Yo. Ilkhamova
Institute for Advanced Studies and Statistical Research
Candidate of Economic Sciences, Associate Professor, Applicant,
State Statistics Committee of the Republic of Uzbekistan

ANALYSIS OF STATISTICAL PACKAGES USED IN THE STATISTICAL RESEARCHS

Abstract: The information flow that has been increasing in the last decade inevitably leads to the accumulation of a huge amount of information, and now the scientific community is faced with the task of analyzing the so-called "big data". It is impossible to analyze such arrays of information without software. Statistical software packages (SPP) are called upon to reveal explicit and hidden patterns. The need for statistical data analysis tools in statistical agencies for statistical data processing is great, which led to the development of the computer software market for statistical data processing. The article discusses the statistical packages used in statistical research.

Key words: statistical package, regression analysis, statistical observation, summary, grouping of statistical observation materials, sample.

Language: English

Citation: Ilkhamova, Yo. (2021). Analysis of statistical packages used in the statistical researchs. *ISJ Theoretical & Applied Science, 06 (98),* 139-143.

Soi: http://s-o-i.org/1.1/TAS-06-98-20   *Doi*: crossref https://dx.doi.org/10.15863/TAS.2021.06.98.20

Scopus ASCC: 2613.

Introduction

Statistics as a body of information deals primarily with the quantitative side of the phenomena and processes of social life. The data of statistical reference books in the language of numbers characterize the sizes and quantitative relationships (volumes, structure, rates of development, etc.) of the phenomena of social life and the regularities manifested in them. A common feature of the information that makes up statistics is that they always refer not to one single (individual) phenomenon, but cover a whole series of such phenomena or, as they say, their totality, with summary characteristics. It should be noted that statistics are based on accounting, but when accounting for statistics, the goal is the final, summarizing characteristic of the entire population as a whole and its individual groups, and each record is a means to achieve this goal. Since statistics deals with quantitative characteristics, it widely uses the positions and methods of mathematics in its research. The theory of probability and mathematical statistics, which are engaged in the study of abstract sets of units and general quantitative laws operating in them, are especially widely used in statistics. The laws, rules

and methods established by these branches of mathematics are used by statistics in solving their specific problems. In particular, the law of large numbers plays an important role in statistics. The automated information system for the collection of state statistical reports in electronic form eStat (hereinafter referred to as the system) is designed to carry out current work on the preparation, filling out and submission of state statistical reports by legal entities to state statistics bodies in electronic form.

The creation of the System pursues a well-grounded solution of the following tasks:

- Introduction of information exchange methods using telecommunication means and related training of specialists involved in exchanging the parties with the proposed technologies for the provision and collection of state statistical reports in electronic form;

- Organized provision of statistical reports by legal entities in this form in electronic form;

- Application and use of modern solutions and technologies for identification and authorization of users using electronic digital signature (EDS) technology. In many areas of natural science,

statistical methods have been and remain an important part of the measurement results processing.

This applies to almost all branches of knowledge: physics, chemistry, biology, geology, meteorology and many others. Modern programs for statistical data processing allow the use of sophisticated modern methods of analysis even in those areas where previously such studies were extremely laborious and, therefore, were carried out quite rarely.

Personal computers open statistical methods to a wide range of users, 90% of which are non-professional statisticians. These methods have become available not theoretically, but in highly integrated environments, where the entire stage of analysis from data entry, importing them from other databases to building auto reports and saving them in HTML format is fully automated.

There are about a thousand packages distributed on the world market that solve statistical data analysis problems in one form or another, in DOS, OS / 2 or Windows (various versions). Known computer systems SAS, SPSS, SYSTAT, Minitab, Statgraphics (or, in other words, STSC). A little aside from them, you can put the Statistica package, which is popular in our country. Of the tools actively used in the world in the field of data analysis, mainly by programmers (developers and researchers), the most famous are the IMSL library of numerical and statistical methods (it has a competitor in numerical methods - NAG), as well as an interactive environment with a built-in OOP language programming. The Russian packages STADIA, OLYMPUS and the Belarusian package ROSTAN can be considered "semi-specialized" (according to the classification) and "semi-universal". Most likely, the American packages ODA, WinSTAT, Statit, UNISTAT, Multivariance 7, JMP, SOLO, STATlab should be referred to the same class [1].

In addition, there are also clearly specialized packages for classification and dimensionality reduction, both domestic: CLASS-MASTER, KVAZAR, PALMODA, Stat-Media, STARC - and foreign ones, for example, MVSP. Quite widely known are packages that solve problems related to classification. They are the American systems BMDP / W, SigmaStat, Statistix, TURBO Spring-Stat-Win, as well as the domestic package "Statistic-Consultant for Windows". In addition, there are statistical expert systems on the market such as STATEKS, Statistical Navigator Pro. Among the non-statistical packages that solve classification problems, one can note the PolyAnalyst packages, DA-system, ARGONAVT, LOREG, OTEKS package and various neural network packages.

General purpose (integral) statistical packages: import or export data in ASCII, dBASE, Lotus 1-2-3 format. In addition, they are able to combine data for analysis from various files, select subsets of data, rank or sort them according to certain conditions. Finally,

it is always possible to edit the data in the package environment and add descriptive text to the parsed datasets (files) [2].

To date, approaches to the statistical assessment of the development of the digital economy have already been formed. In the modern period in the statistical industry of the Republic of Uzbekistan, a large number of statistical calculations and forecasts are carried out on the basis of the use of advanced ICT, there is experience in organizing interaction between information systems (IS) operating in the State Committee of the Republic of Uzbekistan on Statistics. The State Committee of the Republic of Uzbekistan on Statistics has developed and introduced into operation new information systems. Thus, the information system "Consumer Price Index Calculator" was developed to calculate the overall change in the consumer price index (CPI) for a certain period of time and is designed to operate online. The next system, which was developed and implemented in the State Committee of the Republic of Uzbekistan on Statistics in a pilot version, is an information system based on the use of observation tracking technology using tablets - CAPI (Computer Assisted Personal Interviewing) technology.

Statistical package is a software product designed for statistical data processing; usually this package includes business graphics, analysis of variance, regression analysis, time series analysis, etc.

A feature of statistical analysis methods is their complexity due to the variety of forms of statistical patterns, as well as complexity process of statistical research. Statistical research can be carried out through the following methods, such as statistical observation, summary and grouping of statistical observation materials, sampling, etc. Statistical observation is an organized and systematic collection of information aimed mainly at the phenomena of social life and it must be carried out taking into account such requirements as the received data must be accurate and reliable [3].

Typically, statistical observation can take the form of reporting, where information is sent to specific statistical units of the statistical office. In this case, the data is entered into special reports. As you know, taking into account the indicators of the accuracy of the study, a simple and complex summary is distinguished, which is based on a specific grouping feature and the order of formation of groups. In addition, the considered system of indicators allows you to characterize the group and the object as a whole for the developed tables, to present the results of the summary. The variety of statistical packages is due to the diversity of data processing tasks using various types of statistical analysis procedures.

Based on functionality, statistical packages can be divided into 3 main groups [4]:

1. Universal or general-purpose packages (for example, SPSS, STATA, Statistical, S-plus, Stadia,

Statgraphics, Systat, Minitab). These packages are not focused on a specific subject area and can be used to analyze data from various areas of activity. They offer a wide range of statistical methods and have a relatively simple interface. These packages are recommended to work with at the initial stages of working with data, when the statistical methods that will be used to solve a particular issue are not yet clearly defined. The versatility of the universal package allows trial analysis of various types of data using a wide range of statistical methods. Most of the existing generic packages have a lot of overlap in the composition of built-in statistical procedures.

2. Professional packages (for example, SAS, BMDP) differ from universal ones in that they allow you to work with extremely large amounts of data, apply highly specialized analysis methods, and create your own data processing system. In statistical offices, the analyzed data are specific, that special methods of statistical analysis should be applied to them, which are not presented in universal packages.

3. Specialized packages (for example, BioStat, Mesosaur, Datascope) allow analysis using a limited number of specialized statistical methods. Such statistical packages are used by specialists who are well acquainted with the methods of data analysis in the area to which the package is oriented. Thus, the statistical package BioStat was created for the analysis of data in the field of biology and medicine, the statistical package Mesosaur specializes in the analysis of one-dimensional and multivariate time series and the construction of regression models, the statistical package Datascope specializes in the analysis of multivariate data. Specialized packages allow analysis using a limited number of specialized statistical methods. Most of the statistical packages on the market have a flexible modular structure that can be replenished and expanded by custom modules that are additionally purchased or freely available on the Internet. This flexibility allows most packages to be tailored to the needs of a particular user.

Such packages as Statgraphics, SPSS, STATA, SAS, Statistical, EViews, Kxen, S-plus, Deductor, PrognozPlatform, Stadia, Evrista, Mizosaur, Olympus: Stat-Expert, Statistic- Many programs represent not only a tool for statistical data processing, but also a data warehouse system, a reporting system for management, etc. The cost of implementing such packages can reach hundreds of thousands of dollars (SAS, SPSS Modeler), but it pays off in terms of subsequent returns from implementation. By functionality, i.e. in terms of analytical capabilities, a set of tools, they distinguish universal general-purpose or professional packages, these include SPSS, Stata, Statistical, S-Plus, SAS, Deductor, Prognoz Platform, etc., as well as specialized packages: BioStat, EQS, EVRISTA, GWR4, GeoDA, ArrowModel, etc.

SANI, CLASS- .d. Our analysis showed that many packages like R, Python, RapidMiner, BV4.1, GeoDA, Winpepi, EpiInfo, X-12-ARIMA, Deductor, Prognoz Platform, SPSS, STATA, SAS, Statistical, EViews, Maple, Mathematica, Matlab and others are characterized by a relatively small set of statistical methods and are designed to solve specific problems in specific subject areas, for example, when conducting censuses or assessing the epidemiological situation.

The analysis showed that statistical methods are mostly invariants with respect to the initial data, and universal packages can be used for data analysis using a wide range of statistical methods and have a simple interface. The versatility of the universal package allows for a detailed analysis of various types of data using a wide range of statistical methods. Most of the existing universal packages have many overlaps in the composition of built-in statistical procedures and are competing products in relation to each other. Their main difference lies in the way the program interface is implemented.

Specialized packages allow analysis using a limited number of specialized statistical methods. Such statistical packages are used by specialists who are familiar with the methods of data analysis in the area to which the package is oriented. Most of the methodological literature on the SPSS and Statistical packages, but the popularity of the SPSS package is due to the fact that this universal package is actually the standard for processing sociological and marketing data at many commercial enterprises. And the popularity of the Statistical package of the Stat Soft Company has a convenient russified interface and has a lot of reference support.

The advantage of this package is the presence of separate modules with extensive functionality for building neural networks and the package allows you to process data on econometric studies.

In order for the statistical software package to be convenient and effective in its work, it must meet numerous and very stringent requirements:
- contain a significant range of statistical tools (procedures, methods);
- be simple enough for quick learning and easy to use.
-meet high requirements for input, transformation and storage of datasets, also have the ability to export and import data;
- have a wide range of tools for graphical presentation of data and results;
- have detailed documentation (methodological support), sufficient for independent mastering by researchers who are familiar with statistical procedures;
- be affordable, ideally a free product. Among this variety of serious statistical packages, special attention for conducting research in management

deserves, first of all, the SPSS package, for which there are a number of reasons.

The SPSS package, already according to its name (Statistical Package for Social Science), is focused specifically on the application in social sciences, therefore, it contains fields, procedures and methods that understand the peculiarities of the nature of social information, including the primary information that can be obtained during applied research of employees of the enterprise.

This information really has a number of features that distinguish it from ordinary statistical data, and not all statistical processing procedures, and not all calculated statistical indicators may be applicable to it. Let us dwell on them in a little more detail in order to understand the nature of the data and subsequently choose the right processing and analysis procedures [6].

SPSS, unlike many other statistical packages and realizing that many indicators are qualitative, SPSS offers many procedures for working with quality indicators, while the vast majority of other statistical packages are focused on mathematical statistics. In SPSS, there is a concept of "Missing values" - values that do not carry an informative load, they should be excluded from the analysis. But you can work flexibly with these values: include and exclude depending on the procedure used. Another feature of SPSS: the program takes into account that people can give several answers at once. Thus, SPSS, as a program focused on the processing of social science data, is very convenient for applied research, when the source of information is people and the data obtained is of high quality, and not quantitative in nature [7,9].

At the same time SPSS also contains a complete list of mathematical statistics procedures applicable for quantitative data when the unit of observation is not people, but structural divisions, branches or organizations. All this makes SPSS a convenient and efficient data processing and analysis program for applied management research. It should be noted that SPSS, for all its versatility, cannot solve all the problems that may arise in the process of applied research [7]. So, if the collected data are predominantly quantitative in nature, then we can recommend another program that has a very developed mathematical and statistical apparatus, for example, the Statistical program. In the case when data is to be collected, programs are needed that allow organizing and linking different methods of collecting and entering information.

For example, the Vortex program for processing and analyzing sociological and marketing information. Unlike the SPSS program, which does not control the data entry process and works in command mode, Vortex is focused on automating the data collection process (telephone poles, polling on tablets, via the web interface) and quickly and visually presenting data in an interactive mode. At the same time, Vortex is significantly inferior to SPSS in the multivariate data analysis functionality. Therefore, in research practice, Vortex should be used to organize the process of data entry, data quality assessment, transformation and elementary analysis. If more complex data analysis is required, then it is necessary to convert it to SPSS and do the necessary work in SPSS [10,11].

References:

1. Ayvazyan, S. A., & Stepanov, V. S. (1997). Tools for statistical data analysis. *World of PC*, No. 8, pp. 32-41.
2. Adamov, V, E. (1977). *Factor Index Analysis (Methodology and Problems)*. (p.200). Moscow: Statistics.
3. Ayvazyan, S.A., Buchstaber, V.M., Enyukov, I.S., & Meshalkin, L.D. (1989). *Applied Statistics: Classification and Dimension Reduction*. (p.607). Moscow: Finance and Statistics.
4. Ayvazyan, S.A., Enyukov, E.S., & Meshalkin, L.D. (1993). *Applied statistics. Basics of modeling and primary data processing*. (p.471). Moscow: Finance and statistics.
5. Ayvazyan, S.A., & Mkhitaryan, B.C. (1998). *Applied statistics and foundations of econometrics. Textbook for universities*. (p.1022). Moscow: UNITI.
6. Anderson, T. (1976). *Statistical analysis of time series*. Moscow: Mir.
7. Andrukovich, P. (1973). *Application of the method of principal components in practical research.* (p.122). Moscow: Moscow State University Publishing House.
8. Afidi, A., & Eisen, S. (1982). *Statistical analysis: the approach and the use of computers*. (p.488). Moscow: Mir.
9. Molchanov, I. N., & Khadzhiev, V. (2001). Statistical software: trends and features of

development. *Questions of statistics,* No. 1, pp. 44–47.

10. Panarina, D. V. (2015). Use of statistical packages in economic information systems. *Economy and entrepreneurship*, No. 12-2 (65-2), pp. 192-194.

11. Borovikov, V.P., & Ivchenko, G.I. (2006). *Forecasting in the STATISTICA system in the Windows environment. Fundamentals of theory and intensive practice on a computer: Textbook.* 2nd ed., Rev. and add. - Moscow: Finance and Statistics.

Clarivate
Analytics indexed