Excluding Unmatched Observations

By default, DATA step match-merging combines all observations in all input data sets. However, you may want to select only observations that match for two or more specific input data sets.

To exclude unmatched observations from your output data set, you can use the IN= data set option and the subsetting IF statement in your DATA step. In this case, you use

- the IN= data set option to create and name a variable that indicates whether the data set contributed data to the current observation
- the subsetting IF statement to check the IN= values and write to the merged data set only matching observations.

Creating Temporary IN= Variables

General form, IN= data set option:

(IN= variable)

where

- the IN= option, in parentheses, follows the data set name
- variable names the variable to be created.

Within the DATA step, the value of the variable is 1 if the data set contributed data to the current observation. Otherwise, its value is 0.

Example:

```
merge A (in = inA) B (in = inB);
```

Selecting Matching Observations

Next, to select only observations that appear in both data A and B, you specify a subsetting IF statement in the DATA step.

SAS evaluates the expression within an IF statement to produce a result that is either nonzero, zero, or missing. A nonzero and nonmissing result causes the expression to be true; a zero or missing result causes the expression to be false.

Thus, you can specify the subsetting IF statement in either of the following ways. The first IF statement checks specifically for a value of 1. The second IF statement checks for a value that is neither missing nor 0 (which for IN= variables is always 1).

```
if inA=1 and inB=1;
```

```
if inA and inB;
```