

Select Statistical Tests

- 1 Testing for Normality (`normaltest`)
- 2 Kolmogorov Smirnov Test (`kstest`)
- 3 two sample KS test
- 4 Shapiro Test for Normality

`normaltest`

In data analysis, it is often quite useful to know if a data set is normally distributed, as it lets the analysts know which subsequent statistical procedures are valid and which are not.

- `normaltest` tests for normality in an array of data.
- An optional second argument provides the axis to use (default is to use entire array). Returns the test statistic and the p -value of the test.
- This test is a small sample modified version of the *Jarque-Bera* test statistic.

Kolmogorov-Smirnov Test (`kstest`)

- `kstest` implements the *Kolmogorov-Smirnov* test.
- The command requires two inputs, the data to use in the test and the distribution, which can be a string or a frozen random variable object.
- If the distribution is provided as a string, then any required shape parameters are passed in the third argument using a tuple containing these parameters, in order.

```
>>> x = randn(100)
>>> kstest = stats.kstest
>>> stat, pval = kstest(x, 'norm')
>>> stat
0.11526423481470172
>>> pval
0.12963296757465059
>>> ncdf = stats.norm().cdf # No () on cdf to get the function
>>> kstest(x, ncdf)
```

```
(0.11526423481470172, 0.12963296757465059)
>>> x = gamma.rvs(2, size = 100)
>>> kstest(x, 'gamma', (2,)) # (2,) contains the shape parameter
(0.079237623453142447, 0.54096739528138205)
>>> gcdf = gamma(2).cdf
>>> kstest(x, gcdf)
(0.079237623453142447, 0.54096739528138205)
```

`ks_2samp`

`ks_2samp` implements a 2-sample version of the Kolmogorov-Smirnov test. It is called `ks_2samp(x,y)` where both inputs are 1-dimensional arrays, and returns the test statistic and p-value for the null that the distribution of `x` is the same as that of `y`.

Shapiro-Wilk test for normality (`shapiro`)

`shapiro` implements the Shapiro-Wilk test for normality on a 1-dimensional array of data. It returns the test statistic and p-value for the null of normality.

1 Statsmodels

Statsmodels is a Python module that allows users to explore data, estimate statistical models, and perform statistical tests. An extensive list of descriptive statistics, statistical tests, plotting functions, and result statistics are available for different types of data and each estimator. Researchers across fields may find that statsmodels fully meets their needs for statistical computing and data analysis in Python.

Features include:

- Linear regression models
- Generalized linear models
- Discrete choice models
- Robust linear models
- Many models and functions for time series analysis
- Nonparametric estimators
- A collection of datasets for examples
- A wide range of statistical tests
- Input-output tools for producing tables in a number of formats (Text, LaTeX, HTML) and for reading Stata files into NumPy and Pandas.
- Plotting functions
- Extensive unit tests to ensure correctness of results
- Many more models and extensions in development