Combining SAS Data Sets

In SAS programming, a common task is to combine observations from two or more data sets into a new data set. Using the DATA step, you can combine data sets in several ways, first lets talk about one-to-one merging.

We learned how to use the SET statement to read an existing SAS data set. You can also use multiple SET statements in a DATA step to combine data sets. This is called one-to-one merging (or combining). In one-to-one merging you can read different data sets, or you can read the same data set more than once, as if you were reading from separate data sets.

General form, basic DATA step for one-to-one merging:

DATA output-SAS-data-set;
    SET SAS-data-set-1;
    SET SAS-data-set-2;
RUN;

where

- output-SAS-data-set names the data set to be created
- SAS-data-set-1 and SAS-data-set-2 specify the data sets to be read.

How One-to-One Merging Selects Data

When you perform one-to-one merging,

the new data set contains all the variables from all the input data sets. If the data sets contain variables that have the same names, the values that are read from the last data set overwrite the values that were read from earlier data sets.

the number of observations in the new data set is the number of observations in the smallest original data set. Observations are combined based on their relative position in each data set. That is, the first observation in one data set is joined with the first observation in the other, and so on. The DATA step stops after it has read the last observation from the smallest data set.

```
 data one2one;
   set a;
   set b;
run;
```


SAS Data Set A

Num	VarA
1	A1
3	A2
5	A3

SAS Data Set B

Num	VarB
2	B1
4	B2

Merge

Combined SAS Data Set

Num	VarA	VarB
2	A1	B1
4	A2	B2