Learning the Shapiro-Wilk Test: A Practical Guide with Python

Authored by

Mohammed looti

November 7, 2025

The Crucial Role of the Shapiro-Wilk Test in Assessing Normality

The **Shapiro-Wilk test** stands as one of the most reliable and powerful statistical instruments available for rigorously evaluating the assumption of normality within a sampled dataset. It is fundamentally designed to ascertain whether a given set of random observations is statistically likely to have been drawn from a population that adheres to a normal distribution, often referred to as a Gaussian distribution. This preliminary assessment of data distribution is not merely academic; it is a critical prerequisite for the validity of numerous parametric statistical tests, including common procedures like t-tests, Z-tests, and Analysis of Variance (ANOVA).

Statistical testing, particularly the Shapiro-Wilk procedure, begins with the formulation of competing hypotheses. The core premise is defined by the **null hypothesis** (H_0), which conservatively asserts that the sample data is indeed generated from a normally distributed population. Conversely, the alternative hypothesis (H_a) posits that the underlying data distribution deviates significantly from normality. The power of the test lies in calculating a specific test statistic (W) and, more importantly, its corresponding **p-value**. These metrics provide the empirical evidence necessary to decide whether to accept the status quo (H_0) or conclude that the data is statistically non-normal (H_a).

A robust statistical analysis is inextricably linked to understanding the distributional properties of the data being examined. If the Shapiro-Wilk test yields a result that strongly suggests a significant deviation from normality, researchers are faced with a necessary choice: they must either apply mathematical transformations to the data (such as logarithmic or square root transformations) in an attempt to normalize the distribution, or pivot to the use of non-parametric statistical methods. Non-parametric alternatives are valuable because they do not rely on the stringent assumption that the population data follows a Gaussian distribution, thereby preserving the integrity of the analysis when the normality assumption is violated.

Leveraging SciPy for Shapiro-Wilk Implementation in Python

Executing the Shapiro-Wilk test efficiently within the **Python** environment is streamlined through the use of the SciPy library, which is the cornerstone for scientific computing and advanced statistics in Python. The specific function required for this test is `scipy.stats.shapiro()`. This function is meticulously engineered to accept a one-dimensional array or sequence of sample observations and subsequently compute the required statistical metrics for hypothesis testing.

The implementation of this function is remarkably straightforward, demanding only the dataset itself as the primary argument. However, adhering to data integrity standards is essential: the input array must exclusively contain real, finite numerical values. The Shapiro-Wilk test is structurally inappropriate for evaluating categorical data or datasets that contain missing values (NaNs), as its underlying mathematics depends on continuous, countable observations. Therefore, data

preprocessing steps, such as handling missing data and ensuring variable types are numeric, must be completed before invoking the function.

When integrating this test into any Python script or analysis notebook, the general syntax is concise and easily executable, emphasizing the accessibility of powerful statistical tools in the modern data science workflow:

scipy.stats.shapiro(x)

The single required parameter represents the numerical data intended for distribution analysis:

x: This must be a NumPy array, a Pandas Series, or another similar iterable structure containing the sample data points whose normality is being assessed.

Interpreting the Output: The W Statistic and P-Value Decision Rule

The execution of the `scipy.stats.shapiro()` function yields a named tuple containing two paramount values: the test statistic (W) and the associated p-value. The W statistic is an internal measure of how closely the observed data's empirical variance aligns with the variance that would be expected if the data were perfectly normally distributed. Higher values of W (closer to 1.0) suggest a stronger adherence to normality, while values significantly lower than 1.0 indicate substantial non-normal characteristics, such as heavy skewness or kurtosis.

While the W statistic provides context, the ultimate metric for hypothesis decision-making is the p-value. This value quantifies the probability of observing the current sample data, or data that is even more extreme, assuming that the null hypothesis (the assumption of normality) is true. To make a definitive conclusion, we compare this calculated p-value against a predetermined significance level, often denoted as $alpha$ (alpha). The conventional and widely accepted standard for $alpha$ in most scientific fields is 0.05.

The decision rule based on this comparison is unambiguous and dictates the analytical conclusion. If the computed p-value is less than or equal to the chosen significance level ($text{p-value} le alpha$, e.g., 0.05), we possess sufficient statistical evidence to reject the null hypothesis. Rejecting H_0 means we conclude with confidence that the sample data does **not** originate from a normal distribution. Conversely, if the p-value is greater than $alpha$ ($text{p-value} > 0.05$), we fail to reject the null hypothesis. This outcome suggests that the data is statistically consistent with a normal distribution, meaning any observed deviations are likely due to random sampling variability rather than an inherently non-normal population.

Example 1: Demonstrating Normality with Gaussian Data

To properly illustrate the application and interpretation of the Shapiro-Wilk test, we begin by

creating a sample dataset explicitly designed to follow a standard normal distribution. This positive control example helps confirm that the test functions correctly under ideal conditions. We leverage the powerful random number generation capabilities provided by NumPy, creating 100 data points. Crucially, we set a seed for the random generator; this practice ensures that the example is fully reproducible, guaranteeing identical results regardless of when or where the code is executed.

The code snippet below outlines the steps for initializing the NumPy seed and generating the synthetic, normally distributed sample data:

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

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

#generate dataset of 100 random values that follow a standard normal distribution
data = randn(100)
```

Following data generation, we proceed to apply the `scipy.stats.shapiro()` function directly to this array of 100 values. Our primary objective here is to see if the statistical test validates the inherent normality we engineered into the dataset using NumPy's `randn()` function. The following block executes the test and displays the resulting statistic and p-value tuple:

```
from scipy.stats import shapiro

#perform Shapiro-Wilk test
shapiro(data)

ShapiroResult(statistic=0.9926937818527222, pvalue=0.8689165711402893)
```

A thorough analysis of the output reveals a test statistic (W) of approximately **0.9927**, which is extremely close to the ideal value of 1.0, and a corresponding p-value of **0.8689**. Given that this p-value (0.8689) is vastly larger than our standard significance level ($alpha = 0.05$), we must fail to reject the null hypothesis. This powerful statistical conclusion confirms that we lack sufficient evidence to claim the sample data deviates from a normal distribution, aligning perfectly with our expectation based on the data generation method.

Example 2: Detecting Non-Normality with Poisson Distributed Data

In this second crucial example, we shift our focus to demonstrate the test's capability to identify and flag data that clearly violates the assumption of normality. For this purpose, we generate a

sample drawn from a **Poisson distribution**, which is inherently skewed and discrete, thus guaranteed to be non-normal, particularly with a small mean ($lambda$). As in the previous example, we maintain reproducibility by setting the random seed.

The following Python code generates 100 data points configured to follow a Poisson distribution, utilizing a mean (lambda) value of 5:

```
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)
```

We now apply the Shapiro-Wilk test to this newly generated Poisson sample. Since we know the underlying distribution is non-normal, our expectation is a result that strongly compels us to reject the null hypothesis, thereby confirming the test's ability to detect non-Gaussian features in the data.

```
from scipy.stats import shapiro

#perform Shapiro-Wilk test
shapiro(data)

ShapiroResult(statistic=0.9581913948059082, pvalue=0.002994443289935589)
```

The resulting output shows a test statistic (W) of **0.9582** and a corresponding p-value of **0.00299**. Critically, this p-value is significantly smaller than the conventional significance level of $alpha = 0.05$. Based on the decision rule, we firmly reject the null hypothesis. This outcome provides compelling statistical evidence that the sample data is inconsistent with a normal distribution, confirming that the test successfully identified the non-normal nature introduced by the Poisson generation process.

Conclusion and Resources for Further Statistical Exploration

The Shapiro-Wilk test is an essential and highly effective component of any thorough data analysis pipeline, serving to validate the fundamental normality assumption required by powerful parametric statistical methods. By integrating the robust `scipy.stats.shapiro()` function into Python workflows, data analysts can quickly and reliably ascertain the distributional characteristics of their

sample data. The clear interpretation of the resulting p-value against the predefined significance level provides an actionable determination of whether the data is suitably normal for subsequent advanced statistical procedures.

For researchers and practitioners seeking to explore alternative methodologies for assessing data normality, or those working across different statistical computing environments, the following resources offer valuable guidance and comparative tutorials:

How to Perform a Shapiro-Wilk Test in R
How to Perform an Anderson-Darling Test in Python
How to Perform a Kolmogorov-Smirnov Test in Python

The original external links for these related statistical procedures are maintained below for direct reference:

How to Perform a Shapiro-Wilk Test in R
How to Perform an Anderson-Darling Test in Python
How to Perform a Kolmogorov-Smirnov Test in Python