Sociedade Brasileira de Educação Matemática

RIPEM

The understanding of Stochastic brought from basic education by Mathematics prospective teachers in an introductory Statistics course

Maria Helena Monteiro Mendes Baccar
Universidade Federal do Rio de Janeiro
Rio de Janeiro, RJ — Brasil
maria.baccar.1@cp2.edu.br
0000-0001-6102-6667

Luciane de Souza Velasque
Universidade Federal do Estado do Rio de Janeiro
Rio de Janeiro, RJ — Brasil
luciane.velasque@uniriotec.br
0000-0002-4269-4755

Vanessa de Matos Leal
Prefeitura do Rio de Janeiro
Rio de Janeiro, RJ — Brasil
vanessamatosleall@gmail.com
0000-0001-5224-0071

2238-0345

10.37001/ripem.v14i3.3857

Received • 17/03/2024
Approved • 29/04/2024
Published • 20/08/2024

Editor • Gilberto Januario

Abstract: The National Common Curriculum Base made statistics teaching mandatory in basic education. The teaching degree in mathematics began to receive students who possibly had contact with statistics previously. Thus, the research sought to answer the following questions: What understandings about Stochastic[1] are there, and what use do these prospective teachers bring from school? A questionnaire on the topic was administered to a group of mathematics undergraduates in an introductory statistics course. Data analysis indicates that, although most undergraduates had already had some contact with statistics at school and realized its importance, this knowledge was restricted to procedural aspects of descriptive statistics, associated only with some types of graphs and the application of formulas in centrality measures. Talking about variability and uncertainty was also difficult. Besides, although everyone claimed to know probability, there was a lack of knowledge about statistical inference.

Keywords: Statistical knowledge. Stochastic. Statistics for Basic Education. Mathematics Teaching Degree.

La comprensión de la Estocástica aportada desde la educación básica por estudiantes de licenciatura en Matemáticas en un curso de introducción a la Estadística

Resumen: Con la Base Curricular Nacional Común, la enseñanza de estadística pasó a ser obligatoria en la educación básica. La licenciatura en matemáticas comienza a recibir estudiantes que, posiblemente, hayan tenido contacto con la estadística anteriormente. Surge entonces la pregunta: ¿qué conocimientos sobre la Estocástica[2] y su uso traen de la escuela estos futuros profesores? Se administró un cuestionario sobre el tema a un grupo de estudiantes de matemáticas que cursaban un curso de introducción a la estadística. El análisis de los datos indica que, aunque la mayoría de los estudiantes universitarios ya habían tenido algún contacto con la estadística en la escuela y se dieron cuenta de su importancia, este conocimiento se restringió a aspectos procedimentales de la estadística descriptiva, asociados sólo con algunos tipos de gráficos y la aplicación de fórmulas en medidas de centralidad. También se observó:

[1] The term *Stochastic* refers to the knowledge of statistics and probability, considered together.
[2] El término *Estocástico* se refiere al conocimiento de Estadística y Probabilidad, considerados en conjunto.

dificultad para hablar de variabilidad e incertidumbre y desconocimiento sobre inferencia estadística, aunque todos afirmaban conocer probabilidad.

Palabras clave: Conocimiento Estadístico. Estocástica. Estadística para la Educación Básica. Licenciatura en Matemáticas.

As compreensões de Estocástica trazidas da educação básica por licenciandos de Matemática em uma disciplina introdutória de Estatística

Resumo: Com a Base Nacional Comum Curricular, o ensino de estatística tornou-se obrigatório na educação básica. A licenciatura em matemática começa a receber estudantes que, possivelmente, tiveram contato com a estatística anteriormente. Surge, então, o questionamento: quais compreensões sobre a Estocástica[3] e sua utilização esses futuros professores trazem da escola? Aplicou-se um questionário sobre o tema a um grupo de licenciandos em matemática de uma disciplina introdutória de estatística. A análise dos dados indica que, embora a maioria dos licenciandos já tivesse tido algum contato com estatística na escola e percebesse sua importância, esse conhecimento restringia-se a aspectos procedimentais da estatística descritiva, associados apenas a alguns tipos de gráficos e à aplicação de fórmulas em medidas de centralidade. Observou-se, ainda, dificuldade para falar sobre variabilidade e incerteza, bem como desconhecimento sobre inferência estatística, embora todos afirmassem conhecer probabilidade.

Palavras-chave: Conhecimentos Estatísticos. Estocástica. Estatística para a Educação Básica. Licenciatura em Matemática.

1 Introduction

Statistics is an integral part of today's life. As Batanero and Godino (2005) highlight, it provides general methodological tools to "analyze variability, determine relationships between variables, optimally design studies and experiments, and improve predictions and decision-making in situations of uncertainty" (Batanero & Godino, 2005, p. 203, our translation).

In fact, as several authors point out (Batanero & Godino, 2005; Lopes, 2008, 2013; Cazorla, Kataoka, & Silva, 2010), since the end of the 20th century, statistics teaching has been increasingly incorporated into schools and universities, which occurs not only because of its instrumental nature but mainly because of the value that the development of statistical reasoning can have in a society that is constantly bombarded by massive information in the form of data and that must make decisions in environments of uncertainty.

Particularly in basic education, the task of teaching statistics is assigned to the teacher who teaches mathematics. As Lopes (2008, p. 61) points out, working with statistics and probability is relevant because the subjects allow the development of the ability to collect, organize, interpret, and compare data to obtain and substantiate conclusions, which is the basis of the scientific attitude. These actions enable the development of critical analysis of different scientific, technological, and/or social aspects, which is essential in citizenship education.

However, given the specific characteristics of this area of knowledge, adequate formation for prospective mathematics teachers must be provided for the satisfactory teaching of statistics in basic education. This formation can no longer focus solely on formulas, algorithms, and fixation exercises.

Therefore, besides basic education, it is essential to turn our attention to teacher

[3] O termo *Estocástica* refere-se aos saberes de Estatística e Probabilidade, considerados em conjunto.

education, particularly in mathematics, regarding statistics teaching. Costa and Pamplona (2011, p. 898) point out that this formation involves the prospective teachers appropriating the repertoire of professionals in the area, which means acquiring knowledge, practices, values, attitudes, and representations.

As several authors have discussed (Lopes, 2008; Viali, 2008; Costa & Pamplona, 2011), teacher education does not incorporate systematic work on stochastics, making it challenging to develop significant work on this topic in basic education. Statistics subjects in higher education are taught by professors from the most varied backgrounds, such as statistics, mathematics, engineering, and economics. Therefore, it is essential to create spaces for discussion on this subject within the mathematics teaching degree, not only sharing specific knowledge of probability and statistics but also "communicating and reflecting on experiences carried out in statistical education in elementary and secondary education and, above all, about the foundations for its achievement" (Costa & Pamplona, 2011, p. 899).

Thus, the research question is: What understandings[4] about stochastics and its use do these prospective teachers, who are now entering university, bring from school? Seeking to answer this question, this work presents part of a predominantly qualitative research, with reports on statistical knowledge carried out with mathematics undergraduates in 2022 in an introductory statistics course at the Federal University of the State of Rio de Janeiro. During this period, the authors were the professors responsible *for teaching statistics and probability in basic education to* a group of prospective mathematics teachers. Here, we analyze the material collected in the group's first meeting, particularly the participants' answers in the construction of a word cloud and a questionnaire on statistical knowledge.

We initially present the theoretical contribution of our research, in which we discuss some studies relating to the reality of statistics teaching in the education of mathematics teachers and the consequent need to understand the specific nature of statistics. Next, we highlight research that addresses students' statistical knowledge in basic education and preservice teacher education. Afterward, we will indicate the methodological procedures used and present a description and analysis of the data collected with a group of prospective teachers. Finally, we make considerations and show paths for future studies.

2 Literature review

The teaching of statistics has been incorporated into basic education since the end of the 20th century. Following the global trend, Brazil reflected this growing movement in its curricular guidelines and normative documents. Recently, according to Giordano and Vilhena (2020) and Giordano, Souza, Oliveira, and Lima (2022), Brazil has been experiencing a broad curricular reform after the publication of the National Common Curriculum Base (BNCC) in 2018 (Brasil, 2018).

The BNCC is a normative document that delimits the organic and progressive set of essential learning for basic education students. This document expanded the space dedicated to stochastic teaching and learning by determining that the topic must be addressed throughout basic education within the Mathematics curriculum component under the Probability and Statistics title. However, as Giordano et al. (2022) say, there is no counterpart in teachers' preservice and continuing education regarding the teaching and learning of stochastics.

[4] The term *understanding* is used here as the ability to comprehend, understand, or apprhend something. Therefore, considered a cognitive process (related to knowledge), in which the interpretation of a specific thing is necessary for it to be understood by the individual. Understanding is a concept that brings with it the idea of bringing to oneself, apprehending (Adapted from: https://www.significados.com.br/compreensao/ Accessed on: 01 Mar. 2024).

Although present in the normative documents that now govern basic education, according to Costa and Nacarato (2011), stochastics teaching was introduced late in Brazil, in relation to other countries and "without previous formation for teachers to work with these contents in basic schools" (Costa & Nacarato, 2011, p. 368), which is a concern that several authors also highlighted in their research too (Viali, 2008; Giordano & Vilhena, 2020; Giordano et al., 2022). Viali (2008), for example, argues that the low workload of Statistics and Probability subjects in mathematics teaching degree courses in Brazil makes prospective teachers feel unprepared. Bayer, Echeveste, Bittencourt, and Rocha (2005) argue that this lack of training of mathematics teachers to work on statistics means that they often choose not to address the theme in their classes.

This new reality is, in a way, a challenge for the mathematics teacher because, as Lopes (2013, p. 905) points out, "Statistics is a unique scientific discipline separate from mathematics; thus, its study objects are distinctly delineated" (Lopes, 2013, p. 905). To Cobb and Moore (1997), for example, statistics is a methodological discipline that offers other fields of study a set of ideas and tools for dealing with data, considering the ubiquity of variability. This fact differentiates statistics from mathematics and other sciences since "[Statistics] requires a different kind of thinking, because data are not just numbers, they are numbers with a context" (Cobb & Moore, 1997, p. 801).

In the same way, Gal and Garfield (1997, p. 6) list what they call critical differences between statistics and mathematics. In statistics, the context motivates procedures and gives meaning to interpretations. However, it is not always easy to define it (as in mathematics), and the use of technological tools can replace the need for precise application of calculations by hand. Furthermore, many statistical problems do not have a single solution as in mathematics, and precisely, one of the main objectives of teaching statistics is to develop students' ability to describe, judge, and make inferences about data or interpretations made about them.

Therefore, as Batanero (2001) and Shaughnessy (2006) pointed out, we cannot always transfer the general principles of mathematics teaching to statistics teaching. It is necessary to experiment and evaluate specific methods for this, adjusted to the specific nature of that science. As Carvalho (2015) suggests, that difference must be worked on very well in teaching degree courses so that prospective teachers can develop with their students this particular type of thinking and statistical reasoning, considering randomness and variability in opposition to mathematical determinism.

However, as Costa and Nacarato (2011) point out, most probability and statistics courses at the university level are still based on statistical rules and calculations or mathematized introductions of those contents. Therefore, those prospective teachers rarely have the chance to improve their statistical intuitions and those of their students or even address the topic in a broader and more contextualized way (Costa & Nacarato, 2011, p. 375-376).

Therefore, researching how statistics is taught in the preservice education of mathematics teachers is necessary to better understand how prospective teachers build this statistical knowledge that will later be worked on in basic education. Furthermore, they present possibilities of paths to take to change this reality, still based on mathematical determinism. A first step on this journey would be identifying what understandings about stochastics and their use those prospective teachers bring from basic education.

Research by Alencar and Furtado (2013) and Giordano and Coutinho (2019) seeks, for example, to identify statistics knowledge in high school students. Alencar and Furtado (2013) worked with students entering high school in a city in Ceará. The main objective was to analyze those students' level of understanding of statistical concepts and answer the following question:

Do elementary school students enter high school knowing what statistics is? The initial application of a questionnaire probed whether students had already heard of statistics (and what), whether they had used it, or whether they knew how this knowledge could help in everyday life. Although most students had heard of statistics, they did not know how to indicate what they had seen or learned. Therefore, they did not use it in their daily lives. However, they believed in the importance of learning this topic. The research concluded that students enter high school with little or no statistical knowledge and that even those who had some contact with the subject in elementary school do not remember what was studied.

Giordano and Coutinho (2019) conducted a study with third-grade high school students in a public school in São Paulo. Its objective was to diagnose these students' prior knowledge of statistics before carrying out activities related to developing a statistical project. As the authors indicated, in São Paulo's high school curriculum proposal, probability study is limited to two months in the second grade, while statistics is only addressed in the third grade.

Data was collected through a classroom questionnaire to identify students' level of knowledge and credibility about statistics and probability and their uses in everyday life. The results indicated that students give little importance to statistical concepts: "They do not believe in its usefulness in their daily lives or their professional future, and attest to their little knowledge of these concepts, probably implying that they do not remember having already studied the subject [...]" (Giordano & Coutinho, 2019, p. 396).

As the authors point out, probability is not associated with statistics, given the work done in different grades. Furthermore, probability is associated with combinatorics from a classical perspective, with a strong equiprobability bias, leaving aside its frequentist approach. Moreover, descriptive statistics are presented in a decontextualized and isolated way. Furthermore, there is little knowledge of statistical content related to the study of data variation, which would be extremely valuable in everyday decisions. The results thus indicated that students' previous knowledge was fragile and required more in-depth work.

Thus, the authors point out the need to start approaching statistical content in high school carefully and without assuming stable previous knowledge. Despite guidance on working with statistics in elementary education (Brasil, 2018), this content is rarely covered. They also point out the need to make students aware of the nature and relevance of the study of statistics and probability.

Correia (2018), Oliveira and Henriques (2014), Damin (2018), and Santos (2015) present reflections on statistical knowledge in higher education, focusing particularly on undergraduates in mathematics or pedagogy.

Correia (2018) conducted a study with mathematics undergraduates from a federal university in Pernambuco who attended statistics classes scheduled for the second period with a 60-hour workload. The syllabus for this subject contained content on descriptive statistics, counting techniques, and probability, with no statistical inference. The research objective was to identify the statistical knowledge presented by this group of students at the end of the course. The investigation showed that students had many difficulties in descriptive statistics, especially when constructing tables and graphs.

Also carried out with future basic education teachers, the research by Oliveira and Henriques (2018) focused on knowledge to work with statistical investigations. Within a didactics of mathematics subject, the authors developed a work proposal aimed at familiarizing undergraduates with central aspects of the statistical investigation method, such as question formulation, planning, and data collection. After the work and analysis of the written reflections produced, the researchers observed that most undergraduates identified aspects of statistical thinking, such as transnumeration, reasoning with models, and the integration of statistics and context, indicating the development of statistical knowledge beyond formulas and procedures. However, they did not identify data variation in the reflections, which may indicate that this concept was still new for the undergraduates.

Likewise, Damin's (2018) work, carried out with a group of mathematics undergraduates from a public university in Paraná, presents an analysis of the contributions of a statistics teaching project focusing on the development of statistical skills for teaching practice. Data collection occurred from tasks, learning narratives, and reflective memorials during the course. Within the scope of teaching knowledge, Damin (2018) focused on several subcategories, including disciplinary knowledge, which included statistical skills, that is, content-specific knowledge structured in the form of concepts, propositions, and theories. Concerning these statistical skills, data analysis indicated, as in Oliveira and Henriques (2018), that statistical thinking was developed, evidenced by the manifestation of elements such as transnumeration, statistical models, and the investigative cycle by undergraduates.

At the end of the course, the undergraduates also mastered interpreting graphs and tables, calculating measures of central tendency and dispersion, and understanding notions of sampling and estimation. However, the narratives and memorials also highlighted the lack of formative courses and the need to deepen specific and theoretical knowledge. The undergraduate students indicated that they had received flawed teaching in statistics in basic education, which was characterized by many gaps in the content and a mechanical focus that valued techniques and calculations.

Finally, research by Santos (2015) sought to understand early years prospective teachers' knowledge of statistics and teaching after taking subjects on the topic (Escola Superior de Educação de Santarém, Portugal) and during their supervised practicum. It also focused on teaching and learning based on statistical investigations.

Regarding knowledge of statistics, Santos (2015) concluded that the prospective teachers knew data organization instruments (tables and graphs) and measures of central tendency, although markedly procedural (calculations)

> The interpretations prospective teachers make of measures of central tendency are also varied. They mostly show a procedural understanding of the concepts, closely linked to applying calculation procedures and, once again, practicing techniques. The prospective teachers also show that the lack of knowledge or understanding of the measures impairs them from identifying students' errors in tasks addressing these concepts. [...] We conclude that much of the knowledge that prospective teachers reveal is automated and linked to procedures, such as applying concepts already learned and as mandatory steps (Santos, 2015, p. 201, our translation).

Thus, according to Santos (2015), those prospective teachers lack knowledge of statistics, making it necessary to reflect on and intervene in the formative processes, also seeking to develop statistical conceptual knowledge.

Therefore, the research results indicate that prospective teachers generally have gaps in the statistical knowledge they have gained from basic education. Often, they did not have contact with statistics at school, and when this occurred, teaching took a procedural approach associated with calculations to the detriment of conceptual aspects. However, the studies presented also suggest that working with prospective teachers focusing on evolving statistical thinking based on statistical investigations can reverse this situation, helping them understand the true nature of statistics.

3 Methodological procedures

The work presented here is part of an ongoing doctoral investigation that collected data from April to August 2022 in a mandatory and introductory Statistics course in a mathematics teaching degree course at the Federal University of the State of Rio de Janeiro (UNIRIO). This experience was carried out in Shared Teaching Practice (Giraldo & Menezes, 2016) subject, which was conducted by the higher education professor (second author) of statistics education and two basic education mathematics teachers (the first and third authors[5]). In this experience, the three teachers shared the learning space and the entire course organization, implementation, and evaluation to reduce the distance between the school and higher education in teacher qualification.

It was a face-to-face course, with a weekly workload of four hours, spread over two days (Tuesdays and Fridays, from 6 pm to 8 pm). Classes took place in a computer laboratory-type room, where students had access to computers at a rate of two per machine, which helped work with digital tools and access the Internet whenever necessary.

The research population consisted of the nineteen students present in the first class. These students will be indicated as E1 through E19, and the teachers will be called P1 (the second author and higher education professor), P2 (the first author and basic education teacher), and P3 (the third author and basic education teacher). The data was collected using the participant observation technique, and as a methodological procedure to analyze the participants' written production, we used a predominantly qualitative research approach.

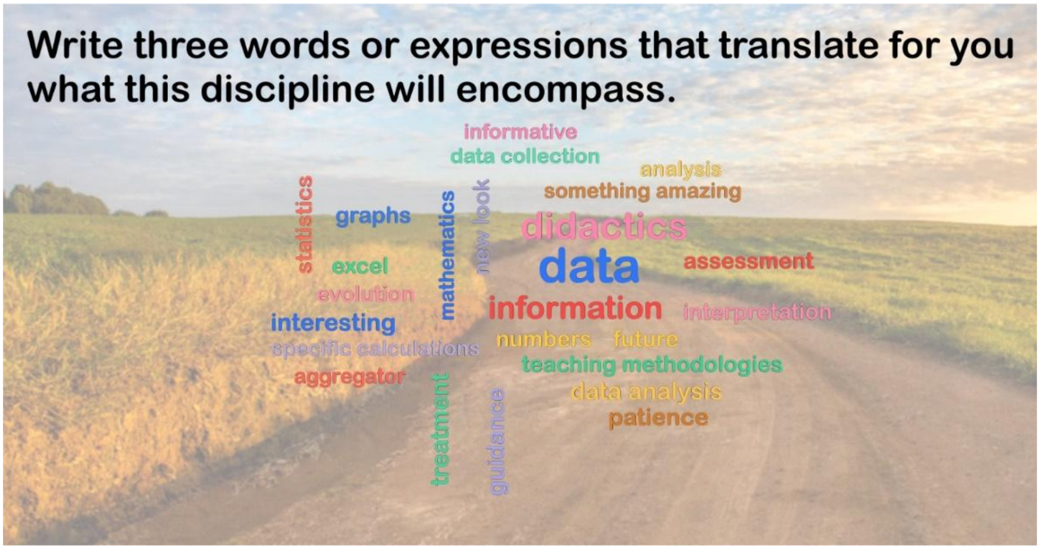
4 Data description and analysis

Our objective was to understand how these graduates arrived at university in terms of statistical knowledge, that is, what understanding of statistics and its use they brought from school. To this end, in the first class, a script was developed to raise awareness on the topic among the nineteen students present. Thus, after the reception, we applied a class dynamics using the word cloud technique. During the activity, students were asked to write three words or expressions that translated what the subject encompassed for them. This word cloud is presented in Figure 1 below.

It is interesting to note that the two most cited terms were *data* and *didactics*, which suggests that maybe the undergraduates were expecting to find some work with data analysis, as the subject was Statistics, however, already identifying the need to think about the approach focusing on teaching at school. Furthermore, there seems to be some previous understanding of Statistics being linked to working with data, given that the cloud has expressions such as *data collection, analysis, data analysis, treatment, interpretation, informatory,* and *information*. There also seems to be a positive predisposition towards what will be studied, represented in the expressions: *something that surprises, new perspective, evolution, future, orientation, interesting,* and *aggregator*. On the other hand, the association with mathematics is present through the expressions *mathematics, numbers, specific calculations*, and *graphs*, perhaps as a rescue of memories from basic education classes.

Figure 1: What do students expect to find in *Statistics and Probability for Basic Education*?

5The third author also participated in this Shared Teaching Practice experience with the aim of collecting data for her ongoing master's thesis, which focuses on the use of digital tools with mathematics prospective teachers for statistics teaching.



Source: Prepared by the authors (2022).

The word cloud was shared and discussed with the class, and then the three teachers introduced themselves, talking about their experiences in their respective segments (higher education and basic education). At that moment, P2 and P3 also discussed their respective research and told them how the classes would be, i.e., that the three teachers would share the teaching practice between them. Organizational aspects of the subject, such as Google Classroom, class format, and assessments, were also presented. Then, the students read and signed the Free and Informed Consent (FIC) corresponding to the research.

The teachers then asked the students to introduce themselves, and then the questionnaire was distributed in Google Forms, *Perceptions of Statistics in Preservice Mathematics Education*, made available through a link[6]. As the questionnaire was individual and there were no computers for all students, some chose to answer on their own mobiles. Afterward, the students answered the questionnaire, and the class ended with an informal chat over coffee and cake.

The virtual questionnaire was structured through twenty-four questions, fifteen objective and nine discursive, divided into five sections. Section 1 presents the research objectives, ensuring the anonymity of the respondents, and an Informed Consent Form, which is required to authorize data use. Section 2 contained three questions to collect personal information from the sample, such as the undergraduate course attended, whether they had experience in basic education teaching, and age. Section 3 was the largest, with twelve questions, and dealt with perceptions and previous knowledge of statistics. Section 4 discussed perceptions and knowledge about technological resources and had six questions. Section 5 was the last and brought the final considerations. There was only one open and optional question, giving space for the student to share additional information about statistics teaching and learning. In this study, our analysis will be related to sections 2, 3, and 5. Section 4 will be covered in another work.

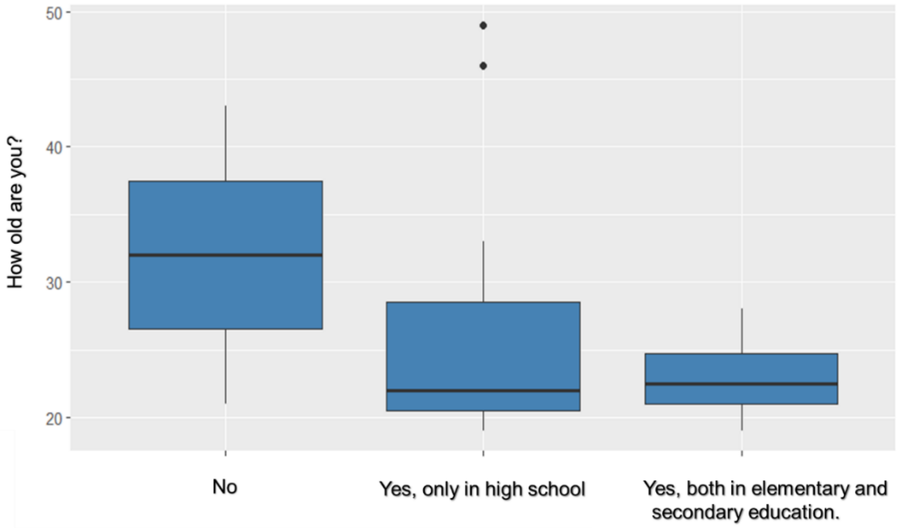
Based on the answers given in Section 2, we initially identified that all nineteen undergraduate students present were from the mathematics teaching degree course, with fourteen having no experience in teaching basic education, probably because it is a subject scheduled in the curriculum for the second course period. Among the other five with some

6Silva, A. S.; Velasque, L. S.; Baccar, M. H. M. & Leal, V. M. (2023). *Research questionnaire "Perceptions of statistics in preservice mathematics education"*. Available in: https://forms.gle/UjcpYySYfdYnEMvW7. Accessed on: June 23 2023.

experience, three had already worked only with middle school, and two with both segments (middle and high school). Concerning the age group, the majority, fourteen students, were between 19 and 25 years old, probably having entered university shortly after finishing high school. The other five were between 28 and 49 years old, and three were over 40 years old. Furthermore, we identified from the presentations at the first meeting that all five of these were already working. The statements also revealed that in the younger age group, some students were already in the job market, and some were taking the subject for the second or third time due to failure or abandonment.

Next, we have Section 3, dealing with previous perceptions and knowledge of statistics. The first question sought to identify whether the student had already had any contact with statistics content during basic education. The majority (seventeen) answered yes, with eleven saying they had had contact with statistics only in high school. Only two students stated that they had not seen statistics in basic education. Graph 1 shows the crossing of this information with the ages of the undergraduates:

Graph 1: Age of undergraduates x statistics in basic education



Source: Prepared by the authors (2023).

We can observe that while two students aged 28 to 43 did not have contact with statistics in basic education, the younger ones indicated that they had, which may indicate that schools are already implementing the guidelines given by BNCC (Brazil, 2018) and working with statistics throughout basic education.

We then asked them to identify in which area(s) of education these students had experienced working with statistics in basic education. The predominant area was mathematics and, except for the two students who had had no contact with statistics at school, all seventeen others stated they had worked on the topic before. Next came natural sciences, with eleven students, and humanities, with seven. Only one student mentioned having worked with statistics in languages. The fact that some students noticed the use of statistics in other areas of knowledge, in addition to mathematics, caught our attention, as it could already indicate some interdisciplinary work in basic education.

The next question was discursive and asked whether and how the students, beyond the school environment and before entering university, had identified the presence of statistics in the world around them. All students demonstrated observing the use of statistics on a daily basis, giving examples of situations in different areas in which there was, in some way, data collection or processing. Ten students also indicated that this information was disseminated through the media. Table 1 presents the contexts highlighted by the prospective teachers in which statistics is present:

Table 1: Contexts in which statistics is present

Area	Answers	Area	Answers
Elections	6	Health	2
Sports	4	Accidents (work, traffic)	2
Opinion/satisfaction surveys in general	4	Crimes	2
IBGE[7] research	3	Meteorology (weather forecast)	2
Games	3	Environment and animals	2
Economy (financial market, trade)	3	Insurance	1

Source: Prepared by the authors (2023).

When identifying these contexts, students seem to strongly associate statistics with the presentation of rates and indices in tables or graphs, as we can see in E11's response:

> *[...] We are exposed to data at all times, whether in its most graphic form [...] or in reports with indexes and presentation of statistical data.*

Other students also indicate, in some way, that this data can be used for decision-making, as we see in E19's response:

> *[...] I noticed when making a decision, looking for references so that I could choose the best option.*

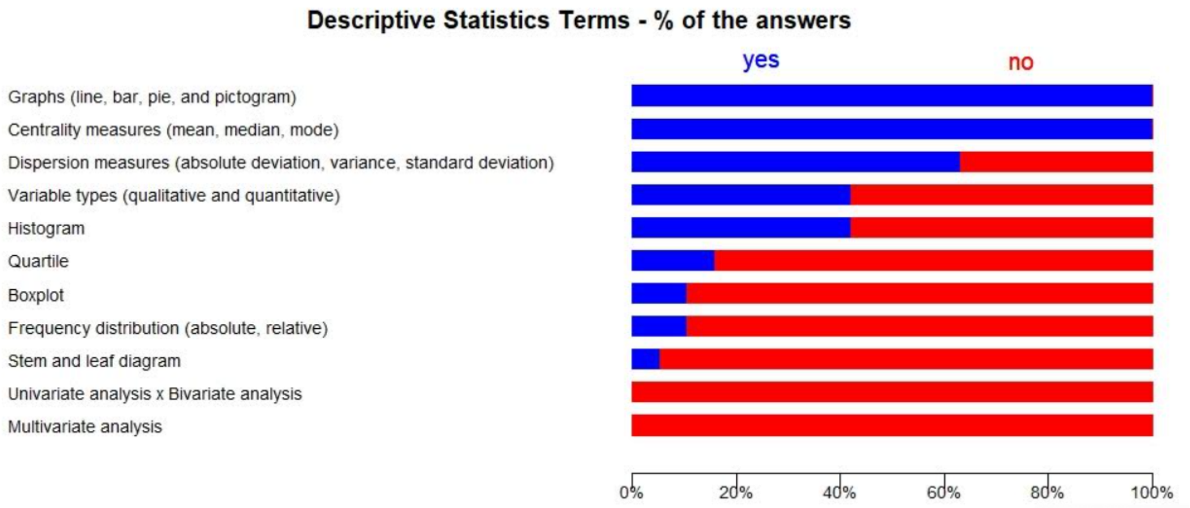
Student E13 highlights the importance of critically examining the information we receive, as statistical data may be wrong or show a trend. Student E14 is the only one who comments on the variability in a data set. Thus, in general, we can say that these students identify the presence of statistics in their daily lives but still strongly associate it with descriptive statistics.

The next question sought to identify which terms of statistics the students already knew. In a list of twenty-three expressions, the student should indicate YES if they already knew the term or NO otherwise. Let us observe, below, the results referring to descriptive statistics in Graph 2.

All students had already had some contact with graphs (line, bar, sector, and pictogram) and centrality measures (mean, median, mode), and many (12) knew dispersion measures, probably because they had seen them in basic education. Most people did not know histograms and types of variables. Few students knew boxplots, stem and leaf diagrams, quartiles, and frequency distributions. And none of them knew the expressions: *univariate*, *bivariate*, and *multivariate analysis*. Regarding probability, we have the results in Graph 3.

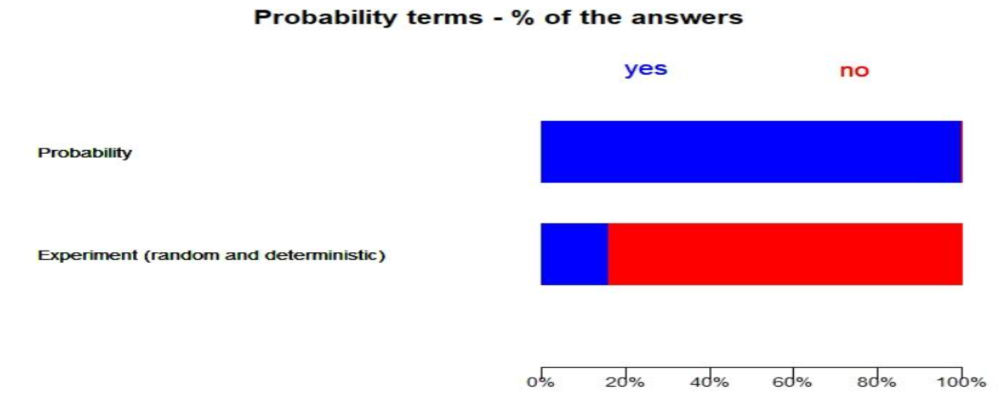
Graph 2: Descriptive statistics terms

[7] IBGE is the acronym for *Brazilian Institute of Geography and Statistics.*



Source: Prepared by the authors (2023).

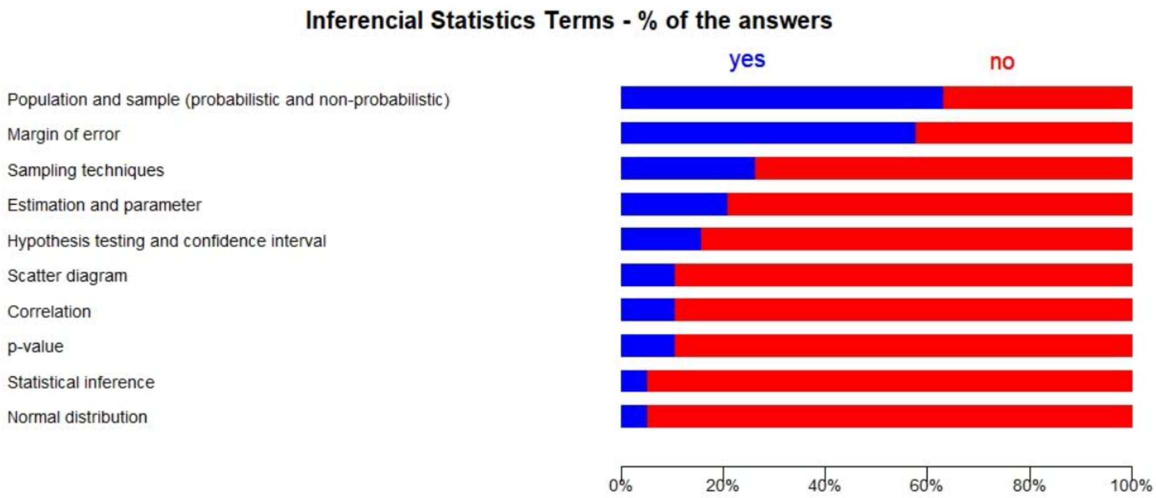
Graph 3: Probability terms - % of the answers



Source: Prepared by the authors (2022).

All undergraduates indicated to have had some contact with probability, but only three had heard about random and deterministic experiments. Regarding inferential statistics, the results are in Graph 4.

Graph 4: Inferential statistics terms



Source: Prepared by the authors (2022).

We can see that, for most of the terms presented in inferential statistics, most students indicated a lack of knowledge, which, in a way, was already expected, as these contents are not covered in school. In particular, only one student claimed to know the expression *statistical inference*. However, two points caught our attention: twelve students knew what population and sample were, and eleven knew the expression *margin of error*, the latter probably from the media, in the dissemination of research in an election year.

The following four discursive questions asked students to explain what the terms *variability*, *randomness*, *uncertainty (or working with uncertainty*), and *statistical inference* meant to them. As they were optional questions, not all participants responded.

Regarding variability, the fourteen answers were very vague, in general, signaling understanding as *something that can change values*, as in E5's and E6's answers:

> *The ability of variation of a result in different analyses of the same data* (E5).
>
> *The ability of something to take on different values* (E6).

Some prospective teachers suggested something more focused on a discrepancy in values, as E17 states:

> *The "space" between the data, which in most cases are numbers, how much the interval varies from one to another.*

However, none explicitly associated variability with the degree of dispersion of the data in a set. Only E5 associated variability with the possibility of variation in the data obtained.

Regarding randomness, there were fourteen answers as well. Seven of them related the theme, in some way, to the idea of different possibilities of occurrence, over which there was no initial control, as we can see in the responses of E7 and E12:

> *The notion that the result of an analysis may vary due to factors beyond the analyst's control* (E7).
>
> *It is the ability of something to assume a random and unpredictable value* (E12).

One of the answers (E11), in fact, roughly defined random as the opposite of something that can be determined.

For the meaning of *uncertainty* (or work with uncertainty), we got fifteen answers. Three students, E3, E7, and E12, associated this concept with randomness. Two others, E8 and E18, related uncertainty to the difficulty of making generalizations. And two others, E13 and E17, associated uncertainty with the notion of estimate. However, none associated uncertainty with an estimate that quantifies the reliability of a measurement result.

Only two students presented any answer regarding the meaning of *statistical inference*: E3 said it was the ability to analyze a set of data, and E17 said it was a set of samples. As it was possible to observe, and as the students indicated when answering a previous question, they really did not know what statistical inference meant.

The next four questions dealt with specific statistics content, the first three on centrality measures, and the last on range and samples in data sets. In the first of these questions, the results of a salary survey conducted with thirty employees were presented as a frequency table. One of the employees had a salary well above what the others received, which was a discrepant value. We then asked the students which centrality measure –mean, mode, or median– should be chosen to better characterize this data set.

Although we expected most students to opt for the mean, only two chose this measure. The rest of the answers were divided between median (9) and mode (8). We consider this result interesting, as it demonstrates that students already perceived the median as a more robust measure of centrality. However, it is essential to highlight that, throughout this first class, during a conversation about what the students had seen in statistics, P1 made comments about centrality measures and presented an example very similar to the one in the questionnaire, which may have induced the students not to select the mean as their answer.

The following question also presented a table, which included the travel time for a given route for three bus companies (A, B, and C), with ten measurements for each of them, as shown in Figure 2:

Figure 2: Travel time of bus companies[8]

COMPANY	TRAVEL TIME (in minutes)										SUM
A	10	11	12	12	13	13	13	14	16	18	132
B	10	12	13	14	14	15	15	15	15	16	139
C	13	13	13	14	14	14	14	14	15	15	139

Source: Prepared by the authors (2023).

Three situations were indicated, and students had to select which company they would choose in each. Table 2 presents the situations and answers obtained:

Table 2: Companies chosen according to each given situation

Situation	A	B	C
My priority is the minimum average travel time.	18	1	0
I will use the company that presents the least variation in journey times.	1	0	18
I prefer to choose the company that offers the same journey time the most times.	1	1	17

Source: Prepared by the authors (2022).

We sought to observe students' understanding of mean, range, and mode concepts, as they often know how to calculate these values but do not understand what they mean. Once again, we were surprised, considering that most students demonstrated knowledge of what was required in the three situations.

Then, in a discursive question, we asked the graduates if company A stopped offering the path, which of the other two companies, B or C, would be chosen. We asked them to justify this choice, but there was no correct answer in this case. It was interesting to observe the students' answers, mainly due to the justifications. Sixteen of them chose C, as it presented less variation in travel time, as explained by E5:

> *C, because it completes the path with less variation of time, which increases my control over when I will reach the end of it.*

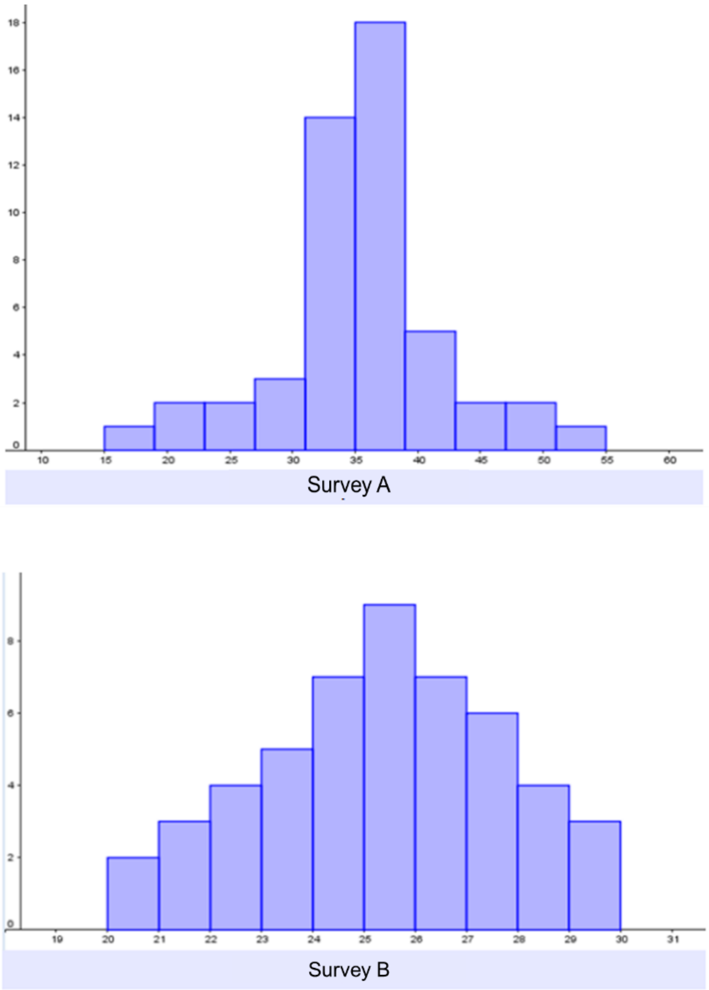
Three chose B, as it had a shorter minimum time, as we observed in E9's response:

[8] A column with the sum of the times for each company is shown in this image. These values would be used to facilitate the calculation of averages, used later.

B because despite presenting a slightly greater variation than company C, it also has the option of completing the path in less time.

The last question in Section 3 presented two surveys (A and B) about age in populations of 50 individuals, with the results presented using histograms. In survey A, ages ranged from 15 to 55 years old, with a high concentration between 30 and 40 years old. In B, they ranged from 20 to 30 years, with a more homogeneous distribution, as shown in Figure 3:

Figure 3: Histograms of Surveys A and B



Source: Prepared by the authors (2022).

Samples would be selected for analysis, and five situations would be presented for the student to indicate which research satisfied each of them (A and/or B or neither), as shown in Table 3:

Table 3: Situations for analysis in research A and B

Situation	A	B	None
Greater age range of the population.	16	3	0
It is possible to collect two random samples of the same size with different means.	14	9	2
A random sample of 10 elements can have the same mean as another random sample with 15 elements.	8	8	5

A sample with five elements may have an average age well below the population average.	10	9	1
The ages of the population elements do not form a homogeneous set of data, focusing specifically on a few age groups.	12	6	1

Source: Produced by the authors (2023).

This question was largely about the topic *samples*, which probably were not worked on in basic education. We observed that the students answered easily to the first statement, which dealt with range, and most of them also demonstrated knowledge about what was covered in the last situation (data homogeneity). However, the answers to specific sample questions varied widely, and we had difficulty identifying whether students already knew the subject. These three questions could probably have been designed differently to make it easier to identify this knowledge.

Ultimately, we asked if the undergraduates would like to share additional information about their perceptions of teaching and learning statistics (feelings, difficulties, experiences, or other considerations). We only got four answers, but they were quite interesting. For example, E11 and E13 find statistics to be a very important tool for developing a critical eye on all types of research that comes to us. As E11 argues:

> *Learning statistics is essential to living in society. It is a very important tool to avoid being treated as a mass of maneuver by those more powerful than us.*

E13 adds, asking that teachers can

> *show students how to understand, at least, the surveys that always appear in the media, knowing how to identify hits and failures, understand graphs, etc.*

Student E5 raises a very relevant question about how statistics is seen at school, like something

> *mechanical and abstract, being completely outside students' realities. This keeps students away from statistics and makes the subject boring for them [...] and harms their learning.*

Below, we present final considerations about the work.

5 Final considerations

This work sought to answer the initial question: What understandings about stochastics and its use do prospective mathematics teachers who now arrive at university bring from school? The word cloud and a questionnaire on the topic, both produced during the first meeting with the prospective teachers, were then analyzed.

Most prospective teachers, aged between 19 and 25, had recently completed basic education. They claimed to have seen statistics at school, always in mathematics, however, in some cases, also in natural sciences and humanities. Contrary to what the works of Alencar and Furtado (2013) and Giordano and Coutinho (2019) pointed out, these prospective teachers brought back memories of statistics teaching in basic education. Perhaps the mandatory approach to this topic in schools, established by the BNCC (Brasil, 2018), has made statistics actually present in school curricula. Furthermore, the presence of statistics in several curriculum components could also indicate some interdisciplinary work in basic education.

The undergraduates have demonstrated that they can identify statistics as a data analysis science and that they must have a specific look at how to address it in basic education. They presented –just like the participants in Alencar and Furtado's (2013) research– a positive predisposition toward learning the topic when pointing out the importance of this knowledge in understanding the world around them.

However, at the same time, as occurred in Damin (2018) and Santos (2015), students indicated that statistics is somehow linked to mathematics, also associating it with specific calculations and graphical representations. Indeed, all students noticed the presence of statistics in their daily lives in the most diverse areas. However, they always associated this presence with rates, indexes, and research –with tables and graphs–, in a strong appeal to descriptive statistics.

As in Giordano and Coutinho's (2019) investigation, students demonstrated knowledge of specific graphs (line, bar, sector, and pictogram) and of centrality measures brought from school, which may have influenced this look directed at descriptive statistics. Thus, data analysis indicated that, although most prospective teachers had already had some contact with statistics at school and realized the need for its daily use, this knowledge was restricted to very procedural aspects of descriptive statistics.

Even though all claimed to have had contact with probability, few knew what random experiments were. Similarly, few associated it with the measurement of uncertainty or decision-making, reinforcing the result Giordano and Coutinho (2019) also presented on teaching probability without connection to statistics at school.

The students also had difficulty discussing variability and uncertainty and lacked a complete understanding of statistical inference. As Oliveira and Henriques (2018) and Giordano and Coutinho (2019) pointed out, students had great difficulty understanding the notion of data variation and its association with the notion of uncertainty, both present in statistical thinking. However, it is important to highlight that students could work with the variation in data and justify their answer choices in the discursive question with multiple answers. Likewise, in the discursive question regarding centrality measures, they presented a conceptual understanding of the measures, going beyond the procedural (calculations).

Finally, we must highlight that this experience is part of ongoing doctoral research that seeks to identify in prospective mathematics teachers, throughout an introductory statistics course, the changes in understanding statistics and what its teaching involves in basic education. Identifying how prospective mathematics teachers conceive statistics and its teaching based on the knowledge brought from school can help us rethink university courses, directing them toward a focus that favors teacher education. We believe that further research on this topic could help in the search for this direction.

As future directions, we intend to observe, in larger samples of undergraduate students, how this transition from basic education to university occurs in terms of statistical knowledge, aiming to expand the dialogue between these two educational levels.

References

Alencar, A. C. & Furtado, A. B. (2013). O nível de conhecimento estatístico de alunos ingressos no ensino médio. In: *Anais do VII Congresso Ibero-Americano de Educação Matemática, Actas* (pp. 2118-2129). Montevideo.

Batanero, C. (2001). *Didáctica de la estadística*. Granada, Universidad de Granada.

Batanero, C. & Godino, J. D. (2005). Perspectivas de la educación estadística como área de investigación. In: R. Luengo. (Ed.). *Líneas de investigación en Didáctica de las Matemáticas* (pp. 203-226). Badajoz: Universidad de Extremadura.

Bayer, A.; Echeveste, S.; Bittencourt, H. & Rocha, J. (2005). Preparação do Formando em Matemática-Licenciatura para lecionar Estatística no Ensino Fundamental e Médio. In: *Anais do V Encontro Nacional de Pesquisa em Educação em Ciências* (pp. 508-518). Bauru, SP.

Brasil. Ministério da Educação. Secretaria de Educação Básica. *Base Nacional Comum Curricular*. Brasília, DF: MEC/SEB.

Carvalho, A. (2015). A importância do ensino de estatística na formação inicial do professor de Matemática. In: *Anais do XIX Encontro Brasileiro de Estudantes de Pós-Graduação em Educação Matemática* (pp. 1-9). Juiz de Fora, MG.

Cazorla, I. M.; Kataoka, V. Y. & Silva, C. B. (2010). Trajetória e Perspectivas da Educação Estatística no Brasil: um olhar a partir do GT-12. In: C. E. Lopes; C. Q. S. Coutinho & S. A. Almouloud. (Org). *Estudos e Reflexões em Educação Estatística* (pp. 19-44). São Paulo, SP: Mercado das Letras.

Cobb, G. W. & Moore, D. S. (1997). Mathematics, Statistics, and Teaching. *The American Mathematical Monthly*, *104*(9), 801-823.

Correia, A. A. G. (2018). *Investigando o conhecimento estatístico de alguns licenciandos em matemática na disciplina de estatística*. 2018. 52f. Trabalho de Conclusão de Curso (Licenciatura em Matemática). Universidade Federal de Pernambuco. Caruaru, PE.

Costa, A. & Nacarato, A. M. (2009). A Estocástica na Formação do Professor de Matemática: percepções de professores e de formadores. *Bolema - Boletim de Educação Matemática*, *24*(39), 367-386.

Costa, W. N. G. & Pamplona, A. S. (2011). Entrecruzando fronteiras: a Educação Estatística na formação de professores de Matemática. *Boletim de Educação Matemática*, *24*(40), 897-911.

Damin, W. (2018). *A educação estatística e a formação de professores de matemática: contribuições de um projeto para a constituição dos saberes docentes.* 2018. 148f. Tese (Doutorado em Ensino de Ciência e Tecnologia). Universidade Tecnológica Federal do Paraná. Ponta Grossa, PR.

Gal, I.; Garfield, J. (Ed.). (1997). *The Assessment Challenge in Statistics Education*, Amsterdam: IOS Press / International Statistical Institute.

Giordano, C. C.; Souza, F. D. S.; Oliveira, P. C. & Lima, R. F. (2022). Formación de profesores que enseñan estadística en Brasil: nuevos desafíos en el escenario pospandémico. In: A. Salcedo & D. Díaz-Levicoy. *Formación del profesorado para enseñar estadística: retos y oportunidades* (pp. 483-502). Talca: Centro de Investigación en Educación Matemática y Estadística.

Giordano, C. C. & Coutinho, C. Q. S. (2019). Conhecimentos prévios de alunos do ensino médio a respeito de estatística: uma análise coesitiva e de similaridades. In: *Anais de X Colloque International Analyse Statistique Implicative X International Conference Statistical Implicative Analysis Belfort* (pp. 380-403). Belfort, France.

Giordano, C. C. & Vilhena, V. D. M. (2020). Educação estatística e a formação de professores que ensinam matemática no Brasil. *Brazilian Journal of Development*, *6*(12), 104137-104148.

Giraldo, V. & Menezes, F. (2016). Práticas Docentes Compartilhadas. In: *Anais de VIII Seminário de Pesquisa em Educação Matemática* (pp. 279-291). Rio de Janeiro, RJ.

Lopes, C. E. (2013). Educação Estatística no Curso de Licenciatura em Matemática. *Bolema*, *27*(47), 901-915.

Lopes, C. E. (2008). O Ensino da Estatística e da Probabilidade na Educação Básica e a Formação dos Professores. *Cadernos CEDES*, *28*(74), 57-73.

Oliveira, H. & Henriques, A. (2014). Um quadro de análise do conhecimento estatístico para ensinar de futuros professores. *Boletim GEPEM*, *64*, 104-115.

Santos, R. F. M. (2015). *O Conhecimento de Estatística e da Sua Didática De Futuros Professores.* 2015. 440f. Tese (Doutorado em Educação), Universidade de Lisboa. Lisboa.

Shaughnessy, J. M. (2006). Research on Students' Understanding of some big concepts in Statistics. In: G. F. Burrill & Portia C. Elliott. (Ed.). *Thinking and Reasoning with data and chance* (pp.77-98). Yearbook, National Council of Teachers of Mathematics, 68th.

Viali, L. (2008). O Ensino de Estatística e Probabilidade nos Cursos de Licenciatura em Matemática. In: *Anais de 18º Simpósio Nacional de Probabilidade e Estatística* (pp. 1-7). Águas de Lindóia, SP.