Introduction to SPSS

The "**Statistical Package for the Social Sciences**" (**SPSS**) is a package of programs for manipulating, analyzing, and presenting data; the package is widely used in the social and behavioral sciences. There are several forms of SPSS. The core program is called *SPSS Base* and there are a number of add-on modules that extend the range of data entry, statistical, or reporting capabilities.

Data Entry

When SPSS 24 for Windows is first opened, a default dialogue box appears that gives the user a number of options. Most likely users will want to **enter data** or **open an existing data file**.
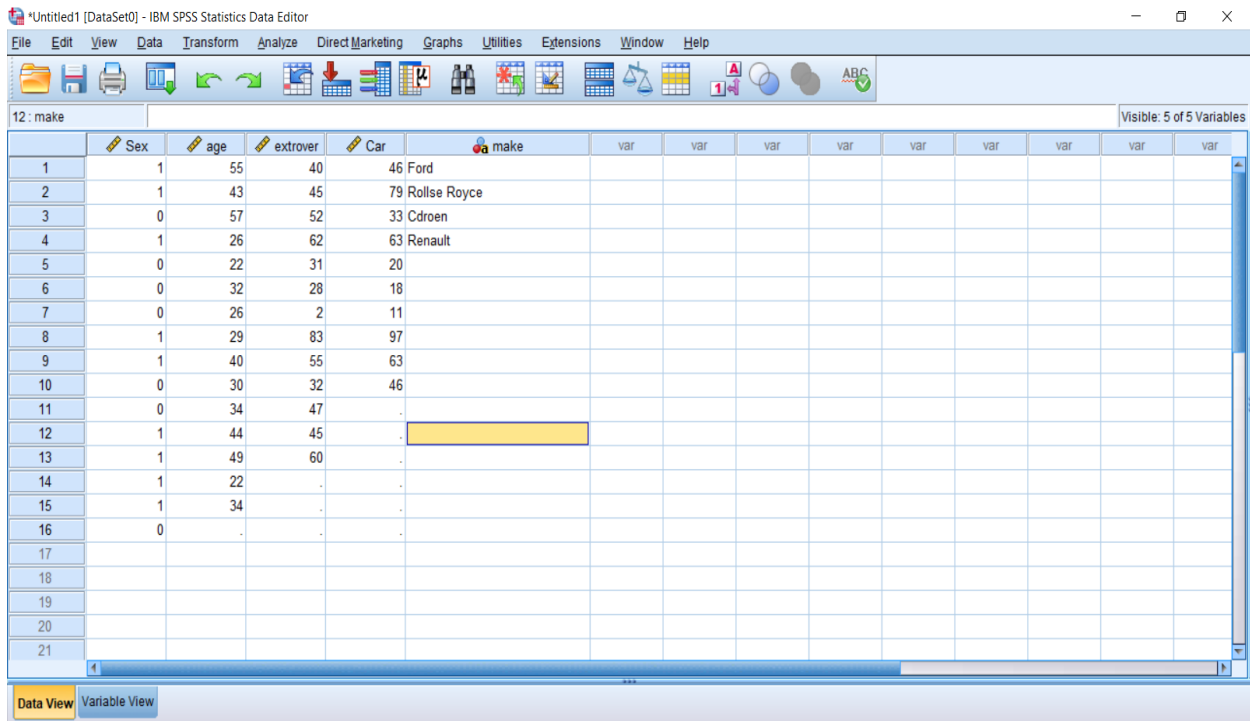
When this type in data is selected, the SPSS Data Editor appears as an empty spreadsheet. At the top of the screen is a menu bar and at the bottom a status bar. The status bar informs the user about facilities currently active.

1. The Data View Spreadsheet

The Data Editor consists of two windows. By default, the **Data View**, which allows the data to be entered and viewed as shown in Figure 1.

 The other window is the **Variable View**, which allows the types of variables to be specified and viewed. The user can toggle between the windows by clicking on the appropriate tabs on the bottom left of the screen.

By default, SPSS aligns numerical data entries to the right-hand side of the cells and text (string) entries to the left-hand side. Here variables sex, age, extrovert, and car take numerical.
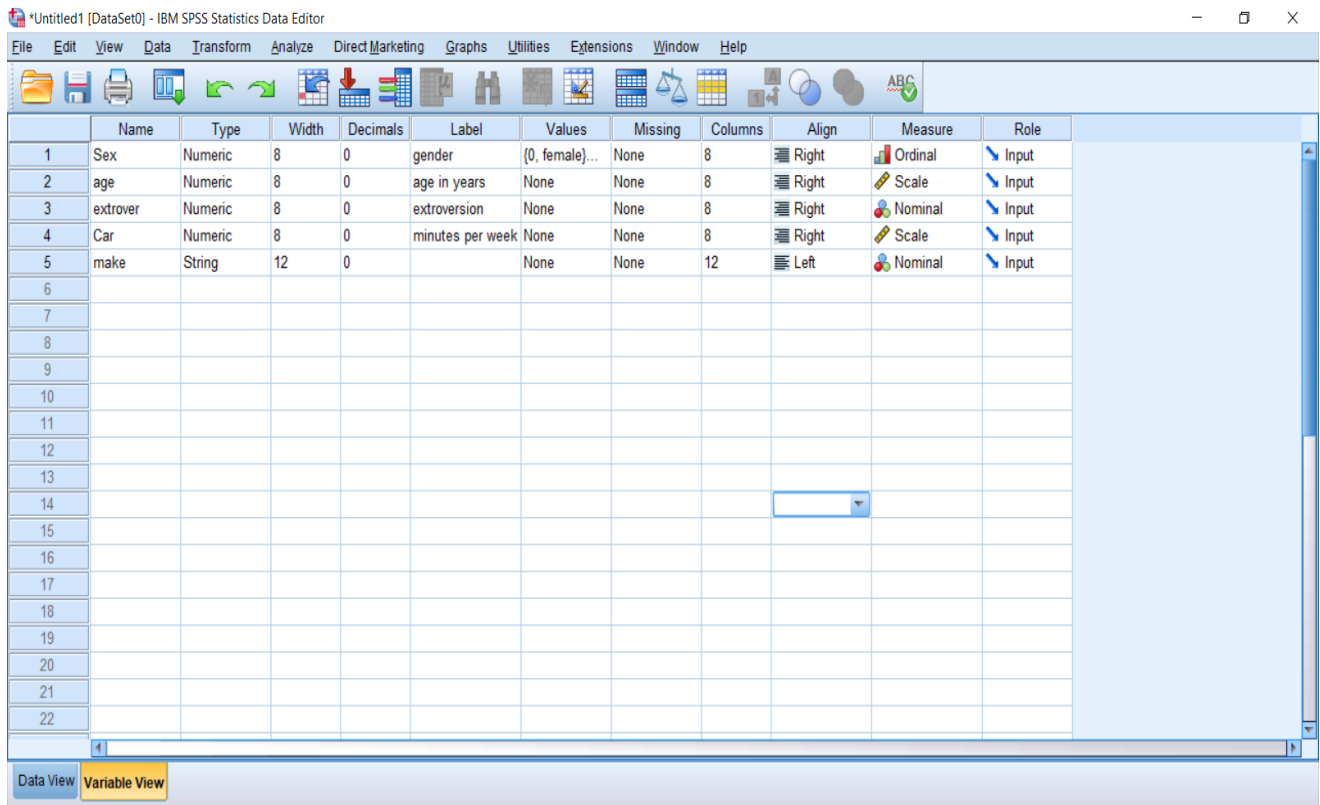
Figure (1): Data View window of the Data Editor

2. The Variable View Spreadsheet

The Variable View spreadsheet serves to define the variables. Each variable definition occupies a row of this spreadsheet. As soon as data is entered under a column in the Data View, the default name of the column occupies a row in the Variable View.

There are 10 characteristics to be specified under the columns of the Variable View (see Figure 2.):

Figure (2): Variable View window of the Data Editor.

1.Name — the chosen variable name. This can be up to eight alphanumeric characters but must begin with a letter. While the underscore (_) is allowed, hyphens (-), ampersands (&), and spaces cannot be used. Variable names are not case sensitive.

2. Type — the type of data. SPSS provides a default variable type once variable values have been entered in a column of the Data View.

The type can be changed by highlighting the respective entry in the second column of the Variable View and clicking the three-periods symbol (…) appearing on the right-hand side of the cell. This results in the Variable Type box being opened, which offers a number of types of data including various formats for numerical data, dates, or currencies.

3. Width — the width of the actual data entries. The default width of numerical variable entries is eight. The width can be increased or decreased by highlighting the respective cell in the third column and employing the upward or downward arrows appearing on the right-hand side of the cell or by simply typing a new number in the cell.

4.Decimals — the number of digits to the right of the decimal place to be displayed for data entries. This is not relevant for string data and for such variables the entry under the fourth column is given as a greyed-out zero. The value can be altered in the same way as the value of Width.

5. Label — a label attached to the variable name. In contrast to the variable name, this is not confined to eight characters and spaces can be used. It is generally a good idea to assign variable labels.

They are helpful for reminding users of the meaning of variables (placing the cursor over the variable name in the Data View will make the variable label appear) and can be displayed in the output from statistical analyses.

6. Values — labels attached to category codes. For categorical variables, an integer code should be assigned to each category and the variable defined to be of type "numeric." When this has been done, clicking on the respective cell under the sixth column of the Variable View makes the three-periods symbol appear, and clicking this opens the Value Labels dialogue box, which in turn allows assignment of labels to category codes as shown in Figure 3.
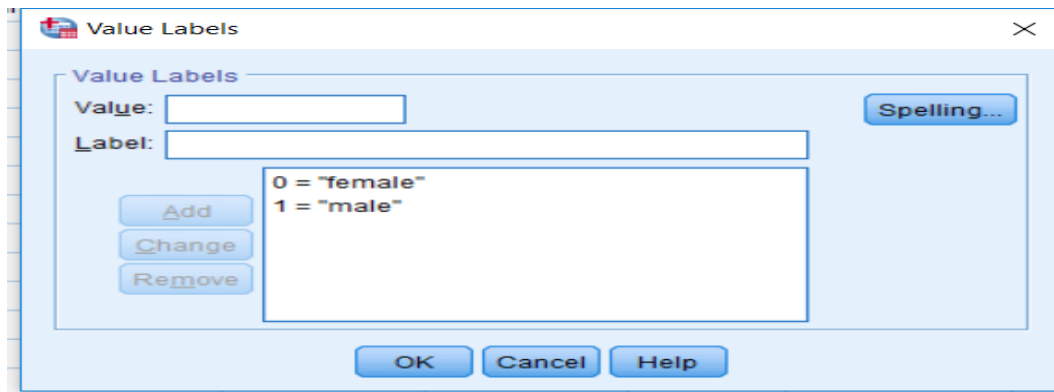
Figure (3): Declaring category code labels

7. Missing — missing value codes. SPSS recognizes the period symbol as indicating a missing value. If other codes have been used (e.g., 99, 999) these have to be declared to represent missing values by highlighting the respective cell in the seventh column.

8. Columns — width of the variable column in the Data View. The default cell width for numerical variables is eight. Note that when the Width value is larger than the Columns value, only part of the data entry might be seen in the Data View. The cell width can be changed in the same way as the width of the data entries or simply by dragging the relevant column boundary.

9. Align — alignment of variable entries. The SPSS default is to align numerical variables to the right-hand side of a cell and string variables to the left. It is generally helpful to adhere to this default; but if necessary, alignment can be changed by highlighting the relevant cell in the ninth column and choosing an option from the drop-down list.

10. Measure — measurement scale of the variable. The default chosen by SPSS depends on the data type. For example, for variables of type "numeric," the default measurement scale is a continuous or interval scale (referred to by SPSS as "scale"). For variables of type "string," the default is a nominal scale. The third option, "ordinal," is for categorical variables with ordered categories but is not used by

default. It is good practice to assign each variable the highest appropriate measurement scale ("scale" > "ordinal" > "nominal").

Since this has implications for the statistical methods that are applicable. The default setting can be changed by highlighting the respective cell in the tenth column and choosing an appropriate option from the drop-down list.