

## Chapter 440

# Tests for Two Means (Simulation)

---

## Introduction

This procedure allows you to study the power and sample size of several statistical tests of the null hypothesis that the difference between two means is equal to a specific value versus the alternative hypothesis that it is greater than, less than, or not equal to that value. This allows the user to conduct inequality, non-inferiority, superiority by a margin, and non-zero null tests. Because the mean represents the center of the population, if the means are different, the populations are different. Other attributes of the two populations (such as the shape and spread) might also be compared, but this module focuses on comparisons of the means only.

Measurements are made on individuals that have been randomly assigned to, or randomly chosen from, one of two groups. This is sometimes referred to as a *parallel-groups* design. This design is used in situations such as the comparison of the income level of two regions, the nitrogen content of two lakes, or the effectiveness of two drugs.

The two-sample t-test is commonly used in this situation. When the variances of the two groups are unequal, Welch's unequal-variance t-test is often used. When the data are not normally distributed, the Mann-Whitney U (or Wilcoxon Rank-Sum) test may be used.

The details of the power analysis of the two-sample t-test using analytic techniques are presented in another **PASS** chapter and they won't be duplicated here. This chapter will only consider power analysis using computer simulation.

---

## Technical Details

*Computer simulation* allows us 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. But, in recent years, as computer speeds have increased, simulation studies can be completed on desktop and laptop computers in a reasonable period of time.

The steps to a simulation study are

1. Specify how the test is carried out. This includes indicating how the test statistic is calculated and how the significance level is specified.
2. Generate random samples from the distributions specified by the alternative hypothesis. Calculate the test statistics from the simulated data and determine if the null hypothesis is accepted or rejected. Tabulate the number of rejections and use this to calculate the test's power.
3. Generate random samples from the distributions specified by the null hypothesis. Calculate each test statistic from the simulated data and determine if the null hypothesis is accepted or rejected. Tabulate the number of rejections and use this to calculate the test's significance level.

## Tests for Two Means (Simulation)

- Repeat steps 2 and 3 several thousand times, tabulating the number of times the simulated data leads to a rejection of the null hypothesis. The power is the proportion of simulated samples in step 2 that lead to rejection. The significance level is the proportion of simulated samples in step 3 that lead to rejection.

## Generating Random Distributions

Two methods are available in **PASS** to simulate random samples. The first method generates the random variates directly, one value at a time. The second method generates a large pool (over 10,000) of random values and then draw the random numbers from this pool. This second method can cut the running time of the simulation by 70%.

As mentioned above, the second method begins by generating a large pool of random numbers from the specified distributions. Each of these pools is evaluated to determine if its mean is within a small relative tolerance (0.0001) of the target mean. If the actual mean is not within the tolerance of the target mean, individual members of the population are replaced with new random numbers if the new random number moves the mean towards its target. Only a few hundred such swaps are required to bring the actual mean to within tolerance of the target mean. This population is then sampled with replacement using the uniform distribution. We have found that this method works well as long as the size of the pool is the maximum of twice the number of simulated samples desired and 10,000.

## Test Statistics

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

### Two-Sample T-Test with Equal Variances

The two-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 statistic is as follows

$$t_{df} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s_{\bar{X}_1 - \bar{X}_2}}$$

where

$$\bar{X}_k = \frac{\sum_{i=1}^{N_k} X_{ki}}{N_k}$$

$$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sum_{i=1}^{N_1} (X_{1i} - \bar{X}_1)^2 + \sum_{i=1}^{N_2} (X_{2i} - \bar{X}_2)^2}{N_1 + N_2 - 2} \left( \frac{1}{N_1} + \frac{1}{N_2} \right)}$$

$$df = N_1 + N_2 - 2$$

The significance of the test statistic is determined by computing the p-value based on the t distribution with degrees of freedom  $df$ . If this p-value is less than a specified level (often 0.05), the null hypothesis is rejected. Otherwise, no conclusion can be reached.

## Tests for Two Means (Simulation)

## Welch's Unequal-Variance T-Test

Welch (1938) proposed the following test for use when the two variances cannot be assumed equal.

$$t_f^* = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_{\bar{X}_1 - \bar{X}_2}^*}$$

where

$$S_{\bar{X}_1 - \bar{X}_2}^* = \sqrt{\left( \frac{\sum_{i=1}^{N_1} (X_{1i} - \bar{X}_1)^2}{N_1(N_1 - 1)} \right) + \left( \frac{\sum_{i=1}^{N_2} (X_{2i} - \bar{X}_2)^2}{N_2(N_2 - 1)} \right)}$$

$$f = \frac{\left( \frac{s_1^2}{N_1} + \frac{s_2^2}{N_2} \right)^2}{\frac{s_1^4}{N_1^2(N_1 - 1)} + \frac{s_2^4}{N_2^2(N_2 - 1)}}$$

$$s_1 = \sqrt{\left( \frac{\sum_{i=1}^{N_1} (X_{1i} - \bar{X}_1)^2}{N_1 - 1} \right)}$$

$$s_2 = \sqrt{\left( \frac{\sum_{i=1}^{N_2} (X_{2i} - \bar{X}_2)^2}{N_2 - 1} \right)}$$

## Trimmed T-Test with Equal Variances

The notion of trimming off a small proportion of possibly outlying observations and using the remaining data to form a t-test was first proposed for one sample by Tukey and McLaughlin (1963). Dixon and Tukey (1968) consider a slight modification of this one sample test, called *Winsorization*, which replaces the trimmed data with the nearest remaining value. The two-sample trimmed t-test was proposed by Yuen and Dixon (1973).

Assume that the data values have been sorted from lowest to highest. The *trimmed mean* is defined as

$$\bar{X}_{tg} = \frac{\sum_{k=g+1}^{N-g} X_k}{h}$$

where  $h = N - 2g$  and  $g = [N(G/100)]$ . Here we use  $[Z]$  to mean the largest integer smaller than  $Z$  with the modification that if  $G$  is non-zero, the value of  $[N(G/100)]$  is at least one.  $G$  is the percent trimming and should usually be less than 25%, often between 5% and 10%. Thus, the  $g$  smallest and  $g$  largest observation are omitted in the calculation.

## Tests for Two Means (Simulation)

To calculate the modified t-test, calculate the *Winsorized mean* and the *Winsorized* sum of squared deviations as follows.

$$\bar{X}_{wg} = \frac{g(X_{g+1} + X_{N-g}) + \sum_{k=g+1}^{N-g} X_k}{N}$$

$$SSD_{wg} = \frac{g(X_{g+1} - \bar{X}_{wg})^2 + g(X_{N-g} - \bar{X}_{wg})^2 + \sum_{k=g+1}^{N-g} (X_k - \bar{X}_{wg})^2}{N}$$

Using the above definitions, the two-sample trimmed t-test is given by

$$T_{tg} = \frac{(\bar{X}_{1tg} - \bar{X}_{2tg}) - (\mu_1 - \mu_2)}{\sqrt{\frac{SSD_{1wg} + SSD_{2wg}}{h_1 + h_2 - 2} \left( \frac{1}{h_1} + \frac{1}{h_2} \right)}}$$

The distribution of this  $t$  statistic is approximately that of a  $t$  distribution with degrees of freedom equal to  $h_1 + h_2 - 2$ . This approximation is often reasonably accurate if both sample sizes are greater than 6.

### Trimmed T-Test with Unequal Variances

Yuen (1974) combines trimming (see above) with Welch's (1938) test. The resulting trimmed Welch test is resistant to outliers and seems to alleviate some of the problems that occur because of skewness in the underlying distributions. Extending the results from above, the trimmed version of Welch's t-test is given by

$$T_{tg}^* = \frac{(\bar{X}_{1tg} - \bar{X}_{2tg}) - (\mu_1 - \mu_2)}{\sqrt{\frac{SSD_{1wg}}{h_1(h_1 - 1)} + \frac{SSD_{2wg}}{h_2(h_2 - 1)}}}$$

with degrees of freedom  $f$  given by

$$\frac{1}{f} = \frac{c^2}{h_1 - 1} + \frac{1 - c^2}{h_2 - 1}$$

where

$$c = \frac{\frac{SSD_{1wg}}{h_1(h_1 - 1)}}{\frac{SSD_{1wg}}{h_1(h_1 - 1)} + \frac{SSD_{2wg}}{h_2(h_2 - 1)}}$$

## Tests for Two Means (Simulation)

**Mann-Whitney U or Wilcoxon Rank-Sum Test**

This test is the nonparametric substitute for the equal-variance t-test. Two key assumptions for this test are that the distributions are at least ordinal and that they are identical under  $H_0$ . This implies that ties (repeated values) are not acceptable. When ties are present, the approximation provided can be used, but know that the theoretic results no longer hold.

The Mann-Whitney test statistic is defined as follows in Gibbons (1985).

$$z = \frac{W_1 - \frac{N_1(N_1 + N_2 + 1)}{2} + C}{s_W}$$

where

$$W_1 = \sum_{k=1}^{N_1} \text{Rank}(X_{1k})$$

The ranks are determined after combining the two samples. The standard deviation is calculated as

$$s_W = \sqrt{\frac{N_1 N_2 (N_1 + N_2 + 1)}{12} - \frac{N_1 N_2 \sum_{i=1} (t_i^3 - t_i)}{12(N_1 + N_2)(N_1 + N_2 - 1)}}$$

where  $t_i$  is the number of observations tied at value one,  $t_2$  is the number of observations tied at some value two, and so forth.

The correction factor,  $C$ , is 0.5 if the rest of the numerator of  $z$  is negative or -0.5 otherwise. The value of  $z$  is then compared to the standard normal distribution.

**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, although the shape parameters are constant, the standard deviations, which are based on both the shape parameter and the mean, are not. Thus, the distributions not only have different means, but different standard deviations!

## Example 1 – Power at Various Sample Sizes

Researchers are planning a parallel-group experiment to test whether the difference in response to a certain drug is zero. The researchers will use a two-sided t-test with an alpha level of 0.05. They want to compare the power at sample sizes of 50, 100, and 200 when the shift in the means is 0.6 from drug 1 to drug 2. They assume that the data are normally distributed with a standard deviation of 2. Since this is an exploratory analysis, they set the number of simulation iterations to 2000.

### 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 .....	<b>Power</b>
Alternative Hypothesis .....	<b>Two-Sided (H1: <math>\mu_1 - \mu_2 \neq \delta_0</math>)</b>
Test Type .....	<b>Equal-Variance T-Test</b>
Simulations .....	<b>2000</b>
Random Seed .....	<b>4045049</b> (for Reproducibility)
Alpha .....	<b>0.05</b>
Group Allocation .....	<b>Equal (N1 = N2)</b>
Sample Size Per Group .....	<b>50 100 200</b>
Input Type .....	<b>Simple (Differences)</b>
Distribution to Simulate .....	<b>Normal</b>
$\delta_0$ (Null Difference) .....	<b>0</b>
$\delta_1$ (Actual Difference) .....	<b>0.6</b>
$\mu_2$ (Group 2 Mean) .....	<b>0</b>
Group Standard Deviations .....	<b>Equal</b>
$\sigma$ (Standard Deviation) .....	<b>2</b>

Tests for Two Means (Simulation)

# Output

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

### Numeric Results

Solve For: [Power](#)  
 Hypotheses:  $H_0: \delta = \delta_0$  vs.  $H_1: \delta \neq \delta_0$   
 Test Statistic: Equal-Variance T-Test  
 Simulated Distribution: Normal

Power	Sample Size			Mean			Difference		Standard Deviation $\sigma$	Alpha	
	N1	N2	N	Null $\mu_{1.0}$	Actual $\mu_{1.1}$	Reference $\mu_2$	Null $\delta_0$	Actual $\delta_1$		Target	Actual
0.3135 (0.0203) [0.2932 0.3338]	50	50	100	0	0.6	0	0	0.6	2	0.05	0.051 (0.01) [0.041 0.061]
0.5640 (0.0217) [0.5423 0.5857]	100	100	200	0	0.6	0	0	0.6	2	0.05	0.052 (0.01) [0.042 0.061]
0.8550 (0.0154) [0.8396 0.8704]	200	200	400	0	0.6	0	0	0.6	2	0.05	0.048 (0.009) [0.038 0.057]

Pool Size: 10000. Simulations: 2000. Run Time: 2.14 seconds.  
 User-Entered Random Seed: 4045049

- 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.
- N1 and N2 The number of items sampled from each population.
- N The total sample size.  $N = N1 + N2$ .
- $\mu_{1.0}$  The mean of group 1 under the null hypothesis.
- $\mu_{1.1}$  The actual mean of group 1 at which power and sample size are calculated.
- $\mu_2$  The mean of group 2 under both the null and alternative hypotheses.
- $\delta_0$  The mean difference under the null hypothesis.  $\delta_0 = \mu_{1.0} - \mu_2$ .
- $\delta_1$  The actual mean difference at which power and sample size are calculated.  $\delta_1 = \mu_{1.1} - \mu_2$ .
- $\sigma$  The assumed population standard deviation for both groups 1 and 2.
- 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 parallel two-group design will be used to test whether the Group 1 mean is different from the Group 2 mean ( $H_0: \delta = 0$  versus  $H_1: \delta \neq 0, \delta = \mu_1 - \mu_2$ ). The comparison will be made using a two-sided, two-sample, equal-variance t-test, with a Type I error rate ( $\alpha$ ) of 0.05. A standard deviation of 2 is assumed for both Group 1 and Group 2. To detect a difference in means of  $\mu_1 - \mu_2 = 0.6 - 0 = 0.6$ , with a sample size of 50 in Group 1 and 50 in Group 2, the power is 0.3135. These results are based on 2000 simulations (Monte Carlo samples) from the Normal distribution.

Tests for Two Means (Simulation)

**Dropout-Inflated Sample Size**

Dropout Rate	Sample Size			Dropout-Inflated Enrollment Sample Size			Expected Number of Dropouts		
	N1	N2	N	N1'	N2'	N'	D1	D2	D
20%	50	50	100	63	63	126	13	13	26
20%	100	100	200	125	125	250	25	25	50
20%	200	200	400	250	250	500	50	50	100

- 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.
- N1, N2, and N      The evaluable sample sizes at which power is computed (as entered by the user). If N1 and N2 subjects are evaluated out of the N1' and N2' subjects that are enrolled in the study, the design will achieve the stated power.
- N1', N2', and N'      The number of subjects that should be enrolled in the study in order to obtain N1, N2, and N evaluable subjects, based on the assumed dropout rate. N1' and N2' are calculated by inflating N1 and N2 using the formulas  $N1' = N1 / (1 - DR)$  and  $N2' = N2 / (1 - DR)$ , with N1' and N2' 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.)
- D1, D2, and D      The expected number of dropouts.  $D1 = N1' - N1$ ,  $D2 = N2' - N2$ , and  $D = D1 + D2$ .

**Dropout Summary Statements**

Anticipating a 20% dropout rate, 63 subjects should be enrolled in Group 1, and 63 in Group 2, to obtain final group sample sizes of 50 and 50, respectively.

**References**

Chow, S.C., Shao, J., Wang, H., and Lohknygina, 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.

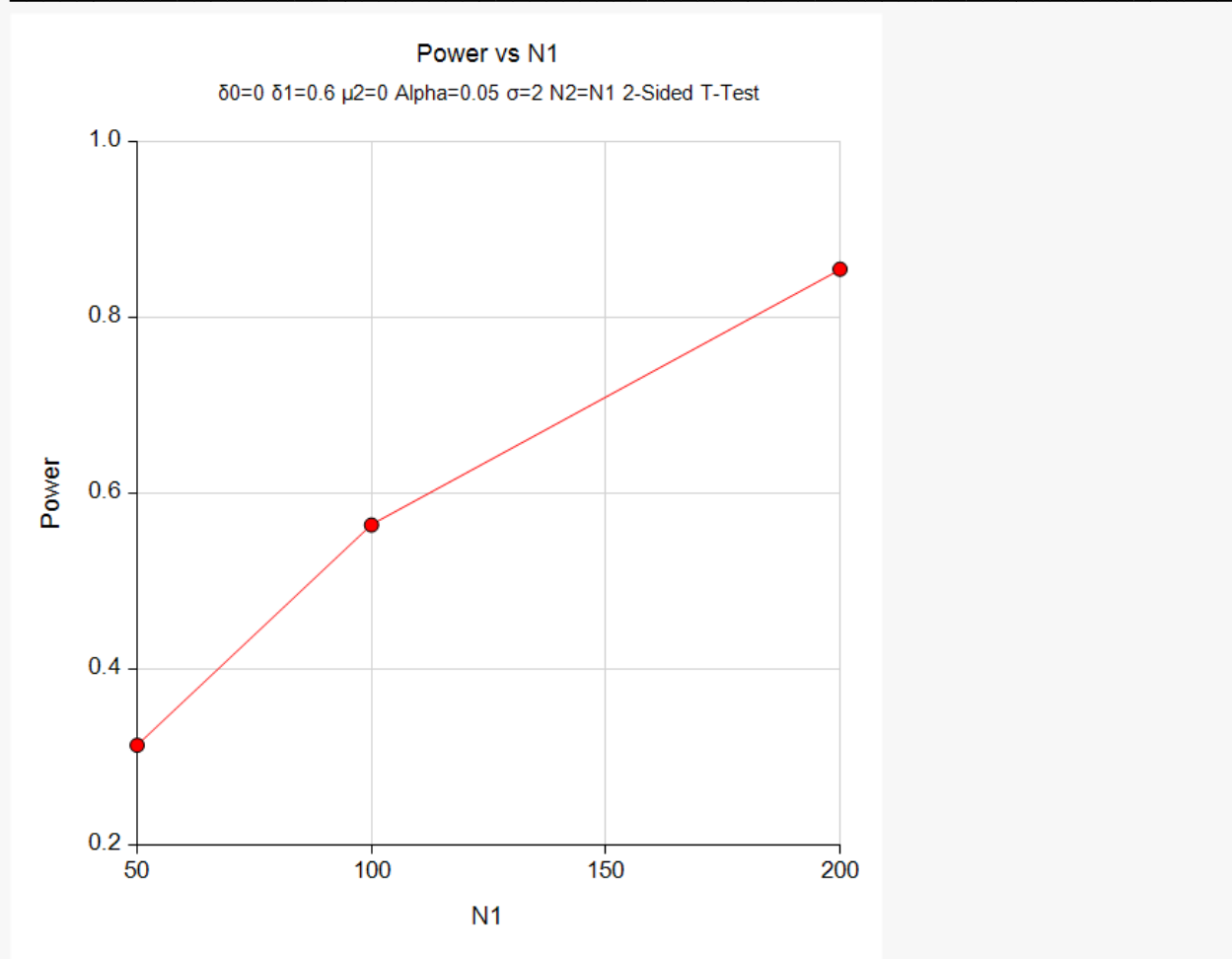
Matsumoto, M. and Nishimura, T. 1998. 'Mersenne twister: A 623-dimensionally equidistributed uniform pseudorandom number generator.' ACM Trans. On Modeling and Computer Simulations.

Zar, Jerrold H. 1984. Biostatistical Analysis (Second Edition). Prentice-Hall. Englewood Cliffs, New Jersey.



Tests for Two Means (Simulation)

Plots



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).

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 intervals will decrease.

## Example 2 – Finding the Sample Size

Continuing with Example 1, the researchers want to determine how large a sample is needed to obtain a power of 0.90.

### 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 .....	<b>Sample Size</b>
Alternative Hypothesis .....	<b>Two-Sided (H1: <math>\mu_1 - \mu_2 \neq \delta_0</math>)</b>
Test Type.....	<b>Equal-Variance T-Test</b>
Simulations .....	<b>2000</b>
Random Seed.....	<b>7544944</b> (for Reproducibility)
Power.....	<b>0.90</b>
Alpha.....	<b>0.05</b>
Group Allocation .....	<b>Equal (N1 = N2)</b>
Input Type.....	<b>Simple (Differences)</b>
Distribution to Simulate .....	<b>Normal</b>
$\delta_0$ (Null Difference).....	<b>0</b>
$\delta_1$ (Actual Difference).....	<b>0.6</b>
$\mu_2$ (Group 2 Mean) .....	<b>0</b>
Group Standard Deviations.....	<b>Equal</b>
$\sigma$ (Standard Deviation).....	<b>2</b>

## Tests for Two Means (Simulation)

## Output

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

### Numeric Results

Solve For: [Sample Size](#)  
 Hypotheses:  $H_0: \delta = \delta_0$  vs.  $H_1: \delta \neq \delta_0$   
 Test Statistic: Equal-Variance T-Test  
 Simulated Distribution: Normal

Power	Sample Size			Mean			Difference		Standard Deviation $\sigma$	Alpha	
	N1	N2	N	Null $\mu_{1.0}$	Actual $\mu_{1.1}$	Reference $\mu_2$	Null $\delta_0$	Actual $\delta_1$		Target	Actual
0.9105 (0.0125) [0.898 0.923]	237	237	474	0	0.6	0	0	0.6	2	0.05	0.052 (0.01) [0.042 0.062]

Pool Size: 10000. Simulations: 2000. Run Time: 13.38 seconds.  
 User-Entered Random Seed: 7544944

The required sample size was 237 which achieved a power of 0.9105. To check the accuracy of this simulation, we ran this scenario through the analytic procedure in **PASS** which gave the sample size as 235 per group. The simulation result of 237 is very close.

## Example 3 – Comparative Results

Continuing with Example 1, the researchers want to study the characteristics of alternative test statistics. They want to compare the results of all test statistics for  $N_1 = 50, 100,$  and  $200$ .

### 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 .....	<b>Power</b>
Alternative Hypothesis .....	<b>Two-Sided (<math>H_1: \mu_1 - \mu_2 \neq \delta_0</math>)</b>
Test Type.....	<b>Equal-Variance T-Test</b>
Simulations .....	<b>2000</b>
Random Seed.....	<b>4061205</b> (for Reproducibility)
Alpha.....	<b>0.05</b>
Group Allocation .....	<b>Equal (<math>N_1 = N_2</math>)</b>
Sample Size Per Group .....	<b>50 100 200</b>
Input Type.....	<b>Simple (Differences)</b>
Distribution to Simulate .....	<b>Normal</b>
$\delta_0$ (Null Difference).....	<b>0</b>
$\delta_1$ (Actual Difference).....	<b>0.6</b>
$\mu_2$ (Group 2 Mean) .....	<b>0</b>
Group Standard Deviations.....	<b>Equal</b>
$\sigma$ (Standard Deviation).....	<b>2</b>

#### Reports Tab

Show Comparative Reports..... **Checked**

#### Comparative Plots Tab

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

## Tests for Two Means (Simulation)

## Output

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

**Power Comparison**

Hypotheses:  $H_0: \delta = \delta_0$  vs.  $H_1: \delta \neq \delta_0$   
 Simulated Distribution: Normal

Sample Size		Difference		Standard Deviation $\sigma$	Target Alpha	Power				
N1	N2	Null $\delta_0$	Actual $\delta_1$			T-Test	Welch	Trimmed T-Test	Trimmed Welch	Mann-Whitney
50	50	0	0.6	2	0.05	0.3065	0.3065	0.3060	0.3055	0.2975
100	100	0	0.6	2	0.05	0.5690	0.5690	0.5305	0.5305	0.5460
200	200	0	0.6	2	0.05	0.8475	0.8475	0.8315	0.8315	0.8390

Pool Size: 10000. Simulations: 2000. Run Time: 39.57 seconds. Percent Trimmed at each end: 10%.  
 User-Entered Random Seed: 4061205

**Alpha Comparison**

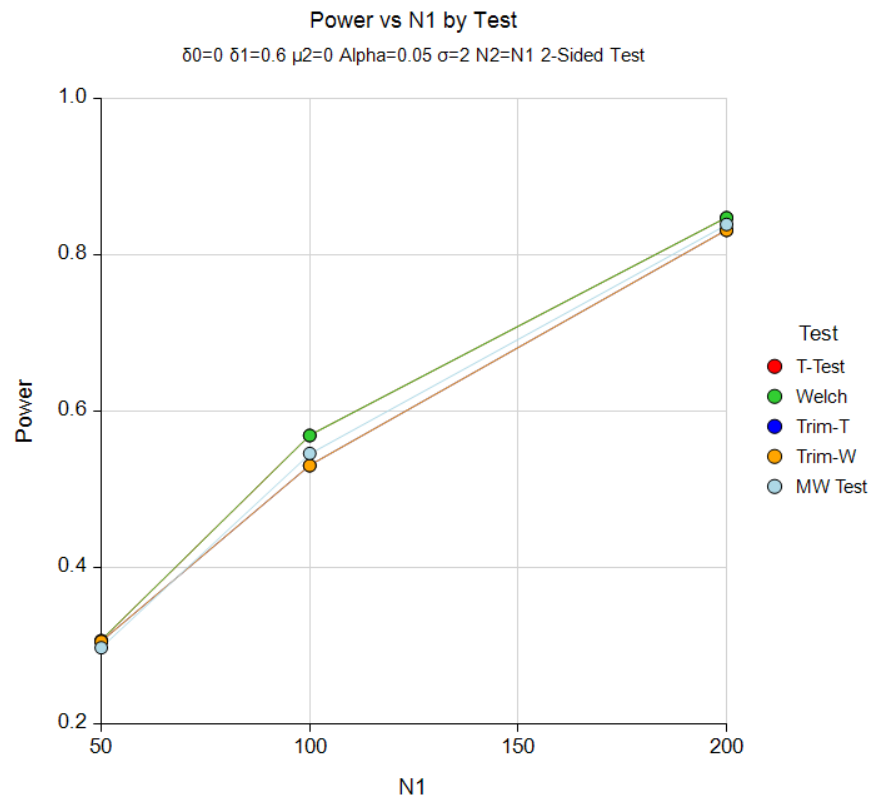
Hypotheses:  $H_0: \delta = \delta_0$  vs.  $H_1: \delta \neq \delta_0$   
 Simulated Distribution: Normal

Sample Size		Difference		Standard Deviation $\sigma$	Target	Alpha				
N1	N2	Null $\delta_0$	Actual $\delta_1$			T-Test	Welch	Trimmed T-Test	Trimmed Welch	Mann-Whitney
50	50	0	0.6	2	0.05	0.041	0.041	0.043	0.043	0.045
100	100	0	0.6	2	0.05	0.055	0.055	0.049	0.049	0.050
200	200	0	0.6	2	0.05	0.048	0.048	0.050	0.050	0.049

Pool Size: 10000. Simulations: 2000. Run Time: 39.57 seconds. Percent Trimmed at each end: 10%.  
 User-Entered Random Seed: 4061205

## Tests for Two Means (Simulation)

## Comparative Plots



These results show that for data that fit the assumptions of the t-test, all five test statistics have accurate alpha values and reasonably close power values. It is interesting to note that the powers of the trimmed procedures are very close to the t-test, even though about 20% of the data were trimmed.

## Example 4 – Validation using Zar (1984)

Zar (1984) page 136 gives an example in which the mean difference is 1, the common standard deviation is 0.7206, the sample sizes are 15 in each group, and the significance level is 0.05. They calculate the power to be 0.96 for a two-sample t-test.

For reproducibility, we'll use a random seed of 6098723.

### 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 .....	<b>Power</b>
Alternative Hypothesis .....	<b>Two-Sided (H1: <math>\mu_1 - \mu_2 \neq \delta_0</math>)</b>
Test Type .....	<b>Equal-Variance T-Test</b>
Simulations .....	<b>10000</b>
Random Seed .....	<b>6098723</b> (for Reproducibility)
Alpha .....	<b>0.05</b>
Group Allocation .....	<b>Equal (N1 = N2)</b>
Sample Size Per Group .....	<b>15</b>
Input Type .....	<b>Simple (Differences)</b>
Distribution to Simulate .....	<b>Normal</b>
$\delta_0$ (Null Difference) .....	<b>0</b>
$\delta_1$ (Actual Difference) .....	<b>1</b>
$\mu_2$ (Group 2 Mean) .....	<b>0</b>
Group Standard Deviations .....	<b>Equal</b>
$\sigma$ (Standard Deviation) .....	<b>0.7206</b>

## Tests for Two Means (Simulation)

## Output

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

### Numeric Results

Solve For: [Power](#)  
 Hypotheses:  $H_0: \delta = \delta_0$  vs.  $H_1: \delta \neq \delta_0$   
 Test Statistic: Equal-Variance T-Test  
 Simulated Distribution: Normal

Power	Sample Size			Mean			Difference		Standard Deviation $\sigma$	Alpha	
	N1	N2	N	Null $\mu_{1.0}$	Actual $\mu_{1.1}$	Reference $\mu_2$	Null $\delta_0$	Actual $\delta_1$		Target	Actual
0.9568 (0.004) [0.9528 0.9608]	15	15	30	0	1	0	0	1	0.7	0.05	0.052 (0.004) [0.047 0.056]

Pool Size: 20000. Simulations: 10000. Run Time: 0.78 seconds.  
 User-Entered Random Seed: 6098723

The power matches the exact value of 0.96.



## Example 5 – Non-Inferiority Test

A non-inferiority test is used to show that a new treatment is not significantly worse than the standard (or reference) treatment. The maximum deviation that is 'not significantly worse' is called the *margin of equivalence*.

Suppose that the mean diastolic BP of subjects on a certain drug is 96mmHg. If the mean diastolic BP of a new drug is not more than 100mmHg, the drug will be considered non-inferior to the standard drug. The standard deviation among these subjects is 6 mmHg.

The developers of this new drug must design an experiment to test the hypothesis that the mean difference between the two mean BP's is less than 4. The statistical hypothesis to be tested is

$$H_0: \mu_N - \mu_S \geq 4 \quad \text{vs.} \quad H_1: \mu_N - \mu_S < 4$$

Notice that when the null hypothesis is rejected, the conclusion is that the average difference is less than 4. Following proper procedure, they use a significance level of 0.025 for this one-sided test to keep it comparable to the usual value of 0.05 for a two-sided test. They decide to find the sample size at which the power is 0.90 when the two means are actually equal.

### 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 .....	<b>Sample Size</b>
Alternative Hypothesis .....	<b>One-Sided (H1: <math>\mu_1 - \mu_2 &lt; \delta_0</math>)</b>
Test Type.....	<b>Equal-Variance T-Test</b>
Simulations .....	<b>2000</b>
Random Seed.....	<b>4086609</b> (for Reproducibility)
Power.....	<b>0.90</b>
Alpha.....	<b>0.025</b>
Group Allocation .....	<b>Equal (N1 = N2)</b>
Input Type.....	<b>Simple (Differences)</b>
Distribution to Simulate .....	<b>Normal</b>
$\delta_0$ (Null Difference).....	<b>4</b>
$\delta_1$ (Actual Difference).....	<b>0</b>
$\mu_2$ (Group 2 Mean) .....	<b>96</b>
Group Standard Deviations.....	<b>Equal</b>
$\sigma$ (Standard Deviation).....	<b>6</b>

## Tests for Two Means (Simulation)

## Output

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

### Numeric Results

Solve For: [Sample Size](#)  
 Hypotheses:  $H_0: \delta \geq \delta_0$  vs.  $H_1: \delta < \delta_0$   
 Test Statistic: Equal-Variance T-Test  
 Simulated Distribution: Normal

Power	Sample Size			Mean			Difference		Standard Deviation $\sigma$	Alpha	
	N1	N2	N	Null $\mu_{1.0}$	Actual $\mu_{1.1}$	Reference $\mu_2$	Null $\delta_0$	Actual $\delta_1$		Target	Actual
0.9115 (0.0124) [0.8991 0.9239]	49	49	98	100	96	96	4	0	6	0.025	0.024 (0.007) [0.017 0.031]

Pool Size: 10000. Simulations: 2000. Run Time: 1.94 seconds.  
 User-Entered Random Seed: 4086609

We see that 49 subjects are required to achieve the desired experimental design.

## Example 6 – Selecting a Test Statistic when the Data Contain Outliers

The two-sample t-test is known to be robust to the violation of some assumptions, but it is susceptible to inaccuracy because the data contain outliers. This example will investigate the impact of outliers on the power and precision of the five test statistics available in **PASS**.

A mixture of two normal distributions will be used to randomly generate outliers. The mixture will draw 95% of the data from a normal distribution with a mean of 0 and a standard deviation of 1. The other 5% of the data will come from a normal distribution with a mean of 0 and a standard deviation that ranges from 1 to 10.

### 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 .....	<b>Power</b>
Alternative Hypothesis .....	<b>Two-Sided (H1: <math>\mu_1 - \mu_2 \neq \delta_0</math>)</b>
Test Type .....	<b>Equal-Variance T-Test</b>
Simulations .....	<b>2000</b>
Random Seed .....	<b>4089500</b> (for Reproducibility)
Alpha .....	<b>0.05</b>
Group Allocation .....	<b>Equal (N1 = N2)</b>
Sample Size Per Group .....	<b>20</b>
Input Type .....	<b>General</b>
Group 1 Distribution H0 .....	<b>Normal(M0 S)[95];Normal(M0 A)[5]</b>
Group 1 Distribution H1 .....	<b>Normal(M1 S)[95];Normal(M1 A)[5]</b>
Group 2 Distribution H0 .....	<b>Normal(M0 S)[95];Normal(M0 A)[5]</b>
Group 2 Distribution H1 .....	<b>Normal(M0 S)[95];Normal(M0 A)[5]</b>
M0 (Mean H0) Parameter Value(s) .....	<b>0</b>
M1 (Mean H1) Parameter Value(s) .....	<b>1</b>
Parameter 1 Label .....	<b>S</b>
Parameter 1 Value(s) .....	<b>1</b>
Parameter 2 Label .....	<b>A</b>
Parameter 2 Value(s) .....	<b>1 5 10</b>

#### Reports Tab

Show Comparative Reports .....	<b>Checked</b>
--------------------------------	----------------

#### Comparative Plots Tab

Show Comparative Plots .....	<b>Checked</b>
------------------------------	----------------

Tests for Two Means (Simulation)

## Output

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

### Power Comparison

Hypotheses: H0:  $\delta = \delta_0$  vs. H1:  $\delta \neq \delta_0$   
 Group 1 Distributions: H0: Normal(M0 S)[95];Normal(M0 A)[5] H1: Normal(M1 S)[95];Normal(M1 A)[5]  
 Group 2 Distributions: H0: Normal(M0 S)[95];Normal(M0 A)[5] H1: Normal(M0 S)[95];Normal(M0 A)[5]

Sample Size		Difference			Power								
N1	N2	Null $\delta_0$	Actual $\delta_1$	Target Alpha	T-Test	Welch	Trimmed T-Test	Trimmed Welch	Mann-Whitney	M0	M1	S	A
20	20	0	1	0.05	0.8620	0.8615	0.8420	0.8395	0.8420	0	1	1	1
20	20	0	1	0.05	0.6395	0.6365	0.8120	0.8100	0.8125	0	1	1	5
20	20	0	1	0.05	0.4875	0.4790	0.7705	0.7685	0.7720	0	1	1	10

Pool Size: 10000. Simulations: 2000. Run Time: 9.00 seconds. Percent Trimmed at each end: 10%.  
 User-Entered Random Seed: 4089500

### Alpha Comparison

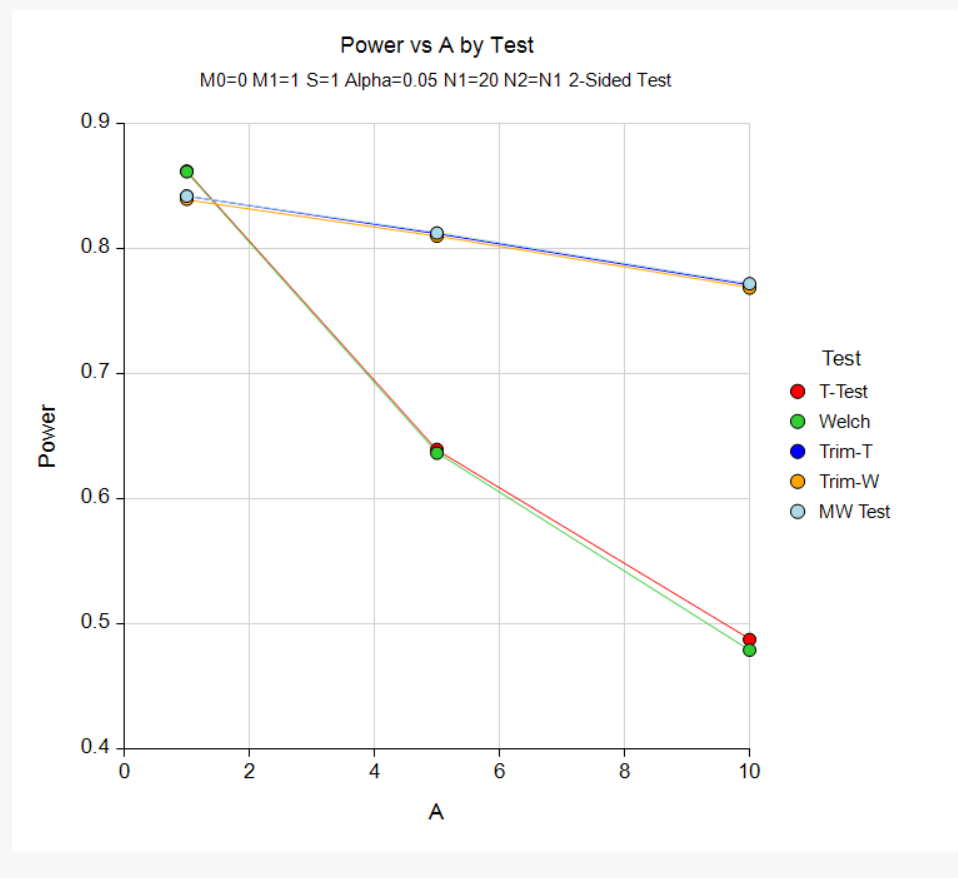
Hypotheses: H0:  $\delta = \delta_0$  vs. H1:  $\delta \neq \delta_0$   
 Group 1 Distributions: H0: Normal(M0 S)[95];Normal(M0 A)[5] H1: Normal(M1 S)[95];Normal(M1 A)[5]  
 Group 2 Distributions: H0: Normal(M0 S)[95];Normal(M0 A)[5] H1: Normal(M0 S)[95];Normal(M0 A)[5]

Sample Size		Difference			Alpha								
N1	N2	Null $\delta_0$	Actual $\delta_1$	Target	T-Test	Welch	Trimmed T-Test	Trimmed Welch	Mann-Whitney	M0	M1	S	A
20	20	0	1	0.05	0.048	0.048	0.050	0.049	0.049	0	1	1	1
20	20	0	1	0.05	0.036	0.035	0.052	0.052	0.051	0	1	1	5
20	20	0	1	0.05	0.035	0.032	0.051	0.051	0.053	0	1	1	10

Pool Size: 10000. Simulations: 2000. Run Time: 9.00 seconds. Percent Trimmed at each end: 10%.  
 User-Entered Random Seed: 4089500

## Tests for Two Means (Simulation)

## Comparative Plots



The first line gives the results for the standard case in which the two standard deviations ( $S$  and  $A$ ) are equal. Note that in this case, the power of the t-test is slightly higher than for the other tests. As the amount of contamination is increased ( $A$  equal 5 and then 10), the power of the trimmed tests and the Mann-Whitney test remain high, but the power of the t-test falls from 88% to 43%. Also, the value of alpha remains constant for the trimmed and nonparametric tests, but the alpha of the t-test becomes very conservative. The conclusion this simulation is that if there is a possibility of outliers, you should use either the nonparametric test or the trimmed test.

## Example 7 – Selecting a Test Statistic when the Data are Skewed

The two-sample t-test is known to be robust to the violation of some assumptions, but it is susceptible to inaccuracy when the underlying distributions are skewed. This example will investigate the impact of skewness on the power and precision of the five test statistics available in **PASS**.

Tukey's lambda distribution will be used because it allows the amount of skewness to be gradually increased.

### 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 .....	<b>Power</b>
Alternative Hypothesis .....	<b>Two-Sided (H1: <math>\mu_1 - \mu_2 \neq \delta_0</math>)</b>
Test Type.....	<b>Equal-Variance T-Test</b>
Simulations .....	<b>2000</b>
Random Seed.....	<b>4097943</b> (for Reproducibility)
Alpha.....	<b>0.05</b>
Group Allocation .....	<b>Equal (N1 = N2)</b>
Sample Size Per Group .....	<b>20</b>
Input Type.....	<b>Simple (Differences)</b>
Distribution to Simulate .....	<b>TukeyGH</b>
G (Skewness) .....	<b>0 0.5 0.9</b>
H (Kurtosis).....	<b>0</b>
$\delta_0$ (Null Difference).....	<b>0</b>
$\delta_1$ (Actual Difference).....	<b>1</b>
$\mu_2$ (Group 2 Mean) .....	<b>0</b>
Group Standard Deviations.....	<b>Equal</b>
$\sigma$ (Standard Deviation).....	<b>1</b>

#### Reports Tab

Show Comparative Reports ..... **Checked**

#### Comparative Plots Tab

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

## Tests for Two Means (Simulation)

## Output

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

## Power Comparison

Hypotheses:  $H_0: \delta = \delta_0$  vs.  $H_1: \delta \neq \delta_0$   
 Simulated Distribution: TukeyGH

Sample Size		Difference		Standard Deviation $\sigma$	Target Alpha	Power						
N1	N2	Null $\delta_0$	Actual $\delta_1$			T-Test	Welch	Trimmed T-Test	Trimmed Welch	Mann-Whitney	G	H
20	20	0	1	1	0.05	0.8590	0.8585	0.8410	0.8385	0.8405	0.0	0
20	20	0	1	1	0.05	0.8570	0.8555	0.8995	0.8985	0.9450	0.5	0
20	20	0	1	1	0.05	0.8935	0.8905	0.9700	0.9690	0.9925	0.9	0

Pool Size: 10000. Simulations: 2000. Run Time: 8.75 seconds. Percent Trimmed at each end: 10%.  
 User-Entered Random Seed: 4097943

## Alpha Comparison

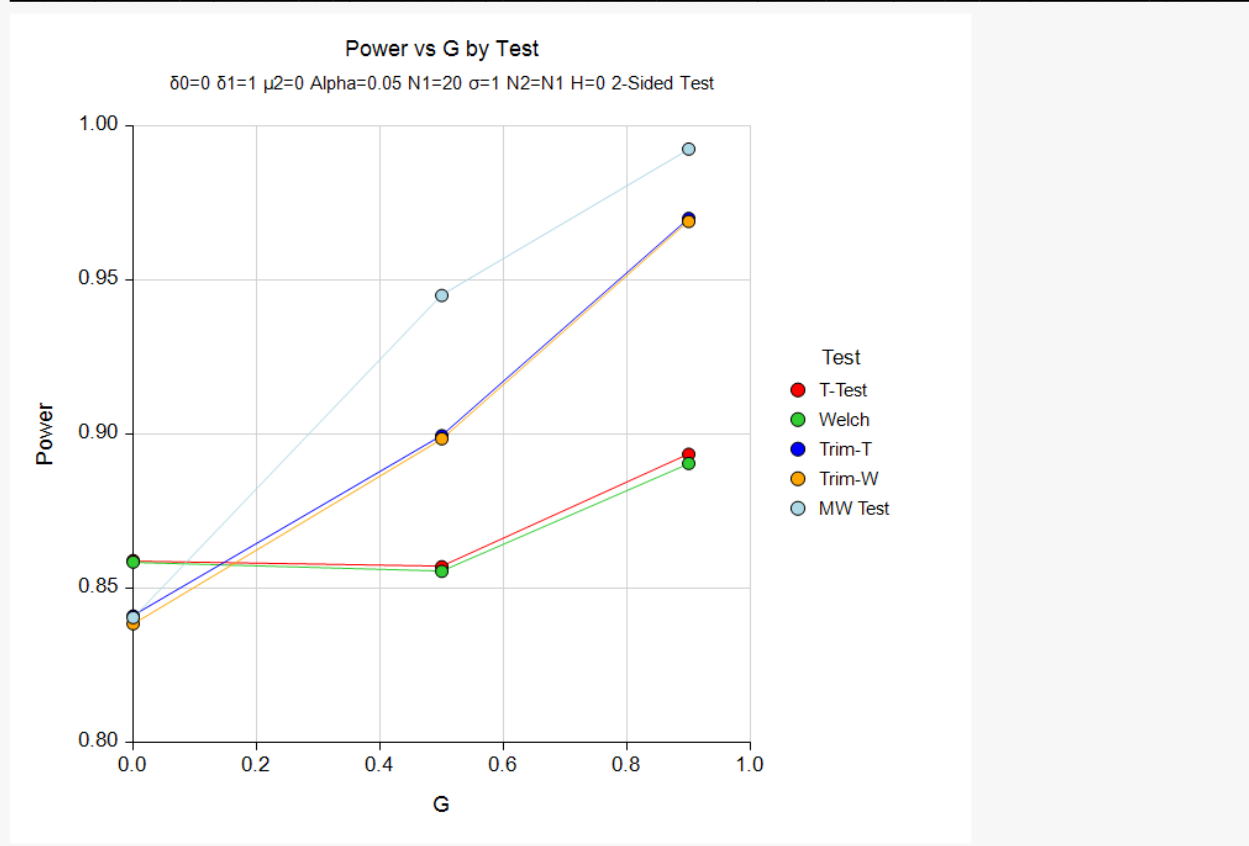
Hypotheses:  $H_0: \delta = \delta_0$  vs.  $H_1: \delta \neq \delta_0$   
 Simulated Distribution: TukeyGH

Sample Size		Difference		Standard Deviation $\sigma$	Target Alpha	Alpha						
N1	N2	Null $\delta_0$	Actual $\delta_1$			T-Test	Welch	Trimmed T-Test	Trimmed Welch	Mann-Whitney	G	H
20	20	0	1	1	0.05	0.058	0.057	0.055	0.053	0.053	0.0	0
20	20	0	1	1	0.05	0.049	0.048	0.053	0.052	0.057	0.5	0
20	20	0	1	1	0.05	0.040	0.036	0.041	0.039	0.046	0.9	0

Pool Size: 10000. Simulations: 2000. Run Time: 8.75 seconds. Percent Trimmed at each end: 10%.  
 User-Entered Random Seed: 4097943

Tests for Two Means (Simulation)

Comparative Plots



The first line gives the results for the standard case in which there is no skewness ( $G = 0$ ). Note that in this case, the power of the t-test is a little higher than that of the other tests. As the amount of skewness is increased ( $G$  equal 0.5 and then 0.9), the power of the trimmed tests and the Mann-Whitney test increases, but the power of the t-test remains about the same. Also, the value of alpha remains constant for all tests. The conclusion of this simulation is that if there is skewness, you will gain power by using the nonparametric or trimmed test.