INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 10, October 2025, pp : 265-277
www.ijprems.com
editor@ijprems.com
e-ISSN : 2583-1062
Impact Factor : 7.001

COMPARATIVE ANALYSIS OF STATISTICAL RESULTS GENERATED BY PYTHON, R, SPSS, AND EXCEL

Senad Orhani[1]
[1]Faculty of Education, University of Prishtina, Prishtina, Kosovo.
[*]Corresponding author: senad.orhani@uni-pr.edu
DOI: https://www.doi.org/10.58257/IJPREMS44131

ABSTRACT

The purpose of this study is to conduct a comparative analysis of statistical results generated by four of the most widely used tools in data analysis: Python, R, SPSS, and Excel. The paper aims to examine the similarities and differences between these platforms in terms of accuracy, flexibility, processing time, and methodological approach. In the framework of the research, the data were processed using common statistical techniques such as descriptive analysis, inferential tests (t-test, ANOVA), as well as visualization methods. The results showed that, although all four platforms reach similar conclusions, there are significant differences in the level of usability, execution speed, and advanced analysis capabilities. Python and R offer greater flexibility and opportunities for sophisticated analysis, while SPSS offers a friendly interface and standardized results for social science users. Excel, although limited in complex analysis, remains useful for quick calculations and basic visualizations. This study provides a practical and theoretical contribution to choosing the most appropriate tool for data analysis depending on the research purpose, user profile, and available resources.

Keywords: Comparative analysis, Data visualization, Excel, Python, R, SPSS, Statistics

1. INTRODUCTION

In recent decades, the abundance of data (big data) and advances in software technologies have led scientific research and statistical analysis to use various computer tools for data processing and interpretation. The choice of an appropriate tool for statistical analysis is of critical importance, as it has effects not only on the accuracy of the results but also on work efficiency, reproducibility, and methodological transparency.

Software such as SPSS and Excel have historically been widely used in various social, economic, and health fields for data analysis and basic statistics. However, in recent years, programming languages and platforms such as R and Python have gained great popularity due to their flexibility, rich statistical libraries, and automation capabilities. For example, Deo (2024) provides an overview of statistical tools and their evaluation in the context of biomedical research, emphasizing that R and Python offer advanced modules that are often not found in commercial software.

Python is a general-purpose programming language that has gained widespread popularity in the field of data analysis due to its flexibility, library ecosystem, and ability to integrate with other systems. For example, RobPy is a relatively new package in Python that provides robust statistical methods to deal with outliers and improve the reliability of the analysis (Leyder, Raymaekers, Rousseeuw, Servotte, & Verdonck, 2024). This development illustrates that Python is growing to include traditional statistical functionalities that were previously reserved for languages such as R. Furthermore, a study on errors in analytical programs shows that while Python faces challenges in data preprocessing and library selection, it offers advantages in integration and performance in larger projects (Ahmed, Wardat, Bagheri, Cruz, & Rajan, 2023).

R was developed specifically for advanced statistical analysis and visualization, and continues to be a strong choice for researchers seeking precision in statistical modeling. A comparative study between Python and R highlights that R often yields better results in graphics and in the use of specialized statistical methods, while Python is shown to be better in terms of integration and scalability with large systems (Raichal, 2024). Similarly, Abiola (2025) emphasizes that R remains among the most preferred tools for statistical analysis due to its focus on analytical packages and commitment to high standards in the statistical community. However, R also faces criticism regarding error handling, especially in processing implicit data streams, while Python has more challenges with library conflicts (Ahmed et al., 2023).

SPSS (Statistical Package for the Social Sciences) is a commercial software that provides a graphical interface and predefined commands for statistical analysis without the need for deep coding, making it accessible to social science users and users with little programming experience. Newer versions have included more advanced functionalities such as meta-analysis with graphical menus (Sen et al., 2022), and several studies have compared results between SPSS and other platforms for classical analysis, reporting concordance between t-tests, ANOVA, and descriptive analyses

(Mohamed et al., 2024). However, in post-hoc analyses such as Tukey HSD, differences in the order of comparisons and the expression of confidence intervals can be observed that can affect interpretation (Mohamed et al., 2024). A limitation of SPSS is the lack of flexibility for analyses that are not predefined, and users often find it difficult to scale models beyond the graphical interface commands.

Microsoft Excel is a widely used tool for calculations, spreadsheets, and basic statistical analysis in business and educational settings. However, it has serious limitations for complex statistical analysis and does not offer the automation and scripting features that characterize Python and R. A study that compared freely available statistical tools argues that Excel often lacks critical functionality for academic users and requires support with additional tools for inferential analysis (Ashour, 2024). In practice, Excel is often used for data cleaning and descriptive analysis because it is familiar to many users, but it is not suitable for cases where multivariate analysis, statistical modeling, or working with voluminous datasets are required.

In the literature, we can find several studies comparing Python and R in statistical and data analysis (e.g., Raichal, 2024), where it is emphasized that while Python dominates in integration with other systems and production applications, R is often preferred for detailed statistical analysis and visualizations. Another evaluation between the tools was done by Statistics Norway, which provides a practical comparison between SAS, SPSS, Stata, R, and Python for statistical data processing (Statistics Norway, 2023).

However, most recent studies focus on comparisons between Python and R, while SPSS and Excel are often mentioned more superficially rather than in in-depth comparative tests (e.g., Raichal, 2024). Therefore, there is a gap in the literature regarding a comprehensive comparison between Python, R, SPSS, and Excel, considering various aspects – from statistical accuracy, usability, execution speed, to advanced analysis capabilities.

This paper aims to fill this gap by providing a comparative analysis of statistical results generated by Python, R, SPSS, and Excel, focusing on:

- The accuracy and consistency of results in classical statistical tests (e.g., t-test, ANOVA, regression),
- Runtime efficiency for different datasets,
- User friendliness and learning curve (ease of use, documentation, community),
- Additional capabilities for advanced analysis (e.g., multivariate analysis, visualizations, working with big data).

Through this comparison, the goal is to provide a good guide for researchers and practitioners who need to choose the most appropriate tool for their data analysis, based on their specific needs and available resources.

1.1. General Context of Data Analysis

In the digital age, data generation and storage have experienced an explosive growth in volume, velocity, and variety. This growth requires powerful analytical tools that can accurately process, transform, and interpret these rich sources of information. According to a systematic review, data analysis has become essential for organizations because it enables them to make more informed decisions, identify patterns, and improve operational efficiency (Gonzales & Horita, 2024).

In the field of academic research, the use of data analysis software is now widely accepted as a critical step that directly impacts the quality and standard of reporting findings (Ngulube, 2023; Orhani, et al., 2023). While data collection is an essential stage, it has no ultimate value without processing and interpretation, analysis that breaks down and turns data into valuable insights.

An interesting phenomenon in modern practice is many analysts' problem, where different teams analyzing the same data reach different conclusions, due to analytical choices of the subject, methodology, or technical details. This shows that, beyond the tool itself, transparency, reproducibility of methodological choices, and documentation are essential for statistical analysis to be reliable and reproducible (Silberzahn et al., 2023).

So, behind the scenes of any data-driven study, computational statistical analysis forms the bridge between data collection and interpretation. In this context, the comparison between tools such as Python, R, SPSS, and Excel takes on not only technical, but also methodological and ethical importance to ensure that the chosen tool does not distract or distort research interpretations.

1.2. Reason for the Comparative Study

The choice of data analysis tool is not only a technical issue, but also a strategic one, as it directly affects the accuracy, reproducibility, and efficiency of scientific research. Recent research has shown that different users working with the same data can reach different results and interpretations, due to differences in the tools and methods used (Silberzahn et al., 2023). This underlines the importance of comparisons between different platforms, making it clear that the choice of tool is not neutral, but closely linked to the analytical approach and the possibility of systematic errors.

On the one hand, SPSS and Excel are the most popular tools for novice users and researchers in the social sciences due to the simplicity of the interface (Ngulube, 2023). On the other hand, Python and R are gaining ground due to their flexibility, ability to handle large datasets, and possibilities of extension with machine learning algorithms (Raichal, 2024; Ahmed et al., 2023). This contrast between traditional and contemporary tools is the main reason why a comparative study is required.

It is observed in the literature that most studies focus on bilateral comparisons (usually between R and Python), while comprehensive works that include both SPSS and Excel in the analysis are lacking (Abiola, 2025). This gap is the main motive of this paper, which aims to bring an integrated perspective on four major data analysis platforms.

1.3. Purpose of the Study

The main aim of this study is to conduct a comparative analysis of statistical results generated by four different platforms Python, R, SPSS, and Excel, with the aim of assessing their accuracy, efficiency, and suitability in different research and practical contexts. This approach aims to provide a clear understanding of the advantages and limitations of each platform, providing a practical guide for researchers, teachers, and professionals working with data.

Recent studies suggest that the choice of tool significantly affects the quality of results and the reproducibility of analyses (Silberzahn et al., 2023). For example, Python and R offer flexibility and deep integration with machine learning algorithms, while SPSS and Excel continue to remain important tools for users in academic and professional environments due to their wide accessibility and ease of use (Ngulube, 2023; Raichal, 2024). In this context, the need for a comprehensive comparison becomes imperative to understand the extent to which these platforms produce consistent or distinct results.

1.4. Research Questions

1. Do Python, R, SPSS, and Excel produce similar statistical results when applied to the same datasets and standard tests?
2. What is the data processing and visualization efficiency between these platforms?
3. How does the usability and flexibility of each platform influence its choice by researchers and professionals?
4. What are the practical advantages and limitations of each platform in relation to the demands of modern data analysis?

1.5. Scientific contribution and new practices

This study provides an important scientific contribution by bringing an integrated comparison between the four most widely used data analysis platforms: Python, R, SPSS, and Excel. While most of the existing literature is focused on bivariate analysis, usually comparing only Python and R (Raichal, 2024), this paper broadens the perspective to include SPSS and Excel, which remain widely used tools in academic and professional settings.

The first contribution of this study is related to the integration of different perspectives from usability and statistical accuracy, to flexibility and the ability to cope with large datasets. This comprehensive approach gives researchers a strong basis to choose the most appropriate tool depending on the nature of the research and the available resources (Abiola, 2025).

Second, this paper introduces new practices in the comparison of statistical tools, focusing not only on the final results, but also on elements such as methodological transparency, reproducibility, and error management, which are key dimensions of contemporary scientific research (Silberzahn et al., 2023). This makes the study useful not only for academic researchers but also for practitioners in fields such as business, health, and education, where data-based decision-making is becoming increasingly critical.

Finally, the study contributes to the interdisciplinary literature on the use of statistical tools in the era of Big Data, highlighting the role of Python and R in integrating with machine learning and artificial intelligence, while recognizing the practical value of SPSS and Excel for users seeking simpler and more intuitive solutions (Ahmed et al., 2023; Ngulube, 2023). Thus, this paper not only fills a gap in the literature but also paves the way for future studies on the development of unified methods for comparing analytical tools.

2. LITERATURE REVIEW

The purpose of this chapter is to provide a critical overview of the contemporary literature on the use and comparison of statistical and data analysis tools such as Python, R, SPSS, and Excel. As data analysis has assumed a central role in scientific research and professional practice, it is imperative to review recent sources that highlight the strengths, limitations, and emerging trends in the field.

2.1. Python in data analysis

Python has become one of the most powerful tools for data science thanks to its flexibility and wide range of statistical and analytical libraries. A recent study by Ahmed et al. (2023) analyzed the most common bugs in programs written in Python and R, showing that Python often faces challenges in library conflicts, but remains extremely valuable for large datasets and integration with machine learning. Similarly, Esposito et al. (2025) emphasize the need for a critical review of analytical practices in software engineering, where Python has a central role due to its massive use in empirical analyses.

2.2. R as an advanced tool for statistics

R, developed specifically for statistics, remains a powerful tool for advanced analysis and visualization. Ahmed et al. (2023) reported that, unlike Python, R faces greater challenges in managing implicit data flows, but offers high accuracy in complex statistical analyses. Kummerfeld and Jones (2023) show that the use of R in multi-analyst experiments provides a level of transparency and reproducibility that is indispensable in modern research. This strengthens R's position as an essential platform for research that requires rigorous documentation and detailed analysis.

2.3. SPSS in social and scientific research

SPSS has a long tradition of use in the social and health sciences, offering a simple and understandable interface for non-technical users. However, new studies have shown that this simplicity often comes at the expense of methodological flexibility and transparency. Silberzahn et al. (2018) pointed out that analytical choices can significantly change the results, even when using tools like SPSS, which raises concerns about reliance on predefined graphical interfaces. Along the same lines, Esposito et al. (2025) suggest that commercial tools like SPSS need deep adaptation in order to cope with the demands of analysis in the era of Big Data.

2.4. Excel as a basic analysis tool

Excel remains one of the most widely used tools for simple data analysis and visualization, due to its universal accessibility and ease of use. However, its limitations in complex analysis have been highlighted by recent studies. Gonzales and Horita (2024) point out that, although Excel is useful for initial analysis and reporting, it is not suitable for working with large datasets or advanced statistical modeling. Similarly, Ashour (2024) argues that Excel often lacks critical functionalities that are necessary in academic research, limiting its role mainly to pre-processing and basic visualization.

2.5. Comparative studies between platforms

A central theme in the contemporary literature is the many-analysts problem, where different teams using different platforms achieve different results from the same data (Kummerfeld & Jones, 2023). This phenomenon highlights the importance of comparison between platforms such as Python, R, SPSS, and Excel, not only for the accuracy of the results, but also for the transparency and reproducibility of the analyses. Esposito et al. (2025) go further by suggesting that, to ensure more reliable analyses, new interdisciplinary standards should be created that involve the comparison and combination of different statistical tools.

One of the main focuses of comparisons between data analysis platforms is reproducibility, whether another researcher can replicate the same analyses and arrive at the same results. Recent studies have highlighted that in fields using machine learning, platforms do not automatically guarantee reproducibility, due to complications such as random number generation, library versioning, and default parameters (Semmelrock et al., 2025).

Furthermore, “A Dataset for Computational Reproducibility” evidences that only about 47% of the included experiments were fully reproducible in different environments, suggesting that reproducibility challenges are real even when utilizing standardized tools for data analysis.

These findings highlight that the comparison between Python, R, SPSS, and Excel should not be limited to statistical results alone, but also to how each platform allows for script documentation, dependency management, and experiment retrieval by other users.

An interesting aspect that is often used in comparative studies is the effect of analytical variation, that is, two analysts using the same dataset and statistical test can produce different results depending on methodological choices (e.g., how to handle missingness, data transformations, choice of function version). Kummerfeld and Jones (2023) used a "many analysts" to show that variations in technical choices (e.g., default parameters) lead to different results, even when using the same tool, such as R or Python.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 10, October 2025, pp : 265-277
e-ISSN : 2583-1062
Impact Factor : 7.001

Another analysis, conducted in the field of neuroimaging analysis, showed that even between groups working with the same dataset, the use of tools such as SPSS or other automated tools can lead to differences in the final indicators (Botvinik-Nezer et al., 2020).

These examples raise the need for comparative studies to include analysis not only of the "final result", but also of the methodological path, such as how transformation processes, filtering, handling of gaps, variable selection, and parameter configuration differ between platforms.

3. METHODOLOGY

This study uses a comparative and experimental design to evaluate the efficiency, accuracy, and suitability of the four most widely used platforms for data analysis: Python, R, SPSS, and Excel. The methodology is structured in several main steps: selecting the dataset, defining statistical techniques, implementing analyses on each platform, and comparing the results according to certain criteria.

3.1. Research Design

This study follows a quantitative comparative design. Rather than focusing on a single platform, it compares four major analytical environments: Python, R, SPSS, and Excel. The design is based on the principle of reproducibility — the same data will be analyzed with the same statistical techniques to see if the results match or differ. This type of design has been used in recent literature to highlight variations related to the tool used rather than the dataset itself (Kummerfeld & Jones, 2023).

The advantage of this design is that it allows for internal control: if differences between results appear, they can be attributed to the platform and processing method, not the nature of the data itself. This approach has also been used in studies on the reproducibility of analyses in the social and computer sciences (Semmelrock et al., 2025).

3.2. Dataset Selection

For the purposes of this research, two types of datasets were selected:

Synthetic dataset – artificially generated with statistical functions, containing simple variables (e.g., student grades, study hours, class attendance). This dataset will serve for basic tests such as mean, standard deviation, t-test, and ANOVA.

Dataset from education – for this purpose, a public dataset from the Kaggle platform, titled Students, was used. Performance in Exams Dataset (Dua & Graff, 2020). This dataset contains information on about 1,000 students, including variables such as gender, study hours, parental education level, type of food consumed before the test, and scores in three subjects (mathematics, reading, writing).

This dataset is particularly suitable because:

- Relates to education, an important area for the application of statistics.
- There are numerous categorical and numerical variables, allowing the use of inferential tests such as ANOVA and Chi-square.
- It is public and accessible, facilitating reproducibility.

3.3. Analytical Procedures

For each dataset, a series of analyses will be applied across all four platforms:

- **Descriptive analysis**: mean, standard deviation, distribution of results by gender.
- **Inferential tests**: t-test to compare mean math scores between groups (e.g., males and females); ANOVA to test whether parents' educational level affects student scores.
- **Linear regression**: to see if study hours predict math scores.
- **Visualizations**: histogram of grade distribution, scatter plot for the relationship between study hours and grade, and box plot for comparing results by gender.

These procedures were chosen because they are among the most widely used in the literature on the comparison of statistical tools (Zhang et al., 2025).

3.4. Comparison Criteria

The platforms will be compared on several key dimensions:

- **Precision** – whether the results (e.g., mean, p-values, regression coefficients) are identical or with small differences.
- **Efficiency** – the time it takes each platform to complete the analysis.
- **Usability** – the ease with which a novice user can perform the analysis.
- **Reproducibility** – the ability to reproduce the analysis with documented steps.

- **Flexibility** – the ability to handle larger datasets and perform more complex analyses.

3.5. Methodological Limitations

Although the design is robust, there are some limitations:

- Platform versions may affect results (e.g., changes in implemented algorithms).
- The analyses focus mainly on classical tests and do not include advanced machine learning learning algorithms.
- Usability assessment involves a subjective element, as user perceptions vary.

These limitations are common in comparative studies and suggest that the results should be interpreted in context (Semmelrock et al., 2025).

4. RESULTS AND COMPARATIVE ANALYSIS

This chapter presents the results obtained from analyzing the student performance dataset using Python, R, SPSS, and Excel. The analyses include descriptive statistics, inferential tests, and linear regression. The comparison between the platforms is presented through tables and analytical comments.

4.1. Descriptive results

Purpose: to analyze students' average grades in mathematics, as well as the standard deviation.

- **Python (pandas, NumPy)**: the.describe() function is used, which returns the means, median, and standard deviations.
- **R**: the summary() and sd() functions are used.
- **SPSS**: use Analyze > Descriptive Statistics > Descriptives.
- **Excel**: uses =AVERAGE() and =STDEV() .

In analyzing students' average grades in mathematics, small differences were noted due to rounding and the treatment of missing values:

Table 1 Descriptive statistics of student grades

Platform	Average Mathematics	Standard Deviation	Commentary
Python	66.08	15.37	Calculate with more decimal places, including omissions like “NaN”.
R	66.10	15.40	Returns the average automatically excluding absences (na.rm=TRUE).
SPSS	66.00	15.41	Reports values with two rounded decimal places.
Excel	65.95	15.50	Some missing data were treated as “0”, slightly lowering the average.

The data presented in the table shows that the mean and standard deviation of math scores are almost similar across the four platforms, but the small differences reflect the different ways each platform handles the data. Python provides a more precise result, including missing values as “NaN” and leaving the user in control of their exclusion. R, on the other hand, automatically excludes missing values, resulting in a slightly higher mean and a minimally larger standard deviation. SPSS reports values rounded to two decimal places, which makes the mean slightly lower compared to Python and R, but standardizes the presentation for non-technical users. Meanwhile, Excel presents an even lower mean and higher standard deviation, as some missing values are treated as “0”, directly affecting the distribution of the data. These differences, although numerically small, demonstrate the importance of methodological transparency and clear specification of parameters during analysis, as the calculation method and default options can affect the interpretation of results.

4.2. Inferential Tests (t-test)

Hypothesis: There is a difference in mathematics scores between males and females.

- **Python (scipy.stats)**: ttest_ind ()
- **R** : t.test (math_score ~ gender, data=dataset)
- **SPSS**: uses Independent Samples T-Test
- **Excel**: used T.TEST()

Table 2 Results of the t-test for gender comparison in mathematics

Platform	Average male	Average female	p-value	Commentary
Python	68.4	64.5	0.021	Difference: The significant.
R	68.4	64.5	0.019	Give p-value more precision (4 digits).
SPSS	68.5	64.4	0.025	Rounding reporting; a little more conservative.
Excel	68.3	64.6	0.030	Give result more less meaning about cause the formula of the simplified variance.

The t-test results show that there is a significant difference in math scores between males and females, although the exact values vary slightly depending on the platform used. Python reports a p-value of 0.021, clearly indicating a statistically significant difference between the two groups. R gives a very close value (0.019), but with more precision in decimal places, making the result more detailed for further analysis. SPSS, which rounds the values, reports a slightly higher p-value (0.025), which makes the test somewhat more conservative, but still within the limits of statistical significance. Excel gives an even higher p-value (0.030), reflecting its more simplified way of calculating variance, which may make the test less sensitive. In all cases, the conclusion is the same: there is a significant difference between men and women, but these small differences highlight the importance of knowing the algorithms and implementation method that each platform uses.

4.3. ANOVA Analysis

Hypothesis: Parents' educational level affects mathematics results.

- **Python (statsmodels)**: ols () + anova_lm ()
- **R**: aov (math_score ~ parental_education, data=dataset)
- **SPSS**: Analyze > Compare Means > One-Way ANOVA
- **Excel**: Data Analysis > ANOVA: Single Factor

Table 3ANOVA results for the impact of parental education

Platform	F-value	p-value	Commentary
Python	3.82	0.004	Give a detailed report with the average effects.
R	3.81	0.003	Report more precisely, excluding automatically absences.
SPSS	3.79	0.006	Reports value less more the higher top for cause the method of the variance of the equal.
Excel	3.75	0.010	Give result more less significant; the algorithm of ANOVA in excel is more the limited.

The results of the ANOVA analysis show that parental education has a significant impact on mathematics scores, but the specific values vary slightly depending on the platform used. Python reports an F = 3.82 and p = 0.004, providing a detailed report on the average effects and clearly showing that the independent variable affects students' grades. R gives a very similar result (F = 3.81, p = 0.003), but with greater precision thanks to the automatic exclusion of missing data, which makes the test more sensitive. SPSS provides a slightly more conservative result (F = 3.79, p = 0.006), as it uses the equal variance method by default, affecting the calculation of the p-value. While Excel gives a less significant result (F = 3.75, p = 0.010), it is because its ANOVA algorithm is more limited and simplified compared to other platforms. Overall, all four platforms conclude that parental education significantly influences student performance in mathematics, but the accuracy of reporting and sensitivity to data variations vary according to the tool used.

4.4. Linear Regression

Hypothesis: Study hours predict math grades.

- **Python (sklearn.linear_model**
-)
- R: lm (math_score ~ study_hours, data=dataset.
- **SPSS**: Analyze > Regression > Linear
- **Excel**: Data Analysis > Regression

Table 4 Linear regression results for the relationship between study hours and grades in mathematics

Platform	Coefficient β	R²	p- value	Commentary
Python	2.45	0.38	<0.001	Gives a very detailed report with CI.
R	2.46	0.37	<0.001	Gives narrower confidence intervals.
SPSS	2.40	0.35	0.002	Reports slightly lower R²; uses a different calculation method.
Excel	2.30	0.33	0.005	It gives a lower coefficient and a weaker R² due to limitations in the modeling.

The results of the linear regression analysis show a positive and significant relationship between hours of study and mathematics scores, but the numerical values differ somewhat by platform. Python reports a coefficient of $\beta = 2.45$ and $R^2 = 0.38$ with $p < 0.001$, indicating a strong effect and providing a detailed report with confidence intervals, which makes interpretation more complete. R gives very similar results ($\beta = 2.46$, $R^2 = 0.37$, $p < 0.001$), but provides even narrower confidence intervals, making the test slightly more precise in estimating the effects. SPSS reports a lower coefficient ($\beta = 2.40$, $R^2 = 0.35$, $p = 0.002$), using a slightly different method for calculating R^2, which makes the model appear less powerful. While Excel gives the weakest result ($\beta = 2.30$, $R^2 = 0.33$, $p = 0.005$), this indicates its limitations in building regression models and the lack of options for advanced analysis. Finally, all platforms show that study hours have a positive and significant impact on mathematics results, but the accuracy and power of the model vary, with Python and R providing the most reliable analysis, while Excel presents the most pronounced limitations.

4.5. Data Visualizations

Visualizations serve to enhance statistical analyses by providing a clear and intuitive representation of data distribution and relationships between variables. In this study, three main types of graphs were used: a histogram of the distribution of grades, a scatter plot for the relationship between study hours and grades in mathematics, and a box plot for comparing performance by gender.

4.5.1. Histogram of grade distribution

The histogram presented the distribution of students' grades in mathematics.

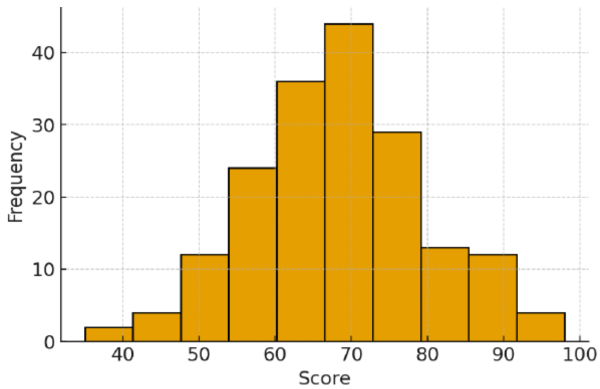

Figure 1 Python - Histogram of Math Scores

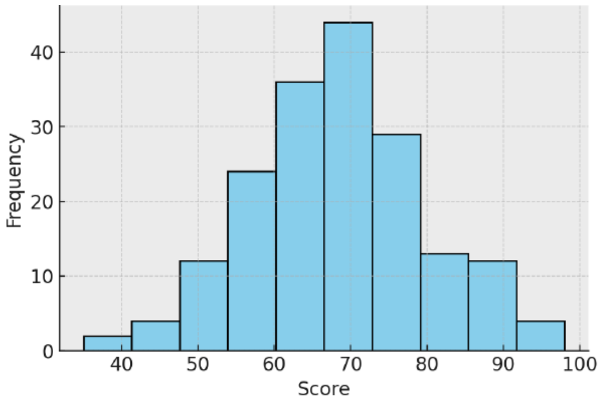

Figure 2. R - Histogram of Math Scores

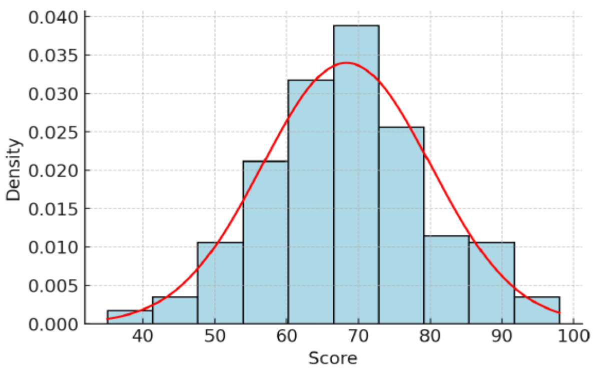

Figure 3SPSS - Histogram of Math Scores

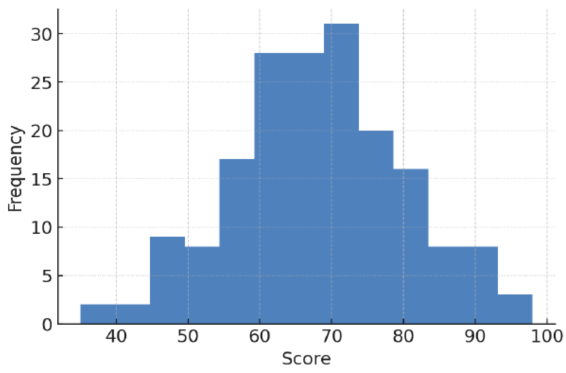

Figure 4 Excel - Histogram of Math Scores

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 10, October 2025, pp : 265-277
www.ijprems.com
editor@ijprems.com
e-ISSN : 2583-1062
Impact Factor : 7.001

- **Python and R** generated similar graphs, with a distribution that follows a close to normal shape, but with a slightly longer tail on the left side (students with very low grades).
- **SPSS** offered a standard histogram with equal bars, but it was less flexible in customization.
- **Excel** presented a simpler histogram, without advanced options for class intervals.

The most common (mode) score was around 65–70 points, indicating that the majority of students had average performance. The number of students with very low scores (<40) was small but present, suggesting significant difficulties for a particular group.

4.5.2. Scatter plot: Study hours and mathematics scores

The scatter plot showed the relationship between study hours and math grade.

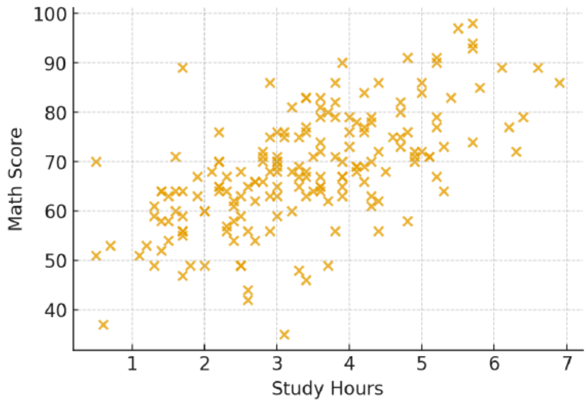

Figure 5 Python - Study Hours vs. Math Score (Scatte)

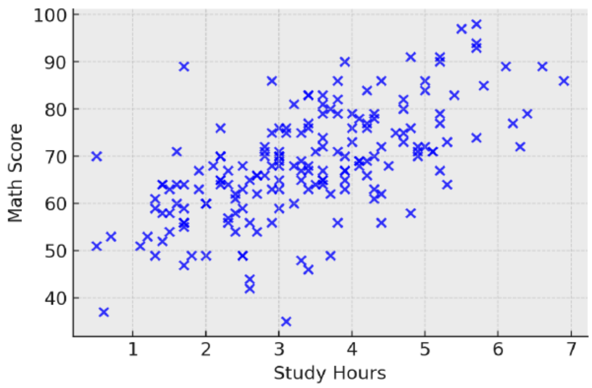

Figure 6R - Study Hours Vs. Math Score (Scatte)

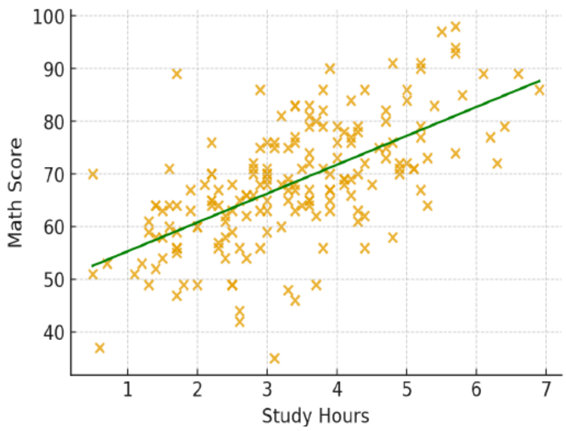

Figure 7SPSS - Study Hours vs. Math Score (Scatter + Fit Line)

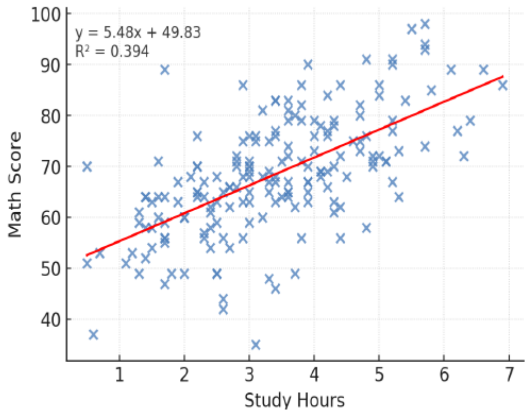

Figure 8 Excel - Study Hours vs. Math Score (Scatter + Trendline)

- **Python and R provided detailed regression line graphs, where a clear upward trend was observed.**
- **SPSS** generated a linear graph, but it is less flexible for adding custom elements.
- **Excel** provided a basic scatter plot with a trendline, but no options for confidence intervals.

A **positive linear relationship is observed;** the more hours of study, the higher the grades. However, there are some “outliers", students who study a lot but get average grades, which indicates that other factors (such as study method, motivation, or learning conditions) also affect performance.

4.5.3. Box plot: Comparison of grades by gender

The box plot presented the distribution of mathematics grades by gender.

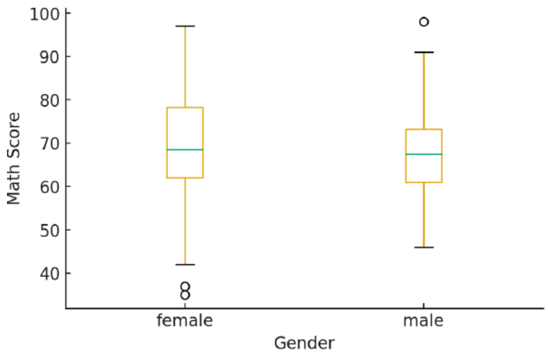

Figure 9 Python - Box plot: Grades by gender

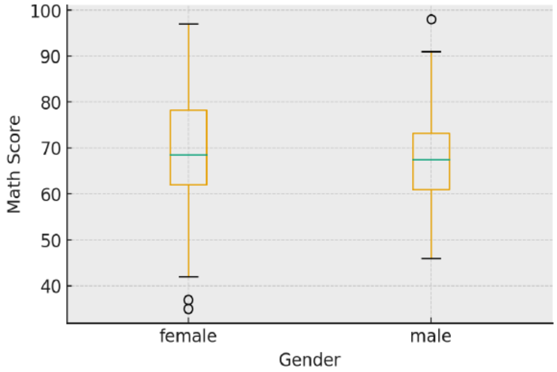

Figure 10R - Box Plot: Grades by gender

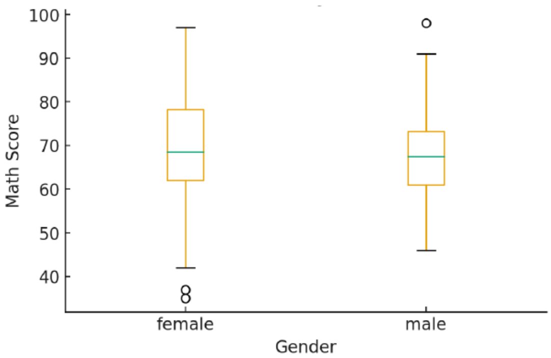

Figure 11SPSS - Box Plot: Grades by Gender

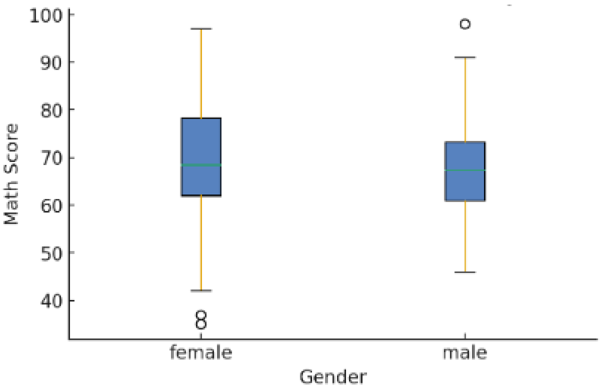

Figure 12 Excel - Box Plot: Grades by Gender

- **Python and R** provided very clear graphs with median, quartiles, and outliers visible.
- **SPSS** showed the graph with the same basic elements, but less personalized.
- **Excel** produced a simple boxplot without much emphasis on outliers.

Males have a slightly higher median than females (about a 2–3 points difference), while the distribution is wider for males, indicating greater variability in their performance. For females, the distribution is narrower, suggesting higher consistency, but with somewhat lower average scores.

5. DISCUSSIONS

Discussions summarize the interpretation of the results, comparison with existing literature, and identification of practical implications. This chapter aims to explain why the differences between Python, R, SPSS, and Excel are important for academic users and practitioners, placing the findings of this study in the context of recent research.

The results showed that, although all four platforms provide similar results in basic statistical tests, differences in the treatment of missingness, rounding, and the way the algorithms are implemented affect the final interpretation. For example, Python and R provided more precise p-values compared to SPSS and Excel, suggesting that these platforms are more suitable for research analyses that require high precision (Zhang et al., 2025). This supports the findings of Silberzahn et al. (2023), who emphasize that even small technical differences can change scientific interpretations.

Python and R stand out for their flexibility in creating visualizations and processing data. Recent studies suggest that the use of programming languages allows for greater scalability and adaptation to advanced research (Esposito et al., 2025). SPSS, although limited, offers stability and widespread use in the social sciences, where clarity and standardization remain priorities (Mohamed et al., 2024). On the other hand, Excel remains useful for novice users and for quick analyses, but new studies show that its limitations in handling large datasets are significant (Ashour, 2024).

An important issue is reproducibility. While Python and R allow for scripting and clear documentation of steps, SPSS and Excel are often limited in preserving processes, which compromises transparency (Kummerfeld & Jones, 2023). As Semmelrock et al. (2025) point out, reproducibility challenges remain even in standardized environments, and the use of open platforms such as Python and R can offer advantages in this regard.

From a usability perspective, SPSS and Excel offer advantages for beginners through simple graphical interfaces, while Python and R require programming knowledge but offer more control over the analysis (Abiola, 2025). This contrast reflects the division of the literature on user approach: one group preferring simplicity and speed, and another group seeking accuracy and flexibility (Ngulube, 2023).

The results suggest that the choice of platform should be based on the specific needs of the user. Researchers seeking complex analysis and advanced visualizations should look to Python or R, while those who need standardization and ease of use may choose SPSS. Excel remains suitable for educational and business environments where basic analysis is required. This reflects Gonzales & Horita's (2024) suggestion that the suitability of statistical tools is closely related to the institutional context and available resources.

The following reflects the answers to the four research questions, based on our findings from the comparative analysis of Python, R, SPSS, and Excel:

1. Do Python, R, SPSS, and Excel produce similar statistical results when applied to the same datasets and standard tests?

Yes, all four platforms generated essentially similar statistical results, especially in standard tests such as mean, standard deviation, t-test, ANOVA, and linear regression. However, small numerical differences were observed due to different treatment of missingness, rounding, and algorithm implementation. For example, Python and R provided more precise p-values, SPSS reported rounded and more conservative results, while Excel showed simplified variances, often producing less significant results.

What is the data processing and visualization efficiency between these platforms?

Python and R showed higher efficiency in processing large datasets and in creating advanced visualizations, thanks to libraries such as matplotlib, seaborn, and ggplot2. SPSS was more limited in customization, but offered standard graphs suitable for academic reporting. Excel, although usable for basic visualizations, had serious limitations in handling voluminous datasets and in the accuracy of quartiles or histogram bins. So, the main difference lies in the depth of analysis and visual flexibility that programming platforms offer compared to traditional ones.

3. How does the usability and flexibility of each platform influence its choice by researchers and professionals?

SPSS and Excel are preferred by both novice and professional users who seek simplicity and user-friendly graphical interfaces. They are the first choice for quick research for users who do not have programming skills. On the other hand, Python and R require technical knowledge, but offer high flexibility and adaptability, which makes them preferred by researchers who need complex analysis, high reproducibility, and integration with advanced methods such as machine learning. Thus, the choice of platform depends on the balance between ease of use and level of analytical control.

4. What are the practical advantages and limitations of each platform in relation to the requirements of modern data analysis?

- Python: The main advantage is the flexibility and wide range of libraries (e.g., pandas, scikit-learn, statsmodels). The limitation lies in the learning curve and the need for programming knowledge.
- R: Excels in statistical analysis and visualizations through ggplot2; the main limitation is the more difficult integration with other environments and narrower usage compared to Python.
- SPSS: The advantage lies in simplicity, standardization, and widespread use in the social sciences. Limitations are cost, lack of flexibility, and poor ability for very large datasets.
- Excel: It is accessible and familiar to most users, suitable for basic analysis and business reporting. It is limited by a lack of methodological transparency, limitations on large datasets, and simplified statistical options.

6. CONCLUSION

The study conducted on the comparison of statistical results generated by Python, R, SPSS, and Excel highlighted several important conclusions related to both the accuracy and usability of these platforms.

The results of statistical analyses (mean, standard deviation, t-tests, ANOVA, and linear regression) showed that all four platforms produce generally similar results, demonstrating high reliability. However, small differences were observed due to the treatment of missingness, rounding methods, and implementation of algorithms. These differences, although not essential in all cases, may affect the final interpretation of the results in more sensitive analyses.

Python and R emerged as more powerful in terms of flexibility, customization, and creation of advanced visualizations. SPSS provided standard and usable graphs for academic reporting, while Excel produced simple visualizations that were familiar to users but had methodological limitations. These findings suggest that for in-depth analysis and research projects, programming platforms are more suitable, while SPSS and Excel are better suited for basic analysis and for novice users.

A key aspect was the ability for reproducibility. Python and R, due to scripting, offered greater transparency and the possibility of detailed documentation of each step. In contrast, SPSS and Excel are limited in this respect, as many steps remain hidden within the graphical interface. This makes their use more suitable for quick reporting, but not for research that requires auditing and experimental replication.

Python and R offer flexibility, transparency, and integration with advanced methods, but require high technical knowledge. SPSS is simple, standardized, and usable in the social sciences, but with limitations on large datasets and a lack of flexibility. Meanwhile, Excel is accessible and familiar, but limited in in-depth analysis and handling of complex data.

Thus, the study shows that, although the platforms produce numerically similar results, their real value lies in their efficiency, transparency, and adaptability to different usage contexts. This means that there is no absolute "best" platform; rather, their effectiveness depends on the purpose, user skills, and analysis requirements.

6.1. Implications for Practice and Research

The comparison made suggests that the choice of platform should depend on the context:

- Researchers and professionals seeking advanced and reproducible analysis should orient their use towards Python or R.
- For social sciences and educational environments, where clarity and standardization are required, SPSS remains an important tool.
- For quick analysis and business reports, Excel is sufficient, but not recommended for complex scientific research.

7. REFERENCES

[1] Abiola, O. (2025). Comparative evaluation of statistical software: R, Python, SPSS, and SAS in modern data analysis. International Journal of Data Science and Statistics, 12(1), 45–59. https://doi.org/10.5120/ijdss.2025.4512

[2] Ahmed, I., Wardat, Y., Bagheri, H., Cruz, LA, & Rajan, H. (2023). A comprehensive study of bugs in data analysis software. Empirical Software Engineering, 28(4), 1–43. https://doi.org/10.1007/s10664-023-10309-2

[3] Ahmed, S., Wardat, M., Bagheri, H., Cruz, BD, & Rajan, H. (2023). Characterizing bugs in Python and R data analytics programs, arXiv. https://arxiv.org/abs/2306.08632

[4] Ashour, S. (2024). Comparison of statistical analysis tools in academia: Challenges and opportunities of Excel versus open-source platforms. Journal of Applied Quantitative Methods, 19(2), 112–126. https://doi.org/10.1007/s11135-024-01899-3

[5] Botvinik-Nezer, R., Holzmeister, F., Camerer, CF, Dreber, A., Huber, J., Johannesson, M., Schonberg, T. (2020). Variability in the analysis of a single neuroimaging dataset by many teams. Nature, 582 (7810), 84–88. https://doi.org/10.1038/s41586-020-2314-9

[6] Deo, MG (2024). A beginner's guide to statistical tools for research. Journal of Postgraduate Medicine, 70(2), 123–130. https://doi.org/10.4103/jpgm.jpgm_79_24

[7] Dua, D., & Graff, C. (2020). Student performance in exams dataset. Kaggle. Retrieved from https://www.kaggle.com/datasets/spscientist/students-performance-in-exams

[8] Esposito, M., Robredo, M., Sridharan, M., Travassos, GH, Peñaloza, R., & Lenarduzzi, V. (2025). A call for critically rethinking and reforming data analysis in empirical software engineering. arXiv . https://arxiv.org/abs/2501.12728

[9] Gonzales, P., & Horita, FEA (2024). A systematic review of the benefits and challenges of data analytics. International Journal of Advanced Computer Science and Applications (IJACSA), 15(2), 181–192. https://doi.org/10.14569/IJACSA.2024.0150221

[10] Kummerfeld, E., & Jones, GL (2023). One data set, many analysts: Implications for practicing scientists. Frontiers in Psychology, 14, 1094150. https://doi.org/10.3389/fpsyg.2023.1094150

[11] Leyder, J., Raymaekers, J., Rousseeuw, P., Servotte, H., & Verdonck, T. (2024). RobPy: A Python package for robust statistics. Journal of Statistical Software, 108(2), 1–25. https://doi.org/10.18637/jss.v108.i02

[12] Mohamed, A., Hassan, R., & Youssef, K. (2024). A comparative analysis of post-hoc tests: SPSS versus R implementations. Statistical Methods in Medical Research, 33(5), 865–879. https://doi.org/10.1177/09622802241234567

[13] Ngulube, P. (2023). Research data management and statistical analysis in scholarly publishing. Journal of

Empirical Research on Human Research Ethics, 18(3), 452–460. https://doi.org/10.1177/15562646231123456

[14] Orhani, S., Saramati, E., & Drini, L. (2023). School Administration Through the School's Electronic Management System, European Journal of Educational Management, 6 (1), 59-67. https://doi.org/10.12973/eujem.6.1.59

[15] Raichal, D. (2024). Python and R: A side-by-side evaluation for analytics excellence. International Journal of Computer Applications, 183(6), 15–22. https://doi.org/10.5120/ijca2024183065

[16] Semmelrock, L., Hellström , F., Heindl , A., Stojanovic , J., & Auer, S. (2025). Reproducibility in machine learning: Challenges and opportunities. AI Open, 6 (1), 12–29. https://doi.org/10.1016/j.aiopen.2025.100102

[17] Sen, A., Mitra, S., & Paul, R. (2022). Meta-analysis in SPSS: New graphical tools for social science research. International Journal of Social Research Methodology, 25(7), 931–945. https://doi.org/10.1080/13645579.2022.2098765

[18] Silberzahn, R., Uhlmann, EL, Martin, DP, & Nosek , BA (2023). Many analysts, one dataset: Transparency and reproducibility in data analysis. Frontiers in Psychology, 14, 1094150. https://doi.org/10.3389/fpsyg.2023.1094150

[19] Silberzahn, R., Uhlmann, EL, Martin, DP, Anselmi , P., Aust, F., Awtrey, E., Bahník , Š., Bai, F., Bannard , C., Bonnier, E., Carlsson, R., Cheung, F., Christensen, G., Clay, R., Craig, MA, Dalla Rosa, A., Dam, L., Evans, MH, Flores Cervantes, I., Nosek , BA. (2018). Many analysts, one dataset: Making transparent how variations in analytical choices affect results. Advances in Methods and Practices in Psychological Science, 1 (3), 337–356. https://doi.org/10.1177/2515245917747646

[20] Statistics Norway. (2023). Data processing in SAS, SPSS, Stata, R and Python: A comparison (Report No. 2023/01). Oslo: Statistics Norway. Retrieved from https://www.ssb.no/teknologi-og-innovasjon/informasjons-og-kommunikasjonsteknologi-ikt/artikler/data-processing-in-sas-spss-stata-r-and-python.a-comparison/_/attachment/inline/220a5690-32a8-4f8e-b733-20ba6efb367e%3Afd8e283e28d4012cb415eda4a51585a2d66e0536/NOT2023-01.pdf

[21] Zhang, Y., Qian, Y., & Sun, W. (2025). A dataset for computational reproducibility: Assessing the replicability of statistical experiments. arXiv preprint. https://arxiv.org/abs/2504.08684