Print

Practice: Exploring Missing Data

The file **Explore Missing.jmp** contains data for seven variables: a Pass/Fail response variable and six process variables.

1. Summarize these data using **Columns Viewer** from the Cols menu. How many variables are missing values? Which variable is missing the most values?

   Four variables are missing values: **X3**, **X4**, **X5**, and **X6**. **X5** is missing 282 values.

2. Use **Missing Data Pattern** from the Tables menu to summarize the missing values. How many complete records do you have? How many rows are missing values for only the last two variables?

   467 rows are not missing any values. (These are complete records.) 175 rows are missing values for only the last two variables. (**Hint**: Look at the Pattern's column and find the pattern that has zeros in every position but the last two.)

3. In the Missing Data Pattern table, run the script for the **treemap**. What percent of records are complete? What percent of the rows are missing values for only the last column?

   Place your mouse pointer on the blocks in the treemap to see the values. Approximately 53% of the rows are complete. Approximately 7.6% are missing values for only the last column.

4. Use **Analyze**, **Distribution** to produce a graph and summary statistics for the following variables: **Response** and **X6**. What do you observe?

   X6 has only one category, *Yes*. There are 252 missing values for **X6**. There are 630 rows in which **Response** is *Pass* and **X6** is *Yes*. There are no values for **X6** when **Response** is *Fail*.

5. Which of the following best describes the missing values for **X6**: missing completely at random, missing at random, or missing not at random? Why?

   The data are missing not at random. Every value of **X6**, when **Response** is *Fail*, is missing.

Hide Solution

Statistical Thinking for Industrial Problem Solving

Close