

Research Article

From Calculators to Datastatpro: The Progression of Data Analysis in the Digital Era

Asif Hanif^{1*}, Tahira Ashraf², Shahzad Ahmad^{3*}, Ali Saad Tariq⁴, Shazia Iqbal⁵, Zohra Shaukat⁶

^{1*}Prof. Department of Biostatistics, Faculty of Medicine, Sakarya University, Sakarya-Türkiye.

²Assistant Professor, Department of Biostatistics, Faculty of Medicine, Sakarya University, Sakarya-Türkiye.

^{3*}Senior Lecturer, Faculty of Medicine and Health Sciences, the University of Buckingham, Buckingham, United Kingdom.

⁴BDS, Dip. OMSF. Assistant Professor, Department of Oral Biology, Islam Dental College, Sialkot, Pakistan.

⁵Senior Lecturer / Unit Lead, Faculty of Medicine and Health Science - The University of Buckingham, UK.

⁶MBBS. Independent Researcher, Vancouver, British Columbia, Canada.

Corresponding Author: Dr. Asif Hanif^{1*}, Shahzad Ahmad^{3*}

^{1*}Department of Biostatistics, Faculty of medicine, Sakarya University, Sakarya-Türkiye.

^{3*}Senior Lecturer, Faculty of Medicine and Health Sciences, the University of Buckingham, Buckingham, United Kingdom.

Email: ^{1*}asifhanif@sakarya.edu.tr, ²tahiraashraf@sakarya.edu.tr,
^{3*}shahzad.ahmad@buckingham.ac.uk, ⁴alisaadtariq@gmail.com, ⁵Iqbalian2002@hotmail.com,
⁶zohra_shaukat@yahoo.com

Received: 22.05.26, Revised: 20.06.26, Accepted: 24.07.26

ABSTRACT

Introduction: Since the inception of manual calculation with the help of mechanical computing machines, major technological advancements have transformed the field of data analysis, including the emergence of standalone digital analytical platforms such as DataStatPro. The purpose of this paper is to examine the progression of data analysis from manual and mechanical approaches to modern digital and AI-assisted analytical systems.

Methods: This paper uses a historical narrative review approach to synthesise key developments in data analysis technologies, including mechanical computation, personal computers, statistical software, open-source tools, big data platforms and artificial intelligence-assisted systems. As this is a conceptual review, no primary data collection or inferential statistical testing was performed.

Results: Initial statistical problems and data processing were done by hand, and early computational machines were used to solve automated computations with low processing depth and speed. This was transformed with the advent of advanced computers and then personal computers, such as tablets and laptops. It is now possible to use powerful software such as the Statistical Analysis System (SAS), Statistical Package for Social Sciences (SPSS), now known as Statistical Product and Service Solutions, and Stata to conduct richer and more efficient analyses. The pace of this growth has been compounded by recent advancements in artificial intelligence, big data technology and open-source tools, which have made it possible to automate analysis, improve reproducibility and support intricate modelling.

Conclusions: Modern data analysis has shifted from isolated computational procedures to integrated end-to-end analytical systems that meet the needs of modern data-driven research and decision-making. AI-assisted analytical systems make use of automation, learning algorithms and subject expertise to accelerate data processing with reduced human error. They enhance decision-making by transforming complex datasets into reproducible, interpretable and actionable insights.

Keywords: Data Analysis, Artificial Intelligence, Datastatpro, Statistical Software, Digital Analytics, Data Science.

INTRODUCTION

Data analysis has evolved from early civilizations' record-keeping to formal statistical theory in the 17th and 19th centuries to real-world applications in the 20th century. People like Graunt, Bayes, Gauss, Nightingale, and

Fisher made significant contributions that influenced contemporary statistical theory and experimental design. Big data, machine learning, and data science were introduced by the digital revolution, while the development of computers and statistical software (SS)

democratized analysis. AI, IoT, real-time analytics, and ethical frameworks are all continuing to expand the application and significance of data analysis in the twenty-first century. (GeeksforGeeks) This article provides a narrative overview of the historical evolution of statistical data analysis tools, recent advances in artificial intelligence-assisted analytics, and the emerging role of integrated analytical platforms in medical and public health research.

Methods

This article was designed as a historical narrative review examining the progression of data analysis from manual and mechanical calculation methods to contemporary statistical software, open-source programming environments, big data systems, artificial intelligence-assisted analytics and integrated platforms such as DataStatPro. The review setting was the use of data-analysis tools and statistical software within medical, public health, biomedical and wider research contexts. Relevant literature and source material were identified through a purposive narrative search of peer-reviewed articles, review papers, software-related publications, historical and technical sources, and available platform documentation. Sources were selected if they contributed to understanding the historical, methodological or practical development of data-analysis tools. The material was synthesised chronologically and thematically. The comparison of statistical tools and DataStatPro features was descriptive, and platform-based information on DataStatPro was interpreted as descriptive rather than independent validation evidence.

The Historical Evolution of Data Analysis

Over the past 60 years, data management and statistical analysis have evolved dramatically from manual paper-based calculations to advanced computational methods. This transformation has largely been driven by developments in information technology, particularly computer hardware and statistical software. During the 1950s and 1960s, researchers such as Fisher used mechanical calculators and early computational systems for statistical analyses. Later developments introduced software packages such as SAS, Genstat, and SPSS, followed by personal computer-based programs such as JMP and Stata. The emergence of R around 2000 initiated a new era of extensible, open-source, and code-based statistical computing,

significantly influencing modern research and data science practices. (Team, 2021)

Big Data and Cloud-Based Analytics

The use of SS in data analysis becomes increasingly important as quantitative research expands. Manual analysis using paper is giving way to more effective digital/electronic analysis using SS, according to researchers. (Abatan and Olayemi, 2014) According to a study, there has been a noticeable transition from manual to digital data analysis, highlighting the increasing significance of SS in quantitative research. According to results from a cross-sectional study of academic staff, the majority of respondents were unable to do analyses without SS, and there was strong agreement (80%) that SS has a favorable impact on research findings. Despite the fact that many of the tools depicted vary in terms of their relevance and adoption to a particular study, SPSS was identified as the most frequently used tool. In general, the research makes it clear that SS is important for increasing the quality and the effectiveness of research. (Abatan and Olayemi, 2014) It further discusses the process of utilizing cloud resources by the mobile cloud computing to circumvent the constraints of devices and highlights the importance of increasing use of mobile devices in the daily lives of people. It talks about how the big data analytics that are defined by volume, diversity, velocity, and veracity enable value extraction from healthcare data. The research also reported the available big data tools and applications besides examined cloud-based and cloudlet-supported networked healthcare infrastructures. Overall, it highlights the potential of mobile cloud computing and big data analytics in the development of future-proof healthcare. Despite these technological advancements, AI-assisted analytical systems also introduce important challenges related to algorithmic bias, data privacy, reproducibility, cybersecurity, and over-reliance on automated interpretations. Ethical considerations and appropriate human oversight remain essential in ensuring the responsible use of AI-driven analytical platforms in healthcare and research. (Lo'ai et al., 2016)

Diversification of Statistical Software and Open-Source Platforms

There are currently about 125 recognized statistical and data-analysis software packages available on open-source, public-domain, freeware, proprietary, and add-on platforms. Their development shows how data science,

machine learning, and domain-specific analytics have replaced traditional statistical inference techniques. Many modern technologies are now based on open-source ecosystems, especially R and Python, which allow for quick methodological advancement. A few well-known programs, such as SAS, SPSS, Stata, and Minitab, continue to rule applied research despite this diversity. It's interesting to note that Excel is still the most popular SS globally, particularly outside of academic settings, despite the availability of hundreds of specialist tools. (Wikipedia, 2025) Strong disciplinary trends in SS use are revealed by an analysis of metadata from more than 10,000 articles; Stata dominates economics, R leads statistics, and political science is expanding quickly. These days, a lot of social science studies use several software tools in one manuscript. Additionally, students expressed increased knowledge of software diversity in quantitative research and better comprehension of research procedures. (Upton et al., 2025)

User-Friendly Statistical Tools in Health Research

Another study emphasizes the importance of statistical analysis in experimental medical research, however researchers often find it challenging to analyze data due to software complexity. It looks at seven easy-to-use, free statistical programs suitable for medical research and outlines their features, benefits, and drawbacks. All of the evaluated tools offer basic statistical analysis, however JASP, Jamovi, and BlueSky Statistics are better suited for more sophisticated methods. G*Power is the suggested tool for determining sample size and power. (Ashour, 2024)

Several free and commercial SS that can be utilized in health research include data management, estimation of sample size, randomization, power analysis, and analysis. These SS are usually affected by cost, usability,

credibility, scope of functionality and capabilities such as audit trails and validation. (Deo and Ranganathan, 2024)

Artificial Intelligence and Data Analysis in Healthcare

This article is about the evaluation as well as development of the Medsensio Screening Tool, a medical platform that integrates medically validated risk prediction algorithms of heart diseases into its functionality. The system was more acceptable and trusted by the users, comparing to the web-based alternatives, and allowed patients to participate and even narrow the screening. The ability of large language models (LLMs) to generate personalized recommendations motivated and boosted the user confidence. On the whole, the findings demonstrate that screening techniques that are based on AI and combined can enhance the prevention of heart diseases and early diagnosis. (Baniya, 2024)

Artificial intelligence and machine learning are increasingly used in medical practice because it can assist doctors in various activities by analyzing clinical, laboratory, and imaging data. These tools are more productive by reducing the load of performing routine recording and interpretation. Obtaining quality medical data that is applicable in training the algorithm is among the hardest jobs and consumes a lot of time on the part for professionals. Medical machine-learning in the future could be enhanced by the unified system of data storage and more effective training of AI. (Lo'ai et al., 2016) Appropriate governance, transparency, and ethical oversight are essential when implementing AI-assisted analytical systems in healthcare and biomedical research. Figure 1 illustrates the progression of data analysis technologies from manual calculation methods to AI-assisted analytical platforms such as DataStatPro.



Fig-1: The Data Analysis Progression in the Digital Era; From Calculators to Datastatpro

DataStatPro as an Integrated Analytical Platform

The approach to data management and evaluation by academics has thus been transformed in the radical manner due to the alteration from manual coding and heterogeneous data analysis to cohesive analytical frameworks. Traditional processes increase the risk of errors and require a significant amount of the research work as they often include long-term coding, regular cleaning of data, and meticulous upkeep in multiple software systems. On the other hand, the DataStatPro simplifies this task because it offers guided data editing, structured administration, and built-in analysis paths which reduce the dependency on complex programming. This change enables researchers, especially in the domains of public health and medicine, to concentrate more on study design, interpretation, and decision-making and less on technological challenges. DataStatPro represents an example of emerging integrated analytical platforms designed to improve workflow efficiency, reduce technical barriers, and support structured data analysis in medical and public health research. (DataStatPro, 2025)

DataStatPro is an integrated statistical platform that aims to provide structured data management, statistical analysis, visualisation, and publication-support tools within a single environment.

Data Management solutions are useful to facilitate the importation, exportation, alteration, transformation, filtering and organizing of variables and datasets across numerous sources to ensure the production of clean and analysis-ready data.

The Descriptive and Inferential Statistics module includes simple summaries, non-parametric tests, frequency distributions, ANOVA, t-tests, and verification of assumptions, and experimental design.

Correlation, Regression, and Time-Series Analysis make it possible to do robust association analysis, prediction, and survival modeling.

Advanced Statistical Analysis includes features related to meta-analysis, reliability testing, factor analysis, multilevel modeling, mediation/moderation, causal inference, and Six Sigma quality.

The Data Visualisation module includes both standard and advanced graphical outputs, including histograms, boxplots, Sankey diagrams, interaction plots, and publication-quality visualisations.

The publication-support tools within DataStatPro are designed to streamline academic reporting through automated methods sections, standardised tables, citation management, flow diagrams, and enhanced figure preparation. (DataStatPro, 2025)

The integrated statistical calculators within DataStatPro are designed to support early stages of study planning and statistical decision-making. These tools include sample size estimation, epidemiological calculations, descriptive summaries, confidence intervals, effect size calculations, and p-value estimation. Such features may assist researchers and healthcare professionals in conducting preliminary analyses more efficiently within a structured analytical environment..(DataStatPro, 2025) At present, formal peer-reviewed validation studies comparing DataStatPro with established statistical software platforms remain limited and warrant further investigation.

Features and Research Workflow Support in DataStatPro

Table 1 summarises the major modules, statistical capabilities, visualisation tools, and publication-support functions available within the DataStatPro platform.

Table -1: Features Of Datastatpro: <https://Datastatpro.Com/App/Dashboard>

Category	Feature Name	Description & Capabilities
Data Management	Data & Variable Editors	Import/Export; Edit/Modify Data; Define Variables; Data Quality Report.
	Transform & Filter	Apply Transformations; Calculations; Sub-Setting And Filtering; Composite Scores
Descriptive Statistics	Descriptives & Frequencies	Mean; Median; SD; Frequency Tables; Distributions.
	Advanced Descriptives	Cross Tabulation; Multiple Response Analysis; Normality Tests.

Inferential Statistics	T-Tests & Non-Parametric	One-Sample; Independent; Paired T-Tests; Mann-Whitney U; Wilcoxon.
	ANOVA & DOE	One/Two-Way; Repeated Measures ANOVA; Design Of Experiments (DOE).
	Assumptions	Automated Verification Of Test Assumptions.
Correlation & Regression	Correlation	Pearson; Spearman; Partial Correlations.
	Regression Models	Linear; Logistic (Binary/Multinomial); Cox (Survival); Time Series.
Data Visualization	Standard Charts	Bar; Pie; Histogram; Box; Scatter; Error Bar Charts.
	Advanced Plots	Rain Cloud; Sankey; Pivot; Shelf-Life; Interaction Plots.
Statistical Calculators	Study Planning	Sample Size Calculators; Epi Calculator; P-Value Calculator; Confidence Interval Calculator; Effect Size Calculator.
	Test Calculators	P-Value; Confidence Intervals (CI); Effect Size; Summary Stats.
Advanced Analysis	Survival & Reliability	Kaplan-Meier; Log-Rank; Cronbach's Alpha; ICC.
	Multivariate & Causal	MANOVA/MANCOVA; Mediation/Moderation; Difference-In-Differences Models; Multilevel Modeling.
	Structural & Factor	EFA & CFA (SEM; Fit Indices); Variable Tree Analysis (Decision Trees/Random Forests).
	Clustering & Meta	K-Means/Hierarchical Clustering; Meta Analysis (Forest Plots).
	Quality Control	Six Sigma (DMAIC; SPC; Process Capability, DOE).
Publication Ready Tools	Results & Tables	Results Manager; Auto-Generated Tables 1; 1a; 1b; 2; 2a; 3; SMD; Regression. Post-Hoc Multiple Comparison Procedures.
	Methodology & Citations	Auto-Methods Generator; APA Converter; Citation Manager.
	Figures & Diagrams	CONSORT Flow Diagrams; Figure Caption Generator; Enhanced Figure Processor.

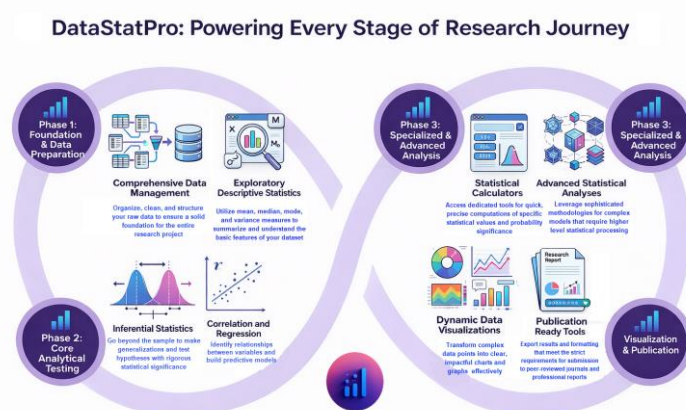


Fig-2: Datastatpro and Stages of Research Journey

Comparison with Existing Statistical Tools

These days the statistical analysis is not restricted to one tool only, but it is a multi-platform process which is conditioned by the abilities, time and needs of the researchers.

Although they may be perceived to be rigid and costly, most physicians and epidemiologists continue to use the popular packages such as SPSS and Stata due to its popularity and ease of use. Nevertheless, even though both R and

Python offer robust chances for automation and reproducibility, the technical complexity of both languages discourages busy medical researchers. Although more advanced inferences and auto interpretations are challenging for more simplistic programs like Excel or JASP, new applications like DataStatPro are being created and are arising to guide researchers through clinically relevant analyses. These platforms are well-built, have a lower technical load and retain scientific rigor, and have contemporary automation to make the experience of data analysis more accessible and friendly. Table 2 compares commonly used statistical software platforms in terms of research applications, strengths, and major limitations. Figure 2 demonstrates the

integration of DataStatPro across different stages of the research journey, from study planning to publication-ready outputs.

Strengths, Challenges and Future Validation Needs

Although integrated analytical platforms offer advantages related to accessibility and workflow efficiency, they may also have limitations, including reduced flexibility for advanced custom analyses, dependency on predefined analytical pipelines, and limited transparency regarding underlying algorithms. Further validation studies and comparative evaluations are required to assess their reproducibility, accuracy, and applicability across different research disciplines.

Table -2: Comparison of Common Statistical Data Analysis Tools and Their Pros and Cons

Software / Tool	Typical Use In Research	Key Strengths	Major Challenges / Limitations
Spss	Clinical, Public Health, Social Sciences	User-Friendly Gui, Widely Accepted In Journals, Strong Classical Statistics	Limited Automation And Reproducibility, Weak Advanced Modeling, License Cost
R / Rstudio	Advanced Statistics, Reproducible Research, Meta-Analysis	Open-Source, Highly Flexible, Excellent Graphics And Reporting	Steep Learning Curve, Coding Required, Package/Version Management
Python	Data Science, Ai/ML, Large Datasets	Powerful Libraries, Ai Integration, Automation	Less Intuitive For Traditional Biostatistics, Verbose Statistical Reporting
Stata	Epidemiology, Econometrics	Clean Syntax, Excellent Regression Tools, Reproducible Workflows	Expensive, Limited Ai/ML, Smaller Visualization Ecosystem
Sas	Clinical Trials, Pharma, Regulatory Studies	Regulatory Acceptance, Robust Data Handling	Very High Cost, Complex Syntax, Declining Academic Use
Jasp	Teaching, Bayesian Analysis, Transparent Statistics	Free, Gui-Based, Bayesian Support	Limited Advanced Models, Smaller Feature Set
Excel	Data Entry, Basic Summaries	Easy Access, Quick Visualization	High Error Risk, Not Suitable For Inferential Research
Matlab	Engineering, Biomechanics, Signal Processing	Numerical Precision, Matrix Operations	Expensive, Not Optimized For Epidemiology
Datastatpro	Applied Medical & Public Health Research	Guided Analysis Workflow, Minimal Coding, Clinical-Friendly Outputs, Integrated Assumptions & Interpretation Support	Currently Limited For Ai/ML, Less Customizable Than R/Python, Dependent On Predefined Pipelines

CONCLUSION

The progression of statistical analysis from manual calculations and mechanical computing devices to contemporary AI-assisted analytical

systems reflects the rapid technological evolution of modern research methodologies. Over recent decades, the development of statistical software such as SAS, SPSS, Stata, R,

and Python has transformed data management, statistical modelling, reproducibility, and accessibility across medical, public health, and multidisciplinary research environments. The emergence of cloud computing, machine learning, open-source ecosystems, and integrated analytical platforms has further expanded the ability of researchers to perform complex analyses with greater efficiency and automation.

The literature reviewed in this article highlights the growing importance of user-friendly, reproducible, and clinically applicable analytical systems capable of supporting modern data-driven decision-making. Emerging platforms such as DataStatPro represent an example of integrated analytical environments designed to reduce technical barriers and streamline data management, statistical analysis, visualisation, and reporting workflows within a single platform. However, despite these advancements, important challenges remain regarding methodological transparency, reproducibility, algorithmic bias, ethical AI implementation, cybersecurity, and the need for appropriate human oversight.

Furthermore, formal peer-reviewed validation and comparative evaluation studies remain necessary to determine the reliability, usability, and scientific robustness of emerging AI-assisted analytical platforms when compared with established statistical software. Future developments in statistical computing should therefore prioritise evidence-based validation, interdisciplinary collaboration, transparency, and responsible AI governance to ensure accurate, reproducible, and ethically sound data analysis practices in healthcare and scientific research.

Ethics and consent to participate

Not applicable. This article is a historical narrative review and did not involve human participants, patient data, animal subjects, or primary data collection. Therefore, ethical approval and consent to participate were not required.

Competing interests

The authors declare that they have no competing interests.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Authors' contributions

All authors contributed to the conception and design of the article. AH, TA, SA, AST, ZS and SI contributed to the literature review, manuscript drafting, critical revision, and approval of the final manuscript. All authors read and approved the final version of the manuscript.

Acknowledgements

The authors would like to acknowledge the developers and publicly available documentation of DataStatPro and other statistical software platforms discussed in this review.

REFERENCES

1. Abatan, S. M. & Olayemi, M. 2014. The role of statistical software in data analysis. *International Journal of Applied Research and Studies (iJARS)* ISSN, 2278-9480.
2. Ashour, L. 2024. A review of user-friendly freely-available statistical analysis software for medical researchers and biostatisticians. *Research in Statistics*, 2, 2322630.
3. Baniya, B. 2024. Integrating and validating heart health calculators in medical platforms. *UiT Norges arktiske universitet*.
4. Datastatpro 2025. DataStatPro Dashboard - Statistical Analysis Platform. [Online available]: <https://datastatpro.com/>, accessed Dec 19, 2025.
5. Deo, V. & Ranganathan, P. 2024. Statistical tools and packages for data collection, management, and analysis - A brief guide for health and biomedical researchers. *Perspect Clin Res*, 15, 209-212.
6. Geeksforgeeks A brief history of data analysis [Internet]. Noida (IN): GeeksforGeeks; 2025 Jul 23 [cited 2025 Dec 19]. Available from: <https://www.geeksforgeeks.org/data-analysis/a-brief-history-of-data-analysis/>.
7. Lo'ai, A. T., Mehmood, R., Benkhelifa, E. & Song, H. 2016. Mobile cloud computing model and big data analysis for healthcare applications. *IEEE Access*, 4, 6171-6180.
8. Team, V. 2021. The evolution of statistical computing: a captivating journey through history Internet [cited 2025 Dec 19]. Available from:

- <https://vsni.co.uk/evolution-of-statistical-computing/>).
9. Upton, E., Cai, X., Jakiela, P., Ozier, O. & Raman, S. 2025. The Software Behind the Stats: A Student Exploration of Software Trends Across Disciplines. arXiv preprint arXiv:2504.06507.
 10. Wikipedia 2025. List of statistical software [Internet]. . Wikimedia Foundation; [updated 2025 Dec; cited 2025 Dec 19]. Available from: https://en.wikipedia.org/wiki/List_of_statistical_software/