Recode made easy

Learn how to change existing variables or to create new variables based on existing data

by John Urbik, Technical Marketing Specialist

Every data set you work with should have all the variables you need. Unfortunately, this is seldom the case, and most projects require you to change existing variables or to create new ones based on existing data. The Recode transformation allows you to do both of these tasks.

If you are using SPSS for Windows, SPSS for the Macintosh or SPSS for UNIX, Recode is found in easy-to-use dialog boxes, under the Data and the Transform menus. If you are using SPSS/PC+, Recode is found under the Modify Data Values menu.

There are two types of Recode: into the same variable and into a different or new variable. Recoding is the ability to collapse or combine values for a specific variable. This can be done to the existing variable or by creating a new variable. Use Recode when you have:

- too many categories
- too few cases in a response group
- to combine similar categories
- to create categories from continuous data

The first choice under Recode is Into Same Variable. This is used to collapse response codes into better groups. Below is the frequency for the JOBCAT variable from the SPSS Bank data file.

In this example, you see there are few people in the College trainee, Exempt employee, MBA trainee and Technical employment categories. You might want to Recode the JOBCAT variable to consolidate these categories, producing the table shown below. Unlike other SPSS procedures, data management tools do not produce any specific output, so we use Frequencies to check our work.

This new group, "College trainee and above," is more useful because it represents a higher category of people with specific managerial and technical skills. It is important to note, after you have recoded this variable that you lose the information from the old categories if you save the data file.

The second Recode is Into a Different Variable. This creates a new variable with values determined by an existing variable. For instance, if we look at the Bank data file (Bank.sav) again, there is a variable called EDLEVEL. This variable contains information about the number of years of school completed.

While having information like this is nice, it is easier to report in more common categories. Using Recode Into Different Variable, we create a new variable called EDCAT, meaning Education Recoded to Categories. To do this, combine the old educational levels "High School or less" (Lowest-12), "Some College" (13-15), "Highest or College or beyond" (16). This produces a new variable seen below.

The hardest decision when using Recode is deciding which choice to use. When making this decision, it is important to remember that once you have recoded into an existing variable, you lose the previous values if you save the data file. This is fine if you are working with a data file where a variable has small categories and you increase category size making the recoded groups better represented.

Recoding into a different variable is best used when taking an item measured on a continuous scale and creating a categorical scale. This is convenient when you want to present data in a table as well as use the information in other types of analysis that require continuous measurement.

These are just two examples of using Recode. For more details, see the Data Transformations chapter of the SPSS Base 7.0 System Users Guide.