

Chapter 2

SPSS Files

2.1 Introduction

Data files play a key role in SPSS. Naturally, the data in these files can be obtained in various ways. To keep things simple, we use data obtained through the evaluations forms that were handed to students after a statistics course was taken. The data of this particular study can be stored in three different ways:

1. The data might have already been stored in an SPSS data file. The file needs to be opened only for modifications and/or for statistical analyses (see Section 2.2).
2. The data may not be digitized; the information may be available only on the filled-out (paper) forms. Thus, a data file must be created (see Section 2.3).
3. The results may be entered into a digital file using a spreadsheet program, like the popular program EXCEL. This file must be converted to an SPSS data file (see Section 2.4).

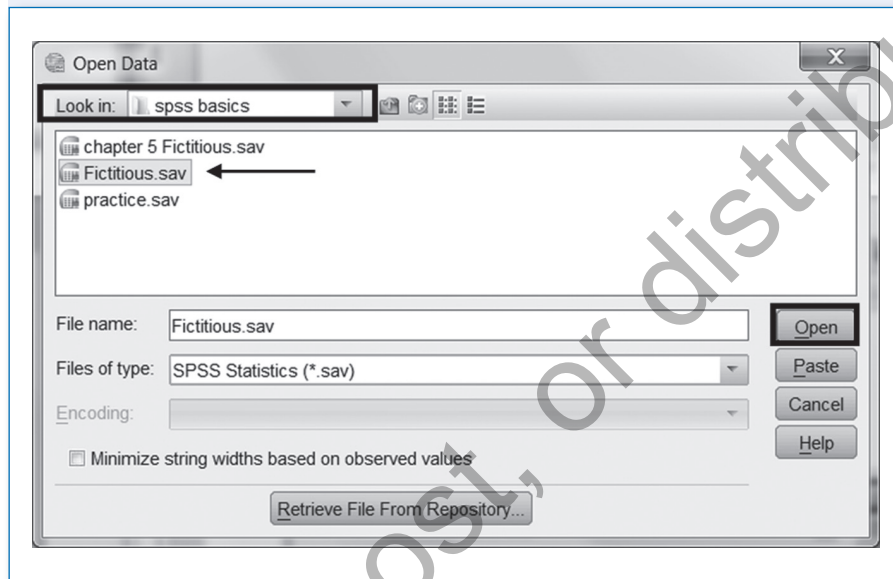
2.2 Open SPSS Files



- Open SPSS using the SPSS icon on your desktop or via Start, Programs ►; click on the IBM SPSS folder; and then on the SPSS program.
- If the SPSS welcome window appears, close it by clicking **Cancel**.
- After the empty SPSS data window appears, go to the menu bar and select:
File
Open ► and click **Data . . .**

The dialog window “Open Data” will appear (see Figure 2.1).

Figure 2.1 Dialog Window “Open Data”



- Select the correct folder in “Look in:” (**SPSS Basics** in this case) and open the file *Fictitious.sav* directly with a double-click on the name or by selecting (one click) and clicking **Open** (see Figure 2.1).

If the output window appears instead of the data window, then

- Click **Window** in the menu bar and click on *Fictitious.sav* to activate the file.

Similarly, the output files can be opened, and they contain the results from commands that have been previously executed. Open the following output file using the instructions in the following box:



- From the menu bar, select **File**
Open ▶ and click **Output . . .**

The accompanying dialog window will appear.

- If necessary, select the right drive and the **SPSS Basics** folder.
- Double-click to open the file *output.spv*.

This file shows the age distribution of 37 students in total. They are aged between 18 and 37 years. Four students are 18 years old, which is 10.8% of the total (37). Two students are 37 years old, which is 5.4% of the total.

2.3 Create and Modify Data Files ▲

Research material supplied as filled-out registration forms on paper must first be converted to digital data sheets. However, before any data can be entered into such a data sheet, variables must be created. This is called *data definition*. This means that the questions from the survey or questionnaire are to be converted to variables, which have to be named, and eventually, more information about the variable must be added (labeling). After all questions have been converted to variables with a name (and a label), you may begin to enter the scores (values) related to those variables into the spreadsheet. In the following sections, we show how this is done.

2.3.1 Define and Create Variables

After starting the program, SPSS opens an empty data window by default. However, in the current situation, the file *fictitious.sav* is open and active (see Section 2.2). To open an empty data window now, follow the instructions below:



- Select from the menu bar **File**
New ▶ Data

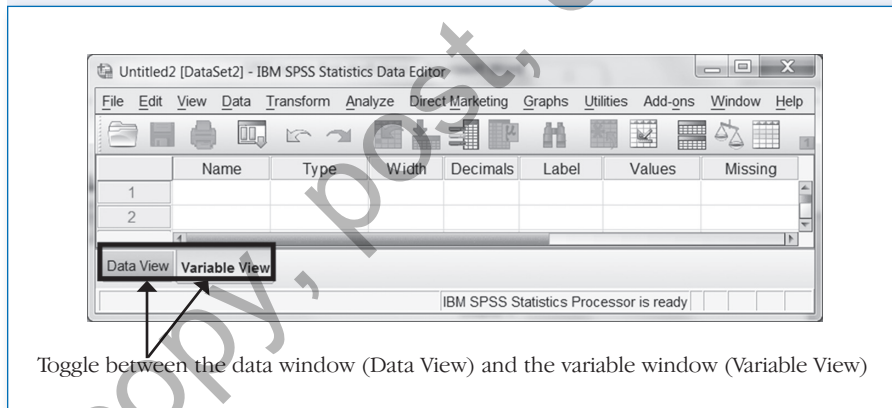
The data view window (Figure 1.2) or the variable window (Figure 2.2) opens, depending on the way SPSS was used in the previous session. If the data view window is open, follow the instruction below first:



From the bottom bar, select **Variable View**.

The data window now has changed to *variable view*, where the information is listed on all variables; see Figure 2.2.

Figure 2.2 An Empty Data Window



Toggle between the data window (Data View) and the variable window (Variable View)

Note: The Variable View window is active.

Before entering the data, it is a standard procedure to first indicate (define) which variables are used. We explain this in the following example.

Example

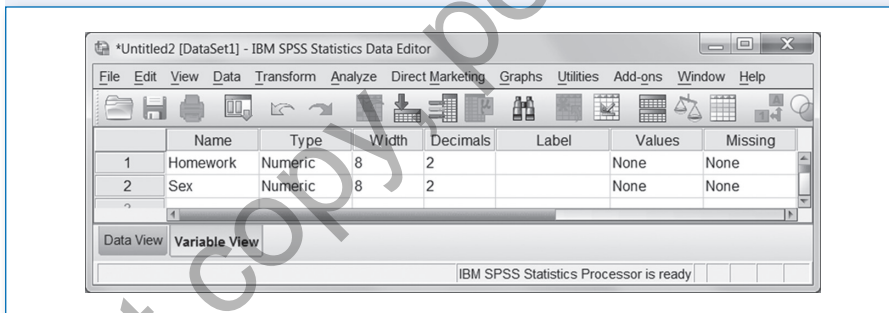
Suppose you want to define two variables related to the survey questions (1) “How much time do you spend per month doing statistics homework?” and (2) “Is the student male or female?” The names for these two variables are chosen to be (1) “homework” and (2) “sex.”



- Double-click on the empty cell beneath **Name** (see Figure 2.2).
- Subsequently, enter the name of the variable (*homework*) and press [Enter] on the keyboard.
- Double-click on the empty cell beneath *homework* and fill in *sex* and again press [Enter].

The results are shown in Figure 2.3. Note that we use only words for the variable names. The variable names in SPSS always begin with a letter (a number as first character is not allowed) and may be succeeded by a maximum of 63 letters or numbers. Furthermore, SPSS does not allow for spaces/blanks or symbols (except an underscore [_]) in a variable name. So “home work,” “home+work,” and “1homework” cannot be used. Once used, you cannot repeat the same name for yet another variable, so each name must be unique.

Figure 2.3 Data Window With Variables “homework” and “sex” Defined



Similarly, other features of variables can be changed (if needed):

- **Type** This is the variable type: numeric (numbers) or—less frequent—string (text); the default type is numeric.
- **Width** The maximum number of characters that can be shown in the Data View window; 8 is the default width.
- **Decimals** The maximum number of decimals that can be shown in the Data View window; 2 is the default number of decimals.

As illustrated in Figure 2.3, SPSS enters the default settings for type = numeric, width = 8, and decimals = 2. These settings may be left unchanged here.

2.3.2 Add Variable Labels

It is common practice to provide a more detailed description of the variables if the variable name itself is not enough. This can be done by using the **Label** option in the Variable View window (see the example below).

Example

The variable that denotes the average time spent by a student on statistics homework (in minutes) has been defined as “homework” in the column **Name** in the data window. The **Label** column indicates the time spent at home in hours (Figure 2.4).

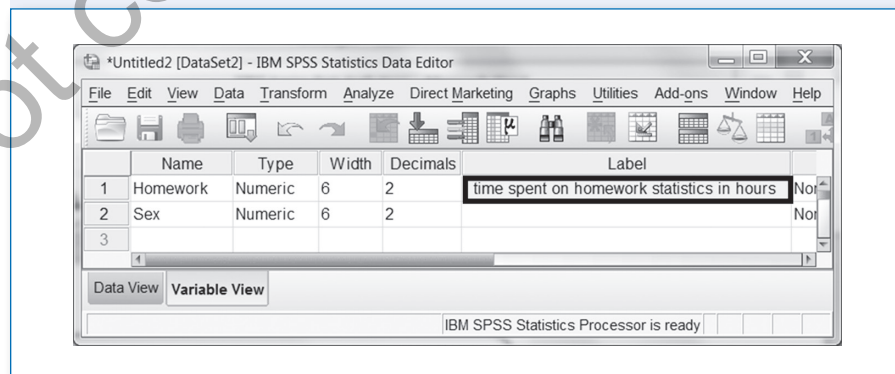
To add labels to variables, follow these instructions:

- Select the first empty cell beneath **Label** with a double-click.

You may enter a more detailed description of the variable in up to 255 characters here: *time spent on homework statistics in hours*.

- Enter our detailed definition and then press [Enter].

Figure 2.4 Add Labels to Variables



2.3.3 Add Value Labels

Besides variable labels, the codes or categories of the variable are also often labeled to avoid confusion.

Example

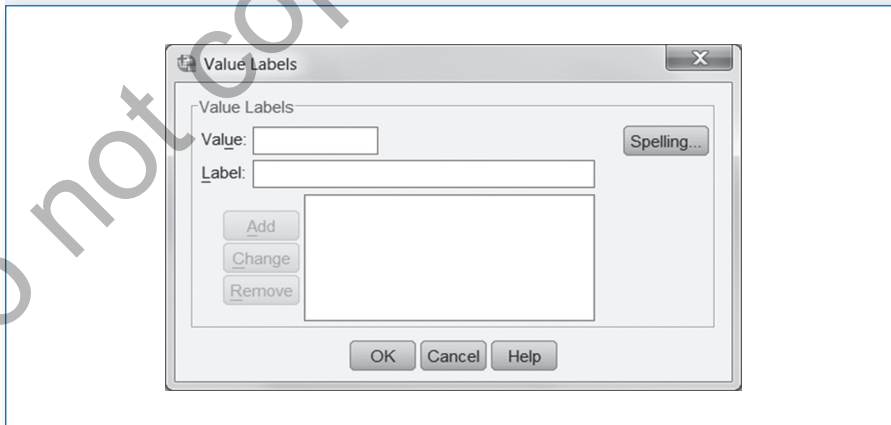
In the data view window, there is a variable named “sex.” This variable, of course, has two categories: (1) male = 0 and (2) female = 1. Labeling the categories as 0 and 1 is useful, given that the information will be displayed in any relevant SPSS output. For example, when viewing a frequency distribution of the variable “sex,” displaying 0 and 1 is not enough. So we have to add the appropriate names to the codes to remove the uncertainty of the meaning of 0 and 1.



- Select **None** in the second cell beneath **Values** (this is a cell from the row marked as “sex”).
- Click on the icon to the right of **None**.

The dialog window “Value Labels” will now appear (see Figure 2.5).

Figure 2.5 Dialog Window “Value Labels”

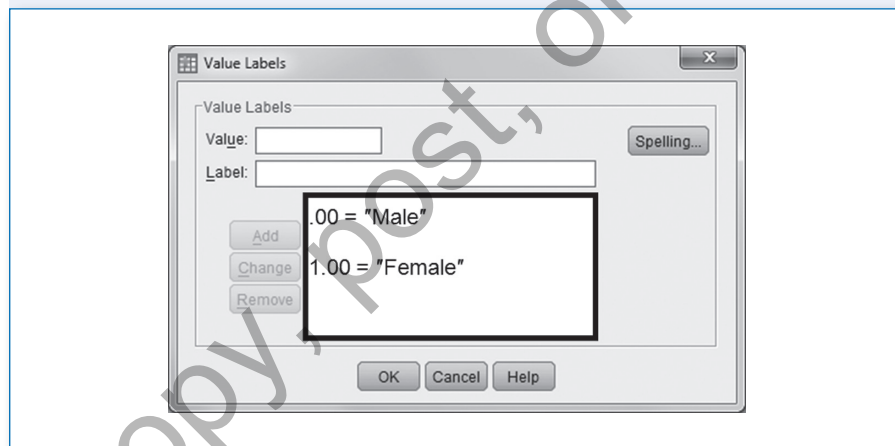




- Enter the number 0 into the empty box right of **Value**. Then, select the box that follows **Label** and enter *male*.
- Click **Add** to label 0 = male.
- Now, enter 1 in the **Value** box and *female* in the **Label** box.
- And again click on **Add** to add 1 = female.

Check whether you have followed the instructions correctly with Figure 2.6 as a reference.

Figure 2.6 Dialog Window “Value Labels” With Labels Added



Click on **OK** to finish (“sex” now has the correct value labels).

2.3.4 Define Missing Values

In research, it often occurs that one or more categories of a variable need to be excluded from analysis. Take, for instance, categories such as “not filled in,” “I don’t know,” or extremely low or high scores (so-called outliers). You

can exclude such categories in SPSS by listing these categories as missing (invalid) values. These values are then not included in any analysis. To list the invalid scores, the dialog window “Missing Values” is used (see Figure 2.7).

Example

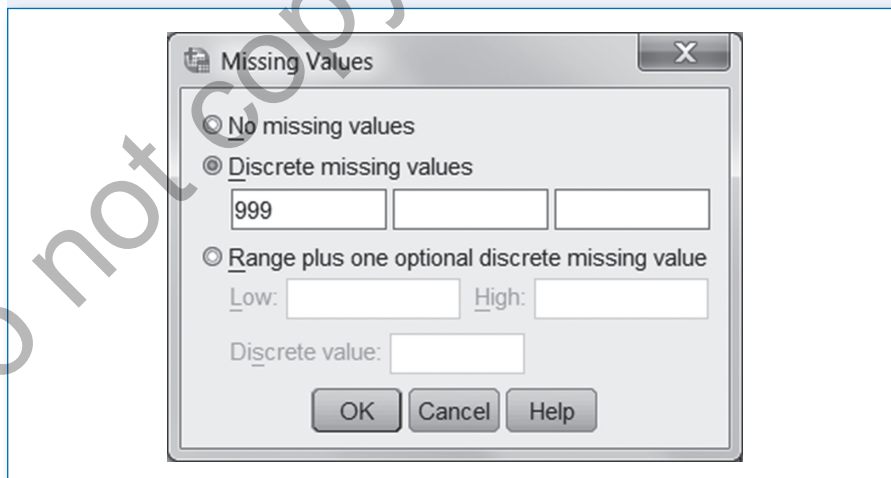
In case of the variable “homework,” where respondents were asked to list the time they spent on statistics homework, a number of them did not answer, maybe because they did not keep track of how much time they spent doing homework or they simply could not remember. For this group of students, we use the value 999, a value that falls way out of the range of possible answers to that question, so it can be easily detected.



- Select **None** in the first cell beneath **Missing** in the Variable View window (this row is named as “homework”).
- Then, click to the right of **None**.

The “Missing Values” dialog window will appear; see Figure 2.7.

Figure 2.7 Dialog Window “Missing Values”

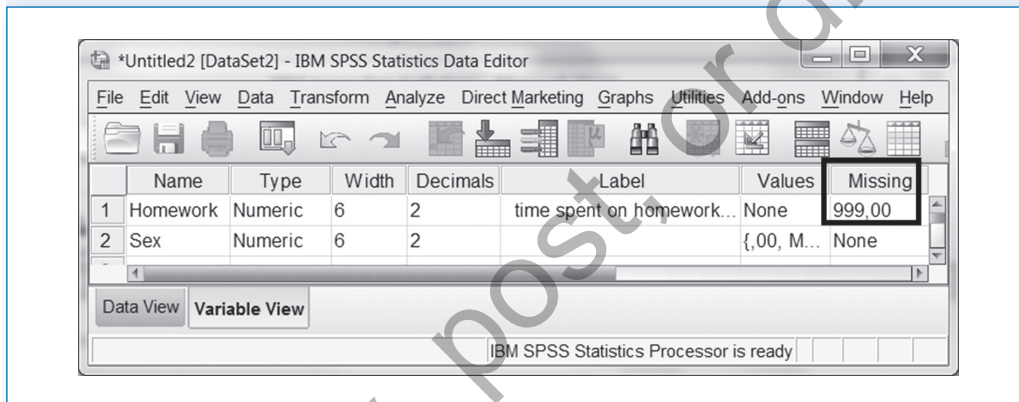




- Select **Discrete missing values**, click on the first empty box below, and enter the numbers 999 (see Figure 2.7).
- Click **OK** to confirm.

If you followed all instructions correctly, the variable view window will look like the one (see the last column [Missing]) in Figure 2.8.

Figure 2.8 Data Window: Variable View (active)



Note: The variables “homework” and “sex” with 999 is defined as “missing” in “homework.”

The information regarding the two variables has now been entered into the file. In practice, of course, a lot more variables must be defined this way. As this is a fairly time-consuming process, we will not pursue this any further. In the next section, a full-fledged file, including all variables of the statistics course evaluation study, is used.

2.3.5 Enter Data

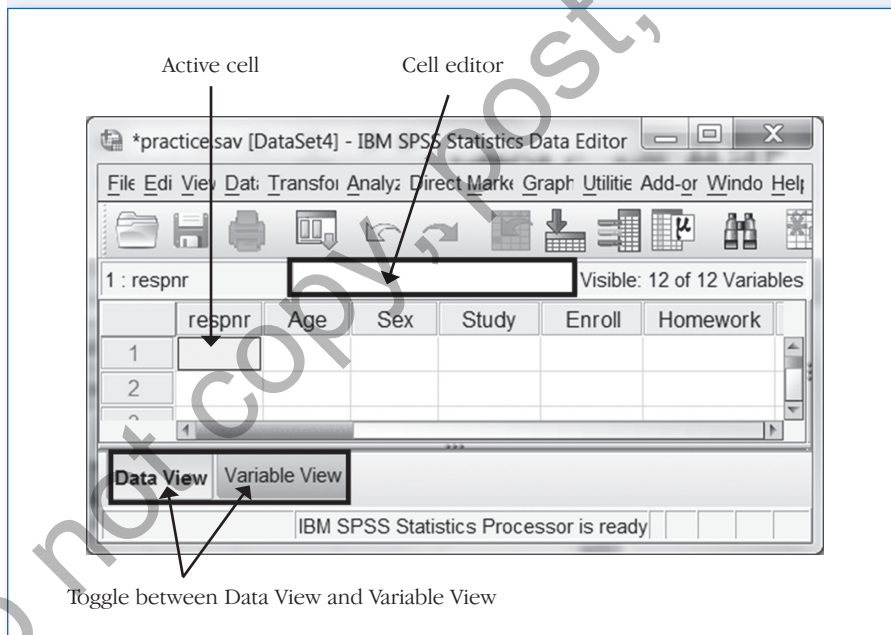


- Open the data file *practice.sav*. See Section 2.2 for the relevant instructions.
- If the Variable View window opens (Figure 2.8), click on **Data View** in the bottom left-hand corner of the variable window.

After this, the SPSS data view window is active (see Figure 2.9). This window is a type of spreadsheet, made up of rows and columns. The respondents (students in this case) are placed in the various rows, with the variables being listed in columns. The places where rows and columns intersect are called *cells*. Each of these cells contains one single score. The cell that is selected with the computer mouse is called the *active cell*. You can activate a cell by pointing the computer mouse on it and then clicking it with the left mouse button.

The active cell is the only cell in which data can be entered or altered. The content of an active cell is always displayed in the *Cell Editor*, which is the white box at the top of the spreadsheet. When you enter data into a cell, it always shows up first in the Cell Editor. In order to actually enter the data into the active cell, you have to hit the [Enter] or [Tab] key or activate another cell with the cursor.

Figure 2.9 Data Window: Data View (active)



The conventional procedure is to enter all the scores (other words are values or codes) case by case. Here, the cases are students. The data for just one case (one student) are normally displayed in one single row of the data sheet.



Select the first empty cell beneath the first variable (“respnr”).

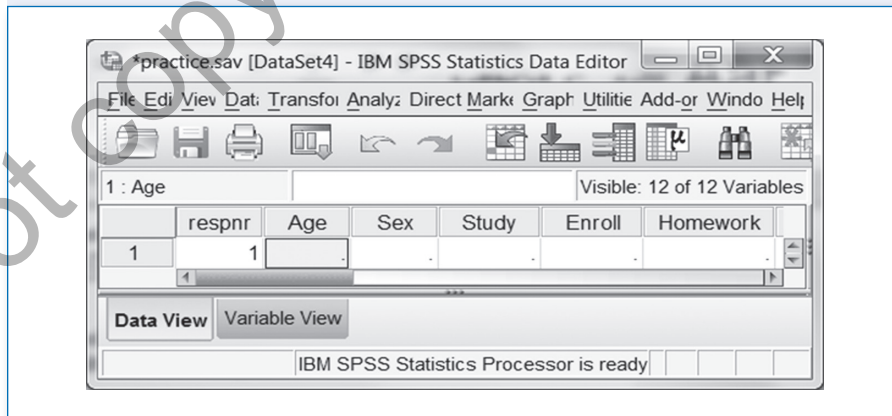
The cell will turn yellow and is activated (see Figure 2.9). You may now enter the score for this variable. In this case, it is the variable called “respnr” (respondent identification number), which serves to identify each student. The value or score for the first student then is 1.



- Enter the value 1 into the empty cell below “respnr.”
- Hit [Tab] on the keyboard.

The value 1 that you have entered is displayed in the active cell, and the first cell beneath “age” is active now (see Figure 2.10).

Figure 2.10 Data Window: First Number Entered (respnr = 1)



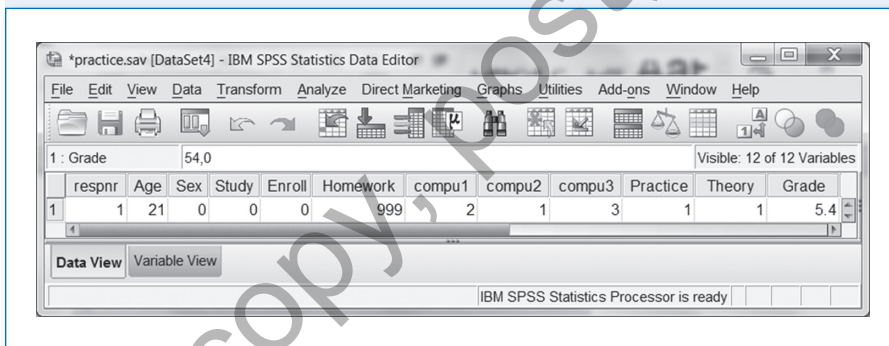
Note: respnr = respondent identification number.



- Our first student is a 21-year-old male, studies pedagogical sciences, and is full-time enrolled. He did not answer how much time he spent on statistics homework, has poor skills working with word processing, has very poor skills working with spreadsheets, has average experience using Windows, attended one practice session and one theory lecture, and received a 5.4 as his final grade. Check pages 23-24 to see how the values/codes match the preceding information and enter them into the correct cells. *Note:* Computers with language settings (see Windows Control Panel) other than English may only accept 5,4 and not 5.4.

After you have entered the values, the data view window will look like the one in Figure 2.11:

Figure 2.11 Data View Window: Data of First Student Entered

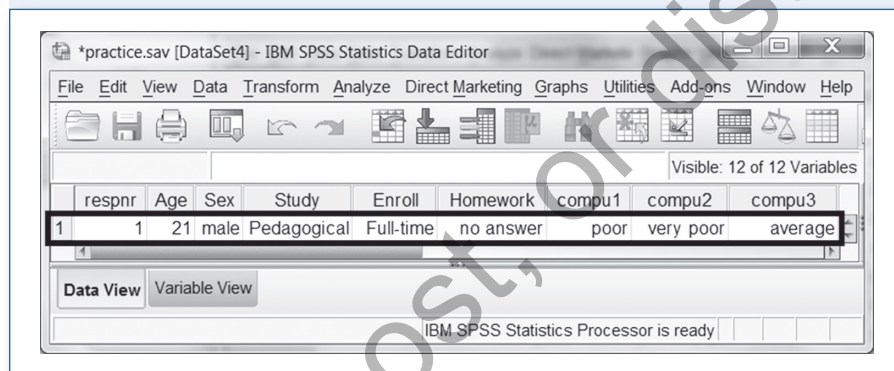


- Go to **View** in the data window menu bar.
- Select **Value Labels**.

Check whether the content of your own data window matches that of Figure 2.12 and verify whether the student (respondent 1) is indeed male,

studies pedagogical sciences, is enrolled full-time, gave no answer about time spent on homework, has poor word processing skills, has very poor experience handling spreadsheets and has average experience working with Windows, attended one practice session, one theory lecture and received a 5.4 as his final grade. If any cells do not match, then change the related cell value (for an explanation, see the instructions below Figure 2.12) until everything is correct. You do not have to save the file; this was just an exercise to learn how to enter data.

Figure 2.12 Data Window: Data View Active and “value labels” Checked



There are several ways to change cell values after you have entered them:

- Modifying cell values

You can modify the values by clicking on the cell in question and correcting the faulty value:

- Deleting a variable or case

You can delete a variable (column) or case (row) by clicking on the variable name at the top of the column or on the row number on the left-hand side and subsequently hitting [**Delete**] on your keyboard.

▶ 2.4 Load Excel Files

It is common practice to enter and store data onto a computer with programs other than SPSS. The fictitious data set we are using in this book was originally just a collection of rows and columns with numbers

(see the following example). There are many so-called spreadsheet programs that can store this type of data. One particularly popular spreadsheet program is EXCEL, which is part of the Microsoft Office package. Such files can easily be converted into SPSS data files.

Basically, EXCEL files are made up of a matrix of numbers in rows and columns. The data of one single respondent is displayed in each row. In our example, it contains the research material from the students who filled in the evaluation form. These data can be used to run statistical analyses not only within EXCEL itself but also in SPSS.

Example

The first five rows of research material (the answers of students 1 to 5) looks like the following:

1	21	0	0	0	999	2	1	3	1	1	5.4
2	32	1	1	1	15	3	1	1	3	4	3.9
3	29	0	1	1	24	4	3	3	2	3	2.8
4	21	0	0	0	36	3	2	3	2	3	7.3
5	23	1	0	0	999	4	4	3	5	5	5.0

The data above are the answers to questions in an evaluation form that have been entered into an EXCEL file. Each column (divided by spaces) provides the data for a single variable. Each row provides the answers for every student separately. In the 12 columns we have the values for the following (12) variables:

1. Student's identifier code (random unique number)
2. Student's age (measured in years)
3. Sex (0 = male, 1 = female)
4. Study (0 = pedagogical sciences, 1 = educational sciences)
5. Enrollment (0 = full-time, 1 = part-time)
6. Time spent on homework (in minutes) (no answer = 999)
7. Word processing skills (Word) (1 = very poor, 2 = poor, 3 = average, 4 = good, 5 = very good)
8. Experience working with spreadsheets (Excel) (1 = very poor, 2 = poor, 3 = average, 4 = good, 5 = very good)
9. Experience working with Windows (1 = very poor, 2 = poor, 3 = average, 4 = good, 5 = very good)

10. Number of practice sessions attended
11. Number of lectures attended
12. Final grade received for the statistics course

Consequently, we can say that the student in the first row on page 23—remember it is the one you had already entered into the *practice.sav* file during an exercise in the previous section—is a 21-year-old male, studies pedagogical sciences, is enrolled full-time, gives no answer on time spent on statistics homework, has poor word processing skills, has very poor skills handling spreadsheets, has average experience working with Windows, attends one practice and one lecture, and received a 5.4 as his final grade for the statistics course.

As mentioned earlier, the data files that have been created with other software can be converted to SPSS data files. This is generally known as *importing* data and basically involves a process that is similar to opening an SPSS data file with a *.sav* extension.

SPSS can handle a great many data file types. The most widely used types are

- SPSS data files for other operating systems, for example, Macintosh and Unix (extension: *.por*);
- text files (extension: *.txt*); and
- EXCEL files (extension: *.xls* or *.xlsx*).

For a complete overview of file types, see Figure 2.13.

As Excel is widely used, we will now demonstrate how Excel files can be imported in SPSS.



- Go to

File

Open ▶ and click Data ...

- If necessary, select the correct drive and folder (*SPSS Basics*) and to the right of **Files of type** click ▼ and select Excel as file type; see Figure 2.13 for reference.
- Now double-click on *Excelfile.xlsx* (see Figure 2.14). The dialog window "Opening Excel Data Source" will appear (see Figure 2.15).

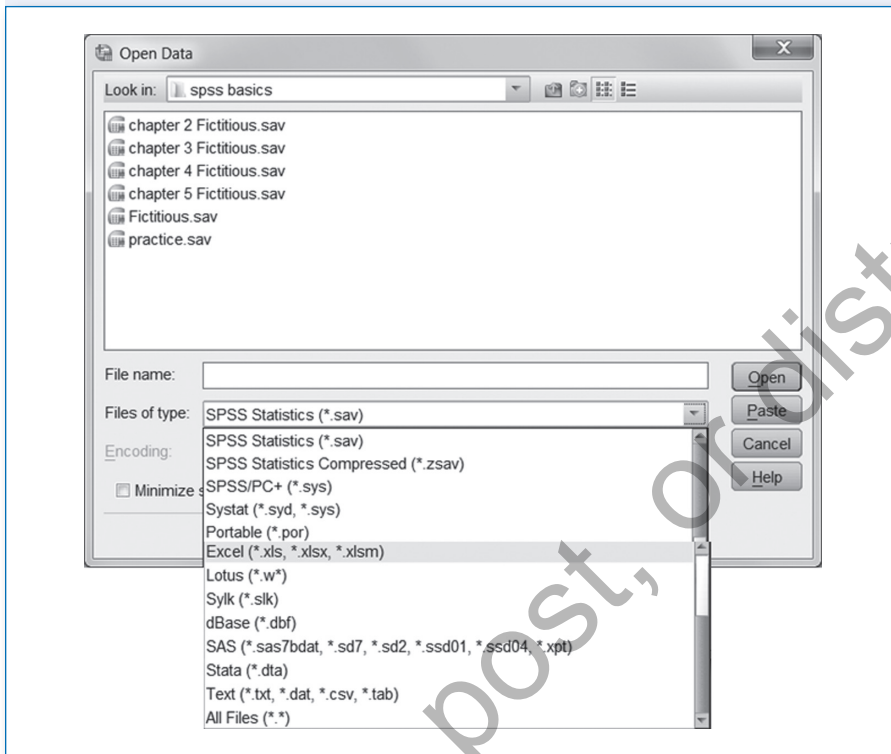
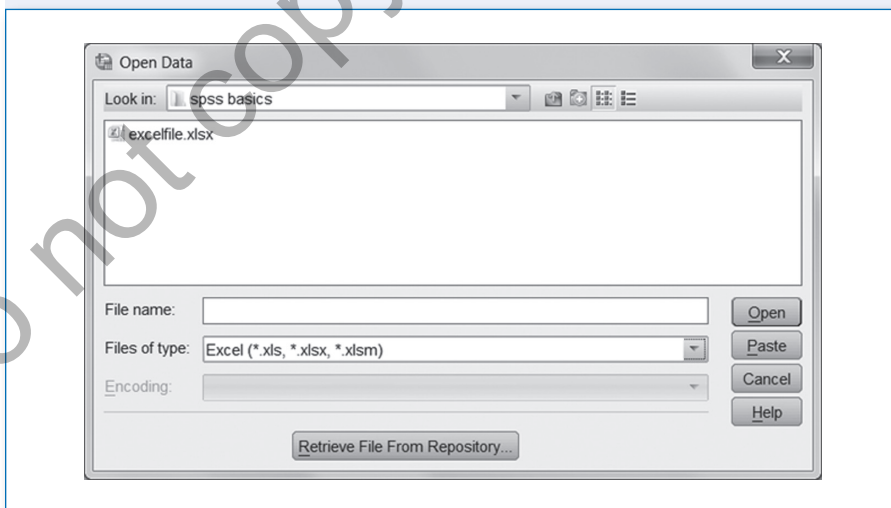
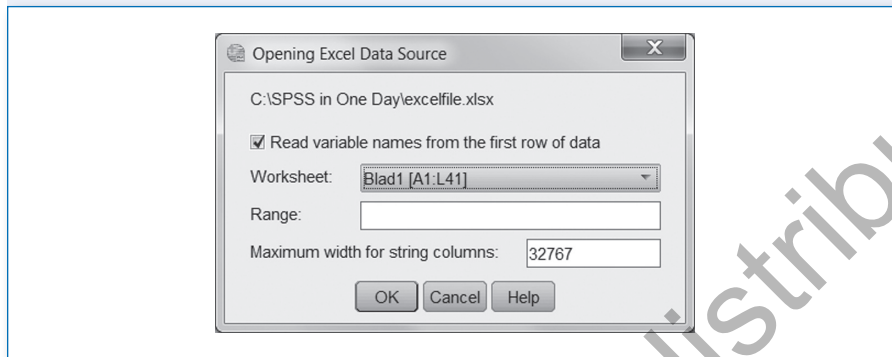
Figure 2.13 Dialog Window “Open Data”**Figure 2.14** Dialog Window “Open Data,” Files of Type: Excel

Figure 2.15 Dialog Window “Opening Excel Data Source”

- Leave “read variable names from the first row of data” checked (the Excel file has already [short] names for the variables in the first row).
- Then, click **OK**. The data will be loaded into SPSS (see Figure 2.16 for reference).

Figure 2.16 Data Window: Excel File Loaded

	Identifier	Age	Sex	Study	Enrollment	Homework	Word
1	1.0	21.0	.0	.0	.0	999.0	2.0
2	2.0	32.0	1.0	1.0	1.0	15.0	3.0
3	3.0	29.0	.0	1.0	1.0	24.0	4.0
4	4.0	21.0	.0	.0	.0	36.0	3.0
5	5.0	23.0	1.0	.0	.0	99.0	4.0
6	6.0	18.0	.0	.0	.0	33.0	3.0
7	7.0	34.0	1.0	1.0	1.0	21.0	4.0
8	8.0	33.0	1.0	1.0	1.0	24.0	3.0
9	9.0	30.0	.0	.0	1.0	24.0	2.0
10	10.0	33.0	1.0	.0	1.0	25.0	1.0

Note: Displays the first 10 rows.

2.5 Save SPSS Files ▲

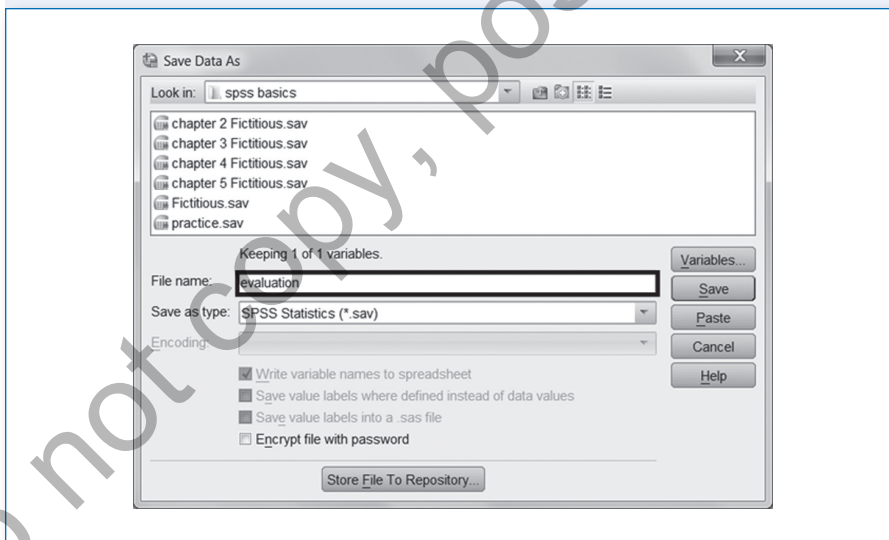
If all data have been entered correctly and have been provided with labels, and so on, then, the entire file can be saved as an SPSS data file.



- Go to **Window** in the menu bar and check whether the imported Excel file (named “Untitled”) is active. SPSS allows multiple data files to be open, although only one can be active at a time—this is indicated by the tick in front of the file name (this chapter has up to four data files open).
- Go to **File** in the menu bar and select **Save As...**

The dialog window “Save Data As” (Figure 2.17) opens:

Figure 2.17 Dialog Window “Save Data As”



Now carry out the following instructions to save the file that was opened in Section 2.4 (name: *excelfile.xlsx*) as a .sav file:



- If needed, select the correct drive and folder, which in this case is **SPSS Basics**.
- Type the relevant file name into the box (*evaluation*).
- You do not have to type the extension **.sav**; this is added automatically by SPSS!
- Now execute the command by clicking **Save**. You have now saved the Excel data file *excelfile.xlsx* as a more accessible and informative SPSS **.sav** file.
- Please note that you can also save the data as an Excel file. You may click on ▼ to the right of **Save as type** to find the possible programs.

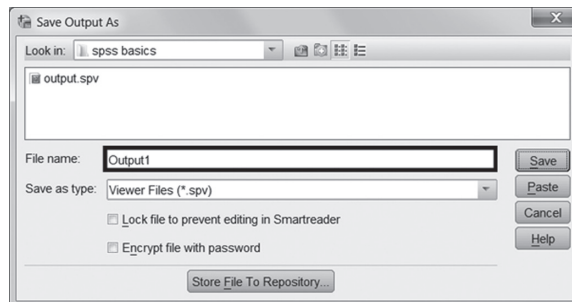
You can save your output in a similar way:



- Go to **Window** in the menu bar and select the output window called **Output1**.
- Then, from the menu bar, go to **File**
Save As...

The dialog window “Save as” will then open (see Figure 2.18).

Figure 2.18 Save SPSS Output





- In the **File name** box, type the relevant file name, which in this case is *firstoutput* (do not add an extension).
- Complete the task by clicking “**Save.**”

In this chapter, you learned how to handle data in SPSS. Since you are new to the program, we guided you through all operations. We believe that you learn SPSS best by doing. The effect is strongest when after practice you do it on your own with minimum guidance. This is why we end every chapter with some exercises that are related to the previous sections with less assistance and guidelines.

2.6 Assignments ▲


1. SPSS allows for multiple data windows to be open at any given time. The file that was opened in Section 2.2, that is, *fictitious.sav*, is still present but not active. You can activate it again by selecting **Window** in the menu bar and subsequently, clicking on *fictitious.sav*. If needed, switch to Variable View (see page 12 and Figure 2.2).
2. The data set is incomplete:
 - First of all, not all names of the variables in the file are correct. Rename V1 into *sex*, V2 into *homework*, and V3 into *practice* (see Section 2.3.1 for instructions).
 - Second, the variables “homework,” “practice,” and “grade” do not have a detailed description yet. Define in **Label** what the variables represent (see Section 2.3.2 for SPSS and page 14 for the definition of the variables).
 - Third, the variables “sex,” “study,” “homework,” and “enroll” do not have value labels (see Section 2.3.3 for SPSS and pages 23–24 for the exact value labels). Add these to the file.
 - Switch to *Data View*, and, if necessary, go to **View** in the menu bar (see instructions on page 21) and switch **value labels** on, to check whether the value labels have been entered correctly.
 - Finally, the file contains the information of 37 students. The data of the three remaining students have not yet been entered. Add them to the file in the empty rows directly beneath the last respondent (see Section 2.3.5). The remaining three students have the following scores on the variables:

respnr	age	sex	study	enroll	homework
6	18	0	0	0	33
9	30	0	0	1	24
40	36	1	1	1	32
compu1	compu2	compu3	practice	theory	grade
3	1	2	2	3	7.3 ^a
2	4	4	5	5	6.2
4	4	3	3	4	7.5

a. Because of the language settings other than English in the Windows control panel, your computer may accept a comma and not a dot.

3. The data file is now ready except for one issue. A number of students did not answer some of the questions. In our data file, students who did not provide answers to specific questions received the value 999. Now this value will still be included in the SPSS calculations. To prevent this from happening, we define it as “missing values” (when not defined thus, SPSS would, for instance, calculate an absurdly high average for the variable “homework”). So define 999 as a missing value for the variable “homework” (see Section 2.3.4).
4. To check whether 999 indeed is a missing value, you may look at the frequency distribution of the variable “homework.” You can create this distribution yourself by using the menu bar. Click on **Analyze, Descriptives Statistics ► Frequencies . . .** Then, double-click on “homework” and finally click **OK**. (We will discuss frequency tables in detail in Chapter 4; for reference, Figure 4.1 on page 51 may suffice here).
5. Look in the output window (see Section 1.5.2) for the frequency table that shows the distribution of “homework” (if necessary, activate this window via **Window** in the menu bar). Verify the following: Students spent between 15 and 41 minutes on statistics homework. Note that the four students with score 999 are no longer part of the “valid” observations as indicated by the differences between “percent” and “valid percent.”
6. Now save the adapted data file fictitious.sav. Make sure to give it a different name: *chapter2fictitious* (preferably store it in the **SPSS**

Basics folder). If the data file is given a new name, the original file is retained, which allows the user to return to the original data in case of a computer malfunction or incorrect changes.

- Save the contents of the output window (i.e., frequency distributions of “age,” “homework,” and “study”) as *outputchapter2* (see Section 2.5 for information on how to save files).
7. Check to see whether all files have actually been saved (see Section 2.2). If everything has been saved with correct names, then you may continue with the next chapter or close SPSS by clicking on the  icon in the top right-hand side of each window. You do not have to save any more files as you have already done so in Assignment 6. Please do not overwrite any files created in this chapter or in the remaining chapters. You can prevent overwriting by giving the file a new, unique name when saved.