IBM SPSS Decision Trees

Easily identify groups and predict outcomes

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).

IBM® SPSS® 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.

Use SPSS 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
- 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

SPSS Decision Trees is available for installation as client-only software but, for greater performance and scalability, a server-based version is also available.
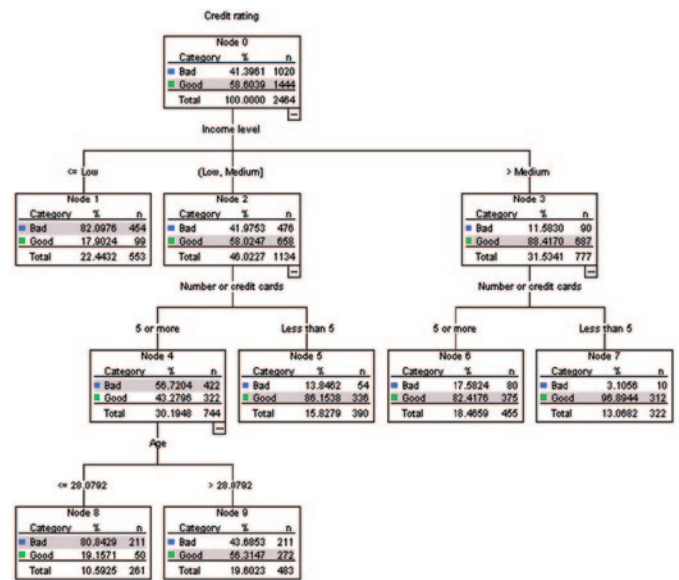

Figure 1: Use the highly visual trees to discover relationships that are currently hidden in your data. The diagrams, tables and graphs in SPSS Decision Trees are easy to interpret.
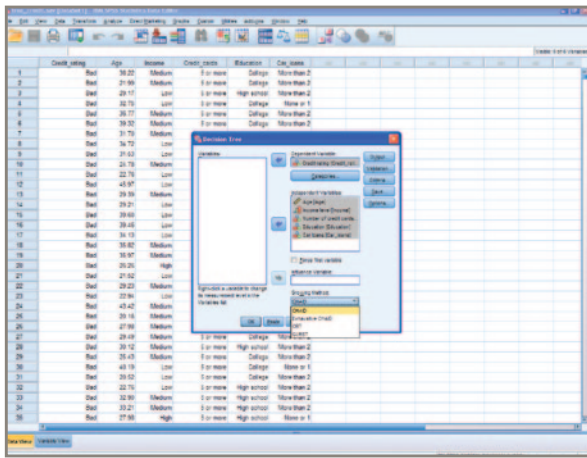
Figure 2: Use tree model results to score cases directly in SPSS Statistics.

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. More information about these valuable capabilities can be found at **ibm.com**/spss/cds.
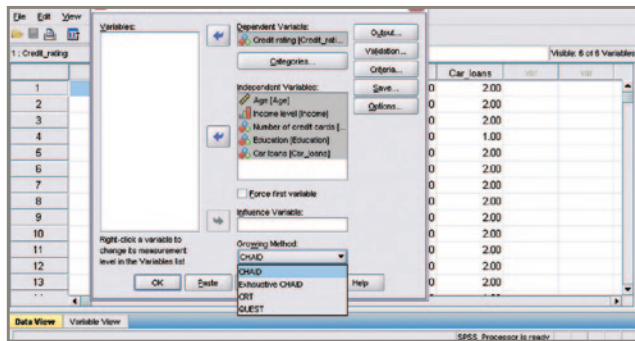
Choose from four decision tree algorithms

SPSS 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.

Our statistical software is available individually as well as 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.

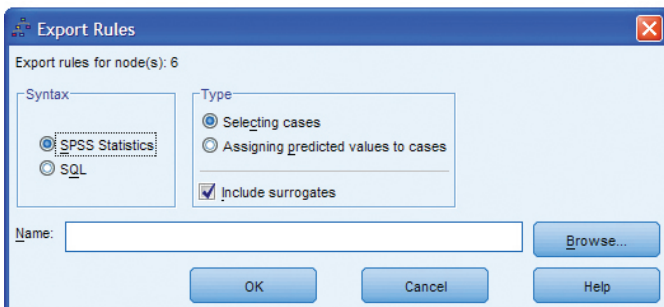*Figure 3*: Create tree models in SPSS Statistics using CHAID,
Exhaustive CHAID, C&RT or QUEST.

Figure 4: Directly select cases or assign predictions in your data from
the model results, or export rules for later use.

Extend your results using IBM SPSS Statistics Base

If you use SPSS Decision Trees with IBM SPSS Statistics Base, 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. If you'd like to use your results to score other data files, you can write information from the tree model directly to your data or create XML models for use in IBM SPSS Statistics Base Server.

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)
- Export trees as XML models for use with SPSS Statistics Server and the SPSS Statistics Portal
- Publish trees as images and tables as static or interactive tables to SPSS Statistics Portal
- For additional insight, select interesting segments in the working data file via tree nodes, and run more analyses

System requirements

- Requirements vary according to platform. For details, see **ibm.com**/spss/requirements.

About IBM Business Analytics

IBM Business Analytics software delivers actionable insights decision-makers need to achieve better business performance. IBM offers a comprehensive, unified portfolio of business intelligence, predictive and advanced analytics, financial performance and strategy management, governance, risk and compliance and analytic applications.

With IBM software, companies can spot trends, patterns and anomalies, compare "what if" scenarios, predict potential threats and opportunities, identify and manage key business risks and plan, budget and forecast resources. With these deep analytic capabilities our customers around the world can better understand, anticipate and shape business outcomes.

For more information

For further information, please visit
ibm.com/business-analytics.

Request a call

To request a call or to ask a question, go to
ibm.com/business-analytics/contactus. An IBM representative will respond to your inquiry within two business days.

Business Analytics software