New Technology in Teaching Statistics

Radim Briš
Faculty of Electrical Engineering and Computer Science, VŠB-TU Ostrava
17. listopadu 15, 708 33 Ostrava-Poruba, Czech Republic
e-mail: radim.bris@vsb.cz, http://www.am.vsb.cz/bris/,

KEYWORDS: *statistics, industrial software, teaching innovation*

ABSTRACT*: Modern computing systems and statistical software have created new demands for the application of statistical tools for data analysis. Many students will confront in practice raw data in a variety of formats and much more extensive than one and two sample examples. Sophisticated software with high quality interactive graphics expert system capabilities includes advanced statistical topics, as for example complete linear models. The output from such software is totally shocking and mysterious to a student who has only taken a traditional statistics course. This is a common problem of inexperienced and untrained people using sophisticated statistical software.*

The objective of the new teaching technology in statistics was to develop the skills necessary to apply statistical methods, tools, and software to data analysis.

Curriculum has been expanded to include new topics like exploratory data analysis and robust methods. Practical work in data analysis in a fairly context has been included. Further, the renewed statistics course gives training in the use of statistical software, as well, etc.

The general approach of the revised curriculum is to present statistics as a science rather than as a deductive mathematical construction. This scientific point of view is central to the creation of a self-referencing context for the course. The approach is also reasonable in the sense that statistical methods are investigative tools of most other sciences. An introductory course in statistics as science should cover similar ground to other introductory courses in chemistry, physics, biology, even economics or psychology. The course should describe or define the subject matter of the science, the methodology and tools which are used to investigate this subject matter, the most important discoveries of the science, and should give the opportunity to learn the practical investigative skills of the science to re-discover some of the science's fundamental knowledge. The basic subject matter of statistics is randomly distributed data and the investigate tools are methods and models for describing, summarizing, and making inferences about randomly distributed data. The discoveries of statistical science concern the performance and properties of statistical methodology, particularly in common practical applications. Thus the lectures are designed to describe statistical methods in a manner motivated by statistical theory without going into the theory too deeply while the exercises are intended to provide practical experience with statistical methods in a laboratory context. Mathematical detail is used only where it is illuminating.

Experiences with revising both the curriculum of the introductory statistics course and the teaching methods to incorporate statistical software are presented within the paper.

1 INTRODUCTION

Introductory statistics has been taught for decades to undergraduate and post-graduate students alike. Almost every student in the physical or biological sciences, business and economics, or the social sciences at some point in their studies will do a course in statistics.

Hundreds of textbooks have been written to introduce statistical methods from every conceivable point of view. Surely in the midst of all this outpouring of activity, another revision of basic statistics would be simply redundant.

Modern computing systems and statistical software have created new demands for the application of statistical tools for data analysis. Many students will confront in practice raw data in a variety of formats and much more extensive than one and two sample examples. Sophisticated software with high quality interactive graphics expert system capabilities includes complete linear models (sometimes robust

alternatives) with secondary analysis facilities, statistical process control charts, and a lot of other advanced statistical topics. The output from such software is totally shocking and mysterious to a student who has only taken a traditional statistics course.

The objective of the new teaching technology in statistics was to develop the skills necessary to apply statistical methods, tools, and software to data analysis. In Section 2 are described main differences between traditional statistics course and requirements in practice. Basic aspects of the new teaching technology are in detail explained in the section 3. Section 4 brings general approach of the revised curriculum. Conclusions are summarized in Section 5.

2 TRADITIONAL COURSE IN BASIC STATISTICS AND CONFRONTATION WITH REQUIREMENTS IN PRACTICE

Traditional Course in Basic Statistics

Traditional course is usually taught from the point of view of mathematics, that is axiomatic. Axiomatic mathematical structure is created in the course, theorems are established, problems are formulated in terms of the mathematical structure and solutions derived. Typical curriculum in the introductory statistics is as follows:

- Descriptive statistics: representing distributions by histograms and box plots, summarizing distributions by mean, median, standard deviation, range.
- Probability: Elementary set theory, basic axioms, theorems (direct consequences of axioms), conditional probability Bayes Theorem.
- Random Variables: Distributions and density functions, expectation operator, sums of independent random variables, generating functions and integral transforms.
- Theoretical distributions: binomial, normal, perhaps Poisson, t, F, and chi-squared.
- Central Limit Theorem
- Inference: hypothesis testing and confidence interval estimation using a strict decision theory approach with applications to one and two samples with large and small sample sizes, binomial approximation using central limit theorem, point estimation with standards errors and sample size determination.

Practice

In practice the student will confront raw data in a variety of formats and much more extensive than one and two sample examples. Sophisticated software with high quality interactive graphics expert system capabilities including complete linear models (sometimes robust alternatives) with secondary analysis facilities, Kaplan-Meier survival curves, statistical process control charts, experimental design libraries, distribution quantile plots and tests (normal, exponential, Weibull), logistic regression, contingency table analysis all implemented in software. The output from such software is totally shocking and mysterious to a student who has only taken a traditional statistics course.

This is a common problem of inexperienced and untrained people using sophisticated statistical software. The consequence is often the growth of a mythology regarding what the statistical output means. This situation is often seen as the problem and the solution adopted is to introduce use of computers into the traditional statistic curriculum. Essentially this entails having the students use a computer to do the more time consuming calculations. A spreadsheet such as Excel is sufficient to perform all the calculations required by a traditional curriculum.

3 DESCRIPTION OF THE NEW TECHNOLOGY IN TEACHING STATISTICS

Objective and problem in teaching statistics

The objective of the new teaching technology in statistics was to develop the skills necessary to apply statistical methods, tools, and software to data analysis. Thus the ultimate goal became teaching data analysis for which statistical methods are tools.

Some difficulties: Like mathematics, nobody believes they can do “statistics”. It is hard, foreign, incomprehensible. But confronted with raw data, everyone believes they can analyze it that they know what the data says or means. That is why traditional statistics courses often seem irrelevant to the student.

Because of the axiomatic application to data analysis is not discussed. And this application is not obvious. It must be taught. Practice is never so straightforward as the axiomatic description of it.

Comparison with elementary mathematics, there is (in technical universities) only ONE statistics course. The substantive courses rarely include applications of statistics or only fairly trivial ones. The student may never see statistics again after the introductory course. Therefore, the elementary statistics course must be fairly self-contained and self-referencing. It must create its own context.

What has been done?

Curriculum has been expanded to include ANOVA and linear regression up to two variables and exploratory data analysis and robust methods. The course includes practical work in data analysis in a fairly context. The course gives training in the use of statistical software. Thus the course teaches much more in the same time, as traditional course. This was achieved by different ways. Obviously many things cannot be treated in the same axiomatic detail but developing a course which successfully develops data analytic skills for an expanded curriculum requires a course design which achieves synergy among the various course components. Course components are:

- Lecture
- Exercise
- Software training
- Projects
- Evaluation

Characterization of the new teaching approach

Ultimately, the key to the course is the exercises. What happens in the exercises is what we want the student to remember. The purpose of the lectures is to serve the exercises. This is exactly the opposite of the axiomatic or didactic approach where the purpose of the exercises is to reinforce the lecture material. This reversal underscores the difference in orientation between the traditional course and a course focused on skill development. The purpose of the lecture is the primarily to give the student a point of view or orientation from which to interpret or understand what happens in the exercise. Of course, the lecture will also transfer some necessary specific information. Consider the analogy of learning to play tennis. The object is to be able to hit the ball. But if you just stand someone up with a tennis racket in hand, it would be a very long time and a lot of experimentation before they developed the skill. In fact, without guidance, they may never develop proper technique. So usually teaching begins with some verbal introduction or explanation is not intended to be an in-depth treatment of aerodynamics of a tennis ball or kinetics of muscle contraction. Rather the explanation would tend to be in the form of metaphorical description which would give the learner something productive to think about as the tennis ball approaches. Once the skill is developed, it is no longer necessary to continuously remember the metaphorical descriptions. So it is with a statistics lecture in this context. Much of what is transmitted in the lecture is ultimately disposable. We do not expect the student to remember much of it.

This new objective for a lecture raises the question of what should be transmitted in the lecture.

- Must every topic be given an axiomatic structure-only if it can be done quickly and the practical implications of the axioms illustrated.
- Must every proposition be derived formally: only if the derivation is short and gives some insight into the idea. Remember, this is disposal knowledge and we cannot expect the student to retain it.
- Must formulas be given in detail: formulas should be given in both schematic and detail form, but the detailed form should conform to the schematic form. Only the basic formulas need be given in detail. The more complex formulas will also follow the schematic form. It is the schematic formula as well as certain fundamental relationships (deviations from mean always sum to zero) that the student should remember/know.

Organization of the Lectures:

Lectures become more effective learning experiences if students are supplied before the lecture with an outline of the lecture containing brief summaries of the main points of the lecture, as well as all formulas, charts of graphs, and where possible a brief description of the derivations and examples. These handouts are usually a copy of the lecture slides which the lecturer uses.

The lecture summary is not intended to be complete. The lecturer discuss the main points in greater detail and give further examples using either the blackboard or a computer file projected onto a screen. This new interactive technology is very promising for lecture use, but requires experimentation and learning on the part of the lecturer to be effective. In truth, this new lecture format will require a major adjustment for a lecturer accustomed to the didactic mathematical approach. Rather than being on the solid ground of deductive exposition, the lecturer will often be just talking, describing in colloquial terms. This requires some experience and a different style of preparation. However, this is to be expected when the emphasis is shifted to skill development. In the practice of data analysis, we cannot talk about right or wrong methods but rather good or bad analyses.

In light of this revised lecture format, the question arises: Will the students even attend the lectures if the notes are given to them? Answer is, that lectures are simply more efficient. If the full lecture text is written out in detail, it will take much longer than 90 minutes to read it and yet much of the feeling, sense, and motivation of the material will still not be conveyed. The flexible, interactive situation which can be created in a lecture transmits more information and sensibility more quickly than can be achieved on the printed page. There is opportunity for emphasis and repetition or re-expression.

Exercises

Under the course design, the most important learning occurs in the exercises. The exercises are designed to serve two primary purposes: develop data analysis skills, and learn to use statistical software. It is these skills which should be the permanent and long term result of the course. Since statistical software is used to perform the data analysis in the exercises, the two objectives are very compatible. A sort of learn by doing approach. Each exercise session consists of some data analysis problems relevant to the lecture of that week. The student will proceed through the analysis with the lecturer. An additional data assignment will be given for students to submit.

Project phases

Making out the project consists of the following phases: problem formulation and data base design; data collection, initial exploratory data analysis, data checking and editing; confirmatory data analysis, interpretation and conclusions. The project component of the course is an important part of both developing an understanding of how statistical tools work in practice and skill in applying these tools to data analysis. The effectiveness of this component depends on the students´ understanding of an interest in the subject area of the project. The students´ knowledge of the subject area of the project will guide analytic decisions, interpretation and formulation of conclusions. In addition, the project component can be organized so that proceeds concurrently with the lectures and at each point, the phase of the project is an application of the current lecture topic.

One difficulty in implementing the project component is the inexperience of the students. Developing a viable project out of the students´ initial conceptions requires a lot of guidance on the part of lecturer. Typically three to four meeting are required during the semester.

Evaluation

Revising the evaluation process is essential to achieving the course objectives. Students will learn what is necessary to achieve a good evaluation. If the course is revised on the manner described above but the evaluation criteria remain the same, students will learn exactly what they were learning before the revision. The evaluation consist of 4 parts: weekly exercises; projects; written examinations; oral examination. The project and oral examination provide an excellent opportunity to evaluate the students´ progress in acquiring data analysis skills. Evaluating skill acquisition in the more constrained format of exercise problems and written examinations is more difficult and requires more creativity. Exercise and written examination problems are presented in a realistic context so than part of the problem is formulation and selection of appropriate statistical tools. The important points of evaluation of the project are: formulation; interpretation; and conclusions. Merely calculating data summaries or performing regression is not sufficient. The oral examination is the best opportunity to assess the students´ data analytic skills. The examination consists of two parts: a discussion of the project and a practical data analysis problem using computer software. For examination purposes, it is not practical to prepare 60 unique data problems. However, it is possible to formulate 5 data analysis situations and then simulate a dozen data sets for each. By adding outliers, or contaminating distributions, each data set is manipulated

to be unique. The tradition of oral examinations at VŠB-TU Ostrava, while extremely demanding for the lecturer, does provide an excellent opportunity to realistically assess the students´ data analytic skills.

4 REVISED CURRICULUM

The general approach of the revised curriculum is to present statistics as a science rather than as a deductive mathematical construction. This scientific point of view is central to the creation of a self-referencing context for the course. The approach is also reasonable in the sense that statistical methods are investigative tools of most other sciences. An introductory course in statistics as science should cover similar ground to other introductory courses in chemistry, physics, biology, even economics or psychology. The course should describe or define the subject matter of the science, the methodology and tools which are used to investigate this subject matter, the most important discoveries of the science, and should give the opportunity to learn the practical investigative skills of the science to re-discover some of the science‘s fundamental knowledge. The basic subject matter of statistics is randomly distributed data and the investigative tools are methods and models for describing, summarizing, and making inferences about randomly distributed data. The discoveries of statistical science concern the performance and properties of statistical methodology, particularly in common practical applications. Thus the lectures are designed to describe statistical methods in a manner motivated by statistical theory without going into the theory too deeply while the exercises are intended to provide practical experience with statistical methods in a laboratory context. Mathematical detail is used only where it is illuminating.

5 CONCLUSIONS

Elementary statistics course was re-designed and innovated to incorporate trends in statistics software technology. The innovation of the course met the basic course objective, that is - to develop in students the following set of skills and knowledge:

- ability to recognize applications of simple probability models in practical situations
- ability to choose appropriate statistical techniques for data summary, analysis, and presentation
- ability to use standard statistical software to implement simple models and techniques
- ability to correctly interpret the output of standard statistical software.

New teaching strategy is used in the course. Explanations are motivated by statistical ideas rather than computational or theoretic considerations to illustrate properties of procedures. These properties are discussed comparatively in the context several techniques. The use of computer software is taught simultaneously with the introduction of the techniques. Computer applications are integrated fully into the course.