ISSN 2456-432X SJIF:4.819 JIF: 0.9127

IJOOAR

INDIAN JOURNAL OF ODYSSEY OF AYURVEDIC RESEARCH

Descriptive study of Statistical Software's as application tool of Data management

1. Dr.Ravasaheb.M.Yallatti (Asso. Prof.),

V. P. Institute of Management Studies and Research Sangli

2. Dr.Shrikant Joshi (MD & CEO)

Joshi Publication House, Pune

ABSTRACT:

Stephanie D. (2009), Standford PhD Statistical Consulting and Karen (2013), Wikipedia library and other scholars have identified different popular statistical software programs, which are **SPSS, Eview, EPI INFO, SAS, MATLAB, MINITAB, STATA etc** and lots more have been utilized by people across all disciplines for many years and are quite user friendly. Statistical The software can either read data directly from EXEL spreadsheet, an user can enter the data directly to software ,or has specialized data entry software to capture data.The statistical softwares then manipulates the information they possess to discover patterns which can help the user uncover the predictions and help to interpret the data.

Keywords : **SPSS, Eview, EPI INFO, SAS, MATLAB, MINITAB, STATA**

INTRODUCTION :

Statistical software's are programs which are used for the statistical analysis of the collection, organization, analysis, interpretation and presentation of data . Statistical analysis is the science of collecting, exploring and presenting large amounts of data to discover underlying patterns and trends and these are applied every day in research, industry and government to become more scientific about decisions that need to be made .Statistical software's helps in analysis of data.

The software can either read data directly from EXEL spreadsheet, an user can enter the data directly to software ,or has specialized data entry software to capture data.

The statistical softwares then manipulates the information they possess to discover patterns which can help the user uncover the predictions and help to interpret the data.

COMMON STATISTICAL SOFTWARE AND THEIR APPLICATION TO DATA ANALYSIS:

Stephanie D. (2009), Standford PhD Statistical Consulting and Karen (2013), Wikipedia library and other scholars have identified different popular statistical software programs, which are **SPSS, Eview, EPI INFO, SAS, MATLAB, MINITAB, STATA etc** and lots more have been utilized by people across all disciplines for many years and are quite user friendly.

FEATURES OF STATISTICAL SOFTWARE:

Statistical software has some common characteristics that make it reliable and suitable for data analysis:

1. Data editor is in rows and columns which make it very easy to enter numeric data.
2. There is availability of menu bar comprises drop-down menu, quick analysis as well as brief user manual.
3. Statistical level of measurement is put into consideration in data entry
4. They follow the initial steps in research project
   (a) Getting your data ready to enter into the software.
   (b) Defining and labeling variable

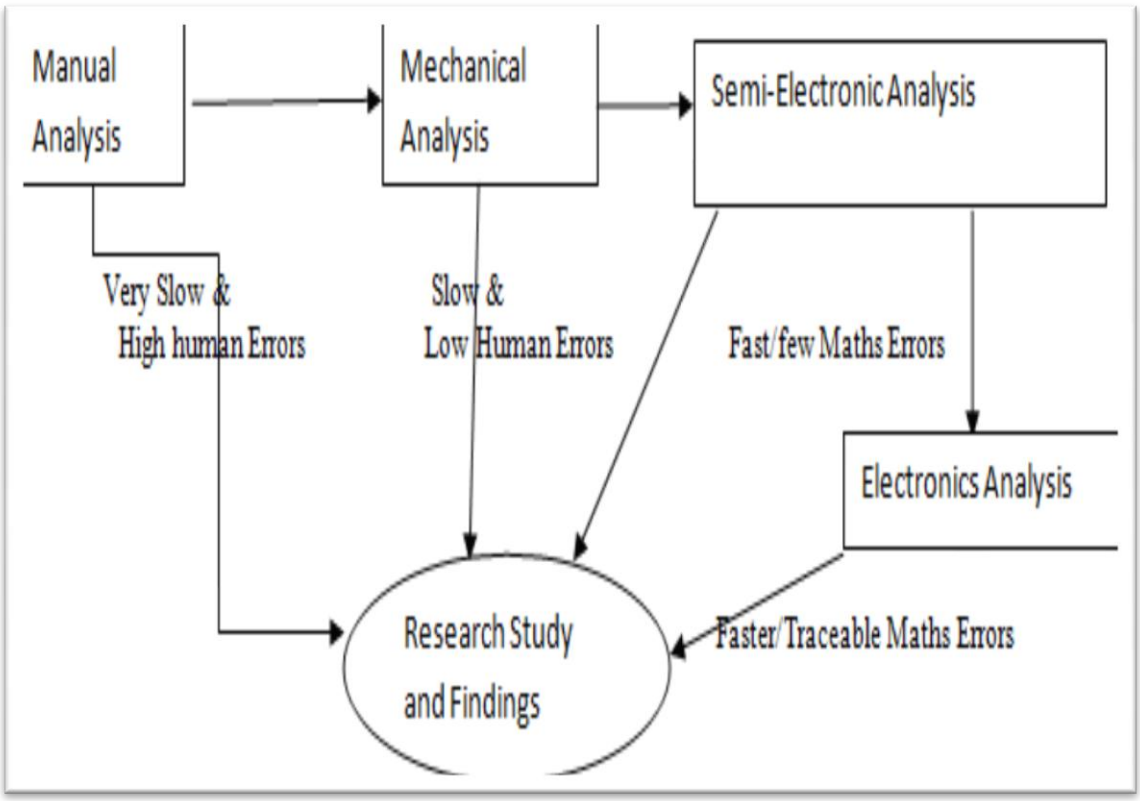
(c) Entering data appropriately with each row containing each case and each column as variable.

(d) Data checking and cleaning is possible.

- All data should be numeric, although it may not be all variables it is not desirable to use letter or word (String variable) as data. This can be achieved by recoding the letter or word (string data) into desirable numeric and labeled appropriately.
- Data exploration can be done to check for errors and other accuracy.
- The statistical level of significance for rejecting null hypothesis (Ho) is when your p-value significance is less than 0.05.

The diagram below shows the frame work of this research study

WHY STATISTICAL SOFTWARE'S?



Researcher discuss here few of them .

1.Microsoft Excel:

History

This is part of the Microsoft Office suite of programs. Excel version 1.0 was first released in 1985, with the latest version Excel 2016. Microsoft Excel 2010 is one of the most popular software applications worldwide and is part of the Microsoft Office 2010 productivity suite. You can use Excel to analyze data, for example, in accounts, budgets, billing and many other areas. Excel allows you to explore the menu bar and the different tasks that can be done with it. You can work on sample spreadsheets doing basic math, adding and deleting columns and rows, and preparing the worksheet for printing. You can run your data visually to show trends, patterns and comparisons between the data in a chart, table or other template, and Excel performs most general statistical analyses but weak in regression, logistic regression, survival analysis, analysis of variance, factor analysis, multivariate analysis.

Advantages:

- Extremely easy to use and interchanges nicely with other Microsoft products
- Excel spreadsheets can be read by many other statistical packages
- Add on module which is part of Excel for undertaking basic statistical analyses
- Can produce very nice graphs

Disadvantages:

- Excel is designed for financial calculations, although it is possible to use it for many other things
- Cannot undertake more sophisticated statistical analyses without purchase of expensive commercial add ons.

Availability

Most computers come with Microsoft software already installed. For blue-plated (UniSA) computers, contact the IT Help Desk to install the latest Microsoft office software. For your own computer, you can always purchase Microsoft Office from a retail store.

2. Statistical Package for the Social Sciences (SPSS):

SPSS

SPSS stands for Statistical Package for the Social Sciences. It was one of the earliest statistical packages with Version 1 being released in 1968, well before the advent of desktop computers. It is now on Version 23. SPSS- (Statistical Package for the Social Sciences now Statistical products and Solution services) is most widely used in social science disciplines and courses. SPSS is the oldest software programs developed and made available in 1960s and has been redeveloped over the years, the latest version is SPSS 20.0 which was produced in August 2013. Many sociologists, psychologists and social workers use this program to enter their research data and formulate results. Although social science uses SPSS more widely than other fields, many find it easy to navigate with SPSS because it is a package that many beginners enjoy due to its very easy to use nature. SPSS has a "point and click" interface that allows you to use pull down menus to select commands that you wish to perform.

SPSS assists the user in describing data, testing hypotheses and looking for a correlation or relationship between one or more variables. SPSS is very suitable for most regression analysis and different kinds of ANOVA (regression, logistic regression, survival analysis, analysis of variance, factor analysis, multivariate analysis but not suitable for time series analysis and multilevel regression analysis)-Wikipedia (2014). Many students, both undergraduate and graduate, are taught SPSS during research analysis classes in demography, psychology, sociology and other social sciences.

Advantages :

- Very easy to learn and use
- Can use either with menus or syntax files
- Quite good graphics
- Excels at descriptive statistics, basic regression analysis, analysis of variance, and some newer techniques such as Classification and Regression Trees (CART)
- Has its own structural equation modelling software AMOS, that dovetails with SPSS

Disadvantage :

- Focus is on statistical methods mainly used in the social sciences, market research and psychology
- Has advanced regression modelling procedures such as LMM and GEE, but they are awful to use with very obscure syntax
- Has few of the more powerful techniques required in epidemiological analysis, such as competing risk analysis or standardised rates

Availability

SPSS is available on blue-plated (UniSA) computers. If it is not on the one that you use, then contact the IT Help Desk to install it. Staff are allowed to use SPSS at home for a cost of $10. Unfortunately, students have no home use rights, but can purchase a pretty much full version called a Premium Grad-pack with a 2-year license for approximately $250 from Hearne software

3. Statistical Analysis System (SAS):

SAS - SAS stands for Statistical Analysis System. It was developed at the North Carolina State University in 1966, so is contemporary with SPSS & which has its latest version produced in December 2011 is a package that many "power users" like because of its power and programmability. SAS is one of the packages that are difficult to learn. To use SAS, you must write SAS programs that manipulate your data and perform your data analyses. If you make a mistake in a SAS program, it can be hard to see where the errors occurred or how to correct it. However, it can take a long time to learn and understand data management in SAS than many other packages like SPSS or STATA with simpler commands line. However, SAS can work with many data files at once SAS can handle enormous data files up to 32,768 variables and the number of records is generally limited to the size of your hard disk.

SAS performs most general statistical analyses (regression, logistic regression, survival analysis, analysis of variance, factor analysis, multivariate analysis). The greatest strengths of SAS are probably in its ANOVA, mixed model analysis and multivariate analysis, while it is probably weakest in ordinal and multinomial logistic regression (because these commands are especially difficult), and robust methods (it is difficult to perform robust regression, or

other kinds of robust methods)- ATS Ucla Edu(2014). While there are some supports for the analysis of survey data, they are quite limited as compared to Stata.

Advantage

- Can use either with menus or syntax files
- Much more powerful than SPSS
- Commonly used for data management in clinical trials

Disadvantage

- Harder to learn and use than SPSS

Availability

Health Sciences has a Division licence for SAS 9.4M3 which is available for the Division's staff and students. To organise installation contact the IT Help Desk. SAS also has a free version SAS University, details are available here: http://www.sas.com/en_us/software/university-edition.html

4.R & MATLAB

R & MATLAB – Stanford (2014**)** identified R and MATLAB as the richest statistical systems by far. They contain an impressive amount of libraries, which is growing each day. Even if a much desired specific model is not part of the standard functionality, you can implement it yourself, because R and Matlab are really programming languages with relatively simple syntaxes. As "languages" they allow you to express any idea. The question is whether you are a good writer or not. In terms of modern applied statistics tools, R libraries are somewhat richer than those of Matlab. Also R is free software. On the flip side, Matlab has much better graphics, which you will not be ashamed to put in a paper or a presentation. MATLAB and R perform most general statistical analyses (regression, logistic regression, survival analysis, analysis of variance, factor analysis, multivariate analysis). The greatest strengths of both are probably in its ANOVA, mixed model analysis and users creative freedom in analysis.

R

S-plus is a statistical programming language developed in Seattle in 1988. R is a free version of S-plus developed in 1996. Since then the original team has expanded to include dozens of individuals from all over the globe. Because it is a programming language and

environment, it is used by giving the software a series of commands, often saved in text documents called syntax files or scripts, rather than having a menu-based system. Because of this, it is probably best used by people already reasonably expert at statistical analaysis, or who have an affinity for computers.

Advantage:

- Very powerful – easily matches or even surpasses many of the models found in SAS or Statas
- Researchers around the world write their own procedures in R, which are then available to all users
- Free!

Disadvantage :

- Much harder to learn and use than SAS or Stata

Availability

R can be downloaded from here:

http://cran.csiro.au/

5 PSPP

PSPP - This software provides a basic set of capabilities: frequencies, cross-tabs comparison of means (t-test and one way ANOVA); linear regression, logistic regression, reliability (Cronbach's Alpha, not failure or Weibull), and re-ordering data, non-parametric tests, factor analysis, cluster analysis, principal components analysis, chi-square analysis and more.

A range of statistical graphs can be produced, such as histograms, pie chart, Scree plots and np-chart.

PSPP can import Gnumeric and Open Document spreadsheets, postgres databases, comma-separated values and ASCII files. It can export files in the SPSS 'portable' and 'system' file formats and to ASCII files. The PSPP project (originally called "Fiasco") was born at the end of the 1990s as a free software replacement for SPSS, which was a data management and analysis tool, at the time produced by SPSS Inc. The nature of SPSS's proprietary licensing and the presence of digital restrictions management

motivated the author to write an alternative which later became functionally identical, but with permission for everyone to copy, modify and share. The latest version was released in April, 2014.

OTHER RELATED INFORMATION & REFERENCES :

There is lots of support available to make you more comfortable with undertaking statistical analyses, including this online course, biostatistician consultants, websites, YouTube tutorials, and even MOOC courses.

- If you would like face-to-face assistance, then information about biostatistical support can be found here:

  http://www.unisa.edu.au/Health-Sciences/Research/Biostatistical-and-epidemiological-support/

- In addition, there are a multitude of statistical software packages available that can do a lot of the work for you – and these are the focus of this current module. However, before we start looking at these, a question that often arises is "How do I get my data into a statistical package?". The good news is that most statistical software can read data directly from an Excel spreadsheet, so using Excel is often the easiest solution. Secondly, you can always enter data directly into a statistical package, since they nearly all have some form of inbuilt spreadsheet.

- Another solution is to use software like SurveyMonkey (https://www.surveymonkey.com/) to collect the data. SurveyMonkey has the facility to convert the data into an Excel spreadsheet or SPSS format. A final solution is to use specialised data entry software. This has the advantage of being able to put things like range checks on data entry fields, so for example, if a data entry field should only have a 0 or 1 entered, if you try and put anything else, it won't let you. A really good and free data entry program is EpiData Entry provided by CDC Atlanta. It is available from here: http://www.epidata.dk/download.php

- There are many commercial statistical packages available, some of which UniSA has licenses for. In addition, there are several free statistical packages available from the internet. For example, PSPP is a clone of SPSS, and can be downloaded here:

https://www.gnu.org/software/pspp/get.html.

- There are also many websites where you can undertake online statistical analyses. A good starting place is:

  http://statpages.info/

- There are also many specialised software programs for things like graphs, sample size calculations, and genetic analyses. Again, some are commercial, but others can be freely downloaded. A good example is the sample size software G*Power, which can be downloaded here: http://www.gpower.hhu.de/en.html

- If you need to calculate the sample size calculations then use the www.samplesurvey.com. Gpower Tools,Satasssist.com etc.

  In fact the diversity and number of software packages and available websites is so large, that reviewing all of them would be a full-time job!

  However, there are some software packages that are readily available and often used at UniSA, including Microsoft Excel, SPSS, SAS, STATA and R, which will briefly overviewed here. Then further details are provided in subsequent modules about each of these packages.