Data files are not always organized in the ideal form for your specific needs. To prepare data for analysis, you can select from a wide range of file transformations, including the ability to:

- **Sort data.** You can sort cases based on the value of one or more variables.
- **Select subsets of cases.** You can restrict your analysis to a subset of cases or perform simultaneous analyses on different subsets.

## Sorting Data

Sorting cases (sorting rows of the data file) is often useful—and sometimes necessary—for certain types of analysis. To reorder the sequence of cases in the data file based on the value of one or more sorting variables, from the menus choose:

Data
Sort Cases...

This opens the Sort Cases dialog box, as shown in Figure 13.1.

**Figure 13.1 Sort Cases dialog box**

If you select multiple sort variables, the order in which they appear on the Sort By list determines the order in which cases are sorted. For example, based on
the Sort By list in Figure 13.1, cases will be sorted by the value of `prevexp` within categories of `jobcat`. For string variables, uppercase letters precede their lowercase counterparts in sort order (for example, the string value `Yes` comes before `yes` in sort order).

**Split-File Processing**

To split your data file into separate groups for analysis, from the menus choose:

Data
  Split File...

This opens the Split File dialog box, as shown in Figure 13.2.

![Split File dialog box](image)

To split the data file into separate groups for analysis:

1. Select Compare groups or Organize output by groups.
2. Select the variable(s) to use to split the file into separate groups.

You can use numeric, short string, and long string variables as grouping variables. A separate analysis is performed for each subgroup defined by the grouping variables. If you select multiple grouping variables, the order in which they appear on the Groups Based On list determines the manner in which cases are grouped. For example, based on the Groups Based On list in Figure 13.2, cases will be grouped by the value of `gender` within categories of `jobcat`. 

The Split File procedure automatically sorts the data file based on the values of the grouping variables. If the original order of cases is important, do not save the file after using the Split File option.
Sorting Cases for Split-File Processing

The Split File procedure creates a new subgroup each time it encounters a different value for one of the grouping variables. Therefore, it is important to sort cases based on the values of the grouping variables before invoking split-file processing.

By default, Split File automatically sorts the data file based on the values of the grouping variables. If the file is already sorted in the proper order, you can save processing time if you select File is already sorted.

Turning Split-File Processing On and Off

Once you invoke split-file processing, it remains in effect for the rest of the session unless you turn it off.
- Analyze all cases. Turns split-file processing off.
- Compare groups and Organize output by groups. Turns split-file processing on.

If split-file processing is in effect, the message Split File on appears on the status bar at the bottom of the application window.

Selecting Subsets of Cases

You can restrict your analysis to a specific subgroup based on criteria that include variables and complex expressions. You can also select a random sample of cases. The criteria used to define a subgroup can include:
- Variable values and ranges
- Date and time ranges
- Case (row) numbers
- Arithmetic expressions
- Logical expressions
- Functions

To select a subset of cases for analysis, from the menus choose:

Data
  Select Cases...
This opens the Select Cases dialog box, as shown in Figure 13.3.

**Figure 13.3 Select Cases dialog box**

**Unselected Cases**

You can choose one of the following alternatives for the treatment of unselected cases:

- **Filtered.** Unselected cases are not included in the analysis but remain in the data file. You can use the unselected cases later in the session if you turn filtering off. If you select a random sample or if you select cases based on a conditional expression, this generates a variable named `filter_`$\_S$ with a value of 1 for selected cases and a value of 0 for unselected cases.

- **Deleted.** Unselected cases are deleted from the data file. By reducing the number of cases, the open data file, you can save processing time. Deleted cases can be recovered only by exiting from the file without saving any changes and then reopening the file. The deletion of cases is permanent if you save the changes to the data file.
Selecting Cases Based on Conditional Expressions

To select cases based on a conditional expression, select "If condition is satisfied" and click "If" in the Select Cases dialog box. This opens the Select Cases If dialog box, as shown in Figure 13.4.

Figure 13.4 Select Cases If dialog box

The conditional expression can use existing variable names, constants, arithmetic operators, logical operators, relational operators, and functions. You can type and edit the expression in the text box just like text in an output window (see Chapter 4). You can also use the calculator pad, variable list, and function list to paste elements into the expression. See Chapter 12 for more information on working with conditional expressions.