# Simplifying Programs with Arrays

Sometimes you want to do the same thing to many variables. You could write a series of assignment statements or IF statements, but if you have a lot of variables to transform, using arrays will simplify and shorten your program.

In SAS, an array is a group of variables. You can define an array to be any group of variables you like, as long as they are either all numeric or all character. The variables can be ones that already exist in your data set, or they can be new variables that you want to create.

Arrays are defined using the ARRAY statement in the DATA step. The ARRAY statement has the following general form:

ARRAY *name* (*n*) $ *variable-list*;

- *name* is a name you give to the array
- *n* is the number of variables in the array
- Following the (*n*) is a list of variable names.
- The number of variables in the list must equal the number given in parentheses. This is called an explicit array, where you explicitly state the number of variables in the array.
- The $ is needed if the variables are character, and is only necessary if the variables have not previously been defined.

The array itself is not stored with the data set; it is defined only for the duration of the DATA step. You can give the array any name, as long as it does not match any of the variable names in your data set or any SAS keywords.

The rules for naming arrays are the same as those for naming variables (must be 32 characters or fewer and start with a letter or underscore followed by letters, numerals, or underscores).

**To reference a variable using the array name**, give the array name and the subscript for that variable. The first variable in the variable list has subscript 1, the second has subscript 2, and so forth. So if you have an array defined as

```
ARRAY SCORE (4) math science language social_study;
```

SCORE(1) is the variable math, SCORE(2) is the variable science, SCORE(3) is the variable language, and SCORE(4) is the variable social_study.