Chapter 410

Tests for One Mean (Simulation)

Introduction

This procedure allows you to study the power and sample size of several statistical tests of the hypothesis that the population mean is equal to a specific value versus the alternative that it is greater than, less than, or not equal to that value. The one-sample t-test is commonly used in this situation, but other tests have been developed for situations where the data are not normally distributed. These additional tests include the Wilcoxon signed-rank test, the sign test, and the computer-intensive bootstrap test. When the population follows the exponential distribution, a test based on this distribution should be used.

The t-test assumes that the data are normally distributed. When this assumption does not hold, the t-test is still used hoping that its robustness will produce accurate results. This procedure allows you to study the accuracy of various tests using simulation techniques. A wide variety of distributions can be simulated to allow you to assess the impact of various forms of non-normality on each test's accuracy.

The details of the power analysis of the t-test using analytic techniques are presented in the **PASS** chapter entitled "One-Sample T-Tests" and will not be duplicated here. This chapter will be confined to power analysis using computer simulation.

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 the method by which the test is to be carried out. This includes specifying how the test statistic is calculated and how the significance level is specified.

2. Generate a random sample, $X_1, X_2, \ldots, X_n$, 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.

3. Generate a random sample, $Y_1, Y_2, \ldots, Y_n$, from the distribution specified by the null 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 significance level of the test.

4. Repeat steps 2 and 3 several thousand 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. The significance level is the proportion of simulated samples in step 3 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 in this procedure.

One-Sample T-Test

The one-sample t-test assumes that the data are a simple random sample from a population of normally distributed values that all have the same mean and variance. This assumption implies that the data are continuous, and their distribution is symmetric. The calculation of the t-test proceeds as follows.

$$t_{n-1} = \frac{\bar{X} - M0}{s_{\bar{X}}/\sqrt{n}}$$

where

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n},$$

$$s_{\bar{X}} = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}},$$

and $M0$ is the value of the mean hypothesized by the null hypothesis.

The significance of the test statistic is determined by computing the p-value. If this p-value is less than a specified level (often 0.05), the null hypothesis is rejected. Otherwise, no conclusion can be reached.

Wilcoxon Signed-Rank Test

The Wilcoxon signed-rank test is a popular, nonparametric substitute for the *t*-test. It assumes that the data follow a symmetric distribution. The test is computed using the following steps.

1. Subtract the hypothesized difference, *M0*, from each data value. Rank the values according to their absolute values.

2. Compute the sum of the positive ranks *Sp* and the sum of the negative ranks *Sn*. The test statistic, W_R, is the minimum of *Sp* and *Sn*.

3. Compute the mean and standard deviation of W_R using the formulas

$$\mu_{W_R} = \frac{n(n+1)}{4}$$

$$\sigma_{W_R} = \sqrt{\frac{n(n+1)(2n+1)}{24} - \frac{\sum t^3 - \sum t}{48}}$$

where t represents the number of times the i^{th} value occurs.

4. Compute the z-value using

$$z_W = \frac{W_R - \mu_{W_R}}{\sigma_{W_R}}$$

For cases when n is less than 38, the significance level is found from a table of exact probabilities for the Wilcoxon test. When n is greater than or equal to 38, the significance of the test statistic is determined by comparing the z value to a normal probability table. If this p-value is less than a specified level (often 0.05), the null hypothesis is rejected. Otherwise, no conclusion can be reached.

Sign Test

The sign test is popular because it is simple to compute. This test assumes that the data all follow the same distribution. The test is computed using the following steps.

1. Count the number of values strictly greater than $M0$. Call this value X.
2. Count the number of values strictly less than $M0$. Call this value Y.
3. Set $m = X + Y$.
4. Under the null hypothesis, X is distributed as a binomial random variable with a proportion of 0.5 and sample size of m.

The significance of X is calculated using binomial probabilities.

Bootstrap Test

The one-sample bootstrap procedure for testing whether the mean is equal to a specific value is given in Efron & Tibshirani (1993), pages 224-227. The bootstrap procedure is as follows.

1. Compute the mean of the sample. Call it $\bar{X}$.
2. Compute the t-value using the standard t-test. The formula for this computation is

$$t_X = \frac{\bar{X} - M0}{s_X/\sqrt{n}}$$

where $M0$ is the hypothesized mean.

3. Draw a random, with-replacement sample of size n from the original X values. Call this sample $Y_1, Y_2, \cdots, Y_n$.

4. Compute the t-value of this bootstrap sample using the formula

$$t_Y = \frac{\bar{Y} - \bar{X}}{s_Y/\sqrt{n}}$$

5. For a two-tailed test, if $|t_Y| > |t_X|$ then add one to a counter variable, *A*.
6. Repeat steps 3 – 5 *B* times. *B* may be anywhere from 100 to 10,000.
7. Compute the *p*-value of the bootstrap test as (*A* + 1) / (*B* + 1)
8. Steps 1 – 7 complete one simulation iteration. Repeat these steps *M* times, where *M* is the number of simulations. The power and significance level are equal to the percent of the time the *p*-value is less than the nominal alpha of the test in their respective simulations.

Note that the bootstrap test is a time-consuming test to analyze, especially if you set *B* to a value much larger than 100.

Exponential Test

The exponential distribution is a highly skewed distribution, so it is very different from the normal distribution. Thus, the t-test does not work well with exponential data.

There is an exact test for the mean of a sample drawn from the exponential distribution. It is well known that a simple function of the mean of exponential data follows the chi-square distribution. This relationship is given in Epstein (1960) as

$$\frac{2n\bar{X}}{M0} \sim \chi^2_{2n}$$

This expression can be used to test hypotheses about the value of the mean, *M*0.

Standard Deviations

Care must be used when either the null or alternative distribution is not normal. In these cases, the standard deviation is usually not specified directly. For example, you might use a gamma distribution with a shape parameter of 1.5 and a mean of 4 as the null distribution and a gamma distribution with the same shape parameter and a mean of 5 as the alternative distribution. This allows you to compare the two means. However, note that although the shape parameters are constant, the standard deviations are not. In cases such as this, the null and alternatives not only have different means, but different standard deviations.

Example 1 – Power at Various Sample Sizes

A researcher is planning an experiment to test whether the mean response level to a certain drug is significantly different from zero. The researcher wants to use a t-test with an alpha level of 0.05. He wants to compute the power at various sample sizes from 10 to 40, assuming the true mean is one. He assumes that the data are normally distributed with a standard deviation of 2.

For reproducibility, we'll use a random seed of 60677.

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 1a** 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**
Alternative Hypothesis	**Two-Sided (H1: μ ≠ μ0)**
Test Type	**T-Test**
Simulations	**2000**
Random Seed	**60677** (for Reproducibility)
Alpha	**0.05**
N (Sample Size)	**10 to 40 by 5**
Input Type	**Simple**
Distribution to Simulate	**Normal**
μ0 (Null or Baseline Mean)	**0**
μ1 (Actual Mean)	**1**
σ (Standard Deviation)	**2**

Output

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

Numeric Reports

Numeric Results

Solve For: Power
Hypotheses: H0: μ = μ0 vs. H1: μ ≠ μ0
Test Statistic: T-Test
Simulated Distribution: Normal

Power	Sample Size N	Mean Null μ0	Mean Actual μ1	Mean Difference μ1 - μ0	Standard Deviation σ	Effect Size |μ1 - μ0| / σ	Alpha Target	Alpha Actual
0.3010	10	0	1	1	2	0.5	0.05	0.049
(0.0201) [0.2809 0.3211]								(0.009) [0.039 0.058]
0.4190	15	0	1	1	2	0.5	0.05	0.051
(0.0216) [0.3974 0.4406]								(0.01) [0.041 0.06]
0.5600	20	0	1	1	2	0.5	0.05	0.053
(0.0218) [0.5382 0.5818]								(0.01) [0.043 0.063]
0.6660	25	0	1	1	2	0.5	0.05	0.057
(0.0207) [0.6453 0.6867]								(0.01) [0.046 0.067]
0.7550	30	0	1	1	2	0.5	0.05	0.050
(0.0188) [0.7362 0.7738]								(0.01) [0.04 0.06]
0.8185	35	0	1	1	2	0.5	0.05	0.055
(0.0169) [0.8016 0.8354]								(0.01) [0.045 0.065]
0.8570	40	0	1	1	2	0.5	0.05	0.054
(0.0153) [0.8417 0.8723]								(0.01) [0.044 0.063]

Simulations: 2000. Run Time: 2.40 seconds.
User-Entered Random Seed: 60677

Power	The probability of rejecting a false null hypothesis when the alternative hypothesis is true. The second row provides the precision and a 95% confidence interval for Power, (Power Precision) [95% LCL and UCL], based on the size of the simulation.
N	The size of the sample drawn from the population.
μ0	The value of the population mean under the null hypothesis.
μ1	The actual value of the population mean at which power and sample size are calculated.
μ1 - μ0	The difference between the actual and null means.
σ	The standard deviation of the population. It measures the variability in the population.
|μ1 - μ0| / σ	The Effect Size, which is the relative magnitude of the effect.
Target Alpha	The probability of rejecting a true null hypothesis. It is set by the user.
Actual Alpha	The alpha level that was actually achieved by the experiment. The second row provides the precision and a 95% confidence interval for Alpha, (Alpha Precision) [95% LCL and UCL], based on the size of the simulation.

Summary Statements

A single-group design will be used to test whether the mean (μ) is different from 0 (H0: μ = 0 versus H1: μ ≠ 0). The comparison will be made using a two-sided, one-sample t-test, with a Type I error rate (α) of 0.05. The standard deviation is assumed to be 2. To detect a mean of 1 with a sample size of 10 subjects, the power is 0.301. These results are based on 2000 simulations (Monte Carlo samples) from the Normal distribution.

Dropout-Inflated Sample Size

Dropout Rate	Sample Size N	Dropout-Inflated Enrollment Sample Size N'	Expected Number of Dropouts D
20%	10	13	3
20%	15	19	4
20%	20	25	5
20%	25	32	7
20%	30	38	8
20%	35	44	9
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 Lokhnygina, Y. (2018) pages 32-33.)
D The expected number of dropouts. D = N' - N.

Dropout Summary Statements

Anticipating a 20% dropout rate, 13 subjects should be enrolled to obtain a final sample size of 10 subjects.

References

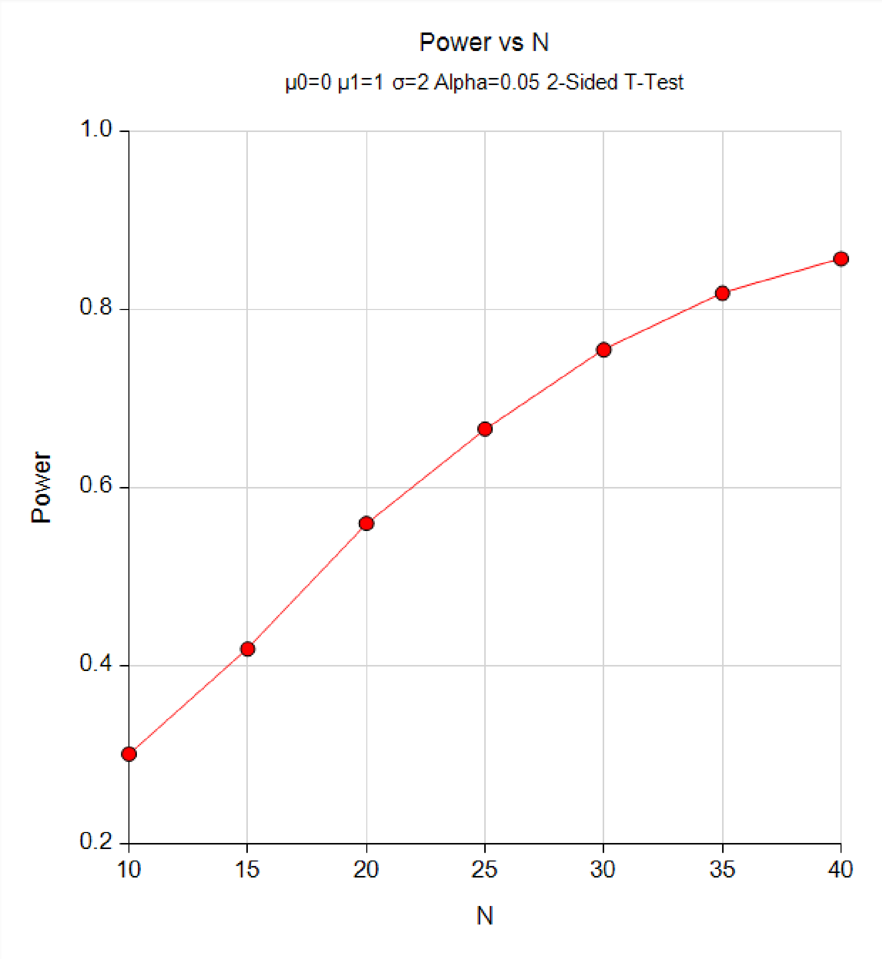
Chow, S.C., Shao, J., Wang, H., and Lokhnygina, Y. 2018. Sample Size Calculations in Clinical Research, Third Edition. Taylor & Francis/CRC. Boca Raton, Florida.
Devroye, Luc. 1986. Non-Uniform Random Variate Generation. Springer-Verlag. New York.
Machin, D., Campbell, M., Fayers, P., and Pinol, A. 1997. Sample Size Tables for Clinical Studies, 2nd Edition. Blackwell Science. Malden, MA.

This report shows the estimated power for each scenario. The first row shows the parameter settings and the estimated power and significance level (Actual Alpha). Note that because these are results of a simulation study, the computed power and alpha will vary from run to run if you use a random seed. Thus, another report obtained using the same input parameters and a random seed will be slightly different than the one above.

The second row shows two 95% confidence intervals in brackets: the first for the power and the second for the significance level. Half the width of each confidence interval is given in parentheses as a fundamental measure of the accuracy of the simulation. As the number of simulations is increased, the width of the confidence interval will decrease.

Plots Section

Plots



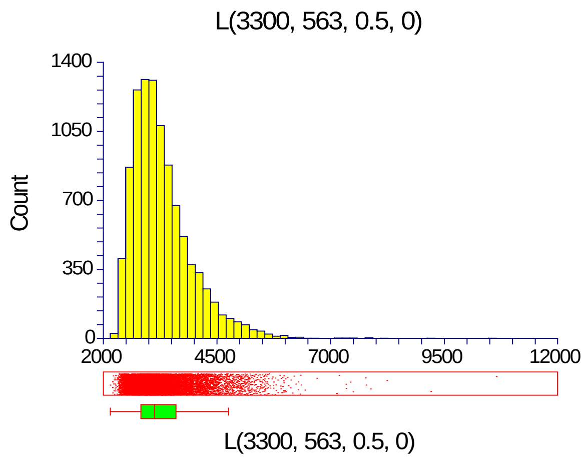
This plot shows the relationship between sample size and power.

Example 2 – Finding the Sample Size for Skewed Data

In studying deaths from SIDS (Sudden Infant Death Syndrome), one hypothesis put forward is that infants dying of SIDS weigh less than normal at birth. Suppose the average birth weight of infants is 3300 grams with a standard deviation of 663 grams. The researchers decide to examine the effect of a skewed distribution on the test used by adding skewness to the simulated data using Tukey's Lambda distribution with a skewness factor of 0.5.

Using the Data Simulator program, the researchers found that the actual standard deviation using the above parameters was almost 800. This occurs because adding skewness changes the standard deviation. They found that setting the standard deviation in Tukey's Lambda distribution to 563 resulted in a standard deviation in the data of about 663.

A histogram of 10,000 pseudo-random values from this distribution appears as follows.



The researchers want to determine how large a sample of SIDS infants will be needed to detect a drop in average weight of 25%? Note that applying this percentage to the average weight of 3300 yields 2475. Use an alpha of 0.05 and 80% power.

Although a one-sided hypothesis might be considered, sample size estimates will assume a two-sided alternative to keep the research design in line with other studies.

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 2a** 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 .... **Sample Size**
Alternative Hypothesis .... **Two-Sided (H1: μ ≠ μ0)**
Test Type .... **T-Test**
Simulations **2000**
Random Seed .... **4513815** (for Reproducibility)
Power **0.80**
Alpha **0.05**
Input Type .... **Simple**
Distribution to Simulate .... **TukeyGH**
G (Skewness) **0.5**
H (Kurtosis) **0**
μ0 (Null or Baseline Mean) **2475**
μ1 (Actual Mean) **3300**
σ (Standard Deviation) **563**

Output

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

Numeric Results of Search for N

Numeric Results

Solve For: Sample Size
Hypotheses: H0: μ = μ0 vs. H1: μ ≠ μ0
Test Statistic: T-Test
Simulated Distribution: TukeyGH

		Mean					Alpha			
Power	Sample Size N	Null μ0	Actual μ1	Difference μ1 - μ0	Standard Deviation σ	Effect Size |μ1 - μ0| / σ	Target	Actual	G	H
0.851	5	2475	3300	825	563	1.465	0.05	0.08	0.5	0
(0.0156) [0.8354 0.8666]								(0.012) [0.068 0.091]		

Simulations: 2000. Run Time: 2.31 seconds.
User-Entered Random Seed: 4513815

The required sample size was 5. Note that the actual alpha value is between 0.068 and 0.091, which is greater than 0.05. This shows one of the problems of using the t-test with a skewed distribution.

To be more accurate and yet avoid the long running time of the search for N, a reasonable strategy would be to run simulations to obtain the powers using N's from 4 to 10. The result of this study is displayed next (**Example 2c**).

Numeric Results of Power Search for Various N

Numeric Results

Solve For: Power
Hypotheses: H0: μ = μ0 vs. H1: μ ≠ μ0
Test Statistic: T-Test
Simulated Distribution: TukeyGH

	Sample	Mean			Standard	Effect	Alpha			
	Size	Null	Actual	Difference	Deviation	Size				
Power	N	μ0	μ1	μ1 - μ0	σ	|μ1 - μ0| / σ	Target	Actual	G	H
0.6130	4	2475	3300	825	563	1.465	0.05	0.081	0.5	0
(0.0213) [0.5917 0.6343]								(0.012) [0.069 0.093]		
0.8240	5	2475	3300	825	563	1.465	0.05	0.081	0.5	0
(0.0167) [0.8073 0.8407]								(0.012) [0.069 0.093]		
0.9465	6	2475	3300	825	563	1.465	0.05	0.092	0.5	0
(0.0099) [0.9366 0.9564]								(0.013) [0.079 0.105]		
0.9830	7	2475	3300	825	563	1.465	0.05	0.081	0.5	0
(0.0057) [0.9773 0.9887]								(0.012) [0.069 0.093]		
0.9925	8	2475	3300	825	563	1.465	0.05	0.084	0.5	0
(0.0038) [0.9887 0.9963]								(0.012) [0.072 0.096]		
0.9960	9	2475	3300	825	563	1.465	0.05	0.082	0.5	0
(0.0028) [0.9932 0.9988]								(0.012) [0.07 0.094]		
0.9995	10	2475	3300	825	563	1.465	0.05	0.081	0.5	0
(0.001) [0.9985 1]								(0.012) [0.069 0.092]		

Simulations: 2000. Run Time: 1.72 seconds.
User-Entered Random Seed: 4538256

The sample sizes of 6 to 10 appear to meet the design parameters the best, but the actual significance level still appears to be greater than 0.05. The researchers decide that they must use a smaller value of Alpha so that the actual alpha is about 0.05. After some experimentation, they find that setting Alpha to 0.025 results in the desired power and significance level (**Example 2d**).

Numeric Results with Alpha = 0.025

Numeric Results

Solve For: Power
Hypotheses: H0: μ = μ0 vs. H1: μ ≠ μ0
Test Statistic: T-Test
Simulated Distribution: TukeyGH

		Mean					Alpha			
Power	Sample Size N	Null μ0	Actual μ1	Difference μ1 - μ0	Standard Deviation σ	Effect Size |μ1 - μ0| / σ	Target	Actual	G	H
0.3495 (0.0209) [0.3286 0.3704]	4	2475	3300	825	563	1.465	0.025	0.041 (0.009) [0.032 0.049]	0.5	0
0.6120 (0.0214) [0.5906 0.6334]	5	2475	3300	825	563	1.465	0.025	0.054 (0.01) [0.044 0.064]	0.5	0
0.8190 (0.0169) [0.8021 0.8359]	6	2475	3300	825	563	1.465	0.025	0.057 (0.01) [0.046 0.067]	0.5	0
0.9130 (0.0124) [0.9006 0.9254]	7	2475	3300	825	563	1.465	0.025	0.059 (0.01) [0.049 0.069]	0.5	0
0.9635 (0.0082) [0.9553 0.9717]	8	2475	3300	825	563	1.465	0.025	0.052 (0.01) [0.042 0.062]	0.5	0
0.9855 (0.0052) [0.9803 0.9907]	9	2475	3300	825	563	1.465	0.025	0.056 (0.01) [0.046 0.066]	0.5	0
0.9920 (0.0039) [0.9881 0.9959]	10	2475	3300	825	563	1.465	0.025	0.050 (0.01) [0.04 0.059]	0.5	0

Simulations: 2000. Run Time: 1.89 seconds.
User-Entered Random Seed: 4538256

It appears that a sample size of 6 with a Target Alpha of 0.025 will result in an experimental design with the characteristics the researchers wanted.

Notice that when working with non-normal distributions, you must change both N and the Target Alpha to achieve the design you want.

Example 3 – Comparative Results with Skewed Data

Continuing with Example 2, the researchers want to study the characteristics of various test statistics as the amount of skewness is increased. To do this, they let the skewness parameter of Tukey's Lambda distribution vary between 0 and 1. The researchers realize that the standard deviation will change as the skewness parameter is increased, but they decide to ignore this complication.

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 3** 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**
Alternative Hypothesis .... **Two-Sided (H1: μ ≠ μ0)**
Test Type .... **T-Test**
Simulations **2000**
Random Seed .... **4572272** (for Reproducibility)
Alpha **0.05**
N (Sample Size) **6**
Input Type .... **Simple**
Distribution to Simulate .... **TukeyGH**
G (Skewness) **0.0 0.2 0.4 0.6 0.8 1.0**
H (Kurtosis) **0**
μ0 (Null or Baseline Mean) **2475**
μ1 (Actual Mean) **3300**
σ (Standard Deviation) **563**

Reports Tab

Show Comparative Reports .... **Checked**
Include T-Test Results .... **Checked**
Include Wilcoxon & Sign Test .... **Checked**

Comparative Plots Tab

Show Comparative Plots .... **Checked**

Output

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

Power Comparison

Hypotheses: H0: μ = μ0 vs. H1: μ ≠ μ0
Simulated Distribution: TukeyGH

Sample Size N	Mean Null μ0	Mean Actual μ1	Standard Deviation σ	Target Alpha	Power T-Test	Power Wilcoxon	Power Sign	G	H
6	2475	3300	563	0.05	0.8130	0.6375	0.6375	0.0	0
6	2475	3300	563	0.05	0.8605	0.7275	0.7275	0.2	0
6	2475	3300	563	0.05	0.9195	0.8980	0.8980	0.4	0
6	2475	3300	563	0.05	0.9470	1.0000	1.0000	0.6	0
6	2475	3300	563	0.05	0.9660	1.0000	1.0000	0.8	0
6	2475	3300	563	0.05	0.9605	1.0000	1.0000	1.0	0

Simulations: 2000. Run Time: 2.14 seconds.
User-Entered Random Seed: 4572272

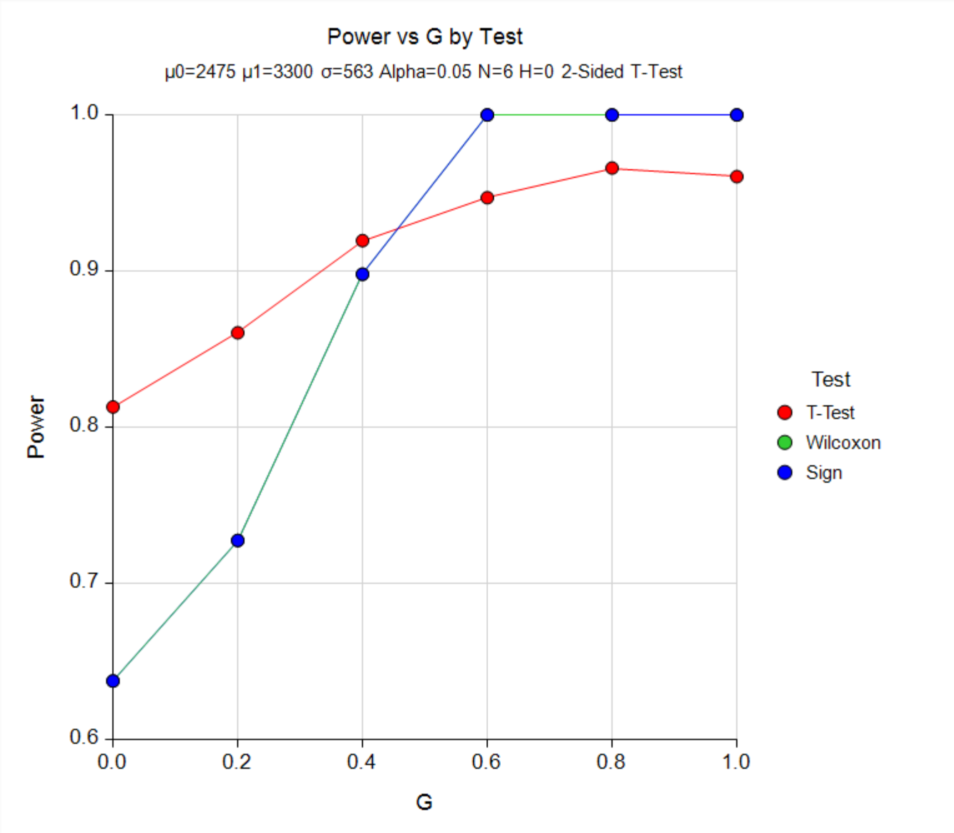
Alpha Comparison

Hypotheses: H0: μ = μ0 vs. H1: μ ≠ μ0
Simulated Distribution: TukeyGH

Sample Size N	Mean Null μ0	Mean Actual μ1	Standard Deviation σ	Alpha Target	Alpha T-Test	Alpha Wilcoxon	Alpha Sign	G	H
6	2475	3300	563	0.05	0.048	0.027	0.027	0.0	0
6	2475	3300	563	0.05	0.058	0.039	0.039	0.2	0
6	2475	3300	563	0.05	0.074	0.044	0.044	0.4	0
6	2475	3300	563	0.05	0.113	0.065	0.065	0.6	0
6	2475	3300	563	0.05	0.124	0.076	0.076	0.8	0
6	2475	3300	563	0.05	0.170	0.109	0.109	1.0	0

Simulations: 2000. Run Time: 2.14 seconds.
User-Entered Random Seed: 4572272

Comparative Plots



Several interesting trends become apparent from this study. First, for a sample size of 6, the power of the Wilcoxon test and the sign test are the same (this is not the case for larger sample sizes). The alpha of the t-test increases as the amount of skewness increases. The alpha of the Wilcoxon and sign tests does not increase as rapidly as it does for the t-test.

Example 4 – Validation using Zar (1984)

Zar (1984), pages 111-112, presents an example in which μ0 = 0.0, μ1 = 1.0, σ = 1.25, alpha = 0.05, and N = 12. Zar obtains an approximate power of 0.72. We will validate this procedure by running this example. To make certain that the results are very accurate, the number of simulations will be set to 10,000.

For reproducibility, we'll use a random seed of 6015683.

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 4** 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**
Alternative Hypothesis .... **Two-Sided (H1: μ ≠ μ0)**
Test Type.. **T-Test**
Simulations **10000**
Random Seed... **6015683** (for Reproducibility)
Alpha.. **0.05**
N (Sample Size).. **12**
Input Type... **Simple**
Distribution to Simulate .... **Normal**
μ0 (Null or Baseline Mean) **0**
μ1 (Actual Mean) **1**
σ (Standard Deviation).................................. **1.25**

Output

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

Numeric Results

Solve For: Power
Hypotheses: H0: μ = μ0 vs. H1: μ ≠ μ0
Test Statistic: T-Test
Simulated Distribution: Normal

		Mean					Alpha	
Power	Sample Size N	Null μ0	Actual μ1	Difference μ1 - μ0	Standard Deviation σ	Effect Size |μ1 - μ0| / σ	Target	Actual
0.7162 (0.0088) [0.7074 0.725]	12	0	1	1	1.3	0.8	0.05	0.049 (0.004) [0.045 0.053]

Simulations: 10000. Run Time: 1.48 seconds.
User-Entered Random Seed: 6015683

This simulation obtained a power of 0.7162 which rounds to the 0.72 computed by Zar.

Example 5 – Validation using Machin (1997)

Machin, et. al. (1997), page 37, present an example in which μ0 = 0.0, μ1 = 0.2, σ = 1.0, alpha = 0.05, and power = 0.80. They obtain a sample size of 199.

For reproducibility, we'll use a random seed of 6030438.

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 5** 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 ....**Sample Size**
Alternative Hypothesis ....**Two-Sided (H1: μ ≠ μ0)**
Test Type...**T-Test**
Simulations **2000**
Random Seed..**6030438** (for Reproducibility)
Power..**0.80**
Alpha...**0.05**
Input Type..**Simple**
Distribution to Simulate ....**Normal**
μ0 (Null or Baseline Mean) **0**
μ1 (Actual Mean) **0.2**
σ (Standard Deviation)..................................**1**

Output

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

Numeric Results

Solve For: Sample Size
Hypotheses: H0: μ = μ0 vs. H1: μ ≠ μ0
Test Statistic: T-Test
Simulated Distribution: Normal

		Mean					Alpha	
Power	**Sample Size N**	**Null μ0**	**Actual μ1**	**Difference μ1 - μ0**	**Standard Deviation σ**	**Effect Size |μ1 - μ0| / σ**	**Target**	**Actual**
0.815	199	0	0.2	0.2	1	0.2	0.05	0.046
(0.017) [0.798 0.832]								(0.009) [0.036 0.055]

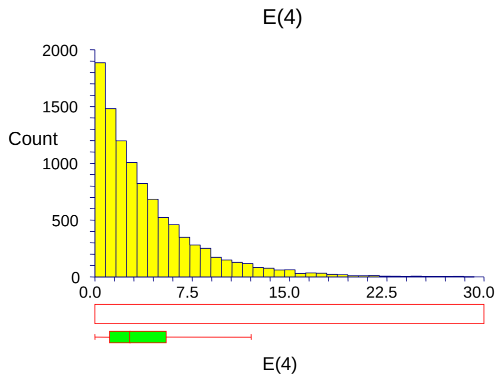
Simulations: 2000. Run Time: 13.10 seconds.
User-Entered Random Seed: 6030438

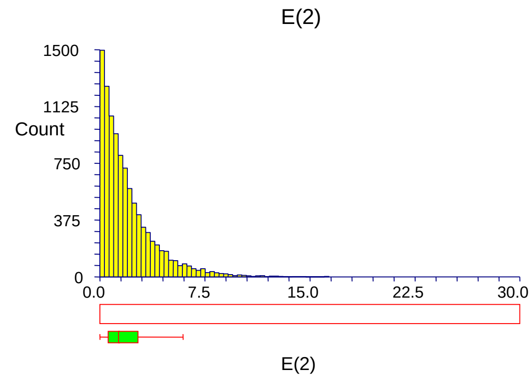
The sample size result matches the value of Machin (1997). If you run the simulation multiple times, you'll come up with values right around 199.

Example 6 – Power of the Wilcoxon Signed-Rank Test

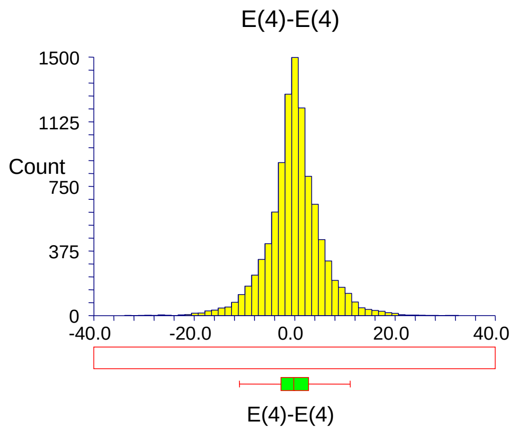
The Wilcoxon nonparametric signed-rank test was designed for data that do not follow the normal distribution but are symmetric. This type of data often occurs when differences between two non-normal variables are taken, as in a study that analyzes differences in pre- and post-test scores.

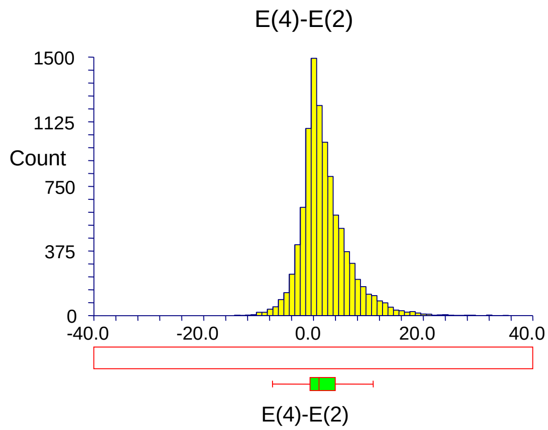
For this example, suppose the pre-test and the post-test scores are exponentially distributed. Here are examples of exponentially distributed data with means of 4 and 2, respectively.





It has been shown that the differences between two identically distributed variables are symmetric. The histogram below on the left shows differences in the null case in which the difference is between two exponential variables both with a mean of 4. The histogram below on the right shows differences in the alternative case in which the difference is between an exponential variable with a mean of 4 and an exponential variable with a mean of 2. Careful inspection shows that the second histogram is skewed to the right and the mean difference is about 2, not 0.





The researchers want to study the power of the two-sided Wilcoxon signed-rank test when sample sizes of 10, 20, 30, and 40 are used, and testing is done at the 5% significance level.

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 6** 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**
Alternative Hypothesis	**Two-Sided (H1: μ ≠ μ0)**
Test Type	**Wilcoxon Signed-Rank Test**
Simulations	**5000**
Random Seed	**8464844** (for Reproducibility)
Alpha	**0.05**
N (Sample Size)	**10 20 30 40 50**
Input Type	**General**
Distribution|H0	**Exponential(M0)-Exponential(M0)**
Distribution|H1	**Exponential(M0)-Exponential(M1)**
M0 (Mean|H0) Parameter Value(s)	**4**
M1 (Mean|H1) Parameter Value(s)	**2**

Output

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

Numeric Results

Solve For: Power
Hypotheses: H0: μ = μ0 vs. H1: μ ≠ μ0
Test Statistic: Wilcoxon Signed-Rank Test
H0 Distribution: Exponential(M0)-Exponential(M0)
H1 Distribution: Exponential(M0)-Exponential(M1)

	Sample	Mean			Alpha		Distribution Parameters	
Power	Size N	Null μ0	Actual μ1	Difference μ1 - μ0	Target	Actual	M0	M1
0.2386	10	0	2	2	0.05	0.048	4	2
(0.0118) [0.2268 0.2504]						(0.006) [0.042 0.054]		
0.4710	20	0	2	2	0.05	0.045	4	2
(0.0138) [0.4572 0.4848]						(0.006) [0.039 0.051]		
0.6620	30	0	2	2	0.05	0.049	4	2
(0.0131) [0.6489 0.6751]						(0.006) [0.043 0.055]		
0.7810	40	0	2	2	0.05	0.045	4	2
(0.0115) [0.7695 0.7925]						(0.006) [0.039 0.051]		
0.8590	50	0	2	2	0.05	0.050	4	2
(0.0096) [0.8494 0.8686]						(0.006) [0.044 0.056]		

Simulations: 5000. Run Time: 6.47 seconds.
User-Entered Random Seed: 8464844

Reasonable power is achieved with N = 50.

Example 7 – Likert-Scale Data

Likert-scale data occurs commonly in survey research. A *Likert Scale* is discrete, ordinal data. It usually occurs when a survey poses a question and the respondent must pick among strongly agree, agree, undecided, disagree, or strongly disagree. The responses are usually coded as 1, 2, 3, 4, and 5.

Likert data can be analyzed in a number of ways. Perhaps the most common is to use a t-test or a Wilcoxon test. (Using the Wilcoxon test is invalid in this case because the data are seldom distributed symmetrically.)

In this example, a questionnaire is planned on which Likert-scale questions will be asked. The researchers want to study the power and actual significance levels of various sample sizes. They decide to look at what happens as the proportion of strongly agree responses is increased beyond a perfectly uniform response pattern. They want to compute the power when the strongly agree response is twice as likely, four times as likely, and eight times as likely. The sample size is 20, alpha is 0.05, and the test is two-sided.

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 7** 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**
Alternative Hypothesis .... **Two-Sided (H1: μ ≠ μ0)**
Test Type.. **T-Test**
Simulations **5000**
Random Seed.. **5487977** (for Reproducibility)
Alpha... **0.05**
N (Sample Size)... **20**
Input Type... **General**
Distribution|H0 .... **Multinomial(M0 1 1 1 1)**
Distribution|H1 .... **Multinomial(M1 1 1 1 1)**
M0 (Mean|H0) Parameter Value(s)................ **1**
M1 (Mean|H1) Parameter Value(s)................ **2 4 8**

Output

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

Numeric Results

Solve For: Power
Hypotheses: H0: μ = μ0 vs. H1: μ ≠ μ0
Test Statistic: T-Test
H0 Distribution: Multinomial(M0 1 1 1 1)
H1 Distribution: Multinomial(M1 1 1 1 1)

		Mean			Alpha		Distribution Parameters	
Power	Sample Size N	Null μ0	Actual μ1	Difference μ1 - μ0	Target	Actual	M0	M1
0.1812 (0.0107) [0.1705 0.1919]	20	3	2.7	-0.3	0.05	0.054 (0.006) [0.048 0.06]	1	2
0.5744 (0.0137) [0.5607 0.5881]	20	3	2.3	-0.8	0.05	0.052 (0.006) [0.046 0.059]	1	4
0.9102 (0.0079) [0.9023 0.9181]	20	3	1.8	-1.2	0.05	0.055 (0.006) [0.049 0.061]	1	8

Simulations: 5000. Run Time: 2.43 seconds.
User-Entered Random Seed: 5487977

Note that *M0* and *M1* are no longer the *H0* and *H1* means. Now, they represent the relative weighting given to the strongly agree response. Under *H0*, the mean (μ0) is 3.0. As *M1* is increased, the mean under *H1* (μ1) changes from 2.7 to 2.3 to 1.8. We note that the actual significance level, alpha, remains close to the target value of 0.05.

Example 8 – Computing the Power after Completing an Experiment

A group of researchers has completed an experiment designed to determine if a particular hormone increases weight gain in rats. The researchers inject 20 rats of the same age with the hormone and measure their weight gain after 1 month. The investigators use the two-sided bootstrap test with alpha = 0.05 and 100 bootstrap samples to determine if the average weight gained by these rats (171 grams) is significantly greater than the known average weight gained by rats of the same age over the same period of time (155 grams). Unfortunately, the results indicate that there is no significant difference between the two means. Therefore, the researchers decide to compute the power achieved by this test for alternative means ranging from 160 to 190 grams. They decide to use 1000 simulations for the study. For comparative purposes, they also decide to look at the power achieved by the bootstrap test in comparison to various other applicable tests. Suppose that they know that the standard deviation for weight gain is 33 grams.

Note that the researchers compute the power for a range of practically significant alternatives. The range chosen should represent likely values based on historical evidence.

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 8** 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**
Alternative Hypothesis .... **Two-Sided (H1: μ ≠ μ0)**
Test Type.. **Bootstrap Test**
Bootstrap Iterations....................................... **100**
Simulations **2000**
Random Seed... **2255448** (for Reproducibility)
Alpha... **0.05**
N (Sample Size).. **20**
Input Type... **Simple**
Distribution to Simulate .... **Normal**
μ0 (Null or Baseline Mean) **155**
μ1 (Actual Mean) .... **160 to 190 by 10**
σ (Standard Deviation).................................. **33**

Reports Tab

Show Comparative Reports .... **Checked**
Include T-Test Results.................................. **Checked**
Include Wilcoxon & Sign Test Results .... **Checked**
Include Bootstrap Test Results .... **Checked**

Comparative Plots Tab

Show Comparative Plots............................... **Checked**

Output

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

Numeric Results for Power of Bootstrap

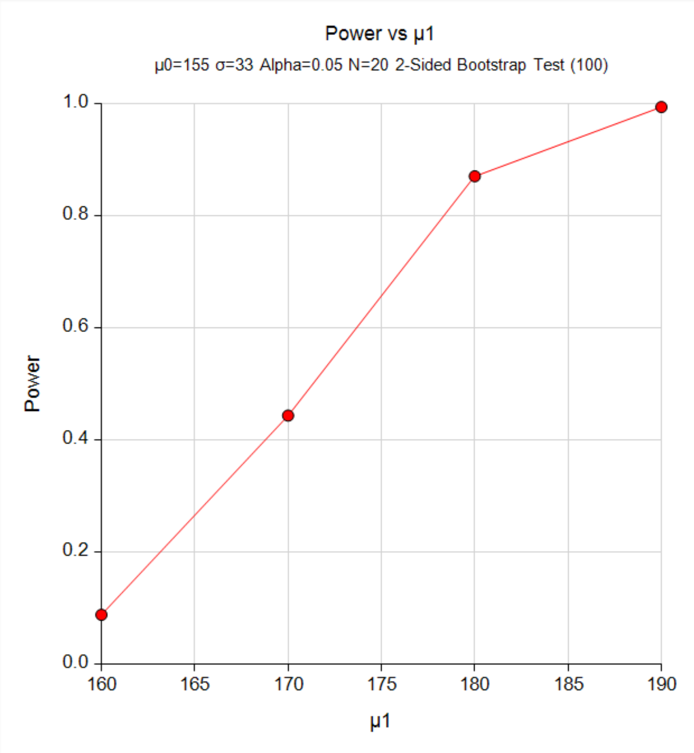
Numeric Results

Solve For: Power
Hypotheses: H0: μ = μ0 vs. H1: μ ≠ μ0
Test Statistic: Bootstrap Test (100)
Simulated Distribution: Normal

Power	Sample Size N	Mean Null μ0	Mean Actual μ1	Mean Difference μ1 - μ0	Standard Deviation σ	Effect Size |μ1 - μ0| / σ	Alpha Target	Alpha Actual
0.0880 (0.0124) [0.0756 0.1004]	20	155	160	5	33	0.152	0.05	0.049 (0.009) [0.04 0.058]
0.4435 (0.0218) [0.4217 0.4653]	20	155	170	15	33	0.455	0.05	0.051 (0.01) [0.041 0.061]
0.8705 (0.0147) [0.8558 0.8852]	20	155	180	25	33	0.758	0.05	0.046 (0.009) [0.037 0.055]
0.9940 (0.0034) [0.9906 0.9974]	20	155	190	35	33	1.061	0.05	0.046 (0.009) [0.036 0.055]

Simulations: 2000. Run Time: 23.80 seconds.
User-Entered Random Seed: 2255448

Plots



Reasonable power is achieved by this test for alternative means larger than 180. The accuracy of these results, of course, depends on the assumption that the data are normally distributed.

Comparative Results for Power of Various Tests

Power Comparison

Hypotheses: H0: μ = μ0 vs. H1: μ ≠ μ0
Simulated Distribution: Normal

Sample Size N	Mean Null μ0	Mean Actual μ1	Standard Deviation σ	Target Alpha	Power T-Test	Power Wilcoxon	Power Sign	Power Bootstrap
20	155	160	33	0.05	0.0945	0.0925	0.0720	0.0880
20	155	170	33	0.05	0.4715	0.4465	0.3120	0.4435
20	155	180	33	0.05	0.8910	0.8735	0.7235	0.8705
20	155	190	33	0.05	0.9965	0.9945	0.9425	0.9940

Simulations: 2000. Run Time: 23.80 seconds.
User-Entered Random Seed: 2255448

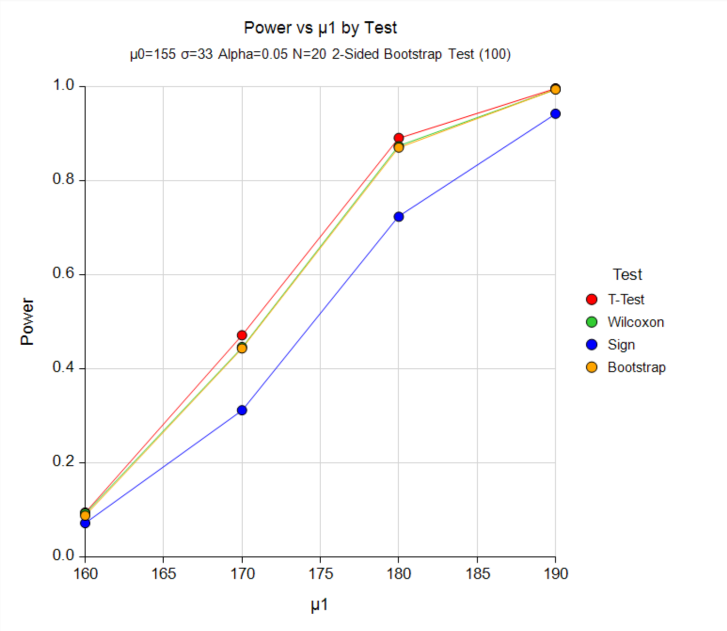
Alpha Comparison

Hypotheses: H0: μ = μ0 vs. H1: μ ≠ μ0
Simulated Distribution: Normal

Sample Size N	Mean Null μ0	Mean Actual μ1	Standard Deviation σ	Alpha Target	Alpha T-Test	Alpha Wilcoxon	Alpha Sign	Alpha Bootstrap
20	155	160	33	0.05	0.055	0.055	0.042	0.049
20	155	170	33	0.05	0.052	0.049	0.043	0.051
20	155	180	33	0.05	0.050	0.050	0.040	0.046
20	155	190	33	0.05	0.048	0.050	0.038	0.046

Simulations: 2000. Run Time: 23.80 seconds.
User-Entered Random Seed: 2255448

Comparative Plots



It is apparent from these results that the bootstrap performs about as well as the t-test and nonparametric tests for this design.

Example 9 – Comparison of Tests for Exponential Data

A researcher is designing an experiment. She believes that the data will follow an exponential distribution. Consequently, she does not believe that the t-test will be useful for her situation. She would like to compare several possible tests to determine which would be best for analyzing exponential data. She is interested in determining the power when the alternative mean is twice the null mean, which is 10. She wants to find the power achieved for sample sizes ranging from 20 to 60 with alpha = 0.05.

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 9** 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**
Alternative Hypothesis .... **Two-Sided (H1: μ ≠ μ0)**
Test Type.. **T-Test**
Simulations **2000**
Random Seed... **1315448** (for Reproducibility)
Alpha... **0.05**
N (Sample Size).. **20 to 60 by 20**
Input Type... **Simple**
Distribution to Simulate .... **Exponential**
μ0 (Null or Baseline Mean) **10**
μ1 (Actual Mean) **20**

Reports Tab

Show Comparative Reports........................... **Checked**
Include T-Test Results.................................. **Checked**
Include Wilcoxon & Sign Test Results .... **Checked**
Include Bootstrap Test Results...................... **Checked**
Include Exponential Test Results.................. **Checked**

Comparative Plots Tab

Show Comparative Plots............................... **Checked**

Output

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

Power Comparison

Hypotheses: H0: μ = μ0 vs. H1: μ ≠ μ0
Simulated Distribution: Exponential

Sample Size N	Mean Null μ0	Mean Actual μ1	Standard Deviation σ	Target Alpha	Power T-Test	Power Wilcoxon	Power Sign	Power Bootstrap	Power Exponential
20	10	20	20	0.05	0.613	0.4525	0.1415	0.4205	0.8905
40	10	20	20	0.05	0.961	0.7760	0.2335	0.8590	0.9900
60	10	20	20	0.05	0.996	0.9065	0.2815	0.9785	0.9990

Simulations: 2000. Run Time: 36.27 seconds.
User-Entered Random Seed: 1315448

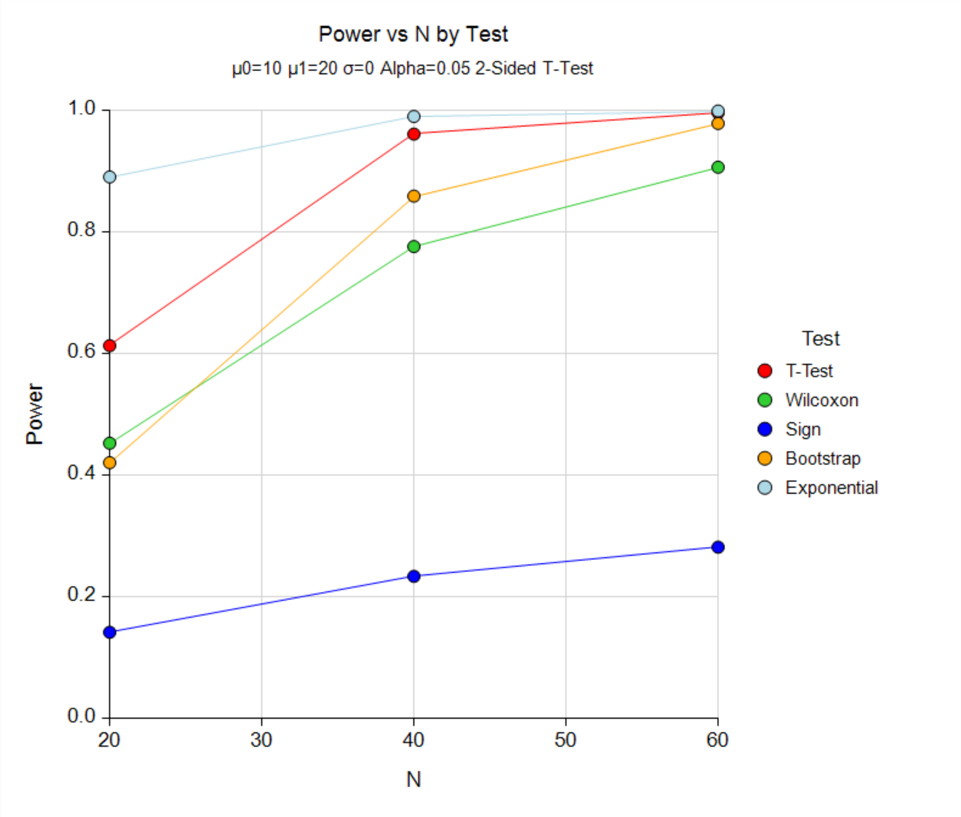
Alpha Comparison

Hypotheses: H0: μ = μ0 vs. H1: μ ≠ μ0
Simulated Distribution: Exponential

Sample Size N	Mean Null μ0	Mean Actual μ1	Standard Deviation σ	Alpha Target	Alpha T-Test	Alpha Wilcoxon	Alpha Sign	Alpha Bootstrap	Alpha Exponential
20	10	20	20	0.05	0.081	0.125	0.198	0.066	0.047
40	10	20	20	0.05	0.066	0.173	0.346	0.052	0.052
60	10	20	20	0.05	0.057	0.243	0.441	0.044	0.049

Simulations: 2000. Run Time: 36.27 seconds.
User-Entered Random Seed: 1315448

Comparative Plots



As would be expected for exponential data, the exponential test performs the best. The bootstrap test performs nearly as well for larger sample sizes. The other tests fail to achieve the target alpha level. Note that these simulation results will vary from run to run because the samples generated are random. The researcher must now decide which test to use based on her level of confidence in the data being truly exponentially distributed and the size of a sample she can afford to take.