Print

Data Tables Essentials

For most statistical analyses, your data must have a basic structure, consisting of rows and columns.

The observations are stored in rows, and the variables are in columns.

An observation is the information about the basic unit of interest. The observational unit, for example, might be a customer, a batch, a part, or a transaction.

A variable is a characteristic that can be measured or recorded for an observation.

For the Impurity data, the observation is the Batch. The variables are the batch number, the day, the percent impurity, and so on.

You have seen that when you are analyzing data, you generally assign your data one of three modeling types: Nominal, Ordinal, and Continuous.

In our example, Day, Outcome, Reactor, and Shift are assigned a Nominal modeling type.

The percent Impurity, Temp, Catalyst Concentration, and Reaction Time have a Continuous modeling type.

Batch, or the batch number, is not used in analysis. Therefore, it is not assigned a modeling type. That is, the modeling type is None.

Throughout this module, you've used data stored in JMP data tables for data summary, visualization, and exploratory data analysis.

All of the examples you've seen are "clean." They have been prepared and structured for analysis.

For example, consider the Impurity data:

Appropriate modeling types have been applied to the variables, there aren't any missing values, and there aren't any typographical errors or strange values.

You can confidently use these data for analysis.

However, real-life data are rarely this clean.

In any project, a majority of your time might be spent collecting, compiling, and preparing your data.

In this lesson, you learn about common data quality issues and steps that you'll need to take to diagnose these issues and prepare your data for analysis.

For a more thorough discussion of data preparation for analysis and modeling, see the Read About It for this module.

Statistical Thinking for Industrial Problem Solving

Close