Chapter 5
Housekeeping:
Transforming Variables and Adding Labels

5.1 The EZDATA File

In this chapter we provide the EZDATA file for you to download. We also describe it's structure and organization, and we explain modifications of the original file that have been made to facilitate it's use in analyzing the results of the leadership project.

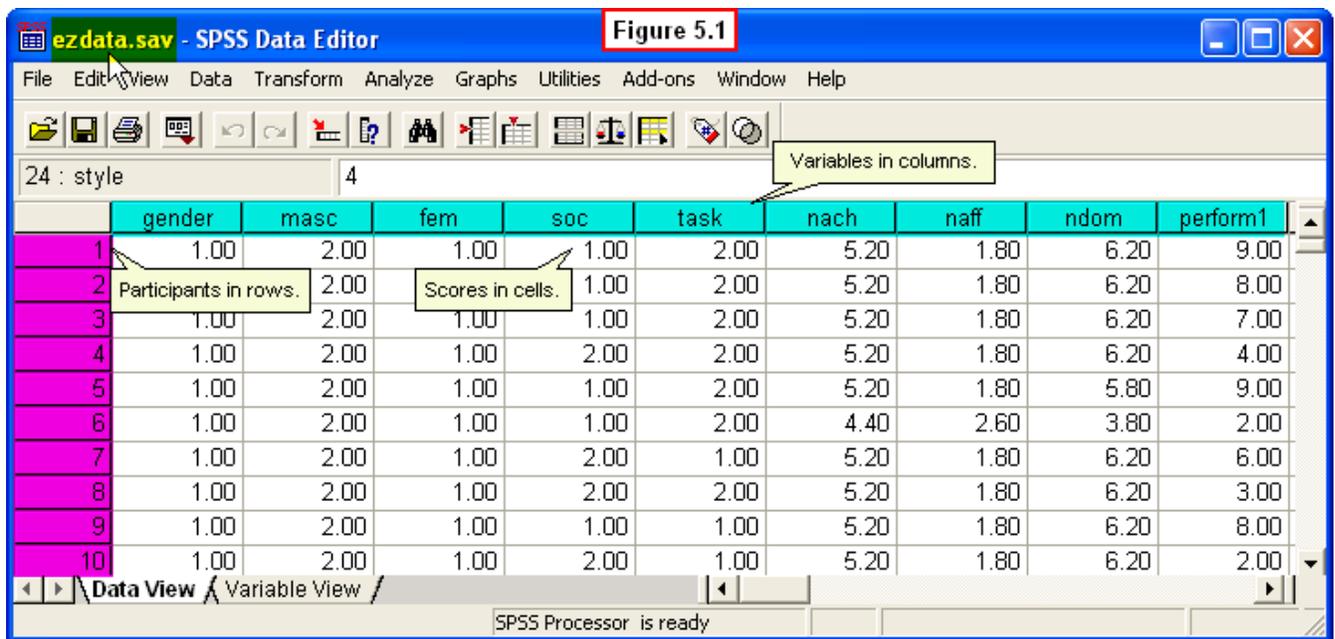### 5.1a Downloading the EZDATA file

Click the **EZDATA button** below to download the EZDATA file we have been describing and will be using for the remainder of this text. Downloading and opening this file will facilitate your understanding as we describe its structure and organization. When you click the button below you will be prompted to either open the file or save it.

Select the save option, then navigate to the ezdata folder you created for storing files. Name the file **ezdata.sav** for easy recall.

Show Me Video!

5.1b Structure of the EZDATA file

After you have downloaded the file, open it in the SPSS Data Editor. When you have opened the file in SPSS, it should look like the one shown in Figure 5.1.

Figure 5.1

	gender	masc	fem	soc	task	nach	naff	ndom	perform1
1	1.00	2.00	1.00	1.00	2.00	5.20	1.80	6.20	9.00
2	2.00	2.00	1.00	1.00	2.00	5.20	1.80	6.20	8.00
3	1.00	2.00	1.00	1.00	2.00	5.20	1.80	6.20	7.00
4	1.00	2.00	1.00	2.00	2.00	5.20	1.80	6.20	4.00
5	1.00	2.00	1.00	1.00	2.00	5.20	1.80	5.80	9.00
6	1.00	2.00	1.00	1.00	2.00	4.40	2.60	3.80	2.00
7	1.00	2.00	1.00	2.00	1.00	5.20	1.80	6.20	6.00
8	1.00	2.00	1.00	2.00	2.00	5.20	1.80	6.20	3.00
9	1.00	2.00	1.00	1.00	1.00	5.20	1.80	6.20	8.00
10	1.00	2.00	1.00	2.00	1.00	5.20	1.80	6.20	2.00

As discussed in Chapter 2.3b, the data in the file are organized with all of our **variables** in the **columns**, and individual **employee participants** (cases) listed in **rows**. The **scores** each employee received on each of the variables are entered in the **cells** along each row.

For example, employee number 1 was assigned a score of 1 for the **gender** variable (to indicate he is a man); he received scores of 1 on the **masc** variable and 2 on the **fem** variable. Scroll to the right and you will see his scores on the other variables in the data file.

As you continue scrolling to the right, you may notice that the listing of the variables in the columns looks different from the listing explained in Chapter 4. You may have also noted that there are some new variables in the file that weren't discussed in Chapter 4. The reason for this is that we have already done some of the housekeeping referred to in the title of the present chapter.

That is, we have completed transformations on the original variables to create the new variables in the file, and have added variable labels for the major variables, as well as value labels for some of them. We did this so that we could provide the file in its complete form for download. This way we can be sure you will be working with exactly the same **ezdata.sav** file that we will be using in the remainder of this book. Also, this should reduce any anxiety you might have about losing or somehow ruining your ezdata file - you can come back to this page anytime and re-download the file.

In this chapter we will explain the procedures for making these changes to the file. Some of them are tedious and error-prone, which is why we have already done them. In the remaining chapters of this book, we *do* suggest that you actually follow along

examples and perform data analysis procedures on the **ezdata.sav** file that we describe.

But in the current chapter, instead of asking you to perform the file modification procedures yourself on the **ezdata.sav** file, we will suggest that you practice these procedures on the simpler **example1.sav** file you created in Chapter 2. That way you will still learn the procedures without having to perform them all on the **ezdata.sav** file.

5.1c So what are these new variables?

If you look back at Table 4.5 and also scroll through the **ezdata.sav** file in the SPSS Data Editor, you will see that the original variables shown in Table 4.5 do exist in the new ezdata file, they are just in different columns than you might have expected. Gender does appear in the first column, but the **masc1-masc5** variables (the self-ratings on the 5 masculine personality traits) do not begin in the second column. Instead, scroll right and you will see them in columns 18-22. The **fem1-fem5** scores appear in columns 23-27.

The variables named **masc** and **fem** that you see in columns 2 and 3 are actually new variables that have been created by transformations of the original **masc1-masc5** and **fem1-fem5** variables. Click the **Variable View** tab at the bottom of the Data Editor for a better view of all of the variables. Recall that this view lists variables in rows instead of columns (Figure 5.2).

Figure 5.2
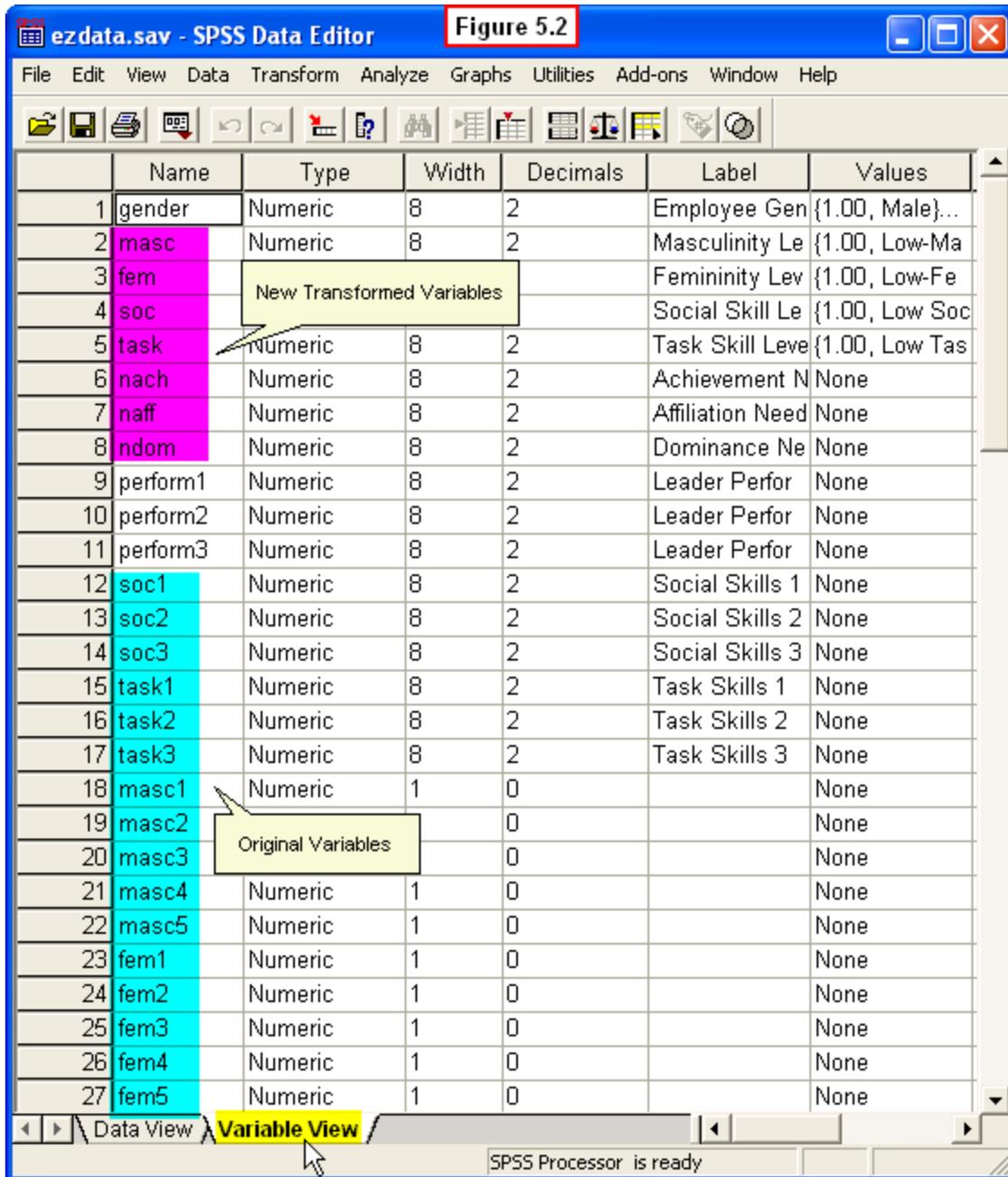
To create the new variables, **masc** and **fem**, we used the SPSS **Transform** procedure to add up the scores on the **masc1-masc5** and **fem1-fem5** variables to create two variables named **masctot** and **femtot**. If you scroll all the way down, you will see these masctot and femtot variables (SPSS tacks newly-created variables at the end of the original data file).

Next, we used the SPSS Transform procedure again to convert the continuous masctot and femtot variables into categorical variables, which we named **masc** and **fem**. Last, we moved these two new variables to rows 2 and 3 for easy access later. If these

variable transformations sound confusing, don't worry - we will explain all of this in the next section!

You will also see the original **soc1-soc3** and **task1-task3** variables (employee's scores on social and task leadership skills) in rows 12-14 and 15-17, respectively. We performed similar transformations on the **soc1** and **task1** variables to create new categorical variables that we named **soc** and **task**, and then placed them in rows 4 and 5 for easy access.

Next, you will see the original **ach1-ach5**, **aff1-aff5**, and **dom1-dom5** variables (employee work motive scores) in rows 28-42. We used the the Transform procedure on these variables to add them up and take their average to create the new variables, **nach**, **naff**, and **ndom** (short for achievement needs, affiliation needs and dominance needs), and we placed these three new variables in rows 6-8 for quick access.

Figure 5.3 graphically depicts these variable transformations. Note that the original variables (in blue) still exist in the data file, and their transformed versions have been given new names (in red), and now exist as separate new variables in the file.

Figure 5.3
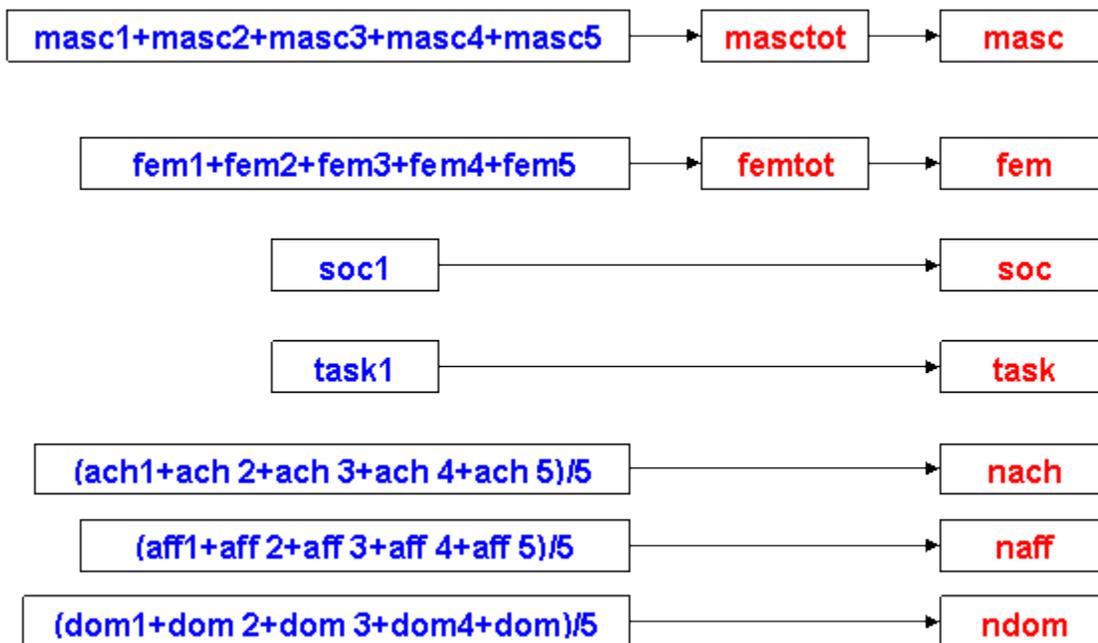
You will find the original **perform1-perform3** variables in rows 9-11. We have not done any transformations on these variables. Last, if you scroll all the way down, you will see yet two more new variables, **sextype** and **leaderstyle**. We also used the Transform

procedure to create these new categorical variables, but we will not discuss how we did this in the present chapter. We explain this procedure in a later chapter since it is a bit more complex than the others we describe in this chapter.

5.2 Creating New Variables using the Transform Procedure

In this section, we explain how the new variables just described were created. All of this was done using an important SPSS feature called **Transform**. This feature allow the researcher to modify existing raw data into more useful forms for analysis.

5.2a The Compute Procedure

Recall that the **masc1-masc5** and **fem1-fem5** variables consist of self-ratings of how descriptive employees thought individual traits (stereotyped masculine and feminine) were of themselves. While some research questions might require the use of individual items such as these, more frequently researchers combine scores on individual items in a questionnaire into one total score. Thus, we combined these individual items into a single summated rating. This is done easily enough using the **Compute** subroutine of the **Transform** procedure. To compute a new variable by summing scores on several variables, simply select **Transform, Compute** from the Data Editor menu (see Figure 5.4).
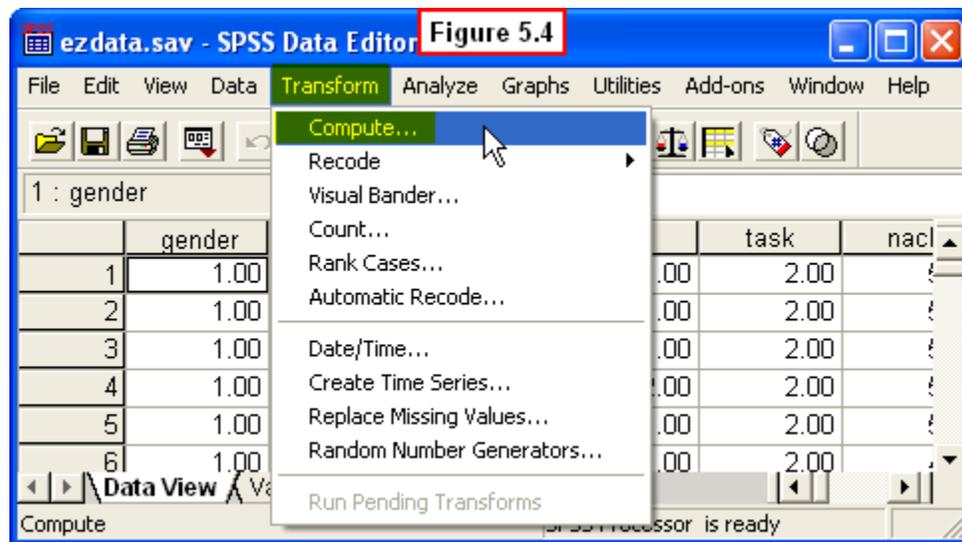
A **Compute Variable** dialog box will appear (see Figure 5.5). To name the first new variable to be created, we typed **masctot** in the **Target Variable:** box. To define the computation to be performed, we typed **masc1+masc2+masc3+masc4+masc5** in the **Numeric Expression** box. After we clicked **OK**, SPSS added the five masculine trait scores for every employee and put the total in a new column labeled **masctot** (our new variable).
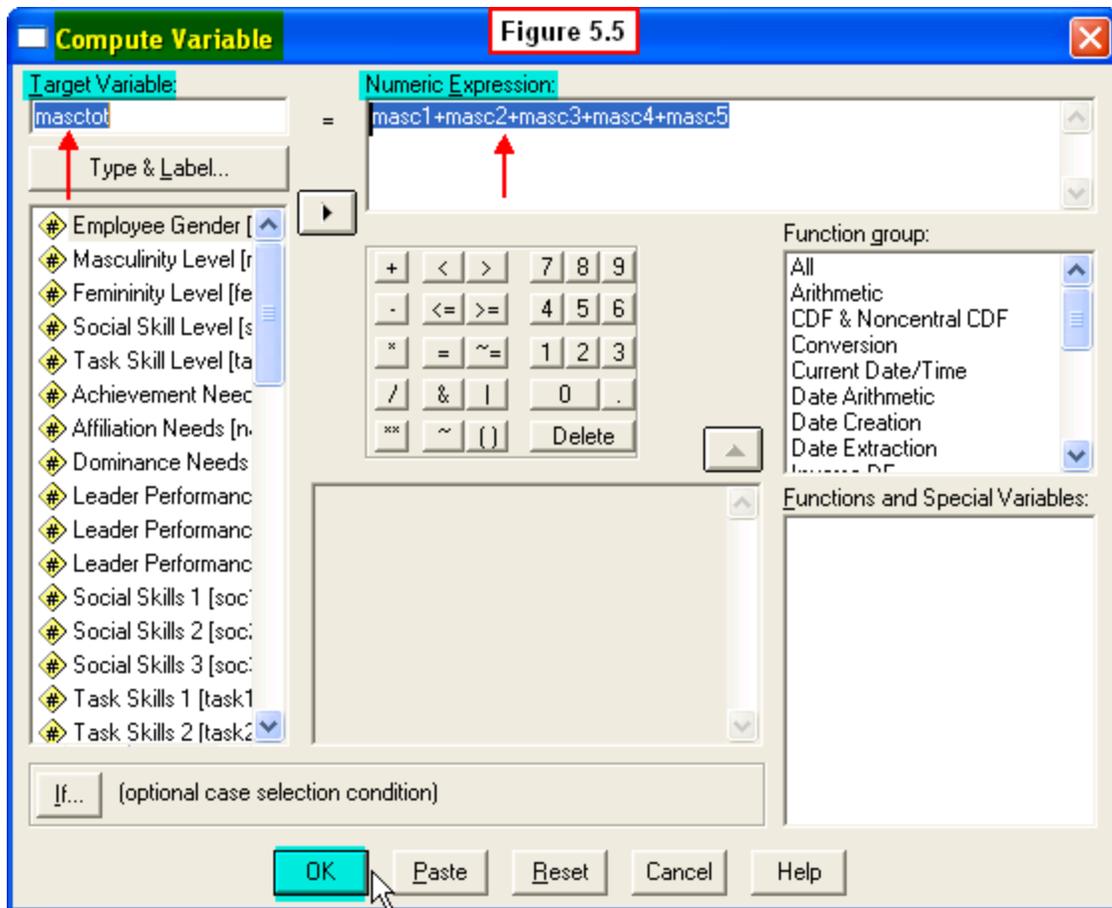
Figure 5.5

As mentioned, SPSS tacks this new variable in the last column of the data file. Note that even though we used the **masc1-masc5** variables to generate the new **masctot** variable, the original variables are still retained in our file. This is convenient should we want to analyze the individual traits at some point. But more importantly, we have also gained a new variable which will be much easier to work with.

Show Me Video!

We followed this same procedure to compute **femtot**. We also used the **Compute** procedure to create the new variables, **nach**, **naff** and **ndom**. This time, however, we typed in the formula for computing the average of the five individual scores on the original **ach1-ach5**, **aff1-aff5** and **dom1-dom5** work motive variables. See Figure 5.6 for an example of this procedure used to compute **nach**.
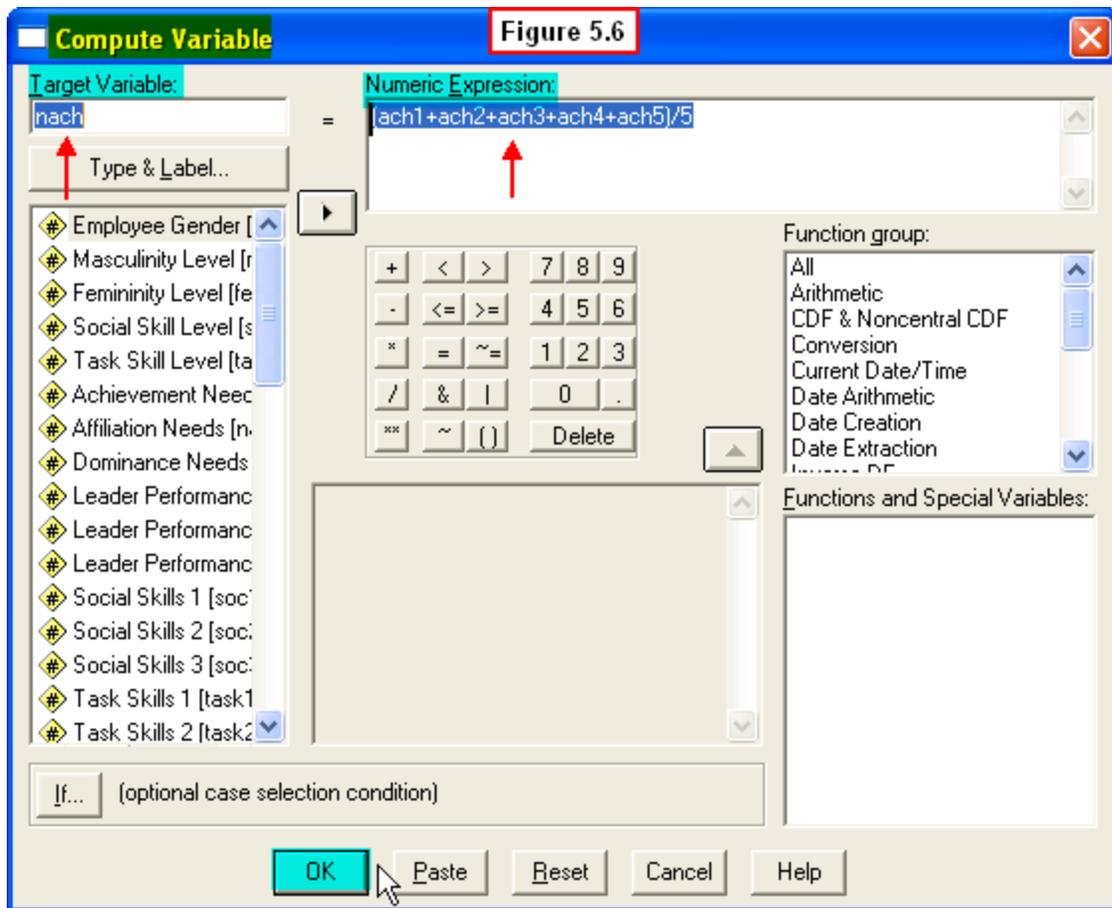
Figure 5.6

Next we turn to another **Transform** procedure subroutine, called **Recode**. This procedure allows us to create a new variable by recoding scores on another variable. This is the procedure we used to convert the continuous variables, **masctot** and **femtot** into the categorical variables, **masc** and **fem**. We also did this to create the new variables, **soc** and **task**.

It is often desirable to transform **continuous variables** (variables whose scores are on a quantitative continuum, such as from 1 to 7) into **dichotomous variables** (variables whose scores represent two different categories). Researchers sometimes do this in order to be able to treat a continuous person variable (e.g., degree of masculinity, as measured in **masctot**) as a **quasi-experimental** independent variable consisting of two groups (e.g., low vs. high masculinity, as we have done with **masc**).

Thus, we transformed the scores on the continuous variable, **masctot**, (which range from 7 to 35) into a new dichotomous variable, **masc**, consisting of just two categories and two score values (1: **low masculinity**; 2: **high masculinity**). Once this is done, we can treat this new**masc** variable as an independent variable to examine differences in leader performance between employees in the high-masculinity category compared to those in the low-masculinity group. Although this may sound complicated, the process is relatively simple using the **Recode** procedure.

5.2b The Recode Procedure

To create this new variable, we used the following recoding rules:

- If an employee's **masctot** score is **17 or lower**, then that score will be transformed to a value of **1** on the new **masc** variable (i.e., s/he will be placed in the **low masculinity** category)
- If an employee's **masctot** score is **18 or higher**, then that score will be transformed a value of **2** on the new **masc** variable (i.e., s/he will be placed in the **high masculinity** category).

To produce this transformation, we selected **Transform, Recode, Into Different Variables** from the menu in the Data Editor (Figure 5.7).
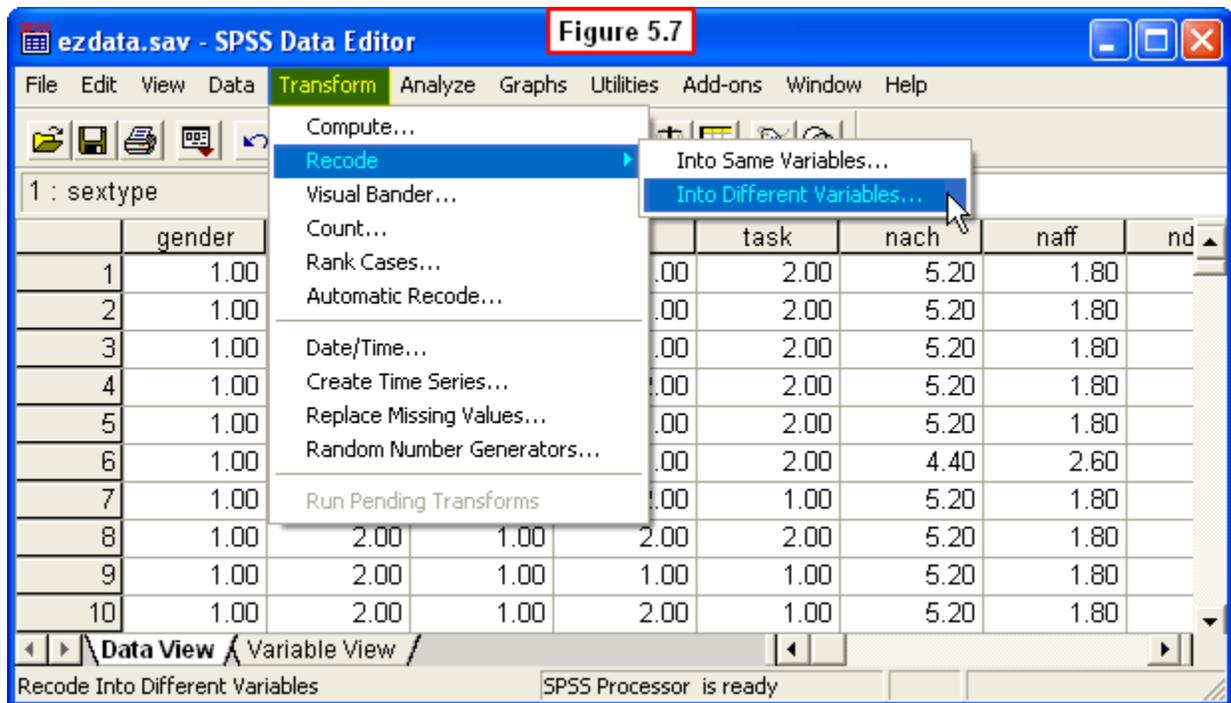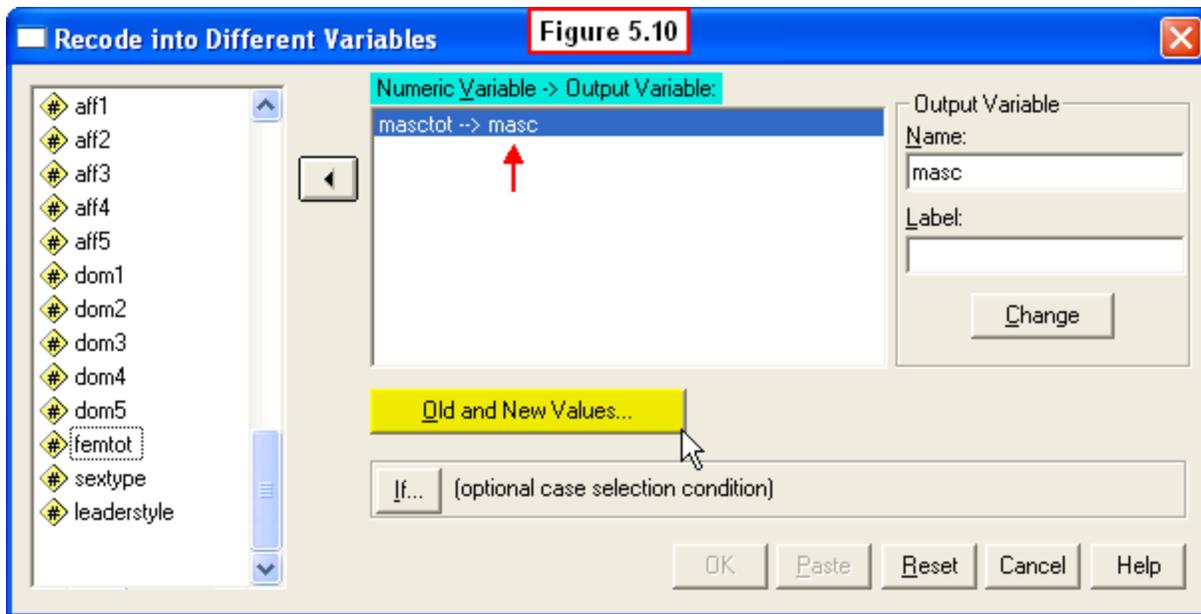

Figure 5.7

This caused a **Recode into Different Variables** dialog window to appear (Figure 5.8). We scrolled down the left pane and highlighted **masctot**, then clicked the right arrow button between panes to move this variable to the **Input Variable - > Output Variable:** pane.
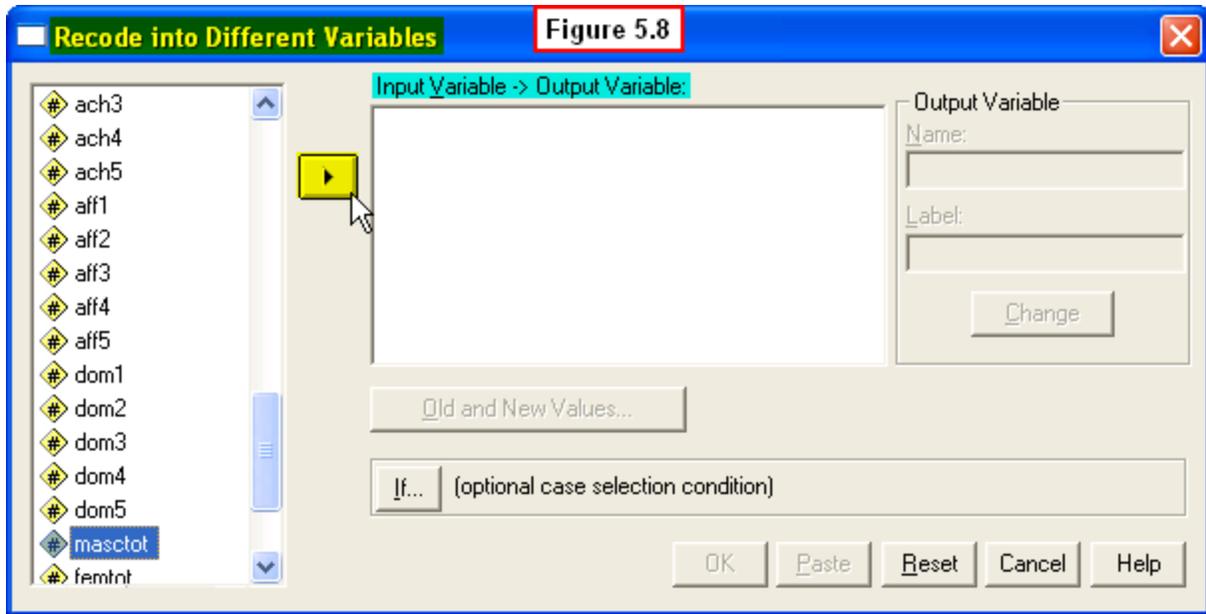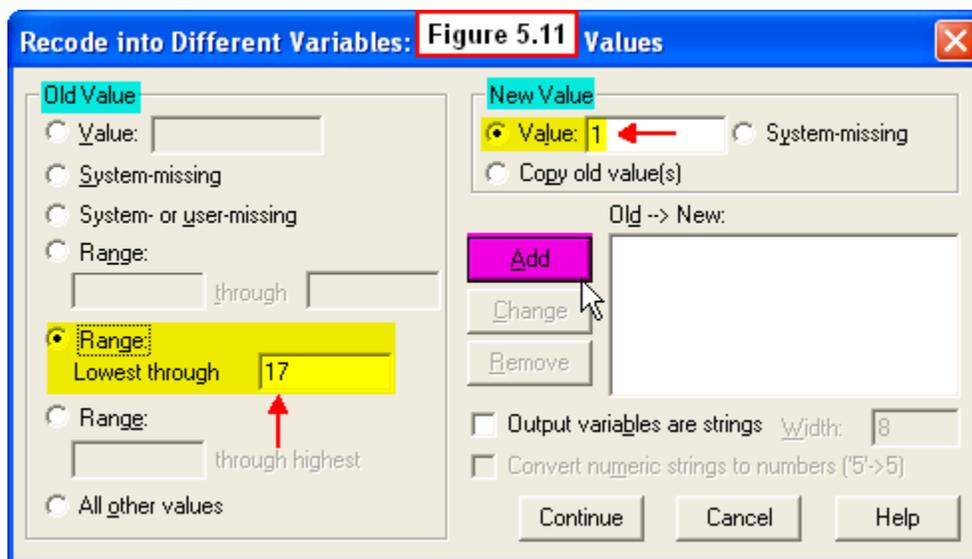
Figure 5.8

SPSS now expects the user to name the new output variable, as can be seen by the **masctot - > ?** question that appears in the **Numeric Variable - > Output Variable:** pane (Figure 5.9). We named the new output variable by typing **masc** into the **Output Variable** box on the right, then clicked the **Change** button to complete the naming.
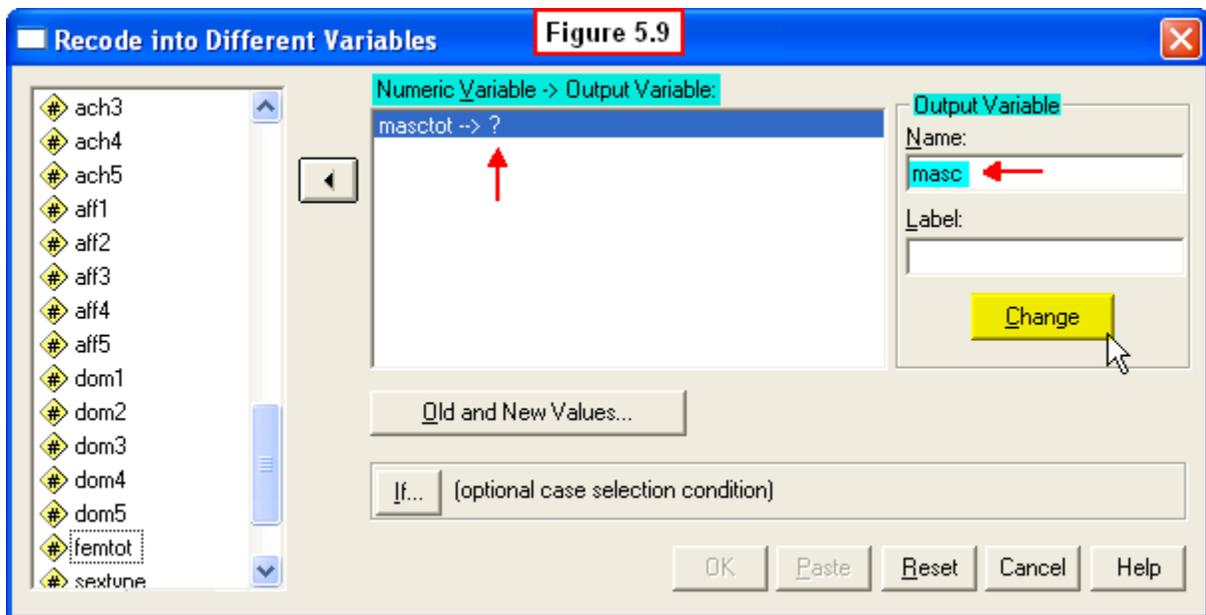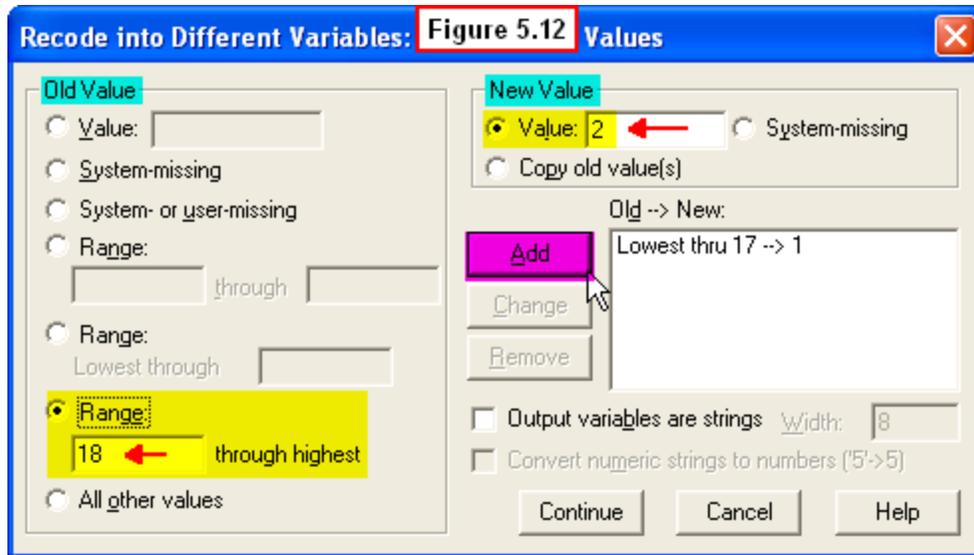

Figure 5.9

The new variable name (masc) was now displayed in the **Numeric Variable - > Output Variable:** pane. Next, to specify the recode rules to be used in creating the new output variable, we clicked the **Old and New Values** button (Figure 5.10).
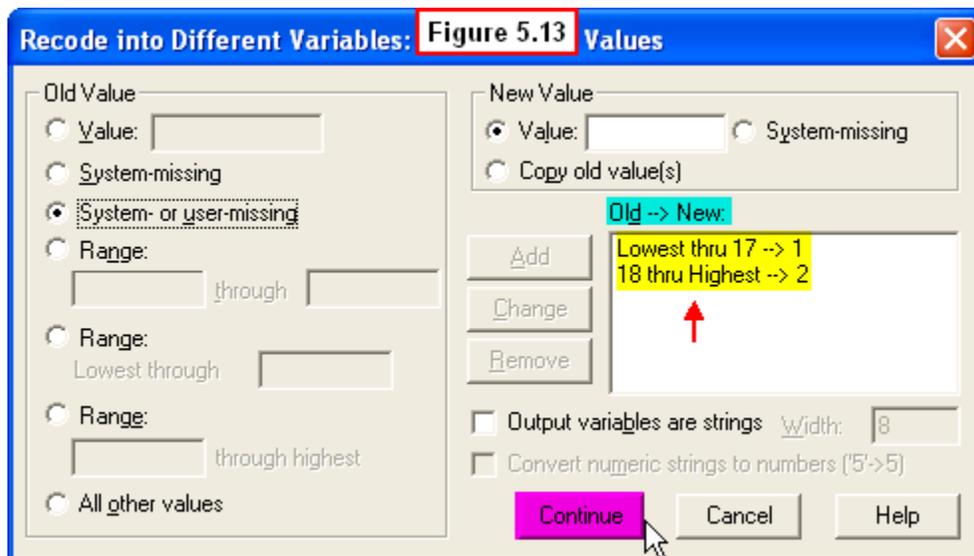
Figure 5.10

This caused yet another dialog window to appear (**Recode into Different Variables: Old and New Values**). As shown in Figure 5.11, we specified **Range Lowest through 17** in the **Old Value** pane, and specified the **New Value** as **1** on the right. This instructs SPSS to recode any employee's **masctot** score that is lower than 18 into the new value of 1 on the new**masc** variable (the **low-masculinity** category). Last, we clicked the **Add** button to complete this variable transformation.


Figure 5.11

As shown in Figure 5.12, we then specified **Range: 18 through highest** in the **Old Value** pane, and specified the **New Value** as **2** on the right. This instructs SPSS to recode any employee's **masctot** score that is higher than 18 into the new value of 2 on the new **masc** variable (the **high-masculinity** category). Again we clicked the **Add** button to complete this variable transformation.

The rules we just specified are now summarized in the **Old -- > New:** pane on the lower right (Figure 5.13). Next, we clicked the **Continue** button at the bottom of this dialog window.

This caused the **Old and New Values** dialog window to close. For the last step, we clicked the **OK** button at the bottom of the **Recode into Different Variables:** dialog window (Figure 5.14).

Figure 5.14
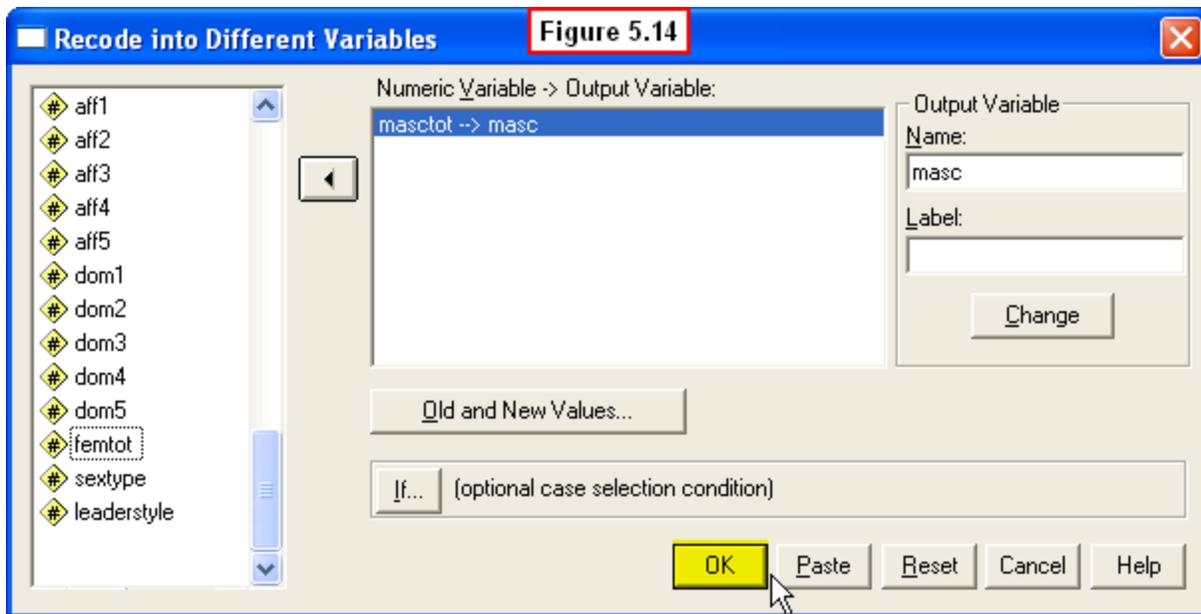
SPSS then performed this transformation and added the new recoded **masc** variable at the end of the data file (recall that we then cut and pasted this new variable so that it now appears as the second variable in the **ezdata.sav** file you downloaded). If you scroll down your Data Editor window, you will see that all the values of this new variable are either **1** (low-masculinity) or **2** (high masculinity).

Show Me Video!

We used exactly the same **Recode** procedure to create the new **fem** variable, using the same recoding rules. That is:

- If the **femtot** score was **17 or less**, then the **fem** score = **1** (**Low-Femininity**)
- If the **femtot** score was **18 or higher**, then the **fem** score = **2** (**High-Femininity**)

Last, we used the **Recode** procedure to create the new categorical variables, **soc** and **task**, by recoding the original **soc1** and **task1** variables. Since soc1 and task1 are variables with score ranges from 1 to 9, we used a different set of recoding rules:

- If the **soc1** score was **4 or less**, then the **soc** score = **1** (**Low Social Skills**)
- If the **soc1** score was **5 or higher**, then the **soc** score = **2** (**High Social Skills**)
- If the **task1** score was **4 or less**, then the **task** score = **1** (**Low Task Skills**)
- If the **task1** score was **5 or higher**, then the **task** score = **2** (**High Task Skills**)

You're probably glad that we performed all of these transformations ourselves in the **ezdata.sav** file that you downloaded! It might also seem like a lot of variables to keep straight, but you will become more familiar with them once we begin doing data analyses involving these variables.

Further, although the **Compute** and **Recode** procedures we have described can seem confusing at first (and tedious to perform), we will give you a chance do some simpler transformations at the end of this chapter. Once you have tried them, you should see that the procedures themselves are fairly straightforward.

Before we do this, in the last section of this chapter we describe one more housekeeping procedure that makes the data file (and outputs from analyses) much easier to understand: providing more descriptive names for variables and adding labels to the levels of categorical variables.

5.3 Adding Variable Labels & Value Labels

One helpful feature of SPSS that reduces confusion and information overload allows the user to provide more descriptive names (labels) for the variables in the file. Recall that SPSS has some restrictive rules for naming variables, which is why we have chosen such short abbreviations for our variable names.

But it can be hard (especially at first) to remember that, for example, the variable **masc** refers to the employee's level of masculinity. Further, it can be hard to remember that a value of 1 on **masc** means low-masculinity and a score of 2 means high masculinity. SPSS allows us to avoid this problem by adding variable and value labels to our data file.

5.3a Adding Variable Labels to the Data File

We have already added variable labels to the major variables in the **ezdata.sav** file that you downloaded. Click the **Variable View** tab of the Data Editor to see these labels (Figure 5.15).

Figure 5.15
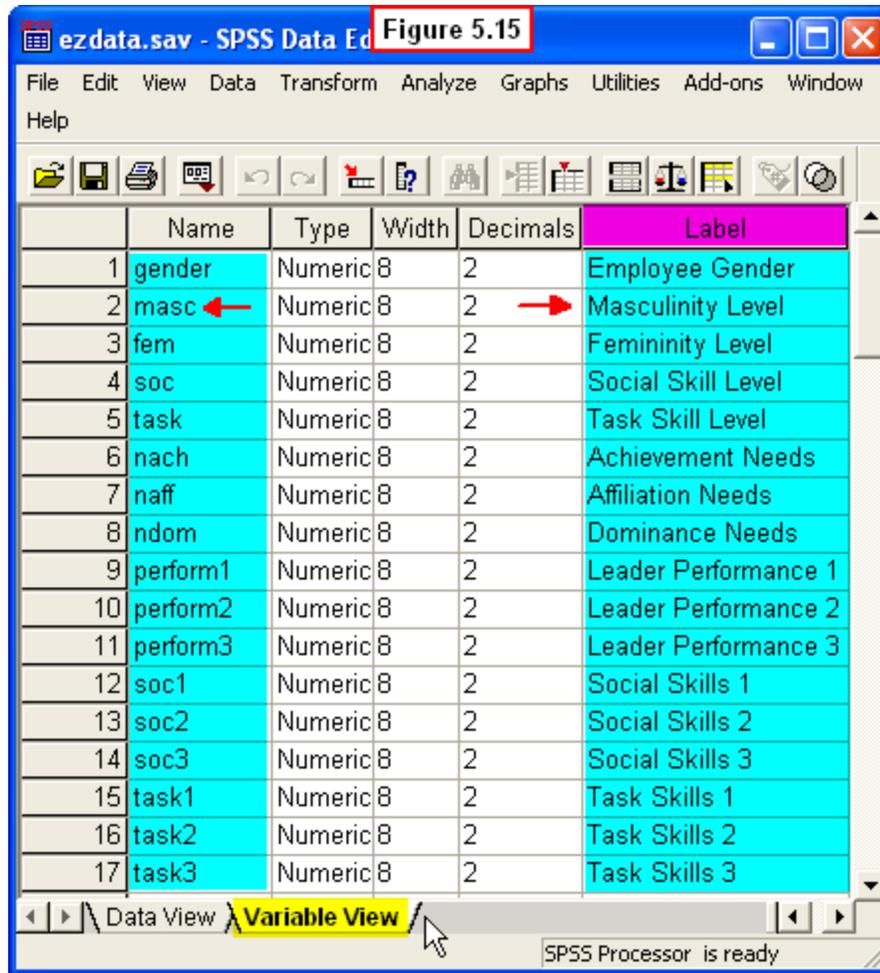
To provide these descriptive variable labels, we simply clicked on the appropriate cell of the **Label** column and typed in the label we wanted to provide for a given variable. For example, we typed **Masculinity Level** as a more descriptive label for the **masc** variable. These labels will appear on output files in place of the shorter variable name. This will make interpreting the output files much easier when we conduct data analyses. Next we turn to the process of adding value labels to the categorical variables in our data file.

As you are probably well aware by now, it can be difficult to remember what the various numerical values assigned to levels of our categorical variables (**gender**, **masc**, **fem**, **soc** and **task**) mean. The Values feature of SPSS allows us to provide descriptive *verbal* labels for these *numerical* values. As with variable labels, value labels will appear in output files. Thus, you will not have to remember that a score of 2 for gender means a female employee - instead of 2, the word **Female** will be displayed in the output file.

5.3b Adding Value Labels to the Data File

This procedure is also accomplished from the **Variable View** in the Data Editor (Figure 5.16).
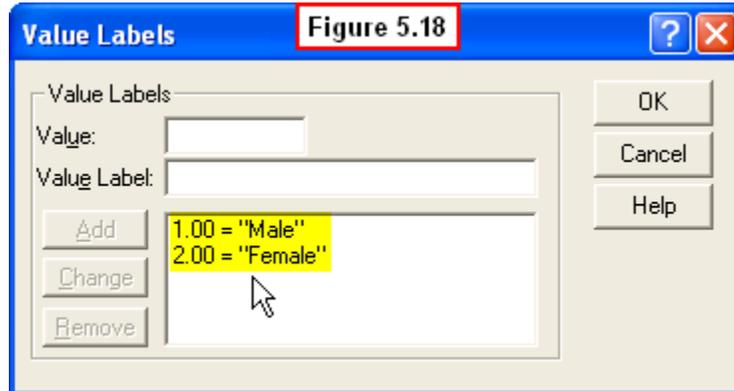

Figure 5.16
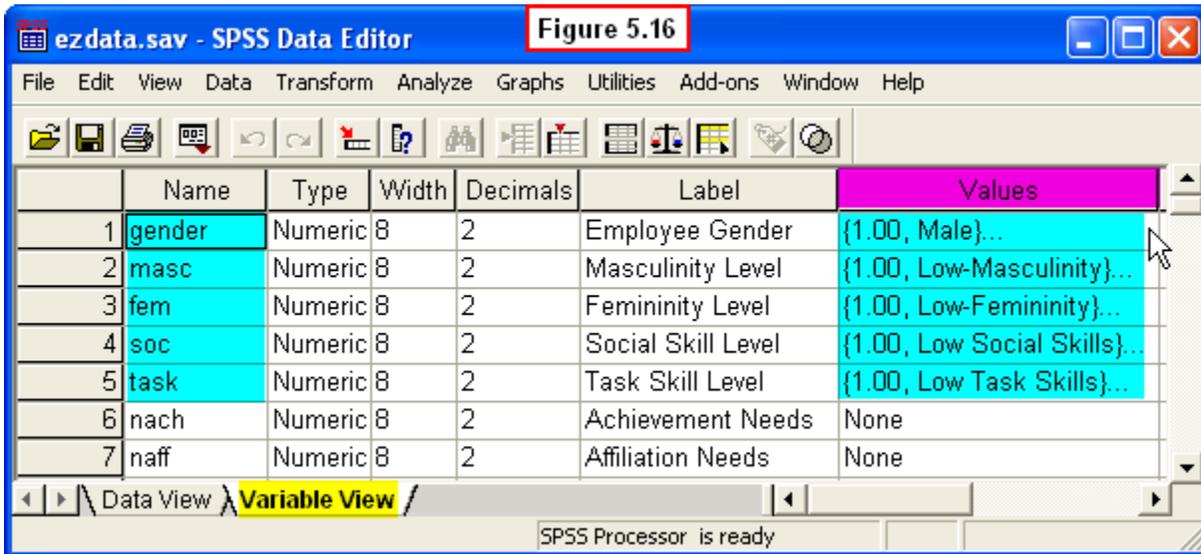
The first part of each value label we added for each variable is shown in the **Values** column of this Data Editor view. As an aside, you may be wondering why there are no value labels shown for other variables in the file (e.g., **nach**). The reason is that the other variables are continuous (as opposed to categorical). Since these types of variables can take on many different values (e.g., an average achievement need of 6.43), it would be unwieldy (and unhelpful) to attempt to provide a verbal label for every possible numerical value of continuous variables.

To see how we added these value labels, click on the first cell in the **Values** column (i.e., the value labels for **gender**). When you do this, a grey box with three dots (**...**) will appear to the right of the **(1.00, Male)...** that you see in this cell (Figure 5.17).
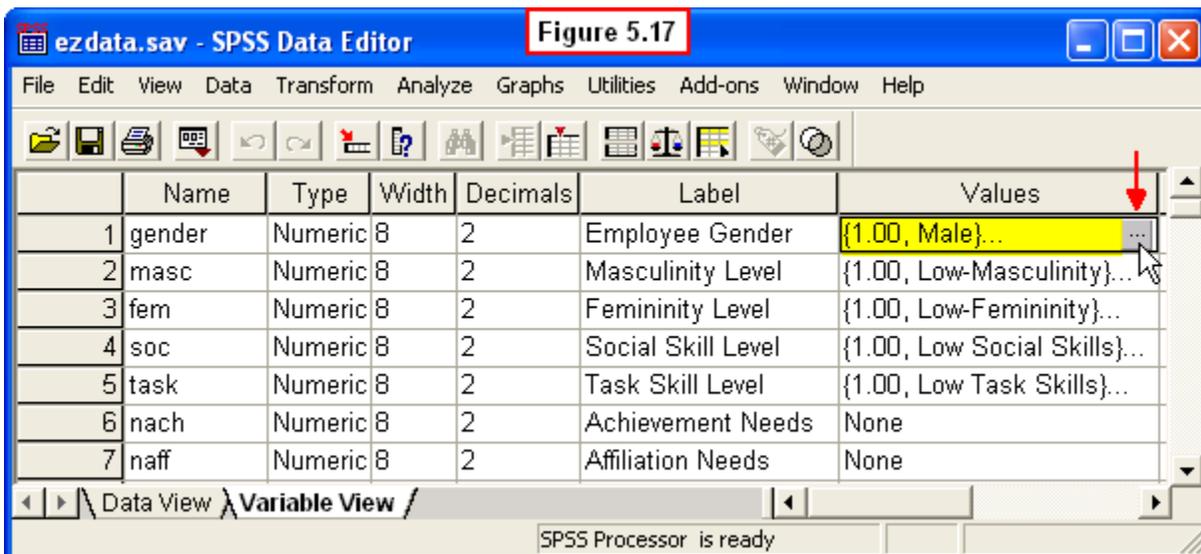
Figure 5.17

Click on the grey box and a **Value Labels** dialog window appears (Figure 5.18).

Figure 5.18

Note that since we have already provided the labels, they appear in the lower pane of Figure 5.18. In Figure 5.19 we illustrate how we created the first label (1.00 = "Male").

Figure 5.19
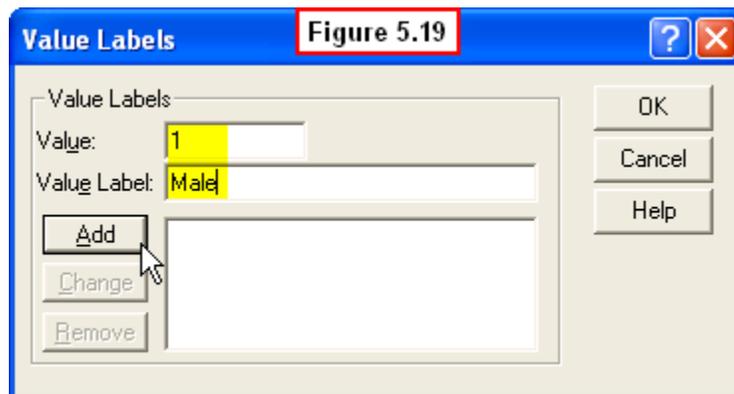
As can be seen, this simply involved typing **1** in the **Value:** box and typing **Male** in the **Value Label:** box, then clicking the **Add** button. To create the second value label, we repeated this procedure, typing **2** for value and **Female** for value label, then clicking the **Add** button again.

The same procedure was followed to create the value labels for the other four categorical variables (**masc**, **fem**, **soc** and **task**). You should click on the cells under the **Labels** column for these other variables to familiarize yourself with the labels we have added for them.

Show Me Video!

In Figure 5.20 we show an example of a frequency table we generated that displays the variable label and value labels for **gender.**

Figure 5.20

Employee Gender — Variable Label

Value Labels

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Male	110	48.2	48.2	48.2
	Female	118	51.8	51.8	100.0
	Total	228	100.0	100.0	

As this figure shows, the variable and value labels are well worth the time to create, since they greatly facilitate the ease of interpreting output of data analyses.

5.4 Chapter Review Video

Review Me!

5.5 Try It! Exercises

1. Creating New Variables, Transforming Variables & Adding Verbal Labels

For these exercises, you will need to open the **example1.sav** file that you created in Chapter 2. You will be asked to type in scores on two new variables in this data file:

- student attitudes towards computer-computation of statistics (**attitude1**)
- student attitudes towards hand-computation of statistics (**attitude2**)

1. Enter the scores on the two new variables we have added to this file (Figure 5.21).

Figure 5.21
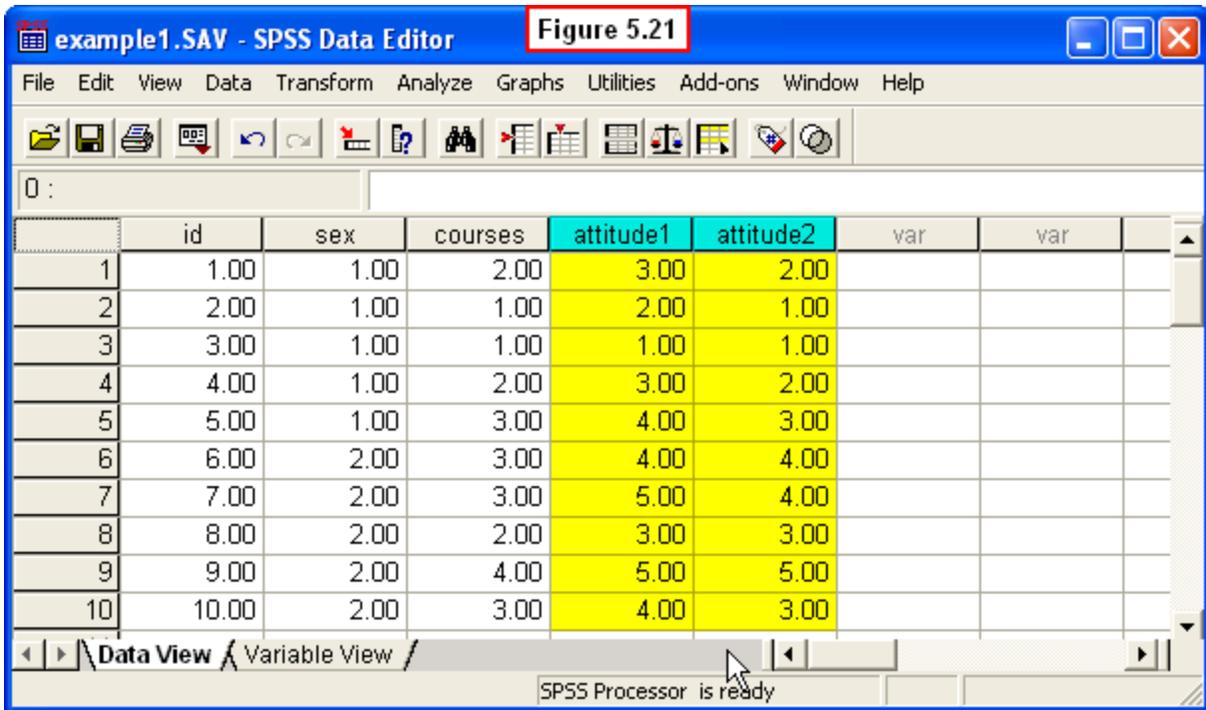
Assume that these scores represent favorableness of the attitude for computer/hand computation:

- **attitude1**: Attitude towards computer computation (1 = Unfavorable; 5 = Favorable)
- **attitude2**: Attitudes towards hand computation (1 = Unfavorable; 5 = Favorable).

Note that before you enter the scores, you should first click the **Variable View** tab and type in the two variable names (Figure 5.22).
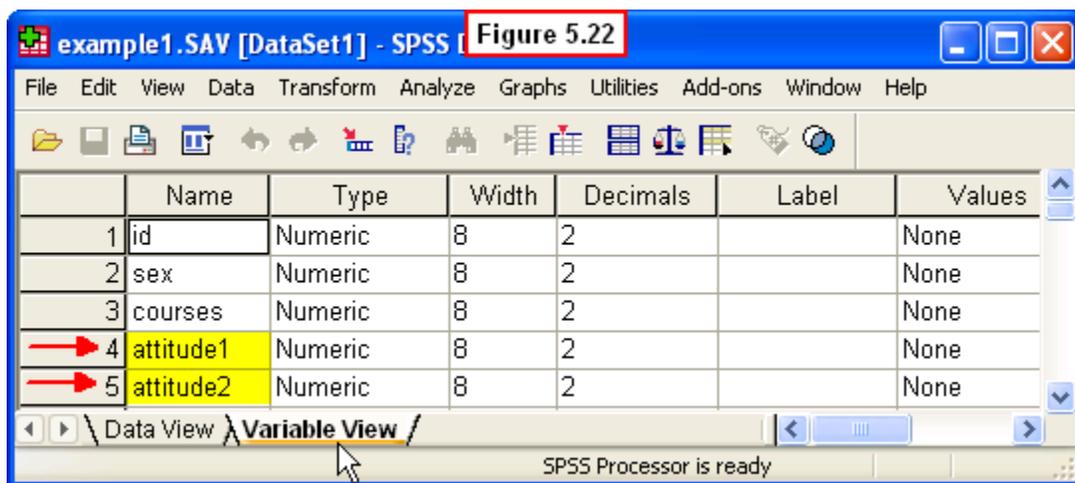

Figure 5.22

Then click the **Data View** tab and enter the scores for these two new variables shown in Figure 5.21.

2. Use the Compute Procedure explained in Section 5.2a to compute a new variable (**attitudetot**) which is the sum of the two new attitude scores:

- Select **Transform, Compute** from the Data Editor menu
- Type **attitudetot** in the **Target Variable:** box of the Compute dialog window
- Type **attitude1+attitude2** in the **Numeric Expression:** box
- Click the **OK** button

When you are finished, the Data Editor should display the new **attitudetot** variable (Figure 5.23).

Figure 5.23
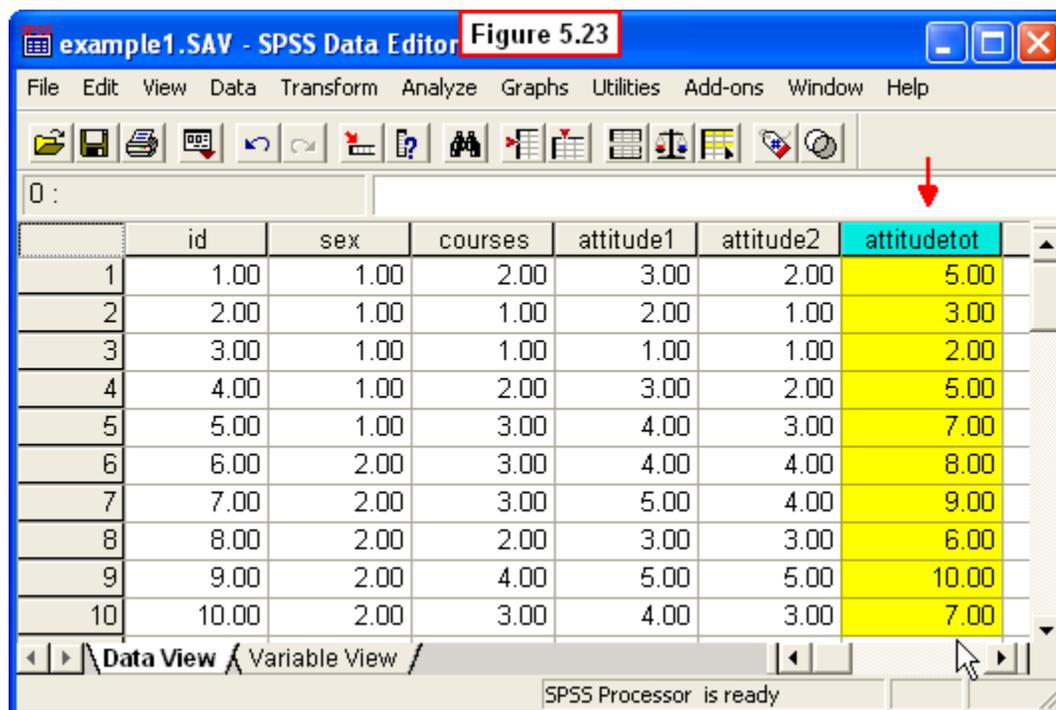
3. Use the Recode Procedure explained in Section 5.2b to transform the continuous variable, **courses**, into a new dichotomous variable, **experience**, with two categories. Use these recoding rules:

- If the **courses** score is **2 or lower**, then **experience** = **1** (Low Experience)
- If the **courses** score is **3 or higher**, then **experience** = **2** (High Experience)

To perform this transformation, select **Transform, Recode, Into Different Variables** from the Data Editor menu. A **Recode into Different Variables** dialog window will appear. Do the following:

- Move the **courses** variable into the **Input Variable - > Output Variable** pane
- Type **experience** in the **Name:** box for the **Output Variable**
- Click the **Change** button, then click the **Old and New Values** button.

The **Old and New Values** dialog window will appear. Do the following:

- In the **Old Value** panel, select the **Range lowest thru__** button, then type **2** in the box
- In the **New Value** panel, type **1** in the box, then click the **Add** button
- Return to the **Old Value** panel, select the **Range __ thru highest** button, then type **3** in the box
- In the **New Value** panel, type **2** in the box, then click the **Add** button, and then click the **Continue** button.

Last, click the **OK** button in the **Recode into Different Variables** dialog window. When you are done the Data Editor window should show the new categorical **experience** variable with scores of either 1 or 2 (Figure 5.24).
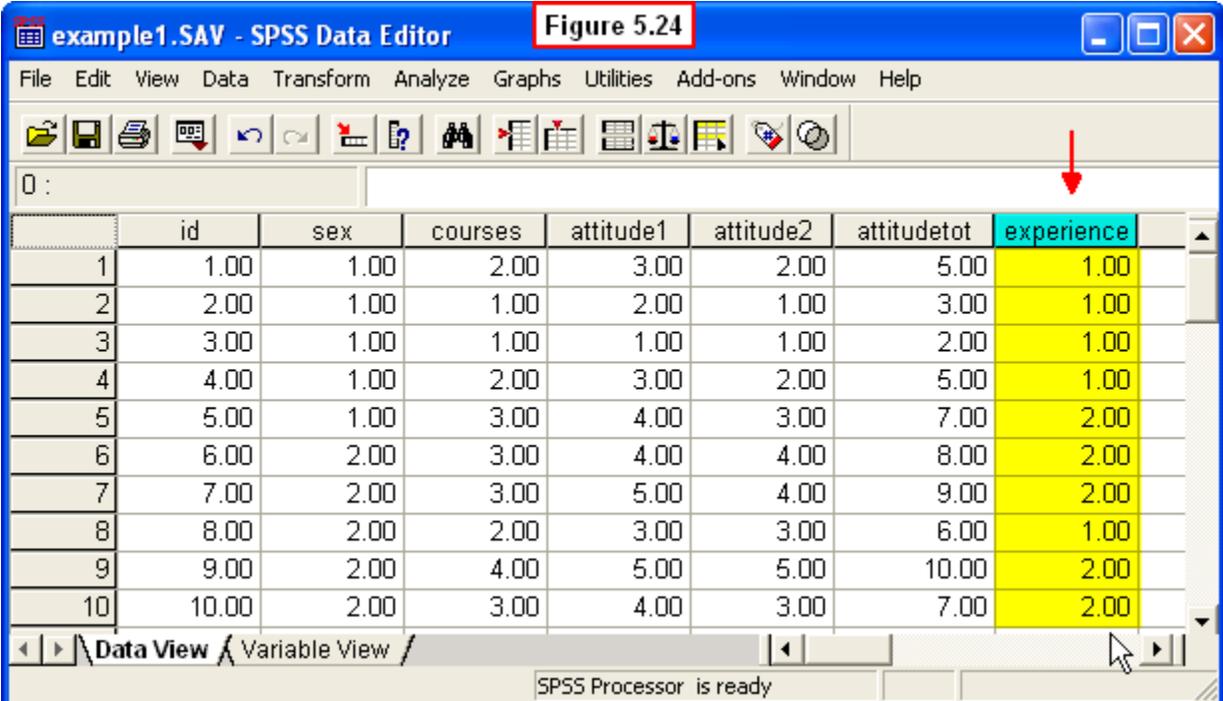

Figure 5.24

	id	sex	courses	attitude1	attitude2	attitudetot	experience
1	1.00	1.00	2.00	3.00	2.00	5.00	1.00
2	2.00	1.00	1.00	2.00	1.00	3.00	1.00
3	3.00	1.00	1.00	1.00	1.00	2.00	1.00
4	4.00	1.00	2.00	3.00	2.00	5.00	1.00
5	5.00	1.00	3.00	4.00	3.00	7.00	2.00
6	6.00	2.00	3.00	4.00	4.00	8.00	2.00
7	7.00	2.00	3.00	5.00	4.00	9.00	2.00
8	8.00	2.00	2.00	3.00	3.00	6.00	1.00
9	9.00	2.00	4.00	5.00	5.00	10.00	2.00
10	10.00	2.00	3.00	4.00	3.00	7.00	2.00

4. Add a descriptive Variable Label and Value Labels for the new **experience** variable using the procedures explained in Section 5.3b. First click the **Variable View** tab in the Data Editor.

- Click on the cell in the **Label** column for the **experience** variable
- Type **Psychology Experience** in this cell.
- Click the cell for this variable in the **Values** column; then click the 3 grey dots (**...**)

A **Value Labels** dialog window will appear. To add the labels:

- Type **1** in the **Value:** box; type **Low Experience** in the **Value Label:** box; click the **Add** button

- Type **2** in the **Value:** box; type **High Experience** in the **Value Label:** box; click the **Add** button
- Click the **OK** button of the **Value Labels** dialog box

When you are finished, the Data Editor window should look like the one in Figure 5.25
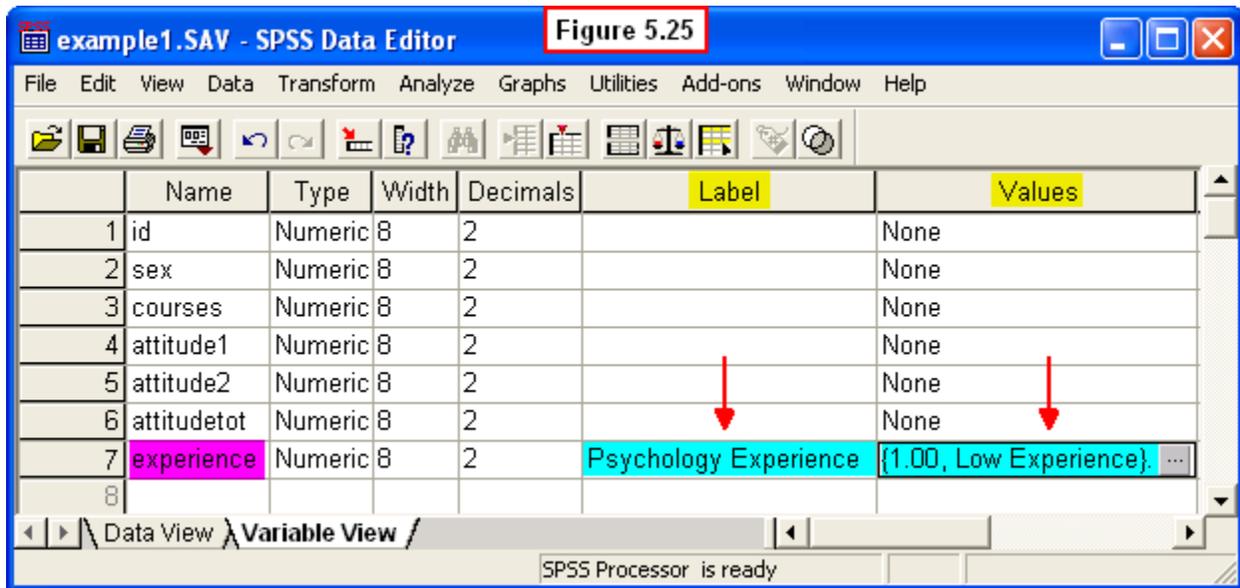
5. Click the **Data View** tab, then save and **print** this file to submit to your instructor.