Chapter 585

Multiple Comparisons of Treatments vs. a Control (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 each treatment group mean versus the control group mean using simulation. These are Dunnett's test and the Kruskal-Wallis test. For each scenario, two simulations are run: one estimates the significance level and the other estimates the power.

The term *multiple comparisons of treatments versus a control* refers to the set of comparisons of each treatment group to a control group. If there are *k* groups of which *k*-1 are treatment groups, there will be *k*-1 tests.

When several comparisons are made among the group means, the interpretation for each comparison becomes more complex because of the problem of *multiplicity. Multiplicity* here refers to the fact that the chances of making at least one incorrect decision increases as the number of statistical tests increases. The method of *multiple comparisons* has 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 comparisons in the family. 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 pairs will be concluded to be different when in fact they are all the same. When the tests are correlated, as they are among a set of pair-wise comparisons, 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 $\mu_1, \mu_1, \cdots, \mu_k$ and unknown variance σ^2. Let $n_1, n_1, \cdots, n_k$ 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

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

and the pooled sample variance

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

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.

The Dunnett procedure and a special version of the Kruskal-Wallis procedure have been developed for this situation. The calculation of each of these tests is given next.

Dunnett's Test

Dunnett (1955) developed a test procedure for simultaneously comparing each treatment with a control group. It uses the critical values of a special *t* distribution given in Dunnett (1955). For each treatment and control pair, the significance test is calculated by rejecting the null hypothesis of mean equality if

$$\frac{|\bar{Y}_i - \bar{Y}_1|}{\sqrt{\hat{\sigma}^2 \left(\frac{1}{n_i} + \frac{1}{n_1}\right)}} \geq t_{1-\alpha/2, k, N-k}^{Dunnett} \quad i = 2, \ldots, k$$

Kruskal-Wallis Test

This test is attributed to Dunn (1964) and is referenced in Gibbons (1976). It is a nonparametric, or distribution-free, test which means that normality is not assumed. It tests whether the medians of each treatment-control pair are equal using a rank test. Sample sizes of at least five (but preferably larger) for each treatment are recommended for the use of this test. The error rate is adjusted on a comparison-wise basis to give the experiment-wise error rate, α_f. Instead of using means, it uses average ranks as the following formula indicates, with $\alpha = \alpha_f / (2(k-1))$. For each treatment and control pair, the significance test is calculated by rejecting the null hypothesis of median equality if

$$\frac{|\bar{R}_i - \bar{R}_1|}{\sqrt{\frac{N(N+1)}{12} \left(\frac{1}{n_i} + \frac{1}{n_1}\right)}} \geq z_\alpha \quad i = 2, \ldots, k$$

Note that, when necessary, the usual adjustment for ties is made.

Definition of Power for Multiple Comparisons

The notion of the power of a test is well-defined for individual tests. Power is the probability of rejecting a false null hypothesis. However, this definition does not extend easily when there are a number of simultaneous tests.

To understand the problem, consider an experiment with four groups labeled, C, A, B, and D. Suppose C is the control group. There are three paired comparisons in this experiment: A-C, B-C, and D-C. How do we define power for these three tests? One approach would be to calculate the power of each of the three tests, ignoring the other two. However, this ignores the interdependence among the three tests. Other definitions of the power of the set of tests might be the probability of detecting at least one of the differing pairs, exactly one of the differing pairs, at least two of the differing pairs, and so on. As the number of pairs increases, the number of possible definitions of power also increases. The two definitions that we emphasize in **PASS** were recommended by Ramsey (1978). They are *any-pair power* and *all-pairs power*. Other design characteristics, such as average-comparison 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-Pair Power

Any-pair power is the probability of detecting at least one of the pairs that are actually different.

All-Pairs Power

All-pairs power is the probability of detecting all of the pairs that are actually different.

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 <u>alternative</u> 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 <u>null</u> 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.

Example 1 – Power at Various Sample Sizes

An experiment is being designed to compare the responses of three different treatments to a control. Previous studies have shown that the standard deviation within a group is 3.0. Researchers want to detect a shift in the mean of at least 2.0. To accomplish this, they set the mean of the control group to zero and the three treatment means to 2.0. They want to investigate sample sizes of 5, 10, 15, and 20 subjects per group.

Their primary analysis will be a set of Dunnett multiple comparison tests. They set the FWER to 0.05.

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 1** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

Solve For ....**Power**
MC Procedure...**Dunnett Test**
Simulations **2000**
Random Seed...**3598403** (for Reproducibility)
FWER (Alpha)...**0.05**
n (Sample Size Multiplier)..............................**5 10 15 20**
Group Sample Size Pattern ....**Equal**
Ctrl Grps **1**
Ctrl Group Distribution(s)|H0.........................**Normal(M0 S)**
Ctrl Group Distribution(s)|H1.........................**Normal(M0 S)**
Set A Grps **3**
Set A Group Distribution(s)|H0 ....**Normal(M0 S)**
Set A Group Distribution(s)|H1 ....**Normal(M1 S)**
Set B Grps **0**
Equivalence Margin **0.1**
M0 (Mean|H0) Parameter Value(s)...............**0**
M1 (Mean|H1) Parameter Value(s)................**2**
Parameter 1 Label ....**S**
Parameter 1 Value(s)....................................**3**

Output

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

Simulation Summary Report

Summary of Simulations of the Control Group and the 3 Treatment Groups

Solve For: Power
MC Procedure: Dunnett's Test

Sim. No.	Any-Pair Power	Group Sample Size n	Total Sample Size N	All-Pairs Power	S.D. of Means Sm|H1	S.D. of Data SD|H1	Actual FWER	Target FWER	M0	M1	S
1	0.192 (0.017) [0.174 0.209]	5	20	0.015 (0.005) [0.01 0.02]	0.9	3	0.051 (0.01) [0.041 0.06]	0.05	0	2	3
2	0.344 (0.021) [0.323 0.365]	10	40	0.046 (0.009) [0.036 0.055]	0.9	3	0.054 (0.01) [0.044 0.063]	0.05	0	2	3
3	0.501 (0.022) [0.479 0.523]	15	60	0.099 (0.013) [0.085 0.112]	0.9	3	0.054 (0.01) [0.044 0.064]	0.05	0	2	3
4	0.653 (0.021) [0.632 0.674]	20	80	0.185 (0.017) [0.168 0.202]	0.9	3	0.051 (0.01) [0.041 0.06]	0.05	0	2	3

Pool Size: 10000. Simulations: 2000. Run Time: 1.47 seconds.
User-Entered Random Seed: 3598403

H0	The null hypothesis that each treatment mean is equal to the control mean.
H1	The alternative hypothesis that at least one treatment mean differs from the control mean.
Pair	Each comparison of a treatment mean and a control mean is a "pair."
Any-Pair Power	The estimated probability of detecting at least one unequal pair. The second row provides the precision and a 95% confidence interval for Any-Pair Power, (Any-Pair Power Precision) [95% LCL and UCL], based on the size of the simulation.
n	The average of the group sample sizes.
N	The combined sample size of all groups.
All-Pairs Power	The estimated probability of detecting all unequal pairs. The second row provides the precision and a 95% confidence interval for All-Pairs Power, (All-Pairs Power Precision) [95% LCL and UCL], based on the size of the simulation.
Sm|H1	The standard deviation of the group means under H1.
SD|H1	The pooled, within-group standard deviation under H1.
FWER	Family-Wise Error Rate. The probability of detecting at least one equal pair assuming H0.
Target FWER	The user-specified FWER.
Actual FWER	The FWER estimated by the alpha simulation. The second row provides the precision and a 95% confidence interval for FWER, (FWER Precision) [95% LCL and UCL], based on the size of the simulation.

Summary Statements

A one-way design with 3 treatment groups and one control group has an average group sample size of 5 for a total sample size of 20. This design achieved an any-pair power of 0.1915 and an all-pairs power of 0.015 using the Dunnett's Test procedure for comparing each treatment mean with the control mean. The target family-wise error rate was 0.05 and the actual family-wise error rate was 0.051. The average within group standard deviation assuming the alternative distribution is 3. These results are based on 2000 Monte Carlo samples from the null distributions: N(M0 S); N(M0 S); N(M0 S); and N(M0 S) and the alternative distributions: N(M0 S); N(M1 S); N(M1 S); and N(M1 S). Other parameters used in the simulation were: M0 = 0, M1 = 2, and S = 3.

Dropout-Inflated Sample Size

Average Group Sample Size n	Group	Dropout Rate	Sample Size Ni	Dropout-Inflated Enrollment Sample Size Ni'	Expected Number of Dropouts Di
5	1 - 4	20%	5	7	2
	Total		20	28	8
10	1 - 4	20%	10	13	3
	Total		40	52	12
15	1 - 4	20%	15	19	4
	Total		60	76	16
20	1 - 4	20%	20	25	5
	Total		80	100	20

n	The average group sample size.
Group	Lists the group numbers.
Dropout Rate	The percentage of subjects (or items) that are expected to be lost at random during the course of the study and for whom no response data will be collected (i.e., will be treated as "missing"). Abbreviated as DR.
Ni	The evaluable sample size for each group at which power is computed (as entered by the user). If Ni subjects are evaluated out of the Ni' subjects that are enrolled in the study, the design will achieve the stated power.
Ni'	The number of subjects that should be enrolled in each group in order to obtain Ni evaluable subjects, based on the assumed dropout rate. Ni' is calculated by inflating Ni using the formula Ni' = Ni / (1 - DR), with Ni' always rounded up. (See Julious, S.A. (2010) pages 52-53, or Chow, S.C., Shao, J., Wang, H., and Lokhnygina, Y. (2018) pages 32-33.)
Di	The expected number of dropouts in each group. Di = Ni' - Ni.

Dropout Summary Statements

Anticipating a 20% dropout rate, group sizes of 7, 7, 7, and 7 subjects should be enrolled to obtain final group sample sizes of 5, 5, 5, and 5 subjects.

.
.
.

References

Devroye, Luc. 1986. Non-Uniform Random Variate Generation. Springer-Verlag. New York.
Hsu, Jason. 1996. Multiple Comparisons: Theory and Methods. Chapman & Hall. London.
Kirk, Roger E. 1982. Experimental Design: Procedures for the Behavioral Sciences. Brooks/Cole. Pacific Grove, California.
Matsumoto, M. and Nishimura,T. 1998. 'Mersenne twister: A 623-dimensionally equidistributed uniform pseudorandom number generator.' ACM Trans. On Modeling and Computer Simulations.
Mukerjee, H., Robertson, T., and Wright, F.T. 1987. 'Comparison of Several Treatments With a Control Using Multiple Contrasts.' Journal of the American Statistical Association, Volume 82, No. 399, pages 902-910.

This report shows the estimated any-pairs power, all-pairs power, and FWER for each scenario. The second row shows three 95% confidence intervals in brackets: the first for the any-pairs power, the second for the all-pairs power, and the third for the FWER. Half the width of each confidence interval is given in parentheses as a fundamental measure of the precision of the simulation. As the number of simulations is increased, the width of the confidence intervals will decrease.

Any-Pairs Power

This is the probability of detecting <u>any</u> of the significant pairs. 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-Pairs Power

This is the probability of detecting <u>all</u> of the significant pairs. 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. This value represents the magnitude of the difference among the means that is being tested. 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

The target value of FWER.

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 4 Groups
───

MC Procedure: Dunnett's Test
───

Sim. No.	No. of Equal Pairs	Mean No. of Type-1 Errors	Prop. Type-1 Errors	Prop. (No. of Type-1 Errors > 0) FWER	Target FWER	Mean Pairs Alpha	Min Pairs Alpha	Max Pairs Alpha
1	3	0.061	0.020	0.051	0.05	0.020	0.020	0.021
2	3	0.062	0.021	0.054	0.05	0.021	0.019	0.023
3	3	0.063	0.021	0.054	0.05	0.021	0.015	0.025
4	3	0.055	0.018	0.051	0.05	0.018	0.016	0.022

Pair	The control group and one treatment group.
Type-1 Error	Rejecting the hypothesis that the pair have equal means when the means are equal.
Family	The set of all pairs (see "Pairs" above).
Sim. No.	Short for simulation, or row, number.
Prop.	Short for "Proportion."
No. of Equal Pairs	Since under H0 all means are equal, this is the number of unique pairs of the groups.
Mean No. of Type-1 Errors	The average number of type-1 errors per set (family).
Prop. (No...) FWER	The probability of falsely detecting at least one pair under H0.
Target FWER	The specified (user-specified) value of the Alpha (FWER).
FWER	Family-wise error rate. The family consists of all possible pairs.
Mean Pairs Alpha	The average of all individual comparison alphas.
Min Pairs Alpha	The minimum of all individual comparison alphas.
Max Pairs Alpha	The maximum of all individual comparison alphas.

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 Equal Pairs

Since under H0 all means are equal, this is the number of unique pairs of the groups. Thus, this is the number of treatment groups.

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 Pairs Alpha

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

Min Pairs Alpha

This is the minimum of all of the individual comparison alphas.

Max Pairs Alpha

This is the maximum of all of the individual comparison alphas.

Error-Rate Summary for H1 Simulation

Error-Rate Summary from H1 (Power) Simulation of 4 Groups

MC Procedure: Dunnett's Test

Sim. No.	No. of Equal/ Uneq. Pairs	Mean No. of False Pos.	Mean No. of False Neg.	Prop. Errors	Prop. Equal that were Detect.	Prop. Uneq. that were Undet.	(FDR) Prop. Detect. that were Equal	Prop. Undet. that were Uneq.	All Uneq. Pairs Power	Any Uneq. Pairs Power	Mean Pairs Power	Min Pairs Power	Max Pairs Power
1	0/3	0	2.74	0.913	0	0.913	0	1	0.015	0.192	0.087	0.079	0.093
2	0/3	0	2.47	0.824	0	0.824	0	1	0.046	0.344	0.176	0.167	0.181
3	0/3	0	2.15	0.716	0	0.716	0	1	0.099	0.501	0.284	0.280	0.290
4	0/3	0	1.76	0.588	0	0.588	0	1	0.185	0.653	0.412	0.405	0.425

Pair	The control group and one treatment group.
Type-1 Error	Rejecting the hypothesis that the pair have equal means when the means are equal.
Family	The set of all pairs (see "Pairs" above).
Sim. No.	Short for simulation, or row, number.
Prop.	Short for "Proportion."
Uneq.	Short for "Unequal." This refers to pairs whose means are actually different under H1.
Equal	Pairs whose means are actually equal under H1.
Detect.	Short for "Detected." This refers to pairs whose means were concluded as being different.
Undet.	Short for "Undetected." This refers to pairs whose means were not concluded as being different.
False Pos.	Equal pairs that were falsely concluded as being different.
False Neg.	Unequal pairs that were falsely not detected.
FDR	False discovery rate.
Mean Pairs Power	The average of all individual comparison powers.
Min Pairs Power	The minimum of all individual comparison powers.
Max Pairs Power	The maximum of all individual comparison powers.

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 Equal Pairs/Unequal Pairs

The first value is the number of pairs for which the control mean and the treatment mean were equal under H1. The second value is the number of pairs for which the means were different under H1.

Mean No. False Positives

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

Mean No. False Negatives

This is the average number of unequal pairs that were not declared as being unequal 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 equal pairs in the H1 simulations that were declared as unequal.

Prop. Uneq. that were Undet.

This is the proportion of the unequal pairs in the H1 simulations that were not declared as being unequal.

Prop. Detect. that were Equal (FDR)

This is the proportion of detected pairs in the H1 simulations that were actually equal. This is often called the *false discovery rate*.

Prop. Undet. that were Uneq.

This is the proportion of undetected pairs in the H1 simulations that were actually unequal.

All Uneq. Pairs Power

This is the probability of detecting <u>all</u> of the pairs that were different in the H1 simulation.

Any Uneq. Pairs Power

This is the probability of detecting <u>any</u> of the pairs that were different in the H1 simulation.

Mean, Min, and Max Pairs Power

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

Detailed Model Reports

Detailed Model Report for Simulation No. 1
───

Target FWER = 0.05, M0 = 0, M1 = 2, S = 3
MC Procedure: Dunnett's Test
───

Hypo. Type	Groups	Group Labels	n/N	Group Mean	Ave. S.D.	Simulation Model
H0	1	Ctrl	5/20	0	3.0	N(M0 S)
H0	2-4	A1-A3	5/20	0	3.0	N(M0 S)
H0	All			Sm=0	3.0	
H1	1	Ctrl	5/20	0	3.1	N(M0 S)
H1	2-4	A1-A3	5/20	2	3.0	N(M1 S)
H1	All			Sm=0.9	3.0	

───

Hypo. Type	H0 for the alpha simulation and H1 for the power simulation.
Groups	The groups reported on this line. 'All' is used for the average of all groups.
Group Labels	The labels that are used in the individual alpha-level reports.
Cntl	The control group or group 1.
n/N	n is the average group size, and N is the combined sample size.
Sm	The standard deviation of the means. This value is tested with the F-test.
Ave. S.D.	The average standard deviation of all groups reported on this line.
Simulation Model	The distribution used to simulate data for the groups reported on this line.

(More Reports Follow)

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. Note that the control group is labeled 'Ctrl'.

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.

Probability of Rejecting Equality

Probability of Rejecting the Equality of Each Pair. Simulation No. 1

Group	Means	Ctrl	A1	A2	A3
Ctrl	0		0.079*	0.093*	0.09*
A1	2	0.020			
A2	2	0.021			
A3	2	0.020			

Individual pairwise powers from the H1 (Power) simulation are shown in the upper-right section.
Individual pairwise significance levels from the H0 (Alpha) simulation are shown in the lower-left section.
* Starred values are the powers of pairs that are unequal under H1.

(More Reports Follow)

This report shows the individual probabilities of rejecting each pair. When a pair was actually different, the value is the power of that test. These power values are starred.

The results shown on the upper-right section of each simulation report are from the H1 simulation. The results shown on the lower-left section of the report are from the H0 simulation.

Plots Section

Plots
_____
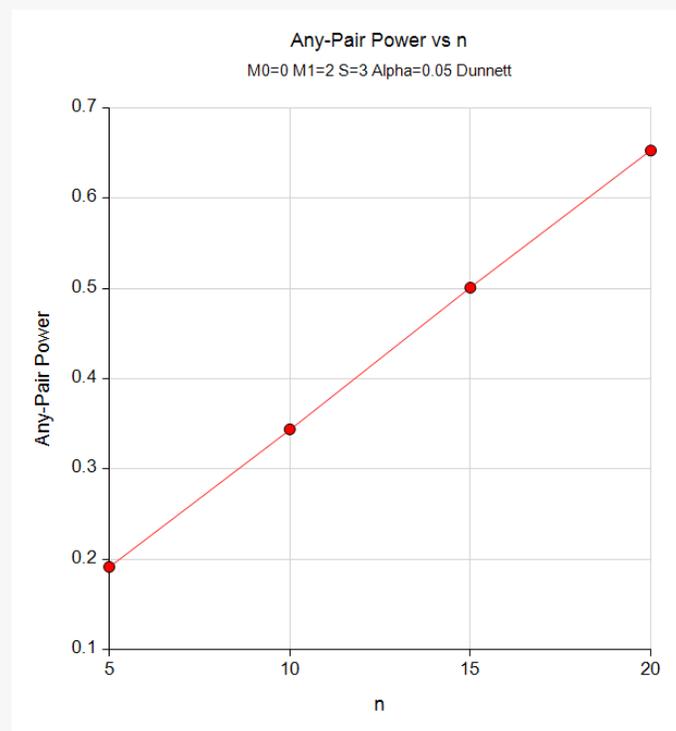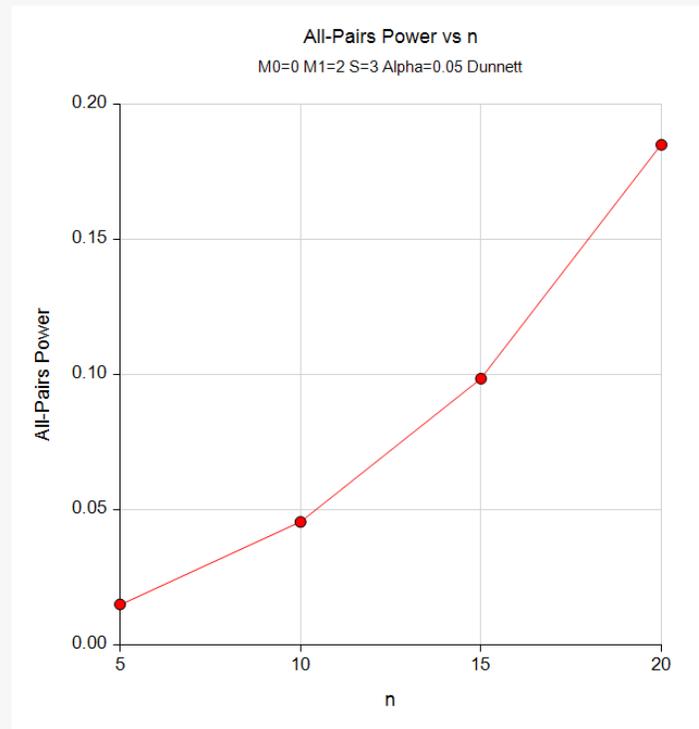
These plots give a visual presentation of the all-pairs power values and the any-pair power values.

Example 2 – Comparative Results

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

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 2** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

Solve For ....**Power**
MC Procedure...**Dunnett Test**
Simulations **2000**
Random Seed...**3645745** (for Reproducibility)
FWER (Alpha)...**0.05**
n (Sample Size Multiplier).............................**5 10 15 20**
Group Sample Size Pattern ....**Equal**
Ctrl Grps **1**
Ctrl Group Distribution(s)|H0........................**Normal(M0 S)**
Ctrl Group Distribution(s)|H1........................**Normal(M0 S)**
Set A Grps **3**
Set A Group Distribution(s)|H0 ....**Normal(M0 S)**
Set A Group Distribution(s)|H1 ....**Normal(M1 S)**
Set B Grps **0**
Equivalence Margin **0.1**
M0 (Mean|H0) Parameter Value(s)...............**0**
M1 (Mean|H1) Parameter Value(s)...............**2**
Parameter 1 Label ....**S**
Parameter 1 Value(s)....................................**3**

Reports Tab

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

Comparative Plots Tab

Comparative All-Pairs Power Plot.................**Checked**
Comparative Any-Pair Power Plot ....**Checked**

Output

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

**Power Comparison for Testing the Control Group Versus the 3 Treatment Groups**

Sim. No.	Total Sample Size	Target Alpha	Dunnett All-Pairs Power	Kruskal Wallis All-Pairs Power	Dunnett Any-Pair Power	Kruskal Wallis Any-Pair Power
1	20	0.05	0.015	0.001	0.176	0.139
2	40	0.05	0.041	0.015	0.371	0.316
3	60	0.05	0.092	0.053	0.512	0.469
4	80	0.05	0.162	0.110	0.630	0.567

Pool Size: 10000. Simulations: 2000. Run Time: 1.59 seconds.
User-Entered Random Seed: 3645745

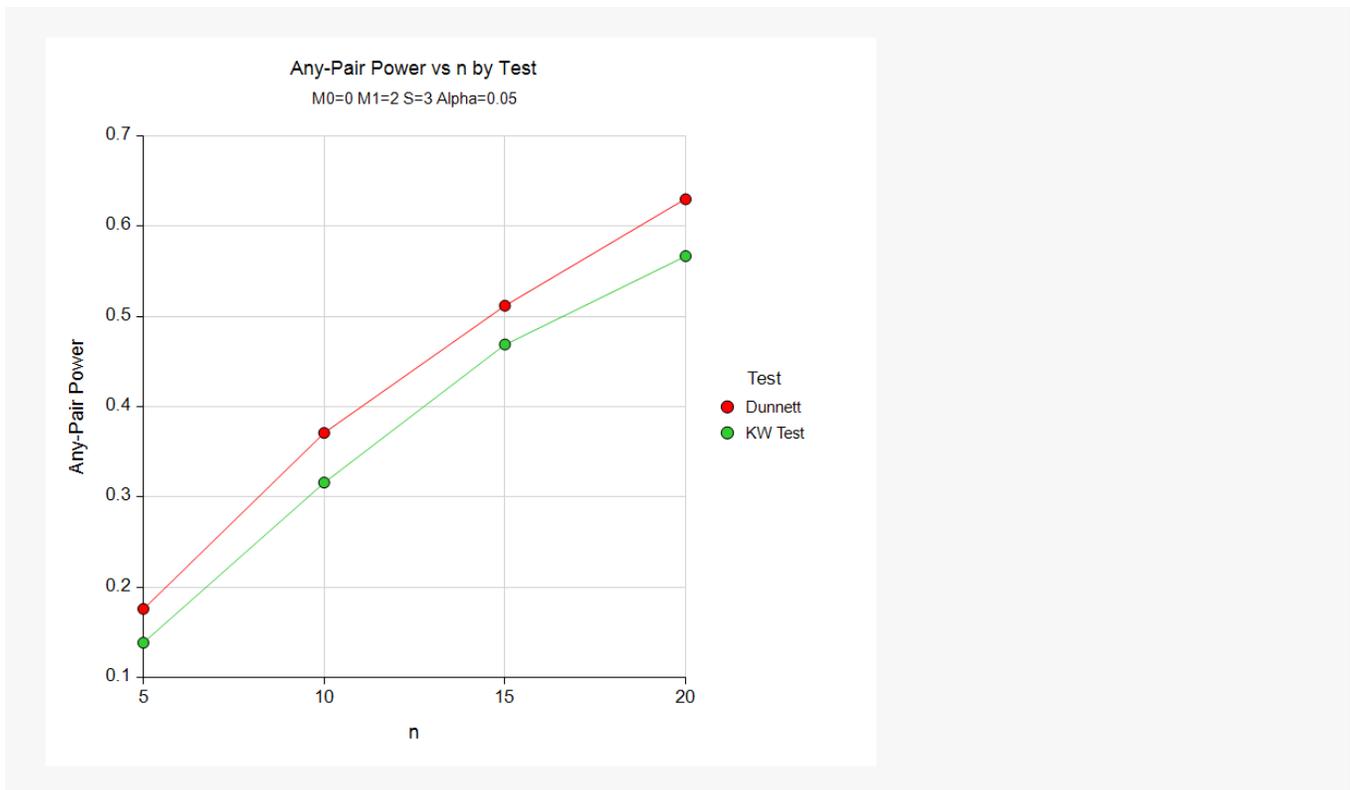**Family-Wise Error-Rate Comparison for Testing the Control Group Versus the 3 Treatment Groups**

Sim. No.	Total Sample Size	Target FWER	Dunnett FWER	Kruskal Wallis FWER
1	20	0.05	0.063	0.050
2	40	0.05	0.053	0.039
3	60	0.05	0.055	0.043
4	80	0.05	0.050	0.043

Plots

All-Pairs Power vs n by Test
M0=0 M1=2 S=3 Alpha=0.05
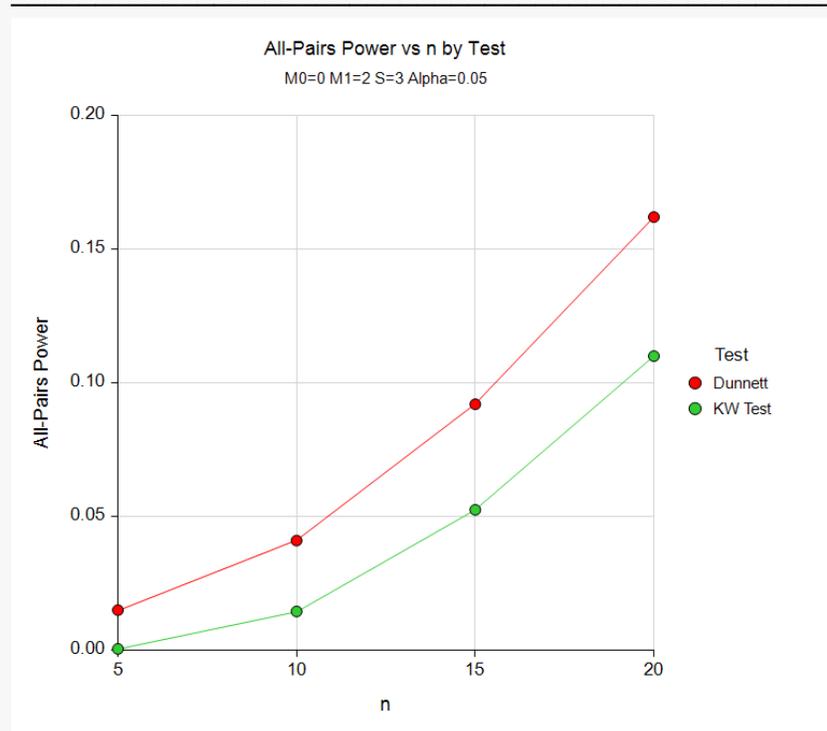
Multiple Comparisons of Treatments vs. a Control (Simulation)

These reports show the power and FWER of both of the multiple comparison procedures. In these simulations of groups from the normal distributions with equal variances, we see that the Dunnett's procedure is the champion.

Example 3 – Validation using Dunnett (1955)

Mukerjee, Robertson, and Wright (1987) page 909 present an example of a sample size calculation which was first discussed on page 1116 of Dunnett (1955). In this example there are five treatments and one control. The control mean and four of the treatment means are the same, while the fifth treatment mean is different.

The value of the within-group standard deviation is 1.0. Four treatment means and the control mean are -0.182574 and the fifth treatment mean is 0.912871. The FWER is 0.05 in the article, but this is for a one-sided test. Since **PASS** is finding the power for two-sided tests, we set FWER at 0.10. When the per-group sample size is 16, the all-pairs power is 0.80. When the per-group sample size is 21, the all-pairs power is 0.90. Note that since there is only one treatment that is different from the control, the all-pairs power is equal to the any-pairs power.

For reproducibility, we'll use a random seed of 5173339.

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 3** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

Solve For	**Power**
MC Procedure...	**Dunnett Test**
Simulations	**2000**
Random Seed...	**5173339** (for Reproducibility)
FWER (Alpha)...	**0.10**
n (Sample Size Multiplier).............................	**16 21**
Group Sample Size Pattern	**Equal**
Ctrl Grps	**1**
Ctrl Group Distribution(s)|H0.........................	**Normal(M0 S)**
Ctrl Group Distribution(s)|H1.........................	**Normal(M0 S)**
Set A Grps	**1**
Set A Group Distribution(s)|H0	**Normal(M0 S)**
Set A Group Distribution(s)|H1	**Normal(M1 S)**
Set B Grps	**4**
Set B Group Distribution(s)|H0	**Normal(M0 S)**
Set B Group Distribution(s)|H1	**Normal(M0 S)**
Equivalence Margin	**0.01**
M0 (Mean|H0) Parameter Value(s)................	**-0.182574**
M1 (Mean|H1) Parameter Value(s)................	**0.912871**
Parameter 1 Label	**S**
Parameter 1 Value(s).....................................	**1**

Output

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

Summary of Simulations of the Control Group and the 5 Treatment Groups

Solve For: Power
MC Procedure: Dunnett's Test

Sim. No.	Any- Pair Power	Group Sample Size n	Total Sample Size N	All- Pairs Power	S.D. of Means Sm|H1	S.D. of Data SD|H1	Actual FWER	Target FWER	M0	M1	S
1	0.798 (0.018) [0.78 0.815]	16	96	0.798 (0.018) [0.78 0.815]	0.4	1	0.097 (0.013) [0.084 0.11]	0.1	-0.2	0.9	1
2	0.906 (0.013) [0.893 0.918]	21	126	0.906 (0.013) [0.893 0.918]	0.4	1	0.111 (0.014) [0.097 0.124]	0.1	-0.2	0.9	1

Pool Size: 10000. Simulations: 2000. Run Time: 1.30 seconds.
User-Entered Random Seed: 5173339

For the first case when n = 16, **PASS** obtained a power of 0.798 which matches the value of 0.80 found by Muterjee et al. (1987). Similarly, the power for the second case when n = 21 is found as 0.906 which matches the article's value of 0.90.