

1.5: Normality

Learning Objectives

At the end of this section you should be able to answer the following questions:

- How would you define the concept of multivariate normality?
- What are three different methods for checking multivariate normality?

There are a number of underlying assumptions that go with parametric statistical testing (which is what we will be focusing on for the majority of this book).

If you are undertaking parametric tests, then one of the key assumptions is *multivariate normality*, or the assumption that the variables in your data are distributed normally.

Chances are you have all encountered an image of the bell curve throughout your academic studies. This bell curve represents a normal distribution.

There are a number of ways you can check for normality.

Your first option is to check multivariate normality by visually examining graphs of the data for each variable. For this type of checking, you will need to create a bar graph or a histogram. A basic visual inspection will often show if the data is normal or near to normal.

You can also check normality by looking at the *skewness* and *kurtosis* (S&K) of the distribution of your variables. Skewness looks at how the data is distributed horizontally. In other words, is the data all bunched up at one end of the graph. Kurtosis is the height of the distribution, and this should be neither too low nor too high. You will need to check if the curve is high and tight or flat and long. In an ideal distribution, the values assigned to S&K would be 0, however, there is nearly always some variance from normality in any dataset. S&K values of less than ± 2.00 are generally considered to be close enough to normal for you to use parametric statistics.

Finally, a third way to check the normality of a data distribution is to use a dedicated *normality test*, which will be used one variable at a time. There are two main normality tests that researchers would typically use: the Kolmogorov-Smirnov test for samples larger than 50, and Shapiro-Wilk tests for samples less than 50. These tests assume that the distribution is normal. Therefore, if these tests are significant (i.e. p value $< .05$) it means that the data varies from the normal model and should be considered not normal.

If a continuous variable is found to be non-normal either from visual inspection, skewness and kurtosis values or a normality test, there are a number of ways to deal with this.

Firstly, you would want to check and see if there are any outliers in your data. If there are, it might be worth deleting them from the data set. You can also transform the variable into a logarithmic scale. Finally, if the violations are not too severe, you could just live with the non-normality of the variable and produce bootstrapped results. In any event, you will need to make mention of how you dealt with any non-normal data in the results section of your report or paper or thesis.

This page titled [1.5: Normality](#) is shared under a [CC BY 4.0](#) license and was authored, remixed, and/or curated by [Erich C Fein](#), [John Gilmour](#), [Tayna Machin](#), and [Liam Hendry](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.