

IAOS Satellite Meeting on Statistics for
the Information Society
August 30 and 31, 2001, Tokyo, Japan

A new approach to development of a statistical software corresponding to customers demands

TAKEUCHI, Akinobu
Rikkyo (St. Paul's) University, Japan

Abstract

In late years the spread of internet becomes remarkable. Many researchers think that developing of the statistical software that used internet is significant so that there are a lot of advantages. In this report, I propose a new approach to publish a statistical software with internet and introduce our project called DLLSA (Dynamic Link Library for Statistical Analysis). The purpose of this project is to DLLize various source programs written in several programming languages for statistical analysis and turn them into a library. This means that individual users will make the interface for input/output according to their needs and working environments while using this DLL as a statistical engine. There are many application software for the Windows system that can call the DLL. It is also easy to DLLize program source written in common programming languages such as Basic, C/C++, Fortran, and Pascal. All contents of our library can be downloaded freely from the DLLSA web site (<http://www.sci.kagoshima-u.ac.jp/~dllsa/>).

1 Introduction

The general public today has a fairly wide knowledge of the internet. Many people can get the massive information from the internet by using the PC, mobile terminal, and a portable telephone. To use the internet has many merits for the statistical software user and the developer. For example, i) using same information everywhere, ii) updating comparatively quickly, iii) making a software expected by analyzing access, iv) getting only the information that a user want by searching, and so on. On the other hand, marketed software includes a lot of analysis methods and options in order to answer a lot of demands of a user. Therefore, a lot of development time and expense are need. Additionally, for the novice users complicated input and output options would only confuse them, which would result in their missing the objectives of lessons. It may also cause errors in the operation. From these things, a lot of developers suggested the statistics software that used internet.

Statistical software is used for a variety of purposes. Naturally, demands for software vary according to the purpose of its users. For educational use, it is not necessary to have many options and elaborate output. On the other hand, software for practical use needs to meet many kinds of user demands, so multiple options and applications for deeper analysis and for presentation are desirable in such software. Currently marketed software, however, does not satisfy all the varying user demands due to the difficulty in creating such software. A general trend in the computer industry has been towards making software more user-friendly. Recent software systems usually have a graphical user interface (GUI), but making a common GUI is very difficult.

Liu et al. (1995) and Yamaguchi et al. (1998) independently proposed solutions to this problem. Their solutions were to create an interface to suit users' purposes using marketed software as a statistical engine. Nakano (1998) tried to develop software with a user interface that was as independent of the computational engine as possible, using a World Wide Web (WWW) interface for this purpose (also see the discussions in Thioulouse and Chevent, 1996; Sato, 1997).

In this report, I propose a new library with the internet, called the DLLSA (Dynamic Link Library (DLL) for statistical analysis) that is available as a statistical engine, as mentioned above, or which can be called from existing general software, such as SAS, SPSS, and S. An example of the interface for cluster

analysis is presented to explain features of the DLLSA.

2 DLLSA

2.1 About DLLs

A DLL can be considered part of a library for applications within the Windows operating system as well as for programming. A DLL contains one or more functions that are compiled, linked, and stored separately from the applications using them. One advantage of a DLL is that multiple applications can access the same DLL. The DLLs in our library can be accessed from any software or spreadsheet-type software if such software has the facility to call functions or subroutines in DLLs. Since the source code for DLLs can be written in several programming languages, such as Basic, C/C++, Fortran, and Pascal, we think it is significant to construct a library using DLLs.

Users can DLLize source programs with commercial software suitable for each programming language. For example, when source programs are written in Fortran, users could make a DLL using Visual Fortran. Table 1 lists commercial software that can create DLLs. For further details, users can refer to manuals for their software according to need.

Table 1 Software to make a DLL

Programing Language	Software
Basic	Visual Basic
C/C++	Visual C++
Fortran	Visual Fortran
Pascal	Delphi

The procedures vary a little, depending on the software used to create the DLL and the software used to call it. The following are the procedures used to call a DLL created in Visual Fortran into Excel. First, the descriptions necessary to call the DLL must be typed into the header of the Excel macro file. As shown in Fig 1, these descriptions include the name of the procedure, the location of the DLL file, and the name of and type of variables that transport the value between the DLL and Excel. Then the `call` command is used at the location for the variables, in order to call the procedure defined in the DLL file (Fig 2).

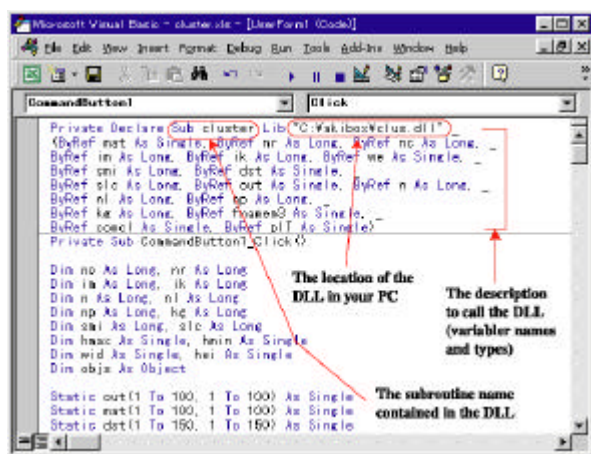


Fig. 1 Header of Excel Macro

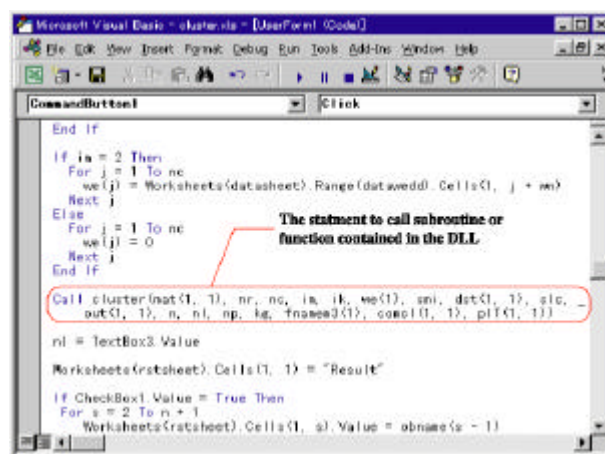


Fig. 2 Calling the DLL in Excel

2.2 DLLSA

It is next to impossible for an independent creator or for a small group of independent creators to develop and publish software that contains a wide variety of statistical methods. From the users' point of view, it is not easy to use a variety of software with different interfaces and varying procedures. The DLLSA project

could offer a solution to this problem. DLLizing existing program sources cuts down the time needed to develop such software, and makes it possible to create a library with procedures that can be used for any software. Currently, the Fortran sources of the NISAN System (Asano and Kimura, 1990) are most often used for the DLLSA.

The DLLs in this library are each made for a different numerical calculation and statistical method. Users can call the DLL from their choice of interface software, such as Excel. In other words, the library can supply what the interface software cannot supply. Users, for instance, could feed data into Excel and then let the DLL perform multivariate analysis, which Excel cannot do. They could then use Excel to view the results of their analysis. Separating the interface from the statistical engine allows users to perform the entire process in a familiar environment, from data input to analysis to output of results. At the same time, the library frees software authors from interface problems, allowing them to concentrate on developing what is most needed.

2.3 Publishing DLLSA

The contents of our library can be downloaded and used freely. Details for the library and published DLLs can be obtained from our web site (Fig 3). The URL is <http://www.sci.kagoshima-u.ac.jp/~dllsa/>. On this web site, we also provide examples of interfaces for several applications, including, Microsoft Excel, S-PLUS, STATISTICA, etc. Examples of specifications for some DLLs are also available.

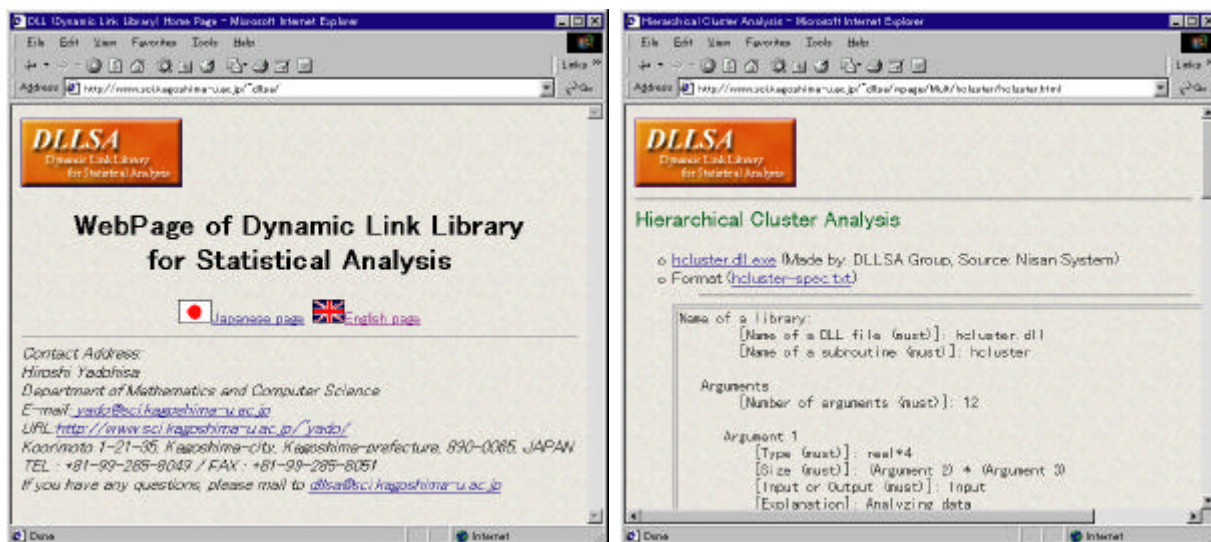


Fig. 3 Top page of DLLSA and the distribution of the DLL (hcluster.dll) page

2.4 DLLSA usage according to purpose

The requirements for statistical software vary according to the goals of its users. In the field of education, in which the level of computer skill varies widely amongst users, it would be more effective if the interface could be changed according to the users' level of computer proficiency. For novice users, complicated input and output options are confusing and often result in users missing the objective of a lesson. They may also cause errors in operation. Dispensing with unnecessary options and simplifying the interface, therefore, would be of greater benefit to this type of user. Using graphics, such as charts, figures, and diagrams, to represent the results of analysis is often more effective than numerical description.

There are also diversified demands for using software for practical purposes. Those users who only do routine work only need the interface necessary for that work, so unused options should be discarded. On the other hand, users who actively use statistical methods for analysis of every kind would need an interface that is applicable to a variety of cases. They demand numerical accuracy for the output of results, in a format that often must be adaptable to further complicated analysis.

The DLLSA proposes separation of the interface and the statistical engine to meet such demands. It provides a statistical engine that satisfies the general user, while the interface can be custom-made according to the users' purposes. As an illustration, the following section examines a sample of the interface for general use of the DLLSA for cluster analysis, which is a very versatile analysis.

3 Concluding Remarks

The features of the DLLSA enable users to change the interface according to their purpose, and can answer diversified demands while using the same DLL. Changing the interface, however, requires a certain level of computer proficiency on the user's part and those who are inexperienced with programming would find it difficult to understand or perform this task. To help such users, an Excel macro that is an example of this interface is published along with the DLL.

In addition, internet includes a lot of merits. I think that I should use these merits of internet more. For example, I will build the library which a user wants by analyzing access.

One goal of this project is to expand this library and to promote joint research. There is room for improvement, such as incorporating the latest statistical methods. Therefore, it is essential to join hands in this project, to expand and improve the library.

References :

- [1] Asano, Ch. and Kimura, H. (1990): A study on system construction of general purpose statistical package for microcomputers --- Micro-NISAN system, *Computational Statistics Quarterly*, **6**, 65 -- 77.
- [2] Liu, L.-M., et al. (1995): A system-independent graphical user interface for statistical software, *Computational Statistics and Data Analysis*, **19**, 23 -- 44.
- [3] Nakano, J. (1998): Graphical user interface for statistical software using internet, in: Payne, R. and Green, P. (Eds.), *Proceedings of Computational Statistics 98*, 407 -- 412: Heidelberg, Physica-Verlag.
- [4] Sato, S. (1997): Introduction to "Web Decomp" ---Seasonal Adjustment System on WWW---, *Proceedings of the Institute of Statistical Mathematics*, **45**, 233 -- 243 (in Japanese).
- [5] Takeuchi, A., Yadohisa, H., Yamaguchi, K., Watanabe, M., and Asano, Ch. (2001): DLLSA: Dynamic link library for statistical analysis, To appear in *Bulletin of The Computational Statistics Society of Japan*, **13** (2), (in Japanese).
- [6] Thioulouse, J. and Chevent, F. (1996): NetMul, a world-wide web user interface for multivariate analysis software, *Computational Statistics and Data Analysis*, **21**, 369 -- 372.
- [7] Thioulouse, J. et al. (1995): ADE software: multivariate analysis and graphical display of environmental data, in: Guariso, G. and Rizzoli, A. (Eds.), *Software per l'ambiente (Pàtron editore Bologna)*, 57 -- 62.
- [8] Yamaguchi, K. et al. (1998): Educational statistical software on the world wide web, *Proceedings of Computational Statistics 98: Short Communications and Posters*, 127 -- 128.