MODERN TENDENCIES IN THE EVOLUTION OF STATISTICAL SOFTWARE

Veselin Hadjiev

In this article are described the principal tendencies in the evolution of statistical software. This is achieved through the following indications: developer, hardware and software dependency, interface, the degree of software usage in the statistical study, specialization, etc. For the convenience of the reader the cited information has been obtained mostly from Internet.

The development and supply of statistical software in the last few years increased so much that it became pointless to count the new products. As major factors for this fast growth can be indicated the progress of information technologies and Internet, the diversity and the expansion of the software market and most importantly the much bigger demand and consumption of statistical products.

The progress and availability of information technologies created a milieu enabling the theoretical methods and techniques from the near past to become real. An additional thrust of development and supply was achieved by means of the progress and availability of Internet. The fast and simple exchange of innovations and modules became possible thanks to the global network. The evolution of electronic and other markets removed the existing boundaries and made the supply of software products available for everyone. Finally the conditions were favorable for the development of the supply and demand of statistical software.

The supply of a large variety of different in structure and functions products gave the consumers the possibility to choose and use the most appropriate for the goals of their study software products. On the other side the variety of statistical products confused the consumers in their choice. The confusion is due to the particular nature of the statistical software, their complexity and the impossibility for them to be assessed before they are bought, the necessity of substantial knowledge in math and statistics, etc. There is obvious need for the principal tendencies in the development of statistical software to be outlined and for the variety of statistical products to be grouped according to their purposes. For the convenience of the reader the cited information has been obtained mostly from Internet.

Catalogues and classifications of statistical software products appear in a number of Internet sites like : *STI Statistics, The Econometrics Journal Online, Q-Stats Software, etc.* Most often the

offered systematizations are made in alphabetical order after the names of the products, the names of the developers and producers, and the included methods of analysis, so eventually they look like catalogues.

Systematizations based on essential indications are missing although they are necessary to describe the modern tendencies in the evolution of statistical software. In this article the progress of statistical software products will be reviewed through the following indications: *developer, hardware and software dependency, interface, the degree of software usage in the statistical study, specialization, etc.*

The first obvious tendency is the permanent expansion of the circle of developers of statistical software. This ascertainment can be explained with the globalization and availability of informational resources and instruments as well as with the increasing needs of statistical software. Whereas the different developers go for different goals while developing a specific product, their products will have different features. On the whole the developers can be divided in the following groups: universities, science institutes, commercial developers, etc.

The completed study shows that universities develop a large number of software statistical products. This ascertainment can be explained with the following premises and conditions. Firstly the universities possess the required science potential to develop either universal or specialized software products. Secondly the universities themselves are big users of statistical software products for the conduct of studies and for the learning process. It's typical for the university products not to be anonymous but to have leading authors – mathematicians, statisticians and specialists in other areas. Some of the most popular university statistical software products are: *Ox* and *PcGive*, Oxford university (**http://hicks.nuff.ox.ac.uk/Users/Doornik/**); *BUGS,* Cambridge university (**http://www.mrc-bsu.cam.ac.uk/bugs**); *COUNT, EI, EzI, Maxlik,* Harvard university (**http://gking.harvard.edu**); *SHAZAM,* British Columbia university (**http://shazam.econ.ubc.ca**); *STAT,* San Diego university, California (**http://www.cis.ohio-state.edu/~perlman/statinfo.html**); etc.

Other major developers of statistical software products are the research institutes. This ascertainment can be explained with the considerable scientific potential of these institutes, as well as with the necessity of developing specific software products, incorporating specific information and specific statistical methods of analysis. Some of the most popular research institutes statistical software products are: *CrossGraphs,* Belmont Research Inc. (**http://www.belmont.com**); *DOE KISS, SPC KISS,* Air Academy Associates (**http://www.airacad.com**); *CPS Utilities,* Unicon Research Corporation (**http://www.unicon.com**); *NOSTRADAMAS,* New Energy Associates

(http://www.newenergieassoc.com/html/nost.html); **_Prophet,_** National Center for Research Resources (http://www.prophet.abtech.com/); **_SCA,_** Scientific Computing Associates Corp. (http://www.scausa.com/); etc.

The third major group of creators of statistical software is that of commercial developers. It's right to be said that almost every developer offer his product on the market. The universities, research institutes, etc. also use their products for other purposes, whereas the commercial developers create their products to sell them. Characteristic of the commercial statistical products is the fact that their creators are trying to include a wider range of statistical methods, regardless of the subject orientation. This is due to the intention of the developers to appeal to a greater range of clients with various needs and interests. Some of the most popular statistical software products are: **_EXPO,_** Leading Market Technologies (http://www.lmt-expo.com/); **_Forecast Pro,_** Business Forecast Systems (http://www.forecastpro.com/); **_LIMDEP,_** Econometric Software (http://www.limdep.com/); **_SAS, JMP, StatView,_** SAS Institute (http://www.sas.com); **_SPSS, SYSTAT, BMDP, SigmaStat,_** SPSS Inc. (http://www.spss.com/); **_Statistica,_** StatSoft (http://www.statsoft.com) and many else.

The fourth group of developers is the group of individual authors. As a rule the creators of software products are leading teachers and scientists in universities and institutes. Of course the separation of this group is relatively conditional because the authors use the resources of the institutions they work in. Characteristic of these products is their coherence with the subjects and methods the authors have specialized in. Some of the most popular products are: **_Microfit,_** Prof. M.H. Pesaran and Dr. Bahram Pesaran (http://www.intecc.co.uk/camfit); **_MLn,_** Prof. Harvey Goldstein (http://www.ioe.ac.uk/multilevel); **_TDA,_** Prof. Dr. Goetz Rohwer and Dr. Ulrich Poetter (http://www.stat.ruhr-uni-bochum.de/tda.html); **_Applied Stat,_** David Cooper and Associates (http://members.aol.com/djpcooper/spc.htm); etc.

An essential tendency in the evolution of statistical software is the growing independence of operational systems and hardware. Almost every developer offers his software compatible with different classes and types of computers, as well as with different operational systems. By doing so the developer widens the range of his potential clients. It is a common practice for one product to be offered in both PC and Macintosh versions, for example the products of ACL Software (http://www.acl.com/), Abacus Concepts Inc. (http://www.abacus.com/), etc. Even more are the products, which come in versions compatible with different operational systems – Windows,

DOS, UNIX, etc. For example the products ***BACC 1999*** (**http://www.econ.umn.edu/~bacc**), ***TSP*** (**www.tspintl.com**), etc.

In addition to these tendencies is the unification of users' interface under different operational systems. The next step for the removal of any differences between operational systems and hardware is the application of the protocols of the global network and the usage of statistical software directly in Internet, like ***SHAZAM.*** Therefore from the user's point of view the factors hardware and operational system lose their influence on the choice of statistical software.

In the evolution of statistical software comes up a tendency the "graphic" and the "shell-oriented" interfaces to become similar. It's done by integration of the advantages of the two types when setting up the communication with the user. On the one hand by means of the possibilities of the menus an intuitive and simple interface is built, common for most Windows applications. Usually this is the way to put in effect the usage of classical statistical methods – descriptive statistics, disperse analysis, regressive analysis, etc. These products are upgraded with modules of applications created on the corresponding computer language. They provide the software product with flexibility, new statistical methods, interactivity with the database, options to use on-line libraries of downloadable upgrades, etc. The products of Stat Soft are an example of the integration of the two types of interface. These products are based on both traditional "graphic" interaction and the computer languages "BASIC" and "STATISTICA COMMAND LANGUAGE". For example the developer offers upgrades of the base configuration, which are able to activate executable files under the loaded statistical software; results of Box-Cox transformations; etc.

Undeniably the shell-oriented interface requires top level of special preparation. It serves to organize the operation of data and calculation of characteristics at lowest level. "Graphic" interface requires more basic level of preparation. It serves to carry out the different stages, methods and stages of methods, but the user does the setup of the whole study. Analysis of the different interfaces used by statistical products reveals the emergence of a new type of interaction with the user. These are products with graphic interface, which are easier to use, and dont require special skills in statistics from the user. ***ViSta***, (**http://forrest.psych.unc.edu**) is one of these products.

A major tendency in the evolution of statistical products is the attempts of some developers to provide a greater degree of backing of the statistical study. They achieve it by developing special products for different stages or universal ones for the whole study. Depending on, what part of the statistical study is provided by the corresponding statistical software products, the following groups are formed:

- statistical software products for collecting and preparing statistical data for processing;
- statistical software products for processing statistical data by the appropriate statistical methods;
- statistical software products capable of completing the whole study;
- software products for converting and exchanging information between data bases, electronic tables and statistical software products.

The software for creating and maintaining statistical databases falls in the first group. Besides the universal *dBASE* and *Paradox*, new specialized software is under development. It can create and maintain data that have the features of statistical information. This software is used for the firsthand collection of data from the object of study. For example surveillance of accomplished sales by bar codes, surveillance of indents, sales and other objects completed in Internet, surveillance of electronic transactions (at stock exchanges for example), features directly measurable by some instruments (hydrological, meteorological and other data), directly measurable data from the industry (like systems for quality control) etc. Products from this group are *SurfReport* (**http://www.bienlogic.com/SurfReport/**), *WatchWise* (**www.watchwise.com**), *Pangsway Limited* (**www.pangsway.demon.co.uk/pangsway.html**) etc.

The second, largest group consists of the products used for processing statistical data by the appropriate methods. The prototypes of statistical software products are in this group too. These products use data from secondary surveillance. The accent is put on the methods of statistical analysis and prognosis. Some of these products are developed by SPSS, StatSoft, Stata Corporation (**www.stata.com**).

In the third group of statistical products the accent is put on the whole study. Through it specifically is realized the whole process of collecting firsthand data, processing, interpretation and recommendation. These systems function like advisory on-line systems and have artificial intelligence elements. These products miss the flexibility, common for the previous group. Some of these products are developed by Raosoft (**www.survey-software.com**), STEN Information Resources (**www.steninfores.com**).

Statistical data are collected and organized by means of various software products operating under different formats. This impedes the exchange of data, raises the cost of informational products, etc. Special converting and translating software products are being developed to solve this problem and facilitate the exchange of data. They support a large spectrum of formats and masks of databases, electronic tables and

statistical software products. This fourth group contains the products *Circle Systems* (**www.stattransfer.com**), *Conceptual Software* (**www.conceptual.com**).

The purpose of using any statistical software is to create direct or indirect conditions for prognosticating the future state and development of the studied object. From this point of view three tendencies in the evolution of statistical software can be conditionally described: descriptive systems, analytical systems and automatic systems for prognostication.

The descriptive systems process the information at lower level, including mostly graphic methods, descriptive statistics, etc. These products, for example *"Analysis ToolPak"* for MS Excel, are good mostly for describing the current condition of the object by means of appropriate statistical features. The analytical systems are used for a careful statistical study, evaluation of different statistical and dynamical features, on the basis of which the user loads variables and extrapolates prognosis for the future condition of the object. Part of this group are the products of Quantitative Micro Software *Eviews* (**http://www.eviews.com**). The automatic systems for prognostication result in evaluating of the future condition of the object. All medial stages of the statistical study remain hidden from the user. The automatic systems for prognostication aren't as flexible as the analytical ones. They can be applied only in cases when the conditions for developing of the object are constant and the variation of prognosticated indications is relatively homogeneous. An example of this system is the product *Autobox,* Automatic Forecasting Systems (**http://www.autobox.com**).

The development of statistical software products shows controversial tendencies about the subject area. On the one hand the development of more specialized products continues. The following products are offered in the field of hydrological study: *Amos,* SmallWaters Corporation (**http://www.smallwaters.com/amos**); *Aqua Chem,* Waterloo Hydrogeologic Inc (**http://www.flowpath.com**); in the field of medical study: *EpiCalc 2000,* Brixton Health (**http://www.myatt.demon.co.uk**); *Epi Info,* Centers for Disease Control and Prevention (**http://www.cdc.gov/epo/epi/epiinfo.htm**); in the field of sport statistics: *Team Sport Software* (**http://www.teamsportsw.com/**); *Sport Shareware* (**www.absnet.com**); etc.

The classification of a given product as specialized for specific field is made on the basis of its methods of statistical analysis. This is not quite relevant because the statistical methods are invariant to the subject field. If a specific statistical information has all the necessary features to be processed with a specific statistical method, it can be done with either

Statistica, SPSS, Eviews or ***BUGS, NOSTRADAMUS,*** etc. From this point of view classifying by subject field according only to statistical methods of analysis isn't enough. Of course, it can't be denied that economical statistical data have specific features, biological statistical data have different features, psychological statistical data have third features, etc. The preparation of data and extracting various features from them are the main criteria for classifying the products in different subject fields. From this point of view strictly specialized products, for example in the field of statistical control of quality are the products of Hertzler Systems Inc (**www.hertzler.com**); IMPACT (**http://www.iaehv.nl/users/mschaeff**); etc.

On the other hand powerful universal products are being developed like the above ***Statistica*** and ***STATGRAPHICS***. The products ought to be specialized because if the methods and techniques included are too many, the products become difficult to use and unacceptably expensive. Usually the developers deal with this problem in two ways. One of them is to create basic software for the most common operations and a number of specialized add-ons. Stat Soft for example creates such products. The other way is to create basic conventions to process data, modules for the calculations and a family of statistical software products with different orientation. Those are the products of SPSS Inc., which consist of the basic product ***SPSS*** and the additional ***SYSTAT, SigmaStat, SigmaPlot, AutoSignal, etc.***

The intensive and multidirectional progress of statistical software makes possible the conduct of more complex research. On the other hand the big diversity of products creates difficulties for the potential buyers to make their choice and brings forward the need of suitable systematization. The above described trends in the evolution of statistical software are starting-points in the creation of classifications of essential features, important for the end-users, and the conduct of comparative research for monotype products. This way they will present the prerequisites for the right choice and use of statistical software products.

BIBLIOGRAPHY: Goev W. What should we know of the statistical software applications? "Statistika", Sofia, 1997, #1, pages 40-45; **http://www.eur.nl**, **http://www.stats.gla.ac.uk/**, **http://www.stat.ufl.edu/**, **http://www.ai-geostats.org/**, **http://www.cpcug.org**, **http://tolstoy.nsw.cmis.csiro.au/**