EMPIRICAL RESEARCH IN EDUCATION: DO STATISTICAL PACKAGES FOR DATA ANALYSIS MATTER?

Bala Bakwai Kwashabawa (Ph.D)
Department of Educational Foundations,
Faculty of Education and Extension Services,
UsmanuDanfodiyo University,
P. M. B. 2346, Sokoto, Nigeria.

ABSTRACT: *This paper is based on review of literature on the relevance of statistical packages in educational research. In this paper the importance of statistical packages for data analysis in educational research examined. The collective importance of statistical software packages is to help improve research expertise, make research work robust and faster, make research work easy and efficient, and minimize human error in data analysis. The paper clarifies the terms empirical research, and statistical data analysis. The statistical packages introduced in this paper include Microsoft Excel (MS-Excel), Minitab, Matlab, Stata, Statistical Package for the Social Sciences (SPSS), R, Statistical Analysis System (SAS),and Econometric Views (Eviews). The common features of statistical software as identified in this paper are that their basic programme are evolving, and they use both menus and/or command languages. The factors limiting the use of statistical packages among educational researchers include the perceived usefulness, statistical software self-efficacy, and statistics anxiety. Ways proposed in this paper to encourage educational researchers embrace the use statistical packages are as follows: Educational researchers should be mandated by the law governing their establishment to be research analyst. Choosing statistical software packages to learn should be based on suitability of the software. Mentorship in using statistical software package is necessary for learners in educational research.*

KEY WORDS: statistics, data analysis, empirical research, statistical package.

INTRODUCTION

Empirical researches are the major source of information in education. Education as a discipline comprises a number of fields of study ranging from business education, educational management, arts education, educational psychology, science education, curriculum studies, adult education, primary education, philosophy of education, test and measurement, educational technology and physical education among others. Apart from the qualitative measurement tools, education employs the scientific or quantitative measurements to generalize samples of study to a larger population just like the pure and applied sciences.

History has shown that different forms of data analysis techniques have been in existence. Initially, it was paper and pen and later the advent of the computer has helped the invention of punching machines and later upgraded to a simple calculator and complex scientific calculator. Later statistical software was introduced which has undoubtedly contributed enormously to the

improvement and easiness of data analysis in education. As quantitative research grows overtime, the application of statistical software becomes a more crucial part of data analysis. Researchers are experiencing a period where manual data analysis using paper and pen is completely taken over by more efficient digital analysis with statistical software.

Although scientific measures in data analysis are becoming most popular and widely accepted and used, many educational researchers face dilemmas in choosing the appropriate software package to be used. One of the most difficult and feared aspects of research is what type of statistical analysis technique and package one shall use. Use of any statistical analysis technique and software package will depend on the type of research questions asked, which in turn depend on the variable of one's study.

Conceptual Clarification

Some concepts considered to be important in this paper are briefly explained. These include:

Research: Kerlinger (1973) defined research as "systematic, controlled, empirical and critical investigation of hypothetical prepositions about the presumed relations among natural phenomena". It is a "careful, systematic, reliable and valid method of investigating knowledge and solving problems" (Wiersma 1991). A meaningful research is central and critical to the science of education, like any of the other sciences. Research is the orderly investigation of a subject matter for the purpose of adding to stock of knowledge. It means investigating a new problem or phenomenon in order to provide solutions to the identified issues. Research can be classified as historical, descriptive, correlational, causal, case study, ethnographic, or development research (Postlethwaite, 2005).

Empirical Research: Empirical research is based on observed and measured phenomena and derives knowledge from actual experience rather than from theory or belief. Empirical research is a type of research methodology that makes use of verifiable evidence in order to arrive at research outcomes. Empirical research is characterized by the following steps. First and foremost, it identifies research issues in terms of general and specific research questions and objective. Secondly, it searches for, and review of, other previous studies that (a) identify controversial debates, and knowledge gaps in the field; (b) elucidate the critical foundations that need to be tested empirically; and/or (c)provide excellent models in terms of design, management, analysis, reporting, and policy impact (Postlethwaite, 2005).Empirical research deals with the overall development of research design including specification of the information that is to be collected from which individuals under what research conditions. It also emphasizes on construction of operation definitions, preparation of instruments (tests, questionnaires, observation schedules, etc.), Pilot testing of instruments, data collection, and data analysis (Postlethwaite, 2005).

Data: We carry out empirical research to test hypotheses, and we do that by getting hold of data. Hopefully, if our experiments are planned and executed correctly, we can get hold of good data that can tell us something unique about the world (Farnsworth, 2019).Data implies information in numerical form that can be digitally transmitted or processed. Data is a factual information (as measurements or statistics) used as a basis for reasoning, discussion, or calculation. It is an

information output by a sensing device or organ that includes both useful and irrelevant or redundant information and must be processed to be meaningful (Winters, Winters, & Amedee, 2010). Data are divided into quantitative (numerical) which is a factual information obtained in numerical form and qualitative (categorical) data which represent characteristics such as gender, status, religion, etc.

Statistics: Statistics is a branch of science that deals with the collection, organisation, and analysis of data and drawing of inferences from the samples to the whole population. Statistics are the results of data analysis - its interpretation and presentation. In other words, some computation has taken place that provides some understanding of what the data means(Winters, Winters, & Amedee, 2010).

Statistical Data Analysis: Data analysis is often seen as the scariest aspect of completing research, but it does not have to be that way. While you will need to understand what to do with the data, and how to interpret the results, software that is designed for statistical analysis can make this process as smooth and as easy as possible (Farnsworth, 2019). Statistical data analysis is the process of collecting, transforming, and organizing data to detect useful information for making a well-informed decision. It gives the researchers real-time data about complex conditions to aid them in making decisions based on facts rather than hunches (Ray, 2019). It is the process of systematically applying statistical and/or logical techniques to describe and illustrate, condense and recap, and evaluate data. Data analysis is the process of evaluating data using the logical and analytical reasoning to carefully examine each component of the data collected or provided. There are two types of data analysis in empirical research: Qualitative data analysis (Thematic analysis, Content Analysis, Narrative Analysis, and Grounded Theory) and quantitative data analysis (Frequencies, Mean, Hypothesis Testing, and Sample Size Determination) (Perez, 2019)

Introduction to Statistical Packages

Statistical package is a software programme that makes the calculation and presentation of statistics relatively easy. It allows researchers to avoid routine mathematical mistakes and produce accurate figures in their research if they input all data correctly. To use most of the multidimensional statistical packages no prior knowledge of programming is required or assumed, but knowledge of basic computer skills and statistics appears necessary. A great number of tools are available in these packages to carry out statistical data analysis, and below is a list of most frequently used multidimensional software packages for educational data analysis:

Microsoft Excel (MS-Excel)

Microsoft Excel 2019 (the latest version as at April, 2020) is one of the most popular computer software programme designed to create spreadsheets and is part of the Microsoft Office productivity suite. Although Excel 2019 is not made for hardcore statistics but perform most popular statistical analysis ranging from mean and standard deviation to multiple regression ANOVA using excel 'Toolpak'. The excel Toolpak is an add-in that you must first install before you can use it. With this tool, you can make simple analysis and create charts about your current statistical data. Excel 2019 also includes many formulas and other tools to complete statistical

analyses (Peck, Olsen & Devore, 2016). Even though Excel perform most general statistical analyses, it is found very weak in regression, logistic regression, survival test, analysis of variance, factor analysis and multivariate analysis (Matthew, & Sunday, 2014).

MINITAB
MINITAB is a complete statistical software used by researchers from different field of study. It was developed around 1990 and remains one of the most popular statistical software programmes available. It is also one of the easiest statistical software programmes to use, and remains a popular choice. MINITAB is compatible with PC, Macintosh, Linux and all other major platforms. With drop-down menus and dialogue boxes describing how and what to do next, MINITAB persists as an academic choice for teaching and learning statistics and data analysis. It performs most general statistical analyses (regression, logistic regression, survival analysis, analysis of variance, factor analysis, but has its weaknesses in the general linear model (GLM) and Multilevel regression) (Matthew & Sunday, 2014). The most recent version of MINITAB is Version 14. There are some fairly substantial changes between Versions 13 and 14, but if you have the earlier version you should not worry. MINITAB uses interactive menus and is easy to use and learn. Unless that it has limited data management facilities and appropriate for very large data sets such as census tapes, etc.

MATLAB
MATLAB, which stands for Matrix and Laboratory, is a numerical computing environment and fourth-generation computer programming language. Developed by Math Works, MATLAB allows matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programmes written in other languages, including C, C++, Java, and Fortran. Although MATLAB is intended primarily for numerical computing, it allows access to symbolic computing capabilities. If one is not a programmer, a programmer can be hired to implement a specific model.

Stanford (2014) identified MATLAB as the richest statistical systems by far. It contains an impressive number of libraries, which is growing each day. Even if a much-desired specific model is not part of the standard functionality, it can be implemented, because MATLAB is really programming language with relatively simple syntaxes. As "languages" it allows expression of any idea. The question is whether one can write or not in programming language. MATLAB has much better graphics, which you will not be ashamed to put in a paper or a presentation. MATLAB performs most general statistical analyses (regression, logistic regression, survival analysis, analysis of variance, factor analysis, multivariate analysis). The greatest strengths of MATLAB is probably in its ANOVA, mixed model analysis and users creative freedom in the analysis.

Statistical Analysis System (SAS)
SAS latest version was released in December 2011. It is a package that many "power users" like because of its power and programmability. SAS also offers an interactive matrix programming language and exploratory data analysis with integration to R which is very beneficial. According to experts, SAS is somehow difficult to learn as compared to other packages. To use it, one must

learn to write its programme that manipulate data and perform data analyses. If one makes a mistake in a SAS programme, it can be very difficult to identify 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. However, SAS can work with many data files. It can handle enormous data files up to 32,768 variables and the number of records is generally limited to the size of a 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.

STATA
The name STATA is a syllabic abbreviation of the words statistics and data. It is a powerful, complete, integrated statistical package that provides everything you need for data analysis, data management, and graphics. STATA is produced by Stata Corp in College Station, TX. The most current professional version is STATA 16. Both STATA 13, 14 and 15 are sufficiently similar so that those who have access to the older version of STATA may feel alright. STATA is not sold in modules, which means you get everything you need in one package. And, one can also choose a perpetual license, with nothing more to buy ever. STATA is most commonly used for cross-sectional and panel data, but it is not as suitable for time-series data as the Eviews. There are several versions of STATA, such as STATA/IC, STATA/SE, and STATA/MP. The difference is basically in terms of the number of variables STATA can handle and the speed at which information is processed. For example, Stata/SE is a special edition that can handle up to 32,766 variables. (Matthew, & Sunday, 2014)

Statistical Package for the Social Sciences (SPSS)
SPSS was released in its first version in 1968 after being developed by Norman H. Nie; Dale H. Bent and C. Hadlai Hull before been acquired by the IBM Corporation as lately in 2009. It is now officially referred to as IBM SPSS. IBM SPSS is a statistical package used by researchers worldwide. It is a user-friendly package and various statistical tests could be conducted using it. It is used by all field of endeavors. SPSS undertakes both comparison and correlational statistical tests in the context of univariate, bivariate and multivariate analysis for both the parametric and non-parametric statistical techniques. The current version as at 2020 is version 26 which was released in 2019.

SPSS assists the user in describing data, testing hypotheses and looking for a correlation 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 (Matthew, & Sunday, 2014).

R
R is a language and environment for statistical computing and graphics. It is a GNU project which is similar to the S language and environment which was developed at Bell Laboratories (formerly

AT&T, now Lucent Technologies) by John Chambers and colleagues. Stanford (2014) identified R as the richest statistical system by far. They contain an impressive number 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 is really programming languages with relatively simple syntaxes. As "languages" it allows you to express any idea. The question is whether you are a good writer or not (of programming language). In terms of modern applied statistics tools, R libraries are somewhat richer. Also, R is a free software which can virtually be downloaded from different sites. R performs 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 user's creative freedom in the analysis. (Matthew, & Sunday, 2014)

Econometric Views (EViews)
Econometric Views (EViews) is a statistical package for windows, 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. The current version of EViews is 12.0, released in March 2019. 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,SAS and Stata (Matthew, & Sunday, 2014)

Importance of Statistical Packages to Educational Research
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 (Matthew, & Sunday, 2014). Statistical software packages have been discovered to help educational researchers in the following ways:
  i.   Statistical software packages help to improve research expertise.
 ii.   Statistical packages make research work robust and faster.
iii.   Statistical packages make research work easy and efficient.
 iv.   Statistical packages minimize human error in data analysis.
  v.   Statistical packages improve the efficiency of educational researchers.
 vi.   Statistical packages make research work easy, suitable and enjoyable.
vii.   Statistical packages are less expensive and some are free.
Features of Statistical Software
Multidimensional statistical packages have so many features that make them reliable and suitable for data analysis. The most common of these features according to Matthew and Sunday, (2014) are:
1.   Many of the basic programme are evolving: Each time a new version is released, problems may be eliminated, error handling may improve, new operating systems may be implemented, or hardware requirements may change.
2.   Most packages use both menus and/or command languages: Statistical packages differ in their use of menus or command languages. Some use menus, some use command, others use both.

Factor Limiting the Use of Statistical Packages by Educational Researchers

According to Brezavscek, Sparl and Znidarsic (2017), there are some factors limiting the use of statistical packages by educational researchers. These include:

Perceived Usefulness

Some educational researchers though that the use of statistical packages cannot in any way affect their productivity. To so many others, not even the knowledge of statistics is necessary. They believe that knowledge of statistics is a specialization made for some special people.

Perceived Ease of Use

Using statistical package to many educational researchers is a myth not something possible. Many educational researchers have phobia to operate even the simplest computer package, not to talk of what they consider as highly complicated as statistical packages. What some educational researchers believed is that statistical packages are specially designed programme for researchers in pure and applied sciences, and for in statistics. According to Matthew and Sunday (2014), most statistical packages are multidimensional and take care of a variety of field of endeavours.

Statistical Software Self-Efficacy

Many educational researchers would have used statistical packages efficiently if they have seen someone else using them before or if someone else had helped them get started. The problem is that many educational researchers are not familiar with statistical packages and they want to start using them, they try them on their own without undergoing training or looking for assistance from the good hands. If they could undergo training or allow somebody to mentor them or call somebody to help them, using the packages would have being very easy for them.

Computer Attitude

Computers are bringing us into a bright new era. The use of computers is enhancing our standard of living. There are unlimited possibilities of computer applications that haven't even been thought of yet. Computers are responsible for many of the good things we enjoy. Working with computers is an enjoyable experience. But still some educational researchers do not want to appreciate computer. Some educational researchers still believe in manually writing their research work and later employing the service of computer operator.

Statistics Anxiety

Some researchers wonder why they have to do all those things in statistics when in actual life they will never use them. Statistics is worthless to they since it's empirical and their area of specialization is something different. They feel statistics is a waste of time, they do not want to learn to like it. they wish the statistics requirement would be removed from their academic programme or research activities. They do not understand why somebody in their field needs statistics. Some researchers do not see why they have to clutter up their head with statistics. They think statistics has no significance to their life work.

Statistics Learning Self-Efficacy

No matter how much effort some researchers put in, they cannot learn statistics. When statistics activities are too difficult, they give up or only do the easy parts. When they find the statistics content of their research difficult, they do not put effort in learning the statistics, rather they drop the idea no matter how big or good it is.

Statistics Learning Value

Some researchers think that learning statistics is not important because there are experts who can do the statistical work for them. One cannot learn statistics unless he believes that he can use it in his daily life to stimulate his thinking, learn to solve his problems, participate in inquiry activities, and satisfy his own curiosity when learning statistics.

Ways to Encourage the Use of Statistical Packages among Educational Researchers

1. Educational researchers should be mandated by the law governing their establishment to be research analyst for their research findings and seize from giving their research data out for others to analyze.
2. Choosing statistical software packages to learn should be based on suitability of the software for all possible analysis researcher may wish to be analyzing. It was therefore advised that at least two packages should be chosen so that learning research analysis would be broaden and robust.
3. Educational institutions should include in the curriculum the practical usage and application of statistical packages in research study. Researchers should be trained with practical application of statistical packages in analyzing data. Research projects at the end of any educational programme should be based on practical application of statistical packages.

CONCLUSIONS

There is a range of different software tools available, and each offers something slightly different to the user – what you choose will depend on a range of factors, including your research question, knowledge of statistics, and experience of coding. The question of which package is suitable for educational researcher depend on: whether one is art, social science or science education researcher; and the type of data and methods on use. Since most educational researches involved a variety of statistical data and methods that cannot be easily handled by a singular statistical package, understanding of more than one package by educational researcher is necessary. This is because some packages are more effective in some type of data than others.

References

Farnsworth, B. (2019). The top 7 statistical tools you need to make your data shine. Retrieved on 13/4/2020 from https://imotions.com/blog/statistical-tools/
Ong, M. H. A. & Puteh, F. (2017). Quantitative data analysis: choosing between SPSS, PLS and AMOS in social science research. *International Interdisciplinary Journal of Scientific Research*, 3(1), 14-25. Retrieved on 13/4/2020 from *www.iijsr.org*
Pallant, J. (2015). *SPSS survival manual*. Berkshire: Open University Press.

Matthew, A. S. & Sunday, O. M. (2014). The role of statistical software in data analysis. *International Journal of Applied Research and Studies, 3(8), 1-15.*

Postlethwaite, T. N. (2005). Educational research: some basic concepts and terminology. In Kenneth N. Ross (Eds) *Quantitative research methods in educational planning*. Paris: International Institute for Educational Planning/UNESCO.

Shamoo, A. E. & Resnik, B. R. (2003). Responsible conduct of research. Oxford University Press. Retrieved on 17/4/2020 from https://ori.hhs.gov/education/products

Perez, E. (2019). What is Data Analysis and Its Methods? Retrieved on 17/4/2020 from https://www.utreee.com/what-is-data-analysis-and-its-methods%EF%BB%BF/

Winters, R.; Winters, A. & Amedee, R. G. (2010). Statistics: A brief overview. *Ochsner Jounal*, (10), 213-216.

Peck, R.; Olsen, C. & Devore, J. (2016). Excel technology manual for introduction to statistics and data analysis. California: Cengage Learning.

Ray, R. (2019). Top 3 benefits of statistical analysis software over manual analysis. Retreived on 19/4/2020 from https://medium.com/datadriveninvestor

Graaf, R. & Vossen, R. (2013). Bits versus brains in content analysis. Comparing the advantages and disadvantages of manual and automated methods for content analysis. *Communications*, 38(4), 433–443.

Brezavšček, A.; Šparl, P. & Žnidaršič, A. (2017). Factors Influencing the Behavioural Intention to Use Statistical Software: The Perspective of the Slovenian Students of Social Sciences. *Journal of Mathematics Science and Technology Education,* 13(3), 953-986. DOI 10.12973/eurasia.2017.00652. Retreived on 19/4/2020 from https://www.ejmste.com/download/factors-influencing-the-behavioural-intention