

2: Entering Data

Objectives

- ◆ Understand the logic of data files
- ◆ Create data files and enter data
- ◆ Insert cases and variables
- ◆ Merge data files
- ◆ Read data into SPSS from other sources

The Logic of Data Files

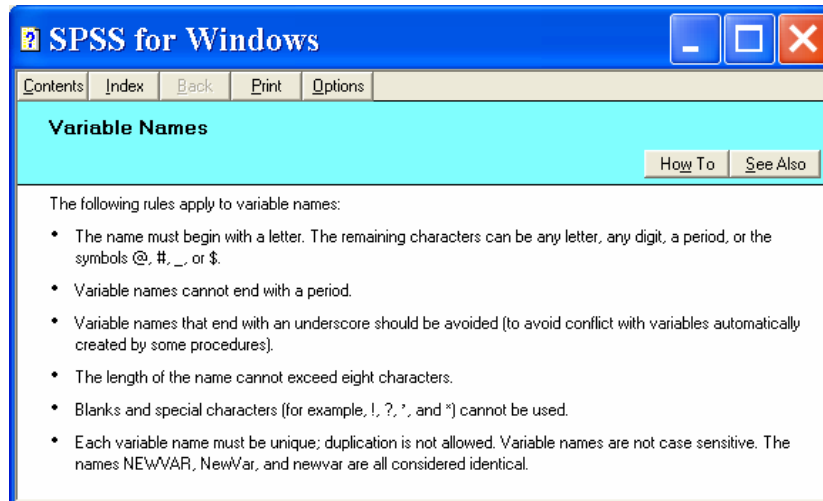
Each row typically represents the data from 1 case, whether that be a person, animal, or object. Each column represents a different variable. A cell refers to the juncture of a specific row and column. For example, the first empty cell in the right hand corner would include the data for case 1, variable 1.

Entering Data

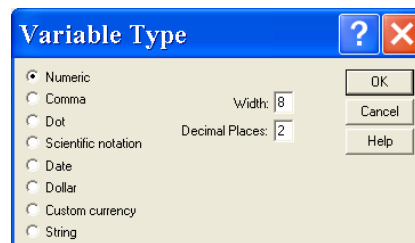
- ✓ Open SPSS and follow along as you read this description.

To enter data, you could simply begin typing information into each cell. If you did so, SPSS would give each column a generic label such as var00001. Clearly this is not desirable because you would have no way of identifying what var00001 meant later on, unless you have a superior memory. Instead, we want to specify names for our variables. To do this, you can either double left click on any column head, this will automatically take you to the Variable View. Alternatively, you can simply click on Variable View on the bottom left hand corner of your screen.

The first column of variable view is **Name**. The default length for variable names is 8 characters. It is very important that you give your variables names that make intuitive sense. This is easy if words under 8 characters come to mind, if not you will need to use thoughtful abbreviations. For example, if I had depression data that was collected at intake, and 1 month, 6 months, and 1 year post intervention, I would name those variables depress0 or depresin (i.e., in for intake), depress1, depress6, and depres12. SPSS also has preferences for variable names. For example, a number cannot begin a variable name (e.g., 12depres would not be a valid name). Error messages will appear if you have selected a name that is “illegal” to SPSS. The rules for variable names appear below. They can be found by typing “variable names” in the **Index** option under **Help** and then selecting rules from the list that appears.



Next, you can select the **Type** of variable. Left click on the empty cell, then left click on the gray box that appears. The dialog box to the right appears. The most commonly used types of data include numeric, date, and string. For numeric data, width and decimal places refer to the number of characters and decimal places that will be displayed in the



Data Editor window. If you entered a value with 3 decimal places, SPSS would save that value, but would only display the value to 2 decimal places.

String variables are those that consist of text. For example, you could type male and female if gender were a variable of interest. It is important to note that SPSS is case sensitive meaning that “female” and “Female” would not be viewed as the same category. Misspellings are also problematic with string data (e.g., “femal” would not be recognized as the intended “female”). For these reasons, it is advantageous to use numbers to represent common categories, and then supply names for those levels as discussed below.

✓ Click on **Date** to see the variety of available formats. Then, click **Cancel**.

The next columns are for **Width** and **Decimals**. You could have set this while specifying your variable type, or you can specify them in these columns. The default for width is 8 characters and the default for decimals is 2. To change this, left click the cell, and up and down arrows will appear, as illustrated below. Left click the up arrow if you want to increase the number, click the down arrow to decrease the value. Alternatively, you can simply type the desired value in the cell.

| | Name | Type | Width | | Missing |
|---|----------|---------|-------|--------------------------------|---------|
| 1 | depresin | Numeric | 8 | Click here to change the width | None |

The next column is **Label**. This is a very nice feature that allows you to provide more information about the variable than you could fit in the 8 character variable name. For example, I could type “Depression assessed at intake” for the example used above. When you hold your cursor over a variable name in the Data View, the full label will appear. (Gee, you must have read a manual. How did you know that??) This is very useful when you need a quick reminder. An example of this feature is below.

The screenshot shows the SPSS Data Editor window titled "Untitled - SPSS Data Editor". The menu bar includes File, Edit, View, Data, Transform, Analyze, Graphs, and Utilities. The toolbar contains various icons for file operations and data manipulation. The data view shows a table with columns: depressin, depress1, depress6, depres12, var, var, and var. The first row of data has values 30.000, 25.0, and a tooltip for the 'depress6' variable label: "depression assessed at 1 month post intervention".

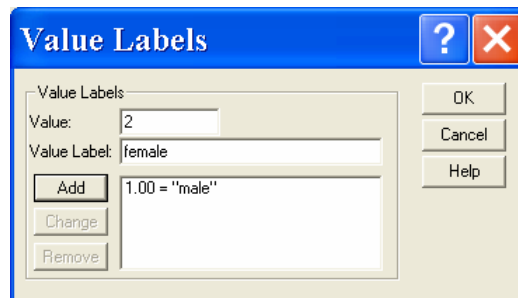
| | depressin | depress1 | depress6 | depres12 | var | var | var |
|---|-----------|----------|--|----------|-----|-----|-----|
| 1 | 30.000 | 25.0 | depression assessed at 1 month post intervention | | | | |
| 2 | 32.000 | 30.00 | 30.00 | 28.00 | | | |
| 3 | 35.000 | 35.00 | 35.00 | 40.00 | | | |
| 4 | 45.000 | 42.00 | 40.00 | 35.00 | | | |

Since labels are so much more detailed than variable names, we can specify that SPSS label variables this way in dialog boxes and output. Let’s do this.

- ✓ Click **Edit/Options/Output Labels** and specify labels for each of the options. Then click **Ok**.

The next column is **Values**. This allows you to assign variable labels. You will typically use this option for categorical variables. For example, we may want the number 1 to represent males and the number 2 to represent females when we enter data on gender. Let’s try this.

- ✓ Type gender in the first **Name** column.
- ✓ Scroll over to the **Values** column and left click. Then, left click on the gray box that appears on the right hand side of the cell. The Value Labels dialog box will appear.



- ✓ Type the numeric value where it says **Value**, then type the **Value Label** or text to explain what it means. Click **Add**. Do this for males and females. When you are done, click **Ok**.

Of the remaining columns, you are most likely to use **Align**, which allows you to specify how the data will appear in the cells. Your choices are left justified, right justified, or centered. This is simply a matter of personal preference.

After you have completed specifying your variables, you can click on Data View and begin entering your data. Put your cursor on the cell in which you want to enter data. Type the value. If you hit Enter the cursor will move to the cell under the one you just filled. You can also use the arrow keys to move to the next cell in any given direction. Typically, you will either enter all of the values in one column by going down or you will enter all of the variables in a row going from left to right.

Let's try this together.


- ✓ On your disk, **Open** *depression scores.sav*. SPSS will ask you if you want to save the file with the gender variable. Click **No**. Then, SPSS will open *depression scores*. As you can see, variables have been named and labeled, but the data have not been entered.
- ✓ Enter the following data that appear on the next page in the Variable View window. Pay attention to your own preferences for data entry (i.e., using the arrows or enter, going down or across). Notice that there is no subject 10.
- ✓ When you are done, click **Save**, but do not close the file. We will continue to use it as an example.

| ID | depressin | depress1 | depress6 | depres12 |
|----|-----------|----------|----------|----------|
| 1 | 30.000 | 25.00 | 23.00 | 20.00 |
| 2 | 32.000 | 30.00 | 30.00 | 28.00 |
| 3 | 35.000 | 35.00 | 35.00 | 40.00 |
| 4 | 45.000 | 42.00 | 40.00 | 35.00 |
| 5 | 45.000 | 45.00 | 38.00 | 40.00 |
| 6 | 25.000 | 25.00 | 20.00 | 20.00 |
| 7 | 60.000 | 45.00 | 30.00 | 40.00 |
| 8 | 55.000 | 50.00 | 40.00 | 35.00 |
| 9 | 40.000 | 40.00 | 35.00 | 30.00 |
| 11 | 37.000 | 30.00 | 25.00 | 20.00 |
| 12 | 30.000 | 25.00 | 22.00 | 20.00 |

Inserting a Variable

After specifying the types of variables for the depression data, I realized I forgot to include a column for ID number. Typically, I like ID to be the first variable in my data file. I can add this in one of two ways.

1. In Variable View, highlight the first row and then click **Insert Variable** on the **Data** menu. This will place a new variable before the selected variable.


2. In Data View, highlight the first variable column and then click the **Insert Variable** icon . This will also place a new variable column at the beginning of the file.

- ✓ Use one of the approaches above to Insert the new variable at the beginning of the file.
- ✓ Name the variable ID, and label it as participant identification number.
- ✓ Enter the ID data that appeared on the previous page.
- ✓ Click **Save**, and leave the file open.

Inserting a Case

As you can see, the data for ID 10 is missing. I found the missing data and want to enter it in the file. I'd like my data to be in order by ID number, so I want to insert a case between the person with ID 9 and ID 11. To do so, I can highlight the row for the case with ID 11, and either:

1. click on **Insert Case** on the **Data** menu or

2. click on the **Insert Case** icon . In either case, a blank row will appear before the highlighted case. Try it yourself.

- ✓ Insert a case for ID 10 using one of the above approaches.
- ✓ Enter the following data: 10, 38, 35, 38, 38 for ID, depressin, depress1, depress6, and depress12 respectively.
- ✓ Check the accuracy of your data entry, then click **Save**.

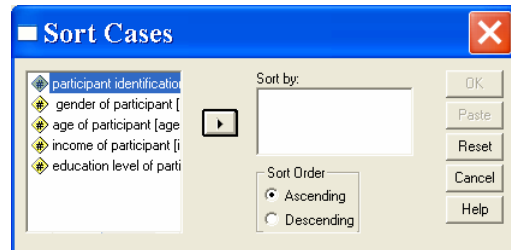
Merging Files

Adding Cases. Sometimes data that are related may be in different files that you would like to combine or merge. For example, in a research methods class, every student may collect and then enter data in their own data file. Then, the instructor might want to put all of their data into one file that includes more cases for data analysis. In this case, each file contains the same variables but different cases. To combine these files, have one of the data files open, then left click on **Merge Files** on the **Data** menu and select **Add Cases**. Then specify the file from which the new data will come and click **Open**. A dialog box will appear showing you which variables will appear in the new file. View it, and if all seems in order, click OK. The two files will be merged. This is fairly simple. See if you can do it yourself in Exercise 3 at the end of this chapter.

Adding Variables. In other cases, you might have different data on the same cases or participants in different files. For example, I may have recorded the demographic information from the participants in my depression study in one file and the depression data in another file. I may want to put them together because I'd like to see if demographic variables, like socioeconomic status or gender are related to depression. In this case, you need to be sure the variables on the same participants end up in the correct row, that is, you want to match the cases. In this case, we will use ID to match cases. SPSS requires that the files you merge be in ascending order by the matching variable. So, in both files, ID must start at 1. You can set this up by sorting cases as discussed above. Then, make sure one of the files is open. Since this procedure is more complicated, let's try this one together.

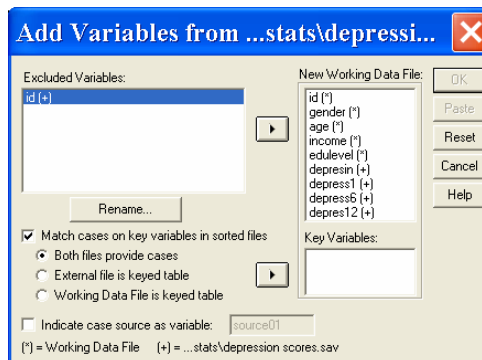
- ✓ Open *depression scores.sav* from your disk (this is the data that you just entered).
- ✓ Check to see if the cases are in ascending order by ID. They should be since we just entered them that way.
- ✓ Now, open *depression demographics.sav*. These data are not in order by ID. To fix this, click **Sort Cases** under the **Data** menu.

- ✓ In the dialog box, select participant identification number and move it into the **Sort by** box by clicking the arrow. Make sure **Ascending** is selected for **Sort Order**. Then click **Ok**.



- ✓ While the demographic file is still open, click on **Merge Files** in the **Data** menu, and select **Add Variables**.
- ✓ The next dialog box will ask you to indicate which file the new variables are coming from. Select *depression scores.sav* and click **Ok**. The following dialog box will appear.

- ✓ Select **Match cases on key variable in sorted files**, then highlight **id** under excluded variables, then click the arrow to move **id** to **Key Variables**. Click **Ok**.



(I have had trouble doing things this way in the past, but succeeded by *not* selecting **Match cases on key variable in sorted files** and just clicking OK. I don't know why, but it worked for me.)

- ✓ A dialog box will appear, reminding you that the files must be sorted. Click **Ok**, and the files will be merged. You may want to do a **Save As** and give the merged file a new name like *depression complete.sav* to help you remember what is in it.

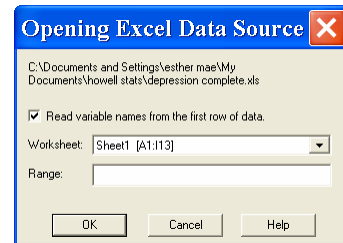
Reading Data In From Other Sources

SPSS can also recognize data from several other sources. For example, you can open data from Microsoft EXCEL in SPSS or you can get SPSS to read data entered in a text file. This is an attractive option, particularly if you do not have your own version of SPSS. It allows you to enter data in other more common programs, save them to disk, and simply open them when you have the opportunity to use a PC that has SPSS on it. Let's try some examples.

Opening data from EXCEL. The complete depression data is also on your disk in a file named *complete depression.xls* (xls is the extension for Excel data files). Take a moment to open this file in Excel and look it over. You will see it looks just like the file we just created except that the variable names are different because they are longer. Since we know SPSS only allows variable names with 8 characters (in older versions), we'll need to change these names. If we don't SPSS will truncate them to 8 characters. This is a problem in this case because all 4 depression variables start with the same 8 characters. SPSS does not allow multiple variables to have the same name, so it would give the last 3 depression variables generic names like var7, var8, and var9 because they are the 7th, 8th, and 9th variables in the file. This is not desirable since we may not remember what they are supposed to represent. Essentially, if you plan to read Excel data into SPSS, you should follow all of the variable name guidelines specified by SPSS (e.g., character length, no numbers beginning a name, etc). Specific instructions follow.

- ✓ **Open** *complete depression.xls*.
- ✓ Rename the variables in Excel to include eight characters. To do so, simply put your cursor in the cell containing the variable name and start typing, then hit the right arrow to move to the next cell. When you are done, **Save** your changes and close the file because SPSS cannot read the file if it is open in another program.
- ✓ Open SPSS and select **Read Text Data** from the **File** menu.
- ✓ A dialog box will appear. Under **Files of type**, select **Excel**. Under **Look in** select your disk (CD?) drive. *Depression complete.xls* should appear. Select it and click **Open**. The following dialog box will appear.

- ✓ Select **Read variable names from the first row of data**, because that is where the names appear in the Excel file. Then, click **Ok**. Check out your new file in SPSS. How does it look? There is no need to save this file since it contains the same information as *depression complete.sav*.

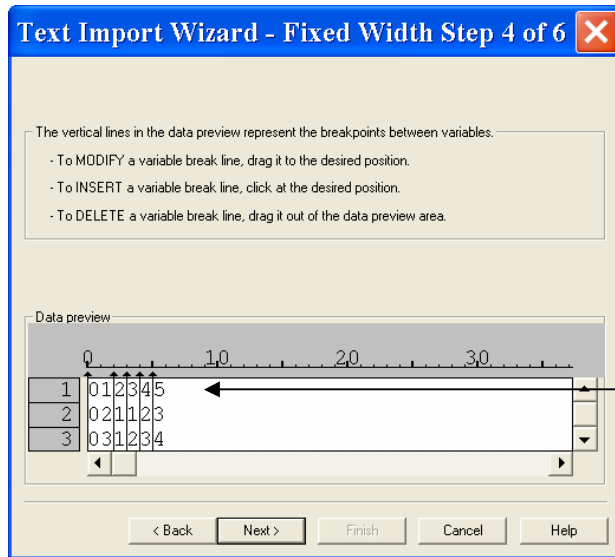


The downside is the new data file does not include variable labels or values, so you would need to add them. You should also make sure that SPSS has identified the variables as the correct type.

Text Data. Now, let's try an example with text data. A text data file can be created in any word processing program or in Notepad or any other text editor. Just be sure to save the file with the .txt or .dat file extension. (I use a .dat extension for the data files on their CD. SPSS can recognize text data in several formats. Let's begin with the simplest example. I have collected data from 11 people and typed it in the following format (this is a sample, not the whole file).

| | |
|--------|---|
| 012345 | The first two digits are the ID number. The next digit is |
| 021123 | gender. Digits 4, 5, and 6 are the responses to the first 3 |
| 031234 | questions on a survey. No characters or spaces separate the |
| 042345 | variables. The data are on your disk in <i>simpletextdata.txt</i> |

- ✓ **Open SPSS.**
- ✓ In the SPSS **File** menu, click **Read text data**.
- ✓ Select *simpletextdata* under **Files of type Text** and click **Open**.
- ✓ In the next dialog box, click **No** for “Does your text file have a predefined format” and click **Next**.
- ✓ In the next dialog box, select **Fixed width** under “How are your variables arranged,” then select **No** for “Are variable names included in the top of your file.” Then click **Next**.
- ✓ In the next dialog box, indicate that the data starts on line **1**, **1** line represents a case, and you want to import **all cases**, then click **Next**. The following dialog box will appear. We need to tell SPSS where to insert breaks for variables.



- ✓ Hold your cursor between the numbers where a break should be inserted. An arrow will appear. Add a break after the 2nd, 3rd, 4th, and 5th columns. When you are done, click **Next**.

- ✓ The next dialog box will show you a draft of what your new data file will look like. Notice, the variables will have generic names like v1, v2, etc. Then click **Next**.
- ✓ At the next dialog box, you can click **Finish** and your new data file will appear. You could then specify variable names, types, and labels as illustrated above.

Let's take one more example. This is based on the same data, but this time the text file is tab delimited (a tab was inserted between each variable) and has variable names at the top. Below, is an example of the first two lines of from this text file.

| ID | Gender | Q1 | Q2 | Q3 |
|----|--------|----|----|----|
| 01 | 2 | 3 | 4 | 5 |

- ✓ On the **File** menu, click **Read text data**.
- ✓ Select *tabtextdata* under **Files of type Text**, then click **Open**.
- ✓ On the next dialog box, you will see a draft that shows a box between each variable to represent the tabs. Are they in the right place? Select **No** for predefined format and then click **Next**.
- ✓ Select **Delimited** for file arrangement and **Yes** for variable names at the top of the file, then click **Next**.
- ✓ In the next dialog box, indicate that the data starts on line **2**, Each line represents a case, and you want to import **all cases**, then click **Next**.
- ✓ In the next dialog box, check **Tab** as the type of delimiter and then click **Next**.
- ✓ You will see a draft of your data file. Review it, and then click **Next**.

✓ Click **Finish** at the next dialog box and your new data file will appear.

One difference between these two examples, is the second included the variable names at the top of the file. This, in my opinion, is the better approach because it reduces the chance of making mistakes later in the process. (Unless you're the person who has to type in all those damn data files.)

This chapter included information about entering data and opening files of various types. This is an important part of the process because data entry errors contribute to inaccurate results. Further, good variable names and labels allow you to perform subsequent analyses more efficiently. Completing the following exercises will help you internalize these processes.

Exercises

1. The data from Appendix D are on the disk in a file named *RxTime.sav*. These are the data from Table 3.1 in your text. Open this file and label the variables and values as described in the Howell text book on page.... Save the changes when you are done because we will use this file in subsequent examples and exercises
2. Read the data from Fig3-2.dat on the website. These are the data on intrusive thoughts shown in Figure 3.2 of your text. These are raw data with variable names in the first line.
3. Review the following data. Then, create your own data file and enter the data. Be sure to include variable and value labels. Then open *exercise2.2.sav* on the disk which includes the same data. Note the similarities and differences between your file and the file on disk. Which do you prefer? Why?

| Age | Gender | Average Hours of Sleep | Number of Classes Missed | Grade in Course |
|-----|--------|------------------------|--------------------------|-----------------|
| 18 | Male | Seven | 0 | A |
| 18 | Female | Four | 1 | C |
| 17 | Female | Six | 2 | B |
| 19 | Female | Ten | 5 | F |
| 20 | Male | Eight | 2 | B |
| 21 | Female | Seven and a half | 3 | C |
| 23 | Male | Nine | 1 | B |
| 22 | Male | Eight | 2 | A |
| 18 | Male | Six | 3 | D |

4. Merge the following files from the disk using the add cases option: *merge1.sav* and *merge2.sav*.
5. Read the following text data file into SPSS: *textdataexercise.txt*. Be sure to open the text file and notice the format before you proceed.
6. Read *readexceleexercise.xls* into SPSS. Note any problems that arise and how you solved them.