Chapter 590

Multiple Contrasts (Simulation)

Introduction

This procedure uses simulation to analyze the power and significance level of two multiple-comparison procedures that perform two-sided hypothesis tests of contrasts of the group means. These are the Dunn-Bonferroni test and the Dunn-Welch test. For each scenario, two simulations are run: one estimates the significance level and the other estimates the power.

The term *contrast* refers to a user-defined comparison of the group means. The term *multiple contrasts* refers to a set of such comparisons. An additional restriction imposed is that the contrast coefficients to sum to zero.

When several contrasts are tested, the interpretation of the results is more complex because of the problem of *multiplicity*. *Multiplicity* here refers to the fact that the probability of making at least one incorrect decision increases as the number of statistical tests increases. Methods for testing *multiple contrasts* have been developed to account for this multiplicity.

Error Rates

When dealing with several simultaneous statistical tests, both individual-wise and experiment wise error rates should be considered.

1. **Comparison-wise error rate**. This is the probability of a type-I error (rejecting a true H0) for a particular test. In the case of the five-group design, there are ten possible comparison-wise error rates, one for each of the ten possible pairs. We will denote this error rate α_c.
2. **Experiment-wise (or family-wise) error rate**. This is the probability of making one or more type-I errors in the set (family) of comparisons. We will denote this error rate α_f.

The relationship between these two error rates when the tests are independent is given by

$$\alpha_f = 1 - (1 - \alpha_c)^C$$

where C is the total number of contrasts. For example, if α_c is 0.05 and C is 10, α_f is 0.401. There is about a 40% chance that at least one of the ten contrasts will be concluded to be non-zero when in fact they are not. When the tests are correlated, as they might be among a set of contrasts, the above formula provides an upper bound to the family-wise error rate.

The techniques described below provide control for α_f rather than α_c.

Technical Details

The One-Way Analysis of Variance Design

The discussion that follows is based on the common one-way analysis of variance design which may be summarized as follows. Suppose the responses Y_{ij} in *k* groups each follow a normal distribution with means $\mu_1, \mu_2, \cdots, \mu_k$ and unknown variance σ^2. Let $n_1, n_2, \cdots, n_k$ denote the number of subjects in each group. The control group is assumed to be group one.

The analysis of these responses is based on the sample means

$$\hat{\mu}_i = \overline{Y}_i = \sum_{j=1}^{n_i} \frac{Y_{ij}}{n_i}$$

and the pooled sample variance

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(Y_{ij} - \overline{Y}_i\right)^2}{\sum_{i=1}^{k}\left(n_i - 1\right)}$$

The *F* test is the usual method of analysis of the data from such a design, testing whether all of the means are equal. However, a significant *F* test does not indicate which of the groups are different, only that at least one is different. The analyst is left with the problem of determining which of the groups are different and by how much.

To determine the mean differences that are most importance, the researcher may specify a set of contrasts. These contrasts, called *a priori*, or, *planned*, contrasts should be specified before the experimental results are viewed.

The Dunn-Bonferroni procedure and the Dunn-Welch procedure have been developed to test these planned contrasts. The calculations associated with each of these tests are given below.

Contrasts

A contrast of the means is a stated difference among the means. The difference is constructed so that it represents a useful hypothesis. For example, suppose there are four groups, the first of which is a control group. It might be of interest to determine which treatments are statistically different from the control. That is, the differences $\mu_2 - \mu_1$, $\mu_3 - \mu_1$, and $\mu_4 - \mu_1$ would be tested to determine if they are non-zero.

Contrasts are often simple differences between two means. However, they may involve more than just two means. For example, suppose the first two groups receive one treatment and the second two groups receive another treatment. The contrast (difference) that would be tested is $\left(\mu_1 + \mu_2\right) - \left(\mu_3 + \mu_4\right)$.

Every contrast can be represented by the list of contrast coefficients: the values by which the means are multiplied. Here are some examples of contrasts that might be of interest when the experiment involves four groups.

Difference	Coefficients
$\mu_2 - \mu_1$	-1, 1, 0, 0
$\mu_3 - \mu_1$	-1, 0, 1, 0
$(\mu_1 + \mu_2) - (\mu_3 + \mu_4)$	1, 1, -1, -1
$(\mu_1 + \mu_1 + \mu_1) - (\mu_2 + \mu_3 + \mu_4)$	3, -1, -1, -1
$(\mu_4 + \mu_4) - (\mu_2 + \mu_3)$	0, -1, -1, 2

Note that in each case, the coefficients sum to zero. This makes it possible to test whether the quantity is different from zero.

A lot is written about *orthogonal contrasts* which have the property that the sum of the products of corresponding coefficients is zero. For example, the sum of the products of the last two contrasts given above is 0(3) + (-1)(-1) + (-1)(-1) + (2)(-1) = 0 + 1 + 1 – 2 = 0, so these two contrasts are orthogonal. However, the first two contrasts are not orthogonal since (-1)(-1) + (1)(0) + (0)(1) + (0)(0) = 1 + 0 + 0 + 0 = 1 (not zero). Orthogonal contrasts have nice properties when the sample sizes are equal. Unfortunately, they lose those properties when the group sample sizes are unequal or when the data are not normally distributed.

The procedures described in this chapter do not require that the contrasts be orthogonal. Instead, you should focus on defining a set of contrasts that answer the research questions of interest.

Dunn-Bonferroni Test

Dunn (1964) developed a procedure for simultaneously testing several contrasts. This method is also discussed in Kirk (1982) pages 106 to 109. The method consists of testing each contrast with Student's *t* distribution with degrees of freedom equal to *N*-*k* with a Bonferroni adjustment of the alpha value. That is, the alpha value is divided by *C*, the number of contrasts simultaneously tested.

The test rejects H0 if

$$\frac{\left|\sum_{i=1}^{k} c_i \overline{Y}_i\right|}{\sqrt{\hat{\sigma}^2\left(\sum_{i=1}^{k} \frac{c_i^2}{n_i}\right)}} \geq \left|t_{1-\alpha/(2C),N-k}\right|$$

Note that this is a two-sided test of the hypothesis that $\sum_{i=1}^{k} c_i \mu_i = 0$ where $\sum_{i=1}^{k} c_i = 0$.

Dunn-Welch Test

Dunn (1964) developed a procedure for simultaneously testing several contrasts. This method, using Welch's (1947) modification for the unequal variances, is discussed in Kirk (1982) pages 100, 101, 106 - 109. The method consists of testing each contrast with Student's *t* distribution with degrees of freedom given below with a Bonferroni adjustment of the alpha value. That is, the alpha value is divided by *C*, the number of contrasts simultaneously tested.

The two-sided test statistic rejects H0 if

$$\frac{\left|\sum_{i=1}^{k} c_i \overline{Y}_i\right|}{\sqrt{\sum_{i=1}^{k} \frac{c_i^2 \hat{\sigma}_i^2}{n_i}}} \geq \left|t_{1-\alpha/(2C),v'}\right|$$

where

$$v' = \frac{\left(\sum_{i=1}^{k} \frac{c_i^2 \hat{\sigma}_i^2}{n_i} \right)^2}{\sum_{i=1}^{k} \frac{c_i^4 \hat{\sigma}_i^4}{n_i^2 (n_i - 1)}}$$

Definition of Power for Multiple Contrasts

The notion of power is well-defined for individual tests. Power is the probability of rejecting a false null hypothesis. However, this definition does not extend directly when there are a number of simultaneous tests. The two definitions that we emphasize in ***PASS*** where recommended by Ramsey (1978). They are *any-contrast power* and *all-contrasts power*. Other design characteristics, such as *average-contrast power* and *false-discovery rate*, are important to consider. However, our review of the statistical literature resulted in our focus on these two definitions of power.

Any-Contrast Power

Any-contrast power is the probability of detecting at least one of the contrasts that are actually non-zero.

All-Contrasts Power

All-contrast power is the probability of detecting all of the contrasts that are actually non-zero.

Simulation 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 each test is to be 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. The number rejected is used to calculate the power of each test.
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. The number rejected is used to calculate the significance level of each test.
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 draws 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.

Procedure Options

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

Design 1 Tab

The Design 1 tab contains most of the parameters and options that you will be concerned with.

Solve For

Solve For

This option specifies the parameter to be solved for: power or sample size (n). If you choose to solve for n, you must choose the type of power you want to solve for: any-contrast power or all-contrasts power. The value of the option *Power* will then represent this type of power.

Any-contrast power is the probability of detecting at least one of the non-zero contrasts. *All-contrast power* is the probability of detecting all non-zero contrasts.

Note that the search for n may take several minutes because a separate simulation must be run for each trial value of n. You may find it quicker and more informative to solve for the Power for a range of sample sizes.

Test

MC Procedure

Specify which multiple contrast procedure is to be reported from the simulations. The choices are

- **Dunn-Bonferroni Test**
  This is the most popular and most often recommended.

- **Dunn-Welch Test**
  This is recommended when the group variances are very different.

Simulations

Simulations

This option specifies the number of iterations, *M*, used in the simulation. As the number of iterations is increased, the running time and accuracy are increased as well.

The precision of the simulated power estimates are calculated using 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 is only used when *Solve For* is set to *Sample Size (All-Contrast)* or *Sample Size (Any-Contrast)*.

Power is defined differently with multiple contrasts. Although many definitions are possible, two are adopted here. *Any-contrast power* is the probability of detecting at least one non-zero contrast. *All-contrasts power* is the probability of detecting all non-zero contrasts. As the number of contrasts is increased, these power probabilities will decrease because more tests are being conducted.

Since this is a probability, the range is between 0 and 1. Most researchers would like to have the power at least at 0.8. However, this may require extremely large sample sizes when the number of tests is large.

FWER (Alpha)

This option specifies one or more values of the *family-wise error rate* (FWER) which is the analog of alpha for multiple contrasts. FWER is the probability of falsely detecting (concluding that the means are different) at least one comparison for which the true means are the same. For independent tests, the relationship between the individual-comparison error rate (ICER) and FWER is given by the formulas

$$FWER = 1 - (1 - ICER)^C$$

or

$$ICER = 1 - (1 - FWER)^{(1/C)}$$

where '^' represents exponentiation (as in 4^2 = 16) and C represents the number of comparisons. For example, if C = 5 and FWER = 0.05, then ICER = 0.0102. Thus, the individual comparison tests must be conducted using a Type-1 error rate of 0.0102, which is much lower than the family-wise rate of 0.05.

The popular value for FWER remains at 0.05. However, if you have a large number of comparisons, you might decide that a larger value, such as 0.10, is appropriate.

Sample Size

n (Sample Size Multiplier)

This is the base, per group, sample size. One or more values separated by blanks or commas may be entered. A separate analysis is performed for each value listed here.

The group samples sizes are determined by multiplying this number by each of the Group Sample Size Pattern numbers. If the Group Sample Size Pattern numbers are represented by *m1, m2, m3, …, mk* and this value is represented by *n*, the group sample sizes *N1, N2, N3, ..., Nk* are calculated as follows:

N1=[n(m1)]

N2=[n(m2)]

N3=[n(m3)]

etc.

where the operator, [*X*] means the next integer after *X*, e.g. [3.1]=4.

For example, suppose there are three groups and the Group Sample Size Pattern is set to *1,2,3*. If n is 5, the resulting sample sizes will be 5, 10, and 15. If n is 50, the resulting group sample sizes will be 50, 100, and 150. If n is set to *2,4,6,8,10*, five sets of group sample sizes will be generated and an analysis run for each. These sets are:

2	4	6
4	8	12
6	12	18
8	16	24
10	20	30

As a second example, suppose there are three groups and the Group Sample Size Pattern is *0.2,0.3,0.5*. When the fractional Pattern values sum to one, n can be interpreted as the total sample size of all groups and the Pattern values as the proportion of the total in each group.

If n is 10, the three group sample sizes would be 2, 3, and 5.

If n is 20, the three group sample sizes would be 4, 6, and 10.

If n is 12, the three group sample sizes would be

(0.2)12 = 2.4 which is rounded up to the next whole integer, 3.

(0.3)12 = 3.6 which is rounded up to the next whole integer, 4.

(0.5)12 = 6.

Note that in this case, 3+4+6 does not equal n (which is 12). This can happen because of rounding.

Group Sample Size Pattern

The purpose of the group sample size pattern is to allow several groups with the same sample size to be generated without having to type each individually.

A set of positive, numeric values (one for each row of distributions) is entered here. Each item specified in this list applies to the whole row of distributions. For example, suppose the entry is *1 2 1* and Grps 1 = 3, Grps 2 = 1, Grps 3 = 2. The sample size pattern used would be *1 1 1 2 1 1*.

The sample size of group *i* is found by multiplying the i^{th} number from this list by the value of *n* and rounding up to the next whole number. The number of values must match the number of groups, *g*. When too few numbers are entered, 1's are added. When too many numbers are entered, the extras are ignored.

- **Equal**

  If all sample sizes are to be equal, enter *Equal* here and the desired sample size in *n*. A set of *g* 1's will be used. This will result in *n1* = *n2* = … = *ng* = *n*. That is, all sample sizes are equal to *n*.

Effect Size

These options specify the distributions to be used in the two simulations. The first option specifies the number of groups represented by the two distributions that follow. The second option specifies the distribution to be used in simulating the null hypothesis to determine the significance level (alpha). The third option specifies the distribution to be used in simulating the alternative hypothesis to determine the power.

Grps [A – C] (Grps D – I are found on the Data 2 tab)

This value specifies the number of groups specified by the H0 and H1 distribution statements to the right. Usually, you will enter '1' to specify a single H0 and a single H1 distribution, or you will enter '0' to indicate that the distributions specified on this line are to be ignored. This option lets you easily specify many identical distributions with a single phrase.

The total number of groups *g* is equal to the sum of the values for the three rows of distributions shown under the Data 1 tab and the six rows of distributions shown under the Data 2 tab.

Note that each item specified in the *Group Sample Size Pattern* option applies to the whole row of entries here. For example, suppose the *Group Sample Size Pattern* was *1 2 1* and Grps 1 = 3, Grps 2 = 1, and Grps 3 = 2. The sample size pattern would be *1 1 1 2 1 1*.

Group Distribution(s)|H0

This entry specifies the distribution of one or more groups under the null hypothesis, H0. The magnitude of the differences of the means of these distributions, which is often summarized as the standard deviation of the means, represents the magnitude of the mean differences specified under H0. Usually, the means are assumed to be equal under H0, so their standard deviation should be zero except for rounding.

These distributions are used in the simulations that estimate the actual significance level. They also specify the value of the mean under the null hypothesis, H0. Usually, these distributions will be identical. 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

Most of 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 Distribution(s)|H1

Specify the distribution of this group under the alternative hypothesis, H1. This distribution is used in the simulation that determines the power. A fundamental quantity in a power analysis is the amount of variation among the group means. In fact, classical power analysis formulas, this variation is summarized as the standard deviation of the means.

The important point to realize is that you must pay particular attention to the values you give to the means of these distributions because they are fundamental to the interpretation of the simulation.

For convenience in specifying a range of values, the parameters of the distribution can be 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 under the alternative hypothesis.

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

Equivalence Margin

Specify the largest difference for which means from different groups will be considered equal. When specifying group distributions, it is possible to end up with scenarios where some means are slightly different from each other, even though they are intended to be equivalent. This often happens when specifying distributions of different forms (e.g. normal and gamma) for different groups, where the means are intended to be the same. The parameters used to specify different distributions do not always result in means that are EXACTLY equal. This value lets you control how different means can be and still be considered equal.

This value is not used to specify the hypothesized mean differences of interest. The hypothesized differences are specified using the means (or parameters used to calculate means) for the null and alternative distributions.

This value should be much smaller than the hypothesized mean differences.

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, C)

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.

Contrasts Tab

Contrasts

Contrasts

These options specify the contrasts. You can specify as many contrasts as are necessary, but a penalty is paid in terms of reduced power for each additional contrast. Thus, the number of contrasts should be limited to those that are most important to the study.

A contrast is a weighted average of the *k* (*k* = number of groups) group means in which the weights (coefficients) sum to zero. Each successive coefficient is applied to the corresponding group mean. For example, suppose *k* = 3 and the first group is a control group. Two contrasts that might be of interest are *-1 1 0* and *-1 0 1*. These are interpreted as *(-1)Mean1 + (1)Mean2 + (0)Mean3* and *(-1)Mean1 + (0)Mean2 + (1)Mean3*, respectively. Notice that the coefficients in each set sum to zero.

Several predefined sets of contrasts are available or you can specify your own. There is no set number of contrasts that must (or may) be specified, but fewer contrasts result in higher power and smaller required samples sizes.

Possible entries are given next.

- **Individual Contrasts**

  Enter a set of numbers, separated by blanks. One coefficient must be entered for each group with one set per box. Examples of valid contrasts are

  -1 1
  -1 0 1
  0 1 -2 1
  -4 1 1 2

- **Each With First**

  This option generates k-1 contrasts appropriate for comparing each of the remaining groups with the first group. This might be used when the first group is a control group. If k = 4, the 3 contrasts are

  -1 1 0 0
  -1 0 1 0
  -1 0 0 1

- **Each With Last**

  This option generates k-1 contrasts appropriate for comparing each of the first k-1 groups with the last group. This might be used when the last group is a control group. If k = 4, the 3 contrasts are

  -1 0 0 1
  0 -1 0 1
  0 0 -1 1

- **Each With Next**

  This option generates k-1 contrasts appropriate for comparing each group with the next group. If k = 4, the 3 contrasts are

  -1 1 0 0
  0 -1 1 0
  0 0 -1 1

- **Each With Remaining**

  Each group mean is compared with the average of those remaining to the right. Suppose k=4, the 3 contrasts are

  -3 1 1 1
  0 -2 1 1
  0 0 -1 1

- **Each With All Others**

  Each group mean is compared with the average of the other groups. Suppose k=4, the 4 contrasts are

  -3 1 1 1
  1 -3 1 1
  1 1 -3 1
  1 1 1 -3

- **Progressive Split**

  The first groups are compared to the last groups. The dividing point moves from left to right. Suppose k=5, the 4 contrasts are

  -4 1 1 1 1
  -3 -3 2 2 2
  -2 -2 -2 3 3
  -1 -1 -1 -1 4.

Options Tab

The Options tab contains a random number pool size option.

Random Numbers

Random Number Pool Size

This is the size of the pool of 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.

Example 1 – Power at Various Sample Sizes

A study is being planned to find the threshold level of a certain drug. Below this threshold level, the response has little change. Once the threshold level is reached, there is a sizeable jump in the mean response rate. Little change in the response occurs as the drug level is increased above the threshold. Scientists believe that the threshold level is between 3 and 7—their best estimate, based on previous studies, is 5. Previous studies have shown that the standard deviation within a group is 3.0.

In order to find the threshold, they design a study with five levels: 3.0, 4.0, 5.0, 6.0, and 7.0. Since there is no trend in the mean value (only a sudden shift) as the dose level is increased, they decide to test the following hypotheses:

Difference	Coefficients
$\mu_2 - \mu_1$	-1, 1, 0, 0, 0
$\mu_3 - \mu_2$	0, -1, 1, 0, 0
$\mu_4 - \mu_3$	0, 0, -1, 1, 0
$\mu_5 - \mu_4$	0, 0, 0, -1, 1

Notice that this set of hypotheses answers the question directly. An overall F-test would test the hypothesis that at least one mean is different, but it would not indicate which is different. The question might be settled by considering all possible pairs, but there are ten pairs, so ten hypothesis tests would have to be considered instead of only four—decreasing the power.

Researchers want to detect a shift in the mean as small as 2.0. Hence, they want to study the power when the means are 0.0, 0.0, 2.0, 2.0, 2.0. They want to investigate sample sizes of 10, 30, 50, and 70 subjects per group.

They have no reason to assume that the variance will change a great deal from group to group, so they decide to analyze the data using the Dunn-Bonferroni procedure. They set the FWER to 0.05. Note that, based on these means, only the second of the four contrasts will be significant, so the any-contrast power will be the same as the all-contrast power.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Multiple Contrasts (Simulation)** procedure window by expanding **Means**, then clicking on **Multiple Comparisons**, and then clicking on **Multiple Contrasts (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 1 Tab	
Solve For	**Power**
MC Procedure	**Dunn-Bonferroni Test**
Simulations	**2000**
FWER (Alpha)	**0.05**
n (Sample Size Multiplier)	**10 30 50 70**
Group Sample Size Pattern	**Equal**
Grps 1	**2**
Control Distribution | H0	**Normal(M0 S)**
Control Distribution | H1	**Normal(M0 S)**
Grps 2	**3**
Group 2 Distribution(s) | H0	**Normal(M0 S)**
Group 2 Distribution(s) | H1	**Normal(M1 S)**
Equivalence Margin	**0.1**
M0 (Mean|H0)	**0**
M1 (Mean|H1)	**2**
S	**3**
Contrasts Tab	
Contrasts	**Each With Next**
Reports Tab	
All reports except Comparative	**Checked**

Annotated Output

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

Simulation Summary Report

Summary of Simulations for Testing Multiple Contrasts of 5 Groups
MC Procedure: Dunn-Bonferroni Test

Sim. No.	Any-Cont. Power	Group Smpl. Size n	Total Smpl. Size N	All-Cont. Power	S.D. of Means Sm|H1	S.D. of Data SD|H1	Actual FWER	Target FWER	M0	M1	S
1	0.136	10.0	50	0.136	1.0	3.0	0.048	0.050	0.0	2.0	3.0
	(0.015)	[0.121	0.151]	(0.015)	[0.121	0.151]	(0.009)	[0.038	0.057]		
2	0.509	30.0	150	0.509	1.0	3.0	0.038	0.050	0.0	2.0	3.0
	(0.022)	[0.487	0.530]	(0.022)	[0.487	0.530]	(0.008)	[0.030	0.046]		
3	0.796	50.0	250	0.796	1.0	3.0	0.048	0.050	0.0	2.0	3.0
	(0.018)	[0.778	0.813]	(0.018)	[0.778	0.813]	(0.009)	[0.038	0.057]		
4	0.924	70.0	350	0.924	1.0	3.0	0.041	0.050	0.0	2.0	3.0
	(0.012)	[0.912	0.936]	(0.012)	[0.912	0.936]	(0.009)	[0.032	0.050]		

Pool Size: 10000. Simulations: 2000. Run Time: 56.58 seconds.

Summary of Simulations Report Definitions
H0: the null hypothesis that the contrast of the means is zero.
H1: the alternative hypothesis that the contrast of the means is not zero.
Cont.: abbreviates 'Contrast'. Refers to a weighted average of the means whose weights sum to zero.
All-Cont. Power: the estimated probability of detecting all unequal contrasts.
Any-Cont. Power: the estimated probability of detecting at least one unequal contrasts.
n: the average of the group sample sizes.
N: the combined sample size of all groups.
Family-Wise Error Rate (FWER): the probability of detecting at least one zero contrast assuming H0.
Target FWER: the user-specified FWE.
Actual FWER: the FWER estimated by the alpha simulation.
Sm|H1: the standard deviation of the group means under H1.
SD|H1: the pooled, within-group standard deviation under H1.
Second Row: provides the precision and a confidence interval based on the size of the simulation for Any-Contrast Power, All-Contrasts Power, and FWER. The format is (Precision) [95% LCL and UCL Alpha].

Summary Statements
A one-way design with 5 groups has an average group sample size of 10.0 for a total sample size of 50. This design achieved an any-contrast power of 0.136 and an all-contrast power of 0.136 using the Dunn-Bonferroni Test procedure for comparing each contrast of the group means with zero. The target family-wise error rate was 0.050 and the actual family-wise error rate was 0.048. The average within group standard deviation assuming the alternative distribution is 3.0. These results are based on 2000 Monte Carlo samples from the null distributions: N(M0 S); N(M0 S); N(M0 S); N(M0 S); and N(M0 S) and the alternative distributions: N(M0 S); N(M0 S); N(M1 S); N(M1 S); and N(M1 S). Other parameters used in the simulation were: M0 = 0.0, M1 = 2.0, and S = 3.0.

This report shows that a group sample size of about 50 will be needed to achieve 80% power or about 70 for 90% power.

Any-Cont. Power

This is the probability of detecting any of the significant contrasts. This value is estimated by the simulation using the H1 distributions.

Note that a precision value (half the width of its confidence interval) and a confidence interval are shown on the line below this row. These values provide the precision of the estimated power.

All- Cont. Power

This is the probability of detecting all of the significant contrasts. This value is estimated by the simulation using the H1 distributions.

Note that a precision value (half the width of its confidence interval) and a confidence interval are shown on the line below this row. These values provide the precision of the estimated power.

Group Sample Size n

This is the average of the individual group sample sizes.

Total Sample Size N

This is the total sample size of the study.

S.D. of Means Sm|H1

This is the standard deviation of the hypothesized means of the alternative distributions. Under the null hypothesis this value is zero. It represents the magnitude of the difference among the means. It is roughly equal to the average difference between the group means and the overall mean.

Note that the effect size is the ratio of Sm|H1 and SD|H1.

S.D. of Data SD|H1

This is the within-group standard deviation calculated from samples from the alternative distributions.

Actual FWER

This is the value of FWER (family-wise error rate) estimated by the simulation using the H0 distributions. It should be compared with the Target FWER to determine if the test procedure is accurate.

Note that a precision value (half the width of its confidence interval) and a confidence interval are shown on the line below this row. These values provide the precision of the Actual FWER.

Target FWER

This is the target value of FWER that was set by the user.

M0

This is the value entered for M0, the group means under H0.

M1

This is the value entered for M1, the group means under H1.

S

This is the value entered for S, the standard deviation.

Error-Rate Summary for H0 Simulation

Error Rate Summary from H0 (Alpha) Simulation of 5 Groups
MC Procedure: Dunn-Bonferroni Test

Sim. No.	No. of Zero Cont.	Mean No. of Type-1 Errors	Prop. Type-1 Errors	Prop. (No. of Type-1 Errors > 0) FWER	Target FWER	Mean Cont. Alpha	Min Cont. Alpha	Max Cont. Alpha
1	4	0.053	0.013	0.048	0.050	0.013	0.010	0.016
2	4	0.042	0.010	0.038	0.050	0.010	0.009	0.012
3	4	0.056	0.014	0.048	0.050	0.014	0.012	0.016
4	4	0.047	0.012	0.041	0.050	0.012	0.010	0.014

This report shows the results of the H0 simulation. This simulation uses the H0 settings for each group. Its main purpose is to provide an estimate of the FWER.

No. of Zero Cont.

Since under H0 all means are equal, this is the number of contrasts.

Mean No. of Type-1 Errors

This is the average number of type-1 errors (false detections) per set (family).

Prop. Type-1 Errors

This is the proportion of type-1 errors (false detections) among all tests that were conducted.

Prop. (No. of Type-1 Errors>0) FWER

This is the proportion of the H0 simulations in which at least one type-1 error occurred. This is called the family-wise error rate.

Target FWER

This is the target value of FWER that was set by the user.

Mean Cont. Alpha

Alpha is the probability of rejecting H0 when H0 is true. It is a characteristic of an individual test. This is the average individual alpha value over all of the contrasts.

Min Cont. Alpha

This is the minimum of all contrast alphas.

Max Cont. Alpha

This is the maximum of all contrast alphas.

Error-Rate Summary for H1 Simulation

Error-Rate Summary from H1 (Power) Simulation of 5 Groups
MC Procedure: Dunn-Bonferroni Test

Sim. No.	No. of Zero/ Non-0 Cont.	Mean No. of False Pos.	Mean No. of False Neg.	Prop. Errors	Prop. Zero that were Detect.	Prop. Non-0. that were Undet.	(FDR) Prop. Detect. that were Zero	Prop. Undet. that were Non-0	All Non-0 Cont. Power	Any Non-0 Cont. Power	Mean Cont. Power	Min Cont. Power	Max Cont. Power
1	3/1	0.04	0.86	0.226	0.013	0.864	0.223	0.226	0.136	0.136	0.044	0.010	0.136
2	3/1	0.03	0.49	0.131	0.011	0.492	0.061	0.142	0.509	0.509	0.135	0.009	0.509
3	3/1	0.04	0.20	0.062	0.014	0.205	0.050	0.065	0.796	0.796	0.209	0.009	0.796
4	3/1	0.03	0.08	0.027	0.010	0.076	0.031	0.025	0.924	0.924	0.239	0.008	0.924

This report shows the results of the H1 simulation. This simulation uses the H1 settings for each group. Its main purpose is to provide an estimate of the power.

No. of Zero/Non-0 Cont.

The first value is the number of contrasts that were zero under H1. The second value is the number of contrasts that were non-zero under H1.

Mean No. False Positives

This is the average number of zero contrasts that were declared as being non-zero by the testing procedure. A *false positive* is a type-1 (alpha) error.

Mean No. False Negatives

This is the average number of non-zero contrasts that were not declared as being non-zero by the testing procedure. A *false negative* is a type-2 (beta) error.

Prop. Errors

This is the proportion of type-1 and type-2 errors.

Prop. Equal that were Detect.

This is the proportion of the zero contrasts in the H1 simulations that were declared as non-zero.

Prop. Uneq. that were Undet.

This is the proportion of non-zero contrasts in the H1 simulations that were not declared as being non-zero.

Prop. Detect. that were Zero (FDR)

This is the proportion of all detected contrasts in the H1 simulations that were actually zero. This is often called the *false discovery rate*.

Prop. Undet. that were Non-0.

This is the proportion of undetected contrasts in the H1 simulations that were actually non-zero.

All Non-0 Cont. Power

This is the probability of detecting all non-zero contrasts in the H1 simulation.

Any Non-0 Cont. Power

This is the probability of detecting any non-zero contrasts in the H1 simulation.

Mean, Min, and Max Cont. Power

These items give the average, the minimum, and the maximum of the contrast powers from the H1 simulation.

Detail Model Report

Detailed Model Report for Simulation No. 1
Target FWER = 0.050, M0 = 0.0, M1 = 2.0, S = 3.0
MC Procedure: Dunn-Bonferroni Test

Hypo. Type	Groups	Group Labels	n/N	Group Mean	Ave. S.D.	Simulation Model
H0	1-2	A1-A2	10/50	0.0	3.1	N(M0 S)
H0	3-5	B1-B3	10/50	0.0	3.0	N(M0 S)
H0	All			Sm=0.0	3.0	
H1	1-2	A1-A2	10/50	0.0	3.0	N(M0 S)
H1	3-5	B1-B3	10/50	2.0	3.0	N(M1 S)
H1	All			Sm=1.0	3.0	

Detailed Model Report for Simulation No. 2

Hypo. Type	Groups	Group Labels	n/N	Group Mean	Ave. S.D.	Simulation Model
H0	1-2	A1-A2	30/150	0.0	3.0	N(M0 S)
H0	3-5	B1-B3	30/150	0.0	3.0	N(M0 S)
H0	All			Sm=0.0	3.0	
H1	1-2	A1-A2	30/150	0.0	3.0	N(M0 S)
H1	3-5	B1-B3	30/150	2.0	3.0	N(M1 S)
H1	All			Sm=1.0	3.0	

Detailed Model Report for Simulation No. 3

Hypo. Type	Groups	Group Labels	n/N	Group Mean	Ave. S.D.	Simulation Model
H0	1-2	A1-A2	30/150	0.0	3.0	N(M0 S)
H0	3-5	B1-B3	30/150	0.0	3.0	N(M0 S)
H0	All			Sm=0.0	3.0	
H1	1-2	A1-A2	30/150	0.0	3.0	N(M0 S)
H1	3-5	B1-B3	30/150	2.0	3.0	N(M1 S)
H1	All			Sm=1.0	3.0	

Detailed Model Report for Simulation No. 4

Hypo. Type	Groups	Group Labels	n/N	Group Mean	Ave. S.D.	Simulation Model
H0	1-2	A1-A2	70/350	0.0	3.0	N(M0 S)
H0	3-5	B1-B3	70/350	0.0	3.0	N(M0 S)
H0	All			Sm=0.0	3.0	
H1	1-2	A1-A2	70/350	0.0	3.0	N(M0 S)
H1	3-5	B1-B3	70/350	2.0	3.0	N(M1 S)
H1	All			Sm=1.0	3.0	

This report shows details of each row of the previous reports.

Hypo. Type

This indicates which simulation is being reported on each row. H0 represents the null simulation and H1 represents the alternative simulation.

Groups

Each group in the simulation is assigned a number. This item shows the arbitrary group number that was assigned.

Group Labels

These are the labels that were used in the individual alpha-level reports.

n/N

n is the average sample size of the groups. N is the total sample size across all groups.

Group Mean

These are the means of the individual groups as specified for the H0 and H1 simulations.

Ave. S.D.

This is the average standard deviation of all groups reported on each line. Note that it is calculated from the simulated data.

Simulation Model

This is the distribution that was used to simulate data for the groups reported on each line.

List of Contrast Coefficients

List of Contrast Coefficients

Contrasts	Groups A1	A2	B1	B2	B3
Con1	-1.0	1.0	0.0	0.0	0.0
Con2	0.0	-1.0	1.0	0.0	0.0
Con3	0.0	0.0	-1.0	1.0	0.0
Con4	0.0	0.0	0.0	-1.0	1.0

The contrasts are shown down the rows. The groups are shown across the columns.
The coefficients (weights) are shown as the body of the table.

This report shows values of the contrast coefficients so you can double-check that they are what was intended.

Probability of Rejecting Individual Contrasts

Probability of Rejecting Individual Contrasts. Simulation No. 1

Contrasts	Alpha	Power
Con1	0.013	0.010
Con2	0.016	0.136
Con3	0.010	0.016
Con4	0.014	0.014

Alpha: probability of rejecting hypothesis that contrast is zero under alpha (H0) simulation.
Power: probability of rejecting hypothesis that contrast is zero under power (H1) simulation.

Probability of Rejecting Individual Contrasts. Simulation No. 2

Contrasts	Alpha	Power
Con1	0.010	0.009
Con2	0.011	0.509
Con3	0.012	0.013
Con4	0.009	0.012

Probability of Rejecting Individual Contrasts. Simulation No. 3

Contrasts	Alpha	Power
Con1	0.013	0.014
Con2	0.015	0.796
Con3	0.016	0.019
Con4	0.012	0.009

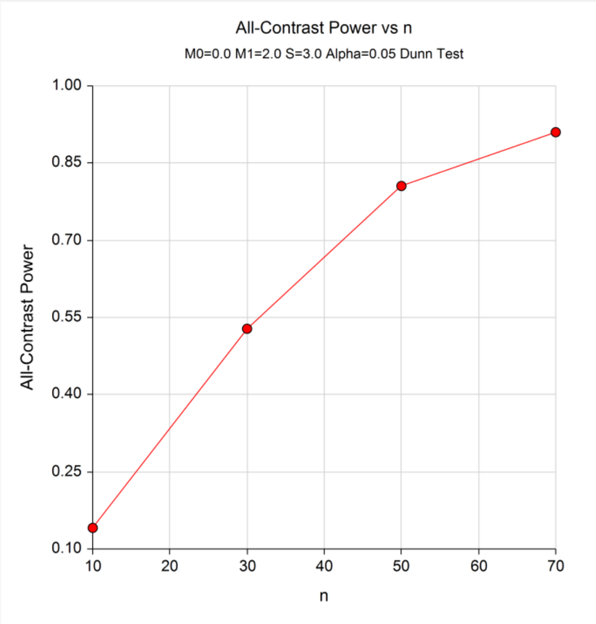
Probability of Rejecting Individual Contrasts. Simulation No. 4

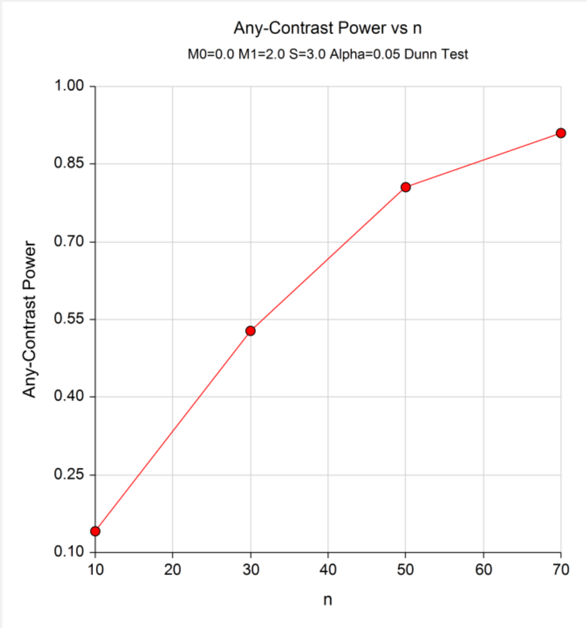
Contrasts	Alpha	Power
Con1	0.010	0.012
Con2	0.014	0.924
Con3	0.011	0.011
Con4	0.012	0.008

This report shows alpha and individual power for each contrast for each simulation that was run.

In this example, only the second contrast was non-zero, so that is the only one which has large values for the power.

Plots Section





These plots give a visual presentation of the all-contrasts power values and the any-contrast power values.

Example 2 – Comparative Results

Continuing with Example 1, the researchers want to study the characteristics of alternative multiple contrast procedures.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Multiple Contrasts (Simulation)** procedure window by expanding **Means**, then clicking on **Multiple Comparisons**, and then clicking on **Multiple Contrasts (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 1 Tab	
Solve For	**Power**
MC Procedure	**Dunn-Bonferroni Test**
Simulations	**2000**
FWER (Alpha)	**0.05**
n (Sample Size Multiplier)	**10 30 50 70**
Group Sample Size Pattern	**Equal**
Grps 1	**2**
Control Distribution | H0	**Normal(M0 S)**
Control Distribution | H1	**Normal(M0 S)**
Grps 2	**3**
Group 2 Distribution(s) | H0	**Normal(M0 S)**
Group 2 Distribution(s) | H1	**Normal(M1 S)**
Equivalence Margin	**0.1**
Design 1 Tab (continued)	
M0 (Mean|H0)	**0**
M1 (Mean|H1)	**2**
S	**3**
Contrasts Tab	
Contrasts	**Each With Next**
Reports Tab	
Comparative Reports	**Checked**
Plots Tab	
Comparative Any-Contrast Power Plot	**Checked**
Comparative All-Contrast Power Plot	**Checked**

Output

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

Numeric Results and Plots

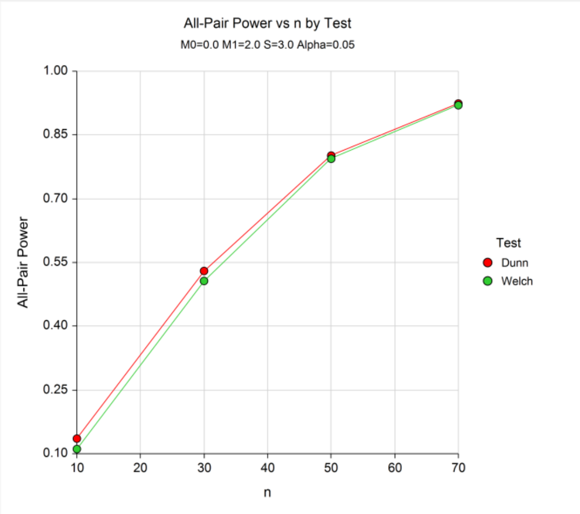
Power Comparison for Simultaneously Testing Multiple Contrasts of 5 Groups

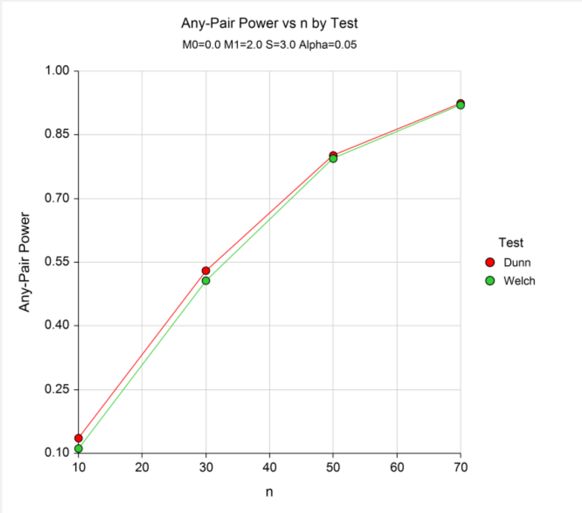
Sim. No.	Total Sample Size	Target Alpha	Dunn Bonferroni All-Cont. Power	Dunn Welch All-Cont. Power	Dunn Bonferroni Any-Cont. Power	Dunn Welch Any-Cont. Power
1	50	0.050	0.145	0.129	0.145	0.129
2	150	0.050	0.535	0.519	0.535	0.519
3	250	0.050	0.806	0.785	0.806	0.785
4	350	0.050	0.928	0.924	0.928	0.924

Pool Size: 10000. Simulations: 2000. Run Time: 5.53 minutes.

Family-Wise Error-Rate Comparison for Simultaneously Testing Multiple Contrasts of 5 Groups

Sim. No.	Total Sample Size	Target FWER	Dunn Bonferroni FWER	Dunn Welch FWER
1	50	0.050	0.039	0.039
2	150	0.050	0.041	0.046
3	250	0.050	0.050	0.047
4	350	0.050	0.051	0.044





These reports show the power and FWER of both multiple contrast procedures. In these simulations of groups from the normal distributions with equal variances, there is little difference in the power of the two procedures.

Example 3 – Validation

We could not find an article that gives power values for this test, so we decided to validate the procedure by comparing its results to those of the one-way ANOVA procedure which allows a single contrast to be tested. Using the settings of Example 1 and using the contrast '0, -1, 1, 0, 0', we obtained the following powers: 0.3085, 0.7274, 0.9131, and 0.9758.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Multiple Contrasts (Simulation)** procedure window by expanding **Means**, then clicking on **Multiple Comparisons**, and then clicking on **Multiple Contrasts (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**
MC Procedure	**Dunn-Bonferroni Test**
Simulations	**2000**
FWER (Alpha)	**0.05**
n (Sample Size Multiplier)	**10 30 50 70**
Group Sample Size Pattern	**Equal**
Grps 1	**2**
Control Distribution | H0	**Normal(M0 S)**
Control Distribution | H1	**Normal(M0 S)**
Grps 2	**3**
Group 2 Distribution(s) | H0	**Normal(M0 S)**
Group 2 Distribution(s) | H1	**Normal(M1 S)**
Equivalence Margin	**0.1**
M0 (Mean|H0)	**0**
M1 (Mean|H1)	**2**
S	**3**
Contrasts Tab	
Contrasts	**0 -1 1 0 0**

Output

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

Numeric Results

Summary of Simulations for Testing Multiple Contrasts of 5 Groups
MC Procedure: Dunn-Bonferroni Test

Sim. No.	Any-Cont. Power	Group Smpl. Size n	Total Smpl. Size N	All-Cont. Power	S.D. of Means Sm|H1	S.D. of Data SD|H1	Actual FWER	Target FWER	M0	M1	S
1	0.297	10.0	50	0.297	1.0	3.0	0.050	0.050	0.0	2.0	3.0
	(0.020)	[0.276	0.317]	(0.020)	[0.276	0.317]	(0.010)	[0.040	0.059]		
2	0.741	30.0	150	0.741	1.0	3.0	0.050	0.050	0.0	2.0	3.0
	(0.019)	[0.721	0.760]	(0.019)	[0.721	0.760]	(0.010)	[0.040	0.060]		
3	0.914	50.0	250	0.914	1.0	3.0	0.061	0.050	0.0	2.0	3.0
	(0.012)	[0.901	0.926]	(0.012)	[0.901	0.926]	(0.010)	[0.050	0.071]		
4	0.974	70.0	350	0.974	1.0	3.0	0.055	0.050	0.0	2.0	3.0
	(0.007)	[0.967	0.981]	(0.007)	[0.967	0.981]	(0.010)	[0.045	0.065]		

Pool Size: 10000. Simulations: 2000. Run Time: 56.58 seconds.

In each case, the confidence interval includes the actual value. That is, 0.3085 is between 0.276 and 0.317, 0.7274 is between 0.721 and 0.760, 0.9131 is between 0.901 and 0.926, and 0.9758 is between 0.967 and 0.981. This validates the procedure.