How can a Shapiro-Wilk test be performed in SAS?

Authored by

stats writer

July 1, 2024

The Shapiro-Wilk test is a statistical test used to determine if a given set of data follows a normal distribution. In SAS, this test can be performed by using the PROC UNIVARIATE procedure. This procedure allows the user to specify the variable of interest and the desired level of significance for the test. The output of the procedure includes the Shapiro-Wilk test statistic, p-value, and a visual plot to assess the normality of the data. By performing the Shapiro-Wilk test in SAS, users can confidently assess the normality of their data and make informed decisions in their statistical analyses.

Perform a Shapiro-Wilk Test in SAS

The Shapiro-Wilk test is used to determine whether or not a dataset follows a normal distribution.

The following step-by-step example shows how to perform a Shapiro-Wilk test for a dataset in SAS.

Step 1: Create the Data

First, we'll create a dataset that contains 15 observations:

```
/*create dataset*/
data my_data;
input x;
datalines;
3
3
4
6
```

```
7
8
8
9
12
14
15
15
17
20
21
;
run;

/*view dataset*/
proc printdata=my_data;
```

Obs	x
1	3
2	3
3	4
4	6
5	7
6	8
7	8
8	9
9	12
10	14
11	15
12	15
13	17
14	20
15	21

Step 2: Perform the Shapiro-Wilk Test

Next, we'll use proc univariate with the normal command to perform a Shapiro-Wilk test for normality:

```
/*perform Shapiro-Wilk test*/proc univariatedata=my_data normal;
run;
```

The UNIVARIATE Procedure
Variable: x

Moments			
N	15	Sum Weights	15
Mean	10.8	Sum Observations	162
Std Deviation	5.96657356	Variance	35.6
Skewness	0.30348727	Kurtosis	-1.1151814
Uncorrected SS	2248	Corrected SS	498.4
Coeff Variation	55.2460514	Std Error Mean	1.54056267

Basic Statistical Measures			
Location		Variability	
Mean	10.80000	Std Deviation	5.96657
Median	9.00000	Variance	35.60000
Mode	3.00000	Range	18.00000
		Interquartile Range	9.00000

Note: The mode displayed is the smallest of 3 modes with a count of 2.

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	7.010426	Pr > |t|	<.0001
Sign	M	7.5	Pr >= |M|	<.0001
Signed Rank	S	60	Pr >= |S|	<.0001

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.936921	Pr < W	0.3452
Kolmogorov-Smirnov	D	0.151886	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.052851	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.331213	Pr > A-Sq	>0.2500

The output provides us with a ton of information, but the only table we need to look at is the one titled Tests for Normality.

This table provides the test statistics and p-values for several normality tests including:

The Shapiro-Wilk TestThe Kolmogorov-Smirnov TestThe Cramer-von Mises TestThe Anderson-Darling Test

From this table we can see that the p-value for the Shapiro-Wilk test is .3452.

Recall that a Shapiro-Wilk test uses the following null and alternative :

H0: The data is normally distributed.HA: The data is *not* normally distributed.

Since the p-value (.3452) is not less than .05, we fail to reject the null hypothesis.

In other words, it's safe to assume that the dataset is normally distributed.

Additional Resources

The following tutorials explain how to perform other common statistical tests in SAS: