaget.

Vergnaud, G. (1998). Comprehensive theory of representation for mathematics education. *Journal of Mathematics Behavior*, *2*(17), 167-181.

Vygotsky, L. (2001). *A construção do pensamento e da linguagem*. Tradução de P. Bezerra. São Paulo, SP: Martins Fontes.

Watson, J.; Fitzallen, N.; Wright, S. & Kelly, B. (2023). Characterizing the ways in which young students recognise, describe, explain, and employ variation when analysing data in a STEM context. *International Journal of Science and Mathematics Education*, *21*, 943-968.

CC BY NC SA
ISSN 2764-3158