THE USE OF FREE WEB-BASED SOFTWARE TO HELP STATISTICS TEACHING

Hernandez, Hugo; Kataoka, Verônica Yumi; Ávila, Roberto
Colegio de Ciencias y Humanidades, Plantel Vallejo, Universidad Nacional Autónoma de México
Universidade Bandeirante de São Paulo
Colegio de Ciencias y Humanidades, Plantel Vallejo, Universidad Nacional Autónoma de México
animal_estocastico@hotmail.com

Teaching of Statistics has evolved to the use of the computer as a supporting tool in classroom. But this carries new challenges, as it is to choose the best software to use in teaching Statistics; and in developing countries with the additional problem of choosing a cheap or, better yet, a free software to help students in learning Statistics. The Colegio de Ciencias y Humanidades, a high school in Mexico, aims to cover in its scholar program topics such Regression and Correlation; since both topics are traditionally complex, it is suggested to support their teaching with the use of the computer in the classroom. A didactical sequence is proposed, in which both items are covered by working with dynamic demos contained in free software intended to support Statistics learning. In this sequence, students are expected to work in small groups, interacting with the software, and under the perspective of Didactical Situations Theory stated by Guy Brousseau.

INTRODUCTION

Teaching of Statistics with the use of the computer as a supporting tool has undergone rapid evolution during recent years. In general, the search for the best software for supporting teaching could be an even greater challenge in public schools in developing countries than in private schools or schools in more wealthy countries, since it can be expected that those public schools will not be able to purchase licences for commercial software.

In Mexico there are a wide range of educational institutions at high-school level. In the particular case of the Colegio de Ciencias y Humanidades (CCH), one of the innovations proposed by this institution relates to the teaching of Statistics. A high school student entering the third year can choose one or two subjects related to mathematics from among three options: Differential & Integral Calculus, Cybernetics & Computer Science, and Statistics & Probability. Students choosing Statistics take a course whose topics range from descriptive statistics to confidence intervals, hypothesis testing, regression, and correlation. In this context, the use of software for teaching Statistics is very important.

On the other hand, the CCH has five different campuses, with a population of about 12,000 students each, which implies an enormous need for computers; in one particular campus there could be about 400 computers available on which the students can work in a class guided by a teacher and supported with some software. But software remains a problem, since it is not possible for the CCH to obtain 400 licenses for different software applications for different subjects. So, one alternative is to use free software. There are different free software applications available on the Internet for teaching Statistics topics, one of them being Winstats, which is available at the URL http://math.exeter.edu/rparris/. As stated on the website, "Winstats provides access to scatter plots, curve fitting, histograms, statistical data, and standard theoretical probability distributions. It performs many statistical tests and calculates confidence intervals. It simulates dealing cards, rolling dice, sampling candy, taking random walks, and tossing darts, needles and coins. There are two least-squares demos and a confidence-interval demo" (Parris, 2011). There are actually four demos: correlation, least-squares, normal samples, and confidence intervals. It's worth pointing that even when Winstats is available in the Internet only, once it has been downloaded, there's no

need to be logged in to the web in order to work with it; furthermore, since Winstats doesn't use so many memory –it's a file of about 1.55 Mb- it's possible to have it not just installed in a computer, but in a USB memory unit also.

This paper proposes a didactical sequence (DS) for teaching regression and correlation using Winstats by exploring the demos of least-squares and correlation from the perspective of Didactical Situations Theory (Brousseau, 1996).

TEACHING REGRESSION AND CORRELATION

Both demos in Winstats for teaching correlation (Correlation) and regression (Least-Squares) are based on graphic exploration, allowing students to change the direction and intensity of correlation between independent and dependent variables as well as joint variation, aspects which have been identified by Estepa and Cobo (2003) as difficulties in teaching these topics and which are therefore important to work with. Lane, Anderson, and Kellam (1985) affirm that estimation of correlation is more accurate when data are presented in a graphic rather than in a table.

Cobb, McClain, and Gravemeijer (2003) and Konold (2002) suggest that the use of technology may help students to make more accurate judgements of covariation. Zieffler (2008) stated that to help students understand and apply covariational reasoning, the use of "particular software that provides flexible, dynamic capabilities and multiple, linked representations – allowing students to display, manipulate and analyze data" is important (p. 295).

DIDACTICAL SEQUENCE

The didactical sequence starts with the steps for downloading Winstats, and it's intended for the students to work in pairs or even larger groups, up to groups of five. Students receive a form with several questions to explore the demo Least-Squares without any help from the teacher; they are guided by the DS in how to use the software. This demo initially shows a line in a coordinate system and two points on it; students may draw points along the system, and the demo will show the square constructed over the perpendicular distance to the line (Figure 1) as well as a bigger square as the sum of all those squares' distances in the lower right corner (Figure 2).

By dragging the points the students can experiment with how to reduce the sum of squares to the minimum (Figure 3). In this same demo it is also possible to obtain the correlation coefficient for the plotted points, so the students may drag those points to produce different shapes of the scatter plot and different values of the correlation coefficient, which is expected to allow them to understand the concept of correlation coefficient and the idea of what that least squares are.

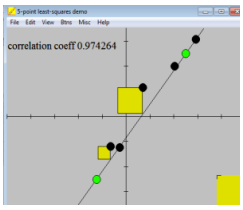

Figure 1

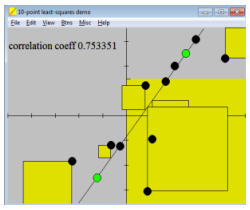

Figure 2

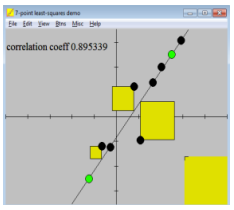

Figure 3

In the second demo, Correlation, a coordinate system is shown, and it is possible for the students to draw and drag as many points as they like as well to obtain the correlation coefficient. The students are asked to start with just two points and later to use a larger number of points, trying to keep a linear distribution, in order to explore the different values of the correlation coefficient; the next step is to drag the points to test their initial ideas (Figure 4). In this demo it is also possible

to show the regression line and its equation, so students may explore the effects of having the points in approximately linear shapes (Figure 5) with both positive and negative slopes and to explore the effects of those shapes on the equation and on the correlation coefficient (Figure 6).

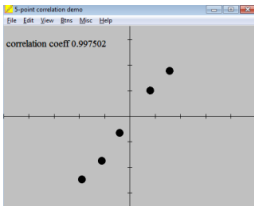

Figure 4

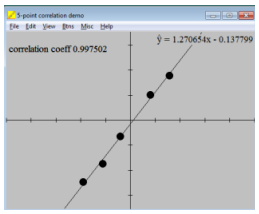

Figure 5

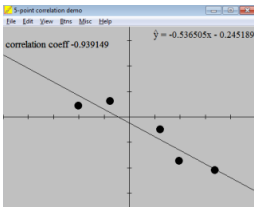

Figure 6

While working with the demos, the students' ideas are discussed to institutionalize all the concepts involved in least squares, the correlation coefficient, and the regression line.

PRACTICAL APLICATION

The final stage of the DS consists in a small report that the students have to write about what they have observed by using both demos, covering what the correlation coefficient, the regression line, and its equation may be and what information they give about the data constructed by the experimentation (all the points).

In general, in the DS's solved and handed by the students, they showed that they understood that the value of the correlation coefficient varies from –1 to 1, showing an almost linear distribution of the experimental points when it is close to –1 or 1. They also realized that the sign of this value is directly related with the sign of the slope in the equation of the regression line and reported that the regression line is the line that best shows how their points would behave if there is any variation. They also reported that the smaller the sum of the squares of the differences, the better the fit of the regression line. All these concepts are basic for understanding the topics, which will help the students to work more efficiently with predictions and inferential aspects involved in this subject.

IMPLICATIONS FOR TEACHING

The topic of regression and correlation is traditionally complicated, even at major level, not only because of all the calculations involved but also because of the difficulty of the concepts. It can be expected then that these problems may be even greater at high-school level. By using the computer, calculations can be performed to enhance acquisition of concepts, and the teacher may focus on concepts in an easier way than the traditional way of teaching these topics.

The proposed DS seems to work properly with high-school students, since as stated above they reported some evidences of learning the concepts related with regression and correlation. Bearing in mind that there is no need to purchase a software licence, this DS could be a good option for teachers working in developing countries for dealing with these topics.

REFERENCES

Brousseau, G. (1996) Fundamentos e Métodos da Didáctica da Matemática. In: Brun, J. (org.). *Didáctica das matemáticas*. Lisboa: Instituto Piaget.

Cobb, P.; McClain, K.; Gravemeijer, K. (2003) Learning about statistical covariation. *Cognition and Instruction*, 21, 1–78.

Estepa, A.; Cobo, F.T.S. (2003) Evaluación de la comprensión de la correlación y regresión a partir de la resolució de problemas. *Statistics Education Research Journal*, 2(1), 54–68.

Konold, C. (2002) Alternatives to scatterplots. In: Phillips, B. (ed), *Proceedings of the Sixth International Conference on Teaching Statistics*. Cape Town, South Africa.

Lane, D.M.; Anderson, C.A.; Kellam, K.L. (1985) Judging the relatedness of variables: The psychophysics of covariation detection. *Journal of Experimental Psychology: Human Perception and Performance*, 11(10), 640–649.

Parris, R. *Software Winstats*.(2011) Phillips Exeter Academy. Mathematics Department. http://math.exeter.edu/rparris/. Accessed February 14th, 2012.

Zeiffler, A. (2008). Developing student´s statistical reasoning*:.Learning to reason about covariation* In: Garfield, J.; BenZvi, D. (eds.) pp. 289–308, Springer.