



HAL
open science

On a statistical Software package for non-expert users

Dimitris Karlis, Gilbert Saporta, Antonis Spinakis, Panagiota Anastasiou,
Luan Jaupi, Pierre-Louis Gonzalez

► **To cite this version:**

Dimitris Karlis, Gilbert Saporta, Antonis Spinakis, Panagiota Anastasiou, Luan Jaupi, et al.. On a statistical Software package for non-expert users. ETK & NTTS 2001, Eurostat, Jun 2001, Hersonissos, Greece. pp.903-905. hal-02468783

HAL Id: hal-02468783

<https://hal-cnam.archives-ouvertes.fr/hal-02468783>

Submitted on 26 Jan 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

On A Statistical Software Package for 'Non-Expert' Users

Dimitris KARLIS¹, Gilbert SAPORTA², Antonis SPINAKIS³ and Panagiota ANASTASIOU⁴

¹*Department of Statistics, AUEB, 76 Pattision st., 10434 Athens, GREECE
karlis@hermes.aueb.gr*

²*Conservatoire National des Arts et Metiers¹ 292 rue Saint Martin, 75141
Paris Cedex03, FRANCE
saporta@cnam.fr*

³*QUANTOS SARL, 191, rue Saint Honoré, 75001 Paris, FRANCE
ansp@quantos-stat.com*

⁴*QUANTOS SARL, 191, rue Saint Honoré, 75001 Paris, FRANCE
ganas@quantos-stat.com*

Abstract: In recent days, statistical methods are applied in a vast extent of different disciplines by a lot of people that, very often, have a rather little knowledge of statistics. This fact can lead to misuse of statistical methods, mainly due to the fact that the users are not able to evaluate the appropriateness and the applicability of the selected methodology, although these methods are rather simple. In addition, these users can be terrified by a statistical output with which they are not familiar. The statistical community has recognized this fact and now statistical packages try to offer some guidance to their users. This paper presents the design of a statistical package, oriented to users with medium or little knowledge of statistics. This package enhances some kind of 'expertise' in order to prevent the users from statistical maltreatment. The results of a survey about the user requirements of such software are presented. The general architecture and the functionality of such a system are provided.

Keywords: semi-expert system; attributes; user requirements; statistical practice

1. Introduction

In recent years, we have seen an explosion of data availability together with an increasing appeal of statistics in the everyday life, especially in socio-economic matters. Large databases can be created containing billions of data and even simple statistical procedures are needed in order to summarise the tremendous amounts of data. Unfortunately, this expansion does not imply that statistics is used in the correct manner. The need for some kind of guidance has been recognised by commercial packages that started offering it via "statistical wizards", like those offered by SPSS or Statgraphics among others. The design of "Statistical Expert Systems" has been discussed many years ago (see [1],[2],[3]). It has been recognised very early that it is quite difficult to enter the statistical knowledge in an

¹ Luan Jaupi and Pierre Louis Gonzalez from CNAM collaborated for the construction of this paper.

expert system. For this reason, the “expertise” has been focused on specific aspects of generic software, or it has been limited to particular application domains. In this paper we discuss the architecture and the design of an ‘expert-system’ for ‘non-expert users’. As ‘non-expert’ users we define all those that use statistical methods but they are not statisticians and they have a small knowledge about statistics, perhaps by some courses offered during their study, or they do not have any knowledge about statistics at all. This software, named “Extended Statistical Information System” (hereafter X-STATIS) aims at helping the ‘non-experts’ to use statistics in a correct manner and being able to understand the results of an analysis.

2. Capturing User Requirements

A user requirements study was carried out at the beginning of this project in order to understand the needs of target group (non-experts) in the market. This study consisted of three major parts. Primarily, a research on several statistical programs had been developed (market survey). This survey extends the results of other authors as it includes general statistical packages as well as packages offering a limited statistical module. In a second step, a qualitative survey (focus groups survey) was conducted, where the main aim was to reveal the underlying factors building opinions, behaviours or attributes of target group regarding the “ideal” statistical software. Finally, a quantitative survey was judged as necessary in order to identify with improved accuracy the requirements of ‘non-expert’ users. The survey was conducted in three European countries (Italy, France, and England) and the target population was employees in a variety of disciplines from medical to economic sectors that have a moderate or small statistical background. The main findings of the survey will be briefly given. The majority of them avoid the use of statistical packages. Microsoft EXCEL was quite popular. The statistical methods that they use were descriptive statistics, including group comparisons, time series and regression methods and rarely more advanced statistical procedures. They reported that guidance for the selection of the appropriate method and the explanation of the results would be quite helpful. They also provided information about IT functionalities that were considered as helpful. The above results provide a guidance of what the ‘non-expert’ wants from statistical software and they have been embodied in the functionalities of the system described in the next section. More details can be found in [4].

3. The Main Functionalities of X-STATIS System

The main goal of the system is to provide users with sufficient information to understand and evaluate the obtained results of statistical analysis. It also aims at creating a user-friendly information system since it will offer to the public facilities that fully will utilize modern IT-based systems and tools and will improve the quality of non-expert work. In the following, we describe some of the main functions of X-STATIS. The functionalities of the proposed package include: friendly interface, data base management, visual query builder, DDE Data manager, automatic descriptive statistical analysis, automatic multiple regression analysis, automatic analysis of time series, automatic elimination of outliers, dynamic graphical methods, manual use of methods and scripting modules among others. The general architecture of the system can be seen in figure 1.

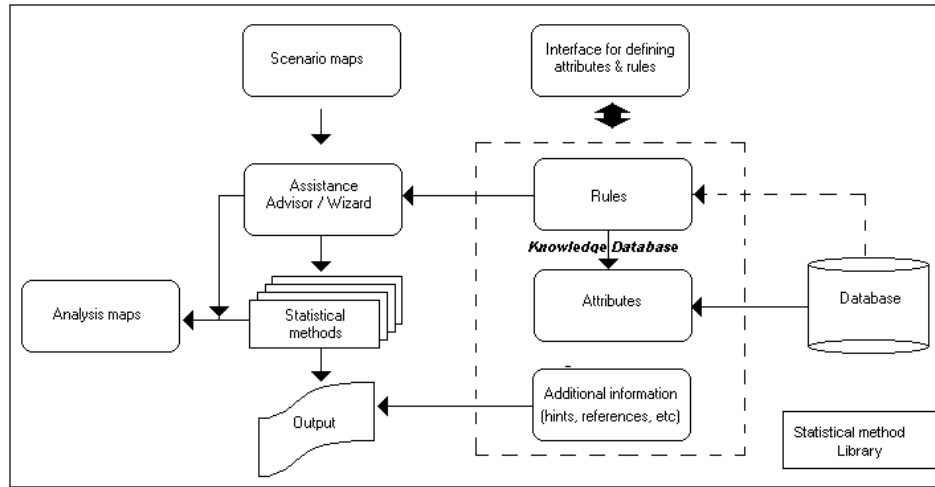


Figure 1. X-STATIS architecture

4. Conclusions

The X-STATIS system can be seen as an expert statistical system, which will address to non-experts data analysts. There is a skepticism whether such software can be of an “expert system” type since statistical knowledge cannot easily be enhanced into a computer program. For example consider the case of detection of heteroscedasticity in linear regression. Statisticians could detect it by a simple glance in the residual plot. This is not easy for the computer and hence we have to provide a measure for detection purposes. Which one is the best? No clear answer exists. Thus, we are limited in using some measure(s) that are known to the literature that ‘work’ but, clearly, controversial issues may arise. Such considerations are real challenges for statisticians. However, taking into consideration that we are in the beginning of this project, our basic aim is to enable non-expert data analysts to make use of modern IT-based systems and tools, and to improve the quality of their work.

Acknowledgements

This paper was written within the X-StatIS project, funded by the European Union through EPROS (project IST-99-1-1A-Proposal No 12134). In X-StatIS consortium participate ATKOSoft SA, QUANTOS SARL, Conservatoire National des Arts et Metiers, UK Office for National Statistics, BVA and National Statistical Service of Greece

References

- [1] Hand, D.J. (1984) Statistical Expert Systems. *Statistician*, 33, 351-369
- [2] Hand, D.J. (1987) The Applications of Artificial Intelligence in Statistics. Paper presented in Doses Seminar, 1-3 December, 1987 Luxembourg
- [3] Hand, D.J. (1986) Expert Systems in Statistics. *Knowledge Engineering Review*, 1,2-10
- [4] Project X-STATIS, D-2.1 Report: Definition and Assessment of User Requirements, including results of Market Survey & Evaluation of existing Statistical Software packages