

# How can a Kolmogorov-Smirnov test be performed in Python?

Authored by  
**stats writer**

April 18, 2024

## RECOMMENDED CITATION

stats writer (2024). *How can a Kolmogorov-Smirnov test be performed in Python?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=136834>

The Kolmogorov-Smirnov test is a statistical test used to determine whether a dataset follows a specific distribution. In Python, this test can be performed by using the "ks\_2samp" function from the SciPy library. This function takes in two arrays, representing the observed data and the expected distribution, and returns the test statistic and p-value. The test statistic measures the maximum distance between the cumulative distribution functions of the two datasets, while the p-value indicates the likelihood of obtaining this result by chance. By comparing the p-value to a predetermined significance level, the test can determine whether the data significantly deviates from the expected distribution. Overall, the Kolmogorov-Smirnov test is a useful tool for analyzing data and can be easily performed in Python using the "ks\_2samp" function.

## Perform a Kolmogorov-Smirnov Test in Python

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

**To perform a Kolmogorov-Smirnov test in Python we can use the `scipy.stats.kstest()` for a one-sample test or `scipy.stats.ks_2samp()` for a two-sample test.**

**This tutorial shows an example of how to use each function in practice.**

### Example 1: One Sample Kolmogorov-Smirnov Test

**Suppose we have the following sample data:**

```
from numpy.random import seed
```

```
from numpy.random import poisson
```

```
#set seed (e.g. make this example reproducible)
```

```
seed(0)
```

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

```
data = poisson(5, 100)
```

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:

```
from scipy.stats import kstest
```

```
#perform Kolmogorov-Smirnov test
```

```
kstest(data, 'norm')
```

```
KstestResult(statistic=0.9072498680518208,  
pvalue=1.0908062873170218e-103)
```

From the output we can see that the test statistic is 0.9072 and the corresponding p-value is 1.0908e-103. 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 also shouldn't be surprising since we generated the sample data using the `poisson()` 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:

```
from numpy.random import seed
from numpy.random import randn
from numpy.random import lognormal

#set seed (e.g. make this example reproducible)
seed(0)

#generate two datasets
data1 = randn(100)
data2 = lognormal(3, 1, 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:

```
from scipy.stats import ks_2samp
```

```
#perform Kolmogorov-Smirnov test  
ks_2samp(data1, data2)
```

```
KstestResult(statistic=0.99,  
pvalue=4.417521386399011e-57)
```

From the output we can see that the test statistic is 0.99 and the corresponding p-value is 4.4175e-57. 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.

This result also shouldn't be surprising since we generated values for the first sample using the standard normal distribution and values for the second sample using the lognormal distribution.

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

**[How to Perform an Anderson-Darling Test in Python](#)**