



SAS/STAT® Software

Providing the foundation for SAS® Analytics

What does SAS/STAT® software do?

From traditional analysis of variance and predictive modeling to exact methods and statistical visualization techniques, SAS/STAT software provides tools for both specialized and enterprisewide analytical needs.

Why is SAS/STAT® software important?

Increasingly organizations are turning to statistical analysis to guide decision-making processes. Using optimal statistical techniques can provide new information that improves processes, drives development and revenue and helps you retain valued and satisfied customers. SAS/STAT software provides a complete, comprehensive set of tools that can meet the data analysis needs of the entire organization.

For whom is SAS/STAT® software designed?

SAS/STAT software is designed for use by business analysts, statisticians, researchers and engineers.

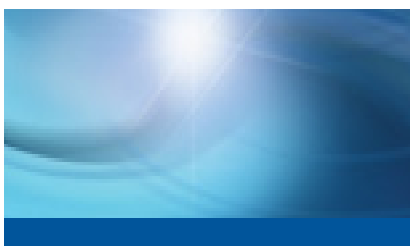
Organizations today depend on data and analysis to provide new insights that deliver competitive advantage. Increased competition in the marketplace demands a deeper understanding of customers' behavior. Pharmaceutical companies must perform additional analyses to comply with federal regulations. And research organizations need cutting-edge statistical techniques as they investigate new areas of scientific discovery.

For many though, the data sources have gotten so large and varied that the majority of statistical software products just can't handle them. On top of that, increasingly advanced statistical methods are called for along with analyses that can be verified and documented to meet government compliance standards.

SAS/STAT software provides extensive statistical capabilities for both specialized and enterprisewide analytic needs. Whether you want to perform customer preference studies for market research, analyze the results of clinical trials, predict credit card usage patterns, model air pollution patterns or produce a sample of people to administer a health behavior survey, SAS/STAT is the software of choice. And, because SAS remains committed to its long tradition of constantly enriching its statistical offerings, you know that you will have access to the most up-to-date statistical techniques, not just today, but well into the future.

Key Benefits

- Take advantage of all data in order to uncover new business opportunities and increase revenue.** SAS/STAT software is designed to handle large data sets from disparate sources, enabling you to take advantage of all data that is available for your analyses. Analysts are freed to focus on analysis rather than data issues.
- Move the scientific discovery process forward by applying the latest statistical techniques.** Statistical procedures in SAS are constantly being updated to reflect the latest advances in statistical methodology, thus enabling you to go beyond the basics for more advanced statistical analyses. In addition, technical support is provided by experienced master's- and doctorate-level statisticians who provide a level of service not often found with other software vendors.
- Achieve corporate and governmental compliance.** SAS has more than 30 years of experience developing analytic software and a proven reputation for delivering superior, reliable results. With SAS/STAT software, you can produce repeatable code that is easily documented and verified for corporate and governmental compliance issues.



Product Overview

SAS/STAT software provides the foundation for many of SAS' analytic offerings. It delivers a complete, comprehensive set of tools that can meet the data analysis requirements of an entire organization.

Analysis of variance

Analysis of variance is a technique for analyzing experimental data. With SAS/STAT software, you can perform analysis of variance for balanced or unbalanced designs, multivariate analysis of variance and repeated measurements analysis of variance.

Mixed models

A mixed model contains both fixed and random effects. SAS/STAT software provides a suite of procedures for fitting both linear and nonlinear mixed models as well as generalized linear mixed models. A procedure for solving large mixed model problems using sparse matrix techniques is also provided.

Regression

Regression analysis examines the relationship between a response variable and a set of explanatory variables. SAS/STAT software's general regression procedure uses least squares to estimate parameters, includes many different model selection methods and produces a variety of diagnostic measures. More specialized procedures fit generalized linear models, mixed linear models, nonlinear models and quadratic response surface models.

Categorical data analysis

In categorical data the outcome of interest reflects categories with data often presented in tabular form, known as contingency tables. With SAS/STAT software, you can investigate the association in a contingency table as well as produce measures that indicate the strength of that relationship. You also can

use parametric models to investigate the variation of a function of the outcome variable across the various levels.

Bayesian analysis

Bayesian analysis provides a natural and principled way of combining prior information with data, within a solid decision-theoretical framework. It provides inferences that are conditional on the data and are exact, without relying on approximation methods. SAS/STAT software provides Bayesian analysis capabilities for generalized linear models, accelerated life failure models and Cox regression models. SAS also provides flexible simulation procedures that are suitable for fitting a wide range of Bayesian models.

Multivariate analysis

Multivariate analyses encompass a variety of methods for modeling data with two or more response variables or for identifying relationships among several variables without designating particular variables as response or explanatory variables. You can use common factor analysis to explain the correlations among a set of variables in terms of a limited number of unobservable, or latent, variables. Principal component analysis summarizes a large number of variables with a small number of linear combinations. Structural equations can express relationships among several variables that can be either directly observed variables or unobserved hypothetical variables.

Survival analysis

Survival analysis concerns data that measures lifetime or time until an event. In many cases, only upper or lower bounds are available for some observations. This is known as censoring. Survival analysis methods correctly use both censored and uncensored observations. SAS/STAT software compares survival distributions for the event-time variable, fits accelerated failure time

models to event-time data and performs regression analysis based on the proportional hazards model.

Psychometric analysis

Psychometric methods are well-suited for analyzing data on human judgment and perception, such as market research data, but can be used for many other types of data. Multidimensional scaling estimates the coordinates of a set of objects and is often used to analyze judgments of product similarity.

Cluster analysis

Cluster analysis places observations into groups based on similarity. SAS/STAT software performs hierarchical clustering of multivariate data or distance data with graphical tree diagrams, disjoint clustering of large data sets and nonparametric clustering with hypothesis tests for the number of clusters.

Nonparametric analysis

Nonparametric analysis analyzes data not requiring specific distributional assumptions such as normality. SAS/STAT software performs nonparametric analysis of variance as well as other rank tests for balanced or unbalanced one-way or two-way designs. Exact probabilities are computed for many nonparametric statistics.

Survey data analysis

Researchers often use sample survey methodology to obtain information about a large population by selecting and measuring a sample from that population. SAS/STAT software provides tools for selecting probability-based random samples from a study population and analyzing the sample survey data. Procedures for computing summary statistics, analyzing contingency tables and building linear and logistic regression models enable you to incorporate the sample design into your analysis.

Multiple imputation

Missing values are an issue in many analyses. SAS/STAT software's multiple imputation procedures enable you to fill in missing values multiple times to generate a series of complete data sets, analyze the data sets using standard statistical analyses and combine the results to produce valid statistical inferences.

Power and sample size computation

Study planning procedures enable you to optimize your resource usage during the design of a study, aid in determining the sample size required to get statistically significant results and help you to detect a meaningful effect. SAS/STAT software includes an easy-to-use interface to these capabilities.

Statistical graphics

SAS software's statistical procedures create statistical graphics as automatically as they create tables. This functionality is referred to as ODS Statistical Graphics and requires minimal syntax to create production-quality graphics. It provides displays commonly needed for data analysis and statistical modeling, including scatter plots, diagnostic plots, histograms, box-and-whisker plots and more.

Postfitting inference

The linear modeling procedures offer a breadth of postfitting analyses once you have fitted your model and estimated its parameters. SAS/STAT software performs postfitting inference with model fit information stored from these same procedures, so you can perform additional analyses without refitting your model. Thus you have the option of saving your model information and doing postfitting later.

Key Features

Analysis of variance

- Balanced and unbalanced designs.
- Multivariate analysis of variance and repeated measurements.
- Linear and nonlinear mixed models.

Mixed models

- Linear mixed models.
- Nonlinear mixed models.
- Generalized linear mixed models.

Regression

- Least squares regression with nine model selection techniques, including stepwise regression.
- Diagnostic measures.
- Robust regression.
- Loess regression.
- Nonlinear regression and quadratic response surface models.
- Partial least squares.
- Quantile regression.

Categorical data analysis

- Contingency tables and measures of association.
- Logistic regression and log linear models; generalized linear models.
- Bioassay analysis.
- Generalized estimating equations.
- Weighted least squares regression.
- Exact methods.
- Zero-inflated Poisson regression.
- Zero-inflated negative binomial regression.

Bayesian

- Bayesian modeling and inference for generalized linear models, accelerated life failure models, Cox regression models and piecewise exponential models.
- General procedure fits Bayesian models with arbitrary priors and likelihood functions.

Multivariate analysis

- Factor analysis.
- Principal components.
- Canonical correlation and discriminant analysis.
- Path analysis.
- Structural equations.

Survival analysis

- Comparison of survival distributions.
- Accelerated failure time models.
- Proportional hazards models.

Psychometric analysis

- Multidimensional scaling.
- Conjoint analysis with variable transformations.
- Correspondence analysis.

Cluster analysis

- Hierarchical clustering of multivariate data or distance data.
- Disjoint clustering of large data sets.
- Nonparametric clustering with hypothesis tests for the number of clusters.

Continued on next page

Key Features (continued from previous page)

Nonparametric analysis

- Nonparametric analysis of variance. Exact probabilities computed for many nonparametric statistics.
- Kruskal-Wallis, Wilcoxon-Mann-Whitney and Friedman tests.
- Other rank tests for balanced or unbalanced one-way or two-way designs.

Survey data analysis

- Sample selection.
- Descriptive statistics and t -tests.
- Linear and logistic regression.
- Frequency table analysis.
- Cox proportional hazards model.

Multiple imputation for missing values

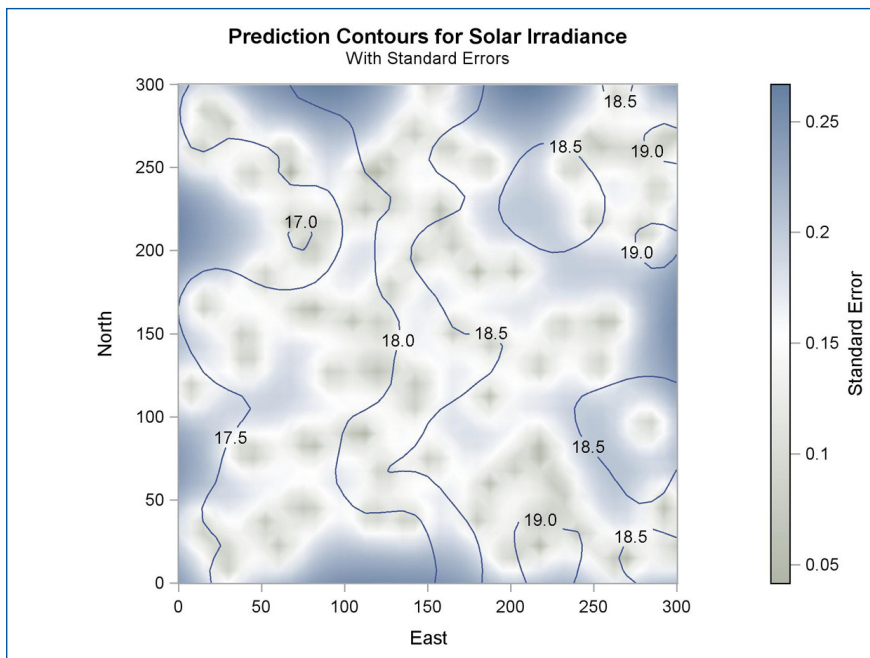
- Regression and propensity scoring for monotone missing patterns.
- MCMC method for arbitrary missing patterns.
- Combine results for statistically valid inferences.

Study planning

- The Power and Sample Size application provides an interface for computation of sample sizes and characterization of power for t -tests, confidence intervals, linear models, tests of proportions and rank tests for survival analysis.

SAS/STAT® System Requirements

To learn more about SAS/STAT system requirements, download white papers, view screenshots and see other related material, please visit www.sas.com/stat.



Statistical procedures create graphics as automatically as they create tables.