Handling Your Data in SPSS 5

Columns, and Labels, and Values . . . Oh My!

You might think that simple intuition will guide you to a useful organization of your data. If you follow that path, you might find yourself ready to analyze your data and unable to do so. In this chapter, we offer advice to help you avoid that dead end and introduce you to the data section of SPSS. The data management tools in SPSS will help you to manage your data so that you will be able to (1) look back years later and still know what your data measured and (2) avoid simple calculation errors. We will show you how to enter your data, provide labels for all of your variables, and perform some simple calculations and data file manipulations. Before we do that, we will briefly introduce you to the structure of the SPSS program.

The Structure of SPSS

You should think about SPSS as having three major parts.

1. The first part is where you store and manage your data. This part of the program is described in some depth in this chapter and works much like any database program and not too unlike most spreadsheet programs.
2. The second part of the program actually consists of many procedures that you can use to calculate descriptive statistics, construct some graphs, and most important, produce a wide range of **inferential statistics**. In Chapters 7, 8, 9, 10, and 11, we will show you how to use many of the statistical analyses that beginning and intermediate researchers need.

3. The third part of the program presents the output or results of your analyses. A useful feature of SPSS is that you can save your output separately from your data. You will find that saving your output can save time later on when you are writing final drafts of your paper, poster, or talk. These files can be read even when you do not have access to SPSS if you download a special reader from IBM. (Search for *IBM SPSS Smartreader*.) However, you will likely find it more convenient to export your output tables to files with more widely used formats, such as Microsoft Word, Excel, HTML, or PDF documents.

When to Create Your Data File: Yes, Even Before Data Collection

5

The first thing you should do is set up your data file. Yes, do this before you collect the first bit of data. It might sound like you are "putting the cart before the horse," but trust us; this is an important order of events. Setting up the file helps you to think about the data you are collecting and whether or not you understand how you will analyze it. And knowing you planned ahead of time will reduce your stress later. You might even avoid a panic moment when you realize you did not collect the right kind of data to answer your research question.

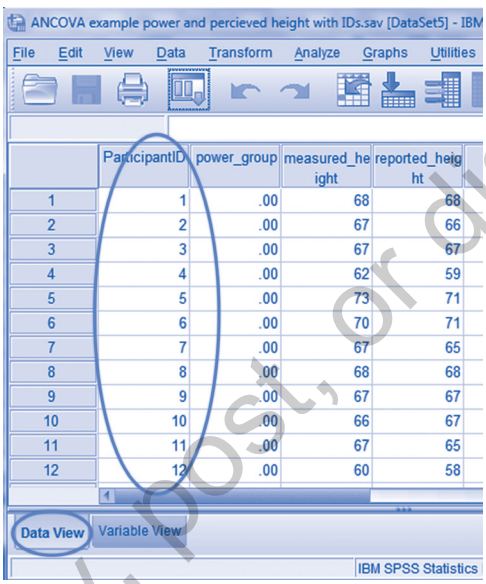
Second, collect your data. Unless your research is completely controlled by a computer (such as an online survey), you will likely have to record some measurements on paper. You want to plan ahead so that it is easy to transfer numbers or other data codes from that paper (or online data collection tool) to the computer file. You also want to make sure that every important bit of data gets recorded and stored on paper as well as in SPSS. We have seen some disasters when this step was overlooked. The worst one was a young researcher who forgot to include a code for level of the IV for each participant! As a result, the experiment had to be conducted a second time with new participants. You can only imagine the student's reaction when she realized she had to restart data collection from scratch. To avoid that type of error, you should, as a general rule, begin entering data for your study after you have tested only two or three participants. When you enter data very early in your project, you can make sure you have included all of the measures necessary to test your hypotheses. This practice, together with setting up the SPSS file before you begin data collection, will avoid time-consuming mistakes.

Setting Up Your Data File

This is probably a good time to tell you that SPSS can import data from many spreadsheet programs like Microsoft Excel. We will show you how that works in the next section. If you are more comfortable typing data into one of those

programs, you may do that and then later import your data to SPSS. We strongly recommend that you only enter the raw data into your spreadsheet and then let SPSS perform any calculations you need. We also recommend that you follow our earlier advice and enter some data early and then import them to SPSS to ensure that you will be able to use the data the way that you intend.

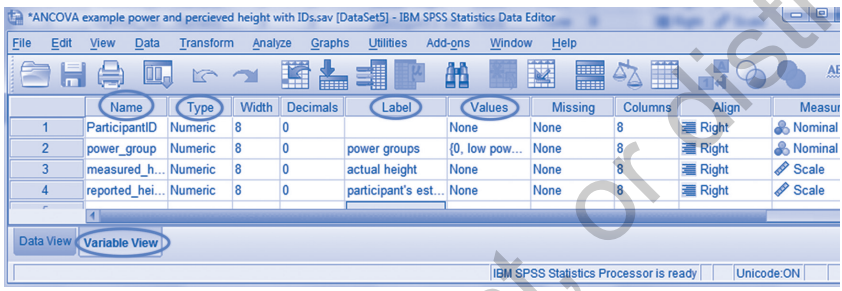
The next screenshot shows you some data that we have entered in SPSS. We are starting out in the ***Data View***.



Each row in your data sheet should represent a single participant or subject, while each column will represent a different variable in your research project. We typically label the first column of our data file as "participant," "ID," or "case," which helps to keep track of each participant's individual data. The data point in this column is usually a number. (Notice that we suggest you do not use names because we often promise to keep data and identities of participants separate from each other!) We include the same identifier, again usually with a number that we assign to a specific participant on every sheet of paper tied to that individual, excluding the informed consent form (again to protect confidentiality as promised to your IRB). This system allows us to find and correct errors in data entry if needed. Every now and again, after we have entered some data, we notice that a score is beyond the range of possible values for that variable. When that happens, if we code each participant's data with an identifying number, we can find the correct data sheet (with the same identifying number) and then double-check all of the data that we have entered into SPSS from that form.

Before you enter any data in your file, you should label each column with a variable name. Take a careful look at the previous screenshot of a small data

file from SPSS. The data set is one that we made up for an example in Chapter 9. In this research design, there are two power groups (low and high) and two dependent measures (reported height and measured height). Notice the two tabs at the bottom of the spreadsheet; one reads *Data View*, and the other *Variable View.* You will use the *Data View* to type in data and see each data point in your data file. You will use the *Variable View* to enter variable names (e.g., measured_height), perform some useful housekeeping (e.g., add labels for your independent variables), and perhaps record some notes about each variable. You get to the *Variable View* by clicking on the tab circled at the bottom of the screenshot below.

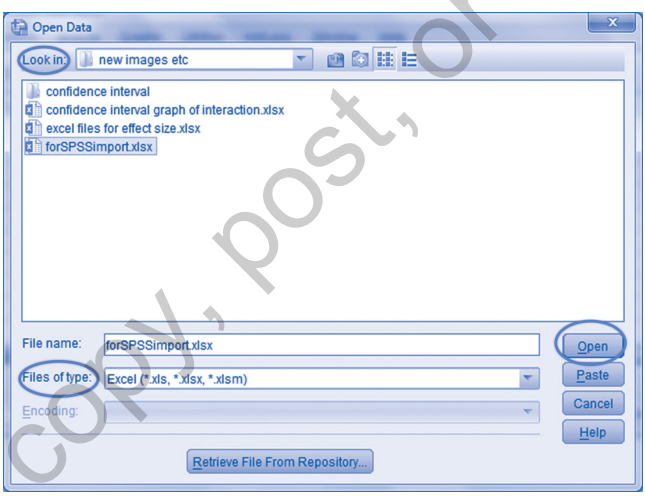


This screenshot shows what the *Variable View* looks like. Notice that the tool bars at the top of the page have not changed but that some icons (Redo, Find, Insert Case, Value Labels) are grayed out. These icons are grayed out because you can only use them in the *Data View.* It might take you a few minutes to get oriented to this different view of your data in SPSS. Some of your disorientation might be produced by the fact that in *Variable View*, a row (rather than a column) represents each variable. Notice that ParticipantID is in the first row of this screenshot, but it was in the first column in the *Data View.* Now look at the row of labels (e.g., *Name*, *Type*, *Width*, *Decimals*, etc.) just above the information that we have entered in the white columns and rows. You will need to be aware of each of those columns in the *Variable View* for your own data, which we describe for you below.

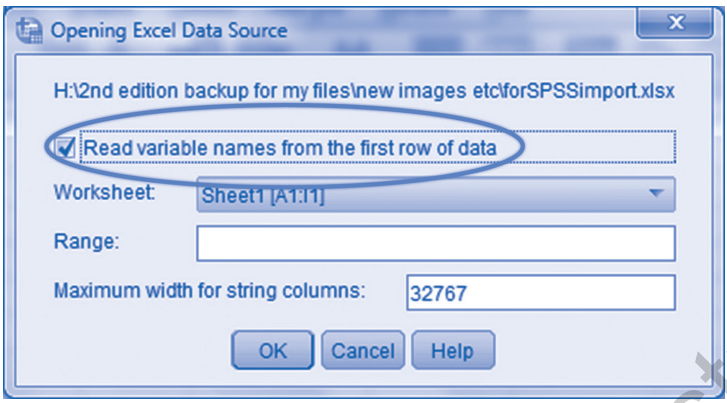
Importing Data

We are going to take a brief detour and show you how to import data from other programs. We often import from Excel or other spreadsheet programs. You can also import data directly from most online survey or other electronic data collection tools. The most general format for moving data files between applications is the CSV or comma-separated value file. The CSV format usually produces a text file with a list of variable names in the first line of text, followed by data for each case on a new line, with commas between each entry on a line. The next screenshot shows you the *Open* file menu in SPSS.

We open this drop-down menu by clicking on *File* and then *Open*, which opens the short list of SPSS file types that appears on the right in the image. We next click on *Data*, which is circled in the above screenshot. In the next screenshot, you can see that the SPSS *Open* file menu looks just as you would expect.

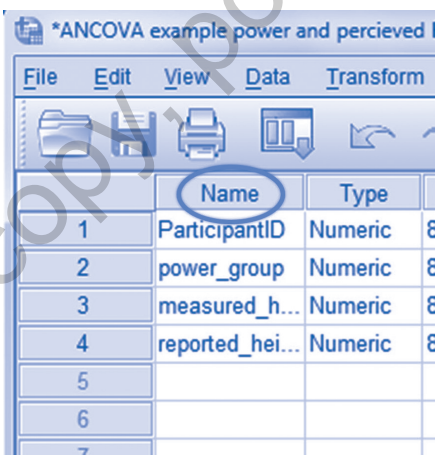


Here, we navigate to the folder where our data file is located by making appropriate selections in the *Look in* box. We select Excel from a drop-down menu in the *Files of type* box. This selection means that only Excel files will show in the dialogue box. That filter makes it easier to find the file we are looking for. Of course, we then click on our file name, which is highlighted in the list of files and appears in the *File name* box. Once our file name appears, we click on the *Open* button to select the desired file. As you can see in the next screenshot, a new dialogue box will appear after you click the *Open* button. The *Read variable names from the first row of data* is automatically checked for us. You should always create these kinds of data files with variable names in the first row of the file. When we click the *OK* button, the data are then imported into SPSS. This action will produce a bit of output. You should immediately save the new data file as an SPSS file.



5 Naming and Labeling Your Variables

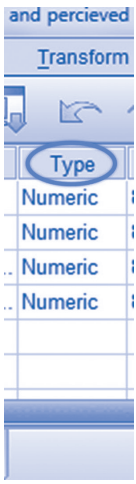
Regardless of how you will enter your data, you want to make sure that you record enough information about your variables so that you won't have to remember too much about your data file. In the next screenshot, we will look again at the *Variable View*. The first column contains the *Name* of your variables, which is circled below.



You probably have also noticed that some of our variable names look a little odd. We agree, and here is why. When SPSS was first written there were severe restrictions on how variables could be named. Many of those restrictions have been removed by recent innovations in programming. However, the program still does not allow a blank space in a variable name. As a result, we typically either run words together without spaces or add an underscore between terms, often creating unusual variable names. For example, you can see in the last screenshot that the first row is labeled "ParticipantID," which is not a word. In the third variable

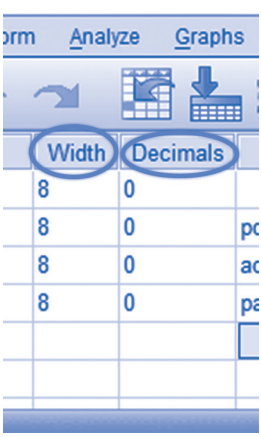
(measured height), we used an underscore (_) in place of a space within the variable's name (i.e., measured_height).

The second column identifies the *Type* of data for each variable.

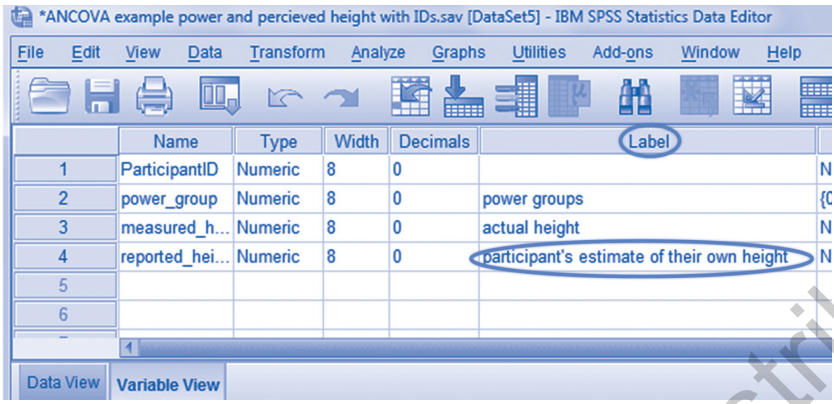


SPSS has several data types to choose from, and you need to choose wisely. Most of your data will be numeric; however, you might also have strings or words (e.g., you might want to record comments that your participants made during the experiment). Be careful about setting a data type to *String* or *Text*, as those variables will not be available for any of the data analysis tools in SPSS.

The next column, *Width*, establishes the maximum number of digits or characters that will be displayed for that variable, and the following column, *Decimals*, sets the number of decimal places that will be displayed. You should set this value to 0 for categorical and integer measures.



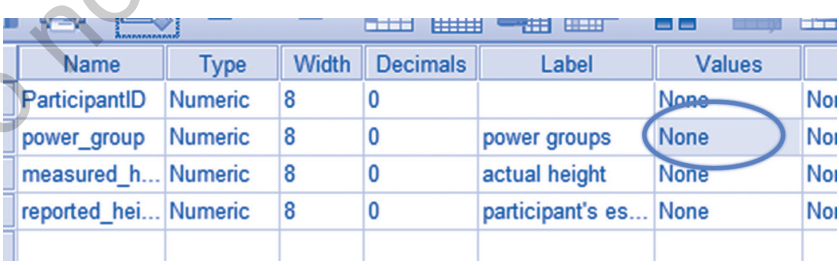
The column called *Label* can be valuable in helping you keep track of exactly how you measured something.



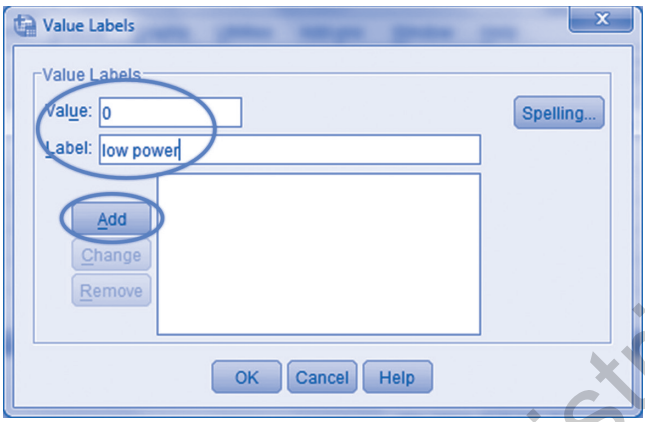
For example, we typically copy the exact wording for any question on a survey into this column to have that information both preserved with the data and included in output tables. You might think of putting any brief notes here that might help you to remember how you conducted this study later on. For example, "DQ scores were measured with the Bailey" or "individuals were randomly assigned to the groups." Notice that in the current example, we made a notation that *reported_height* is the "participant's estimate of their own height." This *Label* will appear on all of our output and we would not need to remember who reported the height.

The *Values* column allows you to enter text labels for specific data values. These labels help you to remember the meaning of numbers that you used to code variables. For example, we might use 0 and 1 to identify females and males in our sample. With these labels, we can make SPSS remember that we coded 0 for females and 1 for males. These labels will show up on our output, too, so we do not have to remember which group we identified as 0 and which as 1.

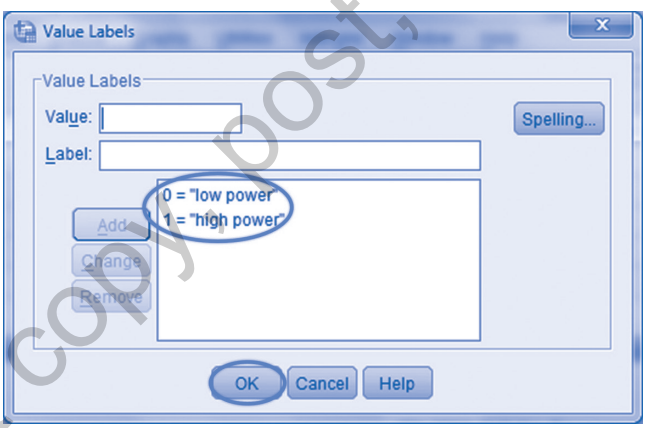
We will look at that option in the next screenshot. To get started, click on the word *None* in the value column next to the variable that you want to work with. The next screenshot shows you where to click to add value labels for the two power groups.



The next two screenshots show you the dialogue box that appears when you want to include labels for a variable.



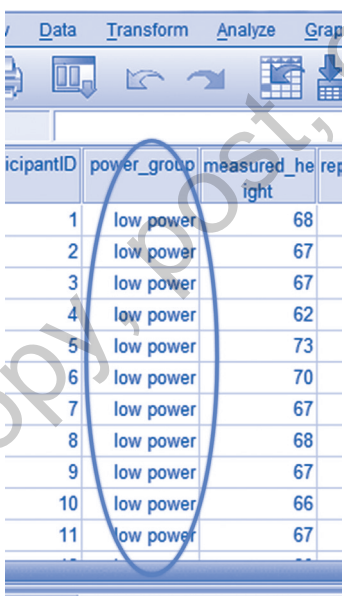
In this first box, we have entered a "0" in the *Value* box and "low power" in the *Label* box. Notice that the *Add* button is ready to be clicked. As you can see in the next screenshot, after you click the *Add* button, the value and label move down as a pair into the larger unlabeled box circled below.



In the completed dialogue box, you can see that "0" represents the low-power group, and a value of "1" represents the high-power group. We added the "high power" label by following the same steps outlined above. Clicking on the *OK* button will take you back to your data file. You probably noticed that we used values of "0" and "1" to designate our two groups. Many researchers use those values, often using "0" to identify a control group. These two values have some added benefit if you will be doing a very advanced **multiple regression** analysis; however, for any other analysis, you may use any two numbers to designate two different groups. In this book, you will notice that we sometimes use "1" and "2." Of course, more value labels are needed when you have more than two groups.

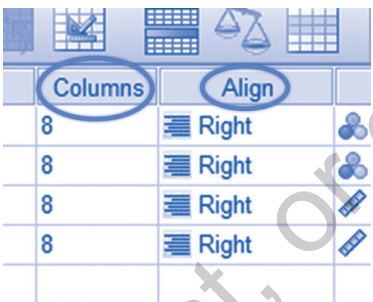
Once you have entered value labels, they will appear on your output so you do not have to remember if you used "0" and "1" or "1" and "2" or which group they indicate. You may also view those labels in the *Data View.* If you switch from *Variable View* to *Data View*, there is a row of icons at the top of the *Data View* spreadsheet. Find the icon that looks like this:

When you click that icon, you will toggle between the number and label view for any variables that have *Value Labels.* The next screenshot shows you what the labels for our power groups look like. Notice that in the *Data View*, the 0's have been replaced with the label "low power."

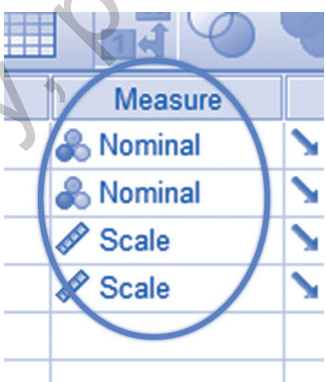


We need to return to the *Variable View* screen.

The next column to the right of *Values*, labeled *Missing*, can be used to specify values that you will enter for missing data. You likely will not want to do that. This tool is most useful when you might have several reasons that data are missing (e.g., participants refused to answer, responded "did not know," or simply left the item blank). For most of your missing data you can simply leave that cell empty in your spreadsheet. The next two columns are used to change the appearance of columns in the *Data View.* You can increase or decrease the space on the data sheet for each variable by changing *Columns*, and you can left adjust, right adjust, or center the values in the *Data View* with *Align*.



That last column, *Measure*, is really important.



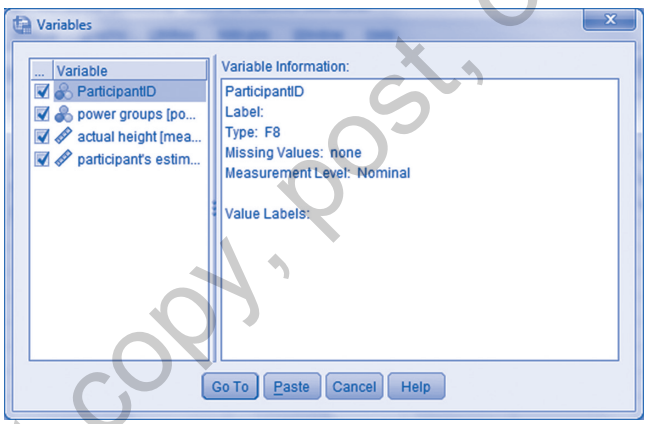
You can see that power_group is a *Nominal* measure, but the other two are *Scale* measures. SPSS treats interval and ratio measures the same way and refers to them as *Scale* measures. (There is a third choice: *Ordinal*.) You can review the distinction among these different kinds of measurements in Chapter 2. Making the wrong choice here might prevent you from conducting the kind of analysis that you planned. You will make changes in *Measure* simply by clicking on the cell and selecting the measure (e.g., *Nominal*, *Ordinal*, or *Scale*) that you want.

How to Keep Track and Remember the Details of Your Data File

This next tool will help you to remember what you did when you look at a data set that you created years ago or to understand data that other researchers have created. In the *Data View*, you will see an icon on the toolbar that looks like this:

5

Clicking on that icon opens a dialogue box that allows you to see all the information that has been entered about any of the variables in the file.



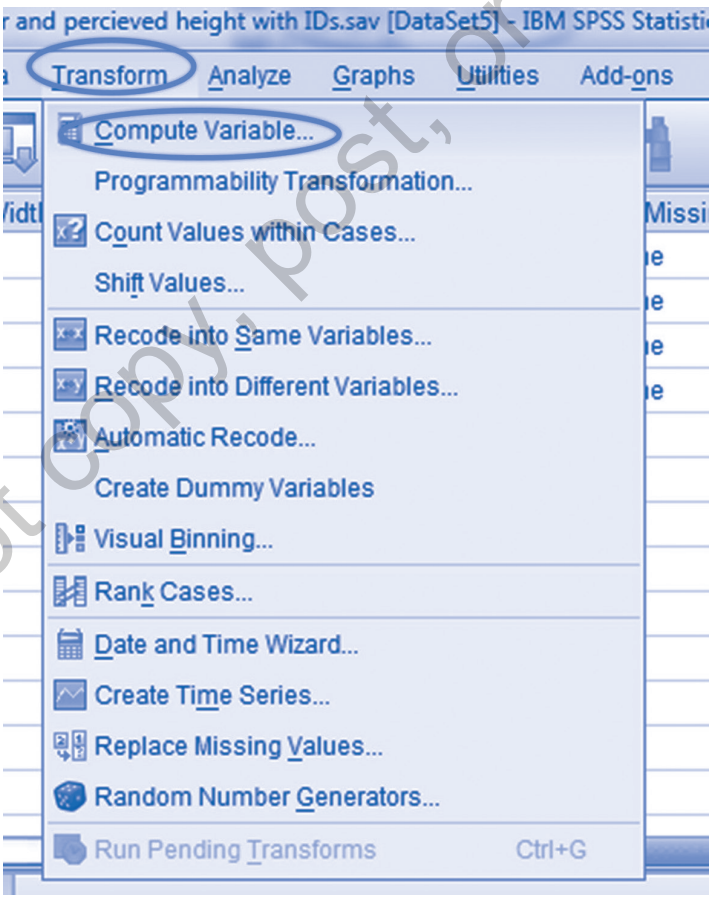
In the example that we have opened, you can see that the "power_group" variable's *Measurement Level* is *Nominal* and that it has the *Value Labels* we created earlier.

Now that you know how to enter your data and the information associated with them into SPSS, you are ready to consider how data can be managed with the program. For example, we will show you some tools that allow you to calculate new measures or divide your data file into parts based on groups. You should always enter every bit of data from a participant on a single row in your data file. Create as many columns as needed for each variable you measured or manipulated. After you have entered all of your data, you will be able to ask SPSS to perform any of these calculations that you need to transform or recode a variable if needed. Read on to see how this all works.

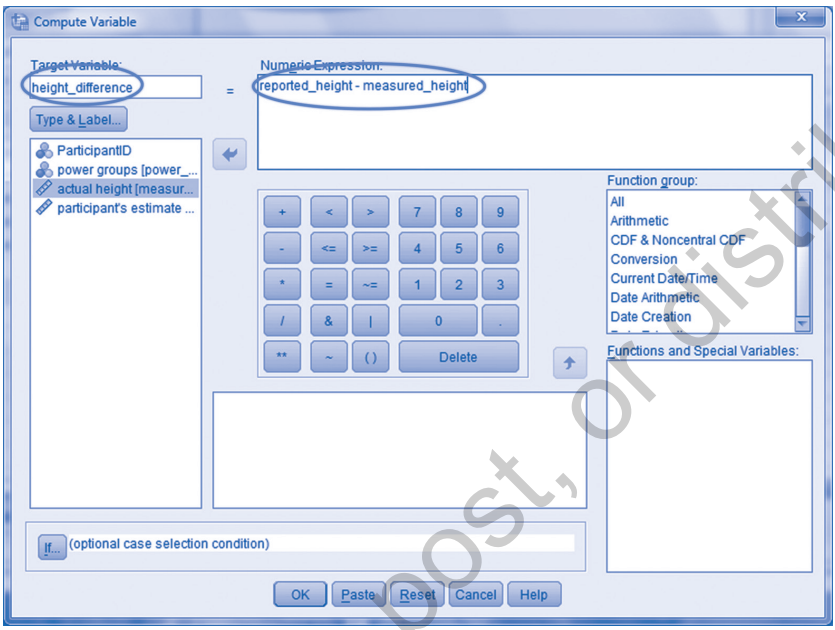
Creating New Variables in Your Data File: Transformations

There are times when you will not need to make any calculations or other changes to your data. However, there are many occasions when you will make calculations. For example, imagine that you have created a new survey to measure food preferences. Imagine that you have asked five questions that evaluate "liking" for junk food and five questions that evaluate "liking" for healthy foods. You will likely want to combine the answers to each of those five questions to get a measure of overall preference for junk and healthy foods. You might even want to calculate the difference in the two measures to evaluate how much your participants prefer one kind of food over the other.

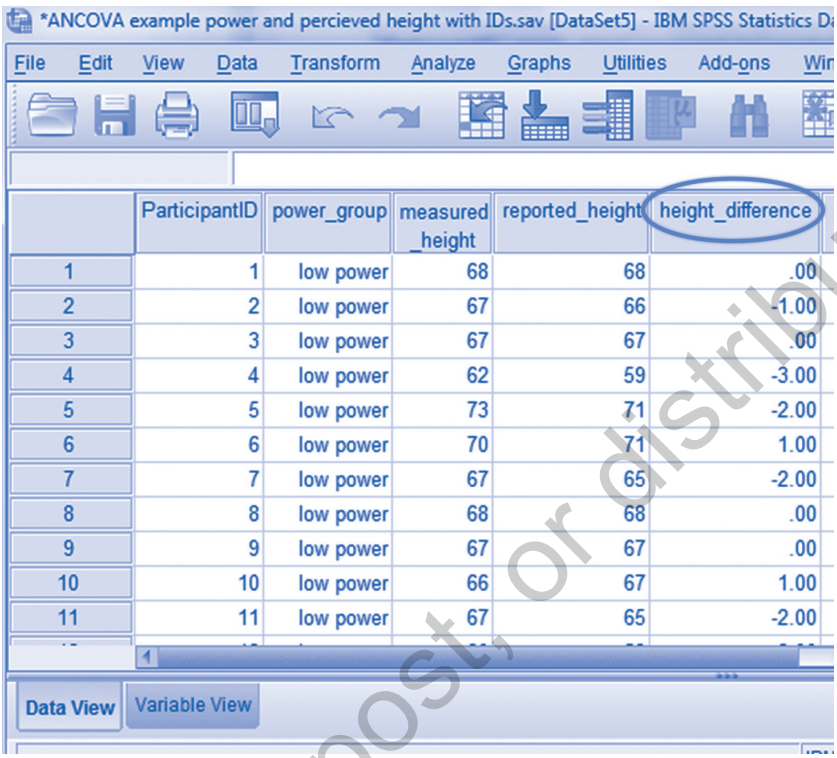
In SPSS, you can conduct calculations by clicking on the *Transform* drop-down menu and then choosing *Compute Variable.*



Choosing *Compute Variable* will open the following dialogue box, which allows you to set up the calculations needed to create new variables using the variables you already entered in your data file. Let us take a look at a simple example.



First, notice that all of the variables in your data file are listed in the large text box at the left. The box at the far right labeled *Function group* contains statistical, mathematical, and other kinds of calculations that you might need for more advanced manipulations of your data. For this example, we wanted to create a new variable measured as the difference between reported height and measured height. To calculate the difference, you must enter the name of your new variable in the box labeled *Target Variable*. We typed the name of a new variable, height_difference, in that box circled above. To calculate this new variable, we moved the names of the two variables necessary for our calculations (reported_height and measured_height) from the list of variables over to the box labeled *Numeric Expression* (circled above). Finally, we placed a minus sign ("–") between the two variables and clicked the *OK* button. After you click the *OK* button, a new column will be created in your data file, and a value for that new variable will be included for each participant. The next screenshot of the data illustrates that column (circled).



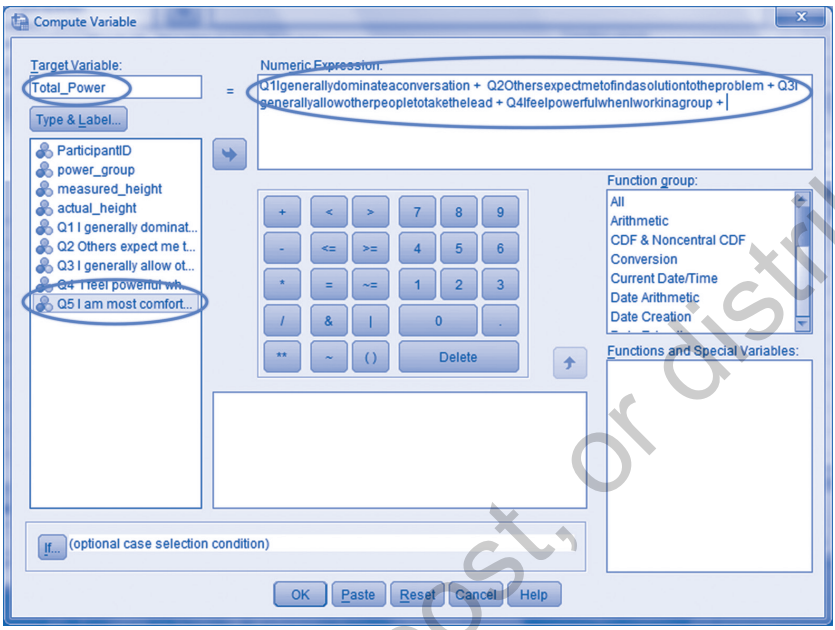
	ParticipantID	power_group	measured_height	reported_height	height_difference
1	1	low power	68	68	.00
2	2	low power	67	66	-1.00
3	3	low power	67	67	.00
4	4	low power	62	59	-3.00
5	5	low power	73	71	-2.00
6	6	low power	70	71	1.00
7	7	low power	67	65	-2.00
8	8	low power	68	68	.00
9	9	low power	67	67	.00
10	10	low power	66	67	1.00
11	11	low power	67	65	-2.00

Again, we recommend that you put all of your "raw" data into your SPSS data files and then use this procedure to perform any calculations that you might need on those raw scores. Following this advice will avoid calculation errors that you might make with a calculator while retaining responses to individual items. In this example, our data file retains the measures of actual and estimated height and includes the additional measure of the difference between the two.

Calculating a Total or Mean Score

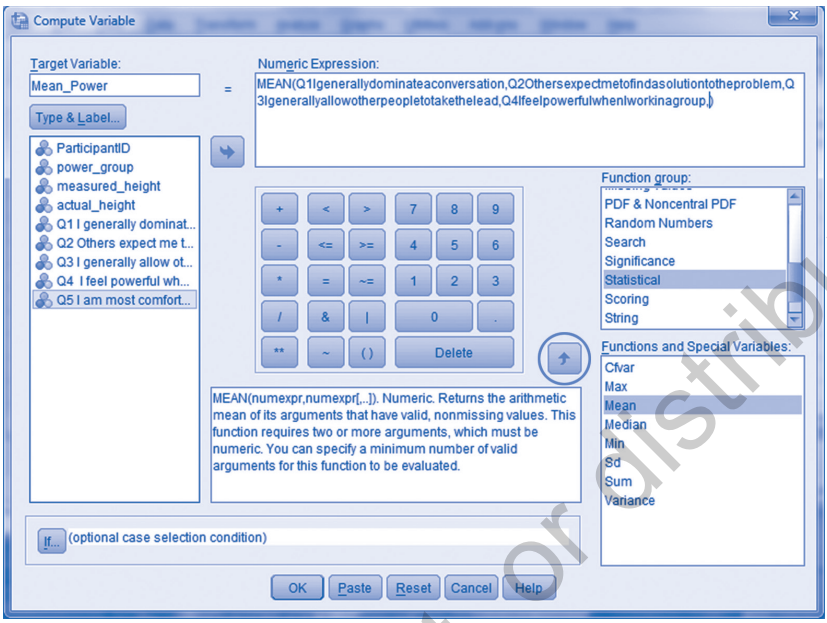
Our students frequently need to calculate an overall score based on responses to several items. Those overall scores might be a simple total of the responses to the group of items or the mean of those responses. You would use the same set of dialogue boxes for those calculations. The only difference is in how you ask SPSS to conduct the calculations.

In the next screenshot, we have asked SPSS to compute the sum of responses to five questions.



You can see that we have named a new *Target Variable* Total_Power. Having done that, we moved the names of each variable into the *Numeric Expression* box and separated those names with plus signs ("+"). Because it is highlighted, you can see that we were ready to move the name of our fifth question into the *Numeric Expression* box at the time that we captured this image. As usual, our final step is to click the *OK* button. That click will produce a little bit of output and will create a new column in our data file called Total_Power.

As you no doubt have noticed, this dialogue box provides us with many tools for performing calculations on our data. In the next screenshot, we will show you the easiest way to compute a mean from several responses.



We used the scroll bar on the right side of the *Function group* box to find the *Statistical* group. When we clicked on that term, a new list appeared in the *Functions and Special Variables* box. We selected *Mean,* and you can see that it is highlighted. We then clicked on the little up arrow that is circled in the screenshot. That click caused *MEAN(,)* to appear in the *Numeric Expression* box. We then moved the variable names for each of our five questions into those parentheses, making sure that each name was separated from the previous name with a comma (","). As you can see, we captured this image as we were ready to move the name for our fifth question into the *Numeric Expression* box. After we move our last variable and type the name of our new variable (Mean_Score), we will click that *OK* button; SPSS will work its magic and create the new variable in our data set.

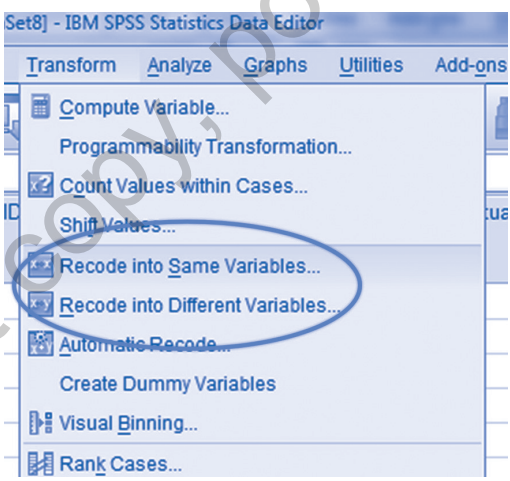
Recoding Variables

We used responses to five questions in the previous examples for how to calculate a total and mean score. Here are the questions we created for those examples:

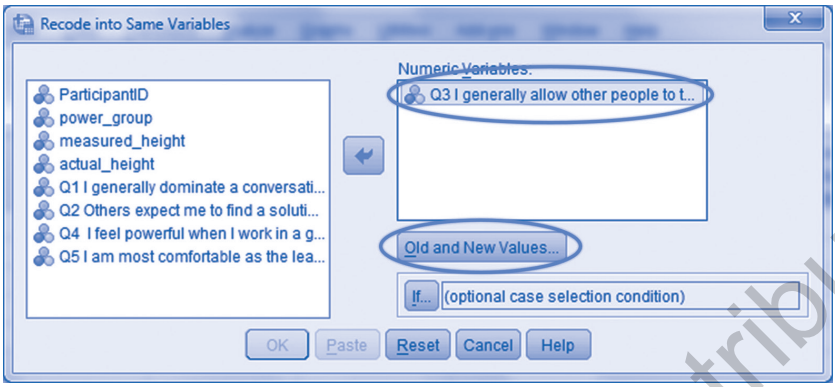
Q1 I generally dominate a conversation.
Q2 Others expect me to find a solution to the problem.
Q3 I generally allow other people to take the lead.
Q4 I feel powerful when I work in a group.
Q5 I am most comfortable as the leader in a group project.

Participants would typically respond to questions like these on a Likert-type scale (with responses indicating the following: 1, strongly disagree; 2, disagree; 3, neither agree nor disagree; 4, agree; and 5, strongly agree). Reread the items with these anchors in mind. Notice that question three (Q3) seems to take a direction that is opposite to the other four questions. An answer of "5" indicates that the respondent believes she is a follower, while answering "5" to the other four questions indicates that she believes she is a leader. As a result, we need to rescore responses to this question so that a "5" indicates the same attitude as it does on the other four questions.

You can easily do this rescoring in SPSS. The first step is to open one of the procedures to recode a variable. As you can see in the next screenshot, you can do that by opening the *Transform* drop-down menu and then clicking on either *Recode into Same Variables* or *Recode into Different Variables*.

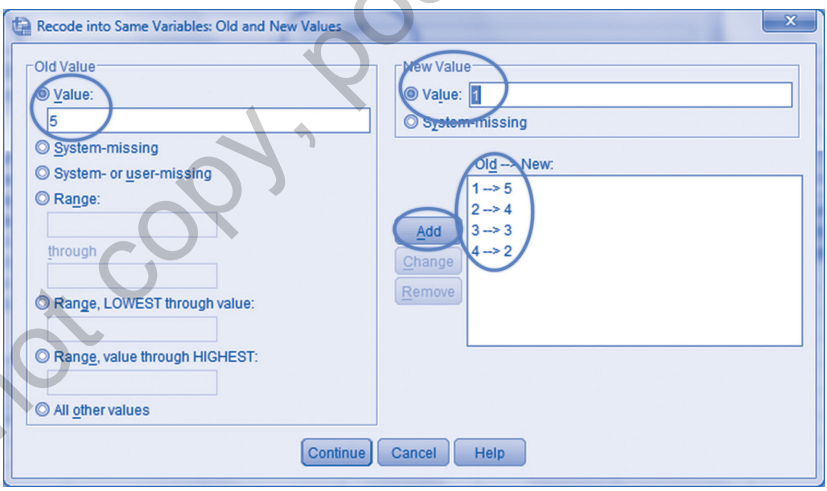


Here, we have circled both choices. There are advantages and disadvantages to each. We suggest you talk with your professor or an experienced researcher about which choice is best for your project. We selected *Recode into Same Variables* for this example, but both procedures work the same way. You will not be surprised to learn that another dialogue box will open when we click on that choice.



We have only a little work to do in this dialogue box. We need to select the variable that we wish to recode and move it over to the *Numeric Variables* box. We have circled this in the screenshot. Notice that you may recode several variables at the same time if you need to. When you have moved those names, you should click on the *Old and New Values* button. Yes, that will open yet one more dialogue box.

In this last dialogue box, we need to tell SPSS what the old values were and what we want each of those values to become.

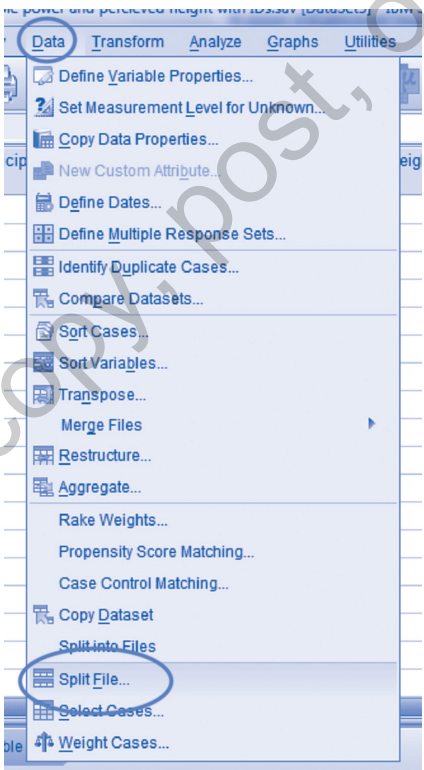


Remember that we are reverse scoring responses to this question, so we want a "5" to become a "1," a "4" to become a "2," and so forth. We have done most of the work needed for that reversal in the above screenshot. Look first at the *Old* → *New* box. You can see that we already have asked that "1" be changed to "5"and "2" be changed to "4." As you anticipate, we did that by typing the existing value in the *Old Value* box and then typing the desire value in the *New Value*

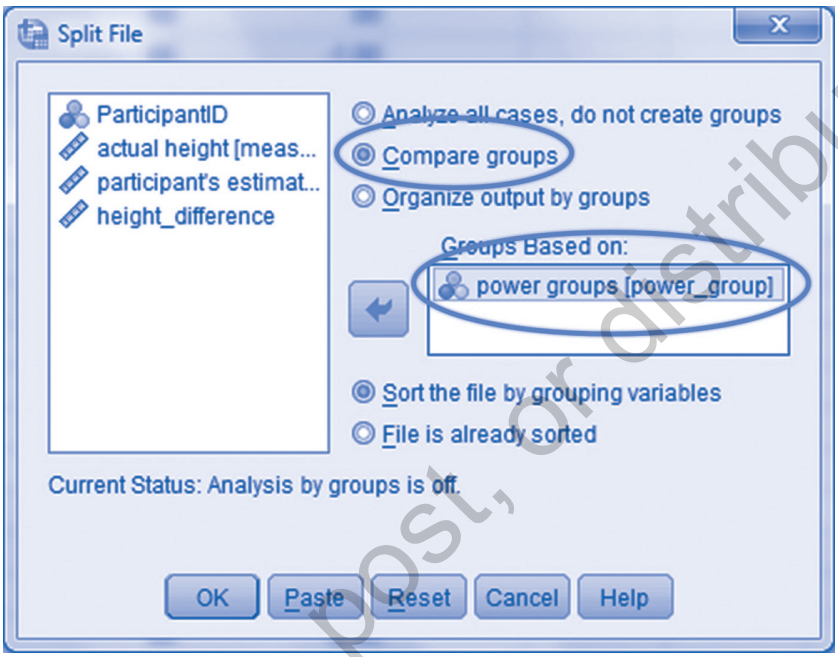
box. You can see that we have typed 5 in the *Old Value* box and 1 in the *New Value* box. When we click the *Add* button, that pair of values will be included in our list and we will click on the *Continue* button. That will take us back to the previous dialogue box, where we click the *OK* button. With that click, SPSS will make the appropriate changes to reverse score this variable.

Conducting Analyses With Only Part of Your Data: Split File and Select Cases

In some of your research, you will need to conduct follow-up analyses (e.g., to examine responses within your groups to see whether they are different or similar). In the example that we are working with in this chapter, we want to calculate the correlation between estimated and actual height separately in the two power groups. In the following screenshot, the *Split File* option (circled) under the *Data* drop-down menu is intuitive.



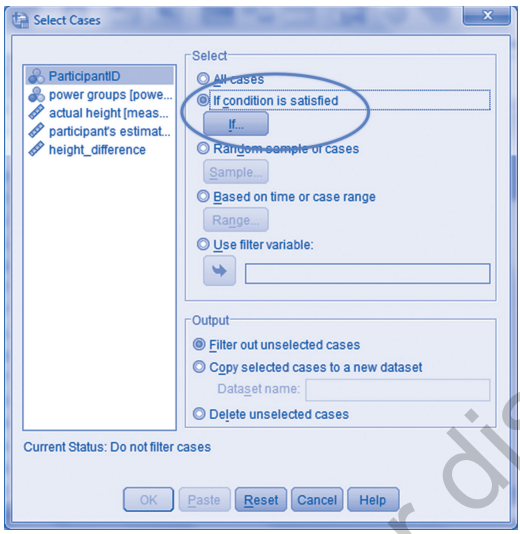
The following window appears. At this point, you must then select either *Compare groups* or *Organize output by groups*. These choices can be seen near the top right of the screenshot.



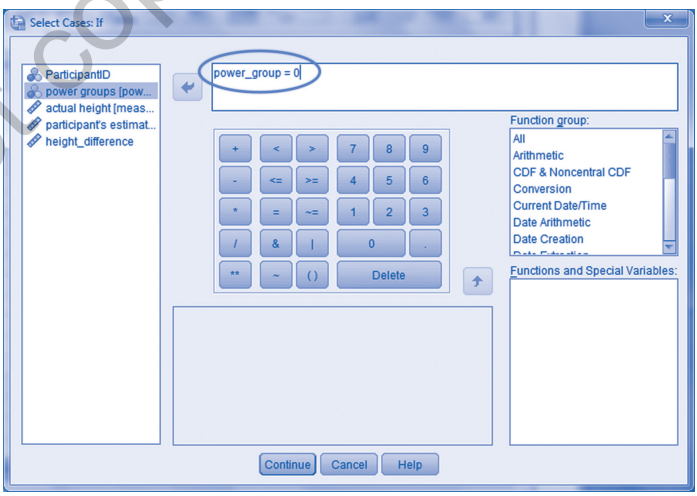
As beginning researchers, your choice here will not make much difference: select *Compare Groups*. Once you have made that selection, you will move the name of your grouping variable into the box labeled *Groups Based on*. As you can see, we selected *Compare groups* and will obtain separate analyses for our two power groups (power_groups). Click the *OK* button to return to the basic SPSS database screen. Later on, if you wish to return to analyses of your complete data file, you return to this procedure (use the *Data/Split File* drop-down menu again), and this time select *Analyze all cases.*

Finally, there might be occasions when you need to eliminate some data from a file. For example, after entering all of your data, you might learn that one of the **experimental groups** obtained knowledge about your manipulation or you might find some outliers that should be eliminated from a second analysis. As a result, you might conduct a second analysis to make sure that the outliers did not change your results. You open the procedure by selecting *Data/Select Cases*.

Choosing *Select Cases* opens the following dialogue box. To select cases, you only need to click on the *If condition is satisfied* button (see the circled portion of the next screenshot).



That click will open a second dialogue box illustrated in the next screenshot. This box looks very much like the *Compute variable* dialogue box that we described earlier in this chapter. To choose certain cases, you will choose a certain variable and value of that variable and enter that information in the text box at the top of the dialogue (see the circled portion of the screenshot). When that function is true, those cases will be included in further analyses. In our example, we entered "power_group = 0." If we click on the *Continue* button and then the *OK* button in the *Select Cases: If* dialogue box below, all further analyses would include only individuals from the low-power group. Again, we can later return to the *Select Cases* dialogue and select *All cases* and then analyze *all* of our data.



Summary

Likely, you will not need every tool that we have described in this chapter for every research project, but we hope that when you do need one of them, you will be ready with a useful bag of tricks. We are confident that if you follow our advice in this chapter, you will make your data-collecting life easier, even in your very first independent research project. You will no doubt need some practice with these tools before you become an expert, so practice away. You can find other tools that could be useful to you in the menus that we introduced here. Explore to your heart's content!