

What is the Kolmogorov-Smirnov Test in R and how can it be implemented? Can you provide examples of its application?

Authored by
stats writer

April 21, 2024

RECOMMENDED CITATION

stats writer (2024). *What is the Kolmogorov-Smirnov Test in R and how can it be implemented? Can you provide examples of its application?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=137602>

The Kolmogorov-Smirnov test is a statistical method used to determine whether a sample or dataset follows a specific distribution. In R, it is implemented through the "ks.test" function, which compares the empirical cumulative distribution function (ECDF) of the data to the theoretical distribution. This test is commonly used to assess the goodness of fit for continuous distributions, such as the normal, exponential, or uniform distribution.

For example, suppose we have a dataset of 1000 observations and we want to test whether it follows a normal distribution. We can use the "ks.test" function in R to compare the ECDF of our data to the theoretical normal distribution. The output of this test will provide a p-value, which can be used to determine whether we can reject or fail to reject the null hypothesis that the data follows a normal distribution.

The Kolmogorov-Smirnov test can also be used to compare two datasets and determine if they have the same underlying distribution. This is done through the "ks.test" function by specifying both datasets as inputs. The output will again provide a p-value, which can be used to determine whether we can reject or fail to reject the null hypothesis that the two datasets have the same distribution.

In summary, the Kolmogorov-Smirnov test in R is a useful tool for assessing the fit of a dataset to a specific distribution or comparing two datasets. It can provide valuable insights into the underlying distribution of a dataset and is commonly used in data analysis and research.

Kolmogorov-Smirnov Test in R (With Examples)

The Kolmogorov-Smirnov test is used to test whether or not or not a sample comes from a certain distribution.

To perform a one-sample or two-sample Kolmogorov-Smirnov test in R we can use the ks.test() function.

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

Example 1: One Sample Kolmogorov-Smirnov Test

Suppose we have the following sample data:

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

```
#generate dataset of 100 values that follow a Poisson  
distribution with mean=5
```

```
data <- rpois(n=20, lambda=5)
```

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

The following code shows how to perform a Kolmogorov-Smirnov test on this sample of 100 data values to determine if it came from a normal distribution:

```
#perform Kolmogorov-Smirnov test  
ks.test(data, "pnorm")
```

One-sample Kolmogorov-Smirnov test

```
data: data
```

```
D = 0.97725, p-value < 2.2e-16
```

```
alternative hypothesis: two-sided
```

From the output we can see that the test statistic is 0.97725 and the corresponding p-value is 2.2e-16. Since the p-value is less than .05, we reject the null hypothesis. We have sufficient evidence to say that the sample data does not come from a normal distribution.

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

Example 2: Two Sample Kolmogorov-Smirnov Test

Suppose we have the following two sample datasets:

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

```
#generate two datasets  
data1 <- rpois(n=20, lambda=5)  
data2 <- rnorm(100)
```

The following code shows how to perform a Kolmogorov-Smirnov test on these two samples to determine if they came from the same distribution:

#perform Kolmogorov-Smirnov test

```
ks.test(data1, data2)
```

Two-sample Kolmogorov-Smirnov test

data: data1 and data2

D = 0.99, p-value = 1.299e-14

alternative hypothesis: two-sided

From the output we can see that the test statistic is 0.99 and the corresponding p-value is 1.299e-14. Since the p-value is less than .05, we reject the null hypothesis. We have sufficient evidence to say that the two sample datasets do not come from the same distribution.

How to Perform a Shapiro-Wilk Test in R

How to Perform an Anderson-Darling Test in R

How to Perform Multivariate Normality Tests in R