COMPARATIVE REVIEW OF STATISTICAL ANALYSIS PROGRAMS

*Kiranpal Singh Virk, **Dr. Munishwar Rai, ***Prabhjot Singh

* Guru Nanak Khalsa College, Yamuna Nagar, Haryana 135001, India,
pursuing Ph.D. from Maharishi Markandeshwar University, Mullana
** M.M. Institute of Computer Technology & Business Management,
Maharishi Markandeshwar University, Mullana (Ambala), Haryana, India
***B.E Third Year, Department of Computer Science and Engg
Dayananda Sagar College Of Engineering, Bangalore – 560078

ABSTRACT

Statistical analysis programs have an indispensible role in the research process and more specifically in data analysis. Over the time, many such programs have surfaced and the contemporary researcher has always been bogged down by the cloud of indecisiveness when it comes to choose a Statistical analysis program for their research. This paper aims to take a detailed feature based study of currently available statistical analysis programs like SPSS and SYSTAT from the perspective of users who analyze surveyed data. It describes the programs' strengths, weaknesses and hardware demands.

Keywords: *statistical analysis programs, SPSS, SYSTAT, research process, data analysis*

INTRODUCTION

A formal research process plays an important role in a good research. It consists of sequence of actions and steps necessary to carry out research efficiently. Figure 1.1 charts out the necessary steps.

Where

(F) = feedback (Helps in controlling the sub system to which it is transmitted)

(FF) = feed forward (Serves the vital function of providing criteria for evaluation)
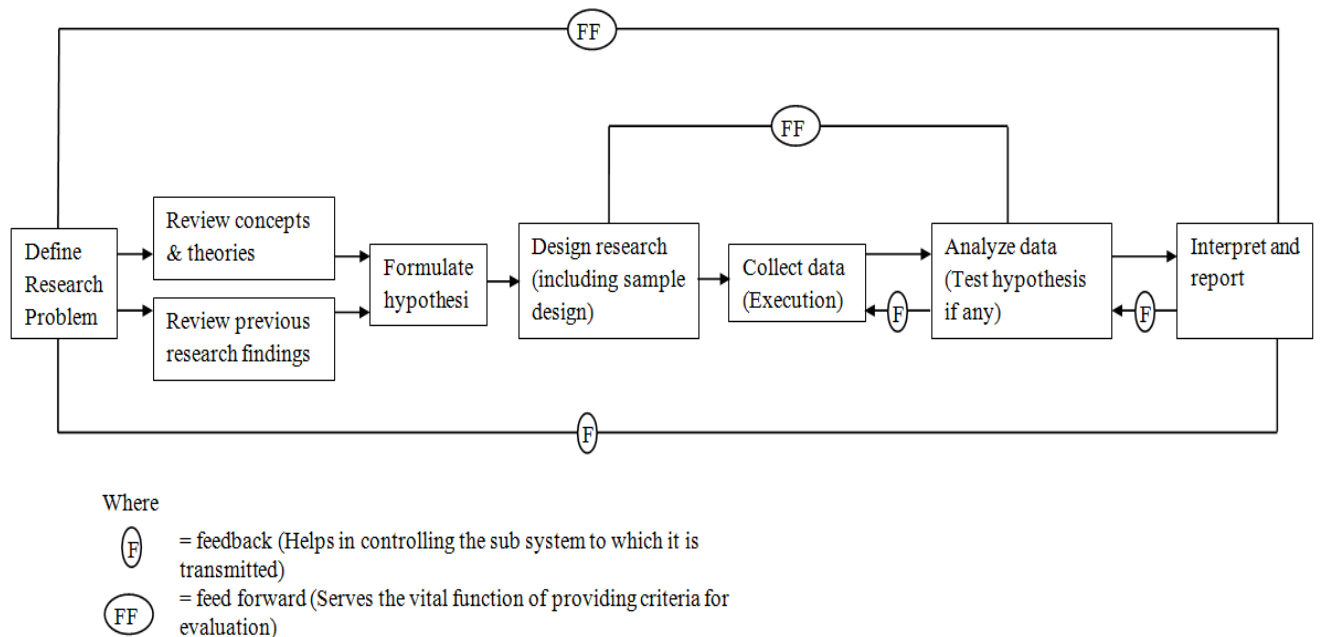
Figure 1.1

Although all steps are equally important; for the scope of this paper we are focussing on data analysis. The task of analysing is an effort consuming activity and involves mathematical and statistical calculations. For example, analysis on tabulation is generally based on the computations of various percentage, coefficients etc by applying well designed statistical formulae. In some other analysis, causal-effect relationships testing (like test of significance) may be required to determine with what validity data can be said to indicate any conclusion.

Analysis being a calculation intensive task, the chance of inadvertent error also increases. And for this the reasons are simple. The foremost been that it is a repetitive process and humans tend to lose interest in the recurring activity sooner or later. To overcome this shortcoming, various computer software based tools and techniques can be employed. Various software available in the market cater to specific needs. Every research being a discovery of new knowledge has unique requirements and thus existing tools may not suffice. A detailed and informed decision is required before a researcher makes use of a tool for the data analysis. This paper is an effort in the same direction.

LITERATURE SURVEY

The past reviews on the comparable apparatuses have been directed every once in a while and a few analysts have inspected measurable programming for microcomputers and offered exceptionally valuable remarks to both clients and sellers. Some of these audits are extensive and general (Searle, S. R. (1989). Some others dissect particular program highlights and distinguish issue zones. For instance, Gerard E. Dallal (1992) distributed an exceptionally succinct paper through the American Statistician titled "The computer

analysis of factorial experiments with nested factors". Okunade, An., and others (1993) looked at the yield of rundown insights of relapse examination in regularly measurable and econometrical packages, for example, SAS, SPSS, SHAZM, TSP, and BMDP. Oster, R. A. (1998) inspected five measurable programming bundles (EPI INFO, EPICURE, EPILOG PLUS, STATA, and TRUE EPISTAT) as per criteria that are of most interest to disease transmission experts, biostatisticians, and others involved in clinical research.

Zhou, X., and others (1999) checked on five programming bundles that can fit a summed up direct blended model for information with more than a two-level structure and a different number of free factors. These five bundles are MLn, MLwiN, SAS Proc Mixed, HLM, and VARCL. The correlation between these bundles was based upon a few elements, for example, information info and administration, statistical model capacities, yield, ease of use, and documentation. Bergmann, R., and others (2000) Compared 11 measurable bundles on a genuine dataset. These bundles are SigmaStat 2.03, SYSTAT 9, JMP 3.2.5, S-Plus 2000, STATISTICA 5.5, UNISTAT 4.53b, SPSS 8, Arcus Quickstat 1.2, Stata 6, SAS 6.12, and StatXact 4. They found that distinctive bundles could give altogether different results for the Wilcoxon-Mann-Whitney test.

In 2009, after IBM acquired SPSS, it has produced varied interest from IBM client group. However the SYSTAT has additionally advanced overtime and has a noteworthy shot when SYSTAT 13 got reviewed in Scientific Computing by Dr. John Wass in 2010.The author, Dr. John Wass, writes in part:

"Despite the power and depth in this version, the most impressive feature is the inherent simplicity. In the early versions the novice could easily get frustrated in deciphering what the dialog box was really asking for and feeding it the proper data types. This is to say nothing about approaching the level of sophistication required to fully understand the depth of the statistical results presented. Now everything seems so intuitive that as with many programs, the learning curve is tilted in favor of the novice (though it still requires some statistical understanding on the part of the analyst!)".

STATISTICAL ANALYSIS PROGRAMS

Many software tools have been available in the market from time to time. From the survey, the information about common tools may be summarized as shown in Table 1.1

Table 1.1

Software Name	Advantages	Disadvantages	Open source	Typical users
R	Library support; visualization	Steep learning curve	Yes	Finance; Statistics
Matlab	Elegant matrix support; visualization	Expensive; incomplete statistics support	No	Engineering
Excel	Easy; visual; flexible	Large datasets	No	Business
SAS	Large datasets	Expensive; outdated programming language	No	Business; Government
SAS (Statistical Analysis System)				
Stata	Easy statistical analysis		No	Science
Stata (The name Stata is a syllabic abbreviation of the words statistics and data)				
SPSS	Easy statistical analysis	Expensive	No	Social Science
SPSS (Statistical Package for the Social Sciences)				
SYSTAT	Easy statistical analysis	Expensive; but not so as compared to SPSS	No; free for academic use by students	Social Science
SYSTAT (System Statistics)				

From the Table 1.1, it is very much evident that many tools are available. However; our paper focuses around two software namely:- SPSS and SYSTAT. First reason for selecting SPSS is that it has very low learning curve in comparison to other tool as per NYU Data Services (Figure 1.2). Secondly the reason for selecting SYSTAT is that it has gained very high popularity recently among the researchers and it is free for academic research purpose.
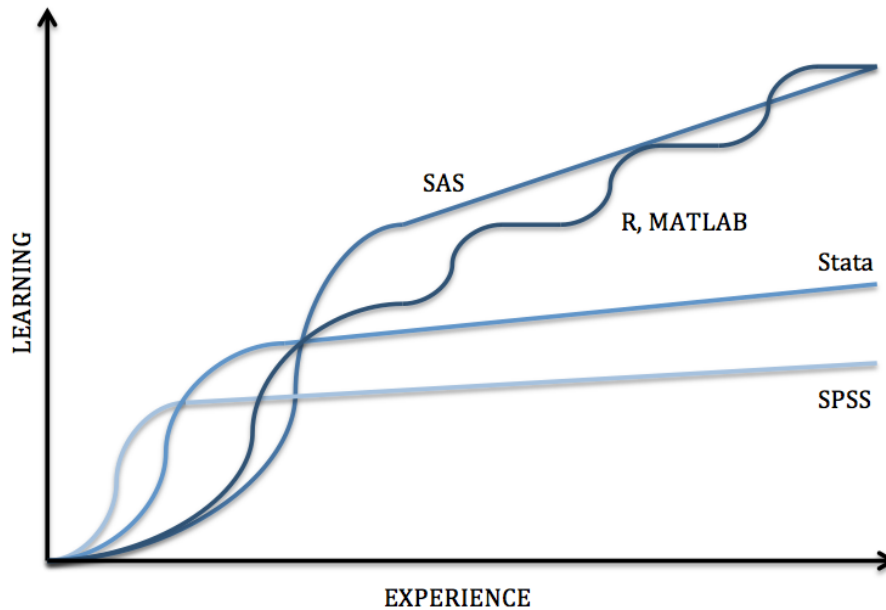
Figure 1.2

SPSS vs SYSTAT

Before we go into the detailed comparative, let us have a look at the feature wise cost comparison of SPSS and SYSTAT as shown in the Table 1.2. From the tabulated information it is clear that the cost factor definitely gives SYSTAT an edge over the SPSS.

Table 1.2

Functionality (Commercial license pricing as on 10-Dec-2015)	SYSTAT (Single User)	SPSS (Single User)
Base Package	$1,499	$1,799
Advanced Statistics Functions (e.g. GLM)	Included	$999
Correspondence Analysis	Included	$969
Conjoint Analysis	Included	$769
Decision Trees	Included	$969
Forecasting	Included	$969
Missing Values	Included	$999
Regression	Included	$999
Bootstraping and Markov Chain Monte Carlo	Included	$999
Robust Regression	Included	Not Offered
Total	$1,499	$9,471

21

When we compare the earlier versions of SYSTAT and SPSS for the feature like data editor, the SYSTAT lags behind the SPSS. However, the Data Editor in SYSTAT 13 looks a lot like SPSS's as shown in Figure 1.3

Figure 1.3

Much like SPSS, SYSTAT variable names can contain up to 256 letters or numbers and must begin with a letter or an underscore. However, names of character, or string, variables must end with a dollar sign ($), which counts as a character. If you create a string variable, SYSTAT will automatically add the dollar sign on the end for you. If you change the variable to numeric, it will remove the dollar sign. And for this reason, the string variables naming in SYSTAT is truncated/limited to 255 characters.

Both SPSS and SYSTAT perform accurately and nearly always reliably. They provide excellent file and data manipulation capabilities, a broad range of basic analytical tools, and many of the more advanced methods like descriptive statistics including frequency distribution, counts and chi-square. However, SYSTAT output is HTML based, so it easily copies into Word and other word processing programs without losing its formatting.

The other features where the earlier versions of the SYSTAT lagged was compatibility with different file formats. With the current version, the SYSTAT 13 wins hands down as it allows opening various file formats as shown in Figure 1.4
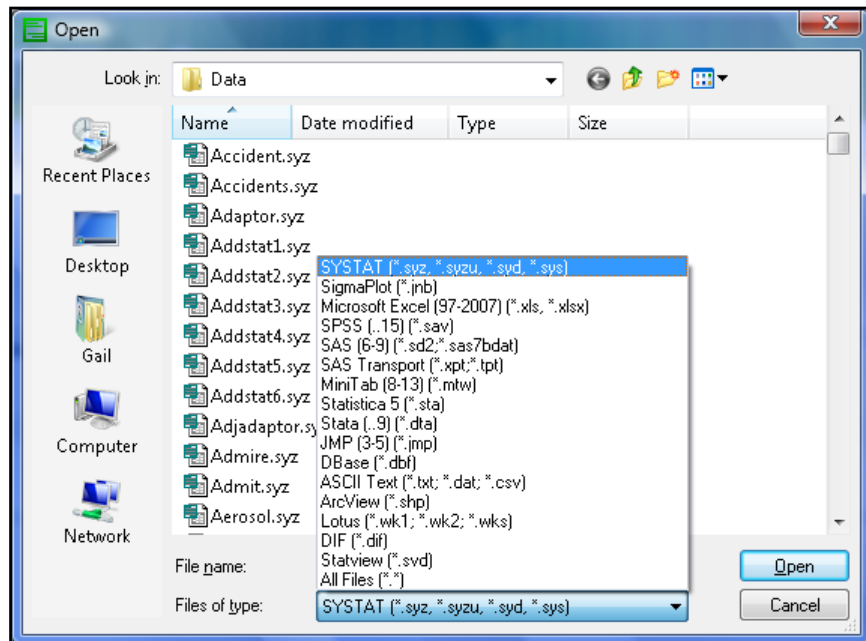
Figure 1.4

In nutshell the features of both the tools could be summarized as shown in Table 1.3-1.6

Table 1.3

General Features		
Product	**SPSS**	**SYSTAT**
Developer	IBM	Systat Software Inc.
Latest version	March 3, 2015	Feb 21, 2007
Open source	No	No
Software license	Proprietary	Proprietary
Interface	CLI/GUI	CLI/GUI
Written in	Java	

Table 1.4

Operating System Support		
Product	SPSS	SYSTAT
Windows	Yes	Yes
Mac OS	Yes	Terminated
Linux	Yes	No
BSD	No	No
Unix	No	No

Table 1.5

ANOVA Methods		
Product	SPSS	SYSTAT
One-way	Yes	Yes
Two-way	Yes	Yes
MANOVA	Yes	Yes
GLM	Yes	Yes
Mixed model	Yes	
Post-hoc	Yes	Yes
Latin squares	Yes	Yes

Table 1.6

Regression Methods		
Product	SPSS	SYSTAT
OLS	Yes	Yes
WLS	Yes	Yes
2SLS	Yes	Yes
NLLS	Yes	Yes
Logistic	Yes	Yes
GLM	Yes	Yes
LAD	No	Yes
Stepwise	Yes	Yes
Quantile	No	No

23

CONCLUSIONS

Based upon the above discussion that focused on the costing aspects and other features aspects like representation of the data sets, descriptive statistical capability, interoperability with numerous file formats, ease of use and free for academic use and research purpose, it can safely concluded that SYSTAT 13 does not lag behind SPSS. Currently SYSTAT development is being taken care of by Cranes Software International lead by its original developer Dr. Wilkinson who along with his team is focussing on the revival of the free MYSTAT student distribution. Both SPSS and SYSTAT 13 are very much similar in the features. However, due to high cost factor SPSS lags behind the SYSTAT13. Hence, it may be safely concluded that SYSTAT 13 may be considered as statistical tool for the academic research.

REFERENCES

1. SPSS Site http:// spss.co.in/
2. SYSTAT Site http://www.systat.com/
3. NYU Data Services http://guides.nyu.edu/dataservices (Data Services is a joint service of New York University's Division of Libraries and Information Technology Services to support quantitative, qualitative, and geographical research at NYU)
4. Oster, R. A. (1998), " An examination of Five Statistical Software Packages for Epidemiology," The American Statistician, 52,267-280.
5. Ibrahim Hassan Ibrahim. Differences Between Statistical Software Packages (SAS, SPSS, and MINITAB) As Applied to Binary Response Variable
6. C.R. Kothari, Research Methodology: Methods and Techniques, New age International (P) Limited Publishers, 2nd Edition, Page 9-13