

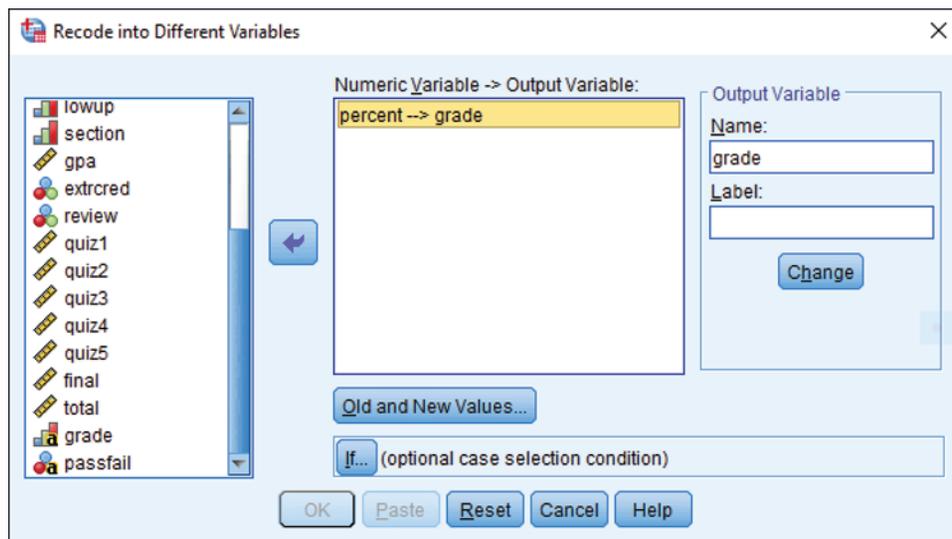
4.5 Recoding Variables

4.5.1 Creating New Variables

This procedure also creates new variables, not by means of calculations with already existing variables but rather by dividing a preexisting variable into categories and coding each category differently. This procedure is ideally suited for a data file that includes class records because at the end of the term the teacher wants to divide scores into different grades. In the present example this coding is based on the class **percent** variable in which percents between 90 and 100 = A, between 80 and 89 = B, between 70 and 79 = C, between 60 and 69 = D, and between 0 and 59 = F. The **percent** variable can also be used to classify subjects into pass/fail categories with codings of greater than or equal to 60 = P and less than 60 = F.

Two different windows control the recoding-into-different-variables process. Screen 4.4 (below) allows you to identify the **Numeric Variable** (**percent** in the present example) and the **Output Variable** (**grade** in this case). A click on **Old and New Values** opens up the second window, Screen 4.5 (following page). This box allows you to identify the ranges of the old variable (90 to 100, 80 to 89, etc.) that code into levels of the new variable (A, B, C, D, and F).

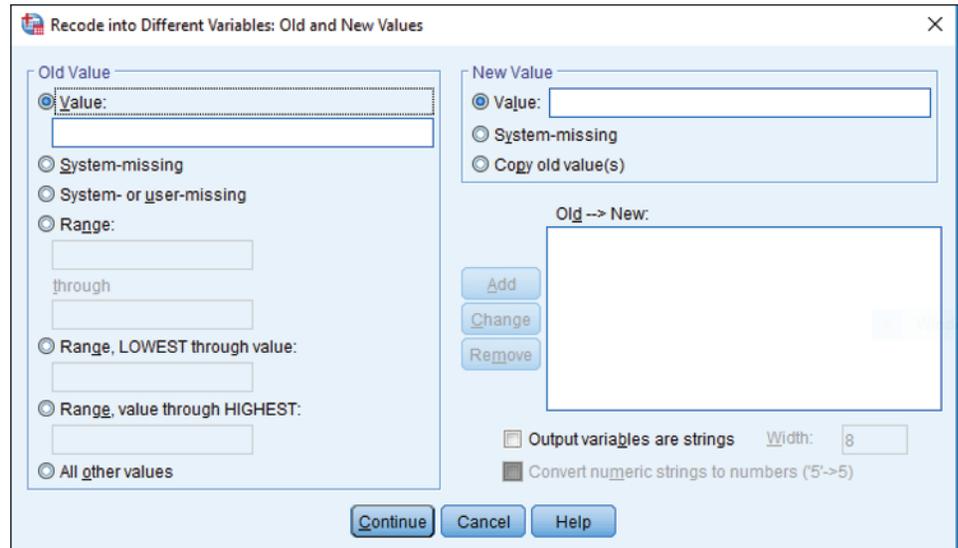
Screen 4.4 The Recode into Different Variables Window



Notice that in Screen 4.4 all numeric variables are listed to the left and the five function buttons are at the bottom of the dialog box. The initial stage (in this example) is to click on **percent** and paste it into the **Numeric Variable** → **Output Variable** box. Then click in the box beneath the **Output Variable** title and type the name of the new variable, **grade**. A click on the **Change** button will move the **grade** variable back to the previous box. The entry in the active box will now read **percent** → **grade**. A click on the **Old and New Values** pushbutton opens the next window.

The dialog box represented by Screen 4.5 allows you to identify the values or range of values in the input variable (**percent**) that is used to code for levels of the Output variable (**grade**). Since we want a *range* of values to represent a certain grade, click on the top **Range** option and indicate the range of values (90–100) associated with the first level (A) of the new variable (**grade**). When the first range is entered, click in the **Value** box and type the letter A. A click on the **Add** button will paste that range and grade into the **Old** → **New** box. We are now ready to perform a similar sequence for each of the other four grades. When all five levels of the new variable are entered (with the associated ranges and letter grades), click **Continue** and Screen 4.4 will reappear. A click on the **OK** button will create the new variable.

Screen 4.5 The Old and New Values Window



A similar process will be conducted for creating the **passfail** variable. When these two new variables are created they will be listed in the last two columns of the data file. Once the new variable is created, you may type in value or variable labels, change the format, or move the variables to a new location in the data file by techniques described in Chapter 3.

The creation of the two new variables is described below in the Step by Step boxes. Remember that these variables are already in the data available by download. If you want to practice with that file, create variables with slightly different names (e.g., **grade2**) and then you can compare your results with those already present.

*To create the new variable **grade** the starting point is any menu screen. Perform the following sequence of steps to create this new variable.*

In Screen	Do This	Step 5d
Menu	Transform → Recode into Different Variables	
4.4	percent → → press TAB → type grade → Change →	
	Old and New Values	
4.5	Output Variables are strings → <i>change 8 to 4 in the Width box</i> →	
	<i>circle to the left of Range (the top one, not the lower Range or Range)</i>	
<i>(Note: before each of the 5 lines that follow, first click in the top Range box)</i>		
	type 90 → press TAB → type 100 → <i>in Value box</i> → type A → Add	
	type 80 → press TAB → type 89.9 → <i>in Value box</i> → type B → Add	
	type 70 → press TAB → type 79.9 → <i>in Value box</i> → type C → Add	
	type 60 → press TAB → type 69.9 → <i>in Value box</i> → type D → Add	
	type 00 → press TAB → type 59.9 → <i>in Value box</i> → type F → Add	
	→ Continue	
4.4	OK	Data

To create the new variable **passfail** the starting point is any menu screen. Perform the following sequence of steps to create this new variable.

In Screen	Do This	Step 5d'
Menu	Transform → Recode into Different Variables	
4.4	percent → → press TAB → type <code>passfail</code> → Change →	
	Old and New Values	
4.5	Output Variables are strings → <i>change 8 to 4 in the Width box</i> →	
	<i>circle to the left of Range (the top one, not the lower Range or Range)</i>	
<i>(Note: before each of the 2 lines that follow, first click in the top Range box)</i>		
	type <code>60</code> → press TAB → type <code>100</code> → <i>in Value box</i> → type <code>P</code> → Add	
	type <code>00</code> → press TAB → type <code>59.9</code> → <i>in Value box</i> → type <code>F</code> → Add	
	→ Continue	
4.4	OK	Data

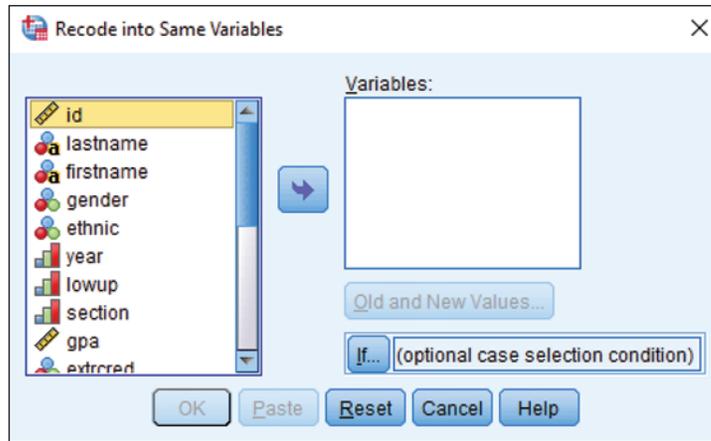
4.5.2 Changing the Coding of Variables

There are times when you may wish to change the coding of your variables. There are usually two reasons why it is desirable to do so. The first is when you wish to reverse coding to be consistent with other data files, or when one coding pattern makes more sense than another. For instance if you had a file in which the gender variable was coded male = 1 and female = 2, you may have other files where it is coded female = 1, male = 2. It would be desirable to recode the files so that they are consistent with each other. Or perhaps you have an **income** variable that has >\$100,000 coded 1 and <\$10,000 coded 7. If you felt that it would make more sense to have the lower incomes coded with the lower numbers, the **Recode** command could accomplish that easily for you.

The second reason for recoding is to group variables that are ungrouped in the original data. For instance, you might have a variable named **marital** that is coded 1 = never married, 2 = divorced, 3 = widowed, 4 = separated, 5 = married. You could use the **Recode** procedure to change the marital variable so that 1 = single (a combination of 1, 2, and 3) and 2 = married (a combination of 4 and 5).

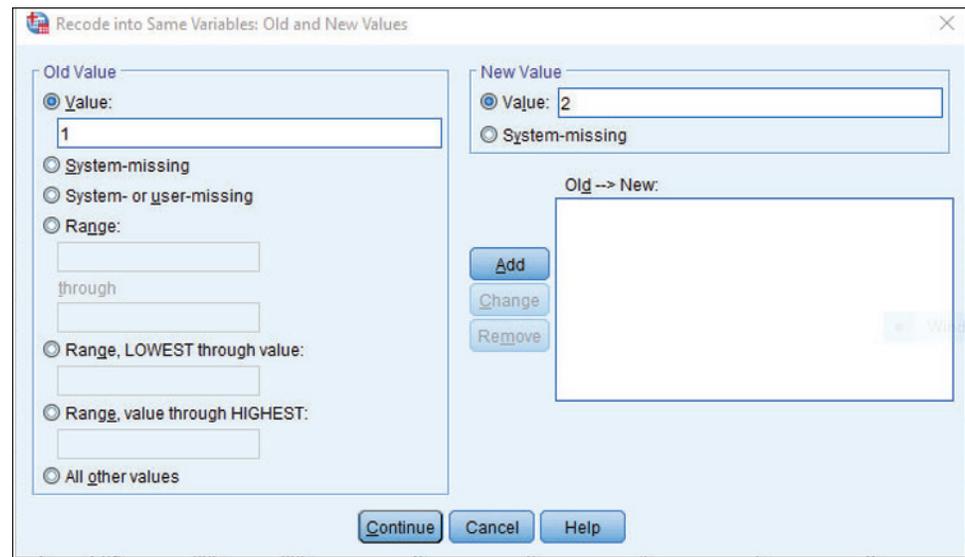
The starting point for recoding variables is a screen that shows the menu of commands across the top. Select the following options to access the Recode screen.

In Screen	Do This	Step 4e
Menu	Transform → Recode into Same Variables	4.6

Screen 4.6 The Recode into Same Variables Window

Two different dialog boxes control (1) the selection of variable(s) to be recoded (Screen 4.6, above) and (2) clarification of old and new coding values (Screen 4.7, below). It is also possible to recode a string variable (such as **grade**) into numeric coding. A researcher might wish to recode the letters A, B, C, D, and F into 4, 3, 2, 1, and 0, so as to run correlations of **grade** with other variables. The **Recode** option can accommodate this comfortably. You can recode several variables in one setting but would need to switch back and forth between the two screens to do so.

In Screen 4.6 (above) the first step is to select the variable(s) of interest and paste them into the active box to the right (titled **Numeric Variables**). A click on **Old and New Values** opens up a new window (Screen 4.7).

Screen 4.7 The Old and New Values Window

Options for **Old Value** include:

- A single value
- System-missing values (blanks in your data set)
- User-missing values (values you have designated as missing for various reasons)
- A range of values (internal range, lowest to specified value, specified value to highest)
- All other values (useful for complex combinations)

For each variable you specify the old value(s) and the new value(s), and then paste them into the **Old** → **New** box one level at a time. To demonstrate both a single switch recode

and a recoding that involves a range of values, we will reverse the coding for the **gender** variable. Then, even though it's not in the **grades.sav** file, we recode the fictitious **marital** variable into levels 1, 2, and 3 representing unmarried and levels 4 and 5 representing married. If you wish to practice the latter procedure using the **grades.sav** file, simply use the **ethnicity** variable for recoding. Like the fictional *marital* variable, it also has five levels.

The starting point is Screen 4.6. Perform the following sequence of steps to reverse the coding on the **gender** variable.

In Screen	Do This	Step 5e
4.6	gender → → Old and New Values	
4.7	in Value box → type 1 → press TAB twice → type 2 → Add	
	in Value box → type 2 → press TAB twice → type 1 → Add →	
	Continue	
4.6	OK	Back1

Perform the following sequence of steps to recode the mythical **marital** variable.

In Screen	Do This	Step 5e'
4.6	marital → → Old and New Values	
4.7	first Range option → in Range box → type 1 → press TAB → type 3	
	→ press TAB twice → type 1 → Add →	
	in same Range box → type 4 → press TAB → type 5 →	
	press TAB twice → type 2 → Add → Continue	
4.6	OK	Back1

Create new value labels! To create new value labels for the recoded variables is an urgent concern. Otherwise Robert will be coded “female,” Will Smith will be coded “Asian,” and Elizabeth Taylor will be coded “never married.” If you wish to create new value labels for the recoded variables, you will need to perform the appropriate operation (click on the **Variable View** tab and the appropriate cell in the **Values** column) described in Chapter 3. Further, if you want these changes to be permanent in your data file, you will need to go through the save data procedure (click **File** → click **Save**).

4.6 The Select Cases Option

SELECTING A PORTION OF THE DATA FOR ANALYSIS To access the necessary initial screen, select the following options:

In Screen	Do This	Step 4f
Menu	Data → Select Cases	4.8