How can we perform multivariate normality tests in R?

Authored by

stats writer

April 20, 2024

Multivariate normality tests in R refer to the statistical methods used to determine whether a set of data follows a multivariate normal distribution. This type of distribution is commonly used in statistical analysis and assumes that the data is normally distributed in multiple dimensions. To perform multivariate normality tests in R, one can use various techniques such as graphical methods, statistical tests, and simulation methods. These tests can be performed using various R packages, such as "MVN," "nortest," and "normtest," which provide functions for assessing the normality of multivariate data. The results of these tests can help researchers determine the appropriateness of using certain statistical methods and models for their data. Overall, performing multivariate normality tests in R is an essential step in ensuring the validity and reliability of statistical analyses.

Perform Multivariate Normality Tests in R

When we'd like to test whether or not a single variable is normally distributed, we can create a Q-Q plot to visualize the distribution or we can perform a formal statistical test like an Anderson Darling Test or a Jarque-Bera Test.

However, when we'd like to test whether or not *several* variables are normally distributed as a group we must perform a multivariate normality test.

This tutorial explains how to perform the following multivariate normality tests for a given dataset in R:

Mardia's TestEnergy TestMultivariate Kurtosis and Skew Tests

If we'd like to identify outliers in a multivariate setting, we can use the Mahalanobis distance.

Example: Mardia's Test in R

Mardia's Test determines whether or not a group of variables follows a multivariate normal distribution. The null and alternative hypotheses for the test are as follows:

H0 (null): The variables follow a multivariate normal distribution.

Ha (alternative): The variables *do not* follow a multivariate normal distribution.

The following code shows how to perform this test in R using the QuantPsyc package:

```
library(QuantPsyc)

#create dataset
set.seed(0)

data <- data.frame(x1 = rnorm(50),
x2 = rnorm(50),
```

x3 = rnorm(50))

#perform Multivariate normality test
mult.norm(data)$mult.test

Beta-hat kappa p-val
Skewness 1.630474 13.5872843 0.1926626
Kurtosis 13.895364 -0.7130395 0.4758213

The mult.norm() function tests for multivariate normality in both the skewness and kurtosis of the dataset. Since both p-values are not less than .05, we fail to reject the null hypothesis of the test. We don't have evidence to say that the three variables in our dataset do not follow a multivariate distribution.

Example: Energy Test in R

An Energy Test is another statistical test that determines whether or not a group of variables follows a multivariate normal distribution. The null and alternative hypotheses for the test are as follows:

H0 (null): The variables follow a multivariate normal distribution.

Ha (alternative): The variables *do not* follow a multivariate normal distribution.

The following code shows how to perform this test in R using the energy package:

```
library(energy)

#create dataset
set.seed(0)

data <- data.frame(x1 = rnorm(50),
x2 = rnorm(50),
x3 = rnorm(50))

#perform Multivariate normality test
mvnorm.etest(data, R=100)
```

Energy test of multivariate normality: estimated parameters

data: x, sample size 50, dimension 3, replicates 100
E-statistic = 0.90923, p-value = 0.31

The p-value of the test is 0.31. Since this is not less than .05, we fail to reject the null hypothesis of the test.

We don't have evidence to say that the three variables in our dataset do not follow a multivariate distribution.

Note: The argument R=100 specifies 100 boostrapped replicates to be used when performing the test. For datasets with smaller sample sizes, you may increase this number to produce a more reliable estimate of the test statistic.

How to Create & Interpret a Q-Q Plot in R
How to Conduct an Anderson-Darling Test in R
How to Conduct a Jarque-Bera Test in R
How to Perform a Shapiro-Wilk Test in R