PENETRATION OF STATISTICS EDUCATION BY INTERACTIVE MATERIALS WITH COMPUTER NETWORK ®

Naoko Sakurai, Michiko Watanabe, Katsuyuki Suenaga, Hiroshi Yadohisa, Akinobu Takeuchi, Kazunori Yamaguchi and Chooichiro Asano
Tokyo University of Information Sciences, Toyo University, Kagoshima Immaculate Heart College, Kagoshima University, Rikkyo University, Rikkyo University, Soka University
Japan

This paper introduces an outline of the integrated web site produced from our faculties' project offering such services as statistics education, electronic books, and statistical libraries for computer use with a search engine for statistics, which we are currently developing. Statistics is a basic and cross-disciplinary study that should be applied to positive analysis in a wide variety of fields. Both the theoretical side and the practical use as a data analysis tool have to be its primary role. Moreover developing statistical analysis software has been the mainstream of disseminating statistics to the various departments. Computers and its network have become quite popular, still use of data processing software such as spreadsheets is expanding among people in the society. In this circumstance, strong demand for basic statistics education is absolutely increasing. In addition as the Internet is supposed to grow as a large-scale communication media, now and in the future, beyond time and space, to increase information about statistical science and education on the Internet promotes the efficient dissemination of statistics.

INTRODUCTION

Our society has been evolving with great success of quality control in manufacturing and finally met an IT revolution in the 1990's, which brought us a hyper-linked computer-networked society. At the same time the big challenges our society has to come to grips with are to accumulate effective data and to analyze the enormous volume of digitized data. For example, database transaction processes between companies (called 'B to B') and between company and consumer (called 'B to C') are now going through computer networks and are recorded via on-line system. Since many companies are considering to build up their own electronic commerce system or managing POS (Point Of Sale) databases, there are increasing expecting needs for data analysis methodologies which will make companies to do decision making by carrying out exploratory searches for valuable information from such large scaled databases.

Moreover, it has become more commonsense for public agencies, several foundations and private companies to publish statistical databases as open source on the web regarding economically responsible social events. Yamada announced that the number of sites of private companies regularly updating statistical data has more than doubled from September 1998 to May 2000 (Yamada, 2000). This trend can be seen to continue. Furthermore, the Japanese ministries are planning to publish a microscopic database, adding to the currently published one, to satisfy the strong demand from universities, economic and other research organizations. In U.S. and European nations, the microscopic data of economic/social survey results have already been made public and its analytical result has been fully utilized via the computer network. In these societies the ability to collect and analyze information through computers is regarded as the key factor in income or social status differentials, which we call 'digital divide'. Thus, under this situation such as 'abundant database' and 'broad-banded computer network', the training opportunities for the statistical analysis ability of administrative officers, business people and university student is regarded as a social need.

Statistics is a basic and cross-disciplinary study applied to positive analysis in a wide variety of social fields. In addition, statistics is more suitable for the online education style where it is possible to grow new potential abilities. This is an inevitable point. The traditional way of learning at university classes mainly explains the mathematical aspects and offers no practical exercises as a data-analyzing tool. Recent network technology makes it possible to develop educational materials using innovative media tools, by making use of multimedia, dynamic charts and hotlinks to related sites. This will change the learning process everywhere from theoretical to practical and from passive to participating. At universities and other educational research

institutions, it has become easier for individual to use data resources on the Internet in daily classes regardless of the discipline, as the network environment has improved. Research Institutes and individuals are not utilizing effectively the interactive and media technologies available through the Internet to develop different kinds of an educational material. In this paper, we introduce our projects for the development of the educational web site for dissemination and saturation of statistics.

ITLS (INTERACTIVE TEXT FOR LEARNING STATISTICS)

When our antecedent project started in 1999, typical lecture style such as mass education in a large classroom was still popular, although network infrastructure at private universities had been improved to some extent. The collaborative project tried to develop and use an educational web site by teaching staff of statistics-related liberal arts. At that time it aimed at realizing practical analytic education in statistics, economics and econometrics, still people did not expect all the educational materials to be on-line. It was a reasonable solution to develop materials and give lectures, making use of the merits of both paper and on-line texts with computer. Therefore, in our lectures we explained the basic tools for analyzing economic data, estimated the finance market, which attracted practical attention in those days and outlined a new frontier style analysis methodology for assessing credit risk. By using online materials through a PC, students were able to actively participate in the learning process. In this environment, students could keep their active will to study, compared with their passive attitude toward conventional style lectures. As a result, it became possible to give participatory and practical education to liberal arts students, who tended not to enter this field. This was exactly a great event. So, we set out new project to develop and publish an interactive textbook for learning statistics along with the progress of computers and network abilities. We standardized the total design, added dynamic charts and improved slides for lectures using Java Applets, and expanded the search function. We offered these materials as lecture notes to teaching staff and also to business people for self-study.

ITLS consists of 'Search Engines and Links', 'Statistics and Excel', 'Data Analysis (Description part and Inference part)', 'Basic Theory of Regression Analysis', 'Analysis of Time-Series and Seasonal Adjustment', 'Lectures on Multivariate Analysis' and 'Statistical Analysis by Statistica'. 'Search Engines and Links' shows a brief explanation of the Internet, how to retrieve econometric and social information (statistical data), and how to download data. 'Statistics and Excel' teaches the fundamental data analysis procedures on the PC through the operation of so-called spreadsheet software, such as Excel. 'Data Analysis' shows the basics of statistical analysis and 'Statistical Analysis by Statistica' introduces one of the popular software programs available for statistical analysis. In order to support the understanding of practical data analysis technologies explained in the online text, the contents of each category offers keyword searching functions, supplementary explanation of appointed keywords, browsing and downloading slides for lectures, links to reference sites, databases for exercises, and Q&A, which cannot be covered by conventional paper textbooks. We try to use short, easy and refined sentences, colors, hyperlinks, images, and dynamic charts, and not to be dependent on only texts and numerical formulas, to promote intuitive understanding of the concept of statistics. On the other hand, theoretical and appropriate explanations using sentences will be made by 'Electronic Books for Statistical Analysis', which will be explained in the next section. Furthermore, we prepare basic operations pages for PC beginners to give some simple explanations with actual screen images from keyboard entry to saving files, web browsing, database searching, sending/receiving e-mail, using word processing, and creating web pages. These will work collaboratively with basic information processing education. We are currently offering approximately 330 slides from lectures, 120 keyword explanations and 17 dynamic charts, using Java Applets.

By posting educational materials open to the public on the Internet, we can offer updated accumulation of resources without posing a limitation on users, place, and time. At the same time, teaching staff at universities and at private companies can share of the data for their lectures. In the near future, a virtual space where teachers can exchange their opinions about education freely will be required on the Internet. Since our site intends to offer educational materials for teaching staff as well as offering educational services to students, a wide variety of materials open to the users will be required. It is not efficient, however, if the materials are collected by the project

members only. So, we are going to publicize the usefulness of this project through contributing papers to domestic and international academic conferences in order to obtain cooperation in sharing resources. From now on, international cooperation is also important. For example, an international multimedia database including animation and Java Applets which is under preparation by Hagen University, registers and exhibits language-wise multimedia Applets including English, German, and Spanish. We are preparing to publish the English and German versions of the Java Applet created under this German project and the Japanese version of the Applet by Mittag (Mittag 2000). We will register them when they are complete and try to share the resource internationally.

Here is one problem. When we use multiple images or dynamic charts and then try to boot up a multimedia site with voice data or animated materials, the response is slow due to the limitation of line speed. Hopefully the inter-university network will be improved in the future. For the current problems such as access congestion caused by the heavy use during the lecture, we will place the contents in the server offered by other organization and make it work as a mirror site. We are also considering the distribution of CD-ROMs to students for self-study. Figure 1 is an example of ITLS contents.

EBSA (ELECTRONIC BOOKS FOR STATISTICAL ANALYSIS)

We introduce here the next project of online book reading system referred to in the previous section. There are plenty of important and valuable books, which are still formally used as technical books in the statistical science area especially in theory or as instruction manuals for many fields of statistical application, and are currently out of print or have no plan of being reprinted in the future. These high quality books are valuable for their universality and usability for their statistical science knowledge, regardless of their publishing age, but they are not currently available and usable. To tackle this problem, we created a site named EBSA in May 2000 digitizing on the web these out-of-print books on theoretical and applicable side of statistical science, after receiving official approval from copyright holders (author or the bereaved family, etc.) and publishing companies.

We scanned all pages of the original books, converted them to PDF format and published them on the web for browsing, reading and printing. We also collected and integrated all index words in every book and created a glossary database, each of which is referred to in all the applicable pages (group of pages). Therefore, if a user enters a keyword of statistical science, the engine searches for the word across all the registered books, or the books user choose. So, the user can meet quickly the exact pages required. In addition, it often excludes heavy image files and enables users to see and print desired pages only. The system also has text-based index pages so that it can be often retrieved by the prevailing search engines. Hopefully people will access these original books more frequently leading to a greater understanding of statistics. Figure 2 shows some parts of the EBSA system.
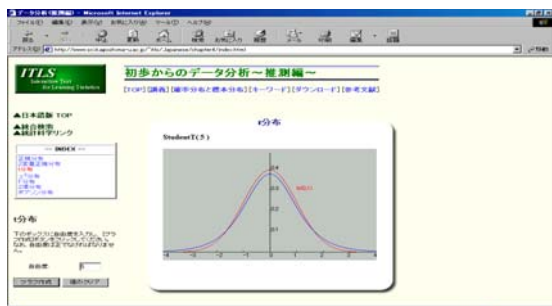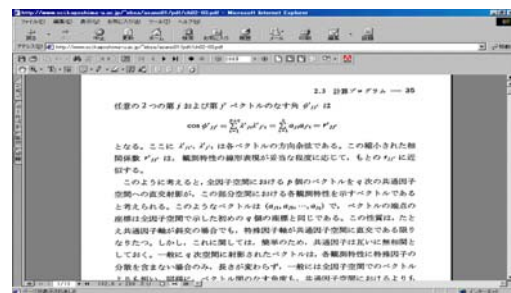

Figure 1. Example from ITLS.

Figure 2. Example from EBSA.

DLLSA (DYNAMIC LINK LIBRARY FOR STATISTICAL ANALYSIS)

One more important consideration in constructing the statistical education system on the Internet web is the existence of scientific statistical analysis software libraries utilized on PC. It is helpful for most learners to analyze practical data by PC being parallel to theoretical learning. In

fact, there have been some reports written about its effectiveness. The 'DLLSA' independent project had already been started earlier when computer use was not so popular among people. DLLSA is composed by subroutine libraries, which can be called from any software on the main operating system such as Windows or through the Internet. Users can call only necessary parts of DLLSA libraries from their application software in their computer. DLLSA system is independent and is also a very excellent point for WWW use.

DLLSA has four main departments:

- Data Management
- Calculation of Statistical Number
- Basic Statistical Quantity
- Multivariate Analysis

In addition, since this system has a usable interface with Microsoft Excel, the user can do statistical analysis by operating DLLSA as if they would manage statistical analysis under the excel environment.

CONCLUSION

This paper has introduced an outline of the integrated web site offering such services as statistics education, electronic books, dynamic link libraries for statistical analysis and a search engine for statistics, which we are currently developing. Statistics is a cross-disciplinary study applied to positive analysis in a wide variety of fields. Practical use as a data analysis tool has to be one of its primary role. Developing statistical analysis software has been the mainstream of disseminating statistics. Computers have become quite popular and the use of data processing software such as spreadsheets is expanding among most people. In these circumstances, demand for basic statistics education becomes more significant. As the Internet is growing as a large-scale communication media increasing information about statistical science on the Internet promotes education and the efficient dissemination of statistics.

For practical and basic education, a system is required which integrates an educational database covering various fields of application, educational software, widely available analytic software, on-line statistics dictionary and analysis sample database. Each site introduced here has not been integrated yet but we believe we are making meaningful progress. Figure 3 shows the step of our trial to integrate the sites.

Web sites introduced here are currently operated on two servers. URLs are:

http://stat.eco.toyo.ac.jp/~stat/ (Toyo University)

http:// www.sci.kagoshima-u.ac.jp/~stat/ (Kagoshima University)

As the inter-university network system is improved, accumulating and sharing digitized educational materials on the Internet will be further required in various areas of education. It is hoped that study of our project working would be of some help.

Figure 3. Top Page of Integration

REFERENCES

Mittag, H.J. (2000). Multimedia and multimedia databases for teaching statistics. CD-ROM version of *Proceedings of ICME-9, Makuhari, Japan*. At *www.fernuni-hagen.de/STATISTIK/*

Yamada, S. (2000). Report on statistical web sites made by private companies (in Japanese). In *Proceedings of the 68th Annual Meeting of Japan Statistical Society* (pp. 228-229).