

Chapter 559

Power Comparison of Tests of Means in One-Way Designs (Simulation)

Introduction

This procedure analyzes the power and significance level of tests which may be used to test statistical hypotheses about location (means and medians) in a one-way experimental design. For each scenario that is set up, two simulations are run. One simulation estimates the significance level and the other estimates the power.

The four tests that are compared in this procedure are

1. One-Way ANOVA F-Test
2. Kruskal-Wallis Test of Ranks
3. Terry-Hoeffding Normal Scores Test
4. Van der Waerden Normal Quantiles Test

Technical Details

Computer simulation allows us to estimate the power and significance level that is actually achieved by a test procedure in situations that are not mathematically tractable. Computer simulation was once limited to mainframe computers. But, in recent years, as computer speeds have increased, simulation studies can be completed on desktop and laptop computers in a reasonable period of time.

The steps to a simulation study are

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

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 at least the maximum of twice the number of simulated samples desired and 10,000.

Tests

The details of the four tests that are compared here are found in the chapters that discuss each procedure and they are not duplicated here.

Example 1 – Power at Various Sample Sizes

This example repeats Example 1 of four individual procedures that are compared in this procedure.

An experiment is being designed to compare the means of four groups using an F test with a significance level of 0.05. Previous studies have shown that the standard deviation within a group is 18. Treatment means of 40, 10, 10, and 10 represent clinically important treatment differences. To better understand the relationship between power and sample size, the researcher wants to compute the power for group sample sizes of 4, 8, and 12. The group sample sizes are equal.

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 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

Alpha.....	0.05
Simulations	5000
Random Seed	5850836 (for Reproducibility)
Number of Groups	4
Group Allocation	Equal (n = n1 = n2 = ...)
n (Group Size)	4 8 12
Group 1 Power Distribution.....	Normal(M1 SD)
Group 2 Power Distribution.....	Normal(M2 SD)
Group 3 Power Distribution.....	Normal(M2 SD)
Group 4 Power Distribution.....	Normal(M2 SD)
Parameter 1 Name	M1
Parameter 1 Value(s).....	40
Parameter 2 Name	M2
Parameter 2 Value(s).....	10
Parameter 6 Name	SD
Parameter 6 Value(s).....	18
Specify Alpha Distributions	All equal to the group 1 distribution of the power simulation

Power Comparison of Tests of Means in One-Way Designs (Simulation)

Output

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

Simulation Summary and Comparison Results

Simulation Summary

Group	Simulation Distribution	
	Power (H1)	Alpha (H0)
1	Normal(M1 SD)	Normal(M1 SD)
2	Normal(M2 SD)	Normal(M1 SD)
3	Normal(M2 SD)	Normal(M1 SD)
4	Normal(M2 SD)	Normal(M1 SD)

Number of Groups	4
Random Number Pool Size	10000
Number of Simulations	5000
Random Seed	5850836 (User-Entered)
Run Time	2.85 seconds

Power Comparison of Location Tests

Scenario	Total Sample Size N	Power				Target Alpha	Standard Deviation of μ 's H1	Average of σ 's H1	M1	M2	SD
		Analysis of Variance F-Test	Kruskal-Wallis Test	Terry-Hoeffding Test	Van der Waerden Test						
1	16	0.519	0.373	0.359	0.365	0.05	13	18.0	40	10	18
2	32	0.900	0.851	0.859	0.861	0.05	13	18.1	40	10	18
3	48	0.991	0.983	0.986	0.986	0.05	13	17.7	40	10	18

Scenario	Identifies this line of the report. It will be used cross-reference this line in other reports.
H0	The null hypothesis that all means are equal.
H1	The alternative hypothesis that at least one mean is different from the others.
N	The total sample size of all groups combined.
Power	The probability of rejecting H0 when it is false calculated by the power simulation. It is usually different from the Target Power because of the discrete nature of the sample size, the violation of test assumptions, and the approximate nature of the test.
AOV F-Test	The usual analysis of variance F-test.
Kruskal-Wallis Test	The nonparametric analog of the F-test based on ranks.
Terry-Hoeffding Test	The Terry-Hoeffding expected normal scores test.
Van der Waerden Test	The Van der Waerden normal quantiles test.
Target Alpha	The desired probability of rejecting a true null hypothesis at which the tests were run.
SD of μ 's H1	The standard deviation of the group means assumed by the power simulation. This is a measure of the effect size.
Average of σ 's H1	The mean of the within-group standard deviations assumed during the power simulation. This measures the variation.
M1, M2, SD, etc.	The value(s) of the user-specified distribution parameters.

Power Comparison of Tests of Means in One-Way Designs (Simulation)

Alpha Comparison of Location Tests

Scenario	Total Sample Size N	Alpha					Standard Deviation of μ 's H1	Average of σ 's H1	M1	M2	SD
		Analysis of Variance F-Test	Kruskal- Wallis Test	Terry- Hoeffding Test	Van der Waerden Test	Target					
1	16	0.050	0.036	0.032	0.033	0.05	0	18.0	40	10	18
2	32	0.050	0.044	0.039	0.040	0.05	0	18.1	40	10	18
3	48	0.053	0.045	0.048	0.048	0.05	0	18.1	40	10	18

Scenario	Identifies this line of the report. It will be used cross-reference this line in other reports.
H0	The null hypothesis that all means are equal.
H1	The alternative hypothesis that at least one mean is different from the others.
N	The total sample size of all groups combined.
Alpha	The probability of rejecting H0 when it is true calculated by the alpha simulation.
AOV F-Test	The usual analysis of variance F-test.
Kruskal-Wallis Test	The nonparametric analog of the F-test based on ranks.
Terry-Hoeffding Test	The Terry-Hoeffding expected normal scores test.
Van der Waerden Test	The Van der Waerden normal quantiles test.
Target Alpha	The desired probability of rejecting a true null hypothesis at which the tests were run.
SD of μ 's H1	The standard deviation of the group means assumed by the power simulation. This is a measure of the effect size.
Average of σ 's H1	The mean of the within-group standard deviations assumed during the power simulation. This measures the variation.
M1, M2, SD, etc.	The value(s) of the user-specified distribution parameters.

References

Fleiss, Joseph L. 1986. The Design and Analysis of Clinical Experiments. John Wiley & Sons. New York.
 Gibbons, J.D. and Chakraborti, S. 2011. Nonparametric Statistical Inference, 5th Edition. CRC Press.
 Devroye, Luc. 1986. Non-Uniform Random Variate Generation. Springer-Verlag. New York.
 Matsumoto, M. and Nishimura, T. 1998. 'Mersenne twister: A 623-dimensionally equidistributed uniform pseudorandom number generator.' ACM Trans. On Modeling and Computer Simulations.

These reports show the output for this run. The definitions of each column are shown below the report.

Detailed Results Reports

Details for Scenario 1 of the Numeric Results

Group	ni	Percent of N	μ H0	μ H1	σ H0	σ H1
1	4	25	40	40	18.0	18.1
2	4	25	40	10	18.0	17.9
3	4	25	40	10	17.9	17.9
4	4	25	40	10	17.9	18.0
Group	N		SD of μ 's H0	SD of μ 's H1	Average of σ 's H0	Average of σ 's H1
All	16	100	0	13	18.0	18.0

(More Reports Follow)

These reports show the details of each scenario.

Scenario (in Title)

This is the row number of the Numeric Results report about which this report gives the details.

Power Comparison of Tests of Means in One-Way Designs (Simulation)

Group

This is the number of the group shown on this line.

 n_i

This is the sample size of each group. This column is especially useful when the sample sizes are unequal.

Percent of N

This is the percentage of the total sample that is allocated to each group.

 $\mu \mid H_0$ and $\mu \mid H_1$

These are the means that were used in the alpha and power simulations, respectively.

 $\sigma \mid H_0$ and $\sigma \mid H_1$

These are the standard deviations that were obtained by the alpha and power simulations, respectively. Note that they often are not exactly equal to what was specified because of the error introduced by simulation.

SD of μ 's $\mid H_0$ or H_1

These are the standard deviations of the means that were obtained by the alpha and power simulations, respectively. Under H_0 , this value should be near zero. The H_0 value lets you determine if your alpha simulation was correctly specified. The H_1 value represents the magnitude of the effect size (when divided by an appropriate measure of the standard deviation).

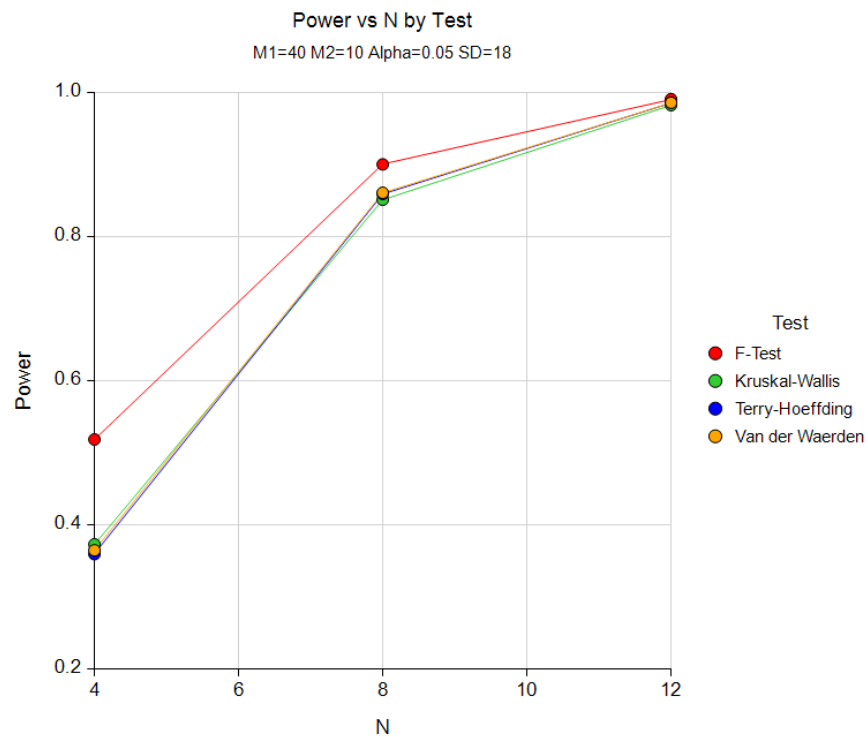
Average of σ 's $\mid H_0$ or H_1

These are the average of the individual group standard deviations that were obtained by the alpha and power simulations, respectively.

Power Comparison of Tests of Means in One-Way Designs (Simulation)

Plots Section

Plots



This plot gives a visual representation of the results in the Numeric Report. We can quickly see the impact on the power of increasing the sample size. We can also see that the F-test is much more power for sample sizes below 8.

Example 2 – Validation using Fleiss (1986)

Fleiss (1986) page 374 presents an example of determining an appropriate sample size when using an F-test in an experiment with 4 groups; means of 9.775, 12, 12, and 14.225; standard deviation of 3; alpha of 0.05, and beta of 0.20. He finds a sample size of 11 per group. The validation runs for the three nonparametric tests found a sample size of 12 per group. Therefore, we generate the results for both 11 and 12 per group.

For reproducibility, we'll use a random seed of 5439669.

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

Alpha.....	0.05
Simulations	5000
Random Seed	5439669 (for Reproducibility)
Number of Groups	4
Group Allocation	Equal (n = n1 = n2 = ...)
n (Group Size)	11 12
Group 1 Power Distribution	Normal(M1 SD)
Group 2 Power Distribution	Normal(M2 SD)
Group 3 Power Distribution	Normal(M3 SD)
Group 4 Power Distribution	Normal(M4 SD)
Parameter 1 Name	M1
Parameter 1 Value(s).....	9.775
Parameter 2 Name	M2
Parameter 2 Value(s).....	12
Parameter 3 Name	M3
Parameter 3 Value(s).....	12
Parameter 4 Name	M4
Parameter 4 Value(s).....	14.225
Parameter 6 Name	SD
Parameter 6 Value(s).....	3
Specify Alpha Distributions	All equal to the group 1 distribution of the power simulation

Power Comparison of Tests of Means in One-Way Designs (Simulation)

Output

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

Simulation Summary

Group	Simulation Distribution	
	Power (H1)	Alpha (H0)
1	Normal(M1 SD)	Normal(M1 SD)
2	Normal(M2 SD)	Normal(M1 SD)
3	Normal(M3 SD)	Normal(M1 SD)
4	Normal(M4 SD)	Normal(M1 SD)

Number of Groups	4
Random Number Pool Size	10000
Number of Simulations	5000
Random Seed	5439669 (User-Entered)
Run Time	2.29 seconds

Power Comparison of Location Tests

Scenario	Total Sample Size N	Power				Target Alpha	Standard Deviation of μ 's H1	Average of σ 's H1	M1	M2	M3	M4	SD
		Analysis of Variance F-Test	Kruskal-Wallis Test	Terry-Hoeffding Test	Van der Waerden Test								
1	44	0.807	0.768	0.776	0.778	0.05	1.6	3	9.8	12	12	14.2	3
2	48	0.843	0.812	0.821	0.822	0.05	1.6	3	9.8	12	12	14.2	3

Alpha Comparison of Location Tests

Scenario	Total Sample Size N	Alpha				Target	Standard Deviation of μ 's H1	Average of σ 's H1	M1	M2	M3	M4	SD
		Analysis of Variance F-Test	Kruskal-Wallis Test	Terry-Hoeffding Test	Van der Waerden Test								
1	44	0.050	0.046	0.043	0.043	0.05	0	3	9.8	12	12	14.2	3
2	48	0.046	0.047	0.043	0.043	0.05	0	3	9.8	12	12	14.2	3

Note that this procedure has also found the power of the F-test to be 0.8 for $n = 11$ ($N = 44$) and the power of the three nonparametric tests to be about 0.8 for $n = 12$ ($N = 48$).

Example 3 – Selecting a Test Statistic when the Data Contain Outliers

The F-test is known to be robust to the violation of some assumptions, but it is susceptible to inaccuracy when the data contain outliers. This example will compare the impact of outliers on the power and alpha of the F-test and the three nonparametric tests.

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 mean 0 and variance 1. The other 5% of the data will come from a normal distribution with mean 0 and variance that ranges from 1 to 10. In the alternative distributions, two will have a mean of 0 and one will have a mean of 1.

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

Alpha.....	0.05
Simulations	5000
Random Seed	1181948 (for Reproducibility)
Number of Groups	3
Group Allocation	Equal (n = n1 = n2 = ...)
n (Group Size)	10 20
Group 1 Power Distribution.....	Normal(M1 SD)[95];Normal(M1 A)[5]
Group 2 Power Distribution.....	Normal(M1 SD)[95];Normal(M1 A)[5]
Group 3 Power Distribution.....	Normal(M2 SD)[95];Normal(M2 A)[5]
Parameter 1 Name	M1
Parameter 1 Value(s).....	0
Parameter 2 Name	M2
Parameter 2 Value(s).....	1
Parameter 3 Name	A
Parameter 3 Value(s).....	1 5 10
Parameter 6 Name	SD
Parameter 6 Value(s).....	1
Specify Alpha Distributions	All equal to the group 1 distribution of the power simulation

Power Comparison of Tests of Means in One-Way Designs (Simulation)

Output

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

Simulation Summary

Simulation Distribution		
Group	Power (H1)	Alpha (H0)
1	Normal(M1 SD)[95];Normal(M1 A)[5]	Normal(M1 SD)[95];Normal(M1 A)[5]
2	Normal(M1 SD)[95];Normal(M1 A)[5]	Normal(M1 SD)[95];Normal(M1 A)[5]
3	Normal(M2 SD)[95];Normal(M2 A)[5]	Normal(M1 SD)[95];Normal(M1 A)[5]
Number of Groups 3		
Random Number Pool Size 10000		
Number of Simulations 5000		
Random Seed 1181948 (User-Entered)		
Run Time 5.88 seconds		

Power Comparison of Location Tests

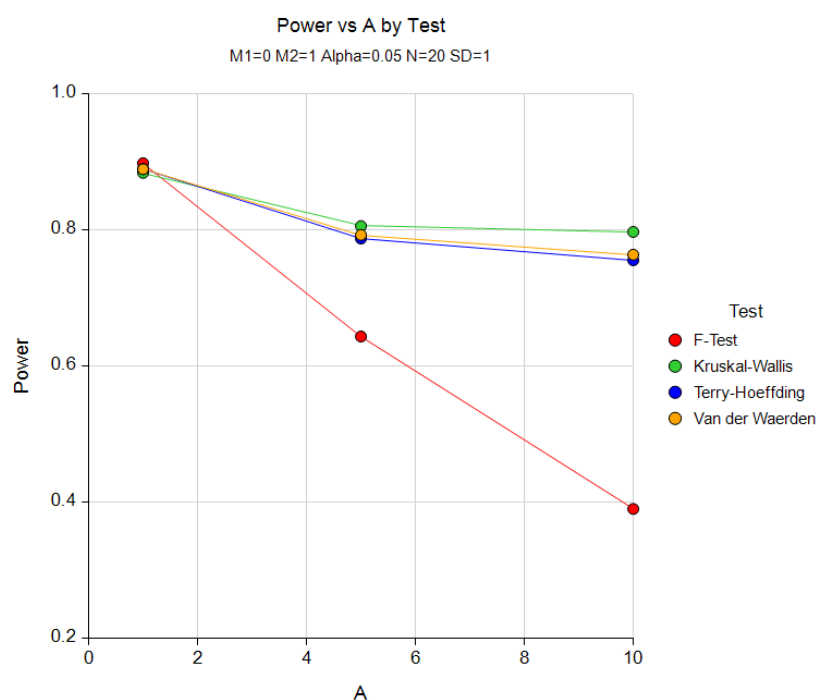
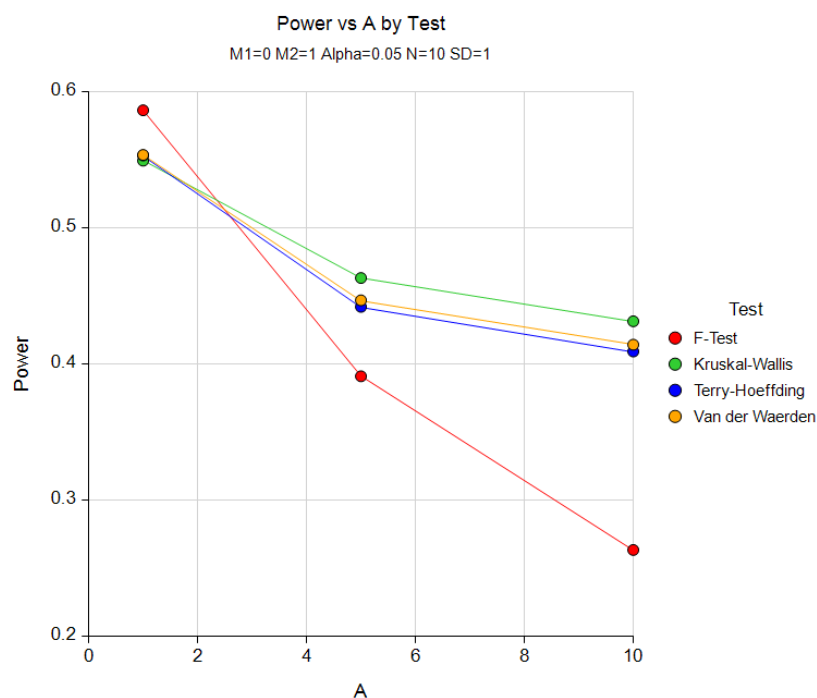
Scenario	Total Sample Size N	Power				Target Alpha	Standard Deviation of μ 's H1	Average of σ 's H1	M1	M2	A	SD
		Analysis of Variance F-Test	Kruskal-Wallis Test	Terry-Hoeffding Test	Van der Waerden Test							
1	30	0.586	0.549	0.553	0.554	0.05	0.5	1.0	0	1	1	1
2	60	0.898	0.884	0.889	0.889	0.05	0.5	1.0	0	1	1	1
3	30	0.391	0.463	0.442	0.447	0.05	0.5	1.5	0	1	5	1
4	60	0.643	0.806	0.787	0.793	0.05	0.5	1.4	0	1	5	1
5	30	0.263	0.431	0.409	0.414	0.05	0.5	2.4	0	1	10	1
6	60	0.390	0.797	0.755	0.764	0.05	0.5	2.5	0	1	10	1

Alpha Comparison of Location Tests

Scenario	Total Sample Size N	Alpha				Target	Standard Deviation of μ 's H1	Average of σ 's H1	M1	M2	A	SD
		Analysis of Variance F-Test	Kruskal-Wallis Test	Terry-Hoeffding Test	Van der Waerden Test							
1	30	0.052	0.049	0.047	0.048	0.05	0	1.0	0	1	1	1
2	60	0.050	0.049	0.047	0.047	0.05	0	1.0	0	1	1	1
3	30	0.038	0.042	0.042	0.043	0.05	0	1.5	0	1	5	1
4	60	0.044	0.051	0.051	0.051	0.05	0	1.5	0	1	5	1
5	30	0.024	0.045	0.045	0.045	0.05	0	2.4	0	1	10	1
6	60	0.027	0.049	0.050	0.050	0.05	0	2.4	0	1	10	1

Power Comparison of Tests of Means in One-Way Designs (Simulation)

Plots



We note that when the variances are equal ($A = 1$), the F-Test is slightly better than the nonparametric tests. However, as the number of outliers is increased, the F-test does increasingly worse than the nonparametric tests, both in terms of power and significance.

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

The F-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 alpha of the F-test and the three nonparametric tests.

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

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 4** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

Alpha.....	0.05
Simulations	5000
Random Seed	2305118 (for Reproducibility)
Number of Groups	3
Group Allocation	Equal (n = n1 = n2 = ...)
n (Group Size)	10 20
Group 1 Power Distribution.....	TukeyGH(M1 SD G 0)
Group 2 Power Distribution.....	TukeyGH(M1 SD G 0)
Group 3 Power Distribution.....	TukeyGH(M2 SD G 0)
Parameter 1 Name	M1
Parameter 1 Value(s).....	0
Parameter 2 Name	M2
Parameter 2 Value(s).....	1
Parameter 3 Name	G
Parameter 3 Value(s).....	0 0.5 0.9
Parameter 6 Name	SD
Parameter 6 Value(s).....	1
Specify Alpha Distributions	All equal to the group 1 distribution of the power simulation

Power Comparison of Tests of Means in One-Way Designs (Simulation)

Output

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

Simulation Summary

Simulation Distribution		
Group	Power (H1)	Alpha (H0)
1	TukeyGH(M1 SD G 0)	TukeyGH(M1 SD G 0)
2	TukeyGH(M1 SD G 0)	TukeyGH(M1 SD G 0)
3	TukeyGH(M2 SD G 0)	TukeyGH(M1 SD G 0)
Number of Groups		
3		
Random Number Pool Size		
10000		
Number of Simulations		
5000		
Random Seed		
2305118 (User-Entered)		
Run Time		
5.79 seconds		

Power Comparison of Location Tests

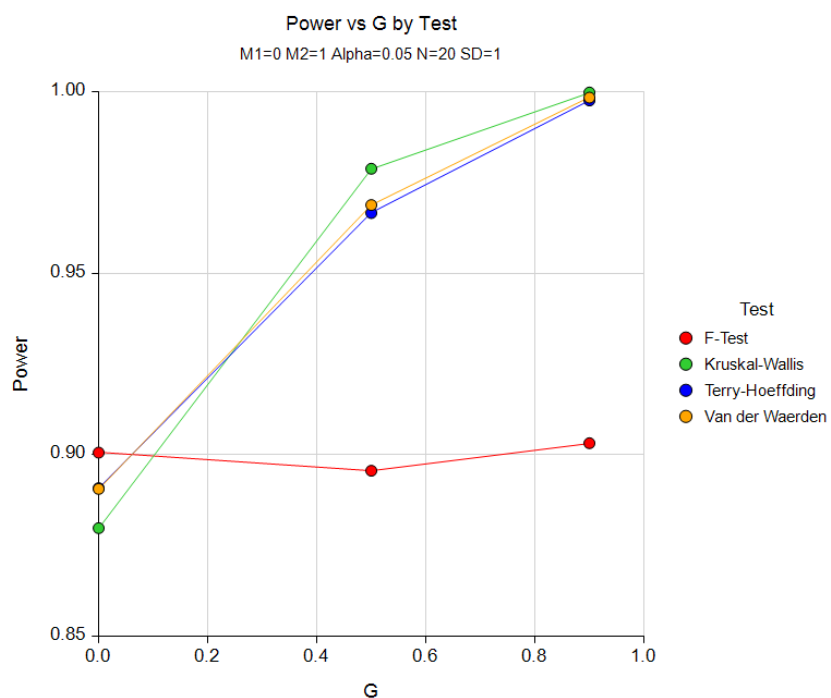
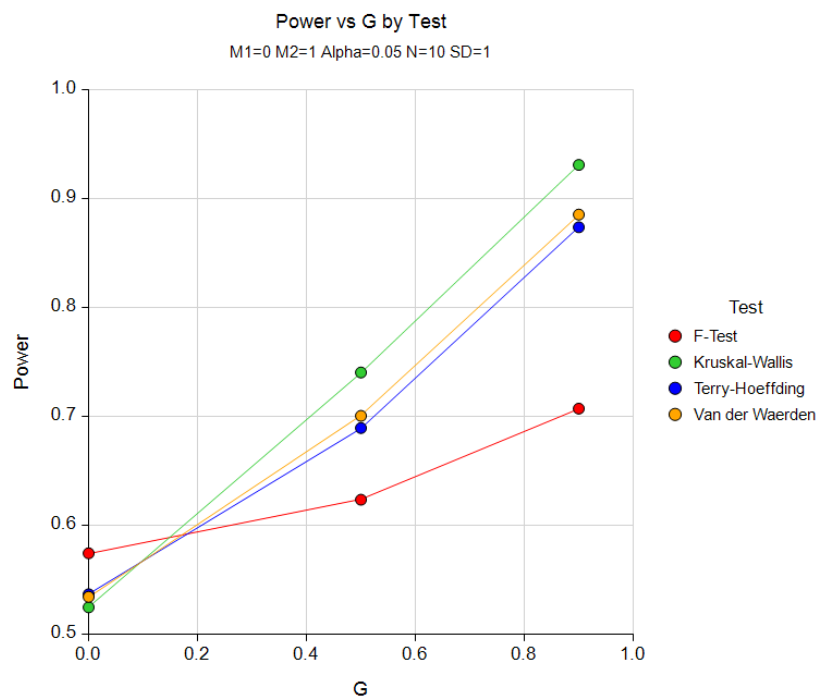
Scenario	Total Sample Size N	Power				Target Alpha	Standard Deviation of μ 's H1	Average of σ 's H1	M1	M2	G	SD
		Analysis of Variance F-Test	Kruskal-Wallis Test	Terry-Hoeffding Test	Van der Waerden Test							
1	30	0.574	0.525	0.537	0.534	0.05	0.5	1	0	1	0.0	1
2	60	0.901	0.880	0.891	0.891	0.05	0.5	1	0	1	0.0	1
3	30	0.624	0.740	0.689	0.701	0.05	0.5	1	0	1	0.5	1
4	60	0.896	0.979	0.967	0.969	0.05	0.5	1	0	1	0.5	1
5	30	0.707	0.931	0.874	0.885	0.05	0.5	1	0	1	0.9	1
6	60	0.903	1.000	0.998	0.998	0.05	0.5	1	0	1	0.9	1

Alpha Comparison of Location Tests

Scenario	Total Sample Size N	Alpha				Target	Standard Deviation of μ 's H1	Average of σ 's H1	M1	M2	G	SD
		Analysis of Variance F-Test	Kruskal-Wallis Test	Terry-Hoeffding Test	Van der Waerden Test							
1	30	0.050	0.046	0.045	0.045	0.05	0	1	0	1	0.0	1
2	60	0.052	0.047	0.047	0.047	0.05	0	1	0	1	0.0	1
3	30	0.042	0.043	0.042	0.042	0.05	0	1	0	1	0.5	1
4	60	0.045	0.050	0.049	0.049	0.05	0	1	0	1	0.5	1
5	30	0.034	0.047	0.045	0.045	0.05	0	1	0	1	0.9	1
6	60	0.037	0.044	0.045	0.045	0.05	0	1	0	1	0.9	1

Power Comparison of Tests of Means in One-Way Designs (Simulation)

Plots



We note that as the skewness increases, the power of the nonparametric tests increase substantially as compared to the F-test. The Kruskal Wallis test is always the best.