

How can I perform a Shapiro-Wilk test in R, and what are some examples of its use?

Authored by
stats writer

April 20, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I perform a Shapiro-Wilk test in R, and what are some examples of its use?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=137449>

The Shapiro-Wilk test is a statistical test used to determine if a given sample of data is normally distributed. This test is commonly used in data analysis and is available in the R programming language. To perform a Shapiro-Wilk test in R, you can use the "shapiro.test()" function, which takes a vector of numerical data as its input. The function will then return a p-value, which can be compared to a chosen significance level to determine if the data is normally distributed or not. This test can be used in various scenarios, such as checking the normality of data before performing further statistical analysis, or determining the appropriateness of using a certain statistical model. Overall, the Shapiro-Wilk test is a useful tool for assessing the distribution of data in R and can aid in making informed decisions in data analysis.

Perform a Shapiro-Wilk Test in R (With Examples)

The Shapiro-Wilk test is a test of normality. It is used to determine whether or not a sample comes from a normal distribution.

This type of test is useful for determining whether or not a given dataset comes from a normal distribution, which is a common assumption used in many statistical tests including regression, ANOVA, t-tests, and many others.

We can easily perform a Shapiro-Wilk test on a given dataset using the following built-in function in R:

shapiro.test(x)

where:

x: A numeric vector of data values.

This function produces a test statistic W along with a corresponding p-value. If the p-value is less than $\alpha = .05$, there is sufficient evidence to say that the sample does not come from a population that is normally distributed.

Note: The sample size must be between 3 and 5,000 in order to use the `shapiro.test()` function.

This tutorial shows several examples of how to use this function in practice.

Example 1: Shapiro-Wilk Test on Normal Data

The following code shows how to perform a Shapiro-Wilk test on a dataset with sample size $n=100$:

```
#make this example reproducible  
set.seed(0)
```

```
#create dataset of 100 random values generated from a  
normal distribution  
data <- rnorm(100)
```

```
#perform Shapiro-Wilk test for normality
```

```
shapiro.test(data)
```

Shapiro-Wilk normality test

```
data: data
```

```
W = 0.98957, p-value = 0.6303
```

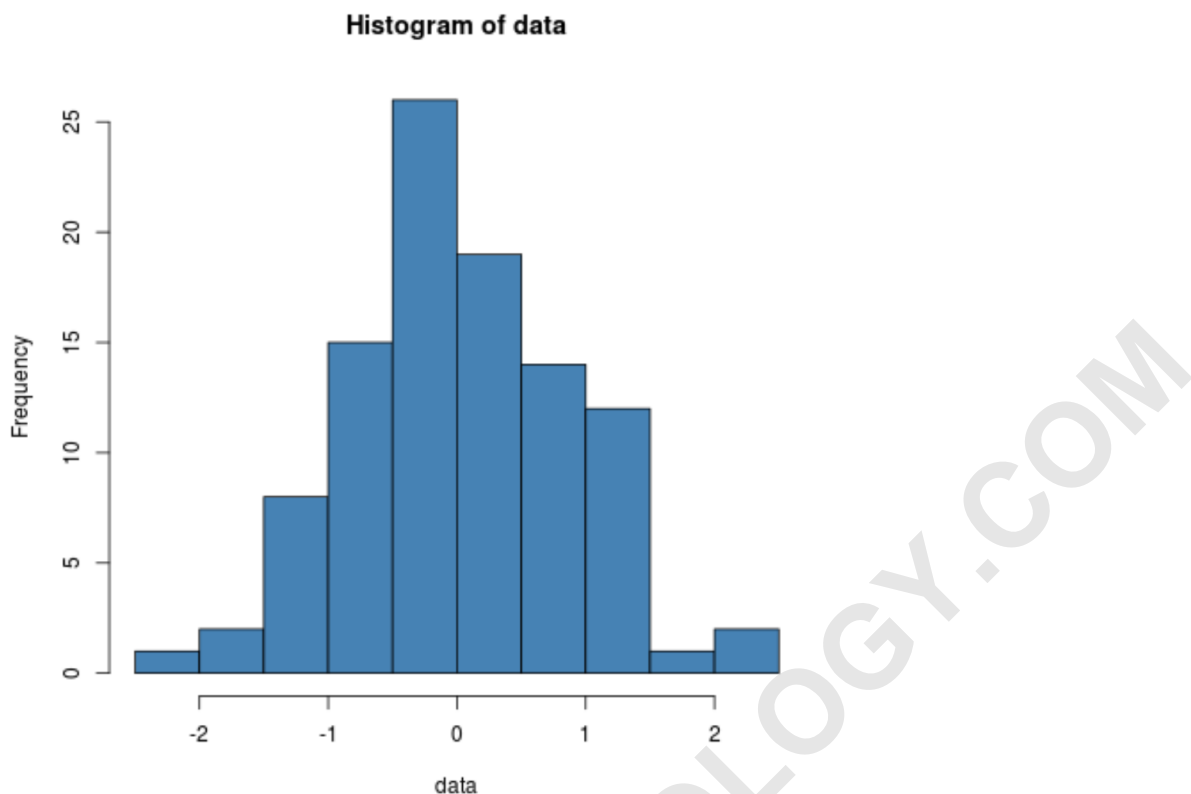
The p-value of the test turns out to be 0.6303. Since this value is not less than .05, we can assume the sample data comes from a population that is normally distributed.

This result shouldn't be surprising since we generated the sample data using the `rnorm()` function, which generates random values from a normal distribution with mean = 0 and standard deviation = 1.

A Guide to `dnorm`, `pnorm`, `qnorm`, and `rnorm` in R

We can also produce a histogram to visually verify that the sample data is normally distributed:

```
hist(data, col='steelblue')
```



We can see that the distribution is fairly bell-shaped with one peak in the center of the distribution, which is typical of data that is normally distributed.

Example 2: Shapiro-Wilk Test on Non-Normal Data

The following code shows how to perform a Shapiro-Wilk test on a dataset with sample size $n=100$ in which the values are randomly generated from a Poisson distribution:

```
#make this example reproducible  
set.seed(0)
```

```
#create dataset of 100 random values generated from a  
Poisson distribution
```

```
data <- rpois(n=100, lambda=3)
```

```
#perform Shapiro-Wilk test for normality
```

```
shapiro.test(data)
```

Shapiro-Wilk normality test

```
data: data
```

```
W = 0.94397, p-value = 0.0003393
```

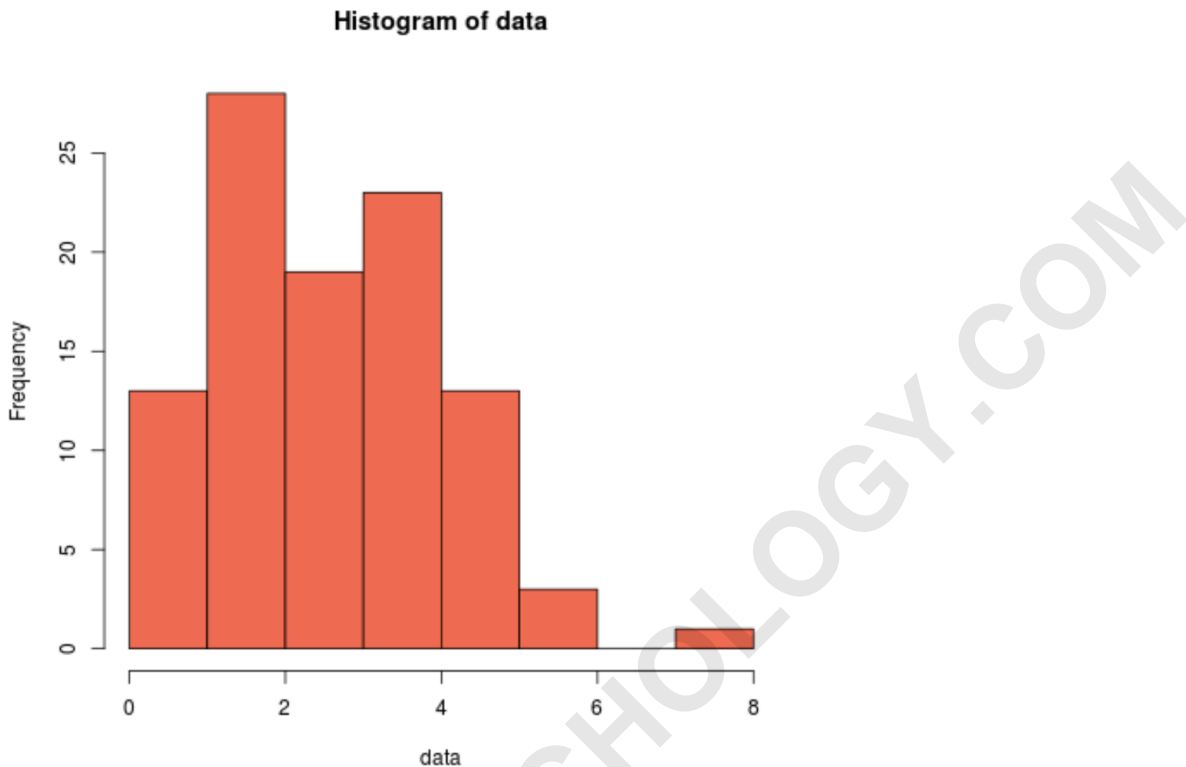
The p-value of the test turns out to be 0.0003393. Since this value is less than .05, we have sufficient evidence to say that the sample data does *not* come from a population that is normally distributed.

This result shouldn't be surprising since we generated the sample data using the `rpois()` function, which generates random values from a Poisson distribution.

A Guide to `dpois`, `ppois`, `qpois`, and `rpois` in R

We can also produce a histogram to visually see that the sample data is not normally distributed:

hist(data, col='coral2')



We can see that the distribution is right-skewed and doesn't have the typical "bell-shape" associated with a normal distribution. Thus, our histogram matches the results of the Shapiro-Wilk test and confirms that our sample data does not come from a normal distribution.

What to Do with Non-Normal Data

If a given dataset is *not* normally distributed, we can often perform one of the following transformations to make it more normal:

1. Log Transformation: Transform the response variable from y to $\log(y)$.

2. Square Root Transformation: Transform the response variable from y to \sqrt{y} .

3. Cube Root Transformation: Transform the response variable from y to $y^{1/3}$.

By performing these transformations, the response variable typically becomes closer to normally distributed.

Check out [this tutorial](#) to see how to perform these transformations in practice.

[How to Conduct an Anderson-Darling Test in R](#)

[How to Perform a Shapiro-Wilk Test in Python](#)