Correlations in SPSS (Quiz)

Correlation practical questions

In this practical we will investigate whether there is a relationship between two variables by looking how correlated they are.

Family background, as well as psychological traits, has also been shown to be an important predictor of educational achievement. Family background encompasses many different dimensions of parental resources which do not necessarily vary perfectly together. In this example you will explore the association between an indicator of a family's wealth and the degree of emotional support provided by parents for a child's learning. PISA's wealth variable, WEALTH, is derived from reports of whether the family owns eight items, such as a car, a computer and a room of the child's own. The parental emotional support variable, EMOSUPS, is derived from four items with which students rated their strength of agreement, e.g. "My parents support my educational efforts and achievements" (see PISA datafile description for further details).

Firstly use SPSS to create a scatterplot of **WEALTH** and **EMOSUPS** and answer the following:

- Question: What does the scatterplot say about the relationship between **WEALTH** and **EMOSUPS**?

Next use SPSS and the Explore screen to create histograms, normality tests and QQ plots of **WEALTH** and **EMOSUPS** and answer the following:

- Question: What do the plots and tests tell us about the normality of **WEALTH**?
- Question: What do the plots and tests tell us about the normality of **EMOSUPS**?

Next use SPSS and the Correlate screen to answer the following:

- Question: What is the Pearson correlation coefficient between **WEALTH** and **EMOSUPS** and is it significant?

Next use SPSS and the Correlate screen to answer the following:

- Question: What is the Spearman correlation coefficient between **WEALTH** and **EMOSUPS** and is it significant?

Next use SPSS and the Correlate screen to answer the following:

- Question: What is the value of the Kendall tau-b correlation coefficient between **WEALTH** and **EMOSUPS** and is it significant?

Solutions to Correlation practical questions

The SPSS instructions are as follows:

- Select **Scatter/Dot** from the **Legacy diagnostics** available from the **Graphs** menu.
- Select Simple Scatter and click on Define to bring up the Simple Scatterplot window.
- Copy the **Family wealth score[WEALTH]** variable into the **Y Axis** box.
- Copy the **Parental emotional support score[EMOSUPS]** variable into the **X Axis** box.
- Click on the **OK** button.

- Question: What does the scatterplot say about the relationship between **WEALTH** and **EMOSUPS**?

Solution: The output from SPSS is as follows:
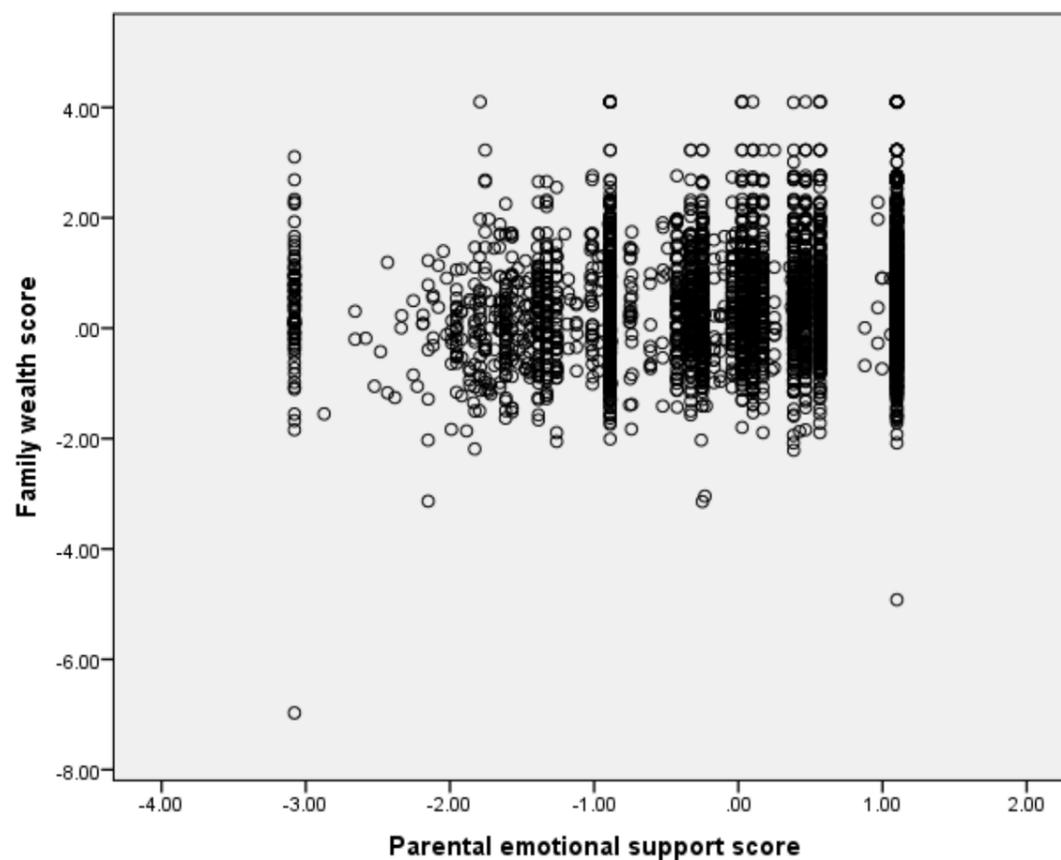
Looking at the scatterplot there appears to be a positive correlation between the variables with larger values of **WEALTH** associated with larger values of **EMOSUPS** (an upward sloping relationship) but this relationship is not that strong with possibly a few more points in the bottom-left and top-right quarters of the plot.

The SPSS instructions are as follows:

- Select **Descriptive Statistics** from the **Analyze** menu.
- Select **Explore** from the **Descriptive Statistics** sub-menu.
- Click on the **Reset** button.
- Copy the **Family wealth score[WEALTH]** and **Parental emotional support score[EMOSUPS]** variables into the **Dependent List:** box.
- Click on the **Plots...** button.
- On the screen that appears select the **Histogram** tick box.
- Unselect the **Stem and leaf** button.
- Select the **Normality plots with tests** button.
- Click on the **Continue** button.
- Click on the **OK** button.

Question: What do the plots and tests tell us about the normality of **WEALTH**?

Solution: The output from SPSS is as follows:

We will first look at a histogram of the variable, **WEALTH**.

Histogram

Mean = .52
Std. Dev. = .981
N = 5,044

Family wealth score
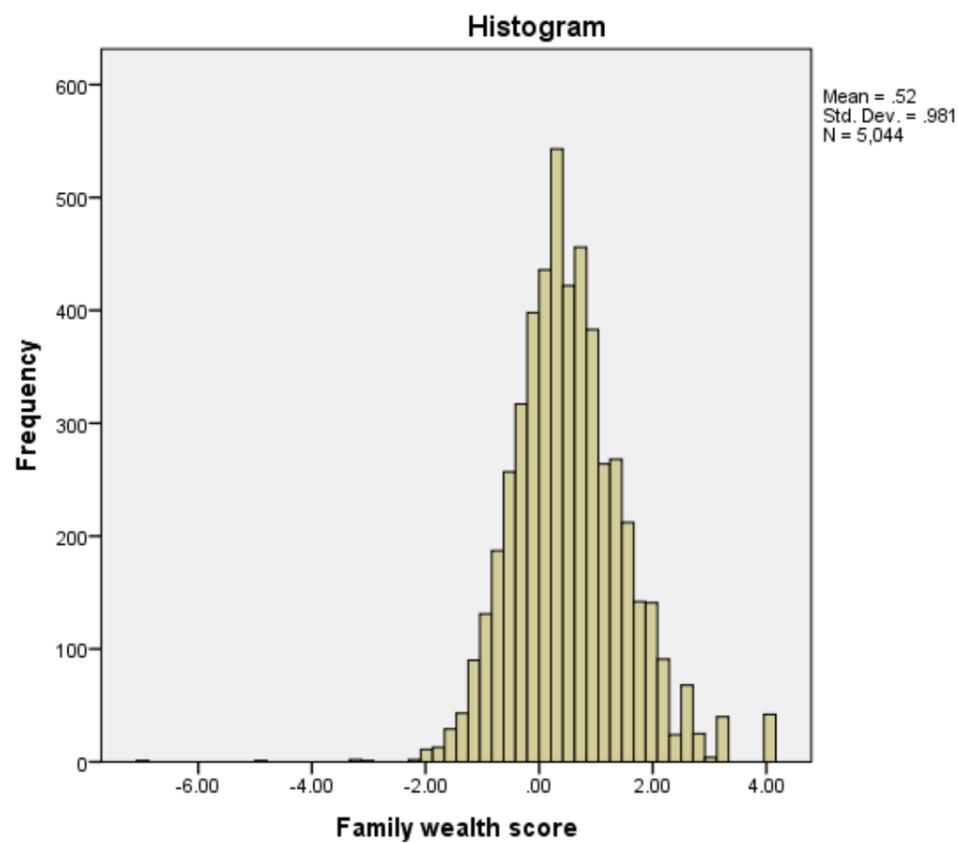
Ideally for a normal distribution this histogram should look symmetric around the mean of the distribution, in this case .5245. This distribution appears to be significantly skewed to the right (positively skewed).

Next we look at the Normality test statistics:

Tests of Normality

|	Kolmogorov-Smirnov[a]		
	Statistic	df	Sig.
Family wealth score	.048	5044	.000
Parental emotional support score	.193	5044	.000

a. Lilliefors Significance Correction

The Kolmogorov Smirnov tests produce test statistics that are used (along with a degrees of freedom parameter) to test for normality. Here we see that the Kolmogorov Smirnov statistic takes value .048 for **WEALTH**. The test has degrees of freedom which equals the number of data points, namely 5044.

For **WEALTH** we see the following: The p value (quoted under Sig. for Kolmogorov Smirnov) is .000 (reported as p < .001) which is less than 0.05. We therefore have significant evidence to reject the null hypothesis that the variable follows a normal distribution.

Although the Kolmogorov Smirnov test tells the researcher whether the distribution followed by a variable is statistically significantly different from a normal distribution one should take care in not overinterpreting such findings. Significance will be strongly effected by the number of observations and so only a small discrepancy from normality will be deemed significant for very large sample sizes whilst very large discrepancies will be required to reject the null hypothesis for small sample sizes.In addition, Pearson's correlation will be robust to non-normality in the data when samples are very large, as is the case here.

For **WEALTH** its Quantile-Quantile plot can be seen below:

Normal Q-Q Plot of Family wealth score
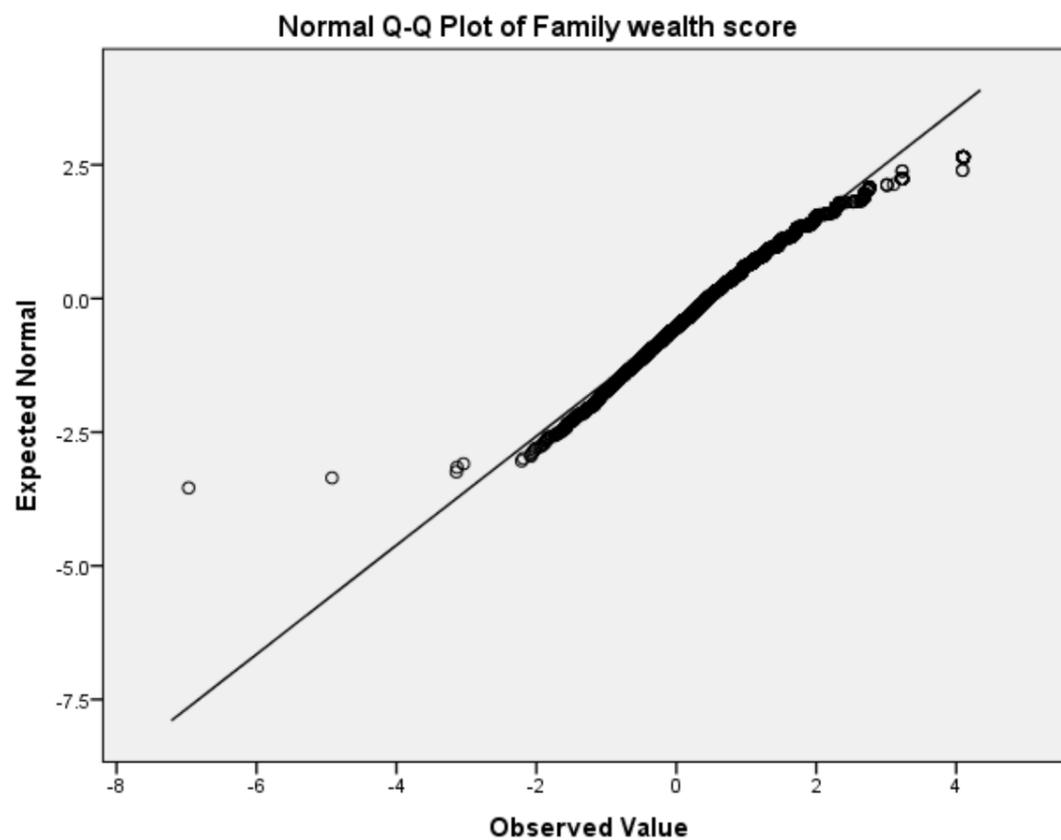
QQ plots can be used to compare the distribution of a variable with a chosen distribution (typically a normal distribution as we are doing here). The data are plotted against a theoretical normal distribution (with the same mean and variance as the sample data) in such a way that the points should form an approximate straight line. Departures from this straight line indicate departures from normality. As we found a significant effect in the Kolmogorov Smirnov test for **WEALTH** we should see the points diverging from the line in the plot above with either some outlying values lying away from the line or even the shape of the points forming a non-linear pattern.

Question: What do the plots and tests tell us about the normality of **EMOSUPS**?

Solution: The output from SPSS is as follows:

We will first look at a histogram of the variable, **EMOSUPS**.

Histogram

Mean = .10
Std. Dev. = .98
N = 5,044
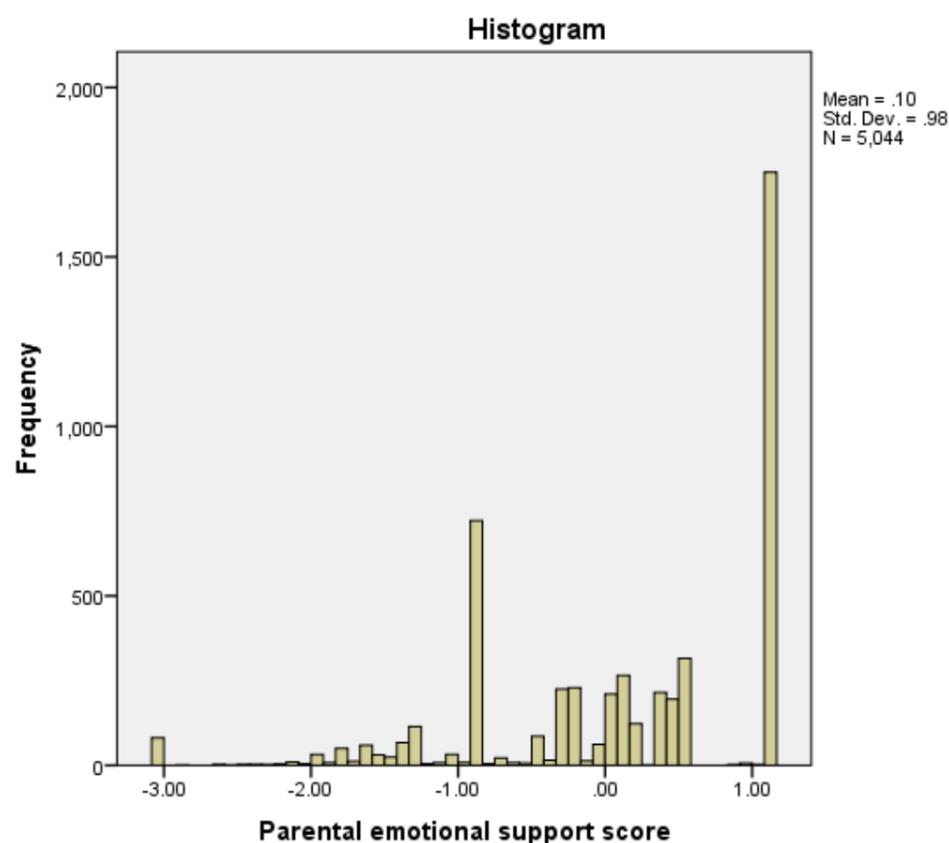
Again for a normal distribution this histogram should look symmetric around the mean of the distribution, in this case .1019. This distribution appears to be significantly skewed to the left (negatively skewed).

Next we look at the Normality test statistics:

Tests of Normality

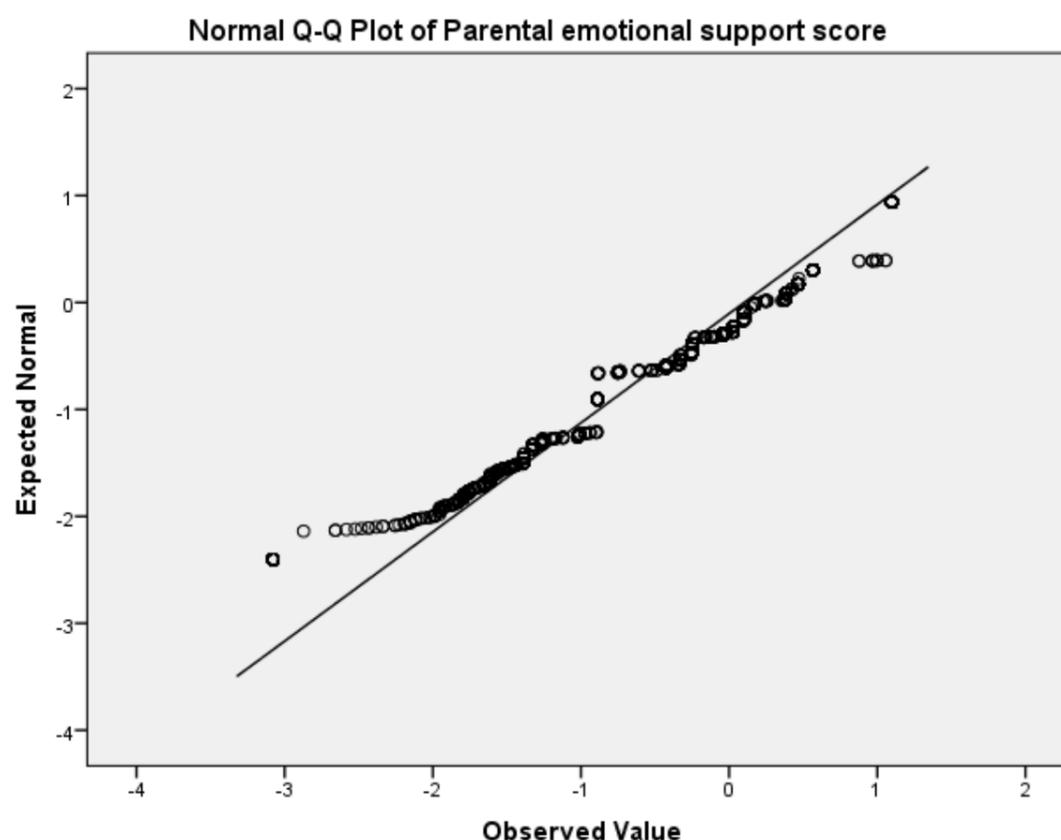|	Kolmogorov-Smirnov[a]		
	Statistic	df	Sig.
Family wealth score	.048	5044	.000
Parental emotional support score	.193	5044	.000

a. Lilliefors Significance Correction

The Kolmogorov Smirnov tests produce test statistics that are used (along with a degrees of freedom parameter) to test for normality. Here we see that the Kolmogorov Smirnov statistic takes value .193 for **EMOSUPS**. The test has degrees of freedom which equals the number of data points, namely 5044.

For **EMOSUPS** we see the following: The p value (quoted under Sig. for Kolmogorov Smirnov) is .000 (reported as p < .001) which is less than 0.05. We therefore have significant evidence to reject the null hypothesis that the variable follows a normal distribution.

Although the Kolmogorov Smirnov test tells the researcher whether the distribution followed by a variable is statistically significantly different from a normal distribution one should take care in not overinterpreting such findings. Significance will be strongly effected by the number of observations and so only a small discrepancy from normality will be deemed significant for very large sample sizes whilst very large discrepancies will be required to reject the null hypothesis for small sample sizes. In addition, Pearson's correlation will be robust to non-normality in the data when samples are very large, as is the case here.

For **EMOSUPS** its Quantile-Quantile plot can be seen below:

Normal Q-Q Plot of Parental emotional support score

As we found a significant effect in the Kolmogorov Smirnov test for **EMOSUPS** we should see the points diverging from the line in the plot above with either some outlying values lying away from the line or even the shape of the points forming a non-linear pattern.
The SPSS instructions are as follows:

- Select **Bivariate...** from the **Correlate** option available from the **Analyse** menu.
- Copy the **Family wealth score[WEALTH]** and the **Parental emotional support score[EMOSUPS]** variables into the **Variables** box.
- Click on the **Options** button and Select the **Means and Standard deviations** tick box.
- Click on the **Continue** button to return to main window.
- Click on the **OK** button.

- Question: What is the Pearson correlation coefficient between **WEALTH** and **EMOSUPS** and is it significant?

Solution: The output from SPSS is as follows:

Descriptive Statistics

	Mean	Std. Deviation	N
Family wealth score	.5213	.98857	5082
Parental emotional support score	.1013	.98013	5053

Correlations

		Family wealth score	Parental emotional support score
Family wealth score	Pearson Correlation	1	.177**
	Sig. (2-tailed)		.000
	N	5082	5044
Parental emotional support score	Pearson Correlation	.177**	1
	Sig. (2-tailed)	.000	
	N	5044	5053

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

Here we are interested in the Pearson correlation between **WEALTH** and **EMOSUPS** which can be found in two places in the table - either in the row for **WEALTH** and column for **EMOSUPS** or the row for **EMOSUPS** and column for **WEALTH**.

In this case the correlation takes value .177. This correlation is small but positive. The correlation is given in the table, along with a significance value and a sample size which in this case is 5044. This is the number of observations in which both **WEALTH** and **EMOSUPS** where observed.

We can test if this correlation is significantly different from zero which will depend on (i) the magnitude of the correlation and (ii) the number of observations on which the correlation is based.

The p value (quoted under Sig. (2-tailed)) is .000 (reported as $p < .001$) which is less than 0.05. We therefore have significant evidence to reject the null hypothesis that the correlation is 0.
The SPSS instructions are as follows:

- Select **Bivariate...** from the **Correlate** option available from the **Analyse** menu.
- Check that the **Family wealth score[WEALTH]** and the **Parental emotional support score[EMOSUPS]** variables are still in the **Variables** box.
- Deselect the **Pearson** tick box.
- Select the **Spearman** tick box.
- Click on the **OK** button.

- Question: What is the Spearman correlation coefficient between **WEALTH** and **EMOSUPS** and is it significant?

Solution: The output from SPSS is as follows:

Correlations

			Family wealth score	Parental emotional support score
Spearman's rho	Family wealth score	Correlation Coefficient	1.000	.178**
		Sig. (2-tailed)	.	.000
		N	5082	5044
	Parental emotional support score	Correlation Coefficient	.178**	1.000
		Sig. (2-tailed)	.000	.
		N	5044	5053

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

Here we are interested in the Spearman correlation between **WEALTH** and **EMOSUPS** which can be found in two places in the table - either in the row for **WEALTH** and column for **EMOSUPS** or the row for **EMOSUPS** and column for **WEALTH**.

In this case the correlation takes value .178. This correlation is small but positive. The correlation is given in the table, along with a significance value and a sample size which in this case is 5044. This is the number of observations in which both **WEALTH** and **EMOSUPS** where observed.

We can test if this correlation is significantly different from zero which will depend on (i) the magnitude of the correlation and (ii) the number of observations on which the correlation is based.

The p value (quoted under Sig. (2-tailed)) is .000 (reported as $p < .001$) which is less than 0.05. We therefore have significant evidence to reject the null hypothesis that the correlation is 0.
The SPSS instructions are as follows:

- Select **Bivariate...** from the **Correlate** option available from the **Analyse** menu.
- Check that the **Family wealth score[WEALTH]** and the **Parental emotional support score[EMOSUPS]** variables are still in the **Variables** box.
- Deselect the **Spearman** tick box.
- Select the **Kendall tau-b** tick box.
- Click on the **OK** button.

- Question: What is the value of the Kendall tau-b correlation coefficient between **WEALTH** and **EMOSUPS** and is it significant?

Solution: The output from SPSS is as follows:

Correlations

			Family wealth score	Parental emotional support score
Kendall's tau_b	Family wealth score	Correlation Coefficient	1.000	.127**
		Sig. (2-tailed)	.	.000
		N	5082	5044
	Parental emotional support score	Correlation Coefficient	.127**	1.000
		Sig. (2-tailed)	.000	.
		N	5044	5053

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

Here we are interested in the Kendall Tau-b correlation between **WEALTH** and **EMOSUPS** which can be found in two places in the table - either in the row for **WEALTH** and column for **EMOSUPS** or the row for **EMOSUPS** and column for **WEALTH**.

In this case the correlation takes value .127. This correlation is small but positive. The correlation is given in the table, along with a significance value and a sample size which in this case is 5044. This is the number of observations in which both **WEALTH** and **EMOSUPS** where observed.

We can test if this correlation is significantly different from zero which will depend on (i) the magnitude of the correlation and (ii) the number of observations on which the correlation is based.

The p value (quoted under Sig. (2-tailed)) is .000 (reported as p < .001) which is less than 0.05. We therefore have significant evidence to reject the null hypothesis that the correlation is 0.

In this example, Pearson's correlation is probably the most appropriate statistic to report, given the continuous nature of the variables and the very large sample size. However, the results all agree that there is a significant positive association between the indicator of parental wealth and the degree of emotional support provided for their child's learning, but this is actually rather weak. Wealthier parents do tend to provide greater learning-related psychological support for their children, but high levels on one construct certainly do not guarantee high levels on the other.