JUST-IN-TIME NETWORK-BASED STATISTICAL LEARNING: TOOLS DEVELOPMENT AND IMPLEMENTATION ®

Lea Vermeire, An Carbonez, Paul Darius
Katholieke Universiteit Leuven, Belgium
Jill Fresen
University of Pretoria, South Africa

The paper reflects on the growing demand for statistics training, in permanent education in general, and in just-in-time training with direct applicability in particular. It reports on work in short-course and in in-company training in statistics. Special attention is given to two cases, on the one hand a self-study course for a government department, on the other hand a system of highly interactive applets for visualization of statistical concepts related to the linear model.

JUST-IN-TIME NETWORK-BASED LEARNING

The existence of a growing demand for permanent education in general, and for just-in-time on-the-job training with direct applicability to ongoing work in particular, cannot be denied. On the demand side are professionals with an academic degree or higher education. They are active in various socio-economic sectors like industry, government, finance and business, health sector or the educational domain; and they have various aims like updating or upgrading their knowledge, skills or competence. On the supply side, universities, private companies and mixed consortiums are the providers of formal and just-in-time education.

Computer based training and/or e-learning (cd-rom and Internet) are becoming increasingly important as learning tools for just-in-time learning. Short focussed study modules for small groups are becoming popular and e-learning is cost-effective when implemented on a large scale. E-learning allows anytime-anywhere learning and stimulates communication and interaction.

Competence-based training reflects on the desired learning outcomes and establishes the means to achieve them. Firstly, the learning contents should be designed to meet the needs of the learners. Secondly, an appropriate didactical approach should be chosen. Thirdly, the available technology should be investigated. Finally, learners and training providers should strive to provide and sustain a stimulating learning environment.

AN INCREASING DEMAND FOR STATISTICS EDUCATION

There is a great need for statistical training, particularly in the field of post-graduate lifelong learning. Reasons for the rising demand for more advanced statistics education are the growth of applicability in almost all sciences, and spectacular evolutions in the types of data sets available, the methods of analysis and computer technologies:

Evolving data sets. There is an increasing availability of data with the potential for filtering out information relevant to decision-makers. Huge data sets and evolving data sets are available for on-line analysis.

Evolving methods. New types of analysis are required to detect structures in the huge data sets and to perform the computations within reasonable time limits. The availability of increasing computer capacity allows for computer intensive methods, such as new robust statistical methods.

Evolving software. The accessibility of user-friendly statistical software has broadened the potential community of users of statistical analysis. Interactive statistical experiments and virtual laboratories enhance the understanding of statistical concepts, and widen the audience with statistical skills, as they make abstract mathematical expressions more concrete and more accessible.

Statistical training is offered by professionals from both the academic and the private sector. Many universities focus on post-graduate education in their mission statement. For example, in recent years all universities in Belgium have implemented a one-year Master's program in statistics or several post-graduate short-courses in statistics, or both. Private stakeholders and producers of statistical software see the potential of new markets.

IN-COMPANY TRAINING IN STATISTICS

Companies realize that just-in-time on-the-job training can be efficient in producing desired learning outcomes and in saving time and money. At the K.U.Leuven, the University Center for Statistics offers a meeting point for mutual support between all statistics groups at the university, in the whole range of statistical activity: from consulting and course delivery, to an international Master of Statistics program, and to education and research projects. This Center offers visibility and an efficient service to internal and external consultation seekers and education prospectors in statistics.

The typical process for an in-company course on statistics starts with an application by the company for a course or an advanced statistics module. The first contact with the company provides information on the statistical techniques they are currently using, and those they intend to use. We ask for the statistical background of the participants, the software package they are using, the size of the data files they usually deal with and the possibility of making use of data files provided by the company. Practical issues, such as timing, full-day or half-day schedule, etc. are settled.

We usually recommend an advanced course should begin with a basic statistics course of a few days as a starter, to refresh the basic concepts and notations. To stimulate interactive teaching and learning, we recommend that the size of the group be restricted to 15 participants. Whenever possible, the statistical techniques are illustrated using data from the company itself, since this improves the motivation of the participants significantly. To illustrate key concepts, interactive Java applets are used and are generally highly appreciated. Almost 50 % of the training time is reserved for guided exercises and case studies. It is essential to give the learners the opportunity to practice during the training session, using their own software package.

Finally, the participants are asked to complete evaluation forms about the pedagogical context of the training, the content of the course, the fulfillment of the objectives etc. The results of these evaluations are implemented to enhance later training sessions.

A SELF-STUDY BASIC STATISTICS COURSE FOR THE NBB

This section reports on the design and the implementation of a course on basic statistics and inference, commissioned by the National Bank of Belgium (NBB). The *instructional design model*, the blueprint that was used in designing and developing the learning material, covered the following steps:

The Instructional Problem

"One of the most damaging traditions in educational technology is our temptation to advocate educational solutions without a clear analysis of the specific problem needing to be solved." (Clark & Estes, 1999, p.6). In this project, the client was able to describe the instructional problem they faced. The need was to teach basic statistical concepts and practical statistical techniques to the personnel in the General Statistics department at the NBB. There were various constraints to be considered, which are described in the following sections.

Goal Analysis

The specific goals of the training package are:

- to provide training in basic concepts in Statistics
- to facilitate understanding of how to use Statistics to address real problems
- to enable learners to think critically about data
- to enable learners to select the most appropriate solution from different techniques at their disposal
- to equip learners with a joint language – for concepts and notations – which facilitates discussions and mutual support in Statistics in the department
- to provide practice in the application of Statistical methods in the work environment, using common spreadsheet software.

Target Population Analysis

According to Harrison, two key factors to consider in designing learning materials are "the needs and constraints of your target group and what you expect them to be able to do" (Harrison, 1999, p.206). The learners in this project are adults working in a national bank. Their ages range from 20 to 60 years. The bank is a bilingual government institution, with languages French and Dutch; therefore, all learning material must be made available in both languages.

The background studies of the learners are heterogeneous, primarily in the field of Economics, not necessarily at university level. Their experience in formal Mathematics is minimal. Furthermore, their aptitude for mathematical and statistical thinking is average, compounded by possible negative perceptions and concerns about the complexity of the subject of Statistics. A self-study package offers this heterogeneous group the advantage of being able to study at their own pace and to revise material as and when necessary. Interactive demonstrations and simulations are necessary to bring meaning to abstract statistical concepts.

The required level of prior knowledge of the students is specified as follows:

- basic PC computer skills: basic Windows skills and use of the mouse
- basic Web knowledge and usage of a Web browser
- basic Excel knowledge – this will be extended further in the examples
- no basic knowledge of Statistics.

Content Analysis

The decision was taken to base the content of the course very closely on the textbook "Introduction to the Practice of Statistics" by Moore and McCabe (1994). (A completely revised 4th edition has since been announced for 2002.) The chapters in the textbook were used as a basis for the syllabus outline. The book is published in English and Dutch. Our developers developed the course in English and Dutch and it was later translated into French. The website of the publishers offers computer-based assessment questions and some of these were modified and used with others in our testing package.

Instructional Approach

The instructional approach (or pedagogical framework) reflects an awareness of adult learners with differing learning styles and differing immediate needs with respect to the learning material. For this reason, a flexible learning model is used, as discussed in the following section.

Interpretation - The learning material concentrates on statistical concepts and the interpretation of statistical data rather than on the underlying mathematical theory. An attempt is made to enhance insight and understanding by highlighting the reasoning behind the concepts. Emphasis is placed on the analysis and meaningful interpretation of statistical results.

Prior knowledge - In each new section of the learning material, links are made to the prior knowledge required in order to master the new material. Regular references are made to learning material previously encountered. In this way the underlying relationships between different concepts are elucidated.

Levels of proficiency - In order to cater for students with more advanced mathematical skills, some sections of the learning material are marked with an asterisk. This indicates content on a more advanced level, which may be omitted without disadvantage to the learner in subsequent sections.

Summaries - Each section ends with a summary of the most important concepts. Thus it is not obligatory for each individual learner to go through all the learning material. Learners who possess an adequate knowledge of particular statistical concepts can, by means of studying the summaries, refresh their memories and master those areas requiring revision.

Guidelines - The emphasis in this course is on reasoning and the development of insight in Statistics. Certain statistical techniques are based on mathematical assumptions and users must be able to judge which procedures may be applied in certain circumstances. Therefore guidelines are given regarding the robustness of specific statistical procedures in various real life situations. Particular emphasis is placed on using actual data sets provided by the National Bank.

Delivery Media

Due to the nature of the target population, the course is a stand alone distance education package, which learners can follow at their own pace and in their own time. Therefore, one of the primary delivery media is a learner-centered Learning Guide, which supports and guides the user through the topics in the textbook. Access to a computer is provided on the intranet at the bank. A facilitator for further support is available during sessions in the computer laboratory. The learning model offers a blend of various different delivery media, as shown in Figure 1.

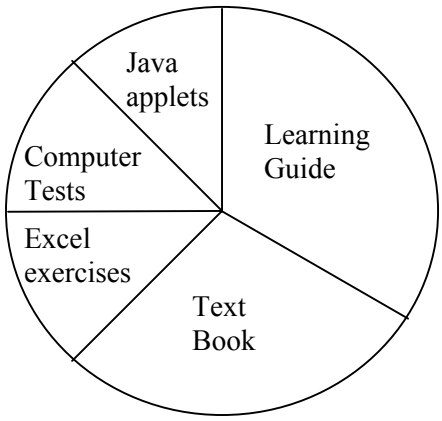

Figure 1. Blend of Delivery Media in the Training Package.

Interactivity

In keeping with the philosophy of elucidating underlying statistical concepts, the course includes several statistical simulations, which were programmed in Java by members of our development team. An Internet browser with Java enablement is required to run the applets.

Some statistical topics, which lend themselves to simulation, are histograms, boxplots, continuous distributions, correlation, regression and confidence intervals. The simulations enable the student to view graphical representations of concepts and relationships. The student is able to interact with the simulation by changing the values of variables and by dragging axes, markers and lines to see the effect of such changes on relationships and on their graphical representations.

Assessment

It is vital that the learners should be able to apply what they have learnt in their real-world job environment. The assessment strategy includes two opportunities for assessment at the end of each chapter in the Learning Guide: self-assessment exercises and computer-based tests.

Self-assessment exercises - The National Bank supplied a cd-rom containing hundreds of large data sets including economic, employment and population data. These data sets were used wherever possible as examples in the Learning Guide, as well as in the self-assessment exercises. Other exercises from the economic sphere are included where no appropriate data was found on the given cd-rom. By using exercises and case studies from the world of work of the participants, it is hoped to maintain their interest and motivation at a high level. The emphasis in the exercises is strongly on the interpretation of results and the formulation of conclusions. Full model answers are provided at the end of each exercise section, so that students can refer to these in order to check their own solutions. The use of Microsoft Excel is encouraged as a tool for statistical calculations. Although basic Excel skills are a prerequisite to the course, Chapter 0 of the Learning Guide gives some basic training in Excel, including screenshots to demonstrate the use of its statistical features. The model answers equally include screenshots from Excel to illustrate its use in solving a particular problem.

Computer-based tests - Formal testing and record keeping is achieved through the use of a computer-based test at the end of each chapter in the Learning Guide. Diagnostic feedback is

given for each question, which indicates which topics have been mastered and which may need revision.

After completing a specific test, the participant and the administrator receive the results immediately, giving a complete record of the participant's progress. Our development team considered using commercial computer testing packages, such as QuestionMark Designer, the Quiz tool in WebCT or Hot Potatoes. Due to budgetary considerations and the fact that our team possessed the required programming skills, it was decided to develop our own simple computer-based testing system in line with the configuration of the client's intranet.

APPLETS FOR BASIC STATISTICAL INFERENCE

Self-study or distance education has become more and more network-based, in the sense of technology-based, form of education, both from the teacher's and the learners' perspectives, all the way from joint course development to learners' support by means of tutoring, illustrations, progress assessment, knowledge certification etc. The variety of applications envolved in web supported distance learning is illustrated by Eisenstadt and Vincent (1998), in a compilation of cases developed at the Knowledge Media Institute of the UK's Open University.

We argue that just-in-time network-based statistical training should be more than a presentation of possible recipes for data analysis. In fact, current technology offers features that can be specifically used for a better understanding of statistical methods and concepts. Interactive Java-applets and virtual statistical laboratory experiments allow for visualization of statistical concepts and hands-on simulations which have a high pedagogical potential:

- to speed up understanding by visualizing basic concepts and basic theorems
- to develop a critical attitude for correct use of methods, through counter examples if the conditions of a theorem are not satisfied
- to foster sustained understanding through experimentation
- to learn by discovery and communication.

Saporta (2001) provides a commented list of Internet resources for statistics. At the autumn 2001 meeting of the EADTU Science and Technology Network in Hagen, Mittag demonstrated a pilot applet follow-up to a widely used multimedia course on descriptive statistics (Mittag & Stemann, 2001). VESTAC is a product developed at K.U.Leuven, in a collaboration between Belgian universities (Visualization of and Experimentation with STAtistical Concepts: www.kuleuven.ac.be/ucs/java/). This is a collection of over thirty highly interactive applets, in the domain of basic statistical inference and the linear model. The flavor and level of the applets is reflected in the following experiments: tail behavior of the key statistical distributions under growing degrees of freedom, statistical tests and their error probabilities, a population regression and a corresponding sample regression, distribution of the estimated regression parameters, exact tests. Special attention is given to the contrast between a fixed population parameter and the sampling variability of a corresponding statistic. Of particular pedagogical value is a multi-window facility, which allows a learner to perform and compare on one screen two or more experiments with new parameter settings and/or new sample sizes. Our educational experience is that the experiments really appeal on the students' curiosity, the beginning of any deeper learning experience. The applets are freely accessible on the web.

CONCLUSION

Just-in-time and anywhere-anytime statistical learning is greatly enhanced through different types of network-based collaboration, from joint preparation of learning materials to web-presented experiments, simulations and assessments.

REFERENCES

Clark, R.E., & Estes, F. (1999). The development of authentic educational technologies. *Educational Technology, 39(2)*, 5-16.

Eisenstadt, M., & Vincent, T. (Eds.) (1998). *The knowledge web: learning and collaborating on the net.* London: Kogan Page.

Harrison, N. (1999). *How to design self-directed and distance learning programs*. New York: McGraw-Hill.

Mittag, H.-J., & Stemann, D. (2001). *Descriptive statistics and exploratory data analysis* (3rd edn). Multimedia software (cd-rom, in German). Hagen (Germany): FernUniversität.

Moore, D.S., & McCabe, G.P. (1994). *Introduction to the practice of statistics* (3rd edn). Academic Press.

Saporta, G. (2001). Teaching statistics with the Internet: available resources and the St@tNet project. *Proceedings of the EADTU Millenium Conference in Paris* (2000). Heerlen (Netherlands): EADTU, 415-419.