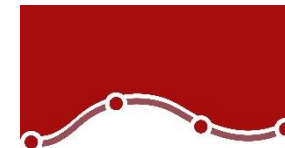




SPSS



**Statistics for
Data Analysis**

WHO WE ARE

SPS is an Italian center of statistical data analysis with more than 20 years of experience.

SPS was born in 1994 as SPSS Italia and it was the only reseller in Italy for SPSS software suite, authorised by SPSS inc.

Today SPS is an IBM Gold Business Partner, Software Support Provider and Expert Level in Data Science & Business Analytics.

CONTACTS

Registered office:
Via Antonio Zanolini, 36 A/B
40126 Bologna (BO)

Operational headquarters:
Via Isonzo, 55/2
40033 Casalecchio di Reno (BO)
P.I. 04222630370
Tel. 051-252573
www.spss.it

DATASHEET

Decision Trees

Statistics for Data Analysis

Organizations can solve a wide array of business and research problems with the solution [Statistics for Data Analysis](#).

Compared to other statistical software, the solution 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 Statistics for Data Analysis to help increase revenue, outmaneuver competitors, conduct research and make better decisions. With decades of built-in expertise and innovation, it's a leading choice for reliable statistical analysis.

Statistics Base is part of the solution Statistics for Data Analysis, which consists of:

- Software license
- Add-On
- SPS Service Program

This comprehensive, easy-to-use solution includes many different procedures and tests to help users solve complex business and research challenges.

Highlights Statistics for Data Analysis

- Get support through every step of the analytical process.
- Carry out essential analyses from an intuitive graphical interface.
- Select from more than a dozen integrated products to make specialized analyses faster and easier.

Statistics for Data Analysis

The solution analytical capabilities to meet the analysis requirements of any type of organization, from basic tools for solving common problems to advanced analytical techniques that enable all type of organization to address complex challenges.

Statistics for Data Analysis can help you:

- Analyze your data with new and advanced statistics, including a variety of new features within UNIANOVA methods
- Integrate better with third-party applications, including stronger integration with Microsoft Office
- Save time and effort with productivity enhancements:
 - More attractive and modern-looking charts in Chartbuilder
 - New groundbreaking features in Statistics Amos 25
 - Data and syntax editor enhancements
 - Accessibility improvements for the visually impaired
 - Updated merge user interface
 - Simplified toolbars

Statistics for Data Analysis can access quickly, manage and analyze any kind of dataset, including survey data, corporate databases or data downloaded from the web.

In addition, the software 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.

Business Benefit Statistics for Data Analysis

- Support business decisions with data-based analytics for improved outcomes.
- Be more confident in your results by incorporating data from many different sources, including geospatial information, 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.

Statistics Decision Trees

Datasheet

Easily identify groups and predict outcomes

Statistics Decision Trees creates classification and decision trees to help you better identify groups, discover relationships between groups and predict future events.

You can use classification and decision trees for:

- Segmentation
- Stratification
- Prediction
- Data reduction and variable screening
- Interaction identification
- Category merging
- Discretizing continuous variables

Highly visual diagrams enable you to present categorical results in an intuitive manner—so you can more clearly explain the results to non-technical audiences. These trees enable you to explore your results and visually determine how your model flows. Visual results can help you find specific subgroups and relationships that you might not uncover using more traditional statistics. Because classification trees break the data down into branches and nodes, you can easily see where a group splits and terminates.

Highlights:

- Identify groups, segments, and patterns in a highly visual manner with classification trees.
- Choose from CHAID, Exhaustive CHAID, C&RT and QUEST to find the best fit for your data.
- Present results in an intuitive manner—perfect for non-technical audiences.
- Save information from trees as new variables in data (information such as terminal node number, predicted value and predicted probabilities).

Statistics Decision Trees

Datasheet

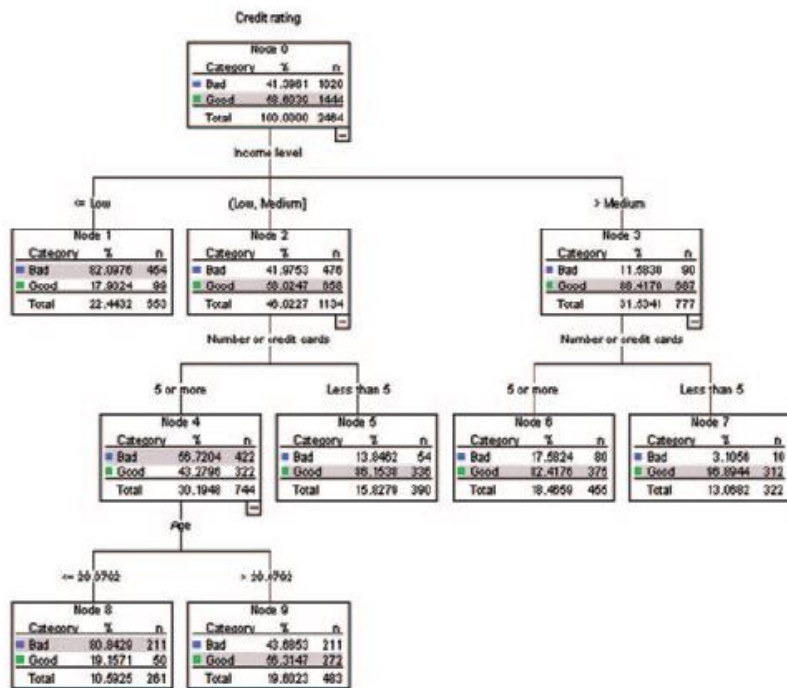


Figure 1: Use the highly visual trees to discover relationships that are currently hidden in your data. The diagrams, tables and graphs in Statistics Decision Trees are easy to interpret.

- Use Statistics Decision Trees in a variety of applications, including:
 - Database marketing:
 - Choose a response variable to segment your customer base (for example, responders/non-responders in a test mailing; high-, medium-, and low-profit customers; or recruits who have extended service versus those who haven't)
 - Profile groups based on other attributes, such as demographics or customer activity
 - Customize new promotions to focus on a specific subgroup, help reduce costs, and improve return on investment (ROI)
- Market research:
 - Perform customer, employee or other types of satisfaction surveys
 - Choose a variable that measures satisfaction (for example, on a "1-5" scale)
 - Profile satisfaction levels according to responses to other questions
 - Change factors, such as work environment or product quality, that can affect satisfaction

Statistics Decision Trees

Datasheet

- Credit risk scoring:
 - Determine risk groups (high, medium or low)
 - Profile risk groups based on customer information, such as account activity
 - Offer the right credit line to the right applicants based on risk group
- Program targeting:
 - Choose a variable with a desirable versus undesirable outcome (for example, successful completion of a welfare-to-work program)
 - Reveal the factors that lead to success, based on applicant information
 - Customize new programs to satisfy the needs of more people
- Marketing in the public sector:
 - Choose a response variable for segmenting your customer base (for example, potential college applicants who actually applied versus those who haven't)
 - Profile groups based on other attributes, such as demographics or customer activity
 - Customize new promotions to focus on a specific subgroup, help reduce costs and improve ROI
 - Statistics Decision Trees is available for installation as client-only software but, for greater performance and scalability, a server-based version is also available.

Statistics Decision Trees

Datasheet

Choose from four decision tree algorithms

Statistics Decision Trees includes four established tree-growing algorithms:

- CHAID—A fast, statistical, multi-way tree algorithm that explores data quickly and efficiently, and builds segments and profiles with respect to the desired outcome
- Exhaustive CHAID—A modification of CHAID that examines all possible splits for each predictor
- Classification and regression trees (C&RT)—A complete binary tree algorithm that partitions data and produces accurate homogeneous subsets
- QUEST—A statistical algorithm that selects variables without bias and builds accurate binary trees quickly and efficiently

With four algorithms, you have the ability to try different tree-growing methods and find the one that best fits your data.

Moreover, you can easily create classification trees and conveniently use the results to segment and group cases directly within the data. Additionally, you can generate selection or classification/prediction rules in the form of SPSS Statistics syntax, SQL statements or simple text (through syntax). You can display these rules in the Viewer and save them to an external file for later use to make predictions about individual and new cases.

Statistics Decision Trees Features

Trees

- Display tree diagrams, tree maps, bar graphs and data tables
- Easily build trees using the comprehensive interface, which enables the setup of:
 - Measurement level (nominal, ordinal, and continuous)
 - Independent variables
 - Dependent variables
 - Influence variables
 - Growing method
 - Output setup, which includes trees, statistics, charts, and rules
 - Split sample validation or cross-validation
 - Stopping criteria
 - Saved variables, including predicted values, probability and XML models
- Choose from four tree-growing methods
- View nodes using one of several methods: Show bar charts or tables of your target variables, or both, in each node
- Collapse and expand branches, and change other cosmetic properties, such as fonts and colors
- View and print trees
- Specify the exact zoom percentage for viewing visual tree models in the interface
- Automate tree building using the production mode
- Automatically generate syntax from the interface
- Force one predictor into the model
- Specify prior probabilities, misclassification costs, revenues, expenses and scale scores

Tree-growing algorithms

- Perform analysis using one of four powerful tree-growing algorithms:
 - CHAID by Kass (1980)
 - Exhaustive CHAID by Biggs, de Ville, and Suen (1991)
 - Classification & regression trees (C&RT) by Breiman, Friedman, Olshen, and Stone (1984)
 - QUEST by Loh and Shih (1997)

- Handle missing predictor data using one of two methods: Assign to a category or impute using a surrogate
- Discretize continuous predictor variables according to the number of categories specified
- Have pruning capabilities for C&RT and QUEST
- Randomly sample source data for split sample validation or use a variable to split the sample

Model evaluation

- Generate risk and classification tables
- Summarize node performance with evaluation graphs and tables to help identify the best segments:
 - Gains
 - Index (lift)
 - Response
 - Mean
 - Average profit
 - ROI
- Partition data between training and test data to verify accuracy
- Display summary graphs or classification rules for selected nodes using the node summary window

Deployment

- Export:
 - Tree diagrams, charts and tables Export formats include: HTML, text, Microsoft Word, Microsoft Excel, RTF and PDF
- Save information from the model as variables in the working data file
- Export decision rules that define selected segments in SQL to score databases, as SPSS Statistics syntax to score SPSS Statistics files, or as simple text (through syntax)
- For additional insight, select interesting segments in the working data file via tree nodes, and run more analyses

Statistics for Data Analysis solution

Add more analytical power, as you need it, with optional modules and stand-alone software from the Statistics for Data Analysis family.

Statistics Base

Statistics Base includes the core capabilities to take the analytical process from start to finish. It is easy to use and includes a broad range of procedures and techniques to increase revenue, outperform competitors, conduct research and make better decisions.

Statistics Advanced

Statistics Advanced includes these powerful multivariate techniques: generalized linear models (GENLIN), generalized estimating equations (GEE), mixed level models, general linear mixed models (GLMM), variance component estimation, MANOVA, Kaplan-Meier estimation, Cox regression, hiloglinear, loglinear and survival analysis.

Statistics Bootstrapping

Statistics Bootstrapping enables researchers and analysts to use bootstrapping techniques on a number of tests contained in Statistics for Data Analysis modules. This provides an efficient way to ensure that your models are stable and reliable. With Statistics 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.

Statistics 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.

Statistics Complex Samples

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

Statistics Conjoint

Statistics 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.

Statistics Tables

Use Statistics 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.

Statistics Preparation

With Statistics 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.

Statistics Decision Trees

Create highly visual classification and decision trees directly within Statistics for Data Analysis 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.

Statistics Direct Marketing

Statistics 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 and access pre-built models.

Statistics Exact Tests

Statistics 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.

Statistics Forecasting

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

Statistics Missing Values

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

Statistics Neural Networks

Use the Statistics 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.

Statistics 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.

Complementary product

Use these products with Statistics for Data Analysis to enhance your analytical results.

Statistics Amos

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. The latest release includes a new nongraphical method of model specification that improves accessibility for users who need scripting capabilities and enables large, complicated models to be run more quickly.