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

Missing Values

Statistics for Data Analysis

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.

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.

SPS S.r.l. P.iva 04222630370
Sede Legale: Via Antonio Zanolini, 36 AB 40126 Bologna (BO) – Sede Operativa: Via Isonzo, 55/2 - 40033 Casalecchio di Reno (BO)
Le idee, i concetti e le informazioni contenute in questo documento costituiscono comunicazione riservata e non possono essere usate o duplicate senza autorizzazione.
La riproduzione di una parte o dell'intero documento senza approvazione di SPS srl è espressamente proibita.

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.

Le idee, i concetti e le informazioni contenute in questo documento costituiscono comunicazione riservata e non possono essere usate o duplicate senza autorizzazione.
La riproduzione di una parte o dell'intero documento senza approvazione di SPS srl è espressamente proibita.

Statistics Missing Values

Datasheet

Build better models when you fill in the blanks

Highlights:

- Easily examine data from different angles.
- Quickly diagnose missing data problems.
- Replace missing values with estimates.
- Display missing data types and any
- extreme values.
- • Remove hidden bias.

Survey and market researchers, social scientists, data miners, and many other professionals rely on Statistics Missing Values to validate their research data.

When you ignore or exclude missing data, you risk obtaining biased or insignificant results. Use Statistics Missing Values to impute your missing data and draw more valid conclusions. Statistics Missing Values is a critical tool for anyone concerned about data validity. You can easily examine your data to uncover missing data patterns, and then estimate summary statistics and impute missing values through statistical algorithms.

For example, improve survey questions that you've identified as possibly confusing based on observed missing data patterns. You can even determine if missing values for one variable are related to missing values of another with the percent mismatch of patterns table.

You might find that respondents who skip a question on income might also bypass a question about education level. Use this information to enhance the quality of your surveys in the future.

SPS S.r.l. P.iva 04222630370
Sede Legale: Via Antonio Zanolini, 36 AB 40126 Bologna (BO) – Sede Operativa: Via Isonzo, 55/2 - 40033 Casalecchio di Reno (BO)
Le idee, i concetti e le informazioni contenute in questo documento costituiscono comunicazione riservata e non possono essere usate o duplicate senza autorizzazione.
La riproduzione di una parte o dell'intero documento senza approvazione di SPS srl è espressamente proibita.

Statistics Missing Values

Datasheet

Quickly and easily diagnose your missing data

You can quickly diagnose a serious missing data problem using the overall summary of missing values report. The missing values pattern report then provides a case-by-case overview of your data. It displays a snapshot of each type of missing value and any extreme values for each case.

Use the flexible separate variance t test and crosstabulation of categorical variables tables to discover if significant differences exist between respondents and non-respondents. These reports help you decide if missing data might cause problems in your analysis.

Receive a summary of missing data patterns and highlights of the variable sets that comprise the patterns with the tabulated pattern reports.

Use multiple imputation to replace missing data values

In Statistics Missing Values, the multiple imputation procedure helps you understand patterns of “missingness” in your dataset and enables you to replace missing values with plausible estimates. It offers a fully automatic imputation mode that chooses the most suitable imputation method based on characteristics of your data, while also allowing you to customize your imputation model.

SPS S.r.l. P.iva 04222630370
Sede Legale: Via Antonio Zanolini, 36 AB 40126 Bologna (BO) – Sede Operativa: Via Isonzo, 55/2 - 40033 Casalecchio di Reno (BO)
Le idee, i concetti e le informazioni contenute in questo documento costituiscono comunicazione riservata e non possono essere usate o duplicate senza autorizzazione.
La riproduzione di una parte o dell’intero documento senza approvazione di SPS srl è espressamente proibita.

Statistics Missing Values

Datasheet

Several complete datasets are generated (typically, three to five), each with a different set of replacement values. Next, you can model the individual datasets using the usual techniques, such as linear regression, to produce parameter estimates for each dataset. Then obtain final parameter estimates. This involves pooling the individual sets of parameter estimates obtained in step two and computing inferential statistics that take into account variation within and between imputations.

Analysis of the individual datasets and pooling of the results are supported via select existing Statistics procedures such as regression. When operating on datasets with imputed values, existing procedures will automatically produce pooled parameter estimates.

Reach more valid conclusions

Replace missing values with estimates and increase the chance of receiving statistically significant results. Remove hidden bias from your data by replacing missing values with estimates to include all groups in your analysis — even those with poor responsiveness.

SPS S.r.l. P.iva 04222630370
Sede Legale: Via Antonio Zanolini, 36 AB 40126 Bologna (BO) – Sede Operativa: Via Isonzo, 55/2 - 40033 Casalecchio di Reno (BO)
Le idee, i concetti e le informazioni contenute in questo documento costituiscono comunicazione riservata e non possono essere usate o duplicate senza autorizzazione.
La riproduzione di una parte o dell'intero documento senza approvazione di SPS srl è espressamente proibita.

Statistics Missing Values 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 non-missing groups for a related variable with the separate variance t test table:
  - t test, degrees of freedom, mean, value and count
- Show differences between present and missing data for categorical variables using the distribution of categorical variables table:
  - Produce crosstabs 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

Statistics

- Univariate: Compute count, mean, standard deviation, and standard error of mean for all cases, excluding those containing missing values, counts, percent of missing values, and extreme values for all variables
- Listwise: Compute mean, covariance matrix, and correlation matrix for all quantitative variables for cases excluding missing values
- Pairwise: Compute frequency, mean, variance, covariance matrix, and correlation matrix

Multiple Imputation

- Specify which variables to impute and specify constraints on the imputed values, such as minimum and maximum value, as well as 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, and predictive mean matching is an option for continuous outcome, ensuring 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 have a monotone pattern of missingness
  - Fully conditional specification (FCS): an iterative Markov Chain Monte Carlo (MCMC) method that is appropriate when the data have an arbitrary (monotone or no monotone) missing pattern
  - Automatic: scans the data to determine the best imputation method (monotone or FCS)

SPS S.r.l. P.iva 04222630370
Sede Legale: Via Antonio Zanolini, 36 AB 40126 Bologna (BO) – Sede Operativa: Via Isonzo, 55/2 - 40033 Casalecchio di Reno (BO)
Le idee, i concetti e le informazioni contenute in questo documento costituiscono comunicazione riservata e non possono essere usate o duplicate senza autorizzazione.
La riproduzione di una parte o dell'intero documento senza approvazione di SPS srl è espressamente proibita.

- Specify:
  - The number of imputations
  - The range of imputed values
  - Whether or not 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
- Specify a variable containing analysis (regression) weights: This procedure incorporates analysis weights in regression and classification models used to impute missing values, and analysis weights are also used in summaries of imputed values (e.g., 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, obtain analysis of missing values by variable as well as tabulated patterns of missing values, or obtain descriptive statistics for imputed values
- Graphically summarize missingness for cases, variables, and individual data (cell) values
- Request an Statistics data file containing imputed values and/or an FCS iteration history
- Multiple imputation datasets 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, crosstabs, correlations, nonparametric correlation, partial correlation
- Comparison of means: means, t test, non-parametric 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 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 due to nonresponse
  - Relative Efficiency: efficiency of estimate for M imputations relative to that for an infinite number of imputations
- Obtain Model PMML for pooled parameter estimates: Linear regression, Generalized Linear Models, multinomial logistic regression, binary logistic regression, discriminant analysis, Cox Regression

Le idee, i concetti e le informazioni contenute in questo documento costituiscono comunicazione riservata e non possono essere usate o duplicate senza autorizzazione.
La riproduzione di una parte o dell'intero documento senza approvazione di SPS srl è espressamente proibita.

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.

Le idee, i concetti e le informazioni contenute in questo documento costituiscono comunicazione riservata e non possono essere usate o duplicate senza autorizzazione.
La riproduzione di una parte o dell'intero documento senza approvazione di SPS srl è espressamente proibita.

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.

SPS S.r.l. P.iva 04222630370
Sede Legale: Via Antonio Zanolini, 36 AB 40126 Bologna (BO) – Sede Operativa: Via Isonzo, 55/2 - 40033 Casalecchio di Reno (BO)
Le idee, i concetti e le informazioni contenute in questo documento costituiscono comunicazione riservata e non possono essere usate o duplicate senza autorizzazione.
La riproduzione di una parte o dell'intero documento senza approvazione di SPS srl è espressamente proibita.

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.

SPS S.r.l. P.iva 04222630370
Sede Legale: Via Antonio Zanolini, 36 AB 40126 Bologna (BO) – Sede Operativa: Via Isonzo, 55/2 - 40033 Casalecchio di Reno (BO)
Le idee, i concetti e le informazioni contenute in questo documento costituiscono comunicazione riservata e non possono essere usate o duplicate senza autorizzazione.
La riproduzione di una parte o dell'intero documento senza approvazione di SPS srl è espressamente proibita.