MODESTUM
EURASIA Journal of Mathematics, Science and Technology Education, 2026, 22(7), em2860
ISSN:1305-8223 (online)
OPEN ACCESS
Literature Review
https://doi.org/10.29333/ejmste/18789

Mathematics education students' utilization of statistical packages for innovation and advancement across STEM fields

Joshua Abah Abah [1*], Omadibi Zahradeen Idris [2]

[1] Department of Mathematics, Science and Technology Education, University of Zululand, Richards Bay, SOUTH AFRICA
[2] Department of Computer Science Education, Joseph Sarwuan Tarka University, Makurdi, Benue State, NIGERIA

Received 18 September 2025 ▪ Accepted 03 March 2026

Abstract

Statistical packages are software tools designed to assist with data analysis, statistical modelling, and visualization. They provide a user-friendly interface for performing complex statistical procedures without requiring advanced programming skills. These packages often include a range of statistical tests, data manipulation functions, and graphing capabilities, making them essential for research in social science discipline such as mathematics education. Some popular statistical packages include SPSS, SAS, R, GenStat, and Stata, each offering unique features and strengths to cater to different types of data analysis tasks. They help researchers, including mathematics education students, to avoid routine mathematical mistakes and produce accurate findings in their research. However, the acceptance and usage of these packages by student-researchers are below expectation given the present dearth of skills and motivation for deploying computational techniques in educational research. Theoretically grounded on self-efficacy theory, technology acceptance model, and expectancy-value theory, this article intends to conceptualize the dynamics of statistical package utilization among mathematics education students by exploring the foundational issues of educational research, statistics, available statistical software (SS), factors limiting usage, and measures for encouraging improved usage. This presentation stands out by addressing gaps in previous research, which often overlooked the unique challenges undergraduates face in adopting SS, focusing on the mathematical rigor and specific needs of mathematics education students, offering discipline-specific insights and strategies. By doing so, it not only enriches understanding but also provides practical recommendations to enhance software proficiency and statistical literacy required for driving innovations and advancements across science, technology, engineering, and mathematics fields.

Keywords: statistics, statistical packages, mathematics education, mathematics education undergraduates, student-researchers, research impact

INTRODUCTION

Statistical software (SS) packages are essential tools in undergraduate education, enabling students to perform complex data analyses and gain practical experience in statistical methods. The integration of these tools into curricula enhances students' analytical skills and prepares them for data-driven decision-making in various fields. The utilization of SS among undergraduates is a critical component of modern education, enhancing students' analytical capabilities and preparing them for data-centric careers. Integrating these tools into curricula, supported by appropriate resources and training, is essential for fostering a comprehensive understanding of statistical methods and their applications in science, technology, engineering, and mathematics (STEM) (Abah, 2018).

Researchers in a wide range of disciplines rely heavily on SS packages to process, analyze, and interpret data. The use of these tools enables researchers to apply complex statistical methods to large datasets, which would be impossible to analyze manually. Over the past few decades, the availability and advancement of SS have revolutionized the way research is conducted,

✉ AbahJ@unizulu.ac.za **(*Correspondence)** ✉ omadibi86@gmail.com

Contribution to the literature

- This review has successfully demonstrated the pivotal role of statistical knowledge in research endeavours and elaborates on the attitude of mathematics education students towards data analysis.
- By re-conceptualizing the self-efficacy theory, the technology acceptance model and expectancy-value model in the context of utilization of statistical packages by mathematics education students, this study connects students-researchers' effort-making to efficient performance in statistical data analysis.
- The recommendations for utilization of statistical packages for innovation and advancement across stem fields are practical nuggets that contribute richly to the body of knowledge on statistical packages utilization by student-researchers.

making it faster, more efficient, and more accurate (Khan & Memon, 2021). As technology continues to evolve, so too does the sophistication and diversity of these tools, contributing to more informed and nuanced insights.

One of the most prominent SS packages used by researchers is statistical package for the social sciences (SPSS). Originally developed for social science research, SPSS has become widely used in various fields such as psychology, healthcare, education, and market research. Its user-friendly interface and ease of use make it particularly attractive for researchers who may not have extensive backgrounds in statistics (Bala, 2016). SPSS provides a variety of statistical tests, data management features, and a graphical user interface that allows users to conduct analyses without needing to write code.

Another popular software tool for researchers is R, an open-source statistical computing environment. R has gained significant traction in academic research due to its versatility, extensive package library, and ability to handle complex analyses. Researchers in fields like biostatistics, economics, and social sciences utilize R for tasks ranging from basic descriptive statistics to advanced modeling techniques. R's flexibility allows researchers to extend its capabilities with custom functions and scripts, making it ideal for researchers who need more control over their analyses (Culpepper & Aguinis, 2011).

For researchers in fields that require heavy data manipulation and analysis, statistical analysis system (SAS) is another widely used tool. SAS is known for its robust data management capabilities and powerful statistical procedures. It is particularly prevalent in areas like epidemiology, clinical research, and finance, where large datasets require sophisticated analyses. While SAS is often associated with corporate and government research, its usage in academia is also substantial, particularly for research requiring high-performance computing and large-scale data processing (Pituch & Stevens, 2015).

Stata, which is widely used in economics, public health, and political science, is another critical tool for researchers. Known for its powerful data manipulation capabilities, Stata provides a variety of statistical tests and advanced econometric methods (Mehmetoglu & Jakobsen, 2022). Researchers frequently turn to Stata for its ability to manage large datasets and its intuitive command syntax, making it easier to learn than some other tools. Stata's popularity also comes from its ability to handle complex survey data, panel data, and longitudinal data analysis.

In addition to these widely known packages, newer and emerging tools such as JASP and Jamovi are becoming more common in research (Kangiwa et al., 2024). These open-source platforms provide an easy-to-use graphical interface while integrating the power of R under the hood. Both JASP and Jamovi focus on making statistical analyses accessible to users who may not have programming expertise (Şahin & Aybek, 2019). This democratization of statistical analysis is especially helpful in fields like psychology and education, where researchers may not have formal training in statistics but still need to conduct robust data analysis.

Similarly, MATLAB, primarily known for its applications in engineering and physical sciences, is also a powerful tool for statistical analysis. Researchers in fields like engineering, physics, and finance often use MATLAB for its computational power and ability to handle complex mathematical models. While it is more programming-intensive than software like SPSS, MATLAB's flexibility and extensive functions make it an essential tool for researchers dealing with simulations, optimization, and other specialized analyses (Niazai et al., 2023).

A growing trend among researchers is the use of Python with statistical libraries such as NumPy, SciPy, and Pandas (Lemenkova, 2019). Python's open-source nature, ease of learning, and extensive libraries make it an attractive choice for researchers in data science, machine learning, and artificial intelligence (AI). Python's ability to perform statistical analyses, combined with its broader capabilities in data cleaning, visualization, and web scraping, makes it increasingly popular in fields where big data and complex algorithms are involved (Harouni, 2024).

Obviously, there exist a countless array of statistical tools for researchers to choose from. The choice of SS can also be influenced by the size and scope of the dataset. For researchers working with large datasets, tools like Hadoop or Apache Spark may be utilized in conjunction with statistical packages to handle the processing and analysis of big data (Ahmed et al., 2020). These tools are essential in fields like genomics, finance, and social media research, where vast amounts of data must be processed efficiently. The integration of these big data technologies with SS packages enables researchers to scale their analyses, ensuring that results are timely and relevant (Singh et al., 2025).

Student researchers also face the challenge of selecting the most appropriate statistical package based on their research objectives. For example, in clinical research, GraphPad Prism is a popular choice due to its specialized capabilities for analyzing biological and medical data (Berkman et al., 2019). In contrast, Minitab is frequently used in industrial research and quality control for its user-friendly interface and reliability in conducting basic statistical analyses (Akers, 2018). Each software package has its own strengths and limitations, and researchers often choose one based on the specific requirements of their study.

Student researchers' use of SS packages is influenced by practical considerations such as cost, licensing, and availability of resources. While open-source tools like R and Python are cost-effective, commercial software packages like SAS and SPSS may require institutional licenses or paid subscriptions. The availability of training resources and technical support also plays a role in software adoption. Many academic institutions provide students and researchers with access to software tools, often through discounted or site-wide licenses, ensuring that researchers have the necessary resources to conduct their analyses effectively. However, there are underfunded institutions where student researchers lack the necessary tools for statistical data analysis. Across many institutions in the Global South, the plight of mathematics education student researchers in this regard is still under-researched.

Mathematics education students are uniquely positioned to leverage statistical packages as powerful tools for innovation and advancement across all STEM fields. Their foundational understanding of mathematical principles, coupled with training in pedagogical approaches, allows them to not only grasp the intricacies of these software applications but also to envision their practical application in diverse scientific and technological contexts. By proficiently utilizing packages like R, Python (with libraries such as NumPy and Pandas), SPSS, or SAS, these students can move beyond theoretical concepts to analyze complex datasets, identify patterns, model phenomena, and derive meaningful insights (Iji & Abah, 2017). This hands-on experience in data manipulation and interpretation is crucial for fostering an innovative mindset, enabling them to contribute to cutting-edge research and development in areas such as biomedical engineering, environmental science, computational finance, and AI.

The ability of mathematics education students to effectively employ statistical packages fosters a collaborative environment, bridging the gap between abstract mathematical theory and real-world scientific inquiry. Their expertise in data analysis allows them to support researchers and practitioners in various STEM disciplines by providing robust statistical methodologies for hypothesis testing, predictive modeling, and data visualization (Falebita & Kok, 2024). For instance, in engineering, they could help optimize designs based on performance data; in biology, analyze genetic sequences; and in computer science, evaluate algorithmic efficiency. This interdisciplinary utility of statistical packages, mastered by mathematics education students, significantly accelerates the pace of discovery and problem-solving, driving advancements that might otherwise be hindered by a lack of sophisticated data analysis capabilities within specialized STEM fields.

Equipping mathematics education students with strong statistical computing skills prepares them to be future educators and leaders who can inspire the next generation of STEM innovators (Abah et al., 2024). By demonstrating the practical applications of mathematics through data-driven projects, they can make abstract concepts more tangible and engaging for their students. This not only enhances mathematical literacy but also cultivates a crucial skill set demanded by an increasingly data-centric world (Asanre et al., 2025). Their proficiency in these packages allows them to design and implement curricula that emphasize computational thinking and data analysis, thereby empowering future students to contribute to the ongoing innovation and advancement across all STEM fields, from academic research to industrial applications (Abah, 2020a).

Considering the centrality of statistical packages to academic and career progress for higher education students, it is surprising that there is a poor attitude towards the utilization of these platforms by students (Soluade et al., 2022). The reported level of utilization of statistical tools by undergraduate students is cause for concern that is only recently being investigated (Sampson et al., 2023). Levels of utilization of SS has been linked to perceived competence, value, interest and cognitive competence (Hasabo et al., 2022; Osakuade & Ifunanya, 2020). Against this backdrop, this presentation attempts to conceptualize the dynamics around mathematics education students' utilization of statistical packages for innovation and advancement across STEM fields. First, the theoretical framework is discussed to explain the role of self-efficacy, technology acceptance, cognitive load and expectancy value. The conceptual framework then looks at the core challenges in educational research, statistical methods, accessible SS, barriers to utilization, and strategies for promoting increased usage. Sound recommendations for effective utilization of statistical packages for mathematics education research were also offered.

THEORETICAL FRAMEWORK

Self-Efficacy Theory

Self-efficacy theory, introduced by Bandura (1977), is a foundational concept in psychology that explains how individuals, beliefs in their ability to succeed in specific tasks influence their actions and outcomes. Bandura (1977) argued that self-efficacy impacts motivation, effort, persistence, and resilience. He identified four key sources that shape self-efficacy: mastery experiences (successfully completing a task), vicarious experiences (observing others succeed), verbal persuasion (encouragement from others), and emotional or physiological states (managing stress and anxiety). These factors collectively determine how confident a person feels about undertaking challenging activities. Bandura's (1977) work highlighted the importance of self-efficacy in education, workplace performance, and personal development. He suggested that individuals with higher self-efficacy are more likely to approach tasks with confidence and persistence, while those with low self-efficacy might avoid challenges or give up easily. The theory has been widely applied in various domains, including health behavior change, career development, and technology adoption.

Mathematics education students, equipped with foundational mathematical knowledge, stand at a unique intersection for contributing to innovation and advancement across STEM fields. The utilization of statistical packages (e.g., R and Python with libraries like NumPy/SciPy/Pandas, SPSS, and SAS) is a critical skill that empowers these students to translate theoretical understanding into practical, data-driven insights. By mastering these tools, they can engage in sophisticated data analysis, modeling, and visualization, which are indispensable for research, development, and problem-solving in areas ranging from engineering and computer science to biology and environmental studies. This proficiency allows them to bridge the gap between abstract mathematical concepts and their tangible applications, fostering interdisciplinary collaboration and driving novel solutions.

However, the effective utilization of these powerful statistical packages is significantly mediated by students' self-efficacy, as posited by Bandura's (1977) self-efficacy theory. Self-efficacy refers to an individual's belief in their capacity to execute behaviors necessary to produce specific performance attainments (Bandura, 1977). For mathematics education students, a high sense of self-efficacy in using SS means they are more likely to approach complex data challenges with confidence, persist in the face of technical difficulties, and ultimately achieve successful analytical outcomes. Conversely, low self-efficacy can lead to avoidance, reduced effort, and a decreased likelihood of engaging with these tools, thereby hindering their potential for innovation and advancement (Hasabo et al., 2022).

The cultivation of self-efficacy in this context can be fostered through several key mechanisms. Mastery experiences, where students successfully complete statistical projects using these packages, are paramount. These successes build confidence and reinforce the belief in their capabilities. Vicarious experiences, such as observing peers or instructors effectively utilize the software, can also inspire and demonstrate that mastery is attainable. Social persuasion, through constructive feedback and encouragement from educators, can further bolster their self-belief. Managing physiological and emotional states, such as reducing anxiety associated with complex software, contributes to a more positive learning environment and higher self-efficacy.

Therefore, for mathematics education programs to truly empower students for innovation and advancement in STEM, it is crucial to integrate hands-on, project-based learning with statistical packages, while also explicitly addressing and nurturing their self-efficacy (Charalambous et al., 2021). Educators must design curricula that provide ample opportunities for authentic data analysis experiences, offer supportive learning environments, and provide targeted feedback that emphasizes students' growth and capabilities. In line with Bandura (1977), by cultivating strong self-efficacy alongside technical proficiency, mathematics education students can confidently leverage statistical tools to drive impactful innovations and contribute significantly to the ongoing evolution of STEM fields.

Building on Bandura's (1977) self-efficacy framework, Walker and Brakke (2017) acknowledged that fostering statistical self-efficacy and positive attitudes about statistics is associated with course performance in psychology. Their study hypothesized that students enrolled in an advanced undergraduate statistics course would have higher self-efficacy for and more positive attitudes about statistical concepts and conducting analyses compared with students in the introductory statistics course. The work assessed 49 introductory and 67 advanced statistics students attending a historically Black college for women at the beginning and end of the semester. Results suggest that self-efficacy increased across the semester for all students, with advanced statistics students having higher self-efficacy at the beginning of the semester than introductory students as measured by the 6-point current statistics self-efficacy (CSSE) scale (mean difference = 0.96, confidence interval [CI] [0.54-1.38], $t = -4.52$, $p < .0001$) (Walker & Brakke, 2017). Baseline attitudes toward statistics were positively correlated with self-efficacy scales at both time points. However, self-efficacy and attitudes were not significantly correlated with final course grade. The results suggest that students hold rather positive attitudes toward statistics and can maintain confidence in their competency for statistical analysis, which may serve them well in future pursuits (Walker & Brakke, 2017).

Evidently, self-efficacy is one of the important factors influencing statistical engagement. Salim et al. (2018) show the influence of statistics self-efficacy towards statistics engagement among undergraduate students taking statistics in applied sciences course in Universiti Putra Malaysia. A total of 293 students were randomly selected from eight different faculties taking this statistics course as a compulsory subject. Descriptive analysis showed that students have low statistics self-efficacy with the median score of 2.00. On the other hand, the median score for statistics engagement is 4.00 indicating that students' statistics engagement was moderate. Spearman's rank correlation analysis showed that there is a positive relationship between statistics self-efficacy and overall statistics engagement ($rs = 0.36$; $p > 0.05$). Further analysis shows that statistics self-efficacy influence statistical engagement among students with a contribution of 25%. These findings indicate the need for students to have statistics self-efficacy because this would influence students' statistics engagement and in turn enables them to achieve good results in statistics (Salim et al., 2018).

Similarly, Kaufmann et al. (2022) present results showing development and validation of a self-assessment questionnaire for examining self-efficacy for statistics in psychology students. Upon using different methodological approaches, we demonstrate that the instrument has

(1) sound psychometric properties, and within the sample of university students and

(2) a robust latent structure disclosing three clearly distinctive profiles that are characterized by a complex and nonlinear interplay between perceived self-efficacy (for basic and advanced statistics), statistics anxiety, and students' belief in the relevance of statistics (Kaufmann et al., 2022).

Bandura's (1977) theory is especially relevant in the context of learning and skill acquisition, as it emphasizes the need for positive reinforcement and supportive environments. For instance, in learning SS, students gain confidence as they experience success in analyzing data, observe peers solving similar problems, and receive guidance from instructors. Barriers like anxiety and fear of failure can be tackled by mathematics educators to help students build stronger self-efficacy. Bandura's (1977) self-efficacy theory remains a cornerstone in psychological and educational research, demonstrating the profound impact of belief systems on behavior and outcomes. Its application extends to various fields, making it a versatile and enduring framework for understanding human motivation and action.

Technology Acceptance Model

Technology acceptance model (TAM) originated from the 1985 doctoral thesis of Fred D. Davis Jr. at the Sloan School of Management, Massachusetts Institute of Technology. The primary aim was to create and validate a theoretical model explaining how system features influence individuals' acceptance of technological systems. The broader goals were twofold: first, to enhance understanding of the user acceptance process by offering a novel theoretical perspective on the successful design of information systems; and second, to establish a theoretical basis for a practical user acceptance test, enabling developers and users to assess and refine processes prior to implementation (Davis, 1989).

The integration of statistical packages into mathematics education is crucial for fostering innovation and advancement across STEM fields, particularly when viewed through the lens of the TAM. TAM posits that perceived usefulness (the degree to which an individual believes that using a particular system would enhance their job performance) and perceived ease of use (the degree to which an individual believes that using a particular system would be free of effort) are key determinants of an individual's intention to use a technology (Davis, 1989). For mathematics education students, a high perceived usefulness of statistical packages, such as GenStat, R, Python libraries (e.g., NumPy, SciPy, and Pandas), or commercial software like SPSS and SAS, would stem from their ability to efficiently analyze complex datasets, visualize patterns, and build predictive models. This direct utility in solving real-world problems encountered in engineering, biological sciences, computer science, and other STEM disciplines significantly motivates their adoption.

The perceived ease of use plays a significant role in overcoming the initial learning curve associated with these powerful statistical tools. Educational institutions are expected to provide adequate training and support, ensuring that students feel confident and competent in navigating the interfaces, understanding the syntax (for code-based packages), and interpreting the outputs. When students find these packages user-friendly, their self-efficacy in data analysis grows, empowering them to explore innovative solutions and contribute meaningfully to interdisciplinary projects. This ease of use directly translates into greater engagement and sustained utilization, moving beyond mere theoretical understanding to practical application, which is vital for genuine advancement in STEM.

Ultimately, the successful utilization of statistical packages by mathematics education students for innovation and advancement across STEM fields hinges on a continuous feedback loop between perceived usefulness and perceived ease of use. As students gain proficiency and witness the tangible benefits of these tools in tackling diverse STEM challenges–from optimizing industrial processes to developing new medical treatments or enhancing AI algorithms–their belief in the technology's utility is reinforced. This positive reinforcement, coupled with ongoing improvements in software accessibility and pedagogical approaches, creates a virtuous cycle where statistical packages become indispensable tools for critical thinking, problem-solving, and driving transformative progress throughout the scientific and technological landscape.

The TAM model has been widely applied in education, business, and healthcare to improve technology adoption. In the context of SS, TAM suggests that students are more likely to develop confidence and proficiency when they find the tools easy to use and understand their practical benefits (Xulu et al., 2025). Training programs that emphasize these aspects can enhance students' technology adoption and learning outcomes (Abah et al., 2025).

Over the years, TAM has evolved, with extensions like TAM2 (Venkatesh & Davis, 2000) and the unified theory of acceptance and use of technology (Venkatesh et al., 2003), further expanding its scope. However, Davis's (1989) original framework remains a seminal contribution to understanding human-technology interactions, allowing researchers to have theoretical basis for empirical investigation into utilization behavior. For instance, educators need to know how to motivate business students (i.e., future business practitioners) to learn and use SS, which can provide the practical skills necessary for business professionals to analyze data and make informed decisions. Using a sample of 207 online MBA students from an AACSB accredited university in the Midwest, a modified TAM model was examined by Hsu et al. (2009) using LISREL 8.80. The empirical results show that both computer attitude and SS self-efficacy have significant positive effects on perceived usefulness (Hsu et al., 2009). In addition, it was found that both perceived usefulness and perceived ease of use positively influence learners' intentions to use SS, whereas their anxiety with statistics has a significant, negative impact on perceived usefulness, perceived ease of use and behavioral intentions (Hsu et al., 2009).

Expectancy-Value Model of Behavior

The expectancy-value model (EVM) has emerged as a model for understanding and predicting behavior in the process of adopting innovations (Wozney et al., 2006). Statistical package utilization within the education system must have a form of appeal as evidenced in its widespread integration by students and teachers alike. The EVM in its present state was developed by Jacquelynne S. Eccles and her colleagues in the early 1980s. In this model choices are assumed to be influenced by both negative and positive task characteristics, and all choices are assumed to have costs associated with them precisely because one choice often eliminates other options (Eccles & Wigfield, 2002). As a result, the relative value and probability of success of various options are key determinants of choice.

Proponents of EVM posit that students' beliefs about how well they will do on a task (called expectation for success) and how they value that task (called subjective task value [STV]) are related. They theorize that students' expectation for success and STV predict their achievement outcomes and that students are more likely to engage in tasks, including supporting their studies with statistical packages, that they value and in which they expect to do well (Ramirez et al., 2010).

A visual representation of EVM is depicted in **Figure 1**. Expectancies and values are assumed to directly influence performance, persistence, and task choice.

Expectancies and values are assumed to be influenced by task-specific beliefs such as perceptions of competence and difficulty of different tasks, and individual's goals and self-schema. According to Eccles et al. (1983), ability beliefs are conceived as broad beliefs about competence in a given domain.

EVM outlined four component factors of STV. These are attainment value, intrinsic value, utility value, and cost (Eccles et al., 1983). The importance a student attaches to the task is called attainment value. Intrinsic value is the students' interest in or enjoyment from engaging in the task. Utility value links the usefulness of the task to the students' future goals such as their careers. The relative cost factor refers to why a student might try to avoid the task, for example, fear of failure, task difficulty, or statistical anxiety.

According to this model, expectation for success are positively influenced by high attainment, intrinsic, and utility values, while high cost imposes a negative influence (Ramirez et al., 2010). These important values are set of stable, general beliefs about what is desirable, which emerge from both society's norms and the individual's core psychological needs and sense of self (Feather, 1988). As such, values are one class of motives that lead individuals to perform acts they think should be done (Eccles et al., 2002).

Consequently, the EVM is one of the most structured models in psychology to predict attitude by measuring attitudinal attributes and relevant external variables (Zhang et al., 2008). Students' choice of additional study strategies like accessing digital statistical platforms for study materials are embedded in the values they placed on such alternatives. Students tend to spend more time in self-initiated utilization of statistical packages because they find personal meaning and relevance in such interconnectivity and computational power (Hulleman & Harackiewicz, 2009). The expectations for success in

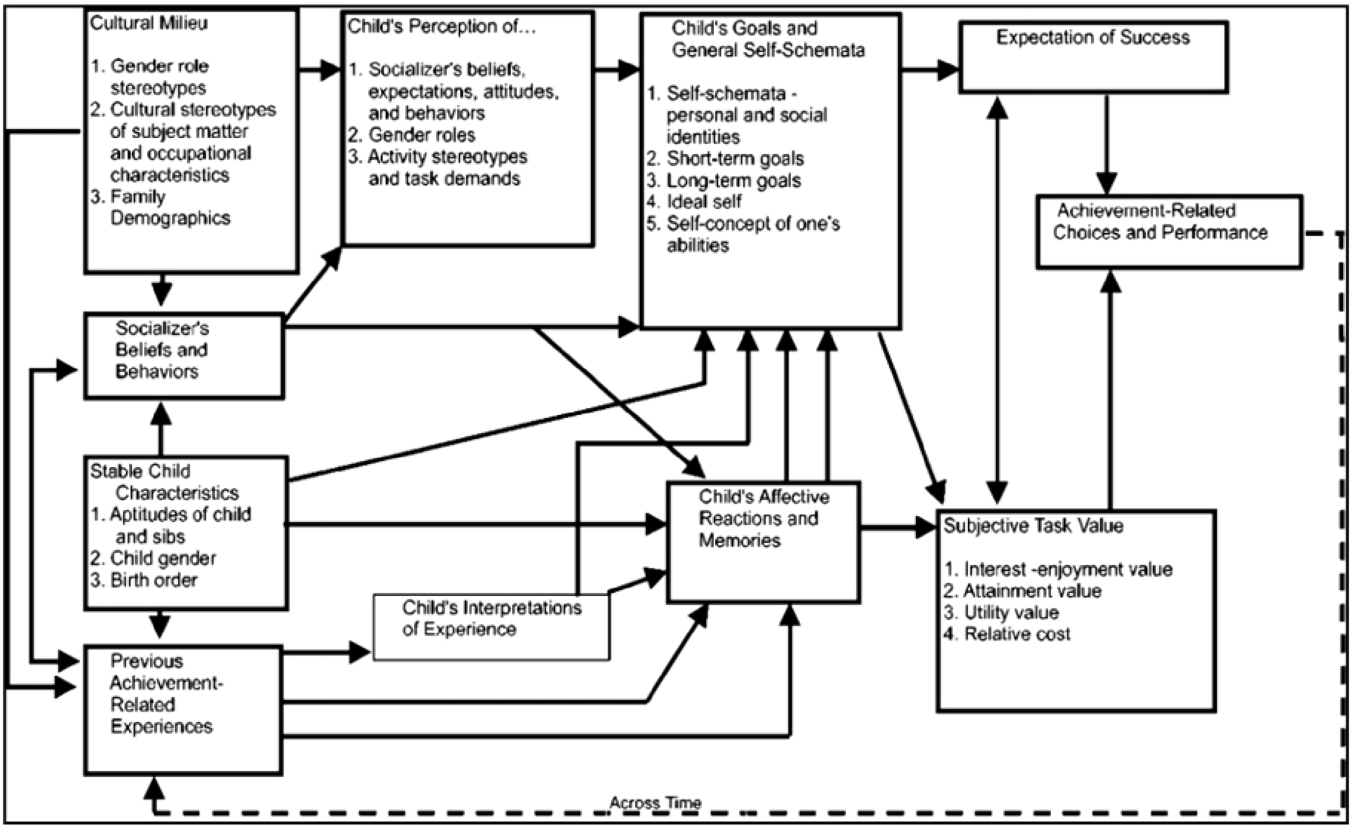

Figure 1. EVM (Eccles & Wigfield, 2002)

the study of statistics provide purposes, or reasons, for task and effort while engaged in the activity.

The use of statistical packages in universities by students of mathematics education constitutes an innovation with enormous potential for both instructors and students. According to the EVM, innovations such as statistical packages are more likely to be adopted if the perceived value of the innovation and the likelihood (or expectancy) of success are high, as well as if these benefits outweigh the perceived costs of implementation (Wozney et al., 2006). Obviously, usage of computational packages by mathematics educators and students enjoys high attainment, intrinsic and utility values at very low relative cost. That is to say, teachers' and students' decisions to use an innovation such as SS to augment classroom instruction delivery relate to

(a) how highly they value the innovation,

(b) how successful they expect their application of the innovation to be, and

(c) how highly they perceive the relative cost of implementation.

The intrinsic value component of the use of statistical packages by mathematics education students in universities points to a single fact: motivation. Motivation arises out of interest and grows into a positive disposition toward the study of mathematics. Hence, the EVM of behavior boldly indicates the development of positive attitude towards mathematics education inherent in the utilization of computational packages.

CONCEPTUAL FRAMEWORK

Statistics

Statistics is a branch of science that deals with the collection, organization, 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 et al., 2010). Statistics is the fulcrum upon which educational research rests.

Educational research is systematic application of scientific method for solving educational problems, regarding students and teachers as well. It attempts to organize data quantitatively and qualitatively to arrive at statistical inferences (Abah, 2020b). Presently, there is continuously growing demand for the statistical data analysis by researchers highlighting their need for statistical methods to be applied in statistical data processing. Statistics is widely used in almost all fields like biology, botany, commerce, medicine, education, physics, chemistry, biotechnology, psychology, zoology, to name a few. While doing research in the above fields, the researchers must have some awareness in using the statistical tools which helps them in drawing rigorous and good conclusions. The most well-known statistical tools are the mean, the arithmetical average of numbers, median and mode, Range, dispersion, standard deviation, inter quartile range, coefficient of variation, and many others (Rahman et al., 2024).

Research is systematic. Mathematics education research can be defined as a formal, systematic application of the scientific approach to the study of a problem within mathematics education to discover new information or expand and verify existing knowledge. Research as used in this definition is an inclusive way to accommodate the range of activities that support original and innovative work to increase the stock of knowledge in the discipline, including knowledge of people (students, teachers, administrators, and others), culture and society, methods, and technological developments. Research can also be defined as any form of disciplined enquiry that aims to contribute not only to knowledge building but also to theory building (Abah, 2018).

What makes research scientific and systematic is the use of the right statistical techniques and tools. Right statistics is the strength of methodology and not vice versa. Statistical procedures, techniques and tools are different from statistical packages/software. Research writing and reporting should transparently project methods of data analysis (statistics) and not versions of software packages.

Researchers are experiencing a transition from manual analysis with paper to more efficient digital/electronic analysis with SS (Abatan & Olayemi, 2014). The emergence of SS has undoubtedly contributed enormously to the progress in research studies in this 21st century. The high premium placed on ICT by society, researchers and organizations has undoubtedly made it a major drive of every nation (Ghaznavi et al., 2011).

SS is a vital tool for research analysis, data validation and findings. At the present day, all kinds of statistical methods are used in various academic fields, depending on the experimental data and the tasks that the researcher has to solve. In educational psychology, there are many areas of statistical methods application, ranging from descriptive statistics to inferential statistics. Descriptive statistics in mathematics Education research include data presentation tools (tables, charts, graphs, etc.), measure of central tendency (mean, mode, and median), measures of dispersion/variability (variance, standard deviation, range, etc.), and measures of association/relationship (correlation and regression).

On the other hand, inferential statistics go beyond mere presentation and description of data characteristics to make projections (hypothesis testing and estimation). Inferential statistical procedures are used to establish scientific facts and make important decisions. Inferential statistical computations are only carried out on data at the interval and ratio measurement scale or level (not meant for nominal, ordinal and categorical data). Inferential statistical tools are either parametric or non-parametric.

Parametric statistical procedures lead to decisions about key population characteristics (e.g., population mean and population variance) by studying related sample characteristics (e.g., sample mean and sample variance). Parametric statistics are used under strong assumptions about the population (assumptions include random distribution, independence of groups, normal distribution, homogeneity of variance, and no extreme outliers). Major parametric tools for mathematics education research include normal distribution test (1 sample), student's t-test (2 independent samples t-test), analysis of variance (ANOVA) (multiple independent samples), analysis of covariance (multiple independent samples with identical covariates), and factor analysis (PCA, EFA, CFA, SEM, path analysis, etc.).

Nonparametric statistics are distribution free procedures that do not rely on assumptions about the structure of the underlying population. They are readily available as alternatives to parametric statistics where assumptions cannot be tested or determined. Some examples of nonparametric statistical approaches are, McNemar test of change, runs test, Kolmogorov-Smirnov test, Chi-square test, median test, Mann-Whitney U test, Wilcoxon signed rank test, Kruskal-Wallis test, and Friedman's test.

Statistical Packages

Descriptive and inferential statistics are often introduced in probability and statistics. The textbook that is required, lectures, and regular chalk and board drills are all common teaching methods in traditional classroom-based statistics courses. However, recent technological advancements give educators a fresh approach to teaching statistics that is pertinent to and adaptable to the needs of the modern educational environment. Technology is increasingly being used in classrooms, and this has several benefits. The use of technology in the classroom can help students learn more actively, collaborate better, become more independent, and focus on task-based instruction, among other things. Nevertheless, despite these activities and movements, statistics is still consistently viewed as a subject that is dull, difficult, and frightening (Jones, 2018). Students often memorize the procedures and formulas needed to perform well on exams. Additionally, memorizing equations without understanding them is low level learning (Leung et al., 1997).

Statistical package is a software program 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.

Spreadsheets have a history spanning over three decades. The first, VisiCalc, was developed by Dan Bricklin in 1978 and released in 1979. Its premise was simple: an endless two-dimensional table where cells could hold text, numbers, or formulas. These formulas combined standard arithmetic operators with built-in functions and could reference the values of other cells (Tang, 2011). Despite its simplicity, the spreadsheet concept proved highly versatile, finding applications in accounting, inventory control, list management, and more. This flexibility turned VisiCalc into the first true "killer app" of the personal computing era (Tang, 2011). In the years that followed, successors such as Lotus 1-2-3 and Microsoft Excel (MS-Excel) introduced gradual enhancements, yet the core idea remained unchanged.

Multidimensional statistical packages have so many features that make them reliable and suitable for data analysis. The most common of these features according to Abatan and Olayemi (2014) are that many of the basic program 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. Similarly, 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 (Abatan & Olayemi, 2014).

Statistical packages feature a data editor for easy input, a wide array of statistical procedures, powerful visualization tools, and data management capabilities like import/export, data preparation, and transformations. They enable descriptive and inferential statistical analysis, predictive modeling, and analysis of complex relationships to facilitate data-driven insights and decision-making. Key functionalities include handling large datasets, generating various chart types, and allowing users to trace command flows and save their work (Deo & Ranganathan, 2024).

According to Adeyemi (2009), the use of appropriate statistical techniques is a critical requirement for effective conduct of social and behavioral research. Regarding the role of SS in data analysis, Abatan and Olayemi (2014) found that some analysis such as post hoc, complex analysis in time series, regression and variance analysis cannot be calculated manually effectively without SS package. They also stated that SS has contributed immensely to social research especially in the area of demographic and data analysis. SS packages have help mathematics education student researchers to improve research expertise; make research work robust and faster; make research work easy and efficient; minimize human error in data analysis; improve the efficiency; make research work easy, suitable and enjoyable; and the packages are less expensive and some are free.

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.

MS-Excel

MS-Excel is one of the most popular computer software programs designed to create spreadsheets and is part of the Microsoft Office productivity suite. Although MS-Excel is not made for hardcore statistics, it can perform most popular statistical analysis ranging from mean and standard deviation to multiple regression ANOVA using excel 'Toolpak'. The MS-Excel ToolPak is an add-in that provides advanced data analysis tools, including statistical, engineering, and financial functions not available in standard Excel. It enables users to perform complex calculations such as regression, ANOVA, and descriptive statistics with ease through a user-friendly interface. 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. Other third-party add-ins such as RealStat are also available and compatible across different versions of MS Excel. Excel also includes many formulas and other tools to complete statistical analyses (Peck et al., 2001). Even though Excel perform most general statistical analyses, on its own, it is found very weak in regression, logistic regression, survival test, ANOVA, factor analysis and multivariate analysis (Abatan & Olayemi, 2014).

MINITAB

MINITAB is a complete SS used by researchers from different field of study. It was developed around 1990 and remains one of the most popular SS programs available. It is also one of the easiest SS programs 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, ANOVA, factor analysis, but has its weaknesses in the general linear model and multilevel regression (Abatan & Olayemi, 2014). 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 results.

MATLAB

MATLAB, which stands for Matrix and Laboratory, is a numerical computing environment and fourth-generation computer programming language. Developed by MathWorks, MATLAB allows matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs 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.

Attaway (2022) 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 a "language," it allows expression of any idea. MATLAB has much better graphics. MATLAB performs most general statistical analyses (regression, logistic regression, survival analysis, ANOVA, factor analysis, and multivariate analysis). The greatest strength of MATLAB is probably in its ANOVA, mixed model analysis and users' creative freedom in the analysis (Martinez, 2021).

SAS

SAS 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 into R which is very beneficial. According to experts, SAS is somehow difficult to learn compared to other packages (Shrink, 2025). To use it, one must learn to write its program that manipulate data and perform data analyses. If one makes a mistake in a SAS program, it can be very difficult to identify where the errors occurred or how to correct them. However, it can take a long time to learn and understand data management in SAS than many other packages (Motiso, 2025). 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, ANOVA, factor analysis, multivariate analysis, etc.). The assertion that SAS's greatest strengths are ANOVA, mixed model, and multivariate analysis, and its weaknesses are ordinal and multinomial logistic regression (Yallati & Joshi, 2020), is partially correct, as SAS excels in advanced statistical modeling such as ANOVA and mixed models, but it can perform ordinal and multinomial logistic regression, and the complexity of these models is a challenge rather than an inherent weakness of the software itself (DexLab, 2025). While older sources might suggest relative difficulty or limitations with certain regression types, modern SAS versions have comprehensive capabilities for these analyses, and the difficulty often stems from the model's complexity and data requirements (SAS Support, 2025).

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. 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 Econometric Views (E-Views). 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.

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 late as 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 in 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. 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, ANOVA, factor analysis, and multivariate analysis but not suitable for time series analysis and multilevel regression analysis).

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. R's richness primarily comes from its vast ecosystem of packages. It has a comprehensive collection of tools for almost any statistical analysis, data visualization, or machine learning task (Code Academy, 2025). This is a huge advantage. 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 a programming language with relatively simple syntaxes. 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, ANOVA, factor analysis, and multivariate analysis). The greatest strengths of both are probably in its ANOVA, mixed model analysis and user's creative freedom in the analysis (Todorov, 2024)

E-Views

E-Views is a statistical package for windows, used mainly for time-series oriented econometrics analysis. It was developed by quantitative micro software and is now a part of IHS. Version 1.0 was released in March 1994. E-Views can be used for general statistical analysis and econometric analyses, such as cross-section and panel data analysis and time series estimation and forecasting. E-Views 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 (Aljandali & Tatahi, 2018).

GenStat

GenStat is a comprehensive SS package widely used for data analysis in research, particularly in agriculture, bioinformatics, and plant breeding. It is known for its powerful capabilities in analyzing complex experimental designs, such as split plot, randomized complete block, and incomplete block designs, which are common in biological and agricultural research. The software offers a rich set of statistical tools, including ANOVA, regression analysis, multivariate analysis, and linear mixed models. One of its key strengths is its ability to handle large datasets and complex data structures efficiently. GenStat also provides a flexible command-line interface, allowing users to write scripts for repetitive tasks and advanced analyses, making it a favorite among experienced statisticians and researchers (Payne, 2009).

In addition to its analytical power, GenStat has a user-friendly graphical interface that simplifies data management, visualization, and routine statistical tests for less experienced users. The software includes specialized routines for genetic analysis, such as the analysis of quantitative trait loci and marker-assisted selection, which are essential for modern breeding programs. It also provides tools for exploratory data analysis, including various plots and charts that help users understand their data better before performing formal statistical tests. GenStat's robust documentation and active user community ensure that users can find support and guidance when needed, making it a reliable and versatile tool for a wide range of statistical applications.

Python

Python serves as an excellent statistical package for mathematics education students, offering a powerful and accessible platform for learning and applying statistical concepts. Unlike traditional calculators or proprietary software, Python is an open-source language with a clear, readable syntax that makes it easier for students to grasp computational thinking. The use of libraries like NumPy, Pandas, and Matplotlib allows students to perform complex statistical analyses, from descriptive statistics and data manipulation to creating insightful visualizations. By writing their own code, students gain a deeper understanding of how statistical formulas and methods work "under the hood," moving beyond simply plugging numbers into a black box. This approach not only reinforces their mathematical knowledge but also develops critical problem-solving and programming skills that are highly valued in many modern careers (Paffenroth & Kong, 2015).

For mathematics education students, the integration of Python into the mathematics curriculum can also bridge the gap between abstract theory and real-world application. For instance, students can collect and analyze their own datasets, such as survey results or public data, to answer questions that are relevant to them. They can explore concepts like probability distributions, regression analysis, and hypothesis testing by building simulations and models. This hands-on, project-based learning fosters a more engaging and active learning environment (Nikula et al., 2007). Moreover, Python's versatility means it can be used for a wide range of mathematical topics beyond statistics, including algebra, calculus, and linear algebra, providing students with a single, powerful tool they can use throughout their mathematical journey (Harvey et al., 2014).

Factor Limiting the Use of Statistical Packages by Mathematics Education Student Researchers

There are several factors limiting the use of statistical packages by student researchers. In the view of Brezavšček et al. (2016), the most influential factors are found to be statistics anxiety and statistics learning value. Several other researchers have explored the factors limiting the use of statistical packages in educational research, with the most critical barriers being the high cost associated with proprietary software such as SPSS and Stata. The high cost of proprietary software like IBM SPSS Statistics and Stata is a significant barrier because these programs require a paid license, which can be a substantial upfront or recurring expense for individuals and organizations. While these packages offer advanced features, a user-friendly interface, and dedicated support, the need to purchase licenses restricts access, especially for students, researchers, and smaller institutions, making the high cost a critical challenge for many users in the field (Ozgur et al., 2015).

While free alternatives like R are available, the lack of comprehensive training programs poses a significant challenge. Educational institutions often fail to provide resources for learning, leaving student researchers unprepared to use these tools effectively. Even with R's open-access appeal, its programming requirements can deter those unfamiliar with coding, further compounding accessibility issues.

Hammerman and Terc (2004) identified the perceived complexity of SS as a major limiting factor. Tools like SAS, with their advanced functionalities and intricate syntax, are often seen as intimidating by non-experts. Researchers without a strong background in statistics may find it difficult to navigate these platforms, leading them to rely on simpler but less effective methods. Additionally, compatibility issues, such as poor performance on older hardware or inefficiencies with large datasets, add to the challenges (Contero et al., 2023).

Broadly, institutional and cultural factors play a substantial role in limiting the use of statistical packages. Many universities in developing countries do not prioritize investments in software licenses or technical support, leaving researchers without essential resources. Moreover, the reliance on traditional tools, such as spreadsheets, persists due to familiarity, even when these tools are insufficient for complex data analysis. Addressing these barriers requires a cultural shift towards embracing modern statistical tools, accompanied by targeted capacity-building initiatives. 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.

Likewise, using statistical package to many educational researchers is a myth; not something possible. Many educational researchers have phobia of operating even the simplest computer package, not to mention what they consider as highly complicated as statistical packages. What some educational researchers believed is that statistical packages are specially designed program for researchers in pure and applied sciences, and for those in core statistics. However, according to Brezavšček et al. (2016), most statistical packages are multidimensional and take care of a variety of field of endeavors.

Evidently, key factors that limit the use of statistical packages include the steep learning curve for users, high software costs, poor data quality, and anxiety related to statistics and software use. For the results to be reliable and actionable, these challenges must be managed effectively. The factors discussed in this section can be summarized into three key categories.

User-related factors

1. **Lack of statistical and technical knowledge:** Many users lack the foundational statistical knowledge to select appropriate tests or interpret results correctly. This can lead to misapplication of methods and misleading conclusions. Furthermore, some packages, especially open-source options like R, require programming knowledge that can be a barrier for non-programmers. Also, some researchers think that learning statistics is not important because there are experts who can do the statistical work for them (Gal & Ginsburg, 1994). 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 (Dani & Al Quraan, 2023).
2. **Statistics anxiety:** A user's fear or anxiety about statistics and mathematical concepts directly influences their willingness to engage with SS. This can be a significant psychological barrier, particularly for students and researchers without extensive quantitative training (Felix et al., 2024).
3. **Steep learning curve:** The sheer number of features, specialized jargon, and complex syntax in SS can be overwhelming for new users. A steep learning curve demands a significant time investment to achieve proficiency, which can deter casual or infrequent users.
4. **Poor communication of results:** Knowing how to generate results is not enough; researchers must also be able to communicate their findings clearly to non-technical audiences. This can be difficult when trying to translate complex statistical outputs into simple, actionable insights (Vandever, 2020).

Software and data-related factors

1. **High cost:** Many commercial statistical packages, such as SPSS and SAS, have high licensing costs, making them inaccessible for individual users, students, and smaller organizations. While free, open-source alternatives like R and Python exist, they may require more time and skill to use.
2. **Poor data quality:** SS can only produce meaningful results with good data. Flawed data, containing inaccuracies, inconsistencies, missing

values, or selection bias, will lead to unreliable output. The time-consuming process of data cleaning is often a major hurdle in data analysis projects.

3. **Difficulty with "big data:"** Some SS packages, such as SPSS, are not designed to handle massive datasets, limiting their use in big data analytics.
4. **Misuse and misinterpretation:** Statistical packages can be misused, either accidentally or intentionally, to produce misleading results. The possibility of misinterpretation is high, as the software's automated functions can hide underlying statistical assumptions from the user (see Abah, 2018 for details).

Organizational and logistical factors

1. **Lack of training and mentorship:** In academic and corporate settings, there is often inadequate training or mentorship available to help users overcome the challenges of statistical analysis. This can result in limited adoption and underutilization of powerful software capabilities.
2. **Shortage of skilled talent:** The market suffers from a shortage of data scientists and analysts with hands-on skills. This lack of talent means organizations may not have the in-house expertise to fully leverage statistical packages.
3. **Poor data governance:** Without a strong organizational culture that values and governs data quality, the accuracy of statistical analysis is undermined. This can lead to inconsistent data collection and analysis, which limits trust in the results.

Research has consistently verified the dynamics of these factors. For instance, Finney and Schraw (2003) developed measures of CSSE and self-efficacy to learn statistics (SELS) to address whether statistics self-efficacy is related to statistics performance, and whether self-efficacy for statistics increases during an introductory statistics course. Both instruments yielded reliable, one-factor solutions that were related positively to each other and to two measures of statistical performance (i.e., specific statistical problems and overall course performance). The CSSE and SELS also were related positively to math self-efficacy and attitudes towards statistics but related negatively to anxiety. Changes between two different testing occasions using the CSSE indicated that statistics self-efficacy increased almost two standard deviations over a 12-week instructional period.

Obviously, when statistical activities are too difficult, many student researchers give up or only do the easy parts. When they find the statistics content of their research difficult, they do not put effort into learning the statistics, rather they drop the idea no matter how big or good it is. In this regard, Perepiczka et al. (2011) emphasized that statistics plays an integral role in graduate programs. However, numerous intra- and interpersonal factors may lead to successful completion of coursework needed in this area. The authors examined the extent of the relationship between SELS and statistics anxiety, attitude towards statistics, and social support of 166 graduate students enrolled in master's and doctoral programs within colleges of education. Results indicated that statistical anxiety and attitude towards statistics were statistically significant predictors of SELS, yet social support was not a statistically significant predictor of self-efficacy. Insight into how student researchers respond to statistics courses and implications for educators are well established in literature (Garfield & Ben-Zvi, 2004; Garfield & Ben-Zvi, 2008; Heaton & Mickelson, 2002; Lovekamp et al., 2017).

WAYS TO ENCOURAGE THE USE OF STATISTICAL PACKAGES AMONG MATHEMATICS EDUCATION STUDENT RESEARCHERS

Encouraging the use of statistical packages among student researchers is essential for enhancing the quality and reliability of research findings. These tools, such as SPSS, R, and STATA, enable researchers to conduct complex data analyses efficiently, allowing them to draw meaningful conclusions from their studies. Key strategies include integrating SS training into research methodology courses, providing access to user-friendly software, and organizing workshops to build technical competence (Cujba & Pifarré, 2024). Institutions can also foster collaboration by creating networks or communities of practice where researchers can share experiences and skills. Additionally, offering incentives, such as funding for software subscriptions and recognizing research excellence through publications, can motivate researchers to embrace these tools. Studies have shown that targeted training programs and consistent institutional support significantly increase the adoption of SS in academic environments (Kwashabawa, 2021).

To encourage mathematics education student researchers to use statistical packages, focus on practical training, emphasize the real-world relevance of data analysis, integrate technology into the curriculum, and foster a supportive environment through mentorship and collaborative learning. Making statistical tools accessible, using student-centered approaches, and highlighting the efficiency and exploratory power of software can also increase adoption and positive attitudes toward statistical analysis (Kallivokas, 2023).

The acquisition of proficiency in statistical analysis software presents several initial challenges for learners. The extensive range of features, functions, and commands can appear overwhelming, particularly to novices who may struggle to identify where to begin. A gradual learning strategy, emphasizing the mastery of fundamental operations before progressing to more advanced tools, has been shown to mitigate these difficulties (Zavez & Harel, 2025). Equally critical is a solid grounding in mathematics and statistics, without which the accurate interpretation of outputs and analytical results remains problematic. Structured learning opportunities, such as formal courses or targeted workshops, can serve to strengthen this foundation and thereby enhance the effective application of SS.

In addition to theoretical preparation, meaningful engagement with authentic datasets is vital for consolidating skills and contextualizing statistical techniques within real-world applications. However, learners often encounter obstacles related to technical terminology and software-specific jargon, which can hinder comprehension. Developing a personalized glossary, consulting technical documentation, and drawing upon community-based resources can address this barrier. Furthermore, the considerable time investment required necessitates careful planning, including the establishment of realistic goals and consistent practice schedules. Active participation in professional or peer-learning networks provides not only practical guidance but also essential social and intellectual support, ultimately facilitating the transition from novice to proficient user of statistical analysis software.

Some Empirical Studies

Several studies have examined statistical package utilization among student researchers. For instance, Osakuade and Ifunanya (2020) undertook a study on the attitudes and competence of Nigerian undergraduates in statistical analysis using computer packages. This descriptive survey study targeted 200 undergraduates across various Nigerian universities, using a structured questionnaire with a Likert-scale design. The data was analyzed using descriptive and inferential statistics. Their findings indicate that most students demonstrated moderate attitudes and competence in using statistical packages like SPSS and MATLAB for data analysis. The study emphasized integrating software training into curricula and providing technical resources.

Similarly, Sampson et al. (2023) studied the utilization of statistical packages for data analysis among postgraduate students in universities in Rivers State. The study was guided by four research questions and their corresponding null hypotheses. The design for the study is descriptive research design. The population of the study is 397 postgraduate students in Universities in Rivers State with sample size of 91. One instrument was used for the study titled "Utilization of statistical packages for data analysis (USPDA)." Research questions were answered using mean and standard deviation. Hypotheses were tested using Z-test at 0.05 level of significance. Results revealed that statistical package is useful to post graduate students and is utilized among male and female post graduate students in universities in Rivers State and is most used by male students. The use of statistical package essentially reduces errors in data analysis, presentation and interpretation of result in research among students. Most schools, companies, business owners' governmental agencies use statistical packages for data analysis and presentation of results in either tabular or graphical form. Sampson et al. (2023) recommended that students as well as teaching and non-teaching staff should learn to improve their efficiency on the usage of statistical packages especially the one that is suitable for all analysis in their areas.

Dani and Al Quraan (2023) investigates whether research students' choice of research methodology is influenced by their attitudes toward statistics. A total of 81 students from three universities participated in a survey that included one open-ended question. Quantitative responses were examined using cluster analysis and independent sample t-tests to compare attitudes between master's and doctoral students. Additional methods, including cross-tabulation, chi-square tests, and ANOVA, were also employed. The open-ended responses were analyzed qualitatively to identify recurring themes such as emphasis on technique, data, or meaning. The findings of Dani and Al Quraan (2023) reveal that postgraduate students in the social sciences are generally less inclined to adopt quantitative research approaches. Moreover, master's students demonstrated different attitudes toward statistics compared to doctoral students. The results further indicate that students who perceive statistics solely as numerical methods tend to avoid it, opting instead for qualitative approaches. Drawing on these insights, the study offers recommendations for curriculum development.

In the same vein, Mutz and Daniel (2013) premised that Psychology students' attitudes toward research methods and statistics are often considered to influence course enrolment, persistence, performance, and the overall learning environment. Yet the variability across institutions has received little attention, despite its significance for teaching purposes given the heterogeneity of the student population. This work of Mutz and Daniel (2013) introduces a scale informed by social psychological research on attitudes, particularly their polar and affective dimensions, combined with a method for assessing first-year university students' views on research methods and statistics. The approach also allows for identifying the proportion of students holding positive attitudes at the institutional level. The study re-analyzed data from a nationwide German survey conducted in August 2000, which included all psychology students who began their studies in the 1999/2000 academic year (N = 1,490) across 44 universities. Using multilevel latent class analysis, the research sought to classify students into distinct attitude profiles while simultaneously categorizing universities according to the prevalence of these profiles. Findings revealed four latent student clusters arranged along a bipolar attitude continuum. Cluster membership was influenced by age, grade point average on the school-leaving exam, and personality traits. Furthermore, two university segments emerged: those with an average proportion of positively inclined students and those with a higher-than-average proportion, labelled as the "excellent" segment. Given the diversity within the psychology student population, the study underscores the need for a variety of learning activities rather than relying solely on traditional lecture-based teaching (Mutz & Daniel, 2013).

Another related study by Akimov et al. (2024) investigates the determinants of student success in a first-year university statistics course. Drawing on a unique sample from Westminster International University in Tashkent, the findings show that student engagement, evident in class attendance and use of online resources, remains a key contributor to academic performance. In addition, sociodemographic characteristics such as age, gender, marital status, employment, and prior schooling's language of instruction significantly influence academic outcomes. These insights carry important policy implications for both the university and the wider national context.

Banda et al. (2021) considered the angle of the teaching methods for statistics instruction. In their view, E-learning is one approach to teaching that actively engages students in learning statistics. Their study was designed to examine both the effects and challenges of e-learning among students at Mukuba University. A quasi-experimental design was employed, with a sample of 60 third-year mathematics students enrolled in a statistics course. Data were collected through a researcher-designed questionnaire and statistics performance tests. Students in the experimental group were taught gamma and beta functions and probability distributions of functions of random variables using an e-learning approach, while the control group received instruction through traditional methods. Data were analyzed using independent sample t-tests, with the null hypothesis (**H0**) tested at the 5% significance level. The results showed that e-learning significantly enhanced students' academic performance. Consequently, the study recommends adopting e-learning as an alternative approach in the teaching of statistics (Banda et al., 2021).

Adamu (2019) identified general broad challenges around software adoption by academic institutions, focusing on exploring the challenges and opportunities associated with the use of software products in Ethiopian higher education institutions. Data were collected from six universities through surveys, in-depth interviews, and observations to assess the extent of software utilization, the challenges faced, and future prospects. The findings reveal that software products are not sufficiently utilized in Ethiopian higher education institutions. Many departments lack software support for task execution and the delivery of standardized services. Several factors contribute to this inadequate use, with a major issue being the lack of awareness or commitment among university managers and leaders regarding the importance of software adoption. Additional challenges include resistance to adopting new technologies, absence of documentation or user manuals, inadequate infrastructure, unreliable internet connectivity, frequent power outages, shortage of trained professionals, and high staff turnover (Adamu, 2019).

CURRENT AND FUTURE DIRECTIONS OF STATISTICAL PACKAGE UTILIZATION IN EDUCATIONAL RESEARCH

In contemporary educational research, statistical packages have become essential tools for data analysis, enabling researchers to uncover trends, test hypotheses, and validate theoretical models (Muenchen, 2012). Packages such as SPSS, R, SAS, STATA, and Python-based libraries are widely used for both descriptive and inferential statistical analyses. These tools facilitate complex quantitative investigations, such as structural equation modeling, multivariate regression, and item response theory analysis. Their growing accessibility and integration with other data management platforms have increased their popularity, particularly in teacher education, policy evaluation, and learning outcomes research.

Currently, researchers leverage statistical packages for handling large datasets derived from assessments, surveys, and longitudinal studies. The rise in mixed-methods research has also encouraged integration between statistical and qualitative analysis tools, such as RQDA (for R) and NVivo, ensuring a comprehensive approach to examining educational phenomena (Onwuegbuzie et al., 2009). User-friendly interfaces in tools like SPSS and Jamovi have further empowered novice researchers to engage with quantitative data without extensive programming knowledge. Meanwhile, open-source platforms such as R and Python offer more customizable, cost-effective, and community-supported alternatives that are gaining traction in academic institutions globally (Fielding, 2012).

Looking toward the future, the role of statistical packages in educational research is expected to evolve alongside advancements in AI, machine learning, and big data analytics (Andreev et al., 2025). These innovations allow for more nuanced analyses, including predictive modeling and learning analytics, which can provide real-time insights into student performance and instructional effectiveness. Tools like Python's Scikit-learn, TensorFlow, and R's machine learning packages are increasingly being integrated into educational data analysis workflows. As educational institutions adopt data-driven decision-making models, the demand for advanced statistical skills and tools will continue to rise.

Furthermore, the future will likely see greater emphasis on cloud-based statistical environments and collaborative tools (Khan et al., 2015). Platforms such as Google Colab, JASP, and RStudio Cloud are already enabling real-time collaboration, version control, and remote data analysis, which are particularly beneficial for large, interdisciplinary research teams (Yadegaridehkordi et al., 2015). In addition, as data privacy and ethics become central concerns, future statistical packages may integrate stronger data governance features, including encryption, anonymization, and compliance tracking, particularly in studies involving minors or sensitive educational contexts (Lee & Jeong, 2016).

Evidently, the utilization of statistical packages in educational research is not only foundational but also rapidly expanding in capability and scope. As educational data grows in volume and complexity, researchers must stay abreast of emerging tools and techniques that enhance analytic accuracy and interpretability (Kalita et al., 2025). Professional development and curriculum revisions in education faculties must prioritize statistical literacy and competency with modern tools. Ultimately, the future of educational research will be shaped by the seamless integration of statistical packages into inquiry processes that are rigorous, ethical, and impact driven (Gul et al., 2025).

RECOMMENDATIONS FOR UTILIZATION OF STATISTICAL PACKAGES FOR INNOVATION AND ADVANCEMENT ACROSS STEM FIELDS

While some student researchers may delegate statistical work to experts, learning statistics is still crucial because it enables researchers to understand the research methodology, critically evaluate results, avoid misinterpretations, and contribute more effectively to the decision-making process that relies on statistical analysis. Relying solely on experts can lead to a superficial understanding and an inability to question or interpret the data and methods used, potentially undermining the validity of the research

Embarking on the journey of mastering statistical analysis software is a critical step for any aspiring mathematics education researcher. The road, however, is fraught with challenges that can be daunting. From the complexity of the software's features to the mathematical rigor required, learners often find themselves facing a steep learning curve. But with the right strategies and mindset, these obstacles can be overcome, paving the way for a rewarding career in mathematics education research that can advance innovation across STEM fields.

Encouraging mathematics education student researchers to use statistical packages is essential for developing data literacy and producing robust empirical studies. Many students experience "statistics anxiety" or a lack of self-efficacy when facing SS, which can be overcome by providing hands-on, contextualized training. The following four recommendations are proffered based on the present deliberation.

Implement Practical, Hands-On Training

1. Educational institutions should offer mandatory software training. They must be intentional in incorporating required modules on SS like R, SPSS, or JASP into research methods courses. The training should move beyond simple instructions and focus on practical application using easy-to-use statistical packages.
2. Schools should provide mentorship programs. They can begin by pairing student researchers with faculty mentors or more experienced graduate students who can provide personalized, one-on-one guidance on how to use statistical packages for specific research projects.
3. Educators can use project-based learning (PBL). This approach will encourage mathematics education students to work on real-world problems and datasets in a project-based learning environment. This makes statistics more relevant and helps students see the value of SS in solving tangible challenges.
4. Research educators should integrate software into assignments. Academic departments must require the use of SS in research proposals and thesis writing. This gives students a clear incentive to master the tools and provides structured opportunities for practice.

Promote Relevance and Reduce Anxiety

1. There is a need for institutional efforts to connect with students' research interests. Such practice would allow student researchers to analyze data related to their specific mathematics education interests. This increases motivation by letting students explore topics they find genuinely engaging.
2. Research methods lecturers must demonstrate the "why." Before teaching the "how," explain why SS is necessary for producing high-quality

research. Connect the software's capabilities to specific research questions and analytical needs in the field of mathematics education.

3. Start with user-friendly software. Introduce students to packages with a gentler learning curve. For instance, JASP provides a user-friendly interface for common statistical tests. As students' confidence grows, they can transition to more complex and powerful software like R or Python.

4. Both tutors and student researchers must normalize mistakes and anxiety. Schools can address statistical anxiety directly by creating a supportive learning environment. Emphasize that making mistakes is a normal part of the process and that the goal is not to be a perfect statistician but an effective user of tools.

Foster a Collaborative Environment

1. Intervention efforts should be geared towards encouraging collaborative work. Instructors are expected to implement collaborative learning strategies where student researchers work in small groups on data analysis projects. This allows students to share knowledge and develop problem-solving strategies together. Collaboration is also welcomed with industry experts and data scientists who are versed in utilization of different statistical packages.

2. Educational research methods academics must create social learning spaces. There is the need to establish a dedicated space, either physical or virtual, where students can discuss statistical challenges and share tips and tricks. This fosters a community of practice and provides a casual, low-pressure setting for students to get help.

3. Educational institutions should host workshops and special interest group discussions with mathematics education students. Universities must consistently organize workshops where faculty and peers can share their expertise and present real-world examples of how SS is used in their research.

Leverage Technology and Resources

1. Academic institutions must make resources accessible. Ensure that students have easy access to the required SS, either through free, open-source options like R or university-provided licenses.

2. Instructors can curate online resources. Teachers can create a repository of online tutorials, video guides, and code examples relevant to mathematics education research. Include guides for using visualization and data exploration tools.

3. Statistics tutors should incorporate reproducible analysis. It is important to teach students how to use reproducible research tools, such as R Markdown, which allow them to integrate their code, output, and commentary into a single document. This improves research rigor and makes sharing and replicating analysis easier.

CONCLUSION

Statistical package utilization has become indispensable in educational research, reshaping how data is collected, analyzed, and interpreted to inform practice and policy. The conceptualization presented thus far has acknowledged that while widely adopted tools such as SPSS continue to dominate due to their user-friendly features and robust analytical capabilities, the rise of open-source alternatives like R and Python reflects a shift toward greater flexibility and innovation in research methodologies. The expanding landscape underscores both opportunities and challenges, particularly the need for enhanced statistical literacy, rigorous training, and careful application to avoid misinterpretation or misuse. As researchers increasingly manage large and complex datasets, the integration of these packages ensures more rigorous evidence-based decision-making within education.

Looking ahead, advancements in AI and machine learning are expected to further transform SS by streamlining data processing, automating analytical tasks, and supporting more sophisticated research inquiries. These innovations, coupled with user-friendly interfaces, collaborative cloud-based solutions, and natural language processing, will make advanced analyses more accessible while demanding heightened attention to ethical considerations such as bias and privacy. Ultimately, the evolving capabilities of statistical packages will empower mathematics education student researchers to conduct more efficient, precise, and insightful studies. However, this potential can only be fully realized through continuous professional development and responsible utilization, ensuring that these powerful tools drive meaningful and sustainable improvements in mathematics education, advancing innovation in STEM generally.

Author contributions: JAA: theoretical framework, formatting; **OZI:** conceptualization; **JAA & OZI:** writing – original draft, writing – review & editing. Both authors agreed with the results and conclusions.

Funding: No funding source is reported for this study.

Ethical statement: The authors stated that the study is a literature review and does not involve contact with human participants. The research was conducted in accordance to set international standards for ethical research.

AI statement: The authors stated that generative AI tools such as ChatGPT and Gemini were minimally utilized for language editing purposes only. The idea, drafting and writing of the article are the original contribution of both authors.

Declaration of interest: No conflict of interest is declared by the authors.

Data sharing statement: Data supporting the findings and conclusions are available upon request from the corresponding author.

REFERENCES

Abah, J. A. (2018). The quest for statistical significance: Ignorance, bias and malpractice of research practitioners. *International Journal of Research & RE-View, 5*(3), 112-129.

Abah, J. A. (2020a). *Research practice and ethics in mathematics education*. Network for Grassroots Science and Mathematics Education.

Abah, J. A. (2020b). *Teaching computational thinking*. Network for Grassroots Science and Mathematics Education.

Abah, J. A., Iji, C. O., & Chinaka, T. W. (2024). Preparing pre-service STEM teachers in an emerging global environment. *Discourse and Communication for Sustainable Education, 15*(2), 62-78. https://doi.org/10.2478/dcse-2024-0016

Abah, J. A., Terungwa, U. D., & Chinaka, T. W. (2025). Conceptualizing the influence of artificial intelligence on students' academic integrity. *Journal of Mathematics Instruction, Social Research and Opinion, 4*(2), 257-290. https://doi.org/10.58421/misro.v4i2.383

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

Adamu, M. A. (2019). Challenges and prospects of using software products in Ethiopian higher educational institutions. *International Journal of System and Software Engineering, 7*(1), 1-12.

Adeyemi, T. O. (2009). Inferential statistics for social and behavioural research. *Research Journal of Mathematics and Statistics, 1*(2), 47-54.

Ahmed, N., Barczak, A. L., Susnjak, T., & Rashid, M. A. (2020). A comprehensive performance analysis of Apache Hadoop and Apache Spark for large scale data sets using HiBench. *Journal of Big Data, 7*(1), 110. https://doi.org/10.1186/s40537-020-00388-5

Akers, M. D. (2018). *Exploring, analysing and interpeting data with MINITAB 18*. Compass Publishing.

Akimov, A., Malin, M., Sargsyan, Y., Suyunov, G., & Turdaliev, S. (2024). Student success in a university first-year statistics course: Do students' characteristics affect their academic performance? *Journal of Statistics and Data Science Education, 32*(1), 24-35. https://doi.org/10.1080/26939169.2023.2184435

Aljandali, A., & Tatahi, M. (2018). *Economic and financial modelling with E-Views. A guide for students and professionals*. Springer. https://doi.org/10.1007/978-3-319-92985-9

Andreev, H., Pericleous, K., & Kosmas, P. (2025). Expanding the frontiers: Revolutionizing quantitative data analysis in social sciences with AI. In P. A. Christou (Ed.), *Artificial intelligence (AI) in social research* (pp. 37-47). CABI. https://doi.org/10.1079/9781800626607.0004

Asanre, A., Sondlo, A., & Abiodun, T. (2025). Exploring the relationship between broken homes and mathematical literacy: A quantitative study of Junior secondary school students. *Journal of Mathematics Instruction, Social Research and Opinion, 4*(2), 333-346. https://doi.org/10.58421/misro.v4i2.404

Attaway, D. C. (2022). *MATLAB: A practical introduction to programming and problem solving*. Butterworth-Heinemann.

Bala, J. (2016). Contribution of SPSS in social sciences research. *International Journal of Advanced Research in Computer Science, 7*(6), 250-255.

Banda, G., Tailoka, F. P., & Muma, E. (2021). Effect of e-learning on the academic performance of third year students in statistics at Mukuba University. *Journal of Mathematics and Science Teacher, 1*(2), 24-56. https://doi.org/10.29333/mathsciteacher/11277

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological RE-View, 84*(2), 191-215. https://doi.org/10.1037/0033-295X.84.2.191

Berkman, S. J., Roscoe, E. M., & Bourret, J. C. (2019). Comparing self-directed methods for training staff to create graphs using Graphpad Prism. *Journal of Applied Behavior Analysis, 52*(1), 188-204. https://doi.org/10.1002/jaba.522

Brezavšček, A., Šparl, P., & Žnidaršič, A. (2016). Factors influencing the behavioural intention to use statistical software: The perspective of the Slovenian students of social sciences. *Eurasia Journal of Mathematics, Science and Technology Education, 13*(3), 953-986. https://doi.org/10.12973/eurasia.2017.00652a

Charalambous, M., Hodge, J. A., & Ippolito, K. (2021). Statistically significant learning experiences: Towards building self-efficacy of undergraduate statistics learners through team-based learning. *Educational Action Research, 29*(2), 226-244. https://doi.org/10.1080/09650792.2020.1782240

Code Academy. (2025). R programming language explained (careers, use cases, difficulty). *Code Academy*. https://www.codecademy.com/article/what-is-r

Contero, M., Pérez-López, D., Company, P., & Camba, J. D. (2023). A quantitative analysis of parametric CAD model complexity and its relationship to perceived modeling complexity. *Advanced Engineering Informatics, 56*, Article 101970. https://doi.org/10.1016/j.aei.2023.101970

Cujba, A., & Pifarré, M. (2024). Enhancing students' attitudes towards statistics through innovative technology-enhanced, collaborative, and data-driven project-based learning. *Humanities and Social Sciences Communications, 11*(1), 1-14. https://doi.org/10.1057/s41599-024-03469-5

Culpepper, S. A., & Aguinis, H. (2011). R is for revolution: A cutting-edge, free, open source statistical package. *Organizational Research Methods, 14*(4), 735-740. https://doi.org/10.1177/1094428109355485

Dani, A., & Al Quraan, E. (2023). Investigating research students' perceptions about statistics and its impact on their choice of research approach. *Heliyon, 9*(10), Article e20423. https://doi.org/10.1016/j.heliyon.2023.e20423

Davis, F. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*, 319-340. https://doi.org/10.2307/249008

Deo, V., & Ranganathan, P. (2024). Statistical tools and packages for data collection, management, and analysis-A brief guide for health and biomedical researchers. *Perspectives in Clinical Research, 15*(4), 209-212. https://doi.org/10.4103/picr.picr_160_24

DexLab. (2025). The right tool for statistical analysis: SAS vs. Stata. *DexLab*. https://m.dexlabanalytics.com/blog/the-right-tool-for-statistical-analysis-sas-vs-stata

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual RE-View of Psychology, 153*, 109-132. https://doi.org/10.1146/annurev.psych.53.100901.135153

Eccles, J. S., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., & Meece, J. L. (1983). Expectancies, values and academic behaviors. In J. T. Spencer (Ed.), *Achievement and achievement motives* (pp. 75-146). Freeman.

Falebita, O. S., & Kok, P. J. (2024). Strategic goals for artificial intelligence integration among STEM academics and undergraduates in African higher education: A systematic rE-View. *Discover Education, 3*(1), Article 151. https://doi.org/10.1007/s44217-024-00252-1

Feather, N. T. (1988). Values, valences, and course enrollment: Testing the role of personal values within an expectancy-value framework. *Journal of Educational Psychology, 80*, 381-391. https://doi.org/10.1037/0022-0663.80.3.381

Felix, C. S., Mocnangan, J. T., Lartec, C. I., Hugo, I. G., & Pinas, A. C. (2024). Challenges in basic statistics: Implications for research accuracy and reliability. *Cognizance Journal of Multidisciplinary Studies, 4*(5), 1-13. https://doi.org/10.47760/cognizance.2024.v04i05.001

Fielding, N. G. (2012). Triangulation and mixed methods designs: Data integration with new research technologies. *Journal of Mixed Methods Research, 6*(2), 124-136. https://doi.org/10.1177/1558689812437101

Finney, S. J., & Schraw, G. (2003). Self-efficacy beliefs in college statistics courses. *Contemporary Educational Psychology, 28*(2), 161-186. https://doi.org/10.1016/S0361-476X(02)00015-2

Gal, I., & Ginsburg, L. (1994). The role of beliefs and attitudes in learning statistics: Towards an assessment framework. *Journal of Statistics Education, 2*(2). https://doi.org/10.1080/10691898.1994.11910471

Garfield, J., & Ben-Zvi, D. (2004). Research on statistical literacy, reasoning, and thinking: Issues, challenges, and implications. In D. Ben-Zvi, & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 397-409). Springer. https://doi.org/10.1007/1-4020-2278-6_17

Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. Springer.

Ghaznavi, M. R., Keikha, A., & Yaghoubi, N. M. (2011). The impact of information and communication technology (ICT) on educational improvement. *International Education Studies, 4*(2), 116-125. https://doi.org/10.5539/ies.v4n2p116

Gul, M. N., Abbasi, W., Babar, M. Z., Aljohani, A., & Arif, M. (2025). Data driven decisions in education using a comprehensive machine learning framework for student performance prediction. *Discover Computing, 28*(1), 1-34. https://doi.org/10.1007/s10791-025-09585-3

Hammerman, J. K., & Terc, A. R. (2004). Strategies for managing statistical complexity with new software tools. *Statistics Education Research Journal, 3*(2), 17-41. https://doi.org/10.52041/serj.v3i2.546

Harouni, I. (2024). The modern methods of data analysis in social research: Python programming language and its pandas library as an example–A theoretic study. *Social Empowerment Journal, 6*(1), 56-70. https://doi.org/10.34118/sej.v6i1.3806

Harvey, P. A., Wall, C., Luckey, S. W., Langer, S., & Leinwand, L. A. (2014). The python project: A unique model for extending research opportunities to undergraduate students. *CBE – Life Sciences Education, 13*(4), 698-710. https://doi.org/10.1187/cbe.14-05-0089

Hasabo, E. A., Ahmed, G. E. M., Alkhalifa, R. M., Mahmoud, M. D., Emad, S., Albashir, R. B., Mansour, M. E., & Hsabo, E. A. (2022). Statistics for undergraduate medical students in Sudan: Associated factors for using statistical analysis software and attitude toward statistics among undergraduate medical students in Sudan. *BMC Medical Education, 22*(1), Article 889. https://doi.org/10.1186/s12909-022-03960-0

Heaton, R. M., & Mickelson, W. T. (2002). The learning and teaching of statistical investigation in teaching and teacher education. *Journal of Mathematics Teacher Education, 5*(1), 35-59. https://doi.org/10.1023/A:1013886730487

Hsu, M. K., Wang, S. W., & Chiu, K. K. (2009). Computer attitude, statistics anxiety and self-efficacy on statistical software adoption behavior: An empirical study of online MBA learners. *Computers in Human Behavior, 25*(2), 412-420. https://doi.org/10.1016/j.chb.2008.10.003

Hulleman, C. S., & Harackiewicz, J. M. (2009). Promoting interest and performance in high school science classes. *Science, 328*(5958), 1410-1412. https://doi.org/10.1126/science.1177067

Iji, C. O., & Abah, J. A. (2017). Exploring gender differences in graduation proficiency in mathematics education using a Markov chain model: Implications for economic growth in Nigeria. *Educational Excellence, 3*(2), 75-86. https://doi.org/10.18562/ijee.028

Jones, A. (2018). *Probability, statistics and other frightening stuff*. Routledge. https://doi.org/10.4324/9781315160061

Kalita, E., Oyelere, S. S., Gaftandzhieva, S., Rajesh, K. N., Jagatheesaperumal, S. K., Mohamed, A., Elbarawy, Y. M., Desuky, A. S., Hussain, S., Cifci, M. A., Theodorou, P., Hilcenko, S., Hazaraki, J., & Ali, T. (2025). Educational data mining: A 10-year rE-View. *Discover Computing, 28*(1), Article 81. https://doi.org/10.1007/s10791-025-09589-z

Kallivokas, D. (2023). Learning statistics using specialized software applications. In *Proceedings of the International Conference on Strategic Innovative Marketing and Tourism* (pp. 749-756). Springer. https://doi.org/10.1007/978-3-031-51038-0_81

Kangiwa, B. I., Ladan, I. M., Nassarawa, H. S., Sabo, S. A., & Umar, M. A. (2024). Free and open-source software for data analysis: Leveraging the potentials of JASP, Jamovi and PSPP in Nigeria tertiary institutions. *International Journal of Multidisciplinary Research in Science, Technology and Innovation, 3*(1), 1-8.

Kaufmann, L., Ninaus, M., Weiss, E. M., Gruber, W., & Wood, G. (2022). Self-efficacy matters: Influence of students' perceived self-efficacy on statistics anxiety. *Annals of the New York Academy of Sciences, 1514*(1), 187-197. https://doi.org/10.1111/nyas.14797

Khan, S., & Memon, M. A. (2021). Data revolution and statistics: Reshaping society and science. In *Proceedings 18th International Conference on Statistical Sciences* (pp. 7-20).

Khan, Z., Anjum, A., Soomro, K., & Tahir, M. A. (2015). Towards cloud based big data analytics for smart future cities. *Journal of Cloud Computing, 4*(1), Article 2. https://doi.org/10.1186/s13677-015-0026-8

Kwashabawa, B. B. (2021). Empirical research in education: Do statistical packages for data analysis matter? *European Journal of Training and Development Studies, 8*(1), 27-35.

Lee, S., & Jeong, T. (2016). Cloud-based parameter-driven statistical services and resource allocation in a heterogeneous platform on enterprise environment. *Symmetry, 8*(10), Article 103. https://doi.org/10.3390/sym8100103

Lemenkova, P. (2019). Processing oceanographic data by Python libraries NumPy, SciPy and Pandas. *Aquatic Research, 2*(2), 73-91. https://doi.org/10.3153/AR19009

Leung, M., Low, R., & Sweller, J. (1997). Learning from equations or words. *Instructional Science, 25*(1), 37-70. https://doi.org/10.1023/A:1002969618881

Lovekamp, W. E., Soboroff, S. D., & Gillespie, M. D. (2017). Engaging students in survey research projects across research methods and statistics courses. *Teaching Sociology, 45*(1), 65-72. https://doi.org/10.1177/0092055X16673136

Martinez, W. L. (2011). Computational statistics in MATLAB®. *Wiley Interdisciplinary RE-Views: Computational Statistics, 3*(1), 69-74. https://doi.org/10.1002/wics.138

Mehmetoglu, M., & Jakobsen, T. G. (2022). *Applied statistics using Stata: A guide for the social sciences*. SAGE. https://doi.org/10.4135/9781036231446

Motiso, D. (2025). SAS vs. R: Definitions, advantages and differences. *Indeed*. https://www.indeed.com/career-advice/career-development/sas-vs-r

Muenchen, R. A. (2012). The popularity of data analysis software. *r4stats*. https://r4stats.com/articles/popularity/

Mutz, R., & Daniel, H. D. (2013). University and student segmentation: Multilevel latent-class analysis of students' attitudes towards research methods and statistics. *British Journal of Educational Psychology, 83*(2), 280-304. https://doi.org/10.1111/j.2044-8279.2011.02062.x

Niazai, S., Rahimzai, A. A., & Atifnigar, H. (2023). Applications of MATLAB in natural sciences: A comprehensive rE-View. *European Journal of Theoretical and Applied Sciences, 1*(5), 1006-1015. https://doi.org/10.59324/ejtas.2023.1(5).87

Nikula, U., Sajaniemi, J., Tedre, M., & Wray, S. (2007). Python and roles of variables in introductory programming: Experiences from three educational institutions. *Journal of Information Technology Education: Research, 6*(1), 199-214. https://doi.org/10.28945/3097

Onwuegbuzie, A. J., Slate, J. R., Leech, N. L., & Collins, K. M. (2009). Mixed data analysis: Advanced integration techniques. International *Journal of Multiple Research Approaches, 3*(1), 13-33. https://doi.org/10.5172/mra.455.3.1.13

Osakuade, J. O., & Ifunanya, M. (2020). Attitudes of Nigerian undergraduates towards statistics and their perceived competence in the use of computer statistical packages for data analysis. *Prestige Journal of Education, 3*(1), 29-39.

Ozgur, C., Kleckner, M., & Li, Y. (2015). Selection of statistical software for solving big data problems: A guide for businesses, students, and universities. *SAGE Open, 5*(2). https://doi.org/10.1177/2158244015584379

Paffenroth, R. C., & Kong, X. (2015). Python in data science research and education. In *SciPy* (pp. 164-170). https://doi.org/10.25080/Majora-7b98e3ed-019

Payne, R. W. (2009). GenStat. *Wiley Interdisciplinary RE-Views: Computational Statistics, 1*(2), 255-258. https://doi.org/10.1002/wics.32

Peck, R., Olsen, C., & Devore, J. L. (2001). *Introduction to statistics and data analysis*. Brooks/Cole.

Perepiczka, M., Chandler, N., & Becerra, M. (2011). Relationship between graduate students' statistics self-efficacy, statistics anxiety, attitude toward statistics, and social support. *Professional Counselor, 1*(2), 99-108. https://doi.org/10.15241/mpa.1.2.99

Pituch, K. A., & Stevens, J. P. (2015). *Applied multivariate statistics for the social sciences: Analyses with SAS and IBM's SPSS*. Routledge. https://doi.org/10.4324/9781315814919

Rahman, A., Abdulla, F., & Hossain, M. M. (2024). *Scientific data analysis with R: Biostatistical applications*. Chapman and Hall/CRC. https://doi.org/10.1201/9781003426189

Ramirez, C., Emmioglu, E., & Schau, C. (2010). *Understanding students' attitudes towards statistics: New perspectives using an expectancy-value model of motivation and survey of attitudes toward statistics* [Paper presentation]. The Joint Statistical Meetings Vancouver.

Salim, N. R., Gopal, K., & Ayub, A. F. M. (2018). The influence of statistics self-efficacy towards statistics engagement among undergraduate students. *AIP Conference Proceedings, 1974*(1), Article 050005. https://doi.org/10.1063/1.5041705

Sampson, A. S., Mutiu, B. M., & Udoh, A. U. (2023). Utilisation of statistical packages for data analysis among post-graduate students in universities in rivers state. *Journal of Education in Developing Areas, 31*(2), 352-366.

SAS Support. (2025). Usage note 22871: Types of logistic (or logit) models that can be fit using SAS®. *SAS Support*. https://support.sas.com/kb/22/871.html

Shrink, A. (2025,). Why SAS programming is challenging for students—And how to overcome it. *Medium*. https://medium.com/@alexshrink123/why-sas-programming-is-challenging-for-students-and-how-to-overcome-it-b41fc84acf72

Singh, S., Alam, M. N., Kaur, B., Kaur, K., Kaur, S., & Hossain, S. (2025). Comparative analysis of Apache Hadoop and Apache Spark for business intelligence. *AIP Conference Proceedings, 3224*(1), Article 020040. https://doi.org/10.1063/5.0246086

Soluade, O., Idowu, S. O., & Sofadekan, A. O. (2022). Attitudes of undergraduate teachers towards the application of statistical package for social science when teaching statistics in social studies. *Journal of Lexicography and Terminology, 6*(1), 134-144.

Şahin, M., & Aybek, E. (2019). Jamovi: An easy to use statistical software for the social scientists. *International Journal of Assessment Tools in Education, 6*(4), 670-692. https://doi.org/10.21449/ijate.661803

Tang, A. (2011). SocialCalc. In A. Brown, & G. Wilson (Eds.), *The architecture of open source applications–Elegance, evolution, and a few fearless hacks* (pp. 251-269).

Todorov, V. (2024). The R package ecosystem for robust statistics. *Wiley Interdisciplinary RE-Views: Computational Statistics, 16*(6), Article e70007. https://doi.org/10.1002/wics.70007

Vandever, C. (2020). Introduction to research statistical analysis: An overview of the basics. *HCA Healthcare Journal of Medicine, 1*(2), Article 71. https://doi.org/10.36518/2689-0216.1062

Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science, 46*, 186-204. https://doi.org/10.1287/mnsc.46.2.186.11926

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly, 27*(3), 425-478. https://doi.org/10.2307/30036540

Walker, E. R., & Brakke, K. E. (2017). Undergraduate psychology students' efficacy and attitudes across introductory and advanced statistics courses. *Scholarship of Teaching and Learning in Psychology, 3*(2), 132-140. https://doi.org/10.1037/stl0000088

Winters, R., Winters, A., & Amedee, R. G. (2010). Statistics: A brief overview. *Ochsner Journal, 10*(3), 213-216.

Wozney, L., Venkatesh, V., & Abrami, P. (2006). Implementing computer technologies: Teachers' perception and practices. *Journal of Technology and Teacher Education, 14*, 173-207.

Xulu, N., Abah, J. A., Chibisa, A. A., & Krishnannair, A. A. (2025). Effects of demographic factors on grade-12 mathematics learners' receptiveness to Siyavula educational application. *International Journal of Instructional Technology and Educational Studies, 6*(3), 1-16. https://doi.org/10.21608/ihites.2025.404601.1255

Yadegaridehkordi, E., Iahad, N. A., & Ahmad, N. (2015). User perceptions of the technology characteristics in a cloud-based collaborative learning environment: A qualitative study. *International Journal of Technology Enhanced Learning, 7*(1), 75-90. https://doi.org/10.1504/IJTEL.2015.071923

Yallati, R. M., & Joshi, S. (2020). Descriptive study of statistical software's as application tool of data management. *Indian Journal of Odyssey of Ayurved Research, 4*(2), 125-135.

Zavez, K., & Harel, O. (2025). Teaching statistical concepts using computing tools: A RE-View of the literature. *Journal of Statistics and Data Science Education*. https://doi.org/10.1080/26939169.2024.2445541

Zhang, X. H., Xie, F., Wee, H. L., Thumboo, J., & Li, S. (2008). Applying the expectancy-value model to understanding health values. *Value in Health, 11*(1), 61-68. https://doi.org/10.1111/j.1524-4733.2008.00368.x