Evaluating Lexical Ambiguities About Brazilian Students in The Fifth Year of Elementary School Towards Statistical Graphs and Tables

Ailton Paulo de Oliveira Júnior
ailton.junior@ufabc.edu.r
https://orcid.org/0000-0002-2721-7192
Universidade Federal do ABC (UFABC)
Santo André, Brazil.

Luzia Roseli da Silva Santos
luziaroselidasilvasantos@gmail.com
https://orcid.org/0000-0001-6930-9215
Universidade Federal do ABC (UFABC)
Santo André, Brazil.

Fátima Aparecida Kian
escritores6@gmail.com
https://orcid.org/0000-0003-0105-7335
Universidade Federal do ABC (UFABC)
Santo André, Brazil.

Abstract

When words that are part of everyday language are used differently from the way they are used in formal classroom language, it is said to present lexical ambiguity. Thus, this study aims to understand possible lexical ambiguities regarding knowledge about representations of statistical graphs and tables of 74 students in the fifth year of elementary school (approximately 10 years) of a municipal school in Barueri, São Paulo, Brazil. This research was exploratory, with a qualitative and quantitative approach, in which data were collected through a questionnaire available on Google Forms. An Automated Content Analysis was carried out using the IraMuTeQ software (R Interface for Multidimensional Text and Questionnaire Analysis) performing multivariate statistical analyzes (Descending Hierarchical Classification - DHC and Factorial Analysis by Correspondence – CFA). The results show the predominance of words and verbal expressions of the common language related mainly to the intuitive meaning and its use in everyday life. It was also noticed indications of the formal construction of tables and graphs associated with guidelines carried out in the classroom.

Keywords: Statistical Tables and Graphs. Elementary School. Automated Content Analysis. Multivariate Textual Analysis.

Evaluando Ambigüedades Léxicas sobre el Conocimiento de Estudiantes Brasileños en el Quinto Año de la Escuela Primaria sobre Gráficos y Tablas Estadísticas

Resumen

Cuando las palabras que forman parte del lenguaje cotidiano se usan de manera diferente a como se usan en el lenguaje formal del aula, se dice que presentan ambigüedad léxica. Por lo tanto, este estudio tiene como objetivo comprender posibles ambigüedades léxicas relacionadas con el conocimiento sobre representaciones de gráficos y tablas estadísticas de 74 estudiantes del quinto año de la escuela primaria (aproximadamente 10 años) de una escuela municipal en Barueri, São Paulo, Brasil. Esta investigación

Ailton Paulo de Oliveira Júnior; Luzia Roseli da Silva Santos & Fátima Aparecida Kian

fue exploratoria, con enfoque cualitativo y cuantitativo, en la que los datos fueron recolectados a través de un cuestionario disponible en Google Forms. Se realizó un Análisis de Contenido Automatizado utilizando el software IraMuTeQ (Interfaz R para texto multidimensional y análisis de cuestionarios) realizando análisis estadísticos multivariados (Clasificación Jerárquica Descendente - DHC y Análisis Factorial por Correspondencia – CFA). Los resultados muestran el predominio de palabras y expresiones verbales del lenguaje común se relacionó principalmente con el significado intuitivo y su uso en la vida cotidiana. También se notaron indicios de la construcción formal de tablas y gráficos asociados a las directrices realizadas en el aula.

Palabras clave: Tablas y Gráficos Estadísticos. Escuela Primaria. Análisis de Contenido Automatizado. Análisis Textual Multivariado.

Avaliando Ambiguidades Lexicais sobre o Conhecimento de Alunos Brasileiros do Quinto Ano do Ensino Fundamental sobre Gráficos e Tabelas Estatísticas

Resumo

Quando palavras que fazem parte da linguagem cotidiana são usadas de maneira diferente da utilizada na linguagem formal da sala de aula, diz-se que apresenta ambiguidade lexical. Dessa, forma, esse estudo destina-se a compreender possíveis ambiguidades lexicais referentes ao conhecimento sobre representações de gráficos e tabelas estatísticas de 74 alunos do quinto ano do ensino fundamental (aproximadamente 10 anos) de uma escola municipal em Barueri, São Paulo, Brasil. Esta pesquisa foi do tipo exploratório, de abordagem qualitativa e quantitativa, na qual os dados foram coletados por meio de questionário disponibilizado no *Google Forms.* Foi realizada uma Análise de Conteúdo Automatizada, por meio do *software* IraMuTeQ (Interface R para Texto Multidimensional e Análise de Questionário) realizando análises estatísticas multivariadas (Classificação Hierárquica Descendente – DHC e Análise Fatorial por Correspondência – CFA). Os resultados mostram a predominância de palavras e expressões verbais da língua comum relacionado principalmente ao significado intuitivo e de sua utilização no dia a dia. Também se percebeu indicações da construção formal de tabelas e gráficos associadas a orientações realizadas em sala de aula.

Palavras chave: Tabelas e Gráficos Estatísticos. Ensino Fundamental. Análise de Conteúdo Automatizada. Análise Textual Multivariada.

Introduction

The inclusion of statistics in curriculum guidelines around the world in elementary education presents several challenges for the school system. In Brazil, the National Common Curricular Base – BNCC (Brazil, 2018), indicates the need to carry out investigative plans, as it is considered that it will help in understanding the role of statistics in the daily lives of students. Thus, the reading, interpretation and construction of tables and graphs play a fundamental role in the students' education.

We believe that graphical and tabular languages are very important, both from a mathematical point of view and for their use in everyday life and, like all languages, it has its own characteristics that students must learn to use it correctly. In addition, by means of graphs and tables, information from an everyday situation can be represented, starting from the

formulation of research questions about a certain content, creating situations to stimulate and make students realize their importance as elements of great informative content. and citizen.

Furthermore, Lemke (1990) noted that, as students begin to be exposed to the vocabulary of specialized content, they still do not speak the language of that discipline. Furthermore, it connects what they are hearing with what they have heard and experienced previously. If a word that is commonly used in everyday life is also used in a technical domain, students hearing the word for the first time in a lesson can incorporate technical usage as a new facet of the word features, they already know.

For Barwell (2005), when the use of specific words that refer to statistical concepts is similar to everyday words (informal language), it can cause students to make incorrect associations between words they know and those that sound and look similar, however, they have different meanings. These words are said to have lexical ambiguity.

Makar and Confrey (2005), in their study on the use of non-standard language by trainee teachers to discuss the concept of variation, concluded that neglecting the use of non-standard language by students makes the subject seem more difficult.

Rangecroft (2002) also raised the issue of the use of words in statistics that have different meanings, whether they are used in everyday or specialized language, concluding that when identifying a problem, the first step is to solve it. If we become more aware of the possibilities of misunderstandings arising from language difficulties, we may be able to recognize them and make the necessary explanations.

Therefore, in this study, we asked students in the fifth year of elementary school at a municipal school in Barueri, São Paulo, Brazil, what are statistical graphs and tables. Through the answers indicated by this group of students, we seek to assess whether there are lexical ambiguities (contrasting the colloquial meanings with the statistical meanings) through multivariate textual analyzes that indicate their knowledge and, based on these results, indicate possible strategies for the teaching of statistical tables and graphs directed to elementary school.

Theoretical Framework

Lexically ambiguous words have a different meaning in scientific fields or in everyday usage (Richardson, Dunn, and Hutchins 2013; Dunn et al. 2016). In scientific language, Kanji (2006) explains that, in statistics and research, a 'graph' or a 'table' is used to summarize data.

Thus, to represent data through graphs or tables, you must learn to: 1) select the appropriate graph or table to summarize the data; 2) interpret them; 3) identify the poorly prepared data representation, substantiating them.

In exploratory data analysis or modeling (Tukey, 1977), tables and graphs occupy a prominent place at the beginning and end of the study. In the development of a statistical analysis, the data, the sources, the unusual characteristics are initially explored, visualizing the data in tabular and/or graphic form; then, when carrying out the further analysis and finalizing the study, it is necessary to communicate the results of the analysis to the target audience in a complete and concise manner.

For Glazer (2011), data interpretation is the ability to build graphs and tables that are essential in the process and production of science, that is, organizing data in graphs and tables is an invaluable method of data representation to find relationships between variables in order to determine patterns, properties and relationships.

Furthermore, statistics is currently considered a transversal and relevant discipline for its contributions in different areas of knowledge (scientific, social and humanistic) (Molina-Portillo et al., 2019). Arteaga et al. (2011) say that this fact is reflected in the large amount of statistical information (in the form of tables, graphs and statistical summaries) that is observed in different media (television, internet, newspapers, etc.).

We also emphasize that it is important to remember its usefulness in various activities of everyday life, for example, the presence of summary measures and graphic and tabular representations in the media (Gal, 2011; Mcconway, 2016; Jurečková and Csachová, 2020).

These elements are part of the statistical culture, that is, the right that every citizen has to read and interpret the statistical information they access in situations of everyday, civic and professional life (Del Pino and Estrella, 2012). On the other hand, Weiland (2017) indicates that the statistical culture goes beyond critically evaluating the information to which one has access, and should be a lens to see the world.

Furthermore, according to Muñoz et al. (2020), a basic skill necessary to achieve a statistical culture is the ability to read and interpret statistical graphs and tables, as these formats are privileged resources for grouping and synthesizing large amounts of information in an efficient and visually attractive way.

Thus, statistical charts and tables are not only widely used by the mass media, but are also an important part of the dissemination of official statistics and investigative reports in a large number of areas of knowledge (Arteaga et al. 2011; Estrella, 2014).

Methodology

This research is exploratory, with a qualitative and quantitative approach, in which data were collected through a questionnaire made available on Google Forms. An Automated Content Analysis - ACA was performed using the IraMuTeQ software (R Interface for Multidimensional Text and Questionnaire Analysis) through multivariate statistical analyzes (Descending Hierarchical Classification - DHC and Correspondence Factor Analysis – CFA) to understand possible ambiguities lexical texts referring to the knowledge of fifth-year elementary school students from a municipal school in Barueri, São Paulo, Brazil (10 to 11 years old) about graphs and statistical tables.

Sample Characterization

In the first part of this study, we established a brief profile of the 74 students participating in the research and that we considered to bring elements both from the personification of the group and in the textual analysis referring to the students' knowledge about statistical tables and graphs. Table 1 shows the distribution of absolute (number of students) and relative (percentage) frequencies of the statistical variables: gender; age years; I like school and mathematics.

The research participants totaled 74 students in the fifth year of Elementary School (aged between 9 and 11 years old) and 62.2% were female. The average age of students is 10.51 years, within the expected age range for this level, with a standard deviation of 0.53 years; a little more than half of them are 11 years old (51.4%).

Still on the age of the students, by calculating the coefficient of variation, which is determined by the ratio between the standard deviation and the mean, the group has low variability or dispersion, that is, the observed values of the ages of the students have little dispersion around the average, showing to be homogeneous.

Table 1 - Profile of students in the fifth year of elementary school at a school in the city of Barueri, São Paulo, Brazil.

Genre	Number of students	Percentage
Female	46	62.2%
Masculine	28	37.8%
Age	**Number of students**	**Percentage**
9 years	1	1.4%
10 years	35	47.3%
11 years	38	51.4%
Mean (Standard deviation):	10,51 years (0,53 years)	
Minimum maximum:	9 years – 11 years	
Median:	11 years	
Do you like school?	**Number of students**	**Percentage**
Yes	71	95.9%
Not	3	4.1%
Do you like math?	**Number of students**	**Percentage**
Yes	59	79.7%
Not	15	20.3%

Source: Prepared by the authors.

It is still verified that almost all students (95.9%) declared to like the school, thus having, even in times of the Covid-19 pandemic, a positive relationship in relation to what it offers. As for mathematics, it indicates that 79.9% of students still like this subject, that is, it seems to us that children have not yet created resistance in relation to mathematics and that it can be used in the teaching and learning process.

Methodological procedures

For this work, we established the following research question: How does the identification of the statistical language used by elementary school students contribute to the understanding of statistical concepts?

The data from this research were analyzed using the technique called Automated Content Analysis - ACA, which according to Chizzoti (2006) is a technique derived from Content Analysis - CA, incorporating the technological possibilities in data processing, supported by software and programs statistics for statistical inference.

AC seeks to interpret the different forms of the content of one, adopting systematic norms to extract thematic meanings (Chizzoti, 2006). In the present study, these thematic meanings were given off when the students conceived about graphs and statistical tables through their day-to-day knowledge and/or what they learned at school. For Chizzoti (2006), the central idea of this analysis technique is to relate the frequency of citation of some themes, words or ideas in a text to measure the relative weight attributed to a certain subject.

We remind you that, although this analysis technique can be performed without the aid of technological instruments, the use of software accelerates this process, corroborating the development of the research. In this sense, ACA appears (Grimmer and Stewart, 2013), which facilitates the process of thematic meaning.

In this study, the IRaMuTeQ software in Version 0.7 Alpha 2 (http://www.iramuteq.org/) was used for the ACA, which is free access software and uses R as a statistical language (Ratinaud, 2009).

As Ratinaud (2019) describes, it is a software that enables different types of textual data analysis, from very simple ones, such as basic lexicography, which mainly covers lemmatization and word frequency calculation; to multivariate analyzes such as DHC of text segments, correspondence analysis and similarity analyses. Through this software, vocabulary distribution can be organized in an easily understandable and visually clear way, with graphical representations based on the analyzes used.

Thus, for textual analyses, using IraMuTeQ, we performed multivariate analyzes such as DHC and CFA (Doise, Clemence and Lorenzi-Cioldi, 1992; Lebart and Salem, 1994).

The specific objective was to create a textual corpus that presents a certain characteristic (knowledge of fifth-year elementary school students about graphs and statistical tables), using a method that is independent of the researcher's subjectivity or prior categorizations of the text.

For this, the first part of the qualification is by classification by Reinert's method (1990), known as DHC. The second is the classification of the parts of the text that contain the terms identified as statistically significant, also according to Reinert's method (1990). It is a method that presents an DHC of the occurrences of terms in a specific segment of the text. Thus, co-occurrences of terms in the same segments are identified, distributing texts in classes by proximity, in addition to hierarchizing the relative presence of each term in the created word classes.

Reinert's proposal (1990) allows for an advance in descriptions, moving from the mere presence and quantity of lexicons to an association with the context of the presence of terms. The researcher proposed the method for a lexical analysis of context by text segment set that allows differentiating contexts according to the positions of terms in the textual segments.

Thus, through the groups (clusters) generated, it was possible to categorize the text by DHC, as well as by CFA. In our research, we used the Simple DHC over Text (ST), as this is

recommended when analyzing short answers from open questionnaires. The significance of clustering was identified through the frequency of citations (words) using the chi-square test that reveals the associative strength between words and their respective class.

This associative strength is analyzed when the test is greater than 3.84, representing $p \leq 0.05$, in which the smallest value of the chi-square represents a smaller relationship between the variables (Lahlou, 2012).

To perform the textual analyses, we identified the coding that each of the participants' responses received. The following statistical variables were considered:

1. Participant: n_01 (student 1) and so on until n_74 (student 74).
2. Age: Id_1 (student aged 9); Id_2 (student aged 10); Id_3 (student aged 11).
3. Gender: Gen_1 (male student); Gen_2 (female student).
4. Do you like school: GE_1 (Yes); GE_2 (No).
5. Likes Mathematics: GM_1 (Yes); GM_2 (No).

Thus, the IRaMuTeQ software was used to carry out a quantitative lexical analysis that considers the word as a unit, also offering its contextualization in the corpus or in the research instrument or questionnaire. Each question in the instrument was composed of semantic contents that formed the database or corpus analyzed by the software.

Therefore, a DHC and a CFA were performed that allowed the analysis of the lexical roots, offering contexts in which the classes (clusters) are inserted, according to the text segment of the research corpus (Camargo and Justo, 2013).

The lexical ambiguities in the knowledge of fifth year elementary school students about statistical tables
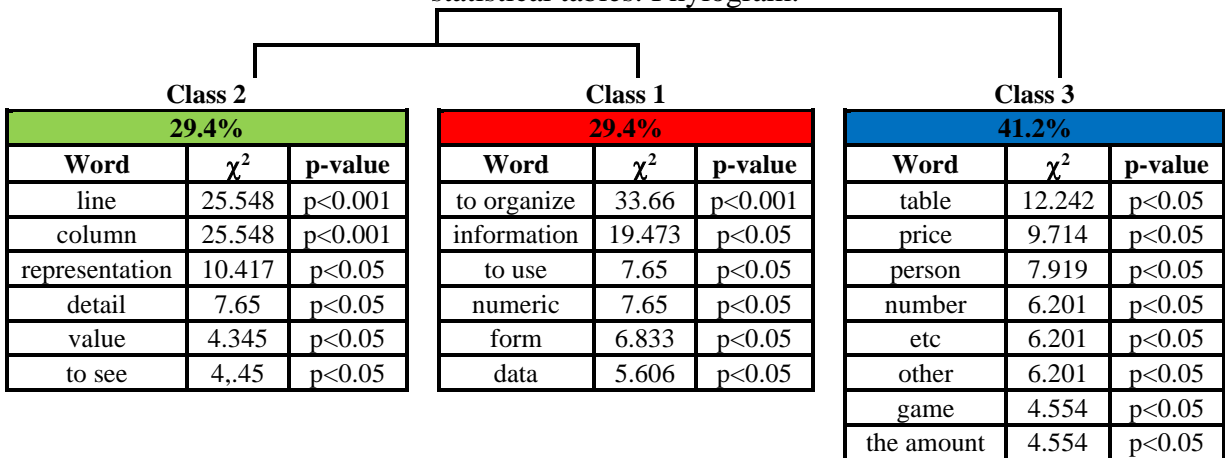
We consider that knowing how to identify tables appropriately for presenting statistical data is an essential part, whether in a press release, an analysis article or a research paper. The use of tables helps to minimize the amount of data in the text and also avoids having to discuss insignificant variables that are not essential for a given topic. Thus, in this topic, we present a multivariate textual analysis to indicate the understanding of 59 students (79.7%) out of a total of 74, regarding statistical tables.

It is important to point out that DHC-type analyses, in order to be useful for classifying any textual material, require a minimum retention of 75% of the text segments, and an analysis lower than this value is not considered an adequate analysis, as it only offers a partial classification (Camargo and Justo, 2013). In this sense, the textual corpus used for the analysis of knowledge about statistical tables is considered representative and useful, as the use was 86.44%.

In Figure 1, we present the phylogram, which separates the main words into classes, allowing to launch hypotheses about possible recurrent themes in the texts generated by the DHC, presenting the partitions that were carried out in the corpus until the final three classes were reached.

A phylogram (tree-shaped diagram) classifies which forms and their respective textual segments are similar, what are their percentage frequencies in the text and how they converge in association or become independent of the others (Salone, 2013).

Figure 1 - Result of Classification by Reinert's Method for the knowledge presented on statistical tables: Phylogram.

| Class 2			Class 1			Class 3		
29.4%			29.4%			41.2%		
Word	χ^2	**p-value**	**Word**	χ^2	**p-value**	**Word**	χ^2	**p-value**
line	25.548	p<0.001	to organize	33.66	p<0.001	table	12.242	p<0.05
column	25.548	p<0.001	information	19.473	p<0.05	price	9.714	p<0.05
representation	10.417	p<0.05	to use	7.65	p<0.05	person	7.919	p<0.05
detail	7.65	p<0.05	numeric	7.65	p<0.05	number	6.201	p<0.05
value	4.345	p<0.05	form	6.833	p<0.05	etc	6.201	p<0.05
to see	4,.45	p<0.05	data	5.606	p<0.05	other	6.201	p<0.05
						game	4.554	p<0.05
						the amount	4.554	p<0.05

Source: IRaMuTeQ software output.

Thus, in the result of Classification by Reinert's Method: Phylogram, Figure 1, at first, the corpus "Body" was divided into two subcorpus, separating class 3 from the rest of the material, which represents 41.2% of the textual corpus (1st partition or iteration). In a second moment, the larger subcorpus was divided, originating classes 1 and 2 that contain, respectively, 29.4% and 29.4% of the textual corpus (2nd partition or iteration).

Continuing with the analyses, we consider the results obtained with the Reinert method (DHC) by the CFA, which bring relationships with the context variables associated with the

classes resulting from the DHC (Figure 2). Remember that the variables used were: Gender; Age; I like school and mathematics.

Figure 2 - Result of Classification by Reinert's Method for the knowledge presented on statistical tables: CFA, associated with the highlighted variables.
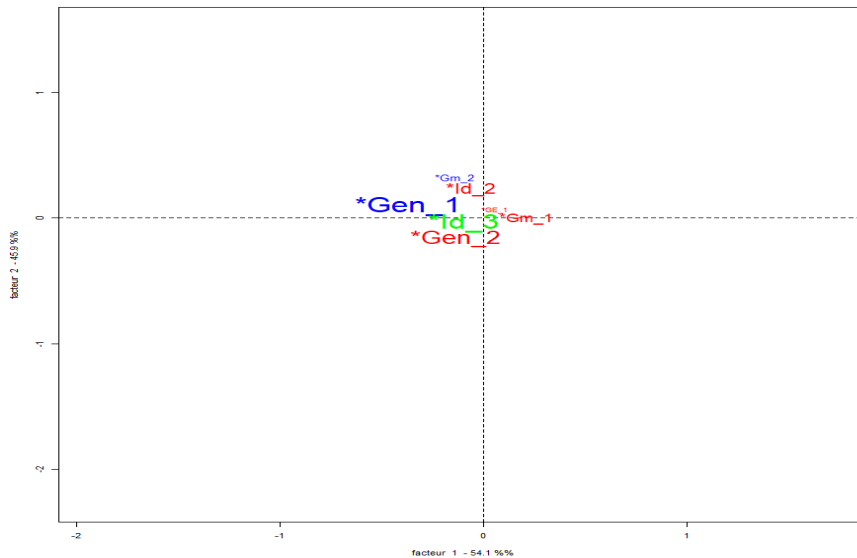
Source: IRaMuTeQ software output.

Taking Figures 1 and 2, we name the three classes, describing what they indicate and, in the sequence, we present some excerpts from the students' speeches to facilitate the understanding of how students perceive the statistical tables. In addition, we established some discussions about possible lexical ambiguities deduced from the knowledge expressed by the 57 students about statistical tables.

Thus, in class 3, which we call "Presenting examples that justify the need to build a table", it is still identified by male students who like mathematics. Taking the combinations of words that present a statistically significant relationship (p ≤0.05), we highlight the following grafts:

> **** *n_37 *Id_3 *Gen_1 *GE_1 *Gm_1
> Price table, chalkboard table. We have several different table types.
> **** *n_38 *Id_2 *Gen_1 *GE_1 *Gm_1
> A football match is there on the table names of people, goals. Gustavo scored 3 goals and Danilo 4.
> **** *n_54 *Id_3 *Gen_1 *GE_1 *Gm_1
> Used in games that football, price list.
> **** *n_56 *Id_3 *Gen_1 *GE_1 *Gm_1
> Standard templates on market lists, price lists and other different ones.

In this first class, we observed from the answers given that students associate statistical tables with everyday situations in which they are necessary for data presentation.

Statistical tables are part of a universal language, a way of presenting data to describe information, with the aim of producing in the researcher, the public or the student a faster and more vivid impression of the subject under study, which nowadays can often be seen occupying a prominent place in the written and spoken media.

Reflecting on representation in tables, Pagan and Magina (2010) argue that citizens need to know how to build them, identifying the best form of representation for the data they are working with.

Thus, to work with tables, you must deal with information from the real world, transform it into numerical data, organize it into tables, to discuss and understand it. We consider that this understanding offers the student subsidies to take a critical position, infer and even make predictions about the fact or phenomenon under study.

In class 1, which we called "To present reasons for the construction of a statistical table", it is identified by female students, aged 10 and who like mathematics. Observing the combinations of words that present a statistically significant relationship ($p \leq 0.05$), we highlight the following grafts:

> **** *n_14 *Id_2 *Gen_2 *GE_1 *Gm_1
> Set of information in order.
> **** *n_22 *Id_2 *Gen_2 *GE_1 *Gm_1
> Way to organize data.
> **** *n_24 *Id_2 *Gen_2 *GE_1 *Gm_1
> Structure we use to organize data, which is information about a certain subject. This data can be qualitative (related to characteristics and attributions) or quantitative (referring to numerical values).
> **** *n_39 *Id_2 *Gen_2 *GE_1 *Gm_1
> Structure we use to organize data, which is information about a certain subject.
> **** *n_52 *Id_2 *Gen_2 *GE_1 *Gm_1
> To organize.
> **** *n_60 *Id_2 *Gen_2 *GE_1 *Gm_1
> They serve to organize textual and numerical information in a clear and convenient way.

In this second group, we noticed that students indicate different reasons that justify the preparation of statistical tables, presenting their perceptions that these representations are ways of organizing data or information.

The tables provide quick and secure information about the variables under study, allowing for more coherent and scientific administrative and pedagogical determinations (Crespo, 2002).

Finally, class 2, which we call "Presenting the formal structure in the construction of a table", is identified by students aged 11 and who like to study. We also present combinations of words that show a statistically significant relationship (p ≤0.05), and we highlight some grafts below:

> **** *n_06 *Id_3 *Gen_2 *GE_1 *Gm_1
> Rows and columns.
> **** *n_19 *Id_3 *Gen_2 *GE_1 *Gm_1
> Frame with lines.
> **** *n_20 *Id_3 *Gen_2 *GE_1 *Gm_1
> It presents data with rows and columns, and is used to see details and compare values.
> **** *n_49 *Id_3 *Gen_2 *GE_1 *Gm_1
> Rows and columns.
> **** *n_61 *Id_3 *Gen_2 *GE_1 *Gm_1
> Rows and columns to organize information clearly and conveniently.
> **** *n_64 *Id_3 *Gen_2 *GE_1 *Gm_1
> Representation formed by rows and columns, used to see details and values.
> ***** *n_71 *Id_3 *Gen_1 *GE_1 *Gm_1
> View in rows and columns.

We consider that this third group assumes that the table is a structure to organize data and is formed by rows and columns, and the junction of rows and columns is called a cell. It is in the cell that the information is inserted, that is, the data. To indicate the number of lines and columns that a table has, you can perform the multiplication between the number of lines and the number of columns. This group of students perceives statistical tables only for their formal aspects (their elaboration structure).

We highlight the indication of the important elements according to Crespo (2005) for the construction of a table, that is, those that ratify the assumptions of Araujo and Flores (2010). In Crespo (2002), the elements that must contain a table for its construction are indicated:

1) Body: set of rows and columns containing information about the variable under study;
2) Header: upper part of the table that specifies the content of each column;
3) Indicator column: part of the table that specifies the content of the columns;

4) Lines: imaginary straight lines that facilitate the reading, in the horizontal direction of data that are inscribed in their intersections with the columns;

5) House or cell: space for a single number;

6) Title: set of information, as complete as possible, and able to answer the questions: What? When? Where? It must be located at the top of the table and it is extremely important, because if we do not place it, readers will not know what the table is talking about.

Thus, although it is important to understand how the tabular representation is developed, as it allows the identification of data more quickly, these should expand the ability to handle statistical information from day-to-day situations, in addition to allowing the establishment of relationships between different types of information.

Therefore, from the early years of elementary school, the student should have contact with different types of tables, from the simplest to the most elaborate (double-entry tables), as their construction is important for the student to develop the ability to observe and representation.

The lexical ambiguities in the knowledge of fifth-year elementary school students about statistical graphs
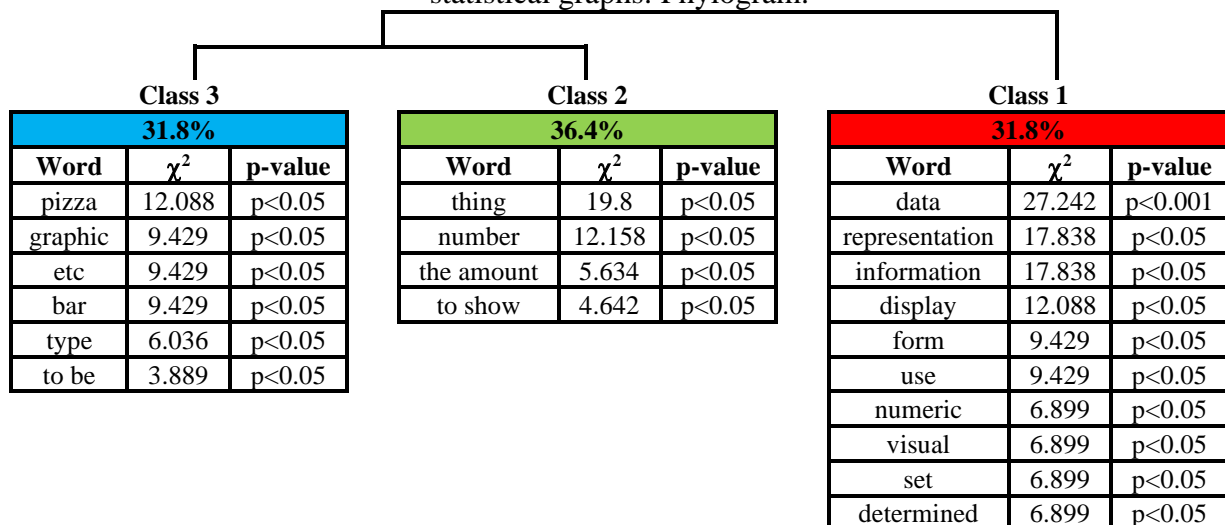
Statistics can often be better understood if they are presented in a graph rather than a table. We consider graphs as a visual representation of statistical data, in which these are represented by symbols, for example, bars or lines. It is a very powerful visual tool as it displays data quickly and easily, allowing comparison and being able to reveal trends and relationships between data.

In general, a graph takes the form of a one-dimensional or two-dimensional figure, such as a bar graph or a line graph. Three-dimensional graphics are also available, although they are generally considered complex.

In this topic, we present a multivariate textual analysis to identify what 67 students (90.5%) out of a total of 74 who participated in the research indicated about their understanding of statistical graphics.

Considering Camargo and Justo (2013), the textual corpus used for the analysis of knowledge of statistical graphics is considered representative and useful, as the use was 75.67%. Again, we remind you that in the DHC tab of the IRaMuTeQ results, it is possible to access the phylogram (Figure 3), which presents the partitions that were made in the corpus until the final three classes were reached.

Figure 3 - Result of Classification by Reinert's Method for the knowledge presented on statistical graphs: Phylogram.

| Class 3				Class 2				Class 1		
31.8%				**36.4%**				**31.8%**		
Word	**χ^2**	**p-value**		**Word**	**χ^2**	**p-value**		**Word**	**χ^2**	**p-value**
pizza	12.088	p<0.05		thing	19.8	p<0.05		data	27.242	p<0.001
graphic	9.429	p<0.05		number	12.158	p<0.05		representation	17.838	p<0.05
etc	9.429	p<0.05		the amount	5.634	p<0.05		information	17.838	p<0.05
bar	9.429	p<0.05		to show	4.642	p<0.05		display	12.088	p<0.05
type	6.036	p<0.05						form	9.429	p<0.05
to be	3.889	p<0.05						use	9.429	p<0.05
								numeric	6.899	p<0.05
								visual	6.899	p<0.05
								set	6.899	p<0.05
								determined	6.899	p<0.05

Source: IRaMuTeQ software output.

Thus, in the result of Classification by Reinert's Method: Phylogram, Figure 3, at first, the corpus "body" was divided into two subcorpus, separating class 1 from the rest of the material, which represents 31.8% of the textual corpus (1st partition or iteration) In a second moment, the larger subcorpus was divided, originating classes 2 and 3 that contain, respectively, 36.4% and 31.8% of the textual corpus (2nd partition or iteration).

Continuing with the analyses, we consider the results obtained with the Reinert method (DHC) by the CFA, which bring relationships with the context variables (Gender; Age; Enjoying School and Mathematics) associated with the classes resulting from the DHC (Figure 4).

Figure 4 - Result of Classification by Reinert's Method for the knowledge presented on statistical graphs: CFA, associated with the highlighted variables.
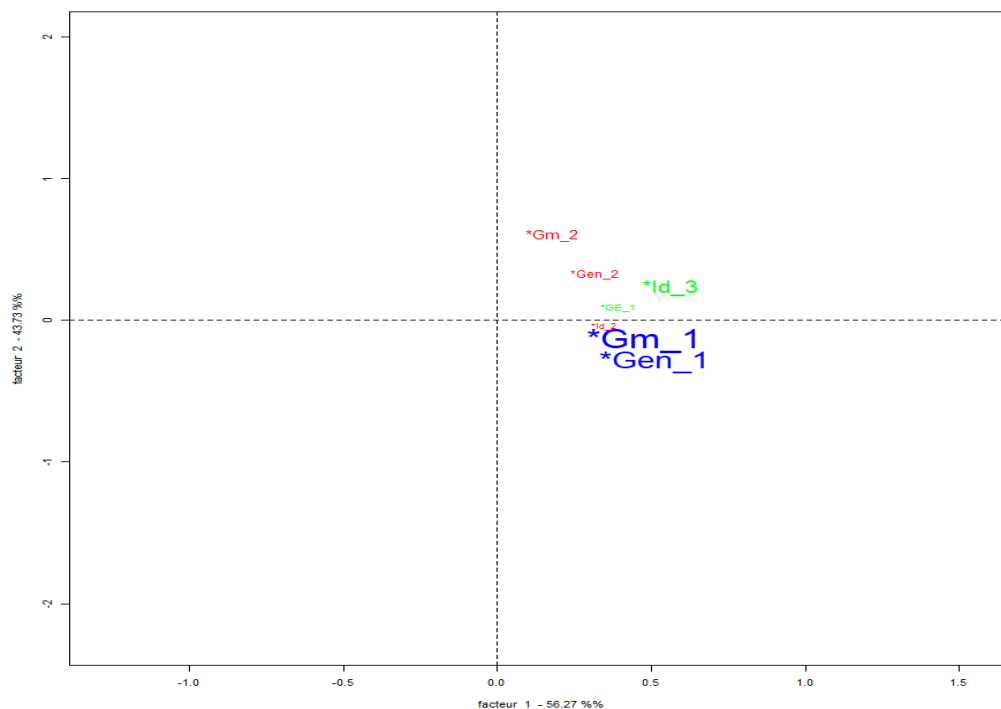
Source: IRaMuTeQ software output.

Taking Figures 3 and 4, we name the three classes, describing what they indicate and, in the sequence, we present some excerpts from the students' speeches to facilitate the understanding of how students perceive the graphs.

In class 1, which we called "Presenting the formal structure in the construction of a graph", it is identified by female students, aged 10 and who do not like mathematics. Taking the combinations of words that present a statistically significant relationship ($p \leq 0.05$), we highlight the following grafts:

> **** *n_07 *Id_2 *Gen_2 *GE_1 *Gm_2
> Representation of information obtained in surveys through geometric shapes to facilitate data reading.
> **** *n_20 *Id_2 *Gen_2 *GE_1 *Gm_2
> Tools used to display data.
> **** *n_38 *Id_2 *Gen_2 *GE_1 *Gm_2
> Geometric representation of a dataset used to facilitate understanding of the information presented in that set.
> **** *n_55 *Id_2 *Gen_2 *GE_1 *Gm_2

Ailton Paulo de Oliveira Júnior; Luzia Roseli da Silva Santos & Fátima Aparecida Kian

Visual representations used to display data about certain information or numerical values.
**** *n_60 *Id_3 *Gen_2 *GE_1 *Gm_2
Geometric representation of a data set to facilitate understanding of the information presented in this set.
**** *n_65 *Id_2 *Gen_2 *GE_1 *Gm_2
Visual representations used to display data.

In this first class, we observed from the answers presented above, that students have the perception of the importance of building statistical graphs, indicating some elements that are essential for their presentation.

It is also noticed that, as the instrument was applied through Google Forms, some students sought and copied definitions that converged to the formal structure in the construction of graphs. In any case, we consider this perception important that, for the construction of a graph, it must have standards that facilitate its presentation and subsequent reading and interpretation of the indicated data.

In Brazil (2018) learning objects are indicated that denote the need for the student to be able to collect, classify and represent the data collected in different types of statistical graphs. Thus, it is important that students have the perception and knowledge that to represent data, for example, a formal structure is necessary so that the graph is presented in a clear and objective way.

In the same line of reasoning, taking the North American document GAISE I (Franklin et al., 2007), it is indicated that students should learn how to collect data, organize their own data or that of third parties and present the data in useful graphs and tables to answer your questions.

In class 2, in green on the phylogram, which we call "Presenting numerical representations in a graph and their applications" is identified by students aged 11 and who like to study. Observing the combinations of words that present a statistically significant relationship ($p \leq 0.05$), we highlight the following grafts:

**** *n_33 *Id_3 *Gen_2 *GE_1 *Gm_1
They serve to show something, for example: the result of a survey, how are sales at a clothing store, how is your performance at school, etc.
**** *n_45 *Id_3 *Gen_2 *GE_1 *Gm_2
Risks that represent one thing, whether they are numbers, calculations or something else.
**** *n_46 *Id_3 *Gen_2 *GE_1 *Gm_2

Lines and numbers that show something.
**** *n_51 *Id_3 *Gen_2 *GE_1 *Gm_2
Number of people who got the answer right.
**** *n_52 *Id_2 *Gen_2 *GE_1 *Gm_1
Things to see something or measure the amount.
**** *n_68 *Id_3 *Gen_2 *GE_1 *Gm_1
Way of showing numbers and types of explanatory drawings of a lesson.
**** *n_72 *Id_3 *Gen_1 *GE_1 *Gm_1
Used to represent quantities of: votes, people, among others.

In this second group, we noticed that students indicate that graphs are used to more easily demonstrate a situation involving numbers, in addition, there is the perception that these numerical representations are associated with everyday situations.

In addition, students indicate variables, both quantitative and qualitative, which seems to show that they perceive, through variables of different types, the possibility of associating a frequency, a number, to the variables.

Statistical graphs, according to Gal (2002), are especially important to achieve a basic statistical culture. These are the knowledge bases that students in the training process should have available, in order to be able to understand, analyze and criticize the statistics that surround us.

For Batanero (2013) it is a component of the statistical sense, that is, working activities involving statistical concepts that are linked to a problem and, for Cazorla (2002), a powerful instrument to communicate and synthesize information from statistical studies.

Finally, class 3 (in blue on the phylogram) which we call "Indicate different types of graphs" is identified by male students who like mathematics. We also present combinations of words that show a statistically significant relationship ($p \leq 0.05$), and we highlight some grafts below:

**** *n_12 *Id_2 *Gen_1 *GE_1 *Gm_1
They can be of many types, they give the data in bars, lines and so on.
**** *n_28 *Id_2 *Gen_1 *GE_1 *Gm_1
Pizza is for you to see how many vegetables, greens and fruits you ate and so on.
**** *n_30 *Id_2 *Gen_1 *GE_1 *Gm_1
They are bars and numbers and they have their value.
**** *n_32 *Id_3 *Gen_1 *GE_1 *Gm_1
It looks like a pizza.
**** *n_43 *Id_2 *Gen_1 *GE_1 *Gm_1
All kinds.
**** *n_47 *Id_3 *Gen_1 *GE_2 *Gm_1

> I think it looks like a pizza.
> **** *n_56 *Id_3 *Gen_1 *GE_1 *Gm_1
> Featured at school and when on TV. Ex. Pizza, bar etc.

We believe that this third group is based on the principle that, on a daily basis, there is access to a high amount of statistical information that can be transmitted through statistical graphics (bars, lines, sectors, pictograms, among others). Graphs are resources used to represent a certain real phenomenon that can be measured, quantified or illustrated in a clear and logical way. They indicate a statistical dimension of a given fact.

For Monteiro and Selva (2001), graphics are a cultural tool that allows expanding the ability to treat statistical information and establish relationships between different types of information. Furthermore, according to Cazorla (2010), in order for us to read, interpret, critically evaluate and build different types of graphic representations, it is necessary to master statistical concepts, promoting citizen and scientific education.

Final Considerations

In this study, based on Ben-Zvi and Garfield (2004), students perceive that the knowledge associated with statistical tables and graphs indicates the organization of data and for its construction there is the perception that there are different types of data representations, as well as understand the symbols used to represent them.

We observed that some students scored, in the case of tabular representations, the identification of specific information present in the intersection of rows and columns, that is, its formal structure when it was built.

This group conceives statistical tables as an ordered presentation of a set of data in cells, which are the intersection of rows and columns. The table includes both the numerical values and the conceptual descriptions to which they refer and particular indications about the origin of the data, as well as specific clarifications that are useful for the user interested in knowing aspects of a conceptual, technical or methodological nature.

In the case of graphs, the students did not indicate that they are ways of presenting statistical data through a design that summarizes the main characteristics of the set of measured observations and that can compare the variables and their categories regarding their magnitudes.

Furthermore, the students referred to statistical tables and graphs associating their life experiences and/or those presented by their teachers in the classroom. It is noticed that students,

in general, both for graphs and for statistical tables, reflected on real data, exposing their ideas and contrasting with their beliefs, understanding that different information can be expressed through these representations.

Starting from the conceptions of students in this study and taking Pereira (2009), tables and graphs are used to summarize a set of information. However, the perception that one should rely on the visual effect was not fully explicit, as this way the information becomes clearer in the presentation of the phenomenon under study.

There is a lexical ambiguity, considering Duval (2003), since the study of graphs and tables must be based on the transit between different types of records, thus providing the visualization of the same object in different ways, leading students to not "enclosure of records", that is, it is about the transformation or conversion of one record to another, from numerical writing to graphic or tabular writing. This last fact leads the individual to have a thought restricted to new possibilities.

It can be seen, through the students' responses, that even with the advancement of schooling, they did not explain that the information contained in statistical tables and graphs is presented in the most diverse means of scientific dissemination and in the media in general. In addition, students are required to list all the data to conclude what they indicate and then allow decision making based on the data.

In view of the results presented, we believe it is necessary to understand statistical tables and graphs: 1) Understand the statistical information provided by the media on various topics and be able to critically evaluate it; 2) Carry out investigations that require the interpretation of data, analyzing them; 3) Discuss or communicate your own opinions about the information that the statistics present.

It should also be considered that the process of statistical training, as well as its conditions and cultural and economic possibilities, must be considered in the teaching and learning process in order to face situations present in the media, research reports or internet access.

Statistical graphs and tables are powerful data visualization tools that allow complex information to be represented in an accessible way, allowing to present information to the user or reader in a clear and precise way, facilitating the comparison and understanding of the evolution of the different variables.

Specifically in relation to the BNCC (Brazil, 2018), for the fifth year of Elementary School, the following skills should be addressed: 1) Interpret statistical data presented in texts, tables and graphs (columns or lines), referring to other areas of knowledge or to other contexts, such as health and traffic, and produce texts with the aim of summarizing conclusions; 2) Carry out research involving categorical and numerical variables, organize data collected through tables, column, pictorial and line graphs, with and without the use of digital technologies, and present written text about the purpose of the research and the synthesis of results.

Also, in Brazil (2018), it is suggested that different types of graphs be analyzed, in particular those that are published in the media, as reading and interpretation develops the skills of questioning, raising, checking hypotheses and looking for relationships between data. When exploring the reading of graphs and tables, questions must be proposed that stimulate their interpretation at different levels of understanding so that the student can relate the data.

In addition, they can be adapted to the level of the target audience, making it a very useful way of offering clear information to anyone, that is, from newspaper and magazine readers, students, teachers, scientific communicators, among others.

Thus, according to Gal (2002), through the insertion of statistics in everyday life, it is important to reflect on the role of the school in helping students to understand basic concepts and procedures in statistics from the early years of elementary school, developing contents and skills that produce statistically literate students. You should know why data are needed and how they can be produced and become familiar with basic concepts related to graphical and tabular representation, highlighting the importance of knowing how to read and interpret information.

References

Araujo, G. E. and Flores, C. R. (2010). O Tratamento da informação nas séries iniciais: uma proposta de formação de professores para o ensino dos gráficos e tabelas. *Anais do 9 Encontro Nacional de Educação Matemática*. Belo Horizonte, Minas Gerais, Brasil.

Arteaga, P., Batanero, C., Cañadas, G. and Contreras, M. (2011). Las Tablas y Gráficas como Objetos Culturales. *Números*, 76, 55-67. Available in http://funes.uniandes.edu.co/3571/1/Arteaga2011LasNumeros76.pdf Access in: 02/02/2023.

Batanero, C. (2013). Sentido estadístico: Componentes y desarrollo. *Actas de 1 Jornadas Virtuales en Didáctica de la Estadística, Probabilidad y Combinatoria* (pp. 55-61). Granada: Grupo de Investigación FQM-126. Available in https://www.ugr.es/~batanero/pages/ARTICULOS/Sentidoestad%C3%ADstico.pdf Access in: 02/02/2023.

Batanero, C., Díaz, C., Contreras, M. and Roa, R. (2013); El sentido estadístico y su desarrollo. *Números*, 83(1), 7-18. Available in http://funes.uniandes.edu.co/3651/1/Batanero2013ElNumeros83.pdf Access in: 02/02/2023.

Brazil. (2018). Ministério da Educação. *Base Nacional Comum Curricular (BNCC):* Educação é a Base. Brasília. Available in http://basenacionalcomum.mec.gov.br/images/BNCC_EI_EF_110518_versaofinal_site .pdf Access in: 01/31/2023.

Camargo, B. V. and Justo, A. M. (2013). Tutorial para uso do software de análise textual IRaMuTeQ. Available in http://www.IRaMuTeQ.org/documentation/fichiers/tutoriel-en-portugais Access in: 01/31/2023.

Cazorla, I. M. (2002). *A relação entre as habilidades viso-pictóricas e o domínio de conceitos estatísticos na leitura de gráficos*. Tese de Doutorado em Educação. Universidade Estadual de Campinas, Campinas, Brasil. Available in https://www.psiem.fe.unicamp.br/pf-psiem/cazorla_irenemauricio_d.pdf Access in: 01/31/2023.

Cazorla, I. M. and Santana, E. (2010). *Do Tratamento da Informação ao Letramento Estatístico*. Via Litterarum.

Crespo, A. (2002). *A Estatística Fácil*. Saraiva.

Crespo, M. I. (2005). Um estudo sobre o comportamento de busca e uso de informação de pesquisadores das áreas de biologia molecular e biotecnologia: impactos do Periódico científico eletrônico. Dissertação de Mestrado em Comunicação e Informação. Faculdade de Biblioteconomia e Comunicação, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brasil. Available in https://www.lume.ufrgs.br/bitstream/handle/10183/4387/000500810.pdf?sequence=1& isAllowed=y Access in: 01/31/2023.

Del Pino, G., and Estrella, S. (2012). Educación estadística: relaciones con la matemática. *Pensamiento Educativo, Revista De Investigación Latinoamericana*, 49(1), 53–64. Available in https://pensamientoeducativo.uc.cl/index.php/pel/article/view/25747/20671 Access in: 01/31/2023.

Doise, W., Clemence, A. and Lorenzi-Cioldi, F. (1992). *Représentations sociales et analyses de données*. P.U.G. Available in https://ccfr.bnf.fr/portailccfr/ark:/06871/00110825154 Access in: 01/31/2023.

Dunn, P. K., Carey, M. D., Richardson, A. M. and McDonald, C. (2016). Learning the language of statistics: Challenges and teaching approaches. *Statistics Education Research Journal*, 15(1), 8-27. Available in https://iase-web.org/ojs/SERJ/article/view/255/156 Access in: 01/31/2023.

Duval, R. (2003). Registro de representações semióticas e funcionamento cognitivo da compreensão em matemática. In Machado, S. D. A. (Org.), Aprendizagem em Matemática: registro de representação semiótica (pp. 11- 33). Papirus. Available in https://1library.org/article/registros-de-representa%C3%A7%C3%B5es-semi%C3%B3ticas-duval.zwveg61q Access in: 01/31/2023.

Estrella, S. (2014). El formato tabular: una revisión de literatura. *Revista Actualidades Investigativas en Educación*, 14(2), 1-23. Available in https://redalyc.org/pdf/447/44731371016.pdf Access in: 01/31/2023.

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M. and Scheaffer, R. (2007). *Guidelines for Assessment and Instruction in Statistics* (GAISE) Education: A Pre-K-12 Curriculum Framework. Alexandria, VA: American Statistical Association. Available in https://www.amstat.org/docs/default-source/amstat-documents/gaiseprek-12_full.pdf Access in: 01/31/2023.

Gal, I. (2002). Adult´s Statistical Literacy: meanings, components, responsibilities. *International Statistical Institute*, 70(1), 1-5. Available in https://www.jstor.org/stable/1403713 Access in: 01/31/2023.

Gal, I. (2011). Does Census at school develop statistical literacy? *Statistical Journal of the IAOS*, 27(3-4), 229-230. Available in https://www.researchgate.net/publication/287265069_Does_CensusAtSchool_develop_statistical_literacy Access in: 01/31/2023.

Glazer, N. Challenges with graph interpretation: a review of the literature. *Studies in Science Education*, 47(2), 183-210. Available in https://www.tandfonline.com/doi/abs/10.1080/03057267.2011.605307?journalCode=rsse20 Access in: 01/31/2023.

Jurečková, M. and Csachová, L. (2020). Statistical literacy of Slovak lower secondary school students. *Technium Social Sciences Journal*, 9, 163-173. Available in https://techniumscience.com/index.php/socialsciences/article/download/966/356 Access in: 01/31/2023.

Kanji, G. K. (2006). *100 statistical tests*. SAGE Publications Ltd.

Lahlou, S. (2001). Text mining methods: an answer to Chartier and Meunier. *Papers on social representations*, 20(38), 1-7. Available in https://eprints.lse.ac.uk/46728/1/Text%20mining%20methods%28lsero%29.pdf Access in: 01/31/2023.

Lebart, L. and Salem, A. (1994). *Statistique textuelle*. DUNOP. Available in https://www.persee.fr/doc/hism_0982-1783_1994_num_9_1_1702 Access in: 01/31/2023.

Mcconway, K. (2016). Statistics and the media: a statistician's view. *Journalism*, 17(1), 49-65. Available in https://journals.sagepub.com/doi/10.1177/1464884915593243 Access in: 01/31/2023.

Molina-Portillo, E. et al. (2019). Statistical literacy in the information society. *Boletín de Estadística e Investigación Operativa*, 35(2), 148-169.

Monteiro, C. and Selva, A. C. V. (2001). Investigando a atividade de interpretação de gráficos entre professores do ensino fundamental. *Anais da XXIV Reunião Anual da Associação Nacional de Pós-Graduação e Pesquisa em Educação*, Caxambu, Brazil.

Muñoz, D. E., Esparza, A. C. M., Ciprés, M. C. and Macías, M. G. M. (2020). Comprehension of statistical graphs and tables by primary school teachers-in-training. *Zetetiké*, 28, 1-17. Available in https://periodicos.sbu.unicamp.br/ojs/index.php/zetetike/article/view/8656854 Access in: 01/31/2023.

Nascimento, A. A. and Menandro, P. R. M. (2006). Análise lexical e análise de conteúdo: uma proposta de utilização conjugada. *Estudos e pesquisas em psicologia*, 6(2), 72-88. Available in http://pepsic.bvsalud.org/pdf/epp/v6n2/v6n2a07.pdf Access in: 01/31/2023.

Pagan, A. and Magina, S. (2010). O ensino de estatística a partir da interdisciplinaridade: um estudo comparativo. Anais do 10 Encontro Nacional de Educação Matemática – ENEM. Salvador, Bahia, Brasil. Available in http://educa.fcc.org.br/pdf/rbep/v92n232/v92n232a15.pdf Access in: 01/31/2023.

Pereira, S. (2009). A leitura e interpretação de tabelas e gráficos para alunos do 6º ano do Ensino Fundamental: uma intervenção de ensino. Dissertação de Mestrado Profissional em Ensino de Matemática. Pontifícia Universidade Católica de São Paulo, São Paulo, Brasil. Available in https://sapientia.pucsp.br/handle/handle/11433?mode=full Access in: 01/31/2023.

Richardson, A. M., Dunn, P. K. and Hutchins, R. (2013). Identification and definition of lexically ambiguous words in statistics by tutors and students. *International Journal of Mathematical Education in Science and Technology*, 44(7), 1007–1019. Available in https://eric.ed.gov/?q=technology+and+education&ff1=dtysince_2004&pg=2485&id=EJ1021774 Access in: 01/31/2023.

Salone, J. J. (2013). Analyse textuelle avec IRaMuTeQ et interpretations référentielles des programes officiels de mathématique en quatrième. *Sciences-Croisées*, 13, 1-13. Available in https://hal.science/hal-01671161 Access in: 01/31/2023.

Tukey, J. W. (1977). Exploratory data analysis. Reading.

Weiland, T. (2017). Problematizing statistical literacy: an intersection of critical and statistical literacies. *Educational Studies in Mathematics*, 96(1), 33-47. Available in https://link.springer.com/article/10.1007/s10649-017-9764-5 Access in: 01/31/2023.

Authors
Ailton Paulo de Oliveira Júnior
Bachelor in Statistics from the National School of Statistical Sciences – ENCE, Rio de Janeiro, Brazil
Master in Operational Research by the Military Institute of Engineering - IME, Rio de Janeiro, Brazil
PhD in Education from the University of São Paulo – USP, São Paulo, Brazil
Associate Professor IV of Federal University of ABC - UFABC
Research line: Statistical Education.
ailton.junior@ufabc.edu.br
https://orcid.org/0000-0002-2721-7192

Luzia Roseli da Silva Santos
Graduated in Pedagogy, Faculty of Aldeia de Carapicuíba - FALC, São Paulo, Brazil.
Master in teaching and history of science and mathematics from the Federal University of ABC - UFABC, São Paulo, Brazil
Doctoral student in the Graduate Program in Teaching and History of Science and Mathematics at the Federal University of ABC - UFABC, São Paulo, Brazil
Research line: Statistical Education.
luziaroselidasilvasantos@gmail.com
https://orcid.org/0000-0001-6930-9215

Fátima Aparecida Kian
Degree in Literature-Portuguese, Faculty of Philosophy and Letters of the Santo André
Foundation, São Paulo, Brazil.
Master in teaching and history of science and mathematics from the Federal University of
ABC - UFABC, São Paulo, Brazil
Doctoral student in the Graduate Program in Teaching and History of Science and
Mathematics at the Federal University of ABC - UFABC, São Paulo, Brazil
Research line: Statistical Education.
escritores6@gmail.com
https://orcid.org/0000-0003-0105-7335