



Conhecimento de estudantes do 3º ano do Ensino Médio brasileiro sobre Amostragem e Curva Normal: uma abordagem por meio da Inferência Informal

Knowledge of 3rd year Brazilian high School students on sampling and normal curve: an approach through informal inference

André Fellipe Queiroz Araújo¹
Diêgo Bezerra de Melo Maciel²

Resumo: O presente estudo objetivou investigar os conhecimentos de estudantes do 3º ano do Ensino Médio brasileiro (16-17 anos) sobre os conceitos de Amostragem e Curva Normal. Para tanto, sua fundamentação teórica está pautada na concepção do Letramento Estatístico e Probabilístico e na Inferência Informal. Essa pesquisa possui abordagem qualitativa, na qual 72 estudantes responderam a um questionário diagnóstico. Os resultados apontaram um domínio sobre os conceitos de amostra e população. Porém, foram observadas lacunas acerca da representatividade e variabilidade amostrais, métodos de amostragem e sobre o modelo da Curva Normal. Diante disso, reforçamos a importância da ampliação de processos de ensino e aprendizagem que contemplem conceitos de Amostragem e Curva Normal alinhados com a Inferência Informal.

Palavras-chave: Educação Estatística. Amostragem. Curva Normal. Inferência Informal. Letramento Estatístico e Probabilístico.

Abstract: The current study aimed to investigate the knowledge of 3rd year Brazilian high school students (ages 16-17) regarding the concepts of sampling and the normal curve. The theoretical foundation is based on the concepts of statistical and probability literacy and informal inference. This research adopts a qualitative approach, with 72 students responding to a diagnostic questionnaire. The results indicated a solid understanding of the concepts of sample and population. However, gaps were observed in areas such as sample representativeness and variability, sampling methods, and the normal curve model. In light of these findings, we emphasize the importance of enhancing teaching and learning processes that address sampling and normal curve concepts aligned with informal inference.

Keywords: Statistical education. Sampling. Normal curve. Informal inference. Statistical and probability literacy.

1 Introdução

Em nosso cotidiano, constantemente nos deparamos com situações que demandam a interpretação de conhecimentos estatísticos e probabilísticos. Por exemplo, durante os processos de investigação estatística, uma das etapas envolve a definição de se a pesquisa será censitária ou amostral. Outra etapa importante é a análise e interpretação dos dados, a qual contempla conceitos e modelos que permitem a realização de inferências estatísticas, ainda que de maneira aproximada (Guimarães & Gitirana, 2013).

Nesse contexto, Gal (2002, 2005) argumenta que para uma cidadania plena é crucial que as pessoas tenham habilidades em Estatística e Probabilidade para interpretar os fenômenos sociais de forma crítica. Logo, é essencial que na Educação Básica sejam sistematizadas

¹ Secretaria de Educação do Estado de Pernambuco • Recife, PE — Brasil • ✉ andrefellipeq93@gmail.com • ORCID <https://orcid.org/0000-0002-7060-0621>

² Universidade Federal de Pernambuco • Recife, PE, Brasil • ✉ diego.mmaciell@ufpe.br • ORCID <https://orcid.org/0000-0002-1198-854X>

possibilidades para o desenvolvimento de competências e habilidades que permitam ao estudante lidar com essas demandas cidadãs e não apenas realizar cálculos matemáticos relativos a esses dois campos de conhecimento. Não se trata de negligenciar as habilidades matemáticas, pelo contrário, mas é preciso que o ensino da Estatística/Probabilidade avance para além das tradicionais memorizações de fórmulas e aplicação de algoritmos. Tem-se, então, uma perspectiva de ensino-aprendizagem voltada para o Letramento Estatístico e Probabilístico.

Nessa direção, a compreensão dos conceitos de Amostragem e Curva Normal são fundamentais para uma formação estudantil letrada, na medida em que tais conceitos estão presentes em muitos fenômenos e práticas do cotidiano. Em linhas gerais, a Amostragem abrange as técnicas para selecionar amostras que sejam representativas de suas populações (Bayer, Echevest, Bittencourt & Rocha, 2005; Triola, 2008). Já a Curva Normal é considerada o principal modelo probabilístico para análise de dados (Batanero, Tauber & Sánchez, 2004). Embora esses conceitos ainda sejam pouco explorados no ensino básico no Brasil, são essenciais para uma compreensão sólida dos dados e para a formulação de informações estatísticas úteis para tomadas de decisão (Luna & Guimarães, 2021).

No que diz respeito a essa temática, a Base Nacional Comum Curricular – BNCC (Brasil, 2018), principal documento orientador para o ensino no Brasil, sugere que durante o Ensino Médio (16-17 anos), os estudantes devem realizar pesquisas estatísticas amostrais e analisar criticamente essas amostras. Além disso, recomenda a resolução de problemas que envolvem medidas de tendência central e de dispersão, bem como a identificação de espaços amostrais em eventos aleatórios para calcular probabilidades em diferentes contextos, o que inclui implicitamente o estudo da Curva Normal.

Naturalmente, todos esses conceitos estão articulados com o ato de inferir (aproximar) alguma característica populacional com base no estudo da amostra. Trata-se da Inferência Estatística. Tradicionalmente, a Inferência Estatística faz parte da divisão curricular moderna da Estatística, sendo trabalhada no Ensino Superior. Sua base teórico-metodológica está assentada em conceitos probabilísticos mais complexos, os quais não estão contemplados na BNCC (Brasil, 2018). Nesse cenário, uma alternativa pedagógica que vem ganhando espaço nas salas de aula da Educação Básica é a Inferência Informal. De acordo com Batanero, Begué e Valenzuela-Ruiz (2022), a Inferência Informal almeja reduzir ao mínimo o instrumental algébrico e conceitual requerido pela Inferência Estatística, permitindo o trabalho docente em uma base matemática e teórica mais simples, com menor nível de formalização.

Em paralelo, diversos estudos vêm apontando certas dificuldades de estudantes do Ensino Médio com os conceitos relacionados à Amostragem e/ou Curva Normal. Por exemplo, Reyes, Ruz, Molina-Portillo e Contreras (2019) mostraram que grande parte de 148 estudantes analisados possuem dificuldades em refletir sobre as implicações da variação do tamanho amostral, o que pode levá-los a erros nas interpretações em análises estatísticas. Jacob e Doerr (2013) mostraram dificuldade, por parte da maioria dos estudantes investigados, sobre a variabilidade no contexto da distribuição amostral. Em um estudo sobre Curva Normal, Herrera, Monrot e Hernández (2018) observaram que a maioria dos estudantes soube aplicar corretamente a fórmula e a tabela da Curva Normal Padronizada para o cálculo da área e consequentemente a probabilidade de intervalos sob a Curva, mas em termos conceituais, exibiram dificuldades em reconhecer o significado e algumas etapas dos procedimentos e a importância da padronização para análises estatísticas em uma distribuição de dados.

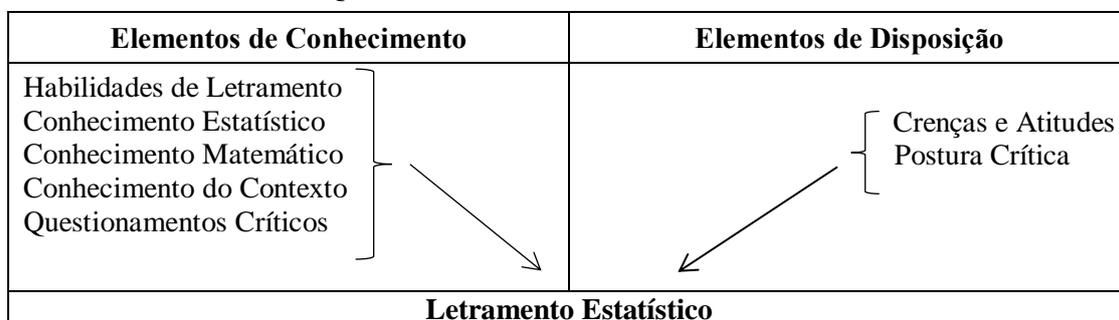
Diante do exposto, o presente estudo investigou o nível de conhecimento de estudantes do 3º ano do Ensino Médio (16-17 anos) sobre a Amostragem e a Curva Normal por meio da

Inferência Informal e foi amparado na perspectiva do Letramento proposta por Gal (2002, 2005). Acreditamos que a realização dessa investigação poderá ampliar as possibilidades didático-pedagógicas para o seu processo de ensino e aprendizagem.

2 Letramento Estatístico e Probabilístico

Na perspectiva de Gal (2002, 2005), o indivíduo é considerado letrado em Estatística e Probabilidade quando detém as competências para interpretar e avaliar criticamente uma informação estatística/probabilística em qualquer contexto e discutir ou comunicar suas reações, crenças e opiniões frente às informações e dados desses campos de conhecimento. O autor enfatiza que esse letramento pode auxiliar os cidadãos a responder demandas sociais em várias perspectivas. O modelo do Letramento Estatístico proposto pelo autor (Quadro 1) contempla dois componentes que se relacionam entre si, a saber: Elementos de Conhecimento e Elementos de Disposição.

Quadro 1: Modelo de Letramento Estatístico



Fonte: Gal (2002).

Explorando cada componente do modelo supracitado, destacamos, em linhas gerais, que os Elementos de Conhecimento estão relacionados com a capacidade de o indivíduo ler, interpretar e analisar de forma crítica uma informação ou dado estatístico. Esse campo contempla cinco elementos que não são independentes e se inter-relacionam (Gal, 2002).

As Habilidades de Letramento dizem respeito às habilidades gerais que os indivíduos devem possuir para se comunicar de forma oral ou escrita e compreender as informações estatísticas de diferentes níveis de complexidade. Já o Conhecimento Estatístico contempla o domínio de conceitos e ideias básicas relacionadas à Estatística Descritiva e à Estatística Inferencial, envolvendo a ideia de sentido numérico dos dados, familiaridade com conceitos das medidas de centralidade e dispersão e as representações gráficas e tabulares (Gal, 2002).

O Conhecimento Matemático, por sua vez, é responsável por dar suporte necessário para o Conhecimento Estatístico e está direcionado para a compreensão da ideia de número, seja ele natural, fracionário, decimal ou percentual com diferentes significados e procedimentos de cálculos relativos às medidas e conceitos. Em continuidade, o Conhecimento de Contexto também é necessário, tendo em vista que toda informação estatística está relacionada a um determinado contexto. Assim, quando o indivíduo o conhece, utiliza esse conhecimento para dar significado e compreender tal informação. Como último Elemento de Conhecimento, o autor aponta que os Questionamentos Críticos compreendem a capacidade de o indivíduo analisar de forma crítica as informações estatísticas veiculadas por diversos meios, avaliando a origem, construção e formulação, como também o propósito e a veracidade dessas informações (Gal, 2002).

Por fim, os Elementos de Disposição referem-se à Postura Crítica, disposição, Crenças e concepções dos indivíduos frente à determinada informação estatística. Gal (2002) enfatiza

que o indivíduo letrado estatisticamente deve se portar criticamente ao lidar com informações relativas ao campo estatístico, confrontando suas crenças com dados e refletindo sobre suas concepções prévias com uma Postura Crítica diante dos mesmos.

De modo análogo, Gal (2005) também apresenta o modelo de Letramento Probabilístico, sistematizado no quadro adiante, que contempla dois componentes essenciais: Os *Elementos Cognitivos* e os *Elementos de Disposição*.

Quadro 2: Modelo do Letramento Probabilístico

Elementos Cognitivos	Elementos de Disposição
1 - Grandes Ideias	1 - Postura Crítica
2 - Figurando Probabilidades	2 - Crenças e Atitudes
3 - Idiomas	3 - Sentimentos pessoais sobre a incerteza e o risco
4 - Contexto	
5 - Questões Críticas	

Fonte: Gal (2005).

Gal (2005) destaca que noções de Probabilidade, incerteza e risco estão presentes em várias informações divulgadas pela mídia e, tanto os cidadãos comuns, quanto os profissionais, necessitam calcular estimativas da Probabilidade de determinados eventos mesmo que de modo informal. Paralelamente, necessitam interpretar e compreender situações que envolvem conceitos probabilísticos para realizarem a tomada de decisões. Porém, o autor argumenta que o pensamento e o comportamento dos indivíduos em situações probabilísticas são afetados por múltiplas bases de conhecimento e disposições. Nesse sentido, o primeiro elemento, *Grandes Ideias*, está relacionado com o domínio de tópicos fundamentais para a compreensão do conceito de Probabilidade, do qual enfatizamos a variação, aleatoriedade e a incerteza. No que diz respeito ao elemento *Figurando Probabilidades*, Gal (2005) argumenta que os indivíduos devem compreender as diferentes formas para o cálculo de Probabilidade de eventos para poder entender e comunicar sobre as informações e dados probabilísticos. De forma paralela, o elemento cognitivo *Idiomas* se faz necessário para o desenvolvimento da linguagem probabilística, pois em nossa sociedade precisamos representar, comunicar e dialogar sobre conceitos e resultados probabilísticos.

Em continuidade, o quarto elemento cognitivo, *Contexto*, está relacionado com a compreensão do papel e dos significados das informações e dados probabilísticos em diferentes contextos, isto é, nas situações em que as noções de acaso e de probabilidade podem aparecer. O último elemento cognitivo do Letramento Probabilístico, *Questões Críticas*, está relacionado com a capacidade dos indivíduos desenvolverem um pensamento crítico sobre as informações e dados probabilísticos. Por fim, consideram-se os *Elementos de Disposição* que, por sua vez, são formados pela *Postura Crítica*, *Crenças e Atitudes* e os *Sentimentos pessoais sobre a incerteza e o risco*. Tais elementos estão relacionados com aspectos pessoais e afetivos dos indivíduos, bem como à disposição para refletir e analisar criticamente as informações relacionadas ao campo probabilístico (Gal, 2005).

Diante do exposto, o autor concebe que os tais Letramentos devem ser construídos pelos indivíduos desde o início de sua escolarização até a vida adulta, tendo em vista que os estudantes já se deparam com situações envolvendo conceitos estatísticos e probabilísticos no começo de suas formações escolares.

3 Amostragem e Curva Normal: Uma abordagem por meio da Inferência Informal

O cerne da investigação de fenômenos (físicos ou sociais), apoiado na perspectiva do Letramento, requer a compreensão do ciclo de investigação estatística como um processo de

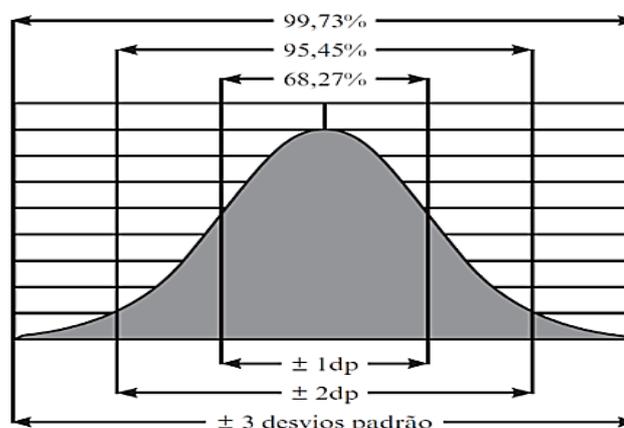
fazer inferências (Cazorla, Utsumi & Monteiro, 2021). Assim, segundo Cazorla, Utsumi e Oliveira (2020, p. 2), “o foco não deve ser nos dados em si, mas nas características e nos processos mais gerais que os criaram. Esse processo é de fato inferencial”. Nisso, o estudo da amostra e de suas técnicas de obtenção são nevrálgicos na Inferência Estatística. Assim, a partir da coleta amostral obtida via processos condizentes com o contexto da análise, são feitas generalizações probabilísticas para determinado aspecto populacional. Quase sempre, essas generalizações estão apoiadas no modelo de distribuição Normal, ou algum outro modelo de probabilidade derivado dele.

Nesse contexto, a Amostragem é responsável pelo desenvolvimento de “*técnicas para seleção das unidades populacionais que formarão a amostra*” (Bayer *et al.*, 2005, p. 2, grifos nossos). O termo população se refere ao conjunto de todos os elementos (indivíduos, animais, objetos) que possuem pelos menos uma característica em comum. Já a amostra é definida como um subconjunto da população (Triola, 2008).

Para que uma amostra seja representativa e permita generalizações probabilísticas para a população, é importante considerar a variabilidade, o tamanho da amostra e o método de Amostragem. Populações mais homogêneas (menor variabilidade) requerem amostras menores, enquanto populações mais heterogêneas (maior variabilidade) exigem amostras maiores para capturar todas as características da população. Já os métodos de Amostragem se ocupam do processo de seleção dos elementos que irão compor a amostra, eles são divididos em dois grupos: não-probabilísticos e probabilísticos. Nos métodos não-probabilísticos, a Probabilidade de seleção dos elementos é desconhecida, podendo alguns ter Probabilidade nula de serem escolhidos. Os métodos não-probabilísticos são: Amostragem por conveniência, julgamento, bola de neve, quotas e resposta voluntária. Nos métodos probabilísticos, a Probabilidade de seleção é conhecida e maior que zero, garantindo representatividade e imparcialidade. Os métodos probabilísticos são: Amostragem aleatória simples, sistemática, estratificada e por conglomerados.

Além disso, a probabilidade possibilita a formulação de técnicas e modelos matemáticos que permitem a caracterização de uma população. A Distribuição Normal, ou Curva Normal, é um modelo essencial na Estatística Inferencial, definida pela média e pelo desvio-padrão dos dados (Batanero *et al.*, 2004). A Curva Normal é simétrica em relação à média e os valores de média, moda e mediana coincidem. A área sob a curva é sempre igual a 1, representando 100% dos dados, e a Probabilidade de uma variável assumir um valor entre dois pontos é dada pela área sob a curva entre esses pontos (Figura 1).

Figura 1: Representação da Curva Normal



Fonte: São Paulo (2014).

O leitor mais atento pode notar que foram colocados diversos conceitos relacionados com Amostragem e Curva Normal, mas com pouca ênfase no arcabouço matemático subjacente. Por exemplo, apresentamos a distribuição normal sem a necessidade de focar na sua função densidade de Probabilidade, a qual apresenta um aspecto pouco convidativo para a maioria dos estudantes de nível médio. Desse modo, a “informalidade” adotada na apresentação dos conceitos possibilita ao docente abordar conteúdos mais complexos (tais como Amostragem e Curva Normal), por meio de um trato condizente com o nível curricular matemático e estatístico da Educação Básica. Isso representa uma das possibilidades da chamada Inferência Informal.

A partir do trabalho seminal de Zieffler, Garfield, Delmas & Reading (2008) tem-se uma sistematização inicial da Inferência Informal. Dentre outros pontos, os autores consideram: i) análise de possíveis características de uma população baseada na observação amostral; ii) avaliação do nível de adequação de determinada amostra, tendo em vista certa expectativa de resultado; e iii) realização de julgamentos, articulações, afirmações ou previsões populacionais baseadas em amostras, mas não utilizando procedimentos formais da Estatística.

A partir de então, diversos autores vêm trazendo uma variedade de propostas para simplificar o ensino da inferência e/ou conceitos relacionados a ela. Assim, para Makar e Rubin (2009), a Inferência Informal é tomada como um processo criativo e indutivo, no qual a experimentação, via observação de padrões, assume o lugar das abordagens formais tipicamente ensinadas no Ensino Superior. Dessa forma, tem-se um processo fundamentado, porém informal, de criar ou testar generalizações a partir dos dados, ou seja, não necessariamente por meio do instrumental estatístico-matemático. Batanero e Borovcnik (2016) têm denominado essa perspectiva como aproximações informais ao ensino (da inferência). Mais recentemente, Cazorla e Giordano (2021) tomam a Inferência Informal para pensar sobre quais respostas serão encontradas com as variáveis levantadas, aspectos relacionados com o delineamento amostral, leitura e interpretação de gráficos, tabelas e medidas estatísticas.

Percebe-se que grande parte das propostas de abordagem da Inferência Informal são fundadas por meio da articulação entre o pensamento inferencial e sua relação com os processos de investigação estatística (Wild & Pfannkuch, 1999). Assim, há um processo de articulação natural entre a investigação e o ato de inferir. Desse modo, sob um cenário investigativo, precisa-se de um tema – ou Contexto - que propicie problemas plausíveis, originadores de questões. As respostas para tais questões – obtidas a partir de dados amostrais, emergem com linguagem probabilística, pois são feitas possíveis generalizações para as características populacionais de interesse. É aí que surge o imbricamento com a Inferência Informal.

Diante disso, pretendemos avaliar o conhecimento dos estudantes sobre Amostragem e Curva Normal, sob a ótica da Inferência informal, direcionando o foco pedagógico para competências e habilidades relacionadas com os preceitos do Letramento Estatístico e Probabilístico presentes em Gal (2002, 2005). Com essa estratégia, construímos um teste de diagnose, a partir de um conjunto de seis questões, que foi trabalhado em uma turma de estudantes do Ensino Médio brasileiro. A próxima seção detalha os aspectos operacionais do estudo e as questões propostas.

4 Método

Em termos metodológicos, esse estudo utiliza uma abordagem qualitativa, definida por Oliveira (2011) como um processo de reflexão e análise detalhada da realidade. Nesse sentido, buscamos investigar os conhecimentos de estudantes do 3º ano do Ensino Médio (16-17 anos) sobre a Amostragem e a Curva Normal, a partir de uma abordagem dos conteúdos sintonizada

com as ideias da Inferência Informal e do Letramento Estatístico e Probabilístico (Gal, 2002, 2005). O universo de participantes dessa pesquisa foi composto por 72 estudantes da referida etapa de escolarização e vinculado à rede pública do estado de Pernambuco, localizado na região Nordeste do Brasil. Nossa escolha por estudantes desse nível de escolaridade está pautada nas orientações curriculares apresentadas pela BNCC (Brasil, 2018), no qual contempla competências e habilidades voltadas para a Estatística e a Probabilidade e, mais especificamente, de forma explícita e implícita, para os fundamentos da Amostragem e da Curva Normal. Para identificar esse conhecimento, utilizamos, como instrumento de coleta de dados um questionário diagnóstico composto por 6 (seis) questões. Mediante as respostas apresentadas pelos participantes, criamos uma escala de gradação para avaliar o desempenho, contendo as seguintes categorias de respostas: i) Adequada; ii) Parcialmente Adequada; e iii) Inadequada. O critério balizador para classificar as respostas fornecidas levou em conta o objetivo e a habilidade explorada em cada questão. A seguir, apresentamos o detalhamento de cada uma.

A primeira questão (Figura 2) objetiva investigar a habilidade dos estudantes em identificar a população e a respectiva amostra. Para isso, foram apresentados três cenários de pesquisas estatísticas. Nossa abordagem segue um viés conceitual, não-paramétrico, focando na capacidade do estudante em aplicar os conceitos de população e amostra em cada contexto apresentado. Assim, não priorizamos questões tipicamente colocadas nos manuais de estatística, as quais trabalham com situações de estimativas paramétricas acerca da média, variância, proporção ou outra quantidade de interesse.

Figura 2: Questão 1 do teste de diagnose

Questão 1		
Para cada uma das pesquisas, a seguir, identifique a população e a amostra correspondente:		
Pesquisa A	Qual é a população?	Qual é a amostra?
A empresa de serviço de proteção ao crédito (SPC-Brasil) fez uma pesquisa com 600 brasileiros para analisar o perfil do consumidor consciente.		
Pesquisa B	Qual é a população?	Qual é a amostra?
Uma pesquisa realizada pelo Sesc analisou 2024 blogs brasileiros com o intuito de verificar as linguagens artísticas utilizadas para a apresentação dos conteúdos.		
Pesquisa C	Qual é a população?	Qual é a amostra?
14 macacos foram submetidos a uma pesquisa que tinha por finalidade testar a eficácia de uma vacina.		

Fonte: Dados da Pesquisa (2024).

A segunda questão (Figura 3) envolve um cenário de realização de pesquisa amostral com uma situação muito familiar aos estudantes: o uso de celular na sala de aula. A questão é composta por dois itens. O primeiro analisa o conhecimento sobre a variabilidade (natural) da amostra, questionando sobre quais características devem ser consideradas para garantir a representatividade. Já o segundo aborda os métodos de Amostragem, mirando acerca das formas pelas quais a amostra deveria ser selecionada. Assim como ocorreu na questão anterior,

não fizemos um tratamento paramétrico. Dessa forma, os estudantes não precisariam conhecer a distribuição amostral das variáveis ou cálculos da margem de erro. Assim, deveriam considerar apenas as respostas apenas os contextos apresentados nos itens da questão.

Figura 3: Questão 2 do teste de diagnose

Questão 2

A direção de uma escola quer saber os hábitos da utilização de celular dos estudantes com a finalidade de discutir com eles o uso consciente. Para isso, ela fará uma pesquisa estatística e decidiu que não irá entrevistar todos, mas, sim, uma amostra de estudantes da escola.

a) Quais características dos estudantes a direção poderia utilizar para selecionar uma amostra de modo que os resultados representem toda a escola? Justifique a escolha.

b) Como a direção poderia selecionar a amostra de estudantes para representar toda a escola?

Fonte: Dados da Pesquisa (2024).

A terceira questão (Figura 4), por sua vez, analisa a capacidade de julgamento dos estudantes com relação à adequação de métodos de Amostragem para se garantir a representatividade amostral em contextos de pesquisa. Para isso, construímos cinco cenários diferentes, em que o método de seleção das amostras variava, mas com base num mesmo objetivo: identificar o percentual de estudantes que desejam estudar francês. Apesar do contexto envolver uma situação paramétrica (estimador da proporção), a habilidade necessária não esteve voltada para tal, pelo contrário. O objetivo pedagógico da questão possui, de certa forma, caráter subjetivo, pois guarda relação com crenças estudantis acerca da viabilidade de determinados tipos de Amostragem. Por isso mesmo não foi exigido conhecimento sobre o cálculo do tamanho amostral, o qual já é fornecido pela questão. Além disso, coloca-se, mais uma vez, um cenário próximo ao estudante (escola de idiomas), mas com um enfoque econômico, na medida em que a obtenção da amostra irá subsidiar uma tomada de decisão empresarial.

Figura 4: Questão 3 do teste de diagnose

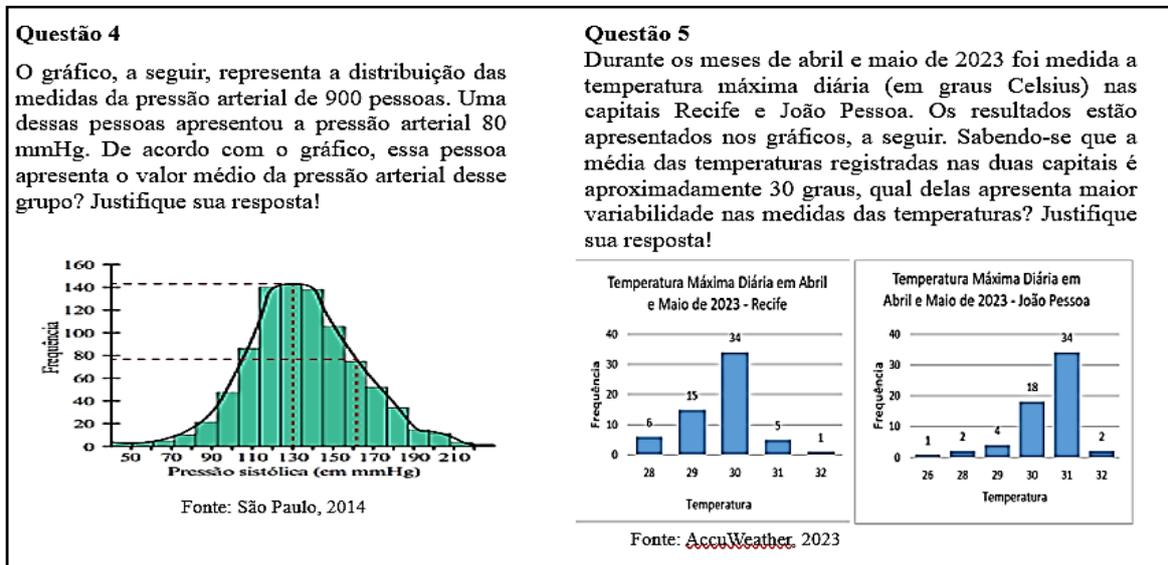
Questão 3	Amostras	Adequada ou Inadequada?	Por quê?
Os donos de uma escola de idiomas (inglês e espanhol) desejam ampliar os serviços oferecidos. Para isso, querem saber o percentual de seus estudantes que tem interesse em se matricular em aulas de francês. Diante disso, cinco pesquisadores entrevistaram 80 estudantes de um total de 800 da escola de idiomas. Porém, cada pesquisador selecionou uma amostra diferente. Julgue cada uma das amostras, a seguir, em adequada ou inadequada para representar o resultado da pesquisa em questão. Além disso, não esqueça de justificar sua resposta!	Amostra 1: Eduardo entrevistou 80 estudantes que estavam passando no refeitório.		
	Amostra 2: Sabendo que a escola tinha o mesmo número de homens e mulheres, Luana entrevistou 80 estudantes, sendo 40 homens e 40 mulheres, escolhendo-os, aleatoriamente, de cada curso de idiomas que a escola possui atualmente (inglês e espanhol).		
	Amostra 3: Natália tinha o nome de todos os 800 estudantes, colocou-os em um chapéu e então tirou, aleatoriamente, 80 deles para realizar a entrevista.		
	Amostra 4: Renato enviou um questionário para todos os estudantes e depois usou os 80 primeiros que foram devolvidos a ele.		
	Amostra 5: Arthur enumerou e ordenou todos os 800 alunos da escola em uma lista. Em seguida, a cada intervalo de 10 alunos, ele selecionou o sexto, totalizando 80 entrevistados.		

Fonte: Dados da pesquisa (2024).

As questões 4, 5 e 6 (quatro, cinco e seis) tratam de aspectos conceituais relacionados com o modelo de distribuição normal e suas propriedades básicas. No caso das duas primeiras questões, não há menção explícita ao modelo normal. O objetivo em ambas é investigar se os

estudantes identificavam o referido modelo probabilístico por meio de uma inspeção gráfica. Caso positivo, deveriam considerar as características centrais e de variabilidade do modelo. Não é necessário realizar nenhum cálculo matemático. As questões 4 e 5 estão reunidas na Figura 5.

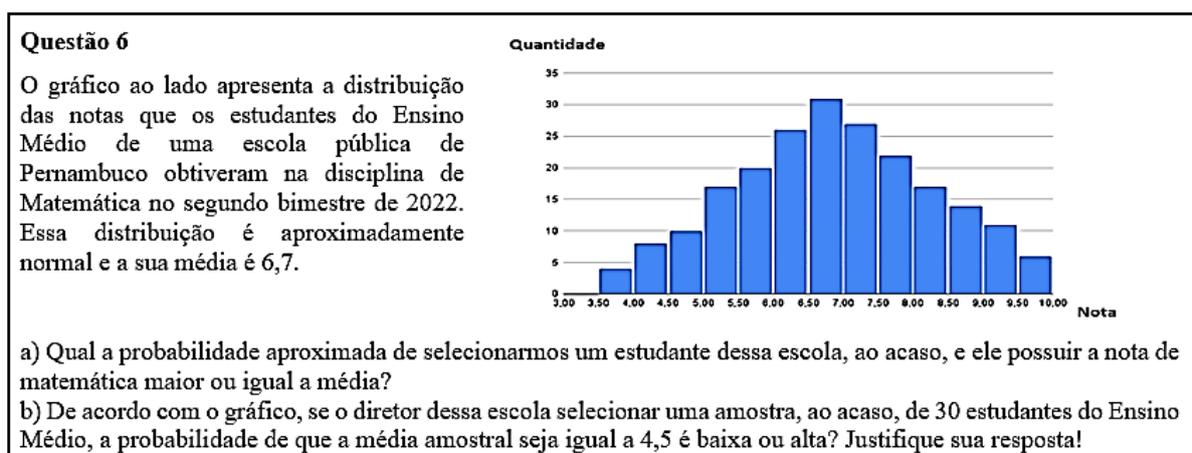
Figura 5: Questões 4 e 5 do teste de diagnose



Fonte: Dados da Pesquisa (2024).

Por fim, a sexta questão (Figura 6) analisa o conceito de Probabilidade associado à simetria da Curva Normal. Para tanto, não é necessário apresentar a função densidade do modelo, nem fazer usos de tabelas estatísticas. Basta considerar que a simetria divide os dados em duas partes iguais, tendo a média como ponto referencial. Logo, a Probabilidade de qualquer valor abaixo ou acima da média é de 50%. Ainda nessa questão foi abordada a relação entre o valor da média amostral com a populacional, mas sem recorrer à complexidade teórica do Teorema do Limite Central, base da inferência clássica.

Figura 6: Questão 6 do teste de diagnose



Fonte: Dados da Pesquisa (2024).

5 Resultados e Discussão sobre o desempenho estudantil

Esta seção analisa as respostas dos estudantes em cada questão, tomando por base a escala de desempenho apresentada anteriormente. De início, a Tabela 1 apresenta o percentual

das respostas classificadas como adequadas (corretas) na questão 1, fazendo um recorte por cada parte da questão (população e amostra). Os dados evidenciaram um bom domínio do Conhecimento Estatístico e de contexto (Gal, 2022) em todos os cenários apresentados. No entanto, destaca-se o menor percentual de acerto (72,5%) na identificação da população pesquisa C. Esse fato pode denotar que alguns estudantes ainda relacionam o fato de população se referir apenas a pessoas, conforme apontado em Gomes (2013).

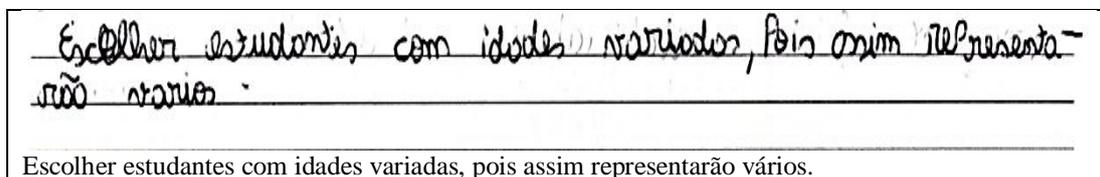
Tabela 1: Percentual de resposta adequada por item da Questão 1

	População	Amostra
Pesquisa A	76,5	80,5
Pesquisa B	74,0	82,0
Pesquisa C	72,5	83,5

Fonte: Dados da Pesquisa (2024).

No que se refere à questão 2, em seu primeiro item, constatamos 35% de respostas adequadas, enquanto as demais se classificaram como inadequadas. Nessa questão, esperávamos que os estudantes pudessem mencionar algumas características importantes para a seleção dos elementos da amostra, como o sexo, ano escolar, idade, se possui celular, dentre outros. Constatamos que a maioria daqueles que responderam de forma inadequada indicou apenas uma característica a ser considerada (Figura 7), o que para as pesquisas amostrais não garante a representatividade estatística. Logo, verificamos fragilidades no Conhecimento Estatístico (Gal, 2002) inerentes à variabilidade amostral relacionada com a representação amostral.

Figura 7: Resposta inadequada de um estudante para a 2ª (A) questão



Escolher estudantes com idades variadas, pois assim representarão vários.

Fonte: Dados da Pesquisa (2024).

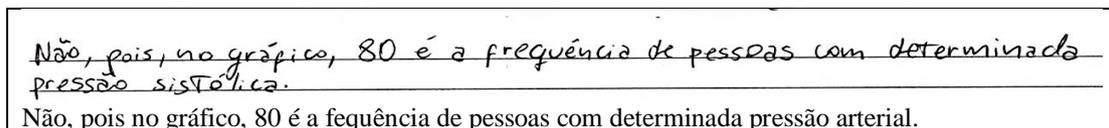
Já no segundo item, os estudantes deveriam indicar um método probabilístico para seleção da amostra. Porém, apenas 10% responderam adequadamente, ao passo que os demais apresentaram respostas inadequadas por não contemplarem elementos da Amostragem em suas respostas. Isso evidencia falta de Conhecimento Estatístico (Gal, 2002) relativo aos métodos de seleção amostral. Percebemos que, em linhas gerais, os estudantes se preocuparam em indicar a quantidade de elementos, mas não um método de seleção específico.

Na questão 3 foram apresentados cinco cenários dos quais as amostras dos cenários um e quatro deveriam ser consideradas como inadequadas para a representatividade, pois envolviam métodos não probabilísticos. No entanto, os estudantes exibiram dificuldades nesse julgamento, na medida em que os percentuais de respostas corretas foram de 15% e 14%, respectivamente. Por outro lado, as amostras dos cenários dois, três e cinco deveriam ser classificadas como adequadas por serem baseadas em métodos probabilísticos. Porém, os percentuais de acerto foram de 34%, 17% e 12%, em respectivo. A partir desses dados, constatamos um desempenho insatisfatório nesse conteúdo, indicando que a maioria dos estudantes analisados desconhecem a validade dos métodos probabilísticos de Amostragem, sobrepondo suas crenças ao Conhecimento Estatístico/probabilístico necessário (Gal, 2002, 2005).

Já a quarta questão teve o objetivo de analisar, a partir da situação apresentada, se os estudantes reconheceriam o gráfico abordado como uma Curva Normal, identificariam a média como o valor que está no centro da curva e julgariam se a medida anunciada corresponderia à média da distribuição. Nessa questão, a maioria dos estudantes (53%) responderam à questão de forma parcialmente adequada, enquanto 17% responderam adequadamente e 30% de modo inadequado. No que se refere a categoria mais frequente, constatamos que os estudantes mencionaram a medida apresentada como não sendo a média da distribuição, mas justificaram com critérios alheios, os quais incluíram a indicação de que a média seria um valor maior, sem defini-lo; como também, de que não era possível, sem especificar todos os valores e suas respectivas frequências, calcular a média.

Isso evidencia que esses estudantes acreditam na média como uma medida restrita ao seu processo do algoritmo, o qual consiste em somar os valores da variável e dividir pela quantidade de dados. Logo, a maioria apresentou fragilidades no Conhecimento Estatístico e Matemático (Gal, 2002). Esse resultado nos mostrou que a maior parte dos estudantes não compreende as características principais da Curva Normal, como o formato do gráfico e as medidas de centralidade e dispersão. Essa conclusão pode ser exemplificada na Figura 8.

Figura 8: Resposta parcialmente adequada de um estudante para a 4ª questão



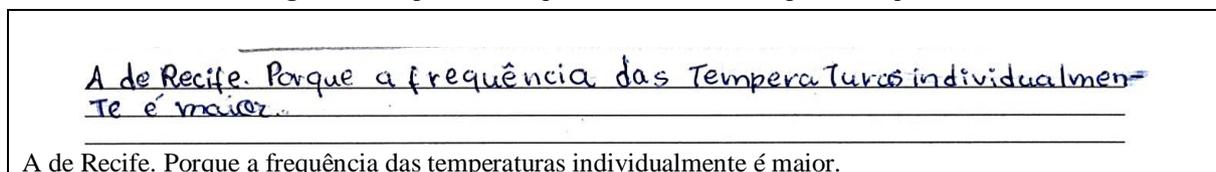
Não, pois, no gráfico, 80 é a frequência de pessoas com determinada pressão sistólica.

Não, pois no gráfico, 80 é a frequência de pessoas com determinada pressão arterial.

Fonte: Dados da Pesquisa (2024).

Em continuidade, a quinta questão foi direcionada para analisar o conhecimento dos estudantes sobre o conceito de variabilidade/dispersão das medidas apresentadas em gráficos que se aproximam de uma Curva Normal. Assim, esperava-se que os estudantes mencionassem que a distribuição mais assimétrica possuiria maior variabilidade; o oposto deveria ser identificado para uma distribuição mais próxima da normal. No entanto, apenas 22% respondeu adequadamente, enquanto os demais tiveram suas respostas classificadas como inadequadas, sobretudo, por atribuírem maior variabilidade ao conjunto de dados mais próximos ao modelo normal. Essa conclusão vai ao encontro dos resultados observados na quarta questão, na qual os estudantes também demonstraram lacunas no Conhecimento Matemático e Estatístico (Gal, 2002) relacionados com a análise do gráfico da Curva Normal. Uma das respostas inadequadas da questão 5 é apresentada na Figura 9.

Figura 9: Resposta inadequada de um estudante para a 5ª questão



A de Recife. Porque a frequência das Temperaturas individualmente é maior.

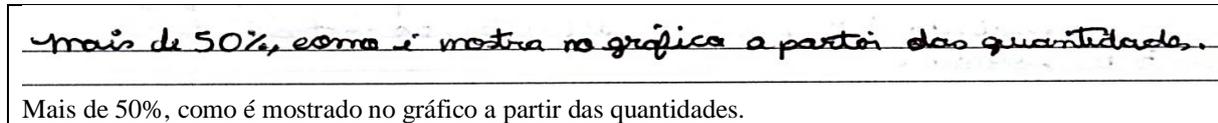
A de Recife. Porque a frequência das temperaturas individualmente é maior.

Fonte: Dados da Pesquisa (2024).

Finalmente, a questão 6 (seis), em seu primeiro item, mostrou que os estudantes apresentaram dificuldades em associar o cálculo de Probabilidades com as características de simetria da distribuição Normal. Assim, constatamos que apenas 10% responderam adequadamente, ao passo que os demais (90%) apresentaram respostas inadequadas por, principalmente, considerar valores de Probabilidade não condizentes com o nível de simetria do modelo normal. Esse resultado reforça os fatos apresentados nas questões 4 e 5 (quatro e cinco), ou seja, que os participantes apresentam lacunas no que diz respeito aos Elementos de Conhecimento (Gal, 2002, 2005) sobre a Curva Normal e suas propriedades, incluindo, agora,

a estimação de Probabilidades. Exibimos uma das respostas considerada inadequada para essa questão (Figura 10).

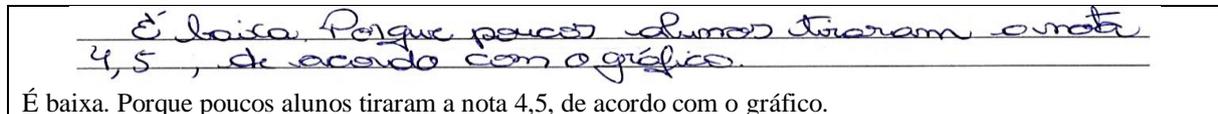
Figura 10: Resposta inadequada de um estudante para a 6ª questão, item a



Fonte: Dados da Pesquisa (2024).

Já o item b dessa mesma questão colocou em cena uma aproximação informal ao Teorema do Limite Central (TCL) sem a necessidade do rigor teórico-metodológico relacionado. Dessa forma, esperávamos que os estudantes diferenciasssem as Probabilidades associadas à média amostral daquelas relacionadas com a variável aleatória de interesse (“nota”), ainda que de maneira informal. Todavia, esse item exibiu o maior percentual de respostas inadequadas (97,5%) dentre todas as questões do estudo. Os estudantes, nitidamente, confundiram a distribuição da variável aleatória “nota” (apresentada no gráfico da questão) com a distribuição amostral das médias das notas, a qual é formulada a partir da Amostragem dos dados (Figura 11). Provavelmente, a crença de que a média amostral é um valor fixo, sem distribuição de Probabilidade associada, pode ajudar a entender o baixo desempenho estudantil nesse item. A média é tipicamente apresentada na Educação Básica a partir de uma fórmula matemática, sem contexto estatístico. Negligencia-se, assim, que no âmbito amostral, a média assume denotação de variável aleatória, possuindo não um valor fixo, mas um conjunto de valores com determinadas Probabilidades associadas. Isso evidencia as crenças dos estudantes relacionadas com o Conhecimento Estatístico e Probabilístico (Gal, 2002, 2005) relacionados ao tema.

Figura 11: Resposta inadequada de um estudante para a 6ª questão, item b



Fonte: Dados da Pesquisa (2024).

6 Considerações finais

No presente estudo tivemos o objetivo de investigar os conhecimentos de estudantes do 3º ano do Ensino Médio brasileiro (16-17 anos) sobre Amostragem e Curva Normal por meio da Inferência Informal e amparado na perspectiva do Letramento proposta por Gal (2002, 2005). Nossa estratégia consistiu em construir um conjunto de seis questões, mirando as competências e habilidades previstas na BNCC, as quais estão alinhadas com os preceitos do Letramento Estatístico e Probabilístico (Gal, 2002, 2005).

A análise dos resultados indicou que a maioria desses estudantes demonstrou domínio sobre o significado da palavra amostra e sua relação com a população. Entretanto, os indivíduos não apresentaram domínio conceitual para além disso. Assim, foram evidenciadas lacunas em tópicos fundamentais da Amostragem, como a compreensão da representatividade da amostra, abarcando a variabilidade amostral e os métodos de Amostragem, de igual modo, o modelo da Curva Normal, contemplando sua representação gráfica e propriedades, a exemplo das medidas de centralidade e dispersão e o cálculo de Probabilidades, ainda que de maneira informal.

Além disso, notou-se que a maior deficiência e dificuldade apresentada pelos estudantes guardou relação com conceitos associados ao TCL, especialmente a compreensão de que a média amostral é uma variável aleatória, possuindo uma distribuição de Probabilidade

associada. Esse diagnóstico, construído a partir dos preceitos da Inferência Informal, indicou que estudantes do Ensino Médio no Brasil necessitam avançar no conhecimento sobre tópicos da Estatística Inferencial, na medida em que não possuem o nível de Letramento Estatístico/Probabilístico recomendado pela BNCC. Logo, reforçamos a importância de se oportunizar um processo de ensino e aprendizagem atrelado ao ciclo investigativo que contemple os fundamentos da Amostragem concatenados ao modelo da Curva Normal. Acreditamos que essa estratégia irá possibilitar o entendimento conceitual desses conteúdos, além da compreensão acerca da natureza aleatória da média amostral. Defendemos, ainda, a viabilidade de uma abordagem pedagógica com utilização de um instrumental estatístico e matemático menos complexo, em sintonia com a Inferência Informal.

Por fim, destacamos a necessidade de investimentos em formações continuadas de professores da Educação Básica, com vistas à ampliação de seus conhecimentos e, por conseguinte, a potencialização do processo de ensino e aprendizagem da Amostragem e Curva Normal de forma articulada em sala de aula.

Referências

- Bayer, A.; Echeveste, S.; Bittencourt, H. R. & Rocha, J. (2005). Preparação do formando em Matemática-licenciatura plena para lecionar estatística no Ensino Fundamental e Médio. In: *Encontro Nacional de Pesquisa em Educação em Ciências* (pp. 1-11). Bauru, SP.
- Batanero, C. & Borovcnik, M. (2016). *Statistics and probability in high school*. Rotterdam: Sense.
- Batanero, C.; Begué, N. & Valenzuela-Ruiz, S. M. (2022). Modelización de la variabilidad muestral en diferentes metodologías de la inferencia. *Revista de Educación Estadística*, 1(1), 1-22.
- Batanero, C.; Tauber, L. & Sánchez, V. (2004). Students' reasoning about the normal distribution. In: D. Ben-Zvi & J. Garfield (Eds.). *The challenge of developing statistical literacy, reasoning and thinking* (pp. 257–276). Dordrecht, Holanda: Springer.
- Brasil. Ministério da Educação. Secretaria de Educação Básica. (2018). *Base Nacional Comum Curricular*. Brasília, DF.
- Cazorla, I. & Giordano, C. (2021). O papel do letramento estatístico na implementação dos temas contemporâneos transversais da BNCC. In: C. Monteiro & L. Carvalho (Orgs.). *Temas emergentes em letramento estatístico* (pp. 88–111). Recife, PE: UFPE.
- 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.
- Cazorla, I., Utsumi, M. & Oliveira, T. (2020). Reflexões sobre o ensino de tabelas de dupla entrada a partir do raciocínio inferencial informal. *Hipátia*, 5(2), 246-270.
- Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International statistical review*, 70(1), 1-25.
- Gal, I. (2005). Towards 'probability literacy' for all citizens. In: G. A. Jones (Ed.). *Exploring probability in school: Challenges for teaching and learning* (pp. 39-63). New York, USA: Springer.
- Gomes, T. M. (2013). *O todo é a soma das partes, mas uma parte representa o todo? Compreensão de Estudantes do 5º e 9º ano sobre Amostragem*. 2012. 110f. Dissertação



- (Mestrado em Educação Matemática e Tecnológica). Universidade Federal de Pernambuco. Recife, PE.
- Guimarães, G. L. & Gitirana, V. (2013). Estatística no Ensino Fundamental: a pesquisa como eixo estruturador. In: Borba, R. E. & Monteiro, C. E. (Org.). *Processos de ensino e aprendizagem em Educação Matemática*. Recife: UFPE.
- Herrera, J. S.; Monrot, J. C. V. & Hernández, U. S. (2018). Un acercamiento a la metodología lesson study para la enseñanza de la distribución normal. In: *Anais do XXII Investigación en Educación Matemática* (pp. 525-534). Gijón, Espanha.
- Jacob, B. & Doerr, H. M. (2013). Students' informal inferential reasoning when working with the sampling distribution. In: *Anais do 8º Congress of European Research in Mathematics Education*. Istanbul, Turquia.
- Luna, L. C. & Guimarães, G. L. (2021) Aprendizagem de amostragem nos PCN e na BNCR e a influência nos livros didáticos. *Revista de Educação, Ciências e Matemática*, 11(1), 1-20.
- Makar, K. & Rubin, A. (2019). A framework for thinking about informal Statistical inference. *Statistics Education Research Journal*, 8(1), 82-105.
- Oliveira, M. M. (2011). *Como fazer projetos, relatórios, monografias, dissertações e teses* (5. ed.). Amsterdã, Holanda: Elsevier.
- Reyes, K. A. R.; Ruz, F.; Molina-Portillo, E. & Contreras, J. M. (2019). Comprensión del concepto de muestra por estudiantes chilenos de cuarto año medio de educación secundaria. In: *Actas del Tercer Congreso Internacional Virtual de Educación Estadística*. Granada, Espanha.
- São Paulo. (2014). *Caderno do Professor*. (v. 2). São Paulo, SP: Secretaria de Educação do estado de São Paulo.
- Triola, M. F. (2008). *Introdução à estatística: atualização da tecnologia*. (v. 10). Rio de Janeiro, RJ: LTC.
- Zieffler A.; Garfield, J.; Delmas, R. & Reading, C. (2008). A framework to support research on informal inferential reasoning. *Statistics Education Research Journal*, 7(2), 40–58.
- Wild, C. & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International statistical review*, 67(3), 223-265.