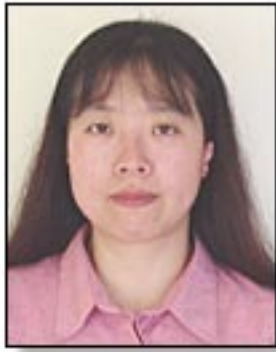# A Short Preview of Free Statistical Software Packages for Teaching Statistics to Industrial Technology Majors

By Ms. Xiaoping Zhu and Dr. Ognjen Kuljaca

NON-Refereed Article

KEYWORD SEARCH

Curriculum
Quality Control
Research Methods
Statistical Methods
Teaching Methods

A Short Preview of Free Statistical Software Packages for Teaching Statistics to Industrial Technology Majors

By Ms. Xiaoping Zhu and Dr. Ognjen Kuljaca

Ms. Xiaoping Zhu is PhD student of Mathematics the Department of Mathematics at the University of Texas at Arlington. Her current research focuses on preliminary test estimators of process capability indices and multivariate regression analysis.

Dr. Ognjen Kuljaca is Assistant Professor of Advanced Technologies at Department of Advanced Technologies at Alcorn State University. His research interests includes intelligent control, fuzzy systems, computer integrated manufacturing, and process identification. Dr. Kuljaca is NAIT Certified Senior Industrial Technologist (CSIT).

Abstract: *The paper gives a short preview of several free statistical packages. The main characteristics of ViSta, AM, IDAMS, IRRISTAT, OS3, PAST and Instat+ are described. The web addresses where the software can be found are given.*

Introduction

The role of statistical analysis and control in modern industrial technology is quite clear. Industrial quality control, control management, market analysis, production analysis, manufacturing analysis, and many other functions cannot be performed without certain knowledge of statistics. This leads to the problem of choosing a tool for teaching statistics and subsequent activities that require statistical content. Excel is a much-used tool and there are add-ins for Excel that can greatly improve its value in statistical analysis. Yet Excel is not in its essence a statistical tool. The other approach is to use languages and packages like R or Scilab that are able to perform statistical functions. However, these languages require extensive learning and training. The user has to learn programming with them well before they can be put into good use. Of course, there are the professional commercial statistics packages that allow command line input and programming as well as use of graphical user interface (GUI) analysis. However, those are often expensive and in many cases beyond the means of the students or individual users.

The authors of this paper performed a small web search to find free statistical packages that could offer functionality and ease of use. There are quite a number of different packages and programs that have statistical functionality and duplicating such a search could be quite time consuming. We hope that the information provided in this paper will help the reader save some time in searching for a statistical tool.

Requirements on the Statistical Software Package

The first question to be answered is what requirements should a statistical software package fulfill to make efficient low-cost teaching possible? There are many software packages with different characteristics and abilities. The three that have established themselves within the statistical community are SPSS, SAS and S-Plus. These three are indeed professional tools with great capabilities. Consequently, they are very expensive tools, sometimes not easy to use, and most of the time beyond the financial capabilities of most industrial technology departments. In addition, unless statistics and data analysis is their primary activity, most companies where students will be professionally employed, will not have these packages.

The fact is that most statistical functions needed in industrial technology are in the areas of statistical process control, process capability indices and sampling, and analysis of samples. This means that the amount of data to be analyzed is not extensively high. There are free statistical tools available able to meet the requirements of statistical analysis in industrial technology. So, here is our first requirement. The package should be free. That makes it more available to students.

The second requirement is ease of use. The learning curve should not take too much time. If a software package is complicated to use and learn, the class will degrade into tackling the software use and programming problems, not investigating statistics, sampling, process control or whatever is the main objective of the class. Ease of use for statistical software generally means that reading the data is simple, at least several formats of input data are available, the analysis is simple with a number of prewritten functions, the package has a GUI which is simple and easy to understand, accompanying documentation is easy to read, and software has good capabilities in producing graphs and plots.

The third requirement is tutorial and teaching components of the software. If there are such functions incorporated into the package, for example explanations of some functions, probability distributions, statistical plots and various exercises, teaching is much easier and consequent study is much simpler for the students.

Of course, a certain number of statistical functions and procedures should be supported so that the most common tasks can be performed without extensive additional programming effort. Linear regression, univariate and multivariate analysis, box-plots, normal, Chi-square and t distribution generation, scatter plots and histograms, frequency tables and ANOVA analysis should be the minimum functions available by the software.

Finally, since most of the students and faculty use Windows operating systems, we have concentrated on packages that can run under Windows. Needless to say, there are many excellent tools for Mac OS or Linux operating systems and these might be a topic of another discussion, but a vast majority of users are running Windows on their computers.

Here, we will describe some characteristics of several free software packages and give some illustrations of how they work and what they can do.
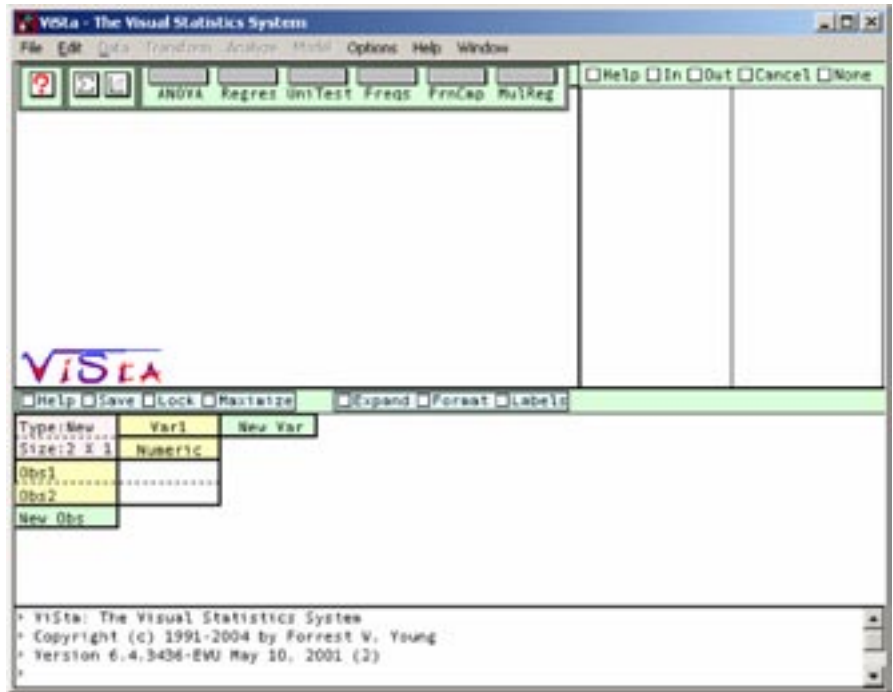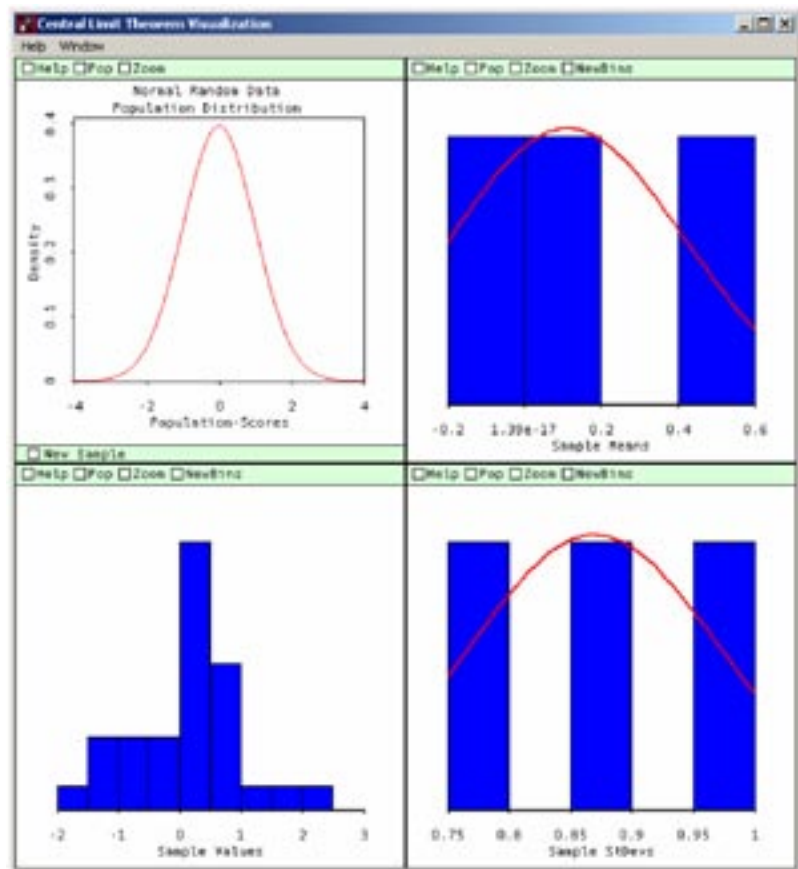
Figure 1: ViSta initial GUI

Figure 2: Central limit theorem visualization - ViSta

Software Packages

Several software packages will be described in this section. We will start with ViSta and continue with AM, WinIDAMS, IRRISTAT, OS3, PAST and Instat+. Of course, there are others, but space and limit the discussion.

ViSta is a very nice graphical tool that can work as standalone application or with Excel, and has nice plotting capabilities. Its GUI could be more appealing (Figure 1). There are some bugs, and to perform multivariate analysis, lognormal analysis, homogeneity analysis, multilevel analysis and application development you have to download plug-ins. The current version of the software is 6. It is developed by Professor Forrest W. Young. It can be downloaded from www.visualstats.org. ViSta runs on Windows, UNIX and Mac OS. It is open and free.

ViSta allows users to define new functions and software, to build web applets for the students and to write scripts. It comes with teaching functions that actually explain what is going on. It can execute LISP, C++ and FORTRAN programs.

Here is an example of a teaching tool and help available from ViSta: If you want to generate and analyze some data, you can choose the "Simulate data" option from File menu, choose your distribution and immediately obtain the central limit theorem visualization (Figure 2). Different multivariate spreadplots are obtained simply by clicking on the Graph button (Figure 3). Each plot window has a Help button. When the Help button is checked, a pop-up window appears explaining exactly what is shown on the plot, how it is constructed, and in what kind of analysis it could be used.

ViSta can read SAS data and Excel spreadsheets and free-form text data files. However, more functionality in the applications' input data files would be welcome.

AM is a statistical package from American Institute for Research. The

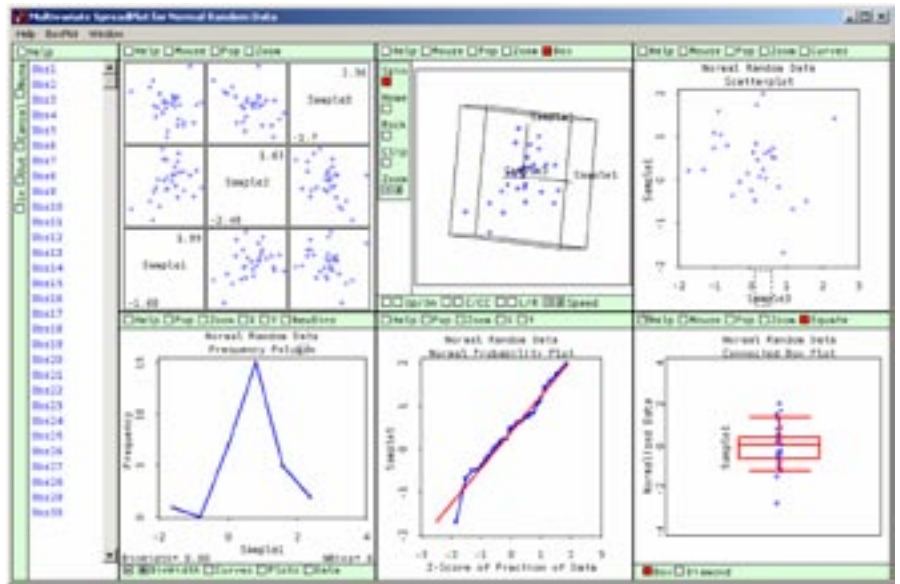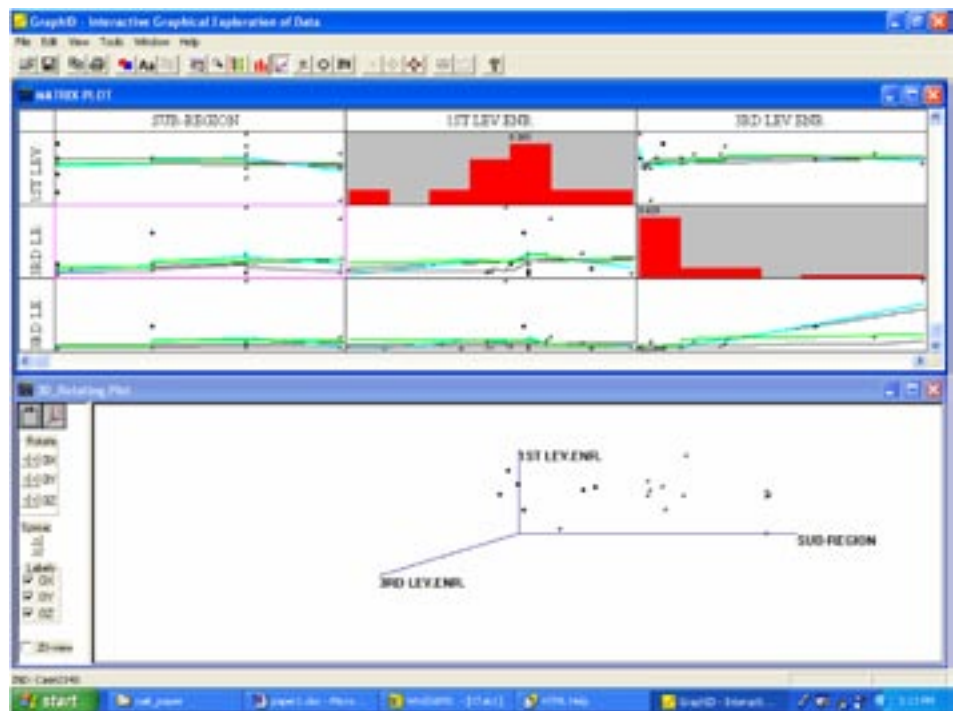*Figure 3: Multivariate spread plot for normal data from 3 samples - ViSta*

Figure 4: Graphical analysis window of WinIdams

software can be used freely for non-commercial purposes. It can be downloaded from http://am.air.org/. As stated at their web page, AM has the following features:

1. AM offers sophisticated statistics with an easy-to-use drag and drop interface, and an integrated help system that explains the statistics as well as how to use the system

2. AM includes statistical procedures for analyzing assessment and survey data, including models estimated via marginal maximum likelihood

(MML). MML procedures recognize that tests can't pinpoint proficiency, but instead define a probability distribution over the proficiency scale. AM will also analyze the "plausible values" used in programs like NAEP.

3. AM includes procedures for analyzing non-assessment survey data as well.
4. All of AM's procedures automatically provide appropriate standard errors for complex samples. All procedures are available with a Taylor-series approximation for the standard errors, and a few offer an option to use a jackknife or other replication technique.
5. AM is completely free and may be shared and copied. It was developed, in part, with funding from the National Center for Education Statistics.

AM can import data from about 100 different file formats, including SAS and SPSS. The manual that comes with the software is well documented and the GUI is easy to use. The manual gives a mathematical description of the functions and procedures. AM is primarily aimed at survey data analysis. It allows univariate and multivariate analysis. The graphing capabilities are limited to bar chart, line chart and density plots. Command line programming, script writing and random number generation are not supported.

WinIdams is a software package for data analysis developed by UNESCO. It allows users to define their own programs, but it is also capable of interactive data analysis through WinIdams GUI. Plotting capabilities are good (Figure 4). The manual in HTML form is easy to read and also comes with an explanation of plots and functions. WinIdams can perform univariate and multivariate analysis, linear regression, ANOVA, frequency tables and the other most common statistical procedures. The results are output in the form and table form. WinIdams is free and can be downloaded from http://www.unesco. org/idams.

IRRISTAT can perform univariate and multivariate analysis, regression, ANOVA and other usual statistical

Figure 5: PAST spreadsheet with plot windows
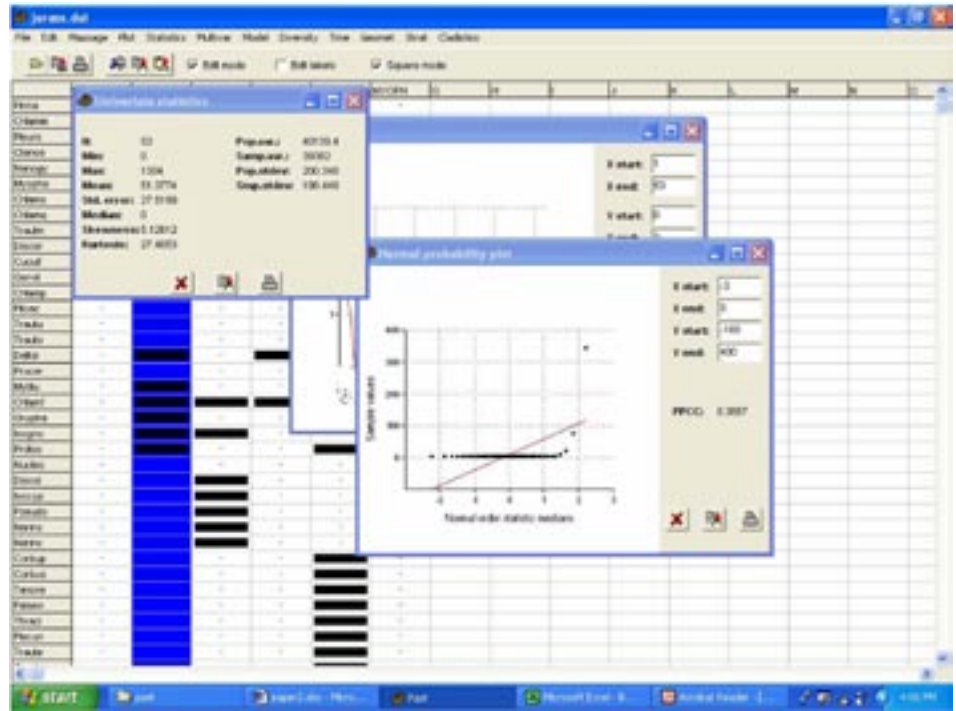
procedures. It comes with examples and a tutorial. However, it does not allow the user to make programs and new functions. Everything is done via a GUI that is well developed and easy to use. The graphical capabilities are not very good though and only scatter and biplots can be generated. IRRISTAT has been developed primarily for the analysis of data from agricultural field trials and that might account for its somewhat limited capabilities in general data analysis. The software can be downloaded from http://www.irri. org/science/software/irristat.asp.

OpenStat3 (OS3) is developed by William G. Miller. It can perform univariate and multivariate analysis and has a set of statistical process control tools. The drawback is that it does not allow programming and defining new functions. Also, only OS3, OS2, comma separated, tab separated and space separated files can be imported. Otherwise, everything is performed through a GUI and pop-up windows. The software is free and can be downloaded from http://www.statpages.org/miller/openstat/OS3.html.

PAST is developed primarily for executing a range of standard numerical analyses and operations used in quantitative paleontology. It integrates spreadsheet type data entry with univariate and multivariate statistics, curve fitting, time-series analysis, and data plotting. Its GUI is natural and easy to use. The package can also perform regression and most of the other standard functions. Plotting capabilities are quite good. However, programming is not allowed and there is no good tutorial or help available. A PAST spreadsheet with plot windows is shown in Figure 5. Package can be downloaded from www.nhm.uio.no/~ohammer/past/.

Instat+ can be downloaded from http://www.reading.ac.uk/ssc. It has somewhat questionable plotting capabilities, but it is very well suited for teaching statistics. In fact, that was one of the main intentions of the authors of the software. All the most common statistics functions are supported, including multivariate analysis but not nonlinear regression. In fact, most free packages do not support nonlinear techniques. Instat+ is free for single noncommercial

users, but it requires a one-time department site license fee. It is very much based on a worksheet approach and it does not allow programming (although macros can be used).

Conclusion

Seven software packages were described. While all of them have comparable GUI capabilities, it appears that ViSta and WinIdams offer the most functionality. The ability of defining user programs offered by both packages is important for more advanced uses. The ease of use by GUI helps in teaching statistics so that students can concentrate on the learning statistical principles and analysis, not spending a lot of time and effort in learning the software.

Graphic abilities of both software packages are quite good. ViSta might have a small advantage in easy access to help functions and tutorial and educational contents. On the other hand, WinIdams is an effort of UNESCO and it might in the future be developed faster or better maintained than ViSta.

The other five software packages have some nice capabilities, but it is questionable how they can be used in more advanced procedures if the user cannot write user-defined programs and scripts.

References

http://am.air.org/, retrieved August 30, 2005

http://www.irri.org/science/software/irristat.asp, retrieved August 30, 2005

www.nhm.uio.no/~ohammer/past/, retrieved August 30, 2005

http://www.reading.ac.uk/ssc, retrieved August 30, 2005

http://www.statpages.org/miller/openstat/OS3.html, retrieved August 30, 2005

http://www.unesco.org/idams, retrieved August 30, 2005

www.visualstats.org, retrieved August 30, 2005