Available online at www.koozakar.com/journals/ajspim

African Journal of Science Policy and Innovation Management

AJSPIM

Assessment of the Proficiency of Students in the use of Mathematical and Statistical Software Packages

Adeyemi O. Binuyo[1*], Folasade E. Akanbi[1], and Gbonjubola O. Binuyo[2]
[1]Ajayi Crowther University, Oyo, Nigeria
[2]Obafemi Awolowo University, Ile-Ife, Nigeria

[*]Corresponding author
Email: ao.binuyo@acu.edu.ng

Article information

https://doi.org/10.69798/66601355

ISSN (Online): 3066-3660
Copyright ©: 2026 The Author(s).

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International (CC-BY-4.0) License, which permits the user to copy, distribute, and transmit the work provided that the original authors and source are credited.

Published by Koozakar LLC. Atlanta GA 30350, United States

A Journal of the African Institute for Science Policy and Innovation, AISPI, Obafemi Awolowo University, Ile-Ife.

Edited by: Prof. Olawale Adejuwon

ABSTRACT

Proficiency with mathematical and statistical software packages is essential for solving complex problems, data analysis and interpretation in a variety of academic disciplines, especially for students enrolled in research-oriented courses. Despite their relevance, many students struggle to use these tools and packages due to lack of exposure, insufficient resources, inadequate training and availability of the software. This study identifies and assesses the socio-economic and institutional factors affecting students' proficiency in mathematical and statistical software packages. Data were collected using a structured questionnaire and administered to 100 randomly selected students of the Department of Statistics at The Polytechnic, Ibadan, Main Campus. Data collected were analyzed using descriptive and regression analysis. Statistical Package for Social Scientists (SPSS) software was used to conduct the statistical analysis. A mathematical model was formed from the data analysis via the use of the regression analysis. Results revealed that participants were familiar with the software packages, with slight variations in usage. SPSS was familiar to 25% of the participants, MATLAB (MATrix LABoratory) to 27%, and both Python and R to 24%. The frequency of statistical package usage further highlights gaps in proficiency. This was measured on a four-point Likert scale (Rarely = 1 to Always = 4). While 42% of participants used statistical packages "Often," only 17% reported "Always" using them. Others reported more sporadic usage, with 23% using them "Sometimes" and 18% "Rarely". Regarding the availability of adequate resources for learning mathematical and statistical software packages, 60% of participants indicated that resources were provided "Often" or "Always," while 40% reported "Rarely" or "Sometimes". From the findings, efforts should be made to motivate students by raising awareness of the importance of mathematical and statistical software skills for their academic and professional success. The institution's curriculum should be reviewed and aligned with industry demands, integrating mathematical and statistical software training as a core component.

Keywords: Proficiency, Regression analysis, Software packages, Mathematical modelling, Educational resources

INTRODUCTION

Solving complex mathematical problems and scientific data analysis aid in the creation of new information or a deeper understanding of natural events. Data analysis is where statistical software comes in handy (Hilary *et al.,* 2021). There are different forms of data analysis. Initially, it was paper and pen, punching machines, and simple, and complex scientific calculators. By advancement in technology, there are now executable mathematical and statistical software packages. Mathematical and Statistical software packages allow researchers to avoid routine mathematical mistakes and produce accurate figures in their research if all data are inputted correctly (Okaygbue *et al*., 2021). According to Abatan and Olayemi (2014), the emergence of mathematical and statistical software packages in the twenty-first century have helped different researchers in the physical and social sciences to improve the quality of research. Most renowned researchers are adopting these software packages in their data analysis and have been able to identify the immense contributions to research findings (Adetola, 2013). The software packages can be used to solve complex equations, analyze data and are capable of developing simple-to-understand solutions while maintaining test accuracy. Some best statistical analysis software packages provide capabilities such as regression analysis, significance tests, T-tests, F-tests, correlation, and statistical process control. Researchers, students, and teachers can use statistical analysis tools to make judgments based on thorough projections and outcomes. Statisticians now face large quantities of data, often in new forms like text or networks, and this data must be obtained—such as from Web services or databases—then managed, wrangled in complex ways, and visualized (Nolan and Temple Lang 2009).

Many statisticians now find themselves delivering not only analyses in the form of reports or presentations on some statistical analysis but also products that are used continually. To build and maintain these products, statisticians need additional skills. This lack of skills can lead to a variety of problems, including inappropriate application of statistical methods, inaccurate interpretation of outcomes, and failure to recognize or address the assumptions that underpin statistical models. These flaws can result in incorrect study findings, poor decision-making, and, eventually, a loss of credibility in the conclusions obtained. This work is significant because of its potential to increase the quality of data analysis and decision-making in a variety of areas. As data become more important in research, commercial strategy, and public policy, statistical analyses must be accurate and reliable. The study aims to identify common issues and faults including insufficient training and institutional factors, evaluate their influence on numerous fields, and provide ways for improving the use and proficiency of mathematical and statistical software packages. The study also intends to identify and assess the socio-economic factors affecting students' proficiency in mathematical and statistical software packages and examine the influence of institutions on students' proficiency in mathematical and statistical software packages.

Mathematical and Statistical Software Packages

Mathematical and Statistical Software Packages are suites of computer programs designed for performing complex mathematical and statistical calculations, data analysis and visualization. They are often used in fields like science, engineering and data sciences. Some popular examples of mathematical and statistical software packages include: SPSS, Excel, Stata, R, MATLAB, Mathematica, Maple, Origin, JMP, Minitab, and so on. These are further described below.

MATLAB: This is an acronym for **MAT**rix **LAB**oratory. It is a numerical computing environment and programming language, popular for data analysis, matrix manipulations, and plotting of graphs.

- **Mathematica:** A powerful software package for symbolic and numerical computation, visualization and programming, suitable for a wide range of mathematical tasks.

- **Excel**: Primarily a spreadsheet program, Excel offers a variety of statistical functions and tools for data analysis and visualization.

- **R**: A free and open-source programming language and software environment for statistical computing and graphics.

- **Minitab:** A statistical software package focused on quality improvement and data analysis, popular in business and education.

- **JMP**: A statistical software package from SAS institute, known for its interactive graphics and data visualization capabilities.

- **SPSS:** A statistical software package for data analysis, modelling and reporting, used in various fields like research, marketing and social sciences.

- **Maple:** A mathematical software package for symbolic computation and mathematical modelling, covering areas like algebra, calculus, and physics.

- **Origin**: A software package for data analysis, graphing and scientific visualization with features for signal processing, image processing and peak processing.

These software packages and many more provide tools to handle large datasets, perform statistical tests, create models, and visualize data, making it easier for users to analyze and interpret information. They streamline complex calculations, reduce the need for manual coding, and allow for efficient data exploration and analysis. Other benefits of mathematical and statistical software include: increased efficiency at work, more accurate data analysis, reduction in sampling error, and more data-driven decisions. They are used across various disciplines, including social sciences, business, finance, engineering and scientific research.

Mathematical Formulation: Model Parameters
Model specification: The model used in this study is multiple linear regression model. In this case, it looks at the effects or the relationship between a dependent (responsible) variable and number of independent (explanatory) variables. Regression analysis is the study of the nature and extent of relationship between two or more variables with a view of predicting the value of one variable from the other. It is used when the focus is on the relationship between a dependent variable and one or more independent variables. More specifically, regression analysis helps to understand how the typical value of the dependent variable changes when one of the independent variables varies, while the other independent variables are held fixed (Montgomery, 2005).

The models to be considered are therefore specified as:

$$P = \beta_{01} + \beta_{11}X_1 + \beta_{21}X_2 + \beta_{31}X_3 + \beta_{41}X_4 + \beta_{51}X_5 + \beta_{61}X_6 + e_{ij}. \quad (1)$$

$$P = \beta_{02}+\beta_{12}Y_1+\beta_{22}Y_2+\beta_{32}Y_3 + \beta_{42}Y_4+e_{ij}. \quad (2)$$

$$P = \beta_{03} + \beta_{13}Z_1 + \beta_{23}Z_2 + \beta_{33}Z_3 + \beta_{43}Z_4 + e_{ij}. \quad (3)$$

Letting Proficiency = P,
X_1 = sex (SEX)
X_2 = age (AGE)
X_3 = level of study (LOS)
X_4 = courses completed (CC)
X_5 = confidence level (CL)
X_6 = familiarity with software (FW)
X_7 = usage frequency (UF)
X_8 = challenges faced (CF)
X_9 = motivation level (ML)
…. **model 1**

Letting Proficiency = P,
Y_1 = Lecturer's enthusiasm for teaching (LE)
Y_2 = Lecturer's explanations of concepts (LEC)
Y_3 = Lecturer's accessibility for support (LAS)
Y_4 = Provision of practical examples (PPE)
Y_5 = Integration of statistical software into curriculum (ISS)
Y_6 = Staying updated with software developments (SU)
Y_7 = Support in overcoming challenges (SO)
Y_8 = Tailoring teaching to learning styles (TT)
Y_9 = Frequency of feedback (FF)
Y_{10}= Use of real-world applications (URWA)
…. **model 2**

Letting Proficiency = P,
Z_1 = Institution provides adequate resources (IAR)
Z_2 = Dedicated facilities (DF)
Z_3 = Accessibility of software (AS)
Z_4 = Workshops or additional training (WT)
Z_5 = Effectiveness of IT support (EITS)
Z_6 = Opportunities for collaboration (OC)
Z_7 = Integration into curriculum (IC)
Z_8 = Approach aligned with industry standards (AAIS)
Z_9 = Access to up-to-date resources (AUDR)
Z_{10} = Overall support for learning (OSL)
…. **model 3**

The models are re-specified as:

$$P = \beta_0 + \beta_1(Sex) + \beta_2(Age) + \beta_3(LOS) + \beta_4(CC) + \beta_5(CL) + \beta_6(FW) + \beta_7(UF) + \beta_8(CF) + \beta_9(ML) + eij \quad \textbf{Model 1}$$

$$P = \beta_0 + \beta_1(LE) + \beta_2(LEC) + \beta_3(LAS) + \beta_4(PPE) + \beta_5(ISS) + \beta_6(SU) + \beta_7(SO) + \beta_8(TT) + \beta_9(FF) + \beta_{10}(URWA) + eij \quad \textbf{Model 2}$$

$$P = \beta_0 + \beta_1(IAR) + \beta_2(DF) + \beta_3(AS) + \beta_4(WT) + \beta_5(EITS) + \beta_6(OC) + \beta_7(IC) + \beta_8(AAIS) + \beta_9(AUDR) + \beta_{10}(OSL) + eij \quad \textbf{Model 3}$$

Where; $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9, \beta_{10}, \beta_{11}$ and β_{12} are the regression coefficients which are estimated from the sample data. The e_{ij} is the random error term.

MATERIALS AND METHODS

Research Design

In this study a Descriptive Research Design was used. The one – group pretest and posttest design was utilised to determine the influence of mathematical and statistical software applications on students' attitude toward Statistics and performance in Statistics.

a) Research instrument: The data were collected using a questionnaire survey on factors affecting student's proficiency in mathematical and statistical packages. The questionnaire consists of five (5) sections. Section A for Student's Information, Section B for Impart of Continuous training, Section C for Lecturers' Related Factors, Section D for Institution Related Factors and Section E for a test evaluation.
b) Sampling Technique: Respondents in this study were students at The Polytechnic, Ibadan, Main Campus who are taking a statistics course. The sampling technique was done by using probability sampling which is simple random sampling technique. In all, 100 students served as respondents for the questionnaire administration.
c) Method of Data Analysis: All data collected were sorted out, edited and collated using a simple table in order to ease developing lassical insights on the data. Tables enable quicker interpretations and drawing help in meaningful conclusions. Statistical Package for Social Scientists (SPSS) software was used to do the statistical analysis.

REULTS

Table 1 shows the results of the socio-economic characteristics of the respondents of the study. The study sample consisted of 100 participants, with a higher proportion of males (57%) compared to females (43%). The majority of the participants fell within the 21 to 25 years of age group (43%), followed by those aged between 26 to 30 years old (35%). Only 19% were between 16 to 20 years of age and a small fraction (3%), were over 30 years old. Regarding their level of study, nearly half of the participants (45%) were in ND2, with 31% in HND1 and 24% in ND1.In terms of academic exposure to statistical packages, most participants had completed two or three related courses, with 28 and 31%, respectively, while 20% had completed just one course, and 21% had taken four courses or more. Participants also reported familiarity with statistical software, with slight variations in usage. SPSS was familiar to 25% of the participants, MATLAB to 27%, and both Python and R to 24% of the respondents. The frequency of statistical package usage further highlights gaps in proficiency. While 42% of participants used statistical packages "Often," only 17% reported "Always" using them. Others reported more sporadic usage, with 23% using them "Sometimes" and 18% "Rarely" Used them. The challenges associated with learning statistical packages were also notable. While 24% found them difficult to comprehend, 22% cited insufficient motivation, 29% attributed their struggles to a lack of time, and 25% expressed outright disinterest in using statistical packages. The current proficiency levels further reflected these challenges, with only 17% of participants rating their proficiency as "High," 23% as "Moderate," 26% as "Low," and 34% indicating they were "Not there" in terms of proficiency in using the packages (Table 1).

The impact of training on participants' proficiency in using statistical packages was analyzed and reported in Table 2. The perception of how training improves academic tasks reflected comparable diversity. While 42% of participants agreed or strongly agreed that training positively impacted their academic tasks, 40% held a contrary view, and 18% were neutral.

Table 1: Socio-economic Characteristics of the Respondents

Variable	Frequency	Percentage
Gender		
Male	57	57
Female	43	43
Age		
16 - 20 years	19	19
21 - 25 years	43	43
26 - 30 years	35	35
Over 30 years	3	3
Level of Study		
ND 1	24	24
ND 2	45	45
HND 1	31	31
Number of courses related to statistical packages completed		
1	20	20
2	28	28
3	31	31
4 and above	21	21
Statistical packages familiar with		
SPSS	25	25
MATLAB	27	27
PYTHON	24	24
R Package	24	24
Frequency of use of statistical packages		
Always	17	17
Often	42	42
Sometimes	23	23
Rarely	18	18
Challenges in learning statistical packages		
It is hard to comprehend	24	24
Not enough motivation	22	22
Lack of time	29	29
Not just interested	25	25
Current level of proficiency on statistical packages		
High	17	17
Moderate	23	23
Low	26	26
Not there	34	34
Each Variable Total	**100**	**100**

Confidence in applying skills gained through training was notably low. Only 32% of participants expressed agreement or strong agreement, while 47% disagreed or strongly disagreed, indicating a lack of confidence among many participants. Conversely, the perception that training improves performance was more positive, with 47% agreeing or strongly agreeing and only 35% disagreeing or strongly disagreeing. Training's impact on enhancing critical thinking revealed a similar pattern. While 36% of participants agreed or strongly agreed, 39% disagreed or strongly disagreed, and 25% remained neutral. Motivation levels after participating in training sessions were generally low, with 44% of participants expressing disagreement or strong disagreement and only 34% indicating agreement or strong agreement (Table 2).

Table 3 reveals the institutional factors affecting students' proficiency in statistical packages and both strengths and areas needing improvement. Regarding the availability of adequate resources for learning statistical packages, 60% of participants indicated that resources were "Often" or "Always" provided, while 40% reported "Rarely" or "Sometimes" Provided. The availability of dedicated facilities for practicing with statistical packages showed similar results, with 62% stating such facilities were "Often" or "Always" available, whereas 38% indicated more limited access.

Accessibility of statistical software for courses was perceived positively by 56% of participants, who reported "Often" or "Always" having access, while 44% reported "Rarely" or "Sometimes." Workshops or additional training on statistical packages were less common, with only 21% indicating "Always" and 38% selecting "Often" had additional training leaving a significant proportion (41%) reporting that these training opportunities were "Rarely" or "Sometimes" available. Opportunities for collaboration on projects involving statistical packages were rated positively by 59% of participants, while 41% indicated limited collaboration opportunities. Institutional efforts to integrate statistical packages into the curriculum were well-rated, with 61% of participants stating that integration "Often" or "Always" occurred, while 39% reported more limited integration. Similarly, 56% of participants felt their institution's approach to teaching statistical packages was aligned with industry standards, while 44% indicated that alignment was less consistent. The availability of up-to-date resources and literature on statistical packages was viewed positively by 58% of participants, who reported these resources were "Often" or "Always" available, while 42% selected "Rarely" or "Sometimes." When considering overall institutional support for learning statistical packages, 55% of participants indicated it was

Table 2: Impact of training on participant's proficiency in Mathematical and statistical package

Impact of Training	Strongly Disagree (%)	Disagree (%)	Neutral (%)	Agree (%)	Strongly agree (%)	Total (%)
Training improves academic tasks	22	18	18	22	20	100
Confidence in applying skills	27	20	21	18	14	100
Training improves performance	15	20	18	22	25	100
Training enhances critical thinking	15	24	25	12	24	100
Motivation after training	23	21	22	21	13	100
Training is valuable for success	18	20	22	22	18	100

Key: 1 = Strongly Disagree; 2 = Disagree; 3 = Neutral, 4 = Agree; 5 = Strongly agree

Table 3: Institutions Related Factors affecting participant's proficiency in Mathematical and statistical package

Institutions Related Factors	Rarely (%)	Sometimes (%)	Often (%)	Always (%)	Total (%)
Institution provides adequate resources for learning statistical packages	13	27	30	30	100
Institution dedicates facilities for practicing statistical packages	11	27	36	26	100
There is accessibility of statistical software for courses	14	30	28	28	100
There is Opportunities for collaboration on statistical package projects	12	29	27	32	100
Institution integrates statistical packages into curriculum	13	26	31	30	100
Institution's approach to teaching statistical packages aligned with industry standards	12	32	34	22	100
Access to up-to-date resources and literature on statistical packages	9	33	30	28	100
Overall support for learning statistical packages	16	29	34	21	100

Key: Rarely = 1, Sometimes = 2, Often = 3, Always =4

provided "Often" or "Always," while 45% felt support was limited to "Sometimes" or "Rarely." These findings indicate that while institutions have made progress in supporting students' proficiency in statistical packages, there remain gaps in providing comprehensive resources, training opportunities, and alignment with industry standards (Table 3).

The results of the theoretical knowledge test in Table 4 below shows that the majority of participants scored between 5 and 8, indicating a moderate level of proficiency in statistical packages. A small percentage of respondents achieved the highest score of 10, suggesting that only a few students possess a strong understanding of statistical software. Conversely, a notable proportion of participants scored below 5, highlighting a gap in foundational knowledge. The distribution of scores suggests that while most students have a basic grasp of statistical concepts, there is still room for improvement in their theoretical understanding.

Table 4: Theoretical Knowledge Test Result of Participants on the use of the software packages

Test Scores	Frequency	Percentage
2	1	1
3	8	8
4	10	10
5	13	13
6	20	20
7	15	15
8	12	12
9	12	12
10	9	9
Total	**100**	**100**

TEST OF HYPOTHESES

Hypothesis 1 (H_1): Socio-economic factors have a significant impact on students' proficiency in using statistical software.

The regression analysis for Hypothesis 1, which examines the impact of socio-economic factors on students' proficiency in statistical packages, shows that the model explains 18.2% of the variation in proficiency levels, as indicated by the R-squared value of 0.182. The adjusted R-squared value of 0.101 suggests that, after accounting for the number of predictors, the model still explains a modest portion of the variance in proficiency. The standard error of the estimate is 1.043, indicating the average deviation of observed proficiency levels from the predicted values. These results imply that socio-economic factors, including motivation level, number of courses completed, challenges faced, frequency of usage, gender, level of study, familiarity, confidence level, and age, have a measurable but limited impact on students' proficiency in statistical software (Table 5).

Table 5: Hypothesis testing 1

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	0.427[a]	0.182	0.101	1.043

a. Predictors: (Constant), Motivation Level, Courses Completed, Challenges, Usage Frequency, Sex, Level of Study, Familiar, Confidence Level, Age
b. Dependent Variable: Proficiency Level

Table 6 reveals the ANOVA results for Hypothesis 1, and it indicates that the regression model is statistically significant, with a p-value of 0.027. This suggests that socio-economic factors collectively have a significant impact on students' proficiency in statistical packages. The regression model accounts for a total sum of squares of 21.835, with 9 degrees of freedom, resulting in a mean square value of 2.426. The F-statistic of 2.231 indicates that the model explains a significant proportion of the variance in proficiency levels compared to the residual variance. The residual sum of squares is 97.875, indicating the proportion of variance not explained by the model. These findings confirm that socio-economic factors play a meaningful role in determining students' proficiency, though other unexplored variables may also contribute to proficiency levels.

Table 6: ANOVA for Hypothesis testing 1

ANOVA[a]

	Model	Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	21.835	9	2.426	2.231	.027[b]
	Residual	97.875	90	1.088		
	Total	119.710	99			

a. Dependent Variable: Proficiency Level
b. Predictors: (Constant)
Motivation Level, Courses Completed, Challenges, Usage Frequency, Sex, Level of Study, Familiar, Confidence Level, Age

Hypothesis 2 (H_2): Lecturer-related factors have a significant impact on students' proficiency in using statistical packages.

The results of the regression analysis for Hypothesis 2 indicate that lecturer-related factors have a weak and non-significant impact on students' proficiency in statistical software. The R-value of 0.225 suggests a weak positive relationship, while the R^2 value of 0.050 implies that only 5% of the variation in students' proficiency can be attributed to lecturer-related factors. The adjusted R^2 of -0.056 further highlights the lack of predictive power of the model. Additionally, the F-statistic value of 0.473 and the p-value of 0.904 confirm that the relationship is not statistically significant. These findings suggest that lecturer-related factors, as measured in this study, do not significantly influence students' proficiency with statistical software (Table 7).

Table 7: Hypothesis testing 2

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	0.225a	0.050	-0.056	1.130

a. Predictors: (Constant)
b. Dependent Variable: Proficiency Level.
Q10: Lecturer's use of real-world applications to make statistical packages relevant, Q5: Lecturer's integration of statistical software into curriculum, Q2: Lecturer's explanations of statistical concepts, Q6: Lecturer's staying updated with developments in statistical software, Q8: Lecturer's tailoring of teaching to accommodate learning styles, Q9: Lecturer's frequency of providing feedback on assignments, Q4: Lecturer's provision of practical examples/exercises, Q3: Lecturer's accessibility for questions/support outside class, CQ1:Lecturer's enthusiasm for teaching statistical packages, Q7: Lecturer's support in helping students overcome challenges

Table 8 shows the analysis of variance (ANOVA) results for Hypothesis 2, it indicates that the regression model examining the impact of lecturer-related factors on students' proficiency is not statistically significant. The regression sums of squares (6.036) and the mean square (0.604) resulted in an F-statistic value of 0.473, with a *p-value* of 0.904. Since the p-value exceeds the conventional threshold of 0.05, we conclude that lecturer-related factors do not significantly influence students' proficiency in statistical software. This further confirms the weak relationship suggested by the model summary.

Table 8: ANOVA for Hypothesis testing 2

ANOVA[a]

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	6.036	10	0.604	0.473	0.904b
	Residual	113.674	89	1.277		
	Total	119.710	99			

a. Dependent Variable: Proficiency Level
b. Predictors: (Constant)
Q10: Lecturer's use of real-world applications to make statistical packages relevant, Q5: Lecturer's integration of statistical software into curriculum, Q2: Lecturer's explanations of statistical concepts, Q6: Lecturer's staying updated with developments in statistical software, Q8: Lecturer's tailoring of teaching to accommodate learning styles, Q9: Lecturer's frequency of providing feedback on assignments, Q4: Lecturer's provision of practical examples/exercises, Q3: Lecturer's accessibility for questions/support outside class, CQ1:Lecturer's enthusiasm for teaching statistical packages, Q7: Lecturer's support in helping students overcome challenges

.

Hypothesis 3 (H_3): Institutional setup has a significant impact on students' proficiency in using statistical packages.

The regression analysis for Hypothesis 3 reveals that institutional setup has a weak and non-significant impact on students' proficiency in statistical software. The R-value of 0.267 indicates a weak positive relationship, and the R^2 value of 0.071 shows that only 7.1% of the variation in students' proficiency can be attributed to institutional factors. The adjusted R^2 of -0.033 further highlights the lack of predictive power in

Table 9: Hypothesis testing 3

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	0.267a	0.071	-0.033	1.118

a. Predictors: (Constant)
b. Dependent Variable: Proficiency Level
Q10: Overall support for learning statistical packages, QD1: Institution provides adequate resources for learning statistical packages, Q5: Effectiveness of institutional support services (IT helpdesk), Q4: Workshops or additional training on statistical packages, Q7: Institution integrates statistical packages into curriculum, Q2: Dedicated facilities for practicing statistical packages, Q8: Institution's approach to teaching statistical packages aligned with industry standards, Q9: Access to up-to-date resources and literature on statistical packages, Q6: Opportunities for collaboration on statistical package projects, Q3: Accessibility of statistical software for courses

the model. These results suggest that institutional setup factors, as measured in this study, do not meaningfully contribute to explaining variations in proficiency levels (Table 9).

The ANOVA results for Hypothesis 3 revealed in Table 10 below shows that the regression model examining the impact of institutional setup on students' proficiency is not statistically significant. The regression sum of squares is 8.538, with a mean square of 0.854, leading to an F-statistic value of 0.684 and a p-value of 0.737. Since the p-value is greater than the standard significance level of 0.05, it indicates that institutional setup factors do not significantly influence students' proficiency in statistical software. This aligns with the weak relationship observed in the model summary.

Table 10: ANOVA for Hypothesis testing 3

ANOVA[a]

	Model	Sum of Squares	Df	Mean Square	F	Sig.
	Regression	8.538	10	0.854	0.684	.737b
1	Residual	111.172	89	1.249		
	Total	119.710	99			

a. Dependent Variable: Proficiency Level
b. Predictors: (Constant)
Q10: Overall support for learning statistical packages, QD1: Institution provides adequate resources for learning statistical packages, Q5: Effectiveness of institutional support services (IT helpdesk), Q4: Workshops or additional training on statistical packages, Q7: Institution integrates statistical packages into curriculum, Q2: Dedicated facilities for practicing statistical packages, Q8: Institution's approach to teaching statistical packages aligned with industry standards, Q9: Access to up-to-date resources and literature on statistical packages, Q6: Opportunities for collaboration on statistical package projects, Q3: Accessibility of statistical software for courses

Model Testing

Hypothesis 1 (H_1): Socio-economic factors have a significant impact on students' proficiency.
The multiple linear regression model for Hypothesis 1, which tests the impact of socio-economic factors on students' proficiency in using statistical packages, is specified as:

$$P = 3.331 - 0.324(Sex) + 0.071(Age) + 0.002(Level\ of\ Study) + 0.091(Courses\ Completed) + 0.136(Confidence\ Level) - 0.194(Familiarity\ with\ Software) - 0.244\ (motivation\ level) + e_{ij} \quad (4)$$

where, *P* represents proficiency, and the independent variables include sex, age, level of study, courses completed, confidence level, and familiarity with software. The constant β_0 is 3.331, which represents the predicted proficiency level when all socio-economic factors are held constant.

The coefficients reveal that challenges faced by students (0.200) have the most significant and positive influence on proficiency, suggesting that overcoming challenges may enhance students' skills. Conversely, motivation level (-0.244) has the most substantial negative influence, indicating that lower motivation correlates with decreased proficiency. Familiarity with statistical software also has a notable negative effect (-0.194), though it is on the borderline of statistical significance with a *p-value* of 0.052. The only statistically significant predictors are challenges faced (p = 0.046) and motivation level (p = 0.019), indicating their strong influence on students' proficiency.

The R-squared value of 0.182 indicates that socio-economic factors explain 18.2% of the variance in students' proficiency levels. The adjusted R-squared value of 0.101 suggests that the model explains a modest proportion of the variance after accounting for the number of predictors. The ANOVA result, with an F-statistic of 2.231 and a p-value of 0.027, confirms that the overall model is statistically significant, implying that socio-economic factors collectively have a significant impact on students' proficiency in statistical packages.

Hypothesis 2 (H_2): Lecturer-related factors have a significant impact on students' proficiency.
The multiple linear regression model for Hypothesis 2, which examines the impact of lecturer-related factors on students' proficiency in statistical packages, is specified as:

P = 4.331 + 0.113 (*Lecturer's enthusiasm for teaching*) + 0.042 (*Lecturer's explanations of concepts*) - 0.004 (*Lecturer's accessibility for support*) - 0.082 (*Provision of practical examples*) + *eij* ______________________ (5)

where P represents students' proficiency, and the independent variables are lecturer-related factors including enthusiasm for teaching statistical packages, clarity of explanations, accessibility for support, and provision of practical examples. The constant β_0 is 4.331, indicating the predicted proficiency level when all these lecturer-related factors are held constant.

The coefficients reveal that lecturer enthusiasm (0.113) has minor positive influences on proficiency, while lecturer's accessibility for support exhibits the largest negative influences. However, none of these factors are statistically significant, as all *p-values* exceed 0.05.

The R-squared value of 0.050 shows that lecturer-related factors explain only 5% of the variation in students' proficiency levels, while the adjusted R-squared of -0.056 indicates that the inclusion of multiple predictors does not enhance the model's explanatory power.
The ANOVA result, with an F-statistic of 0.473 and a *p-value* of 0.904, confirms that the overall model is not statistically significant, suggesting that lecturer-related factors do not have a substantial impact on students' proficiency in statistical packages.

Hypothesis 3 (H_3): Institutional setup has a significant impact on students' proficiency.
For Hypothesis 3, which tests whether institutional setup has a significant impact on students' proficiency in statistical packages, the multiple linear regression model is specified as:

P = 3.282 + 0.000(*Institution provides adequate resources*) - 0.062(*Dedicated facilities*) + 0.039(*Accessibility of software*) + 0.174(*Workshops or additional training*) + *eij*__________________________________ (6)

where P represents proficiency, and the independent variables include institution-related factors such as provision of adequate resources, dedicated facilities, accessibility of statistical software, and workshops or training. The constant β_0 is 3.282, indicating the predicted proficiency level when all institutional factors are held constant.

The coefficients reveal that workshops or additional training (0.174) and accessibility of software (0.039) have the most positive influence on proficiency, while dedicated facilities (-0.062) exhibit the most notable negative influences. However, none of these factors are statistically significant, as all p-values are greater than 0.05.

The model's R-squared value of 0.071 indicates that institutional-related factors explain only 7.1% of the variance in students' proficiency levels. The adjusted R-squared of -0.033 suggests that the inclusion of these predictors did not improve the model's explanatory power. The ANOVA result, with an F-statistic of 0.684 and a p-value of 0.737, shows that the overall model is not statistically significant, implying that institutional factors do not have a substantial impact on students' proficiency in statistical packages.

DISCUSSION

The findings of this study provide valuable insights into the factors influencing students' proficiency in mathematical and statistical software packages, including socio-economic, lecturer-related, and institutional variables. The significant impact of socio-economic factors on students' proficiency underscores the critical role of intrinsic motivation, confidence, and academic exposure. The regression analysis revealed that, socio-economic factors explained 18.2% of the variance in proficiency, highlighting their importance as foundational determinants. This aligns with Abatan and Olayemi (2014) who emphasised that, access to statistical education and confidence-building initiatives significantly enhance student competency.

Furthermore, Adetola (2013) supports this result by asserting that individuals' belief in their ability to succeed in specific tasks directly impacts their performance. The descriptive results highlight significant challenges among students. Previous research by Nolan and Temple Lang (2009) noted that, despite the increased accessibility of mathematical and statistical software, many students struggle with higher-order analytical skills, often relying on surface-level familiarity with tools like SPSS. This finding emphasizes the need for pedagogical strategies that move beyond rote learning and focus on cultivating problem-solving and critical reasoning skills. Also, institutional setup, encompassing factors such as resource availability, IT support, and curriculum alignment with industry standards, demonstrated a weak and non-significant impact on proficiency. Despite positive descriptive ratings, the regression analysis did not show a meaningful contribution of these factors to students' skill development. The lack of workshops and limited hands-on opportunities may have contributed to this result. This finding aligns with Mendezabal and Tindowen (2017), who argued that institutional support must be holistic and proactive. Merely providing access to resources without ensuring their alignment with industry demands or fostering a culture of continuous learning limits their overall effectiveness. Additionally, workshops or specialized training sessions, which were found to be lacking in this study, have been shown by Bandura (1977) to significantly enhance students' ability to apply statistical tools in practical contexts.

The theoretical knowledge test results further underscore these challenges, with most participants achieving only moderate scores of between 5 and 8. This supports findings by Baumer *et al.* (2015). who highlight a persistent gap in translating theoretical understanding into practical competency among students. The low levels of motivation observed in this study—where 57% of participants reported being either "Not motivated" or "Not just interested"—further exacerbate this issue. According to Sakamaki *et al.* (2022) motivation acts as a critical mediator between institutional support and individual achievement, underscoring the need for institutions to adopt strategies that actively engage and inspire students. Taken together, the results suggest that while socio-economic factors play a measurable role in shaping proficiency, the impact of institutional factors remains constrained by systemic gaps in implementation. The reliance on traditional teaching methods and the limited availability of workshops highlights the need for a more integrated approach to mathematical and statistical education. Bandura (1977) proposed that embedding statistical software education within interdisciplinary projects can bridge the gap between theoretical and applied learning. Similarly, fostering peer collaboration and mentorship, as suggested by Bandura (1997) can enhance students' confidence and deepen their engagements with mathematical and statistical tools.

CONCLUSION

The study demonstrated that socio-economic factors significantly influence students' proficiency in mathematical and statistical software, with key determinants including motivation, confidence, and academic exposure. While these factors collectively explained a modest proportion of the variation in proficiency levels, they highlight the critical role of individual attributes in shaping students' abilities. In contrast, lecturer-related and institutional factors did not show a statistically significant impact on proficiency. Although descriptive data suggested positive perceptions of teaching efforts and institutional support, these measures were insufficient to drive measurable improvements in proficiency. This points to a disconnect between the resources provided and their effectiveness in fostering practical skill development. The findings further revealed moderate proficiency levels among students, as evidenced by the theoretical knowledge test results. This underscores the need to bridge the gap between theoretical understanding and hands-on application of statistical tools. Overall, the study highlights the importance of aligning teaching practices, institutional support, and student engagement strategies to better address the practical demands of statistical education.

RECOMMENDATIONS

Based on the findings of this study, the following recommendations are proposed to enhance students' proficiency in mathematical and statistical software packages:

i. Increase practical training opportunities: Institutions should organize regular hands-on workshops and training sessions focused on statistical software. These sessions should emphasize practical applications to help students bridge the gap between theoretical knowledge and real-world usage.
ii. Enhance lecturer engagement: Lecturers should adopt interactive teaching methods, such as problem-based learning and collaborative projects, to actively involve students in the learning process. Providing detailed feedback and mentoring can also strengthen students' confidence and skills.
iii. Improve institutional support: Institutions must ensure the availability of up-to-date statistical software, dedicated computer labs, and comprehensive support services such as IT helpdesks. Collaborations with industry professionals can also help students gain exposure to practical, industry-standard applications.
iv. Motivate students through awareness and incentives: Efforts should be made to motivate students by raising awareness of the importance of statistical software skills for their academic and professional success. Incentives such as certifications for completing workshops or outstanding performance can further encourage participation.

Develop tailored curriculum: The curriculum should be reviewed and aligned with industry demands, integrating statistical software training as a core component. This includes embedding practical assignments and projects that require students to apply statistical tools to real-world problems.

REFERENCES

Abatan, A., & Olayemi, O. (2014). The Role of Statistical Software in Data Analysis: A twenty-first-century perspective. *International Journal of Applied Research and Studies, 3*(8) https://doi.org/10.2139/ssrn.1234567

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

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

Baumer, E. P. S. & Çetinkaya-Rundel, M. (2015). Problem-based learning in statistics: Enhancing critical thinking and technical proficiency. *Journal of Statistical Education, 23*(2), 1–20. https://doi.org/10.1080/10691898.2015.11889757

Hilary I. O., Pelumi E. O., Emmanuela C. O. and Elvir M. A., (2021). Trends and usage pattern of SPSS and Minitab Software in Scientific Research. *Journal of Physics: Conference Series 1734* (1): 012017-2021. International Conference on Recent Trends in Applied Research, 2020, 14-15 August, 2

Mendezabal, M. J. N., & Tindowen, D. J. C. (2017). Institutional factors affecting students' statistical proficiency. *International Journal of Educational Research*, *10*(2), 50–68.

Montgomery D. C. (2005). A Supplement for Using JMP (R). SAS Institute Inc., Cary, North Carolina, USA.

Nolan, D. & Temple L. D. (2009). Computing in the statistics curricula. The American Statistician, 63(2), 89–97. https://doi.org/10.1198/tast.2009.06134

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

Sakamaki, K., Taguri, M., & Akimoto, Y. (2022). Experience of distance education for project-based learning in data science. *International Journal of Statistics and Data Science Education*, 5(1), 22–35. https://doi.org/10.1007/s42081-022-001542.