Performing the data manipulations described in Section 1.4 of the chapter require minimal computations, easily handled with a pencil, sheet of paper, and a calculator. However, social research projects often involve many subjects and many variables, making these tasks very cumbersome to perform without the help of a computer. Statistical software programs serve fundamentally to perform the types of statistical analyses described in upcoming chapters of this book. However, basic manipulations of data provide a good point of introduction into using such programs.

Many novice researchers have experience with spreadsheets, which resemble statistical software programs in basic appearance. Spreadsheets can organize data, perform basic analyses, and produce summary graphics of data. But, they can perform only cursory assessments of data, providing a standard set of analysis options used most often by novice researchers. Statistical software programs provide more advanced statistical tests and more options for evaluation than spreadsheets do. Therefore, in the interest of maximizing the analysis capabilities, then, subsequent descriptions of software-based analyses refer to statistical software programs rather than spreadsheets.

**SPSS Basics**

Many statistical software programs exist and each has its own benefits. One of the most popular programs used by social scientists, the Predictive Analysis Software (SPSS) has a wide range of functions, including organizing data, obtaining and displaying descriptive statistics, and, as discussed later in the book, performing inferential statistical tests.

The SPSS program, greets users with a Data View screen, on which users input raw data. This screen resembles a spreadsheet.
TABLE 1.9 – SPSS DATA VIEW SCREEN
The highlighted "Data View" tab at the bottom of the screen identifies the contents of the worksheet. Each column represents an individual variable. The numbers on the left of the screen refer to subjects. Thus, reading across a row, one can see a particular subject’s scores on all variables for which data exists.

By clicking the “Variable View” tab at the bottom of the screen, the user can access another window in which he or she can provide information about the data.

TABLE 1.10 – SPSS VARIABLE VIEW SCREEN
The highlighted "Variable View" tab at the bottom of the screen identifies the information that the page contains. On this screen, users can supply names and information for the data supplied by each variable.

On this page, researchers can
- identify a name for each variable in the column entitled “name.” (Variable names cannot begin with a number and cannot contain any non-alpha-numeric characters. Also, some versions of SPSS limit variable names to eight characters.)
- provide a description of each variable in the column entitled “label.”
- provide a coding scheme for qualitative data in the column entitled, “values.” Use the following steps to do so.

1. Click on the relevant cell in the “values” column, and then the grey box that becomes visible in the cell. A window, entitled Value Labels should appear.
2. Input the coding scheme by identifying each numerical code as the value and each category name as the label. Click the “Add” button after entering the information for each category. The code for each newly-added category should appear in the large untitled box below the value and label areas.

3. Click OK.
- adjust the appearance of numerical values in the column entitled, “decimals.”
- inform SPSS of any units of measurements for the data in the column entitled, “type.”

Supplying this information permits the SPSS program to correctly interpret the data as it is entered onto the Data View screen.

After you enter variable names on the Variable View screen, the names of the variable should appear at the top of a column on the Data View screen. You simply need to enter the data for each variable in the cells below the appropriate variable name. Because SPSS understands only numbers, all data must take numerical form. You must code qualitative data and input the coding scheme into the values column on the Variable View screen. To enter a word or name, for example, to identify subjects by initials, you should identify that variable as a “string” in the type column on the Variable View screen. SPSS cannot interpret these strings; they have meaning only for the user.

**Data Manipulation in SPSS**
Various SPSS utilities allow the user to manipulate inputted data. Options within the program’s Data and Transform pull-down menus perform tasks such as sorting data, selecting cases, weighting cases. The toolbar also contains icons that lead to shortcuts for
some of these menus’ most commonly-used applications. However, for purposes of providing full understandings of the processes involved, instructions in the following sections present the formal methods of handling manipulations of data.

The Sort Function

**Sorting** involves rearranging subjects and their associated data in order of increasing or decreasing values for a chosen variable. One may sort data for a variety of reasons including the need to easily find particular data points and to make any gaps in the progression of continuous data values evident. Most researchers do not sort data by hand because, especially with large data sets, doing so takes a very long time. However, SPSS can sort data within a matter of seconds.

The process of sorting data in SPSS requires five steps.

1. Begin by selecting Data from the options at the top of the Data View or Variable View screen. A pull-down menu should appear.
2. From the pull-down menu, select “Sort Cases.” A new window entitled Sort Cases should appear.

![FIGURE 1.13 – SPSS SORT CASES WINDOW](image)

The user sorts data by selecting the appropriate variable from those listed in the box above. The designation in the “Sort Order” portion of the window instructs SPSS whether to arrange the values in increasing or decreasing order.

3. An untitled box in the Sort Cases window contains the names of all variables for which data exists in the file. Indicate the variable that should serve as the basis for sorting by clicking on its name and clicking on the arrow to the right of the box. The name of the variable should move from its original position to the box labeled “Sort by.”
4. If necessary, change the designation in the “Sort Order” box from ascending to descending.
5. Click “OK.”

Upon completing this process, the Data View screen should reappear with data sorted according to the values of the variable specified.

The Filter Function

The Data option at the top of the Data View and Variable View screens also contains the command to filter data. The following steps present the process for doing so.

1. Begin by selecting Data from the options at the top of the Data View or Variable View screen. A pull-down menu should appear.
2. From the pull-down menu, select “Select Cases.” A new window entitled Select Cases should appear.

![FIGURE 1.14 – SPSS SELECT CASES WINDOW](image)

The filtering process begins with the user informing SPSS that he or she will describe the data that should remain viable for future analyses. Selecting the “If condition is satisfied” option accomplishes this task.

3. Click on the “If” button located under the “If condition is satisfied” prompt located in the “select” portion of the box. Another window, entitled Select Cases :If, should appear.
FIGURE 1.15 – SPSS SELECT CASES: IF WINDOW
The user filters data by selecting the appropriate variable from those listed in the box above and then defining the conditions for inclusion in the subsequent analysis. Until the user removes the filter, SPSS ignores all other data points.

4. An untitled box in the Select Cases: If window contains the names of all variables for which data exists in the file. Indicate the variable that should serve as the basis for filtering by clicking on its name and clicking on the arrow to the right of the box. The name of the variable should move from its original position to the box on the right.

5. Use the keypad located in the Select Cases: If window to describe the values of the selected variable that should remain. For example, a researcher who wishes to omit subjects coded with a “1” from subsequent analyses would select click on the “>” and the “1” buttons. A mathematical statement identifying the selected values for the filter variable (e.g. VAR00002 > 1) should appear in the box above the keypad.

6. Click “Continue” to return to the Select Cases window.

7. Click “OK.”

Upon completing this process, the Data View screen should reappear. Diagonal lines through case numbers identifies the data to eliminate from subsequent analyses. Also, a new column, entitled “filter_$ should appear. This column assigns values of 0 to cases that SPSS has “filtered out” and 1 to the cases that remain.
The diagonal lines through subject numbers 1, 9, and 10 indicate that subsequent analyses will not data from these individuals. These subjects also receive the designation of “0” in the newly-created “filter_$” variable.

To remove the filter, the user need only choose the “All Cases” option from the “Select” portion of the Select Cases window.

The Split File Function
With SPSS’s split file function enabled, SPSS executes all commands separately for each category of a particular variable. Researchers should use the following procedure to split a file.

1. Begin by selecting “Data” from the options at the top of the Data View or Variable View screen. A pull-down menu should appear.
The user splits data by selecting the appropriate variable from those listed in the box above and, then, indicating that SPSS should organize its output by groups for that variable. SPSS splits the file until the user disables the function.

3. Select “Organize output by groups” from the options in the center of the Split Files window. The originally faded box marked “Groups Based on,” should become clearly visible.

4. An untitled box in the Split Files window contains the names of all variables for which data exists in the file. Indicate the variable that should serve as the basis for splitting the file by clicking on its name and clicking on the arrow to the right of the box. The name of the variable should move from its original position to the box marked “Groups Based on.”

5. Click OK.

Upon returning to the Data View page, the user sees no immediate evidence of splitting a file, as he or she sees after sorting of filtering data. The results of having split a file do not become apparent until the user creates SPSS output, when he or she can see separate statistics or graphics for each identified category of subjects.

To disable the split files, command, the user must revisit the Split File window to instruct the program to analyze all cases of the data. Also, for good measure, the user should remove the name of the variable that served as the factor for splitting the file from the “Groups Based on” box, by highlighting its name and clicking on the arrow the left of the box.

The Compute Function
Performing a computation in SPSS results in the creation of a new variable. However, SPSS does not self-define the computed variable as it does for the filtering variable. One instructs SPSS how to determine the values for the computed variable using the following steps.

1. Begin by selecting Transform from the options at the top of the Data View or Variable View screen. A pull-down menu should appear.
2. From the pull-down menu, select “Compute Variable.” A new window entitled *Compute Variable* should appear.

![SPSS Compute Variable Window](image)

**FIGURE 1.17 – SPSS COMPUTE VARIABLE WINDOW**
The user calculates a new variable from existing values by, first, creating a name for the new variable and, then, inputting a formula to compute the new variable. The “Function group” box contains commands for more advanced mathematical operations than those available from the keypad in the center of the window.

3. Select a name for the new variable that SPSS will calculate. Enter this variable name as the target variable.

4. An untitled box in the *Compute Variable* window contains the names of all variables for which data exists in the file. Use these variable names and the operators located in the center of the window to indicate the desired calculations. Move the name of a variable to the “Numeric Expression” box by clicking on its name and then clicking on the arrow to the right of the untitled box. Move an operator to the “Numeric Expression” box by clicking on its key.

5. Click “OK.”
The newly-created variable should appear on the Data View page.

<table>
<thead>
<tr>
<th></th>
<th>VAR000001</th>
<th>VAR000002</th>
<th>newvariable</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>52.00</td>
<td>1.00</td>
<td>52.00</td>
</tr>
<tr>
<td>2</td>
<td>79.00</td>
<td>1.00</td>
<td>79.00</td>
</tr>
<tr>
<td>3</td>
<td>52.00</td>
<td>1.00</td>
<td>52.00</td>
</tr>
<tr>
<td>4</td>
<td>68.00</td>
<td>2.00</td>
<td>34.00</td>
</tr>
<tr>
<td>5</td>
<td>36.00</td>
<td>2.00</td>
<td>18.00</td>
</tr>
<tr>
<td>6</td>
<td>58.00</td>
<td>2.00</td>
<td>29.00</td>
</tr>
<tr>
<td>7</td>
<td>74.00</td>
<td>3.00</td>
<td>24.67</td>
</tr>
<tr>
<td>8</td>
<td>29.00</td>
<td>3.00</td>
<td>9.67</td>
</tr>
<tr>
<td>9</td>
<td>85.00</td>
<td>3.00</td>
<td>28.33</td>
</tr>
<tr>
<td>10</td>
<td>66.00</td>
<td>3.00</td>
<td>22.00</td>
</tr>
</tbody>
</table>

**TABLE 1.12 – COMPUTED DATA**

Values produced by the formula defined in SPSS’s *Compute Variable* window appear on the Data View screen. The column entitled “newvariable” contains values obtained by dividing VAR000001 numbers by VAR000002 numbers.