Sociedade Brasileira de Educação Matemática

RIPEM

How High School students understand statistical variability based on a contextual teaching situation

Edicarlos Pereira de Sousa
Universidade Federal do Cariri
Brejo Santo, CE — Brasil
edicarlos.pereira@ufca.edu.br
0000-0003-4262-544X

Carlos Eduardo Ferreira Monteiro
Universidade Federal de Pernambuco
Recife, PE — Brasil
carlos.fmonteiro@ufpe.br
0000-0003-4355-0793

2238-0345

10.37001/ripem.v14i3.3848

Received • 15/03/2024
Approved • 13/06/2024
Published • 16/08/2024

Editor • Gilberto Januario

Abstract: The teaching of statistics in basic education is important due to the growing needs of contemporary society. This article presents elements of a study that investigated how students perceive, represent, compare, and identify patterns related to statistical variability, in a teaching situation related to electricity consumption. The research had a qualitative and exploratory approach, based on a theoretical perspective that considers different aspects inherent to the conceptual understanding of statistical variability. The participants were 38 third-grade high school students from a state public school in the Cariri region, in the south of Ceará. The results suggested that the perception and identification of variability patterns were the aspects best understood by the students. The answers related to comparison and the representation of variability indicated conceptual mistakes. The analysis of the results emphasized the importance of the context for the students' learning and statistical literacy process.

Keywords: Contextual Activities. Understanding Statistical Variability. Aspects of Variability. Statistical Education.

Cómo los estudiantes de Secundaria entienden la variabilidad estadística a partir de una situación de enseñanza contextual

Resumen: La enseñanza de la estadística en la educación básica es importante debido a las necesidades crecientes de la sociedad contemporánea. Este artículo presenta elementos de un estudio que investigó cómo los estudiantes perciben, representan, comparan e identifican patrones relacionados con la variabilidad estadística, en una situación de enseñanza relacionada con el consumo de electricidad. La investigación tuvo un enfoque cualitativo y exploratorio, basado en una perspectiva teórica que considera diferentes aspectos inherentes a la comprensión conceptual de la variabilidad estadística. Participaron 38 estudiantes de tercer año de secundaria de una escuela pública estatal de la región de Cariri, en el sur de Ceará. Los resultados sugirieron que la percepción e identificación de patrones de variabilidad fueron los aspectos mejor comprendidos por los estudiantes. Las respuestas relacionadas con la comparación y representación de la variabilidad indicaron errores conceptuales. El análisis de los resultados enfatizó la importancia del contexto para el proceso de aprendizaje y alfabetización estadística de los estudiantes.

Palabras clave: Actividades Contextuales. Comprensión de la Variabilidad Estadística. Aspectos de Variabilidad. Educación Estadística.

Como estudantes do Ensino Médio compreendem a variabilidade estatística com base numa situação de ensino contextual

Resumo: O ensino de Estatística na Educação Básica se apresenta como relevante devido às crescentes necessidades da sociedade contemporânea. Este artigo traz elementos de um estudo que investigou como estudantes percebem, representam, comparam e identificam padrões relacionados à variabilidade estatística, numa situação de ensino relacionada ao consumo de energia elétrica. A pesquisa teve uma abordagem qualitativa e exploratória, fundamentando-se numa perspectiva teórica que considera diferentes aspectos inerentes à compreensão conceitual da variabilidade estatística. Participaram 38 estudantes do 3º ano do Ensino Médio de uma escola pública estadual da região do Cariri, no sul do Ceará. Os resultados sugeriram que a percepção e a identificação de padrões de variabilidade foram os aspectos mais bem compreendidos pelos estudantes. As respostas relacionadas à comparação e à representação da variabilidade indicaram equívocos conceituais. As análises dos resultados enfatizaram a importância do contexto para o processo de aprendizagem e de letramento estatístico dos estudantes.

Palavras-chave: Atividades Contextuais. Compreensões de Variabilidade Estatística. Aspectos da Variabilidade. Educação Estatística.

1 Introduction

Statistics teaching has been recognized as a relevant curriculum content inside and outside Brazil. This recognition occurred as a response to contemporary society's demands, which increasingly clamored for knowledge from people inserted in varied contexts and exposed to countless pieces of information. This scenario was highlighted by Isoda, Chitmun, and González (2018), who noted that it is no surprise that, in recent years, statistics have started to play a more prominent role in the curricula of many countries.

For Batanero (2019), the growing movement to approach statistical concepts in global curricula is due to the identification of the significance of pedagogical proposals that help students develop statistical thinking from the first years of their school life. Such changes imply the need for educational institutions to think about and implement educational processes that are updated to everyday challenges.

Within the scope of Brazilian basic education, according to Cazorla, Utsumi, and Monteiro (2021a), statistics contents began to be addressed based on the National Curriculum Parameters — PCN (Brasil, 1997) and were expanded with the publication of the Common National Curriculum Base — BNCC (Brasil, 2017). Both documents aim to guide public and private basic education schools and, therefore, support the implementation or reformulation of national public policies in education. Frei, Rosa, and Biazi (2023) emphasize that there was no break between these regulations, just a content distribution reorganization.

Lima, Paula, and Giordano (2022), Portiúncula and Batisti (2023), Santos (2015), Schreiber and Porciúncula (2019), and Silva, Curi, and Schimiguel (2017), among others, underscore advances in scientific production in the field of statistical education, which show fruitful results at different levels of education, in addition to an increase in the number of research and researchers. However, these authors reinforce the need for more studies on the processes of statistical education to assist structured teaching practices in challenging situations that may lead students to problematize and investigate the realities of their school, their neighborhood, and the country in which they live.

Despite recognizing the importance of teaching statistics, some factors make learning statistical concepts difficult. On the one hand, teachers have conceptual gaps arising from their initial education (Conti, 2016). On the other hand, students cannot access sufficient statistical knowledge. To Cazorla, Henriques, Correia, and Santana (2021), one factor that causes difficulties in teaching and learning statistical concepts is the lack of understanding of how data can and should be collected and how raw data are transformed into statistics, usually represented by tables, graphs, and summary measures. Furthermore, in situations designed to promote statistics learning, students must consider the role of the nature of the variables analyzed, paying attention to the appropriate way of representing them.

This article presents elements of a study that aimed to investigate how high school students understand different aspects of statistical variability (perception, representation, comparison, and identification of patterns) in a contextual teaching situation about electrical energy consumption.

2 Statistical variability as an inherent characteristic of data sets

The ubiquity of variability is considered the main characteristic of statistical variables, as understanding how and why data vary is part of the essence of statistics. To Cazorla, Utsumi, and Monteiro (2021b), statistical variables are marked by variability, which differentiates them from mathematical variables. This difference can be explained because while mathematics is based on the deterministic nature of its results, statistics presuppose uncertainty and attempt to reduce error in estimates. It encompasses the concept of possibility and seeks to understand the phenomenon of data variation. Therefore, statistics is situated from a non-deterministic perspective. Thus, the objects of knowledge of mathematics and statistics are in different epistemic fields.

Understanding variability encompasses both informal aspects, for example, when we perceive that data varies and verify differences in their values and formal aspects through the understanding and interpretation of variability measures, such as range, interquartile range, and standard deviation, among others. According to Garfield et al. (2008), although many students can calculate formal measures of variability, they rarely understand what these summary statistics represent, whether numerically or graphically, and do not understand their importance and connection with other statistical concepts. In the same sense, Lee and Meletiou-Mavrotheris (2003) showed that students presented mistaken ideas or incorrect reasoning about the concept of variability present in graphic representations.

Variability is one of the characteristics of statistical variables, and therefore, they need to be treated differently from mathematical variables (Cazorla et al., 2021b). These authors emphasize that when we study the relationship between statistical variables, we need to understand the role that each plays. In other words, it is essential to highlight the dependent and the independent variables, as separating these roles is not always clear and easy, as is the case with mathematical functions.

In Brazil, ideas involving variability are included in several BNCC knowledge objects. Cazorla et al. (2021) described all the objects of statistics knowledge that should be taught to Brazilian students during their basic education journey:

> [...] sample and census research; qualitative and quantitative variables, population and sample; statistical tables and frequency distribution tables (FDT), single and double entry; bar/column, line, circular, pictogram, stem and leaf diagram, box-plot and histogram graphs; summary measures: absolute and relative frequency; measures of central tendency — MCT (mean, median and mode); measures of dispersion measures — MD (amplitude and standard deviation). (Cazorla et al., 2021, p. 25)

Among the objects of knowledge listed above, we highlight graphical representations, measures of central tendency, and measures of dispersion as examples in which the conception of variability is intrinsically related. It should also be noted that although the quartiles and interquartile range were not listed directly in the document, these are statistical concepts preceding the construction and understanding of the box plot. Additionally, although the BNCC does not include the dot plot, rod or stick diagram, and coefficient of variation (CV), Cazorla et al. (2021a) advocate for its teaching, preferably before presenting the box plot.

The studies by Pereira, Rodrigues, and Souza (2020) and Sousa, Cazorla, and Monteiro (2024) discuss the importance of investigating variability and its teaching in basic education. They carried out a survey of scientific production involving variability statistics, and the results showed few studies that focused on this topic.

3 The role of recognizing the nature of statistical variables

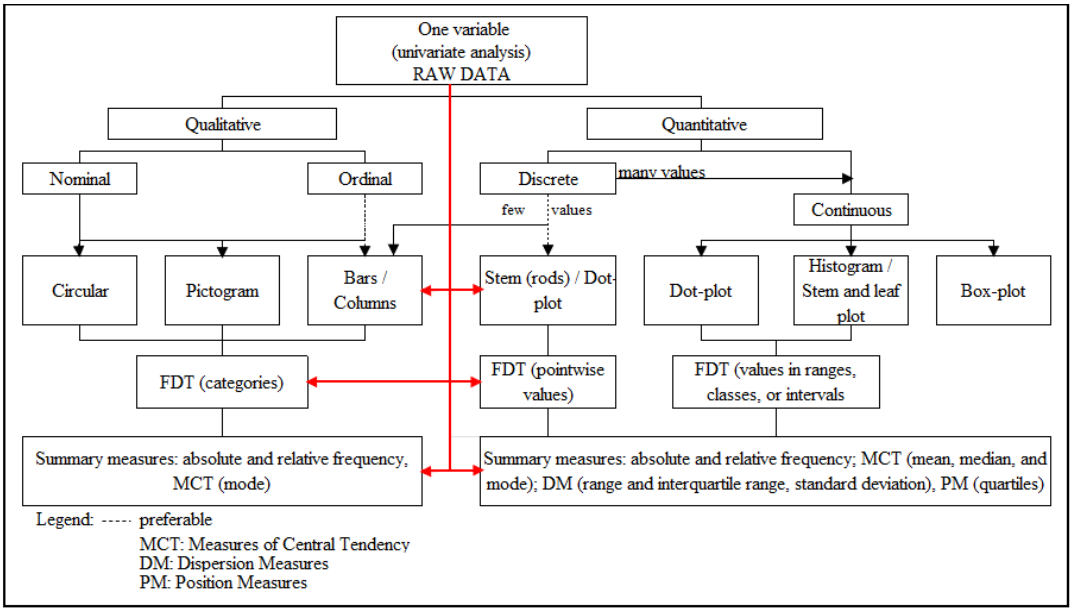
Statistical variables are classified as quantitative and qualitative, depending on their nature. Cazorla et al. (2021b) highlight that the BNCC adopted the terms categorical (qualitative) and numerical (quantitative). Data from counting or measurement represent quantitative variables. This type of variable is subdivided into discrete (when it takes positive integer values, for example, the number of dengue cases in a municipality) and continuous (when it results from a measurement, such as a student's height). Qualitative variables manifest characteristics of the observed phenomenon, usually grouped into categories, and may be nominal (when they do not present ordering in their categories, for example, the choice of streaming preferred among categories such as films, series, documentaries, and podcasts) or ordinal (when they express an ordered process, such as the level of satisfaction with the cell phone operator).

Regarding the observation reference, Cazorla and Oliveira (2010) point out that statistical variables can receive another form of classification: empirical, when they have a directly observable referent in the empirical world (for example, people's age), generally associated with the phenomena of physical, natural and experimental sciences; and conceptual, when they cannot be directly observable, generally associated with human behavior (for example, habits to prevent Covid-19 contagion). Conceptual appropriation and carrying out sufficient statistical procedures become important to organize and synthesize raw data, expressing summary measures (of central tendency or dispersion) and representing them graphically in an appropriate manner.

Graphical representation is one of the most common ways of presenting data, as it helps understanding the underlying information. Therefore, it assumes great relevance in research in statistical education. Figure 1 shows the path of raw data in studies with a single variable (univariate) during transformation into statistics.

When analyzing Figure 1, we realize that to conveniently choose which statistical graph to construct, it is crucial that we have established the concept of statistical variable and recognized its nature. Otherwise, mistaken representations may be made, causing conceptual difficulties and inferring superficial or inadequate conclusions about the data. Cazorla et al. (2021a) suggest that we must recognize the nature of the variable and the data generated to make appropriate decisions about its statistical treatment. The authors believe that a statistical variable and how it should be represented/operationalized must be known.

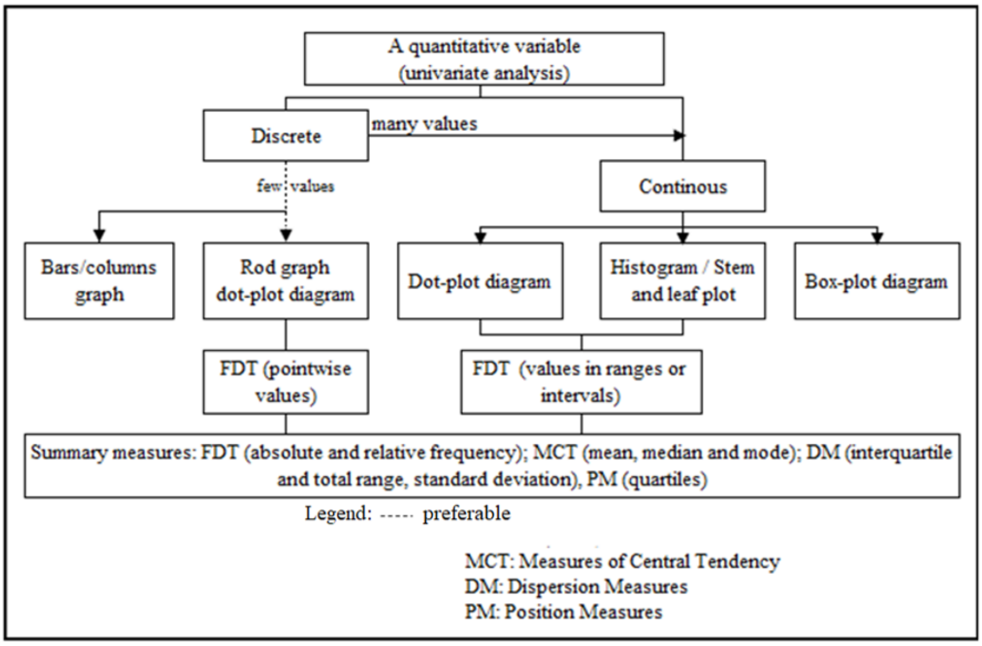
Figure 1: Types of graphical representations of a statistical variable



Source: Cazorla et al. (2021a, p. 115)

In this work, the approach to representing quantitative variables occurred only in the univariate context. Figure 2 shows the different types of charts recommended in this situation.

Figure 2: Types of graphical representations of a statistical variable



Source: Cazorla et al. (2021, p. 26)

It is noteworthy that qualitative variables, conceptual variables, and bivariate contexts, as they are complex, would require specific approaches and, therefore, are not included in this discussion. According to Cazorla et al. (2021b) and in line with the BNCC (Brazil, 2018), except for box plots, histograms, and stems and leaves, which are suitable for use in high school (15 to 17 years of age) and in teaching situations involving variables continuous graphs that present many values, the other graphs must be introduced since elementary school (6 to 14 years of age).

The possibilities for transforming and representing raw data on quantitative variables can be organized into three groups: (1) univariate analysis of discrete variables that take few values; (2) univariate analysis of continuous or discrete variables that take many values; and (3) bivariate analysis of quantitative variables (Cazorla et al., 2021). Given the statistical nature of the variable used in the proposed teaching situation and the objective listed above, this work was restricted to the theoretical framework of the second group.

4 Methodological procedures

Aiming to contribute to the discussion, we investigated students' understanding of the statistical variability of a data set on monthly residential electricity consumption. The study explored whether students understood how and why the data varied; when they varied, what factors interfered with the variation, and how to represent them graphically, based on the recognition of the nature of the variable investigated.

The methodological choices defined in a study constitute opportunities for organizing thought, provoking reflections on the investigations to be developed throughout the process (Ghedin & Franco, 2011). To these authors, following specific methodological itineraries means being aware that this implies a series of attitudes, positions, and procedures, all coherently planned and exhaustively consistent with established convictions, giving meaning to the approaches taken and intentions of the research. Therefore, we chose to emphasize the qualitative aspect of the research, regardless of whether elements of a quantitative nature appear in some moments of data collection, systematization, or analysis. Marconi and Lakatos (2017) point out that the qualitative approach aims at a particular understanding of the object it investigates within the context in which it arises. In this way, we sought to investigate how high school students understand the statistical variability of monthly electricity consumption in a teaching situation with actual data, in addition to understanding how they argue about the different graphical representations of these data. When we investigate the meanings, beliefs, values, and attitudes students attribute to the proposed questions, we can analyze them as statistically literate subjects or not because the concept of statistical variability is at the heart of understanding several other objects of knowledge in this field.

We also adopted an exploratory nature, as we intended to approach the problem investigated, that is, students' understanding of the variation in electricity consumption and associated factors, to make it more explicit, in line with Gil (2017). For this author, many research designs consider the objectives, the environment of its development, and the forms of data collection and analysis. For this reason, we initially carried out the activity by recording facts but without intervening in the students' initial responses or interpretations. We must highlight that the discussions presented here result from the first moment with students, among several stages planned for the first author's ongoing doctoral research.

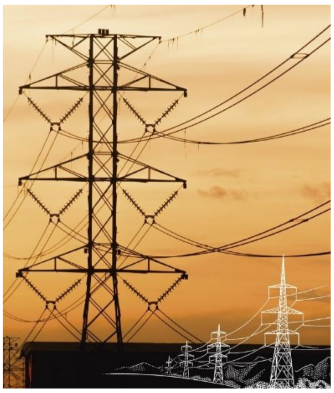
The activity took place in November 2023 in a state public school in the south of Ceará. Thirty-eight students from the third grade of high school participated voluntarily after authorization through a Free and Informed Consent Form (FIC) signed by all students over 18 years of age and by a guardian of each minor. Participants completed an activity composed of several items that addressed different aspects of statistical variability regarding electricity consumption data. The activity used Garfield and Ben-Zvi (2005) as a reference; they present several components (aspects) of an epistemological model that constitutes a deep understanding of statistical variability. According to the authors, the proposed model helps in planning and developing actions aimed at conceptual learning for students at different levels of education.

5 Analyzing students' understanding of variations in electrical energy data

The results of this discussion come from the activity *"Electrical energy: why did the prices go up?"*, prepared based on an article published by the Ceará newspaper *Diário do Nordeste* in the second half of 2021 to highlight the increase in costs with essential expenses. Figure 3 shows the context of the teaching situation proposed to students.

Figure 3: Context of the teaching situation proposed to students

"ELECTRICITY: Why Did It Increase?"



"In addition to the high costs of food, consumers increasingly have to spend more on essential expenses such as electricity and water and sewage fees. The situation worsened with the pandemic, as consumption increased due to people spending more time at home."

In April, for example, with the annual tariff adjustment, residential electricity bills in Ceará increased by 7.5%. On average, the value revision was 8.95% for the entire state.

The costs became even higher this month with a 52% increase in the value of the red flag level 2 tariff.

The fee, which previously cost R$ 6.24, increased to R$ 9.49 per 100 kilowatt-hours consumed."

Source: Diário do Nordeste newspaper, August 2, 2021.

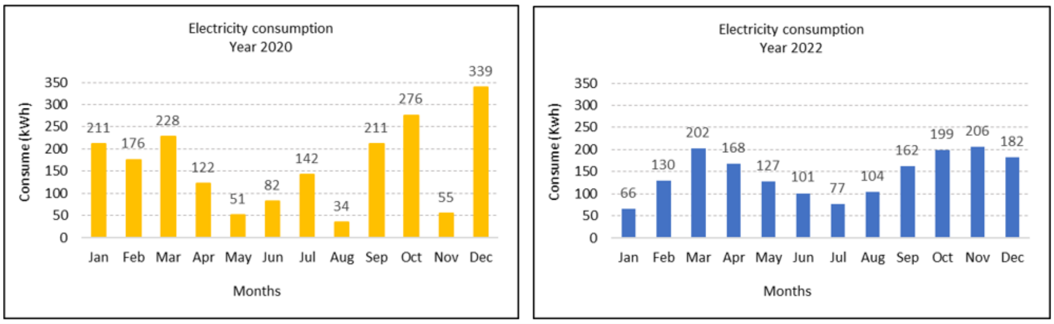
The article above was published by the Diário do Nordeste newspaper to highlight the increase in the costs of essential expenses, such as residential electricity bills, across the state of Ceará.

Source: Adapted from Carvalho (2021)

After reading the article, students were invited to analyze the electricity consumption of a residence in the Cariri region of Ceará in 2020 and 2022 (Figure 4).

Figure 4: Proposed activity on electricity consumption

In this context, we present some information related to the monthly electricity consumption of a residence located in the municipality of Brejo Santo, in the Cariri region of Ceará, for the years 2020 and 2022.



Source: Research Data

The investigation considered the aspects of perception, representation, comparison, and identification of patterns of statistical variability. Therefore, we present students' understanding of just a few items in the proposed activity. A complete discussion, including all items of the activity and covering other aspects of variability, according to Garfield and Ben-Zvi (2005), became unfeasible, as it would require greater depth and space to discuss each.

To understand the students' perception of the variability of electricity consumption, they were asked which statistical variable was being represented and were asked to justify the answer presented. Despite the item being focused only on the issue of representation, the students' answers could also indicate whether they understood the concept of a statistical variable, in its essence, and whether they had intuitive ideas about variability. Chart 1 presents some participants' answers.

Chart 1: Students' answers to the question: *According to the graphs above, which statistical variable is being represented?*

Examples of answers that correctly identified the statistical variable	Examples of answers that confused the statistical variable with the unit of measurement
MCSB: Electricity consumption during the months of 2020 and 2022.	*ARAFS: The variable represented in the text is kwh consumption [...].*
MSS: Energy consumption, as the graph is structured to discover the monthly energy consumption.	*CMD: "Kwh" electric consumption and time.*
MSP: The electrical energy consumption of a residence.	*KOL: Kilowatt-hour variable.*

Source: Research Data

Most students could correctly identify the statistical variable of *electric power consumption*. However, some left the item blank or gave answers that revealed misunderstandings between what would essentially be the variable and its unit of measurement, the kilowatt-hour (kWh).

Regarding the perception of variability and the contribution of representation to understanding the phenomenon investigated, some answers that highlighted the presence of both aspects stand out:

> EAML: *The consumption/month and the variation for the years 2020 and 2022 are being represented;*
>
> MMCF: *The electricity consumption that can be observed through the numbers in the graph and the comparison of years;*
>
> SAS*: Energy consumption is being represented whether it increased or decreased between 2020 and 2022;*
>
> SRXS*: There was a drop in consumption during the months because the columns for the years 2020 and 2022 differed regarding kWh.*

We can highlight that the students mobilized the perception and representation aspects to argue about the variation in the data. In this sense, several authors emphasize that much of the incentive to investigate statistical variability in the school environment arises from the difficulties usually presented by students concerning their understanding and the constant tensions that their perception and representation offer regarding other statistical ideas, such as distribution, centrality, sample, and representativeness (Chan & Ismail, 2013; Estrella, Vergara & González, 2021; Garfield, Le, Zieffler & Ben-Zvi, 2015).

Students also mentioned the source of data variability:

> ARGL: *[...] in 2020, consumption was higher; due to the pandemic, people spent more time at home [...] in 2022, the pandemic was over, citizens returned to their normal tasks [...] so they stopped consuming the energy they used so much in 2020;*

JVAS*: [...] energy has a variable value since the expense revolves around the time used;*

VFV: *The energy consumption statistics, because 2020 was the year of the pandemic and people were more at home, by 2022, the pandemic had already ended and the numbers dropped;*

CE: *Monthly consumption in kWh [...] and the post-pandemic effects caused by this consumption.*

Students appropriated relevant issues specific to the activity as it occurred, which Watson, Fitzallen, Wright, and Kelly (2023) define as contextual variation, which includes recognizing themes related to personal practical experience. For these authors, the student must understand the essence of the nature of the context in the construction of the statistical understanding regarding the specific and general aspects of variation. Students highlighted the pandemic context and the time of use of electrical appliances and lamps as factors responsible for the variation in consumption. From this perspective, Garfield and Ben-Zvi (2005) highlight the need to discuss the sources that lead to variability in a data set instead of just collecting data to analyze in graphs and summary measures or presenting data to be summarized in this way.

When asked how they would explain the abrupt variation in electricity consumption from November to December 2020, the students explained possible factors that they believe are responsible for the peak in energy consumption at the end of 2020, as seen in Chart 2.

Chart 2: Student answers to the question: *How would you explain the variation in energy consumption from November to December 2020?*

End of year festivities	Vacation
LKSN: I think that at the end of the year, parties and enjoyment require more energy; *SRXS: Because of Christmas since, throughout December, houses are decorated with celebratory lights and probably the large number of these lights, added to conventional appliances, increased energy consumption.*	*JVAS: Well, it's vacation time. The more people at home, the more consumption [...];* *KOL: I imagine that because December was a holiday month, people spent more time at home, thus increasing energy consumption.*
Issues related to local weather or others	**Because of the pandemic**
ARAFS: December is one of the hot months [...]; *MESF: [...] climate change, which I believe affects consumption.*	*ROS: Due to the pandemic, which made people stay at home and use more energy;* *MCSB: The end of the pandemic.*
The increase in the cost of energy supply	**Other factors**
MSP: The energy bill may have increased and then decreased; *MESF: I believe that the readjustment in value, or awareness [...].*	*EICC: The family may have taken a long trip or cut back on unnecessary energy costs.*

Source: Research Data

Students went beyond the explicit data, identifying their variability and indicating possible reasons for changes in the monthly electricity consumption values, such as end-of-year parties, vacation periods, local climate issues, and the pandemic context. However, regarding supply costs, students confused the variable *consumption* (in kWh) with the variable of *expenditure on energy* (in reais), which depends on the price established by the energy concessionaire. The graph presented did not deal with the value of the energy bill; it only showed how consumption varied during the period considered.

The students seem to have pointed out factors responsible for the variation in electricity consumption, citing the importance of the context of the data investigated and the need for a critical stance towards them from the perspective of statistical literacy (Gal, 2002). Corrêa and Lopes (2020) reinforce that statistics consider the data part of a context. Therefore, understanding, describing, and questioning real situations are part of an investigation whose interpretation works with possible answers without the determinism of a single solution, as the data presents variability.

In the participants' opinion, other factors may also have influenced the variation in the data. One student suggested that the data variation pattern may have occurred due to changes in the family's routine during the pandemic. Indeed, this happened to the family in the researched residence. During 2020, for personal reasons, the family made quick trips between the states of Ceará and Paraíba. However, students were not informed of that, although a participant suggested it to explain the changes in monthly electricity consumption.

Considering the role of the nature of the variable in understanding the data and its adequate representation, we sought to analyze whether the students recognized the studied variable. In this case, they answered the following question: *For you, can this variable be understood as qualitative or quantitative? Justify your answer*. The answers are detailed in Chart 3.

Chart 3: Student answers to the question: *Should this variable be understood as qualitative or quantitative? Justify your answer*

Classification of energy consumption as a quantitative variable	Classification of energy consumption as a qualitative variable
RRP: Quantitative, as it shows the amount of kWh consumed in each consecutive month; *MSS: Quantitative, as it is determined by a measurement, which, in this case, is the kWh;* *LPS: Quantitative, because what changes are the values;* *KPV: Quantitative. This is the amount of energy (kWh) consumed.*	*CMD: Qualitative, as consumption will vary over the months;* *MSP: [...] Qualitative because, through the data, it is possible to get an idea of the financial quality of the residents [...].*

Source: Research Data

Most students provided appropriate answers to understanding energy consumption as a quantitative variable. The justifications were associated with the idea of measure or quantity of consumption. However, we noted that some students did not answer the question or stated that it was a qualitative variable. Thus, a contradiction is perceived when, for example, a student justified the variation over the months as the main reason for classifying the variable as qualitative. What the other understands by *dwellers' financial quality* is also vague. However, this does not change the quantitative nature of energy consumption, which highlights a conceptual misunderstanding about the types of variables, particularly for these participants. These cases confirm what Silva Junior and Lopes (2014) pointed out: they noticed that when undergraduate students were questioned about qualitative, quantitative, discrete, and continuous variables, they declared not knowing how to conceptualize them or ignored the best way to represent them.

Graphical representation is relevant for understanding research data. Therefore, students were asked about the type of graph used in the proposed situation: *A column graph presented the energy consumption of the analyzed residence in both years. Would you use any other type* *of graph to represent these data?* Table 1 presents the students' answers to this question.

Table 1: Types of graphs suggested by students to represent electrical energy consumption

Would you change to a different graph?		Type of graph students suggested	Frequency	Percentage (%)
No	18	Does not apply		47.4
No answer	04	Does not apply		10.5
Yes	16	Bar	03	42.1
		Line	05	
		Bar and Line	01	
		Circular ("Pizza")	06	
		Dot	01	
Total	38	—		100

Source: Research Data

The students gave different answers to the question of representation, which were gathered into three groups: those who believed that the column chart should be maintained as the most appropriate way to visualize the data, those who proposed another type of chart to represent the monthly data on electricity consumption and, finally, those who did not offer any way of graphically representing the variable, similar to the results obtained by Silva Junior and Lopes (2014). This shows that students' conceptions about representation are not well defined. Furthermore, other graphic forms also emerged in their answers and deserve to be discussed in the classroom. According to Fernandes, Martinho, and Gonçalves (2020), the type of statistical variable plays a vital role in conditioning the appropriate possibilities to represent the data when selecting the type of graph.

Regarding the choice of graph type, we must clarify some points. Although we are working with a quantitative variable, which could include, for example, a dot plot or even a box plot, which the participants did not even mention, both representations would not be adequate to describe the behavior of this variable, as it is indexed in time, that is, it is a time series. In the specific case of electricity consumption, we are facing a variable influenced by the seasons, which is why it does not allow us to remove the influence of seasonality to better understand the data investigated.

Column charts are often used in textbooks, television, and research studies, which may explain why almost half of the class maintained this choice. However, it is worth highlighting that this type of graph becomes unfeasible with more data or when one wants to superimpose the data for a year, distinguishing the years by superimposed lines, for example. In this sense, we can observe that the eighteen students who answered *No* are partly correct, as the column chart is commonly used for energy consumption data. However, the variable suffers the interference of time. Therefore, the most appropriate graph for this type of situation is the line graph, as it allows one to visually explore possible interferences from the seasons or cultural events in a location, such as school holidays. Of the students who responded *Yes*, five chose the line graph, which is correct.

One student chose bars and lines, and three others chose just bars. To the students who opted for a bar graph, such as the one with horizontal bars, we could ask them about the suitability of using this type of graph to represent a time series, considering that the practice is to represent time on the abscissa, from left (oldest) to right (most current), and very rarely in ordinate.

Although most of the class defined energy consumption as a quantitative variable, six students suggested changing the form of the graphical representation to a circular graph, known to them as a *pie chart*. In this case, it was evident that there are conceptual difficulties with representation, as the richness of the data information could be lost. In this type of graph, how much was consumed in a given month in relation to the year would be represented, i.e., a part-whole analysis, which does not make the circular graph appropriate enough to describe the essence of the behavior of the variable under study. This type of graph would be more suitable to represent a nominal qualitative variable (Figure 1). During the process of defining the appropriate graphical representations for each situation, observing whether the data is qualitative or quantitative, discrete or continuous, small or high in number, and with many or few repetitions has different implications for choosing the appropriate graph to represent the data analyzed in each context (Fernandes et al., 2020).

A student marked the dot graph, and, as we have already mentioned, it makes sense when we remove the interference of time in the behavior of the variable from the analysis. This type of graph is advantageous when we want to analyze data distribution, concentration, and dispersion. It is an auxiliary graph for data analysis. We must highlight the uniqueness of a student choosing this option, considering that the BNCC does not include the dot diagram, which makes it little known by most students and teachers.

The comparison between two or more data sets is highlighted by Garfield and Ben-Zvi (2005) as one of the key ideas for understanding statistical variability. Therefore, we asked students: Which year had the highest energy consumption, looking at the graphs for 2020 and 2022? Why did sudden drops occur between some consecutive months of a single year, such as October/November 2020? Practically all students said that 2020 presented the greatest variation. Some mobilized visual elements from the graphs to make the comparison.

> EICC: *2020. Because visibly the first graph seems higher most of the time;*
>
> KOL: *2020. Because the graphs are being represented by columns and in 2020, the numbers are higher;*
>
> LKSN*: 2020. Because the changes in the charts are much bigger;*
>
> EAML*: 2020. Bearing in mind the graph itself shows more high variations in 2020 than in 2022, thus leading to this answer;*
>
> ALSF*: 2020. Comparing one year to another, the difference is striking; an example of this is when we calculate the difference in the month of December in both years.*

In other cases, students performed calculations to extract information about the variation. We highlight, for example, one of the answers that pointed out a difference of more than 200 kWh between the household's annual consumption in the two years studied. Table 2 presents the main descriptive statistics on the two data sets.

It is worth highlighting that the variability of energy consumption must be considered together with its center and dispersion measures. According to Garfield and Ben-Zvi (2005), to evaluate reasoning about variability, students must experience situations with graphs that show more or less variability, mobilizing the perception of the shape, center, and different data measures of dispersion. Variability is a global characteristic of random phenomena and can be analyzed based on different statistics, such as amplitude, variance, and standard deviation, but it cannot be reduced to just a single measurement (Estrella et al., 2021; Pfannkuch & Wild, 2005). By synthesizing data according to the statistics summarized in Table 2, students increase their chances of understanding the most unstable year. Therefore, despite 2020 having the most significant variation in energy consumption, focusing on specific values, comparing similar months in the two graphs, and observing exclusively the height of the columns, as pointed out by some students, consist only of initial understandings about variability or even conceptual mistakes, depending on how these reflections are expressed.

Table 2: Descriptive statistics of energy consumption in the two years analyzed

Statistics	2020	2022
Mean	160.6	143.7
Minimum	34.0	66.0
Maximum	339.0	206.0
Amplitude	305.0	140.0
Standard deviation	92.3	47.5

Source: Research Data

Furthermore, the two years could have the same total annual consumption but present a different internal variation, i.e., a divergence in the intra-group variation. The same reasoning can be applied, as an example, to average consumption. The sets could have equal monthly averages but different variations for each year. Therefore, it is not enough to consider just one measure of central tendency to analyze the variability of a data set since distributions can have equal center measures but different dispersions.

Two students wrongly argued about the year that suffered the most variation in energy consumption. One of them claimed that 2022 was the year that presented the most months with much higher consumption than 2020. In this case, there is a misconception of variability as a measure of the length of the vertical axis. Further discussions about students' understanding of the comparison of two data sets can be seen in González (2021), who obtained results similar to those presented in this work.

The second item, which involved the comparison aspect, dealt with the abrupt variations between October and November 2020. The answers focused on climate issues, family habits, the more severe pandemic context in the first year and more stability in the second year, and personal factors, similar to the discussion related to Chart 2.

The last aspect of statistical variability discussed was pattern identification. We asked students to analyze the behavior of the variable in the two years, asking them which year had the most stable consumption. Only one student failed to answer this question, and two others did not clearly argue about the existence or absence of a data pattern:

> CMD: *2020. Because this year had a lower consumption rate, so consumption is stable;*
>
> RRP: *2020. In some months, energy was used much less.*

One cited the concept of a consumption index; however, no statistical treatment was carried out on the data, and no index was defined. Paying attention to low energy consumption in a specific month of 2020 is also not appropriate to prove the existence of a general pattern in the data.

For the other students, we presented excerpts in which, in some way, identifiable patterns were highlighted.

> MESF: *2022. No month presented too high consumption, nor any month too low consumption, remaining considerably stable;*
>
> JMISS: *2022. Because the variation from one month to the next is not that great;*
>
> LKSN: *Both years show different patterns. Because of the graph, the columns are very different;*
>
> SRXS: *[...] the columns for 2020 present higher values than the values for 2022;*
>
> JMISS*: [...] one year shows stability in energy expenditure, the other shows variation in the months;*
>
> MCSB*: [...] the 2020 values change a lot, up and down. In 2022, the values remain similar.*

One of the main characteristics extracted from the students' justifications is related, as well as in the comparison aspect, to the conception of variability based on the columns of the graphs, but no justifications related to formal statistical measures were presented to support the initial understandings. On the other hand, specific conceptions were possible to observe in terms of variability, as reported below.

> ARAFS: *[...] consumption is in a range, there is no big drop or big increase;*
>
> JVAS: *[...] it was a year in which monthly consumption was almost the same during the 12 months;*
>
> EICC*: [...] the graph is higher, your monthly energy average will be higher;*
>
> SRXS*: [...] due to the size of the columns referring to the month, as in December 2020 alone the value was 339;*
>
> LTL*: [...] it was the top peak of energy expenditure.*

Some students used the idea of a range of energy consumption to suggest a pattern of variation in the data, while others identified similar values throughout the year. In this way, we noted that students mobilized different conceptions of variability, i.e., the variability of the entire range of data and variability as change over time, corroborating Shaughnessy (2007). To this author, there is also a conception of variability related to particular values, including extreme or atypical values. Students who paid greater attention to consumption peaks, as in December 2020 (339 kWh), mobilized such knowledge. In this case, this conception of variability also dialogues with that of Garfield and Bem-Zvi (2005), when students focus on specific data values, such as some measure of central tendency, often on very large or very small values, or very odd individual values in a graph or data set.

6 Final considerations

In this article, we investigated how high school students understood aspects of statistical variability based on a teaching situation with real data on electricity consumption. The results indicated differences in the aspects mobilized and the understanding of a given aspect by different students. This showed that participants do not present a linear understanding of aspects of statistical variability. Conceptual errors in terms of statistical knowledge were also identified in some situations.

The students mobilized several explanations to justify the statistical variability in the data, reinforcing the perception aspect and the sources of variability: the pandemic period, which kept many people in social isolation and generated high energy consumption, was frequently highlighted in the answers; consumption that is not always conscious, in which unnecessary expenses were made without taking into account the resulting financial impact; the increase in energy company supply costs; changes in the family routine; end-of-year parties and get-togethers; holiday periods, when people stayed at home longer and ended up consuming more energy; and seasonal issues, highlighting students' relationship with the local climate, which required more expressive use of devices such as air-conditioning fans in hotter periods. Therefore, the importance of the context in stimulating discussions and mobilizing students to critical reflection and extrapolation of purely mathematical procedures from a statistical literacy perspective became evident.

Understanding the concept of a statistical variable was investigated in the research and presented satisfactorily by most participants, although some expressed difficulties in separating the variable from its measurement unit. The type of statistical variable was also included in some activity items. In this topic, more than half of the students presented appropriate answers when classifying energy consumption as a quantitative variable.

Regarding data graphical representation, students presented divergent answers. Some insisted on the column chart as appropriate to the data type addressed, but others suggested bar, bar-and-line, line, circular, or dot charts. The different answers indicated that the graphical representation is not consolidated among students and that associating it with the nature of the variable and the size of the data set is fundamental for making the right choices to better elucidate the characteristics underlying data.

The comparison was approached because it is one of the key ideas for understanding statistical variability. In this way, using visual graphic elements helped participants analyze the difference in consumption between years. Some correctly identified 2020 as the year with the broadest variability. However, the emphasis on point values and column heights revealed early-stage understandings or conceptual misunderstandings when associating the variability in the data with the extension of the vertical axis in the graph. Intragroup variability was also disregarded in some cases, which showed the need to explore this situation with students further.

Participants also pointed out conceptions of variability as being the entire range of data and particular values. They identified them during the investigation of the existence of patterns in energy consumption. Consumption ranges, mention of specific or atypical values, or some measure of central tendency, such as the average, were some of the arguments students used. In general terms, identifying patterns, similar to perception, was one of the aspects detected by practically all participants, as opposed to representing and comparing variability, which generated greater difficulties.

Finally, this research presented contributions to the field of research in statistical education. Statistical variability is conceived as the heart of statistical thinking, which justifies the importance of this type of study with students and teachers. Therefore, school curricula and pedagogical practices must stimulate structured teaching activities based on meaningful contexts for students, allowing them to dialogue and favoring the learning of objects of statistical knowledge aligned with the different social demands inherent to the context in which they live, locally or globally.

Acknowledgments

We would like to thank Prof. Dr. Irene Mauricio Cazorla from the State University of Santa Cruz (UESC) for her important contributions in preparing this article and co-supervising the doctoral project to which the research discussed refers. Her efforts to promote statistical

education to students, teachers, and society inspire mathematical and statistical educators.

References

Batanero, C. (2019). Treinta años de investigación en educación estadística: reflexiones y desafíos. In: *Actas del Tercer Congreso Internacional Virtual de Educación Estadística* (pp. 1-15). Granada, AN.

Brasil. Ministério da Educação. 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.

Brasil. Ministério da Educação. Secretaria de Educação Básica. (2018). *Base Nacional Comum Curricular: Ensino Médio*. Brasília, DF.

Carvalho, L. (2021). *Por que os custos com energia elétrica subiram mais de 24% em Fortaleza?* Diário do Nordeste. Disponível em: https://diariodonordeste.verdesmares.com.br/negocios/por-que-os-custos-com-energia-eletrica-subiram-mais-de-24-em-fortaleza-1.3104771 Acesso em 6 nov. 2023.

Cazorla, I. M. & Oliveira, M. S. (2010). Para saber mais. In: I. M. Cazorla & E. Santana, E. (Org.). *Do tratamento da informação ao letramento estatístico* (pp. 113-144). Itabuna, BA: Via Litterarum.

Cazorla, I. M.; Henriques, A.; Correia, G. S. & Santana, C. V. (2021). O papel dos ostensivos na compreensão das variáveis estatísticas quantitativas. *Acta Scientiae*, *23*(4), 16-51.

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

Cazorla, I. M.; Utsumi, M. C. & Monteiro, C. E. F. (2021b). Variáveis estatísticas e suas representações em gráficos: reflexões para seu ensino. *Números*, *106*, 23-32.

Chan, S. & Ismail, Z. (2013). Assessing misconceptions in reasoning about variability among high school students. *Procedia - Social and Behavioral Sciences*, *93*, 1478-1483.

Conti, K. C. (2016). Educação Estatística num contexto colaborativo: ensinar e aprender probabilidade. *Educação Matemática Pesquisa*, *18*(3), 1117-1140.

Corrêa, S. A. & Lopes, C. E. (2020). A insubordinação criativa e o processo dialógico na Educação Estatística na infância. *Revista Internacional de Pesquisa em Educação Matemática*, *10*(1), 95-107.

Estrella, S.; Vergara, A. & González, O. (2021). El desarrollo del sentido del dato: haciendo inferencias desde la variabilidad de los tsunamis en primaria. *Statistics Education Research Journal*, *20*(2), 1-14.

Fernandes, J. A.; Martinho, M. H. & Gonçalves, G. (2020). Uso de gráficos estatísticos por futuros professores dos primeiros anos na realização de trabalhos de projeto. *Jornal Internacional de Estudos em Educação Matemática*, *13*(4), 394-401.

Frei, F.; Rosa, J. S. & Biazi, A. H. (2023). Professores de Matemática estão preparados para o ensino de Estatística e Probabilidade? *Revista Internacional de Pesquisa em Educação Matemática*, *13*(2), 1-17.

Gal, I. (2002). Adult's Statistical Literacy: meanings, components, responsibilities. *International Statiscal Review*, *70*(1), 1-25.

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.

Ghedin, E. & Franco, M. A. S. (2011). *Questões de método na construção da pesquisa em educação*. (2. ed.). São Paulo, SP: Cortez.

Gil, A. C. (2017). *Como elaborar projetos de pesquisa*. (5. ed.). São Paulo, SP: Atlas.

González, O. (2021). Teachers' conceptions and professional knowledge of variability from their interpretation of histograms: the case of venezuelan in-service secondary mathematics teachers. *Statistics Education Research Journal*, *20*(2), 1-13.

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.

Lee, C. & Meletiou-Mavrotheris, M. (2003). Some difficulties of learning histograms in introductory statistics. *Joint Statistical Meeting: Section on Statistical Education*, 2326-2333.

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), 1-19.

Marconi, M. A. & Lakatos, E. M. (2017). *Metodologia científica*. (7. ed.). São Paulo, SP: Atlas.

Pereira, F. A.; Rodrigues, C. K. & Souza, F. S. (2020). Pesquisa sobre variabilidade na Educação Estatística: uma revisão sistemática da literatura. *Zetetiké*, *28*, 1-14.

Pfannkuch, M. & Wild, C. (2005). Towards an understanding of statistical thinking. In: Ben-Zvi, D. & Garfield, J. (Ed.). *The challenge of developing statistical literacy, reasoning and thinking* (pp. 17-46). Dordrecht, The Netherlands: Springer.

Porciúncula, M. & Batisti, I. (2023). Estado do conhecimento acerca da Educação Estatística no contexto da Educação Infantil. *Ensino em Re-Vista*, *30*, 1-28.

Santos, R. M. (2015). *Estado da arte e história da pesquisa em Educação Estatística em programas brasileiros de pós-graduação*. 2015. Tese (Doutorado em Educação). Universidade Estadual de Campinas. Campinas, SP.

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*, 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.

Silva Junior, G. B. & Lopes, C. E. (2014). Representações gráficas e variáveis estatísticas: reflexões para a formação de engenheiros de produção. *Revista Paranaense de Educação Matemática*, *3*(5), 100-114.

Silva, J. F.; Curi, E. & Schimiguel, J. (2017). Um cenário sobre a pesquisa em Educação Estatística no Boletim de Educação Matemática – BOLEMA, de 2006 até 2015. *Bolema*, *31*(58), 679-698.

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*, 1-24.

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.