SAS® Visual Statistics
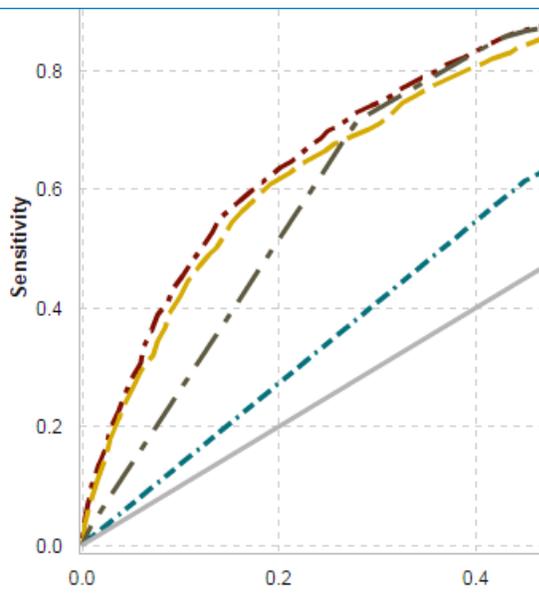
Interactively create, run and evaluate high-performance analytic models for instant results

§sas
THE POWER TO KNOW®

What does SAS® Visual Statistics do?

It provides a drag-and-drop web browser interface that empowers multiple users to explore massive data, and then interactively and iteratively create descriptive and predictive models. Distributed, in-memory processing dramatically shortens model development time so you can run complex analytic computations in just minutes.

Why is SAS® Visual Statistics important?

Your data scientists, statisticians and business analysts can build well-qualified predictive models by segments, test new ideas and fine-tune their models on the fly. Even the most difficult problems can be solved and new opportunities identified faster for better-informed decision making.

For whom is SAS® Visual Statistics designed?

SAS Visual Statistics is primarily designed for use by statisticians, data scientists, business analysts, etc., who need to analyze diverse data and interactively create and evaluate predictive models to quickly get precise insights.

Growing volumes and varieties of big data make it difficult to visualize and understand valuable relationships in data and get the analytically based answers you need to take the best actions. Traditional IT infrastructures are just not designed for rapid, iterative analytical processing and on-the-fly changes to predictive models.

It's hard for statisticians, data scientists and business analysts to build the number of models that are needed. They can't easily experiment with segments or groups, or quickly refine their models to identify the best one.

SAS® Visual Statistics solves these issues. As an add-on to SAS Visual Analytics, it combines interactive data exploration and discovery with the ability to easily build and adjust huge numbers of predictive models on the fly. It's really that easy – no coding is required. The in-memory engine reads data into memory once, putting an end to constant and expensive data shuffling.

Benefits

- **Gain new insights and identify areas of opportunity before your competitors do.** Quickly surface insights hidden in vast stores of data. Discover, analyze and evaluate new opportunities. Find new ways to grow revenue. Powerful, predictive analytics with visual data exploration capabilities enable business analysts and statisticians do more with data than ever before.

- **Run more models faster for more precision.** How long does it take to run your models? Hours? Our multicore processing environment reduces that to minutes. Build models to target specific groups or segments, and run numerous multiple scenarios simultaneously. Analytical professionals can ask more what-if questions and get fast answers. Refined models produce better results.

- **Stay agile with in-memory computing.** SAS Visual Statistics performs complex analytic computations using an in-memory engine. Modelers can quickly test new ideas, try different sophisticated modeling techniques and refine models on the fly – all using data volumes never before possible.

- **Improve the productivity of your analytical teams.** Multiple users can quickly and interactively customize their models – adding or changing variables, removing outliers, etc. – and instantaneously see how those changes affect model outcomes. Which model provides the most predictive power? It's now easy to find out. Organizations realize more value from their big data analytics.

Overview

SAS Visual Statistics provides an interactive, intuitive, drag-and-drop, web-browser interface for creating descriptive and predictive models on data of any size – rapidly! It takes advantage of SAS® LASR™ Analytic Server to persist and analyze data in memory and deliver near-instantaneous results.

When combined with SAS Visual Analytics , you get a fast, single environment for interactive data exploration and model development.

SAS Visual Statistics is designed for statisticians, data scientists and business analysts who want to visually and instantly interact with and analyze complex data. The easy-to-use, drag-and-drop interface provides nonprogramming access to powerful SAS statistical modeling and machine-learning techniques. These techniques are used to predict outcomes that result in better, more targeted actions.

Interactive data exploration and modeling environment

SAS Visual Statistics is an add-on to SAS Visual Analytics Explorer. The common SAS Visual Analytics Explorer environment provides interactive data exploration and analytical modeling capabilities. You can quickly identify predictive drivers among multiple exploratory variables, and interactively discover outliers and data discrepancies. Then, use that information to populate your interactive environment for sophisticated predictive modeling.

The web browser interface makes it a simple drag-and-drop process to create powerful descriptive and predictive models. Multiple users can easily collaborate to build and refine the best models. Interactive processing is very fast so users can quickly and easily experiment with different techniques.

Descriptive modeling

Clustering is the task of segmenting a heterogeneous population into a number of more homogenous subgroups or clusters based on self-similarity, and it is often
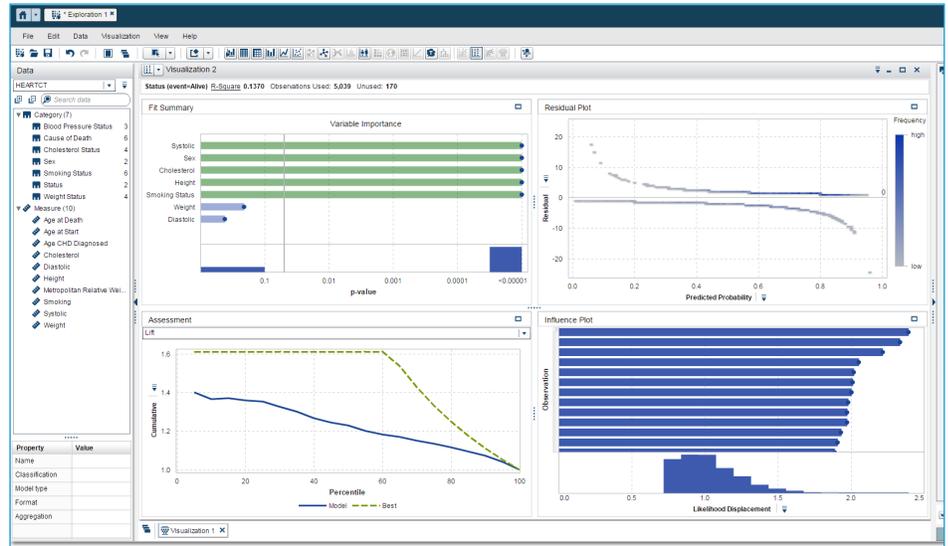

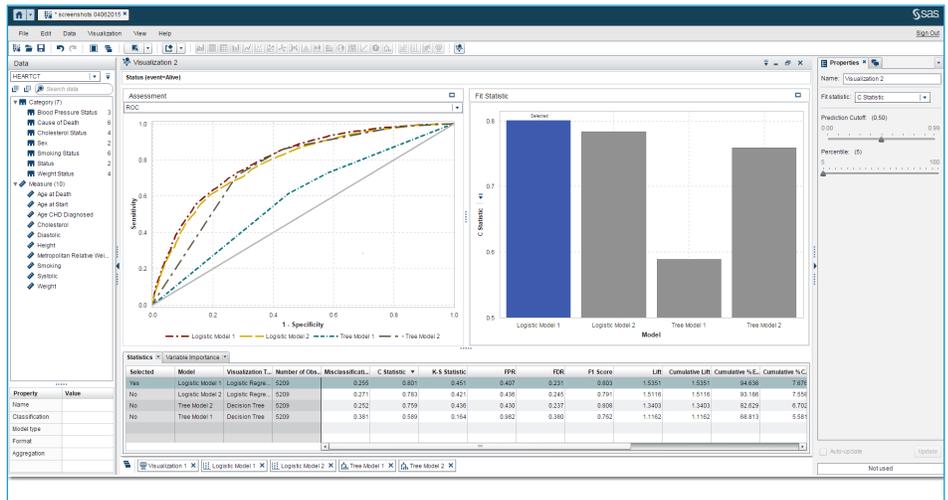Figure 1: Predict binary outcomes using logistic regression.

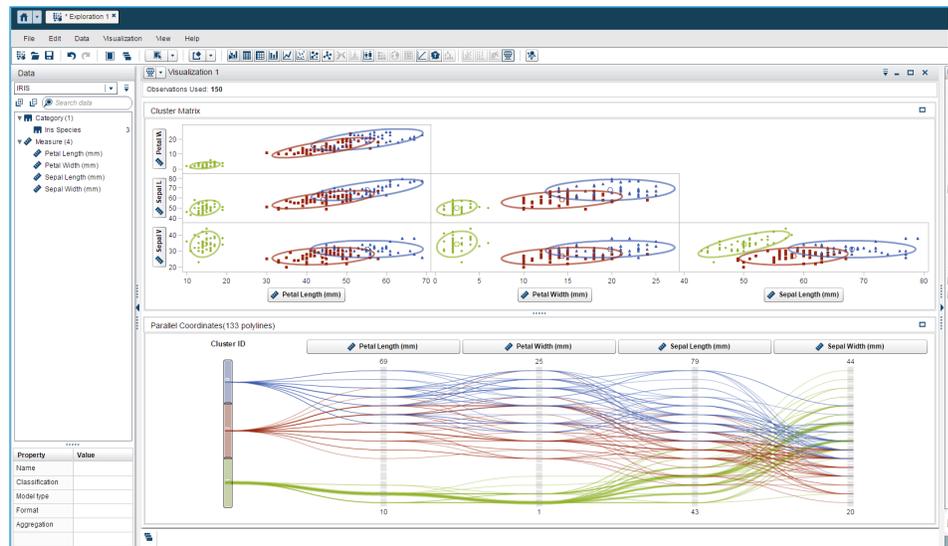Figure 2: Compare competing models to determine which will produce the best outcome.

Figure 3: Use k-means clustering to segment data.

performed as a prelude to some other type of data mining. For example, market segmentation clusters customers with similar buying habits to find out which promotion will work best. SAS Visual Statistics lets you visually explore and evaluate segments for further analysis using k-means clustering, scatter plots and detailed summary statistics.

Predictive modeling

Prediction deals with classifying records according to a future behavior or estimated future value. SAS Visual Statistics makes it easy to create predictive models using techniques such as linear regression, generalized linear modeling, logistic regression and classification trees. Classification helps with prediction of discrete outcomes such as fraud or no fraud, credit extension or no credit extension, etc. Multilevel classification helps you make determinations like voluntary churn, involuntary churn or active customer.

Dynamic group-by processing

Concurrently build numerous models and process results for each group or segment without having to sort or index data each time. The grouping variables, or their properties, can change from one action to the next, and groups are processed without reordering the data. This means more results can be generated for each group on the fly without additional processing overhead.

In-memory analytical processing

Build models faster. With the SAS LASR Analytic Server, there's no need to write data to disk or perform data shuffling. SAS Visual Statistics loads all data into memory once and interacts with the data without reloading it each time a new task is performed. This means the impact of changes to models (e.g., adding new variables or removing outliers) is instantly visible. Because it is designed for concurrent processing, many users can create and run complex models simultaneously. Data and analytic workloads are performed in distributed form across multiple server nodes, and are multi-threaded on each node, for blazingly fast speeds.

Key Features

Exploratory Modeling (available through SAS® Visual Analytics Explorer)
- Quickly interpret complex relationships or key variables influencing predictive modeling outcomes within large data sets.
- Understand a variable's level of influence on the overall model lift.
- Detect outliers and/or influence points to help you determine, capture and remove them for further analysis.
- Explore data using bar charts, histograms, box plots, heat maps, bubble plots, geographic maps and more (with SAS Visual Analytics Explorer).
- Derive models directly from correlation matrices, scatter plots, heat maps and box plots.
- Derive predictive outputs or segmentations that can be used directly in other modeling or visualization tasks. Derived outputs can be saved and passed to a user who lacks a model building role and those capabilities.
- Generate model score codes. Score codes are automatically concatenated if a model contains the derived output from other models (leaf ID, cluster ID, etc.).

Interactive Descriptive and Predictive Modeling Techniques
- Filtering of observations.
- Interactively exclude outliers; model is automatically refit.
- User-defined event level.
- Auto-update model toggling for all of the modeling techniques below.

Clustering
- k-means clustering.
- Parallel coordinate plots to interactively evaluate cluster membership.
- Scatter plots of inputs with cluster profiles overlaid for small data sets and heat maps with cluster profiles overlaid for large data sets.
- Ability to export cluster segment variables for further analysis.
- Detailed summary statistics (means of each cluster, number of observations in each cluster, etc.).
- Generate on-demand cluster ID as a new column.

Classification tree
- Based on the C4.5 algorithm (information gain or information gain ratio).
- Interactively grow and prune a tree.
- Set tree depth, max branch, leaf size, aggressiveness of tree pruning and more.
- Bin continuous response with a user-specified number of bins.
- Use tree map and tree overview displays to interactively navigate the tree structure.
- Generate on-demand cluster ID as a new column.

Generalized linear model
- Distributions supported include beta, normal, binary, exponential, gamma, geometric, Poisson, inverse Gaussian and negative binomial.
- Set convergence and iteration criteria.
- Offset variable support.
- Frequency and weight variables.
- Residual diagnostics.
- Summary table includes model summary, iteration history, fit statistics, Type III test table and parameter estimates.
- Informative missing option for treatment of missing values on the predictor variable.
- Generate on-demand predicted values and residuals as new columns.

Logistic regression
- Models for binary data with logit and probit link functions.
- Influence statistics.
- Variable selection.
- Offset variable support.

Model comparison and assessment

Generate model comparison summaries such as lift charts, ROC charts, concordance statistics and misclassification tables on one or more models. An interactive slider lets you manipulate cutoff thresholds so you can easily and visually evaluate lift at different percentiles. Combine model fitting with model diagnostics to quickly see and understand impacts on performance.

Model scoring

Export your models as SAS DATA step code, and apply them to new data.

Run in the cloud

Subscribe to SAS Visual Statistics for SAS Cloud to begin exploring data and building and refining your models faster than ever. Analytics users can rely on an always-on, cloud-based infrastructure to quickly build well-qualified predictive models by groups or segments, test new ideas and fine-tune models on the fly.

Key Features (continued)

- Frequency and weight variables.
- Residual diagnostics.
- Summary table includes model dimensions, iteration history, fit statistics, convergence status, Type III tests, parameter estimates and response profile.
- Informative missing option for treatment of missing values on the predictor variable.
- Generate on-demand predicted labels and predicted event probabilities as new columns. Adjust the prediction cutoff to label an observation as event or non-event.

Linear regression
- Influence statistics.
- Variable selection.
- Frequency and weight variables.
- Residual diagnostics.
- Summary table includes overall ANOVA, model dimensions, fit statistics, model ANOVA, Type III test and parameter estimates.
- Informative missing option for treatment of missing values on the predictors.
- Generate on-demand predicted values and residuals as new columns.

Group-By Processing
- Build models, compute and process results on the fly for each group or segment without having to sort or index the data each time.
- Build segment-based models instantly (i.e., stratified modeling) from a decision tree or clustering analysis.

Model Assessment and Comparison
- Generate model comparison summaries such as lift charts, ROC charts, concordance statistics and misclassification tables for one or more models.
- Interactively slide the prediction cutoff for automatic updating of assessment statistics and misclassification tables.
- Interactively evaluate lift at different percentiles.

Model Scoring
- Export models as SAS DATA step code to integrate models with other applications. Score codes are automatically concatenated if a model is using the derived outputs from other models.

Highly Scalable In-Memory Processing
- An in-memory analytic engine provides a secure, scalable, multiple-user environment.
- Reads data into memory once and persists it for multiple users to interact with in real time. No need to write data to disk or perform data shuffling during different analytical processing steps.
- Available as a single-server mode for smaller organizations or departments, or for distributed environments that take advantage of multinode, multithreaded processing.

To contact your local SAS office, please visit: sas.com/offices

§sas
THE POWER TO KNOW®