Training Medical Students in Biostatistics by Applying Computer Simulations to Perform Scientific Research through Processing True (Valid) Data from Medical Practice

Nedialka Petrova1, Ralitsa Raycheva2, Ani Kevorkyan3, Jordanka Stoilova4
(Department of Social Medicine and Public Health, Faculty of Public Health, Medical University Plovdiv, Bulgaria)

Abstract: Aim: The purpose of using simulations in Biostatistics education is to give students basic knowledge of the principles of experimental data collection, statistical processing and analyses of results. These are activities where knowledge is acquired through trial and error. The simulation as a didactic tool directly influences on the students’ minds and involve them in the research. Students play the role of researchers, through the imitation of actual situations in order to describe and model complex medical trails. During the practical exercises, in order to create the closest to reality research model, students use real medical records databases, materials from websites, books and further reading.

The goals of implementing this training approach could be summarized in the following aspects: to encourage creativity, to develop associative thinking and intuition in acquiring statistical knowledge. This is a new concept of human capabilities and its “smart extensions” such as computers and communication technologies.

Methods: Students develop their own questionnaire — an electronic patient record as a dynamic scheme. In the next step the young “explorers” select significant features for the cases studied as an input data: gender, age, social status, suspected diagnosis, co-morbidities, type of therapy, complications, condition after treatment. Students apply different computer techniques such as sorting, filtering by various criteria and etc. From a demo copy the researchers extract a representative sample, which contains factorial and consequential variables.

Results: To justify the effectiveness of incorporating simulations as teaching tools analyses of the test results from the final monitoring and evaluation of knowledge were performed. We used relative criterion to examine the level of cognitive and achievement abilities. Confidence interval for the average students’ marks is respectively 13.28 ± 0.83 for the experimental group and 4.6±0.70 for control group. The proportion of students from the experimental group who have learners’ rate/coefficient of learners 0.6÷0.79 is 25% ± 5.75%, for the control group is 4% ± 2.5%. The difference was statistically significant – U = 8.34.01, p < 0.01.

Key words: biostatistics, simulation, medical education

Nedialka Petrova, Ph.D., Associated Professor, Department of Social Medicine and Public Health, Medical University of Plovdiv; research areas: biostatistics. E-mail: dirdriem@gmail.com.

Ralitsa Dimitrova Raycheva, Masters of Economics, Assistant Professor, Department of Social Medicine and Public Health, Medical University of Plovdiv; research areas: Biostatistics, Health Economics.

Ani Kevork Kevorkyan, M.D., Ph.D., Assistant Professor, Department of Epidemiology and Medicine of disastrous events, Medical University of Plovdiv; research areas: epidemiology of infectious diseases.

Jordanka Dimitrova Stoilova, M.D., Ph.D., Professor, Department of Epidemiology and Medicine of disastrous events, Medical University of Plovdiv; research areas: epidemiology of infectious diseases.
1. Introduction

In search of effective methods for teaching students in biostatistics an educational simulation of medical scientific research was applied. The Educational Simulation as a didactic tool provides direct influence on the students’ sensorium and involves them in the scientific research rapidly. This is a new concept of human abilities and their “smart extensions” such as computers and communication technologies (Kirova L., 2001; 2001).

In our study we have reviewed the criteria and mechanisms of the computer simulation we used in practical lessons to address the stages of medical scientific research (Petrova N., Dimitrov I. & Todorova M., 2004; 2005).

Learning through these deductive tools facilitates students in acquiring substantial knowledge about statistical analyzes and offers opportunities for:

- using specialized software for information processing;
- proofing existing patterns, by employing examples from actual databases of medical practice.
- obtaining routine and confidence in making independent decisions in different situations during the scientific research.

Meeting academic goals in the application of computer simulation in training requires careful selection of:

- the functions to simulate;
- the theoretical basis to ensure basic knowledge of statistical analysis;
- the appropriate training stage to be used in the methodology;
- the precise combinations of scientific objectives and practical skills obtained.

In our practical education our students practice computer simulation in the last stage of the teaching process. The students replicate certain situations selected from the key stages for conducting scientific research:

- determination of the study design;
- development of document for collecting information (short modified questionnaire and respective reduction of the variables in the database);
- test of scientific hypothesis and results analysis in decision making;
- demonstration of the association between medical events and the significance of statistical dependencies;
- definition of the trend parameters for a medical phenomenon.

Simulation as a teaching tool for modeling reality, it provides students with productive learning time, accelerates and stimulates mental activity. Clark Aldrich is an author and consultant on e-learning by developing strategies for learning by doing. He led an international team to create a virtual learning system, using simulations, computer games, and pedagogy in e-learning. Clark Aldrich outlines six criteria that are crucial not only for simulation but also for any successful educational activities.

Three of the criteria are focused on the description of the curriculum, and the other three secure (delivery) items.

The main criteria for content are:

1. Linen content.
2. Systematic content.
3. Cyclical content.

Additional criteria for the delivery of items are:

4. using elements of simulation as a model of reality.
5. using game elements that provide familiar and fun interactions.
(6) pedagogy (teaching) elements that provide a productive time spent by students.

In terms of educational objective linear content and teaching tools are dominant and powerful. But skilful combination of six criteria guarantees the transformations in the philosophy of actual educational practices (Aldrich C., 2004).

The imitation of the medical scientific study met these 6 criteria of educational simulation. By linear content students are prepared for algorithmic thinking. Each successive portion used previous knowledge, in order to build up logical components:

The algorithm for checking the scientific assumption includes some of the steps for calculating the range of the confidence interval for a true population value.

In practice, the implementation of this criterion appears as follows:

Problem: Calculate the range of 95% CI for key statistical parameter of the population.

A. The algorithm’s initial data is the frequency distribution of the cases in the sample allocated by the values of the tested variables.

In order to group the individuals in the database the students execute computer commands — filtering by “AND” logic (Boolean algebra). In sequential progression the students perform the following commands — Data/Filter/AutoFilter.

1. The criterion for one-dimensional distribution is one (e.g. growth). A simple 2x2 cross tabulation requires two criteria (e.g., gender and nosological unit).

2. The next step involves the transformation of tabular data to graphic construction — the bell shape is a clear evidence for normal Gauss Laplace frequency distribution, i.e., descriptive statistics is possible to be applied during processing the sample’s data.

3. Researched contingent distribution by age for average age to be calculated.

4. Calculation of the range of the CI for a statistical parameter of the population based on the sample’s data.

5. Researched contingent alignment by gender, diagnosis — compilation of a complex table and application of parametric and nonparametric hypothesis tests in order association between the phenomena to be demonstrated. Null Hypothesis — Gender is not a factor for any specific disease.

These are the key steps of the cyclic criterion of the computer simulation, incorporated to obtain the final results — routine skills for working in dialog rhythm and combination of statistical knowledge with computer skills.

Consequently, the accumulated knowledge on the main course — Biostatistics, helps students to acquire specific skills for managing the training content. This enables trainees to set their priorities and enlighten them how to achieve these priorities in actual educational practice.

Statistical processing of input data is performed by implementing exact actions in EXCEL virtual environment. While modeling the reality by virtue of simulation elements students apply different analyzes upon same data sets. To discover and prove relationships and dependencies between medical phenomena, through practical examples allows the learners dynamic and flexible utilization of scientific knowledge from three disciplines — medicine, information technology and Biostatistics.

The application of other elements assists the students to create simulation model for medical phenomenon’s dynamic changes. For this purpose they must extract the method of least squares from the basic statistical knowledge they have already mastered and to create the model of the trend. The improved computer skills facilitate the trainees in solving the two equation system with two unknowns — variables of the linear model.
The process of the study simulation is permanently accompanied by various game elements, as students work in small groups. They change their roles and convey knowledge between them, while explaining problems to one another; they are able to elaborate situations with varying complexity such as:

- changing the criteria (variables) while grouping the cases – by immune response, by the length of the service, by gender, etc.
- compete and motivate each other to achieve the speed and apply a different logic. Each of the students seeks to perform the analysis; this reduces the need for constant teacher intervention and increases the pleasure and satisfaction from their experience.

Activities of cyclic criterion are also used for:

1. Demonstrating the tendency in the dynamics and obtain the linear model of the phenomenon by using registered formulas for solving two equation system with two variables of the model (smoothing and curve fitting techniques of the method of least squares).
2. Using built-in functions of the application to automatically obtain the linear equation of the model and the graphical image base on provided time series data.

The presence of demo materials, specialized literature, medical data bases and Internet access create opportunities for improving end-users' competence and developing exploration-based learning.

The learners experience pioneer spirit, experiment with concrete examples, are active and focused on constructing model’s separate parts from the overall data, use interdisciplinary knowledge, if necessary, display the results graphically, through computer simulation of a real medical object.

During the simulation students cultivate their responsibility as they work in teams. Trainees interact with one another; discuss successful and unsuccessful algorithms, which are mandatory backed up. Learners discuss their achievements, mistakes and the possible ways to adjust them, which further motivates them. They implement training activities to accelerate the process of memorization and perform steps to improve individually. Thus the acquisition of knowledge and skills becomes more interactive.

The sense of spacious imagination and fantasy of the young “researchers” is due to shortening the time for data processing by applying different designs for the experiment. Students explore the phenomenon repeatedly by using the same design on different types of variables; they state different formulation of scientific hypotheses and thus perform multiple algorithms for testing them and making the final decisions.

Each new experiment increases the realized students‘ necessity to accumulate theoretical knowledge, the lack of which caused the previous failure. Errors result in perceived need for increased knowledge and improve skills in all three combined areas — individual work with computers, use of medical problems and Biostatistician analyzes applied in the actual implementation of a virtual experiment.

Virtual electronic library offers diverse examples from university hospital cases and university theoretical departments scientific studies and thus provides the opportunity for the different groups to model examples from various fields of medical practice — neurology, psychiatry, obstetrics and gynecology, surgery, epidemiology, etc.

According to generally accepted didactic requirements each practical lesson in the students’ program includes:

- Identification and discussion of the problem;
- Theoretical explanations and illustrations by the teacher;
- Assignment’s modeling by the students;
- Experiment’s conduction and outcomes’ discussion by the trainees.
The first two steps are failed in protocol, while the next, the experimental phases are stored in virtual notebooks and on computer hard disks—folder with student’s name and group.

The content of the practical lesson includes:

AIM: Simulation of scientific research:

1. The risk for occupational exposure in three categories of medical staff.
2. Hepatitis B incidence trend line modeling.

2. Accomplishment of the Objectives

1. The Lecturer Perform A Preliminary Explanations and Illustrations of the Chosen Medical Problem
   a. From 1970 to 1996 the Center for Disease Control in the U.S. (CDC) issued six consecutive manual to prevent transmission of infectious agents in healthcare units. The standard prevention measures, introduced by CDC in 1996, aim all patients with anticipated contact with blood, body fluids, mucous membranes and non-intact skin. The application of the measures focuses on reducing the risk of microorganisms’ transmission of and remains one of the best practices for protection of medical staff.

   In 2009 Dr. Ani Kevorkian performed anonymous survey among healthcare employees in University Hospital “St. George”, Plovdiv, Bulgaria (Kevorkyan A., Teoharov P., Petrova N., Stoilova J., Plachkova A., Angelova N., 2011; Kevorkyan A., Petrova N., Angelova N., 2012). The research includes 680 surveyed objects from the hospital personnel: 186 doctors, 330 nurses and 164 attendants. The developed questionnaire contains 14 questions, grouped into three groups. First group of questions were related to the length of service in the hospital and if immunization with hepatitis B vaccine was performed. A second group of questions were related to risk exposures incurred with biological material at work within one calendar year. A third group of questions were related to professional staff’s behavior after the incident occurrence.

   b. The Department of Social Medicine Electronic Library’s database contains researches of Prof. J. Stoilova, which refer to all cases of hepatitis A in Plovdiv region for ten year period, from 1990 to 2000 (Stoilova I., Popivanova N. & Petrova N., 2002; Rakadjieva T., Stoilovaq J. & Petrov A., 2002). The main goal of the research project is to model the dynamics of the phenomenon and characterize the trend of the event — the arithmetic mean of the cases and the slope of the trend line.

2. In connection with the “A” problem the lecturer demonstrates specific computer techniques.

3. Practical Session’s Key Steps

The series of actions that require the preparation of a simulation game begin with selecting the right situation. This largely depends on the opportunities provided by the curriculum material, the school system and the skills of the students. These settings give the framework, where to look for the theme and the adequate for it role-playing situation that gives the richest communication solutions. Leading for this choice are the results from the analysis of current and future interests and needs of the group trained (Rodriguez R. J. & White R. H., 1983).

The scientific research’s computer simulation is a creative process of consideration, focused on particular problem and on its omnifarious facets — the occupational exposure risk. Students are aware that correspondence between the aim and the objectives is sine qua non. The first stage of the simulation has staging nature. The trainees identify the object of the study, specify the concepts, and analyze the possible outcomes after hypothesis testing, the associations, the trends, and potential relations. They combine the emerging issues that are the active
component of research with essential role in deduction and decision making.

The lecturer explains the necessity of applying the definition of the CDC relating to the risk for occupational exposure:

Definition: Occupational Exposure Risk occurs in the workplace through percutaneous inoculation (needle stick injury, cut with a sharp instrument) or through contact of mucous membranes or non-intact skin (with cracks, bruising, inflammation, etc.), with blood, tissue, or other body fluids that are potentially infectious (Communicable Disease Control Manual, 2003).

The study objects were divided into three groups, according to the “length of service”:

(1) Group I – length of service up to 5 years (n = 119)
(2) Group II – length of service from 5 to 10 years (n = 103)
(3) Group III – length of service more than 10 years (n = 458)

The respondents were grouped into seven groups by medical unit:

(1) In surgery units (n = 424)
(2) In obstetrics and gynecology (n = 96)
(3) In anesthesiology, resuscitation and intensive therapy unit (n = 54)
(4) In therapeutic clinics (n = 34)
(5) In cauterization unit (n = 32)
(6) In hemodialysis unit (n = 26)
(7) In toxicology unit (n = 14).

- filtration (grouping) of the input data according to specified criteria (variables) with relational database EXCEL (Figure 1);

Data/Filter/AutoFilter

Specific criteria:

• “length of service” — factorial function;
• profile of the unit in which the researched objects work — factorial variable;
• “profession”;
• “injuries” — resultant variable.
  clinic = 7 and profession = 3 and length of service = 3.

-transcoding the variable “length of service” — from quantitative to qualitative:
If ....then
  Replace ....../Find and replace .....with/replace all
  - analysis of the empirical frequency distributions – descriptive statistics (Figure 2).
  • calculating the range of the confidence interval for the central tendency of the population;
  • formulation of a working hypothesis in order to prove the association between medical phenomena.
  Application of parametric and nonparametric hypothesis tests;
  • In connection with “B” problem a model of trend line of medical phenomenon — the incidence of hepatitis B in Plovdiv region — is necessary to be constructed.
  -fitting of a straight line by the “method of least squares”
  • by applying the interactive computer mode;
  • by registering arithmetic operations to determine the parameters’ model of the phenomenon.
  -analysis of the simulation outcomes and comparison with those from the actual scientific study.
Students apply built-in tools for calculating the range of the confidence interval for the average length of service of the population.

Learners execute commands from the system menu — **TOOLS/DATA ANALYSIS/DESCRIPTIVE STATISTICS** for determining statistical parameter “average length of service”, researched objects’ “average age”. Trainees perform data pre-filtering — the group consists only staff members on medical orderly position (code = 3) in toxicology unit (code = 7).
• formulation of a working hypothesis in order to prove the association between medical phenomena. Application of parametric and nonparametric hypothesis tests;

Working hypothesis (H_0): The profession of medical staff is not a factor for the injuries' occurrence at work.

For parametric test the working hypothesis (H_0) students apply SPSS v.11 licensed software. Since the frame of their survey card is built by using EXCEL spreadsheet, they convert the file of format from *.xls format to *.sav.

The data processing proceeds through the following stages:

• construction of a two-dimensional cross table of the frequency distribution’s values of the categorical variables “PROFESSION” and “INJURES”.

Results of applying the Pearson criterion for cases’ distribution and hypothesis testing.

![Figure 3 Pearson Criterion for Cases’ Distribution and Hypothesis Testing](image)

Results’ analysis: The hypothesis is rejected, based on the assumption that the calculated value of the criterion is higher than the critical tabular value, where the degrees of freedom K = 2 and the error’s chosen level \( \alpha = 0.05 \). The decision that students have to take is — The “researchers” accept the alternative hypothesis H1: The profession of medical staff is a factor for the injuries’ occurrence at work.

Development of micros that contains the key steps for determining the parameters of the linear model for the dynamics of the incidence of hepatitis B in the Plovdiv region for ten year period.

During the simulation, students use the actual data from the retrospective research of Prof. J. Stoilova. In addition to the calculated parameters of the two equation system with two variables, learners apply the built-in graphics options of EXCEL product and add the trend line to visualize the trend of the existing data or forecasts of future data (Stoilova I., Popivanova N. & Petrova N., 2002; Rakadjieva T., Stoilovaq J. & Petrov A., 2002).

Hepatitis A incidence’s dynamics in the Plovdiv region of Bulgaria for ten year period (1990–2000)

According to Robert De Beaugrande true intelligence is “a small amount of productive skills, which to be applied appropriately and to be tuned to the particular task”. Simulation games as a balance between typology and
improvisation developed precisely this kind of abilities (Robert De Beaugrande, Wolfgang U. Dressler, 1981).

References


