

SPSS - Exploring Normality (Quiz)



Checking for Normality practical questions

In the example you will explore the distribution of a variable capturing the extent of a student's engagement in out-of-school science activities. The variable SCIEACT is a continuous score derived by PISA from responses to nine items on how often the student engages in a particular science activity, such as watching TV programmes about science and attending a science club (see PISA datafile description for further details).

Use the Explore window to construct plots and tests for normality to answer the following questions:

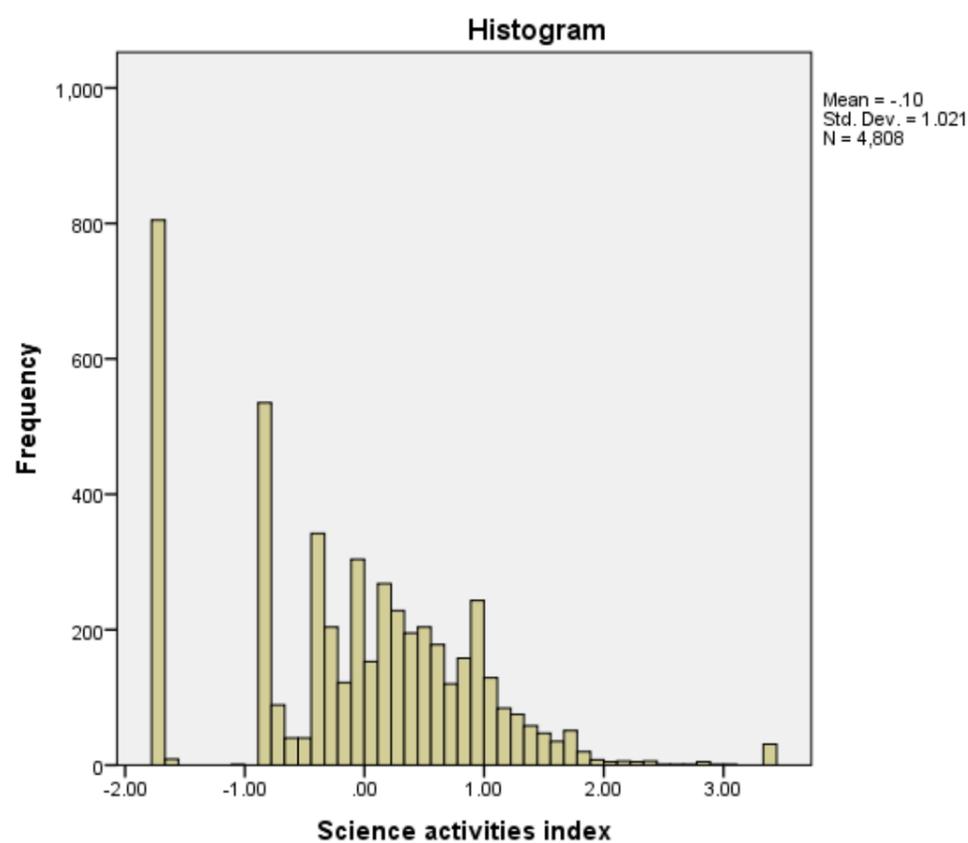
- Question: Firstly considering the histogram for the variable, **SCIEACT**. Does this histogram suggest normality?
- Question: Next looking at the two Normality test statistics do they suggest normality?
- Question: Finally does the QQ plot produced by SPSS suggest normality?

Solutions to Checking for Normality practical questions

- Select **Descriptive Statistics** from the **Analyze** menu.
- Select **Explore** from the **Descriptive Statistics** sub-menu.
- Click on the **Reset** button.
- Copy the **Science activities index[SCIEACT]** 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: Firstly considering the histogram for the variable, **SCIEACT**. Does this histogram suggest normality?

Solution: The output histogram from SPSS looks as follows:



Ideally for a normal distribution this histogram should look symmetric around the mean of the distribution, in this case $-.1025$. This distribution appears to be reasonably symmetric so there is no obvious skewness to make us question normality.

- Question: Next looking at the two Normality test statistics do they suggest normality?

Solution: The output of the test statistics from SPSS is as follows

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Science activities index	.111	4808	.000	.955	4808	.000

a. Lilliefors Significance Correction

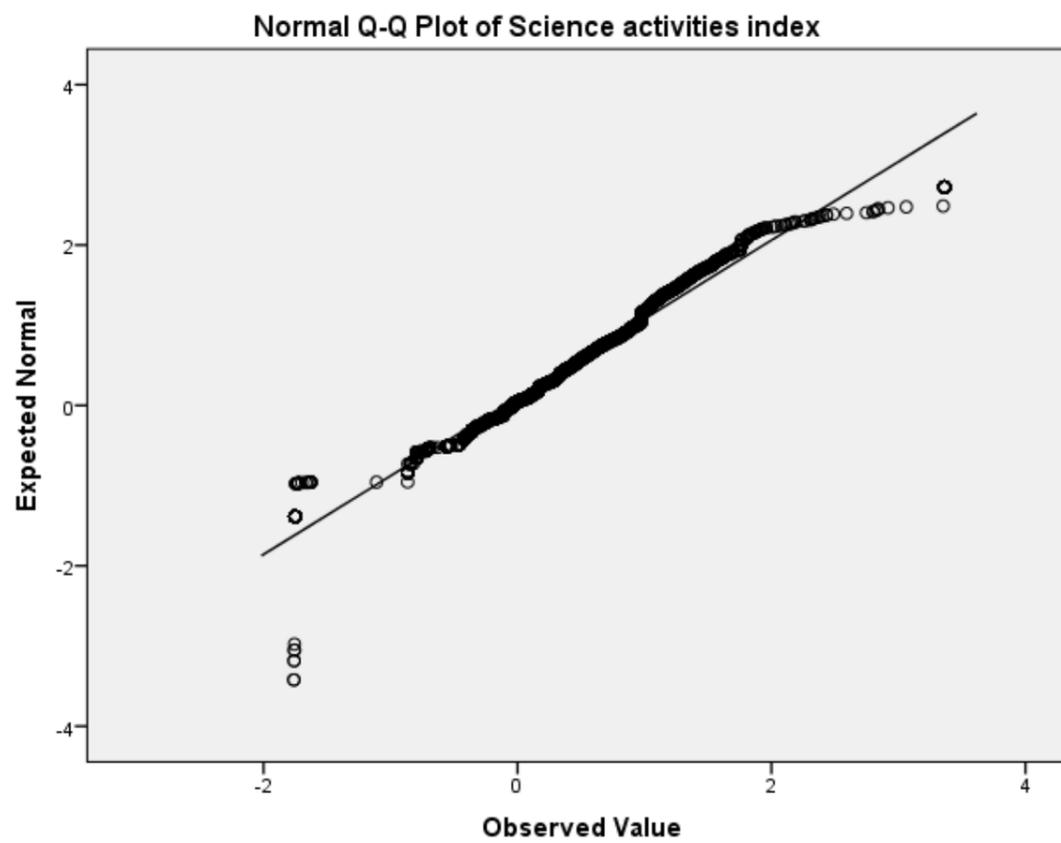
Here we see that the Kolmogorov Smirnov statistic takes value .111 whilst the Shapiro-Wilks statistic takes value .955. Both tests have the same degrees of freedom which equals the number of data points, namely 4808.

Here we see the p-value provided by SPSS (quoted under Sig. for Kolmogorov-Smirnov) is .000 (reported as $p < .001$). We therefore have significant evidence to reject the null hypothesis that the variable follows a normal distribution. The Shapiro-Wilks p value agrees with the Kolmogorov-Smirnov p value that the null hypothesis can be rejected.

As explained in the practical, formal tests of normality are only one piece of evidence that should be used in judgements of whether the distributional assumptions of parametric tests are likely to be violated. Visual inspection and consideration of sample sizes are equally, if not more, important.

- Question: Finally does the QQ plot produced by SPSS suggest normality?

Solution: The output from SPSS is as follows



In a QQ plot of normally distributed data all the points will lie on the diagonal black line. As we found a significant effect in the Kolmogorov Smirnov test we should see 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.

Inspection of the histogram of scores on the science activities index shows a large mass of observations at the lowest point on the scale, with a slightly smaller mass at another point that is considerably below the mean. This is likely to reflect the fact that a substantial number of students engage in none or one of the nine science-related activities that are used to form the index, leading to what is known as a floor effect. In a small sample, this distribution may lead to caution in the interpretation of parametric statistical tests that rely on normality, but in a sample this large these concerns are less of an issue.