

Descriptive Statistics

Using a simple data set, this tutorial will show you how to produce your first set of statistical output using SPSS.

The data in this example is loosely based on the evaluation of the Schools Linking Network. As the name suggests, this project links schools in different communities to put 'the contact hypothesis' into practice. The contact hypothesis suggests that by increasing contact between people from diverse backgrounds, prejudice can be reduced and positive attitudes towards 'out-groups' can be fostered.

The data can be found in the SPSS file: 'Week 4 data file.sav' and looks like this:

	Id_No	Gender	Ethnicity	Enjoyment	Respect1	Respect2	var	var
1	1	2	1	5	65	75		
2	2	1	1	3	81	90		
3	3	2	5	1	66	77		
4	4	2	1	5	80	87		
5	5	2	1	4	91	96		
6	6	2	1	3	55	74		
7	7	2	2	5	66	72		
8	8	2	1	2	56	64		
9	9	2	1	5	75	83		
10	10	2	1	3	92	96		
11	11	2	1	1	56	61		
12	12	1	3	3	73	82		
13	13	2	3	3	78	85		
14	14	1	1	4	63	72		
15	15	2	3	4	84	92		
16	16	1	5	2	68	81		
17	17	2	1	4	76	84		

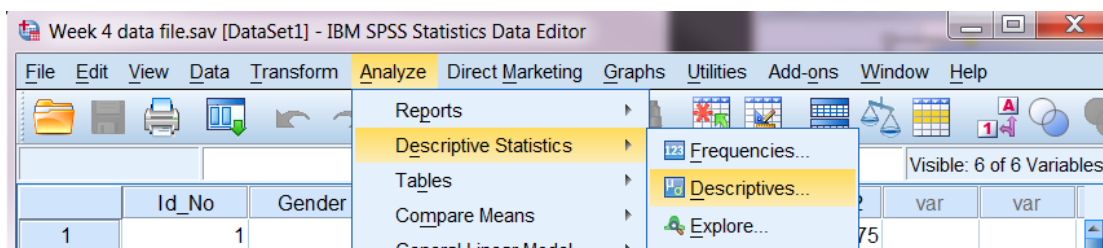
Each column represents a different variable, and each row contains the data from one participant. The different columns display the following data:

- **ID_No:** This refers to the ID number assigned to the participants. We use these numbers as identifiers instead of participant names, as this allows us to collect data while keeping the participants anonymous. As anonymity is ethically important in psychology research, this is generally considered good practice.
- **Gender:** This column contains information about participants' gender. As this is a categorical (or nominal) variable, the data needs to be entered using numerical codes to represent different groups. In this case we have used the code '1' for males and '2' for females. Refer to the earlier tutorial 'Adding Variables' to see how this coding is done.

- **Ethnicity:** This column contains information about participants' ethnicity. Like gender, this is a categorical (or nominal) variable, as participants belong to different groups or categories. Each category is coded with a number, and this number is used to represent each ethnic group.
- **Enjoyment:** This variable measures pupil's enjoyment of meeting new people through the Linking School Network on a scale of 1-5 (where 1 = did not enjoy at all and 5 = really enjoyed). As this is a single scale with 5 ordered categories to choose from, it's an *ordinal* variable.
- **Respect1:** The Linking School Network aims to reduce prejudice between different groups. To assess this, a questionnaire that measures participants' respect for the rights of others was given to students both before and after their participation in the program. This variable represents their respect for others *before* the intervention. As this is measured using a standardised questionnaire, this is an interval variable.
- **Respect2:** Following Respect1, this variable represents participants' respect for others *after* the intervention. Again, this is an interval variable.

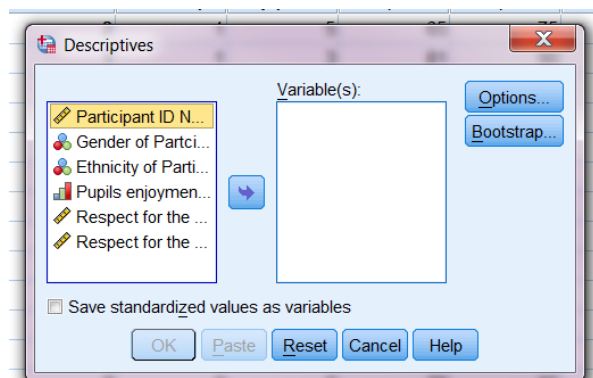
Now you have had a chance to explore what the different variables are, you need a way to inspect and summarise the data, so you can get a better idea of any patterns that may exist within it. You can do this by looking at descriptive statistics for the data.

To obtain descriptive statistics you first need to **CLICK** on the **Analyze** menu, select the **Descriptive Statistics** option, and then the **Descriptives** sub-option.



The following **Descriptives** dialog box should now open:

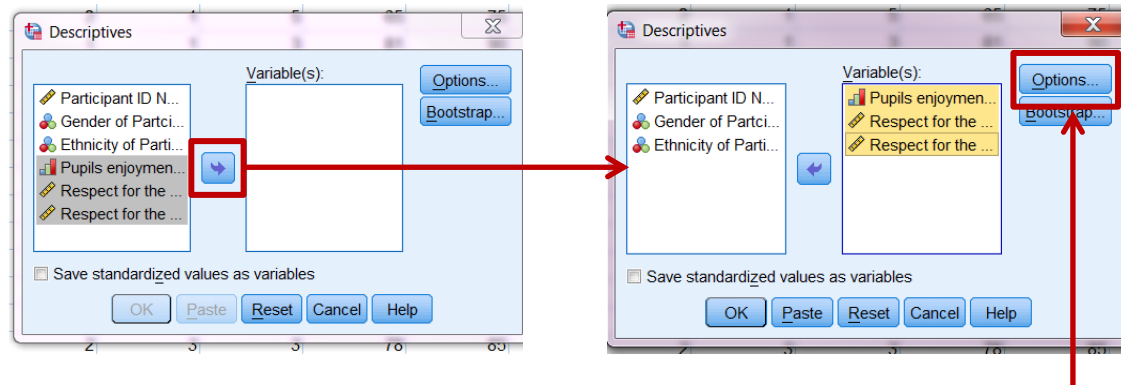
It has the names of your variables in the pane on the left hand side; and an empty pane on the right hand side labelled **Variable(s)**. To tell SPSS which variables you want summary scores for, you need to move the variables from the left pane to the right. This process is quite common for many of the analyses you will undertake using SPSS.



When producing descriptive statistics, you only want to do so for interval and ordinal variables. It doesn't make sense to ask SPSS to produce means and standard deviations for categorical (or nominal) data.

Interval variables are represented by a ruler symbol ; ordinal variables by a bar chart ; and categorical by the three circles .

Holding down the CTRL key on your keyboard, select your three variables of interest (Enjoyment, Respect1 and Respect2) and move them across to the right-hand pane using the blue arrow button in the centre of the dialog box.



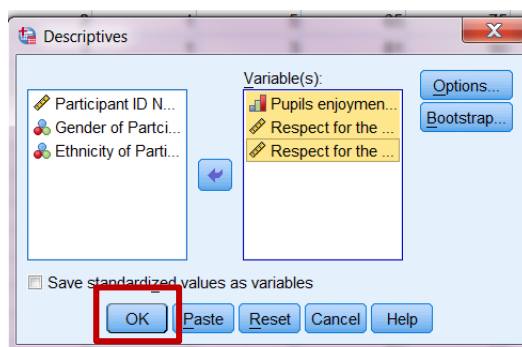
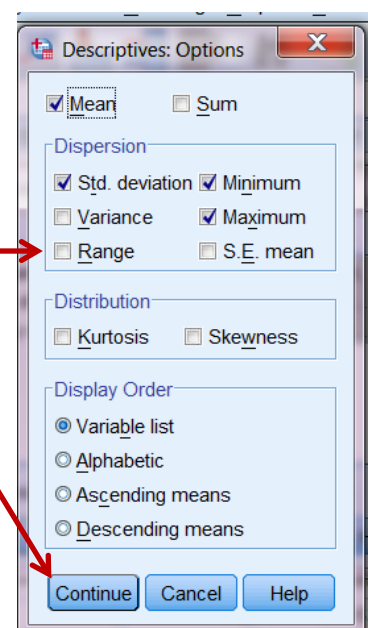
Once you have selected all of the variables you want to explore, click the **Options...** button to select and/or change the statistics SPSS will display.

This opens another dialog box up. Here you can see which descriptive statistics are displayed by default, and change them if you wish.

For the purpose of this tutorial, let's add the **Range** item to our output. To do this, **CLICK** on the check box to the left of this option, so the box becomes ticked.

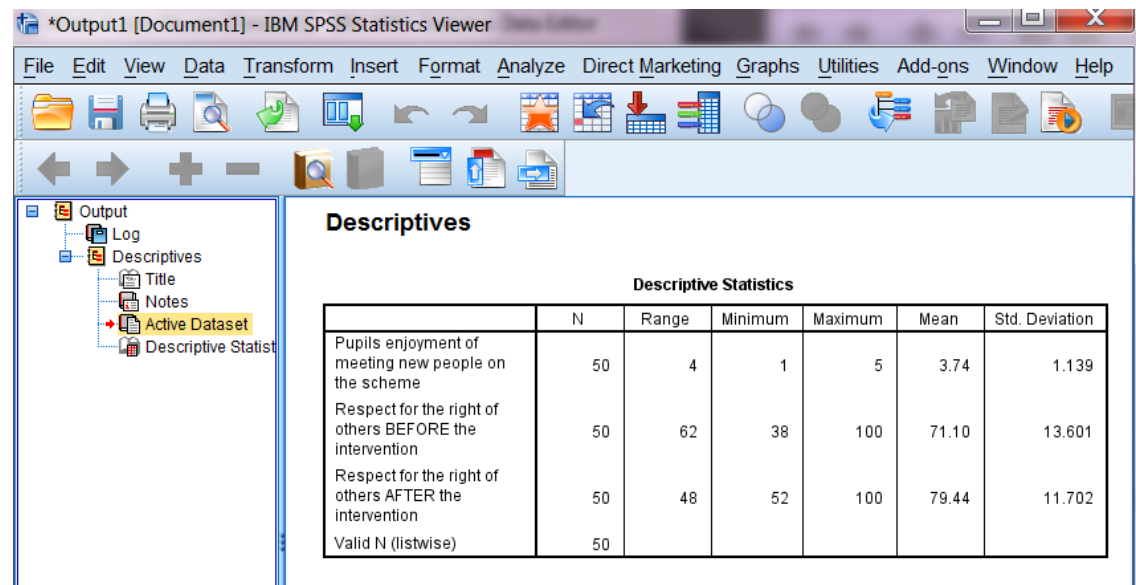
To save any changes you have made, click on the **Continue** button.

Note: if you click **Cancel** the Range will no longer be selected and so this descriptive statistic will not be displayed in your SPSS Output.



This will return you back to your **Descriptives** dialog box, where you need to **CLICK** on the **OK** button to display your statistics.

Once you have told SPSS which statistics you would like to see, the program displays them in a separate Output window like this one:



The screenshot shows the IBM SPSS Statistics Viewer window with the title '*Output1 [Document1] - IBM SPSS Statistics Viewer'. The menu bar includes File, Edit, View, Data, Transform, Insert, Format, Analyze, Direct Marketing, Graphs, Utilities, Add-ons, Window, and Help. The toolbar contains various icons for file operations and data manipulation. On the left, a tree view shows the output structure: Output, Log, Descriptives, Title, Notes, Active Dataset (highlighted), and Descriptive Statistics. The main area displays the 'Descriptives' table.

	Descriptive Statistics					
	N	Range	Minimum	Maximum	Mean	Std. Deviation
Pupils enjoyment of meeting new people on the scheme	50	4	1	5	3.74	1.139
Respect for the right of others BEFORE the intervention	50	62	38	100	71.10	13.601
Respect for the right of others AFTER the intervention	50	48	52	100	79.44	11.702
Valid N (listwise)	50					

The **Descriptive Statistics** table displays all of the information that you have requested. The left-most column tells you which row relates to which variable. The full Variable Labels (rather than abbreviated Variable Names) are displayed by default here. 'Valid N (listwise)' is a label for the number of 'valid' cases in the dataset. The rest of the columns represent the following:

- **N:** 'N' stands for 'number of participants'. This column simply tells you how many participants you have data for for each variable (excluding missing data). 'Valid N (listwise)' displays the number of participants who you have data for on all measures.
- **Range:** This column measures the spread of the scores obtained. Essentially the range is the difference between the lowest and the highest value.
- **Minimum:** This refers to the lower end of the range. In the minimum column is the lowest score of the variable in each row.
- **Maximum:** This refers to the upper end of the range. In the maximum column is the highest score of the variable in each row.
- **Mean:** The Mean column shows the mathematical average for each of the variables. This is the sum of the scores divided by the number of scores (the N).
- **Std. Deviation:** This column displays the Standard Deviation for each variable. This refers to the spread of the scores around the mean, and represents how much variation you have in the data. The larger the values relative to the means, the more dispersed the scores are.

When carrying out research you should always report the means and standard deviations of your variables, as it gives instant meaning to your data. They are usually reported to 2 decimal places.

Descriptive Statistics						
	N	Range	Minimum	Maximum	Mean	Std. Deviation
Pupils enjoyment of meeting new people on the scheme	50	4	1	5	3.74	1.139
Respect for the right of others BEFORE the intervention	50	62	38	100	71.10	13.601
Respect for the right of others AFTER the intervention	50	48	52	100	79.44	11.702
Valid N (listwise)	50					

In this case, looking at the table you could say:

- On average pupil's seemed to enjoy the Linking Schools Project, with a mean Enjoyment score of **3.74** (SD = **1.14**).
- In addition, participant's Respect scores before taking part in the project appeared lower (mean = **71.10**; SD = **13.60**) than afterwards (mean = **79.44**; SD = **11.70**).

Now you have been shown how to enter data and produce descriptive statistics, why don't you try creating your own data set and calculating the means, standard deviations and other descriptive statistics? Or download the data file used in this tutorial (Week 4 data file.sav) and see if you can produce the same output yourself.

There are also a number of other ways you can produce summary statistics for your variables. For example, you can use the following menu options:

- **Analyse -> Descriptive Statistics -> Frequencies...**
- **Analyse -> Descriptive Statistics -> Explore...**

Why not explore these options and see what output you can produce?

Practicing using SPSS yourself will help to increase your confidence with using the program and analysing statistics, so we really do recommend that you take advantage of any opportunities for practice that are available!