STATISTICS WITH DYNAMIC GEOMETRY

José Alexandre dos Santos Vaz Martins
Escola Superior de Turismo e Telecomunicações de Seia - Instituto Politécnico da Guarda Rua Dr. José António Fernandes Camelo - Arrifana - 6270-372 Seia – Portugal
jasvm@ipg.pt

ABSTRACT

Students from this generation, learn best by seeing. Nowadays, visualisation is extremely important to help young students to catch the real meaning of some concepts. Therefore, this paper explores the application of a dynamic geometry software (Cabri-Géomètre II) to illustrate basic statistical concepts (like mean, median and mode), their properties and graphical representation, in a way that allows the teacher to explore, discover and uncover those ideas that will get the messages to students and stimulate interactive work in the classroom. In this presentation a main goal is showing a way to approach statistics through the use of a computational tool, with emphasis in the visual exploration of statistical concepts and focus on the improvement of statistical literacy, reasoning and thinking.

INTRODUCTION

A recent trend from that clearly emerge visualisation with a brand new role in mathematic and statistic teaching. This legacy allows us, today, with the computers help, to go further, as learners as well as teachers.

The truth is that, in Mathematics and in Statistics, some ideas, concepts and methods present a great deal of richness of visual contents. From this point of view, Dynamic Geometry Software (DGS) could be very useful, because they have constructive, measurement and dynamic proprieties that allow the creation of some applications with a huge potential in terms of the visualisation characteristics.

The intention of this paper is to present the DGS potential at the service of the Descriptive Statistic basic concepts apprehension, contributing, in this way, to improve statistics teaching, that has reached a overwhelming importance in a curricular and social perspective. It is also intended to facilitate the teacher's work showing ideas, trails and new ways of reaching such objectives. In this sense, some Cabri applications built to get a visual stimulation and to facilitate mean, median and mode concepts acquisition, trough a geometric interpretation of them and of theirs proprieties will be presented, hence dynamic manipulation and exploration of geometric properties may play an important role in teaching and learning statistical concepts and to experience the effect of data values changes on statistical measures. Some Cabri applications with some type of graphics to explore and to establish interrelationships will also be present.

APLICATIONS PRESENTATION

All of the applications presented in this paper are constructed in the dynamic geometry software, Cabri Géomètre II, and most of them could be implemented with rudimentary knowledge of Cabri tools.

Median

According to Cobo and Batanero (2000) the usual median definitions present a few difficulties and ambiguities that persist in the calculus work. The same authors state that the median calculus is based in cumulative frequency display comprehension and in a correct proportional reasoning by the students and also in the understanding of triangle similarity concept.

For the median, the application purposed shows how to get it from an cumulative frequency display (the x corresponding to the 50% y), establishes the relation between the histogram and the cumulative frequency display and, based in triangles relations, deduces the formula for the median calculus in the continuity case, as it is possible to see in figure1.

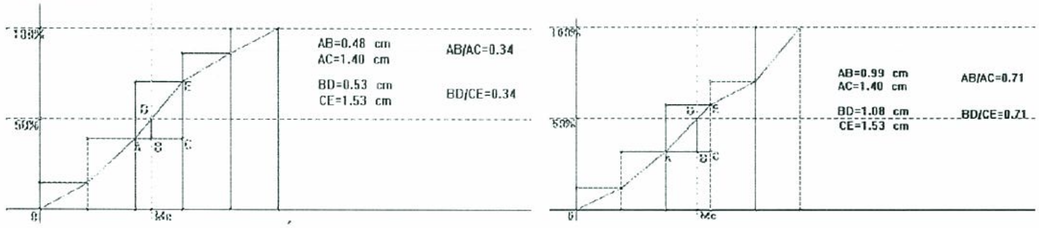

Figure 1 – Obtaining geometrically median and its formula

Besides, it is possible to explore some numerical, algebraic, statistical and geometrical properties. One example is a dynamic verification of the equality of the median right and left areas in the histogram as it is shown in figure 2.

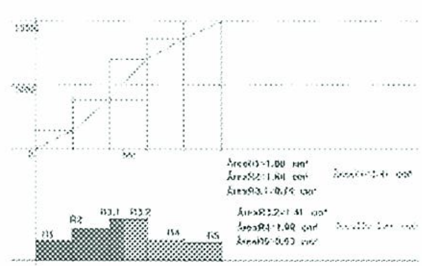

Figure 2 – Equality of the median right and left area

Mode

As it is known, there are several formulas to establish a value for the mode in the modal interval. This value is an approximation, and therefore needs a careful interpretation. In this sense, it is very important to have a better knowledge about how those formulas work. So, based in the principle that, in the modal interval, the mode should be closer to the contiguous interval with the greatest frequency now, like it is visible in figure 3, two geometric processes will be approached to obtain the mode value and its formula of calculus. From a duplicated histogram in a Cabri application, it is possible to compare the values of the mode obtained by the two different ways mentioned. One can also study and understand the behaviour of each geometric process; especially with extreme values in the neighbours' intervals and to find in what situation both values are equal. From triangle relations, it is also possible to deduce the formulas to calculate a value for the mode and confirm that in a dynamic way.

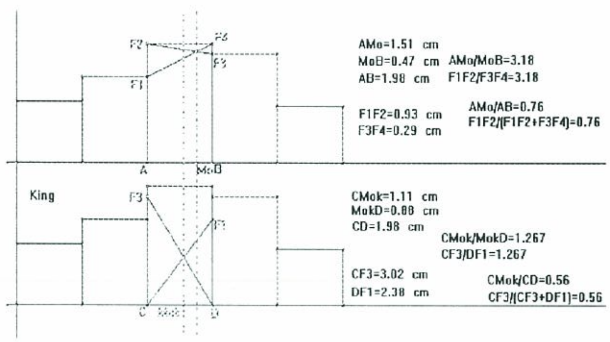

Figure 3 – Geometric processes to obtain the mode

Mean

As it is well known, the mean could have a physical interpretation as the gravity centre. So, based in the arithmetic mean property $\sum(x_i - \bar{x}) = 0$, in the Cabri application proposed a scale was constructed where the respective arithmetic mean represents the equilibrium point, as it is possible to confirm in figure 4. In this application, beside the physical interpretation, it is also possible to confirm analytically the value of arithmetic mean and it could help as a dynamic approach to some of the difficulties related to the arithmetic mean that were mentioned by Batanero (2000).

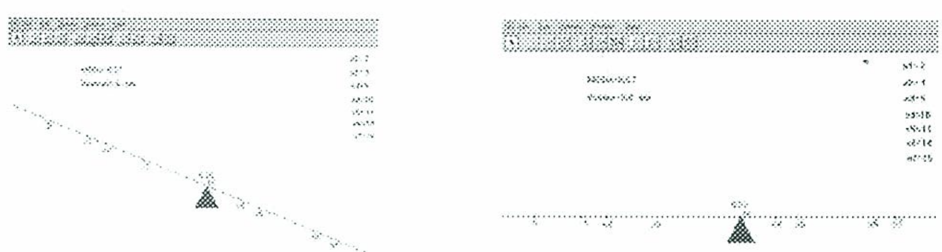

Figure 4 – The arithmetic mean scale

Variance

For the discrete case, the variance formula has a geometric interpretation as the mean of the square areas with length side equal to the distance between each data and the data mean. Therefore, variance is represented by a square area and the standard deviation corresponds to the length side of that square, as it's possible to see in figure 5.

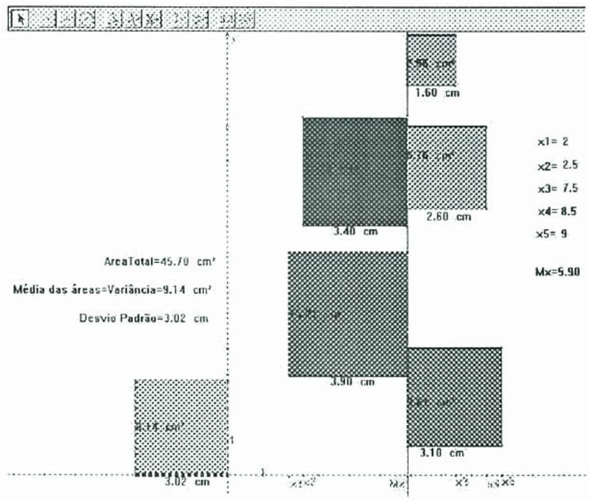

Figure 5 – Geometric representation of the variance

For the continuous case, and with a Cabri application, it's possible to establish the relation between the histogram and the variance value, to stimulate intuitive aspects and critical ability in the students exploring frequency changes dynamically and to show that variance it's highly sensible to data changes and dependent of the order of data values, as suggested in figure 6.

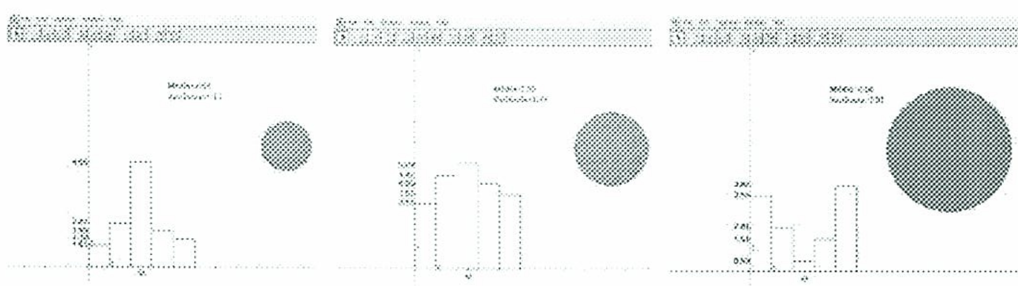

Figure 6 – Relation between histogram and variance

Variance can also be seen like a mean, which has a geometric interpretation that uses the mean property, $\sum(x_i - \bar{x}) = 0$, where the new data are the squares of the differences between the initial data values and the respective data mean.

Linear Regression

The objective of the linear least squares regression application is dynamically fitting a straight line to some bivariate data, obtaining the regression equation by minimizing the sum of the squares of the residuals. This dynamic process is obtained through the displacement of a line, and it's possible to certificate that one of the points of this line is $(\bar{x}, \bar{y})$ as is represented in figure 7.

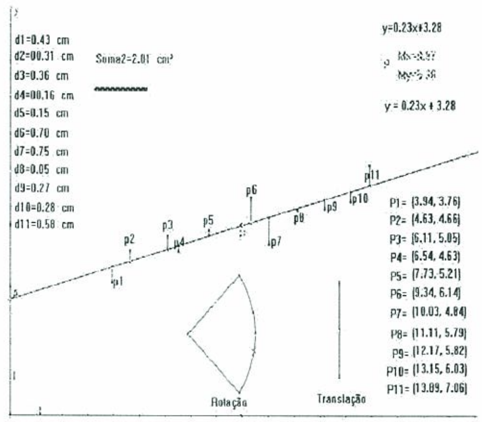

Figure 7– Regression line

Graphics

One example of the use of Cabri applied to graphic issues is one that remarks the relation between a bar graphic and the correspondent pie graphic as it is shown in figure 8. Finally, other Cabri application dynamically shows the relation between histogram, cumulative frequency curve (ogive) and box plot, like it is visible in figure 9.

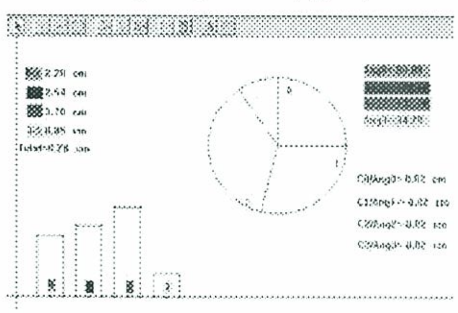

Figure 8 – Relation between bar and pie graphics

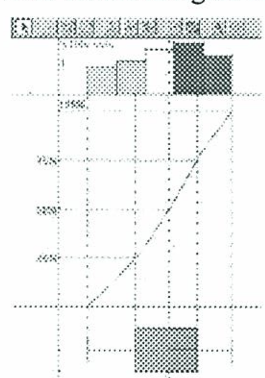

Figure 9 – Histogram, cumulative frequency display and box plot

CONCLUSION

An improvement in statistics teaching using DGS in a conscientious, reflected and careful way is achievable. There are many possibilities to explore, improve and add other potentialities to the presented Cabri applications.