## 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.

### Data Entry

When SPSS 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. The Data Editor consists of two windows **Data View** and **Variable View.** The user can toggle between the windows by clicking on the appropriate tabs on the bottom left of the screen.

## 1. The Data View Spreadsheet

The **Data View**, which allows the data to be entered and viewed as shown in Figure 1

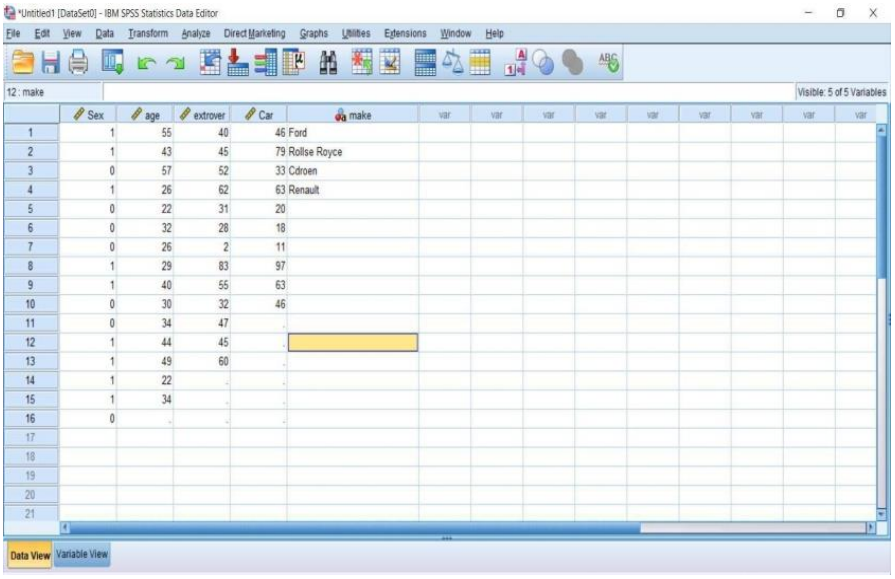



Figure (1): Data View window of the Data Editor

## 2. The Variable View Spreadsheet

The **Variable View spreadsheet** is a window, which allows the types of variables to be specified and viewed. 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.):

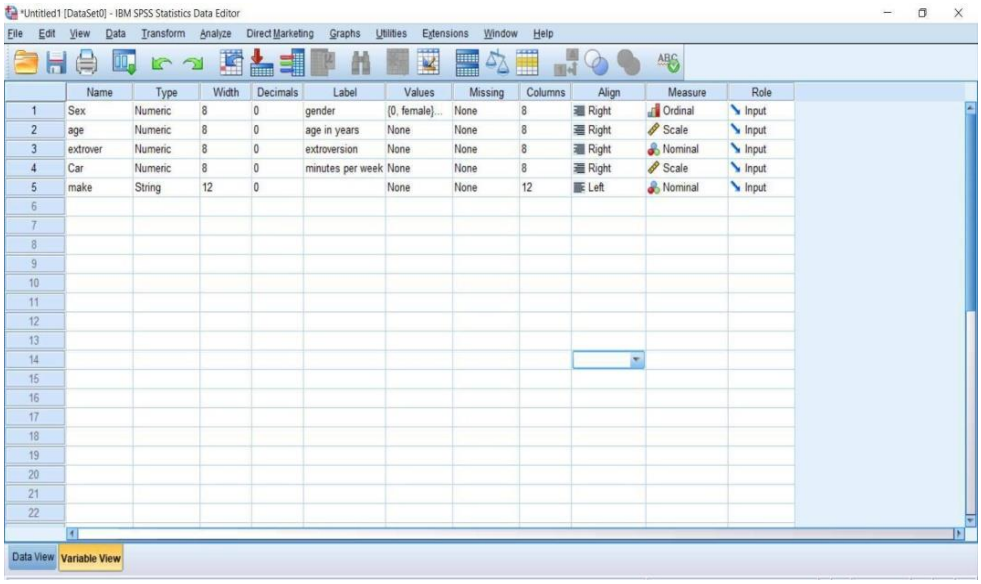



**Figure (2):** Variable View window of the Data Editor.

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.

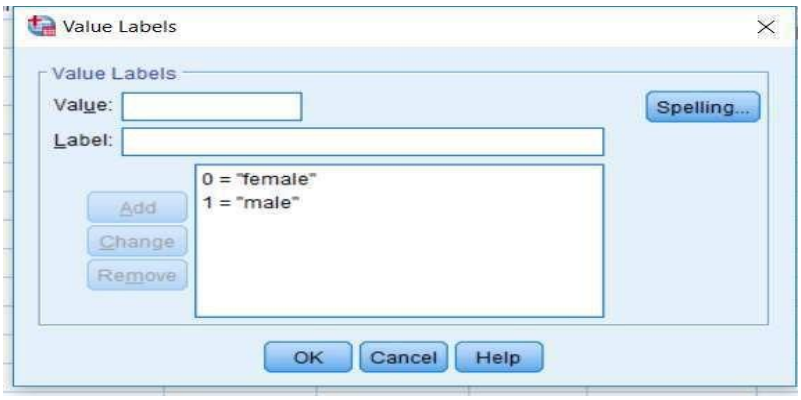



Figure (3): Declaring category code labels

**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 level of the variable. The default chosen by SPSS depends on the data type. For example, for variables of type "numeric," the default measurement level is scale which is a continuous or interval type. For variables of type "string," the default is a nominal . 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" > "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

## Frequencies Distributions

This is the most important univariate procedure. Conducting it properly, and interpreting the output rigorously, will enable you to understand the major attributes of the frequency distribution of each variable.

1. **The distribution of variables –frequency, statistics, charts**

GO TO Analyze/ Descriptive Statistics/ Frequencies. Select the variables and move them into the box "Variable(s) as shown in Figure 4."

Figure (4): Frequencies Interface

Now for constracting a chart in SPSS with frequency table (as Histogram), click on the button “Charts.” Choose to draw a histogram with a normal curve - select the option “Histogram” and click on the box to the left of the title “With normal curve. " Click on “Continue.”

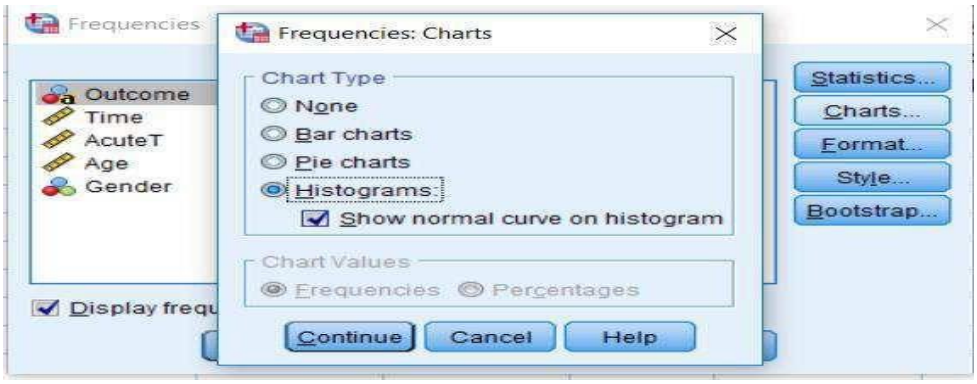



Figure (5): Selecting type of chart

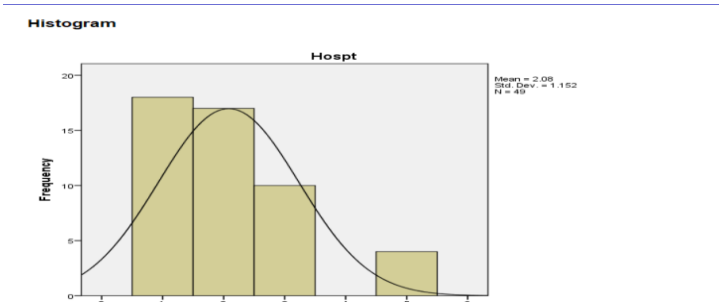



Figure (6): Histogram with normal distribution

We also want to obtain **descriptives**. Click on the button "Statistics." As in Figure 8. Select the options as shown in Figure 9. These statistics cover the list of "descriptive statistics.

The options under "Percentile values" can assist in learning about the spread of the variable across its range.

Figure (8): Applying descriptives by selecting statistics

- Click on "Continue."
- Click on OK.

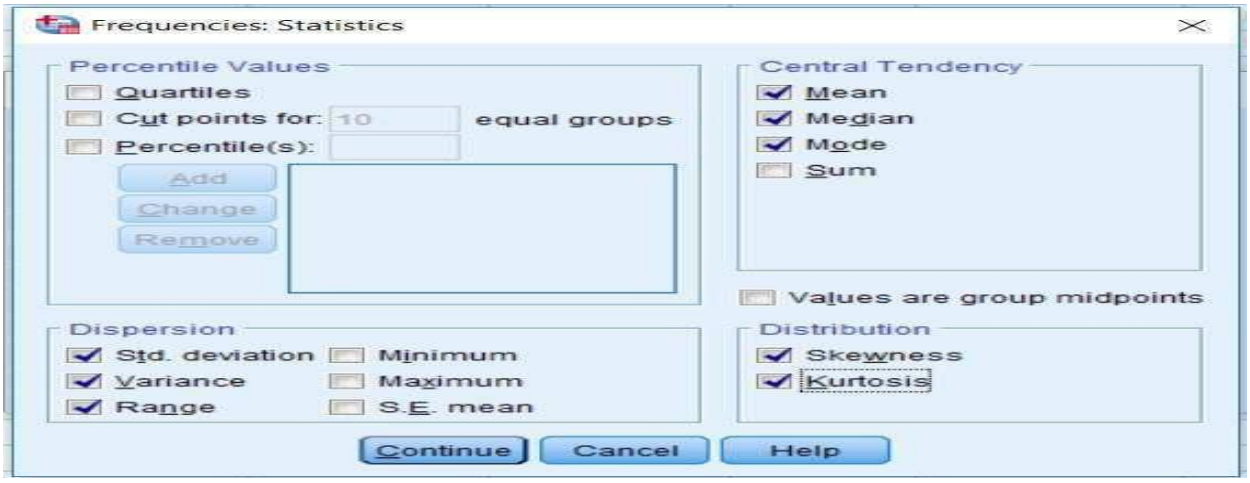



Figure (9): selecting required statistics

The output will have one frequency table for all the variables and statistics chosen and one histogram for each variable.

**Frequencies**

**Warnings**

Treat is a string so a histogram cannot be produced.

**Statistics**

| | | Hospt | Treat |
|---|---|---|---|
| N | Valid | 49 | 49 |
| | Missing | 0 | 0 |
| Mean | | 2.08 | |
| Median | | 2.00 | |
| Mode | | 1 | |
| Std. Deviation | | 1.152 | |
| Variance | | 1.327 | |
| Skewness | | 1.200 | |
| Std. Error of Skewness | | .340 | |
| Kurtosis | | 1.122 | |
| Std. Error of Kurtosis | | .668 | |
| Range | | 4 | |

Figure (10): output of the selected statistics

**Frequency Table**

**Hospt**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 1 | 18 | 36.7 | 36.7 | 36.7 |
| | 2 | 17 | 34.7 | 34.7 | 71.4 |
| | 3 | 10 | 20.4 | 20.4 | 91.8 |
| | 5 | 4 | 8.2 | 8.2 | 100.0 |
| | Total | 49 | 100.0 | 100.0 | |

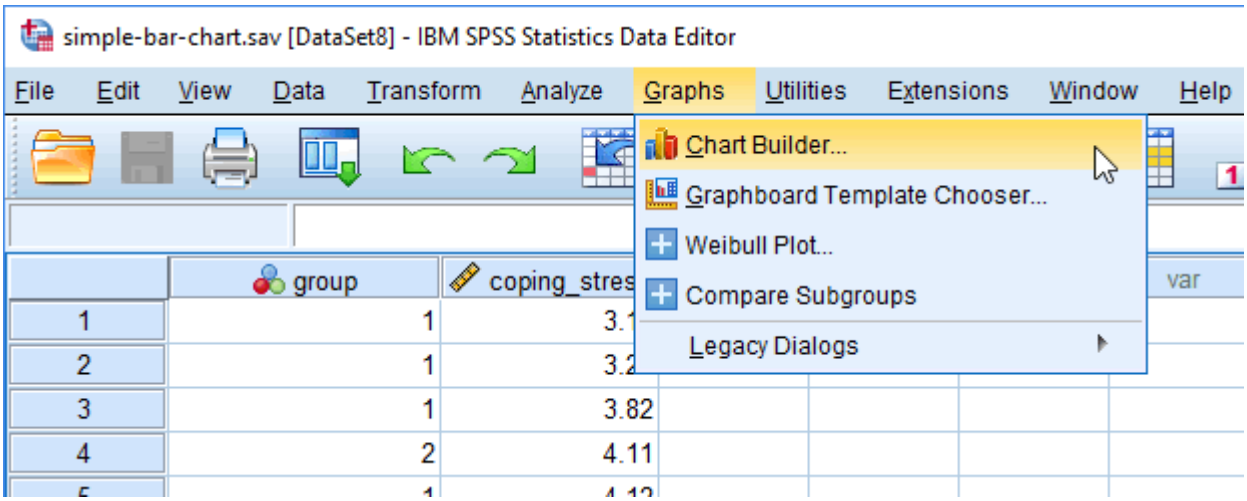
**Treat**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | Imipramine | 18 | 36.7 | 36.7 | 36.7 |
| | Lithium | 18 | 36.7 | 36.7 | 73.5 |
| | Placebo | 13 | 26.5 | 26.5 | 100.0 |
| | Total | 49 | 100.0 | 100.0 | |

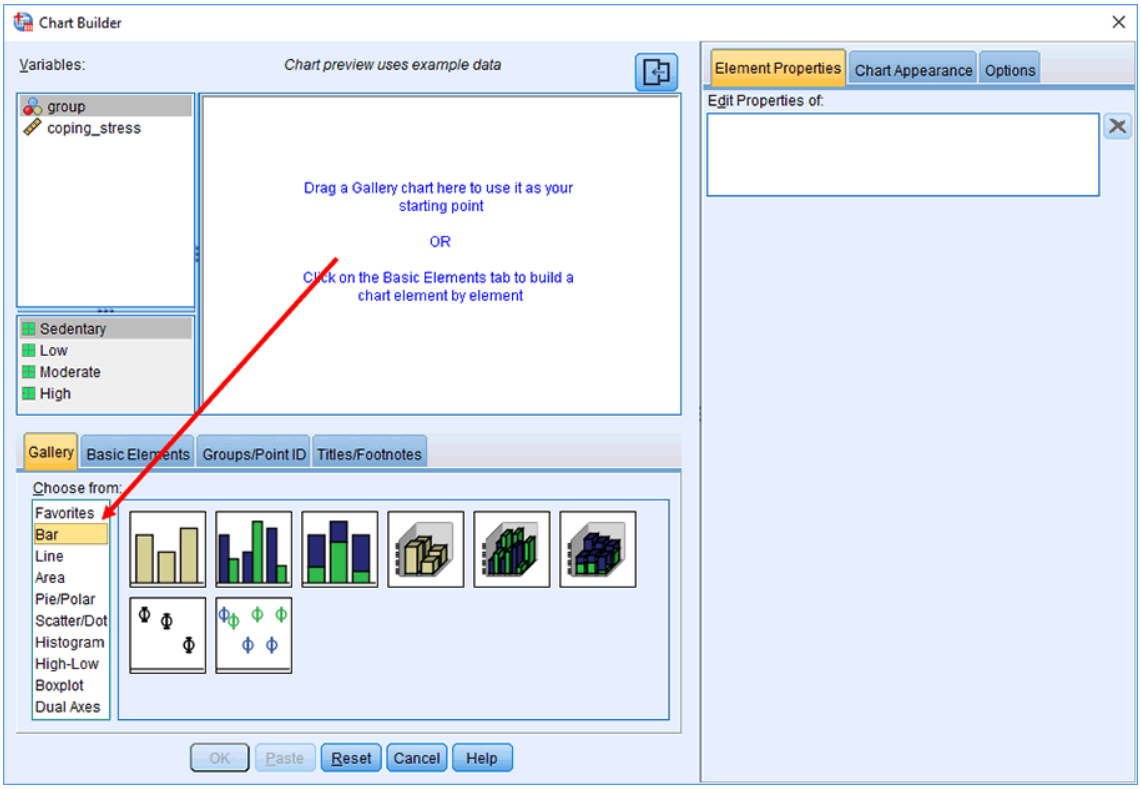
Figure (11): Frequency table of variables

**Graph in SPSS**

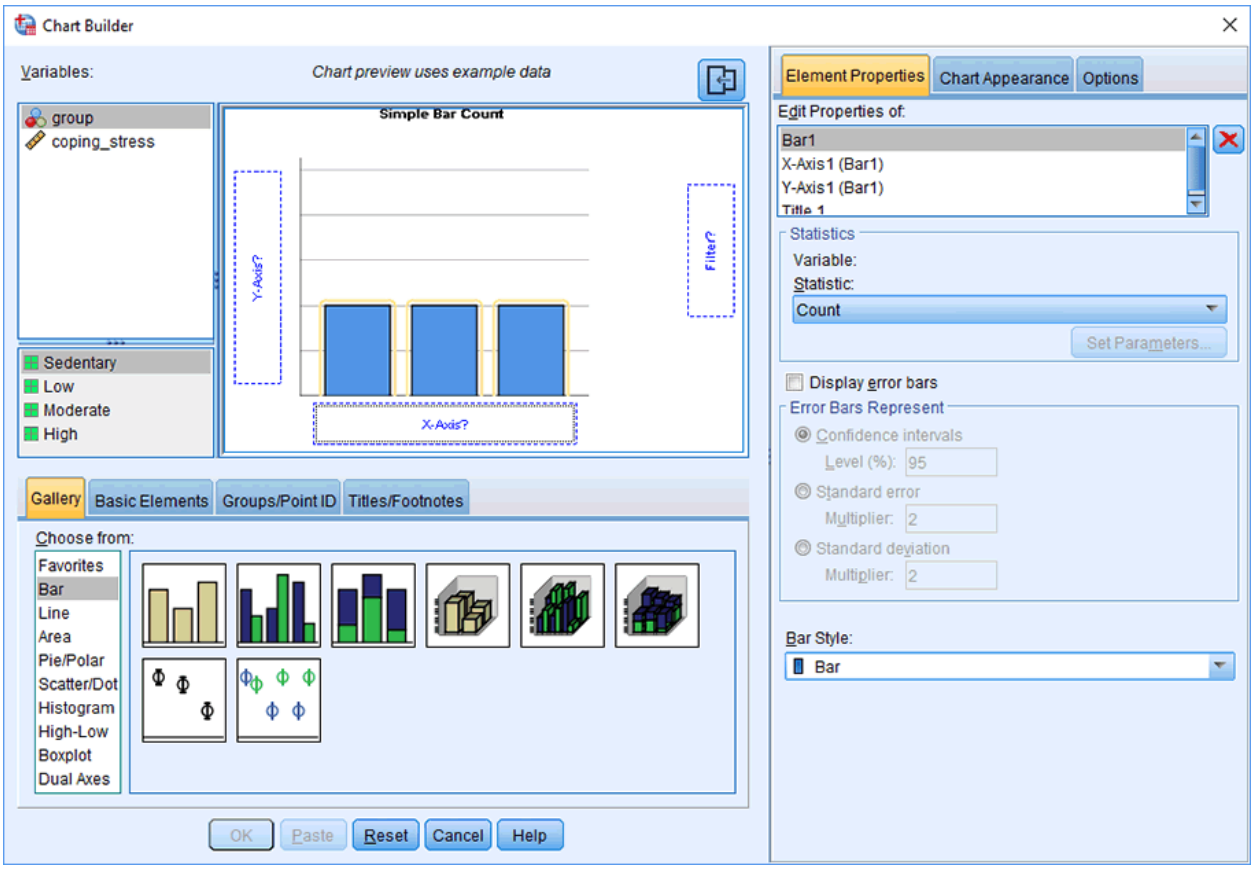
**In drawing the graph as Bar chart,** , Click **Graphs > Chart Builder...** on the main menu, as shown below:




You will be presented with the **Chart Builder** dialogue box, select "**Bar**" from the Choose from: box in the bottom-left-hand corner, as highlighted below:



Drag-and-drop the top-left-hand option (you will see it labelled as "**Simple Bar**" if you hover your mouse over the box) into the main chart preview pane, you will be presented with the screen below, which shows a simple bar chart in the main chart preview pane with boxes for the y-axis ("**Y-Axis?**") and x-axis ("**X-Axis?**") for you to populate with the appropriate variables




Drag-and-drop the independent variable, group, from the Variables: box into the "**X-Axis?**" box in the main chart preview screen and do the same for the dependent variable, coping_stress, but into the "**Y-Axis?**" box. You should end up with a screen like the one below:

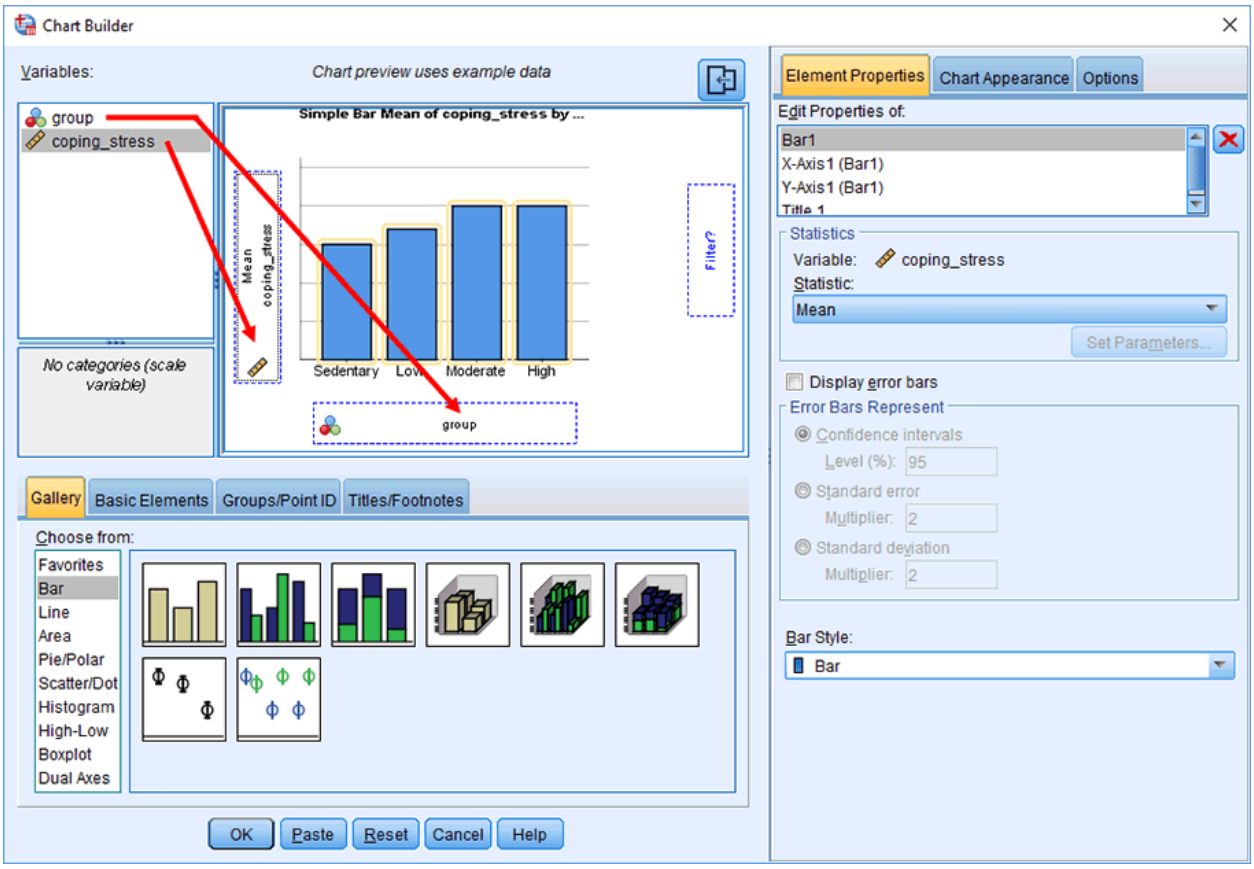


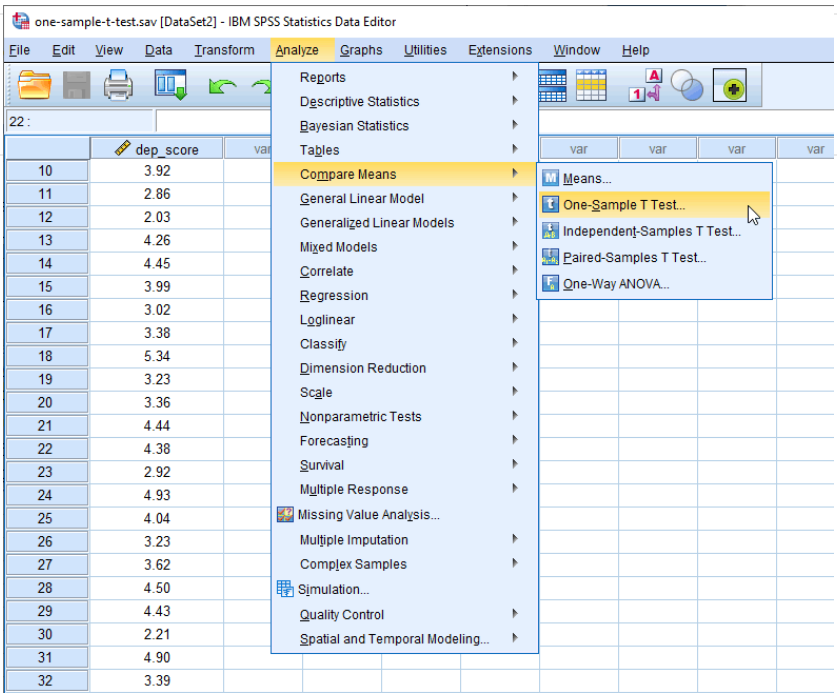
Chart Builder
Variables:
Chart preview uses example data
group
coping_stress
Simple Bar Mean of coping_stress by ...
Mean coping_stress
Sedentary
Low
Moderate
High
group
Filter?
No categories (scale variable)
Element Properties
Chart Appearance
Options
Edit Properties of:
Bar1
X-Axis1 (Bar1)
Y-Axis1 (Bar1)
Title 1
Statistics
Variable: coping_stress
Statistic:
Mean
Set Parameters...
Display error bars
Error Bars Represent
Confidence intervals
Level (%): 95
Standard error
Multiplier: 2
Standard deviation
Multiplier: 2
Bar Style:
Bar
Gallery
Basic Elements
Groups/Point ID
Titles/Footnotes
Choose from:
Favorites
Bar
Line
Area
Pie/Polar
Scatter/Dot
Histogram
High-Low
Boxplot
Dual Axes
OK
Paste
Reset
Cancel
Help

## One Sample *t* test

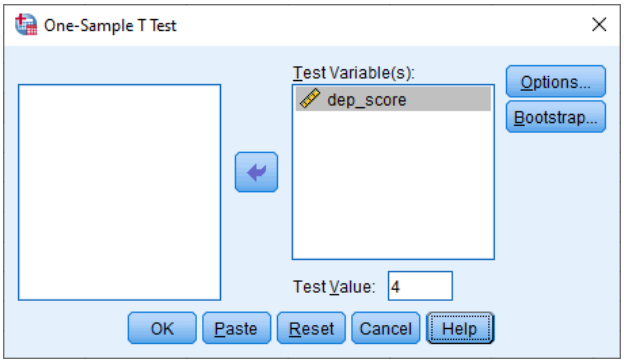
A **One Sample *t* test** compares a sample average to a population mean.

### How to Do a One Sample T Test

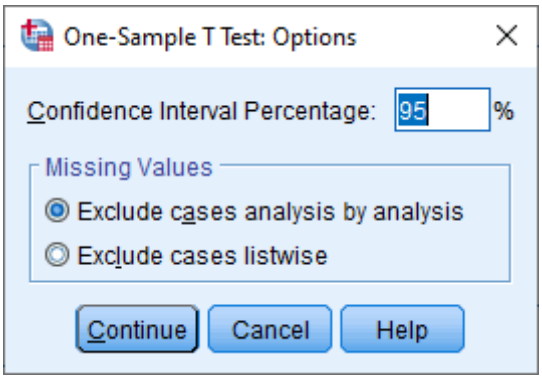
Click **Analyze > Compare Means > One-Sample T Test...** on the main menu:




You will be presented with the **One-Sample T Test** dialogue box. Then transfer the dependent variable into the Test Variable(s)




Click on the"Options" button. You will be presented with the **One-Sample T Test: Options** dialogue box, as shown below:




**The Output is as follows**

**One-Sample Test**

| | Test Value = 4 | | | | | |
|---|---|---|---|---|---|---|
| | t | df | Sig. (2-tailed) | Mean Difference | 95% Confidence Interval of the Difference | |
| | | | | | Lower | Upper |
| dep_score | -2.381 | 39 | .022 | -.27750 | -.5132 | -.0418 |

Moving from left-to-right, you are presented with the observed *t*-value ("**t**" column), the degrees of freedom ("**df**"), and the statistical significance (*p*-value) ("**Sig. (2-tailed)**") of the one-sample t-test. In this example, $p < .05$ (it is $p = .022$). Therefore, it can be concluded that the population means are statistically significantly different. If $p > .05$, the difference between the sample-estimated population mean and the comparison population mean would not be statistically significantly different.

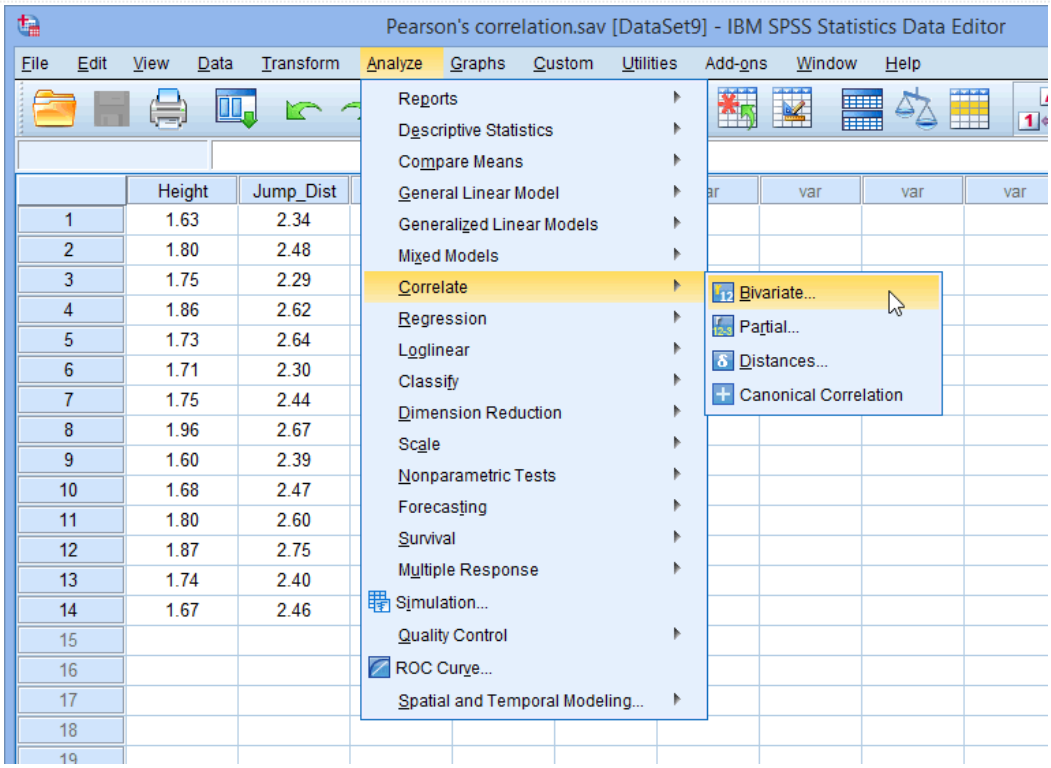
## Pearson's Correlation

The Pearson correlation coefficient is a measure of the strength and direction of association that exists between two variables measured in scale.

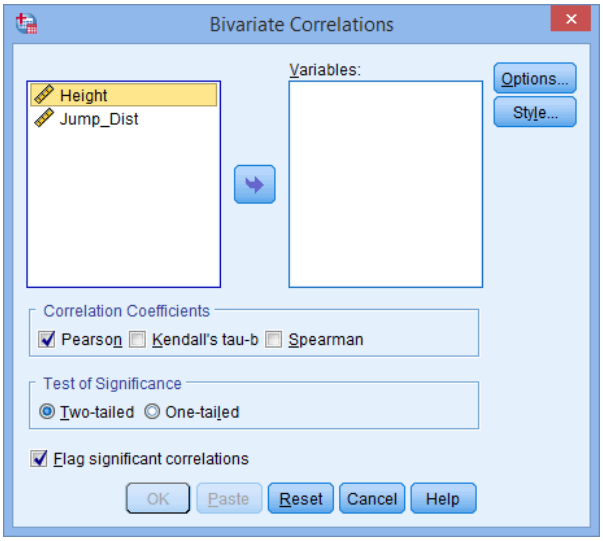
For example, you could use a Pearson's correlation to understand whether there is an association between exam performance and time spent revising. You could also use a Pearson's correlation to understand whether there is an association between depression and length of unemployment.

### Pearson's Correlation in SPSS Statistics

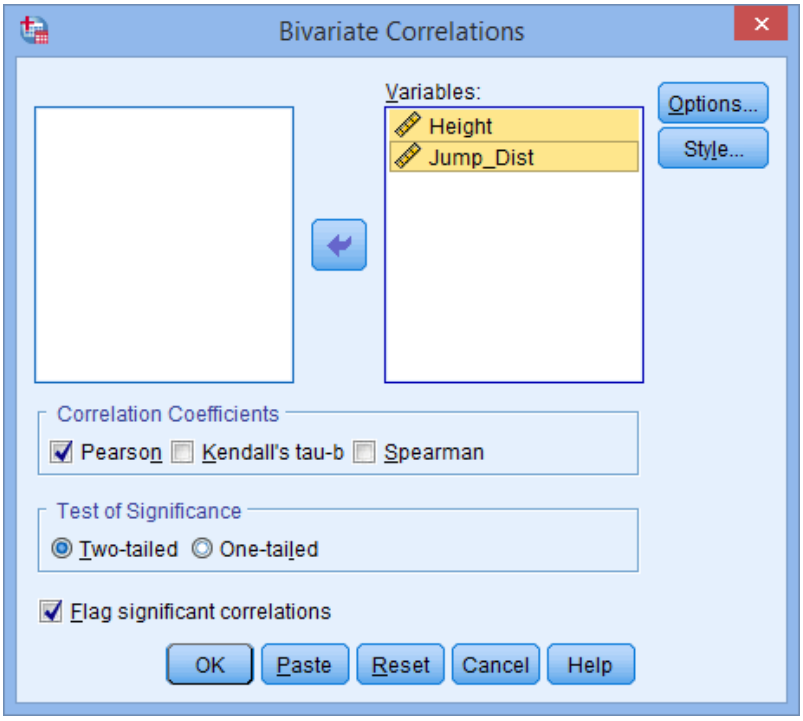
**Click** Analyze > Correlate > Bivariate...**on the main menu, as shown below:**




You will be presented with the **Bivariate Correlations** dialogue box:




Transfer the variables Height and Jump_Dist into the Variables: box by dragging-and-dropping them or by clicking on them and then clicking on the button. You will end up with a screen similar to the one below:



The output will be

**Correlations**

| | | Height | Jump_Dist |
|---|---|---|---|
| Height | Pearson Correlation | 1 | .706** |
| | Sig. (2-tailed) | | .005 |
| | N | 14 | 14 |
| Jump_Dist | Pearson Correlation | .706** | 1 |
| | Sig. (2-tailed) | .005 | |
| | N | 14 | 14 |

**. Correlation is significant at the 0.01 level (2-tailed).

we can see that the Pearson correlation coefficient( in this example), *r*, is 0.706, and that it is statistically significant (*p* = 0.005).