IBM SPSS
Missing Values

Build better models when you fill in the blanks

Highlights

Make IBM SPSS Missing Values part of your data management and preparation, and ensure that you take missing values into account.

- 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 IBM® SPSS® Missing Values* to validate their research data.

When you ignore or exclude missing data, you risk obtaining biased or insignificant results. Use IBM SPSS Missing Values to impute your missing data and draw more valid conclusions. IBM SPSS Missing Values is a critical tool for anyone concerned about data validity. You can easily examine your data to uncover missing data patterns. 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.

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.
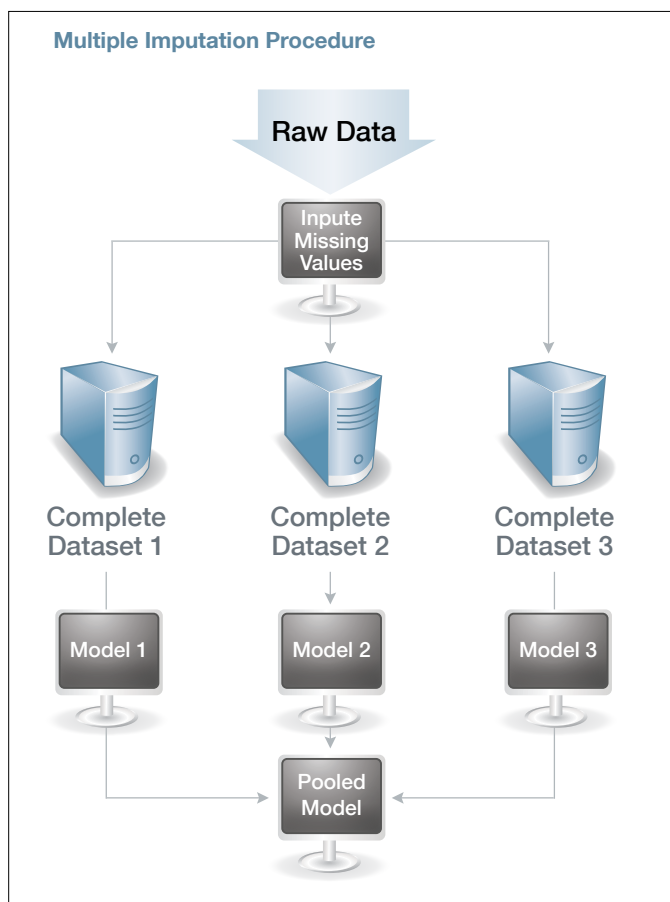
* IBM SPSS Missing Values was formerly called PASW® Missing Values.

Use multiple imputation to replace missing data values

In 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.

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.

Multiple Imputation Procedure

Raw Data

Inpute Missing Values

Complete Dataset 1   Complete Dataset 2   Complete Dataset 3

Model 1   Model 2   Model 3

Pooled Model

Analysis of the individual datasets and pooling of the results are supported via select existing IBM SPSS 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.

Our suite of statistical software is now available 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.

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

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 (formerly PASW® Collaboration and Deployment Services). More information about these valuable capabilities can be found at **www.spss.com/software/ deployment/cds**.

About IBM Business Analytics

IBM Business Analytics software delivers complete, consistent and accurate information that decision-makers trust to improve business performance. A comprehensive portfolio of business intelligence, predictive analytics, financial performance and strategy management, and analytic applications provides clear, immediate and actionable insights into current performance and the ability to predict future outcomes. Combined with rich industry solutions, proven practices and professional services, organizations of every size can drive the highest productivity, confidently automate decisions and deliver better results.

As part of this portfolio, IBM SPSS Predictive Analytics software helps organizations predict future events and proactively act upon that insight to drive better business outcomes. Commercial, government and academic customers worldwide rely on IBM SPSS technology as a competitive advantage in attracting, retaining and growing customers, while reducing fraud and mitigating risk. By incorporating IBM SPSS software into their daily operations, organizations become predictive enterprises – able to direct and automate decisions to meet business goals and achieve measurable competitive advantage.

* *IBM SPSS Statistics and IBM SPSS Statistics Base were formerly called PASW® Statistics and PASW® Statistics Base.*

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 values. You can also 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. Predictive mean matching is an option for continuous outcomes; this ensures 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)

- 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
- You can also specify a variable containing analysis (regression) weights. The procedure incorporates analysis weights in regression and classification models used to impute missing values. 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. You can obtain analysis of missing values by variable as well as tabulated patterns of missing values. Optionally, you can obtain descriptive statistics for imputed values.
- Graphically summarize missingness for cases, variables, and individual data (cell) values
- Request an IBM SPSS 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

*** Multi-threaded procedures*

- 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

System requirements
- Requirements vary according to platform. For details, see www.spss.com/statistics.

Features subject to change based on final product release

Business Analytics software