Sorting Your Data with PROC SORT

Often we need to sort our data, we can use Proc Sort to do that. The basic form of this procedure is

```
PROC SORT;
    BY variable-list;
```

- The variables named in the BY statement are called BY variables.
- You can specify as many BY variables as you wish.
  - With one BY variable, SAS sorts the data based on the values of that variable.
  - With more than one variable, SAS sorts observations by the first variable, then by the second variable within categories of the first variable, and so on.
- A BY group is all the observations that have the same values of BY variables.

Controlling the Input and Output Data Set

```
PROC SORT DATA = input-data OUT = output-data;
```

- The DATA= and OUT= options specify the input and output data sets.
- If you don't specify the DATA= option, then SAS will use the most recently created data set.
- If you don't specify the OUT= option, then SAS will replace the original data set with the newly sorted version.

This sample statement tells SAS to sort the data set named MESSY, and then put the sorted data into a data set named NEAT:

```
PROC SORT DATA = messy OUT = neat;
```

The NODUPKEY option tells SAS to eliminate any duplicate observations that have the same values for the BY variables. If you specify the DUPOUT= option, then SAS will put the deleted observations in that data set.

To use these options, just add them to the PROC SORT statement:

```
PROC SORT DATA = messy OUT = neat  NODUPKEY  DUPOUT = extraobs;
```

Ascending versus Descending Sorts

By default SAS sorts data in ascending order, from lowest to highest values. To have your data sorted in the opposite order, add the keyword DESCENDING to the BY statement before each variable that should be sorted in reverse order.

This statement tells SAS to sort first by State (from A to Z – default ascending order) and then by City (from Z to A – descending order) within State:

```
BY State DESCENDING City;
```