What's next for IBM SPSS Statistics

The next evolution of IBM SPSS Statistics is here : Version 31

Highlights

Enables businesses to deliver personalized promotions and real-time navigation assistance

Measure dependencies beyond simple correlations

Visualize multi-factor connections to find hidden groupings and relationships

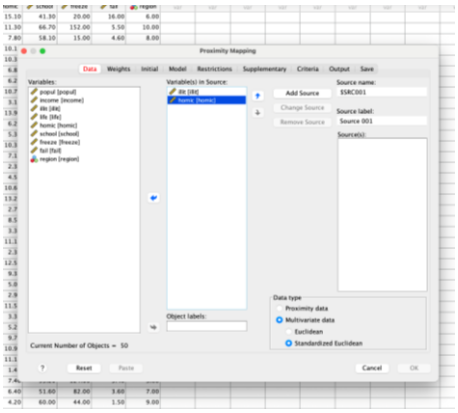
Clear noise from time data for better trend identification and prediction

IBM® SPSS® Statistics is an end-to-end on-premise statistical solution that simplifies advanced statistical analysis across industries. It is widely used for data preparation and management, advanced data analysis, forecasting models, and data visualizations. It offers comprehensive resources, expert support, and proven reliability to transform complex data into impactful decisions. It is designed to help not only researchers, data analysts, statisticians, and educators but also business professionals extract actionable insights from their data. Our solution offers a powerful suite of tools that can address any type of analysis, from basic descriptive statistics to complex multivariate methods.

Widely leveraged by both individuals and organizations spanning across Retail, Consumer Packaged Goods, E-Commerce, Healthcare, Government , Wholesale Distribution and Services , Manufacturing , and Supply Chain etc. who seek an advanced statistical solution that can simply complex statistical test through an easy to use, accurate ,reliable and secure solution.

With support for multiple data formats and an intuitive interface, IBM SPSS Statistics transforms data into valuable insights and engaging visualizations through graphs and charts. Its comprehensive features in data preparation, statistical modeling, and reporting enables organizations and data professionals to make informed, data-driven decisions.

IBM® SPSS® Statistics Version 31 introduces a comprehensive set of new algorithms, best-in-class features and product enhancements, designed to optimize both data analysis capabilities and user experience. It enables organizations to evaluate data with ad hoc analysis, hypothesis testing, and predictive analytics by offering a powerful set of tools to validate assumptions, analyse past performance, and forecast trends. It brings new capabilities such as proximity mapping, distance correlation, time series filtering, STATS Earth and conditional inference trees along with new exciting introduction such as Curated Help.



Proximity Mapping

PROXMAP or proximity mapping refers to a technique that is used in various fields like geography, biology, and technology, to analyze and visualize the relationships or distances between different elements or entities. Proximity mapping helps researchers, planners, and analysts understand the spatial or relational dynamics of the elements that are being studied. This study aids decision-making, identifying patterns, optimizing systems and more.

It involves measuring the distance between geographic features to understand their spatial relationships. This includes calculating distances between points, lines, or polygons, and is commonly used in Geographic Information Systems (GIS) to analyse spatial patterns and interactions.

This feature enables businesses to deliver personalized promotions and real-time navigation assistance, enhancing the customer experience in retail stores, malls, airports, and hospitals. By understanding customer behaviour through location data, companies can implement highly targeted marketing campaigns, increasing conversion rates and prompting impulse purchases. Additionally, proximity mapping allows for optimized inventory management and more efficient staff allocation, ultimately improving operational efficiency. Early adopters gain a competitive edge by offering advanced, personalized experiences that foster customer loyalty, while the rich behavioural insights gathered help businesses identify and adapt to emerging trends swiftly.

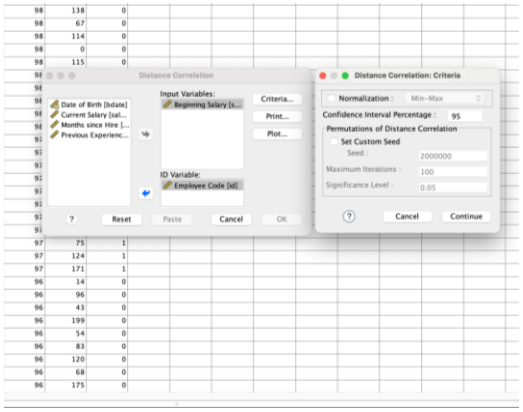
Key Benefits
Proximity mapping in Multidimensional Scaling (MDS) holds significant business value across various domains. This technique allows businesses to visualize and analyse complex relationships between data points, leading to several key benefits:

- **Customer insights and segmentation:** By employing MDS-based proximity mapping on customer data, businesses can identify clusters of customers with similar preferences, behaviours, or characteristics. This segmentation is invaluable for targeted marketing strategies, personalized offerings, and improved customer satisfaction

- **Market understanding and competitive analysis:** Proximity mapping helps businesses gain insights into market dynamics, competitive positioning, and brand perception. By mapping products, services, or brands based on customer perceptions or market attributes, businesses can identify market trends, competitive advantages, and areas for differentiation

- **Supply chain optimization:** Proximity mapping is instrumental in optimizing supply chain operations. By mapping the spatial relationships between suppliers, distribution centres, and demand points, businesses can streamline logistics, reduce transportation costs, minimize lead times, and improve overall supply chain efficiency

- **Risk management and decision making:** Proximity mapping facilitates risk assessment and decision making across various business functions. In finance and investment sectors, it helps in portfolio diversification, asset allocation, risk mitigation, and identifying correlations between assets for informed investment decisions



Distance Correlation

Distance Correlation is a measure of statistical dependence between two random variables or datasets that captures both linear and nonlinear relationships. It is a robust and universal way to check if there is a relation between two numeric variables. In statistical analysis, understanding the relationship between two variables is fundamental. Traditional methods, such as Pearson's correlation coefficient are limited to capturing linear associations. However, real-world data, especially in the context of financial behavior, often exhibit nonlinear or complex dependencies that linear methods might fail to detect. This is where distance correlation comes in handy as it is a more versatile metric that detects any form of statistical dependence between variables (linear or nonlinear)

Unlike pearson correlation which only measures linear dependence, distance correlation accounts for complex associations, making it useful in high-dimensional data analysis. It equals zero if and only if the variables are statistically independent. Therefore, distance correlation is an effective tool for uncovering complex interactions in multidimensional datasets as it can analyze behavioral or demographic variables where patterns may not follow a straight-line relationship.

Key Benefits

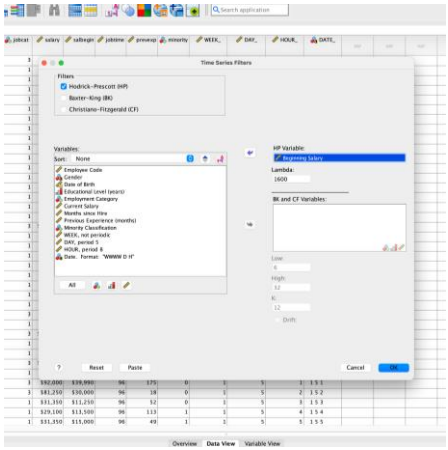
- **Captures non-linear dependencies**: Computes pairwise correlations for all variable combinations, detecting both linear and non-linear relationships, unlike traditional methods

- **Flexible normalization:** Offers multiple methods (MINMAX, ZSCORE, ROBUST, LOG) to normalize data, enhancing adaptability to diverse distributions and outliers

- **Robust significance testing:** Supports permutation testing with customizable options for reliable inference on small or non-normal datasets

- **Comprehensive outputs:** Provides detailed tables and scatterplot visualizations for clear insights into data relationships

- **Wide application scope**: Ideal for fields like bioinformatics (e.g., gene expression analysis), finance (e.g., portfolio risk assessment), machine learning (e.g., feature dependency analysis), social sciences (e.g., survey data relationships), and environmental science (e.g., climate variable interactions), due to its ability to handle high-dimensional, non-linear data



Time Series Filters

Time Series Filter is basically designed to address the challenge of noisy time-dependent data. By employing advanced filters such as Hodrick-Prescott (HP) filter, Baxter-King (BK) filter and Christiano-Fitzgerald (CF) filter, it allows users to gain a clearer understanding of their data's underlying behavior, leading to more accurate forecasting, improved anomaly detection, and enhanced real-time monitoring capabilities.

Time series filtering offers several broad-level advantages that are crucial for effective data analysis and forecasting. It aids in identifying long-term trends, cyclical patterns and seasonal patterns, which are essential for understanding the overall direction and periodic behaviors of the data. Additionally, it enhances the accuracy of predictive models and helps in detecting anomalies, ensuring data integrity and early identification of potential issues. These benefits collectively contribute to more reliable and insightful time series analysis.

Key Benefits

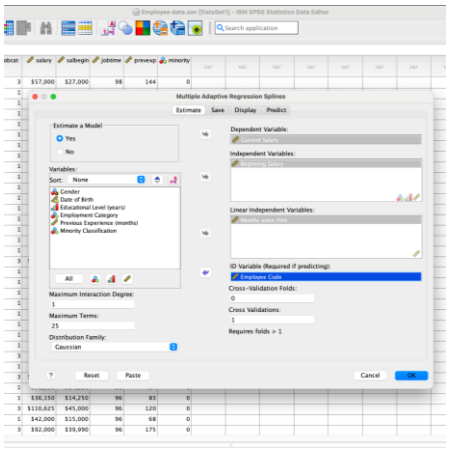
- **Better forecasting:** Cleaned data, free from distracting noise, leads to more accurate predictions of future trends
- **Improved anomaly detection:** Effectively filters out normal fluctuations, making it easier to detect genuine deviations and unusual events in your time series
- **Efficient decision-making:** Reveals underlying trends and patterns that are otherwise unknown, enabling businesses to react swiftly and strategically to changes
- **Clearer trend identification:** Makes long-term pattern and cycles in your data more visible, facilitating a deeper understanding of underlying dynamics
- **Enhanced real-time monitoring:** Provides a smoother, more stable view of streaming data, improving the accuracy of real-time alerts and insights

Conditional Inference Trees

It is a type of decision tree that uses statistical tests to select variables for splitting the data. This method avoids biases associated with traditional decision trees, which often use measures like the Gini coefficient. Conditional inference trees are better at determining the true effect of predictors by considering all other effects simultaneously

Key Benefits

- **Reduced bias:** Unlike traditional decision trees that may rely on biased measures, these use statistical tests to select variables, reducing bias in the selection process
- **Statistical significance:** This method ensures that the splits in the data are statistically significant, which enhances the robustness of the model and helps in making more confident decisions
- **Avoiding overfitting:** By using statistical tests to guide the splitting process, conditional inference trees are less prone to overfitting compared to traditional decision trees, leading to models that generalize better to new data



STATS EARTH

Multivariate Adaptive Regression Splines (MARS) Overview The MARS procedure, implemented in SPSS as STATS EARTH, constructs predictive models by fitting piecewise linear segments and one-hot encoded dummy variables to model nonlinearities and interactions.

Curated Help

When a certain statistical procedure is performed over a dataset, this feature will help users get more information about certain values in the table (output). This feature is implemented for Correlation Procedures.

Grey Out functionality

The non licensed list of algorithms/features will appear in a dark greyed font with a lock icon which enables the user to understand that there are advanced algorithms they ought to purchase for advanced analysis. The user will also be given the steps to make a purchase of the greyed-out algorithms/feature.

User Experience Enhancements

Enhanced Dark Mode experience
The dark mode interface has been significantly improved to offer a more visually comfortable and immersive experience. Enhancements include refined contrast levels, consistent color palettes across all UI components, and optimized readability for tables, charts, and text.

Customizable excel import with header row selection
Users now have the ability to specify which row(s) in an excel file should be treated as headers during the import process. This is particularly useful for datasets that include introductory notes or metadata above the actual data. It ensures accurate variable naming and data alignment, streamlining the data preparation workflow.

Search Bar for design of experiments (DoE) Techniques
A new search bar has been added to the DoE module, allowing users to quickly locate and access various experimental design techniques. This feature enhances discoverability and efficiency, especially for users working with complex or unfamiliar design methodologies.

Create Output Themes for Pivot Tables, Charts, and Viewer Text
Users can now create and save custom output themes that bundle their preferred styles for pivot tables, charts, and viewer text. These themes can be reused across projects to promote consistency in reporting and reducing time spent on formatting.

Persistent Column Widths in Variable View
Adjustments made to column widths in the Variable View are now saved and persist across sessions. This ensures a consistent and personalized layout, eliminating the need for repetitive manual resizing each time the software is launched.

Mnemonic Key Shortcuts for Tabbed Controls in CDB
Support for mnemonic key shortcuts has been added to tabbed controls in the Custom Dialog Builder (CDB). Users can now navigate between tabs using keyboard shortcuts, improving accessibility and speeding up dialog creation.

Redesigned Splash Screen Aligned with IBM Branding
The splash screen displayed during the launch of SPSS Statistics has been redesigned in accordance with IBM's latest branding guidelines.

Status Bar Displays Selected Case Count with Filters
When a filter is applied to the dataset, the status bar now shows the number of cases currently selected. This provides immediate feedback and helps users understand the scope of their filtered data, aiding in validation and analysis.

Add Background Images to Charts and Export in Multiple Formats
Users can now enhance their charts by adding background images, such as logos or watermarks and export in a wide range of formats including .png, .jpg, .svg, .bmp, .tif, and document formats, supporting diverse presentation and publication needs.

Coefficient of Variation (CV) Added
The Coefficient of Variation (CV) has been added as a statistical measure under both "Descriptives" and "Frequencies." The same label CV has been added across modules including Complex Samples and Ratio Statistics, ensuring consistency and clarity in statistical outputs.

Guidance for Completing Mandatory Fields When Paste is Disabled
When users attempt to paste syntax but required fields are incomplete, the system now provides contextual guidance to help identify and complete those fields. This reduces user frustration and ensures smoother execution of procedures.

Improved Variable Combination and Type Conversion
Enhancements have been made to simplify the process of combining multiple variables into one, changing variable types, and applying measurement levels to multiple variables simultaneously. These improvements streamline data preparation and reduce manual effort.

Chi-Square Test Promoted to Standalone Menu Item
The Chi-Square Test has been elevated to a dedicated menu item under Analyze → Descriptive Statistics, complete with its own icon and syntax support. This change reflects its importance in categorical data analysis and improves accessibility for frequent users.

Enhanced Trial Experience with Improved UI and CTAs
The trial version of the software now features a more intuitive user interface and clearer call-to-action buttons. These enhancements aim to provide a smoother onboarding experience, guiding new users through key features and encouraging deeper engagement with the product.

Security Enhancements

- JRE/JDK upgrade to 17.0.13.0
- Replaced old jai_*.jar files (Image library) with commons-imaging
- Addressed the difficulty of combining several variables into one variable, changing type and measuring multiple variables at once
- Python Upgrade to v3.13

Conclusion

IBM SPSS Statistics v31 introduces new algorithms, features and enhancements such as proximity mapping, distance correlation, curated help and many more. The new capabilities streamline the analysis process, providing intuitive visualization and tailored outputs that improve measurement reliability. By providing deeper insights into complex datasets, IBM SPSS Statistics v31 empowers professionals to make data-driven decisions that enhances the quality of their work across various industries.

For more information

To learn more about IBM SPSS Statistics, contact your IBM representative or IBM Business Partner us here, or visit here

IBM Corporation
New Orchard Road
Armonk, NY 10504

Produced in the
United States of America
June 2025