IBM SPSS Statistics Base

Be confident in your analytical results and in the business decisions you make

Highlights

- Provides support through every step of the analytical process

- Empowers analysts to carry out essential analyses from an intuitive graphical interface

- Select from more than a dozen integrated products to make specialized analyses faster and easier

- Add power when you need it, and connect data to decision-making through IBM SPSS Collaboration and Deployment Services

Organizations can solve a wide array of business and research problems with IBM SPSS Statistics. This groundbreaking suite of analytical products has been used worldwide for more than 40 years under the SPSS brand name.

Compared to other statistical software, IBM SPSS Statistics is easier to use, has a lower total cost of ownership and more comprehensively addresses the entire analytical process, from planning to data collection to analysis, reporting and deployment.

Organizations of all types rely on IBM SPSS Statistics to increase revenue, outmaneuver competitors, conduct research and make better decisions. With decades of built-in expertise and innovation, it's the world's number one choice for reliable statistical analysis.

IBM SPSS Statistics consists of IBM® SPSS® Statistics Base* plus more than a dozen fully integrated products that offer specialized functionality. Additional related products can easily share data and results with IBM SPSS Statistics but are operated through separate interfaces. Most are available on your choice of platforms: Microsoft® Windows®, Apple® Mac® or Linux®.**

Our suite of statistical software is now available in three editions: IBM SPSS Statistics Standard, IBM SPSS Statistics Professional and IBM SPSS Statistics Premium. By grouping essential capabilities, these editions provide an efficient way to ensure that your entire team or department has the features and functionality they need to perform the analyses that contribute to your organization's success.

You can choose either a client-only version of IBM SPSS Statistics Base or a server-based version, which provides more powerful capabilities, increased performance and scalability and more efficient administration.

* *IBM SPSS Statistics and IBM SPSS Statistics Base were formerly called PASW® Statistics and PASW® Statistics Base.*
** *Related products include IBM® SPSS® Amos™ for structural equation modeling, IBM® SPSS® SamplePower® for estimating sample size, IBM® SPSS® Text Analytics for Surveys for analyzing text data and IBM® SPSS® Visualization Designer for creating customized visualization.*
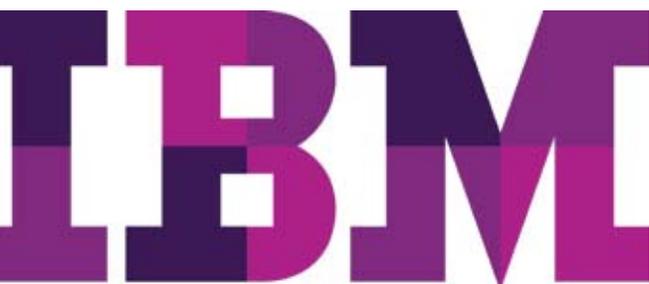
Business Benefits:

- Support business decisions with data-based analytics for improved outcomes

- Be more confident in your results by incorporating data from many different sources in your analysis and using proven, tested techniques to perform your analysis

- Save time and effort with capabilities that enable experienced analysts to develop procedures or dialogs that others can use to speed through repetitive tasks

- Give results greater impact by using visualization capabilities that clearly show others the significance of your findings

Access and analyze massive datasets quickly

IBM SPSS Statistics makes it easy for you to quickly access, manage and analyze any kind of dataset, including survey data, corporate databases or data downloaded from the web. In addition, IBM SPSS Statistics can process Unicode data. This eliminates variability in data due to language-specific encoding and enables your organization to view, analyze and share data written in multiple languages.

Prepare your data for analysis quickly and easily

Before you can analyze your data, you need to prepare it for analysis. Numerous techniques and features built into Base enable easy data preparation. Following are summaries of just a few data management highlights. With Base, you can easily set up data dictionary information (for example, value labels and variable types) and prepare your data for analysis more quickly using the Define Variable Properties tool. IBM SPSS Statistics Base presents a list of values and counts of those values so you can add this information. Once the data dictionary is set up, you can apply it using the Copy Data Properties tool. The data dictionary acts as a template, so you can apply it to other data files and to other variables within the same file.

IBM SPSS Statistics Base makes it easy for you to identify duplicate cases, so you can eliminate them prior to your analysis. Use the Identify Duplicate Cases tool to set parameters and flag duplicates so that you can keep track of them for the record.

Additionally, Base makes it easy to prepare continuous-level data for analysis. For example, the Visual Binner enables you to easily break income into "bands" of 10,000 or break ages into groups. A data pass provides a histogram that enables you to specify cutpoints in an intelligent manner. You can then automatically create value labels from the specified cutpoints (for example, "21-30").

Create your own dictionary information for variables with Custom Attributes. For example, create a custom attribute that represents the full text of a survey question when a code name such as "demo01" is used as the variable name. You can also create custom attributes describing transformations for a derived variable with information explaining how you transformed the variable.

Enhancements in Version 19

IBM SPSS Statistics Base 19 includes major performance enhancements, improved procedures and new ease-of-use features, including:

- Faster Performance – Up to 200% faster performance for creating reports with large tables or a large number of smaller tables

- Automatic Linear Models - An easy, automated way to build powerful linear models

- Syntax Editor - More than a dozen enhancements to make writing syntax faster and easier

- Data Scoring - Easily score new customer data, access pre-built models and interface directly with data in Salesforce.com

You can open multiple datasets within a single session. This enables you to save time and condense steps when merging data files. It also helps you maintain consistency when copying data dictionary information between multiple files. Or, if you prefer, you can suppress the number of active datasets.

IBM SPSS Statistics Base enables you to restructure your data files to prepare them for analysis. For example, take a data file that has multiple cases per subject and restructure the data to put all data for each subject into a single record. You also can complete the reverse action – you can take a data file that has a single case per subject and spread the data across multiple cases.

Use the Date and Time Wizard to make calculations with dates and times, create date/time variables from strings containing date variables (such as "03/29/10") and bring date/time data from a variety of sources into Base. You can also parse individual date/time units, such as year, from date/time variables to apply filters.

Analyze data with comprehensive techniques

Go beyond summary statistics and row-and-column math. Base gives you a wide range of statistical procedures for basic analysis, including counts, crosstabs, cluster, descriptives, factor analysis, linear regression, cluster analysis, ordinal regression and nearest neighbor analysis.

Once you complete your analysis, you can write data back to your database with ease by using the Export to Database Wizard. For even more analytical power, use Base with other modules, such as IBM® SPSS® Regression*** and IBM® SPSS® Advanced Statistics***, that focus on data analysis (details start on page 17).

Build charts more easily with sophisticated reporting capabilities

Create commonly used charts, such as SPLOMs (scatterplot matrices), histograms and population pyramids more easily with Chart Builder. This highly visual chart creation interface enables you to create a chart by dragging variables and elements onto a chart creation canvas. Optionally, use a shortcut method based on an existing chart in the Gallery. You will see a limited preview of the chart as it is being built. Advanced users can attain a broader range of chart and option possibilities by using the Graphics Production Language (GPL).

Those working with statistical process control charts can request rule-checking on both primary and secondary control charts, which provides greater accuracy and a better understanding of whether a process is operating normally.

***IBM SPSS Regression and IBM SPSS Advanced Statistics were formerly called PASW® Regression and PASW® Advanced Statistics.

The presentation graphics system in IBM SPSS Statistics Base gives you control at both the creation and edit stages, to help ease your workload in a production setting. Create a chart once, and then use your specifications to create hundreds more just like it.

Present your best results with report OLAP

OLAP technology transforms the way you create and share information. Report OLAP in Base provides you with a fast, flexible way to create, distribute, and manipulate information for ad hoc decision-making. Create tables, graphs and report cubes that feature unique, award-winning pivoting technology and enable you to discover new insights in your data. Swap rows, columns and layers of report cubes – or quickly change information and statistics in graphs – for new levels of understanding. You can even convert a table to a graph with just a few mouseclicks.

Custom Dialog Builder

The Custom Dialog Builder enables more experienced users to make existing dialogs easier for business users, and create dialogs for custom features built through programmability. The Custom Dialog Builder enables your organization's less experienced users to quickly learn how to perform routine operations efficiently and gives programmers a way to deploy their work effectively.

Gain greater value with collaboration

To share and re-use assets efficiently, protect them in ways that meet internal and external compliance requirements, and publish results so that a greater number of business users can view and interact with them, consider augmenting your IBM SPSS Statistics software with IBM® SPSS® Collaboration and Deployment Services (formerly PASW® Collaboration and Deployment Services).

More information about these valuable capabilities can be found at **www.spss.com/software/deployment/cds**.

Features
General operations

- Switch user interface language (for example, switch between English and Japanese)
- Apply splitters through the Data Editor to more quickly and easily understand wide and long datasets
- Select the customizable toolbar feature to:
  – Assign procedures, scripts or other software products
  – Select from standard toolbar icons or create your own
- Work with multidimensional pivot tables/report cubes to:
  – Rearrange columns, rows and layers by dragging icons for easier ad hoc analyses
  – Toggle between layers by clicking on an icon for easier comparison between subgroups
  – Enable online statistical help for choosing statistical procedures or chart types and interpreting results; realistic application examples are included
- Change text attributes such as fonts, colors, bolding, italics and others
- Change table attributes such as number formats, line styles, line width, column alignments, background/foreground shading, enable or disable lines, and more
- Selectively display or hide rows, columns or labels to highlight important findings
- Enable task-oriented help with step-by-step instructions:
  – View case studies that show you how to use selected statistics and interpret results
  – Select the Statistics Coach™, which helps you choose the best statistical procedure or graph
  – Work through tutorials

  – Select "Show Me" buttons, which link to the tutorial for more in-depth help when you need it
  – Use "What's This?" help, which provides pop-up definitions of statistical terms and rules of thumb
- Use formatting capabilities for output in order to:
  – Transform a table into a graph for more visually compelling communication
  – Show correlation coefficients together with their significance level (as well as n) in correlations using the default output display
  – Control whether, upon activation, a table is opened in place or in its own window
  – Stamp date and time into the journal file for easy reference
  – Right-click on an IBM SPSS Statistics syntax file icon to run a command file without needing to go through production mode
  – Use drop-down lists for easier access to different layers
  – Set permanent page settings
  – Set a column width for all pivot tables and define text wrapping
  – Choose whether to use scientific notation to display small numbers
  – Control number of digits of precision in presentations
  – Add footnotes and annotations
  – Reorder categories within a table to display results most effectively
  – Group or ungroup multiple categories in rows or columns under a single heading that spans the rows or columns
  – Use one of 16 pre-formatted TableLooks™ for quick and consistent formatting of results

  – Create and save customized formats as TableLooks for your own personalized style
  – Display values or labels
  – Rotate table labels
  – Interact with reports and use models and code created by others in your organization with the optional addition of IBM SPSS Collaboration and Deployment Services
- Work with the Viewer to organize, view and move through results
  – Keep a record of your work using the "append" default in journal files
  – Use outline representation to quickly determine output location
  – Select an icon in the outline and see corresponding results displayed in the content pane
  – Reorder charts, tables and other objects by dragging icons in the outline
  – Selectively collapse or expand the outline to view or print selected results
  – Contain tables, charts and objects in a single content pane for easy review and access
  – Right-justify, left-justify or center output
  – Search and replace information in the Viewer of the contents pane, the outline pane, or both
- Create and save analysis specifications for repetitive tasks or unattended processing
- Use the enhanced production mode facility with dialog interface and macros for easier periodic reporting
- Have full control over table splitting with improved pagination and printing
- Refer to explanations of statistical terms through the on-screen statistical glossary

- Work with data more easily, thanks to:
  - Resizable dialog boxes
  - Drag-and-drop in dialogs
- Export output to Microsoft Word
  - Convert pivot tables to Word tables with all formatting saved
  - Convert graphics into static pictures
  - Wrapping and shrinking of wide tables
  - Syntax to automate report production
- Export output to PowerPoint® (Windows only)
  - Convert pivot tables to tables in PowerPoint with all formatting saved
  - Convert graphics into static pictures
  - Wrapping and shrinking of wide tables
  - Syntax to automate report production
  - Modify an existing worksheet by appending rows or columns
- Export output to Excel®
  - Export only the current view or all layers of an IBM SPSS Statistics pivot table
  - Place each pivot table layer on the same sheet or on separate sheets within one Excel workbook
  - Syntax to automate report production
  - Create a new worksheet within an existing workbook
  - Modify an existing worksheet by appending rows or columns

- Export output to PDF
  - Choose to optimize the PDF for Web viewing
  - Control whether PDF-generated bookmarks correspond to Navigator Outline entries in the Output Viewer. Bookmarks facilitate navigation of large documents.
  - Control whether fonts are embedded in the document. Embedded fonts ensure that the reader of your document sees the text in its original font, preventing font substitution.
  - Syntax to automate report production
- Easily open/save and create new output files through syntax
- Receive wheel mouse support for Output Viewer scroll
- Switch output languages (for example, switch between Japanese and English)
- Use the scripting facility to:
  - Create, edit and save scripts
  - Build customized form interfaces
  - Assign scripts to toolbar icons or menus
  - Automatically execute scripts whenever certain events occur
  - Support Python 2.6 to make scripting easier and more reliable
- Use automation to:
  - Integrate IBM SPSS Statistics with other desktop applications
  - Build custom applications using Visual Basic®, PowerBuilder® and C++
  - Integrate IBM SPSS Statistics into larger custom applications (such as Word or Excel)

- Use the HOST command to take advantage of the operating system functionality in IBM SPSS Statistics. This command enables applications to "escape" to the operating system and execute other programs in sync with the IBM SPSS Statistics session.
- Prevent syntax jobs from breaking when you create a common or main project directory that enables you to include transformations for multiple projects
  - Better manage multiple projects, syntax files and datasets
- Specify interactive syntax rules using the INSERT command
- Command syntax editor for the easy creation of syntax with new features such as:
  - Auto-completion
  - Color coding of syntax
  - Error coding of syntax
  - Gutter to display line numbers and break point
  - Stepping through of syntax jobs
  - Auto-indentation - button that automatically indents command contents.
  - Indent and outdent buttons
  - Uncomment/comment toggle
  - Ability to split the Syntax Editor window
  - Improved scrolling
- Custom Dialog Builder to create user-defined interfaces for existing procedures and user-defined procedures
- IBM SPSS Smartreader to share IBM SPSS Statistics output with those who do not have IBM SPSS Statistics

Graphic capabilities
- Categorical charts
  - 3-D Bar: Simple, cluster and stacked
  - Bar: Simple, cluster, stacked, drop- shadow and 3-D
  - Line: Simple, multiple and drop-line
  - Area: Simple and stacked
  - Pie: Simple, exploding and 3-D effect
  - High-low: High-low-close, difference area and range bar
  - Boxplot: Simple and clustered
  - Error bar: Simple and clustered
  - Error bars: Add error bars to bar, line and area charts; confidence level; standard deviation; and standard error
  - Dual-Y axis and overlay
- Scatterplots
  - Simple, grouped, scatterplot matrix and 3-D
  - Fit lines: Linear, quadratic or cubic regression, and Lowess smoother; confidence interval control for total or subgroups; and display spikes to line
  - Bin points by color or marker size to overlap
- Density charts
  - Population pyramids: Mirrored axis to compare distributions; with or without normal curve
  - Dot charts: Stacked dots show distribution; symmetric, stacked, and linear
  - Histograms: With or without normal curve; custom binning options
- Quality control charts
  - Pareto
  - X-Bar
  - Range
  - Sigma
  - Individuals
  - Moving range

  - Control chart enhancements include automatic flagging of points that violate Shewhart rules, the ability to turn off rules and the ability to suppress charts
- Rule-checking on secondary SPC charts
- Diagnostic and exploratory charts
  - Caseplots and time-series plots
  - Probability plots
  - Autocorrelation and partial autocorrelation function plots
  - Cross-correlation function plots
  - Receiver-Operating Characteristics (ROC)
- Multiple use charts
  - 2-D line charts (both axes can be scale axes)
  - Charts for multiple response sets
- Custom charts
  - Graphics Production Language (GPL), a custom chart creation language, enables advanced users to attain a broader range of chart and option possibilities than the interface supports
- Graphboard integration allows graph templates created in IBM SPSS Vizualization Designer to be accessed through IBM SPSS Statistics Base
- Editing options
  - Automatically reorder categories in differing order (descending or ascending) or by different sort methods (value, label or summary statistic)
  - Create data value labels
  - Drag to any position on your chart, add connecting lines, and match font color to subgroup
  - Select and edit specific elements directly within a chart: Colors, text and styles
  - Choose from a wide range of line styles and weights

  - Display gridlines, reference lines, legends, titles, footnotes and annotations
  - Include an Y=X reference line
- Layout options
  - Paneled charts: Create a table of subcharts, one panel per level or condition, showing multiple rows and columns
  - 3-D effects: Rotate, modify depth and display backplanes
- Chart templates
  - Save selected characteristics of a chart and apply them to others automatically. You can apply the following attributes either at creation or when editing: Layout, titles, footnotes and annotations, chart element styles, data element styles, axis scale range, axis scale settings, fit and reference lines and scatterplot point binning
  - Tree-view layout and finer control of template bundles
- Graph export: BMP, EMF, EPS, JPG, PCT, PNG, TIF and WMF
- iGRAPH conversion utility for opening files in Version 15.0 and earlier

Analysis
Descriptive statistics

Reports
- OLAP cubes enable you to:
  - Quickly estimate changes in the mean or sum between any two related variables using percent change. For example, easily see how sales increase from quarter to quarter.
  - Create case summaries
  - Create report summaries
  - Generate presentation-quality reports using numerous formatting options
  - Generate case listing and case summary reports with statistics on break groups

Codebook
- Control the variable information included in the results: position, label, type, format, measurement level, value labels, missing values, custom attributes, reserved attributes
- Control the file information order in the results: name, location, number of cases, file label, user defined custom attributes, data file document text, weight status, reserved data file attributes
- Control summary statistics: number of cases in each category, percent of cases in each category, mean, standard deviation, quartile
- Control display order: file order, alphabetic order by variable name, order in which the variables and multiple response sets are listed in the command, measurement level, user-defined custom attribute name and value

Frequencies
- Frequency tables: Frequency counts, percent, valid percent and cumulative percent
- Option to order your output by analysis or by table
- More compact output tables by eliminating extra lines of text where they're not needed
- Central tendency: Mean, median, mode and sum
- Dispersion: Maximum, minimum, range, standard deviation, standard error and variance
- Distribution: Kurtosis, kurtosis standard error, skewness and skewness standard error
- Percentile values: Percentiles (based on actual or grouped data), quartiles and equal groups
- Format display: Condensed or standard, sorted by frequency or values, or index of tables
- Charts: Bar, histogram or pie chart

Descriptives
- Central tendency: Mean and sum
- Dispersion: Maximum, minimum, range, standard deviation, standard error and variance
- Distribution: Kurtosis and skewness
- Z scores: Compute and save as new variables
- Display order: Ascending or descending order on means and variable name

Explore
- Confidence intervals for mean
- Descriptives: Interquartile range, kurtosis, kurtosis standard error, median, mean, maximum, minimum, range, skewness, skewness standard error, standard deviation, standard error, variance, five percent trimmed mean and percentages
- M-estimators: Andrew's wave estimator, Hampel's M-estimator, Huber's M-estimator and Tukey's biweight estimator
- Extreme values and outliers identified
- Grouped frequency tables: Bin center, frequency, percent, valid and cumulative percent
- Plots: Construct plots with uniform scale or dependence on data values
  - Boxplots: Dependent variables and factor levels together
  - Descriptive: Histograms and stem-and-leaf plots
  - Normality: Normal probability plots and detrended probability plots with Kolmogorov-Smirnov and Shapiro-Wilk statistics
  - Spread versus level plots using Levene's test: Power estimation, transformed or untransformed
  - Shapiro-Wilk test of normality in EXAMINE allows for 5,000 cases when weights are not specified

Crosstabs
- Three-way relationships in categorical data with Cochran's and Mantel-Haenszel statistics allow you to go beyond the limits of a two-way crosstab
- Counts: Observed and expected frequencies
- Percentages: Column, row and total
- Long string variables
- Residuals: Raw, standardized and adjusted standardized
- Marginals: Observed frequencies and total percentages
- Tests of independence: Pearson and Yates corrected Chi-square, likelihood ratio Chi-square and Fisher's exact test
- Test of linear association: Mantel-Haenszel Chi-square
- Measure of linear association: Pearson r
- Nominal data measures: Contingency coefficient, Cramer's V, Phi, Goodman and Kruskal's Lambda (asymmetric and symmetric), Tau (column or row dependent) and uncertainty coefficient (asymmetric and symmetric)
- Ordinal data measures: Goodman and Kruskal's Gamma, Kendall's Tau-b and Tau-c, Somers' D (asymmetric and symmetric) and Spearman's Rho
- Nominal by interval measure: Eta
- Measure of agreement: Cohen's Kappa
- Relative risk estimates for case control and cohort studies
- Display tables in ascending or descending order
- Frequency counts written to file
- McNemar's test
- Option to use integer or non-integer weights

Descriptive ratio statistics
- Help for understanding your data using:
  – Coefficient of dispersion
  – Coefficient of variation
  – Price-related differential (PRD)
  – Average absolute deviance

Compare means

Means
- Create better models with harmonic and geometric means
- Cells: Count, mean, standard deviation, sum and variance
- All-ways totals
- Measure of analysis with Eta and Eta^2
- Test of linearity with R and R^2
- Results displayed in report, crosstabular or tree format
- Statistics computed for total sample *t* test
- One sample *t* test to compare sample mean to a reference mean of your choice
- Independent sample statistics: Compare sample means of two groups for both pooled and separate-variance estimates with Levene's test for equal variances
- Paired sample statistics: Correlation between pairs, difference between means, and two-tailed probability for test of no difference and for test of zero correlation between pairs
- Statistics: Confidence intervals, counts, degrees of freedom, mean, two-tailed probability, standard deviation, standard errors, and *t* statistic

One-way ANOVA
- Contrasts: Linear, quadratic, cubic, higher-order and user-defined
- Range tests: Duncan, LSD, Bonferroni, Student-Newman-Keuls, Scheffé, Tukey's alternate test, and Tukey's HSD

- Post hoc tests: Student-Newman-Keuls, Tukey's honestly significant difference, Tukey's *b*, Duncan's multiple comparison procedure based on the Studentized range test, Scheffé's multiple comparison *t* test, Dunnett's two-tailed *t* test, Dunnett's one-tailed *t* test, Bonferroni t test, least significant difference *t* test, Sidak *t* test, Hochberg's GT2, Gabriel's pairwise comparisons test based on the Studentized maximum modulus test, Ryan-Einot-Gabriel-Welsch's multiple stepdown procedure based on an F test, Ryan-Einot-Gabriel-Welsch's multiple stepdown procedure based on the Studentized range test, Tamhane's T^2, Tamhane's T^3, Games and Howell's pairwise comparisons test based on the Studentized range test, Dunnett's C, and Waller-Duncan *t* test
- ANOVA statistics: Between- and within-groups sums of squares, degrees of freedom, mean squares, F ratio, and probability of F
- Fixed-effects measures: Standard deviation, standard error and 95 percent confidence intervals
- Random effects measures: Estimate of variance components, standard error, and 95 percent confidence intervals
- Group descriptive statistics: Maximum, mean, minimum, number of cases, standard deviation, standard error and 95 percent confidence interval
- Homogeneity of variance test: Levene's test
- Read and write matrix materials
- Equality of means: Reach accurate results when variances and sample sizes vary across different groups
  – Brown-Forsythe test
  – Welch test

ANOVA models—simple factorial
- Create custom models without limits on maximum order of interaction

- Work faster because you don't have to specify ranges of factor levels
- Choose the right model using four types of sum of squares
- Increase certainty with better data handling in empty cells
- Perform lack-of-fit tests to select your best model
- Choose from one of two designs: Balanced or unbalanced
- Use analysis of covariance with up to 10 covariate methods: Classic experimental, hierarchical and regression
- Enter covariates control: Before, with or after main effects
- Set interaction to: None, 2-, 3-, 4- or 5-way
- Select from the following statistics: ANOVA, means and counts table, multiple classification analysis, unstandardized regression coefficients and n-way cell means
- Choose up to 10 independent variables
- Reach predicted values and deviations from the mean in MCA table

Correlate†

Bivariate
- Pearson r, Kendall's Tau-b, and Spearman
- One- and two-tailed probabilities
- Means, number of non-missing cases, and standard deviations
- Cross-product deviations and covariances
- Coefficients displayed in matrix or serial format

Partial†
- One- and two-tailed probabilities
- Mean, number of non-missing cases and standard deviation
- Zero-order correlations
- Up to 100 control variables
- Up to five order values
- Correlations displayed in matrix or serial string format, lower triangular or rectangular correlation matrix

† *Multithreaded algorithm, resulting in improved performance and scalability on multiprocessor or multicore machines.*

Distances
- Compute proximities between cases or variables
- Dissimilarity measures
  - Interval measure: Euclidean and squared Euclidean distance, Chebychev distance metric, city-block or Manhattan distance, distance in an absolute Minkowski power metric and customized
  - Counts measures: Chi-square and Phi-square
  - Binary measures: Euclidean and squared Euclidean distance; size, pattern and shape difference; variance dissimilarity measure; and Lance and Williams nonmetric
- Similarity measures
  - Interval measures: Pearson correlation and cosine
  - Binary measures: Russell and Rao; simple matching; Jaccard; dice (or Czekanowski or Sorenson); Rodgers and Tanimoto; Sokal and Sneath 1 through 5; Kulczynski 1 and 2; Hamann; Goodman and Krusal Lambda; Anderberg's D; Yule's coefficient of colligation; Yule's Q; Ochiai; dispersion similarity measure; and fourfold point correlation
- Standardize data values: Z scores, range of -1 to 1, range of 0 to 1, maximum magnitude of 1, mean of 1 and standard deviation of 1
- Transform measures: Absolute values, dissimilarities into similarities, similarities into dissimilarities and rescale proximity values to a range of 0 to 1
- Identification variable specification
- Printed matrix of proximities between items
- Improved scalability for proximities between variable matrices

Automatic Linear Modeling (ALM)
Automates the prediction of numeric outcomes
- Automated data preparation to improve predictive power
- Boosting to enhance accuracy
- Bagging to enhance stability
- Interactive visual output
- Variable selection algorithms such as best subsets and forward stepwise
- Improved performance when building models on very large datasets. Data passes are reduced by building models on subsets of the data which are then combined. (Statistics Server only)

Regression—linear regression[†]
- Methods: Backward elimination, forced entry, forced removal, forward entry, forward stepwise selection, and R^2 change/test of significance
- Equation statistics: Akaike information criterion (AIC), Ameniya's prediction criterion, ANOVA tables (F, mean square, probability of F, regression, and residual sum of squares), change in R^2, F at step, Mallow's Cp, multiple R, probability of F, R^2, adjusted R^2, Schwarz Bayesian criterion (SBC), standard error of estimate, sweep matrix and variance-covariance matrix
- Descriptive statistics: Correlation matrix, covariance matrix, cross-product deviations from the mean, means, number of cases used to compute correlation coefficients, one-tailed probabilities of correlation coefficients, standard deviations and variances
- Independent variable statistics: Regression coefficients, including B; standard errors of coefficients, standardized regression coefficients, approximate standard error of standardized regression coefficients and t; tolerances; zero-order; part and partial correlations; and 95 percent confidence interval for unstandardized regression coefficient

- Variables not in equation: Beta or minimum tolerance
- Durbin-Watson
- Collinearity diagnostics: Condition indexes, eigenvalues, variance inflation factors, variance proportions and tolerances
- Plots: Casewise, histogram, normal probability, detrended normal, partial, outlier and scatterplots
- Create and save variables:
  - Prediction intervals: Mean and individual
  - Predicted values: Unstandardized, standardized, adjusted and standard error of mean
  - Distances: Cook's distances, Mahalanobis' distance and leverage values
  - Residuals: Unstandardized, standardized, Studentized, deleted and Studentized deleted
  - Influence statistics: dfbetas, standardized dfbetas, dffits, standardized dffits and covariance ratios
- Option controls: F-to-enter, F-to-remove, probability of F-to-enter, probability of F-to-remove, suppress the constant, regression weights for weighted least-squares model, confidence intervals, maximum number of steps, replace missing values with variable mean and tolerance
- Regression coefficients displayed in user-defined order
- System files can contain parameter estimates and their covariance and correlation matrices through the OUTFILE command
- Solutions can be applied to new cases or used in further analysis
- Decision making can be further improved throughout your organization when you export your models via XML

[†] *Multithreaded algorithm, resulting in improved performance and scalability on multiprocessor or multicore machines.*

Ordinal regression—PLUM †
- Predict ordinal outcomes
  - Seven options to control the iterative algorithm used for estimation, to specify numerical tolerance for checking singularity and to customize output
  - Five link functions to specify the model: Cauchit, complementary log-log, logit, negative log-log and probit
  - Location subcommand to specify the location model: Intercept, main effects, interactions, nested effects, multiple-level nested effects, nesting within an interaction, interactions among nested effects and covariates
  - Print: Cell information, asymptotic correlation matrix of parameter estimates, goodness-of-fit statistics, iteration history, kernel of the log-likelihood function, test of parallel lines assumption, parameter statistics and model summary
  - Save casewise post-estimation statistics into the active file: Expected probabilities of classifying factor/covariate patterns into response categories and response categories with the maximum expected probability for factor/covariate patterns
  - Customize your hypotheses tests by directly specifying null hypotheses as linear combinations of parameters using the TEST subcommand (syntax only)

Curve estimation
- Eleven types of curves are available for specification

- Regression summary displays: Curve type, R^2 coefficient, degrees of freedom, overall F test and significance level and regression coefficients
- Trend-regression models available: Linear, logarithmic, inverse, quadratic, cubic, compound, power, S, growth, exponential and logistic

Nonparametric tests
The tests listed below have been enhanced to allow multiple comparisons and to operate on large datasets more efficiently.

- Chi-square: Specify expected range (from data or user-specified) and frequencies (all categories equal or user-specified)
- Binomial: Define dichotomy (from data or cutpoint) and specify test proportion
- Runs: Specify cutpoints (median, mode, mean or specified)
- One sample: Kolmogorov-Smirnov, uniform, normal and Poisson
- Two independent samples: Mann-Whitney U, Kolmogorov-Smirnov Z, Moses extreme and Wald-Wolfowitz runs
- k-independent samples: Kruskal-Wallis H and median
- 2-related samples: Wilcoxon, sign and McNemar
- k-related samples: Friedman, Kendall's W and Cochran's Q
- Descriptives: Maximum, mean, minimum, number of cases and standard deviation

Multiple response
- Crosstabulation tables: Cell counts, cell percentages based on cases or responses, column and row and two-way table percentages
- Frequency tables: Counts, percentage of cases or responses
- Both multiple-dichotomy and multiple- response groups can be handled

Data reduction

Factor †
- Number of cases and variable labels for analysis can be displayed
- Input from correlation matrix, factor, loading matrix, covariance matrix or raw data case file
- Output of correlation matrix or factor matrix
- Seven extraction methods available for use when analysis is performed on correlation matrices or raw data files: Principal component, principal axis, Alpha factoring, image factoring, maximum likelihood, unweighted least squares and generalized least squares
- Rotation methods: Varimax, equamax, quartimax, promax and oblimin
- Display: Initial and final communalities, eigenvalues, percent variance, unrotated factor loadings, rotated factor pattern matrix, factor transformation matrix, factor structure and correlation matrix (oblique rotations only)
- Covariance matrices can be analyzed using three extraction methods: Principal component, principal axis and image
- Factor scores: Regression, Bartlett and Anderson-Rubin
- Factor scores saved as active variables
- Statistics available: Univariate correlation matrix, determinant and inverse of correlation matrix, anti-image correlation and covariance matrices, Kaiser-Meyer-Olkin measure of sampling adequacy, Bartlett's test of sphericity, factor pattern matrix, revised communalities, eigenvalues and percent variance by eigenvalue, reproduced and residual correlations and factor score coefficient matrix
- Plots: Scree plot and plot of variables in factor space
- Matrix input and output

† *Multithreaded algorithm, resulting in improved performance and scalability on multiprocessor or multicore machines.*

- Post-rotational calculated through sum-of-squares loadings
- Solutions applied to new cases or to use in further analysis with the SELECT subcommand
- Factor score coefficient matrix exported to score new data (syntax only)

Classify

TwoStep cluster analysis
- Group observations into clusters based on a nearness criterion. This procedure uses a hierarchical agglomerative clustering procedure in which individual cases are successively combined to form clusters whose centers are far apart. This algorithm is designed to cluster large numbers of cases. It passes the data once to find cluster centers and again to assign cluster memberships. Cluster observations by building a data structure called the CF Tree, which contains the cluster centers. The CF Tree is grown during the first stage of clustering and values are added to its leaves if they are close to the cluster center of a particular leaf.
  - Categorical-level and continuous-level data can be used
  - Distance measures: Euclidean distance and the likelihood distance
  - Criteria command tunes the algorithm so that:
    - The initial threshold can be specified to grow a CF Tree
    - The maximum number of child nodes a leaf node may have can be set
    - The maximum number of levels a CF Tree may have can be set

- HANDLENOISE subcommand enables you to treat outliers in a special manner during clustering. The default value of noise percent is zero, equivalent to no noise handling. The value can range between zero and 100.
- INFILE subcommand allows the algorithm to update a cluster model in which a CF Tree is saved as an XML file using the OUTFILE subcommand
- MEMALLOCATE subcommand specifies the maximum amount of memory in megabytes (MB) that the cluster algorithm should use
- Missing data: Exclude both user-missing and system-missing values, or let user-missing values be treated as valid
- Option to standardize continuous-level variables or leave them at the original scale
- Ability to specify the number of clusters, specify the maximum number of clusters or let the number of clusters be chosen automatically
  - Algorithms available for determining the number of clusters: BIC or AIC
- Output written to a specified filename as XML
- Final model output saved, or use an option that updates the model later with more data
- Plots:
  - Bar chart of frequencies for each cluster
  - Pie chart showing observation percentages and counts within each cluster
  - Importance of each variable within each cluster: The output is sorted by the importance rank of each variable

- Plot options:
  - Comparisons (one plot per cluster or one plot per variable)
  - Measure of variable importance (parametric or non-parametric)
  - Ability to specify Alpha level when considering importance
- Print options:
  - AIC or BIC for different numbers of clusters
  - Two tables describing the variables in each cluster. In one table, means and standard deviations are reported for continuous variables. The other table reports frequencies of categorical variables. All values are separated by cluster.
  - List of clusters and number of observations in each cluster
- Cluster number saved for each case to the working data file

Cluster
- Use one of six linkage methods to determine clusters: Single linkage (nearest neighbor), average linkage between groups, centroid (average linkage within groups), complete linkage (farthest neighbor), median and Ward
- Provide the same set of similarity and dissimilarity measures as in proximity
- Save cluster memberships as new variables
- Save distance matrices for use in other procedures
- Display: Agglomeration schedules, cluster membership and distance matrices
- Use proximities between variable matrices for improved scalability
- Choose from the following plots: Horizontal and vertical icicle plots and dendrogram plots of cluster solutions

- Specify case identifiers for tables and plots
- Have the ability to accept matrix input and produce matrix output

Quick cluster
- Squared Euclidean distance
- Centers selected by widely spaced cases, first K cases or direct specification
- Cluster membership saved as a variable
- Two methods provided for updating cluster centers
- K-means clustering algorithms

Nearest Neighbor analysis
- Can be used for prediction (outcome specified) or classification (no outcome specified)
- Mark cases of specific interest
- Rescale covariates
- Use three methods of partitioning the active dataset into training and holdout samples: specify the relative number of cases in the active dataset to randomly assign to the training sample; specify the relative number of cases in the active dataset to randomly assign to the holdout sample; specify a variable that assigns each case in the active dataset to the training or holdout sample.
- Specify the nearest neighbor "model"
    - Specify the distance metric used to measure the similarity of cases
    - Whether to use automatic selection of the number of nearest neighbors
    - Whether to use automatic selection of features (predictors)
- Specify computational and resource settings for the KNN procedure; in particular:
    - How automatic feature selection should select the number of features
    - The function used to compute the predicted value of scale response variables.

- Whether to weight features by their normalized importance when computing distances
- Specify settings for performing v-fold cross-validation to determine the "best" number of neighbors
- Control whether user-missing values for categorical variables are treated as valid values
- Control options for display of model-related output, including tables and charts
- Write optional temporary variables to the active dataset
- Save an XML-format file containing the nearest neighbor model. It also saves an IBM SPSS Statistics-format data file containing distances from focal cases.

Discriminant
- Variable selection methods: Direct entry, Wilks' Lambda minimization, Mahalanobis' distance, smallest F ratio, minimization of sum of unexplained variation for all pairs and largest increase in Rao's V
- Statistics:
    - Summary: Eigenvalues, percent and cumulative percent of variance, canonical correlations, Wilks' Lambda and Chi-square tests
    - At each step: Wilks' Lambda, equivalent F, degrees of freedom and significance of F for each step; F-to-remove; tolerance; minimum tolerance; F-to-enter; and value of statistic for each variable not in equation
    - Final: Standardized canonical discriminant function coefficients, structure matrix of discriminant functions and functions evaluated within group means

- Optional: Means, standard deviations, univariate F ratios, pooled within-groups covariance and correlation matrices, matrix of pairwise F ratios, Box's M test, group and total covariance matrices, unstandardized canonical discriminant functions, classification results table and classification function coefficients
- Rotation of coefficient (pattern) and structure matrices
- Output displayed step by step and/or in summary form
- In classification stage: Prior probabilities, equal, proportion of cases or user-specified All groups, cases, territorial maps and separate groups plotted
- Casewise results saved to system file for further analysis
- Matrix files read/written, including additional statistics: Counts, means, standard deviations and Pearson correlation coefficients
- Solutions applied to new cases or for use in further analysis
- Jacknife estimates provided for misclassified error rate
- Decision-making further improved by exporting your models throughout your organization via XML

Scaling
- Reduce your data and improve measurement with reliability
- Find the hidden structure in your similarity data using ALSCAL multidimensional scaling

Matrix operations
- Write your own statistical routines in the compact language of matrix algebra

Data management
- New! Use the default measurement feature to identify the type of data in your dataset

- Prepare continuous-level data for analysis with the Visual Binner
  - Specify cutpoints in an intelligent manner using a histogram created through a data pass
  - Automatically create value labels based on your cutpoints
  - Copy bins to other variables
- Create your own custom programs with the Output Management System (OMS). Turn output from IBM SPSS Statistics procedures into data (IBM SPSS Statistics data files, XML or HTML) and create your programs for bootstrapping, jacknifing and leaving-one-out methods and Monte Carlo simulations
  - Create custom programs in IBM SPSS Statistics, even if you have little or no experience with IBM SPSS Statistics syntax, using the Output Management System Control Panel
- Easily clean your data when you identify duplicate records through the user interface with the Identify Duplicate Cases tool
- Make sense and keep track of your data files by adding notes to them with the Data File Comments command
- Prevent the accidental destruction of data by making the dataset read-only
- Easily set up all of your value labels to prepare your data for analysis using the Define Variable Properties tool
  - Set up data dictionary information, including value labels and variable types
  - Intelligently add labels because a data pass made first enables IBM SPSS Statistics to present a list of values and counts of those values
  - Save time by being able to enter data and value labels directly onto the grid rather than having to use nested dialogs

- Save work by easily copying dictionary information from one variable to another and from one dataset to another using the Copy Data Properties tool
  - Copy dictionary information (such as variable and value labels) between variables and datasets using the template facility
  - Receive a ready means of cloning dictionaries
- Analyze more data, more efficiently – file size considerations are practically eliminated (especially when used in conjunction with the optional IBM SPSS Statistics Base Server) Assign like variable attributes to multiple variables simultaneously
- Easily select rows and columns to paste information elsewhere
- Easily reorder your variables
- Save time by sorting data directly in the Data Editor
- Avoid reformatting column widths for each new session
- Increase speed by creating customized keyboard options
- Restructure data files that have multiple cases per subject and restructure data to put all data for each subject into a single record (restructure data files from a univariate form to a multivariate form)
- Restructure data files that have a single case per subject and spread data across multiple cases (restructure data files from a multivariate form to a univariate form)
- When saving data files, keep variables using an intuitive graphical interface
- Identify and select variables using your own organization scheme as you sort variables according to variable labels in a list box
- Display variable labels in a dialog; use up to 256 characters
- Display variable labels as a tool tip in the Data Editor
- Save SQL queries for later use
- Create prompted queries

- Select data more easily using the "where" clause
- Set any character or combination of characters as the delimiter between fields in an ASCII text file
- Create your own dictionary information for variables by using Custom Attributes. For example, create a custom attribute describing transformations for a derived variable with information explaining how it was transformed.
- Customize the viewing of extremely wide files with Variable Sets. You can instantly reduce the variables shown in the Variable View and Data View windows to a subset while keeping the entire file loaded and available for analysis.
- Write IBM SPSS Statistics data files from within other applications, such as Excel, using the IBM SPSS Statistics ODBC driver
- Use virtually unlimited numbers of variables and cases
- Specify and work with subsets of variables
- Enter, edit and browse data in the Data Editor's spreadsheet format
- Easily work with dates and times using the Date and Time Wizard
  - Create a date/time variable from a string containing a date/time variable
  - Create a date/time variable from variables that include individual date units, such as month or year
  - Parse individual date/time units from date/time variables
  - Calculate with dates and times
    - Round instead of truncating date/time information, if desired
    - Add decimal places to time data, if desired
- Display values or value labels in Data Editor cells

- With a right mouseclick, receive direct access to variable information within dialog boxes
- Rename and reorder variables
- Sort cases
- Choose from several data formats: Numeric, comma, dot, scientific notation, date, dollar, custom currency and string
- Set an option to show currency as comma- or decimal-delimited
- Choose system missing and up to three user-defined missing values per variable
- Create value labels of up to 120 characters (double that of versions prior to SPSS 13.0)
- Create variable labels of up to 256 characters
- Insert and delete variables and cases
- Search for values of a selected variable
- Transpose working files
- Clone or duplicate datasets
- Apply an extended Variable Properties command to customize properties for individual users
- Aggregate data using an extensive set of summary functions
    - Save aggregated values directly to your active file
    - Aggregate by string for source variables (within the interface)
        ◦ Allow the use of long strings as a break variable (e.g., if gender is the break variable, then males and females aggregate separately)
        ◦ Allow the use of strings as the aggregated variable
- Split files to apply analyses and operations to subgroups
- Select cases either permanently or temporarily
- Process first n cases
- Select random samples of cases for analysis
- Select subsets of cases for analysis

- Weigh cases by values of a selected variable
- Specify random number seeds
- Rank data
- Use neighboring observations for smoothing, averaging, and differencing fast Fourier transformations and their inverse
- More accurately describe your data using longer variable names (up to 64 bytes)
    - Work more easily with data from databases and spreadsheets that include longer variable names than allowed in versions earlier than SPSS 12.0
- Ensure data containing longer text strings (up to 32,767 bytes) is not truncated or lost when working with open-ended question responses, data from other software that allows long text strings or other types of long text strings
- Find and replace information using the Data Editor
- Save time with spell checking of value labels and variable labels and text strings
- Easily inspect data dictionary information in the Variable View of the Data Editor, since you can configure (show only certain attributes) and sort by Variable name, by Type, by Format, etc.
- Easily navigate the Data View in the Data Editor by going directly to a variable
- Add missing values and value labels for strings of any length
- Change string length and variable type through syntax

File management
- Use Unicode when working with multi-lingual data, thus eliminating variability in data due to language-specific encodings. Save the data file either as a Unicode file or as a codepage file (for backwards

compatibility with earlier versions of IBM SPSS Statistics).
- Truly minimize data handling with conversion-free/copy-free data access in SQL databases. Save time by not needing to convert data into IBM SPSS Statistics format (especially when used in conjunction with the optional IBM SPSS Statistics Base Server)
- Set a permanent default starting folder
- Easily write back to databases from IBM SPSS Statistics by using the Database Wizard.
- For example, you can:
    - Create a new table and export it to your database
    - Add new rows to an existing table
    - Add new columns to an existing table
    - Export data to existing columns in a table
- Import data (including compound documents) from current versions of Excel without needing the Database Wizard
    - Read columns that contain mixed data types without any loss of data
    - Automatically read columns with mixed data types as string variables and read all values as valid string variables
- Open multiple datasets within a single IBM SPSS Statistics session or suppress the number of datasets in the user interface
- Directly import data from IBM SPSS Data Collection‡ products, including IBM SPSS Data Collection Web Interviews‡, and traditional market research products, including Quanvert™ *
- Export data from IBM SPSS Statistics to IBM SPSS Data Collection products*
- Import from OLE DB data sources without having to go through ODBC

---

‡   *IBM SPSS Data Collection and IBM SPSS Data Collection Web Interviews were formerly called PASW® Data Collection and PASW® Data Collection Interviewer Web.*

*   *Supported only on Windows platform*

- Read/write Stata® files
- Work more efficiently as you run multiple sessions on one desktop. For example, on lengthy jobs, you can use IBM SPSS Statistics in another session as long as the licenses are available.

Easily read and define ASCII data using a Text Wizard similar to the one provided in Excel

- Use text qualifiers to make reading in data even easier
- Increase the accuracy and repeatability of your syntax files with search and replace enhancements
- Read database tables using the Database Wizard
    – Drag-and-drop join support
- Export tables and text as ASCII output
- Save tables as HTML and charts as JPG formats to post IBM SPSS Statistics results on the Internet or your intranet
- Gain quick access to the IBM SPSS Developer Central Web site through the IBM SPSS Statistics Help menu
- Read/write Excel 2007 files
- Translate files to and from Excel, Lotus® 1-2-3® and dBASE®
- Read and write data to and from fixed, free-field or tab-delimited ASCII files
- Write data to fixed-format or tab-delimited ASCII files
- Read complex file structures: Hierarchical files, mixed record types, repeating data and non-standard file structures
- Read and write SPSS/PC+™ system files
- Merge files
- Display and apply data definitions from an IBM SPSS Statistics data file to a working file
- Update master files using transaction files
- Read and write data matrices

- Save many intermediate results for further analysis
- Read recent versions of SAS® files
- Export data files to SAS
- Export data files to current versions of Excel
- Save comma-separated value (CSV) text files from IBM SPSS Statistics data files
- "File in use" message to reduce errors in data created by more than one user writing to a IBM SPSS Statistics file at once

Transformations
- Compute new variables using arithmetic, cross-case, date and time, logical, missing-value, random-number, and statistical or string functions
- Create new variables that contain the values of existing variables from preceding or subsequent cases
- Count occurrences of values across variables
- Recode string or numeric values
- Automatically convert string variables to numeric variables using the autorecode command
    – Use an autorecode template to append existing recode schemes
    – Recode multiple variables simultaneously
    – Autorecode blank strings so that they are defined as "user-missing"
- Create conditional transformations using do if, else if, else and end if structures
- Use programming structures such as do repeat-end repeat, loop-end loop and vectors
- Make transformations permanent or temporary
- Execute transformations immediately, in batch mode or on demand
- Easily find and replace text strings in your data using the find/replace function

- Use cumulative distribution, inverse cumulative distribution and random number generator functions: Beta, Cauchy, Chi-square, Exponential, F, Gamma, Laplace, logistic, lognormal, Normal, Pareto, Student t, uniform and Weibull
    – Standard bivariate normal distribution with correlation r, Half Normal, inverse Gaussian, Studentized range and Studentized maximum modulus
- Work with cumulative distribution and the random number generator for discrete distribution functions: Bernoulli, binomial, geometric, hypergeometric, negative binomial and Poisson
- Use cumulative distribution for non-central distribution: Non-central Beta, non-central Chi-square, non-central F and non-central T
- Use density/probability functions for:
    – Continuous distributions: Beta, standard bivariate normal with correlation R, Cauchy, Chi-square, exponential, F, Gamma, half normal random, inverse Gaussian, Laplace, logistic, lognormal, normal, Pareto, Student t, uniform and Weibull
    – Discrete distributions: Bernoulli, binomial, geometric, hypergeometric, negative binomial and Poisson
- Use non-central density/probability functions for: Non-central Beta, non-central Chi-square, non-central F distribution and non-central t distribution
- Select two-tail probabilities: Chi-square and F
- Use auxiliary function: Logarithm of the complete Gamma function

System requirements

IBM SPSS Statistics Base 19 for Windows

- Operating system: Microsoft Windows XP (Professional, 32-bit), Vista® (Home, Business, 32- or 64-bit) or Windows 7 (32- or 64-bit)
- Hardware:
  - Intel® or AMD x86 processor running at 1 GHz or higher
  - Memory: 1 GB RAM or more recommended
  - Minimum free drive space: 800 MB***
  - CD-ROM drive
  - Super VGA (800 x 600) or higher-resolution monitor
- For connecting with Statistics Base Server, a network adapter running the TCP/IP network protocol
- Web browser: Internet Explorer 6 or 7

IBM SPSS Statistics Base 19 for Mac

- Operating system: Apple Mac OS 10.5 (Leopard™) or 10.6 (Snow Leopard) (32-bit or 64-bit versions)
- Hardware:
  - Intel processor
  - Memory: 1 GB RAM or more recommended
  - Minimum free drive space: 800 MB***
  - CD-ROM drive
  - Super VGA (800 x 600) or higher-resolution monitor
- Web browser: Safari™
- Java™ Standard Edition 5.0 (J2SE 5.0)

IBM SPSS Statistics Base 19 for Linux

- Operating system: Any Linux OS that meets the following requirements (32-bit only):
  - Kernel 2.6.3
  - glibc 2.8
  - libstdc++5
  - XFree86-4.7

- Hardware:
  - Processor: Intel or AMD x86 processor running at 1 GHz or higher
  - Memory: 1 GB RAM or more recommended
  - Minimum free drive space: 800 MB***
  - CD-ROM drive
  - Super VGA (800 x 600) or a higher-resolution monitor
- Web browser: Mozilla® Firefox®

Enterprise products

IBM SPSS Statistics Server†

IBM SPSS Statistics Server enables Statistics users in your organization to work with large data files for better decision making. The client/server version delivers additional capabilities, enterprise-strength scalability and enhanced performance. For even greater scalability and security, it is available on IBM System z® running Linux.

IBM SPSS Statistics Family

Add more analytical power, as you need it, with optional modules and standalone software from the Statistics family.

IBM SPSS Direct Marketing†

IBM SPSS Direct Marketing helps marketers perform various kinds of analyses easily and confidently, without requiring a detailed understanding of statistics. They can conduct recency, frequency and monetary value (RFM) analysis, cluster analysis, and prospect profiling. They can also improve marketing campaigns through postal code analysis, propensity scoring, and control package testing. And they can easily score new customer data, access pre-built models, and interface directly with data in Salesforce.com (these capabilities are also available in Base.)

IBM SPSS Bootstrapping†

Bootstrapping enables researchers and analysts to use bootstrapping techniques on a number of tests contained in Statistics modules. This provides an efficient way to ensure that your models are stable and reliable. With Bootstrapping, you can reliably estimate the standard errors and confidence intervals of a population parameter like a mean, median, proportion, odds ratio, correlation coefficient, regression coefficient and numerous.

IBM SPSS Statistics Programmability Extension†

Expanded programmability functionality helps make Statistics one of the most powerful statistical development platforms. You can use the external programming language Python® to develop new procedures and applications, including those written in R. You'll enjoy improved tools for adding these procedures, namely a new user interface and the ability to deliver results to pivot tables in the IBM SPSS Output Viewer. Visit IBM SPSS Developer Central at www.spss.com/devcentral to share code, tools, and programming ideas.

IBM SPSS Regression†

Predict behavior or events when your data go beyond the assumptions of linear regression techniques. Perform multinomial or binary logistic regression and nonlinear regression, weighted least squares, two-stage least squares and probit analysis.

IBM SPSS Advanced Statistics†

Advanced Statistics' powerful multivariate techniques include generalized linear models (GENLIN), generalized estimating equations (GEE), mixed level models, general linear mixed models (GLMM),

*** Installing Help in all languages requires 1.1 GB free drive space

variance component estimation, MANOVA, Kaplan-Meier estimation, Cox regression, hiloglinear, loglinear and survival analysis.

IBM SPSS Custom Tables[†]

Use Custom Tables to present survey, customer satisfaction, polling and compliance reporting results. Features such as a table builder preview, included inferential statistics and data management capabilities make it easy to clearly communicate your results.

IBM SPSS Decision Trees[†]

Create highly visual classification and decision trees directly within Statistics for segmentation, stratification, prediction, data reduction and variable screening, interaction identification, category merging and discretizing continuous variables. Highly visual trees enable you to present results in an intuitive manner.

IBM SPSS Exact Tests (now available for Windows, Mac and Linux)[†]

Exact Tests always provides you with correct p values, regardless of your data structure, even if you have a small number of cases, have subset your data into fine breakdowns or have variables where 80 percent or more of the responses are in one category.

IBM SPSS Categories[†]

Unleash the full potential of your categorical data through perceptual maps with optimal scaling and dimension reduction techniques. This add-on module provides you with everything you need to analyze and interpret multivariate data and their relationships more completely.

IBM SPSS Forecasting[†]

Improve forecasting with complete time-series analyses, including multiple curve-fitting and smoothing models and methods for estimating autoregressive

functions. Use the Expert Modeler to automatically determine which ARIMA (autoregressive integrated moving average) process or exponential smoothing model best fits your time-series and independent variables, eliminating selection through trial and error.

IBM SPSS Conjoint[†]

Conjoint helps market researchers develop successful products. By performing conjoint analysis, you learn what product attributes are important in the consumer's mind and what the most preferred attribute levels are, and can perform pricing studies and brand equity studies.

IBM SPSS Missing Values[†]

If values are missing from your data, this procedure may find some relationships between the missing values and other variables. In addition, the missing values procedure can estimate what the value would be if data weren't missing.

IBM SPSS Data Preparation[†]

With Data Preparation, you gain several procedures that facilitate the data preparation process. This add-on module enables you to easily identify suspicious and invalid cases, variables and data values; view patterns of missing data; summarize variable distributions to get your data ready for analysis; and more accurately work with algorithms designed for nominal attributes.

IBM SPSS Neural Networks[†]

Use the Neural Networks module to model complex relationships between inputs and outputs or to discover patterns in your data. Choose from algorithms that can be used for classification (categorical outcomes) and prediction (numerical outcomes). The two available algorithms are

Multilayer Perceptron and Radial Basis Function.

IBM SPSS Complex Samples[†]

Incorporate complex sample designs into data analysis for more accurate analysis of complex sample data. Complex Samples, with specialized planning tools and statistics, reduces the risk of reaching incorrect or misleading inferences for stratified, clustered or multistage sampling.

IBM SPSS Amos™[†] (Windows only)

Support your research and theories by extending standard multivariate analysis methods when using this stand-alone software package for structural equation modeling (SEM). Build attitudinal and behavioral models that more realistically reflect complex relationships, because any numeric variable, whether observed or latent, can be used to predict any other numeric variable. New capabilities in this latest release include bootstrapping of user-defined functions of the model parameter for increased model stability.

IBM SPSS Statistics Developer[†]

With Statistics Developer, R algorithms can be easily "wrapped" in Statistics syntax so that they take on the appearance of standard IBM SPSS Statistics procedures, which can easily be invoked through an interface that is indistinguishable from Statistics' built-in dialogs. Non-specialists will be able to access and use the entire array of free statistical functions and procedures available in R. At the same time, those who are dedicated to R and want to use the language to do groundbreaking work will find it easier to do so.

IBM SPSS Text Analytics for Surveys†

Text Analytics for Surveys is a standalone software package that offers a combination of linguistic technologies and manual techniques to categorize responses to open-ended questions. To enhance your quantitative analysis, you can export the results as categories or dichotomies for analysis in Statistics Base, IBM SPSS Data Collection or Excel.

IBM SPSS Data Collection Data Entry† and IBM SPSS Data Collection† products

SPSS, an IBM Company, offers a variety of stand-alone products that help you enter and capture data for survey research. Data Collection Data Entry provides you with options for desktop- or web-based data entry. Data Collection gives you the ability to automatically capture data online, by telephone, through handheld devices or when using paper forms that you scan. All of these products work with IBM SPSS Statistics, enabling you to seamlessly analyze your survey data.

IBM SPSS Visualization Designer†

This product makes it easy to create compelling visualizations that can be saved as templates and reused within IBM SPSS products.

About IBM Business Analytics

IBM Business Analytics software delivers complete, consistent and accurate information that decision-makers trust to improve business performance. A comprehensive portfolio of business intelligence, predictive analytics, financial performance and strategy management, and analytic applications provides clear, immediate and actionable insights into current performance and the ability to predict future outcomes. Combined with rich industry solutions, proven practices and professional services, organizations of every size can drive the highest productivity, confidently automate decisions and deliver better results.

As part of this portfolio, IBM SPSS Predictive Analytics software helps organizations predict future events and proactively act upon that insight to drive better business outcomes. Commercial, government and academic customers worldwide rely on IBM SPSS technology as a competitive advantage in attracting, retaining and growing customers, while reducing fraud and mitigating risk. By incorporating IBM SPSS software into their daily operations, organizations become predictive enterprises – able to direct and automate decisions to meet business goals and achieve measurable competitive advantage.

---

† *IBM® SPSS® Statistics Server, IBM® SPSS® Direct Marketing, IBM® SPSS® Bootstrapping, IBM® SPSS® Statistics Programmability Extension, IBM® SPSS® Regression, IBM® SPSS® Advanced Statistics, IBM® SPSS® Custom Tables, IBM® SPSS® Decision Trees, IBM® SPSS® Exact Tests, IBM® SPSS® Categories, IBM® SPSS® Forecasting, IBM® SPSS® Conjoint, IBM® SPSS® Missing Values, IBM® SPSS® Data Preparation, IBM® SPSS® Neural Networks, IBM® SPSS® Complex Samples, IBM® SPSS® Amos, IBM® SPSS® Statistics Developer, IBM® SPSS® Text Analytics for Surveys, IBM® SPSS® Data Collection Data Entry, IBM® SPSS® Data Collection and IBM® SPSS® Visualization Designer were formerly known as PASW® Statistics Server, PASW® Direct Marketing, PASW® Bootstrapping, PASW® Statistics Programmability Extension, PASW® Regression, PASW® Advanced Statistics, PASW® Custom Tables, PASW® Decision Trees, PASW® Exact Tests, PASW® Categories, PASW® Forecasting, PASW® Conjoint, PASW® Missing Values, PASW® Data Preparation, PASW® Neural Networks, PASW® Complex Samples, Amos™, PASW® Statistics Developer, PASW® Text Analytics for Surveys, PASW® Data Collection Data Entry, PASW® Data Collection and PASW® Viz Designer.*

IBM.

Business Analytics software

IMD14291USEN-02