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. 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 t-test is often used. When the data are not normally distributed, the Mann-Whitney (Wilcoxon signed-ranks) U 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 <u>alternative</u> 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 <u>null</u> 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.

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.

Test Statistics

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

Two-Sample T-Test

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

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

where

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

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

$$df = N_1 + N_2 - 2$$

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.

Welch's T-Test

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

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

where

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

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

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

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 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

$$\overline{X}_{tg} = \frac{\sum\limits_{k=g+1}^{N-g} X_k}{h}$$

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.

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

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

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

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

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

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 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

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

with degrees of freedom f given by

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

where

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

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 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).

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

where

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

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

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

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!

Procedure Options

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

Design Tab

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

Solve For

Solve For

This option specifies whether you want to find *Power* or *Sample Size (N1)* from the simulation. Select *Power* when you want to estimate the power of a certain scenario. Select *Sample Size (N1)* when you want to determine the sample size needed to achieve a given power and alpha error level. Finding *Sample Size (N1)* is very computationally intensive, and so it may take a long time to complete.

Test and Simulations

Alternative Hypothesis

This option specifies the alternative hypothesis, H1. This implicitly specifies the direction of the hypothesis test. With Diff = $\mu_1 - \mu_2$, the null hypothesis is always H0: Diff = Diff0.

Note that the alternative hypothesis enters into power calculations by specifying the rejection region of the hypothesis test. Its accuracy is critical.

Tests for Two Means (Simulation)

Possible selections are:

- **Difference ≠ Diff0**

    This is the most common selection. It yields a *two-tailed test*. Use this option when you are testing whether the true difference is different from a specified value Diff0, but you do not want to specify beforehand whether it is smaller or larger. Most scientific journals require two-tailed tests.

- **Difference < Diff0**

    This option yields a *one-tailed test*. Use it when you want to test whether the true difference is less than Diff0.

- **Difference > Diff0**

    This option yields a *one-tailed test*. Use it when you want to test whether the true difference is greater than Diff0. Note that this option could be used for a **non-inferiority test**.

Test Type

Specify which test statistic is to be used in the simulation. Although the t-test is the most commonly used test statistic, it is based on assumptions that may not be viable in many situations. For your data, you may find that one of the other tests is more accurate (actual alpha = target alpha) and more precise (better power).

Simulations

This option specifies the number of iterations, *M*, used in the simulation. As the number of iterations is increased, the accuracy and running time of the simulation will be increased also.

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

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

Notice that a simulation size of 1000 gives a precision of plus or minus 0.01 when the true power is 0.95. Also note that as the simulation size is increased beyond 5000, there is only a small amount of additional accuracy achieved.

Power and Alpha

Power

This option specifies one or more values for power. Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected. In this procedure, a type-II error occurs when you fail to reject the null hypothesis of equal means when in fact the means are different.

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

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

Alpha

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected. In this procedure, a type-I error occurs when you reject the null hypothesis of equal means when in fact the means are equal.

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

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

Sample Size (When Solving for Sample Size)

Group Allocation

Select the option that describes the constraints on *N1* or *N2* or both.

The options are

- **Equal (N1 = N2)**

   This selection is used when you wish to have equal sample sizes in each group. Since you are solving for both sample sizes at once, no additional sample size parameters need to be entered.

- **Enter N2, solve for N1**

   Select this option when you wish to fix *N2* at some value (or values), and then solve only for *N1*. Please note that for some values of *N2*, there may not be a value of *N1* that is large enough to obtain the desired power.

- **Enter R = N2/N1, solve for N1 and N2**

   For this choice, you set a value for the ratio of *N2* to *N1*, and then PASS determines the needed *N1* and *N2*, with this ratio, to obtain the desired power. An equivalent representation of the ratio, *R*, is

$$N2 = R * N1.$$

- **Enter percentage in Group 1, solve for N1 and N2**

   For this choice, you set a value for the percentage of the total sample size that is in Group 1, and then PASS determines the needed *N1* and *N2* with this percentage to obtain the desired power.

N2 (Sample Size, Group 2)

This option is displayed if Group Allocation = "Enter N2, solve for N1"

N2 is the number of items or individuals sampled from the Group 2 population.

N2 must be ≥ 2. You can enter a single value or a series of values.

R (Group Sample Size Ratio)

This option is displayed only if Group Allocation = "Enter R = N2/N1, solve for N1 and N2."

R is the ratio of *N2* to *N1*. That is,

$$R = N2 / N1.$$

Use this value to fix the ratio of *N2* to *N1* while solving for *N1* and *N2*. Only sample size combinations with this ratio are considered.

N2 is related to *N1* by the formula:

$$N2 = [R \times N1],$$

where the value *[Y]* is the next integer $\geq Y$.

For example, setting $R = 2.0$ results in a Group 2 sample size that is double the sample size in Group 1 (e.g., *N1* = 10 and *N2* = 20, or *N1* = 50 and *N2* = 100).

R must be greater than 0. If $R < 1$, then *N2* will be less than *N1*; if $R > 1$, then *N2* will be greater than *N1*. You can enter a single or a series of values.

Percent in Group 1

This option is displayed only if Group Allocation = "Enter percentage in Group 1, solve for N1 and N2."

Use this value to fix the percentage of the total sample size allocated to Group 1 while solving for *N1* and *N2*. Only sample size combinations with this Group 1 percentage are considered. Small variations from the specified percentage may occur due to the discrete nature of sample sizes.

The Percent in Group 1 must be greater than 0 and less than 100. You can enter a single or a series of values.

Sample Size (When <u>Not</u> Solving for Sample Size)

Group Allocation

Select the option that describes how individuals in the study will be allocated to Group 1 and to Group 2.

The options are

- **Equal (N1 = N2)**

   This selection is used when you wish to have equal sample sizes in each group. A single per group sample size will be entered.

- **Enter N1 and N2 individually**

   This choice permits you to enter different values for *N1* and *N2*.

- **Enter N1 and R, where N2 = R * N1**

   Choose this option to specify a value (or values) for *N1*, and obtain *N2* as a ratio (multiple) of *N1*.

- **Enter total sample size and percentage in Group 1**

   Choose this option to specify a value (or values) for the total sample size (*N*), obtain *N1* as a percentage of *N*, and then *N2* as *N - N1*.

Sample Size Per Group

This option is displayed only if Group Allocation = "Equal (N1 = N2)."

The Sample Size Per Group is the number of items or individuals sampled from each of the Group 1 and Group 2 populations. Since the sample sizes are the same in each group, this value is the value for *N1*, and also the value for *N2*.

The Sample Size Per Group must be ≥ 2. You can enter a single value or a series of values.

N1 (Sample Size, Group 1)

This option is displayed if Group Allocation = "Enter N1 and N2 individually" or "Enter N1 and R, where N2 = R * N1."

N1 is the number of items or individuals sampled from the Group 1 population.

N1 must be ≥ 2. You can enter a single value or a series of values.

N2 (Sample Size, Group 2)

This option is displayed only if Group Allocation = "Enter N1 and N2 individually."

N2 is the number of items or individuals sampled from the Group 2 population.

N2 must be ≥ 2. You can enter a single value or a series of values.

R (Group Sample Size Ratio)

*This option is displayed only if Group Allocation = "Enter N1 and R, where N2 = R * N1."*

R is the ratio of *N2* to *N1*. That is,

$$R = N2/N1$$

Use this value to obtain *N2* as a multiple (or proportion) of *N1*.

N2 is calculated from *N1* using the formula:

$$N2=[R \times N1],$$

where the value *[Y]* is the next integer $\geq Y$.

For example, setting *R* = 2.0 results in a Group 2 sample size that is double the sample size in Group 1.

R must be greater than 0. If *R* < 1, then *N2* will be less than N1; if *R* > 1, then *N2* will be greater than *N1*. You can enter a single value or a series of values.

Total Sample Size (N)

This option is displayed only if Group Allocation = "Enter total sample size and percentage in Group 1."

This is the total sample size, or the sum of the two group sample sizes. This value, along with the percentage of the total sample size in Group 1, implicitly defines *N1* and *N2*.

The total sample size must be greater than one, but practically, must be greater than 3, since each group sample size needs to be at least 2.

You can enter a single value or a series of values.

Percent in Group 1

This option is displayed only if Group Allocation = "Enter total sample size and percentage in Group 1."

This value fixes the percentage of the total sample size allocated to Group 1. Small variations from the specified percentage may occur due to the discrete nature of sample sizes.

The Percent in Group 1 must be greater than 0 and less than 100. You can enter a single value or a series of values.

Effect Size

Group 1 & 2 Distributions|H0 (Null)

These options specify the distributions of the two groups under the null hypothesis, H0. The difference between the means of these two distributions is the difference that is tested, *Diff0*.

Usually, these two distributions will be identical and *Diff0* = 0. However, if you are planning a non-inferiority test, the means will be different.

Several of the distributions are parameterized so that the mean is entered first. For example, if you wanted to specify that the mean of a normally-distributed variable is to be five, you could enter Normal(5, S) or Normal(M0, S) here and *M0* = 5 below.

The parameters of each distribution are specified using numbers or letters. If letters are used, their values are specified in the boxes below. The value *M0* is reserved for the value of the mean under the null hypothesis.

The parameters of each distribution are specified using numbers or letters. If letters are used, their values are specified in the boxes below. The value *M0* is reserved for the value of the mean under the null hypothesis.

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

Distributions with Common Parameters

Beta(Shape1, Shape2, Min, Max)

Binomial(P, N)

Cauchy(Mean, Scale)

Constant(Value)

Exponential(Mean)

Gamma(Shape, Scale)

Gumbel(Location, Scale)

Laplace(Location, Scale)

Logistic(Location, Scale)

Lognormal(Mu, Sigma)

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

Normal(Mean, Sigma)

Poisson(Mean)

TukeyGH(Mu, S, G, H)

Uniform(Min, Max)

Weibull(Shape, Scale)

Distributions with Mean and SD Parameters

BetaMS(Mean, SD, Min, Max)

BinomialMS(Mean, N)

GammaMS(Mean, SD)

GumbelMS(Mean, SD)

LaplaceMS(Mean, SD)

LogisticMS(Mean, SD)

LognormalMS(Mean, SD)

UniformMS(Mean, SD)

WeibullMS(Mean, SD)

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

Finding the Value of the Mean of a Specified Distribution

Except for the multinomial distribution, the distributions have been parameterized in terms of their means, since this is the parameter being tested. The mean of a distribution created as a linear combination of other distributions is found by applying the linear combination to the individual means. However, the mean of a distribution created

by multiplying or dividing other distributions is not necessarily equal to applying the same function to the individual means. For example, the mean of 4 Normal(4, 5) + 2 Normal(5, 6) is 4*4 + 2*5 = 26, but the mean of 4 Normal(4, 5) * 2 Normal(5, 6) is not exactly 4*4*2*5 = 160 (although it is close).

Group 1 & 2 Distributions|H1 (Alternative)

These options specify the distributions of the two groups under the alternative hypothesis, H1. The difference between the means of these two distributions is the difference that is assumed to be the true value of the difference. That is, this is the difference at which the power is computed.

Usually, the mean difference is specified by entering *M0* for the mean parameter in the distribution expression for group 1 and *M1* for the mean parameter in the distribution expression for group 2. The mean difference under H1 then becomes the value of *M0*– *M1*.

The parameters of each distribution are specified using numbers or letters. If letters are used, their values are specified in the boxes below. The value *M1* is reserved for the value of the mean of group 2 under the alternative hypothesis.

A list of the distributions that are available and the syntax used to specify them is given above.

Effect Size – Distribution Parameters

M0 (Mean|H0)

These values are substituted for *M0* in the distribution specifications given above. *M0* is intended to be the value of the mean hypothesized by the null hypothesis, H0.

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

M1 (Mean|H1)

These values are substituted for *M1* in the distribution specifications given above. Although it can be used wherever you want, *M1* is intended to be the value of the mean hypothesized by the alternative hypothesis, H1.

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

Parameter Values (S, A, B)

Enter the numeric value(s) of the parameters listed above. These values are substituted for the corresponding letter in all four distribution specifications.

You can enter a list of values for each letter using the syntax *0 1 2 3* or *0 to 3 by 1*.

You can also change the letter that is used as the name of this parameter using the pull-down menu to the side.

Options Tab

The Options tab contains calculation options.

Random Numbers

Random Number Pool Size

This is the size of the pool of random values from which the random samples will be drawn. Pools should be at least the maximum of 10,000 and twice the number of simulations. You can enter *Automatic* and an appropriate value will be calculated.

If you do not want to draw numbers from a pool, enter 0 here.

Trimmed T-Test

Percent Trimmed at Each End

Specify the percent of each end of the sorted data that is to be trimmed (constant *G* above) when using the trimmed means procedures. This percentage is applied to the sample size to determine how many of the lowest and highest data values are to be trimmed by the procedure. For example, if the sample size (N1) is 27 and you specify 10 here, then [27*10/100] = 2 observations will be trimmed at the bottom and the top. For any percentage, at least one observation is trimmed from each end of the sorted dataset.

The range of possible values is 0 to 25.

Reports Tab

The Reports tab contains settings about the format of the numeric reports.

Select Numeric Report Output

Show Inc's & 95% C.I.

Checking this option causes an additional line to be printed showing a 95% confidence interval for both the power and actual alpha and half the width of the confidence interval (the increment).

Show Comparative Reports

These options let you specify whether you want to generate reports that compare test statistics.

Comparative Report/Plot Options

Include T-Test Results – Include Mann-Whitney-Test Results

These options let you specify whether to include each test statistic in the comparative reports. These options are only used if comparative reports and/or plots are generated.

Plots Tab

The Plots tab contains settings about the format of the plots.

Plots

Show Comparative Plots

These options let you specify whether you want to generate plots that compare the test statistics that are available.

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

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

Option	Value
Design Tab	
Solve For	**Power**
Alternative Hypothesis............................	**Diff ≠ Diff0**
Test Type...	**T-Test**
Simulations	**2000**
Alpha..	**0.05**
Group Allocation	**Equal (N1 = N2)**
Sample Size Per Group...........................	**50 100 200**
Group 1 Distribution	H0	**Normal(M0 S)**
Group 2 Distribution	H0	**Normal(M0 S)**
Group 1 Distribution	H1	**Normal(M0 S)**
Group 2 Distribution	H1	**Normal(M1 S)**
M0 (Mean|H0)..	**0**
M1 (Mean|H1)..	**0.6**
S..	**2**

Annotated Output

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

Numeric Results and Plots

Numeric Results for Testing Mean Difference = Diff0.   Hypotheses: H0: Diff1 = Diff0; H1: Diff1 ≠ Diff0
H0 Dist's: Normal(M0 S) & Normal(M0 S)
H1 Dist's: Normal(M0 S) & Normal(M1 S)
Test Statistic: T-Test

Power	N1/N2	H0 Diff0	H1 Diff1	Target Alpha	Actual Alpha	Beta	M0	M1	S
0.324	50/50	0.0	-0.6	0.050	0.056	0.676	0.0	0.6	2.0
(0.021)	[0.303	0.345]			(0.010)	[0.045	0.066]		
0.563	100/100	0.0	-0.6	0.050	0.047	0.437	0.0	0.6	2.0
(0.022)	[0.541	0.585]			(0.009)	[0.038	0.056]		
0.855	200/200	0.0	-0.6	0.050	0.045	0.145	0.0	0.6	2.0
(0.015)	[0.840	0.870]			(0.009)	[0.035	0.054]		

Tests for Two Means (Simulation)

Notes
Pool Size: 10000. Simulations: 2000. Run Time: 34.78 seconds.

Report Definitions
Power is the probability of rejecting a false null hypothesis.
N1 is the size of the sample drawn from population 1.
N2 is the size of the sample drawn from population 2.
Diff0 is the mean difference between (Grp1 - Grp2) assuming the null hypothesis, H0.
Diff1 is the mean difference between (Grp1 - Grp2) assuming the alternative hypothesis, H1.
Target Alpha is the probability of rejecting a true null hypothesis. It is set by the user.
Actual Alpha is the alpha level that was actually achieved by the experiment.
Beta is the probability of accepting a false null hypothesis.
Second Row: (Power Prec.) [95% LCL and UCL Power]   (Alpha Prec.) [95% LCL and UCL Alpha]

Summary Statements
Group sample sizes of 50 and 50 achieve 32% power to detect a difference of -0.6 between the
null hypothesis mean difference of 0.0 and the actual mean difference of -0.6 at the 0.050
significance level (alpha) using a two-sided T-Test. These results are based on 2000 Monte
Carlo samples from the null distributions: Normal(M0 S) and Normal(M0 S), and the alternative
distributions: Normal(M0 S) and Normal(M1 S).

Chart Section
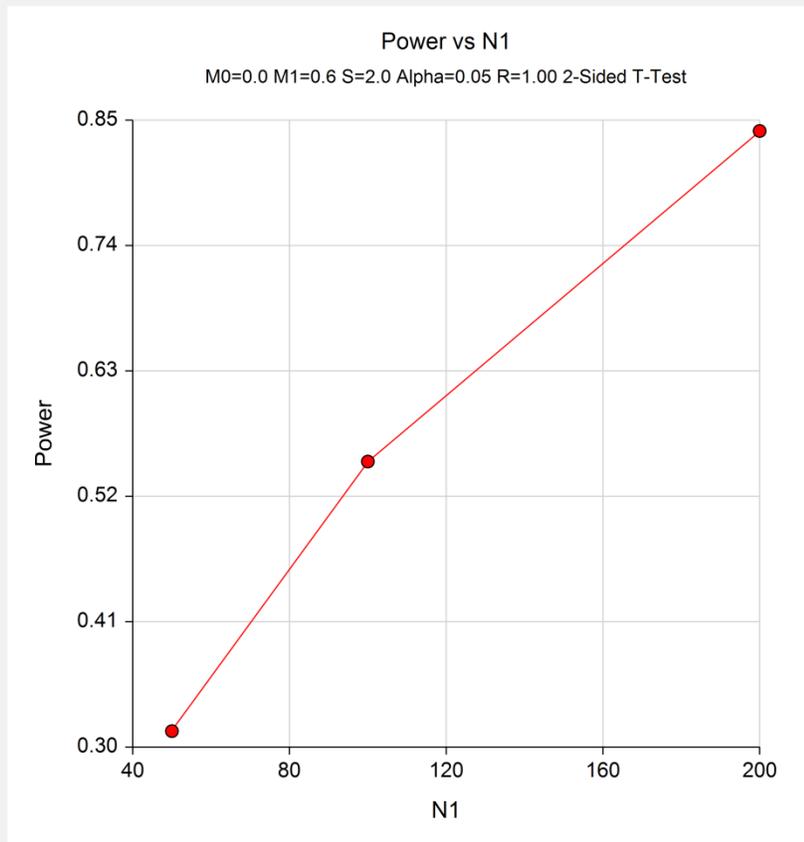
This report shows the estimated power for each scenario. The first row shows the parameter settings and the estimated power and significance level (Actual Alpha).

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 Example1, the researchers want to determine how large a sample is needed to obtain a power of 0.90.

Setup

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

Option	Value
Design Tab	
Solve For	**Sample Size**
Alternative Hypothesis	**Diff ≠ Diff0**
Test Type...	**T-Test**
Simulations	**2000**
Power..	**0.90**
Alpha...	**0.05**
Group Allocation	**Equal (N1 = N2)**
Group 1 Distribution	H0	**Normal(M0 S)**
Group 2 Distribution	H0	**Normal(M0 S)**
Group 1 Distribution	H1	**Normal(M0 S)**
Group 2 Distribution	H1	**Normal(M1 S)**
M0 (Mean|H0)	**0**
M1 (Mean|H1)	**0.6**
S..	**2**

Output

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

Numeric Results of Search for N

Power	N1/N2	H0 Diff0	H1 Diff1	Target Alpha	Actual Alpha	Beta	M0	M1	S
0.904	231/231	0.0	-0.6	0.050	0.053	0.097	0.0	0.6	2.0
(0.013)	[0.891 0.916]				(0.010)	[0.043 0.063]			

The required sample size was 231 which achieved a power of 0.904. To check the accuracy of this simulation, we ran this scenario through the analytic procedure in **PASS** which gave the sample size as 234 per group. The simulation answer of 231 was reasonably close.

Example 3 – Comparative Results

Continuing with Example 2, the researchers want to study the characteristics of alternative test statistics. They want to compare the results of all test statistics for N1 = 50, 100, and 200.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Two Means (Simulation)** procedure window by expanding **Means**, then **Two Independent Means**, then clicking on **Non-Normal Data**, and then clicking on **Tests for Two Means (Simulation)**. You may then make the appropriate entries as listed below, or open **Example 3** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Design Tab	
Solve For	**Power**
Alternative Hypothesis	**Diff ≠ Diff0**
Test Type	**T-Test**
Simulations	**2000**
Alpha	**0.05**
Group Allocation	**Equal (N1 = N2)**
Sample Size Per Group	**50 100 200**
Group 1 Distribution | H0	**Normal(M0 S)**
Group 2 Distribution | H0	**Normal(M0 S)**
Group 1 Distribution | H1	**Normal(M0 S)**
Group 2 Distribution | H1	**Normal(M1 S)**
M0 (Mean|H0)	**0**
M1 (Mean|H1)	**0.6**
S	**2**
Reports Tab	
Show Comparative Reports	**Checked**
Plots Tab	
Show Comparative Plots	**Checked**

Output

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

Numeric Results

Power Comparison for Testing Mean Difference = Diff0.   Hypotheses: H0: Diff1 = Diff0; H1: Diff1 ≠ Diff0
H0 Dist's: Normal(M0 S) & Normal(M0 S)
H1 Dist's: Normal(M0 S) & Normal(M1 S)

N1/N2	H0 Diff (Diff0)	H1 Diff (Diff1)	Target Alpha	T-Test Power	Welch Power	Trim. T-Test Power	Trim. Welch Power	Mann Whit'y Power	M0	M1	S
50/50	0.0	-0.6	0.050	0.304	0.303	0.283	0.283	0.288	0.0	0.6	2.0
100/100	0.0	-0.6	0.050	0.577	0.577	0.538	0.538	0.544	0.0	0.6	2.0
200/200	0.0	-0.6	0.050	0.859	0.859	0.848	0.848	0.850	0.0	0.6	2.0

Pool Size: 10000. Simulations: 2000. Run Time: 81.03 seconds. Percent Trimmed at each end: 10.

Tests for Two Means (Simulation)

Alpha Comparison for Testing Mean Difference = Diff0.   Hypotheses: H0: Diff1 = Diff0; H1: Diff1 ≠ Diff0
H0 Dist's: Normal(M0 S) & Normal(M0 S)
H1 Dist's: Normal(M0 S) & Normal(M1 S)

	H0 Diff (Diff0)	H1 Diff (Diff1)	Target Alpha	T-Test Alpha	Welch Alpha	Trim. T-Test Alpha	Trim. Welch Alpha	Mann Whit'y Alpha			
N1/N2									**M0**	**M1**	**S**
50/50	0.0	-0.6	0.050	0.048	0.047	0.048	0.048	0.045	0.0	0.6	2.0
100/100	0.0	-0.6	0.050	0.048	0.048	0.049	0.049	0.048	0.0	0.6	2.0
200/200	0.0	-0.6	0.050	0.054	0.054	0.054	0.054	0.053	0.0	0.6	2.0

Pool Size: 10000. Simulations: 2000. Run Time: 81.03 seconds. Percent Trimmed at each end: 10.

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, when N1 = 50, are only 7% less than that of the t-test, even though about 20% of the data were trimmed.

Example 4 – Validation using Zar

Zar (1984) page 136 give 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.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Two Means (Simulation)** procedure window by expanding **Means**, then **Two Independent Means**, then clicking on **Non-Normal Data**, and then clicking on **Tests for Two Means (Simulation)**. You may then make the appropriate entries as listed below, or open **Example 4** by going to the **File** menu and choosing **Open Example Template**.

Option   **Value**

Design Tab
Solve For .... **Power**
Alternative Hypothesis .... **Diff ≠ Diff0**
Test Type... **T-Test**
Simulations **10000**
Alpha... **0.05**
Group Allocation .... **Equal (N1 = N2)**
Sample Size Per Group.......................... **15**
Group 1 Distribution | H0 .... **Normal(M0 S)**
Group 2 Distribution | H0 .... **Normal(M0 S)**
Group 1 Distribution | H1 .... **Normal(M0 S)**
Group 2 Distribution | H1 .... **Normal(M1 S)**
M0 (Mean|H0).. **0**
M1 (Mean|H1) **1**
S... **0.7206**

Output

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

Numeric Results

Numeric Results for Testing Mean Difference = Diff0.   Hypotheses: H0: Diff1 = Diff0; H1: Diff1 ≠ Diff0
H0 Dist's: Normal(M0 S) & Normal(M0 S)
H1 Dist's: Normal(M0 S) & Normal(M1 S)
Test Statistic: T-Test

		H0	H1	Target	Actual				
Power	**N1/N2**	**Diff0**	**Diff1**	**Alpha**	**Alpha**	**Beta**	**M0**	**M1**	**S**
0.956	15/15	0.0	-1.0	0.050	0.045	0.044	0.0	1.0	0.7
(0.004)	[0.952	0.960]			(0.004)	[0.041	0.049]		

Notes:
Pool Size: 20000. Simulations: 10000. Run Time: 10.14 seconds.

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

$$H_0: \mu_N - \mu_S \geq 4 \text{ versus } H_1: \mu_N - \mu_S < 4$$

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

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Two Means (Simulation)** procedure window by expanding **Means**, then **Two Independent Means**, then clicking on **Non-Normal Data**, and then clicking on **Tests for Two Means (Simulation)**. You may then make the appropriate entries as listed below, or open **Example 5** by going to the **File** menu and choosing **Open Example Template**.

Option   **Value**

Design Tab
Solve For .... **Sample Size**
Alternative Hypothesis .... **Diff < Diff0**
Test Type .... **T-Test**
Simulations **2000**
Power **0.90**
Alpha **0.025**

Tests for Two Means (Simulation)

Design Tab (continued)

Group Allocation .... **Equal (N1 = N2)**
Group 1 Distribution | H0 .... **Normal(M1 S)**
Group 2 Distribution | H0 .... **Normal(M0 S)**
Group 1 Distribution | H1 .... **Normal(M0 S)**
Group 2 Distribution | H1 .... **Normal(M0 S)**
M0 (Mean|H0) **96**
M1 (Mean|H1) **100**
S... **6**

Output

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

Numeric Results

Power	N1/N2	H0 Diff0	H1 Diff1	Target Alpha	Actual Alpha	Beta	M0	M1	S
0.918	49/49	4.0	0.0	0.025	0.024	0.083	96.0	100.0	6.0

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

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Two Means (Simulation)** procedure window by expanding **Means**, then **Two Independent Means**, then clicking on **Non-Normal Data**, and then clicking on **Tests for Two Means (Simulation)**. You may then make the appropriate entries as listed below, or open **Example 6** by going to the **File** menu and choosing **Open Example Template**.

Option   **Value**

Design Tab
Solve For .... **Power**
Alternative Hypothesis .... **Diff ≠ Diff0**
Test Type... **T-Test**
Simulations **2000**
Alpha.. **0.05**
Group Allocation .... **Equal (N1 = N2)**
Sample Size Per Group **20**

Tests for Two Means (Simulation)

Design Tab (continued)

Group 1 Distribution | H0 .... **Normal(M0 S)[95];Normal(M0 A)[5]**
Group 2 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 | H1 .... **Normal(M1 S)[95];Normal(M1 A)[5]**
M0 (Mean|H0) **0**
M1 (Mean|H1) **1**
S... **1**
A... **1 5 10**

Reports Tab

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

Plots Tab

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

Output

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

Numeric Results

Power Comparison for Testing Mean Difference = Diff0.   Hypotheses: H0: Diff1 = Diff0; H1: Diff1 ≠ Diff0
H0 Dist's: Normal(M0 S)[95];Normal(M0 A)[5] & Normal(M0 S)[95];Normal(M0 A)[5]
H1 Dist's: Normal(M0 S)[95];Normal(M0 A)[5] & Normal(M1 S)[95];Normal(M1 A)[5]

	H0 Diff	H1 Diff	Target	T-Test	Welch	Trim. T-Test	Trim. Welch	Mann Whit'y				
N1/N2	(Diff0)	(Diff1)	Alpha	Power	Power	Power	Power	Power	M0	M1	S	A
20/20	0.0	-1.0	0.050	0.865	0.864	0.835	0.835	0.841	0.0	1.0	1.0	1.0
20/20	0.0	-1.0	0.050	0.638	0.637	0.789	0.787	0.781	0.0	1.0	1.0	5.0
20/20	0.0	-1.0	0.050	0.469	0.463	0.778	0.775	0.776	0.0	1.0	1.0	10.0

Pool Size: 10000. Simulations: 2000. Run Time: 1.77 minutes. Percent Trimmed: 10.

Alpha Comparison for Testing Mean Difference = Diff0.   Hypotheses: H0: Diff1 = Diff0; H1: Diff1 ≠ Diff0
H0 Dist's: Normal(M0 S)[95];Normal(M0 A)[5] & Normal(M0 S)[95];Normal(M0 A)[5]
H1 Dist's: Normal(M0 S)[95];Normal(M0 A)[5] & Normal(M1 S)[95];Normal(M1 A)[5]

	H0 Diff	H1 Diff	Target	T-Test	Welch	Trim. T-Test	Trim. Welch	Mann Whit'y				
N1/N2	(Diff0)	(Diff1)	Alpha	Alpha	Alpha	Alpha	Alpha	Alpha	M0	M1	S	A
20/20	0.0	-1.0	0.050	0.046	0.046	0.045	0.044	0.047	0.0	1.0	1.0	1.0
20/20	0.0	-1.0	0.050	0.040	0.039	0.045	0.044	0.048	0.0	1.0	1.0	5.0
20/20	0.0	-1.0	0.050	0.037	0.034	0.054	0.052	0.061	0.0	1.0	1.0	10.0

Pool Size: 10000. Simulations: 2000. Run Time: 1.77 minutes. Percent Trimmed: 10.

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 a little 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 86% to 47%. 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

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Two Means (Simulation)** procedure window by expanding **Means**, then **Two Independent Means**, then clicking on **Non-Normal Data**, and then clicking on **Tests for Two Means (Simulation)**. You may then make the appropriate entries as listed below, or open **Example 7** by going to the **File** menu and choosing **Open Example Template**.

Option   **Value**

Design Tab
Solve For .... **Power**
Alternative Hypothesis .... **Diff ≠ Diff0**
Test Type.. **T-Test**
Simulations **2000**
Alpha.. **0.05**
Group Allocation .... **Equal (N1 = N2)**
Sample Size Per Group **20**
Group 1 Distribution | H0 .... **TukeyGH(M0 S G 0)**
Group 2 Distribution | H0 .... **TukeyGH(M0 S G 0)**
Group 1 Distribution | H1 .... **TukeyGH(M0 S G 0)**
Group 2 Distribution | H1 .... **TukeyGH(M1 S G 0)**
M0 (Mean|H0)... **0**
M1 (Mean|H1) **1**
S.. **1**
G **0 0.5 0.9**

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

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

Output

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

Numeric Results

Power Comparison for Testing Mean Difference = Diff0.   Hypotheses: H0: Diff1 = Diff0; H1: Diff1 ≠ Diff0
H0 Dist's: TukeyGH(M0 S G 0) & TukeyGH(M0 S G 0)
H1 Dist's: TukeyGH(M0 S G 0) & TukeyGH(M1 S G 0)

N1/N2	H0 Diff (Diff0)	H1 Diff (Diff1)	Target Alpha	T-Test Power	Welch Power	Trim. T-Test Power	Trim. Welch Power	Mann Whit'y Power	M0	M1	S	G
20/20	0.0	-1.0	0.050	0.869	0.867	0.833	0.833	0.838	0.0	1.0	1.0	0.0
20/20	0.0	-1.0	0.050	0.880	0.879	0.923	0.922	0.948	0.0	1.0	1.0	0.5
20/20	0.0	-1.0	0.050	0.867	0.866	0.963	0.960	0.993	0.0	1.0	1.0	0.9

Pool Size: 10000. Simulations: 2000. Run Time: 16.46 seconds. Percent Trimmed: 10.

Alpha Comparison for Testing Mean Difference = Diff0.   Hypotheses: H0: Diff1 = Diff0; H1: Diff1 ≠ Diff0
H0 Dist's: TukeyGH(M0 S G 0) & TukeyGH(M0 S G 0)
H1 Dist's: TukeyGH(M0 S G 0) & TukeyGH(M1 S G 0)

N1/N2	H0 Diff (Diff0)	H1 Diff (Diff1)	Target Alpha	T-Test Alpha	Welch Alpha	Trim. T-Test Alpha	Trim. Welch Alpha	Mann Whit'y Alpha	M0	M1	S	G
20/20	0.0	-1.0	0.050	0.051	0.051	0.043	0.043	0.045	0.0	1.0	1.0	0.0
20/20	0.0	-1.0	0.050	0.039	0.038	0.043	0.041	0.044	0.0	1.0	1.0	0.5
20/20	0.0	-1.0	0.050	0.050	0.049	0.051	0.047	0.054	0.0	1.0	1.0	0.9

Pool Size: 10000. Simulations: 2000. Run Time: 16.46 seconds. Percent Trimmed: 10.

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.