

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.

Normality Tests (Simulation)

- 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.

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.

Normality Tests (Simulation)

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

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

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

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

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

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

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

Procedure Options

This section describes the options that are specific to this procedure. These are located on the Design tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Design Tab

The Design tab contains most of the parameters and options that you will be concerned with.

Solve For

Solve For

This option specifies the parameter to be calculated using the values of the other parameters. Under most conditions, you would select either *Power* or *Sample Size*.

Select *Power* when you want to estimate the power for a specific scenario.

Select *Sample Size* when you want to determine the sample size needed to achieve a given power and alpha error level. This option can be very computationally intensive, and may take considerable time to complete.

Normality Tests (Simulation)

Test and Simulations

Search/Report Test

Specify the specific normality test to be used for searching (if Solve For = 'Sample Size') and reporting. Note that each test was developed for a specific situation and that no one test is best in all situations. A reasonable choice, if you have no other information, might be the Shapiro-Wilk test.

Simulations

This option specifies the number of iterations, M , used in the simulation. Larger numbers of iterations result in longer running time and more accurate results.

The precision of the simulated power estimates can be determined by recognizing that they follow the binomial distribution. Thus, confidence intervals may be constructed for power estimates. The following table gives an estimate of the precision that is achieved for various simulation sizes when the power is either 0.50 or 0.95. The table values are interpreted as follows: a 95% confidence interval of the true power is given by the power reported by the simulation plus and minus the 'Precision' amount given in the table.

Simulation Size M	Precision when Power = 0.50	Precision when Power = 0.95
100	0.100	0.044
500	0.045	0.019
1000	0.032	0.014
2000	0.022	0.010
5000	0.014	0.006
10000	0.010	0.004
50000	0.004	0.002
100000	0.003	0.001

Notice that a simulation size of 1000 gives a precision of plus or minus 0.014 when the true power is 0.95. Also note that as the simulation size is increased beyond 5000, there is only a small amount of additional precision achieved.

Power and Alpha

Power

This option specifies one or more values for power. Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as *0.8 to 0.95 by 0.05* may be entered.

Alpha

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected.

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as *0.01 0.05 0.10* or *0.01 to 0.10 by 0.01*.

Normality Tests (Simulation)

Sample Size

N (Sample Size)

This option specifies one or more values of the sample size, the number of individuals in the study. This value must be an integer greater than one. Note that you may enter a list of values using the syntax *50 100 150 200 250* or *50 to 250 by 50*.

Effect Size

Data Distribution

This option specifies distribution that is being compared to normality. That is, this is the actual (true) distribution of the data from which the power is computed. Usually, the mean is specified by entering 'M1' for the mean parameter in the distribution expression and then entering values for the M1 parameter below. All of the distributions are parameterized so that the mean is entered first.

The parameters of each distribution are specified using numbers or letters. If letters are used, their values are specified in the boxes below. The values 'M0' and 'M1' are usually used to specify the value of the mean.

Following is a list of the distributions that are available and the syntax used to specify them. Each of the parameters should be replaced with a number or parameter name.

Distributions with Common Parameters

Beta(Shape1, Shape2, Min, Max)

Binomial(P, N)

Cauchy(Mean, Scale)

Constant(Value)

Exponential(Mean)

Gamma(Shape, Scale)

Gumbel(Location, Scale)

Laplace(Location, Scale)

Logistic(Location, Scale)

Lognormal(Mu, Sigma)

Multinomial(P1, P2, P3, ..., Pk)

Normal(Mean, Sigma)

Poisson(Mean)

TukeyGH(Mu, S, G, H)

Uniform(Min, Max)

Weibull(Shape, Scale)

Normality Tests (Simulation)

Distributions with Mean and SD Parameters

BetaMS(Mean, SD, Min, Max)

BinomialMS(Mean, N)

GammaMS(Mean, SD)

GumbelMS(Mean, SD)

LaplaceMS(Mean, SD)

LogisticMS(Mean, SD)

LognormalMS(Mean, SD)

UniformMS(Mean, SD)

WeibullMS(Mean, SD)

Details of writing mixture distributions, combined distributions, and compound distributions are found in the chapter on *Data Simulation* and will not be repeated here.

Effect Size – Distribution Parameters

M0 and M1

These values are substituted for the M0 and M1 in the distribution specifications given above. M0 is intended to be the value of the mean.

You can enter a list of values using syntax such as *0 1 2 3* or *0 to 3 by 1*.

Parameter Values (S, A, B, C)

Enter the numeric value(s) of parameter listed above. These values are substituted for the corresponding letter in the Data Distribution specifications.

You can enter a list of values using syntax such as *0 1 2 3* or *0 to 3 by 1*.

You can also change the letter that is used as the name of this parameter.

Normality Tests (Simulation)

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

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Normality Tests (Simulation)** procedure window by clicking on **Normality**, and then clicking on **Normality Tests (Simulation)**. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Power
Search/Report Test.....	Shapiro-Wilk
Simulations	1000
Alpha.....	0.05
N (Sample Size).....	5 10 15 20 30 40
Data Distribution	Exponential(M1)
M0	0
M1	1

Annotated Output

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

Numeric Results

Numeric Results for the Shapiro-Wilk Normality Test Actual Distribution: Exponential(M1)

	Lower 95% C.L. of Power	Upper 95% C.L. of Power	N	Alpha	Beta	M1	No.
0.175	0.152	0.200	5	0.050	0.825	1.0	1
0.442	0.411	0.473	10	0.050	0.558	1.0	2
0.690	0.660	0.719	15	0.050	0.310	1.0	3
0.840	0.816	0.862	20	0.050	0.160	1.0	4
0.968	0.955	0.978	30	0.050	0.032	1.0	5
0.996	0.990	0.999	40	0.050	0.004	1.0	6

Notes

Simulations: 1000. Run Time: 1.98 seconds.

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.

Normality Tests (Simulation)

Report Definitions

Power is the probability of rejecting a false null hypothesis.
 Lower/Upper 95% C.L. are the boundaries of an exact binomial confidence interval for power.
 N is the size of the sample drawn from the population.
 Alpha is the probability of rejecting when the distribution is normal.
 Beta is the probability of concluding normality when the distribution is actually the one stated.
 No. is the sequence number of this combination of parameter values.

Summary Statements

A sample size of 5 achieves 18% power to detect non-normality using the Shapiro-Wilk test when the significance level is 0.050 and the actual distribution is Expo(M1). These results are based on 1000 Monte Carlo samples from this distribution.

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.

Comparative Results

Power of Various Normality Tests Actual Distribution: Exponential(M1)

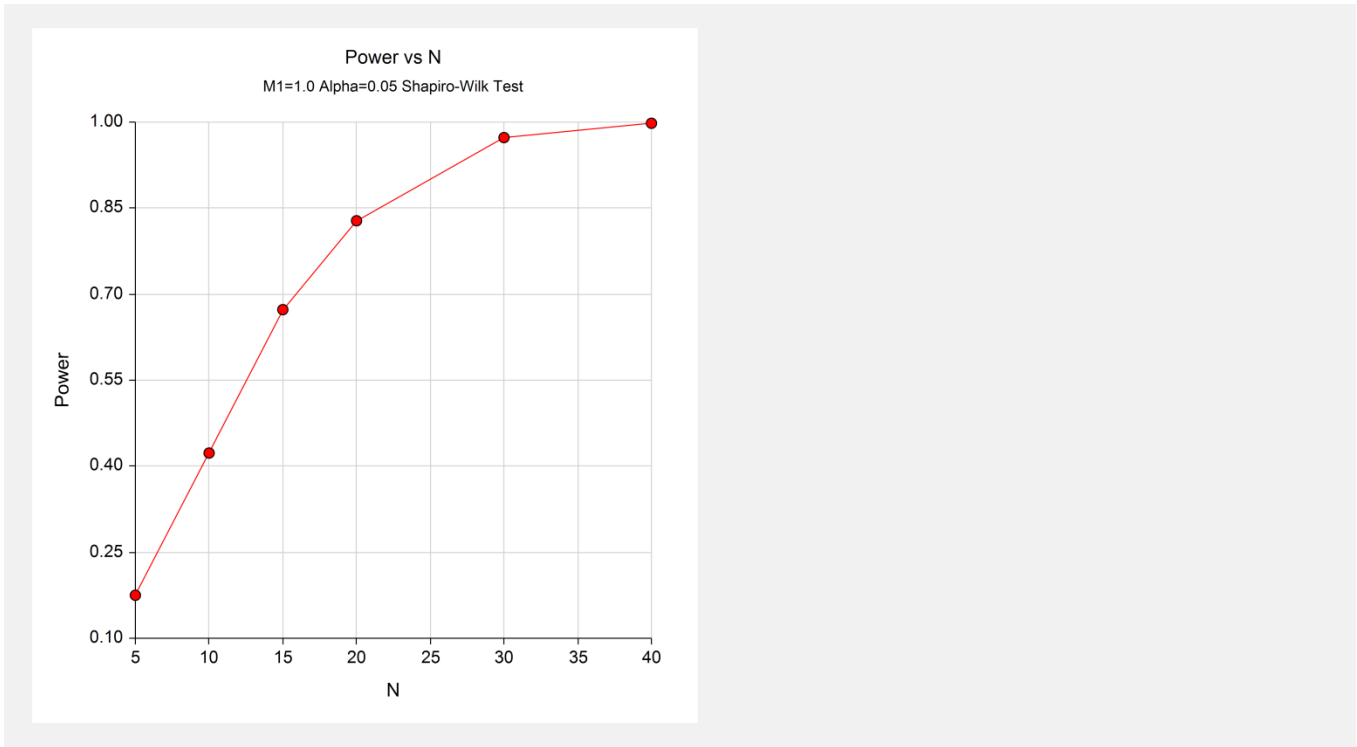
Sequence No.	N	Anderson Darling	Kolmo'v Smirnov	Kurtosis	Martinez Iglewicz	Omnibus	Range /SD	Shapiro Wilk	Skewness	Any Test
1	5		0.124		0.225		0.076	0.175		0.290
		Lower 95% C.L.	0.104		0.199		0.060	0.152		0.262
		Upper 95% C.L.	0.146		0.252		0.094	0.200		0.319
2	10	0.401	0.297	0.215	0.311	0.315	0.062	0.442	0.366	0.549
		Lower 95% C.L.	0.370	0.190	0.282	0.286	0.048	0.411	0.336	0.518
		Upper 95% C.L.	0.432	0.242	0.341	0.345	0.079	0.473	0.397	0.580
3	15	0.627	0.448	0.290	0.456	0.474	0.068	0.690	0.562	0.750
		Lower 95% C.L.	0.596	0.417	0.425	0.443	0.053	0.660	0.531	0.722
		Upper 95% C.L.	0.657	0.479	0.319	0.487	0.085	0.719	0.593	0.777
4	20	0.784	0.597	0.349	0.598	0.607	0.074	0.840	0.713	0.871
		Lower 95% C.L.	0.757	0.319	0.567	0.576	0.059	0.816	0.684	0.849
		Upper 95% C.L.	0.809	0.379	0.629	0.637	0.092	0.862	0.741	0.891
5	30	0.937	0.775	0.485	0.735	0.777	0.131	0.968	0.895	0.975
		Lower 95% C.L.	0.920	0.454	0.706	0.750	0.111	0.955	0.874	0.963
		Upper 95% C.L.	0.951	0.516	0.762	0.802	0.154	0.978	0.913	0.984
6	40	0.987	0.886	0.588	0.839	0.894	0.142	0.996	0.958	0.998
		Lower 95% C.L.	0.978	0.557	0.815	0.873	0.121	0.990	0.944	0.993
		Upper 95% C.L.	0.993	0.619	0.861	0.912	0.165	0.999	0.970	1.000

This report shows the estimated power with its confidence interval (using the binomial distribution) for each combination 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.

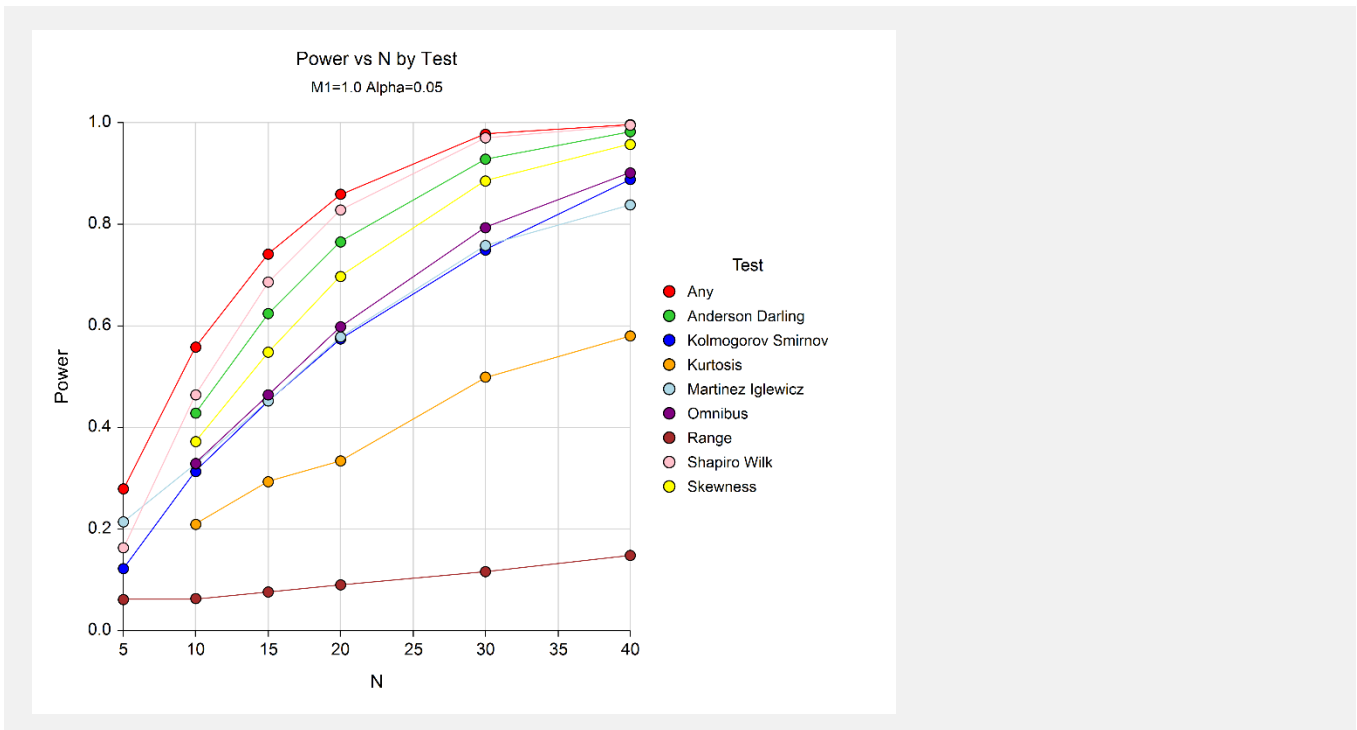
Note the variation in the power values. In this case, the Martinez-Iglewicz is the champion for $N = 5$, but the Shapiro-Wilk test has greater power for $N = 10$.

Normality Tests (Simulation)

Plots Section



This plot shows the relationship between sample size and power.



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.

Normality Tests (Simulation)

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.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Normality Tests (Simulation)** procedure window by clicking on **Normality**, and then clicking on **Normality Tests (Simulation)**. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Power
Search/Report Test.....	Shapiro-Wilk
Simulations	10000
Alpha.....	0.05
N (Sample Size).....	10 30 50
Data Distribution	Normal(M1 S)
M1	1
S.....	1

Output

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

Numeric Results

Numeric Results for the Shapiro-Wilk Normality Test Actual Distribution: Normal(M1 S)

	Lower 95% C.L. of Power	Upper 95% C.L. of Power	N	Alpha	Beta	M1	S	No.
0.052	0.047	0.056	10	0.050	0.948	1.0	1.0	1
0.047	0.043	0.051	30	0.050	0.953	1.0	1.0	2
0.053	0.048	0.057	50	0.050	0.948	1.0	1.0	3

Power of Various Normality Tests Actual Distribution: Normal(M1 S)

Sequence No.	N	Anderson Darling	Kolmo'v Smirnov	Kur- tosis	Martinez Iglewicz	Omnibus	Range /SD	Shapiro Wilk	Skew- ness	Any Test
1	10	0.053	0.027	0.043	0.051	0.058	0.049	0.052	0.052	0.142
	Lower 95% C.L.	0.048	0.024	0.039	0.046	0.053	0.045	0.047	0.048	0.135
	Upper 95% C.L.	0.057	0.030	0.047	0.055	0.062	0.053	0.056	0.057	0.148
2	30	0.048	0.029	0.053	0.049	0.053	0.053	0.047	0.047	0.151
	Lower 95% C.L.	0.044	0.026	0.049	0.045	0.048	0.049	0.043	0.043	0.145
	Upper 95% C.L.	0.053	0.033	0.058	0.054	0.057	0.058	0.051	0.052	0.159
3	50	0.052	0.032	0.055	0.049	0.057	0.054	0.053	0.048	0.159
	Lower 95% C.L.	0.048	0.028	0.050	0.045	0.052	0.050	0.048	0.044	0.152
	Upper 95% C.L.	0.056	0.035	0.059	0.053	0.061	0.059	0.057	0.053	0.167

Simulations: 10000. Run Time: 10.67 seconds.

Normality Tests (Simulation)

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.