

Teaching Statistics with Internet: a Survey of Available Resources and the St@tNet Project

Gilbert Saporta

Conservatoire National des Arts et Métiers, Chaire de Statistique Appliquée

292 rue Saint Martin

75141 Paris cedex03, France

saporta@cnam.fr

Internet supplies useful resources for teachers and students for an interactive education (data, applets, softwares, web links, on-line electronic courses) and may be a vector for distance education: St@tnet project developed within the framework of the French-speaking Virtual University is an illustration of it.

1. A few lists of links

Internet provides a lot of material useful for teachers and students and this is a growing phenomenon. The most useful resources are data tables, textbooks, electronic journals, applets and free softwares. There is subsequently an increasing number of sites giving lists of links, frequently maintained by academic researchers. It is actually impossible to have a complete and up-to-date list of references, and the following addresses contain almost surely some erroneous links, (see Morin and al., 1998).

The IASC web site (<http://www.stat.unipg.it/iasc/>) is among the best websites maintained by association or consortiums where one can easily access to a wide range of links. An other interesting entry point is the CTI-Stats, which is a part of a wide project between british universities for developing the use of computers in education (<http://www.stats.gla.ac.uk/cti/>). Let us mention also the Cirdis site in Italy (<http://www.stat.unipg.it/cirdis/>) or the australian "Guide to the Web for Statisticians" (<http://www.maths.uq.oz.au/~gks/webguide/>) and the Statlib (<http://lib.stat.cmu.edu/>) of Carnegie Mellon.

Some private companies provide useful links to software, data-base and other material, for instance STATSNET (<http://www.statsnet.net/>)

However the most comprehensive sites are maintained by individual researchers:

M.Friendly from York University (<http://www.math.yorku.ca/SCS/StatResource.html>) gives very complete informations and J.Puranen (<http://noppa5.pc.helsinki.fi/links.html>) has one of the most complete list of links for teaching material.

D.Stockburger provides in his electronic textbook about Multivariate Analysis some list of links (<http://www.psychstat.smsu.edu/scripts/dws148f/statisticsresourcesmain.asp>) with short comments. A synthetic interactive reference paper is the conference given by Robin H. Lock's (1998).

2. Applets, free software and computer environments

Teachers may find free Java applets, designed for illustration purposes (eg: central limit theorem, influential points in regression) which may run on almost any computer. One of the best known site is WebStat (<http://www.stat.sc.edu/~west/webstat/>) created by W.West from the University of South Carolina. "Many times statisticians develop procedures in languages such as Splus, SAS, Minitab, etc. Students and other potential users may not have access to these languages. By using Java and the World Wide Web, WebStat should reach the broadest possible audience of

any statistical software of its kind". See also the paper by West and Ogden (1998) in the electronic Journal of Statistics Education.

Other resources for Java applets are GASP (Globally Accessible Statistical Procedures) (<http://www.stat.sc.edu/rsrch/gasp/>) and WISE (<http://acad.cgu.edu/wise/appletsf.shtml>).

If applets are very useful to illustrate concepts, the need of statistical software to process real data may be satisfied by software of high quality such as ViSta or StatEasy which are freely downloadable:

ViSta developed by Forrest W. Young (<http://forrest.psych.unc.edu/vista/>) is available in English, French and Spanish. It comprises Introductory Statistics, Multivariate Statistics, Computational Statistics, and Graphical Statistics, including advanced topics such as robust and monotonic regression, biplots, multidimensional scaling, correspondence analysis.

StatEasy by A. DiCiaccio is a freeware italian software oriented to research and didactics in descriptive and inferential statistics (<http://space.tin.it/scuola/adiciacc/>).

The computer environment XploRe (<http://www.xplores-stat.de/>) has been developed for data analysis, research and teaching. Its main features are an extensive set of parametric and nonparametric methods incorporating kernel estimation and smoothing, generalized additive models, nonlinear time series analysis, modern regression techniques, wavelets and neural networks with a high level object oriented programming language, networking facilities providing web-based use of methods and a set of tutorials. An experience using XploRe in a classroom for survey methodology has been related by M.Lejeune (1998).

3. Data bases

Maybe the most well known collection of datasets are The Data and Story Library known as DASL (<http://lib.stat.cmu.edu/DASL/>) and the Chance database which contains materials designed to help teach a "Chance course" or a more standard introductory probability or statistics course. (<http://www.dartmouth.edu/~chance/>). The Chance course is a case study quantitative literacy course developed cooperatively by three north-american universities. The aim of Chance is to make students more informed, and critical, readers of current news that uses probability and statistics as reported in daily newspapers such as "The New York Times" and current journals and magazines such as "Chance", "Science", "Nature", and the "New England Journal of Medicine".

It may be noticed that NSI's generally do not provide free online datasets, despite they are funded on public money.

4. From electronic textbooks to Internet courses

Electronic textbooks with hyperlinks have been written for some years and distributed as CD-Roms. The new generation on Internet provides of course web links and makes available the use of distant resources like Java applets. A typical example is "Multivariate statistics, concepts, models and applications" by David W. Stockburger (1998) from Southwest Missouri State University (<http://www.psychstat.smsu.edu/multibook/mlt00.htm>) where one can find a course on discriminant analysis, cluster analysis, multiple regression, analysis of variance and experimental designs.

The electronic textbook of StatSoft Inc (Developer of Statistica) deserves a special mention for its completeness (<http://www.statsoft.com/textbook/stathome.html>).

More ambitious is the UCLA electronic textbook called "Statistics. The study of stability in variation" developed by J. de Leeuw (1997) and his team (<http://www.stat.ucla.edu/textbook/>). The goal of the authors is to write a textbook freely available to everyone on the internet, independent of the level of the student, interactive, using graphics and demos and which covers most of the statistical theory. The contents of the book starts from an introduction, the analysis of a single variable, till multivariate analysis. The book provides in a very user-friendly way a statistics glossary, a statistics toolbox and Xlisp-Stat demos.

The other three noticeable projects concerns Introductory Statistics.

“SurfStat” (<http://surfstat.newcastle.edu.au/surfstat/main/surfstat.html>) started in 1994 with *“the brave but naive intent of making an existing set of course notes available online as hypertext”*. It has since grown to include an extensive glossary, interactive exercises, JavaScript functions replacing statistical probability tables, and the beginnings of a set of Java applets demonstrating statistical concepts through dynamic graphics. It is the primary learning resource for students taking a basic course at the University of Newcastle, Australia.

The goal of the project called “Virtual Laboratories in Probability and statistics” by Kyle Siegrist from the University of Alabama in Huntsville is to provide interactive, web-based modules for students and teachers of probability and statistics (<http://www.math.uah.edu/~stat/>). It covers geometric models, Bernoulli trials, basic probability and basic statistics. This project is divided into chapters, similar to chapters in a conventional book. Each chapter is divided into web “pages” similar to sections in a conventional book. The pages have three basic components: Hypertext, Applets, Data Sets.

The Claremont Colleges' WISE “Web Interface for Statistics Education” (<http://acad.cgu.edu/wise/>) concerns an Introductory Statistics course, especially for social sciences. WISE is conceived as a supplement to traditional teaching materials, addressing specific topics that instructors have difficulty in presenting in traditional classroom. The tool serves to promote self-paced learning and to provide a means for advanced students to review concepts. The goal of the authors is *“to provide easy entry to internet resources that can be used in support of statistics education and applications. The site includes the following: interactive statistics tutorials, answers to questions, a jump-off point for connecting to other resources in statistics on the Internet. Instructors will be able to utilize the WWW presentation in conjunction with their classes: to supplement their text assignments, to provide common examples and practice problems, where students can receive immediate feedback. Instructors can ask students to use the pages as a supplement for the textbook and classroom, and for homework”*.

5. Distance teaching and the St@tNet project

Despite of their quality, the former products are tailored for teachers' use in a classroom or for self-learning. Surprisingly, it seems that the traditional Distance Teaching Universities do not use the resources provided by Internet apart from e-mail and the possibility of downloading paper documents. There is consequently a potential for a combination of distance teaching and Internet statistics resources.

Partly funded by AUPELF-UREF, the Academic Agency of French-speaking world (<http://www.aupelf-uref.org/UVF/accueil.htm>), St@tNet is a part of the future “Université Virtuelle Francophone”. The purpose of UVF is to make teaching resources available to french-speaking students, of all over the world, but more particularly of southern countries, through internet and intranet systems. The contents of UVF will be freely accessible, but registration and some fees will be needed, both to the UVF and to an university of the network, in order to pass examination and get a valid credit. In this framework, a consortium of four academic institutions (CNAM, Agro-Montpellier, Lebanese University, Kenitra University) is developing a course in Introductory Statistics, which consists in an interactive book, with animations, where the student has to answer questions and solve exercises. This course is based on a previous experience with the CD-Rom “CNAM Media”, but the adaptation to Internet needed a complete rewriting. Downloadable facilities are offered for text courses, exams, softwares as well as links to interesting websites. The course will be offered in 1999-2000 as an on-line distance teaching at CNAM for continuous education. Students might register from anywhere and a tutorship will be organized with the following features: discussion forum with a virtual class-room, e-mail with a teacher for 25 students, round-up sessions organized locally every two months.



Figure 1. Provisional homepage of St@tNet

6. Concluding remarks

Its very low usage cost makes Internet a very attractive medium, especially for developing countries and distance teaching. Despite Internet's ability of connecting people, human teachers remain necessary, see Velleman and Moore (1996), and it is not sure that all concepts, especially the most abstract ones, could be only taught with new technologies. Internet offers teachers the possibility of comparing their practice with other ones and improving it, but will put them in competition: students, especially in continuous education, will soon have choice between courses from all the world.

REFERENCES

- De Leeuw J. (1997) The UCLA Statistics Textbooks and Modules 1997, ISI 51st session, Istanbul 55-59
- Lejeune M. (1998): Statistique et Internet, in a One-Day Conference of Société Française de Statistique, 11 rue P. et M. Curie, F 75005 Paris
- Lock, Robin H. (1998) : WWW resources for teaching statistics, <http://it.stlawu.edu/~rlock/tise98>
- Morin, A., and Saporta, G. and Spinakis, A. (1998): Computer Assisted Training in Statistics: Internet and Multimedia: a Survey of Existing Tools, in Proceedings of NTTS'98, Sorrento, Italy (eds. P. Nanopoulos, P. Garonna, C. Lauro), vol 2, 333-338.
- Velleman, D.S. and Moore, D. (1996) : Multimedia for teaching statistics : promises and pitfalls, *The American Statistician*, 3, 217-225
- West, R.W. and Ogden, R.T. (1998): Interactive Demonstrations for Statistics Education on the WorldWide Web, *Journal of Statistics Education*, <http://www.stat.ncsu.edu/info/jse/v6n3/west.html>

RÉSUMÉ

Internet fournit des ressources utilisables par les enseignants et les étudiants pour un enseignement interactif (données, applets, liens divers, cours électroniques en ligne), et peut devenir un vecteur de l'enseignement à distance: le projet St@tmet développé dans le cadre de l'Université Virtuelle Francophone de l'AUPELF-UREF en est une illustration.