

AN OVERVIEW OF STATISTICAL PACKAGES USED IN ENVIRONMENTAL PROTECTION RESEARCHES

Popovici Diana*

* University of Oradea, Faculty of Environmental Protection, 26 Gen. Magheru St., 410048, Oradea, Romania, e-mail: pdiana@uoradea.ro

Abstract

The importance of use a quantitative analyze software instrument is no more necessary to prove. In this order, the present paper intend to prezent a group of most common statistical package software and its use in the field of Environmental Protection.

Key words: statistical package software, environmental protection.

INTRODUCTION

The research methodology has reached with a large number of statistical package software, that are used in the field of Social Sciences (Sociology, Psychology, Political Sciences), Economic Sciences (like Marketing), Sciences of Life (Biology, Medicine) and Sciences of Earth (Environmental Protection). Besides the variety domain enumerated above, this software packages essentially is used for analyze quantitative data. In our Faculty of Environmental Protection, we have many field that need to analyze numerical data: Agriculture, Agro-Tourism, Animal Breeding, Control and Expertise of the Alimentary Products, Engineering and Management of Public Alimentation, Forestry, Forest Exploitation, Horticulture, Technology of Processing Agricultural Products and Wood Processing, all benefit of statistical packages use.

MATERIAL AND METHOD

For begin, we describe a number of these software products, proprietary and free, desktop or web-based, that more frequently is used. Also we reviewed some studies that assess the reliability of statistical packages.

The second goal is to assess the degree in which, the researchers in the field above described, used this instruments. In that purpose we counted the number of papers of Annals of the University of Oradea: Environmental Protection, The Ecotoxicology, Animal Breeding and Technology of Food Processing Fascicle, that was published in the first half of 2012 year.

RESULTS AND DISCUSSIONS

Assessment

Firstly is important to discuss the characteristics that make this software useful for research and teaching. In 2005, Zhu (Zhu, 2005) looking for statistical packages for use in education, found 4 criteria for assessment:

- to be free of charge;
- to be ease to learn and friendly to use, so the students can focus on the processing data and not on the learning the tool;
- to have good tutorial and teaching components;
- to contain the most common tests like Chi-square, frequency tables, univariate and multivariate analysis, ANOVA analysis, box-plots, scatter plots, histograms, and Linear regression;
- to be Windows compatible.

Another problem is the reliability of these packages, that should be demonstrate by the manufacturer. Among studies after 2000 that we find regarding the subject, there is also (Kitchen, 2003) which has assessed 2 web-based statistical software, WebStat and Statlets, concluding that they are good only for teaching.

In 2007, Keeling K. (Keeling, Pavur, 2007) has tested the reliability of 9 statistical packages using The American National Institute of Standards and Technology (NIST) data sets, regarding distribution tests, univariate summary statistics, one-way ANOVA, linear and nonlinear regression; one of these packages is Microsoft Excel in two variants Excel 2000/XP and Excel 2003; another group is of proprietary software: JMP 5.0, Minitab 14.0, SAS 9.1, Splus 6.2, SPSS 12.0 and Stata 8.1; free software assesses was R 1.9.1 and StatCrunch 3.0 (online). Among his findings are also some problems with nonlinear regression at Excel 2000 and 2003 Spreadsheets.

In 2010, Odeh O. (Odeh et al, 2010) has tested, with the same NIST linear and nonlinear least squares datasets and models, following packages: Excel 2007, GAMS 23.4, GAUSS 9.0, LIMDEP 8.0, Mathematica 7.0, MATLAB 7.5, R 2.10, SAS 9.1, SHAZAM 10, and Stata 10; according to his findings, first is necessary to test with benchmark software packages the tool before use and second is necessary to use more than one package to solve the complex problem.

In 2011 the same author couple (Keeling, Pavur, 2007) has tested six spread sheet packages, 3 proprietary: Excel 2007 and 2010 (Microsoft), Numbers '09 (Apple) and Quattro Pro X4 (Corel), respectively 3 free products: Google Docs, Gnumeric 1.10.14 and OpenOffice Calc 3.3.0, using the same data sets. They find that Excel 2010 was improved, but the most reliable spreadsheet of the studied group is Gnumeric, while the least remain Google Docs (which lack nine tested distribution and ANOVA procedure).

Description

We choose for a short description 7 of the most popular statistical packages used in the publication that we consult, 4 of them proprietary software and 3 free, respectively 6 of them desktop and 2 with online versions.

SPSS¹: is a strong package, now produced by IBM Company. Is has implemented Data Mining models, like Decision trees and Nonlinear regression methods (Gorunescu, 2006); It has 6 components: IBM Analytic Answers, IBM Analytical Decision Management, IBM Social Media Analytics, IBM SPSS Data Collection, IBM SPSS Modeler and IBM SPSS Statistics (IBM, 2013). It accepts data bases from dBase, MS Excel and MS Access. In Romania is sold by IBM Romanian subsidiary and Romsym Data Company.

Statistica: Like SPSS, Statistica has implemented Data Mining models Decision trees and Nonlinear regression methods (Gorunescu, 2006); is created by StatSoft Company, actually running the 10's version. He has no Romanian office (Statsoft, 2013).

Minitab: Is produced by Minitab Inc. and it has two European Offices at Coventry (GB) and Paris (Minitab, 2013). It has an interesting animated video guide for learning, named Quality Trainer. Romsym Data and XTC Computers sell Minitab 16 in Romania too.

Table 1

More often used Statistical packages classification

	Desktop	Online version
Proprietary	SPSS Statistica Minitab GraphPad Prism	GraphPad QuickCalcs
Free	R GPower Epi Info	OpenEpi

GraphPad Software: It has 3 components, GraphPad Prism 6 (statistical tests), InStat (for beginners) and StatMate (for choosing the sample). It has an online calculation option, GraphPad QuickCalcs. It can be buy directly thru download from the producers site (Graph Pad, 2013).

R-Project: Is a GNU Project; GNU was a project to produce a clone of Unix Operating System for free use; afterwards the promoter of the project, Free Software Foundation, has elaborated the GNU General Public License that established the status of free software. Recently, in April 2013, was release the R 3.0.0 version of product (R-Project, 2013).

¹ Statistical Package for Social Science

GPower: It's a German package, created at University of Dusseldorf and reached version 3. It has still unfinished documentation (GPower, 2013).

Epi Info 7: It is a free statistical package for epidemiology, created at Center for Disease Control, Atlanta, Georgia, SUA; It has an open version, OpenEpi than can use thru a web connection or in a Smartphone browser (Epi Info, 2013).

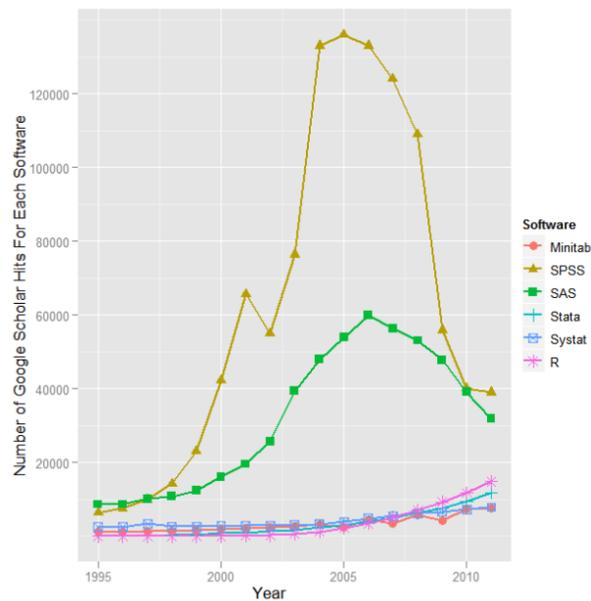


Fig. 1. Use of data analysis software in academic publications as measured by hits on Google Scholar (Source: Muenchen R., <http://r4stats.com/articles/popularity/>)

Popularity

About the popularity, a great numbers of information was collected and updated by Muenchen Robert (Statistical Consulting Center, University of Tennessee, USA) and published on a website (Muenchen, 2013); in number of software used in data analysis competitions in 2011, lead R followed by Matlab and SAS, in a survey of Rexer Analytics with the question 'What Data mining/analytic tools did you use in 2009?', lead R followed by SAS and IBM SPSS, and in number of use in academic publication as measured by hits on Google Scholar, lead SPSS followed by SAS, Fig. 1 and Fig. 2.

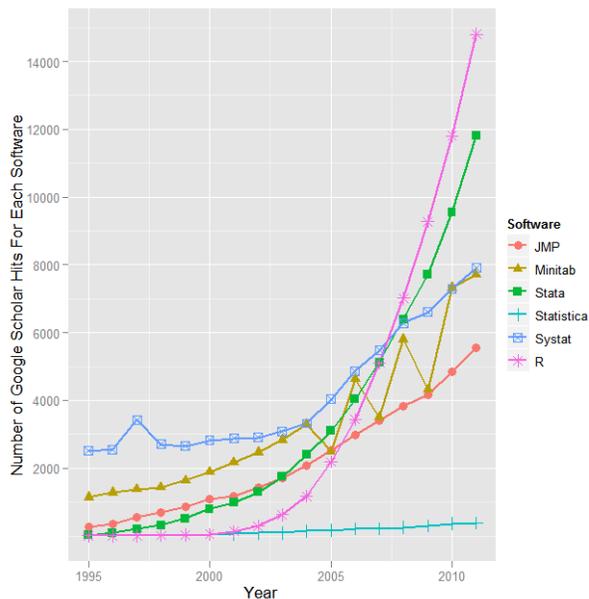


Fig. 2. Use of data analysis software in academic publications as measured by hits on Google Scholar, excluding SAS and SPSS (Source: Muenchen R., <http://r4stats.com/articles/popularity/>)

Use of Statistical Packages

The research was conducted on a total of 31 papers; in the Figure 3, we can see the repartition of papers after the following criteria:

- A: Use of statistical processing, Tests and Package/Spreadsheet specified (3%);
- B: Use of statistical processing, Tests specified, Package/Spreadsheet unspecified (13%);
- C: Use of statistical processing but Tests and Package/Spreadsheet unspecified; use of statistical indicators (10%);
- D: Statistical processing not use (74%).

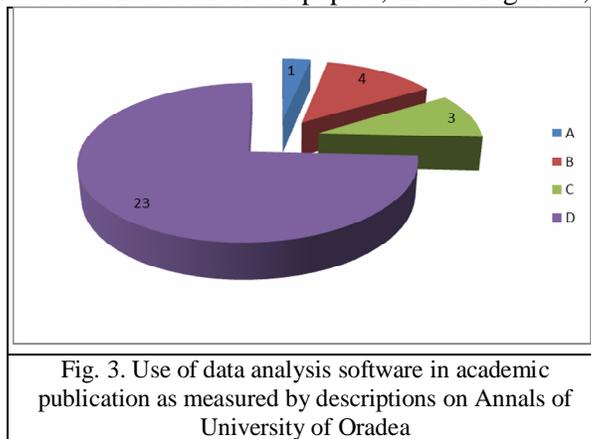


Fig. 3. Use of data analysis software in academic publication as measured by descriptions on Annals of University of Oradea

CONCLUSIONS

From this little research it follows that a large number of statistical package software are available, including free online versions, relatively reliable; despite this, in academic publication Annals of University of Oradea we counted a low level of usage of quantitative analysis software; next step is to study the other publications with the same profile, to compare the statistical tests usage. Beside, this conclusion must be interpreted cautiously because were not excluded from total of papers those, which are not an experimental work.

REFERENCES

1. Gorunescu F., 2006, Data Mining- concepte, modele si tehnici, Ed. Albastră, Cluj-Napoca
2. Keeling K., B, Pavur R., J., 2007, A comparative study of the reliability of nine statistical software packages, Computational Statistics & Data Analysis 51 (2007), pp. 3811-3831
3. Keeling K., B., Pavur R., J., 2011, Statistical Accuracy of Spreadsheet Software, The American Statistician, 65:4, pp. 265-273
4. Kitchen A., M., Drachenberg R., Symanzik J., 2003, Assessing the reliability of web-based statistical software, Computational Statistics, March 2003, Volume 18, Issue 1, pp. 107-122
5. Odeh O., Featherstone A., Bergtold J., 2010, Reliability of Statistical Software, American Journal of Agricultural Economics, 10/2010, Volume 92, Issue 5, pp. 1472-1489
6. Zhu X., Kuljaca O., 2005, A Short Preview of Free Statistical Software Packages for Teaching Statistics to Industrial Technology Majors, Journal of Industrial Technology, Volume 21, No. 2
7. * * * Epi Info, <http://www.cdc.gov/epiinfo/>
8. * * * Graph Pad, <http://www.graphpad.com/demos/>
9. * * * IBM, <http://www-01.ibm.com/software/analytics/spss/>, consulted on 1 march 2013
10. * * * Minitab Inc., <http://www.minitab.com/en-RO/default.aspx>
11. * * * Muenchen R., <http://r4stats.com/articles/popularity/>, consulted on 1 march 2013
12. * * * R-Project, <http://www.r-project.org/>
13. * * * Statsoft, <http://www.statsoft.com/#>