

The Role of Statistical Software in Data Analysis

Authors:

¹Abatan S. Matthew*, ²Olayemi Micheal Sunday

Address For correspondence:

¹Department of Demography and Social Statistics, Federal University Oye-Ekiti, Ekiti State

²Department of Mathematics and Statistics, Kogi State Polytechnic, Lokoja Kogi State

Abstract:

As quantitative research grows, application of statistical software (SS) becomes a more crucial part of data analysis. Researchers are experiencing a transition from manual analysis with paper to more efficient digital/electronic analysis with statistical software (SS). It identifies the prerequisites of producing world-class studies by using modern SS solutions. SS has contributed immensely in improving not only in demography studies and social investigation but also among other professionals' researches in Nigeria and the world at large.

A cross-sectional survey of 5 lecturers each were selected from 8 departments in the two faculties in Oye-Ekiti campus and were given questionnaire base on their availability and interest. A total sample size of forty (40) academic staff were selected but thirty (30) eventually responded which comprises fifteen less experienced staff with (0-5yrs) and 15 more experienced staff with (6yrs and above) as well as 10 non-response. The small sample size was due to few lecturers in some departments as at the time of this study. Data were analyzed using SPSS package.

A univariate and bivariate analysis was done and findings of the study revealed that impact of statistical software on research results give Mean (M) =4.80 and Standard Deviation (SD)=0.41, on a (1-5) Likert scales with 80% Strongly Agree that SS has positive impact on their research result. Respondents category and running analysis without SS shows mean (M) =2.27 and Standard deviation (SD)=1.37 on a (1-5) Likert scales with 66% admitting that they cannot run analysis without SS. Some SS are suitable for some kind of analysis than others for instance, while SPSS, STATA, SAS, MATLAB and R are 100% suitable for ANOVA, Eview, SAS, STATA, R and MATLAB are 100% suitable for time series analysis. Furthermore, STATA, SAS, Eview, MATLAB and R are 100% suitable for different kind of regression analysis among others. In FUOYE, while SPSS has 92.9% knowledge and usage, others have usage and knowledge as follows; STATA (57.1%), SAS (15.4%), MiniTab (0%), Ms-Excel (76.9%), MATLAB (28.6%), R (0%), Epi-info (16.7%), and PSPP (8.3%). The paper concludes

sunday.abatan@fuoye.edu.ng *Corresponding Author E-Mail Id

among other recommendations that academic staff/researchers should improve in their SS workshops training and further recommends integration of SS application in academic curriculum just like other compulsory courses. More so choosing statistical software packages to learn should be based on suitability of the software for all possible analysis you may wish to be analyzing

Keywords: Statistical Software Usage, Academic Research and Data Analysis.

I. INTRODUCTION

The emergence of statistical software has undoubtedly contributed enormously to the development in research studies in this 21st century. The high premium placed on ICT by human beings, researchers and organizations has undoubtedly made it a major drive of every nation (Eshasrenan, 2006: Akindutire 2013). Statistical Software (SS) is a vital tool for research analysis, data validation and findings.

‘Over the course of history, different forms of data analysis methods have been in existence. Initially, it was paper and pen and later the advent of which computer has helped invention of punching machines and later upgraded to simple calculator and complex scientific calculator. Nevertheless, pundits have revealed that statistical software is a software program that makes the calculation and presentation of statistics relatively easy (McDaniel, 2010). Statistical software allows researchers to avoid routine mathematical mistakes and produce accurate figures in their research if they input all data correctly. Research has been identified as one of the important duties of academic staff (NUC, condition of service: FUOYE, 2013) and the common saying *No publish! You perish*. This has led many lecturers to look for different means of ensuring their work is given excellent analysis with Statistical Packages. Development of statistical software allows academic researchers to conduct more quantitative studies easily (Chris McGann, 2009). Many researchers, professionals, scientists and business managers also can clearly present accurately prediction of the future using statistical software. Many proprietary and freeware statistical software packages are available that are suitable for different statistical analysis, depending on the user's needs: Some of the proprietary Software are SPSS, STATA, EVIEW, MINITAB, BMDP etc., and few among open or free Statistical software are R, EPI-INFO, CS-PRO, to mention but few.

The emergence of statistical software in the twenty-first century has helped different researchers in the physical and social science to improve in the quality of research. Most renowned researchers in adopting this software in their data analysis have been able to identify the immense contribution to research findings (Adetola, 2013). Any quantitative research cannot be done effectively without SS. Moreover, it enables pundits faction research data for easy presentation. It helps professionals to interact with data thereby paving way for creativity and innovation. Some are user friendly interface with drop-down tips for beginners (ATS, Ucla Edu, 2014). Advances in technology have improved all our lives (Akindutire, 2013) and has allowed experts greater freedom to come out with results within a twinkle of an eye than ever before where it takes time to finish analysis. This same technology has offer pundits tremendous opportunity to research and keep research as a more interesting

field of study. This study is significant, because it is an attempt to measure the direct impact of different statistical software on research analysis.

II. 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, SAS, MATLAB, MINITAB, STATA, Mathematica and lots more have been utilized by people across all disciplines for many years and are quite user friendly.

Statistical Package for the Social Sciences (SPSS)

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. Odusina (2011) disclosed that working with SPSS demand some background knowledge of statistics. There are slight variations in the difference version of SPSS e.g. version 10, 11, 12, 13, 14, 15, 16, etc.

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.

Statistical Analysis System (SAS)

SAS - 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.

Econometric Views (EViews)

EViews is a statistical package for window, used mainly for time-series oriented econometrics analysis. It was developed by Quantitative Micro Software (QMS) and now a part of IHS. Version 1.0 was released in March 1994, and has replaced MicroTSP. The current version of EViews is 8.0, released in March 2013. EViews can be used for general statistical analysis and econometric analyses, such as cross-section and panel data analysis and time series estimation and forecasting. EViews relies heavily on a proprietary and undocumented file format for data storage. However, for input and output it supports numerous formats, including databank format, Ms- Excel format, SPSS/PSPP, DAP/SAS, STATA, RATS, and TSP. EViews can access ODBC databases.

MINITAB

MINITAB is statistical software used by educators, students, scientists, business associates and researchers to provide statistical software in a multitude of areas. MINITAB, developed around 1990, and remains one of the oldest statistical software programs available. MINITAB has compatibility with PC, Macintosh, Linux and all other major platforms.

As one of the easiest statistical software programs to use, MINITAB remains a popular choice with those new statistical software. With drop-down menus and dialog boxes describing how and what to do next, MINITAB persists as a popular choice for teaching students about statistics and data analysis. MINITAB primarily has a user base of educators using the program to show students research methods and analysis in college and graduate-level courses. MINITAB performs most general statistical analyses (regression, logistic regression, survival analysis, analysis of variance, factor analysis, but has its weaknesses in general linear model (GLM) and Multilevel regression).

STATA

STATA is a powerful statistical package with smart data-management facilities, a wide array of up-to-date statistical techniques, and an excellent system for producing publication-quality graphs. Stata latest version was produced June 24, 2013 which is a fast and easy to use data management package. Stata is available for Windows, Unix, and Mac computers. The standard version is called Stata/IC (or Intercooled Stata) and can handle up to 2,047 variables. There is a special edition called Stata/SE that can handle up to 32,766 variables (and also allows longer string variables and larger matrices), and a version for multicore/multiprocessor computers called Stata/MP, which has the same limits but is substantially faster. STATA performs most general statistical analyses (regression, logistic regression, survival analysis, analysis of variance, factor analysis, multivariate analysis and time series analysis).

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.

MS-EXCEL

MS-EXCEL -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.

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.

EPI-INFO

Epi Info is public domain statistical software for epidemiology developed by Centers for Disease Control and

Prevention (CDC) in Atlanta, Georgia (USA). Epi Info has been in existence for over 20 years and is currently available for Microsoft windows. The program allows for electronic survey creation, data entry, and analysis. Within the analysis module, analytic routines include t-tests, ANOVA, nonparametric statistics, cross tabulations and stratification with estimates of odds ratios, risk ratios, and risk differences, logistic regression (conditional and unconditional), survival analysis (Kaplan Meier and Cox proportional hazard), and analysis of complex survey data. The software is in the public domain, free, and can be downloaded from <http://www.cdc.gov/epiinfo>.

The objective of this study is to examine the perception of researchers in working with statistical software and their attitude towards using SS in analysis, based on their level of expertise in statistical software usage in quantitative research, and the impact factors on the speed of their results. The level of efficiency in using SPSS, Eview, SAS, STATA, MINITAB, MATLAB, MS-Excel, R, Epi-info and PSPP was measured.

Table I: Showing 10 Statistical Packages and means of Accessing

Statistical Packages	Latest Version	Open Source Software	Proprietary/ Licensed Software	Available in Linux/ Support OS	Available and Support Window OS
SPSS	August 2013		YES	Yes	Yes
Eview	February 2013		YES		Yes
SAS	December 2011		YES	Yes	Yes
STATA	June 24, 2013.		YES	Yes	Yes
Epi-Info	January 26, 2011	YES			Yes
MATLAB	New releases twice per year		YES	Yes	Yes
MINITAB	February 18, 2014		YES		Yes
R	April 3, 2013	YES		Yes	Yes
MS-EXCEL	January, 2010		YES	Yes	Yes
PSPP	January 13, 2014	YES		Yes	Yes

Source: Wikipedia Articles, the Online Free Encyclopedia visited on 28/04/2014 Edited by Researcher

Statistical Software packages are in different kinds: free and open-software and proprietary software. Chris Bolu (2013) Free and open-source operating systems such as Linux and OpenBSD are widely utilized today, powering

millions of servers, desktops, smart phones (e.g. Google Android), and other devices. Free software and open-source licenses are used by many software packages and their adoption have grown in leaps and bounds during the last two decades, because they are accessed freely. Proprietary software on the other hand has its source code in control and secured for profitability. All cannot be accessed unless it is been paid for. This is probably done to ensure quality and avoid pirate.

III. 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 (H_0) is when your p-value significance is less than 0.05.

IV. METHODOLOGY

In this study the staffs in Federal University Oye-Ekiti (FUOYE) were selected from individual departments in the two faculties in Oye Campus excluding Ikole Campus because of time and cost. A cross-sectional survey of 5 lecturers each were selected from 8 departments in the two faculties in were given questionnaire base on their availability and interest. A total sample size of thirty (30) academic staff responded to the questionnaire out of 40 sent out, which comprises fifteen less experienced staff with (0-5yrs) and 15 more experienced staff with (6yrs and above). The small sample size was due to few lecturers in some departments as at the time of this study. They were questioned about their knowledge of different statistical software, ever used and reason for usage of SPSS, Eview, STATA, MINITAB, MATLAB, MS-EXCEL, R-PACKAGE, EPI-INFO, PSPP and SAS. The academic staffs were asked on their self confidence and cognitive skills on how it enhanced their understanding of statistical concepts in research analyses and on what they perceived as the added value of statistical software as a tool for "doing statistics." The impact of statistical software on their research studies was measured by judgments made on a 5-point scale, where 1 = strongly disagree, 2=disagree, 3=no opinion, 4=agree and 5 = strongly agree. Staffs rated their perception of the positive impact of working with statistical software on their attitude towards statistics, on their self-confidence in using statistical software on their interest in quantitative

scientific research, and on their skills needed to analyze the research question, to select a procedure, and to guide and control their thinking process. The level of efficiency in using SPSS, Eview, SAS, STATA, MINITAB, MATLAB, MS-Excel, R, Epi-info and PSpP were measured on 5 level of expertise; 1= Novice, 2=Amateur, 3=Average, 4=Intermediate and 5=Expert.

40 questionnaires were distributed by the researcher but 30 questionnaires were returned with valid response. Twenty (20) questionnaires each was sent to four (4) departments in the faculty of social sciences and 4 departments in faculty of sciences.

V. CONCEPTUAL FRAME WORK

The diagram below shows the frame work of this research study.

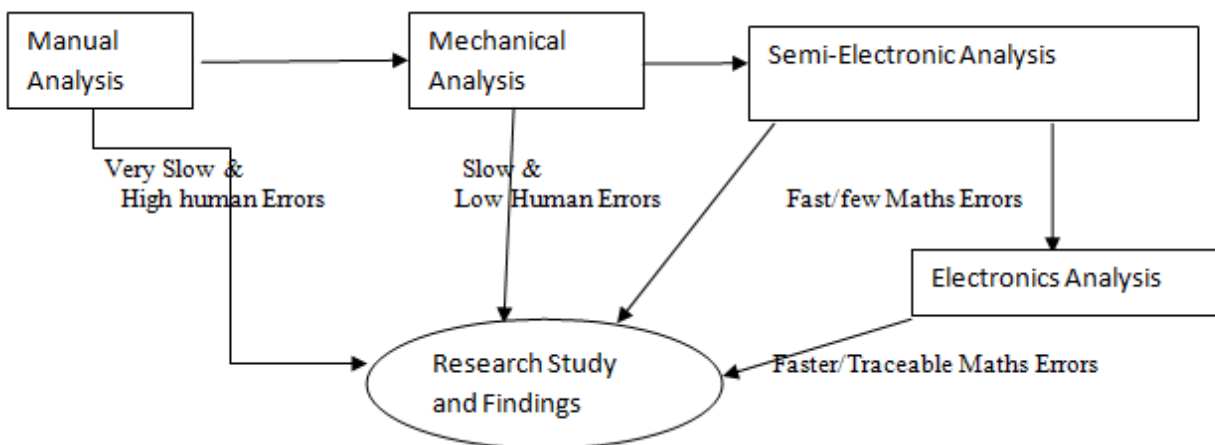


Figure 1

VI. TEN (10) STATISTICAL PACKAGES AND THEIR LEVEL OF EFFECTIVENESS IN DIFFERENT KINDS OF STATISTICAL ANALYSIS

1. ANOVA: Wikipedia Articles, on comparison of different statistical software visited on 28th April, 2014 disclosed that considering the effectiveness statistical package on different kinds of Analysis of Variance e.g One-way ANOVA, 2-way ANOVA, MANOVA, GLM, Mixed Model, Post-Hoc analysis on ANOVA, Latin Square; SPSS, SAS and STATA as well as MATLAB have 100% efficiency each under proprietary statistical software while R which has 100% effectiveness is the only open source statistical software suitable to perform all analysis of variation (ANOVA) you may think of. Although Epi-info with (28.6%) can do one way and two ways ANOVA but it cannot do multiple analysis of variance, General linear model, mixed model, post-hoc analysis

and the likes. Ms-Excel having (57.1%) can do some of the analysis of variance but cannot perform post-Hoc analysis which thereby making it less effective. Eview is the packages that are not really suitable for analysis of variance because it can only perform one way ANOVA.

2. Regression Analysis such as Ordinary Least Square method, Logistics Regression, General Linear model, Multi-level, step-wise regression model: SPSS has 80% efficiency in these areas of regression analysis because it can hardly run multi level regression analysis. Eview, SAS, STATA and MATLAB have 100% efficiency each in all these areas of regression analysis. MINITAB has (60%) efficiency but cannot run GLM and MLR. Among the free statistical software R has (100%) efficiency and can run any regression analysis, Epi-info with (40%) can run only two regression analyses which are Ordinary Least Square (OLS) and Logistics Regression.

3. Time Series Analysis such as ARMA, GARCH, Unit Root test, Test of Cointegration, VAR, Multivariate GARCH: Eview, SAS and STATA have 100% efficiency each are in time series analysis while others like SPSS and Ms-Excel can perform on autoregressive moving average (ARIMA). MATLAB with (83.3%) is efficient too but cannot run multivariate GARCH.

Among the free statistical software R with (100%) can run any time series analysis, while Epi-info and PSPP with (0%) cannot run any time series analysis.

VII. DATA ANALYSIS

Table II: Demographic Analysis of respondents under review

Sex	Frequency	Percentage
Male	23	76.7
Female	7	23.3
Total	30	100.0

Note: 25% of the total questionnaires are non response

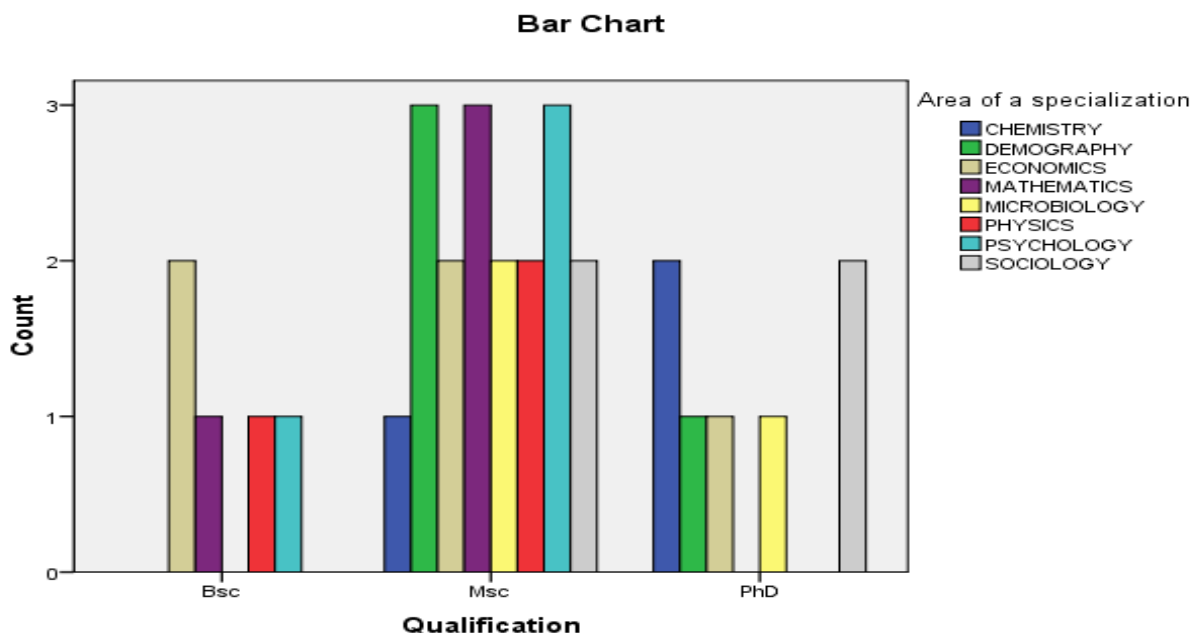
Source: Research Questionnaire results (Abatan, 2014)

The percentage of male that is higher than that of female show that actually women are few in academic/lecturing profession in FUYOE. Therefore the survey ensures that female are duly represented having the one-third of the total sample size for this study.

The respondents for this study are Chemistry 10%, Demography and Social Statistics 13.3%, Economics and Development Studies 16.7%, Mathematics 13.3%, Microbiology 10.0% Physics 10.0%, Psychology 13.3%, Sociology 13.3%. The Economics department that has higher percentage was due full response to the questionnaire sent to them; others with less percentage responded less.

The qualifications of respondents for this study are as follows: BSc. 23.3%, MSc 60% and PhD 16.7%. No professors in the respondents because they were not available.

Chart I: Showing Qualification and Area of a specialization of Respondents



Source: Research Questionnaire results (Abatan, 2014)

The knowledge and Usage of Statistical Packages

The knowledge of SPSS is 100% in Federal University Oye-Ekiti but the usage is 92.9% among academic staff just few percentage i.e 7.1% of staff who know but do not use SPSS.

The knowledge of STATA is 71.4 but the usage of Stata is very high at 80.0%.

The knowledge of SAS is weak (46%) and usage is very poor among those who know SAS. The overall usage of SAS is 15.4% as compare with 84.6% non-usage.

The knowledge of MINITAB is very poor in FUYOYE with 15.4% and 84.6% do not know the package. Among those that have the knowledge of it none of them has ever used it i.e 100% non usage. The overall knowledge of Ms-Excel as statistical software is 92.3% while overall usage is 76.9% The overall knowledge of MATLAB is 57.1% and the usage is 28.6% among FUYOYE staff.

The knowledge of R as a statistical package is very poor and usage is zero percent in FUYOYE with 63.6% no knowledge of R and 100% do not use R. Although there is 50% knowledge of EPi-Info the usage of Epi-Info is quite poor having 16.7% of ever used. The knowledge and usage of PSpP among academic staff in FUYOYE is very poor having 91.7% no knowledge and just 8.3% ever used.

Table III: Table Showing Descriptive analysis of Impact of Statistical Software on Research Analysis

SN	Impact Variables	Frequency	Strongly Disagr	Disagree	No Opinid	Agree	Strongly Agree	Mean	Standard Deviation
1	Research Results	Frequency	0	0	0	6	24	4.80	0.41
		Percentage	0%	0%	0%	20%	80%		
2	Manual Analysis	Frequency	12	8	2	6	2	2.27	1.36
		Percentage	40%	26.7%	6.7%	20%	6.7%		
3	Analysis with Statistical Software	Frequency	18	10	0	2	0	1.53	0.82
		Percentage	60%	33.3%	0%	6.7%	0%		
4	Training workshop	Frequency	0	0	0	20	10	4.33	0.48
		Percentage	0%	0%	0%	66.7%	33.3%		
5	Research Speed Presentation	Frequency	0	0	0	10	20	4.67	0.48
		Percentage	0%	0%	0%	33.3%	66.7%		

Source: Research Questionnaire results (Abatan, 2014)

The results of the questionnaire administered indicate that Statistical software has effect on research results which has Mean (M) as 4.8 and standard deviation (SD) = 0.41; (on a scale of 1-5) with the staff in Federal university Oye Ekiti actually recognizing the impact of statistical software on their research result.

The total sum of (40% Strongly Disagree and 26% Disagree) shows and overall report that 66% of staff admitted that they cannot run all their analysis without statistical software. This is more explained as the (Mean=2.27 Standard deviation (SD) = 1.37; on a scale of 1-5)

The total sum of (60% Agree and 33.3% strongly Agree) shows an overall report that 93% of staff agreed that they can improve their research expertise by attending training and workshop on statistical packages usage. This is more explained as the (Mean=4.33 standard deviation (SD) = 0.48; on a scale of 1-5)

The total sum of (33.3% Agree and 66.7% strongly Agree) shows an overall report that 100% of staff agreed that statistical software makes their research work look robust and faster. This is more explained as the (Mean=4.67 standard deviation (SD) = 0.48; on a scale of 1-5)

Correlation Analyses on years of Experience and Efficiency in Statistical Software usage

The year of experience and SPSS efficiency has positive weak relationship although not significant with correlation (R) =0.065 and p-value=0.77. The year of experience and Eview efficiency has weak relationship although not significant with correlation (R) =0.183 and p-value=0.351.

The year of experience and STATA efficiency have weak relationship although not significant with correlation (R) =0.086 and p-value=0.675. The year of experience and SAS efficiency has average relationship and very significant with correlation (R) =0.448 and p-value=0.017. The year of experience and Ms-Excel efficiency have an average relationship and very significant with correlation (R)=0.523 and p-value=0.004. Minitab, Matlab, R, Epi-info and PSPP has no significant relationship with years of experience of staff in FUOYE with correlation (R) =0

Correlation analyses on area of specialization (Field of studies) and statistical software usage

Area of Specialization and SPSS usage have a positive weak relationship among staff in FUOYE with correlation (R) = 0.29 and p-value = 0.12; although the relationship is insignificant but at least SPSS has some level of usage in the different field of studies in FUOYE.

Area of Specialization and Eview usage have a negative weak relationship among staff in FUOYE with correlation (R) = -0.26 and p-value = 0.20; although the relationship is insignificant but at least Eview have some level of usage in the different field of studies in FUOYE.

Area of Specialization and STATA usage has a positive weak relationship among staff in FUOYE with correlation (R) = 0.168 and p-value = 0.39; although the relationship is insignificant but at least STATA has some level of usage in the different field of studies in FUOYE.

Area of Specialization and SAS usage has a negative weak relationship among staff in FUOYE with correlation (R) = -0.208 and p-value = 0.307; although the relationship is insignificant but at least SAS has some level of usage in the different field of studies in FUOYE.

Area of Specialization and MS-Excel usage has a average relationship among staff in FUOYE with correlation (R) = 0.417 and p-value = 0.03; the relationship has a significant level of usage in the different field of studies in FUOYE.

Minitab has no relationship because there is no usage of this packages in FUOYE staff this is similar to Matlab, R, Epi-info and PSPP has no significant relationship with years of experience of staff in FUOYE with correlation (R) = 0.

The surprising situation is that most of these packages are open source and free download.

Reasons for Using Statistical Software

The result shows that 95.0% of those that have ever used any statistical software admitted that it is very easy to use. With mean (0.952) and Standard deviation (SD) = 0.218 (under a likert scale of 0-1). Further more the result shows that 77.8% of those that has ever used statistical software admitted that some stat software are cheap. With mean (0.700) and Standard deviation (SD) = 0.483

The result shows that 100.0% of those that have ever used statistical software admitted that they use statistical software because it is suitable for their analysis with mean (0.966) and Standard deviation (SD) = 0.185

Reason for not using Some Statistical software

The result shows that 65.0% of those that have ever used statistical software disclosed that statistical software is expensive and 35.0% concealed that it was due to expensive nature that make them not to use some software for analysis with mean (0.38) and Standard deviation (SD) = 0.498

The result shows that 75.0% of those that has ever used statistical software disclosed that statistical software difficult to use and 25.0% concealed that is not difficult issues that make them not to use some software for analysis. The mean (M) = (0.231) and Standard deviation (SD) = 0.438. The result shows that 27.3% of those that has ever used statistical software disclosed that they have not given time to statistical software training and 72.7% concealed that they have not given enough time learning it The mean (M) = (0.33) and Standard deviation (SD) = 0.48.

The result shows that 57.1% of those that has ever used statistical software disclosed that the reason for not using some statistical software was that they are not suitability for their analysis. While 42.9% declared that it was not suitability. The mean (M) = (0.55) and Standard deviation (SD) = 0.506.

VIII. FINDINGS AND DISCUSSIONS

Statistical software packages have been recognized to have contributed immensely to research analysis by helping to minimizing human and experimental errors in data analysis.

1. It has been discovered that some analysis such as post Hoc, complex analysis in time series, regression and variance analysis cannot be calculated manually effectively without statistical software packages.
2. Statistical software packages have been discovered to help academic staffs in higher institution to improve their research expertise by attending training on usage of statistical packages.
3. Statistical packages make research work robust and faster.
4. It was discovered that 81% efficiency of staff in statistical software is determine by the years of experience in usage and the area of specialization.
5. Most reason for using statistical software was it easy usage, suitability for many statistical analysis and the likes.
6. While reason for non usage range from lack of attention to learn, difficult usage, cost of licensing, etc. It was discover that statistical software are not expensive neither are they too difficult to use but people do not give attention to its learning.

IX. CONCLUSION

Conclusively, it is the opinion of the researcher, after analyzing the field data to a very large extent, that statistical software has contributed immensely to social research especially in the area of demographic and data analysis. This was achieved by using a scientific approach to solve fundamental problems in research which is data analysis. Although some other factors contributed to the quality of research work such as literature review, methodology and findings, it is quite clear that the impact of statistical software packages on analysis and findings of research cannot be over estimated. This research is not conclusive, similar research can be carried out in other institutions especially older universities in Nigeria due to its limitation of few samples in a new university and other areas can be explored for further studies on their impact on research work.

X. RECOMMENDATIONS

Based on afore mention findings this study recommends as follow:

Academic staff should learn to improve their efficiency on the usage of statistical software packages especially the one that is suitable for all analysis in their areas of studies.

1. Researchers should be the research analyst for their research findings and seize from giving their research data out for another to analyze.
2. Choosing statistical software packages to learn should be based on suitability of the software for all possible analysis you may wish to be analyzing. It was therefore recommended that at least two (2) should be chosen so that learning research analysis would be broaden and robust.
3. Nigeria institutions should encourage separate curriculum on the usage and application of statistical packages in research study just like other compulsory courses where students are trained with practical application of statistical packages in analyzing data because research project/study at the end of a discipline should be based on pure or applied field work that solve existing problem in our society.

REFERENCES:

1. Adetola O.G (2013) Learning Statistical Package Workbook. Nigeria(Unpublised).
2. Akindutire, A.F. (2013) Usage of Mobile Phone: Deception, Deviance and Fraud (A case study of Ambrose-Alli)
3. Chris McGann, Role of Statistical Software, eHow Contributor http://www.ehow.com/facts_6923793_role-statistical-software.html#ixzz2xoZFcTZX visited on 3rd April, 2014.
4. Daniel, S. (2009). *The Most Popular Statistical Software*. accessed on 3rd April, 2014: eHow Contribution Eview Users Manual (2009) History and usage of Eview visited on 28th April, 2014
5. Importance of Statistical Analysis by Data Collection <http://highqualityarticles.com/> online article

6. Karen (2013) Choosing a Statistical Software Package or Two
Source:<http://www.theanalysisfactor.com/choosing-statistical-software/> VISITED ON 7TH April, 2014
7. McGann, C. (2009). Role of Statistical Software . eHow Contribution accessed by 3rd April, 2014
8. McCullough, B.D. (1999). "Econometric software reliability: EViews, LIMDEP, SHAZAM and TSP". *Journal of Applied Econometrics* 14 (2):191–202. HYPERLINK
9. "http://en.wikipedia.org/wiki/Digital_object_identifier" \o "Digital object identifier" doi : Odusina E.K. (2011) Computer Application for Population analysis. JABU, Osun state.
10. Stanford P. (2013) Statistical & Financial Consulting http://stanfordphd.com/Statistical_Software.html visited 6th April, 2014
11. STATA PRESS site (2014) Stata history and usage. visited 28th April, 2014.
12. Zaslavsky Alan (2014) HYPERLINK "<http://www.hcp.med.harvard.edu/statistics/survey-soft/> Summary of Survey Analysis Software wikipedia library



***Authors Copy**