

How can I calculate the skewness and kurtosis of a dataset in R?

Authored by
stats writer

April 21, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I calculate the skewness and kurtosis of a dataset in R?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=137684>

Calculating the skewness and kurtosis of a dataset is an important step in understanding its distribution and identifying any potential outliers. In R, this can be done using the functions 'skewness' and 'kurtosis' from the 'e1071' package. These functions take in a dataset as input and provide numerical values for skewness and kurtosis, which can then be interpreted to determine the shape and symmetry of the dataset. Skewness measures the degree of asymmetry in the dataset, while kurtosis measures the peakedness or flatness of the distribution. By calculating these values, users can gain insights into the characteristics of their dataset and make informed decisions in their data analysis.

Calculate Skewness & Kurtosis in R

In statistics, skewness and kurtosis are two ways to measure the shape of a distribution.

Skewness is a measure of the asymmetry of a distribution. This value can be positive or negative.

A negative skew indicates that the tail is on the left side of the distribution, which extends towards more negative values. A positive skew indicates that the tail is on the right side of the distribution, which extends towards more positive values. A value of zero indicates that there is no skewness in the distribution at all, meaning the distribution is perfectly symmetrical.

Kurtosis is a measure of whether or not a distribution is heavy-tailed or light-tailed relative to a normal distribution.

The kurtosis of a normal distribution is 3. If a given distribution has a kurtosis less than 3, it is said to be *platykurtic*, which means it tends to produce fewer and less extreme outliers than the normal distribution. If a given distribution has a kurtosis greater than 3, it is said to be *leptokurtic*, which means it tends to produce more outliers than the normal distribution.

Note: Some formulas (Fisher's definition) subtract 3 from the kurtosis to make it easier to compare with the normal distribution. Using this definition, a distribution would have kurtosis greater than a normal distribution if it had a kurtosis value greater than 0.

This tutorial explains how to calculate both the skewness and kurtosis of a given dataset in R.

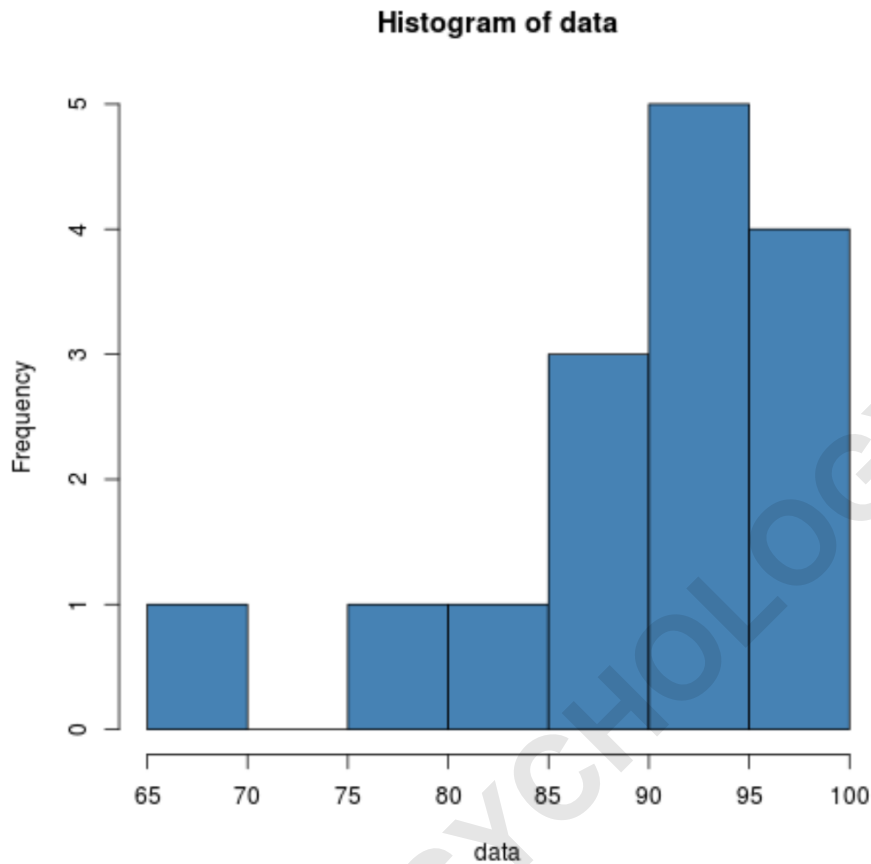
Example: Skewness & Kurtosis in R

Suppose we have the following dataset:

```
data = c(88, 95, 92, 97, 96, 97, 94, 86, 91, 95, 97, 88, 85, 76, 68)
```

We can quickly visualize the distribution of values in this dataset by creating a histogram:

hist(data, col='steelblue')



From the histogram we can see that the distribution appears to be left-skewed. That is, more of the values are concentrated on the right side of the distribution.

To calculate the skewness and kurtosis of this dataset, we can use `skewness()` and `kurtosis()` functions from the moments library in R:

```
library(moments)
```

```
#calculate skewness
```

```
skewness(data)
```

```
-1.391777
```

```
#calculate kurtosis
```

```
kurtosis(data)
```

```
4.177865
```

The skewness turns out to be **-1.391777** and the kurtosis turns out to be **4.177865**.

Since the kurtosis is greater than 3, this indicates that the distribution has more values in the tails compared to a normal distribution.

The moments library also offers the `jarque.test()` function, which performs a goodness-of-fit test that determines whether or not sample data have skewness and kurtosis that matches a normal distribution. The null and alternative hypotheses of this test are as follows:

Null Hypothesis: The dataset has a skewness and kurtosis that matches a normal distribution.

Alternative Hypothesis: The dataset has a skewness and kurtosis that *does not* match a normal distribution.

The following code shows how to perform this test:

```
jarque.test(data)
```

Jarque-Bera Normality Test

```
data: data
```

```
JB = 5.7097, p-value = 0.05756
```

```
alternative hypothesis: greater
```

The p-value of the test turns out to be 0.05756. Since this value is not less than $\alpha = .05$, we fail to reject the null hypothesis. We do not have sufficient evidence to say that this dataset has a skewness and kurtosis that is different from the normal distribution.

You can find the complete documentation for the moments library [here](#).

Bonus: Skewness & Kurtosis Calculator

You can also calculate the skewness for a given dataset using the [Statology Skewness and Kurtosis Calculator](#),

which automatically calculates both the skewness and kurtosis for a given dataset.

ARABPSYCHOLOGY.COM