User's Guide - III

ANOVA, Multiple Comparisons, Simulator, Variances, Survival Analysis, Correlations, Regression, and Helps

PASS
Power Analysis and Sample Size System

Published by
NCSS
Dr. Jerry L. Hintze
Kaysville, Utah

PASS User’s Guide - III

Direct inquiries to:

NCSS
329 North 1000 East
Kaysville, Utah 84037
Phone (801) 546-0445
Fax (801) 546-3907
Email: support@ncss.com

About this manual

Congratulations on your purchase of the ***PASS*** package! ***PASS*** offers:

- Easy parameter entry.
- A comprehensive list of power analysis routines that are accurate and verified, yet are quick and easy to learn and use.
- Straightforward procedures for creating paper printouts and file copies of both the numerical and graphical reports.

Our goal is that with the help of these user's guides, you will be up and running on ***PASS*** quickly. After reading the quick start manual (at the front of User's Guide I) you will only need to refer to the chapters corresponding to the procedures you want to use. The discussion of each procedure includes one or more tutorials that will take you step-by-step through the tasks necessary to run the procedure.

I believe you will find that these user’s guides provides a quick, easy, efficient, and effective way for first-time ***PASS*** users to get up and running.

I look forward to any suggestions you have to improve the usefulness of this manual and/or the ***PASS*** system. Meanwhile, good computing!

Jerry Hintze, Author

PASS License Agreement

Important: The enclosed Power Analysis and Sample Size software program (PASS) is licensed by NCSS to customers for their use only on the terms set forth below. Purchasing the system indicates your acceptance of these terms.

1. **LICENSE.** NCSS hereby agrees to grant you a non-exclusive license to use the accompanying PASS program subject to the terms and restrictions set forth in this License Agreement.

2. **COPYRIGHT.** PASS and its documentation are copyrighted. You may not copy or otherwise reproduce any part of PASS or its documentation, except that you may load PASS into a computer as an essential step in executing it on the computer and make backup copies for your use on the same computer.

3. **BACKUP POLICY.** PASS may be backed up by you for your use on the same machine for which PASS was purchased.

4. **RESTRICTIONS ON USE AND TRANSFER.** The original and any backup copies of PASS and its documentation are to be used only in connection with a single user. This user may load PASS onto several machines for his/her convenience (such as a desktop and laptop computer), but only for use by the licensee. You may physically transfer PASS from one computer to another, provided that PASS is used in connection with only one user. You may not distribute copies of PASS or its documentation to others. You may transfer this license together with the original and all backup copies of PASS and its documentation, provided that the transferee agrees to be bound by the terms of this License Agreement. PASS licenses may not be transferred more frequently than once in twelve months. Neither PASS nor its documentation may be modified or translated without written permission from NCSS.

*You may not use, copy, modify, or transfer **PASS**, or any copy, modification, or merged portion, in whole or in part, except as expressly provided for in this license.*

5. **NO WARRANTY OF PERFORMANCE.** NCSS does not and cannot warrant the performance or results that may be obtained by using PASS. Accordingly, PASS and its documentation are licensed "as is" without warranty as to their performance, merchantability, or fitness for any particular purpose. The entire risk as to the results and performance of PASS is assumed by you. Should PASS prove defective, you (and not NCSS or its dealer) assume the entire cost of all necessary servicing, repair, or correction.

6. **LIMITED WARRANTY ON CD.** To the original licensee only, NCSS warrants the medium on which PASS is recorded to be free from defects in materials and faulty workmanship under normal use and service for a period of ninety days from the date PASS is delivered. If, during this ninety-day period, a defect in a cd should occur, the cd may be returned to NCSS at its address, or to the dealer from which PASS was purchased, and PASS will replace the cd without charge to you, provided that you have sent a copy of your receipt for PASS. Your sole and exclusive remedy in the event of a defect is expressly limited to the replacement of the cd as provided above.

Any implied warranties of merchantability and fitness for a particular purpose are limited in duration to a period of ninety (90) days from the date of delivery. If the failure of a cd has resulted from accident, abuse, or misapplication of the cd, NCSS shall have no responsibility to replace the cd under the terms of this limited warranty. This limited warranty gives you specific legal rights, and you may also have other rights which vary from state to state.

7. **LIMITATION OF LIABILITY.** Neither NCSS nor anyone else who has been involved in the creation, production, or delivery of PASS shall be liable for any direct, incidental, or consequential damages, such as, but not limited to, loss of anticipated profits or benefits, resulting from the use of PASS or arising out of any breach of any warranty. Some states do not allow the exclusion or limitation of direct, incidental, or consequential damages, so the above limitation may not apply to you.

8. **TERM.** The license is effective until terminated. You may terminate it at any time by destroying PASS and documentation together with all copies, modifications, and merged portions in any form. It will also terminate if you fail to comply with any term or condition of this License Agreement. You agree upon such termination to destroy PASS and documentation together with all copies, modifications, and merged portions in any form.

9. **YOUR USE OF PASS ACKNOWLEDGES** that you have read this customer license agreement and agree to its terms. You further agree that the license agreement is the complete and exclusive statement of the agreement between us and supersedes any proposal or prior agreement, oral or written, and any other communications between us relating to the subject matter of this agreement.

Dr. Jerry L. Hintze & **NCSS**, Kaysville, Utah

Preface

PASS (**P**ower **A**nalysis and **S**ample **S**ize) is an advanced, easy-to-use statistical analysis software package. The system was designed and written by Dr. Jerry L. Hintze over the last fifteen years. Dr. Hintze drew upon his experience both in teaching statistics at the university level and in various types of statistical consulting.

The present version, written for 32-bit versions of Microsoft Windows (98, 2000, ME, NT, XP, etc.) computer systems, is the result of several iterations. Experience over the years with several different types of users has helped the program evolve into its present form.

NCSS maintains a website at WWW.NCSS.COM where we make the latest edition of ***PASS*** available for free downloading. The software is password protected, so only users with valid serial numbers may use this downloaded edition. We hope that you will download the latest edition routinely and thus avoid any bugs that have been corrected since you purchased your copy.

We believe ***PASS*** to be an accurate, exciting, easy-to-use program. If you find any portion which you feel needs to be changed, please let us know. Also, we openly welcome suggestions for additions and enhancements.

Verification

All calculations used in this program have been extensively tested and verified. First, they have been verified against the original journal article or textbook that contained the formulas. Second, they have been verified against second and third sources when these exist.

Table of Contents

Chapter 550

One-Way ANOVA

Introduction

A common task in research is to compare the averages of two or more populations (groups). We might want to compare the income level of two regions, the nitrogen content of three lakes, or the effectiveness of four drugs. The one-way analysis of variance compares the means of two or more groups to determine if at least one mean is different from the others. The *F* test is used to determine statistical significance. *F* tests are nondirectional in that the null hypothesis specifies that all means are equal and the alternative hypothesis simply states that at least one mean is different.

The methods described here are usually applied to the one-way experimental design. This design is an extension of the design used for the two-sample *t* test. Instead of two groups, there are three or more groups. With careful modifications, this procedure may be used to test interaction terms as well.

Planned Comparisons

PASS performs power and sample size calculations for user-specified contrasts.

The usual *F* test tests the hypothesis that all means are equal versus the alternative that at least one mean is different from the rest. Often, a more specific alternative is desired. For example, you might want to test whether the treatment means are different from the control mean, the low dose is different from the high dose, a linear trend exists across dose levels, and so on. These questions are tested using planned comparisons.

We call the comparison *planned* because it was determined before the experiment was conducted. We planned to test the comparison.

A comparison is a weighted average of the means, in which the weights may be negative. When the weights sum to zero, the comparison is called a *contrast*. ***PASS*** provides results for contrasts. To specify a contrast, we need only specify the weights. Statisticians call these weights the *contrast coefficients*.

For example, suppose an experiment conducted to study a drug will have three dose levels: none (control), 20 mg., and 40 mg. The first question is whether the drug made a difference. If it did, the average response for the two groups receiving the drug should be different from the control. If we label the group means M0, M20, and M40, we are interested in comparing M0 with M20 and M40. This can be done in two ways. One way is to construct two tests, one comparing M0 and M20 and the other comparing M0 and M40. Another method is to perform one test comparing M0 with the average of M20 and M40. These tests are conducted using planned comparisons. The coefficients are as follows:

M0 vs. M20

To compare M0 versus M20, use the coefficients -1,1,0. When applied to the group means, these coefficients result in the comparison M0(-1)+M20(1)+M40(0) which reduces to M20-M0. That is, this contrast results in the difference between the two group means. We can test whether this difference is non-zero using the *t* test (or *F* test since the square of the *t* test is an *F* test).

M0 vs. M40

To compare M0 versus M40, use the coefficients -1,0,1. When applied to the group means, these coefficients result in the comparison M0(-1)+M20(0)+M40(1) which reduces to M40-M0. That is, this contrast results in the difference between the two group means.

M0 vs. Average of M20 and M40

To compare M0 versus the average of M20 and M40, use the coefficients -2,1,1. When applied to the group means, these coefficients result in the comparison M0(-2)+M20(1)+M40(1) which is equivalent to M40+M20-2(M0).

To see how these coefficients were obtained, consider the following manipulations. Beginning with the difference between the average of M20 and M40 and M0, we obtain the coefficients above—aside from a scale factor of one-half.

$$\begin{aligned}\frac{M20+M40}{2}-M0&=\frac{M20}{2}+\frac{M40}{2}-\frac{M0}{1}\\&=\frac{1}{2}M20+\frac{1}{2}M40-M0\\&=\frac{1}{2}\left(M20+M40-2M0\right)\end{aligned}$$

Assumptions

Using the *F* test requires certain assumptions. One reason for the popularity of the *F* test is its robustness in the face of assumption violation. However, if an assumption is not even approximately met, the significance levels and the power of the *F* test are invalidated. Unfortunately, in practice it often happens that several assumptions are not met. This makes matters even worse. Hence, steps should be taken to check the assumptions before important decisions are made.

The assumptions of the one-way analysis of variance are:

1. The data are continuous (not discrete).
2. The data follow the normal probability distribution. Each group is normally distributed about the group mean.
3. The variances of the populations are equal.
4. The groups are independent. There is no relationship among the individuals in one group as compared to another.
5. Each group is a simple random sample from its population. Each individual in the population has an equal probability of being selected in the sample.

Technical Details for the One-Way AOV

Suppose k groups each have a normal distribution and equal means $(\mu_1 = \mu_2 = \cdots = \mu_k)$. Let $n_1 = n_2 = \cdots = n_k$ denote the number of subjects in each group and let N denote the total sample size of all groups. Let $\bar{\mu}_W$ denote the weighted mean of all groups. That is

$$\bar{\mu}_W = \sum_{i=1}^{k} \left(\frac{n_i}{N} \right) \mu_i$$

Let σ denote the common standard deviation of all groups.

Given the above terminology, the ratio of the mean square between groups to the mean square within groups follows a central F distribution with two parameters matching the degrees of freedom of the numerator mean square and the denominator mean square. When the null hypothesis of mean equality is rejected, the above ratio has a noncentral F distribution which also depends on the noncentrality parameter, λ. This parameter is calculated as

$$\lambda = N \frac{\sigma_m^2}{\sigma^2}$$

where

$$\sigma_m = \sqrt{\sum_{i=1}^{k} \frac{n_i \left(\mu_i - \bar{\mu}_W \right)^2}{N}}$$

Some authors use the symbol ϕ for the noncentrality parameter. The relationship between the two noncentrality parameters is

$$\phi = \sqrt{\frac{\lambda}{k}}.$$

The process of planning an experiment should include the following steps:

1. Determine an estimate of the within group standard deviation, σ. This may be done from prior studies, from experimentation with the Standard Deviation Estimation module, from pilot studies, or from crude estimates based on the range of the data. See the chapter on estimating the standard deviation for more details.

2. Determine a set of means that represent the group differences that you want to detect.

3. Determine the appropriate group sample sizes that will ensure desired levels of α and β. Although it is tempting to set all group sample sizes equal, it is easy to show that putting more subjects in some groups than in others may have better power than keeping group sizes equal (see Example 4).

Power Calculations for One-Way AOV

The calculation of the power of a particular test proceeds as follows:

1. Determine the critical value, $F_{k-1,N-k,\alpha}$ where α is the probability of a type-I error and k and N are defined above. Note that this is a two-tailed test as no direction is assigned in the alternative hypothesis.
2. From a hypothesized set of μ_i's, calculate the noncentrality parameter λ based on the values of N, k, σ_m, and σ.
3. Compute the power as the probability of being greater than $F_{k-1,N-k,\alpha}$ on a noncentral-F distribution with noncentrality parameter λ.

Technical Details for a Planned Comparison

The terminology of planned comparisons is identical to that of the one-way AOV, so the notation used above will be repeated here.

Suppose you want to test whether the contrast C

$$C = \sum_{i=1}^{k} c_i \mu_i$$

is significantly different from zero. Here the c_i's are the contrast coefficients.

Define

$$\sigma_{mc} = \frac{\left| \sum_{i=1}^{k} c_i \mu_i \right|}{\sqrt{N \sum_{i=1}^{k} \frac{c_i^2}{n_i}}}$$

Define the noncentrality parameter λ_c, as

$$\lambda_C = N \frac{\sigma_{mc}^2}{\sigma^2}$$

Power Calculations for Planned Comparisons

The calculation of the power of a particular test proceeds as follows:

1. Determine the critical value, $F_{1,N-k,\alpha}$ where α is the probability of a type-I error and k and N are defined above. Note that this is a two-tailed test as no direction is assigned in the alternative hypothesis.
2. From a hypothesized set of μ_i's, calculate the noncentrality parameter λ_c based on the values of N, k, σ_{mc}, and σ.
3. Compute the power as the probability of being greater than $F_{1,N-k,\alpha}$ on a noncentral-*F* distribution with noncentrality parameter λ_c.

Procedure Options

This section describes the options that are unique to this procedure. These are located on the panels associated with the Data, Options, and Reports tabs. To find out more about using the other tabs such as Plot Text, Axes, and Template, turn to the chapter entitled Procedure Templates.

Data Tab

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

Find

This option specifies the parameter to be solved for from the other parameters. The parameters that may be selected are *SM*, *S*, *k*, *n*, Alpha, and Beta. Under most situations, you will select either Beta for a power analysis or *n* for sample size determination.

k (Number of Groups)

This is the number of group means being compared. It must be greater than or equal to two.

You can enter a list of values, in which case, a separate analysis will be calculated for each value. Commas or blanks may separate the numbers. A TO-BY list may be used.

Note that the number of items used in the Hypothesized Means box and the Group Sample Size Pattern box is controlled by this number.

Examples:

2,3,4

2 3 4

2 to 10 by 2

Hypothesized Means

Enter a set of hypothesized means, one for each group. These means represent the group centers under the alternative hypothesis (the null hypothesis is that they are equal). The standard deviation of these means (*SM*) is used in the power calculations to represent the average size of the differences among the means. The standard deviation of the means is calculated using the formula:

$$\sigma_m = \sqrt{\sum_{i=1}^{k} \frac{(\mu_i - \overline{\mu})^2}{k}}$$

This quantity gives the magnitude of the differences among the group means. Note that when all means are equal, σ_m is zero.

You should enter a set of means that give the pattern of differences you expect or the pattern that you wish to detect. For example, in a particular study involving three groups, your research might be “meaningful” if either of two treatment means is 50% larger than the control mean. If the control mean is 50, then you would enter *50,75,75* as the three means.

It is usually more intuitive to enter a set of mean values. However, it is possible to enter the standard deviation of the means directly by placing an *S* in front of the number (see below).

Some might wish to specify the alternative hypothesis as the effect size, *f*, which is defined as

$$f = \frac{\sigma_m}{\sigma}$$

If so, set $\sigma = 1$ and $\sigma_m = f$. Cohen (1988) has designated values of *f* less than 0.1 as *small*, values around 0.25 to be *medium*, and values over 0.4 to be *large*.

Entering a list of means

If a set of numbers is entered without a leading *S*, they are assumed to be the hypothesized group means under the alternative hypothesis. Their standard deviation will be calculated and used in the calculations. Blanks or commas may separate the numbers. Note that it is not the values of the means themselves that is important, but only their differences. Thus, the mean values *0,1,2* produce the same results as the values *100,101,102*.

If too few means are entered to match the number of groups, the last mean is repeated. For example, suppose that four means are needed and you enter *1,2* (only two means). ***PASS*** will treat this as *1,2,2,2*. If too many values are entered, ***PASS*** will truncate the list to the number of means needed.

Examples:

5 20 60

2,5,7

-4,0,6,9

S Option

If an *S* is entered before the list of numbers, they are assumed to be values of σ_m, the standard deviations of the group means. A separate power calculation is made for each value. Note that this list can be a TO-BY phrase.

Examples:

S 4.7

S 4.3 5.7 4.2

S 10 to 20 by 2

Contrast Coefficients

If you want to analyze a specific planned comparison, enter a set of contrast coefficients here. The calculations will then refer to the hypothesis that the corresponding contrast of the means is zero versus the alternative that it is non-zero (two-sided test). These are often called Planned Comparisons.

A contrast is a weighted average of the means in which the weights sum to zero. For example, suppose you are studying four groups and that the main hypothesis of interest is whether there is a linear trend across the groups. You would enter -*3, -1, 1, 3* here. This would form the weighted average of the means:

-3(Mean1)-(Mean2)+(Mean3)+3(Mean3)

The point to realize is that these numbers (the coefficients) are used to calculate a specific weighted average of the means which is to be compared against zero using a standard *F* (or *t*) test.

NONE or blank

When the box is left blank or the word *None* is entered, this option is ignored.

Linear Trend

A set of coefficients is generated appropriate for testing the alternative hypothesis that there is a linear (straight-line) trend across the means. These coefficients assume that the means are equally spaced across the trend variable.

Quadratic

A set of coefficients is generated appropriate for testing the alternative hypothesis that the means follow a quadratic model. These coefficients assume that the means are equally spaced across the implicit *X* variable.

Cubic

A set of coefficients is generated appropriate for testing the alternative hypothesis that the means follow a cubic model. These coefficients assume that the means are equally spaced across the implicit *X* variable.

First Against Others

A set of coefficients is generated appropriate for testing the alternative hypothesis that the first mean is different from the average of the remaining means. For example, if there were four groups, the generated coefficients would be -*3, 1, 1, 1*.

List of Coefficients

A list of coefficients, separated by commas or blanks, may be entered. If the number of items in the list does not match the number of groups (*k*), zeros are added or extra coefficients are truncated.

Remember that these coefficients must sum to zero. Also, the scale of the coefficients does not matter. That is *0.5,0.25,0.25*; *-2,1,1*; and *-200,100,100* will yield the same results.

To avoid rounding problems, it is better to use *-3,1,1,1* than the equivalent *-1,0.333,0.333,0.333*. The second set does not sum to zero.

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 times 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

A set of positive, numeric values, one for each group, is entered here. The sample size of group *i* is found by multiplying the i^{th} number from this list times the value of n and rounding up to the next whole number. The number of values must match the number of groups, *k*. 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 *k* 1's will be used. This will result in $N1 = N2 = N3 = n$. That is, all sample sizes are equal to *n*.

Alpha

This option specifies one or more values for the probability of a type-I error (alpha). A type-I error occurs when you reject the null hypothesis of equal means when in fact the means are equal.

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

Beta

This option specifies one or more values for the probability of a type-II error (beta). A type-II error occurs when you fail to reject the null hypothesis of equal means when in fact the means are different.

Values must be between zero and one. Historically, the value of 0.20 was often used for beta. However, you should pick a value for beta that represents the risk of a type-II error you are willing to take.

Power is defined as one minus beta. Power is equal to the probability of rejecting a false null hypothesis. Hence, specifying the beta error level also specifies the power level. For example, if you specify beta values of 0.05, 0.10, and 0.20, you are specifying the corresponding power values of 0.95, 0.90, and 0.80.

S - Std Dev of Subjects

This is σ, the standard deviation within a group. It represents the variability from subject to subject that occurs when the subjects are treated identically. It is assumed to be the same for all groups. This value is approximated in an analysis of variance table by the square root of the mean square error.

Since they are positive square roots, the numbers must be strictly greater than zero. You can press the *SD* button to obtain further help on estimating the standard deviation.

Note that if you are using this procedure to test a factor (such as an interaction) from a more complex design, the value of standard deviation is estimated by the square root of the mean square of the term that is used as the denominator in the *F* test.

You can enter a list of values separated by blanks or commas, in which case, a separate analysis will be calculated for each value.

Examples of valid entries:

1,4,7,10

1 4 7 10

1 to 10 by 3

Example 1 - Finding the Statistical Power

An experiment is being designed to compare the means of four groups using an *F* test with a significance level of either 0.01 or 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 several group sample sizes between 2 and 14. The sample sizes will be equal across all groups.

Setup

Option	Value
Data Tab	
Find	**Beta and Power**
k (Number of Groups)	**4**
Hypothesized Means	**40 10 10 10**
Contrast Coefficients	**None**
n (Sample Size Multiplier)	**2 to 14 by 2**
Group Sample Size Pattern	**Equal**
Alpha	**0.01, 0.05**
Beta	*Ignored since this is the Find setting*
S (Std Dev of Subjects)	**18**

Annotated Output

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

Numeric Results Report

Numeric Results

Power	Average n	k	Total N	Alpha	Beta	Std Dev of Means (Sm)	Standard Deviation (S)	Effect Size
0.04238	2.00	4	8	0.01000	0.95762	12.99	18.00	0.7217
0.17513	2.00	4	8	0.05000	0.82487	12.99	18.00	0.7217
0.23886	4.00	4	16	0.01000	0.76114	12.99	18.00	0.7217
0.52165	4.00	4	16	0.05000	0.47835	12.99	18.00	0.7217
0.50581	6.00	4	24	0.01000	0.49419	12.99	18.00	0.7217
0.77327	6.00	4	24	0.05000	0.22673	12.99	18.00	0.7217
0.72695	8.00	4	32	0.01000	0.27305	12.99	18.00	0.7217
0.90642	8.00	4	32	0.05000	0.09358	12.99	18.00	0.7217
0.86702	10.00	4	40	0.01000	0.13298	12.99	18.00	0.7217
0.96514	10.00	4	40	0.05000	0.03486	12.99	18.00	0.7217
0.94143	12.00	4	48	0.01000	0.05857	12.99	18.00	0.7217
0.98802	12.00	4	48	0.05000	0.01198	12.99	18.00	0.7217
0.97623	14.00	4	56	0.01000	0.02377	12.99	18.00	0.7217
0.99614	14.00	4	56	0.05000	0.00386	12.99	18.00	0.7217

Report Definitions

Power is the probability of rejecting a false null hypothesis. It should be close to one.
n is the average group sample size.
k is the number of groups.
Total N is the total sample size of all groups.
Alpha is the probability of rejecting a true null hypothesis. It should be small.
Beta is the probability of accepting a false null hypothesis. It should be small.
Sm is the standard deviation of the group means under the alternative hypothesis.
Standard deviation is the within group standard deviation.
The Effect Size is the ratio of Sm to standard deviation.

Summary Statements

In a one-way ANOVA study, sample sizes of 2, 2, 2, and 2 are obtained from the 4 groups whose means are to be compared. The total sample of 8 subjects achieves 4% power to detect differences among the means versus the alternative of equal means using an F test with a 0.01000 significance level. The size of the variation in the means is represented by their standard deviation which is 12.99. The common standard deviation within a group is assumed to be 18.00.

This report shows the numeric results of this power study. Following are the definitions of the columns of the report.

Power

The probability of rejecting a false null hypothesis.

Average n

The average of the group sample sizes.

k

The number of groups.

Total N

The total sample size of the study.

Alpha

The probability of rejecting a true null hypothesis. This is often called the significance level.

Beta

The probability of accepting a false null hypothesis that *Sm* is zero when *Sm* is actually equal to the value shown in the next column.

Std Dev of Means

This is the standard deviation of the hypothesized means. It was computed from the hypothesized means. It is roughly equal to the average difference between the group means and the overall mean.

Once you have computed this, you can enter a range of values to determine the effect of the hypothesized means on the power.

Standard Deviation (S)

This is the within-group standard deviation. It was set in the Data window.

Effect Size

The effect size is the ratio of *SM* to *S*. It is an index of relative difference between the means that can be compared from study to study.

Detail Results Report

Details when Alpha = 0.01000, Power = 0.04238, SM = 12.99, S = 18.00

Group	Ni	Percent Ni of Total Ni	Mean	Deviation From Mean	Ni Times Deviation
1	2	25.00	40.00	22.50	45.00
2	2	25.00	10.00	7.50	15.00
3	2	25.00	10.00	7.50	15.00
4	2	25.00	10.00	7.50	15.00
ALL	8	100.00	17.50		

Details when Alpha = 0.05000, Power = 0.17513, SM = 12.99, S = 18.00

Group	Ni	Percent Ni of Total Ni	Mean	Deviation From Mean	Ni Times Deviation
1	2	25.00	40.00	22.50	45.00
2	2	25.00	10.00	7.50	15.00
3	2	25.00	10.00	7.50	15.00
4	2	25.00	10.00	7.50	15.00
ALL	8	100.00	17.50		

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

Group

The number of the group shown on this line. The last line, labeled *ALL*, gives the average or the total as appropriate.

Ni

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

Percent Ni of Total Ni

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

Mean

The is the value of the Hypothesized Mean. The final row gives the average for all groups.

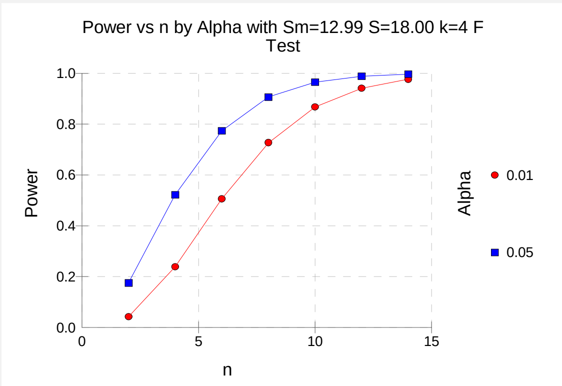
Deviation From Mean

This is the absolute value of the mean minus the overall mean. Since *Sm* is the sum of the squared deviations, these values show the relative contribution to *Sm*.

Ni Times Deviation

This is the group sample size times the absolute deviation. It shows the combined influence of the size of the deviation and the sample size on Sm.

Plot Section



This plot gives a visual presentation to the results in the Numeric Report. We can quickly see the impact on the power of increasing the sample size and increase the significance level.

When you create one of these plots, it is important to use trial and error to find an appropriate range for the horizontal variable so that you have results with both low and high power.

Example 2 - Power after a Study

This example will cover the situation in which you are calculating the power of a one-way analysis of variance *F* test on data that have already been collected and analyzed.

An experiment included a control group and two treatment groups. Each group had seven individuals. A single response was measured for each individual and recorded in the following table.

Control	T1	T2
452	646	685
674	547	658
554	774	786
447	465	536
356	759	653
654	665	669
558	767	557

When analyzed using the one-way analysis of variance procedure in ***NCSS***, the following results were obtained.

Analysis of Variance Table

Source Term	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level
A (...)	2	75629.8	37814.9	3.28	0.061167
S(A)	18	207743.4	11541.3		
Total (Adjusted)	20	283373.3			
Total	21				

Means Section

Group	Count	Mean
Control	7	527.8571
T1	7	660.4286
T2	7	649.1429

The significance level (Prob Level) was only 0.061—not enough for statistical significance. The researcher had hoped to show that the treatment groups had higher response levels than the control group. He could see that the group means followed this pattern since the mean for *T1* was about 25% higher than the control mean and the mean for *T2* was about 23% higher than the control mean. He decided to calculate the power of the experiment using these values of the means. (We do not recommend this approach because the power should be calculated for the minimum difference among the means that is of interest, not at the values of the sample means.)

The data entry for this problem is simple. The only entry that is not straight forward is finding an appropriate value for the standard deviation. Since the standard deviation is estimated by the square root of the mean square error, it is calculated as $\sqrt{11541.3} = 107.4304$.

Setup

You can enter these values yourself or load the Example2 template from the Template tab.

Option	Value
Data Tab	
Find	**Beta and Power**
k (Number of Groups)	**3**
Hypothesized Means	**527.8571 660.4286 649.1429**
Contrast Coefficients	**None**
n (Sample Size Multiplier)	**7**
Group Sample Size Pattern	**Equal**
Alpha	**0.05**
Beta	*Ignored since this is the Find setting*
S (Std Dev of Subjects)	**107.4304**

Annotated Output

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

Numeric Results

Numeric Results

Power	Average n	k	Total N	Alpha	Beta	Std Dev of Means (Sm)	Standard Deviation (S)	Effect Size
0.54788	7.00	3	21	0.05000	0.45212	60.01	107.43	0.5586

The power is only 0.55. That is, there was only a 55% chance of rejecting the false null hypothesis. It is important to understand this power statement is conditional, so we will state it in detail. Given that the population means are equal to the sample means (that *Sm* is 60.01) and the population standard deviation is equal to 107.43, the probability of rejecting the false null hypothesis is 0.55. If the population means are different from the sample means (which they must be), the power is different. However, the sample means provide a reasonable place to begin.

Example 3 - Finding the Sample Size Necessary to Reject

Continuing with the last example, we will determine how large the sample size would need to have been for alpha = 0.05 and beta = 0.20.

Setup

This section presents the values of each of the parameters needed to run this example. You can enter these values yourself or load the Example3 template from the Template tab.

Option	Value
Data Tab	
Find	**n (Sample Size)**
k (Number of Groups)	**3**
Hypothesized Means	**527.8571 660.4286 649.1429**
Contrast Coefficients	**None**
n (Sample Size Multiplier)	*Ignored*
Group Sample Size Pattern	**Equal**
Alpha	**0.05**
Beta	**0.20**
S (Std Dev of Subjects)	**107.4304**

Annotated Output

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

Numeric Results

Numeric Results for One-Way Analysis of Variance

Power	Average n	k	Total N	Alpha	Beta	Std Dev of Means (Sm)	Standard Deviation (S)	Effect Size
0.82511	12.00	3	36	0.05000	0.17489	60.01	107.43	0.5586

The required sample size is 12 per group or 36 subjects.

Example 4 - Using Unequal Sample Sizes

Continuing with the last example, consider the impact of allowing the group sample sizes to be unequal. Since the control group is being compared to two treatment groups, the mean of the control group is assumed to be different from those of the treatment groups. In this situation, experience has shown that adding extra subjects to the control group can increase the power. In a separate analysis, the power with 11 subjects per group was found to be 0.7851—not quite the required 0.80.

We will try moving two subjects from each treatment group into the control group. This will give an experimental design with 15 in the control group and 9 in each of the treatment groups.

Setup

Pay particular attention to how the sample size parameters were changed. The value of n is set to one so that it is essentially ignored. The Group Sample Size Pattern contains the three unequal sample sizes.

Option	Value
Data Tab	
Find	**Beta and Power**
k (Number of Groups)	**3**
Hypothesized Means	**527.8571 660.4286 649.1429**
Contrast Coefficients	**None**
n (Sample Size Multiplier)	**1**
Group Sample Size Pattern	**15 9 9**
Alpha	**0.05**
Beta	**0.20**
S (Std Dev of Subjects)	**107.4304**

Annotated Output

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

Numeric Results

Numeric Results

Power	Average n	k	Total N	Alpha	Beta	Std Dev of Means (Sm)	Standard Deviation (S)	Effect Size
0.82967	11.00	3	33	0.05000	0.17033	63.34	107.43	0.5896

The power of 0.82967 achieved with the 33 subjects in this design is slightly higher than the power of 0.82511 that was achieved with the 36 subjects in the equal group size design. Apparently, unequal sample allocation can achieve better power!

We suggest that you try several different sample allocations. You will find that the optimum sample allocation depends on the values of the hypothesized means.

You should keep in mind that power may not be the only goal of the experiment. Other goals may include finding confidence intervals for each of the group means. And the narrowness of the width of the confidence interval is directly related to the sample size.

Example 5 - Minimum Detectable Difference

Background

It may be useful to determine the minimum detectable difference among the means that can be found at the experimental conditions. This amounts to finding σ_m (which we call *Sm* on the windows and printouts).

Problem Statement

Continuing with the previous example, find *Sm* for a wide range of sample sizes when alpha is 0.05 and beta is 0.10 or 0.20.

Setup

This section presents the values of each of the parameters needed to run this example. You can enter these values yourself or load the Example5 template from the Template tab.

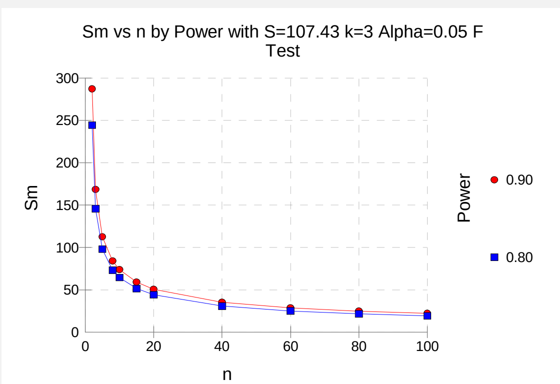
Option	Value
Data Tab	
Find	**Sm (Std Dev of Means)**
k (Number of Groups)	**3**
Hypothesized Means	*Ignored*
Contrast Coefficients	**None**
n (Sample Size Multiplier)	**2 3 5 8 10 15 20 40 60 80 100**
Group Sample Size Pattern	**Equal**
Alpha	**0.05**
Beta	**0.10 0.20**
S (Std Dev of Subjects)	**107.4304**

Annotated Output

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

Numeric Results

Power	Average n	k	Total N	Alpha	Beta	Std Dev of Means (Sm)	Standard Deviation (S)	Effect Size
0.90000	2.00	3	6	0.05000	0.10000	287.18	107.43	2.6732
0.80000	2.00	3	6	0.05000	0.20000	244.31	107.43	2.2741
0.90000	3.00	3	9	0.05000	0.10000	168.33	107.43	1.5669
0.80000	3.00	3	9	0.05000	0.20000	145.82	107.43	1.3573
0.90000	5.00	3	15	0.05000	0.10000	112.62	107.43	1.0483
0.80000	5.00	3	15	0.05000	0.20000	98.08	107.43	0.9130
0.90000	8.00	3	24	0.05000	0.10000	83.98	107.43	0.7817
0.80000	8.00	3	24	0.05000	0.20000	73.23	107.43	0.6817
0.90000	10.00	3	30	0.05000	0.10000	73.86	107.43	0.6875
0.80000	**10.00**	**3**	**30**	**0.05000**	**0.20000**	**64.42**	**107.43**	**0.5997**
0.90000	15.00	3	45	0.05000	0.10000	59.07	107.43	0.5499
0.80000	15.00	3	45	0.05000	0.20000	51.54	107.43	0.4797
0.90000	20.00	3	60	0.05000	0.10000	50.67	107.43	0.4716
0.80000	20.00	3	60	0.05000	0.20000	44.21	107.43	0.4115
0.90000	40.00	3	120	0.05000	0.10000	35.34	107.43	0.3289
0.80000	40.00	3	120	0.05000	0.20000	30.83	107.43	0.2870
0.90000	60.00	3	180	0.05000	0.10000	28.73	107.43	0.2674
0.80000	60.00	3	180	0.05000	0.20000	25.07	107.43	0.2333
0.90000	80.00	3	240	0.05000	0.10000	24.82	107.43	0.2311
0.80000	80.00	3	240	0.05000	0.20000	21.66	107.43	0.2016
0.90000	100.00	3	300	0.05000	0.10000	22.18	107.43	0.2064
0.80000	100.00	3	300	0.05000	0.20000	19.35	107.43	0.1801



This plot shows the relationships between power, sample size, and detectable difference. Several conclusions are possible, but the most impressive is the sharp elbow in the curve that occurs near n = 10 when *Sm* is about 64.

How do you interpret an *Sm* of 64? The easiest way is to generate a set of means that have a standard deviation of 64. To do this, press the SD button in the lower right corner of the One Way ANOVA panel to load the Standard Deviation Estimator module. Set *N* = 3, Mean = 0, and Standard Deviation = 64. Press the Two Unique Values button. This results in three means equal to -91, 45, and 45. The differences among these means are the minimum detectable differences that can be detecting with a sample size of 9 when the power is 80%.

Example 6 - Validation using Fleiss

Background

Fleiss (1986) page 374 presents an example of determining a sample size 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.

Setup

This section presents the values of each of the parameters needed to run this example. You can enter these values yourself or load the Example6 template from the Template tab.

Option	Value
Data Tab	
Find	**n (Sample Size)**
k (Number of Groups)	**4**
Hypothesized Means	**9.775 12 12 14.225**
Contrast Coefficients	**None**
n (Sample Size Multiplier)	*Ignored*
Group Sample Size Pattern	**Equal**
Alpha	**0.05**
Beta	**0.20**
S (Std Dev of Subjects)	**3**

Annotated Output

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

Numeric Results

Numeric Results

Power	Average n	k	Total N	Alpha	Beta	Std Dev of Means (Sm)	Standard Deviation (S)	Effect Size
0.80273	11.00	4	44	0.05000	0.19727	1.57	3.00	0.5244

Details when Alpha = 0.05000, Power = 0.80273, SM = 1.57, S = 3.00

Group	Ni	Percent Ni of Total Ni	Mean	Deviation From Mean	Ni Times Deviation
1	11	25.00	9.78	2.23	24.48
2	11	25.00	12.00	0.00	0.00
3	11	25.00	12.00	0.00	0.00
4	11	25.00	14.23	2.23	24.48
ALL	44	100.00	12.00		

PASS also found n = 11. Note that Fleiss used calculations based on a normal approximation, but ***PASS*** uses exact calculations based on the noncentral F distribution.

Example 7 - Validation using Desu

Background

Desu (1990) page 48 presents an example of determining a sample size in an experiment with 3 groups; means of 0, -0.2553, and 0.2553; standard deviation of 1; alpha of 0.05, and beta of 0.10. He finds a sample size of 99 per group.

Setup

This section presents the values of each of the parameters needed to run this example. You can enter these values yourself or load the Example7 template from the Template tab.

Option	Value
Data Tab	
Find	**n (Sample Size)**
k (Number of Groups)	**3**
Hypothesized Means	**0 -0.2553 0.2553**
Contrast Coefficients	**None**
n (Sample Size Multiplier)	*Ignored*
Group Sample Size Pattern	**Equal**
Alpha	**0.05**
Beta	**0.10**
S (Std Dev of Subjects)	**1**

Annotated Output

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

Numeric Results

Numeric Results

Power	Average n	k	Total N	Alpha	Beta	Std Dev of Means (Sm)	Standard Deviation (S)	Effect Size
0.90285	99.00	3	297	0.05000	0.09715	0.21	1.00	0.2085

Details when Alpha = 0.05000, Power = 0.90285, SM = 0.21, S = 1.00

Group	Ni	Percent Ni of Total Ni	Mean	Deviation From Mean	Ni Times Deviation
1	99	33.33	0.00	0.00	0.00
2	99	33.33	-0.26	0.26	25.27
3	99	33.33	0.26	0.26	25.27
ALL	297	100.00	0.00		

PASS also found *n* = 99.

Example 8 - Validation using Kirk

Background

Kirk (1982) pages 140-144 presents an example of determining a sample size in an experiment with 4 groups; means of 2.75, 3.50, 6.25, and 9.0; standard deviation of 1.20995; alpha of 0.05, and beta of 0.05. He finds a sample size of 3 per group.

Setup

This section presents the values of each of the parameters needed to run this example. You can enter these values yourself or load the Example8 template from the Template tab.

Option	Value
Data Tab	
Find	**n (Sample Size)**
k (Number of Groups)	**4**
Hypothesized Means	**2.75 3.5 6.25 9**
Contrast Coefficients	**None**
n (Sample Size Multiplier)	*Ignored*
Group Sample Size Pattern	**Equal**
Alpha	**0.05**
Beta	**0.05**
S (Std Dev of Subjects)	**1.20995**

Annotated Output

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

Numeric Results

Numeric Results

Power	Average n	k	Total N	Alpha	Beta	Std Dev of Means (Sm)	Standard Deviation (S)	Effect Size
0.99767	3.00	4	12	0.05000	0.00233	2.47	1.21	2.0376

PASS also found *n* = 3.

Example 9 - Power of a Planned Comparison

An experiment is being designed to study the response to different doses of a drug. Three groups, receiving a dose of 0, 10, and 20 milligrams of the drug, are anticipated. An *F* test will be used to test the hypothesis that the means exhibit a linear trend across the doses. The significance level is 0.05. Previous studies have shown the within group standard deviation to be 18. Treatment means of 5, 16, and 30 represent clinically important treatment differences. To better understand the relationship between power and sample size, the researcher wants to compute the power for several group sample sizes between 2 and 18. The sample sizes will be equal across all groups.

Setup

Option | **Value**

Data Tab

Find .... **Beta and Power**
k (Number of Groups) **3**
Hypothesized Means **5 16 30**
Contrast Coefficients.......................... **Linear Trend**
n (Sample Size Multiplier) .... **2 to 18 by 2**
Group Sample Size Pattern .... **Equal**
Alpha.. **0.05**
Beta... *Ignored since this is the Find setting*
S (Std Dev of Subjects)..................... **18**

Axes Tab

Vertical Range .... **User (Given Below)**
Minimum **0**
Maximum **1**

Annotated Output

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

Numeric Results Report

Numeric Results

Power	Average n	k	Total N	Alpha	Beta	Std Dev of Means (Sm)	Standard Deviation (S)	Effect Size
0.16781	2.00	3	6	0.05000	0.83219	10.21	18.00	0.5670
0.41889	4.00	3	12	0.05000	0.58111	10.21	18.00	0.5670
0.61410	6.00	3	18	0.05000	0.38590	10.21	18.00	0.5670
0.75458	8.00	3	24	0.05000	0.24542	10.21	18.00	0.5670
0.84932	10.00	3	30	0.05000	0.15068	10.21	18.00	0.5670
0.91013	12.00	3	36	0.05000	0.08987	10.21	18.00	0.5670
0.94768	14.00	3	42	0.05000	0.05232	10.21	18.00	0.5670
0.97017	16.00	3	48	0.05000	0.02983	10.21	18.00	0.5670
0.98329	18.00	3	54	0.05000	0.01671	10.21	18.00	0.5670

Summary Statements

In a one-way ANOVA study, sample sizes of 2, 2, and 2 are obtained from the 3 groups whose means are to be compared using a planned comparison (contrast). The total sample of 6 subjects achieves 17% power to detect a non-zero contrast of the means versus the alternative that the contrast is zero using an F test with a 0.05000 significance level. The value of the contrast of the means is 25.00. The common standard deviation within a group is assumed to be 18.00.

This report shows the numeric results of this power study. Most of the definitions are the same as with the one-way AOV test. Following are the definitions that are different.

Std Dev of Means

When displaying results for planned comparisons, this is not the standard deviation of the hypothesized means. Instead, it is a special function of the coefficients and the hypothesized means given by the equation

$$\sigma_{mc} = \frac{\left|\sum_{i=1}^{k} c_i \mu_i\right|}{\sqrt{N \sum_{i=1}^{k} \frac{c_i^2}{n_i}}}$$

Effect Size

The effect size is the ratio of SM to S. It is an index of relative difference between the means that can be compared from study to study.

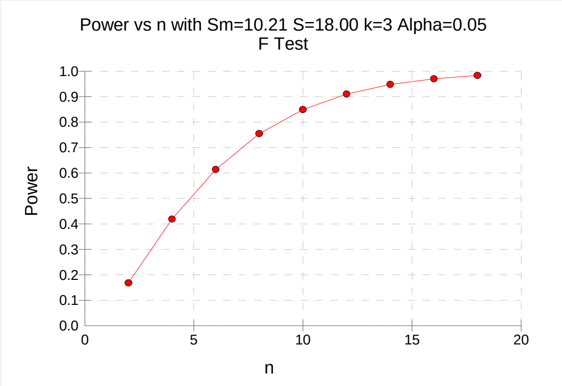
Detail Results Report

Details when Alpha = 0.05000, Power = 0.16781, SM = 10.21, S = 18.00

Group	Ni	Percent Ni of Total N	Mean	Contrast Coefficient	Mean Times Contrast
1	2	33.33	5.00	-1.000	-5.00
2	2	33.33	16.00	0.000	0.00
3	2	33.33	30.00	1.000	30.00
ALL	6	100.00	17.00	0.00	25.00

This report shows the details of each row of the previous report. It is especially useful because it shows the values of the contrast coefficients and the contrast (which is the value in the lower right corner of the table).

Detail Results Report



This plot gives a visual presentation to the results in the Numeric Report. We can quickly see the impact on the power of increasing the sample size.

When you create one of these plots, it is important to use trial and error to find an appropriate range for the horizontal variable so that you have results with both low and high power.

Chapter 555

One-Way ANOVA using Simulation

Introduction

This procedure analyzes the power and significance level of the parametric F-Test and the nonparametric Kruskal-Wallis test which are used to test statistical hypotheses 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.

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 draw the random numbers from this pool. This second method can cut the running time of the simulation by 70%.

As mentioned above, the second method begins by generating a large pool of random numbers from the specified distributions. Each of these pools is evaluated to determine if its mean is within a small relative tolerance (0.0001) of the target mean. If the actual mean is not within the tolerance of the target mean, individual members of the population are replaced with new random numbers if the new random number moves the mean towards its target. Only a few hundred such swaps are required to bring the actual mean to within tolerance of the target mean. This population is then sampled with replacement using the uniform distribution. We have found that this method works well as long as the size of the pool is the maximum of twice the number of simulated samples desired and 10,000.

Test Statistics

Suppose *g* groups each have a normal distribution and means $\mu_1, \mu_2, \cdots, \mu_g$ and common standard deviation σ. Let $n_1, n_2, \cdots, n_g$ denote the number of subjects in each group and let *N* denote the total sample size of all groups. The tests that follow assume that the data are obtained by taking simple random samples from the *g* populations.

F-Test

The formula for the calculation of the F-test is

$$F_{g-1,N-g} = \frac{MSR}{MSE}$$

where

$$MSR = \frac{\sum_{k=1}^{g} n_k \left(\overline{X}_k - \overline{\overline{X}} \right)^2}{g-1}$$

$$MSE = \frac{\sum_{k=1}^{g} \sum_{j=1}^{n_k} \left(X_{kj} - \overline{X}_k \right)^2}{N-g}$$

$$\overline{X}_k = \frac{\sum_{j=1}^{n_k} X_{kj}}{n_k}$$

$$\overline{\overline{X}} = \frac{\sum_{k=1}^{g} n_k \overline{X}_k}{N}$$

$$N = \sum_{k=1}^{g} n_k$$

If the assumptions are met, the distribution of this test statistic follows the *F* distribution with degrees of freedom *g*-1 and *N*-*g*.

Kruskal-Wallis Test

The Kruskal-Wallis test corrected for ties is calculated using the formula

$$W = \frac{H}{T_C}$$

where

$$H = \frac{12}{N(N+1)} \sum_{k=1}^{g} \frac{R_k^2}{n_k} - 3(N+1)$$

$$T_C = 1 - \frac{\sum t(t^2 - 1)}{N(N^2 - 1)}$$

R_k is the sum of the ranks of the k^{th} group, and *t* is the count of a particular tie. The distribution of *W* is approximately Chi-square with *g*-1 degrees of freedom.

Procedure Options

This section describes the options that are unique to this procedure. These are located on the panels associated with the Data, Options, and Reports tabs. To find out more about using the other tabs such as Plot Text, Axes, and Template, turn to the chapter entitled Procedure Templates.

Data Tab

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

Find

This option specifies the parameter to be calculated using the values of the other parameters. Under most conditions, you would select either *Power* or *n*.

Select *Power* when you want to estimate the power of a certain scenario.

Select *n* when you want to determine the sample size needed to achieve a given power and alpha error level. This option is very computationally intensive, so it may take a long time to complete.

Simulations

This option specifies the number of iterations, *M*, used in the simulation. The larger the number of iterations, the longer the running time, and, the more accurate the results.

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

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

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

Alpha

This option specifies one or more values for the probability of a type-I error (alpha). A type-I error occurs when you reject the null hypothesis of equal means when in fact the means are equal.

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

Beta

This option specifies one or more values for the probability of a type-II error (beta). A type-II error occurs when you fail to reject the null hypothesis of equal means when in fact the means are different.

Values must be between zero and one. The values of 0.10 and 0.20 are often used for beta. However, you should pick a value for beta that represents the risk of a type-II error you are willing to take.

Power is defined as one minus beta. Power is equal to the probability of rejecting a false null hypothesis. Hence, specifying the beta error level also specifies the power level. For example, if you specify beta values of 0.05, 0.10, and 0.20, you are specifying the corresponding power values of 0.95, 0.90, and 0.80, respectively.

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 times each of the Group Sample Size Pattern numbers. If the Group Sample Size Pattern numbers are represented by

m1, m2, m3, …, mg and this value is represented by *n*, the group sample sizes

n1, n2, n3, ...,ng 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. This is required since sample sizes must be whole numbers.

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 *N* 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

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 $n1 = n2 = \ldots = ng = n$. That is, all sample sizes are equal to *n*.

Test Statistic

Specify which test is to be simulated. Although the F-test is the most commonly used test, it is based on assumptions that may not be viable in some situations. For your data, you may find that the Kruskal-Wallis test is more accurate (actual alpha = target alpha) and more precise (better power).

Specifying Simulation Distributions

These options specify the distributions to be used in the two simulations, one set per row. 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 k (k = 1 to 9)

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 Data1 tab and the six rows of distributions shown under the Data2 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 k 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. Note that, except for the multinomial, the distributions are parameterized so that the mean is entered first.

Beta=A(M0,A,B,Minimum)
Binomial=B(M0,N)
Cauchy=C(M0,Scale)
Constant=K(Value)
Exponential=E(M0)
F=F(M0,DF1)
Gamma=G(M0,A)
Multinomial=M(P1,P2,…,Pk)
Normal=N(M0,SD)
Poisson=P(M0)
Student's T=T(M0,D)
Tukey's Lambda=L(M0,S,Skewness,Elongation)
Uniform=U(M0,Minimum)
Weibull=W(M0,B)

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

Finding the Value of the Mean of a Specified Distribution

Except for the multinomial distribution, the distributions have been parameterized in terms of their means, since this is the parameter being tested. The mean of a distribution created as a linear combination of other distributions is found by applying the linear combination to the individual means. However, the mean of a distribution created by multiplying or dividing other distributions is not necessarily equal to applying the same function to the individual means. For example, the mean of 4N(4, 5) + 2N(5, 6) is 4*4 + 2*5 = 26, but the mean of 4N(4, 5) * 2N(5, 6) is not exactly 4*4*2*5 = 160 (although it is close).

Group k 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.

Following is a list of the distributions that are available and the syntax used to specify them. Note that, except for the multinomial, the distributions are parameterized so that the mean, *M1*, is entered first.

Beta=A(M1,A,B,Minimum)
Binomial=B(M1,N)
Cauchy=C(M1,Scale)
Constant=K(Value)
Exponential=E(M1)
F=F(M1,DF1)
Gamma=G(M1,A)
Multinomial=M(P1,P2,…,Pk)
Normal=N(M1,SD)
Poisson=P(M1)
Student's T=T(M1,D)
Tukey's Lambda=L(M1,S,Skewness,Elongation)
Uniform=U(M1,Minimum)
Weibull=W(M1,B)

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

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.

Options Tab

The Options tab contains limits on the number of iterations and various options about individual tests.

Maximum Iterations

Specify the maximum number of iterations before the search for the sample size is aborted. When the maximum number of iterations is reached without convergence, the sample size is left blank. We recommend a value of at least 500.

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.

Reports Tab

The Reports tab contains settings about the format of the output.

Show Numeric Reports & Plots

These options let you specify whether you want to generate the standard reports and plots.

Show Inc’s & 95% C.I.

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

Show Comparative Reports & Plots

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

Example1 - Power at Various Sample Sizes

For this first example we repeat Example 1 of the regular One-Way ANOVA procedure. This will allow you to compare the values obtained by simulation with the actual values obtained from the theoretical results.

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

This section presents the values of each of the parameters needed to run this example. You can make these changes directly on your screen or you can load the template entitled Example1 by clicking the Template tab and loading this template.

Option	Value
Data Tab	
Find	**Power**
Simulations	**2000**
Alpha	**0.05**
Beta	*Ignored since this is the Find setting*
n (Multiplier)	**4 8 12**
Group Sample Size Pattern	**Equal**
Test Statistic	**F-Test**
Grps 1	**1**
Group 1 Distribution(s) | H0	**N(M0 S)**
Group 1 Distribution(s) | H1	**N(M1 S)**
Grps 2	**3**
Group 2 Distribution(s) | H0	**N(M0 S)**
Group 2 Distribution(s) | H1	**N(M0 S)**
Grps 3	**0**
M0 (Mean under H0)	**10**
M1 (Mean under H1)	**40**
S	**18**
Option Tab	
Random Number Pool Size	**Automatic**

Annotated Output

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

Numeric Results Report

Numeric Results for Testing the g = 4 Group Means
Test Statistic: F-Test

Power	Average Group Size n	Total Sample Size N	Target Alpha	Actual Alpha	Beta	S.D. of Means Sm|H1	S.D. of Data SD|H1	M0	M1	S
0.536	4.0	16	0.050	0.047	0.465	13.0	18.0	10.0	40.0	18.0
(0.022)	[0.514	0.557]		(0.009)	[0.038	0.056]				
0.896	8.0	32	0.050	0.053	0.105	13.0	18.1	10.0	40.0	18.0
(0.013)	[0.882	0.909]		(0.010)	[0.043	0.063]				
0.982	12.0	48	0.050	0.051	0.019	13.0	18.1	10.0	40.0	18.0
(0.006)	[0.976	0.987]		(0.010)	[0.041	0.061]				

Notes:
Pool Size: 10000. **Simulations**: 2000. **Run Time**: 12.84 seconds.
H0 Distributions: Normal(M0 S); Normal(M0 S); Normal(M0 S); and Normal(M0 S)
H1 Distributions: Normal(M1 S); Normal(M0 S); Normal(M0 S); and Normal(M0 S)

Report Definitions
H0 represents the null hypothesis.
H1 represents the alternative hypothesis.
Power is the probability of rejecting a false null hypothesis. It should be close to one.
'n' is the average of the group sample sizes.
Total N is the total sample size of all groups combined.
Target Alpha is the desired probability of rejecting a true null hypothesis.
Actual Alpha is the alpha achieved by this simulation.
Beta is the probability of accepting a false null hypothesis.
Sm|H1 is the standard deviation of the group means under the alternative hypothesis.
SD|H1 is the within group standard deviation under the alternative hypothesis.
Second Row: (Power Prec.) [95% LCL and UCL Power] (Alpha Prec.) [95% LCL and UCL Alpha]

Summary Statements
A one-way design with 4 groups has sample sizes of 4, 4, 4, and 4. The null hypothesis is that the standard deviation of the group means is 0.1 and the alternative standard deviation of the group means is 13.0. The total sample of 16 subjects achieves a power of 0.536 using the F-Test with a target significance level of 0.050 and an actual significance level of 0.047. The average within group standard deviation assuming the alternative distribution is 18.0. These results are based on 2000 Monte Carlo samples from the null distributions: Normal(M0 S); Normal(M0 S); Normal(M0 S); and Normal(M0 S) and the alternative distributions: Normal(M1 S); Normal(M0 S); Normal(M0 S); and Normal(M0 S). Other parameters used in the simulation were: M0 = 10.0, M1 = 40.0, and S = 18.0.

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

The second row shows two 95% confidence intervals in brackets: the first for the power and the second for the significance level. Half the width of each confidence interval is given in parentheses as a fundamental measure of the accuracy of the simulation. As the number of simulations is increased, the width of the confidence intervals will decrease.

Power

This is the probability of rejecting a false null hypothesis. 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 are provided to help you understand the precision of the estimated power.

Average Group Size n

This is the average of the group sample sizes.

Total Sample Size N

This is the total sample size of the study.

Target Alpha

The target value of alpha: the probability of rejecting a true null hypothesis. This is often called the significance level.

Actual Alpha

This is the value of alpha estimated by the simulation using the H0 distributions. It should be compared with the Target Alpha to determine if the test statistic is accurate in this scenario.

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 are provided to help you understand the precision of the Actual Alpha.

Beta

Beta is the probability of accepting a false null hypothesis. This is the value of beta estimated by the simulation using the H1 distributions.

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. So 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.

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.

Detail Results Report

Details when Target Alpha = 0.050, M0 = 10.0, M1 = 40.0, S = 18.0
Test Statistic: F-Test

Group	Ni	Percent Ni of N	H0 Mean	H1 Mean	H0 S.D.	H1 S.D.	H0 Sm	H1 Sm	Actual Alpha	Power
All	16	100.00	10.0	17.5	18.1	18.0	0.0	13.0	0.051	0.536
1	4	25.00	10.0	40.0	17.7	17.9				
2	4	25.00	10.0	10.0	18.4	18.2				
3	4	25.00	10.0	10.0	18.2	17.8				
4	4	25.00	10.0	10.0	18.0	18.1				

Details when Target Alpha = 0.050, M0 = 10.0, M1 = 40.0, S = 18.0
Test Statistic: F-Test

Group	Ni	Percent Ni of N	H0 Mean	H1 Mean	H0 S.D.	H1 S.D.	H0 Sm	H1 Sm	Actual Alpha	Power
All	32	100.00	10.0	17.5	18.0	18.1	0.0	13.0	0.051	0.896
1	8	25.00	10.0	40.0	17.9	18.1				
2	8	25.00	10.0	10.0	18.1	18.4				
3	8	25.00	10.0	10.0	18.1	18.1				
4	8	25.00	10.0	10.0	18.0	18.0				

Details when Target Alpha = 0.050, M0 = 10.0, M1 = 40.0, S = 18.0
Test Statistic: F-Test

Group	Ni	Percent Ni of N	H0 Mean	H1 Mean	H0 S.D.	H1 S.D.	H0 Sm	H1 Sm	Actual Alpha	Power
All	48	100.00	10.0	17.5	18.1	18.1	0.0	13.0	0.051	0.982
1	12	25.00	10.0	40.0	18.3	18.0				
2	12	25.00	10.0	10.0	18.1	18.4				
3	12	25.00	10.0	10.0	18.2	18.0				
4	12	25.00	10.0	10.0	17.9	17.9				

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

Group

This is the number of the group shown on this line. The first line, labeled *All*, gives the average or the total as appropriate.

Ni

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

Percent Ni of N

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

H0 and H1 Means

These are the means that were used in the simulations for H0 and H1, respectively.

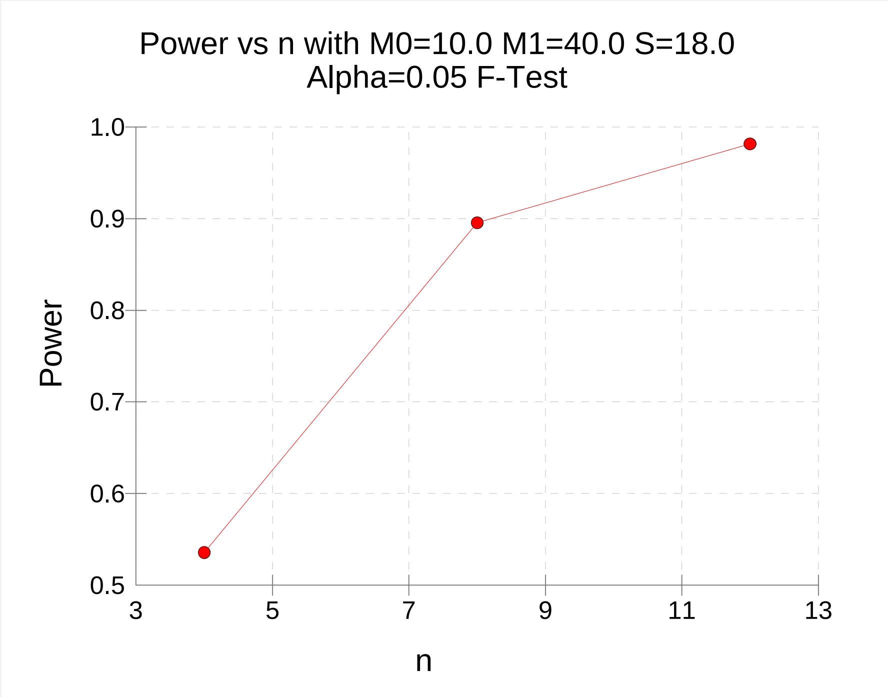
H0 and H1 S.D.'s

These are the standard deviations that were obtained by the simulations for H0 and H1, respectively. Note that they often are not exactly equal to what was specified because of simulation error.

H0 and H1 Sm's

These are the standard deviations of the means that were obtained by the simulations for H0 and H1, respectively. Under H0, the value of Sm should be near zero. It lets you determine if your simulation of H0 was correctly specified.

Plot Section



This plot gives a visual presentation to the results in the Numeric Report. We can quickly see the impact on the power of increasing the sample size.

Example2 - Comparative results

Continuing with Example1, the researchers want to study the characteristics of alternative test statistics. They want to compare the results of the F-test and the Kruskal-Wallis test.

Setup

This section presents the values of each of the parameters needed to run this example. You can enter these values yourself or load the Example2 template from the Template tab.

Option	Value
Data Tab	
Find	**Power**
Simulations	**2000**
Alpha	**0.05**
Beta	*Ignored since this is the Find setting*
n (Multiplier)	**4 8 12**
Group Sample Size Pattern	**Equal**
Test Statistic	**F-Test**
Grps 1	**1**
Group 1 Distribution(s) | H0	**N(M0 S)**
Group 1 Distribution(s) | H1	**N(M1 S)**
Grps 2	**3**
Group 2 Distribution(s) | H0	**N(M0 S)**
Group 2 Distribution(s) | H1	**N(M0 S)**
Grps 3	**0**
M0 (Mean under H0)	**10**
M1 (Mean under H1)	**40**
S	**18**
Reports Tab	
Show Comparative Reports	**Checked**
Show Comparative Plots	**Checked**

Annotated Output

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

Numeric Results

Numeric Results for Testing the g = 4 Group Means
Test Statistic: F-Test

Power	Average Group Size n	Total Sample Size N	Target Alpha	Actual Alpha	Beta	S.D. of Means Sm|H1	S.D. of Data SD|H1	M0	M1	S
0.510	4.0	16	0.050	0.051	0.490	13.0	17.8	10.0	40.0	18.0
0.901	8.0	32	0.050	0.047	0.100	13.0	18.0	10.0	40.0	18.0
0.986	12.0	48	0.050	0.046	0.015	13.0	18.1	10.0	40.0	18.0

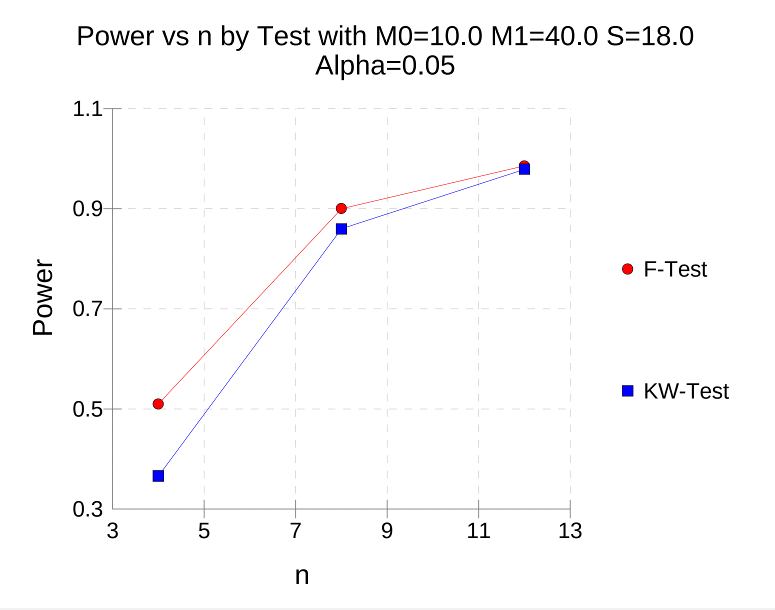
Notes:
Pool Size: 10000. Simulations: 2000. Run Time: 21.02 seconds.
H0 Distributions: Normal(M0 S); Normal(M0 S); Normal(M0 S); and Normal(M0 S)
H1 Distributions: Normal(M1 S); Normal(M0 S); Normal(M0 S); and Normal(M0 S)

Power Comparison for Testing the g = 4 Group Means

Total Sample Size	S.D. of Means Sm|H1	S.D. of Data SD|H1	Target Alpha	F-Test Power	Kruskal Wallis Power	M0	M1	S
16	13.0	17.8	0.050	0.510	0.366	10.0	40.0	18.0
32	13.0	18.0	0.050	0.901	0.860	10.0	40.0	18.0
48	13.0	18.1	0.050	0.986	0.979	10.0	40.0	18.0

Alpha Comparison for Testing the g = 4 Group Means

Total Sample Size	S.D. of Means Sm|H1	S.D. of Data SD|H1	Target Alpha	F-Test Alpha	Kruskal Wallis Alpha	M0	M1	S
16	13.0	16.5	0.050	0.042	0.026	10.0	40.0	18.0
16	13.0	17.8	0.050	0.051	0.041	10.0	40.0	18.0
32	13.0	18.0	0.050	0.047	0.042	10.0	40.0	18.0
48	13.0	18.1	0.050	0.046	0.039	10.0	40.0	18.0



We notice that the power of the F-test is much greater than the Kruskal-Wallis test for $n = 4$. However, when $n = 12$, the powers of the two tests are almost equal. Note that the alpha value of the Kruskal-Wallis test is almost half that of the F-test for $n = 4$. This is probably why the power is also low.

Example3 - Validation using Fleiss

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.

Setup

This section presents the values of each of the parameters needed to run this example. You can enter these values yourself or load the Example3 template from the Template tab.

Option	**Value**
Data Tab	
Find	**Power**
Simulations	**2000**
Alpha	**0.05**
Beta	**0.20**
n (Multiplier)	*Ignored since this is the Find setting*
Group Sample Size Pattern	**Equal**
Test Statistic	**F-Test**
Grps 1	**1**
Group 1 Distribution(s) | H0	**N(M0 S)**
Group 1 Distribution(s) | H1	**N(9.775 S)**
Grps 2	**2**
Group 2 Distribution(s) | H0	**N(M0 S)**
Group 2 Distribution(s) | H1	**N(M0 S)**
Grps 3	**1**
Group 3 Distribution(s) | H0	**N(M0 S)**
Group 3 Distribution(s) | H1	**N(14.225 S)**
M0 (Mean under H0)	**12**
S	**3**

Annotated Output

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

Numeric Results

Numeric Results for Testing the g = 4 Group Means
Test Statistic: F-Test

Power	Average Group Size n	Total Sample Size N	Target Alpha	Actual Alpha	Beta	S.D. of Means Sm|H1	S.D. of Data SD|H1	M0	M1	S
0.822	11.0	44	0.050	0.052	0.179	1.6	3.0	12.0	3.0	
(0.017)	[0.805	0.838]		(0.010)	[0.042	0.061]				

Notes:
Pool Size: 10000. Simulations: 2000. Run Time: 18.88 seconds.
H0 Distributions: Normal(M0 S); Normal(M0 S); Normal(M0 S); and Normal(M0 S)
H1 Distributions: Normal(9.775 S); Normal(M0 S); Normal(M0 S); and Normal(14.225 S)

Details when Target Alpha = 0.050, M0 = 12.0, S = 3.0
Test Statistic: F-Test

Group	Ni	Percent Ni of N	H0 Mean	H1 Mean	H0 S.D.	H1 S.D.	H0 Sm	H1 Sm	Actual Alpha	Power
All	44	100.00	12.0	12.0	3.0	3.0	0.0	1.6	0.052	0.822
1	11	25.00	12.0	9.8	3.0	3.0				
2	11	25.00	12.0	12.0	3.0	3.0				
3	11	25.00	12.0	12.0	3.0	3.0				
4	11	25.00	12.0	14.2	3.0	3.0				

Note that ***PASS*** has also found the group sample size to be 11.

Example4 - 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 investigate the impact of outliers on the power and precision of the F-test and the Kruskal-Wallis test.

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

Setup

This section presents the values of each of the parameters needed to run this example. You can enter these values yourself or load the Example4 template from the Template tab.

Option	Value
Data Tab	
Find	**Power**
Simulations	**2000**
Alpha	**0.05**
Beta	*Ignored since this is the Find setting*
n (Multiplier)	**10 20**
Group Sample Size Pattern	**Equal**
Test Statistic	**F-Test**
Grps 1	**2**
Group 1 Distribution(s) | H0	**N(M0 S)[95];N(M0 A)[5]**
Group 1 Distribution(s) | H1	**N(M0 S)[95];N(M0 A)[5]**
Grps 2	**1**
Group 2 Distribution(s) | H0	**N(M0 S)[95];N(M0 A)[5]**
Group 2 Distribution(s) | H1	**N(M1 S)[95];N(M1 A)[5]**
Grps 3	**0**
M0 (Mean under H0)	**0**
M1 (Mean under H1)	**1**
S	**1**
A	**1 5 10**
Reports Tab	
Show Comparative Reports	**Checked**
Show Comparative Plots	**Checked**

Annotated Output

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

Numeric Results for Testing the g = 3 Group Means
Test Statistic: F-Test

Power	Average Group Size n	Total Sample Size N	Target Alpha	Actual Alpha	Beta	S.D. of Means Sm|H1	S.D. of Data SD|H1	M0	M1	S	A
0.588	10.0	30	0.050	0.054	0.413	0.5	1.0	0.0	1.0	1.0	1.0
0.381	10.0	30	0.050	0.035	0.619	0.5	1.4	0.0	1.0	1.0	5.0
0.321	10.0	30	0.050	0.026	0.680	0.5	2.4	0.0	1.0	1.0	10.0
0.897	20.0	60	0.050	0.049	0.104	0.5	1.0	0.0	1.0	1.0	1.0
0.618	20.0	60	0.050	0.047	0.383	0.5	1.5	0.0	1.0	1.0	5.0
0.408	20.0	60	0.050	0.026	0.592	0.5	2.4	0.0	1.0	1.0	10.0

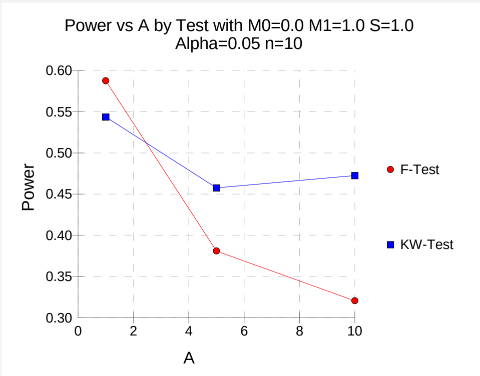
Notes:
Pool Size: 10000. Simulations: 2000. Run Time: 40.30 seconds.
H0 Distributions: Normal(M0 S)[95];Normal(M0 A)[5]; Normal(M0 S)[95];Normal(M0 A)[5]; and Normal(M0 S)[95];Normal(M0 A)[5]
H1 Distributions: Normal(M1 S)[95];Normal(M1 A)[5]; Normal(M0 S)[95];Normal(M0 A)[5]; and Normal(M0 S)[95];Normal(M0 A)[5]

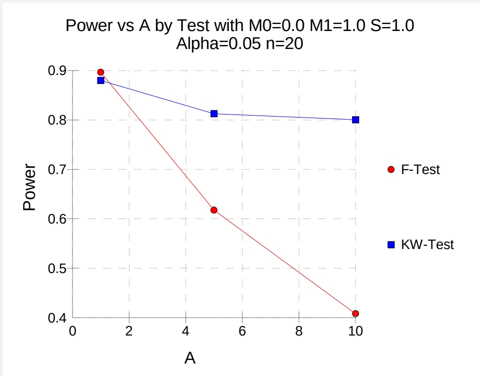
Power Comparison for Testing the g = 3 Group Means

Total Sample Size	S.D. of Means Sm|H1	S.D. of Data SD|H1	Target Alpha	F-Test Power	Kruskal Wallis Power	M0	M1	S	A
30	0.5	1.0	0.050	0.588	0.544	0.0	1.0	1.0	1.0
30	0.5	1.4	0.050	0.381	0.458	0.0	1.0	1.0	5.0
30	0.5	2.4	0.050	0.321	0.473	0.0	1.0	1.0	10.0
60	0.5	1.0	0.050	0.897	0.880	0.0	1.0	1.0	1.0
60	0.5	1.5	0.050	0.618	0.813	0.0	1.0	1.0	5.0
60	0.5	2.4	0.050	0.408	0.801	0.0	1.0	1.0	10.0

Alpha Comparison for Testing the g = 3 Group Means

Total Sample Size	S.D. of Means Sm|H1	S.D. of Data SD|H1	Target Alpha	F-Test Alpha	Kruskal Wallis Alpha	M0	M1	S	A
30	0.5	1.0	0.050	0.054	0.055	0.0	1.0	1.0	1.0
30	0.5	1.4	0.050	0.035	0.041	0.0	1.0	1.0	5.0
30	0.5	2.4	0.050	0.026	0.041	0.0	1.0	1.0	10.0
60	0.5	1.0	0.050	0.049	0.050	0.0	1.0	1.0	1.0
60	0.5	1.5	0.050	0.047	0.054	0.0	1.0	1.0	5.0
60	0.5	2.4	0.050	0.026	0.048	0.0	1.0	1.0	10.0





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

Example5 - 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 precision of the F-test and the Kruskal-Wallis test.

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

Setup

This section presents the values of each of the parameters needed to run this example. You can enter these values yourself or load the Example5 template from the Template tab.

<u>Option</u>	<u>Value</u>
Data Tab	
Find	**Power**
Simulations	**2000**
Alpha	**0.05**
Beta	*Ignored since this is the Find setting*
n (Multiplier)	**10 20**
Group Sample Size Pattern	**Equal**
Test Statistic	**F-Test**
Grps 1	**2**
Group 1 Distribution(s) | H0	**L(M0 S G 0)**
Group 1 Distribution(s) | H1	**L(M0 S G 0)**
Grps 2	**1**
Group 2 Distribution(s) | H0	**L(M0 S G 0)**
Group 2 Distribution(s) | H1	**L(M1 S G 0)**
Grps 3	**0**
M0 (Mean under H0)	**0**
M1 (Mean under H0)	**1**
S	**1**
G	**0 0.5 0.9**
Reports Tab	
Show Comparative Reports	**Checked**
Show Comparative Plots	**Checked**

Annotated Output

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

Numeric Results for Testing the g = 3 Group Means
Test Statistic: F-Test

Power	Average Group Size n	Total Sample Size N	Target Alpha	Actual Alpha	Beta	S.D. of Means Sm|H1	S.D. of Data SD|H1	M0	M1	S	G
0.596	10.0	30	0.050	0.050	0.405	0.5	1.0	0.0	1.0	1.0	0.0
0.615	10.0	30	0.050	0.034	0.386	0.5	1.0	0.0	1.0	1.0	0.5
0.719	10.0	30	0.050	0.036	0.281	0.5	1.0	0.0	1.0	1.0	0.9
0.899	20.0	60	0.050	0.046	0.101	0.5	1.0	0.0	1.0	1.0	0.0
0.902	20.0	60	0.050	0.050	0.098	0.5	1.0	0.0	1.0	1.0	0.5
0.895	20.0	60	0.050	0.034	0.105	0.5	1.0	0.0	1.0	1.0	0.9

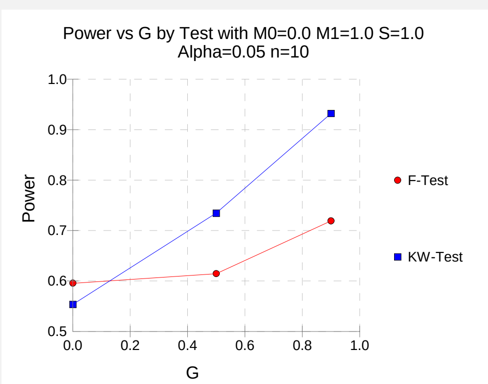
Notes:
Pool Size: 10000. Simulations: 2000. Run Time: 40.30 seconds.
H0 Distributions: Tukey(M0 S G 0); Tukey(M0 S G 0); and Tukey(M0 S G 0)
H1 Distributions: Tukey(M0 S G 0); Tukey(M0 S G 0); and Tukey(M1 S G 0)

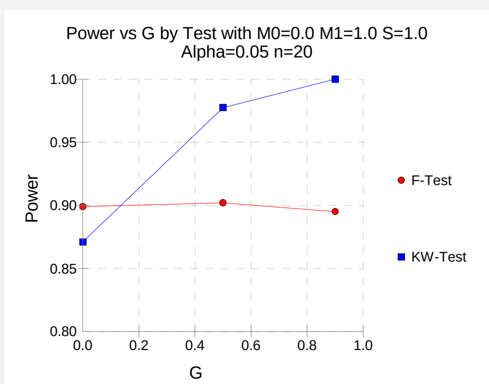
Power Comparison for Testing the g = 3 Group Means

Total Sample Size	S.D. of Means Sm|H1	S.D. of Data SD|H1	Target Alpha	F-Test Power	Kruskal Wallis Power	M0	M1	S	G
30	0.5	1.0	0.050	0.596	0.554	0.0	1.0	1.0	0.0
30	0.5	1.0	0.050	0.615	0.735	0.0	1.0	1.0	0.5
30	0.5	1.0	0.050	0.719	0.932	0.0	1.0	1.0	0.9
60	0.5	1.0	0.050	0.899	0.871	0.0	1.0	1.0	0.0
60	0.5	1.0	0.050	0.902	0.978	0.0	1.0	1.0	0.5
60	0.5	1.0	0.050	0.895	1.000	0.0	1.0	1.0	0.9

Alpha Comparison for Testing the g = 3 Group Means

Total Sample Size	S.D. of Means Sm|H1	S.D. of Data SD|H1	Target Alpha	F-Test Alpha	Kruskal Wallis Alpha	M0	M1	S	G
30	0.5	1.0	0.050	0.050	0.045	0.0	1.0	1.0	0.0
30	0.5	1.0	0.050	0.034	0.039	0.0	1.0	1.0	0.5
30	0.5	1.0	0.050	0.036	0.047	0.0	1.0	1.0	0.9
60	0.5	1.0	0.050	0.046	0.041	0.0	1.0	1.0	0.0
60	0.5	1.0	0.050	0.050	0.046	0.0	1.0	1.0	0.5
60	0.5	1.0	0.050	0.034	0.055	0.0	1.0	1.0	0.9





We note that as the skewness increases, the power of the Kruskal-Wallis test increases substantially as compared to the F-test.

Chapter 560

Fixed Effects Analysis of Variance

Introduction

A common task in research is to compare the average response across levels of one or more factor variables. Examples of factor variables are income level of two regions, nitrogen content of three lakes, or drug dosage. The fixed-effects analysis of variance compares the means of two or more factors. *F* tests are used to determine statistical significance of the factors and their interactions. The tests are nondirectional in that the null hypothesis specifies that all means are equal and the alternative hypothesis simply states that at least one mean is different. This ***PASS*** module performs power analysis and sample size estimation for an analysis of variance design with up to three factors.

In the following example, the responses of a weight loss experiment are arranged in a two-factor, fixed-effect, design. The first factor is diet (D1 and D2) and the second factor is dose level of a dietary drug (low, medium, and high). The twelve individuals available for this study were assigned at random to one of the six treatment groups (cells) so that there were two per group. The response variable was an individual's weight loss after four months.

Table of Individual Weight Losses			
	Dietary Drug Dose Level		
Diet	Low	Medium	High
D1	14, 16	15, 18	23, 28
D2	18, 21	18, 22	38, 39

Important features to note are that each table entry represents a different individual and that the response variable (weight loss) is continuous, while the factors (Diet and Dose) are discrete.

Means can be calculated for each cell of the table. These means are shown in the table below. Note that we have added an additional row and column for the row, column, and overall means. The six means in the interior of this table are called the *cell means*.

Table of Means

	Dietary Drug Dose Level			
Diet	**Low**	**Medium**	**High**	**Total**
D1	15.00	16.50	25.50	19.00
D2	19.50	20.00	38.50	26.00
Total	17.25	18.25	32.00	22.50

The Linear Model

A mathematical model may be formulated that underlies this experimental design. This model expresses each cell mean, μ_{ij}, as the sum of parameters called *effects*. A common linear model for a two-factor experiment is

$$\mu_{ij} = m + a_i + b_j + (ab)_{ij}$$

where i = 1, 2, ..., I and j = 1, 2, ..., J. This model expresses the value of a cell mean as the sum of four components:

m the grand mean.

a_i the effect of the i^{th} level of factor A. Note that $\sum a_i = 0$.

b_j the effect of the j^{th} level of factor B. Note that $\sum b_j = 0$.

ab_{ij} the combined effect of the i^{th} level of factor A and the j^{th} level of factor B. Note that $\sum (ab)_{ij} = 0$.

Another way of stating this model for the two factor case is

Cell Mean = Overall Effect + Row Effect + Column Effect + Interaction Effect.

Since this model is the sum of various constants, it is called a *linear model*.

Calculating the Effects

We will now calculate the effects for our example. We will let Drug Dose correspond to factor A and Diet correspond to factor B.

Step 1 - Remove the grand mean

Remove the grand mean from the table of means by subtracting 22.50 from each entry. The values in the margins are the *effects* of the corresponding factors.

Table of Mean Weight Losses After Subtracting the Grand Mean				
	Dietary Drug Dose Level			
Diet	**Low**	**Medium**	**High**	**Overall**
D1	-7.50	-6.00	3.00	-3.50
D2	-3.00	-2.50	16.00	3.50
Overall	-5.25	-4.25	9.50	22.50

Step 2 - Remove the effects of factor B (Diet)

Subtract the Diet effects (-3.50 and 3.50) from the entries in those rows.

Table of Mean Weight Losses After Subtracting the Diet Effects				
	Dietary Drug Dose Level			
Diet	**Low**	**Medium**	**High**	**Overall**
D1	-4.00	-2.50	6.50	-3.50
D2	-6.50	-6.00	12.50	3.50
Overall	-5.25	-4.25	9.50	22.50

Step 3 - Remove the effects of factor A (Drug Dose)

Subtract the Drug Dose effects (-5.25, -4.25, and 9.50) from the rest of the entries in those columns. This will result in a table of effects.

Table of Effects				
	Dietary Drug Dose Level			
Diet	**Low**	**Medium**	**High**	**Overall**
D1	1.25	1.75	-3.00	-3.50
D2	-1.25	-1.75	3.00	3.50
Overall	-5.25	-4.25	9.50	22.50

We have calculated a table of effects for the two-way linear model. Each cell mean can calculated by summing the appropriate entries from this table.

The estimated linear effects are:

m = 22.50

a1 = -5.25 a2 = -4.25 a3 = 9.50

b1 = -3.50 b2 = 3.50

ab11 = 1.25 ab21 = 1.75 ab31 = -3.00

ab12 = -1.25 ab22 = -1.75 ab32 = 3.00.

The six cell means are calculated from these effects as follows:

15.00 = 22.50 - 5.25 - 3.50 + 1.25

19.50 = 22.50 - 5.25 + 3.50 - 1.25

16.50 = 22.50 - 4.25 - 3.50 + 1.75

20.00 = 22.50 - 4.25 + 3.50 - 1.75

25.50 = 22.50 + 9.50 - 3.50 - 3.00

38.50 = 22.50 + 9.50 + 3.50 + 3.00

Analysis of Variance Hypotheses

The hypotheses that are tested in an analysis of variance table concern the effects, so in order to conduct a power analysis you must have a firm grasp of their meaning. For example, we would usually test the following hypotheses:

1. Are there differences in weight loss among the three drug doses? That is, are the drug dose effects all zero? This hypothesis is tested by the *F* test for factor *A*, which tests whether the standard deviation of the a_i is zero.

2. Is there a difference in weight loss between the two diets? That is, are the diet effects all zero? This hypothesis is tested by the *F* test for factor *B*, which tests whether the standard deviation of the b_j is zero.

3. Are there any diet-dose combinations that exhibit a weight loss that cannot be explained by diet and/or drug dose singly? This hypothesis is tested by the *F* test for the *AB* interaction, which tests whether the standard deviation of the $(ab)_{ij}$ is zero.

Each of these hypotheses can be tested at a different alpha level and different precision. Hence each can have a different power. One of the tasks in planning such an experiment is to determine a sample size that yields necessary power values for each of these hypothesis tests. This is accomplished using this program module.

Definition of Terms

Factorial designs evaluate the effect of two or more categorical variables (called *factors*) on a response variable by testing hypotheses about various averages. These designs are popular because they allow experimentation across a wide variety of conditions and because they evaluate the *interaction* of two or more factors. Interaction is the effect that may be attributed to a combination of two or more factors, but not to one factor singly.

A *factor* is a variable that relates to the response. Either the factor is discrete by nature (as in location or gender) or has been made discrete by collapsing a continuous variable (as in income level or age group). The term *factorial* implies that all possible combinations of the factors being studied are included in the design.

A *fixed* factor is one in which all possible *levels* (categories) are considered. Examples of fixed factors are gender, dose level, and country of origin. They are different from *random* factors which represent a random selection of individuals from the population described by the factor. Examples of random factors are people living within a region, a sample of schools in a state, or a selection of labs. Again, a fixed factor includes the range of interest while a random factor includes only a sample of all possible levels.

A factorial design is analyzed using the analysis of variance. When only fixed factors are used in the design, the analysis is said to be a *fixed-effects analysis of variance*. Other types of designs will be discussed in later chapters.

Suppose a group of individuals have agreed to be in a study involving six treatments. In a *completely randomized factorial design*, each individual is assigned at random to one of the six groups and then the treatments are applied. In some situations, the randomization occurs by randomly selecting individuals from the populations defined by the treatment groups. The designs analyzed by this module are completely randomized factorial designs.

Power Calculations

The calculation of the power of a particular test proceeds as follows

1. Determine the critical value, $F_{df1,df2,\alpha}$ where *df1* is the numerator degrees of freedom, *df2* is the denominator degrees of freedom, and α is the probability of a type-I error (significance level). Note that the *F* test is a two-tailed test as no logical direction is assigned in the alternative hypothesis.

2. Calculate the standard deviation of the hypothesized effects, using the formula:

$$\sigma_m = \sqrt{\frac{\sum_{i=1}^{k}\left(e_i - \bar{e}\right)^2}{k}}$$

   where the e_i are effect values and *k* is the number of effects. Note that the average effect will be zero by construction, so this formula reduces to

$$\sigma_m = \sqrt{\frac{\sum_{i=1}^{k}\left(e_i\right)^2}{k}}$$

3. Compute the noncentrality parameter λ using the relationship:

$$\lambda = N\frac{\sigma_m^2}{\sigma^2}$$

   where *N* is the total number of subjects.

4. Compute the power as the probability of being greater than $F_{df1,df2,\alpha}$ on a noncentral-F distribution with noncentrality parameter λ.

Example

In the example discussed earlier, the standard deviation of the dose effects is

$$\sigma_m(A) = \sqrt{\frac{(-5.25)^2 + (-4.25)^2 + 9.50^2}{3}}$$
$$= 6.729908$$

the standard deviation of the diet effects is

$$\sigma_m(B) = \sqrt{\frac{(-3.5)^2 + 3.5^2}{2}}$$
$$= 3.5$$

and the standard deviation of the interaction effects is

$$\sigma_m(AB) = \sqrt{\frac{1.25^2 + (-1.25)^2 + 1.75^2 + (-1.75)^2 + (-3.00)^2 + 3.00^2}{6}}$$
$$= 2.131119$$

Change in Calculation from PASS 6.0

In ***PASS 6.0***, we used the approach of Cohen (1988) to calculate λ. However, we have found that Cohen's method is less accurate in some situations. Here's why. Cohen produced a set of tables for the one-way AOV which he extended to the two-way and three-way cases by adjusting the per group sample size (his n') so that the denominator degrees of freedom were accurate. However, his adjustment also causes a change in λ which can cause a substantial difference in the calculated power. By using the formula

$$\lambda = N \frac{\sigma_m^2}{\sigma^2}$$

we now calculate the correct power. This is why our calculations differ from that of Cohen (1988) for fixed factorial models.

Standard Deviation of Effects (of Means)

In the two-sample t-test case, the alternative hypothesis was represented as the difference between two group means. Unfortunately, for three or more groups, there is no simple extension of the two group difference. Instead, you must hypothesize a set of effects and calculate the value of σ_m.

Some might wish to specify the alternative hypothesis as the effect size, *f*, which is defined as

$$f = \frac{\sigma_m}{\sigma}$$

where σ is the standard deviation of values within a cell (see Sigma below). If you want to use *f*, set $\sigma = 1$ and then f is always equal to σ_m so that the values you enter for σ_m will be the values of *f*. Cohen (1988) has designated values of *f* less than 0.1 as *small*, values around 0.25 to be *medium*, and values over 0.4 to be *large*. You should come up with your own cutoff values for low, medium, and high.

When you are analyzing the power of an existing analysis of variance table, you can compute the values of σ_m for each term from its mean square or *F* ratio using the following formulas:

$$\sigma_m = \sqrt{\frac{df_{numerator} MS_{numerator}}{N}}$$

or

$$\sigma_m = \sqrt{\frac{df_{numerator}(F)(MSE)}{N}}$$

where *N* is the total number of observations, *MSE* is the mean square error, *df* is the numerator degrees of freedom, *MS* is the mean square of the term, and *F* is the *F* ratio of the term. If you do this, you are setting the sample effects equal to the population effects for the purpose of computing the power.

Procedure Options

This section describes the options that are unique to this procedure. These are located on the panels associated with the Data and Reports tabs. To find out more about using the other tabs, turn to the chapter entitled Procedure Templates.

Data Tab

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

Factors (A, B, C) & Interactions (AB, …, ABC)

These check boxes specify which terms are included in the analysis of variance model. Check a term to signify that it must be included in the analysis.

The three factors are assigned the labels *A*, *B*, and *C*. The interaction between factors *A* and *B* is labeled *AB*. The three-way interaction is labeled *ABC*.

You cannot include an interaction term without including all shorter terms that make up that interaction. For example, if you include the interaction *AC*, you must also include the terms *A* and *C*. Similarly, if you include the term *ABC*, you must also include the terms *A, B, C, AB, AC*, and *BC*.

Categories (A, B, and C)

These options specify the number of categories (levels) contained in each factor. Since the total sample size is equal to the product of the number of levels in each factor and the number of observations per cell (N Per Cell), increasing the number of levels of a factor increases the total sample size of the experiment.

Hypothesized Means (A, B, C)

Enter a set of hypothesized means (or effects), one for each factor level. The standard deviation of these means is used in the power calculations. The standard deviation is calculated using the formula:

$$\sigma_m = \sqrt{\sum_{i=1}^{k} \frac{\left(\mu_i - \bar{\mu}\right)^2}{k}}$$

where k is the number of effects. Note that the standard deviation will be the same whether you enter means or effects since the average of the effects is zero by definition.

Enter a set of means that give the pattern of differences you expect or the pattern that you wish to detect. For example, in a particular study involving a factor with three categories, your research might be meaningful if either of two treatment means is 50% larger than the control mean. If the control mean is 50, then you would enter *50,75,75* as the three means.

It is usually more intuitive to enter a set of mean values. However, it is possible to enter the standard deviation of the means directly by placing an *S* in front of the number.

Entering a list of means

If numbers are entered without a leading *S*, they are assumed to be the hypothesized group means under the alternative hypothesis. Their standard deviation will be calculated and used in the calculations. Blanks or commas may separate the numbers. Note that it is not the values of the means themselves that is important, but only their differences. Thus, the mean values *0,1,2* produce the same results as the values *100,101,102*.

If not enough means are entered to match the number of groups, the last mean is repeated. For example, suppose that four means are needed and you enter *1,2* (only two means). ***PASS*** will treat this as *1,2,2,2*. If too many values are entered, ***PASS*** will truncate the list to the number of means needed.

Examples:

5 20 60

2,5,7

-4,0,6,9

S Option

If an *S* is entered before a number, the number is assumed to be the value of σ_m, the standard deviation of the means.

Examples:

S 4.7

S 5.7

Hypothesized Effects

Specify the standard deviation of the interaction effects using one of the following methods:

1. Enter a set of effects and let the program calculate their standard deviation (see below).
2. Enter the standard deviation directly.
3. Instruct the program to make the standard deviation proportional to one of the main effect terms.

The standard deviation of the effects is calculated using the formula:

$$\sigma_m = \sqrt{\sum_{i=1}^{k} \frac{\left(e_i - \bar{e}\right)^2}{k}}$$

$$= \sqrt{\sum_{i=1}^{k} \frac{e_i^2}{k}}$$

where k is the number of effects and $e_1, e_2, \cdots, e_k$ are the effect values. The value of $\bar{e}$ may be ignored because it is zero by definition.

Syntax

Entering a list of effects

If numbers are entered without a leading letter, they are assumed to be the hypothesized effects under the alternative hypothesis (they are all assumed to be zero under the null hypothesis). Their standard deviation will be calculated and used in the calculations. Blanks or commas may separate the numbers.

If not enough effects are entered to match the number of levels in the term, the last effect is repeated. For example, suppose that four effects are needed and you enter *1,2* (only two effects). ***PASS*** will treat this as *1,2,2,2*. If too many values are entered, ***PASS*** will truncate the list to the number of effects needed.

For interactions, the number of effects is equal to the product of the number of levels of each factor in the interaction. For example, suppose a two-factor design has one factor with three levels and another factor with five levels. The number of effects in the two-factor interaction is (3)(5) = 15.

Examples (note that they sum to zero):

-1 1 -3 3

2 2 0 -1 -1 -2

-4,0,1,3

S Option

If an *S* is followed by a number, the number is assumed to be the value of σ_m, the standard deviation of the effects.

When a set of effects are equal to either *e* or -*e*, the formula for the standard deviation may be simplified as follows:

$$\begin{aligned}\sigma_m &= \sqrt{\sum_{i=1}^{k} \frac{(e_i - 0)^2}{k}} \\ &= \sqrt{\sum_{i=1}^{k} \frac{e^2}{k}} \\ &= e\end{aligned}$$

Hence, another interpretation of σ_m is the absolute value of a set of effects that are equal, except for the sign.

Example:

S 4.7

Enter a Term Followed by a Percentage

You can enter the name of a previous term followed by a percentage. This instructs the program to set this standard deviation to *x*% of the term you specify, where *x* is a positive integer. This allows you to set the magnitude of the interaction standard deviation as a percentage of another term without specifying the interaction in detail.

Note that the term you are taking a percentage of must appear above the term you are specifying. That is, you cannot specify *AB 50* for factor *C* (since only *A* and *B* occur above *C* on the screen).

For example, if the standard deviation of factor *A* is 16, the command

A 75

will set the standard deviation of the current term to (16)(75)/(100) = 12.0.

Other examples of this syntax are:

A 50

B 25

AB 125

AC 150

Discussion

The general formula for the calculation of the standard deviation is

$$\sigma_m = \sqrt{\frac{\sum_{i=1}^{k}(e_i)^2}{k}}$$

where *k* is the number of effects. In the case of a two-way interaction, the standard deviation is calculated using the formula:

$$\sigma_m(AB) = \sqrt{\frac{\sum_{i=1}^{I}\sum_{j=1}^{J}\left(\mu_{ij} - \mu_{i\bullet} - \mu_{\bullet j} + \overline{\mu}\right)^2}{IJ}}$$

where *i* is the factor *A* index (from 1 to *I*), *j* is the factor *B* index (from 1 to *J*), μ_{ij} is the mean in the ij^{th} *cell,* $\mu_{i\bullet}$ is the i^{th} mean of factor *A* across all levels of other factors, $\mu_{\bullet j}$ is the j^{th} mean when factor *B* across all levels of other factors, and $\overline{\mu}$ is the overall mean of the means.

To see how this works, consider the following table of means from an experiment with I = 2 and J = 3:

		i		
		1	2	
	1	2.0	4.0	| 3.0
j	2	4.0	6.0	| 5.0
	3	6.0	11.0	| 8.5
	---	---	---	| ---
	Total	4.0	7.0	| 5.5

Now, if we subtract the factor A means, subtract the factor B means, and add the overall mean, we get the interaction effects:

0.5	-0.5
0.5	-0.5
-1.0	1.0

Next, we sum the squares of these six values:

$$(0.5)^2 + (-0.5)^2 + (0.5)^2 + (-0.5)^2 + (-1.0)^2 + (1.0)^2 = 3$$

Next we divide this value by (2)(3) = 6:

$$3/6 = 0.5$$

Finally, we take the square root of this value:

$$\sqrt{0.5} = 0.7071$$

Hence, for this configuration of means,

$$\sigma_m(AB) = 0.7071\,.$$

Notice that the average of the absolute values of the interaction effects is:

$$[0.5 + 0.5 + 0.5 + 0.5 + 1.0 + 1.0]/6 = 0.6667$$

We see that SD(interaction) is close to the average absolute interaction effect. That is, 0.7071 is close to 0.6667. This will usually be the case. Hence, one way to interpret the interaction standard deviation is as a number a little larger than the average absolute interaction effect.

Alpha

These options specify the significance levels (the probability of a type-I error) of each term. A type-I error occurs when you reject the null hypothesis of that all effects are zero when in fact they are.

Since they are probabilities, alpha values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This value may be interpreted as meaning that about one *F* test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You can select different alpha values for different terms. For example, although you have three factors in an experiment, you might be mainly interested in only one of them. Hence, you could increase the alpha level of the tests from, for example, 0.05 to 0.10 and thereby increase their power. Also, you may want to increase the alpha level of the interaction terms, since these will often have poor power otherwise.

N per Cell

This is the sample size within a cell. Fractional values are allowed. When you have an unequal number of observations per cell, enter the average cell sample size.

If you enter more than one value, a separate analysis will be generated for each value.

S (Std Dev of Subjects)

This option specifies the value of the standard deviation (σ) within a cell (the analysis of variance assumes that σ is constant across all cells). Since they are positive square roots, the numbers must be strictly greater than zero. You can press the SD button to obtain further help on estimating the standard deviation.

This value may be estimated from a previous analysis of variance table by the square root of the mean square error.

If you want to use the effect size, *f*, as the measure of the variability of the effects, you can use 1.0 for σ.

Example 1 - Power after a study

This example will explain how to calculate the power of *F* tests from data that have already been collected and analyzed.

Analyze the power of the experiment that was given at the beginning of this chapter. These data were analyzed using the analysis of variance procedure in ***NCSS*** and the following results were obtained.

Analysis of Variance Table

Source Term	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level	Power (Alpha=0.05)
A (Dose)	2	543.5	271.75	50.95	0.000172*	1.000000
B (Diet)	1	147	147	27.56	0.001920*	0.990499
AB	2	54.5	27.25	5.11	0.050629	0.588884
S	6	32	5.333333			
Total (Adjusted)	11	777				
Total	12					

* Term significant at alpha = 0.05

Means and Effects Section

Term	Count	Mean	Standard Error	Effect
All	12	22.50		22.50
A: Dose				
High	4	32.00	1.154701	9.50
Medium	4	18.25	1.154701	-4.25
Low	4	17.25	1.154701	-5.25
B: Diet				
D1	6	19.00	0.942809	-3.50
D2	6	26.00	0.942809	3.50
AB: Dose,Diet				
High,D1	2	25.50	1.632993	-3.00
High,D2	2	38.50	1.632993	3.00
Low,D1	2	15.00	1.632993	1.25
Low,D2	2	19.50	1.632993	-1.25
Medium,D1	2	16.50	1.632993	1.75
Medium,D2	2	20.00	1.632993	-1.75

Setup

To analyze these data, we can enter the means for factors *A* and *B* as well as the *AB* interaction effects.

Alternatively, we could have calculated the standard deviation of the interaction. This can be done in either of two ways.

Using mean square for *AB* (27.25), the degrees of freedom for *AB* (2), and the total sample size (12), the standard deviation of the *AB*-interaction effects is calculated as follows

$$\sigma_m(AB) = \sqrt{\frac{2(27.25)}{12}} = 2.1311$$

Using the formula based on the effects, the standard deviation of the *AB*-interaction effects is calculated as follows

$$\sigma_m(AB) = \sqrt{\frac{3^2 + 3^2 + 1.25^2 + 1.25^2 + 1.75^2 + 1.75^2}{6}} = 2.1311$$

The value of σ is estimated from the square root of the mean square error:

$$\sigma = \sqrt{5.333333} = 2.3094$$

You can enter these values yourself or load the Example1 template from the Template tab.

Option	Value
Data Tab	
Terms (A, B, AB)	***Checked***
Terms (C, AC, BC, ABC)	***Not checked***
Categories (A)	**3**
Categories (B)	**2**
Hypothesized Means (A)	**17.25 18.25 32**
Hypothesized Means (B)	**19 26**
Hypothesized Effects (AB)	**-3 3 1.25 -1.25 1.75 -1.75**
N Per Cell	**2**
S	**2.3094**
Alpha Level	***All are set to 0.05***
Report Tab	
Report Prob Decimals	**6**
Std Dev Decimals	**4**

Annotated Output

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

Numeric Results

Numeric Results

Term	Power	n	Total N	df1	df2	Std Dev of Means (Sm)	Effect Size	Alpha	Beta
A	1.000000	2.00	12	2	6	6.7299	2.914	0.050000	0.000000
B	0.990499	2.00	12	1	6	3.5000	1.516	0.050000	0.009501
AB	0.588914	2.00	12	2	6	2.1311	0.923	0.050000	0.411086

Standard Deviation Within Subjects = 2.3094

Summary Statements

A factorial design with two factors at 3 and 2 levels has 6.0 cells (treatment combinations). A total of 12.0 subjects are required to provide 2.0 subjects per cell. The within-cell standard deviation is 2.3094. This design achieves 100% power when an F test is used to test factor A at a 5% significance level and the actual standard deviation among the appropriate means is 6.7299 (an effect size of 2.914), achieves 99% power when an F test is used to test factor B at a 5% significance level and the actual standard deviation among the appropriate means is 3.5000 (an effect size of 1.516), and achieves 59% power when an F test is used to test the AB interaction at a 5% significance level and the actual standard deviation among the appropriate means is 2.1312 (an effect size of 0.923).

This report shows the power for each of the three factors. Note that these power values match those given by the ***NCSS*** program in the analysis of variance report.

It is important to emphasize that these power values are for the case when the effects associated with the alternative hypotheses are equal to those given by the data. It will often be informative to calculate the power for other values as well.

Term

This is the term (main effect or interaction) from the analysis of variance model being displayed on this line.

Power

This is the power of the *F* test for this term. Note that since adding and removing terms changes the denominator degrees of freedom (*df2*), the power depends on which other terms are included in the model.

n

This is the sample size per cell (treatment combination). Fractional values indicate an unequal allocation among the cells.

Total N

This is the total sample size for the complete design.

df1

This is the numerator degrees of freedom of the *F* test.

df2

This is the denominator degrees of freedom of the *F* test. This value depends on which terms are included in the AOV model.

Std Dev of Means (Sm)

This is the standard deviation of the means (or effects). It represents the size of the differences among the effects that is to be detected by the analysis. If you have entered hypothesized means, only their standard deviation is displayed here.

Effect Size

This is the standard deviation of the means divided by the standard deviation of subjects. It provides an index of the magnitude of the difference among the means that can be detected by this design.

Alpha

This is the significance level of the *F* test. This is the probability of a type-I error given the null hypothesis of equal means and zero effects.

Beta

This is the probability of the type-II error for this test given the sample size, significance level, and effect size.

Example 2 - Finding the sample size

Background

In this example, we will investigate the impact of increasing the sample size on the power of each of the seven tests in the analysis of variance table of a three factor experiment. The first factor (*A*) has two levels, the second factor (*B*) has three levels, and the third factor (*C*) has four levels. This creates a design with 2 x 3 x 4 = 24 treatment combinations.

All values of σ_m will be set equal to 0.2, σ is set equal to 1.0, and alpha is set to 0.05.

Setup

This section presents the values of each of the parameters needed to run this example. You can enter these values yourself or load the Example2 template from the Template tab.

Option	Value
Data Tab	
Factors and Interactions	***All Checked***
Categories (A)	**2**
Categories (B)	**3**
Categories (C)	**4**
Hypothesized Means (A, B, & C)	**S 0.2**
Hypothesized Effects (AB to ABC)	**A 100** ***(so they will equal that of factor A)***
N Per Cell	**2 8 16 22**
S	**1.0**
Alpha Level	***All are set to 0.05***

Annotated Output

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

Numeric Results

Numeric Results

Term	Power	n	Total N	df1	df2	Std Dev of Means (Sm)	Effect Size	Alpha	Beta
A	0.26502	2.00	48	1	24	0.2	0.200	0.05000	0.73498
B	0.19674	2.00	48	2	24	0.2	0.200	0.05000	0.80326
C	0.16369	2.00	48	3	24	0.2	0.200	0.05000	0.83631
AB	0.19674	2.00	48	2	24	0.2	0.200	0.05000	0.80326
AC	0.16369	2.00	48	3	24	0.2	0.200	0.05000	0.83631
BC	0.11945	2.00	48	6	24	0.2	0.200	0.05000	0.88055
ABC	0.11945	2.00	48	6	24	0.2	0.200	0.05000	0.88055
A	0.78682	8.00	192	1	168	0.2	0.200	0.05000	0.21318
B	0.69038	8.00	192	2	168	0.2	0.200	0.05000	0.30962
C	0.62299	8.00	192	3	168	0.2	0.200	0.05000	0.37701
AB	0.69038	8.00	192	2	168	0.2	0.200	0.05000	0.30962
AC	0.62299	8.00	192	3	168	0.2	0.200	0.05000	0.37701
BC	0.49353	8.00	192	6	168	0.2	0.200	0.05000	0.50647
ABC	0.49353	8.00	192	6	168	0.2	0.200	0.05000	0.50647
A	0.97434	16.00	384	1	360	0.2	0.200	0.05000	0.02566
B	0.94723	16.00	384	2	360	0.2	0.200	0.05000	0.05277
C	0.92061	16.00	384	3	360	0.2	0.200	0.05000	0.07939
AB	0.94723	16.00	384	2	360	0.2	0.200	0.05000	0.05277
AC	0.92061	16.00	384	3	360	0.2	0.200	0.05000	0.07939
BC	0.84559	16.00	384	6	360	0.2	0.200	0.05000	0.15441
ABC	0.84559	16.00	384	6	360	0.2	0.200	0.05000	0.15441
A	0.99569	22.00	528	1	504	0.2	0.200	0.05000	0.00431
B	0.98880	22.00	528	2	504	0.2	0.200	0.05000	0.01120
C	0.98045	22.00	528	3	504	0.2	0.200	0.05000	0.01955
AB	0.98880	22.00	528	2	504	0.2	0.200	0.05000	0.01120
AC	0.98045	22.00	528	3	504	0.2	0.200	0.05000	0.01955
BC	0.95001	22.00	528	6	504	0.2	0.200	0.05000	0.04999
ABC	0.95001	22.00	528	6	504	0.2	0.200	0.05000	0.04999

Standard Deviation of Subjects = 1.0

A few interesting features of this report stand out. First note the range of power values across the range of sample size values tested. Reasonable power is not reached until *n* is 16. Also note that as the number of numerator degrees of freedom (*df1*) increases, the power decreases, other things being equal. We must use this knowledge when planning for appropriate power in tests of important interaction terms.

There are a lot of additional runs that you might try. For example, you might look at the impact of setting the alpha level of interaction terms 0.08. You might look at varying σ_m across the different terms. You might try varying the number of levels of a factor. All of these will impact the power of the *F* tests and will thus be important to consider during the planning stage of an experiment.

Example 3 - Latin Square

Background

This example shows how to study the power of a complicated experimental design like a Latin square. Suppose you want to run a Five-Level Latin square design. Recall that a Five-Level Latin square design consists of three factors each at five levels. One factor is associated with the columns of the square, a second factor is associated with the rows of the square, and a third factor is associated with the letters of the square. In all there are only 5 x 5 = 25 observations used instead of the 5 x 5 x 5 = 125 that would normally be required. The Latin square design has reduced the number of observations by 80%.

The 80% decrease in observations comes at a price—the interaction terms must be ignored. If you can legitimately assume that the interactions are zero, the Latin square (or some other design which reduces the number of observations) is an efficient design to use. We will now show you how to analyze the power of the *F* tests from such a design.

The key is to enter 0.2 (which is 25/125) for *n* and set all the interaction indicators off.

Problem Statement

Since all three factors have five levels, the power of the three *F* tests will be the same if σ_m is the same. Hence, we can try three different sets of hypothesized means. The first set will be five means 0.1 units apart. The second set will be five means 0.5 units apart. The third set will be five means 1.0 unit apart. The standard deviation will be set to 1.0. All alpha levels will be set at 0.05.

The sample size per cell is set at 0.2 and 0.4. This will result in total sample sizes of 25 (one replication) and 50 (two replications).

Setup

This section presents the values of each of the parameters needed to run this example. You can enter these values yourself or load the Example3 template from the Template tab.

Option	Value
Data Tab	
Factors (A, B, C)	***Checked***
Interactions (AB, AC, BC, ABC)	***Not checked***
Categories (A, B, C)	**5**
Hypothesized Means (A)	**1.0 1.1 1.2 1.3 1.4**
Hypothesized Means (B)	**1.0 1.5 2.0 2.5 3.0**
Hypothesized Means (C)	**1.0 2.0 3.0 4.0 5.0**
N Per Cell	**0.2 0.4**
S	**1.0**
Alpha Level	***All are set to 0.05***

Annotated Output

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

Numeric Results

Term	Power	n	Total N	df1	df2	Std Dev of Means (Sm)	Effect Size	Alpha	Beta
A	0.06807	0.20	25	4	12	0.141	0.141	0.05000	0.93193
B	0.63675	0.20	25	4	12	0.707	0.707	0.05000	0.36325
C	0.99867	0.20	25	4	12	1.414	1.414	0.05000	0.00133
A	0.09842	0.40	50	4	37	0.141	0.141	0.05000	0.90158
B	0.97743	0.40	50	4	37	0.707	0.707	0.05000	0.02257
C	1.00000	0.40	50	4	37	1.414	1.414	0.05000	0.00000

Standard Deviation of Subjects = 1.000

In the first design in which N = 25, only the power of the test for C is greater than 0.8. Of course, this power value also depends on the value of the standard deviation of subjects within a cell.

It is interesting to note that doubling the sample size did not double the power!

Example 4 - Validation using Winer

Winer (1991) pages 428-429 presents the power calculations for a two-way design in which factor *A* has two levels and factor *B* has three levels. Winer provides estimates of the sum of squared *A* effects (1.0189), sum of squared *B* effects (5.06), and sum of squared interaction effects (42.11). The mean square error is 8.83 and the per cell sample size is 3. All alpha levels are set to 0.05.

Winer's results are approximate because he has to interpolate in the tables that he is using. He finds the power of the *F* test for factor *A* to be between 0.10 and 0.26. He estimates it as 0.17. The exact power of the *F* test for factor *B* is not given. Instead, the range is found to be between 0.26 and 0.36. The power of the *F* test for the *AB* interaction is "approximately" 0.86.

Setup

This section presents the values of each of the parameters needed to run this example. You can enter these values yourself or load the Example4 template from the Template tab.

<u>**Option**</u>	<u>**Value**</u>
Data Tab	
Factors (A and B)	***Checked***
Factor (C)	***Not checked***
Interaction (AB)	***Checked***
Interactions (AC, BC, ABC)	***Not checked***
Categories (A)	**2**
Categories (B)	**3**
Hypothesized Means (A)	**S 0.714**
Hypothesized Means (B)	**S 1.3**
Hypothesized Effects (AB)	**S 2.65**
N Per Cell	**3**
S	**2.97**
Alpha Level	***All are set to 0.05***

Annotated Output

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

Numeric Results

Term	Power	n	Total N	df1	df2	Std Dev of Means (Sm)	Effect Size	Alpha	Beta
A	0.15576	3.00	18	1	12	0.714	0.240	0.05000	0.84424
B	0.29178	3.00	18	2	12	1.300	0.438	0.05000	0.70822
AB	0.85338	3.00	18	2	12	2.650	0.892	0.05000	0.14662

Standard Deviation of Subjects = 2.970

The power of the test for factor A is 0.16 which is between 0.10 and 0.26. It is close to the interpolated 0.17 that Winer obtained from his tables.

The power of the test for factor B is 0.29 which is between 0.26 and 0.36.

The power of the test for the AB interaction is 0.85 which is close to the interpolated 0.86 that Winer obtained from his tables.

Example 5 - Validation using Prihoda

Prihoda (1983) pages 7-8 presents the power calculations for a two-way design with the following pattern of means:

		Factor B				
		1	2	3	4	All
Factor A	1	41	34	30	27	33
	2	33	24	22	29	27
All		37	29	26	28	30

The means may be manipulated to show the overall mean, the main effects, and the interaction effects:

		Factor B				
		1	2	3	4	All
Factor A	1	1	2	1	-4	3
	2	-1	-2	-1	4	-3
All		7	-1	-4	-2	30

Based on the above effects, Prihoda calculates the power of the interaction test when the sample size per cell is 6, 8, 10, 12, and 14 to be 0.34, 0.45, 0.56, 0.65, and 0.73. The mean square error is 64 and the alpha level is 0.05.

Setup

This section presents the values of each of the parameters needed to run this example. You can enter these values yourself or load the Example5 template from the Template tab.

Option	Value
Data Tab	
Factors (A and B)	***Checked***
Factor (C)	***Not checked***
Interaction (AB)	***Checked***
Interactions (AC, BC, ABC)	***Not checked***
Categories (A)	**2**
Categories (B)	**4**
Hypothesized Means (A)	**33 27**
Hypothesized Means (B)	**37 29 26 28**
Hypothesized Effects (AB)	**1 -2 2 -2 1 -1 -4 4**
N Per Cell	**6 8 10 12 14**
S	**8**
Alpha Level	***All are set to 0.05***

Annotated Output

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

Numeric Results

Term	Power	n	Total N	df1	df2	Std Dev of Means (Sm)	Effect Size	Alpha	Beta
A	0.71746	6.00	48	1	40	3.000	0.375	0.05000	0.28254
B	0.83676	6.00	48	3	40	4.183	0.523	0.05000	0.16324
AB	0.33722	6.00	48	3	40	2.345	0.293	0.05000	0.66278
A	0.83848	8.00	64	1	56	3.000	0.375	0.05000	0.16152
B	0.93871	8.00	64	3	56	4.183	0.523	0.05000	0.06129
AB	0.45099	8.00	64	3	56	2.345	0.293	0.05000	0.54901
A	0.91134	10.00	80	1	72	3.000	0.375	0.05000	0.08866
B	0.97917	10.00	80	3	72	4.183	0.523	0.05000	0.02083
AB	0.55558	10.00	80	3	72	2.345	0.293	0.05000	0.44442
A	0.95292	12.00	96	1	88	3.000	0.375	0.05000	0.04708
B	0.99346	12.00	96	3	88	4.183	0.523	0.05000	0.00654
AB	0.64749	12.00	96	3	88	2.345	0.293	0.05000	0.35251
A	0.97568	14.00	112	1	104	3.000	0.375	0.05000	0.02432
B	0.99807	14.00	112	3	104	4.183	0.523	0.05000	0.00193
AB	0.72541	14.00	112	3	104	2.345	0.293	0.05000	0.27459

Standard Deviation of Subjects = 8.000

Prihoda only presents the power for the interaction test at each sample size. You can check to see that the results match Prihoda's exactly.

Example 6 - Validation using Neter, Kutner, Nachtsheim, and Wasserman

Neter, Kutner, Nachtsheim, and Wasserman (1996) page 1057 presents a power analysis of a two-factor experiment in which factor *A* has three levels and factor *B* has two levels. The significance level is 0.05, the standard deviation is 3.0, and *N* is 2. They calculate a power of about 0.89 for the test of factor *A* when the three means are 50, 55, and 45.

Setup

This section presents the values of each of the parameters needed to run this example. You can enter these values yourself or load the Example5 template from the Template tab.

Option	Value
Data Tab	
Factors (A and B)	***Checked***
Factor (C)	***Not checked***
Interaction (AB)	***Checked***
Interactions (AC, BC, ABC)	***Not checked***
Categories (A)	**3**
Categories (B)	**2**
Hypothesized Means (A)	**50 55 45**
Hypothesized Means (B)	**S 1**
Hypothesized Effects (AB)	**S 1**
N Per Cell	**2**
S	**3.0**
Alpha Level	***All are set to 0.05***

Annotated Output

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

Numeric Results

Term	Power	n	Total N	df1	df2	Std Dev of Means (Sm)	Effect Size	Alpha	Beta
A	0.90162	2.00	12	2	6	4.082	1.361	0.05000	0.09838
B	0.16479	2.00	12	1	6	1.000	0.333	0.05000	0.83521
AB	0.11783	2.00	12	2	6	1.000	0.333	0.05000	0.88217

Standard Deviation of Subjects = 3.000

Note that the power of 0.90 that ***PASS*** has calculated is within rounding of the 0.89 that Neter *et al.* calculated.

Chapter 560

Randomized Block Analysis of Variance

Introduction

This module analyzes a randomized block analysis of variance with up to two treatment factors and their interaction. It provides tables of power values for various configurations of the randomized block design.

The Randomized Block Design

The randomized block design (*RBD*) may be used when a researcher wants to reduce the experimental error among observations of the same treatment by accounting for the differences among blocks. If three treatments are arranged in two blocks, the *RBD* might appear as follows:

Block A	Block B
Treatment 1	Treatment 2
Treatment 3	Treatment 1
Treatment 2	Treatment 3

This diagram shows the main features of a *RBD*:

1. Each block is divided into *k* subblocks, where *k* is the number of treatments.
2. Each block receives all the treatments.
3. The treatments are assigned to the subblocks in random order.
4. There is some reason to believe that the blocks are the same internally, but different from each other.

RBD Reduces Random Error

The random error component of a completely randomized design (such as a one-way or a fixed-effects factorial design) represents the influence of all possible variables in the universe on the response except for the controlled (treatment) variables. This random error component is called the standard deviation or σ (sigma).

As we have discussed, the sample size required to meet alpha and beta error requirements depends directly on the standard deviation. As the standard deviation increases, the sample size increases. Hence, researchers are always looking for ways to reduce the standard deviation. Since the random error component contains the variation due to all possible variables other than treatment variables, one of the most obvious ways to reduce the standard deviation is to remove one or more of these *nuisance* variables from the random error component. One of the simplest ways of doing this is by blocking on them.

For example, an agricultural experiment is often blocked on fields so that differences among fields are explicitly accounted for and removed from the error component. Since these field differences are caused by variations in variables such as soil type, sunlight, temperature, and water, blocking on fields removes the influence of several variables.

Blocks are constructed so that the response is as alike (homogeneous) as possible within a block, but as different as possible between blocks. In many situations, there are obvious natural blocking factors such as schools, seasons, individual farms, families, times of day, etc. In other situations, the blocks may be somewhat artificially constructed.

Once the blocks are defined, they are divided into *k* smaller sections called *subblocks*, where *k* is the number of treatment levels. The *k* treatments are randomly assigned to the subblocks, one block at a time. Hence the order of treatment application will be different from block to block.

Measurement of Random Error

The measurement of the random error component (σ) is based on the assumption that there is no fundamental relationship between the treatment variable and the blocking variable. When this is true, the interaction component between blocks and treatment is zero. If the interaction component is zero, then the amount measured by the interaction is actually random error and can be used as an estimate of σ.

Hence, the randomized block design makes the assumption that there is no interaction between treatments and blocks. The block by treatment mean square is still calculated, but it is used as the estimated standard deviation. This means that the degrees of freedom associated with the block-treatment interaction are the degrees of freedom of the error estimate. If the experimental design has *k* treatments and *b* blocks, the interaction degrees of freedom are equal to $(k-1)(b-1)$. Hence the sample size of this type of experiment is measured in terms of the number of blocks.

Treatment Effects

Either one or two treatment variables may be specified. If two are used, their interaction may also be measured. The null hypothesis in the *F* test states that the effects of the treatment variable are zero. The magnitude of the alternative hypothesis is represented as the size of the standard deviation (σ_m) of these effects. The larger the size of the effects, the larger their standard deviation.

When there are two factors, the block-treatment interaction may be partitioned just as the treatment may be partitioned. For example, if we let *C* and *D* represent two treatments, an analysis of variance will include the terms *C*, *D*, and *CD*. If we represent the blocking factor as *B*, there will be three interactions with blocks: *BC*, *BD*, and *BCD*. Since all three of these terms are assumed to measure the random error, the overall estimate of random error is found by averaging (or *pooling*) these three interactions. The pooling of these interactions increases the power of the experiment by effectively increasing the sample size on which the estimate of σ is based. However, it is based on the assumption that $\sigma = \sigma_{BC} = \sigma_{BD} = \sigma_{BCD}$, which may or may not be true.

An Example

Following is an example of data from a randomized block design. The block factor has four blocks (*B1*, *B2*, *B3*, *B4*) while the treatment factor has three levels (low, medium, and high). The response is shown within the table.

Randomized Block Example			
	Treatments		
Blocks	Low	Medium	High
B1	16	19	20
B2	18	20	21
B3	15	17	22
B4	14	17	19

Analysis of Variance Hypotheses

The *F* test for treatments in a randomized block design tests the hypothesis that the treatment effects are zero. (See the beginning of the Fixed-Effects Analysis of Variance chapter for a discussion of the meaning of effects.)

Single-Factor Repeated Measures Designs

The randomized block design is often confused with a single-factor repeated measures design because the analysis of each is similar. However, the randomization pattern is different. In a randomized block design, the treatments are applied in random order within each block. In a repeated measures design, however, the treatments are usually applied in the same order through time. You should not mix the two. If you are analyzing a repeated measures design, we suggest that you use that module of ***PASS*** to do the sample size and power calculations.

Procedure Options

This section describes the options that are unique to this procedure. These are located on the panels associated with the Data and Reports tabs. To find out more about using the other tabs, turn to the chapter entitled Procedure Templates.

Data Tab

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

Terms (A, B, and AB)

These check boxes specify which terms are included in the analysis of variance model. Check a term to indicate that it is included.

The two factors are assigned the labels *A* and *B*. The interaction between factors *A* and *B* is labeled *AB*. You cannot include the interaction term without including both *A* and *B*.

Categories (A and B)

This option specifies the number of categories (levels) contained in each factor. Since the effective sample size is equal to the product of the number of levels in each factor and the number of blocks, increasing the number of levels of a factor increases the sample size of the experiment.

Hypothesized Means (A and B)

Enter a set of hypothesized means (or effects), one for each factor level. The standard deviation of these means is used in the power calculations. The standard deviation is calculated using the formula:

$$\sigma_m = \sqrt{\sum_{i=1}^{k} \frac{(e_i - \bar{e})^2}{k}}$$

where k is the number of levels. Note that the standard deviation will be the same whether you enter means or effects since the average of the effects is zero by definition.

Enter a set of means that give the pattern of differences you expect or the pattern that you wish to detect. For example, in a particular study involving a factor with three categories, your research might be meaningful if either of two treatment means is 50% larger than the control mean. If the control mean is 50, then you would enter *50,75,75* as the three means.

It is usually more intuitive to enter a set of mean values. However, it is possible to enter the standard deviation of the means directly by placing an *S* in front of the number.

Entering a list of means

If numbers are entered without a leading *S*, they are assumed to be the hypothesized group means under the alternative hypothesis. Their standard deviation will be calculated and used in the calculations. Blanks or commas may separate the numbers. Note that it is not the values of the means themselves that is important, but only their differences. Thus, the mean values *0,1,2* produce the same results as the values *100,101,102*.

If not enough means are entered to match the number of groups, the last mean is repeated. For example, suppose that four means are needed and you enter *1,2* (only two means). ***PASS*** will treat this as *1,2,2,2*. If too many values are entered, ***PASS*** will truncate the list to the number of means needed.

Examples:

5 20 60

2,5,7

-4,0,6,9

S Option

If an *S* is entered before a number, the number is assumed to be the value of σ_m, the standard deviation of the means.

Examples:

S 4.6

S 5.8

Hypothesized Effects

Specify the standard deviation of the interaction effects using one of the following methods:

1. Enter a set of effects and let the program calculate their standard deviation.
2. Enter the standard deviation directly.
3. Instruct the program to make the standard deviation proportional to one of the main effect terms.

The standard deviation of the effects is calculated using the formula:

$$\sigma_m = \sqrt{\sum_{i=1}^{k} \frac{(e_i - \bar{e})^2}{k}}$$

where k is the number of effects and $e_1, e_2, \cdots, e_k$ are the effect values. The value of $\bar{e}$ may be ignored because it is zero by definition.

Syntax

Entering a list of effects

If numbers are entered without a leading letter, they are assumed to be the hypothesized effects under the alternative hypothesis (they are all assumed to be zero under the null hypothesis). Their standard deviation will be calculated and used in the calculations. Blanks or commas may separate the numbers.

If not enough effects are entered to match the number of levels in the term, the last effect is repeated. For example, suppose that four effects are needed and you enter *1,2* (only two effects). ***PASS*** will treat this as *1,2,2,2*. If too many values are entered, ***PASS*** will truncate the list to the number of effects needed.

For interactions, the number of effects is equal to the product of the number of levels of each factor in the interaction. For example, suppose a two-factor design has one factor with three levels and another factor with five levels. The number of effects in the two-factor interaction is (3)(5) = 15.

Examples (note that they sum to zero):

-1 1 -3 3

2 2 0 -1 -1 -2

-4,0,1,3

S Option

If an *S* is followed by a number, the number is assumed to be the value of σ_m, the standard deviation of the effects.

When a set of effects are equal to either *e* or -*e*, the formula for the standard deviation may be simplified as follows:

$$\sigma_m = \sqrt{\sum_{i=1}^{k} \frac{(e_i - \bar{e})^2}{k}}$$

Hence, another interpretation of σ_m is the absolute value of a set of effects that are equal, except for the sign.

Example:

S 4.7

Enter a Term Followed by a Percentage

You can enter the name of a previous term followed by a percentage. This instructs the program to set this standard deviation to *x*% of the term you specify, where *x* is a positive integer. This allows you to set the magnitude of the interaction standard deviation as a percentage of the standard deviation of one of the factors without specifying the interaction in detail.

For example, if the standard deviation of factor *A* is 16, the command

A 75

will set the standard deviation of the current term to (16)(75)/(100) = 12.0.

Other examples of this syntax are:

A 50

B 25

Discussion

The general formula for the calculation of the standard deviation is

$$\sigma_m = \sqrt{\frac{\sum_{i=1}^{k}(e_i)^2}{k}}$$

where *k* is the number of effects. In the case of a two-way interaction, the standard deviation is calculated using the formula:

$$\sigma_m(AB) = \sqrt{\frac{\sum_{i=1}^{I}\sum_{j=1}^{J}\left(\mu_{ij} - \mu_{i\bullet} - \mu_{\bullet j} + \overline{\mu}\right)^2}{IJ}}$$

where *i* is the factor A index (from 1 to *I*), *j* is the factor B index (from 1 to *J*), μ_{ij} is the mean in the ij^{th} *cell,* $\mu_{i\bullet}$ is the i^{th} mean of factor *A* across all levels of other factors, $\mu_{\bullet j}$ is the j^{th} mean when factor *B* across all levels of other factors, and $\overline{\mu}$ is the overall mean of the means.

To see how this works, consider the following table of means from an experiment with *I* = 2 and *J* = 3:

		i		
		1	2	
	1	2.0	4.0	| 3.0
j	2	4.0	6.0	| 5.0
	3	6.0	11.0	| 8.5
	---	---	---	| ---
	Total	4.0	7.0	| 5.5

Now, if we subtract the factor *A* means, subtract the factor *B* means, and add the overall mean, we get the interaction effects:

0.5	-0.5
0.5	-0.5
-1.0	1.0

Next, we sum the squares of these six values:

$$(0.5)^2 + (-0.5)^2 + (0.5)^2 + (-0.5)^2 + (-1.0)^2 + (1.0)^2 = 3$$

Next we divide this value by (2)(3) = 6:

$$3 / 6 = 0.5$$

Finally, we take the square root of this value:

$$\sqrt{0.5} = 0.7071$$

Hence, for this configuration of means,

$$\sigma_m(AB) = 0.7071.$$

Notice that the average of the absolute values of the interaction effects is:

$$[0.5 + 0.5 + 0.5 + 0.5 + 1.0 + 1.0]/6 = 0.6667.$$

We see that SD(interaction) is close to the average absolute interaction effect. That is, 0.7071 is close to 0.6667. This will usually be the case. Hence, one way to interpret the interaction standard deviation is as a number a little larger than the average absolute interaction effect.

Alpha

This option specifies the probability of a type-I error (alpha) for each term. A type-I error occurs when you reject the null hypothesis that the effects are zero when in fact they are.

Since they are probabilities, alpha values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This value may be interpreted as meaning that about one *F* test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You can select different alpha values for different terms. For example, although you have three factors in an experiment, you might be mainly interested in only one of them. Hence, you could increase the alpha level of the tests from, for example, 0.05 to 0.10 and thereby increase their power. Also, you may want to increase the alpha level of the interaction terms, since these will often have poor power otherwise.

Number of Blocks

This specifies one or more values for the number of blocks. If a list of values is entered, a separate calculation will be made for each value.

S (Standard Deviation)

This option specifies the value of the standard deviation. In a randomized block design, this value is estimated by the square root of the mean square error (which may be listed as the mean square of the block-by-treatment interaction). This value will usually have to be determined from a previous study.

Assuming that each block is divided into several subblocks, this is an estimate of the standard deviation that would result when the subblocks within the same block received the same treatment.

If you want to use the effect size, f, as the measure of the variability of the effects, you can use 1.0 for σ.

Estimating the standard deviation is discussed in detail in the Sigma Calculator chapter.

Example1 - Power after a study

This example will explain how to calculate the power of *F* tests from data that have already been collected and analyzed.

We will analyze the power of the experiment that was given at the beginning of this chapter.

These data were analyzed using the analysis of variance procedure in ***NCSS*** and the following results were obtained.

Analysis of Variance Table

Source Term	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level	Power (Alpha=0.05)
A (Blocks)	3	13.66667	4.555555			
B (Treatment)	2	45.16667	22.58333	19.83	0.002269*	0.991442
AB	6	6.833333	1.138889			
S	0	0				
Total (Adjusted)	11	65.66666				
Total	12					

* Term significant at alpha = 0.05

Means and Effects Section

Term	Count	Mean	Standard Error	Effect
B: Treatment				
High	4	20.5	0.5335937	2.333333
Low	4	15.75	0.5335937	-2.416667
Medium	4	18.25	0.5335937	8.333334E-02

We will now calculate the power of the *F* test. Note that factor *B* in this printout becomes factor *A* on the ***PASS*** template.

Setup

To analyze these data, we enter the means for factor *A*. The value of σ is estimated as the square root of the mean square error:

$$\sigma = \sqrt{1.138889} = 1.0672$$

You can enter these values yourself or load the Example1 template from the Template tab.

<u>**Option**</u>	<u>**Value**</u>
Data Tab	
Terms (A)	***Checked***
Terms (B, AB)	***Not checked***
Categories (A)	**3**
Hypothesized Means	**15.75 18.25 20.50**
Number of Blocks	**2 3 4 5**
Sigma	**1.0672**
Alpha Level	***All are set to 0.05***

Annotated Output

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

Numeric Results

Numeric Results

Term	Power	Blocks	Units	df1	df2	Std Dev of Means (Sm)	Effect Size	Alpha	Beta
A	0.42132	2	6	2	2	1.940	1.8179	0.05000	0.57868
A	0.89376	3	9	2	4	1.940	1.8179	0.05000	0.10624
A	0.99144	4	12	2	6	1.940	1.8179	0.05000	0.00856
A	0.99956	5	15	2	8	1.940	1.8179	0.05000	0.00044

Standard Deviation Within Blocks (block-treatment interaction) = 1.067

Summary Statements

A randomized-block design with one treatment factor at 3 levels has 2.0 blocks each with 3.0 treatment combinations. The square root of the block-treatment interaction is 1.067. This design achieves 42% power when an F test is used to test factor A at a 5% significance level and the actual standard deviation among the appropriate means is 1.940 (an effect size of 1.8179).

This report shows the power for each of the five block counts. We see that adequate power of about 0.9 would have been achieved by three blocks.

It is important to emphasize that these power values are for the case when the effects associated with the alternative hypotheses are equal to those given by the data. It will often be informative to calculate the power for other values as well.

Term

This is the term (main effect or interaction) from the analysis of variance model being displayed on this line.

Power

This is the power of the *F* test for this term. Note that since adding and removing terms changes the denominator degrees of freedom (*df2*), the power depends on which other terms are included in the model.

Blocks

This is the number of blocks in the design.

Units

This is the number of subblocks (plots) in the design. It is the product of the number of treatment levels and the number of blocks.

df1

This is the numerator degrees of freedom of the *F* test.

df2

This is the denominator degrees of freedom of the *F* test. This value depends on which terms are included in the AOV model.

Std Dev of Means (Sm)

This is the standard deviation of the means (or effects). It represents the size of the differences among the effects that is to be detected by the analysis. If you have entered hypothesized means, only their standard deviation is displayed here.

Effect Size

This is the standard deviation of the means divided by the standard deviation of subjects. It provides an index of the magnitude of the difference among the means that can be detected by this design.

Alpha

This is the significance level of the *F* test. This is the probability of a type-I error given the null hypothesis of equal means and zero effects.

Beta

This is the probability of the type-II error for this test given the sample size, significance level, and effect size.

Example2 - Validation with Prihoda

Prihoda (1983) presents details of an example that is given in Odeh and Fox (1991). In this example, *Alpha* is 0.025, *Sm* of *A* is 0.577, the number of treatments in factor *A* is 6, the number of treatments in factor *B* is 3, S is 1.0, and the Number of Blocks is 2, 3, 4, 5, 6, 7, and 8. Prihoda gives the power values for the *F* test on factor *A* as 0.477, 0.797, 0.935, 0.982, 0.995, 0.999, and 1.000.

You can enter these values yourself or load the Example2 template from the Template tab.

Option	Value
Data Tab	
Terms (A, B, AB)	***Checked***
Categories (A)	**6**
Categories (B)	**3**
Hypothesized Means (A)	**S 0.577**
Hypothesized Means (B)	**S 1**
Hypothesized Effects (AB)	**S 1**
Number of Blocks	**2 3 4 5 6 7 8**
Sigma	**1.0**
Alpha Level	***All are set to 0.025***

Annotated Output

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

Numeric Results

Numeric Results

Term	Power	Blocks	Units	df1	df2	Std Dev of Means (Sm)	Effect Size	Alpha	Beta
A	**0.47622**	2	36	5	17	0.577	0.5770	0.02500	0.52378
B	0.99697	2	36	2	17	1.000	1.0000	0.02500	0.00303
AB	0.85337	2	36	10	17	1.000	1.0000	0.02500	0.14663
A	**0.79521**	3	54	5	34	0.577	0.5770	0.02500	0.20479
B	0.99999	3	54	2	34	1.000	1.0000	0.02500	0.00001
AB	0.99615	3	54	10	34	1.000	1.0000	0.02500	0.00385
A	**0.93479**	4	72	5	51	0.577	0.5770	0.02500	0.06521
B	1.00000	4	72	2	51	1.000	1.0000	0.02500	0.00000
AB	0.99995	4	72	10	51	1.000	1.0000	0.02500	0.00005
A	**0.98226**	5	90	5	68	0.577	0.5770	0.02500	0.01774
B	1.00000	5	90	2	68	1.000	1.0000	0.02500	0.00000
AB	1.00000	5	90	10	68	1.000	1.0000	0.02500	0.00000
A	**0.99573**	6	108	5	85	0.577	0.5770	0.02500	0.00427
B	1.00000	6	108	2	85	1.000	1.0000	0.02500	0.00000
AB	1.00000	6	108	10	85	1.000	1.0000	0.02500	0.00000
A	**0.99907**	7	126	5	102	0.577	0.5770	0.02500	0.00093
B	1.00000	7	126	2	102	1.000	1.0000	0.02500	0.00000
AB	1.00000	7	126	10	102	1.000	1.0000	0.02500	0.00000
A	**0.99981**	8	144	5	119	0.577	0.5770	0.02500	0.00019
B	1.00000	8	144	2	119	1.000	1.0000	0.02500	0.00000
AB	1.00000	8	144	10	119	1.000	1.0000	0.02500	0.00000

Standard Deviation Within Blocks (block-treatment interaction) = 1.000

We have bolded the power values on this report that should match Prihoda's results. You see that they do match.

Chapter 570

Repeated Measures Analysis of Variance

Introduction

This module calculates the power for *repeated measures* designs having up to three within factors and up to three between factors. It computes power for various test statistics including the *F* test with the Geisser-Greenhouse correction, Wilks' lambda, Pillai-Bartlett trace, and Hotelling-Lawley trace. It can be used to calculate the power of *crossover* designs. .

Repeated measures designs are popular because they allow a subject to serve as their own control. This usually improves the precision of the experiment. However, when the analysis of the data uses the traditional *F* tests, additional assumptions concerning the structure of the error variance must be made. When these assumptions do not hold, the Geisser-Greenhouse correction provides reasonable adjustments so that significance levels are accurate.

An alternative to using the *F* test with repeated measures designs is to use one of the multivariate tests: Wilks' lambda, Pillai-Bartlett trace, or Hotelling-Lawley trace. These alternatives are appealing because they do not make the strict, often unrealistic, assumptions about the structure of the error variance. Unfortunately, they may have less power than the *F* test and they cannot be used in all situations.

An example of a two-factor repeated measures design that can be analyzed by this procedure is shown by the following diagram.

Group 1			Group 2	
Subject 1	**Subject 2**	**Month**	**Subject 3**	**Subject 4**
Treatment L	Treatment L	1	Treatment L	Treatment L
Treatment M	Treatment M	2	Treatment M	Treatment M
Treatment H	Treatment H	3	Treatment H	Treatment H

Groups 1 and 2 form the *between* factor. The within factor has three levels: *L*, *M*, and *H* (low, medium, and high). There are four subjects in this experiment. The three treatments are applied to each subject, one treatment per month. Note that the three treatments are applied to each subject in the same order. Although the order of treatment application should be randomized, it is often the same for all subjects.

This diagram shows the main features of a repeated measures design, which are

1. Each subject receives all treatments.
2. The treatments are applied through time. When the treatments are applied in the same order across all subjects, it is impossible to separate the treatment effects from the sequence effects. Some processes that can cause *sequence effects* are learning, practice, or fatigue—any pattern in the responses across time that occurs without the treatment. If you think the possibility for sequence effects exists, you must make sure that the effects of prior treatments have been washed out before applying the next treatment.
3. Unlike other designs, the repeated measures design has two experimental units: *between* and *within*. In this example, the first (between*)* experimental unit is a subject. Subject-to-subject variability is used to test the between factor (groups). The second (within) experimental unit is the time period. In the above example, the month to month variability within a subject is used to test the treatment. The important point to realize is that the repeated measures design has two error components, the between and the within.

Assumptions

The following assumptions are made when using the *F* test to analyze a factorial experimental design.

1. The response variable is continuous.
2. The residuals follow the normal probability distribution with mean equal to zero and constant variance.
3. The subjects are independent.

Since in a within-subject design responses coming from the same subject are not independent, assumption 3 must be modified for responses within a subject. Independence between subjects is still assumed.

4. The within-subject covariance matrices are equal for all between-subject groups. In this type of experiment, the repeated measurements on a subject may be thought of as a multivariate response vector having a certain covariance structure. This assumption states that these covariance matrices are constant from group to group.
5. When using an *F* test, the within-subject covariance matrices are assumed to be *circular*. One way of defining circularity is that the variances of differences between any two measurements within a subject are constant for all measurements. Since responses that are close together in time often have a higher correlation than those that are far apart, it is common for this assumption to be violated. This assumption is not necessary for the validity of the three multivariate tests: Wilks' lambda, Pillai-Bartlett trace, or Hotelling-Lawley trace.

Advantages of Within-Subjects Designs

Because the response to stimuli usually varies less within an individual than between individuals, the within-subject variability is usually less than (or at most equal to) the between-subject variability. By reducing the underlying variability, the same power can be achieved with a smaller number of subjects.

Disadvantages of Within-Subjects Designs

1. *Practice effect*. In some experiments, subjects systematically improve as they practice the task being studies. In other cases, subjects may systematically get worse as the get fatigued or bored with the experimental task. Note that only the treatment administered first is immune to practice effects. Hence, experimenters should make an effort to balance the number of subjects receiving each treatment first.

2. *Carryover effect*. In many drug studies, it is important to wash out the influence of one drug completely before the next drug is administered. Otherwise, the influence of the first drug carries over into the response to the second drug.
3. *Statistical analysis.* The statistical model is more restrictive than in a regular factorial design since the individual responses must have certain mathematical properties.

Even in the face of all these disadvantages, repeated measures (within-subject) designs are popular in many areas of research. It is important that you recognize these problems going in so you can make sure that the design is appropriate, rather than learning of them later after the research has been conducted.

Technical Details

General Linear Multivariate Model

This section provides the technical details of the repeated measures designs that can be analyzed by ***PASS***. The approximate power calculations outlined in Muller, LaVange, Ramey, and Ramey (1992) are used. Using their notation, for N subjects, the usual general linear multivariate model is

$$\underset{(N\times p)}{Y} = \underset{(N\times q\times p)}{XM} + \underset{(N\times p)}{R}$$

where each row of the residual matrix R is distributed as a multivariate normal

$$row_k\left(R\right) \sim N_p\left(0,\Sigma\right)$$

Note that p is the product of the number of levels of each of the within-subject factors, q is the number of design variables, Y is the matrix of responses, X is the design matrix, M is the matrix of regression parameters (means), and R is the matrix of residuals.

Hypotheses about various sets of regression parameters are tested using

$$H_0\!: \underset{a\times b}{\Theta} = \Theta_0$$

$$\underset{a\times q\times p\times b}{CMD} = \Theta$$

where C and D are orthonormal contrast matrices and Θ_0 is a matrix of hypothesized values, usually zeros. Note that C defines contrasts among the between-subject factor levels and D defines contrast among the within-subject factor levels.

Tests of the various main effects and interactions may be constructed with suitable choices for C and D. These tests are based on

$$\hat{M} = \left(X'X\right)^{-} X'Y$$

$$\hat{\Theta} = C\hat{M}D$$

$$\underset{b\times b}{H} = \left(\hat{\Theta} - \Theta_0\right)' \left[C\left(X'X\right)^{-} C'\right]^{-1} \left(\hat{\Theta} - \Theta_0\right)$$

$$\underset{b\times b}{E} = D'\hat{\Sigma}D \cdot \left(N - r\right)$$

$$\underset{b\times b}{T} = H + E$$

where r is the rank of X.

Geisser-Greenhouse F Test

Upon the assumption that Σ has compound symmetry, a size α test of $H_0 : \Theta = \Theta_0$ is given by the F ratio

$$F = \frac{\mathrm{tr}\left(H\right) / ab}{\mathrm{tr}\left(E\right) / \left[b\left(N - r\right)\right]}$$

with degrees of freedom given by

$$df1 = ab$$

$$df2 = b\left(N - r\right)$$

and noncentrality parameter

$$\lambda = df1\left(F\right)$$

The assumption that Σ has compound symmetry is usually not viable. Box (1954a,b) suggested that adjusting the degrees of freedom of the above F-ratio could compensate for the lack of compound symmetry in Σ. His adjustment has become known as the Geisser-Greenhouse adjustment. Under this adjustment, the modified degrees of freedom and noncentrality parameter are given by

$$df1 = ab\varepsilon$$

$$df2 = b\left(N - r\right)\varepsilon$$

$$\lambda = df1\left(F\right)\varepsilon$$

where

$$\varepsilon = \frac{\mathrm{tr}\left(D'\hat{\Sigma}D\right)^2}{b\ \mathrm{tr}\left(D'\hat{\Sigma}DD'\hat{\Sigma}D\right)}$$

The range of ε is $\frac{1}{b-1}$ to 1. When $\varepsilon = 1$, the matrix is *spherical*. When $\varepsilon = \frac{1}{b-1}$, the matrix differs maximally from sphericity.

Note that the Geisser-Greenhouse adjustment is only needed for testing main effects and interactions involving within-subject factors. Main effects and interactions that involve only between-subject factors need no such adjustment.

The critical value F_{Crit} is computed using the expected value of ε to adjust the degrees of freedom. That is, the degrees of freedom of F_{Crit} are given by

$$df1 = ab\mathrm{E}(\varepsilon)$$

$$df2 = b(N-r)\mathrm{E}(\varepsilon)$$

where

$$E(\hat{\varepsilon}) = \begin{cases} \varepsilon + \dfrac{g_1}{N-r} & \text{if } \varepsilon > \dfrac{g_1}{N-r} \\ \varepsilon / 2 & \text{otherwise} \end{cases}$$

$$g_1 = \sum_{i=1}^{T} f_{ii}\xi_i^2 + \sum\sum_{i \neq j} \frac{f_i \xi_i \xi_j}{\left(\xi_i - \xi_j\right)}$$

$$\begin{aligned} f_i &= \frac{\partial \varepsilon}{\partial \xi_i} \\ &= \frac{2\sum \xi_j}{df_1 \sum \xi_j^2} - \frac{2\lambda_i \left(\sum \xi_j\right)^2}{df_1 \left(\sum \xi_j^2\right)^2} \end{aligned}$$

$$\begin{aligned} f_{ii} &= \frac{\partial^{(2)} \varepsilon}{\partial \xi_i^{(2)}} \\ &= 2h_1 - 8h_2 + 8h_3 - 2h_4 \end{aligned}$$

$$h_1 = \frac{2}{df_1 \sum \xi_j^2}$$

$$h_2 = \frac{\xi_i \left(\sum \xi_j\right)}{df_1 \left(\sum \xi_j^2\right)^2}$$

$$h_3 = \frac{\xi_i^2 \left(\sum \xi_j\right)^2}{df_1 \left(\sum \xi_j^2\right)^3}$$

$$h_4 = \frac{\left(\sum \xi_j\right)^2}{df_1 \left(\sum \xi_j^2\right)^2}$$

where the ξ_j's are the ordered eigenvalues of $D'\Sigma D$.

Wilks' Lambda Approximate F Test

The hypothesis $H_0 : \Theta = \Theta_0$ may be tested using Wilks' likelihood ratio statistic *W*. This statistic is computed using

$$W = \left| E T^{-1} \right|$$

An *F* approximation to the distribution of *W* is given by

$$F_{df_1, df_2} = \frac{\eta / df_1}{(1 - \eta) / df_2}$$

where

$$\lambda = df_1 F_{df_1, df_2}$$

$$\eta = 1 - W^{1/g}$$

$$df1 = ab$$

$$df2 = g\left[(N - r) - (b - a + 1)/2\right] - (ab - 2)/2$$

$$g = \left(\frac{a^2 b^2 - 4}{a^2 + b^2 - 5} \right)^{\frac{1}{2}}$$

Pillai-Bartlett Trace Approximate F Test

The hypothesis $H_0 : \Theta = \Theta_0$ may be tested using the Pillai-Bartlett Trace. This statistic is computed using

$$T_{PB} = tr\left(H T^{-1} \right)$$

A noncentral *F* approximation to the distribution of T_{PB} is given by

$$F_{df_1, df_2} = \frac{\eta / df_1}{(1 - \eta) / df_2}$$

where

$$\lambda = df_1 F_{df_1, df_2}$$

$$\eta = \frac{T_{PB}}{s}$$

$$s = \min(a, b)$$

$$df1 = ab$$

$$df2 = s\left[(N - r) - b + s\right]$$

Hotelling-Lawley Trace Approximate F Test

The hypothesis $H_0: \Theta = \Theta_0$ may be tested using the Hotelling-Lawley Trace. This statistic is computed using

$$T_{HL} = tr\left(HE^{-1}\right)$$

An *F* approximation to the distribution of T_{HL} is given by

$$F_{df_1, df_2} = \frac{\eta / df_1}{\left(1 - \eta\right) / df_2}$$

where

$$\lambda = df_1 F_{df_1, df_2}$$

$$\eta = \frac{\frac{T_{HL}}{s}}{1 + \frac{T_{HL}}{s}}$$

$$s = \min\left(a, b\right)$$

$$df1 = ab$$

$$df2 = s\left[\left(N - r\right) - b + s\right]$$

M (Mean) Matrix

In the general linear multivariate model presented above, *M* represents a matrix of regression coefficients. Since you must provide the elements of *M*, we will discuss its meaning in more detail. Although other structures and interpretations of *M* are possible, in this module we assume that the elements of *M* are the cell means. The rows of *M* represent the between-subject categories and the columns of *M* represent the within-group categories.

The *q* rows of *M* represent the *q* groups into which the subjects can be classified. For example, if a design includes three between-subject factors with 2, 3, and 4 categories, the matrix *M* would have 2 *x* 3 *x* 4 = 24 rows. That is, $q = 24$. Similarly, if a design has three within-subject factors with 3, 3, and 3 categories, the matrix *M* would have 3 *x* 3 *x* 3 = 27 columns. That is, $p = 27$.

Consider now an example in which $q = 3$ and $p = 4$. That is, there are three groups into which subjects can be placed. Each subject is measured four times. The matrix *M* would appear as follows.

$$M = \begin{bmatrix} \mu_{11} & \mu_{12} & \mu_{13} & \mu_{14} \\ \mu_{21} & \mu_{22} & \mu_{23} & \mu_{24} \\ \mu_{31} & \mu_{32} & \mu_{33} & \mu_{34} \end{bmatrix}$$

For example, the element μ_{12} is the mean of the second measurement of subjects in the first group. To calculate the power of this design, you would need to specify appropriate values of all twelve means.

As a second example, consider a design with three between-subject factors and three within-subject factors, all of which have two categories. The *M* matrix for this design would be as follows.

$$M = \left[\begin{array}{ccc|cccccccc}
 & & W1 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 \\
 & & W2 & 1 & 1 & 2 & 2 & 1 & 1 & 2 & 2 \\
 & & W3 & 1 & 2 & 1 & 2 & 1 & 2 & 1 & 2 \\
B1 & B2 & B3 & & & & & & & & \\
\hline
1 & 1 & 1 & \mu_{111111} & \mu_{111112} & \mu_{111121} & \mu_{111122} & \mu_{111211} & \mu_{111212} & \mu_{111221} & \mu_{111222} \\
1 & 1 & 2 & \mu_{112111} & \mu_{112112} & \mu_{112121} & \mu_{112122} & \mu_{112211} & \mu_{112212} & \mu_{112221} & \mu_{112222} \\
1 & 2 & 1 & \mu_{121111} & \mu_{121112} & \mu_{121121} & \mu_{121122} & \mu_{121211} & \mu_{121212} & \mu_{121221} & \mu_{121222} \\
1 & 2 & 2 & \mu_{122111} & \mu_{122112} & \mu_{122121} & \mu_{122122} & \mu_{122211} & \mu_{122212} & \mu_{122221} & \mu_{122222} \\
2 & 1 & 1 & \mu_{211111} & \mu_{211112} & \mu_{211121} & \mu_{211122} & \mu_{211211} & \mu_{211212} & \mu_{211221} & \mu_{211222} \\
2 & 1 & 2 & \mu_{212111} & \mu_{212112} & \mu_{212121} & \mu_{212122} & \mu_{212211} & \mu_{212212} & \mu_{212221} & \mu_{212222} \\
2 & 2 & 1 & \mu_{221111} & \mu_{221112} & \mu_{221121} & \mu_{221122} & \mu_{221211} & \mu_{221212} & \mu_{221221} & \mu_{221222} \\
2 & 2 & 2 & \mu_{222111} & \mu_{222112} & \mu_{222121} & \mu_{222122} & \mu_{222211} & \mu_{222212} & \mu_{222221} & \mu_{222222}
\end{array}\right]$$

The subscripts for each mean follow the pattern $\mu_{B1\,B2\,B3W1W2W3}$. The first three subscripts indicate the between-subject categories and the second three subscripts indicate the within-subject categories. Notice that the first three subscripts are constant in each row and the second three subscripts are constant in each column.

Specifying the M Matrix

When computing the power in a repeated measures analysis of variance, the specification of the *M* matrix is one of your main tasks. The program cannot do this for you. The calculated power is directly related to your choice. So your choice for the elements of *M* must be selected carefully and thoughtfully. When authorization and approval from a government organization is sought, you should be prepared to defend your choice of *M*. In this section, we will explain how you can specify *M*.

Before we begin, it is important that you have in mind exactly what *M* is. *M* is a table of means that represent the size of the differences among the means that you want the study or experiment to detect. That is, *M* gives the means under the alternative hypothesis. Under the null hypothesis, these means are assumed to be equal. Because of the complexity of the repeated measures design, it is often difficult to choose reasonable values, so ***PASS*** will help you. But it is important to remember that you are responsible for these values and that the sample sizes calculated are based on them.

One way to specify the *M* matrix is to do so directly into the spreadsheet. You might do this if you are calculating the ‘retrospective’ power of a study that has already been completed, or if it is simply easier to write the matrix directly. Usually, however, you will specify the *M* matrix in portions.

We will begin our discussion of specifying the *M* matrix with an example. Consider a study of two groups of subjects. Each subject was tested, then a treatment was administered, then the subject was tested again at the ten minute mark, and then tested a third time after sixty minutes. The researchers wanted the sample size to be large enough to detect the following pattern in the means.

Table of Hypothesized Means				
	Time Period			
Group	T0	T10	T60	Average
A	100	130	100	110
B	120	180	120	140
Average	110	155	110	125

To understand how they derived this table, we will perform some basic arithmetic on it.

Step 1 - Remove the overall mean

Subtract 125, the overall mean, from each of the individual means.

Table of Hypothesized Means Adjusted for Overall Mean				
	Time Period			
Group	T0	T10	T60	Average
A	-25	5	-25	-15
B	-5	55	-5	15
Average	-15	30	-15	125

Step 2 - Remove the group effect

Subtract -15 from the first row and 15 from the second row.

Table of Hypothesized Means Adjusted for Group				
	Time Period			
Group	T0	T10	T60	Total
A	-10	20	-10	-15
B	-20	40	-20	15
Total	-15	30	-15	125

Step 3 - Remove the time effect

Subtract -15 from the first column, 30 from the second column, and -15 from the third column.

Table of Hypothesized Means Adjusted for Group and Time					
	Time Period				
Group	**T0**	**T10**	**T60**	**Effect**	**Effect + Overall**
A	5	-10	5	-15	110
B	-5	10	-5	15	140
Effect	-15	30	-15		
Effect + Overall	110	155	110		125

This table, called an effects table, lets us see the individual effect of each component of the model. For example, we can see that the hypothesized pattern across time is that T10 is 45 units higher than either endpoint. Similarly, we note that the hypothesized pattern for the two groups is that Group B is 30 units larger than Group A.

Understanding the interaction is more difficult. One interpretation focuses on T10. We note that in Group A the response for T10 is 10 less than expected while in Group B the response for T10 is 10 more than expected.

Entering this information into *PASS*

Rather than enter the individual values into ***PASS***, you can enter the group, time, and interaction effects directly. For this example, you could enter '110 140' or '-15 15' for the hypothesized means of the between factor and '110 155 110' or '-15 30 -15' for the hypothesized means of the within factor. For the interaction, you would enter the six interaction values '5 -10 5 -5 10 -5'.

Another way to enter the interaction information would be to indicate that the size of the interaction to be detected is about half that of the group factor or about a third of the time factor. For a complete discussion of the interpretation of various interactions, we suggest that you look at Kirk (1982).

C Matrix for Between-Subject Contrasts

The *C* matrix is comprised of contrasts that are applied to the rows of *M*. That is, these are between-group contrasts. You do not have to specify these contrasts. They are generated for you. You should understand that a different *C* matrix is generated for each between-subject term in the model. For example, in the six factor example above, the *C* matrix that will be generated for testing the between-subject factor B1 is

$$C_{B1} = \begin{bmatrix} \frac{-1}{\sqrt{8}} & \frac{-1}{\sqrt{8}} & \frac{-1}{\sqrt{8}} & \frac{-1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} \end{bmatrix}$$

Note that the divisor $\sqrt{8}$ is used so that the total of the squared elements is one. This is required so that the contrast matrix is *orthonormal*.

When creating a test for B1, the matrix *D* is created to average across all within-subject categories.

$$D_{B1} = \begin{bmatrix} \frac{1}{\sqrt{8}} \\ \frac{1}{\sqrt{8}} \\ \frac{1}{\sqrt{8}} \\ \frac{1}{\sqrt{8}} \\ \frac{1}{\sqrt{8}} \\ \frac{1}{\sqrt{8}} \\ \frac{1}{\sqrt{8}} \\ \frac{1}{\sqrt{8}} \end{bmatrix}$$

Generating the C Matrix when there are Multiple Between Factors.

Generating the *C* matrix when there is more than one between factor is more difficult. We like the method of O'Brien and Kaiser (1985) which we briefly summarize here.

Step 1. Write a complete set of contrasts suitable for testing each factor separately. For example, if you have three factors with 2, 3, and 4 categories, you might use

$$\ddot{C}_{B1} = \begin{bmatrix} \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix},\ \ddot{C}_{B2} = \begin{bmatrix} \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ 0 & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix},\ \text{and}\ \ddot{C}_{B3} = \begin{bmatrix} \frac{-3}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} \\ 0 & \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ 0 & 0 & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}.$$

Step 2. Define appropriate J_k matrices corresponding to each factor. These matrices comprised of one row and *k* columns whose equal element is chosen so that the sum of its elements squared is one. In this example, we use

$$J_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix},\; J_3 = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix},\; J_4 = \begin{bmatrix} \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} \end{bmatrix}$$

Step 3. Create the appropriate contrast matrix using a direct (Kronecker) product of either the $\ddot{C}_{Bi}$ matrix if the factor is included in the term or the J_i matrix when the factor is not in the term. Remember that the direct product is formed by multiplying each element of the second matrix by all members of the first matrix. Here is an example

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 & -1 \\ 0 & 2 & 0 \\ -1 & 0 & 3 \end{bmatrix}$$
$$= \begin{bmatrix} 1 & 2 & 0 & 0 & -1 & -2 \\ 3 & 4 & 0 & 0 & -3 & -4 \\ 0 & 0 & 2 & 4 & 0 & 0 \\ 0 & 0 & 6 & 8 & 0 & 0 \\ -1 & -2 & 0 & 0 & 3 & 6 \\ -3 & -4 & 0 & 0 & 9 & 12 \end{bmatrix}$$

As an example, we will compute the C matrix suitable for testing factor $B2$

$$C_{B2} = J_2 \otimes \ddot{C}_{B2} \otimes J_4$$

Expanding the direct product results in

$$C_{B2} = J_2 \otimes \ddot{C}_{B2} \otimes J_4$$
$$= \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \otimes \begin{bmatrix} \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ 0 & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \otimes \begin{bmatrix} \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} \end{bmatrix}$$
$$= \begin{bmatrix} \frac{-2}{\sqrt{12}} & \frac{-2}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} \\ 0 & 0 & \frac{-1}{\sqrt{4}} & \frac{-1}{\sqrt{4}} & \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} \end{bmatrix} \otimes \begin{bmatrix} \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} \end{bmatrix}$$
$$= \begin{bmatrix} \frac{-2}{\sqrt{48}} & \frac{-2}{\sqrt{48}} & \frac{1}{\sqrt{48}} & \frac{1}{\sqrt{48}} & \frac{1}{\sqrt{48}} & \frac{1}{\sqrt{48}} & \frac{-2}{\sqrt{48}} & \frac{-2}{\sqrt{48}} & \frac{1}{\sqrt{48}} & \frac{1}{\sqrt{48}} & \frac{1}{\sqrt{48}} & \frac{1}{\sqrt{48}} & \frac{-2}{\sqrt{48}} & \frac{-2}{\sqrt{48}} & \frac{1}{\sqrt{48}} & \frac{1}{\sqrt{48}} & \frac{1}{\sqrt{48}} & \frac{1}{\sqrt{48}} & \frac{-2}{\sqrt{48}} & \frac{-2}{\sqrt{48}} & \frac{1}{\sqrt{48}} & \frac{1}{\sqrt{48}} & \frac{1}{\sqrt{48}} & \frac{1}{\sqrt{48}} \\ 0 & 0 & \frac{-1}{\sqrt{16}} & \frac{-1}{\sqrt{16}} & \frac{1}{\sqrt{16}} & \frac{1}{\sqrt{16}} & 0 & 0 & \frac{-1}{\sqrt{16}} & \frac{-1}{\sqrt{16}} & \frac{1}{\sqrt{16}} & \frac{1}{\sqrt{16}} & 0 & 0 & \frac{-1}{\sqrt{16}} & \frac{-1}{\sqrt{16}} & \frac{1}{\sqrt{16}} & \frac{1}{\sqrt{16}} & 0 & 0 & \frac{-1}{\sqrt{16}} & \frac{-1}{\sqrt{16}} & \frac{1}{\sqrt{16}} & \frac{1}{\sqrt{16}} \end{bmatrix}$$

Similarly, the C matrix suitable for testing interaction $B2B3$ is

$$C_{B2B3} = J_2 \otimes \ddot{C}_{B2} \otimes \ddot{C}_{B3}$$

We leave the expansion of this matrix *PASS*, but we think you have the idea.

D Matrix for Within-Subject Contrasts

The *D* matrix is comprised of contrasts that are applied to the columns of *M*. That is, these are within-group contrasts. You do not have to specify these contrasts either. They will be generated for you. Specification of the *D* matrix is similar to the specification of the *C* matrix, except that now the matrices are all transposed.

Interactions of Between-Subject and Within-Subject Factors

Interactions that include both between-subject factors and within-subject factors require that between-subject portion be specified by the *C* matrix and the within-subject portion be specified with the *D* matrix.

Power Calculations

To calculate statistical power, we must determine distribution of the test statistic under the alternative hypothesis which specifies a different value for the regression parameter matrix *B*. The distribution theory in this case has not been worked out, so approximations must be used. We use the approximations given by Mueller and Barton (1989) and Muller, LaVange, Ramey, and Ramey (1992). These approximations state that under the alternative hypothesis, F_U is distributed as a noncentral *F* random variable with degrees of freedom and noncentrality shown above. The calculation of the power of a particular test may be summarized as follows

1. Specify values of X, M, Σ, C, D, and Θ_0.
2. Determine the critical value using $F_{crit} = FINV(1-\alpha, df1, df2)$, where *FINV*( ) is the inverse of the central *F* distribution and α is the significance level.
3. Compute the noncentrality parameter λ.
4. Compute the power as

$$Power = 1 - NCFPROB\left(F_{crit}, df1, df2, \lambda\right)$$

where *NCFPROB*( ) is the noncentral *F* distribution.

Covariance Matrix Assumptions

The following assumptions are made when using the *F* test. These assumptions are not needed when using one of the three multivariate tests.

In order to use the *F* ratio to test hypotheses, certain assumptions are made about the distribution of the residuals e_{ijk}. Specifically, it is assumed that the residuals for each subject, $e_{ij1}, e_{ij2}, \cdots, e_{ijT}$, are distributed as a multivariate normal with means equal to zero and covariance matrix Σ_{ij}. Two additional assumptions are made about these covariance matrices. First, they are assumed to be equal for all subjects. That is, it is assumed that $\Sigma_{11} = \Sigma_{12} = \cdots = \Sigma_{Gn} = \Sigma$. Second, the covariance matrix is assumed to have a particular form called *circularity*. A covariance matrix is *circular* if there exists a matrix *A* such that

$$\Sigma = A + A' + \lambda I_T$$

where I_T is the identity matrix of order *T* and λ is a constant.

This property may also be defined as

$$\sigma_{ii} + \sigma_{jj} - 2\sigma_{ij} = 2\lambda$$

One type of matrix that is circular is one that has *compound symmetry*. A matrix with this property has all elements on the main diagonal equal and all elements off the main diagonal equal. An example of a covariance matrix with compound symmetry is

$$\Sigma = \begin{bmatrix} \sigma^2 & \rho\sigma^2 & \rho\sigma^2 & \cdots & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \rho\sigma^2 & \cdots & \rho\sigma^2 \\ \rho\sigma^2 & \rho\sigma^2 & \sigma^2 & \cdots & \rho\sigma^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho\sigma^2 & \rho\sigma^2 & \rho\sigma^2 & \cdots & \sigma^2 \end{bmatrix}$$

or, with actual numbers,

$$\begin{bmatrix} 9 & 2 & 2 & 2 \\ 2 & 9 & 2 & 2 \\ 2 & 2 & 9 & 2 \\ 2 & 2 & 2 & 9 \end{bmatrix}$$

An example of a matrix which does not have compound symmetry but is still circular is

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 2 & 2 & 2 \\ 3 & 3 & 3 & 3 \\ 4 & 4 & 4 & 4 \end{bmatrix} + \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{bmatrix} + \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 4 & 3 & 4 & 5 \\ 3 & 6 & 5 & 6 \\ 4 & 5 & 8 & 7 \\ 5 & 6 & 7 & 10 \end{bmatrix}$$

Needless to say, the need to have the covariance matrix circular is a very restrictive assumption.

Between Standard Deviation

The subject-to-subject variability is represented by $\sigma^2_{Between}$. In the repeated measures AOV table, this quantity is estimated by the between subjects mean square (*MSB*). This quantity is calculated from Σ using the formula

$$\sigma^2_{Between} = \frac{\sum_{i=1}^{T}\sum_{j=1}^{T}\sigma_{ij}}{T} = \frac{\sum_{i=1}^{T}\sum_{j=1}^{T}\sigma_{ii}\sigma_{jj}\rho_{ij}}{T}$$

When Σ has compound symmetry, which requires all $\sigma_{ii} = \sigma$ and all $\rho_{ij} = \rho$, the above formula reduces to

$$\sigma^2_{Between} = \sigma^2\left(1+\left(T-1\right)\rho\right)$$

Note that *F* tests of between factors and their interactions do not require the circularity assumption so the Geisser-Greenhouse correction is not applied to these tests.

Within Standard Deviation

The within-subject variability is represented by σ^2_{Within} . In the repeated measures AOV table, this quantity is estimated by the within-subjects mean square (*MSW*). This quantity is calculated from Σ using the formula

$$\sigma^2_{Within} = \frac{\sum_{i=1}^{T}\sigma_{ii}}{T} - \frac{2\sum_{i=1}^{T}\sum_{j=i+1}^{T}\sigma_{ij}}{T\left(T-1\right)} = \frac{\sum_{i=1}^{T}\sigma_{ii}}{T} - \frac{2\sum_{i=1}^{T}\sum_{j=i+1}^{T}\rho_{ij}\sqrt{\sigma_{ii}\sigma_{jj}}}{T\left(T-1\right)}$$

When Σ has compound symmetry, which requires all $\sigma_{ii} = \sigma$ and all $\rho_{ij} = \rho$, the above formula reduces to

$$\sigma^2_{Within} = \sigma^2\left(1-\rho\right)$$

Estimating Sigma and Rho from Existing Data

Using the above results for existing data, approximate values for σ and ρ may be estimated from a previous analysis of variance table that provides estimates of MSB and MSW. Solving the above equations for σ and ρ yields

$$\rho = \frac{\sigma_{Between}^2 - \sigma_{Within}^2}{\sigma_{Between}^2 + (T-1)\sigma_{Within}^2}$$

$$\sigma^2 = \frac{\sigma_{Within}^2}{1-\rho}$$

Substituting MSB for $\sigma_{Between}^2$ and MSW for σ_{Within}^2 yields the estimates

$$\hat{\rho} = \frac{MSB - MSW}{MSB + (T-1)MSW}$$

$$\hat{\sigma}^2 = \frac{MSW}{1-\hat{\rho}}$$

Note that these estimators assume that the design meets the circularity assumption, which is usually not the case. However, they provide crude estimates that can be used in planning.

Procedure Options

This section describes the options that are unique to this procedure. To find out more about using the other tabs, turn to the chapter entitled Procedure Templates.

Data 1 & 2 Tabs

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

Find (Solve For)

This option specifies the parameter to be solved for. If you choose to solve for *n* (sample size), you must also specify which test statistic you want to use.

When you choose to solve for *n*, the program searches for the lowest sample size that meets the alpha and beta criterion you have specified for each of the terms. If you do not want a term to be used in the search, set its alpha and beta values to 0.99.

Also, when the '= n's' box is not checked, the search is made using unequal group sample sizes. The relative proportion of the sample in each group is set by the values of *n* given in the Subjects Per Group box. For example, if your design has three groups and you entered '1 1 2' in the Subjects Per Group box, the search will only consider designs in which the size of the last group is twice the rest. That is, it will consider '2 2 4', '3 3 6', '4 4 8', etc.

Note: no plots are generated when you solve for *n*.

n (Subjects Per Group)

Specify one or more values for the number of subjects per group. The total sample size is the sum of the individual group sizes across all groups.

You can specify a list like '2 4 6'. The items in the list may be separated with commas or blanks. The interpretation of the list depends on the =n's check box. When the =n's box is checked, a separate analysis is calculated for each value of *n*. When the =n's box is not checked, ***PASS*** uses the *n's* as the actual group sizes. In this case, the number of items entered must match the number of groups in the design.

When you choose to solve for *n* and the '= n's' box is not checked, the search is made using unequal group sample sizes. The relative proportion of the sample in each group is set by the values of *n* given in this box. For example, if your design has three groups and you enter '1 1 2' here, the search will only consider designs in which the size of the last group is twice the rest. That is, it will consider '2 2 4', '3 3 6', '4 4 8', etc.

= n's

This option controls whether the number of subjects per group is to be equal for all groups or not. When checked, the number of subjects per group is equal for all groups. A list of values such as '5 10 15' represents three designs: one with five per group, one with ten per group, and one with fifteen per group.

When this option is not checked, the n's are assumed to be unequal. A list of values represents the size of the individual groups. For example, '5 10 15' represents a single, three-group design with five in the first group, ten in the second group, and fifteen in the third group.

Means Matrix

Use this option to the specify spreadsheet columns containing a hypothesized means matrix that will be used to compute the Sm values. All individual Sm values are ignored. You can obtain the spreadsheet by selecting 'Window', then 'Data', from the menus.

The between factors are represented across the columns of the spreadsheet and the within factors are represented down the rows. The number of columns specified must equal the number of groups. The number of rows with data in these columns must equal the number of times. For example, suppose you are designing an experiment that is to have two between factors (A & B) and two within factors (D & E). Suppose each of the four factors has two levels. The columns of the spreadsheet would be

A1B1 A1B2 A2B1 A2B2.

The rows of the spreadsheet would represent

D1E1

D1E2

D2E1

D2E2

Example

To see how this option works, consider the following table of hypothesized means for an experiment with one between factor (A) having two groups and one within factor (B) having three time periods. The values in columns C1 and C2 of the spreadsheet are

C1	C2
2.0	4.0
4.0	6.0
6.0	11.0

By subtracting the appropriate means, the following table of effects results

	C1	C2	Means	Effects
Row1	0.5	-0.5	3.0	-2.5
Row2	0.5	-0.5	5.0	-0.5
Row3	-1.0	1.0	8.5	3.0
	---	---	---	
Means	4.0	7.0	5.5	
Effects	-1.5	1.5		

The standard deviation of the A effects is calculated as

$$\begin{aligned}\sigma_A &= \sqrt{\frac{(-1.5)^2 + (1.5)^2}{2}} \\ &= \sqrt{2.25} \\ &= 1.5\end{aligned}$$

The standard deviation of the B effects is calculated as

$$\begin{aligned}\sigma_B &= \sqrt{\frac{(-2.5)^2+(-0.5)^2+(3.0)^2}{3}}\\ &= \sqrt{\frac{15.5}{3}}\\ &= 2.27\end{aligned}$$

The standard deviation of the interaction effects is found to be

$$\begin{aligned}\sigma_{AB} &= \sqrt{\frac{(0.5)^2+(0.5)^2+(-1.0)^2+(-0.5)^2+(-0.5)^2+(1.0)^2}{6}}\\ &= \sqrt{\frac{3.0}{6}}\\ &= 0.71\end{aligned}$$

These three standard deviations are used to represent the effect sizes of the corresponding terms.

Discussion

When using this option, it is less confusing to concentrate on a single term at a time. For example, consider a 2-by-4 design in which your primary interest is in testing the AB interaction. Instead of trying to determine a means matrix the will represent factor A, factor B, and the AB interaction, ignore factor A and factor B and just consider the interaction. You might want to consider the following pattern

C1	C2
0.0	0.0
0.0	1.0
0.0	2.0
0.0	3.0

That is, the first group remains constant while the second group increases for 0.0 to 3.0.

By specifying various values for K (the means multiplier), you can study to impact of increasing the values. For example, when K is set to 2, the above means matrix becomes

C1	C2
0.0	0.0
0.0	2.0
0.0	4.0
0.0	6.0

Thus, by simply changing K, several scenarios may be studied. (We wish to thank Keith Muller for suggesting this method of specifying the Sm values.)

K (Means Multipliers)

These values are multiplied times the means matrix to give you various effect sizes. A separate power calculation is generated for each value of K. These values become the horizontal axis in the second power chart. For example, if an Sm value is 80, setting this option to ‘50 100 150’ would result in three Sm values: 40, 80, and 120. If you want to ignore this setting, enter ‘1’.

Note that when you enter Sm values directly, ***PASS*** generates an appropriate means matrix and then multiplies this matrix by each of these K values.

Labels

Specify a label for this factor. Although we suggest that only a single letter be used, the label can consist of several letters. When several letters are used, the labels for the interactions may be extra long and confusing. Of course, you must be careful not to use the same label for two factors.

One of the easiest sets of labels is to use A, B, and C for the between factors and D, E, and F for the within factors. A useful alternative is to use B1, B2, and B3 for the between factors and W1, W2, and W3 for the within factors.

Levels

Specify the number of levels (categories) in this factor. Typical values are from 2 to 8. Set this to a blank (or 0) to ignore the factor in the design.

Alpha

These options specify the probability of a type-I error (alpha) for each factor and interaction. A type-I error occurs when you reject the null hypothesis of zero effects when in fact they are zero. Since they are probabilities, alpha values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This value may be interpreted as meaning that about one F-test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You can specify different alpha values for different terms. For example, although you have three terms in an experiment, you might be mainly interested in only one of them. Hence, you could increase the alpha level of the tests of the other terms and thereby increase their power. Also, you may want to increase the alpha level of the interaction terms, as these will often have poor power otherwise.

Beta

These options specify the probability of a type-II error (beta) for each factor and interaction. A type-II error occurs when you fail to reject the null hypothesis of equal effects when in fact they are different.

Values must be between zero and one. Historically, the value of 0.20 was often used for beta. Now, 0.10 is becoming more common. You should pick a value for beta that represents the risk of a type-II error you are willing to take.

Power is defined as one minus beta. Power is equal to the probability of rejecting a false null hypothesis. Hence, specifying the beta error level also specifies the power level. For example, if you specify a beta value of 0.10, you are specifying the corresponding power value of 0.90.

Sm (SD of Effects)

Enter the standard deviation of the effects ($\sigma_{effects}$) for this factor or interaction. This value represents the magnitude of the differences among the means (effects) that is to be detected.

The value of Sm may be entered in several ways: directly, as a list of numbers, or as a percentage of another term.

Directly

You can enter the value of Sm directly by specifying a single number. If only a single number is entered, it becomes the value of Sm.

You can use the Standard Deviation Estimator window to calculate the value of Sm for various sets of means. This window is obtained by selecting PASS, then Other, and then Standard Deviation Estimator from the menus.

List of Numbers

When a list of numbers is entered, the standard deviation of those numbers is computed and used as the value of Sm. The numbers in the list may represent means or effects. The list may be a simple list, the STEP command, or the RANGE command.

Simple List

A Simple List is a set of numbers separated by blanks or commas. Examples a simple list are

5 20 60

2,5,7

-4,0,6,9

STEP Command

The syntax of the STEP command is *STEP Start Inc*. The list begins at *Start* and increases by *Inc*. The number of values generated is determined by the number of levels in the term. Examples of the STEP command for a term with four levels are

‘STEP 0 2’ results in ‘0 2 4 6’.

‘STEP 1 -1’ results in ‘1 0 -1 -2’.

‘STEP 1 0.5’ results in ‘1 1.5 2 2.5’.

RANGE Command

The syntax of the RANGE command is *RANGE Minimum Maximum*. The list of numbers generated increase steadily from *Minimum* to *Maximum*. The RANGE command is handy when you want to vary the number of levels while keeping the values in a known range. Examples of the RANGE command for a term with four levels are

‘RANGE 10 70’ results in ‘10 30 50 70’.

‘RANGE 0 1’ results in ‘0 0.33 0.67 1’.

‘RANGE 1 4’ results in ‘1 2 3 4’.

Percentage of Another Term

You can specify the value as a percentage of another term. The syntax of this command is *TERM PCT* where *TERM* is any other main effect or interaction in the model and *PCT* is a percentage. This method is often used to specify interaction Sm's. Examples of the PERCENTAGE command are

'A 100' results in an Sm equal to the Sm of factor A.

'B 50' results in an Sm equal to one-half of the Sm of factor B.

Interactions

The values of Alpha, Beta, and Sm are entered for various groups of interactions.

Covariance Tab

This tab specifies the covariance matrix.

Specify Which Covariance Matrix Input Method to Use

This option specifies which method will be used to define the covariance matrix.

Standard Deviations and Correlations

This option generates a covariance matrix based on the settings for the standard deviations (SD's) and the pattern of autocorrelations as specified in the options on this screen down to and including 'R2'. More about this option is given below.

Covariance Matrix Variables

When this option is selected, the covariance matrix is read in from the columns of the spreadsheet. This is the most flexible method, but specifying a covariance matrix is tedious. You will usually only use this method when a specific covariance is given to you. More about this option is given below.

Note that the spreadsheet is shown by selecting the menus: 'Window' and then 'Data'.

1) Specify Covariance Matrix Using SD's and Autocorrelations

The parameters in this section provide a flexible way to specify Σ , the covariance matrix. Because the covariance matrix is symmetric, it can be represented as

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1T} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1T} & \sigma_{2T} & \cdots & \sigma_{TT} \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_{12} & \cdots & \sigma_1\sigma_T\rho_{1T} \\ \sigma_1\sigma_2\rho_{12} & \sigma_2^2 & \cdots & \sigma_2\sigma_T\rho_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_1\sigma_T\rho_{1T} & \sigma_2\sigma_T\rho_{2T} & \cdots & \sigma_T^2 \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_T \end{bmatrix} \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1T} \\ \rho_{12} & 1 & \cdots & \rho_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1T} & \rho_{2T} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_T \end{bmatrix}$$

where T is the product of the number of levels of all of the within factors.

Thus, the covariance matrix can be represented with complete generality by specifying the standard deviations $\sigma_1, \sigma_2, \cdots, \sigma_T$ and the correlation matrix

$$R = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1T} \\ \rho_{12} & 1 & \cdots & \rho_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1T} & \rho_{2T} & \cdots & 1 \end{bmatrix}.$$

Specify How the SD's Change Across Time

This option specifies the method used to specify the standard deviations $\sigma_1, \sigma_2, \cdots, \sigma_T$. Based on the method selected, the actual values are specified using SD1 and, in some cases, SD2.

Each σ is an estimate of the standard deviation that occurs when the same individual is measured at the same point in time under identical treatment conditions. It is a measure of the within-subject variability.

The available options are

Constant (Use SD1. Ignore SD2)

When this option is selected, the standard deviations are assumed to be equal. That is, it is assumed that $\sigma_1 = \cdots = \sigma_T$. The value of σ_i is specified in the SD1 field. The value in the SD2 field is ignored.

List of SD's (Use list in SD1. Ignore SD2.)

When this option is selected, a list of standard deviations can be entered in SD1. The items in the list can be separated by commas or blanks. The first value in the list becomes σ_1, the second value becomes σ_2, and so on. If the number of values in the list is less than the number of standard deviations required, the last value in the list is repeated. Note that all standard deviations in the list must be positive numbers.

Range from SD1 to SD2 using the Time Metric

When this option is selected, the standard deviations are spread between σ_1 and σ_T according to the spread in the Time Metric. The value in SD1 becomes σ_1 and the value in SD2 becomes σ_T.

For example, suppose SD1 = 100, SD2 = 200, and the Time Metric values are 0, 2, 4, 10. The standard deviations would be

$$\sigma_1 = 100.0$$
$$\sigma_2 = 120.0$$
$$\sigma_3 = 140.0$$
$$\sigma_4 = 200.0$$

Time Metric

This option is used when the 'How the SD's Change Across Time' option is set to 'Range from SD1 to SD2 using the Time Metric' to help define the covariance matrix. It specifies a sequential list of time points at which measurements of the subjects are made. Often, measurements are made at equally-spaced points through time. This is not always the case. It is important to define a time metric that corresponds to the study. For example, measurements might be planned at the beginning, after one day, after one week, and after one month.

The number of time points is the product of the number of levels of all within factors.

The time metric influences the values of the SD's as well as the correlations between two measurements on the same individual.

Entering a list of times

A list of times can be entered in which the time values are separated by blanks or commas. The time metric can be defined in any time scale desired. For example, you could enter 0, 0.143, 1, 2, 3 if times were 0 weeks, 1/7 week (day 1), 1 week, 2 weeks, 3 weeks. The same values in days would be 0, 1, 7, 14, 21.

Using the RANGE command

The RANGE command can be used to specify a list of times. The syntax of the RANGE command is

RANGE Minimum Maximum

A set of equal-spaced time points is generated between Minimum and Maximum. The number of time points depends on the number of within-factor levels.

This setting is very useful when you want to study the impact of increasing/decreasing the number of measurements per subject during the same period of time. That is, if the study will last five weeks, will the power of the statistical tests increase if you take ten measurements rather than five?

For example, suppose there are six times. Entering

RANGE 0 10

will generate the time metric: 0, 2, 4, 6, 8, 10. If the number of times is changed to eleven, the time metric will become: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.

Using the STEP command

The STEP command can be used to specify a list of times. The syntax of the STEP command is

STEP Start, Inc

This command generates time points beginning at *Start* and incrementing by *Inc*. For example,

STEP 0 2

would generate the values 0, 2, 4, 6, 8, …

SD1 (Standard Deviation 1)

This option is used to generate the covariance matrix. Its interpretation depends on the ‘How the SD’s Change Across Time’ option’s setting. Fundamentally, this is the standard deviation that occurs when the same individual is measured at the same point in time under identical treatment conditions. It is a measure of the within-subject variability.

You may want to use the special window that has been prepared to estimate SD1 from the mean square between (MSB) and the mean square within (MSW) of an existing table. To display this special window, from the menus select ‘PASS’, then ‘Other’, and then ‘Standard Deviation Estimator’. Click on the ‘Covariance Matrix’ tab. Enter the values from the ANOVA table. The resulting value of ‘Sigma’ should be placed here.

How the SD's Change Across Time = Constant

The value entered here is used as the standard deviation for all time points.

How the SD's Change Across Time = List of SD’s

The values in the list entered here become the values of the standard deviations. If the number in the list is less than the number required, the last value in the list is repeated.

How the SD's Change Across Time = Range

The value entered here is used as a beginning standard deviation, the value in SD2 is used as an ending standard deviation, and the intermediate standard deviations are spaced between SD1 and SD2 proportional to the values of the Time Metric.

For example, suppose SD1 = 100, SD2 = 200, and the Time Metric values are 0,2,4,10. The standard deviation values would be:

S(1)=100

S(2)=120

S(3)=140

S(4)=200.

SD2 (Standard Deviation 2)

This parameter is used when 'How the SD's Change Across Time' option is set to 'Range…'. In that case, this option specifies the value of σ_T .

Specify How the Autocorr's Change Across Time

This option specifies the pattern of the autocorrelations in the variance-covariance matrix. Three options are possible:

Constant

The value of R1 is used as the constant autocorrelation until the maximum time difference is reached, then the value of R2 is used. For example, if the maximum time difference is 3, R1 = 0.6, R2 = 0.1, and T = 6, the correlation matrix would appear as

$$R = \begin{bmatrix} 1 & 0.600 & 0.600 & 0.600 & 0.100 & 0.100 \\ 0.600 & 1 & 0.600 & 0.600 & 0.600 & 0.100 \\ 0.600 & 0.600 & 1 & 0.600 & 0.600 & 0.600 \\ 0.600 & 0.600 & 0.600 & 1 & 0.600 & 0.600 \\ 0.100 & 0.600 & 0.600 & 0.600 & 1 & 0.600 \\ 0.100 & 0.100 & 0.600 & 0.600 & 0.600 & 1 \end{bmatrix}$$

Note that when all correlations are equal, this is the correlation pattern that is assumed by the repeated measure ANOVA F-test. It may be a good first approximation, but many researchers believe the next option (first-order autocorrelation) is more realistic.

1st Order Autocorrelation

The value of R1 is used as the base autocorrelation in a first-order, serial correlation pattern. For example, if the maximum time difference is 3, R1 = 0.6, R2 = 0.1, and T = 6, the correlation matrix would appear as

$$R = \begin{bmatrix} 1 & 0.600 & 0.360 & 0.216 & 0.100 & 0.100 \\ 0.600 & 1 & 0.600 & 0.360 & 0.216 & 0.100 \\ 0.360 & 0.600 & 1 & 0.600 & 0.360 & 0.216 \\ 0.216 & 0.360 & 0.600 & 1 & 0.600 & 0.360 \\ 0.100 & 0.216 & 0.360 & 0.600 & 1 & 0.600 \\ 0.100 & 0.100 & 0.216 & 0.360 & 0.600 & 1 \end{bmatrix}$$

This pattern is often chosen as the most realistic when little is known about the correlation pattern.

Custom

The values of R1, R2, A, V, and Max Time Diff are used to generate a custom autocorrelation pattern. This relationship is modeled using the equation

$$\begin{aligned} Corr\left(Y_{si}, Y_{sj}\right) &= \rho_{ij} \\ &= d\left(R1^{1-A+A\left|t_i - t_j\right|^V}\right) + (1-d)R2 \end{aligned}$$

where R1 is the base correlation, t_i and t_j are two time points, and A and V are specified constants. The variable d is one if $\left|t_i - t_j\right|$ is less than Max Time Diff and zero otherwise.

Machin, Campbell, Fayers, and Pinol (1997) state that values of R1 between 0.60 and 0.75 are common.

We will present some examples to show you how this formula may be interpreted. For the moment, assume that the time metric is four, equally space time points of 1, 2, 3, and 4. Also, assume that Max Time Diff is set to 20.

Example 1

Let A = 0, V = 1, and $\rho_1 = \rho$. The correlation matrix becomes

$$R = \begin{bmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{bmatrix}$$

Example 2

Let A = 1, V = 1, and $\rho_1 = \rho$. The correlation matrix becomes

$$R = \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

which is the first-order autoregression model, a popular model in time series analysis.

Example 3

Let A = 1, V = 2, and $\rho_1 = \rho$. The correlation matrix becomes

$$R = \begin{bmatrix} 1 & \rho^2 & \rho^4 & \rho^6 \\ \rho^2 & 1 & \rho^2 & \rho^2 \\ \rho^4 & \rho^2 & 1 & \rho^2 \\ \rho^6 & \rho^4 & \rho^2 & 1 \end{bmatrix}$$

which is similar to Example 2 except that the correlations die out much more quickly.

Example 4

Let A = 0.5, V = 1, and $\rho_1 = \rho$. The correlation matrix becomes

$$R = \begin{bmatrix} 1 & \rho^{0.5} & \rho^{1.0} & \rho^{1.5} \\ \rho^{0.5} & 1 & \rho^{0.5} & \rho^{1.0} \\ \rho^{1.0} & \rho^{0.5} & 1 & \rho^{0.5} \\ \rho^{1.5} & \rho^{1.0} & \rho^{0.5} & 1 \end{bmatrix}$$

which is similar to Example 2 except that the correlations die out much more slowly.

Example 5

Let A = 1 and V = 1. For this example, set the Max Time Diff option to 2. The correlation matrix becomes

$$R = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \rho_2 \\ \rho_1 & 1 & \rho_1 & \rho_2 \\ \rho_2 & \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_2 & \rho_1 & 1 \end{bmatrix}$$

Notice that this scenario lets you create a banded correlation matrix with two unique correlations.

Example 6

This example shows how this formula works when the Max Time Diff is set to 7 and the time metric is 1, 2, 7, 15. Let A = 1 and V = 1. The correlation matrix becomes

$$R = \begin{bmatrix} 1 & \rho_1 & \rho_1^6 & \rho_2 \\ \rho_1 & 1 & \rho_1^5 & \rho_2 \\ \rho_1^6 & \rho_1^5 & 1 & \rho_2 \\ \rho_2 & \rho_2 & \rho_2 & 1 \end{bmatrix}$$

R1 (Autocorrelation)

This is the autocorrelation, r1, between two measurements made on a subject at two time points that differ by one time unit. This value is combined with the other parameters in this section to form the covariance matrix.

Since this is a type of correlation, possible values range from -1 to 1. However, in this situation, a positive value is usually assumed, so the range is 0 to 1. A value near 0 indicates low autocorrelation. A value near 1 indicates high autocorrelation.

The value of this parameter depends on the Time Metric that is defined. Normally, you would expect a larger autocorrelation if the time metric units were in hours rather than days. In their book on sample size, Machin and Campbell comment the values between 0.60 and 0.75 are typical.

It is reasonable to assume that there is a correlation between two measurements made on the same subject at two points in time. It is often reasonable to assume that the size of this correlation diminishes as the two time points are further and further apart. That is, you would expect a much larger autocorrelation between two measurements taken one minute apart than between two measurements taken one week apart.

You may want to use the special window that has been prepared to estimate R1 from the mean square between (MSB) and the mean square within (MSW) of an existing table. To display this special window, from the menus select ‘PASS’, then ‘Other’, and then ‘Standard Deviation Estimator’. Click on the ‘Covariance Matrix’ tab. Enter the values from the ANOVA table. The resulting value of ‘Rho should be placed here.

R2 (Second AC)

This is the value of the secondary autocorrelation, R2. This value is used when the difference between two time points (see Time Metric) is greater than the value of Max Time Diff. Hence, if you set Max Time Diff to zero, this value will be used to calculate all correlations in the covariance matrix. When used, think of R2 as the correlation between measurements made on the same subject, regardless Of how far apart in time they are. Since we are assuming a positive autocorrelation, this value ranges between 0 and 1.

2) Specify Covariance Matrix using Spreadsheet Columns

This option instructs the program to read the covariance matrix from the spreadsheet.

Spreadsheet Columns Containing the Covariance Matrix

This option designates the columns on the current spreadsheet holding the covariance matrix. It is used when the ‘Specify Which Covariance Matrix Input Method to Use’ option is set to *Covariance Matrix Variables*.

The number of columns and number of rows must match the number of time periods at which the subjects are measured.

Reports Tab

This tab specifies which reports and graphs are displayed as well as their format.

Skip Line After

The names of the terms can be too long to fit in the space provided. If the name contains more characters than this, the rest of the output is placed on a separate line. Enter ‘1’ when you want every term’s results printed on two lines. Enter ‘100’ when you want every variable’s results printed on one line.

Test in Summary Statement

Indicate the test that is to be reported on in the Summary Statements.

Maximum Term-Order Reported

Indicate the maximum order of terms to be reported on. Occasionally, higher-order interactions are of little interest and so they may be omitted. For example, enter a ‘2’ to limit output to individual factors and two-way interactions.

Example 1 - Determining Sample Size

Researchers are planning a study of the impact of a drug on heart rate. They want to evaluate the differences in heart rate among three age groups: 20-40, 41-60, and over 60. Their experimental protocol calls for a baseline heart rate measurement, followed by administration of a certain level of the drug, followed by three additional measurements 30 minutes apart. They want to be able to detect a 10% difference in heart rate among the age groups. They want to detect 5% difference in heart rate within an individual across time. They decide the experiment should detect interaction effects of the same magnitude as the time factor. They plan to analyze the data using a Geisser-Greenhouse corrected F test.

Similar studies have found an average heart rate of 93, a standard deviation of 4, and an autocorrelation between adjacent measurements on the same individual of 0.7. The researchers assume that first-order autocorrelation adequately represents the autocorrelation pattern.

From a heart rate of 93, a 10% reduction gives 84. They decide on the age-group means of 93, 87, and 84. Similarly, a 5% reduction within a subject would result in a heart rate of 88. They decide on time means of 93, 89, 88, and 91.

How many subjects per age group are needed to achieve 95% power and a 0.05 significance level?

Setup

You can enter these values yourself or load the Example1 template from the Template tab.

Option	**Value**
Data 1 Tab	
Find	**Beta and Power**
Means Matrix	blank
n	**2 to 8 by 1**
=n's	**checked**
K	**1.0**
For First Between-Subject Factor	
Label	**B**
Levels	**3**
Alpha	**0.05**
Beta	*Ignored since this is the Find setting*
Sm	**93 87 84**
For First Within-Subject Factor	
Label	**W**
Levels	**4**
Alpha	**0.05**
Beta	*Ignored since this is the Find setting*
Sm	**93 89 88 91**

Data 2 Tab

2-Way(Mixed) Interaction

Alpha...**0.05**

Beta..*Ignored since this is the Find setting*

Sm..**W 100**

Covariance Tab

Specify Covariance Method..............**1) Standard Deviations and Autocorrelations**

How SD's Change Across Time........**Constant**

Time Metric.......................................**STEP 0 1**

SD1...**4**

Specify How Autocorr's Change.......**1st Order Autocorr**

R1...**0.7**

Max Time Diff....................................**100** (This large value will cause R2 to be ignored.)

Report Tab

Numeric Results by Term.................**Checked**

Numeric Results by Design...............**Checked**

Regular F Test..................................**Not checked**

GG F Test...**Checked**

Wilks' Lambda..................................**Not checked**

Pillai-Bartlett....................................**Not checked**

Hotelling-Lawley................................**Not checked**

GG Detail Report..............................**Checked**

Means Matrix....................................**Checked**

Covariance Matrix.............................**Checked**

Show Plot 1.......................................**Checked**

Show Plot 2.......................................**Not checked**

Test in Summary Statement.............**GG F Test**

Max Term-Order Reported................**2**

Plot Setup Tab

Max Term-Order Plotted...................**2**

Test That is Plotted...........................**GG F Test**

Annotated Output

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

Design Report

Term	Test	Power	n	N	Multiply Means By	SD of Effects (Sm)	Standard Deviation (Sigma)	Effect Size	Alpha	Beta
B(3)	GG F	0.3096	2	6	1.00	3.74	3.29	1.14	0.0500	0.6904
W(4)	GG F	0.3266	2	6	1.00	1.92	1.31	1.47	0.0500	0.6734
BW	GG F	0.1588	2	6	1.00	1.92	1.31	1.47	0.0500	0.8412
B(3)	GG F	0.6459	3	9	1.00	3.74	3.29	1.14	0.0500	0.3541
W(4)	GG F	0.8099	3	9	1.00	1.92	1.31	1.47	0.0500	0.1901
BW	GG F	0.6267	3	9	1.00	1.92	1.31	1.47	0.0500	0.3733
B(3)	GG F	0.8478	4	12	1.00	3.74	3.29	1.14	0.0500	0.1522
W(4)	GG F	0.9486	4	12	1.00	1.92	1.31	1.47	0.0500	0.0514
BW	GG F	0.8598	4	12	1.00	1.92	1.31	1.47	0.0500	0.1402
B(3)	GG F	0.9415	5	15	1.00	3.74	3.29	1.14	0.0500	0.0585
W(4)	GG F	0.9871	5	15	1.00	1.92	1.31	1.47	0.0500	0.0129
BW	GG F	0.9535	5	15	1.00	1.92	1.31	1.47	0.0500	0.0465
B(3)	GG F	0.9793	6	18	1.00	3.74	3.29	1.14	0.0500	0.0207
W(4)	GG F	0.9970	6	18	1.00	1.92	1.31	1.47	0.0500	0.0030
BW	GG F	0.9860	6	18	1.00	1.92	1.31	1.47	0.0500	0.0140
B(3)	GG F	0.9931	7	21	1.00	3.74	3.29	1.14	0.0500	0.0069
W(4)	GG F	0.9993	7	21	1.00	1.92	1.31	1.47	0.0500	0.0007
BW	GG F	0.9961	7	21	1.00	1.92	1.31	1.47	0.0500	0.0039
B(3)	GG F	0.9978	8	24	1.00	3.74	3.29	1.14	0.0500	0.0022
W(4)	GG F	0.9999	8	24	1.00	1.92	1.31	1.47	0.0500	0.0001
BW	GG F	0.9990	8	24	1.00	1.92	1.31	1.47	0.0500	0.0010

The *Design Report* gives the power for each term in the design for each value of *n*. It is useful when you want to compare the powers of the terms in the design at a specific sample size.

In this example, the design goals of 0.95 power on all terms are achieved for $n = 6$.

The definitions of each of the columns of the report are as follows.

Term
This column contains the identifying label of the term. The number of levels for a factor is given in parentheses.

Test
This column identifies the test statistic. Since the power depends on the test statistic, you should make sure that this is the test statistic that you will use.

Power
This is the computed power for the term.

n
The value of *n* is the number of subjects per group.

N
The value of *N* is the total number of subjects in the study.

Multiply Means By

This is the value of the means multiplier, *K*.

SD of Effects (Sm)

This is the standard deviation of the effects σ_m for this term.

Standard Deviation

This is the value of σ, the random variation that σ_m is compared against by the *F* test. See the Technical Details for details on how these values are calculated.

Effect Size

The Effect Size is calculated by the expression σ_m / σ. It is an index of the size of the effect values relative to the standard deviation. Its value may be compared from experiment to experiment, regardless of the scale of the response variable.

Alpha

Alpha is the significance level of the test

Beta

Beta is the probability of failing to reject the null hypothesis when the alternative hypothesis is true.

Term Reports

Results for Factor B (Levels = 3)

Test	Power	n	N	Multiply Means By	SD of Effects (Sm)	Standard Deviation (Sigma)	Effect Size	Alpha	Beta
GG F	0.3096	2	6	1.00	3.74	3.29	1.14	0.0500	0.6904
GG F	0.6459	3	9	1.00	3.74	3.29	1.14	0.0500	0.3541
GG F	0.8478	4	12	1.00	3.74	3.29	1.14	0.0500	0.1522
GG F	0.9415	5	15	1.00	3.74	3.29	1.14	0.0500	0.0585
GG F	0.9793	6	18	1.00	3.74	3.29	1.14	0.0500	0.0207
GG F	0.9931	7	21	1.00	3.74	3.29	1.14	0.0500	0.0069
GG F	0.9978	8	24	1.00	3.74	3.29	1.14	0.0500	0.0022

Results for Factor W (Levels = 4)

Test	Power	n	N	Multiply Means By	SD of Effects (Sm)	Standard Deviation (Sigma)	Effect Size	Alpha	Beta
GG F	0.3266	2	6	1.00	1.92	1.31	1.47	0.0500	0.6734
GG F	0.8099	3	9	1.00	1.92	1.31	1.47	0.0500	0.1901
GG F	0.9486	4	12	1.00	1.92	1.31	1.47	0.0500	0.0514
GG F	0.9871	5	15	1.00	1.92	1.31	1.47	0.0500	0.0129
GG F	0.9970	6	18	1.00	1.92	1.31	1.47	0.0500	0.0030
GG F	0.9993	7	21	1.00	1.92	1.31	1.47	0.0500	0.0007
GG F	0.9999	8	24	1.00	1.92	1.31	1.47	0.0500	0.0001

Results for Term BW

Test	Power	n	N	Multiply Means By	SD of Effects (Sm)	Standard Deviation (Sigma)	Effect Size	Alpha	Beta
GG F	0.1588	2	6	1.00	1.92	1.31	1.47	0.0500	0.8412
GG F	0.6267	3	9	1.00	1.92	1.31	1.47	0.0500	0.3733
GG F	0.8598	4	12	1.00	1.92	1.31	1.47	0.0500	0.1402
GG F	0.9535	5	15	1.00	1.92	1.31	1.47	0.0500	0.0465
GG F	0.9860	6	18	1.00	1.92	1.31	1.47	0.0500	0.0140
GG F	0.9961	7	21	1.00	1.92	1.31	1.47	0.0500	0.0039
GG F	0.9990	8	24	1.00	1.92	1.31	1.47	0.0500	0.0010

The *Term Reports* provide a complete report for each term at all sample sizes. They are especially useful when you are only interested in the power of one or two terms.

The definitions of each of the columns of the report are identical to the corresponding columns in the *Design Report*, so they are not repeated here.

Geisser-Greenhouse Correction Detail Report

Term (Levels)	Power	Alpha	Critical F	Lambda	df1|df2	Epsilon	E(Epsilon)	G1
n = 2 N = 6 Means x 1								
B (3)	0.3096	0.0500	9.55	7.74	2|3	1.00	1.00	0.00
W (4)	0.3266	0.0500	7.60	12.88	3|9	0.77	0.43	-1.01
BW	0.1588	0.0500	6.92	12.88	6|9	0.77	0.43	-1.01
n = 3 N = 9 Means x 1								
B (3)	0.6459	0.0500	5.14	11.62	2|6	1.00	1.00	0.00
W (4)	0.8099	0.0500	4.12	19.32	3|18	0.77	0.60	-1.01
BW	0.6267	0.0500	3.46	19.32	6|18	0.77	0.60	-1.01
n = 4 N = 12 Means x 1								
B (3)	0.8478	0.0500	4.26	15.49	2|9	1.00	1.00	0.00
W (4)	0.9486	0.0500	3.58	25.76	3|27	0.77	0.66	-1.01
BW	0.8598	0.0500	2.95	25.76	6|27	0.77	0.66	-1.01
n = 5 N = 15 Means x 1								
B (3)	0.9415	0.0500	3.89	19.36	2|12	1.00	1.00	0.00
W (4)	0.9871	0.0500	3.36	32.20	3|36	0.77	0.68	-1.01
BW	0.9535	0.0500	2.75	32.20	6|36	0.77	0.68	-1.01
n = 6 N = 18 Means x 1								
B (3)	0.9793	0.0500	3.68	23.23	2|15	1.00	1.00	0.00
W (4)	0.9970	0.0500	3.25	38.64	3|45	0.77	0.70	-1.01
BW	0.9860	0.0500	2.64	38.64	6|45	0.77	0.70	-1.01
n = 7 N = 21 Means x 1								
B (3)	0.9931	0.0500	3.55	27.11	2|18	1.00	1.00	0.00
W (4)	0.9993	0.0500	3.17	45.07	3|54	0.77	0.71	-1.01
BW	0.9961	0.0500	2.57	45.07	6|54	0.77	0.71	-1.01
n = 8 N = 24 Means x 1								
B (3)	0.9978	0.0500	3.47	30.98	2|21	1.00	1.00	0.00
W (4)	0.9999	0.0500	3.12	51.51	3|63	0.77	0.72	-1.01
BW	0.9990	0.0500	2.52	51.51	6|63	0.77	0.72	-1.01

This report gives the details of the components of the Geisser-Greenhouse correction for each term and sample size. It is useful when you want to compare various aspects of this test.

The definitions of each of the columns of the report are as follows.

Term

This column contains the identifying label of the term. For factors, the number of levels is also given in parentheses.

Power

This is the computed power for the term.

Alpha

Alpha is the significance level of the test.

Critical F

This is the critical value of the F statistic. An *F* value computed from the data that is larger than this value is statistically significant at the alpha level given.

Lambda

This is the value of the noncentrality parameter λ of the approximate noncentral *F* distribution.

df1|df2

These are the values of the numerator and denominator degrees of freedom of the approximate *F* test that is used. These values are useful when comparing various designs. Other things being equal, you would like to have df2 large and df1 small.

Epsilon

The Geisser-Greenhouse epsilon is a measure of how far the covariance matrix departs from the assumption of circularity.

E(Epsilon)

This is the expected value of epsilon. It is a measure of how far the covariance matrix departs from the assumption of circularity.

G1

G1 is part of a correction factor used to convert ε to $E\left(\hat{\varepsilon}\right)$. It is reported for your convenience.

Summary Statements

A repeated measures design with 1 between factor and 1 within factor has 3 groups with 2 subjects each for a total of 6 subjects. Each subject is measured 4 times. This design achieves 31% power to test factor B if a Geisser-Greenhouse Corrected F Test is used with a 5% significance level and the actual effect standard deviation is 3.74 (an effect size of 1.14), achieves 33% power to test factor W if a Geisser-Greenhouse Corrected F Test is used with a 5% significance level and the actual effect standard deviation is 1.92 (an effect size of 1.47), and achieves 16% power to test the BW interaction if a Geisser-Greenhouse Corrected F Test is used with a 5% significance level and the actual effect standard deviation is 1.92 (an effect size of 1.47).

A summary statement can be generated for each sample size that is used. This statement gives the results in sentence form. The number of designs reported on textually is controlled by the Summary Statement option on the Report tab.

Means Matrix

Name	B1	B2	B3
W1	-10.62	5.19	-2.72
W2	1.46	3.97	2.72
W3	-4.58	4.58	0.00
W4	-4.58	4.58	0.00

This report shows the means matrix that was read in from the spreadsheet or generated by the Sm values that were given. It may be used to get an impression of the magnitude of the difference among the means that is being studied. When a Means Multiplier, *K*, is used, each value of *K* is multiplied times each value of this matrix.

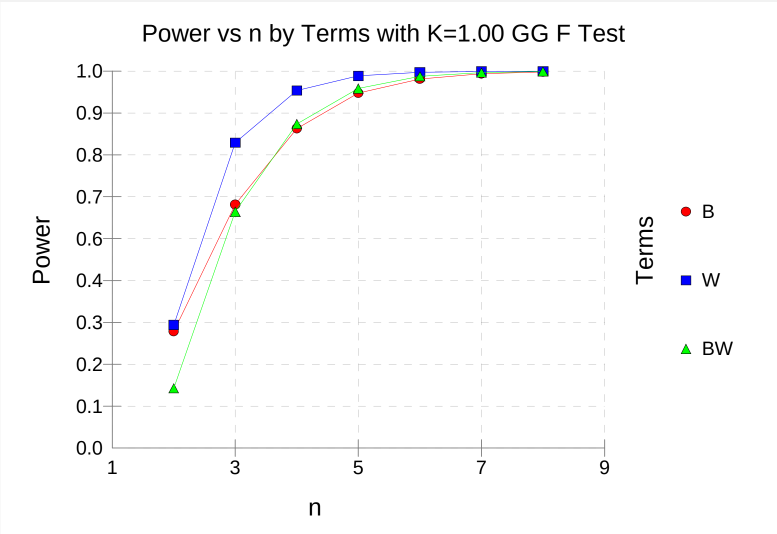
Variance-Covariance Matrix Section

Variance-Covariance Matrix Section

Time	W1	W2	W3	W4
W1	4.00	0.70	0.49	0.34
W2	0.70	4.00	0.70	0.49
W3	0.49	0.70	4.00	0.70
W4	0.34	0.49	0.70	4.00

This report shows the variance-covariance matrix that was read in from the spreadsheet or generated by the settings of on the Covariance tab. The standard deviations are given on the diagonal and the autocorrelations are given off the diagonal.

Chart Section



The chart shows the relationship between power and *n* for the terms in the design. Note that high-order interactions may be omitted from the plot by reducing the Max Term-Order Plotted option on the Plot Setup tab.

Example 2 - Varying the Difference Between the Means

Continuing with Example 1, the researchers want to evaluate the impact on power of varying the size of the difference among the means for a range of sample sizes from 2 to 8 per groups. The researchers could try calculating various multiples of the means, inputting them, and recording the results. This can be accomplished very easily by using the *K* option.

Keeping all other settings as in Example 1, the value of *K* is varied from 0.2 to 3.0 in steps of 0.2. We determined these values by experimentation so that a full range of power values are shown on the plots.

In the output to follow, we only display the plots. You may want to display the numeric reports as well, but we do not here in order to save space.

Setup

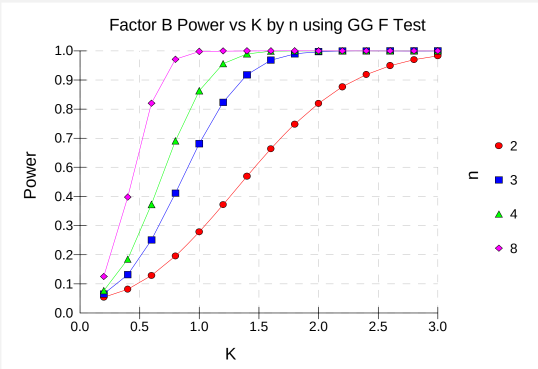
You can enter these values yourself or load the Example2 template from the Template tab.

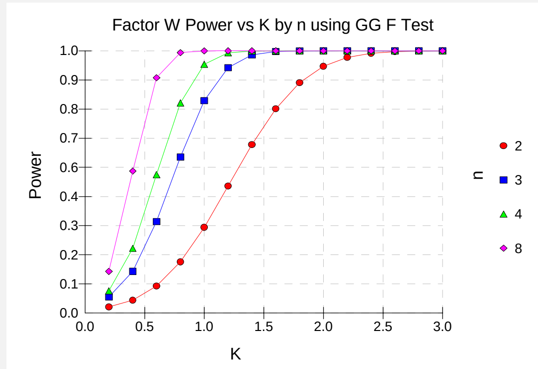
Option	**Value**
Data 1 Tab	
Find	**Beta and Power**
Means Matrix	blank
n	**2 3 4 8**
=n's	**checked**
K	**0.2 to 3.0 by 0.2**
For First Between-Subject Factor	
Label	**B**
Levels	**3**
Alpha	**0.05**
Beta	*Ignored since this is the Find setting*
Sm	**93 87 84**
For First Within-Subject Factor	
Label	**W**
Levels	**4**
Alpha	**0.05**
Beta	*Ignored since this is the Find setting*
Sm	**93 89 88 91**
Data 2 Tab	
2-Way(Mixed) Interaction	
Alpha	**0.05**
Beta	*Ignored since this is the Find setting*
Sm	**W 100**

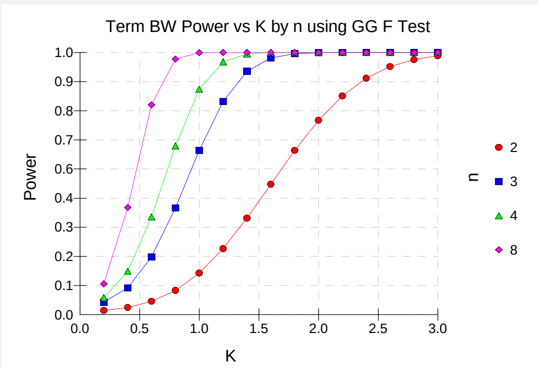
Covariance Tab
Specify Covariance Method ....**1) Standard Deviations and Autocorrelations**
How SD's Change Across Time........**Constant**
Time Metric ....**STEP 0 1**
SD1 ....**4**
Specify How Autocorr's Change ....**1st Order Autocorr**
R1..**0.7**
Max Time Diff. ....**100** (This large value will cause R2 to be ignored.)
Report Tab
Numeric Results by Term..................**Not checked**
Numeric Results by Design...............**Not checked**
Regular F Test ....**Not checked**
GG F Test ....**Not checked**
Wilks' Lambda...................................**Not checked**
Pillai-Bartlett ....**Not checked**
Hotelling-Lawley................................**Not checked**
GG Detail Report...............................**Not checked**
Means Matrix.....................................**Not checked**
Covariance Matrix ....**Not checked**
Show Plot 1 ....**Not checked**
Show Plot 2 ....**Checked**
Plot Setup Tab
Max Term-Order Plotted....................**2**
Test That is Plotted ....**GG F Test**

Annotated Output

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







These charts show how the power depends on the relative size of *Sm* (i.e. *K*) as well as the group sample size *n*.

Example 3 - Impact of the Number of Repeated Measurements

Continuing with Example 1, the researchers want to study the impact on power of changing the number of measurements made on each individual. Their experimental protocol calls for four measurements that are 30 minutes apart. They want to see the impact of taking twice that many measurements. To keep the output simple and two the point, they decide to look at the case when n = 4 and K = 1.

Setup

You can enter these values yourself or load the Example3 template from the Template tab.

Option	Value
Data 1 Tab	
Find	**Beta and Power**
Means Matrix	blank
n	**4**
=n's	**checked**
K	**1.0**
For First Between-Subject Factor	
Label	**B**
Levels	**3**
Alpha	**0.05**
Beta	*Ignored since this is the Find setting*
Sm	**93 87 84**
For First Within-Subject Factor	
Label	**W**
Levels	**4**
Alpha	**0.05**
Beta	*Ignored since this is the Find setting*
Sm	**RANGE 88 93**
Data 2 Tab	
2-Way(Mixed) Interaction	
Alpha	**0.05**
Beta	*Ignored since this is the Find setting*
Sm	**W 100**
Covariance Tab	
Specify Covariance Method	**1) Standard Deviations and Autocorrelations**
How SD's Change Across Time	**Constant**
Time Metric	**RANGE 0 3**
SD1	**4**
Specify How Autocorr's Change	**1st Order Autocorr**
R1	**0.7**
Max Time Diff.	**100** (This large value will cause R2 to be ignored.)

Report Tab

Numeric Results by Term .... **Not checked**
Numeric Results by Design............... **Checked**
Regular F Test .... **Not checked**
GG F Test .... **Checked**
Wilks' Lambda................................... **Not checked**
Pillai-Bartlett...................................... **Not checked**
Hotelling-Lawley................................ **Not checked**
GG Detail Report .... **Not checked**
Means Matrix .... **Not checked**
Covariance Matrix .... **Not checked**
Show Plot 1....................................... **Not checked**
Show Plot 2....................................... **Not checked**
Max Term-Order Reported................ **2**

Annotated Output

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

Design Report

Term	Test	Power	n	N	Multiply Means By	SD of Effects (Sm)	Standard Deviation (Sigma)	Effect Size	Alpha	Beta
(Results with 4 measurements)										
B(3)	GG F	0.8478	4	12	1.00	3.74	3.29	1.14	0.0500	0.1522
W(4)	GG F	0.9350	4	12	1.00	1.86	1.31	1.42	0.0500	0.0650
BW	GG F	0.8348	4	12	1.00	1.86	1.31	1.42	0.0500	0.1652
(Results with 8 measurements)										
B(3)	GG F	0.8375	4	12	1.00	3.74	3.34	1.12	0.0500	0.1625
W(8)	GG F	0.9354	4	12	1.00	1.64	0.83	1.97	0.0500	0.0646
BW	GG F	0.8321	4	12	1.00	1.64	0.83	1.97	0.0500	0.1679

Notice that the power of the between subjects factor decreased slightly, the power of the within-subjects factor increased slightly, and the power of the interaction test decreased slightly. This pattern of increase or decrease depends on all the settings.

We tried varying the value of the autocorrelation from 0.7 to 0.1 and found this to impact the direction of the change in the number of measurements. Hence, our conclusion is that there is no single answer. Changing the number of measurements may increase or decrease the power of a specific test depending on the values of the other parameters.

Example 4 - Power after a study

This example will show how to calculate the power of *F* tests from data that have already been collected and analyzed using the analysis of variance. The following results were obtained using the analysis of variance procedure in ***NCSS***. In this example, Gender is the between factor with two levels and Treatment is the within factor with three levels. The experiment was conducted with two subjects per group, but there is interest in the power for 2, 3, and 4 subjects per group. All significance levels are set to 0.05.

Analysis of Variance Table

Source Term	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level
A (Gender)	1	21.33333	21.33333	32.00	0.029857
B(A)	2	1.333333	0.6666667		
C (Treatment)	2	5.166667	2.583333	6.20	0.059488
AC	2	5.166667	2.583333	6.20	0.059488
BC(A)	4	1.666667	0.4166667		
Total (Adjusted)	11	34.66667			
Total	12				

Means and Effects Section

Term	Count	Mean	Standard Error
All	12	17.33333	
A: Gender			
Females	6	16	0.3333333
Males	6	18.66667	0.3333333
C: Treatment			
L	4	16.75	0.3227486
M	4	17	0.3227486
H	4	18.25	0.3227486
AC: Gender,Treatment			
Females,L	2	14.5	0.4564355
Females,M	2	16	0.4564355
Females,H	2	17.5	0.4564355
Males,L	2	19	0.4564355
Males,M	2	18	0.4564355
Males,H	2	19	0.4564355

Note that the treatment means (L, M, and H) show an increasing pattern from 16.75 to 18.25, but the hypothesis test of this factor is not statistically significant at the 0.05 level. We will now calculate the power of the three *F* tests using ***PASS***. We will use the regular F test since that is what was used in the above table.

From the above printout, we note that MSB = 0.6666667 and MSW = 0.4166667. Plugging these values into the estimating equations

$$\hat{\rho} = \frac{MSB - MSW}{MSB + (T-1)MSW}$$

$$\hat{\sigma}^2 = \frac{MSW}{1-\hat{\rho}}$$

yields

$$\hat{\rho} = \frac{0.6666667 - 0.4166667}{0.6666667 + (3-1)0.4166667} = 0.16666667$$

$$\hat{\sigma}^2 = \frac{0.4166667}{1-0.16666667} = 0.5$$

so that

$$\hat{\sigma} = \sqrt{0.5} = 0.70710681$$

With these values calculated, we can setup ***PASS*** to calculate the power of the three *F* tests as follows.

Setup

You can enter these values yourself or load the Example4 template from the Template tab.

Option **Value**

Data 1 Tab

Find .... **Beta and Power**

Means Matrix .... **FEMALES-MALES**

n **2 3 4**

=n's .... **checked**

K... **1.0**

For First Between-Subject Factor

Label .... **B**

Levels.. **2**

Alpha... **0.05**

Beta... *Ignored since this is the Find setting.*

Sm... *Ignored since the Means Matrix is loaded.*

For First Within-Subject Factor

Label .... **W**

Levels.. **3**

Alpha... **0.05**

Beta... *Ignored since this is the Find setting.*

Sm... *Ignored since the Means Matrix is loaded.*

Data 2 Tab

2-Way(Mixed) Interaction

Alpha... **0.05**

Beta... *Ignored since this is the Find setting.*

Sm... *Ignored since the Means Matrix is loaded.*

Covariance Tab

Specify Covariance Method .... **1) Standard Deviations and Autocorrelations**

How SD's Change Across Time........ **Constant**

Time Metric .... **STEP 0 1**

SD1 .... **.70710681**

Specify How Autocorr's Change .... **Constant**

R1 .... **0.16666667**

Max Time Diff.................................... **100** (This large value will cause R2 to be ignored.)

Report Tab

Numeric Results by Term..................**Checked**
Numeric Results by Design...............**Not Checked**
Regular F Test ....**Checked**
GG F Test ....**Not checked**
Wilks' Lambda..................................**Not checked**
Pillai-Bartlett ....**Not checked**
Hotelling-Lawley................................**Not checked**
GG Detail Report...............................**Not checked**
Means Matrix.....................................**Not checked**
Covariance Matrix ....**Not checked**
Show Plot 1 ....**Not checked**
Show Plot 2 ....**Not checked**
Test in Summary Statement.............**Regular F Test**
Max Term-Order Reported................**2**

Annotated Output

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

Numeric Results

Term	Test	Power	n	N	Multiply Means By	SD of Effects (Sm)	Standard Deviation (Sigma)	Effect Size	Alpha	Beta
B(2)	F	0.8004	2	4	1.00	1.33	0.47	2.83	0.0500	0.1996
W(3)	F	0.5536	2	4	1.00	0.66	0.37	1.76	0.0500	0.4464
BW	F	0.5536	2	4	1.00	0.66	0.37	1.76	0.0500	0.4464
B(2)	F	0.9985	3	6	1.00	1.33	0.47	2.83	0.0500	0.0015
W(3)	F	0.8933	3	6	1.00	0.66	0.37	1.76	0.0500	0.1067
BW	F	0.8933	3	6	1.00	0.66	0.37	1.76	0.0500	0.1067
B(2)	F	1.0000	4	8	1.00	1.33	0.47	2.83	0.0500	0.0000
W(3)	F	0.9801	4	8	1.00	0.66	0.37	1.76	0.0500	0.0199
BW	F	0.9801	4	8	1.00	0.66	0.37	1.76	0.0500	0.0199

You can see that the power of the tests on W and BW was only 0.55 for an *n* of 2. However, if *n* would have been 3, a much more reasonable power of 0.89 would have been achieved.

Example5 - Cross-over Design

A *crossover design* is a special type of repeated measures design in which the treatments are applied to the subjects in different orders. The between-subjects (grouping) factor is defined by the specific sequence in which the treatments are applied. For example, suppose the treatments are represented by B1 and B2. Further suppose that half the subjects receive treatment B1 followed by treatment B2 (sequence B1B2), while the other half receive treatment B2 followed by treatment B1 (sequence B2B1). This is a two-group crossover design.

Crossover designs assume that a long enough period elapses between measurements so that the effects of one treatment are *washed out* before the next treatment is applied. This is known as the assumption of no *carryover* effects.

When a crossover design is analyzed using repeated measures, the interaction is the only term of interest. The *F* test on the between factor tests whether averages across each sequence are equal—a test of little interest. The *F* test on the within factor tests whether the response is different across the time periods—also of little interest. The *F* test for interaction tests whether the change in response across time is the same for both sequences. The interaction can only be significant if the treatments effect the outcome differently. Hence, to specify a crossover design requires the careful specification of the interaction effects.

With this background, we present an example. Suppose researchers want to investigate the reduction in heart-beat rate caused by the administration of a certain drug using a simple two-period crossover design. The researchers want a sample size large enough to detect a drop in heart-beat rate from 95 to 90 with a power of 90% at the 0.05 significance level. Previous studies have shown a within-patient autocorrelation of 0.50 and a standard deviation of 3.98. They decide to consider sample sizes between 2 and 8.

The hypothesized interaction is specified by entering the mean heart-beat rates of the four treatment groups as 95, 90, 90, and 95. Since the standard deviation of these values is all that is used, the order of these values does not matter. In this case the sequences means are both 92.5 and the average time-period means are both 92.5. Hence, the interaction effects are 2.5, -2.5, -2.5, and 2.5. You can check that the set of numbers ‘95, 90, 90, 95’ has the same standard deviation as the set ‘2.5, -2.5, -2.5, 2.5’ or even ‘5, 0, 0, 5’. All of these will work.

Setup

You can enter these values yourself or load the Example5 template from the Template tab.

<u>**Option**</u>	<u>**Value**</u>
Data 1 Tab	
Find	**Beta and Power**
Means Matrix	blank
n	**2 to 8 by 1**
=n’s	**checked**
K	**1.0**

For First Between-Subject Factor
Label ....**Seq**
Levels...**2**
Alpha..**0.05**
Beta...*Ignored since this is the Find setting*
Sm..**0.1** (This value is arbitrary.)
For First Within-Subject Factor
Label ....**Time**
Levels...**2**
Alpha..**0.05**
Beta...*Ignored since this is the Find setting*
Sm..**0.1** (This value is arbitrary.)
Data 2 Tab
2-Way(Mixed) Interaction
Alpha..**0.05**
Beta...*Ignored since this is the Find setting*
Sm..**95 90 90 95**
Covariance Tab
Specify Covariance Method ....**1) Standard Deviations and Autocorrelations**
How SD's Change Across Time........**Constant**
Time Metric ....**STEP 0 1**
SD1 ....**3.98**
Specify How Autocorr's Change ....**1st Order Autocorr**
R1...**0.5**
Max Time Diff. ....**100** (This large value will cause R2 to be ignored.)
Report Tab
Numeric Results by Term..................**Checked**
Numeric Results by Design...............**Not checked**
Regular F Test ....**Not checked**
GG F Test ....**Checked**
Wilks' Lambda..................................**Not checked**
Pillai-Bartlett ....**Not checked**
Hotelling-Lawley................................**Not checked**
GG Detail Report...............................**Not checked**
Means Matrix....................................**Not checked**
Covariance Matrix ....**Not checked**
Show Plot 1 ....**Not checked**
Show Plot 2 ....**Not checked**
Max Term-Order Reported................**2**

Annotated Output

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

Numeric Results

Results for Term SeqTime

Test	Power	n	N	Multiply Means By	SD of Effects (Sm)	Standard Deviation (Sigma)	Effect Size	Alpha	Beta
GG F	0.3017	2	4	1.00	2.50	1.99	1.26	0.0500	0.6983
GG F	0.6401	3	6	1.00	2.50	1.99	1.26	0.0500	0.3599
GG F	0.8395	4	8	1.00	2.50	1.99	1.26	0.0500	0.1605
GG F	0.9338	5	10	1.00	2.50	1.99	1.26	0.0500	0.0662
GG F	0.9742	6	12	1.00	2.50	1.99	1.26	0.0500	0.0258
GG F	0.9903	7	14	1.00	2.50	1.99	1.26	0.0500	0.0097
GG F	0.9965	8	16	1.00	2.50	1.99	1.26	0.0500	0.0035

We only display the interaction term since that is the only term of interest. A quick glance at the plot shows that 90% power is achieved when *n* is five. This corresponds to a total sample size of ten subjects.

Example 6 - Power of a Completed Crossover Design

The following analysis of variance table was generated by ***NCSS*** for a set of crossover data. Find the power of the interaction *F* test assuming a significance level of 0.05.

Analysis of Variance Table

Source Term	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level
A: Sequence	1	89397.6	89397.6	1.19	0.285442
B(A): Subject	28	2110739	75383.54		
C: Period	1	117395.3	117395.3	1.40	0.246854
AC	1	122401.7	122401.7	1.46	0.237263
BC(A)	28	2349752	83919.72		
Total (Adjusted)	59	4789686			
Total	60				

Means Section

Term	Count	Mean	Standard Error
All	60	492.2000	
A: Sequence			
1	30	453.6000	50.12768
2	30	530.8000	50.12768
C: Period			
1	30	447.9667	52.88973
2	30	536.4333	52.88973
AC: Sequence,Period			
1,1	15	364.2000	74.79738
1,2	15	543.0000	74.79738
2,1	15	531.7333	74.79738
2,2	15	529.8666	74.79738

One difficulty in analyzing an existing crossover design is determining an appropriate value for the hypothesized interaction effects. One method is to find the standard deviation of the interaction effects by taking the square root of the Sum of Squares for the interaction divided by the total number of observations. In this case,

$$\begin{aligned}\sigma_{Interaction} &= \sqrt{\frac{122401.7}{60}} \\ &= 45.1667\end{aligned}$$

Another method is to find the individual interaction effects by subtraction. This method proceeds as follows.

First, subtract the Period means from the Sequence by Period means.

$$\begin{bmatrix} 364.2000 & 531.7333 \\ 543.0000 & 529.8666 \end{bmatrix} - \begin{bmatrix} 447.9667 \\ 536.4333 \end{bmatrix} = \begin{bmatrix} -83.7667 & 83.7667 \\ 6.5667 & -6.5667 \end{bmatrix}$$

Next, compute the column means and subtract them from the current values. This results in the effects.

$$\begin{bmatrix} -83.7667 & 83.7667 \\ 6.5667 & -6.5667 \end{bmatrix} - \begin{bmatrix} -38.6000 & 38.6000 \\ -38.6000 & 38.6000 \end{bmatrix} = \begin{bmatrix} -45.1667 & 45.1667 \\ 45.1667 & -45.1667 \end{bmatrix}$$

Finally, compute the standard deviation of the effects. Since the mean of the effects is zero, the standard deviation is

$$\sigma_{Interaction} = \sqrt{\frac{(-45.1667)^2 + (45.1667)^2 + (45.1667)^2 + (-45.1667)^2}{4}}$$
$$= 45.1667$$

Another difficulty that must be solved is to estimate the autocorrelation and within-subject standard deviation. From the above printout, we note that MSB = 75383.54 and MSW = 83919.72. Plugging these values into the estimating equations

$$\hat{\rho} = \frac{MSB - MSW}{MSB + (T-1)MSW}$$

$$\hat{\sigma}^2 = \frac{MSW}{1-\hat{\rho}}$$

yields

$$\hat{\rho} = \frac{75383.54 - 83919.72}{75383.54 + (2-1)83919.72} = -0.05358447$$

$$\hat{\sigma}^2 = \frac{83919.72}{1+0.05358447} = 79651.63$$

so that

$$\hat{\sigma} = \sqrt{79651.63} = 282.2262$$

Setup

You can enter these values yourself or load the Example6 template from the Template tab.

<u>Option</u> **<u>Value</u>**

Data 1 Tab

Find .... **Beta and Power**
Means Matrix .... blank
n **15**
=n's .... **checked**
K... **1.0**
For First Between-Subject Factor
Label .... **S**
Levels.. **2**
Alpha **0.05**
Beta... *Ignored since this is the Find setting*
Sm... **453.6000 530.8000**

For First Within-Subject Factor
Label ....**P**
Levels...**2**
Alpha **0.05**
Beta..*Ignored since this is the Find setting*
Sm..**447.9667 536.4333**
Data 2 Tab
2-Way(Mixed) Interaction
Alpha **0.05**
Beta..*Ignored since this is the Find setting*
Sm..**45.1667**
Covariance Tab
Specify Covariance Method ....**1) Standard Deviations and Autocorrelations**
How SD's Change Across Time........**Constant**
Time Metric ....**STEP 0 1**
SD1 ....**282.2262**
Specify How Autocorr's Change ....**1st Order Autocorr**
R1...**-0.05358447**
Max Time Diff. ....**100** (This large value will cause R2 to be ignored.)
Report Tab
Numeric Results by Term..................**Not checked**
Numeric Results by Design...............**Checked**
Regular F Test ....**Not checked**
GG F Test ....**Checked**
Wilks' Lambda...................................**Not checked**
Pillai-Bartlett ....**Not checked**
Hotelling-Lawley................................**Not checked**
GG Detail Report...............................**Not checked**
Means Matrix.....................................**Not checked**
Covariance Matrix ....**Not checked**
Show Plot 1 ....**Not checked**
Show Plot 2 ....**Not checked**
Summary Statement Rows................**0**
Max Term-Order Reported................**2**

Annotated Output

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

Numeric Results

Term	Test	Power	n	N	Multiply Means By	SD of Effects (Sm)	Standard Deviation (Sigma)	Effect Size	Alpha	Beta
S(2)	GG F	0.1832	15	30	1.00	38.60	194.14	0.20	0.0500	0.8168
P(2)	GG F	0.2078	15	30	1.00	44.23	204.84	0.22	0.0500	0.7922
SP	GG F	0.2147	15	30	1.00	45.17	204.84	0.22	0.0500	0.7853

Notice that these power values are low. Fifteen was not a large enough sample size to detect Sm values near 40.

Example 7 - Validation using O'Brien and Muller

O'Brien and Muller's article in the book edited by Edwards (1993) analyze the power of a two-group repeated-measures experiment in which three measurements are made on each subject.

The hypothesized means are

	Group 1	Group 2
Time 1	3	1
Time 2	12	5
Time 3	8	7

The covariance matrix is

	Time 1	Time 2	Time 3
Time 1	25	16	12
Time 2	16	64	30
Time 3	12	30	36

With *n*'s of 12, 18, and 24 and an alpha of 0.05, they obtained power values using the Wilks' Lambda test. Their reported power values are

	Power Values for each Term		
n	**Group**	**Time**	**Interaction**
12	0.326	0.983	0.461
18	0.467	0.999	0.671
24	0.589	0.999	0.814

O'Brien, in a private communication, re-ran these data using the Geisser-Greenhouse correction. His results were as follows:

	Power Values for each Term		
n	**Group**	**Time**	**Interaction**
12	0.326	0.993	0.486
18	0.467	0.999	0.685
24	0.589	0.999	0.819

In order to run this example in ***PASS***, the values of the means and the covariance matrix (given above) must be entered on a spreadsheet. We have loaded these values into the database called OBRIEN. Either enter the values yourself, or load the OBRIEN database which should be in the Data directory. The instructions below assume that the means are in columns one and two, while the covariance matrix is in columns four through six of the current database.

Setup

You can enter these values yourself or load the Example7 template from the Template tab.

Option	Value
Data 1 Tab	
Find	**Beta and Power**
Means Matrix	**M1-M2**
n	**12 18 24**
=n's	**checked**
K	**1.0**
For First Between-Subject Factor	
Label	**G**
Levels	**2**
Alpha	**0.05**
Beta	*Ignored since this is the Find setting*
Sm	*Ignored*
For First Within-Subject Factor	
Label	**T**
Levels	**3**
Alpha	**0.05**
Beta	*Ignored since this is the Find setting*
Sm	*Ignored*
Data 2 Tab	
2-Way(Mixed) Interaction	
Alpha	**0.05**
Beta	*Ignored since this is the Find setting*
Sm	*Ignored*
Covariance Tab	
Specify Covariance Method	**2) Covariance Matrix Variables**
Spreadsheet Columns	**S1-S3**
Report Tab	
Numeric Results by Term	**Checked**
Numeric Results by Design	**Not checked**
Regular F Test	**Not checked**
GG F Test	**Checked**
Wilks' Lambda	**Checked**
Pillai-Bartlett	**Not checked**
Hotelling-Lawley	**Not checked**
GG Detail Report	**Not checked**
Means Matrix	**Not checked**
Covariance Matrix	**Not checked**
Show Plot 1	**Not checked**
Show Plot 2	**Not checked**
Summary Statement Rows	**0**
Max Term-Order Reported	**2**

Annotated Output

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

Numeric Results

Results for Factor G (Levels =2)

Test	Power	n	N	Multiply Means By	SD of Effects (Sm)	Standard Deviation (Sigma)	Effect Size	Alpha	Beta
GG F	0.3263	12	24	1.00	1.67	5.17	0.32	0.0500	0.6737
Wilks	0.3263	12	24	1.00	1.67	5.17	0.32	0.0500	0.6737
GG F	0.4673	18	36	1.00	1.67	5.17	0.32	0.0500	0.5327
Wilks	0.4673	18	36	1.00	1.67	5.17	0.32	0.0500	0.5327
GG F	0.5889	24	48	1.00	1.67	5.17	0.32	0.0500	0.4111
Wilks	0.5889	24	48	1.00	1.67	5.17	0.32	0.0500	0.4111

Results for Factor T (Levels =3)

Test	Power	n	N	Multiply Means By	SD of Effects (Sm)	Standard Deviation (Sigma)	Effect Size	Alpha	Beta
GG F	0.9933	12	24	1.00	2.86	2.73	1.05	0.0500	0.0067
Wilks	0.9825	12	24	1.00	2.86	2.73	1.05	0.0500	0.0175
GG F	0.9999	18	36	1.00	2.86	2.73	1.05	0.0500	0.0001
Wilks	0.9995	18	36	1.00	2.86	2.73	1.05	0.0500	0.0005
GG F	1.0000	24	48	1.00	2.86	2.73	1.05	0.0500	0.0000
Wilks	1.0000	24	48	1.00	2.86	2.73	1.05	0.0500	0.0000

Results for Term GT

Test	Power	n	N	Multiply Means By	SD of Effects (Sm)	Standard Deviation (Sigma)	Effect Size	Alpha	Beta
GG F	0.4861	12	24	1.00	1.31	2.73	0.48	0.0500	0.5139
Wilks	0.4605	12	24	1.00	1.31	2.73	0.48	0.0500	0.5395
GG F	0.6850	18	36	1.00	1.31	2.73	0.48	0.0500	0.3150
Wilks	0.6706	18	36	1.00	1.31	2.73	0.48	0.0500	0.3294
GG F	0.8193	24	48	1.00	1.31	2.73	0.48	0.0500	0.1807
Wilks	0.8136	24	48	1.00	1.31	2.73	0.48	0.0500	0.1864

PASS agrees exactly with O'Brien's calculations.

Example 8 - Unequal Group Sizes

Usually, in the planning stages, the group sample sizes are equal. Occasionally, however, you may want to plan for a situation in which one group will have a much larger sample size than the others. Also, when doing a power analysis on a study that has already been conducted, the group sample sizes are often unequal.

In this example, we will re-analyze the Example 4. However, we will now assume that there were four subjects in group 1 and eight subjects in group 2. The setup and output for this example are as follows. Remember that you must open the database PASS RM EXAMPLE2 before running this example.

Setup

You can enter these values yourself or load the Example8 template from the Template tab.

Option	Value
Data 1 Tab	
Find	**Beta and Power**
Means Matrix	**FEMALES-MALES**
n	**4 8**
=n's	**Not checked**
K	**1.0**
For First Between-Subject Factor	
Label	**B**
Levels	**2**
Alpha	**0.05**
Beta	*Ignored since this is the Find setting.*
Sm	*Ignored since the Means Matrix is loaded.*
For First Within-Subject Factor	
Label	**W**
Levels	**3**
Alpha	**0.05**
Beta	*Ignored since this is the Find setting.*
Sm	*Ignored since the Means Matrix is loaded.*
Data 2 Tab	
2-Way(Mixed) Interaction	
Alpha	**0.05**
Beta	*Ignored since this is the Find setting.*
Sm	*Ignored since the Means Matrix is loaded.*

Covariance Tab
Specify Covariance Method ....**1) Standard Deviations and Autocorrelations**
How SD's Change Across Time........**Constant**
Time Metric ....**STEP 0 1**
SD1 ....**.70710681**
Specify How Autocorr's Change ....**Constant**
R1...**0.16666667**
Max Time Diff. ....**100** (This large value will cause R2 to be ignored.)
Report Tab
Numeric Results by Term..................**Checked**
Numeric Results by Design...............**Not Checked**
Regular F Test ....**Not checked**
GG F Test ....**Checked**
Wilks' Lambda...................................**Not checked**
Pillai-Bartlett ....**Not checked**
Hotelling-Lawley................................**Not checked**
GG Detail Report...............................**Not checked**
Means Matrix.....................................**Not checked**
Covariance Matrix ....**Not checked**
Show Plot 1 ....**Not checked**
Show Plot 2 ....**Not checked**
Max Term-Order Reported................**2**

Annotated Output

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

Numeric Results

Term	Test	Power	n	N	Multiply Means By	SD of Effects (Sm)	Standard Deviation (Sigma)	Effect Size	Alpha	Beta
B(2)	GG F	1.0000	6.0	12	1.00	1.26	0.47	2.67	0.0500	0.0000
W(3)	GG F	0.9983	6.0	12	1.00	0.62	0.37	1.66	0.0500	0.0017
BW	GG F	0.9983	6.0	12	1.00	0.62	0.37	1.66	0.0500	0.0017

n's: 4 8

Notice that the values of *n* are now shown to one decimal place. That is because the value reported is the average value of *n*. The actual *n*'s are shown following the report.

Example 9 - Designs with More Than Two Factors

Occasionally, you will have a design that has more than two factors. We will now show you how to compute the necessary sample size for such a design.

Suppose your design calls for two between-subject factors, Age (A) and Gender (G), and two within-subject factors, Dose-Level (D) and Application-Method (M). Suppose the number of levels of these four factors are, respectively, 3, 2, 4, and 2.

Our first task is to determine appropriate values of Sm for each of the terms. We decide to ignore the interactions during the planning and only consider the factors themselves. The desired difference to be detected among the three age groups can be represented by the means 80, 88, and 96. The desired difference to be detected among the two genders can be represented by the means 80 and 96. The desired difference to be detected among the four dose levels is represented by the means 80, 82, 84, and 86. The desired difference to be detected among the two application methods is represented by the means 80 and 86.

Our next task is to specify the covariance matrix. From previous experience, we have found that a constant value of 20.0 is appropriate for SD1. An autocorrelation of 0.5 with a first-order autocorrelation pattern is also appropriate.

Finally, we decide to calculate the power using the GG F test at the following sample sizes: 2, 4, 6, 8, 10, 20, 30, and 40.

Setup

You can enter the following values yourself or load the Example9 template.

<u>**Option**</u>	<u>**Value**</u>
Data 1 Tab	
Find	**Beta and Power**
Means Matrix	*blank*
n	**2 4 6 8 10 20 30 40**
=n's	**Checked**
K	**1.0**
For First Between-Subject Factor	
Label	**A**
Levels	**3**
Alpha	**0.05**
Beta	*Ignored since this is the Find setting.*
Sm	**80 88 96**
For Second Between-Subject Factor	
Label	**G**
Levels	**2**
Alpha	**0.05**
Beta	*Ignored since this is the Find setting.*
Sm	**80 96**
For First Within-Subject Factor	

Label ....**D**
Levels..**4**
Alpha **0.05**
Beta..*Ignored since this is the Find setting.*
Sm..**80 82 84 86**
For Second Within-Subject Factor
Label ....**M**
Levels..**2**
Alpha **0.05**
Beta..*Ignored since this is the Find setting.*
Sm..**80 86**

Data 2 Tab

All Interactions
Alpha **0.05**
Beta..*Ignored since this is the Find setting.*
Sm..**D 100**

Covariance Tab

Specify Covariance Method ....**1) Standard Deviations and Autocorrelations**
How SD's Change Across Time........**Constant**
Time Metric ....**STEP 0 1**
SD1 ....**20**
Specify How Autocorr's Change ....**1st Order Autocorr**
R1...**0.5**
Max Time Diff. ....**100** (This large value will cause R2 to be ignored.)

Report Tab

Numeric Results by Term..................**Not checked**
Numeric Results by Design...............**Checked**
Regular F Test ....**Not checked**
GG F Test ....**Checked**
Wilks' Lambda...................................**Not checked**
Pillai-Bartlett ....**Not checked**
Hotelling-Lawley................................**Not checked**
GG Detail Report...............................**Not checked**
Means Matrix.....................................**Not checked**
Covariance Matrix ....**Not checked**
Show Plot 1 ....**Checked**
Show Plot 2 ....**Not checked**
Test in Summary Statement..............**Regular F Test**
Max Term-Order Reported................**1**

Plot Setup Tab

Max Term-Order Plotted....................**1**
Test That is Plotted ....**GG F Test**

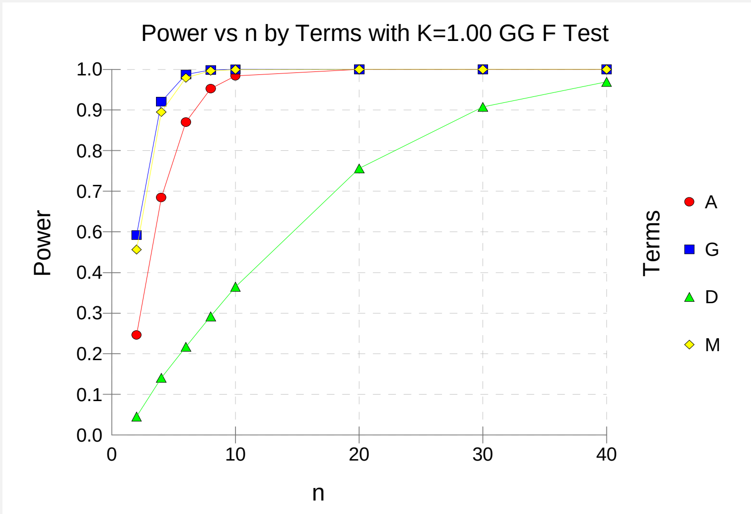
Annotated Output

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

Numeric Results

Term	Test	Power	n	N	Multiply Means By	SD of Effects (Sm)	Standard Deviation (Sigma)	Effect Size	Alpha	Beta
A(3)	GG F	0.2735	2	12	1.00	6.53	11.18	0.58	0.0500	0.7265
G(2)	GG F	0.5465	2	12	1.00	8.00	11.18	0.72	0.0500	0.4535
D(4)	GG F	0.0506	2	12	1.00	2.24	7.64	0.29	0.0500	0.9494
M(2)	GG F	0.5072	2	12	1.00	3.00	4.41	0.68	0.0500	0.4928
A(3)	GG F	0.6493	4	24	1.00	6.53	11.18	0.58	0.0500	0.3507
G(2)	GG F	0.9118	4	24	1.00	8.00	11.18	0.72	0.0500	0.0882
D(4)	GG F	0.1563	4	24	1.00	2.24	7.64	0.29	0.0500	0.8437
M(2)	GG F	0.8833	4	24	1.00	3.00	4.41	0.68	0.0500	0.1167
A(3)	GG F	0.8556	6	36	1.00	6.53	11.18	0.58	0.0500	0.1444
G(2)	GG F	0.9858	6	36	1.00	8.00	11.18	0.72	0.0500	0.0142
D(4)	GG F	0.2412	6	36	1.00	2.24	7.64	0.29	0.0500	0.7588
M(2)	GG F	0.9767	6	36	1.00	3.00	4.41	0.68	0.0500	0.0233
A(3)	GG F	0.9471	8	48	1.00	6.53	11.18	0.58	0.0500	0.0529
G(2)	GG F	0.9980	8	48	1.00	8.00	11.18	0.72	0.0500	0.0020
D(4)	GG F	0.3245	8	48	1.00	2.24	7.64	0.29	0.0500	0.6755
M(2)	GG F	0.9959	8	48	1.00	3.00	4.41	0.68	0.0500	0.0041
A(3)	GG F	0.9823	10	60	1.00	6.53	11.18	0.58	0.0500	0.0177
G(2)	GG F	0.9997	10	60	1.00	8.00	11.18	0.72	0.0500	0.0003
D(4)	GG F	0.4054	10	60	1.00	2.24	7.64	0.29	0.0500	0.5946
M(2)	GG F	0.9993	10	60	1.00	3.00	4.41	0.68	0.0500	0.0007
A(3)	GG F	1.0000	20	120	1.00	6.53	11.18	0.58	0.0500	0.0000
G(2)	GG F	1.0000	20	120	1.00	8.00	11.18	0.72	0.0500	0.0000
D(4)	GG F	0.7286	20	120	1.00	2.24	7.64	0.29	0.0500	0.2714
M(2)	GG F	1.0000	20	120	1.00	3.00	4.41	0.68	0.0500	0.0000
A(3)	GG F	1.0000	30	180	1.00	6.53	11.18	0.58	0.0500	0.0000
G(2)	GG F	1.0000	30	180	1.00	8.00	11.18	0.72	0.0500	0.0000
D(4)	GG F	0.8971	30	180	1.00	2.24	7.64	0.29	0.0500	0.1029
M(2)	GG F	1.0000	30	180	1.00	3.00	4.41	0.68	0.0500	0.0000
A(3)	GG F	1.0000	40	240	1.00	6.53	11.18	0.58	0.0500	0.0000
G(2)	GG F	1.0000	40	240	1.00	8.00	11.18	0.72	0.0500	0.0000
D(4)	GG F	0.9658	40	240	1.00	2.24	7.64	0.29	0.0500	0.0342
M(2)	GG F	1.0000	40	240	1.00	3.00	4.41	0.68	0.0500	0.0000

This report gives the power values for the various terms and sample sizes that were entered. It is much easier to consider the following plot to interpret the results.



From this chart, we can see that the first within-subject factor, dose level, has a power much lower than the other factors. Looking at the Sm values in the numeric table, we find that the Sm value for factor D is much less than for the other values. This explains why its power is so poor. Our options are to either increase the sample size or increase the value of Sm for factor D.

Chapter 575

Multiple Comparisons

Introduction

This module computes sample sizes for multiple comparison procedures. The term "*multiple comparison*" refers to the individual comparison of two means selected from a larger set of means. The module emphasizes one-way analysis of variance designs that use one of three multiple-comparison methods: Tukey's all pairs (MCA), comparisons with the best (MCB), or Dunnett's all versus a control (MCC). Because these sample sizes may be substantially different from those required for the usual *F* test, a separate module is provided to compute them.

There are only a few articles in the statistical literature on the computation of sample sizes for multiple comparison designs. This module is based almost entirely on the book by Hsu (1996). We can give only a brief outline of the subject here. Users who want more details are referred to Hsu's book.

Although this module is capable of computing sample sizes for unbalanced designs, it emphasizes balanced designs.

Technical Details

The One-Way Analysis of Variance Design

The summarized discussion that follows is based on the common, one-way analysis of variance design. Suppose the responses Y_{ij} in k groups each follow a normal distribution with respective 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 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 for analyzing such a design, and tests 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 group(s) is(are) different and by how much.

The probability statement associated with the *F* test is simple, straightforward, and easy to interpret. However, when several simultaneous comparisons are made among the group means, the interpretation of individual probability statements becomes much more complex. This is called 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. The method of *multiple comparisons* has been developed to account for such multiplicity.

Power Calculations for Multiple Comparisons

For technical reasons, the definition of power in the case of multiple comparisons is different from the usual definition. Following Hsu (1996) page 237, power is defined as follows.

Using a $1-\alpha$ simultaneous confidence interval multiple comparison method, power is the probability that the confidence intervals cover the true parameter values and are sufficiently narrow. Power is still defined to be $1-\beta$. Note that $1-\beta<1-\alpha$. Here, *narrow* refers to the width of the confidence intervals. The definition says that the confidence intervals should be as narrow as possible while still including the true parameter values. This definition may be restated as the probability that the simultaneous confidence intervals are *correct* and *useful*.

The parameter ω represents the maximum width of any of the individual confidence intervals in the set of simultaneous confidence intervals. Thus, ω is used to specify the narrowness of the confidence intervals.

Multiple Comparisons with a Control (MCC)

A common experimental design compares one or more *treatment* groups with a *control* group. The control group may receive a placebo, the standard treatment, or even an experimental treatment. The distinguishing feature is that the mean response of each of the other groups is to be compared with this control group.

We arbitrarily assume that the last group (group *k*), is the control group. The *k*-1 parameters of primary interest are

$$\begin{aligned}\delta_1 &= \mu_1 - \mu_k\\ \delta_2 &= \mu_2 - \mu_k\\ &\vdots\\ \delta_i &= \mu_i - \mu_k\\ &\vdots\\ \delta_{k-1} &= \mu_{k-1} - \mu_k\end{aligned}$$

In this situation, Dunnett's method provides simultaneous, one- or two-sided confidence intervals for all of these parameters.

The one-sided confidence intervals, $\delta_1,\cdots,\delta_{k-1}$, are specified as follows:

$$\Pr\left(\delta_i > \hat{\mu}_i - \hat{\mu}_k - q_{\alpha,df,\lambda_1,\cdots,\lambda_{k-1}}\hat{\sigma}\sqrt{\frac{1}{n_i}+\frac{1}{n_k}}\ \ for\ i=1,\cdots,k-1\right)=1-\alpha$$

where

$$\lambda_i = \sqrt{\frac{n_i}{n_i + n_k}}$$

and q, found by numerical integration, is the solution to

$$\int_0^\infty \int_{-\infty}^\infty \prod_{i=1}^{k-1} \left[\Phi\left(\frac{\lambda_i z + qs}{\sqrt{1-\lambda^2}} \right) \right] d\Phi(z)\gamma(s)ds = 1-\alpha$$

where $\Phi(z)$ is the standard normal distribution function and $\gamma(z)$ is the density of $\hat{\sigma}/\sigma$.

The two-sided confidence intervals for $\delta_1, \cdots, \delta_{k-1}$ are specified as follows:

$$\Pr\left(\delta_i \in \hat{\mu}_i - \hat{\mu}_k - \left| q_{\alpha, df, \lambda_1, \cdots, \lambda_{k-1}} \right| \hat{\sigma} \sqrt{\frac{1}{n_i} + \frac{1}{n_k}} \;\; for\ i = 1, \cdots, k-1 \right) = 1-\alpha$$

where

$$\lambda_i = \sqrt{\frac{n_i}{n_i + n_k}}$$

and $|q|$, found by numerical integration, is the solution to

$$\int_0^\infty \int_{-\infty}^\infty \prod_{i=1}^{k-1} \left[\Phi\left(\frac{\lambda_i z + |q|s}{\sqrt{1-\lambda^2}} \right) - \Phi\left(\frac{\lambda_i z - |q|s}{\sqrt{1-\lambda^2}} \right) \right] d\Phi(z)\gamma(s)ds = 1-\alpha$$

where $\Phi(z)$ is the standard normal distribution function and $\gamma(z)$ is the density of $\hat{\sigma}/\sigma$.

Interpretation of Dunnett's Simultaneous Confidence Intervals

There is a specific interpretation given for Dunnett's method. It provides a set of confidence intervals calculated so that, if the normality and equal-variance assumptions are valid, the probability that all of the k-1 confidence intervals enclose the true values of $\delta_1, \cdots, \delta_{k-1}$ is $1-\alpha$. The presentation below is for the two-sided case. The one-sided case is the same as for the MCB case.

Sample Size and Power - Balanced Case

Using the modified definition of power, the two-sided case is outlined as follows.

$$\begin{aligned}
&\Pr\left[(\text{simultaneous coverage}) \text{ and } (\text{narrow}) \right] \\
&= \Pr\left[\left(\mu_i - \mu_j \in \hat{\mu}_i - \hat{\mu}_j \pm |q| \hat{\sigma} \sqrt{2/n} \text{ for } i = 1, \cdots, k-1 \right) \text{ and } \left(|q| \hat{\sigma} \sqrt{2/n} < \omega / 2 \right) \right] \\
&= k \int_0^u \int_{-\infty}^\infty \left[\Phi\left(z + \sqrt{2}|q|s \right) - \Phi\left(z - \sqrt{2}|q|s \right) \right]^{k-1} d\Phi(z)\gamma(s)ds \\
&\geq 1-\beta
\end{aligned}$$

where

$$u = \frac{\omega / 2}{\sigma|q|\sqrt{\frac{2}{n}}}$$

This calculation is made using the algorithm developed by Hsu (1996).

To reiterate, this calculation requires you to specify the minimum value of $\omega = \mu_i - \mu_j$ that you want to detect, the group sample size, *n*, the power, $1-\beta$, the significance level, α, and the within-group standard deviation, σ.

Sample Size and Power - Unbalanced Case

Using the modified definition of power, the unbalanced case is outlined as follows.

$$\begin{aligned}
&\Pr\left[\left(\text{simultaneous coverage}\right) \text{ and } \left(\text{narrow}\right)\right] \\
&= \Pr\left[\left(\mu_i - \mu_k \in \hat{\mu}_i - \hat{\mu}_k \pm |q|\hat{\sigma}\sqrt{\frac{1}{n_i}+\frac{1}{n_k}} \text{ for } i = 1,...,k-1\right) \text{ and } \left(\min_{i<k}\left[|q|\hat{\sigma}\sqrt{\frac{1}{n_i}+\frac{1}{n_k}}\right] < \omega / 2\right)\right] \\
&= k\int_0^u \int_{-\infty}^{\infty}\left[\Phi(z) - \Phi\left(z - \sqrt{2}|q|s\right)\right]^{k-1} d\Phi(z)\gamma(s)ds \\
&\geq 1-\beta
\end{aligned}$$

where

$$u = \frac{\omega / 2}{\min_{i<k}\left[\sigma|q|\sqrt{\frac{1}{n_i}+\frac{1}{n_k}}\right]}$$

This calculation is made using the algorithm developed by Hsu (1996).

To reiterate, this calculation requires you to specify the minimum value of $\omega = \mu_i - \mu_j$ that you want to detect, the group sample sizes, $n_1, n_2, \cdots, n_k$, the power, $1-\beta$, the significance level, α, and the within-group standard deviation, σ.

Multiple Comparisons with the Best (MCB)

The method of multiple comparisons with the best (champion) is used in situations in which the best group (we will assume the best is the largest, but it could just as well be the smallest) is desired. Because of sampling variation, the group with the largest sample mean may not actually be the group with the largest population mean. The following methodology has been developed to analyze data in this situation.

Perhaps the most obvious way to define the parameters in this situation is as follows

$$\max_{j=1,\cdots,k} \mu_j - \mu_i, \ \mathit{for}\ i = 1,\cdots,k$$

Obviously, the group for which all of these values are positive will correspond to the group with the largest mean.

Another way of looking at this, which has some advantages, is to use the parameters

$$\theta_i = \mu_i - \max_{i \neq j} \mu_j, \ \mathit{for}\ i = 1,\cdots,k$$

since, if $\theta_i > 0$, group *i* is the best.

Hsu (1996) recommends using constrained MCB inference in which the intervals are constrained to include zero. Hsu recommends this because inferences about which group is best are sharper. For example, a confidence interval for θ_i whose lower limit is 0 indicates that group *i* is the best. Similarly, a confidence interval for θ_i whose upper limit is 0 indicates that group *i* is not the best.

Hsu (1996) shows that $100(1-\alpha)\%$ simultaneous confidence intervals for θ_i are given by

$$-\min\left[0,\left(\hat{\theta}_i - q^i_{\alpha,df,\lambda_1,\cdots,\lambda_{k-1}}\hat{\sigma}\sqrt{\frac{1}{n_i}+\frac{1}{n_k}}\right)\right],\ \max\left[0,\left(\hat{\theta}_i + q^i_{\alpha,df,\lambda_1,\cdots,\lambda_{k-1}}\hat{\sigma}\sqrt{\frac{1}{n_i}+\frac{1}{n_k}}\right)\right],\ i=1,\cdots,k$$

where q^i is found using Dunnett's one-sided procedure discussed above assuming that group *i* is the control group.

Sample Size and Power - Balanced Case

Using the modified definition of power, the balanced case is outlined as follows

$$\begin{aligned}
&\Pr\left[(\text{simultaneous coverage}) \text{ and } (\text{narrow})\right] \\
&= \Pr\left[\begin{array}{l}
\left(\begin{array}{l}
-\min\left(0, \hat{\mu}_i - \max\limits_{j\neq i}\left(\hat{\mu}_j\right) - q\hat{\sigma}\sqrt{2/n}\right) \le \\
\mu_i - \max\limits_{j\neq i}\left(\mu_j\right) \le \\
\max\left(0, \hat{\mu}_i - \max\limits_{j\neq i}\left(\hat{\mu}_j\right) + q\hat{\sigma}\sqrt{2/n}\right) \text{ for } i=1,\cdots,k
\end{array}\right) \\
\text{and } \left(q\hat{\sigma}\sqrt{2/n} < \omega/2\right)
\end{array}\right] \\
&= k\int_0^u \int_{-\infty}^{\infty} \left[\Phi\left(z+\sqrt{2}qs\right)\right]^{k-1} d\Phi(z)\gamma(s)ds \\
&\ge 1-\beta
\end{aligned}$$

where

$$u = \frac{\omega/2}{\sigma q\sqrt{\frac{2}{n}}}$$

This calculation is made using the algorithm developed by Hsu (1996).

To reiterate, this calculation requires you to specify the minimum value of $\omega = \mu_i - \mu_j$ that you want to detect, the group sample size, *n*, the power, $1-\beta$, the significance level, α, and the within-group standard deviation, σ.

Sample Size and Power - Unbalanced Case

Using the modified definition of power, the unbalanced case is outlined as follows

$$\Pr\left[(\text{simultaneous coverage}) \text{ and } (\text{narrow})\right]$$
$$= \Pr\left[\begin{array}{l}\left(\begin{array}{l}-\min\left(0, \hat{\mu}_i - \max_{j\neq i}\left(\hat{\mu}_j\right) - q^i\hat{\sigma}\sqrt{\frac{1}{n_i}+\frac{1}{n_j}}\right) \leq \\ \mu_i - \max_{j\neq i}\left(\mu_j\right) \leq \\ \max\left(0, \hat{\mu}_i - \max_{j\neq i}\left(\hat{\mu}_j\right) + q^i\hat{\sigma}\sqrt{\frac{1}{n_i}+\frac{1}{n_j}}\right) \text{ for } i = 1,\cdots,k\end{array}\right) \\ \text{and } \left(\min_{j\neq i}\left[q^i\hat{\sigma}\sqrt{\frac{1}{n_i}+\frac{1}{n_j}}\right] < \omega/2\right)\end{array}\right]$$
$$= k\int_0^u \int_{-\infty}^{\infty}\left[\Phi\left(z+\sqrt{2}|q|s\right)\right]^{k-1} d\Phi(z)\gamma(s)ds$$
$$\geq 1-\beta$$

where

$$u = \max_{i\neq j}\left(\frac{\omega/2}{\sigma|q^i|\sqrt{\frac{1}{n_i}+\frac{1}{n_j}}}\right)$$

This calculation is made using the algorithm developed by Hsu (1996).

To reiterate, this calculation requires you to specify the minimum value of $\omega = \mu_i - \mu_j$ that you want to detect, the group sample sizes, $n_1, n_2, \cdots, n_k$, the power, $1-\beta$, the significance level, α, and the within-group standard deviation, σ.

All-Pairwise Comparisons (MCA)

In this case you are interested in all possible pairwise comparisons of the group means. There are $k(k$-1)/2 such comparisons. A popular method in this case is that developed by Tukey.

Balanced Case

The Tukey method provides simultaneous, two-sided confidence intervals. They are specified as follows

$$\Pr\left(\mu_i - \mu_j \in \hat{\mu}_i - \hat{\mu}_k \pm |q^*|\hat{\sigma}\sqrt{\frac{2}{n}} \;\; for\; i \neq j\right) = 1 - \alpha$$

When all the sample sizes are equal, q^* may be found by numerical integration as the solution to the equation

$$\int_0^\infty \int_{-\infty}^\infty \prod_{i=1}^{k-1} \left[\Phi(z) - \Phi\left(z - \sqrt{2}|q^*|s\right)\right] d\Phi(z)\gamma(s)ds = 1 - \alpha$$

where $\Phi(z)$ is the standard normal distribution function and $\gamma(z)$ is the density of $\hat{\sigma}/\sigma$. Note that $q' = \sqrt{2}|q^*|$ is the critical value of the Studentized range distribution.

Sample Size and Power

Using the modified definition of power, the balanced case is outlined as follows

$$\begin{aligned}
&\Pr\left[(\text{simultaneous coverage}) \text{ and } (\text{narrow})\right] \\
&= \Pr\left[\left(\mu_i - \mu_j \in \hat{\mu}_i - \hat{\mu}_j \pm |q^*|\hat{\sigma}\sqrt{2/n} \text{ for all } i \neq j\right) \text{and} \left(|q^*|\hat{\sigma}\sqrt{2/n} < \omega/2\right)\right] \\
&= k\int_0^u \int_{-\infty}^\infty \left[\Phi(z) - \Phi\left(z - \sqrt{2}|q^*|s\right)\right]^{k-1} d\Phi(z)\gamma(s)ds \\
&\geq 1 - \beta
\end{aligned}$$

where

$$u = \frac{\omega/2}{\sigma|q^*|\sqrt{\frac{2}{n}}}$$

This calculation is made using the algorithm developed by Hsu (1996).

To reiterate, this calculation requires you to specify the minimum value of $\omega = \mu_i - \mu_j$ that you want to detect, the group sample size, n, the power, $1-\beta$, the significance level, α, and the within-group standard deviation, σ.

Unbalanced Case

The simultaneous, two-sided confidence intervals are specified as follows

$$\Pr\left(\mu_i - \mu_j \in \hat{\mu}_i - \hat{\mu}_k \pm \left|q^e\right| \hat{\sigma} \sqrt{\frac{1}{n_i} + \frac{1}{n_k}} \;\; for\ i \neq j \right) = 1 - \alpha$$

Unfortunately, $\left|q^e\right|$ cannot be calculated as a double integral as in previous cases. Instead, the Tukey-Kramer approximate solution is used. Their proposal is to use $|q*|$ in place of $\left|q^e\right|$.

Sample Size and Power

Using the modified definition of power, the unbalanced case is outlined as follows

$$\begin{aligned}
&\Pr\left[(\text{simultaneous coverage}) \text{ and } (\text{narrow})\right] \\
&= \Pr\left[\left(\mu_i - \mu_j \in \hat{\mu}_i - \hat{\mu}_j \pm |q*| \hat{\sigma} \sqrt{\frac{1}{n_i} + \frac{1}{n_j}} \text{ for all } i \neq j \right) \text{ and } \left(\min_{i \neq j} \left[|q*| \hat{\sigma} \sqrt{\frac{1}{n_i} + \frac{1}{n_j}} \right] < \omega / 2 \right) \right] \\
&= k \int_0^u \int_{-\infty}^{\infty} \left[\Phi(z) - \Phi\left(z - \sqrt{2}|q*|s\right) \right]^{k-1} d\Phi(z) \gamma(s) ds \\
&\geq 1 - \beta
\end{aligned}$$

where

$$u = \frac{\omega / 2}{\min\limits_{i \neq j} \left[\sigma |q*| \sqrt{\frac{1}{n_i} + \frac{1}{n_j}} \right]}$$

This calculation is made using the algorithm developed by Hsu (1996).

To reiterate, this calculation requires you to specify the minimum value of $\omega = \mu_i - \mu_j$ that you want to detect, the group sample sizes, $n_1, n_2, \cdots, n_k$, the power, $1 - \beta$, the significance level, α, and the within-group standard deviation, σ.

Procedure Options

This section describes the options that are unique to this procedure. These are located on the panels associated with the Data, Options, and Reports tabs. To find out more about using the other tabs such as Plot Text, Axes, and Template, refer to the chapter entitled Procedure Templates.

Data Tab

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

Find

This option specifies the parameter to be solved for using the other parameters. Under most situations you will select either Beta for a power analysis or *n* for sample size determination.

k (Number of Groups)

This is the number of group means being compared. It must be greater than or equal to three.

You can enter a list of values, in which case a separate analysis will be calculated for each value. Commas or blanks may separate the numbers. A TO-BY list may also be used.

Note that the number of items used in the Hypothesized Means box and the Group Sample Size Pattern box is controlled by this number.

Examples:

3,4,5

4 5 6

3 to 11 by 2

Minimum Detectable Difference

Specify one or more values of the minimum detectable difference. This is the smallest difference between any two group means that is to be detectable by the experiment. Note that this is a positive amount. This value is set by the researcher to represent the smallest difference between two means that will be of practical significance to other researchers. Note that in the case of Dunnett’s test, this is the maximum difference between each treatment and the control.

Examples:

3

3 4 5

10 to 20 by 2

Type of Multiple Comparison

Specify the type of multiple comparison test to be analyzed. The tests available are

Tukey-Kramer: All possible paired comparisons.

Hsu Best: Constrained comparisons with the best.

Dunnett: Two-sided versus a control group.

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. If the Group Sample Size Pattern is *1,2,3*, these sets would be

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,

and (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

A set of positive, numeric values (one for each group) is entered here. The sample size of group *i* is found by multiplying the i^{th} number from this list times the value of *n* and rounding up to the next whole number. The number of values must match the number of groups, *k*. When too few numbers are entered, 1's are added. When too many numbers are entered, the extras are ignored.

Note that in the case of Dunnett's test, the last group corresponds to the control group. Thus, if you want to study the implications of having a group with a different sample size for the control group, you should specify that sample size in the last position

Equal

If all sample sizes are to be equal, enter "Equal" here and the desired sample size in n. A set of *k* 1's will be used. This will result in $N1 = N2 = N3 = n$. That is, all sample sizes are equal to *n*.

Alpha

This option specifies one or more values for the probability of a type-I error (alpha). A type-I error occurs when you reject the null hypothesis of equal means when in fact the means are equal.

Values must be between zero and one. However, since alpha must be less than or equal to beta, the upper limit is beta, not one.

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

Beta

This option specifies one or more values for beta. In the case of multiple comparisons, power is the probability that the confidence intervals will cover the true parameter values and be sufficiently narrow to be useful. This is a modified definition of power; since beta equals 1-power, it is has a modified definition as well.

Values must be between zero and one. However, since alpha must be less than or equal to beta, the lower limit is alpha, not zero.

Historically, the value of 0.20 was often used for beta. However, you should pick a value for beta that represents the risk you are willing to take.

Power is defined as 1- beta. Specifying the beta error level also specifies the power level. For example, if you specify beta values of 0.05, 0.10, and 0.20, you are specifying the corresponding power values of 0.95, 0.90, and 0.80, respectively.

S - Std Dev of Subjects

This option refers to σ, the standard deviation within a group. It represents the variability from subject to subject that occurs when the subjects are treated identically. It is assumed to be the same for all groups. This value is approximated in an analysis of variance table by the square root of the mean square error.

Since they are positive square roots, the numbers must be strictly greater than zero. You can press the *SD* button to obtain further help on estimating the standard deviation.

Note that if you are using this procedure to test a factor (such as an interaction) from a more complex design, the value of standard deviation is estimated by the square root of the mean square of the term that is used as the denominator in the *F* test.

You can enter a list of values separated by blanks or commas, in which case a separate analysis will be calculated for each value.

Examples of valid entries:

1,4,7,10

1 4 7 10

1 to 10 by 3

Options Tab

The Options tab contains parameters that control the progress and termination of the iterative procedures.

Maximum Iterations

Specify the maximum number of iterations before the search for the criterion of interest is aborted. When the maximum number of iterations is reached without convergence, the criterion is left blank.

Search Precision

Specify the precision to be used in the routines that search for the value of the minimum detectable difference. Note that the closer this value is to zero, the longer a search will take.

Example1 - Calculating Power

An experiment is being designed to compare the means of four groups using the Tukey-Kramer pairwise multiple comparison test with a significance level of 0.05. Previous studies indicate that the standard deviation is 5.3. The typical mean response level is 63.4. The researcher believes that a 25% increase in the mean will be of interest to others. Since 0.25(63.4) = 15.85, this is the number that will be used as the minimum detectable difference.

To better understand the relationship between power and sample size, the researcher wants to compute the power for several group sample sizes between 2 and 14. The sample sizes will be equal across all groups.

Setup

This section presents the values of each of the parameters needed to run this example. You can make these changes directly on your screen or you can load the template 'Example1' by clicking the Template tab and loading this template.

Option	Value
Data Tab	
Find	**Beta and Power**
k (Number of Groups)	**4**
Minimum Detectable Difference	**15.85**
Type of Multiple Comparison	**All Pairs - Tukey Kramer**
n (Sample Size Multiplier)	**2 to 14 by 2**
Group Sample Size Pattern	**Equal**
Alpha	**0.05**
Beta	*Ignored since this is the Find setting*
S (Std Dev of Subjects)	**5.3**

Annotated Output

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

Numeric Results Report

Numeric Results for Multiple Comparison Test: Tukey-Kramer (Pairwise)

Power	Average Size (n)	k	Total N	Alpha	Beta	Minimum Detectable Difference	Standard Deviation (S)	Diff / S
0.0113	2.00	4	8	0.0500	0.9887	15.85	5.30	2.9906
0.0666	4.00	4	16	0.0500	0.9334	15.85	5.30	2.9906
0.3171	6.00	4	24	0.0500	0.6829	15.85	5.30	2.9906
0.7371	8.00	4	32	0.0500	0.2629	15.85	5.30	2.9906
0.9301	10.00	4	40	0.0500	0.0699	15.85	5.30	2.9906
0.9497	12.00	4	48	0.0500	0.0503	15.85	5.30	2.9906
0.9500	14.00	4	56	0.0500	0.0500	15.85	5.30	2.9906

Report Definitions
Power is the probability of rejecting a false null hypothesis. It should be close to one.
n is the average group sample size.
k is the number of groups.
Total N is the total sample size of all groups.
Alpha is the probability of rejecting a true null hypothesis. It should be small.
Beta is the probability of accepting a false null hypothesis. It should be small.
The Minimum Detectable Difference between any two group means.
S is the within group standard deviation.
The Diff / S is the ratio of Min. Detect. Diff. to standard deviation.

Summary Statements
In a single factor ANOVA study, sample sizes of 2, 2, 2, and 2 are obtained from the 4 groups whose means are to be compared. The total sample of 8 subjects achieves 1% power to detect a difference of at least 15.85 using the Tukey-Kramer (Pairwise) multiple comparison test at a 0.0500 significance level. The common standard deviation within a group is assumed to be 5.30.

This report shows the numeric results of this power study. Following are the definitions of the columns of the report.

Power

This is the probability that the confidence intervals will cover the true parameter values and be sufficiently narrow to be useful.

Average n

The average of the group sample sizes.

k

The number of groups.

Total N

The total sample size of the study.

Alpha

The probability of rejecting a true null hypothesis. This is often called the significance level.

Beta

Equal to 1- power. Note the definition of power above.

Minimum Detectable Difference

This is the value of the minimum detectable difference. This is the minimum difference between two means that is thought to be of practical importance. Note that in the case of Dunnett's test, this is the minimum difference between a treatment mean and the control mean that is of practical importance.

Standard Deviation (S)

This is the within-group standard deviation. It was set in the Data window.

Diff / S

This is an index of relative difference between the means standardized by dividing by the standard deviation. This value can be used to make comparisons among studies.

Detail Results Report

Tukey-Kramer Test Details

Group	n	Percent n of Total N	Alpha	Power	Minimum Detectable Difference	Standard Deviation
1	2	25.00	0.0500	0.0113	15.85	5.30
2	2	25.00				
3	2	25.00				
4	2	25.00				
Total	8	100.00				

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

Group

The group identification number is shown on each line. The second to the last line represents the last group. When Dunnett's test has been selected, this line represents the control group.

The last line, labeled *Total*, gives the total for all the groups.

n

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

Percent n of Total N

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

Alpha

The probability of rejecting a true null hypothesis. This is often called the significance level.

Power

This is the probability that the confidence intervals will cover the true parameter values and be sufficiently narrow to be useful.

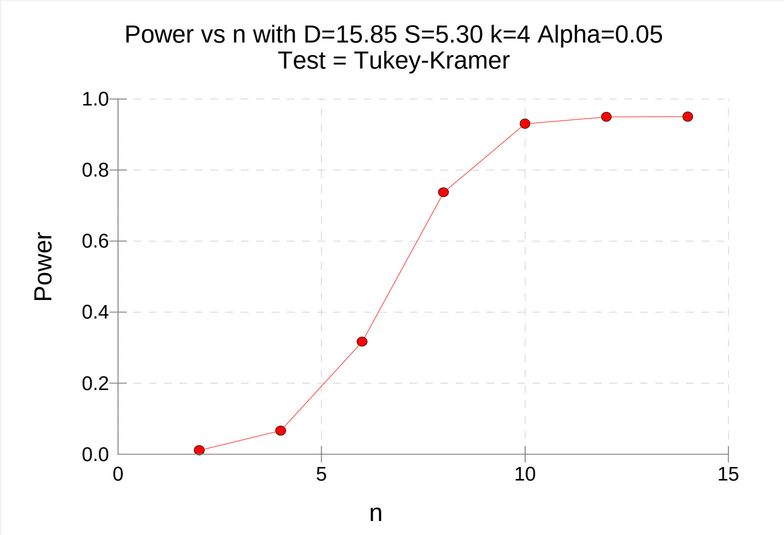
Minimum Detectable Difference

This is the value of the minimum detectable difference. This is the minimum difference between two means that is thought to be of practical importance. Note that in the case of Dunnett's test, this is the minimum difference between a treatment mean and the control mean that is of practical importance.

Standard Deviation (S)

This is the within-group standard deviation. It was set in the Data window.

Plot Section



This plot gives a visual presentation to the results in the Numeric Report. We can see the impact on the power of increasing the sample size.

When you create one of these plots, it is important to use trial and error to find an appropriate range for the horizontal variable so that you have results with both low and high power.

Example 2 - Power after Dunnett's Test

This example covers the situation in which you are calculating the power of Dunnett's test on data that have already been collected and analyzed.

An experiment included a control group and two treatment groups. Each group had seven individuals. A single response was measured for each individual and recorded in the following table.

Control	T1	T2
554	774	786
447	465	536
356	759	653
452	646	685
674	547	658
654	665	669
558	767	557

When analyzed using the one-way analysis of variance procedure in ***NCSS***, the following results were obtained.

Analysis of Variance Table

Source Term	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level
A (...)	2	75629.8	37814.9	3.28	0.061
S(A)	18	207743.4	11541.3		
Total (Adjusted)	20	283373.3			
Total	21				

Dunnett's Simultaneous Confidence Intervals for Treatment vs. Control
Response: Control,T1,T2
Term A:
Control Group: Control
Alpha=0.050 Error Term=S(A) DF=18 MSE=11541.3 Critical Value=2.3987

Treatment Group	Count	Mean	Lower 95.0% Simult.C.I.	Difference With Control	Upper 95.0% Simult.C.I.	Test Result
T1	7	660.43	-5.17	132.57	270.31	
T2	7	649.14	-16.46	121.29	259.03	

The significance level (Prob Level) was only 0.061—not enough for statistical significance. Since the lower confidence limits are negative and the upper confidence limits are positive, Dunnett's two-sided test did not find a significant difference between either treatment and the control group.

The researcher had hoped to show that the treatment groups had higher response levels than the control group. He could see that the group means followed this pattern since the mean for *T1* was about 25% higher than the control mean and the mean for *T2* was about 23% higher than the control mean. He decided to calculate the power of the experiment.

Setup

The data entry for this problem is simple. The only entry that is not straight forward is finding an appropriate value for the standard deviation. Since the standard deviation is estimated by the square root of the mean square error, it is calculated as $\sqrt{11541.3} = 107.4304$.

You can enter these values yourself or load the Example2 template from the Template tab.

<u>**Option**</u>	<u>**Value**</u>
Data Tab	
Find	**Beta and Power**
k (Number of Groups)	**3**
Minimum Detectable Difference	**133**
Type of Multiple Comparison	**With Control - Dunnett**
n (Sample Size Multiplier)	**7**
Group Sample Size Pattern	**Equal**
Alpha	**0.05**
Beta	*Ignored since this is the Find setting*
S (Std Dev of Subjects)	**107.4304**

Annotated Output

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

Numeric Results

Power	Average Size (n)	k	Total N	Alpha	Beta	Minimum Detectable Difference	Standard Deviation (S)	Diff / S
0.0002	7.00	3	21	0.0500	0.9998	133.00	107.43	1.2380

The power is only 0.0002. Hence, there was little chance of detecting a difference of 133 between a treatment and a control group.

It was of interest to the researcher to determine how large of a sample was needed if the power was to be 0.90. Setting Beta equal to 0.90 and ‘Find/Solve For’ to *n* resulted in the following report.

Power	Average Size (n)	k	Total N	Alpha	Beta	Minimum Detectable Difference	Standard Deviation (S)	Diff / S
0.9042	33.00	3	99	0.0500	0.0958	133.00	107.43	1.2380

We see that instead of 7 per group, 33 per group were needed.

It was also of interest to the research to determine how large of a difference between the means could have been detected. Setting ‘Find’ to Min Detectable Difference resulted in the following report.

Power	Average Size (n)	k	Total N	Alpha	Beta	Minimum Detectable Difference	Standard Deviation (S)	Diff / S
0.9000	7.00	3	21	0.0500	0.1000	348.81	107.43	3.2468

We see that a study of this size with these parameters could only detect a difference of 348.8. This explains why the results were not significant.

Example3 - Using Unequal Sample Sizes

It is usually advisable to design experiments with equal sample sizes in each group. In some cases, however, it may be necessary to allocate subjects unequally across the groups. This may occur when the group variances are unequal, the costs per subject are different, or the dropout rates are different. This module can be used to study the power of unbalanced experiments.

In this example which will use Dunnett's test, the minimum detectable difference is 2.0, the standard deviation is 1.0, alpha is 0.05, and *k* is 3. The sample sizes are 7, 7, and 14.

Setup

This section presents the values of each of the parameters needed to run this example. You can make these changes directly on your screen or you can load the template Example3a by clicking the Template tab and loading this template.

Option	Value
Data Tab	
Find	**Beta and Power**
k (Number of Groups)	**3**
Minimum Detectable Difference	**2**
Type of Multiple Comparison	**With Control - Dunnett**
n (Sample Size Multiplier)	**1**
Group Sample Size Pattern	**7 7 14**
Alpha	**0.05**
Beta	*Ignored since this is the Find setting*
S (Std Dev of Subjects)	**1**

Annotated Output

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

Numeric Results

Power	Average Size (n)	k	Total N	Alpha	Beta	Minimum Detectable Difference	Standard Deviation (S)	Diff / S
0.2726	9.33	3	28	0.0500	0.7274	2.00	1.00	2.0000

Alternatively, this problem could have been set up as follows (Example3b):

n (Sample Size Multiplier)	**7**
Group Sample Size Pattern	**1 1 2**

The advantage of this method is that you can try several values of *n* while keeping the same allocation ratios.

Example4 - Validation using Hsu

Hsu (1996) page 241 presents an example of determining the sample size in an experiment with 8 groups. The minimum detectable difference is 10,000 psi. The standard deviation is 3,000 psi. Alpha is 0.05 and beta is 0.10. He finds a sample size of 10 per group for the Tukey-Kramer test, a sample size of 8 for Hsu's test, and a sample size of 8 for Dunnett's test.

Setup

This section presents the values of each of the parameters needed to run this example. You can enter these values yourself or load and run each of the Example4 templates from the Template tab.

Option	Value
Data Tab	
Find	**n**
k (Number of Groups)	**8**
Minimum Detectable Difference	**10000**
Type of Multiple Comparison	**All Pairs - Tukey Kramer**
n (Sample Size Multiplier)	*Ignored since this is the Find setting*
Group Sample Size Pattern	**Equal**
Alpha	**0.05**
Beta	**0.1**
S (Std Dev of Subjects)	**3000**

Annotated Output

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

Numeric Results for Tukey-Kramer Test

Power	Average Size (n)	k	Total N	Alpha	Beta	Minimum Detectable Difference	Standard Deviation (S)	Diff / S
0.9397	10.00	8	80	0.0500	0.0603	10000.00	3000.00	3.3333

PASS also found *n* = 10.

Numeric Results for Hsu's Test

Power	Average Size (n)	k	Total N	Alpha	Beta	Minimum Detectable Difference	Standard Deviation (S)	Diff / S
0.9087	6.00	8	48	0.0500	0.0913	10000.00	3000.00	3.3333

PASS also found *n* = 6.

Numeric Results for Dunnett's Test

Power	Average Size (n)	k	Total N	Alpha	Beta	Minimum Detectable Difference	Standard Deviation (S)	Diff / S
0.9434	8.00	8	64	0.0500	0.0566	10000.00	3000.00	3.3333

PASS found *n* = 8.

Example 5 - Validation using Pan and Kupper

Pan and Kupper (1999, page 1481) present examples of determining the sample size using alternative methods. It is interesting to compare the method of Hsu (1996) with theirs. Although the results are not exactly the same, they are very close.

In the example of Pan and Kupper, the minimum detectable difference is 0.50. The standard deviation is 0.50. Alpha is 0.05, and beta is 0.10. They find a sample size of 51 per group for Dunnett's test and a sample size of 60 for the Tukey-Kramer test.

Setup

This section presents the values of each of the parameters needed to run this example. You can enter these values yourself or load and run each of the Example5 templates from the Template tab.

Option	Value
Data Tab	
Find	**n**
k (Number of Groups)	**4**
Minimum Detectable Difference	**0.50**
Type of Multiple Comparison	**With Control - Dunnett**
n (Sample Size Multiplier)	*Ignored since this is the Find setting*
Group Sample Size Pattern	**Equal**
Alpha	**0.05**
Beta	**0.1**
S (Std Dev of Subjects)	**0.50**

Annotated Output

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

Numeric Results for Dunnett's Test

Power	Average Size (n)	k	Total N	Alpha	Beta	Minimum Detectable Difference	Standard Deviation (S)	Diff / S
0.9147	53.00	4	212	0.0500	0.0853	0.50	0.50	1.0000

PASS found *n* = 53. This is very close to the 51 that Pan and Kupper found using a slightly different method.

Numeric Results for Tukey-Kramer Test

Power	Average Size (n)	k	Total N	Alpha	Beta	Minimum Detectable Difference	Standard Deviation (S)	Diff / S
0.9057	62.00	4	248	0.0500	0.0943	0.50	0.50	1.0000

PASS also found *n* = 62. This is very close to the 60 that Pan and Kupper found using a slightly different method.

Chapter 580

Pair-Wise Multiple Comparisons using Simulation

Introduction

This procedure uses simulation analyze the power and significance level of three pair-wise multiple-comparison procedures: Tukey-Kramer, Kruskal-Wallis, and Games-Howell. For each scenario, two simulations are run: one estimates the significance level and the other estimates the power.

The term *multiple comparisons* refers to a set of two or more statistical hypothesis tests. The term *pair-wise multiple comparisons* refers to the set of all pairs of means that can be generated among the means of *k* groups. For example, suppose the levels of a factor with five groups are labeled A, B, C, D, and E. The ten possible paired-comparisons that could be made among the five groups are A-B, A-C, A-D, A-E, B-C, B-D, B-E, C-D, C-E, and D-E.

As the number of groups increases, the number of comparisons (pairs) increases dramatically. For example, a 5 group design has 10 pairs, a 10 group design has 45 pairs, and a 20 group design has 190 pairs. When several comparisons are made among the group means, the determination of the significance level of each individual comparison is much 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 $\mu_1, \mu_2, \cdots, \mu_k$ and unknown variance σ^2. Let $n_1, n_2, \cdots, n_k$ denote the number of subjects in each group.

The 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.

The Tukey-Kramer procedure, the Kruskal-Wallis procedure, and the Games-Howell procedure are the pair-wise multiple-comparison procedures that have been developed for this situation. The calculation of each of these tests is given next.

Tukey-Kramer

This test is referenced in Kirk (1982). It uses the critical values from the studentized-range distribution. For each pair of groups, the significance test between any two groups *i* and *j* is calculated by rejecting the null hypothesis of mean equality if

$$\frac{\left| \overline{Y}_i - \overline{Y}_j \right|}{\sqrt{\frac{\hat{\sigma}^2}{2} \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}} \geq q_{\alpha_f, k, v}$$

where

$$\begin{aligned} v &= \sum_{i=1}^{k} n_i - k \\ &= N - k \end{aligned}$$

Kruskal-Wallis

This test is attributed to Dunn (1964) and is referenced in Gibbons (1976). It is a nonparametric, or distribution-free, test for which the assumption of normality is not necessary. It tests whether pairs of medians are equal using a rank test. Sample sizes of at least five (but preferably larger) for each treatment are recommended for 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 $\alpha = \alpha_f \ / \left(k\left(k-1\right)\right)$. For each pair of groups, *i* and *j*, the null hypothesis of equality is rejected if

$$\frac{\left|\overline{R}_i - \overline{R}_j\right|}{\sqrt{\frac{n(n+1)}{12}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}} \geq z_\alpha$$

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

Games-Howell

This test is referenced in Kirk (1982) page 120. It was developed for the case when the individual group variances cannot be assumed to be equal. It also uses critical values from the studentized-range distribution. For each pair of groups, *i* and *j*, the null hypothesis of equality is rejected if

$$\frac{\left|\overline{Y}_i - \overline{Y}_j\right|}{\sqrt{\frac{1}{2}\left(\frac{\hat{\sigma}_i^2}{n_i} + \frac{\hat{\sigma}_j^2}{n_j}\right)}} \geq q_{\alpha_f,k,v'}$$

where

$$v' = \frac{\left(\frac{\hat{\sigma}_i^2}{n_i} + \frac{\hat{\sigma}_j^2}{n_j}\right)^2}{\frac{\hat{\sigma}_i^4}{n_i^2\left(n_i - 1\right)} + \frac{\hat{\sigma}_j^4}{n_j^2\left(n_j - 1\right)}}$$

If any of the following conditions hold, then $v' = n_i + n_2 - 2$:

1. $9/10 \leq n_i \ / \ n_j \leq 10/9$

2. $9/10 \leq \left(\frac{\hat{\sigma}_i^2}{n_i}\right) / \left(\frac{\hat{\sigma}_j^2}{n_j}\right) \leq 10/9$

3. $4/5 \leq n_i \ / \ n_j \leq 5/4$ and $1/2 \leq \left(\frac{\hat{\sigma}_i^2}{n_i}\right) / \left(\frac{\hat{\sigma}_j^2}{n_j}\right) \leq 2$

4. $2/3 \leq n_i \ / \ n_j \leq 3/2$ and $3/4 \leq \left(\frac{\hat{\sigma}_i^2}{n_i}\right) / \left(\frac{\hat{\sigma}_j^2}{n_j}\right) \leq 4/3$

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 three groups labeled, A, B, and C. There are three paired comparisons in this experiment: A-B, A-C, and B-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.

Procedure Options

This section describes the options that are unique to this procedure. These are located on the panels associated with the Data, Options, and Reports tabs. To find out more about using the other tabs such as Plot Text, Axes, and Template, turn to the chapter entitled Procedure Templates.

Data Tab

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

Find

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-pair power or all-pairs power. The value of the option *Power* will then represent this type of power.

Any-pair power is the probability of detecting at least one pair from among those that are actually different. *All-pairs power* is the probability of detecting all pairs that are actually different.

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.

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.

FWER (Alpha)

This option specifies one or more values of the *family-wise error rate* (FWER) which is the analog of alpha for multiple comparisons. 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.

Power

This option is only used when *Find (Solve For)* is set to *n (All-Pairs)* or *n (Any-Pair)*.

Power is defined differently with multiple comparisons. Although many definitions are possible, two are adopted here. *Any-pair power* is the probability of detecting at least one pair of the means that are different. *All-pairs power* is the probability of detecting all pairs of means that are truly different. As the number of groups 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.

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, …, mg and this value is represented by *n*, the group sample sizes

n1, *n2*, *n3*, ...,*ng* 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. This is required because sample sizes must be whole numbers.

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 *N* 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 $n1 = n2 = \ldots = ng = n$. That is, all sample sizes are equal to *n*.

MC Procedure

Specify which pair-wise multiple comparison procedure is to be reported from the simulations. The choices are

Tukey-Kramer

This is the most popular and the most often recommended.

Kruskal-Wallis

This is recommended when a nonparametric procedure is wanted.

Games-Howell

This is recommended when the variances of the groups are unequal.

Specifying Simulation Distributions

These options specify the distributions to be used in the two simulations, one set per row. 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 i (i = 1 to 9)

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 Data1 tab and the six rows of distributions shown under the Data2 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*.

Note that since the first group is the control group, the value for Grps 1 is usually set to one.

Group i 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. Note that, except for the multinomial, the distributions are parameterized so that the mean is entered first.

Beta=A(M0,A,B,Minimum)
Binomial=B(M0,N)
Cauchy=C(M0,Scale)
Constant=K(Value)
Exponential=E(M0)
F=F(M0,DF1)
Gamma=G(M0,A)
Multinomial=M(P1,P2,…,Pk)
Normal=N(M0,SD)
Poisson=P(M0)
Student's T=T(M0,D)
Tukey's Lambda=L(M0,S,Skewness,Elongation)
Uniform=U(M0,Minimum)
Weibull=W(M0,B)

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

Finding the Value of the Mean of a Specified Distribution

Except for the multinomial distribution, the distributions have been parameterized in terms of their means since this is the parameter being tested. The mean of a distribution created as a linear combination of other distributions is found by applying the linear combination to the individual means. However, the mean of a distribution created by multiplying or dividing other distributions is not necessarily equal to applying the same function to the individual means. For example, the mean of 4N(4, 5) + 2N(5, 6) is 4*4 + 2*5 = 26, but the mean of 4N(4, 5) * 2N(5, 6) is not exactly 4*4*2*5 = 160 (although it is close).

Group i 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.

Following is a list of the distributions that are available and the syntax used to specify them. Note that, except for the multinomial, the distributions are parameterized so that the mean, M1, is entered first.

Beta=A(M1,A,B,Minimum)
Binomial=B(M1,N)
Cauchy=C(M1,Scale)
Constant=K(Value)
Exponential=E(M1)
F=F(M1,DF1)
Gamma=G(M1,A)
Multinomial=M(P1,P2,…,Pk)
Normal=N(M1,SD)
Poisson=P(M1)
Student's T=T(M1,D)
Tukey's Lambda=L(M1,S,Skewness,Elongation)
Uniform=U(M1,Minimum)
Weibull=W(M1,B)

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

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.

Minimum Difference

Specify the smallest difference between any two means that is to be detectable by the experiment. Pairs with mean differences smaller than this amount are to be considered equal.

Options Tab

The Options tab contains limits on the number of iterations and various options about individual tests.

Maximum Iterations

Specify the maximum number of iterations before the search for the sample size is aborted. When the maximum number of iterations is reached without convergence, the sample size is left blank. We recommend a value of at least 500.

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.

Reports Tab

The Reports tab contains settings about the format of the output.

Show Various Reports & Plots

These options let you specify whether you want to generate the standard reports and plots.

Show Inc's & 95% C.I.

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

Show Comparative Reports & Plots

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

Example1 - Power at Various Sample Sizes

An experiment is being designed to investigate the variety of response when an experiment is replicated under five different conditions. Previous studies have shown that the standard deviation within a group is 3.0. Researchers want to detect a shift in the mean of 3.0 or more. To accomplish this, they set the means of the first four groups to zero and the mean of the fifth group to 3.0. They want to investigate sample sizes of 5, 10, 15, and 20 subjects per group.

Although they will conduct an F-test on the data, their primary analysis will be a set of Tukey-Kramer multiple comparison tests. They set the FWER to 0.05.

Setup

This section presents the values of each of the parameters needed to run this example. You can make these changes directly on your screen or you can load the template entitled Example1 by clicking the Template tab and loading this template.

Option	Value
Data Tab	
Find	**Power**
Simulations	**2000**
FWER	**0.05**
Power	*Ignored since this is the Find setting*
n (Multiplier)	**5 10 15 20**
Group Sample Size Pattern	**Equal**
MC Procedure	**Tukey-Kramer**
Grps 1	**4**
Group 1 Distribution(s) | H0	**N(M0 S)**
Group 1 Distribution(s) | H1	**N(M0 S)**
Grps 2	**1**
Group 2 Distribution(s) | H0	**N(M0 S)**
Group 2 Distribution(s) | H1	**N(M0 S)**
M0	**0**
M1	**3**
S	**3**

Annotated Output

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

Simulation Summary Report

Summary of Simulations of 5 Groups
MC Procedure: Tukey-Kramer M.C. Test

Sim. No.	Any-Pairs Power	Group Smpl. Size n	Total Smpl. 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	0.251	5.0	25	0.014	1.2	3.0	0.043	0.050	0.0	3.0	3.0
	(0.019)	[0.232	0.269]	(0.005)	[0.008	0.019]	(0.009)	[0.034	0.051]		
2	0.562	10.0	50	0.077	1.2	3.0	0.052	0.050	0.0	3.0	3.0
	(0.022)	[0.540	0.583]	(0.012)	[0.065	0.088]	(0.010)	[0.042	0.062]		
3	0.770	15.0	75	0.200	1.2	3.0	0.057	0.050	0.0	3.0	3.0
	(0.018)	[0.751	0.788]	(0.018)	[0.182	0.217]	(0.010)	[0.046	0.067]		
4	0.898	20.0	100	0.356	1.2	3.0	0.060	0.050	0.0	3.0	3.0
	(0.013)	[0.884	0.911]	(0.021)	[0.335	0.377]	(0.010)	[0.049	0.070]		

Pool Size: 10000. Simulations: 2000. Run Time: 26.45 seconds.

Summary of Simulations Report Definitions
H0: the null hypothesis that each pair of group means are equal.
H1: the alternative hypothesis that at least one pair of group means are not equal.
All-Pairs Power: the estimated probability of detecting all unequal pairs.
Any-Pairs Power: the estimated probability of detecting at least one unequal pair.
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 equal pair assuming H0.
Target FWER: the user-specified FWER.
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-Pairs Power, All-Pairs 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 5.0 for a total sample size of 25. This design achieved an any-pair power of 0.2505 and an all-pair power of 0.0135 using the Tukey-Kramer M.C. Test with a target family-wise error rate of 0.050 and an actual target family-wise error rate 0.043. 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(M0 S); N(M0 S); and N(M1 S). Other parameters used in the simulation were: M0 = 0.0, M1 = 3.0, and S = 3.0.

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 all 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

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: Tukey-Kramer M.C. 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	10	0.057	0.006	0.043	0.050	0.006	0.005	0.007
2	10	0.072	0.007	0.052	0.050	0.007	0.005	0.011
3	10	0.075	0.008	0.057	0.050	0.008	0.006	0.011
4	10	0.075	0.008	0.060	0.050	0.008	0.004	0.010

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 pair-wise multiple comparisons.

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 5 Groups
MC Procedure: Tukey-Kramer M.C. 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	6/4	0.04	3.58	0.362	0.007	0.896	0.091	0.375	0.014	0.251	0.046	0.004	0.109
2	6/4	0.04	2.85	0.289	0.007	0.711	0.035	0.323	0.077	0.562	0.120	0.006	0.302
3	6/4	0.03	2.11	0.214	0.006	0.527	0.017	0.261	0.200	0.770	0.192	0.002	0.478
4	6/4	0.03	1.41	0.144	0.005	0.352	0.012	0.191	0.356	0.898	0.262	0.003	0.658

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 means 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 all 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.

Detail Model Report

Detailed Model Report for Simulation No. 1
Target FWER = 0.050, M0 = 0.0, M1 = 3.0, S = 3.0
MC Procedure: Tukey-Kramer M.C. Test

Hypo. Type	Groups	Group Labels	n/N	Group Mean	Ave. S.D.	Simulation Model
H0	1-4	A1-A4	5/25	0.0	2.9	N(M0 S)
H0	5	B1	5/25	0.0	3.0	N(M0 S)
H0	All			Sm=0.0	3.0	
H1	1-4	A1-A4	5/25	0.0	3.0	N(M0 S)
H1	5	B1	5/25	3.0	3.0	N(M1 S)
H1	All			Sm=1.2	3.0	

Detailed Model Report for Simulation No. 2

Hypo. Type	Groups	Group Labels	n/N	Group Mean	Ave. S.D.	Simulation Model
H0	1-4	A1-A4	10/50	0.0	3.0	N(M0 S)
H0	5	B1	10/50	0.0	3.0	N(M0 S)
H0	All			Sm=0.0	3.0	
H1	1-4	A1-A4	10/50	0.0	3.0	N(M0 S)
H1	5	B1	10/50	3.0	3.0	N(M1 S)
H1	All			Sm=1.2	3.0	

Detailed Model Report for Simulation No. 3

Hypo. Type	Groups	Group Labels	n/N	Group Mean	Ave. S.D.	Simulation Model
H0	1-4	A1-A4	15/75	0.0	3.0	N(M0 S)
H0	5	B1	15/75	0.0	3.0	N(M0 S)
H0	All			Sm=0.0	3.0	
H1	1-4	A1-A4	15/75	0.0	3.0	N(M0 S)
H1	5	B1	15/75	3.0	3.0	N(M1 S)
H1	All			Sm=1.2	3.0	

Detailed Model Report for Simulation No. 4

Hypo. Type	Groups	Group Labels	n/N	Group Mean	Ave. S.D.	Simulation Model
H0	1-4	A1-A4	20/100	0.0	3.0	N(M0 S)
H0	5	B1	20/100	0.0	3.0	N(M0 S)
H0	All			Sm=0.0	3.0	
H1	1-4	A1-A4	20/100	0.0	3.0	N(M0 S)
H1	5	B1	20/100	3.0	3.0	N(M1 S)
H1	All			Sm=1.2	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.

Probability of Rejecting Equality

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

Group	Means	A1	A2	A3	A4	B1
A1	0.0		0.006	0.006	0.011	0.109*
A2	0.0	0.006		0.006	0.004	0.106*
A3	0.0	0.005	0.006		0.008	0.097*
A4	0.0	0.005	0.007	0.007		0.107*
B1	3.0	0.005	0.005	0.007	0.006	

Probability of Rejecting the Equality of Each Pair. Simulation No. 2

Group	Means	A1	A2	A3	A4	B1
A1	0.0		0.007	0.007	0.008	0.294*
A2	0.0	0.006		0.007	0.006	0.280*
A3	0.0	0.011	0.007		0.006	0.302*
A4	0.0	0.005	0.007	0.005		0.281*
B1	3.0	0.008	0.009	0.008	0.007	

Probability of Rejecting the Equality of Each Pair. Simulation No. 3

Group	Means	A1	A2	A3	A4	B1
A1	0.0		0.005	0.002	0.007	0.478*
A2	0.0	0.006		0.006	0.007	0.477*
A3	0.0	0.007	0.008		0.005	0.473*
A4	0.0	0.009	0.008	0.007		0.465*
B1	3.0	0.006	0.011	0.007	0.008	

Probability of Rejecting the Equality of Each Pair. Simulation No. 4

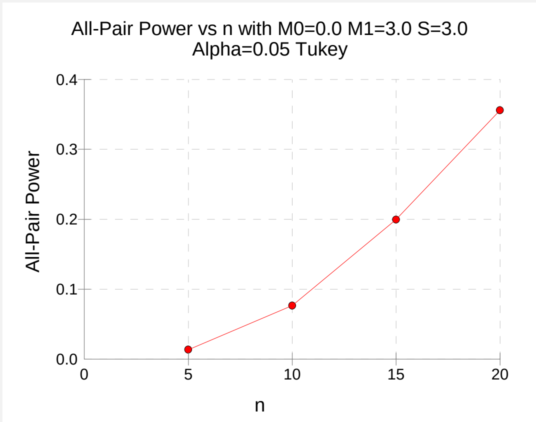
Group	Means	A1	A2	A3	A4	B1
A1	0.0		0.003	0.005	0.004	0.645*
A2	0.0	0.008		0.008	0.007	0.650*
A3	0.0	0.006	0.008		0.004	0.640*
A4	0.0	0.010	0.007	0.009		0.658*
B1	3.0	0.010	0.007	0.004	0.007	

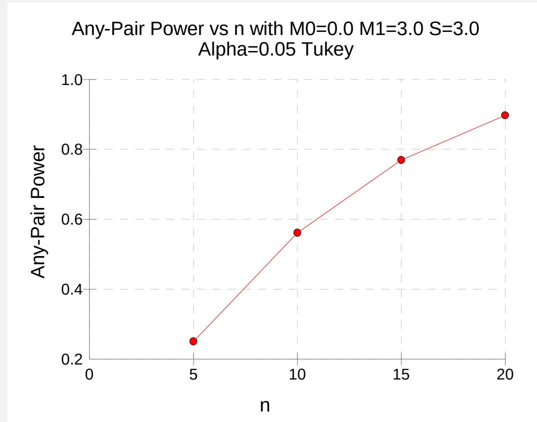
Individual pairwise powers from the H1 (Power) simulation are shown in the upper-right section.
Individual pairwise significance levels from the H0 (Alpha) simulation are shown in the lower-left section.
* Starred values are the powers of pairs that are unequal under H1.

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.

Plot Section





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

Example2 - Comparative results

Continuing with Example1, the researchers want to study the characteristics of alternative multiple comparison procedures.

Setup

This section presents the values of each of the parameters needed to run this example. You can enter these values yourself or load the Example2 template from the Template tab.

Option	Value
Data Tab	
Find	**Power**
Simulations	**2000**
FWER	**0.05**
Power	*Ignored since this is the Find setting*
n (Multiplier)	**5 10 15 20**
Group Sample Size Pattern	**Equal**
MC Procedure	**Tukey-Kramer**
Grps 1	**4**
Group 1 Distribution(s) | H0	**N(M0 S)**
Group 1 Distribution(s) | H1	**N(M0 S)**
Grps 2	**1**
Group 2 Distribution(s) | H0	**N(M0 S)**
Group 2 Distribution(s) | H1	**N(M1 S)**
M0	**0**
M1	**3**
S	**3**
Reports Tab	
Comparative Reports	**Checked**
Comparative Any-Pair Power Plot	**Checked**
Comparative All-Pair Power Plot	**Checked**

Annotated Output

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

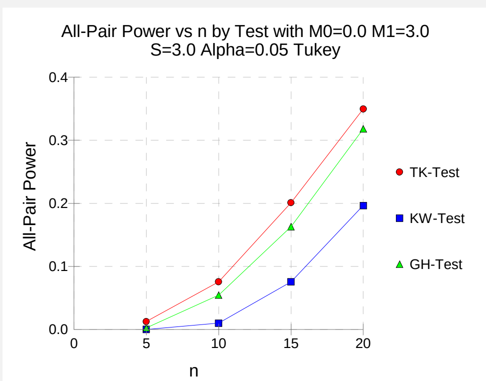
Power Comparison for Testing the 5 Group Means

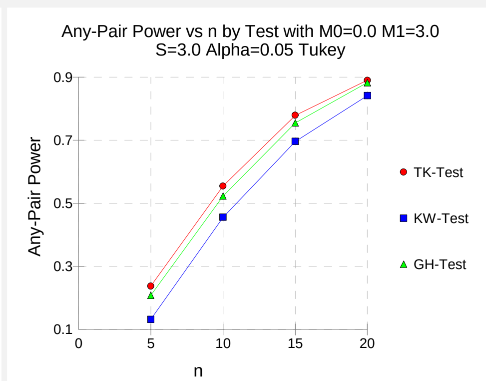
Sim. No.	Total Sample Size	Target Alpha	Tukey Kramer All-Pair Power	Kruskal Wallis All-Pair Power	Games Howell All-Pair Power	Tukey Kramer Any-Pair Power	Kruskal Wallis Any-Pair Power	Games Howell Any-Pair Power
1	25	0.050	0.013	0.000	0.003	0.238	0.132	0.209
2	50	0.050	0.076	0.010	0.055	0.555	0.456	0.523
3	75	0.050	0.201	0.076	0.163	0.779	0.697	0.755
4	100	0.050	0.350	0.197	0.318	0.890	0.842	0.883

Pool Size: 10000. Simulations: 2000. Run Time: 33.82 seconds.

Family-Wise FWER Comparison for Testing the 5 Group Means

Sim. No.	Total Sample Size	Target FWER	Tukey Kramer FWER	Kruskal Wallis FWER	Games Howell FWER
1	25	0.050	0.039	0.023	0.064
2	50	0.050	0.050	0.039	0.059
3	75	0.050	0.051	0.033	0.055
4	100	0.050	0.051	0.040	0.058





These reports show the power and FWER of each of the three multiple comparison procedures. In these simulations of groups from the normal distributions with equal variances, we see that the Tukey-Kramer procedure is the champion.

Example3 - Validation using Ramsey

Ramsey (1978) presents the results of a simulation study that compared the all-pair power of several different multiple comparison procedures. On page 483 of this article, he presents the results of a simulation in which there were four groups: two with means of -0.7 and two with means of 0.7. The standard deviation was 1.0 and the FWER was 0.05. Tukey's multiple comparison procedure was used in the simulation. The sample size was 16 per group. Using a simulation of 1000 iterations, the all-pairs power was calculated as 0.723. Note that a confidence interval for this estimated all-pairs power is (0.703 to 0.759).

Setup

This section presents the values of each of the parameters needed to run this example. You can enter these values yourself or load the Example3 template from the Template tab.

Option	Value
Data Tab	
Find	**Power**
Simulations	**2000**
FWER	**0.05**
Power	*Ignored since this is the Find setting*
n (Multiplier)	**16**
Group Sample Size Pattern	**Equal**
MC Procedure	**Tukey-Kramer**
Grps 1	**2**
Group 1 Distribution(s) | H0	**N(M0 S)**
Group 1 Distribution(s) | H1	**N(M0 S)**
Grps 2	**2**
Group 2 Distribution(s) | H0	**N(M0 S)**
Group 2 Distribution(s) | H1	**N(M1 S)**
M0	**-0.7**
M1	**0.7**
S	**1**

Annotated Output

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

Numeric Results

Summary of Simulations of 5 Groups
MC Procedure: Tukey-Kramer M.C. Test

Sim. No.	Any-Pairs Power	Group Smpl. Size n	Total Smpl. 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	0.996	16.0	64	0.729	0.7	1.0	0.052	0.050	-0.7	0.7	1.0
	(0.004)	[0.992	1.000]	(0.028)	[0.701	0.757]	(0.014)	[0.038	0.066]		

Pool Size: 10000. Simulations: 1000. Run Time: 4.28 seconds.

Note that the value found by ***PASS*** of 0.729 is very close to the value of 0.723 found by Ramsey (1978). More importantly, the value found by ***PASS*** is inside the confidence limits of Ramsey's study.

We ran the simulation five more times and obtained 0.727, 0.723, 0.714, 0.740, and 0.732. We also ran the simulation with 10,000 iterations and obtained a power of 0.736 with a confidence interval of (0.727 to 0.745).

Example4 - Selecting a multiple comparison procedure when the data contain outliers

This example will investigate the impact of outliers on the power and precision of the various multiple comparison procedures when there are five groups.

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

Setup

This section presents the values of each of the parameters needed to run this example. You can enter these values yourself or load the Example4 template from the Template tab.

Option	Value
Data Tab	
Find	**Power**
Simulations	**2000**
FWER	**0.05**
Power	*Ignored since this is the Find setting*
n (Multiplier)	**10 20**
Group Sample Size Pattern	**Equal**
Multiple Comparison Procedure	**Tukey Kramer**
Grps 1	**2**
Group 1 Distribution(s) | H0	**N(M0 S)[95];N(M0 A)[5]**
Group 1 Distribution(s) | H1	**N(M0 S)[95];N(M0 A)[5]**
Grps 2	**3**
Group 2 Distribution(s) | H0	**N(M0 S)[95];N(M0 A)[5]**
Group 2 Distribution(s) | H1	**N(M1 S)[95];N(M1 A)[5]**
S	**1**
A	**1 5 10**
Reports Tab	
Show Comparative Reports	**Checked**
Show Comparative Plots	**Checked**

Annotated Output

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

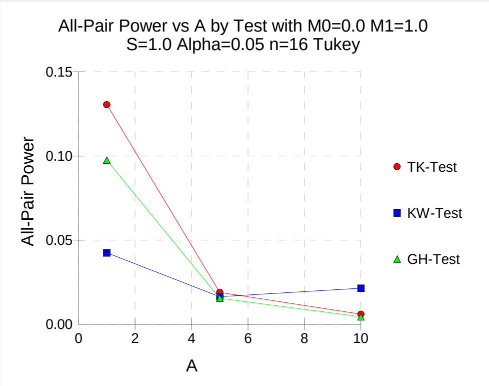
Power Comparison for Testing the 5 Group Means

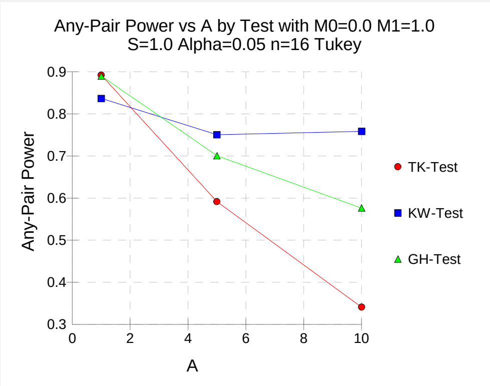
Sim. No.	Total Sample Size	Target Alpha	Tukey Kramer All-Pair Power	Kruskal Wallis All-Pair Power	Games Howell All-Pair Power	Tukey Kramer Any-Pair Power	Kruskal Wallis Any-Pair Power	Games Howell Any-Pair Power
1	80	0.050	0.131	0.043	0.098	0.892	0.837	0.890
2	80	0.050	0.019	0.017	0.016	0.592	0.751	0.701
3	80	0.050	0.006	0.022	0.005	0.341	0.759	0.577

Pool Size: 10000. Simulations: 2000. Run Time: 33.82 seconds.

Family-Wise Error-Rate Comparison for Testing the 5 Group Means

Sim. No.	Total Sample Size	Target FWER	Tukey Kramer FWER	Kruskal Wallis FWER	Games Howell FWER
1	80	0.050	0.046	0.036	0.049
2	80	0.050	0.044	0.044	0.040
3	80	0.050	0.033	0.038	0.025





These reports show the power and FWER of each of the three multiple comparison procedures. We note that when the variances are equal (A = 1), the Tukey-Kramer procedure performs only slightly better than the others. However, as the number of outliers is increased, the Kruskal-Wallis procedure emerges as the better choice. Also note that in the case with many outliers (Simulation 3), the FWER of all procedures is below the target value.

Chapter 585

Multiple Comparisons of Treatments versus a Control using 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 $\mu_1, \mu_2, \cdots, \mu_k$ 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.

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

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

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 $\alpha = \alpha_f / \left(2\left(k-1\right)\right)$. For each treatment and control pair, the significance test is calculated by rejecting the null hypothesis of median equality if

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

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 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 unique to this procedure. These are located on the panels associated with the Data, Options, and Reports tabs. To find out more about using the other tabs such as Plot Text, Axes, and Template, turn to the chapter entitled Procedure Templates.

Data Tab

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

Find

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-pair power or all-pairs power. The value of the option *Power* will then represent this type of power.

Any-pair power is the probability of detecting at least one pair from among those that are actually different. *All-pairs power* is the probability of detecting all pairs that are actually different.

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.

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.

FWER (Alpha)

This option specifies one or more values of the *family-wise error rate* (FWER) which is the analog of alpha for multiple comparisons. 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

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

or

$$ICER = 1 - (1 - FWER)^\wedge(1/C)$$

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.

Power

This option is only used when *Find (Solve For)* is set to *n (All-Pairs)* or *n (Any-Pair)*.

Power is defined differently with multiple comparisons. Although many definitions are possible, two are adopted here. *Any-pair power* is the probability of detecting at least one pair of the means that are different. *All-pairs power* is the probability of detecting all pairs of means that are truly different. As the number of groups 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.

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

$m1, m2, m3, \ldots, mg$ and this value is represented by n, the group sample sizes

$n1, n2, n3, \ldots, ng$ 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. This is required because sample sizes must be whole numbers.

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 N 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 $n1 = n2 = \ldots = ng = n$. That is, all sample sizes are equal to *n*.

MC Procedure

Specify which pair-wise multiple comparison procedure is to be reported from the simulations. The choices are

Dunnett

This is the most popular and the most often recommended.

Kruskal-Wallis

This is recommended when a nonparametric procedure is wanted.

Specifying Simulation Distributions

These options specify the distributions to be used in the two simulations, one set per row. 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.

Note that group number one is the control group.

Grps i (i = 1 to 9)

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 Data1 tab and the six rows of distributions shown under the Data2 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*.

Note that since the first group is the control group, the value for Grps 1 is usually set to one.

Group i 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. Note that, except for the multinomial, the distributions are parameterized so that the mean is entered first.

Beta=A(M0,A,B,Minimum)
Binomial=B(M0,N)
Cauchy=C(M0,Scale)
Constant=K(Value)
Exponential=E(M0)
F=F(M0,DF1)
Gamma=G(M0,A)
Multinomial=M(P1,P2,…,Pk)
Normal=N(M0,SD)
Poisson=P(M0)
Student's T=T(M0,D)
Tukey's Lambda=L(M0,S,Skewness,Elongation)
Uniform=U(M0,Minimum)
Weibull=W(M0,B)

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

Finding the Value of the Mean of a Specified Distribution

Except for the multinomial distribution, the distributions have been parameterized in terms of their means since this is the parameter being tested. The mean of a distribution created as a linear combination of other distributions is found by applying the linear combination to the individual means. However, the mean of a distribution created by multiplying or dividing other distributions is not necessarily equal to applying the same function to the individual means. For example, the mean of 4N(4, 5) + 2N(5, 6) is 4*4 + 2*5 = 26, but the mean of 4N(4, 5) * 2N(5, 6) is not exactly 4*4*2*5 = 160 (although it is close).

Group i 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.

Following is a list of the distributions that are available and the syntax used to specify them. Note that, except for the multinomial, the distributions are parameterized so that the mean, M1, is entered first.

Beta=A(M1,A,B,Minimum)
Binomial=B(M1,N)
Cauchy=C(M1,Scale)
Constant=K(Value)
Exponential=E(M1)
F=F(M1,DF1)
Gamma=G(M1,A)
Multinomial=M(P1,P2,…,Pk)
Normal=N(M1,SD)
Poisson=P(M1)
Student's T=T(M1,D)
Tukey's Lambda=L(M1,S,Skewness,Elongation)
Uniform=U(M1,Minimum)
Weibull=W(M1,B)

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

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.

Minimum Difference

Specify the smallest difference between any two means that is to be detectable by the experiment. Pairs with mean differences smaller than this amount are to be considered equal.

Options Tab

The Options tab contains limits on the number of iterations and various options about individual tests.

Maximum Iterations

Specify the maximum number of iterations before the search for the sample size is aborted. When the maximum number of iterations is reached without convergence, the sample size is left blank. We recommend a value of at least 500.

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.

Reports Tab

The Reports tab contains settings about the format of the output.

Show Various Reports & Plots

These options let you specify whether you want to generate the standard reports and plots.

Show Inc's & 95% C.I.

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

Show Comparative Reports & Plots

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

Example1 - 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

This section presents the values of each of the parameters needed to run this example. You can enter these values yourself or load the Example1 template from the Template tab.

Option	Value
Data Tab	
Find	**Power**
Simulations	**2000**
FWER	**0.05**
Power	*Ignored since this is the Find setting*
n (Multiplier)	**5 10 15 20**
Group Sample Size Pattern	**Equal**
MC Procedure	**Dunnett**
Grps 1	**1**
Control Distribution | H0	**N(M0 S)**
Control Distribution | H1	**N(M0 S)**
Grps 2	**3**
Group 2 Distribution(s) | H0	**N(M0 S)**
Group 2 Distribution(s) | H1	**N(M1 S)**
M0	**0**
M1	**2**
S	**3**
Report Tab	
All reports except Comparative	**checked**

Annotated Output

Click the Run 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
MC Procedure: Dunnett's M.C. Test

Sim. No.	Any-Pairs Power	Group Smpl. Size n	Total Smpl. 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	0.187	5.0	20	0.016	0.9	3.0	0.055	0.050	0.0	2.0	3.0
	(0.017)	[0.170	0.204]	(0.005)	[0.010	0.021]	(0.010)	[0.045	0.064]		
2	0.358	10.0	40	0.045	0.9	3.0	0.050	0.050	0.0	2.0	3.0
	(0.021)	[0.337	0.379]	(0.009)	[0.036	0.054]	(0.010)	[0.040	0.059]		
3	0.507	15.0	60	0.098	0.9	3.0	0.042	0.050	0.0	2.0	3.0
	(0.022)	[0.485	0.529]	(0.013)	[0.085	0.111]	(0.009)	[0.033	0.051]		
4	0.623	20.0	80	0.162	0.9	3.0	0.045	0.050	0.0	2.0	3.0
	(0.021)	[0.602	0.644]	(0.016)	[0.145	0.178]	(0.009)	[0.035	0.054]		

Pool Size: 10000. Simulations: 2000. Run Time: 26.45 seconds.

Summary of Simulations Report Definitions
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'.
All-Pairs Power: the estimated probability of detecting all unequal pairs.
Any-Pairs Power: the estimated probability of detecting at least one unequal pair.
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 equal pair 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 aconfidence interval based on the size of the simulation for Any-Pairs Power, All-Pairs Power, and FWER. The format is (Precision) [95% LCL and UCL Alpha].

Summary Statements
A one-way design with 3 treatment groups and one control group has an average group sample size of 5.0 for a total sample size of 20. This design achieved an any-pair power of 0.187 and an all-pair power of 0.0155 using the Dunnett's Test procedure for comparing each treatment mean with the control mean. The target family-wise error rate was 0.050 and the actual family-wise error rate was 0.055. 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); 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 = 0.0, M1 = 2.0, and S = 3.0.

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 all 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 M.C. 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	0.067	0.022	0.055	0.050	0.022	0.020	0.025
2	3	0.060	0.020	0.050	0.050	0.020	0.017	0.023
3	3	0.048	0.016	0.042	0.050	0.016	0.016	0.017
4	3	0.055	0.018	0.045	0.050	0.018	0.018	0.019

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 M.C. 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	0.00	2.74	0.913	0.000	0.913	0.000	1.000	0.016	0.187	0.088	0.082	0.092
2	0/3	0.00	2.45	0.816	0.000	0.816	0.000	1.000	0.045	0.358	0.184	0.182	0.188
3	0/3	0.00	2.13	0.711	0.000	0.711	0.000	1.000	0.098	0.507	0.289	0.286	0.292
4	0/3	0.00	1.84	0.613	0.000	0.613	0.000	1.000	0.162	0.623	0.388	0.380	0.392

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 all of the pairs that were different in the H1 simulation.

Any Uneq. Pairs Power

This is the probability of detecting any 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.

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: Dunnett's M.C. Test

Hypo. Type	Groups	Group Labels	n/N	Group Mean	Ave. S.D.	Simulation Model
H0	1	Cntl	5/20	0.0	3.0	N(M0 S)
H0	2-4	B1-B3	5/20	0.0	3.0	N(M0 S)
H0	All			Sm=0.0	3.0	
H1	1	Cntl	5/20	0.0	3.0	N(M0 S)
H1	2-4	B1-B3	5/20	2.0	3.0	N(M1 S)
H1	All			Sm=0.9	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	Cntl	10/40	0.0	3.0	N(M0 S)
H0	2-4	B1-B3	10/40	0.0	3.0	N(M0 S)
H0	All			Sm=0.0	3.0	
H1	1	Cntl	10/40	0.0	3.0	N(M0 S)
H1	2-4	B1-B3	10/40	2.0	3.0	N(M1 S)
H1	All			Sm=0.9	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	Cntl	15/60	0.0	3.0	N(M0 S)
H0	2-4	B1-B3	15/60	0.0	3.1	N(M0 S)
H0	All			Sm=0.0	3.0	
H1	1	Cntl	15/60	0.0	2.9	N(M0 S)
H1	2-4	B1-B3	15/60	2.0	3.0	N(M1 S)
H1	All			Sm=0.9	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	Cntl	20/80	0.0	3.0	N(M0 S)
H0	2-4	B1-B3	20/80	0.0	3.0	N(M0 S)
H0	All			Sm=0.0	3.0	
H1	1	Cntl	20/80	0.0	3.0	N(M0 S)
H1	2-4	B1-B3	20/80	2.0	3.0	N(M1 S)
H1	All			Sm=0.9	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. Note that the control group is labeled 'Cntl'.

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	Cntl	B1	B2	B3
Cntl	0.0		0.082*	0.092*	0.090*
B1	2.0	0.025			
B2	2.0	0.020			
B3	2.0	0.022			

Probability of Rejecting the Equality of Each Pair. Simulation No. 2

Group	Means	Cntl	B1	B2	B3
Cntl	0.0		0.188*	0.184*	0.182*
B1	2.0	0.023			
B2	2.0	0.020			
B3	2.0	0.017			

Probability of Rejecting the Equality of Each Pair. Simulation No. 3

Group	Means	Cntl	B1	B2	B3
Cntl	0.0		0.292*	0.286*	0.289*
B1	2.0	0.016			
B2	2.0	0.016			
B3	2.0	0.017			

Probability of Rejecting the Equality of Each Pair. Simulation No. 4

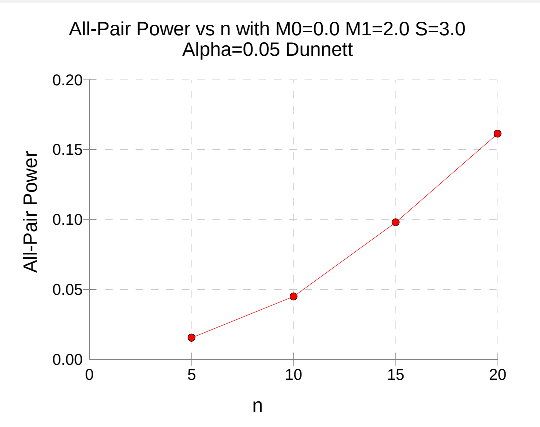
Group	Means	Cntl	B1	B2	B3
Cntl	0.0		0.392*	0.380*	0.392*
B1	2.0	0.019			
B2	2.0	0.018			
B3	2.0	0.018			

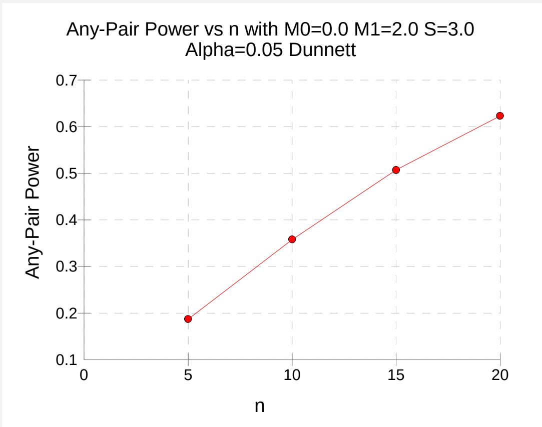
Individual pairwise powers from the H1 (Power) simulation are shown in the upper-right section.
Individual pairwise significance levels from the H0 (Alpha) simulation are shown in the lower-left section.
* Starred values are the powers of pairs that are unequal under H1.

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.

Plot Section





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

Example2 - Comparative Results

Continuing with Example1, the researchers want to study the characteristics of alternative multiple comparison procedures.

Setup

This section presents the values of each of the parameters needed to run this example. You can enter these values yourself or load the Example2 template from the Template tab.

Option	Value
Data Tab	
Find	**Power**
Simulations	**2000**
FWER	**0.05**
Power	*Ignored since this is the Find setting*
n (Multiplier)	**5 10 15 20**
Group Sample Size Pattern	**Equal**
MC Procedure	**Dunnett**
Grps 1	**1**
Control Distribution | H0	**N(M0 S)**
Control Distribution | H1	**N(M0 S)**
Grps 2	**3**
Group 2 Distribution(s) | H0	**N(M0 S)**
Group 2 Distribution(s) | H1	**N(M1 S)**
M0	**0**
M1	**2**
S	**3**
Reports Tab	
Comparative Reports	**Checked**
Comparative Any-Pair Power Plot	**Checked**
Comparative All-Pair Power Plot	**Checked**

Annotated Output

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

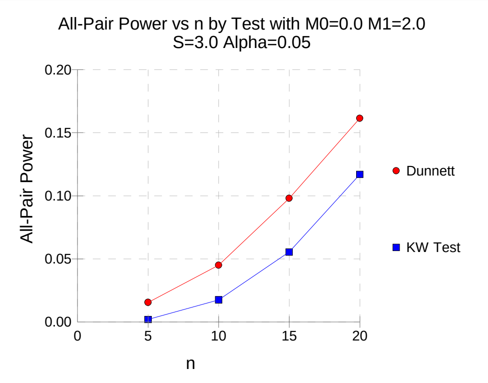
Power Comparison for Testing the Control Group Versus the 3 Treatment Groups

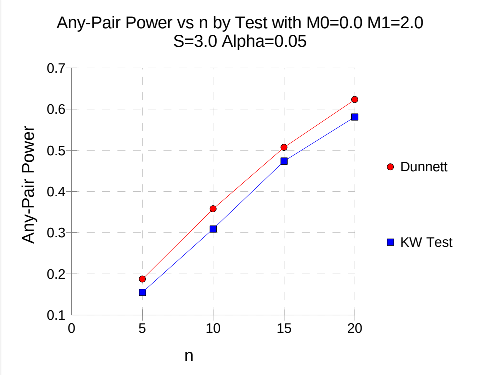
Sim. No.	Total Sample Size	Target Alpha	Dunnett All-Pair Power	Kruskal Wallis All-Pair Power	Dunnett All-Pair Power	Kruskal Wallis Any-Pair Power
1	20	0.050	0.016	0.002	0.187	0.155
2	40	0.050	0.045	0.018	0.358	0.309
3	60	0.050	0.098	0.056	0.507	0.474
4	80	0.050	0.162	0.117	0.623	0.581

Pool Size: 10000. Simulations: 2000. Run Time: 26.18 seconds.

Family-Wise FWER 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	0.050	0.055	0.044
2	40	0.050	0.050	0.040
3	60	0.050	0.042	0.034
4	80	0.050	0.045	0.038





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.

Example3 - Validation using Dunnett

Murkerjee, 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.

Setup

This section presents the values of each of the parameters needed to run this example. You can enter these values yourself or load the Example3 template from the Template tab.

Option	Value
Data Tab	
Find	**Power**
Simulations	**2000**
FWER	**0.10**
Power	*Ignored since this is the Find setting*
n (Multiplier)	**16 21**
Group Sample Size Pattern	**Equal**
MC Procedure	**Dunnett**
Grps 1	**1**
Control Distribution(s) | H0	**N(M0 S)**
Control Distribution(s) | H1	**N(M0 S)**
Grps 2	**1**
Group 2 Distribution(s) | H0	**N(M0 S)**
Group 2 Distribution(s) | H1	**N(M1 S)**
Grps 3	**4**
Group 3 Distribution(s) | H0	**N(M0 S)**
Group 3 Distribution(s) | H1	**N(M0 S)**
M0	**-0.182574**
M1	**0.912871**
S	**1.0**

Annotated Output

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

Numeric Results

Summary of Simulations of the Control Group and the 5 Treatment Groups
MC Procedure: Dunnett's M.C. Test

Sim. No.	Any-Pairs Power	Group Smpl. Size n	Total Smpl. 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	0.795	16.0	96	0.795	0.4	1.0	0.095	0.100	-0.2	0.9	1.0
	(0.018)	[0.777	0.812]	(0.018)	[0.777	0.812]	(0.013)	[0.082	0.107]		
2	0.894	21.0	126	0.894	0.4	1.0	0.112	0.100	-0.2	0.9	1.0
	(0.014)	[0.880	0.907]	(0.014)	[0.880	0.907]	(0.014)	[0.098	0.125]		

Pool Size: 10000. Simulations: 1000. Run Time: 4.28 seconds.

For the first case when n = 16, ***PASS*** obtained a power of 0.795 which is very close to the value of 0.80 found by Muterjee et al. (1987). Indeed, 0.80 is within the confidence limits of 0.777 to 0.812. Similarly, the power for the second case when n = 21 is found as 0.894 which is very close to the article's value of 0.90.

Chapter 590

Multiple Contrasts using 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 (1961) 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 (1961) 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 unique to this procedure. These are located on the panels associated with the Data, Options, and Reports tabs. To find out more about using the other tabs such as Plot Text, Axes, and Template, turn to the chapter entitled Procedure Templates.

Data Tab

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

Find

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.

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.

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

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

or

$$ICER = 1 - (1 - FWER)^\wedge(1/C)$$

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.

Power

This option is only used when *Find (Solve For)* is set to *n (All-Contrast)* or *n (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.

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, …, mg and this value is represented by *n*, the group sample sizes

n1, n2, n3, ...,ng 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. This is required because sample sizes must be whole numbers.

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 *N* 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 $n1 = n2 = \ldots = ng = n$. That is, all sample sizes are equal to *n*.

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

This is recommended when the group variances are very different.

Specifying Simulation Distributions

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 i (i = 1 to 9)

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 Data1 tab and the six rows of distributions shown under the Data2 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 i 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. Note that, except for the multinomial, the distributions are parameterized so that the mean is entered first.

Beta=A(M0,A,B,Minimum)
Binomial=B(M0,N)
Cauchy=C(M0,Scale)
Constant=K(Value)
Exponential=E(M0)
F=F(M0,DF1)
Gamma=G(M0,A)
Multinomial=M(P1,P2,…,Pk)
Normal=N(M0,SD)
Poisson=P(M0)
Student's T=T(M0,D)
Tukey's Lambda=L(M0,S,Skewness,Elongation)
Uniform=U(M0,Minimum)
Weibull=W(M0,B)

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

Finding the Value of the Mean of a Specified Distribution

Except for the multinomial distribution, the distributions have been parameterized in terms of their means since this is the parameter being tested. The mean of a distribution created as a linear combination of other distributions is found by applying the linear combination to the individual means. However, the mean of a distribution created by multiplying or dividing other distributions is not necessarily equal to applying the same function to the individual means. For example, the mean of 4N(4, 5) + 2N(5, 6) is 4*4 + 2*5 = 26, but the mean of 4N(4, 5) * 2N(5, 6) is not exactly 4*4*2*5 = 160 (although it is close).

Group i 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.

Following is a list of the distributions that are available and the syntax used to specify them. Note that, except for the multinomial, the distributions are parameterized so that the mean, M1, is entered first.

Beta=A(M1,A,B,Minimum)
Binomial=B(M1,N)
Cauchy=C(M1,Scale)
Constant=K(Value)
Exponential=E(M1)
F=F(M1,DF1)
Gamma=G(M1,A)
Multinomial=M(P1,P2,…,Pk)
Normal=N(M1,SD)
Poisson=P(M1)
Student's T=T(M1,D)
Tukey's Lambda=L(M1,S,Skewness,Elongation)
Uniform=U(M1,Minimum)
Weibull=W(M1,B)

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

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.

Minimum Difference

Specify the smallest difference between any two means that is to be detectable by the experiment. Contrasts with values smaller than this amount are to be considered equal.

Contrasts Tab

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 limits on the number of iterations and various options about individual tests.

Maximum Iterations

Specify the maximum number of iterations before the search for the sample size is aborted. When the maximum number of iterations is reached without convergence, the sample size is left blank. We recommend a value of at least 500.

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.

Reports Tab

The Reports tab contains settings about the format of the output.

Show Various Reports & Plots

These options let you specify whether you want to generate the standard reports and plots.

Show Inc’s & 95% C.I.

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

Show Comparative Reports & Plots

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

Example1 - 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. You can enter these values yourself or load the Example1 template from the Template tab.

Option	Value
Data Tab	
Find	**Power**
Simulations	**2000**
FWER	**0.05**
Power	*Ignored since this is the Find setting*
n (Multiplier)	**10 30 50 70**
Group Sample Size Pattern	**Equal**
MC Procedure	**Dunn-Bonferroni**
Grps 1	**2**
Control Distribution | H0	**N(M0 S)**
Control Distribution | H1	**N(M0 S)**
Grps 2	**3**
Group 2 Distribution(s) | H0	**N(M0 S)**
Group 2 Distribution(s) | H1	**N(M1 S)**
M0	**0**
M1	**2**
S	**3**
Contrasts Tab	
Contrasts	**Each With Next**
Report Tab	
All reports except Comparative	**checked**

Annotated Output

Click the Run 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 <u>any</u> 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 <u>all</u> 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

	Groups				
Contrasts	**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

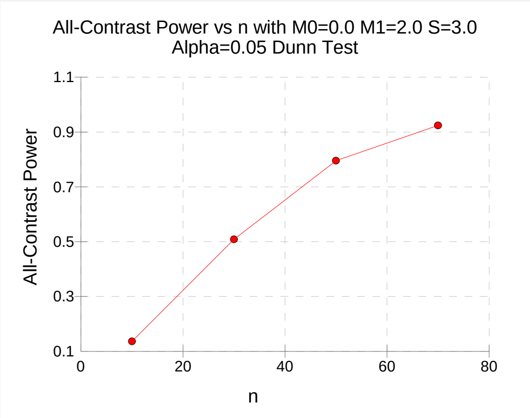
Probability of Rejecting Individual Contrasts. Simulation No. 4

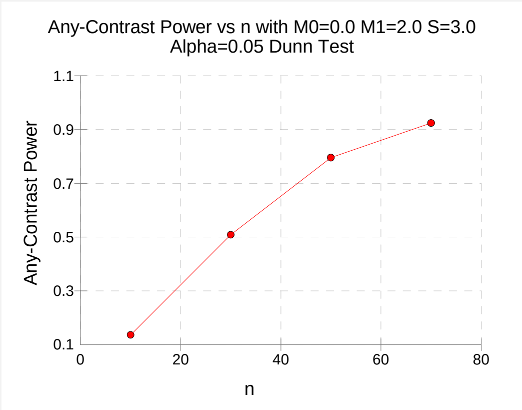
Contrasts	Alpha	Power
Con1	0.010	0.012
Con2	0.014	0.924
Con3	0.011	0.011
Con4	0.012	0.008

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.

Plot Section





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

Example2 - Comparative Results

Continuing with Example1, 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. You can enter these values yourself or load the Example2 template from the Template tab.

Option	Value
Data Tab	
Find	**Power**
Simulations	**2000**
FWER	**0.05**
Power	*Ignored since this is the Find setting*
n (Multiplier)	**10 30 50 70**
Group Sample Size Pattern	**Equal**
MC Procedure	**Dunn-Bonferroni**
Grps 1	**2**
Control Distribution | H0	**N(M0 S)**
Control Distribution | H1	**N(M0 S)**
Grps 2	**3**
Group 2 Distribution(s) | H0	**N(M0 S)**
Group 2 Distribution(s) | H1	**N(M1 S)**
M0	**0**
M1	**2**
S	**3**
Contrasts Tab	
Contrasts	**Each With Next**
Reports Tab	
Comparative Reports	**Checked**
Comparative Any-Contrast Power Plot	**Checked**
Comparative All-Contrast Power Plot	**Checked**

Annotated Output

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

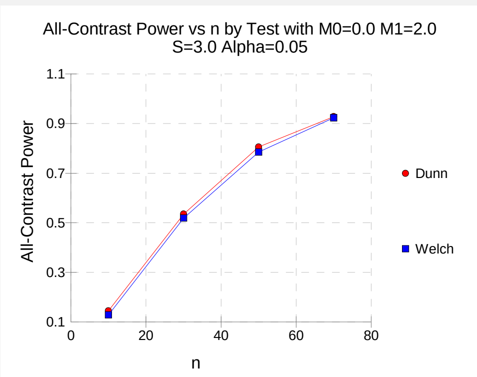
Power Comparison for Simultaneously Testing Multiple Contrasts of 5 Groups

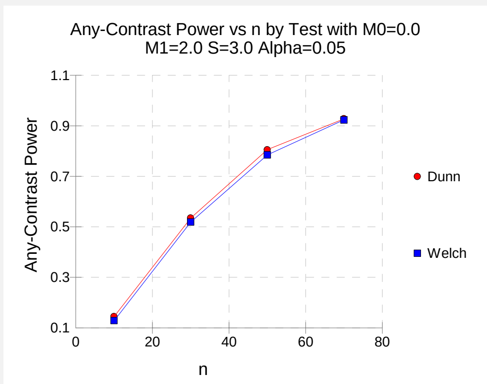
Sim. No.	Total Sample Size	Target Alpha	Dunn Bonferroni All-Cont. Power	Dunn Welch All-Cont. Power	Dunn Bonferroni Any-Cont. Power	Dunn Welch Any-Cont. Power
1	50	0.050	0.145	0.129	0.145	0.129
2	150	0.050	0.535	0.519	0.535	0.519
3	250	0.050	0.806	0.785	0.806	0.785
4	350	0.050	0.928	0.924	0.928	0.924

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. You can enter these values yourself or load the Example3 template from the Template tab.

Option	Value
Data Tab	
Find	**Power**
Simulations	**2000**
FWER	**0.05**
Power	*Ignored since this is the Find setting*
n (Multiplier)	**10 30 50 70**
Group Sample Size Pattern	**Equal**
MC Procedure	**Dunn-Bonferroni**
Grps 1	**2**
Control Distribution | H0	**N(M0 S)**
Control Distribution | H1	**N(M0 S)**
Grps 2	**3**
Group 2 Distribution(s) | H0	**N(M0 S)**
Group 2 Distribution(s) | H1	**N(M1 S)**
M0	**0**
M1	**2**
S	**3**
Contrasts Tab	
Contrasts	**0 -1 1 0 0**

Annotated Output

Click the Run 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.

Chapter 600

Hotelling's T2

Introduction

This module calculates power for Hotelling's one-group, and two-group, T-squared (T2) test statistics. Hotelling's T2 is an extension of the univariate t-tests in which the number of response variables is greater than one. In the two-group case, these results may also be obtained using ***PASS's*** MANOVA test.

Assumptions

The following assumptions are made when using Hotelling's T2 to analyze one or two groups of data.

1. The response variables are continuous.
2. The residuals follow the multivariate normal probability distribution with mean zero and constant variance-covariance matrix.
3. The subjects are independent.

Technical Details

The formulas used to perform a Hotelling's T2 power analysis provide exact answers if the above assumptions are met. These formulas can be found in many places. We use the results in Rencher (1998). We refer you to that reference for more details.

One-Group Case

In this case, a set of N observations is available on p response variables. We assume that all N observations have the same multivariate normal distribution with mean vector μ and variance covariance matrix Σ and that Hotelling's $T2$ is used for testing the null hypothesis that $\mu = \mu_0$ versus the alternative that $\mu = \mu_A$ where at least one component of μ_A is different from the corresponding component of μ_0. Usually, the vector μ_0 is a vector of zeros.

The value of $T2$ is computed using the formula

$$T^2_{p,N-1} = N\left(\bar{y} - \mu_0\right)' S^{-1}\left(\bar{y} - \mu_0\right)$$

where $\bar{y}$ is the vector of sample means and S is the sample variance-covariance matrix.

To calculate power we need the noncentrality parameter for this distribution. This noncentrality parameter is defined as follows

$$\begin{aligned}\lambda &= N\left(\mu_A - \mu_0\right)' \Sigma^{-1}\left(\mu_A - \mu_0\right) \\ &= N\Delta^2\end{aligned}$$

where

$$\Delta = \sqrt{\left(\mu_A - \mu_0\right)' \Sigma^{-1}\left(\mu_A - \mu_0\right)}$$

We define Δ as *effect size* because it provides a expression for the magnitude of the standardized difference between the null and alternative means.

Using this noncentrality parameter, the power of the Hotelling's T2 may be calculated for any value of the means and standard deviations. Since there is a simple relationship between the noncentral T2 and the noncentral *F*, calculations are actually based on the noncentral *F* using the formula

$$\beta = \Pr\left(F' < F'_{\alpha, df1, df2, \lambda}\right)$$

where

$$\begin{aligned} df1 &= p \\ df2 &= N - p \end{aligned}$$

Two-Group Case

In this case, sets of *N1* observations from group 1 and *N2* observations from group 2 are available on *p* response variables. We assume that all observations have the multivariate normal distribution with common variance covariance matrix Σ. The mean vectors of the two groups are assumed to be μ_1 and μ_2 under the alternative hypothesis. Under the null hypothesis, these mean vectors are assumed to be equal.

The value of $T2$ is computed using the formula

$$T^2_{p, N1+N2-2} = \frac{N1N2}{N1+N2}\left(\bar{y}_1 - \bar{y}_2\right)' S_{sp}^{-1}\left(\bar{y}_1 - \bar{y}_2\right)$$

where $\bar{y}_1$ and $\bar{y}_2$ are the vectors sample mean vectors of the two groups and S_{pl} is the pooled sample variance-covariance matrix.

To calculate power we need the noncentrality parameter for this distribution. This noncentrality parameter is defined as follows

$$\begin{aligned}\lambda &= \frac{N1N2}{N1+N2}\left(\mu_1-\mu_2\right)' \Sigma^{-1}\left(\mu_1-\mu_2\right) \\ &= \frac{N1N2}{N1+N2}\Delta^2\end{aligned}$$

where

$$\Delta = \sqrt{\left(\mu_1-\mu_2\right)' \Sigma^{-1}\left(\mu_1-\mu_2\right)}$$

We define Δ as *effect size* because it provides a expression for the magnitude of the standardized difference between the null and alternative means.

Using this noncentrality parameter, the power of the Hotelling's T2 may be calculated for any value of the means and standard deviations. Since there is a simple relationship between the noncentral T2 and the noncentral *F*, calculations are actually based on the noncentral *F* using the formula

$$\beta = \Pr\left(F' < F'_{\alpha,df1,df2,\lambda}\right)$$

where

$$\begin{aligned}df1 &= p \\ df2 &= N1+N2-p-1\end{aligned}$$

Procedure Options

This section describes the options that are unique to this procedure. To find out more about using the other tabs, turn to the chapter entitled Procedure Templates.

Data Tab

The Data tab contains many of the options that you will be primarily concerned with.

Find (Solve For)

This option specifies the parameter to be solved for.

When you choose to solve for *N* (one-group) or *N1* (two-group), the program searches for the lowest sample size that meets the alpha and beta criterion you have specified.

Groups

Specify whether the analysis is for one group or two groups. If '1' is selected, N1 is used as the sample size (N) and the values of N2 and R are ignored.

Response Variables

Enter the number of response (dependent or Y) variables. For a true multivariate test, this value will be greater than one.

The number of mean differences entered in the Mean Differences box or in the Means column must equal this value. If you read-in the covariance matrix from the spreadsheet, the number of columns specified must equal this value.

Mean Differences

Enter a list of values representing the mean differences under the alternative hypothesis. Under the null hypothesis, these values are all zero. The values entered here represent the differences that you want the experiment (study) to be able to detect.

Note that the number of values must match the number of Response Variables.

If you like, you can enter these values in a column on the spreadsheet. This column is specified using the 'Means Column' option. When that option is specified, any values entered here are ignored.

Means Column

Use this option to specify the spreadsheet column containing the hypothesized mean differences. The response variables are represented down the rows. The number of rows with data must equal the number of response variables. When this option is used, the 'Mean Differences' box is ignored.

You can obtain the spreadsheet by selecting 'Window', then 'Data', from the menus.

K (Means Multipliers)

These values are multiplied times the mean differences to give you various effect sizes. A separate power calculation is generated for each value of K. If you want to ignore this setting, enter '1'.

Alpha (Significance Level)

This option specifies the probability of a type-I error (alpha). A type-I error occurs when you reject the null hypothesis that all mean differences are zero this is true. Since these are probabilities, alpha values must be between zero and one.

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

Beta (1 - Power)

These options specify the probability of a type-II error (beta) for each factor and interaction. A type-II error occurs when you fail to reject the null hypothesis of equal effects when in fact they are different.

Values must be between zero and one. Historically, the value of 0.20 was often used for beta. Now, 0.10 is becoming more common. You should pick a value for beta that represents the risk of a type-II error you are willing to take.

Power is defined as one minus beta. Power is equal to the probability of rejecting a false null hypothesis. Hence, specifying the beta error level also specifies the power level. For example, if you specify a beta value of 0.10, you are specifying the corresponding power value of 0.90.

N1 (Sample Size Group 1)

Enter a value (or range of values) for the sample size of group one. For the one-group case, this is the value of N. For the two-group case, this is the value of $N1$.

You may enter a range of values such as '10 to 100 by 10'.

N2 (Sample Size Group 2)

In the two-group case, enter the value (or range of values) for the sample size of group two. In the one-group case, this value is ignored.

Use R

Enter 'Use R' if you want $N2$ to be calculated using the formula: $N2=[R \text{ x } N1]$ where R is the Sample Allocation Ratio and $[Y]$ is the first integer >= Y. For example, if you want $N1=N2$, select 'Use R' here and set R equal to one.

Covariance Tab

This tab specifies the covariance matrix.

Specify Which Covariance Matrix Input Method to Use

This option specifies which method will be used to define the covariance matrix.

Standard Deviation and Correlation

This option generates a covariance matrix based on the settings for the standard deviation (SD) and the pattern of correlations as specified in the Correlation Pattern and R options.

Covariance Matrix Variables

When this option is selected, the covariance matrix is read in from the columns of the spreadsheet. This is the most flexible method, but specifying a covariance matrix is tedious. You will usually only use this method when a specific covariance is given to you.

Note that the spreadsheet is shown by selecting the menus: 'Window' and then 'Data'.

1) Specify Covariance Matrix Using SD's and Correlations

The parameters in this section provide a flexible way to specify Σ, the covariance matrix. Because the covariance matrix is symmetric, it can be represented as

$$\begin{aligned}
\Sigma &= \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \cdots & \sigma_{pp} \end{bmatrix} \\
&= \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_{12} & \cdots & \sigma_1\sigma_p\rho_{1p} \\ \sigma_1\sigma_2\rho_{12} & \sigma_2^2 & \cdots & \sigma_2\sigma_p\rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_1\sigma_p\rho_{1p} & \sigma_2\sigma_p\rho_{2p} & \cdots & \sigma_p^2 \end{bmatrix} \\
&= \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_p \end{bmatrix} \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{12} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1p} & \rho_{2p} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_p \end{bmatrix}
\end{aligned}$$

where p is the number of response variables.

Thus, the covariance matrix can be represented with complete generality by specifying the standard deviations $\sigma_1, \sigma_2, \cdots, \sigma_p$ and the correlation matrix

$$R = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{12} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1p} & \rho_{2p} & \cdots & 1 \end{bmatrix}.$$

SD (Common Standard Deviation)

This value is used to generate the covariance matrix. This option specifies a single standard deviation to be used for all response variables. The square of this value becomes the diagonal elements of the covariance matrix. Since this is a standard deviation, it must be greater than zero.

This option is only used when the first Covariance Matrix Input Method is selected.

Specify Correlation Pattern

This option specifies the pattern of the correlations in the variance-covariance matrix. Two options are available:

Constant

The value of R is used as the constant correlation. For example, if R = 0.6 and p = 6, the correlation matrix would appear as

$$R = \begin{bmatrix} 1 & 0.600 & 0.600 & 0.600 & 0.600 & 0.600 \\ 0.600 & 1 & 0.600 & 0.600 & 0.600 & 0.600 \\ 0.600 & 0.600 & 1 & 0.600 & 0.600 & 0.600 \\ 0.600 & 0.600 & 0.600 & 1 & 0.600 & 0.600 \\ 0.600 & 0.600 & 0.600 & 0.600 & 1 & 0.600 \\ 0.600 & 0.600 & 0.600 & 0.600 & 0.600 & 1 \end{bmatrix}$$

1st Order Autocorrelation

The value of R is used as the base autocorrelation in a first-order, serial correlation pattern. For example, R = 0.6 and p = 6, the correlation matrix would appear as

$$R = \begin{bmatrix} 1 & 0.600 & 0.360 & 0.216 & 0.130 & 0.078 \\ 0.600 & 1 & 0.600 & 0.360 & 0.216 & 0.130 \\ 0.360 & 0.600 & 1 & 0.600 & 0.360 & 0.216 \\ 0.216 & 0.360 & 0.600 & 1 & 0.600 & 0.360 \\ 0.130 & 0.216 & 0.360 & 0.600 & 1 & 0.600 \\ 0.078 & 0.130 & 0.216 & 0.360 & 0.600 & 1 \end{bmatrix}$$

This pattern is often chosen as the most realistic when little is known about the correlation pattern and the responses variables are measured across time.

R (Correlation)

Specify a correlation to be used in calculating the off-diagonal elements of the covariance matrix. Since this is a correlation, it must be between -1 and 1. This option is only used when the first Covariance Matrix Input Method is selected.

2) Specify Covariance Matrix using Spreadsheet Columns

This option instructs the program to read the covariance matrix from the spreadsheet.

Spreadsheet Columns Containing the Covariance Matrix

This option designates the columns on the current spreadsheet holding the covariance matrix. It is used when the 'Specify Which Covariance Matrix Input Method to Use' option is set to *Covariance Matrix Variables*. The number of columns and number of rows must match the number of response variable at which the subjects are measured.

Reports Tab

This tab specifies which reports and graphs are displayed as well as their format.

Numeric Results

Specify whether to display the numeric report.

Means Matrix

Specify whether to display the mean differences.

Covariance Matrix

Specify whether to display the covariance matrix.

Show Definitions

Specify whether to display a list of definitions for each value in the numeric report.

Show Plot

Specify whether to display this plot.

Report Prob. Dec's - K Dec's

Specify the number of decimal places on these items in the reports.

Summary Statement Rows

The program will output a text statement summarizing the results for each scenario. This option specifies the number of scenarios (rows) from the Numerical Report that will have a summary statement displayed. Select 0 to omit the summary statement.

Example 1 - Power in the One-Group Case and Validation

Rencher (1998) page 106 presents an example of power calculations for the one-group case in which the mean differences are both 1.88 and the covariance matrix is

$$\Sigma = \begin{bmatrix} 56.78 & 11.98 \\ 11.98 & 29.28 \end{bmatrix}$$

When *N* is 25 and the significance level is 0.05, Rencher calculated the power to be 0.3397.

To allow for a nice chart, we will calculate the power for several samples sizes and for *K* equal 1.0 and 1.5.

For your convenience, the covariance matrix has been stored in a spreadsheet called RENCHER2.S0. You must open that spreadsheet to run this example.

Setup

You can enter these values yourself or load the Example1 template from the Template tab.

Option **Value**

Data Tab

Find ....**Beta and Power**
Groups **1**
Response Variables **2**
Mean Differences **1.88 1.88**
Means Column ....*blank*
K...**1.0 1.5**
Alpha **0.05**
Beta ....*Ignored since this is the Find setting*
N1...**5 15 25 35 50 75 100 150**

Covariance Tab

Specify Covariance Method ....**2) Covariance Matrix Variables**
Spreadsheet Columns.......................**VC1-VC2**

Reports Tab

Numeric Results ....**Checked**
Means Matrix.....................................**Checked**
Covariance Matrix ....**Checked**
Show Definitions ....**Checked**
Show Plot ....**Checked**
Summary Statement Rows................**1**

Annotated Output

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

Numeric Report

Power	N	Multiply Means By	Alpha	Beta	Effect Size	Number of Y's (DF1)	DF2
0.0737	5	1.0000	0.0500	0.9263	0.38	2	3
0.1996	15	1.0000	0.0500	0.8004	0.38	2	13
0.3379	25	1.0000	0.0500	0.6621	0.38	2	23
0.4707	35	1.0000	0.0500	0.5293	0.38	2	33
0.6409	50	1.0000	0.0500	0.3591	0.38	2	48
0.8311	75	1.0000	0.0500	0.1689	0.38	2	73
0.9282	100	1.0000	0.0500	0.0718	0.38	2	98
0.9895	150	1.0000	0.0500	0.0105	0.38	2	148
0.1040	5	1.5000	0.0500	0.8960	0.38	2	3
0.4033	15	1.5000	0.0500	0.5967	0.38	2	13
0.6635	25	1.5000	0.0500	0.3365	0.38	2	23
0.8302	35	1.5000	0.0500	0.1698	0.38	2	33
0.9475	50	1.5000	0.0500	0.0525	0.38	2	48
0.9943	75	1.5000	0.0500	0.0057	0.38	2	73
0.9995	100	1.5000	0.0500	0.0005	0.38	2	98
1.0000	150	1.5000	0.0500	0.0000	0.38	2	148

Report Definitions
Power is the probability of rejecting a false null hypothesis. Note that Power = 1 - Beta.
N is the sample size, the number of subjects in the experiment or study.
K is a constant by which all means are multiplied.
Alpha is the probability of rejecting a true null hypothesis.
Beta is the probability of accepting a false null hypothesis. Note that Beta = 1 - Power.
Effect Size is a standardized version of T2 under the alternative hypothesis.
DF1 is the first degrees of freedom of T2. It is the number of response variables.
DF2 is the second degrees of freedom of T2.

Summary Statements
A sample size of 5 achieves 7% power to detect an effect size of 0.38 which represents the differences between the null and alternative means of the 2 response variables, adjusted by the variance-covariance matrix. The one-sample Hotelling's T-squared test statistic is used with a significance level of 0.0500.

This report gives the power for each value of *N* and *K*. Notice that the power for *K* = 1 and *N* = 25 is 0.3379. This is slightly different than the 0.3397 obtained by interpolation by Rencher.

Means Matrix

Means Matrix Section

Name	Mean
Y1	1.8800
Y2	1.8800

This report shows the mean differences that were read in. When a Means Multiplier, *K*, is used, each value of *K* is multiplied times each of these values.

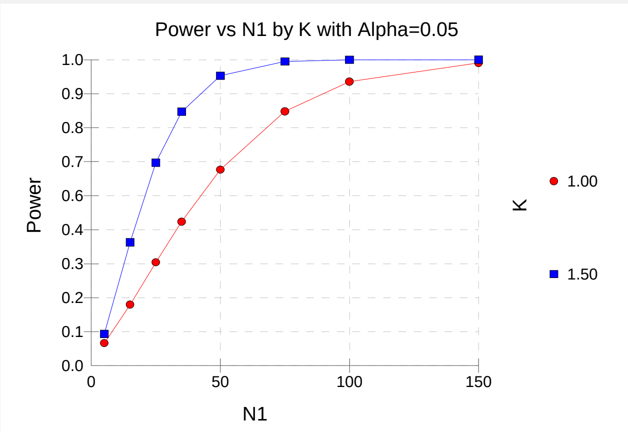
Variance-Covariance Matrix Section

Variance-Covariance Matrix Section

Response	Y1	Y2
Y1	7.5353	0.2938
Y2	0.2938	5.4111

This report shows the variance-covariance matrix that was read in from the spreadsheet or generated by the settings of on the Covariance tab. The standard deviations are given on the diagonal and the correlations are given off the diagonal.

Chart Section



This chart shows the relationship between power and *N* for each value of *K*.

Example2 - Power in the Two-Group Case and Validation

Rencher (1998) pages 107-108 presents an example of power calculations for the two-group case in which the mean differences and covariance matrix are

$$\mu_1 - \mu_2 = \begin{bmatrix} 3 \\ -2 \\ 3 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 6 & -3 & 3 \\ -3 & 5 & -6 \\ 3 & -6 & 9 \end{bmatrix}$$

When $N1 = N2 =$ 10, 12, 14, 16 and the significance level is 0.05, Rencher calculated the power to be 0.6438, 0.7520, 0.8329, 0.8936, respectively.

For your convenience, the mean differences and covariance matrix have been stored in a spreadsheet called RENCHER2.S0. You must open that spreadsheet to run this example.

Setup

You can enter these values yourself or load the Example2 template from the Template tab.

Option	Value
Data Tab	
Find	**Beta and Power**
Groups	**2**
Response Variables	**3**
Mean Differences	*blank*
Means Column	**Differences**
K	**1.0**
Alpha	**0.05**
Beta	*Ignored since this is the Find setting*
N1	**10 12 14 16**
N2	**Use R**
R	**1.0**
Covariance Tab	
Specify Covariance Method	**2) Covariance Matrix Variables**
Spreadsheet Columns	**VC_1-VC_3**
Reports Tab	
Numeric Results	**Checked**
Means Matrix	**Checked**
Covariance Matrix	**Checked**
Show Definitions	**Checked**
Show Plot	**Checked**
Summary Statement Rows	**1**

Annotated Output

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

Numeric Report

Power	N	Multiply Means By	Alpha	Beta	Effect Size	Number of Y's (DF1)	DF2	
0.6442	10	10	1.0000	0.0500	0.3558	1.41	3	16
0.7546	12	12	1.0000	0.0500	0.2454	1.41	3	20
0.8361	14	14	1.0000	0.0500	0.1639	1.41	3	24
0.8936	16	16	1.0000	0.0500	0.1064	1.41	3	28

Note that the power values obtained here are very close to those obtained by Rencher. We feel that our results are more accurate since Rencher’s results were obtained by interpolation from Tang’s tables.

Chapter 605

MANOVA

Introduction

This module calculates power for multivariate analysis of variance (MANOVA) designs having up to three factors. It computes power for three MANOVA test statistics: Wilks' lambda, Pillai-Bartlett trace, and Hotelling-Lawley trace.

MANOVA is an extension of common analysis of variance (ANOVA). In ANOVA, differences among various group means on a single-response variable are studied. In MANOVA, the number of response variables is increased to two or more. The hypothesis concerns a comparison of vectors of group means. The multivariate extension of the F-test is not completely direct. Instead, several test statistics are available. The actual distributions of these statistics are difficult to calculate, so we rely on approximations based on the F-distribution.

Assumptions

The following assumptions are made when using MANOVA to analyze a factorial experimental design.

1. The response variables are continuous.
2. The residuals follow the multivariate normal probability distribution with mean zero and constant variance-covariance matrix.
3. The subjects are independent.

Technical Details

General Linear Multivariate Model

This section provides the technical details of the MANOVA designs that can be analyzed by ***PASS***. The approximate power calculations outlined in Muller, LaVange, Ramey, and Ramey (1992) are used. Using their notation, for N subjects, the usual general linear multivariate model is

$$\underset{(N\times p)}{Y} = \underset{(N\times q\times p)}{XM} + \underset{(N\times p)}{R}$$

where each row of the residual matrix R is distributed as a multivariate normal

$$row_k(R) \sim N_p(0,\Sigma)$$

Note that p is the number of response variables and q is the number of design variables, Y is the matrix of responses, X is the design matrix, M is the matrix of regression parameters (means), and R is the matrix of residuals.

Hypotheses about various sets of regression parameters are tested using

$$H_0: \underset{a\times p}{\Theta} = \Theta_0$$

$$\underset{a\times q\times p}{CM} = \Theta$$

where C is an orthonormal contrast matrix and Θ_0 is a matrix of hypothesized values, usually zeros. Note that C defines contrasts among the factor levels. Tests of the various main effects and interactions may be constructed with suitable choices of C. These tests are based on

$$\hat{M} = (X'X)^{-} X'Y$$

$$\hat{\Theta} = C\hat{M}$$

$$\underset{p\times p}{H} = \left(\hat{\Theta} - \Theta_0\right)' \left[C(X'X)^{-} C'\right]^{-1} \left(\hat{\Theta} - \Theta_0\right)$$

$$\underset{p\times p}{E} = \hat{\Sigma} \cdot (N - r)$$

$$\underset{p\times p}{T} = H + E$$

where r is the rank of X.

Wilks' Lambda Approximate F Test

The hypothesis $H_0 : \Theta = \Theta_0$ may be tested using Wilks' likelihood ratio statistic *W*. This statistic is computed using

$$W = \left| ET^{-1} \right|$$

An *F* approximation to the distribution of *W* is given by

$$F_{df_1, df_2} = \frac{\eta / df_1}{(1-\eta) / df_2}$$

where

$$\lambda = df_1 F_{df_1, df_2}$$

$$\eta = 1 - W^{1/g}$$

$$df1 = ap$$

$$df2 = g\left[(N-r) - (p-a+1)/2\right] - (ap-2)/2$$

$$g = \left(\frac{a^2 p^2 - 4}{a^2 + p^2 - 5} \right)^{\frac{1}{2}}$$

Pillai-Bartlett Trace Approximate F Test

The hypothesis $H_0:\Theta = \Theta_0$ may be tested using the Pillai-Bartlett Trace. This statistic is computed using

$$T_{PB} = tr\left(HT^{-1}\right)$$

A noncentral *F* approximation to the distribution of T_{PB} is given by

$$F_{df_1,df_2} = \frac{\eta / df_1}{\left(1-\eta\right) / df_2}$$

where

$$\lambda = df_1 F_{df_1,df_2}$$

$$\eta = \frac{T_{PB}}{s}$$

$$s = \min\left(a, p\right)$$

$$df1 = ap$$

$$df2 = s\left[\left(N - r\right) - p + s\right]$$

Hotelling-Lawley Trace Approximate F Test

The hypothesis $H_0: \Theta = \Theta_0$ may be tested using the Hotelling-Lawley Trace. This statistic is computed using

$$T_{HL} = tr\left(HE^{-1}\right)$$

An F approximation to the distribution of T_{HL} is given by

$$F_{df_1, df_2} = \frac{\eta / df_1}{(1-\eta) / df_2}$$

where

$$\lambda = df_1 F_{df_1, df_2}$$

$$\eta = \frac{\frac{T_{HL}}{s}}{1 + \frac{T_{HL}}{s}}$$

$$s = \min(a, p)$$

$$df1 = ap$$

$$df2 = s\left[(N - r) - p + s\right]$$

M (Mean) Matrix

In the general linear multivariate model presented above, *M* represents a matrix of regression coefficients. Although other structures and interpretations of *M* are possible, in this module we assume that the elements of *M* are the cell means. The rows of *M* represent the factor categories and the columns of *M* represent the response variables. (Note that this is just the opposite of the orientation used when entering *M* into the spreadsheet.)

The *q* rows of *M* represent the *q* groups into which the subjects can be classified. For example, if a design includes three factors with 2, 3, and 4 categories, the matrix *M* would have 2 *x* 3 *x* 4 = 24 rows. That is, $q = 24$.

Consider now an example in which $q = 3$ and $p = 4$. That is, there are three groups into which subjects can be placed. Each subject has four made. The matrix *M* would appear as follows.

$$M = \begin{bmatrix} \mu_{11} & \mu_{12} & \mu_{13} & \mu_{14} \\ \mu_{21} & \mu_{22} & \mu_{23} & \mu_{24} \\ \mu_{31} & \mu_{32} & \mu_{33} & \mu_{34} \end{bmatrix}$$

For example, the element μ_{12} is the mean of the second response of subjects in the first group. To calculate the power of this design, you would need to specify appropriate values of all twelve means.

C Matrix - Contrasts

The *C* matrix is comprised of contrasts that are applied to the rows of *M*. You do not have to specify these contrasts. They are generated for you. You should understand that a different *C* matrix is generated for each term in the model.

Generating the C Matrix when there are Multiple Between Factors.

Generating the *C* matrix when there is more than one factor is more difficult. We use the method of O'Brien and Kaiser (1985) which we briefly summarize here.

Step 1. Write a complete set of contrasts suitable for testing each factor separately. For example, if you have three factors with 2, 3, and 4 categories, you might use

$$\ddot{C}_{B1} = \begin{bmatrix} \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix},\ \ddot{C}_{B2} = \begin{bmatrix} \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ 0 & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}, \text{ and } \ddot{C}_{B3} = \begin{bmatrix} \frac{-3}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} \\ 0 & \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ 0 & 0 & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}.$$

Step 2. Define appropriate J_k matrices corresponding to each factor. These matrices comprised of one row and *k* columns whose equal element is chosen so that the sum of its elements squared is one. In this example, we use

$$J_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix},\ J_3 = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix},\ J_4 = \begin{bmatrix} \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} \end{bmatrix}$$

Step 3. Create the appropriate contrast matrix using a direct (Kronecker) product of either the $\ddot{C}_{Bi}$ matrix if the factor is included in the term or the J_i matrix when the factor is not in the term. Remember that the direct product is formed by multiplying each element of the second matrix by all members of the first matrix. Here is an example

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 & -1 \\ 0 & 2 & 0 \\ -1 & 0 & 3 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 2 & 0 & 0 & -1 & -2 \\ 3 & 4 & 0 & 0 & -3 & -4 \\ 0 & 0 & 2 & 4 & 0 & 0 \\ 0 & 0 & 6 & 8 & 0 & 0 \\ -1 & -2 & 0 & 0 & 3 & 6 \\ -3 & -4 & 0 & 0 & 9 & 12 \end{bmatrix}$$

As an example, we will compute the C matrix suitable for testing factor $B2$

$$C_{B2} = J_2 \otimes \ddot{C}_{B2} \otimes J_4$$

Expanding the direct product results in

$$C_{B2} = J_2 \otimes \ddot{C}_{B2} \otimes J_4$$

$$= \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \otimes \begin{bmatrix} \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ 0 & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \otimes \begin{bmatrix} \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{-2}{\sqrt{12}} & \frac{-2}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} \\ 0 & 0 & \frac{-1}{\sqrt{4}} & \frac{-1}{\sqrt{4}} & \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} \end{bmatrix} \otimes \begin{bmatrix} \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{-2}{\sqrt{48}} & \frac{-2}{\sqrt{48}} & \frac{1}{\sqrt{48}} & \frac{1}{\sqrt{48}} & \frac{1}{\sqrt{48}} & \frac{1}{\sqrt{48}} & \frac{-2}{\sqrt{48}} & \frac{-2}{\sqrt{48}} & \frac{1}{\sqrt{48}} & \frac{1}{\sqrt{48}} & \frac{1}{\sqrt{48}} & \frac{1}{\sqrt{48}} & \frac{-2}{\sqrt{48}} & \frac{-2}{\sqrt{48}} & \frac{1}{\sqrt{48}} & \frac{1}{\sqrt{48}} & \frac{1}{\sqrt{48}} & \frac{1}{\sqrt{48}} & \frac{-2}{\sqrt{48}} & \frac{-2}{\sqrt{48}} & \frac{1}{\sqrt{48}} & \frac{1}{\sqrt{48}} & \frac{1}{\sqrt{48}} & \frac{1}{\sqrt{48}} \\ 0 & 0 & \frac{-1}{\sqrt{16}} & \frac{-1}{\sqrt{16}} & \frac{1}{\sqrt{16}} & \frac{1}{\sqrt{16}} & 0 & 0 & \frac{-1}{\sqrt{16}} & \frac{-1}{\sqrt{16}} & \frac{1}{\sqrt{16}} & \frac{1}{\sqrt{16}} & 0 & 0 & \frac{-1}{\sqrt{16}} & \frac{-1}{\sqrt{16}} & \frac{1}{\sqrt{16}} & \frac{1}{\sqrt{16}} & 0 & 0 & \frac{-1}{\sqrt{16}} & \frac{-1}{\sqrt{16}} & \frac{1}{\sqrt{16}} & \frac{1}{\sqrt{16}} \end{bmatrix}$$

Similarly, the C matrix suitable for testing interaction $B2B3$ is

$$C_{B2B3} = J_2 \otimes \ddot{C}_{B2} \otimes \ddot{C}_{B3}$$

We leave the expansion of this matrix *PASS*, but we think you have the idea.

Power Calculations

To calculate statistical power, we must determine distribution of the test statistic under the alternative hypothesis which specifies a different value for the regression parameter matrix *B*. The distribution theory in this case has not been worked out, so approximations must be used. We use the approximations given by Mueller and Barton (1989) and Muller, LaVange, Ramey, and Ramey (1992). These approximations state that under the alternative hypothesis, F_U is distributed as a noncentral *F* random variable with degrees of freedom and noncentrality shown above. The calculation of the power of a particular test may be summarized as follows.

1. Specify values of X, M, Σ, C, and Θ_0.
2. Determine the critical value using $F_{crit} = FINV\left(1-\alpha, df1, df2\right)$, where *FINV*( ) is the inverse of the central *F* distribution and α is the significance level.
3. Compute the noncentrality parameter λ.
4. Compute the power as

$$Power = 1 - NCFPROB\left(F_{crit}, df1, df2, \lambda\right)$$

where *NCFPROB*( ) is the noncentral *F* distribution.

Procedure Options

This section describes the options that are unique to this procedure. To find out more about using the other tabs, turn to the chapter entitled Procedure Templates.

Data Tab

The Data tab contains many of the options that you will be primarily concerned with.

Find (Solve For)

This option specifies the parameter to be solved for. If you choose to solve for *n* (sample size), you must also specify which test statistic you want to use.

When you choose to solve for *n*, the program searches for the lowest sample size that meets the alpha and beta criterion you have specified for each of the terms. If you do not want a term to be used in the search, set its alpha and beta values to 0.99.

Also, when the '= n's' box is not checked, the search is made using unequal group sample sizes. The relative proportion of the sample in each group is set by the values of *n* given in the Subjects Per Group box. For example, if your design has three groups and you entered '1 1 2' in the Subjects Per Group box, the search will only consider designs in which the size of the last group is twice the rest. That is, it will consider '2 2 4', '3 3 6', '4 4 8', etc.

Note: no plots are generated when you solve for *n*.

n (Subjects Per Group)

Specify one or more values for the number of subjects per group. The total sample size is the sum of the individual group sizes across all groups.

You can specify a list like '2 4 6'. The items in the list may be separated with commas or blanks. The interpretation of the list depends on the =n's check box. When the =n's box is checked, a separate analysis is calculated for each value of *n*. When the =n's box is not checked, ***PASS*** uses the *n's* as the actual group sizes. In this case, the number of items entered must match the number of groups in the design.

When you choose to solve for *n* and the = n's box is not checked, the search is made using unequal group sample sizes. The relative proportion of the sample in each group is set by the values of *n* given in this box. For example, if your design has three groups and you enter '1 1 2' here, the search will only consider designs in which the size of the last group is twice the rest. That is, it will consider '2 2 4', '3 3 6', '4 4 8', etc.

= n's

This option controls whether the number of subjects per group is to be equal for all groups or not. When checked, the number of subjects per group is equal for all groups. A list of values such as '5 10 15' represents three designs: one with five per group, one with ten per group, and one with fifteen per group.

When this option is not checked, the *n*'s are assumed to be unequal. A list of values represents the size of the individual groups. For example, '5 10 15' represents a single, three-group design with five in the first group, ten in the second group, and fifteen in the third group.

Means Matrix

Use this option to the specify spreadsheet columns containing a hypothesized means matrix that is used to specify the alternative hypothesis. You can obtain the spreadsheet by selecting ‘Window’, then ‘Data’, from the menus.

The factors are represented across the columns of the spreadsheet and the response variables are represented down the rows. The number of columns specified must equal the number of groups. The number of rows with data in these columns must equal the number of response variables. For example, suppose you are designing an experiment that is to have two factors (A and B) and two response variables (Y1 and Y2). Suppose each of the factors has two levels. The four columns of the spreadsheet would represent

A1B1 A1B2 A2B1 A2B2.

The two rows of the spreadsheet would represent

Y1

Y2

K (Means Multipliers)

These values are multiplied times the means matrix to give you various effect sizes. A separate power calculation is generated for each value of K. These values become the horizontal axis in the second power chart. If you want to ignore this setting, enter ‘1’.

Labels

Specify a label for this factor. Although we suggest that only a single letter be used, the label can consist of several letters. When several letters are used, the labels for the interactions may be extra long and confusing. Of course, you must be careful not to use the same label for two factors.

One of the easiest sets of labels is to use A, B, and C for the factors.

Levels

Specify the number of levels (categories) in this factor. Typical values are from 2 to 8. Set this to a blank (or 0) to ignore the factor in the design.

Alpha

These options specify the probability of a type-I error (alpha) for each factor and interaction. A type-I error occurs when you reject the null hypothesis of zero effects when in fact they are zero. Since they are probabilities, alpha values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This value may be interpreted as meaning that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You can specify different alpha values for different terms. For example, although you have three terms in an experiment, you might be mainly interested in only one of them. Hence, you could increase the alpha level of the tests of the other terms and thereby increase their power. Also, you may want to increase the alpha level of the interaction terms, as these will often have poor power otherwise.

Beta

These options specify the probability of a type-II error (beta) for each factor and interaction. A type-II error occurs when you fail to reject the null hypothesis of equal effects when in fact they are different.

Values must be between zero and one. Historically, the value of 0.20 was often used for beta. Now, 0.10 is becoming more common. You should pick a value for beta that represents the risk of a type-II error you are willing to take.

Power is defined as one minus beta. Power is equal to the probability of rejecting a false null hypothesis. Hence, specifying the beta error level also specifies the power level. For example, if you specify a beta value of 0.10, you are specifying the corresponding power value of 0.90.

Response Variables

Enter the number of response variables in your design. For a true MANOVA, this value must be greater than one. The number of rows in the means matrix must equal this value. If you specify a covariance matrix, the number of columns specified must equal this value.

Covariance Tab

This tab specifies the covariance matrix.

Specify Which Covariance Matrix Input Method to Use

This option specifies which method will be used to define the covariance matrix.

Standard Deviation and Correlation

This option generates a covariance matrix based on the settings for the standard deviation (SD) and the pattern of correlations as specified in the Correlation Pattern and R options.

Covariance Matrix Variables

When this option is selected, the covariance matrix is read in from the columns of the spreadsheet. This is the most flexible method, but specifying a covariance matrix is tedious. You will usually only use this method when a specific covariance is given to you.

Note that the spreadsheet is shown by selecting the menus: ‘Window’ and then ‘Data’.

1) Specify Covariance Matrix Using SD’s and Correlations

The parameters in this section provide a flexible way to specify Σ , the covariance matrix. Because the covariance matrix is symmetric, it can be represented as

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \cdots & \sigma_{pp} \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_{12} & \cdots & \sigma_1\sigma_p\rho_{1p} \\ \sigma_1\sigma_2\rho_{12} & \sigma_2^2 & \cdots & \sigma_2\sigma_p\rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_1\sigma_p\rho_{1p} & \sigma_2\sigma_p\rho_{2p} & \cdots & \sigma_p^2 \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_p \end{bmatrix} \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{12} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1p} & \rho_{2p} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_p \end{bmatrix}$$

where p is the number of response variables.

Thus, the covariance matrix can be represented with complete generality by specifying the standard deviations $\sigma_1, \sigma_2, \cdots, \sigma_p$ and the correlation matrix

$$R = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{12} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1p} & \rho_{2p} & \cdots & 1 \end{bmatrix}.$$

SD (Standard Deviation)

This value is used to generate the covariance matrix. This option specifies a standard deviation to be used for all response variables. The square of this value becomes the diagonal elements of the covariance matrix. Since this is a standard deviation, it must be greater than zero.

This option is only used when the first Covariance Matrix Input Method is selected.

Specify Correlation Pattern

This option specifies the pattern of the correlations in the variance-covariance matrix. Two options are available:

Constant

The value of R is used as the constant correlation. For example, if R = 0.6 and *p* = 6, the correlation matrix would appear as

$$R = \begin{bmatrix} 1 & 0.600 & 0.600 & 0.600 & 0.600 & 0.600 \\ 0.600 & 1 & 0.600 & 0.600 & 0.600 & 0.600 \\ 0.600 & 0.600 & 1 & 0.600 & 0.600 & 0.600 \\ 0.600 & 0.600 & 0.600 & 1 & 0.600 & 0.600 \\ 0.600 & 0.600 & 0.600 & 0.600 & 1 & 0.600 \\ 0.600 & 0.600 & 0.600 & 0.600 & 0.600 & 1 \end{bmatrix}$$

1st Order Autocorrelation

The value of R is used as the base autocorrelation in a first-order, serial correlation pattern. For example, R = 0.6 and *p* = 6, the correlation matrix would appear as

$$R = \begin{bmatrix} 1 & 0.600 & 0.360 & 0.216 & 0.130 & 0.078 \\ 0.600 & 1 & 0.600 & 0.360 & 0.216 & 0.130 \\ 0.360 & 0.600 & 1 & 0.600 & 0.360 & 0.216 \\ 0.216 & 0.360 & 0.600 & 1 & 0.600 & 0.360 \\ 0.130 & 0.216 & 0.360 & 0.600 & 1 & 0.600 \\ 0.078 & 0.130 & 0.216 & 0.360 & 0.600 & 1 \end{bmatrix}$$

This pattern is often chosen as the most realistic when little is known about the correlation pattern and the responses variables are measured across time.

R (Correlation)

Specify a correlation to be used in calculating the off-diagonal elements of the covariance matrix. Since this is a correlation, it must be between -1 and 1. This option is only used when the first Covariance Matrix Input Method is selected.

2) Specify Covariance Matrix using Spreadsheet Columns

This option instructs the program to read the covariance matrix from the spreadsheet.

Spreadsheet Columns Containing the Covariance Matrix

This option designates the columns on the current spreadsheet holding the covariance matrix. It is used when the ‘Specify Which Covariance Matrix Input Method to Use’ option is set to *Covariance Matrix Variables*. The number of columns and number of rows must match the number of response variable at which the subjects are measured.

Reports Tab

This tab specifies which reports and graphs are displayed as well as their format.

Skip Line After

The names of the terms can be too long to fit in the space provided. If the name contains more characters than this, the rest of the output is placed on a separate line. Enter ‘1’ when you want every term’s results printed on two lines. Enter ‘100’ when you want every variable’s results printed on one line.

Test in Summary Statement

Indicate the test that is to be reported on in the Summary Statements.

Maximum Term-Order Reported

Indicate the maximum order of terms to be reported on. Occasionally, higher-order interactions are of little interest and so they may be omitted. For example, enter a ‘2’ to limit output to individual factors and two-way interactions.

Example 1 - Determining Power

Researchers are planning a study of the impact of a drug. They want to evaluate the differences in heart rate and blood pressure among three age groups: 20-40, 41-60, and over 60. They want to be able to detect a 10% change in heart rate and in blood pressure among the age groups. They plan to analyze the data using Wilks' lambda.

Previous studies have found an average heart rate of 93 with a standard deviation of 4 and an average blood pressure of 130 with a standard deviation of 5. The correlation between the two responses will be set at 0.7.

From a heart rate of 93, a 10% reduction gives 84. They want to be able to detect age-group heart-rate means the range from 93 to 84. From a blood pressure of 130, a 10% reduction gives 117. The want to be able to detect age-group blood-pressure means that range from 130 to 117. Hence, the means matrix that they will use is

C1	C2	C3
93	88	84
130	124	117

Base on the standard deviation settings that they chose to use, the covariance matrix will be

C4	C5
16	14
14	25

In order to understand the relationship between power and sample size, they decide to calculate power values for sample sizes between 2 and 12, using a 0.05 significance level.

For your convenience, the Means Matrix and Covariance Matrix have been stored in a spreadsheet called PASSMANOVA1.S0. You can enter the above values yourself or you can open that spreadsheet.

Setup

You can enter these values yourself or load the Example1 template from the Template tab.

Option	Value
Data Tab	
Find	**Beta and Power**
Means Matrix	**C1-C3**
n	**2 4 6 8 10 12**
=n's	**checked**
K	**0.5 1 1.5**
Response Variables	**2**
For Factor F1	
Label	**A**
Levels	**3**
Alpha	**0.05**
Beta	*Ignored since this is the Find setting*

Covariance Tab
Specify Covariance Method .... **2) Covariance Matrix Variables**
Spreadsheet Columns **C4-C5**
Reports Tab
Numeric Results by Term .... **Checked**
Numeric Results by Design............... **Not Checked**
Wilks' Lambda................................... **Checked**
Pillai-Bartlett...................................... **Not checked**
Hotelling-Lawley................................ **Not checked**
Means Matrix .... **Checked**
Covariance Matrix .... **Checked**
Show Plot 1....................................... **Checked**
Show Plot 2....................................... **Checked**
Test in Summary Statement .... **Wilks Lambda**
Max Term-Order Reported................ **2**
Plot Setup Tab
Max Term-Order Plotted **2**
Test That is Plotted........................... **Wilks Lambda**

Annotated Output

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

Term Report

Results for Factor A (Levels = 3)

Test	Power	n	N	Multiply Means By	Test Statistic	Approx. F Statistic	DF1|DF2	Alpha	Beta
Wilks	0.0729	2	6	0.50	0.2115	0.27	4|4	0.0500	0.9271
Wilks	0.1291	2	6	1.00	0.4654	0.87	4|4	0.0500	0.8709
Wilks	0.2046	2	6	1.50	0.6185	1.62	4|4	0.0500	0.7954
Wilks	0.1888	4	12	0.50	0.1562	0.74	4|16	0.0500	0.8112
Wilks	0.5749	4	12	1.00	0.3853	2.51	4|16	0.0500	0.4251
Wilks	0.8722	4	12	1.50	0.5443	4.78	4|16	0.0500	0.1278
Wilks	0.3191	6	18	0.50	0.1437	1.17	4|28	0.0500	0.6809
Wilks	0.8548	6	18	1.00	0.3648	4.02	4|28	0.0500	0.1452
Wilks	0.9916	6	18	1.50	0.5241	7.71	4|28	0.0500	0.0084
Wilks	0.4488	8	24	0.50	0.1382	1.60	4|40	0.0500	0.5512
Wilks	0.9603	8	24	1.00	0.3554	5.51	4|40	0.0500	0.0397
Wilks	0.9997	8	24	1.50	0.5147	10.61	4|40	0.0500	0.0003
Wilks	0.5678	10	30	0.50	0.1351	2.03	4|52	0.0500	0.4322
Wilks	0.9907	10	30	1.00	0.3500	7.00	4|52	0.0500	0.0093
Wilks	1.0000	10	30	1.50	0.5092	13.49	4|52	0.0500	0.0000
Wilks	0.6704	12	36	0.50	0.1331	2.46	4|64	0.0500	0.3296
Wilks	0.9981	12	36	1.00	0.3465	8.48	4|64	0.0500	0.0019
Wilks	1.0000	12	36	1.50	0.5057	16.37	4|64	0.0500	0.0000

Summary Statements

A MANOVA design with 1 factor and 2 response variables has 3 groups with 2 subjects each for a total of 6 subjects. This design achieves 7% power to test factor A if a Wilks' Lambda Approximate F Test is used with a 5% significance level.

This report gives the power for each value of *n* and *K*. It is useful when you want to compare the powers of the terms in the design at a specific sample size.

In this example, for $K = 1$, the design goal of 0.95 power is achieved for $n = 8$.

The definitions of each of the columns of the report are as follows.

Test

This column identifies the test statistic. Since the power depends on the test statistic, you should make sure that this is the test statistic that you will use in your analysis.

Power

This is the computed power for the term.

n

The value of *n* is the number of subjects per group.

N

The value of *N* is the total number of subjects in the study.

Multiply Means By

This is the value of the means multiplier, *K*.

Test Statistic

This is the value of the test statistic computed at the hypothesized values. The name of the statistic is identified in the Test column. Possible values are Wilks’ lambda, Pillai-Bartlett trace, or Hotelling-Lawley trace. The actual formulas used were given earlier in the Technical Details section.

Approx. F Statistic

This is the value of the *F* statistic that is used to compute the probability levels. This value is calculated using the hypothesized values. The actual formulas used were given earlier in the Technical Details section.

DF1|DF2

These are the numerator and denominator degrees of freedom of the approximating *F* distribution.

Alpha

Alpha is the significance level of the test.

Beta

Beta is the probability of failing to reject the null hypothesis when the alternative hypothesis is true.

Means Matrix

Means Matrix Section

Name	A1	A2	A3
Y1	93.00	88.00	84.00
Y2	130.00	124.00	117.00

This report shows the means matrix that was read in. It may be used to get an impression of the magnitude of the difference among the means that is being studied. When a Means Multiplier, *K*, is used, each value of *K* is multiplied times each value of this matrix.

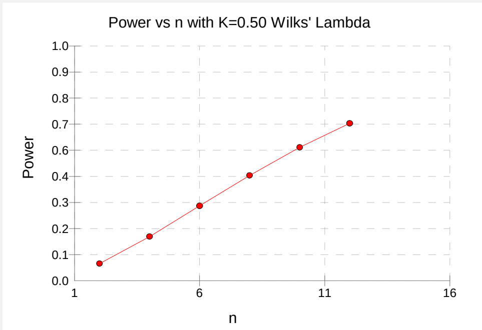
Variance-Covariance Matrix Section

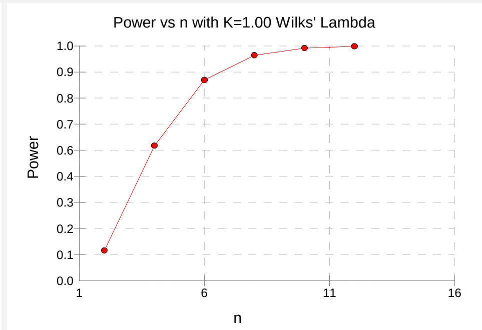
Variance-Covariance Matrix Section

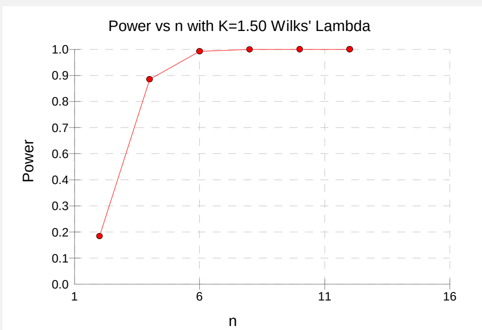
Response	Y1	Y2
Y1	4.00	0.70
Y2	0.70	5.00

This report shows the variance-covariance matrix that was read in from the spreadsheet or generated by the settings of on the Covariance tab. The standard deviations are given on the diagonal and the correlations are given off the diagonal.

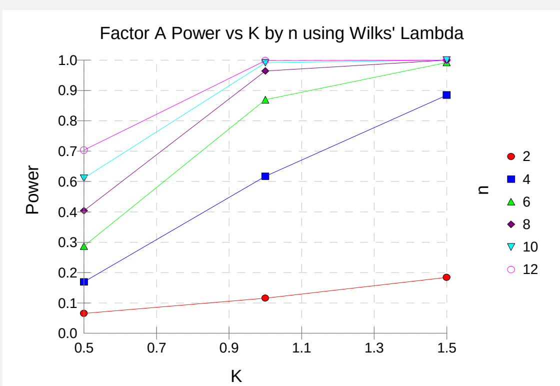
Chart Section







These chart show the relationship between power and *n* for each value of *K*. Note that high-order interactions may be omitted from the plot by reducing the Max Term-Order Plotted option on the Plot Setup tab.



These charts show the relationship between power and *K* for each value of *n*. Remember that *K* is the mean multiplier. It changes the effect size.

Example 2 - Validation

In this example, we will set p = 2, q = 3, alpha = 0.05, and n = 4. The mean and covariance matrices are

$$M = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 2 \end{bmatrix} \text{ and } \Sigma = \begin{bmatrix} 4 & 1 \\ 1 & 4 \end{bmatrix}$$

The contrast matrix C is

$$C = \begin{bmatrix} \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ 0 & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

The $X'X$ matrix is

$$X'X = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix}$$

The matrix Θ is

$$\begin{aligned} \Theta &= CM \\ &= \begin{bmatrix} \frac{3}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \end{aligned}$$

The matrix H is

$$\begin{aligned} H &= \left(\hat{\Theta} - \Theta_0\right)' \left[C\left(X'X\right)^{-} C'\right]^{-1} \left(\hat{\Theta} - \Theta_0\right) \\ &= \begin{bmatrix} 8 & 4 \\ 4 & 8/3 \end{bmatrix} \end{aligned}$$

The matrix E is

$$\begin{aligned} E &= \hat{\Sigma} \cdot \left(N - r\right) \\ &= \begin{bmatrix} 36 & 9 \\ 9 & 36 \end{bmatrix} \end{aligned}$$

The matrix T is

$$\begin{aligned} T &= H + E \\ &= \begin{bmatrix} 44 & 13 \\ 13 & 38\tfrac{2}{3} \end{bmatrix} \end{aligned}$$

Using these matrices, we can calculate the values of the test statistics. We will only calculate the results for Wilks' lambda. We have

$$W = \det\left(ET^{-1}\right)$$
$$= 0.79290842$$
$$a = q - 1$$
$$= 2$$
$$g = \left(\frac{a^2 p^2 - 4}{a^2 + p^2 - 5}\right)^{\frac{1}{2}}$$
$$= \left(\frac{2^2 2^2 - 4}{2^2 + 2^2 - 5}\right)^{\frac{1}{2}}$$
$$= 2$$
$$\eta = 1 - W^{1/g}$$
$$= 1 - \sqrt{0.79290842}$$
$$= 0.10954595$$
$$df1 = ap$$
$$= 4$$
$$df2 = g\left[\left(N - r\right) - \left(p - a + 1\right)/2\right] - \left(ap - 2\right)/2$$
$$= 2\left[\left(12 - 3\right) - \left(2 - 2 + 1\right)/2\right] - \left(4 - 2\right)/2$$
$$= 16$$
$$F_{df_1, df_2} = \frac{\eta / df_1}{\left(1 - \eta\right)/df_2}$$
$$= \frac{0.10954595/4}{\left(1 - 0.10954595\right)/16}$$
$$= 0.49209030$$
$$\lambda = df_1 F_{df_1, df_2}$$
$$= 4\left(0.49209030\right)$$
$$= 1.96836120$$

For an F with 4 and 16 degrees of freedom, the 5% critical value is 3.0069172799. Finally, compute the power using a noncentral F with 4 and 16 degrees of freedom and noncentrality parameter

$$Power = \Pr\left(f > F | df_1 = 4, df_2 = 16, \lambda = 1.96836120\right)$$
$$= 0.1370631884$$

In order to run this example in ***PASS***, the values of the means and the covariance matrix (given above) must be entered on a spreadsheet. We have loaded these values into the database called PASSMANOVA2. Either enter the values yourself, or load the PASSMANOVA2 database which should be in the Data directory. The instructions below assume that the means are in columns C1-C3, while the covariance matrix is in columns C4-C5.

Setup

You can enter these values yourself or load the Example2 template from the Template tab.

Option	Value
Data Tab	
Find	**Beta and Power**
Means Matrix	**C1-C3**
n	**4**
=n's	**checked**
K	**1**
Response Variables	**2**
For Factor F1	
Label	**A**
Levels	**3**
Alpha	**0.05**
Beta	*Ignored since this is the Find setting*
Covariance Tab	
Specify Covariance Method	**2) Covariance Matrix Variables**
Spreadsheet Columns	**C4-C5**
Reports Tab	
Numeric Results by Term	**Checked**
Numeric Results by Design	**Not Checked**
Wilks' Lambda	**Checked**
Pillai-Bartlett	**Not checked**
Hotelling-Lawley	**Not checked**
Means Matrix	**Checked**
Covariance Matrix	**Checked**

Annotated Output

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

Term Report

Results for Factor A (Levels = 3)

Test	Power	n	N	Multiply Means By	Test Statistic	Approx. F Statistic	DF1|DF2	Alpha	Beta
Wilks	0.1371	4	12	1.0000	0.7929	0.4921	4|16	0.0500	0.8629

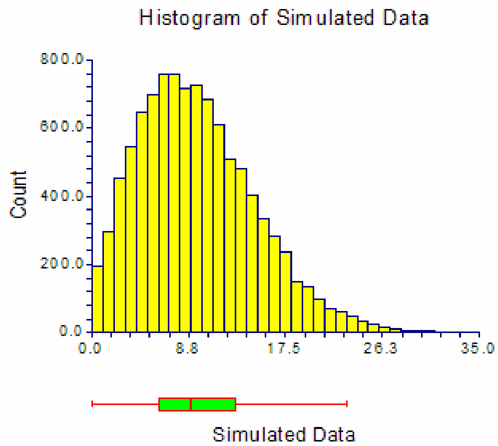
As you can see, the power computed here matches the results we computed manually.

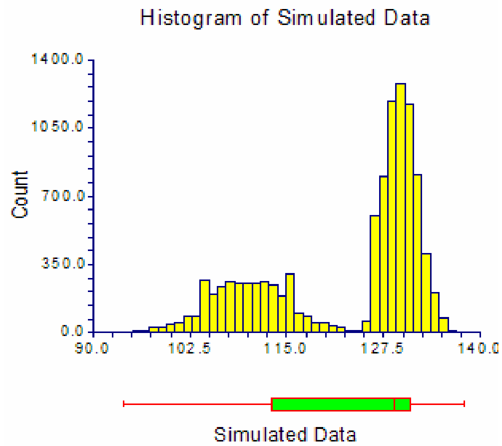
Chapter 630

Data Simulator

Introduction

Because of mathematical intractability, it is often necessary to investigate the properties of a statistical procedure using *simulation* (or *Monte Carlo*) techniques. In power analysis, *simulation* refers to the process of generating several thousand random samples that follow a particular distribution, calculating the test statistic from each sample, and tabulating the distribution of these test statistics so that the significance level and power of the procedure may be investigated. This module creates a histogram of a specified distribution as well as a numerical summary of simulated data. By studying the histogram and the numerical summary, you can determine if the distribution has the characteristics you desire. The distribution formula can then be used in procedures that use simulation, such as the new t-test procedures. Below are examples of two distributions that were generated with this procedure.



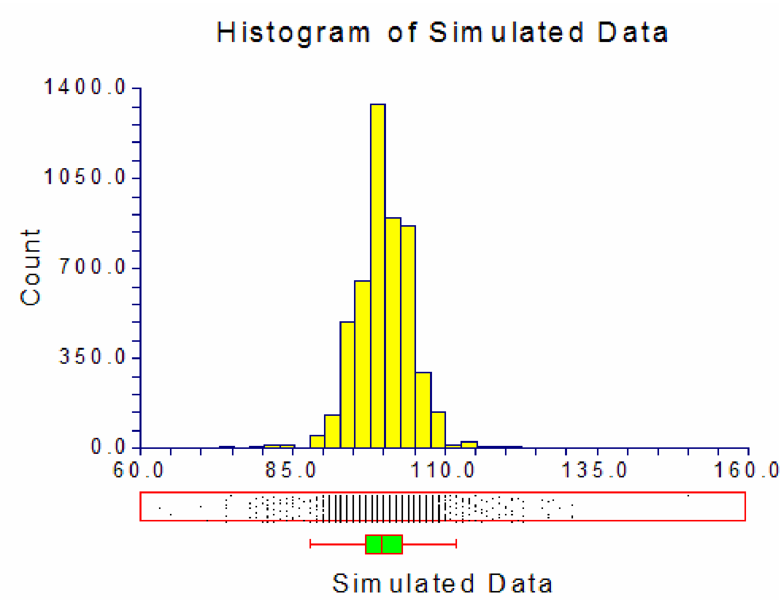


Technical Details

A random variable's probability distribution specifies its probability over its range of values. Examples of common continuous probability distributions are the normal and uniform distributions. Unfortunately, experimental data often do not follow these common distributions, so other distributions have been proposed. One of the easiest ways to create distributions with desired characteristics is to combine simple distributions. For example, outliers may be added to a distribution by mixing it with data from a distribution with a much larger variance. Thus, to simulate normally distributed data with 5% outliers, we could generate 95% of the sample from a normal distribution with mean 100 and standard deviation 4 and then generate 5% of the sample from a normal distribution with mean 100 and standard deviation 16. Using the standard notation for the normal distribution, the composite distribution of the new random variable Y could be written as

$$Y \sim \delta(0 \le X < 0.95)N(100,4) + \delta(0.95 \le X \le 1.00)N(100,16)$$

where X is a uniform random variable between 0 and 1, $\delta(z)$ is 1 or 0 depending on whether *z* is true or false, N(100,4) is a normally distributed random variable with mean 100 and standard deviation 4, and N(100,16) is a normally distributed random variable with mean 100 and standard deviation 16. The resulting distribution is shown below. Notice how the tails extend in both directions.



The procedure for generating a random variable, *Y*, with the mixture distribution described above is

1. Generate a uniform random number, *X*.

2. If *X* is less than 0.95, *Y* is created by generating a random number from the *N*(100,4) distribution.

3. If *X* is greater than or equal to 0.95, *Y* is created by generating a random number from the *N*(100,16) distribution.

Note that only one uniform random number and one normal random number are generated for any particular random realization from the mixture distribution.

In general, the formula for a mixture random variable, Y, which is to be generated from two or more random variables defined by their distribution function $F_i(Z_i)$ is given by

$$Y \sim \sum_{i=1}^{k} \delta(a_i \leq X < a_{i+1}) F_i(Z_i), \quad a_1 = 0 < a_2 < \cdots < a_{K+1} = 1$$

Note that the a_i's are chosen so that weighting requirements are met. Also note that only one uniform random number and one other random number actually need to be generated for a particular value. The $F_i(Z_i)$'s may be any of the distributions which are listed below.

Since the test statistics which will be simulated are used to test hypotheses about one or more means, it will be convenient to parameterize the distributions in terms of their means.

Beta Distribution

The beta distribution is given by the density function

$$f(x) = \frac{\Gamma(A+B)}{\Gamma(A)\Gamma(B)} \left(\frac{x-C}{D-C}\right)^{A-1} \left(1 - \frac{x-C}{D-C}\right)^{B-1}, \quad A, B > 0, C \leq x \leq D$$

where A and B are shape parameters, C is the minimum, and D is the maximum. In statistical theory, C and D are usually zero and one, respectively, but the more general formulation used here is more convenient for simulation work. In this program module, a beta random variable is specified as ***A*(*M*, *A*, *B*, *C*)**, where M is the mean which is

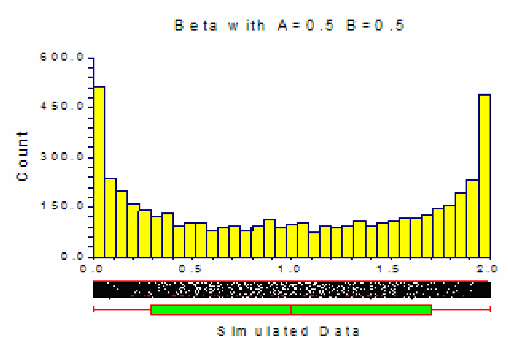
$$E(X) = M = (D-C)\left[\frac{A}{A+B}\right] + C$$

The parameter D is obtained from M, A, B, and C using the relationship

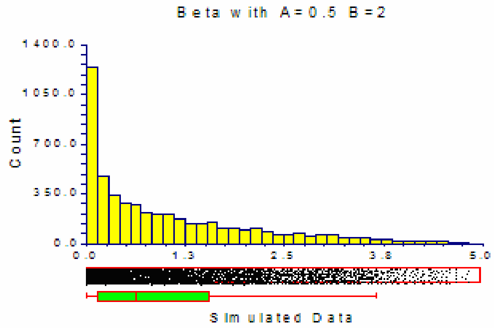
$$D = \frac{(M-C)(A+B)}{A} + C.$$

The beta density can take a number of shapes depending on the values of A and B:

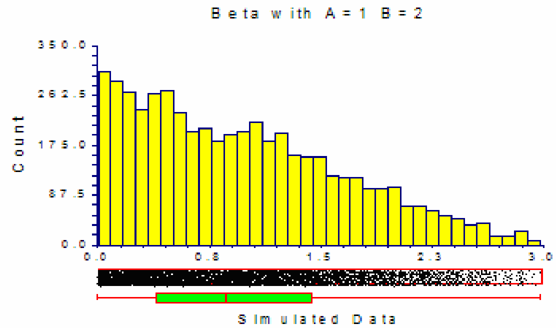
1. When A<1 and B<1 the density is U-shaped.



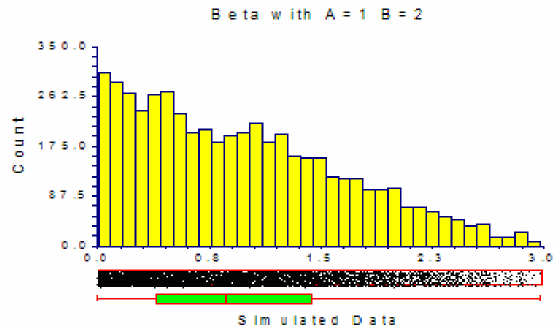
2. When $0 < A < 1 \leq B$ the density is J-shaped.



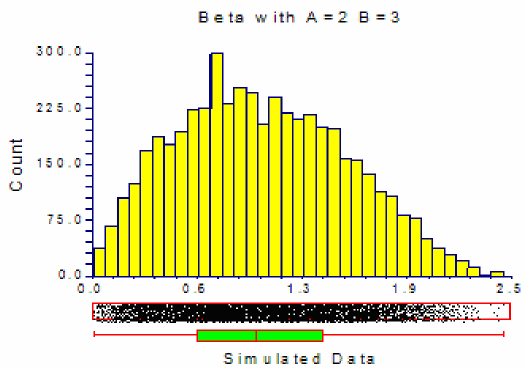
3. When A=1 and B>1 the density is bounded and decreases monotonically to 0.



4. When A=1 and B=1 the density is the uniform density.



5. When A>1 and B>1 the density is unimodal.



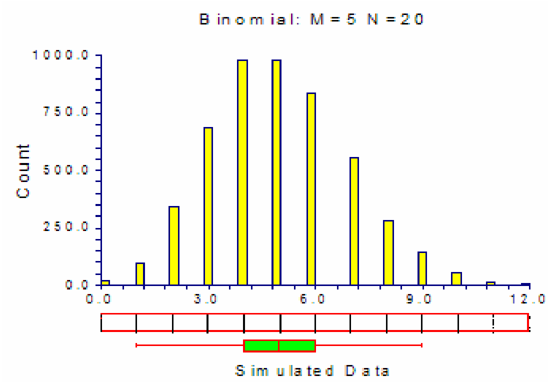
Beta random variates are generated using Cheng's rejection algorithm as given on page 438 of Devroye (1986).

Binomial Distribution

The binomial distribution is given by the function

$$\Pr(X = x) = \binom{n}{x} \pi^x (1-\pi)^{n-x}, \quad x = 0,1,2,\ldots,n$$

In this program module, the binomial is specified as ***B*(*M*, *n*)**, where *M* is the mean which is equal to $n\pi$ and *n* is the number of trials. The probability of a positive response, π, is not entered directly, but is obtained using the relationship $\pi = M / n$. For this reason, $0 < M < n$.



Binomial random variates are generated using the inverse CDF method. That is, a uniform random variate is generated, and then the CDF of the binomial distribution is scanned to determine which value of *X* is associated with that probability.

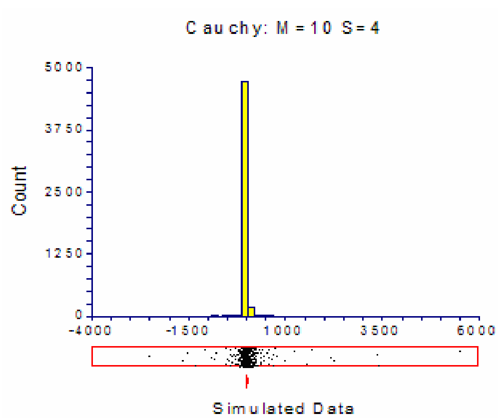
Cauchy Distribution

The Cauchy distribution is given by the density function

$$f(x) = \left[S\pi \left(1 + \left\{ \frac{X-M}{S} \right\}^2 \right) \right]^{-1}, \quad S > 0$$

Although the Cauchy distribution does not possess a mean and standard deviation, *M* and *S* are treated as such. Cauchy random numbers are generated using the algorithm given in Johnson, Kotz, and Balakrishnan (1994), page 327.

In this program module, the Cauchy is specified as ***C*(*M*, *S*)**, where *M* is a location parameter (median), and *S* is a scale parameter.



Constant Distribution

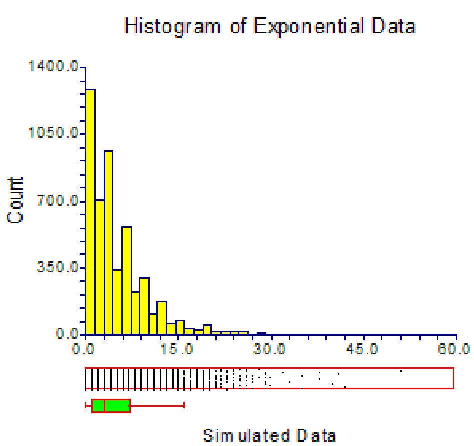
The *constant* distribution occurs when a random variable can only take a single value, *X*. The constant distribution is specified as ***K*(*X*)**, where *X* is the value.

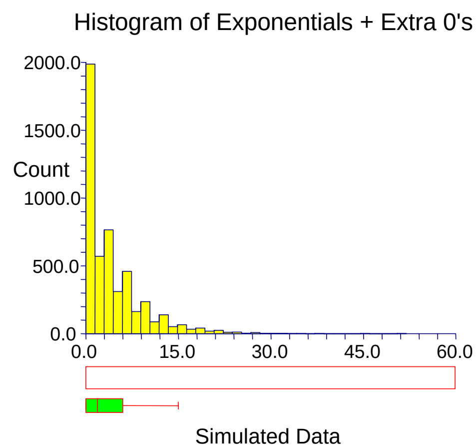
Data with a Many Zero Values

Sometimes data follow a specific distribution in which there is a large proportion of zeros. This can happen when data are counts or monetary amounts. Suppose you want to generate exponentially distributed data with an extra number of zeros. You could use the following simulation model:

K(0)[2]; E(5)[9]

The exponential distribution alone was used to generate the histogram below on the left. The histogram below on the right was simulated by adding extra zeros to the exponential data.



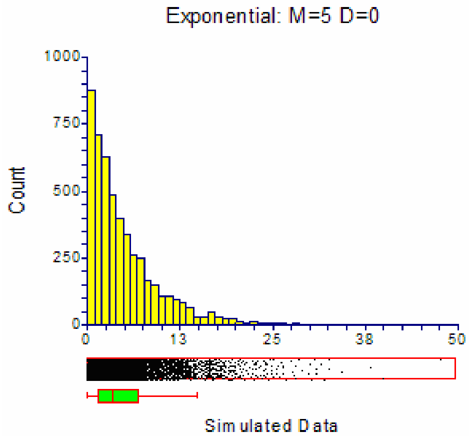


Exponential Distribution

The exponential distribution is given by the density function

$$f(x) = \frac{1}{M} e^{-\frac{x}{M}}, \quad x > 0$$

In this program module, the exponential is specified as ***E*(*M*)**, where M is the mean.



Random variates from the exponential distribution are generated using the expression $-M\ln(U)$, where U is a uniform random variate.

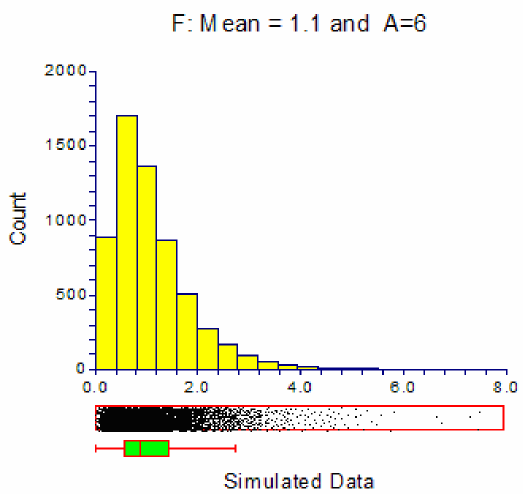
F Distribution

Snedecor's *F* distribution is the distribution of the ratio of two independent chi-square variates. The degrees of freedom of the numerator chi-square variate is *A*, while that of the denominator chi-square is *D*. The *F* distribution is specified as ***F*(*M, A*)**, where *M* is the mean and *A* is the degrees of freedom of the numerator chi-square. The value of *M* is related to the denominator chi-square degrees of freedom using the relationship *M=D/(D-2)*.

F variates are generated by first generating a symmetric beta variate, *B*(*A*/2,*D*/2), and transforming it into an *F* variate using the relationship

$$F_{A,D} = \frac{BD}{A - BA}$$

Below is a histogram for data generated from an *F* distribution with a mean of 1.1 and *A* = 6.



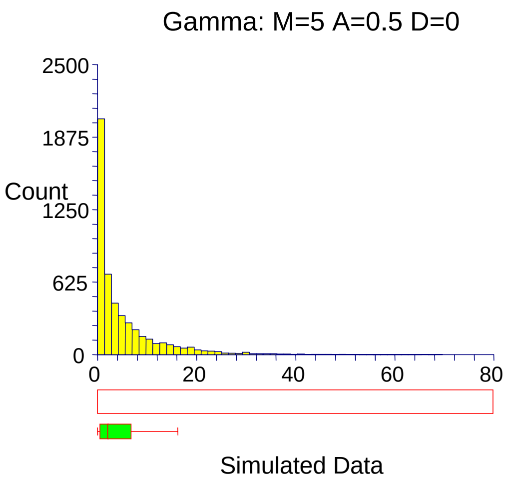
Gamma Distribution

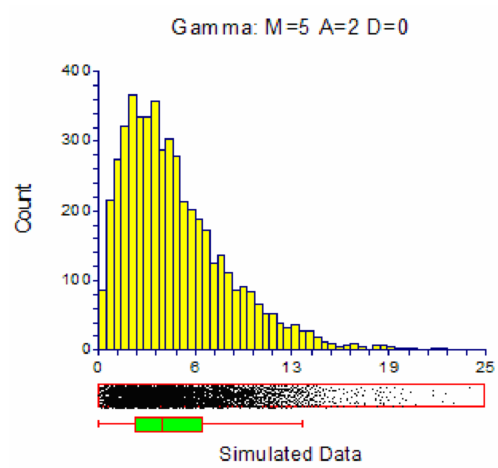
The three parameter gamma distribution is given by the density function

$$f(x)=\frac{(x)^{A-1}}{B^{A}\Gamma(A)}e^{-\frac{x}{B}},\quad x>0,\ A>0,\ B>0$$

where A is a shape parameter and B is a scale parameter. In this program module, the gamma is specified as ***G*(*M*, *A*)**, where M is the mean, given by $M=AB$. The parameter B may be obtained using the relationship $B = M/A$.

Gamma variates are generated using the exponential distribution when $A = 1$, Best's XG algorithm given in Devroye (1986), page 410, when $A > 1$, and Vaduva's algorithm given in Devroye (1986), page 415, when $A < 1$.





Multinomial Distribution

The *multinomial* distribution occurs when a random variable has only a few discrete values such as 1, 2, 3, 4, and 5. The multinomial distribution is specified as ***M(P1, P2, …, Pk)***, where *Pi* is the is the probability of that the integer *i* occurs. Note that the values start at one, not zero.

For example, suppose you want to simulate a distribution which has 50% 3's and 1's, 2's, 4's, and 5's all with equal percentages. You would enter *M(1 1 4 1 1)*.

As a second example, suppose you wanted to have a equal percentage of 1's, 3's, and 7's, and none of the other percentages. You would enter *M(1 0 1 0 0 0 1)*.

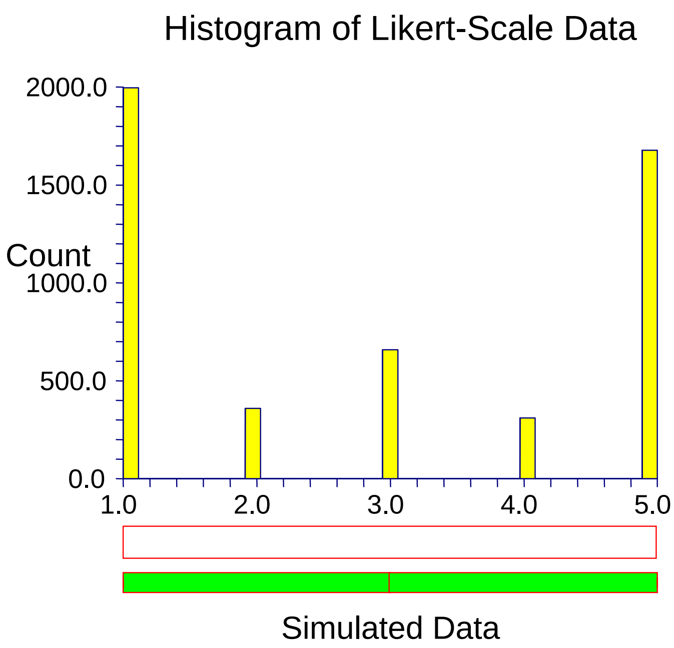
Likert-Scale Data

Likert-scale data are common in surveys and questionnaires. To generate data from a five-point Likert-scale distribution, you could use the following simulation model:

M(6 1 2 1 5)

Note that the weights are relative—they do not have to sum to one. The program will make the appropriate weighting adjustments so that they do sum to one.

The above expression generated the following histogram.



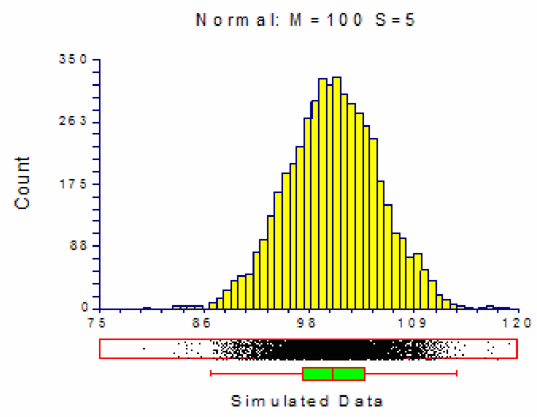
Normal Distribution

The normal distribution is given by the density function

$$f(x) = \phi\left(\frac{x-\mu}{\sigma}\right), \quad -\infty \le x \le \infty$$

where $\phi(z)$ is the usual standard normal density. The normal distribution is specified as ***N*(*M*, *S*)**, where *M* is the mean and *S* is the standard deviation.

The normal distribution is generated using the Marsaglia and Bray algorithm as given in Devroye (1986), page 390.



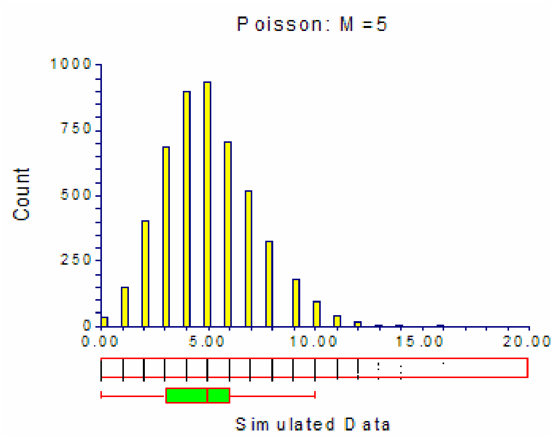
Poisson Distribution

The Poisson distribution is given by the function

$$\Pr(X = x) = \frac{e^{-M} M^{x}}{x!}, \quad x = 0,1,2,\ldots, \; M > 0$$

In this program module, the Poisson is specified as ***P*(*M*)**, where *M* is the mean.

Poisson random variates are generated using the inverse CDF method. That is, a uniform random variate is generated and then the CDF of the Poisson distribution is scanned to determine which value of *X* is associated with that probability.



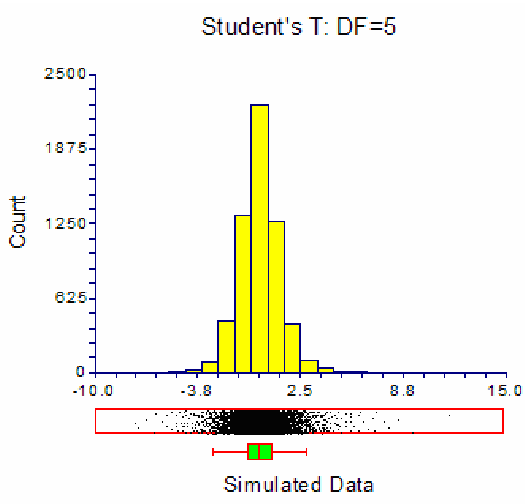
Student's T Distribution

Student's *T* distribution is the distribution of the ratio of a unit normal variate and the square root of an independent chi-square variate. The degrees of freedom of the chi-square variate are the degrees of freedom of the *T* distribution. The *T* is specified as ***T*(*M*, *A*)**, where *M* is the mean and *A* is the degrees of freedom. The central *T* distribution generated in this program has a mean of zero, so, to obtain a mean of *M*, *M* is added to every data value.

T variates are generated by first generating a symmetric beta variate, B(A/2,A/2), with mean equal to 0.5. This beta variate is then transformed into a *T* variate using the relationship

$$T = \sqrt{A}\frac{X - 0.5}{\sqrt{X(1 - X)}}$$

Here is a histogram for data generated from a *T* distribution with mean 0 and 5 degrees of freedom.



Tukey's Lambda Distribution

Hoaglin (1985) presents a discussion of a distribution developed by John Tukey for allowing the detailed specification of skewness and kurtosis in a simulation study. This distribution is extended in the work of Karian and Dudewicz (2000). Tukey's idea was to reshape the normal distribution using functions that change the skewness and/or kurtosis. This is accomplished by multiplying a normal random variable by a skewness function and/or a kurtosis function. The general form of the transformation is

$$X = A + B\left\{G_g(z)H_h(z)z\right\}$$

where *z* has the standard normal density. The skewness function Tukey proposed is

$$G_g(z) = \frac{e^{gz} - 1}{gz}$$

The range of *g* is typically -1 to 1. The value of $G_0(z) \equiv 1$. The kurtosis function Tukey proposed is

$$H_h(z) = e^{hz^2/2}$$

The range of *h* is also -1 to 1.

Hence, if both *g* and *h* are set to zero, the variable *X* follows the normal distribution with mean *A* and standard deviation *B*. As *g* is increased toward 1, the distribution is increasingly skewed to the right. As *g* is decreased towards -1, the distribution is increasingly skewed to the left. As *h* is increased toward 1, the data are stretched out so that more extreme values are probable. As *h* is decreased toward -1, the data are concentrated around the center—resulting in a beta-type distribution.

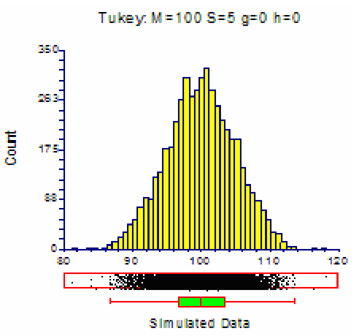
The mean of this distribution is given by

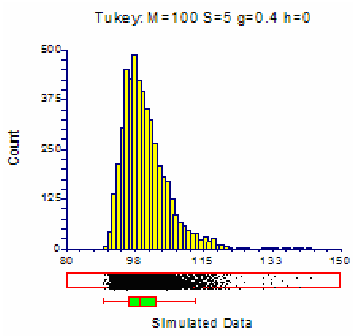
$$M = A + B\left(\frac{e^{g^2/2(1-h)} - 1}{g\sqrt{1-h}}\right), \quad 0 \le h < 1$$

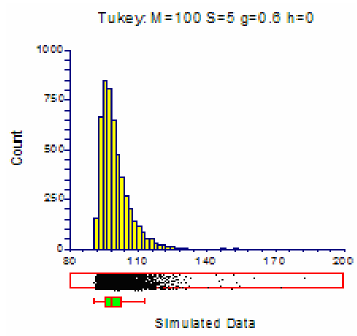
which may be easily solved for *A*.

Tukey's lambda is specified in the program as ***L*(*M, B, g, h*)** where *M* is the mean, *B* is a scale factor (when *g*=*h*=0, *B* is the standard deviation), *g* is the amount of skewness, and *h* is the amount of kurtosis.

Random variates are generated from this distribution by generating a random normal variate and then applying the skewness and kurtosis modifications. Here are some examples as *g* is varied from 0 to 0.4 to 0.6. Notice how the amount of skewness is gradually increased. Similar results are achieved when *h* is varied from 0 to 0.5.





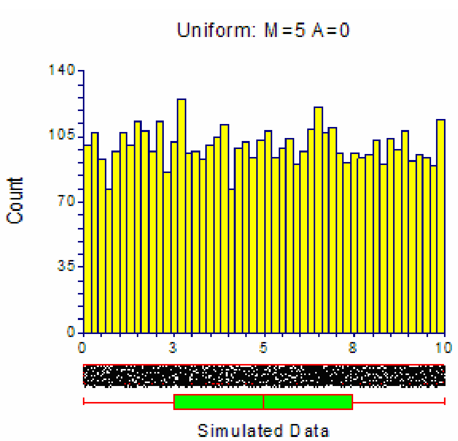


Uniform Distribution

The uniform distribution is given by the density function

$$f(x) = \frac{1}{B - A}, \quad A \leq x \leq B$$

In this program module, the uniform is specified as ***U*(*M*, *A*)**, where *M* is the mean which is equal to (*A*+*B*)/2 and *A* is the minimum of *x*. The parameter *B* is obtained using the relationship *B=2M-A*.



Uniform random numbers are generated using Makoto Matsumoto's Mersenne Twister uniform random number generator which has a cycle length greater than (that's a one followed by 6000 zeros).

Weibull Distribution

The Weibull distribution is indexed by a shape parameter, *B*, and a scale parameter, *C*. The Weibull density function is written as

$$f(x|B,C) = \frac{B}{C}\left(\frac{z}{C}\right)^{(B-1)} e^{-\left(\frac{x}{C}\right)^{B}}, \quad B > 0, C > 0, \; x > 0.$$

Shape Parameter - B

The shape parameter controls the overall shape of the density function. Typically, this value ranges between 0.5 and 8.0. One of the reasons for the popularity of the Weibull distribution is that it includes other useful distributions as special cases or close approximations. For example, if

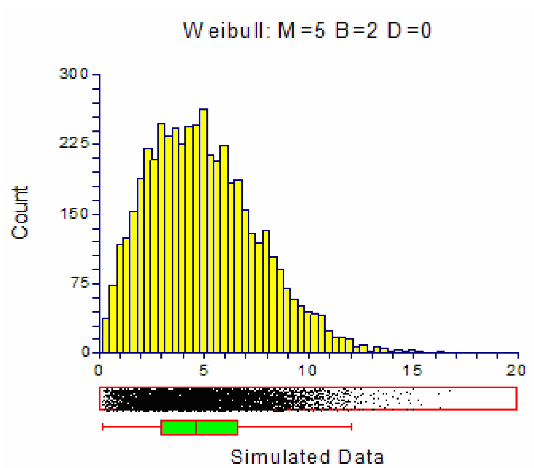
B = 1	The Weibull distribution is identical to the exponential distribution.
B = 2	The Weibull distribution is identical to the Rayleigh distribution.
B = 2.5	The Weibull distribution approximates the lognormal distribution.
B = 3.6	The Weibull distribution approximates the normal distribution.

Scale Parameter - C

The scale parameter only changes the scale of the density function along the *x* axis. Some authors use 1/*C* instead of *C* as the scale parameter. Although this is arbitrary, we prefer dividing by the scale parameter since that is how one usually scales a set of numbers.

The Weibull is specified in the program as ***W*(*M*, *B*)**, where *M* is the mean which is given by

$$M = C\,\Gamma\left(1 + \frac{1}{B}\right).$$



Combining Distributions

This section discusses how distributions may be combined to form new distributions. Combining may be done in the form of algebraic manipulation, mixtures, or both.

Creating New Distributions using Expressions

The set of probability distributions discussed above provides a basic set of useful distributions. However, you may want to mimic reality more closely by combining these basic distributions. For example, paired data is often analyzed by forming the differences of the two original variables. If the original data are normally distributed, then the differences are also normally distributed. Suppose, however, that the original data are exponential. The difference of two exponentials is not a common distribution.

Expression Syntax

The basic syntax is

C1 D1 operator1 C2 D2 operator2 C3 D3 operator3 …

where C1, C2, C3, etc. are coefficients (numbers), D1, D2, D3, etc. are probability distributions, and *operator* is one of the four symbols: +, -, *, /. Parentheses are only permitted in the specification of distributions.

Examples of valid expressions include

N(4, 5) – N(4, 5)

2E(3) – 4E(4) + 2E(5)

N(4, 2)/E(4)-K(5)

Notes about the coefficients: C1, C2, C3

The coefficients may be positive or negative decimal numbers such as 2.3, 5, or -3.2. If no coefficient is specified, the coefficient is assumed to be one.

Notes about the Distributions: D1, D2, D3

The distributions may be any of the distributions listed above such as normal, exponential, or beta. The expressions are evaluated by generating random values from each of the distributions specified and then combining them according to the operators.

Notes about the operators: +, -, *, /

All multiplications and divisions are performed first, followed by any additions and subtractions.

Note that if only addition and subtraction are used in the expression, the mean of the resulting distribution is found by applying the same operations to the individual distribution means. If the expression involves multiplication or division, the mean of the resulting distribution is usually difficult to calculate directly.

Creating New Distributions using Mixtures

Mixture distributions are formed by sampling a fixed percentage of the data from each of several distributions. For example, you may model outliers by obtaining 95% of your data from a normal distribution with a standard deviation of 5 and 5% of your data from a distribution with a standard deviation of 50.

Mixture Syntax

The basic syntax of a mixture is

D1[W1]; D2[W2]; …; Dk[Wk]

where the D's represent distributions and the *W*'s represent weights. Note that the weights must be positive numbers. Also note that semi-colons are used to separate the components of the mixture.

Examples of valid mixture distributions include

N(4, 5)[19]; N(4, 50)[1]. 95% of the distribution is N(4, 5), and the other 5% is N(4, 50).

W(4, 3)[7]; K(0)[3]. 70% of the distribution is W(4, 3), and the other 30% is made up of zeros.

N(4, 2)-N(4,3)[2]; E(4)*E(2)[8]. 20% of the distribution is N(4, 2)-N(4,3), and the other 80% is E(4)*E(2).

Notes about the distributions

The distributions D1, D2, D3, etc. may be any valid distributional expression.

Notes about the weights

The weights w1, w2, w3, etc. need not sum to one (or to one hundred). The program uses these weights to calculate new, internal weights that do sum to one. For example, if you enter weights of 1, 2, and 1, the internal weights will be 0.25, 0.50, and 0.25.

When a weight is not specified, it is assumed to have the value of '1.' Thus

N(4, 5)[19]; N(4,50)[1]

is equivalent to

N(4, 5)[19]; N(4,50)

Procedure Options

This section describes the options that are specific to this procedure. These are located on the Data tab. To find out more about using the other tabs such as Labels or Plot Setup, turn to the chapter entitled Procedure Templates.

Data Tab

The Data tab contains the parameters used to specify a probability distribution.

Probability Distribution to be Simulated

Enter the components of the probability distribution to be simulated. One or more components may be entered from among the continuous and discrete distributions listed below the data-entry boxes. The model is specified in the first two boxes. The model is formed by joining the contents of these two boxes.

The *W* parameter gives the relative weight of that component. For example, if you entered

P(5)[1];K(0)[2], about 33% of the random numbers would follow the P(5) distribution, and 67% would be 0. When only one component is used, the value of *W* may be omitted. For example, to generate data from the normal distribution with mean of five and standard deviation of one, you would enter N(5, 1), not N(5, 1)[1].

Each of the possible components were discussed earlier in the chapter.

Numbers in Histogram

This is the number of values generated from the probability distribution for display in the histogram. We recommend a value of about 5000.

Note that the histogram, box plot, and dot plot row limits must be set higher than this amount or the corresponding plot will not be displayed. These limits are modified by selecting Edit, Options, and Limits from the spreadsheet menu.

Numbers Stored

This is the number of values that are stored in the current database for use by NCSS in the column designated under the *Storage* tab. If this value is larger than the *Numbers in Histogram* value, it will be reset to that value.

If no column is specified in the *Simulated Values* box of the *Storage* tab, or if this value is zero, nothing will be stored on the database.

Usually, you will not need to store values on a database.

Force Positive

Checking this box will force all generated values to be positive. This is accomplished by taking the absolute value of each generated value.

Round to Integer

When checked, each value will be rounded off to an integer.

Example1 – Generating Normal Data

In this example, 5000 values will be generated from the standard normal (mean zero, variance one) distribution. These values will be displayed in a histogram and summarized numerically.

Setup

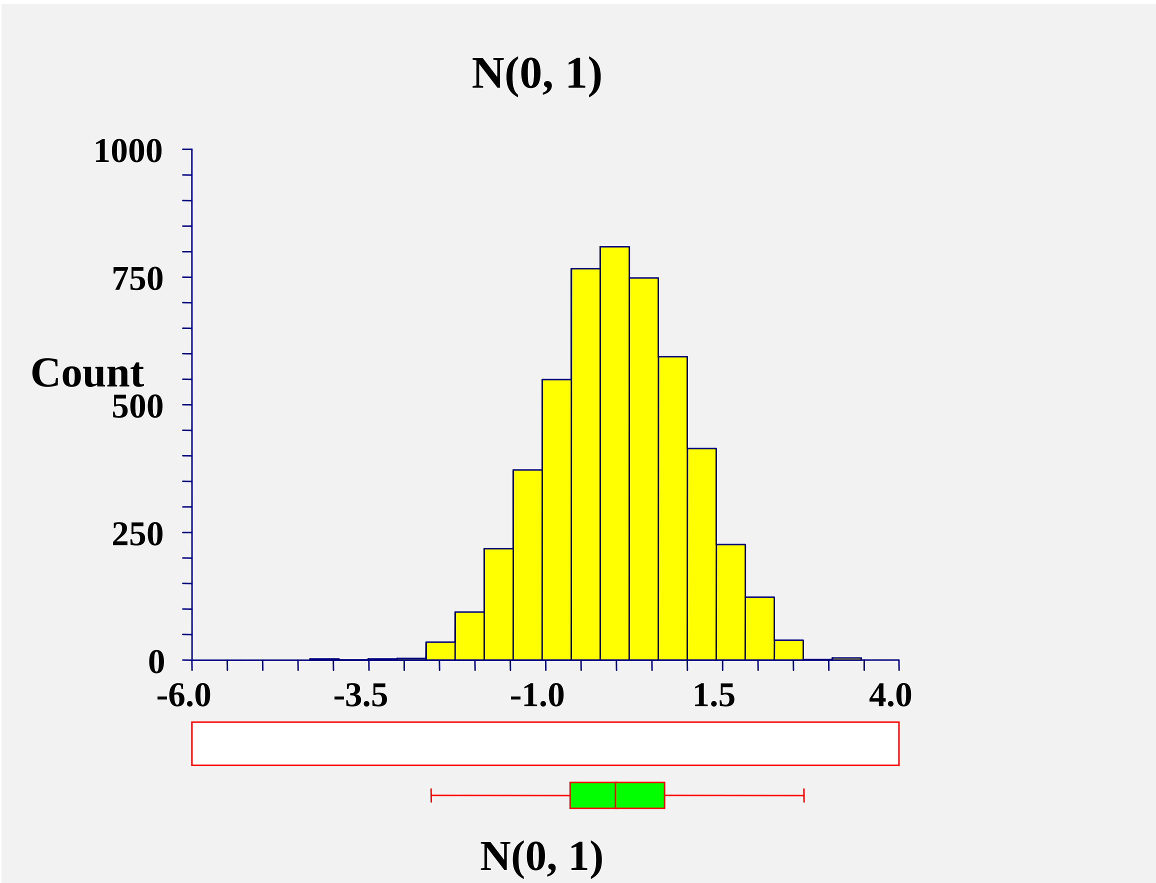
This section presents the values of each of the parameters needed to run this example. You can make these changes directly on your screen or you can load the template entitled Example1 by clicking the Template tab and loading this template.

Option	Value
Data Tab	
Probability Distribution	**N(0, 1)**
Numbers in Histogram	**5000**
Numbers Stored	**0**
Force Positive	**not checked**
Round to Integer	**not checked**

Annotated Output

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

Histogram



Descriptive Statistics of Simulated Data

Statistic	Value	Statistic	Value
Mean	1.228682E-02	Minimum	-4.328354
Standard Deviation	0.9888594	1st Percentile	-2.234667
Skewness	-1.794159E-02	5th Percentile	-1.642934
Kurtosis	2.881271	10th Percentile	-1.276706
Coefficient of Variation	80.48133	25th Percentile	-0.6501855
Count	5000	Median	1.085566E-02
		75th Percentile	0.6852141
		90th Percentile	1.321415
		95th Percentile	1.673769
		99th Percentile	2.220547
		Maximum	3.468608

This report shows the histogram and a numerical summary of the 5000 simulate normal values. It is interesting to check how well the simulation did. Theoretically, the mean should be zero, the standard deviation one, the skewness zero, and the kurtosis three. Of course, your results will vary from these because these are based on generated random numbers.

Example2 – Generating Data from a Contaminated Normal

In this example, we will generate data from a contaminated normal. This will be accomplished by generating 95% of the data from a N(100,3) distribution and 5% from a N(110,15) distribution.

Setup

This section presents the values of each of the parameters needed to run this example. You can make these changes directly on your screen or you can load the template entitled Example2 by clicking the Template tab and loading this template.

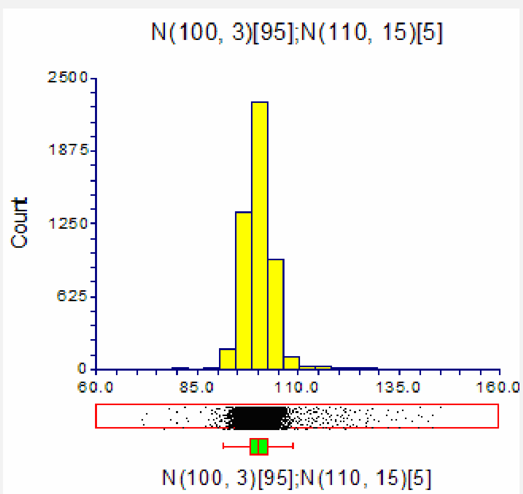
Option **Value**

Data Tab

Probability Distribution ....**N(100, 3)[95];N(110, 15)[5]**

Annotated Output

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



Descriptive Statistics of Simulated Data

Statistic	Value	Statistic	Value
Mean	100.3549	Minimum	71.11177
Standard Deviation	4.716588	1st Percentile	92.40327
Skewness	2.67692	5th Percentile	94.82752
Kurtosis	23.11323	10th Percentile	96.01687
Coefficient of Variation	4.699908E-02	25th Percentile	97.88011
Count	5000	Median	100.0328
		75th Percentile	102.18
		90th Percentile	104.266
		95th Percentile	105.9543
		99th Percentile	120.4605
		Maximum	145.2731

This report shows the data from the contaminated normal. The mean is close to 100, but the standard deviation, skewness, and kurtosis have non-normal values. Note that there are now some very large outliers.

Example3 – Likert-Scale Data

In this example, we will generate data following a discrete distribution on a Likert scale. The distribution of the Likert scale will be 30% 1's, 10% 2's, 20% 3's, 10% 4's, and 30% 5's.

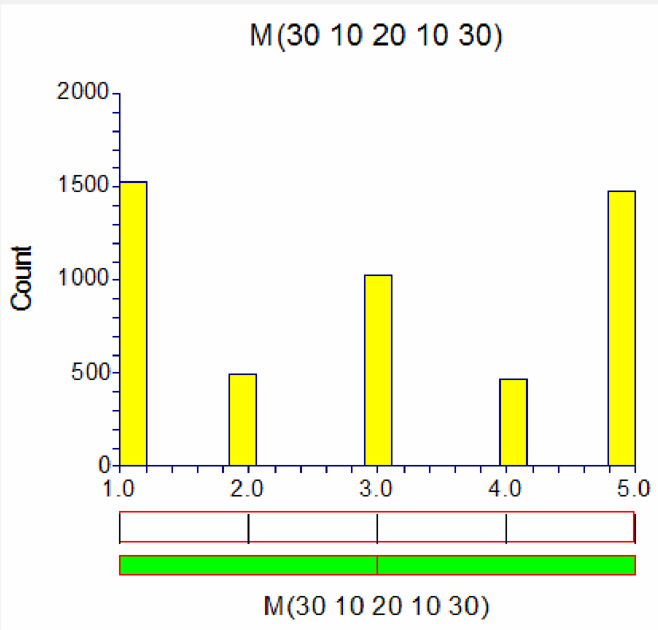
Setup

This section presents the values of each of the parameters needed to run this example. You can make these changes directly on your screen or you can load the template entitled Example3 by clicking the Template tab and loading this template.

Option	Value
Data Tab	
Probability Distribution	**M(30 10 20 10 30)**

Annotated Output

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



Descriptive Statistics of Simulated Data

Statistic	Value	Statistic	Value
Mean	2.9858	Minimum	1
Standard Deviation	1.622506	1st Percentile	1
Skewness	1.169226E-02	5th Percentile	1
Kurtosis	1.436589	10th Percentile	1
Coefficient of Variation	0.5434074	25th Percentile	1
Count	5000	Median	3
		75th Percentile	5
		90th Percentile	5
		95th Percentile	5
		99th Percentile	5
		Maximum	5

This report shows the data from a Likert scale.

Example4 –Bimodal Data

In this example, we will generate data that have a bimodal distribution. We will accomplish this by combining data from two normal distributions, one with a mean of 10 and the other with a mean of 30. The standard deviation will be set at 4.

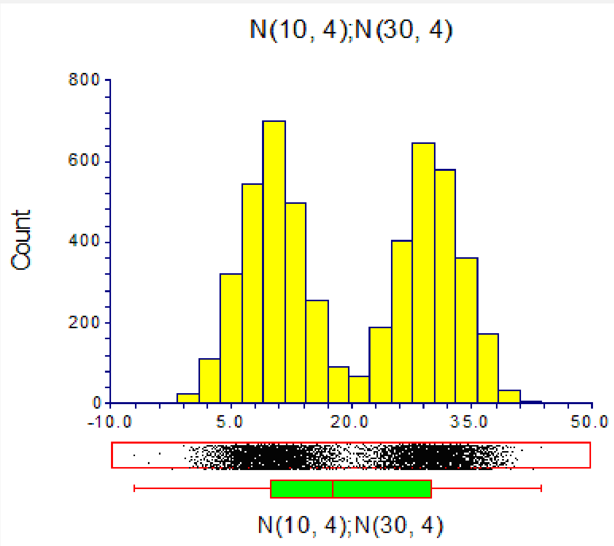
Setup

This section presents the values of each of the parameters needed to run this example. You can make these changes directly on your screen or you can load the template entitled Example4 by clicking the Template tab and loading this template.

Option	**Value**
Data Tab	
Probability Distribution	**N(10, 4);N(30, 4)**

Annotated Output

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



Descriptive Statistics of Simulated Data

Statistic	Value	Statistic	Value
Mean	19.82392	Minimum	-7.067288
Standard Deviation	10.81724	1st Percentile	1.819361
Skewness	3.503937E-02	5th Percentile	4.949873
Kurtosis	1.500108	10th Percentile	6.473907
Coefficient of Variation	0.545666	25th Percentile	9.863808
Count	5000	Median	17.73235
		75th Percentile	29.94195
		90th Percentile	33.4243
		95th Percentile	35.11861
		99th Percentile	38.04005
		Maximum	43.71141

This report shows the results for the simulated bimodal data.

Example5 – Gamma Data with Extra Zeros

In this example, we will generate data that have a gamma distribution, except that we will force there to be about 30% zeros. The gamma distribution will have a shape parameter of 5 and a mean of 10.

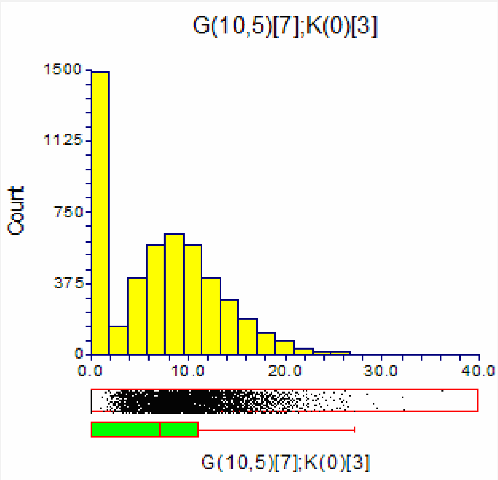
Setup

This section presents the values of each of the parameters needed to run this example. You can make these changes directly on your screen or you can load the template entitled Example5 by clicking the Template tab and loading this template.

Option	**Value**
Data Tab	
Probability Distribution	**G(10,5)[7];K(0)[3]**

Annotated Output

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



Descriptive Statistics of Simulated Data

Statistic	Value	Statistic	Value
Mean	7.0965	Minimum	0
Standard Deviation	5.960041	1st Percentile	0
Skewness	0.4635893	5th Percentile	0
Kurtosis	2.730829	10th Percentile	0
Coefficient of Variation	0.8398563	25th Percentile	0
Count	5000	Median	7.24275
		75th Percentile	11.09861
		90th Percentile	14.94062
		95th Percentile	17.41686
		99th Percentile	22.18812
		Maximum	36.13172

This report shows the results for the simulated bimodal data.

Example6 – Mixture of Two Poisson Distributions

In this example, we will generate data that have a mixture of two Poisson distributions. 60% of the data will be from a Poisson distribution with a mean of 10 and 40% from a Poisson distribution with a mean of 20.

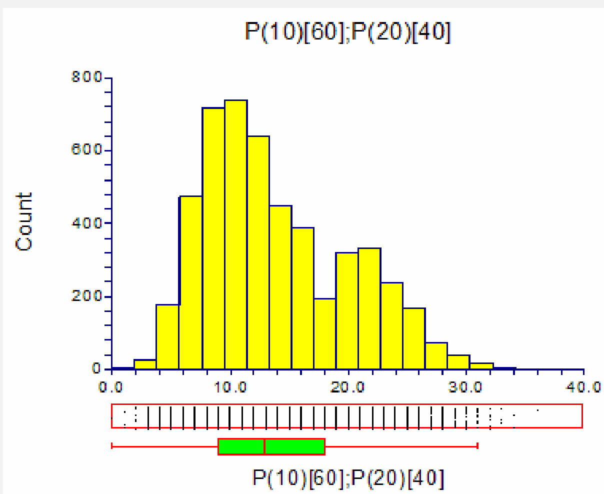
Setup

This section presents the values of each of the parameters needed to run this example. You can make these changes directly on your screen or you can load the template entitled Example6 by clicking the Template tab and loading this template.

Option	Value
Data Tab	
Probability Distribution	**P(10)[60];P(20)[40]**

Annotated Output

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



Descriptive Statistics of Simulated Data

Statistic	Value	Statistic	Value
Mean	13.8662	Minimum	0
Standard Deviation	6.088589	1st Percentile	4
Skewness	0.5726398	5th Percentile	6
Kurtosis	2.580317	10th Percentile	7
Coefficient of Variation	0.4390957	25th Percentile	9
Count	5000	Median	13
		75th Percentile	18
		90th Percentile	23
		95th Percentile	25
		99th Percentile	29
		Maximum	36

This report shows the results for the simulated mixture-Poisson data.

Example7 – Difference of Two Identically Distributed Exponentials

In this example, we will demonstrate that the difference of two identically distributed exponential random variables follows a symmetric distribution. This is particularly interesting because the exponential distribution is skewed. In fact, the difference between any two identically distributed random variables follows a symmetric distribution.

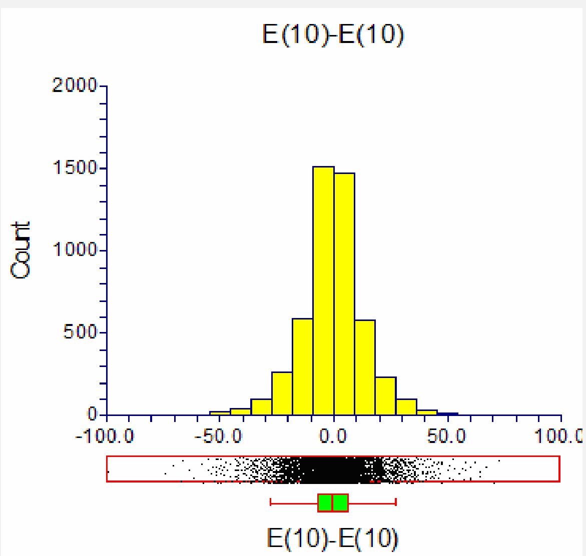
Setup

This section presents the values of each of the parameters needed to run this example. You can make these changes directly on your screen or you can load the template entitled Example7 by clicking the Template tab and loading this template.

Option	Value
Data Tab	
Probability Distribution	**E(10)-E(10)**

Annotated Output

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



Descriptive Statistics of Simulated Data

Statistic	Value	Statistic	Value
Mean	-0.3431585	Minimum	-99.37727
Standard Deviation	14.20288	1st Percentile	-41.02162
Skewness	-7.391038E-02	5th Percentile	-23.19279
Kurtosis	6.115685	10th Percentile	-16.70725
Coefficient of Variation	-41.38868	25th Percentile	-7.131404
Count	5000	Median	-0.1380393
		75th Percentile	6.575472
		90th Percentile	15.78555
		95th Percentile	22.12444
		99th Percentile	38.82129
		Maximum	72.36995

This report shows demonstrates that the distribution of the difference is symmetric..

Chapter 650

One Variance

Introduction

Occasionally, researchers are interested in the estimation of the variance (or standard deviation) rather than the mean. This module calculates the sample size and performs power analysis for hypothesis tests concerning a single variance.

Technical Details

Assuming that a variable X is normally distributed with mean μ and variance σ^2, the sample variance is distributed as a Chi-square random variable with N - 1 degrees of freedom, where N is the sample size. That is,

$$\mathrm{X}^2 = \frac{(N-1)s^2}{\sigma^2}$$

is distributed as a Chi-square random variable. The sample statistic, s^2, is calculated as follows

$$s^2 = \frac{\sum_{i=1}^{N}\left(X_i - \overline{X}\right)^2}{N-1}.$$

The power or sample size of a hypothesis test about the variance can be calculated using the appropriate one of the following three formulas from Ostle (1988) page 130.

Case 1: $H_0\colon \sigma_1^2 = \sigma_0^2 \quad \textit{versus} \quad H_a\colon \sigma_1^2 \neq \sigma_0^2$

$$\beta = P\left(\frac{\sigma_0^2}{\sigma_1^2}\chi^2_{\alpha/2,N-1} < \chi^2 < \frac{\sigma_0^2}{\sigma_1^2}\chi^2_{1-\alpha/2,N-1}\right)$$

Case 2: $H_0\colon \sigma_1^2 = \sigma_0^2 \quad \textit{versus} \quad H_a\colon \sigma_1^2 > \sigma_0^2$

$$\beta = P\left(\chi^2 < \frac{\sigma_0^2}{\sigma_1^2}\chi^2_{1-\alpha,N-1}\right)$$

Case 3: $H_0\colon \sigma_1^2 = \sigma_0^2 \quad \textit{versus} \quad H_a\colon \sigma_1^2 < \sigma_0^2$

$$\beta = P\left(\chi^2 > \frac{\sigma_0^2}{\sigma_1^2}\chi^2_{\alpha,N-1}\right)$$

Procedure Options

This section describes the options that are unique to this procedure. These are located on the panels associated with the Data, Options, and Reports tabs. To find out more about using the other tabs such as Plot Text, Axes, and Template, turn to the chapter entitled Procedure Templates.

Data Tab

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

Find

This option specifies the parameter to be solved for from the other parameters.

V0 (Baseline Variance)

Enter one or more value(s) of the baseline variance. This variance will be compared to the alternative variance. It must be greater than zero.

Actually, only the ratio of the two variances (or standard deviations) is used, so you can enter a one here and enter the ratio value in the *V1* box.

If Scale is *Standard Deviation* this value is treated as a standard deviation rather than a variance.

V1 (Alternative Variance)

Enter one or more value(s) of the alternative variance. This variance will be compared to the baseline variance. It must be greater than zero.

Actually, only the ratio of the two variances (or standard deviations) is used, so you can enter a one for *V0* and enter a ratio value here.

If Scale is *Standard Deviation* this value is treated as a standard deviation rather than a variance.

N (Sample Size)

This is the number of observations in the study.

Known Mean

The degrees of freedom of the Chi-square test is N - 1 if the mean is calculated from the data (this is usually the case) or it is N if the mean is known. Check this box if the mean is known. This will cause an increase of the sample size by one.

Scale

Specify whether *V0* and *V1* are variances or standard deviations.

Alternative Hypothesis

This option specifies the alternative hypothesis. This implicitly specifies the direction of the hypothesis test. The null hypothesis is always $H_0: \sigma_0^2 = \sigma_1^2$.

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

Possible selections are:

Ha: V0 <> V1. This selection yields a *two-tailed* test. Use this option when you are testing whether the variances are different but you do not want to specify beforehand which variance is larger.

Ha: V0 > V1. The options yields a *one-tailed* test. Use it when you are only interested in the case in which *V1* is less than *V0*.

Ha: V0 < V1. This option yields a *one-tailed* test. Use it when you are only interested in the case in which *V1* is greater than *V0*.

Alpha

This option specifies one or more values for the significance level--probability of a type-I error (alpha). A type-I error occurs when you reject the null hypothesis when in fact it is true.

Values between 0.001 and 0.100 are common. The value of 0.05 has become the standard. This means that about one test in twenty will falsely reject the null hypothesis when it is actually true. Although 0.05 is the standard value, you should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

Note that you can enter a range of values such as *0.01,0.05* or *0.01 to 0.05 by 0.01*.

Beta (1-Power)

This option specifies one or more values for beta (the probability of accepting a false null hypothesis). Since statistical power is equal to one minus beta, specifying beta implicitly specifies the power. For example, setting beta at 0.20 also sets the power to 0.80.

Values must be between zero and one. The value of 0.20 has often used for beta. However, you should pick a value for beta that represents the risk of this type of error you are willing to take.

Note that you can enter a range of values such as *0.10,0.20* or *0.05 to 0.20 by 0.05*.

If your only interest is in determining the appropriate sample size for a confidence interval, set beta to 0.5.

Options Tab

This tab sets a couple of options used in the iterative procedures.

Maximum Iterations

Specify the maximum number of iterations allowed before the search for the criterion of interest is aborted. When the maximum number of iterations is reached without convergence, the criterion is left blank. A value of 500 is recommended.

Example1 - Calculating the Power

A machine used to perform a particular analysis is to be replaced with a new type of machine if the new machine reduces the variation in the output. The current machine has been tested repeatedly and found to have an output variance of 42.5. The new machine will be cost effective if it can reduce the variance by 30% to 29.75. If the significance level is set to 0.05, calculate the power for sample sizes of 10, 50, 90, 130, 170, 210, and 250.

Setup

You can enter these values yourself or load the Example1 template from the Template tab.

Option	Value
Data Tab	
Find	**Beta and Power**
V0	**42.5**
V1	**29.75**
N	**10, 50, 90, 130, 170, 210, 250**
Known Mean	**Not checked**
Scale	**Variance**
Alternative Hypothesis	**Ha: V0>V1**
Alpha	**0.05**
Beta	**Ignored**

Annotated Output

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

Numeric Results

Numeric Results when H0: V0 = V1 versus Ha: V0>V1

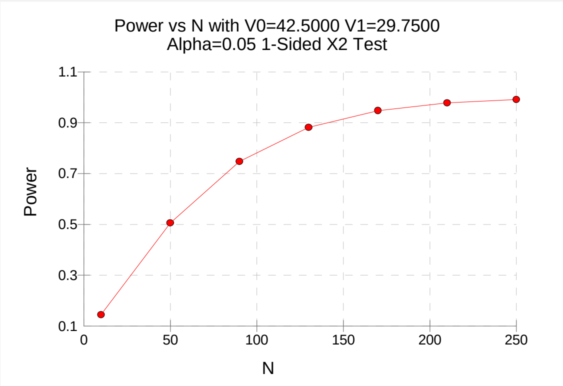
Power	N	V0	V1	Alpha	Beta
0.144479	10	42.5000	29.7500	0.050000	0.855521
0.505556	50	42.5000	29.7500	0.050000	0.494444
0.747747	90	42.5000	29.7500	0.050000	0.252253
0.881740	130	42.5000	29.7500	0.050000	0.118260
0.947851	170	42.5000	29.7500	0.050000	0.052149
0.978059	210	42.5000	29.7500	0.050000	0.021941
0.991108	250	42.5000	29.7500	0.050000	0.008892

Report Definitions
Power is the probability of rejecting a false null hypothesis. It should be close to one.
N is the size of the sample drawn from the population.
V0 is the value of the population variance under the null hypothesis.
V1 is the value of the population variance under the alternative hypothesis.
Alpha is the probability of rejecting a true null hypothesis. It should be small.
Beta is the probability of accepting a false null hypothesis. It should be small.

Summary Statements
A sample size of 10 achieves 14% power to detect a difference of 12.7500 between the null hypothesis variance of 42.5000 and the alternative hypothesis variance of 29.7500 using a one-sided, Chi-square hypothesis test with a significance level (alpha) of 0.050000.

This report shows the calculated power for each scenario.

Plot Section



This plot shows the power versus the sample size. We see that a sample size of about 150 is necessary to achieve a power of 0.90.

Example2 - Calculating Sample Size

Continuing with the previous example, the analyst wants to find the necessary sample sizes to achieve a power of 0.9, for two significance levels, 0.01 and 0.05, and for several variance values. To make interpreting the output easier, the analyst decides to switch from the absolute scale to a ratio scale. To accomplish this, the baseline variance is set at 1.0 and the alternative variances of 0.2, 0.3, 0.4, 0.5, 0.6, and 0.7 are tried.

Setup

You can enter these values yourself or load the Example2 template from the Template tab.

Option	Value
Data Tab	
Find	**N**
V0	**1.0**
V1	**0.2 to 0.7 by 0.1**
N	**Ignored**
Known Mean	**Not checked**
Scale	**Variance**
Alternative Hypothesis	**Ha: V0>V1**
Alpha	**0.05**
Beta	**0.10**

Annotated Output

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

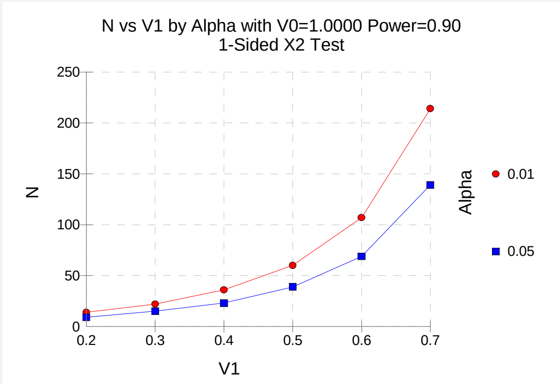
Numeric Results

Numeric Results when H0: V0 = V1 versus Ha: V0>V1

Power	N	V0	V1	Alpha	Beta
0.909022	9	1.0000	0.2000	0.050000	0.090978
0.919352	15	1.0000	0.3000	0.050000	0.080648
0.900670	23	1.0000	0.4000	0.050000	0.099330
0.904235	39	1.0000	0.5000	0.050000	0.095765
0.900775	69	1.0000	0.6000	0.050000	0.099225
0.901171	139	1.0000	0.7000	0.050000	0.098829

This report shows the necessary sample size for each scenario.

Plot Section



This plot shows the necessary sample size for various values of *V1*. Note that as *V1* gets further away from zero, the sample size is increased.

Example3 - Validation using Zar

Zar (1994) page 117 presents an example with *V0* = 1.5, *V1* = 2.6898, *N* = 40, *Alpha* = 0.05, and *Power* = 0.84. We will run this example through ***PASS***.

Setup

You can enter these values yourself or load the Example3 template from the Template tab.

Option	Value
Data Tab	
Find	**Beta and Power**
V0	**1.5**
V1	**2.6898**
N	**40**
Known Mean	**Not checked**
Scale	**Variance**
Alternative Hypothesis	**Ha: V0<V1**
Alpha	**0.05**
Beta	**0.16**

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

Numeric Results

Numeric Results when H0: V0 = V1 versus Ha: V0<V1

Power	N	V0	V1	Alpha	Beta
0.835167	40	1.5000	2.6898	0.050000	0.164833

PASS calculated the power at 0.835167 which matches Zar's result of 0.84 within rounding.

Example4 - Validation using Davies

Davies (1971) page 40 presents an example of determining *N* when (in the standard deviation scale) *V0* = 0.04, *V1* = 0.10, Alpha = 0.05, and Power = 0.99. Davies calculates *N* to be 13. We will run this example through ***PASS***.

Setup

You can enter these values yourself or load the Example4 template from the Template tab.

Option	Value
Data Tab	
Find	**N**
V0	**0.04**
V1	**0.10**
N	**Ignored**
Known Mean	**Not checked**
Scale	**Standard Deviation**
Alternative Hypothesis	**Ha: V0<V1**
Alpha	**0.05**
Beta	**0.01**

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

Numeric Results

Numeric Results when H0: S0 = S1 versus Ha: S0<S1

Power	N	S0	S1	Alpha	Beta
0.992381	13	0.0400	0.1000	0.050000	0.007619

PASS calculated an *N* of 13 which matches Davies' result.

Chapter 655

Two Variances

Introduction

Occasionally, researchers are interested in comparing the variances (or standard deviations) of two groups rather than their means. This module calculates the sample sizes and performs power analyses for hypothesis tests concerning two variances.

Technical Details

Assuming that variables *X1* and *X2* are normally distributed variances σ_1^2 and σ_2^2 (the means are ignored), the distribution of the ratio of the sample variances follows the *F* distribution. That is,

$$F = \frac{s_1^2}{s_2^2}$$

is distributed as an *F* random variable with $N_1 - 1$ and $N_2 - 1$ degrees of freedom. The sample statistic, s_j^2, is calculated as follows

$$s_j^2 = \frac{\sum_{i=1}^{N} \left(X_{ji} - \overline{X}_j \right)^2}{N_j - 1}.$$

The power or sample size of a hypothesis test about the variance can be calculated using the appropriate one of the following three formulas:

Case 1: $H_0\!: \sigma_1^2 = \sigma_2^2 \quad versus \quad H_a\!: \sigma_1^2 \neq \sigma_2^2$

$$\beta = P\left(\frac{\sigma_1^2}{\sigma_2^2} F_{\alpha/2, N1-1, N2-1} < F < \frac{\sigma_1^2}{\sigma_2^2} F_{1-\alpha/2, N1-1, N2-1} \right)$$

Case 2: $H_0\!: \sigma_1^2 = \sigma_2^2 \quad versus \quad H_a\!: \sigma_1^2 > \sigma_2^2$

$$\beta = P\left(F > \frac{\sigma_1^2}{\sigma_2^2} F_{\alpha, N1-1, N2-1} \right)$$

Case 3: $H_0\!: \sigma_1^2 = \sigma_2^2 \quad versus \quad H_a\!: \sigma_1^2 < \sigma_2^2$

$$\beta = P\left(F < \frac{\sigma_1^2}{\sigma_2^2} F_{1-\alpha, N1-1, N2-1} \right)$$

Procedure Options

This section describes the options that are unique to this procedure. These are located on the panels associated with the Data, Options, and Reports tabs. To find out more about using the other tabs such as Plot Text, Axes, and Template, turn to the chapter entitled Procedure Templates.

Data Tab

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

Find

This option specifies the parameter to be solved for from the other parameters.

V1 and V2

Enter one or more value(s) for the variances of the groups, σ_1^2 and σ_2^2. All entries must be greater than zero.

Note that since the ratio of these variances is all that is used in the power equations, you can specify the problem in terms of the variance ratio instead of the two variances. To do this, enter 1.0 for *V2* and enter the desired variance ratio in *V1*.

If Scale is *Standard Deviation* this value is the standard deviation rather than the variance.

Alpha

This option specifies one or more values for the significance level--probability of a type-I error (alpha). A type-I error occurs when you reject the null hypothesis when in fact it is true.

Values between 0.001 and 0.100 are acceptable. The value of 0.05 has become the standard. This means that about one test in twenty will falsely reject the null hypothesis when it is actually true. Although 0.05 is the standard value, you should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

Note that you can enter a range of values such as *0.01,0.05* or *0.01 to 0.05 by 0.01*.

Beta (1-Power)

This option specifies one or more values for beta (the probability of accepting a false null hypothesis). Since statistical power is equal to one minus beta, specifying beta implicitly specifies the power. For example, setting beta at 0.20 also sets the power to 0.80.

Values must be between zero and one. The value of 0.20 has often used for beta. However, you should pick a value for beta that represents the risk of this type of error you are willing to take.

Note that you can enter a range of values such as *0.10,0.20* or *0.05 to 0.20 by 0.05*.

If your only interest is in determining the appropriate sample size for a confidence interval, set beta to 0.5.

Scale

Specify whether *V1* and *V2* are variances or standard deviations.

Alternative Hypothesis

This option specifies the alternative hypothesis. This implicitly specifies the direction of the hypothesis test. The null hypothesis is always $H_0: V_1 = V_2$.

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

Possible selections are:

Ha: V1 <> V2. This selection yields a *two-tailed* test. Use this option when you are testing whether the variances are different but you do not want to specify beforehand which variance is larger.

Ha: V1 > V2. The options yields a *one-tailed* test. Use it when you are only interested in the case in which *V2* is less than *V1*.

Ha: V1 < V2. This option yields a *one-tailed* test. Use it when you are only interested in the case in which *V2* is greater than *V1*.

N1 (Sample Size Group 1)

Enter a value (or range of values) for the sample size of this group. Note that these values are ignored when you are solving for *N1*. You may enter a range of values such as *10 to 100 by 10*.

N2 (Sample Size Group 2)

Enter a value (or range of values) for the sample size of group 2 or enter *Use R* to base *N2* on the value of *N1*. You may enter a range of values such as *10 to 100 by 10*.

Use R

When *Use R* is entered, *N2* is calculated using the formula

$$N2 = [R\ N1]$$

where *R* is the Sample Allocation Ratio and [*Y*] is the first integer greater than or equal to *Y*. For example, if you want *N1 = N2*, select *Use R* here and set *R* = 1.

R (Sample Allocation Ratio)

Enter a value (or range of values) for *R*, the allocation ratio between samples. This value is only used when *N2* is set to *Use R*.

When used, *N2* is calculated from *N1* using the formula: *N2*=[*R N1*] where [*Y*] is the next integer greater than or equal to *Y*. Note that setting *R* = 1.0 forces *N2* = *N1*.

Options Tab

This tab sets a couple of options used in the iterative procedures.

Maximum Iterations

Specify the maximum number of iterations allowed before the search for the criterion of interest is aborted. When the maximum number of iterations is reached without convergence, the criterion is left blank. A value of 500 is recommended.

Example1 - Calculating the Power

A machine used to perform a particular analysis is to be replaced with a new type of machine if the new machine reduces the variance in the output by 50%. If the significance level is set to 0.05, calculate the power for sample sizes of 5, 10, 20, 35, 50, 90, 130, and 200.

Setup

You can enter these values yourself or load the Example1 template from the Template tab.

Option	Value
Data Tab	
Find	**Beta and Power**
V1	**1.0**
V2	**0.5**
Alpha	**0.05**
Beta	**Ignored**
Scale	**Variance**
Alternative Hypothesis	**Ha: V1>V2**
N1	**5, 10, 20, 35, 50, 90, 130, 200**
N2	**Use R**
R	**1.0**

Annotated Output

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

Numeric Results

Numeric Results when H0: V1 = V2 versus Ha: V1>V2

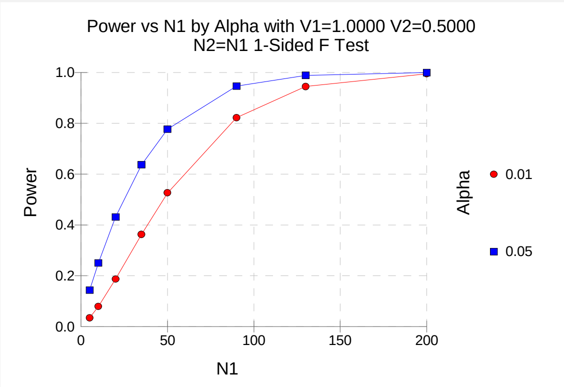
Power	N1	N2	V1	V2	Alpha	Beta
0.034378	5	5	1.0000	0.5000	0.010000	0.965622
0.079391	10	10	1.0000	0.5000	0.010000	0.920609
0.187122	20	20	1.0000	0.5000	0.010000	0.812878
0.362650	35	35	1.0000	0.5000	0.010000	0.637350
0.526210	50	50	1.0000	0.5000	0.010000	0.473790
0.821599	90	90	1.0000	0.5000	0.010000	0.178401
0.944261	130	130	1.0000	0.5000	0.010000	0.055739
0.994526	200	200	1.0000	0.5000	0.010000	0.005474
0.143437	5	5	1.0000	0.5000	0.050000	0.856563
0.250418	10	10	1.0000	0.5000	0.050000	0.749582
0.431043	20	20	1.0000	0.5000	0.050000	0.568957
0.636863	35	35	1.0000	0.5000	0.050000	0.363137
0.776507	50	50	1.0000	0.5000	0.050000	0.223493
0.946023	90	90	1.0000	0.5000	0.050000	0.053977
0.988497	130	130	1.0000	0.5000	0.050000	0.011503
0.999364	200	200	1.0000	0.5000	0.050000	0.000636

Report Definitions
Power is the probability of rejecting a false null hypothesis. It should be close to one.
N1 is the size of the sample drawn from the population 1.
N2 is the size of the sample drawn from the population 2.
V1 is the value of the population variance of group 1.
V2 is the value of the population variance of group 2.
Alpha is the probability of rejecting a true null hypothesis. It should be small.
Beta is the probability of accepting a false null hypothesis. It should be small.

Summary Statements
Group sample sizes of 5 and 5 achieve 3% power to detect a ratio of 2.0000 between the group one variance of 1.0000 and the group two variance of 0.5000 using a one-sided F test with a significance level (alpha) of 0.010000.

This report shows the calculated power for each scenario.

Plot Section



This plot shows the power versus the sample size for the two significance levels. It is now easy to determine an appropriate sample size to meet both the alpha and beta objectives of the study.

Example2 - Calculating Sample Size

Continuing with the previous example, the analyst wants to find the necessary sample sizes to achieve a power of 0.9 for two significance levels, 0.01 and 0.05, and for several variance ratio values of 0.2, 0.3, 0.4, 0.5, 0.6, and 0.7.

Setup

You can enter these values yourself or load the Example2 template from the Template tab.

<u>**Option**</u>	<u>**Value**</u>
Data Tab	
Find	**N1**
V1	**1.0**
V2	**0.2 to 0.7 by 0.1**
Alpha	**0.05**
Beta	**0.10**
Scale	**Variance**
Alternative Hypothesis	**Ha: V0>V1**
N1	**Ignored**
N2	**Use R**
R	**1.0**

Annotated Output

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

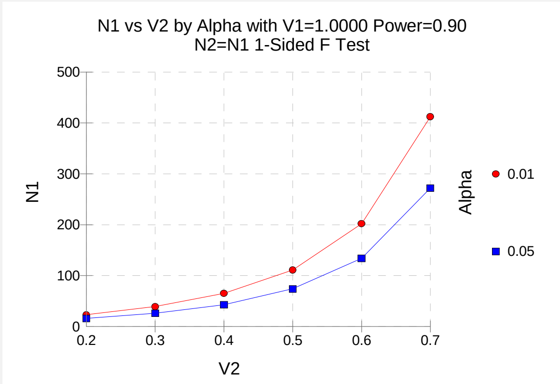
Numeric Results

Numeric Results when H0: V1 = V2 versus Ha: V1>V2

Power	N1	N2	V1	V2	Alpha	Beta
0.911109	23	23	1.0000	0.2000	0.010000	0.088891
0.916243	16	16	1.0000	0.2000	0.050000	0.083757
0.907774	39	39	1.0000	0.3000	0.010000	0.092226
0.905324	26	26	1.0000	0.3000	0.050000	0.094676
0.904075	65	65	1.0000	0.4000	0.010000	0.095925
0.902069	43	43	1.0000	0.4000	0.050000	0.097931
0.901318	111	111	1.0000	0.5000	0.010000	0.098682
0.902949	74	74	1.0000	0.5000	0.050000	0.097051
0.900454	202	202	1.0000	0.6000	0.010000	0.099546
0.901654	134	134	1.0000	0.6000	0.050000	0.098346
0.900417	412	412	1.0000	0.7000	0.010000	0.099583
0.900817	272	272	1.0000	0.7000	0.050000	0.099183

This report shows the necessary sample size for each scenario.

Plot Section



This plot shows the necessary sample size for various values of *V2*. Note that as *V2* nears *V1*, the sample size is increased.

Example3 - Validation Using Davies

Problem Statement

Davies (1971) page 41 presents an example with *V1* = 4, *V1* = 1, *Alpha* = 0.05, and *Power* = 0.99 in which the sample sizes, *N1* and *N2*, are calculated to be 36. We will run this example through ***PASS***.

Setup

You can enter these values yourself or load the Example3 template from the Template tab.

Option	Value
Data Tab	
Find	**N1**
V1	**4**
V2	**1**
Scale	**Variance**
Alternative Hypothesis	**Ha: V1>V2**
Alpha	**0.05**
Beta	**0.01**
N1	**Ignored**
N2	**Use R**
R	**1.0**

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

Numeric Results

Numeric Results when H0: V1 = V2 versus Ha: V1>V2

Power	N1	N2	V1	V2	Alpha	Beta
0.991423	36	36	4.0000	1.0000	0.050000	0.008577

PASS calculated the *N1* and *N2* to be 36 which matches Davies' result.

Chapter 700

Log-Rank Test (Simple)

Introduction

This module allows the sample size and power of the log-rank test to be analyzed under the assumption of proportional hazards. Time periods are not stated. Rather, it is assumed that enough time elapses to allow for a reasonable proportion of responses to occur. If you want to study the impact of accrual and follow-up time, you should use the advanced log-rank module also contained in ***PASS***. The formulas used in this module come from Machin *et al.* (1997).

A clinical trial is often employed to test the equality of survival distributions for two treatment groups. For example, a researcher might wish to determine if Beta-Blocker A enhances the survival of newly diagnosed myocardial infarction patients over that of the standard Beta-Blocker B. The question being considered is whether the pattern of survival is different.

The two-sample t-test is not appropriate for two reasons. First, the data consist of the length of survival (time to failure), which is often highly skewed, so the usual normality assumption cannot be validated. Second, since the purpose of the treatment is to increase survival time, it is likely (and desirable) that some of the individuals in the study will survive longer than the planned duration of the study. The survival times of these individuals are then said to be *censored*. These times provide valuable information, but they are not the actual survival times. Hence, special methods have to be employed which use both regular and censored survival times.

The log-rank test is one of the most popular tests for comparing two survival distributions. It is easy to apply and is usually more powerful than an analysis based simply on proportions. It compares survival across the whole spectrum of time, not at just one or two points.

The power calculations used here assume an underlying exponential distribution. However, we are rarely in a position to assume exponential survival times in an actual clinical trial. How do we justify the exponential survival time assumption? First, the log-rank test and the test derived using the exponential distribution have nearly the same power when the data are in fact exponentially distributed. Second, under the proportional hazards model (which is assumed by the log-rank test), the survival distribution can be transformed to be exponential and the log-rank test remains the same under monotonic transformations.

Technical Details

In order to define the input parameters, we will present below some rather complicated looking formulas. You need not understand the formulas. However, you should understand the individual parameters used in these formulas.

We assume that a study is to be made comparing the survival of an existing (control) group with an experimental group. The control group consists of patients that will receive the existing treatment. In cases where no existing treatment exists, the control group consists of patients that will receive a placebo. This group is arbitrarily called group one. The experimental group will receive the new treatment. It is called group two.

We assume that the critical event of interest is death and that two treatments have survival distributions with instantaneous death (hazard) rates, λ_1 and λ_2. These hazard rates are an subject's probability of death in a short period of time. We want to test hypotheses about these hazard rates.

Hazard Ratio

There are several ways to express the difference between two hazard rates. One way is to calculate the difference, $\lambda_1 - \lambda_2$. Another way is to form the hazard ratio (*HR*):

$$HR = \frac{\lambda_2}{\lambda_1}.$$

Note that since HR is formed by dividing the hazard rate of the experimental group by that of the control group, a treatment that does better than the control will have a hazard ratio that is less than one.

The hazard ratio may be formulated in other ways. If the proportions surviving during the study are called *S1* and *S2* for the control and experimental groups, the hazard ratio is given by

$$HR = \frac{Log(S2)}{Log(S1)}.$$

Furthermore, if the median survival times of the two groups are *M1* and *M2*, the hazard ratio is given by

$$HR = \frac{M1}{M2}.$$

Each of these expressions for the difference between hazards rates is useful in some situations. There is no one best way, so you will have to be a little flexible.

Log-Rank Test

We assume that the log-rank test will be used to analyze the data once they are collected. Basing the power calculations on the log-rank test, we arrive at the following formula that gives the power based on several other parameters:

$$z_{1-\beta} = \frac{|HR-1|\sqrt{N(1-w)\varphi\left[(1-S1)+\varphi(1-S2)\right]/(1+\varphi)}}{(1+\varphi HR)} - z_{1-\alpha/k}$$

where k is 1 for a one-sided hypothesis test or 2 for a two-sided test, α and β are the error rates defined as usual, the z's are the usual points from the standard normal distribution, w is the proportion that are lost to follow up, and φ represents the sample size ratio between the two groups. That is, is *p1* is the proportion of the total sample size that is in the control group, φ is given by

$$\varphi = \frac{1-p1}{p1}$$

Note that the hypothesis being tested is that the hazard rates are equal:

$$H_0{:}\lambda_1 = \lambda_2$$

This formulation assumes that the hazard rates are proportional. It does not assume that the failure times are exponentially distributed.

Procedure Options

This section describes the options that are unique to this procedure. These are located on the panels associated with the Data, Options, and Reports tabs. To find out more about using the other tabs such as Plot Text, Axes, and Template, turn to the chapter entitled Procedure Templates.

Data Tab

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

Find

This option specifies the parameter to be solved for from the other parameters. The parameters that may be selected are Proportion Surviving 1, Proportion Surviving 2, *N*, Alpha, or Beta. Under most situations, you will select either Beta or *N*.

Select *N* when you want to calculate the sample size needed to achieve a given power and alpha level.

Select *Beta* when you want to calculate the power of an experiment that has already been run.

If you select Survive 1, the search is made for values that are less then *S2*. Likewise, if you select Survive 2, the search is made for values that are greater than *S1*.

S1, Proportion Surviving 1

S1 is the proportion of patients belonging to group 1 (controls) that are expected to survive during the study. Since *S1* is a proportion, it must be between zero and one. A value for *S1* must be determined either from a pilot study or from previous studies.

S2, Proportion Surviving 2

S2 is the proportion of patients belonging to group 2 (experimental) that are expected to survive during the study. Since *S2* is a proportion, it must be between zero and one.

This value is not necessarily the expected survival proportion under the treatment. Rather, you may set it to that value that, if achieved, would be of special interest. Values below this amount would not be of interest.

For example, if the standard 1-year survival proportion is 0.2 and the new treatment raises this proportion to 0.3 (a 50% increase in the proportion surviving), others may be interested in using it.

Sometimes, researchers wish to state the alternative hypothesis in terms of the hazard ratio, *HR*, rather than the value of *S2*. Using the fact that

$$HR = Log(S2)/Log(S1)$$

we can solve for *S2* to obtain

$$S2 = Exp(Log(S1)HR).$$

Using this equation, you can determine an appropriate value for *S2* from a value for *HR*.

For example, suppose *S1* is 0.4 and you have determined that if the hazard rate is reduced by half, the improvement is sufficient to justify a change in treatment. The appropriate value for *S2* is Exp(Log(0.4)(0.5)) = 0.632.

When you do not have an anticipated value for *HR*, you can find an estimate base on the Median Survival Times (*M1* and *M2*) since *HR = M1/M2*.

Total Sample Size (N)

The combined sample size of both groups.

Proportion in Group 1

The proportion of the sample size (*N*) that is assigned to the first group. The rest of the patients are assigned to the second group. When exact values are not available, the number in group one is found by rounding up to the next integer. The number in group two is found by subtraction.

Alpha

This option specifies one or more values for the probability of a type-I error, alpha. This is also called the *significance level* or *test size*. A type-I error occurs when you reject the null hypothesis of equal survival curves when in fact the curves are equal.

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

The check box lets you indicate whether you want to run a one-sided or two-sided test. In most cases, you should run a two-sided test unless you have a legitimate reason to run the one-sided test.

Beta (1-Power)

This option specifies one or more values for the probability of a type-II error (beta). A type-II error occurs when you fail to reject the null hypothesis of survival curves when in fact the curves are different.

Values must be between zero and one. Historically, the value of 0.20 was often used for beta. Now, 0.10 is becoming common. You should pick a value that represents the risk of a type-II error you are willing to take.

Power is defined as one minus beta. Power is equal to the probability of rejecting a false null hypothesis. Hence, specifying the beta error level also specifies the power level. For example, if you specify beta values of 0.05, 0.10, and 0.20, you are specifying the corresponding power values of 0.95, 0.90, and 0.80.

Example 1 - Finding the Power

An experiment has been conducted to test the effectiveness of a new treatment on a total of 100 patients. The current treatment for this disease achieves 50% survival after two years. The proportion in the treatment group that survived two years was 0.70. Testing was done at the 0.05 significance level. Even though there was an increase of 0.20 for 0.50 to 0.70, the log-rank test did not reject the hypothesis of equal hazard rates for the two treatments. Study the power of this test.

Setup

You can make these changes directly on your screen or you can load the template entitled Example1 by clicking the Template tab and loading this template.

Option	Value
Data Tab	
Find	**Beta and Power**
S1	**0.5**
S2	**0.7**
One-Sided	**Not Checked**
Alpha	**0.05 0.10**
N	**50 to 300 by 50**
Proportion in Group 1	**0.5**
Proportion Lost in Follow Up	**0**
Beta	*Ignored*

Annotated Output

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

Numeric Results

Numeric Results

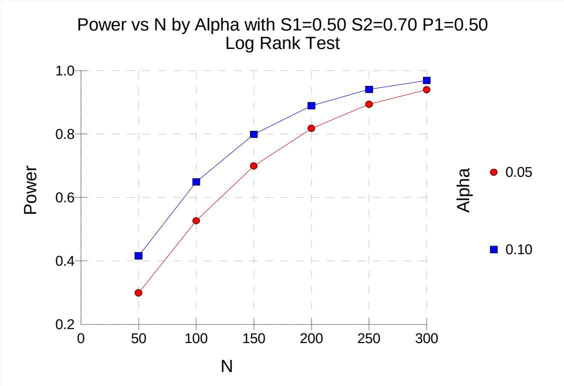
Power	N	N1	N2	S1	S2	Hazard Ratio	Two-Sided Alpha	Beta
0.2992	50	25	25	0.5000	0.7000	0.5000	0.0500	0.7008
0.4162	50	25	25	0.5000	0.7000	0.5000	0.1000	0.5838
0.5267	100	50	50	0.5000	0.7000	0.5000	0.0500	0.4733
0.6488	100	50	50	0.5000	0.7000	0.5000	0.1000	0.3512
0.6994	150	75	75	0.5000	0.7000	0.5000	0.0500	0.3006
0.7989	150	75	75	0.5000	0.7000	0.5000	0.1000	0.2011
0.8177	200	100	100	0.5000	0.7000	0.5000	0.0500	0.1823
0.8891	200	100	100	0.5000	0.7000	0.5000	0.1000	0.1109
0.8934	250	125	125	0.5000	0.7000	0.5000	0.0500	0.1066
0.9406	250	125	125	0.5000	0.7000	0.5000	0.1000	0.0594
0.9395	300	150	150	0.5000	0.7000	0.5000	0.0500	0.0605
0.9690	300	150	150	0.5000	0.7000	0.5000	0.1000	0.0310

Report Definitions
Power is the probability of rejecting a false null hypothesis.
N is the combined sample size.
N1 sample size in group 1.
N2 sample size in group 2.
S1 is the proportion surviving in group 1
S2 is the proportion surviving in group 2
The Hazard Ratio is the ratio of hazard2 and hazard1. It is Log(S2)/Log(S1).
Alpha is the probability of rejecting a true null hypothesis.
Beta is the probability of accepting a false null hypothesis.

Summary Statements
A two-sided log rank test with an overall sample size of 50 subjects (of which 25 are in group 1 and 25 are in group 2) achieves 30% power at a 0.0500 significance level to detect a difference of 0.2000 between 0.5000 and 0.7000--the proportions surviving in groups 1 and 2, respectively. This corresponds to a hazard ratio of 0.5000. The proportion of patients lost during follow up was 0.0000. These results are based on the assumption that the hazard rates are proportional.

This report shows the values of each of the parameters, one scenario per row. The values from this table are in the chart below.

Plots Section



This plot shows the relationship between alpha and power in this example. We notice that for alpha = 0.05, a power of 0.80 is reached when the sample size is about 200. A power of 90% is reached when the sample size is 250. Hence, we realize that this study should have had at least twice the number of patients that it had.

Example 2 - Finding the Sample Size

Continuing with our example, the researcher decides that he wants to do it right this time. He believes that if the survival proportion of the treatment group is 0.6 or better, people will begin to use his treatment. He wants to know how many subjects he needs.

Setup

You can enter these values yourself or load the Example2 template from the Template tab.

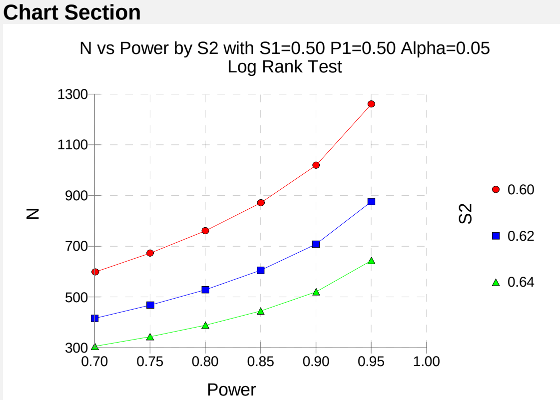
Option	Value
Data Tab	
Find	**N**
S1	**0.5**
S2	**0.60 0.62 0.64**
One-Sided	**Not Checked**
Alpha	**0.05**
N	*Ignored*
Proportion in Group 1	**0.5**
Proportion Lost in Follow Up	**0**
Beta	**0.05 to 0.30 by 0.05**

Annotated Output

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

Results

Chart Section



We will consider the chart since that allows us to understand the patterns more quickly. We note that changing *S2* from 0.60 (the top line) to 0.62 (the middle line) decreases the sample size requirements by almost half. Our researcher decides to take a sample of 500 patients. This will achieve almost 80% power in detecting a shift from 0.50 to 0.62 in survivability.

Example 3 - Validation using Machin

Machin *et al.* (1997) page 180 gives an example in which *S1* is 0.25, *S2* is 0.50, the one-sided significance level is 0.05, and the power is 90%. The total sample size is 124. We will now run this example through ***PASS***.

Setup

You can enter these values yourself or load the Example3 template from the Template tab.

Option	Value
Data Tab	
Find	**N**
Proportion Surviving 1	**0.25**
Proportion Surviving 2	**0.50**
One-Sided	**Checked**
Alpha	**0.05**
N	*Ignored*
Proportion in Group 1	**0.5**
Proportion Lost in Follow Up	**0**
Beta	**0.1**

Annotated Output

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

Results

Numeric Results

Power	N	N1	N2	S1	S2	Hazard Ratio	One-Sided Alpha	Beta
0.9014	124	62	62	0.2500	0.5000	0.5000	0.0500	0.0986

PASS also calculated the total sample size to be 124.

Chapter 705

Log-Rank Survival Test (Advanced)

Introduction

This module computes the sample size and power of the log-rank test for equality (or non-inferiority) of survival distributions under the assumption of proportional hazards. Accrual time and follow-up time are included among the parameters to be set.

A clinical trial is often employed to test the equality of survival distributions for two treatment groups. For example, a researcher might wish to determine if Beta-Blocker A enhances the survival of newly diagnosed myocardial infarction patients over that of the standard Beta-Blocker B. The question being considered is whether the pattern of survival is different.

The two-sample t-test is not appropriate for two reasons. First, the data consist of the length of survival (time to failure), which is often highly skewed, so the usual normality assumption cannot be validated. Second, since the purpose of the treatment is to increase survival time, it is likely (and desirable) that some of the individuals in the study will survive longer than the planned duration of the study. The survival times of these individuals are then said to be *censored*. These times provide valuable information, but they are not the actual survival times. Hence, special methods have to be employed which use both regular and censored survival times.

The log-rank test is one of the most popular tests for comparing two survival distributions. It is easy to apply and is usually more powerful than an analysis based simply on proportions. It compares survival across the whole spectrum of time, not just at one or two points. This module allows the sample size and power of the log-rank test to be analyzed under very general conditions.

Power and sample size calculations for the log-rank test have been studied by several authors. ***PASS*** uses the methods of Lachin and Foulkes (1986) because of their generality. Although small differences among methods can be found depending upon which assumptions are adopted, there is little practical difference between the techniques.

The power calculations used here assume an underlying exponential distribution. However, this assumption is rarely accurate in an actual clinical trial. How is this assumption justified? First, the log-rank test and the test derived using the exponential distribution have nearly the same power when the data are in fact exponentially distributed. Second, under the proportional hazards model (which is assumed by the log-rank test), the survival distribution can be transformed to exponential and the log-rank test remains the same under monotonic transformations.

Technical Details

Some rather complicated formulas are used to define the input parameters. You need not understand the formulas. However, you should understand the individual parameters used in these formulas.

Basic Model

Suppose a clinical trial consists of two independent treatment groups labeled '1' and '2' (you could designate group one as the control group and group two as the treatment group). If the total sample size is *N*, the sizes of the two groups are n_1 and n_2. Usually, you would plan to have $n_1 = n_2$. Define the proportion of the total sample in each group as

$$Q_i = \frac{n_i}{N}, \quad i = 1,2$$

Individuals are recruited during an accrual period of *R* years (or months or days). They are followed for an additional period of time until a total of *T* years is reached. Hence, the follow-up period is *T*-*R* years. At the end of the study, the log-rank test is conducted at significance level α with power $1-\beta$.

If we assume an exponential model with hazard rates λ_1 and λ_2 for the two groups, Lachin and Foulkes (1986, Eq. 2.1) establish the following equation relating *N* and power:

$$\sqrt{N}\left|\lambda_1 - \lambda_2\right| = Z_\alpha \sqrt{\phi(\bar{\lambda})\left(\frac{1}{Q_1}+\frac{1}{Q_2}\right)} + Z_\beta \sqrt{\frac{\phi(\lambda_1)}{Q_1}+\frac{\phi(\lambda_2)}{Q_2}}$$

where

$$\bar{\lambda} = Q_1\lambda_1 + Q_2\lambda_2$$

$$\phi(\lambda) = N\sigma^2\left(\hat{\lambda}\right)$$

$$Z_\theta = \Phi(1-\theta)$$

$\Phi(z)$ is the area to the left of *z* under the standard normal density, $\hat{\lambda}$ is the maximum likelihood estimate of λ, and that $\sigma^2\left(\hat{\lambda}\right)$ represents the variance of $\hat{\lambda}$.

Exponential Distribution

The hazard rate from the exponential distribution, λ, is usually estimated using maximum likelihood techniques. In the planning stages, you have to obtain an estimate of this parameter. To see how to accomplish this, let's briefly review the exponential distribution. The density function of the exponential is defined as

$$f(t) = \lambda e^{-\lambda t}, \quad t \geq 0, \lambda > 0.$$

The cumulative survival distribution function is

$$S(t) = e^{-\lambda t}, \quad t \geq 0.$$

Solving this for λ yields

$$\lambda = -\frac{\log(S(t))}{t}$$

Note that $S(t)$ gives the probability of surviving t years. To obtain a planning estimate of λ, you need only know the proportion surviving during a particular time period. You can then use the above equation to calculate λ.

Patient Entry

Patients are enrolled during the accrual period. ***PASS*** lets you specify the pattern in which subjects are enrolled. Suppose patient entry times are distributed as $g(t)$ where t_i is the entry time of the i^{th} individual and $0 \leq t_i \leq R$. Let $g(t)$1 follow the truncated exponential distribution with parameter A, which has the density

$$g(t) = \begin{cases} \dfrac{Ae^{-At}}{1-e^{-AR}}, & 0 \leq t \leq R, \quad A \neq 0 \\ 1/R, & 0 \leq t \leq R, \quad A = 0 \end{cases}$$

Note that R is accrual time. The corresponding cumulative distribution function is

$$G(t) = \begin{cases} \dfrac{1-e^{-At}}{1-e^{-AR}}, & 0 \leq t \leq R, \quad A \neq 0 \\ t/R, & 0 \leq t \leq R, \quad A = 0 \end{cases}$$

A is interpreted as follows.

$A > 0$ results in a convex (faster than expected) entry distribution.

$A < 0$ results in a concave (slower than expected) entry distribution.

$A = 0$ results in the uniform entry distribution in which $g(t) = 1/R$.

Rather than specify A directly, ***PASS*** has you enter the percentage of the accrual time that will be needed to enroll 50% of the subjects. Using an iterative search, the value of A corresponding to this percentage is calculated and used in the calculations.

Losses to Follow-Up

The staggered patient entry over the accrual period results in censoring times ranging from *T - R* to *T* years during the follow-up period. This is often referred to as administrative censoring, since it is caused by the conclusion of the study rather than by some random factor working on an individual. To model the losses to follow-up which come from other causes, we use the exponential distribution with hazard rates η_1 and η_2. Since these rates are difficult and confusing to define directly, ***PASS*** lets you input the proportion lost due to other causes over a specified period of time and uses the following equation to determine the hazard rates:

$$\eta = -\frac{\log(1-P(t))}{t}$$

General Model

Combining all these parameters into the model results in

$$\phi(\lambda,\eta,\gamma) = \lambda^2\left(\frac{\lambda}{\lambda+\eta} + \frac{\lambda\gamma e^{-(\lambda+\eta)T}\left[1-e^{(\lambda+\eta-\gamma)R}\right]}{\left(1-e^{-\gamma R}\right)(\lambda+\eta)(\lambda+\eta-\gamma)}\right)^{-1}.$$

This expression may then be used in an equation that relates these parameters to sample size and power

$$\sqrt{N}|\lambda_1-\lambda_2| = Z_\alpha\sqrt{\phi(\bar{\lambda})\left(\frac{1}{Q_1}+\frac{1}{Q_2}\right)} + Z_\beta\sqrt{\frac{\phi(\lambda_1)}{Q_1}+\frac{\phi(\lambda_2)}{Q_2}}$$

Procedure Options

This section describes the options that are unique to this procedure. These are located on the panels associated with the Data, Options, and Reports tabs. To find out more about using the other tabs such as Plot Text, Axes, and Template, turn to the chapter entitled Procedure Templates.

Data Tab

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

Find

This option specifies the parameter to be solved for from the other parameters. The parameters that may be selected are *Proportion Surviving 1*, *Proportion Surviving 2*, *Accrual*, *Follow-Up*, *Alpha*, *Beta*, and *N*. Under most situations, you will select either *Beta* or *N*.

Select *N* when you want to calculate the sample size needed to achieve a given power and alpha level.

Select *Beta* when you want to calculate the power.

S1 and S2 (Proportion Surviving Past T0)

Specify the proportion of patients in each group that survive until after *T0*. These quantities are called *S1* and *S2*. The value of *T0* is given in the Fixed Time Point box. For example, if the Fixed Time parameter is 3 and this value is set to 0.7, then 70% of the patients survive at least 3 years. Note that since this is a proportion, values must be between zero and one.

These quantities are used as a convenient method of entering the hazard rates λ_1 and λ_2. The first group is arbitrarily designated as the control group and the second group is arbitrarily designated as the treatment group.

The hazard rates are calculated from the proportions surviving using the formulas

$$\lambda_1 = -\frac{\log(S1)}{T0} \text{ and } \lambda_2 = -\frac{\log(S2)}{T0}$$

If your sample size problem is cast in terms of hazard rates, the proportions surviving are calculated using the formula

$$S1 = \exp(-\lambda_1 t) \text{ and } S2 = \exp(-\lambda_2 t)$$

Accrual Time

The accrual time is the length of time during which patients enter the study. It is the value of *R*.

% Time Until 50% Accrual

This specifies the percentage of the accrual time needed to enroll 50% of the patients. This value is converted into a value for *A*, the patient entry parameter. Use this option to indicate how subjects are enrolled during the accrual period. For example, in one study, it may be possible to enroll a number of patients early on, while in another study, most of the subjects will be enrolled near the end of the accrual period.

Values between 1 and 97 may be entered.

If you expect uniform patient entry, enter 50. Unless you know that patient enrollment will not be uniform during the accrual period, you should use this amount.

If you expect more patients to enter during the early part of the accrual period, enter an amount less than 50 such as 30. A 30 here means that 50% of the patients will have been enrolled when 30% of the accrual time has elapsed.

If you expect more patients to enter during the later part of the accrual period, enter an amount greater than 50 such as 70. A 70 here means that 50% of the patients will have been enrolled when 70% of the accrual time has elapsed.

PASS assumes that patient entry times follow the truncated exponential distribution. This parameter controls the shape and scale of that distribution.

Follow-Up Time

The *follow-up time* is the length of time between the entry of the last individual into the study and the end of the study. Since *T* is the total length of the study and *R* is the accrual time, the follow-up time is *T-R*.

Follow-Up Loss (1 and 2)

This is the proportion in this group that is lost to follow-up during the Fixed Time Period, *T0*. Using this value, the lost-to-follow-up hazard rate is calculated using the exponential distribution as it is with *S*(*t*). The equation used to convert these proportions into η_1 and η_2 is:

$$\eta_i = -\frac{\log\left(1 - P_i\left(T0\right)\right)}{T0}$$

Here *P*(*T0*) is the proportion lost to follow up from the beginning of the study until *T0*. So if *T0* is 3 years and 10% of the patients in group one are lost to follow-up

$$\begin{aligned}\eta_1 &= -\log(1.0 - 0.1) / 3 \\ &= 0.03512\end{aligned}$$

Values between zero and one are valid. Zero is used to indicate no loss to follow-up.

Alternative Hypothesis

This option specifies the alternative hypothesis in terms of the proportion surviving in each group. This implicitly specifies the direction of the hypothesis test. The null hypothesis is always $H_0: S_1 = S_2$.

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

Possible selections are:

Ha: S_1 <> S_2. This is the usual selection. It yields the *two-tailed* t-test. Use this option when you are testing whether the survival curves are different, but you do not want to specify beforehand which curve is better.

Ha: S_1 < S_2. This option yields a *one-tailed* test. Use it when you are only interested in the case in which the survival in group one is less than that for group two.

Ha: S_1 > S_2. This option yields a *one-tailed* test. Use it when you are only interested in the case in which the survival in group one is greater than that for group two.

T0 (Fixed Time Point)

This is the time period used to convert the proportions given in Proportion Surviving 1, Proportion Surviving 2, Follow-Up Loss1, and Follow-Up Loss2 into hazard rates. If you enter *0.4* in Proportion Survive1 and a *3* here, you are indicating that 40% survive longer than 3 years.

N (Total Sample Size)

This is the combined sample size of both groups. This amount is divided between the two groups using the value of the Proportion in Group 1.

Proportion in Group 1

This is the proportion of *N* in group one. If this value is labeled p_1, the sample size of group one is Np_1 and the sample size of group two is $N - Np_1$. Note that the value of Np_1 is rounded to the nearest integer.

Alpha (Significance Level)

This option specifies one or more values for the probability of a type-I error (alpha). A type-I error occurs when you reject the null hypothesis of equal survival curves when in fact the curves are equal.

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

Beta (1 - Power)

This option specifies one or more values for the probability of a type-II error (beta). A type-II error occurs when you fail to reject the null hypothesis of equal survival curves when in fact the curves are different.

Values must be between zero and one. Historically, the value of 0.20 was often used for beta. Now, 0.10 is becoming more popular. However, you should pick a value for beta that represents the risk of a type-II error you are willing to take.

Power is defined as one minus beta. Power is equal to the probability of rejecting a false null hypothesis. Hence, specifying the beta error level also specifies the power level. For example, if you specify beta values of 0.05, 0.10, and 0.20, you are specifying the corresponding power values of 0.95, 0.90, and 0.80, respectively.

Example 1 - Finding the Power

A researcher is planning a clinical trial using a parallel, two-group, equal sample allocation design to compare the survivability of a new treatment with that of the current treatment. The proportion surviving one-year after the current treatment is 0.50. The new treatment will be adopted if the proportion surviving after one year increases to 0.75.

The trial will include a recruitment period of one-year after which participants will be followed for an additional two-years. It is assumed that patients will enter the study uniformly over the accrual period. The researcher estimates a loss-to-follow rate of 15% during the first year in both the control and the experimental groups.

The researcher decides to investigate various sample sizes between 10 and 250 at both the 0.01 and 0.05 significance levels.

Setup

This section presents the values of each of the parameters needed to run this example. You can make these changes directly on your screen or you can load the template entitled Example1 by clicking the Template tab and loading this template.

Option	Value
Data Tab	
Find	**Beta and Power**
S1	**0.5**
S2	**0.75**
R	**1**
% Time Until 50% Accrual	**50**
Follow-Up Time	**2**
Proportion Lost to Follow-Up 1	**0.15**
Proportion Lost to Follow-Up 2	**0.15**
Alternative Hypothesis	**Ha: S1 <> S2**
T0	**1**
N	**10 25 50 100 150 200 250**
Proportion in Group 1	**0.5**
Alpha	**0.01 0.05**
Beta	*Ignored since this is the Find setting*

Annotated Output

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

Numeric Results

Numeric Results with Ha: S1<>S2

Power	N	Surviving Group 1	Surviving Group 2	Accrual Time	Follow Up Time	Prop in Group 1	Alpha	Beta
0.06718	10	0.50000	0.75000	1.00	2.00	0.50000	0.01000	0.93282
0.18406	10	0.50000	0.75000	1.00	2.00	0.50000	0.05000	0.81594
0.17527	25	0.50000	0.75000	1.00	2.00	0.50000	0.01000	0.82473
0.36633	25	0.50000	0.75000	1.00	2.00	0.50000	0.05000	0.63367
0.38357	50	0.50000	0.75000	1.00	2.00	0.50000	0.01000	0.61643
0.61606	50	0.50000	0.75000	1.00	2.00	0.50000	0.05000	0.38394
0.72756	100	0.50000	0.75000	1.00	2.00	0.50000	0.01000	0.27244
0.88428	100	0.50000	0.75000	1.00	2.00	0.50000	0.05000	0.11572
0.90273	150	0.50000	0.75000	1.00	2.00	0.50000	0.01000	0.09727
0.97052	150	0.50000	0.75000	1.00	2.00	0.50000	0.05000	0.02948
0.96998	200	0.50000	0.75000	1.00	2.00	0.50000	0.01000	0.03002
0.99328	200	0.50000	0.75000	1.00	2.00	0.50000	0.05000	0.00672
0.99167	250	0.50000	0.75000	1.00	2.00	0.50000	0.01000	0.00833
0.99858	250	0.50000	0.75000	1.00	2.00	0.50000	0.05000	0.00142

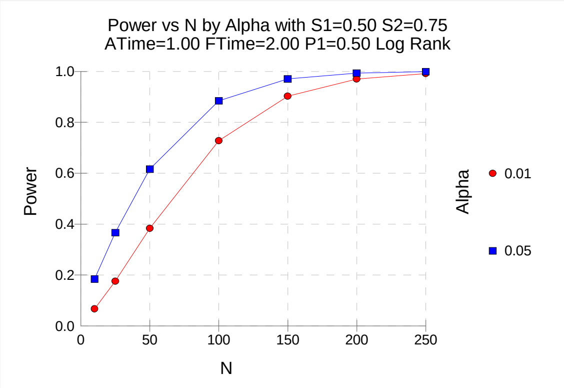
Base Time 1.00
Proportion loss to follow up in group 1 0.15000
Proportion loss to follow up in group 2 0.15000
Percent of accrual time until 50% enrollment is reached: 50.00%

Report Definitions
Power is the probability of rejecting a false null hypothesis. Power should be close to one.
Alpha is the probability of rejecting a true null hypothesis. It should be small.
Beta is the probability of accepting a false null hypothesis. It should be small.

Summary Statements
A two-sided log rank test with an overall sample size of 10 subjects (of which 5 are in group 1 and 5 are in group 2) achieves 7% power at a 0.01000 significance level to detect a difference of 0.25000 between 0.50000 and 0.75000--the proportions surviving in groups 1 and 2 after 1.00 time periods. Patients entered the study during an accrual period of 2.00 time periods. 50% of the enrollment was complete when 50.00% of the accrual time had past. A follow-up period of 1.00 time periods had a 15.0% loss from group 1 and a 15.0% loss from group 2.

This report shows the values of each of the parameters, one scenario per row. Note that approximately 100 patients, 50 per group, will be needed to achieve about 90% power at the 0.05 significance level.

Plots Section



This plot shows the relationship between sample size and power for the two significance levels.

Example 2 - Finding the Sample Size

Continuing with the previous example, the researcher wants to investigate the sample size necessary to achieve 80% or 90% power for various values of the proportion surviving in the treatment group from 0.55 to 0.80 at the 0.05 significance level. All other parameters will remain the same.

Setup

This section presents the values of each of the parameters needed to run this example. You can make these changes directly on your screen or you can load the template entitled Example2 by clicking the Template tab and loading this template.

Option	Value
Data Tab	
Find	**Beta and Power**
S1	**0.5**
S2	**0.55 to 0.80 by 0.05**
R	**1**
% Time Until 50% Accrual	**50**
Follow-Up Time	**2**
Proportion Lost to Follow-Up 1	**0.15**
Proportion Lost to Follow-Up 2	**0.15**
Alternative Hypothesis	**Ha: S1 <> S2**
T0	**1**
N	*Ignored since this is the Find setting*
Proportion in Group 1	**0.5**
Alpha	**0.05**
Beta	**0.2 0.1**

Annotated Output

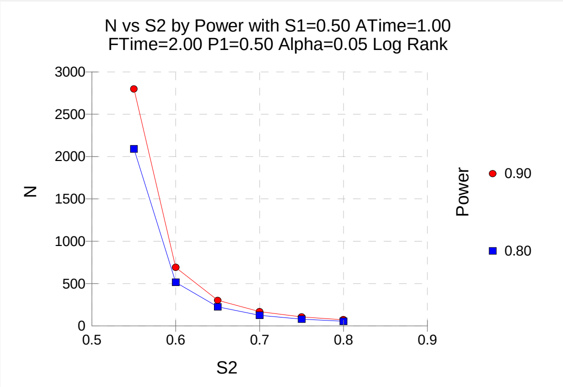
Click the Run button to perform the calculations and generate the following output.

Numeric Results

Numeric Results with Ha: S1<>S2

Power	N	Surviving Group 1	Surviving Group 2	Accrual Time	Follow Up Time	Prop in Group 1	Alpha	Beta
0.90004	2798	0.50000	0.55000	1.00	2.00	0.50000	0.05000	0.09996
0.80017	2090	0.50000	0.55000	1.00	2.00	0.50000	0.05000	0.19983
0.90024	690	0.50000	0.60000	1.00	2.00	0.50000	0.05000	0.09976
0.80050	515	0.50000	0.60000	1.00	2.00	0.50000	0.05000	0.19950
0.90001	302	0.50000	0.65000	1.00	2.00	0.50000	0.05000	0.09999
0.80010	225	0.50000	0.65000	1.00	2.00	0.50000	0.05000	0.19990
0.90098	168	0.50000	0.70000	1.00	2.00	0.50000	0.05000	0.09902
0.80177	125	0.50000	0.70000	1.00	2.00	0.50000	0.05000	0.19823
0.90107	106	0.50000	0.75000	1.00	2.00	0.50000	0.05000	0.09893
0.80357	79	0.50000	0.75000	1.00	2.00	0.50000	0.05000	0.19643
0.90274	73	0.50000	0.80000	1.00	2.00	0.50000	0.05000	0.09726
0.80432	54	0.50000	0.80000	1.00	2.00	0.50000	0.05000	0.19568

Base Time1.00
Proportion loss to follow up in group 10.15000
Proportion loss to follow up in group 20.15000
Percent of accrual time until 50% enrollment is reached: 50.00%



This study shows the huge increase in sample size necessary to detect values of *S2* below 0.65. It also shows that roughly 35% more participants are required for 90% power than for 80% power in this situation.

Example 3 - Validation using Lachin and Foulkes

Lachin and Foulkes (1986), the developers of formulas used in this routine, give an example on page 509 in which $\lambda_1 = 0.30$, $\lambda_2 = 0.20$, $N = 378$, $\alpha = 0.05$ (one-sided), $R = 3$, $T = 3$, and $\beta = 0.10$. There is no loss to follow up and uniform patient entry is assumed.

The first step is to determine the proportion surviving at the end of one year for each group using the formula:

$$S(t) = e^{-\lambda t}$$

For group 1 we have

$$\begin{aligned} S_1(1) &= e^{-0.3(1)} \\ &= 0.74081822 \end{aligned}$$

For group 2 we have

$$\begin{aligned} S_2(1) &= e^{-0.2(1)} \\ &= 0.81873075 \end{aligned}$$

Setup

This section presents the values of each of the parameters needed to run this example. You can make these changes directly on your screen or you can load the template entitled Example3 by clicking the Template tab and loading this template.

Option	Value
Data Tab	
Find	**Beta and Power**
S1	**0.74081822**
S2	**0.81873075**
R	**3**
% Time Until 50% Accrual	**50**
Follow-Up Time	**2**
Follow-Up Loss (1 and 2)	**0.0**
Alternative Hypothesis	**Ha: S1 < S2**
T0	**1**
N	**378**
Proportion in Group 1	**0.5**
Alpha	**0.05**
Beta	*Ignored since this is the Find Setting*

Annotated Output

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

Numeric Results

Numeric Results with Ha: S1<S2

Power	N	Surviving Group 1	Surviving Group 2	Accrual Time	Follow Up Time	Prop in Group 1	Alpha	Beta
0.90123	378	0.74082	0.81873	3.00	2.00	0.50000	0.05000	0.09877

Base Time1.00
Proportion loss to follow up in group 10.00000
Proportion loss to follow up in group 20.00000
Percent of accrual time until 50% enrollment is reached: 50.00%

The power of 0.90 matches the value published in Lachin's article.

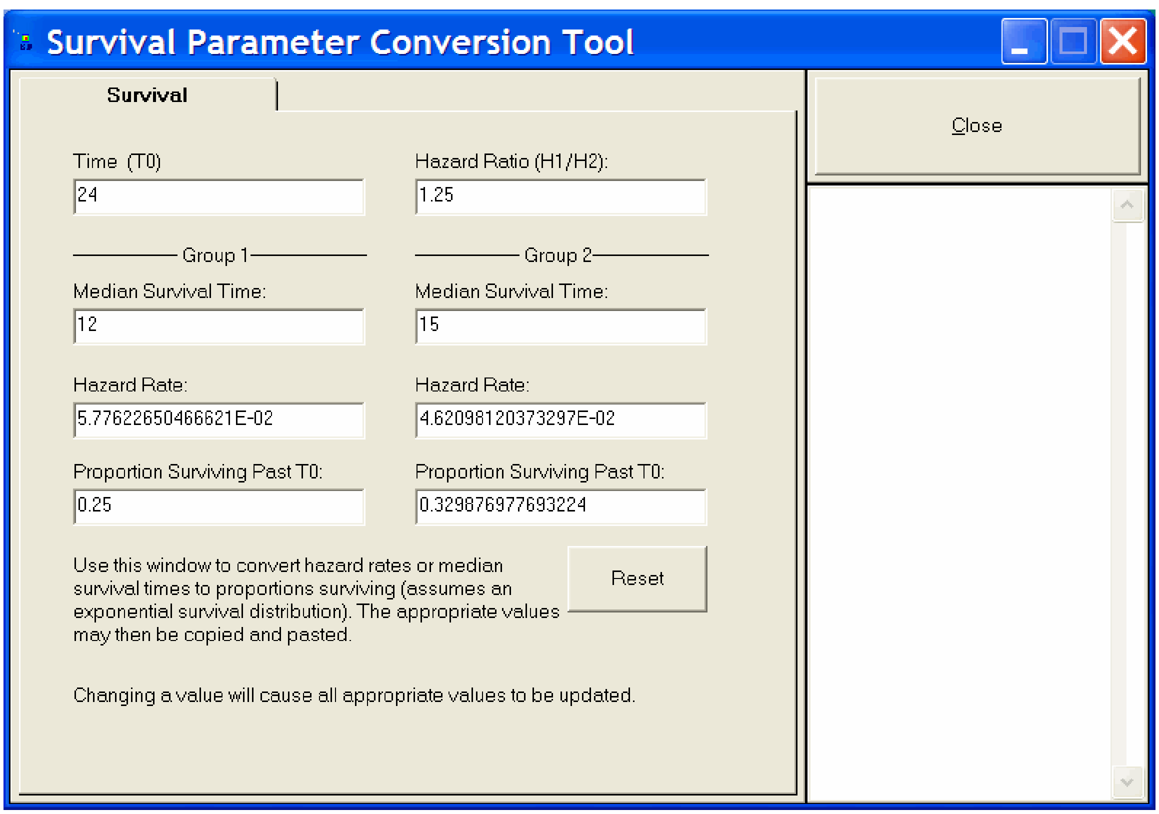
Example 4 - Non-Inferiority Test

This example will show how to use this module to calculate power and sample size for non-inferiority trials. Remember that non-inferiority is established by showing that the new treatment is no worse than the standard treatment, except for a small amount called the margin of equivalence.

Consider the following example. Suppose the median survival time of the standard drug is 15 months. Unfortunately, this drug has serious side effects. A promising new drug has been developed that has much milder side effects. A non-inferiority trial is to be designed to show that the new drug is not inferior to the standard drug. The margin of equivalence is set at 3 months, so, to establish non-inferiority, the study has to conclude that the median survival time of the new drug is at least 12 months.

The trial will accept subjects for 18 months. It will continue for an additional 6 months after the accrual period. Although the study planners anticipate some dropout, they want to begin their analysis without considering dropout. They set the significance level at 0.05. They want to determine the necessary sample size to achieve 80% and 90% power. The calculations proceed as follows.

Use the *Survival Parameter Conversion Window* to convert the 'median survival rates' to 'proportions surviving' so that they can be entered into the panel. Press the **Sv** button that is above the S1 box to display this window. Enter **24** for T0, **12** for Median Survival Time 1, and **15** for Median Survival Time 2. The screen will appear as follows.



Enter the converted values into the panel as described below.

Option	**Value**
Data Tab	
Find	**N**
S1	**0.25**
S2	**0.329876977693224**
R	**18**
% Time Until 50% Accrual	**50**
Follow-Up Time	**6**
Follow-Up Loss (1 and 2)	**0.0**
Alternative Hypothesis	**Ha: S1 < S2**
T0	**24**
N	*Ignored since this is the Find Setting*
Proportion in Group 1	**0.5**
Alpha	**0.05**
Beta	**0.10 0.20**

Annotated Output

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

Numeric Results

Power	N	Surviving Group 1	Surviving Group 2	Accrual Time	Follow Up Time	Prop in Group 1	Alpha	Beta
0.90018	1326	0.25000	0.32988	18.00	6.00	0.50000	0.05000	0.09982
0.80030	957	0.25000	0.32988	18.00	6.00	0.50000	0.05000	0.19970

Thus, a sample size of about 663 per group would be needed for 80% power, and a sample of 479 per group would be needed for 90% power.

Chapter 710

Group Sequential Tests of Two Survival Curves

Introduction

Clinical trials are longitudinal. They accumulate data sequentially through time. The participants cannot be enrolled and randomized on the same day. Instead, they are enrolled as they enter the study. It may take several years to enroll enough patients to meet sample size requirements. Because clinical trials are long term studies, it is in the interest of both the participants and the researchers to monitor the accumulating information for early convincing evidence of either harm or benefit. This permits early termination of the trial.

Group sequential methods allow statistical tests to be performed on accumulating data while a phase III clinical trial is ongoing. Statistical theory and practical experience with these designs have shown that making four or five *interim analyses* is almost as effective in detecting large differences between treatment groups as performing a new analysis after each new data value. Besides saving time and resources, such a strategy can reduce the experimental subject's exposure to an inferior treatment and make superior treatments available sooner.

When repeated significance testing occurs on the same data, adjustments have to be made to the hypothesis testing procedure to maintain overall significance and power levels. The landmark paper of Lan & DeMets (1983) provided the theory behind the *alpha spending function* approach to group sequential testing. This paper built upon the earlier work of Armitage, McPherson, & Rowe (1969), Pocock (1977), and O'Brien & Fleming (1979). ***PASS*** implements the methods given in Reboussin, DeMets, Kim, & Lan (1992) to calculate the power and sample sizes of various group sequential designs.

This module calculates sample size and power for group sequential designs used to compare two survival curves. Other modules perform similar analyses for the comparison of means and proportions. The program allows you to vary number and times of interim tests, type of alpha spending function, and test boundaries. It also gives you complete flexibility in solving for power, significance level, sample size, or effect size. The results are displayed in both numeric reports and informative graphics.

Technical Details

In many clinical trials, patients are recruited and randomized to receive a particular treatment, either experimental or control, and then monitored until either a critical event occurs or the study is ended. The length of time the patient is monitored until the critical event occurs is called the *follow-up time*. After the study is ended, the follow-up times of the patients in the two groups are compared using the *logrank test* in what is often called a *survival analysis*.

When the critical event does not occur for a patient by the time the study is ended, the follow-up time is said to have been *censored*. Although the actual event time is not known for this patient, it is known that the event time will be greater than the follow-up time. Hence, some information is gleaned from these participants. Because of this censoring, the usual tests of means or proportions cannot be used. The logrank test was developed to provide a statistical test comparing the efficacy of the two treatments.

The Hazard Ratio (HR)

Suppose the critical event is death. The survival distribution of each treatment can be characterized by the *instantaneous death rates*, λ_1 and λ_2 . An instantaneous death rate, often called the *hazard*, is the probability of death in a short interval of time. The comparison of the efficacies of the two treatments is often formalized by considering the ratio of the hazard, or *hazard ratio* (HR).

$$HR = \frac{\lambda_2}{\lambda_1}$$

Although the logrank test concerns the ratio of the hazard rates of the two groups, when planning a study, it may be easier to obtain information about the expected proportion surviving during the trial. It turns out that the hazard ratio can be computed from the survival proportions, *S1* and *S2*, using the equation

$$HR = \frac{\log(S2)}{\log(S1)}$$

when the hazard ratio is constant through time.

Assuming that group one is the control group, it may be easiest during the planning stages of a study to find *S1* and state the minimum value of *HR* that would make the experimental treatment useful. The last equation can then be manipulated to calculate a value for *S2* as follows

$$S2 = \exp\left\{HR\left(\log\left(S1\right)\right)\right\}.$$

Sometimes it is more convenient to state hazard ratio in terms of the median survival times. In this case, the hazard ratio is estimated using

$$HR = \frac{M_1}{M_2}$$

when the hazard ratio is constant for different times.

The Logrank Statistic

The following results are excerpted from Reboussin (1992). The logrank statistic is given by the equation

$$L(d) = \sum_{i=1}^{d} \left(\frac{x_i r_{ic}}{r_{ic} + r_{it}} \right)$$

where *d* is the number of events, x_i is 1 if the event at time t_i is in the control group and 0 if it is in the treatment group, r_{ic} is the number of patients in the control group at risk just be before t_i, and r_{it} is the corresponding number of patients at risk in the treatment group.

If $r_{ic} \approx r_{it}$ and *HR* is close to 1, then the sequential logrank statistic is (approximately)

$$z_k = 2\frac{L(d_k)}{\sqrt{d_k}}.$$

The subscript *k* indicates that the computations use all data that are available at the time of the k^{th} interim analysis or k^{th} *look* (*k* goes from 1 to K).

Spending Functions

Lan and DeMets (1983) introduced alpha spending functions, $\alpha(\tau)$, that determine a set of boundaries $b_1, b_2, \cdots, b_K$ for the sequence of test statistics $z_1, z_2, \cdots, z_K$. These boundaries are the critical values of the sequential hypothesis tests. That is, after each interim test, the trial is continued as long as $|z_k| < b_k$. When $|z_k| \geq b_k$, the hypothesis of equal means is rejected and the trial is stopped early.

The time argument τ either represents the proportion of elapsed time to the maximum duration of the trial or the proportion of the sample that has been collected. When elapsed time is used, it is referred to as *calendar time*. When time is measured in terms of the sample, it is referred to as *information time*. Since it is a proportion, τ can only vary between zero and one.

Alpha spending functions have the characteristics:

$$\alpha(0) = 0$$

$$\alpha(1) = \alpha$$

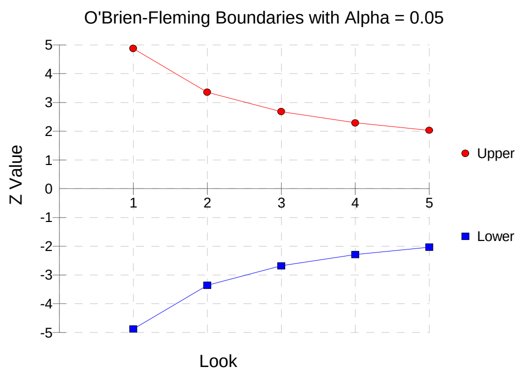
The last characteristic guarantees a fixed α level when the trial is complete. That is,

$$\Pr\left(|z_1| \geq b_1 \text{ } or \text{ } |z_2| \geq b_2 \text{ } or \cdots or \text{ } |z_k| \geq b_k\right) = \alpha(\tau)$$

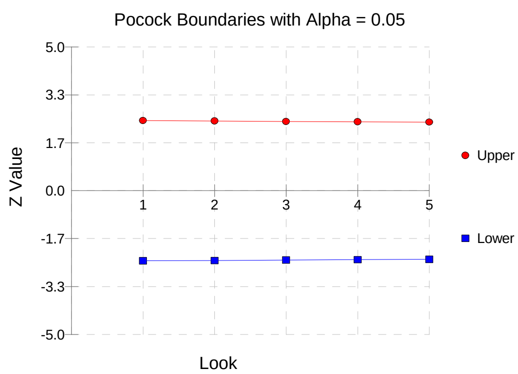
This methodology is very flexible since neither the times nor the number of analyses must be specified in advance. Only the functional form of $\alpha(\tau)$ must be specified.

PASS provides five popular spending functions plus the ability to enter and analyze your own boundaries. These are calculated as follows:

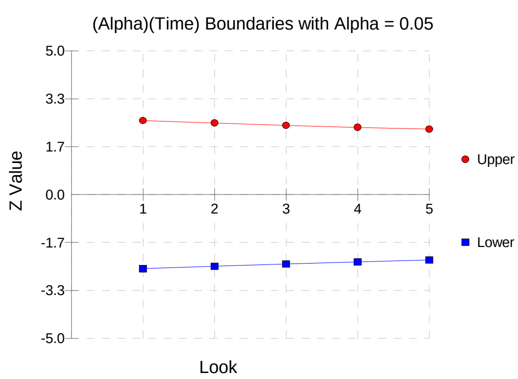
1. O'Brien-Fleming $2-2\Phi\left(\frac{Z_{\alpha/2}}{\sqrt{\tau}}\right)$



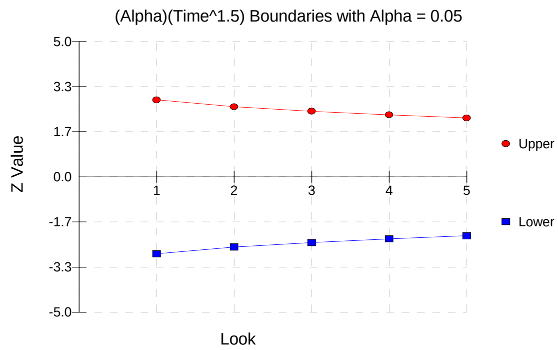
2. Pocock $\alpha\ln\left(1+(e-1)\tau\right)$



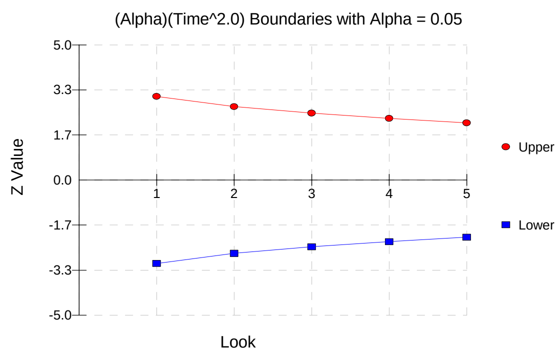
3. Alpha * time $\alpha\tau$



4. Alpha * time^1.5 $\alpha\tau^{3/2}$



5. Alpha * time^2 $\alpha\tau^{2}$



6. User Supplied

A custom set of boundaries may be entered.

The O'Brien-Fleming boundaries are commonly used because they do not significantly increase the overall sample size and because they are conservative early in the trial. Conservative in the sense that the means must be extremely different before statistical significance is indicated. The Pocock boundaries are nearly equal for all times. The Alpha*t boundaries use equal amounts of alpha when the looks are equally spaced. You can enter your own set of boundaries using the User Supplied option.

Sequential Theory

A detailed account of the methodology is contained in Lan & DeMets (1983), DeMets & Lan (1984), Lan & Zucker (1993), and DeMets & Lan (1994). A brief summary of the theoretical basis of the method will be presented here.

Group sequential procedures for interim analysis are based on their equivalence to discrete boundary crossing of a Brownian motion process with drift parameter θ. The test statistics z_k follow the multivariate normal distribution with means $\theta\sqrt{\tau_k}$ and, for $j \le k$, covariances $\sqrt{\tau_k / \tau_j}$. The drift is related to the parameters of the z-test through one of the equations

$$\theta = \frac{\log(HR)\sqrt{d_k}}{2} \text{ (exponential survival) or } \theta = \frac{|1-HR|\sqrt{d_k}}{(1+HR)} \text{ (proportional hazards).}$$

These equations may be solved for d_k, the required number of events, giving

$$d_k = \frac{4\theta^2}{[\log(HR)]^2} \text{ (exponential survival) or } d_k = \left(\frac{(1+HR)\theta}{1-HR}\right)^2 \text{ (proportional hazards).}$$

In survival analysis, the size of a sample is measured in terms of number of events rather than number of patients because it is probable that many of the patients will be censored—their event times are not known. In order to ensure that the sample size produces the required number of events, it must be inflated by the event rates.

The expected number of events can be computed from the proportion surviving using the equation

$$d_k = \frac{N(1-S_1) + N(1-S_2)}{2}$$

where *N* is the total sample size (assumed to be split evenly between groups). This can be solved for *N* to give the sample size as

$$N = \frac{2d_k}{2 - S_1 - S_2}.$$

Hence, the algorithm is as follows:

1. Compute boundary values based on a specified spending function and alpha value.
2. Calculate the drift parameter based on those boundary values and a specified power value.
3. Use the drift parameter and the above equation to calculate the appropriate event size per group d_k.
4. Use the event size to compute the appropriate sample size, *N*.

Procedure Tabs

This section describes the options that are unique to this procedure. These are located on the panels associated with the Data, Options, and Reports tabs. To find out more about using the other tabs such as Plot Text, Axes, and Template, turn to the chapter entitled Procedure Templates.

Data Tab

The Data tab contains the parameters associated with the z-test such as the survival rates, sample sizes, alpha, and beta.

Find

This option specifies the parameter to be solved for from the other parameters. The parameters that may be selected are *S1*, *S2*, *Alpha*, *Beta*, or *N*. Under most situations, you will select either Beta or *N*.

Select *N* when you want to calculate the sample size needed to achieve a given power and alpha level.

Select *Beta* when you want to calculate the power of an experiment.

S1 (Proportion Surviving 1)

S1 is the proportion of patients belonging to group 1 (controls) that are expected to survive during the study. Since *S1* is a proportion, it must be between zero and one. In many studies, this is the proportion surviving in the regular population with the standard treatment.

You may enter a range of values such as *0.1, 0.2, 0.3* or *0.1 to 0.9 by 0.2*.

S2 (Proportion Surviving 2)

S2 is the proportion of patients in the experimental group that survive during the study. This is not necessarily the expected proportion. Rather, you may set it to that proportion that, if achieved, would be of special interest. Values below (or above) this amount would not be of interest. For example, if the standard one-year survival proportion is 0.2 and the new treatment raises this proportion to 0.3 (a 50% increase in the proportion surviving), others may be interested in adopting this new treatment.

Sometimes, researchers wish to state the alternative hypothesis in terms of the hazard ratio, *HR*, rather than the value of *S2*. Using the fact that

$$HR = \frac{\log(S2)}{\log(S1)},$$

An appropriate value for *S2* may be calculated from *S1* and *HR* using the equation

$$S2 = \exp\left\{HR\left(\log(S1)\right)\right\}.$$

Sometimes it is more convenient to state hazard ratio in terms of the median survival times. In this case, the hazard ratio is estimated using

$$HR = \frac{M_1}{M_2}$$

and the above equation is used to find *S2*.

Since *S2* is a proportion, it must remain between zero and one. You may enter a range of values such as *0.1, 0.2, 0.3* or *0.1 to 0.9 by 0.2*.

N (Total Sample Size)

Enter a value (or range of values) for the total sample size. Each group is assumed to have a sample size of N/2. You may enter a range of values such as *100 to 1000 by 100*.

Note that the sample size is based implicitly on the length of the study since we used the equation

$$N = \frac{2d_k}{2 - S_1 - S_2}$$

to estimate the necessary sample size.

Alternative Hypothesis

Specify whether the test is one-sided or two-sided. When a two-sided hypothesis is selected, the value of alpha is halved. Everything else remains the same.

Note that the accepted procedure is to use Two-Sided unless you can justify using a one-sided test.

Alpha

This option specifies one or more values for the probability of a type-I error, alpha. This is also called the *significance level* or *test size*. A type-I error occurs when you reject the null hypothesis of equal proportions when in fact the proportions are equal.

Values of alpha must be between zero and one. Often, the value of 0.05 is used for alpha since this value is spread across several interim tests. This means that about one trial in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error that you are willing to take.

Beta (1-Power)

This option specifies one or more values for the probability of a type-II error (beta). A type-II error occurs when you fail to reject the null hypothesis of equal proportions when in fact they are different.

Values must be between zero and one. Historically, the value of 0.20 was often used for beta. Now, 0.10 is more common. You should pick a value that represents the risk of a type-II error you are willing to take.

Power is defined as one minus beta. Power is equal to the probability of rejecting a false null hypothesis. Hence, specifying the beta error level also specifies the power level. For example, if you specify beta values of 0.05, 0.10, and 0.20, you are specifying the corresponding power values of 0.95, 0.90, and 0.80.

Sequential Tab

The Sequential tab contains the parameters associated with Group Sequential Design such as the type of spending function, the times, and so on.

Number of Looks

This is the number of interim analyses (including the final analysis). For example, a five here means that four interim analyses will be run in addition to the final analysis.

Spending Function

Specify which alpha spending function to use. The most popular is the O'Brien-Fleming boundary that makes early tests very conservative. Select *User Specified* if you want to enter your own set of boundaries.

Boundary Truncation

You can truncate the boundary values at a specified value. For example, you might decide that no boundaries should be larger than 4.0. If you want to implement a boundary limit, enter the value here.

If you do not want a boundary limit, enter *None* here.

Times

Enter a list of time values here at which the interim analyses will occur. These values are scaled according to the value of the Max Time option.

For example, suppose a 48-month trial calls for interim analyses at 12, 24, 36, and 48 months. You could set Max Time to 48 and enter *12,24,36,48* here or you could set Max Time to *1.0* and enter *0.25,0.50,0.75,1.00* here.

The number of times entered here must match the value of the Number of Looks.

Equally Spaced

If you are planning to conduct the interim analyses at equally spaced points in time, you can enter *Equally Spaced* and the program will generate the appropriate time values for you.

Max Time

This is the total running time of the trial. It is used to convert the values in the Times box to fractions. The units (months or years) do not matter, as long as they are consistent with those entered in the Times box.

For example, suppose Max Time = 3 and Times = 1, 2, 3. Interim analyses would be assumed to have occurred at 0.33, 0.67, and 1.00.

Informations

You can weight the interim analyses on the amount of information obtained at each time point rather than on actual calendar time. If you would like to do this, enter the information amounts here. Usually, these values are the sample sizes obtained up to the time of the analysis.

For example, you might enter *50, 76, 103, 150* to indicate that 50 individuals where included in the first interim analysis, 76 in the second, and so on.

Upper and Lower Boundaries

If the Spending Function is set to *User Supplied* you can enter a set of lower test boundaries, one for each interim analysis. The lower boundaries should be negative and the upper boundaries should be positive. Typical entries are *4,3,3,3,2* and *4,3,2,2,2*.

Symmetric

If you only want to enter the upper boundaries and have them copied with a change in sign to the lower boundaries, enter *Symmetric* for the lower boundaries.

Options Tab

The Options tab controls the convergence of the various iterative algorithms used in the calculations.

Max Iterations 1

Specify the maximum number of iterations to be run before the search for the criterion of interest (Alpha, Beta, etc.) is aborted. When the maximum number of iterations is reached without convergence, the criterion is left blank.

Recommended: 500 (or more).

Max Iterations 2

This is the maximum number of iterations used in the Lan-DeMets algorithm during its search routine. We recommend a value of at least 200.

Probability Tolerance

During the calculation of the probabilities associated with a set of boundary values, probabilities less than this are assumed to be zero.

We suggest a value of 0.00000000001.

Power Tolerance

This is the convergence level for the search for the spending function values that achieve a certain power. Once the iteration changes are less than this amount, convergence is assumed. We suggest a value of 0.0000001.

If the search is too time consuming, you might try increasing this value.

Alpha Tolerance

This is the convergence level for the search for a given alpha value. Once the changes in the computed alpha value are less than this amount, convergence is assumed and iterations stop. We suggest a value of 0.0001.

This option is only used when you are searching for alpha.

If the search is too time consuming, you may try increasing this value.

Bnd Axes Tab

The Bnd Axes tab, short for Boundary Axes tab, allows the axes of the spending function plots to be set separately from those of the power plots. The options are identical to those of the Axes tab.

Example1 - Finding the Sample Size

A clinical trial is to be conducted over a two-year period to compare the hazard rate of a new treatment to that of the current treatment. The proportion surviving for two years using the current treatment is 0.3. The health community will be interested in the new treatment if the proportion surviving is increased to 0.45, a 50% increase. So that the sample size requirements for several survival proportions can be compared, it is also of interest to compute the sample size at response rates of 0.30, 0.35, 0.40, and 0.50. Assume the survival times are exponential.

Testing will be done at the 0.05 significance level and the power should be set to 0.10. A total of four tests are going to be performed on the data as they are obtained. The O'Brien-Fleming boundaries will be used.

Find the necessary sample sizes and test boundaries assuming equal sample sizes for each arm and two-sided hypothesis tests.

Setup

You can enter these values yourself or load the Example1 template from the Template tab.

Option **Value**

Data Tab

Find .... **N**
S1.. **0.3**
S2.. **0.35, 0.40, 0.45, 0.50**
N.. *Ignored*
Alternative Hypothesis .... **Two-Sided**
Alpha.. **0.05**
Beta.. **0.10**

Sequential Tab

Number of Looks............................... **4**
Spending Function .... **O'Brien-Fleming**
Times .... **Equally Spaced**
Max Time **2**

Annotated Output

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

Numeric Results

Numeric Results for Two-Sided LogRank Test (Assuming Exponential Survival)

Power	Total Sample Size (N)	Total Required Events	Alpha	Beta	Proportion Surv. (S1)	Proportion Surv. (S2)	Hazard Ratio
0.900003	3378	2280	0.050000	0.099997	0.3000	0.3500	0.8720
0.900267	884	575	0.050000	0.099733	0.3000	0.4000	0.7611
0.900626	407	254	0.050000	0.099374	0.3000	0.4500	0.6632
0.900018	234	140	0.050000	0.099982	0.3000	0.5000	0.5757

Report Definitions
Power is the probability of rejecting a false null hypothesis. Power should be close to one.
N is the number of items sampled from each group.
Events is the number of events that must occur in each group.
Alpha is the probability of rejecting a true null hypothesis in at least one of the sequential tests.
Beta is the probability of accepting a false null hypothesis at the conclusion of all tests.
S1 is the proportion surviving in group 1.
S2 is the proportion surviving in group 2.
HR is the hazard ratio. It is calculated using Log(S2)/Log(S1).

Summary Statements
A total sample size of 3378 (split equally between the two groups), or 2280 events, achieves 90% power to detect a hazard rate of 0.8720 when the proportions surviving in each group are 0.3000 and 0.3500 at a significance level (alpha) of 0.050000 using a two-sided log rank test. These results assume that 4 sequential tests are made using the O'Brien-Fleming spending function to determine the test boundaries and that the survival times are exponential.

This report shows the values of each of the parameters, one scenario per row. Note that 254 events are required when $S2 = 0.45$. Based on the expected survival proportions, this many events will occur if the overall sample size is 407.

Total Sample Size (N)

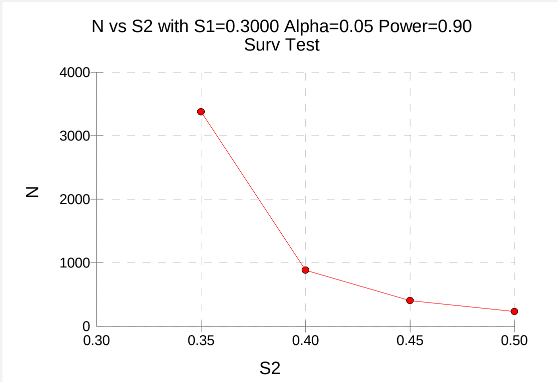
This is the estimated sample size that is needed to obtain the necessary number of events.

Total Required Events

This is the number of events (deaths, etc.) that are required to achieve the desired power levels.

Plots Section

The values from the above table are displayed in the chart below. Note that this plot actually occurs further down in the report.



This plot shows that an increase in sample size from under 1000 to well over 3000 is necessary when the detectable proportion surviving is reduced from 0.4 to 0.35.

Details Section

Details when Spending = O'Brien-Fleming, N = 407, d = 254, S1 = 0.3000, S2 = 0.4500

Look	Time	Lower Bndry	Upper Bndry	Nominal Alpha	Inc Alpha	Total Alpha	Inc Power	Total Power
1	0.5000	-4.33263	4.33263	0.000015	0.000015	0.000015	0.003516	0.003516
2	1.0000	-2.96311	2.96311	0.003045	0.003036	0.003051	0.255183	0.258699
3	1.5000	-2.35902	2.35902	0.018323	0.016248	0.019299	0.427665	0.686364
4	2.0000	-2.01406	2.01406	0.044003	0.030701	0.050000	0.214262	0.900626
Drift	3.27466							

This report shows information about the individual interim tests. One report is generated for each scenario.

Look

These are the sequence numbers of the interim tests.

Time

These are the time points at which the interim tests are conducted. Since the Max Time was set to 2 (for two years), these time values are in years. Hence, the first interim test is at half a year, the second at one year, and so on.

We could have set Max Time to 24 so that the time scale was in months.

Lower and Upper Boundary

These are the test boundaries. If the computed value of the test statistic *z* is between these values, the trial should continue. Otherwise, the trial can be stopped.

Nominal Alpha

This is the value of alpha for these boundaries if they were used for a single, standalone, test. Hence, this is the significance level that must be found for this look in a standard statistical package that does not adjust for multiple looks.

Inc Alpha

This is the amount of alpha that is *spent* by this interim test. It is close to, but not equal to, the value of alpha that would be achieved if only a single test was conducted. For example, if we lookup the third value, 2.35902, in normal probability tables, we find that this corresponds to a (two-sided) alpha of 0.018323. However, the entry is 0.016248. The difference is due to the correction that must be made for multiple tests.

Total Alpha

This is the total amount of alpha that is used up to and including the current test.

Inc Power

These are the amounts that are added to the total power at each interim test. They are often called the exit probabilities because they give the probability that significance is found and the trial is stopped, given the alternative hypothesis.

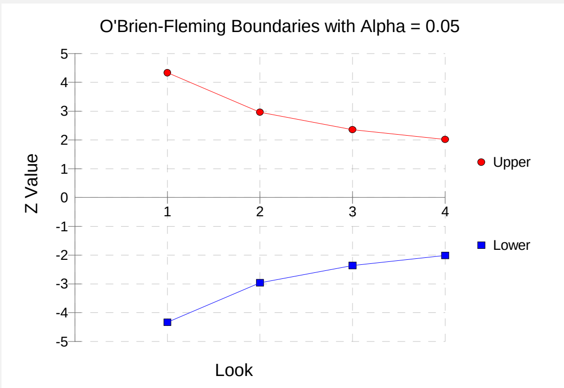
Total Power

These are the cumulative power values. They are also the cumulative exit probabilities. That is, they are the probability that the trial is stopped at or before the corresponding time.

Drift

This is the value of the Brownian motion drift parameter.

Boundary Plots



This plot shows the interim boundaries for each look. This plot shows very dramatically that the results must be extremely significant at early looks, but that they are near the single test boundary (1.96 and -1.96) at the last look.

Example2 - Finding the Power

Continuing the scenario began in Example1, the researcher wishes to calculate the power of the design at sample sizes 50, 250, 450, 650, and 850. Testing will be done at the 0.01, 0.05, 0.10 significance levels and the overall power will be set to 0.10. Find the power of these sample sizes and test boundaries assuming equal sample sizes per arm and two-sided hypothesis tests.

Proceeding as in Example1, we decide to translate the mean and standard deviation into a percent of mean scale.

Setup

You can enter these values yourself or load the Example2 template from the Template tab.

Option **Value**

Data Tab

Find ....**Beta and Power**
S1..**0.3**
S2..**0.45**
N..**50 to 850 by 200**
Alternative Hypothesis .... **Two-Sided**
Alpha...**0.01, 0.05, 0.10**
Beta...*Ignored*

Sequential Tab

Number of Looks...............................**4**
Spending Function ....**O’Brien-Fleming**
Times ....**Equally Spaced**
Max Time **2**

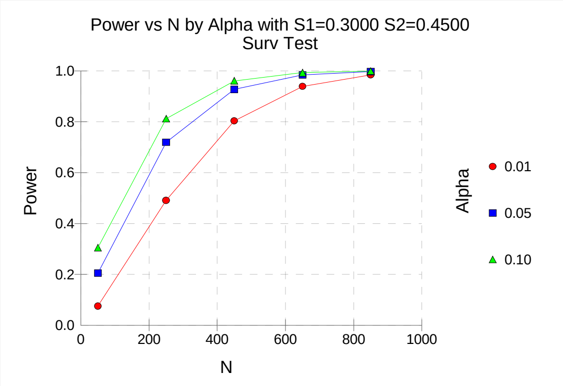
Annotated Output

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

Numeric Results

Numeric Results for Two-Sided LogRank Test (Assuming Exponential Survival)

Power	Total Sample Size (N)	Total Required Events	Alpha	Beta	Proportion Surv. (S1)	Proportion Surv. (S2)	Hazard Ratio
0.075632	50	31	0.010000	0.924368	0.3000	0.4500	0.6632
0.491011	250	156	0.010000	0.508989	0.3000	0.4500	0.6632
0.802921	450	281	0.010000	0.197079	0.3000	0.4500	0.6632
0.938923	650	406	0.010000	0.061077	0.3000	0.4500	0.6632
0.983780	850	531	0.010000	0.016220	0.3000	0.4500	0.6632
0.205032	50	31	0.050000	0.794968	0.3000	0.4500	0.6632
0.719225	250	156	0.050000	0.280775	0.3000	0.4500	0.6632
0.926894	450	281	0.050000	0.073106	0.3000	0.4500	0.6632
0.984044	650	406	0.050000	0.015956	0.3000	0.4500	0.6632
0.996907	850	531	0.050000	0.003093	0.3000	0.4500	0.6632
0.305105	50	31	0.100000	0.694895	0.3000	0.4500	0.6632
0.812103	250	156	0.100000	0.187897	0.3000	0.4500	0.6632
0.960490	450	281	0.100000	0.039510	0.3000	0.4500	0.6632
0.992807	650	406	0.100000	0.007193	0.3000	0.4500	0.6632
0.998800	850	531	0.100000	0.001200	0.3000	0.4500	0.6632



These data show the power for various sample sizes and alphas. It is interesting to note that once the sample size is greater than about 450, the value of alpha has comparatively little difference on the value of power.

Example3 - Effect of Number of Looks

Continuing with examples one and two, it is interesting to determine the impact of the number of looks on power. ***PASS*** allows only one value for the Number of Looks parameter per run, so it will be necessary to run several analyses. To conduct this study, set alpha to 0.05, *N* to 407, and leave the other parameters as before. Run the analysis with Number of Looks equal to 1, 2, 3, 4, 6, 8, 10, and 20. Record the power for each run.

Setup

You can enter these values yourself or load the Example3 template from the Template tab.

<u>**Option**</u>	<u>**Value**</u>
Data Tab	
Find	**Beta and Power**
S1	**0.3**
S2	**0.45**
N	**407**
Alternative Hypothesis	**Two-Sided**
Alpha	**0.05**
Beta	*Ignored*
Sequential Tab	
Number of Looks	**1 (Also run with 2, 3, 4, 6, 8, 10, and 20)**
Spending Function	**O'Brien-Fleming**
Times	**Equally Spaced**
Max Time	**2**

Numeric Results

Numeric Results for Two-Sided Logrank Test (Assuming Exponential Survival)

Power	N	Events	Alpha	Beta	S1	S2	Looks
0.905693	407	254	0.050000	0.094307	0.3000	0.4500	1
0.904719	407	254	0.050000	0.095281	0.3000	0.4500	2
0.902412	407	254	0.050000	0.097588	0.3000	0.4500	3
0.900626	407	254	0.050000	0.099374	0.3000	0.4500	4
0.898243	407	254	0.050000	0.101757	0.3000	0.4500	6
0.896763	407	254	0.050000	0.103237	0.3000	0.4500	8
0.895758	407	254	0.050001	0.104242	0.3000	0.4500	10
0.893398	407	254	0.050001	0.106602	0.3000	0.4500	20

This analysis shows how little the number of looks impacts the power of the design. The power of a study with no interim looks is 0.905693. When twenty interim looks are made, the power falls to 0.893398—a very small change.

Example4 - Studying a Boundary Set

Continuing with the previous examples, suppose that you are presented with a set of boundaries and want to find the quality of the design (as measured by alpha and power). This is easy to do with ***PASS***. Suppose that the analysis is to be run with five interim looks at equally spaced time points. The upper boundaries to be studied are 3.5, 3.5, 3.0, 2.5, 2.0. The lower boundaries are symmetric. The analysis would be run as follows.

Setup

You can enter these values yourself or load the Example4 template from the Template tab.

Option	Value
Data Tab	
Find	**Beta and Power**
S1	**0.30**
S2	**0.45**
N	**407**
Alternative Hypothesis	**Two-Sided**
Alpha	**0.05 (will be calculated from boundaries)**
Beta	*Ignored*
Sequential Tab	
Number of Looks	**5**
Spending Function	**User Supplied**
Times	**Equally Spaced**
Lower Boundaries	**Symmetric**
Upper Boundaries	**3.5, 3.5, 3.0, 2.5, 2.0**
Max Time	**2**

Numeric Results

Numeric Results for Two-Sided Logrank Test (Assuming Exponential Survival)

Power	Total Sample Size (N)	Total Required Events	Alpha	Beta	Proportion Surv. (S1)	Proportion Surv. (S2)	Hazard Ratio
0.900746	407	254	0.048157	0.099254	0.3000	0.4500	0.6632

Details when Spending = User Supplied, N = 407, d=254, S1 = 0.3000, S2 = 0.4500

Look	Time	Lower Bndry	Upper Bndry	Nominal Alpha	Inc Alpha	Total Alpha	Inc Power	Total Power
1	0.4000	-3.50000	3.50000	0.000465	0.000465	0.000465	0.020899	0.020899
2	0.8000	-3.50000	3.50000	0.000465	0.000408	0.000874	0.063132	0.084031
3	1.2000	-3.00000	3.00000	0.002700	0.002410	0.003284	0.243522	0.327553
4	1.6000	-2.50000	2.50000	0.012419	0.010331	0.013615	0.343072	0.670625
5	2.0000	-2.00000	2.00000	0.045500	0.034542	0.048157	0.230122	0.900746
Drift	3.27466							

The power for this design is about 0.90. This value depends on both the boundaries and the sample size. The alpha level is about 0.048. This value only depends on the boundaries.

Example5 - Validation using O'Brien-Fleming Boundaries

Reboussin (1992) presents an example for binomial distributed data for a design with two-sided O'Brien-Fleming boundaries, looks = 3, alpha = 0.05, beta = 0.10, *S1* = 0.30, *S2* = 0.786 (which gives a hazard ratio of 0.20). They compute a drift of 3.261 and the number of events at 16.42. The upper boundaries are: 4.8769, 3.3569, 2.6803, 2.2898, 2.0310.

To test that ***PASS*** provides the same result, enter the following.

Setup

You can enter these values yourself or load the Example5 template from the Template tab.

Option **Value**

Data Tab

Find ....**N**
S1...**0.30**
S2...**0.786**
N...*Ignored*
Alternative Hypothesis .... **Two-Sided**
Alpha..**0.05**
Beta..**0.10**

Sequential Tab

Number of Looks...............................**5**
Spending Function ....**O'Brien-Fleming**
Times ....**Equally Spaced**
Max Time **1**

Numeric Results

Numeric Results for Two-Sided Logrank Test (Assuming Exponential Survival)

Power	Total Sample Size (N)	Total Required Events	Alpha	Beta	Proportion Surv. (S1)	Proportion Surv. (S2)	Hazard Ratio
0.905173	37	17	0.050000	0.099434	0.3000	0.7860	0.2000

Details when Spending = O'Brien-Fleming, N = 37, d = 17, S1 = 0.3000, S2 = 0.7860

Look	Time	Lower Bndry	Upper Bndry	Nominal Alpha	Inc Alpha	Total Alpha	Inc Power	Total Power
1	0.2000	-4.87688	4.87688	0.000001	0.000001	0.000001	0.000341	0.000341
2	0.4000	-3.35695	3.35695	0.000788	0.000787	0.000788	0.102760	0.103101
3	0.6000	-2.68026	2.68026	0.007357	0.006828	0.007616	0.352450	0.455551
4	0.8000	-2.28979	2.28979	0.022034	0.016807	0.024424	0.298953	0.754504
5	1.0000	-2.03100	2.03100	0.042255	0.025576	0.050000	0.150669	0.905173
Drift	3.30902							

The number of events, rounded to 17, matches the 16.42 reported in Reboussin (1992).

Chapter 800

Correlation Coefficient

Introduction

The correlation coefficient, ρ (rho), is a popular statistic for describing the strength of the relationship between two variables. The correlation coefficient is the slope of the regression line between two variables when both variables have been standardized by subtracting their means and dividing by their standard deviations. The correlation ranges between plus and minus one.

When ρ is used as a descriptive statistic, no special distributional assumptions need to be made about the variables (Y and X) from which it is calculated. When hypothesis tests are made, you assume that the observations are independent and that the variables are distributed according to the bivariate-normal density function. However, as with the t-test, tests based on the correlation coefficient are robust to moderate departures from this normality assumption.

The population correlation ρ is estimated by the sample correlation coefficient *r*. Note we use the symbol *R* on the screens and printouts to represent the population correlation.

Difference between Linear Regression and Correlation

The correlation coefficient is used when both X and Y are from the normal distribution (in fact, the assumption actually is that X and Y follow a bivariate normal distribution). The point is, X is assumed to be a random variable whose distribution is normal. In the linear regression context, no statement is made about the distribution of X. In fact, X is not even a random variable. Instead, it is a set of fixed values such as 10, 20, 30 or -1, 0, 1. Because of this difference in definition, we have included both Linear Regression and Correlation algorithms. This module deals with the Correlation (random X) case.

Test Procedure

The testing procedure is as follows. H_0 is the null hypothesis that the true correlation is a specific value, ρ_0 (usually, $\rho_0 = 0$). H_A represents the alternative hypothesis that the actual correlation of the population is ρ_1, which is not equal to ρ_0. Choose a value R_α, based on the distribution of the sample correlation coefficient, so that the probability of rejecting H_0 when H_0 is true is equal to a specified value, α . Select a sample of *n* items from the population and compute the sample correlation coefficient, r_S. If $r_S > R_\alpha$ reject the null hypothesis that $\rho = \rho_0$ in favor of an alternative hypothesis that $\rho = \rho_1$, where $\rho_1 > \rho_0$. The power is the probability of rejecting H_0 when the true correlation is ρ_1.

All calculations are based on the algorithm described by Guenther (1977) for calculating the cumulative correlation coefficient distribution.

Calculating the Power

Let $R(r|N,\rho)$ represent the area under a correlation density curve to the left of *r*. *N* is the sample size and ρ is the population correlation. The power of the significance test of $\rho_1 > \rho_0$ is calculated as follows:

1. Find r_α such that $1-R(r_\alpha|N,\rho_0)=\alpha$.

2. Compute the power = $1-R(r_\alpha|N,\rho_1)$.

Notice that the calculations follow the same pattern as for the t-test. First find the rejection region by finding the critical value (r_α) under the null hypothesis. Next, calculate the probability that a sample of size *N* drawn from the population defined by setting the correlation to ρ_1 is in this rejection region. This is the power.

Procedure Options

This section describes the options that are unique to this procedure. These are located on the panels associated with the Data, Options, and Reports tabs. To find out more about using the other tabs such as Plot Text, Axes, and Template, turn to the chapter entitled Procedure Templates.

Data Tab

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

Find

This option specifies the parameter to be calculated from the values of the other parameters. Under most conditions, you would either select *Beta* or *N*.

Select *N* when you want to determine the sample size needed to achieve a given power and alpha error level.

Select *Beta* when you want to calculate the power.

R0 (Baseline Correlation)

Specify the value of ρ_0. Note that the range of the correlation is between plus and minus one. This value is usually set to zero.

R1 (Alternative Correlation)

Specify the value of ρ_1, the population correlation under the alternative hypothesis. Note that the range of the correlation is between plus and minus one. The difference between *R0* and *R1* is being tested by this significance test.

You can enter a range of values separated by blanks or commas.

N (Sample Size)

The number of observations in the sample. Each observation is made up of two values: one for *X* and one for *Y*.

Alternative Hypothesis

This option specifies the alternative hypothesis. This implicitly specifies the direction of the hypothesis test. The null hypothesis is $H_0: \rho_0 = \rho_1$.

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

Possible selections are:

Ha: R0 <> R1. This is the most common selection. It yields the *two-tailed* test. Use this option when you are testing whether the correlation values are different, but you do not want to specify beforehand which correlation is larger.

Ha: R0 < R1. This option yields a *one-tailed* test.

Ha: R0 > R1. This option also yields a *one-tailed* test.

Alpha (Significance Level)

This option specifies one or more values for the probability of a type-I error (alpha). A type-I error occurs when you reject the null hypothesis of equal correlations when in fact they are equal.

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

Beta (1 - Power)

This option specifies one or more values for the probability of a type-II error (beta). A type-II error occurs when you fail to reject the null hypothesis of equal correlations when in fact they are different.

Values must be between zero and one. Historically, the value of 0.20 was often used for beta. However, you should pick a value for beta that represents the risk of a type-II error you are willing to take.

Power is defined as one minus beta. Power is equal to the probability of rejecting a false null hypothesis. Hence, specifying the beta error level also specifies the power level. For example, if you specify beta values of 0.05, 0.10, and 0.20, you are specifying the corresponding power values of 0.95, 0.90, and 0.80, respectively.

Example 1 - Finding the Power

Suppose a study will be run to test whether the correlation between forced vital capacity (X) and forced expiratory value (Y) in a particular population is 0.30. Find the power when alpha is 0.01, 0.05, and 0.10 and the N = 20, 60, 100.

Setup

This section presents the values of each of the parameters needed to run this example. You can make these changes directly on your screen or you can load the template entitled Example1 by clicking the Template tab and loading this template.

Option	Value
Data Tab	
Find	**Beta and Power**
R0	**0.0**
R1	**0.3**
N	**20 60 100**
Alternative Hypothesis	**Ha: R0 <> R1**
Alpha	**0.01 0.05 0.10**
Beta	*Ignored since this is the Find setting*

Annotated Output

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

Numeric Results

Numeric Results for Ha: R0<>R1

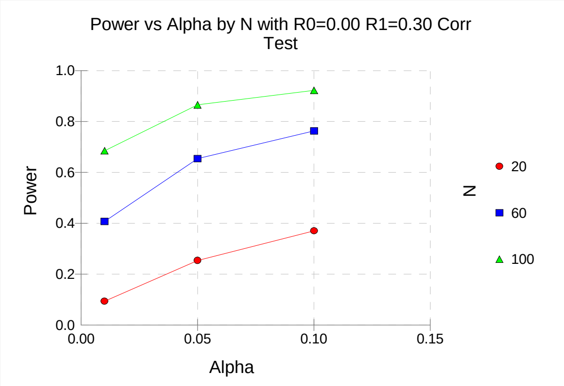
Power	N	Alpha	Beta	R0	R1
0.09401	20	0.01000	0.90599	0.00000	0.30000
0.40755	60	0.01000	0.59245	0.00000	0.30000
0.68475	100	0.01000	0.31525	0.00000	0.30000
0.25394	20	0.05000	0.74606	0.00000	0.30000
0.65396	60	0.05000	0.34604	0.00000	0.30000
0.86524	100	0.05000	0.13476	0.00000	0.30000
0.37052	20	0.10000	0.62948	0.00000	0.30000
0.76282	60	0.10000	0.23718	0.00000	0.30000
0.92230	100	0.10000	0.07770	0.00000	0.30000

Report Definitions
Power is the probability of rejecting a false null hypothesis. It should be close to one.
N is the size of the sample drawn from the population. To conserve resources, it should be small.
Alpha is the probability of rejecting a true null hypothesis. It should be small.
Beta is the probability of accepting a false null hypothesis. It should be small.
R0 is the value of the population correlation under the null hypothesis.
R1 is the value of the population correlation under the alternative hypothesis.

Summary Statements
A sample size of 20 achieves 9% power to detect a difference of -0.30000 between the null hypothesis correlation of 0.00000 and the alternative hypothesis correlation of 0.30000 using a two-sided hypothesis test with a significance level of 0.01000.

This report shows the values of each of the parameters, one scenario per row. The values from this table are plotted in the chart below.

Plots Section



This plot shows the relationship between alpha and power in this example.

Example 2 - Finding the Sample Size

Continuing with the last example, find the sample size necessary to achieve a power of 90% with a 0.05 significance level.

Setup

This section presents the values of each of the parameters needed to run this example. You can make these changes directly on your screen or you can load the template entitled Example2 by clicking the Template tab and loading this template.

Option	Value
Find	**N**
R0	**0.0**
R1	**0.3**
N	*Ignored since this is the Find setting*
Alternative Hypothesis	**Ha: R0 <> R1**
Alpha	**0.05**
Beta	**0.20**

Annotated Output

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

Numeric Results

Numeric Results for Ha: R0<>R1

Power	N	Alpha	Beta	R0	Ra
0.90081	112	0.05000	0.09919	0.00000	0.30000

The required sample size is 112. You would now experiment with the parameters to find out how much varying each will influence the sample size.

Example 3 - Validation using Zar

Zar (1984) page 312 presents an example in which the power of a correlation coefficient is calculated. If *N* = 12, *alpha* = 0.05, *R0* = 0, and *R1* = 0.866, Zar calculates a power of 98% for a two-sided test.

Setup

This section presents the values of each of the parameters needed to run this example. You can make these changes directly on your screen or you can load the template entitled Example3 by clicking the Template tab and loading this template.

Option	Value
Find	**Beta and Power**
R0	**0.0**
R1	**0.866**
N	**12**
Alternative Hypothesis	**Ha: R0 <> R1**
Alpha	**0.05**
Beta	*Ignored since this is the Find setting*

Annotated Output

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

Numeric Results

Numeric Results for Ha: R0<>R1

Power	N	Alpha	Beta	R0	R1
0.98398	12	0.05000	0.01602	0.00000	0.86600

The power of 0.98 matches Zar’s results.

Example 4 - Validation using Graybill

Graybill (1961) pages 211-212 presents an example in which the power of a correlation coefficient is calculated when the baseline correlation is different from zero. Let *N* = 24, *alpha* = 0.05, and *R0* = 0.5. Graybill calculates the power of a two-sided test when *R1* = 0.2 and 0.3 to be 0.363 and 0.193.

Setup

This section presents the values of each of the parameters needed to run this example. You can make these changes directly on your screen or you can load the template entitled Example4 by clicking the Template tab and loading this template.

Option	Value
Find	**Beta and Power**
R0	**0.5**
R1	**0.2 0.3**
N	**24**
Alternative Hypothesis	**Ha: R0 <> R1**
Alpha	**0.05**
Beta	*Ignored since this is the Find setting*

Annotated Output

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

Numeric Results

Numeric Results for Ha: R0<>R1

Power	N	Alpha	Beta	R0	R1
0.36583	24	0.05000	0.63417	0.50000	0.20000
0.19950	24	0.05000	0.80050	0.50000	0.30000

The power values match Graybill’s results to two decimal places.

Chapter 805

Two Correlations

Introduction

The correlation coefficient (or correlation), ρ, is a popular parameter for describing the strength of the association between two variables. The correlation coefficient is the slope of the regression line between two variables when both variables have been standardized. It ranges between plus and minus one. This chapter covers the case in which you want to test the difference between two correlations, each coming from a separate sample.

Since the correlation is the standardized slope between two variables, you could also apply this procedure to the case in which you want to test whether the slopes in two groups are equal.

Test Procedure

In the following discussion, ρ is the population correlation coefficient and *r* is the value calculated from a sample. The testing procedure is as follows. H_0 is the null hypothesis that $\rho_1 = \rho_2$. H_A represents the alternative hypothesis that $\rho_1 \neq \rho_2$ (one-tailed hypotheses are also available). To construct the hypothesis test, transform the correlations using the Fisher-*z* transformation.

$$z_i = \frac{1}{2}\log\left(\frac{1+r_i}{1-r_i}\right)$$

$$Z_i = \frac{1}{2}\log\left(\frac{1+\rho_i}{1-\rho_i}\right)$$

This transformation is used because the combined distribution of r_1 and r_2 is too difficult to work with, but the distributions of z_1 and z_2 are approximately normal.

Note that the reverse transformation is

$$r_i = \frac{e^{z_i} - e^{-z_i}}{e^{z_i} + e^{-z_i}}$$

Once the correlations have been converted into *z* values, the normal distribution may be used to conduct the test of $Z_1 - Z_2$. The standard deviation of the difference is given by

$$\sigma_{z_1 - z_2} = \sqrt{\frac{1}{N_1 - 3} + \frac{1}{N_2 - 3}}$$

The test statistic is given by

$$z = \frac{(z_1 - z_2) - (Z_1 - Z_2)}{\sigma_{z_1 - z_2}}$$

Note that the lower case *z's* represent the values calculated from the two samples and the upper case *Z's* represent the hypothesized population values.

Calculating the Power

1. Find z_α such that $1 - \Phi(z_\alpha) = \alpha$, where $\Phi(x)$ is the area under the standardized normal curve to the left of *x*.
2. Calculate: $Z_1 = \frac{1}{2}\log\left(\frac{1+\rho_1}{1-\rho_1}\right)$
3. Calculate: $Z_2 = \frac{1}{2}\log\left(\frac{1+\rho_2}{1-\rho_2}\right)$
4. Calculate: $\sigma_{z_1 - z_2} = \sqrt{\frac{1}{N_1 - 3} + \frac{1}{N_2 - 3}}$
5. Calculate: $x_a = Z_1 - Z_2 + z_\alpha \sigma_{z_1 - z_2}$
6. Calculate: $z_a = \frac{x_a}{\sigma_{z_1 - z_2}}$
7. Calculate: $\text{Power} = 1 - \Phi(z_a)$

Procedure Options

This section describes the options that are unique to this procedure. These are located on the panels associated with the Data, Options, and Reports tabs. To find out more about using the other tabs such as Plot Text, Axes, and Template, turn to the chapter entitled Procedure Templates.

Data Tab

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

Find

This option specifies the parameter to be solved for from the other parameters. The parameters that may be selected are *R1*, *R2*, *N1*, *N2*, *Alpha*, and *Beta*. Under most situations, you will select either *Beta* or *N1*.

Select *N1* when you want to calculate the sample size needed to achieve a given power and alpha level.

Select *Beta* when you want to calculate the power of an experiment.

R1 (Correlation Group 1)

Specify the value of the population correlation coefficient of group one. Possible values range between plus and minus one.

You can enter a single value or a range of values separated by commas or blanks.

Note that the power depends on the specific values of *R1* and *R2*, not just their difference. Hence, *R1* = 0 and *R2* = 0.3 will have a different power from *R1* = 0.3 and *R2* = 0.6.

R2 (Correlation Group 2)

Specify the value of the population correlation coefficient from group two under the alternative hypothesis. Possible values range between plus and minus one.

You can enter a single value or a range of values separated by commas or blanks.

Alpha (Significance Level)

This option specifies one or more values for the probability of a type-I error (alpha). A type-I error occurs when you reject the null hypothesis of equal correlations when in fact they are equal.

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

Beta (1 - Power)

This option specifies one or more values for the probability of a type-II error (beta). A type-II error occurs when you fail to reject the null hypothesis of equal correlations when in fact they are different.

Values must be between zero and one. Historically, the value of 0.20 was often used for beta. However, you should pick a value for beta that represents the risk of a type-II error you are willing to take.

Power is defined as one minus beta. Power is equal to the probability of rejecting a false null hypothesis. Hence, specifying the beta error level also specifies the power level. For example, if you specify beta values of 0.05, 0.10, and 0.20, you are specifying the corresponding power values of 0.95, 0.90, and 0.80.

Alternative Hypothesis

This option specifies the alternative hypothesis. This implicitly specifies the direction of the hypothesis test. The null hypothesis is always $H_0: \rho_1 = \rho_2$.

Possible selections are:

Ha: R1 <> R2. This is the most common selection. It yields the *two-tailed* test. Use this option when you are testing whether the correlation values are different, but you do not want to specify beforehand which value is larger.

Ha: R1 < R2. This option yields a *one-tailed* test. When you use this option, you should be careful to enter values for *R1* and *R2* that follow this relationship.

Ha: R1 > R2. This option yields a *one-tailed* test. When you use this option, you should be careful to enter values for *R1* and *R2* that follow this relationship.

N1 (Sample Size Group 1)

Enter a value (or range of values) for the sample size of this group. Note that these values are ignored when you are solving for *N1*. You may enter a range of values such as *10 to 100 by 10*.

N2 (Sample Size Group 2)

Enter a value (or range of values) for the sample size of group 2 or enter *Use R* to base *N2* on the value of *N1*. You may enter a range of values such as *10 to 100 by 10.*

Use R

When *Use R* is entered here, *N2* is calculated using the formula

$$N2 = [R(N1)]$$

where *R* is the Sample Allocation Ratio and [*Y*] is the first integer greater than or equal to *Y*. For example, if you want *N1 = N2*, select *Use R* and set *R* = 1.

R (Sample Allocation Ratio)

Enter a value (or range of values) for *R*, the allocation ratio between samples. This value is only used when *N2* is set to *Use R*.

When used, *N2* is calculated from *N1* using the formula: *N2*=[*R*(*N1*)] where [*Y*] is the next integer greater than or equal to *Y*. Note that setting *R* = 1.0 forces *N2 = N1*.

Example 1 - Finding the Power

A researcher wants to compare the relationship between weight and heart rate in males and females. If the correlation between weight and heart rate is 0.3 in a sample of 100 males and 0.5 in a sample of 100 females, what is the power of a two sided test for the difference between correlations at the 0.01 and 0.05 significance levels? Also compute the power for samples of 20, 200, 300, 400, and 600.

Setup

This section presents the values of each of the parameters needed to run this example. You can make these changes directly on your screen or you can load the template entitled Example1 by clicking the Template tab and loading this template.

Option	**Value**
Data Tab	
Find	**Beta and Power**
R1	**0.3**
R2	**0.5**
Alpha	**0.01 0.05**
Beta	*Ignored since this is the Find setting.*
Alternative Hypothesis	**Ha: R1 <> R2**
N1	**20 100 200 300 400 600**
N2	**Use R**
R	**1.0**

Annotated Output

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

Numeric Results

Numeric Results when Ha: R1<>R2

Power	N1	N2	Allocation Ratio	R1	R2	Difference (R1-R2)	Alpha	Beta
0.03081	20	20	1.000	0.30000	0.50000	-0.20000	0.01000	0.96919
0.18250	100	100	1.000	0.30000	0.50000	-0.20000	0.01000	0.81750
0.42230	200	200	1.000	0.30000	0.50000	-0.20000	0.01000	0.57770
0.63541	300	300	1.000	0.30000	0.50000	-0.20000	0.01000	0.36459
0.78888	400	400	1.000	0.30000	0.50000	-0.20000	0.01000	0.21112
0.94144	600	600	1.000	0.30000	0.50000	-0.20000	0.01000	0.05856
0.10760	20	20	1.000	0.30000	0.50000	-0.20000	0.05000	0.89240
0.38603	100	100	1.000	0.30000	0.50000	-0.20000	0.05000	0.61397
0.66271	200	200	1.000	0.30000	0.50000	-0.20000	0.05000	0.33729
0.83200	300	300	1.000	0.30000	0.50000	-0.20000	0.05000	0.16800
0.92196	400	400	1.000	0.30000	0.50000	-0.20000	0.05000	0.07804
0.98548	600	600	1.000	0.30000	0.50000	-0.20000	0.05000	0.01452

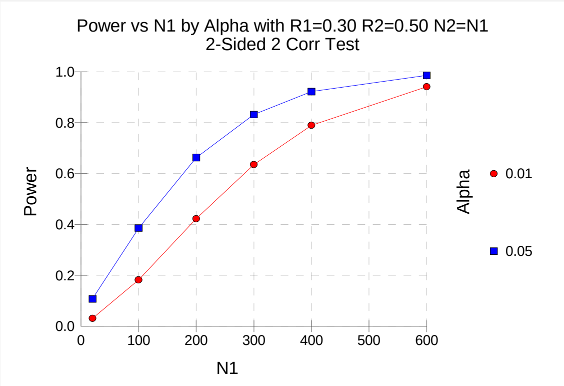
Report Definitions
Power is the probability of rejecting a false null hypothesis. It should be close to one.
N1 and N2 are the sizes of the samples drawn from the two populations. To conserve resources, it should be small.
Allocation Ratio is N1/N2 so that N2 = N1 x R.
Alpha is the probability of rejecting a true null hypothesis. It should be small.
Beta is the probability of accepting a false null hypothesis. It should be small.
R1 is the value of both correlations under the null hypothesis.
R2 is the correlation in group two under the alternative hypothesis.

Summary Statements
Group sample sizes of 20 and 20 achieve 3% power to detect a difference of 0.20000 between the null hypothesis that both group correlations are 0.30000 and the alternative hypothesis that the correlation in group 2 is 0.50000 using a two-sided z test (which uses Fisher's z-transformation) with a significance level of 0.01000.

This report shows the values of each of the parameters, one scenario per row. The definitions of each column are given in the Report Definitions section of the report, so they will not be repeated here.

The values from this table are plotted in the chart below.

Plots Section



This plot shows the relationship between alpha, power, and sample size in this example.

Example 2 - Finding the Sample Size

Continuing with the previous example, suppose the researchers want to determine the exact sample size necessary to achieve 90% power at a 0.05 significance level.

Setup

This section presents the values of each of the parameters needed to run this example. You can make these changes directly on your screen or you can load the template entitled Example2 by clicking the Template tab and loading this template.

Option	Value
Data Tab	
Find	**N1**
R1	**0.3**
R2	**0.5**
Alpha	**0.05**
Beta	**0.10**
Alternative Hypothesis	**Ha: R1 <> R2**
N1	*Ignored since this is the Find setting*
N2	**Use R**
R	**1.0**

Annotated Output

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

Numeric Results

Numeric Results when Ha: R1<>R2

Power	N1	N2	Allocation Ratio	R1	R2	Difference (R1-R2)	Alpha	Beta
0.90040	369	369	1.000	0.30000	0.50000	-0.20000	0.05000	0.09960

PASS has calculated the sample size as 369 per group.

Example 3 - Validation using Zar

Zar (1984) page 314 presents an example of calculating the power for a test of two correlations. In his example, when *N1* = 95, *N2* = 98, *R1* = 0.84, *R2* = 0.78, and *alpha* = 0.05, the power is 22% for a two-sided test.

Setup

This section presents the values of each of the parameters needed to run this example. You can make these changes directly on your screen or you can load the template entitled Example3 by clicking the Template tab and loading this template.

Option	Value
Data Tab	
Find	**Beta and Power**
R1	**0.84**
R2	**0.78**
Alpha	**0.05**
Beta	*Ignored since this is the Find setting*
Alternative Hypothesis	**Ha: R1 <> R2**
N1	**95**
N2	**98**
R	**1.0**

Annotated Output

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

Numeric Results

Numeric Results when Ha: R1<>R2

Power	N1	N2	Allocation Ratio	R1	R2	Difference (R1-R2)	Alpha	Beta
0.22498	95	98	1.032	0.84000	0.78000	0.06000	0.05000	0.77502

PASS has also calculated the power to be 22%.

Chapter 810

Intraclass Correlation

Introduction

The intraclass correlation coefficient is often used as an index of reliability in a measurement study. In these studies, there are *N* observations made on each of *K* individuals. These individuals represent a factor observed at random. This design arises when *K* subjects are each rated by *N* raters.

The intraclass correlation coefficient may be thought of as the correlation between any two observations made on the same subject. When this correlation is high, the observations on a subject tend to match, and the measurement reliability is 'high.'

Technical Details

Our formulation comes from Walter, Eliasziw, and Donner (1998) and Winer (1991). Denote response *j* of subject *i* by Y_{ij}, where $i = 1, 2, \ldots, K$ and $j = 1, 2, \ldots, N$. The model for this situation is

$$Y_{ij} = \mu + a_i + e_{ij}$$

where the random subject effects a_i are normally distributed with mean 0 and variance σ_a^2 and the measurement errors, e_{ij} are normally distributed with mean 0 and variance σ_e^2. We assume that the subject effects and the measurement errors are independent. The intraclass correlation is then defined as

$$\rho = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2}$$

The hypothesis test is stated formally as

$$H_0\colon \rho = \rho_0$$
$$H_1\colon \rho = \rho_1 > \rho_0$$

This hypothesis is tested from the data of a one-way analysis of variance table using the value: $\frac{MS_a}{MS_e}$. The critical value for the test statistic is

$$C\left(F_{1-\alpha/2,df1,df2}\right)$$

where

$$C = 1 + \left[\frac{N\rho_0}{1-\rho_0}\right]$$

$$df1 = K - 1$$

$$df2 = K\left(N-1\right)$$

The power of this test procedure is given by

$$Power = 1 - P(F \geq C_0 F_{1-\alpha/2,df1,df2})$$

where

$$C_0 = \frac{1+N\rho_0/\left(1-\rho_0\right)}{1+N\rho_1/\left(1-\rho_1\right)}$$

Procedure Options

This section describes the options that are unique to this procedure. These are located on the panels associated with the Data, Options, and Reports tabs. To find out more about using the other tabs such as Plot Text, Axes, and Template, turn to the chapter entitled Procedure Templates.

Data Tab

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

Find

This option specifies the parameter to be solved for from the other parameters. The parameters that may be selected are *R0*, *R1*, *K*, *N*, Alpha, and Beta (or Power).

Under most situations, you will select either Beta to calculate power or *N* to calculate sample size.

Note that the value selected here always appears as the vertical axis on the charts.

The program is set up to evaluate beta directly. For the other parameters, a search is made using an iterative procedure until an appropriate value is found.

Alpha (Significance Level)

This option specifies one or more values for the probability of a type-I error (alpha). A type-I error occurs when you reject the null hypothesis when in fact it is true.

Values between 0.001 and 0.100 are acceptable. The value of 0.05 has become the standard. This means that about one test in twenty will falsely reject the null hypothesis. Although 0.05 is the standard value, you should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

Note that you can enter a range of values such as *0.01,0.05* or *0.01 to 0.05 by 0.01*.

Beta (1-Power)

This option specifies one or more values for beta (the probability of accepting a false null hypothesis). Since statistical power is equal to one minus beta, specifying beta implicitly specifies the power. For example, setting beta at 0.20 also sets the power to 0.80.

Values must be between zero and one. The value of 0.20 has often used for beta. However, you should pick a value for beta that represents the risk of this type of error you are willing to take.

Note that you can enter a range of values such as *0.10,0.20* or *0.05 to 0.20 by 0.05*.

If your only interest is in determining the appropriate sample size for a confidence interval, set beta to 0.5.

R0 (Intraclass Correlation 0)

This is the value(s) of the intraclass correlation coefficient when the null hypothesis is true. You may enter a single value or a list of values. The range of *R0* is between zero and *R1*.

The intraclass correlation is calculated as $V(A)/[V(A)+V(E)]$ where $V(E)$ is the variation within a subject and $V(A)$ is the variation between subjects. It is a measure of the extent to which the observations within a subject are similar (or dependent) relative to observations from other subjects.

R1 (Intraclass Correlation 1)

This is the value(s) of the intraclass correlation coefficient when the alternative hypothesis is true. You may enter a single value or a list of values. The range of *R1* is between *R0* and one.

K (Number of Subjects)

Enter a value (or range of values) for the number of subjects, *K*, that are measured.

You may enter a range of values such as *50,150,250* or *50 to 300 by 50*.

N (Observations Per Subject)

Enter a value (or range of values) for the number of observations, *N*, per subjects. In a reliability study, this is the number of raters (assuming each subject is rated by all raters).

You may enter a range of values such as *2,3,4* or *2 to 12 by 2*.

Options Tab

This tab sets a couple of options used in the iterative procedures.

Maximum Iterations

Specify the maximum number of iterations allowed before the search for the criterion of interest is aborted. When the maximum number of iterations is reached without convergence, the criterion is left blank. A value of 500 is recommended.

Example1 - Calculating Power

Suppose that a study is to be conducted in which *R0* = 0.2; *R1* = 0.3; *K* = 50 to 250 by 100; Alpha = 0.05; and *N* = 2 to 5 by 1 and beta is to be calculated.

Setup

You can enter these values yourself or load the Example1 template from the Template tab.

Option	Value
Data Tab	
Find	**Beta and Power**
Alpha	**0.05**
Beta	**Calculated**
R0	**0.2**
R2	**0.3**
K	**50 to 250 by 100**
N	**2 to 5 by 1**

Annotated Output

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

Numeric Results

Numeric Results

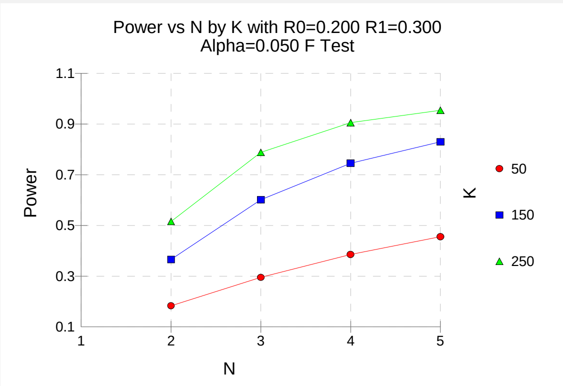
Power	Number of Subjects	Observations Per Subject	Intraclass Correlation 0	Intraclass Correlation 1	Alpha	Beta
0.18333	50	2	0.200	0.300	0.05000	0.81667
0.29534	50	3	0.200	0.300	0.05000	0.70466
0.38528	50	4	0.200	0.300	0.05000	0.61472
0.45522	50	5	0.200	0.300	0.05000	0.54478
0.36558	150	2	0.200	0.300	0.05000	0.63442
0.60094	150	3	0.200	0.300	0.05000	0.39906
0.74538	150	4	0.200	0.300	0.05000	0.25462
0.83005	150	5	0.200	0.300	0.05000	0.16995
0.51549	250	2	0.200	0.300	0.05000	0.48451
0.78790	250	3	0.200	0.300	0.05000	0.21210
0.90522	250	4	0.200	0.300	0.05000	0.09478
0.95403	250	5	0.200	0.300	0.05000	0.04597

Report Definitions
Power is the probability of rejecting a false null hypothesis. It should be close to one.
K is the number of subjects.
N is the number of observations per subject in the sample.
R0 is intraclass correlation assuming the null hypothesis.
R1 is intraclass correlation assuming the alternative hypothesis.
Alpha is the probability of rejecting a true null hypothesis. It should be small.
Beta is the probability of accepting a false null hypothesis. It should be small.

Summary Statements
A sample size of 50 subjects with 2 observations per subject achieves 18% power to detect an intraclass correlation of 0.300 under the alternative hypothesis when the intraclass correlation under the null hypothesis is 0.200 using an F-test with a significance level of 0.05000.

This report shows the power for each of the scenarios.

Plot Section



This plot shows the relation between power, number of subjects (*K*), and observations per subject (*N*).

Example2 - Validation using Walter

Walter *et al.* (1998) page 106 give a table of sample sizes. When R0 is 0.2, R1 is 0.3, K is 544, N is 2, and Alpha is 0.05, Beta is 0.80.

Setup

You can enter these values yourself or load the Example2 template from the Template tab.

Option	Value
Data Tab	
Find	**Beta and Power**
Alpha	**0.05**
Beta	**Calculated**
R0	**0.2**
R2	**0.3**
K	**50 to 250 by 100**
N	**2 to 5 by 1**

Annotated Output

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

Numeric Results

Numeric Results

Power	Number of Subjects	Observations Per Subject	Intraclass Correlation 0	Intraclass Correlation 1	Alpha	Beta
0.80033	544	2	0.200	0.300	0.05000	0.19967

PASS has also calculated the power as 0.80.

Chapter 815

Coefficient Alpha: One Set

Introduction

Coefficient alpha, or *Cronbach's alpha*, is a measure of the reliability of a scale consisting of k parts. The k parts usually often represent k items on a questionnaire or k raters. This module calculates power and sample size for testing whether coefficient alpha, ρ, is different from a given value such as zero.

Technical Details

Feldt et al. (1987) has shown that if $\hat{\rho}$ is the estimated value of coefficient alpha computed from a sample of size N questionnaires with k items, the statistic W is distributed as an F ratio with degrees of freedom N-1 and (k-1)(N-1), where

$$W = \frac{1-\rho_0}{1-\hat{\rho}}$$

and ρ_0 is the value of ρ assumed by the null hypothesis, H0.

Calculating the Power

Using the above definition of W, the power of the significance test of $\rho > \rho_0$ is calculated as follows:

1. Find F_α such that $\text{Prob}\left(F_{1-\alpha,N-1,(k-1)(N-1)}\right) = 1-\alpha$
2. Compute $\rho_C = \dfrac{F_\alpha + \rho_0 - 1}{F_\alpha}$
3. Compute $W_1 = \dfrac{1-\rho_1}{1-\rho_c}$, where ρ_1 is the value of ρ at which the power is calculated.
4. Compute the power = $1 - \Pr\left(W_1 > F_{N-1,(k-1)(N-1)}\right)$

Procedure Options

This section describes the options that are unique to this procedure. These are located on the panels associated with the Data, Options, and Reports tabs. To find out more about using the other tabs such as Plot Text, Axes, and Template, turn to the chapter entitled Procedure Templates.

Data Tab

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

Find

This option specifies the parameter to be calculated from the values of the other parameters. Under most conditions, you would either select *Beta* or *N*.

Select *N* when you want to determine the sample size needed to achieve a given power and alpha error level.

Select *Beta* when you want to calculate the power of an experiment.

CA0 (Coefficient Alpha|H0)

Specify the value of ρ_0, the value of coefficient alpha under the null hypothesis. Usually, this value will be zero, but any value between -1 and 1 is valid as long as it is not equal to CA1.

You may enter a list of values separated by blanks such as *0 0.1 0.2*.

CA1 (Coefficient Alpha|H1)

Specify the value of ρ_1, the value of coefficient alpha at which the power is computed. Usually, this value is positive, but any value between -1 and 1 is valid as long as it is not equal to CA0.

You may enter a list of values separated by blanks such as *0.1 0.2 0.3*.

N (Sample Size)

Specify the number of observations in the sample. You may enter a range such as *10 to 100 by 10* or a list of values separated by commas or blanks such as *20 50 100*.

K (Number of Items or Raters)

K is the number of items or raters in the study. Since it is a count, it must be an integer greater than one. You may enter a list of values separated by blanks.

Alternative Hypothesis

This option specifies whether the alternative hypothesis is one-sided or two-sided. It also specifies the direction of the hypothesis test. The null hypothesis is $H_0: \rho_0 = \rho$.The alternative hypothesis enters into power calculations by specifying the rejection region of the hypothesis test. Its accuracy is critical.

Possible selections are:

H1: CA0 <> CA1. This is the most common selection. It yields the *two-tailed* test. Use this option when you are testing whether values are different, but you do not want to specify beforehand which is larger.

H1: CA0 < CA1. This option yields a *one-tailed* test.

H1: CA0 > CA1. This option also yields a *one-tailed* test.

Alpha (Significance Level)

This option specifies one or more values for the probability of a type-I error (alpha). A type-I error occurs when you reject the null hypothesis when it is true.

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

Beta (1 - Power)

This option specifies one or more values for the probability of a type-II error (beta). A type-II error occurs when you fail to reject a false null hypothesis.

Values must be between zero and one. Historically, the value of 0.20 was often used for beta. However, you should pick a value for beta that represents the risk of a type-II error you are willing to take.

Power is defined as one minus beta. Power is the probability of rejecting a false null hypothesis. Hence, specifying the beta error level also specifies the power level. For example, if you specify beta values of 0.05, 0.10, and 0.20, you are specifying the corresponding power values of 0.95, 0.90, and 0.80, respectively.

Example1 - Finding the Power

Suppose a study is being designed to test whether the coefficient alpha is 0.6 against the two-sided alternative. Find the power when K = 20, alpha = 0.05, CA1 = 0.65 0.70 0.75, and N = 50 100 200 300 500 700 1000 and 1400.

Setup

This section presents the values of each of the parameters needed to run this example. You can make these changes directly on your screen or you can load the template entitled Example1 by clicking the Template tab and loading this template.

Option	Value
Data Tab	
Find	**Beta**
CA0	**0.6**
CA1	**0.65 0.70 0.75**
N	**50 100 200 300 500 700 1000 1400**
Alternative Hypothesis	**H1: CA0 <> CA1**
Alpha	**0.05**
Beta	*Ignored since this is the Find setting*
K	**20**

Annotated Output

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

Numeric Results

Numeric Results when H1: CA0<>CA1

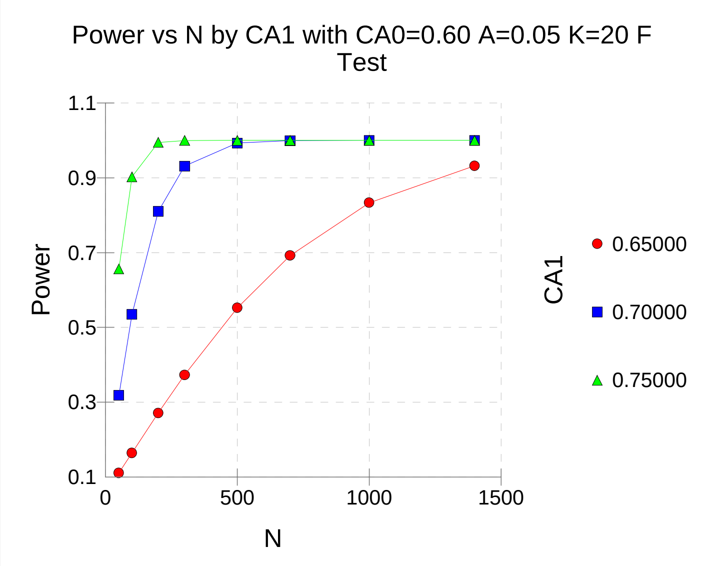
Power	Sample Size (N)	Number of Items (K)	Coefficient Alpha|H1 (CA1)	Coefficient Alpha|H0 (CA0)	Signif. Level (Alpha)	Beta
0.11084	50	20	0.65000	0.60000	0.05000	0.88916
0.16444	100	20	0.65000	0.60000	0.05000	0.83556
0.27111	200	20	0.65000	0.60000	0.05000	0.72889
0.37314	300	20	0.65000	0.60000	0.05000	0.62686
0.55224	500	20	0.65000	0.60000	0.05000	0.44776
0.69191	700	20	0.65000	0.60000	0.05000	0.30809

Report Definitions
Power is the probability of rejecting a false null hypothesis.
N is the total sample size.
K is the number of items or raters.
CA1 is the value of coefficient alpha at which the power is computed.
CA0 is the value of coefficient alpha under the null hypothesis.
Alpha is the probability of rejecting a true null hypothesis. It should be small.
Beta is the probability of accepting a false null hypothesis. It should be small.
H0 is the null hypothesis that coefficient alpha equals CA0.
H1 is the alternative hypothesis that coefficient alpha does not equal CA0.

Summary Statements
A sample of 50 subjects each responding to 20 items achieves 11% power to detect the difference between the coefficient alpha under the null hypothesis of 0.60000 and the coefficient alpha under the alternative hypothesis of 0.65000 using a two-sided F-test with a significance level of 0.05000.

This report shows the values of each of the parameters, one scenario per row. The values from this table are plotted in the chart below.

Plots Section



This plot shows the relationship between CA1, N, and power.

Example2 - Finding the Sample Size

Continuing with the last example, find the sample size necessary to achieve a power of 90% with a 0.05 significance level.

Setup

This section presents the values of each of the parameters needed to run this example. You can make these changes directly on your screen or you can load the template entitled Example2 by clicking the Template tab and loading this template.

Option	Value
Data Tab	
Find	**N**
CA0	**0.6**
CA1	**0.65 0.70 0.75**
N	*Ignored since this is the Find setting*
Alternative Hypothesis	**H1: CA0 <> CA1**
Alpha	**0.05**
Beta	**0.10**
K	**20**

Annotated Output

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

Numeric Results

Numeric Results when H1: CA0<>CA1

Power	Sample Size (N)	Number of Items (K)	Coefficient Alpha|H1 (CA1)	Coefficient Alpha|H0 (CA0)	Signif. Level (Alpha)	Beta
0.90022	1233	20	0.65000	0.60000	0.05000	0.09978
0.90073	265	20	0.70000	0.60000	0.05000	0.09927
0.90261	100	20	0.75000	0.60000	0.05000	0.09739

This report shows the dramatic increase in sample size that is needed to achieve the desired sample power as CA1 gets closer to CA0.

Example 3 - Validation using Bonett

Bonett (2002) page 337 presents a table in which the sample sizes were calculated for several parameter configurations. When CA0 = 0, CA1 = 0.50, alpha = 0.10, beta = 0.05, and *k* = 2, 5, 10, and 100, he finds *N* to be 93, 59, 52, and 48, respectively.

Setup

This section presents the values of each of the parameters needed to run this example. You can make these changes directly on your screen or you can load the template entitled Example3 by clicking the Template tab and loading this template.

Option	Value
Find	**N**
CA0	**0**
CA1	**0.5**
N	*Ignored since this is the Find setting*
Alternative Hypothesis	**H1: CA0 <> CA1**
Alpha	**0.1**
Beta	**0.05**
K	**2 5 10 100**

Annotated Output

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

Numeric Results

Numeric Results when H1: CA0<>CA1

Power	Sample Size (N)	Number of Items (K)	Coefficient Alpha|H1 (CA1)	Coefficient Alpha|H0 (CA0)	Signif. Level (Alpha)	Beta
0.95176	93	2	0.50000	0.00000	0.10000	0.04824
0.95253	59	5	0.50000	0.00000	0.10000	0.04747
0.95047	52	10	0.50000	0.00000	0.10000	0.04953
0.95213	48	100	0.50000	0.00000	0.10000	0.04787

The sample sizes match Bonett’s results exactly.

Chapter 820

Coefficient Alpha: Two Sets

Introduction

Coefficient alpha, or *Cronbach's alpha*, is a popular measure of the reliability of a scale consisting of *k* parts. The *k* parts often represent *k* items on a questionnaire (scale) or *k* raters. This module calculates power and sample size for testing whether two coefficient alphas are different when the two samples are either dependent or independent.

Technical Details

Feldt et al. (1999) presents methods for testing one-, or two-, sided hypotheses about two coefficient alphas, which we label ρ_1 and ρ_2. The results assume that N_1 observations for each of k_1 items are available for one scale and N_2 observations for each of k_2 items are available for another scale. These sets of observations may either be from two independent groups of subjects (independent case) or two sets of observations on each subject (dependent case). In the dependent case, $N_1 = N_2$ and the correlation coefficient between the overall scores of each scale is represented by ϕ. For the independent case $\phi = 0$.

Suppose $\hat{\rho}_1$ and $\hat{\rho}_2$ are the sample estimates of ρ_1 and ρ_2, respectively. Hypothesis tests are based on the result that the test statistic,

$$\begin{aligned} W &= \left(\frac{1-\hat{\rho}_2}{1-\hat{\rho}_1}\right)\left(\frac{1-\rho_1}{1-\rho_2}\right) \\ &= \hat{\delta}\left(\frac{1-\rho_1}{1-\rho_2}\right) \end{aligned},$$

is approximately distributed as a central *F* variable with degrees of freedom ν_1 and ν_2. The values of ν_1 and ν_2 depend on N_1, N_2, k_1, k_2, and ϕ.

Also define

$$c_i = (N_i - 1)(k_i - 1), \quad i = 1,2$$

Independent Case

When the two scales are independent, there are two situations that must be considered separately. If $c_i > 1000 \textit{ and } k_i > 25$, the values of ν_1 and ν_2 are computed using

$$\nu_1 = N_1 - 1$$

$$\nu_2 = N_2 - 1$$

otherwise, they are computed using

$$\nu_1 = \frac{2A^2}{2B - AB - A^2}$$

$$\nu_2 = \frac{2A}{A-1}$$

where

$$A = \frac{c_1(N_2 - 1)}{(c_1 - 2)(N_2 - 3)}$$

$$B = \frac{(N_1 + 1)(N_2 - 1)^2(c_2 + 2)c_1^2}{(N_2 - 3)(N_2 - 5)(N_1 - 1)(c_1 - 2)(c_1 - 4)c_2}$$

Dependent Case

When the two scales are dependent, it follows that $N_1 = N_2 = N$. There are two situations that must be considered separately.

If $c_i > 1000 \textit{ and } k_i > 25$, the values of ν_1 and ν_2 are computed using

$$\nu_1 = \nu_2 = \frac{N - 1 - 7\phi^2}{1 - \phi^2}$$

otherwise, they are computed using

$$\nu_1 = \frac{2M^2}{V(2 - M) - M^2(M - 1)}$$

$$\nu_2 = \frac{2M}{M-1}$$

where

$$M = A - \frac{2\phi^2}{N-1}$$

$$V = B - A^2 - \frac{4\phi^2}{N-1}$$

Calculating the Power

Let ρ_{20} be the value of coefficient alpha in the second set under H0, ρ_{21} be the value of coefficient alpha in the second set at which the power is calculated, and ρ_1 be the value of coefficient alpha in the first set. The power of the one-sided hypothesis that $H_0 : \rho_{20} \le \rho_1$ versus the alternative that $H_1 : \rho_{20} > \rho_1$ is calculated as follows:

1. Find F_α such that $\mathrm{Prob}\left(F < F_{\alpha,\nu 1,\nu 2}\right) = \alpha$
2. Compute $\delta' = \frac{1}{F_\alpha}\left(\frac{1-\rho_1}{1-\rho_{20}}\right)$
3. Compute $W_1 = \left(\frac{1-\rho_1}{1-\rho_{21}}\right)\delta'$
4. Compute the power = $1 - \Pr\left(W_1 > F_{\nu 1,\nu 2}\right)$

Procedure Options

This section describes the options that are unique to this procedure. These are located on the panels associated with the Data, Options, and Reports tabs. To find out more about using the other tabs such as Plot Text, Axes, and Template, turn to the chapter entitled Procedure Templates.

Data Tab

The Data tab contains most of the parameters and options of interest for this procedure.

Find

This option specifies the parameter to be calculated from the values of the other parameters. Under most conditions, you would either select *Beta* or *N1*.

Select *N1* when you want to determine the sample size needed to achieve a given power and alpha error level.

Select *Beta* when you want to calculate the power of an experiment.

CA1 (Coefficient Alpha Set 1)

Specify the value of ρ_1, the value of coefficient alpha, for dataset one. Often, this value will be zero, but any value between -1 and 1 is valid as long as it is not equal to CA21.

You may enter a list of values separated with blanks or a range of values.

CA20 (Coefficient Alpha Set 2|H0)

Enter the value of coefficient alpha in dataset two under the null hypothesis. The null hypothesis is H0: CA1=CA20, so often you will set this value equal to CA1. Any value between -1 and 1 (non-inclusive) is valid as long as it is not equal to CA21.

You may enter *CA1* to indicate that you want the value of CA1 copied here. You may also enter a multiple of CA1 such as *1.5CA1*.

You may enter a list of values separated with blanks or a range of values.

CA21 (Coefficient Alpha Set 2|H1)

Enter the value of coefficient alpha in dataset two at which the power is computed. Any value between -1 and 1 (non-inclusive) is valid as long as it is not equal to CA20. The values of CA20 and CA21 should match the direction set by the ‘Alternative Hypothesis’ option.

You may enter a list of values separated with blanks or a range of values.

Phi (Correlation Between Sets)

This option implicitly specifies the two datasets as being either independent (datasets on different subjects) or dependent (both datasets on the same subjects). Suppose you calculate the average score for each subject for both dataset one and dataset two. This parameter is the correlation between those two averages over all subjects. If the correlation is zero, the two datasets are assumed to be independent. That is, it is assumed that they come from different sets of subjects.

If the correlation is non-zero, the two datasets are assumed to be dependent. That is, it is assumed that both sets of items were measured on the same subjects. Typical values in this case are between 0.2 and 0.7. When the datasets are dependent, it is assumed that N1=N2.

Since this is a correlation, the theoretical range is from -1 to 1. Typical values are 0.0 for independent designs and between 0.2 and 0.7 for dependent designs.

K1 (Items/Scale in Set 1)

K1 is the number of items or raters in dataset one. Since it is a count, it must be an integer. K1 must be greater than or equal to two.

You may enter a list of values separated with blanks or a range of values.

N1 (Sample Size in Set 1)

N1 is the sample size (number of observations or subjects) in dataset one. Feldt (1999) states that it is ill-advised to use sample sizes less than 30.

You may enter a list of values separated with blanks or a range of values.

K2 (Items/Scale in Set 2)

K2 is the number of items or raters in dataset two. Since it is a count, it must be an integer greater than or equal to two.

You may enter a list of values separated with blanks or a range of values.

For ease of input, you can enter multiples of K1. For example, all of the following entries are allowed:

K1 2K1 0.5K1

N2 (Sample Size in Set 2)

N2 is the sample size (number of observations or subjects) in dataset two. Feldt (1999) states that it is ill-advised to use sample sizes less than 30. Note that if phi is non-zero, this value will be forced equal to N1 regardless of what is entered here.

For ease of input, when phi is non-zero, you can enter multiples of N1. For example, all of the following entries are allowed:

N1 2N1 0.5N1

You may enter a list of values separated with blanks or a range of values.

Alternative Hypothesis

This option specifies whether the alternative hypothesis is one-sided or two-sided. It also specifies the direction of the hypothesis test. The null hypothesis is $H_0: \rho_1 = \rho_2$. The alternative hypothesis enters into power calculations by specifying the rejection region of the hypothesis test. Its accuracy is critical.

Possible selections are:

H1: CA1 <> CA2. This is the most common selection. It yields the *two-tailed* test. Use this option when you are testing whether values are different, but you do not want to specify beforehand which is larger.

H1: CA1 < CA2. This option yields a *one-tailed* test.

H1: CA1 > CA2. This option also yields a *one-tailed* test.

Alpha (Significance Level)

This option specifies one or more values for the probability of a type-I error (alpha). A type-I error occurs when you reject the null hypothesis when in fact it is true.

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

Beta (1 - Power)

This option specifies one or more values for the probability of a type-II error (beta). A type-II error occurs when you fail to reject the null hypothesis of equality when it is false.

Values must be between zero and one. Historically, the value of 0.20 was often used for beta. However, you should pick a value for beta that represents the risk of a type-II error you are willing to take.

Power is defined as one minus beta. Power is equal to the probability of rejecting a false null hypothesis. Hence, specifying the beta error level also specifies the power level. For example, if you specify beta values of 0.05, 0.10, and 0.20, you are specifying the corresponding power values of 0.95, 0.90, and 0.80, respectively.

Example1 - Finding the Power

Suppose a study is being designed to compare the coefficient alphas of two scales. The researchers are going to use a two-sided F-test at a significance level of 0.05. Past experience has shown that CA1 is approximately 0.4. The researchers will use different subjects in each dataset. Find the power when K1 = K2 = 10, CA20 = CA1, N1 = 50, 100, 150, 200, 250, and 300, N2 = N1, and CA21 = 0.6 and 0.7.

Setup

This section presents the values of each of the parameters needed to run this example. You can make these changes directly on your screen or you can load the template entitled Example1 by clicking the Template tab and loading this template.

<u>**Option**</u>	<u>**Value**</u>
Data Tab	
Find	**Beta and Power**
CA1	**0.4**
CA20	**CA1**
CA21	**0.60 0.70**
K1	**10**
N1	**50 to 300 by 50**
K2	**K1**
N2	**N1**
Alternative Hypothesis	**H1: CA1 <> CA2**
Phi	**0**
Alpha	**0.05**
Beta	*Ignored since this is the Find setting*

Annotated Output

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

Numeric Results

Numeric Results for Comparing Two Coefficient Alphas
H1: CA1 <> CA2

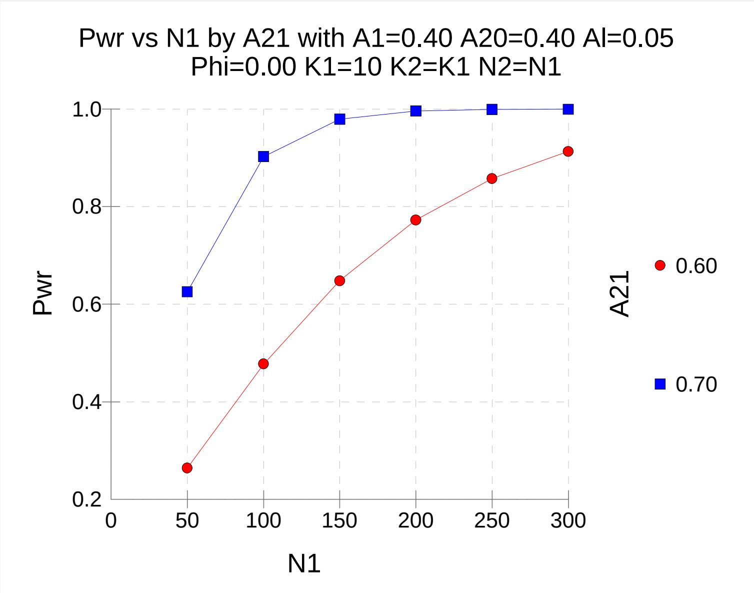
Power	Sample Sizes N1/N2	Number of Items K1/K2	Coef. Alpha Set 2|H1 (CA21)	Coef. Alpha Set 2|H0 (CA20)	Coef. Alpha Set 1 (CA1)	Signif. Level (Alpha)	Corr. Between Datasets (Phi)	Beta
0.2642	50/50	10/10	0.6000	0.4000	0.4000	0.0500	0.0000	0.7358
0.4775	100/100	10/10	0.6000	0.4000	0.4000	0.0500	0.0000	0.5225
0.6481	150/150	10/10	0.6000	0.4000	0.4000	0.0500	0.0000	0.3519
0.7725	200/200	10/10	0.6000	0.4000	0.4000	0.0500	0.0000	0.2275
0.8576	250/250	10/10	0.6000	0.4000	0.4000	0.0500	0.0000	0.1424
0.9132	300/300	10/10	0.6000	0.4000	0.4000	0.0500	0.0000	0.0868
0.6253	50/50	10/10	0.7000	0.4000	0.4000	0.0500	0.0000	0.3747
0.9026	100/100	10/10	0.7000	0.4000	0.4000	0.0500	0.0000	0.0974
0.9793	150/150	10/10	0.7000	0.4000	0.4000	0.0500	0.0000	0.0207
0.9961	200/200	10/10	0.7000	0.4000	0.4000	0.0500	0.0000	0.0039
0.9993	250/250	10/10	0.7000	0.4000	0.4000	0.0500	0.0000	0.0007
0.9999	300/300	10/10	0.7000	0.4000	0.4000	0.0500	0.0000	0.0001

Report Definitions
H0, H1 abbreviate the null and alternative hypotheses, respectively.
Power is the probability of rejecting H0 when it is false. It should be close to one.
N1 & N2 are the sample sizes of datasets one and two, respectively.
K1 & K2 are the number of items in datasets one and two, respectively.
CA21 is the coefficient alpha in dataset two at which the power is calculated.
CA20 is the coefficient alpha in dataset two under H0.
CA1 is the coefficient alpha in dataset one.
Phi is the correlation between the average scores of each of the two datasets.
Alpha is the probability of a type-I error: rejecting H0 when it is true.
Beta is the probability of a type-II error: accepting H0 when it is false. Power = 1 - Beta.

Summary Statements
Samples of 10 items on 50 subjects in dataset one and 10 items on 50 subjects in dataset two achieve 26% power to detect the difference between the coefficient alphas in the two datasets. Under the null hypothesis, the coefficient alphas in datasets one and two are 0.4000 and 0.4000, respectively. The power is computed assuming that the coefficient alpha of dataset two is actually 0.6000. The test statistic used is the two-sided F-test. The significance level of the test was 0.0500.

This report shows the values of each of the parameters, one scenario per row. The values from this table are plotted in the chart below.

Plots Section



This plot shows the relationship between CA21, N1, and power.

Example2 - Finding the Sample Size

Continuing with the last example, find the sample size necessary to achieve a power of 90% at the 0.05 significance level.

Setup

This section presents the values of each of the parameters needed to run this example. You can make these changes directly on your screen or you can load the template entitled Example2 by clicking the Template tab and loading this template.

Option	Value
Data Tab	
Find	**N1**
CA1	**0.4**
CA20	**CA1**
CA21	**0.60 0.70**
K1	**10**
N1	*Ignored since this is the Find setting*
K2	**K1**
N2	**N1**
Alternative Hypothesis	**H1: CA1 <> CA2**
Phi	**0**
Alpha	**0.05**
Beta	**0.10**

Annotated Output

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

Numeric Results for Comparing Two Coefficient Alphas
H1: CA1 <> CA2

Power	Sample Sizes N1/N2	Number of Items K1/K2	Coef. Alpha Set 2|H1 (CA21)	Coef. Alpha Set 2|H0 (CA20)	Coef. Alpha Set 1 (CA1)	Signif. Level (Alpha)	Corr. Between Datasets (Phi)	Beta
0.9000	286/286	10/10	0.6000	0.4000	0.4000	0.0500	0.0000	0.1000
0.9026	100/100	10/10	0.7000	0.4000	0.4000	0.0500	0.0000	0.0974

This report shows that 286 subjects per dataset are needed when CA21 is 0.60 and 100 subjects per dataset are needed when CA21 is 0.70.

Example3 - Validation using Feldt

Feldt et al. (1999) presents an example in which CA1 = 0, CA20 = 0, CA21 = 0.5, alpha = 0.05, Phi = 0, N1 = N2 = 60,and *k* = 5. They find the power of a one-sided test to be 0.761.

Setup

This section presents the values of each of the parameters needed to run this example. You can make these changes directly on your screen or you can load the template entitled Example3 by clicking the Template tab and loading this template.

Option	Value
Find	**Beta and Power**
CA1	**0.0**
CA20	**CA1**
CA21	**0.5**
K1	**5**
N1	**60**
K2	**K1**
N2	**N1**
Alternative Hypothesis	**H1: CA1 < CA2**
Phi	**0**
Alpha	**0.05**
Beta	*Ignored since this is the Find setting*

Annotated Output

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

Numeric Results

Numeric Results when H1: CA0<>CA1

Power	Sample Sizes N1/N2	Number of Items K1/K2	Coef. Alpha Set 2|H1 (CA21)	Coef. Alpha Set 2|H0 (CA20)	Coef. Alpha Set 1 (CA1)	Signif. Level (Alpha)	Corr. Between Datasets (Phi)	Beta
0.7655	60/60	5/5	0.5000	0.0000	0.0000	0.0500	0.0000	0.2345

Note that ***PASS's*** result is slightly different from Feldt's because ***PASS*** uses fractional degrees of freedom and Feldt rounds to the closest integer. Although the difference in power is small, allowing fractional degrees of freedom is more accurate.

Chapter 850

Cox Proportional Hazards Regression

Cox proportional hazards regression models the relationship between the hazard function $\lambda(t|X)$ of survival time and *k* covariates using the following formula

$$\log\left(\frac{\lambda(t|X)}{\lambda_0(t)}\right) = \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k$$

where $\lambda_0(t)$ is the baseline hazard. Note that the covariates may be discrete or continuous.

This procedure calculates power and sample size for testing the hypothesis that $\beta_1 = 0$ versus the alternative that $\beta_1 = B$. Note that β_1 is the change in log hazards for a one-unit change in X_1 when the rest of the covariates are held constant. The procedure assumes that this hypothesis will be tested using the Wald (or score) statistic

$$z = \frac{\hat{\beta}_1}{\sqrt{Var(\hat{\beta}_1)}}$$

Power Calculations

Suppose you want to test the null hypothesis that $\beta_1 = 0$ versus the alternative that $\beta_1 = B$. Hsieh and Lavori (2000) gave a formula relating sample size, α, β, and *B* when X_1 is normally distributed. The sample size formula is

$$D = \frac{\left(z_{1-\alpha/2} + z_{1-\beta}\right)^2}{\left(1 - R^2\right)\sigma^2 B^2}$$

where *D* is the number of events, σ^2 is the variance of X_1, and R^2 is the proportion of variance explained by the multiple regression of X_1 on the remaining covariates. It is interesting to note that the number of censored observations does not enter in to the power calculations. To obtain a formula for the sample size, *N*, we inflate *D* by dividing by *P*, the proportion of subjects that fail.

Thus, the formula for *N* is

$$N = \frac{\left(z_{1-\alpha/2} + z_{1-\beta}\right)^2}{P\left(1 - R^2\right)\sigma^2 B^2}$$

This formula is an extension of an earlier formula for the case of a single, binary covariate derived by Schoenfeld (1983). Thus, it may be used with discrete or continuous covariates.

Assumptions

It is important to note that this formulation assumes that proportional hazards model with *k* covariates is valid. However, it does not assume exponential survival times.

Procedure Options

This section describes the options that are unique to this procedure. These are located on the panels associated with the Data, Options, and Reports tabs. To find out more about using the other tabs such as Plot Text, Axes, and Template, turn to the chapter entitled Procedure Templates.

Data Tab

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

Find

This option specifies the parameter to be solved for from the other parameters. Under most situations, you will select either *Beta* for a power analysis or *N* for sample size determination.

Select *N* when you want to calculate the sample size needed to achieve a given power and alpha level.

Select *Beta* when you want to calculate the power of an experiment.

B (Log Hazard Ratio)

This procedure calculates power or sample size for testing the hypothesis that $\beta_1 = 0$ versus the alternative that $\beta_1 = B$ in a Cox regression. Enter one or more values of *B* here.

B is the predicted change in log (base e) hazards corresponding to a one unit change in X1 when the other covariates are held constant. Thus, if you want to detect a hazard ratio of 1.5, enter ln(1.5) = 0.4055. Although any non-zero value may be entered, common values are between -3 and 3.

S (Std Deviation of X1)

Enter an estimate of the standard deviation of X1, the predictor variable of interest. The formulation used here assumes that X1 follows the normal distribution. However, you can obtain approximate results for non-normal variables by putting in the correct value here. For example, if X1 is binary, the standard deviation is given by $\sqrt{p(1-p)}$ where *p* is the proportion of either of the binary values in the population of X1.

If you don’t have an estimate, you can press the SD button to obtain a window that will help you determine a rough estimate of the standard deviation.

R-Squared Other X’s

This is the R-Squared that is obtained when X1 is regressed on the other X’s (covariates) in the model. Use this to account for the influence on power and sample size of adding other covariates. Note that the number of additional variables does not matter in this formulation. Only their overall relationship with X1 through this R-Squared value is used.

Of course, this value is restricted to being greater than or equal to zero and less than one. Use zero when there are no other covariates.

P (Overall Event Rate)

Enter one or more values for the event rate. The event rate is the proportion of subjects in which the event of interest occurs during the duration of the study. This is the proportion of non-censored subjects. Since the values entered here are proportions, they must be in the range $0 < P \le 1$.

Note that when this value is set to 1.0, the sample size is the number of events (deaths).

Hypothesis Test

Specify whether the test is one-sided or two-sided. When a two-sided hypothesis is selected, the value of alpha is halved by ***PASS***. Everything else remains the same.

Note that the accepted procedure is to use the Two Sided option unless you can justify using a one-sided test.

N (Sample Size)

This option specifies the total number of observations in the sample. You may enter a single value or a list of values.

Note that when the Overall Event Rate is set to 1.0, the sample size becomes the number of events.

Alpha (Significance Level)

This option specifies one or more values for the probability of a type-I error (alpha). A type-I error occurs when you reject the null hypothesis of equal probabilities when in fact they are equal.

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

Beta (1 - Power)

This option specifies one or more values for the probability of a type-II error (beta). A type-II error occurs when you fail to reject the null hypothesis of equal probabilities of the event of interest when in fact they are different.

Values must be between zero and one. Historically, the value of 0.20 was used for beta. Now, 0.10 is more popular. You should pick a value for beta that represents the risk of a type-II error you are willing to take.

Power is defined as one minus beta. Power is equal to the probability of rejecting a false null hypothesis. Hence, specifying the beta error level also specifies the power level. For example, if you specify beta values of 0.05, 0.10, and 0.20, you are specifying the corresponding power values of 0.95, 0.90, and 0.80.

Example1 - Power for Several Sample Sizes

Cox regression will be used to analyze the power of a survival time study. From past experience, the researchers want to evaluate the sample size needs for detecting regression coefficients of 0.2 and 0.3 for the independent variable of interest. The variable has a standard deviation of 1.20. The *R*-squared of this variable with seven other covariates is 0.18.

The event rate is thought to be 70% over the 3-year duration of the study. The researchers will test their hypothesis using a 5% significance level with a two-sided Wald test. They decide to calculate the power at sample sizes between 5 and 250.

Setup

You can enter these values yourself or load the Example1 template from the Template tab.

Option	Value
Data Tab	
Find	**Beta and Power**
B (Log Hazard Ratio)	**0.2 0.3**
S (Std Deviation of X1)	**1.2**
R-Squared Other X's	**0.18**
P (Overall Event Rate)	**0.70**
Hypothesis Test	**Two-Sided**
N	**5 to 250 by 40**
Alpha	**0.05**
Beta	*Ignored since this is the Find setting*

Annotated Output

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

Numeric Results

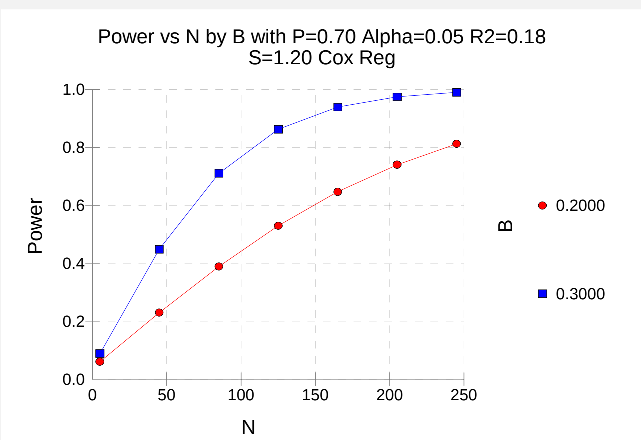
Power	Sample Size (N)	Reg. Coef. (B)	S.D. of X1 (SD)	Event Rate (P)	R-Squared X1 vs Other X's (R2)	Two-Sided Alpha	Beta
0.06017	5	0.2000	1.2000	0.7000	0.1800	0.05000	0.93983
0.22959	45	0.2000	1.2000	0.7000	0.1800	0.05000	0.77041
0.38837	85	0.2000	1.2000	0.7000	0.1800	0.05000	0.61163
0.52908	125	0.2000	1.2000	0.7000	0.1800	0.05000	0.47092
0.64643	165	0.2000	1.2000	0.7000	0.1800	0.05000	0.35357
0.74004	205	0.2000	1.2000	0.7000	0.1800	0.05000	0.25996
0.81223	245	0.2000	1.2000	0.7000	0.1800	0.05000	0.18777
0.08849	5	0.3000	1.2000	0.7000	0.1800	0.05000	0.91151
0.44815	45	0.3000	1.2000	0.7000	0.1800	0.05000	0.55185
0.71043	85	0.3000	1.2000	0.7000	0.1800	0.05000	0.28957
0.86202	125	0.3000	1.2000	0.7000	0.1800	0.05000	0.13798
0.93865	165	0.3000	1.2000	0.7000	0.1800	0.05000	0.06135
0.97412	205	0.3000	1.2000	0.7000	0.1800	0.05000	0.02588
0.98953	245	0.3000	1.2000	0.7000	0.1800	0.05000	0.01047

Report Definitions
Power is the probability of rejecting a false null hypothesis. It should be close to one.
N is the size of the sample drawn from the population.
B is the size of the regression coefficent to be detected.
SD is the standard deviation of X1.
P is the event rate.
R2 is the R-squared achieved when X1 is regressed on the other covariates.
Alpha is the probability of rejecting a true null hypothesis.
Beta is the probability of accepting a false null hypothesis.

Summary Statements
A Cox regression of the log hazard ratio on a covariate with a standard deviation of 1.2000 based on a sample of 5 observations achieves 6% power at a 0.93983 significance level to detect a regression coefficient equal to 0.2000. The sample size was adjusted since a multiple regression of the variable of interest on the other covariates in the Cox regression is expected to have an R-Squared of 0.1800. The sample size was adjusted for an anticipated event rate of 0.7000.

This report shows the power for each of the scenarios.

Plot Section



Example2 - Validation Using Hsieh

Hsieh and Lavori (2000) present an example which we will use to validate this program. In this example, B = 1.0, S = 0.3126, $R2$ = 0.1837, P = 0.738, one-sided alpha = 0.05, and power = 0.80. They calculated N = 107.

Setup

You can enter these values yourself or load the Example2 template from the Template tab.

Option	Value
Data Tab	
Find	**N**
B (Log Hazard Ratio)	**1.0**
S (Std Deviation of X1)	**0.3126**
R-Squared Other X's	**0.1837**
P (Overall Event Rate)	**0.738**
Hypothesis Test	**One-Sided**
N	*Ignored since this is the Find setting*
Alpha	**0.05**
Beta	**0.20**

Annotated Output

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

Numeric Results

Power	Sample Size (N)	Reg. Coef. (B)	S.D. of X1 (SD)	Event Rate (P)	R-Squared X1 vs Other X's (R2)	One-Sided Alpha	Beta
0.80321	106	1.000	0.313	0.738	0.184	0.05000	0.19679

Note that ***PASS*** calculated 106 rather than the 107 calculated by Hsieh and Lavori (2000). The discrepancy is due to the intermediate rounding that they did. To show this, we will run a second example from Hsieh and Lavori in which $R2$ = 0 and P = 1.0. In this case, N = 64.

Numeric Results with R2 = 0 and P = 1.0

Power	Sample Size (N)	Reg. Coef. (B)	S.D. of X1 (SD)	Event Rate (P)	R-Squared X1 vs Other X's (R2)	One-Sided Alpha	Beta
0.80399	64	1.000	0.313	1.000	0.000	0.05000	0.19601

Note that ***PASS*** also calculated 64. Hsieh and Lavori obtained the 107 by adjusting this 64 for P first and then for $R2$. ***PASS*** does both adjustments at once, obtaining the 106. Thus, the difference is due to intermediate rounding.

Example3 - Validation for Binary X1 using Schoenfeld

Schoenfeld (1983), page 502, presents an example for the case when X1 is binary. In this example, B = ln(1.5) = 0.4055, S = 0.5, $R2$ = 0.0, P = 0.71, one-sided alpha = 0.05, and power = 0.80. Schoenfeld calculated N = 212.

Setup

You can enter these values yourself or load the Example3 template from the Template tab.

Option	Value
Data Tab	
Find	**N**
B (Log Hazard Ratio)	**0.4055**
S (Std Deviation of X1)	**0.5**
R-Squared Other X's	**0.0**
P (Overall Event Rate)	**0.71**
Hypothesis Test	**One-Sided**
N	*Ignored since this is the Find setting*
Alpha	**0.05**
Beta	**0.20**

Annotated Output

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

Numeric Results

Power	Sample Size (N)	Reg. Coef. (B)	S.D. of X1 (SD)	Event Rate (P)	R-Squared X1 vs Other X's (R2)	One-Sided Alpha	Beta
0.80028	212	0.406	0.500	0.710	0.000	0.05000	0.19972

Note that ***PASS*** also obtains N = 212.

Chapter 855

Linear Regression

Introduction

Linear regression is a commonly used procedure in statistical analysis. One of the main objectives in linear regression analysis is to test hypotheses about the slope (sometimes called the regression coefficient) of the regression equation. This module calculates power and sample size for testing whether the slope is a value other than the value specified by the null hypothesis.

Difference between Linear Regression and Correlation

The correlation coefficient is used when both *X* and *Y* are from the normal distribution (in fact, the assumption actually is that *X* and *Y* follow a bivariate normal distribution). That is, *X* is assumed to be a random variable whose distribution is normal. In the linear regression context, no statement is made about the distribution of *X*. In fact, *X* is not even a random variable. Instead, it is a set of fixed values such as 10, 20, 30 or -1, 0, 1. Because of this difference in definition, we have included both Linear Regression and Correlation algorithms. They gave different results. This module deals with the Linear Regression (fixed *X*) case.

Technical Details

Suppose that the dependence of a variable *Y* on another variable *X* can be modeled using the simple linear equation

$$Y = A + BX$$

In this equation, *A* is the *Y*-intercept, *B* is the slope, *Y* is the dependent variable, and *X* is the independent variable.

The nature of the relationship between *Y* and *X* is studied using a sample of *N* observations. Each observation consists of a data pair: the *X* value and the *Y* value. The values of *A* and *B* are estimated from these observations. Since the linear equation will not fit the observations exactly, estimated values of *A* and *B* must be used. These estimates are found using the method of least squares. Using these estimated values, each data pair may be modeled using the equation

$$Y_i = a + bX_i + e_i$$

Note that *a* and *b* are the estimates of the population parameters *A* and *B*. The *e* values represent the discrepancies between the estimated values (*a* + *bX*) and the actual values *Y*. They are called the errors or residuals.

If it is assumed that these *e* values are normally distributed, tests of hypotheses about *A* and *B* can be constructed. Specifically, we can employ an *F* ratio to test the null hypothesis that the slope is *B0* versus the alternative hypothesis that the slope is *B* (*B* not *B0*). The power function of this F test can be written

$$Power = \Pr\left(F > F_{\alpha}\right)$$

where F_{α} is the critical value based on the central-*F* distribution with 1 and *N* - 2 degrees of freedom and the significant level α and *F* is distributed as a non-central *F* with degrees of freedom 1 and *N* - 2 and non-centrality parameter λ. The value of λ is

$$\lambda = N\left(\frac{SX\left(B - B0\right)}{\sigma}\right)^{2}$$

where σ^2 is the variance of the *e*'s and

$$SX = \sqrt{\frac{\sum_{i=1}^{N}\left(X_i - \overline{X}\right)^2}{N}}$$

The values of each of the parameters α, β, σ^2, *N*, *SX*, and *B* can be determined from the others using the above formulation.

Note that the power for a one-sided test may be found by using 2α for α in the above formulation.

Procedure Options

This section describes the options that are unique to this procedure. These are located on the panels associated with the Data, Options, and Reports tabs. To find out more about using the other tabs such as Plot Text, Axes, and Template, turn to the chapter entitled Procedure Templates.

Data Tab

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

Find

This option specifies the parameter to be solved for from the other parameters.

B0 (Slope|H0)

This is the value of the slope assumed by the null hypothesis. Often this value is set to zero, but this is not necessary. The alternative hypothesis is that the slope is not equal to this value.

B (Slope|H1)

This is the value of the slope at which the power is computed. The hypothesis being tested is that the slope is some value other than B0.

SX (Std Deviation of X's)

This is the standard deviation of the *X* values in the sample. It is not necessarily the standard deviation of *X* in the population. For example, suppose the *X* values are five -1's and five 1's. The computed standard deviation of these values (dividing by *N* rather than *N* - 1) is 1.0.

You will often want to compute the power or sample size for a specific set of *X* values. Instead of computing *SX* by hand, you can use the keyword *XS* (short for *X*'s) followed by the list of *X* values. For example, the phrase

XS -1,1

is translated into a 1.0 (which is the standard deviation of these two values). This calculation assumes that the sample is allocated equally to the two values. Hence, an *N* of 10 implies that five are assigned to -1 and five to 1.

If you are planning a study involving two random variables, X and Y, that come from a bivariate normal population, you should enter the actual standard deviation of X here.

Residual Variance Method

The standard deviation of the residuals is needed for the power and sample size calculations. These residuals are the e_i in the regression model

$$Y_i = a + bX_i + e_i$$

However, their standard deviation is not available until after the study is complete. ***PASS*** provides three methods for specifying the standard deviation of the residuals: *Std Deviation of Y*, *Correlation*, and specifying it directly.

SY (Std Deviation of Y)

Enter an estimate of the standard deviation of *Y*. This standard deviation ignores *X*. An estimate of this value must be found from previous studies, pilot studies, or using your best guess. This option is used when 'Residual Variance Method' is set to 'SY'.

When this value is used, the standard deviation of the residuals is computed using the relationship

$$\sigma = \sqrt{\sigma_Y^2 - B^2 SX^2}$$

This value must be greater than zero.

Correlation

Enter an estimate of the correlation between *Y* and the *X* values. An estimate of this correlation must be found from previous studies, pilot studies, or using your best guess. This value must be greater than zero and less than one—negative values are allowed. This option is used when 'Residual Variance Method' is set to 'R'.

When this method is used, the standard deviation of the residuals is computed using the relationship

$$\sigma = B(SX)\sqrt{1/R^2 - 1}$$

where *R* is the correlation.

S (Std Dev of Residuals)

Enter an estimate for the value of the standard deviation of the residuals. This option is used when 'Residual Variance Method' is set to 'S'.

N (Sample Size)

This is the number of observations in the study.

Alpha

This option specifies one or more values for the significance level--probability of a type-I error (alpha). A type-I error occurs when you reject the null hypothesis when in fact it is true.

The *F* test corresponds to a two-sided test of the hypothesis that slopes are equal versus the alternative that they are unequal.

Values between 0.001 and 0.100 are acceptable. The value of 0.05 has become the standard. This means that about one test in twenty will falsely reject the null hypothesis. Although 0.05 is the standard value, you should pick a value for alpha that represents the risk of a type-I error you are willing to take.

Note that you can enter a range of values such as *0.01,0.05* or *0.01 to 0.05 by 0.01*.

Beta (1-Power)

This option specifies one or more values for beta (the probability of accepting a false null hypothesis). Since statistical power is equal to one minus beta, specifying beta implicitly specifies the power. For example, setting beta at 0.20 also sets the power to 0.80.

Values must be between zero and one. The value of 0.20 has often used for beta. However, you should pick a value for beta that represents the risk of this type of error you are willing to take.

Note that you can enter a range of values such as *0.10,0.20* or *0.05 to 0.20 by 0.05*.

If your only interest is in determining the appropriate sample size for a confidence interval, set beta to 0.5.

Options Tab

This tab sets a couple of options used in the iterative procedures.

Maximum Iterations

Specify the maximum number of iterations allowed before the search for the criterion of interest is aborted. When the maximum number of iterations is reached without convergence, the criterion is left blank. A value of 500 is recommended.

Example1 - Calculating the Power

Suppose a power analysis must be conducted for a linear regression study that will test the relationship between two variables, *Y* and *X*. The test will look at the power using two significance levels, 0.01 and 0.05 and several sample sizes between 5 and 85. Based on previous studies, the standard deviation of *Y* will be assumed to be 1.0. The standard deviation of the *X*'s in the sample will also be assumed as 1.0. The experimenter decides that unless the slope is at least 0.5, the relationship between *X* and *Y* is too weak to be considered.

Setup

You can enter these values yourself or load the Example1 template from the Template tab.

Option	Value
Data Tab	
Find	**Beta and Power**
B0	**0.0**
B	**0.5**
SX	**1**
Residual Variance Method	**Std Deviation of Y**
SY	**1**
N	**5 to 85 by 10**
Alpha	**0.01,0.05**
Beta	*Ignored*

Annotated Output

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

Numeric Results

Numeric Results for Two-Sided Testing of B=B0 where B0 = 0.00

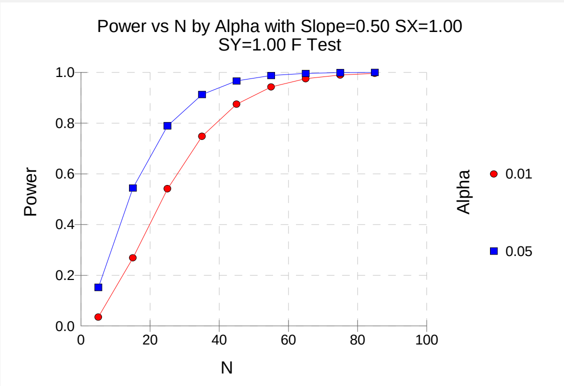
Power	Sample Size (N)	Slope (B)	Standard Deviation of X (SX)	Standard Deviation of Y (SY)	Standard Deviation of Residuals (S)	Alpha	Beta
0.03533	5	0.50	1.00	1.00	0.87	0.01000	0.96467
0.15194	5	0.50	1.00	1.00	0.87	0.05000	0.84806
0.26836	15	0.50	1.00	1.00	0.87	0.01000	0.73164
0.54369	15	0.50	1.00	1.00	0.87	0.05000	0.45631
0.54097	25	0.50	1.00	1.00	0.87	0.01000	0.45903
0.78944	25	0.50	1.00	1.00	0.87	0.05000	0.21056
0.74759	35	0.50	1.00	1.00	0.87	0.01000	0.25241
0.91225	35	0.50	1.00	1.00	0.87	0.05000	0.08775
0.87415	45	0.50	1.00	1.00	0.87	0.01000	0.12585
0.96601	45	0.50	1.00	1.00	0.87	0.05000	0.03399
0.94183	55	0.50	1.00	1.00	0.87	0.01000	0.05817
0.98755	55	0.50	1.00	1.00	0.87	0.05000	0.01245
0.97470	65	0.50	1.00	1.00	0.87	0.01000	0.02530
0.99564	65	0.50	1.00	1.00	0.87	0.05000	0.00436
0.98953	75	0.50	1.00	1.00	0.87	0.01000	0.01047
0.99853	75	0.50	1.00	1.00	0.87	0.05000	0.00147
0.99585	85	0.50	1.00	1.00	0.87	0.01000	0.00415
0.99952	85	0.50	1.00	1.00	0.87	0.05000	0.00048

Report Definitions
Power is the probability of rejecting a false null hypothesis. It should be close to one.
N is the size of the sample drawn from the population. To conserve resources, it should be small.
B0 is the slope under the null hypothesis.
B is the slope at which the power is calculated.
SX is the standard deviation of the X values.
SY is the standard deviation of Y.
Alpha is the probability of rejecting a true null hypothesis. It should be small.
Beta is the probability of accepting a false null hypothesis. It should be small.

Summary Statements
A sample size of 5 achieves 4% power to detect a change in slope from 0.00 under the null hypothesis to 0.50 under the alternative hypothesis when the standard deviation of the X's is 1.00, the standard deviation of Y is 1.00, and the two-sided significance level is 0.01000.

This report shows the calculated sample size for each of the scenarios.

Plot Section



This plot shows the power versus the sample size for the two values of alpha.

Example2 - Validation using Neter, Wasserman, and Kutner

Problem Statement

Neter, Wasserman, and Kutner (1983) pages 71 and 72 present a power analysis when N = 10, *Slope* = 0.25, alpha = 0.05, $SX = \sqrt{(3400/10)} = 18.439$, and $SY = \sqrt{10 + (0.25)^2(3400/10)} = 5.59015$. They found the power to be approximately 0.97.

Setup

You can enter these values yourself or load the Example2 template from the Template tab.

Option	**Value**
Data Tab	
Find	**Beta and Power**
B0	**0.00**
B	**0.25**
SX	**18.439**
Residual Variance Method	**Std Deviation of Y**
SY	**5.59015**
N	**10**
Alpha	**0.05**
Beta	**Ignored**

Annotated Output

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

Numeric Results

Numeric Results for Two-Sided Testing of B=B0 where B0 = 0.00

Power	Sample Size (N)	Slope (B)	Standard Deviation of X (SX)	Standard Deviation of Y (SY)	Standard Deviation of Residuals (S)	Alpha	Beta
0.97975	10	0.25	18.44	5.59	3.16	0.05000	0.02025

The power of 0.97975 matches their result to two decimals. Note that they used interpolation from a table to obtain their answer.

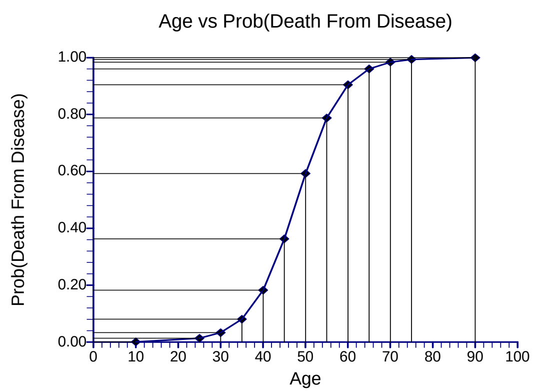
Chapter 860

Logistic Regression

Introduction

Logistic regression expresses the relationship between a binary response variable and one or more independent variables called *covariates*. A covariate can be discrete or continuous.

Consider a study of death from disease at various ages. This can be put in a logistic regression format as follows. Let a binary response variable Y be one if death has occurred and zero if not. Let X be the individual's age. Suppose a large group of various ages is followed for ten years and then both Y and X are recorded for each person. In order to study the pattern of death versus age, the age values are grouped into intervals and the proportions that have died in each age group are calculated. The results are displayed in the following plot.



As you would expect, as age increases, the proportion dying of disease increases. However, since the proportion dying is bounded below by zero and above by one, the relationship is approximated by an "S" shaped curve. Although a straight-line might be used to summarize the relationship between ages 40 and 60, it certainly could not be used for the young or the elderly.

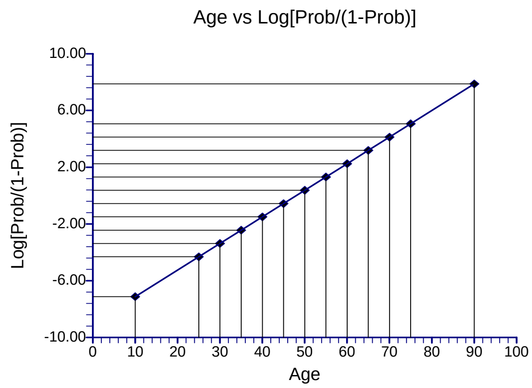
Under the logistic model, the proportion dying, P, at a given age can be calculated using the formula

$$P = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}}$$

This formula can be rearranged so that it is linear in X as follows

$$Log\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 X$$

Note that the left side is the logarithm of the odds of death versus non-death and the right side is a linear equation for *X*. This is sometimes called the *logit* transformation of *P*. When the scale of the vertical axis of the plot is modified using the logit transformation, the following straight-line plot results.



In the logistic regression model, the influence of *X* on *Y* is measured by the value of the slope of *X* which we have called β_1. The hypothesis that $\beta_1 = 0$ versus the alternative that $\beta_1 = B \neq 0$ is of interest since if $\beta_1 = 0$, *X* is not related to *Y*.

Under the alternative hypothesis that $\beta_1 = B$, the logistic model becomes

$$\log\left(\frac{P_1}{1 - P_1}\right) = \beta_0 + BX$$

Under the null hypothesis, this reduces to

$$\log\left(\frac{P_0}{1 - P_0}\right) = \beta_0$$

To test whether the slope is zero at a given value of *X*, the difference between these to quantities is formed giving

$$\beta_0 + BX - \beta_0 = \log\left(\frac{P_1}{1 - P_1}\right) - \log\left(\frac{P_0}{1 - P_0}\right)$$

which reduces to

$$\begin{aligned} BX &= \log\left(\frac{P_1}{1 - P_1}\right) - \log\left(\frac{P_0}{1 - P_0}\right) \\ &= \log\left(\frac{P_1 / \left(1 - P_1\right)}{P_0 / \left(1 - P_0\right)}\right) \\ &= \log\left(OR\right) \end{aligned}$$

where *OR* is odds ratio of P_1 and P_0. This relationship may be solved for *OR* giving

$$OR = e^{BX}$$

This shows that the odds ratio of P_1 and P_0 is directly related to the slope of the logistic regression equation. It also shows that the value of the odds ratio depends on the value of *X*. For a given value of *X*, testing that *B* is zero is equivalent to testing *OR* is one. Since *OR* is commonly used and well understood, it is used as a measure of effect size in power analysis and sample size calculations.

Power Calculations

Suppose you want to test the null hypothesis that $\beta_1 = 0$ versus the alternative that $\beta_1 = B$. Hsieh, Block, and Larsen (1998) have presented formulae relating sample size, α, power, and *B* for two situations: when X_1 is normally distributed and when X_1 is binomially distributed.

When X_1 is normally distributed, the sample size formula is

$$N = \frac{\left(z_{1-\alpha/2} + z_{1-\beta}\right)^2}{P^*\left(1-P^*\right)B^2}$$

where P^* is the event rate (probability that $Y = 1$) at the mean of X_1. Note that *B* is defined in terms of an increase of one standard deviation of X_1 above the mean.

When X_1 is binomially distributed and $X_1 = 0$ or 1, the sample size formula is

$$N = \frac{\left(z_{1-\alpha/2}\sqrt{\frac{\overline{P}\left(1-\overline{P}\right)}{R}} + z_{1-\beta}\sqrt{P_0\left(1-P_0\right) + \frac{P_1\left(1-P_1\right)\left(1-R\right)}{R}}\right)^2}{\left(P_0 - P_1\right)^2\left(1-R\right)}$$

where P_0 is the event rate at $X_1 = 0$ and P_1 is the event rate at $X_1 = 1$, *R* is the proportion of the sample with $X_1 = 1$, and $\overline{P}$ is the overall event rate given by

$$\overline{P} = \left(1-R\right)P_0 + R\left(P_1\right).$$

Multiple Logistic Regression

The multiple logistic regression model relates the probability distribution of Y to two or more covariates $X_1, X_2, \cdots, X_k$ by the formula

$$\log\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 X_1 + ... + \beta_k X_k$$

where P is the probability that $Y = 1$ given the values of the covariates. It is a simple extension of the simple logistic regression model that was just presented. In power analysis and sample size work, attention is placed on a single covariate while the influence of the other covariates is statistically removed by placing them at their mean values.

When there are multiple covariates, the following adjustment was given by Hsieh (1998) to give the total sample size, N_m

$$N_m = \frac{N}{1-\rho^2}$$

where ρ is the multiple correlation coefficient between X_1 (the variable of interest) and the remaining covariates. Notice that the number of extra covariates does not matter in this approximation.

Procedure Options

This section describes the options that are unique to this procedure. These are located on the panels associated with the Data, Options, and Reports tabs. To find out more about using the other tabs such as Plot Text, Axes, and Template, turn to the chapter entitled Procedure Templates.

Data Tab

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

Find

This option specifies the parameter to be solved for from the other parameters. The parameters that may be selected are *P1*, *N*, *Alpha*, and *Beta*. Under most situations, you will select either *Beta* for a power analysis or *N* for sample size determination.

Select *N* when you want to calculate the sample size needed to achieve a given power and alpha level.

Select *Beta* when you want to calculate the power of an experiment.

X (Independent Variable)

This option specifies whether the covariate is binary (binomial) or continuous (normal). This is a very important distinction since the sample size required for a particular power level is much larger for a binary covariate than for a continuous covariate.

This selection also changes the meaning of *P0* and *P1*.

P0 (Baseline Prob Y=1)

This option specifies one or more P_0 values. The interpretation of P_0 depends on whether X_1 is binary or continuous.

Binomial Covariate

When X_1 is binary, P_0 is the probability that $Y = 1$ when $X_1 = 0$. All other covariates are assumed to be equal to their mean values. In this case, the logistic equation reduces to

$$\log\left(\frac{P_0}{1-P_0}\right) = \beta_0$$

so that

$$P_0 = \frac{e^{\beta_0}}{1+e^{\beta_0}}$$

Normal Covariate

When X_1 is normally distributed, P_0 is the probability that Y = 1 when $X_1 = \mu_{X_1}$, where μ_{X_1} is the mean of X_1. That is, P_0 is the baseline probability that Y = 1 when X_1 is ignored. All other covariates are assumed to be equal to their mean values. In this case, the logistic equation reduces to

$$\log\left(\frac{P_0}{1 - P_0}\right) = \beta_0 + \beta_1 \mu_{X_1}$$

so that

$$P_0 = \frac{e^{\beta_0 + \beta_1 \mu_{X_1}}}{1 + e^{\beta_0 + \beta_1 \mu_{X_1}}}$$

Use P1 or Odds Ratio

This option specifies the whether to specify *P1* directly or to specify it by specifying the odds ratio. Since the relationship between the odds ration, *P1*, and *P0* is given by

$$OR = \frac{P_1 / \left(1 - P_1\right)}{P_0 / \left(1 - P_0\right)}$$

specifying *OR* and *P0* implicitly specifies *P1*.

This options lets you specify whether you want to state the alternative hypothesis in terms of *P1* or the odds ratio.

P1 (Alt. Prob Y =1)

This option specifies the effect size to be detected by specifying P_1. As was shown earlier, the slope of the logistic regression can be expressed in terms of P_0 and P_1. Hence, by specifying P_1, you are also specifying the slope.

This option is only used when the User P1 or Odds Ratio option is set to *P1*. Its interpretation depends on whether X_1 is binomial or normal.

Binomial Covariate

When X_1 is binary, *P1* is the probability that Y = 1 when X_1 = 1. All other covariates are assumed to be equal to their mean values. In this case, the logistic equation reduces to

$$\log\left(\frac{P_1}{1 - P_1}\right) = \beta_0 + \beta_1$$

since X_1 = 1.

Normal Covariate

When X_1 is normally distributed, *P1* is the probability that *Y* = 1 when $X_1 = \mu_{x_1} + \sigma_{x_1}$. That is, when x_1 is one standard deviation above the mean. All other covariates are assumed to be equal to their mean values. In this case, the logistic equation reduces to

$$\log\left(\frac{P_1}{1 - P_1}\right) = \beta_0 + \beta_1 x_1$$

so that

$$P_1 = \frac{e^{\beta_0 + \beta_1 x_1}}{1 + e^{\beta_0 + \beta_1 x_1}}$$

Odds Ratio (Odds1/Odds0)

This option specifies the odds ratio to be detected by the study. As was shown earlier, the slope of the logistic regression can be expressed in terms of P_0 and the odds ratio. Hence, by specifying *OR*, you are also specifying the slope. Using the formula

$$P_1 = \frac{OR(P_0)}{1 - P_0 + OR(P_0)}$$

specifying *OR* and P_0 implicitly specifies P_1.

This option is only used when the User P1 or Odds Ratio option is set to *Odds Ratio*. Its interpretation depends on whether X_1 is binomial or normal.

Binomial Covariate

When X_1 is binary, this option gives the odds ratio of P_1 and P_0. All other covariates are assumed to be equal to their mean values. In this case, the logistic equation reduces to

$$\log\left(\frac{P_1}{1 - P_1}\right) = \beta_0 + \beta_1$$

since X_1 = 1.

This odds ratio compares the odds of obtaining *Y* = 1 when X_1 = 1 to the odds of obtaining *Y* = 1 when X_1 = 0.

Normal Covariate

When X_1 is normally distributed, this option gives the odds ratio of P_1 and P_0, where P_1 is the probability that $Y = 1$ when $X_1 = x_1$, where x_1 is a value other than μ_{x_1}. All other covariates are assumed to be equal to their mean values. In this case, the logistic equation reduces to

$$\log\left(\frac{P_1}{1 - P_1}\right) = \beta_0 + \beta_1 x_1$$

so that

$$P_1 = \frac{e^{\beta_0 + \beta_1 x_1}}{1 + e^{\beta_0 + \beta_1 x_1}}$$

This odds ratio compares the odds of obtaining $Y = 1$ when $X_1 = x_1$ to the odds of obtaining Y when $X_1 = \mu_{x_1}$.

R-Squared Other X's

This is the R-Squared that is obtained when X_1 is regressed on the other *X's* (covariates) in the model. Use this to study the influence on power and sample size of adding other covariates. Note that the number of additional variables does not matter in this formulation. Only their overall relationship with X_1 through this R-Squared value is used.

Of course, this value is restricted to being greater than or equal to zero and less than one. Use zero when there are no other covariates.

Hypothesis Test

Specify whether the test is one-sided or two-sided. When a two-sided hypothesis is selected, the value of alpha is halved by ***PASS***. Everything else remains the same.

Note that the accepted procedure is to use the Two Sided option unless you can justify using a one-sided test.

N (Sample Size)

This option specifies the total number of observations in the sample. You may enter a single value or a list of values.

% N with X = 1 (Binary Only)

When X_1 is binary, this option specifies the proportion, *R*, of the sample in which $X_1 = 1$. Note that the value is specified as a percentage.

Alpha (Significance Level)

This option specifies one or more values for the probability of a type-I error (alpha). A type-I error occurs when you reject the null hypothesis of equal probabilities when in fact they are equal.

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

Beta (1 - Power)

This option specifies one or more values for the probability of a type-II error (beta). A type-II error occurs when you fail to reject the null hypothesis of equal probabilities of the event of interest when in fact they are different.

Values must be between zero and one. Historically, the value of 0.20 was used for beta. Now, 0.10 is becoming more popular. You should pick a value for beta that represents the risk of a type-II error you are willing to take.

Power is defined as one minus beta. Power is equal to the probability of rejecting a false null hypothesis. Hence, specifying the beta error level also specifies the power level. For example, if you specify beta values of 0.05, 0.10, and 0.20, you are specifying the corresponding power values of 0.95, 0.90, and 0.80, respectively.

Example1 - Power for a Continuous Covariate

A study is to be undertaken to study the relationship between post-traumatic stress disorder and heart rate after viewing video tapes containing violent sequences. Heart rate is assumed to be normally distributed. The event rate is thought to be 7% among soldiers. The researchers want a sample size large enough to detect an odds ratios of 1.5 or 2.0 with 90% power at the 0.05 significance level with a two-sided test. They decide to calculate the power at level sample sizes between 20 and 1200.

Setup

You can enter these values yourself or load the Example1 template from the Template tab.

Option	Value
Data Tab	
Find	**Beta and Power**
P0	**0.07**
Use P1 or Odds Ratio	**Odds Ratio**
Odds Ratio	**1.5 2.0**
R-Squared Other X's	**0**
X (Independent Variable)	**Continuous (Normal)**
Hypothesis Test	**Two-Sided**
N	**20 50 100 200 300 500 700 1000 1200**
% N with X = 1	*Ignored since X is continuous*
Alpha	**0.05**
Beta	*Ignored since this is the Find setting*

Annotated Output

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

Numeric Results

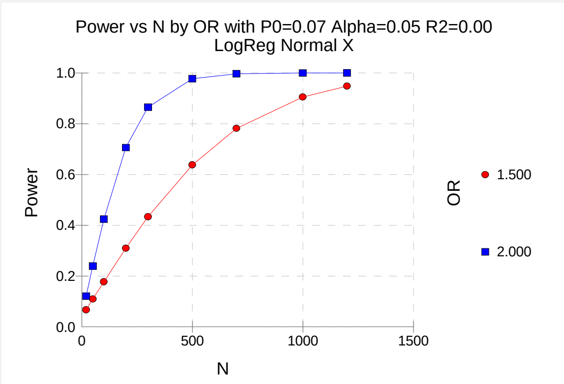
Power	N	P0	P1	Odds Ratio	R Squared	Alpha	Beta
0.06716	20	0.070	0.101	1.500	0.000	0.05000	0.93284
0.10964	50	0.070	0.101	1.500	0.000	0.05000	0.89036
0.17737	100	0.070	0.101	1.500	0.000	0.05000	0.82263
0.30962	200	0.070	0.101	1.500	0.000	0.05000	0.69038
0.43325	300	0.070	0.101	1.500	0.000	0.05000	0.56675
0.63808	500	0.070	0.101	1.500	0.000	0.05000	0.36192
0.78147	700	0.070	0.101	1.500	0.000	0.05000	0.21853
0.90516	1000	0.070	0.101	1.500	0.000	0.05000	0.09484
0.94779	1200	0.070	0.101	1.500	0.000	0.05000	0.05221
0.12119	20	0.070	0.131	2.000	0.000	0.05000	0.87881
0.23903	50	0.070	0.131	2.000	0.000	0.05000	0.76097
0.42410	100	0.070	0.131	2.000	0.000	0.05000	0.57590
0.70579	200	0.070	0.131	2.000	0.000	0.05000	0.29421
0.86504	300	0.070	0.131	2.000	0.000	0.05000	0.13496
0.97696	500	0.070	0.131	2.000	0.000	0.05000	0.02304
0.99673	700	0.070	0.131	2.000	0.000	0.05000	0.00327
0.99986	1000	0.070	0.131	2.000	0.000	0.05000	0.00014
0.99998	1200	0.070	0.131	2.000	0.000	0.05000	0.00002

Report Definitions
Power is the probability of rejecting a false null hypothesis. It should be close to one.
N is the size of the sample drawn from the population.
P0 is the response probability at the mean of the covariate, X.
P1 is the response probability when X is increased to one standard deviation above the mean.
Odds Ratio is the odds ratio when P1 is on top. That is, it is [P1/(1-P1)]/[P0/(1-P0)].
R-Squared is the R2 achieved when X is regressed on the other independent variables in the regression.
Alpha is the probability of rejecting a true null hypothesis.
Beta is the probability of accepting a false null hypothesis.

Summary Statements
A logistic regression of a binary response variable (Y) on a continuous, normally distributed, independent variable (X) with a sample size of 20 observations achieves 7% power at a 0.050 significance level to detect a change in Prob(Y=1) from the value of 0.070 at the mean of X to 0.101 when X is increased to one standard deviation above the mean.This change corresponds to an odds ratio of 1.500.

This report shows the power for each of the scenarios. The report shows that a power of 90% is reached at a sample size of about 300 for an odds ratio of 2.0 and 1000 for an odds ratio of 1.5.

Plot Section



This plot shows the power versus the sample size for the two values of the odds ratio.

Example2 - Sample Size for a Continuous Covariate

Continuing with the previous study, determine the exact sample size necessary to attain a power of 90%.

Setup

You can enter these values yourself or load the Example2 template from the Template tab.

Option	Value
Data Tab	
Find	**N**
P0	**0.07**
Use P1 or Odds Ratio	**Odds Ratio**
Odds Ratio	**1.5 2.0**
R-Squared Other X's	**0**
X (Independent Variable)	**Continuous (Normal)**
Hypothesis Test	**Two-Sided**
N	*Ignored since this is the Find setting*
% N with X = 1	*Ignored since X is continuous*
Alpha	**0.05**
Beta	**0.10**

Annotated Output

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

Numeric Results

Power	N	P0	P1	Odds Ratio	R Squared	Alpha	Beta
0.89978	981	0.070	0.101	1.500	0.000	0.05000	0.10022
0.89920	335	0.070	0.131	2.000	0.000	0.05000	0.10080

This report shows the power for each of the scenarios. The report shows that a power of 90% is achieved at a sample size of 335 for an odds ratio of 2.0 and 981 for an odds ratio of 1.5.

Example3 - Effect Size for a Continuous Covariate

Continuing the previous study, suppose the researchers can only afford a sample size of 500 individuals. They want to determine if a meaningful odds ratio can be detected with this sample size.

Setup

You can enter these values yourself or load the Example3 template from the Template tab.

Option	**Value**
Data Tab	
Find	**P1>P0 or Odds Ratio>1**
P0	**0.07**
Use P1 or Odds Ratio	**Odds Ratio**
Odds Ratio	*Ignored since this is the Find setting*
R-Squared Other X's	**0**
X (Independent Variable)	**Continuous (Normal)**
Hypothesis Test	**Two-Sided**
N	**500**
% N with X = 1	*Ignored since X is continuous*
Alpha	**0.05**
Beta	**0.10**

Annotated Output

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

Numeric Results

Power	N	P0	P1	Odds Ratio	R Squared	Alpha	Beta
0.90000	500	0.070	0.117	1.765	0.000	0.05000	0.10000

This report shows that this experimental design can detect an odds ratio of 1.765. That is, it can detect a shift in the event rate from 0.070 to 0.117.

Example4 - Sample Size for a Binary Covariate

A study is to be undertaken to study the relationship between post-traumatic stress disorder and gender. The event rate is thought to be 7% among males. The researchers want a sample size large enough to detect an odds ratio of 1.5 with 90% power at the 0.05 significance level with a two-sided test.

Setup

You can enter these values yourself or load the Example4 template from the Template tab.

Option	Value
Data Tab	
Find	**N**
P0	**0.07**
Use P1 or Odds Ratio	**Odds Ratio**
Odds Ratio	**1.5**
R-Squared Other X's	**0**
X (Independent Variable)	**Binary (X=0 or 1)**
Hypothesis Test	**Two-Sided**
N	*Ignored since this is the Find setting*
% N with X = 1	**50**
Alpha	**0.05**
Beta	**0.10**

Annotated Output

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

Numeric Results

Power	N	Pcnt N X=1	P0	P1	Odds Ratio	R Squared	Alpha	Beta
0.89997	3326	50.000	0.070	0.101	1.500	0.000	0.05000	0.10003

The sample size is estimated at 3326. This should be evenly divided among males and females.

Example5 - Validation for a Continuous Covariate

Hsieh (1998) page 1628 gives the power as 95% when N = 317, alpha = 0.05 (two-sided), $P0$ = 0.5, and the odds ratio is 1.5. The covariate is assumed to be continuous.

Setup

You can enter these values yourself or load the Example5 template from the Template tab.

Option	Value
Data Tab	
Find	**Beta and Power**
P0	**0.50**
Use P1 or Odds Ratio	**Odds Ratio**
Odds Ratio	**1.5**
R-Squared Other X's	**0**
X (Independent Variable)	**Continuous (Normal)**
Hypothesis Test	**Two-Sided**
N	**317**
Alpha	**0.05**
Beta	*Ignored since this is the Find setting*

Annotated Output

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

Numeric Results

Power	N	P0	P1	Odds Ratio	R Squared	Alpha	Beta
0.95049	317	0.500	0.600	1.500	0.000	0.05000	0.04951

PASS calculates a power of 0.95049 which matches Hsieh.

Example6 - Validation for a Binary Covariate

Hsieh (1998) page 1626 gives the power as 95% when *N* = 1282 (equal sample sizes for both groups), alpha = 0.05 (two-sided), *P0* = 0.4, and the *P1* = 0.5. The covariate is assumed to be binary.

Setup

You can enter these values yourself or load the Example5 template from the Template tab.

Option	Value
Data Tab	
Find	**Beta and Power**
P0	**0.4**
Use P1 or Odds Ratio	**P1**
P1	**0.5**
R-Squared Other X's	**0**
X (Independent Variable)	**Binary (X=0 or 1)**
Hypothesis Test	**Two-Sided**
N	**1282**
% N with X = 1	**50**
Alpha	**0.05**
Beta	*Ignored since this is the Find setting*

Annotated Output

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

Numeric Results

Power	N	Pcnt N X=1	P0	P1	Odds Ratio	R Squared	Alpha	Beta
0.95021	1282	50.000	0.400	0.500	1.500	0.000	0.05000	0.04979

PASS calculates a power of 0.95021 which matches Hsieh.

Chapter 865

Multiple Regression

Introduction

The Linear Regression chapter dealt with the regression of a dependent variable on a single independent variable. This module deals with the general case of multiple regression in which there are two or more independent variables.

Instead of using the correlation coefficient as the index of interest, the square of the correlation coefficient, R^2, has been adopted, mainly because of its ease of interpretation and its ability to be partitioned among various subgroups of the independent variables. R^2 ranges from zero to one.

When performing a regression analysis, a typical hypothesis involves testing the significance of a subgroup of the independent variables after considering a second, nonoverlapping, group of independent variables. For example, suppose you have five independent variables. One common hypothesis tests whether a certain variable is influential (has a nonzero coefficient in the regression equation) after considering the other four variables. To perform this test, you partition the R^2 from fitting all five variables into the R^2 of the first four variables and the R^2 added by the fifth variable. An *F*-test is constructed that tests whether this second R^2 value is significantly different from zero. If it is, the fifth variable is significant after adjusting for the other four variables.

Definition of Terms

Cohen (1988) provides a comprehensive discussion of this topic. We strongly suggest that you refer to the examples given there for a thorough understanding of this subject. We will provide a brief introduction, concentrating on defining the terms and statistics that are used.

Consider a sample of *N* rows of data. Suppose each row consists of the value of a dependent variable, *Y*, followed by the values of *k* independent variables, $X_1, X_2, \cdots, X_k$. The multiple regression equation of *Y* on the *X's* is

$$Y_j = \beta_0 + \beta_1 X_{1j} + \beta_2 X_{2j} + \cdots + \beta_k X_{kj} + \varepsilon_j$$

The $\beta's$ are called the population regression coefficients or beta weights. They represent the true values in the population from which the sample was taken. Their sample estimates, $b_0, b_1, b_2, \cdots, b_k$, are calculated from the data using least squares methodology. The estimated regression equation becomes

$$Y_j = b_0 + b_1 X_{1j} + b_2 X_{2j} + \cdots + b_k X_{kj} + e_j$$

Upon removing the residuals, e_j, the predicted values are calculated as

$$\hat{Y}_j = b_0 + b_1 X_{1j} + b_2 X_{2j} + \cdots + b_k X_{kj}$$

R^2 is defined as the proportion of the variation of *Y* that is explained (accounted for) by the variation in the *X*'s. The equation for R^2 is

$$R^2 = \frac{\sum_{i=1}^{N} \left(\hat{Y}_i - \overline{Y}\right)^2}{\sum_{i=1}^{N} \left(Y_i - \overline{Y}\right)^2}$$

or

$$R^2 = 1 - \frac{\sum_{i=1}^{N} e_i^2}{\sum_{i=1}^{N} \left(Y_i - \overline{Y}\right)^2}$$

Notice that the numerator sum of squares in the above equations depend on the *X*'s through $\hat{Y}_i$. Therefore, each time you change the *X*'s included in the regression equation, you will obtain a different value of R^2. Hence, you need special notation to indicate on which *X*'s the particular R^2 is based.

Let *C* and *T* refer to subgroups of the *X*'s. Define $R^2_{T|C} = R^2_{TC} - R^2_C$ to be the R^2 added when *Y* is regressed on the variables in set *T* after the variables in set *C*. Define R^2_C to be the R^2 achieved when *Y* is regressed on those in set *C* ignoring the variables in set *T*. Define R^2_{TC} to be the R^2 achieved when *Y* is regressed on the variables in both sets *C* and *T*.

Using the above notation, you can construct *F*-tests that will test whether the $\beta' s$ corresponding to certain sets of *X's* are simultaneously zero while controlling for other variables. For example, to test the significance of the *X's* in set *T* while removing the influence of the *X's* in set *C* from experimental error, you would use:

$$F_{u,v} = \frac{\left(R^2_{T|C}\right)/u}{\left(1 - R^2_C - R^2_{T|C}\right)/v}$$

where *u* is the number of variables in *T*, *v* = *N*-*k*-*1*, and *k* is the total number of variables in *C* and *T*. Note that this formula includes the effect size, f^2.

$$f^2 = \left(\frac{R^2_{T|C}}{1 - R^2_C - R^2_{T|C}}\right)$$

Most significance tests in regression analysis, correlation analysis, analysis of variance, and analysis of covariance may be constructed using these F-ratios.

For example, suppose the seven *X's* available are dividing into two sets: set *C* includes variables X_1, X_3, X_6 and set *T* includes X_2, X_4, X_5, X_7. The above *F*-ratio tests the null hypothesis that $\beta_2 = \beta_4 = \beta_5 = \beta_7 = 0$ after adjusting for X_1, X_3, X_6.

Calculating the Power

The calculation of the power of a particular test proceeds as follows:

1. Determine the critical value, $F_{u,v,\alpha}$ where *u* is the numerator degrees of freedom, *v* is the denominator degrees of freedom, and α is the probability of a type-I error.
2. Calculate the noncentrality parameter λ using the formula:

$$\lambda = N \frac{R^2_{T|C}}{\left(1 - R^2_C - R^2_{T|C}\right)}$$

3. Compute the power as the probability of being greater than $F_{u,v,\alpha}$ on a noncentral-*F* distribution with noncentrality parameter λ.

Note that the formula for λ is different from that used in ***PASS 6.0***. The algorithm used in ***PASS 6.0*** was based on formula (9.3.1) in Cohen (1988) which gives approximate answers. This version of ***PASS*** using an algorithm that gives exact answers.

Procedure Options

This section describes the options that are unique to this procedure. These are located on the panels associated with the Data, Options, and Reports tabs. To find out more about using the other tabs such as Plot Text, Axes, and Template, turn to the chapter entitled Procedure Templates.

Data Tab

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

Find

This option specifies the parameter to be solved for from the other parameters. The parameters that may be selected are *R2*, *N*, *Alpha*, and *Beta*. Under most situations, you will select either *Beta* or *N*.

Select *N* when you want to calculate the sample size needed to achieve a given power and alpha level.

Select *Beta* when you want to calculate the power of an experiment.

T: Variables Tested

These options refer to the set of independent variables (*X's*) whose coefficients are to be tested to determine if they all are zero. That is, this is the set of variables to be tested for statistical significance.

Number

This option specifies the number of variables in *T*. This is *u*. This number must be greater than or equal to one. Note that the total number of variables specified in the two Number options must be less than $N - 1$.

R2 Added

This box specifies the increase in R^2 due to the variables in set *T* after including the variables in set *C* in the regression equation. Note that this amount must be between zero and one and that the total of the two R^2 values must be less than one.

C: Variables Controlled

These options refer to the independent variables (*X's*) to be controlled (adjusted) for (these are the variables in set *C*). That is, the influence of these variables is statistically removed. These are variables that are included in the regression equation but are not being tested for statistical significance.

Number

This option specifies the number of variables in *C*. This number must be greater than or equal to zero. Note that the total number of variables specified in the two Number options must be less than *N-1*.

R2 Added

This box specifies the R^2 achieved by the variables in set *C* when they are fit alone in the regression equation. Note that this amount must be between zero and one and that the total of the two R^2 values must be less than one.

N (Sample Size)

This option specifies the value(s) for *N*, the sample size. Note that this value must be at least one greater than the total number of variables specified in sets *C* and *T*.

Alpha (Significance Level)

This option specifies one or more values for the probability of a type-I error (alpha). A type-I error occurs when you reject the null hypothesis when in fact it is true.

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

Beta (1 - Power)

This option specifies one or more values for the probability of a type-II error (beta). A type-II error occurs when you fail to reject the null hypothesis of equal correlations when in fact they are different.

Values must be between zero and one. Historically, the value of 0.20 was often used for beta. However, you should pick a value for beta that represents the risk of a type-II error you are willing to take.

Power is defined as one minus beta. Power is equal to the probability of rejecting a false null hypothesis. Hence, specifying the beta error level also specifies the power level. For example, if you specify beta values of 0.05, 0.10, and 0.20, you are specifying the corresponding power values of 0.95, 0.90, and 0.80.

Example 1 - Testing the addition or deletion of a single variable

This example calculates the power of an *F* test constructed to test a fifth variable which adds 0.05 to R^2 after considering four other variables whose combined R^2 value is 0.50. Sample sizes from 10 to 150 will be investigated. The significance level is 0.05.

Setup

This section presents the values of each of the parameters needed to run this example. You can make these changes directly on your screen or you can load the template entitled Example1 by clicking the Template tab and loading this template.

Option	Value
Data Tab	
Find	**Beta and Power**
C: Number	**4**
C: R2(C)	**0.50**
T: Number	**1**
T: R2(T|C)	**0.05**
N	**10 to 150 by 20**
Alpha	**0.05**
Beta	*Ignored since this is the Find setting.*

Annotated Output

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

Numeric Results

Numeric Results

				Ind. Variables		Ind. Variables	
					Tested		Controlled
Power	**N**	**Alpha**	**Beta**	**Cnt**	**R2**	**Cnt**	**R2**
0.13037	10	0.05000	0.86963	1	0.05000	4	0.50000
0.41799	30	0.05000	0.58201	1	0.05000	4	0.50000
0.63511	50	0.05000	0.36489	1	0.05000	4	0.50000
0.78431	70	0.05000	0.21569	1	0.05000	4	0.50000
0.87818	90	0.05000	0.12182	1	0.05000	4	0.50000
0.93366	110	0.05000	0.06634	1	0.05000	4	0.50000
0.96493	130	0.05000	0.03507	1	0.05000	4	0.50000
0.98192	150	0.05000	0.01808	1	0.05000	4	0.50000

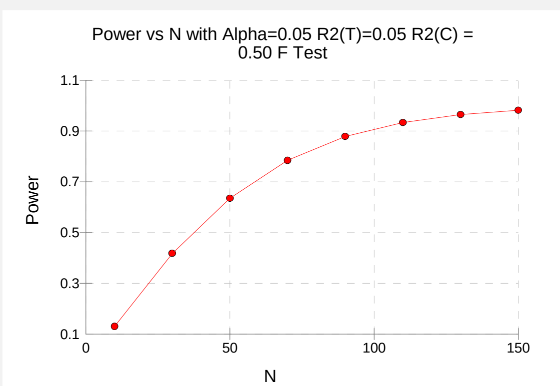
Report Definitions
Power is the probability of rejecting a false null hypothesis.
N is the number of observations on which the multiple regression is computed.
Alpha is the probability of rejecting a true null hypothesis. It should be small.
Beta is the probability of accepting a false null hypothesis. It should be small.
Cnt refers to the number of independent variables in that category.
R2 is the amount that is added to the overall R-Squared value by these variables.
Ind. Variables Tested are those variables whose regression coefficients are tested against zero.
Ind. Variables Controlled are those variables whose influence is removed from experimental error.

Summary Statements
A sample size of 10 achieves 13% power to detect an R-Squared of 0.05000 attributed to 1 independent variable(s) using an F-Test with a significance level (alpha) of 0.05000. The variables tested are adjusted for an additional 4 independent variable(s) with an R-Squared of 0.50000.

This report shows the values of each of the parameters, one scenario per row. The definitions of each of the columns is given in the Report Definitions section.

Note that in this particular example, a reasonable power of 0.90 is not reached until the sample size is 110.

Plots Section



This plot shows the relationship between sample size and power.

Example 2 - Minimum Detectable R-Squared

Suppose the researcher in Example1 can only afford a sample size of 30. He wants to know the minimum detectable R^2 that can be detected if the power is 80% and 90%.

Setup

This section presents the values of each of the parameters needed to run this example. You can make these changes directly on your screen or you can load and run the Example2 template.

Option	Value
Data Tab	
Find	**R-Squared of IV's Tested**
C: Number	**4**
C: R2(C)	**0.50**
T: Number	**1**
T: R2(T|C)	*Ignored since this is the Find setting*
N	**30**
Alpha	**0.05**
Beta	**0.10 0.20**

Annotated Output

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

Numeric Results

Numeric Results

				Ind. Variables	Tested	Ind. Variables	Controlled
Power	**N**	**Alpha**	**Beta**	**Cnt**	**R2**	**Cnt**	**R2**
0.90000	30	0.05000	0.10000	1	0.13787	4	0.50000
0.80000	30	0.05000	0.20000	1	0.11066	4	0.50000

This report shows that at 90% power, a sample size of 30 cannot detect an R^2 less than 0.13787.

Example 3 - Sample Size, Many X's

Suppose you have 25 observations on 19 independent variables. You run a regression analysis with all 19 variables and obtain an R^2 value of 0.90. You find that 4 of these variables will result in an R^2 of 0.60, leaving an R^2 increment of 0.30 for the other 15. You decide to test the significance of these 15 variables. The *F* ratio is

$$\begin{aligned} F_{15,5} &= \frac{(0.90-0.60)/15}{(1.0-0.90)/5} \\ &= 1.0 \end{aligned}$$

Obviously, an *F* of 1.0 is nonsignificant, so you might be tempted to discard these variables and proceed with the four that are significant. Before doing so, you decided to do a power analysis on this *F* test.

Setup

This section presents the values of each of the parameters needed to run this example. You can make these changes directly on your screen or you can load the template entitled Example3 by clicking the Template tab and loading this template.

Option	Value
Data Tab	
Find	**Beta and Power**
C: Number	**4**
C: R2(C)	**0.60**
T: Number	**15**
T: R2(T|C)	**0.3**
N	**25**
Alpha	**0.05**
Beta	*Ignored since this is the Find setting.*

Annotated Output

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

Numeric Results

Numeric Results

				Ind. Variables Tested		Ind. Variables Controlled	
Power	**N**	**Alpha**	**Beta**	**Cnt**	**R2**	**Cnt**	**R2**
0.72022	25	0.05000	0.27978	15	0.30000	4	0.60000

The power is only 0.72022, which is a little low. Hence, you cannot be certain whether these 15 variables are unimportant or if you just do not have a large enough sample.

Example 4 - Validation

Ralph O'Brien, in a private communication to Jerry Hintze, gave the result that when alpha = 0.05, N = 15, K = 2, and R-Squared = 0.6, the power is 0.96829.

Setup

This section presents the values of each of the parameters needed to run this example. You can make these changes directly on your screen or you can load the template entitled Example4 by clicking the Template tab and loading this template.

Option	Value
Data Tab	
Find	**Beta and Power**
C: Number	**0**
C: R2(C)	**0.0**
T: Number	**2**
T: R2(T|C)	**0.6**
N	**15**
Alpha	**0.05**
Beta	*Ignored since this is the Find setting.*

Annotated Output

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

Numeric Results

Numeric Results

				Ind. Variables Tested		Ind. Variables Controlled	
Power	**N**	**Alpha**	**Beta**	**Cnt**	**R2**	**Cnt**	**R2**
0.96829	15	0.05000	0.03171	2	0.60000	6	0.00000

The power of 0.96829 matches O'Brien's result.

Chapter 870

Poisson Regression

Poisson regression is used when the dependent variable is a count. Following the results of Signorini (1991), this procedure calculates power and sample size for testing the hypothesis that $\beta_1 = 0$ versus the alternative that $\beta_1 = B$. Note that e^{β_1} is the change in the rate for a one-unit change in X_1 when the rest of the covariates are held constant. The procedure assumes that this hypothesis will be tested using the score statistic

$$z = \frac{\hat{\beta}_1}{\sqrt{Var\left(\hat{\beta}_1\right)}}$$

The Poisson Distribution

The Poisson distribution models the probability of *y* events (i.e. failure, death, or existence) with the formula

$$\Pr\left(Y = y|\mu\right) = \frac{e^{-\mu}\mu^y}{y!} \quad \left(y = 0,1,2,...\right)$$

Notice that the Poisson distribution is specified with a single parameter μ. This is the mean incidence rate of a rare event per unit of *exposure*. Exposure may be time, space, distance, area, volume, or population size. Because exposure is often a period of time, we use the symbol *t* to represent the exposure. When no exposure value is given, it is assumed to be one.

The parameter μ may be interpreted as the risk of a new occurrence of the event during a specified exposure period, *t*. The probability of *y* events is then given by

$$\Pr\left(Y = y|\mu,t\right) = \frac{e^{-\mu t}\left(\mu t\right)^y}{y!} \quad \left(y = 0,1,2,...\right)$$

The Poisson distribution has the property that its mean and variance are equal.

The Poisson Regression Model

In Poisson regression, we suppose that the Poisson incidence rate μ is determined by a set of k regressor variables (the *X's*). The expression relating these quantities is

$$\mu = \exp\left(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k\right)$$

The regression coefficients $\beta_0, \beta_1, \beta_2, \cdots, \beta_k$ are unknown parameters that are estimated from a set of data. Their estimates are labeled $b_0, b_1, \cdots, b_k$.

Using this notation, the fundamental Poisson regression model for an observation i is written as

$$\Pr\left(Y_i = y_i | \mu_i, t_i\right) = \frac{e^{-\mu_i t_i}\left(\mu_i t_i\right)^{y_i}}{y_i!}$$

where

$$\mu_i = \lambda\left(\mathbf{x}_i'\boldsymbol{\beta}\right)$$
$$\lambda\left(\mathbf{x}_i'\boldsymbol{\beta}\right) = \exp\left(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki}\right)$$

That is, for a given set of values of the regressor variables, the outcome follows the Poisson distribution.

Power Calculations

Suppose you want to test the null hypothesis that $\beta_1 = 0$ versus the alternative that $\beta_1 = B1$. Signorini (1991) gives the formula relating sample size, α , β, and B1 when X_1 is the only covariate of interest as

$$N = \phi \frac{\left(z_{1-\alpha/2}\sqrt{V\left(b_1|\beta_1 = 0\right)} + z_{1-\beta}\sqrt{V\left(b_1|\beta_1 = B1\right)}\right)^2}{\mu_T e^{\beta_0} B1^2}$$

where N is the sample size, ϕ is a measure of over-dispersion, μ_T is the mean exposure time, and z is the standard normal deviate. Following the extension used in Hsieh, Block, and Larsen (1998) and Hsieh and Lavori (2000), when there are other covariates, ***PASS*** uses the approximation

$$N = \phi \frac{\left(z_{1-\alpha/2}\sqrt{V\left(b_1|\beta_1 = 0\right)} + z_{1-\beta}\sqrt{V\left(b_1|\beta_1 = B1\right)}\right)^2}{\mu_T e^{\beta_0} B1^2\left(1 - R^2\right)}$$

where R^2 is the square of the multiple correlation coefficient when the covariate of interest is regressed on the other covariates. The variance in the null case is given by

$$V\left(b_1|\beta_1 = 0\right) = \frac{1}{Var\left(X1\right)}$$

The variance for the non-null case depends on the underlying distribution of X. Common choices are given next.

Normal

$$V\left(b_1|\beta_1 = B1\right) = e^{-\left(B1\mu_{X1} + \frac{B1^2\sigma_{X1}^2}{2}\right)}$$

$$V\left(X1\right) = \sigma_{X1}^2$$

Exponential

$$V\left(b_1|\beta_1 = B1\right) = \frac{\left(\lambda_{X1} - B\right)^3}{\lambda_{X1}}$$

$$V\left(X1\right) = \lambda_{X1}^{-2}$$

Uniform, on the interval [C,D]

$$V\left(b_1|\beta_1 = B1\right) = \frac{m}{m\left(m_{11}\right) - m_1^2}$$

$$V\left(X1\right) = \frac{\left(D-C\right)^2}{12}$$

where

$$m = \frac{e^{B1D} - e^{B1C}}{\left(D-C\right)B1}$$

$$m_1 = \frac{e^{B1D}\left(B1D-1\right) - e^{B1C}\left(B1C-1\right)}{\left(D-C\right)B1^2}$$

$$m_{11} = \frac{e^{B1D}\left(2-2B1D+B1^2D^2\right) - e^{B1C}\left(2-2B1C+B1^2C^2\right)}{\left(D-C\right)B1^3}$$

Binomial, Parameter π_{X1}

$$V\left(b_1|\beta_1 = B1\right) = \frac{1}{1-\pi_{X1}} + \frac{1}{\pi_{X1}e^{B1}}$$

$$V\left(X1\right) = \pi_{X1}\left(1-\pi_{X1}\right)$$

Procedure Options

This section describes the options that are unique to this procedure. These are located on the panels associated with the Data, Options, and Reports tabs. To find out more about using the other tabs such as Plot Text, Axes, and Template, turn to the chapter entitled Procedure Templates.

Data Tab

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

Find

This option specifies the parameter to be solved for from the other parameters. Under most situations, you will select either *Beta* for a power analysis or *N* for sample size determination.

Select *N* when you want to calculate the sample size needed to achieve a given power and alpha level. Select *Beta* when you want to calculate the power of an experiment.

Exp(B1)/Exp(B0) (Rate Ratio)

B1 is the value of the regression coefficient under the alternative hypothesis. In Poisson regression, it is more natural to specify a value for exp(*B1*)/exp(*B0*) than for *B1* because exp(*B1*)/exp(*B0*) represents the ratio of the response rate when *X1* is increased one unit and all other covariates are constant to its baseline rate. Depending on the dataset being analyzed, the response rate might be the hazard rate, death rate, survival rate, or accident rate.

For example, suppose the baseline response rate (the value of $\exp(\beta_0)$ is 0.70 and you want the study to be large enough to detect a 30% increase in the response rate. The value entered here would be 1.3.

Alpha (Significance Level)

This option specifies one or more values for the probability of a type-I error (alpha). A type-I error occurs when you reject the null hypothesis of equal probabilities when in fact they are equal.

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

Beta (1 - Power)

This option specifies one or more values for the probability of a type-II error (beta). A type-II error occurs when you fail to reject the null hypothesis of equal probabilities of the event of interest when in fact they are different.

Values must be between zero and one. Historically, the value of 0.20 was used for beta. Now, 0.10 is more popular. You should pick a value for beta that represents the risk of a type-II error you are willing to take.

Power is defined as one minus beta. Power is equal to the probability of rejecting a false null hypothesis. Hence, specifying the beta error level also specifies the power level. For example, if you specify beta values of 0.05, 0.10, and 0.20, you are specifying the corresponding power values of 0.95, 0.90, and 0.80.

R-Squared Other X's

This is the *R*-Squared that is obtained when X_1 is regressed on the other *X*'s (covariates) in the model. Use this to account for the influence on power and sample size of adding other covariates. Note that the number of additional variables does not matter in this formulation. Only their overall relationship with X_1 through this R-Squared value is used.

Of course, this value is restricted to being greater than or equal to zero and less than one. Use zero when there are no other covariates.

Phi (Over-Dispersion Parameter)

Phi is the over-dispersion parameter for Poisson regression. When there is no over-dispersion, set this value to one.

Exp(B0) (Baseline Rate)

Exp(B0) is the baseline response rate. This is the response rate that occurs when all covariates are equal to zero. Depending on the dataset being analyzed, the response rate might be the hazard rate, death rate, survival rate, or accident rate.

Mu T (E Time)

This is the mean exposure time, μ_T, over which the response rate is calculated. If the response rates are for one year, enter 1 here. If they are for 30 days, enter 30 here.

Hypothesis Test

Specify whether the test is one-sided or two-sided. When a two-sided hypothesis is selected, the value of alpha is halved by ***PASS***. Everything else remains the same.

Note that the accepted procedure is to use the Two-Sided option unless you can justify using a one-sided test.

N (Sample Size)

This option specifies the total number of observations in the sample. You may enter a single value or a list of values.

Note that when the Overall Event Rate is set to 1.0, the sample size becomes the number of events.

X Distribution

This option specifies the distribution of the covariate being analyzed. You can choose from the normal, exponential, uniform, and binomial distributions. Once you have selected the anticipated distribution of the covariate, you must also specify the distribution's parameter(s).

Normal(M)

This option is used when the X Distribution is set to Normal(M,S) to specify the mean of the covariate distribution.

Normal(S)

This option is used when the X Distribution is set to Normal(M,S) to specify the standard deviation of the covariate distribution.

Uniform(C)

This option is used when the X Distribution is set to Uniform(C,D) to specify the minimum of the covariate distribution. That is, the distribution of the covariate is assumed to range from *C* to *D*.

Uniform(D)

This option is used when the X Distribution is set to Uniform(C,D) to specify the maximum of the covariate distribution. That is, the distribution of the covariate is assumed to range from *C* to *D*.

Binomial(P)

This option is used when the X Distribution is set to Binomial(P) to specify the proportion associated with the covariate. For example, if the covariate being studied is an indicator variable of treatment and control and you anticipated having an equal number of treatments and controls, this value should be set to 0.50.

Exponential(L)

This option is used when the X Distribution is set to Exponential(L) to specify the *lambda* associated with the covariate. Note that the pdf of the exponential distribution is assumed to be (*lambda*) exp(-*lambda X*).

Example1 - Power for Several Sample Sizes

Poisson regression will be used to analyze the power for a study of the relationship between the number of flaws on a manufactured article and the experience (measured in years) of the operator. The researchers want to evaluate the sample size needs for detecting ratios in response rates of 1.3 and 1.5. The experience of an operator is assumed to be normally distributed with mean 3.2 and standard deviation 2.1. No other covariates will be included in the analysis. The researchers will test their hypothesis using a 5% significance level with a two-sided Wald test. They decide to calculate the power at sample sizes between 5 and 50.

Setup

You can enter these values yourself or load the Example1 template from the Template tab.

Option	Value
Data Tab	
Find	**Beta and Power**
Exp(B1)	**1.3 1.5**
Alpha	**0.05**
Beta	*Ignored since this is the Find setting*
R2 Other X's	**0.0**
Phi	**1.0**
Exp(B0)	**1.0**
Mu T	**1.0**
Hypothesis Test	**Two-Sided**
N	**5 to 50 by 5**
X Distribution	**Normal(M,S)**
Normal(M)	**3.2**
Normal(S)	**2.1**

Annotated Output

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

Numeric Results

Numeric Results when X1 is Normal with Mean = 3.2 and Sigma = 2.1
And Phi (Over-Dispersion Parameter) = 1.0000

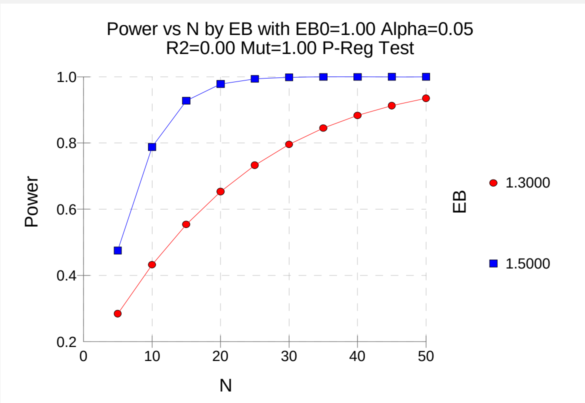
Power	Sample Size (N)	Response Rate Ratio	Baseline Rate Exp(B0)	Mean Exposure Time (MuT)	R-Squared X1 vs Other X's (R2)	Two-Sided Alpha	Beta
0.28466	5	1.3000	1.0000	1.0000	0.0000	0.05000	0.71534
0.43245	10	1.3000	1.0000	1.0000	0.0000	0.05000	0.56755
0.55407	15	1.3000	1.0000	1.0000	0.0000	0.05000	0.44593
0.65321	20	1.3000	1.0000	1.0000	0.0000	0.05000	0.34679
0.73282	25	1.3000	1.0000	1.0000	0.0000	0.05000	0.26718
0.79585	30	1.3000	1.0000	1.0000	0.0000	0.05000	0.20415
0.84516	35	1.3000	1.0000	1.0000	0.0000	0.05000	0.15484
0.88334	40	1.3000	1.0000	1.0000	0.0000	0.05000	0.11666
0.91262	45	1.3000	1.0000	1.0000	0.0000	0.05000	0.08738
0.93491	50	1.3000	1.0000	1.0000	0.0000	0.05000	0.06509
0.47562	5	1.5000	1.0000	1.0000	0.0000	0.05000	0.52438
0.78817	10	1.5000	1.0000	1.0000	0.0000	0.05000	0.21183
0.92798	15	1.5000	1.0000	1.0000	0.0000	0.05000	0.07202
0.97821	20	1.5000	1.0000	1.0000	0.0000	0.05000	0.02179
0.99394	25	1.5000	1.0000	1.0000	0.0000	0.05000	0.00606
0.99842	30	1.5000	1.0000	1.0000	0.0000	0.05000	0.00158
0.99961	35	1.5000	1.0000	1.0000	0.0000	0.05000	0.00039
0.99991	40	1.5000	1.0000	1.0000	0.0000	0.05000	0.00009
0.99998	45	1.5000	1.0000	1.0000	0.0000	0.05000	0.00002
1.00000	50	1.5000	1.0000	1.0000	0.0000	0.05000	0.00000

Report Definitions
Power is the probability of rejecting a false null hypothesis. It should be close to one.
N is the size of the sample drawn from the population.
Exp(B1)/Exp(B0) is the response rate ratio due to a one-unit change in X1.
Exp(B0) is the response rate when all covariates have a value of zero.
Phi is the over-dispersion parameter used when the Poisson model does not fit.
R2 is the R-squared achieved when X1 is regressed on the other covariates.
Alpha is the probability of rejecting Exp(B1)/Exp(B0) is one.
Beta is the probability of accepting a false null hypothesis.

Summary Statements
A Poisson regression of a dependent variable of counts on normally distributed independent variable with mean = 3.2 and standard deviation = 2.1 using a sample of 5 observations achieves 28% power at a 0.71534 significance level to detect a response rate ratio of at least 1.3000 due a a one-unit change in the IV. The baseline rate is 1.0000 and the mean exposure time is 1.0000.

This report shows the power for each of the scenarios. Note that if you were interested in $B1$ instead of Exp($B1$), you would simply take the natural logarithm of the value of Exp($B1$).

Plot Section



Example2 - Validation Using Signorini

Signorini (1991), page 449, presents an example which we will use to validate this program. In the example, Exp(*B*1)/Exp(B0)= 1.3, Exp(*B*0)= 0.85, *R*2 = 0.0, *Mu T* = 1.0, and *Phi* = 1.0. The independent variable is assumed to be binomial with proportion 0.5. A one-sided test with alpha = 0.05 will be used. Sample sizes for power = 0.80, 0.90, and 0.95 are calculated to be 406, 555, and 697.

Setup

You can enter these values yourself or load the Example2 template from the Template tab.

Option	Value
Data Tab	
Find	**N**
Exp(B1)	**1.3**
Alpha	**0.05**
Beta	**0.05 0.10 0.20**
R2 Other X's	**0.0**
Phi	**1.0**
Exp(B0)	**0.85**
Mu T	**1.0**
Hypothesis Test	**One-Sided**
N	***Ignored since this is what is being solved for.***
X Distribution	**Binomial(P)**
Binomial(P)	**0.5**

Annotated Output

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

Numeric Results

Numeric Results when X1 is Binomial with Proportion = 0.5
And Phi (Over-Dispersion Parameter) = 1.0000

Power	Sample Size (N)	Response Rate Ratio	Baseline Rate Exp(B0)	Mean Exposure Time (MuT)	R-Squared X1 vs Other X's (R2)	Two-Sided Alpha	Beta
0.95012	697	1.3000	0.8500	1.0000	0.0000	0.05000	0.04988
0.90030	556	1.3000	0.8500	1.0000	0.0000	0.05000	0.09970
0.80016	406	1.3000	0.8500	1.0000	0.0000	0.05000	0.19984

Note that ***PASS*** calculated 556 rather than the 555 calculated by Signorini (1991). The discrepancy is due to rounding.

Chapter 900

Chi-Square Effect Size Estimator

Introduction

This procedure calculates the effect size of the Chi-square test for use in power and sample size calculations. Based on your input, the procedure provides effect size estimates for Chi-square goodness-of-fit tests and for Chi-square tests of independence.

The *Chi-square test* is often used to test whether sets of frequencies or proportions follow certain patterns. The two most common cases are in tests of goodness of fit and tests of independence in contingency tables.

The *Chi-square goodness-of-fit* test is used to test whether a set of data follows a particular distribution. For example, you might want to test whether a set of data comes from the normal distribution.

The *Chi-square test for independence* in a contingency table is another common application of this test. Here individuals (people, animals, or things) are classified by two (nominal or ordinal) classification variables into a two-way contingency table. This table contains the counts of the number of individuals in each combination of the row categories and column categories. The Chi-square test determines if there is dependence (association) between the two classification variables.

Effect Size

For each cell of a table containing m cells, there are two proportions considered: one specified by a null hypothesis and the other specified by the alternative hypothesis. Usually, the proportions specified by the alternative hypothesis are those occurring in the data. Define p_{0i} to be the proportion in cell i given by the null hypothesis and p_{1i} to be the proportion in cell i according to the alternative hypothesis. The effect size, W, is calculated using the following formula

$$W = \sqrt{\sum_{i=1}^{m} \frac{(p_{0i} - p_{1i})^2}{p_{0i}}}.$$

The formula for computing the Chi-square value, χ^2, is

$$\chi^2 = \sum_{i=1}^{m} \frac{(O_i - E_i)^2}{E_i}$$

$$= N \sum_{i=1}^{m} \frac{(p_{0i} - p_{1i})^2}{p_{0i}},$$

where N is the total count in all the cells. Hence, the relationship between W and χ^2 is

$$\chi^2 = NW^2$$

or

$$W = \sqrt{\frac{\chi^2}{N}}$$

Contingency Table Tab

This window allows you to enter up to an eight-by-eight contingency table. You can enter percentages or counts. If you enter counts, the Chi-Square and Prob Level values are correct and may be used to test the independence of the row and column variables. If you enter percentages, you should ignore the Chi-Square and Prob Level values.

Note that if you are entering percentages, it does not matter whether you enter table percentages or row (or column) percentages as long as you are consistent.

Example

Suppose you are planning a survey with the primary purpose of testing whether marital status is related to gender. You decide to adopt four marital status categories: never married, married, divorced, widowed. In the population you are studying, previous studies have found the following percentages in each of these categories:

Never Married	27%
Married	39%
Divorced	23%
Widowed	11%

You decide that you want to calculate the effect size when the individual percentages for males and females are

Gender	Male	Female
Never Married	22%	32%
Married	46%	33%
Divorced	22%	24%
Widowed	10%	11%

To complete this example, you would load the Chi-Square Effect Size Estimator procedure from the PASS-Other menu and enter “22 46 22 10” across the top row and “32 33 24 11” across the next row. The value of *W* turns out to be 0.143626.

Note that even though a Chi-square value (4.13) and probability level (0.248) are displayed, you would ignore them since you have entered percentages, not counts, into the table. If you had entered counts, these results could be used to test the hypothesis of independence.

Multinomial Table Tab

This window allows you to enter a multinomial table with up to fourteen cells. You can enter percentages or counts. If you enter counts, the Chi-Square and Prob Level values are correct and may be used to test the statistical significance of the table. If you enter percentages, you should ignore the Chi-Square and Prob Level values.

Note that if you are using the window to perform a goodness-of-fit test on a set of data, you will need to adjust the degrees of freedom for the number of parameters you are estimating. For example, if you are testing whether the data are distributed normally and you estimate the mean and standard deviation from the data, you will need to reduce the degrees of freedom by two.

Example

Suppose you are going to use the Chi-square goodness-of-fit statistic calculated from a multinomial table to test whether a set of exponential data follow the normal distribution. That is, you want to find a reasonable effect size for comparing exponentially distributed data to the normal distribution.

You decide to divide the data into five groups: 5 or less, 5-10, 10-15, 15-20, 20+

Using tables for the normal and exponential distributions, you find that the probabilities for each group are

Category	Normal	Exponential
5 or Less	11%	39%
5 to 10	20%	26%
10 to 15	38%	18%
15 to 20	20%	11%
Above 20	11%	6%

To complete this example, you would set the Chi-Square Effect Size Estimator procedure to the Multinomial Test tab and enter “11 20 38 20 11” down the first column and “39 26 18 11 6” down the second column. The calculated value of *W* is 0.948271. You would enter this value into the Effect Size option of the Chi-Square Test window in ***PASS*** to determine the necessary sample size.

Chapter 905

Standard Deviation Estimator

Introduction

Even though it is not of primary interest, an estimate of the *standard deviation* (*SD*) is needed when calculating the power or sample size of an experiment involving one or more means. Finding such an estimate is difficult not only because the estimate is required before the data are available, but also because the interpretation of the standard deviation is vague and our experience with it is low. How do you estimate a quantity without data and without a clear understand of what the quantity is? This section will acquaint you with the standard deviation and offer several ways to obtain a rough estimate of it before the experiment begins using the Standard Deviation Estimator.

The Standard Deviation Estimator can also be used to calculate the standard deviation of the means, a quantity used in estimating sample sizes in analysis of variance designs.

Understanding the Standard Deviation

How well do you understand the standard deviation from the formula below? How would you describe it to another person? That's a tough task. Everyone understands an average, but what about a standard deviation?

In the paragraphs that follow, we will explain that the standard deviation has two general interpretations. **First, it may be thought of as the average difference between an observation and the mean, ignoring the sign. Second, that it may be thought of as the average difference between any two data values, ignoring the sign.**

The population standard deviation is calculated using the formula:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}\left(X_i - \mu\right)^2}{N}}$$

where *N* is the number of items in the population, *X* is the variable being measured, and μ is the mean of *X*. This formula indicates that the standard deviation is the square root of an average. An average of what? An average of the squared differences between each value and the mean. The differences are squared to remove the sign so that negative values will not cancel out positive values. After summing up these squared differences and dividing by *N*, the square root is taken to put the result back in the original scale. Bottom line—the standard deviation can be thought of as the average difference between the data values and their mean. (We are using the terms *mean* and *average* here. Both refer to the same operation.)

An Example

Consider the following two sets of numbers

A: 1, 5, 9

B: 4, 5, 6

Both sets have the same mean of 5. However, their standard deviations are quite different. Subtracting the mean and squaring the three items in each set results in

Set A

(1-5)(1-4) = 16

(5-5)(5-5) = 0

(9-5)(9-5) = 16

Sum = 32

SD = SQRT(32/3) = 3.266

Set B

(4-5)(4-5) = 1

(5-5)(5-5) = 0

(6-5)(6-5) = 1

Sum = 2

SD = SQRT(2/3) = 0.8165

The standard deviations show that the data in set A vary more than the data in set B.

Divide by N or N-1?

Note that we are dividing by *N*, not *N-1* as you usually see. When the standard deviation is computed using all values in the population, *N* is used as the divisor. However, when the standard deviation is calculate from a sample, *N-1* is used as the divisor. In fact, we stress that the results are for a sample by using the lower-case *n* and naming the *sample standard deviation S*. The value of *S* is computed from a sample of *n* values using the formula

$$S = \sqrt{\frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n-1}}$$

The *n-1* is used instead of the *n* to correct for bias that statisticians have discovered. That is, over the long run, dividing by *n-1* provides a better estimate of *SD* than does dividing by *n*. However, dividing by *n-1* instead of *n* just confuses our understanding of the standard deviation even further. For the purposes of interpretation, assume that the divisor is *n* so that the operation can be thought of as computing an average.

Average Absolute Deviation

If you had to devise a measure of variability, many of you would employ the average absolute deviation (AD) which is computed by forming the deviations from the mean, taking their absolute values, and computing their average. The absolute value has been applied to get rid of the negative signs that will occur. The formula for AD is

$$AD = \frac{\sum_{i=1}^{N}\left|X_i - M\right|}{N}$$

Notice how appealing this measure is! No squaring. No square roots. Just the direct calculation of the average absolute deviation about the mean. This is a simple average so we can understand it. Why isn't this measure used? The answer is that even though the average absolute deviation is simple to understand and calculate, it is difficult to work with mathematically. Its distribution function is difficult to work with. The standard deviation is just the opposite. It is difficult to understand, but it can be worked with mathematically in statistical problems. Hence, we use the standard deviation, but don't understand it!

Comparing AD and SD

Fortunately, the average absolute deviation and the standard deviation are usually close in value. Mathematically, you can show that *AD* is always less than or equal to *SD*. A small simulation study is summarized below. It shows the relationship between *AD* and *SD* for data generated from various distributions.

Distribution	Percent SD > AD	Characteristics
Uniform	15%	Level
Normal	20%	Bell-Shaped
Gamma(5)	30%	Moderately Skewed Right
Gamma(5)^2	45%	Extremely Skewed Right

These distributions were selected for study because they represent a wide range of possibilities. The table shows that, for typical datasets, the standard deviation is from 15 to 30 percent larger than the average absolute deviation. And in the case of the normal distribution, the *SD* is about 20% higher than *AD*.

Hence, for planning purposes, you can think of the standard deviation as an inflated version of the average absolute deviation.

Example

In out example, we can compute *AD* for datasets A and B as follows.

Set A

AD = (4+0+4)/3 = 8/3 = 2.667. Note that SD = 3.266.

Set B

AD = (1+0+1)/3 = 2/3 = 0.667. Note that SD = 0.8165.

We see that the values are close, although not exactly the same. The degree of difference is certainly within the error that we would expect during the planning phase.

SD as the Average Difference Between Values

The above discussion and formula have pointed out that the standard deviation may be thought of as an average deviation from the mean. In this section, a second interpretation of the standard deviation will be given.

We can manipulate the formula for the sum of squared deviations to show that

$$\sum_{i=1}^{N}\left(X_i - M\right)^2 = \frac{\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\left(X_i - X_j\right)^2}{N}$$

This formula shows that the squared deviations from the mean are proportional to the squared deviations of each observation from every other observation. Note that the mean is not involved in the expression on the right. Think about that for a moment. What better measure of variability than one that compares each observation with every other observation?

Using the above relationship, the standard deviation may be calculated using the formula

$$SD = \sqrt{\frac{\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\left(X_i - X_j\right)^2 / N}{N}}$$

Example

Consider again our simple example of sets A and B. Applying this operation to the three possible pairs of the data in set A (1, 5, 9) gives

(1-5)(1-5) = 16

(1-9)(1-9) = 64

(5-9)(5-9) = 16

The sum is 96. Dividing 96 by 3 (the number of pairs) again yields 32. Hence, the standard deviation is computed as *SD* = SQRT(96/9) = 3.266 (which matches the previous result).

Likewise, for set B (4, 5, 6), the formula results in

(4-5)(4-5) = 1

(4-6)(4-6) = 4

(5-6)(5-6) = 1

so that SD = SQRT(6/9) = 0.8165.

Estimating the Standard Deviation

Our task is to find a rough estimate of the standard deviation. Several possible methods are available in the Standard Deviation Estimator procedure which may be loaded from the PASS-Other menu. ***PASS*** provides a panel that implements each of these methods for you.

Estimate SD From A Set of Data Values (Data Tab)

One method of estimating the standard deviation is to put in a typical set of values and calculate the standard deviation.

This window is also used when you need the standard deviation of a set of hypothesized means in an analysis of variance sample size study.

Pros and Cons of This Method

This method lets you experiment with several different data values. It lets you determine the influence of different data configurations on the standard deviation. In so doing, you can come up with a likely range of SD values.

However, investigators tend to pick trial numbers that are closer to the mean and more uniform than will result in practice. This results in SD's that are underestimated. So if you use this method, you must be careful that your range of possible SD values is wide enough to be accurate.

Example 1 - SD for a set of values

As an example, suppose that you decide that the following values represent a typical set of data that you would anticipate for one group of individuals:

10, 12, 14, 10, 11, 10, 12, 13, 9, 13, 15, 11.

To calculate the appropriate standard deviation, do the following.

1. Load the **Standard Deviation Estimator** window and click on the **Data** tab.
2. The order that the data are entered in does not matter. However, to show the use of the Counts column, we count up the number of times each value occurs. The values and their frequency counts are then entered into the **Values** and **Counts** columns. The data entry goes as follows:

9	**1**
10	**3**
11	**2**
12	**2**
13	**2**
14	**1**
15	**1**

3. Check the **Use N-1 as divisor** box. (We use the N-1 divisor when estimating sigma from a set of data.)
4. Press the **Calculate Statistics** button. The standard deviation is 1.825742. We probably would round this value up to 2.0 for planning purposes.

Example 2 - SD for a set of means

In this example, we will show you how to obtain the standard deviation of a set of hypothesized means. Care must be taken that you select the correct divisor—*N*, not *N*-1.

In this example, a researcher is studying the influence of a drug on heart rate. He estimates that the average heart rate of his group without the drug is 80. His experimental design will apply three different doses. The first dose is expected to lower the heart rate by 10%, the second by 20%, and the third by 30%. Hence, the hypothesized means for the four groups are 80, 80(0.9) = 72, 80(0.8) = 64, and 80(0.7) = 56.

To calculate the appropriate standard deviation, do the following.

1. Load the **Standard Deviation Estimator** window and click on the **Data** tab.
2. Enter the four means into the **Values** column. The Counts column is left blank.
   80
   72
   64
   56
3. Make sure the **Use N-1 as divisor** box is not checked since we want the population standard deviation.
4. Press the **Calculate Statistics** button. The standard deviation of the means is 8.944272.

Estimate SD From The Standard Error (Standard Error Tab)

If the value of the standard error of the mean is available from another experiment, it may be used to estimate the standard deviation. The formula estimating the standard deviation from the standard error is

$$SD = SE\sqrt{N}$$

where *N* is the sample size.

Pros and Cons of This Method

This method is only useful when you have a standard error value available.

Example

To calculate the appropriate standard deviation when a previous study of 23 individuals had a standard error of the mean of 2.7984, do the following.

1. Load the **Standard Deviation Estimator** window and click on the **Standard Error** tab.
2. Enter **23** for **N**.
3. Enter **2.7984** for **Standard Error**.
4. Press the **Calculate Standard Deviation** button. The standard deviation is 1.825742. We probably would round this value up to 2.0 for planning purposes.

Estimate SD From The Range (Range Tab)

There are two cases in which the range may be used to estimate the standard deviation. In the first case, the sample size and data range may be available from a previous study. In the second case, a reasonable estimate of the population range may be obtainable. This window allows you to estimate SD in both of these situations.

The basic formula for estimating the standard deviation from the range is

$$SD = \frac{Range}{C}$$

where *C* is determined by the situation.

Determining C

If the data range is available from a previous study, the constant *C* is the median of the distribution of the range for that sample range. This distribution assumes that the data themselves are normally distributed. The median of the distribution of the range is calculated in NCSS using numerical methods. You can find it in our Probability Calculator. The calculated value of C is shown to the right of the Standard Deviation box. To use this method, leave the Population Divisor box empty.

If the a population range can be established, it may be used to estimate sigma by dividing by an appropriate constant. To determine an appropriate value of the constant, statisticians use the fact that most of the data is contained within three standard deviations of the mean—so they set *C* to six. However, consultants have found that for one reason (understated range) or another (population not normal) dividing by six tends to understate the standard deviation. So they divide by five or even four. Dividing by a smaller number increases the estimated standard deviation. Our recommendation is to divide by four. To use this method, enter the divisor (4, 5, or 6) in the Population Divisor box.

Pros and Cons of This Method

The range is a poor substitute for having the standard deviation. This method should be used as a 'last resort.'

Example 1 - Previous Sample Available

To calculate an estimate of the standard deviation when a previous study of 20 animals had a minimum value of 15.3 and a maximum of 18.7, do the following.

1. Load the **Standard Deviation Estimator** window and click on the **Range** tab.
2. Enter **20** for **N**.
3. Enter **3.4** for **Range**. This is 18.7 minus 15.3.
4. Be sure that the **Population Divisor** box is blank.
5. Press the **Calculate Standard Deviation** button. The estimate of the standard deviation is 0.92243. The value of C used is 3.685916.

Example 2 - Population Range 'Known'

To calculate an estimate of the standard deviation when the population range is known to be 150, do the following.

1. Load the **Standard Deviation Estimator** window and click on the **Range** tab.
2. Enter **150** for **Range**.
3. Enter **4** for **Population Divisor**.
4. Press the **Calculate Standard Deviation** button. The estimate of the standard deviation is 37.5. The value of C used is 4.

Estimate SD From Two Percentiles (Percentiles Tab)

If you are willing to assume that the population values are normally distributed (bell-shaped), you can use the values of two percentiles to estimate the standard deviation.

The basic formula for estimating the standard deviation from two percentiles is

$$SD = \left| \frac{X2 - X1}{Z\left(\frac{P2}{100}\right) - Z\left(\frac{P1}{100}\right)} \right|$$

where *X1* and *X2* are the two percentiles, *P1* and *P2* are the two percentages, and *Z(P)* is the standard normal deviate that has a tail area of *P* to the left.

Pros and Cons of This Method

This method works well if you have two accurate percentiles available and the underlying distribution of the data is normal.

Example

To calculate an estimate of the standard deviation when you know that the 25th percentile of the population is 80.25, the 75th percentile of the population is 116.38, and that the population is normally distributed, do the following.

1. Load the **Standard Deviation Estimator** window and click on the **Percentiles** tab.
2. Enter **25** for **Percentage 1**.
3. Enter **80.25** for **Percentile Value 1**.
4. Enter **75** for **Percentage 2**.
5. Enter **116.38** for **Percentile Value 2**.
6. Press the **Calculate Standard Deviation** button. The estimate of the standard deviation is 26.78321.

Estimate SD From The Coefficient of Variation (COV Tab)

The coefficient of variation (*COV*) is equal to SD divided by the mean. Hence, if you know the coefficient of variation and the mean, you can estimate SD.

Note that t-tests and the analysis of variance assume that the standard deviations are equal for all groups. If the standard deviations are proportional to their group means, you should use a data transformation (such as the square root or the logarithm) to make the standard deviations equal.

The basic formula for estimating the standard deviation from the *COV* is

$$SD = (COV)(Mean)$$

Pros and Cons of This Method

This method works well if you have estimates of the mean and COV.

Example

To calculate an estimate of the standard deviation when you know that the mean is 127 and the COV is 0.832, do the following.

1. Load the **Standard Deviation Estimator** window and click on the **COV** tab.
2. Enter **0.832** for **COV**.
3. Enter **127** for **Mean**.
4. Press the **Calculate Standard Deviation** button. The estimate of the standard deviation is 105.664.

Confidence Interval for a Sample Standard Deviation (Confidence Limits Tab)

Often, you will obtain an estimate of the standard deviation from a previous study or a pilot study. Since this estimate is based on a sample, it is important to understand its precision. This can easily be calculated since the square of the sample standard deviation follows a chi-squared distribution. This confidence interval does assume that the population you are sampling from is normally distributed.

Once a confidence interval has been obtained, it would be wise to enter both values into the appropriate place in the sample size calculations to provide a range of possible sample size values (or of statistical power).

Example

Suppose a pilot study of 10 individuals yields a standard deviation of 83.21. Calculate a 95% confidence interval for the population standard deviation using these results.

1. Load the **Standard Deviation Estimator** window and click on the **Confidence Limits** tab.
2. Enter **10** for **N**.
3. Enter **83.21** for **Standard Deviation**.
4. Enter **0.05** for **Alpha**.
5. Press the **Calculate Confidence Limits** button. The confidence limits are 57.23477 and 151.909.

Chapter 910

Odds Ratio Estimator

Introduction

This procedure calculates any one of three parameters, odds ratio, p_1, or p_2, from the other two parameters. Note that p_1 and p_2 are the proportions in groups one and two, respectively. This provides you with a tool to study the relationship between this three parameters. This procedure is most often used when planning the sample size for a test involving two proportions. The procedure may be loaded by selecting *Odds Ratio Estimator* from the *PASS-Other* menu.

When planning studies involving two proportions, two parameters, p_1 and p_2, need to be specified. At times, it may be difficult to propose a value for p_2. In these cases, it might be easier to propose a value for the *odds ratio* (*OR*).

The odds of obtaining the response of interest in group 1 are $p_1 / (1 - p_1)$ and the odds of obtaining the response in group 2 are $p_2 / (1 - p_2)$. The ratio of these odds, called the odds ratio, is defined as

$$
\begin{aligned}
OR &= \frac{p_2 / (1 - p_2)}{p_1 / (1 - p_1)} \\
&= \frac{p_2(1 - p_1)}{p_1(1 - p_2)}
\end{aligned}
$$

To understand better how to interpret an odds ratio, consider the following example. Suppose the proportion dying from a particular disease during the first five years is 80%. The odds of dying are thus 0.8 / 0. 2 = 4.0 Suppose a treatment reduces the death rate from 80% to 60%. The odds of dying are now 0.6 / 0.4 = 1.5. The odds ratio is 4.0 / 1.5 = 2.7. That is, the odds of dying have been reduced by a factor of 2.67.

The odds ratio is reversible. In this example, suppose we talk in terms of surviving instead of dying. The odds of the two groups are now 0.2 / 0.8 = 0.25 and 0.4 / 0.6 = 0.67. The odds ratio is 0.67 / 0.25 = 2.67. The odds of surviving have increased by a factor of 2.67.

In some situations, it may be easier to define a meaning treatment effect in terms of the odds ratio. That is, it might be meaningful to say that a certain treatment increases the odds of survival by 50% (*OR* = 1.5) or by 200% (*OR* = 2.0). If this is the case, a value for p_2 can be calculated from p1 and *OR* by solving the above equation for p_2 to find that

$$
p_2 = \frac{p_1(OR)}{1 - p_1 + p_1(OR)}
$$

Hence, given a value for p_1 and *OR*, you can calculate an appropriate value for p_2 .

Odds Ratio Tab

This window lets you enter values for two of the three parameters used to form an odds ratio and calculate the remaining parameter.

Example 1 - Solving for P1

Suppose you know that $p_2 = 0.8$ and that *OR* = 4 and you want to find the corresponding value of p_1 .

1. Load the Odds Ratio Estimator procedure by selecting it from the *PASS-Other* menu.
2. Set P2 equal to *0.8*.
3. Set Odds Ratio equal to *4*.
4. Press the *Calculate P1* button.
5. Read the result in the P1 box. The result is 0.5

Example 2 - Solving for P2

Suppose you know that $p_1 = 0.4$ and that *OR* = 1.5 and you want to find the corresponding value of p_2 .

1. Load the Odds Ratio Estimator procedure by selecting it from the *PASS-Other* menu.
2. Set P1 equal to *0.4*.
3. Set Odds Ratio equal to *1.5*.
4. Press the *Calculate P2* button.
5. Read the result in the P2 box. The result is 0.5.

Example 3 - Solving for Odds Ratio

Suppose you know that $p_1 = 0.4$ and that $p_2 = 0.8$ and you want to find the corresponding value of the odds ratio.

1. Load the Odds Ratio Estimator procedure by selecting it from the *PASS-Other* menu.
2. Set P1 equal to *0.4*.
3. Set P2 equal to *0.8*.
4. Press the *Calculate Odds Ratio* button.
5. Read the result in the Odds Ratio box. The result is 6.

References

Agresti, A. and Coull, B. 1998. "Approximate is Better than 'Exact' for Interval Estimation of Binomial Proportions," *American Statistician*, Volume 52 Number 2, pages 119-126.

A'Hern, R. P. A. 2001. "Sample size tables for exact single-stage phase II designs." *Statistics in Medicine*, Volume 20, pages 859-866.

AIAG (Automotive Industry Action Group). 1995. *Measurement Systems Analysis*. This booklet was developed by Chrysler/Ford/GM Supplier Quality Requirements Task Force. It gives a detailed discussion of how to design and analyze an R&R study. The book may be obtained from ASQC or directly from AIAG by calling 801-358-3570.

Albert, A. and Harris, E. 1987. *Multivariate Interpretation of Clinical Laboratory Data*. Marcel Dekker, New York, New York. This book is devoted to a discussion of how to apply multinomial logistic regression to medical diagnosis. It contains the algorithm that is the basis of our multinomial logistic regression routine.

Allen, D. and Cady, F.. 1982. *Analyzing Experimental Data by Regression*. Wadsworth. Belmont, Calif. This book works completely through several examples. It is very useful to those who want to see complete analyses of complex data.

Al-Sunduqchi, Mahdi S. 1990. *Determining the Appropriate Sample Size for Inferences Based on the Wilcoxon Statistics*. Ph.D. dissertation under the direction of William C. Guenther, Dept. of Statistics, University of Wyoming, Laramie, Wyoming.

Altman, Douglas. 1991. *Practical Statistics for Medical Research*. Chapman & Hall. New York, NY. This book provides an introductory discussion of many statistical techniques that are used in medical research. It is the only book we found that discussed ROC curves.

Andersen, P.K., Borgan, O., Gill, R.D., and Keiding, N. 1997. *Statistical Models Based on Counting Processess*. Springer-Verlag, New York. This is an advanced book giving many of the theoretically developments of survival analysis.

Anderson, R.L. and Hauck, W.W. 1983. "A new Procedure for testing equivalence in comparative bioavailability and other clinical trials." *Commun. Stat. Theory Methods.*, Volume 12, pages 2663-2692.

Anderson, T.W. and Darling, D.A. 1954. "A test of goodness-of-fit." *J. Amer. Statist. Assoc*, Volume 49, pages 765-769.

Andrews, D.F., and Herzberg, A.M. 1985. *Data*. Springer-Verlag, New York. This book is a collection of many different data sets. It gives a complete description of each.

Armitage. 1955. "Tests for linear trends in proportions and frequencies." *Biometrics*, Volume 11, pages 375-386.

Armitage, P., and Colton, T. 1998. *Encyclopedia of Biostatistics*. John Wiley, New York.

Armitage,P., McPherson, C.K., and Rowe, B.C. 1969. "Repeated significance tests on accumulating data." *Journal of the Royal Statistical Society, Series A*, 132, pages 235-244.

Atkinson, A.C. 1985. *Plots, Transformations, and Regression*. Oxford University Press, Oxford (also in New York). This book goes into the details of regression diagnostics and plotting. It puts together much of the recent work in this area.

Atkinson, A.C., and Donev, A.N. 1992. *Optimum Experimental Designs*. Oxford University Press, Oxford. This book discusses D-Optimal designs.

Bain, L.J. and Engelhardt, M. 1991. *Statistical Analysis of Reliability and Life-Testing Models*. Marcel Dekker. New York. This book contains details for testing data that follow the exponential and Weibull distributions.

Baker, Frank. 1992. *Item Response Theory*. Marcel Dekker. New York. This book contains a current overview of IRT. It goes through the details, providing both formulas and computer code. It is not light reading, but it will provide you with much of what you need if you are attempting to use this technique.

Barnard, G.A. 1947. "Significance tests for 2 x 2 tables." *Biometrika* 34:123-138.

Barrentine, Larry B. 1991. *Concepts for R&R Studies*. ASQC Press. Milwaukee, Wisconsin. This is a very good applied work book on the subject of repeatability and reproducibility studies. The ISBN is 0-87389-108-2. ASQC Press may be contacted at 800-248-1946.

Bartholomew, D.J. 1963. "The Sampling Distribution of an Estimate Arising in Life Testing." *Technometrics*, Volume 5 No. 3, 361-374.

Bartlett, M.S. 1950. "Tests of significance in factor analysis." *British Journal of Psychology (Statistical Section)*, 3, 77-85.

Beal, S. L. 1987. "Asymptotic Confidence Intervals for the Difference between Two Binomial Parameters for Use with Small Samples." *Biometrics*, Volume 43, Issue 4, 941-950.

Belsley, Kuh, and Welsch. 1980. *Regression Diagnostics*. John Wiley & Sons. New York. This is the book that brought regression diagnostics into the main-stream of statistics. It is a graduate level treatise on the subject.

Blackwelder, W.C. 1993. "Sample size and power in prospective analysis of relative risk." *Statistics in Medicine*, Volume 12, 691-698.

Blackwelder, W.C. 1998. "Equivalence Trials." In *Encyclopedia of Biostatistics*, John Wiley and Sons. New York. Volume 2, 1367-1372.

Bloomfield, P. 1976. *Fourier Analysis of Time Series*. John Wiley and Sons. New York. This provides a technical introduction to fourier analysis techniques.

Bock, R.D., Aiken, M. 1981. "Marginal maximum likelihood estimation of item parameters. An application of an EM algorithm. *Psychometrika*, 46, 443-459.

Bonett, Douglas. 2002. “Sample Size Requirements for Testing and Estimating Coefficient Alpha.” *Journal of Educational and Behavioral Statistics*, Vol. 27, pages 335-340.

Box, G.E.P. and Jenkins, G.M. 1976. *Time Series Analysis - Forecasting and Control*. Holden-Day.: San Francisco, California. This is the landmark book on ARIMA time series analysis. Most of the material in chapters 6 - 9 of this manual comes from this work.

Box, G.E.P. 1949. "A general distribution theory for a class of likelihood criteria." *Biometrika*, 1949, **36**, 317-346.

Box, G.E.P. 1954a. "Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variable Problems: I." *Annals of Mathematical Statistics*, **25**, 290-302.

Box, G.E.P. 1954b. "Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variable Problems: II." *Annals of Mathematical Statistics*, **25**, 484-498.

Box, G.E.P., Hunter, S. and Hunter. 1978. *Statistics for Experimenters*. John Wiley & Sons, New York. This is probably the leading book in the area experimental design in industrial experiments. You definitely should acquire and study this book if you plan anything but a casual acquaintance with experimental design. The book is loaded with examples and explanations.

Breslow, N. E. and **Day, N. E.** 1980. *Statistical Methods in Cancer Research: Volume 1. The Analysis of Case-Control Studies*. Lyon: International Agency for Research on Cancer.

Brush, Gary G. 1988. *Volume 12: How to Choose the Proper Sample Size*, American Society for Quality Control, 310 West Wisconsin Ave, Milwaukee, Wisconsin, 53203. This is a small workbook for quality control workers.

Burdick, R.K. and Larsen, G.A. 1997. “Confidence Intervals on Measures of Variability in R&R Studies.” *Journal of Quality Technology, Vol. 29, No. 3, Pages 261-273*. This article presents the formulas used to construct confidence intervals in an R&R study.

Bury, Karl. 1999. *Statistical Distributions in Engineering.*. Cambridge University Press. New York, NY. (www.cup.org).

Cameron, A.C. and Trivedi, P.K. 1998. *Regression Analysis of Count Data*. Cambridge University Press. New York, NY. (www.cup.org).

Carmines, E.G. and Zeller, R.A. 1990. *Reliability and Validity Assessment*. Sage University Paper. 07-017. Newbury Park, CA.

Casagrande, J. T., Pike, M.C., and Smith, P. G. 1978. "The Power Function of the "Exact" Test for Comparing Two Binomial Distributions," *Applied Statistics*, Volume 27, No. 2, pages 176-180. This article presents the algorithm upon which our Fisher's exact test is based.

Cattell, R.B. 1966. "The scree test for the number of factors." *Mult. Behav. Res.* 1, 245-276.

Cattell, R.B. and Jaspers, J. 1967. "A general plasmode (No. 30-10-5-2) for factor analytic exercises and research." *Mult. Behav. Res. Monographs*. 67-3, 1-212.

Chambers, J.M., Cleveland, W.S., Kleiner, B., and Tukey, P.A. 1983. *Graphicals Methods for Data Analysis.* Duxbury Press, Boston, Mass. This wonderful little book is full of examples of ways to analyze data graphically. It gives complete (and readable) coverage to such topics as scatter plots, probability plots, and box plots. It is strongly recommended.

Chatfield, C. 1984. *The Analysis of Time Series*. Chapman and Hall. New York. This book gives a very readable account of both ARMA modeling and spectral analysis. We recommend it to those who wish to get to the bottom of these methods.

Chatterjee and Price. 1979. *Regression Analysis by Example*. John Wiley & Sons. New York. A great hands-on book for those who learn best from examples. A newer edition is now available.

Chen, K.W.; Chow, S.C.; and Li, G. 1997. "A Note on Sample Size Determination for Bioequivalence Studies with Higher-Order Crossover Designs" *Journal of Pharmacokinetics and Biopharmaceutics*, Volume 25, No. 6, pages 753-765.

Chen, T. T. 1997. "Optimal Three-Stage Designs for Phase II Cancer Clinical Trials." *Statistics in Medicine*, Volume 16, pages 2701-2711.

Chen, Xun. 2002. "A quasi-exact method for the confidence intervals of the difference of two independent binomial proportions in small sample cases." *Statistics in Medicine*, Volume 21, pages 943-956.

Chow, S.C. and Liu, J.P. 1999. *Design and Analysis of Bioavailability and Bioequivalence Studies*. Marcel Dekker. New York.

Chow, S.C.; Shao, J.; Wang, H. 2003. *Sample Size Calculations in Clinical Research*. Marcel Dekker. New York.

Cochran and Cox. 1992. *Experimental Designs. Second Edition*. John Wiley & Sons. New York. This is one of the classic books on experimental design, first published in 1957.

Cohen, Jacob. 1988. *Statistical Power Analysis for the Behavioral Sciences*, Lawrence Erlbaum Associates, Hillsdale, New Jersey. This is a very nice, clearly written book. There are MANY examples. It is the largest of the sample size books. It does not deal with clinical trials.

Cohen, Jacob. 1990. "Things I Have Learned So Far." *American Psychologist*, December, 1990, pages 1304-1312. This is must reading for anyone still skeptical about the need for power analysis.

Collett, D. 1991. *Modelling Binary Data.* Chapman & Hall, New York, New York. This book covers such topics as logistic regression, tests of proportions, matched case-control studies, and so on.

Collett, D. 1994. *Modelling Survival Data in Medical Research.* Chapman & Hall, New York, New York. This book covers such survival analysis topics as Cox regression and log rank tests.

Conlon, M. and Thomas, R. 1993. "The Power Function for Fisher's Exact Test." *Applied Statistics*, Volume 42, No. 1, pages 258-260. This article was used to validate the power calculations of Fisher's Exact Test in PASS. Unfortunately, we could not use the algorithm to improve the speed because the algorithm requires equal sample sizes.

Conover, W.J. 1971. *Practical Nonparametric Statistics*. John Wiley & Sons, Inc. New York.

Conover, W.J., Johnson, M.E., and **Johnson, M.M.** 1981. *Technometrics*, **23,** 351-361**.**

Cook, D. and Weisberg, S. 1982. *Residuals and Influence in Regression*. Chapman and Hall. New York. This is an advanced text in the subject of regression diagnostics.

Cooley, W.W. and Lohnes, P.R. 1985. *Multivariate Data Analysis*. Robert F. Krieger Publishing Co. Malabar, Florida.

Cox, D. R. 1972. "Regression Models and life tables." *Journal of the Royal Statistical Society, Series B*, Volume 34, Pages 187-220. This article presents the proportional hazards regression model.

Cox, D. R. 1975. "Contribution to discussion of Mardia (1975a)." *Journal of the Royal Statistical Society, Series B,* Volume 37, Pages 380-381.

Cureton, E.E. and D'Agostino, R.B. 1983. *Factor Analysis - An Applied Approach*. Lawrence Erlbaum Associates. Hillsdale, New Jersey. (This is a wonderful book for those who want to learn the details of what factor analysis does. It has both the theoretical formulas and simple worked examples to make following along very easy.)

D'Agostino, R.B., Belanger, A., D'Agostino, R.B. Jr. 1990."A Suggestion for Using Powerful and Informative Tests of Normality.", *The American Statistician*, November 1990, Volume 44 Number 4, pages 316-321. This tutorial style article discusses D'Agostino's tests and tells how to interpret normal probability plots.

D'Agostino, R.B., Chase, W., Belanger, A. 1988."The Appropriateness of Some Common Procedures for Testing the Equality of Two Independent Binomial Populations.", *The American Statistician*, August 1988, Volume 42 Number 3, pages 198-202.

Dallal, G. 1986. "An Analytic Approximation to the Distribution of Lilliefors's Test Statistic for Normality," *The American Statistician*, Volume 40, Number 4, pages 294-296.

Daniel, C. and Wood, F. 1980. *Fitting Equations to Data*. John Wiley & Sons. New York. This book gives several in depth examples of analyzing regression problems by computer.

Davies, Owen L. 1971. *The Design and Analysis of Industrial Experiments*. Hafner Publishing Company, New York. This was one of the first books on experimental design and analysis. It has many examples and is highly recommended.

Davis, J. C. 1985. *Statistics and Data Analysis in Geology*. John Wiley. New York. (A great layman's discussion of many statistical procedures, including factor analysis.)

Davison, A.C. and Hinkley, D.V. 1999. *Bootstrap Methods and their Applications*. Cambridge University Press. NY, NY. This book provides and detailed account of bootstrapping.

Davison, Mark. 1983. *Multidimensional Scaling*. John Wiley & Sons. NY, NY. This book provides a very good, although somewhat advanced, introduction to the subject.

DeLong, E.R., DeLong, D.M., and Clarke-Pearson, D.L. 1988. "Comparing the Areas Under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach." *Biometrics,* 44, pages 837-845.

DeMets, D.L. and Lan, K.K.G. 1984. "An overview of sequential methods and their applications in clinical trials." *Communications in Statistics, Theory and Methods*, 13, pages 2315-2338.

DeMets, D.L. and Lan, K.K.G. 1994. "Interim analysis: The alpha spending function approach." *Statistics in Medicine*, 13, pages 1341-1352.

Desu, M. M. and Raghavarao, D. 1990. *Sample Size Methodology*. Academic Press. New York. (Presents many useful results for determining sample sizes.)

DeVor, Chang, and Sutherland. 1992. *Statistical Quality Design and Control*. Macmillan Publishing. New York. This is a comprehensive textbook of SPC including control charts, process capability, and experimental design. It has many examples. 800 pages.

Devroye, Luc. 1986. *Non-Uniform Random Variate Generation.* Springer-Verlag. New York. This book is currently available online at http://jeff.cs.mcgill.ca/~luc/rnbookindex.html.

Dillon, W. and Goldstein, M. 1984. *Multivariate Analysis - Methods and Applications*. John Wiley. NY, NY. This book devotes a complete chapter to loglinear models. It follows Fienberg's book, providing additional discussion and examples.

Dixon, W. J. and Tukey, J. W. 1968. “Approximate behavior of the distribution of Winsorized t,” *Technometrics*, Volume 10, pages 83-98.

Dodson, B. 1994. *Weibull Analysis*. ASQC Quality Press. Milwaukee, Wisconsin. This paperback book provides the basics of Weibull fitting. It contains many of the formulas used in our Weibull procedure.

Donnelly, Thomas G. 1980. “ACM Algorithm 462: Bivariate Normal Distribution,” *Collected Algorithms from ACM*, Volume II, New York, New York.

Donner, Allan. 1984. “Approaches to Sample Size Estimation in the Design of Clinical Trials--A Review,” *Statistics in Medicine*, Volume 3, pages 199-214. This is a well done review of the clinical trial literature. Although it is becoming out of date, it is still a good place to start.

Donner, A. and Klar, N. 1996. “Statistical Considerations in the Design and Analysis of Community Intervention Trials.” *The Journal of Clinical Epidemiology*, Vol. 49, No. 4, 1996, pages 435-439.

Donner, A. and Klar, N. 2000. *Design and Analysis of Cluster Randomization Trials in Health Research.* Arnold. London.

Draper, N.R. and Smith, H. 1966. *Applied Regression Analysis*. John Wiley & Sons. New York. This is a classic text in regression analysis. It contains both in depth theory and applications. This text is often used in graduate courses in regression analysis.

Draper, N.R. and Smith, H. 1981. *Applied Regression Analysis - Second Edition*. John Wiley & Sons. New York, NY. This is a classic text in regression analysis. It contains both in-depth theory and applications. It is often used in graduate courses in regression analysis.

du Toit, S.H.C., Steyn, A.G.W., and Stumpf, R.H. 1986. *Graphical Exploratory Data Analysis.* Springer-Verlag. New York. This book contains examples of graphical analysis for a broad range of topics.

Dunn, O. J. 1964. “Multiple comparisons using rank sums,” *Technometrics*, Volume 6, pages 241-252.

Dunnett, C. W. 1955. “A Multiple comparison procedure for Comparing Several Treatments with a Control,” *Journal of the American Statistical Association*, Volume 50, pages 1096-1121.

Dunteman, G.H. 1989. *Principal Components Analysis*. Sage University Papers, 07-069. Newbury Park, California. Telephone (805) 499-0721. This monograph costs only $7. It gives a very good introduction to PCA.

Dupont, William. 1988. “Power Calculations for Matched Case-Control Studies,” *Biometrics*, Volume 44, pages 1157-1168.

Dupont, William and **Plummer, Walton D.** 1990. “Power and Sample Size Calculations--A Review and Computer Program,” *Controlled Clinical Trials*, Volume 11, pages 116-128. Documents a nice public-domain program on sample size and power analysis.

Durbin, J. and Watson, G. S. 1950. “Testing for Serial Correlation in Least Squares Regression - I,” *Biometrika*, Volume 37, pages 409-428.

Durbin, J. and Watson, G. S. 1951. “Testing for Serial Correlation in Least Squares Regression - II,” *Biometrika*, Volume 38, pages 159-177.

Dyke, G.V. and Patterson, H.D. 1952. “Analysis of factorial arrangements when the data are proportions.” *Biometrics*. Volume 8, pages 1-12. This is the source of the data used in the LLM tutorial.

Eckert, Joseph K. 1990. *Property Appraisal and Assessment Administration*. International Association of Assessing Officers. 1313 East 60th Street. Chicago, IL 60637-2892. Phone: (312) 947-2044. This is a how-to manual published by the IAAO that describes how to apply many statistical procedures to real estate appraisal and tax assessment. We strongly recommend it to those using our *Assessment Model* procedure.

Edgington, E. 1987. *Randomization Tests*. Marcel Dekker. New York. A comprehensive discussion of randomization tests with many examples.

Edwards, L.K. 1993. *Applied Analysis of Variance in the Behavior Sciences*. Marcel Dekker. New York. Chapter 8 of this book is used to validate the repeated measures module of PASS.

Efron, B. and Tibshirani, R. J. 1993. *An Introduction to the Bootstrap*. Chapman & Hall. New York.

Elandt-Johnson, R.C. and Johnson, N.L. 1980. *Survival Models and Data Analysis*. John Wiley. NY, NY. This book devotes several chapters to population and clinical life-table analysis.

Epstein, Benjamin. 1960. "Statistical Life Test Acceptance Procedures." *Technometrics*. Volume 2.4, pages 435-446.

Everitt, B.S. and Dunn, G. 1992. *Applied Multivariate Data Analysis*. Oxford University Press. New York. This book provides a very good introduction to several multivariate techniques. It helps you understand how to interpret the results.

Farrington, C. P. and Manning, G. 1990. "Test Statistics and Sample Size Formulae for Comparative Binomial Trials with Null Hypothesis of Non-Zero Risk Difference or Non-Unity Relative Risk." *Statistics in Medicine*, Vol. 9, pages 1447-1454. This article contains the formulas used for the Equivalence of Proportions module in PASS.

Feldt, L.S.; Woodruff, D.J.; & Salih, F.A. 1987. "Statistical inference for coefficient alpha." *Applied Psychological Measurement*, Vol. 11, pages 93-103.

Feldt, L.S.; Ankenmann, R.D. 1999. "Determining Sample Size for a Test of the Equality of Alpha Coefficients When the Number of Part-Tests is Small." *Psychological Methods*, Vol. 4(4), pages 366-377.

Fienberg, S. 1985. *The Analysis of Cross-Classified Categorical Data*. MIT Press. Cambridge, Massachusetts. This book provides a very good introduction to the subject. It is a must for any serious student of the subject.

Finney, D. 1971. *Probit Analysis*. Cambridge University Press. New York, N.Y.

Fisher, N.I. 1993. *Statistical Analysis of Circular Data*. Cambridge University Press. New York, New York.

Fisher, R.A. 1936. "The use of multiple measurements in taxonomic problems." *Annuals of Eugenics*, Volume 7, Part II, 179-188. This article is famous because in it Fisher included the 'iris data' that is always presented when discussing discriminant analysis.

Fleiss, Joseph L. 1981. *Statistical Methods for Rates and Proportions*. John Wiley & Sons. New York. This book provides a very good introduction to the subject.

Fleiss, J. L., Levin, B., Paik, M.C. 2003. *Statistical Methods for Rates and Proportions. Third Edition*. John Wiley & Sons. New York. This book provides a very good introduction to the subject.

Fleiss, Joseph L. 1986. *The Design and Analysis of Clinical Experiments*. John Wiley & Sons. New York. This book provides a very good introduction to clinical trials. It may be a bit out of date now, but it is still very useful.

Fleming, T. R. 1982. "One-sample multiple testing procedure for Phase II clinical trials." *Biometrics*, Volume 38, pages 143-151.

Flury, B. and Riedwyl, H. 1988. *Multivariate Statistics: A Practical Approach*. Chapman and Hall. New York. This is a short, paperback text that provides lots of examples.

Flury, B. 1988. *Common Principal Components and Related Multivariate Models*. John Wiley & Sons. New York. This reference describes several advanced PCA procedures.

Gans. 1984. “The Search for Significance: Different Tests on the Same Data.” *The Journal of Statistical Computation and Simulation*, 1984, pages 1-21.

Gart, John J. and Nam, Jun-mo. 1988. "Approximate Interval Estimation of the Ratio in Binomial Parameters: A Review and Corrections for Skewness." *Biometrics*, Volume 44, Issue 2, 323-338.

Gart, John J. and Nam, Jun-mo. 1990. "Approximate Interval Estimation of the Difference in Binomial Parameters: Correction for Skewness and Extension to Multiple Tables." *Biometrics*, Volume 46, Issue 3, 637-643.

Gehlback, Stephen. 1988. *Interpreting the Medical Literature: Practical Epidemiology for Clinicians*. Second Edition. McGraw-Hill. New York. Telephone: (800)722-4726. The preface of this book states that its purpose is to provide the reader with a useful approach to interpreting the quantitative results given in medical literature. We reference it specifically because of its discussion of ROC curves.

Gentle, James E. 1998. *Random Number Generation and Monte Carlo Methods*. Springer. New York.

Gibbons, J. 1976. *Nonparametric Methods for Quantitative Analysis*. Holt, Rinehart and Winston. New York.

Gleason, T.C. and Staelin, R. 1975. "A proposal for handling missing data." *Psychometrika*, 40, 229-252.

Goldstein, Richard. 1989. “Power and Sample Size via MS/PC-DOS Computers,” *The American Statistician*, Volume 43, Number 4, pages 253-260. A comparative review of power analysis software that was available at that time.

Gomez, K.A. and Gomez, A. A. 1984. *Statistical Procedures for Agricultural Research*. John Wiley & Sons. New York. This reference contains worked-out examples of many complex ANOVA designs. It includes split-plot designs. We recommend it.

Graybill, Franklin. 1961. *An Introduction to Linear Statistical Models*. McGraw-Hill. New York, New York. This is an older book on the theory of linear models. It contains a few worked examples of power analysis.

Greenacre, M. 1984. *Theory and Applications of Correspondence Analysis*. Academic Press. Orlando, Florida. This book goes through several examples. It is probably the most complete book in English on the subject.

Greenacre, Michael J. 1993. *Correspondence Analysis in Practice*. Academic Press. San Diego, CA. This book provides a self-teaching course in correspondence analysis. It is the clearest exposition on the subject that I have every seen. If you want to gain an understanding of CA, you must obtain this (paperback) book.

Griffiths, P. and Hill, I.D. 1985. *Applied Statistics Algorithms*, The Royal Statistical Society, London, England. See page 243 for ACM algorithm 291.

Gross and Clark 1975. *Survival Distributions*: Reliability Applications in Biomedical Sciences. John Wiley, New York.

Guenther, William C. 1977. “Desk Calculation of Probabilities for the Distribution of the Sample Correlation Coefficient,” *The American Statistician*, Volume 31, Number 1, pages 45-48.

Guenther, William C. 1977. *Sampling Inspection in Statistical Quality Control*. Griffin’s Statistical Monographs, Number 37. London.

Haberman, S.J. 1972. “Loglinear Fit of Contingency Tables.” *Applied Statistics*. Volume 21, pages 218-225. This lists the fortran program that is used to create our LLM algorithm.

Hahn, G. J. and Meeker, W.Q. 1991. *Statistical Intervals*. John Wiley & Sons. New York.

Hambleton, R.K; Swaminathan, H; Rogers, H.J. 1991. *Fundamentals of Item Response Theory*. Sage Publications. Newbury Park, California. Phone: (805)499-0721. Provides an inexpensive, readable introduction to IRT. A good place to start.

Hamilton, L. 1991. *Regression with Graphics: A Second Course in Applied Statistics*. Brooks/Cole Publishing Company. Pacific Grove, California. This book gives a great introduction to the use of graphical analysis with regression. It is a must for any serious user of regression. It is written at an introductory level.

Hand, D.J. and Taylor, C.C. 1987. *Multivariate Analysis of Variance and Repeated Measures*. Chapman and Hall. London, England.

Hanley, J. A. and McNeil, B. J. 1982. "The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve." *Radiology*, 143, 29-36. April, 1982.

Hanley, J. A. and McNeil, B. J. 1983. "A Method of Comparing the Areas under Receiver Operating Characteristic Curves Derived from the Same Cases." *Radiology*, 148, 839-843. September, 1983.

Hartigan, J. 1975. *Clustering Algorithms*. John Wiley. New York. (This is the "bible" of cluster algorithms. Hartigan developed the K-means algorithm used in **NCSS**.)

Haupt, R.L. and Haupt, S.E. 1998. *Practical Genetic Algorithms*. John Wiley. New York.

Hernandez-Bermejo, B. and Sorribas, A. 2001. "Analytical Quantile Solution for the S-distribution, Random Number Generation and Statistical Data Modeling." *Biometrical Journal* 43, 1007-1025.

Hintze, J. L. and Nelson, R.D. 1998. "Violin Plots: A Box Plot-Density Trace Synergism." *The American Statistician* 52, 181-184.

Hoaglin, Mosteller, and Tukey. 1985. *Exploring Data Tables, Trends, and Shapes*. John Wiley. New York.

Hoaglin, Mosteller, and Tukey. 1983. *Understanding Robust and Exploratory Data Analysis*. John Wiley & Sons. New York.

Hoerl, A.E. and Kennard, R.W. 1970. "Ridge Regression: Biased estimation for nonorthogonal problems." *Technometrics* 12, 55-82.

Hoerl, A.E. and Kennard R.W. 1976. "Ridge regression: Iterative estimation of the biasing parameter." *Communications in Statistics* A5, 77-88.

Howe, W.G. 1969. "Two-Sided Tolerance Limits for Normal Populations—Some Improvements." *Journal of the American Statistical Association*, 64, 610-620.

Hosmer, D. and Lemeshow, S. 1989. *Applied Logistic Regression*. John Wiley & Sons. New York. This book gives an advanced, in depth look at logistic regression.

Hosmer, D. and Lemeshow, S. 1999. *Applied Survival Analysis*. John Wiley & Sons. New York.

Hotelling, H. 1933. "Analysis of a complex of statistical variables into principal components." *Journal of Educational Psychology* 24, 417-441, 498-520.

Hsieh, F.Y. 1989. "Sample Size Tables for Logistic Regression," *Statistics in Medicine*, Volume 8, pages 795-802. This is the article that was the basis for the sample size calculations in logistic regression in PASS 6.0. It has been superceded by the 1998 article.

Hsieh, F.Y., Block, D.A., and Larsen, M.D. 1998. "A Simple Method of Sample Size Calculation for Linear and Logistic Regression," *Statistics in Medicine*, Volume 17, pages 1623-1634. The sample size calculation for logistic regression in PASS are based on this article.

Hsieh, F.Y. and Lavori, P.W. 2000. "Sample-Size Calculations for the Cox Proportional Hazards Regression Model with Nonbinary Covariates," *Controlled Clinical Trials*, Volume 21, pages 552-560. The sample size calculation for Cox regression in PASS are based on this article.

Hsu, Jason. 1996. *Multiple Comparisons: Theory and Methods.* Chapman & Hall. London. This book gives a beginning to intermediate discussion of multiple comparisons, stressing the interpretation of the various MC tests. It provides details of three popular MC situations: all pairs, versus the best, and versus a control. The power calculations used in the MC module of PASS came from this book.

Jackson, J.E. 1991. *A User's Guide To Principal Components.* John Wiley & Sons. New York. This is a great book to learn about PCA from. It provides several examples and treats everything at a level that is easy to understand.

James, Mike. 1985. *Classification Algorithms*. John Wiley & Sons. New York. This is a great text on the application of discriminant analysis. It includes a simple, easy-to-understand, theoretical development as well as discussions of the application of discriminant analysis.

Jammalamadaka, S.R. and SenGupta, A. 2001. *Topics in Circular Statistics*. World Scientific. River Edge, New Jersey.

Jobson, J.D. 1992. *Applied Multivariate Data Analysis - Volume II: Categorical and Multivariate Methods*. Springer-Verlag. New York. This book is a useful reference for loglinear models and other multivariate methods. It is easy to follows and provides lots of examples.

Jolliffe, I.T. 1972. "Discarding variables in a principal component analysis, I: Artifical data." *Applied Statistics*, 21:160-173.

Johnson, N.L., Kotz, S., and Kemp, A.W. 1992. *Univariate Discrete Distributions, Second Edition*. John Wiley & Sons. New York.

Johnson, N.L., Kotz, S., and Balakrishnan, N. 1994. *Continuous Univariate Distributions Volume 1, Second Edition*. John Wiley & Sons. New York.

Johnson, N.L., Kotz, S., and Balakrishnan, N. 1995. *Continuous Univariate Distributions Volume 2, Second Edition*. John Wiley & Sons. New York.

Jolliffe, I.T. 1986. *Principal Component Analysis*. Springer-Verlag. New York. This book provides an easy-reading introduction to PCA. It goes through several examples.

Julious, Steven A. 2004. "Tutorial in Biostatistics. Sample sizes for clinical trials with Normal data." *Statistics in Medicine*, 23:1921-1986.

Juran, J.M. 1979. *Quality Control Handbook*. McGraw-Hill. New York.

Kaiser, H.F. 1960. "The application of electronic computers to factor analysis." *Educational and Psychological Measurement*. 20:141-151.

Kalbfleisch, J.D. and Prentice, R.L. 1980. *The Statistical Analysis of Failure Time Data*. John Wiley, New York.

Karian, Z.A and Dudewicz, E.J. 2000. *Fitting Statistical Distributions.* CRC Press, New York.

Kaufman, L. and Rousseeuw, P.J. 1990. *Finding Groups in Data*. John Wiley. New York. This book gives an excellent introduction to cluster analysis. It treats the forming of the distance matrix and several different types of cluster methods, including fuzzy. All this is done at an elementary level so that users at all levels can gain from it.

Kay, S.M. 1988. *Modern Spectral Estimation*. Prentice-Hall: Englewood Cliffs, New Jersey. A very technical book on spectral theory.

Kendall,M. and Ord, J.K. 1990. *Time Series.* Oxford University Press. New York. This is theoretical introduction to time series analysis that is very readable.

Kendall,M. and Stuart, A. 1987. *Kendall's Advanced Theory of Statistics. Volume 1: Distribution Theory.* Oxford University Press. New York. This is a fine math-stat book for graduate students in statistics. We reference it because it includes formulas that are used in the program.

Keppel, Geoffrey. 1991. *Design and Analysis - A Researcher's Handbook.* Prentice Hall. Englewood Cliffs, New Jersey. This is a very readable primer on the topic of analysis of variance. Recommended for those who want the straight scoop with a few, well-chosen examples.

Kirk, Roger E. 1982. *Experimental Design: Procedures for the Behavioral Sciences.* Brooks/Cole. Pacific Grove, California. This is a respected reference on experimental design and analysis of variance.

Klein, J.P. and Moeschberger, M.L.. 1997. *Survival Analysis.* Springer-Verlag. New York. This book provides a comprehensive look at the subject complete with formulas, examples, and lots of useful comments. It includes all the more recent developments in this field. I recommend it.

Koch, G.G.; Atkinson, S.S.; Stokes, M.E. 1986. *Encyclopedia of Statistical Sciences.* Volume 7. John Wiley. New York. Edited by Samuel Kotz and Norman Johnson. The article on Poisson Regression provides a very good summary of the subject.

Kotz and Johnson. 1993. *Process Capability Indices.* Chapman & Hall. New York. This book gives a detailed account of the capability indices used in SPC work. 207 pages.

Kraemer, H. C. and **Thiemann, S.** 1987. *How Many Subjects*, Sage Publications, 2111 West Hillcrest Drive, Newbury Park, CA. 91320. This is an excellent introduction to power analysis.

Kruskal, J. 1964. "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis." *Psychometrika* 29, pages 1-27, 115-129. This article presents the algorithm on which the non-metric algorithm used in NCSS is based.

Kruskal, J. and Wish, M. 1978. *Multidimensional Scaling*. Sage Publications. Beverly Hills, CA. This is a well-written monograph by two of the early pioneers of MDS. We suggest it to all serious students of MDS.

Lachenbruch, P.A. 1975. *Discriminant Analysis.* Hafner Press. New York. This is an in-depth treatment of the subject. It covers a lot of territory, but has few examples.

Lachin, John M. 2000. *Biostatistical Methods.* John Wiley & Sons. New York. This is a graduate-level methods book that deals with statistical methods that are of interest to biostatisticians such as odds ratios, relative risks, regression analysis, case-control studies, and so on.

Lachin, John M. and **Foulkes, Mary A. 1986.** "Evaluation of Sample Size and Power for Analyses of Survival with Allowance for Nonuniform Patient Entry, Losses to Follow-up, Noncompliance, and Stratification," *Biometrics*, Volume 42, September, pages 507-516.

Lan, K.K.G. and DeMets, D.L. 1983. "Discrete sequential boundaries for clinical trials." *Biometrika*, 70, pages 659-663.

Lan, K.K.G. and Zucker, D.M. 1993. "Sequential monitoring of clinical trials: the role of information and Brownian motion." *Statistics in Medicine*, 12, pages 753-765.

Lance, G.N. and Williams, W.T. 1967. "A general theory of classificatory sorting strategies. I. Hierarchical systems." *Comput. J.* 9, pages 373-380.

Lance, G.N. and Williams, W.T. 1967. "Mixed-data classificatory programs I. Agglomerative systems." *Aust.Comput. J.* 1, pages 15-20.

Lawless, J.F. 1982. *Statistical Models and Methods for Lifetime Data*. John Wiley, New York.

Lawson, John. 1987. *Basic Industrial Experimental Design Strategies*. Center for Statistical Research at Brigham Young University. Provo, Utah. 84602. This is a manuscript used by Dr. Lawson in courses and workshops that he provides to industrial engineers. It is the basis for many of our experimental design procedures.

Lebart, Morineau, and Warwick. 1984. *Multivariate Descriptive Statistical Analysis*. John Wiley & Sons. This book devotes a large percentage of its discussion to correspondence analysis.

Lee, E.T. 1974. "A Computer Program for Linear Logistic Regression Analysis" in *Computer Programs in Biomedicine*, Volume 4, pages 80-92.

Lee, E.T. 1980. *Statistical Methods for Survival Data Analysis*. Lifetime Learning Publications. Belmont, California.

Lee, E.T. 1992. *Statistical Methods for Survival Data Analysis*. Second Edition. John Wiley & Sons. New York. This book provides a very readable introduction to survival analysis techniques.

Lee, S. K. 1977. "On the Asymptotic Variances of u Terms in Loglinear Models of Multidimensional Contingency Tables." *Journal of the American Statistical Association*. Volume 72 (June, 1977), page 412. This article describes methods for computing standard errors that are used in the LLM section of this program.

Lenth, Russell V. 1987. "Algorithm AS 226: Computing Noncentral Beta Probabilities," *Applied Statistics*, Volume 36, pages 241-244.

Lenth, Russell V. 1989. "Algorithm AS 243: Cumulative Distribution Function of the Non-central t Distribution," *Applied Statistics*, Volume 38, pages 185-189.

Lesaffre, E. and Albert, A. 1989. "Multiple-group Logistic Regression Diagnostics" *Applied Statistics*, Volume 38, pages 425-440. See also Pregibon 1981.

Levene, H. 1960. In *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, I. Olkin et al., eds. Stanford University Press, Stanford Calif., pp. 278-292.

Lewis, J.A. 1999. "Statistical principles for clinical trials (ICH E9) an introductory note on an international guideline." *Statistics in Medicine*, 18, pages 1903-1942.

Lipsey, Mark W. 1990. *Design Sensitivity Statistical Power for Experimental Research*, Sage Publications, 2111 West Hillcrest Drive, Newbury Park, CA. 91320. This is an excellent introduction to power analysis.

Little, R. and Rubin, D. 1987. *Statistical Analysis with Missing Data*. John Wiley & Sons. New York. This book is completely devoted to dealing with missing values. It gives a complete treatment of using the EM algorithm to estimate the covariance matrix.

Lu, Y. and Bean, J.A. 1995. "On the sample size for one-sided equivalence of sensitivities based upon McNemar's test," *Statistics in Medicine*, Volume 14, pages 1831-1839.

Lui, J., Hsueh, H., Hsieh, E., and Chen, J.J. 2002. "Tests for equivalence or non-inferiority for paired binary data," *Statistics in Medicine*, Volume 21, pages 231-245.

Lloyd, D.K. and Lipow, M. 1991. *Reliability: Management, Methods, and Mathematics*. ASQC Quality Press. Milwaukee, Wisconsin.

Locke, C.S. 1984. "An exact confidence interval for untransformed data for the ratio of two formulation means," *J. Pharmacokinet. Biopharm.*, Volume 12, pages 649-655.

Lockhart, R. A. & Stephens, M. A. 1985. "Tests of fit for the von Mises distribution." *Biometrika* 72, pages 647-652.

Machin, D., Campbell, M., Fayers, P., and Pinol, A. 1997. *Sample Size Tables for Clinical Studies, 2nd Edition*. Blackwell Science. Malden, Mass. A very good & easy to read book on determining appropriate sample sizes in many situations.

Makridakis, S. and Wheelwright, S.C. 1978. *Iterative Forecasting*. Holden-Day.: San Francisco, California. This is a very good book for the layman since it includes several detailed examples. It is written for a person with a minimum amount of mathematical background.

Manly, B.F.J. 1986. *Multivariate Statistical Methods - A Primer*. Chapman and Hall. New York. This nice little paperback provides a simplified introduction to many multivariate techniques, including MDS.

Mardia, K.V. and Jupp, P.E. 2000. *Directional Statistics*. John Wiley & Sons. New York.

Marple, S.L. 1987. *Digital Spectral Analysis with Applications*. Prentice-Hall: Englewood Cliffs, New Jersey. A technical book about spectral analysis.

Martinez and Iglewicz. 1981. "A test for departure from normality based on a biweight estimator of scale." *Biometrika*, 68, 331-333).

Marubini, E. and **Valsecchi, M.G.** 1996. *Analysing Survival Data from Clinical Trials and Observational Studies*. John Wiley: New York, New York.

Mather, Paul. 1976. *Computational Methods of Multivariate Analysis in Physical Geography*. John Wiley & Sons. This is a great book for getting the details on several multivariate procedures. It was written for non-statisticians. It is especially useful in its presentation of cluster analysis. Unfortunately, it is out-of-print. You will have to look for it in a university library (it is worth the hunt).

Matsumoto, M. and Nishimura,T. 1998. "Mersenne twister: A 623-dimensionally equidistributed uniform pseudorandom number generator" *ACM Trans. On Modeling and Computer Simulations*.

Mauchly, J.W. 1940. "Significance test for sphericity of a normal n-variate distribution." *Annals of Mathematical Statistics*, 11: 204-209

McCabe, G.P. 1984. "Principal variables." *Technometrics*, 26, 137-144.

McClish, D.K. 1989. "Analyzing a Portion of the ROC Curve." *Medical Decision Making*, 9: 190-195

McHenry, Claude. 1978. "Multivariate subset selection." *Journal of the Royal Statistical Society, Series C*. Volume 27, No. 23, pages 291-296.

McNeil, D.R. 1977. *Interactive Data Analysis*. John Wiley & Sons. New York.

Mendenhall, W. 1968. *Introduction to Linear Models and the Design and Analysis of Experiments*. Wadsworth. Belmont, Calif.

Metz, C.E. 1978. "Basic principles of ROC analysis." *Seminars in Nuclear Medicine,* Volume 8, No. 4, pages 283-298.

Miettinen, O.S. and Nurminen, M. 1985. "Comparative analysis of two rates." *Statistics in Medicine* 4: 213-226.

Milliken, G.A. and Johnson, D.E. 1984. *Analysis of Messy Data, Volume 1*. Van Nostrand Rienhold. New York, NY.

Milne, P. 1987. *Computer Graphics for Surveying.* E. & F. N. Spon, 29 West 35th St., NY, NY 10001

Montgomery, Douglas. 1984. *Design and Analysis of Experiments*. John Wiley & Sons, New York. A textbook covering a broad range of experimental design methods. The book is not limited to industrial investigations, but gives a much more general overview of experimental design methodology.

Montgomery, Douglas and Peck. 1992. *Introduction to Linear Regression Analysis*. A very good book on this topic.

Montgomery, Douglas C. 1991. *Introduction to Statistical Quality Control.* Second edition. John Wiley & Sons. New York. This is a comprehensive textbook of SPC including control charts, process capability, and experimental design. It has many examples. 700 pages.

Moore, D. S. and McCabe, G. P. 1999. *Introduction to the Practice of Statistics*. W. H. Freeman and Company. New York.

Mosteller, F. and Tukey, J.W. 1977. *Data Analysis and Regression.* Addison-Wesley. Menlo Park, California. This book should be read by all serious users of regression analysis. Although the terminology is a little different, this book will give you a fresh look at the whole subject.

Motulsky, Harvey. 1995. *Intuitive Biostatistics.* Oxford University Press. New York, New York. This is a wonderful book for those who want to understand the basic concepts of statistical testing. The author presents a very readable coverage of the most popular biostatistics tests. If you have forgotten how to interpret the various statistical tests, get this book!

Moura, Eduardo C. 1991. *How To Determine Sample Size And Estimate Failure Rate in Life Testing*. ASQC Quality Press. Milwaukee, Wisconsin.

Mueller, K. E., and Barton, C. N. 1989. "Approximate Power for Repeated-Measures ANOVA Lacking Sphericity." *Journal of the American Statistical Association*, Volume 84, No. 406, pages 549-555.

Mueller, K. E., LaVange, L.E., Ramey, S.L., and Ramey, C.T. 1992. “Power Calculations for General Linear Multivariate Models Including Repeated Measures Applications.” *Journal of the American Statistical Association*, Volume 87, No. 420, pages 1209-1226.

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.

Myers, R.H. 1990. *Classical and Modern Regression with Applications*. PWS-Kent Publishing Company. Boston, Massachusetts. This is one of the bibles on the topic of regression analysis.

Nam, Jun-mo. 1992. “Sample Size Determination for Case-Control Studies and the Comparison of Stratified and Unstratified Analyses,” *Biometrics*, Volume 48, pages 389-395.

Nam, Jun-mo. 1997. “Establishing equivalence of two treatments and sample size requirements in matched-pairs design,” *Biometrics*, Volume 53, pages 1422-1430.

Nam, J-m. and Blackwelder, W.C. 2002. “Analysis of the ratio of marginal probabilities in a matched-pair setting,” *Statistics in Medicine*, Volume 21, pages 689-699.

Nash, J. C. 1987. *Nonlinear Parameter Estimation*. Marcel Dekker, Inc. New York, NY.

Nash, J.C. 1979. *Compact Numerical Methods for Computers*. John Wiley & Sons. New York, NY.

Nel, D.G. and van der Merwe, C.A. 1986. “A solution to the multivariate Behrens-Fisher problem.” *Communications in Statistics—Series A, Theory and Methods*, 15, pages 3719-3735.

Nelson, W.B. 1982. *Applied Life Data Analysis*. John Wiley, New York.

Nelson, W.B. 1990. *Accelerated Testing*. John Wiley, New York.

Neter, J., Kutner, M., Nachtsheim, C., and Wasserman, W. 1996. *Applied Linear Statistical Models*. Richard D. Irwin, Inc. Chicago, Illinois. This mammoth book covers regression analysis and analysis of variance thoroughly and in great detail. We recommend it.

Neter, J., Wasserman, W., and Kutner, M. 1983. *Applied Linear Regression Models*. Richard D. Irwin, Inc. Chicago, Illinois. This book provides you with a complete introduction to the methods of regression analysis. We suggest it to non-statisticians as a great reference tool.

Newcombe, Robert G. 1998a. "Two-Sided Confidence Intervals for the Single Proportion: Comparison of Seven Methods." *Statistics in Medicine*, Volume 17, 857-872.

Newcombe, Robert G. 1998b. "Interval Estimation for the Difference Between Independent Proportions: Comparison of Eleven Methods." *Statistics in Medicine*, Volume 17, 873-890.

Newcombe, Robert G. 1998c. "Improved Confidence Intervals for the Difference Between Binomial Proportions Based on Paired Data." *Statistics in Medicine*, Volume 17, 2635-2650.

Newton, H.J. 1988. *TIMESLAB: A Time Series Analysis Laboratory*. Wadsworth & Brooks/Cole: Pacific Grove, California. This book is loaded with theoretical information about time series analysis. It includes software designed by Dr. Newton for performing advanced time series and spectral analysis. The book requires a strong math and statistical background.

O’Brien, P.C. and Fleming, T.R. 1979. “A multiple testing procedure for clinical trials.” *Biometrics*, 35, pages 549-556.

O’Brien, R.G. and Kaiser, M.K. 1985. “MANOVA Method for Analyzing Repeated Measures Designs: An Extensive Primer.” *Psychological Bulletin*, 97, pages 316-333.

Obuchowski, N. 1998. “Sample Size Calculations in Studies of Test Accuracy.” *Statistical Methods in Medical Research*, 7, pages 371-392.

Obuchowski, N. and McClish, D. 1997. “Sample Size Determination for Diagnostic Accuracy Studies Involving Binormal ROC Curve Indices.” *Statistics in Medicine*, 16, pages 1529-1542.

Odeh, R.E. and Fox, M. 1991. *Sample Size Choice*. Marcel Dekker, Inc. New York, NY.

O’Neill and Wetherill. 1971 “The Present State of Multiple Comparison Methods,” *The Journal of the Royal Statistical Society*, Series B, vol.33, 218-250).

Orloci, L. & Kenkel, N. 1985. *Introduction to Data Analysis*. International Co-operative Publishing House. Fairland, Maryland. This book was written for ecologists. It contains samples and BASIC programs of many statistical procedures. It has one brief chapter on MDS, and it includes a non-metric MDS algorithm.

Ostle, B. 1988. *Statistics in Research. Fourth Edition*. Iowa State Press. Ames, Iowa. A comprehension book on statistical methods.

Ott, L. 1977. *An Introduction to Statistical Methods and Data Analysis*. Wadsworth. Belmont, Calif. Use the second edition.

Ott, L. 1984. *An Introduction to Statistical Methods and Data Analysis, Second Edition*. Wadsworth. Belmont, Calif. This is a complete methods text. Regression analysis is the focus of five or six chapters. It stresses the interpretation of the statistics rather than the calculation, hence it provides a good companion to a statistical program like ours.

Owen, Donald B. 1956. "Tables for Computing Bivariate Normal Probabilities," *Annals of Mathematical Statistics*, Volume 27, pages 1075-1090.

Owen, Donald B. 1965. "A Special Case of a Bivariate Non-Central t-Distribution," *Biometrika*, Volume 52, pages 437-446.

Pandit, S.M. and Wu, S.M. 1983. *Time Series and System Analysis with Applications*. John Wiley and Sons. New York. This book provides an alternative to the Box-Jenkins approach for dealing with ARMA models. We used this approach in developing our automatic ARMA module.

Parmar, M.K.B. and Machin, D. 1995. *Survival Analysis*. John Wiley and Sons. New York.

Parmar, M.K.B., Torri, V., and Steart, L. 1998. "Extracting Summary Statistics to Perform Meta-Analyses of the Published Literature for Survival Endpoints." *Statistics in Medicine* 17, 2815-2834.

Pearson, K. 1901. "On lines and planes of closest fit to a system of points in space." *Philosophical Magazine* 2, 557-572.

Pan, Z. and Kupper, L. 1999. "Sample Size Determination for Multiple Comparison Studies Treating Confidence Interval Width as Random." *Statistics in Medicine* 18, 1475-1488.

Pedhazur, E.L. and Schmelkin, L.P. 1991. *Measurement, Design, and Analysis: An Integrated Approach.* Lawrence Erlbaum Associates. Hillsdale, New Jersey. This mammoth book (over 800 pages) covers multivariate analysis, regression analysis, experimental design, analysis of variance, and much more. It provides annotated output from SPSS and SAS which is also useful to our users. The text is emphasizes the social sciences. It provides a "how-to," rather than a theoretical, discussion. Its chapters on factor analysis are especially informative.

Phillips, Kem F. 1990. "Power of the Two One-Sided Tests Procedure in Bioequivalence," *Journal of Pharmacokinetics and Biopharmaceutics*, Volume 18, No. 2, pages 137-144.

Pocock, S.J. 1977. "Group sequential methods in the design and analysis of clinical trials." *Biometrika*, 64, pages 191-199.

Press, S. J. and Wilson, S. 1978. "Choosing Between Logistic Regression and Discriminant Analysis." *Journal of the American Statistical Association*, Volume 73, Number 364, Pages 699-705. This article details the reasons why logistic regression should be the preferred technique.

Press, William H. 1986. *Numerical Recipes*, Cambridge University Press, New York, New York.

Pregibon, Daryl. 1981. "Logistic Regression Diagnostics." *Annals of Statistics*, Volume 9, Pages 705-725. This article details the extensions of the usual regression diagnostics to the case of logistic regression. These results were extended to multiple-group logistic regression in Lesaffre and Albert (1989).

Prihoda, Tom. 1983. "Convenient Power Analysis For Complex Analysis of Variance Models." *Poster Session of the American Statistical Association Joint Statistical Meetings*, August 15-18, 1983, Toronto, Canada. Tom is currently at the University of Texas Health Science Center. This article includes FORTRAN code for performing power analysis.

Ramsey, Philip H. 1978 “Power Differences Between Pairwise Multiple Comparisons,” *JASA*, vol. 73, no. 363, pages 479-485.

Rao, C.R. , Mitra, S.K., & Matthai, A. 1966. *Formulae and Tables for Statistical Work*. Statistical Publishing Society, Indian Statistical Institute, Calcutta, India.

Ratkowsky, David A. 1989. *Handbook of Nonlinear Regression Models*. Marcel Dekker. New York. A good, but technical, discussion of various nonlinear regression models.

Rawlings John O. 1988. *Applied Regression Analysis: A Research Tool*. Wadsworth. Belmont, California. This is a readable book on regression analysis. It provides a thorough discourse on the subject.

Reboussin, D.M., DeMets, D.L., Kim, K, and Lan, K.K.G. 1992. “Programs for computing group sequential boundaries using the Lan-DeMets Method.” Technical Report 60, Department of Biostatistics, University of Wisconsin-Madison.

Rencher, Alvin C. 1998. *Multivariate Statistical Inference and Applications*. John Wiley. New York, New York. This book provides a comprehensive mixture of theoretical and applied results in multivariate analysis. My evaluation may be biased since Al Rencher took me fishing when I was his student.

Robins, Greenland, and Breslow. 1986 “A General Estimator for the Variance of the Mantel-Haenszel Odds Ratio,” *American Journal of Epidemiology*, vol.42, pages 719-723.

Robins, Breslow, and Greenland. 1986 “Estimators of the Mantel-Haenszel variance consisten in both sparse data and large-strata limiting models,” *Biometrics*, vol. 42, pages 311-323.

Ryan, Thomas P. 1989. *Statistical Methods for Quality Improvement*. John Wiley & Sons. New York. This is a comprehensive treatment of SPC including control charts, process capability, and experimental design. It provides many rules-of-thumb and discusses many non-standard situations. This is a very good ‘operators manual’ type of book. 446 pages.

Ryan, Thomas P. 1997. *Modern Regression Methods*. John Wiley & Sons. New York. This is a comprehensive treatment of regression analysis. The author often deals with practical issues that are left out of other texts.

Sahai, Hardeo & Khurshid, Anwer. 1995. *Statistics in Epidemiology*. CRC Press. Boca Raton, Florida.

Schiffman, Reynolds, & Young. 1981. *Introduction to Multidimensional Scaling*. Academic Press. Orlando, Florida. This book goes through several examples.

Schilling, Edward. 1982. *Acceptance Sampling in Quality Control*. Marcel-Dekker. New York.

Schlesselman, Jim. 1981. *Case-Control Studies*. Oxford University Press. New York. This presents a complete overview of case-control studies. It was our primary source for the Mantel-Haenszel test.

Schmee and Hahn. November, 1979. "A Simple Method for Regression Analysis." *Technometrics*, Volume 21, Number 4, pages 417-432.

Schoenfeld, David A. 1983**.** “Sample-Size Formula for the Proportional-Hazards Regression Model” *Biometrics*, Volume 39, pages 499-503.

Schoenfeld, David A. and **Richter, Jane R.** 1982**.** “Nomograms for Calculating the Number of Patients Needed for a Clinical Trial with Survival as an Endpoint,” *Biometrics*, March 1982, Volume 38, pages 163-170.

Schork, M. and Williams, G. 1980. "Number of Observations Required for the Comparison of Two Correlated Proportions." *Communications in Statistics-Simula. Computa.*, B9(4), 349-357.

Schuirmann, Donald. 1981**.** “On hypothesis testing to determine if the mean of a normal distribution is continued in a known interval,” *Biometrics*, Volume 37, pages 617.

Schuirmann, Donald. 1987**.** “A Comparison of the Two One-Sided Tests Procedure and the Power Approach for Assessing the Equivalence of Average Bioavailability,” *Journal of Pharmacokinetics and Biopharmaceutics*, Volume 15, Number 6, pages 657-680.

Seber, G.A.F. 1984. *Multivariate Observations*. John Wiley & Sons. New York. (This book is an encyclopedia of multivariate techniques. It emphasizes the mathematical details of each technique and provides a complete set of references. It will only be useful to those comfortable with reading mathematical equations based on matrices.)

Seber, G.A.F. and Wild, C.J. 1989. *Nonlinear Regression*. John Wiley & Sons. New York. This book is an encyclopedia of nonlinear regression.

Senn, Stephen. 1993. *Cross-over Trials in Clinical Research*. John Wiley & Sons. New York.

Senn, Stephen. 2002. *Cross-over Trials in Clinical Research*. Second Edition. John Wiley & Sons. New York.

Shapiro, S.S. and Wilk, M.B. 1965 "An analysis of Variance test for normality." *Biometrika*, Volume 52, pages 591-611.

Shuster, Jonathan J. 1990. *CRC Handbook of Sample Size Guidelines for Clinical Trials*. CRC Press, Boca Raton, Florida. This is an expensive book ($300) of tables for running log-rank tests. It is well documented, but at this price it better be.

Signorini, David. 1991. "Sample size for Poisson regression," *Biometrika,* Volume 78, 2, pages 446-450.

Simon, Richard. "Optimal Two-Stage Designs for Phase II Clinical Trials," *Controlled Clinical Trials,* 1989, Volume 10, pages 1-10.

Snedecor, G. and Cochran, Wm. 1972. *Statistical Methods*. The Iowa State University Press. Ames, Iowa.

Sorribas, A., March, J., and Trujillano, J. 2002. "A new parametric method based on S-distributions for computing receiver operating characteristic curves for continuous diagnostic tests." *Statistics in Medicine* 21, 1213-1235.

Spath, H. 1985. *Cluster Dissection and Analysis.* Halsted Press. New York. (This book contains a detailed discussion of clustering techniques for large data sets. It contains some heavy mathematical notation.)

Sutton, A.J., Abrams, K.R., Jones, D.R., Sheldon, T.A., and Song, F. 2000. *Methods for Meta-Analysis in Medical Research.* John Wiley & Sons. New York.

Swets, John A. 1996. *Signal Detection Theory and ROC Analysis in Psychology and Diagnostics - Collected Papers.* Lawrence Erlbaum Associates. Mahway, New Jersey.

Tabachnick, B. and Fidell, L. 1989. *Using Multivariate Statistics.* Harper Collins. 10 East 53d Street, NY, NY 10022. This is an extremely useful text on multivariate techniques. It presents computer printouts and discussion from several popular programs. It provides checklists for each procedure as well as sample written reports. <u>I strongly encourage you to obtain this book!</u>

Tango, Toshiro. 1998. "Equivalence Test and Confidence Interval for the Difference in Proportions for the Paired-Sample Design." *Statistics in Medicine*, Volume 17, 891-908.

Therneau, T.M. and Grambsch, P.M. 2000. *Modeling Survival Data*. Springer: New York, New York. A the time of the writing of the Cox regression procedure, this book provides a thorough, up-to-date discussion of this procedure as well as many extensions to it. Recommended, especially to those with at least a masters in statistics.

Thomopoulos, N.T. 1980. *Applied Forecasting Methods*. Prentice-Hall: Englewood Cliffs, New Jersey. This book contains a very good presentation of the classical forecasting methods discussed in chapter two.

Thompson, Simon G. 1998. *Encyclopedia of Biostatistics, Volume 4*. John Wiley & Sons. New York. Article on Meta-Analysis on pages 2570-2579.

Tiku, M. L. 1965. "Laguerre Series Forms of Non-Central X^2 and F Distributions," *Biometrika*, Volume 42, pages 415-427.

Torgenson, W.S. 1952. “Multidimensional scaling. I. Theory and method.” *Psychometrika* 17, 401-419. This is one of the first articles on MDS. There have been many advances, but this article presents many insights into the application of the technique. It describes the algorithm on which the metric solution used in this program is based.

Tubert-Bitter, P., Manfredi,R., Lellouch, J., Begaud, B. 2000. “Sample size calculations for risk equivalence testing in pharmacoepidemiology.” *Journal of Clinical Epidemiology* 53, 1268-1274.

Tukey, J.W. and McLaughlin, D.H. 1963. “Less Vulnerable confidence and significance procedures for location based on a single sample: Trimming/Winsorization.” *Sankhya, Series A* 25, 331-352.

Tukey, J.W. 1977. *Exploratory Data Analysis*. Addison-Wesley Publishing Company. Reading, Mass.

Upton, G.J.G. 1982."A Comparison of Alternative Tests for the 2 x 2 Comparative Trial.", *Journal of the Royal Statistical Society,* Series A,, Volume 145, pages 86-105.

Upton, G.J.G. and Fingleton, B. 1989. *Spatial Data Analysis by Example: Categorical and Directional Data. Volume 2.* John Wiley & Sons. New York.

Velicer, W.F. 1976. "Determining the number of components from the matrix of partial correlations." *Psychometrika*, 41, 321-327.

Velleman, Hoaglin. 1981. *ABC's of Exploratory Data Analysis*. Duxbury Press, Boston, Massachusets.

Voit, E.O. 1992. "The S-distribution. A tool for approximation and classification of univariate, unimodal probability distributions." *Biometrical J.* 34, 855-878.

Voit, E.O. 2000. "A Maximum Likelihood Estimator for Shape Parameters of S-Distributions." *Biometrical J.* 42, 471-479.

Voit, E.O. and Schwacke, L. 1998. "Scalability properties of the S-distribution." *Biometrical J.* 40, 665-684.

Voit, E.O. and Yu, S. 1994. "The S-distribution. Approximation of discrete distributions." *Biometrical J.* 36, 205-219.

Walter, S.D., Eliasziw, M., and Donner, A. 1998. "Sample Size and Optimal Designs For Reliability Studies.” *Statistics in Medicine*, 17, 101-110.

Welch, B.L. 1938. "The significance of the difference between two means when the population variances are unequal.” *Biometrika*, 29, 350-362.

Westgard, J.O. 1981. “A Multi-Rule Shewhart Chart for Quality Control in Clinical Chemistry,” *Clinical Chemistry*, Volume 27, No. 3, pages 493-501. (This paper is available online at the www.westgard.com).

Westlake, W.J. 1981. “Bioequivalence testing—a need to rethink,” *Biometrics*, Volume 37, pages 591-593.

Whittemore, Alice. 1981. “Sample Size for Logistic Regression with Small Response Probability,” *Journal of the American Statistical Association*, Volume 76, pages 27-32.

Wickens, T.D. 1989. *Multiway Contingency Tables Analysis for the Social Sciences.* Lawrence Erlbaum Associates. Hillsdale, New Jersey. A thorough book on the subject. Discusses loglinear models in depth.

Wilson, E.B.. 1927. “Probable Inference, the Law of Succession, and Statistical Inference,” *Journal of the American Statistical Association*, Volume 22, pages 209-212. This article discusses the ‘score’ method that has become popular when dealing with proportions.

Winer, B.J. 1991**.** *Statistical Principles in Experimental Design (Third Edition)*. McGraw-Hill. New York, NY. A very complete analysis of variance book.

Woolson, R.F., Bean, J.A., and Rojas, P.B. 1986. “Sample Size for Case-Control Studies Using Cochran’s Statistic,” *Biometrics*, Volume 42, pages 927-932.

Yuen, K.K. and Dixon, W. J. 1973. "The approximate behavior and performance of the two-sample trimmed t," *Biometrika*, Volume 60, pages 369-374.

Yuen, K.K. 1974. "The two-sample trimmed t for unequal population variances," *Biometrika*, Volume 61, pages 165-170.

Zar, Jerrold H. 1984**.** *Biostatistical Analysis (Second Edition)*. Prentice-Hall. Englewood Cliffs, New Jersey. This introductory book presents a nice blend of theory, methods, and examples for a long list of topics of interest in biostatistical work.

Zhou, X., Obuchowski, N., McClish, D. 2002**.** *Statistical Methods in Diagnostic Medicine*. John Wiley & Sons, Inc. New York, New York. This is a great book on the designing and analyzing diagnostic tests. It is especially useful for its presentation of ROC curves.

Index

Note that most index entries are of the form “chapter-page”. If no chapter is given, the entry refers to the introductory chapters. A list of chapters is given in the Table of Contents.