

1 Testing for Normality using SPSS Statistics

1.1 Introduction

An assessment of the normality of data is a prerequisite for many statistical tests because normal data is an underlying assumption in parametric testing. There are two main methods of assessing normality: graphically and numerically. This "quick start" guide will help you to determine whether your data is normal, and therefore, that this assumption is met in your data for statistical tests. The approaches can be divided into two main themes: relying on statistical tests or visual inspection. Statistical tests have the advantage of making an objective judgement of normality, but are disadvantaged by sometimes not being sensitive enough at low sample sizes or overly sensitive to large sample sizes.

- As such, some statisticians prefer to use their experience to make a subjective judgement about the data from plots/graphs.
- Graphical interpretation has the advantage of allowing good judgement to assess normality in situations when numerical tests might be over or under sensitive, but graphical methods do lack objectivity.
- If you do not have a great deal of experience interpreting normality graphically, it is probably best to rely on the numerical methods.
- For each statistical test where you need to test for normality, we show you, step-by-step, the procedure in SPSS Statistics, as well as how to deal with situations where your data fails the assumption of normality (e.g., where you can try to "transform" your data to make it "normal"; something we also show you how to do using SPSS Statistics).

1.2 Methods of assessing normality

- SPSS Statistics allows you to test all of these procedures within **Explore...** command.
- The **Explore...** command can be used in isolation if you are testing normality in one group or splitting your dataset into one or more groups.
- For example, if you have a group of participants and you need to know if their height is normally distributed, everything can be done within the **Explore...** command.
- If you split your group into males and females (i.e., you have a categorical independent variable), you can test for normality of height within both the male group and the female group using just the **Explore...** command.
- This applies even if you have more than two groups.
- However, if you have 2 or more categorical, independent variables, the **Explore...** command on its own is not enough and you will have to use the **Split File...** command also.

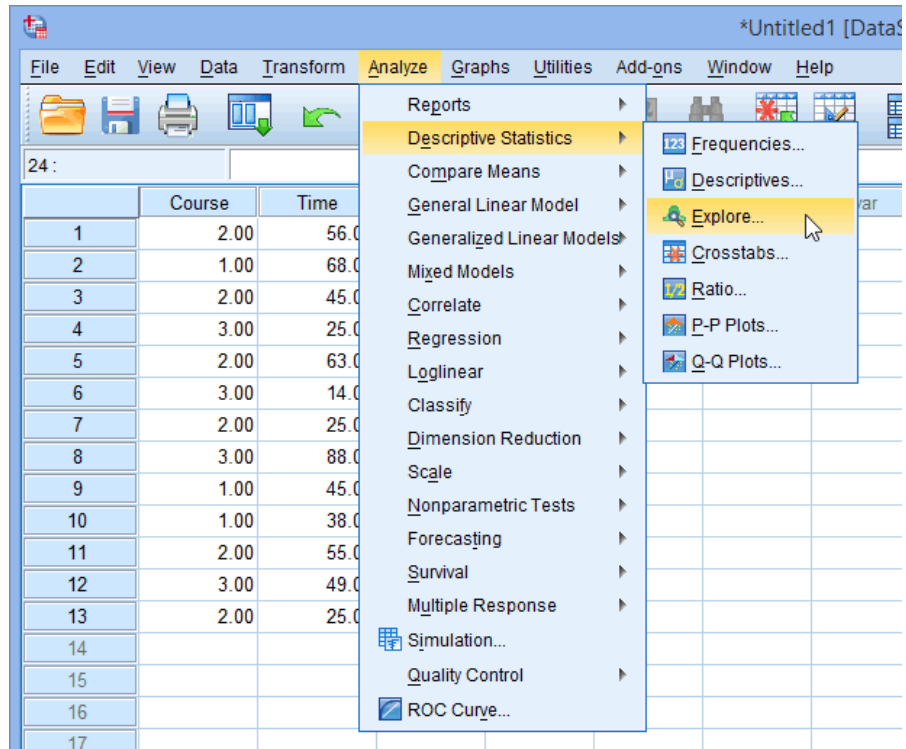


Figure 1:

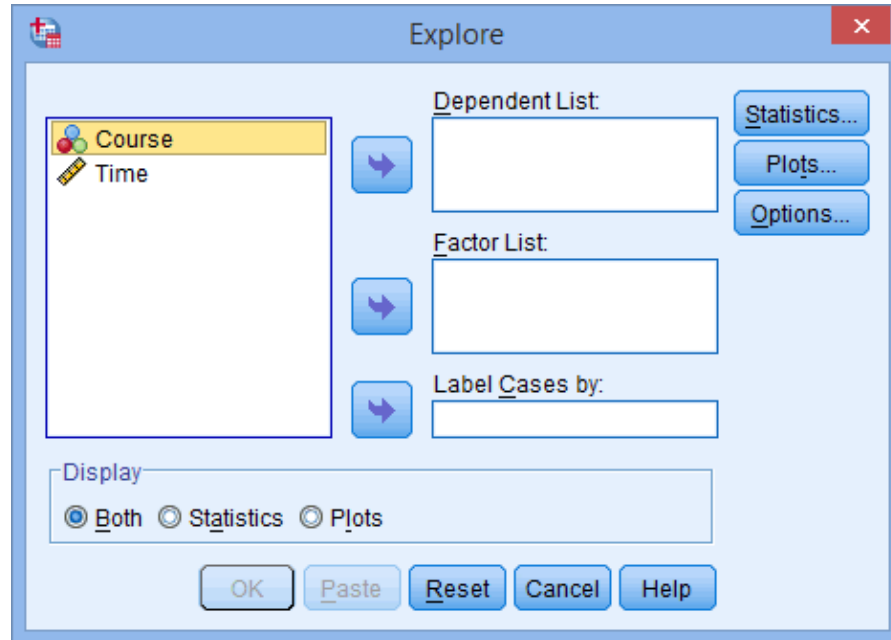


Figure 2:

1.3 Procedure for none or one grouping variable

The following example comes from our guide on how to perform a one-way ANOVA in SPSS Statistics.

Click Analyze > Descriptive Statistics > Explore... on the top menu, as shown below:

You will be presented with the Explore dialogue box, as shown below:

Transfer the variable that needs to be tested for normality into the Dependent List: box by either drag-and-dropping or using the SPSS Arrow Right Button button. In this example, we transfer the Time variable into the Dependent List: box. You will then be presented with the following screen:

[Optional] If you need to establish if your variable is normally distributed for each level of your independent variable, you need to add your independent variable to the Factor List: box by either drag-and-dropping or using the SPSS Arrow Right Button button. In this example, we transfer the Course variable into the Factor List: box. You will be presented with the following screen:

Click the SPSS Statistics Button button. You will be presented with the Explore: Statistics dialogue box, as shown below:

Leave the above options unchanged and click the SPSS Continue Button button.

Click the SPSS Plots Button button. Change the options so that you are presented with the following screen:

Click the button.

Click the button.

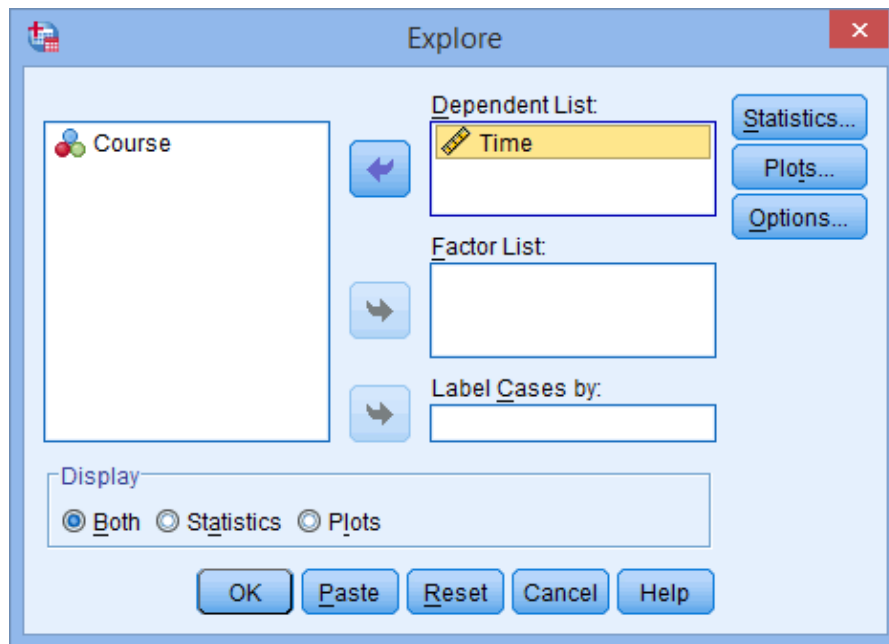


Figure 3:

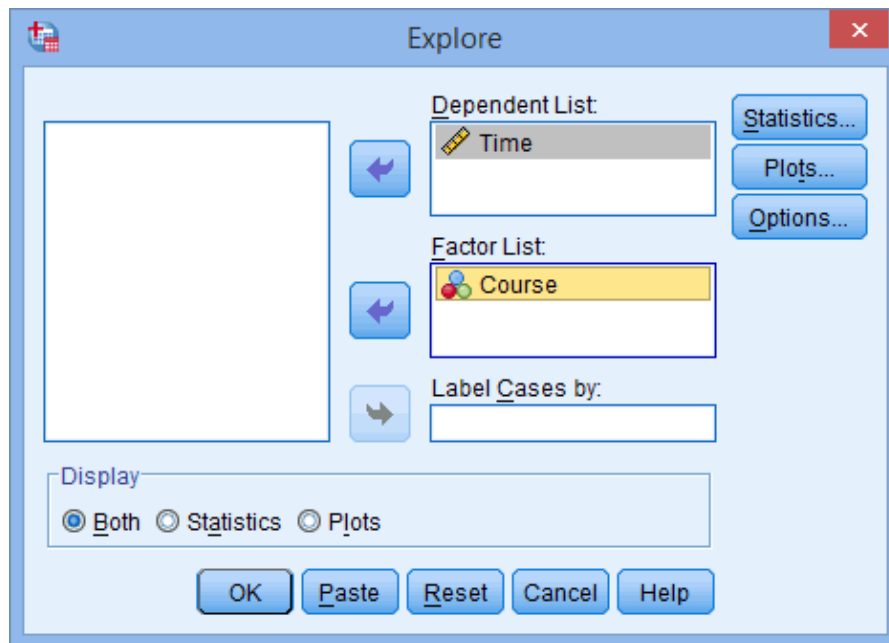


Figure 4:

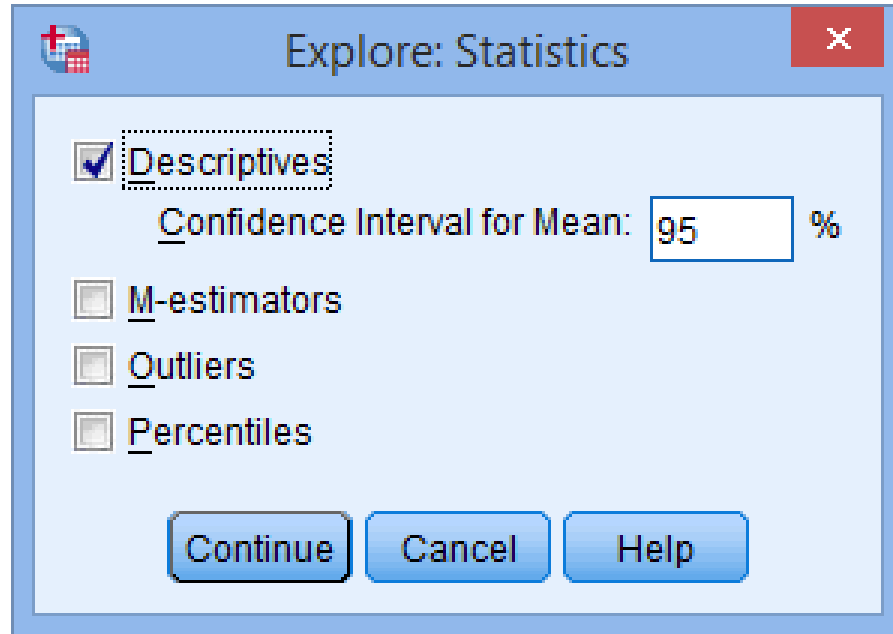


Figure 5:

1.4 Output

SPSS Statistics outputs many table and graphs with this procedure. One of the reasons for this is that the `Explore...` command is not used solely for the testing of normality, but in describing data in many different ways. When testing for normality, we are mainly interested in the Tests of Normality table and the Normal Q-Q Plots, our numerical and graphical methods to test for the normality of data, respectively.

1.5 Shapiro-Wilk Test of Normality

- The above table presents the results from two well-known tests of normality, namely the Kolmogorov-Smirnov Test and the Shapiro-Wilk Test. The Shapiro-Wilk Test is more appropriate for small sample sizes ($n < 50$ samples), but can also handle sample sizes as large as 2000. For this reason, we will use the Shapiro-Wilk test as our numerical means of assessing normality.
- We can see from the above table that for the "Beginner", "Intermediate" and "Advanced" Course Group the dependent variable, "Time", was normally distributed.
- **IMPORTANT** How do we know this? If the Sig. value of the Shapiro-Wilk Test is greater than 0.05, the data is normal. If it is below 0.05, the data significantly deviate from a normal distribution.

1.6 Normal Q-Q Plot

- In order to determine normality graphically, we can use the output of a normal Q-Q Plot. If the data are normally distributed, the data points will be close to the diagonal line.

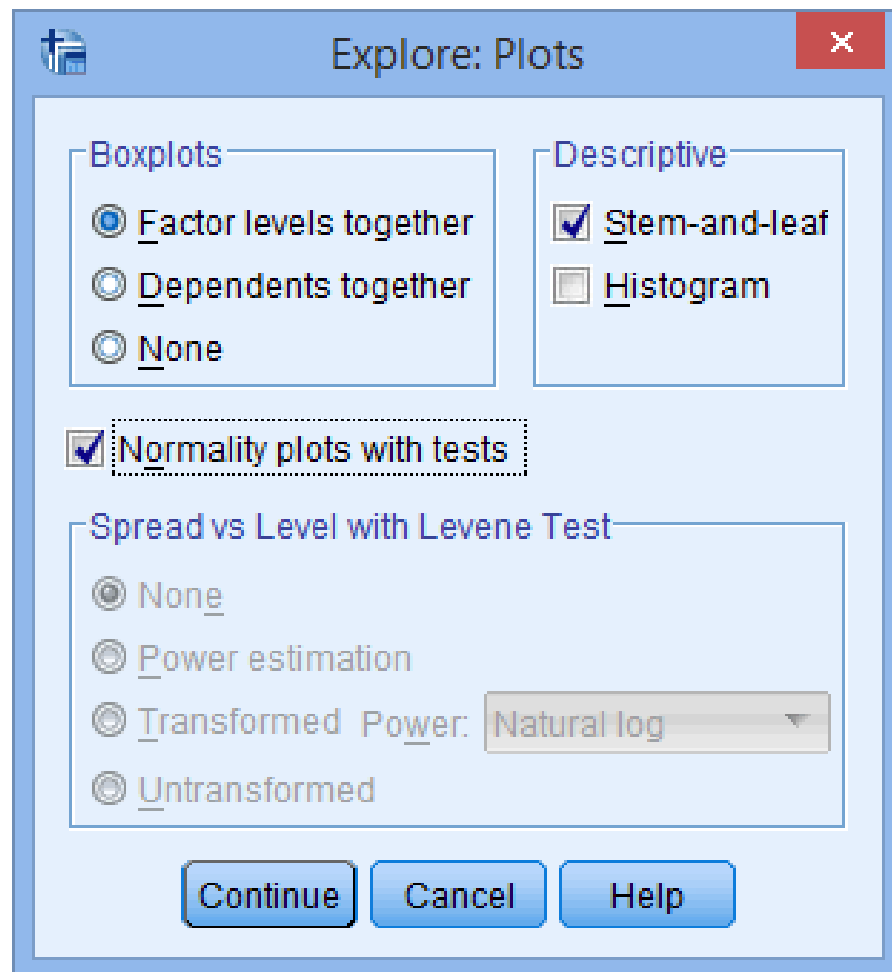


Figure 6:

Tests of Normality							
Course		Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Time	Beginner	.177	10	.200 [*]	.964	10	.827
	Intermediate	.166	10	.200 [*]	.969	10	.882
	Advanced	.151	10	.200 [*]	.965	10	.837

a. Lilliefors Significance Correction

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

Figure 7:

- If the data points stray from the line in an obvious non-linear fashion, the data are not normally distributed.
- As we can see from the normal Q-Q plot below, the data is normally distributed.
- If you are at all unsure of being able to correctly interpret the graph, rely on the numerical methods instead because it can take a fair bit of experience to correctly judge the normality of data based on plots.

Normality Screenshot If you need to know what Normal Q-Q Plots look like when distributions are not normal (e.g., negatively skewed), you will find these in our enhanced testing for normality guide. You can learn more about our enhanced content [here](#).

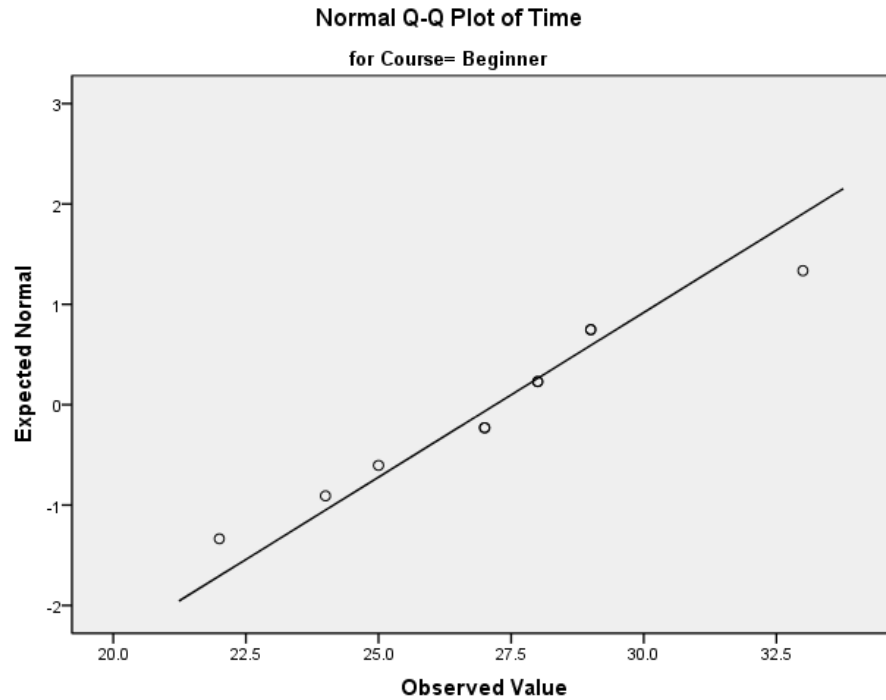


Figure 8:

2 Procedure when there are two or more independent variables

The Explore... command on its own cannot separate the dependent variable into groups based on not one but two or more independent variables. However, we can perform this feat by using the Split File... command.

Click Data \downarrow Split File... on the top menu as shown below:

You will be presented with the Split File dialogue box, as shown below:

Normality Screenshot

- Click the radio option, Organize output by groups. Transfer the independent variables you wish to categorize the dependent variable on into the Groups Based on: box.
- In this example, we want to know whether interest in politics (Int_Politics) is normally distributed when grouped/categorized by Gender AND Edu_Level (education level). You will be presented with the following screen:

Normality Screenshot

Click the SPSS OK Button button.

Note: Your file is now split and the output from any tests will be organized into the groups you have selected.

Click Analyze \downarrow Descriptive Statistics \downarrow Explore... on the top menu as shown below:

Normality Screenshot

You will be presented with the Explore dialogue box, as shown below:

Normality Screenshot Leave the above options unchanged and click the SPSS Continue Button button.

- Transfer the variable that needs to be tested for normality into the Dependent List: box by either drag-and-dropping or using the SPSS Right Arrow Button button. In this example, we transfer the `Int.Politics` variable into the Dependent List: box. You will then be presented with the following screen:

Normality Screenshot Published with written permission from SPSS Statistics, IBM Corporation.

Note: There is no need to transfer the independent variables `Gender` and `Edu.Level` into the Factor List: box because this has been accomplished with the Split File... command. We cannot simply transfer these two independent variables into the Factor List: box because this will not achieve the desired result. It will first analyse `Int.Politics` for normality with respect to `Gender` and then with respect to `Edu.Level`. It does NOT analyse `Int.Politics` for normality by grouping individuals into both `Gender` and `Edu.Level` AT THE SAME TIME.

Click the SPSS Statistics Button button. You will be presented with the Explore: Statistics dialogue box, as shown below:

Normality Screenshot

Leave the above options unchanged and click the SPSS Continue Button button.

Click the SPSS Plots Button button. Change the options so that you are presented with the following screen:

Normality Screenshot Click the button.

Click the button.

3 Output

You will now see that the output has been split into separate sections based on the combination of groups of the two independent variables. As an example we show the tests of normality when the dependent variable, "`Int.Politics`", is categorized into the first "`Gender`" group (male) and first "`Edu.Level`" group (School). All other possible combinations are also presented in the full output but we will not shown them here for clarity.

Normality Screenshot

Under this above category you are presented with the Tests of Normality table as shown below:

Normality Screenshot Published with written permission from SPSS Statistics, IBM Corporation.

- The Shapiro-Wilk test is now analysing the normality of "`Int.Politics`" on the data of those individuals that are classified as both "male" in the independent variable, "`Gender`", and "school" in the independent variable "`Edu.Level`". As the Sig. value under the Shapiro-Wilk column is greater than 0.05, we can conclude that "`Int.Politics`" for this particular subset of individuals is normally distributed.
- The same data from the same individuals are now also being analysed to produce a Normal Q-Q Plot as below. From this graph, we can conclude that the data appears to be normally distributed as it follows the diagonal line closely and does not appear to have a non-linear pattern.

Normality Screenshot