

Chapter 670

Normality Tests (Simulation)

Introduction

This procedure allows you to study the power and sample size of eight statistical tests of normality. Since there are no formulas that allow the calculation of power directly, simulation is used. This gives you the ability to compare the adequacy of each test under a wide variety of solutions.

The reason there are so many different normality tests is that there are many different forms of normality. Thode (2002) presents the following recommendations concerning which tests to use for each situation. Note that the details of each test will be presented later.

Normal vs. Long-Tailed Symmetric Alternative Distributions

The Shapiro-Wilk and the kurtosis tests have been found to be best for normality testing against long-tailed symmetric alternatives.

Normal vs. Short-Tailed Symmetric Alternative Distributions

The Shapiro-Wilk and the range tests have been found to be best for normality testing against short-tailed symmetric alternatives.

Normal vs. Asymmetric Alternative Distributions

The Shapiro-Wilk and the skewness tests have been found to be best for normality testing against asymmetric alternatives.

Technical Details

Computer simulation allows one to estimate the power and significance level that is actually achieved by a test procedure in situations that are not mathematically tractable. Computer simulation was once limited to mainframe computers. Currently, due to increased computer speeds, simulation studies can be completed on desktop and laptop computers in a reasonable period of time.

The steps to a simulation study are as follows.

1. Specify which the normality test is to be used. This includes specifying the significance level.
2. Generate a random sample, X_1, X_2, \dots, X_n , from the distribution specified by the alternative hypothesis. Calculate the test statistic from the simulated data and determine if the null hypothesis is accepted or rejected. Each of these samples is used to calculate the power of the test. Note that if the alternative distribution is set to normal, the power is the significance level of the test.
3. Repeat step 2 several hundred times, tabulating the number of times the simulated data lead to a rejection of the null hypothesis. The power is the proportion of simulation samples in step 2 that lead to rejection.

Data Distributions

A wide variety of distributions may be studied. These distributions can vary in skewness, elongation, or other features such as bimodality. A detailed discussion of the distributions that may be used in the simulation is provided in the chapter 'Data Simulator'.

Test Statistics

This section describes the test statistics that are available of study in this procedure.

Anderson-Darling Test

This test, developed by Anderson and Darling (1954), is a popular normality test based on EDF statistics. In some situations, it has been found to be as powerful as the Shapiro-Wilk test. This test is available when n is greater than or equal to 8.

Kolmogorov-Smirnov

This test for normality is based on the maximum difference between the observed distribution and expected cumulative-normal distribution. Since it uses the sample mean and standard deviation to calculate the expected normal distribution, the Lilliefors' adjustment is used. The Lilliefors' adjusted critical values used are those given by Dallal (1986).

This test is available when n is greater than or equal to 3.

This test has been shown to be less powerful than the other tests in most situations. It is included because of its historical popularity.

Kurtosis

D'Agostino et al. (1990) describes a normality test based on the kurtosis coefficient, b_2 . Recall that for the normal distribution, the theoretical value of b_2 is 3. Hence, a test can be developed to determine if the value of b_2 is significantly different from 3. If it is, the data are obviously non-normal. The statistic, z_k , is, under the null hypothesis of normality, approximately normally distributed for sample sizes $n > 20$. This test is available when n is greater than or equal to 8.

The calculation of this test proceeds as follows:

$$z_k = \frac{\left(1 - \frac{2}{9A}\right) - \left(\frac{1 - \frac{2}{A}}{1 + G\sqrt{\frac{2}{A-4}}}\right)^{1/3}}{\sqrt{\frac{2}{9A}}}$$

where

$$b_2 = \frac{m_4}{m_2^2}$$

Normality Tests (Simulation)

$$G = \frac{b_2 - \left(\frac{3n-3}{n+1}\right)}{\sqrt{\frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)}}}$$

$$E = \frac{6(n^2 - 5n + 2)}{(n+7)(n+9)} \sqrt{\frac{6(n+3)(n+5)}{n(n-2)(n-3)}}$$

$$A = 6 + \frac{8}{E} \left(\frac{2}{E} + \sqrt{1 + \frac{4}{E^2}} \right)$$

Martinez-Iglewicz

This test for normality, developed by Martinez and Iglewicz (1981), is based on the median and a robust estimator of dispersion. They have shown that this test is very powerful for heavy-tailed symmetric distributions as well as a variety of other situations. A value of the test statistic that is close to one indicates that the distribution is normal. This test is recommended for exploratory data analysis by Hoaglin et al. (1983). The formula for this test is:

$$I = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{(n-1)s_{bi}^2}$$

where s_{bi}^2 is a biweight estimator of scale.

This test is available when n is greater than or equal to 3.

Omnibus

D'Agostino et al. (1990) describes a normality test that combines the tests for skewness and kurtosis. The statistic, K^2 , is approximately distributed as a chi-square with two degrees of freedom.

After calculating z_s and z_k , calculate K^2 as follows:

$$K^2 = z_s^2 + z_k^2$$

This test is available when n is greater than or equal to 8.

Range Test

The range test, u , was created to test for normality when the alternative distribution is actually the uniform distribution. It is calculated by dividing the range by the standard deviation. Tables of critical values of the range test are given in Pearson and Hartley (1976) for n equal 3 to 1000.

Normality Tests (Simulation)

Shapiro-Wilk W Test

This test for normality, developed by Shapiro and Wilk (1965), has been found to be the most powerful test in most situations. It is the ratio of two estimates of the variance of a normal distribution based on a random sample of n observations. The numerator is proportional to the square of the best linear estimator of the standard deviation. The denominator is the sum of squares of the observations about the sample mean. W may be written as the square of the Pearson correlation coefficient between the ordered observations and a set of weights which are used to calculate the numerator. Since these weights are asymptotically proportional to the corresponding expected normal order statistics, W is roughly a measure of the straightness of the normal quantile-quantile plot. Hence, the closer W is to one, the more normal the sample.

The probability values for W are valid for samples where n is greater than or equal to 3.

Skewness

D'Agostino et al. (1990) describes a normality test based on the skewness coefficient, $\sqrt{b_1}$. Recall that because the normal distribution is symmetrical, $\sqrt{b_1}$ is equal to zero for normal data. Hence, a test can be developed to determine if the value of $\sqrt{b_1}$ is significantly different from zero. If it is, the data are obviously non-normal. The statistic, z_s , is, under the null hypothesis of normality, approximately normally distributed. The computation of this statistic, which is restricted to sample sizes $n > 8$, is

$$z_s = d \ln \left(\frac{T}{a} + \sqrt{\left(\frac{T}{a} \right)^2 + 1} \right)$$

where

$$b_1 = \frac{m_3^2}{m_2^3}$$

$$T = \sqrt{b_1 \left(\frac{(n+1)(n+3)}{6(n-2)} \right)}$$

$$C = \frac{3(n^2 + 27n - 70)(n+1)(n+3)}{(n-2)(n+5)(n+7)(n+9)}$$

$$W^2 = -1 + \sqrt{2(C-1)}$$

$$a = \sqrt{\frac{2}{W^2 - 1}}$$

$$d = \frac{1}{\sqrt{\ln(W)}}$$

Example 1 – Power at Various Sample Sizes

A researcher is planning an experiment to test whether a certain data distribution is reasonably close to normality. He will begin his research by generating data from the exponential distribution. The researcher is particularly interested in the Shapiro-Wilk normality test, but he wants to see the power of various other choices when the data actually come from an exponential distribution. Alpha is set to 0.05. He wants to compute the power at various sample sizes from 5 to 40. Since this is an exploratory analysis, he sets the number of simulation iterations to 1000.

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 1** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

```

Design Tab
-----
Solve For ..... Power
Search/Report Test..... Shapiro-Wilk
Simulations ..... 1000
Random Seed..... 3801875 (for Reproducibility)
Alpha..... 0.05
N (Sample Size)..... 5 10 15 20 30 40
Data Distribution ..... Exponential(M1)
M1 Parameter Value(s)..... 1
    
```

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

```

Numeric Results
-----
Solve For:      Power
Test Type:     Shapiro-Wilk Normality Test
Actual Distribution: Exponential(M1)
-----


| Power                 | 95% Confidence Interval<br>Limits for Power |       | Sample<br>Size<br>N | Alpha | Beta  | M1 | Number |
|-----------------------|---------------------------------------------|-------|---------------------|-------|-------|----|--------|
|                       | Lower                                       | Upper |                     |       |       |    |        |
| <a href="#">0.150</a> | 0.128                                       | 0.174 | 5                   | 0.05  | 0.850 | 1  | 1      |
| <a href="#">0.468</a> | 0.437                                       | 0.499 | 10                  | 0.05  | 0.532 | 1  | 2      |
| <a href="#">0.676</a> | 0.646                                       | 0.705 | 15                  | 0.05  | 0.324 | 1  | 3      |
| <a href="#">0.819</a> | 0.794                                       | 0.842 | 20                  | 0.05  | 0.181 | 1  | 4      |
| <a href="#">0.964</a> | 0.951                                       | 0.975 | 30                  | 0.05  | 0.036 | 1  | 5      |
| <a href="#">0.993</a> | 0.986                                       | 0.997 | 40                  | 0.05  | 0.007 | 1  | 6      |


-----
Simulations: 1000. Run Time: 1.32 seconds.
User-Entered Random Seed: 3801875
    
```

Normality Tests (Simulation)

Power	The probability of rejecting a false null hypothesis when the alternative hypothesis is true.
Lower and Upper 95% C.L.	The boundaries of an exact binomial confidence interval for power.
N	The size of the sample drawn from the population.
Alpha	The probability of rejecting when the distribution is normal.
Beta	The probability of concluding normality when the distribution is actually the one stated.
M0, M1, S, etc.	The value(s) of the user-specified distribution parameters.
Number	The sequence number of this combination of parameter values.

Summary Statements

A single-group design will be used to test whether the distribution is non-normal. The comparison will be made using a Shapiro-Wilk test with a Type I error rate (α) of 0.05. To detect the Exponential(M1) distribution with a sample size of 5, the power is 0.15. These results are based on 1000 simulations (Monte Carlo samples) from the Exponential(M1) distribution.

Dropout-Inflated Sample Size

Dropout Rate	Sample Size N	Dropout-Inflated Enrollment Sample Size N'	Expected Number of Dropouts D
20%	5	7	2
20%	10	13	3
20%	15	19	4
20%	20	25	5
20%	30	38	8
20%	40	50	10

Dropout Rate	The percentage of subjects (or items) that are expected to be lost at random during the course of the study and for whom no response data will be collected (i.e., will be treated as "missing"). Abbreviated as DR.
N	The evaluable sample size at which power is computed (as entered by the user). If N subjects are evaluated out of the N' subjects that are enrolled in the study, the design will achieve the stated power.
N'	The total number of subjects that should be enrolled in the study in order to obtain N evaluable subjects, based on the assumed dropout rate. N' is calculated by inflating N using the formula $N' = N / (1 - DR)$, with N' always rounded up. (See Julious, S.A. (2010) pages 52-53, or Chow, S.C., Shao, J., Wang, H., and Lohknygina, Y. (2018) pages 32-33.)
D	The expected number of dropouts. $D = N' - N$.

Dropout Summary Statements

Anticipating a 20% dropout rate, 7 subjects should be enrolled to obtain a final sample size of 5 subjects.

References

Thode, Henry C. Jr. 2002. Testing for Normality. Marcel Dekker. New York.
 Devroye, Luc. 1986. Non-Uniform Random Variate Generation. Springer-Verlag. New York.

This report shows the estimated power with its confidence interval (using the binomial distribution) for each combination of parameter values. Note that because these are results of a simulation study, the computed power and alpha will vary from run to run. Thus, another report obtained using the same input parameters will be slightly different from the one above. If the researcher wants 80% power, he will need an N of 20. To achieve 90% power, he will need an N of 30.

Normality Tests (Simulation)

Comparative Results

Power of Various Normality Tests

Actual Distribution: Exponential(M1)

Number	Sample Size N	Power of Individual Normality Tests								Any Test Power
		Anderson-Darling	Kolmogorov-Smirnov	Kurtosis	Martinez-Iglewicz	Omnibus	Range	Shapiro-Wilk	Skewness	
1	5		0.137		0.216		0.045	0.150		0.263
	Lower 95% C.L.		0.116		0.191		0.033	0.128		0.236
	Upper 95% C.L.		0.160		0.243		0.060	0.174		0.291
2	10	0.441	0.316	0.224	0.344	0.330	0.060	0.468	0.369	0.560
	Lower 95% C.L.	0.410	0.287	0.199	0.315	0.301	0.046	0.437	0.339	0.529
	Upper 95% C.L.	0.472	0.346	0.251	0.374	0.360	0.077	0.499	0.400	0.591
3	15	0.628	0.452	0.307	0.471	0.486	0.078	0.676	0.546	0.750
	Lower 95% C.L.	0.597	0.421	0.279	0.440	0.455	0.062	0.646	0.515	0.722
	Upper 95% C.L.	0.658	0.483	0.337	0.502	0.517	0.096	0.705	0.577	0.777
4	20	0.764	0.550	0.339	0.560	0.564	0.095	0.819	0.684	0.859
	Lower 95% C.L.	0.736	0.519	0.310	0.529	0.533	0.078	0.794	0.654	0.836
	Upper 95% C.L.	0.790	0.581	0.369	0.591	0.595	0.115	0.842	0.713	0.880
5	30	0.939	0.785	0.483	0.737	0.778	0.110	0.964	0.884	0.971
	Lower 95% C.L.	0.922	0.758	0.452	0.709	0.751	0.091	0.951	0.863	0.959
	Upper 95% C.L.	0.953	0.810	0.514	0.764	0.803	0.131	0.975	0.903	0.980
6	40	0.983	0.845	0.559	0.823	0.881	0.159	0.993	0.951	0.994
	Lower 95% C.L.	0.973	0.821	0.528	0.798	0.859	0.137	0.986	0.936	0.987
	Upper 95% C.L.	0.990	0.867	0.590	0.846	0.900	0.183	0.997	0.964	0.998

Simulations: 1000. Run Time: 0.85 seconds.

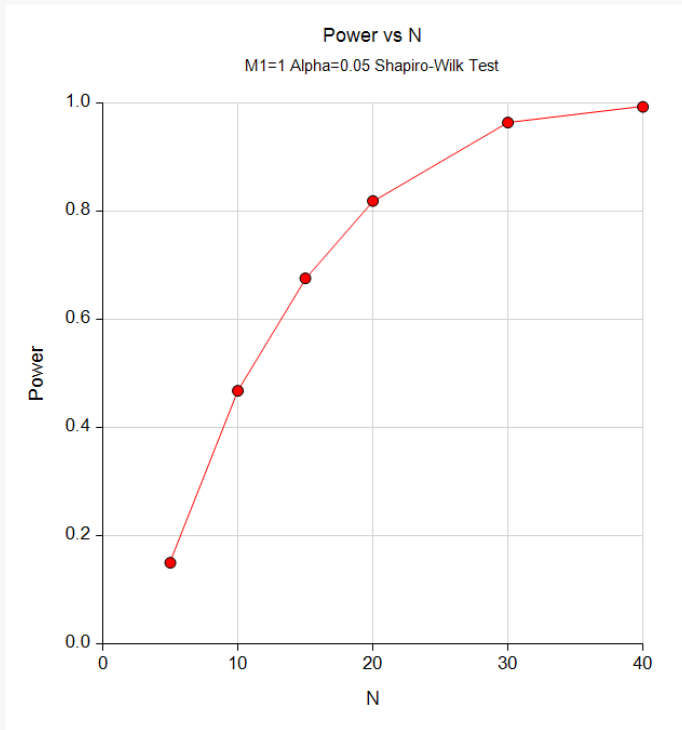
User-Entered Random Seed: 3801875

This report shows the estimated power with its confidence interval (using the binomial distribution) for each combination of each of the normality tests. The last test, called *Any Test*, is calculated by combining the results of the eight normality tests into a single test.

Note the variation in the power values. In this case, the Martinez-Iglewicz is the champion for N = 5, but the Shapiro-Wilk test has greater power for N = 10.

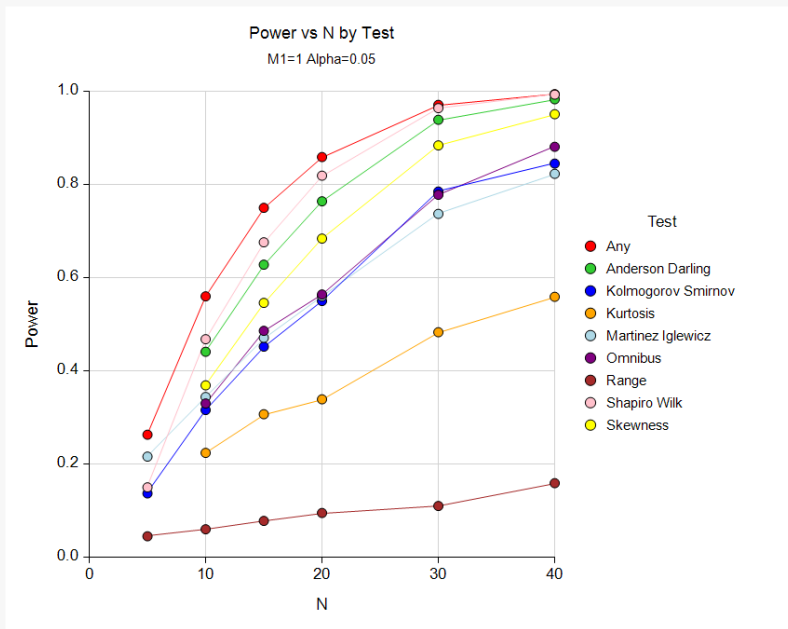
Plots Section

Plots



This plot shows the relationship between sample size and power.

Comparative Plots



This plot shows performance of each of the normality tests. The champion is the Shapiro-Wilk test. The range test, which was designed for uniform alternatives, does very poorly with exponential alternatives.

Example 2 – Validation using the Normal Distribution

We will validate this procedure by setting the normal distribution as the alternative distribution. In this case, the power should be equal to the alpha value. Alpha is set to 0.05. The sample sizes will be set to 10, 30, and 50. The number of simulation iterations to 10,000.

For reproducibility, we'll use a random seed of 5400722.

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 2** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

Solve For **Power**
 Search/Report Test..... **Shapiro-Wilk**
 Simulations **10000**
 Random Seed..... **5400722** (for Reproducibility)
 Alpha..... **0.05**
 N (Sample Size)..... **10 30 50**
 Data Distribution **Normal(M1 S)**
 M1 Parameter Value(s)..... **1**
 Parameter 1 Label **S**
 Parameter 1 Value(s)..... **1**

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Solve For: [Power](#)
 Test Type: Shapiro-Wilk Normality Test
 Actual Distribution: Normal(M1 S)

Power	95% Confidence Interval Limits for Power		Sample Size N	Alpha	Beta	M1	S	Number
	Lower	Upper						
0.047	0.043	0.052	10	0.05	0.953	1	1	1
0.051	0.047	0.056	30	0.05	0.949	1	1	2
0.048	0.044	0.052	50	0.05	0.952	1	1	3

Simulations: 10000. Run Time: 6.78 seconds.
 User-Entered Random Seed: 5400722

Normality Tests (Simulation)

Power of Various Normality Tests

Actual Distribution: Normal(M1 S)

Number	Sample Size N	Power of Individual Normality Tests								Any Test Power
		Anderson-Darling	Kolmogorov-Smirnov	Kurtosis	Martinez-Iglewicz	Omnibus	Range	Shapiro-Wilk	Skewness	
1	10	0.048	0.024	0.039	0.048	0.056	0.050	0.047	0.051	0.134
	Lower 95% C.L.	0.044	0.021	0.035	0.044	0.051	0.046	0.043	0.047	0.128
	Upper 95% C.L.	0.052	0.027	0.043	0.052	0.060	0.054	0.052	0.056	0.141
2	30	0.051	0.032	0.053	0.051	0.057	0.049	0.051	0.049	0.151
	Lower 95% C.L.	0.047	0.028	0.049	0.047	0.053	0.045	0.047	0.045	0.144
	Upper 95% C.L.	0.055	0.035	0.058	0.056	0.062	0.054	0.056	0.054	0.158
3	50	0.051	0.030	0.051	0.046	0.056	0.051	0.048	0.050	0.156
	Lower 95% C.L.	0.047	0.027	0.046	0.042	0.052	0.047	0.044	0.045	0.149
	Upper 95% C.L.	0.055	0.033	0.055	0.050	0.061	0.055	0.052	0.054	0.164

Simulations: 10000. Run Time: 6.88 seconds.

User-Entered Random Seed: 5400722

This report confirms that the simulation results are quite accurate. Most of the power values are very close to 0.05. The Kolmogorov-Smirnov test appears to be low, and the omnibus test appears to be high. Perhaps this is the reason that the Shapiro-Wilk and the Anderson-Darling tests are usually recommended for general normality testing.