

WEB ENVIRONMENT FOR LEARNING, EXPERIMENTING AND KNOWLEDGE EVALUATION IN REGRESSION

Cristian Marinoiu, Iuliana Dobre, Laurențiu-Nicușor Pană

ABSTRACT

Regression is a powerful instrument of statistics, largely used for solving various problems in economics, agriculture, sociology, psychology, biology, educational sciences etc. Also, in the last years, many modern technologies in data mining are essentially based on regression concepts.

This paper presents a web-based environment for better understanding and appropriate regression. The users can learn (self-assessment included) and experiment regression concepts and techniques. Moreover s/he gets support for solving real-world problems using regression techniques. This environment helps the user understand what is behind the scene in order to appropriate the power of regression and of its tools. The presentation focuses on the pedagogical benefits obtained from the use of some opportunities provided by modern computing technologies in this work.

KEYWORDS

Web environment, regression, passive learning, active learning, learning tools, self-evaluation, simulation techniques.

INTRODUCTION

This paper is based on the results obtained in (Marinoiu and Dobre, 2001) and (Marinoiu and Dobre, 2003). Both described applications fully use the graphic and simulation abilities that modern computers have in order to improve the learning process. The questionnaires that have been administered to the students prove the students' interest for this new way of approaching concepts from statistics.

The final aim of our synopsis is to obtain an integrated learning and evaluation environment which would be able to allow a better efficiency of the teaching process. In our opinion, this new version represents a decisive step in order to achieve this aim. In the proposed and implemented learning model self-assessment and evaluation seem to be very important factors of the learning process. They are adjusting factors of this process, but in the same time, profound study and consolidation factors of knowledge.

PROPOSED LEARNING MODEL

The proposed learning model is based on the following general schema:

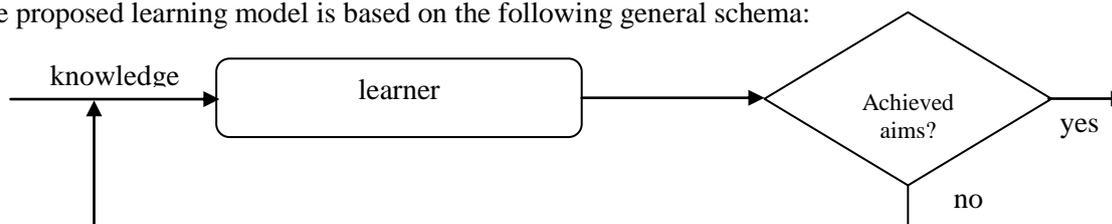


Figure 1. The general schema of the proposed learning model

As shown in the schema, the subject enters a learning loop until the aims are achieved or he gives up. The learning process being extremely complex, it always presumes activities of memorizing, clearing up and profound study of the existing notions. It is obvious that in a learning process these activities always overlap. Typically, one can distinguish two kinds of scenarios which can be viewed as stages in the learning process: passive stage and active stage.

IMPLEMENTATION OF THE PROPOSED MODEL FOR THE REGRESSION COURSE

In this chapter, we present the structure of the regression course and the ways of implementing the above mentioned learning model. The course is divided in several modules as in the example shown below:

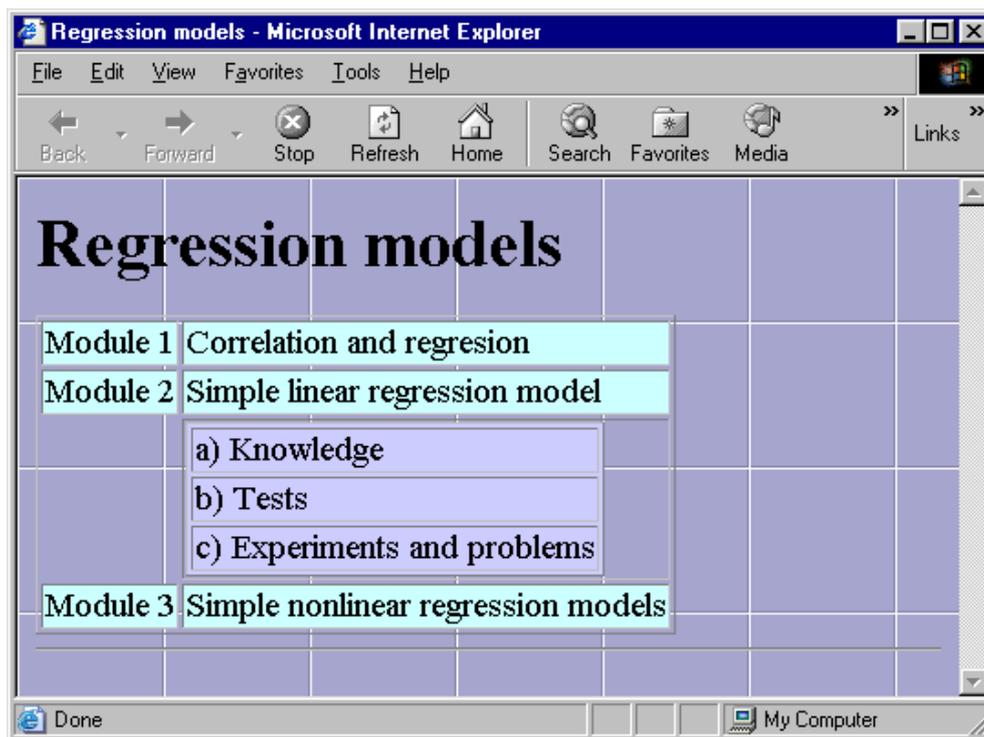


Figure 2. The structure of the regression course

Every course module has a standard structure: *a*) Knowledge; *b*) Tests; *c*) Experiments and problems. Essentially the item *a* implements the passive stage, while the items *b*) and *c*) implement the active stage of the proposed learning model.

Passive learning stage

Selecting the item *a* the subject accesses a web page which contains specified basic knowledge of the respective module. Typically such web pages contain: definitions, propositions, theorems, examples and explanations.

There are at least two facilities by using the computer in this stage: the hyperlinks and Java applets. The hyperlinks allow that the demonstrations of the propositions and the theorems to be hidden in other pages. As a consequence, in this initial stage the subject can avoid technical details and focus on the main results of the present module. The graphics, especially the simulations – practically impossible to be realized without the computer - consolidate the assimilated knowledge by their suggestive power.

Comments

Despite the facilities offered by the computer, this stage consists in a passive examination of the proposed learning material and its aim is to make certain notion and concepts become more familiar.

The aims which adjust the learning loop in this stage (see Figure 1) are defined by the subject itself; therefore, there is not an objective measure of the degree of the assimilation of the knowledge.

Active learning stage

The tests (the item *b* - *Tests*) and the proposed experiments and problems (the item *c* – *Experiments and problems*) make the subject become an active participant in the learning process

Selecting the item *b* the subject must answer to a set of questions randomly selected from an electronic collection of tests. On the basis of the answers, the computer automatically provides the mark. The aim which adjusts the proposed learning process at this stage (see Figure 1) is to obtain a mark bigger than some fixed threshold.

The use of the computer in this stage allows a quick random selection of the questions and the immediate marking.

Comments

The tests are not just an instrument for self-evaluation. It is an active way to learn, because trying to choose the correct answer the subject must revise and clarify some notions and concepts from the course.

Selecting the item *c* the subject accesses web pages where the proposed problems and experiments are randomly selected from an electronic collection. The subject tries to solve the problems and to make the experiments. The results are electronically sent to the tutor. He assesses their correctness and sends the proposed marks and possible comments on the received results.

Comments

Establishing if the subject achieved or not the specified aims, the tutor acts as an adjusting element in this stage of learning model (see Figure 1).

The computer, by its implemented software, represents the essential part, which allows the realization of experiments. These experiments are indispensably connected by the graphical and simulation abilities of the computer in the same way in which a classical experiment is possible if there is an adequate material basis. In our case, the used ‘ingredients’ are not, for example chemical substances, but random number generators, functions that trace random straight lines, regression lines, functions which add or delete points etc. (the so called *instruments*).

Unlike REGREX version used in (Marinoiu and Dobre, 2003) the present implementation brings about a major change in the management and realization of experiments. In this version the stress is not laid on the experiment, but on the instrument that helps in order to realize it. The subject does only receive tailor made experiments. He receives instruments which he may use during the experiment and which are imagined by him. The experiments proposed by the subjects are sent to the tutor, who validates them or not. The validates experiments are added to the collection of experiments and those that are not notified are not destroyed, but separately archived so that they can be consulted by those who want to; although rejected, they might contain potential useful ideas, which were not exploited enough. This new approach has important advantages: it changes the working environment into an open one, permanently subject to improvements proposed by its beneficiaries, to whom this activity represents a genuine creativity exercise. An example of experiment is presented bellow:

Experiment

1. Draw the true regression of Y on X, knowing:
 - (X, Y) is a two-dimensional random variables normally distributed;
 - the expected values and the variances of the random variables X and Y;
 - the correlation coefficient of the random variables X and Y.
2. Generate a sample of size $n=5$ for the two-dimensional random variables (X, Y), normally distributed.
3. Draw the ordinary least squares (OLS) regression line of Y on X.
4. Repeat the steps 2 and 3 for $n=10, 15, 20$ and so on.

5. Describe what you see on the scatterplot and provide necessary explanations.

As a result of making this experiment the subject should notice that while the selection volume n increases, the regression line OLS gets closer and closer to the true regression line $y = \beta_0 + \beta_1 x$. The reason of this proximity lies in the fact that the regression coefficients $\hat{\beta}_0, \hat{\beta}_1$ of the OLS regression line $y = \hat{\beta}_0 + \hat{\beta}_1 x$ are unbiased estimators for the coefficients β_0, β_1 of the true regression line $y = \beta_0 + \beta_1 x$.

Technologies used for building teaching web environment

The teaching environment is built using the actual web technologies. All interfaces (courses, test pages, experiments and problems) with the actors (trainer and learners) that interact with the teaching environment are realised as web pages implemented by means of (D)HTML language. In order to build visual dynamic elements (animated graphic examples, tools used for realising experiments) the Java technology is used by means of applets. An Apache web server sustains all web interfaces. In order to store test libraries, experiments and problems as well as students' profiles a relational database is used, which explores the facilities offered by the database server MySQL. The web server interacts with the database by means of a software interface realised by the scripting language PHP.

Two types of actions may act upon the environment:

1. learning realised by a student supervised by a tutor. The general schema of the learning is the following:

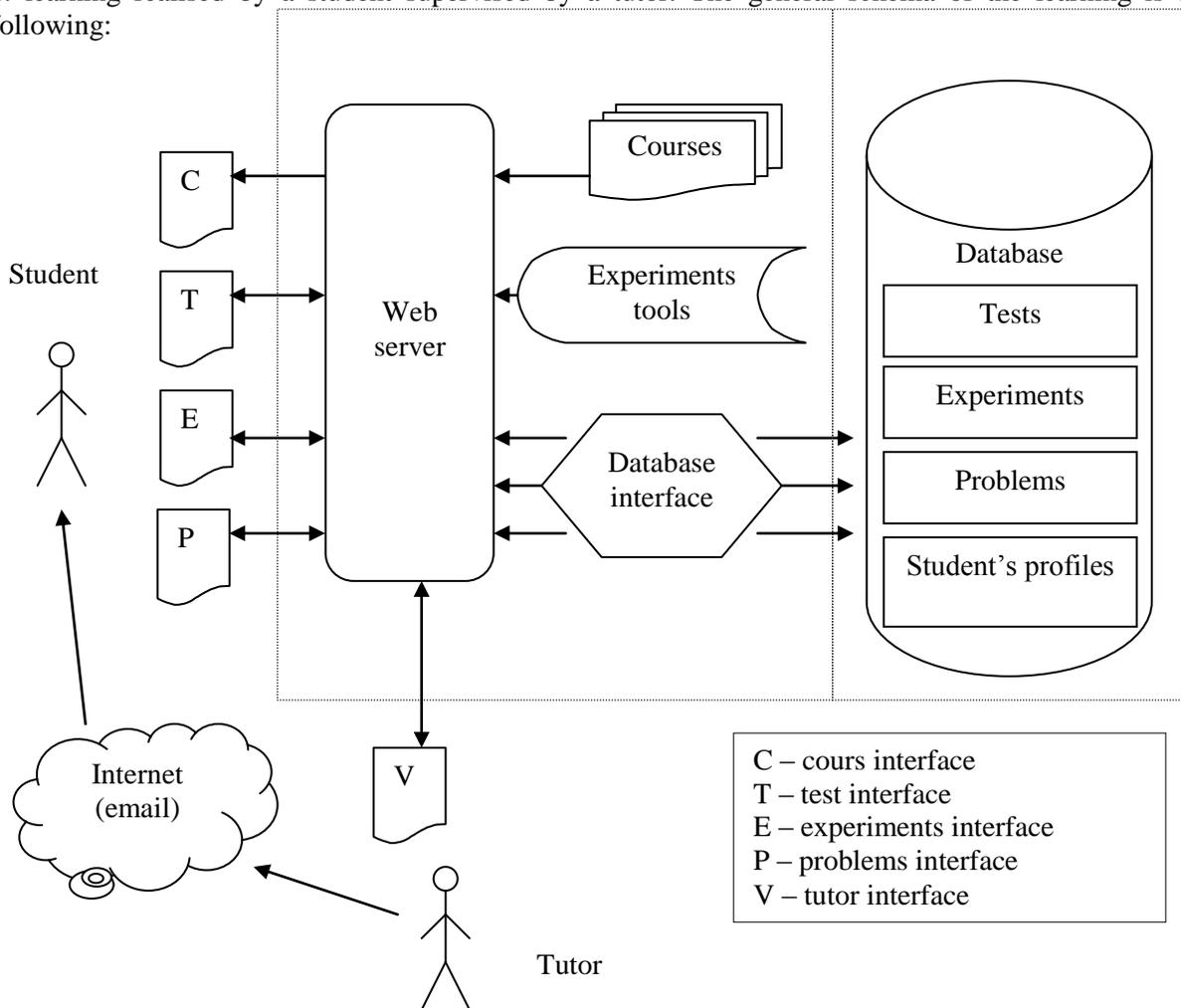


Figure 3. The general schema of the web learning environment

2. upgrade of the tests library, experiments and problems realised by the trainer or sometimes by the student himself. If the learner wants to propose new way of assessment he/she may do it, but his/her proposals are accepted only if a trainer has advised them. The general schema of the database upgrade is the following:

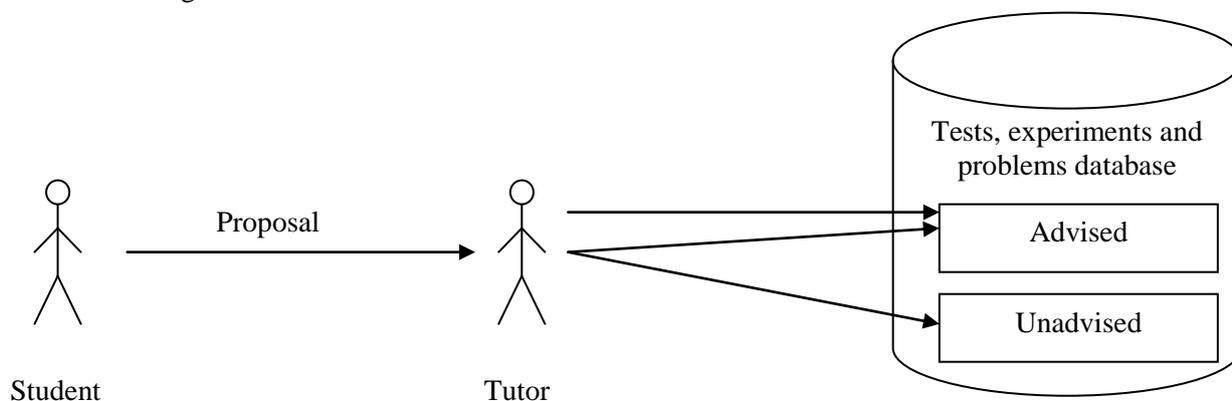


Figure 4. The schema of the upgrade of the database (tests, problems and experiments)

CONCLUSIONS AND FUTURE WORK

The proposed teaching environment gives the learner:

- the opportunity of realising statistical experiments in the same manner in which experiments in the field of natural sciences are realised (realising experiments in statistics is quite impossible without the help of the computers);
- the possibility of getting familiar with the degree of assimilation and understanding of knowledge, using as feedback the answers from the tests and the tutor's remarks;
- the opportunity to propose and solve problems and experiments under the tutor's supervision, increasing creativity and the learner's initiative spirit.

The improvement of the teaching module package will take into account at least two aspects:

- the enlargement of the approached statistics subjects
- the development of new tools (instruments) in order to enlarge experimenting possibilities and in this respect, training the students from graduate years of the mathematics – informatics specialization of the Petroleum-Gas University.

REFERENCES

Marinoiu, C., Dobre, I. (2001). Teaching and learning with Regress, Proceedings of 5th the International Conference Computer Based Learning In Science, Brno, Czech Republic, paper C5.

Marinoiu, C., Dobre, I. (2003). REGREX: a Java-based package for teaching and active learning in regression, Proceedings of 6th the International Conference Computer Based Learning In Science, Nicosia, Cyprus, 281-288.

Cristian Marinoiu
Associate Professor
Department of Informatics
Petroleum-Gas University of Ploiesti
39, Bucuresti bd.,
100680, Ploiesti,
Romania
Email: marinoiu_c@yahoo.com

Iuliana Dobre
Lecturer
Department of Informatics
Petroleum-Gas University of Ploiesti
39, Bucuresti bd.,
100680, Ploiesti,
Romania
Email: iulianadobre@yahoo.com

Laurențiu-Nicușor Pană
Assistant
Department of Informatics
Petroleum-Gas University of Ploiesti
39, Bucuresti bd.,
100680, Ploiesti,
Romania
Email: laurpana@upg-ploiesti.ro