From **Brunelli, Lina & Cicchitelli, Giuseppe (editors). Proceedings of the First Scientific Meeting (of the IASE).** Università di Perugia (Italy), 1994. Pages 277-281. This pdf file is from the IASE website at http://www.stat.auckland.nz/~iase/publications/proc1993. Copies of the complete Proceedings are available for 10 Euros from the ISI (International Statistical Institute). See http://isi.cbs.nl/sale-iase.htm for details.

TEACHING DATA ANALYSIS WITH THE HELP OF SURVO

Juha Puranen
Department of Statistics, PL33 (Aleksanterinkatu 7)
00014 University of Helsinki, Finland

1. Survo

The Department of Statistics at the University of Helsinki uses various statistical software packages such as Survo, SAS, SPSS, SPLUS, SAZAM, PC-GIVE, GLIM, EGRET... Among these, the local statistical software package SURVO, created by professor Seppo Mustonen (Mustonen, 1992), is the most widely used.

Outstanding in Survo is the editorial user interface in which all the functions are carried out. Outcomes from various statistical analyses are also recorded in the edit field.

Using Survo's editorial capabilities, it is possible to generate reports, incorporate Survo-generated graphics, and send the overall report to the printer. For example the 500 page Survo User's Guide has been created in this way, saved onto disk, and taken to the printer as a Postscript file.

Comparing Survo to other statistical software, one can see the resemblance to S-PLUS, when this latter one is used within the GNU-Emacs environment. The number of different statistical techniques is slightly higher in S-PLUS. However, the user interface in Survo is much better.

Students can get 'SURVOS', a freeware version of Survo, for their personal use. Although the number of statistical procedures in 'SURVOS' is limited, it is far more complete than most existing freeware programs.

2. Survo in teaching

During the last 9 years I have used Survo's capabilities for almost all my lectures. By means of Survo I have:

handled all my text processing;
created transparencies;
planned and solved exercises;
built up all required teaching material.

3. Survo in data analysis

In this paper I will present the use of Survo for teaching data analysis where emphasis is focused on the handling of statistical data, the selection of the right statistical method and the interpretation of the final results.

The background of the students attending the course is usually merely the introductory course (40 hrs lectures + 20 hrs workshops) and the supplementary course (40 hrs lectures + 20 hrs workshops) in statistics. Almost 95% of the students have at least basic knowledge of computers but hardly any knowledge of real statistical data analysis.

The course on data-analysis is put together from three parts:

1 - Lectures 20 hrs (theory);
2 - Survo laboratory 15 hrs;
3 - Individual case study.

During the lectures, we work through various data-analytical problems, such as how to describe data, what statistical method to use, and how to interpret the results.

During the Survo laboratory, students learn how to analyse data in practice. All the teaching takes place in the computer laboratory, equipped with eight 486-PCs and the teacher's personal PC connected to a Data-show projector. For the laboratory the students are divided into small groups of 8 to 16 so that there is no more than 1 or 2 students sharing one computer.

Only a few students have their own data sets for a genuine case study. For the majority of students a common data set is selected. Last year's data set dealt with people's support of Finish political parties. For the case study, each student selected a personal random sample of 120 rural and urban districts. The main object of the case study is to explore the distributions of support for the parties they selected. This support, of course, depends on various confounding variables (demographic, socioeconomic, etc.).

Some students prefer to use SAS or SPSS to carry out their case study and therefore do not participate in the Survo laboratory.

4. How to use Survo in teaching data analysis

4.1 *Lectures*

A few years ago I used transparencies in all my teaching. "God knows how many times the right transparency was seemingly missing in the

disordered pile of transparencies". With the versatility of Survo, I was able to start a computerized and more disciplined teaching presentation.

A teaching presentation is usually about 5 to 10 minutes long. It is built up from texts, graphics and simulations. Because the length of each teaching presentation is short, it is very easy to pause, start it again, or reorganise it.

For the presentation of the teaching material a portable computer and color Data-show projector are used. In some special situations, extra transparencies or a blackboard might be used.

4.2 *Survo laboratory*

In the Survo laboratory, students learn how to analyse data in practical situations and how to use the statistical package Survo.

Usually the students have a very limited experience in statistical computing. Therefore, teaching should concentrate on both the use of the Survo package and data analysis methodology. Since the Survo laboratory duration is limited to only fifteen hours, teaching focuses on both simultaneously.

The Survo laboratory includes the following topics:

1 - Text processing and data input;
2 - Data file management;
3 - Data aggregation and graphical presentations;
4 - Hypothesis testing;
5 - Report printing.

A collection of short tasks (30 in all) is used for teaching purposes. The tasks are chosen so that they support as much as possible the final case study.

The Survo laboratory is intensive. There are more tasks than the students can perform under normal circumstances. In this way both advanced and less advanced students are kept under constant pressure! At the end of each session about 1 or 2 tasks might remain undone for some students. These are, however, to be completed before the next workshop.

Students learn the use of Survo as a by-product in the process of solving their practical assignments. In this way, emphasis is directed entirely to data analysis by means of Survo. Experience shows that students are more motivated to learn about the software Survo, when they are faced with real research situations.

In order to be able to control heterogeneous groups of students with different levels of skill, two teachers follow the students' progress. A principal teacher shows how to carry out analysis and how to interpret the

results. The second teacher is either an assistant professor or an advanced student who watches constantly the progress of all the students in their problem solving attempts.

Because students have a limited knowledge of Survo, they can get stuck in more or less simple matters. In such a case, they can get instant advice either from the main teacher or his assistant. This allows them to concentrate on analysing the data and, therefore, avoid conflict with the system they use.

4.3 *Case study*

The case study analysis starts immediately after the laboratory is over. Since students' ability to analyse data at this stage is rather uncertain, a help desk is organised where they can get thorough advice whenever necessary.

5. Why Survo?

Survo is an excellent tool for this type of teaching. Survo's unique user interface makes possible the control of text, data, graphics and statistical analysis.

Also the use of live simulations and the alteration or addition of a single or a multiple set of observations give a dynamic aspect and a new dimension to students'comprehension of the course. This undoubtedly motivates the students more towards learning than does the traditional ready-made examples method.

By means of Survo the students can perform their analyses step by step and learn more about statistical methods. Survo is not a 'black box' where one inputs data from one end and gets the outcomes from the other end. In fact, a knowledge-based procedure is required from any prospective student in order to get the best results using Survo.

By means of the editorial user interface, students can make personal notes in the edit field. They can save extra information made up of 'whats and whys' they might use later on. They also write into the edit field what methods were used and how the results were interpreted. The resulting document can be saved on disk and/or printed out and used later on as a model when analysing the case study.

6. Future

In the future, the improvement of the teaching material concerning the

theoretical part of the course will continue. In addition, I am planning to install the latest version of the teaching material in the local computer network. In this way, students will be able to access the teaching material through LAN and make this part of the course without a need to be present during the main lectures.

Bibliography

Mustonen S. (1992), SURVO *An Integrated Environment for Statistical Computing and Related Areas*, Survo Systems Ltd, Helsinki.