

THE ST@TNET PROJECT FOR TEACHING STATISTICS

Saporta G. and Bourdeau M.

Key words: Teaching, Statistics, Information Society.

COMPSTAT 2004 section: Teaching Statistics.

Abstract: This paper describes the design and implementation of St@tNet, an internet environment for the teaching of basic Applied Statistics. St@tNet has been developed by a consortium of French-speaking universities. After a few general considerations on education for the Information Society, and more specifically for the teaching of Statistics, we will present our product in its present state of development.

1 Means and ends

The title of this session is about teaching Statistics for the Information Society. Well, the Information Society began with the invention of the printing press with moveable type, and that has profoundly modified the formal education process. In essence, it has permitted widespread knowledge dissemination. For a few centuries things have stayed more or less the same, until the invention of mass media. Starting with the radio, then television, it became apparent that the world had once more profoundly changed. Their consequences in the formal and informal education processes were no doubt far-reaching, but now that we have entered the computer age, we have passed into speed *Warp Five* to speak StarTrek lingo; in the last few years the Internet development has brought about a genuine revolution in education thinking, actually a totally new *zeitgeist*.

Neil Postman [1931-2003], one of the keener observers of the evolution of education, and of society in general, has fully explored the consequences of this information revolution [2, 3]. He reports, and it is a common observation, that the situation of teachers and professors has become precarious: they are worried, even anxious, about their role and their immediate future in the Information Society.

Topping all this, governing bodies in many of the developed countries have nowadays become obstinate in drastically reducing budgets, with the elusive hope that the new technologies will give rise to an unprecedented increase in productivity: lesser means, greater expectations... Illusion, reality, who can tell? And what is the end of education finally?

Dwindling budgets are especially noticeable in higher education, where the knowledge transmission process in itself has lost ground to research, knowledge production, essentially applied, where most of the time the bottom line is the amount of hoped for funding and money gains. "Glory" for universities comes from their research productions, not their teaching quality!

At first sight, it might appear that the new Information technologies (ITs) will lead to the end of the profession: only a few great courses will remain, developed and maintained, like computer software, by superior experts and educators, appearing through virtual media with a very widespread audience across the whole world: think here in terms of added sales and to what amounts this could add up! At this journey's end, all the transmission of knowledge would originate from a few specialized quarters far away from students, the "new" teachers and professors remaining for good in their ivory towers, pedagogical encounters would be virtual, the Internet being the sole communication channel. Universities and colleges would supply themselves for knowledge transmission and certification from those virtual hyper-classrooms.

ITs could be a boon for education, a giant economy for education boards, but could entail the disappearance of most teachers and professors.

From an overview of some recent and very successful pedagogical experiments in Québec universities using ITs, one can suspect that things will not be that simple [1]. The same situation, it is easy to confirm, is prevalent the world over. Actually, getting an education is a form a travelling. And quality travelling often imply personal guides, at least human encounters, not just guidebooks and TV documentaries though they can be illuminating and irreplaceable. In our experience, all the pedagogies devised with the ITs in mind have always implied more personal contacts with students, less mass dispensing of knowledge[1]...

Maybe with the Internet have we entered an era of renaissance for the true pedagogical relation, not the opposite. Generally speaking, we could very well be on the way of building a web of human interactions much more intense than ever in History. The brain consists of billions of synaptic links between neurons, the Internet world could very well consist of billions of synaptic links between brains. Is this not what is meant by the Information Society? This has far reaching implications, as we will see, for teachers and students reciprocal relations.

1.1 Teaching Statistics in the Information Age

Concerning Statistics and Data analysis, there is no gloom and doom scenario in view: there is a huge increase of informations that have to be processed. As John Wilder Tukey [1915-2000] has so correctly noted "The best thing about being a Statistician is that you get to play in everybody's backyard."

Better tools of analysis are badly needed, and, since there is already a widespread availability of data sources and an increased appetite for synthetic information, an important increase in Statistics literacy is urgently needed

[1] All the relevant documents upon which rests this assertion and that have been used for [1], are located on the following web pages
http://www.mgi.polymtl.ca/marc.bourdeau/InfAgeTeaching .

for an ever increasing number of people. Think, among other things, of the amount of information stored and available in official Statistics Offices the world over. All newspapers and mass media are now replete with reports of polls, of official statistics on the economy and society in general. Think also of the huge amount of business informations stored in Data Warehouses that come with an abundance of *Data Mining* softwares recently marketed. Making sense out of this "chaos" [6] is a huge undertaking. We are heading towards a knowledge-based society where statisticians will be ever more in demand.

So, much more educational needs than ever are emerging, not only for the mass media personnel and specialists of data management, but also for the whole population in general. And clearly, many new means ensuring a much increased level of knowledge have to be devised. Already, most of the new pedagogical material created now use the indispensable Internet facilities and new technologies.

We report here on the education material for the teaching of Statistics produced in our universities. See Saporta [4] for an overview of some of the web facilities for the teaching of Statistics[2]. See also the remarkable paper by Velleman & Moore for the ins and outs for the use of ITs in the teaching of Statistics [7].

2 The St@tNet project

The St@tNet project is developed at the *Conservatoire National des Arts et Métiers* (Cnam), a major public institution for continuous education and an integral part of the French Ministry of Education, Research and Technology. The Cnam was founded in 1794 to "enlighten ignorance that does not yet know, and poverty which cannot afford knowledge." More than 70 000 adult students attend its courses each year in numerous fields, two-thirds of them have already had two years of 'higher' education, one third are women.

Courses are given mainly in the evenings and in Saturday classes for credits leading towards a degree, as well as through in-service training during working hours, and, finally, through distance-learning. The Cnam links a network of 150 towns and is organized around a 'main' complex in Paris, 22 regional centers, plus some centers in overseas territories. One can begin a program anywhere in the network and continue in any other center. Graduate studies leading to Masters and PhDs are available in many disciplines.

St@tNet follows a series of previous developments of teaching materials for introductory Statistics that date back to the early nineties. Previous courses were available on diskettes and CD-roms [5]. The actual web-course version was financed by the *Agence Universitaire de la Francophonie* (*AUF*) and the French *Ministère de l'Éducation Nationale*. It is operational since 2002, and can be obtained also on a CD-rom version.

[2] Also available in the web pages just referred.

St@tNet is the only web resource proposed at the Cnam for distance learning for the much needed Introductory Statistics. It is freely accessible[3]. Indeed, having been financed by public funds, and for the advancement of public learning in conformity with its founding principles that go back to the Enlightenment Age, the decision of the free access of St@tNet was finally agreed upon after fierce debates, but registering at a cost of 250 euros is mandatory for certification purposes and the use of usual facilities: tutorship (one tutor per 25 students), an Internet access on a virtual teaching environment (VTE), an e-mail, etc. This fee comprises the CD-rom that avoids most of the Internet costs and waiting times, especially in distant locations. St@tNet is now also implemented on the virtual campuses of the *Agence Universitaire de la Francophonie* where it is one of the two most popular resources for self-education. Starting in the Fall of 2004, the Cnam will organize a certification system for the AUF courses. St@tNet is a complementary resource for the *École Militaire*, it is also recommended by the French association of mathematics teachers as an aid to school teachers who have to adapt themselves to new curricula that include elements of Probability and Statistics.

With its network of institutions, the Cnam is an ideal ground for the development of pedagogy and teaching material using ITs. Modern teaching of Applied Statistics requires the use of specialized software, and should be data based, centered on case studies for more advanced material and hands-on training. Applied Statistics is indeed much more than a set of mathematical formulae: its learning implies the development of "statistical thinking", requires the understanding of difficult concepts such as variation, randomness, laws of chance — a difficult oxymoron at first glance —, probable errors, risks, etc. Animations and various graphical tools provide efficient means of learning.

Depending on the level, one can think of various designs for the Internet environments and interactions. Up to now, there are two stages planned in the St@tNet project, the first one is fully operational, the second in development, but with partial versions tested in ordinary classrooms.

For the first stage, at the very basic level of statistical knowledge, St@tNet has opted for a complete Html environment. The advantage of this choice is that interactions of the students with the environment are quite easy to realize: this course is by no means a paper-course translated into Html, as one can still see quite often, but a full-fledged Html environment with frequent short interactions inserted by design into the course.

For higher levels of knowledge, where short interactions are much less needed, St@tNet has opted for a downloadable Latex-Pdf text, with full hyper-referencing possibilities, and many of the hyper-references are internal.

[3] http://www.agro-montpellier.fr/cnam-lr/statnet/.

2.1 First stage: the basics

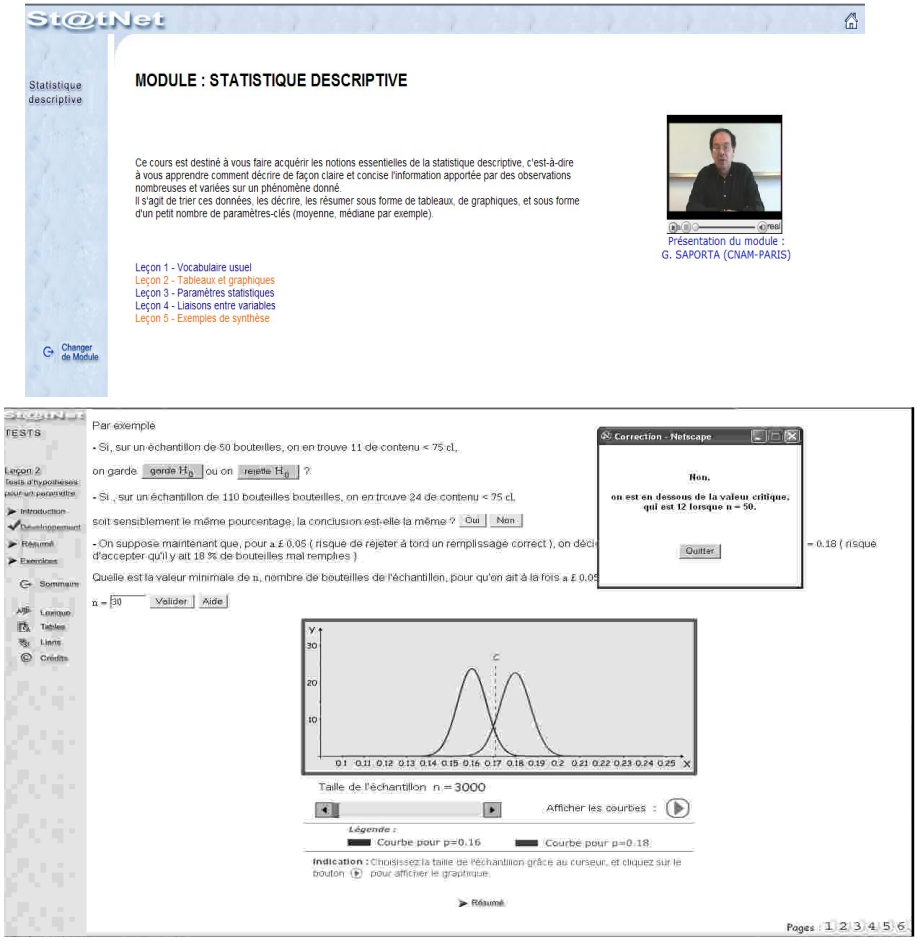


Figure 1: Upper: The entry for the module *Statistiques descriptives* (Descriptive statistics), with its introductory video. Lower: Part of the development section for Lesson 1 of the module *Tests* (Tests), with a pop-up window obtained with a wrong answer.

The first stage of the project, the one for the really basic knowledge, is now fully operational. It consists of six modules: data description, probability, random variables, sampling and estimation, tests, basic linear regression. Each of the modules is introduced by a video file (Figure 1, upper part) and is composed of lessons, all of which are of the same structure: Introduction, development, synopsis, exercices. A glossary of terms is accessible within each lesson, as well as all the necessary Statistical Tables and Internet links.

Once in a module, and after viewing its presentation video, the user can pick a *lesson* of his choice: indeed, the learning progression is not designed with a linear structure in mind. Most of our students detest such a progression that do not correspond to their needs.

The lower part of Figure 1 shows part of a page of the *Développement* (development) section of *Leçon* 1 (Lesson 1) of the module *Tests* (tests), with the pop-up window produced when a wrong answer is given by the reader. Upon a wrong answer, the reader can either correct his answer or get the right one with a short explanation.

Similarly to what is represented in this last Figure, lessons are interspersed with questions to the reader to check if the elements of learning have been correctly assimilated. The pages are designed accordingly. All lessons end with a page of summary (Figure 2), and a few more elaborate exercices, again with answers given directly on the page, with pop-ups for feedback. A pop-up Glossary, the same for all lessons is hyper-referenced, and, finally, a page of links is available, with some of them referring to Java applets useful for the learning.

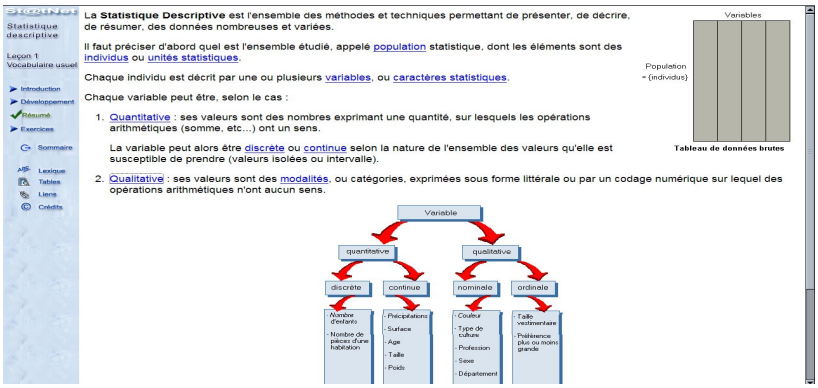


Figure 2: The summary page from Lesson 1 of the module *Statistiques descriptives* (Descriptive Statistics).)

A new audience has been reached by this approach, and the rate of retention and success is better than for traditional courses. This last point might be the consequence of the type of students (a "sampling bias" !) interested in such an environment.

2.2 Second stage: applied linear models

After the first stage of the project was carried out, and after a decision was made to embark on a large project concerning applied linear models, consisting of the standard curriculum completed by methodologies for categorical data, like the logistic and log-linear models, reflection was given as to what

format would be appropriate for more advanced learning.

The advanced learner of a given discipline, especially at the Cnam, has very different needs than the learner of the elements. More often than not, a first course in Probability-Statistics is mandatory. A second course is taken by those who feel a greater relevance of the material taught to their actual work. Hence a truer motivation. In any case, to get the attention of a student, any student, one has to pay heed to its needs, to speak his language.

In Applied Statistics, the actual practice requires the continuous use of a Statistics software on real data — real as opposed to simulated, with all the complexities then of reality —, and an important part of the work consists of careful questioning from the analyst and writing of the facts found during the process.

All this points to a pedagogy that rests principally on case studies, probably the natural points of entry to the curriculum for many students in the engineering and management sciences to whom this course is destined. Theory, the mathematical derivations — and they show a complexity far beyond that found in the elements —, are seen as answers to specific questioning on their part. Thus bigger and more mathematical chunks of material in more advanced studies, instead of the tidbits of the elements.

Another important point in our view of things, there should be a constant preoccupation from the designers of Internet courses, all courses for that matter, to instill into the students the art of questioning. We refer here also to Postman in his last essay ([2] p.161 *seq*): "(...) question-asking is the most significant intellectual skill available to human beings," and it is extremely strange that, especially in the Sciences whether hard or applied, it is not taught in schools!

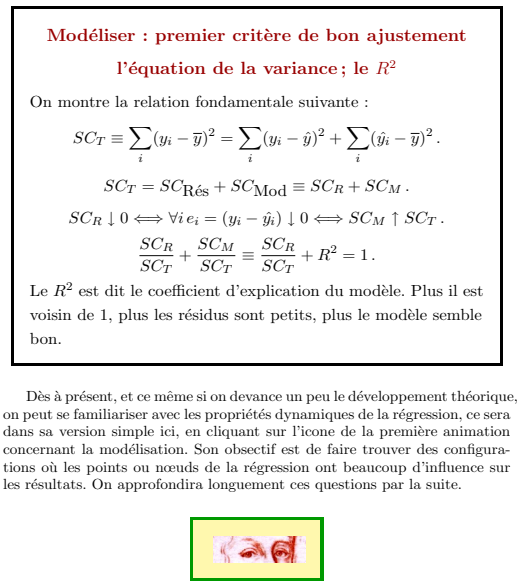
Finally, and this also harks back to Postman in all of his books on Education, we have written historical notes on all the principal aspects on the origin of the need of statistical models for reality. It is a fact that there is with History notes a sort of holographic phenomenon: even when one starts from hard sciences' bits of knowledge, exploring how things came to be, where ideas came from and how we came by them, provides, if propelled by a sense of questioning, an insight on the whole of societies, on all of Human nature. This constitutes an essential part for any formation. Education after all is not only about information, but first and foremost about the formation or casting of minds, young ones in particular.

In summary, due to the mathematical sophistication of this material there is a need for textbook typography, as well as, as usual, a need for a complete system of inner referencing and outer or hyper-referencing facilities. This leaves nowadays almost no choice: such a course must be written in Latex-Pdf typeset. The Pdf-files are virus-proof, they can be readily printed on paper with textbook color quality, their use on computer screens is very confortable, moreover providing some annotating facilities, and, finally, inner links and hyperlinks are manipulated with extreme ease.

This second stage of St@tNet is not, as yet, fully operational, but a demo-version ia available, and parts of the material, especially some case studies, were tested with great success in standard classrooms[4]. In the following pages we present some its highlights.

In Figure 3, we can see part of ordinary page of the course file. At the bottom of the page an icon referring to a Flash animation, an image of which appears on Figure 4.

The reader can flip back and forth from any page giving internal links to an equation, a table, a figure. He can also, if he subscribes to an Internet server, readily access a certain number of hyper-links to whatever sites deemed interesting by the authors. These pages will be added automatically at the end of the pdf-file file which can be saved with the added information. The Adobe-reader provides also various facilities to annotate the file pages.



Modéliser : premier critère de bon ajustement

l'équation de la variance ; le R^2

On montre la relation fondamentale suivante :

$$SC_T \equiv \sum_i (y_i - \overline{y})^2 = \sum_i (y_i - \hat{y})^2 + \sum_i (\hat{y}_i - \overline{y})^2 .$$

$$SC_T = SC_{\text{Rés}} + SC_{\text{Mod}} \equiv SC_R + SC_M .$$

$$SC_R \downarrow 0 \Longleftrightarrow \forall i\, e_i = (y_i - \hat{y}_i) \downarrow 0 \Longleftrightarrow SC_M \uparrow SC_T .$$

$$\frac{SC_R}{SC_T} + \frac{SC_M}{SC_T} \equiv \frac{SC_R}{SC_T} + R^2 = 1 .$$

Le R^2 est dit le coefficient d'explication du modèle. Plus il est voisin de 1, plus les résidus sont petits, plus le modèle semble bon.

Dès à présent, et ce même si on devance un peu le développement théorique, on peut se familiariser avec les propriétés dynamiques de la régression, ce sera dans sa version simple ici, en cliquant sur l'icone de la première animation concernant la modélisation. Son obsectif est de faire trouver des configurations où les points ou nœuds de la régression ont beaucoup d'influence sur les résultats. On approfondira longuement ces questions par la suite.

Figure 3: Part of a typical page in stage two, with the icon referring to a Flash animation.

Many Flash animations are included in the system. They constitute a remarkable tool to ease the learning. Each one of them comes with a certain number of controllable buttons one of which is an audio file. In the example (Figure 4), the nodes of the regression are mobile and new nodes can be added, the confidences bands resulting from the least squares results have a button to control their level, and whenever a change is made, directly with the mouse on the computer screen, the new regression line and confidence

[4] http://www.mgi.polymtl.ca/marc.bourdeau/InfAgeTeaching .

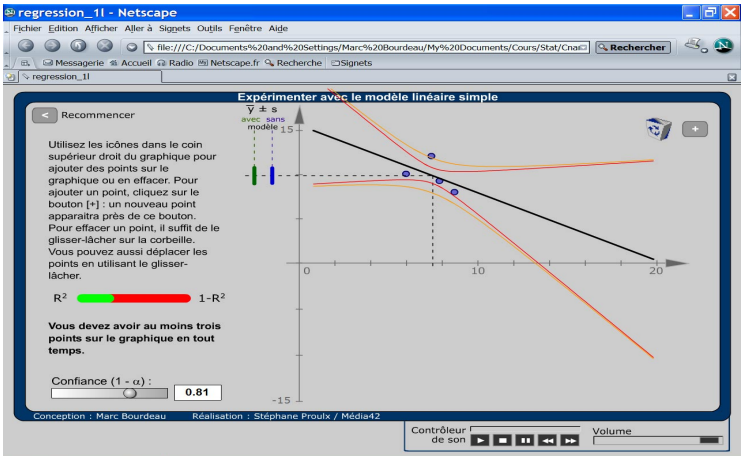


Figure 4: A page from one of the course Flash animations, with its controllable buttons, one of which (bottom) is for an audio file.

bands with the other numerical parameters promptly appear on the screen. The audio file provides instructions for the use of the animation, a few explanations, and always, this is very important, a questioning that the animation brings out.

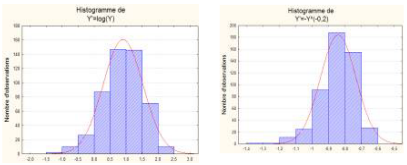


FIG. 5 – Diverses transformations de Y (Fig. 4) : à la suite $Y^{0,5}$, $Y^{0,3}$, $Y^{0,25}$, $Y^{0,2}$, $\log(Y)$, $-Y^{-0,2}$.

Exercices.

1. Utilisez les données de cet exemple (cliquer ci-contre) pour vérifier les effets des transformations (observés comme sur la Fig. 5 et le Tab. 1) sur des sous-échantillons de quelques dizaines de sujets.


Data

Figure 5: A typical page of a case study, with an icon to import the data.

In Figure 5, we show a typical page of a case study. Remember: case studies are the backbone of our pedagogy. A case study is usually several pages long and is built along a certain questioning on a data set that is usually quite complicated. Its flow is very progressive, and generally requires a few dozen hours of work with the writing of a roughly 20 page report.

This task imperatively implies team work. And true collaboration is necessary: a case study is not composed of a certain number of unrelated problems, like the standard homeworks found in most curricula, but has a synthetic character where each part responds or resonates with other ones. There is a wide landscape built in every case study, several chapters of Statistics are brought to bear. This precludes the "usual" split up of the work... The work required is similar to actual data analysis required by engineers and scientists, and ordinary work for that matter.

In a standard classroom, students often have a natural peer group and team formation is quite easy — though there are more optimal ways for this selection —, but for distance learning things are not that easy. However, in most organizations nowadays, the new ITs (chats, forums, etc.) already allow team virtual meeting and working that take up a very large, if not the larger part of the work process!

Students tend to use all the modern hyper-communicating computer facilities that are usually available on all recent computers, as well as within the VTEs in use in most universities. In most North American universities students are in constant Internet contact with their peer-groups, almost day and night, sending each other files of their works, of their thoughts, comments on the courses, etc. And Internet real time voice communication facilities are rapidly spreading. Writing, however, thanks to the Internet, has regained much luster: it constitutes indeed an essential tool for the unveiling of one's real thoughts. Team work is constant. On a less bright side, homeworks and exams tend to be freely accessible to all...

Professors must adapt to this situation. The Internet has and will profoundly change the learning world. Thus ITs can compensate the isolation of students of the past in distance learning, as they have already done so in the traditional classroom. But our type of pedagogy implies a much greater contact not only horizontally, from students to students as we just noted, but also vertically. For professors of Statistics, it is not the transmission of bits of knowledge that will constitute their main task, it is the statistical cast of mind itself that will be more and more the focal point of teaching. And statistical thinking, like all casts of mind, is best transferred by apprenticeship.

3 Conclusion

The Information Age offers mind-boggling perspectives and cannot but have a profound impact on the pedagogy of whatever discipline there is, but first and foremost for those that present a technical character, and Statistics is one of them. All the presentations in this session will no doubt show the diversity of options.

The end of the journey for teachers? At first sight, it might appear that all these new facilities lead to the disappearance of teachers and professors. But many very successful pedagogical experiments have shown that human

pedagogical guides are more necessary than ever, and that ITs provide an indispensable structure for more interactions between them and the students. Il would not be surprising that the new pedagogical paradigm would be that of apprentices and masters. In all the pedagogical experiences we have seen, not only in Statistics, not only our own, there is a greater need than ever for human personal transmission. The role of professors becomes more and more that of a *personne ressource*, a guide so to speak, and less and less that of a knowledge dispenser. Pure knowledge transmission is not the principal role of professors anymore: this has now been more or less automated thanks to the new ITs. Transmission is required now at a much higher cognitive level. And written words for their precision, as well as oral contacts, play — ITs in the background again! — a crucial role. On the Internet, all courses tend to become tutorials! And this is the expensive form of teaching... It can explain why the pedagogical interaction has become so much more demanding than ever.

On the other hand, many governments nowadays tend easily to believe that education and other public services are not of primary importance and cost too much. For reasons of globalization and so forth, they preside over decreasing public spending. The Internet Age can readily provide very low quality and very low cost formative material — garbage-in, garbage-out —, as well as higher than ever quality education. The latter being the kind needed in an increasingly complex world. But the wheel of Fortune spins faster than ever, and, it has always been the case, the outcomes are not totally random: the better educated no doubt reap the profits.

The question of what's in store for pedagogues in the future will be with us for some time.

What about St@tNet's journey? The conception, development and implementation of St@tNet required considerable resources, human as well as financial. The end product could constitute a complete curriculum in French for Applied Statistics.

The first stage was conceived with a playful spirit in mind, to which the elementary concepts of Probability and Statistics lend themselves fairly easily. But putting it into service required a considerable amount of work, so much more than the writing of a standard chalk and blackboard course, or of a set of telegraphic computer slides. The second stage is much more difficult to conceive if one does not care for a standard run of the mill product, but strives after something more pedagogically efficient. The deeper one goes into the discipline, the more difficult the task.

The question is not whether or not there will be a need for St@tNet or its successors in the future, but what form they will take, and what resources will it be necessary to put into action? A knowledge-based society will indeed bring no dearth of work for statisticians and teachers of the discipline (cf. the 6th European research program FP6). However, at the same time that technology's pace shows no sign of slowing down and that the demand

is growing rapidly, the human and financial resources might become more difficult to muster... International cooperation and sharing of the new IT products catering to the needs of students of Statistics as well as greater imagination and dedication on the part of teachers of Statistics will no doubt be necessary.

References

[1] Bourdeau M. (2003). *L'enseignement supérieur et les TICE (Technologies de l'information et des communications en enseignement). Big bang ou méga flop ?* Conférence pour l'inauguration de la mission TICE, Université de Bretagne Sud, 4 mars 2003. http://www.mgi.polymtl.ca/marc.bourdeau/InfAgeTeaching.

[2] Postman N. (1999). *Building a Bridge to the 18th Century. How the Past can Improve Our Future.* Alfred A. Knopf, New York.

[3] Postman N. (1995). *The End of Education. Redefining the Value of School.* Alfred A. Knopf, New York.

[4] Saporta G. (2002). *St@tNet, an Internet Based Software for Teaching Introductory Statistics.* Proceedings Icots 6, Sixth International Conference on Teaching Statistics, Capetown, July 8-12 2002.

[5] Saporta G. & Morin A. (1996). *Interactive software for Learning Statistics.* Compstat 96, Barcelona, 26-30 August 1996.

[6] Serban A.N. & Luan J. (2002). *Overview of Knowledge Management.* New Directions for Institutional Research, **2002**(113), 5 – 12.

[7] Velleman Paul F. & Moore David S. (1996). *Multimedia for Teaching Statistics: Promises and pitfall.* The American Statistician, **50**(3), 217 – 225.

Acknowledgement: St@tNet was developed with the generous support of the *Agence Universitaire pour la Francophonie* and of the French *Ministère de l'Éducation Nationale*. Deep gratitude to all those who are working on the project.

Address: Gilbert Saporta is *Titulaire de la Chaire de Statistique Appliquée*, at the *Conservatoire National des Arts et Métiers*, 292, Rue Saint-Martin, F-75003 Paris, France. http://cedric.cnam.fr/∼saporta.

Marc Bourdeau is Professor of Applied Statistics at the *École Polytechnique de Montréal*, http://www.mgi.polymtl.ca/marc.bourdeau.

E-mail: Saporta@cnam.fr ; Marc.Bourdeau@polymtl.ca