Rearranging Your Data Using PROC SORT

The order of the data observations usually doesn't matter for many of the statistical analyses researchers perform. Often, however, you may want the data arranged in some specific order. For example, you may wish to print the data set example1 in alphabetical order by name. This will aid in finding records for specific individuals. Or you may wish to perform a "matched merge" on two data sets. In this situation, both sets of data need to be arranged in the same sorted order. Or you may wish to analyze your data in subsets.

To sort the data use a PROC SORT statement followed by a BY statement that specifies the variable(s) by which you wish to sort the data. For example, to sort the data set example1 in alphabetical order using the variable name:

```
PROC SORT DATA= example1;
   BY name;
```

You may specify more than one variable in the BY statement. The first variable in the BY statement is the primary sort key, the second variable in the BY statement is the secondary sort key, and so on. The data set is sorted according to the primary sort key. For records with the same value for the primary sort key, these records will be sorted using the secondary sort key. An example:

```
PROC SORT;
   BY lastname  firstnam ;
```

The sorted data set will be arranged similar to the list of names in the telephone directory. Those records with the same lastname will be sorted alphabetically by firstnam.

SAS normally reads the data set identified by the DATA= option in the PROC SORT statement (or the most recently created data set if the DATA= option is omitted from the PROC SORT statement). SAS rearranges the data set in the order of the variable(s) specified in the BY statement and stores the rearranged data set with the same name as the original (unsorted) data set. The unsorted version of the data "disappears".

1
You may wish to keep both the sorted and unsorted versions of the data set. This can be accomplished by using the OUT= option in the PROC SORT statement. For example,

```
PROC SORT DATA=example1 OUT= ex1sort ;
   BY sex;
```

The data set `ex1sort` contains the same observations as the data set `example1`, but the observations are sorted by the values of the variable `sex`.

If the sort key is a character variable, SAS will sort the data in alphabetic order (ascending order). If you wish to sort the data in reverse alphabetic order, you must specify the option DESCENDING before the name of the character variable in the BY statement. For example,

```
PROC SORT DATA= example1;
   BY  DESCENDING  name;
```

will rearrange the observations such that those observations whose value for name begins with a "Z" will be placed first in the sorted data set, and those observations whose value for name begins with an "A" will be placed last in the sorted data set. The option DESCENDING applies only to the variable which immediately follows.

Similarly, data sorted using a numeric variable are arranged in ascending order (low values to high values). The DESCENDING option may also be used for a numeric sort key:

```
PROC SORT DATA= example1;
   BY DESCENDING height;
```

It should be noted that PROC SORT does not produce any printed output, although it does print a message in the SAS log file indicating the amount of storage and time was used to sort the data, and how many observations the sorted data set contains. In order to print the sorted data set you must use the PRINT procedure.