ISSN: 3048-8451 (ONLINE)

INTERNATIONAL RESEARCH JOURNAL OF SCIENTIFIC STUDIES

OCTOBER 2025
VOLUME 2, ISSUE 8

https://www.irjss.com

irjss.editor@gmail.com

Published by Sciencia Publishers

International Research Journal of Scientific Studies
e-ISSN: 3048-8451

RESEARCH ARTICLE OCTOBER 2025, VOLUME 2, ISSUE 8

NovoNumeric: An Accessible, High-Performance Statistical Software for Advancing Medical Research on macOS

Mohammed Jaseem Ibrahim K. [1]

[1] NovoNumeric Software, Thiruvananthapuram, India

✉ Corresponding author email: drjaseemibrahim@gmail.com

Cite this article: Mohammed Jaseem Ibrahim K., (2025). NovoNumeric: An Accessible, High-Performance Statistical Software for Advancing Medical Research on macOS. International Research Journal of Scientific Studies, 2(8), 1–9. https://doi.org/10.64383/irjss.20250924

Abstract: *The field of statistical analysis is increasingly reliant on software tools that are not only powerful but also accessible and intuitive. This paper introduces NovoNumeric, a novel statistical analysis application engineered specifically for the macOS platform. Developed natively in SwiftUI, NovoNumeric provides a comprehensive, end-to-end analytical workflow within a single, responsive interface. The application's architecture is centered around a stateless analysis engine that leverages Apple's high-performance Accelerate and LAPACK frameworks for its core numerical computations, ensuring both speed and accuracy. Key features include a versatile data management suite, a library of 18 parametric and non-parametric statistical tests, advanced multivariate procedures such as Principal Component Analysis (PCA) and K-Means Clustering, and a dynamic visualization engine built on the native Charts framework. Notably, the application integrates essential diagnostic procedures, including Variance Inflation Factor (VIF) for regression and post-hoc tests for ANOVA, promoting statistical best practices. By combining a robust computational backend with a user-centric design, NovoNumeric aims to bridge the gap between complex analytical power and user accessibility, providing a valuable tool for researchers, students, and data analysts in the Apple ecosystem.*

Keywords: *Statistical Software, Data Analysis, macOS, SwiftUI, Computational Statistics, High-Performance Computing, Accelerate Framework, Data Visualization, Principal Component Analysis, Linear Regression.*

1. Introduction

The proliferation of data across scientific disciplines has underscored the necessity for robust, efficient, and accessible statistical software. While a number of powerful analytical environments exist, such as R [1] and the Python scientific stack [2], these often present a steep learning curve, particularly for users whose primary expertise is not in computer programming. Conversely, many graphical user interface (GUI) based statistical packages can be limited in their capabilities, platform availability, or may rely on non-native, cross-platform frameworks that result in a suboptimal user experience and performance degradation.

A specific gap exists for researchers and analysts operating within the macOS ecosystem who seek a tool that combines the performance and usability of a native application with a comprehensive suite of statistical functionalities. While established tools like SPSS or SAS are powerful, their cross-platform nature often leads to a user experience on macOS that can feel less fluid and disconnected from the native environment. By being engineered specifically for macOS using native frameworks like SwiftUI and Charts, NovoNumeric offers a more intuitive and responsive interface that adheres to the platform's established design conventions, providing a significant advantage in user experience over non-native competitors.

To address this need, we have developed NovoNumeric, a modern statistical analysis application engineered from the ground up for macOS. NovoNumeric is designed to provide an integrated environment that supports the entire research pipeline—from data ingestion and preprocessing to complex analysis and visualization. Its design philosophy is centered on three core principles:

- Native Performance: Leveraging system-level frameworks to ensure that computations are fast and the user interface is fluid and responsive.
- Statistical Rigor: Implementing established algorithms and incorporating essential diagnostic tests to ensure the validity of analytical results.
- User Accessibility: Offering an intuitive, graphical interface that lowers the barrier to entry for advanced statistical methods, complemented by an integrated help system.

This paper details the software architecture, methodological implementation, and key features of NovoNumeric, demonstrating its utility as a powerful new tool for the scientific community.

2. System Architecture and Design

NovoNumeric is architected using the Model-View-ViewModel (MVVM) pattern, a modern design paradigm that promotes a clean separation of concerns. The application is built entirely in SwiftUI [3], Apple's declarative UI framework, which allows for a reactive and maintainable user interface. The architecture comprises four primary components:

- **Views (UI Layer):** A set of declarative SwiftUI View structs that define the application's layout and user interaction elements. This layer is lightweight and stateless, simply reflecting the current state of the View Model.
- **View Model (DataManager.swift):** A centralized ObservableObject that serves as the single source of truth for the application. It manages the active dataset, user selections, and all data transformation logic. Its @Published properties ensure that any change to the data automatically propagates to the View layer.
- **Models (Models.swift):** A collection of struct types that define the application's data entities, such as the results from a t-test (TTestResult) or the coefficients from a regression model (RegressionCoefficient). This ensures type safety and data consistency.
- **Analysis Engine (Logic Layer)**: A stateless computational engine composed of static methods (AnalysisEngine.swift,LinearAlgebra. swift). This engine accepts data as input, performs a calculation, and returns a result model. Its stateless nature—meaning each calculation is self-contained and does not depend on prior states—makes it highly predictable, reusable, and straightforward to test. This design choice promotes a clearer separation of concerns and enhances performance by simplifying concurrent operations, a key benefit for a computing-focused application.

A key design decision was to leverage Apple's native frameworks for performance-critical tasks. All intensive numerical computations are delegated to the Accelerate framework [4], which provides highly optimized, low-level access to the CPU for vector and matrix mathematics.

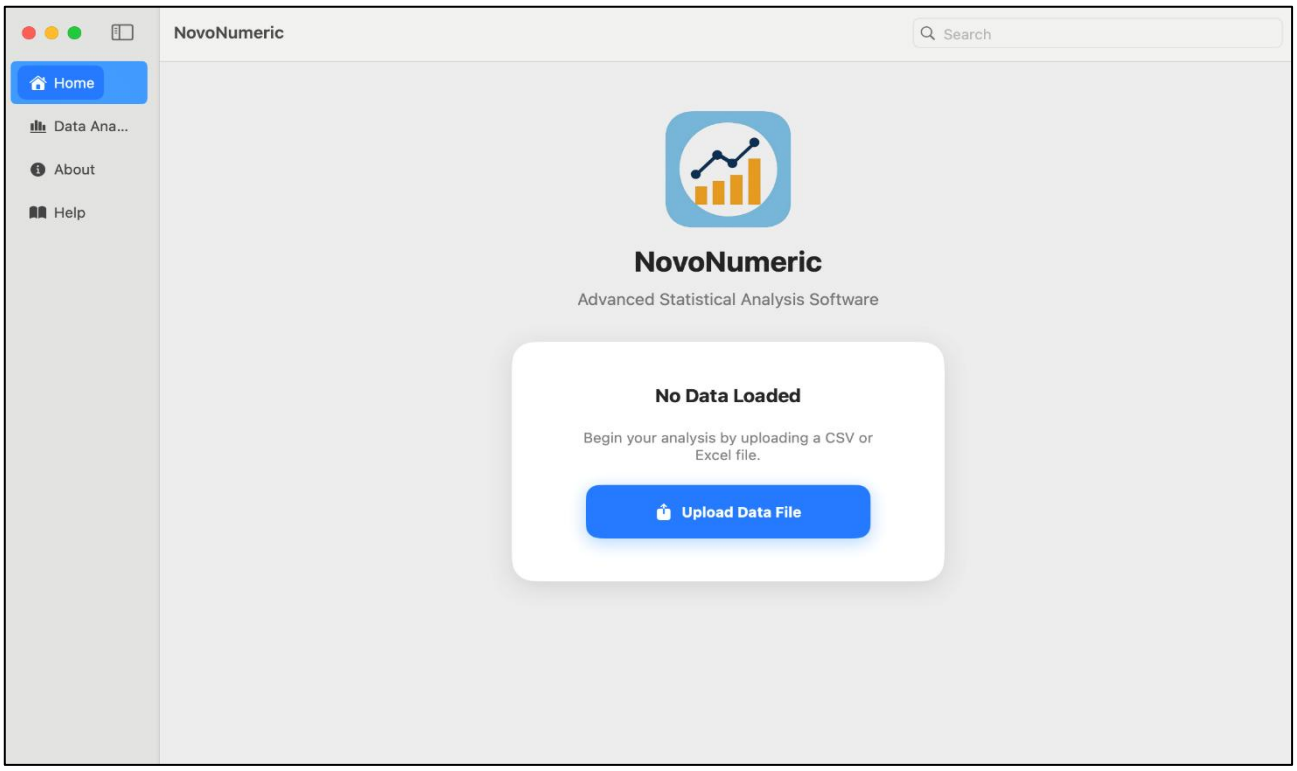


Figure 1: Apps user interface (Home screen)

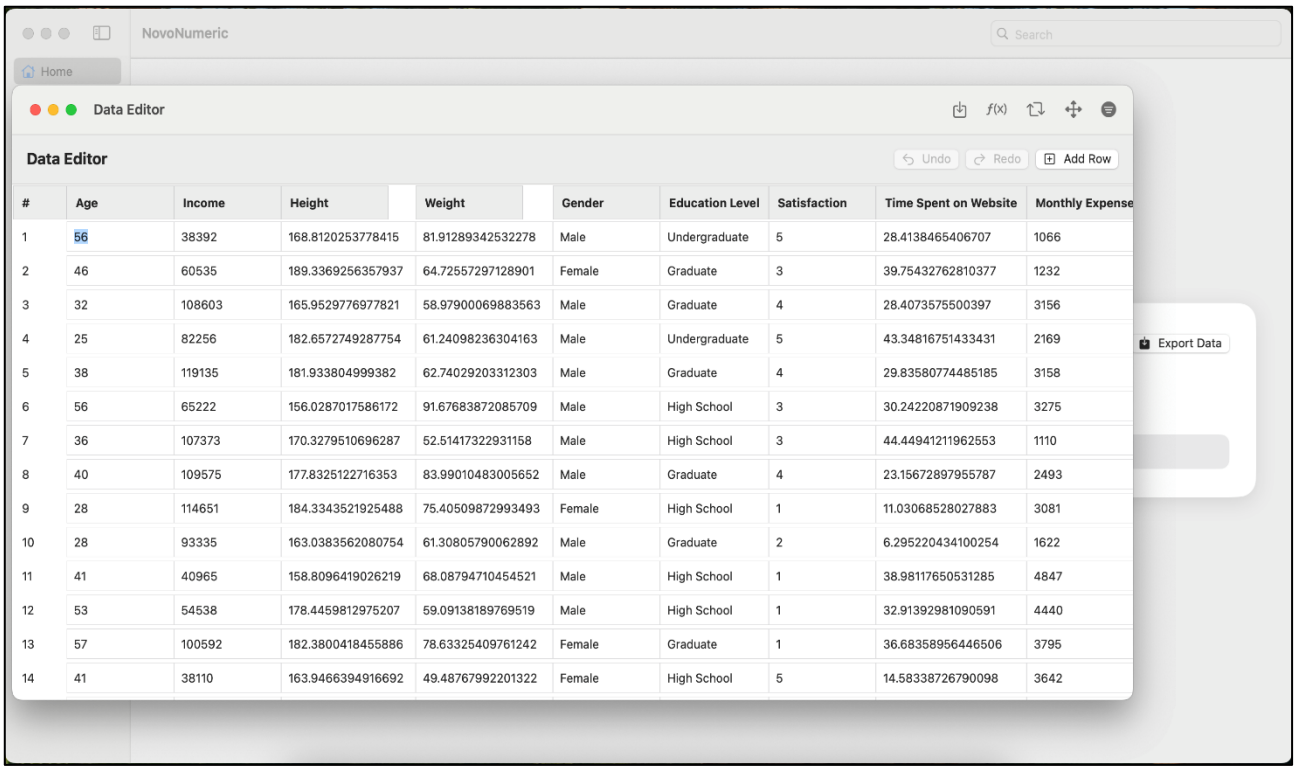
3. Methodological Implementation

The validity of any statistical software rests on the correctness of its algorithmic implementations. NovoNumeric implements established statistical formulas and relies on industry-standard numerical libraries for its core computations.

3.1 Data Management and Preprocessing

NovoNumeric provides a suite of tools for data preparation, a critical first step in any analysis.

- **Data Import:** The application uses Apple's TabularData framework for parsing CSV files and the CoreXLSX library for .xlsx files. This process is handled asynchronously on a background thread to ensure the UI remains responsive, even with large datasets.
- **Data Transformation:** Users can compute new variables from existing columns, add new rows, recode continuous variables into categorical ones, and filter the dataset based on logical conditions.
- **Normalization:** The application provides both Min-Max Scaling $(x - min) / (max - min)$ and Z-Score Standardization $(x - \mu) / \sigma$ to prepare data for distance-sensitive algorithms like PCA and K-Means, as described by Hastie et al. [5].



#	Age	Income	Height	Weight	Gender	Education Level	Satisfaction	Time Spent on Website	Monthly Expense
1	56	38392	168.8120253778415	81.91289342532278	Male	Undergraduate	5	28.4138465406707	1066
2	46	60535	189.3369256357937	64.72557297128901	Female	Graduate	3	39.75432762810377	1232
3	32	108603	165.9529776977821	58.97900069883563	Male	Graduate	4	28.4073575500397	3156
4	25	82256	182.6572749287754	61.24098236304163	Male	Undergraduate	5	43.34816751433431	2169
5	38	119135	181.933804999382	62.74029203312303	Male	Graduate	4	29.83580774485185	3158
6	56	65222	156.0287017586172	91.67683872085709	Male	High School	3	30.24220871909238	3275
7	36	107373	170.3279510696287	52.51417322931158	Male	High School	3	44.44941211962553	1110
8	40	109575	177.8325122716353	83.99010483005652	Male	Graduate	4	23.15672897955787	2493
9	28	114651	184.3343521925488	75.40509872993493	Female	High School	1	11.03068528027883	3081
10	28	93335	163.0383562080754	61.30805790062892	Male	Graduate	2	6.295220434100254	1622
11	41	40965	158.8096419026219	68.08794710454521	Male	High School	1	38.98117650531285	4847
12	53	54538	178.4459812975207	59.09138189769519	Male	High School	1	32.91392981090591	4440
13	57	100592	182.3800418455886	78.63325409761242	Female	Graduate	1	36.68358956446506	3795
14	41	38110	163.9466394916692	49.48767992201322	Female	High School	5	14.58338726790098	3642

Figure 2: Data Management

3.2 Statistical Analysis Engine

The core analytical capabilities are implemented in the stateless AnalysisEngine.

- **Linear and Multiple Regression:** To solve for the coefficient vector β in the ordinary least squares (OLS) equation $y = Xβ$, NovoNumeric solves the normal equation $(X'X)β = X'y$. For maximum numerical stability, particularly in the presence of multicollinearity, the system does not compute the matrix inverse $(X'X)^{-1}$ directly. Instead, it solves the system of linear equations using the LAPACK function dgesv_ [6], which employs LU decomposition. Standard errors are subsequently calculated using the diagonal of the variance-covariance matrix $σ^2(X'X)^{-1}$, where the inverse is computed via dgetrf_ and dgetri_.
- **Regression Diagnostics:** A critical feature for multiple regression is the automatic calculation of the Variance Inflation Factor (VIF) for each predictor. For each predictor X_j, an ancillary regression is performed against all other predictors to obtain an R_j^2 value. The VIF is then calculated as $1 / (1 - R_j^2)$. A high VIF (>5) alerts the user to potential multicollinearity issues, a technique detailed in Belsley et al. [7].
- **Principal Component Analysis (PCA):** The PCA implementation is centered on the Singular Value

Decomposition (SVD) of the standardized (Z-score) data matrix. This is performed by the LAPACK function dgesvd_ [6]. The component loadings are derived from the resulting right singular vectors (the VT matrix), and the variance explained by each component is calculated from the singular values. The results are visualized using a Scree Plot, which includes a Kaiser criterion line to aid in determining the number of components to retain.

- **ANOVA and Non-Parametric Tests:** One-Way ANOVA is implemented by partitioning the total sum of squares and calculating the F-statistic. A significant result is automatically followed by a Tukey's HSD post-hoc test [8] to identify which specific group means differ. Its non-parametric equivalent, the Kruskal-Wallis test, is also available and is followed by a Dunn's test [9] with Bonferroni correction for pairwise comparisons. The normality assumption for these parametric tests is checked using the Shapiro-Wilk test [10].
- **K-Means Clustering:** The application implements the K-Means++ algorithm for centroid initialization, which offers advantages over random seeding by selecting initial centroids with probabilities proportional to their contribution to the overall inertia. This method was proposed by Arthur and Vassilvitskii [11] to improve both the speed and accuracy of the clustering results.

3.3 Integrated Visualization

All visualizations are generated using Apple's native Charts framework [12]. This ensures high-quality, performant graphics that are tightly integrated with the application's data flow. The ChartContainer view is a dynamic component that can render any of the 13 chart types based on the user's selection and the results of the current analysis.

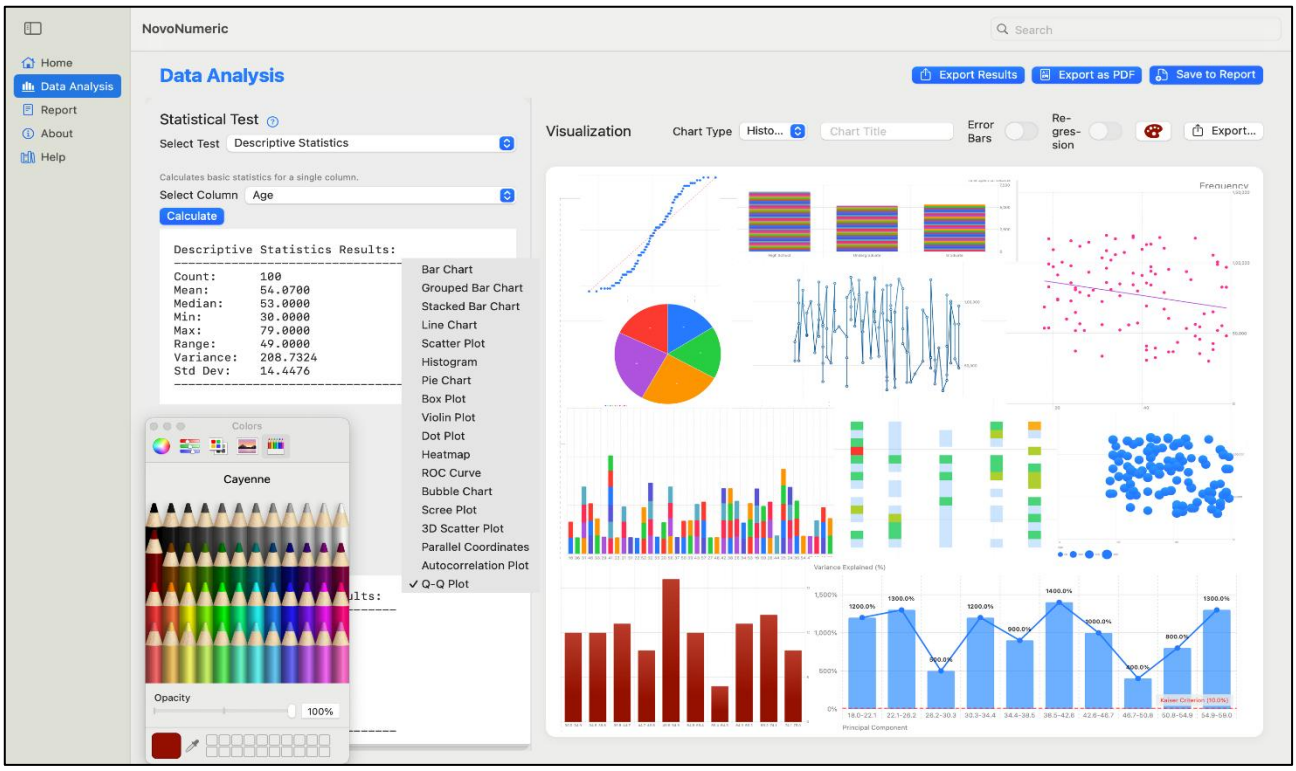


Figure 3: Visualizations in the app

4. Performance and Validation

The direct use of the Accelerate framework provides a significant performance advantage over solutions that rely on interpreted languages or cross-platform libraries for numerical computation. Matrix multiplication (vDSP_mmulD), linear system solving (dgesv_), and SVD (dgesvd_) are executed with near-native hardware speed. From a validation perspective, the software has been verified by comparing its output for standard test datasets against established statistical packages like R. The inclusion of diagnostic tools like VIF, post-hoc tests, and assumption checks (e.g., Shapiro-Wilk test for normality) provides users with the necessary tools to validate their own models and adhere to statistical best practices.

5. Illustrative Case Study

Consider a researcher investigating the predictors of a student's final exam score. The researcher imports an Excel file containing columns for Final_Score, Hours_Studied, Previous_GPA, and Attendance_Rate.

- ✓ **Import & Explore:** The researcher loads the data via the HomeView. They use the Descriptive Statistics tool to summarize each variable.
- ✓ **Analysis:** In the DataAnalysisView, they select "Multiple Linear Regression". They set Final_Score as the dependent variable and the other three as independent variables.

- ✓ **Validation:** Upon clicking "Calculate", NovoNumeric provides the regression coefficients, standard errors, and p-values. It also displays the VIF for each predictor. The researcher notes that all VIFs are below 2.0, indicating no significant multicollinearity.
- ✓ **Interpretation:** The results show that Hours_Studied and Previous_GPA are significant predictors ($p < 0.05$), while Attendance_Rate is not.
- ✓ **Visualization & Export:** The researcher creates a scatter plot of Hours_Studied vs. Final_Score, adds a regression line, gives it a title, and exports the chart as a PNG for their report. They also export the numerical results table as a text file.

This entire workflow is completed within a single, intuitive interface without writing any code.

6. Discussion and Future Directions

NovoNumeric provides a powerful and accessible tool for a wide range of common statistical analyses. It excels by offering a highly integrated, fluid, and performant experience that is often lacking in non-native or purely programmatic alternatives on the macOS platform. However, it is not intended to be a complete replacement for programmatic environments like R or Python, which offer unparalleled flexibility for highly specialized or novel statistical methods.

Future development will focus on expanding the library of statistical tests to include more advanced models, such as mixed-effects models, factor analysis with rotations other than Varimax, and more time-series forecasting techniques (e.g., ARIMA). Additionally, the integration of a minimal scripting interface is being considered to allow for the automation of repetitive analyses.

Conclusion

NovoNumeric represents a significant contribution to the ecosystem of statistical software available on the macOS platform. By combining a high-performance computational engine, a comprehensive suite of validated statistical methods, and an intuitive, native user interface, it successfully lowers the barrier to entry for sophisticated data analysis. It empowers students, educators, and researchers to conduct their work efficiently and rigorously, making it a valuable addition to the modern analyst's toolkit.

References:

[1] R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2023. Available from: https://www.R-project.org/

[2] Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods. 2020;17(3):261-272.

[3] Apple Inc. SwiftUI. Apple Developer Documentation. [Cited 2025 Jul 10]. Available from: https://developer.apple.com/xcode/swiftui/

[4] Apple Inc. Accelerate. Apple Developer Documentation. [Cited 2025 Jul 10]. Available from: https://developer.apple.com/documentation/accelerate

[5] Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed. New York: Springer; 2009.

[6] Anderson E, Bai Z, Bischof C, Blackford S, Demmel J, Dongarra J, et al. LAPACK Users' guide. 3rd ed. Philadelphia, PA: Society for Industrial and Applied Mathematics; 1999.

[7] Belsley DA, Kuh E, Welsch RE. Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. New York: John Wiley & Sons; 1980.

[8] Tukey JW. The problem of multiple comparisons. Princeton, NJ: Princeton University; 1953.

[9] Dunn OJ. Multiple Comparisons Among Means. J Am Stat Assoc. 1961;56(293):52-64.

[10] Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). Biometrika. 1965;52(3/4):591-611.

[11] Arthur D, Vassilvitskii S. k-means++: The advantages of careful seeding. In: Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms. 2007. p. 1027-1035.

[12] Apple Inc. Charts. Apple Developer Documentation. [Cited 2025 Jul 10]. Available from: https://developer.apple.com/documentation/charts

Author(s) Profile:

1. Dr. Mohammed Jaseem Ibrahim K. is a medical doctor (pharmacologist) and medical researcher specializing in creating high-performance computational tools. With a keen interest in research and a passion for technology, his work focuses on bridging the gap between complex statistical methods and accessible user interfaces. He is the creator of NovoNumeric, a native macOS application designed to make advanced statistical analysis intuitive for researchers and students. His interests include computational statistics, native software development, and improving analytical workflows in the scientific community.

Article History:

Submitted: 2025-09-12
Reviewed: 2025-09-27
Accepted: 2025-10-07