

ST@TNET, AN INTERNET BASED SOFTWARE FOR TEACHING INTRODUCTORY STATISTICS

Gilbert Saporta

Conservatoire National des Arts et Métiers
France

This paper will report about the realisation and use of St@tNet, an interactive internet course on introductory statistics. St@tNet has been realised by a consortium of several French-speaking universities for distance education purposes. After a survey of the resources provided by Internet for teaching statistics, and especially of similar courses, we will present the product and its current use.

INTRODUCTION

The Conservatoire National des Arts et Métiers (CNAM) is a major public institution and an integral part of the French Ministry of Education, Research and Technology devoted to vocational training. The CNAM has a unique, three-fold purpose: teaching, research and conservation and was established in 1794 by the Convention in order to "enlighten ignorance that as yet knows not, and poverty which cannot afford knowledge". More than 70 000 adult students take its courses each year in various fields. What sorts of people study here? Mostly young people: average age 30, increasingly well qualified: two-thirds have had 2 years' higher education, one third are women.

Courses are given mainly in evening and Saturday classes, for credits towards a Degree, but also in-service training during working hours, and distance learning. CNAM is a network of 150 towns organized as follows: a large complex in Paris, 22 regional centres, plus some centers in overseas territories. Thanks to the CNAM network, one can begin one's course of study anywhere, and continue in any other French town. Over 7000 degrees are awarded each year, 1000 of them specialised Engineering degrees, others being Bachelor and Masters' degrees (equivalent to 2 or 4 years' higher education) in Social Science, Economics, Science and Technology, and postgraduate degrees.

The course on introductory statistics is present in most centers and there was a need to propose specific products for distance teaching and the opportunity to promote a modern way of teaching statistics. Modern teaching requires the use of statistical software, case studies, simulations, because statistics as an applied discipline is much more than a set of mathematical tools: the development of "statistical thinking" necessitates the understanding of concepts such as randomness, probable errors, risks etc. We had a previous experience of producing CD-Roms since 1990 and in 1998 we decided to transform completely our course into an Internet one; since Internet provides enormous resources for teachers and students (datasets, applets, free softwares, web links, on-line electronic courses) and we considered it as the best support for interactive education and distance teaching. St@tNet started thanks to a funding of AUF (Agence Universitaire de la Francophonie) the organisation of French-speaking universities throughout the world, who was willing to promote a "virtual university".

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 are subsequently 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 et 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. 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, database 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 information 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 (1998) from the TISE program (Technology in Statistics Education) (<http://it.stlawu.edu/~rlock/tise98>)

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 sites is WebStat (<http://www.stat.sc.edu/~west/webstat/>) created by W. West from the University of South Carolina. 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>).

Let us also mention the VESTAC collection of Java Applets of Leuven University for visualisation of statistical concepts www.kuleuven.ac.be/ucs/java/index.htm and the project developed by a team of INRIA about probability <http://www-sop.inria.fr/mefisto/java/tutorial1/tutorial1.html>.

It should be of a high interest to create a repository of such material or a multimedia catalogue as it is proposed by H.J.Mittag (2002) in this conference. 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 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. Win-IDAMS is a free software package for the validation, manipulation, statistical analysis of data developed by UNESCO <http://www.unesco.org/idams>. Besides downloadable software, one can process one's own data free of charge with distant software such as Webstat or Statlets (<http://www.sgcorp.com/statlets.htm>).

DATA BASES

It is very important that students use real data in a statistics course. Books on introductory statistics are often accompanied with a CD-ROM or with a website where one can find the data used in the book for re-use and practice. For an internet course, students should have access to collection of data: 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 most NSI's generally do not provide free online datasets, despite they are funded on public money. Let us mention however the efforts done by StatCan and by the INEI (Portugal) towards training of school students. The French Statistical System, with the CNRS is creating the "Institut Quételet": teachers and researchers will have a free access to a data base of surveys and a sample of 60 000 anonymous units of the census.

A SHORT SURVEY OF INTERNET COURSES

Electronic textbooks with hyperlinks have been written for some years and distributed as CD-Roms. The new generation on Internet provides 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 four 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 "Virtual Laboratories in Probability and statistic" by Kyle Siegrist from the University of Alabama in Huntsville aims 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.
- "Hyperstat" (<http://davidmlane.com/hyperstat/index.html>) is an interactive book with many links to other websites.

All the above-mentioned products are in English. In French there were no available products at the beginning of our project. Since that time, two courses have been released: MM*Stat developed in several languages by the Humboldt University of Berlin which is more of the kind of an electronic textbook (http://ise.wiwi.hu-berlin.de/~scripts/mdstat/mm_fre/cdrom/start.html) and SEL (Statistique en ligne) (<http://www.inrialpes.fr/sel>) which only covers a part of St@tNet.

Most of 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 all the resources provided by Internet apart from e-mail and the possibility of downloading paper documents. One exception is the multimedia course (in German) developed by H.J. Mittag at FernUniversität, Hagen (<http://www.fernuni-hagen.de/STATISTIK>). There was consequently a potential for a combination of distance teaching and Internet statistics resources.

THE ST@TNET PROJECT

St@tNet (demo web site: <http://www.cnam.agropolis.fr/statnet>) is a complete, interactive, web-based, course on introductory statistics with six modules: data description, probability, random variables, sampling and estimation, tests, linear regression. The course corresponds to a classical credit of 50 hours of teaching.

St@tNet is composed with three main parts:

- A course where each chapter is itself divided into four parts a presentation, a development, a summary, interactive exercises, a glossary, ...
- Downloadable files: course notes, exercises, datasets
- A set of tools including links, statistical tables, a glossary.

Each chapter is accompanied by a short video introduction (about 2 minutes) presented by the teacher who was in charge of this part.

As we already stated it, St@tNet (actually the first chapter on data description) was first funded by a grant of AUF to be a part of the UVF "Université Virtuelle Francophone". The purpose of UVF was 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. UVF project has been renamed "NTIC and knowledge acquisition" in order to avoid any confusion: AUF is not creating a new distance university but provides resources to its members. The French Ministry of Education selected our project to be funded, which allowed the team to develop the other chapters. The project is now completed and the first release has been delivered to french authorities in January 2002.

Since St@tNet has been financed on public funds its contents are freely accessible to anyone and the product will be distributed to universities. But St@tNet is conceived as a part of a distance teaching organisation which includes tutorship, discussion groups, assessment. St@tNet being platform independent (we use only HTML, Java and Flash animations) may be integrated in any learning system like WebCT for instance. Any French speaking university could thus include St@tNet in a distance teaching offer: of course students willing to pass examination and get a valid credit have to pay specific fees. These fees enable them to take advantage of the tutorship.

At CNAM the course is offered since 1999 as an on-line distance teaching for continuous education. Each regional centre of CNAM is able to propose this course. Students may register from anywhere; they receive a CD-rom with the full contents and a password to have access to specific services : tutorship with discussion forum in a virtual class-room, e-mail with a teacher for 25 students. Registered students have also access to previous exams subjects. We deliver a CD-rom to registered students since it is not necessary to be connected to the Internet all the time, which would be too costly for most students. Connections are necessary only for activating links, like external Java applets, and for online services. Moreover downloading video presentations is still very low.

Figure 1 presents a few screens, but of course only the demonstration of the product may reveal interactive aspects and animations. Most of the animations were written with Flash and have the purpose of illustrating concepts in a (hopefully) pleasant manner. It is necessary to alternate course, questions, exercises, animations in order to have an attractive and interactive product. It is to be noticed that the mathematical proofs which are too long to be presented on the main screen are in the form of external files or pop-up windows. We have also developed some specific applications in the form of Java applets, in order to perform simulations and that we also advise have put links to applets already written by other authors (see a previous section).

We have programmed the analysis of answers to questions boxes, but the student is never stopped in his learning progression: we are not in favour of systems where one can go to a part of a course only if one has succeeded to understand the preceding part. It is our experience that most adult people refuse this way of teaching.

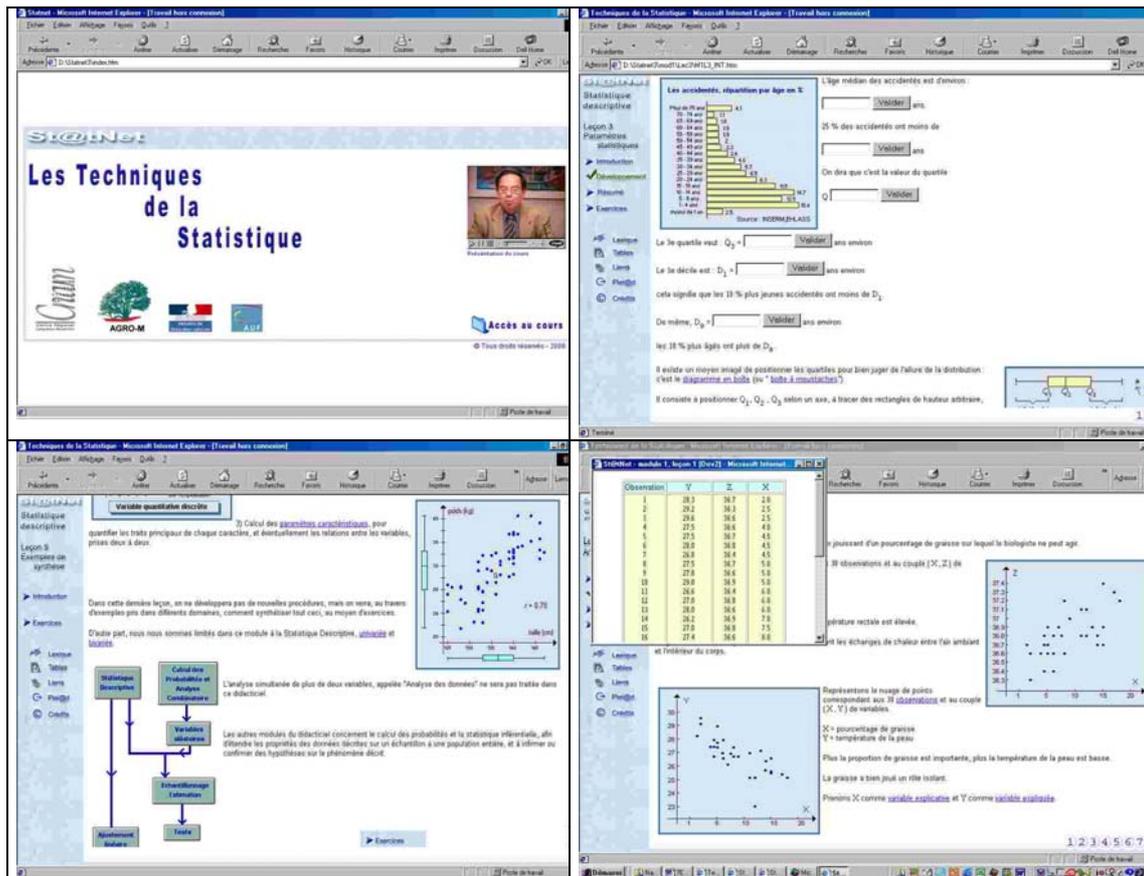


Figure 1. Examples of Web Pages.

CONCLUDING REMARKS

St@tNet is the first complete course on introductory statistics available in French on Internet. It is our opinion that the use of multimedia and distributed resources gives new possibilities to statistical teaching and makes it more attractive. Our product was designed for distance teaching purposes, but can be used also in classrooms and as a complementary resource for students. First experiences at CNAM show that a new audience has been met: people who would not be students in evening courses have registered. Success to the examinations is similar, if not better than for a classical course. A selection bias is likely to exist, since people connected to internet have a higher level of education than the average population of our evening courses.

Another potential audience for our course is also the population of teachers in secondary schools, since most of them need to complete their training in probability and statistics which have been recently introduced in school curricula. We hope that new funding and cooperation will make it possible to complete this introductory course with new chapters (time series, quality control, experimental designs, multivariate analysis).

ACKNOWLEDGEMENTS

St@tNet is realized through grants of AUF and of the French Ministry of Education and is authored by F.Kazi-Aoual, M.Janvier, under the scientific supervision of G.Caraux (Agro-Montpellier), G.Saporta, A.Brethon and E.Verdoire (CNAM-Languedoc-Roussillon). G.Chirici, M.Marco, F.Pintus did the programmation work. Other partners are Kenitra University, Morocco (Y.El Kettani), Lebanese University (M.Hakim), Yaoundé Polytechnic School (H.Gwet).

REFERENCES

De Leeuw, J. (1997). The UCLA statistics textbooks and modules. In *Proceedings of the 51st Session of the International Statistical Institute* (pp. 55-59). Istanbul.
 Mittag, H.J. (2002) *Java applets and multimedia catalogue for statistics education*. In ICOTS6 proceedings.

- Morin, A., Saporta, G., & Spinakis, A. (1998). Computer assisted training in statistics: internet and multimedia: A survey of existing tools. In P.Nanopoulos, P.Garonna, and C.Lauro (Eds.), *Proceedings of NTTS'98, Sorrento Italy, Vol. 2*, (pp. 333-338).
- Velleman, D.S., & Moore, D. (1996). Multimedia for teaching statistics: promises and pitfalls. *The American Statistician*, 3, 217-225.
- West, R.W., & Ogden, R.T. (1998). Interactive demonstrations for statistics education on the WorldWide Web. *Journal of Statistics Education*. Available at:
<http://www.stat.ncsu.edu/info/jse/v6n3/west.html>