Introduction to correlation

Testing whether or not a correlation is signficant

Null Hypothesis: There is no correlation between two variables X and Y

Alternative Hypothesis: There is a correlation between two variables X and Y

Degrees of Freedom: Number of data points minus two (lose a degree of freedom for calculating each mean)

To test whether or not to reject the null hypothesis, we can obtain a p-value in SPSS or use a table of critical values to look up the value of r for a specific degrees of freedom

Testing whether or not a correlation is signficant

We often want to test whether or not the correlation between two variables is significant

- Test of Pearson product moment correlation assumes variables are normally distributed
- Test of Spearman's rank correlation coefficient (i.e., correlation of ranks) does not assume normality

To test whether or not two variables are correlated, we first must test the null hypothesis that the two variables are normally distributed.

A data set of soil depths and root densities

Sample number	Soil depth (m)	Root density (g per m^3)
1	0.8	13
2	2.0	8
3	2.3	4
4	2.7	6
5	0.5	18
6	1.8	7
7	1.5	9
8	2.1	3
9	1.2	7
10	1.1	10

In the above table, soil depth is measured in metres and root density is measured in grams per cubic metre.

Testing for normality in SPSS

Select ‘Analyse > Descriptive Statistics > Explore’, the move both variables to the Dependent list, click ‘Plots’ and make sure to select ‘Normality plots with tests’ as below

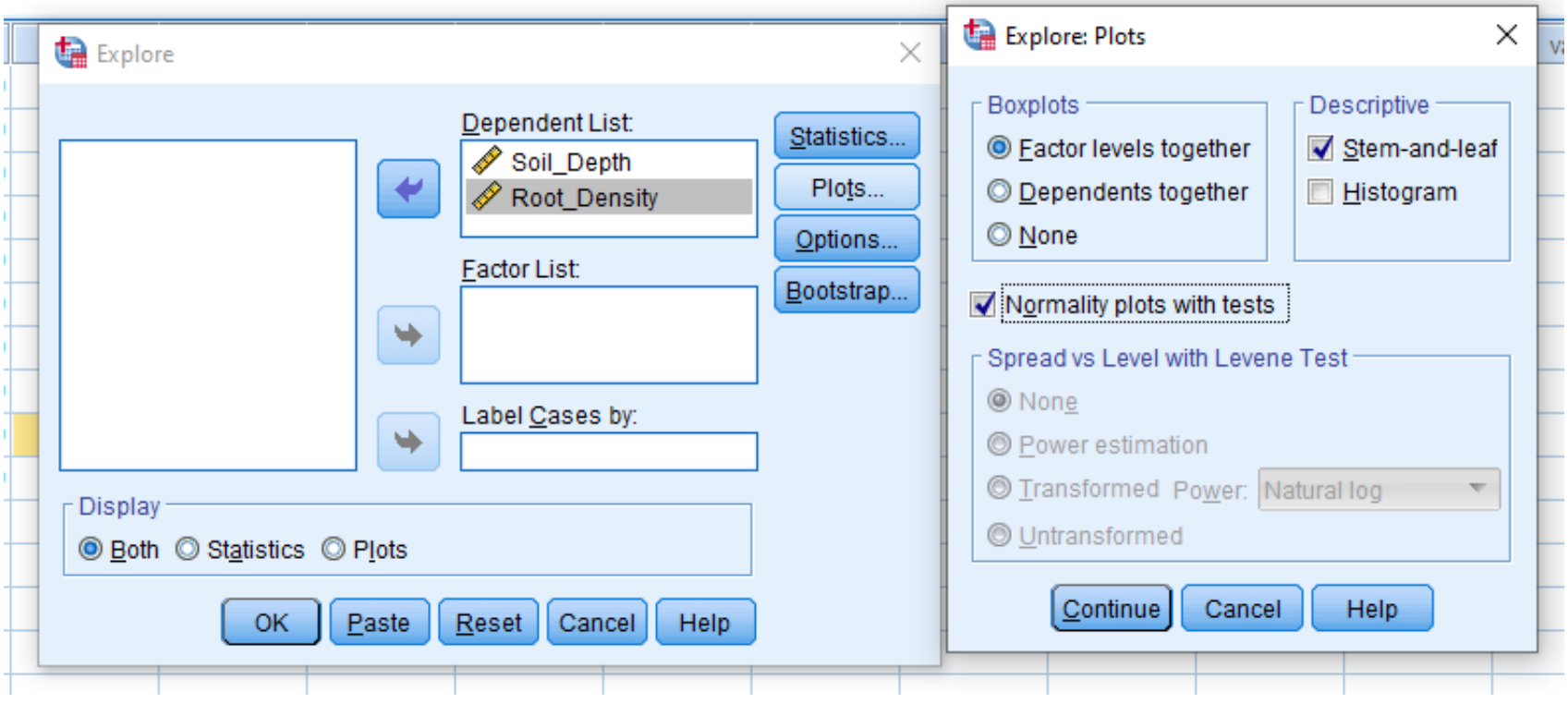

Figure 1: Windows in SPSS for testing the null hypothesis that variables are normally distributed.

Testing for normality in SPSS

Below shows the output of the tests for normality in SPSS (try to recreate these for yourself).

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Soil_Depth	.116	10	.200*	.979	10	.960
Root_Density	.167	10	.200*	.926	10	.407

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Figure 2: SPSS output for tests of normality on two variables.

Plotting soil depth versus root density in SPSS

How to make a scatterplot in SPSS

- Select 'Graph > Legacy Dialogues > Scatter/Dot'
- A small box should pop up, and we can select 'Simple Scatter',
- Click 'Define'.
- Select 'Soil_Depth' for the x axis and 'Root_Density' for the y-axis
- Click 'OK'.

Plotting soil depth versus root density in SPSS

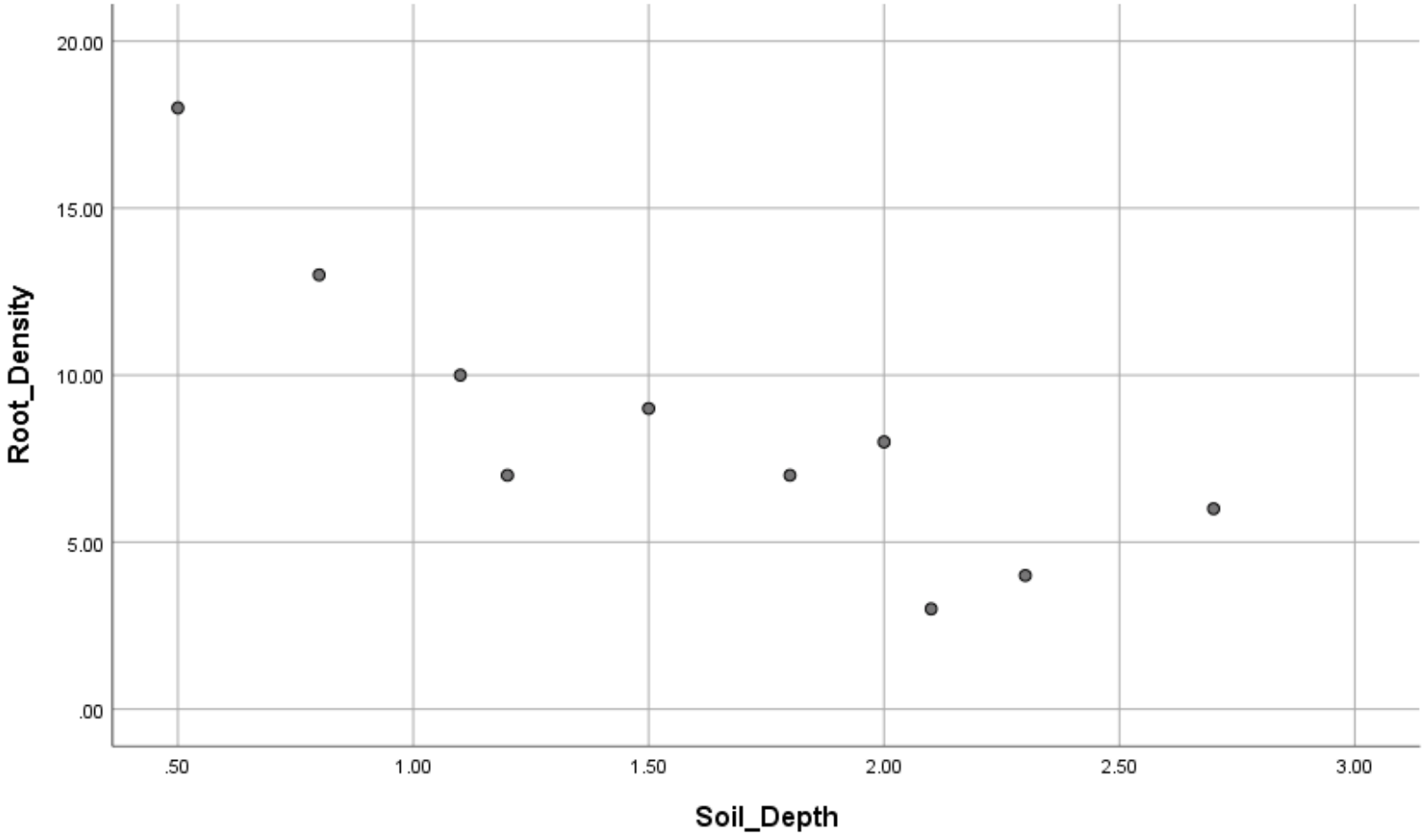

Figure 3: SPSS output of a scatterplot for Soil Depth versus Root Density

Testing whether soil depth and root density are correlated

Test whether or not our variables 'soil_depth' and 'root_density' are correlated.

Hypothesis for Pearson's correlation coefficient

- **Null:** There is no correlation between root density and soil depth
- **Alternative:** There is a significant correlation between root density and soil depth

We will reject the null hypothesis if, assuming that the null hypothesis is true, the probability of getting an r value as or more extreme than the one we obtained from our sample (i.e., the p-value) is less than or equal to 0.05.

Testing whether soil depth and root density are correlated

Test the null hypothesis that this correlation is not significant in SPSS

- Selecting 'Analyse > Correlate > Bivariate Correlation'
- Move both variables into the 'Variables' box
- Make sure 'Pearson' selected for Correlation Coefficients
- Use a two-tailed test for statistical significance
- Ignore the 'Flag significant correlations' (does not matter)

Testing whether soil depth and root density are correlated

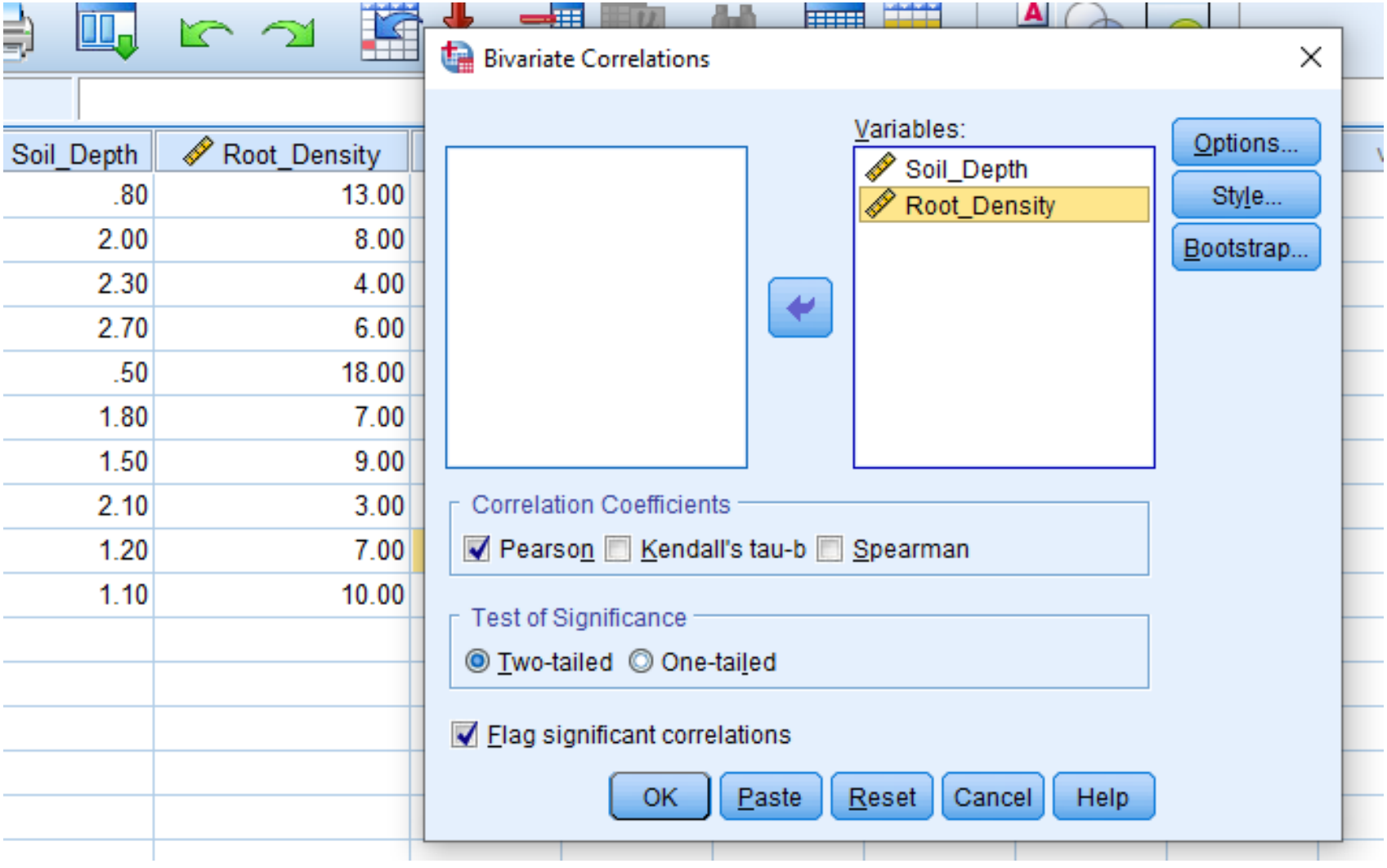

Figure 4: SPSS box showing how to run a test of the correlation coefficient.

Testing whether soil depth and root density are correlated

A table of output that looks like the one below.

Correlations

		Soil_Depth	Root_Density
Soil_Depth	Pearson Correlation	1	-.843**
	Sig. (2-tailed)		.002
	N	10	10
Root_Density	Pearson Correlation	-.843**	1
	Sig. (2-tailed)	.002	
	N	10	10

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

Figure 5: SPSS table showing output of a parameteric test of the significance of a correlation coefficient.

Spearman rank correlation coefficient

If either variable is not normally distributed, we need a non-parametric test

- The Spearman rank correlation coefficient is a non-parametric alternative.
- Calculate the correlation of the **ranks** of the values
- Test whether this Spearman rank correlation coefficient is significant

Consider some measurements of per cent dissolved oxygen and ammonia concentration (in mg per litre) from eight locations in Scotland.

Spearman rank correlation coefficient

Sample	%O2	Rank %O2	NH3 Conc.	Rank NH3 Conc.
1	95.9	8	0.080	2
2	81.9	3	0.100	3
3	80.9	2	0.210	6
4	77.9	1	0.370	8
5	90.7	6	0.250	7
6	88.2	4	0.130	5
7	93.6	7	0.070	1
8	89.1	5	0.121	4

Testing whether per cent O2 and Ammonia are correlated

Test whether or not our variables per cent O2 and Ammonia concentration are correlated.

Hypothesis for Spearman's correlation coefficient

- **Null:** There is no correlation between per cent dissolved oxygen and ammonia concentration
- **Alternative:** There is a significant correlation between per cent dissolved oxygen and ammonia concentration

We will reject the null hypothesis if, assuming that the null hypothesis is true, the probability of getting an r value as or more extreme than the one we obtained from our sample (i.e., the p-value) is less than or equal to 0.05.

Testing whether per cent O2 and Ammonia are correlated

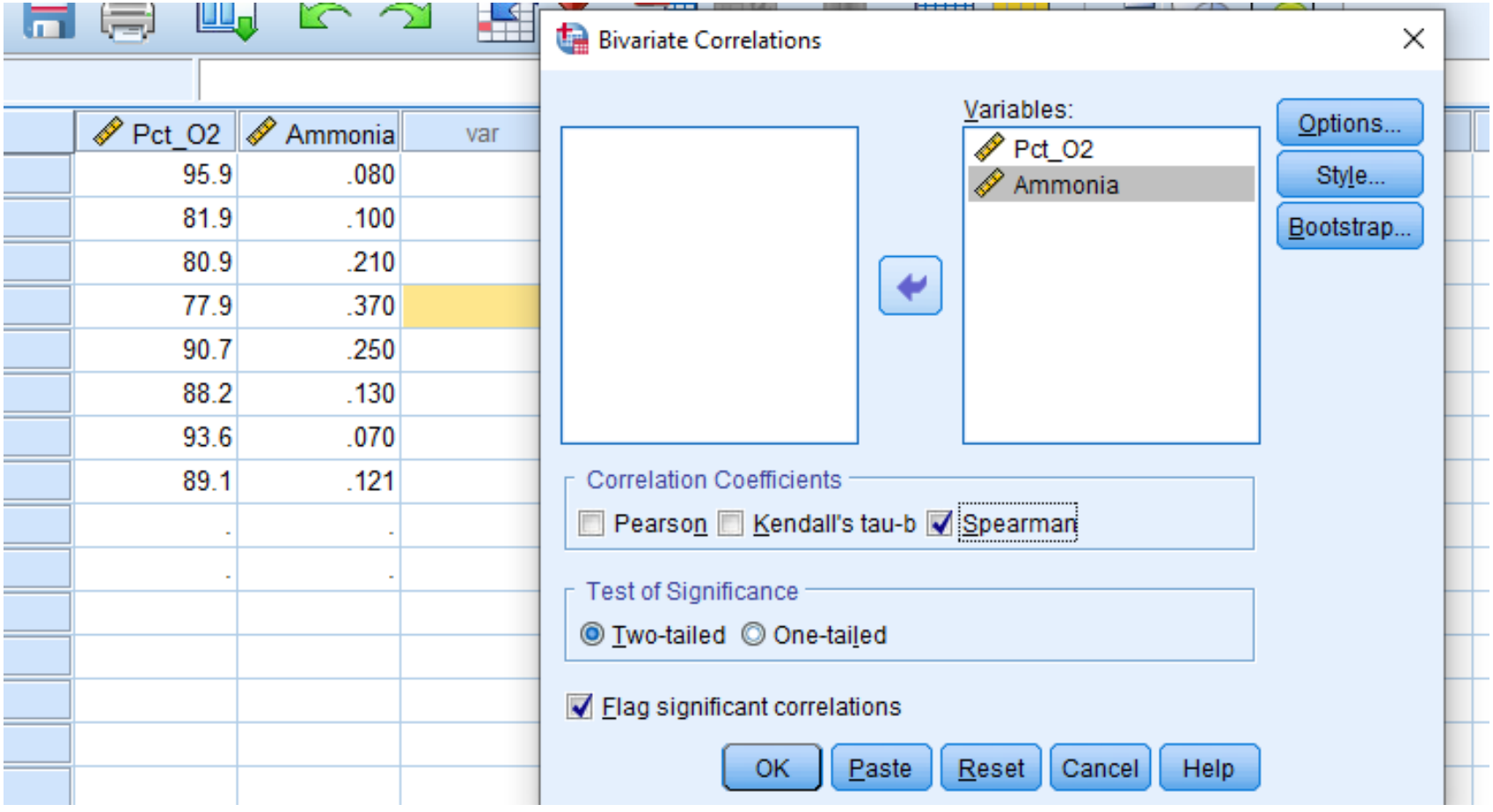

Figure 6: SPSS box showing how to run a test of the Spearman rank correlation coefficient.

Testing whether per cent O2 and Ammonia are correlated

Spearman rank correlation coefficient between the variables per cent dissolved oxygen and ammonia concentration is -0.667, and the p-value for this test is 0.071, meaning that we cannot reject our null hypothesis that the two variables are uncorrelated.

Correlations

			Pct_O2	Ammonia
Spearman's rho	Pct_O2	Correlation Coefficient	1.000	-.667
		Sig. (2-tailed)	.	.071
		N	8	8
	Ammonia	Correlation Coefficient	-.667	1.000
		Sig. (2-tailed)	.071	.
		N	8	8

Figure 7: SPSS table showing output of a parameteric test of the significance of a Spearman rank correlation coefficient.