

A TOOL FOR BASIC STATISTICS LEARNING ON THE WEB

Renata Grigório Silva Gomes¹, Ronei Marcos de Moraes² and Liliane dos Santos Machado³

Abstract — *This paper presents an architecture for support basic statistics learning on the Web. Because of its features it can also be used to assist traditional learning methods and it can be applied to other similar courses. From that architecture it was implemented a tool for statistics teaching. In a first moment the tool seems to be similar to an ordinary calculator but it presents some restrictions of use that forces the student to correctly apply the operators in a statistical analysis. This way, future mistakes and analysis errors are prevented. The tool is composed by a client application that treats parameters and restrictions and sends queries to a server application. The server runs the computational and statistics methods necessary for the execution of the queries. The tool was developed using the JAVA language and it implements statistical measures of center and spread of distribution.*

Index Terms — *Distance learning, JAVA application, R software, Statistics learning, Teaching by Web.*

INTRODUCTION

The profusion of inadequate statistical analyses in the scientific community demonstrates the lack of attention in the manipulation of research data. Several works published in newspapers evidence the necessity of professionals from the statistics area. Mainly because insufficient analyses and the wrong use of statistical tools can seriously affected researches conclusions. These deficiencies are originated from a faulty learning of statistics in the academic middle, specially because the contents of statistics disciplines can not include all necessary topics of each area. This way, most of the professionals know about the existence of the science, but they don't know deeply enough how it could be really useful [7] [8].

This problematic became worst by the profusion of statistical software capable of complex analyses, but that don't limit or guide user's actions. The several available statistical systems, commercial or academics, are merely programs that compute statistics, not endowed with any intelligence. So, those software can not aid the user in the use of a technique in detriment of other and not even they can guide the user to not use a mistaken technique in his analysis.

A student can not has a teacher always in his side to solve his doubts. Besides that, assistance systems dedicated to statistics teaching are not present in commercial systems, as SAS, SPSS, Statistica, Splus or Minitab, which are capable to make complex analyses [7]. Even recently developed software as R software [11] or statistical analysis systems at distance, as Rweb [1], do not limit or guide the users in their actions, allowing mistaken actions.

The distance learning is not a recent idea in education or in statistics teaching. In the United States, for example, education at the distance by regular mail dates from 1800. In the 20's courses transmitted by radio appeared and in the 30's they started to be transmitted by television. In 60's, courses of statistics at distance were introduced in some universities, standing out the University of Colorado and University of Iowa [6, 10].

The literature contains few examples of use of a device or other tool by the Internet to support the teaching of statistics or even to teach statistics. For example, Oeste and Ogden [9] discuss the use of *Java applets* for execution of interactive demonstrations and tools for the Internet to support teaching of statistics at distance. Leon and Parr [4] discuss about advantages of present learning supported by specific didactical materials by the Web.

This way, once there is no computational systems to guide and assist the use of statistics tools we developed an intelligent system called Statistics Calculator. The Statistics Calculator has as function to help students pointing out their mistakes and guiding them in the use of statistical techniques.

METHODOLOGY

The goal of this work is the development of an architecture for a system that teach students correct sequences of a statistical analysis. The main properties of this architecture are:

- Do not allow the execution of techniques without their pre-requirements have been verified.
- Teach students about correct sequences step-by-step.
- Points out deficiencies in the student's knowledge.
- Have a simple, intuitive and friendly interface.
- Execute commands through Web.

¹ Renata Grigório Silva Gomes, Scientific Initiation Student (PIBIC/CNPq/UFPB). Department of Statistics, Federal University of Paraíba, Cidade Universitária s/n CEP 58.051-900 João Pessoa – PB - Brazil, tel.: +55 83 216-7075, renatagrigoio@yahoo.com.br

² Ronei Marcos de Moraes, Department of Statistics, Federal University of Paraíba, Cidade Universitária s/n CEP 58.051-900 João Pessoa – PB - Brazil, tel.: +55 83 216-7075, ronei@de.ufpb.br

³ Liliane dos Santos Machado, Department of Informatics, Federal University of Paraíba, Cidade Universitária s/n CEP 58.051-900 João Pessoa – PB - Brazil, liliane@di.ufpb.br

- Modular implementation to favor future expansions of the system and code reusability.

With all these features, a system implemented from that architecture is able to help a student at distance to understand and memorize statistical concepts, learn correct phases of statistical analysis and solve some of his deficiencies.

To implement that architecture some pre-requisites must be solved:

- Use of a high level programming language that allows to access contents through Web, with portability to several operational systems.
- Use of a statistical software with good accuracy and reliability, which allows future expansions.
- The programming language used for the implementation and the statistical software should communicate by the Web to assist the distant students' solicitations.
- Preference for free software to do a low cost implementation.

To serve all those pre-requisites, it was chosen the JAVA programming language for implementations on the Web, the statistical software R for statistical calculation and the Rserve for translate solicitation from Web to the R software. In the next sections we present some details about them.

JAVA programming language

JAVA is an object oriented programming language. In the beginning JAVA was projected as a language for embedded electronic systems. However, that project does not works and for years JAVA was restructured to be the language as we know today.

The main characteristics of JAVA language are [3]:

- It is an object oriented programming language.
- It allows distributive programming.
- It is multi-user.
- It is portable by use of JAVA Virtual Machine.
- It is interpreted.
- It is secure.
- It works for Web and can be used for communications between two computer that use JAVA Virtual Machine.

For our purposes the JAVA language serves all the specifications.

R Software

R software is a computational system for statistical analysis. R provides a programming language, a graphical run-time environment and a debugger environment. It can run stored programs as such as it allows access to other

functions. R is as an implementation of the S language, which was developed at Bell Laboratories by Rick Becker, John Chambers and Allan Wilks, and also forms the basis of the S-PLUS systems [11].

R software includes several statistical procedures pre-programmed. They can be used in new programs as libraries call. For example, there are procedures for linear, non-linear and generalized linear models, time series analysis, parametric and nonparametric testing of hypothesis. There are modules for data treatment and presenting too. Besides, other contributed modules can add new features [11].

The statistical functionalities from R software are completely enough for our purposes. As additional benefits, it can be used for future expansions of the system.

Rserve

Rserve is a system that allows others programs to use R software functionalities remotely, without necessity of its initialization. It is very convenient for system like ours, because each user can use different functionalities of R software without users cause problems to each other.

Rserve use the TCP/IP protocol for communication of data. TCP/IP is a set of protocols for transmission of data with control of transmission. Its great advantage is to allow intercommunication with heterogeneous computers, that is, different operational systems.

In our system, Rserve is responsible for establishing the communication between distant user, called client and the R software in the server. It makes the administration and translation of instructions of client's language to the R software and vice-versa. After that, it is also responsible for the sending of these instructions for the software R. The R software executes the instructions and returns the answers to Rserve. The Rserve makes the inverse way of information to the client and waits for new instructions. The blocks diagram of the Figure 1 shows that interactions.

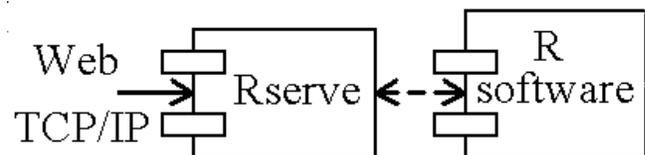


FIGURE. 1

DIAGRAM SHOWS INTERACTION AMONG RSERVE, R SOFTWARE AND WEB.

IMPLEMENTATION

From the architecture, pre-requisites and components presented above we can present a system implemented using all that. This system was called Statistics Calculator and it has two distinct subsystems: a Client and a Server. When a user accesses the Statistics Calculator's Web page, he receives an applet JAVA, which establishes a link by Web,

using TCP/IP protocol, with the Server subsystem. Figure 2 shows the blocks diagram of this implementation, according to the architecture proposed. The Client subsystem is hosted in user PC and Rserve and R software are hosted in Server. The Client subsystem has as function to execute commands solicited by the user using JAVA applet. On the other hand, the Server subsystem receives that request and must process it using Rserve and R software. After that, the required results are sending back to the user.

On the side of Client subsystem, in the first moment, the user can read and send his data to Server. The JAVA applet first verifies data integrity and sends data to Server. If something is wrong, the user receives a message with information about the error and how he can solve it.

After the data was sent, the user can ask for predefined operations over his data. The applet verifies restrictions over that operation in its internal controls and if it is allowed, the applet sends the request to the Server. Otherwise, a message returns to the user with information about the error and how he can solve it. If the user insists in his incorrect request, the Client subsystem can indicate for the user a study reference.

On the side of Server, a request from user by Web is received by Rserve. Rserve translates the JAVA request for the R software and sends it. R software processes the request and the results are sent to Rserve. Rserve must translate the R software results for JAVA and send back to the user by Web. As it can be seen, Rserve is necessary because JAVA and the R software cannot communicate to each other directly.

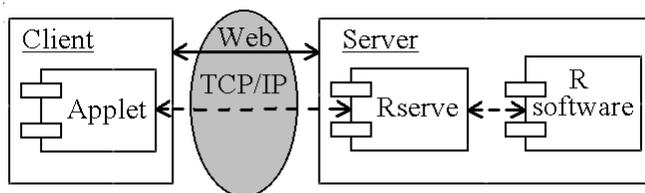


FIGURE. 2

DIAGRAM SHOWS INTERACTION BETWEEN CLIENT AND SERVER BY WEB.

RESULTS

In the first version of the Statistics Calculator, a module for basic statistics with some statistical measures was implemented. The measures presented in this module are: mean, median, mode, maximum, minimum, variance, standard deviation, total amplitude and variation coefficient.

When a user accesses the Statistics Calculator's Web page, he receives an applet JAVA. That applet has a set of methods, which contain components for:

- Reading and sending of data.
- Identifying of all variables and their categories.

- Verifying of data integrity.
- Choice of statistical measures of center and spread of distribution for each variable.
- Verifying of restrictions over each operation.
- Information about errors and how they can be solved.
- Indication of references to the user study about persistent errors.

The graphical interface of the Statistics Calculator is composed by a set of buttons, similar to a common calculator. Bellow, Statistics Calculator window has a message bar, where the messages from the Client and Server system are showed. The results of requests over data are showed in new Java applet windows.

The Client subsystem monitors each action of user and verifies if his request satisfies the restrictions for that request in its files of controls. If restrictions are satisfied, the request is sent to Server. If restrictions are not satisfied, the user receives an error message, with details about it and how he can fix that problem. The main restrictions are:

- Data is not available.
- Category of variable does not allow that measure.
- It is necessary to calculate other measure before the requested measure.

The restrictions are beneficial to the user, because they allow him to learn the correct sequence of a statistical analysis. It will avoid the wrong use of the technique in the future.

For that reason a system constructed over that architecture is important for distance learning. As the system permits its use only in correct sequences, the student can note his deficiencies and solve them.

FINAL CONCLUSIONS

In this paper, we presented a new architecture to teach students at distance about correct sequences of a statistical analysis. Because of its features it can also be used to assist traditional learning methods and it can be applied to other similar courses. Using that architecture, a system was implemented and was called Statistics Calculator. This system provides a basic statistical module as an applet JAVA that runs in PC, with operations over measures of center and spread of distribution.

The main feature of that system is to permit to the student solve his doubts and to verify his knowledge about correct sequence of a statistical analysis. It is done by the use of restriction to the student to access only valid operations.

As future works, we intend to expand the system with new modules to complete and to enlarge the use possibilities of this tool.

REFERENCES

- [1]. Banfield, J. "Statistical Analysis on the Web" Internet site: <http://www.montana.edu/Rweb/>. Access in March, 26, 2004.
- [2]. Coob, G. W., "Reconsidering Statistics Education: A National Science Foundation Conference" *Journal of Statistics Education*, Vol 1, N°1
- [3]. Deitel, H. M.; Deitel, P.J., "Java-Como Programar", Porto Alegre, Bookman, 2001.
- [4]. Leon, R. V. ; Parr, W. C., "Use of Course Home Page in Teaching Statistics.", *The American Statistician*,; Fevereiro, Vol 54, N°1, pp.44-48.
- [5]. Malone, C. J., "Statistics Course Web Sites: Beyond syllabus.html" , *Journal of Statistics Education*, Vol 9, N°2.
- [6]. Moore, D. S., "A Generation of Statistics Education: An Interview with Frederick Mosteller", *Journal of Statistics Education*, 1993, Vol 1, N°1.
- [7]. Moraes, R. M. ; Zuffo, J. A., "Intelligent Systems with Visual Interface Applied to the Basic Teaching of Statistics ", *Anais 2001 International Conference on Artificial Intelligence(IC-AT2001)*, Las Vegas, EUA, Junho , pp 464-468.
- [8]. Moraes, R. M. ; Zuffo, J. A., "An Intelligent Tool for Distance Learning support", *Anais 2001 IASTED International Conference on Computers and Advanced Technology in Education(CATE 2001)*, Banff, Canadá, Junho, pp 241-244.
- [9]. Oeste, R. W.; Ogden, R.T., "Interactive Demonstrations for Statistics Education on the Word Wide web", *Journal of Statistics Education*, 1998, Vol 6, N°3.
- [10]. Stephenson, W. R. "Statistics at a Distance", *Journal of Statistics Education* , Vol 9, N°3
- [11]. Venables, B.; Smith, D.; Gentleman, R. e Ihaka, R. (1998) Notes on R: A Programming Environment for Data Analysis and Graphics} Internet site: <http://bayes.math.montana.edu/Rweb/Rnotes/R.html>. Access in March, 26, 2004.