

THE CALCULADORA ESTATÍSTICA FOR PRESENT AND DISTANCE LEARNING

Renata G. S. Gomes, Liliane dos S. Machado and Ronei M. de Moraes
renatagrigorio@yahoo.com.br, *liliane@di.ufpb.br*, *ronei@de.ufpb.br*
Federal University of Paraíba
Brazil

Discussions about distance learning are not a novelty. In the last years, several systems to support learning were proposed and developed, some of them based on the Internet. For statistics teaching, as in other areas, there are many systems for distance learning in the web that only present a digital version of the traditional classes' content. Besides of the information access, these systems do not present a function to help the student in his doubts or a function to teach step-by-step for some concepts. Systems with this approach only present the simulation of some situations, or only detect mistakes, or refuse the operation and provide information about the user mistakes. In this paper we present a system to support statistics teaching that joins all of these features: help students by pointing out his mistakes and by teaching the correct steps. The system can be used for present or distance learning and it is named Calculadora Estatística (Statistics Calculator).

INTRODUCTION

According to Moore (1993), discussions about distance learning are not a novelty. But when we compare the distance learning projects along the time, many of the differences among them are related to technologic advances. In the last years occurred relevant advances related to the use of learning systems to support teaching, mainly in the based on Internet (Silveira, 2002) (Moraes & Zuffo, 2001). A learning system is a set of programs whose goal is to transmit knowledge and teach specific concepts. Almendra (2002) and Moraes & Zuffo (2001) affirm that there are several systems to support learning running over the Internet dedicated to Statistics and other areas.

The system developed by Kugushev (2005), the CyberStat, provides an environment with interactive content presented by simulation and interactive quizzes. By the other side, the approach of Lock (1998) offers a good relation among texts, theoretical material, links and Statistical demonstration software over the Internet. The system developed by Lane (1995), the HyperStat, provides access to several Statistics learning applets with graphic simulation. Besides that, the system also provides theoretical content and exercises. The STEPS (1995) consists on a system to help the knowledge construction based on problem solution.

Systems like the developed by Kugushev (2005), Lock (1998), Lane (1995) and STEPS Consortium (1995) are examples of systems to help students in their knowledge process. But, these systems do not help the student to solve his doubts and they do not even teach the correct steps to solve a problem. They also do not limit student's actions or guide him. In Statistical analyses systems, as the developed by Firth (2003), West and Ogden (1997), Banfield (1999) and Leew (1997), there aren't any type of use limitation or restrictions in user queries. The few systems that provide any kind of use restriction perform parameters and pre-requirements analyses followed by an error message. Many times, these messages do not allow the identification of the real reason of the error. An example of that are the systems developed by Firth (2000) and Lumley (2004). Firth (2000) developed a Statistical analyses system, the QV Calculator, which does not perform statistical calculus of queries identified with errors of data or parameters. The same happens in Lumley (2004) package, but it also verifies all the pre-requirements of the query.

Because there isn't a system that helps the student to identify the real reasons of his errors and also teach the correct steps to solve a problem, we developed the Calculadora Estatística. The Calculadora Estatística is a system dedicated to the present and distance learning. It consists on an intelligent tool composed by an interactive environment, as the developed by Kugushev (2005) and Lane (1995). The Calculadora Estatística guides the user during the correct steps of a statistical analysis by limitations present in each query when a pre-requirement is not satisfied. This way, the user is informed about the reason of the error, what helps and stimulates

him in the knowledge construction, as proposed by the STEPS Consortium (1995). With the Calculadora Estadística the student will be able to solve his doubts, learn and fix the statistical concepts, at the same time that previous doubts in his knowledge will be solved.

METHODOLOGY

This paper presents the Calculadora Estadística system developed for present and distance learning of statistics. The system is based on an architecture that helps the user/student in the learning and memorization of statistics concepts by the verification of the correct steps of a statistics analysis. The main features of this architecture are:

- To allow the execution of statistics calculations only when the requirements are valid;
- To teach, step-by-step the correct sequence of an statistics analysis;
- To detect deficiencies in students knowledge;
- To execute commands over the Internet;
- To provide a modular implementation (allows future expansions)

The pedagogical methodology used in the conception of the architecture follows the Constructivism (Azenha, 2001), based on Piaget studies (Wadsworth, 1995). It has as a base that to learn is to construct one's own knowledge from experiences and from the present knowledge. So, during the conception of the Calculadora Estadística system was implemented a structure to evaluate the user's commands. This evaluation is performed by the verification of requirements or restrictions related with the execution of an operation. The requirements evaluation is related with the student's retained knowledge. It means that, if the student has a consistent knowledge, he will not make errors or misunderstandings when using a specific technique. This methodology stimulates the student to ponder and makes him responsible for his own learning. As a consequence, the student will construct his own knowledge. This approach prevents future wrong uses of statistics techniques and analysis errors. Besides, it stimulates the sedimentation of a precise knowledge.

For the Calculadora Estadística implementation some requirements were defined:

- Use of a high-level programming language, portable to several operating systems and that allows content access through the Web;
- Use of a statistics analysis system that allows expansion;
- Use of a free software with low implementation cost.

So, the system was implemented using the Java language, the R software for statistics calculations and the Rserve to translate the commands through the Internet. Java is an object-oriented high-level programming language that provides resources to create interactive and dynamic content. Besides that, it is portable (Deitel and Deitel, 2003). R is a free software for statistical analysis used for data manipulation, calculations and graphics generation (Venables at al., 1998). Rserve is a kind of translator between R and other languages (Urbanek, 2003a) that allows the use of R without previous initializations (Urbanek, 2003b).

ARCHITECTURE

The Calculadora Estadística communication uses a client/server protocol. It works over the Internet and the user can access the page where the system is hosted. At this moment he will receive an applet that connects him to the server by the TCP/IP protocol. Figure 1 presents the architecture of the Calculadora Estadística.

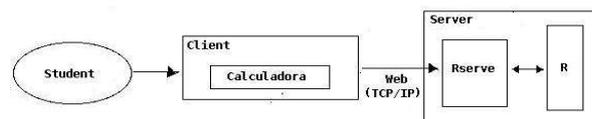


Figure 1. The Calculadora Estadística architecture.

The Client application is started in the user computer by an applet. It is responsible for the execution of requested commands and for the communication with the Server to send that requests. Initially the Client application verifies if the restrictions and requirements of the request are satisfied by the analysis of precedence for each step. If everything is right, the request is sent

to the Server. If the requirements were not satisfied, the Client returns an error message to the user and also returns possible solutions for the error. With this set of actions the student is compelled to remember his actions, to detect and to correct the error.

The Server application is responsible for receiving the Client's requests, processes and sends them through the Rserve for the R software. For that, the Rserve receives a request with a set of parameters and commands written in Java. This set is translated and transmitted (over the Internet) by the Rserve to the Server and sent to the R software. After the R software processes the requests, they are translated again and sent back to the Client.

RESULTS

The first prototype of the Calculadora Estatística system implements three modules. One module is for basic statistics and contains center and spread distribution measures. The other modules allow the generation of graphics and frequency tables. The functions presented in these modules are: mean, median, mode, maximum, minimum, variance, standard deviation, total amplitude and variation coefficient. Other functions offer the generation of histograms, box plots, spread plots, bars plots, line plots and frequency tables for one or two variables. Figure 2 presents the (a) main window of the Calculadora Estatística and the (b) graphics interface of the Calculadora Estatística.



Figure 2. Main window of the Calculadora Estatística (a) and an example of its interface (b).

To use the Calculadora Estatística, the user should access an Internet page where it is hosted. At this moment he will receive a program (the Client application) that can process:

- Capture and sending of data inserted by the user;
- Variables identification and their classification;
- Restrictions evaluation;
- Sending of requests;
- User support with errors information and ways of solution;
- Bibliographical references that can be used by user to understand and solve his errors.

After the user request, the Client application will verify features about:

- Category of the variable that will be used in the analysis;
- Reading of the variables data files;
- Fulfillment of request requirements.

For users not familiarized with computer systems, the graphics interface of the Calculadora Estatística is very intuitive. Figure 3 presents the set of structures visualized by the user when a request is formulated. It is possible to observe that the sequence of steps is simplified and dynamic.

The Calculadora Estatística can be used, as example, to generate a histogram (Figure 2). In this case, the user should initially read and classify the variable data according to its type (quantitative or qualitative levels). If the user chooses categorical data, the system will return an error message to report: "categorical data is not appropriated to generate this kind of graphic". The same happens in other queries that involve statistical measures as variance and standard deviation. In these cases, besides of the necessity of non-categorical data, previous operations are checked as mean and variance calculus.

CONCLUSION

In this paper we present a new architecture for distance learning composed by a system to teach, step-by-step, the correct sequence of a statistics analysis. All the features presented

allow the architecture use for present learning as for distance learning. Using this architecture we developed the Calculadora Estatística system. This system provides a basic statistics module locally executed in the user computer that can perform calculations of center and spread distribution measures and process graphics and tables. The main feature of the system is to help students to solve their doubts and verify their knowledge about the correct sequence of a statistical analysis. It is reached by the use of requirements evaluation that can block the execution of invalid operations.

The benefits provided by the Calculadora Estatística are:

- Aid in the absorption of knowledge in an efficient and interactive way;
- Correction of deficiencies coming from a defective learning;
- Fixation of the correct execution steps of a technique;
- Aid the identification and the correction of mistakes;
- Can be used locally or at distance.

The system is a novelty due to its intelligent module capable to teach the correct steps of an analysis. Until this moment, a distance learning system with this kind of treatment was not found in the literature.

ACKNOWLEDGMENTS

Renata G. S. Gomes is a scientific initiation student supported by PIBIC/CNPq.

REFERENCES

- Almendra, C. C., Vidal, C. A., Serra, A. B., Júnior, A. J. M. L. & Santos, E. M. (2002). Integration of Web Virtual Reality Environments to Web Learning Environments: A Case-Study. *Proceedings of the V Symposium on Virtual Reality*. Fortaleza, Brazil. (In Portuguese)
- Azenha, M.G. (2001) *Construtivismo: De Piaget a Emilia Ferreiro*. Ed. Atica. 7ª edição.
- Banfield, J. (1999). Rweb: Web-based Statistical Analysis, *Journal Software Statistical*, 4(1).
- Deitel, H. M. & Deitel, P.J. (2003). *Java – How to Program*. Bookman.
- Firth, D. (2000). QV Calculator: Quasi-variances in Xlisp-Stat and on the Web, *Journal Software Statistical*, 5(4).
- Firth, D. (2003). CGI with R: Facilities for processing web forms using R, *Journal Software Statistical*, 8(10).
- Kugushev, A. (2005). CyberStats. <http://statistics.cyberk.com/splash/>.
- Lane, D. M. (2005). HyperStat Online Textbook. <http://davidmlane.com/hyperstat/index.html>.
- Leeuw, J. (1997). Server-Side Statistics Scripting in PHP, *Journal Software Statistical*, 2(1).
- Lock, R. H. (1998) WWW Resources For Teaching Statistics. *1998 Technology in Statistics Education Conference*.
- Lumley, T. (2004). Analysis of Complex Survey Samples, *Journal Software Statistical*, 9(8).
- Moore, D. S. (1993). A Generation of Statistics Education: An Interview with Frederick Mosteller, *Journal of Statistics Education*, 1(1).
- Moraes, R. M. & Zuffo, J. A. (2001). An Intelligent Tool for Distance Learning Support. *Proc. of CATE 2001* (pp. 241-244). Paper presented at CATE'2001: Banff, Canada.
- Silveira, I. F. & Ferreira, M. A. (2002). Piaget – Um Ambiente de Realidade Virtual Distribuída Aplicada: A Aprendizagem Colaborativa, *Proc. V Symp. on Virtual Reality*. Fortaleza, Brazil.
- STEPS Consortium (1995). STatistical Education through Problem Solving: The STEPS Project. http://www.stats.gla.ac.uk/steps/about_steps.html
- Urbanek, S. (2003a). Rserve – a Fast Way to Provide R functionality to Applications, 2003 *Proc. 3rd Int.Work. on Distributed Statistical Computing*. Paper presented at DSC 2003: Austria.
- Urbanek, S. (2003b). Rserve on line documentation. <http://stats.math.uni-augsburg.de/Rserve>.
- Venables, B., Smith, D., Gentleman, R. & Ihaka, R. (1998). Notes on R. <http://bayes.math.mon-tana.edu/Rweb/Rnotes/R.html>.
- Wadsworth, B.J. (1995) *Piaget's Theory of Cognitive and Affective Development – Foundations of Constructivism*. 5th edition. Allyn & Bacon.
- West, R. W. & Ogden, R. T. (1997). Statistical Analysis with WebStat, A Java applet for the World Wide Web, *Journal of Statistical Software*, 2(3).