Application of Statistical Software in Quantitative Data Analysis

Olayemi Michael Sunday*, Olajide Oluwamayowa Opeyimika and Ajayi Omolola Felicia
Department of Mathematics/Statistics, Kogi State Polytechnic, Lokoja, Kogi State, Nigeria.
*Corresponding author: apostlemike2@yahoo.com

Abstract

TThe paper expounds on the application of Statistical Software (SS) in quantitative data analysis. Research questions were formulated and questionnaire were used for data collection. The collected data were analysed using percentage, mean and regression. From the analysis carried out, we discovered that women are fewer in the academic profession. We also find out that the most common statistical software is MsExcel, followed by Statistical Package for Social Sciences(SPSS) and the less popular statistical software is R. Also, the frequently used statistical software is SPSS, followed by MsExcel, while the less usage statistical software is R. In this paper work, it was shown that there is a positive impact of statistical software in obtaining a fast, accurate and efficient result in academic staff research work in Kogi State Polytechnic (KSP), Lokoja. The following statistical software - MsExcel, SPSS, Graphpad, Minitab and R, were highlighted and the regression analysis carried out for each statistical software were screenshot and presented. It was discovered that the five (5) different SS used gave the same result. The simple regression model obtained is Number of Death = 44+ 2.17Malaria. Recommendations were made accordingly.
Key words: Statistical Software (SS), Quantitative Data, Regression Model

INTRODUCTION

Quantitative data are any information about quantities; that is, information that can be measured and written down with numbers (continuous). Examples of quantitative data are height, shoe size, the length of your fingernails. While discrete are information that can be counted such as the number of academic staff in Kogi State, number of desk top computer in the statistical laboratory, number of students admitted into Kogi State Polytechnic, etc. Olayemi (2019). On daily basis, we encounter data that applicable in life but it remains meaningless except it is collected, presented, tabulated and analyzed.

Errors in human data entry can sabotage results and conclusions. A single data entry error can cause a moderate correlation to vanish and a significant t-test to become non-significant. As a result, scientists should create and use human-computer interface (Statistical Software) that reduce data entry errors. Scientific data analysis aids in the creation of new information or a deeper understanding of natural events. Data analysis is where statistical software comes in handy. SPSS and Minitab appear to be the most popular, especially among those who can't code or aren't mathematically oriented enough to use complex tools like R, MATLAB, or Maple Hilary *et. al.* (2021).

The aim of this research work is to compare statistical software in quantity data analysis. The objectives are to determine the level of awareness and usage of statistical software, evaluate the relevance of statistical software in data analysis and to derive a multiple regression model.

Over the years, statistical analysis has been carried out using manual analysis methods which is cumbersome and prone to a lot of human errors especially in a case where we have large information. This led to the emergence of Statistical Software (SS) which has undoubtedly contributed enormously to the development in research in this 21st century Kinmberly and Larry (2011).

There are different forms of data analysis methods. Initially, it was paper and pen, punching machines, simple calculator, and complex scientific calculator. By advancement in technology, we now have executable statistical program software. Statistical software allows researchers to avoid routine mathematical mistakes and produce accurate figures in their research if all data are inputted correctly Okaygbue *et. al* (2021).

Research has been identified as one of the important duties of Academic Staff (National Board of Technical Education condition of service, Kogi State Polytechnic Revised Edition 2017) and the popular saying that 'no publication no promotion' has made it mandatory for Academic Staff to continuously delve into academic research work. This has led many academic staff

to seek for different means of ensuring that their research work is given the best statistical analysis and interpretation NBTE (2017) edition.

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, GRAPHPAD, MINITAB, EXCEL, R etc. https://www.predictiveanalyticstoday.com.

According to Abatan and Olayemi (2014), the emergence of statistical software in the twenty- first century has helped different researchers in the physical and social science to improve 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). This study is significant, because it is an attempt to consider the application of SS on quantitative data analysis.

Hilary *et. al* (2021) carried out research on Trends and usage pattern of SPSS and Minitab software in scientific research. It was discovered from their findings that SPSS and Minitab seems to be more popular in term of usage as compare to the use of other statistical software like MATLAB, Maple, and R, particularly to the researchers who are not statistically inclined to handle advance statistical software or those who could not code or write statistical programming. It was also discovered that the usage of SPSS and Minitab is steadily increasing though the spreading of Minitab is not as rapid as that of SPSS. Finally, the researchers established that SPSS are commonly used in the fields of medicine, social science, and engineering while Minitab is used mostly in engineering, materials science, and Computer Science.

Selected Types of Statistical Software (SS)
The following are some selected SS programs, which are Microsoft Excel, SPSS, GRAPHPAD, MINITAB, and R.

Microsoft Excel
This is part of the Microsoft Office suite of programs. Excel version 1.0 was first released in 1985, with the latest version Excel 2018. It is extremely easy to use and interchanges nicely with other Microsoft products. Many other statistical packages can read

Excel spreadsheets. One of the disadvantages is that it was designed for financial calculations and cannot undertake more sophisticated statistical analyses without purchase of expensive commercial add ons.

Most computers come with Microsoft software already installed. For your own computer, you can always purchase Microsoft Excel from a retail store. To install, insert the purchased disk containing the software and follow the instruction or contact the IT Help Desk in your institution.

Statistical Package for the Social Sciences (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 24. It is very easy to learn and use. It can use either with menus or syntax files and has good graphics.

One of the disadvantages is that it focused on statistical methods mainly used in the social sciences, market research and psychology. Others are that it has advanced regression modeling procedures such as LMM and GEE, but they are awful to use with very obscure syntax.

For your own computer, you can always purchase SPSS from a retail store. To install, insert the purchased disk containing the software and follow the instruction or contact the IT Help Desk in your institution.

GraphPad
GraphPad InStat is a commercial scientific statistical software available for both Windows and Macintosh computers. InStat is designed for the novice in handling statistical packages. It helps you pick an appropriate test by asking questions about the data. It presents results in simple paragraphs, with a minimum of statistical jargon. InStat's help screens review the use of each test and explain every portion of the results, and provide unique analysis checklists to confirm that the data have not violated any assumptions of the test, and that the test answers the right question.

The arguments for GraphPad InStat is extremely easy to use software and limited to the essence of statistical analysis. It is quite suitable for beginners. For your own computer, you can always purchase GraphPad from a retail store. To install, insert the purchased disk containing the software and follow the instruction or contact the IT Help Desk in your institution.

Minitab

Minitab is a statistics package developed at the Pennsylvania State University by researchers Barbara F. Ryan, Thomas A Ryan, Jr., and Brian L. Joiner in 1972. It began as a light version of OMNITAB, a statistical analysis program by NIST. Minitab automates calculations and the creation of graphs, allowing the user to focus more on the analysis of data and the interpretation of results.

Minitab is a versatile statistics package that is cheaper and requires less disk space than its heavy-weight competitor like SPSS. However, this comes at a cost, and Minitab is out-performed by its rivals in a number of areas.

Although Minitab is generally considered easy to use, and operates through an interface that is intuitive to anyone familiar with other statistics packages, it does suffer from some drawbacks in this area. Like the SPSS data view, the worksheet window in Minitab uses a fixed structure that is more difficult to manipulate than in spreadsheet programs like Microsoft Excel. Also, Minitab has poor compatibility with other statistics programs, making file imports more difficult. For your own computer, you can always purchase Minitab from a retail store. For installation, insert the purchased disk containing the software and follow the instruction or contact the IT Help Desk in your institution.

R-Package

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 analysis, or who have an affinity for computers. R is a free software and can be downloaded from http://cran.csiro.au/

General Attribute of Statistical Software

SS has common characteristics that make it reliable and suitable for data analysis, some of these are: Data editor is in rows and columns which make it very easy to enter numeric data. There is availability of menu bar comprises drop-down menu, quick analysis as well as brief user manual. Statistical level of measurement is put into consideration in data entry.

The following are the initial steps in quantitative data analysis in any research project work.

- Getting your data ready to enter into the software
- Defining and labeling variable
- Entering data appropriately with each row containing each case and each column as variable
- 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 recording the letter or word (string data) into desirable numeric and labelled appropriately.
- Data exploration can be done to check for errors and other accuracy.
- The statistical level of significance for rejecting null hypothesis (H_o) is when your p-value significance is less than 0.05.

Materials and Methods

In this study, some staff of Kogi State Polytechnic were selected from five Schools in Lokoja Campus. This research work excludes Osara/Itakpe campus because of time and cost. A cross sectional survey of academic staff from various departments in the five schools was given questionnaire based on their availability and interest. A total sample size of 120 academic staff responded to the questionnaire out of 130 sent out.

Commonly used Statistical Tools
Descriptive Statistics

These are used to describe the basic features of the data in a study. They provide simple summaries about the sample and the measures.

There are four major types of descriptive statistics:

Measures of Frequency

- Count, Percent, Frequency
- Shows how often something occurs
- Use this when you want to show how often a response is given

Measures of Central Tendency

- Mean, Median, and Mode
- Locates the distribution by various points
- Use this when you want to show how an average or most commonly indicated response

Measures of Dispersion or Variation

- Range, Variance, Standard Deviation
- Identifies the spread of scores by stating intervals
- Range = High/Low points
- Variance or Standard Deviation = difference between observed score and mean
- Use this when you want to show how "spread out" the data are. It is helpful to know when your data are so spread out that it affects the mean

Measures of Position
- Percentile Ranks, Quartile Ranks
- Describes how scores fall in relation to one another. Relies on standardized scores
- Use this when you need to compare scores to a normalized score (e.g., a national norm)

Regression Analysis

Ben (2018) defined regression analysis as a reliable method of identifying which variables have impact on a topic of interest. The process of performing a regression allows you to confidently determine which factors matter most, which factors can be ignored, and how these factors influence each other.

In order to understand regression analysis fully, it's essential to comprehend the following terms:
- Dependent Variable: This is the main factor that you're trying to understand or predict.
- Independent Variables: These are the factors that you hypothesize have an impact on your dependent variable.

According to Olayemi and Alabi (2018), simple linear regression can be mathematically expressed as:

$$Y_t = \beta_0 + \beta_1 X_t + e_t$$

where β_0 is the intercept and β_1 is the slope of the line and e_t is the error term. Where we have more than one explanatory variable, we apply multiple linear regression. A multiple linear regression is carried out to predict the values of a dependent variable (y), given a set of explanatory variables (xi, i = 1, 2, 3, …, n). Mathematically, it can be represented in the form:

$$Y_t = \beta_0 + \beta_i x_i + e_t$$

where the are the parameters for i = 1, 2, 3, …, n and are the explanatory variables, and is the error term.

Results

Table I: Demographic Analysis of Respondents under review

Sex	Frequency	Percentage (%)
Male	78	65
Female	42	35
Total	**120**	**100**

Source: Researcher's field work, 2021

The percentage of male is higher than that of female. This shows that women are few in the academic profession in Kogi State Polytechnic, Lokoja Campus. Therefore, the survey ensures that female was duly represented having the one-third of the total sample size for this study.

Table II: Awareness and usage of Statistical Software

S/n	Software Package	Awareness	Percentage (%)	Usage	Percentage (%)
1	MsExcel	110	92	43	36
2	SPSS	96	80	67	56
3	Graphpad	25	21	08	07
4	Minitab	10	08	03	3
5	R	05	04	02	2

Source: Researcher's field work, 2021.

Discussion

From table II above, we discovered that the most common SS is MS Excel with 92% awareness from respondents, followed by SPSS with 80% awareness and the less popular SS is R with 4% respondents. Also, the frequently used SS is SPSS with 56% respondents, followed by MS Excel with 36% respondents, while the less usage SS is R with 2% respondents.

Table III: Analysis of respondent' on the relevance of SS in academic research work

S/N	Validating Statement	SA	A	D	SD	N	T	Mean	Decision
1	The use of SS give better research result	38	69	10	3	120	382	3.18	Accepted
2	The time of execution of manual analysis faster than SS analysis	1	14	41	64	120	192	1.6	Rejected
3	The result obtained from SS analysis is difficult to interpret for a non-statistician	43	57	15	5	120	378	3.15	Accepted
4	There is need for regular training workshop on the use of SS	56	50	11	3	120	399	3.33	Accepted
5	There is positive impact on the application of SS in contemporary research work	53	41	21	5	120	382	3.18	Accepted

Source: Researcher's field work, 2021.

The data obtained were analyzed using the mean to estimate the degree of agreement or disagreement in each of the questionnaire item. Four-point rating scale was adopted with responses and numerical values attached as Strongly Agreed (SA) = 4; Agreed (A) = 3; Disagreed (D) = 2; Strongly Disagreed (SD) =1. By adding an interval scale of 0.5 to the mean numerical values assigned to scaling items which was 2.5. Therefore, the cut-off point is 3.0.

Decision rule: Accept item with mean of 3.0 and above, otherwise Reject.

From the above table item 1, 3, 4, & 5 were accepted while item 2 is rejected. This shows that SS has a positive impact on contemporary academic research work.

Table IV: Number of deaths caused by Malaria

S/N	Year	No of death	Malaria
1	2002	66	24
2	2003	52	18
3	2004	60	15
4	2005	72	13
5	2006	66	10
6	2007	64	20
7	2008	125	30
8	2009	99	10
9	2010	94	15
10	2011	129	35
11	2012	111	27
12	2013	95	23
13	2014	96	21
14	2015	105	31
15	2016	135	40
16	2017	124	18
17	2018	95	15
18	2019	69	20
19	2020	54	10
20	2021	60	12

Source: Records department, Kabba Zonal Hospital.

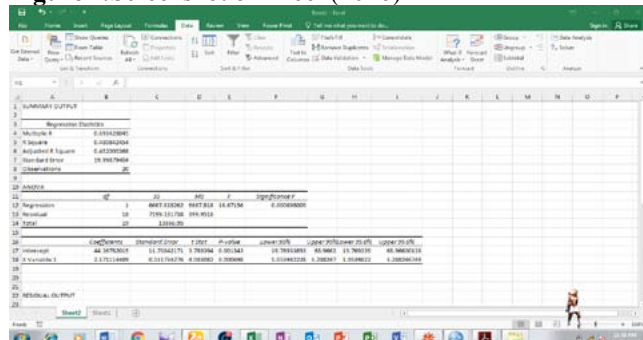**Figure I:** Screenshot of Excel (2016)

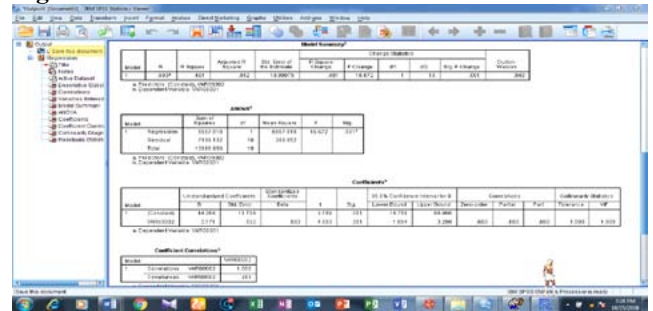
Figure II: Screenshot of SPSS

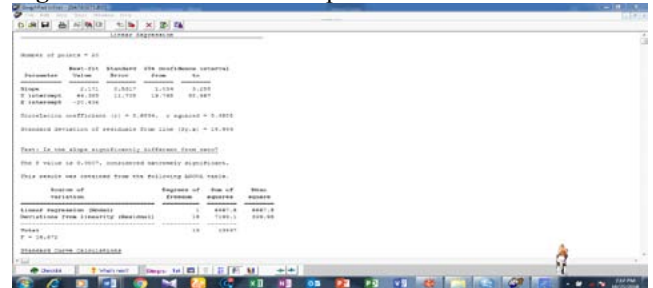**Figure III:** Screenshot of GraphPad
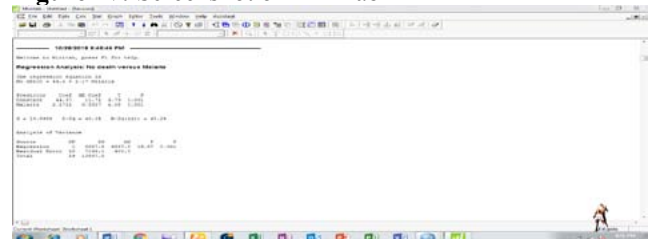

Figure IV: Screenshot of MiniTab

Figure V: Screenshot of R-package

From the above figure I to figure IV, SS were used to run regression model as one of the statistical tools discussed in this paper to ascertain the efficiency of SS. The result obtained were display and screenshot as shown in figure I to IV. It was discovered that the five different SS used for analysis gave the same result. The simple regression model obtained is expressed as follows:

Number of Death = 44+ 2.17 Malaria

CONCLUSION

The use of SS for academic research work is a necessity for all academic researchers. In this paper, we discovered that women are few in the academic profession. We also find out that the most common SS is MS Excel with 92% awareness from respondents, followed by SPSS with 80% awareness and the less popular SS is R with 4% respondents. Also, the frequently used SS is SPSS with 56% respondents, followed by MS Excel with 36% respondents, while the less usage SS is R with 2% respondents.

This paper also shows that there is a positive impact of Statistical Software in obtaining a fast and accurate result in academic staff research work in Kogi State Polytechnic (KSP). The following SS (MS Excel, SPSS, Graphpad, Minitab and R) were highlighted. Some of these packages can be easily downloaded from online and installed.

Output screenshot displayed gotten from the five selected SS were presented. It was discovered that the five different SS used for analysis gave the same result. The simple regression model obtained is expressed as follows: Number of Death = 44+ 2.17 Malaria

RECOMMENDATIONS

From the findings in this study, it is obvious that there is a positive impact of SS in academic research work. Thus, the following are recommended:

- The Management of Kogi State Polytechnic should organize training workshop on the use of SS on a regular basis.
- Academic Staff of the institution are encouraged to use SS for their quantitative data analysis in place of the manual or formula method.

REFERENCES

Adetola, O. G (2013). Learning Statistical Package Workbook. Nigeria (unpublished).

Abatan, S. M and Olayemi, M. S (2014). The role of statistical software in Data Analysis. IJARS, Vol. 3 (8).

Ben F. (2018). What is regression analysis and why should I use it? Retrieved from www.surveygizmo:com/resources/blog/regression- analysis.

Chris McGann (2014). Role of Statistical Software, EHow Contributor.

Hilary I.O., Pelumi E.O., Emmanuela C.O. and Elvir M.A., (2021). Trends and usage pattern of SPSS and Minitab Software in Scientific Research. *Journal of Physics: Conference Series* 1734 (1): 012017-2021

Top 48 Free Statistical Software in 2022-PAT Research. https://www.predictiveanalyticstoday.com.

Importance of Statistical Analysis by Data Collection. http://highqualityarticles.com/. online. Accessed on April, 2018.

Kimberly, A.B and Larry, A.P (2011). Preventing human error: The impact of data entry methods on data accuracy and statistical results. *Computers in Human Behavior* 27(5):1834-1839, DOI: 10.1016/j.chb.2011.04.004.

Kogi State Polytechnic, Lokoja Revised Conditions and Scheme of Service for Staff of the Polytechnic September, 2017.

Minitab Software. http://en.wikipedia.org/wiki/minitab. Accessed on February, 2022.

Okaygbue, H.I., Oguntunde, P.E., Obasi, E.C. and Akhmetshin, E.M., (2021). Trends and usage pattern of SPSS and Minitab software in Scientific Research. In *Journal of Physics: Conference Series* (Vol. 1734 (1), P.012017). IOP Publishing.

Olayemi, M. S (2019). Statistical Concept Textbook. Agunbay Publishers. ISBN: 978-978-50454-5-3.

Olayemi, M. S and Alabi, T.J. (2018). Multiple Regression Analysis on Selected Diseases: A case study of Zonal Hospital, Kabba, Kogi State. (Unpublished).