

NETWORK FACILITIES FOR TEACHING STATISTICS:
A DATA ARCHIVE DRIVEN BY INTERACTIVE SOFTWARE

Gianfranco Galmacci and Giuseppe Cicchitelli
Dipartimento di Scienze Statistiche - Perugia, Italy

The aim of this paper is to illustrate a project that is going on as a cooperation among the Department of Statistics of the University of Perugia, CIRDIS and ISTAT (Italian National Institute of Statistics) for the creation of data archives to be used as a basis for practice on the main topics of a first course of statistics.

The archive has been partially set up and it is interfaced by many programs showing the main characteristics of the data as well as the information they contain. A short documentation (meta-data) is available for each data set as well as a discussion of the main results of the analysis suggested. This service is available on the World Wide Web.

INTRODUCTION

Computer technology has recently been heavily involved in new proposals and initiatives for teaching statistics. In fact, hardware and software facilities can represent a valid and useful support for transferring many statistical concepts to students by using approaches that are easy, quick, efficient and, often, fun. However, in general teachers need good and appropriate software and a wide set of data as the basis for covering the various topics of statistics. In many countries, today, secondary schools and university departments are connected to the Internet, this makes the dissemination of software and data possible for any purpose, including teaching.

Recently CIRDIS (Inter-university Centre of Research for Teaching Statistics), the Department of Statistics of the University of Perugia, Italy, jointly with the Italian National Institute of Statistics (ISTAT), have set up a new project to create a WWW data archive for teaching purposes. This archive contains many kinds of selected data for student practice and takes account of the main statistical methods. A further contribution of this project is an interface that allows a WWW browser to use statistical software (such as XlispStat) to explore the main characteristics of each data set, either graphically or by computing some basic statistics. The interface, by means of a friendly dialogue on the browser screen, will allow the user to carry out a preliminary analysis even if she/he does not know how to use the specific statistical software.

MOTIVATIONS AND PURPOSES

With the modifications in the Italian educational system several years ago, the teaching of statistics was introduced into many types of schools at the pre-university level; this discipline is usually part of traditional mathematics courses, although it is sometimes presented as a separate subject. However, in many cases this initiative has not met with an adequately prepared teaching staff to carry out this new task. Indeed, teaching of statistics in secondary schools is usually entrusted to mathematics teachers who have most often obtained degrees in subjects (mathematics, physics, engineering or some type of science) in which statistics examinations do not exist, are not compulsory or are either generally very theoretical (mathematics) or oriented toward applications in the sphere of a specific subject (physics, engineering, etc.). Therefore, the average teacher's level of preparation is neither enough to allow him to have an organic and operative vision of statistical methods (as should be required of a secondary school teacher) nor to have the knowledge necessary to analyse data and interpret results. Even worse, in this situation themes such as *official sources and data quality* are almost always unknown. All this often determines inadequate teaching and, therefore, a lack of practical application which could help students to become familiar with the instruments they use and to understand the nature of statistics and its role.

Other evaluations must be added to the preceding considerations, which constitute some of the main reasons for this project, that are of no less importance. Introductory university courses require a great amount of possibly up-to-date data to use as examples in the classroom to support theories, for exercises or for the preparation of examinations.

Beyond to main purpose of the course, a suitable choice of data allows the progressive acquisition of knowledge regarding the dimensions of certain phenomena that characterise our society. Thus, the student can become acquainted with the magnitude of certain phenomena, with the way they evolve, how they are distributed throughout the territory, etc. This is one of the results most hoped for in an introductory course of statistics.

In the light of these considerations, CIRDIS and the Department of Statistical Sciences of the University of Perugia have undertaken the initiative of activating a service for teachers of statistics with the aim of facilitating their task by putting data at their disposal which have been appropriately selected, documented and equipped with specifically developed software. The project has been promoted in collaboration with ISTAT (Italian National Statistical Institute) and is well underway.

THE DATA

As the purpose of the project is to make *real data* available for an introductory course of statistics and to provide a clear idea of the Italian situation by means of official statistics, the criteria with which the user may wish to have access to the data were kept in mind while choosing and organizing the archives: *by statistic topics* and *by type of phenomenon*. Using the first method, the user can have direct access to the series of data which best present the type of methodology he intends to use. Indeed, data selection was carried out keeping in mind the various types of statistics and the teaching methods used in Italian schools: *graphic representations, statistical ratios, index numbers, simple and multiple distributions, mean values, measures of variability and concentration, measures of dependence, correlation and regression, probability distributions, parameter estimation, hypotheses testing*.

Furthermore, archives with the original records of a subgroup of censored data are made available. Samples can be easily extracted from them for any purpose.

The second method for accessing data was designed with the intention of providing a *picture* of Italy from a statistical point of view, bearing in mind several sectors of interest created by ISTAT for the dissemination of data: *population, health, culture (education), family and society (consumption), justice, national accounts (revenue and production), labour, prices, services (tourism), foreign trade*.

The selection of one of these sectors allows access to the most significant series of data, the ones able to describe its main aspects. For every type of data, historical series are made available whose length may vary according to the nature of the data.

The series chosen refer to data gathered at the regional level. This choice was prompted by objective considerations: first of all, regional series have a limited dimension and are therefore easier to read and analyse; secondly, such information is often encountered in the discussions dealing with the country's problems promoted by the media. Therefore, it is more suitable to consider data collected at this level, rather than great issues, as it is more stimulating for secondary school students.

CHARACTERISTICS OF THE SERVICE

The archives will be made available through World Wide as well as through traditional channels (diskettes, CDRom). However, as obsolete data would thwart the

objective of supplying up-to-date images of phenomena, the WWW will be the preferred one. This choice seems right in spite of the fact that at present the number of schools and teachers that have access to the Internet in Italy is not particularly high. However, one can easily foresee that the trend of growth and diffusion of the network will progressively and rapidly increase.

In any case, making data available is not enough to supply teaching support when teachers have not received specific instruction. Sometimes they find themselves in a situation where they must teach a subject which is not only limited to theoretical aspects but rather assumes importance at the moment of its application (and this is particularly important in introductory courses).

The particular versatility of the World Wide Web, its hypertextual nature and the possibility of having access these services using any type of computer have prompted the development of specific software which allows data to be explored by means of the most appropriate kind of method and graphics. Therefore, every series is furnished with a brief introduction which explains its nature and source (meta-data), one or more procedures that implement the correct methods (with relative graphics) and several comments which describe the main results. In this way the user can decide whether the data are adequate for the use he intends them for and acquire practice in statistical analysis if necessary.

SOFTWARE ARCHITECTURE

First of all, and for reasons that will become clearer later, statistical software for implementing the procedures was chosen keeping in mind that it was probably available for the most common platforms, free of charge or was in widespread use (and therefore presumably already installed on the potential user's machine). Such considerations led to the choice of two products: XlispStat (Tierney, 1990) and Microsoft Excel. The first offers the advantage of being an object-oriented language implementing very powerful tools for dynamic graphics; the second is an easy-to-use product and presumably already well known to those who use the computer for teaching support.

In this regard, one should note that although using XlispStat requires knowledge of the Lisp language (knowledge which is uncommon among teachers and even less common among secondary school students), using the procedures set up for this purpose requires no specific knowledge.

Before illustrating the architecture designed for this project, several considerations of a general nature must be made which conditioned the choices made. One of the most critical factors of the Internet is its present intense traffic which is too great for the usual band width available. This considerably limits projects which provide for the intense use of graphics. Indeed, they require the transfer of large amounts of data from the server to the client, causing further link stoppage and reply levels which are often unacceptable.

Setting up a service with the previously described characteristics requires, among other things, a considerable use of graphics; but this seriously clashes with the present limits imposed by the network. To circumvent this obstacle several techniques are used that allow an advanced configuration of the World Wide Web services taking advantage of the network distributed computing resources. The most commonly used browsers (e.g. Netscape, Microsoft Explorer) offer interesting solutions which overcome some of the limits imposed by the html language and allow the processor of the client machine to be exploited for the local execution of procedures. In point of fact, it is possible to identify a file whose name has a predefined extension to a specific application by means of suitable specific MIME (Multi-purpose Internet Mail Extensions) definition. Suppose, for example, that a file contains commands for a certain language (or application); when a browser receives one of such files, it identifies its nature with the extension of its name and thus calls that particular application by executing all of the commands in the file. This technique can be used on most platforms and allows the organisation of services in the WWW framework which are based on architecture that provides the remote processing of programs and, therefore, the exploitation of every potentiality: the advantages are evident and easily intuitable.

Based on this assumption, such technology can be used to enrich archives with data and software procedures, allowing the user to acquire official data and procedures at the same time as well as a local framework in which he can gain experience. It should be stressed that this technique has proved to be particularly effective for using dynamic graphics which are especially useful in the field of teaching.

From the technical point of view, once the remote machine (client) has been suitably configured and the necessary software installed (the steps to follow are easy and illustrated in detail on the home pages), the server can limit itself to only sending data, procedures and possible comments; all of the computations are born by the local

computer and, if necessary, the user may fully interact with the analyses. For example, when the execution of an Excel procedure is requested by the home page, an Excel window opens in the local computer with the data and results; at this point the user will be able to intervene personally on the electronic page to analyse more aspects of the data, create graphics, and save the work space on a disk. On the other hand, with local execution of XlispStat the interactivity offered by dynamic graphics can be used and, therefore, interventions can be made on the graphics or on the control tools (such as sliders) for linking, labelling, rotating, etc. The management of the service described is entrusted to a series of appropriately written programs which support the WWW server in analysing and satisfying requests made by the user through the home page.

The general organisation of the system can be synthesized as follows. A group of programs in C language converse with the HTTP server through Common Gateway Interface (CGI), interfacing the archives of data, procedures and document files in such a way as to make the necessary documents available according to the user's request.

The functions of the management programs are as follows: i) automatic on the fly generation of HTML pages to show the files containing the descriptions of the data; ii) research and sending of Lisp or Excel macros to the client in relation to the type of data requested; iii) automatic generation of the HTML pages that describe the results. In this way the archive structure is completely free from the WWW service's management file and the archives can be updated independently of the rest of the system.

At present a revision of the software is underway which includes the reorganisation of the archive management procedures using Java as its language.

CONCLUSIONS

The work carried out constitutes an innovation in the sphere of services that can be produced for statistics through the World Wide Web. Indeed, besides having the normal functions of the WWW frameworks, this kind of archive can be considered as a distributed computing frameworks and enhance traditional services with many new potentialities. In this context, the use of systems oriented toward dynamic graphics opens a vast horizon in which initiatives in the field of teaching assume completely new dimensions and significance.