Revista Internacional de Educação Superior
International Journal of Higher Education

Teaching of probability and statistics by exploratory data analysis and resolution of problems

Fabiano dos Santos Souza[1]
[1] Universidade Federal Fluminense

ABSTRACT

The present paper is based on studies, research and reflections of a group of research of teaching and learning in mathematics and statistics, which were converted into a methodological approach to the teaching of probability and statistics implemented a workshop given at the III Day of Mathematics of Laboratory of Mathematics Education (LABEM) of the University Federal Fluminense. In the study participated in 24 subjects of the course of degree in mathematics, in order to assess the proposal in relation to the teaching and learning process of the basic content of statistics and probability. Theoretical and methodological issues on teaching probability and statistics were discussed, reflected and worked during the workshop, which was organized under the perspective of Batanero (2001) and through the troubleshooting Walle (2009). The main objectives of this workshop were: (i) identify the statistic as a powerful tool for analysis of data, which can be used to balance techniques and concepts; (ii) provide participants a reflection space for the construction of a collective work, that connect academic and practical knowledge. The results of this study found that the teaching of probability and statistics when is performed through the methodology of problem solving and exploratory data analysis, becomes more significant, causing the knowledge gained in this process are connected with the daily life of students and with their respective previous knowledge, providing the construction of statistical literacy.

KEYWORDS

Education statistics. Probability and statistics teaching. Teacher education.

Corresponding Author
[1] Fabiano dos Santos Souza
E-mail: fabiano_souza@id.uff.br
Universidade Federal Fluminense, Brasil
CV Lattes
http://lattes.cnpq.br/0468143359513798

Submmited: 14 Jun. 2018
Accepted: 19 Sept. 2018
Published: 10 Oct. 2018

doi 10.20396/riesup.v5i0.8652708
e-location: e019002
ISSN 2446-9424

Anti-plagiarism check
turnitin

Ensino de probabilidade e estatística por meio da análise exploratória de dados e resolução de problemas

RESUMO

O presente artigo baseia-se em estudos, pesquisas e reflexões de um grupo de pesquisa de Ensino e Aprendizagem em Matemática e Estatística, os quais foram convertidos em uma proposta metodológica para o ensino de probabilidade e estatística implementada numa oficina ministrada no III Dia da Matemática do Laboratório de Educação Matemática (LABEM) da Universidade Federal Fluminense. No estudo participaram 24 sujeitos do curso de licenciatura em Matemática, com objetivo de avaliar a proposta em relação ao processo de ensino e aprendizagem dos conteúdos básicos de estatística e probabilidade. Questões teóricas e metodológicas sobre o ensino de probabilidade e estatística foram discutidas, refletidas e trabalhadas durante a oficina, que foi organizada sob a perspectiva de Batanero (2001) e por meio da resolução de problemas Walle (2009). Os objetivos principais desta oficina foram: (i) identificar a Estatística como sendo uma poderosa ferramenta para análise de dados, que pode ser utilizada de forma a equilibrar técnicas e conceitos; (ii) proporcionar aos participantes um espaço de reflexão para a construção de um trabalho coletivo, que interligue saberes acadêmicos e práticos. Os resultados deste trabalho constataram que o ensino de probabilidade e estatística quando é realizado por meio da metodologia da resolução de problemas e da análise exploratória de dados, torna-se mais significativa, fazendo com que os conhecimentos adquiridos nesse processo estejam conectados com o cotidiano dos alunos e com os seus respectivos conhecimentos prévios, proporcionando a construção do letramento estatístico.

PALAVRAS-CHAVE

Educação estatística. Ensino de probabilidade e Estatística. Formação de professores.

Enseñanza de probabilidad y estadística por medio del análisis exploratorio de datos y resolución de problemas

RESUMEN

El presente artículo se basa en estudios, investigaciones y reflexiones de un grupo de investigación de Enseñanza y Aprendizaje en Matemática y Estadística, los cuales fueron convertidos en una propuesta metodológica para la enseñanza de probabilidad y estadística implementada en un taller impartido en el III Día de las Matemáticas del Laboratorio de Educación Matemática (LABEM) de la Universidad Federal Fluminense. En el estudio participaron 24 sujetos del curso de licenciatura en Matemáticas, con el objetivo de evaluar la propuesta en relación al proceso de enseñanza y aprendizaje de los contenidos básicos de estadística y probabilidad. Las cuestiones teóricas y metodológicas sobre la enseñanza de probabilidad y estadística fueron discutidas, reflejadas y trabajadas durante el taller, que fue organizada bajo la perspectiva de Batanero (2001) y por medio de la resolución de problemas Walle (2009). Los objetivos principales de este taller fueron: (i) identificar la Estadística como una poderosa herramienta para análisis de datos, que puede ser utilizada de forma a equilibrar técnicas y conceptos; (ii) proporcionar a los participantes un espacio de reflexión para la construcción de un trabajo colectivo, que interconecte saberes académicos y prácticos. Los resultados de este trabajo constataron que la enseñanza de probabilidad y estadística cuando es realizada por medio de la metodología de la resolución de problemas y del análisis exploratorio de datos, se vuelve más significativa, haciendo que los conocimientos adquiridos en ese proceso estén conectados con el cotidiano de los alumnos y con sus respectivos conocimientos previos, proporcionando la construcción del texto estadístico.

PALABRAS CLAVE

Educación estadística. Enseñanza de probabilidad y Estadística. Formación de profesores.

Initial Considerations

With the advancement of technology and computational resources, Statistics has become one of the most important sciences in the last decade, being present in almost all areas of knowledge, in the media, where we find graphs, tables, statistical data, information that many are not understood by many citizens. In this direction, we defend a statistical formation from the initial series, so that our children learn from an early age to deal with real situations and that they can: read, understand, analyze, interpret statistical information of daily life, helping in decision making from the apprehension and understanding of the variability contained in the data, becoming critical citizens. Thus, the understanding of these everyday situations requires the citizen to reason statistically and obtain a level of statistical literacy.

In this perspective, Garfield (2002) defines statistical reasoning as the way that a person reasons with statistical ideas and makes sense with statistical information. Statistical literacy is understood not only literacy, but the correct use of concepts and statistical procedures by the subject as pointed out by Coutinho (2013).

Sá, Silva, and Samá (2015, p. 155) emphasize the need to have a formation that concretises the statistical literacy in Basic Education, "since the understanding and the statistical understanding are fundamental for the individual to be able to make inferences, act as a prudent consumer, and make decisions in their professional or personal lives".

In view of this scenario, it is important to highlight that the Federal Constitution of 1998 and the Brazilian educational policy, such as the National Education Guidelines and Bases Law (LDBEN 9394/96), was fundamental for the creation of National Curricular Parameters (NCP) and Secondary Education in 1997, as well as the creation of the National Curriculum Guidelines for Secondary Education. These documents emphasize the importance of the students of Basic Education developing the probabilistic thinking, combinatorial since the initial grades. For this purpose, the Information Processing Block for Basic Education and Data Analysis for High School was created, seeking an integration of basic notions of Descriptive Statistics with Combinatorial Analysis and Probability, with the intention that the student builds his knowledge through through everyday situations. Therefore, teachers of Basic Education should mediate learning management situations that provide students with a statistical culture.

In December 2017, recently approved the National Curricular Common Base (NCCB) for Early Childhood Education and Elementary Education provides guidelines for the teaching of Statistics and Probability for teachers to carry out with their students' research processes involving social issues, their daily lives and interdisciplinary. In this way, students are expected to be able to develop the Statistical Literacy in an appropriate way, becoming a critical, reflective and participatory citizen (SANTOS; SANTOS JUNIOR; VELASQUE, 2018).

> The National Curricular Common Base, in relation to Mathematics, also approaches National Curricular Parameters, considering that these documents aim at the construction of a referential that orient the school practice in order to contribute so that all the Brazilian students have access to a mathematical knowledge that enables them, in fact, to insert them as citizens in the world of work, social relations and culture. (BRAZIL, 2016, p. 134).

In this regard, Gal (2002) reveals the need to provide a statistical culture for society where they can develop the capacity to interpret and critically evaluate statistical information, the ability to discuss and speak their opinions about statistical information.

On the development of a statistical culture, Coutinho and Souza (2015, p. 121) emphasize “the importance of reflection on the teaching and learning process of the statistical contents that allow the development of this type of culture, important both for professional practice and for the full exercise of citizenship”.

According to Batanero (2001), probability and statistics can be applied to reality as directly as elementary arithmetic since they do not require complicated mathematical techniques. The researcher emphasizes that, one must make that its teaching is realized through a heuristic and active methodology, emphasizing the experimentation and the resolution of problems. For the author, probability teaching should be able to improve probabilistic reasoning so that students can better deal with the random situations of their day to day and thus improve their probabilistic intuition.

Regarding Statistics Teaching, Lopes (2005) emphasizes that it is important to develop statistical thinking, which allows students to be able to use statistical ideas and assign meaning to the desired information. The author stresses the importance of teaching statistics and probability through experimentation, observations, records, collections and data analysis.

In this study, the definition adopted by Garfield, delMas e Chance (2003, apud BEN-ZVI and GARFIELD, 2004, p. 7) is used, which was adopted as the base:

> Statistical Literacy includes basic and important skills that can be used to understand statistical information or search results. These skills include being able to organize data, build and display tables, and work with different representations of the data. Statistical Literacy also includes an understanding of concepts, vocabulary, and symbols, and includes an understanding of probability as a measure of uncertainty.

From this perspective, a 6-hour workshop was developed in two days, whose main objective of this didactic-pedagogical proposal was to teach through situations of learning management, probability and statistical activities, using as methodology to solve problems and the exploratory analysis of data from the perspective of Batanero, Estepa and Godino (1991), who point out that the exploratory analysis of data is effective when the main focus is the development of statistical literacy in the approach of statistical content.

Activities for the Teaching of Probability and Statistics

The workshop titled: Activities for Teaching Probability and Statistics was given by the author at the Federal Fluminense University, on the III Mathematics Day, which is one of the activities of the Extension Program of the Mathematics Education Laboratory (LABEM) of this university in partnership with the Education and Training Group and Learning in Mathematics and Statistics of CNPQ.

The day of Mathematics at LABEM proposed to hold lectures, workshops, exhibitions for the students of the Degree in Mathematics, students and teachers of Basic Education. The aim of the workshop was to provide a space for reflection and construction of student and teacher knowledge, based on the discussions on the different methodologies and research related to Statistical Education and on teaching of probability and statistics.

Initially, the students were explained the importance of the tests they would carry out and of the workshop, as they would be part of a research to be developed by the teacher.

The following activities are described in the workshop: Population and sample, variable classification, deterministic and random experiment, concept and calculation of mean, mode, median, amplitude, standard deviation and variance for isolated data, sample space, independent and dependent events, definition of probability, probability of union of two events and conditional probability.

Activity 1 - *Introduction* - In this stage of the workshop, in addition to explaining how the study/work would be carried out, a dialogue was established with the students regarding the basic concepts of Statistics and how they were articulated with everyday life. The knowledge involved in this activity started from statistical literacy (GAL, 2002), which would be important for the understanding of statistical information and everyday situations, consequently the development of statistical thinking and the respective statistical literacy. Initially, it is necessary to emphasize that some basic concepts and definitions must precede the statistics calculations.

Activity 2 - *Job selection* - Matshow sent resumes to various colleges after the selection stages was called to work in three companies - A, B and C. Matshow very smartly hired a consultant to interview five professors from each company, each occupying the same function that Matshow will occupy.

	P1	**P2**	**P3**	**P4**	**P5**	**Average**	**Median**	**Mode**	**Standard Deviation**
School A	2200	2200	2200	2200	2200	2200	2200	_	0
School B	1700	2700	2500	2000	2100	2200	2100	_	400

School C	5100	1900	1333	1333	1333	2200	1333	1333	1640

Which company / school should Matshow choose?

This activity was developed through solving a problem by analyzing the following statistics: average, median, mode, standard deviation of the results obtained with salary values of the 15 teachers participating in the sample by the consulting company. The purpose of this activity was to use the results of measures of central tendency and position in this situation described to assist the teacher in decision making.

From this question, we began to understand the context of the situation of the problem, the studies carried out, alongside statistical techniques aided in decision-making. Statistical reasoning that is defined by Garfield (2002), as the way an individual reasons with statistical ideas and makes sense with statistical information. Thus, the knowledge worked on this activity were: uncertainty, sampling, measures of central tendency, variability and interpretation of the obtained results. The concepts and techniques for the calculation of average, median, mode and standard deviation, their characteristics and properties were considered, it was emphasized, therefore, the importance of working with measures of central tendency and dispersion to aid in decision making.

Activity 3 - *Comparison between average, mode and median* - For this activity, the exploratory data analysis was used where some ages of the participants were collected, which were divided in two large groups, including in the data collection and analysis. In the second time, the age of the teacher who was teaching the workshop was added, that is, an outlier (discrepant value, atypical value) of the data set, since the participants were undergraduates and there were none older than 25 years. After data collection by the students, the groups were asked to calculate some statistics: the mean, median and mode of the sample of ages of a class. Then, the following questions were asked: (i) Which measure of central tendency best describes a typical input of this data set? (ii) Is there any discrepancy? (iii) With the age of the teacher, is there any difference value? (iv) What happened to the data after the age of the teacher was shown? (v) Which measure (s) of central tendency underwent change (s) at the second moment?

The subjects, in groups, performed the calculations of the central tendency statistics but did not know how to best choose the data set. In this sense, we used the analysis of the best measure of central tendency by means of a bar graph with all the ages of the participants, in order to choose the best measure of central tendency between the average, mode and median for the obtained data the participants.

Sometimes a graphical comparison can help you decide which core trend measure best represents a set of data. In this way, the presence of the discrepant value that is also called outlier, which affects the average, but does not affect the fashion nor the median is emphasized, since fashion depends only on the value or values that show higher frequency, and the median depends on the number of elements in the data set. The average is affected by each value in the data set.

Activity 4 - *Dispersion Measures* - This activity was elaborated through two problems; the first one contained information from two patients (A and B), where a nurse checked their pulse three times a day. Patient A accused the rates of 72, 76 and 74; and patient B accused 72, 91, and 59. It was asked: which patient was the most stable and with what statistics could they use to justify the patient´s choice of?

In this question, both patients had the same average 74, there was no mode and the median of Patient A was 74 and of Patient B was 72, making it impossible to say which was the more stable of the two. However, the analysis using the measures of variability could confirm which of the two patients was the most stable. Thus, by calculating the amplitude of both, it is concluded that the amplitude of Patient A is equal to 4 and that of Patient B equals 32, that is, the amplitude of Patient B was 8 times greater than that of Patient A, therefore, Patient A was the most stable. In this way, he asked himself: Which patient was more stable and how did they do the analysis?

The second problem - A company opens a recruitment process to hire an employee, and at the end of the competition there are two candidates left for a single vacancy. Then, there were 4 equal tasks for each one, where they had as recording the time (in minutes) of execution.

Tasks	**1**	**2**	**3**	**4**
Worker 1	55	45	52	48
Worker 2	30	70	40	60

The problem was selecting the best candidate for the company. In the resolution of this problem, a line graph was plotted and the concepts of variability, amplitude, average and variance were worked out, and with the help of the graph and of these measures the students were able to choose the worker with the statistics at hand.

Activity 5 - *Exploratory data analysis* - Students were divided into three groups of eight students, who were asked to collect data of the following variables: name, age, height (cm), mass (kg), university time, work (yes or no), time spent with study per week, scholarship (yes or no), and whether to pursue a teaching career (yes or no). After data collection, some statistics were requested: younger age, oldest age, age range, average age, standard deviation of age, higher height, maximum time at university, average time spent with studies, relative frequency in percentage: work, scholarship and teaching career, and two graphs for two distinct variables. At the end of these stages, each group presented its results to the participants, after the presentations, a discussion and reflection about the results was made, and comparisons and analysis were carried out between the groups.

Activity 6 - *Probabilities* - This activity began with a little history of mathematics on the calculation of probabilities, types of probabilities, the idea of chance, uncertainty, returned to the concepts of deterministic and random until arriving at the definition Laplacian probability. With the aid of problem solving, we approached the Fundamental Principle of Counting, after this activity, it became a question of simple calculation of probability involving independent events.

Some probability properties were addressed, dependent and independent events. Therefore, we worked with three vestibular questions that involved these concepts, one of the questions, was solved through the tree of possibilities.

The activities were related, meeting the objectives of this study, measuring this methodological proposal for teaching probability and statistics through problem solving and data analysis.

Methodological Procedures and Hypotheses of the Study

In this study, the methodological procedure employed was based on the construction of pedagogical actions integrated in permanent interlocution with the subjects participating in the workshop. The workshop was held on two consecutive days with 3 hours each day. On the first day, after the explanation about the activities and the proposal of methodological adequacy for the teaching of probability and statistics, the subjects were submitted to a pre-test before starting the proposed activities of the workshop. In the second moment, the participants after the workshop finished the post-test.

The pre-test consisted of a questionnaire containing 19 multiple-choice questions. By favoring the analyzes were used as model those proposed by Lopes (2010).

After the activities of the workshop, a post-test was applied containing the same questions of the pre-test. The last question in the questionnaire asked whether the subjects had any contact with the contents of probability and statistics in Basic Education. Of the 24 participants, only seven said they had seen the contents covered in the workshop, two reported that they had had no contact and 15 had partial contact with the contents.

The application of the tests to the participants was aimed at analyzing the adequacy and/or feasibility of this methodological teaching proposal through problem solving and exploratory data analysis. The validation, in this case, was based on the analysis of the following hypotheses:

(i) Is there a difference between the average number of correct answers per student in both tests?
(ii) Is there a difference between the number of hits per question in both tests?

Statistical inference was used by means of tests considering a p value equal to or less than 0.05 as being statistically significant at the level of significance ($\alpha = 5\% = 0.05$), with a confidence interval of 95 %.

The analyzes and graphs of the information collected with the students were supported by a statistical software R version 3.4 that is a language and environment for statistical computation and for graphs, in the program the tools of R Commander (Rcmdr) are used, which is a complementary package for the R. This Rcmdr package, best known as R Commander, was developed in 2000 and has a great advantage over the work itself in R, since the package offers several menus with statistical methods without having to program them through the rows of command. Further details on R software and its Rcmdr package can be found in (DALGAARD, 2002; FOX, 2004, 2006, 2011; PETERNELLI and MELLO, 2011).

For Souza (2013, p. 87) “R software is a tool for statistical treatment of data, which offers well-designed graphs with excellent quality for publication, and may include mathematical symbols and formulas when necessary”.

Initially, the normality of the series was tested (number of correct answers per student in the pre-test, number of correct answers per student in the post-test, number of correct answers per question in the pre-test and number of correct answers per question in the post-test), using the R software the Shapiro-Wilk test was applied and the following results were obtained.

Table 1. Summary of normality tests performed by the Shapiro-Wilk test in the Rcmdr of the statistical series under analysis.

Series	**Test statistical for Shapiro-Wilk Test**	**p-value (Pre-test)**	**p-value (Post-test)**
number of correct answers per student in the pre-test	W = 0.9513	0.2888	_
number of correct answers per student in the post-test	W = 0.9267	_	0.08231
number of correct answers per question in the pre-test	W = 0.9179	0.1036	_
number of correct answer per question in the post-test	W = 0.9436	_	0.3062

Source: Author (2018).

Based on the two test results (W and p-value), which were presented in table 1, we conclude that in all the analyzed series the p-value was greater than 0.05, that is, these values are higher than the significance level of 5% ($\alpha = 0.05$), with a confidence level of 95%, therefore, it can be stated that the data come from a population with normal distribution, being possible to application of a parametric statistical test.

Thus, the parametric statistical test chosen was the *Paired t-test*, since the sample during the study was paired. One of the characteristics of the paired samples is that for each number of correct answer in the pre-test, there is a corresponding post-test and these measurements are taken in a single subject at two different time points.

Analysis, Discussion and Results

Before answering the first hypothesis: **(i)** *Is there a difference between the average number of students' correct answers in both tests?* The descriptive measures of the series were obtained (number of correct answers per student in the pre-test, number of correct answers per student in the post-test), according to table 2.

Table 2. Descriptive statistics of the number of correct answers per student in the pre-test and post-test.

Statistics\Test	**Pre-test**	**Post-test**
Average	12.08	16
Minimum	5	13
1st Quartile (25%)	9	14,75
Median (50%)	13	16
3rd Quartile (75%)	14,25	17
Maximum	17	19
Standard Deviation	3.5252	1.8415
Variance	12.4275	3.3913

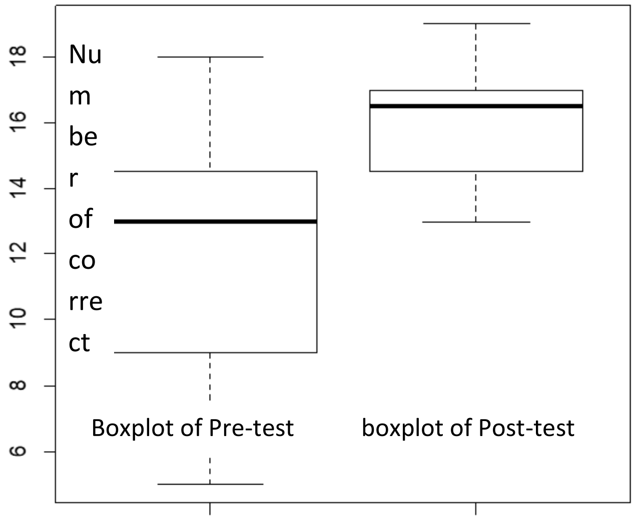
Source: Author (2018).

These statistics revealed that there are indications that the methodology applied to the teaching of statistics implemented in the workshop produced an improvement in the results obtained by the subjects in the pre-test in relation to the post-test. One of these statistical results, the average, in the post-test increased by about 32.45% over the pre-test average, the standard deviation in the post-test decreased by 47.76% and the variance 72.71%, i.e., the data became more homogeneous and more concentrated around the average.

The boxplot graph, also called boxes and mustaches (chart 1), compares the data obtained in the pre-test and post-test by highlighting the values of the following statistics: minimum, first quartile, median, third quartile, maximum.

The boxplot is a graph that makes it possible to work with two types of apprehensions: the perceptive and the discursive. The perceptive apprehension in the field of Statistical Education for Almouloud (2007) refers to the interpretation of the shapes of the figure in the context of solving problems that involve the analysis of statistical data. Discursive apprehension is essential in the explication of mathematical properties. By making an analogy to Statistics, one can make a discursive apprehension in the boxplot graph describing and explaining its properties.

Chart 1. Boxplot of the number of correct answers by students in the pre-test and post-test.

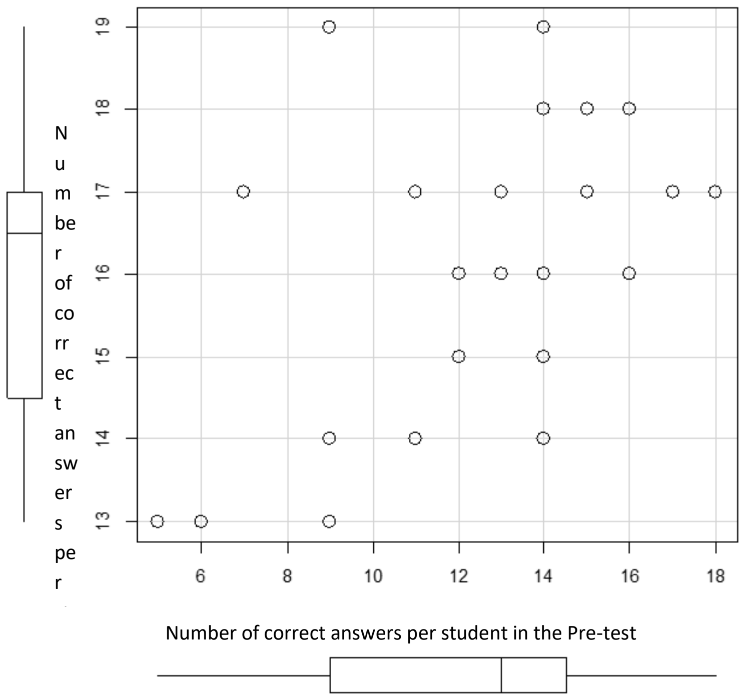


Source: Author (2018).

Note that the left boxplot for the pre-test and the right of the post-test, the more "spread" box indicates that the dispersion of the data around the average is larger than the smaller box, as shown in the table 2. The minimum score in the pre-test was 5 and in post-test 13. That is, the minimum value in the post-test was equal to the value of the median in the pre-test. It should be noted that the median value is the one that occupies the position of 50% of the data, and the maximum value in the post-test was 19.

The interquartile range which is the difference between the value that occupies the position of 75% (third quartile) of the data with the position of 25% (first quartile) was 5.25 in the pre-test and in the post-test 2.25, that is, a smaller amplitude for the post-test. This means that 50% of the data had a maximum amplitude of 2.25.

Graph 2. Comparison of the number of correct answers per student in the pre-test and in the post-test.



Source: Author (2018).

Graph 2 provides a comparison at two different times in relation to the same student. That is, it compares the number of hits per student in the pre-test and in the post-test. In the bars of the horizontal and vertical axes we have the marginal boxplot relative to the observed data, showing all measures of the quartiles, minimum, maximum, interquartile amplitude and revealing that the variance decreased in the post-test. It is necessary to reveal that of the 24 students only four students scored the same amount of questions in both tests, the other 21 students answered more questions in the post-test. The results obtained in table 2 and in the tests of graphs 1 and 2, provide a clue that the average number of correct answers in the pre-test and in the post-test is statistically significant.

To test this conjecture, the Paired T-Test was performed as described previously. Among other values is presented the value of the statistic t = 6.0415, and the p-value = 1.837e-06. Considering that the p-value is less than 0.05 we rejected the null hypothesis that the difference between the averages is equal to zero, so there is a significant difference between the average of correct answers in both tests. That is, there was a better performance of the students after the activities proposed in the workshop, at a significance level of 5%.

Following the analysis of the second hypothesis: **(ii)** Is there a difference between the number of hits per question in both tests? Initially, the answers obtained by the participants in the pre-test and post-test questionnaire were surveyed.

Frame 1. Student responses by question: number of correct answers, errors and of students who did not know in the pre-test and post-test.

Question	Correct answers per question Pre-test	Wrong answers per question Pre-test	Do not know Pre-test	Correct answers per question Post-test	Wrong answers per question Pre-test	Do not know Post-test
1	8	16	0	20	4	0
2	22	2	0	24	0	0
3	11	13	0	17	7	0
4	13	2	9	20	3	1
5	17	3	4	20	3	1
6	7	9	8	14	9	1
7	14	1	9	24	0	0
8	19	3	2	21	2	1
9	18	3	3	20	3	1
10	18	5	1	17	7	0
11	17	6	1	21	3	0
12	19	3	2	23	1	0
13	9	3	12	22	1	1
14	16	6	2	19	5	0
15	12	2	10	19	3	2
16	7	11	6	16	8	0
17	21	1	2	22	1	1
18	20	1	3	23	1	0
19	20	2	3	23	1	0

Source: Author (2018).

The first three questions of the questionnaire were multiple choice and dealt with definitions of population, sample, and descriptive statistics.

Questions 4 to 19 were made affirmative about contents, concepts of statistics and probability, containing options for answers: Yes, No and Do not Know.

The twentieth question had as objective to know if the students had already studied the contents covered in the questionnaire. The results obtained in the twentieth question were: of the 24 subjects, seven answered that YES, 14 answered that they had already studied partially and only 3 had never studied these contents.

The first question: population definition - In the pre-test only 8 students matched the definition of population, while 20 students qualified in the post-test, and of the 16 who had failed in the pre-test only 4 continued without learning the definition and concept population. Given this picture there was learning in the activity proposal 1.

Second question: sample definition - Although the participants initially presented a misconception of the concept and definition of population, since 16 subjects believed it could only be a set of people, in the definition and concept of sample 22 participants presented the concept of sample in the pre- test, and with activity 1, all participants learned this concept, which was evidenced with the 100% correct answers in the post-test.

Third question: was asked which part of the statistic is concerned only with the description of certain characteristics of a group, without drawing conclusions about a larger group, and had as options the classifications of the branches of statistics such as descriptive statistics, inferential statistics, in the pre-test 11 students got the correct answer and 13 missed this question, but after the application of activity 1, these results were modified 17 correct answers and 7 errors. There is an existential difficulty about these branches of statistics since in Basic Education, students only study the contents coming from Descriptive Statistics.

Question 4: it was about the definition and the concept of mode, where it was affirmed false that the moda it is a measure of position that allows dividing the distribution into two equal parts of frequency. Although mode is a measure of central tendency easy to be understood by the students and be a content proposed to be taught in elementary school, 9 participants did not know how to answer this question, 2 missed this question and only 13 got the correct aswer. Activities 2 and 3 proposed a significant increase to 20 correct answers, and only 3 students missed and 1 continued without answering the item. The purpose of these activities was to balance the technique of calculating measures of central tendency with their meanings, and concepts.

The fifth question addressed one of the properties of the average asserting that it is influenced by the extreme values of the distribution. It should be emphasized here, that the average has a self-explanatory concept, what was done with the students of mathematics were their properties and the importance of associating the average with the dispersion measures.

In this item, the number of correct answers increased from 17 to 20 students, 3 students missed both tests, and from 4 decreased to 1 that did not know how to answer the item.

Question 6: There was the same affirmative changing only the measures of central tendency, average was substituted by mode and median. The results of the tests were: from 7 to 14 correct answers, of 8 only 1 student still did not know how to answer the question, and 9 students continued to answer erroneously.

In view of the above, it is necessary to emphasize that teachers of Mathematics who teach the contents of statistics in Basic Education emphasizing the procedures of calculations of measures of central tendency, leaving a little aside the concepts and their meanings, as well as their properties

The seventh question was about a very important statement about the dispersion measures, stating that in a distribution where the variance is small, we can say that the data are quite homogeneous. Of the 24 students only 14 were right, 1 was wrong and 9 did not know in the pre-test. In the post-test, after the implementation, reflection and discussion of activities 2, 3 and 4, all subjects answered this question.

Questions 8, 9, 10, 11 and 12 contained assumptions about calculating probabilities for independent events with increasing difficulty level. As only 3 subjects did not see the content covered, and probability is a subject much charged in the exams that give access to the higher level, it is believed that this fact made that in the pre-test presented bigger numbers of correct answers. The only question that got a lower percentage in the post-test than in the pre-test was question number ten. This question asked the following question: Is it more likely to have equal faces throwing two coins than equal faces in the roll of two dices? In the first moment, only one student did not know the question and the same one ended up failing in the post-test, and another student who had answered correctly in the pre-test ended up failing in the post-test. This fact may have happened because of the following conjectures: (i) the students did not pay attention to the sample spaces of the events, since there were two coins, so there are four possibilities and with two dices there are 36 possibilities; (ii) another hypothesis was that: at the time of calculating the probabilities they ended up misreading the calculations or simply responded intuitively. In the other questions, the percentage of correct answers in the post-test was greater or equal to 20 subjects.

Questions 13, 14, 15 and 16 were related to a set of data that contained six references of a student during a semester in a discipline, based on this set there were some affirmations about: breadth, average, mode and median. In question 13, half of the students did not know the concept and definition of amplitude. After the activities carried out and mainly the one that contained graphs, only 1 student did not know to answer this question in the post-test, whereas 22 of them got right giving a significant increase.

In question 14, two students did not know how to calculate the average in the pre-test, and 5 students ended up calculating the wrong average in the post-test, those participants who missed this item were in the basic operations with real numbers, this may have occurred due to the fact that they were tired of the activity, since it was during the night, or because of lack of attention, since it was not allowed to use the calculator.

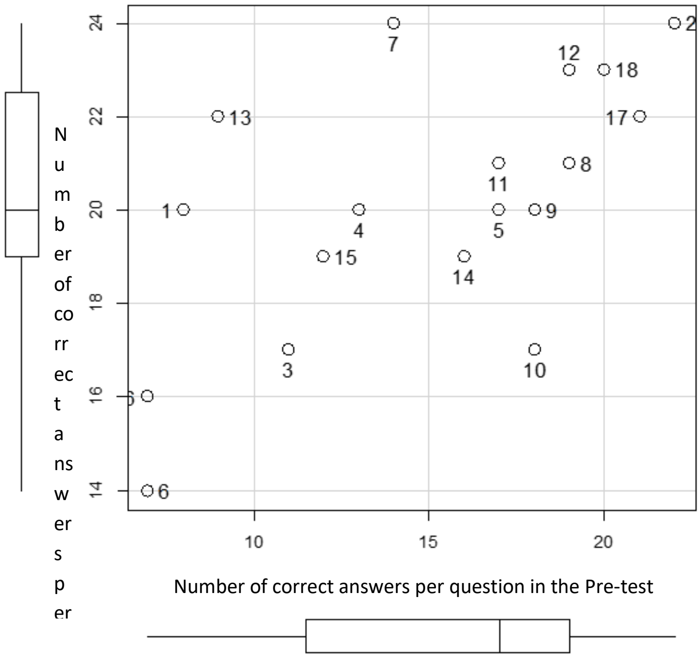
Already question 15, was about mode, 10 students did not know how to calculate the mode in the pre-test and 2 answered erroneously, only 12 were correct, that is, half of the participants did not know how to answer the question correctly. This picture decreased after the activities worked with the participants, since 19 subjects were correct, 2 remained unresponsive and 3 had answered erroneously the item in the post-test.

Question 16 asserted a value for the median score, in the pre-test 10 participants could not answer, while 2 missed, only 12 had the right answer, in the post-test only 2 remained unresponsive, 19 had the correct answer and 3 missed.

Questions 17, 18 and 19 dealt with a pie chart of a 250-person opinion poll, where relative frequency was basically considered. The title of the question: "Do braziliana give importance to environmental preservation? In the legend of this chart the following information was found: A - Yes, B - No: 30% of the people, C - I do not know: 40 people. With this information, it was stated in question 17: seventy people interviewed answered "NO". As the students already had previous knowledge of percentage in the three questions, the number of correct answers were 21, 20 and 20 respectively in the pre-test, while in the post-test these values were 22, 22 and 23 respectively, with only one student failing in each issue.

Graph 3 presents a comparison of the number of correct answers per question of the 24 subjects who participated in both tests. This graph gives a clue that the average number of correct answers in the pre-test and in the post-test is statistically significant.

Graph 3. Number of hits per question in the pre-test and post-test.



Source: Author (2018).

In graph 3, each question is identified at both study moments, in the pre-test (x-axis) and in the post-test (y-axis). Analyzing table 1 and graph 3 after the workshop approach, of the 19 questions applied in the pre-test and post-test, there was an increase in the number of answers in eighteen questions in the questionnaire.

To test this conjecture, the Paired T-Test was performed as described previously. Among other values is presented the value of the statistic t = 5.813, and the p-value = 8.283e-06. Considering that the p-value is less than 0.05 the null hypothesis that the difference between the average values is equal to zero is rejected, so there is a significant difference between the averages of right answers per question in both tests, that is , there was a better performance of the students after the methodology applied, at a significance level of 5%.

The analysis of the hypotheses tested and the confirmation of the hypothesis of the improvement of the students' performance after the workshop implementation, offer us indications to affirm that the methodology used for teaching probability and statistics through the workshop, resulted in effective learning because the activities proposed and implemented provided an improvement in post-test results.

It is argued that this "teaching practice" for Teaching Probability and Statistics should be introduced from the discussion of contextualized examples in the areas of interest of the students, working whenever possible with the methodology of problem solving, and with data from real situations.

Final Considerations

In Basic Education, those who teach the contents of probability and statistics are the teachers of Mathematics, who emphasize the mathematical aspect, valuing the logic and the algorithms, to the detriment of using Statistics as a data analysis tool, promoting a mediation between the techniques and the meanings of the concepts.

The teaching of probability and statistics at the various levels should be rethought in order to achieve a balance between techniques and meanings, and also of the deterministic/random, in which the world is seen most of the time as if it were deterministic, when in reality it is fundamentally random.

In higher education, teachers also exaggerate in the emphasis of their mathematical aspect, where they should carry out more empirical studies of random phenomena, sample variability and their probabilistic interpretation. Mathematicians and Mathematical Educators must "give up" demonstrations, proofs of Theorems, in order to teach probability and statistical classes in a more "dynamic" way in order to learn more meaningfully.

According to Magalhães (2015), to be a full citizen is necessary to know basic concepts of Statistics. Understanding tables and simple graphs, perceiving limitations in the numerical information available on the subjects, understanding the existence of randomness in events are some of those skills that, unfortunately, are not yet part of the daily routine of the majority of the population.

Moreover, on learning, it is argued that it becomes more meaningful when the content taught is appropriate to the student's knowledge structure and constructs meanings in the dialectical relation between the content learned with his previous knowledge. Otherwise, learning becomes mechanical as Ausubel points out (1982).

This way, it is corroborated with Freire (2014) when he says that the school must respect the knowledge of all learners that come to it, these knowledges socially constituted in their community practice, also stressing the need to take advantage of this experience that students have in the educational process.

In this perspective, it is suggested that teachers should work with projects or from contextualized examples, valuing group work with the use of software, such as Excel, R, Geogebra, with their referred statistical packages, which allow the effective participation of the student in the process of teaching learning.

References

AUSUBEL, David Paul. **A aprendizagem significativa:** a teoria de David Ausubel. São Paulo: Moraes, 1982.

ALMOULOUD, Saddo Ag. **Fundamentos da didática da matemática.** Curitiba: Editora da UFPR, 2007.

BATANERO, Carmen. **Didáctica de la estadística.** Universidade de Granada: Espanha. Disponível em: <http://www.ugrs.es/~batanero>. Acesso em: 20 mar. 2011.

BATANERO, Carmen. **Estadística y probabilidad en la enseñanza no universitária**: oportunidades y desafíos actuales. São Paulo: Uniban,: 2011. (III SIEMAT).

BATANERO, Carmen; ESTEPA, Antonio; GODINO, Juan D. Análisis exploratorio de datos: sus posibilidades en la enseñanza secundaria. **Suma**, n. 9, p. 25-31, 1991. Disponível em: http://www.ugr.es/~batanero. Acesso em: 05 jun. 2013.

BRASIL. Ministério da Educação (MEC). Lei n.º 9394 de 20 de dezembro de 1996. **Estabelece as Diretrizes e Bases da Educação Nacional.** Diário Oficial da República Federativa do Brasil, Brasília, c 2048, p. 27833-27841, 23 dez. 1996.

BRASIL. Ministério da Educação (MEC). Secretaria de Educação Básica (SEB), Departamento de Políticas de Ensino Médio. **Orientações Curriculares do Ensino Médio.** Brasília: MEC/SEB, 2006.

BRASIL. Ministério da Educação (MEC). Secretaria de Educação Básica (SEB). **Parâmetros Currículares Nacionais (5ª a 8ª séries)**: matemática. Brasília: MEC/SEF, 1997.

BRASIL. Ministério da Educação (MEC). Secretaria de Educação Média e Tecnológica (Semtec). **Parâmetros Curriculares Nacionais para o Ensino Médio.** Brasília: MEC/Semtec, 1999.

CAMPOS, Celso Ribeiro; WODEWOTZKI, Maria Lúcia Lorenzzetti; JACOBINI, Otávio Roberto. **Educação estatística**: teoria e prática em ambientes de modelagem matemática. Belo Horizonte: Autêntica, 2011.

COUTINHO, Cileda de Queiroz e Silva. **Introdução ao conceito de probabilidade por uma visão frequentista**: estudo epistemológico e didático. Programa de Pós Graduação da PUC-SP (SP), 1994.

COUTINHO, Cileda de Queiroz e Silva; SOUZA, Fabiano dos Santos, Potencialidades do Uso do Geogebra e do R na Construção e Interpretação de Gráficos Estatísticos. In: SILVA, Mauren Porciúncula Moreira da (Org.). **Educação estatística**: ações e estratégias pedagógicas no ensino básico e superior. Suzi Samá; Curitiba: CRV, 2015.

COUTINHO, Cileda de Queiroz e Silva; SOUZA, Fabiano dos Santos, Aprendizagem Da Estatística e o Uso De Ambientes Computacionais: Uma Análise Didática De Programas Para Construção De Gráficos Estatísticos. In: CONGRESO IBEROAMERICANO DE EDUCACIÓN MATEMÁTICA, 7., Motivideo, Uruguay, 2013. **Actas del…** Montivideo: [s.n.], 2013. v.1, p.6240 - 6247.

DALGAARD, Peter. **Introductory satatistics with R**. New York: Springer-Velarg, 2002.

FOX, John. The R Commander: A Basic-Statistics Graphical User Interface to R. **Journal of Statistical Software**, v.14, n.9, p.1-42, 2004.

FOX, John. **Getting started with R Commander.** 2006. Disponível em: <http://socserv.mcmaster.ca/jfox/Misc/Rcmdr/Getting-Started-with-the-Rcmdr.pdf.>. Acesso em: 10 abr. 2012.

FOX, John. **Package Rcmdr.** 2011. Disponível em: <http://cran.r-project.org/web/packages/Rcmdr/Rcmdr.pdf>. Acesso em: 10 abr. 2012.

FREIRE, Paulo. **Pedagogia da autonomia**: saberes necessários à prática educativa**.** 48. ed. Rio de Janeiro: Paz e Terra, 2014.

GAFIELD, Joan. **The** challenge of developin statistical reasoning. **Journal of Statistis Education,** v. 10, n. 3, 2002. Disponível em: <www.amstat.org.publications/jse/v10n3/chance.html>. Acesso em: 10 jan. 2012.

GAL, Iddo. Adults' statistical literacy: meanings, components, responsibilities. **International Statistical Review**, v.70, n.1, p. 1-25, 2002.

LOPES, Celi Espasandin; CARVALHO, Carolina. **Literacia estatística na educação básica, leituras e escritas na educação matemática.** Belo Horizonte: Autêntica, p. 77-92, 2005.

LOPES, Celi Espasandin; COUTINHO, Cileda de Queiroz e Silva; ALMOULOUD, Saddo Ag. **Estudos e reflexões em educação estatística.** Campinas, SP: Mercado de Letras, 2010.

MAGALHÃES, Marcos Nascimento. Desafios do Ensino de Estatística na Licenciatura em Matemática. In: SILVA, Mauren Porciúncula Moreira da (Org.). **Educação estatística**: ações e estratégias pedagógicas no ensino básico e superior. Suzi Samá; Curitiba: CRV, 2015.

PETERNELLI, L. A.; MELLO, M. P. **Conhecendo o R**: uma visão estatística. Viçosa, MG: Ed. UFV, 2011. Disponível em: http://www.R-project.org. Acesso em: 10 jun. 2012.

SÁ, D. L.; SILVA; M. P. M.; SAMÁ, S. Análise do letramento estatístico de estudantes concluintes do ensino médio. In: SILVA, M. P. M. (Org.). **Educação estatística**: ações e estratégias pedagógicas no ensino básico e superior. Curitiba: Suzi Samá; CRV, 2015.

SANTOS, W. D.; SANTOS JUNIOR, J.; VELASQUE, L. S. O Desenvolvimento do Letramento Estatístico pelos Livros Didáticos e a Base Nacional Comum Curricular. **REnCiMa**, v. 9, n. 2, p. 210-229, 2018. Disponível em: <http://revistapos.cruzeirodosul.edu.br/index.php/rencima/issue/view/58/showToc>. Acesso em: 10 maio. 2018.

SOUZA, Fabiano dos Santos. A Importância das Atividades para o Ensino de Estatística na Educação Básica com o Uso do Software R e seu Pacote de Dados Rcmdr no III SIPEMAT. **Perspectivas da Educação Matemática**, Campo Grande, MS, v.6, n. 11, p. 85-92, jan/jun. 2013.

WALLE, John A. Van De. **Matemática no ensino fundamental.** Artmed Editora S.A., São Paulo, 2009.