

Chapter 2

Introduction to SPSS

Abstract This chapter introduces several basic SPSS procedures that are used in the analysis of a data set. The chapter explains the structure of SPSS data files, how to open an SPSS data file, and how to import into SPSS data contained in an Excel file. The chapter also explains how to select cases for an analysis, display variables listed in dialog boxes in alphabetical order, label and print output, paste output into a Microsoft Word document, and save data and output as SPSS or Excel files.

2.1 Overview

Throughout the book we take a hands-on approach to teaching statistics by asking you to carry out many of the statistical procedures yourself with SPSS. In this chapter, we show you how to open an SPSS data file and how to import data into an SPSS file from a Microsoft Excel spreadsheet. We show you how to modify and save a data file, how to display variables listed in dialog boxes in alphabetical order, how to label and print output, how to paste your output into a Microsoft Word document, and how to export your output into Excel.

Often it is necessary to limit an analysis of data to a subset of respondents. For example, you might want to include only women in your analysis or only respondents whose answers to a particular question were within a certain range. In this chapter, we show you one way by which you can select out a subset of cases for analysis.

The data that we will use are responses of residents of New York state in 2005 to telephone interview questions asked by the Centers for Disease Control and Prevention (CDC) Behavioral Risk Factor Surveillance System (BRFSS). BRFSS has been conducting annual cross-sectional studies of health conditions and risk behaviors in the USA since 1984. As we will soon see, the data set from 2005 consists of a large number of *categorical* and *quantitative* variables. Examples of categorical variables are the respondent's sex, marital status, and educational level. Examples of quantitative variables are the respondent's age, body mass index, and the number of days per week of exercise.



Fig. 2.1 Accessing the Open Data dialog

2.2 Opening SPSS Data Files

We will begin by opening an SPSS data file, **CDC BRFSS.sav** [1].

Double-Clicking The easiest way to open an SPSS data file is to navigate to it (e.g., with Window’s *Explore* utility) and double-click it. In a few moments, SPSS will open and display the data file in its *Data Editor*.

From Within SPSS If SPSS is already running, you can open a data file within SPSS by following the sequence displayed in Figs. 2.1 and 2.2. Within SPSS, select its *File menu* at the top of the screen and then choose **Open** and **Data**. (To save some effort on our part, hereafter we will refer to a sequence of keystrokes such as this one as **File > Open > Data**.) An alternative method for opening the dialog box is to click the *Open data document* icon, located just beneath the *File menu* tab. SPSS will then display the *Open Data* dialog box. In the *Look in* window, navigate to the location of the data file, click the file so that it appears in the *File name* window, and then click **Open**.

Importing an Excel Spreadsheet It is often convenient to build a data set with Excel, and then analyze those data with SPSS. To import an Excel spreadsheet into SPSS, use the *Open Data* dialog box. There in the *Files of type* window, click on the downward pointing arrow and select *Excel*. Any Excel files at the location to which you have navigated will now appear. Click on the file you wish to open to move it to the *File name* window. Click **Open** and the *Opening Excel data source* dialog box will appear. If the first row of the Excel file contains the names of the variables (a good idea, by the way), then be sure that *Read variable names from the first row of data* is checked. Otherwise, uncheck this instruction. Click **OK**. This sequence of keystrokes is displayed in Figs. 2.3, 2.4, and 2.5. SPSS will now convert the Excel spreadsheet to an SPSS data file.

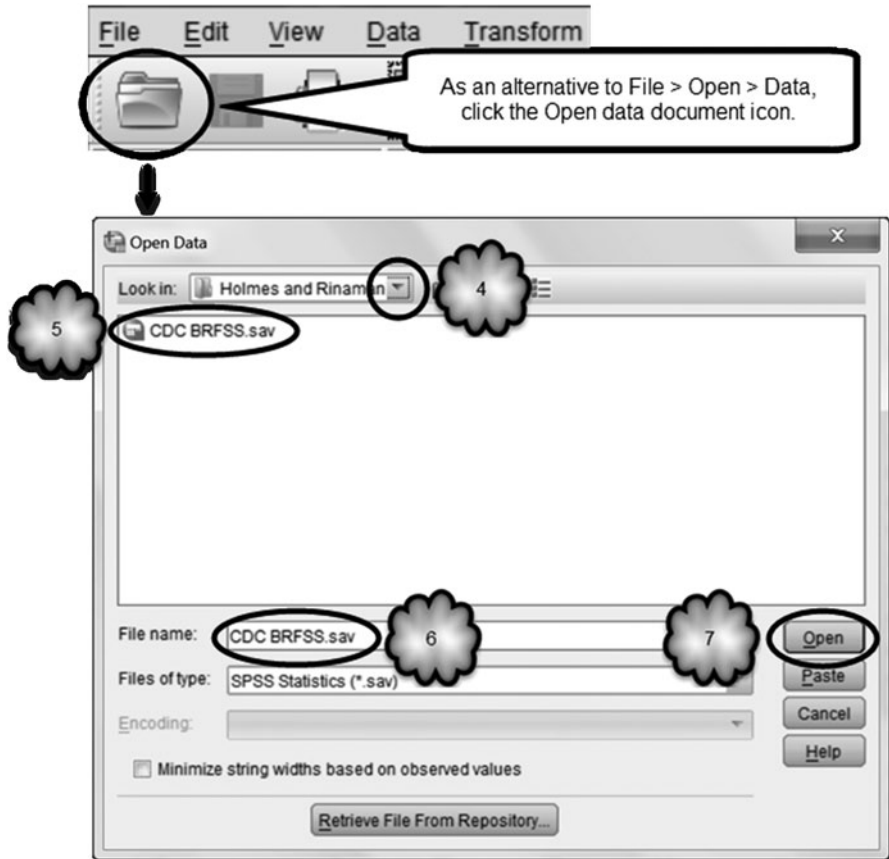


Fig. 2.2 Opening a data set with the Open Data dialog

2.3 Structure of SPSS Data Files: Data and Variable Views

Regardless of how you open a data file, once it is open you will see in the lower left-hand corner two tabs labeled *Data View* and *Variable View*, as shown in Fig. 2.6.

One of these two views will be active, and the tab for the active view will be highlighted. In Fig. 2.6, *Data View* is active. The view that is selected when a data file is opened depends on which view was active when the file was last saved.

Data View Click the *Data View* tab if it is not currently active. We will look at the *Variable View* window in a moment. A portion of the *Data View* page of the file is shown in Fig. 2.7. Note the layout of the data: Variables run across the top of the columns while respondents (SPSS refers to these as *cases*) run down the rows.

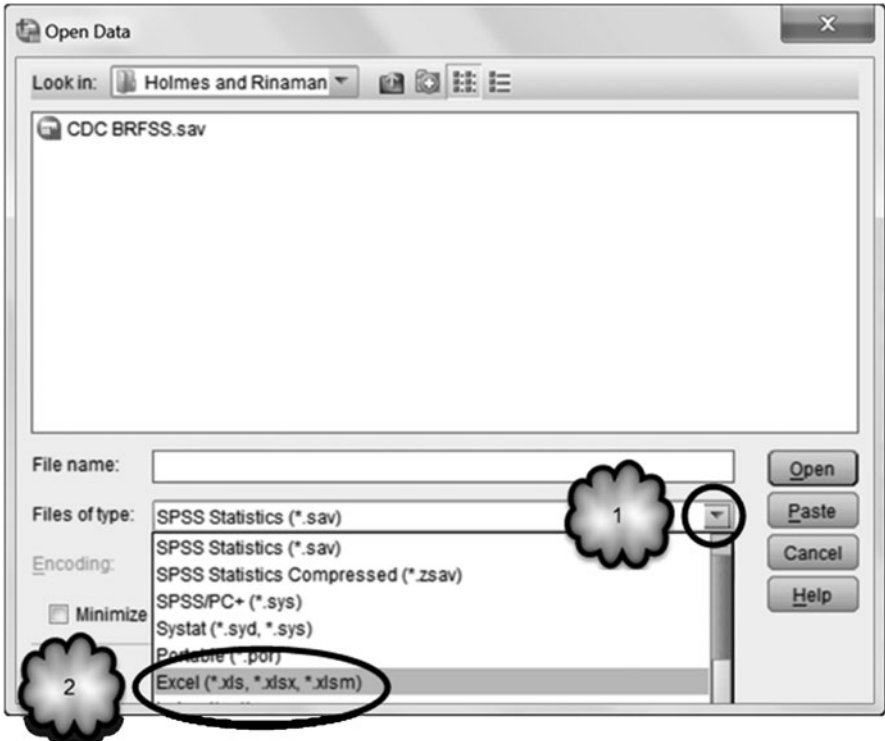


Fig. 2.3 Accessing data file types and selecting Excel as the file type

Variable View Click the *Variable View* tab in the lower left-hand corner and study the internal structure of the file. In Variable View, each row represents each of the variables and its associated properties. There are 11 properties altogether. The first five are displayed in Fig. 2.8. The 11 properties are as follows.

- **Name** This is the name that appears at the top of each column of the *Data View*. There are some limitations on names. Names must begin with a letter or one of the characters @, #, or \$, and can contain no more than 64 characters. Only @ can be used in variable names that you define. The rest of the variable name can be a combination of letters, numbers, and underbars, but they cannot contain any blank spaces or other unusual symbols. Underscore, @, #, and \$ are not considered unusual characters.
- **Type** There are different data types. The most common with this kind of data file are *String* and *Numeric*. String refers to variables that contain text. Numeric variables contain values that are numbers. To change the data type, click the *Type* cell. A button with an ellipsis will appear. Click the button to bring up the *Variable Type* dialog box, select the appropriate data type, and then click **OK**. These steps are displayed in Figs. 2.9 and 2.10.

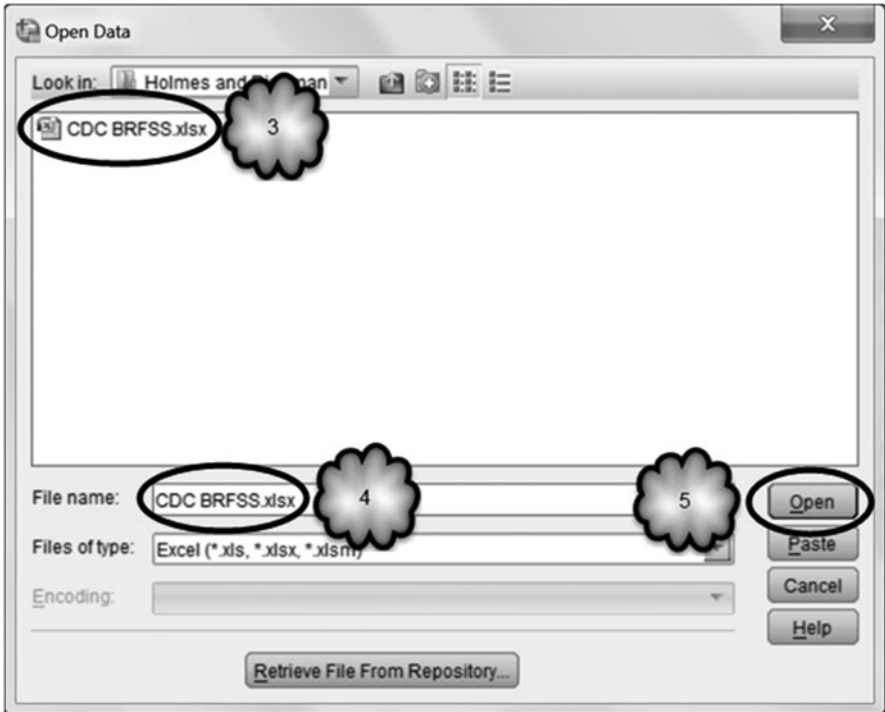


Fig. 2.4 Opening the Excel file

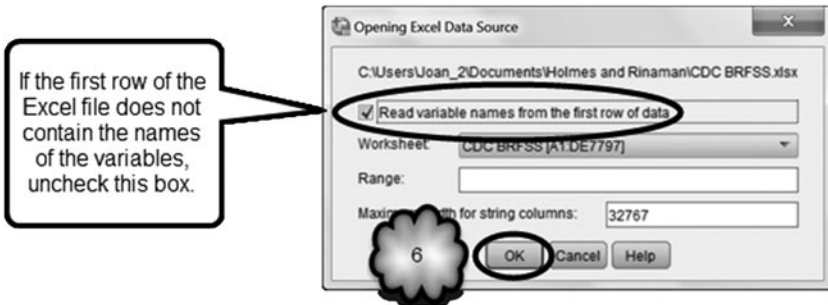


Fig. 2.5 Completing the import process

Fig. 2.6 Data view and variable view tabs



	@_GEOSTR	IMONTH	GENHLTH	PHYSHLTH	MENTHLTH	POORHLTH
1	1 07		2	88	88	.
2	1 12		3	88	88	.
3	1 02		2	88	88	.
4	1 02		2	88	2	1
5	1 04		1	88	88	.
6	1 06		2	88	1	88
7	1 03		3	88	6	7
8	1 07		2	88	28	10
9	1 06		1	88	88	.
10	1 12		2	88	15	88

Fig. 2.7 Data view

	Name	Type	Width	Decimals	Label
1	@_GEOSTR	Numeric	2	0	STRATUM CODE
2	IMONTH	String	6	0	MONTH OF INTERVIEW
3	GENHLTH	Numeric	1	0	GENERAL HEALTH
4	PHYSHLTH	Numeric	2	0	NUMBER OF DAYS PHYSICAL HEALTH NOT GOOD
5	MENTHLTH	Numeric	2	0	NUMBER OF DAYS MENTAL HEALTH NOT GOOD
6	POORHLTH	Numeric	2	0	DAYS HEALTH IMPAIRED LAST MO
7	HLTHPLAN	Numeric	1	0	HAVE HEALTH CARE COVERAGE

Fig. 2.8 Variable view

Fig. 2.9 Accessing the Variable Type dialog

	Name	Type
1	@_GEOSTR	Numeric
2	IMONTH	String
3	GENHLTH	Numeric
4	PHYSHLTH	Numeric
5	MENTHLTH	Numeric

	Name	Type
1	@_GEOSTR	Numeric
2	IMONTH	String
3	GENHLTH	Numeric ...
4	PHYSHLTH	Numeric
5	MENTHLTH	Numeric

- Width** This is the number of characters that SPSS will allow to be entered for the variable. It is set automatically when you enter data. You change this by clicking the *Width* cell for that variable and then clicking the up or down arrow to get the desired width. Figure 2.11 displays an example for a variable named **GENHLTH**.

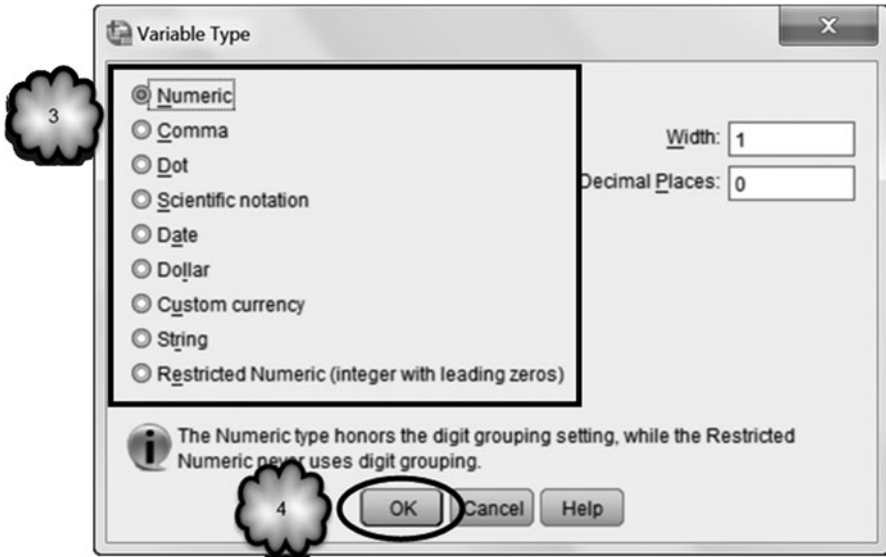


Fig. 2.10 Assigning a variable type

Fig. 2.11 Example of a numeric variable whose values consist of one character

	Name	Type	Width
1	@_GEOSTR	Numeric	2
2	IMONTH	String	6
3	GENHLTH	Numeric	1

Fig. 2.12 Changing the number of decimal places

	Name	Type	Width	Decimals
1	@_GEOSTR	Numeric	2	0
2	IMONTH	String	6	0
3	GENHLTH	Numeric	1	0

- **Decimals** This value is the number of decimal places that are displayed in the *Data View* for each value of that variable. It does not alter the values that are actually stored in the data set. As shown in Fig. 2.12, you change the number of decimal places by clicking on the *Decimals* cell and then clicking the up or down arrow to obtain the desired number of decimal places.
- **Label** This gives a more descriptive name for the variable. If this cell is not empty, the label will be displayed in any output. Figure 2.13 displays the variable label for **GENHLTH**.

It is highly recommended that every variable has a descriptive label. Because some variables names can be cryptic or uninformative, we will refer to variables by their labels rather than by their names. When we refer to a variable for the first time, we will also include the variable name in brackets and the variable number in parentheses. For example, **NUMBER OF DAYS PHYSICAL HEALTH**

Name	Label
1 @_GEOSTR	STRATUM CODE
2 IMONTH	MONTH OF INTERVIEW
3 GENHLTH	GENERAL HEALTH

Fig. 2.13 Example of a variable label

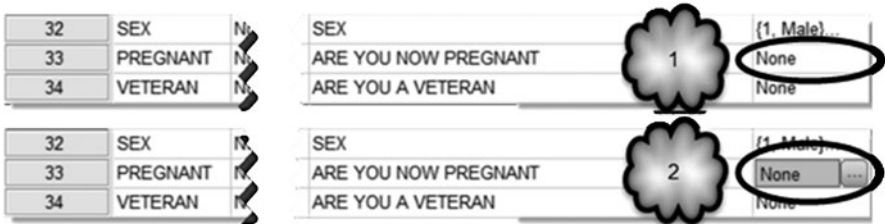


Fig. 2.14 Accessing the value labels dialog

NOT GOOD is the fourth variable in the data set, so on our first mention of it we would refer to it as **NUMBER OF DAYS PHYSICAL HEALTH NOT GOOD** [*PHYSHLTH*] (variable 4). Sometimes we will also include the value labels as well. For example, **SEX** [*SEX*] (variable 32; 1=Male; 2=Female).

- **Values** Categorical data are stored in the *Data View* with numerical values representing each category. The *Values* cell allows you to associate each numerical value with a plain language value. This should be done for all categorical variables. For example, female respondents were asked if they were pregnant. Their responses were entered into the variable, **ARE YOU NOW PREGNANT** [*PREGNANT*] (variable 33; 1=Yes, 2=No, 7=Do not know/Not Sure, 9=Refused). To enter the value labels, you would click the *Values* cell of this variable and then the button with the ellipsis to bring up the *Value Labels* dialog box. Next, you would enter each numerical value for the variable in the *Value* box, the plain language label in the *Label* box, and click the **Add** button. When you finish, you would click **OK**. Figures 2.14, 2.15, and 2.16, and 2.17 display these steps.

Practicing Entering Value Labels

Since understanding the results of data analysis is easier if value labels are used, practice entering a set of value labels for the categorical variable, **HAVE HEALTH CARE COVERAGE** [*HLTHPLAN*] (variable 7; 1=Yes; 2=No; 7=Do not know/Not sure; 9=Refused). This variable stores answers to the question, “Do you have any kind of health care coverage, including health insurance, prepaid plans such as HMOs, or government plans such as Medicare?”

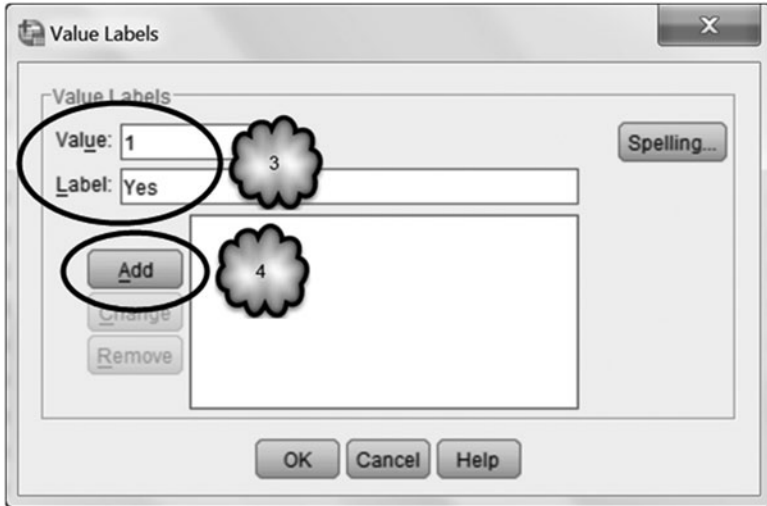


Fig. 2.15 Labeling a value

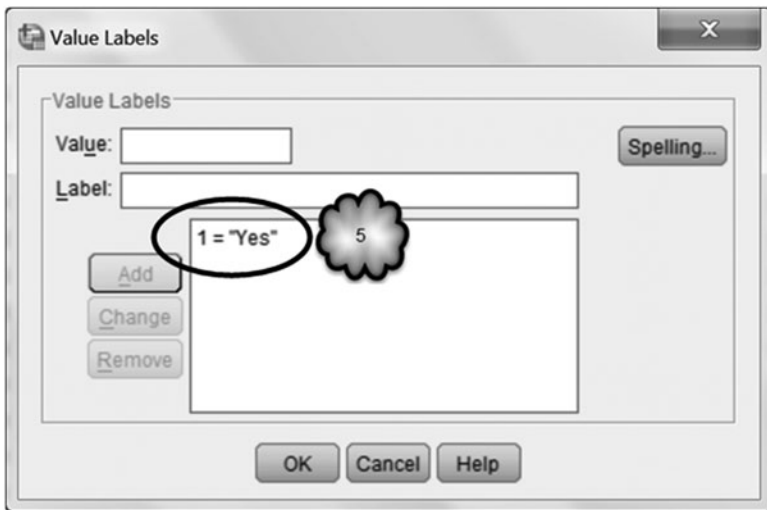


Fig. 2.16 Value Labels dialog showing one value labeled

- **Missing** Usually missing data appear in the *Data View* as periods. Sometimes special numerical values, such as 9999, are used to indicate missing data. You can set these by clicking the *Missing* cell for that variable and then the button with the ellipsis to bring up the *Missing Values* dialog box. If you have three or fewer missing value codes, you would check *Discrete missing values* and enter one code in each of the boxes. If there are more than three missing value codes, but they are within a range that does not contain any nonmissing value codes,

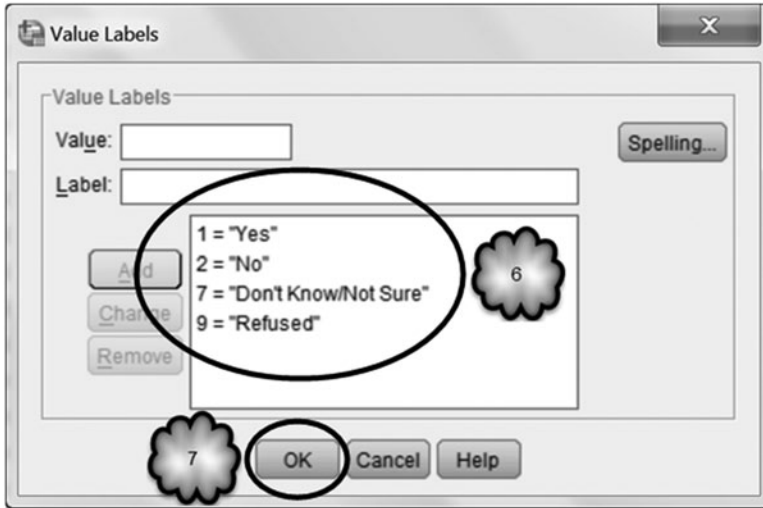
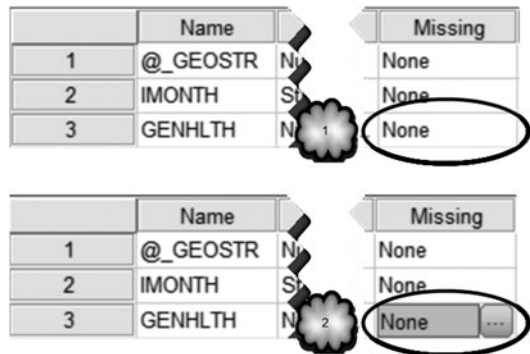


Fig. 2.17 Value Labels dialog showing all values labeled

Fig. 2.18 Accessing the Missing Values dialog



then you would check *Range plus one optional discrete missing value* and enter the low and high ends of the range in the *Low* and *High* boxes. Figures 2.18 and 2.19 show what to do when there are two missing values codes. The codes happened to be 7 and 9.

The BRFSS data set contains examples of treating some responses as missing. This is because participants sometimes responded to interview questions by saying that they were not sure or by refusing to answer. We saw an example when we entered value labels for the variable, **ARE YOU NOW PREGNANT**. Most women replied either yes or no, but some said that they did not know or were not sure, and others would not give an answer. The CDC recorded all of the responses but considered not knowing, not being sure, and refusing to answer as missing data.



<http://www.springer.com/978-3-319-12549-7>

Statistical Literacy for Clinical Practitioners

Holmes, W.H.; Rinaman, W.C.

2014, XI, 485 p. 340 illus., Hardcover

ISBN: 978-3-319-12549-7