Selection and Application of graphical and numerical statistical tools by prospective primary school teachers

José António Fernandes[a]
Adelaide Freitas[b]

[a] Universidade do Minho, Instituto de Educação, Braga, Portugal
[b] Universidade de Aveiro, Departamento de Matemática, Aveiro, Portugal

ABSTRACT

Teaching statistics in the early years requires that teachers at this school level develop skills to analyze small collections of data. Given a collection of quantitative data (12 observations), in this paper we looked at how students and prospective primary school teachers select and make appropriate graphs and identify and determine statistical measures suitable for summarizing the data, including the interpretation of the third quartile. The study involved 50 students who were attending the 2nd year of the Basic Education Bachelor's program at a university in northern Portugal. The collected data correspond to the answers given by the students in a formal examination in a Probability and Statistics course. An analysis of the answers showed that the students had difficulties in both the selection and application of statistical methods, which were more pronounced when they had to identify the appropriate graphs to represent the data and to determine the quartiles and to interpret the third quartile, and less pronounced in the case of determining other statistical measures.

Keywords: statistical graphs; statistical measures; prospective primary school teachers.

Escolha e aplicação de ferramentas estatísticas gráficas e numéricas por futuros professores dos primeiros anos

RESUMO

O ensino da Estatística nos primeiros anos de escolaridade requer que os professores desse nível escolar desenvolvam competências de análise de pequenas coleções de dados. Neste artigo, estudamos a habilidade de, perante uma coleção de 12 observações de tipo quantitativo, estudantes, futuros professores dos primeiros anos escolares, selecionarem e construírem gráficos adequados e identificarem e determinarem medidas estatísticas convenientes à sumarização dos dados, incluindo a interpretação do 3.º quartil. No estudo participaram 50 estudantes que se encontravam a frequentar o 2.º ano da Licenciatura em Educação Básica numa universidade do norte de Portugal. Os dados recolhidos correspondem às respostas dadas pelos estudantes numa prova de avaliação formal da disciplina de Probabilidades e Estatística. Em termos de resultados, uma análise das respostas

---

Corresponding author: Adelaide Freitas. Email: adelaide@ua.pt

permitiu verificar que os estudantes revelam dificuldades tanto ao nível da escolha como da aplicação dos métodos estatísticos, mais acentuadas quando se tratava de identificar os gráficos adequados para representar os dados, e de determinar os quartis e interpretar o 3.º quartil, e menos acentuadas no caso da determinação de outras medidas estatísticas.

Palavras-chave: gráficos estatísticos; medidas estatísticas; futuros professores dos primeiros anos.

INTRODUCTION

We have been witnessing an increasing use of Statistics in different areas of society both at the personal and professional level, which explains its more in-depth teaching at schools. In Portugal, the content of Probability and Statistics courses which are part of the Data Organization and Treatment subject, are included in the Mathematics program in all school grades, both Elementary Education (Ministério da Educação e Ciência, 2013) and in Secondary Education (Ministério da Educação e Ciência, 2014).

Therefore, in line with school programs, a more in-depth teaching of Probability and Statistics at school requires that teachers also get education in mathematics that enables them to teach these subjects according to the needs of their students and the education system. It is particularly important for teachers in the first school years to learn the basics of Descriptive Statistics for their teaching and the basic skills to design or prepare simple research activities involving standard, graphical and numerical statistical tools to summarize qualitative and quantitative data (discrete or continuous data).

Around one decade ago, Batanero (2009) warned that the teaching of Probability and Statistics was usually not part of the initial education for all teachers. For primary school teachers, this issue is possibly even more serious because only recently, more precisely in the beginning of the century (Ministério da Educação, 2007), these contents were included in school programs and, as a result, in the initial education of these teachers.

Therefore, in the context of the initial education of prospective primary school teachers, this study looks at the statistics knowledge and skills of these prospective teachers in Descriptive Statistics. More specifically, it assesses the suitability of the exploratory statistical methods selected by the prospective teacher for a quantitative data analysis involving graphs and statistical measures, as well as the appropriate application of these methods.

The right selection of exploratory methods for preliminary data analysis is of utmost importance because, on the one hand, it is the first step to explore more open tasks, as is the case with investigative projects in Statistics and, on the other hand, it underlies the understanding of strategies used to solve even simple exercises and problems, which can lead to a deeper understanding of concepts or of the relationships between concepts. In a previous study, Fernandes, Batanero and Gea (2019) looked at the suitability of the statistical methods chosen by prospective primary school teachers to analyse a data set related to a qualitative (nominal) variable and concluded that they had more difficulties in selecting the statistical methods.

In the present study, which is a follow-up study aiming to complement the previously mentioned study, we analyse the knowledge and skills of prospective primary school teachers in selecting the statistical methods (graphs and measures) to analyse a data set related to a continuous quantitative variable. The type of statistical variable, i.e., either a qualitative variable (nominal or ordinal) or a quantitative variable (discrete or continuous) is extremely relevant in a statistical study because it dictates the statistical methods that can be applied.

In the next sections, we will present the theoretical framework and the background, reporting and discussing some studies related to the issue investigated here; the research method, where we explain the type of study, the participants and the data collection and analysis methods; the presentation of results according to the graphical contents and statistical measures; and, finally, we summarize and discuss the study conclusions.

THEORETICAL FRAMEWORK AND BACKGROUND

Research on the knowledge teachers should have to successfully teach Mathematics has resulted in several frameworks. For the purpose of our study, from the theoretical models most frequently mentioned in the literature (Gonzalez & Eudave, 2018), we chose the Mathematical Knowledge for Teaching (MKT) framework, which is widely quoted and originally proposed by Ball (2000). This framework characterizes the mathematical knowledge teachers should have to carry out their teaching activities and enhance the learning of mathematics, classifying the mathematical knowledge of teachers into two categories: Content Knowledge and Pedagogical Content Knowledge and is based on Shulman's framework (1986) which classifies teacher's knowledge into three categories: Content knowledge, pedagogical content knowledge and curriculum knowledge. Content Knowledge includes mathematical knowledge and the transformation of this knowledge to make it accessible to students, while in Pedagogical Content Knowledge relationships are established between the mathematical knowledge and several factors such as students, teaching and the curriculum.

Since in this study we are investigating the mathematical knowledge of prospective teachers, we have limited the investigation to Content Knowledge, which consists of the three following subcategories (Hill, Ball & Schilling, 2008):

— Common Content Knowledge (CCK), which consists of the knowledge adults use to solve mathematical problems;

— Specialized Content Knowledge (SCK), which relates to the special knowledge teachers need to plan and implement their teaching sequences;

— Knowledge at the mathematical horizon, which refers to the knowledge of the sequence of mathematical topics in the curriculum, with teachers relating what their students are learning with what they will learn in the future.

Of these three subcategories we will discuss Common Content Knowledge and Knowledge at the Mathematical Horizon, because in this study we will not look at issues related to Specialized Content Knowledge.

Since we are dealing here with prospective primary school teachers, it is extremely important that during their education they acquire skills that will enable them to decide which methods to use in (simple) statistical data analyses, e.g., in investigative projects. According to Batanero, Díaz, Contreras and Arteaga (2011), open tasks are widely used today because they enhance students' motivation and reveal the context and its realistic nature. Furthermore, these tasks enable students to acquire strategic knowledge that is added to the technical knowledge applied in conventional tasks.

Additionally, statistics studies, such as investigative projects, involving more open tasks, can have several steps, as advocated by Wild and Pfannkuch (1999) when they proposed the PPDAC framework, which they call investigative cycle, divided into five steps: problem (P); plan (P); data (D); analysis (A) and conclusions (C). We will focus particularly on the second and fourth steps, related to the planning and implementation of tasks designed to solve a problem or give an answer to the investigation questions defined in the previous step. Decisions are made in the second step regarding which statistical analysis/methods or statistical measures are appropriate for the study, while in the fourth step the previously selected methods/procedures are applied.

The type of statistical variable being studied determines, to a great extent, the statistical methods that can be used in the data analysis regarding frequency, statistical measures or the statistical graphs, thus impacting the planning and analysis steps. Regarding frequency, it is essential for their teaching that prospective teachers are able to relate the concept to the procedure. While absolute and relative frequencies can be determined for any type of variable, accumulated absolute and relative frequencies cannot be determined in nominal statistical variables because these frequencies require the establishment of an order relationship. Thus, accumulated frequencies can only be determined for an ordinal qualitative variable or quantitative variable. However, we know that if some statistical methods cannot be used in certain situations, because they require numerical data, students incorrectly use frequencies to overcome the issue of the values of qualitative statistical variables (Fernandes & Barros, 2005; Fernandes, Carvalho & Ribeiro, 2007; Fernandes, Carvalho & Correia, 2011).

Additionally, the choice of graph used to represent data is very important because it may impact the success of the next PPDAC steps. Regarding the media, it is expected that any individual with mathematical education will be able to interpret the various types of graphs; but making graphs or deciding which one is more adequate considering a given data set (qualitative or quantitative, discrete or continuous, a few or a lot of data) requires a deeper knowledge and is not always easy to realize. For example, Morais and Fernandes (2011) and Fernandes, Morais and Lacaz (2011) found that 9th grade students had difficulties when choosing the right graphs to represent data. In the three items proposed to students, the best performance was achieved in the one that involved the graphical representation of a discrete quantitative variable (age), followed by the one that

asked a graphical representation to compare the values of the previous variable according to the variables male and female gender and, finally, a very poor performance in the item that required the graphical representation of a continuous quantitative variable.

Students clearly preferred to make simple bar graphs. When this graph was adequate to represent the variable, the proportion of correct or partially correct answers was high; however, when this graph was not appropriate to represent the variable, there was a significant reduction in the number of correct and partially correct answers. This reduction was particularly pronounced when the item involved making a histogram, when almost half of the students did not answer and except for one student who gave a correct answer and another student who gave a partially correct answer, all other students made bar graphs, pie charts, line graphs and Cartesian graphs. According to Fernandes et al. (2011),

> the prevalence of simple bar graphs could result from the fact that this type of graph was the one most practiced by these students during the classes and also because it is easier to make that the other types of graph as, for example, the grouped or stacked bar graphs or histograms. (p. 11)

In addition to choosing a graph that was not adequate to represent the data, the graphs made by the students presented many flaws described in the literature (e.g. Espinel, González, Bruno & Pinto, 2009; Ruiz, Arteaga & Batanero, 2009), like the absence of a title and axis labels, inadequate scales and careless graph making. According to Friel, Curcio and Bright (2001), these aspects, which they call graph structure, are important because they provide information on the type of measurement that is being used and the data that are being measured. We should stress here that changing the scales can distort the shape of a bar graph and of a histogram, which in turn can result in wrong interpretations by individuals with less mathematical education. Also in the representation of data in a box plot, a less conventional graph that is widely used in exploratory studies in several areas of knowledge, there are well-recognized difficulties, particularly regarding its interpretation (Carvalho, Fernandes & Freitas, 2019).

Similarly, the choice of statistical measures to summarize data depends on the type of statistical variable being studied. Except for mode, which is a kind of statistics that can be applied to any type of variable, none of the other measures of location (median, quartile and mean) and dispersion (amplitude, variance and standard deviation) can be applied to every type of statistical variable. However, as shown in the literature (e.g., Boaventura & Fernandes 2004; Fernandes et al., 2007), when it is not possible to determine the median or the mean, because it is a nominal qualitative variable or a non-quantitative variable, respectively, students use frequencies to obtain numerical values that enable them to apply formulas to calculate these statistics.

In the study by Fernandes et al. (2007), when the teacher asked 7th grade students to calculate the mean, the mode and the median of the variable color of the eyes of the students

in the class, she found that students were confused and found it difficult to calculate the median and the mean, with one student suggesting: "We could transfer this to a number: we could have blue 1, green 2 and so on" (page 49). Then, when the teacher reminded them that to calculate the median they should order data, another student added: "Only if we did it by alphabetical order." In these excerpts we can see how students insist to wrongly overcome the non-existence of numerical values which they know are required to determine statistics.

In the study by Boaventura and Fernandes (2004), given the distribution of a nominal qualitative variable (the students' favorite beverage), 12th grade students were asked to determine, if possible, the mode, median and mean, and they found that many students had difficulties solving the task, particularly the median and the mean. Regarding the median and the mean, students used the frequencies of the variable values to determine these statistics and, for the mode, many students indicated the frequency instead of the variable value. Fernandes and Barros (2005) also found similar results, this time with prospective primary school teachers.

In a more recent study involving the same nominal qualitative variable (the students' favorite beverage), Fernandes, Batanero and Gea (2019) found that around half of the students, who were prospective primary school teachers, determined accumulated absolute and relative frequencies, which is not right, around three out of four students made a bar or pie chart, which is adequate, almost all of them (82%) indicated the mode value, which is equally right and many (78%) determined the value of the mean, the median, the standard deviation, quartiles, amplitude or variance, which are clearly not the right statistical measures for the study variable. Students found it more difficult to choose the statistical methods than to apply them, particularly in the case of frequencies and statistical measures.

In our study we looked at the extent to which prospective primary school teachers select and apply statistical analysis methods that are adequate to a given data distribution that involves a continuous quantitative variable regarding graphs and statistical measures.

METHOD

In this mainly descriptive study we investigated the choices of statistical methods regarding graphs and statistical measures made by students, prospective primary school teachers and the later application of these methods in the analysis of a statistics task.

The study included 50 students (E_i, with $i = 1, 2, ..., 50$) from a university in northern Portugal who were attending a Probability and Statistics course in the 2nd year of a Basic Education Bachelor's program. When entering the university, these students had very different levels of mathematics education, which also explains the diverse perceptions of their difficulties in the Mathematics courses they had attended so far at the university.

Data collection was made through a formal exam that students took after the end of the Probability and Statistics course. This curriculum unit is focused on content knowledge and it does not explore teaching aspects. Of the several tasks included in the exam, we looked at just one (Figure 1), focused on a set of 12 observations of a continuous quantitative variable. To solve this task, students could use a calculator or a computer calculation sheet.

The chart below shows the grades, in a 0 to 100 scale, obtained on a Mathematics exam by twelve 5th grade students.

55	60	84	45	48	64
90	65	55	58	70	50

a) Which type of graphs properly represents the data? Choose one of these graphs and represent the data.

b) If possible, identify the mode and determine the mean, quartile and standard deviation values. In general, what is the meaning of the third quartile?

Figure 1. Task statement proposed to the students

In terms of data analysis and organization, we started by listing the chosen statistical graphs and then proceeded to identify failures in the graphs made by the students regarding question a); then we assessed the determination of the six statistical measures required (mode, mean, 1st quartile, median, 3rd quartile and standard deviation) according to the type of answer (correct and incorrect) and the interpretation of the third quartile in question b). If applicable, we determined the students' frequencies according to the different types of answer (correct or incorrect) and not answered, information is summarized in tables and examples of problem-solving by the students are presented to clarify their reasoning.

RESULTS AND ANALYSIS

In this section, we present the results obtained according to the content of the task questions: a) graphs; and b) statistical measures.

Graphs

In question a) students are asked which graphs are adequate to represent the provided data and, then, they are asked to make one of these graphs to represent the data. Taking into account the continuous nature of the data and the lessons learned by the students, the most adequate graphs would be histograms or box plots and given the reduced number of observations (12), a stem-and-leaf plot. In the problem context, students could also group data in classes/categories and, thus, select the graphs most suitable to the qualitative variables, namely a bar or pie chart. Although the latter two charts were marked, students

did not point to any type of previous data categorization. In addition to these graphs, students indicated and/or made other inadequate graphs, as shown in Table 1.

Table 1
Frequencies of the types of graphs indicated and made by the students.

Type of graph	Number of students (%)	
	Graphs indicated	Graphs made
Histogram	35(70)	24(48)
Stem-and-leaf plot	3(6)	2(4)
Bar chart	36(72)	12(24)
Line graph	7(14)	3(6)
Pie chart	5(10)	—
Scatter diagram	5(10)	—
Frequency table	2(4)	1(2)
Grouped bar chart	1(2)	—
Not answered	1(2)	8(16)

By analysing the choice of graphs made by the students, we concluded that, on average, each student indicated two different types of graphs to represent the data. However, more inadequate (56) than adequate graphs (38) were mentioned and the box plot was not mentioned as an adequate graph to represent the data.

Regarding the graphs that were considered adequate we highlight the histogram, which was mentioned by over half of the students (70%), followed by the stem-and-leaf plot, although at a much lower frequency (6%).

Regarding the graphs that were considered inadequate, the bar chart was mentioned by over half of the students (72%), being the most frequently chosen, followed by the line graph (14%), the pie chart and the scatter diagram (both with 10%), the frequency table (4%) and the grouped bar chart (2%).

Therefore, in general students performed poorly when asked to indicate which graphs were adequate to represent the data. In addition to the bar chart, the line graph and the pie chart, which are not adequate because data are almost all distinct, it is even more difficult to understand why they mentioned the scatter diagram, which applies to two-dimensional distributions, thus involving two statistical variables, and the frequency tables.

In the graphs made by students to represent the data, we highlight the use of the histogram by around half of the students (48%), which is a graph that is adequate to the given situation. However, many students used many different classes, 10 classes (8%), 6 classes (14%), 5 classes (16%), 4 classes (2%) and 3 classes (8%), when, for a set of twelve pieces of data, it would be adequate to use 3 or 4 classes (using the Table by Truman L. Kelly or the rule of the sample size square root). In three of these graphs separate bars

were considered. Figure 2 shows an example of a histogram with 10 classes. In the case of stem-and-leaf plots, the students did not specify either the stem or the leaf units.
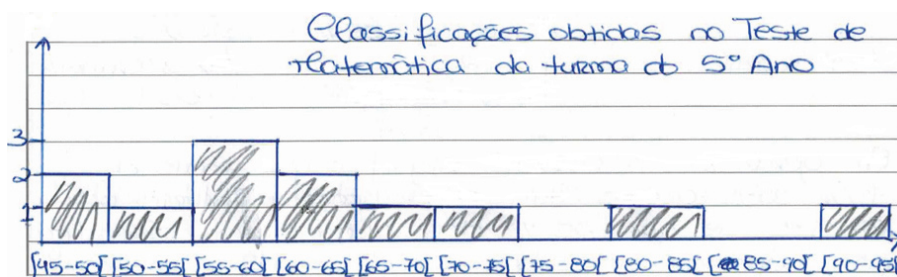

Figure 2. Graph made by student E41 for question a)

Although it was not adequate, a reasonable percentage of students (24%) made a bar chart to represent the data, as shown in Figure 3. We realize that student E4 did not pay attention to the scale of observations and considered only 11 pieces of data, forgetting datum number 58.
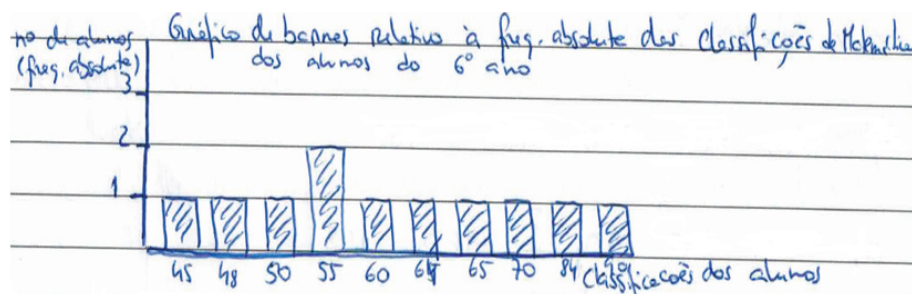

Figure 3. Graph made by student E4 for question a)

The fact that there was almost no data repetition, as is the case in this situation, results in the fact that the distribution given by the bar chart is not very different from the given gross data, with the graphical representation achieving a minimal data reduction. Some of these students (12%) consider on the horizontal axis the data labels (1 to 12) and on the vertical axis their classifications, as shown in Figure 4.

In this representation there is no data reduction whatsoever, all that was done was to represent the data provided in the task statement in a bar chart. This kind of difficulty was also observed by Fernandes and Correia (2009) in some practicing primary school teachers.
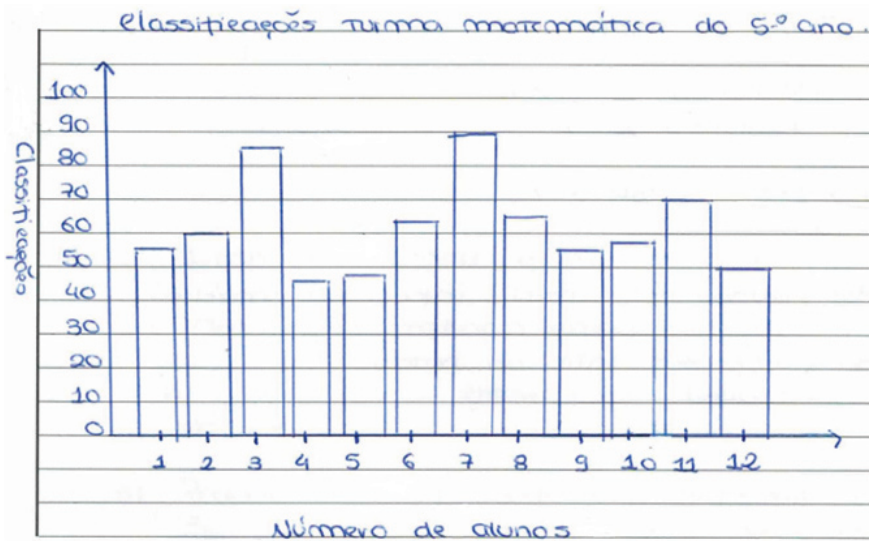
Figure 4. Graph made by student E4 for question a)

Additionally, we found that students often either did not gave a title to the graph or named the respective coordinated axes. Of all graphs, 32% did not have a title, 26% did not name the horizontal axis and 24% did not name the vertical axis.

Statistical measures

In general, in question b) students considered that it would be possible to determine the mode, the mean, the quartiles and the standard deviation. Table 2 shows the frequencies of students according to the type of response (correct and incorrect) and of questions not answered.

Table 2
Frequencies of types of answers in the different statistics

Type of answer	Number of students (%)			
	Mode	Mean	Quartiles	Standard deviation
Correct	50(100)	43(86)	14(28)	26(52)
Incorrect	—	6(12)	29(58)	16(32)
Not answered	—	1(2)	7(14)	8(16)

Except for one student, who mentioned the modal class, all other students mentioned the mode value of the simple data. Similarly, almost all students (86%) properly determined the mean value, out of whom two (4%) used the class marks instead of the values of the statistical variable; and some answers were wrong because of calculation errors (12%).

Determining quartiles was a difficult task for the students, with few of them (28%) correctly determining the three quartiles. The majority (58%) determined at least one quartile incorrectly and some not even answered this question (14%). Regarding these answers, we would like to highlight the number of students who did not determine correctly any of the quartiles (16%) and those who determined only the second quartile correctly (38%). In the latter case, students started by determining the second quartile (or the median) correctly, corresponding to the average of the two central values and then, they eliminated these values when determining the first and third quartile, as shown in Figure 5. Thus, these students determine the first and third quartiles as being the central value of the odd data set, when they should determine the arithmetic mean of the two central values of the even data set.
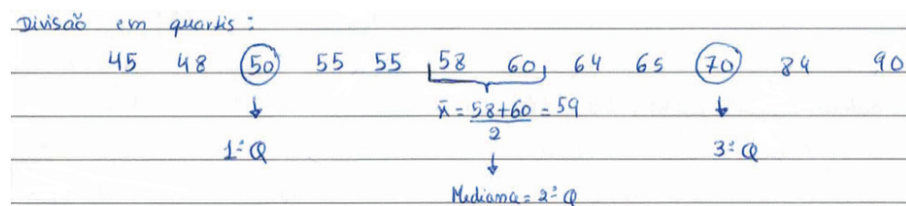

Figure 5. Quartile determination by student E7 for question b)

Around half of the students (52%) determined the standard deviation correctly, with two of them (4%) using class marks, while the rest made mistakes in their determination (32%) or did not answer (16%). Regarding incorrect answers, 12% were due to calculation errors and mistakes resulting from the use of an inadequate formula to determine the standard deviation, such as mistaking standard deviation for variance (4%), not dividing by the sample size (4%), determining squares of the differences between absolute frequencies and variable values (2%), dividing by the value of the mean and not by the sample size (2%) and considering 11 pieces of data instead of 12 (2%).

Finally, in the interpretation of the third quartile, less than half of the students (38%) used the division of data set ordered in four parts, each containing around 25% of data, to determine its meaning. Some students (22%) gave the correct answer simultaneously referring to around 75% of data less than or equal to and 25% of data greater than or equal to the third quartile, as shown in Figure 5.

O 3º Q significa que cerca de 75% dos dados são inferiores ou
iguais ao valor do 3ºQ e cerca de 25% dos dados são superiores
ou iguais ao valor do 3ºQ.
Neste caso, 75% dos dados são inferiores ou iguais a 67,5% e
25% dos dados são superiores ou iguais a 67,5%.

Figure 5. Quartile determination by student E7 for question b)

Less students (16%) only said that around 75% of data are less than or equal to the third quartile. As in the previous case, this was also considered correct, because the value of the third quartile is not a value found in the collected data, and therefore, under these circumstances, the percentage of values greater than or equal to the value of the third quartile is the complement to 100% of the 75% (for further details on the issue of the correct interpretation of quartiles linked to the existence of observations equal to the value of the quartile in the collected data, see, for example, Freitas, Cruz & Silva (2017) for the specific case of the second quartile).

In addition to the high percentage of unanswered questions (34%), incorrect assignment of meanings was mainly due to inaccurate statements made by the students (14%), by confusing the third quartile with the first quartile or the median (8%), by confusing the third quartile with the mean or the mode (4%) and by confusing the third quartile with an extreme value (2%). These results show that, just like in the determination of quartile values, students also have many difficulties to interpret the third quartile.

CONCLUSIONS

In the present study, we investigated students and prospective primary school teachers regarding two fundamental steps in a statistics assignment: the selection of methods (graphical and numerical) used in a statistical analysis, followed by the application of these methods.

Considering the overall performance of the students in question a), we found that 76% indicated an adequate graph, but out of those, only 22% indicated only one adequate graph, meaning that most of them presented, in addition to the adequate graph, at least one graph that was not adequate for the given situation; 52% made an adequate graph which could, however, have flaws in its structure. In question b), considering the determination of statistical measures and the interpretation of the third quartile, on average, 60.8% gave a correct answer.

These results show that the question involving the choice of graphs that would properly represent data and making one of these graphs caused the greatest difficulties for the students, and the choice of inadequate graphs and the omission of structural aspects in the graphs, like the title, the scale and axis labels (Friel at al., 2001) were found in

studies with 9th grade students (Fernandes et al., 2011, Morais & Fernandes, 2011) and prospective primary school teachers (Fernandes et al., 2019). Many students chose to make bar charts in which, due to the virtual absence of repetition, the graphical representation cannot be distinguished from the given data. According to Fernandes and Correia (2009), graphical representation of each individual piece of data, as in this case, is the most basic level of data representation. What we have here is a graphical representation that does not result in any data reduction.

The question involving the determination of statistical measures and the interpretation of the third quartile had the highest percentage of correct answers. Generally, in the determination of statistical measures, students performed better than in the choice and interpretation of these statistics (Boaventura & Fernandes, 2004; Fernandes et al., 2007) which, among other possible reasons, is due to a teaching more directed to techniques and procedures (Fernandes et al., 2011). Compared to other statistics, the greatest difficulty in the determination of quartiles generalizes the difficulties students have to determine the median, as shown in several studies (e.g., Carvalho, Fernandes & Freitas, 2019; Fernandes & Barros, 2005).

Comparing the results of the present study, whose data refer to a quantitative variable, with those obtained in the study by Fernandes et al. (2019), whose data refer to a nominal qualitative variable, studies with the participation of prospective primary school teachers, we find that students have not chosen very different statistical methods. However, the fact that data relate to a nominal qualitative variable or to a quantitative variable means the use of frequently different statistical methods, which apparently many students did not realize. Therefore, these students determine accumulated frequencies and several statistical measures in nominal qualitative variables, use frequencies instead of variable values for statistics and represent the data using graphs that do not significantly reduce the data, thus enabling the recognition of patterns and trends.

Therefore, regarding the identification and application of statistical methods, it seems that students think that they can apply these methods to any type of statistical variable, overgeneralizing the use of these methods.

Thus, this study concludes that prospective primary school teachers should explore, during their early education, tasks such as those proposed in this study, particularly those involving decision-making about the statistical methods to be used and the interpretation of statistical results. The practice of these tasks by prospective teachers is extremely important if we want them to explore these tasks with their future students, because these educational experiences will influence their future pedagogical practices (Almeida & Fernandes, 2010).

Considering the difficulties students had, the solution is to face instead of ignoring them, because the choice of the statistical methods and the interpretation of results are activities that are closely related to the exploration of more open tasks, application and investigative projects (Batanero et al. 2011). Not letting students decide which statistical methods they should use, because they are already specified in the task statements and

not encouraging interpretation show that Statistics teaching and learning is focused on formulas and calculations, which is a limited approach considering the current recommendations for the teaching of Statistics (Fernandes et al., 2007; MacGillivray & Pereira-Mendoza, 2011).

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTION STATEMENTS

JAF was more involved with planning the study and organizing the paper, while AF had a major role regarding the statistical aspects. Both authors equally participated in writing the paper, namely the theoretical framework and background, method, data analysis and conclusions.

DATA AVAILABILITY STATEMENT

The authors agree to make available the data that support the results of this study by means of reasonable request from a reader and the authors will determine at their own discretion whether the request is reasonable or not.

REFERENCES

Almeida, M. G., & Fernandes, J. A. (2010). A comunicação promovida por futuros professores na aula de Matemática. *Zetetiké*, *18*(34), 109-154.

Ball, D. L. (2000). Bridging Practices. Intertwining content and pedagogy in teaching and learning to teach. *Journal of Teacher Education*, *51*(3), 241–247.

Batanero, C. (2009). Retos para la formación estadística de los profesores. In J. A. Fernandes, M. H. Martinho, F. Viseu & P. F. Correia (Eds.), *Actas do II Encontro de Probabilidades e Estatística na Escola* (pp. 52-71). Braga: Centro de Investigação em Educação da Universidade do Minho.

Batanero, C., Díaz, C., Contreras, J. M., & Arteaga, P. (2011). Enseñanza de la Estadística a través de proyectos. In C. Batanero & C. Díaz (Eds.), *Estadística con proyectos* (pp. 9-46). Granada: Universidad de Granada.

Boaventura, M. G., & Fernandes, J. A. (2004). Dificuldades de alunos do 12.º ano nas medidas de tendência central: O contributo dos manuais escolares. In J. A. Fernandes, M. V. Sousa & S. A. Ribeiro (Eds.), *Actas do I Encontro de Probabilidades e Estatística na Escola* (pp. 103-126). Braga: Centro de Investigação em Educação da Universidade do Minho.

Carvalho, M. J., Fernandes, J. A., & Freitas, A. (2019). Dificuldades de alunos do 8.º ano na construção de diagramas de extremos e quartis. In J. M. Contreras, M. M. Gea, M. M. López-Martín, & E. Molina- Portillo (Eds.), *Actas del Tercer Congreso Internacional Virtual de Educación Estadística.* Granada: Universidade de Granada.

Espinel, M. C., González, M. T., Bruno, A., & Pinto, J. (2009). Las gráficas estadísticas. In L. Serrano (Ed.), *Tendencias actuales de la investigación en educación estocástica* (pp.57-74). Málaga: Gráficas San Pancracio.

Fernandes, J. A., & Barros, P. M. (2005). Dificuldades de futuros professores do 1.º e 2.º ciclos em estocástica. In *Actas do V Congresso Ibero-Americano de Educação Matemática* (CIBEM) (13 pp.), Faculdade de Ciências da Universidade do Porto, 17-22 de julho.

Fernandes, J. A., & Correia, P. F. (2009). Variação em contexto de probabilidades na perspetiva de professores de Matemática. In A. Gomes (Ed.), *EME 2008 – Elementary Mathematics Education* (pp. 197-208). Braga: Universidade do Minho e Associação para a Educação Matemática Elementar.

Fernandes, J. A., Batanero, C., & Gea, M. M. (2019). Escolha e aplicação de métodos estatísticos por futuros professores dos primeiros anos. In J. M. Contreras, M. M. Gea, M. M. López-Martín, & E. Molina-Portillo (Eds.), *Actas del Tercer Congreso Internacional Virtual de Educación Estadística*. Granada: Universidade de Granada.

Fernandes, J. A., Carvalho, C., & Correia, P. F. (2011). Contributos para a caracterização do ensino da Estatística nas escolas. *Boletim de Educação Matemática (BOLEMA)*, *24*(39), 585-606.

Fernandes, J. A., Carvalho, C., & Ribeiro, S. A. (2007). Caracterização e implementação de tarefas de Estatística: um exemplo no 7.º ano de escolaridade. *Zetetiké*, *15*(28), 27-61.

Fernandes, J. A., Morais, P. C., & Lacaz, T. V. S. (2011). Representação de dados através de gráficos estatísticos por alunos do 9º ano de escolaridade. *Anais da XIII Conferência Interamericana de Educação Matemática*, Recife, Brasil, 26-30 junho de 2011.

Freitas, A., Cruz, J. P., & Silva, N. (2017) Mediana de dados não agrupados: a questão de ser pelo menos 50%. *Educação e Matemática*, 143, 18-21.

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.

Gonzalez, J. F., & Eudave, D. (2018). Modelos de análisis del conocimiento matemático y didáctico para la enseñanza de los profesores. *Unión: Revista Iberoamericana de Educación Matemática*, *54*, 25-45.

Hill, H. C., Ball, D. L., & Schilling, S. G. (2008). Unpacking pedagogical content knowledge: Conceptualizing and measuring teachers' topic-specific knowledge of students. *Journal for Research in Mathematics Education, 39*(4), 372-400.

MacGillivray, H., & Pereira-Mendoza, L. (2011). Teaching statistical thinking through investigative projects. In C. Batanero, G. Burril & C. Reading (Eds.), *Teaching statistics*

in school mathematics – Challenges for teaching and teacher education: A joint ICMI/ IASE study (pp. 109-120). New York: Springer.

Ministério da Educação (2007). *Programa de matemática do ensino básico*. Lisboa: Autor.

Ministério da Educação e Ciência (2014). *Programa de Matemática A – Ensino Secundário*. Lisboa: Autor.

Ministério da Educação e Ciência. (2013). *Programa de matemática para o ensino básico*. Lisboa: Autor.

Morais, P. C., & Fernandes, J. A. (2011). Realização de duas tarefas sobre construção, leitura e interpretação de gráficos estatísticos por alunos do 9º ano. In *Actas do XXII Seminário de Investigação em Educação Matemática (XXII SIEM)*. Lisboa: Associação de Professores de Matemática.

Ruiz, B., Arteaga, P., & Batanero, C. (2009). Competencias de futuros profesores en la comparación de datos. In L. Serrano (Ed.), *Tendencias actuales de la investigación en educación estocástica* (pp. 57-74). Melilla: Grupo de Investigación en Educación Estadística.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, *15*(2), 4-14.

Wild, C., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, *67*(3), 223-248.