

# IBM SPSS Missing Values

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

When you ignore missing data, you risk obtaining biased or insignificant results. Use SPSS Missing Values to impute your missing data and draw more valid conclusions. SPSS 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.

This improvement in data modelling brings several potential benefits. For example, you could improve survey questions that you've identified as possibly confusing based on observed missing data patterns—or even use a percent mismatch of patterns table to determine if missing values for one variable are related to missing values of another.

This improved insight could show you, for example, that respondents who skip a question on income might also bypass a question about education level—and this information can 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.

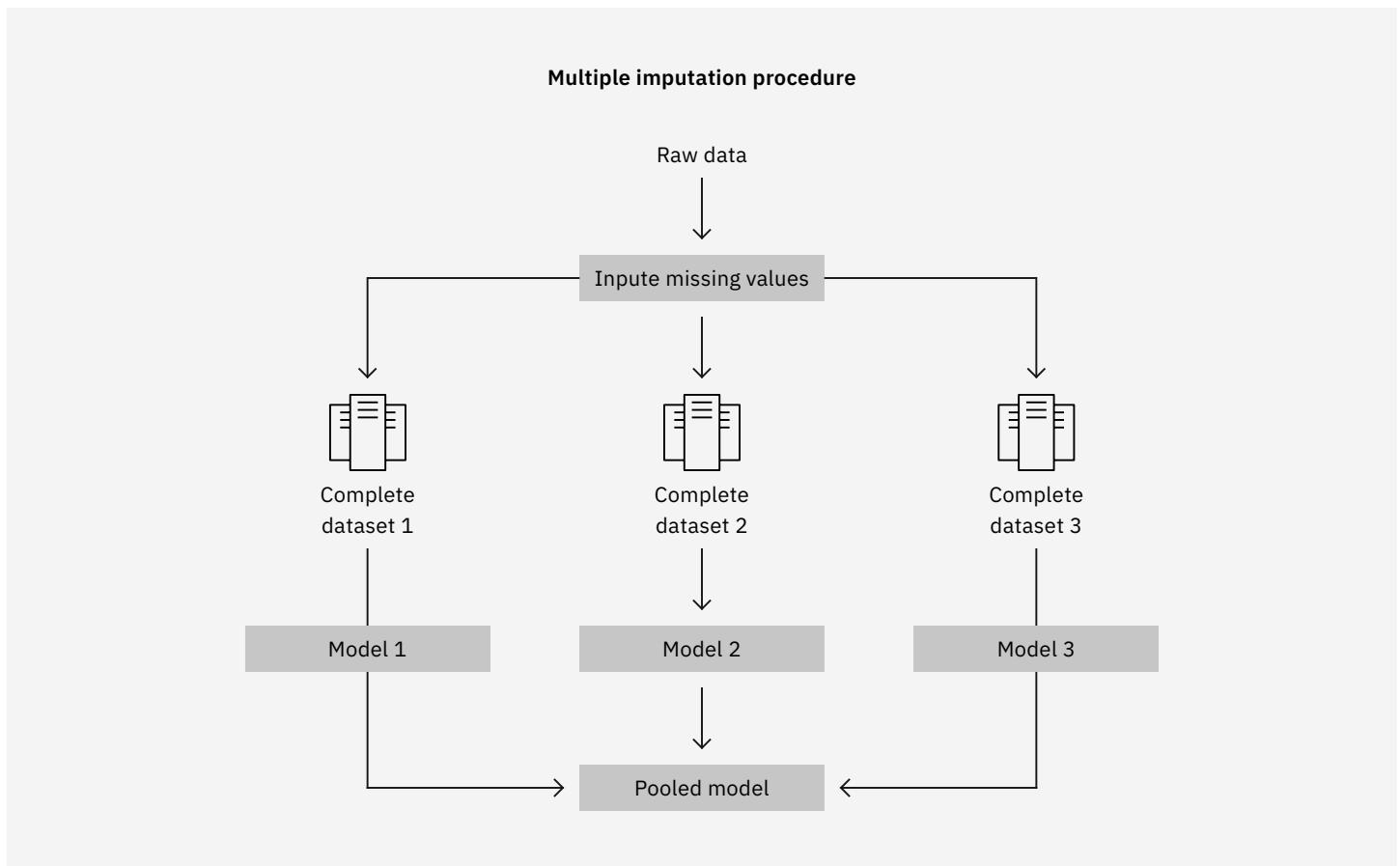
Receive a summary of missing data patterns and highlights of the variable sets that comprise the patterns with the tabulated pattern reports.

## Use multiple imputation to replace missing data values

In SPSS 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.

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

SPSS Missing Values is part of the IBM SPSS Statistics suite of software and is available within the IBM SPSS Statistics [Professional](#) and IBM SPSS Statistics [Premium](#) editions. 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.

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 SPSS Statistics software with [IBM SPSS Collaboration and Deployment Services](#).

## 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:**

- Assess 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)

### 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 such as 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 SPSS Statistics data file containing imputed values, FCS iteration history, or both.

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.

## Why IBM?

IBM SPSS Statistics is a comprehensive software solution that addresses all facets of the analytical process from data preparation and management to analysis and reporting. You can improve forecasts and plans by inputting missing values with expected values, deliver tables and visualizations to communicate results effectively, classify cases into groups and predict values of target variables based on values of predictor variables, and accurately model linear and nonlinear relationships. Organizations of all types have relied on IBM SPSS Statistics to increase revenue, outmaneuver competitors, conduct research, and drive decision-making with data.

## Additional information

To learn more about IBM SPSS Statistics, start a free trial, or make a purchase, visit the [product page](#).

### Next steps

[Discover IBM SPSS Statistics 28.](#)  
[Start a no-cost 30-day trial.](#)

If you'd like to purchase, or upgrade to SPSS Statistics 28 through an IBM Business Partner, please contact a partner from the following list: [IBM SPSS Business Partners](#).



© Copyright IBM Corporation 2021

IBM Corporation  
New Orchard Road  
Armonk, NY 10504

Produced in the United States of America  
September 2021

IBM, and the IBM logo are trademarks or registered trademarks of International Business Machines Corporation, in the United States and/or other countries. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on [ibm.com/trademark](http://ibm.com/trademark).

This document is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates.

It is the user's responsibility to evaluate and verify the operation of any other products or programs with IBM products and programs. THE INFORMATION IN THIS DOCUMENT IS PROVIDED "AS IS" WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT. IBM products are warranted according to the terms and conditions of the agreements under which they are provided.