

> Build Better Models When You Fill in the Blanks

When you ignore or exclude missing data, you risk finding invalid and insignificant results. Use SPSS Missing Value Analysis to impute your missing data and draw more valid conclusions. SPSS Missing Value Analysis is a critical tool for anyone concerned about data validity. You can easily examine your data from several angles using six diagnostic reports to uncover missing data patterns. Then, estimate summary statistics and impute missing values through statistical algorithms.

Quickly and easily diagnose your missing data

You can quickly diagnose a serious missing data problem using the data patterns report, which provides a case-by-case overview of your data. The report helps you determine the extent of missing 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 nonrespondents. 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 pattern with the tabulated pattern reports. For example, in a consumer survey, quickly notice that 98 out of 100 people didn't respond to the question set on preferred features and price.

Improve survey questions that you've identified as possibly troublesome or 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. For example, respondents who didn't answer a question on income might also skip

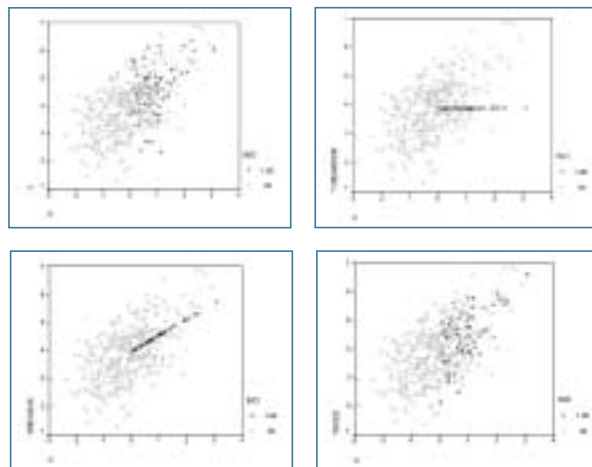
a question about education level. Use this information to enhance the quality of your surveys in the future by improving these questions.

Receive better summary statistics

Since summary statistics are often the starting point for other analyses, SPSS Missing Values Analysis allows you to adjust for missing data when working with them. Choose from four methods: Listwise deletion, pairwise deletion, expectation maximization (EM), and covariance matrix.

Improve the likelihood of finding statistically significant results

Use all of your data instead of limiting analysis to complete cases. 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 so all groups are represented in your analysis—even those with poor responsiveness. Choose the powerful EM or regression algorithm to predict missing values based on data you already have.



Missing data (top left) can leave you with invalid or erroneous results. Mean substitution (top right) and a fairly simple regression (bottom left) show that these methods provide an inaccurate or insignificant way to impute missing values. SPSS Missing Values Analysis (bottom right) provides the best method for imputing missing values. As shown here, it provides a scatterplot of YMISS and Y with imputed 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, p-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, count, and 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
- EM algorithm
 - Estimate the means, covariance matrix, and correlation matrix of quantitative variables with missing values, assuming normal distribution, t-distribution with degrees of freedom, or a mixed-normal distribution with any mixture proportion and any standard deviation ratio
 - Impute missing data and save the completed data as a file

- Regression algorithm
 - Estimate the means, covariance matrix, and correlation matrix of variables set as dependent; set number of predictor variables; and set random elements as normal, t, residuals, or none
 - Impute missing data and save the completed data as a file

Data management

- Handle all character variables as categorical variables
- Use the first eight characters of a string variable when defined as categorical
- Save the completed data matrix as an external file

System requirements

- Software: SPSS Base 13.0
- Minimum free drive space: 1MB
- Other system requirements vary according to platform

Features subject to change based on final product release.

	N	Mean	Std. Deviation	Missing		No. of Extremes ^a	
				Count	Percent	Low	High
Population	108	47723.88	146726.364	0	.0	0	11
Population density	108	202.415	675.7052	0	.0	0	13
% people in cities	108	56.53	24.203	1	.9	0	0
Life expectancy_MALE	108	70.18	10.572	0	.0	9	0
Life expectancy_FEMALE	108	64.92	9.273	0	.0	6	0
Literacy	107	78.34	22.893	2	1.9	0	0
Infant mortality	108	42.313	38.8792	0	.0	0	1
GDP per capita	108	5656.88	6479.936	0	.0	0	13
Daily calory intake (avg)	79	2753.83	587.826	34	31.2	0	0
Literacy_MALE	88	78.73	20.445	24	22.0	0	0
Literacy_FEMALE	88	67.28	28.807	24	22.0	0	0
Region	108			0	.0		
Region1	108			1	.9		

^a. Number of cases outside the range (Q1 - 1.5*QR, Q3 + 1.5*QR)

Use univariate statistics to compute mean and standard deviation, and display a separate tally of missing values and extreme values.

To learn more, please call Technologies4Targeting Ltd +44 (0)1733 890790 or visit www.tech4t.co.uk/spss

SPSS is a registered trademark and the other SPSS products named are trademarks of SPSS Inc. All other names are trademarks of their respective owners. © Copyright 2004 SPSS Inc. SMV13SPC-0804

SPSS