Data Manipulations II

Conditional Merge

Understanding the IN= Data Step

When you create a new data set by combining observations from two or more data sets, knowing which data set an observation came from can be useful.

For example, you might want to perform a calculation based on which data set contributed an observation. Otherwise, you might lose important contextual information that you need for later processing.

You can determine which data set contributed a particular observation by using the IN= data set option.

The syntax for this option on the SET statement is:

DATA combined;
SET data1 (IN=variable1) data2 (IN=variable2);
BY ID;

When you use the IN= option with a data set in a SET, MERGE statement, SAS creates a temporary variable associated with that data set. The value of variable is 1 if the data set has contributed to the observation currently in the program data vector. The value is 0 if it has not contributed. You can use the IN= option with any or all the data sets you name in a SET, MERGE statement, but use a different variable name in each case.

Note: The IN= variable exists during the execution of the DATA step only; it is not written to the output data set that is created.