Chapter 465

Equivalence Tests for Two Means (Simulation)

Introduction

This procedure allows you to study the power and sample size of an equivalence test comparing two means from independent groups. Schuirmann's (1987) two one-sided tests (TOST) approach is used to test equivalence. The t-test is commonly used in this situation, but other tests have been developed for use when the t-test assumptions are not met. These additional tests include the Mann-Whitney U test, Welch's unequal variance test, and trimmed versions of the t-test and the Welch test.

Measurements are made on individuals that have been randomly assigned to, or randomly chosen from, one of two groups. This *parallel-groups* design may be analyzed by a TOST equivalence test to show that the means of the two groups do not differ by more than a small amount, called the margin of equivalence.

The two-sample t-test is commonly used in this situation. When the variances of the two groups are unequal, Welch's t-test is often used. When the data are not normally distributed, the Mann-Whitney (Wilcoxon signed-ranks) U test and, less frequently, the trimmed t-test may be used.

The details of the power analysis of equivalence test using analytic techniques are presented in another **PASS** chapter and they will not 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 as follows:

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.

© NCSS, LLC. All Rights Reserved.

4. 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.

Simulating Data for an Equivalence Test

Simulating equivalence data is more complex than simulating data for a regular two-sided test. An equivalence test essentially reverses the roles of the null and alternative hypothesis. The null hypothesis becomes

$$H0: (\mu_1 - \mu_2) \leq -D \quad \text{or} \quad (\mu_1 - \mu_2) \geq D$$

where *D* is the margin of equivalence. Thus, the null hypothesis is made up of two simple hypotheses:

$$H0_1: (\mu_1 - \mu_2) \leq -D$$
$$H0_2: (\mu_1 - \mu_2) \geq D$$

The additional complexity comes in deciding which of the two null hypotheses are used to simulate data for the null hypothesis situation. The choice becomes more problematic when asymmetric equivalence limits are chosen. In this case, you may want to try simulating using each simple null hypothesis in turn.

To generate data for the null hypotheses, generate data for each group. The difference in the means of these two groups will become one of the equivalence limits. The other equivalence limit will be determined by symmetry and will always have a sign that is the opposite of the first equivalence limit.

Test Statistics

This section describes the test statistics that are available in this procedure. Note that these test statistics are computed on the differences. Thus, when the equation refers to an X value, this X value is assumed to be a difference between two individual variates.

© NCSS, LLC. All Rights Reserved.

Two-Sample T-Test

The t-test assumes that the data are simple random samples from populations of normally distributed values that 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 a p-value which is based on the t distribution with appropriate degrees of freedom. If this p-value is less than a specified level (often 0.05), the null hypothesis is rejected. Otherwise, no conclusion can be reached.

Welch's T-Test

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

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

where

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

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

$$s_1 = \sqrt{\left(\frac{\sum_{i=1}^{N_1}(X_{1i} - \bar{X}_1)^2}{N_1 - 1}\right)}, \quad s_2 = \sqrt{\frac{\sum_{i=1}^{N_2}(X_{2i} - \bar{X}_2)^2}{N_2 - 1}}$$

© NCSS, LLC. All Rights Reserved.

Trimmed T-Test assuming 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 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* – 2*g* 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.

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

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

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

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

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

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.

© NCSS, LLC. All Rights Reserved.

Trimmed T-Test assuming 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

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

with degrees of freedom *f* given by

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

where

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

Mann-Whitney U 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 H0. This means that ties (repeated values) are not acceptable. When ties are present, an approximation can be used, but the theoretic results no longer hold.

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

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

where

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

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

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

© NCSS, LLC. All Rights Reserved.

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 an experiment to determine if the response to a new drug is equivalent to the response to the standard drug. The average response level to the standard drug is known to be 63 with a standard deviation of 5. The researchers decide that if the average response level to the new drug is between 60 and 66, they will consider it to be equivalent to the standard drug.

The researchers decide to use a parallel-group design. The response level for the standard drug will be measured for each subject. They will analyze the data using an equivalence test based on the t-test with an alpha level of 0.05. They want to compare the power at sample sizes of 10, 30, 50, and 70. They assume that the data are normally distributed and that the true difference between the mean response of the two drugs is zero. Since this is an exploratory analysis, the number of simulation iterations is set 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 1** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

Solve For	**Power**
Equivalence Limit	**Symmetric**
Test Type	**T-Test**
Simulations	**2000**
Random Seed	**4426805** (for Reproducibility)
Alpha	**0.05**
Group Allocation	**Equal (N1 = N2)**
Sample Size Per Group	**10 30 50 70**
Group 1 Distribution|H0	**Normal(M0 S)**
Group 1 Distribution|H1	**Normal(M0 S)**
Group 2 Distribution|H0	**Normal(M1 S)**
Group 2 Distribution|H1	**Normal(M0 S)**
M0 (Mean|H0) Parameter Value(s)	**63**
M1 (Mean|H1) Parameter Value(s)	**66**
Parameter 1 Label	**S**
Parameter 1 Value(s)	**5**

© NCSS, LLC. All Rights Reserved.

Output

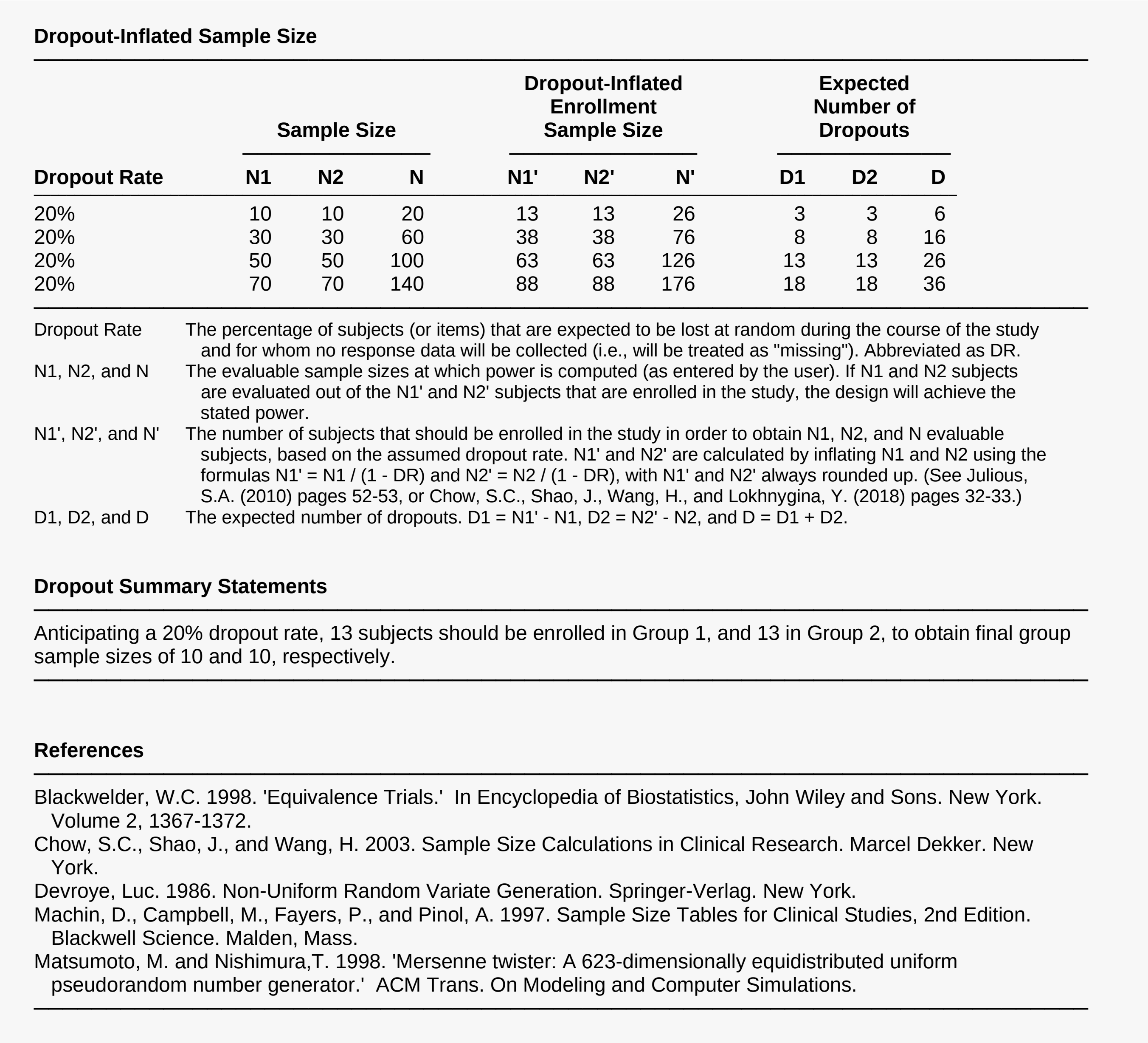
Click the Calculate button to perform the calculations and generate the following output.

```
Numeric Results
──────────────────────────────────────────────────────────────────────────────────────────
Solve For:          Power
Hypotheses:         H0: Diff ≥ |Diff0|   vs.   H1: Diff < |Diff0|
H0 Distributions:   Normal(M0 S) & Normal(M1 S)
H1 Distributions:   Normal(M0 S) & Normal(M0 S)
Test Statistic:     T-Test
──────────────────────────────────────────────────────────────────────────────────────────
                                              Equivalence
                    Sample          Mean         Limits                 Alpha
                      Size    Difference   ──────────────       ──────────────────
Power        N1 | N2 | N         Diff1      Lower     Upper      Target    Actual     M0     M1     S
─────────────────────────────────────────────────────────────────────────────────────────────────────
0.012         10 | 10 | 20            0        -3         3        0.05     0.003     63     66     5
(0.005) [0.007 0.017]                                                 (0.002) [0.001 0.005]

0.473         30 | 30 | 60            0        -3         3        0.05     0.041     63     66     5
(0.022) [0.451 0.494]                                                 (0.009) [0.032 0.05]

0.834        50 | 50 | 100            0        -3         3        0.05     0.055     63     66     5
(0.016) [0.818 0.85]                                                  (0.01) [0.045 0.064]

0.943        70 | 70 | 140            0        -3         3        0.05     0.059     63     66     5
(0.01) [0.933 0.953]                                                  (0.01) [0.048 0.069]
──────────────────────────────────────────────────────────────────────────────────────────
Pool Size: 10000. Simulations: 2000. Run Time: 1.34 seconds.
User-Entered Random Seed: 4426805

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 | N2 | N              The size of the samples drawn from groups 1, 2, and both.
Diff1                    The difference between the means (Group 1 - Group 2) under the alternative hypothesis, H1.
Lower Equivalence Limit  The lower limit on the mean difference that is considered as equivalent.
Diff0                    The difference between the means (Group 1 - Group 2) assuming the null hypothesis, H0. This is
                           one of the equivalence limits.
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.
Parameters               The additional columns in the report represent the distribution parameters used (if any) in the
                           calculations.
```

Summary Statements

A parallel, two-group design will be used to test whether the Group 1 mean (μ1) is equivalent to the Group 2 mean (μ2), with lower and upper mean difference equivalence limits of -3 and 3 (H0: δ ≤ -3 or δ ≥ 3 versus H1: -3 < δ < 3, δ = μ1 - μ2). The comparison will be made using two one-sided t-tests, with an overall Type I error rate (α) of 0.05. To detect a mean difference of 0 with sample sizes of 10 for Group 1 and 10 for Group 2, the power is 0.012. These results are based on 2000 simulations (Monte Carlo samples) from the null distributions: Normal(M0 S) and Normal(M1 S), and the alternative distributions: Normal(M0 S) and Normal(M0 S), with M0 = 63, M1 = 66, S = 5.

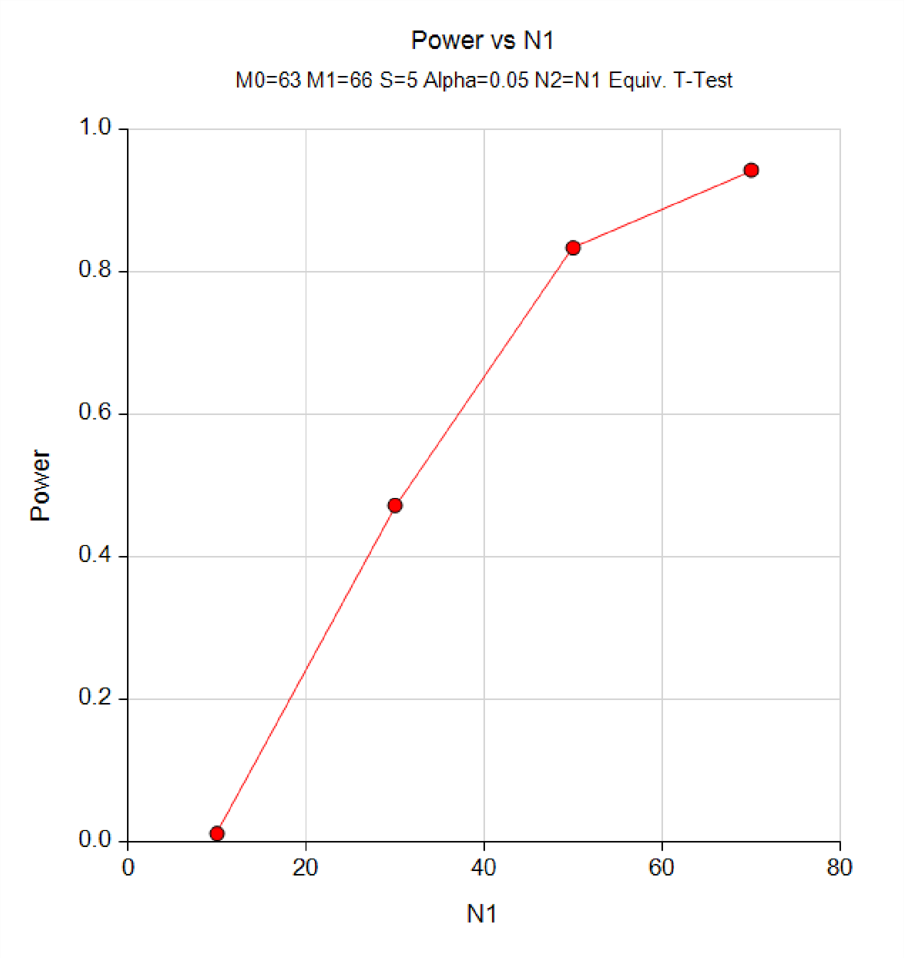
© NCSS, LLC. All Rights Reserved.



```
Dropout-Inflated Sample Size
─────────────────────────────────────────────────────────────────────────────────────────
                                           Dropout-Inflated          Expected
                                              Enrollment             Number of
                        Sample Size          Sample Size             Dropouts
                    ──────────────────    ──────────────────    ────────────────
Dropout Rate         N1      N2      N      N1'     N2'     N'      D1      D2      D
─────────────────────────────────────────────────────────────────────────────────────
20%                  10      10     20       13      13     26       3       3      6
20%                  30      30     60       38      38     76       8       8     16
20%                  50      50    100       63      63    126      13      13     26
20%                  70      70    140       88      88    176      18      18     36
─────────────────────────────────────────────────────────────────────────────────────────

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 Lokhnygina, 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, 13 subjects should be enrolled in Group 1, and 13 in Group 2, to obtain final group
sample sizes of 10 and 10, respectively.
─────────────────────────────────────────────────────────────────────────────────────────


References
─────────────────────────────────────────────────────────────────────────────────────────
Blackwelder, W.C. 1998. 'Equivalence Trials.'  In Encyclopedia of Biostatistics, John Wiley and Sons. New York.
   Volume 2, 1367-1372.
Chow, S.C., Shao, J., and Wang, H. 2003. Sample Size Calculations in Clinical Research. Marcel Dekker. New
   York.
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, Mass.
Matsumoto, M. and Nishimura,T. 1998. 'Mersenne twister: A 623-dimensionally equidistributed uniform
   pseudorandom number generator.'  ACM Trans. On Modeling and Computer Simulations.
─────────────────────────────────────────────────────────────────────────────────────────
```

© NCSS, LLC. All Rights Reserved.

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 2** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

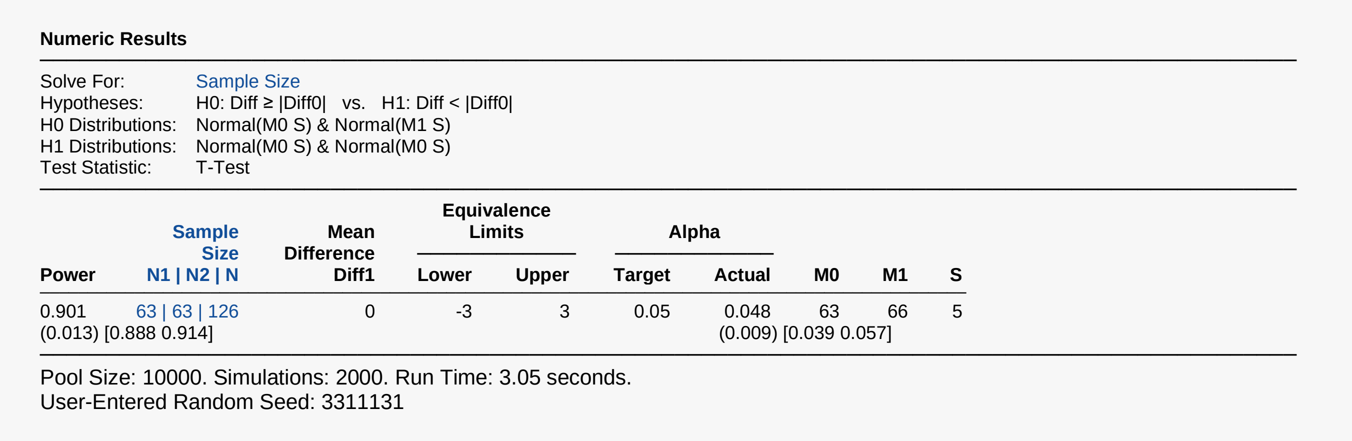
Design Tab

Solve For	**Sample Size**
Equivalence Limit	**Symmetric**
Test Type	**T-Test**
Simulations	**2000**
Random Seed	**3311131** (for Reproducibility)
Power	**0.90**
Alpha	**0.05**
Group Allocation	**Equal (N1 = N2)**
Group 1 Distribution|H0	**Normal(M0 S)**
Group 1 Distribution|H1	**Normal(M0 S)**
Group 2 Distribution|H0	**Normal(M1 S)**
Group 2 Distribution|H1	**Normal(M0 S)**
M0 (Mean|H0) Parameter Value(s)	**63**
M1 (Mean|H1) Parameter Value(s)	**66**
Parameter 1 Label	**S**
Parameter 1 Value(s)	**5**

© NCSS, LLC. All Rights Reserved.

Output

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



```
Numeric Results
─────────────────────────────────────────────────────────────────────────────────────────
Solve For:          Sample Size
Hypotheses:         H0: Diff ≥ |Diff0|   vs.   H1: Diff < |Diff0|
H0 Distributions:   Normal(M0 S) & Normal(M1 S)
H1 Distributions:   Normal(M0 S) & Normal(M0 S)
Test Statistic:     T-Test
─────────────────────────────────────────────────────────────────────────────────────────
                                        Equivalence
                Sample        Mean         Limits               Alpha
                  Size  Difference  ──────────────────  ─────────────────
Power      N1 | N2 | N        Diff1   Lower      Upper    Target     Actual      M0      M1     S
───────────────────────────────────────────────────────────────────────────────────────
0.901      63 | 63 | 126          0      -3          3      0.05      0.048      63      66     5
(0.013) [0.888 0.914]                                              (0.009) [0.039 0.057]
─────────────────────────────────────────────────────────────────────────────────────────
Pool Size: 10000. Simulations: 2000. Run Time: 3.05 seconds.
User-Entered Random Seed: 3311131
```

The required sample size is 63 per group.

© NCSS, LLC. All Rights Reserved.

Example 3 – Comparative Results when the Data Contain Outliers

Continuing Example 1, this example will investigate the impact of outliers on the characteristics of the various test statistics. The two-sample t-test is known to be robust to the violation of some assumptions, but it is susceptible to inaccuracy when the data contains 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 standard distribution. The other 5% of the data will come from a normal distribution with the same mean but with a standard deviation that is one, five, and ten times larger than that of the standard.

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**
Equivalence Limit	**Symmetric**
Test Type	**T-Test**
Simulations	**2000**
Random Seed	**9879778** (for Reproducibility)
Alpha	**0.05**
Group Allocation	**Equal (N1 = N2)**
Sample Size Per Group	**40**
Group 1 Distribution|H0	**Normal(M0 S)[95];Normal(M0 A)[5]**
Group 1 Distribution|H1	**Normal(M0 S)[95];Normal(M0 A)[5]**
Group 2 Distribution|H0	**Normal(M1 S)[95];Normal(M1 A)[5]**
Group 2 Distribution|H1	**Normal(M0 S)[95];Normal(M0 A)[5]**
M0 (Mean|H0) Parameter Value(s)	**63**
M1 (Mean|H1) Parameter Value(s)	**66**
Parameter 1 Label	**S**
Parameter 1 Value(s)	**5**
Parameter 2 Label	**A**
Parameter 2 Value(s)	**5 25 50**

Reports Tab

Show Comparative Reports	**Checked**

Comparative Plots Tab

Show Comparative Plots	**Checked**

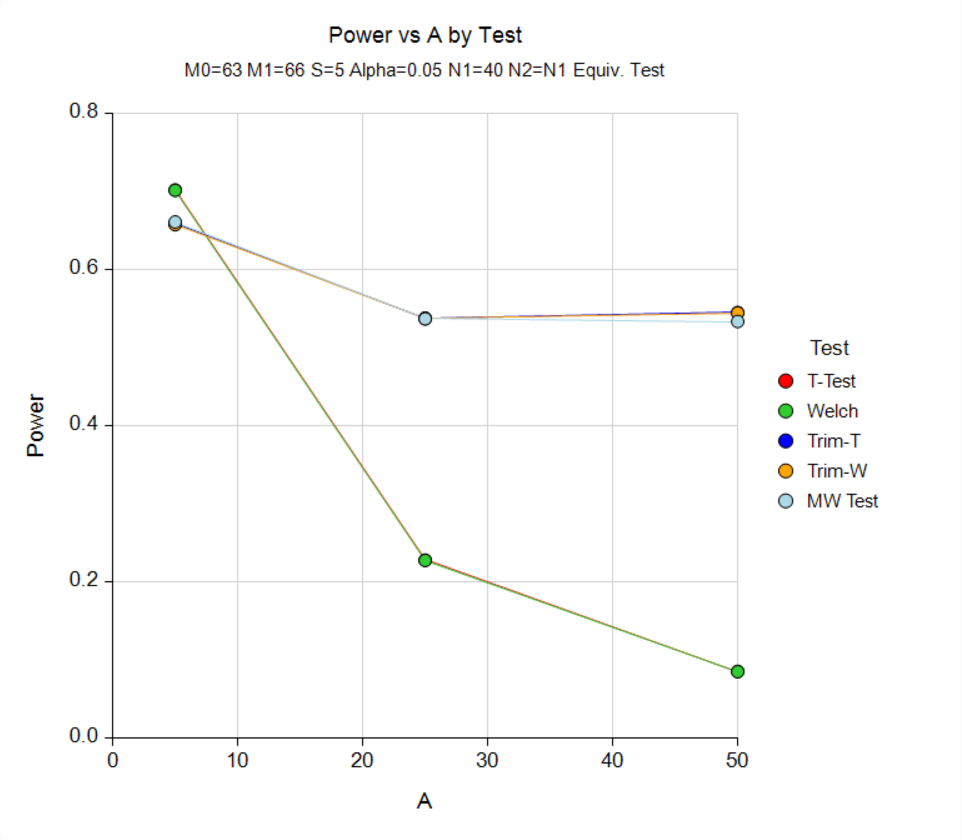
© NCSS, LLC. All Rights Reserved.

Output

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

```
Power Comparison
─────────────────────────────────────────────────────────────────────────────────────────────────────────────
Hypotheses:        H0: Diff ≥ |Diff0|   vs.   H1: Diff < |Diff0|
H0 Distributions:  Normal(M0 S)[95];Normal(M0 A)[5] & Normal(M1 S)[95];Normal(M1 A)[5]
H1 Distributions:  Normal(M0 S)[95];Normal(M0 A)[5] & Normal(M0 S)[95];Normal(M0 A)[5]
─────────────────────────────────────────────────────────────────────────────────────────────────────────────
                             Equivalence                                 Power
Sample              Mean       Limits                ─────────────────────────────────────────────────
Size          Difference  ──────────────                                 Trimmed   Trimmed    Mann-
N1 | N2 | N        Diff1   Lower   Upper   Alpha   T-Test   Welch         T-Test     Welch  Whitney   M0   M1   S    A
──────────────────────────────────────────────────────────────────────────────────────────────────────────────────────
40 | 40 | 80           0      -3       3    0.05    0.702   0.702          0.659     0.658    0.661   63   66   5    5
40 | 40 | 80           0      -3       3    0.05    0.228   0.228          0.538     0.537    0.537   63   66   5   25
40 | 40 | 80           0      -3       3    0.05    0.086   0.085          0.545     0.545    0.534   63   66   5   50
─────────────────────────────────────────────────────────────────────────────────────────────────────────────
Pool Size: 10000. Simulations: 2000. Run Time: 15.71 seconds. Percent Trimmed at each end: 10.
User-Entered Random Seed: 9879778


Alpha Comparison
─────────────────────────────────────────────────────────────────────────────────────────────────────────────
Hypotheses:        H0: Diff ≥ |Diff0|   vs.   H1: Diff < |Diff0|
H0 Distributions:  Normal(M0 S)[95];Normal(M0 A)[5] & Normal(M1 S)[95];Normal(M1 A)[5]
H1 Distributions:  Normal(M0 S)[95];Normal(M0 A)[5] & Normal(M0 S)[95];Normal(M0 A)[5]
─────────────────────────────────────────────────────────────────────────────────────────────────────────────
                             Equivalence                              Alpha
Sample              Mean       Limits        ────────────────────────────────────────────────────────────
Size          Difference  ──────────────                                  Trimmed   Trimmed    Mann-
N1 | N2 | N        Diff1   Lower   Upper   Target   T-Test   Welch         T-Test     Welch  Whitney   M0   M1   S    A
──────────────────────────────────────────────────────────────────────────────────────────────────────────────────────
40 | 40 | 80           0      -3       3     0.05    0.055   0.055          0.060     0.060    0.060   63   66   5    5
40 | 40 | 80           0      -3       3     0.05    0.029   0.029          0.063     0.062    0.061   63   66   5   25
40 | 40 | 80           0      -3       3     0.05    0.014   0.014          0.059     0.059    0.061   63   66   5   50
─────────────────────────────────────────────────────────────────────────────────────────────────────────────
Pool Size: 10000. Simulations: 2000. Run Time: 15.71 seconds. Percent Trimmed at each end: 10.
User-Entered Random Seed: 9879778
```

Comparative Plots



When A = 5, there are no outliers and the power of the nonparametric test, and the trimmed tests are a little less than that of the t-test. When A = 25, the distortion of the t-test caused by the outliers becomes apparent. In this case, the powers of the standard t-test and Welch's t-test are 0.228, but the powers of the nonparametric Mann-Whitney test and the trimmed tests are about 0.537. When A = 50, the standard t-test only achieves a power of 0.086, but the trimmed and nonparametric tests achieve powers of about 0.54.

Looking at the second table, we see that the true significance level of the t-test is distorted by the outliers, while the significance levels of the other tests remain close to the target value.

© NCSS, LLC. All Rights Reserved.

Example 4 – Selecting a Test Statistic when the Data Are Skewed

Continuing Example 3, this example will investigate the impact of skewness in the underlying distribution on the characteristics of the various test statistics.

Tukey's G-H 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 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**
Equivalence Limit	**Symmetric**
Test Type	**T-Test**
Simulations	**2000**
Random Seed	**4450651** (for Reproducibility)
Alpha	**0.05**
Group Allocation	**Equal (N1 = N2)**
Sample Size Per Group	**40**
Group 1 Distribution|H0	**TukeyGH(M0 S G 0)**
Group 1 Distribution|H1	**TukeyGH(M0 S G 0)**
Group 2 Distribution|H0	**TukeyGH(M1 S G 0)**
Group 2 Distribution|H1	**TukeyGH(M0 S G 0)**
M0 (Mean|H0) Parameter Value(s)	**63**
M1 (Mean|H1) Parameter Value(s)	**66**
Parameter 1 Label	**S**
Parameter 1 Value(s)	**5**
Parameter 2 Label	**G** (note that A has been changed to G)
Parameter 2 Value(s)	**0 0.5 0.9**

Reports Tab

Show Comparative Reports	**Checked**

Comparative Plots Tab

Show Comparative Plots	**Checked**

Output

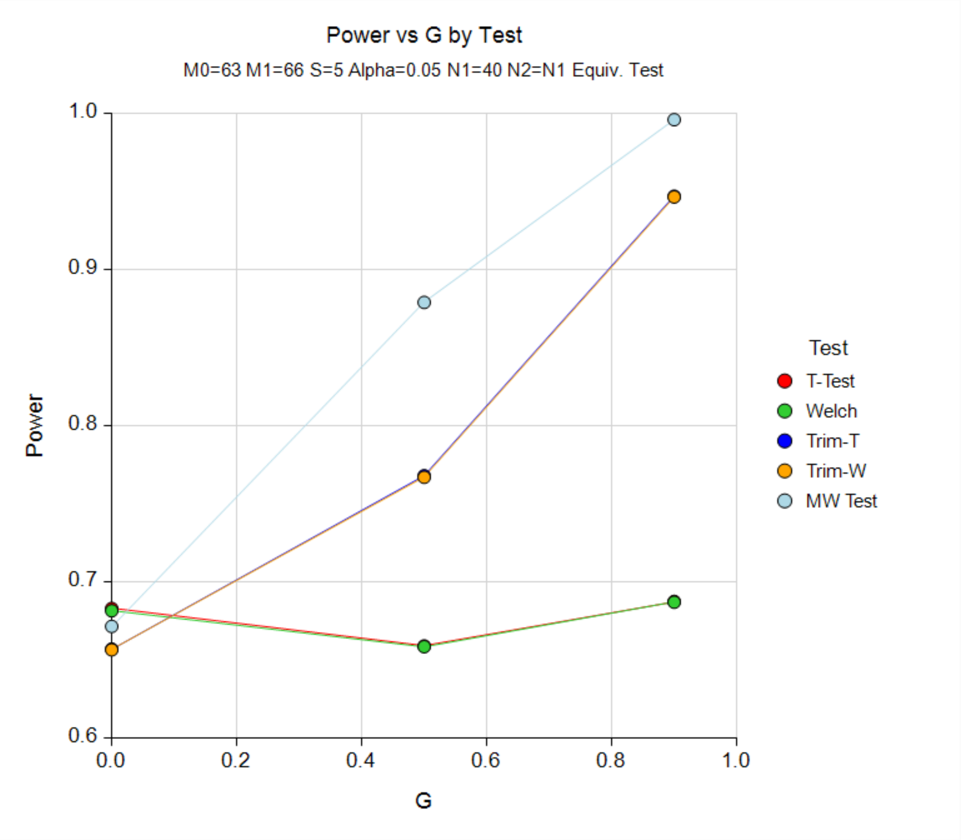
Click the Calculate button to perform the calculations and generate the following output.

```
Power Comparison
─────────────────────────────────────────────────────────────────────────────────────────────────────────────
Hypotheses:        H0: Diff ≥ |Diff0|   vs.   H1: Diff < |Diff0|
H0 Distributions:  TukeyGH(M0 S G 0) & TukeyGH(M1 S G 0)
H1 Distributions:  TukeyGH(M0 S G 0) & TukeyGH(M0 S G 0)
─────────────────────────────────────────────────────────────────────────────────────────────────────────────
                             Equivalence                                 Power
Sample           Mean           Limits                ─────────────────────────────────────────
Size       Difference    ─────────────                                 Trimmed    Trimmed     Mann-
N1 | N2 | N      Diff1    Lower    Upper   Alpha   T-Test    Welch      T-Test      Welch    Whitney    M0    M1    S      G
────────────────────────────────────────────────────────────────────────────────────────────────────────────────────
40 | 40 | 80          0       -3        3    0.05    0.683    0.682       0.657      0.657      0.672    63    66    5    0.0
40 | 40 | 80          0       -3        3    0.05    0.659    0.659       0.768      0.767      0.879    63    66    5    0.5
40 | 40 | 80          0       -3        3    0.05    0.688    0.687       0.947      0.947      0.996    63    66    5    0.9
─────────────────────────────────────────────────────────────────────────────────────────────────────────────
Pool Size: 10000. Simulations: 2000. Run Time: 16.46 seconds. Percent Trimmed at each end: 10.
User-Entered Random Seed: 4450651
```

```
Alpha Comparison
─────────────────────────────────────────────────────────────────────────────────────────────────────────────
Hypotheses:        H0: Diff ≥ |Diff0|   vs.   H1: Diff < |Diff0|
H0 Distributions:  TukeyGH(M0 S G 0) & TukeyGH(M1 S G 0)
H1 Distributions:  TukeyGH(M0 S G 0) & TukeyGH(M0 S G 0)
─────────────────────────────────────────────────────────────────────────────────────────────────────────────
                             Equivalence                                 Alpha
Sample           Mean           Limits          ──────────────────────────────────────────────────
Size       Difference    ─────────────                                  Trimmed    Trimmed     Mann-
N1 | N2 | N      Diff1    Lower    Upper   Target   T-Test    Welch      T-Test      Welch    Whitney    M0    M1    S      G
────────────────────────────────────────────────────────────────────────────────────────────────────────────────────
40 | 40 | 80          0       -3        3     0.05    0.047    0.047       0.051      0.051      0.049    63    66    5    0.0
40 | 40 | 80          0       -3        3     0.05    0.057    0.057       0.051      0.051      0.051    63    66    5    0.5
40 | 40 | 80          0       -3        3     0.05    0.060    0.060       0.066      0.065      0.067    63    66    5    0.9
─────────────────────────────────────────────────────────────────────────────────────────────────────────────
Pool Size: 10000. Simulations: 2000. Run Time: 16.46 seconds. Percent Trimmed at each end: 10.
User-Entered Random Seed: 4450651
```

© NCSS, LLC. All Rights Reserved.

Comparative Plots



We see that as the degree of skewness is increased, the power of the t-test increases slightly, but the powers of the trimmed and nonparametric tests improve dramatically. The significance levels do not appear to be adversely impacted.

Example 5 – Validation using Machin et al. (1997)

Machin *et al.* (1997) page 107 present an example of determining the sample size for a parallel-group design in which the reference mean is 96, the treatment mean is 94, the standard deviation is 8, the limits are plus or minus 5, the power is 80%, and the significance level is 0.05. They calculate the sample size to be 88. It is important to note that Machin *et al.* use an approximation, so their results cannot be expected to exactly match those of **PASS**.

For reproducibility, we'll use a random seed of 5067146.

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.

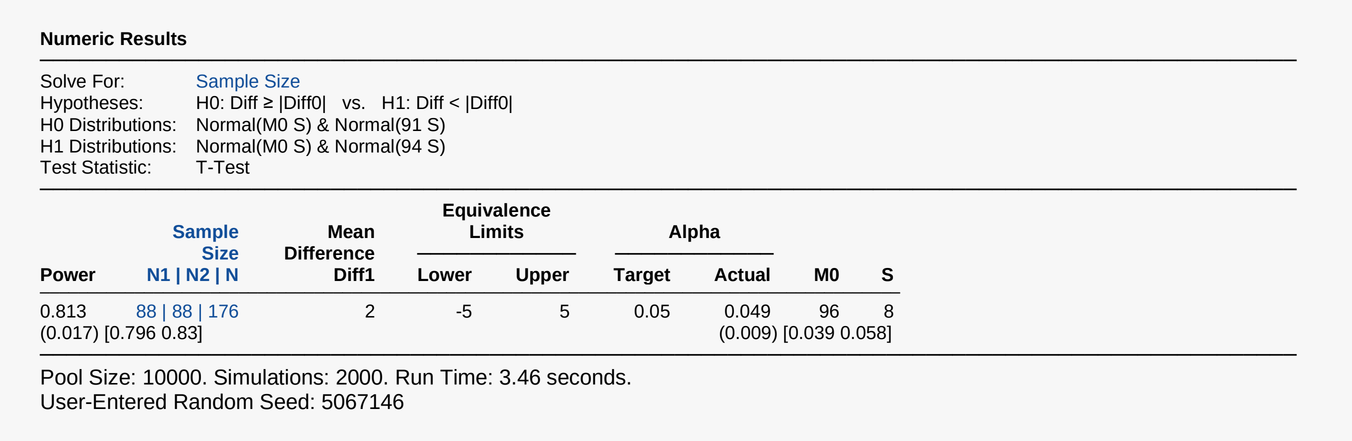
Design Tab

Solve For	**Sample Size**
Equivalence Limit	**Symmetric**
Test Type	**T-Test**
Simulations	**2000**
Random Seed	**5067146** (for Reproducibility)
Power	**0.80**
Alpha	**0.05**
Group Allocation	**Equal (N1 = N2)**
Group 1 Distribution|H0	**Normal(M0 S)**
Group 1 Distribution|H1	**Normal(M0 S)**
Group 2 Distribution|H0	**Normal(91 S)**
Group 2 Distribution|H1	**Normal(94 S)**
M0 (Mean|H0) Parameter Value(s)	**96**
M1 (Mean|H1) Parameter Value(s)	**1**
Parameter 1 Label	**S**
Parameter 1 Value(s)	**8**

© NCSS, LLC. All Rights Reserved.

Output

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



```
Numeric Results
──────────────────────────────────────────────────────────────────────────────────────────────
Solve For:           Sample Size
Hypotheses:          H0: Diff ≥ |Diff0|   vs.   H1: Diff < |Diff0|
H0 Distributions:    Normal(M0 S) & Normal(91 S)
H1 Distributions:    Normal(M0 S) & Normal(94 S)
Test Statistic:      T-Test
──────────────────────────────────────────────────────────────────────────────────────────────
                                        Equivalence
                Sample         Mean        Limits              Alpha
                  Size   Difference   ─────────────    ───────────────
Power      N1 | N2 | N        Diff1   Lower   Upper    Target   Actual     M0     S
──────────────────────────────────────────────────────────────────────────────────
0.813      88 | 88 | 176          2      -5       5      0.05    0.049     96     8
(0.017) [0.796 0.83]                                           (0.009) [0.039 0.058]
──────────────────────────────────────────────────────────────────────────────────────────────
Pool Size: 10000. Simulations: 2000. Run Time: 3.46 seconds.
User-Entered Random Seed: 5067146
```

The sample size of 88 per group is the same as the analytic answer of 88.