

## Do you need to see it to believe it? Let's see statistics and geometry dynamically together!

José Alexandre Martins<sup>1</sup>, Assumpta Estrada Roca<sup>2</sup>, Maria Manuel Nascimento<sup>3</sup>

<sup>1</sup>UDI-IPG, Instituto Politécnico da Guarda, Guarda, Portugal

<sup>2</sup>Universitat de Lleida, Lleida, Spain

<sup>3</sup>CM-UTAD, Universidade de Trás-os-Montes e Alto Douro, Vila Real, Portugal

For correspondence: jasvm@ipg.pt

### Abstract

Statistical graphs, measures of central tendency and measures of spread are key concepts in the statistics curriculum, so we present here a dynamic method (software) that may be used in the classroom. In this work we begin with an introductory approach. This is done to emphasize the importance of stimulating the visualization of statistical measures by using available technology as a means of overcoming difficulties and errors related to interpretations and also to motivate students in the classroom.

With the help of the dynamic geometry software Cabri-Géomètre II Plus, some applications will be presented from a didactic viewpoint as a means to give visual stimulation, to motivate and to facilitate the familiarization with statistical concepts involved in statistical graphs, measures of central tendency and measures of spread.

The applications presented here may be implemented by any teacher with basic knowledge of Cabri-Géomètre or any other dynamic geometry software. We hope that they will also be elements that promote further interaction within the classroom.

**Keywords:** Statistics, education, concepts, visualization, simulation, Dynamic Geometry Software (DGS).

### Introduction

The nature of mathematics means, as stated by Guzmán (2001), that visualization plays an important role both in introducing mathematical concepts and in searching for relationships between them.

Caraça (1998) describes the evolution of the confrontation between numbers and geometrical figures over almost twenty centuries. The gap between geometry and arithmetic led to the virtual exclusion of any hint of movement in the field of geometry, thus putting aside anything linked to quantitative and dynamic conceptions. From the 10<sup>th</sup> to the 16<sup>th</sup> centuries the gap between the geometric and the analytic fields started to narrow. For the last four or five centuries, this reconciliation has brought about significant progress in the history of knowledge and has confirmed the importance of visualization in the field of mathematics.

Additionally, in the last few decades, “the belief that technology can positively impact student learning has led many governments to create programs for the integration of technology in their schools” (Hew and Brush, 2006). Certainly, the use of technological tools gives rise to new teaching possibilities that make it possible for students to have a different, more efficient and effective, learning experience, as “it has been argued that supplementing the traditional material with tools based on a visual approach and a more active form of learning could improve the effectiveness of the teaching” (Darius et al, 2002). But this poses a great challenge and a huge workload for the teacher, as has been pointed out by Godino, Ruiz, Roa, Pareja and Recio (2003) regarding the use of applets with young learners in the classroom.

Mathematical ideas, concepts and methods are often introduced with a wealth of visual content that can be represented geometrically. The usefulness of this visual content, especially in a dynamic

environment, is great, both for introducing and manipulating specific concepts and also for theoretical demonstrations and problem solving.

Clearly, visualization is another step towards improving teaching that requires in-depth thought and research to make it suitable for the teaching of specific concepts.

In a complementary way, as stated by Martins (2003, 2006a, 2006b) and by Martins and Nascimento (2009), the statistical curriculum allows the development of the (dynamic) visualization of several concepts.

There has been a shift from the need to give people basic knowledge to the current aim of fostering people's statistical literacy in today's world. In the last 50 years, there have been increasing numbers of students studying statistics, and of degrees and courses on statistics, and at the same time statistical learning has been introduced at a non-university level, from high school down to elementary education, because "statistics provides a good way of achieving communication skills, information processing, troubleshooting, computer usage, cooperative and group work needed in today's curricula" (Batanero, 2001, p. 118), or, in other words, because of "the need for increased statistical content at school levels to improve statistical literacy in young students around the world" (Batanero et al, 2011, p. 407).

On the other hand, there is a growing feeling of disaffection towards mathematics and also towards statistics. This phenomenon indicates a need for change and is a source of controversy amongst mathematics and statistics teachers. Therefore, a teaching method based only on theoretical exposition, manipulation and computation becomes inadequate. This calls for a teaching method that incorporates a hands-on approach based on real life, and that encourages real and integrated acquisition of basic concepts while developing statistical reasoning. So, in order to empower statistics teachers, some training approaches are suggested: "promoting teachers' statistical literacy and statistical reasoning; engaging teachers with real data and training teachers with project work and statistical investigations; working with technology; and connecting teacher education to their own practice" (Batanero et al., 2011, p. 412).

In order to overcome the formal-theoretical approach and the obstacles it puts in the way of the implementation of the actual paradigm of teaching statistics (Carvalho, 2006), the teacher plays a crucial role in changing his or her classes in order to build the students' correct conceptions of statistics and to develop their statistical reasoning.

Considering the huge potential of graphic and visual exploration in the current technological environment and the distinctive nature of statistics, visualization tends to be perceived as an immensely important aspect both for building up and transmitting concepts, and also for discovering new relationships (Guzmán, 2001; Knipping and Manyá-Ndjeka, 2009). Thus, it is essential to stimulate the visualization of graphs and statistical measures and to try to overcome such issues as difficulties in interpretation and/or motivation, using the technologies currently available.

This paper aims to: disseminate the potential of dynamic geometry software, which supplies a dynamic and visual feedback for teaching statistics; facilitate teachers' work by revealing ideas, paths and methods for using dynamic geometry software to teach statistics; and create dynamic visual forms to reinforce and enable students to assimilate concepts and explore extreme situations in order to consolidate the knowledge they have acquired and therefore to increase their critical thinking abilities.

The applications built into the dynamic geometry software Cabri-Géomètre II Plus allow visual motivation and facilitate students' contact with graphic representations and with concepts such as mean, median, mode and dispersion and with some of their properties. We also think that it has the potential to stimulate students' participation in classroom discussions (smaller groups or the whole class), with a playful component for young students.

## Measures of central tendency

As pointed out by Batanero (2000), “statistical concepts, even simple ones such as mean, median and mode have a complex meaning and require a more extensive teaching period throughout primary and secondary education to achieve a progressive connection between the personal meanings built up by the students and the institutional meanings that we intend for them to acquire”. Taking into consideration this difficulty, the following statistical measures will be tackled using Cabri software.

### *Arithmetic mean*

Measures of central tendency, like the mean, have a special interest as they are frequently found in daily life and as an understanding of them is an important component of statistical literacy (Groth, 2006). Consequently, despite the vast research work with students and pre-service and in-service teachers about this concept, there are still some problems and issues to be solved. For instance, even Grade 1 exemplary teachers do not have a conceptual knowledge of some basic statistical concepts like the mean (Jacobbe, 2008). A frequent mistake in computing the mean was found by Martins, Pires and Barros (2009): the mean was computed using absolute frequencies in coded qualitative variables. In addition, incorrect interpretations are given from the maximum value, the minimum value, a specific value, the median and the mode (Chatzivasileiou et al, 2010; Leavy and O’Loughlin, 2006). In order to enlighten the understanding of the mean, Leavy and O’Loughlin (2006) indicate that there are two types of understanding of the mean – one is conceptual and the other is procedural. Interpretations of the mean as the value that represents the data set as if all data were equal (fair share), or as the point where higher values compensate lower values (balance point), show a conceptual understanding of the mean. When considering how the properties of the mean are taught and learnt, Strauss and Bichler (1988, as cited in Batanero et al, 1994) studied the evolution in the awareness of the concept of the mean in students aged from 8 to 12, and identified the following properties:

- The mean is a data point located between the extreme values of a distribution;
- The sum of the deviations of data points about the mean equals zero;
- The mean value is influenced by particular values in a data set;
- The average does not have to be one of the values in the set;
- When the mean is calculated, the result may be a fraction (although this may not make sense for the variable under consideration);
- When the mean is calculated, any value of zero must be also taken into account;
- The mean value is a representative of the values that were averaged.

In order to get young learners to “build” the concept of the mean and deduce some of its properties, we will present Cabri applications for a game and a scale that we devised. Obviously, a teacher can adapt the applications for use in class, tailoring them to the students’ profile and the course objectives.

### *The up and down game*

The game (Figure 1) is based on the idea that the arithmetic mean represents a given value such that if all the elements of a set had that value, their total sum would equal the sum of all values of the original data. The arithmetic mean is computed by adding all the values of the data and dividing it by the number of elements. Therefore, in the first stage of the game six equal “recipients” with a maximum capacity of 5 and an initial score of zero points are presented. As shown in Figure 1, each subsequent stage begins from the level of the previous stage, and only two symmetric changes can be made in the  $v_i$ ,  $i = 1, 2, \dots, 6$ ; if non-symmetric changes are made, the error no longer has the value 0 (and this is accompanied by the representation of a segment proportional to the error). In addition, the score is calculated as the learner progresses (i.e. the closer the learner is to the solution, the higher the score). The maximum score is the sum of the absolute values of the deviation from the original data in relation to the mean. The game ends at the fourth stage.

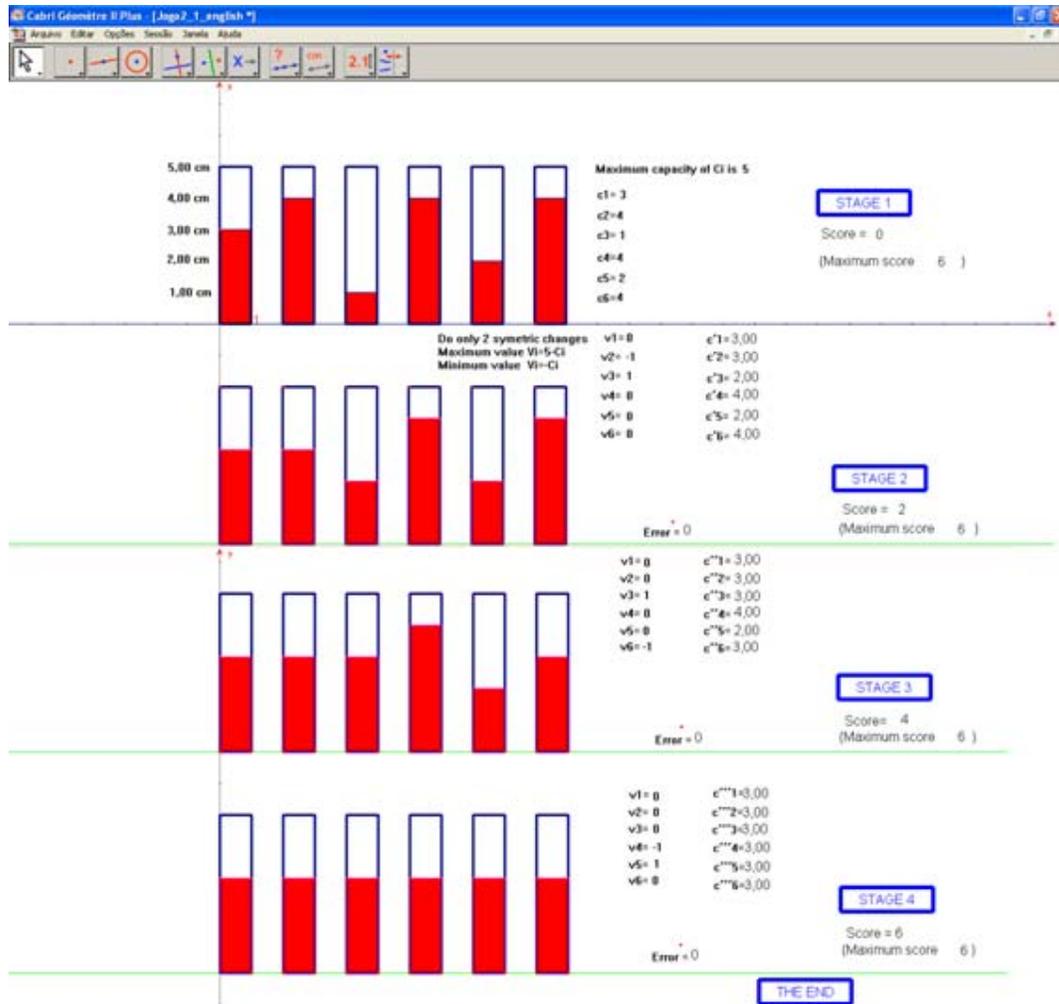


Figure 1: Simulation of the four stages of the game

This simple game allows a real-life experience to be simulated, with an instantaneous monitoring system to deal with measurements and an accurate computation of the arithmetic mean. By means of cooperation in small groups or in the whole class, confrontation emerges when several strategies and experimental results may be discussed. Depending on the students' ages, the use of negative numbers may be a controversial issue to be explored by the teacher, since it calls for full attention, and help from the teacher may be needed in order to tackle these problems, using several examples.

*The teeter-totter scale*

For non-classified data, one of the expressions for the arithmetic mean is  $\bar{x} = \left( \sum_{i=1}^N x_i \right) / N$ .

The central moment of order one is equal to zero (0); that is:  $\sum_{i=1}^n (x_i - \bar{x}) = 0$ .

The mean may be viewed as the fulcrum of a data set. In other words, the sum of the deviations of the data points about the mean equals zero, or "the arithmetic average is the score around which deviations in one direction exactly equal deviations in another direction" (Leavy and O'Loughlin, 2006). The arithmetic mean can therefore be interpreted as an x value (fulcrum) that is the centre of gravity of a system made up of N points with each assuming a unitary weight.

Taking the relationship (the fulcrum) between mean and geometry as the basic principle, the Cabri-based application scale was created. There are six points on a straight line:  $x_1, \dots, x_5$ , and the M point, which can be moved along the straight line. When the M point is moved, the value Vsum appears on the screen, representing the measure of the sum vector of all the differences between the x points and M. The mean can then be obtained dynamically by moving the M point along the straight line in such a way that the value Vsum becomes zero. As the M point is moved, one can look at the “teeter-totter scale”, which is evenly balanced whenever M has the same value (i.e. the x coordinate value) as the mean, and goes off balance whenever M moves away from the mean. This simple graphic visualization shows the relationship between “weight” (using the centre of gravity, the fulcrum) and the mean (Figure 2). Finally, this application allows students to see the analytical validation for the arithmetic mean value through the M point.

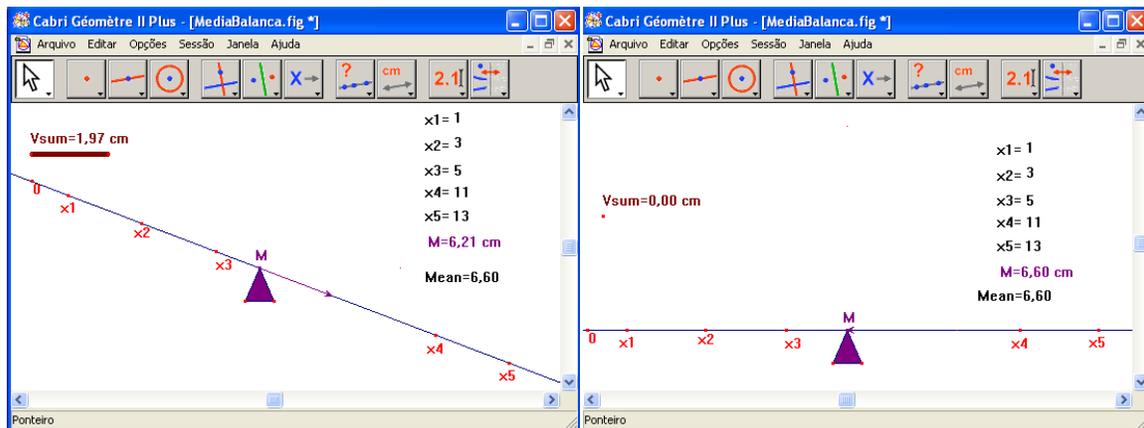


Figure 2: Mean on the teeter-totter scales

Other possibilities include placing the points on a fixed straight line, assuming unitary or non-unitary frequencies and using a teeter-totter scale design that represents the equilibrium whenever the M point coincides with the mean, and goes proportionally out of balance (non-equilibrium) as the M point moves away from the mean (Figure 3). With these non-unitary frequencies seven points were used.

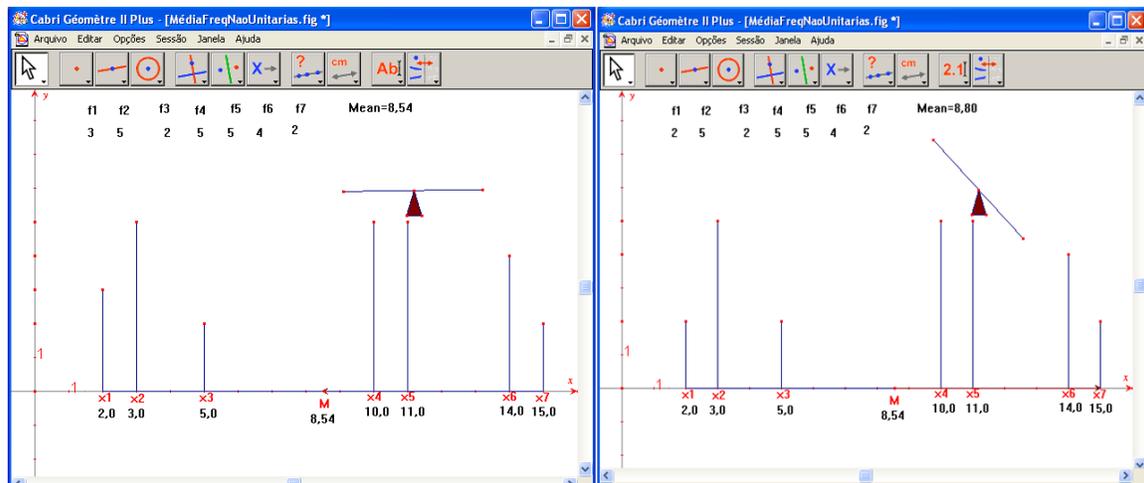


Figure 3: Mean (non-unitary frequencies) and the teeter-totter scales

Moreover, another application was created for continuous data, as a simulator of a histogram and with the same approach used in the previous applications.

Those applications may be used by teachers (or their students) for different grades. The concept of arithmetic mean and its properties (Batanero et al., 1994) is dynamically explored for different data types.

*The stats middle number*

The importance attached to the study of the statistics of order values such as median and percentiles is high, as exploratory data analysis is also based on these values. This is due to their “robustness”, that is, they are less sensitive to minor data changes and outliers than other measures of central tendency; besides, they are the basis of some non-parametric methods. Furthermore, they can be applied more widely, despite being less powerful.

Cobo and Batanero (2000) point to four different, but equivalent, ways of defining median, each one with its own difficulties and ambiguities. Another difficulty noted by these authors was that there are different ways of computing of the median, not only for different types of variable (discrete and continuous), but also depending on the sample size (even or odd in the discrete case) and the way the variables are presented. These authors also describe some numerical, algebraic and statistical properties of the median. With the Cabri application it is easy to work out the numerical properties, and we highlight these now:

- The median may not coincide with any of the data values;
- The computation of the median does not include all data values;
- The median is invariant if data below it or data above it change.

From the viewpoint of didactics, “the study of statistics of order values shows difficulties, both at procedural and conceptual level” (Batanero et al., 1994). These authors referred to Barr (1980), Estepa (1990) and Schuyten (1991) to point out the following difficulties and errors in computing median, percentiles and percentile range:

- Different algorithms are taught for the case in which the statistical variables are grouped together into intervals and the case in which they are ungrouped, and even college students have difficulty in accepting that two different algorithms can be used for the same average value and that two different values can result for the same parameter, if the width of the class intervals is varied;
- There is a difference between the concept of median and the computation method used to obtain its value. From the definition of median as the numerical value of the statistical variable separating the higher half of the sample of a population from the lower half when data have been arranged in ascending order, to its computation based on the graphs of cumulative frequencies, there is a series of steps that is not always entirely understood;
- The interpretation of the graph of cumulative frequencies of discrete variables causes difficulties. These graphs show a discontinuous distribution and it is not possible to use them in reverse: a point can have more than one image, or several points may have the same image;
- There is a lack of understanding as regards the numerical set for which the central value is taken. Students can interpret the median as the central value of variable values, variable frequencies or even variables in a series of data prior to ordering it.

In an attempt to overcome some of these difficulties, a Cabri application was created to explore the underlying geometry of the concept, computation and properties of the median. We start with a brief introduction to the concept of median, followed by the presentation of the application that should be used under the supervision of a teacher.

It is widely accepted that the median is a localization measure defined by its position in a succession of observations or a frequency distribution.

In the case of continuous variables, the median is the abscissa that corresponds to the 50% cumulative frequency on the graph. This establishes a geometric and graphic process to obtain the (approximate) value of the median for the continuous case.

On the other hand, a histogram shows frequencies interpreted as areas, and the median is defined as the variable value whose corresponding ordinate on the abscissa axis divides the area of the histogram into two equal areas.

Finally, assuming the hypothesis that frequencies are uniformly distributed within each class (i.e., the cumulative frequencies show a linear variation within each class and use similar triangles), the plotted graph of cumulative frequencies gives the formula for the median computation.

The median computation of a continuous data set applies all the geometrical concepts presented. We used it in the Cabri Stats Middle Number application (Figure 4), to work out the characteristics in a dynamic way:

- Dynamic graphic localization of the median;
- Equality, in the histogram, of the left-hand and right-hand areas of the median, and a visualization of the relationship between the histogram and the cumulative frequency graph;
- The similarity of the triangles is the origin for the formula for analytical computation of the median, which is immediately seen.

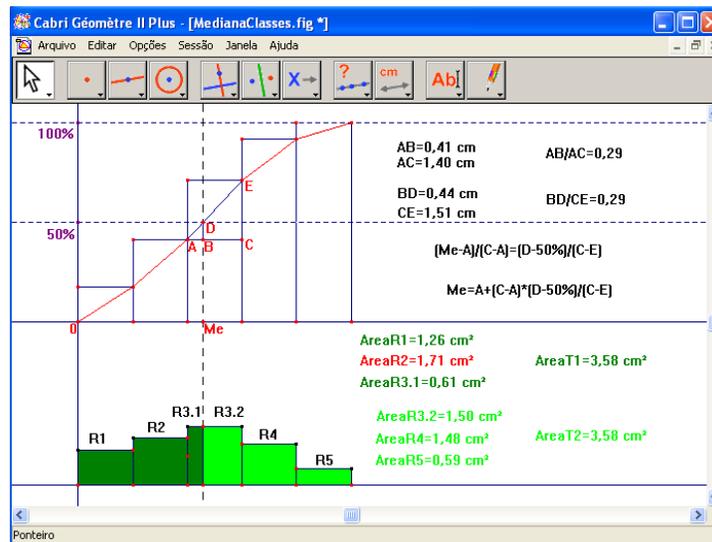


Figure 4: Graphic localization of the median

#### The stats "common" number

We begin by stating that the whole of our presentation supposes that for continuous variables presented in intervals (grouped data), the intervals are all of equal width. Although the modal interval in the frequency distribution of a continuous variable, that is, the interval with the highest frequency, is a familiar concept, determining the mode (the most common value or the stats "common" number) may create some computational problems. There are several empirical formulae to identify the mode within a modal interval, but these are approximate values. It is therefore important to be better acquainted with how these formulae work.

A geometric process to obtain the mode or "common" number is based on the principle that, within the modal interval, the mode is located closest to the adjacent interval with the higher frequency, and similar triangles are used to obtain the computation formulae.

Another grouped data formula to obtain the mode, based on the same principle as the previous formula, is King's formula (Murteira and Black, 1983). Its analytical expression is also obtained through valid relationships of similar triangles.

These are the conditions for building an application that aims to test the relationships mentioned above and to visualize the two processes of obtaining the mode value dynamically, making them easier to understand.

In this respect, visualization of constant relationships can be explored with one single application that looks at the opposing triangles that constitute the basis of the formulae for computing the mode for the grouped data.

For this and other applications, an equal width histogram can be constructed to obtain the mode through an analytical process, that is, to confirm geometrically and dynamically the relationships that immediately allow us to identify the modal value that we have computed. Besides, by duplicating the histogram, we may obtain the mode value using King's formula and compare the two estimated values (Figure 5).

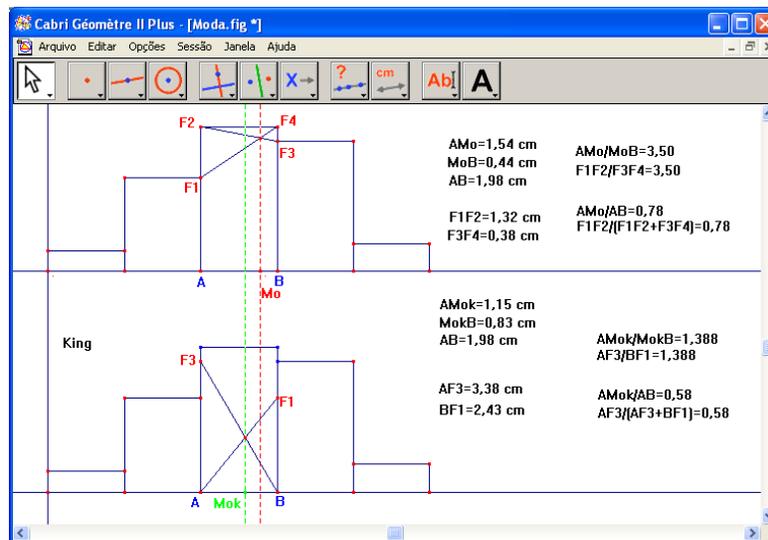


Figure 5: Computing the mode: analytical process versus King's process

This allows the relationships mentioned above to be tested, and gives a dynamic visualization of the two processes for obtaining the modal value for the grouped data. Finally, when comparing the processes, one can verify that, on the one hand, the analytical process is more sensitive to frequency variations of the adjacent classes if they are higher (closer to the modal frequency), and that, on the other hand, King's process is more sensitive to frequency variations of the adjacent classes if they are lower (closer to zero frequency). It is interesting to explore the extreme cases for the frequency values of adjacent intervals, and to compare the results obtained using the two methods, and also to simulate examples with the same mode values in both models.

### Measures of spread

Despite the importance of statistical variation, the approach to it in the school curriculum is often limited to the computation of measures, which does not fully reflect the variation concept. Some explanations for this situation are pointed out by Shaughnessy (1997):

- Computation and interpretation are difficult;

- There is a lack of didactic models to give meaning to the variation measures, in comparison with the models that help motivate the concepts of central measures (scales, equilibrium point, etc.).

We should consider statistical variation as a key concept in statistical thinking that should lead us to an in-depth reflection on the role given to it in teaching and learning statistics (Wild and Pfannkuch, 1999). The approach to the variation concept should lead to enriched work within the statistical classroom. So, we will use Cabri for a different exploration of the relationship between frequencies and the value of variance, the graphical representation of variance and the idea of variance as a mean value.

### *Circling the Variance*

In the case of continuous variables, the first aim is to relate the histogram to the variance value and to visualize it dynamically. To achieve this, a histogram is built that allows the frequencies to be changed dynamically and the corresponding variance value to be verified, by means of the area of its associated circle.

This allows students to explore changes in frequencies and to follow their results. It is particularly interesting to consider situations with equal means but different variances (Figure 6). It is also possible to try different “shapes” for the histogram (asymmetry versus symmetry, and also different kurtosis) and immediately to discuss the effect of these shapes on the variance. Another way to use this application is to play with students, asking them, for instance, to predict the effect on the variance value of a change in the frequency of one of the intervals.

These experiments make it possible to understand the complexity of the concept of variance, and to stimulate the students’ intuition and critical skills as regards dispersion and measures of spread.

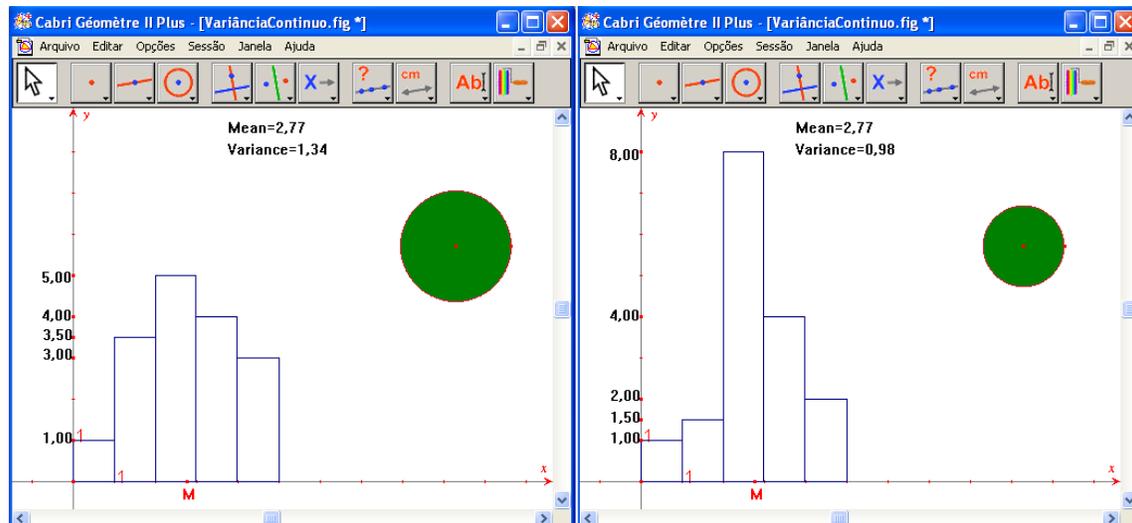


Figure 6: Variance and frequency

### *Variance as a mean?*

At this point, the application aims to visualize the variance value and to give a dynamic visualization of a discrete variable based on the variance formulae. Here the variance is considered as a mean, since it is the mean of the squared differences from the arithmetic mean. In this way, we remind students that the differences from the arithmetic mean add up to zero (the above  $V_{sum}=0$ ). Therefore, considering a new variable  $y_i$  as the squared difference between a value and the arithmetic mean, that is  $y_i = (x_i - \bar{x})^2, i = 1, \dots, n$ , one can say that the variance of the data  $x_i$ 's is equal to the mean of the data

$y_i$ 's, that is,  $\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{n} = \frac{\sum y_i}{n} = \bar{y}$ . Thus,  $\sum_{i=1}^n (y_i - \sigma^2) = \sum_{i=1}^n (y_i - \bar{y}) = 0$ . So, in Figure 7,  $y_i = (V_i)^2$  and  $\sigma^2 = V$  only when  $V_{sum} = 0$ . With the Cabri-Géomètre software it is possible simultaneously to calculate the mean and the data variance, using the same property as the mean for the two related data sets.

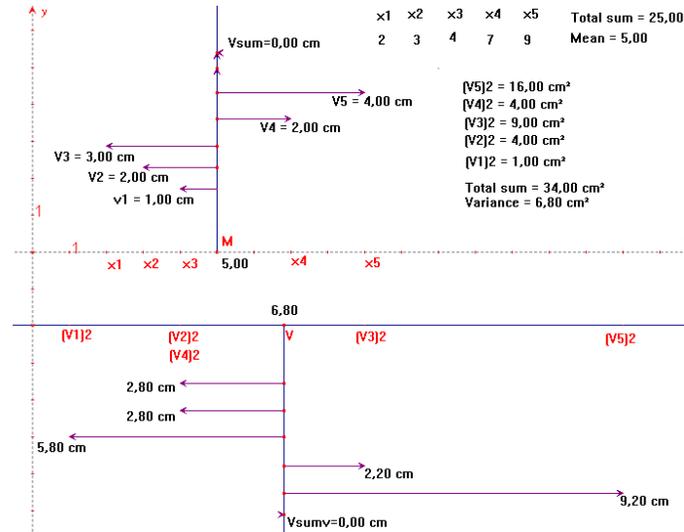


Figure 7: Variance as a mean

*Tree of squares*

This application aims to visualize, by means of squares (with their areas and the lengths of their sides), the variance value and its changes, for a discrete variable (with five values), using the variance formula that states that the variance is the sum of the squared differences from the mean, divided by the number of points in the sample.

In the geometrical context, variance comes from the (arithmetic) mean of the areas of a square with sides equal to the distance between each individual data point and the arithmetical mean.

Based on this geometrical interpretation, the tree of squares (Figure 8) represents, on the left of the mean (Mx) the squares from the points with lower data values and on the right of the mean (Mx) the squares with higher data values, thus creating a picture similar to a tree. Finally, the application shows a grey square with an area equal to the mean of the areas of the aforementioned squares and with a side equal to the standard deviation.

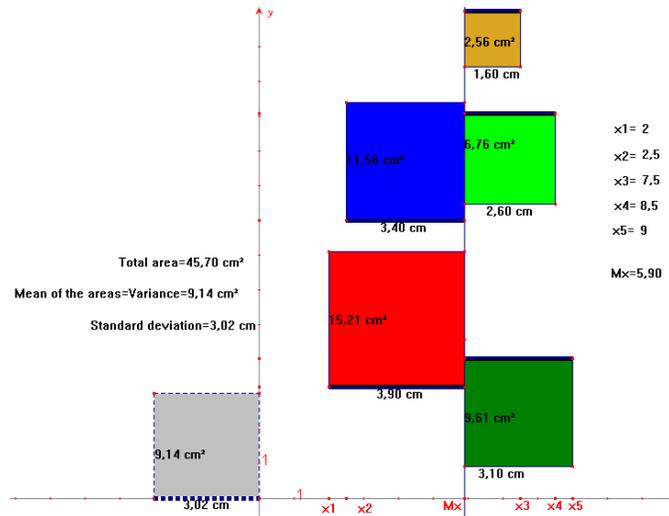


Figure 8: Graphic representation of variance

This application allows a better visualization of the geometric variance concept, and helps one to explore the changes in the variable values. It reveals that the variance is highly sensitive and that its value mainly depends on the order of magnitude of the data set values. Using this Cabri application, students may discover what happens when they impose different conditions, for instance, with equal distances between the data set values (such as 3, 5, 7, 9 and 11, a difference of two, or a difference of one such as 3, 4, 5, 6 and 7). Another interesting data set may be chosen with all the numbers except one (the outlier) very close together; students can “guess” and then “see” the effect on the value of the variance. Other situations may be devised by the teacher or even suggested by the students.

### Pictures of data

According to Arteaga, Batanero, Contreras and Cañadas (2012, p. 261) “graphs are pervasive in our society and can be used to communicate information efficiently, as a tool for data analysis (Cazorla, 2002). Moreover, in many school topics graphs are used to visualize abstract concepts and relationships (Postigo and Pozo, 2000)”. The same authors also stand that the “construction and interpretation of statistical graphs is also an important part of statistical literacy according to Watson (2006). Despite this relevance, educational research alerts us that the competence related to statistical graphic language is neither achieved in the students (Friel, Curcio and Bright, 2001) nor in the preparation of future primary school teachers (González, Espinel and Ainley, 2011).” (Arteaga et al., 2012, p. 261).

To help in overcoming this issue, the applications presented in this section were devised to stimulate data visualization through graphs and to establish relationships between different types of graphs when students or teachers are manipulating different data sets.

### Histogram pictures

This application shows that, in the case of equal-width intervals, the areas of the histogram and the frequency polygon are equal (Figure 9). The use of colours enables the dynamic visualization of this equality.

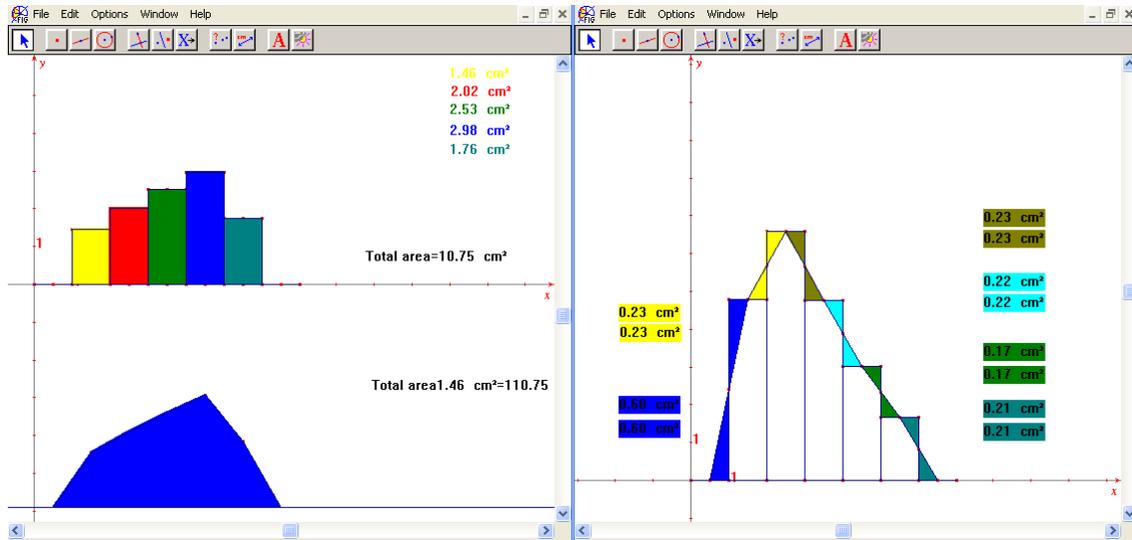


Figure 9: Histogram and frequency polygon (two different perspectives)

*Pie picture*

Another application we have developed shows the relationship between a bar graph and its corresponding pie graph, by presenting its corresponding angles and verifying that there is a constant relationship between the bar areas and their corresponding angles (Figure 10).

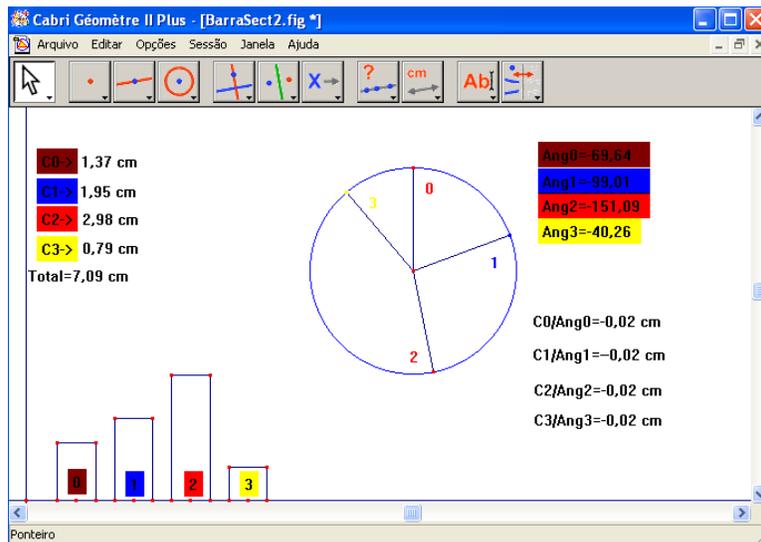


Figure 10: Bar graph and pie chart

*Box-plot picture*

Finally, another application shows the quartiles, using a histogram and a cumulative relative frequencies polygon, as well as the construction of the box-plot (Figure 11), giving a dynamic vision of a difficult graph such as the one highlighted by Lem, Onghena, Verschaffel and Van Dooren (2011). The same authors indicate that students' major misconception when interpreting box plots is that "they tended to interpret the area of box plots incorrectly by assuming that a larger area represented more observations than a smaller area".

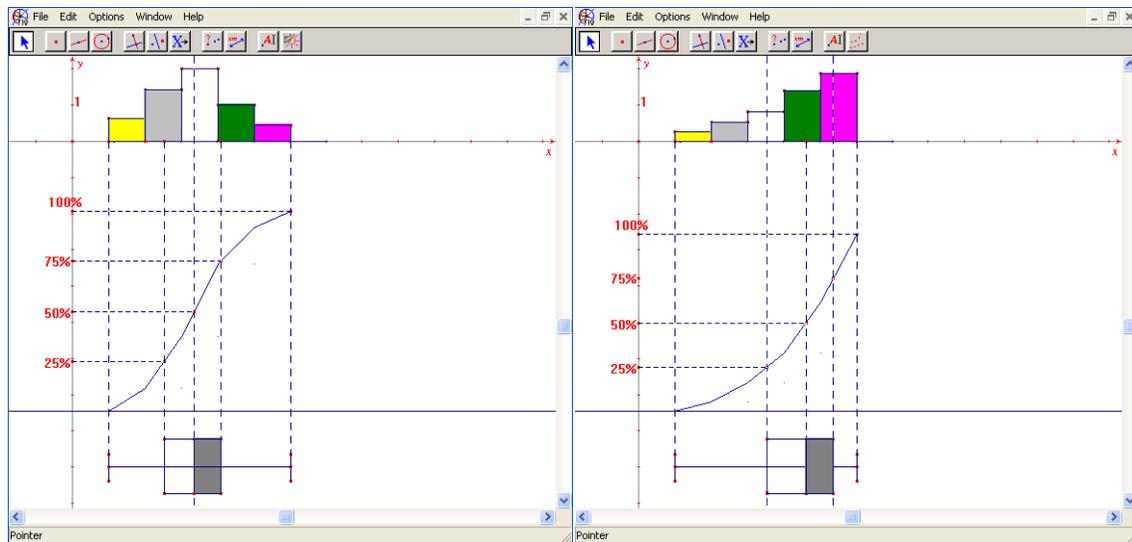


Figure 11: Box-plot

### Final remarks

We think that we have achieved our main goal – to see statistics and geometry dynamically together through examples that use versatile Cabri-Géomètre applications. With the Cabri-Géomètre applications we had devised, we showed the potential of dynamic geometry to encourage and to help the interpretation and the understanding of basic key statistical concepts. In this sense, we believe that the applications we have presented could enhance students' learning and help to overcome some of the known difficulties of students with these topics. These same applications may also be able to promote a more interactive and really engaged class for the teacher and his or her students.

As the experiences that we had before with this Cabri applications were very positive both with students and teachers, the next step will be to promote its use in classes and design studies in order to access if the use of GDS applications in classes - such as this ones that we presented here - will contribute in a significant way to motivate and improve either the comprehension or its practical use (data analysis) of the referred statistical concepts.

Finally, we are sure that the examples presented offer a wide range of possibilities to explore, to delve into and to improve. Other dynamic geometry applications could also be implemented focusing those and other statistical topics. The only requirement for this purpose is using one's curiosity, imagination, reflection and willingness, since the applications presented may easily be implemented by any teacher with a basic knowledge of Cabri-Géomètre or any other dynamic geometry software.

### Acknowledgements

This research was supported by the project EDU2010-14947 (MICIIN, Spain), by Centro de Matemática da UTAD (CM-UTAD) and by the project PEst-OE/EGE/UI4056/2011 financed by Fundação para a Ciência e a Tecnologia (Science and Technological Foundation, FCT, MCTES, Portugal).

### References

- Arteaga, P., Batanero, C., Contreras, J.M., and Cañadas, G. (2012). Understanding statistical graphs: A research survey. *Boletín de Estadística e Investigación Operativa (BEIO)*, 28(3), 261-277.
- Barr, G. V. (1980). Some student ideas on the median and the mode. *Teaching Statistics*, 2(2), 38-41.
- Batanero, C., Godino, J., Green, D., Holmes, P., and Vallecillos, A. (1994). Errors and difficulties in understanding elementary statistical concepts. *International Journal of Mathematics Education in Science and Technology*, 25(4), 527-547.
- Batanero, C. (2000). Significado y comprensión de las medidas de posición central. [Meaning and understanding of measures of central tendency]. *UNO*, 25, 41-58.

- Batanero, C. (2001). *Didáctica de la Estadística*. [Didactics of Statistics]. Department of Mathematical Didactics, University of Granada: Granada, Spain:
- Batanero, C., Burrill, G., and Reading, C. (2011). *Teaching statistics in school mathematics - Challenges for teaching and teacher education*. A Joint ICMI/IASE Study. ICMI Study, Vol. 14. New York: Springer.
- Caraça, B. (1998). *Conceitos Fundamentais da Matemática* [Fundamental Concepts of Mathematics] Lisbon, Portugal: Gradiva.
- Carvalho, C. (2006). Desafios à educação estatística [Challenges to Statistics Education]. *Boletim da Sociedade Portuguesa de Estatística – Ensino e Aprendizagem da Estatística*, Outono 2006, 7-9.
- Cazorla, I. (2002). *A relação entre as habilidades viso-pictóricas e o domínio de conceitos estatísticos na leitura de gráficos* [Relating visual-pictoricability and knowledge of statistics in reading graphs]. Unpublished Ph.D. University of Campinas, Brazil.
- Cobo, B. and C. Batanero (2000) La mediana en la educación secundaria obligatoria: ¿un concepto sencillo? [The median in the secondary education: a simple concept?]. *UNO*, 23, pp. 85-96.
- Chatzivasilieiou, E., Michalis, J., and Tsaliki, C. (2010). Elementary school students' understanding of concept of arithmetic mean. In C. Reading (Ed.), *Proceedings of the 8th International Conference on Teaching Statistics*. Ljubljana, Slovenia.
- Darius, P., Michiels, S., and Raeymaekers, B. (2002). Applets for Experimenting with Statistical Concepts. In B. Phillips (Ed.), *Proceedings of the 6th International Conference on Teaching Statistics*. Cape Town, South Africa.
- Estepa, A. (1990). *Enseñanza de la estadística basada en el uso de ordenadores: Un Estudio exploratorio*. [Teaching of statistics based on computer use: An exploratory study.]. (Unpublished PhD thesis). Department of Mathematical Didactics, University of Granada: Granada, Spain.
- Friel, S., Curcio, F., and Bright, G. (2001). Making sense of graphs: critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education*, 32(2), 124-158.
- González, M. T., Espinel, M. C. and Ainley, J. (2011). *Teachers' graphical competence*. In C. Batanero, G. Burrill, and C. Reading (Eds.), *Teaching statistics in school mathematics-Challenges for teaching and teacher education*. A Joint ICMI/IASE Study. New York: Springer.
- Groth, R. E. (2006). An exploration of students' statistical thinking. *Teaching Statistics*, 28(1), 17-21.
- Hew, K. and Brush, T. (2006). Integrating technology into K-12 teaching and learning: Current knowledge gaps and recommendations for future research. *Education Tech Research*, 55, 223-252.
- Godino, J., Ruiz, F., Roa, R., Pareja, J., and Recio, A. (2003). Analysis of two internet interactive applets for teaching statistics in schools. In G. Aydynly, W. Hardle, and B. Ronz, (Eds.), *Proceedings of the IAASE Satellite Conference Statistics Education and the Internet*, (pp.1-10). Berlin, Germany.
- Guzmán, M. (2001). *El rincón de la pizarra – ensayos de visualización en análisis matemático – elementos básicos del análisis*. [The corner of the blackboard – display testing mathematical analysis - basic elements of the analysis]. Madrid, Spain: Ediciones Pirámide.
- Jacobbe, T. (2008) "Elementary school teachers' understanding of the mean and median  
[https://www.stat.auckland.ac.nz/~iase/publications/rt08/T2P13\\_Jacobbe.pdf](https://www.stat.auckland.ac.nz/~iase/publications/rt08/T2P13_Jacobbe.pdf)
- Knipping, L. and Manyá-Ndjeka, E.. (2009). The role and potential of visualization in the mathematical education of engineers. In *Proceedings of the 2009 ASEE Annual Conference*. American Society for Engineering Education, Austin, Texas, USA.
- Leavy, A., and O'Loughlin, N. (2006). Pre-service teacher understanding of the mean: Moving beyond the arithmetic average. *Journal of Mathematics Teacher Education*, 9(1), 53-90.
- Lem, S., Onghena, P., Verschaffel, L., and Van Dooren, W. (2011) Coordinating between histograms and box plots  
<http://www.conkerstatistics.co.uk/iase/papers/IASE2011PaperPoster6Lemetal.pdf>
- Martins, C., Pires, M. V., and Barros, P. M. (2009). Conhecimento estatístico: Um estudo com futuros professores. [Statistical Knowledge: A study with future teachers]. In C. Costa, E. Mamede, and F. Guimarães (Eds.) *Proceedings of the 2009 EIEM: Números e Estatística*. Vila Real, Portugal.
- Martins, J. (2003). Estadística con Geometría Dinámica [Statistics with dynamic geometry]. In Department of Statistics and Operations Research of Vigo University (Ed.), *Actas del VI Congreso Galego de Estatística e Investigación de Operacións*, Spain, (pp.73-80).
- Martins, J. (2006a). Statistics visualization with dynamic geometry, In B. Chance and A. Rossman, (Eds.) *Proceedings of the 7th International Conference on Teaching Statistics*. Bahia, Brazil.
- Martins, J. (2006b). Dispersión estadística con Cabri [Statistical Dispersion with Cabri], In Technological Institute of Teaching Mathematics (Eds.) *Actas del 3er Congreso Internacional IberoCabri 2006*. Bogotá, Colombia.
- Martins, J. and Nascimento, M. (2009). Estadística con software de Geometría Dinámica [Statistical with Dynamic Geometry Software], School of Mathematics of the Technological Institute of Costa Rica (Ed.) *Actas del I Encuentro de Didáctica de la Estadística, la Probabilidad y el Análisis de Datos (I EDEPA)*. Cartago, Costa Rica.
- Murteira, B. and Black, G. (1983). *Estatística Descritiva* [Descriptive Statistics]. MacGraw-Hill. Lisbon, Portugal.
- Postigo, Y. and Pozo, J. I. (2000). Cuando una gráfica vale más que 1000 datos: la interpretación de gráficas por alumnos adolescentes [When a graph is worth 1000 data: interpretation of graphs by adolescent students]. *Infancia y Aprendizaje*, 90, 89-110.
- Schuyten, G. (1991). Statistical thinking in psychology and education. In D. Vere-Jones (Ed.) *Proceeding of the 3rd International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute, (pp.486-490).
- Shaughnessy, J. M. (1997). Missed opportunities in research on the teaching and learning of data and chance. In F. Biddulph and K. Carr (Eds.) *People in Mathematics Education*, Vol. 1, (pp.177 –197). Waikato, New Zealand: Mathematics Education Research Group of Australia.
- Watson, J. M. (2006). *Statistical literacy at school: Growth and goals*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Wild, C. and Pfannkuch, M. (1999), Statistical Thinking in Empirical Enquiry. *International Statistical Review*, 67, 223–265.