SPSS Version 29.0

› Detail Module Specification

IBM SPSS Statistics helps data analysts, planners, forecasters, survey researchers, program evaluators and database marketers among others more easily accomplish tasks at virtually every phase of the analytical process. It includes a broad array of integrated SPSS Statistics capabilities and related products for specialized analytical tasks across the enterprise. The software can improve productivity significantly and help achieve superior results for specific projects and business goals.

TABLE OF CONTENTS

 # Independent Product
* Represents newly added features in SPSS Statistics version 29.0
** Represents newly added features in SPSS Statistics version 28.0

1

> IBM SPSS Statistics Base*

Key Features

Descriptive statistics
- Cross-tabulations
- Frequencies, descriptives, explore, descriptive, ratio statistics

Bivariate statistics
- Means, t tests, , correlation (bivariate, partial, distances) and nonparametric tests

Prediction for numerical outcomes and identifying groups
- Factor analysis
- K-means cluster analysis
- Hierarchical cluster analysis
- Two-step cluster analysis
- Discriminant
- Linear regression (Linear, Ordinal, Partial Least square and Curve Estimation)
- Ordinal regression—PLUM
- Multithreaded algorithms: SORT, correlation, partial correlation, linear regression, factor analysis
- Nearest neighbour analysis, which can be used for prediction or for classification
- Nonparametric tests provide multiple comparisons and perform efficiently on large data sets
- Monte Carlo simulation

Data editor
- Easily work with data containing time and dates in IBM SPSS Statistics software
    - Create a time or date variable from a string containing a date variable
    - Create a time or date variable from variables that include individual date units, such as month or year
    - Calculate times and dates
    - Separate a date unit from a time or date variable
- Make sense and keep track of your data files by adding notes to them with the Data File Comments command.
- Easily eliminate duplicate records with Identify Duplicate cases tool.
- Create read-only data sets
- More accurately describe your data using longer variable names (up to 64 bytes)
- Create value labels up to 120 characters
- Clone or duplicate data sets
- Apply an extended Variable Properties command to customize properties for individual users
- Longer text strings (up to 32,000 bytes)
- Define Variables Properties tool
- Right-click on the variable to choose its descriptive statistics
- Copy Data Properties tool
- Data Restructure wizard
- Aggregate data to external or to the active data file Automatically convert string variables to numeric with auto recode
    - Spell-checking of long text strings.
- Date and Time wizard:
    - Easily work with data containing time and dates.
    - Create a time or date variable from a string containing a date variable
    - Create a time or date variable from variables that include individual date units, such as month or year
    - Calculate times and dates
    - Separate a date unit from a time or date variable

- Apply splitters in the data editor for easier viewing of wide or long data files
- Create your own dictionary information for variables by using custom attributes
- Customize the viewing of extremely wide files with variable sets
- Use syntax to change string length and basic data type Set a permanent default working directory

Transformations
- Factor analysis
- More easily find and replace text strings in your data using the find and replace function
- Recode string or numeric value
- Recode values into consecutive integers
- Create conditional transformations using DO IF, ELSE IF, ELSE and END IF statements
- Use programming structures, such as do repeat-end repeat, loop-end loop and vectors
- Compute variables using arithmetic, cross-case, date and time, logical, missing-value, random-number, statistical, or string functions
- Create variables that contain the values of existing variables from preceding or subsequent cases
- Count occurrences of values across variables
- Make transformations permanent or temporary
- Execute transformations immediately, batched or on demand

Reporting
- OLAP
- Case summaries
- Report summaries

Geospatial Analysis
- Temporal casual modeling
- Spatio-temporal predictions

Graphs
- Categorical charts
    - 3-D bar: Simple, cluster and stacked
    - Bar: Simple, cluster, stacked, dropped 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
    - Box plot: Simple and clustered
    - Error bars: Add to bar, line and area charts; confidence level; S.D.; or S.E.
    - Dual-Y axes and overlay subgroups, display spikes to line.
- Bubble Chart
    - The size of each bubble or circle is proportional to its value.
- Violin Plots*
    - Density of each variable can be depicted as well as comparison of distribution between multiple groups can also be done.
- Relationship Maps**
    - The thickness of the connections and the size of the categories indicate the extent of association between the variables.
- Scatterplots
    - Simple, grouped, scatterplot matrix and 3-D
    - Fit lines: Linear, quadratic or cubic regression; Lowess smoother; confidence interval control for total or bivariate statistics
    - Bin points by color or marker size to prevent 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, individual chart or moving range chart
  - Rule-checking performed on primary and secondary charts
  - Automatic flagging of points that violate Shewhart rules, the ability to turn off rules and the ability to suppress charts
- Diagnostic and exploratory charts
  - Case plots and time-series plots
  - Probability plots
  - Autocorrelation and partial autocorrelation function plots
  - Cross-correlation function plots
  - Receiver-operating characteristics
- Multiple use charts
  - 2-D line charts (with two-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 to create mixed charts and more
- Layout options
  - Paneled charts: Create a table of subcharts, one panel per level or condition; multiple row and columns
  - 3-D effects: Rotate, modify depth and display backplanes
- Chart templates
  - PaneledSave selected characteristics of a chart and apply them to others automatically
  - Apply the following attributes at creation or edit time: 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.

Reliability Analysis

- Interrater Agreement: Fleiss' Kappa
  - Assesses the interrater agreement to determine the reliability among the various raters. A higher agreement provides more confidence in the ratings reflecting the true circumstance, generalized the unweighted kappa statistic to measure the agreement among any constant number of raters while assuming:
  - At least two item variables must be specified to run any reliability statistic.
  - At least two ratings variables must be specified.
  - The variables selected as items can also be selected as ratings.
  - There is no connection between raters.
  - The number of raters is a constant.
  - Each subject is rated by the same group containing only a single rater.
  - No weights can be assigned to the various disagreements.
- Hotelling's T-square
  - Produces a multivariate test of the null hypothesis that all items on the scale have the same mean.
- Tukey's test of additivity
  - Produces a test of the assumption that there is no multiplicative interaction among the items.
- Intraclass correlation coefficient
  - Produces measures of consistency or agreement of values within cases.

- Model
  - Select the model for calculating the intraclass correlation coefficient. Available models are Two-Way Mixed, Two-Way Random, and One-Way Random. Select Two-Way Mixed when people effects are random and the item effects are fixed, select Two-Way Random when people effects and the item effects are random, or select One-Way Random when people effects are random.
- Type
  - Select the type of index. Available types are Consistency and Absolute Agreement.
- Confidence interval (%)
  - Specify the level for the confidence interval. The default is 95%.
- Test value
  - Specify the hypothesized value of the coefficient for the hypothesis test. This value is the value to which the observed value is compared. The default value is 0.
- Display agreement on individual categories
  - Specifies whether or not to output the agreement on individual categories. By default, the output suppresses the estimation on any individual categories. When enabled, multiple tables display in the output.
- Ignore string case
  - Controls whether or not the string variables are case sensitive. By default, string rating values are case sensitive.
- String category labels are displayed in uppercase
  - Controls whether the category labels in the output tables are displayed in uppercase or lowercase. The setting is enabled by default, which displays the string category labels in uppercase.
- Asymptotic significance level (%)
  - Specifies the significance level for the asymptotic confidence intervals. 95is the default setting.
- Missing
  - Exclude both user-missing and system missing values
  - Controls the exclusion of user-missing and system-missing values. By default, user-missing and system-missing values are excluded.
  - User-missing values are treated as valid
  - When enabled, treats user-missing and system-missing values as valid data. The setting is disabled by default.

ROC Analysis

  - Assesses the accuracy of model predictions by plotting sensitivity versus (1-specificity) of a classification test (as the threshold varies over an entire range of diagnostic test results).
  - ROC Analysis supports the inference regarding a single AUC, precision-recall (PR) curves, and provides options for comparing two ROC curves that are generated from either independent groups or paired subjects.

3

Linear OLS Alternatives*

Three new procedures are included under the Linear OLS Alternatives Option within regression, namely:

- Ridge Regression*
  - Ridge regression uses the L2 regularization technique i.e. it uses the penalty based on the sum of the squared coefficient values. Ridge regression extension provides the users the option to generate trace plots along with the option to enter the set of alpha values as well as a specified value of alpha.
- Lasso Regression*
  - Lasso regression uses the L1 regularization technique i.e. it uses the penalty based on the absolute sum of the coefficient values. The Lasso option extension provides the user to draw the trace plot and also allows the user to enter the set of alpha values as well as a specified value of alpha.
- Elastic Net Regression*
  - Elastic Net regression uses the weighted combination of L1 and L2 regularization technique i.e. the penalty introduced in Elastic Net takes into consideration the penalties used in both Ridge and Lasso regression.

Meta-analysis**

- Meta-analysis is the scientific process to synthesize and combine different results from different studies. Helps in understanding the magnitude of the effect size and the causes of potential variation. Provides more robust point estimates with higher power as compared to the individual studies. Following are the options under Meta Analysis:
  - Meta-analysis of continuous outcomes: Raw data, pre-calculated effect size data
  - Meta-analysis of binary outcomes: Raw data, pre-calculated effect size data
  - Meta-regression

Power Analysis**

- Help in determining the optimum sample size for a study or project. This feature includes the graphical capabilities for expressing power as a function of sample size and effect size. It includes the 3-dimensional plots also. Following are the Power analysis procedures:
  - One sample t-test
  - Paired sample t-test
  - Independent sample t-test
  - One-Way ANOVA.
  - One sample binomial test
  - Related-samples binomial test
  - Independent samples binomial test.
  - Pearson's product-moment correlation test
  - Spearman's rank-order correlation test
  - Partial Pearson correlation test
- All the procedures under power analysis now provide the option to enter a series of power values rather than a single one. These can take the form of separate (point) values, a series of values incremented by a specific amount, or a combination of both.

Weighted Kappa procedure

- Helps in understanding the closeness of agreement between raters. Following are the weighted versions of Cohen's weighted kappa that are being introduced:
  - Linear weighting
  - Quadratic weighting
  - Confidence intervals for all kappa coefficients

Effect Sizes

- This enhancement to the t-test and one-way procedures add effect size estimates and confidence intervals for the options One-sample t-tests, Independent-sample t-tests, Paired samples t-tests, One-Way Analysis of Variance (ANOVA), Custom Contrasts in One-Way ANOVA.

Ratio Statistics**

- Introduction of the coefficient of price related bias or PRB. PRB comes with an associated confidence interval. This index (PRB) tells us whether the assessment ratios are systematically higher or lower for high-priced properties.

⟩ IBM SPSS
Advanced Statistics*

Key Features

In SPSS Statistics, you get to analyze data with some new features:

GENLINMIXED now provides:
- Random effects solution results (EBLUPs)
- Continuous time spatial covariance structures
- GLM/UNIANOVA now provides:
- Profile plots with error bars, bar and line charts, and an option to include grand mean option to force charts to include 0 on the y-axis
- New tests for heteroskedasticity, including White's test
- Robust standard errors
- Modified versions of Levene's test

Bayesian Statistics
SPSS Statistics supports for Bayesian Statistics, which includes both new syntax and GUI elements that are as easy to run as traditional p-value statistics. Some of the features available within Bayesian Statistics include:
- One Sample and Pair Sample T-tests
- Binomial Proportion tests
- Poisson Distribution Analysis
- Independent Samples T-tests
- Pairwise Correlation
- Simple and Multiple Linear Regression
- Analysis of variance (ANOVA)

GLMM
Extends the linear model so that the target is linearly related to the factors and covariates through a specified link function, the target can have a non-normal distribution, and the observations can be correlated. GLMM covers a wide variety of models, from simple linear regression to complex multilevel models for nonnormal longitudinal data.
- Specify the subject structure for repeated measurements and how the errors of the repeated measurements are correlated
- Choose among the eight covariance types
- Specify the target, optional offset and optional analysis (regression) weight
- Choose among the following probability distributions: binomial, gamma, inverse Gaussian, multinomial, negative binomial, normal, Poisson
- Choose among the following link functions: identity, complementary log-log, log-link, log complement, logit, negative log-log, power, probit.

GENLIN and GEE
GENLIN and GEE procedures provide a unifying framework for a wide variety of model types. Together, they enable you to predict more types of outcomes, including:
- Ordinal outcomes such as customer satisfaction
- Outcomes that are a combination of discrete and continuous outcomes, such as claim amount, with a Tweedie distribution

- Provide a common framework for the following outcomes: continuous outcomes, count data, event or trial data, claim data, ordinal outcomes, combination of discrete and continuous outcomes, and correlated responses within subjects

MIXED
Expands the general linear model used in the GLM procedure so that data can exhibit correlation and no constant variability
- Fit the following types of models:
  - Fit Fixed effects ANOVA model, randomized complete blocks design, split-plot design, purely random effects model, random coefficient model, multilevel analysis, unconditional linear growth model, linear growth model with person-level covariate, repeated measures analysis, and repeated measures analysis with time-dependent covariate
- Use one of six covariance structures offered
- Select from 11 nonspatial covariance types

GLM
Describe the relationship between a dependent variable and a set of independent variables
- Select univariate and multivariate lack-of-fit tests
- Regression model
- Fixed effect ANOVA, ANCOVA, MANOVA and MANCOVA
- Random or mixed ANOVA and ANCOVA
- Repeated measures: Univariate or multivariate
- Doubly multivariate design

VARCOMP
Variance component estimation (VARCOMP)
- Estimation methods: ANOVA MINQUE, maximum likelihood and restricted maximum likelihood
- Type I and Type III sums of squares for the ANOVA method
- Choices of zero-weight or uniform-weight methods
- Choices of ML and REML calculation methods
- Save variance components estimates and covariance matrices

LOGLINEAR and HILOGLINEAR*
- Life Tables
- Kaplan-Meier Survival Analysis
- Cox regression
- Parametric Accelerated Failure Time (AFT) Models*

5

Bayesian Statistics
- One-way Repeated Measures ANOVA
- One Sample Binomial enhancements.
- One Sample Poisson enhancements

MATRIX-END MATRIX

The MATRIX and END MATRIX commands enclose statements that are executed by the matrix processor. Using matrix programs, you can write your own statistical routines in the compact language of matrix algebra. Matrix programs can include mathematical calculations, control structures, display of results, and reading and writing matrices as character files or data files.

GENLINMIXED

GENLINMIXED is available in the Advanced Models option.
The GENLINMIXED procedure fits generalized linear mixed models.
The KRONECKER_MEASURES keyword should be only when COVARIANCE_TYPE is defined as UN_AR1, UN_CS, or UN_UN.
When both KRONECKER_MEASURES and REPEATED_MEASURES keywords are in effect, they may or may not have common fields, but their values cannot be exactly the same (even when the values are not in the same order).

MIXED

MIXED is available in SPSS® Statistics Standard Edition or the Advanced Statistics Option.
The MIXED procedure fits a variety of mixed linear models. The mixed linear model expands the general linear model used in the GLM procedure in that the data are permitted to exhibit correlation and non-constant variability.
† covstruct can take the following values: AD1, AR1, ARH1, ARMA11, CS, CSH, CSR, DIAG, FA1,FAH1, HF, ID, TP, TPH, UN, UNR, VC.
For the REPEATED subcommand, covstruct can take the following additional values: SP_POWER, SP_EXPONENTIAL, SP_GAUSSIAN, SP_LINEAR, SP_LINEARLOG, SP_SPHERICAL, UN_AR1, UN_CS, UN_UN.
The KRONECKER keyword should be used only when COVTYPE is defined as UN_AR1, UN_CS, or UN_UN.

When both KRONECKER and REPEATED are in effect, they may or may not have common fields, but their values cannot be exactly the same (even when the values are not in the same order).

⟩ IBM SPSS
Bootstrapping

Key Features

IBM SPSS Bootstrapping software provides the ability to bootstrap a number of analytical procedures found throughout the IBM SPSS Statistics product family, including:

Descriptive procedures

IBM SPSS Statistics Base:
- Descriptives
- Frequencies
- Examine
- Means
- Crosstabs
- t tests
- Correlations or nonparametric correlations
- Partial correlations

Modeling procedures
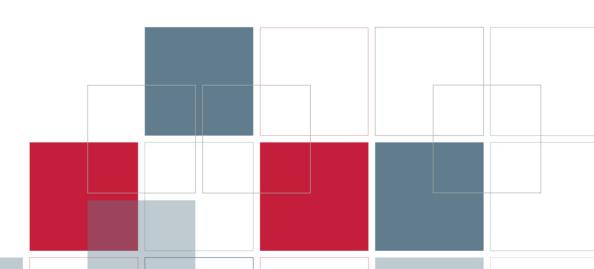
IBM SPSS Statistics Base:
- One-way
- UNIANOVA
- PLUM
- Discriminant

IBM SPSS Advanced Statistics:

- GLMM
- GENLIN
- Linear mixed models
- Cox regression

IBM SPSS Regression:

- Regression
- Nominal regression
- Logistic regression

⟩ IBM SPSS
Direct Marketing

Key Features

Analytical capabilities

- Compute RFM scores from a data set in which each row contains the aggregated data for one customer or the data for one transaction
- Accept recency data in the form of a transaction date or the time interval since the transaction
- Generate scores indicating which contacts are most likely to respond to similar campaigns
- Create training and testing groups for diagnostic purposes
- Recode the response field to represent positive or negative responses
- Cluster analysis works with continuous and categorical fieldsContact profiling works with nominal, ordinal, string or numeric fields
- Profile customers or contacts

Data output

- Illustrate in an average monetary value chart how recency, frequency and spending are related in the sample
- Display counts and percentages of positive and negative responses for each group and identify which are significant

- Examine descriptions of each profile group, response rates and cumulative response rates using the contact profiles feature
- Export all analytics to Excel
- Leverage improved, descriptive text to help explain—in everyday language—the output that results from executed procedures

User options

- Export Write the computed scores (and ID variables, where necessary) to the active data set, a new data set or a file
- Append the RFM results directly to your data or a new data file to quickly identify and build lists of high-value customers
- Create descriptive profiles that indicate who responded to which previous campaigns

Zip code response

- Export Create a new data set that contains response rates by postal code
- Choose general response rates based on N characters, three digits, five digits or the complete value of a postal code.

⟩ IBM SPSS
Conjoint

Key Features

- Orthoplan: Generates orthogonal main effects fractional factorial designs.
- Specify the desired number of cards for the plan
- Generate holdout cards to test the fitted conjoint model
- Orthoplan can mix the training and holdout cards or can stack the holdout cards after the training cards
- Plan cards: A utility procedure used to produce printed cards for a conjoint experiment; the printed cards are used as stimuli to be sorted, ranked or rated by the subjects
- Specify the variables to be used as factors and the order in which their labels are to appear in the output.
- Choose a format
  – Listing-file format
  – Card format
- Conjoint: Performs an ordinary least squares analysis of preference or rating data.
- Work with the plan file generated by plan cards or a plan file input

- Work with individual level rank or rating data Provide individual level and aggregate results
- Treat the factors in a number of ways; conjoint indicates reversals
- Experimental cards have one of three scenarios: Training, holdout and simulation
- Three conjoint simulation methods: Max utility; Bradley- Terry-Luce (BTL); and logit
- Print results
  – Attribute importance
  – Utility (part worth) and standard error
  – Graphical indication of most-preferred to least-preferred levels of each attribute
  – Counts of reversals and reversal summary Pearson R for training and holdout data
  – Kendall's tau for training, holdout data simulation results and simulation summary.

IBM SPSS
Exact Tests

Key Features

Tests and statistics

Pearson Chi-square test, Likelihood ratio and Fisher's Exact
- Exact one-tailed, two-tailed p values for 2x2 table
- Exact two-tailed p values for general RxC table
- Monte Carlo (MC) two-tailed p values and general RxC table

Linear-by-linear association
- Exact one-tailed and two-tailed p values and exact point probability.
- MC one-tailed and two-tailed p values and Cls

Contingency coefficient, Phi, Cramer's V, Goodman and Kruskal Tau, uncertainty coefficient—symmetric or asym¬metric, Kappa, Gamma, Kendall's Tau-b and Tau-c, Somers' D—symmetric and asymmetric, Pearson's R and Spearman correlation
- Exact two-tailed p values
- MC two-tailed p values and Cls

McNemar
- Exact one-tailed and two-tailed p values and point probability

Sign and Wilcoxon signed-rank
- Exact one-tailed and two-tailed p values and point probability

- MC one-tailed and two-tailed p values and Cls

Marginal homogeneity
- Asymptotic, exact, MC one-tailed and two-tailed p values and point probability

Two-sample Kolmogorov-Smirnov
- Exact two-tailed p values and point probability
- MC two-tailed p values and Cls

Mann-Whitney U or Wilcoxon Rank-sum W
- Exact two-tailed p values, exact one-tailed p values and point probability
- MC one-tailed and two-tailed p values and CIs

Wald-Wolfowitz Runs
- Exact one-tailed p value and point probability
- MC one-tailed p value and Cls

Jonckheere-Terpstra
- Asymptotic, exact, MC one-tailed and two-tailed p values and point probability

IBM SPSS
Categories

8

Key Features

PREFSCAL

- Multidimensional unfolding analysis
- Read one or more rectangular matrices of proximities
- Read weights, initial configurations and fixed coordinates

PROXSCAL (Multidimensional Scaling)
- Statistics: Iteration history, stress measures, stress decomposition, coordinates of the common space and object distances within the final configuration
- Plots: Stress plots, common space scatterplots, individual space weight scatterplots and individual space scatterplots

CATPCA (Categorical Principal Component analysis)
- Statistics: Frequencies, missing values, optimal scaling level, mode, variance accounted for by centroid coordinates, vector coordinates, total per variable and per dimension, component loadings for vector quantified variables, category quantifications and coordinates, and iteration history
- Plots: Joint category plots, transformation plots, residual plots, projected centroid plots, object plots, biplots, triplots and component loadings plots

CORRESPONDENCE (Correspondence analysis)
- Statistics: Correspondence measures; row and column profiles; singular values; row and column scores; and inertia, mass
- Plots: Transformation plots, row point plots, column point plots and biplots

CATREG (Categorical regression)
- Statistics: Frequencies, regression coefficients, ANOVA table, iteration history and category quantifications
- Plots: Correlations between untransformed formed predictors, correlations between transformed predictors, residual plots and transformation plots

- Three regularization methods: Ridge regression, the Lasso and the Elastic Net

Improve prediction accuracy by stabilizing the parameter estimates
- Analyse high-volume data (more variables than objects)
- Obtain automatic variable selection from the predictor set
- Write regularized models and coefficients to a new data set for later use
- Two model selection and predictive accuracy assessment methods: the .632 bootstrap and cross-validation (CV) Find the model that is optimal for prediction with the .632 or .632+ bootstrap and CV options
- Obtain nonparametric estimates of the standard errors of the coefficients with the bootstrap
- Systematic multiple starts

MULTIPLE CORRESPONDENCE ANALYSIS

Statistics
- Model summary
- Distory statistics
- Descriptive statistics
- Discrimination measures
- Category quantifications
- Inertia of the categories
- Contribution of the categories to the inertia of the dimensions and contribution of the dimensions to the inertia of the categories; and iteration history

Plot
- Object points, category points (centroid coordinates), discrimination measures and transformation (optimal category quantifications against category indicators)

OVERALS
- Statistics: Frequencies, centroids, iteration history, object scores, category quantifications, weights, component loadings, and single and multiple fit

⟩ IBM SPSS
Neural Networks

Key Features

MLP

Fits an MLP neural network, which uses a feed-forward architecture
- Can have multiple hidden layers
- One or more dependent variables may be specified— scale, categorical or a combination of these
- EXCEPT subcommand excludes selected variables
- RESCALE subcommand rescales covariates or scale-dependent variables
- PARTITION subcommand specifies the method of partitioning the active data set into training, testing and holdout samples
- ARCHITECTURE subcommand specifies the network architecture
- The number of hidden layers
- The activation function to use for all units in the hidden layers
- The activation function to use for all units in the output layer
- CRITERIA subcommand is used to specify computational resources
- STOPPINGRULES subcommand specifies the rules that determine when to stop training the network
- MISSING subcommand controls whether user-missing values for categorical variables are treated as valid values

- PRINT subcommand indicates the tabular output to display and can be used to request a sensitivity analysis
- PLOT subcommand indicates the chart output to display
- SAVE subcommand writes temporary variables to the active data set
- OUTFILE subcommand saves XML format files containing the synaptic weights

RBF

- Fits an RBF network, which uses a feed-forward, supervised architecture
- Has only one hidden layer
- Trains the network in two stages and is faster than an MLP network.
- Subcommands listed for the MLP procedure (above) perform similar functions for the RBF procedure, except that:
  - When using the ARCHITECTURE subcommand, users can specify the Gaussian RBF used in the hidden layer: either normalized RBF or ordinary RBF.
  - When using the CRITERIA subcommand, users can specify the computation settings for the RBF procedures, specifying the hidden-unit overlapping factor that controls how much overlap occurs among the hidden units.

⟩ IBM SPSS
Custom Tables

Key features

Graphical user interface
- Simple drag-and-drop table builder interface allows you to preview tables as you select variables and options
- Single, unified table builder instead of multiple menu choices and dialog boxes for different table types

Control contents
- Create tables with up to three display dimensions: Rows (stub), columns (banner) and layers
- Nest variables to any level in all dimensions
- Cross-tabulate multiple independent variables in the same table
- Display frequencies for multiple variables side by side with tables of frequencies
- Display all categories when multiple variables are included in a table even if a variable has a category without responses
- Display multiple statistics in rows, columns or layers
- Place totals in any row, column or layer
- Create subtotals for subsets of categories of a categorical variable
- Perform calculations (for example, sum, difference, percentage difference) on output categories, and display the results in new fields created directly in custom tables output
  - No limit on the number of calculated fields
- Show significance test results directly in custom tables output instead of in a separate table
  - No need to combine findings in a Word document
  - Complies with American Psychological Association guidlines
- Control category display order and the ability to selectively show or hide categories.

Statistics
- Select from more than 160 summary statistics
- Calculate statistics for each cell, subgroup or table
- Calculate percentages at any or all levels for nested variables
- Calculate counts and percentages for multiple-response variables based on the number of responses or the number of cases

- Select percentage bases for missing values to include or exclude missing responses

Formatting controls
- Sort categories by any summary statistic in the table
- Hide the categories that make up subtotals—remove a category from the table without removing it from the subtotal calculation
- Directly edit any table element, including formatting and labels
- Sort tables by cell contents in ascending or descending order
- Automatically display labels instead of coded values
- Specify minimum and maximum width of table columns (overrides TableLooks)
- Show a name, label or both for each table variable
- Display missing data as blank, zero, "." or any other user-defined term
- Add titles and captions
- Output as IBM SPSS Statistics pivot tables
- Specify corner labels
- Customize labels for statistics
- Display the entire label for variables, values and statistics
- Choose from a variety of numerical formats
- Apply preformatted TableLooks
- Define the set of variables that are related to multiple response data and save it with your data definition for sub-sequent analysis
- Use both long-string and short-string elementary variables
- Define an unlimited number of sets or variables that can exist in a set
- Tests of significance:
  - Chi-square
  - Column means
  - Column proportions
- Exclude categories from significance tests
- Significance tests for multiple response variables
- Syntax and printing formats
- Simpler, easy-to-understand syntax
- Syntax converter (for upgrade users)
- Specify page layout: Top, bottom, left and right margins and page length
- Use the global break command to produce a table for each value of a variable when the variable is used in a series of tables

IBM SPSS
Data Preparation

Key features

Automated data preparation

- Recommends steps to speed up model building and improve predictive power
- Determine objective, prepare dates and times for modeling, exclude low-quality input fields, prepare fields to improve data quality, rescale fields, continuous input and target fields, transform fields, perform feature selection and construction, name fields and apply transformations to data

Validate data

Validate data in the working data file

- Basic checks
  - Maximum percent of missing values, single category cases and cases with a count of one
  - Minimum coefficient of variation
  - Minimum standard deviation
  - Flag incomplete IDs, duplicate IDs and empty cases
- Standard rules: Describe the data, view single variable rules and apply them to analysis variables
  - Description of data
    - Distribution: shows a thumbnail-size bar chart for categorical variables or histogram for scale variables.
    - Minimum and maximum data values shown
- Single-variable rules:
  - Apply rules to identify missing or invalid values
  - User-defined
- Custom rules: Define cross-variable rule expressions in which respondents' answers violate logic
- Output: Reports for invalid data
  - Casewise report, specify by case
    - Specify the minimum number of violations needed for a case
    - Specify the maximum number of cases in the report
  - Standard validation rules reports
    - Summarize violations by analysis variable and rule
    - Display descriptive statistics
- Save: Save variables that record rule violations and use them to clean data and filter out bad cases
  - Summary variables:
    - Empty case indicator
    - Duplicate ID indicator
    - Incomplete ID indicator
    - Validation rule violation
- Indicator variables that record all validation rule violations

Identify unusual cases

Anomaly detection searches for unusual cases based on deviations from peer group and reasons for deviations

- VARIABLES subcommand: Specify categorical, continuous and ID variables and list variables that are excluded from the analysis
- HANDLEMISSING subcommand: Specify the methods of handling missing values in this procedure
- The CRITERIA subcommand specifies the following settings:
  - Number of peer groups
  - Adjustment weight on the measurement level
  - Number of reasons in the anomaly list
  - Percentage and number of cases considered as anomalies and included in the anomaly list
  - Cut point of the anomaly index to determine whether a case is considered as an anomaly
- Save additional variables to the working data file including:
  - Anomaly index
  - Peer group ID, size and size in percentage
  - Variable, variable impact measure, variable value and norm value associated with a reason
- OUTFILE subcommand: Write a model to a file name as XML
- PRINT subcommand prints:
  - Case-processing summary
  - Anomaly index list, anomaly peer ID list and anomaly reason list
  - The Continuous Variable Norms table for continuous variable and the Categorical Variable Norms table for categorical variable
  - Anomaly index summary
  - Reason Summary table

Optimal binning

Preprocess data with optimal binning. Categorizes one or more continuous variables by distributing the values of each into bins

- Select from the following methods:
  - Unsupervised binning via the equal frequency algorithm: It uses the equal frequency algorithm to discretize the binning input variables. Guide variable not required.
  - Supervised binning via the MDLP (Minimum Description Length Principle) algorithm: Discretizes binning input variables using the MDLP algorithm without any preprocessing. Ideal for small data sets. Guide variable required.
  - Hybrid MDLP binning: Involves preprocessing via the equal frequency algorithm followed by the MDLP algorithm. Ideal for large data sets. Guide variable required.
- Specify the following criteria:
  - How to define the minimum and maximum cut point for each binning input variable and the lower limit of an interval
  - Whether to force-merge sparsely populated bins
  - Whether missing values uses listwise or pairwise deletion
- Save new variables with binned values and syntax to an IBM SPSS Statistics syntax file.
- PRINT subcommand prints:
  - The binning input variables' cut point sets
  - Descriptive information for all binning input variables
  - Model entropy for binned variables.

⟩ IBM SPSS Decision Trees

Key features

Create tree-based classification models for:
- Segmentation
- Stratification
- Prediction
- Data reduction and variable screening
- Interaction identification
- Category merging and discretizing continuous variables
- Classify cases into groups or predict values of a dependent (target) variable based on values of independent (predictor) variables
- Validation tools for exploratory and confirmatory classification analysis
- View nodes using one of several methods: Show bar charts of your target variables, tables or both in each node
- Collapse and expand branches without deleting the model
- Generate syntax automatically from the user interface (UI)
- Rerun tree building using syntax in production mode
- Score data based on results, or use results in further analysis using other IBM SPSS Statistics procedures

Algorithms
- Four powerful tree-modeling algorithms: CHAID by Kass (1980)
- Exhaustive CHAID by Biggs, de Ville and Suen (1991)
- Classification and regression trees C&RT by Breiman, Friedman, Olshen and Stone (1984)
- QUEST by Loh and Shih (1997)

Evaluation
- Evaluation graphs enable visual representation of gains summary tables
- Misclassification functionality
- Gains chart: Identify segments by highest (and lowest) contribution

Deployment
- Export output objects to any of IBM SPSS Statistics software's available output formats
- Generate rules that define selected segments in SQL to score databases or IBM SPSS Statistics syntax to score IBM SPSS Statistics files
- Export XML models to score new cases

⟩ IBM SPSS Regression

Key Features

Multinomial logistic regression (MLR)
- Control the values of the algorithm-tuning parameters
- Include interaction terms
- Customize hypotheses by directly specifying null hypotheses as linear combinations of parameters
- Specify a dispersion scaling value
- Build equations with or without a constant
- Use a confidence interval for odds ratios
- Save the following statistics: predicted probability, predicted response category, probability of the predicted response category and probability of the actual response category
- Find the best predictor from dozens of possible predictors using stepwise functionality
- Use Score and Wald methods to quickly reach results with a large number of predictors
- Assess model fit
- Diagnostics for the classification table

Binary logistic regression (BLR)
- Forward and backward stepwise and forced entry modeling
- Transform categorical variables by using deviation contrasts, simple comparison, difference (reverse Helmert) contrasts, Helmert contrasts, polynomial contrasts, comparison of adjacent categories, user-defined contrasts or indicator variables
- Criteria for model building: probability of score statistic for entry, probability of Wald or likelihood ratio statistic for removal
- Save the following statistics: Predicted probability and group, residuals, deviance values, logit, studentized and standardized residuals, leverage value, analog of Cook's influence statistic, and difference in Beta
- Export the model using XML

Constrained nonlinear regression (CNLR)
- Save predicted values, residuals and derivatives
- Choose numerical or user-specified derivatives

Nonlinear regression (NLR)
- Specify loss function options
- Use bootstrap estimates of standard errors

Weighted least Square (WLS)
- Calculate weights based on source variable and Delta values or apply from an existing series
- Output for calculated weights: Log-likelihood functions for each value of Delta; R, R2, adjusted R2, standard errors, analysis of variance and t tests of individual coefficient for Delta value with maximized log-likelihood function
- Display output in pivot tables

Two-stage least squares (2SLS)
Structural equations and instrumental variables
Control for correlations between predictor variables and error terms
Display output in pivot tables

Probit analysis
- Transform predictors: Base 10, natural or user-specified base
- Natural response rate estimates
- Algorithm control parameters: Convergence, iteration limit and heterogeneity criterion probability
- Statistics: Frequencies, fiducial confidence intervals, relative median potency, test of parallelism, plots of observed probits or logits
- Display output in pivot tables

Quantile Regression
- Estimation Method:
  – Simplex algorithm
  – Automatic
  – Fisch Newton interion point non Linear optimization
- Post Estimation(custom select cases are IID)
  – Bandwidth Type
  – Bofinger
  – Hall Sheathe

11

IBM SPSS Forecasting

Key Features

TSMODEL

Model a set of time-series variables by using the expert modeler or specifying the structure of an ARIMA or exponential smoothing model

- Allow expert modeler to select ideal predictor variables and models
- Limit search space to only ARIMA or only exponential smoothing models
- Treat independent variables as events
- Specify custom ARIMA models
- Produces maximum likelihood estimates for seasonal and nonseasonal univariate models
- General or constrained models specified by autoregressive or moving average order, order of differencing, seasonal autoregressive or moving average order, and seasonal differencing
- Two dependent variable transformations: Square root and natural log
- Automatically detect or specify outliers: Additive, level shift, innovational, transient, seasonal additive, local trend and additive patch
- Specify seasonal and nonseasonal numerator, denominator and difference transfer function orders and transformations for each independent variable
- Specify custom exponential smoothing models
- Four nonseasonal model types: Simple, Holt's linear trend, Brown's linear trend and damped trend
- Three seasonal model types: Simple seasonal, Winters' additive and Winters' multiplicative
- Two dependent variable transformations: Square root and natural log
- Display forecasts, fit measures, Ljung-Box statistic, parameter estimates and outliers by model
- Generate tables and plots to compare statistics across models
- Eight goodness-of-fit measures available: stationary R^2, R^2, root mean square error, mean absolute percentage error, mean absolute error, maximum absolute percentage error, maximum absolute error and normalized Bayesian Information Criterion (BIC)
- Tables and plots of residual autocorrelation function (ACF) and partial autocorrelation function (PACF)
- Plot observed values, forecasts, fit value, confidence intervals for forecasts and confidence intervals for fit values for each series

- Filter output to fixed number or percentage of best-fitting or worst-fitting models
- Save predicted values, lower confidence limits, upper confidence limits and noise residuals for each series back to the data set
- Specify forecast period, treatment of user-missing values and confidence intervals
- Export models to an XML file for later use by TSAPPLY

TSAPPLY

- Apply saved models to new or updated data
- Simultaneously apply models from multiple XML files created with TS-MODEL
- Reestimate model parameters and goodness-of-fit measures from the data or load from the saved model file
- Selectively choose saved models to apply
- Override the periodicity (seasonality) of the active data set
- Same output, fit measure, statistics and options as TSMODEL
- Export reestimated models to an XML file.

SEASON

Estimates multiplicative or additive seasonal factors for peri¬odic time series
- Multiplicative or additive model
- Moving averages, ratios, seasonal and seasonal adjustment factors, seasonally adjusted series, smoothed trend-cycle components, and irregular components

SPECTRA

Decomposes a time series into its harmonic components, a set of regular periodic functions at different wavelengths or periods
- Produces or plots univariate or bivariate periodogram and spectral density estimate
- Bivariate spectral analysis
- Smooth periodogram values with weighted moving averages
- Spectral data windows available for smoothing: Tukey- Hamming, Tukey, Parzen, Bartlett, equal weight, no smoothing and user-specified weights
- High-resolution charts available: Periodogram, spectral and cospectral density estimate, squared coherency, quadrature spectrum estimate, phase spectrum, cross-amplitude and gain.

IBM SPSS Complex Samples

Key Features

Complex Samples Plan (CSPLAN)
- Provides a common place to specify the sampling frame to create a complex sample design or analysis design used by procedures in IBM SPSS Complex Samples software.

Complex Samples Selection (CSSELECT)
- Selects complex, probability-based samples from a population. It chooses units according to a sample design created through the CSPLAN procedure.

Complex Samples Descriptives (CSDESCRIPTIVES)
- Estimates means, sums and ratios and computes their standard errors, design effects, confidence intervals and hypothesis tests.

Complex Samples Tabulate (CSTABULATE)
- Displays one-way frequency tables or two-way cross-tabulations and associated standard errors, design effects, confidence intervals and hypothesis tests.

Complex Samples General Linear Model (CSGLM)
- Enables you to build linear regression, analysis of variance (ANOVA) and analysis of covariance (ANCOVA) models.

Complex Samples Ordinal (CSORDINAL)
- Performs regression analysis on a binary or ordinal polytomous dependent variable using the selected cumulative link function.

Complex Samples Logistic Regression (CSLOGISTIC)
- Performs binary logistic regression analysis as well as multinomial logistic regression analysis.

Complex Samples Cox Regression (CSCOXREG)
- Applies Cox proportional hazards regression to analysis of survival times-that is, the length of time before the occurrence of an event.

⟩ IBM SPSS
Missing Values

Key Features

Analyze patterns

Display missing data and extreme cases for all cases and all variables using the data patterns table
- Display system-missing and three types of user-defined missing values
- Sort in ascending or descending order
- Display actual values for specified variables

Display patterns of missing values for all cases that have at least one missing value using the missing patterns table
- Group similar missing value patterns together
- Sort by missing patterns and variables
- Display actual values for specified variables

Determine differences between missing and nonmissing groups for a related variable with the separate variance t test table
- t test, degrees of freedom, mean, p value and count

Show differences between present and missing data for categorical variables using the distribution of categorical variables table
- Produce cross-tabs showing product and missing data for each category of one variable by the other variables

Assess how much missing data for one variable relates to the missing data of another variable using the percent mismatch of patterns table
- Sort matrices by missing value patterns or variables

Identify all unique patterns with the tabulated patterns table, which summarizes each missing data pattern and displays the count for each pattern plus means and frequencies for each variable
- Display count and averages for each missing value pattern using the summary of missing value patterns table

Multiple imputation

- Specify which variables to impute and specify constraints on the imputed values, such as minimum and maximum values. You can also specify which variables are used as predictors when imputing missing values of other variables.
- Impute values for categorical and continuous variables.
- Logistic regression is used for categorical variables and linear regression for continuous variables.
- Predictive mean matching is an option for continuous outcomes; this helps ensure that the imputed values are reasonable (within the range of the original data).
- Missing data pattern detection helps determine which imputation method to use.

Three imputation methods are offered:

- Monotone: An efficient method for data that has a monotone pattern of missingness
- Fully conditional specification (FCS): An iterative MCMC method that is appropriate when the data has an arbitrary (monotone or no monotone) missing pattern

- Automatic: Scans the data to determine the best imputation method (monotone or FCS)

Specify
- The number of imputations
- The range of imputed values
- Whether interaction effects are used when imputing
- Optionally, turn off imputation for variables that have a higher percentage of missing values
- Tolerance levels, to check for singularity
- You can also specify a variable containing analysis (regression) weights. The procedure incorporates analysis weights in regression and classification models used to impute missing values. Analysis weights are also used in summaries of imputed values (for example, mean, standard deviation and standard error).
- Display an overall summary of missingness in your data as well as an imputation summary and the imputation model for each variable whose values are imputed. You can obtain analysis of missing values by variable as well as tabulated patterns of missing values. Optionally, you can obtain descriptive statistics for imputed values.
- Graphically summarize missingness for cases, variables and individual data (cell) values.
- Request an IBM SPSS Statistics data file containing imputed values or an FCS iteration history.
- Multiple imputation data sets can be analyzed using supported analysis procedures to obtain final (combined) parameter estimates that take into account the inherent uncertainty in the various sets of imputed values.

Analysis
- Supported analysis procedures for multiple imputation (Note: You must have purchased the proper module in which the procedure is located.)
- Descriptive procedures: Frequencies, descriptives, cross-tabs, correlations, nonparametric correlation, partial correlation
- Comparison of means: Means, t test, nonparametric tests, one-way ANOVA, univariate ANOVA
- Models: General linear models, generalized linear models, linear regression, multinomial logistic regression, binary logistic regression, discriminant analysis, ordinal regression, linear mixed models
- Survival analysis techniques: Cox regression

Pooling
- Pooling of output: Output is pooled using one of two levels of pooling, which produces pooled parameters
- Pooling diagnostics Relative increase in variance: Measure of relative variability in parameter estimate across imputations
- Fraction of missing information: Relative increase in variance scaled as a proportion; a measure of uncertainty as a result of nonresponse
- Relative efficiency: Efficiency of estimate for M imputations relative to that for an infinite number of imputations
- Obtain model predictive model markup language (PMML) for pooled parameter estimates: Linear regression, generalized linear models, multinomial logistic regression, binary logistic regression, discriminant analysis, Cox regression.

13

> IBM SPSS Amos[#]
SPSS Product for Structural Equation Modeling (SEM)

Key Features
- Automatically displayed output on the path diagram.
- The ability to specify a model without drawing a path diagram via syntax.
- With new ground-breaking features in SPSS Amos, you are now able to: Perform D-separation analyses.
- Read R data files (*.rds and *.RData files) even if R is not installed.**
- Execute Amos from within an R program. (This is similar to the ability to execute Amos from the IBM SPSS Statistics main menu).
- Click File > File Explorer to show the current path diagram (*.amw) file in Windows File Explorer. If you haven't saved your path diagram yet, File Explorer shows the default folder for saving path diagrams.
- Get your own copy of the example files within the installation directory.
- Locate important Amos folders through environment variables.
- View list of variables displayed by clicking View > Variables in Dataset, which shows additional information about each variable and allows sorting by variable name and variable label.
- View SPSS Statistics data files (.sav files) in the SPSS Amos View Data window even if SPSS Statistics is not installed.

Modeling capabilities
- Create structural equation models with observed and latent variables
- Specify each individual candidate model as a set of equality constraints on model parameters
- Analyze data from several populations at once
- Save time by combining factor and regression models into a single model and then fit them simultaneously

Bayesian estimation
- Fit models with ordered-categorical and censored data
- MCMC simulation

Computationally intensive modeling
- MCMC simulation
- Evaluate parameter estimates with normal or non-normal data using powerful bootstrapping.

Data imputation
- Impute numerical values for ordered-categorical and censored data
- Impute missing values and latent variable scores
- Choose from three different methods: Regression, stochastic regression and Bayesian

Analytical capabilities and statistical functions
- Determine probable values for missing or partially missing data values in a latent variable model
- Use full information maximum likelihood estimation in missing data situations for more efficient and less biased estimates
- Use a variety of estimation methods, including maximum likelihood, unweighted least squares, generalized least squares, Browne's asymptotically distribution-free criterion and scale-free least squares
- Evaluate models using more than two dozen fit statistics, including Chi-square; Akaike Information Criterion (AIC); Bayes and Bozdogan information criteria; Browne-Cudeck (BCC); ECVI, RMSEA and PCLOSE criteria; root mean square residual; Hoelter's critical n; and Bentler-Bonett and Tucker-Lewis indices
- Bootstrapping of user-defined functions of the model parameters, using the new User-Defined Estimands shortcut in the Start menu.
- Introduces the ability to export many models to the Bayesian modeling program Stan.
- A new version history option is being added that allows the user to view the previous versions of the current model. Using this option one can revert to any previous version in the version history.*