From Data Analysis to Inference: A Research Project on Students' Understanding of The Normal Distribution

Carmen Batanero
Universidad de Granada, Departamento de Didáctica de la Matemática
Campus de Cartuja,
18071 Granada, Spain
batanero@goliat.ugr.es

Liliana Tauber and Roberto Meyer
Universidad del Litoral, Departamento de Matemáticas
Ciudad Universitaria
3000 Santa Fe, Argentina
rmeyer@fce.unl.edu.ar

1. Background

This research continues our previous work on the teaching of statistics with the use of computers (Batanero, Estepa, & Godino, 199,; Batanero, Godino, & Estepa, 1998). This research is receiving a considerable interest, as it is shown in Garfield, & Burrill, (1997), and Biehler (1997), who pointed out that computers provide powerful calculation and graphical resources that serve to enlarge the meaning of the concepts we show to the students. We have focused on the normal distribution, whose relevance in statistics is pointed out by Wilensky (1997), who suggest its relevance in the understanding of statistics models implicit in scientific research; its role in acquiring a coherent knowledge of randomness; the interface it establish between probability and statistics and the possibilities of serving to guide students towards significant changes in their learning experiences.

2. Analysis of a teaching experience

The experiences have been carried out in a free- choice course of *Data Analysis and its Teaching*, with 9 credits length, at the Faculty of Education, University of Granada. A total of 60 students took the course in 1997-98 and 78 in 1998-99. Most of them where majoring in Education, though there were also students from Psychology, Psychopedagogy, Social workers, Engineering and Business. The students' statistical knowledge was very varied, although in general they had never being introduced to the use of the statistical software, or to the practical aspects of the application of statistics.

Students had been given a written summary of the theoretical content, and they could also study the topic with the help of a basic book. The theoretical lessons were mainly devoted to solving exercises that required an intuitive introduction to normal distributions, as a function whose graph approaches the frequency histogram for some statistical variables, when we progressively increase the number of observations. Interpretative activities and applications to real problems were also emphasised. The practical lessons were carried out in a computer lab, where the students worked in pairs with the statistical package Statgraphics. Other didactical tools were the data files and activities proposed. The option of DESCRIPTIVE ANALYSIS for one numerical variable, which the student already knew, was used. They were also introduced to a new option of FITTING DISTRIBUTIONS and to graphical representations and use of tables of normal distributions (PLOT option; probability distributions). The lecturer provided some data files, which contained variables that could be well approximated by a normal distribution and others where this was not possible, because of their asymmetry, for being qualitative variables, or because of the fact that their probabilities in given intervals were markedly different from those given by normal distributions.

3. Assessment and discussion

At the end of the course and as part of the final assessment, we asked the students to analyse a new data file to analyse that included qualitative and quantitative, discreet and continuous variables. One question consisted of finding a variable in the new data file that could be well approximated by a normal distribution and another where this approximation was not adequate. We also asked the students to justify the criteria followed and Statgraphics procedures used to select the normal and non normal variables. Each students worked alone with the program StatGraphics and they were free to solve the problem using the diverse tools they knew.

The students' solutions were recorded in the statfolios and afterwards were transcribed and classified in the following way: a) *Graphical procedures*: The students could represent the variable distribution using histograms, frequency polygons, cumulative curves, box-plots, stem and leaf, or density functions. A more specific representation for this problem are the symmetry plot or the normal probability plot. b) *Numerical summaries:* The students could compute the mean, median and mode, and compare their relative position; from the typical deviation they check the percentage of cases in the intervals $(\mu-k\sigma, \mu + k\sigma)$, the kurtosis and skewness coefficients could be compared with those of the normal distribution. c) *Computing areas:* Other analytical tools provided by StatGraphics, are: tables of areas under the normal curve and tables of critical values, where the area under the curve is smaller than a given probability.

Among the most significant errors we found students who considered a qualitative variable to be normal, and others who considered normal those distributions with little dispersion. A significant proportion of students used only graphs to assess normality, without reference to areas or critical value. More detailed results will be given in the presentation.

Acknowledgement: This work is part of the project PB96-1411 (DGES. MEC, Madrid).

REFERENCES

Batanero, C., Estepa, A., & Godino, J. D. (1997). Evolution of students' understanding of statistical association in a computer based teaching environment. In J.B. Garfield, & G. Burrill (Eds.), *Research on the Role of Technology in Teaching and Learning Statistics* (pp. 191-205). Voorburg, The Netherlands: International Statistical Institute

Batanero, C., Godino, J. D.. & Estepa. A. (1998). Building the meaning of association through data analysis activities. Research Forum. In A. Olivierr, & K. Newstead (Eds.), *Proceedings of the 22 Conference of the International Group for the Psychology of Mathematics Education, (v. 1, pp. 221-236).* University of Stellembosch, South Africa.

Biehler, R. (1997). Software for learning and for doing statistics. *International Statistical Review*, 65(2), 167-190.

Garfield, J.B., & Burrill, G. (1997) (Eds.), *Research on the role of technology in teaching and learning statistics. Proceeding of the IASE 1996 Roundtable Conference.* Voorburg, The Netherlands: International Statistical Institute.

Wilensky, U. (1997). What is normal anyway? therapy for epistemological anxiety. *Educational Studies in Mathematics*, 33, 171-202

RÉSUMÉ

Nous décrivons un projet de recherche sur l'enseignement de la distribution gaussienne dans les premiers années universitaires à l'aide du logiciel Statgraphics. Nous présentons aussi les premiers résultats obtenus.