Zetetiké

DOI: http://dx.doi.org/10.20396/

Taxonomy of errors and difficulties when constructing and interpreting frequency tables

Taxonomía de errores y dificultades en la construcción e interpretación de tablas de frecuencia

Ingrith Álvarez Alfonso[1]

Yuly Guerrero Gutiérrez[2]

Yessica Torres López[3]

Abstract

This article describes a qualitative research process carried out with secondary school students in Colombia. The research aimed at further exploring a set of errors and difficulties which were identified through the teaching experience and for which no particular theoretical support was found. It established the need to focus on the structural elements of frequency tables and their relation to errors in the corresponding statistical graphs so that a theoretical explanation for the errors could be found. Data was collected from a total of 100 students by using instruments designed purposefully to obtain evidence that allowed formulating a taxonomy of errors and difficulties associated with the construction and interpretation of frequency tables. The taxonomy represents a theoretical contribution that may be relevant in the study of teaching processes related to this subject matter, and the design of teaching activities to overcome or mitigate such errors and difficulties.

Keywords: Errors and Difficulties; Construction & Interpretation of Frequency Tables.

Resumen

Se describe un proceso de investigación cualitativa llevada a cabo con estudiantes de educación básica y media en Colombia. Se tuvo como objetivo corroborar un conjunto de dificultades y errores que se preveían a partir de la práctica pedagógica para los cuales no se hallaba sustento teórico particular. Se estableció la necesidad de centrar la atención en los elementos estructurales de las tablas de frecuencias y su relación con los errores reportados para gráficos estadísticos con el fin de encontrar una explicación teórica de los errores asociados a tablas. Los datos fueron recolectados de un total de 100 estudiantes, mediante el uso de instrumentos diseñados con el propósito de obtener evidencia que permitiera la formulación de una taxonomía de los errores y las dificultades asociadas en la construcción e interpretación de tablas de frecuencia. La taxonomía representa una contribución teórica que puede ser relevante en el estudio de los procesos pedagógicos relacionados con este objeto de estudio, y el diseño gestión de actividades para superar o mitigar tales errores y sus dificultades.

Palabras clave: Errores; Dificultades; Construcción tablas de frecuencia; Interpretación tablas de frecuencia.

Submetido em: 04/09/2019 – **Aceito em**: 12/02/2020 – **Publicado em**: 14/04/2020

[1] Magister en Docencia de la Matemática de la Universidad Pedagógica Nacional. Profesora Investigadora de la Universidad Pedagógica Nacional, Colombia. Email: ialvarez@pedagogica.edu.co

[2] Magister en Docencia de la Matemática de la Universidad Pedagógica Nacional. Profesora de educación básica, Colombia. Email:yaguerrerog@upn.edu.co

[3] Licenciada en Matemáticas de la Universidad Pedagógica Nacional, Colombia. Email:dma_ytorres438@pedagogica.edu.co

DOI: http://dx.doi.org/10.20396/

Introduction

The development of statistical thinking plays a fundamental role in the training of statistically-literate citizens. This kind of thinking intends for everyone to be able to interpret and analyze critically information displayed in different media –either through graphs, frequency tables or statistical data within texts, such as press articles, editorials or blogs. This literacy also regards the importance of sufficient and necessary knowledge so that individuals are not only critical consumers of information but can also take on roles as producers. This means being able to collect and organize statistical data in counting tables such as frequency and contingency tables, and as stated by Tauber (2010), it also means to produce statistical graphs associated with such data sets. Thus, the concepts of statistical variable, variable value, frequency and its different types, and the construction and interpretation of statistical tables (frequency tables) have been positioned within the Colombian school curriculum.

As described by the Ministry of National Education [MEN] (1998) with regards to 'Random Thinking and Data Systems', treating daily life situations or other sciences must occur through exploration and research by collecting, organizing, representing and analyzing data through tables whose interpretation guides decision making (p.47). The Ministry, therefore, proposes that students be able to i) classify and organize data according to its qualities, and represent it using tables and graphs (pictograms, bar charts, line graphs, pie charts); ii) interpret information displayed in statistical tables and graphs; and iii) solve and formulate problems from a data set presented in tables, bar charts, pie charts and others (MEN, 2006). This panorama about training individuals implies that secondary school statistics professors[4] must address topics such as frequency tables for grouped and ungrouped data (a conceptual basis for the construction of statistical graphs) and promote the use of strategies that contribute to the development of random thinking, particularly to the construction and interpretation skills of this type of tables.

Notwithstanding these national curriculum guidelines, during teaching experiences professors face scenarios slightly positive for learning such topics. These scenarios sometimes cannot be explained by existing theories, even though some authors highlight the importance of recognizing and analyzing errors, difficulties and obstacles in teaching-learning processes. For instance, Batanero (2001) characterizes errors, difficulties and obstacles as vital inputs for professors, since they serve as a basis for (re)guiding teaching processes. Furthermore, Socas (1997) particularly states that the characterization of these elements is useful to design activities that might contribute to overcome or avoid their emergence.

Given the importance of having references related to errors and difficulties in the construction and interpretation of frequency tables, qualitative-descriptive research was

[4] The structure of the Colombian education system is made up as follows: early childhood education (before the age of 4), preschool education (between 4 and 6 years old), elementary education (five grades – between 6 and 10 years old), secondary education (four grades – between 11 and 14 years old), middle high school education (two grades and ends with the bachelor’s degree – between 15 and 16 years old), and higher education.

DOI: http://dx.doi.org/10.20396/

conducted to describe the process carried out with secondary school students in Colombia. Such research aimed at providing and explaining a theoretical framework about difficulties and errors perceived in classrooms and for which no specific theoretical support was found.

Two instruments were designed and implemented to collect information (excerpts shown in Appendices) as part of the research process. Data collected through them provided evidence to contrast hypothesis coming from teaching practices with some theoretical analogies, to describe in detail errors and difficulties related to the construction and interpretation of frequency tables, and to formulate the taxonomy (a theoretical contribution to the research field).

Conceptual framework

The conceptual framework is based on the conceptions known as obstacle, difficulty and error within the conceptual structures developed when learning mathematics. Conceptions that are later associated with the field of random thinking and, specifically, with inherent errors when constructing statistical graphs. This is due to the fact that traditionally, data representations are assumed from graphs, without focusing on the structural and fundamental elements of tables. Accordingly, focusing on the construction of graphs leads to assume the levels of interpretation and the abilities for their comprehension proposed by Friel, Curcio and Bright (2001) –the starting point to set up the mentioned instruments of data collection.

Difficulties, obstacles and errors

Socas (1997) points out that difficulties are connected and strengthened in complex conceptual networks, which are consolidated in practice into obstacles to end up being manifested in form of errors. Keeping such relationship, these ideas (difficulties, obstacles, errors) are studied in the mentioned order. Thus, **difficulties** are understood as a conflict generated during the process of learning mathematics and it is assumed that these difficulties may be of different nature, as it is presented in Table 1.

Table 1 – Types of difficulties

Difficulties associated with the complexity of the mathematical object	These difficulties are intrinsically related to the conceptual structure of the mathematical object (graphic or symbolic representations, definitions, applications, etc.); to the level of abstraction reached; and to the variety of meanings that the terminology may have for students according to their everyday life since assuming a new meaning for a specific expression implies cognitive modification and adaptation.
Difficulties associated with mathematical thinking processes	This category is associated with the necessary ruptures (modification and adaptation) in relation to the types of mathematical thinking. It is also associated with the need to achieve larger processes involving formal logic deduction and master algorithmic procedures to reach a certain level of mathematical competences.
Difficulties associated with the processes of teaching mathematics	Such difficulties are directly associated with decisions concerning schools, curriculum, teaching methods, administrative decisions and methodological approaches assumed in the classroom. These components may contribute to or

DOI: http://dx.doi.org/10.20396/

	impact student learning.
Difficulties associated with students' cognitive development processes	Difficulties framed in this category are directly related to the impact that i. the levels of reasoning reached by students, ii. the general stages of intellectual development, and iii. each student's learning processes according to their ages and levels of training have on pedagogic processes.
Difficulties associated with affective and emotional attitudes towards mathematics	Difficulties framed in this category are related to students' distaste for mathematics, which affects teaching-learning processes and may be due to different factors, namely: professors' negative attitude, anxiety about completing an unpleasant task, fear of failure or negative attitudes and beliefs towards mathematics transmitted by others.

Source: Own elaboration. Inspired by Socas (1997)

These types of difficulties lead to the emergence of **obstacles**. From Bachelard's (1988) perspective, they are conceptual stagnations caused by confusions that occur when it is not possible to develop a task by using the knowledge already possessed. In consequence, an obstacle is identified once new knowledge to be acquired conflicts with previous knowledge that could have been learnt wrongly or partially right. This knowledge was functional on some occasions, but by the time it is wanted to be used for another task, it ends up being not strong enough to face the conceptual demands of the new task and conflicts with the student's cognitive structure.

In this context, psychogenetic, didactic and epistemological obstacles are recognized. Psychogenetic obstacles are related to the individual's development stage that challenges apprenticeship. Didactic obstacles, on the other hand, are linked to the methodology used in the teaching-learning process and are mainly the result of the teaching process carried out by professors. Finally, epistemological obstacles are associated with concept difficulty, which could be identified through its historical development.

According to Socas (1997), an **error** is generated by the persistence of difficulties and obstacles in the teaching process and an inadequate cognitive scheme. This is understood as obstacle external manifestation (written –symbolic, algorithmic, graphic–, verbal, attitudinal, etc.). The error is visible to the eyes of professors but rarely to students', since, for them, their knowledge is still efficient despite the fact that the task is not solved satisfactorily.

Errors and difficulties: Statistical graphs

Rodríguez and Sandoval (2012) pose that data representation systems, particularly frequency tables, play a significant role in both data analysis and their communication in different educational and social contexts. On that basis, it can be said that statistical tables and graphs share the same nature, which makes it necessary to take –as initial research support– some of the errors that students make when they construct graphs. According to Batanero and Godino (2002, p. 727), some of the errors are as follows:

- Incorrect choice of the type of graph regarding the statistical variable under study, the information to be presented and the data set.
- The chosen scales of representation are inadequate in comparison with the amount of data presented.
- The scales of either or both the vertical and horizontal axes are omitted.

DOI: http://dx.doi.org/10.20396/

- The origin of coordinates is not specified either for frequencies or for values taken by the statistical variable.
- Insufficient divisions in axis scales are provided; therefore, information represented lacks clarity.
- Combination of non-comparable data in the same graph.

Statistical graphs: Teaching and learning

In the same way that statistical tables were assumed as forms of graphical representation of data, elements which impact the teaching-learning process of statistical graphs, such as the need for previous knowledge, the skills required for understanding graphs and data comprehension levels are taken into account as part of the reference framework. In that sense, according to Batanero and Godino (2002), some aspects that may affect the understanding of statistical graphs are related to previous knowledge with regard to:

- The topic which the graph is related to. If students are not familiar with the context of the data set or data just does not make sense or is not important to them, the understanding of statistical information that is intended to be reported in the respective graph may be affected due to lack of context.
- Mathematical content within the graph, that is, numerical concepts (percentages, decimals, scales, etc.), order relation, operations, proportions, areas, etc.
- The type of statistical graph used (bar charts, histograms, pictograms, box plots, cartograms, etc.) depending on the type of statistical variables presented or the data to be related in it.

According to Friel, Curcio, and Bright (2001), skills such as translation, interpretation, extrapolation, and interpolation are required to reach the understanding of statistical graphs (Table 2).

Table 2 - Skills for understanding statistical graphs

Translation	Ability to transit among different types of representation of the same data set. That means, being able to go from raw data tables (without any order) to contingency, frequency or counting tables; from statistical tables to graphs; and being able to use different statistical graphs to represent an equivalent data set.
Interpretation	Expertise to establish relationships among data, considering the most representative ones to understand the information represented.
Extrapolation/ Interpolation	Ability to identify trends, regularities or atypical situations presented implicitly within a data set.

Source: Own elaboration. Inspired by Friel, Curcio and Bright (2001)

Depending on previous knowledge, skills and conceptual construction developed around the subject matter, individuals may be positioned at a certain level of interpretation of statistical graphs. According to Friel, Curcio and Bright (2001), a wide description of each of the four levels is presented in Table 3.

Table 3 - Levels of interpretation of statistical graphs

Level 1.	Literal (basic or superficial) reading of the graph is conducted. There is no interpretation at

DOI: http://dx.doi.org/10.20396/

Reading the data	this level. The reader simply identifies factual information found in the graph (quantities, labels, frequencies, variables, values or categories for the variable, title, etc.).
Level 2. Reading within the data	Integration and interpretation of data presented in the graph are conducted. More detailed reading is conducted to establish relationships among the data based on previous knowledge regarding mathematical concepts and skills. Order relations are established between frequencies, the type of variable is recognized and any data basic behavior is identified from central tendency (the mode).
Level 3. Reading beyond the data	Predictions and inferences are conducted regarding information that is not directly displayed in the data presented in the graph: hypotheses are formulated about the future or previous behavior of the data set. It is possible to establish relationships among two or more statistical variables implicit in a graph.
Level 4. Reading behind the data	The reader assesses completeness and reliability of data, criticizes its origin and the intention of its representation, and establishes questions that require information on other statistical variables associated with the study to make sense of it.

Source: Own elaboration. Inspired by Friel, Curcio and Bright (2001)

Analysis of difficulties, obstacles and errors; abilities for comprehension of statistical graphs; reading and interpretation levels; and transfer of these perspectives towards the theoretical construct of frequency tables guide the design of instruments to collect information and to be able to characterize errors and difficulties of the object associated with this research.

Methodological research strategy

Qualitative-descriptive research is carried out in four phases. At the first phase, a set of errors and difficulties is formulated for each study process (construction and interpretation). Such compilation is based on experiences obtained through the teaching practice, in particular, the teaching of frequency tables as well as on theoretical references assumed as the framework for this research. At the second phase, two questionnaires are prepared in order to collect evidence so that they become formal inputs. The first questionnaire was designed to identify errors and difficulties presented in the construction of frequency tables, which included a total of 12 situations (some of them are described in Appendix A). The second questionnaire included 8 situations in order to identify errors and difficulties related to the interpretation of frequency tables (some of these situations are described in Appendix B). Both cases (construction and interpretation) involve grouped and ungrouped data sets.

At the third phase, management is carried out in four schools located in Cundinamarca, Colombia. The chosen institutions are those in which the research proposal was presented and accepted by professors and directors[5]. The chosen grades are those which had already studied the topic under study, and whose students would, therefore, be able to sort out the questionnaires. As per these conditions, 7th and 9th grades were not included, since professors did not grant access or the participants did not have previous knowledge on the subject. Hence, the interpretation questionnaire is applied in two different schools with a

[5] The proposal was presented to 10 schools, but only four of them approved the research implementation.

DOI: http://dx.doi.org/10.20396/

total of 65 students between 11 and 16 years of age, (6th and 8th grades, basic secondary School in Colombia). The questionnaire aimed at identifying errors in construction was applied to 100 students from three different institutions. They attended tenth or eleventh grade (middle school in Colombia) and were 15 to 18 years of age. The questionnaires were implemented in diverse schools (state and private schools, with technical, teaching or academic high school diploma), school grades (6th, 8th, 10th and 11th), and age ranges (between 11 and 18 years of age). A large variety of answers from students immersed in different contexts were collected (Table 4), and consequently associated with different teaching-learning processes.

Table 4 – Participants' characterization

STUDENTS		GRADE	TYPE OF INSTITUTION	INSTRUMENT COMPLETED
QUANTITY	AGE			
1 19 17 2	15 Years 16 Years 17 Years 18 Years	11°	State school: Academic High School Diploma	Questionnaire 1: Construction of frequency tables
17 11 3	15 Years 16 Years 17 Years	10°	State school: Academic High School Diploma – Minor: Teaching	
2 23 5	15 Years 16 Years 17 Years	11°	Private school: Academic High School Diploma	
19 7 3 2	11 Years 12 Years 13 Years 14 Years	6°	State school: Academic High School Diploma	Questionnaire 2: Interpretation of frequency tables
1 15 11 5 2	12 Years 13 Years 14 Years 15 Years 16 Years	8°	State school: Commercial and Technical High School Diploma	

Source: Own elaboration

Finally, at the fourth phase, each answer obtained in the questionnaires is reviewed and analyzed by converting the information into data. According to Batanero and Godino (2002), data is initially organized by transferring errors in construction of graphs to construction of tables. As common answers (errors) are found and the teaching experience is addressed regarding this object, data classification is refined and analytical categories are generated. Categories allow not only categorizing data in favor of those fundamental elements of frequency tables but also characterizing population against errors and difficulties in construction and interpretation of frequency tables. Finally, taking into account the processes and concepts involved in frequency tables, a deep and differentiated definition of such categories leads to the consolidation of a particular taxonomy for such object: the main conclusion and theoretical contribution of this research to the Statistics Education context.

DOI: http://dx.doi.org/10.20396/

Analysis and Results

As far as the construction of tables is concerned, participants built a frequency table with ungrouped data despite the large amount of data they had to manipulate. Other participants determined inadequate width of intervals and therefore obtained too many or too few (they determine not only intervals of very large width, but also with different width), which reveals that they did not take into account data dispersion. They also made wrong calculations of frequencies and among frequencies (subtracted simple frequency of two consecutive values of the variable to complete the cumulative frequency column); omitted statistical variable values (excluded or repeated data); exchanged frequencies (e.g. simple and cumulative); included information displayed in the table that should not be classified as a variable; and even repeated data in frequency distribution (Figure 1).

Lack of previous knowledge, partial or superficial reading of context, and lack of representativeness of interval width may be presented as difficulties associated with the processes of teaching mathematics. Dificulties related to previous knowledge about the use of intervals in frquency distribution were observed as well. However, it is possible to establish that in no case these students omitted the value of the statistical variable with zero frequency which demonstrates that this is not a common error when constructing frequency tables, as it was assumed from teaching experience.

Interpretation activities allow identifying that pupils use either frequency interchangeably to conclude or to give answers. They get confused between relative simple frequency and cumulative frequency. Others claim not to remember or not to understand the meaning of determinate frequency, which leads them to confuse a frequency with the value of the corresponding statistical variable. Students probably choose results based on their subjectivity and not on data. They also omit width difference shown by intervals used in the table. In addition, they confuse variable greater value and frequency greater value, and use one of the interval ends to represent it. Figure 2 shows the aforementioned errors, among others.

DOI: http://dx.doi.org/10.20396/

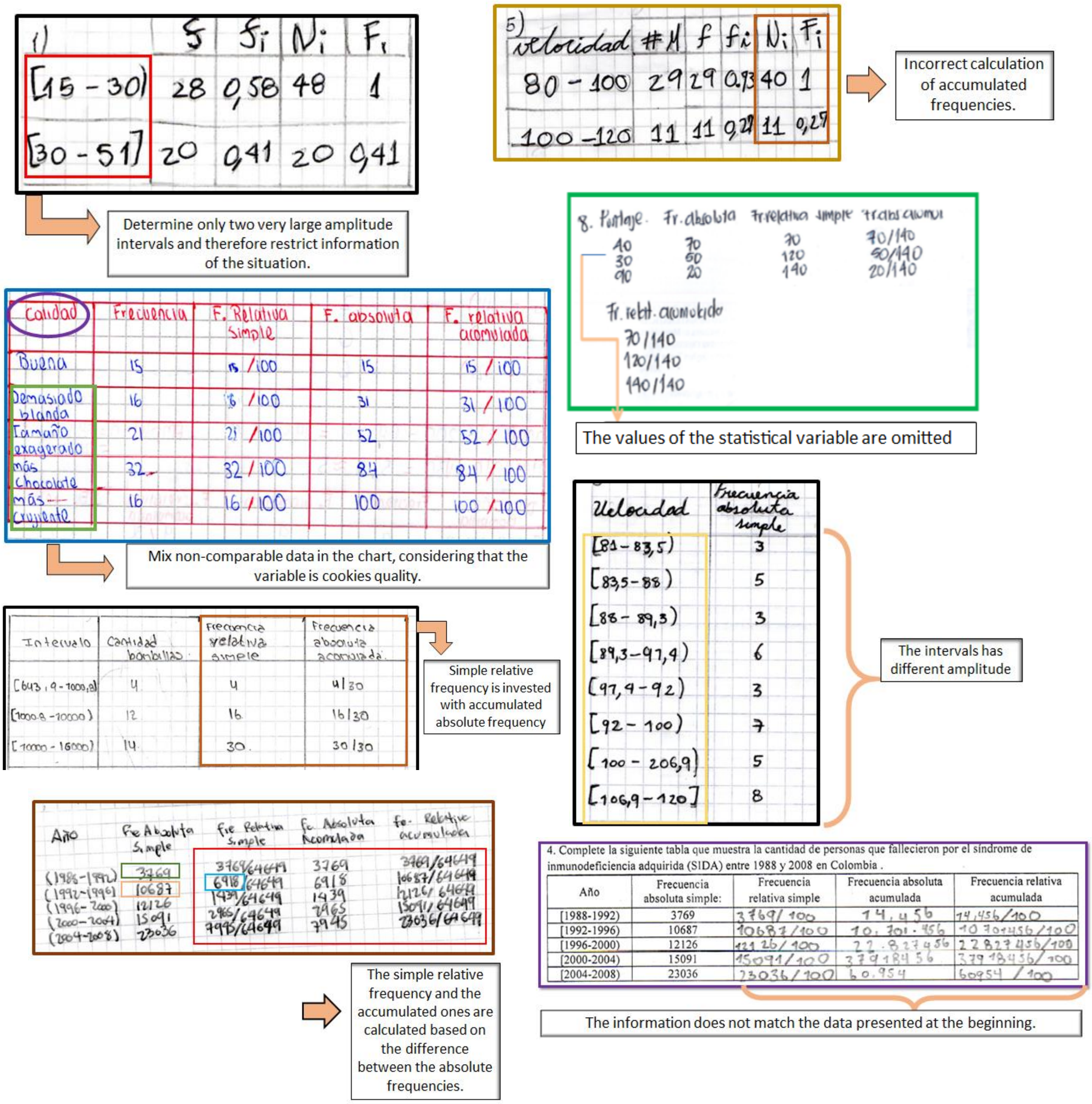


Figure 1 – Examples of errors when constructing frequency tables
Source: Own elaboration

Most participants limit data analysis / interpretation reported in frequency tables likely due to superficial or partial visualization of the information displayed in them. Students answer as long as questions inquire about a particular datum explicitly stated (what they may visually identify) in the table. This reveals little ability to view data sets as a whole and, therefore, lack of ability to make predictions and inferences and to analyze such sets as part of a group.

zetetiké

DOI: http://dx.doi.org/10.20396/

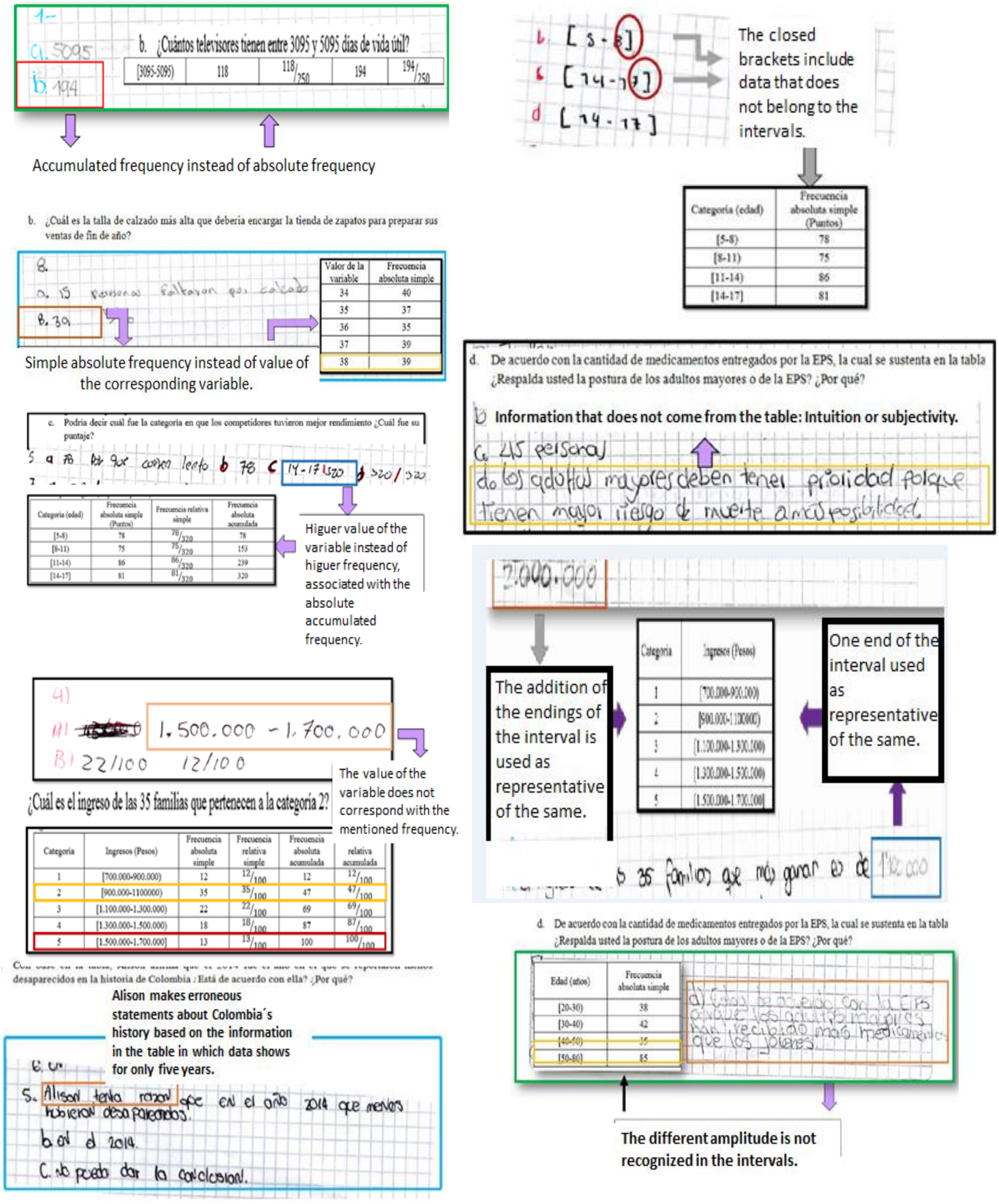


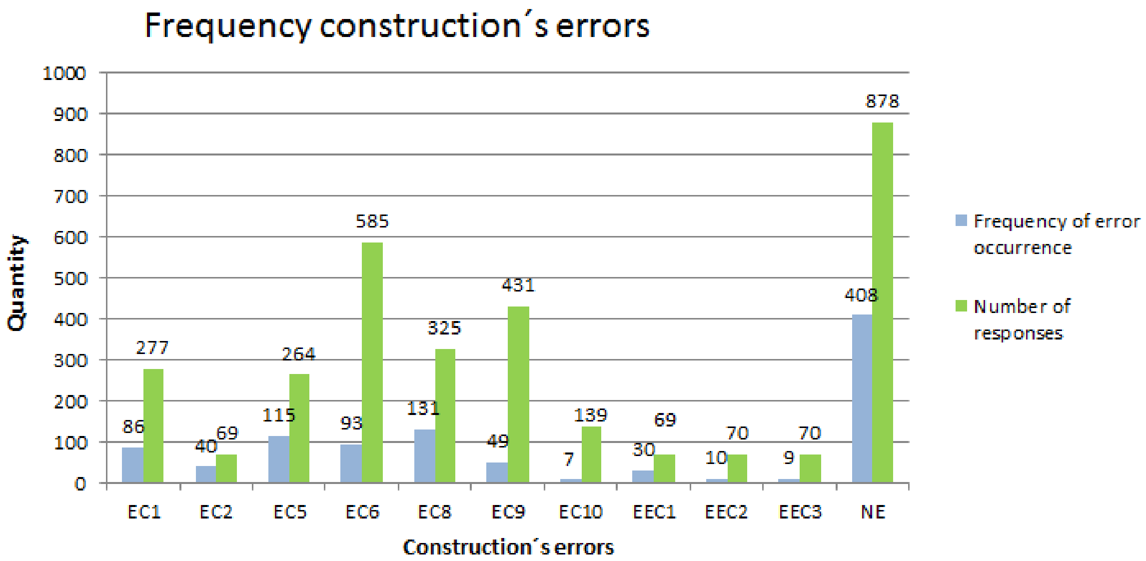
Figure 2 – Examples of errors when interpreting frequency tables
Source: Owen elaboration

Conversely, it is evident that participants do not recognize nor do they assess information provided by the absolute cumulative frequency. For example, when students are asked about an absolute cumulative frequency explicitly stated in the table, they repeat the algorithmic process (performing the addition of the corresponding absolute simple frequencies) and ignore representativeness of the information already provided by the table.

DOI: http://dx.doi.org/10.20396/

These errors are linked to the lack of knowledge about frequency types (absolute, relative, cumulative) and their corresponding meaning, and are caused by difficulties associated with teaching mathematics processes. In particular, both lack of previous knowledge and anxiety about completing a task are observed, in other words, difficulties are linked to emotions towards mathematics.

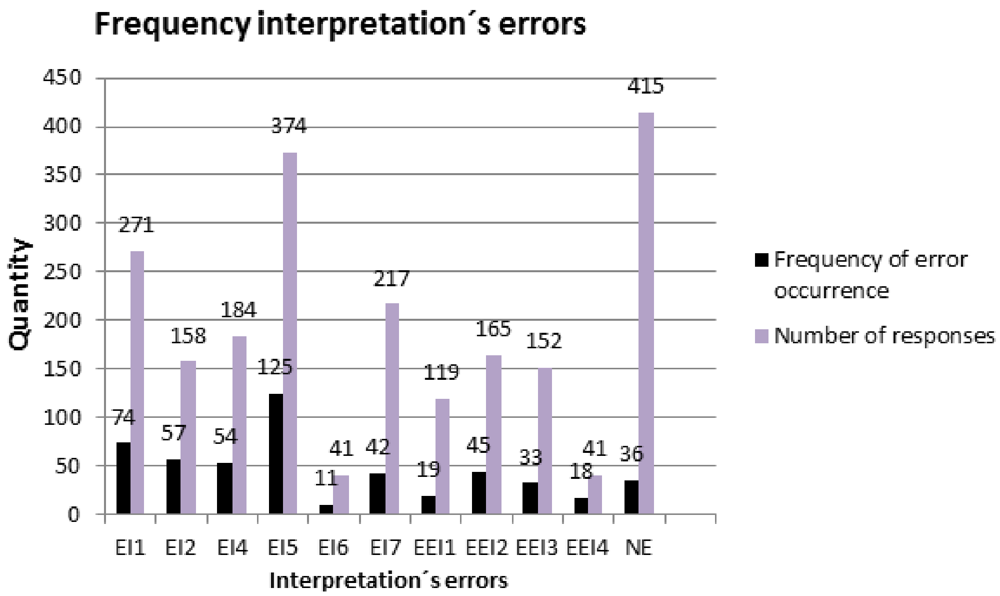
Once the whole data set has been analyzed, they are synthesized in Graph 1 and Graph 2, showing the frequency of occurrence for each error according to the process (construction or interpretation). The number of times that an error appears in all instruments is quantified regardless of the number of participants who participated in the activities. Hence, the number of times the error is displayed is greater than the number of participating students. It could be that the error was repeated several times while applying one or several activities, or different errors occurred in the same activity. Error frequency is contrasted with the total number of responses in which it could have been identified.



Graph 1. Occurrence of errors when constructing frequency tables
Source: Own elaboration

Graph 1 shows that a total of 86 evidences were obtained for the first construction error [EC1] out of a total of 277 answers (responses from all participants, as well as those responses associated with all the situations in which the error could have been made). Conversely, Graph 2 shows that 74 out of a total of 271 answers obtained for all the situations addressing the first interpretation error [IE1] show the error.

DOI: http://dx.doi.org/10.20396/



Graph 2. Occurrence of errors when interpreting frequency tables
Source: Own elaboration

Once the evidence has been collected, organized and analyzed, it is possible to consolidate a taxonomy to characterize errors and potential difficulties that generate them. Table 5 shows not only the errors evidenced by instruments but also those errors deemed relevant and identified during teaching pedagogical practices. These errors did not have resonance in the research although they were contemplated to emerge through situations proposed in instruments.

Table 5 – Taxonomy of errors and difficulties when constructing and interpreting frequency tables

CONSTRUCTION PROCESS	
ERRORS	**DIFFICULTIES**
EC_1.Ignore data dispersion by inadequately choosing interval width. Provide too many or too few intervals when working with grouped data.	DC_1.Omit statistical data representativeness.
EC_2.Combine non-comparable data (e.g. include in the table values that the statistical variable cannot take due to its qualitative or quantitative nature).	DC_2.Ignore coherence between the situation in which statistical data is framed and statistical variable values.
EC_3.Invert elements involved in the algorithm to calculate relative simple or cumulative frequency (numerator- denominator).	DC_3.Get confused about the definition for relative frequency as a part-whole relation or in the division algorithm.
EC_4.Omit an interval for having zero frequency (grouped data).	DC_4.Do not recognize zero as the possible frequency value of a data subset nor its meaning as value absence for the variable.
EC_5.Determine intervals with different widths in the same frequency distribution of grouped data.	DC_5.Ignore equality among lengths of intervals.
EC_6.Invert types of frequencies when constructing the table (e.g. place relative simple frequencies in the column for absolute simple frequencies).	DC_6.Lack of knowledge about frequency types.

DOI: http://dx.doi.org/10.20396/

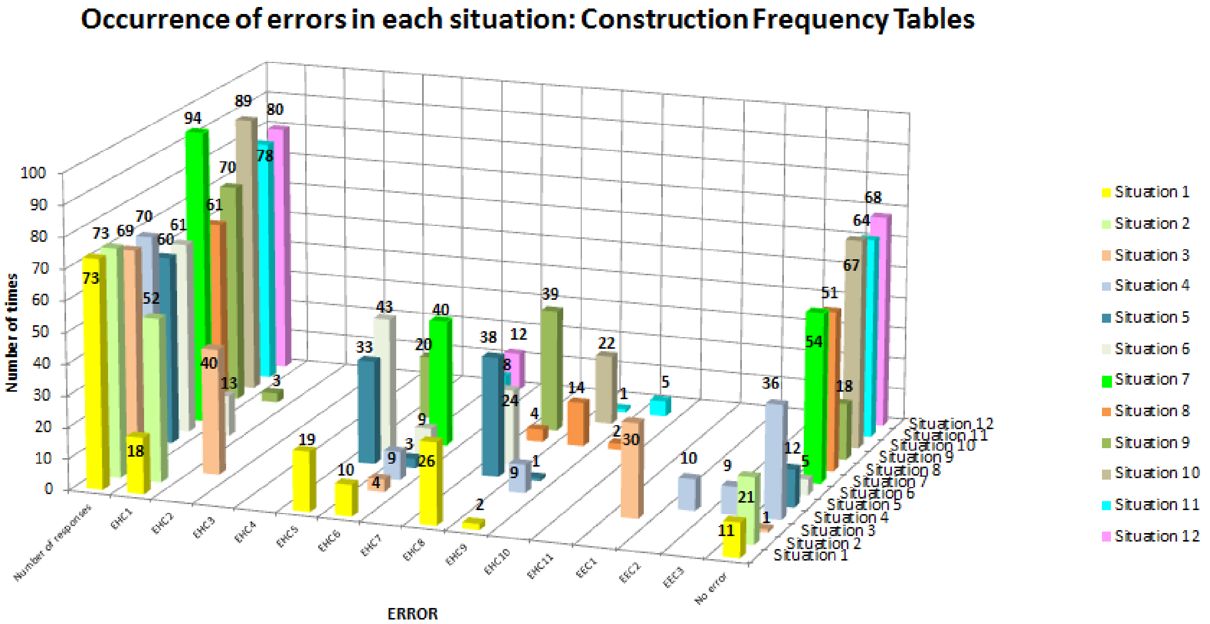
EC_7.Invert the values of a statistical variable with their respective frequencies.	DC_7.Get confused about the meaning of "value of the statistical variable". DC_8.Interpret wrongly the statistical data displayed in texts (news, blogs).
EC_8.Exclude or repeat data when constructing frequency tables for grouped data sets.	DC_9.Use wrongly open intervals or closed intervals.
EC_9.Calculate incorrectly cumulative frequency.	DC_{10}.Lack of knowledge about addition algorithms.
EC_{10}.Omit statistical variable values when constructing frequency tables.	DC_{11}.Misinterpret statistical data from verbal representation.
EC_{11}.Invert frequencies of two or more statistical variable values.	DC_{12}.Ignore the correspondence between data counting and the value of the statistical variable.
EC_{12}.Include information from personal perceptions without drawing attention to statistical information about the situation.	DC_{13}.Ignore or omit information about the situation; subjectivity is prioritized.
EC_{13}.Construct frequency tables whose information does not match the information presented in the situation.	
EC_{14}.Determine cumulative frequency based on subtracting frequencies.	DC_{14}.Lack of knowledge about the meaning of cumulative frequency.
INTERPRETATION PROCESS	
EI_1.Provide the frequency of the variable value which is not the type of frequency required.	DI_1.Lack of knowledge about the types of frequency (absolute, relative, cumulative).
EI_2.Establish inconsistent relationships among the frequencies of two or more statistical variable values.	DI_2.Lack of knowledge about numeric relationships among two or more frequencies.
EI_3.Establish trends in a data set incomprehensibly through the information presented in the frequency table.	DI_3.Visualize part of the statistical data set.
EI_4.Associate the value of the statistical variable with one of the frequencies provided in the table that does not correspond to it.	DI_4.Ignore conventions when interpreting statistical tables. Read horizontally table rows since the type of frequency provided matches the one requested despite the fact that it is not the corresponding value.
EI_5.Conclude information that does not belong to statistical data provided in the frequency table.	DI_5.Induce carelessly, trying to generalize the information.
EI_6.Ignore dispersion of statistical data.	DI_6.Omit data representativeness.
EI_7.Change the values of a statistical/quantitative variable for the values of their respective frequencies.	DI_7.Get confused about the meaning of "value of the quantitative variable" versus frequency of the variable value.
EI_8.Identify the required value of the variable; however, include or exclude values of grouped data.	DI_8.Interpret wrongly open or closed intervals in tabular representations of grouped data.
EI_9.Choose an interval representative (e.g. initial value, final value, substraction or addition of the ends) as a value of the quantitative variable.	DI_9.Lack of knowledge about the meaning of value of the statistical variable.
EI_{10}.Confuse the greatest value of the variable and the greatest frequency, in the case of the quantitative variable.	DI_{10}.Lack of knowledge about or confuse the meaning of value of the statistical/quantitative variable with absolute frequency.
EI_{11}.Conclude information from personal perceptions without taking into consideration data provided in the table.	DI_{11}.Ignore or omit information from the table or the situation; subjectivity is prioritized.

Source: Own elaboration

Graphs 3 and 4 show general results of analyses. These graphs not only display all the answers obtained for each situation but also represent the frequency of occurrence of each error for each situation. Based on these data, it is possible to establish a ratio between the number of answers obtained and the number of responses with errors.

DOI: http://dx.doi.org/10.20396/

For instance, Graph 3 shows that error [EC6] was identified in 12 out of 80 answers collected for Situation 12; there was no error in 68 answers. Only 11 out of 73 answers collected for Situation 1 are correct. It is, therefore, inferred that there were 62 answers incorrect, but 75 errors were calculated (EC1, EC5, EC6, EC8 and EC9), which indicates that the same error was repeated more than once in this situation.

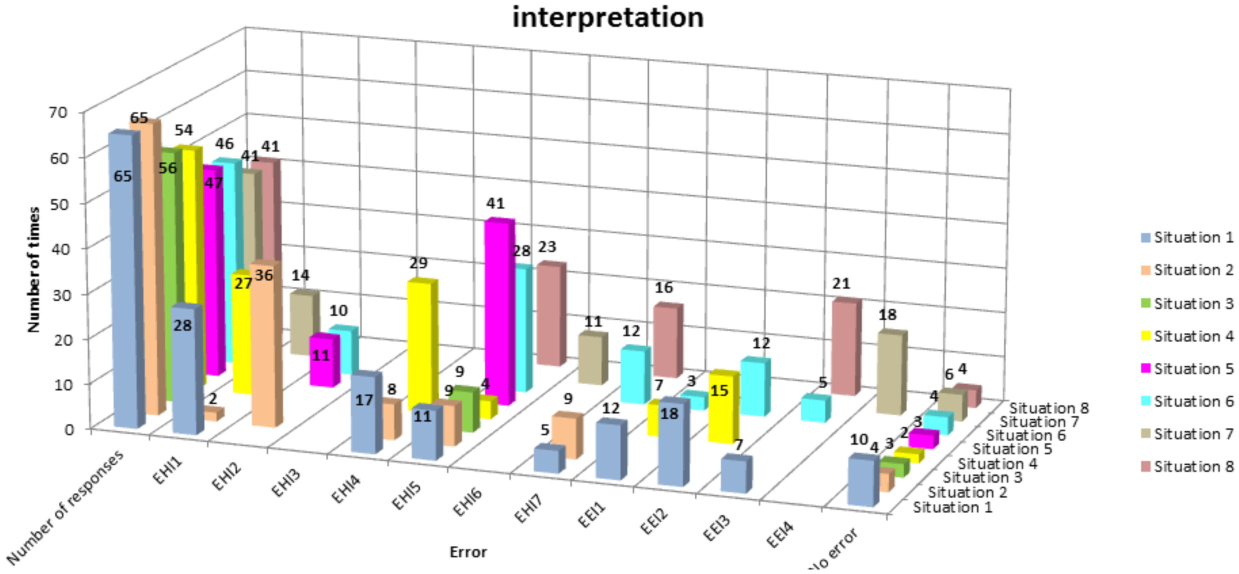


Graph 3. Occurrence of errors in each situation: Construction of frequency tables
Source: Own elaboration

Based on these data, it can be demonstrated that i. the existence of errors is particularly related to the construction of frequency tables ii. errors are different to those reported in literature for the construction of statistical graphs. Particular elements and fundamental concepts of frequency tables are revealed through such errors and their respective difficulties. These results may lead statistics educators to draw their attention on the Taxonomy set out in Table 5, and based on it, to generate strategies to avoid such difficulties and, in consequence, overcome those errors reported in this research.

DOI: http://dx.doi.org/10.20396/

Graph 4. Occurrence of errors in each situation: Frequency tables interpretation



Graph 4. Occurrence of errors in each situation: Interpretation of frequency tables
Source: Own elaboration

One of the strengths of this research is data potential to be able to characterize participants in detail with respect to the proposed aim. Regarding the errors made while interpreting frequency tables, it was thus verified that most students repeatedly make error EI_5, that is, to conclude information that does not match the data presented in the table. As far as grouped data is concerned, recurring errors are: error EI_8, identify variable value requested (values of grouped data are included or excluded) and error EI_9, choose an interval representative as the quantitative variable value (e.g. initial value, final value, substraction or addition of ends).

Based on the analyses carried out on each of the answers collected, it is noticeable that most of the difficulties identified have the same origin. Table 6 shows that difficulties are associated with mathematics teaching processes. They are specifically related to previous knowledge about the use and interpretation of intervals in frequency tables. This can be stated since all students made error EC_5 (Determine intervals with different widths) at least once, or error EC_8 (Exclude or repeat data when constructing the table). These results suggest the need to deepen in the construction of frequency tables with grouped data, making their elements and the relation among each of them explicit, namely, elements related to intervals, and the difference between frequency, variable value and representative data.

Table 6 – Type of difficulties when constructing and interpreting frequency tables

DIFFICULTIES WHEN CONSTRUCTING [DC_x] AND INTERPRETING [DI_x]	TYPE OF DIFFICULTY (Socas, 1997)
DC_1, DC_2, DC_3, DC_4, DC_6, DC_7, DC_8, DC_9, DC_{10}, DC_{11}, DC_{12}, DC_{13}, DC_{12}, DC_{13}. DI_1, DI_2, DI_3, DI_4, DI_5, DI_6, DI_8, DI_9, DI_{10}.	Associated with teaching processes developed for learning mathematics: in particular those related to previous knowledge.

DOI: http://dx.doi.org/10.20396/

DC_5. DI_7.	Associated with object complexity in mathematics: Variety of meanings.
DC_8, DC_9, DC_{13}, DEC_{12}. DI_4, DI_5, DEI_{11}	Associated with affective and emotional attitudes towards mathematics: anxiety about completing a task.

Source: Own elaboration

Besides results associated with the objective under study (Taxonomy and characterization of participants), this research provides another added value: a compilation of situations presented in both questionnaires. That means that instruments for collecting information become tools for professors, and can be replicated in order to characterize diverse school populations. Based on the results obtained from the research, it is an invitation to refine instruments, and define or expand the Taxonomy, as it seeks to enrich the theoretical contribution of this research.

Conclusions

Information collected from several education institutions and different school levels provides a strong basis to make a complete and wide characterization of errors and their difficulties. Therefore, it is possible to state that occurrence of these errors is independent from teacing processes conducted at every institution.

It might be specified that errors and inherent difficulties to the statistical object and its linked processes are mainly conditioned by lack of previous knowledge, the context provided in some situations and the complexity of frequency tables constructs. This reveals a close relationship between difficulty, obstacle and error, and exposes a potential path to study the characterization of obstacles generated by errors mentioned herein and the typology of such obstacles.

In accordance with several references mentioned in the conceptual framework, frequency tables are recognized and accepted as one of the main ways to organize, summarize and represent data. Nevertheless, it is observed that frequency tables do not have the same nature as statistical graphs despite the fact that they share some conceptual elements (statistical variables, value or categories of the variable, frequencies, etc.).

It is evident that errors and difficulties address particular frequency table structural concepts such as interval definition and construction for grouping data; differentiation of frequency and variable value; arithmetic calculations to determine types of frequency; different types of frequency and its use when analyzing data, etc. This allows stating that errors in statistical graph teaching-learning are not directly and distinctively transferable to frequency tables. However, it is necessary to corroborate the following hypothesis: if errors reported in the construction of graphs are the result of incorrect construction or interpretation of frequency tables that serve as the basis for their creation, do errors in the construction and interpretation of frequency tables turn into obstacles when constructing or reading statistical graphs?

DOI: http://dx.doi.org/10.20396/

Since several errors can be observed and, consequently, difficulties in both the construction and interpretation of frequency tables are observable as well, there is an immediate need to implement strategic training plans regarding the conceptual structure (frequency tables). This is a crucial issue for information organization and representation statistical processes. The objective is to support training of statistically-literate citizens using interpretation processes by means of frequency table critical reading and data organization and representation via such tables.

Acknowledgments

We would like to thank *Pedagógica Nacional University* of Colombia and particularly the Mathematics Department for their support during the research process. We would also like to thank the Research Incubator in Statistics Education for its accompaniment and support in the production of this article.

References

Bachelard, G. (1988). *La formación del espíritu científico*. México: Siglo XXI.

Batanero, C. (2001). *Didáctica de la Estadística*. Departamento de Didáctica de la Matemática. Facultad de Ciencias de la Educación. Universidad de Granada. Granada, España.

Batanero, C., & Godino, J. (2002). *Estocástica y su Didáctica para maestros*. Granada, España: Universidad de Granada, Departamento de Didáctica de la Matemática.

Friel, S., Curcio, F., & Bright, G. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education*, *32*(2), 124-158.

Ministerio de Educación Nacional (MEN). (1998). *Lineamientos Curriculares de Matemáticas*. MEN. Bogotá D.C., Colombia.

Ministerio de Educación Nacional (MEN). (2006). *Estándares Básicos de competencias en Matemáticas*. MEN. Bogotá D.C., Colombia.

Rodríguez, F., & Sandoval, P. (2012). Habilidades de codificación y decodificación de gráficos estadísticos: Un estudio comparativo en profesores y alumnos de pedagogía en enseñanza básica. *Revista da Avaliação da Educação Superior*. *17*(2), 207-235.

Socas, M. (1997). Dificultades, obstáculos y errores en el aprendizaje de las matemáticas en la educación secundaria. En L, Rico (Ed.), *La Educación Matemática en la Enseñanza Secundaria* (pp. 125-154). Barcelona, España: Horsori.

Tauber, L. (2010). Análisis de elementos básicos de alfabetización estadística en tareas de interpretación de gráficos y tablas descriptivas. *Ciencias Económicas*, (1), 53-74.

DOI: http://dx.doi.org/10.20396/

APPENDIX A. Excerpt from a questionnaire for identifying errors in frequency table construction

1. En un estudio se pretende identificar a qué edad suelen presentarse mayores accidentes en moto, para ello se encuestaron 48 motociclistas. Aquí se muestra la información recopilada.

16	28	26	15	33	18	16	30	22	27	21	42
19	46	34	23	40	16	32	27	25	40	30	37
41	21	22	30	45	51	26	48	18	39	50	36
47	48	20	22	47	19	16	51	20	17	27	32

Construya la tabla de distribución de frecuencia absoluta que represente la información presentada anteriormente usando intervalos.

2. Una compañía de galletas desea saber la calidad (buenas, muy buenas regulares o malas) de unas galletas que recién salieron al mercado y el gusto de la población hacia ellas para determinar su continuidad; para ello se ofrecieron muestras gratis a la salida de un supermercado y se realizó una encuesta a quienes recibían la muestra. Cinco horas después se recopiló la información obtenida, que se muestra a continuación:

- 15 personas la clasificaron como buena
- 16 personas dijeron que era demasiado blanda
- 21 personas concluyeron que su tamaño era exagerado
- 32 personas dijeron que tenía buen sabor pero en su opinión podría agregarse más chocolate
- 16 personas la clasificaron en muy buena, aunque podría ser más crujiente

Construya la tabla de frecuencias que representa la información suministrada por las personas encuestadas, en relación con la calidad de las galletas.

3. Complete la siguiente tabla que muestra la cantidad de personas que fallecieron por el síndrome de inmunodeficiencia adquirida (SIDA) entre 1988 y 2008 en Colombia.

Año	Frecuencia absoluta simple:	Frecuencia relativa simple	Frecuencia absoluta acumulada	Frecuencia relativa acumulada
[1988-1992)	3769			
[1992-1996)	10687			
[1996-2000)	12126			
[2000-2004)	15091			
[2004-2008]	23036			

4. En una estación de policía se tiene el siguiente reporte de las velocidades a las que iban 40 conductores multados por exceso de velocidad, en zonas donde el máximo permitido es 79 km/h.

120	90	90	91	92	91,4	89,3	92	83	107
88,9	106,9	92	85	87	88,3	104,3	90	90	82
83,5	87	95	96	119,3	100	99	88	91,8	115,7
91,5	120	85	100	102,7	92	116	80	103,6	116

Construya la tabla de distribución de frecuencia absoluta que represente la información respecto a las velocidades a las que se multaron los conductores, usando 8 intervalos.

DOI: http://dx.doi.org/10.20396/

5. En una fábrica se quiere establecer el tiempo de duración de una nueva bombilla, para ello se escoge una muestra al azar de 30 bombillas y se anota su tiempo de duración en horas, en seguida se muestran los resultados:

1000, 8	830,4	12300	14962,3	11623.2	8020,3	1592,9	9849
643,9	9700	11121	15000	14328,3	13002	11740,6	2535,7
1287,5	14728,6	15.200	1493,8	15000,6	12506,4	10002,4	
13621,7	2381,6	650,9	5291,4	9983,4	9630,9	7499,2	

Construya la tabla de distribución de frecuencia absoluta que represente la información respecto al tiempo de duración de las bombillas usando intervalos.

6. En una institución educativa se realizó una prueba de matemáticas a 200 estudiantes de grado undécimo con el fin de prepararlos para la aplicación de las Pruebas Saber cuya puntuación máxima es de 100 puntos. Para incentivar a los estudiantes en el desarrollo consiente de la prueba se prometió otorgar el 30% de descuento de la excursión a los mejores puntajes, mientras que los estudiantes con puntajes más bajos tendrían que asistir a clases de refuerzo y mejoramiento durante todo el año. Teniendo en cuenta que al obtener los resultados de la prueba se realizaron 20 descuentos por el puntaje correspondiente a 90 puntos, que el puntaje más bajo fue de 30 puntos y que fue necesario preparar 50 planes de mejoramiento; y por otro lado que a los 70 estudiantes que obtuvieron 40 puntos se solicitó un certificado de clase particulares para obtener mejores resultados en la prueba; y que los estudiantes restantes obtuvieron la mitad de la puntuación total, para ellos se dispuso un cuestionario que permitiera saber porque solo obtuvieron este puntaje, construya la tabla de frecuencias correspondiente.

APPENDIX B. Excerpt from a questionnaire for identifying errors when interpreting frequency tables

1. A continuación se presenta el tiempo de vida útil de una muestra de 250 televisores producidos por una empresa.

Vida útil (días)	Frecuencia absoluta simple	Frecuencia relativa simple	Frecuencia absoluta acumulada	Frecuencia relativa acumulada
[1095-3095)	76	$^{76}/_{250}$	76	$^{76}/_{250}$
[3095-5095)	118	$^{118}/_{250}$	194	$^{194}/_{250}$
[5095-7095]	56	$^{56}/_{250}$	250	$^{250}/_{250}$

Con base en la información anterior, responda:

a. ¿Cuál es el tiempo de vida útil de los televisores cuya frecuencia absoluta simple es 118?
b. ¿Cuántos televisores tienen entre 3095 y 5095 días de vida útil?
c. En el caso de la frecuencia relativa simple ¿Cuántos televisores representan mayor cantidad en cuanto al tiempo en vida útil respecto al total? ¿Por qué?
d. ¿Cuál es el número total de televisores que duran entre 1095 y 5095 días? Explique cómo obtuvo ese valor

2. En un centro comercial se recolectó la siguiente información respecto a las fechas favoritas de 900 personas:

DOI: http://dx.doi.org/10.20396/

Fecha	Frecuencia absoluta simple	Frecuencia relativa simple	Frecuencia absoluta acumulada	Frecuencia relativa acumulada
Amor y amistad	178	$178/900$	178	$178/900$
Mes del padre	228	$228/900$	406	$406/900$
Mes de la madre	290	$290/900$	696	$696/900$
Halloween	122	$122/900$	818	$818/900$
Navidad	82	$82/900$	900	$900/900$

Responda:

a. Con base en la frecuencia absoluta simple ¿Cuál es la fecha favorita de las personas encuestadas?
b. Con base en la frecuencia relativa simple ¿Cuál es la fecha favorita de las personas encuestadas?
c. Encuentra alguna relación entre las respuestas en los ítems anteriores
d. De acuerdo con la información planteada en la columna de frecuencia relativa simple, en relación con la encuesta ¿Qué época es mejor Halloween o Navidad?
e. ¿Cuántas personas prefieren las fechas especiales anteriores al mes de octubre? ¿Cómo obtuvo el resultado?
f. Con base en las frecuencias relativas ¿las personas que prefieren amor y amistad y mes del padre, representan más respecto al total de las que prefieren Halloween? Explique su respuesta

3. La siguiente tabla presenta información en relación con los ingresos mensuales de 100 familias residentes en la localidad de Engativá. Los resultados se organizaron en cinco categorías, como se muestra en la siguiente tabla.

Categoría	Ingresos (Pesos)	Frecuencia absoluta simple	Frecuencia relativa simple	Frecuencia absoluta acumulada	Frecuencia relativa acumulada
1	[700.000-900.000)	12	$12/100$	12	$12/100$
2	[900.000-1100000)	35	$35/100$	47	$47/100$
3	[1.100.000-1.300.000)	22	$22/100$	69	$69/100$
4	[1.300.000-1.500.000)	18	$18/100$	87	$87/100$
5	[1.500.000-1.700.000]	13	$13/100$	100	$100/100$

Teniendo en cuenta la información de la tabla anterior, responda:

a. ¿Cuál es el ingreso de las 35 familias que pertenecen a la categoría 2?
b. ¿Cuál es el número total de familias que ganan entre $700000 y $1300000? Explique su respuesta.
c. ¿Cuántas familias ganan entre $1300000 y $1500000?
d. Con base en la frecuencia relativa simple, ¿Cuál es el número total de familias que ganan entre $1500000y $1700000? Explique su respuesta.
e. ¿En qué categoría se encuentran las familias que ganan $1.100.000? ¿Cuántas familias pertenecen a esta categoría?

4. En la siguiente tabla se muestra información respecto al número de personas reportadas como desaparecidas en los últimos cinco años en Colombia

DOI: http://dx.doi.org/10.20396/

Año	Frecuencia absoluta simple	Frecuencia relativa simple	Frecuencia absoluta acumulada	Frecuencia relativa acumulada
2012	5452	$5452/_{21500}$	5452	$5452/_{21500}$
2013	3994	$3994/_{21500}$	9446	$9446/_{21500}$
2014	3867	$3867/_{21500}$	13313	$13313/_{21500}$
2015	4239	$4239/_{21500}$	17552	$17552/_{21500}$
2016	3948	$3948/_{21500}$	21500	$21500/_{21500}$

Responda:

a. Con base en la tabla, Alison afirma que el 2014 fue el año en el que se reportaron menos desaparecidos en la historia de Colombia ¿Está de acuerdo con ella? ¿Por qué?
b. ¿En qué año, de los incluidos en la tabla, hubo más personas reportadas como desaparecidas?
c. Escriba una conclusión con base en la columna frecuencia absoluta simple de la tabla.

5. En una escuela de patinaje inscribieron a sus estudiantes en un torneo en el cual las categorías se determinan de acuerdo a la edad de los competidores. Los directivos de la escuela reciben el siguiente reporte respecto al rendimiento de sus estudiantes, la valoración se presenta sobre 100 puntos y es el puntaje total obtenido por categoría (no por competidor).

Categoría (edad)	Frecuencia absoluta simple (Puntos)	Frecuencia relativa simple	Frecuencia absoluta acumulada	Frecuencia relativa acumulada
[5-8)	78	$78/_{320}$	78	$78/_{320}$
[8-11)	75	$75/_{320}$	153	$153/_{320}$
[11-14)	86	$86/_{320}$	239	$239/_{320}$
[14-17]	81	$81/_{320}$	320	$320/_{320}$

Con base en la información presentada en la tabla anterior responda:

a. Los directivos quieren identificar las edades de los tres competidores que obtuvieron más bajo rendimiento para aumentar su tiempo de entrenamiento, podría decir ¿Quiénes deberán entrenar más tiempo? ¿Por qué?
b. Los directivos quieren identificar la categoría cuyos competidores tuvieron más bajo rendimiento ¿Cuál fue? ¿Por qué?
c. Podría decir cuál fue la categoría en que los competidores tuvieron mejor rendimiento ¿Cuál fue su puntaje?
d. El próximo torneo se llevará a cabo en la ciudad de Medellín tres meses después de haber recibido los resultados de esta competencia. ¿Podría determinar cuál categoría obtendrá más puntos en dicho torneo? ¿Por qué?

6. Un grupo de adultos mayores (50 años en adelante) ha demandado a la EPS “Salud para todos” alegando que durante la entrega de medicamentos del último mes se dio prioridad a los más jóvenes (personas entre 20 y 49 años). Sin embargo la EPS ha presentado en su defensa la información, respecto a la entrega de medicamentos, según se contempla en la siguiente tabla.

Edad (años)	Frecuencia absoluta simple	Frecuencia relativa simple	Frecuencia absoluta acumulada	Frecuencia relativa acumulada
[20-30)	38	$38/_{200}$	38	$38/_{200}$
[30-40)	42	$42/_{200}$	80	$80/_{200}$

DOI: http://dx.doi.org/10.20396/

[40-50)	35	$35/_{200}$	115	$115/_{200}$
[50-80]	85	$85/_{200}$	200	$200/_{200}$

Con base en la información presentada en la tabla responda

a. ¿Cuántas personas de 30 años reclamaron medicamentos el mes pasado?
b. ¿Cuántas personas entre 20 y 29 años reclamaron medicamentos el mes pasado?
c. ¿Cuántas personas entre 50 y 80 años reclamaron medicamentos el mes pasado?
d. De acuerdo con la cantidad de medicamentos entregados por la EPS, la cual se sustenta en la tabla ¿Respalda usted la postura de los adultos mayores o de la EPS? ¿Por qué?

7. En seguida se presenta una tabla de frecuencias que muestra información respecto a la variable "talla de calzado" de las 190 mujeres adultas que habitan el municipio de Miraflores, Guaviare. Observe la tabla y responda:

Valor de la variable	Frecuencia absoluta simple	Frecuencia relativa simple	Frecuencia absoluta acumulada	Frecuencia relativa acumulada
34	40	$40/_{190}$	40	$40/_{190}$
35	37	$37/_{190}$	77	$77/_{190}$
36	35	$35/_{190}$	112	$112/_{190}$
37	39	$39/_{190}$	151	$151/_{190}$
38	39	$39/_{190}$	190	$190/_{190}$

a. Suponga que por las fiestas decembrinas el único almacén de calzado de la zona ha agotado todos los zapatos talla 35 para dama. Si sólo 15 de las mujeres cuya talla de calzado es 35 alcanzaron a comprar los zapatos para navidad y todas estaban interesadas en hacerlo. ¿Cuántas se vieron afectadas por la falta de calzado?
b. ¿Cuál es la talla de calzado más alta que debería encargar la tienda de zapatos para preparar sus ventas de fin de año?