User's Guide I

Quick Start, Proportions, and Diagnostic Tests

PASS
Power Analysis
and
Sample Size
System

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

PASS User's Guide I

Copyright © 2011
Dr. Jerry L. Hintze
Kaysville, Utah 84037

All Rights Reserved

Direct inquiries to:

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

PASS is a trademark of Dr. Jerry L. Hintze and NCSS.

Warning:

This software and manual are both protected by U.S. Copyright Law (Title 17 United States Code). Unauthorized reproduction and/or sales may result in imprisonment of up to one year and fines of up to $10,000 (17 USC 506). Copyright infringers may also be subject to civil liability.

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, President

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. Your purchase and use of the PASS 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 nor 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 NCSS 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 and other members of the development team over the last twenty 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 and 64-bit versions of Microsoft Windows (7, Vista, XP) 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.

User's Guide I
Table of Contents

Quick Start

1 Installation
2 Running PASS
3 The PASS Home Window
4 The Procedure Window
5 The Output Window
6 Introduction to Power Analysis
7 Proportions
8 Means

Quick Start Index

Proportions

One Proportion

100 Inequality Tests
103 Non-Zero Null Tests
105 Non-Inferiority Tests
110 Equivalence Tests
115 Confidence Intervals
120 Single-Stage Phase II Clinical Trials
125 Two-Stage Phase II Clinical Trials
130 Three-Stage Phase II Clinical Trials
135 Post-Marketing Surveillance

Two Correlated Proportions

150 Inequality Tests (McNemar Test)
155 Inequality Tests (Matched Case-Control Design)
160 Non-Inferiority Tests
165 Equivalence Tests

Two Independent Proportions

200 Inequality Tests
201 Inequality Tests (Repeated Measures Design)
205 Non-Zero Null Tests
210 Non-Inferiority Tests
215 Equivalence Tests
216 Confidence Intervals
220 Group-Sequential Tests
221 Group-Sequential Tests (Simulation)
222 Non-Zero Null Group-Sequential Tests (Simulation)
223 Group-Sequential Non-Inferiority Tests (Simulation)
225 Inequality Tests (Stratified Design – Cochran-Mantel-Haenszel Test)

Two Independent Proportions in a Cluster-Randomized Design

230 Inequality Tests
233 Non-Zero Null Tests
235 Non-Inferiority Tests
240 Equivalence Tests

Many Proportions (Contingency Tables)

250 Chi-Square Tests
253 Two Ordered Categorical Variable Tests
255 Cochran-Armitage Test for Trend in Proportions

ROC Curves

260 Inequality Tests for One ROC Curve
265 Inequality Tests for Two ROC Curves

Sensitivity and Specificity

270 Sensitivity and Specificity Test for One Group
275 Sensitivity Test for Two Groups
276 Tests for Paired Sensitivities

References and Index

References
Index

User's Guide II
Table of Contents

Control Charts

290 Control Charts for Process Means (Simulation)
295 Control Charts for Process Variation (Simulation)

Means

One Mean

400 Inequality Tests
405 Inequality Tests (Exponential Data)
410 Inequality Tests (Simulation)
412 Inequality Tests (Poisson Data)
413 Non-Zero Null Tests
415 Non-Inferiority Tests
420 Confidence Intervals
421 Confidence Intervals with Tolerance Probability

Two Independent Means

430 Inequality Tests using Differences (Two-Sample T-Test)
431 Inequality Tests (Repeated Measures Design)
435 Inequality Tests (Exponential Data)
437 Inequality Tests (Poisson Data)
440 Inequality Tests (Simulation)
445 Inequality Tests using Ratios
448 Non-Zero Null Tests using Differences
450 Non-Inferiority Tests using Differences
453 Non-Unity Null Tests using Ratios
455 Non-Inferiority Tests using Ratios
460 Equivalence Tests using Differences
465 Equivalence Tests (Simulation)
470 Equivalence Tests using Ratios
471 Confidence Intervals for the Difference
472 Confidence Intervals for the Difference with Tolerance Probability
475 Group-Sequential Tests
476 Group-Sequential Tests (Simulation)
477 Group-Sequential Tests assuming Normality (Simulation)
478 Group-Sequential Non-Inferiority Tests (Simulation)
480 Inequality Tests (Cluster-Randomized Design)

Two Correlated (Paired) Means

485 Inequality Tests
490 Inequality Tests (Simulation)
495 Equivalence Tests (Simulation)
496 Confidence Intervals
497 Confidence Intervals with Tolerance Probability

Two Independent Means in a 2×2 Cross-Over Design

500 Inequality Tests using Differences
505 Inequality Tests using Ratios
508 Non-Zero Null Tests using Differences
510 Non-Inferiority Tests using Differences
513 Non-Unity Null Tests using Ratios
515 Non-Inferiority Tests using Ratios
520 Equivalence Tests using Differences
525 Equivalence Tests using Ratios

Two Independent Means in a Higher-Order Cross-Over Design

528 Non-Zero Null Tests using Differences
530 Non-Inferiority Tests using Differences
533 Non-Unity Null Tests using Ratios
535 Non-Inferiority Tests using Ratios
540 Equivalence Tests using Differences
545 Equivalence Tests using Ratios

Many Means (ANOVA)

550 One-Way Analysis of Variance
551 Analysis of Covariance (ANCOVA)
555 One-Way Analysis of Variance (Simulation)
560 Fixed Effects Analysis of Variance
565 Randomized Block Analysis of Variance
570 Repeated Measures Analysis of Variance

References and Index

References
Index

User's Guide III
Table of Contents

Means (Continued)

Mixed Models

571 Mixed Models

Multiple Comparisons

575 Multiple Comparisons
580 Pair-Wise Multiple Comparisons (Simulation)
585 Multiple Comparisons of Treatments vs. a Control (Simulation)
590 Multiple Contrasts (Simulation)
595 Williams Test for the Minimum Effective Dose

Multivariate Means

600 Hotelling's T2
605 Multivariate Analysis of Variance (MANOVA)

Microarrays

610 One-Sample or Paired T-Test
615 Two-Sample T-Test

Standard Deviations

640 Confidence Intervals for One Standard Deviation using Standard Deviation
641 Confidence Intervals for One Standard Deviation with Tolerance Probability
642 Confidence Intervals for One Standard Deviation using Relative Error

Variances

One Variance

650 Inequality Tests
651 Confidence Intervals using Variance
652 Confidence Intervals with Tolerance Probability
653 Confidence Intervals using Relative Error

Two Variances

655 Inequality Tests
656 Confidence Intervals for the Ratio of Two Variances using Variances
657 Confidence Intervals for the Ratio of Two Variances using Relative Error

Normality Tests

670 Normality Tests (Simulation)

Survival Analysis

700 Logrank Tests (Freedman)
705 Logrank Tests (Lachin and Foulkes)
706 Logrank Tests for Non-Inferiority
710 Group-Sequential Logrank Tests
711 Group-Sequential Logrank Tests (Simulation)
715 Logrank Tests (Lakatos)

Correlations

800 Inequality Tests for One Correlation
801 Confidence Intervals for One Correlation
805 Inequality Tests for Two Correlations
810 Inequality Tests for Intraclass Correlation
811 Kappa Test for Agreement Between Two Raters
815 Inequality Tests for One Coefficient Alpha
820 Inequality Tests for Two Coefficient Alphas

Regression

850 Cox Regression
855 Linear Regression
856 Confidence Intervals for Linear Regression Slope
860 Logistic Regression
865 Multiple Regression
870 Poisson Regression

User's Guide III
Table of Contents (Continued)

Design of Experiments

880 Randomization Lists
881 Two-Level Designs
882 Fractional Factorial Designs
883 Balanced Incomplete Block Designs
884 Latin Square Designs
885 Response Surface Designs
886 Screening Designs
887 Taguchi Designs
888 D-Optimal Designs
889 Design Generator

Tools, Helps, and Aids

900 Chi-Square Effect Size Estimator
903 Survival Parameter Conversion Tool
905 Standard Deviation Estimator
910 Odds Ratio and Proportions Estimator
912 Standard Deviation of Means Estimator
915 Probability Calculator
920 Data Simulator
925 Spreadsheets
930 Macros

Plots

940 Scatter Plots
945 Histograms

References and Index

References
Index

Chapter 1

Installation

Before You Install

Minimum System Requirements

In order to run **PASS**, your computer must meet the following minimum standards.

Processor:	450 MHz or faster processor 32-bit (x86) or 64-bit (x64) processor
RAM:	256 MB (512 MB recommended)
Operating Systems:	Windows XP with Service Pack 2 or higher Windows Vista Windows 7 or later Windows Server 2003 Windows Server 2008 Windows Server 2008 R2 or later
Privileges:	Administrative rights required during installation only
Third Party Software:	Microsoft .NET 3.5 SP1 (included with **PASS** CD, comes pre-installed with Windows 7 and Windows Server 2008 R2) or higher Microsoft Windows Installer 3.1 or higher Adobe Reader 7 or higher (required for the Help System only)
Hard Disk Space:	120 MB for **PASS** (plus space for Microsoft .NET 3.5 SP1 if not already installed)
Printer:	Any Windows-compatible inkjet or laser printer

Installing PASS

The installation procedure creates the necessary folders and copies the **PASS** program from the installation file. The files in the installation file are compressed, so the installation program decompresses these files as it copies them to your hard disk. During the installation wizard dialog, you will be able to choose the **PASS** installation location.

This section gives instructions for installing **PASS** on your computer system. You must use the **PASS** setup program to install **PASS**. The files are compressed, so you cannot simply copy the files to your hard drive.

Follow these basic steps to install **PASS** onto your computer system:

1. If you are installing from a CD, insert the CD in the disc drive on your computer. The installation program should start automatically. If it does not, use your file browser to explore the CD, locate the executable **PASS** setup file (ends with .exe), and run the file to start the installation. If you are installing from a download, simply run the downloaded installation file.
2. Once the setup starts, follow the instructions on the screen. **PASS** will be installed to the drive and folder you designate during the installation process.

If Something Goes Wrong During Installation

The installation procedure is automatic. If something goes wrong during installation, restart the installation procedure from the beginning. It may help to restart your computer before retrying the installation. If trouble persists, contact our technical support staff.

Starting PASS

PASS may be started using your keyboard or your mouse using the same techniques that you use to start any other Windows application. You can start **PASS** from your Start menu using standard mouse or keyboard operations. The installation procedure will also place a shortcut to **PASS** on your desktop for your convenience.

The first time you run **PASS**, enter your serial number in the pop-up window that appears when the program begins. After entering a serial number, the PASS Home window will appear.

Entering Your Serial Number

The first time you run **PASS**, enter your serial number in the pop-up window that appears. If you do not enter a serial number, you will be able to navigate among **PASS** windows, but none of the procedures will run and you will not be able to obtain output. If you would like to evaluate **PASS** procedures and reports, you must first obtain an evaluation serial number from NCSS. For more information about starting your trial or purchasing a license, visit http://www.ncss.com.

While in trial mode, **PASS** is fully-functional and gives you access to all procedures, reports, and graphics. You can enter a serial number anytime by going to the **Help** menu and choosing **Registration (Serial Number)**. Once you enter a purchased license serial number, the trial version will become the professional version. You do not need to reinstall **PASS**.

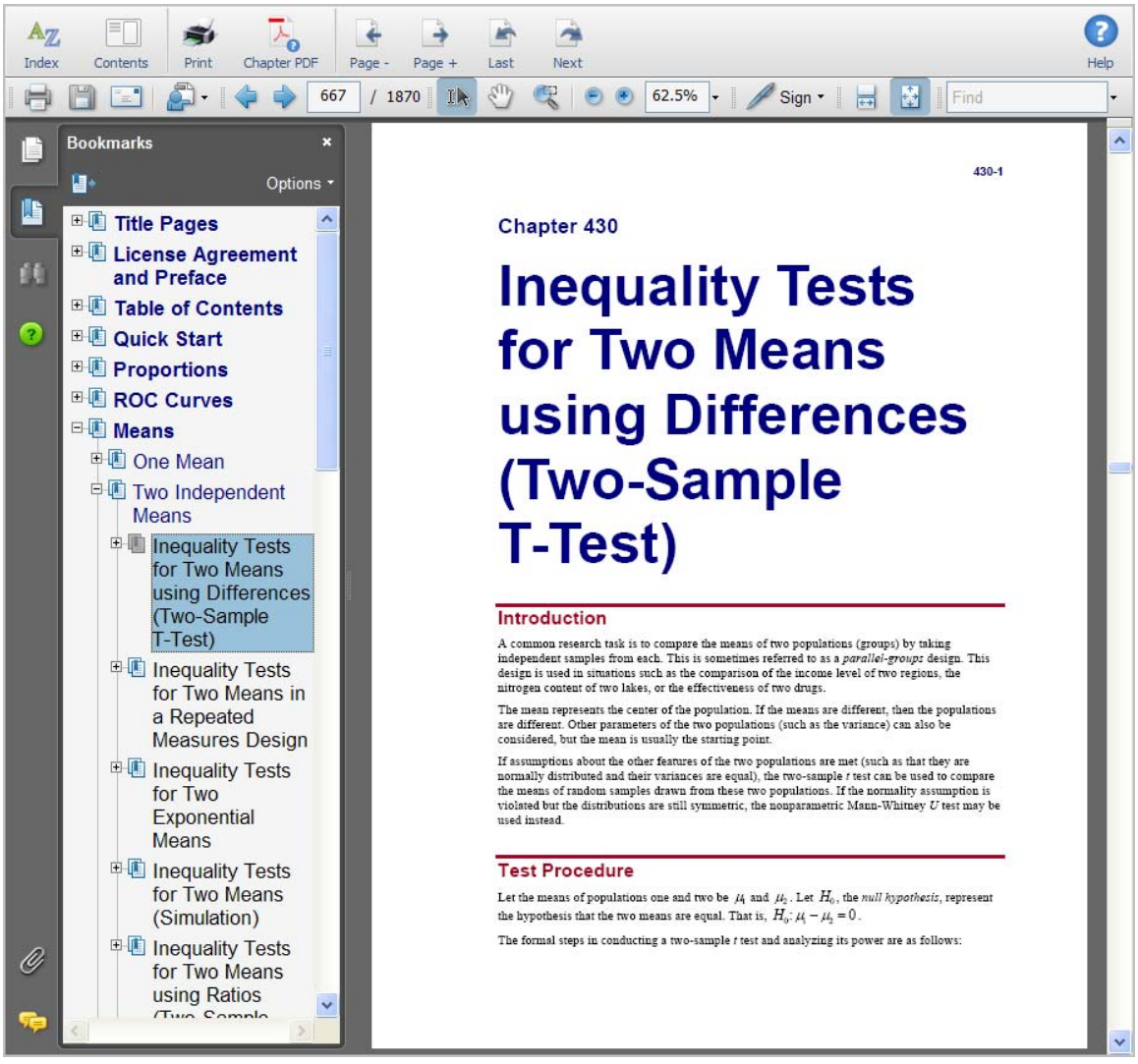
Updating PASS

NCSS periodically releases updates that fix bugs or provide enhancements to your current version of **PASS**. These updates are provided with all purchased licenses, free of additional charge. You can view the problems that each update fixes and download updates from our website, http://www.ncss.com. To determine whether you have the most recent update, click on **About** from the **Help** menu to find the current version and release date of your installed software. Compare this information to that of the updates available at our website.

Obtaining Help

The PASS Help System

To help you learn and use **PASS** efficiently, the material in this manual is included in the **PASS** Help System. The Help System is started from the Help menu or by clicking on the help icon on the right side of the toolbar. Adobe Acrobat or Adobe Reader version 7 or later is required to view the help system. You can download Adobe Reader for free by going to http://www.adobe.com. An Adobe Reader setup file is included in the *Utilities* folder on your **PASS** installation CD.

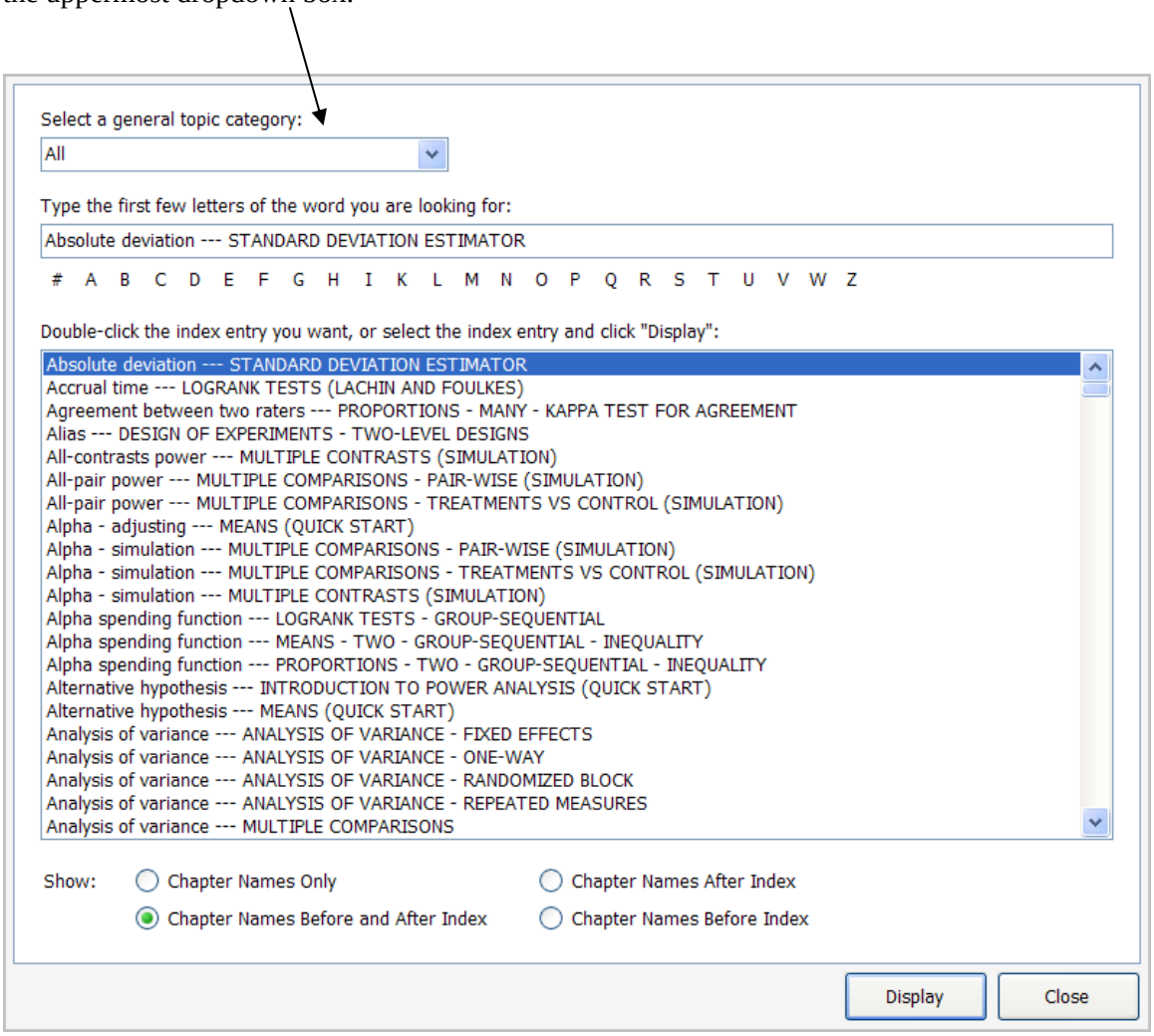


Navigating the Help System

There are a few key features of our help system that will let you use the help system more efficiently. We will now explain each of these features.

Index Window

The Index Window can be launched at any time by clicking on the Index button on the **PASS** Help System display window. The index allows you to quickly locate keywords and/or statistical topics. You can narrow the list of index entries displayed by selecting a specific topic category in the uppermost dropdown box.



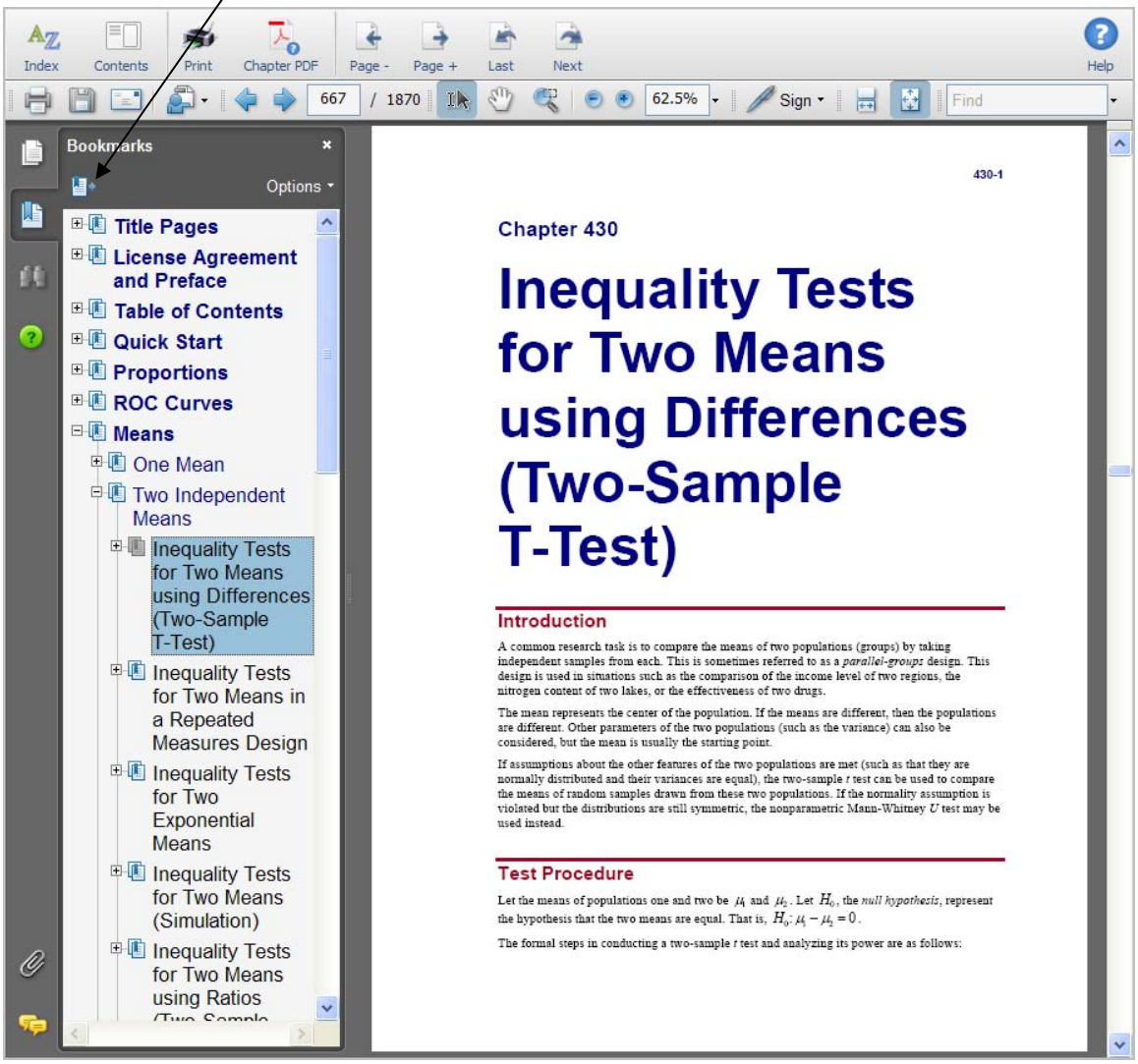
Index entries are displayed in the format

Index Entry --- CHAPTER or CHAPTER --- Index Entry.

You can control which entries are displayed by clicking on the radio buttons at the bottom of the window.

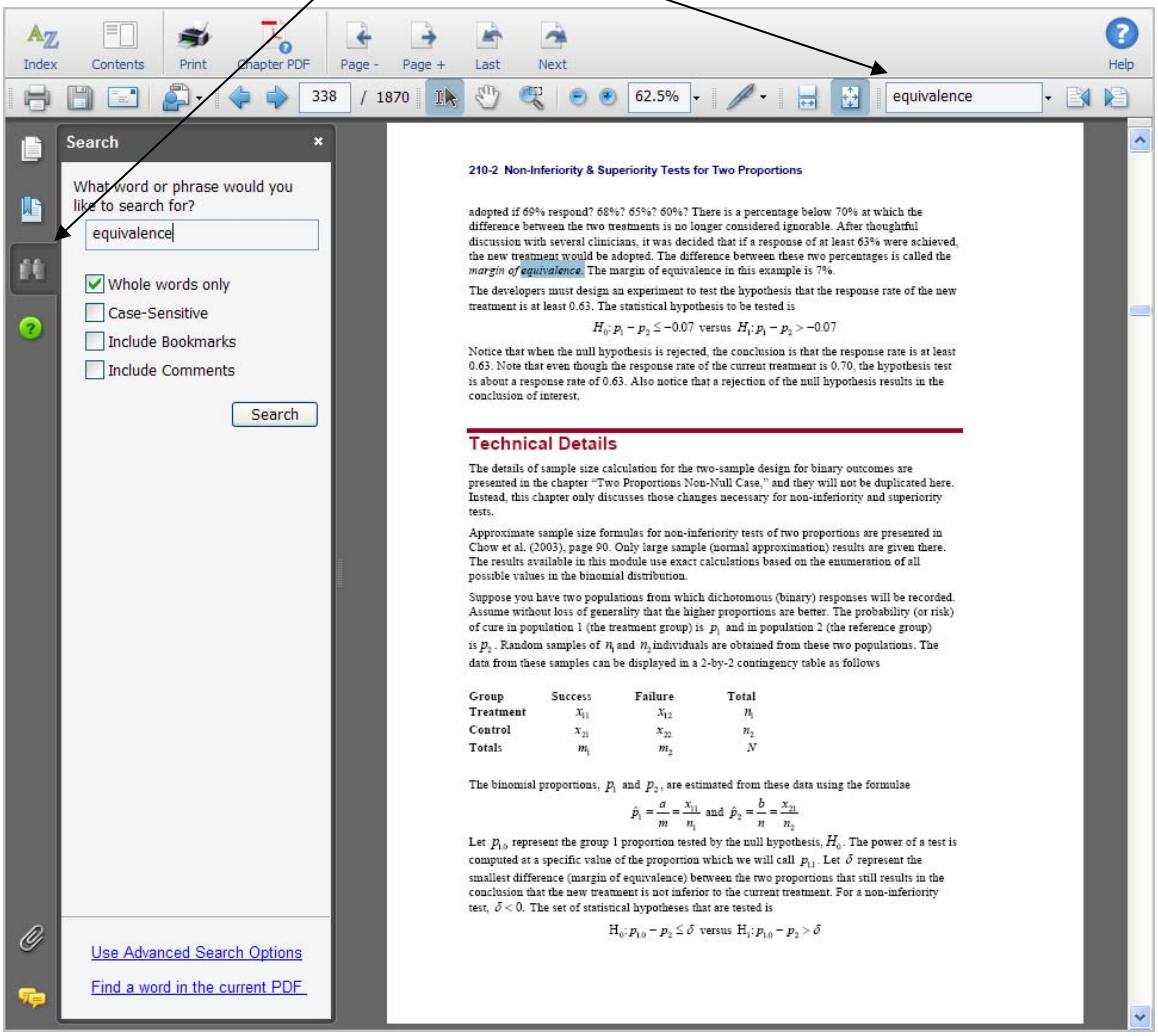
Contents Window

Clicking on the Contents button opens the Contents (Bookmarks) Window of the viewer. From this window you can expand the table of contents to view nested headings. You can click on the "Expand Current Bookmark" icon to instantly find the bookmark location for the currently-displayed page in the help document.



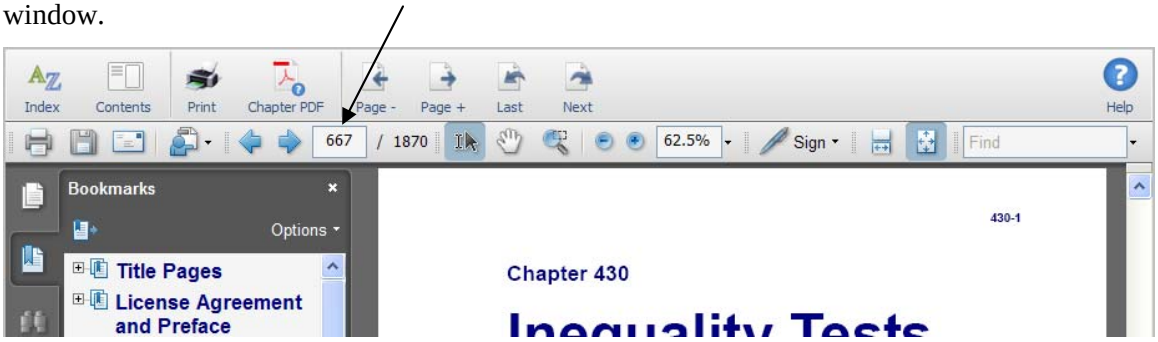
Search Window

To initiate a search of the help system for any word or phrase, type your search word(s) in the "Find" box in the viewer or click on the "Find" tab icon and enter a search term.



Printing the Documentation

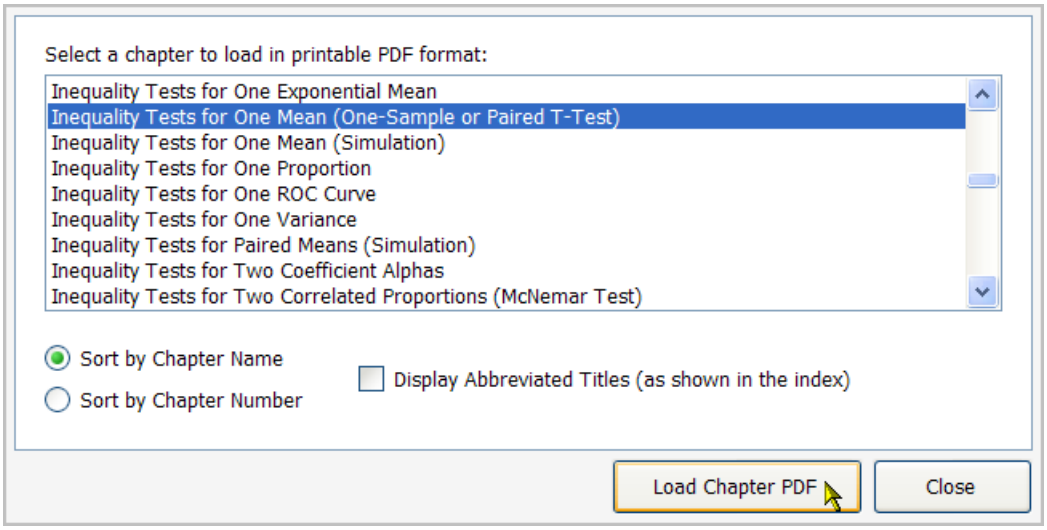
To print pages from the documentation, click on the **Print** button on the **PASS** Help System toolbar. This will launch the print dialogue screen. You can choose to print a single page or a range of pages from the help file. When entering page numbers, remember to use the actual PDF file page numbers (e.g., 667-670) and not the page numbers found in the document pages (e.g., 430-1 to 430-4 is not a valid page range). The actual page numbers can be seen in the viewer window.



One of the benefits of the **PASS** Help System is the ease with which you can print any chapter or topic from the electronic help manual. To print a single chapter or topic using your default PDF viewer, take the following steps:

1. Click on the **Chapter PDF** icon in the **PASS** Help System toolbar.

2. Choose the chapter you would like to print from the list and click **Load Chapter PDF**. This will launch the individual chapter PDF in a separate window using your default PDF viewer (e.g., Adobe Reader).



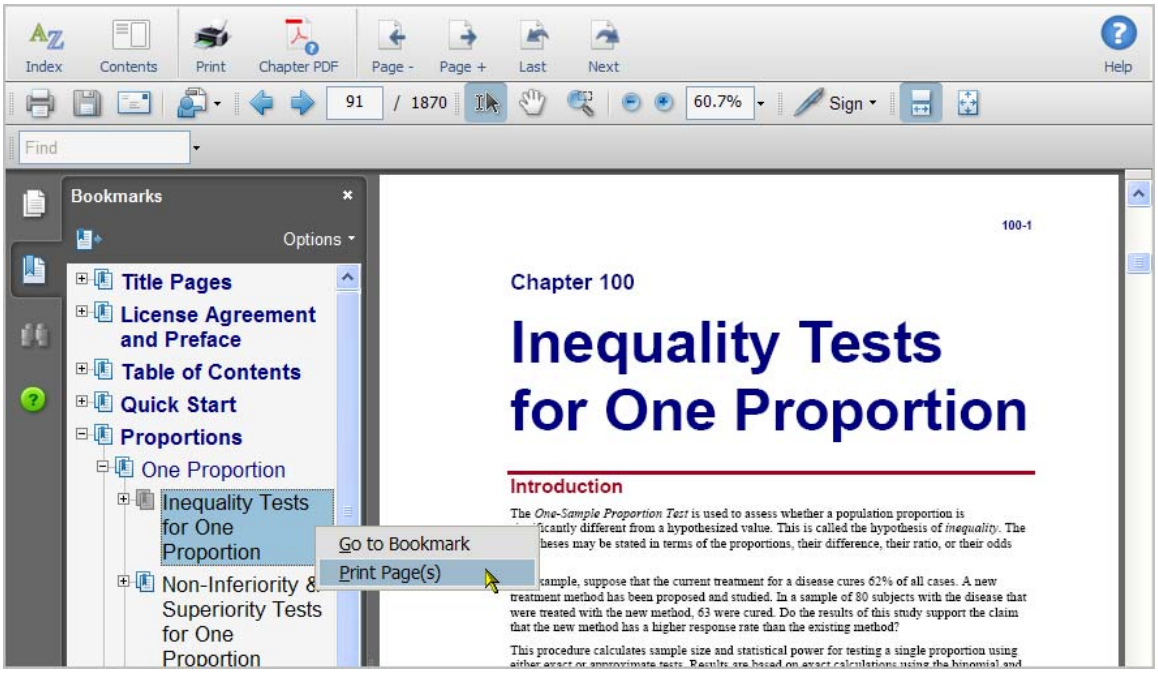
3. Use the **Print** function of your PDF viewer to print the entire chapter or individual pages from the chapter.

If you have Adobe Reader 8 or later, you can print entire chapters using an alternative method as follows (**This will not work with Adobe Reader 7**):

1. Open the Contents (Bookmarks) Window by clicking on the **Contents button** at the top of the **PASS** Help System display window.

2. Expand the bookmarks to display the chapter or topic name you wish to print (e.g., the Two-Sample T-Test Chapter). Then, **highlight** the chapter name, **right-click** on the highlighted selection (or select Options in the panel above), and select **Print Page(s)**. This will automatically print only the pages from the selected chapter.

   CAUTION: When you click Print Page(s), the command is sent to the printer automatically without any intermediate Print Setup window being displayed. Make sure that you have selected only the topic you want before clicking Print Page(s).



If you do not want to print the entire chapter, continue to expand the bookmark tree to the topic you wish to print before completing step 2. The Print Page(s) command prints all pages containing bookmarks that are nested within the highlighted bookmark.

Technical Support

If you have a question about **PASS**, you should first look to the printed documentation and the included Help system. If you cannot find the answer there, look for help on the web at www.ncss.com/support.html. If you are unable to find the answer to your question by these means, contact NCSS technical support for assistance by calling (801) 546-0445 between 8 a.m. and 5 p.m. (MST). You can contact us by email at support@ncss.com or by fax at (801) 546-3907. Our technical support staff will help you with your question.

If you encounter problems or errors while using **PASS**, please view our list of recent corrections before calling by going to our website to find out if you problem or error has been corrected by an update. If updating your software does not correct the problem, contact us by phone or email.

To help us answer your questions more accurately, we may need to know about your computer system. Please have pertinent information about your computer and operating system available. You can obtain system information from the **Help** menu by choosing **About** and then clicking on the **System Info** button.

Chapter 2

Running PASS

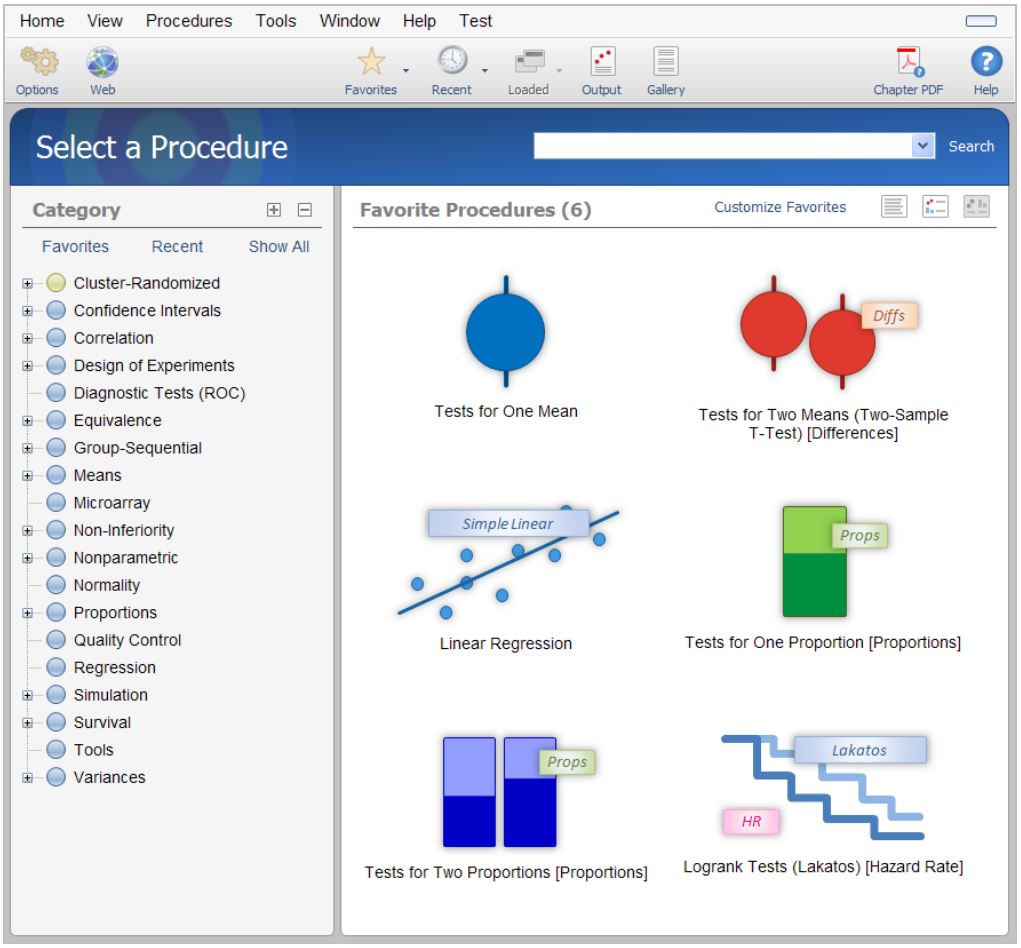
About This Chapter

This chapter will show you how to start up and run a power analysis of the two-sample *t* test. It will give you a brief introduction to the windows used in **PASS**: the *PASS Home* window, the *Procedure* window, and the *Output* window.

Starting PASS

To start **PASS**, select the shortcut from the Windows Start menu or double-click the shortcut icon on your desktop. If you are licensed for **PASS**, the following PASS Home window will appear.

PASS Home

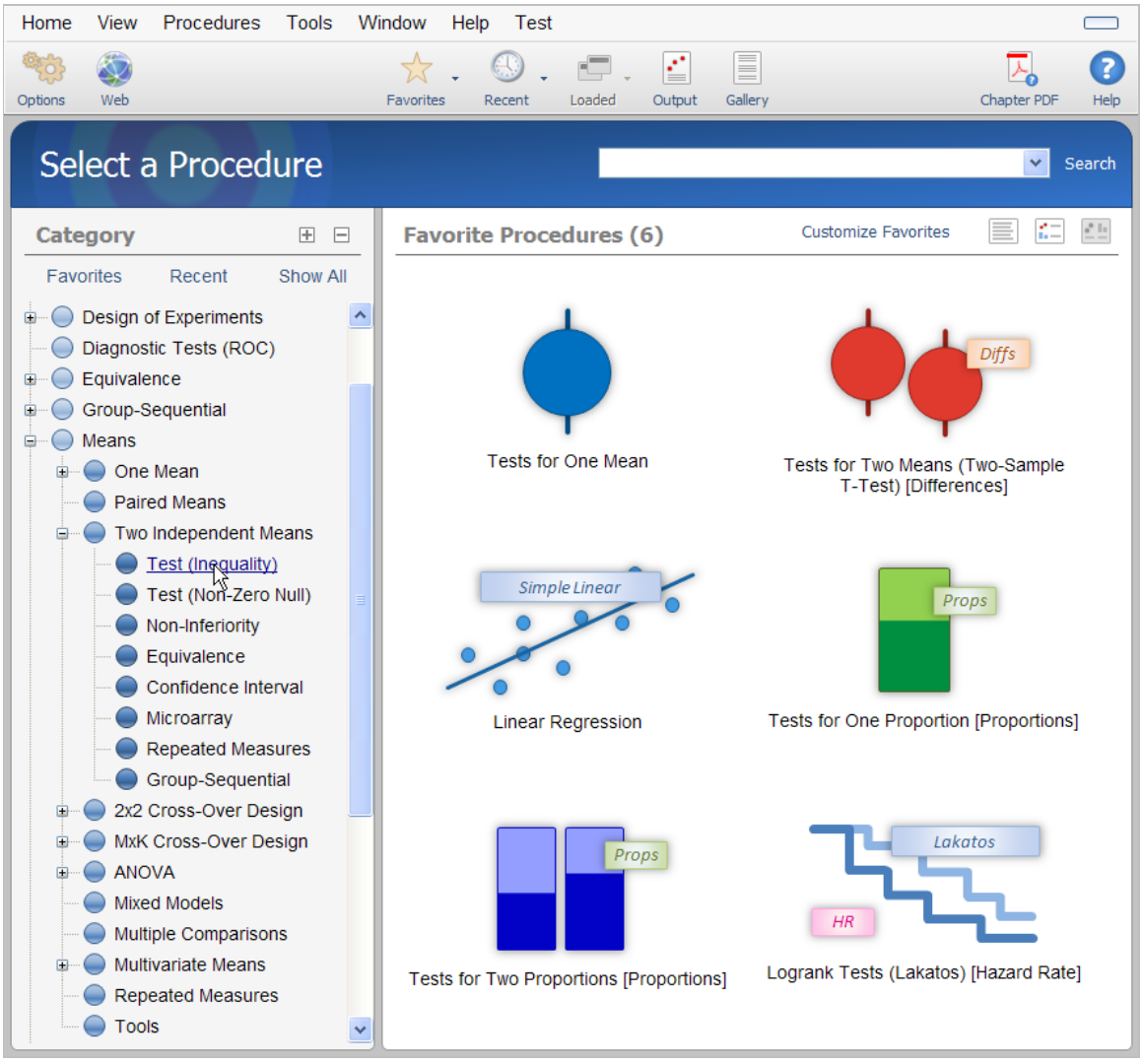


This window gives you access to all of the **PASS** procedures. Clicking on the plus sign or double-clicking on a phrase will expand the list so that you can see the procedures in that group. To load a specific procedure window, double-click on it or highlight it and click the Load button.

Loading a Procedure

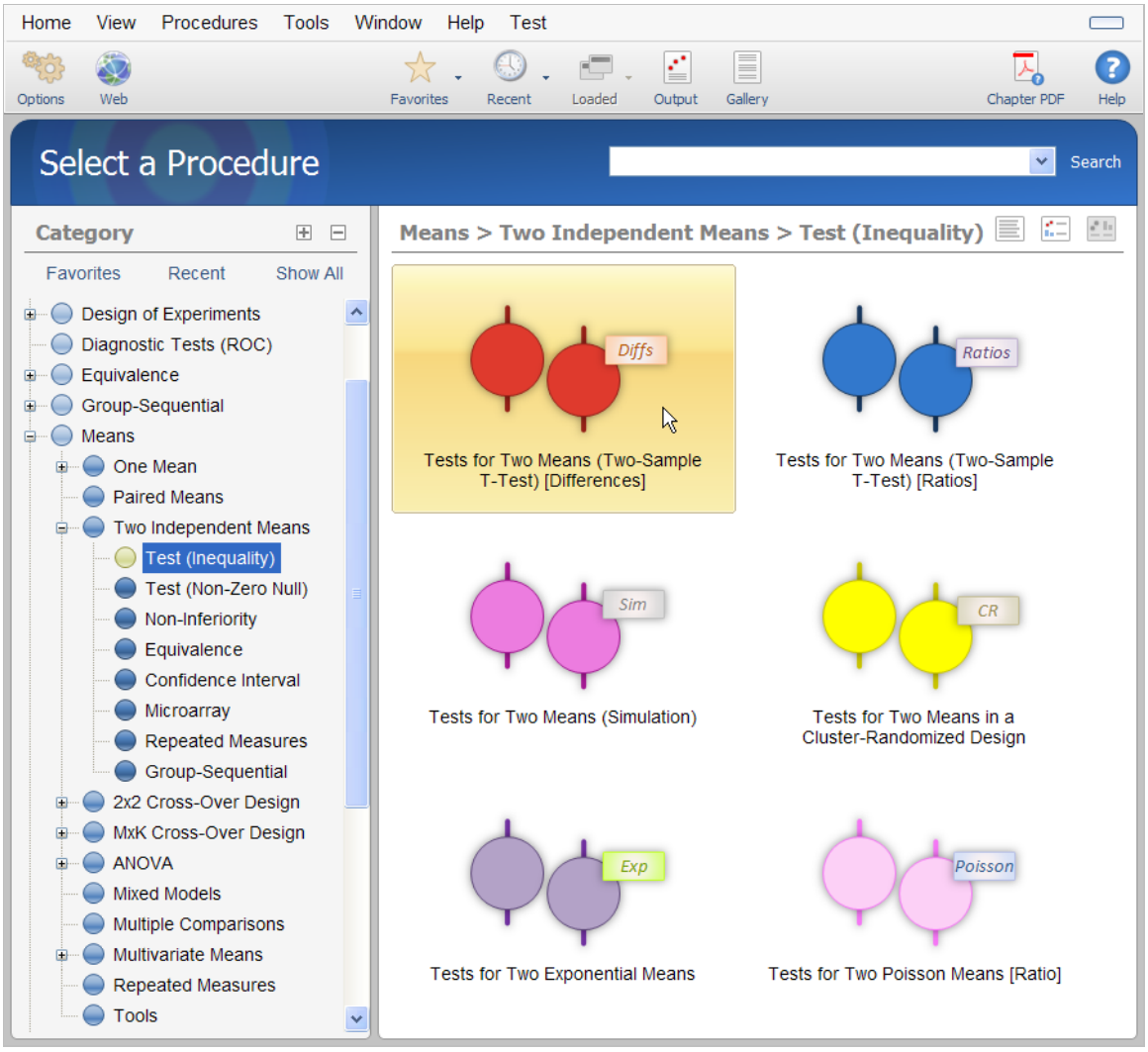
The *Two-Sample T-test* is a procedure to test the inequality of two means from independent samples. Take the following steps to load this procedure. Expand the Means clicking the + next to the word **Means**. Then expand **Two Independent Means**, and then click on **Test (Inequality)**.

PASS Home



The top left button is **Tests for Two Means (Two-Sample T-Test) [Differences]**. Click this button.

PASS Home

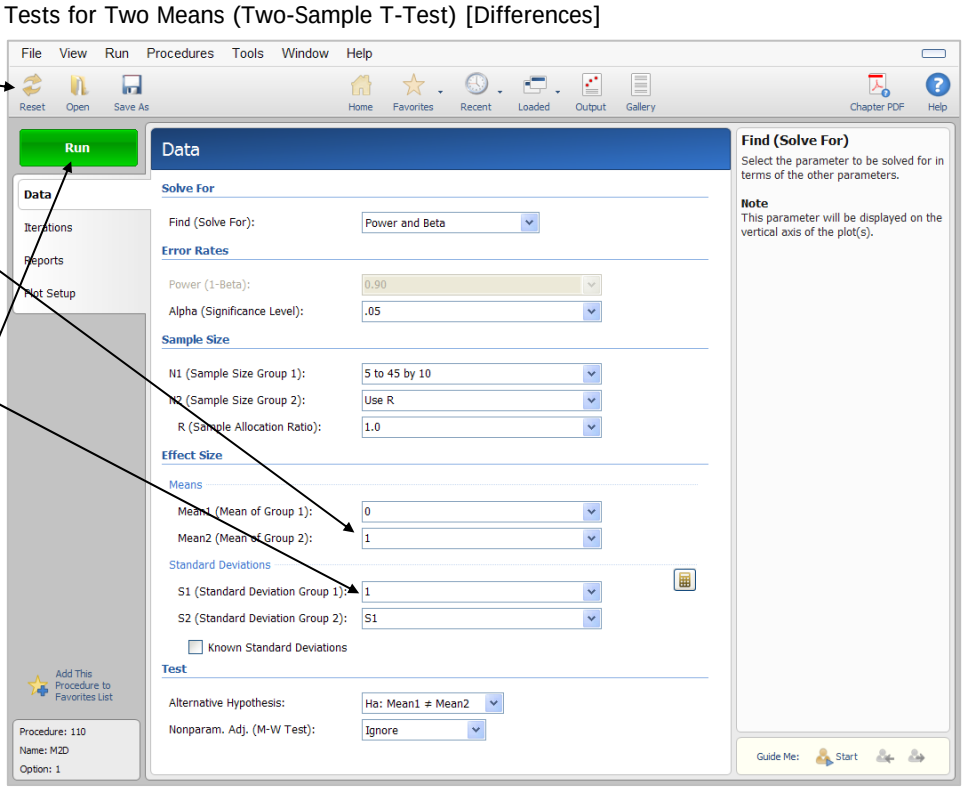


The **Tests for Two Means (Two-Sample T-Test) [Differences]** procedure window will appear. Procedure windows let you specify parameters, load and save templates, and run the analyses.

Entering Parameters and Running the Procedure

We will run a power analysis using the default values except that the value of **Mean2** will be **2** and the value of **S1** will be **3**.

1. Click the **Reset** button to set all options to their default values.
2. Click in the box next to **Mean2**.
3. Enter **2**.
4. Click in the box next to **S1**.
5. Enter **3**.
6. Click the **Run** button to run the procedure

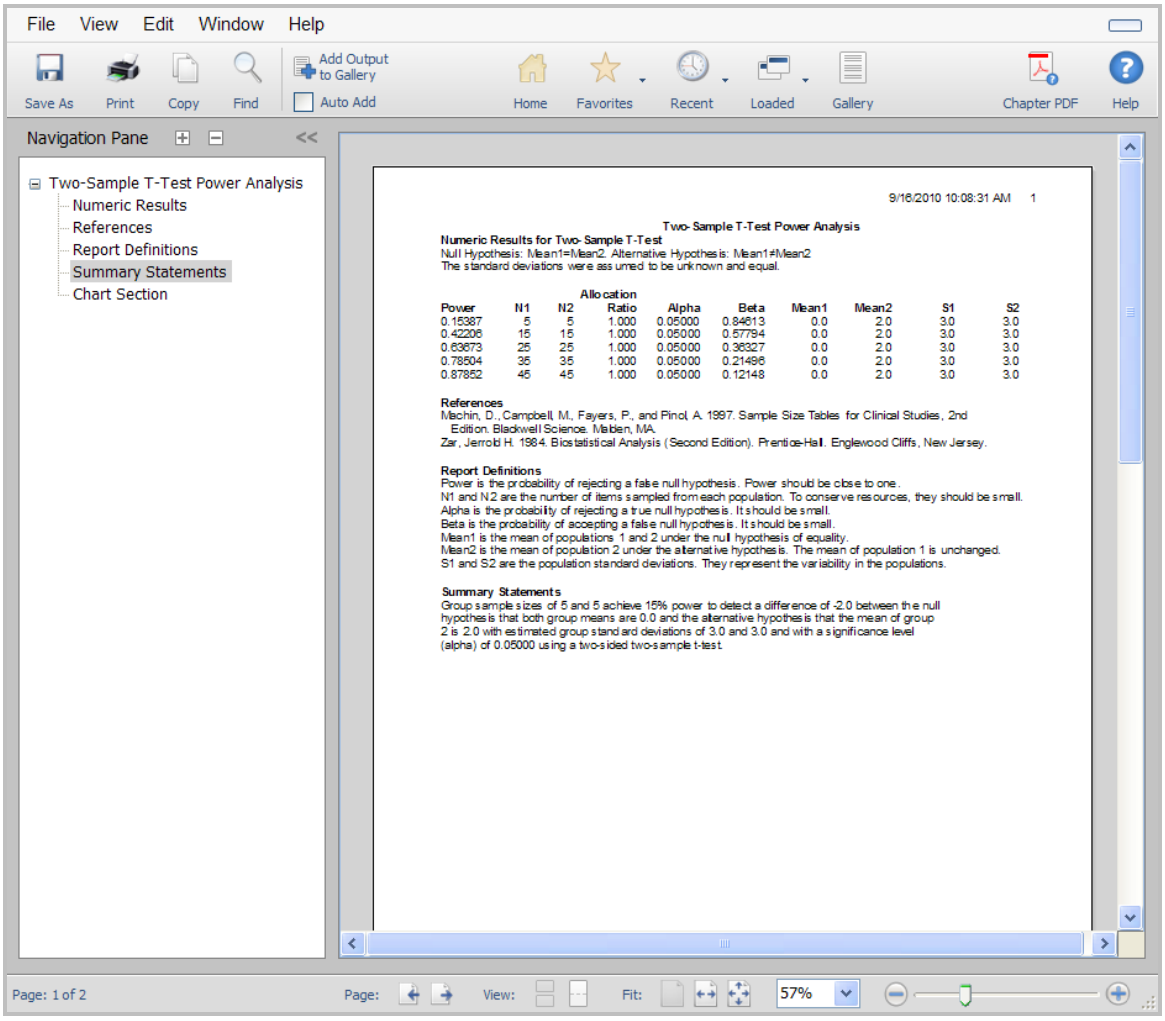


Viewing the Output

The Output window displays the output of the power analysis. It allows you to view, edit, save, and print your output. You may want to scroll down to view the graph at the end of the report.

When you are finished, you can quit **PASS** by selecting **Exit PASS** from the Home menu.

PASS Output



2-6 Quick Start – Running PASS

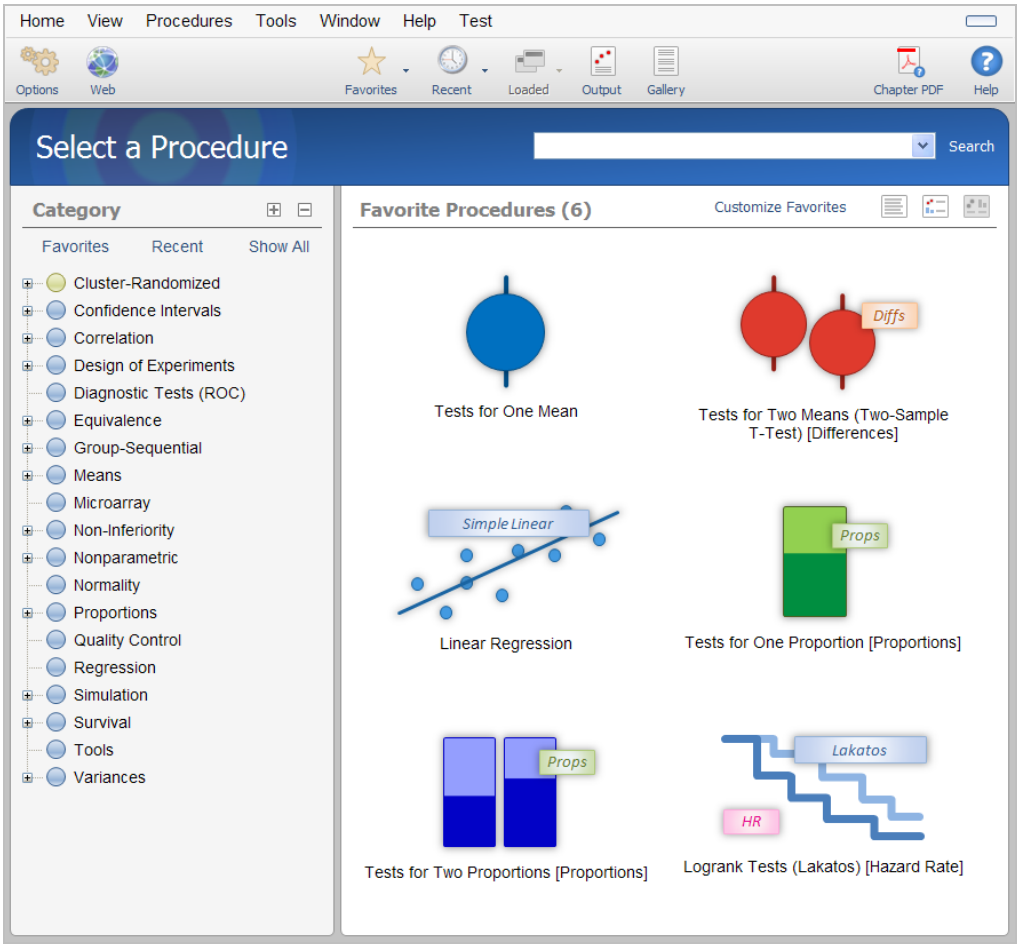
Chapter 3

The PASS Home Window

Introduction

The **PASS** Home window lets you quickly and easily find the appropriate procedure to be loaded.

PASS Home



Procedure categories and subcategories are given on the left. Clicking on any category or subcategory gives a list of all procedures in that category or subcategory. The list of procedures can be shown as large buttons, medium tile buttons, or small list buttons. Favorite procedures and

recently run procedures can be shown by clicking on the respective buttons. Procedures may also be found by entering keywords and clicking on Search.

Using the PASS Home Window

The **PASS** Home window was designed to be easy to use. The window has a set of menus, a toolbar, and a large display area. On the left side of the display area is an outline list of the major categories of procedures in **PASS**. The right side displays the procedures corresponding to the chosen category or the results of the search.

Menus

Most of the selections available in the menus are also available through the toolbar buttons and other buttons.

Home Menu

This menu is used to view or change system options or to exit **PASS**.

Options

A System Options window is opened. This window gives general options about the appearance of the reports, calculation constants, default spreadsheet numbers, whether power or beta is the primary input, and other minor options.

Exit PASS

This option closes the PASS Home window and exits **PASS**.

View Menu

This menu is used to change the appearance of the procedure buttons or to expand and collapse the tree of procedure categories.

Procedures Menu

The procedure menu allows you to quickly find and load **PASS** power analysis procedures.

Tools Menu

The tools menu contains links to various **PASS** utilities.

Window Menu

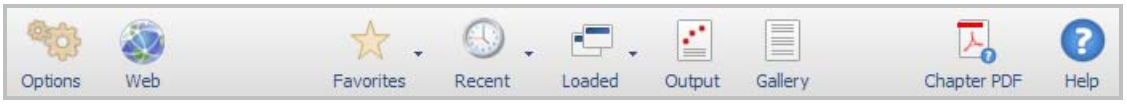
This menu allows you to open other windows in the **PASS** system such as the the Output window, the Output Spreadsheet, or recent, favorite, or currently loaded procedures.

Help Menu

This menu allows you to view the **PASS** Help System, modify your serial number, get information about the program and load various portions of the printable PDF documentation.

Toolbar

The toolbar gives you one-click access to several of the menu items. Notes about each button are displayed when the mouse is held over the button for a few seconds.



Options. This button opens the System Options window.

Web. This button sends the user to the support web site.

Favorites. This drop-down button shows a list of procedures that have been added to the favorites list.

Recent. This drop-down button shows a list of procedures that have been recently run.

Loaded. This drop-down button shows a list of procedures that are currently loaded.

Output. This takes the user to the primary output window.

Gallery. This takes the user to window with output that has been sent from the primary output window.

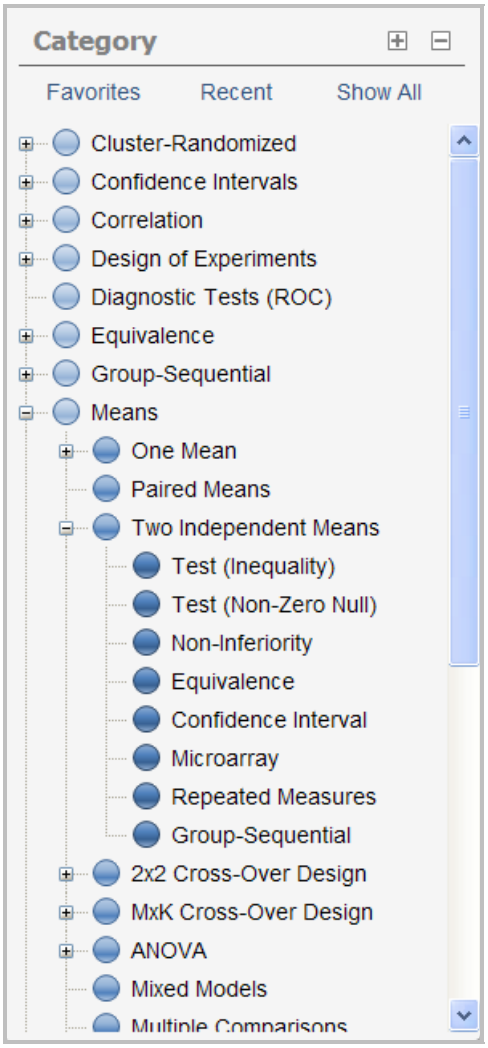
Chapter PDF. This opens the PASS Home Window chapter (this chapter) as an individual PDF.

Help. This opens the PASS Help System and goes to the current topic.

Category Tree

The Category Tree expands and contracts as you either click on a plus or minus sign at the left. This gives you quick, intuitive access to all of the procedures in **PASS**. Procedures are not selected here directly, but instead a category is chosen. When you click on a category or subcategory, all the procedures corresponding to that category or subcategory are displayed to the right.

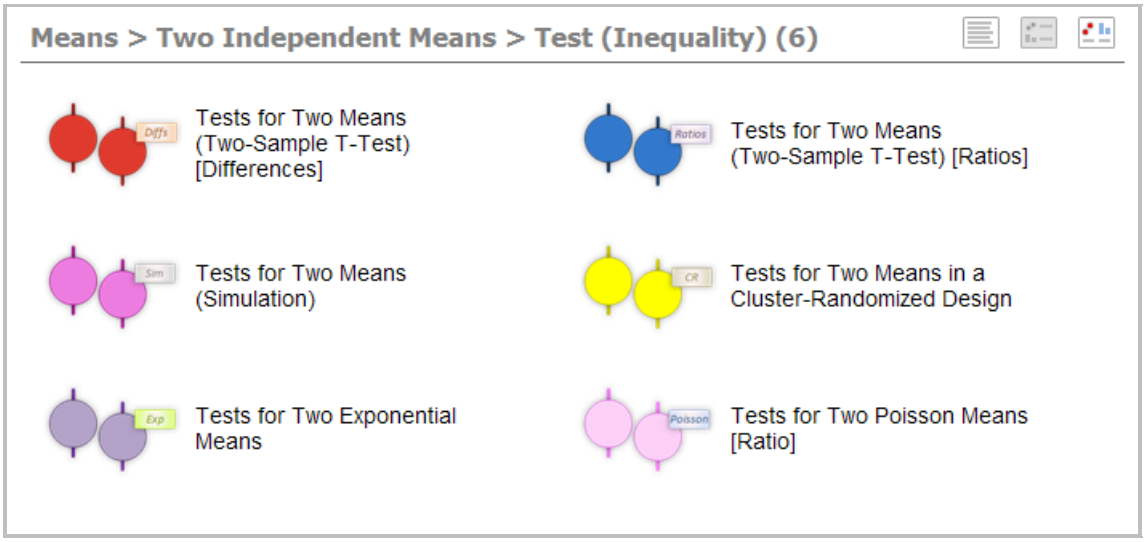
The plus and minus buttons at the top right can be used to expand and contract the entire tree, respectively.



The Favorites, Recent, and Show All buttons can be used to display the corresponding procedures in the box on the right.

Procedure List

The procedure list shows all the procedures corresponding to the selected category or all the procedures with keywords matching those of the search. The procedure list below is the result of clicking on the Test (Inequality) node of the Two Independent Means branch of the Means root of the Category Tree.



The number in parentheses shows the number of procedures in the group. The three buttons on the top right are used to change the view of the list of procedures. The left button gives a list, the middle button gives tiles (above), and the right button gives large buttons.

The procedure is loaded by a single-click on the procedure button. The user can right-click the procedure button to load the procedure, add or remove the procedure to/from the favorites list, or to view the PDF documentation for the procedure.

Search for a Procedure

The search bar is used to find a procedure. The program contains a list of key words for each procedure. As you type a key word in the search bar, a list of words will be generated below the bar. The down arrow or the mouse can be used to select these words, or the word(s) can be typed.

When the desired search words are in the box, the Search button is pressed to generate a list of procedures corresponding to the words. These procedures can be viewed or opened as described in the Procedure List section above.

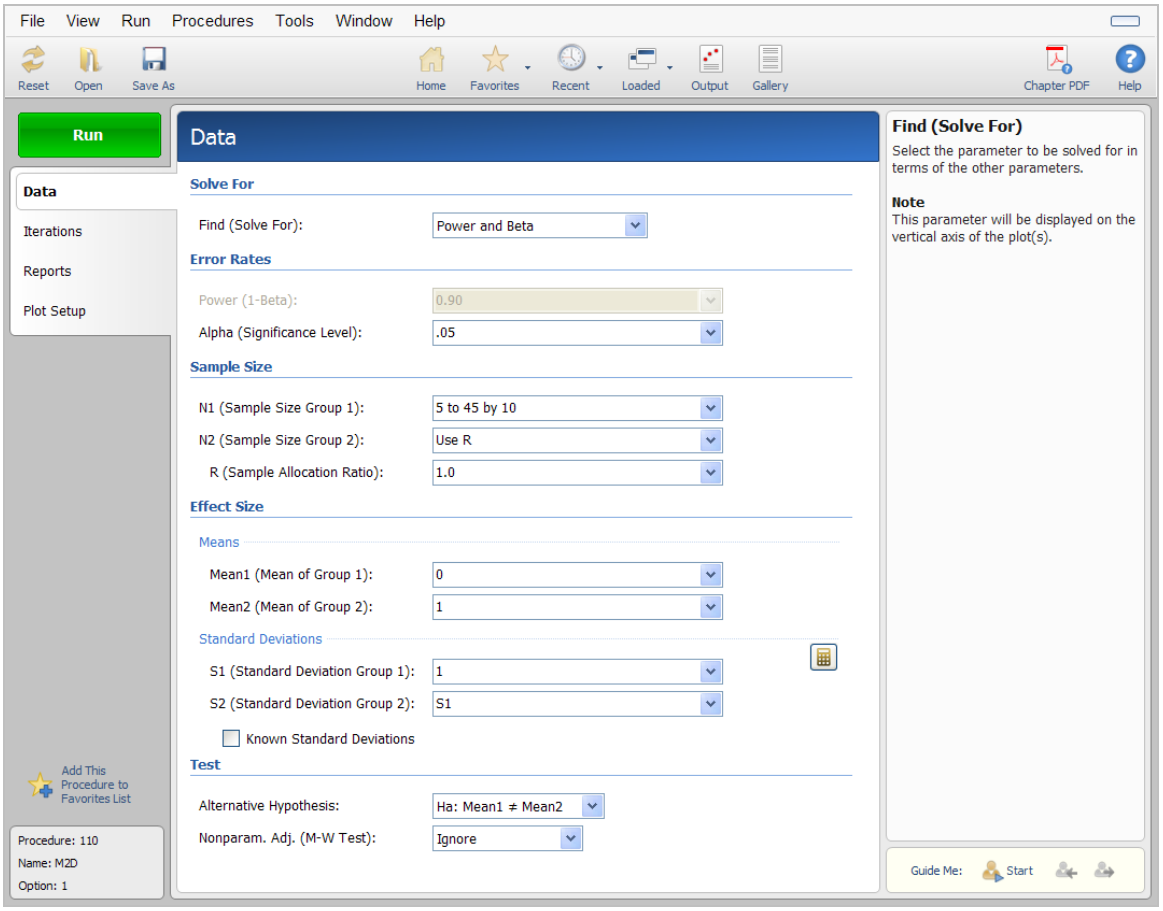
Chapter 4

The Procedure Window

Introduction

All **PASS** procedures are controlled by a *procedure* window that contains all of the settings, options, and parameters that determine the input and output of the program. These options are separated into groups called *panels* or *tabs*. A particular panel is viewed by pressing the corresponding tab that appears at the left side of the window. For example, in the window below, the Data panel is active. Other common panels include Reports and Plot Setup.

Tests for Two Means (Two-Sample T-Test) [Differences]



The values of all options available for a procedure are referred to as a *template*. A template may be stored for future use in a *template file*. By creating and saving template files (often referred to as *templates*), you can tailor each procedure to your own specific needs. Each time you use a procedure, you simply load your template and run the analysis you have preset. You do not have to set all the options every time. The specific operations needed to do this are shown later.

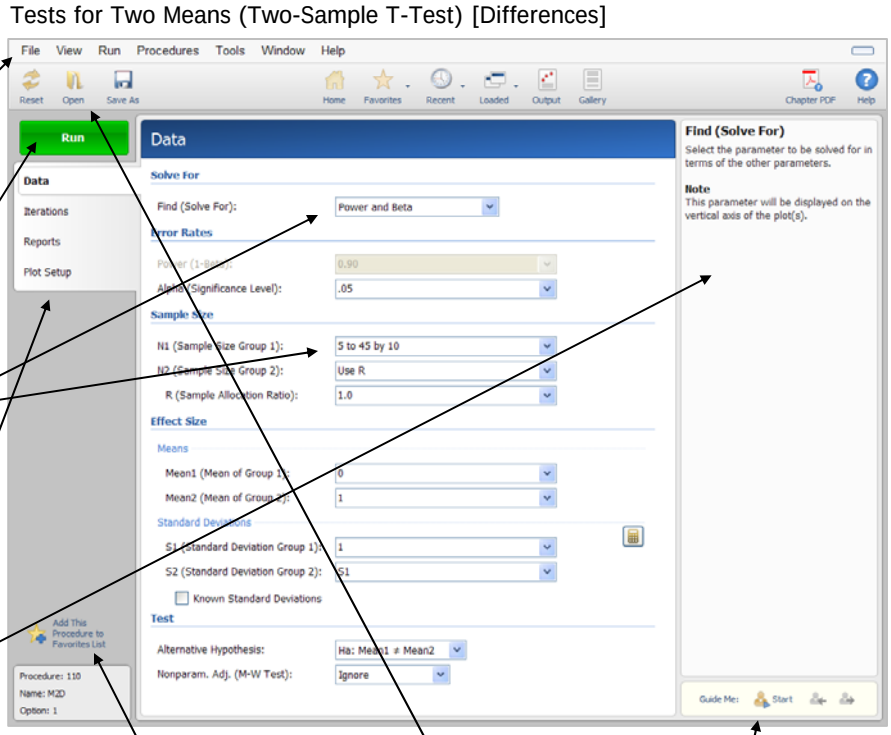
At most six procedure windows can be opened at a time. You can widen the window to increase the size of the immediate help window by dragging the corners of the window.

Default Template

Whenever you close a procedure, the current settings are automatically saved in a default template file. This template file is automatically loaded when the procedure is next opened.

Procedure Window Anatomy

This section explains the various objects found on the procedure window.



Menus. The menus allow you to open/ save templates, go to other windows, get help, etc.

Run. Clicking this button runs the program and generates output.

Options. These fields set values that control the analysis.

Tabs. These let you view different groups of options analysis.

Immediate Help. This box displays a brief help message about the field over which the mouse is currently positioned.

Add Procedure to Favorites. This button adds this procedure to the favorites list for easy access later.

Reset, Open, Save As. These buttons are used to set or save a group of options for the whole procedure at once. A group of saved options is called a template.

Guide Me. These buttons instruct the program to step you through the main options that must be set for an analysis.

Menus

The menus provide a convenient way to transfer from module to module within the ***PASS*** system. Each set of menus will be briefly described here.

File Menu

The File Menu is used for initializing, loading, and saving a file of a template. Each set of options for a procedure, called a *template*, may be saved for future use. In this way, you do not have to set the options every time you use a procedure. Instead, you set the options the first time, save them as a template, and re-use the template whenever you re-use the procedure.

New Template (Reset)

This menu item resets all options to their default values. It performs the same function as the Reset button.

Open Template

This option is used to open an existing template file that has been saved by the user.

Open Example Template

This option is used to open an example template file that corresponds to one of the examples at the end of each chapter.

Save Template As

This option is used to name and save a template file of the current options to the location of your choice.

Options

A System Options window is opened. This window gives general options about the appearance of the reports, calculation constants, default spreadsheet numbers, whether power or beta is the primary input, and other minor options.

Close Procedure

This option closes this procedure window.

Exit PASS

This option terminates the **PASS** system. Before using this option, you should save all templates and output documents that you want to keep.

View Menu

This menu controls two viewing aspects of the procedure window.

Show/Hide Toolbar

This option controls whether the toolbar is shown.

Show/Hide Procedure Info

This option controls whether a box is shown in the bottom left corner giving the procedure number, procedure short name, and the option number when each option is moused over.

Run Menu

This menu controls the execution of the program.

Run Procedure

The Run Procedure option runs the analysis, displaying the output in the Output document.. After you have set all options to their appropriate values, select this option to perform the analysis.

Note that the procedure may also be run by pressing the green run button.

Procedures and Tools Menus

These menus allow you to transfer to various **PASS** procedures and utilities.

Window Menu

This menu lets you display any of the other windows in the **PASS** system that are currently open such as the Output window, the Output Spreadsheet window, the PASS Home window, or favorite, recent, or currently open procedure windows.

PASS Home

Select this option to display the PASS Home window.

Output

Select this option to display the output window.

Output Gallery

Select this option to display the output gallery window.

Output Spreadsheet

Select this option to display the output spreadsheet.

Favorite Procedures

Select this option to display a list of the procedure that have been added to the favorites list.

Recent Procedures

Select this option to display a list of the recently run procedures.

Currently Loaded Procedures

The last menu options contain the procedures that are currently open.

Help Menu

This menu gives you access to the **PASS** Help System and PDF documentation, including references and the Quick Start Manual. This menu also gives the citation information for PASS and access to the website and other support information.

Toolbar

The toolbar is a series of buttons that appear just below the menus at the top of the procedure window. Each of these buttons provides quick access to a menu item.

Reset. This button sets all parameters to their default values.

Open. This button opens an open file window so you can load a saved template.

Save As. This button is used to name and save a template file of the current options to the location of your choice.

Home. This button brings the PASS Home Window to the front.

Favorites. This drop-down button shows a list of procedures that have been added to the favorites list.

Recent. This drop-down button shows a list of procedures that have been recently run.

Loaded. This drop-down button shows a list of procedures that are currently loaded.

Output. This takes the user to the primary output window.

Gallery. This takes the user to window with output that has been sent from the primary output window.

Chapter PDF. This opens the chapter for the current procedure as an individual PDF.

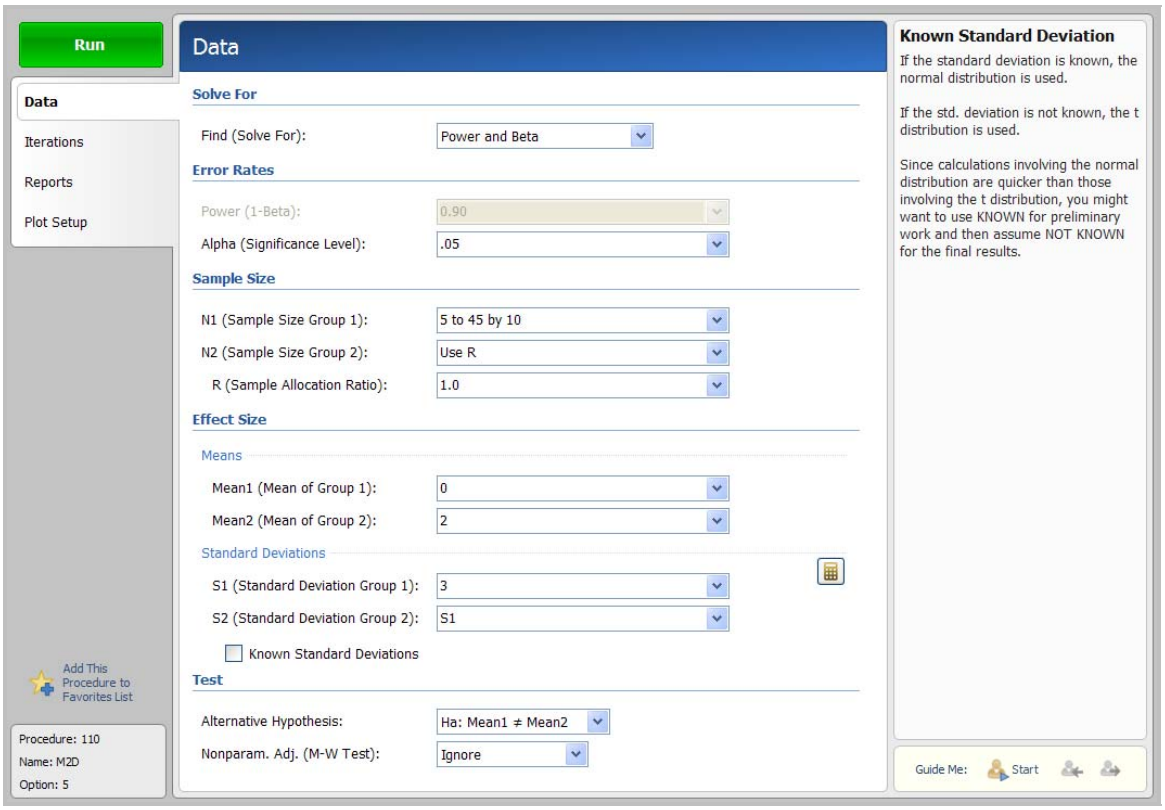
Help. This opens the PASS Help System and goes to the current topic.

Procedure Window Tabs

The procedure window contains several sets of options on panels (or tabs). Each panel is displayed by clicking on the appropriate tab. We will now describe the purpose and operation of each tab.

Data Tab

The Data tab displays most of the options specific to the procedure. This is where you set the values of power, sample size, alpha, etc. These options are described in detail in the chapters corresponding to each procedure. Once you have set the options, click the **Run** button to generate the output.



Entering Multiple Values

In most cases, boxes that are extra wide allow you to enter multiple values. When this is done, a separate analysis is done for each combination of all multiple values. For example, if you enter four sample sizes and three alpha values, the resulting report will contain $3 \times 4 = 12$ rows, one for each combination.

You can enter multiple options using list or the *to-by* syntax. The *to-by* syntax is most easily described by an example.

The *to-by* phrase *20 to 100 by 20* is translated to the values: *20 40 60 80 100*.

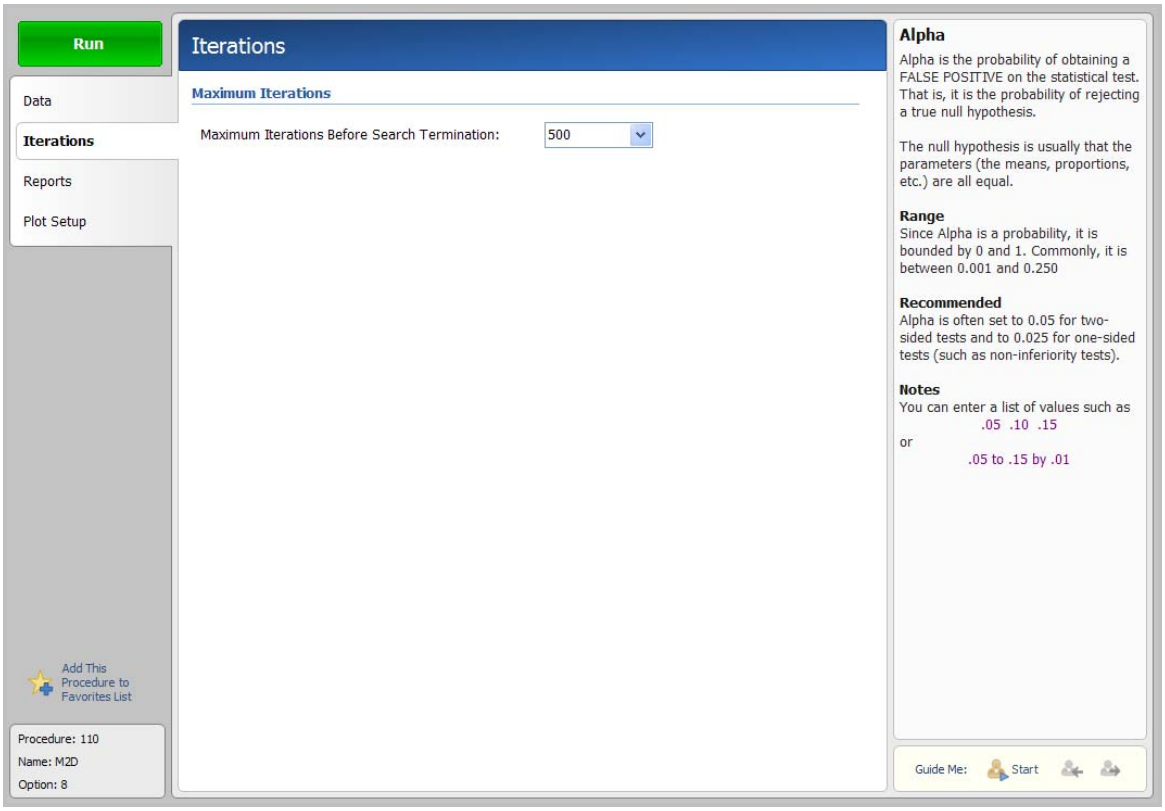
Find (Solve For)

Specify the parameter that is to be solved for in terms of the other parameters. For example, you might want to solve for power or sample size.

In most cases, the algorithm for the calculating the power is programmed within ***PASS***. When other parameters (such as sample size or difference) are selected, a binary search is conducted using the power algorithm. These searches can be time consuming, so the best place to start is with **Find (Solve For)** set to *Power and Beta*.

Iterations Tab

The Iterations tab presents the parameters that control the searching process.



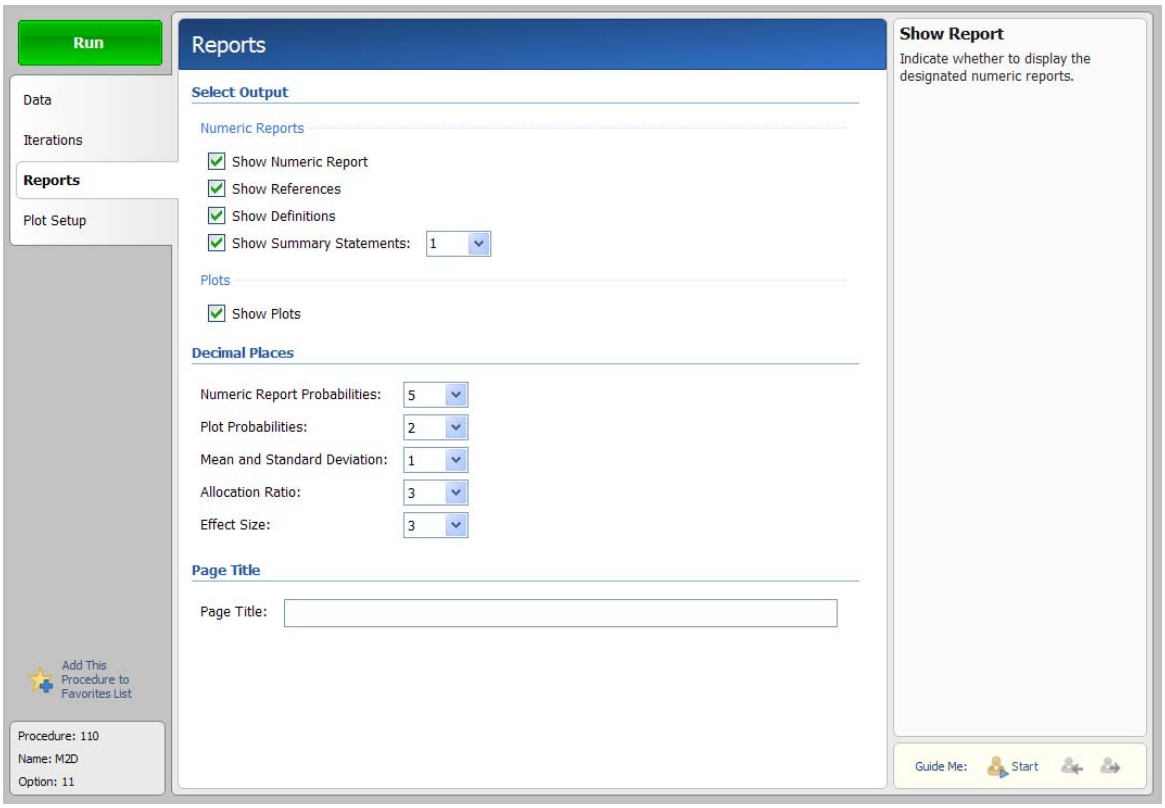
Maximum Iterations

Maximum Iterations Before Search Termination

This option specifies the maximum number of iterations before a search for the parameter of interest is halted. When the maximum number of iterations is reached without convergence, the criterion is left blank. We recommend that at least 500 iterations be specified.

Reports Tab

The Reports tab displays the options that control the output reports.



Select Output – Numeric Reports

Show Numeric Report

Determines whether the numeric report is displayed in the output.

Show References

Check this box to cause the literature reference(s) to be displayed on the report.

Show Definitions

Check this box to show the definitions at the end of the numeric report. Although these definitions are helpful at first, they tend to clutter the output and this option lets you skip them.

Show Summary Statements

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 Output – Plots

Show Plots

Check this box to display plots in the output.

Decimal Places

Decimals

These options set the number of decimal places in corresponding values of the numeric and graphic output.

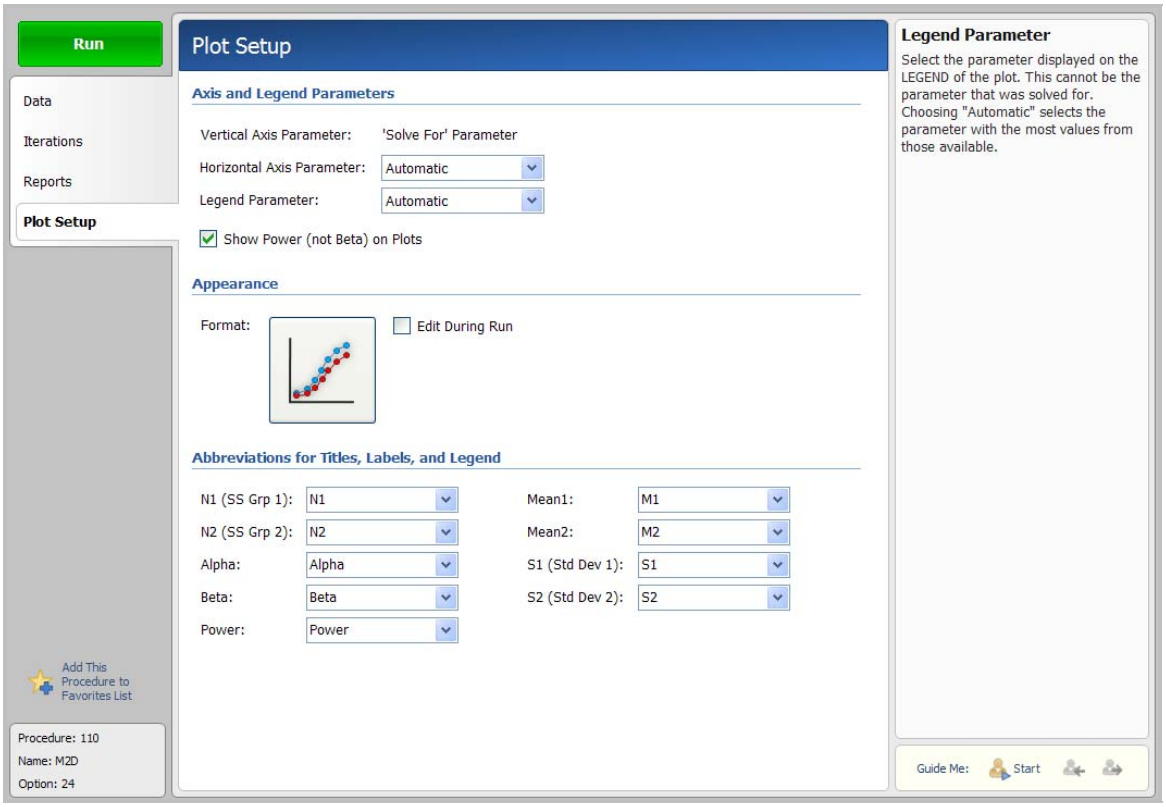
Page Title

Page Title

This option allows you to enter an option title phrase that will appear in the heading of each page of the output.

Plot Setup Tab

The Plot Setup tab displays the options that control the output plots.



Axis and Legend Parameters

Vertical and Horizontal Axis Parameters

This option selects which of the parameters from the Data tab is displayed on the horizontal axis. The vertical axis always contains the **Solve For** parameter, so you cannot select the parameter that was listed in the Find option. Also, you would normally only select a parameter that has multiple entries.

When this option is set to *Automatic*, the parameter with the most values is selected.

Legend Parameter

A separate line is drawn for each value of this parameter. The lines are labeled in the legend. When this option is set to *Automatic*, the parameter with the second most values is selected.

Show Power (not Beta) on Plots

When checked, power is displayed on the plots. When unchecked, power quantities are displayed on the plots as beta using the relationship: **Beta = 1 – Power**.

Appearance

Format

Click this button to edit the format of the plots with random data. The Scatter Plot chapter is devoted to the options available in this window.

Edit During Run

This option controls whether the plot may be reformatted interactively after it has been generated. When checked, this option allows charts to be formatted interactively using a plot-editing window.

Abbreviations for Titles, Labels, and Legend

Parameter Abbreviations

These options specify the abbreviations that are used for the parameters in the titles of the plots and the axis labels. It is usually necessary to keep these abbreviations as short as possible since the titles may become very large if the abbreviations are large.

Chapter 5

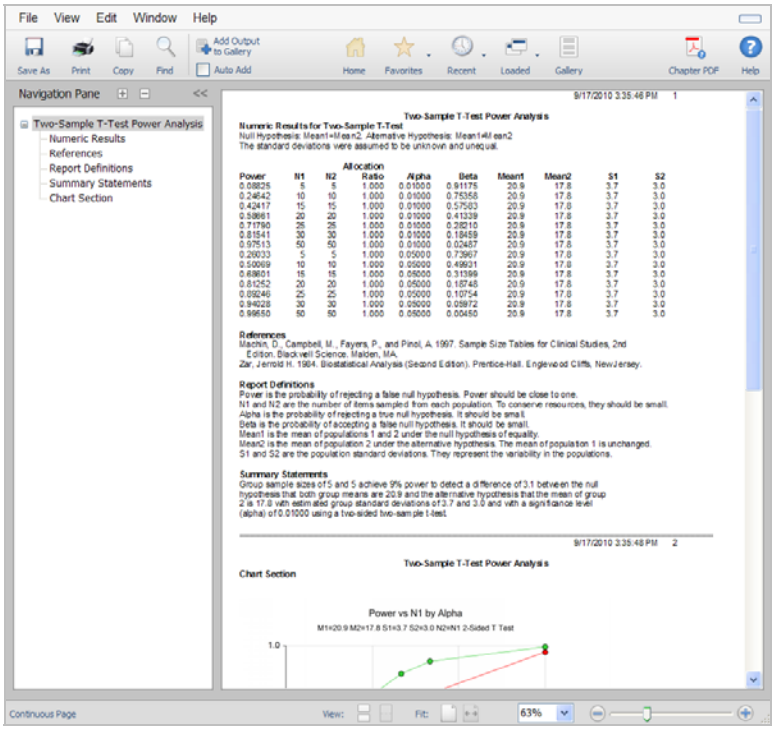
The Output Window

Introduction

PASS sends all statistics and graphics output to its built-in output viewer from where they can be viewed, edited, printed, or saved. Reports and graphs are saved in rich text format (RTF). Since RTF is a standard document transfer format, these files may be loaded directly into your word processor for further processing. This chapter covers the basics of the built-in output viewer in **PASS**.

Viewing the Output

The output of the Example 1 template of the Two-Sample T-Test procedure is shown below. The output window is in full-screen mode. The screen will look similar to this. Note that the actual size of your screen depends on the resolution of your monitor, so it may vary.



Output and Gallery

The **PASS** output system maintains two documents: *Output* and *Output Gallery*. Although both of these documents allow you to view your data, the *Output* document serves as a viewer while the *Output Gallery* document serves as a recorder.

You can load additional documents as well. For example, you might want to view the output from a previous analysis to compare the results with the current analysis. To do this, you open a third document that has been saved from a previous analysis.

All **PASS** documents are stored in the RTF format. This is a common format that is used by most word processors, including Microsoft Word. When you save a **PASS** report, you will be able to load that report directly into your own word processor. All text, formatting, and graphics will appear in your word processor ready for further editing. You can then save the document in your word processor's native format. In this way, you can easily transfer the output of a **PASS** procedure to almost any format you desire.

Output Document

The Output document displays the output report from the current analysis. Whenever you run a **PASS** procedure, the resulting reports and graphs are displayed in the Output document. Each new run clears the existing Output document, so if you want to save a report, you must do so before running the next report.

The Output document provides four main functions: display, print, add to gallery document, and save as an RTF file.

Output Gallery Document

The Output Gallery document provides a place to store a permanent record of your analysis. Since the Output document is erased by each new analysis, you need a place to store your permanent work. The Output Gallery document serves this purpose. When you have a report or graph that you want to keep, copy it from the Output document to the Output Gallery document.

The Output Gallery document provides four main word processing functions: load, display and edit, print, and save. When you load a file into the Output Gallery document, you can add new output to it. In this way, you can record your work on a project in a single file, even though your work on that project is spread out over several days.

Menus

The menus provide a convenient way to perform several functions or transfer from module to module within the **PASS** system. Each set of menus will be briefly described here.

File Menu

The File Menu is used for opening, saving, and printing **PASS** word processor files. All options apply to the currently active document (the document whose title bar is selected). We will now discuss each of the options on this menu.

New (Clear This Window)

This option clears the output document.

Open

This option opens a saved document. This option is only available on the Gallery document window and on additional document windows. You cannot open a saved report into the Output document window.

Recent

This option opens a saved document from the list of recently saved documents. This option is only available on the Gallery document window and on additional document windows. You cannot open a saved report into the Output document window.

Open a New Report Window

This option opens an empty document. You might use this when you want to start a new project.

Open a Report File in a New Window

This option opens a saved document in a new window.

Add Output to Gallery

Selecting this option copies the contents of the Output document to the Gallery document. The Output document remains unchanged. This allows you to save the current output document for further use.

Auto-Add Output to Gallery

When Auto-Add is on, the contents of the Output document are automatically copied to the Gallery document. The Output document remains unchanged. If you want to keep a copy of all the output that has been placed in the Gallery document, you will still need to save it manually.

This function allows you to automatically save all output for further use.

Save

This option lets you save the contents of the currently active document. This option is only available on the Gallery document window and on additional document windows.

Save As

This option lets you save the contents of the currently active document to a designated file using the RTF format. Note that only the active document is saved. Also note that all file names should have the “RTF” extension so that other systems can recognize their format.

Export

This option lets you save the contents of the currently active document to a designated file using the RTF, doc, docx, pdf or other formats.

Print

This option lets you print the entire document or a range of pages. When you select this option, a Print Dialog box will appear that lets you control which pages are printed.

Print Preview

This option displays the output report as it will appear on the printed page. Use it to preview your report before printing it out.

Report System Options

This option displays the system options window. The Reports tab is used to set up the pages of the output for procedure runs. These options take effect when a procedure is run. Examples of options are titles, page numbers, font, page breaks, page size, and page orientation.

Hide This Window

This option minimizes the document that is currently being viewed. Note that this option does not clear the document, it simply minimizes it.

Exit PASS

This option exits the **PASS** system. All documents, spreadsheets, and procedure windows are closed.

View Menu

This menu lets you designate the views of the output page.

Show/Hide Toolbar

This option controls whether the toolbar at the top is displayed.

Show/Hide Navigation Pane

This option controls whether the navigation pane at the left is displayed.

Show/Hide Bottom Bar

This option controls whether the bar at the bottom is displayed.

Show/Hide Horizontal Ruler

This option controls whether the ruler and tabs bar is displayed. The ruler displays the physical dimensions of the document and lets you set the margins and tabs of your document. Only the currently selected part of your document is affected by a change in the tabs and margins.

Expand Navigation Headings

This option expands all headings in the navigation pane to show sub-headings (if the navigation pane is showing).

Collapse Navigation Headings

This option collapses all headings in the navigation pane to show only headings (if the navigation pane is showing).

Page Layout

Click here to show the output divided into the pages as specified in the System Options when the procedure was run.

Continuous Page

Click here to show the output as a single, continuous page.

Maintain Page Size

If this option is selected, the size of the document remains unchanged when the Output window is resized.

Fit to Width

If this option is selected, the size of the document grows or shrinks to fit the width of the window when the Output window is resized.

Fit to Width and Height

If this option is selected, the size of the document grows or shrinks to fit the width and height of the window when the Output window is resized.

Zoom

Use this option to specify a specific viewing size.

Previous Page

Click to move backward one page.

Next Page

Click to advance forward one page.

Edit Menu

This menu contains options that let you edit a document.

Undo

This item reverses the last edit action. It is particularly useful for replacing something that was accidentally deleted.

Redo

This item un-reverses the last edit action.

Cut

This item copies the currently selected text to the Windows clipboard and erases it from the document. You can paste the information from the clipboard to a different location in the current document, into another document, into a datasheet in the spreadsheet, or into another application. The selected text is erased.

Copy

This item copies the currently selected text from the document to the Windows clipboard. You can paste this information from the clipboard to a different location in the current document, into another document, into a datasheet in the spreadsheet, or into another application. The selected text is not modified.

Paste

This item copies the contents of the clipboard to the current document at the insertion point. This command is especially useful for moving selected information from the Output document to the Log document.

Font

This option is used to specify the font of the currently selected text region.

Delete

This option deletes the currently selected region.

Select All

This item selects the entire document. Although you can select a portion of the document using the mouse or a shift-arrow key, this is much faster if you want to select the entire document.

Insert Page Break

This option inserts a page break at the current cursor location.

Find

This item opens the Search dialog box. You can specify text that you want to search for. This is especially useful when you are looking for a certain topic or data value in a large report.

Replace

This item opens the Search and Replace dialog box. This allows you to make repetitive changes. For example, you might want to change the name of one of the variables to a more useful name.

Window Menu

This menu lets you display any of the other windows in the **PASS** system that are currently open such as the Output Gallery window, the Output Spreadsheet window, the PASS Home window, or favorite, recent, or currently open procedure windows.

PASS Home

Select this option to display the PASS Home window.

Output Gallery

Select this option to display the output gallery window.

Output Spreadsheet

Select this option to display the output spreadsheet.

Favorite Procedures

Select this option to display a list of the procedure that have been added to the favorites list.

Recent Procedures

Select this option to display a list of the recently run procedures.

Currently Loaded Procedures

The last menu options contain the procedures that are currently open.

Help Menu

This menu gives you access to the **PASS** Help System and PDF documentation, including references and the Quick Start Manual. This menu also gives the citation information for PASS and access to the website and other support information.

Toolbar

A toolbar is a series of small buttons that appear just below the menus at the top of the procedure window. The output window has two toolbars. Each button on the toolbar provides quick access to a menu item.

Save As. This button lets you save the contents of the currently active document to a designated file using the RTF format.

Print. This button lets you print the entire document or a range of pages. When you select this option, a Print Dialog box will appear that lets you control which pages are printed.

Copy. This button copies the currently selected text from the document to the Windows clipboard. You can paste this information from the clipboard to a different location in the current document, into another document, into a datasheet in the spreadsheet, or into another application. The selected text is not modified.

Find. This button opens the Search dialog box. You can specify text that you want to search for. This is especially useful when you are looking for a certain topic or data value in a large report.

Add Output to Gallery. Selecting this option copies the contents of the Output document to the Gallery document. The Output document remains unchanged. This allows you to save the current output document for further use.

Auto-Add Output to Gallery. When Auto-Add is on, the contents of the Output document are automatically copied to the Gallery document. The Output document remains unchanged. If you want to keep a copy of all the output that has been placed in the Gallery document, you will still need to save it manually.

Home. This button brings the PASS Home Window to the front.

Favorites. This drop-down button shows a list of procedures that have been added to the favorites list.

Recent. This drop-down button shows a list of procedures that have been recently run.

Loaded. This drop-down button shows a list of procedures that are currently loaded.

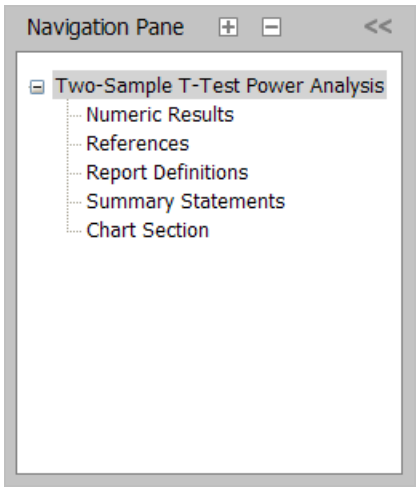
Gallery. This takes the user to window with output that has been sent from the primary output window.

Chapter PDF. This opens the chapter for the output window (this chapter) as an individual PDF.

Help. This opens the PASS Help System and goes to the current topic.

Navigation Pane

The navigation pane is a tree outline that is gives the user quick access to major sections of the output.



The plus and minus buttons at the top can be used to expand and contract the entire tree, respectively. The arrow button at the top right can be pressed to hide the entire Navigation Pane.

Pressing any node of the navigation tree will jump the output to the top of that section. Clicking anywhere in the output will highlight the corresponding section title in the navigation pane.

You can right-click the nodes to go to a location or cut, copy, or delete a section of the output.

Bottom Bar (View Buttons)

The bottom bar of the output window gives quick access to many of the options available in the View menu.

Previous Page. Click to move backward one page.

Next Page. Click to advance forward one page.

Page Layout. Click here to show the output divided into the pages as specified in the System Options when the procedure was run.

Continuous Page. Click here to show the output as a single, continuous page.

Maintain Page Size. If this option is selected, the size of the document remains unchanged when the Output window is resized.

Fit to Width. If this option is selected, the size of the document grows or shrinks to fit the width of the window when the Output window is resized.

Fit to Width and Height. If this option is selected, the size of the document grows or shrinks to fit the width and height of the window when the Output window is resized.

Zoom Box. Use this option to specify a specific viewing size.

Zoom Slider. Sliding along this bar changes the zoom percentage.

Chapter 6

Introduction to Power Analysis

Overview

A statistical test's *power* is the probability that it will result in statistical significance. Since statistical significance is the desired outcome of a study, planning to achieve high power is of prime importance to the researcher. Because of its complexity, however, an analysis of power is often omitted.

PASS calculates statistical power and determines sample sizes. It does so for a broad range of statistical techniques, including the study of means, variances, proportions, survival curves, correlations, bioequivalence, analysis of variance, log rank tests, multiple regression, and contingency tables.

PASS was developed to meet several goals, including ease of learning, ease of use, accuracy, completeness, interpretability, and appropriateness. It lets you study the influence of sample size, effect size, variability, significance level, and power on your statistical analysis.

Brief Introduction to Power Analysis

Statistical power analysis must be discussed in the context of *statistical hypothesis testing*. Hence, this discussion starts with a brief introduction to statistical hypothesis testing, paying particular attention to topics that relate to power analysis and sample size determination. Although the theory behind hypothesis testing is general, its concepts can be reviewed by discussing simple case: testing whether a proportion is greater than a known standard.

Following the usual terminology of statistical hypothesis testing, define two complementary hypotheses

$$H_0: P \leq P0 \text{ vs. } H_1: P > P0$$

where P is the response proportion in the population of interest and $P0$ is the known standard value.

H_0 is called the *null hypothesis* because it specifies that the difference between the two proportions is zero (null).

H_1 is called the *alternative hypothesis*. This is the hypothesis of interest to us. Our motivation for conducting the study is to provide evidence that the alternative (or research) hypothesis is true. We do this by showing that the null hypothesis is unlikely—thus establishing that the alternative hypothesis (the only possibility left) is likely.

Outcomes from a statistical test may be categorized as follows:

1. Reject H_0 when H_0 is true. That is, conclude that H_0 is unlikely when it is true. This constitutes a decision error known as the *Type-I error*. The probability of this error is alpha (α) and is often referred to as the *significance level* of the hypothesis test.
2. Do not reject H_0 when H_0 is false. That is, conclude that H_0 is likely when it is false. This constitutes a decision error known as the *Type-II error*. The probability of this error is beta (β). *Power* is $1-\beta$. It is the probability of rejecting H_0 when it is false.
3. Reject H_0 when H_0 is false. This is a correct decision.
4. Do not reject H_0 when H_0 is true. This is also a correct decision.

The basic steps in conducting a study that is analyzed with a hypothesis test are:

1. Specify the statistical hypotheses, H_0 and H_1.
2. Run the experiment on a given number of subjects.
3. Calculate the value of a test statistic, such as the sample proportion.
4. Determine whether the sample values favor H_0 or H_1.

Binomial Probability Table

In the current example, suppose that a random sample of ten individuals is selected, i.e. *N* = 10. The number of individuals, *R*, with the characteristic of interest is counted. Hence, *R* is the test statistic. A table of binomial probabilities gives the probability that *R* takes on each of its eleven possible values for various values for *P*.

	P								
R	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
0	0.349	0.107	0.028	0.006	0.001	0.000	0.000	0.000	0.000
1	0.387	0.376	0.121	0.040	0.010	0.002	0.000	0.000	0.000
2	0.194	0.302	0.233	0.121	0.044	0.011	0.001	0.000	0.000
3	0.057	0.201	0.267	0.215	0.117	0.042	0.009	0.001	0.000
4	0.011	0.088	0.200	0.251	0.205	0.111	0.037	0.006	0.000
5	0.001	0.026	0.103	0.201	0.246	0.201	0.103	0.026	0.001
6	0.000	0.006	0.037	0.111	0.205	0.251	0.200	0.088	0.011
7	0.000	0.001	0.009	0.042	0.117	0.215	0.267	0.201	0.057
8	0.000	0.000	0.001	0.011	0.044	0.121	0.233	0.302	0.194
9	0.000	0.000	0.000	0.002	0.010	0.040	0.121	0.376	0.387
10	0.000	0.000	0.000	0.000	0.001	0.006	0.028	0.107	0.349

Let us discuss in detail the interpretation of the values in this table for the simple case in which a coin is flipped ten times and the number of heads is recorded. The column parameter *P* is the probability of obtaining a head on any one toss of the coin. When dealing with coin tossing, one would usually set *P* = 0.5, but this does not have to be the case. The row parameter *R* is the number of heads obtained in ten tosses of a coin.

The body of the table gives the probability of obtaining a particular value of *R*. One way to interpret this probability value is as follows: conduct a simulation in which this experiment is repeated a million times for each value of *P*. Using the results of this simulation, calculate the proportion of experiments that result in each value of *R*. This proportion is recorded in this table. For example, when the probability of obtaining a head on a single toss of a coin is 0.5, ten flips of a coin would result in five heads 24.6% of the time. That is, as the procedure is repeated (flipping a coin ten times) over and over, 24.6% of the outcomes would be five heads.

Calculating the Significance Level, Alpha

We will now explain how the above table is used to set the significance level (the probability of a type-I error) to a pre-specified value. Recall that a type-I error occurs when an experiment results in the rejection of the null hypothesis when, in fact, the null hypothesis is true. By studying the table, the impact of using different *rejection regions* can be determined. A rejection region is a simple rule that states which values of the test statistic will result in the null hypothesis being rejected.

For example, suppose we want to test *P*0 = 0.5. That is, the null hypothesis is that *P* = 0.5 and the alternative hypothesis is that *P* > 0.5. Suppose the rejection region is *R* equal to 8, 9, or 10. That is, H_0 is rejected if *R* = 8, 9, or 10. From the above table, the probability of obtaining 8, 9, or 10 heads in 10 tosses when *P* = 0.5 is calculated as follows:

$$\begin{aligned}\Pr(R = 8,9,10 \mid P = 0.5) &= \Pr(R = 8 \mid P = 0.5) + \Pr(R = 9 \mid P = 0.5) + \Pr(R = 10 \mid P = 0.5) \\ &= 0.044 + 0.010 + 0.001 \\ &= 0.055\end{aligned}$$

That is, 5.5% of these coin tossing experiments using this decision rule result in a type-I error. By setting the rejection criterion to *R* = 8, 9, or 10, alpha has been set to 0.055.

It is extremely important to understand what alpha means, so we will go over its interpretation again. If the probability of obtaining a head on a single toss of a coin is 0.5, then 5.5% of the experiments that use the rejection criterion of *R* = 8, 9, or 10 will result in the false conclusion that *P* > 0.5.

The key features of this definition that are often overlooked by researchers are:

1. **The value of alpha is based on a particular value of *P*.** Note that we used the assumption "if the probability of obtaining a head is 0.5" in our calculation of alpha. Hence, if the actual value of *P* is 0.4, our calculations based on the assumption that *P* is 0.5 are wrong. Mathematicians call this a conditional probability since it is based on the condition that *P* is 0.5. Alpha is 0.055 if *P* is 0.5.

   Often, researchers think that setting alpha to 0.05 means that the probability of rejecting the null hypothesis is 0.05. Can you see what is wrong with this statement? They have forgotten to mention the key fact that this statement is based on the assumption that *P* is 0.5!

2. **Alpha is a statement about a proportion in multiple experiments.** Alpha tells us what percentage of a large number of experiments will result in 8, 9, or 10 heads. Alpha is a statement about what to expect from future experiments. It is not a statement about *P*. Occasionally, researchers conclude that the alpha level is the probability that *P* = 0.5. This is not what is meant. Alpha is not a statement about *P*. It is a statement about future experiments, given a particular value of *P*.

Interpreting *P* Values

The term *alpha value* is often used interchangeably with the term *p value*. Although these two terms are closely related, there is an important distinction between them. A *p* value is the largest value of alpha that would result in the rejection of the null hypothesis for a particular set of data. Hence, while the value of alpha is set during the planning of an experiment, the *p* value is calculated from the data after experiment has been run.

Calculating Power and Beta

We will now explain how to calculate the power. Recall that power is the probability of rejecting a false null hypothesis. A false H_0 means that *P* is some value other than *P*0. In order to compute power, we must know the actual value of *P*.

Returning to our coin tossing example, suppose the actual value of *P* is 0.7. What is the power and beta value of this testing procedure? The decision rule is to reject the null hypothesis when *R* is 8, 9, or 10. From the above probability table, the probability of obtaining 8, 9, or 10 heads in 10 tosses of a coin when probability of a head is actually 0.7 is

$$\begin{aligned}\Pr(R = 8,9,10 \mid P = 0.7) &= \Pr(R = 8 \mid P = 0.7) + \Pr(R = 9 \mid P = 0.7) + \Pr(R = 10 \mid P = 0.7) \\ &= 0.233 + 0.121 + 0.028 \\ &= 0.382\end{aligned}$$

This is the power. The value of a type-II error is 1.000 - 0.382, which is 0.618. That is, if *P* is 0.7, then 38.2% of these coin tossing experiments will reject H_0, while 61.8% of them will result in a type-II error.

It is extremely important to understand what beta means, so we will go over its interpretation again. If the probability of obtaining a head on the toss of a coin is 0.7, then 61.8% of the experiments that use the rejection criterion of *R* = 8, 9, or 10 will result in the false conclusion that *P* = 0.5.

The key features of this definition that are often overlooked by researchers are:

1. **The value of beta is based on a particular value of *P*.** Note that we used the assumption “if the probability of obtaining a head is 0.7” in our calculation of beta. Hence, if the actual value of *P* is 0.6, our calculation based on the assumption that *P* was 0.7 is wrong.
2. **Beta is a statement about the proportion of experiments.** Beta tells us what percentage of a large number of experiments will result in 8, 9, or 10 heads. Beta is a statement about what we can expect from future experiments. It is not a statement about *P*.
3. **Beta depends on the value of alpha.** Since the rejection region (8, 9, or 10 heads) depends on the value of alpha, beta depends on alpha.
4. **You cannot make both errors at the same time.** A type-II error can only occur when a type-I error did not occur, and vice versa.

Specifying Alternative Values of the Parameters

We have noted a great deal of confusion about specifying the values of the parameters under the alternative hypothesis. The alternative hypothesis is usually that the value of one parameter is different from another. The hypothesis does not usually specify how different. It simply gives the direction of the difference. The power is calculated at specified alternative values. These values should be considered as values at which the power is calculated, not as the true value.

Effect Size

The *effect size* is the size of the change in the parameter of interest that can be detected by an experiment. For example, in the coin tossing example, the parameter of interest is *P*, the probability of a head. In calculating the sample size, we would need to state what the baseline probability is (probably 0.5) and how large of a deviation from *P* that we want to detect with our experiment. We would expect that it would take a much larger sample size to detect a deviation of 0.01 than it would to detect a deviation of 0.40.

Selecting an appropriate effect size is difficult because it is subjective. The question that must be answered is: what size change in the parameter would be of interest? Note that, in power analysis, the effect size is not the actual difference. Instead, *the effect size is the change in the parameter that is of interest* or is to be detected. This is a fundamental concept that is often forgotten after the experiment is run.

After an experiment is run that leads to non-significance, researchers often ask, "What was the experiment's power?" and "How large of a sample size would have been needed to detect significance?" To compute the power or sample size, they set the effect size equal to the amount that was seen in their experiment. This is incorrect. *When performing a power analysis after an experiment has completed, the effect size is still the change in the parameter that would be of interest to other scientists*. It is not the change that was actually observed!

Often, the effect size is stated as a percentage change rather than an absolute change. If this is the case, you must convert the percentage change to an absolute change. For example, suppose that you are designing an experiment to determine if tossing a particular coin has exactly a 50% chance of yielding a head. That is, *P*0 is 0.50. Suppose your gambling friends are interested in whether a certain coin has a 10% greater chance. That is, they are concerned with the case where *P* is 0.55 or greater. The effect size is |0.50 - 0.55| or 0.05.

Types of Power Analyses

There are several types of power analyses. Often, power analysis is performed during the design phase of a study to determine the sample size. This type of study would determine the value of *N* for set values of alpha and beta. Another type of power analysis is a post hoc analysis, which is done after the study is concluded. A post hoc analysis studies such questions as:

1. What sample size would have been needed to detect a specific effect size?
2. What is the smallest effect size that could be detected with this sample size?
3. What was the power of the test procedure?

These and similar questions may be answered using power analysis. By considering these kinds of questions after a study is concluded, you can gain important insights into how to make your research more efficient and effective.

Nuisance Parameters

Statistical hypotheses usually make statements about one or more parameters from a set of one or more probability distributions. Often, the hypotheses leave other parameters of the probability distribution unspecified. These unspecified parameters are called 'nuisance' parameters.

For example, a common clinical hypothesis is that the response proportions of two drugs are equal. The null hypothesis is that the difference between these two drugs is zero. The alternative is that the difference is non-zero. Note that the actual values of the two proportions are not stated in the hypothesis—just their difference. The actual values of the proportions will be needed to compute the power. That is, different powers will result for the case when $P1 = 0.05$ and $P2 = 0.25$ and for the case $P1 = 0.50$ and $P2 = 0.70$. In this example, the proportion difference ($D = P1 - P2$) is the parameter of interest. The baseline proportion, $P1$, is a nuisance parameter.

Another example of a nuisance parameter occurs when using the t-test to test whether the mean is equal to a particular value. When computing the power or sample size for this test, the hypothesis specifies the value of the mean. However, the value of the standard deviation is also required. In this case, the standard deviation is a nuisance parameter.

When performing a power analysis, you should state all your assumptions, including the values of any nuisance parameters that were used. When you do not have any idea as to reasonable values for nuisance parameters, you should use a range of possible values so that you can analyze how sensitive the results are to the values of the nuisance parameters. Also, do not be tempted to use the nuisance parameter's value from a previous (or pilot) study. Instead, a reasonable strategy is to compute a confidence interval and use the confidence limit that results in the largest sample size.

Choice of Test Statistics

Many hypothesis tests can be tested with a variety of test statistics. For example, statisticians often have to decide between the t-test and the Wilcoxon test when testing means. Similarly, when testing whether two proportions are equal, they have to decide whether to use a z-test or an exact test. If they choose a z-test, they have to decide whether to apply a continuity correction.

In most cases, each test statistic will have a different power. Thus, it should be obvious that *you must compute the power of the test statistic that will be used in the analysis*. A sample size based on the t-test will not be accurate for a nonparametric test.

The next question is usually "Which test statistic should I use?" You might say "They one that requires the smallest sample size." However, other issues besides power must be considered. For example, consideration must be given to whether the assumptions of the test statistic will be met be the data. If your data is binary, it is probably unreasonable to assume that they are continuous.

These are simple principles, but they are often overlooked.

Types of Hypotheses

Hypothesis tests work this way. If the null hypothesis if rejected, the alternative hypothesis is concluded to be true. However, if null hypothesis is not rejected, no conclusion is reached--the null hypothesis *is not* concluded to be true. The only way that a conclusion is reach is if the null hypothesis is rejected.

Because of this, it is very important that the null and alternative hypotheses be constructed so that the conclusion of interest is associated with the alternative hypothesis. That way, if the null hypothesis is rejected, the study reaches the desired conclusion.

There are several types of hypotheses. These include inequality, equivalence, non-inferiority, and superiority hypotheses. In the statistical literature, these terms are used with completely different meanings, so it is important to define what is meant by each. We have tried to adopted names that are associated with the alternative hypothesis, since this is the hypothesis of interest.

It is important to note that even though two sets of hypotheses may be similar, they often will have different power and sample size requirements. For example, an equivalence test (see below) appears to be the simple reverse of a two-sided test of inequality, yet the equivalence test requires a much larger sample size to achieve the same power as the inequality test. Hence, you cannot select the sample size for an inequality test and then later decide to run an equivalence test.

Each of the sections to follow will give a brief definition along with an example based on the difference between two proportions.

Inequality Hypothesis

The term ‘inequality’ is represents the classical one-sided and two-sided hypotheses in which the alternative hypothesis is simply that the two values are unequal. These hypotheses are called tests of superiority by Julious (2004), emphasizing the one-sided versions.

Two-Sided

When the null hypothesis is rejected, the conclusion is simply that the two parameters are unequal. No statement is made about how different. For example, 0.501 and 0.500 are unequal, as are 0.500 and 0.800. Obviously, even though the former are different, the difference is not large enough to be of practical importance in most situations.

$$H_0: p_1 - p_2 = 0 \text{ vs. } H_1: p_1 - p_2 \neq 0 \text{ or } H_1: p_1 - p_2 < 0 \text{ or } p_1 - p_2 > 0$$

One-Sided

These tests offer a little more information than the two sided tests since the direction of the difference is given. Again, no indication is made about how much higher (or lower) the superior value is to the inferior.

$$H_0: p_1 - p_2 \leq 0 \text{ vs. } H_1: p_1 - p_2 > 0 \text{ or } H_0: p_1 - p_2 \geq 0 \text{ vs. } H_1: p_1 - p_2 < 0$$

Non-Inferiority Hypothesis

These tests are a special case of the one-sided inequality tests. The term ‘non-inferiority’ is used to indicate that one treatment is not worse than another treatment. That is, one proportion is not less than another proportion by more than a trivial amount called the ‘margin of equivalence’.

For example, suppose that a new drug is being developed that is less expensive and has fewer side effects than the standard drug. Producers must show that its effectiveness is no worse than the drug it is to replace.

When testing two proportions in which a higher proportion is better, the non-inferiority of treatment 1 as compared to treatment 2 is expressed as

$$\mathrm{H}_0: p_1 - p_2 \le -\delta \text{ vs. } \mathrm{H}_1: p_1 - p_2 > -\delta \text{ or } \mathrm{H}_0: p_1 \le p_2 - \delta \text{ vs. } \mathrm{H}_1: p_1 > p_2 - \delta,$$

where $\delta > 0$ is called the margin of equivalence. Note that when H_0 is rejected, the conclusion is that the first proportion is not less than the second proportion by more than δ.

Perhaps an example will help introduce this type of test. Suppose that the current treatment for a disease works 70% of the time. Unfortunately, this treatment is expensive and occasionally exhibits serious side-effects. A promising new treatment has been developed to the point where it can be tested. One of the first questions that must be answered is whether the new treatment is as good as the current treatment. In other words, do at least 70% of subjects respond to the new treatment?

Because of the many benefits of the new treatment, clinicians are willing to adopt the new treatment even if it is slightly less effective than the current treatment. They must determine, however, how much less effective the new treatment can be and still be adopted. Should it be adopted if 69% respond? 68%? 65%? 60%? There is a percentage below 70% at which the difference between the two treatments is no longer considered ignorable. After thoughtful discussion with several clinicians, it is decided that if a response of at least 63% is achieved, the new treatment will be adopted. The difference between these two percentages is called the *margin of equivalence*. The margin of equivalence in this example is 7% (which is ten percent of the original 70%).

The developers must design an experiment to test the hypothesis that the response rate of the new treatment is at least 0.63. The statistical hypothesis to be tested is

$$H_0: p_1 - p_2 \le -0.07 \text{ versus } H_1: p_1 - p_2 > -0.07.$$

Notice that when the null hypothesis is rejected, the conclusion is that the response rate is at least 0.63. Note that even though the response rate of the current treatment is 0.70, the hypothesis test is about a response rate of 0.63. Also, notice that a rejection of the null hypothesis results in the conclusion of interest.

Non-zero Null Hypothesis

These tests are a special case of the one-sided inequality tests. The term ‘non-zero null’ is used to indicate that one treatment is better than another by more than a trivial amount. For example, suppose that a new drug is being developed that is thought to have superior performance to the existing drug. Producers must show that its effectiveness is better than the drug it is to replace.

When testing two proportions in which a higher proportion is better, the superiority of treatment 1 over treatment 2 is expressed as

$$\mathrm{H}_0: p_1 - p_2 \le \delta \text{ vs. } \mathrm{H}_1: p_1 - p_2 > \delta \text{ or } \mathrm{H}_0: p_1 \le p_2 + \delta \text{ vs. } \mathrm{H}_1: p_1 > p_2 + \delta,$$

where $\delta > 0$ is called the difference margin. Note that when H_0 is rejected, the conclusion is that the first proportion is higher than the second proportion by more than δ.

Equivalence Hypothesis

The term ‘equivalence’ is used here to represent tests designed to show that response rates of two treatments do not differ by more than a trivial amount called the ‘margin of equivalence’. These tests are the reverse of the two-sided inequality test.

The typical set of hypotheses are

$$\mathrm{H}_0: p_1 - p_2 \le \delta_L \text{ or } p_1 - p_2 \ge \delta_U \text{ vs. } \mathrm{H}_1: \delta_L \le p_1 - p_2 \le \delta_U ,$$

where $\delta_L < 0$ and $\delta_U > 0$ are called the *equivalence limits*.

Suppose 70% of subjects with a certain disease respond to a certain drug. The company that produces the drug has decided to open a new facility in another city. They must show that the drug produced in the new facility is equivalent (all most the same) as that produced in existing facilities. After thoughtful discussion with several clinicians and regulatory agencies, it is decided that if the response rate of the drug produced at the new facility is between 65% and 75%, the new facility will go into production. In this case, the *margin of equivalence* is 5%.

The statistical hypothesis to be tested is

$$H_0: |p_1 - p_2| \ge 0.05 \text{ vs. } H_1: |p_1 - p_2| < 0.05 .$$

Chapter 7

Proportions

Introduction

This chapter introduces power analysis and sample size calculation for proportions. When the response is binary, the results for each group may be summarized as proportions. Usually, hypothesis tests are conducted to compare two proportions.

There are many issues that must be considered when designing experiments that use proportions. For example, will the proportions be analyzed directly, or will they be converted into differences, ratios, or odds ratios? Which test statistic will be used? Will the design use independent groups or will subjects be measured twice? Will the study have an active control?

The various answers to these and other questions result in a large number of situations. This chapter will introduce you to the issues that are common to all the tests of proportions.

Designs

There are several experimental designs for comparing two proportions. You can you a one-sample design to compare a sample proportion to a specific value. You can compare proportions from two independent samples using independent, stratified, cluster-randomized, or group-sequential designs. You can compare two correlated proportions. And finally, you can compare several categories in a contingency table.

Comparing Proportions

Two proportions may be compared by considering their difference, ratio, or odds ratio. Each of these cases may lead to different test statistics with different powers and sample size requirements.

Assume that p_1 is the response proportion of the experimental group and p_2 is the response proportion of the control (standard or reference) group. Mathematically, these alternative parameterizations are

Parameter	**Computation**
Difference	$\delta = p_1 - p_2$
Ratio	$\phi = p_1 / p_2$
Odds Ratio	$\psi = \dfrac{p_1 / (1 - p_1)}{p_2 / (1 - p_2)}$

Once you know p_1 and p_2, you can calculate any of these values, and you can easily change from once parameterization to another. Thus, the choice of which parameter you use may seem arbitrary. However, since different parameterizations lead to different test statistics, the choice can lead to a different power and sample size. These parameterizations will be discussed next.

Difference

The difference $\delta = p_1 - p_2$ is perhaps the most common method of comparing two proportions. This parameter is easy to interpret and communicate. It gives the absolute impact of the treatment. However, there are subtle difficulties that can arise with its use.

One difficulty occurs when the event of interest is rare. If a difference of 0.001 is reported for an event with a baseline probability of 0.40, we would dismiss this as being trivial. That is, there is usually little interest in a treatment that decreases the probability from 0.400 to 0.399. However, if the baseline probability of a disease is 0.002, a 0.001 decrease in the disease probability would represent a 50% reduction. Thus, the interpretation of the difference depends on the baseline probability of the event.

When planning studies, the value of p_2 is usually known and the power is calculated at various values of δ. The value of p_1 is then calculated using $p_1 = p_2 + \delta$. Because of the requirement that $0 < p_1 < 1$, the values of δ are restricted to the interval $-p_2 < \delta < 1 - p_2$, not $-1 < \delta < 1$ as you might expect. Likewise, the values of p_2 are restricted to $0 < p_2 < 1 - \delta$ if $\delta > 0$ and $-\delta < p_2 < 1$ if $\delta < 0$.

Because typical values of δ are usually between -0.2 and 0.2, the allowable values of p_2 are restricted to be between 0.2 and 0.8. When the values of p_2 are outside this range, it may be more convenient to use the odds ratio.

Ratio

The (risk) ratio, $\phi = p_1 / p_2$, gives the relative change in the probability of the outcome under each of the hypothesized values. This parameter is direct and easy to interpret. To compare the ratio with the difference, examine the case where p_1 = 0.1437 and p_2 = 0.0793. One should consider which number is more enlightening, δ = -0.0644, or ϕ = 55.18%. In many cases, the relative change (the ratio) is easier to interpret that the absolute change (the difference).

When planning studies, the value of p_2 is usually known and the power is calculated at various values of ϕ. The value of p_1 is then calculated using $p_1 = p_2 \times \delta$. Because of the requirement that $0 < p_1 < 1$, the values of ϕ are restricted to the interval $0 < \phi < 1 / p_2$, not $0 < \phi < \infty$ as you might expect. Likewise, the values of p_2 are restricted to $0 < p_2 < 1 / \phi$ if $\phi > 1$ and $0 < p_2 < 1$ if $\phi < 1$.

Because typical values of ϕ are usually between 0.5 and 1.5, the values of p_2 are restricted to be between 0 and 0.667. When the values of p_2 are outside this range, it may be more convenient to use the odds ratio.

Odds Ratio

Chances are usually communicated as long-term proportions or probabilities. In betting, chances are often given as odds. For example, the odds of a horse winning a race might be set at 10-to-1 or 3-to-2. Odds can easily be translated into probabilities, and vice versa. An odds of 3-to-2 means that the event is expected to occur three out of five times. That is, an odds of 3-to-2 (1.5) translates to a probability of winning of 0.60.

The odds of an event are calculated by dividing the event risk by the non-event risk. Thus the odds are

$$Odds_1 = \frac{p_1}{1-p_1} \text{ and } Odds_2 = \frac{p_2}{1-p_2}$$

For example, if *p* is 0.6, the odds are 0.6/0.4 = 1.5. Rather than represent the odds as a decimal amount, it is re-scaled into whole numbers. Thus, instead of presenting the odds as 1.5-to-1, they present as 3-to-2.

Two odds could be compared by considering their difference, but it is more convenient in many situations to form their ratio. Thus, another way to compare proportions is to compute the ratio of their odds. The odds ratio is

$$\psi = \frac{Odds_1}{Odds_2} = \frac{\frac{p_1}{1-p_1}}{\frac{p_2}{1-p_2}}$$

Unlike the difference and the ratio, the odds ratio is not restricted by the value of p_2. The range of possible values of the odds ratio is $-\infty < \psi < \infty$. Because of the freedom in specifying the parameters, the odds ratio is a popular parameterization, even though it is not as easy to interpret as the difference and the ratio.

Specifying the Proportions – Important!

It is important to understand the interpretation of p_1 and p_2 within **PASS**. Suppose p_1 represents the proportion in the treatment group and p_2 represents the proportion in the control group. In most hypothesis tests, these values are equal under the null hypothesis and different under the alternative hypothesis. Thus, under the null hypothesis, all that is needed is the value of p_1 or p_2, but not both. Under the alternative hypothesis, both values are necessary. So, when the input screen asks for p_2 and the difference, these values should be interpreted as follows. The value of p_2 is actually the value of both p_1 and p_2 under H_0. Under H_1, the value of p_1 is calculated from p_2 and δ using the formula $p_1 = p_2 + \delta$.

Also, it is important to understand what we mean by 'under H_1' in the above discussion. Notice that H_1 does not specify the exact value of p_1. Instead, it specifies a range of values. For

example, H_1 might be $p_1 > p_2$ or $p_1 - p_2 > \delta$. However, even though H_1 gives a range of values for p_1, the power is computed at a specific value of p_1.

Selecting an appropriate value for p_1 must be done very carefully. We recommend the following approach. Select a value of p_1 that represents the change from p_2 that you want the experiment to detect. When you calculate a sample size, it is interpreted as the sample size necessary to detect a difference of at least $p_1 - p_2$ when the significance level is α and the power is $1 - \beta$.

The important point is that p_1 is not the value you anticipate obtaining from an experiment. Instead, it is that value of p_1 at which you want to compute the power. This is a very important distinction! This is why, when investigating the power after an experiment is run, we recommend that you do not simply plug in the values of p_1 and p_2 from that experiment. Rather, you use values that represent the size of the difference that you want to detect.

Chapter 8

Means

Introduction

This chapter introduces power analysis and sample size calculation for tests that compare means. In many situations, the results for each treatment group are summarized as means. There are many issues that must be considered when designing experiments for comparing means. For example, are the means independent or correlated? Which test statistic to use? Will a parametric or nonparametric test be used? Are the data normally distributed? Are there more than two treatment groups? The answers to these and other questions result in a large number of situations.

Specifying the Means

Assume that μ_1 is the mean of an experimental group and μ_2 is the mean of a control (standard or reference) group. Suppose δ represents their difference. That is, $\delta = \mu_1 - \mu_2$. In most hypothesis tests, the null hypothesis (H_0) is $\delta = 0$ and the alternative hypothesis (H_1) is $\delta \neq 0$. Since H_0 assumes that $\delta = 0$, all that is really needed to compute the power is the value of δ under H_1. So, when the input screen asks for μ_1 and μ_2, these values should be interpreted as follows. The value of μ_1 is actually the value of both μ_1 and μ_2 under H_0. Under H_1, the values of μ_1 and μ_2 provide the value of δ at which the power is calculated.

The above discussion is summarized in the following chart:

Input Parameter	**Under H_0**	**Under H_1**
Mean1	μ_1, μ_2	μ_1
Mean2	ignored	μ_2

Also, it is important to understand what we mean by 'under H_1' in the above discussion. H_1 defines a range of values for δ at which the power can be computed. To compute the power, the specific values of δ must be determined. Thus, there is not a single power value. Instead, there are an infinite number of power values possible, depending on the value of δ.

Selecting an appropriate value for μ_1 must be done very carefully. We recommend the following approach. Select a value of μ_1 that represents the change from μ_2 that you want the experiment to detect. When you calculate a sample size, it is interpreted as the sample size necessary to detect a difference of at least δ when the significance level is α and the power is $1-\beta$.

It is important to realize that δ is not the value you anticipate obtaining from the experiment. Instead, it is that value of δ at which you want to compute the power. This is a very important

distinction! This is why, when investigating the power after an experiment is run, we recommend that you do not simply plug in the values of μ_1 and μ_2 from that experiment. Rather, you use values that represent the size of the difference that you want to detect.

Specifying the Standard Deviation

Usually, statistical hypotheses about the means make no direct statement about the standard deviation. However, the standard deviation is a parameter in the normal distribution, so its value must be specified. For this reason, it is called a *nuisance parameter*.

Even though it is not of primary interest, an estimate of the standard deviation is necessary to perform a power analysis. Finding such an estimate is difficult not only because it is required before the data are available, but also because the physical interpretation of the standard deviation is vague. How do you estimate a quantity without data and without a clear understanding of what it is? This section will try to help.

Understanding the Standard Deviation

The standard deviation has two general interpretations. First, it is similar to the average absolute difference between each observation and the mean. Second, it is the average absolute difference between every pair of observations.

The standard deviation of a population of values is calculated using the formula

$$\sigma_X = \sqrt{\frac{\sum_{i=1}^{N}\left(X_i - \mu_X\right)^2}{N}}$$

where N is the number of items in the population, X is the variable being measured, and μ_X is the mean of X. This formula indicates that the standard deviation is the square root of 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 absolute difference between the data values and their mean.

Estimating the Standard Deviation

Our task is to find a rough estimate of the standard deviation to use in a power analysis. Several possible methods could be used. These include using the results of a previous study or a pilot study, using the range, using the coefficient of variation, etc. **PASS** includes a Standard Deviation Estimator procedure that will help you obtain a standard deviation estimate based on these methods. It is loaded from the Tools menu. Remember that as the standard deviation increases, the power decreases. Hence, an increase in the standard deviation will cause an increase in the sample size. To be conservative in sample size calculation, you should use a large value for the standard deviation.

Simulations

Most of the formulas used in **PASS** were derived by analytic methods. That is, based on a series of assumptions, a formula for the power and sample size is derived mathematically. This formula is then programmed and made available in **PASS**. Unfortunately, the formula is only as realistic as the assumptions upon which it is based. If the assumptions are inaccurate in a certain situation, the power calculations may also be inaccurate. An alternative to using analytic methods is to use *simulation* (or *Monte Carlo*) techniques. Because of the speed of modern computers, simulations can now be run in minutes that would have taken days or weeks only a few years ago.

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 estimated.

The steps to a simulation study are

1. Specify how the study is carried out. This includes specifying the randomization procedure, the test statistic that is used, and the significance level that will be used.
2. 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.
3. 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.
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 significance level is the proportion of simulated samples in step 2 that lead to rejection. The power is the proportion of simulated samples in step 3 that lead to rejection.

How Large Should the Simulation Be?

As the number of simulations is increased, the precision and running time of the simulation will be increased also. This section provides a method for estimating of the number simulations needed to achieve a given precision.

Each simulation iteration (or simulation) generates a binary outcome: either the null hypothesis is rejected or not. Thus, the significance level and power estimates each follow the binomial distribution. This knowledge makes it a simple matter to compute confidence intervals for the significance level and power values.

The following table gives one-half the width of a 95% confidence interval for the power when the estimated value is either 0.50 or 0.95.

Simulation Size M	Half-Width when Power = 0.50	Half-Width 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 0.014 when the true power is 0.95. Also, as the simulation size is increased beyond 5000, there is only a small amount of additional accuracy achieved. Since most sample-size studies require an accuracy of within one or two percentage points, simulation sizes from 2000 to 10000 should be ample.

You are Running Two Simulations

It is important to realize that when you run a simulation in **PASS**, you are actually running two separate simulations: one to estimate the significance level and the other to estimate the power. The significance-level simulation is defined by the input parameters labeled "|H0". The power simulation is defined by the input parameters labeled "|H1". Even though you have complete flexibility as to what distributions you use in each of these simulations, it usually makes sense to use the same distributions for both simulations—only changing the values of the means.

Unequal Standard Deviations

One of the subtle problems that can make the results of a simulation study misleading is to specify unequal standard deviations unknowingly when you are investigating another feature, such as the amount of skewness. It is well known that if the standard deviations differ (a situation called heteroskedasticity), the accuracy of the significance level and power is doubtful. When investigating the power of the t or F tests in non-normal situations, care must be taken to insure that the standard deviations of the groups remain about the same. Otherwise, the effects of skewness and heteroskedasticity cannot be separated.

Finding the Hypothesized Means

It is important to set the mean difference of each simulation carefully. In the case of analytic formulas, the mean difference is specified easily and directly. Usually, the mean difference is set to zero under the null hypothesis and to a non-zero value under the alternative hypothesis. You must make certain that you follow this pattern when setting up a simulation.

For most distributions, the means are set explicitedly (the exception is the multinomial distribution, where this is impossible). Hence, for both the null and alternative simulations, it is relatively simple to calculate the mean difference. You must do this! We will now present two examples showing how this is done.

For the first example, consider the case of a simulation being run to compare two independent group means using the two-sample t-test. Suppose the **PASS** setup is as follows. Note that N(40 2) stands for a normal distribution with a mean of 40 and a standard deviation of 2.

Distribution	*PASS* Input	Mean Value of Simulated Data
Group 1 Distribution | H0	N(40 2)	40.0
Group 2 Distribution | H0	N(40 2)	40.0
Group 1 Distribution | H1	N(42 2)	42.0
Group 2 Distribution | H1	N(40 2)	40.0

The mean difference under H0 is 40 – 40 = 0, which is as it should be. The mean difference under H1 is 42 – 40 = 2. Hence, the power is being estimated for a mean difference of 2.

Next we will consider a more complicated example. Suppose the **PASS** setup is as follows. Note that N(40 2)[70];K(0)[30] specifies a mixture distribution made up of 70% from a normal distribution with a mean of 40 and a standard deviation of 2 and 30% from a constant distribution with a value of 30.

Distribution	*PASS* Input	Mean Value of Simulated Data
Group 1 Distribution | H0	N(40 2) [70];K(0)[30]	40(0.7) + 30(0.3) = 37.0
Group 2 Distribution | H0	N(40 2) [70];K(0)[30]	40(0.7) + 30(0.3) = 37.0
Group 1 Distribution | H1	N(42 2) [70];K(0)[30]	42(0.7) + 30(0.3) = 38.4
Group 2 Distribution | H1	N(40 2) [70];K(0)[30]	40(0.7) + 30(0.3) = 37.0

The mean difference under H0 is 37.0 – 37.0 = 0, which is as it should be for the null hypothesis. The mean difference under H1 is 38.4 – 37.0 = 1.4. Hence, the power is being estimated by simulation for a mean difference of 1.4.

You must always be aware of what the mean differences are under both the null and alternative hypotheses.

Adjusting the Significance Level

When faced with the task of designing an experiment that will have a specific significance level for a situation that does not meet the usual assumptions, there are several possibilities.

1. A statistician could be hired to find an appropriate testing procedure.
2. A nonparametric test could be run that (hopefully) corrects for the implausible assumptions.
3. The regular parametric test could be run, relying on the test's 'robustness' to correct for the implausible assumptions.
4. A simulation study could be conducted to determine an appropriate adjustment to the significance level so that the actual significance level is at the required value.

We will now present an example of how to do the simulation adjustment alluded to in item 4, above.

The two-sample t-test is known to be robust to the violation of some assumptions, but it is susceptible to inaccuracies when the data contain outliers. A mixture of two normal distributions

will be used to generate data with outliers. The mixture will draw 95% of the data from a normal distribution with a mean of 0 and a standard deviation of 1. The other 5% of the data will come from a normal distribution with a mean of zero and a standard deviation of 10. A simulation study using 10,000 iterations and a sample size of 100 per group produced the following results when the nominal significance level was set to 0.05.

Nominal Alpha	Actual Alpha	Lower 95% Confidence Limit	Upper 95% Confidence Limit	Power
0.050	0.045	0.041	0.049	0.816
0.055	0.051	0.047	0.055	0.843
0.060	0.057	0.053	0.062	0.835

The actual alpha level of the t-test is 0.045, which is below the target value of 0.50. When the nominal alpha level is increased to 0.055, the actual alpha is 0.051—close to the desired level of 0.05. Hence, an adjustment could be applied as follows. Analyze the data with the two-sample t-test even though they contain outliers. However, instead of using an alpha of 0.050, use an alpha of 0.055. When this is done, the simulation shows that the actual alpha will be at the desired 0.05 level.

There is one limitation to this method: the resulting test procedure is not necessarily efficient. That is, it may be possible to derive a testing procedure that is more efficient (requires a smaller sample size to achieve the same power). For example, in this example, a test based on the trimmed mean may be more efficient in the presence of outliers. However, if you do not have the time or ability to derive an alternative test, this adjustment allows you to obtain reasonable testing procedure that achieves a desired significance level and power.

Quick Start Index

Index entries are of the form “chapter-page”. A list of chapters is given in the Table of Contents.

A

Abbreviations, 4-10
Alpha, 6-3
  adjusting, 8-5
Alternative hypothesis, 6-1, 8-1
Axis
  parameters, 4-10

B

Beta
  calculating, 6-4
Binomial probabilities, 6-2
Bottom bar
  output window, 5-10
Buttons
  output window, 5-8
  PASS home window, 3-3
  procedure window, 4-5

D

Data tab, 4-6
Default template, 4-2
Difference
  proportions, 7-2
Documentation
  printing, 1-7
Downloading updates, 1-2

E

Edit menu
  output window, 5-5
Effect size, 6-5
Entering procedure options, 4-1
Entering your serial number, 1-2
Equivalence hypothesis, 6-9
Errors, 6-2
Exiting PASS, 4-3

F

File menu
  output window, 5-2
  procedure window, 4-3

G

Gallery, 5-2

H

Help menu
  output window, 5-7
  procedure window, 4-5
Help system, 1-3
  contents window, 1-5
  index window, 1-4
  navigating, 1-4
  printing documentation, 1-7
  search window, 1-6
Hypothesis
  equivalence, 6-9
  inequality, 6-7
  introduction, 6-1
  means, 8-1
  non-inferiority, 6-7
  superiority, 6-8
  types, 6-6
Hypothesis testing
  introduction, 6-1

I

Icons
  output window, 5-8
  procedure window, 4-5
Inequality hypothesis, 6-7
Installation, 1-1
Introduction to power analysis, 6-1
Iterations
  maximum, 4-7
Iterations tab, 4-7

L

Legend
  parameter, 4-10
Loading a procedure, 2-3

M

Maximum iterations, 4-7
Means
  introduction, 8-1
Menus
  output window, 5-2
  PASS home window, 3-2
  procedure window, 4-3
Minimum system requirements, 1-1
Monte Carlo, 8-3

N

Navigating the help system, 1-4
Navigation pane
  output window, 5-9
New template, 4-3
Non-inferiority hypothesis, 6-7
Nuisance parameter, 8-2
Nuisance parameters, 6-6
Null hypothesis, 6-1, 8-1

O

Odds ratio
  proportions, 7-3
Open example template, 4-3
Open template, 4-3
Operating system, 1-1

Output, 2-5
Output Gallery, 5-2
Output window, 5-1
  bottom bar, 5-10
  edit menu, 5-5
  file menu, 5-2
  help menu, 5-7
  navigation pane, 5-9
  toolbar, 5-8
  view menu, 5-4
  window menu, 5-7

P

P value, 6-4
Panel, 4-1
Parameters
  abbreviations, 4-10
  axis, 4-10
  entering, 4-6
  legend, 4-10
PASS help system, 1-3
PASS home window, 3-1
  Category Tree, 3-4
  Find, 3-5
  Menus, 3-2
  Procedure list, 3-5
  Search, 3-5
  toolbar, 3-3
Plot setup tab, 4-9
Power
  calculating, 6-4
  introduction, 6-1
  means, 8-1
Printing documentation, 1-7
Printing output, 5-4, 5-8
Procedure menus
  procedure window, 4-4
Procedure options, 4-1
Procedure window, 4-1
  file menu, 4-3
  help menu, 4-5
  procedure menus, 4-4
  run menu, 4-4
  tabs, 4-6
  toolbar, 4-5
  tools menu, 4-4
  view menu, 4-4
  window menu, 4-4
Program updates, 1-2
Proportions
  comparing, 7-1
  difference, 7-2
  interpretation, 7-3
  introduction, 7-1
  odds ratio, 7-3
  ratio, 7-2

Q

Quitting PASS, 4-3

R

Ratio
  proportions, 7-2
Rejection region, 6-3
Reports tab, 4-8
Requirements, 1-1
Resetting a template, 4-3
RTF files, 5-1
Ruler, 5-4
Run menu
  procedure window, 4-4
Running PASS, 2-1

S

Sample size
  introduction, 6-1
Save template, 4-3
Search
  PASS home window, 3-5
Serial number, 1-2
Significance level, 6-3
  adjusting, 8-5
Simulation, 8-3
  size, 8-3
Standard deviation, 8-2
Starting PASS, 1-2, 2-1
Superiority hypothesis, 6-8
System requirements, 1-1

T

Tabs
  data, 4-6
  iterations, 4-7
  plot setup, 4-9
  reports, 4-8
Tabs on the procedure window, 4-6
Technical support, 1-9
Templates, 4-1
  automatic, 4-2
  creating a new, 4-3
  default, 4-2
  loading, 4-3
  loading examples, 4-3
  opening, 4-3
  opening examples, 4-3
  saving, 4-3
Test statistics, 6-6
Toolbar
  output window, 5-8
  PASS home window, 3-3
  procedure window, 4-5
Tools menu
  procedure window, 4-4
Type-I error, 6-2
Type-II error, 6-2

U

Updates, 1-2

V

View menu
  output window, 5-4
  procedure window, 4-4
Viewing output, 2-5

W

Window menu
  output window, 5-7
  procedure window, 4-4

Chapter 100

Inequality Tests for One Proportion

Introduction

The *One-Sample Proportion Test* is used to assess whether a population proportion is significantly different from a hypothesized value. This is called the hypothesis of *inequality*. The hypotheses may be stated in terms of the proportions, their difference, their ratio, or their odds ratio.

For example, suppose that the current treatment for a disease cures 62% of all cases. A new treatment method has been proposed and studied. In a sample of 80 subjects with the disease that were treated with the new method, 63 were cured. Do the results of this study support the claim that the new method has a higher response rate than the existing method?

This procedure calculates sample size and statistical power for testing a single proportion using either exact or approximate tests. Results are based on exact calculations using the binomial and hypergeometric distributions. Because the analysis of several different test statistics is available, their statistical power may be compared to find the most appropriate test for a given situation.

Some sample size programs use the normal approximation to the binomial distribution for power and sample size estimates. This approximation is useful for rough hand calculations, but more accurate results are easily obtainable with today's software. When the normal approximation to the binomial is used, issues such as the need for continuity correction come into play. We avoid these issues by calculating exact results. Programs that use these approximations will often give different answers. Our calculations are exact, not approximate.

Four Procedures Documented Here

There are four procedures in the menus that use the program module described in this chapter. These procedures are identical except for the type of parameterization. The parameterization can be in terms of proportions, differences in proportions, ratios of proportions, and odds ratios. Each of these options is listed separately on the menus.

Binomial Model

A binomial variable should exhibit the following four properties:

1. The variable is binary --- it can take on one of two possible values.
2. The variable is observed a known number of times. Each observation or replication is called a Bernoulli trial. The number of replications is *n*. The number of times that the outcome of interest is observed is *r*. Thus *r* takes on the possible values 0, 1, 2, ..., *n*.
3. The probability, *P*, that the outcome of interest occurs is constant for each trial.
4. The trials are independent. The outcome of one trial does not influence the outcome of the any other trial.

A binomial probability is calculated using the formula

$$b(r;n,P)=\binom{n}{r}P^r(1-P)^{n-r}$$

where

$$\binom{n}{r}=\frac{n!}{r!(n-r)!}$$

The Hypergeometric Model

When samples are taken without replacement from a population of known size, *N*, the *hypergeometric* distribution should be used in place of the binomial distribution. The properties of a variable that is distributed according to the hypergeometric distribution are

1. The variable is binary--it can take on one of two possible values.
2. The variable is observed a known number of times. Each observation or replication is called a Bernoulli trial. The number of replications is *n*. The number of times that the outcome of interest is observed is *r*. Thus *r* takes on the possible values 0, 1, 2, ..., *n*.
3. The total number of items is *N*. The proportion of items with the characteristic of interest is *P*.

The hypergeometric probability of obtaining exactly *r* of *n* items with the characteristic of interest is calculated using

$$h(r;N,n,P)=\frac{\binom{NP}{r}\binom{N-NP}{n-r}}{\binom{N}{n}}$$

Note that the quantity *NP* is rounded to the nearest integer.

Hypothesis Testing

Steps to Calculate Power

The testing procedure is as follows. Let P represent the true probability that an item selected at random from a population will have a characteristic of interest.

1. State the Hypotheses

H0 is the null hypothesis that the proportion is P0.

H1 is the alternative hypothesis that the proportion is P1.

Three sets of statistical hypotheses may be formulated:

1. $H0: P = P0$ versus $H1: P = P1 \neq P0$. This hypothesis results in a two-tailed test.
2. $H0: P \leq P0$ versus $H1: P = P1 > P0$. This hypothesis results in a one-tailed test.
3. $H0: P \geq P0$ versus $H1: P = P1 < P0$. This hypothesis results in a one-tailed test.

2. Find the Critical Value

For an upper-tailed test with a given sample size find the critical value, Pc, based on the binomial (or hypergeometric) distribution, so that the probability of rejecting H0 when H0 is true is equal to a specified significance level, α.

3. Evaluate the Sample

Select a sample of *n* items from the population and compute the sample proportion, p = r / n. If p > Pc then reject the null hypothesis that P = P0 in favor of an alternative hypothesis that P = P1 > P0.

4. Calculate the Power

The power is the probability of rejecting H0 when the true proportion is P1. That is, the power is the probability that p > Pc calculated from a binomial (or hypergeometric) distribution in which P = P1.

Similar steps are used for the lower-tail and two-tailed tests.

Test Statistics

Many different test statistics have been proposed for testing a single proportion. Most of these were proposed before computers or hand calculators were widely available. Although these legacy methods are still presented in textbooks, their power and accuracy should be compared against modern exact methods before they are adopted for serious research. To make this comparison easy, the power and significance of several tests of a single proportion are available in this procedure.

Exact Binomial Test

The test statistic is *r*, the number of successes in *n* trials. This test should be the standard against which other test statistics are judged. The significance level and power are computed by

enumerating the possible values of *r*, computing the probability of each value, and then computing the corresponding value of the test statistic. Hence the values that are reported in the output for these tests are exact, not approximate.

Z Test

Several *z* statistics have been proposed that use the central limit theorem. This theorem states that for large sample sizes, the distribution of the *z* statistic is approximately normal. All of these tests take the following form:

$$z = \frac{p - P0}{s}$$

Although these *z* tests were developed because the distribution of *z* is approximately normal in large samples, the actual significance level and power can be computed exactly using the binomial distribution.

We include four *z* tests which are based on two methods for computing *s* and whether a continuity correction is applied.

Z Test using S(P0)

This test statistic uses the value of P0 to compute *s*.

$$z_1 = \frac{p - P0}{\sqrt{P0(1 - P0) / n}}$$

Z Test using S(P0) with Continuity Correction

This test statistic is similar to the one above except that a continuity correction is applied to make the normal distribution more closely approximate the binomial distribution.

$$z_2 = \frac{(p - P0) + c}{\sqrt{P0(1 - P0) / n}}$$

where

$$c = \begin{cases} \frac{-1}{2n} & \text{if } p > P0 \\ \frac{1}{2n} & \text{if } p < P0 \\ 0 & \text{if } |p - P0| < \frac{1}{2n} \end{cases}$$

Z Test using S(P-hat)

This test statistic uses the value of *p* to compute *s*.

$$z_3 = \frac{p - P0}{\sqrt{p(1 - p) / n}}$$

Z Test using S(P-hat) with Continuity Correction

This test statistic is similar to the one above except that a continuity correction is applied to make the normal distribution more closely approximate the binomial distribution.

$$z_4 = \frac{(p - P0) + c}{\sqrt{p(1-p)/n}}$$

where

$$c = \begin{cases} \frac{-1}{2n} & \text{if } p > P0 \\ \frac{1}{2n} & \text{if } p < P0 \\ 0 & \text{if } |p - P0| < \frac{1}{2n} \end{cases}$$

T Test

The one-sample t-test may be applied to this design. This is accomplished by considering the *n* trials as the outcomes of a numeric variable in which a success is coded as a '1' and a failure is coded as a '0'. The standard t-test may then be computed on these data values.

Parameterizations of the Proportions

There are several ways to specify the proportions under the null and the alternative hypotheses. The most direct is to simply give values for P0 and P1. However, it is often more meaningful to specify P0 and then specify the alternative as the difference, the ratio, or the odds ratio. The value of P1 is calculated from these values.

Mathematically, these alternative parameterizations are

Parameter	Computation
Difference	$\delta = P1 - P0$
Ratio	$\phi = P1 / P0$
Odds Ratio	$\psi = \frac{P1/Q1}{P0/Q0} = \frac{P1Q0}{P0Q1}$

Difference

The (risk) difference, $\delta = P1 - P0$, is perhaps the most direct method of comparison between the two proportions. This parameter is easy to interpret and communicate. It gives the absolute impact of the treatment. However, there are subtle difficulties that can arise with its interpretation.

One interpretation difficulty occurs when the event of interest is rare. If a difference of 0.001 is reported for an event with a baseline probability of 0.40, we would dismiss this as being trivial. That is, there is usually little interest in a treatment that decreases the probability from 0.400 to 0.399. However, if the baseline probability of a disease is 0.002, a 0.001 decrease in the disease

probability would represent a reduction of 50%. The interpretation depends on the baseline probability of the event.

Ratio

The (risk) ratio, $\phi = P1 / P0$, gives the relative change in the probability of the outcome under each of the hypothesized values. This parameter is direct and easy to interpret. To compare the ratio with the difference, examine the case where P0 = 0.1437 and P1 = 0.0793. One should consider which number is more enlightening, the difference of -0.0644, or the ratio of 55.18%. In many cases, the ratio communicates the change in proportion in a manner that is more appropriate than the difference.

Odds Ratio

Chances are usually communicated as long-term proportions or probabilities. In betting, chances are often given as odds. For example, the odds of a horse winning a race might be set at 10-to-1 or 3-to-2. Odds can easily be translated into probability. An odds of 3-to-2 means that the event is expected to occur three out of five times. That is, an odds of 3-to-2 (1.5) translates to a probability of winning of 0.60.

The odds of an event are calculated by dividing the event risk by the non-event risk. Thus the odds are

$$O1 = \frac{P1}{1 - P1} \text{ and } O0 = \frac{P0}{1 - P0}$$

For example, if P1 is 0.60, the odds are 0.60/0.4 = 1.5. Rather than represent the odds as a decimal amount, it is re-scaled into whole numbers. Thus, instead of saying the odds are 1.5-to-1, we say they are 3-to-2.

Thus, another way to compare proportions is to compute the ratio of their odds. The odds ratio of two proportions is

$$\psi = \frac{O1}{O0} = \frac{\dfrac{P1}{1 - P1}}{\dfrac{P0}{1 - P0}}$$

Procedure Options

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

Data Tab (Common Options)

The Data tab contains the parameters associated with this test such as the proportions, sample sizes, alpha, and power. This chapter covers four procedures, each of which has different options. This section documents options that are common to all four procedures. Later, unique options for each procedure will be documented.

Solve For

Find (Solve For)

This option specifies the parameter to be solved for using the other parameters. The parameters that may be selected are *P0*, *P1*, *Alpha*, *Power and Beta*, and *n*. Under most situations, you will select either *Power and Beta* or *n*.

Select *n* when you want to calculate the sample size needed to achieve a given power and alpha level.

Select *Power and Beta* when you want to calculate the power.

Error Rates

Power or Beta

This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected.

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

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

Alpha (Significance Level)

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

Note that because of the discrete nature of the binomial distribution, the alpha level rarely will be achieved exactly.

A single value may be entered here or a range of values such as *0.05 to 0.2 by 0.05* may be entered.

Sample Size

n (Sample Size)

This option specifies the total number of observations in the sample. Values must be integers greater than one.

You may enter a single value or a range of values such as *10, 50, 100* or *10 to 100 by 10*.

Effect Size

P0 (Null Proportion)

Enter a value (or range of values) for the population proportion under the null hypothesis, P0. This is the baseline proportion, the proportion that exists in the general population. The proportion estimated from the data will be compared to this value by the statistical test.

Proportions must be between zero and one.

You may enter a single value or a range of values such as *0.1 0.2 0.3* or *0.1 to 0.9 by 0.1*.

Test

Test Type

Specify the type of test that will be used in reporting. Note that *C.C.* is short for *Continuity Correction*. This refers to the adding or subtracting 1/(2*n*) to (or from) the numerator of the z-value to bring the normal approximation closer to the binomial distribution.

In most situations, you would select the 'Exact Test' option. The other options are provided for comparative purposes.

N (Population Size)

Enter the total number of items in the population from which the sample of *n* items is selected. Enter *Infinite* to signify an infinite population so that no correction factor is applied. An *infinite* population is generally one in which the number in the population is large and unknown.

Note that *N* must be greater than *n*.

When samples are drawn from a very large (infinite) population, calculations are based on the binomial distribution.

When samples are drawn from a population of known size, specified here as *N*, calculations are based on the hypergeometric distribution.

Data Tab (Proportions)

This section documents options that are used when the parameterization is in terms of the alternative proportion, P1. P0 is the value of the proportion assumed by the null hypothesis and P1 (or P) is the value of the proportion at which the power is calculated.

Effect Size

P1 (Alternative Proportion)

Enter a value (or range of values) for the value of the binomial proportion at which the power is calculated. This is labeled P1 on the screen. Power calculations assume that this is the true value of the proportion.

This value cannot be equal to P0 since, by definition, it must be an alternative.

Test

H1 (Alternative Hypothesis)

This option specifies the alternative hypothesis, H1. This implicitly specifies the direction of the hypothesis test. Note that the null hypothesis, H0, is the opposite of H1.

P represents the actual value of the proportion and P0 represents the specific value of the proportion assumed by the null hypothesis, H0.

Possible selections are

- **H1: P <> P0**

  This is the most common selection. It yields the *two-tailed* test. Use this option when you are testing whether the proportions are different, but you do not want to specify beforehand which proportion is larger. By tradition, most studies are two-tailed unless there is a strong reason to make them one-tailed.

- **H1: P < P0**

  This option yields a *one-tailed* test.

- **H1: P > P0**

  This option also yields a *one-tailed* test.

Data Tab (Differences)

This section documents options that are used when the parameterization is in terms of the difference, P1 – P0. P0 is the value of the proportion assumed by the null hypothesis and P1 (or P) is the value of the proportion at which the power is calculated. Once P0 and the difference are given, the value of P1 is found by the formula: P1 = difference + P0.

Effect Size

Alternative Difference (P1-P0)

This option implicitly specifies the value of P1 (the proportion at which the power is calculated) by explicitly specifying the difference. The difference is used with P0 to calculate the value of P1 using the formula, P1 = diff + P0.

Since P1 is a proportion, the difference must be between –P0 and 1 – P0. By definition, the difference cannot be zero since P1 is an 'alternative' to P0.

A single value or a range of values may be entered here.

Test

H1 (Alternative Hypothesis)

This option specifies the alternative hypothesis, H1. This implicitly specifies the direction of the hypothesis test. Note that the null hypothesis, H0, is the opposite of H1.

Possible selections are

- **H1: Difference <> 0**

  This is the most common selection. It yields the *two-tailed* test. Use this option when you are testing whether the proportions are different, but you do not want to specify beforehand which proportion is larger. By tradition, most studies are two-tailed unless there is a strong reason to make them one-tailed.

- **H1: Difference < 0**

  This option yields a *one-tailed* test.

- **H1: Difference > 0**

  This option also yields a *one-tailed* test.

Data Tab (Ratios)

This section documents options that are used when the parameterization is in terms of the ratio, P1 / P0. P0 is the value of the proportion assumed by the null hypothesis and P1 (or P) is the value of the proportion at which the power is calculated. Once P0 and the ratio are given, the value of P1 is found by the formula: P1 = (P0) x (ratio).

Effect Size

Alternative Ratio (P1/P0)

This option implicitly specifies the value of P1 (the proportion at which the power is calculated) by explicitly specifying the ratio. The ratio is used with P0 to calculate the value of P1 using the formula, P1 = (P0) x (ratio).

Since P1 is a proportion, the ratio must be between 0 and 1 / P0. By definition, the ratio cannot be one since P1 is an 'alternative' to P0.

A single value or a range of values may be entered here.

Test

H1 (Alternative Hypothesis)

This option specifies the alternative hypothesis, H1. This implicitly specifies the direction of the hypothesis test. Note that the null hypothesis, H0, is the opposite of H1.

Possible selections are

- **H1: Ratio <> 1**

  This is the most common selection. It yields the *two-tailed* test. Use this option when you are testing whether the proportions are different, but you do not want to specify beforehand which proportion is larger. By tradition, most studies are two-tailed unless there is a strong reason to make them one-tailed.

- **H1: Ratio < 1**

  This option yields a *one-tailed* test.

- **H1: Ratio > 1**

  This option also yields a *one-tailed* test.

Data Tab (Odds Ratios)

This section documents options that are used when the parameterization is in terms of the odds ratio, O1 / O0 where O1 = P1 / (1 – P1) and O0 = P0 / (1 – P0). P0 is the value of the proportion assumed by the null hypothesis and P1 (or P) is the value of the proportion at which the power is calculated. Once P0 and the odds ratio are given, the value of P1 is found by the formula P1 = A / (1 + A) where A = (O0) x (odds ratio).

Effect Size

Alternative Odds Ratio (O1/O0)

This option implicitly specifies the value of P1 (the proportion at which the power is calculated) by explicitly specifying the odds ratio. Since P1 is a proportion, the odds ratio must be greater than zero. By definition, the odds ratio cannot be one since P1 is an 'alternative' to P0.

A single value or a range of values may be entered here.

Test

H1 (Alternative Hypothesis)

This option specifies the alternative hypothesis, H1. This implicitly specifies the direction of the hypothesis test. Note that the null hypothesis, H0, is the opposite of H1.

Possible selections are

- **H1: Odds Ratio <> 1**

  This is the most common selection. It yields the *two-tailed* test. Use this option when you are testing whether the proportions are different, but you do not want to specify beforehand which proportion is larger. By tradition, most studies are two-tailed unless there is a strong reason to make them one-tailed.

- **H1: Odds Ratio < 1**

  This option yields a *one-tailed* test.

- **H1: Odds Ratio > 1**

  This option also yields a *one-tailed* test.

Iterations Tab

The Iterations tab allows for specification of the maximum number of iterations to be used in searches.

Maximum Iterations

Maximum Iterations Before Search Termination

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 not reported.

Example 1 – Finding the Power

Suppose 50% of patients with a certain type of cancer survive five years. Researchers have developed a new treatment to increase the percentage of individuals that survive five years. Although the researchers do not know the true percentage of patients that will survive with the new treatment, they would like to examine the power that is achieved if the percentage under the new treatment is 60%. The power will be determined for trials with sample sizes of 50, 100, 200, 300, 500, or 800 and a significance level of 0.05. For comparative purposes, the power is also to be calculated for alternative proportions of 55% and 65%.

This is an example of a *historically controlled* trial. *Historically controlled* means that no control group is formed for the current study. Instead, the rates reported from previous studies or that are known to exist in the general population are used. Because of the many advantages that occur when an actual control group is used, historically controlled trials should only be used when a control group is either impossible to obtain or unethical.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for One Proportion [Proportions]** procedure window by expanding **Proportions**, then **One Proportion**, then clicking on **Test (Inequality)**, and then clicking on **Tests for One Proportion [Proportions]**. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power or Beta	*Ignored since this is the Find setting*
Alpha	**0.05**
n (Sample Size)	**50 100 200 300 500 800**
P0 (Null Proportion)	**0.50**
P1 (Alternative Proportion)	**0.55 0.60 0.65**
H1 (Alternative Hypothesis)	**H1: P ≠ P0**
Test Type	**Exact Test**
N (Population)	**Infinite**

Annotated Output

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

Numeric Results

Numeric Results when H0: P = P0 versus H1: P = P<>P0
Test Statistic: Exact Test

Power	N	Proportion Given H0 (P0)	Proportion Given H1 (P1)	Target Alpha	Actual Alpha	Beta	Reject H0 If <=R|>=R
0.0788	50	0.5000	0.5500	0.0500	0.0328	0.9212	17|33
0.1352	100	0.5000	0.5500	0.0500	0.0352	0.8648	39|61
0.2620	200	0.5000	0.5500	0.0500	0.0400	0.7380	85|115
0.3867	300	0.5000	0.5500	0.0500	0.0431	0.6133	132|168
0.5895	500	0.5000	0.5500	0.0500	0.0441	0.4105	227|273
0.7932	800	0.5000	0.5500	0.0500	0.0438	0.2068	371|429
0.2371	50	0.5000	0.6000	0.0500	0.0328	0.7629	17|33
0.4621	100	0.5000	0.6000	0.0500	0.0352	0.5379	39|61
0.7868	200	0.5000	0.6000	0.0500	0.0400	0.2132	85|115
0.9291	300	0.5000	0.6000	0.0500	0.0431	0.0709	132|168
0.9937	500	0.5000	0.6000	0.0500	0.0441	0.0063	227|273
0.9999	800	0.5000	0.6000	0.0500	0.0438	0.0001	371|429
0.5060	50	0.5000	0.6500	0.0500	0.0328	0.4940	17|33
0.8276	100	0.5000	0.6500	0.0500	0.0352	0.1724	39|61
0.9884	200	0.5000	0.6500	0.0500	0.0400	0.0116	85|115
0.9995	300	0.5000	0.6500	0.0500	0.0431	0.0005	132|168
1.0000	500	0.5000	0.6500	0.0500	0.0441	0.0000	227|273
1.0000	800	0.5000	0.6500	0.0500	0.0438	0.0000	371|429

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.
P0 is the value of the population proportion under the null hypothesis.
P1 is the value of the population proportion under the alternative hypothesis.

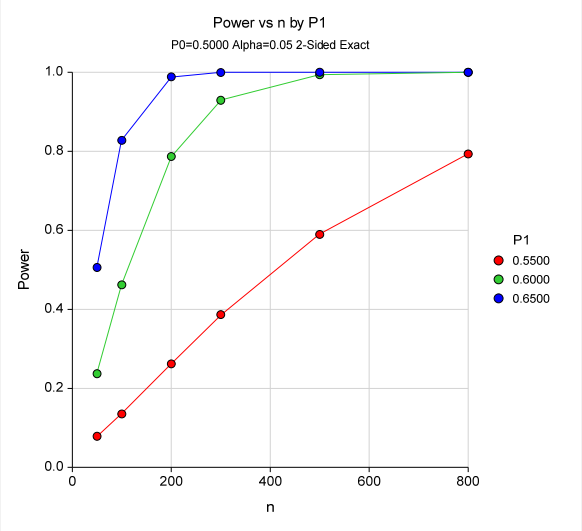
Summary Statements
A sample size of 50 achieves 8% power to detect a difference (P1-P0) of 0.0500 using a two-sided binomial test. The target significance level is 0.0500. The actual significance level achieved by this test is 0.0328. These results assume that the population proportion under the null hypothesis is 0.5000.

This report shows the values of each of the parameters, one scenario per row. Because of the discrete nature of the binomial distribution, the stated (Target) alpha is usually greater than the actual alpha. Hence, we also show the Actual Alpha along with the rejection region.

The symbol, *R*, stands for the number of items with the characteristic of interest out of the *n* items sampled. Hence, for the scenario presented on the first line, an exact test does not exist for these parameters at the target alpha of 0.05. The closest that can be achieved is an alpha of 0.0328. In this case, we would reject the null hypothesis in any sample of size 50 in which the count of individuals with the characteristic of interest is less than or equal to 17 or greater than or equal to 33.

The values from this table are plotted in the chart below.

Plots Section



This plot shows the relationship between power, sample size, and P1 in this example. We note that 80% power is achieved with a sample size of about 200 when P1 is 0.60, which was the specific value of interest.

Example 2 – Finding the Sample Size

Continuing with Example 1, suppose you want to study the impact of various choices for P1 on sample size. Using a significance level of 0.05 and 90% power, find the sample size when P1 is 0.55, 0.60, 0.65, 0.70, 0.75, and 0.80. Assume that an exact, two-tailed binomial test will be used.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for One Proportion [Proportions]** procedure window by expanding **Proportions**, then **One Proportion**, then clicking on **Test (Inequality)**, and then clicking on **Tests for One Proportion [Proportions]**. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**n**
Power or Beta	**0.90**
Alpha	**0.05**
n (Sample Size)	*Ignored since this is the Find setting*
P0 (Null Proportion)	**0.50**
P1 (Alternative Proportion)	**0.55 to 0.80 by 0.05**
H1 (Alternative Hypothesis)	**H1: P ≠ P0**
Test Type	**Exact Test**
N (Population)	**Infinite**

Output

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

Numeric Results

Numeric Results when H0: P = P0 versus H1: P <>P0

Power	N	Proportion Given H0 (P0)	Proportion Given H1 (P1)	Target Alpha	Actual Alpha	Beta	Reject H0 If <=R|>=R
0.9003	1055	0.5000	0.5500	0.0500	0.0487	0.0997	495|560
0.9022	263	0.5000	0.6000	0.0500	0.0483	0.0978	115|148
0.9015	114	0.5000	0.6500	0.0500	0.0487	0.0985	46|68
0.9100	65	0.5000	0.7000	0.0500	0.0464	0.0900	24|41
0.9195	42	0.5000	0.7500	0.0500	0.0436	0.0805	14|28
0.9100	28	0.5000	0.8000	0.0500	0.0357	0.0900	8|20

This report shows the sample sizes corresponding to various values of P1. Notice that a sample size of only 28 is needed to detect the difference between 0.5 and 0.8, but a sample size of 1055 is needed to detect a difference between 0.50 and 0.55.

Example 3 – Investigating the Saw-Tooth Power Function

After releasing the first version of ***PASS***, we received many inquiries about the strange shape of the relationship between power and sample size when testing a single proportion using the exact binomial test. This example will show why this strange shape occurs.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for One Proportion [Proportions]** procedure window by expanding **Proportions**, then **One Proportion**, then clicking on **Test (Inequality)**, and then clicking on **Tests for One Proportion [Proportions]**. You may then make the appropriate entries as listed below, or open **Example 3** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power or Beta	*Ignored since this is the Find setting*
Alpha	**0.05**
n (Sample Size)	**51 to 60 by 1**
P0 (Null Proportion)	**0.60**
P1 (Alternative Proportion)	**0.70**
H1 (Alternative Hypothesis)	**H1: P ≠ P0**
Test Type	**Exact Test**
N (Population Size)	**Infinite**

Output

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

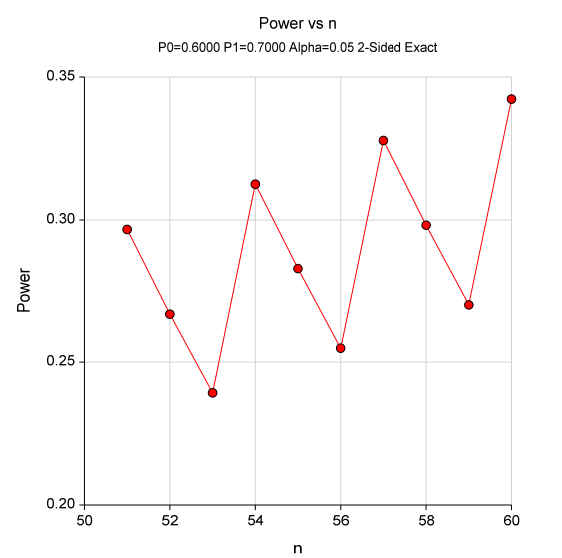
Numeric Results

Numeric Results for H0: P = P0 versus H1: P <>P0

Power	N	Proportion Given H0 (P0)	Proportion Given H1 (P1)	Target Alpha	Actual Alpha	Beta	Reject H0 If <=R|>=R
0.2966	51	0.6000	0.7000	0.0500	0.0443	0.7034	23|38
0.2669	52	0.6000	0.7000	0.0500	0.0328	0.7331	23|39
0.2393	53	0.6000	0.7000	0.0500	0.0348	0.7607	24|40
0.3124	54	0.6000	0.7000	0.0500	0.0371	0.6876	24|40
0.2828	55	0.6000	0.7000	0.0500	0.0379	0.7172	25|41
0.2549	56	0.6000	0.7000	0.0500	0.0281	0.7451	25|42
0.3277	57	0.6000	0.7000	0.0500	0.0417	0.6723	26|42
0.2981	58	0.6000	0.7000	0.0500	0.0314	0.7019	26|43
0.2701	59	0.6000	0.7000	0.0500	0.0327	0.7299	27|44
0.3423	60	0.6000	0.7000	0.0500	0.0354	0.6577	27|44

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



Notice that the power decreases as *n* increases from 51 to 52 and continues to decrease as *n* increases to 53. Usually, the power increases as the sample size increases.

To understand why this happens, look at the last column and at the Actual Alpha column. Note that at *n* = 51, the actual alpha is 0.0443 and at *n* = 52, the actual alpha has decreased to 0.0328. Remember that as alpha decreases, power decreases as well. Hence, increasing the sample size from 51 to 52 was not enough to counterbalance the effect on power of a decrease in alpha from 0.04428 to 0.03281. Hence, the power drops from 0.29656 to 0.26688.

This phenomenon usually occurs for relatively small values of *n*.

Example 4 – Step by Step Calculations

In this example, we will take you step by step through the calculations necessary to compute the power of a specific scenario. We will set n = 10, P0 = 0.50, P1 = 0.80, and alpha = 0.05. We will compute the power of the two-tailed test.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for One Proportion [Proportions]** procedure window by expanding **Proportions**, then **One Proportion**, then clicking on **Test (Inequality)**, and then clicking on **Tests for One Proportion [Proportions]**. You may then make the appropriate entries as listed below, or open **Example 4** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power or Beta	*Ignored since this is the Find setting*
Alpha	**0.05**
n (Sample Size)	**10**
P0 (Null Proportion)	**0.5**
P1 (Alternative Proportion)	**0.8**
H1 (Alternative Hypothesis)	**H1: P ≠ P0**
Test Type	**Exact Test**
N (Population Size)	**Infinite**
Reports Tab	
Numeric Report Probability Decimals	**6**
Report and Plot Proportion Decimals	**6**

Output

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

Numeric Results

Numeric Results when H0: P = P0 versus H1: P <>P0.

Power	N	Proportion Given H0 (P0)	Proportion Given H1 (P1)	Target Alpha	Actual Alpha	Beta	Reject H0 If <=R|>=R
0.375814	10	0.500000	0.800000	0.050000	0.021484	0.624186	1|9

We will now proceed through the calculations necessary to compute this power value.

We first construct a table of binomial probabilities for n = 10 and p = 0.5 using the formula

$$b(r;10,0.5) = \binom{10}{r} 0.5^r (1-0.5)^{10-r}$$

Binomial Probabilities for *n* = 10 and *p* = 0.5

R	Prob(r = R)	Cumulative Total	1 – Cumulative Total
0	0.000977	0.000977	0.999023
1	0.009766	0.010742	0.989258
2	0.043945	0.054688	0.945313
3	0.117188	0.171875	0.828125
4	0.205078	0.376953	0.623047
5	0.246094	0.623047	0.376953
6	0.205078	0.828125	0.171875
7	0.117188	0.945313	0.054688
8	0.043945	0.989258	0.010742
9	0.009766	0.999023	0.000977
10	0.000977	1.000000	0.000000

When we construct a two-tailed test, we split the alpha value evenly between the two tails. Hence, we place $\alpha / 2$ (or 0.025) in each tail. Moving down from the top, we find that the cumulative probability is 0.010742 for $R = 1$ and 0.054688 for $R = 2$. Since 0.054688 is greater than 0.025, we adopt $R = 1$ as our lower rejection value. Likewise, we find that $R = 9$ is the upper rejection value.

Our testing strategy is

1. Draw a sample of 10 items and count the number with the characteristic of interest. Call this value *r*.
2. If $r = 0$, 1, 9, or 10, reject the null hypothesis that $p = 0.5$ in favor of the alternative hypothesis that $p \neq 0.5$.

Now, to compute the power for P1 = 0.8, we must compute another table of binomial probabilities, this time for $p = 0.8$ using the formula.

$$b(r;10,0.8) = \binom{10}{r} 0.8^r (1-0.8)^{10-r}$$

Binomial Probabilities for *n* = 10 and *p* = 0.8

R	Prob(r = R)	Cumulative Total	1 - Cumulative Total
0	0.000000	0.000000	1.000000
1	0.000004	0.000004	0.999996
2	0.000074	0.000078	0.999922
3	0.000786	0.000864	0.999136
4	0.005505	0.006369	0.993631
5	0.026424	0.032793	0.967207
6	0.088080	0.120874	0.879126
7	0.201327	0.322200	0.677800
8	0.301990	0.624190	0.375810
9	0.268435	0.892626	0.107374
10	0.107374	1.000000	0.000000

The power is the probability of rejecting the null hypothesis. This occurs when $r = 0$, 1, 9, or 10. From the above table, we compute the power as 0.000000 + 0.000004 + 0.268435 + 0.107374 = 0.375813. This matches the calculated power value as displayed in the results above to within rounding error.

Example 5 – Validation using Zar

Zar (1984) page 388 gives the results of a power analysis. When *n* = 12, P0 = 0.50, P1 = 0.83, and alpha = 0.05 using a one-sided test, Zar reports a power of 0.666.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for One Proportion [Proportions]** procedure window by expanding **Proportions**, then **One Proportion**, then clicking on **Test (Inequality)**, and then clicking on **Tests for One Proportion [Proportions]**. You may then make the appropriate entries as listed below, or open **Example 5** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power or Beta	*Ignored since this is the Find setting*
Alpha	**0.05**
n (Sample Size)	**12**
P0 (Null Proportion)	**0.5**
P1 (Alternative Proportion)	**0.83**
H1 (Alternative Hypothesis)	**H1: P > P0**
Test Type	**Exact Test**
N (Population Size)	**Infinite**

Output

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

Numeric Results

Numeric Results for H0: P = P0 versus Ha: P = P1>P0

Power	N	Proportion Given H0 (P0)	Proportion Given H1 (P1)	Target Alpha	Actual Alpha	Beta	Reject H0 If R>=This
0.6656	12	0.5000	0.8300	0.0500	0.0193	0.3344	10

PASS calculated the power as 0.6656, which agrees with Zar's value of 0.666.

Example 6 – Comparing Test Statistics

One important decision that must be made before conducting the experiment is to decide which of the available test statistics to use. This procedure makes it easy to make this comparison. The parameter settings will be set as they were in Example 1 except that the alternative proportion is set to 0.60 and the sample sizes are 10, 11, 12, 25, 50, and 70.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for One Proportion [Proportions]** procedure window by expanding **Proportions**, then **One Proportion**, then clicking on **Test (Inequality)**, and then clicking on **Tests for One Proportion [Proportions]**. You may then make the appropriate entries as listed below, or open **Example 6** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power or Beta	*Ignored since this is the Find setting*
Alpha	**0.05**
n (Sample Size)	**10 11 12 25 50 70**
P0 (Null Proportion)	**0.50**
P1 (Alternative Proportion)	**0.60**
H1 (Alternative Hypothesis)	**H1: P ≠ P0**
Test Type	**Exact Test**
N (Population Size)	**Infinite**
Reports Tab	
Show Comparative Reports	**Checked**
Show Comparative Plots	**Checked**

Output

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

Numeric Results and Plots

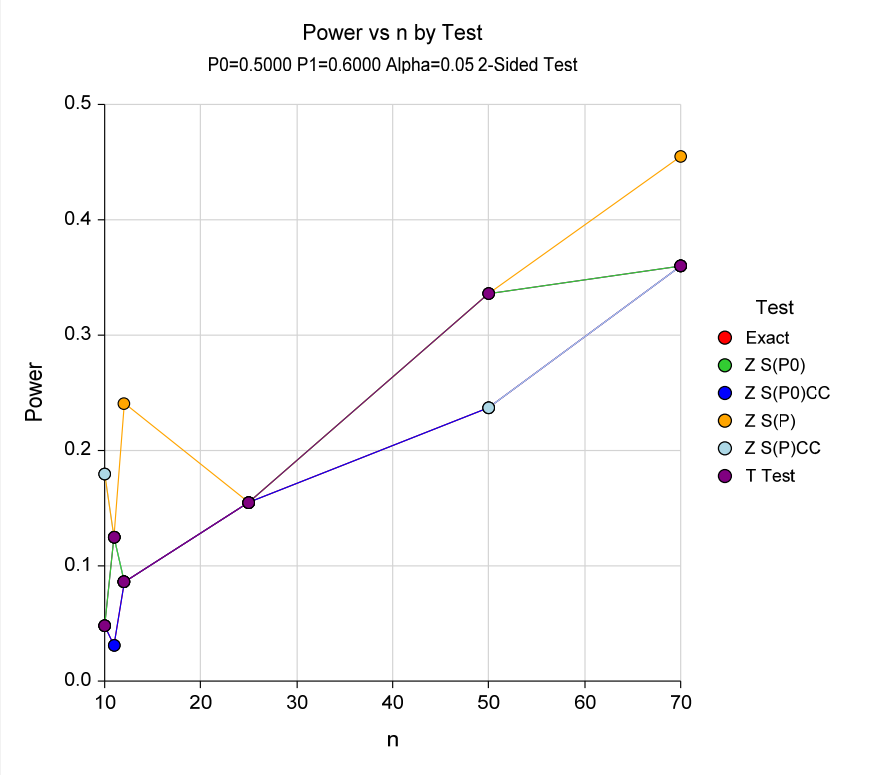
Power Comparison for Methods of Testing H0: P = P0 versus H1: P <> P0

N	P0	P1	Alpha	Exact Test Power	Z-Test S(P0) Power	Z-Test S(P0)C Power	Z-Test S(P) Power	Z-Test S(P)C Power	T-Test Power
10	0.5000	0.6000	0.0500	0.0480	0.0480	0.0480	0.1796	0.1796	0.0480
11	0.5000	0.6000	0.0500	0.0310	0.1248	0.0310	0.1248	0.1248	0.1248
12	0.5000	0.6000	0.0500	0.0863	0.0863	0.0863	0.2406	0.0863	0.0863
25	0.5000	0.6000	0.0500	0.1548	0.1548	0.1548	0.1548	0.1548	0.1548
50	0.5000	0.6000	0.0500	0.2371	0.3361	0.2371	0.3361	0.2371	0.3361
70	0.5000	0.6000	0.0500	0.3601	0.3601	0.3601	0.4549	0.3601	0.3601

Actual Alpha Comparison for Methods of Testing H0: P = P0 versus H1: P <> P0

N	P0	P1	Target Alpha	Exact Test Alpha	Z-Test S(P0) Alpha	Z-Test S(P0)C Alpha	Z-Test S(P) Alpha	Z-Test S(P)C Alpha	T-Test Alpha
10	0.5000	0.6000	0.0500	0.0215	0.0215	0.0215	0.1094	0.1094	0.0215
11	0.5000	0.6000	0.0500	0.0117	0.0654	0.0117	0.0654	0.0654	0.0654
12	0.5000	0.6000	0.0500	0.0386	0.0386	0.0386	0.1460	0.0386	0.0386
25	0.5000	0.6000	0.0500	0.0433	0.0433	0.0433	0.0433	0.0433	0.0433
50	0.5000	0.6000	0.0500	0.0328	0.0649	0.0328	0.0649	0.0328	0.0649
70	0.5000	0.6000	0.0500	0.0414	0.0414	0.0414	0.0722	0.0414	0.0414

Chart Section



An examination of the first report shows that for most sample sizes, the power is different for at least one of the tests. Also, notice that the exact test always has the minimum power in each row. This would lead us discard this test statistic. However, consider the second report, which shows the actual alpha level (the target was 0.05) for each test. By inspecting corresponding entries in both tables, we can see that whenever a test statistic achieves a better power than the exact test, it also exceeds the target alpha. For example, look at the powers for n = 12. The z test using s(p hat) has an unusually large power of 0.2406. This is a much larger power than the exact test's value of 0.0863. However, note that the actual alpha level for this test is 0.1460, which is much higher than the target of 0.05 and the actual value of the other tests, which is 0.0386.

We conclude that indeed, the exact test is consistently the best test since it always achieves a significance level that is less than the target value.

Example 7 – Finding the Power using Ratios

Suppose that only 5% of patients with an aggressive type of cancer respond to the standard treatment. Researchers have found a new treatment which could be widely used if the percentage of patients responding is at least 0.5 times greater than the proportion responding to the standard treatment, i.e. P1 = 1.5(P0), or in terms of ratios, P1/P0 = 1.5. What power will be achieved for trials with sample sizes of 200, 300, 500, or 800 and a significance level of 0.05? For comparative purposes, also calculate the power for alternative ratios of 1.25 and 1.75.

Setup

This section presents the values of each of the parameters for this example. From the PASS Home window, load the **Tests for One Proportion [Ratios]** procedure window by expanding **Proportions**, then **One Proportion**, then clicking on **Test (Inequality)**, and then clicking on **Tests for One Proportion [Ratios]**. You may then make the appropriate entries listed below, or open **Example 7** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power or Beta	*Ignored since this is the Find setting*
Alpha	**0.05**
n (Sample Size)	**200 300 500 800**
P0 (Null Proportion)	**0.05**
Alternative Ratio (P1/P0)	**1.25 1.50 1.75**
H1 (Alternative Hypothesis)	**H1: Ratio ≠ 1**
Test Type	**Exact Test**
N (Population Size)	**Infinite**

Output

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

Numeric Results

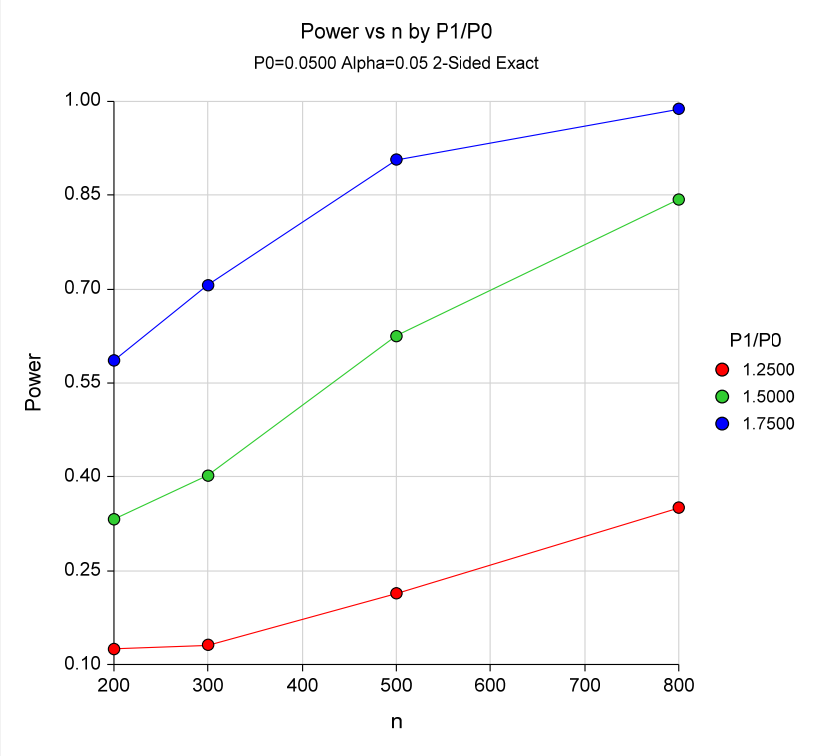
Numeric Results for testing H0: P = P0 versus H1: P <> P0
Test Statistic: Exact Test

Power	N	Proportion Given H0 (P0)	Proportion Given H1 (P1)	Ratio (P1 / P0)	Target Alpha	Actual Alpha	Beta	Reject H0 If <=R|>=R
0.1247	200	0.0500	0.0625	1.2500	0.0500	0.0328	0.8753	3|17
0.1315	300	0.0500	0.0625	1.2500	0.0500	0.0328	0.8685	7|24
0.2138	500	0.0500	0.0625	1.2500	0.0500	0.0395	0.7862	15|36
0.3509	800	0.0500	0.0625	1.2500	0.0500	0.0420	0.6491	27|53
0.3322	200	0.0500	0.0750	1.5000	0.0500	0.0328	0.6678	3|17
0.4019	300	0.0500	0.0750	1.5000	0.0500	0.0328	0.5981	7|24
0.6248	500	0.0500	0.0750	1.5000	0.0500	0.0395	0.3752	15|36
0.8432	800	0.0500	0.0750	1.5000	0.0500	0.0420	0.1568	27|53
0.5861	200	0.0500	0.0875	1.7500	0.0500	0.0328	0.4139	3|17
0.7062	300	0.0500	0.0875	1.7500	0.0500	0.0328	0.2938	7|24
0.9072	500	0.0500	0.0875	1.7500	0.0500	0.0395	0.0928	15|36
0.9882	800	0.0500	0.0875	1.7500	0.0500	0.0420	0.0118	27|53

This report shows the values of each of the parameters, one scenario per row. Because of the discrete nature of the binomial distribution, the stated (Target) alpha is usually greater than the actual alpha. Hence, we also show the Actual Alpha along with the rejection region.

The values from this table are plotted in the chart below.

Plots Section



This plot shows the relationship between power and P1/P0 in this example. We note that 80% power is achieved with a sample size of about 720 when P1/P0 is 1.50, which was the specific ratio of interest.

Example 8 – Determining the Power after Completing an Experiment

A group of researchers is studying the effects of a new diet on cholesterol levels in high-risk patients. The researchers had hypothesized that the cholesterol level would be reduced to a safe level in more than 70% of subjects following the new diet. They are confident that the proportion will be no less than 0.70. To test this one-sided hypothesis, they randomly sampled 200 individuals with dangerously high cholesterol and put them on the new diet. After the period of the study, the researchers determined that 150 of the 200 patients sampled (75%) had reduced their cholesterol level while on the new diet. Statistical analysis using the exact test and an alpha level of 0.05, however, resulted in failure to reject the null hypothesis that the proportion is 0.70. The researchers desire now to compute the power of their study for true proportions ranging from 0.71 to 0.80.

Note that a range of proportions is considered for power calculations instead of just 0.75, the sample proportion found in the experiment. While it is tempting to use the sample proportion as the true proportion in post-experiment power calculations, it is more informative to review a range of possible alternatives representing practically significant differences from the null value.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for One Proportion [Proportions]** procedure window by expanding **Proportions**, then **One Proportion**, then clicking on **Test (Inequality)**, and then clicking on **Tests for One Proportion [Proportions]**. You may then make the appropriate entries as listed below, or open **Example 8** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power or Beta	*Ignored since this is the Find setting*
Alpha	**0.05**
n (Sample Size)	**200**
P0 (Null Proportion)	**0.70**
P1 (Alternative Proportion)	**0.71 to 0.80 by 0.01**
H1 (Alternative Hypothesis)	**H1: P > P0**
Test Type	**Exact Test**
N (Population Size)	**Infinite**

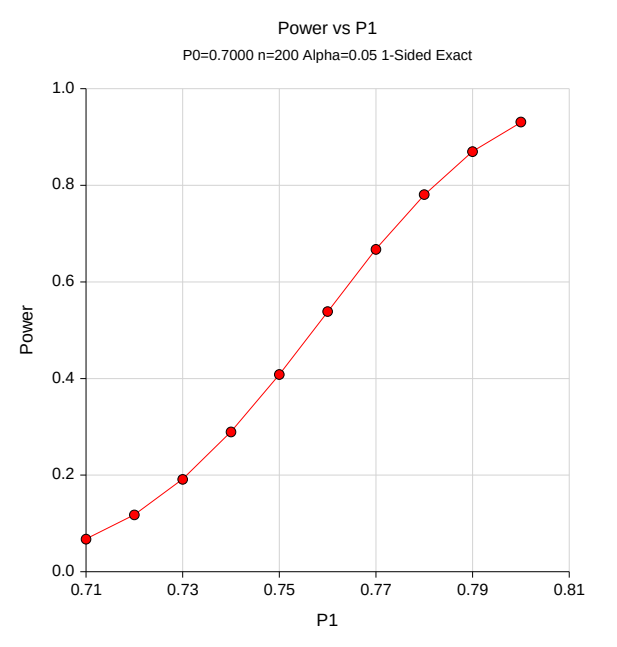
Output

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

Numeric Results

Numeric Results for testing H0: P = P0 versus H1: P > P0
Test Statistic: Exact Test

Power	N	Proportion Given H0 (P0)	Proportion Given H1 (P1)	Target Alpha	Actual Alpha	Beta	Reject H0 If R>=This
0.0675	200	0.7000	0.7100	0.0500	0.0359	0.9325	152
0.1178	200	0.7000	0.7200	0.0500	0.0359	0.8822	152
0.1913	200	0.7000	0.7300	0.0500	0.0359	0.8087	152
0.2894	200	0.7000	0.7400	0.0500	0.0359	0.7106	152
0.4083	200	0.7000	0.7500	0.0500	0.0359	0.5917	152
0.5386	200	0.7000	0.7600	0.0500	0.0359	0.4614	152
0.6673	200	0.7000	0.7700	0.0500	0.0359	0.3327	152
0.7807	200	0.7000	0.7800	0.0500	0.0359	0.2193	152
0.8696	200	0.7000	0.7900	0.0500	0.0359	0.1304	152
0.9310	200	0.7000	0.8000	0.0500	0.0359	0.0690	152



Power ranges from 0.0675 for a true proportion of 0.71 to 0.9310 for a true proportion of 0.80.

Chapter 103

Tests for One Proportion with Non-Zero Null Hypothesis

Introduction

This module provides power analysis and sample size calculation for one-sample proportion tests in which the researcher is testing whether the proportion is different from a hypothesized proportion by some specified amount. Users may choose from among six popular test statistics commonly used for running the hypothesis test.

The details of sample size calculation for the one-sample design for binary outcomes are presented in the chapter Inequality Tests for One Proportion and they will not be duplicated here. Instead, this chapter focuses on those changes necessary for non-zero null tests.

Approximate sample size formulas for non-inferiority tests of a single proportion are presented in Chow et al. (2003) page 83. However, only large sample (normal approximation) results are given there. The results available in this module use exact calculations based on the enumeration of all possible values of the binomial distribution.

Four Procedures Documented Here

There are four procedures in the menus that use the program module described in this chapter. These procedures are identical except for the type of parameterization. The parameterization can be in terms of proportions, differences in proportions, ratios of proportions, and odds ratios. Each of these options is listed separately on the menus.

Example

A non-zero null test example will set the stage for the discussion of the terminology that follows. Suppose that the current treatment for a disease is effective 70% of the time. This treatment is inexpensive and exhibits few side-effects. A promising new treatment has been developed to the

point where it can be tested, but the treatment is more expensive and may have more side-effects. One of the basic questions that must be answered is whether the new treatment is effective enough to justify its use over the treatment with the known effectiveness of 70%. In other words, do more than 70% of treated subjects respond to the new treatment?

After thoughtful discussion with several clinicians, it was decided that if a response of at least 78% was achieved, the new treatment would be adopted. The difference between these two percentages may be called the *margin of superiority*. The margin of superiority in this example is 8%.

The developers must design an experiment to test the hypothesis that the response rate of the new treatment is at least 0.78. The statistical hypothesis to be tested is

$$H0: P \le 0.78 \text{ versus } H1: P > 0.78$$

Notice that when the null hypothesis is rejected, the conclusion is that the response rate is at least 0.78. Note that even though the response rate of the current treatment is 0.70, the hypothesis test is about a response rate of 0.78. Also notice that a rejection of the null hypothesis results in the conclusion of interest.

Technical Details

In the discussion that follows, let *P* represent the proportion responding as a success. That is, *P* is the actual probability of a *success* in a binomial experiment. Let PB represent the *baseline* proportion. In a non-zero null experiment, the baseline proportion is the response rate of the current treatment. Furthermore, let P0 represent the response proportion that is tested in the null hypothesis, H0. The power of a test is computed at a specific value of the proportion. Let P1 represent the proportion at which the power is computed.

Let *PS* represent the smallest value of *P* that still results in the conclusion that the new treatment is superior to the current treatment. The statistical hypotheses that are tested are

$$H0: P \le PS \text{ versus } H1: P > PS$$

There are three common methods of specifying the margin of superiority. The most direct is to simply assign values for PB and PS. However, it is often more meaningful to identify PB and then specify PS implicitly by giving their difference, ratio, or odds ratio. Mathematically, the definitions of these parameterizations are

Parameter	Computation	Hypotheses
Difference	$d0 = PS - PB$	$H0: P \le PB + d0$ vs $H1: P > PB + d0$
Ratio	$r0 = PS / PB$	$H0: P \le r0(PB)$ vs $H1: P > r0(PB)$
Odds Ratio	$o0 = OddsS / OddsB$	$H0: P \le A$ vs $H_1: P > A$

where $A = \dfrac{(o0)(PB)}{1 + PB(o0 - 1)}$.

Difference

The difference is perhaps the most direct method of comparison between two proportions. It is easy to interpret and communicate. It gives the absolute impact of the treatment. However, there are subtle difficulties that can arise with its interpretation.

One difficulty arises when the event of interest is rare. If a difference of 0.001 occurs when the baseline probability is 0.40, it would be dismissed as being trivial. That is, there is usually little interest in a treatment that only decreases the probability from 0.400 to 0.399. However, if the baseline probability of a disease is 0.002, a 0.001 decrease would represent a reduction of 50%. Thus, interpretation of the difference depends on the baseline probability of the event. As a rule of thumb, the difference is best suited for those cases in which 0.20 < P < 0.80.

Superiority using a Difference

The following example is intended to help you understand the concept of *superiority*. Suppose 60% of patients respond to the current treatment method (PB = 0.60). If the response rate of the new treatment is at least ten percentage points better (d0 = 0.10), it will be considered to be superior to the existing treatment. Substituting these figures into the statistical hypotheses gives $H0: d \le 0.10$ versus $H1: d > 0.10$. The relationship P0 = PB + d0 gives $H0: P \le 0.70$ versus $H1: P > 0.70$.

In this example, when the null hypothesis is rejected, the concluded alternative is that the response rate is at least 0.70. That is, the conclusion of superiority is that the new treatment's response rate is at least 0.10 more than that of the existing treatment.

Ratio

The ratio r0 = PS / PB gives the relative change in the probability of the response. Testing non-zero null hypotheses use the same formulation $H0: r \le r0$ versus $H1: r > r0$.

Superiority using a Ratio

The following example is intended to help you understand the concept of *superiority* as it applies to the ratio. Suppose that 60% of patients (PB = 0.60) respond to the current treatment method. If a new treatment increases the response rate by at least 10% (r0 = 1.10), it will be considered to be superior to the existing treatment. Substituting these figures into the statistical hypotheses gives $H0: r \le 1.10$ versus $H1: r > 1.10$. The relationship P0 = (r0)(PB) gives $H0: P \le 0.66$ versus $H1: P > 0.66$.

In this example, when the null hypothesis is rejected, the concluded alternative is that the response rate is at least 66%. That is, the conclusion of superiority is that the new treatment's response rate is at least 10% more than that of the existing treatment.

Odds Ratio

The odds ratio, o0 = (PE / (1 – PE)) / (PB / (1 – PB)), gives the relative change in the odds of the response. Testing non-zero null hypotheses uses the similar formulation $H0: o \le o0$ versus $H1: o > o0$.

Power and Sample Size Calculation

Historically, power and sample size calculations for a one-sample proportion test have been based on normal approximations to the binomial. However, with the speed of modern computers using the normal approximation is unnecessary, especially for small samples. Rather, the significance level and power can be computed using complete enumeration of all possible values of *x*, the number of successes in a sample of size *n*.

This is done as follows.

1. The critical value of the test is computed using standard techniques.
2. For each possible value of *x*, the value of the test statistic (*z* test, *t* test, or exact test) is computed along with its associated probability of occurrence.
3. The significance level and power are computed by summing the probabilities of occurrence for all values of the test statistic that are greater than (or less than) the critical value. Each probability of occurrence is calculated using P0 for the significance level and P1 for the power.

Other variables such as the sample size are then found using an efficient search algorithm. Although this method is not as elegant as a closed-form solution, it is completely accurate.

Test Statistics

The test statistics used are listed in the Inequality Tests for One Proportion chapter. They will not be repeated here.

Procedure Options

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

Data Tab (Common Options)

The Data tab contains the parameters associated with this test such as the proportions, sample sizes, alpha, and beta. This chapter covers four procedures, each of which has different options. This section documents options that are common to all four procedures. Later, unique options for each procedure will be documented.

Solve For

Find (Solve For)

This option specifies the parameter to be solved for using the other parameters. The parameters that may be selected are *Superiority Value*, *Actual Value*, *n*, *Alpha*, and *Power and Beta*. In most situations, you will select either *Power and Beta* or *n*.

Select *n* when you want to calculate the sample size needed to achieve a given power and alpha level.

Select *Power and Beta* when you want to calculate the power of an experiment.

Error Rates

Power or Beta

This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected.

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

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

Alpha (Significance Level)

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected.

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.

Note that because of the discrete nature of the binomial distribution, the alpha level rarely will be achieved exactly.

A single value may be entered here or a range of values such as *0.05 to 0.2 by 0.05* may be entered.

Sample Size

n (Sample Size)

Enter a value (or range of values) for the sample size *n*. This is the number of individuals sampled in the study. Values must be integers greater than one.

You may a single value or a range of values such as *10, 50, 100* or *10 to 100 by 10*.

Effect Size

PB (Baseline Proportion)

Enter a value (or range of values) for the baseline proportion. In a non-inferiority study, this is the response rate of the standard (existing) treatment. Note that this is not the value of P0. Instead, this value is used in the calculation of P0.

Proportions must be 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 0.1*.

Test

Higher Proportions Are

This option specifies whether proportions represent successes or failures.

- **Better**

  When proportions represent successes, the higher proportions are better. In this case, a superior treatment is one whose proportion is higher than the baseline. The alternative hypothesis of superiority is H1: P > P0, where P0 is greater than PB.

- **Worse**

  When proportions represent failures, the lower proportions are better. In this case, a superior treatment is one whose proportion is lower than the baseline. The alternative hypothesis of superiority is H1: P < P0, where P0 is lower than PB.

Test Type

Specify the type of test that will be used in searching and reporting. Note that *C.C.* is an abbreviation for *Continuity Correction*. This refers to the adding or subtracting of 1/(2*n*) to (or from) the numerator of the z-value to bring the normal approximation closer to the binomial distribution.

In most situations, you would select the Exact Test option. The other options are provided for comparative purposes.

N (Population Size)

Enter the total number of items in the population from which the sample of *n* items is selected. Enter *Infinite* to signify an infinite population so that no correction factor is applied. An *infinite* population is one in which the number in the population is large and unknown.

Note that *N* must be greater than *n*.

When samples are drawn from a very large (infinite) population, calculations are based on the binomial distribution.

When samples are drawn from a population of known size *N*, calculations are based on the hypergeometric distribution.

Data Tab (Proportions)

This section documents options that are used when the parameterization is given directly in terms of the proportions P0 and P1.

Effect Size

P0 (Superiority Proportion)

Enter the value of the proportion that is the superiority cutoff.

If Higher Proportions are "Better", proportions higher than this proportion are considered superior. If Higher Proportions are "Worse", proportions lower than this proportion are considered superior.

The reference proportion (PB) is not needed for calculation of power using this module.

Since this value is a proportion, it must be a positive value less than one.

P1 (Actual Proportion)

This is the value of the proportion (P1) at which the power calculations are made.

If Higher Proportions are "Better", P1 should be larger than P0. If Higher Proportions are "Worse", P1 should be less than P0.

Proportions must be 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 0.1*.

Data Tab (Differences)

This section documents options that are used when the parameterization is in terms of the difference, P1 – P0. P0 is the value of the proportion assumed by the null hypothesis and P1 (or P) is the value of the proportion at which the power is calculated. Once P0 and the difference are given, the value of P1 is found by the formula: P1 = difference + P0.

Effect Size

d0 (Superiority Difference)

Enter the difference from PB (Baseline Proportion) that defines the superiority cutoff.

The resulting superiority cut-off is P0 = PB + d0.

If Higher Proportions are "Better", d0 should be positive. If Higher Proportions are "Worse", d0 should be negative.

PB + d0 must be between 0 and 1.

d1 (Actual Difference)

This option specifies the value of P1 (the actual proportion) by specifying the difference between the two proportions P1 and PB. This difference is used with PB to calculate the value of P1 using the formula: P1 = PB + d1.

The power calculations assume that P1 is the actual value of the proportion.

If Higher Proportions are "Better", d1 should be positive and greater than d0. If Higher Proportions are "Worse", d1 should be negative and less than d0.

Differences must be between -1 and 1.

You may enter a range of values such as *.03 .05 .10* or *.01 to .05 by .01*.

Data Tab (Ratios)

This section documents options that are used when the parameterization is in terms of the ratio, P1 / P0.

Effect Size

r0 (Superiority Ratio)

Enter the ratio of P0 to PB (Baseline Proportion) that defines the superiority cutoff.

The resulting superiority cut-off is P0 = PB * r0.

If Higher Proportions are "Better", r0 should be greater than one. If Higher Proportions are "Worse", r0 should be less than one.

PB * r0 must be between 0 and 1.

r1 (Actual Ratio)

This option specifies the value of P1 (the actual proportion) by specifying the ratio of P1 to PB. This ratio is used with PB to calculate the value of P1 using the formula: P1 = PB*r1.

The power calculations assume that P1 is the actual value of the proportion.

If Higher Proportions are "Better", r1 should be greater than r0. If Higher Proportions are "Worse", r1 should be less than r0.

Ratios must greater than zero. Note that the ratios must be small enough so that P1 is less than one.

You may enter a range of values such as *.5 .6 .7 .8* or *1.25 to 2.0 by .25*.

Data Tab (Odds Ratios)

This section documents options that are used when the parameterization is in terms of the odds ratio, Odds1 / Odds0 where Odds1 = P1 / (1 – P1) and Odds0 = P0 / (1 – P0).

Effect Size

o0 (Superiority Odds Ratio)

Enter the odds ratio of P0 to PB (Baseline Proportion) that defines the superiority cutoff.

o0 = [P0(1-P0)]/[PB(1-PB)]

If Higher Proportions are "Better", o0 should be greater than one. If Higher Proportions are "Worse", o0 should be less than one.

o0 must be such that P0 is between 0 and 1.

o1 (Actual Odds Ratio)

This option specifies the value of P1 (the actual proportion) by specifying the odds ratio of P1 to PB. This ratio is used with PB to calculate the value of P1 using the formula:

o1 = [P1(1-P1)]/[PB(1-PB)]

The power calculations assume that P1 is the actual value of the proportion.

If Higher Proportions are "Better", o1 should be greater than o0. If Higher Proportions are "Worse", o1 should be less than o0.

You may enter a range of values such as ".03 .05 .10" or ".01 to .05 by .01"

o1 must be such that P1 is between 0 and 1.

Iterations Tab

The Iterations tab allows for specification of the maximum number of iterations to be used in searches.

Maximum Iterations

Maximum Iterations Before Search Termination

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 not reported.

Example 1 – Finding the Power

Suppose 50% of patients with a certain type of cancer survive two years using the current treatment. The current treatment is inexpensive and has few severe side effects. A new treatment has more side effects and is more expensive. A superiority trial is to be conducted to show that the two-year survival rate of the new treatment is better than the current treatment. After serious consideration, the margin of superiority is set at 5%. What power will be achieved by sample sizes of 50, 100, 200, 300, 500, or 800 and a significance level of 0.05? For comparative purposes, also calculate the power for a margin of superiority of 10%. Assume that the true survival rate of the new treatment is 62%.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for One Proportion (Non-Zero Null Hypothesis) [Differences]** procedure window by expanding **Proportions**, then **One Proportion**, then clicking on **Test (Non-Zero Null)**, and then clicking on **Tests for One Proportion (Non-Zero Null Hypothesis) [Differences]**. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power or Beta	*Ignored since this is the Find setting*
Alpha	**0.05**
n (Sample Size)	**50 100 200 300 500 800**
PB (Baseline Proportion)	**0.50**
d0 (Superiority Difference)	**0.05 0.10**
d1 (Actual Difference)	**0.12**
Higher Proportions Are	**Better**
Test Type	**Exact Test**
N (Population Size)	**Infinite**
Reports Tab	
Show Numeric Reports	**Checked**
Show Comparative Reports	**Not checked**
Show Definitions	**Checked**
Show Plots	**Checked**
Show Comparative Plots	**Not checked**

Annotated Output

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

Numeric Results

Numeric Results for testing H0: P = P0 versus H1: P > P0
Higher Proportions are Better
Test Statistic: Exact Test

Power	N	Superiority Difference (d0)	Actual Difference (d1)	Baseline Proportion (PB)	Target Alpha	Actual Alpha	Beta	Reject H0 If R>=This
0.2351	50	0.0500	0.1200	0.5000	0.0500	0.0427	0.7649	34
0.3816	100	0.0500	0.1200	0.5000	0.0500	0.0429	0.6184	64
0.5886	200	0.0500	0.1200	0.5000	0.0500	0.0373	0.4114	123
0.7808	300	0.0500	0.1200	0.5000	0.0500	0.0458	0.2192	180
0.9352	500	0.0500	0.1200	0.5000	0.0500	0.0478	0.0648	294
0.9907	800	0.0500	0.1200	0.5000	0.0500	0.0472	0.0093	464
0.0519	50	0.1000	0.1200	0.5000	0.0500	0.0280	0.9481	37
0.0890	100	0.1000	0.1200	0.5000	0.0500	0.0398	0.9110	69
0.1370	200	0.1000	0.1200	0.5000	0.0500	0.0475	0.8630	132
0.1560	300	0.1000	0.1200	0.5000	0.0500	0.0429	0.8440	195
0.2172	500	0.1000	0.1200	0.5000	0.0500	0.0450	0.7828	319
0.2931	800	0.1000	0.1200	0.5000	0.0500	0.0445	0.7069	504

Report Definitions
Power is the probability of rejecting a false null hypothesis.
N is the size of the sample drawn from the population.
d0 is the distance above PB that is considered superior.
d1 is the value of the difference at which the power is calculated.
Alpha is the probability of rejecting a true null hypothesis.
Beta is the probability of accepting a false null hypothesis.
PB is the baseline or standard value of the proportion.

Summary Statements
A sample size of 50 achieves 24% power to detect a superiority difference (P0-PB) of 0.0500 using a one-sided binomial test for superiority. The target significance level is 0.0500. The actual significance level achieved by this test is 0.0427. These results assume a baseline proportion (PB) of 0.5000 and that the actual difference (P1-PB) is 0.1200.

This report shows the values of each of the parameters, one scenario per row. Because of the discrete nature of the binomial distribution, the target alpha is usually greater than the actual alpha. Hence, the actual alpha is also shown.

Power

Power is the probability of concluding superiority when the treatment is indeed superior.

N

This is the sample size.

Superiority Difference (or Proportion, Ratio, or Odds Ratio)

This difference is the minimum difference from the baseline proportion PB that is still considered important. This value is used to calculate P0.

Actual Difference (or Proportion, Ratio, or Odds Ratio)

The actual difference is difference between the true proportion, P1, and the baseline proportion, PB.

Baseline Proportion

The baseline proportion is the response rate that is achieved by the current (standard) treatment.

Target Alpha

This is the target (set in the design) value of the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected. That is, this is the probability of concluding non-inferiority when in fact the new treatment is inferior. Because of the discreteness of the binomial distribution from which this value is calculated, the target value is seldom achieved exactly.

Actual Alpha

This is the actual value of alpha (see Target Alpha) that is achieved by the design. Note that lower values of alpha imply lower power.

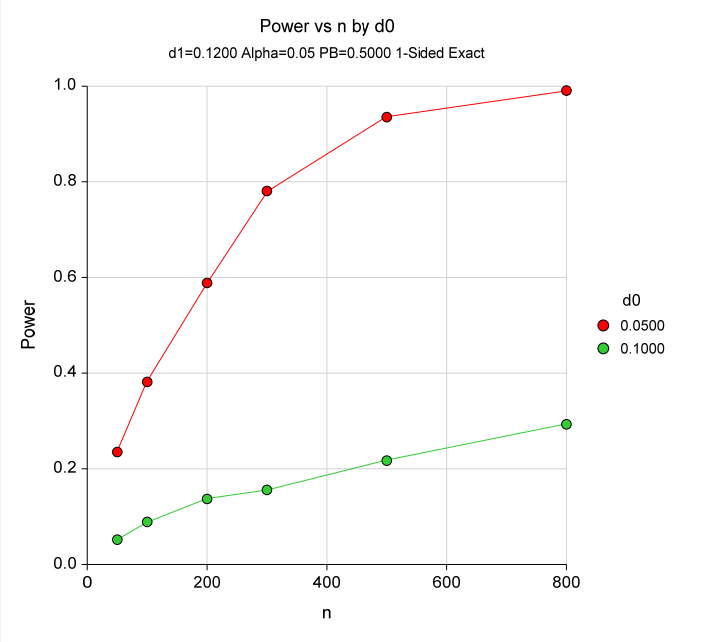
Beta

Beta is the probability of accepting a false null hypothesis. It is the opposite of power.

Reject H0 if R>=This

The symbol *R* stands for the number of items with the characteristic of interest out of the *n* items sampled. For the scenario presented on the first line, an exact test does not exist for these parameters at the target alpha of 0.05. The closest that can be achieved is an alpha of 0.0314. In this case, we would reject the null hypothesis in any sample of size 50 in which the count of individuals with the characteristic of interest is greater than or equal to 27.

Plots Section



This plot shows the relationship between power, sample size, and the minimum difference.

Example 2 – Finding the Sample Size

Continuing from Example 1, suppose you want to find the exact sample size necessary to achieve 90% power when the superiority difference is 0.05. Assume that an exact binomial test will be used.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Non-Inferiority & Superiority Tests for One Proportion [Differences]** procedure window by clicking on **Proportions**, then **One Proportion**, then **Non-Inferiority & Superiority Tests**, then **Specify using Differences**. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**n**
Power or Beta	**0.90**
Alpha	**0.05**
n (Sample Size)	*Ignored since this is the Find setting*
PB (Baseline Proportion)	**0.50**
d0 (Equivalence Difference)	**0.05**
d1 (Actual Difference)	**0**
Higher Proportions Are	**Better**
Test Type	**Exact Test**
N (Population Size)	**Infinite**
Reports Tab	
Show Numeric Reports	**Checked**
Show Comparative Reports	**Not checked**
Show Definitions	**Not checked**
Show Plots	**Not checked**
Show Comparative Plots	**Not checked**

Output

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

Numeric Results

Numeric Results for testing H0: P = P0 versus H1: P > P0
Higher Proportions are Better
Test Statistic: Exact Test

Power	N	Superiority Difference (d0)	Actual Difference (d1)	Baseline Proportion (PB)	Target Alpha	Actual Alpha	Beta	Reject H0 If R>=This
0.9030	430	0.0500	0.1200	0.5000	0.0500	0.0493	0.0970	254

This report shows that a sample size of 430 will be necessary to achieve the design requirements.

Example 3 – Validation

These procedures use the same mechanics as the Non-Inferiority Tests for One Proportion procedures. We refer the user to Example 4 of Chapter 105 for the validation.

Chapter 105

Non-Inferiority Tests for One Proportion

Introduction

This module provides power analysis and sample size calculation for non-inferiority tests in one-sample designs in which the outcome is binary. Users may choose from among six popular test statistics commonly used for running the hypothesis test.

The details of sample size calculation for the one-sample design for binary outcomes are presented in the chapter Inequality Tests for One Proportion and they will not be duplicated here. Instead, this chapter focuses on those changes necessary for non-inferiority tests.

Approximate sample size formulas for non-inferiority tests of a single proportion are presented in Chow et al. (2003) page 83. However, only large sample (normal approximation) results are given there. The results available in this module use exact calculations based on the enumeration of all possible values of the binomial distribution.

Four Procedures Documented Here

There are four procedures in the menus that use the program module described in this chapter. These procedures are identical except for the type of parameterization. The parameterization can be in terms of proportions, differences in proportions, ratios of proportions, and odds ratios. Each of these options is listed separately on the menus.

Example

A non-inferiority test example will set the stage for the discussion of the terminology that follows. Suppose that the current treatment for a disease is effective 70% of the time. Unfortunately, this treatment is expensive and occasionally exhibits serious side-effects. A promising new treatment has been developed to the point where it can be tested. One of the first questions that must be answered is whether the new treatment is as good as the current treatment. In other words, do at least 70% of treated subjects respond to the new treatment?

Because of the many benefits of the new treatment, clinicians are willing to adopt the new treatment even if it is slightly less effective than the current treatment. They must determine,

however, how much less effective the new treatment can be to still be adopted. Should it be adopted if 69% respond? 68%? 65%? 60%? There is a percentage below 70% at which the difference between the two treatments is no longer considered ignorable. After thoughtful discussion with several clinicians, it was decided that if a response of at least 63% was achieved, the new treatment would be adopted. The difference between these two percentages is called the *margin of equivalence*. The margin of equivalence in this example is 7%.

The developers must design an experiment to test the hypothesis that the response rate of the new treatment is at least 0.63. The statistical hypothesis to be tested is

$$H0: P \le 0.63 \text{ versus } H1: P > 0.63$$

Notice that when the null hypothesis is rejected, the conclusion is that the response rate is at least 0.63. Note that even though the response rate of the current treatment is 0.70, the hypothesis test is about a response rate of 0.63. Also notice that a rejection of the null hypothesis results in the conclusion of interest.

Technical Details

In the discussion that follows, let *P* represent the proportion responding as a success. That is, *P* is the actual probability of a *success* in a binomial experiment. Let PB represent the *baseline* proportion. In a non-inferiority experiment, the baseline proportion is the response rate of the current treatment. Furthermore, let P0 represent the response proportion that is tested in the null hypothesis, H0. The power of a test is computed at a specific value of the proportion. Let P1 represent the proportion at which the power is computed.

Let *PN* represent the smallest value of *P* that still results in the conclusion that the new treatment is non-inferior to the current treatment. The statistical hypotheses that are tested are

$$H0: P \le PN \text{ versus } H1: P > PN$$

There are three common methods of specifying the margin of non-inferiority. The most direct is to simply assign values for PB and PN. However, it is often more meaningful to identify PB and then specify PN implicitly by giving their difference, ratio, or odds ratio. Mathematically, the definitions of these parameterizations are

Parameter	**Computation**	**Hypotheses**
Difference	$d0 = PN - PB$	$H0: P \le PB + d0$ vs $H1: P > PB + d0$
Ratio	$r0 = PN / PB$	$H0: P \le r0(PB)$ vs $H1: P > r0(PB)$
Odds Ratio	$o0 = OddsN / OddsB$	$H0: P \le A$ vs $H_1: P > A$

where $A = \dfrac{(o0)(PB)}{1 + PB(o0 - 1)}$.

Difference

The difference is perhaps the most direct method of comparison between two proportions. It is easy to interpret and communicate. It gives the absolute impact of the treatment. However, there are subtle difficulties that can arise with its interpretation.

One difficulty arises when the event of interest is rare. If a difference of 0.001 occurs when the baseline probability is 0.40, it would be dismissed as being trivial. That is, there is usually little interest in a treatment that only decreases the probability from 0.400 to 0.399. However, if the baseline probability of a disease is 0.002, a 0.001 decrease would represent a reduction of 50%. Thus, interpretation of the difference depends on the baseline probability of the event. As a rule of thumb, the difference is best suited for those cases in which 0.20 < P < 0.80.

Non-Inferiority using a Difference

The following example might help you understand the concept of a *non-inferiority* test. Suppose 60% of patients respond to the current treatment method (PB = 0.60). If the response rate of the new treatment is no less than five percentage points worse (d0 = -0.05) than the existing treatment, it will be considered noninferior. Substituting these figures into the statistical hypotheses gives $H0: d \le -0.05$ versus $H1: d > -0.05$. The relationship P0 = PB + d0 gives $H0: P \le 0.55$ versus H1: P > 0.55.

In this example, when the null hypothesis is rejected, the concluded alternative is that the response rate is at least 55%.

Ratio

The ratio r0 = PE / PB gives the relative change in the probability of the response. Testing non-inferiority and superiority use the same formulation $H0: r \le r0$ versus $H1: r > r0$.

Non-Inferiority using a Ratio

The following example might help you understand the concept of *non-inferiority* as defined by the ratio. Suppose that 60% of patients (PB = 0.60) respond to the current treatment method. If a new treatment decreases the response rate by no more than 10% (r0 = 0.90), it will be considered to be noninferior to the standard treatment. Substituting these figures into the statistical hypotheses gives $H0: r \le 0.90$ versus $H1: r > 0.90$. The relationship P0 = (r0)(PB) gives $H0: P \le 0.54$ versus $H1: P > 0.54$.

In this example, when the null hypothesis is rejected, the concluded alternative is that the response rate is at least 54%. That is, the conclusion of non-inferiority is that the new treatment's response rate is no worse than 10% less than that of the standard treatment.

Odds Ratio

The odds ratio, o0 = (PE / (1 – PE)) / (PB / (1 – PB)), gives the relative change in the odds of the response. Testing non-inferiority and superiority use the same formulation $H0: o \le o0$ versus $H1: o > o0$. The only difference is that for non-inferiority tests o0 <1, while for superiority tests o0 > 1.

Power and Sample Size Calculation

Historically, power and sample size calculations for a one-sample proportion test have been based on normal approximations to the binomial. However, with the speed of modern computers using the normal approximation is unnecessary, especially for small samples. Rather, the significance level and power can be computed using complete enumeration of all possible values of *x*, the number of successes in a sample of size *n*.

This is done as follows.

1. The critical value of the test is computed using standard techniques.
2. For each possible value of *x*, the value of the test statistic (*z* test, *t* test, or exact test) is computed along with its associated probability of occurrence.
3. The significance level and power are computed by summing the probabilities of occurrence for all values of the test statistic that are greater than (or less than) the critical value. Each probability of occurrence is calculated using P0 for the significance level and P1 for the power.

Other variables such as the sample size are then found using an efficient search algorithm. Although this method is not as elegant as a closed-form solution, it is completely accurate.

Test Statistics

The test statistics used are listed in the Inequality Tests for One Proportion chapter. They will not be repeated here.

Procedure Options

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

Data Tab (Common Options)

The Data tab contains the parameters associated with this test such as the proportions, sample sizes, alpha, and beta. This chapter covers four procedures, each of which has different options. This section documents options that are common to all four procedures. Later, unique options for each procedure will be documented.

Solve For

Find (Solve For)

This option specifies the parameter to be solved for using the other parameters. The parameters that may be selected are *Non-Inferiority Value*, *Actual Value*, *n*, *Alpha*, and *Power and Beta*. In most situations, you will select either *Power and Beta* or *n*.

Select *n* when you want to calculate the sample size needed to achieve a given power and alpha level.

Select *Power and Beta* when you want to calculate the power of an experiment.

Error Rates

Power or Beta

This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected.

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

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

Alpha (Significance Level)

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected.

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.

Note that because of the discrete nature of the binomial distribution, the alpha level rarely will be achieved exactly.

A single value may be entered here or a range of values such as *0.05 to 0.2 by 0.05* may be entered.

Sample Size

n (Sample Size)

Enter a value (or range of values) for the sample size *n*. This is the number of individuals sampled in the study. Values must be integers greater than one.

You may a single value or a range of values such as *10, 50, 100* or *10 to 100 by 10*.

Effect Size

PB (Baseline Proportion)

Enter a value (or range of values) for the baseline proportion. In a non-inferiority study, this is the response rate of the standard (existing) treatment. Note that this is not the value of P0. Instead, this value is used in the calculation of P0.

Proportions must be 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 0.1*.

Test

Higher Proportions Are

This option specifies whether proportions represent successes or failures.

- **Better**

  When proportions represent successes, the higher proportions are better. In this case, a non-inferior treatment is one whose proportion is at least as high as the baseline. The alternative hypothesis of non-inferiority is H1: P > P0, where P0 is slightly less than PB.

- **Worse**

  When proportions represent failures, the lower proportions are better. In this case, a noninferior treatment is one whose proportion is at least as low as the baseline or lower. The alternative hypothesis of non-inferiority is H1: P < P0, where P0 is slightly greater than PB.

Test Type

Specify the type of test that will be used in searching and reporting. Note that *C.C.* is an abbreviation for *Continuity Correction*. This refers to the adding or subtracting of 1/(2*n*) to (or from) the numerator of the z-value to bring the normal approximation closer to the binomial distribution.

In most situations, you would select the Exact Test option. The other options are provided for comparative purposes.

N (Population Size)

Enter the total number of items in the population from which the sample of *n* items is selected. Enter *Infinite* to signify an infinite population so that no correction factor is applied. An *infinite* population is one in which the number in the population is large and unknown.

Note that *N* must be greater than *n*.

When samples are drawn from a very large (infinite) population, calculations are based on the binomial distribution.

When samples are drawn from a population of known size *N*, calculations are based on the hypergeometric distribution.

Data Tab (Proportions)

This section documents options that are used when the parameterization is given directly in terms of the proportions P0 and P1.

Effect Size

P0 (Non-Inferiority Proportion)

This option sets the smallest value which is still trivially different from PB by directly setting the value of P0. If ‘Higher Proportions Are’ is set to ‘Better’, specify a value of P0 that is less than PB.

For example, if PB (baseline proportion) is 0.50, you might consider 0.49, 0.48, 0.47, and 0.46 to be close enough so that the fact that they are less than 0.50 can be overlooked. However, you

might decide that if the value is 0.45 or less, the treatment is inferior. Thus, this value would be set to 0.45.

Since this value is a proportion, it must be a positive value less than one. It cannot be equal to PB.

P1 (Actual Proportion)

This is the value of the proportion (P1) at which the power is calculated. The power calculations assume that this is the actual value of the proportion. For non-inferiority tests, this value is often set equal to PB.

Proportions must be 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 0.1*.

Data Tab (Differences)

This section documents options that are used when the parameterization is in terms of the difference, P1 – P0. P0 is the value of the proportion assumed by the null hypothesis and P1 (or P) is the value of the proportion at which the power is calculated. Once P0 and the difference are given, the value of P1 is found by the formula: P1 = difference + P0.

Effect Size

d0 (Non-Inferiority Difference)

This option sets the smallest value which is still trivially different from PB by setting the difference between P0 and PB. If 'Higher Proportions Are' is set to 'Better', specify a difference that is less than zero.

For example, if PB (baseline proportion) is 0.50, you might consider -.01, -.02, or -.04 to be small enough so that the fact that P0 is less than 0.50 can be overlooked. However, you might decide that if the difference is -.05 or less, the treatment is inferior. Thus, this value would be set to -.05.

Since this value is a difference between two proportions, it must be between -1 and 1.

d1 (Actual Difference)

This option specifies the value of P1 (the proportion at which the power is calculated) by specifying the difference between the two proportions, P1 and PB. This difference is used with PB to calculate the value of P1 using the formula, P1 = PB + Difference. This value is often set equal to zero.

Differences must be between -1 and 1.

You may enter a range of values such as *.03 .05 .10* or *.01 to .05 by .01*.

Data Tab (Ratios)

This section documents options that are used when the parameterization is in terms of the ratio, P1 / P0.

Effect Size

r0 (Non-Inferiority Ratio)

This option sets the smallest value which is still trivially different from PB by setting the ratio of P0 to PB. If 'Higher Proportions Are' is set to 'Better', specify a ratio that is less than one.

For example, if PB (baseline proportion) is 0.50, you might consider ratios of 0.99, 0.98, or even 0.96 to be small enough so that the fact that P0 is less than PB can be overlooked (the difference is trivial). However, you might decide that if the ratio is 0.95 or less, the treatment is inferior. Thus, this value would be set to 0.95.

Since this value is a ratio between two proportions, it must be positive. Since it is a margin, it cannot be one. It cannot be so large that the calculated value of P0 is greater than one.

r1 (Actual Ratio)

This option specifies the value of P1 (the actual proportion) by specifying the ratio between the two proportions, P1 and PB. This ratio is used with PB to calculate the value of P1 using the formula, P1 = (PB)(Ratio). This value is often set equal to one.

Ratios must greater than zero. Note that the ratios must be small enough so that P1 is less than one.

You may enter a range of values such as *.5 .6 .7 .8* or *1.25 to 2.0 by .25*.

Data Tab (Odds Ratios)

This section documents options that are used when the parameterization is in terms of the odds ratio, Odds1 / Odds0 where Odds1 = P1 / (1 – P1) and Odds0 = P0 / (1 – P0).

Effect Size

o0 (Non-Inferiority Odds Ratio)

This option sets the smallest value that is still trivially different from PB by setting the odds ratio of P0 and PB. If 'Higher Proportions Are' is set to 'Better', specify a ratio that is less than one.

For example, if PB (baseline proportion) is 0.50, you might consider odds ratios of 0.99, 0.90, or even 0.81 to be small enough so that the fact that P0 is less than PB can be overlooked (the difference is trivial). However, you might decide that if the odds ratio is 0.80 or less, the treatment is inferior. Thus, this value would be set to 0.80.

Since this value is a ratio between two odds, it must be positive. Since it is a margin, it cannot be one.

o1 (Actual Odds Ratio)

This option specifies the value of P1 (the actual proportion) by specifying the odds ratio between the two proportions, P1 and PB. This ratio is used with PB to calculate the value of P1. For non-inferiority tests, this value is often set equal to one.

Odds ratios must greater than zero.

You may enter a range of values such as *.5 .6 .7 .8* or *1.25 to 2.0 by .25*.

Iterations Tab

The Iterations tab allows for specification of the maximum number of iterations to be used in searches.

Maximum Iterations

Maximum Iterations Before Search Termination

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 not reported.

Example 1 – Finding the Power

Suppose 50% of patients with a certain type of cancer survive two years using the current treatment. The current treatment is expensive and has several severe side effects. A new treatment has fewer side effects and is less expensive. A non-inferiority trial is to be conducted to show that the two-year survival rate of the new treatment is as good as the current treatment. After serious consideration, the margin of non-inferiority is set at 5%. What power will be achieved by sample sizes of 50, 100, 200, 300, 500, or 800 and a significance level of 0.05? For comparative purposes, also calculate the power for a margin of non-inferiority of 10%. Assume that the true survival rate of the new treatment is the same as that of the current (baseline) treatment.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Non-Inferiority Tests for One Proportion [Differences]** procedure window by expanding **Proportions**, then **One Proportion**, then clicking on **Non-Inferiority**, and then clicking on **Non-Inferiority Tests for One Proportion [Differences]**. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power or Beta	*Ignored since this is the Find setting*
Alpha	**0.05**
n (Sample Size)	**50 100 200 300 500 800**
PB (Baseline Proportion)	**0.50**
d0 (Equivalence Difference)	**-.10 -.05**
d1 (Actual Difference)	**0**
Higher Proportions Are	**Better**
Test Type	**Exact Test**
N (Population Size)	**Infinite**
Reports Tab	
Show Numeric Reports	**Checked**
Show Comparative Reports	**Not checked**
Show Definitions	**Checked**
Show Plots	**Checked**
Show Comparative Plots	**Not checked**

Annotated Output

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

Numeric Results

Numeric Results when H0: P = P0 versus H1: P = P>P0
Test Statistic: Exact Test

Power	N	Non-Inf. Difference (d0)	Actual Difference (d1)	Baseline Proportion (PB)	Target Alpha	Actual Alpha	Beta	Reject H0 If R>=This
0.3359	50	-0.1000	0.0000	0.5000	0.0500	0.0314	0.6641	27
0.6178	100	-0.1000	0.0000	0.5000	0.0500	0.0423	0.3822	49
0.8854	200	-0.1000	0.0000	0.5000	0.0500	0.0492	0.1146	92
0.9633	300	-0.1000	0.0000	0.5000	0.0500	0.0443	0.0367	135
0.9976	500	-0.1000	0.0000	0.5000	0.0500	0.0461	0.0024	219
1.0000	800	-0.1000	0.0000	0.5000	0.0500	0.0453	0.0000	344
0.1611	50	-0.0500	0.0000	0.5000	0.0500	0.0444	0.8389	29
0.2421	100	-0.0500	0.0000	0.5000	0.0500	0.0441	0.7579	54
0.3619	200	-0.0500	0.0000	0.5000	0.0500	0.0381	0.6381	103
0.5230	300	-0.0500	0.0000	0.5000	0.0500	0.0465	0.4770	150
0.7195	500	-0.0500	0.0000	0.5000	0.0500	0.0484	0.2805	244
0.8783	800	-0.0500	0.0000	0.5000	0.0500	0.0476	0.1217	384

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.
Equiv. is the maximum value that is still considered unimportant.
Actual is the value of this parameter given the alternative hypothesis is true.
PB is the baseline or standard value of the proportion. This is the value under the current treatment.
d0 is the smallest difference from PB which is still considered as equivalent.
d1 is the value of the difference 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 50 achieves 34% power to detect a difference (P0-PB) of -0.1000 using a one-sided binomial test. The target significance level is 0.0500. The actual significance level achieved by this test is 0.0314. These results assume a baseline proportion (PB) of 0.5000 and that the actual difference (P1-PB) is 0.0000.

This report shows the values of each of the parameters, one scenario per row. Because of the discrete nature of the binomial distribution, the target alpha is usually greater than the actual alpha. Hence, the actual alpha is also shown.

Power

Power is the probability of concluding non-inferiority when the treatment is indeed non-inferior.

N

This is the sample size.

Non-Inferiority Difference (or Proportion, Ratio, or Odds Ratio)

This difference is the maximum difference from the baseline proportion PB that is still considered unimportant or trivial. This value is used to calculate P0.

Actual Difference (or Proportion, Ratio, or Odds Ratio)

The actual difference is difference between the true proportion, P1, and the baseline proportion, PB.

Baseline Proportion

The baseline proportion is the response rate that is achieved by the current (standard) treatment.

Target Alpha

This is the target (set in the design) value of the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected. That is, this is the probability of concluding non-inferiority when in fact the new treatment is inferior. Because of the discreteness of the binomial distribution from which this value is calculated, the target value is seldom achieved exactly.

Actual Alpha

This is the actual value of alpha (see Target Alpha) that is achieved by the design. Note that lower values of alpha imply lower power.

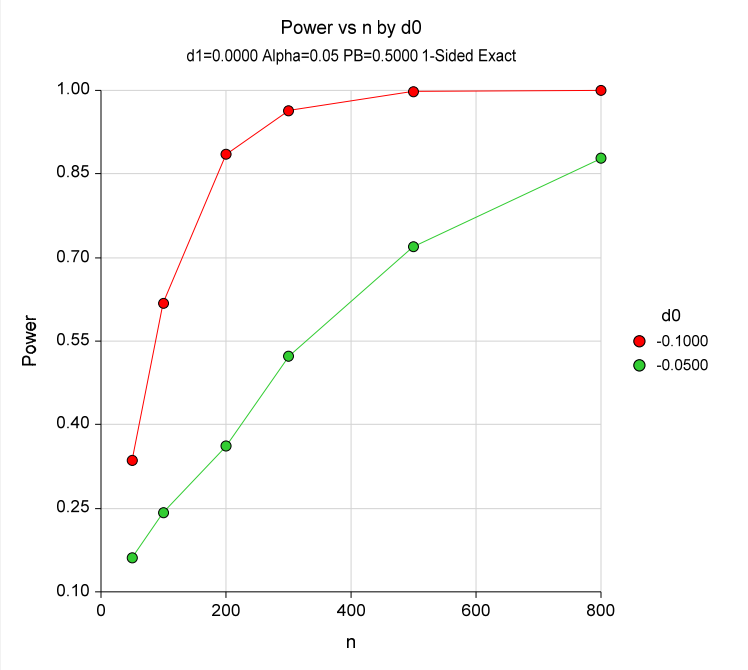
Beta

Beta is the probability of accepting a false null hypothesis. It is the opposite of power.

Reject H0 if R>=This

The symbol *R* stands for the number of items with the characteristic of interest out of the *n* items sampled. For the scenario presented on the first line, an exact test does not exist for these parameters at the target alpha of 0.05. The closest that can be achieved is an alpha of 0.0314. In this case, we would reject the null hypothesis in any sample of size 50 in which the count of individuals with the characteristic of interest is greater than or equal to 27.

Plots Section



This plot shows the relationship between power, sample size, and the trivial difference. Note that 90% power is achieved with an *n* of about 200 when the trivial difference is -.10 and about 800 when the trivial difference is -.05.

Example 2 – Finding the Sample Size

Continuing from Example 1, suppose you want to find the exact sample size necessary to achieve 90% power when the equivalence difference is -.05. Assume that an exact binomial test will be used.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Non-Inferiority Tests for One Proportion [Differences]** procedure window by expanding **Proportions**, then **One Proportion**, then clicking on **Non-Inferiority**, and then clicking on **Non-Inferiority Tests for One Proportion [Differences]**. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**n**
Power or Beta	**0.90**
Alpha	**0.05**
n (Sample Size)	*Ignored since this is the Find setting*
d0 (Non-Inferiority Difference)	**-.05**
d1 (Actual Difference)	**0**
PB (Baseline Proportion)	**0.50**
Higher Proportions Are	**Better**
Test Type	**Exact Test**
N (Population Size)	**Infinite**
Reports Tab	
Show Numeric Reports	**Checked**
Show Comparative Reports	**Not checked**
Show Definitions	**Not checked**
Show Plots	**Not checked**
Show Comparative Plots	**Not checked**

Output

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

Numeric Results

Numeric Results when H0: P = P0 versus H1: P = P>P0
Test Statistic: Exact Test

Power	N	Non-Inf. Difference (d0)	Actual Difference (d1)	Baseline Proportion (PB)	Target Alpha	Actual Alpha	Beta	Reject H0 If R>=This
0.9024	861	-0.0500	0.0000	0.5000	0.0500	0.0499	0.0976	412

This report shows that a sample size of 861 will be necessary to achieve the design requirements.

Example 3 – Comparing Test Statistics

Continuing Example 1, suppose the researchers want to investigate which of the six test statistics to use. This is an important question since choosing the wrong test statistic may increase the sample size, reduce power or inflate the actual alpha level. The differences in the characteristics of test statistics are most noticeable in small samples. Hence, the investigation done here is for sample sizes of 20 to 200 in steps of 20. The trivial difference will be set to -.10. All other settings are as given in Example 1.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Non-Inferiority Tests for One Proportion [Differences]** procedure window by expanding **Proportions**, then **One Proportion**, then clicking on **Non-Inferiority**, and then clicking on **Non-Inferiority Tests for One Proportion [Differences]**. You may then make the appropriate entries as listed below, or open **Example 3** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power or Beta	*Ignored since this is the Find setting*
Alpha	**0.05**
n (Sample Size)	**20 to 200 by 20**
PB (Baseline Proportion)	**0.50**
d0 (Non-Inferiority Difference)	**-.10**
d1 (Actual Difference)	**0**
Higher Proportions Are	**Better**
Test Type	**Exact Test**
N (Population Size)	**Infinite**
Reports Tab	
Show Numeric Reports	**Not checked**
Show Comparative Reports	**Checked**
Show Definitions	**Not checked**
Show Plots	**Not checked**
Show Comparative Plots	**Checked**

Output

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

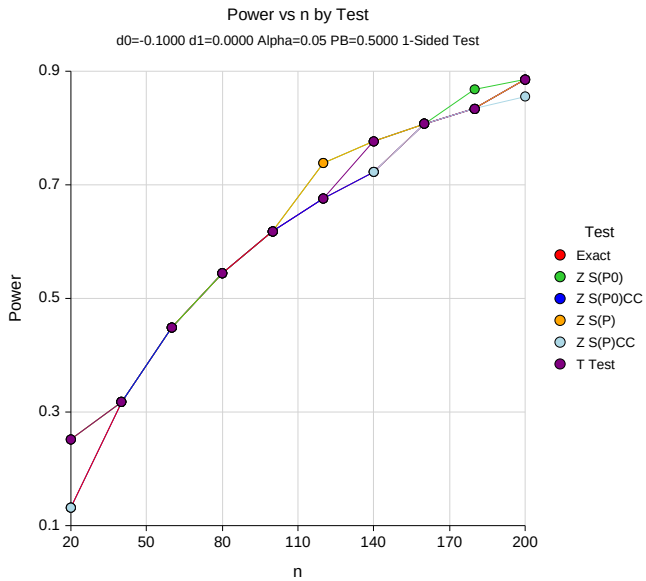
Numeric Results

Power Comparison for Methods of Testing H0: P = P0 versus H1: P > P0

N	Non-Inf. Diff. (d0)	Actual Diff. (d1)	Baseline Prop. (PB)	Target Alpha	Exact Test Power	Z-Test S(P0) Power	Z-Test S(P0)C Power	Z-Test S(P) Power	Z-Test S(P)C Power	T-Test Power
20	-0.1000	0.0000	0.5000	0.0500	0.1316	0.2517	0.1316	0.2517	0.1316	0.2517
40	-0.1000	0.0000	0.5000	0.0500	0.3179	0.3179	0.3179	0.3179	0.3179	0.3179
60	-0.1000	0.0000	0.5000	0.0500	0.4487	0.4487	0.4487	0.4487	0.4487	0.4487
80	-0.1000	0.0000	0.5000	0.0500	0.5445	0.5445	0.5445	0.5445	0.5445	0.5445
100	-0.1000	0.0000	0.5000	0.0500	0.6178	0.6178	0.6178	0.6178	0.6178	0.6178
120	-0.1000	0.0000	0.5000	0.0500	0.6759	0.7385	0.6759	0.7385	0.6759	0.6759
140	-0.1000	0.0000	0.5000	0.0500	0.7229	0.7765	0.7229	0.7765	0.7229	0.7765
160	-0.1000	0.0000	0.5000	0.0500	0.8077	0.8077	0.8077	0.8077	0.8077	0.8077
180	-0.1000	0.0000	0.5000	0.0500	0.8337	0.8683	0.8337	0.8337	0.8337	0.8337
200	-0.1000	0.0000	0.5000	0.0500	0.8854	0.8854	0.8854	0.8854	0.8556	0.8854

Actual Alpha Comparison for Methods of Testing H0: P = P0 versus H1: P > P0

N	Non-Inf. Diff. (d0)	Actual Diff. (d1)	Baseline Prop. (PB)	Target Alpha	Exact Test Alpha	Z-Test S(P0) Alpha	Z-Test S(P0)C Alpha	Z-Test S(P) Alpha	Z-Test S(P)C Alpha	T-Test Alpha
20	-0.1000	0.0000	0.5000	0.0500	0.0210	0.0565	0.0210	0.0565	0.0210	0.0565
40	-0.1000	0.0000	0.5000	0.0500	0.0392	0.0392	0.0392	0.0392	0.0392	0.0392
60	-0.1000	0.0000	0.5000	0.0500	0.0445	0.0445	0.0445	0.0445	0.0445	0.0445
80	-0.1000	0.0000	0.5000	0.0500	0.0445	0.0445	0.0445	0.0445	0.0445	0.0445
100	-0.1000	0.0000	0.5000	0.0500	0.0423	0.0423	0.0423	0.0423	0.0423	0.0423
120	-0.1000	0.0000	0.5000	0.0500	0.0392	0.0575	0.0392	0.0575	0.0392	0.0392
140	-0.1000	0.0000	0.5000	0.0500	0.0358	0.0514	0.0358	0.0514	0.0358	0.0514
160	-0.1000	0.0000	0.5000	0.0500	0.0459	0.0459	0.0459	0.0459	0.0459	0.0459
180	-0.1000	0.0000	0.5000	0.0500	0.0408	0.0558	0.0408	0.0408	0.0408	0.0408
200	-0.1000	0.0000	0.5000	0.0500	0.0492	0.0492	0.0492	0.0492	0.0363	0.0492



The first numeric report shows the power for each test statistic. The second shows the actual alpha achieved by the design.

An examination of the first report shows that the power is often different for at least one of the tests. Also notice that the exact test always has the minimum power in each row. This would lead us discard this test statistic. However, consider the second report which shows the actual alpha level (the target was 0.05) for each test. By inspecting corresponding entries in both tables, we see that whenever a test statistic achieves a better power than the exact test, it also yields an actual alpha level larger than the target alpha.

For example, look at the powers for n = 20. The z test using s(p hat) has an unusually large power = 0.2517. This is a much larger power than the exact test's value of 0.1316. However, note that the actual alpha for this test is 0.0560 which is larger than the target alpha of 0.05.

We conclude that indeed, the exact test is consistently the best test since it always achieves a significance level that is less than the target value.

Example 4 – Validation using Chow, Shao, and Wang

The only appropriate example we have found is Chow, Shao, and Wang (2003) page 85, which gives the result of a sample size calculation using an asymptotic formula. They calculate a sample size of 22 when alpha = 0.05, beta = 0.20, PB = 0.30, Equiv. difference = -.10, and actual proportion = 0.50. As we shall see, ***PASS*** obtains a different answer.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Non-Inferiority Tests for One Proportion [Differences]** procedure window by expanding **Proportions**, then **One Proportion**, then clicking on **Non-Inferiority**, and then clicking on **Non-Inferiority Tests for One Proportion [Differences]**. You may then make the appropriate entries as listed below, or open **Example 4** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**n**
Power or Beta	**0.80**
Alpha	**0.05**
n (Sample Size)	*Ignored since this is the Find setting*
PB (Baseline Proportion)	**0.30**
d0 (Non-Inferiority Difference)	**-.10**
d1 (Actual Difference)	**0.20**
Higher Proportions Are	**Better**
Test Type	**Exact Test**
N (Population Size)	**Infinite**
Reports Tab	
Show Numeric Reports	**Checked**

Output

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

Numeric Results

Power	N	Non-Inf. Difference (d0)	Actual Difference (d1)	Baseline Proportion (PB)	Target Alpha	Actual Alpha	Beta	Reject H0 If R>=This
0.8338	17	-0.1000	0.5000	0.3000	0.0500	0.0377	0.1662	7

PASS calculated a sample size of only 17 while Chow's formula obtained 22. The difference occurs because ***PASS*** uses exact calculations based on the binomial distribution, while Chow et al. use a large-sample approximation based on the normal approximation to the binomial. To see that 17 is indeed the correct answer, enter the values into ***PASS***'s one-sample proportion test. The necessary values are P0 = 0.20, P1 = 0.50, alpha = 0.05, and beta = 0.10. These values result in a sample size of 17.

We have found that the approximate results are closer to the exact results when the sample sizes are over 200. For sample sizes less than 50, there can be significant error in the approximate formulas.

Example 5 – Finding Power after an Experiment

The proportion of successes of the current treatment is known to be 0.74 based on years of treatment use. Researchers have developed a new method of treatment which costs about half the current treatment price. Before the new treatment can be approved it must be shown that the success of the proposed treatment is not inferior to that of the current treatment. It is determined that use of the new treatment is justifiable if it is shown that it is effective more than 70% of the time. Sixty individuals are randomly selected to receive the new method of treatment. Forty-three (71.67%) of the 60 individuals responded positively to the treatment. The p-value for the test based on exact binomial probabilities is 0.4514. Because the researchers were unable to show the new treatment is non-inferior, they desire to know the power of the test.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Non-Inferiority Tests for One Proportion [Differences]** procedure window by expanding **Proportions**, then **One Proportion**, then clicking on **Non-Inferiority**, and then clicking on **Non-Inferiority Tests for One Proportion [Differences]**. You may then make the appropriate entries as listed below, or open **Example 5** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power or Beta	*Ignored since this is the Find setting*
Alpha	**0.05**
n (Sample Size)	**60**
PB (Baseline Proportion)	**0.74**
d0 (Non-Inferiority Difference)	**-.04**
d1 (Actual Difference)	**0.00**
Higher Proportions Are	**Better**
Test Type	**Exact Test**
N (Population Size)	**Infinite**

Output

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

Numeric Results

Numeric Results for testing H0: P = P0 versus H1: P > P0
Test Statistic: Exact Test

Power	N	Non-Inf. Difference (d0)	Actual Difference (d1)	Baseline Proportion (PB)	Target Alpha	Actual Alpha	Beta	Reject H0 If R>=This
0.1112	60	-0.0400	0.0000	0.7400	0.0500	0.0295	0.8888	49

This report shows that the power for the test was only 0.1112.

Example 6 – Finding Sample Size based on the Odds Ratio

The odds for success of the current treatment is known be 4.31. A new treatment is developed to compete with the current treatment with respect to cost and reduction in side effects. It must be shown to be non-inferior to the current treatment. The researchers want to determine the sample size necessary to achieve 80% power in this test of non-inferiority. The researchers determine that the new treatment will be considered non-inferior if the odds for success are no less than 90% the odds for success of the current treatment. The baseline proportion is calculated as PB = odds/(1+odds) = 4.31/(1+4.31) = 0.8117.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Non-Inferiority Tests for One Proportion [Odds Ratios]** procedure window by expanding **Proportions**, then **One Proportion**, then clicking on **Non-Inferiority**, and then clicking on **Non-Inferiority Tests for One Proportion [Odds Ratios]**. You may then make the appropriate entries as listed below, or open **Example 6** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**n**
Power or Beta	**0.8**
Alpha	**0.05**
n (Sample Size)	*Ignored since this is the Find setting*
PB (Baseline Proportion)	**0.8117**
o0 (Non-Inferiority Odds Ratio)	**0.9**
o1 (Actual Odds Ratio)	**1.0**
Higher Proportions Are	**Better**
Test Type	**Exact Test**
N (Population Size)	**Infinite**

Output

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

Numeric Results

Numeric Results for testing H0: P = P0 versus H1: P > P0
Test Statistic: Exact Test

Power	N	Non-Inf. Odds Ratio (o0)	Actual Odds Ratio (o1)	Baseline Proportion (PB)	Target Alpha	Actual Alpha	Beta	Reject H0 If R>=This
0.8004	3547	0.9000	1.0000	0.8117	0.0500	0.0499	0.1996	2860

A sample size of 3547 is required to show non-inferiority under these conditions.

Chapter 110

Equivalence Tests for One Proportion

Introduction

This module provides power analysis and sample size calculation for equivalence tests in one-sample designs in which the outcome is binary. Users may choose from among commonly-used test statistics.

The details of sample size calculations for the one-sample design for binary outcomes are presented in the chapter entitled Inequality Tests for One Proportion and will not be repeated here. Instead, this chapter discusses those changes necessary for equivalence tests.

Approximate sample size formulas for equivalence tests of a single proportion are presented in Chow et al. (2003) page 83. However, only large sample (normal approximation) results are given there. The results available in this module use exact calculations based on the enumeration of all possible values for the binomial distribution.

Four Procedures Documented Here

There are four procedures in the menus that use the program module described in this chapter. These procedures are identical except for the type of parameterization. The parameterization can be in terms of proportions, differences in proportions, ratios of proportions, and odds ratios. Each of these options is listed separately on the menus.

Example

An equivalence test example will set the stage for the discussion of the terminology that follows. Suppose that the current treatment for a disease is effective 70% of the time. Unfortunately, this treatment is expensive and occasionally exhibits serious side-effects. A promising new treatment has been developed to the point where it can be tested. One of the questions that must be answered is whether the new treatment is equivalent to the current treatment. In other words, do about 70% of treated subjects respond to the new treatment?

It is known that the new treatment will not have a response rate that is exactly the same as that of the standard treatment. After careful consideration, they decide that the margin of equivalence is plus or minus 10%. That is, if the response rate of the new treatment is between 60% and 80% it will be deemed equivalent to the standard treatment.

The developers must design an experiment to test the hypothesis that the response rate of the new treatment is within 10% of the standard (baseline) treatment. The statistical hypotheses to be tested are

$$H_0: |P - PB| \geq 0.1 \text{ versus } H_1: |P - PB| < 0.1$$

Notice that when the null hypothesis is rejected the conclusion is that the response rate is between 0.6 and 0.8.

Technical Details

In the discussion that follows, let *P* represent the proportion being investigated. That is, *P* is the actual probability of a *success* in a binomial experiment. Often, this proportion is a response rate, cure rate, or survival rate. Let PB represent the *baseline* proportion. In an equivalence trial, the baseline proportion is the response rate of the current (standard) treatment. Let PL represent the smallest value of *P* that still results in the conclusion that the new treatment is equivalent to the current treatment. Similarly, let PU represent the largest value of *P* that still results in the conclusion that the new treatment is equivalent to the current treatment. Note that PB will be between PL and PU. The power of a test is computed at a specific value of the proportion, P1.

The statistical hypotheses that are tested are

$$H0: P \leq PL \text{ or } P \geq PU \text{ versus } H1: PL < P < PU$$

This unusual hypothesis test can be broken down into two, one-sided hypothesis tests (TOST) as follows

$$H0: P \leq PL \text{ versus } H1: P > PL$$

and

$$H0: P \geq PU \text{ versus } H1: P < PU$$

If both of these one-sided tests are rejected at significance level α, then equivalence can be concluded at significance level α. Note that we do not conduct the individual tests at $\alpha / 2$.

There are three common methods of specifying the margin of equivalence. The most direct is to simply assign values for PL and PU. However, it is often more meaningful to identify PB and then specify PL and PU implicitly by giving a difference, ratio, or odds ratio. Mathematically, the definitions of these parameterizations are

Parameter	Computation	Hypotheses
Difference	$d0 = \lvert PL - P_B \rvert = PU - PB$	$H0: d \geq d0$ vs $H1: d < d0$
Ratio	$r0 = \dfrac{1}{PL / PB} = PU / PB$	$H0: r \geq r0$ vs $H1: r < r0$
Odds Ratio	$o0 = \dfrac{1}{OddsL / OddsB} = \dfrac{OddsU}{OddsB}$	$H0: o \geq o0$ vs $H1: o < o0$

where

Difference $d = |P - PB|$

Ratio $r = \begin{cases} P / PB & \text{if } P > PB \\ PB / P & \text{if } P < PB \end{cases}$

Odds Ratio $o = \begin{cases} Odds / OddsB & \text{if } P > PB \\ OddsB / Odds & \text{if } P < PB \end{cases}$

Difference

The difference is perhaps the most direct method of comparison between two proportions. It is easy to interpret and communicate. It gives the absolute impact of the treatment. However, there are subtle difficulties that can arise with its use.

One difficulty arises when the event of interest is rare. If a difference of 0.001 occurs when the baseline probability is 0.40, it would be dismissed as being trivial. That is, there is usually little interest in a treatment that only decreases the probability from 0.400 to 0.399. However, if the baseline probability of a disease is 0.002, a 0.001 decrease would represent a reduction of 50%. Thus, interpretation of the difference depends on the baseline probability of the event. As a rule of thumb, the difference is best suited for those cases in which.

Equivalence Test using a Difference

The following example might be instructive. Suppose 60% of patients respond to the current treatment method (PB = 0.60). If the response rate of the new treatment is no less than five percentage points different (d0 = 0.05) from the existing treatment, it will be considered to be equivalent. Substituting these figures into the statistical hypotheses gives

$$H0: d \geq 0.05 \text{ versus } H1: d < 0.05$$

where d = |P – PB|.

The resulting joint hypotheses are

$$H0: P \leq 0.55 \text{ versus H1: P > 0.55.}$$

and

$$H0: P \geq 0.65 \text{ versus H1: P < 0.65.}$$

In this example, when both null hypotheses are rejected, the concluded alternative is that the response rate is between 55% and 65%.

Ratio

The ratio r0 = PE / PB denotes the relative change in the probability of the response. Testing equivalence uses the hypotheses

$$H0: r \leq r0 \text{ versus } H1: r > r0$$

where r = P / PB if P > PB or r = PB / P if P < PB.

Equivalence Test using a Ratio

The following example might help to understand the concept of *equivalence* as defined by the ratio. Suppose that 60% of patients (PB = 0.60) respond to the current treatment method. If a new treatment changes the response rate by no more than 10% (r0 = 1.1), it will be considered to be equivalent to the standard treatment. Substituting these figures into the statistical hypotheses gives

$$H0: r \geq 1.1 \text{ versus } H1: r < 1.1$$

The relationship P0 = (r0)(PB) gives the two, one-sided, hypotheses

$$H0: P \leq 0.54 \text{ versus } H1: P > 0.54$$

$$H0: P \geq 0.66 \text{ versus } H1: P < 0.66$$

In this example, when the null hypothesis is rejected, the concluded alternative is that the response rate is between 54% and 66%.

Odds Ratio

The odds ratio, o0 = (PE / (1 – PE)) / (PB / (1 – PB)), gives the relative change in the odds of the response. Testing noninferiority and superiority use the same formulation, namely

$$H0: o \leq o0 \text{ versus } H1: o > o0$$

where o = Odds / OddsB if P > PB or o = OddsB / Odds if P < PB.

Power and Sample Size Calculation

Historically, power and sample size calculations for a one-sample proportion test have been based on normal approximations to the binomial. However, with the speed of modern computers, using the normal approximation is unnecessary, especially for small samples. Rather, the significance level and power can be computed using complete enumeration of all possible values of x, the number of successes in a sample of size n.

This is done as follows.

1. The critical value of the test is computed using standard techniques.
2. For each possible value of x, the value of the test statistic (z test, t test, or exact test) is computed along with its associated probability of occurrence.
3. The significance level and power are computed by summing the probabilities of occurrence for all values of the test statistic that are greater than (or less than) the critical value. Each probability of occurrence is calculated using P0 for the significance level and P1 for the power.

Other variables such as the sample size are then found using an efficient search algorithm. Although this method is not as elegant as a closed-form solution, it is completely accurate.

Examples of Power Calculation for the Exact Test

Suppose the baseline proportion, PB, is 0.50, the sample size is 10, and the target alpha level is 0.05. A typical value for the equivalence difference is 0.05. However, because the example is for a small sample size, the equivalence difference will be set to 0.4 (which is, of course, a very unrealistic figure) for illustrative purposes. Calculate the power of this design to detect equivalence if the actual difference between the proportions is 0.10.

The first step is to find the rejection region under the null hypothesis. In this example, the null hypothesis is $H0: P \le 0.1$ or $H0: P > 0.9$ and the alternative hypothesis is $H1: 0.1 < P < 0.9$ $H_1: 0.1 < P < 0.9$. This composite hypothesis breaks down into the following two, one-sided, simple hypotheses

1. $H0: P \le 0.1$ versus $H1: P > 0.1$
2. $H0: P \ge 0.9$ versus $H1: P < 0.9$

The rejection regions for the both tests are determined from the following table of cumulative binomial probabilities for $N = 10$. The first column of probabilities is for r greater than or equal to R while the second two columns of probabilities are for r less than or equal to R.

Table of Binomial Probabilities for N = 10 and P = 0.1, 0.9, and 0.6

R	Pr(r>=R |P=0.1)	Reject Test1	Pr(r<=R |P=0.9)	Reject Test2	Reject Both	Pr(r<=R|P=0.6)
0	1.0000	No	0.0000	Yes	No	0.0001
1	0.6513	No	0.0000	Yes	No	0.0017
2	0.2639	No	0.0000	Yes	No	0.0123
3	0.0702	No	0.0000	Yes	No	**0.0548**
4	**0.0128**	Yes	0.0001	Yes	Yes	0.1662
5	0.0016	Yes	0.0016	Yes	Yes	0.3669
6	0.0001	Yes	**0.0128**	Yes	Yes	**0.6177**
7	0.0000	Yes	0.0702	No	No	0.8327
8	0.0000	Yes	0.2639	No	No	0.9536
9	0.0000	Yes	0.6513	No	No	0.9940
10	0.0000	Yes	1.0000	No	No	1.0000

The second column gives the value of alpha for the first test ($H0: P \le 0.1$ versus $H1: P > 0.1$). The rejection region for this test is all values of R greater than or equal to 4. The fourth column gives the values of alpha for the second test. The rejection region for the second test is all values of R less than or equal to 6. The rejection region for both tests is those values of R values that result in rejection of both individual tests. These are the R values 4, 5, and 6. The power is computed using the final column of the table which gives cumulative binomial probabilities for $P = 0.5 + 0.1 = 0.6$. The power is probability for the cases 4, 5, and 6. It is calculated as 0.6177 – 0.0548 = 0.5629.

It is informative to consider what happens when the equivalence difference is reduced from 0.4 to 0.2. The following table gives the appropriate cumulative binomial probabilities for this case.

Table of Binomial Probabilities for N = 10 and P = 0.3, 0.7, and 0.6

R	Pr(r>=R |P=0.3)	Reject Test1	Pr(r<=R |P=0.7)	Reject Test2	Reject Both	Pr(r<=R|P=0.6)
0	1.0000	No	0.0000	Yes	No	0.0001
1	0.9718	No	0.0001	Yes	No	0.0017
2	0.8507	No	0.0016	Yes	No	0.0123
3	0.6172	No	0.0106	Yes	No	0.0548
4	0.3504	No	**0.0473**	Yes	No	0.1662
5	0.1503	No	0.1503	No	No	0.3669
6	**0.0473**	Yes	0.3504	No	No	0.6177
7	0.0106	Yes	0.6172	No	No	0.8327
8	0.0016	Yes	0.8507	No	No	0.9536
9	0.0001	Yes	0.9718	No	No	0.9940
10	0.0000	Yes	1.0000	No	No	1.0000

The second column gives the value of alpha for the first test. The rejection region for this test is all values of *R* greater than or equal to 6. The fourth column gives the values of alpha for the second test. The rejection region for the second test is all values of *R* less than or equal to 4. The rejection region for both tests together is empty! There is no *R* for which both tests will be rejected. Hence, the alpha level and the power will both be 0.0.

Examples of Power Calculation for the Z Test

The following example illustrates how to calculate the power of an approximate *z* test. There are several *z* tests to choose from. We will use the following test.

$$z = \frac{p - P0}{\sqrt{P0(1 - P0) / n}}$$

Calculating the rejection region for the *z* test is based on a table of normal probabilities. For the target alpha level of 0.05, the critical value is 1.6449. That is, the first hypothesis test that $H0: P \le 0.1$ versus $H1: P > 0.1$ is rejected if the resulting calculated *z* value is greater than 1.6449. Similarly, the second hypothesis test that $H0: P \ge 0.9$ versus $H1: P < 0.9$ is rejected when the calculated *z* value is less than -1.6449. The rejection regions for the both tests are shown in the following table of binomial probabilities for $N = 10$.

Table Showing Both One-Sided Z Tests for N = 10 and P = 0.1, 0.9, and 0.6

R	Z for P = 0.1	Reject Test1	Z for P = 0.9	Reject Test2	Reject Both	Pr(r<=R|P=0.6)
0	-1.0541	No	-9.4868	Yes	No	0.0001
1	0.0000	No	-8.4327	Yes	No	0.0017
2	**1.0541**	No	-7.3786	Yes	No	**0.0123**
3	2.1082	Yes	-6.3246	Yes	Yes	0.0548
4	3.1623	Yes	-5.2705	Yes	Yes	0.1662
5	4.2164	Yes	-4.2164	Yes	Yes	0.3669
6	5.2705	Yes	-3.1623	Yes	Yes	0.6177
7	6.3246	Yes	-2.1082	Yes	Yes	**0.8327**
8	7.3786	Yes	-1.0541	No	No	0.9536
9	8.4327	Yes	0.0000	No	No	0.9940
10	9.4868	Yes	1.0541	No	No	1.0000

Note that the null hypothesis is rejected for the equivalence test when *R* is 3, 4, 5, 6, and 7. The power is the probability of these values calculated using *P* = 0.60. It is calculated as 0.8327 – 0.0123 = 0.8204. Notice that this is much larger than 0.5629 which was the power for the exact test. The reason for this discrepancy is that the approximate test is actually testing at a larger alpha than the target of 0.05. The actual alpha is the maximum of the two individual alphas. From the first table, we can see that the actual alpha for the first test is Pr(r>=3|P=0.1) = 0.0702. Similarly, the actual alpha for the second test is Pr(r<=7|P=0.9) = 0.0702. Hence the alpha level is 0.0702. The actual alpha of the exact test was 0.0128.

Test Statistics

The test statistics used are given in the chapter entitled Inequality Tests for One Proportion. They will not be repeated here.

Procedure Options

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

Data Tab (Common Options)

The Data tab contains the parameters associated with this test such as the proportions, sample sizes, alpha, and beta. This chapter covers four procedures, each of which has different options. This section documents options that are common to all four procedures. Later, unique options for each procedure will be documented.

Solve For

Find (Solve For)

This option specifies the parameter to be solved for from the other parameters. The parameters that may be selected are *Alpha*, *Power and Beta*, and *n*. In most situations, you will select either *Power and Beta* or *n*.

Select *n* when you want to calculate the sample size needed to achieve a given power and alpha level.

Select *Power and Beta* when you want to calculate the power of an experiment.

Error Rates

Power or Beta

This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected.

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

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

Alpha (Significance Level)

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected.

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.

Note that because of the discrete nature of the binomial distribution, the alpha level rarely will be achieved exactly.

A single value may be entered here or a range of values such as *0.05 to 0.2 by 0.05* may be entered.

Sample Size

n (Sample Size)

Enter a value (or range of values) for the sample size *n*. This is the number of individuals sampled in the study. Values must be integers greater than one.

You may enter a range such as *10, 50, 100* or *10 to 100 by 10*.

Effect Size

PB (Baseline Proportion)

Enter a value (or range of values) for the baseline proportion. In an equivalence study, this is the response rate of the standard (existing) treatment. Note that this is not the value of P0. Instead, this value is used in the calculation of P0.

Proportions must be 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 0.1*.

Test

Test Type

Specify the type of test that will be used in searching and reporting. Note that *C.C.* is an abbreviation for *Continuity Correction*. This refers to the adding or subtracting of $1/(2n)$ to (or from) the numerator of the z-value to bring the normal approximation closer to the binomial distribution.

In most situations, you would select the ‘Exact Test’ option. The other options are provided for comparative purposes.

Data Tab (Proportions)

This section documents options that are used when the parameterization is given directly in terms of the proportions PL, PU, PB, and P1.

Effect Size

P1 (Actual Proportion)

This is the value of the proportion, P1, at which the power is calculated. The power calculations assume that this is the actual value of the proportion. For noninferiority tests, this value is often set equal to PB.

Proportions must be 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 0.1*.

Upper and Lower Equivalence Proportions

These options set the smallest and largest values which are still to be considered trivially different from PB. Note that the lower proportion must be less than PB, and the upper proportion must be greater than PB. Since these values are proportions, they must be positive values less than one. They cannot be equal to PB.

Data Tab (Differences)

This section documents options that are used when the parameterization is in terms of the difference of two proportions.

Effect Size

d0 (Equivalence Difference)

This option sets the smallest value which is still trivially different from PB by setting the magnitude of the difference between P0 and PB. For example, if PB (baseline proportion) is 0.50, you might consider differences of 0.01, 0.02, or 0.04 to be small enough so that the fact that P0 is different from 0.50 can be overlooked. However, you might decide that if the difference is 0.05 or more, the treatment is not equivalent. Thus, this value would be set to 0.05.

Since this value is an absolute difference between two proportions, it must be between 0 and 1.

d1 (Actual Difference)

This option specifies the value of P1 (the actual proportion) by specifying the difference between the two proportions, P1 and PB. This difference is used with PB to calculate the value of P1 using the formula: P1 = PB + difference. For equivalence tests, this value is often set equal to zero.

Differences must be between -1 and 1.

You may enter a range of values such as *.03 .05 .10* or *.01 to .05 by .01*.

Data Tab (Ratios)

This section documents options that are used when the parameterization is in terms of the ratio of two proportions.

Effect Size

r0 (Equivalence Ratio)

This option sets the value which is still trivially different from PB by setting the ratio between P0 and PB. For P0 example, if PB (baseline proportion) is 0.50, you might consider ratios of 0.99, 0.98, or even 0.96 to be small enough so that the fact that P0 is less than PB can be overlooked (the difference is trivial). However, you might decide that if the ratio is 0.95 or less, the treatment is not equivalent. Thus, this value would be set to 0.95.

Since this value is a ratio between two proportions, it must be positive. Since it is a margin, it cannot be one. Also, it cannot be so large that the calculated value of P0 is greater than one.

r1 (Actual Ratio)

This option specifies the value of P1 (the actual proportion) by specifying the ratio between the two proportions, P1 and PB. This ratio is used with PB to calculate the value of P1 using the formula: P1 = (Ratio)(PB). For equivalence tests, this value is often set equal to one.

Ratios must be greater than zero. Note that the ratios must be small enough so that P1 is less than one.

You may enter a range of values such as *.5 .6 .7 .8* or *1.25 to 2.0 by .25*.

Data Tab (Odds Ratios)

This section documents options that are used when the parameterization is in terms of the odds ratios.

Effect Size

o0 (Equivalence Odds Ratio)

This option sets the value which is still trivially different from PB by setting the odds ratio of P0 and PB. For example, if PB (baseline proportion) is 0.50, you might consider odds ratios of 0.99, 0.98, or even 0.96 to be small enough so that the fact that P0 is less than PB can be overlooked (the difference is trivial). However, you might decide that if the odds ratio is 0.80 or less, the treatment is inferior. Thus, this value would be set to 0.80.

Since this value is a ratio between two odds, it must be positive. Because it is a margin, it cannot be one.

o1 (Actual Odds Ratio)

This option specifies the value of P1 (the actual proportion) by specifying the odds ratio between the two proportions, P1 and PB. This ratio is used with PB to calculate the value of P1. For noninferiority tests, this value is often set equal to one.

Odds ratios must be greater than zero. You may enter a range of values such as *.5 .6 .7 .8* or *1.25 to 2.0 by .25*.

Iterations Tab

The Iterations tab allows for specification of the maximum number of iterations to be used in searches.

Maximum Iterations

Maximum Iterations Before Search Termination

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 not reported.

Example 1 – Finding the Power

Suppose 50% of patients with a certain type of cancer survive two years using the current treatment. The current treatment is expensive and has several severe side effects. A new treatment has fewer side effects and is less expensive. An equivalence trial is to be conducted to show that the two-year survival rate of the new treatment is the same as the current treatment. After serious consideration, the margin of equivalence is set at 5%. What power will be achieved by sample sizes of 50, 100, 200, 300, 500, or 800 and a significance level of 0.05? For comparative purposes, also calculate the power for margin of equivalence of 10%. Assume that the true survival rate of the new treatment is the same as that of the current (baseline) treatment.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Equivalence Tests for One Proportion [Differences]** procedure window by expanding **Proportions**, then **One Proportion**, then clicking on **Equivalence**, and then clicking on **Equivalence Tests for One Proportion [Differences]**. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power	*Ignored since this is the Find setting*
Alpha	**0.05**
n (Sample Size)	**50 100 200 300 500 800**
d0 (Equivalence Difference)	**0.05 0.10**
d1 (Actual Difference)	**0**
PB (Baseline Proportion)	**0.50**
Test Type	**Exact Test**
Reports Tab	
Show Numeric Reports	**Checked**
Show Comparative Reports	**Not checked**
Show Definitions	**Checked**

Show Plots .....**Checked**
Show Comparative Plots.....**Not checked**

Annotated Output

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

Numeric Results

Numeric Results when H0: Non-Equivalence versus H1: Equivalence
Test Statistic: Exact Test

Power	N	Equiv. Diff. (d0)	Lower Equiv. Prop. (P0L)	Upper Equiv. Prop. (P0U)	Actual Diff. (d1)	Baseline Prop. (PB)	Target Alpha	Actual Alpha	Beta	Reject H0 if R1<=R<=R2 (R1|R2)
0.0000	50	0.0500	0.4500	0.5500	0.0000	0.5000	0.0500	0.0000	1.0000	29|21
0.0000	100	0.0500	0.4500	0.5500	0.0000	0.5000	0.0500	0.0000	1.0000	54|46
0.0000	200	0.0500	0.4500	0.5500	0.0000	0.5000	0.0500	0.0000	1.0000	103|97
0.0460	300	0.0500	0.4500	0.5500	0.0000	0.5000	0.0500	0.0465	0.9540	150|150
0.4390	500	0.0500	0.4500	0.5500	0.0000	0.5000	0.0500	0.0484	0.5610	244|256
0.7567	800	0.0500	0.4500	0.5500	0.0000	0.5000	0.0500	0.0476	0.2433	384|416
0.0000	50	0.1000	0.4000	0.6000	0.0000	0.5000	0.0500	0.0000	1.0000	27|23
0.2356	100	0.1000	0.4000	0.6000	0.0000	0.5000	0.0500	0.0423	0.7644	49|51
0.7708	200	0.1000	0.4000	0.6000	0.0000	0.5000	0.0500	0.0492	0.2292	92|108
0.9267	300	0.1000	0.4000	0.6000	0.0000	0.5000	0.0500	0.0443	0.0733	135|165
0.9952	500	0.1000	0.4000	0.6000	0.0000	0.5000	0.0500	0.0461	0.0048	219|281
0.9999	800	0.1000	0.4000	0.6000	0.0000	0.5000	0.0500	0.0453	0.0001	344|456

Report Definitions
Power is the probability of concluding equivalence when the proportions are equivalent.
N is the size of the sample drawn from the population.
The equivalence difference is the maximum value of the difference that is still considered unimportant.
The actual difference is the value of the difference under the alternative hypothesis.
PB is the baseline or standard value of the proportion. This is the value under the current treatment.
P0L and P0U are the limits between which an equivalent proportion must fall.
d0 is the smallest absolute difference that is still considered equivalent.
d1 is the value of the difference under the alternative hypothesis.
Alpha is the probability of concluding equivalence when the proportions are non-equivalent.
Beta is the probability concluding non-equivalence when the proportions are equivalent.

Summary Statements
A sample size of 50 achieves 0% power to detect a difference (P0-PB) of 0.0500 using a two-sided binomial test. The target significance level is 0.0500. The actual significance level achieved by this test is 0.0000. These results assume a baseline proportion (PB) of 0.5000 and that the actual difference (P1-PB) is 0.0000.

This report shows the values of each of the parameters, one scenario per row. Because of the discrete nature of the binomial distribution, the target alpha is usually different than the actual alpha. Hence, the actual alpha is also shown.

Power

Power is the probability of concluding equivalence when the treatment is indeed equivalent.

N

This is the sample size.

Equivalence Difference (or Proportion, Ratio, or Odds Ratio)

The equivalence difference is the maximum difference from the baseline proportion, PB, that is still considered as unimportant or trivial. This value is used to calculate P0.

Equivalence Upper and Lower Proportions

If the true proportion is between these two limits, the treatment is considered to be equivalent to the baseline proportion. These are the bounds of equivalence.

Actual Difference (or Proportion, Ratio, or Odds Ratio)

The actual difference is the difference between the actual proportion, P1, and the baseline proportion, PB.

Baseline Proportion

The baseline proportion, PB, is the response rate that is achieved by the current (standard) treatment.

Target Alpha

This is the target (set in the design) value of the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected. That is, this is the probability of concluding equivalence when in fact the new treatment is not equivalent. Because of the discreteness of the binomial distribution from which this value is calculated, the target value is seldom achieved.

Actual Alpha

This is the actual value of alpha (see Target Alpha) that is achieved by the design. Note that low values of alpha reduce the power.

Beta

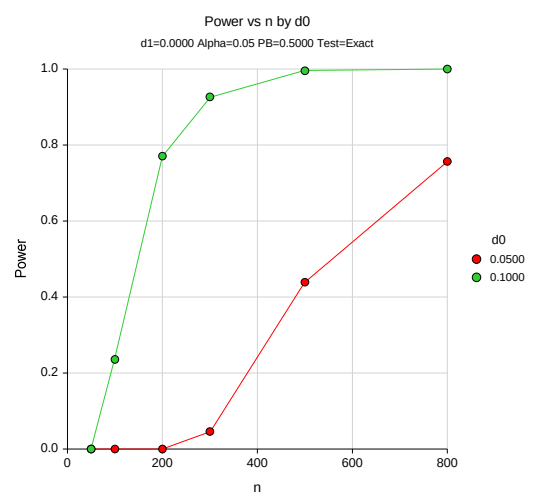
Beta is the probability of accepting a false null hypothesis. It is the opposite of power.

Reject H0 if R1<=R<=R2

This value provides the bounds between which equivalence is concluded. For example, if *n* is 50, then a value here of 29|31 means that the null hypothesis of non-equivalence is rejected when the number of items with the characteristic of interest is 29, 30, or 31.

When the second number is less than the first as it is in the first line (29|21), the design can never reject the null hypothesis. These designs should never be used.

Plots Section



This plot shows the relationship between power, sample size, and the trivial difference. Note that 80% power is achieved with a sample size of about 210 when the trivial difference is 0.10 and over 800 when the trivial difference is 0.05.

Example 2 – Finding the Sample Size

Continuing from Example1, suppose you want to find the exact sample size necessary to achieve 90% power when the trivial difference is 0.05. Assume that an exact binomial test will be used.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Equivalence Tests for One Proportion [Differences]** procedure window by expanding **Proportions**, then **One Proportion**, then clicking on **Equivalence**, and then clicking on **Equivalence Tests for One Proportion [Differences]**. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**n**
Power	**0.90**
Alpha	**0.05**
n (Sample Size)	*Ignored since this is the Find setting*
d0 (Equivalence Difference)	**0.05**
d1 (Actual Difference)	**0**
PB (Baseline Proportion)	**0.50**
Test Type	**Exact Test**
Reports Tab	
Show Numeric Reports	**Checked**
Show Comparative Reports	**Not checked**
Show Definitions	**Not checked**
Show Plots	**Not checked**
Show Comparative Plots	**Not checked**

Output

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

Numeric Results

Numeric Results when H0: Non-Equivalence versus H1: Equivalence
Test Statistic: Exact Test

Power	N	Equiv. Diff. (d0)	Lower Equiv. Prop. (P0L)	Upper Equiv. Prop. (P0U)	Actual Diff. (d1)	Baseline Prop. (PB)	Target Alpha	Actual Alpha	Beta	Reject H0 if R1<=R<=R2 (R1|R2)
0.9040	1092	0.0500	0.4500	0.5500	0.0000	0.5000	0.0500	0.0498	0.0960	519|573

This report shows that a sample size of 1092 will be necessary to achieve the design requirements.

Example 3 – Comparing Test Statistics

Continuing Example1, suppose the researchers want to investigate which of the five test statistics to use. This is an important question since choosing the wrong test statistic can increase sample size and reduce power. The differences in the characteristics of test statistics are most noticeable in small samples. Hence, the investigation done here is for sample sizes of 20 to 200 in steps of 20. The trivial difference will be set to 0.10. All other settings are as given in Example 1.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Equivalence Tests for One Proportion [Differences]** procedure window by expanding **Proportions**, then **One Proportion**, then clicking on **Equivalence**, and then clicking on **Equivalence Tests for One Proportion [Differences]**. You may then make the appropriate entries as listed below, or open **Example 3** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power	*Ignored since this is the Find setting*
Alpha	**0.05**
n (Sample Size)	**20 to 200 by 20**
d0 (Equivalence Difference)	**0.10**
d1 (Actual Difference)	**0**
PB (Baseline Proportion)	**0.50**
Test Type	**Exact Test**
Reports Tab	
Show Numeric Reports	**Not checked**
Show Comparative Reports	**Checked**
Show Definitions	**Not checked**
Show Plots	**Not checked**
Show Comparative Plots	**Checked**

Output

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

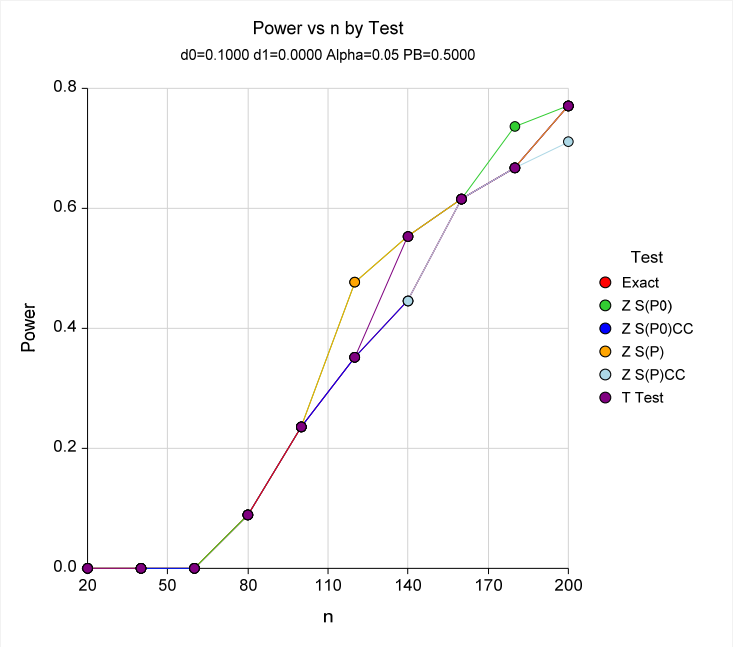
Numeric Results and Plots

Power Comparison for Methods of Testing H0: Non-Equivalence versus H1: Equivalence

N	Equiv. Diff. (d0)	Actual Diff. (d1)	Baseline Prop. (PB)	Target Alpha	Exact Test Power	Z-Test S(P0) Power	Z-Test S(P0)C Power	Z-Test S(P) Power	Z-Test S(P)C Power	T-Test Power
20	0.1000	0.0000	0.5000	0.0500	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
40	0.1000	0.0000	0.5000	0.0500	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
60	0.1000	0.0000	0.5000	0.0500	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
80	0.1000	0.0000	0.5000	0.0500	0.0889	0.0889	0.0889	0.0889	0.0889	0.0889
100	0.1000	0.0000	0.5000	0.0500	0.2356	0.2356	0.2356	0.2356	0.2356	0.2356
120	0.1000	0.0000	0.5000	0.0500	0.3517	0.4770	0.3517	0.4770	0.3517	0.3517
140	0.1000	0.0000	0.5000	0.0500	0.4457	0.5530	0.4457	0.5530	0.4457	0.5530
160	0.1000	0.0000	0.5000	0.0500	0.6154	0.6154	0.6154	0.6154	0.6154	0.6154
180	0.1000	0.0000	0.5000	0.0500	0.6674	0.7365	0.6674	0.6674	0.6674	0.6674
200	0.1000	0.0000	0.5000	0.0500	0.7708	0.7708	0.7708	0.7708	0.7112	0.7708

Actual Alpha Comparison for Methods of Testing H0: Non-Equivalence versus H1: Equivalence

N	Equiv. Diff. (d0)	Actual Diff. (d1)	Baseline Prop. (PB)	Target Alpha	Exact Test Alpha	Z-Test S(P0) Alpha	Z-Test S(P0)C Alpha	Z-Test S(P) Alpha	Z-Test S(P)C Alpha	T-Test Alpha
20	0.1000	0.0000	0.5000	0.0500	0.0000	0.0565	0.0210	0.0565	0.0210	0.0565
40	0.1000	0.0000	0.5000	0.0500	0.0000	0.0392	0.0392	0.0392	0.0392	0.0392
60	0.1000	0.0000	0.5000	0.0500	0.0000	0.0445	0.0445	0.0445	0.0445	0.0445
80	0.1000	0.0000	0.5000	0.0500	0.0445	0.0445	0.0445	0.0445	0.0445	0.0445
100	0.1000	0.0000	0.5000	0.0500	0.0423	0.0423	0.0423	0.0423	0.0423	0.0423
120	0.1000	0.0000	0.5000	0.0500	0.0392	0.0575	0.0392	0.0575	0.0392	0.0392
140	0.1000	0.0000	0.5000	0.0500	0.0358	0.0514	0.0358	0.0514	0.0358	0.0514
160	0.1000	0.0000	0.5000	0.0500	0.0459	0.0459	0.0459	0.0459	0.0459	0.0459
180	0.1000	0.0000	0.5000	0.0500	0.0408	0.0558	0.0408	0.0408	0.0408	0.0408
200	0.1000	0.0000	0.5000	0.0500	0.0492	0.0492	0.0492	0.0492	0.0363	0.0492



The first report shows the power for each test statistic. The second report shows the actual alpha achieved by the design.

An examination of the first report shows that once non-zero powers are obtained, they are often different for at least one of the tests. Also notice that the exact test always has the minimum power in each row. This would lead us discard this test statistic. However, consider the second report which shows the actual alpha level (the target was 0.05) for each test. By inspecting corresponding entries in both tables, we see that whenever a test statistic achieves a better power than the exact test, it also yields an actual alpha level larger than the target alpha.

For example, look at the powers for n = 120. The z test using s(P0) has an unusually large power = 0.4770. This is a much larger power than the exact test's value of 0.3517. However, note that the actual alpha for this test is 0.0575 which is larger than the target alpha of 0.05 and the exact test's alpha of 0.0392.

We conclude that indeed, the exact test is consistently the best test since it always achieves a significance level that is less than the target value.

Example 4 – Validation

We could not find a worked example for this situation in the literature. Therefore, we will use the example that was worked 'by hand' earlier in this chapter to validate the program. In that example, the baseline proportion was 0.50, alpha was 0.05, n was 10, the actual difference was 0.10, and the trivial difference was 0.40. The power was calculated to be 0.5629.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Equivalence Tests for One Proportion [Differences]** procedure window by expanding **Proportions**, then **One Proportion**, then clicking on **Equivalence**, and then clicking on **Equivalence Tests for One Proportion [Differences]**. You may then make the appropriate entries as listed below, or open **Example 4** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power	*Ignored since this is the Find setting*
Alpha	**0.05**
n (Sample Size)	**10**
d0 (Equivalence Difference)	**0.40**
d1 (Actual Difference)	**0.10**
PB (Baseline Proportion)	**0.50**
Test Type	**Exact Test**
Reports Tab	
Show Numeric Reports	**Checked**

Output

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

Numeric Results

Numeric Results when H0: Non-Equivalence versus H1: Equivalence
Test Statistic: Exact Test

Power	N	Equiv. Diff. (d0)	Lower Equiv. Prop. (P0L)	Upper Equiv. Prop. (P0U)	Actual Diff. (d1)	Baseline Prop. (PB)	Target Alpha	Actual Alpha	Beta	Reject H0 if R1<=R<=R2 (R1|R2)
0.5630	10	0.4000	0.1000	0.9000	0.1000	0.5000	0.0500	0.0128	0.4370	4|6

PASS has obtained the same answer within rounding error.

Example 5 – Computing the Power after Completing the Experiment

Researchers are testing a generic drug to determine if it is equivalent to the name-brand alternative. Equivalence is declared if the success rate of the generic brand is no more than 10% from that of the name-brand drug. Suppose that the name-brand drug is known to have a success rate of 60%. In a study of 500 individuals, they find that 265, or 53%, are successfully treated using the generic brand. An equivalence test (exact test) with alpha = 0.05 failed to declare that the two drugs are equivalent. The researchers would now like to compute the power for actual differences ranging from 0 to 9%.

Note that the power is not calculated solely at the difference observed in the study, 7%. It is more informative to study a range of values with practical significance.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Equivalence Tests for One Proportion [Differences]** procedure window by expanding **Proportions**, then **One Proportion**, then clicking on **Equivalence**, and then clicking on **Equivalence Tests for One Proportion [Differences]**. You may then make the appropriate entries as listed below, or open **Example 5** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power	*Ignored since this is the Find setting*
Alpha	**0.05**
n (Sample Size)	**500**
d0 (Equivalence Difference)	**0.10**
d1 (Actual Difference)	**0.0 to 0.09 by 0.01**
PB (Baseline Proportion)	**0.60**
Test Type	**Exact Test**

Reports Tab

Show Numeric Reports ..... **Checked**
Show Comparative Reports ..... **Not checked**
Show Definitions ..... **Not checked**
Show Plots ..... **Checked**
Show Comparative Plots..... **Not checked**

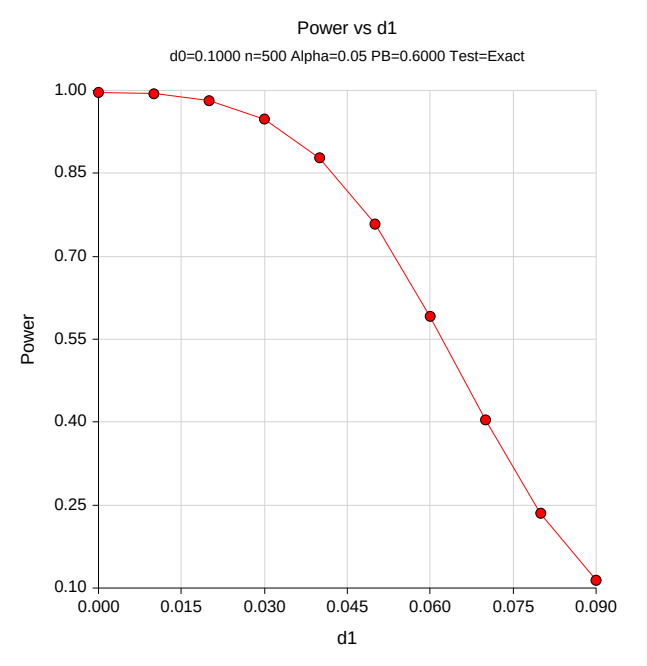
Output

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

Numeric Results and Plots

Numeric Results when H0: Non-Equivalence versus H1: Equivalence
Test Statistic: Exact Test

Power	N	Equiv. Diff. (d0)	Lower Equiv. Prop. (P0L)	Upper Equiv. Prop. (P0U)	Actual Diff. (d1)	Baseline Prop. (PB)	Target Alpha	Actual Alpha	Beta	Reject H0 if R1<=R<=R2 (R1|R2)
0.9965	500	0.1000	0.5000	0.7000	0.0000	0.6000	0.0500	0.0489	0.0035	269|332
0.9940	500	0.1000	0.5000	0.7000	0.0100	0.6000	0.0500	0.0489	0.0060	269|332
0.9815	500	0.1000	0.5000	0.7000	0.0200	0.6000	0.0500	0.0489	0.0185	269|332
0.9482	500	0.1000	0.5000	0.7000	0.0300	0.6000	0.0500	0.0489	0.0518	269|332
0.8783	500	0.1000	0.5000	0.7000	0.0400	0.6000	0.0500	0.0489	0.1217	269|332
0.7583	500	0.1000	0.5000	0.7000	0.0500	0.6000	0.0500	0.0489	0.2417	269|332
0.5914	500	0.1000	0.5000	0.7000	0.0600	0.6000	0.0500	0.0489	0.4086	269|332
0.4041	500	0.1000	0.5000	0.7000	0.0700	0.6000	0.0500	0.0489	0.5959	269|332
0.2352	500	0.1000	0.5000	0.7000	0.0800	0.6000	0.0500	0.0489	0.7648	269|332
0.1139	500	0.1000	0.5000	0.7000	0.0900	0.6000	0.0500	0.0489	0.8861	269|332



The range in power is quite large. The power is relatively high and constant if the true difference is less than or equal to 4%, but it decreases rapidly as the differences increase from there.

Example 6 – Finding the Sample Size using Ratios

Researchers would like to compare a new treatment to an existing standard treatment. The new treatment will be deemed equivalent to the standard treatment if the response rate is changed by no more than 20%, hence, r = 1.20. It is known that 60% of patients respond to the standard treatment. If the researchers use the exact test and a significance level of 0.05, how large of a sample must they take to achieve 90% power if the actual ratio is 1.0?

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Equivalence Tests for One Proportion [Ratios]** procedure window by expanding **Proportions**, then **One Proportion**, then clicking on **Equivalence**, and then clicking on **Equivalence Tests for One Proportion [Ratios]**. You may then make the appropriate entries as listed below, or open **Example 6** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**n**
Power	**0.90**
Alpha	**0.05**
n (Sample Size)	*Ignored since this is the Find setting*
r0 (Equivalence Ratio)	**1.2**
r1 (Actual Ratio)	**1.0**
PB (Baseline Proportion)	**0.60**
Test Type	**Exact Test**
Reports Tab	
Show Numeric Reports	**Checked**
Show Comparative Reports	**Not checked**
Show Definitions	**Not checked**
Show Plots	**Not checked**
Show Comparative Plots	**Not checked**

Output

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

Numeric Results

Numeric Results when H0: Non-Equivalence versus H1: Equivalence
Test Statistic: Exact Test

Power	N	Equiv. Ratio (r0)	Lower Equiv. Prop. (P0L)	Upper Equiv. Prop. (P0U)	Actual Ratio (r1)	Baseline Prop. (PB)	Target Alpha	Actual Alpha	Beta	Reject H0 if R1<=R<=R2 (R1|R2)
0.9014	228	1.2000	0.5000	0.7200	1.0000	0.6000	0.0500	0.0488	0.0986	127|152

They must sample 228 individuals to achieve just over 90% power for an actual ratio of 1.0 and equivalence ratio of 1.20.

Chapter 115

Confidence Intervals for One Proportion

Introduction

This routine calculates the sample size necessary to achieve a specified interval width or distance from the sample proportion to the confidence limit at a stated confidence level for a confidence interval for one proportion.

Caution: This procedure assumes that the proportion of the future sample will be the same as the proportion that is specified. If the sample proportion is different from the one specified when running this procedure, the interval width may be narrower or wider than specified.

Technical Details

Many methods have been devised for computing confidence intervals for a single proportion. Five of these methods are available in this procedure. The five confidence interval methods are

1. Exact (Clopper-Pearson)
2. Score (Wilson)
3. Score with continuity correction
4. Simple Asymptotic
5. Simple Asymptotic with continuity correction

For a comparison of methods, see Newcombe (1998a).

Confidence Interval Formulas

For each of the following methods, let p be the population proportion, and let r represent the number of successes from a sample of size N. Let $\hat{p} = r / N$.

Exact (Clopper-Pearson)

Using a mathematical relationship (see Fleiss et al (2003), p. 25) between the F distribution and the cumulative binomial distribution, the lower and upper confidence limits of a 100(1-α)% exact confidence interval for the true proportion p are given by

$$\left[\frac{r}{r+(n-r+1)F_{1-\alpha/2;2(n-r+1),2r}}, \frac{(r+1)F_{1-\alpha/2;2(r+1),2(n-r)}}{(n-r)+(r+1)F_{1-\alpha/2;2(r+1),2(n-r)}}\right]$$

One-sided limits may be obtained by replacing $\alpha/2$ by α.

Score (Wilson)

The Wilson Score confidence interval, which is based on inverting the z-test for a single proportion, is calculated using

$$\frac{\left(2n\hat{p}+z^2_{1-\alpha/2}\right) \pm z_{1-\alpha/2}\sqrt{z^2_{1-\alpha/2}+4n\hat{p}(1-\hat{p})}}{2\left(n+z^2_{1-\alpha/2}\right)}$$

One-sided limits may be obtained by replacing $\alpha/2$ by α.

Score with Continuity Correction

The Score confidence interval with continuity correction is based on inverting the z-test for a single proportion with continuity correction. The 100(1-α)% limits are calculated by

$$\text{Lower Limit} = \frac{\left(2n\hat{p}+z^2_{1-\alpha/2}-1\right) - z_{1-\alpha/2}\sqrt{z^2_{1-\alpha/2}-\{2+(1/n)\}+4\hat{p}\{n(1-\hat{p})+1\}}}{2\left(n+z^2_{1-\alpha/2}\right)}$$

$$\text{Upper Limit} = \frac{\left(2n\hat{p}+z^2_{1-\alpha/2}+1\right) + z_{1-\alpha/2}\sqrt{z^2_{1-\alpha/2}+\{2-(1/n)\}+4\hat{p}\{n(1-\hat{p})-1\}}}{2\left(n+z^2_{1-\alpha/2}\right)}$$

One-sided limits may be obtained by replacing $\alpha/2$ by α.

Simple Asymptotic

The simple asymptotic formula is based on the normal approximation to the binomial distribution. The approximation is close only for very large sample sizes. The 100(1-α)% confidence limits are given by

$$\hat{p} \pm z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

One-sided limits may be obtained by replacing $\alpha/2$ by α.

Simple Asymptotic with Continuity Correction

This formula is identical to the previous one, but with continuity correction. The 100(1-α)% confidence limits are

$$\left(\hat{p} - z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} - \frac{1}{2n}, \hat{p} + z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} + \frac{1}{2n} \right)$$

One-sided limits may be obtained by replacing $\alpha/2$ by α.

For two-sided intervals, the distance from the sample proportion to each of the limits may be different. Thus, instead of specifying the distance to the limits we specify the width of the interval, *W*.

The basic equation for determining sample size for a two-sided interval when *W* has been specified is

$$W = U - L$$

For one-sided intervals, the distance from the sample proportion to limit, *D*, is specified.

The basic equation for determining sample size for a one-sided upper limit when *D* has been specified is

$$D = U - \hat{p}$$

The basic equation for determining sample size for a one-sided lower limit when *D* has been specified is

$$D = \hat{p} - L$$

Each of these equations can be solved for any of the unknown quantities in terms of the others.

Confidence Level

The confidence level, 1 – α, has the following interpretation. If thousands of samples of *n* items are drawn from a population using simple random sampling and a confidence interval is calculated for each sample, the proportion of those intervals that will include the true population proportion is 1 – α.

Procedure Options

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

Data Tab

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

Solve For

Find (Solve For)

This option specifies the parameter to be solved for from the other parameters.

Confidence

Confidence Level

The confidence level, 1 - α, has the following interpretation. If thousands of samples of n items are drawn from a population using simple random sampling and a confidence interval is calculated for each sample, the proportion of those intervals that will include the true population proportion is 1 – α.

Often, the values 0.95 or 0.99 are used. You can enter single values or a range of values such as *0.90, 0.95 or 0.90 to 0.99 by 0.01*.

Sample Size

N (Sample Size)

Enter one or more values for the sample size. This is the number of individuals selected at random from the population to be in the study.

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

Precision

Confidence Interval Width (Two-Sided)

This is the distance from the lower confidence limit to the upper confidence limit.

You can enter a single value or a list of values. The value(s) must be between 0 and 1.

Distance from P to Limit (One-Sided)

This is the distance from the sample proportion to the lower or upper limit of the confidence interval, depending on whether the Interval Type is set to Lower Limit or Upper Limit.

You can enter a single value or a list of values. The value(s) must be between 0 and 1.

Proportion

P (Proportion)

Enter an estimate of the proportion. The sample size and width calculations assume that the value entered here is the proportion estimate that is obtained from the sample. If the sample proportion is different from the one specified here, the width may be narrower or wider than specified.

You can enter a range of values such as *.1 .2 .3* or *.1 to .5 by .1*.

Confidence Interval Method

Confidence Interval Formula

Specify the formula to be in used in calculation of confidence intervals.

- **Exact (Clopper-Pearson)**
  The exact formula uses the binomial probabilities directly.
- **Score (Wilson)**
  This formula is based on inverting a critical ratio test.
- **Score (Continuity Correction)**
  This formula is based on inverting a critical ratio test with continuity correction.
- **Simple Asymptotic**
  The simple asymptotic formula is based on the normal approximation to the binomial.
- **Simple Asymptotic (Continuity Correction)**
  This formula is based on the normal approximation to the binomial with continuity correction.

One-Sided or Two-Sided Interval

Interval Type

Specify whether the interval to be used will be a two-sided confidence interval, an interval that has only an upper limit, or an interval that has only a lower limit.

Iterations Tab

This tab sets an option used in the iterative procedures.

Maximum Iterations

Maximum Iterations Before Search Termination

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.

Example 1 – Calculating Sample Size

Suppose a study is planned in which the researcher wishes to construct a two-sided 95% exact (Clopper-Pearson) confidence interval for the population proportion such that the width of the interval is no wider than 0.06. The anticipated proportion estimate is 0.3, but a range of values from 0.1 to 0.5 will be included to determine the effect of the proportion estimate on necessary sample size. Instead of examining only the interval width of 0.06, widths of 0.04 and 0.10 will also be considered.

The goal is to determine the necessary sample size.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Confidence Intervals for One Proportion** procedure window by expanding **Proportions**, then **One Proportion**, then clicking on **Confidence Interval**, and then clicking on **Confidence Intervals for One Proportion**. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**N (Sample Size)**
Confidence Level	**0.95**
N (Sample Size)	*Ignored since this is the Find setting*
Confidence Interval Width (Two-Sided)	**0.04 0.06 0.10**
P (Proportion)	**0.1 to 0.5 by 0.1**
Confidence Interval Formula	**Exact (Clopper-Pearson)**
Interval Type	**Two-Sided**

Annotated Output

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

Numeric Results

Numeric Results for Two-Sided Confidence Intervals for One Proportion
Confidence Interval Formula: Exact (Clopper-Pearson)

Confidence Level	Sample Size (N)	Target Width	Actual Width	Proportion (P)	Lower Limit	Upper Limit	Width if P = 0.5
0.950	914	0.040	0.040	0.100	0.081	0.121	0.066
0.950	1585	0.040	0.040	0.200	0.181	0.221	0.050
0.950	2065	0.040	0.040	0.300	0.280	0.320	0.044
0.950	2353	0.040	0.040	0.400	0.380	0.420	0.041
0.950	2449	0.040	0.040	0.500	0.480	0.520	0.040
0.950	417	0.060	0.060	0.100	0.073	0.133	0.098
0.950	715	0.060	0.060	0.200	0.171	0.231	0.075
0.950	928	0.060	0.060	0.300	0.271	0.331	0.065
0.950	1056	0.060	0.060	0.400	0.370	0.430	0.061
0.950	1098	0.060	0.060	0.500	0.470	0.530	0.060
0.950	158	0.100	0.100	0.100	0.058	0.158	0.161
0.950	264	0.100	0.100	0.200	0.153	0.253	0.124
0.950	341	0.100	0.100	0.300	0.252	0.352	0.109
0.950	387	0.100	0.100	0.400	0.351	0.451	0.102
0.950	402	0.100	0.100	0.500	0.450	0.550	0.100

References

Fleiss, J. L., Levin, B., Paik, M.C. 2003. Statistical Methods for Rates and Proportions. Third Edition. John Wiley & Sons. New York.
Newcombe, R. G. 1998. 'Two-Sided Confidence Intervals for the Single Proportion: Comparison of Seven Methods.' Statistics in Medicine, 17, pp. 857-872.

Report Definitions

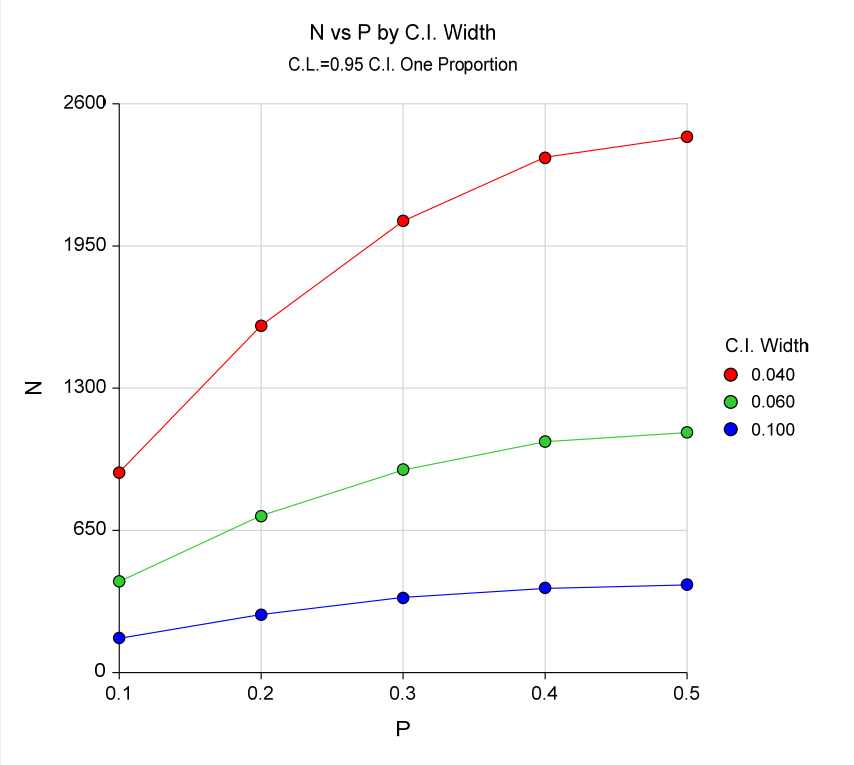
Confidence level is the proportion of confidence intervals (constructed with this same confidence level, sample size, etc.) that would contain the population proportion.
N is the size of the sample drawn from the population.
Width is the distance from the lower limit to the upper limit.
Target Width is the value of the width that is entered into the procedure.
Actual Width is the value of the width that is obtained from the procedure.
Proportion (P) is the assumed sample proportion.
Lower Limit is the lower limit of the confidence interval.
Upper Limit is the upper limit of the confidence interval.
Width if P = 0.5 is the maximum width for a confidence interval with sample size N.

Summary Statements

A sample size of 914 produces a two-sided 95% confidence interval with a width equal to 0.040 when the sample proportion is 0.100.

This report shows the calculated sample size for each of the scenarios.

Plots Section



This plot shows the sample size versus the sample proportion for the three confidence interval widths.

Example 2 – Validation using Fleiss, Levin, and Paik

Fleiss, Levin, and Paik (2003), pages 22-23, give an example of a calculation for an exact (Clopper-Pearson) one-sided lower limit confidence interval for a single proportion when the confidence level is 95%, the sample proportion is 0.92, and the distance from the lower limit to the sample proportion is 0.15104. The necessary sample size is 25.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Confidence Intervals for One Proportion** procedure window by expanding **Proportions**, then **One Proportion**, then clicking on **Confidence Interval**, and then clicking on **Confidence Intervals for One Proportion**. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**N (Sample Size)**
Confidence Level	**0.95**
N (Sample Size)	*Ignored since this is the Find setting*
Distance from P to Limit (One-Sided)	**0.15104**
P (Proportion)	**0.92**
Confidence Interval Formula	**Exact (Clopper-Pearson)**
Interval Type	**Lower Limit**

Output

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

Numeric Results

Confidence Level	Sample Size (N)	Target Distance from P to Lower Limit	Actual Distance from P to Lower Limit	Proportion (P)	Lower Limit	Upper Limit	Distance from P to Limit if P = 0.5
0.950	25	0.151	0.151	0.920	0.769	1	0.177

PASS also calculated the necessary sample size to be 25.

Example 3 – Validation using Newcombe

Newcombe (1998a), pages 860-861, gives an example of a calculation for a two-sided confidence interval for a single proportion for each of the methods when the confidence level is 95%. Here we validate the score method with continuity correction. The sample proportion is 0.034483, and the interval width is 0.1945. The necessary sample size is 29.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Confidence Intervals for One Proportion** procedure window by expanding **Proportions**, then **One Proportion**, then clicking on **Confidence Interval**, and then clicking on **Confidence Intervals for One Proportion**. You may then make the appropriate entries as listed below, or open **Example 3** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**N (Sample Size)**
Confidence Level	**0.95**
N (Sample Size)	*Ignored since this is the Find setting*
Confidence Interval Width (Two-Sided)	**0.1945**
P (Proportion)	**0.034483**
Confidence Interval Formula	**Score (Contin. Correction)**
Interval Type	**Two-Sided**

Output

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

Numeric Results

Confidence Level	Sample Size (N)	Target Width	Actual Width	Proportion (P)	Lower Limit	Upper Limit	Width if P = 0.5
0.950	29	0.195	0.194	0.034	0.002	0.196	0.372

PASS also calculated the necessary sample size to be 29.

Example 4 – Zero Events, Validation using Lachin

Lachin (2000), page 19, gives an example of a calculation for a one-sided upper limit exact confidence interval for a single proportion when the confidence level is 95%, the sample proportion is 0, and the upper bound is 0.01. The necessary sample size is 299.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Confidence Intervals for One Proportion** procedure window by expanding **Proportions**, then **One Proportion**, then clicking on **Confidence Interval**, and then clicking on **Confidence Intervals for One Proportion**. You may then make the appropriate entries as listed below, or open **Example 4** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**N (Sample Size)**
Confidence Level	**0.95**
N (Sample Size)	*Ignored since this is the Find setting*
Distance from P to Limit (One-Sided)	**0.01**
P (Proportion)	**0**
Confidence Interval Formula	**Exact (Clopper-Pearson)**
Interval Type	**Upper Limit**

Output

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

Numeric Results

Confidence Level	Sample Size (N)	Target Distance from P to Upper Limit	Actual Distance from P to Upper Limit	Proportion (P)	Lower Limit	Upper Limit	Distance from P to Limit if P = 0.5
0.950	299	0.010	0.010	0.000	0	0.010	0.049

PASS also calculated the necessary sample size to be 299.

Chapter 120

Single-Stage Phase II Clinical Trials

Introduction

Phase II clinical trials determine whether a drug or regimen has sufficient activity against disease to warrant more extensive study and development. In a single-stage design, a single group of patients is studied. Usually, investigators will know the response rate of other drugs against the disease. Unless the current drug can be shown to be significantly more effective, its use will not be pursued.

This module finds designs that meet the error rate (alpha and beta) criterion and minimize the sample size when an exact test of proportions is used. The algorithm, discussed by A'Hern (2001), is an exact version of the algorithm of Fleming (1982).

Technical Details

Phase I clinical trials are designed to provide information about the maximum tolerated dose levels of a treatment. They consist of three to six patients at each dose level and provide little information about the effectiveness of the treatment.

Phase II trials obtain initial estimates of the degree of treatment activity. A patient's response may be measured by the decrease in the size of a tumor. For example, a patient may be considered to have responded to treatment if the tumor shrinks by 50% or more. There is no control group in these designs. Rather, the purpose of the trial is to determine if the drug shows enough activity against disease to warrant a full-scale, phase III clinical trial.

Let *P0* be the largest response proportion that, if true, clearly implies that the treatment does not warrant further study. *P0* is sometimes called the response rate of a poor treatment. For example, for a new anti-tumor drug, this may be set to 0.10.

Let *P1* be the smallest response proportion that, if true, clearly implies that the treatment does warrant further study. *P1* is sometimes called the response rate of a good treatment. For example, for a new anti-tumor drug, this may be set to 0.30.

A statistical test of hypothesis may be conducted to test the null hypothesis that $P \leq P0$ versus the alternative hypothesis that $P \geq P1$ (P is the true proportion responding to the treatment in the

population). Let α be the probability of rejecting the null hypothesis when it is true. Let β be the probability of rejecting the alternative hypothesis when it is true.

A single-stage phase II design can be represented by two numbers: *N* and *R*. *N* is the sample size. *R* is the critical value. If *R* or fewer responses occur in the *N* patients, the drug is rejected. The design is found by searching for the minimum value of *N* for which a value for *R* can be found such that the following two error rate constraints are met:

$$\Pr\left(reject|P0,R,N\right)\geq 1-\alpha$$

and

$$\Pr\left(reject|P1,R,N\right)\leq \beta$$

Limiting the Range of the Search

Because of the discrete nature of the binomial distribution by which these error rates are calculated, there is no closed-formed solution and so a search among possible values of *N* must be conducted. In order to speed up the search, only values of *N* between 0.8*F* and 4*F* are considered. *F* is the sample size based on the normal approximation to the binomial, suggested by Fleming (1982).

Procedure Options

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

Data Tab

Error Rates

Power or Beta

Power is the probability of rejecting the null hypothesis that the proportion responding to the treatment is less than or equal to *P0* when this hypothesis is false. That is, Power = Pr(rejecting *P<=P0|P>=P1*).

Beta is the probability of not rejecting the hypothesis that the proportion responding to the treatment is less than or equal to *P0* when this hypothesis is false. That is, Beta = Pr(not rejecting *P<=P0|P>=P1*).

The common range of power is 0.6 to 0.999 (Beta = 0.001 to 0.4). Popular values for power are 0.80 and 0.90 (Beta = 0.1 and 0.2).

Alpha

Alpha is the probability of rejecting the hypothesis that the proportion responding to the treatment is less than or equal to *P0* when this hypothesis is actually true. That is, Alpha = Pr(Rejecting *P<=P0 | P<=P0*).

The range of Alpha is 0.001 to 0.25. Popular values are 0.05 and 0.10.

Effect Size

P0 (Maximum Response Rate of a Poor Treatment)

Enter one or more response proportions of a poor drug. If the true proportion responding to the treatment is less than this amount, study of the treatment will not be recommended.

This value must be less than *P1* and greater than zero.

P1 (Minimum Response Rate of a Good Treatment)

Enter one or more response proportions of a good drug. If the true proportion responding to the treatment is greater than or equal to this amount, study of the treatment can be recommended.

This value must be greater than P0 and less than one.

Example 1 – Validation using A'Hern

A'Hern (2001) presents tables of sample sizes for various values of the design parameters. Setting alpha = 0.05, beta = 0.20, *P0* = 0.05, and *P1* = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9, A'Hern finds the corresponding sampling plans to be (using the notation *R*+1/*N*) 14/169, 4/27, 3/14, 2/7, 2/5, 2/4, 2/4, 1/1, and 1/1. This would be set up as follows.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Single-Stage Phase II Clinical Trials** procedure window by expanding **Proportions**, then **One Proportion**, then clicking on **Multi-Stage Trials**, and then clicking on **Single-Stage Phase II Clinical Trials**. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Power	**0.80**
Alpha	**0.05**
P0	**0.05**
P1	**.1 .2 .3 .4 .5 .6 .7 .8 .9**

Annotated Output

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

Numeric Results

Single Stage Design for Testing H0: P <= P0 versus H1: P >= P1

P0	P1	Alpha	Beta	Cut-Off R + 1	N	Actual Alpha	Actual Beta
0.050	0.100	0.050	0.200	14	169	0.045	0.194
0.050	0.200	0.050	0.200	4	27	0.044	0.182
0.050	0.300	0.050	0.200	3	14	0.030	0.161
0.050	0.400	0.050	0.200	2	7	0.044	0.159
0.050	0.500	0.050	0.200	2	5	0.023	0.188
0.050	0.600	0.050	0.200	2	4	0.014	0.179
0.050	0.700	0.050	0.200	2	4	0.014	0.084
0.050	0.800	0.050	0.200	1	1	0.050	0.200
0.050	0.900	0.050	0.200	1	1	0.050	0.100

Report Definitions
P0 is the maximum response proportion of a poor drug.
P1 is the minimum response proportion of a good drug.
N is the sample size.
If the number of responses >= R+1, P0 is rejected.
If the number of responses <= R, P1 is rejected.
Alpha is the probability of rejecting that P<=P0 when this is true.
Beta is the probability of rejecting that P>=P1 when this is true.

Summary Statements
A study requires 169 subjects to decide whether the proportion responding, P, is less than or equal to 0.050 or greater than or equal to 0.100. If the number of responses is 14 or more, the hypothesis that P <= 0.050 is rejected with a target error rate of 0.050 and an actual error rate of 0.045. If the number of responses is 13 or less, the hypothesis that P >= 0.100 is rejected with a target error rate of 0.200 and an actual error rate of 0.194.

Note that the designs match those of A'Hern (2001) exactly.

Chapter 125

Two-Stage Phase II Clinical Trials

Introduction

Phase II clinical trials determine whether a drug or regimen has sufficient activity against disease to warrant more extensive study and development. In a two-stage design, the patients are divided into two groups or stages. At the completion of the first stage, an interim analysis is made to determine if the second stage should be conducted. If the number of patients responding is greater than a certain amount, the second stage is conducted. Otherwise, it is not.

This module finds designs that meet the error rate (alpha and beta) criterion and minimize the expected sample size. The algorithm is discussed in Simon (1989). Extending Simon's work, our algorithm allows the investigation of near-optimal designs that may have other useful properties.

Technical Details

Phase I clinical trials are designed to provide information about the maximum tolerated dose levels of a treatment. They consist of three to six patients at each dose level and provide little information about the effectiveness of the treatment.

Phase II trials obtain initial estimates of the degree of treatment activity. A patient's response may be measured by the decrease in the size of a tumor. For example, a patient may be considered to have responded to treatment if the tumor shrinks by 50% or more. There is no control group in these designs. Rather, the purpose of the trial is to determine if the drug shows enough activity against disease to warrant a full-scale, phase III clinical trial.

Let *P0* be the largest response proportion which, if true, clearly implies that the treatment does not warrant further study. *P0* is sometimes called the response rate of a poor treatment. For a new anti-tumor drug, this may be set to 0.10.

Let *P1* be the smallest response proportion which, if true, clearly implies that the treatment does warrant further study. *P1* is sometimes called the response rate of a good treatment. For a new anti-tumor drug, this may be set to 0.30.

A statistical test of hypothesis may be conducted to test the null hypothesis that $P \leq P0$ versus the alternative hypothesis that $P \geq P1$ (P is the true proportion responding to the treatment in the population). Let α be the probability of rejecting the null hypothesis when it is true. Let β be the probability of rejecting the alternative hypothesis when it is true.

A phase II design can be represented by four numbers: *N1*, *R1*, *N*, and *R*. *N1* is the sample size in the first stage. *R1* is the critical value in the first stage. If *R1* or fewer responses occur in the *N1*

patients, the drug is rejected. N is the combined sample size for both the first and second stages. R is the critical value in the combined sample. If R or fewer of the N patients respond, the drug is rejected.

The expected (or average) sample size of this design is

$$E(N_E) = N1 + (1 - PET)(N - N1)$$

where PET is the probability of early termination of the study.

The probability of rejecting a drug with success probability P can be found using the binomial distribution. The formulation is

$$\Pr(reject|P, N1, R1, R, N) = B(R1|P, N1) + \sum_{X=R1+1}^{\min(N1,R)} b(X|P, N1)B(R - X|P, N - N1)$$

where

$$b(X|P, N) = \frac{N!}{X!(N - X)!} P^X (1 - P)^{N-X}$$

$$B(X|P, N) = \sum_{r=0}^{X} b(R|P, N)$$

The two error rate constraints are

$$\Pr(reject|P0, N1, R1, R, N) \geq 1 - \alpha$$

and

$$\Pr(reject|P1, N1, R1, R, N) \leq \beta$$

Optimum Design

The optimum design minimizes the average sample size, $E(N)$, while meeting the error rate constraints. This design is found through an exhaustive search of all possible designs. This search may take several minutes to complete.

Designs Other Than Optimal

The optimal design minimizes the average sample size. There are examples where a less-than optimal design may be more desirable. For example, suppose the optimal design were $N1 = 5$ and $N = 25$. This design is poor because only 5 patients are obtained during the first stage, but 20 are needed during the second stage. Most researchers would rather have more balance in the sample sizes of the two stages. Because of this, the actual optimal design may be rejected on other grounds.

Design Flexibility

Dealing with sequential designs is complicated. It may be difficult to achieve exactly the number of patients proscribed for each phase. However, it should be remembered that the validity of the probability statements depends on the sample size requirements being met exactly. This is

because the interpretation of an error rate probability statement is for repeated studies conducted in exactly the same way. We envision that if many studies of the same drug are conducted using the specific sampling plan *N1*, *R1*, *N*, *R* when *P = P0*, a proportion α of them will be falsely terminated due to chance occurrences.

The point is, the interpretation of the error rates is for a large number of identical studies in which the sampling plan is identical and as proscribed. If the sampling plan is allowed to vary, this interpretation is invalid. Of course, the degree of possible error in interpretation depends on the degree to which the sampling plan is changed. We recognize that when dealing with human subjects, flexibility must be maintained. However, the scientist must also recognize that when the sampling plan is changed, the exact probability statements can no longer be calculated.

Procedure Options

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

Data Tab

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

Designs

Designs to Display

This parameter specifies which designs are displayed. Since several thousand designs may be considered during the search for the optimum, it is important to limit the number of designs reported on.

The options are:

- **All designs**

  All designs considered are output. This option should only be used in special cases in which a small number of designs are tested. Otherwise, hundreds of pages of output will be generated.

- **Only designs that meet alpha & beta constraints**

  Only designs that meet the alpha and beta constraints are shown. This allows you to consider many near optimal designs which may be selected on grounds other than expected sample size.

- **Optimum designs only**

  Only the optimum design, the minimax design, and the single stage design are displayed.

Error Rates

Power or Beta

Power is the probability of rejecting the null hypothesis that the proportion responding to the treatment is less than or equal to *P0* when this hypothesis is false. That is, Power = Pr(rejecting *P<=P0|P>=P1*).

Beta is the probability of not rejecting the hypothesis that the proportion responding to the treatment is less than or equal to *P0* when this hypothesis is false. That is, Beta = Pr(not rejecting *P<=P0|P>=P1*).

The common range of power is 0.6 to 0.999 (Beta = 0.001 to 0.4). Popular values for power are 0.80 and 0.90 (Beta = 0.1 and 0.2).

Alpha

Alpha is the probability of rejecting the hypothesis that the proportion responding to the treatment is less than or equal to *P0* when this hypothesis is actually true. That is, Alpha = Pr(Rejecting *P<=P0 | P<=P0*).

The range of Alpha is 0.001 to 0.25. Popular values are 0.05 and 0.10.

Effect Size

P0 (Poor)

This is the response proportion of a poor drug. If the true proportion responding to the treatment is less than this amount, study of the treatment would not be recommended.

This value must be less than *P1* and greater than zero.

Only one value can be entered.

P1 (Good)

This is the response proportion of a good drug. If the true proportion responding to the treatment is greater than or equal to this amount, study of the treatment would be recommended.

This value must be greater than *P0* and less than one.

Only one value can be entered.

Search Parameters – N (Combined Sample Size)

Min

N is the combined sample size of the two stages of the design. This parameter sets the minimum value of *N* that is used during the search. The optimum value of *N* must be between N Min and N Max or it will not be found.

The keyword MIN indicates that the value used is the minimum of the smallest sample size from a single stage design and *MIN*2 where *MIN*2 is calculated using

$$MIN2 = \frac{p_0 + p_1}{2}\left(1 - \frac{p_0 + p_1}{2}\right)\left[\frac{z_{1-\alpha} + z_{1-\beta}}{p_1 - p_0}\right]^2$$

Since it is unlikely that the two stage sample size will be less MIN, this provides a reasonable starting point for a search for *N*. **However, experience has shown that you should use a small number such as 2 to insure that you obtain the optimum.**

You can also enter a value like MIN-x where *x* is a positive integer. This will cause the search to begin *x* units below the MIN.

The problem here is that this procedure may take a long time to run. Specifying a good starting value significantly reduces the running time.

Examples of valid entries are

2, 10, 20, MIN, MIN-1, MIN-15.

Max

N is the combined sample size of the two stages of the design. This parameter sets the maximum value of *N* used during the search. The optimum value of *N* should be between N Min and N Max or it will not be found.

The keyword BEST+X indicates that the search should try at least *X* units above the latest optimum value of *N*. For example, suppose the N Min is set at 10. The search algorithm begins at 10, and then continues by examining 11, 12, and so. Suppose that the search finds a candidate optimum at *N* = 13. To make sure that 13 is the optimum, the search continues on from 13 to 13+*X* (if, for example, *X* = 5, this value is 18). If no new optimum designs are found, the design at *N* = 13 is selected.

When using this option, *X* should be set large enough to guarantee that the true optimum can be found, but small enough so that the search does not take hours to complete. Our experience is that *X* should be greater than or equal to 8.

Examples of valid entries are for this parameter are:

20

30

BEST+8

BEST 8 (the plus sign is optional)

BEST 3

Best 4 (capitalization is not necessary)

Step

This parameter sets the step size in the search for *N*. Usually, you would enter *1* here. Occasionally, you may want to increase the step size during the initial part of your search to speed up convergence. Once you have determined a likely range, you can tighten up the search boundaries and reset this value to 1.

Search Parameters – R (Rejection Number)

Min

R is the treatment rejection number for the combined samples. If the total number of patients responding to the treatment is not greater than *R*, the treatment is deemed unworthy of further

study. R Min sets the lower boundary for *R* during the search for the optimum design. The optimum design must have an *R* value between R Min and R Max.

The recommended value for this parameter is zero. Its range is from zero to N.

Max

R is the rejection number for the combined samples. If the total number of patients responding to the treatment is not greater than *R*, the treatment is deemed unworthy of further study.

R Max sets the upper boundary for *R* during the search for the optimum design. The optimum design must have an *R* value between R Min and R Max.

Since the upper value is *N* and *N* is also a varying parameter, you can set this parameter to *MAX* or *MAX-X* (replacing *X* with an appropriate integer like 1, 2, or 3). This causes the maximum value of *R* to be set to the current value of *N-X* during each iteration of the search.

Step

This parameter sets the step size in the search for *R*. Usually, you would enter *1* here. Occasionally, you may want to increase the step size during the initial part of your search to speed up convergence. Once you have determined a likely range, you can tighten up the search boundaries and reset this value to 1.

Search Parameters – N1 (First Stage Sample Size)

Min

N1 is the sample size of the first stage. This value sets the minimum value of *N1* that is used during the search. The optimum value must be between N1 Min and N1 Max or it will not be found.

Although, in theory, the sample first stage design may have only 1, 2, or 3 patients, you may want to ignore such designs from consideration by setting this value to 4 or 5.

The actual range of this parameter is from 1 to *N*.

Max

N1 is the sample size of the first stage of the design. This parameter sets the maximum value of *N1* used during the search. The optimum value of *N1* should be between N Min and N Max or it will not be found. Although, in theory, the sample first stage design may have *N*-3, *N*-2, or *N*-1 patients, you may want to ignore such designs from consideration by setting this value to a smaller number.

Since the upper value is *N*-1 and *N* is also a varying parameter, you can set this parameter to MAX or MAX-X (replacing *X* with an appropriate value like 1 or 2). This causes the maximum value of *N1* to be set to the current value of *N-X*.

Examples:

10

20

MAX

Max-2

Max-4

Step

This parameter sets the step size in the search for *N1*. Usually, you would enter a *1* here. Occasionally, you may want to increase the step size during the initial part of your search to speed up convergence. Once you have determined a likely range, you can tighten up the search boundaries and reset this value to 1.

Search Parameters – R1 (First Stage Rejection Number)

Min

R1 is the drug rejection number for the first stage. If the number of patients responding to the treatment in the first stage is not greater than *R1*, the treatment is deemed unworthy of further study. This parameter sets the lower boundary for *R1* during the search for the optimum design. The optimum design must have an *R1* value between R Min and R Max.

The recommended value for this parameter is zero. Its range is from zero to *N1*.

Max

R1 is the rejection number for the first stage. If the number of patients responding to the treatment in the first stage is not greater than *R1*, the treatment is deemed unworthy of further study. This parameter sets the upper boundary for *R1* in the search for the optimum design.

Since the upper value is *N1* and *N1* is a varying parameter, you can set this parameter to *MAX* or MAX-X (replacing *X* with an integer like 1,2, or 3). This causes the maximum value of *R1* to be set to the current value of *N1-X*.

The valid range of *R1* is between zero and *N1*.

Step

This parameter sets the step size in the search for *R1*. Usually, you would enter a *1* here. Occasionally, you may want to increase the step size during the initial part of your search to speed up convergence. Once you have determined a likely range, you can tighten up the search boundaries and reset this value to 1.

Iterations Tab

This tab sets a couple of options used in the iterative procedures.

Maximum Iterations

Maximum Iterations Before Search Termination

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.

Example 1 – Calculating the Power

Suppose a design is wanted for the case Alpha = 0.05, Beta = 0.20, *P0* = 0.05, and *P1* = 0.25. This would be set up as follows:

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Two -Stage Phase II Clinical Trials** procedure window by expanding **Proportions**, then **One Proportion**, then clicking on **Multi-Stage Trials**, and then clicking on **Two -Stage Phase II Clinical Trials**. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Designs to Display	**Optimum designs only**
Power	**0.80**
Alpha	**0.05**
P0 (Poor)	**0.05**
P1 (Good)	**0.25**
N Min	**Min-1**
N Max	**Best+8**
N Step	**1**
R Min	**0**
R Max	**Max-3**
R Step	**1**
N1 Min	**1**
N1 Max	**Max-4**
N1 Step	**1**
R1 Min	**0**
R1 Max	**Max-1**
R1 Step	**1**

Annotated Output

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

Numeric Results

Possible Designs For P0=0.050, P1=0.250, Alpha=0.050, Beta=0.200

N1	R1	PET	N	R	Ave N	Alpha	Beta	Constraints Satisfied
16	2	0.000	16	2	16.00	0.043	0.197	Single Stage
12	0	0.540	16	2	13.84	0.043	0.199	Minimax
9	0	0.630	17	2	11.96	0.047	0.188	Optimum

Report Definitions
N1 is the sample size in the first stage.
R1 is the drug rejection number in the first stage.
PET is the probability of early termination of the study.
N is the combined sample size of both stages.
R is the combined drug rejection number after both stages.
Ave N is the average sample size if this design is repeated many times.
Alpha is the probability of rejecting that P<=P0 when this is true.
Beta is the probability of rejecting that P>=P1 when this is true.
P0 is the response proportion of a poor drug.
P1 is the response proportion of a good drug.

Summary Statements
The optimal two-stage design to test the null hypothesis that P<=0.050 versus the alternative that P>=0.250 has an expected sample size of 11.96 and a probability of early termination of 0.630. If the drug is actually not effective, there is a 0.047 probability of concluding that it is (the target for this value was 0.050). If the drug is actually effective, there is a 0.188 probability of concluding that it is not (the target for this value was 0.200). After testing the drug on 9 patients in the first stage, the trial will be terminated if 0 respond. If the trial goes on to the second stage, a total of 17 patients will be studied. If the total number responding is less than or equal to 2, the drug is rejected.

This report shows three designs. The first is the smallest single stage design. The second is the Minmax solution. This is the design with the smallest total sample size (*N*). The third is the optimum design—the one that minimizes the average sample size.

Example 2 – Validation using Simon

Simon (1989) page 4 in his Table 1 presents designs for several scenarios. The first row of the table sets *P0* to 0.05, *P1* to 0.25, Alpha to 0.10, and Beta to 0.10. The optimal design is *N1* = 9, *R1* = 0, *N* = 24, and *R* = 2. The minimax design is *N1* = 13, *R1* = 0, *N* = 20, and *R* = 2. We will now run this example through ***PASS***.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Two -Stage Phase II Clinical Trials** procedure window by expanding **Proportions**, then **One Proportion**, then clicking on **Multi-Stage Trials**, and then clicking on **Two -Stage Phase II Clinical Trials**. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Which Designs	**Optimum designs only**
Power	**0.90**
Alpha	**0.10**
P0 (Poor)	**0.05**
P1 (Good)	**0.25**
N Min	**Min-1**
N Max	**Best+8**
N Step	**1**
R Min	**0**
R Max	**Max-3**
R Step	**1**
N1 Min	**1**
N1 Max	**Max-4**
N1 Step	**1**
R1 Min	**0**
R1 Max	**Max-1**
R1 Step	**1**

Output

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

Numeric Results

Possible Designs For P0=0.050, P1=0.250, Alpha=0.050, Beta=0.200

N1	R1	PET	N	R	Ave N	Alpha	Beta	Constraints Satisfied
20	2	0.000	20	2	20.00	0.075	0.091	Single Stage
13	0	0.513	20	2	16.41	0.074	0.097	Minimax
9	0	0.630	24	2	14.55	0.093	0.097	Optimum

PASS has calculated exactly the same optimal design and minimax design.

Chapter 130

Three-Stage Phase II Clinical Trials

Introduction

Phase II clinical trials determine whether a drug or regimen has sufficient activity against disease to warrant more extensive study and development. In a three-stage design, the patients are divided into three groups or stages. At the completion of the first stage, an interim analysis is made to determine if the second stage should be conducted. If the number of patients responding is greater than a certain amount, the second stage is conducted. Otherwise, it is not. A similar interim analysis is conducted at the end of the second stage.

This module finds designs that meet the error rate (alpha and beta) criterion and minimize the expected sample size. The formulation is given in Chen (1997). Extending Chen's work, our algorithm allows the investigation of near-optimal designs that may have other useful properties.

Technical Details

Phase I clinical trials are designed to provide information about the maximum tolerated dose levels of a treatment. They consist of three to six patients at each dose level and provide little information about the effectiveness of the treatment.

Phase II trials obtain initial estimates of the degree of treatment activity. A patient's response may be measured by the decrease in the size of a tumor. For example, a patient may be considered to have responded to treatment if the tumor shrinks by 50% or more. There is no control group in these designs. Rather, the purpose of the trial is to determine if the drug shows enough activity against disease to warrant a full-scale, phase III clinical trial.

Let *P0* be the largest response proportion which, if true, clearly implies that the treatment does not warrant further study. *P0* is sometimes called the response rate of a *poor* treatment. For a new anti-tumor drug, this may be set to 0.10.

Let *P1* be the smallest response proportion which, if true, clearly implies that the treatment does warrant further study. *P1* is sometimes called the response rate of a *good* treatment. For a new anti-tumor drug, this may be set to 0.30.

A statistical test of hypothesis may be conducted to test the null hypothesis that $P \le P0$ versus the alternative hypothesis that $P \ge P1$ (P is the true proportion responding to the treatment in the population). Let α be the probability of rejecting the null hypothesis when it is true. Let β be the probability of rejecting the alternative hypothesis when it is true.

A three-stage phase II design can be represented by six numbers: *R1*, *N1*, *R2*, *N2*, *R3* and *N3*. *N1* is the sample size in the first stage. *R1* is the critical value in the first stage. If *R1* or fewer responses occur in the *N1* patients, the drug is rejected. *N2* is the total sample size of stages one and two. *R2* is the critical value in the second stage. If *R2* or fewer responses occur in the *N2* patients, the drug is rejected. *N*3 is the combined sample size of all three stages. *R3* is the critical value in the combined sample. If *R3* or fewer of the *N3* patients respond, the drug is rejected.

The expected (or average) sample size of this design is

$$E(N_E) = N1 + (1 - PET1)(N2 - N1) + (1 - PET2)(N3 - N2)$$

where *PET1* is the probability of early termination of the study after stage one and *PET2* is the probability of early termination after stage two.

The probability of rejecting a drug with success proportion *P* can be found using the binomial distribution. The formulation is

$$\Pr(reject|P, N1, R1, R2, N2, R3, N3) = PET1 + PET2 + PET3$$

where

$$PET1 = B(R1|P, N1)$$

$$PET2 = \sum_{X1=R1+1}^{\min(N1,R2)} b(X1|P, N1)B(R2 - X1|P, N2 - N1)$$

$$PET3 = \sum_{X1=R1+1}^{\min(N1,R3)} b(X1|P, N1) \sum_{X2=R2+1-X1}^{\min(N3-N2,R3-X1)} b(X2|P, N2 - N1)B(R3 - X1 - X2|P, N3 - N2)$$

$$b(X|P, N) = \frac{N!}{X!(N - X)!} P^X (1 - P)^{N-X}$$

$$B(X|P, N) = \sum_{r=0}^{X} b(R|P, N)$$

The two error rate constraints are

$$\Pr(reject|P0, N1, R1, R2, N2, R3, N3) \ge 1 - \alpha$$

and

$$\Pr(reject|P1, N1, R1, R2, N2, R3, N3) \le \beta$$

Optimum Design

The optimum design minimizes the average sample size, $E(N)$, while meeting the error rate constraints. This design is found through an exhaustive search of all possible designs. This search may take several minutes to complete.

Designs Other Than Optimal

The optimal design minimizes the average sample size. There are examples where a less-than optimal design may be more desirable. For example, suppose the optimal design were $N1 = 5$, $N2 = 25$, and $N3 = 26$. This design is poor because the bulk of the subjects are tested in the second phase. Most researchers would rather have more balance in the sample sizes of the three stages. For reasons like this, the actual optimal design may be replaced by another, sub-optimal, design.

Design Flexibility

Dealing with sequential designs is complicated. It may be difficult to achieve exactly the number of patients proscribed for each phase. However, it should be remembered that the validity of the probability statements depends on the sample size requirements being met exactly. This is because the interpretation of an error rate probability statement is for repeated studies conducted in exactly the same way. We envision that if many studies of the same drug are conducted using the specific sampling plan when $P = P0$, a proportion α of them will be falsely terminated due to chance occurrences.

The point is, the interpretation of the error rates is for a large number of identical studies in which the sampling plan is identical and as proscribed. If the sampling plan is allowed to vary, this interpretation is invalid. Of course, the degree of possible error in interpretation depends on the degree to which the sampling plan is changed. We recognize that when dealing with human subjects, flexibility must be maintained. However, the researcher must also recognize that when the sampling plan is changed, the exact probability statements can no longer be calculated.

Procedure Options

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

Data Tab

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

Designs

Designs to Display

This parameter controls which designs are displayed. Since several thousand designs may be considered during the search for the optimum, it is important to limit the number of designs reported on.

The options are:

- **All designs**

  All designs considered are output. This option should only be used in special cases in which a small number of designs are tested. Otherwise, hundreds of pages of output will be generated.

- **Only designs that meet alpha & beta constraints**

  Only designs that meet the alpha and beta constraints are shown. This allows you to consider many near optimal designs which may be selected on grounds other than expected sample size.

- **Optimum designs only**

  Only the optimum design, the minimax design, and the single stage design are displayed.

Error Rates

Power or Beta

Power is the probability of rejecting the null hypothesis that the proportion responding to the treatment is less than or equal to *P0* when this hypothesis is false. That is, Power = Pr(rejecting *P<=P0|P>=P1*).

Beta is the probability of not rejecting the hypothesis that the proportion responding to the treatment is less than or equal to *P0* when this hypothesis is false. That is, Beta = Pr(not rejecting *P<=P0|P>=P1*).

The common range of power is 0.6 to 0.999 (Beta = 0.001 to 0.4). Popular values for power are 0.80 and 0.90 (Beta = 0.1 and 0.2).

Alpha

Alpha is the probability of rejecting the hypothesis that the proportion responding to the treatment is less than or equal to *P0* when this hypothesis is actually true. That is, Alpha = Pr(Rejecting *P<=P0 | P<=P0*).

The range of Alpha is 0.001 to 0.25. Popular values are 0.05 and 0.10.

Effect Size

P0 (Poor)

This is the response proportion of a poor drug. If the true proportion responding to the treatment is less than this amount, study of the treatment would not be recommended.

This value must be less than *P1* and greater than zero.

Only one value can be entered.

P1 (Good)

This is the response proportion of a good drug. If the true proportion responding to the treatment is greater than or equal to this amount, study of the treatment would be recommended.

This value must be greater than *P0* and less than one.

Only one value can be entered.

Search Parameters – N (Combined Sample Size)

Min

N is the combined sample size of the three stages of the design. This parameter sets the minimum value of *N3* that is used during the search. The optimum value of *N3* must be between N Min and N Max or it will not be found.

The keyword MIN indicates that the value used is the minimum of the smallest sample size from a single stage design and *MIN*2 where *MIN*2 is calculated using

$$MIN2 = \frac{p_0 + p_1}{2}\left(1 - \frac{p_0 + p_1}{2}\right)\left[\frac{z_{1-\alpha} + z_{1-\beta}}{p_1 - p_0}\right]^2$$

Since it is unlikely that the three stage sample size will be less MIN, this provides a reasonable starting point for a search for *N*. You can also enter a value like MIN-x where *x* is a positive integer. This will cause the search to begin *x* units below the MIN.

This procedure may take a long time to run. Specifying a good starting value significantly reduces the running time.

Examples of valid entries are

2, 10, 20, MIN, MIN-1, MIN-15.

Max

N is the combined sample size of the three stages of the design. This parameter sets the maximum value of *N3* used during the search. The optimum value of *N3* should be between N Min and N Max or it will not be found.

The keyword BEST+X indicates that the search should try at least *X* units above the latest optimum value of *N3*. For example, suppose the N Min is set at 10. The search algorithm begins at 10, and then continues by examining 11, 12, and so. Suppose that the search finds a candidate optimum at *N* = 13. To make sure that 13 is the optimum, the search continues on from 13 to 13+*X* (if, for example, *X* = 5, this value is 18). If no new optimum designs are found, the design at *N3* = 13 is selected.

When using this option, *X* should be set large enough to guarantee that the true optimum can be found, but small enough so that the search does not take hours to complete. Our experience indicates that *X* should be greater than or equal to 8.

Examples of valid entries for this parameter are:

20

30

BEST+8

BEST 8 (the plus sign is optional)

BEST 3

Best 4 (capitalization is not necessary)

Step

This parameter sets the step size in the search for *N*. Usually, you would enter *1* here. Occasionally, you may want to increase the step size during the initial part of your search to speed up convergence. Once you have determined a likely range, you can tighten up the search boundaries and reset this value to 1.

Search Parameters – R (Rejection Number)

Step

This parameter sets the step size in the search for *R*. Usually, you would enter *1* here. Occasionally, you may want to increase the step size during the initial part of your search to speed up convergence. Once you have determined a likely range, you can tighten up the search boundaries and reset this value to 1.

Search Parameters – M1 (First Stage Sample Size)

Min

M1 is the sample size in the first stage. Thus, M1=N1.This option sets a minimum value for M1 so that designs with a very small value are not considered. For example, if N3=100, you might want to only consider designs with at least M1=25. You can enter either an actual amount (no percent sign) or a percentage of N3. If you enter '%' after the value, the value is taken to mean a percentage of N3. For example, suppose that N3=50. Then the entry of 30% would result in M1=15. However, if you did not enter the '%', M1=30.

Note that it up to you to enter meaningful values. For example, you would not want to enter values for M1-Min, M2-Min, and M3-Min that total to more than 100% because no designs would be considered.

Step

This parameter sets the step size in the search for *M1*. Usually, you would enter a *1* here. Occasionally, you may want to increase the step size during the initial part of your search to speed up convergence. Once you have determined a likely range, you can tighten up the search boundaries and reset this value to 1.

Search Parameters – M2 (Second Stage Sample Size)

Min

M2 is the number of subjects added in the second stage. Thus, M2 = N2 - N1 and M1+M2+M3=N3. This option sets a minimum value for M2 so that designs with a very small value are not considered. For example, if N3=100, you might want to only consider designs with at least M2=25. You can enter either an actual amount (no percent sign) or a percentage of N3. If you enter '%' after the value, the value is taken to mean a percentage of N3. For example, suppose that N3=50. Then the entry of 30% would result in M2=15. However, if you did not enter the '%', M2=30.

Note that it up to you to enter meaningful values. For example, you would not want to enter values for M1-Min, M2-Min, and M3-Min that total to more than 100% because no designs would be considered.

Step

This parameter sets the step size in the search for *M2*. Usually, you would enter a *1* here. Occasionally, you may want to increase the step size during the initial part of your search to speed up convergence. Once you have determined a likely range, you can tighten up the search boundaries and reset this value to 1.

Search Parameters – M3 (Third Stage Sample Size)

Min

M3 is the number of subjects added in the third stage. Thus, M3 = N3 - N2 and M1+M2+M3=N3. This option sets a minimum value for M3 so that designs with a very small value are not considered. For example, if N3=100, you might want to only consider designs with at least M3=25. You can enter either an actual amount (no percent sign) or a percentage of N3. If you enter '%' after the value, the value is taken to mean a percentage of N3. For example, suppose that N3=50. Then the entry of 30% would result in M3=15. However, if you did not enter the '%', M3=30.

Note that it up to you to enter meaningful values. For example, you would not want to enter values for M1-Min, M2-Min, and M3-Min that total to more than 100% because no designs would be considered.

Iterations Tab

This tab sets a couple of options used in the iterative procedures.

Maximum Iterations

Maximum Iterations Before Search Termination

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 results are left blank. A value of 500 is recommended.

Example 1 – Calculating the Power and Validation using Chen

Chen (1997) provides the minimax and optimum design for the case Alpha = 0.05, Beta = 0.20, *P0* = 0.05, and *P1* = 0.25. The optimum design is 0/8, 1/13, and 2/19. The minimax design is 0/12, 1/15, and 2/16.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Three-Stage Phase II Clinical Trials** procedure window by expanding **Proportions**, then **One Proportion**, then clicking on **Multi-Stage Trials**, and then clicking on **Three-Stage Phase II Clinical Trials**. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Designs to Display	**Optimum designs only**
Power	**0.80**
Alpha	**0.05**
P0 (Poor)	**0.05**
P1 (Good)	**0.25**
N Min	**Min**
N Max	**Best+2**

Note that the search may take several minutes to run, depending on the speed of your computer.

Annotated Output

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

Numeric Results

Possible Designs For P0=0.050, P1=0.250, Alpha=0.050, Beta=0.200

Stage 1 R1/N1	Stage 2 R2/N2	Stage 3 R3/N3	Ave N	Stage 1 Pet P0	Overall Pet P0	Alpha	Beta	Constraints Satisfied
2/16	2/16	2/16	16.00	0.000	0.000	0.043	0.197	Single Stage
0/12	1/15	2/16	13.55	0.833	0.869	0.043	0.199	Minimax
0/8	1/13	2/19	10.41	0.663	0.880	0.049	0.195	Optimum

Report Definitions
N1 is the sample size in the first stage.
R1 is the drug rejection number in the first stage.
N2 is the sample size in the first and second stages.
R2 is the drug rejection number in the second stage.
N3 is the combined sample size of all three stages.
R3 is the drug rejection number in the third stage.
Stage 1 PET P0 is the probability of early termination at the first stage.
Stage 2 PET P0 is the probability of early termination at the second stage.
Ave N is the average sample size if this design is repeated many times.
Alpha is the probability of rejecting that P<=P0 when this is true.
Beta is the probability of rejecting that P>=P1 when this is true.
P0 is the response proportion of a poor drug.
P1 is the response proportion of a good drug.

This report shows three designs. The first is the smallest single stage design. The second is the Minmax solution. This is the design with the smallest total sample size (*N*). The third is the optimum design—the one that minimizes the average sample size.

Note that ***PASS*** matches the results of Chen.

Chapter 135

Post-Marketing Surveillance

Introduction

Post-marketing surveillance refers to the search for adverse reactions to drugs that have been cleared for general use. Two types of study designs are often used: the cohort study and the case-control study. In a cohort design, a large group of treated patients are studied to determine the frequency of any adverse reactions. In a case-control study, patients who have experienced the adverse reaction are matched with other treated patients who have not.

Technical Details

This section presents the formulas used to calculate sample size and power in four situations. The theory and formulas provided by Machin *et al.* (1997) are used.

Design Type 1 – Cohort Study, No Background Incidence of Adverse Reactions

Let the anticipated incidence rate of adverse reactions be *R0*, the number of occurrences of a particular adverse reaction be *A*, the number of patients be *N*, and the probability that you will not find *A* reactions in the sample of *N* patients be β. If *R0* is small, the occurrence of an adverse reaction may be assumed to follow the Poisson distribution. If this is the case, the relationship among the above parameters is

$$\beta = \sum_{i=0}^{A-1} \frac{N^i (R0)^i e^{-N(R0)}}{i!}$$

Using numerical search techniques, ***PASS*** is able to solve any one of these parameters in terms of the others.

Design Type 2 – Cohort Study, Known Background Incidence of Adverse Reactions

Let the anticipated incidence rate of adverse reactions be *R0*, let the additional incidence rate caused by the drug be *D*, and let the number of patients be *N*. For a given significance level α and power $1-\beta$, the relationship between these parameters is

$$z_{1-\beta} = \frac{D\sqrt{N} - z_{1-\alpha}\sqrt{R0}}{\sqrt{R0 + D}}$$

Design Type 3 – Cohort Study, Unknown Background Incidence of Adverse Reactions

A control group is needed when the background incidence rate is not known. In post-marketing surveillance studies, the control group is usually made up of untreated individuals. Let the anticipated incidence rate of adverse reactions be *R0*, let the additional incidence rate caused by the drug be *D*, let the number of patients be *N*, and let the number of control patients for each treated patient be *M*. Thus the number of control patients is *NM*. For a given significance level α and power $1-\beta$, the relationship between these parameters is

$$z_{1-\beta} = \frac{D\sqrt{MN} - z_{1-\alpha}\sqrt{(M+1)R(1-R)}}{\sqrt{MR0(1-R0) + (R0+D)(1-R0-D)}}$$

where

$$R = \frac{R0 + M(R0+D)}{1+M}$$

Design Type 4 – Matched Case-Control Study

A case-control design involves identifying a group of patients that have experienced the reaction of interest and then obtaining matched control patients that have not experienced the reaction.

Let the anticipated incidence rate of adverse reactions be *R0*, let the additional incidence rate caused by the drug be *D*, let the number of patients be *N*, and let the number of control patients matched with each treated patient be *M*. For a given significance level α and power $1-\beta$, the relationship between these parameters is

$$z_{1-\beta} = \frac{|R0-\Omega|\sqrt{N} - z_{1-\alpha}\sqrt{\left(1+\frac{1}{M}\right)\Pi(1-\Pi)}}{\sqrt{\frac{R0(1-R0)}{M} + (\Omega)(1-\Omega)}}$$

where

$$\Omega = \frac{R0 + D}{1 + D}$$

$$\Pi = \left(\frac{R0}{1 + D}\right)\left(M + \frac{\Omega}{R0}\right)$$

Procedure Options

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

Data Tab

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

Design

Design Type

This parameter specifies which of the four possible designs is to be analyzed.

The possible designs are:

- **[1] Cohort - No Background Incidence**
  This is a cohort design in which the adverse effect does not occur except when caused by the drug.
- **[2] Cohort - Known Background Incidence**
  This is a cohort design. The adverse effect can occur without being related to the drug. The incidence rate of the adverse effect is known.
- **[3] Cohort - Unknown Background Incidence**
  This is a cohort design. Although the adverse effect can occur, its incidence rate is not known. A control group must be followed to determine the background incidence rate.
- **[4] Matched Case-Control Study**
  One or more control patients is matched with each case patient. All of these patients are in the study. This is different from design type 3 in that the controls are matched with the cases. In design type 3, no matching is done.

Solve For

Find (Solve For)

This option specifies the parameter to be solved for from the other parameters.

Error Rates

Power or Beta

This option specifies one or more values for power or beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected.

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

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

If your only interest is in determining the appropriate sample size for a confidence interval, set power or beta to 0.5.

Note that the interpretation of Power or Beta is a little different when the Design Type is 1.

Alpha

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected.

Values between 0.001 and 0.100 are most common. The value of 0.05 is often a standard. This means that about one test in twenty will falsely reject the null hypothesis. Although 0.05 is a 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 0.10* or *0.01 to 0.05 by 0.01*.

Sample Size

N (Patients)

This is the number of patients in the cohort being studied. In the case-control designs, this is the number of cases. For case-control studies, the total number of patients in the study is $N(1+M)$ where M is the number of control patients per case patient.

M (Controls Per Case)

This is the number of control patients for each case patient. For case-control studies, the total number of patients in the study is $N(1+M)$ where N is the number of case patients.

Effect Size – Incidence Rates

R0 (Incidence Rate)

This is the background incidence rate of the adverse reaction. This is the rate that occurs in the population without the drug being monitored. Since this is an incidence rate, and hence a proportion, it should be less than one. Also, this value plus *D* must be less than one.

D (Additional Incidence Rate)

This is the additional incidence rate of the adverse reaction that can be attributed to the drug being studied. Since this is a rate, it should not be greater than one. Also, this value plus *R0* should be less than one.

Effect Size – Occurrences

A (Number of Occurrences)

This is the number of occurrences of the adverse reaction of interest in the *N* patients being monitored. Sometimes, a single drug-related adverse reaction (such as death) might be enough to make the drug unacceptable. The acceptable range is one or greater.

Test

One-Sided

This option lets you designate whether the test will be one-sided (checked) or two-sided (unchecked). When the two-sided box is indicated, the alpha value is simply divided by two.

Iterations Tab

This tab sets a couple of options used in the iterative procedures.

Maximum Iterations

Maximum Iterations Before Search Termination

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.

Iterative Precision

When a search is made for the precision value, this is the cutoff value used to terminate the search. In most cases, a value of 0.0001 will be more than sufficient.

Example 1 – Calculating the Power

Suppose 1 in 10,000 people receiving a certain drug are expected to have an irregular heart beat. A researcher decides that if the irregular heart beat occurs in three patients, the drug will have to be withdrawn. Study the sample size necessary to achieve 99% probability of success.

In order to do this, sample sizes between 1000 and 21,000 will be considered.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Post-Marketing Surveillance** procedure window by expanding **Proportions**, then **One Proportion**, then clicking on **Post-Marketing Surveillance**, and then clicking on **Post-Marketing Surveillance**. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Design Type	**[1] Cohort - No Background Incidence**
Find (Solve For)	**Power and Beta**
Power	*Ignored*
Alpha	*Ignored*
N (Patients)	**1000 to 21000 by 4000**
M (Controls Per Case)	*Ignored*
R0 (Background Incidence Rate)	*Ignored*
D (Additional Incidence Rate)	**0.0001**
A (Number of Occurrences)	**1 2 3**
One-Sided	**Checked**
Reports Tab	
Incidence Rate Decimals	**5**

Annotated Output

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

Numeric Results

Numeric Results (Cohort Study with No Background Incidence)

Power	Sample Size (N)	Additional Incidence Rate (D)	Number of Occurrences (A)	Beta
0.09516	1000	0.00010	1	0.90484
0.00468	1000	0.00010	2	0.99532
0.00015	1000	0.00010	3	0.99985
0.39347	5000	0.00010	1	0.60653
0.09020	5000	0.00010	2	0.90980
0.01439	5000	0.00010	3	0.98561
0.59343	9000	0.00010	1	0.40657
0.22752	9000	0.00010	2	0.77248
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.

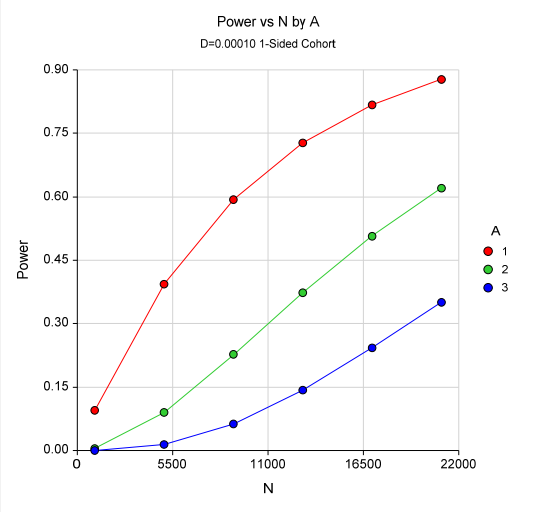
(output continues)

Report Definitions
Power is 1 - Beta.
N is the number of patients monitored.
D is the expected incidence rate of adverse reactions.
A is the number of adverse reactions.
Beta is the probability that A reactions will not be found in the N patients.

Summary Statements
In a cohort study with no background incidence of a particular adverse reaction, the probability that 1 or more adverse reactions will not occur in a sample of 1000 patients with an anticipated incidence rate of 0.00 is 0.90484. The power of this study is 10%.

This report shows the calculated sample size for each of the scenarios.

Plots Section



This plot shows the power versus the sample size for three values of *A*.

Example 2 – Validation using Machin

Machin *et al.* (1997) page 147 give an example of a cohort design with no background incidence in which *N* is 30000, incidence is 0.0001, and *A* is 1 or 2. When *A* is 1, the power is 95%. When *A* is 2, the power is 80%.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Post-Marketing Surveillance** procedure window by expanding **Proportions**, then **One Proportion**, then clicking on **Post-Marketing Surveillance**, and then clicking on **Post-Marketing Surveillance**. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Design Type	**[1] Cohort - No Background Incidence**
Find (Solve For)	**Power and Beta**
Power	*Ignored*
Alpha	*Ignored*
N (Patients)	**30000**
M (Controls Per Case)	*Ignored*
R0 (Background Incidence Rate)	*Ignored*
D (Additional Incidence Rate)	**0.0001**
A (Number of Occurrences)	**1 2**
One-Sided	**Checked**
Reports Tab	
Incidence Rate Decimals	**5**

Output

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

Numeric Results

Numeric Results (Cohort Study with No Background Incidence)

Power	Sample Size (N)	Additional Incidence Rate (D)	Number of Occurrences (A)	Beta
0.95021	30000	0.00010	1	0.04979
0.80085	30000	0.00010	2	0.19915

PASS has calculated the same power values as did Machin *et al* (1997).

Example 3 – Validation using Machin

Machin *et al.* (1997) page 148 give an example of a cohort design with unknown background incidence in which *N* is 8500, *R0* is 0.01, *D* is 0.005, and *A* is 1. The power is 90%.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Post-Marketing Surveillance** procedure window by expanding **Proportions**, then **One Proportion**, then clicking on **Post-Marketing Surveillance**, and then clicking on **Post-Marketing Surveillance**. You may then make the appropriate entries as listed below, or open **Example 3** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Design Type	**[3] Cohort - Unknown Background Incidence**
Find (Solve For)	**Power and Beta**
Power	*Ignored*
Alpha	**0.05**
N (Patients)	**8500**
M (Controls Per Case)	**1**
R0 (Background Incidence Rate)	**0.01**
D (Additional Incidence Rate)	**0.005**
A (Number of Occurrences)	*Ignored*
One-Sided	**Checked**
Reports Tab	
Incidence Rate Decimals	**5**

Output

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

Numeric Results

Numeric Results (Cohort Study with No Background Incidence)

Power	Sample Size (N)	Controls Per Case (M)	Background Incidence Rate (R0)	Additional Incidence Rate (D)	One-Sided Alpha	Beta
0.90136	8500	1	0.01000	0.00500	0.05000	0.09864

PASS has calculated the same power value as did Machin *et al* (1997).

Chapter 150

Inequality Tests for Two Correlated Proportions (McNemar Test)

Introduction

McNemar's test compares the proportions for two correlated dichotomous variables. These two variables may be two responses on a single individual or two responses from a matched pair (as in matched case-control studies).

This procedure is similar to the Matched Case-Control procedure also available in ***PASS***. It differs from that procedure in three basic ways:

1. The results are calculated exactly using an unconditional formula rather than using conditional, normal approximations to the binomial.
2. It only deals with the case of a matched pair: one case and one control (the Matched Case-Control procedure lets you match several controls with each case).
3. It is based directly on the 2-by-2 contingency table.

To fix these ideas, consider the following fictitious data concerning the relationship between smoking and lung cancer. Suppose that a sample of $N = 100$ cases of identical twins in which only one twin has lung cancer is selected for further study. The twin with lung cancer is the *case*. The other twin serves as the *control*. Each pair of twins is surveyed to determine if they smoke tobacco. The results are summarized in the following two-way table:

	No Lung Cancer Twin (Control)	
Lung Cancer Twin (Case)	**Smokes = Yes**	**Smokes = No**
Smokes = Yes	16	21
Smokes = No	4	59

There is a basic difference between this table and the more common two-way table. In the matched-paired case, the count represents the number of pairs, not the number of individuals.

The investigator wishes to compare the proportion of cases that smoke with the proportion of controls that smoke. The proportion of controls who smoke is (16+4)/100 = 0.20. The proportion of cases who smoke is (16+21)/100 = 0.37.

Dividing each of the entries in the table by *N* gives the proportions:

	No Lung Cancer Twin (Control)		
Lung Cancer Twin (Case)	**Smokes = Yes**	**Smokes = No**	**Total**
Smokes = Yes	0.16	0.21	0.37
Smokes = No	0.04	0.59	0.63
Total	0.20	0.80	0.63

Symbolically, this table is represented as:

	No Lung Cancer Twin (Control)		
Lung Cancer Twin (Case)	**Smokes = Yes**	**Smokes = No**	**Total**
Smokes = Yes	P11	P10	Pt
Smokes = No	P01	P00	1-Pt
Total	Ps	1-Ps	1

Formally, the hypothesis of interest is that *Pt* equals *Ps*. A little algebra shows that *Pt* = *Ps* is equivalent to *P10* = *P01*, since *P11* is common to both. Thus, the null hypothesis of McNemar's test is *P10* = *P01* and the alternative is that they are unequal. The alternative hypothesis may be one-sided (such as *P10*>*P01*) or two-sided (*P10* ≠ *P01*).

The null hypothesis may also be stated in terms of the odds ratio as *OR* = 1. The odds ratio is computed differently in the matched pairs case. The formula is:

$$OR = \frac{P10}{P01}$$

Notice that the values of *P11* and *P00* are not used in these hypotheses. It turns out that their individual values are not needed, but their sum is.

For this example, the odds ratio is computed as 21/4 = 5.25.

Technical Details

Consider the matched-pairs table again:

	Controls		
Cases	**Yes**	**No**	**Total**
Yes	P11	P10	Pt
No	P01	P00	1-Pt
Total	Ps	1-Ps	1

Pairs with the same response from cases and controls (Yes-Yes and No-No) are called *concordant* pairs. Pairs with different responses (Yes-No and No-Yes) are call *discordant* pairs. McNemar's test statistic is the estimated odds ratio:

$$Mc = \frac{P10}{P01}$$

The sample size problem thus reduces to a study of how many Yes-No's and No-Yes's are needed. Once this has been determined, the overall sample size is found be estimating the proportion of discordant pairs and inflating the sample size appropriately.

Some power analysis programs follow an approximate procedure. Since the McNemar statistic follows the binomial probability distribution for a fixed number of discordant pairs, they use formulas that use the normal approximation to the binomial and then adjust the sample size depending on the proportion of discordant pairs, *PD=P10+P01*. This is called the conditional procedure.

One such approximate formula is given by Machin, Campbell, Fayers, and Pinol (1997).

$$N_{pairs} = \frac{\left\{z_{1-\alpha/s}(OR+1) + z_{1-\beta}\sqrt{(OR+1)^2 - (OR-1)^2 PD}\right\}^2}{(OR-1)^2 PD}$$

where *s* is the number of sides to the test (one or two), $OR = \frac{P10}{P01}$, $PD = P10 + P01$, and alpha and beta are the usual error rate probabilities.

However, Schork and Williams (1980) published a formula which provides the exact results for the unconditional case. This is the formulation that is used in ***PASS***.

$$Power = \sum_{R=r}^{N} \sum_{n_{12}=0}^{IR} \frac{N!}{(N-R)!n_{12}!(R-n_{12})!}(1-PD)^{N-R}\left(\frac{D+PD}{2}\right)^{n_{12}}\left(\frac{PD-D}{2}\right)^{R-n_{12}}$$

where

PD = P10+P01

D = P10-P01

N is total of all four cells (*N11+N12+N21+N22*)

r is the smallest integer for which $\left(\frac{1}{2}\right)^r \le \alpha$

IR is the largest integer such that $\sum_{j=0}^{IR}\binom{R}{j}\left(\frac{1}{2}\right)^R \le \alpha$

Difference or Odds Ratio

The formula given above is parameterized in terms of the difference. This formula is also used when odds ratios are specified. The program simply converts the OR value into its corresponding D value.

Discussion

The exact algorithm works for $N < 2000$. Above 2000, computing time goes up and the algorithm has numerical problems. ***PASS*** lets you select either the exact, or the approximate, solutions. We

have found that the approximate solution tends to use a sample size that is about 10% less than the exact solution.

Because of the lengthy computer time necessary to compute the exact answer when $N > 1500$, we suggest that you use the approximate formula to begin with and then revert to the exact formula when you are ready for your final results. This is based on the value of *N* specified in the 'Use Approximations if *N* >' box of the Options tab.

Procedure Options

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

Data Tab (Common Options)

The Data tab contains the parameters associated with this test such as the proportions, sample sizes, alpha, and beta. This chapter covers two procedures, which have different options. This section documents options that are common to both procedures. Later, unique options for each procedure will be documented.

Solve For

Find (Solve For)

This option specifies the parameter to be solved for from the other parameters. The parameters that may be selected are *Odds Ratio*, *Difference*, *N*, *Alpha*, and *Power and Beta*. Under most situations, you will select either *Power and Beta* to calculate power or *N* to calculate sample size.

The program is set up to evaluate power directly. For the other parameters, a search is made using an iterative procedure until an appropriate value is found. Two solutions can often be found when searching for the *Odds Ratio* or the *Difference*. You may specify the region in which you want the solution to be searched for. For example, you may search for an odds ratio either above or below one. Also note that the parameter selected must match the procedure you are using. For example, if you are searching for the odds ratio, you must be in the odds ratio window.

Note that the value selected here always appears as the vertical axis on the charts.

Error Rates

Power or Beta

This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected.

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

If your only interest is in determining the appropriate sample size for a confidence interval, set power or beta to 0.5.

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 a true null hypothesis is rejected.

Sample Size

N (Number of Pairs)

Enter a value (or range of values) for the sample size, *N*, the number of pairs in the study.

You may enter a range of values such as *100,200,300* or *200 to 400 by 50*.

Effect Size

Proportion Discordant (P10+P01)

This is the proportion of discordant pairs (*P10 + P01*). This value will be difficult to specify unless you have previous studies that give you some idea of what to expect. When you have no idea, Machin, Campbell, Fayers, and Pinol (1997) suggest the following strategy. Estimate values of *Pt* and *Ps*. Calculate the proportion of discordant pairs using the approximation

$$PD = Pt(1 - Ps) + Ps(1 - Pt)$$

This approximation assumes that the two responses are independent in each subject, which will usually not be true. However, it may be the only way of determining a ball park value for this parameter.

Test

Alternative Hypothesis

Specify whether the test is one-sided or two-sided. A one-sided hypothesis uses an inequality as in *P10 > P01* or Odds Ratio > 1. A two-sided hypothesis states that the proportions are not equal without specifying which is greater. If you do not have any special reason to do otherwise, you should use the two-sided option.

Data Tab (Differences)

This section documents options that are used when the parameterization is in terms of the difference, P10 - P01.

Effect Size

Difference (P10-P01)

The difference, *P10-P01*, is a popular parameter to specify because it comes directly from the alternative hypothesis that *Pt* not equal *Ps*. If *Pt* and *Ps* are not equal, the obvious question is, by how much? The answer is, by *P10-P01* since *Pt-Ps = (P11+P10)-(P11+P01) = P10-P01*. Hence, this is a value that may easily be set.

The range of values is between -1 and 1. You may enter a list of value list *0.1, 0.15, 0.2* or *0.05 to 0.20 by 0.05*.

Data Tab (Odds Ratios)

This section documents options that are used when the parameterization is in terms of the odds ratio, P10 / P01.

Effect Size

Odds Ratio (P10/P01)

The odds ratio is a popular parameter to specify because of its simple interpretability and close relationship to the relative risk. An odds ratio of 2.0 means that the odds of the numerator variable is twice the odds of the denominator variable. Note that several values of *P10* and *P01* can yield the same odds ratio. For example, 0.2/0.1 and 0.4/0.2 both have an odds ratio of 2.0, but are based on very different values of *P10* and *P01*. Under the null hypothesis, the odds ratio is one. Only positive values are allowed.

You may enter a list of values like *1.5,2.0,3.0* or *1.2 to 2.4 by 0.2*.

Options Tab

This tab sets a couple of options used in the iterative procedures and approximations.

Maximum Iterations

Maximum Iterations Before Search Termination

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.

Approximations

Use Approximations if N is greater than

Below this value of N, the exact power calculation formula based on the binomial is used. Above this value of N, the approximate formula based on the normal approximation to the binomial is used. The exact formula suffers from numerical problems when N is greater than 2000. On the other hand, the approximate formula tends to underestimate the N necessary to achieve a certain beta value by about 5%.

You control which formula is used by setting this value.

Example 1 – Calculating Power using Odds Ratios

This example will show how to calculate the power of a retrospective study for several sample sizes and odds ratio values. Suppose that a matched case-control study is to be run in which the odds ratios might be 1.5, 2.5, or 3.5; *PD* is 0.3, *N* = 50 to 300 by 50; alpha is 0.05; and beta is to be calculated.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Two Correlated Proportions (McNemar Test) [Odds Ratios]** procedure window by expanding **Proportions**, then **Two Correlated Proportions**, then clicking on **Test (Inequality)**, and then clicking on **Tests for Two Correlated Proportions (McNemar Test) [Odds Ratios]**. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power	*Ignored since this is the Find setting*
Alpha	**0.05**
N (Number of Pairs)	**50 to 200 by 50**
Odds Ratio	**1.5 2.0 2.5**
Proportion Discordant	**0.3**
Alternative Hypothesis	**Two-Sided**
Options Tab	
Use Approximations if N >	**1500**

Annotated Output

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

Numeric Results

Numeric Results for Two-Sided Test

Power	N	P10	P01	Difference (P10-P01)	Proportion Discordant	Odds Ratio	Alpha	Beta
0.07467	50	0.180	0.120	0.060	0.300	1.500	0.05000	0.92533
0.14214	100	0.180	0.120	0.060	0.300	1.500	0.05000	0.85786
0.22451	150	0.180	0.120	0.060	0.300	1.500	0.05000	0.77549
0.29790	200	0.180	0.120	0.060	0.300	1.500	0.05000	0.70210
0.17849	50	0.200	0.100	0.100	0.300	2.000	0.05000	0.82151
0.37305	100	0.200	0.100	0.100	0.300	2.000	0.05000	0.62695
0.56457	150	0.200	0.100	0.100	0.300	2.000	0.05000	0.43543
0.70340	200	0.200	0.100	0.100	0.300	2.000	0.05000	0.29660
0.29601	50	0.214	0.086	0.129	0.300	2.500	0.05000	0.70399
0.59122	100	0.214	0.086	0.129	0.300	2.500	0.05000	0.40878
0.80105	150	0.214	0.086	0.129	0.300	2.500	0.05000	0.19895
0.90751	200	0.214	0.086	0.129	0.300	2.500	0.05000	0.09249

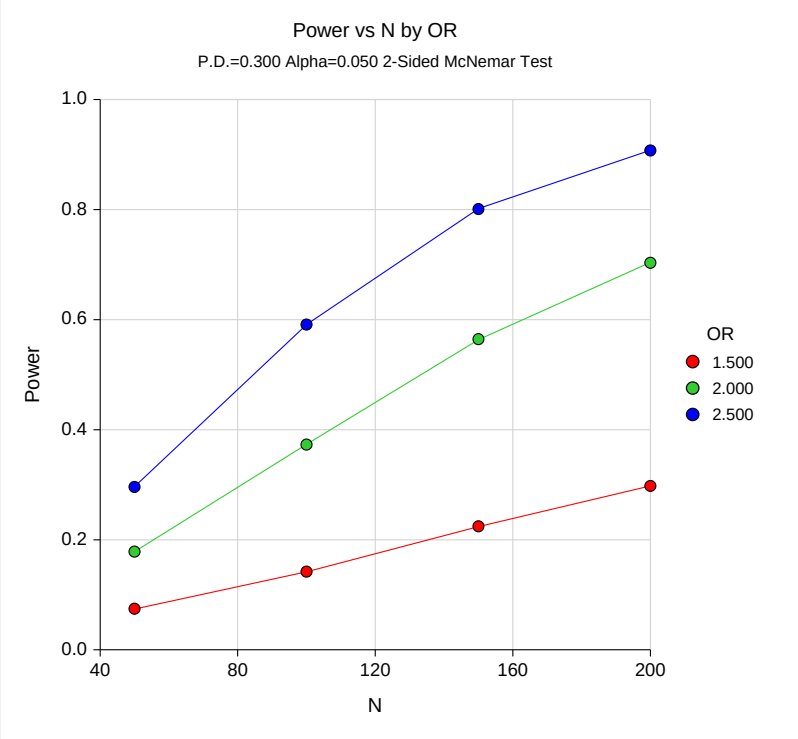
150-8 Inequality Tests for Two Correlated Proportions (McNemar Test)

Report Definitions
Power is the probability of rejecting a false null hypothesis. It should be close to one.
N is the number of pairs in the sample.
P10 is the proportion of pairs in cell 1,2 of the 2x2 table.
P01 is the proportion of pairs in cell 2,1 of the 2x2 table.
Difference is the difference between proportions parameter under the alternative hypothesis.
Proportion Discordant' is the total of P10 and P01.
Odds Ratio is the value of this parameter 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 50 pairs achieves 7% power to detect an odds ratio of 1.500 using a two-sided McNemar test with a significance level of 0.05000. The odds ratio is equivalent to a difference between two paired proportions of 0.060 which occurs when the proportion in cell 1,2 is 0.180 and the proportion in cell 2,1 is 0.120. The proportion of discordant pairs is 0.300.

This report shows the power for each of the scenarios.

Plot Section



This plot shows the power versus the sample size for the three odds ratios.

Example 2 – Validation using Schork and Williams

Schork and Williams (1980) page 354 present a table of sample sizes for various combinations of the other parameters. When the difference is 0.2, the proportion discordant is 0.7, the power is 80%, and the one-sided significance level is 0.025, the sample size is 144.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Two Correlated Proportions (McNemar Test) [Differences]** procedure window by expanding **Proportions**, then **Two Correlated Proportions**, then clicking on **Test (Inequality)**, and then clicking on **Tests for Two Correlated Proportions (McNemar Test) [Differences]**. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power	*Ignored since this is the Find setting*
Alpha	**0.025**
N (Number of Pairs)	**144**
Difference	**0.2**
Proportion Discordant	**0.7**
Alternative Hypothesis	**One-Sided**

Output

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

Numeric Results

Numeric Results for Two-Sided Test

Power	N	P10	P01	Difference (P10-P01)	Proportion Discordant	Odds Ratio	Alpha	Beta
0.80092	144	0.450	0.250	0.200	0.700	1.800	0.02500	0.19908

PASS has also found the power to be about 80%.

150-10 Inequality Tests for Two Correlated Proportions (McNemar Test)

Chapter 155

Inequality Tests for Two Correlated Proportions in a Matched Case-Control Design

Introduction

A 2-by-*M* case-control study investigates a risk factor relevant to the development of a disease. A population of *case* patients with a disease and *control* patients without the disease is considered. Some of these patients have had exposure to a risk factor of interest. A random sample of *N* case patients is selected. Patients are stratified by the levels of a confounding variable (such as age, gender, etc.). For each selected case patient, a random sample of *M* matched control patients is drawn from the same strata (group). An estimate of the odds ratio, *OR*, of developing the disease in exposed and unexposed patients who have equal values of the confounding variable is desired. This odds ratio is assumed to be constant across all levels of the confounding variables.

To fix these ideas, consider the following fictitious data concerning the relationship between smoking and lung cancer. Suppose that a sample of $N = 100$ cases of identical twins in which only one twin has lung cancer is selected. The second twin serves as the control. Each pair of twins is surveyed to determine if either, both, or none smoke tobacco. The results are summarized in the following two-way table:

	No Lung Cancer Twin (Control)	
Lung Cancer Twin (Case)	Smokes = Yes	Smokes = No
Smokes = Yes	16	21
Smokes = No	4	59

Note that the values in this table are counts of pairs, not individuals. The proportion of controls who smoke is (16+4)/100 = 0.20. The proportion of cases who smoke is (16+21)/100 = 0.37. The

odds ratio is 21/4 = 5.25. That is, the twin who smoked is 5.25 times more likely to have lung cancer than the twin who did not.

This procedure is similar to the McNemar procedure also available in ***PASS***. It differs from that procedure in three basic ways:

1. The results are based on the normal approximation to the binomial.
2. This procedure lets you have multiple controls for each case. The McNemar procedure only allows one control per case.
3. The input parameters are different.

Technical Details

The following results summarize the article by Dupont (1988) upon which this module is based. The probabilities that the data fall into various categories are:

1. The probability that a case patient was exposed to the risk factor is p_1.
2. The probability that a control patient was exposed to the risk factor is p_0.
3. The probability that a case patient was not exposed to the risk factor is $q_1 = 1 - p_1$.
4. The probability that a control patient was not exposed to the risk factor is $q_0 = 1 - p_0$.

The odds ratio, *OR*, is defined as

$$OR = \frac{p_1 / q_1}{p_0 / q_0}$$

Assume that you use a χ^2 test for the null hypothesis that *OR* = 1, that is, that $p_0 = p_1$. Such a test is given by Breslow and Day (1980).

Let $x_k = 1 \, or \, 0$ if the k^{th} sampled case patient was or was not exposed, respectively. Let $y_k = 1 \, or \, 0$ if the corresponding first matched control patient was or was not exposed. Let $p_{ij} = \Pr(x_k = i \text{ and } y_k = j)$. For example, p_{10} is the probability that the case patient was exposed to the risk factor while the corresponding first control patient was not. The relationships between these probabilities are

$$p_1 = p_{11} + p_{10}$$

and

$$p_0 = p_{11} + p_{01}$$

Define ϕ to be the correlation between x_k and y_k. It can be shown that

$$\phi = \frac{p_{11}p_{00} - p_{10}p_{01}}{\sqrt{p_1 q_1 p_0 q_0}}$$

A little algebra will show that

$$p_{11} = p_1 p_0 + \phi\sqrt{p_1 q_1 p_0 q_0}$$

$$p_{10} = p_1 q_0 - \phi\sqrt{p_1 q_1 p_0 q_0}$$

$$p_{01} = q_1 p_0 - \phi\sqrt{p_1 q_1 p_0 q_0}$$

$$p_{00} = q_1 q_0 + \phi\sqrt{p_1 q_1 p_0 q_0}$$

$$p_{0+} = \frac{p_{11}}{p_1}$$

$$p_{0-} = \frac{p_{01}}{q_1}$$

$$q_{0+} = 1 - p_{0+}$$

$$q_{0-} = 1 - p_{0-}$$

and

$$t_k = p_1 \binom{M}{k-1} p_{0+}^{k-1} q_{0+}^{M-k+1} + q_1 \binom{M}{k} p_{0-}^{k} q_{0-}^{M-k}; \quad k = 1, \ldots, M$$

Let n_{ij} represent the number of matched sets of subjects in which the case patient was (i = 1) or was not (i = 0) exposed and j of the M control subjects were exposed. Let

$$y = \sum_{m=1}^{M} n_{1,m-1}$$

be the number of discordant sets in which the case patient was exposed and let

$$T_m = n_{1,m-1} + n_{0,m}$$

be the number of sets in which m subjects were exposed. The expected value of T_m is Nt_m.

Let

$$e_{OR} = \sum_{m=1}^{M} \frac{m t_m(OR)}{m(OR) + M - m + 1}$$

$$v_{OR} = \sum_{m=1}^{M} \frac{m t_m(OR)(M - m + 1)}{\left(m(OR) + M - m + 1\right)^2}$$

Dupont (1988) provides the following formula relating α, β, p_0, ϕ, OR, N, and M.

$$1 - \beta = \Phi\left(\frac{\sqrt{N}(e_1 - e_{OR}) - z_{\alpha/2}\sqrt{v_1}}{\sqrt{v_{OR}}}\right) + 1 - \Phi\left(\frac{\sqrt{N}(e_1 - e_{OR}) + z_{\alpha/2}\sqrt{v_1}}{\sqrt{v_{OR}}}\right)$$

This equation may be used to make power and sample size calculations.

Estimating the Sample Control Exposure Probability

To calculate power and sample size, a value for the probability that a sample control patient was exposed to the risk factor (p_0) must be estimated. Remember that the control sample is not a random sample of the population. Instead, it is matched to a random sample of case patients. Hence, the sample does not necessarily provide an unbiased estimate of p_0. Care should be taken to provide an accurate estimate of the probability that a matched control patient was exposed, not the probability that someone was exposed in the general population. However, when there is little association between the confounding (matching) variable and the exposure variable in the control population, the baseline probability of the exposure variable may be used.

Estimating the Correlation, Φ

Previous matched 2-by-2 contingency tables can be used to estimate ϕ using the relationship

$$\phi = \sqrt{\frac{\chi_u^2}{N}}$$

where

$$\chi_u^2 = \frac{N\left(n_{00}n_{11} - n_{01}n_{10}\right)^2}{n_{0\bullet}n_{1\bullet}n_{\bullet 0}n_{\bullet 1}}$$

When no previous data are available about ϕ, Dupont (1988) suggests using a value of 0.2 rather than 0.

Procedure Options

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

Data Tab

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

Solve For

Find (Solve For)

This option specifies the parameter to be solved for from the other parameters. The parameters that may be selected are *Odds Ratio*, *Probability Exposed*, *Correlation*, *N*, *M*, *Alpha*, and *Power and Beta*. Under most situations, you will select either *Power and 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 *Power and Beta* when you want to calculate the power of an experiment that has already been run.

Error Rates

Power or Beta

This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected.

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

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

Alpha (Significance Level)

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected. In this procedure, the test is two-sided.

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.

A single value may be entered here or a range of values such as *0.05 to 0.2 by 0.05* may be entered.

Sample Size

N (Number of Case Patients)

The number of case patients to be sampled. This is the number of pairs. Since there are M controls for each case patient, the total number of patients is $N(M + 1)$.Values of one or greater are allowed here.

M (Number of Controls per Case)

The number of control patients matched with each case patient. When M is one, McNemar's test is used to analyze the data.

The total number of patients in the study is $N(M + 1)$. Values of one or greater are allowed here.

Effect Size

P0 (Probability that a Control is Exposed)

Specify the value of p_0, the probability that a sample control patient is exposed to the risk factor. If the matching variables are independent of the risk factor, you can use the baseline rate of exposure to the risk factor in the general population.

Since this is a probability, it must be between zero and one.

You can enter a list of values separated by commas or blanks.

OR (Odds Ratio)

This option sets the value of the odds ratio, *OR*, which is the ratio of the disease odds of individuals exposed to the risk factor to the disease odds of individuals not exposed to the risk factor.

For example, an odds ratio of 2.0 means that subjects exposed to the risk factor are twice as likely of developing the disease as are unexposed subjects.

A value greater than one is usually used. The value must be greater than zero. The null hypothesis is that the odds ratio is one. You should use a value that will be of interest to others, such as 1.5 or 2.0.

You can enter a list of values.

Phi (Correlation between Case and Control)

This is the correlation for exposure between a case subject and the first of the corresponding control subjects. A value of zero here indicates independence between exposure rates for case and controls. Often, assuming complete independence is unrealistic, so when no other information is available, Dupont (1988) suggests using a value of 0.2.

Correlations can range between -1 and 1. However, only positive correlations should be used.

Example 1 – Calculating Power

This example will show how to calculate the power of a retrospective study for several sample sizes and odds ratios.

Suppose that a matched case-control study is to be run in which the odds ratios of interest are 1.5, 2.5, or 3.5, *P0* = 0.6, correlation = 0.2, *M* = 1, *N* = 25 50 100 150 200, alpha = 0.05, and power is to be found.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Two Correlated Proportions in a Matched Case-Control Design** procedure window by expanding **Proportions**, then **Two Correlated Proportions**, then clicking on **Test (Inequality)**, and then clicking on **Tests for Two Correlated Proportions in a Matched Case-Control Design**. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power	*Ignored since this is the Find setting*
Alpha	**0.05**
N (Case Patients)	**25 50 100 150 200**
M (Controls Per Case)	**1**
P0 (Prob Control Exposed)	**0.6**
OR (Odds Ratio)	**1.5 2.5 3.5**
Phi (Correlation of Case and Control)	**0.2**

Annotated Output

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

Numeric Results

Numeric Results

Power	Cases (N)	Controls Per Case (M)	Odds Ratio (OR)	Probability Exposed (P0)	Correlation (Phi)	Alpha	Beta
0.08863	25	1	1.50	0.60000	0.20000	0.05000	0.91137
0.13364	50	1	1.50	0.60000	0.20000	0.05000	0.86636
0.22622	100	1	1.50	0.60000	0.20000	0.05000	0.77378
0.31832	150	1	1.50	0.60000	0.20000	0.05000	0.68168
0.40652	200	1	1.50	0.60000	0.20000	0.05000	0.59348
0.23067	25	1	2.50	0.60000	0.20000	0.05000	0.76933
0.44278	50	1	2.50	0.60000	0.20000	0.05000	0.55722
0.75646	100	1	2.50	0.60000	0.20000	0.05000	0.24354
0.90966	150	1	2.50	0.60000	0.20000	0.05000	0.09034
0.97004	200	1	2.50	0.60000	0.20000	0.05000	0.02996
0.36379	25	1	3.50	0.60000	0.20000	0.05000	0.63621
0.68570	50	1	3.50	0.60000	0.20000	0.05000	0.31430
0.95159	100	1	3.50	0.60000	0.20000	0.05000	0.04841
0.99482	150	1	3.50	0.60000	0.20000	0.05000	0.00518
0.99956	200	1	3.50	0.60000	0.20000	0.05000	0.00044

Report Definitions

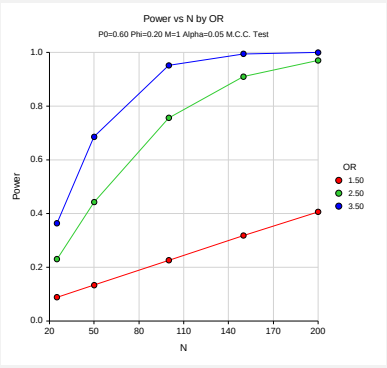
Power is the probability of rejecting a false null hypothesis.
N is the size of the sample drawn from the treatment (case) group.
M is the number of matching control patients drawn for each case patient.
OR is the odds ratio of for subjects exposed to the risk factor.
P0 is the probability of exposure among sampled control patients.
Phi is the correlation of exposure between matched individuals.
Alpha is the probability of rejecting a true null hypothesis.
Beta is the probability of accepting a false null hypothesis.

Summary Statements

In a matched case-control study, the probability of exposure among sampled control patients is 0.60000 and the correlation coefficient for exposure between matched case and control patents is 0.20000. A sample of 25 case patients is obtained. For each case patient, a matching sample of 1 control patient(s) is also obtained. This sample of 50 patients achieves 9% power to detect an odds ratio of 1.50 versus the alternative of equal odds using a Chi-Square test with a 0.05000 significance level.

This report shows the power for each of the scenarios.

Plots Section



This plot shows the power versus the sample size for the three odds ratios.

Example 2 – Calculating Sample Size

Suppose that a matched case-control study is planned to study the relationship between smoking and a certain kind of cancer. Researchers want to have a sample large enough to detect an odds ratio of 2.0. During the power analysis, the researchers also want to calculate the required sample size for odds ratios of 1.5 and 2.5.

The probability that a sampled control (non-cancer) patient smokes is estimated at 0.3. The correlation of smoking between cases and controls is 0.2. The researchers want samples sizes large enough to achieve 80% power at the 0.05 significance levels. In an effort to reduce the number of cancer patients that must be enrolled, the researchers want to try several values for the number of controls per case between 1 and 20.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Two Correlated Proportions in a Matched Case-Control Design** procedure window by expanding **Proportions**, then **Two Correlated Proportions**, then clicking on **Test (Inequality)**, and then clicking on **Tests for Two Correlated Proportions in a Matched Case-Control Design**. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**N (Cases)**
Power	**0.80**
Alpha	**0.05**
N (Case Patients)	*Ignored since this is the Find setting*
M (Controls Per Case)	**1 2 3 4 5 10 20**
P0 (Prob Control Exposed)	**0.3**
OR (Odds Ratio)	**1.5 2.0 2.5**
Phi (Correlation of Case and Control)	**0.2**

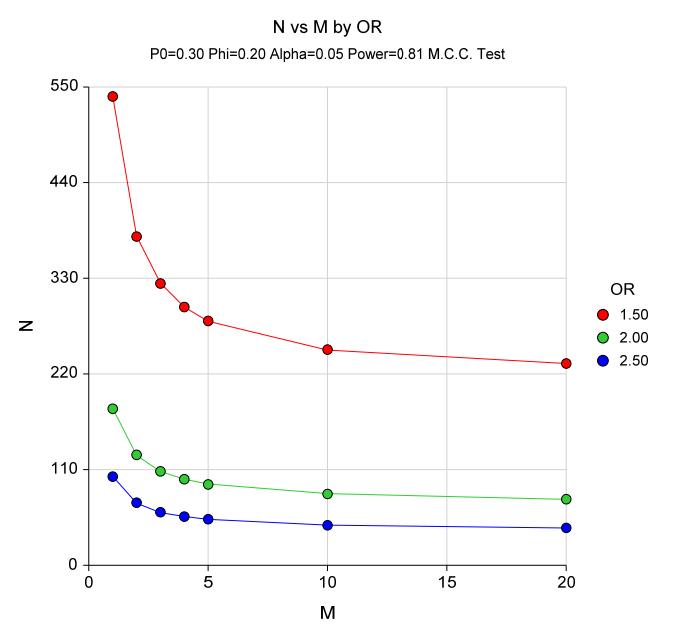
Output

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

Numeric Results and Plots

Numeric Results

Power	Cases (N)	Controls Per Case (M)	Odds Ratio (OR)	Probability Exposed (P0)	Correlation (Phi)	Alpha	Beta
0.80064	539	1	1.50	0.30000	0.20000	0.05000	0.19936
0.80021	378	2	1.50	0.30000	0.20000	0.05000	0.19979
0.80025	324	3	1.50	0.30000	0.20000	0.05000	0.19975
0.80038	297	4	1.50	0.30000	0.20000	0.05000	0.19962
0.80076	281	5	1.50	0.30000	0.20000	0.05000	0.19924
0.80009	248	10	1.50	0.30000	0.20000	0.05000	0.19991
0.80044	232	20	1.50	0.30000	0.20000	0.05000	0.19956
0.80100	180	1	2.00	0.30000	0.20000	0.05000	0.19900
0.80249	127	2	2.00	0.30000	0.20000	0.05000	0.19751
0.80077	108	3	2.00	0.30000	0.20000	0.05000	0.19923
0.80220	99	4	2.00	0.30000	0.20000	0.05000	0.19780
0.80087	93	5	2.00	0.30000	0.20000	0.05000	0.19913
0.80261	82	10	2.00	0.30000	0.20000	0.05000	0.19739
0.80125	76	20	2.00	0.30000	0.20000	0.05000	0.19875
0.80156	102	1	2.50	0.30000	0.20000	0.05000	0.19844
0.80262	72	2	2.50	0.30000	0.20000	0.05000	0.19738
0.80028	61	3	2.50	0.30000	0.20000	0.05000	0.19972
0.80327	56	4	2.50	0.30000	0.20000	0.05000	0.19673
0.80559	53	5	2.50	0.30000	0.20000	0.05000	0.19441
0.80351	46	10	2.50	0.30000	0.20000	0.05000	0.19649
0.80672	43	20	2.50	0.30000	0.20000	0.05000	0.19328



This report shows the sample size for each of the scenarios. Notice that the required number of cancer patients (*N*) drops off drastically as more controls are added. However, using more than five controls seems to only moderately reduce the sample size necessary sample size.

Also notice that the difference in sample size is much larger when moving from an odds ratio of 2.0 to 1.5 than from 2.5 to 2.0.

Example 3 – Validation using Dupont

The formulas used in this module were given in Dupont (1988). He gives an example on page 1164 of the article in which *P0* is 0.6, *Phi* is 0.2, *OR* is 3.0, alpha is 0.05, and beta is 0.2. Dupont finds the sample size for $M = 1$ to be 80 and for $M = 3$ to be 50.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Two Correlated Proportions in a Matched Case-Control Design** procedure window by expanding **Proportions**, then **Two Correlated Proportions**, then clicking on **Test (Inequality)**, and then clicking on **Tests for Two Correlated Proportions in a Matched Case-Control Design**. You may then make the appropriate entries as listed below, or open **Example 3** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**N (Cases)**
Power	**0.80**
Alpha	**0.05**
N (Case Patients)	*Ignored since this is the Find setting*
M (Controls Per Case)	**1 3**
P0 (Prob Control Exposed)	**0.6**
OR (Odds Ratio)	**3.0**
Phi (Correlation of Case and Control)	**0.2**

Output

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

Numeric Results

Numeric Results

Power	Cases (N)	Controls Per Case (M)	Odds Ratio (OR)	Probability Exposed (P0)	Correlation (Phi)	Alpha	Beta
0.80149	80	1	3.00	0.60000	0.20000	0.05000	0.19851
0.80052	50	3	3.00	0.60000	0.20000	0.05000	0.19948

Note that values of 80 and 50 for *N* agree exactly with Dupont.

Chapter 160

Non-Inferiority Tests for Two Correlated Proportions

Introduction

These modules compute power and sample size for non-inferiority tests in which two dichotomous responses are measured on each subject. When one is interested in showing that the true proportions are different, the data are often analyzed with McNemar's test. However, we are interested in showing non-inferiority rather than difference. For example, suppose a diagnostic procedure is accurate, but is expensive to apply or has serious side effects. A replacement procedure is sought which is no less accurate, but is less expensive or has fewer side effects. In this case, we are not interested in showing that the two diagnostic procedures are different, but rather that the second is no worse than the first. *Non-inferiority tests* were designed for this situation.

These tests are often divided into two categories: *equivalence* (two-sided) tests and *non-inferiority* (one-sided) tests. Here, the term *equivalence tests* means that we want to show that two diagnostic procedures are equivalent—that is, their accuracy is about the same. This requires a two-sided hypothesis test. On the other hand, *non-inferiority tests* are used when we want to show that a new (experimental) procedure is no worse than the existing (reference or gold-standard) one. This requires a one-sided hypothesis test. The procedures discussed in this chapter deal with the non-inferiority (one-sided) case.

Technical Details

The results of a study in which two dichotomous responses are measured on each subject can be displayed in a 2-by-2 table in which one response is shown across the columns and the other is shown down the rows. In the discussion to follow, the columns of the table represent the standard (reference or control) response and the rows represent the treatment (experimental) response. The outcome probabilities can be classified into the following table.

Experimental Diagnosis	Standard Diagnosis Yes	No	Total
Yes	p_{11}	p_{10}	P_T
No	p_{01}	p_{00}	$1-P_T$
Total	P_S	$1-P_S$	1

In this table, $p_{ij} = p_{Treatment,\,Standard}$. That is, the first subscript represents the response of the new, experimental procedure while the second subscript represents the response of the standard procedure. Thus, p_{01} represents the proportion having a negative treatment response and a positive standard response.

Sensitivity, Specificity, and Prevalence

To aid in interpretation, analysts have developed a few proportions that summarize the table. Three of the most popular ratios are *sensitivity*, *specificity*, and *prevalence*.

Sensitivity

Sensitivity is the proportion of subjects with a positive standard response who also have a positive experimental response. In terms of proportions from a 2-by-2 table,

$$\text{Sensitivity} = p_{11} / \left(p_{01} + p_{11}\right) = p_{11} / P_S$$

Specificity

Specificity is the proportion of subjects with a negative standard response who also have a negative experimental response. In terms of proportions from a 2-by-2 table,

$$\text{Specificity} = p_{00} / \left(p_{10} + p_{00}\right)$$

Prevalence

Prevalence is the overall proportion of individuals with the disease (or feature of interest). In terms of proportions from a 2-by-2 table,

$$\text{Prevalence} = P_S$$

Table Probabilities

The outcome counts from a sample of *n* subjects can be classified into the following table.

Experimental Diagnosis	Standard Diagnosis Yes	No	Total
Yes	n_{11}	n_{10}	n_T
No	n_{01}	n_{00}	$n-n_T$
Total	n_S	$n-n_S$	n

Note that $n_{11} + n_{00}$ is the number of matches (*concordant pairs*) and $n_{10} + n_{01}$ is the number of *discordant pairs*.

The hypothesis of interest concerns the two marginal probabilities P_T and P_S. P_S represents the accuracy or success of the standard test and P_T represents the accuracy or success of the new, experimental test. Non-inferiority is defined in terms of either the difference, $D = P_T - P_S$, or the relative risk ratio, $R = P_T / P_S$, of these two proportions. The choice between *D* and *R* will usually lead to different sample sizes to achieve the same power.

Non-Inferiority Hypotheses using Differences

This section is based on Liu, Hsueh, Hsieh and Chen (2002). Refer to that paper for complete details.

The null and alternative hypotheses of non-inferiority in terms of the difference are

$$H_0 : P_T - P_S \le -D_E \quad \text{versus} \quad H_1 : P_T - P_S > -D_E$$

To demonstrate non-inferiority, one desires to reject the null hypothesis and thus conclude that the experimental treatment is not worse than the standard by as much or more than D_E. In the context of the preceding statement, D_E is defined to be positive. The choice of an appropriate D_E may be difficult. It should be clinically meaningful so that clinicians are willing to conclude that the experimental treatment is acceptable if the difference is less than D_E. From a statistical perspective, D_E should be less than the effect, if known, of the standard treatment compared to placebo.

Liu et al. (2002) discuss the RMLE-based (score) method for constructing these confidence intervals. This method is based on (developed by, described by) Nam (1997).

Asymptotic Tests

An asymptotic test is given by

$$Z_L = \frac{\hat{D} + D_E}{\hat{\sigma}} = \frac{c + nD_E}{\sqrt{d - n\hat{D}^2}} \ge z_\alpha$$

where

$$\hat{D} = \frac{n_T}{n} - \frac{n_S}{n} = \frac{n_{10}}{n} - \frac{n_{01}}{n}$$

$$d = n_{10} + n_{01}$$

$$c = n_{10} - n_{01}$$

and z_α is the standard normal deviate having α in the right tail.

An estimate of $\hat{\sigma}$ based on the RMLE-based (score) procedure of Nam (1997) uses the estimates

$$\tilde{\sigma}_L = \sqrt{\frac{\tilde{p}_{L,10} + \tilde{p}_{L,01} - D_E^2}{n}}$$

where

$$\tilde{p}_{L,10} = \frac{-\tilde{a}_L + \sqrt{\tilde{a}_L^2 - 8\tilde{b}_L}}{4}$$

$$\tilde{p}_{L,01} = \tilde{p}_{L,10} - D_E$$

$$\tilde{a}_{L,01} = -\hat{D}(1 - D_E) - 2(\hat{p}_{01} + D_E)$$

$$\tilde{b}_{L,01} = D_E(1 + D_E)\hat{p}_{01}$$

Power Formula

The power when the actual difference is D_A can be evaluated exactly using the multinomial distribution. However, when the sample size is above a user-set level, we use a normal approximation to this distribution which leads to

$$1 - \beta(D_A) = \begin{Bmatrix} 1 - \Phi(c_L) & \text{if } D_A > -D_E \\ 0 & \text{otherwise} \end{Bmatrix}$$

where

$$c_L = \frac{-D_A}{\sigma} - \frac{D_E}{\sigma} + \frac{z_\alpha}{w_L}$$

$$\sigma = \sqrt{\frac{p_{01} + p_{10} - D_A}{n}}$$

$$w_L = \sqrt{\frac{2p_{01} + D_A - D_A^2}{2\bar{p}_{L,01} - D_E - D_E^2}}$$

$$\bar{p}_{L,01} = \frac{-a_L + \sqrt{a_L^2 - 8b_L}}{4}$$

$$a_L = -D_A(1 - D_E) - 2(p_{01} + D_E)$$

$$b_L = D_E(1 + D_E)p_{01}$$

Non-Inferiority Hypotheses using Ratios

The following is based on Nam and Blackwelder (2002). We refer you to this paper for the complete details of which we will only provide a brief summary here.

When $R_E < 1$, the statistical hypotheses of non-equivalence are

$$H_0: P_T / P_S \le R_E \text{ versus } H_1: P_T / P_S > R_E$$

Test Statistics

The test statistic for an asymptotic test based on constrained maximum likelihood for large *n* is given by

$$Z(R_E) = \sqrt{\frac{n\left(\hat{P}_\mathrm{T} - R_E\hat{P}_\mathrm{S}\right)}{R_E\left(\tilde{p}_{10} + \tilde{p}_{01}\right)}}$$

where

$$\tilde{p}_{10} = \frac{-\hat{P}_\mathrm{T} + R_E^2\left(\hat{P}_\mathrm{S} + 2\hat{p}_{10}\right) + \sqrt{\left(\hat{P}_\mathrm{T} - R_E^2\hat{P}_\mathrm{S}\right)^2 + 4R_E^2\hat{p}_{10}\hat{p}_{01}}}{2R_E\left(R_E + 1\right)}$$

$$\tilde{p}_{01} = R_E\tilde{p}_{10} - \left(R_E - 1\right)\left(1 - \hat{p}_{00}\right)$$

$$\hat{p}_{01} = \frac{n_{01}}{n},\ \hat{p}_{10} = \frac{n_{10}}{n},\ \hat{P}_\mathrm{T} = \frac{n_{10} + n_{11}}{n},\ \hat{P}_\mathrm{S} = \frac{n_{01} + n_{11}}{n}$$

Power Formula

The power when the true value of the relative risk ratio is R_E can be evaluated exactly using the multinomial distribution. When *n* is large, we use a normal approximation to the multinomial distribution which leads to

$$\beta(R_A) = \Phi(c_U)$$

where

$$c_U = \frac{z_{1-\alpha}\sqrt{\overline{V}_0(T_0)} - E_1(T_0)}{\sqrt{V_1(T_0)}}$$

$$\overline{V}_0(T_0) = \frac{R_E\left(\overline{p}_{10} + \overline{p}_{01}\right)}{n}$$

$$E_1(T_0) = \left(R_A - R_E\right)P_\mathrm{S}$$

$$V_1(T_0) = \frac{\left(R_A + R_E^2\right)P_\mathrm{S} - 2R_E p_{11} - \left(R_A - R_E\right)^2 P_\mathrm{S}^2}{n}$$

$$\overline{p}_{10} = \frac{-P_\mathrm{T} + R_E^2\left(P_\mathrm{S} + 2p_{10}\right) + \sqrt{\left(P_\mathrm{T} - R_E^2 P_\mathrm{S}\right)^2 + 4R_E^2 p_{10}p_{01}}}{2R_E\left(R_E + 1\right)}$$

$$\overline{p}_{01} = R_E\overline{p}_{10} - \left(R_E - 1\right)\left(1 - p_{00}\right)$$

Nuisance Parameter

Unfortunately, the 2-by-2 table includes four parameters p_{11}, p_{10}, p_{01}, and p_{00}, but the power specifications above only specify two parameters: P_S and D_A or R_A. A third parameter is

defined implicitly since the sum of the four parameters is one. One parameter, known as a nuisance parameter, remains unaccounted for. This parameter must be addressed to fully specify the problem. This fourth parameter can be specified by specifying any one of the following: p_{11}, p_{10}, p_{01}, p_{00}, $p_{10} + p_{01}$, $p_{11} + p_{00}$, or the sensitivity of the experimental response, p_{11} / P_S.

It may be difficult to specify a reasonable value for the nuisance parameter since its value may not be even approximately known until after the study is conducted. Because of this, we suggest that you calculate power or sample size for a range of values of the nuisance parameter. This will allow you to determine how sensitive the results are to its value.

Procedure Options

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

Data Tab (Common Options)

The Data tab contains the parameters associated with this test such as the proportions, sample sizes, alpha, and power. This chapter covers two procedures which have different options. This section documents options that are common to both procedures. Later, unique options for each procedure will be documented.

Solve For

Find (Solve For)

This option specifies the parameter to be solved for from the other parameters. The parameters that may be selected are *Power and Beta* or *N*.

Error Rates

Power or Beta

This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected. Here, a type-II error occurs when you fail to conclude non-inferiority when in fact it is true.

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

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

Alpha (Significance Level)

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected. Here, a type-I error occurs when you falsely conclude non-inferiority.

Sample Size

N (Sample Size)

Enter a value for the sample size. This value must be greater than two. You may enter a range of values such as *10 to100 by 10*.

Effect Size – Standard Proportion

Ps (Standard Proportion)

This is the proportion of yes's (or successes), P_S, when subjects received the standard treatment. This value or a good estimate is often available from previous studies.

You may enter a set of values separated by blank spaces. For example, you could enter '0.50 0.60 0.70'. Values between, but not including, 0 and 1 are permitted.

Effect Size – Nuisance Parameter

Nuisance Parameter Type

Enter the type of nuisance parameter here. Unfortunately, the 2-by-2 table cannot be completely specified by using only the parameters Ps and Da or Ps and Ra. One other parameter must be specified. This additional parameter is called a 'nuisance' parameter. It will be assumed to be a known quantity. Several possible choices are available. This option lets you specify which parameter you want to use. In all cases, the value you specify is a proportion.

- **P11**
  The proportion of subjects that are positive on both tests.
- **P00**
  The proportion of subjects that are negative on both tests.
- **P01**
  The proportion of subjects that are negative on the treatment, but positive on the standard.
- **P10**
  The proportion of subjects that are positive on the treatment, but negative on the standard.
- **P11+P00**
  The proportion of matches (concordant pairs).
- **P01+P10**
  The proportion of non-matches (discordant pairs).

- **P11/Ps**
  The sensitivity.

Nuisance Parameter Value

Enter the value of the nuisance parameter that you specified in the 'Nuisance Parameter Type' box. This value is a proportion, so it must be between 0 and 1.

Data Tab (Differences)

This section documents options that are used when the parameterization is in terms of the difference, P1 – P2. P1.0 is the value of P1 assumed by the null hypothesis and P1.1 is the value of P1 at which the power is calculated. Once P2, D0, and D1 are given, the values of P1.1 and P1.0 can be calculated.

Effect Size – Differences

|De| (Equivalence Difference)

De is the maximum allowable difference between the standard and treatment proportions that will still result in the conclusion of equivalence. In order to ensure that De is positive, the difference is computed in reverse order. That is, $D_E = P_S - P_T$. This parameter is only used when the Test Statistic option is set to 'Difference'.

Only positive values can be entered here. Typical values for this difference are 0.05, 0.10, and 0.20. For two-sided tests, you must have |Da| < De. For one-sided tests, you must have Da > -De.

Da (Actual Difference)

Da is the actual difference between the treatment and standard proportions $D_A = P_T - P_S$. Da may be positive, negative, or (usually) zero. This parameter is only used when the Test Statistic option is set to 'Difference'.

For two-sided tests, you must have |Da| < De. For one-sided tests, you must have Da > -De.

Data Tab (Ratios)

This section documents options that are used when the parameterization is in terms of the ratio, P1 / P2. P1.0 is the value of P1 assumed by the null hypothesis and P1.1 is the value of P1 at which the power is calculated. Once P2, R0, and R1 are given, the values of P1.0 and P1.1 can be calculated.

Effect Size – Ratios

Re (Equivalence Ratio)

Re is the minimum size of the relative risk ratio, P_T / P_S, that will still result in the conclusion of equivalence. Both equivalence and non-inferiority trials use a value that is less than one. Typical values for this ratio are 0.8 or 0.9.

This parameter is only used when the Test Statistic option is set to 'Ratio'.

Ra (Actual Ratio)

Enter a value for Ra, the actual relative risk ratio P_T / P_S. This value is used to generate the value of P_T using the formula $P_T = P_S R_a$. Often this value is set equal to one, but this is not necessary.

Options Tab

This tab sets a couple of options used in the iterative procedures.

Maximum Iterations

Maximum Iterations Before Search Termination

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.

Approximations

Use Approximations if N is greater than

Specify the maximum value of N (sample size) for which you would like an exact power calculation based on the multinomial distribution. Sample sizes greater than this value will use the asymptotic approximation given in the documentation. The exact calculation of the multinomial distribution becomes very time consuming for N > 200. For most cases, when N > 200, the difference between the exact and approximate calculations is small. For N > 200, the length of time needed to calculate the exact answer may become prohibitive. However, as the speed of computers increases, it will become faster and easier to calculate the exact power for larger values of N.

If you want all calculations to use exact results, enter ‘1000’ here.

If you want all calculations to use the quick approximations, enter ‘1’ here.

Example 1 – Finding Power

A clinical trial will be conducted to show that a non-invasive MRI test is not inferior to the invasive CTAP reference test. Historical data suggest that the CTAP test is 80% accurate. After careful discussion, the researchers decide that if the MRI test is 75% accurate or better, it will be considered non-inferior. They decide to use a difference test statistic. Thus, the equivalence difference is 0.05. They want to study the power for various sample sizes between 20 and 1000 at the 5% significance level.

They use P01 as the nuisance parameter and look at two values: 0.05 and 0.10.

Setup

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

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power	*Ignored since this is the Find setting*
Alpha	**0.05**
N (Sample Size)	**20 100 200 300 450 600 800 1000**
|De| (Equivalence Difference)	**0.05**
Da (Actual Difference)	**0.00**
Ps (Standard Proportion)	**0.80**
Nuisance Parameter Type	**P01**
Nuisance Parameter Value	**0.05 0.10**
Options Tab	
Use Approximations if N is >	**100**

Annotated Output

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

Numeric Results

Numeric Results for a Non-Inferiority (One-Sided) Test of a Difference

Power	Sample Size (N)	Equiv. Difference (De)	Actual Difference (Da)	Treatment Proportion (Pt)	Standard Proportion (Ps)	Nuisance Parameter (P01)	Alpha	Beta
0.22938	20	0.05000	0.00000	0.80000	0.80000	0.05000	0.05000	0.77062
0.13717	20	0.05000	0.00000	0.80000	0.80000	0.10000	0.05000	0.86283
0.43625	100	0.05000	0.00000	0.80000	0.80000	0.05000	0.05000	0.56375
0.28895	100	0.05000	0.00000	0.80000	0.80000	0.10000	0.05000	0.71105
0.67771	200	0.05000	0.00000	0.80000	0.80000	0.05000	0.05000	0.32229
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.

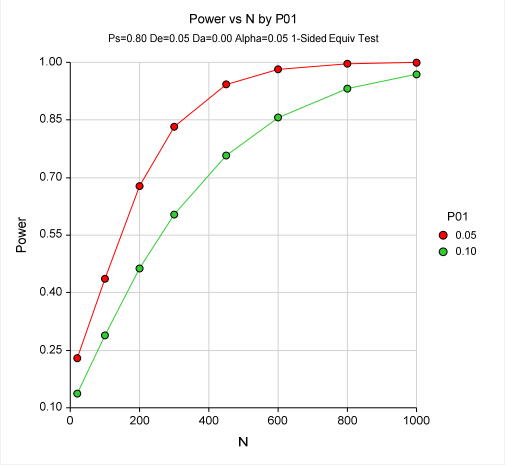
(report continues)

Report Definitions
Power is the probability of rejecting a false null hypothesis.
N is the number of subjects, the sample size.
De is the maximum difference between the two proportions that is still called 'equivalent'.
Da is the actual difference between Pt and Ps. That is, Da = Pt-Ps.
Pt is the response proportion to the treatment (experimental or new) test.
Ps is the response proportion to the standard (reference or old) test.
The Nuisance Parameter is a value that is needed, but is not a direct part of the hypothesis.
Alpha is the probability of rejecting a true null hypothesis.
Beta is the probability of accepting a false null hypothesis.

Summary Statements
A sample size of 20 subjects achieves 23% power at a 5% significance level using a one-sided equivalence test of correlated proportions when the standard proportion is 0.80000, the maximum allowable difference between these proportions that still results in equivalence (the range of equivalence) is 0.05000, and the actual difference of the proportions is 0.00000.

This report shows the power for the indicated scenarios. All of the columns are defined in the ‘Report Definitions’ section.

Plots Section



This plot shows the power versus the sample size for the two values of P01. In this example, we see that the value of the nuisance parameter has a large effect on the calculated sample size.

Example 2 – Finding Sample Size

Continuing with Example1, the analysts want to determine the exact sample size necessary to achieve 90% power for both values of the nuisance parameter.

Setup

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

Option	Value
Data Tab	
Find (Solve For)	**Sample Size (N)**
Power	**0.90**
Alpha	**0.05**
N (Sample Size)	*Ignored since this is the Find setting*
|De| (Equivalence Difference)	**0.05**
Da (Actual Difference)	**0.00**
Ps (Standard Proportion)	**0.80**
Nuisance Parameter Type	**P01**
Nuisance Parameter Value	**0.05 0.10**
Options Tab	
Use Approximations if N is >	**100**

Output

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

Numeric Results

Numeric Results for a Non-Inferiority (One-Sided) Test of a Difference

Power	Sample Size (N)	Equiv. Difference (De)	Actual Difference (Da)	Treatment Proportion (Pt)	Standard Proportion (Ps)	Nuisance Parameter (P01)	Alpha	Beta
0.90026	374	0.05000	0.00000	0.80000	0.80000	0.05000	0.05000	0.09974
0.90014	699	0.05000	0.00000	0.80000	0.80000	0.10000	0.05000	0.09986

This report shows that the sample size required nearly doubles when P01 is changed from 0.05 to 0.10.

Example 3 – Validation using Liu

Liu *et al*. (2002) give an example in which P01 is 0.05, P10 is 0.05, Da is 0.00, De is 0.05, the significance level is 0.025, and the power is 80%. From their Table III, the sample size is 350.

In this example, the value of Ps is arbitrary. We set it at 0.50.

Setup

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

Option	Value
Data Tab	
Find (Solve For)	**Sample Size (N)**
Power	**0.80**
Alpha	**0.025**
N (Sample Size)	*Ignored since this is the Find setting*
|De| (Equivalence Difference)	**0.05**
Da (Actual Difference)	**0.00**
Ps (Standard Proportion)	**0.50**
Nuisance Parameter Type	**P01**
Nuisance Parameter Value	**0.05**
Options Tab	
Use Approximations if N is >	**100**

Output

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

Numeric Results

Numeric Results for a Non-Inferiority (One-Sided) Test of a Difference

Power	Sample Size (N)	Equiv. Difference (De)	Actual Difference (Da)	Treatment Proportion (Pt)	Standard Proportion (Ps)	Nuisance Parameter (P01)	Alpha	Beta
0.80046	350	0.05000	0.00000	0.50000	0.50000	0.05000	0.02500	0.19954

The calculated sample size of 350 matches the results of Liu et al. (2002).

Example 4 – Validation using Nam and Blackwelder

Nam and Blackwelder (2002) give an example in which Ps is 0.80, P10 is 0.05, Ra is 1.00, Re is 0.80, the significance level is 0.05, and the power is 80%. From their Table III, the sample size is 34.

Note that their calculations use the approximate formula, so we will set the value of 'Use Approximations if N is greater than' to '1' so that only the approximate formula is used.

Setup

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

Option	Value
Data Tab	
Find (Solve For)	**Sample Size (N)**
Power	**0.80**
Alpha	**0.05**
N (Sample Size)	*Ignored since this is the Find setting*
Re (Equivalence Ratio)	**0.80**
Ra (Actual Ratio)	**1.00**
Ps (Standard Proportion)	**0.80**
Nuisance Parameter Type	**P10**
Nuisance Parameter Value	**0.05**
Options Tab	
Use Approximations if N is >	**1**

Output

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

Numeric Results

Numeric Results for a Non-Inferiority (One-Sided) Test of a Ratio

Power	Sample Size (N)	Equiv. Ratio (Re)	Actual Ratio (Ra)	Treatment Proportion (Pt)	Standard Proportion (Ps)	Nuisance Parameter (P10)	Alpha	Beta
0.80050	34	0.80	1.00	0.80000	0.80000	0.05000	0.05000	0.19950

The calculated sample size of 34 matches the results of Nam and Blackwelder (2002).

Example 5 – Finding Sample Size for a Non-Inferiority Test

Researchers have developed a new treatment for migraine headaches which is less expensive than a current standard. The researchers need to show that the proportion of individuals who respond to the new treatment is not inferior to the standard treatment. They want to determine the minimum number of subjects required to achieve 90% power for the test of non-inferiority. The new treatment will be considered non-inferior if its success rate is no less than 90% of the success rate of the standard, which is about 0.65. The sample size required is evaluated for various values (0.3 to 0.9) of the nuisance parameter: P11/Ps = sensitivity.

Setup

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

Option	Value
Data Tab	
Find	**Sample Size (N)**
Power	**0.90**
Alpha	**0.05**
N (Sample Size)	*Ignored since this is the Find setting*
Re (Equivalence Ratio)	**0.95**
Ra (Actual Ratio)	**1.0**
Ps (Standard Proportion)	**0.65**
Nuisance Parameter Type	**P11/Ps (Sensitivity)**
Nuisance Parameter Value	**0.3 to 0.9 by 0.1**
Options Tab	
Use Approximations if N is >	**1**

Output

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

Numeric Results for a Non-Inferiority (One-Sided) Test of a Ratio

Power	Sample Size (N)	Equiv. Ratio (Re)	Actual Ratio (Ra)	Treatment Proportion (Pt)	Standard Proportion (Ps)	Nuisance Parameter (P11/Ps)	Alpha	Beta
	13	0.95	1.00	0.65000	0.65000	0.30000	0.05000	
	13	0.95	1.00	0.65000	0.65000	0.40000	0.05000	
0.90004	5013	0.95	1.00	0.65000	0.65000	0.50000	0.05000	0.09996
0.90001	4013	0.95	1.00	0.65000	0.65000	0.60000	0.05000	0.09999
0.90004	3015	0.95	1.00	0.65000	0.65000	0.70000	0.05000	0.09996
0.90011	2020	0.95	1.00	0.65000	0.65000	0.80000	0.05000	0.09989
0.90016	1035	0.95	1.00	0.65000	0.65000	0.90000	0.05000	0.09984

These scenarios require a large sample size. In fact, the first two rows are blank because the sample size is so large it can’t be determined.

Chapter 165

Equivalence Tests for Two Correlated Proportions

Introduction

The two procedures described in this chapter compute power and sample size for testing equivalence using differences or ratios in designs in which two dichotomous responses are measured on each subject. Each of these options is listed separately on the menus.

When one is interested in showing that two correlated proportions are different, the data are often analyzed with McNemar's test. However, the procedures discussed here are interested in showing equivalence rather than difference. For example, suppose a diagnostic procedure is accurate, but is expensive to apply or has serious side effects. A replacement procedure may be sought which is equally accurate, but is less expensive or has fewer side effects. In this case, we are not interested in showing that the two diagnostic procedures are different, but rather that they are the same. Equivalence tests were designed for this situation.

These tests are often divided into two categories: *equivalence* (two-sided) tests and *non-inferiority* (one-sided) tests. Here, the term *equivalence tests* means that we want to show that two diagnostic procedures are equivalent—that is, their accuracy is about the same. This requires a two-sided hypothesis test. On the other hand, *non-inferiority tests* are used when we want to show that a new (experimental) procedure is no worse than the existing (reference or gold-standard) one. This requires a one-sided hypothesis test.

Technical Details

The results of a study in which two dichotomous responses are measured on each subject can be displayed in a 2-by-2 table in which one response is shown across the columns and the other is shown down the rows. In the discussion to follow, the columns of the table represent the standard (reference or control) response and the rows represent the treatment (experimental) response. The outcome probabilities can be classified into the following table.

Experimental Diagnosis	Standard Diagnosis Yes	No	Total
Yes	p_{11}	p_{10}	P_T
No	p_{01}	p_{00}	$1 - P_T$
Total	P_S	$1 - P_S$	1

In this table, $p_{ij} = p_{Treatment,\,Standard}$. That is, the first subscript represents the response of the new, experimental procedure while the second subscript represents the response of the standard procedure. Thus, p_{01} represents the proportion having a negative treatment response and a positive standard response.

Sensitivity, Specificity, and Prevalence

To aid in interpretation, analysts have developed a few proportions that summarize the table. Three of the most popular ratios are *sensitivity*, *specificity*, and *prevalence*.

Sensitivity

Sensitivity is the proportion of subjects with a positive standard response who also have a positive experimental response. In terms of proportions from the 2-by-2 table,

$$\text{Sensitivity} = p_{11} / \left(p_{01} + p_{11}\right) = p_{11} / P_S$$

Specificity

Specificity is the proportion of subjects with a negative standard response who also have a negative experimental response. In terms of proportions from the 2-by-2 table,

$$\text{Specificity} = p_{00} / \left(p_{10} + p_{00}\right)$$

Prevalence

Prevalence is the overall proportion of individuals with the disease (or feature of interest). In terms of proportions from the 2-by-2 table,

$$\text{Prevalence} = P_S$$

Table Probabilities

The outcome counts from a sample of *n* subjects can be classified into the following table.

Experimental Diagnosis	Standard Diagnosis Yes	No	Total
Yes	n_{11}	n_{10}	n_T
No	n_{01}	n_{00}	$n - n_T$
Total	n_S	$n - n_S$	n

Note that $n_{11} + n_{00}$ is the number of matches (*concordant pairs*) and $n_{10} + n_{01}$ is the number of *discordant pairs*.

The hypothesis of interest concerns the two marginal probabilities P_T and P_S. P_S represents the accuracy or success of the standard test and P_T represents the accuracy or success of the new, experimental test. Equivalence is defined in terms of either the difference, $D = P_T - P_S$, or the relative risk ratio, $R = P_T / P_S$, of these two proportions. The choice between *D* and *R* will usually lead to different sample sizes to achieve the same power.

Equivalence Hypotheses using Differences

This section is based on Liu, Hsueh, Hsieh and Chen (2002). We refer you to that paper for complete details.

The hypotheses of equivalence in terms of the difference are

$$H_0: P_T - P_S \geq D_E \quad \text{or} \quad H_0: P_T - P_S \leq -D_E \quad \text{versus} \quad H_1: -D_E < P_T - P_S < D_E$$

These hypotheses can be decomposed into two sets of one-sided hypotheses

$$H_{0L}: P_T - P_S \leq -D_E \quad \text{versus} \quad H_{1L}: P_T - P_S > -D_E$$

and

$$H_{0U}: P_T - P_S \geq D_E \quad \text{versus} \quad H_{1U}: P_T - P_S < D_E$$

The hypothesis test of equivalence with type I error rate α is conducted by computing a $100(1-2\alpha)\%$ confidence interval for $P_T - P_S$ and determining if this interval is wholly contained between $-D_E$ and D_E. This confidence interval approach is often recommended by regulatory agencies.

Liu et al. (2002) discuss the RMLE-based (score) method for constructing these confidence intervals. This method is based on (developed by, described by) Nam (1997).

Asymptotic Tests

An asymptotic test for testing H_{0L} versus H_{1L} is given by

$$Z_L = \frac{\hat{D} + D_E}{\hat{\sigma}} = \frac{c + nD_E}{\sqrt{d - n\hat{D}^2}} \geq z_\alpha$$

where

$$\hat{D} = \frac{n_T}{n} - \frac{n_S}{n} = \frac{n_{10}}{n} - \frac{n_{01}}{n}$$

$$d = n_{10} + n_{01}$$

$$c = n_{10} - n_{01}$$

and z_α is the standard normal deviate having α in the right tail.

Similarly, an asymptotic test for testing H_{0U} versus H_{1U} is given by

$$Z_U = \frac{\hat{D} - D_E}{\hat{\sigma}} = \frac{c - nD_E}{\sqrt{d - n\hat{D}^2}} \leq -z_\alpha .$$

Equivalence is concluded if both the tests on Z_L and Z_U are rejected.

An estimate of $\hat{\sigma}$ based on the RMLE-based (score) procedure of Nam (1997) uses the estimates

$$\tilde{\sigma}_L = \sqrt{\frac{\tilde{p}_{L,10} + \tilde{p}_{L,01} - D_E^2}{n}}$$

and

$$\tilde{\sigma}_U = \sqrt{\frac{\tilde{p}_{U,10} + \tilde{p}_{U,01} - D_E^2}{n}}$$

where

$$\tilde{p}_{L,10} = \frac{-\tilde{a}_L + \sqrt{\tilde{a}_L^2 - 8\tilde{b}_L}}{4}$$

$$\tilde{p}_{L,01} = \tilde{p}_{L,10} - D_E$$

$$\tilde{p}_{U,10} = \frac{-\tilde{a}_U + \sqrt{\tilde{a}_U^2 - 8\tilde{b}_U}}{4}$$

$$\tilde{p}_{U,01} = \tilde{p}_{U,10} + D_E$$

$$\tilde{a}_{L,01} = -\hat{D}(1 - D_E) - 2(\hat{p}_{01} + D_E)$$

$$\tilde{b}_{L,01} = D_E(1 + D_E)\hat{p}_{01}$$

$$\tilde{a}_{U,01} = -\hat{D}(1 + D_E) - 2(\hat{p}_{01} - D_E)$$

$$\tilde{b}_{U,01} = -D_E(1 - D_E)\hat{p}_{01}$$

Note that the ICH E9 guideline (see Lewis (1999)) suggests using a significance level of $\alpha / 2$ when testing this hypothesis.

Power Formula

The power when the actual difference is D_A can be evaluated exactly using the multinomial distribution. However, when the sample size is above a user-set level, we use a normal approximation to this distribution which leads to

$$1 - \beta(D_A) = \begin{Bmatrix} \Phi(c_U) - \Phi(c_L) & \text{if } c_U - c_L > 0 \\ 0 & \text{otherwise} \end{Bmatrix}$$

where

$$c_U = \frac{-D_A}{\sigma} + \frac{D_E}{\sigma} - \frac{z_\alpha}{w_U}$$

$$c_L = \frac{-D_A}{\sigma} - \frac{D_E}{\sigma} + \frac{z_\alpha}{w_L}$$

$$\sigma = \sqrt{\frac{p_{01} + p_{10} - D_A}{n}}$$

$$w_L = \sqrt{\frac{2p_{01} + D_A - D_A^2}{2\bar{p}_{L,01} - D_E - D_E^2}}$$

$$w_U = \sqrt{\frac{2p_{01} + D_A - D_A^2}{2\bar{p}_{U,01} + D_E - D_E^2}}$$

$$\bar{p}_{L,01} = \frac{-a_L + \sqrt{a_L^2 - 8b_L}}{4}$$

$$\bar{p}_{U,01} = \frac{-a_U + \sqrt{a_U^2 - 8b_U}}{4}$$

$$a_L = -D_A(1 - D_E) - 2(p_{01} + D_E)$$

$$b_L = D_E(1 + D_E)p_{01}$$

$$a_U = -D_A(1 + D_E) - 2(p_{01} - D_E)$$

$$b_U = -D_E(1 - D_E)p_{01}$$

Equivalence Hypotheses using Ratios

For the two-sided (equivalence) case when $R_E < 1$, the statistical hypotheses are

$$H_0: P_T / P_S \le 1/R_E \text{ or } P_T / P_S \ge R_E \quad \text{versus} \quad H_1: R_E < P_T / P_S < 1/R_E$$

These can be decomposed into two sets of one-sided hypotheses

$$H_{0L}: P_T / P_S \le R_E \quad \text{versus} \quad H_{1L}: P_T / P_S > R_E$$

and

$$H_{0U}: P_T / P_S \ge 1/R_E \quad \text{versus} \quad H_{1U}: P_T / P_S < 1/R_E$$

Note that the first set of one-sided hypotheses, H_{0L} versus H_{1L}, is referred to as the hypotheses of non-inferiority.

The following is based on Nam and Blackwelder (2002). We refer you to this paper for the complete details of which we will only provide a brief summary here.

Test Statistics

The test statistic for an asymptotic test based on constrained maximum likelihood for large *n* is given by

$$Z(R_E) = \sqrt{\frac{n\left(\hat{P}_\mathrm{T} - R_E\hat{P}_\mathrm{S}\right)}{R_E\left(\tilde{p}_{10} + \tilde{p}_{01}\right)}}$$

where

$$\tilde{p}_{10} = \frac{-\hat{P}_\mathrm{T} + R_E^2\left(\hat{P}_\mathrm{S} + 2\hat{p}_{10}\right) + \sqrt{\left(\hat{P}_\mathrm{T} - R_E^2\hat{P}_\mathrm{S}\right)^2 + 4R_E^2\hat{p}_{10}\hat{p}_{01}}}{2R_E\left(R_E + 1\right)}$$

$$\tilde{p}_{01} = R_E\tilde{p}_{10} - \left(R_E - 1\right)\left(1 - \hat{p}_{00}\right)$$

$$\hat{p}_{01} = \frac{n_{01}}{n},\ \hat{p}_{10} = \frac{n_{10}}{n},\ \hat{P}_\mathrm{T} = \frac{n_{10} + n_{11}}{n},\ \hat{P}_\mathrm{S} = \frac{n_{01} + n_{11}}{n}$$

Note that the above applies to a one-sided test. When using a two-sided test, we calculate both $Z(R_E)$ and $Z(1/R_E)$ using the above formula.

Power Formula

The power of the one-sided procedure when the true value of the relative risk ratio is R_E can be evaluated exactly using the multinomial distribution. When *n* is large, we use a normal approximation to the multinomial distribution which leads to

$$\beta(R_A) = \Phi(c_U)$$

where

$$c_U = \frac{z_{1-\alpha}\sqrt{\overline{V}_0(T_0)} - E_1(T_0)}{\sqrt{V_1(T_0)}}$$

$$\overline{V}_0(T_0) = \frac{R_E\left(\overline{p}_{10} + \overline{p}_{01}\right)}{n}$$

$$E_1(T_0) = \left(R_A - R_E\right)P_\mathrm{S}$$

$$V_1(T_0) = \frac{\left(R_A + R_E^2\right)P_\mathrm{S} - 2R_E p_{11} - \left(R_A - R_E\right)^2 P_\mathrm{S}^2}{n}$$

$$\overline{p}_{10} = \frac{-P_\mathrm{T} + R_E^2\left(P_\mathrm{S} + 2p_{10}\right) + \sqrt{\left(P_\mathrm{T} - R_E^2P_\mathrm{S}\right)^2 + 4R_E^2p_{10}p_{01}}}{2R_E\left(R_E + 1\right)}$$

$$\overline{p}_{01} = R_E\overline{p}_{10} - \left(R_E - 1\right)\left(1 - p_{00}\right)$$

Nuisance Parameter

Unfortunately, the 2-by-2 table includes four parameters p_{11}, p_{10}, p_{01}, and p_{00}, but the power specifications above only specify two parameters: P_S and D_A or R_A. A third parameter is defined implicitly since the sum of the four parameters is one. One parameter, known as a nuisance parameter, remains unaccounted for. This parameter must be addressed to fully specify the problem. This fourth parameter can be specified by specifying any one of the following: p_{11}, p_{10}, p_{01}, p_{00}, $p_{10} + p_{01}$, $p_{11} + p_{00}$, or the sensitivity of the experimental response, p_{11} / P_S.

It may be difficult to specify a reasonable value for the nuisance parameter since its value may not be even approximately known until after the study is conducted. Because of this, we suggest that you calculate power or sample size for a range of values of the nuisance parameter. This will allow you to determine how sensitive the results are to its value.

Procedure Options

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

Data Tab (Common Options)

The Data tab contains the parameters associated with this test such as the proportions, sample sizes, alpha, and power. This chapter covers two procedures which have different options. This section documents options that are common to both procedures. Later, unique options for each procedure will be documented.

Solve For

Find (Solve For)

This option specifies the parameter to be solved for from the other parameters. The parameters that may be selected are *Power and Beta* or *N*.

Error Rates

Power or Beta

This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when you fail to conclude equivalence when in fact it is true.

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

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

Alpha (Significance Level)

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected. Here, a type-I error occurs when you falsely conclude equivalence.

Sample Size

N (Sample Size)

Enter a value for the sample size. This value must be greater than two. You may enter a range of values such as *10 to100 by 10*.

Effect Size – Standard Proportion

Ps (Standard Proportion)

This is the proportion of yes's (or successes), P_S, when subjects received the standard treatment. This value or a good estimate is often available from previous studies.

Note that this value does not matter when the Nuisance Parameter Type is set to 'P01' (or 'P10'), as long as it is greater than P01 (or P10).

You may enter a set of values separated by blank spaces. For example, you could enter *0.50 0.60 0.70*. Values between, but not including, 0 and 1 are permitted.

Effect Size – Nuisance Parameter

Nuisance Parameter Type

Enter the type of nuisance parameter here. Unfortunately, the 2-by-2 table cannot be completely specified by using only the parameters Ps and Da or Ps and Ra. One other parameter must be specified. This additional parameter is called a 'nuisance' parameter. It will be assumed to be a known quantity. Several possible choices are available. This option lets you specify which parameter you want to use. In all cases, the value you specify is a proportion.

- **P11**
  The proportion of subjects that are positive on both tests.
- **P00**
  The proportion of subjects that are negative on both tests.
- **P01**
  The proportion of subjects that are negative on the treatment, but positive on the standard.
- **P10**
  The proportion of subjects that are positive on the treatment, but negative on the standard.
- **P11+P00**
  The proportion of matches (concordant pairs).
- **P01+P10**
  The proportion of non-matches (discordant pairs).

- **P11/Ps**

  The sensitivity.

Nuisance Parameter Value

Enter the value of the nuisance parameter that you specified in the 'Nuisance Parameter Type' box. This value is a proportion, so it must be between 0 and 1.

Data Tab (Differences)

This section documents options that are used when the parameterization is in terms of the difference, P1 – P2. P1.0 is the value of P1 assumed by the null hypothesis and P1.1 is the value of P1 at which the power is calculated. Once P2, D0, and D1 are given, the values of P1.1 and P1.0 can be calculated.

Effect Size – Differences

|De| (Equivalence Difference)

De is the maximum allowable difference between the standard and treatment proportions that will still result in the conclusion of equivalence. In order to ensure that De is positive, the difference is computed in reverse order. That is, $D_E = P_S - P_T$. This parameter is only used when the Test Statistic option is set to *Difference*.

Only positive values can be entered here. Typical values for this difference are 0.05, 0.10, and 0.20. For two-sided tests, you must have |Da| < De. For one-sided tests, you must have Da > -De.

Da (Actual Difference)

Da is the actual difference between the treatment and standard proportions $D_A = P_T - P_S$. Da may be positive, negative, or (usually) zero. This parameter is only used when the Test Statistic option is set to *Difference*.

For two-sided tests, you must have |Da| < De. For one-sided tests, you must have Da > -De.

Data Tab (Ratios)

This section documents options that are used when the parameterization is in terms of the ratio, P1 / P2. P1.0 is the value of P1 assumed by the null hypothesis and P1.1 is the value of P1 at which the power is calculated. Once P2, R0, and R1 are given, the values of P1.0 and P1.1 can be calculated.

Effect Size – Ratios

Re (Equivalence Ratio)

Re is the minimum size of the relative risk ratio, P_T / P_S, that will still result in the conclusion of equivalence. Both equivalence and non-inferiority trials use a value that is less than one. Typical values for this ratio are 0.8 or 0.9.

This parameter is only used when the Test Statistic option is set to *Ratio*.

Ra (Actual Ratio)

Enter a value for Ra, the actual relative risk ratio P_T / P_S. This value is used to generate the value of P_T using the formula $P_T = P_S R_a$. Often this value is set equal to one, but this is not necessary.

This parameter is only used when the Test Statistic option is set to *Ratio*.

Options Tab

This tab sets a couple of options used in the iterative procedures.

Maximum Iterations

Maximum Iterations Before Search Termination

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.

Approximations

Use Approximations if N is greater than

Specify the maximum value of N (sample size) for which you would like an exact power calculation based on the multinomial distribution. Sample sizes greater than this value will use the asymptotic approximation given in the documentation. The exact calculation of the multinomial distribution becomes very time consuming for N > 200. For most cases, when N > 200, the difference between the exact and approximate calculations is small. For N > 200, the length of time needed to calculate the exact answer may become prohibitive. However, as the speed of computers increases, it will become faster and easier to calculate the exact power for larger values of N.

If you want all calculations to use exact results, enter ‘1000’ here.

If you want all calculations to use the quick approximations, enter ‘1’ here.

Example 1 – Finding Power

A clinical trial will be conducted to show that a non-invasive MRI test is equivalent to the invasive CTAP reference test. Historical data suggest that the CTAP test is 80% accurate. After careful discussion, the researchers decide that if the MRI test is five percentage points of the CTAP, it will be considered equivalent. They decide to use a difference test statistic. Thus, the equivalence difference is 0.05. They want to study the power for various sample sizes between 20 and 1000 at the 5% significance level. They decide to use the approximate power calculations, so they set the 'Use Approximations if N is greater than' option of the Options tab to 2.

They use P01 as the nuisance parameter and look at two values: 0.05 and 0.10.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Equivalence Tests for Two Correlated Proportions [Differences]** procedure window by expanding **Proportions**, then **Two Correlated Proportions**, then clicking on **Equivalence**, and then clicking on **Equivalence Tests for Two Correlated Proportions [Differences]**. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power	*Ignored since this is the Find setting*
Alpha	**0.05**
N (Sample Size)	**20 100 200 300 450 600 800 1000**
|De| (Equivalence Difference)	**0.05**
Da (Actual Difference)	**0.00**
Ps (Standard Proportion)	**0.80**
Nuisance Parameter Type	**P01**
Nuisance Parameter Value	**0.05 0.10**
Options Tab	
Use Approximations if N is >	**2**

Annotated Output

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

Numeric Results

Numeric Results for an Equivalence (Two-Sided) Test of a Difference

Power	Sample Size (N)	Equiv. Difference (De)	Actual Difference (Da)	Treatment Proportion (Pt)	Standard Proportion (Ps)	Nuisance Parameter (P01)	Alpha	Beta
0.00000	20	0.05000	0.00000	0.80000	0.80000	0.05000	0.05000	1.00000
0.00000	20	0.05000	0.00000	0.80000	0.80000	0.10000	0.05000	1.00000
0.00000	100	0.05000	0.00000	0.80000	0.80000	0.05000	0.05000	1.00000
0.00000	100	0.05000	0.00000	0.80000	0.80000	0.10000	0.05000	1.00000
0.35542	200	0.05000	0.00000	0.80000	0.80000	0.05000	0.05000	0.64458
0.00000	200	0.05000	0.00000	0.80000	0.80000	0.10000	0.05000	1.00000
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.

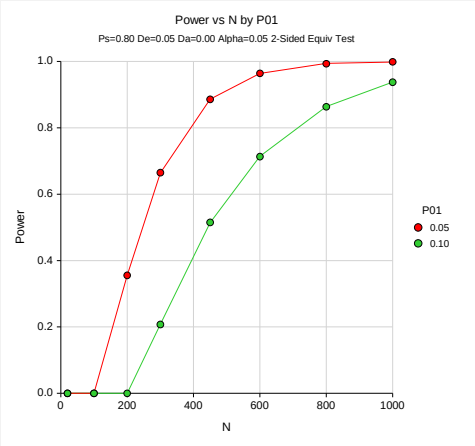
(report continues)

Report Definitions
Power is the probability of rejecting a false null hypothesis.
N is the number of subjects, the sample size.
De is the maximum difference between the two proportions that is still called 'equivalent'.
Da is the actual difference between Pt and Ps. That is, Da = Pt-Ps.
Pt is the response proportion to the treatment (experimental or new) test.
Ps is the response proportion to the standard (reference or old) test.
The Nuisance Parameter is a value that is needed, but is not a direct part of the hypothesis.
Alpha is the probability of rejecting a true null hypothesis.
Beta is the probability of accepting a false null hypothesis.

Summary Statements
A sample size of 20 subjects achieves 0% power at a 5% significance level using a two-sided equivalence test of correlated proportions when the standard proportion is 0.80000, the maximum allowable difference between these proportions that still results in equivalence (the range of equivalence) is 0.05000, and the actual difference of the proportions is 0.00000.

This report shows the power for the indicated scenarios. All of the columns are defined in the ‘Report Definitions’ section.

Plots Section



This plot shows the power versus the sample size for the two values of P01. In this example, we see that the value of the nuisance parameter has a large effect on the calculated sample size.

Example 2 – Finding Sample Size

Continuing with Example1, the analysts want to determine the exact sample size necessary to achieve 90% power for both values of the nuisance parameter.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Equivalence Tests for Two Correlated Proportions [Differences]** procedure window by expanding **Proportions**, then **Two Correlated Proportions**, then clicking on **Equivalence**, and then clicking on **Equivalence Tests for Two Correlated Proportions [Differences]**. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**Sample Size (N)**
Power	**0.90**
Alpha	**0.05**
N (Sample Size)	*Ignored since this is the Find setting*
|De| (Equivalence Difference)	**0.05**
Da (Actual Difference)	**0.00**
Ps (Standard Proportion)	**0.80**
Nuisance Parameter Type	**P01**
Nuisance Parameter Value	**0.05 0.10**
Options Tab	
Use Approximations if N is >	**2**

Output

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

Numeric Results

Numeric Results for an Equivalence (Two-Sided) Test of a Difference

Power	Sample Size (N)	Equiv. Difference (De)	Actual Difference (Da)	Treatment Proportion (Pt)	Standard Proportion (Ps)	Nuisance Parameter (P01)	Alpha	Beta
0.90019	468	0.05000	0.00000	0.80000	0.80000	0.05000	0.05000	0.09981
0.90002	881	0.05000	0.00000	0.80000	0.80000	0.10000	0.05000	0.09998

This report shows that the sample size required nearly doubles when P01 is changed from 0.05 to 0.10.

Example 3 – Validation using Liu

Liu *et al*. (2002) page 238 give a table of power values for sample sizes of 50, 100, and 200 when the significance level is 0.05. From this table, we find that when P01 is 0.10, P10 is 0.10, Da = P01 - P10 = 0.00, and De is 0.10, and the three power values are 0.026, 0.417, and 0.861 for the column head 'RMLE-based Without CC' (this is the case we use).

In their calculations, they round the z value to 1.64. This corresponds to an alpha value of 0.0505025835. So that our results match, we will use this value for alpha rather than 0.05.

In this example, the value of Ps is not used. We set it at 0.50. Also, we set the 'Use Approximations if N is greater than' value of the Options tab to 200 so that the exact values will be calculated.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Equivalence Tests for Two Correlated Proportions [Differences]** procedure window by expanding **Proportions**, then **Two Correlated Proportions**, then clicking on **Equivalence**, and then clicking on **Equivalence Tests for Two Correlated Proportions [Differences]**. You may then make the appropriate entries as listed below, or open **Example 3** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power	*Ignored since this is the Find setting*
Alpha	**0.0505025835**
N (Sample Size)	**50 100 200**
|De| (Equivalence Difference)	**0.1**
Da (Actual Difference)	**0.0**
Ps (Standard Proportion)	**0.5**
Max N Using Exact Power	**200**
Nuisance Parameter Type	**P01**
Nuisance Parameter Value	**0.1**

Output

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

Numeric Results

Numeric Results for an Equivalence (Two-Sided) Test of a Difference

Power	Sample Size (N)	Equiv. Difference (De)	Actual Difference (Da)	Treatment Proportion (Pt)	Standard Proportion (Ps)	Nuisance Parameter (P01)	Alpha	Beta
0.02614	50	0.10000	0.00000	0.50000	0.50000	0.10000	0.05050	0.97386
0.41741	100	0.10000	0.00000	0.50000	0.50000	0.10000	0.05050	0.58259
0.86080	200	0.10000	0.00000	0.50000	0.50000	0.10000	0.05050	0.13920

As you can see, the values computed by ***PASS*** match the results of Liu et al. (2002).

Example 4 – Finding Power Following an Experiment

An experiment involving a single group of 57 subjects was run to show that a new treatment was equivalent to a previously used standard. Historically, the standard treatment has had a 48% success rate. The new treatment is known to have similar side effects to the standard, but is much less expensive. The treatments were to be considered equivalent if the success rate of the new treatment is within 10% of the success rate of the standard.

To compare the new and standard treatments, each of the 57 subjects received both treatments with a washout period between them. Thus, the proportions based on the two treatments are correlated. Of the 57 subjects, 18 responded to both treatments, 20 did not respond to either treatment, 9 responded to the new treatment but not the standard, and 10 responded to the standard but not the new treatment. The proportion responding to the new treatment is (18+9)/57 = 0.4737. The proportion responding to the standard is (18+10)/57 = 0.4912. The difference is 0.0175, lower than the threshold for equivalence, but the resulting p-value was 0.3358, indicating the two treatments could not be deemed equivalent at the 0.05 level. Note that McNemar's test only uses the discordant pairs, so the effective size of this study is really only 9 + 10 = 19, although 57 subjects were investigated. The researchers want to know the power of the test they used.

It may be the inclination of the researchers to use the observed difference in proportions for calculating power. The p-value, however, is based on the maximum allowable difference for equivalence, which is 10% of 0.48, or 0.048. This is the number that should be used in the power calculation. The experiment gave a value of P01 of 10/28 = 0.36. The power of the experiment is near zero for all values of P01 less than 0.10. We calculate the power for a variety of nuisance parameter values (P01 = 0.01, 0.03, 0.05, and 0.10) to monitor its effect. Because it is in fact believed that the success rates are equivalent for the two treatments, the specified actual difference is set to 0.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Equivalence Tests for Two Correlated Proportions [Differences]** procedure window by expanding **Proportions**, then **Two Correlated Proportions**, then clicking on **Equivalence**, and then clicking on **Equivalence Tests for Two Correlated Proportions [Differences]**. You may then make the appropriate entries as listed below, or open **Example 4** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power	*Ignored since this is the Find setting*
Alpha	**0.05**
N (Sample Size)	**57**
|De| (Equivalence Difference)	**0.048**
Da (Actual Difference)	**0.00**
Ps (Standard Proportion)	**0.48**
Nuisance Parameter Type	**P01**
Nuisance Parameter Value	**0.01 0.03 0.05 0.10**
Options Tab	
Use Approximations if N is >	**200**

Output

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

Numeric Results

Numeric Results for an Equivalence (Two-Sided) Test of a Difference

Power	Sample Size (N)	Equiv. Difference (De)	Actual Difference (Da)	Treatment Proportion (Pt)	Standard Proportion (Ps)	Nuisance Parameter (P01)	Alpha	Beta
0.31614	57	0.04800	0.00000	0.48000	0.48000	0.01000	0.05000	0.68386
0.02940	57	0.04800	0.00000	0.48000	0.48000	0.03000	0.05000	0.97060
0.00247	57	0.04800	0.00000	0.48000	0.48000	0.05000	0.05000	0.99753
0.00000	57	0.04800	0.00000	0.48000	0.48000	0.10000	0.05000	1.00000

Note that there is no power for value of P01 greater than 0.05. This is probably due to the low number of discordant pairs.

Chapter 200

Inequality Tests for Two Proportions

Introduction

This module computes power and sample size for hypothesis tests of the difference, ratio, or odds ratio of two independent proportions. The test statistics analyzed by this procedure assume that the difference between the two proportions is zero or their ratio is one under the null hypothesis. The *non-null* (*offset*) *case* is discussed in another procedure. This procedure computes and compares the power achieved by each of several test statistics that have been proposed.

For example, suppose you want to compare two methods for treating cancer. Your experimental design might be as follows. Select a sample of patients and randomly assign half to one method and half to the other. After five years, determine the proportion surviving in each group and test whether the difference in the proportions is significantly different from zero.

The power calculations assume that random samples are drawn from two separate populations.

Four Procedures Documented Here

There are four procedures in the menus that use the program module described in this chapter. These procedures are identical except for the type of parameterization. The parameterization can be in terms of proportions, differences in proportions, ratios of proportions, and odds ratios. Each of these options is listed separately on the menus.

Technical Details

Suppose you have two populations from which dichotomous (binary) responses will be recorded. The probability (or risk) of obtaining the event of interest in population 1 (the treatment group) is p_1 and in population 2 (the control group) is p_2. The corresponding failure proportions are given by $q_1 = 1 - p_1$ and $q_2 = 1 - p_2$.

The assumption is made that the responses from each group follow a binomial distribution. This means that the event probability, p_i, is the same for all subjects within the group and that the response from one subject is independent of that of any other subject.

Random samples of *m* and *n* individuals are obtained from these two populations. The data from these samples can be displayed in a 2-by-2 contingency table as follows

Group	Success	Failure	Total
Treatment	*a*	*c*	*m*
Control	*b*	*d*	*n*
Total	*s*	*f*	*N*

The following alternative notation is also used.

Group	Success	Failure	Total
Treatment	x_{11}	x_{12}	n_1
Control	x_{21}	x_{22}	n_2
Total	m_1	m_2	N

The binomial proportions p_1 and p_2 are estimated from these data using the formulae

$$\hat{p}_1 = \frac{a}{m} = \frac{x_{11}}{n_1} \text{ and } \hat{p}_2 = \frac{b}{n} = \frac{x_{21}}{n_2}$$

Comparing Two Proportions

When analyzing studies such as this, one usually wants to compare the two binomial probabilities, p_1 and p_2. Common measures for comparing these quantities are the difference and the ratio. If the binomial probabilities are expressed in terms of odds rather than probabilities, another common measure is the odds ratio. Mathematically, these comparison parameters are

Parameter	**Computation**
Difference	$\delta = p_1 - p_2$
Risk Ratio	$\phi = p_1 / p_2$
Odds Ratio	$\psi = \frac{p_1 / (1 - p_1)}{p_2 / (1 - p_2)} = \frac{p_1 q_2}{p_2 q_1}$

The tests analyzed by this routine are for the *null case*. This refers to the values of the above parameters under the null hypothesis. In the *null case*, the difference is zero and the ratios are one under the null hypothesis. In the *non-null case*, discussed in another chapter, the difference is some value other than zero and the ratios are some value other than one. The non-null case often appears in equivalence and non-inferiority testing.

Hypothesis Tests

Several statistical tests have been developed for testing the inequality of two proportions. For large samples, the powers of the various tests are about the same. However, for small samples, the differences in the powers can be quite large. Hence, it is important to base the power analysis on the test statistic that will be used to analyze the data. If you have not selected a test statistic, you may wish to determine which one offers the best power in your situation. No single test is the champion in every situation, so you must compare the powers of the various tests to determine which to use.

Difference

The (risk) difference, $\delta = p_1 - p_2$, is perhaps the most direct measure for comparing two proportions. Three sets of statistical hypotheses can be formulated:

1. $H_0: p_1 - p_2 = 0$ versus $H_1: p_1 - p_2 \neq 0$; this is often called the *two-tailed test*.
2. $H_0: p_1 - p_2 \leq 0$ versus $H_1: p_1 - p_2 > 0$; this is often called the *upper-tailed test*.
3. $H_0: p_1 - p_2 \geq 0$ versus $H_1: p_1 - p_2 < 0$; this is often called the *lower-tailed test*.

The traditional approach for testing these hypotheses has been to use the Pearson chi-square test for large samples, the Yates chi-square for intermediate sample sizes, and the Fisher Exact test for small samples. Recently, some authors have begun questioning this solution. For example, based on exact enumeration, Upton (1982) and D'Agostino (1988) conclude that the Fisher Exact test and Yates test should never be used.

Ratio

The (risk) ratio, $\phi = p_1 / p_2$, is often preferred to the difference when the baseline proportion is small (less than 0.1) or large (greater than 0.9) because it expresses the difference as a percentage rather than an amount. In this null case, the null hypothesized ratio of proportions, ϕ_0, is one. Three sets of statistical hypotheses can be formulated:

1. $H_0: p_1 / p_2 = \phi_0$ versus $H_1: p_1 / p_2 \neq \phi_0$; this is often called the *two-tailed test*.
2. $H_0: p_1 / p_2 \leq \phi_0$ versus $H_1: p_1 / p_2 > \phi_0$; this is often called the *upper-tailed test*.
3. $H_0: p_1 / p_2 \geq \phi_0$ versus $H_1: p_1 / p_2 < \phi_0$; this is often called the *lower-tailed test*.

Odds Ratio

The odds ratio, $\psi = \dfrac{o_1}{o_2} = \dfrac{p_1 / (1 - p_1)}{p_2 / (1 - p_2)} = \dfrac{p_1 q_2}{p_2 q_1}$, is sometimes used to compare the two proportions because of its statistical properties and because some experimental designs require its use. In this null case, the null hypothesized odds ratio, ψ_0, is one. Three sets of statistical hypotheses can be formulated:

1. $H_0: \psi = \psi_0$ versus $H_1: \psi \neq \psi_0$; this is often called the *two-tailed test*.

2. $H_0: \psi \leq \psi_0$ versus $H_1: \psi > \psi_0$; this is often called the *upper-tailed test*.
3. $H_0: \psi \geq \psi_0$ versus $H_1: \psi < \psi_0$; this is often called the *lower-tailed test*.

Power Calculation

The power for a test statistic that is based on the normal approximation can be computed exactly using two binomial distributions. The following steps are taken to compute the power of such a test.

1. Find the critical value (or values in the case of a two-sided test) using the standard normal distribution. The critical value, $z_{critical}$, is that value of z that leaves exactly the target value of alpha in the appropriate tail of the normal distribution. For example, for an upper-tailed test with a target alpha of 0.05, the critical value is 1.645.
2. Compute the value of the test statistic, z_t, for every combination of x_{11} and x_{21}. Note that x_{11} ranges from 0 to n_1, and x_{21} ranges from 0 to n_2. A small value (around 0.0001) can be added to the zero cell counts to avoid numerical problems that occur when the cell value is zero.
3. If $z_t > z_{critical}$, the combination is in the rejection region. Call all combinations of x_{11} and x_{21} that lead to a rejection the set *A*.
4. Compute the power for given values of p_1 and p_2 as

$$1-\beta = \sum_A \binom{n_1}{x_{11}} p_1^{x_{11}} q_1^{n_1-x_{11}} \binom{n_2}{x_{21}} p_2^{x_{21}} q_2^{n_2-x_{21}}$$

5. Compute the actual value of alpha achieved by the design by substituting p_2 for p_1

$$\alpha^* = \sum_A \binom{n_1}{x_{11}} p_1^{x_{11}} q_1^{n_1-x_{11}} \binom{n_2}{x_{21}} p_1^{x_{21}} q_1^{n_2-x_{21}}$$

When the values of n_1 and n_2 are large (say over 200), these formulas may take a little time to evaluate. In this case, a large sample approximation may be used.

Test Statistics

The various test statistics that are available in this routine are listed next.

Fisher's Exact Test

The most useful reference we found for power analysis of Fisher's Exact test was in the StatXact 5 (2001) documentation. The material present here is summarized from Section 26.3 (pages 866 – 870) of the StatXact-5 documentation. In this case, the test statistic is

$$T = -\ln\left[\frac{\binom{n_1}{x_1}\binom{n_2}{x_2}}{\binom{N}{m}}\right]$$

The null distribution of T is based on the hypergeometric distribution. It is given by

$$\Pr\left(T \ge t | m, H_0\right) = \sum_{A(m)} \left[\frac{\binom{n_1}{x_1}\binom{n_2}{x_2}}{\binom{N}{m}} \right]$$

where

$$A(m) = \left\{ \text{all pairs } x_1, x_2 \text{ such that } x_1 + x_2 = m, \text{ given } T \ge t \right\}$$

Conditional on *m*, the critical value, t_α, is the smallest value of *t* such that

$$\Pr\left(T \ge t_\alpha | m, H_0\right) \le \alpha$$

The power is defined as

$$1 - \beta = \sum_{m=0}^{N} P(m) \Pr\left(T \ge t_\alpha | m, H_1\right)$$

where

$$\Pr\left(T \ge t_\alpha | m, H_1\right) = \sum_{A(m, T \ge t_\alpha)} \left[\frac{b(x_1, n_1, p_1) b(x_2, n_2, p_2)}{\sum_{A(m)} b(x_1, n_1, p_1) b(x_2, n_2, p_2)} \right]$$

$$\begin{aligned} P(m) &= \Pr\left(x_1 + x_2 = m | H_1\right) \\ &= b(x_1, n_1, p_1) b(x_2, n_2, p_2) \end{aligned}$$

$$b(x, n, p) = \binom{n}{x} p^x (1-p)^{n-x}$$

When either group's sample size is greater than the Maximum N1 or N2 limit, an approximation is used based on the pooled, continuity corrected Chi-Square test.

Chi-Square Test (Pooled and Unpooled)

This test statistic was first proposed by Karl Pearson in 1900. Although this test is usually expressed directly as a Chi-Square statistic, it is expressed here as a *z* statistic so that it can be more easily used for one-sided hypothesis testing.

Both *pooled* and *unpooled* versions of this test have been discussed in the statistical literature. The pooling refers to the way in which the standard error is estimated. In the pooled version, the two proportions are averaged, and only one proportion is used to estimate the standard error. In the unpooled version, the two proportions are used separately.

The formula for the test statistic is

$$z_t = \frac{\hat{p}_1 - \hat{p}_2}{\hat{\sigma}_D}$$

Pooled Version

$$\hat{\sigma}_D = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

$$\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1+n_2}$$

Unpooled Version

$$\hat{\sigma}_D = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

Power

The power of this test is computed using the enumeration procedure described above. For large sample sizes, the following approximation is used.

1. Find the critical value (or values in the case of a two-sided test) using the standard normal distribution. The critical value is that value of z that leaves exactly the target value of alpha in the tail.
2. Use the normal approximation to binomial distribution to compute binomial probabilities, Compute the power using

$$1-\beta = \Pr\left(Z < z_\alpha\sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} + \frac{p_1-p_2}{\sqrt{\frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}}}\right)$$

where

$$\bar{p} = \frac{n_1p_1 + n_2p_2}{n_1+n_2}$$

$$\bar{q} = 1-\bar{p}$$

Chi-Square Test with Continuity Correction

Frank Yates is credited with proposing a correction to the Pearson Chi-Square test for the lack of continuity in the binomial distribution. However, the correction was in common use when he proposed it in 1922.

Both *pooled* and *unpooled* versions of this test have been discussed in the statistical literature. The pooling refers to the way in which the standard error is estimated. In the pooled version, the two proportions are averaged, and only one proportion is used to estimate the standard error. In the unpooled version, the two proportions are used separately.

The continuity corrected z-test is

$$z = \frac{\left(\hat{p}_1-\hat{p}_2\right) + \frac{F}{2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}{\hat{\sigma}_D}$$

where F is -1 for lower-tailed, 1 for upper-tailed, and both -1 and 1 for two-sided hypotheses.

Pooled Version

$$\hat{\sigma}_D = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

$$\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1+n_2}$$

Unpooled Version

$$\hat{\sigma}_D = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

Power

The power of this test is computed using the enumeration procedure described for the *z*-test above. For large samples, approximate results based on the normal approximation to the binomial are used.

Conditional Mantel Haenszel Test

The conditional Mantel Haenszel test, see Lachin (2000) page 40, is based on the *index frequency*, x_{11}, from the 2x2 table. The formula for the z-statistic is

$$z = \frac{x_{11} - E(x_{11})}{\sqrt{V_c(x_{11})}}$$

where

$$E(x_{11}) = \frac{n_1 m_1}{N}$$

$$V_c(x_{11}) = \frac{n_1 n_2 m_1 m_2}{N^2(N-1)}$$

Power

The power of this test is computed using the enumeration procedure described above.

Likelihood Ratio Test

In 1935, Wilks showed that the following quantity has a chi-square distribution with one degree of freedom. Using this test statistic to compare proportions is presented, among other places, in Upton (1982). The likelihood ratio test statistic is computed as

$$LR = 2\begin{bmatrix} a\ln(a)+b\ln(b)+c\ln(c)+d\ln(d)+ \\ N\ln(N)-s\ln(s)-f\ln(f)-m\ln(m)-n\ln(n) \end{bmatrix}$$

Power

The power of this test is computed using the enumeration procedure described above. When large sample results are needed, the results for the *z* test are used.

T-Test

Based on a study of the behavior of several tests, D'Agostino (1988) and Upton (1982) proposed using the usual two-sample *t*-test for testing whether two proportions are equal. One substitutes a '1' for a success and a '0' for a failure in the usual, two-sample *t*-test formula. The test statistic is computed as

$$t_{N-2} = (ad - bc)\left(\frac{N-2}{N(nac + mbd)}\right)^{\frac{1}{2}}$$

which can be compared to the *t* distribution with *N*-2 degrees of freedom.

Power

The power of this test is computed using the enumeration procedure described above, except that the *t* tables are used instead of the standard normal tables.

Procedure Options

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

Data Tab (Common Options)

The Data tab contains the parameters associated with this test such as the proportions, sample sizes, alpha, and power. This chapter covers four procedures, each of which has different options. This section documents options that are common to all four procedures. Later, unique options for each procedure will be documented.

Solve For

Find (Solve For)

This option specifies the parameter to be solved for using the other parameters. The parameters that may be selected are *P1*, *Power and Beta*, *N1*, and *N2*. Under most situations, you will select either *Power and Beta* or *N1*.

Select *N1* when you want to calculate the sample size needed to achieve a given power and alpha level.

Select *Power and Beta* when you want to calculate the power of an experiment.

Error Rates

Power or Beta

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

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

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

Alpha (Significance Level)

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

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

Sample Size

N1 (Sample Size Group 1)

Enter a value (or range of values) for the sample size of this group. 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.

Effect Size

P2 (Control Group Proportion)

Specify the value of p_2, the control, baseline, or standard group's proportion. The null hypothesis is that the two proportions, p_1 and p_2, are both equal to this value.

Since these values are proportions, values must be 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 0.1*.

Test

Test Type

Specify which test statistic will be used in searching and reporting.

Note that 'C.C.' is an abbreviation for *Continuity Correction*. This refers to the adding or subtracting 2/(N1+N2) to (or from) the numerator of the z-value to bring the normal approximation closer to the binomial distribution.

Data Tab (Proportions)

This section documents options that are used when the parameterization is in terms of the values of the two proportions, P1 and P2. P1 is the value of the P1 at which the power is calculated.

Effect Size

P1 (Treatment Group Proportion |H1)

This is the value of P1 under the alternative hypothesis, H1. The power calculations assume that this is the actual value of this proportion.

You may enter a range of values such as *0.1 0.2 0.3* or *0.1 to 0.9 by 0.1*.

Note that values must be between zero and one and cannot be equal to P2.

Test

H1 (Alternative Hypothesis)

This option specifies the type of alternative hypothesis. The null hypothesis is H0: P1 = P2.

One-Sided (H1:P1<P2) refers to a one-sided test in which the alternative hypothesis is of the form H1: P1<P2.

One-Sided (H1:P1>P2) refers to a one-sided test in which the alternative hypothesis is of the form H1: P1>P2.

Two-Sided refers to a two-sided test in which the alternative hypothesis is of the type H1: P1<>P2. Here '<>' means unequal.

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

Data Tab (Differences)

This section documents options that are used when the parameterization is in terms of the difference, P1 – P2. P1 is the value of the group 1 proportion at which the power is calculated. Once P2 and D1 are given, the value of P1 can be calculated.

Effect Size

D1 (Difference|H1 = P1-P2)

This option specifies the difference between the two proportions under the alternative hypothesis, H1. This difference is used with P2 to calculate the value of P1 using the formula: P1 = D1 + P2. Differences must be between -1 and 1. They cannot take on the values -1, 0, or 1.

The power calculations assume that P1 is the actual value of the proportion in group 1 (the experimental or treatment group).

You may enter a range of values such as *0.03 0.05 0.10* or *0.01* to *0.05 by 0.01*.

Test

Alternative Hypothesis (H1)

This option specifies the type of alternative hypothesis. The null hypothesis is H0: P1 = P2.

One-Sided (H1:D1<0) refers to a one-sided test in which the alternative hypothesis is of the form H1: D1<0.

One-Sided (H1:D1>0) refers to a one-sided test in which the alternative hypothesis is of the form H1: D1>0.

Two-Sided refers to a two-sided test in which the alternative hypothesis is of the type H1: D1<>0. Here '<>' means unequal.

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

Data Tab (Ratios)

This section documents options that are used when the parameterization is in terms of the ratio, P1 / P2. P1 is the value of the group 1 proportion at which the power is calculated. Once P2 and R1 are given, the value of P1 can be calculated.

Effect Size

R1 (Ratio|H1 = P2/P1)

This option specifies the ratio between the two proportions, P1 and P2. This ratio is used with P2 to calculate the value of P1 at which the power is calculated using the formula: P1=(R1) x (P2). The power calculations assume that P1 is the actual value of the proportion in group 1, which is the experimental, or treatment, group.

You may enter a range of values such as *0.5 0.6 0.7 0.8* or *1.25 to 2.0 by 0.25*. Ratios must greater than zero. They cannot take on the value of one.

Test

Alternative Hypothesis (H1)

This option specifies the type of alternative hypothesis. The null hypothesis is H0: P1 = P2.

One-Sided (H1:R1<1) refers to a one-sided test in which the alternative hypothesis is of the form H1: R1<1.

One-Sided (H1:R1>1) refers to a one-sided test in which the alternative hypothesis is of the form H1: R1>1.

Two-Sided refers to a two-sided test in which the alternative hypothesis is of the type H1: R1<>1, Here '<>' means unequal.

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

Data Tab (Odds Ratios)

This section documents options that are used when the parameterization is in terms of the odds ratios. Note that the odds are defined as O2 = P2 / (1 – P2) and O1 = P1 / (1 – P1). Once P2 and OR1 are given, the value of P1 can be calculated.

Effect Size

OR1 (Odds Ratio|H1 = O1/O2)

This option specifies the odds ratio of the two proportions, P1 and P2. This odds ratio is used with P2 to calculate the value of P1. The power calculations assume that P1 is the actual value of the proportion in group 1, which is the experimental, or treatment, group.

You may enter a range of values such as *0.5 0.6 0.7 0.8* or *1.25 to 2.0 by 0.25*. Odds ratios must greater than zero. They cannot take on the value of one.

Test

Alternative Hypothesis (H1)

This option specifies the type of alternative hypothesis. The null hypothesis is H0: P1 = P2.

One-Sided (H1:OR1<1) refers to a one-sided test in which the alternative hypothesis is of the form H1: P1<P2, H1: D1<0, H1: R1<1, or H1: OR1<1.

One-Sided (H1: OR1>1) refers to a one-sided test in which the alternative hypothesis is of the form H1: P1>P2, H1: D1>0, H1: R1>1, or H1: OR1>1.

Two-Sided refers to a two-sided test in which the alternative hypothesis is of the type H1: OR1<>1. Here '<>' means unequal.

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

Options Tab

The Options tab contains various limits and options.

Maximum Iterations

Maximum Iterations Before Search Termination

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 not reported. A value of at least 500 is recommended.

Zero Counts

Zero Count Adjustment Method

Zero cell counts often cause calculation problems. To compensate for this, a small value (called the Zero Count Adjustment Value) can be added either to all cells or to all cells with zero counts. This option specifies whether you want to use the adjustment and which type of adjustment you want to use. We recommend that you use the option *Add to zero cells only*.

Zero cell values often do not occur in practice. However, since power calculations are based on total enumeration, they will occur in power and sample size estimation.

Adding a small value is controversial, but can be necessary for computational considerations. Statisticians have recommended adding various fractions to zero counts. We have found that adding 0.0001 seems to work well.

Zero Count Adjustment Value

Zero cell counts cause many calculation problems when computing power or sample size. To compensate for this, a small value may be added either to all cells or to all zero cells. This value indicates the amount that is added. We have found that 0.0001 works well.

Be warned that the value of the ratio and the odds ratio will be affected by the amount specified here!

Exact Test Options

Maximum N1 or N2 for Exact Calculations

When either N1 or N2 is above this amount, power calculations are based on the normal approximation to the binomial. In this case, the actual value of alpha is not calculated. Currently, for three-gigahertz computers, a value near 200 is reasonable. As computers increase in speed, this number may be increased.

Calculate Exact Test Results

When checked, the power of Fisher’s Exact Test will be calculated for the comparative reports, even if the ‘Test Statistic’ option is not set to *Fisher’s Exact Test*.

This option is provided because calculations for Fisher's Exact Test can become lengthy for large sample sizes.

Example 1 – Finding Power

A study is being designed to study the effectiveness of a new treatment. Historically, the standard treatment has enjoyed a 60% cure rate. Researchers want to compute the power of the two-sided *z*-test at group sample sizes ranging from 50 to 650 for detecting differences of 0.05 and 0.10 in the cure rate at the 0.05 significance level.

Setup

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

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power	*Ignored since this is the Find setting*
Alpha	**0.05**
N1 (Sample Size Group 1)	**50 to 650 by 100**
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
D1 (Difference|H1 = P1 – P2)	**0.05 0.10**
P2 (Control Group Proportion)	**0.6**
H1 (Alternative Hypothesis)	**Two-Sided**
Test Type	**Z Test (Pooled)**
Options Tab	
Maximum N1 or N2 Exact	**400**

Annotated Output

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

Numeric Results

Numeric Results of Tests Based on the Difference: P1 – P2
H0: P1-P2=0. H1: P1-P2<>D1. Test Statistic: Z test with pooled variance

Power	Sample Size Grp 1 N1	Sample Size Grp 2 N2	Prop|H1 Grp 1 or Trtmnt P1	Prop Grp 2 or Control P2	Diff if H0 D0	Diff if H1 D1	Target Alpha	Actual Alpha	Beta
0.08173	50	50	0.65000	0.60000	0.00000	0.05000	0.05000	0.05239	0.91827
0.14469	150	150	0.65000	0.60000	0.00000	0.05000	0.05000	0.05173	0.85531
0.20852	250	250	0.65000	0.60000	0.00000	0.05000	0.05000	0.04981	0.79148
0.27586	350	350	0.65000	0.60000	0.00000	0.05000	0.05000	0.04946	0.72414
0.34064	450	450	0.65000	0.60000	0.00000	0.05000	0.05000		0.65936
0.40234	550	550	0.65000	0.60000	0.00000	0.05000	0.05000		0.59766
0.46095	650	650	0.65000	0.60000	0.00000	0.05000	0.05000		0.53905
0.18042	50	50	0.70000	0.60000	0.00000	0.10000	0.05000	0.05239	0.81958

0.43968	150	150	0.70000	0.60000	0.00000	0.10000	0.05000	0.05173	0.56032
0.65180	250	250	0.70000	0.60000	0.00000	0.10000	0.05000	0.04981	0.34820
0.79585	350	350	0.70000	0.60000	0.00000	0.10000	0.05000	0.04946	0.20415
0.88326	450	450	0.70000	0.60000	0.00000	0.10000	0.05000		0.11674
0.93640	550	550	0.70000	0.60000	0.00000	0.10000	0.05000		0.06360
0.96636	650	650	0.70000	0.60000	0.00000	0.10000	0.05000		0.03364

Note: exact results based on the binomial were only made when both N1 and N2 were less than 400.

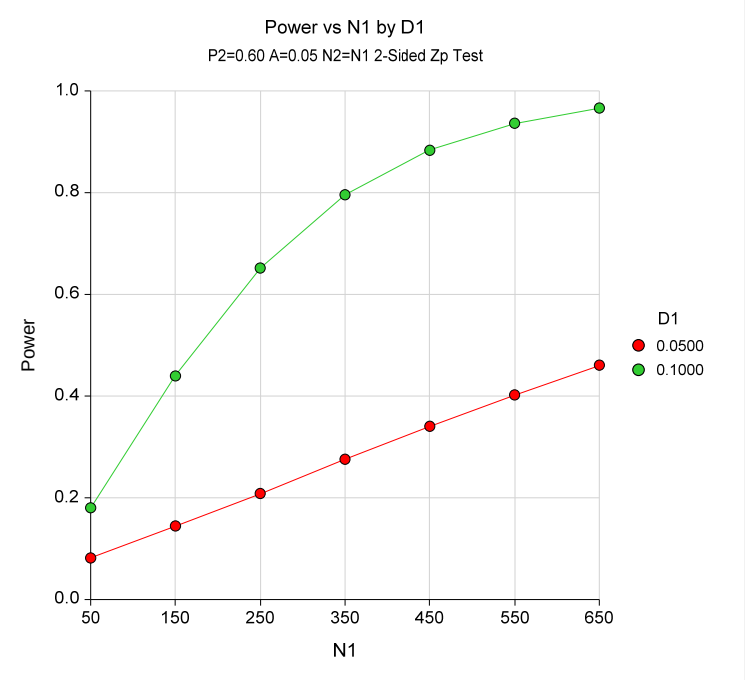
Report Definitions
'Power' is the probability of rejecting a false null hypothesis. It should be close to one.
'N1 and N2' are the sizes of the samples drawn from the corresponding populations.
'P1' is the proportion for group one under H1. This is the treatment or experimental group.
'P2' is the proportion for group two. This is the standard, reference, or control group
'D1: Diff. if H1' is the difference P1 – P2 assuming the alternative hypothesis.
'Target Alpha' is the probability of rejecting a true null hypothesis that was desired.
'Actual Alpha' is the value of alpha that is actually achieved.
'Beta' is the probability of accepting a false null hypothesis.

Summary Statements
Group sample sizes of 50 in group one and 50 in group two achieve 8% power to detect a difference between the group proportions of 0.05000. The proportion in group one (the treatment group) is assumed to be 0.60000 under the null hypothesis and 0.65000 under the alternative hypothesis. The proportion in group two (the control group) is 0.60000. The test statistic used is the two-sided Z test . The significance level of the test was targeted at 0.05000. The significance level actually achieved by this design is 0.05239.

This report shows the values of each of the parameters, one scenario per row. Note that the actual alpha value is blank for sample sizes greater than 400, which was the limit set for exact computation.

The values from this table are plotted in the chart below.

Plots Section



Example 2 – Finding the Sample Size

A clinical trial is being designed to test effectiveness of new drug in reducing mortality. Suppose the current cure rate during the first year is 0.44. The sample size should be large enough to detect a difference in the cure rate of 0.10. Assuming the test statistic is a two-sided *z*-test with a significance level of 0.05, what sample size will be necessary to achieve 90% power?

Setup

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

Option	Value
Data Tab	
Find (Solve For)	**N1**
Power	**0.90**
Alpha	**0.05**
N1 (Sample Size Group 1)	*Ignored since this is the Find setting*
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
D1 (Difference|H1 = P1 – P2)	**0.10**
P2 (Control Group Proportion)	**0.44**
H1 (Alternative Hypothesis)	**Two-Sided**
Test Type	**Z Test (Pooled)**
Options Tab	
Maximum N1 or N2 Exact	**100 (Set low for a rapid search.)**

Output

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

Numeric Results

Numeric Results of Tests Based on the Difference: P1 – P2
H0: P1-P2=0. H1: P1-P2<>D1. Test Statistic: Z test with pooled variance

Power	Sample Size Grp 1 N1	Sample Size Grp 2 N2	Prop|H1 Grp 1 or Trtmnt P1	Prop Grp 2 or Control P2	Diff if H0 D0	Diff if H1 D1	Target Alpha	Actual Alpha	Beta
0.90050	524	524	0.54000	0.44000	0.00000	0.10000	0.05000		0.09950

The required sample size is 524 per group. These results use the large sample approximation. As an exercise, reset the Maximum N1 or N2 Exact parameter to 600 so that exact results can be calculated. When this is done, the sample size is 521—not much of a difference from the 524 that

was found by approximate methods. The actual alpha is 0.04930 which is very close to the target of 0.05.

Example 3 – Comparing the Power of Several Test Statistics

Researchers want to determine which of the eight test statistics to adopt using the comparative reports and charts that ***PASS*** produces. They want to detect a difference of 0.20 when the response rate of the control group is 0.30. The significance level is 0.05. They want to study sample sizes from 10 to 100.

Setup

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

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power	*Ignored since this is the Find setting*
Alpha	**0.05**
N1 (Sample Size Group 1)	**10 to 100 by 10**
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
D1 (Difference|H1 = P1 – P2)	**0.2**
P2 (Control Group Proportion)	**0.3**
H1 (Alternative Hypothesis)	**Two-Sided**
Test Type	**Z Test (Pooled)**
Reports Tab	
Show Numeric Report	**Not checked**
Show Comparative Reports	**Checked**
Show Definitions	**Not checked**
Show Plots	**Not checked**
Show Comparative Plots	**Checked**
Show Summary Statements	**Not checked**
Options Tab	
Maximum N1 or N2 Exact	**400**
Calculate Exact Test Results	**Checked**

Output

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

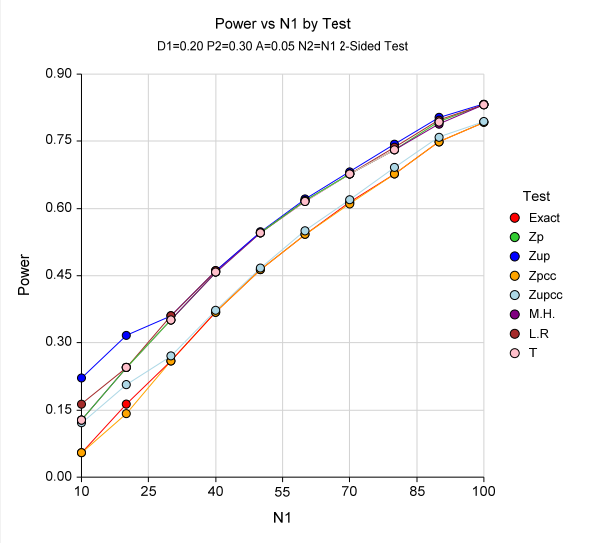
Numeric Results and Plots

Power Comparison of Tests Based on the Difference: P1 – P2
H0: P1-P2=0. H1: P1-P2<>D1.

N1/N2	P1	P2	Target Alpha	Exact Test Power	Z(P) Test Power	Z(UnP) Test Power	Z(P) cc Test Power	Z(UnP) cc Test Power	Mantel Hnzl. Power	Like. Ratio Power	T Test Power
10/10	0.5000	0.3000	0.0500	0.0547	0.1275	0.2215	0.0547	0.1215	0.1275	0.1629	0.1275
20/20	0.5000	0.3000	0.0500	0.1632	0.2452	0.3167	0.1419	0.2067	0.2452	0.2452	0.2452
30/30	0.5000	0.3000	0.0500	0.2594	0.3511	0.3604	0.2594	0.2708	0.3511	0.3604	0.3511
40/40	0.5000	0.3000	0.0500	0.3683	0.4581	0.4612	0.3683	0.3728	0.4581	0.4612	0.4581
50/50	0.5000	0.3000	0.0500	0.4635	0.5455	0.5481	0.4635	0.4671	0.5455	0.5455	0.5455
60/60	0.5000	0.3000	0.0500	0.5424	0.6177	0.6214	0.5424	0.5501	0.6157	0.6177	0.6157
70/70	0.5000	0.3000	0.0500	0.6138	0.6771	0.6815	0.6101	0.6195	0.6771	0.6771	0.6771
80/80	0.5000	0.3000	0.0500	0.6773	0.7310	0.7435	0.6773	0.6917	0.7310	0.7368	0.7310
90/90	0.5000	0.3000	0.0500	0.7485	0.7930	0.8036	0.7485	0.7589	0.7882	0.7969	0.7930
100/100	0.5000	0.3000	0.0500	0.7924	0.8320	0.8328	0.7924	0.7942	0.8316	0.8320	0.8316

Actual Alpha Comparison of Tests Based on the Difference: P1 - P2
H0: P1-P2=0. H1: P1-P2=D1<>0.

N1/N2	P1	P2	Target Alpha	Exact Test Alpha	Z(P) Test Alpha	Z(UnP) Test Alpha	Z(P) cc Test Alpha	Z(UnP) cc Test Alpha	Mantel Hnzl. Alpha	Like. Ratio Alpha	T Test Alpha
10/10	0.5000	0.3000	0.0500	0.0119	0.0371	0.0949	0.0119	0.0258	0.0371	0.0771	0.0371
20/20	0.5000	0.3000	0.0500	0.0248	0.0533	0.0686	0.0214	0.0267	0.0533	0.0534	0.0533
30/30	0.5000	0.3000	0.0500	0.0261	0.0487	0.0583	0.0261	0.0321	0.0487	0.0583	0.0487
40/40	0.5000	0.3000	0.0500	0.0282	0.0484	0.0541	0.0276	0.0317	0.0484	0.0541	0.0484
50/50	0.5000	0.3000	0.0500	0.0307	0.0498	0.0554	0.0307	0.0334	0.0498	0.0498	0.0498
60/60	0.5000	0.3000	0.0500	0.0308	0.0525	0.0552	0.0308	0.0353	0.0483	0.0525	0.0491
70/70	0.5000	0.3000	0.0500	0.0330	0.0516	0.0549	0.0318	0.0348	0.0516	0.0516	0.0516
80/80	0.5000	0.3000	0.0500	0.0331	0.0513	0.0518	0.0331	0.0350	0.0493	0.0516	0.0493
90/90	0.5000	0.3000	0.0500	0.0344	0.0497	0.0525	0.0344	0.0365	0.0497	0.0500	0.0497
100/100	0.5000	0.3000	0.0500	0.0348	0.0510	0.0529	0.0348	0.0373	0.0494	0.0517	0.0494



It is interesting to note that the power of Fisher's Exact Test and the *z*-test with continuity correction are consistently lower than the other tests. This occurs because the actual alpha achieved by these tests is much lower than that of the other tests. An interesting finding of this short study was that the regular *t*-test performed better than the more popular *z*-test.

Example 4 – Validation using Fleiss with Equal Sample Sizes

Fleiss (2003), page 74, presents a sample size study in which P1 = 0.7, P2 = 0.6, alpha = 0.01, and beta = 0.05. Assuming two-sided testing and equal sample allocation, Fleiss finds the necessary sample size to be 827 in each group. The calculations of Fleiss (2003) included an adjustment for continuity correction. This continuity correction is not necessary here when exact calculations are made. However, when the sample size is large enough so that approximate calculations are used, the continuity correction must be applied to obtain the same results. This is done by setting the Test Statistic to 'Z Test C.C.'. Note that this adjustment is used here to keep our results identical to those of Fleiss (2003). In practice, this adjustment is not recommended because it reduces the power and the actual alpha of the test procedure.

Setup

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

Option	Value
Data Tab	
Find (Solve For)	**N1**
Power	**0.95**
Alpha	**0.01**
N1 (Sample Size Group 1)	*Ignored since this is the Find setting*
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
D1 (Difference|H1 = P1 – P2)	**0.10**
P2 (Control Group Proportion)	**0.60**
H1 (Alternative Hypothesis)	**Two-Sided**
Test Type	**Z Test C.C. (Pooled)**
Options Tab	
Maximum N1 or N2 Exact	**100 (Set low for a rapid search.)**

Output

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

Numeric Results

Numeric Results of Tests Based on the Difference: P1 – P2
H0: P1-P2=0. H1: P1-P2<>D1. Test Statistic: Z test with continuity correction and pooled variance

Power	Sample Size Grp 1 N1	Sample Size Grp 2 N2	Prop|H1 Grp 1 or Trtmnt P1	Prop Grp 2 or Control P2	Diff if H0 D0	Diff if H1 D1	Target Alpha	Actual Alpha	Beta
0.95001	827	827	0.70000	0.60000	0.00000	0.10000	0.01000		0.04999

PASS found the required sample size to be 827 which corresponds to Fleiss.

Example 5 – Validation using Fleiss with Unequal Sample Sizes

Fleiss (2003), pages 76-77, presents a sample size study in which P1 = 0.25, P2 = 0.40, alpha = 0.01, and beta = 0.05. Assuming two-sided testing with half as many in the second group as the first, Fleiss finds the sample sizes to be 530 in the first group and 265 in the second.

Note that half as many in the second group is achieved by setting *R* to 0.5.

Setup

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

Option	Value
Data Tab	
Find (Solve For)	**N1**
Power	**0.95**
Alpha	**0.01**
N1 (Sample Size Group 1)	*Ignored since this is the Find setting*
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**0.5**
D1 (Difference|H1 = P1 – P2)	**-0.15**
P2 (Control Group Proportion)	**0.40**
H1 (Alternative Hypothesis)	**Two-Sided**
Test Type	**Z Test C.C. (Pooled)**
Options Tab	
Maximum N1 or N2 Exact	**100 (Set low for a rapid search.)**

Output

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

Numeric Results

Numeric Results of Tests Based on the Difference: P2 - P1
H0: P2-P1=0. H1: P2-P1<>D1. Test Statistic: Z test with continuity correction and pooled variance

Power	Sample Size Grp 1 N1	Sample Size Grp 2 N2	Prop|H1 Grp 1 or Trtmnt P1	Prop Grp 2 or Control P2	Diff if H0 D0	Diff if H1 D1	Target Alpha	Actual Alpha	Beta
0.95066	531	266	0.25000	0.40000	0.00000	-0.15000	0.01000		0.04934

PASS found the required sample sizes to be 531 and 266 which nearly corresponds to Fleiss's results. Fleiss computed 530 instead of 531. The number 531 is correct because the power for 530 is slightly less than the required 0.95.

Example 6 – Determining the Power after Completing an Experiment

A study has just been completed aimed at determining the effectiveness of a new treatment for cancer. Because of the cost of administering the new treatment, they would adopt the new treatment only if the difference between the proportion cured by the new treatment and that cured by the standard treatment is at least 0.10. The researchers enrolled 200 cancer patients in the study (100 for each treatment) and found that 51% were cured by the standard treatment, while 62% were cured by the new treatment. These results, however, showed no statistically significant difference based on the pooled z-test with continuity correction and alpha = 0.05. Therefore, the researchers want to compute the power of this test for detecting a difference of 0.10 for standard treatment proportions ranging from 0.40 to 0.60.

Note that the power was not exclusively computed at the observed sample proportion for the standard treatment group, 0.51. It is more informative to compute the power for a range of likely values suggested by historical evidence.

Setup

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

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power	*Ignored since this is the Find setting*
Alpha	**0.05**
N1 (Sample Size Group 1)	**100**
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
D1 (Difference|H1 = P1 – P2)	**0.10**
P2 (Control Group Proportion)	**0.40 to 0.60 by 0.04**
H1 (Alternative Hypothesis)	**Two-Sided**
Test Type	**Z Test C.C. (Pooled)**
Options Tab	
Maximum N1 or N2 Exact	**400**

Output

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

Numeric Results

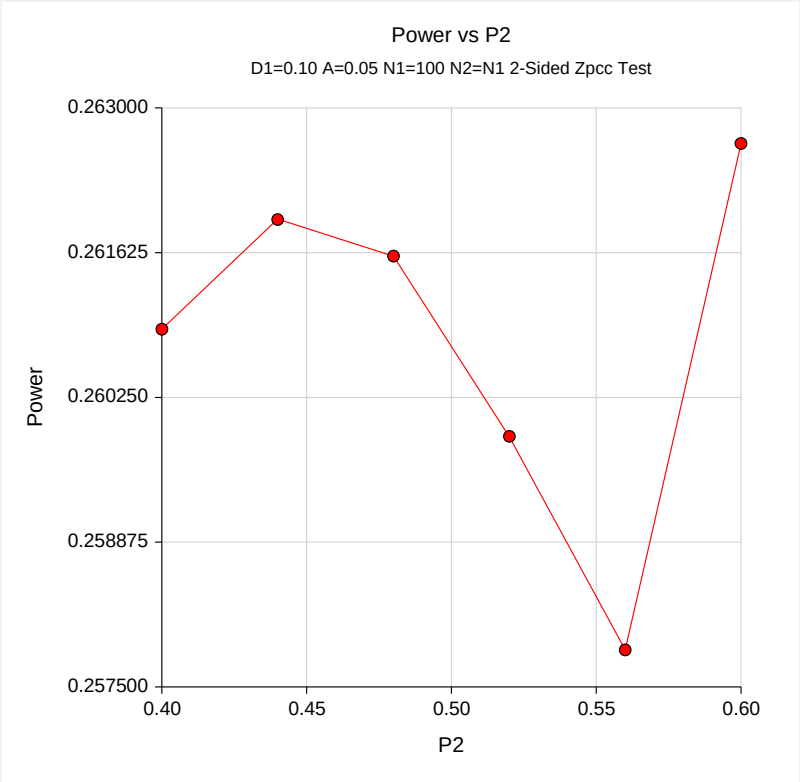
Numeric Results of Tests Based on the Difference: P1 - P2
H0: P1-P2=0. H1: P1-P2=D1<>0. Test Statistic: Z test with continuity correction and pooled variance

Power	Sample Size Grp 1 N1	Sample Size Grp 2 N2	Prop|H1 Grp 1 or Trtmnt P1	Prop Grp 2 or Control P2	Diff if H0 D0	Diff if H1 D1	Target Alpha	Actual Alpha	Beta
0.26090	100	100	0.50000	0.40000	0.00000	0.10000	0.05000	0.03628	0.73910
0.26194	100	100	0.54000	0.44000	0.00000	0.10000	0.05000	0.03861	0.73806
0.26159	100	100	0.58000	0.48000	0.00000	0.10000	0.05000	0.03988	0.73841
0.25988	100	100	0.62000	0.52000	0.00000	0.10000	0.05000	0.03988	0.74012
0.25785	100	100	0.66000	0.56000	0.00000	0.10000	0.05000	0.03861	0.74215
0.26266	100	100	0.70000	0.60000	0.00000	0.10000	0.05000	0.03628	0.73734

This report shows the values of each of the parameters, one scenario per row. The power over the entire range of the likely standard treatment proportions is relatively constant at 0.26.

The values from this table are plotted in the chart below.

Plots Section



It is evident from these results that the test performed by the researchers had very low power to detect a difference of 0.10 with the sample size used. The power is only 0.26 for a large range of standard treatment proportions. Note that the fluctuation in power is related to the value of alpha.

Example 7 – Finding the Sample Size using Ratios

Researchers would like to design an experiment to compare the infection rate of a rare disease among two populations. More specifically, they would like to determine how many subjects they need to sample from each population to determine if the disease rate in population 1 is at least three times that of population 2 with 80% power. Suppose that the researchers are confident from previous studies that the infection rate in population 2 is 0.025. The researchers plan to use the likelihood ratio test and alpha = 0.05.

Setup

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

Option	Value
Data Tab	
Find (Solve For)	**N1**
Power	**0.80**
Alpha	**0.05**
N1 (Sample Size Group 1)	*Ignored since this is the Find setting*
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
R1 (Ratio|H1 = P1 / P2)	**3**
P2 (Control Group Proportion)	**0.025**
H1 (Alternative Hypothesis)	**Two-Sided**
Test Type	**Likelihood Ratio Test**
Options Tab	
Maximum N1 or N2 Exact	**100**

Output

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

Numeric Results

Numeric Results of Tests Based on the Ratio: P1 / P2
H0: P1/P2=1. H1: P1/P2=R1<>1. Test Statistic: Likelihood Ratio test

Power	Sample Size Grp 1 N1	Sample Size Grp 2 N2	Prop|H1 Grp 1 or Trtmnt P1	Prop Grp 2 or Control P2	Ratio if H0 R0	Ratio if H1 R1	Target Alpha	Actual Alpha	Beta
0.8012	298	298	0.0750	0.0250	1.000	3.000	0.0500		0.1988

The researchers must sample 298 individuals from each population to achieve 80% power to detect a ratio of 3.0.

Chapter 201

Inequality Tests for Two Proportions in a Repeated Measures Design

Introduction

This module calculates the power for testing the time-averaged difference (TAD) between two proportions in a *repeated measures* design. A repeated measures design is one in which subjects are observed repeatedly over time. Measurements may be taken at pre-determined intervals (e.g. weekly or at specified time points following the administration of a particular treatment), or at random times with variable intervals between repeated measurements.

This type of time-averaged difference analysis is often used when the outcome to be measured varies with time. For example, suppose that you want to compare two treatment groups based on a certain binary response variable such as the presence (or absence) of a disease. The disease status may change over time, depending on various factors unrelated to the treatment. The precision of the experiment is increased by taking multiple measurements from each individual and comparing the time-averaged difference in proportions between the two groups. Care must be taken in the analysis because of the correlation that is introduced when several measurements are taken from the same individual. The covariance structure may take on several forms depending on the nature of the experiment and the subjects involved. This procedure allows you to calculate sample sizes and power using four different covariance patterns: Compound Symmetry, AR(1), Banded(1), and Simple.

This procedure can be used to calculate sample size and power for tests of pairwise contrasts in a mixed models analysis of repeated measures data. Mixed models analysis of repeated measures data is also employed to provide more flexibility in covariance specification and a greater degree of robustness in the presence of missing data, provided that the data can be assumed to be missing at random.

Two Procedures Documented Here

There are two procedures that use the program module described in this chapter. These procedures are identical except for the type of parameterization. The parameterization can be in terms of proportions or odds ratios. Each of these options is listed separately on the menus.

Technical Details

Two Test Statistics

This routine has the capability of calculating power and sample size for testing time-averaged difference in proportions based on two different test statistics. The first test statistic is presented in Liu and Wu (2005) and Diggle et al. (1994). The test statistic is based on the difference in proportions:

$$d = p_1 - p_2 ,$$

and has the form

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\text{var}(\hat{p}_1 - \hat{p}_2)}} .$$

The second type of test statistic, presented in Brown and Prescott (2006), has application to mixed models analysis of repeated measures data where there aren't any random effects other than the subjects themselves, and is based on the difference in proportions defined on the logit link scale:

$$d = \log(OR) = \log\left(\frac{p_1}{1-p_1}\right) - \log\left(\frac{p_2}{1-p_2}\right) = \text{logit}(p_1) - \text{logit}(p_2) ,$$

and has the form

$$z = \frac{\text{logit}(\hat{p}_j) - \text{logit}(\hat{p}_h)}{\sqrt{\text{var}(\text{logit}(\hat{p}_j) - \text{logit}(\hat{p}_h))}} .$$

Testing the Time-Averaged Difference between Two Proportions

Theory and Notation

The following derivation is based on the results in Liu and Wu (2005). For a study with n_1 subjects in group 1, having success proportion p_1, and n_2 subjects in group 2, having success proportion p_2 (for a total of N subjects), each measured m times, the time-averaged difference ($d = p_1 - p_2$) in proportions between the two groups can be estimated using the following model:

$$E(y_{ij} \mid x_{ij}) = \Pr(y_{ij} = 1 \mid x_{ij}) = \beta_0 + \beta_1 x_{ij}, \quad i = 1,\cdots,N; j = 1,\cdots,m ,$$

where

y_{ij} is the j^{th} binary response from subject i,

β_0 is the model intercept,

β_1 is the treatment effect or the time-averaged difference in proportions between groups 1 and 2 (i.e. $\beta_1 = d$),

x_{ij} is a binary group assignment variable, which is equal to 1 if the i^{th} subject is in group 1 and equal to 0 if the i^{th} subject is in group 2.

The proportions used to find the difference might be expressed directly as p_1 and p_2, or indirectly as p_2 and an odds ratio

$$\psi = \frac{p_1(1-p_1)}{p_2(1-p_2)} = \frac{p_1 q_2}{p_2 q_1}.$$

The proportion from group 1 can then be computed as

$$p_1 = \frac{\psi p_2}{1 - p_2 + \psi p_2}.$$

Accounting for the relationship between repeated measurements, the model presented above can be written in matrix form as

$$E(\mathbf{y}_i \mid x_i) = \mathbf{X}_i'\boldsymbol{\beta},$$

where

$\mathbf{y}_i = \begin{pmatrix} y_{i1} & y_{i2} & \cdots & y_{im} \end{pmatrix}'$ is an $m \times 1$ vector of responses from subject i,

$\mathbf{X}_i = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \end{pmatrix}_{m \times 2}$ if the i^{th} subject is in group 1,

$\mathbf{X}_i = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \end{pmatrix}_{m \times 2}$ if the i^{th} subject is in group 2, and

$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$ is the vector of model parameters.

We can stack the data in a single vector and matrix form as follows:

$$\mathbf{y} = (\mathbf{y}_1', \mathbf{y}_2', \ldots, \mathbf{y}_N')'$$
$$\mathbf{X} = (\mathbf{X}_1, '\mathbf{X}_2, '\cdots, \mathbf{X}_N')'$$

and the model for the N equations can be compressed into one as

$$E(\mathbf{y} \mid \mathbf{x}) = \mathbf{X}'\boldsymbol{\beta},$$

with

$$\begin{aligned}\mathbf{V} &= \text{var}(\mathbf{y}) \\ &= \sigma^2 \begin{pmatrix} \mathbf{R}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{R}_N \end{pmatrix} \\ &= \sigma^2 \mathbf{R}\end{aligned}$$

as the covariance (or variance-covariance) matrix.

Model Estimation

With $\hat{\mathbf{V}} = \hat{\sigma}^2 \hat{\mathbf{R}}$, then estimates of the regression coefficients from the above regression model are given as

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} \\ &= (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{y},\end{aligned}$$

and the variance of $\hat{\boldsymbol{\beta}}$ is estimated as

$$\begin{aligned}\text{var}(\hat{\boldsymbol{\beta}}) &= \begin{pmatrix} \text{var}(\hat{\beta}_0) & \text{cov}(\hat{\beta}_0, \hat{\beta}_1) \\ \text{cov}(\hat{\beta}_0, \hat{\beta}_1) & \text{var}(\hat{\beta}_1) \end{pmatrix} \\ &= (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1} \\ &= \hat{\sigma}^2 (\mathbf{X}'\hat{\mathbf{R}}^{-1}\mathbf{X})^{-1}.\end{aligned}$$

Since the data are binary, the variance term σ^2 depends on the proportions p_1 and p_2. Under the null hypothesis, H_0, the estimate of σ^2 is

$$\begin{aligned}\hat{\sigma}_0^2 &= \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}\left(1 - \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}\right) \\ &= \frac{(n_1 p_1 + n_2 p_2)(n_1 q_1 + n_2 q_2)}{(n_1 + n_2)^2},\end{aligned}$$

where $q_1 = 1 - p_1$ and $q_2 = 1 - p_2$. Under the alternative hypothesis, H_1, the estimate of σ^2 is

$$\begin{aligned}\hat{\sigma}_1^2 &= \frac{n_1}{n_1 + n_2} p_1 q_1 + \frac{n_2}{n_1 + n_2} p_2 q_2 \\ &= \frac{n_1 p_1 q_1 + n_2 p_2 q_2}{n_1 + n_2}.\end{aligned}$$

The estimated variance of $\hat{\beta}_1$ under the null hypothesis is

$$\text{var}(\hat{\beta}_1 \mid H_0) = \hat{\sigma}^2_{\hat{\beta}_1, H_0} = \hat{\sigma}_0^2 \left[(\mathbf{X}'\hat{\mathbf{R}}^{-1}\mathbf{X})^{-1}\right]_{11},$$

and the estimated variance of $\hat{\beta}_1$ under the alternative hypothesis is

$$\text{var}(\hat{\beta}_1 \mid H_1) = \hat{\sigma}^2_{\bar{\beta}_1, H_1} = \hat{\sigma}_1^2 \left[(\mathbf{X}' \hat{\mathbf{R}}^{-1} \mathbf{X})^{-1} \right]_{11},$$

where $[A]_{11}$ denotes the lower right-hand element of a 2 × 2 matrix, A.

Hypothesis Test

A two-sided test of the null hypothesis that the time-averaged difference in proportions is equal to zero is equivalent to the test of $H_0 : \beta_1 = 0$ vs. $H_1 : \beta_1 \neq 0$. Similarly, the upper and lower one-sided tests are $H_0 : \beta_1 \leq 0$ vs. $H_1 : \beta_1 > 0$ and $H_0 : \beta_1 \geq 0$ vs. $H_1 : \beta_1 < 0$, respectively. The test can be carried out using the test statistic

$$z = \frac{\hat{\beta}_1}{\sqrt{\text{var}(\hat{\beta}_1)}} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\text{var}(\hat{p}_1 - \hat{p}_2)}} \rightarrow N(0,1).$$

Power Calculations

Sample sizes for repeated measures studies are often calculated as if a simple trial with no repeated measures was planned, which results in a higher calculated sample size than would be found if the correlation between repeated measures were taken into consideration. With an idea of the correct covariance structure, and an estimate of the within-patient correlation, you can get a better estimate of the power and sample size necessary to achieve your objectives. If you have no indication of the correct covariance structure for the experiment, then the compound symmetry (program default) is likely to be adequate. If you have no previous estimate of the within-patient correlation, then Brown and Prescott (2006) suggest using a conservative prediction of the correlation, i.e. a higher correlation than anticipated.

For a two-sided test where it is assumed that $d > 0$ (without loss of generality),

$$\begin{aligned}
\text{Power} = 1 - \beta &= \Pr(\text{rejecting } H_0 \mid H_1) \\
&= \Pr\left(\left| \frac{\hat{\beta}_1}{\sqrt{\text{var}(\hat{\beta}_1)}} \right| > z_{1-\alpha/2} \mid H_1 \right) \\
&\approx \Pr\left(\frac{\hat{\beta}_1}{\sqrt{\text{var}(\hat{\beta}_1)}} > z_{1-\alpha/2} \mid H_1 \right) \text{ since it is assumed that } d > 0 \\
&= \Pr\left(\frac{\hat{\beta}_1 - d}{\hat{\sigma}_{\bar{\beta}_1, H_0}} > z_{1-\alpha/2} - \frac{d}{\hat{\sigma}_{\bar{\beta}_1, H_0}} \mid H_1 \right) \\
&= \Pr\left(\frac{\hat{\beta}_1 - d}{\hat{\sigma}_{\bar{\beta}_1, H_1}} \cdot \frac{\hat{\sigma}_{\bar{\beta}_1, H_1}}{\hat{\sigma}_{\bar{\beta}_1, H_0}} > z_{1-\alpha/2} - \frac{d}{\hat{\sigma}_{\bar{\beta}_1, H_0}} \mid H_1 \right) \\
&= \Pr\left(\frac{\hat{\beta}_1 - d}{\hat{\sigma}_{\bar{\beta}_1, H_1}} > \frac{\hat{\sigma}_{\bar{\beta}_1, H_0}}{\hat{\sigma}_{\bar{\beta}_1, H_1}} \cdot z_{1-\alpha/2} - \frac{d}{\hat{\sigma}_{\bar{\beta}_1, H_1}} \mid H_1 \right) \\
&= 1 - \Phi\left(\frac{\hat{\sigma}_{\bar{\beta}_1, H_0}}{\hat{\sigma}_{\bar{\beta}_1, H_1}} \cdot z_{1-\alpha/2} - \frac{d}{\hat{\sigma}_{\bar{\beta}_1, H_1}} \right),
\end{aligned}$$

where Φ() is the standard normal density function, and α and β are the probabilities of type I and type II error, respectively. For a one-sided test, α is used in place of $\alpha/2$.

Testing Two Proportions using the Time-Averaged Difference defined on the Logit Link Scale (Testing Pairwise Contrasts of Fixed Effects in Mixed Models)

Mixed Models Theory and Notation

The following derivation is based on the results in Brown and Prescott (2006) and Liu and Wu (2005). A generalized linear mixed model incorporates both fixed and random effects. Fixed effects are those effects in the model whose values are assumed constant, or unchanging. Random effects are those effects in the model that are assumed to have arisen from a distribution, resulting in another source of random variation other than residual variation. For an experiment with N subjects, p fixed effect parameters, and q random effect parameters, the generalized linear mixed model can be expressed using matrix notation as

$$\mathbf{y}_i = \boldsymbol{\mu}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, \cdots, N$$

where

$\mathbf{y}_i$ is an $n_i \times 1$ vector of responses for subject i,

$\boldsymbol{\mu}_i$ is an $n_i \times 1$ vector of expected means for subject i, and is linked to the model parameters by a link function, g:

$$g(\boldsymbol{\mu}_i) = \text{logit}(\boldsymbol{\mu}_i) = \begin{pmatrix} \log(\pi_i/(1-\pi_i)) \\ \log(\pi_i/(1-\pi_i)) \\ \vdots \\ \log(\pi_i/(1-\pi_i)) \end{pmatrix}_{n_i \times 1} = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{u}_i, \quad i = 1, \cdots, N$$

where

π_i is the probability of success from a bernoulli distribution for individual i,

$\mathbf{X}_i$ is an $n_i \times p$, full-rank design matrix of fixed effects for subject i,

$\boldsymbol{\beta}$ is a $p \times 1$ vector of fixed effects parameters,

$\mathbf{Z}_i$ is an $n_i \times q$ design matrix of the random effects for subject i,

$\mathbf{u}_i$ is a $q \times 1$ vector of random effects for subject i which has means of zero and scaled covariance matrix $\mathbf{G}$,

$\boldsymbol{\varepsilon}_i$ is an $n_i \times 1$ vector of errors for subject i with zero mean and scaled covariance $\boldsymbol{\Sigma}_i$.

We can stack the data in a single vector and matrix form as follows:

$$\begin{aligned}
\mathbf{y} &= (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N)' \\
\boldsymbol{\mu} &= (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_N)' \\
\mathbf{X} &= (\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_N)' \\
\mathbf{Z} &= \begin{pmatrix} \mathbf{Z}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{Z}_N \end{pmatrix} \\
\mathbf{u} &= (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N)' \\
\boldsymbol{\varepsilon} &= (\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \ldots, \boldsymbol{\varepsilon}_N)'
\end{aligned}$$

and the mixed model for the N equations can be compressed into one as

$$\mathbf{y} = \boldsymbol{\mu} + \boldsymbol{\varepsilon}$$

with

$$g(\boldsymbol{\mu}) = \text{logit}(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}\,.$$

The covariance of $\mathbf{y}$, $\text{var}(\mathbf{y}) = \mathbf{V}$, can then be written as

$$\begin{aligned}
\mathbf{V} &= \begin{pmatrix} \mathbf{V}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{V}_N \end{pmatrix} \\
&= \text{var}(\boldsymbol{\mu} + \boldsymbol{\varepsilon}) \\
&= \text{var}(\boldsymbol{\mu}) + \boldsymbol{\Sigma} \\
&\approx \mathbf{BZGZ}'\mathbf{B} + \mathbf{B}^{1/2}\mathbf{R}\mathbf{B}^{1/2},
\end{aligned}$$

where

$$\mathbf{B} = \begin{pmatrix} \pi_1(1-\pi_1) & 0 & 0 & 0 \\ 0 & \pi_2(1-\pi_2) & 0 & 0 \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & \cdots & \pi_N(1-\pi_N) \end{pmatrix}$$

R is the correlation matrix defined on the linear scale.

Mixed Models Estimation

In order to fit the generalized linear mixed model, a pseudo-variable **z** must be introduced to transform **y** onto the linear scale. More specifically,

$$\begin{aligned}
\mathbf{z} &= g(\boldsymbol{\mu}) + (\mathbf{y} - \boldsymbol{\mu})\mathbf{B}^{-1} \\
&= \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + (\mathbf{y} - \boldsymbol{\mu})\mathbf{B}^{-1}
\end{aligned}$$

and **z** has variance

$$\begin{aligned}
\mathbf{V}_\mathbf{z} &= \text{var}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}) + \mathbf{B}^{-1}\,\text{var}(\mathbf{y} - \boldsymbol{\mu})\mathbf{B}^{-1} \\
&= \mathbf{ZGZ}' + \mathbf{B}^{-1/2}\mathbf{R}\mathbf{B}^{-1/2}
\end{aligned}$$

If $\mathbf{ZGZ}' = \mathbf{0}$ (which is the case when no random effects are included in the model), then

$$\mathbf{V_z} = \mathbf{B}^{-1/2}\mathbf{R}\mathbf{B}^{-1/2}.$$

Estimates of the variance components are found using maximum likelihood (ML) or restricted/residual maximum likelihood (REML) methods. The fixed effects are then estimated as

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\hat{\mathbf{V}}_\mathbf{z}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}_\mathbf{z}^{-1}\mathbf{y}$$

with the variance estimated as

$$\text{var}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\hat{\mathbf{V}}_\mathbf{z}^{-1}\mathbf{X})^{-1}$$

These estimation equations are nearly identical to the TAD estimation equations presented earlier, except for the fact that $\boldsymbol{\beta}$ may contain more than two parameters, i.e. a parameter for each fixed effect being modeled. In the TAD model presented above, β_1 represents the difference between two treatment proportions. In the generalized mixed model formulation presented here, β_1, β_2, etc. represent individual proportions defined on the logit link scale.

Testing Fixed Effects

Significance tests for fixed or random effects can be done using tests based on the *t* distribution. We can define tests of fixed and random effects as contrasts

$$\mathbf{C} = \mathbf{L}'\hat{\boldsymbol{\beta}} = \mathbf{0},$$

respectively. For example, in a trial containing three treatments A, B, and C, a pairwise comparison of treatments A and C is given by the contrast

$$\mathbf{C}_{AC} = \mathbf{L}'\hat{\boldsymbol{\beta}} = (0 \quad 1 \quad 0 \quad -1)\hat{\boldsymbol{\beta}} = \hat{\beta}_A - \hat{\beta}_C,$$

where the first term in $\boldsymbol{\beta}$ is the intercept term, and the other three terms are the treatment effects.

For a single comparison, the Wald test statistic is given by

$$\begin{aligned} z &= \frac{\mathbf{L}'\hat{\boldsymbol{\beta}}}{\sqrt{\text{var}(\mathbf{L}'\hat{\boldsymbol{\beta}})}} \\ &= \frac{\hat{\beta}_j - \hat{\beta}_h}{\sqrt{\text{var}(\hat{\beta}_j - \hat{\beta}_h)}} \\ &= \frac{\text{logit}(\hat{p}_j) - \text{logit}(\hat{p}_h)}{\sqrt{\text{var}(\text{logit}(\hat{p}_j) - \text{logit}(\hat{p}_h))}} \rightarrow N(0,1), \end{aligned}$$

where $\hat{\beta}_j$ and $\hat{\beta}_h$ ($j \neq h$) are estimated treatment effects defined on the logit link scale and p_j and p_h are the proportions from groups *j* and *h*, respectively.

Since the data are binary, $\text{var}(\hat{\beta}_j - \hat{\beta}_h)$ depends on the proportions p_j and p_h. Under the null hypothesis, H_0, the estimate of $\text{var}(\hat{\beta}_j - \hat{\beta}_h)$ is

$$\begin{aligned}\text{var}(\hat{\beta}_j - \hat{\beta}_h \mid H_0) &= \hat{\sigma}^2_{\hat{\beta}_j - \hat{\beta}_h, H_0} \\ &= \mathbf{L}'(\mathbf{X}'\hat{\mathbf{V}}_{\mathbf{z}}^{-1}\mathbf{X})^{-1}\mathbf{L} \\ &= \frac{(n_j + n_h)^2}{(n_j p_j + n_h p_h)(n_j q_j + n_h q_h)}\mathbf{L}'(\mathbf{X}'\hat{\mathbf{R}}^{-1}\mathbf{X})^{-1}\mathbf{L},\end{aligned}$$

where $q_k = 1 - p_k$. Under the alternative hypothesis, H_1, the estimate of $\text{var}(\hat{\beta}_j - \hat{\beta}_h)$ is

$$\begin{aligned}\text{var}(\hat{\beta}_j - \hat{\beta}_h \mid H_1) &= \hat{\sigma}^2_{\hat{\beta}_j - \hat{\beta}_h, H_1} \\ &= \mathbf{L}'(\mathbf{X}'\hat{\mathbf{V}}_{\mathbf{z}}^{-1}\mathbf{X})^{-1}\mathbf{L} \\ &= \frac{n_j + n_h}{n_j p_j q_j + n_h p_h q_h}\mathbf{L}'(\mathbf{X}'\hat{\mathbf{R}}^{-1}\mathbf{X})^{-1}\mathbf{L},\end{aligned}$$

In practice, the test is often performed using software containing generalized linear models capability, such as SAS® `PROC GLIMMIX` or SAS® `PROC GENMOD` with a `REPEATED` statement. The test of the difference in proportions is generated with an estimation statement such as

```
ESTIMATE 'A-C' treat 1 0 -1;   or   LSMEANS treat/ PDIFF;
```

The latter statement would produce tests of all pairwise comparisons of the levels of the treatment variable, defined on the logit link scale. The former would only test the difference between groups A and C. Of course, these comparison statements must be used in conjunction with appropriate model and class statements.

Power Calculations

Sample sizes for repeated measures studies are often calculated as if a simple trial with no repeated measures was planned, which results in a higher calculated sample size than would be found if the correlation between repeated measures were taken into consideration. With an idea of the correct covariance structure, and an estimate of the within-patient correlation, you can get a better estimate of the power and sample size necessary to achieve your objectives. If you have no indication of the correct covariance structure for the experiment, then the compound symmetry (program default) is likely to be adequate. If you have no previous estimate of the within-patient correlation, then Brown and Prescott (2006) suggest using a conservative prediction of the correlation, i.e. a higher correlation than anticipated.

For a two-sided test where it is assumed that $\hat{\beta}_j - \hat{\beta}_h > 0$ (without loss of generality),

$$
\begin{aligned}
\text{Power} = 1-\beta &= \Pr(\text{rejecting } H_0 \mid H_1) \\
&= \Pr\left(\left| \frac{\hat{\beta}_j - \hat{\beta}_h}{\sqrt{\text{var}(\hat{\beta}_j - \hat{\beta}_h)}} \right| > z_{1-\alpha/2} \mid H_1 \right) \\
&\approx \Pr\left(\frac{\hat{\beta}_j - \hat{\beta}_h}{\sqrt{\text{var}(\hat{\beta}_j - \hat{\beta}_h)}} > z_{1-\alpha/2} \mid H_1 \right) \text{ since it is assumed that } d > 0 \\
&= \Pr\left(\frac{\hat{\beta}_1 - d}{\hat{\sigma}_{\hat{\beta}_j - \hat{\beta}_h, H_0}} > z_{1-\alpha/2} - \frac{d}{\hat{\sigma}_{\hat{\beta}_j - \hat{\beta}_h, H_0}} \mid H_1 \right) \\
&= \Pr\left(\frac{\hat{\beta}_j - \hat{\beta}_h - d}{\hat{\sigma}_{\hat{\beta}_j - \hat{\beta}_h, H_1}} \cdot \frac{\hat{\sigma}_{\hat{\beta}_j - \hat{\beta}_h, H_1}}{\hat{\sigma}_{\hat{\beta}_j - \hat{\beta}_h, H_0}} > z_{1-\alpha/2} - \frac{d}{\hat{\sigma}_{\hat{\beta}_j - \hat{\beta}_h, H_0}} \mid H_1 \right) \\
&= \Pr\left(\frac{\hat{\beta}_1 - d}{\hat{\sigma}_{\hat{\beta}_j - \hat{\beta}_h, H_1}} > \frac{\hat{\sigma}_{\hat{\beta}_j - \hat{\beta}_h, H_0}}{\hat{\sigma}_{\hat{\beta}_j - \hat{\beta}_h, H_1}} \cdot z_{1-\alpha/2} - \frac{d}{\hat{\sigma}_{\hat{\beta}_j - \hat{\beta}_h, H_1}} \mid H_1 \right) \\
&= 1 - \Phi\left(\frac{\hat{\sigma}_{\hat{\beta}_j - \hat{\beta}_h, H_0}}{\hat{\sigma}_{\hat{\beta}_j - \hat{\beta}_h, H_1}} \cdot z_{1-\alpha/2} - \frac{d}{\hat{\sigma}_{\hat{\beta}_j - \hat{\beta}_h, H_1}} \right),
\end{aligned}
$$

where Φ() is the standard normal density function, and α and β are the probabilities of type I and type II error, respectively. For a one-sided test, α is used in place of α/2.

Covariance Patterns

In a repeated measures design with *N* subjects, each measured *m* times, observations from a single subject may be correlated and a pattern for their covariance is specified. In this case, **V** will have a block-diagonal form and can be written as

$$
\mathbf{V} = \sigma^2 \mathbf{R} = \begin{pmatrix}
\mathbf{V}_1 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\
\mathbf{0} & \mathbf{V}_2 & \mathbf{0} & \cdots & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{V}_3 & \cdots & \mathbf{0} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{V}_N
\end{pmatrix}
$$

where $\mathbf{V}_i$ are $m \times m$ covariance matrices corresponding to the i^{th} subject.The **0**'s represent $m \times m$ matrices of zeros giving zero covariances for observations on different subjects. This routine allows the specification of four different covariance matrix types: Compound Symmetry, AR(1), Banded(1), and Simple.

Compound Symmetry

A compound symmetry covariance model assumes that all covariances are equal, and all variances on the diagonal are equal. That is

$$\mathbf{V}_i = \sigma^2 \begin{pmatrix} 1 & \rho & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \rho & \cdots & \rho \\ \rho & \rho & 1 & \rho & \cdots & \rho \\ \rho & \rho & \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \rho & \rho & \cdots & 1 \end{pmatrix}_{m \times m}$$

where $\sigma^2 = \text{var}(y_{ij})$ and ρ is the correlation between observations on the same subject.

AR(1)

An AR(1) (autoregressive order 1) covariance model assumes that all variances on the diagonal are equal and that covariances *t* time periods apart are equal to $\sigma^2 \rho^t$. That is

$$\mathbf{V}_i = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \ldots & \rho^{m-1} \\ \rho & 1 & \rho & \rho^2 & \ldots & \rho^{m-2} \\ \rho^2 & \rho & 1 & \rho & \ldots & \rho^{m-3} \\ \rho^3 & \rho^2 & \rho & 1 & \ldots & \rho^{m-4} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{m-1} & \rho^{m-2} & \rho^{m-3} & \rho^{m-4} & \ldots & 1 \end{pmatrix}_{m \times m}$$

where $\sigma^2 = \text{var}(y_{ij})$ and ρ is the correlation between observations on the same subject.

Banded(1)

A Banded(1) (banded order 1) covariance model assumes that all variances on the diagonal are equal, covariances for observations one time period apart are equal to $\sigma^2 \rho$, and covariances for measurements greater than one time period apart are equal to zero. That is

$$\mathbf{V}_i = \sigma^2 \begin{pmatrix} 1 & \rho & 0 & 0 & \cdots & 0 \\ \rho & 1 & \rho & 0 & \cdots & 0 \\ 0 & \rho & 1 & \rho & \cdots & 0 \\ 0 & 0 & \rho & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{pmatrix}_{m \times m}$$

where $\sigma^2 = \text{var}(y_{ij})$ and ρ is the correlation between observations on the same subject.

Simple

A simple covariance model assumes that all variances on the diagonal are equal and that all covariances are equal to zero. That is

$$\mathbf{V}_i = \sigma^2 \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{pmatrix}_{m \times m}$$

where $\sigma^2 = \text{var}(y_{ij})$.

Procedure Options

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

Data Tab (Common Options)

The Data tab contains the parameters associated with this test such as the proportions, sample sizes, alpha, and power. This chapter covers two procedures, each of which has different options. This section documents options that are common to both procedures. Later, unique options for each procedure will be documented.

Solve For

Find (Solve For)

This option specifies the parameter to be solved for. 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. The "solve for" parameter is displayed on the vertical axis of the plot.

Error Rates

Power or Beta

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

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

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

Alpha (Significance Level)

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

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

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

Sample Size

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 the operator [*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*.

Effect Size – Group 2 Proportion

P2 (Group 2 Proportion)

Enter a value for P2, the baseline proportion, or the proportion of "successes" from group 2. You may enter a single value or a range of values such as *0.1 0.2 0.3* or *0.1 to 0.3 by 0.05*. The items in the list may be separated with commas or blanks.

Effect Size – Repeated Measurements

M (Number of Time Points)

Enter a value for the number of time points (repeated measurements) for which each subject will be observed. You may enter a single value or a range of values such as '3 5 7' or '2 to 8 by 1'. The items in the list may be separated with commas or blanks.

Effect Size – Covariance Structure

Covariance Type

Select the within-subject covariance structure that will be used in the mixed models analysis. The options are:

- **Compound Symmetry**
  All variances on the diagonal of the within-subject variance-covariance matrix are equal to σ^2, and all covariances are equal to $\rho\sigma^2$.

- **AR(1)**
  All variances on the diagonal of the within-subject variance-covariance matrix are equal to σ^2, and the covariance between observations *t* time periods apart is $\rho^t\sigma^2$.

- **Banded(1)**
  All variances on the diagonal of the within-subject variance-covariance matrix are equal to σ^2, and the covariance between observations one time period apart is $\rho\sigma^2$. Covariances between observations more than one time period apart are equal to zero.

- **Simple**
  All variances are equal to σ^2, and all covariances are equal to zero.

Rho (Autocorrelation)

Enter a value for the correlation between observations on the same subject. When no previous estimate of the within-patient correlation is available, you should use a conservative prediction of the correlation, i.e. a correlation that is higher than anticipated. You may enter a single value or a range of values such as *0.5 0.6 0.7* or *0.4 to 0.9 by 0.1*. The items in the list may be separated with commas or blanks.

Test

Test Statistic Based On

This option specifies the type of test statistic for which power is calculated. This routine has the capability of calculating power and sample size for testing time-averaged differences for binary data based on two different test statistics. The options are as follows:

- **Difference: P1-P2**
  This option calculates the power for the test statistic based on the difference in proportions:

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\text{var}(\hat{p}_1 - \hat{p}_2)}}.$$

- **Log(OR): logit(P1)-logit(P2)**
  This option calculates the power for the test statistic based on the difference defined on the logit link scale:

$$z = \frac{\text{logit}(\hat{p}_j) - \text{logit}(\hat{p}_h)}{\sqrt{\text{var}(\text{logit}(\hat{p}_j) - \text{logit}(\hat{p}_h))}}.$$

Alternative Hypothesis

This option specifies the alternative hypothesis. This implicitly specifies the direction of the hypothesis test. The null hypothesis is always $H_0 : \beta_1 = 0$.

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:

- **One-Sided**

  This option yields a *one-tailed* test. Use it for testing the alternative hypotheses $H_1 : \beta_1 > 0$ or $H_1 : \beta_1 < 0$.

- **Two-Sided**

  This is the most common selection. It yields the *two-tailed* test. Use this option when you are testing whether the means are different, but you do not want to specify beforehand which mean is larger.

Data Tab (Odds Ratios)

Effect Size – Odds Ratio to Detect

OR (Odds Ratio|H1)

Enter a value for the odds ratio for that is to be detected. If you know the expected proportions for the two treatment groups but do not know the odds ratio, click on the "OR" button to the right to calculate the corresponding odds ratio. The odds ratio is used along with P2 to calculate P1 using the formula: P1 = (OR*P2)/(1 - P2 + OR*P2). You may enter a single value or a range of values such as *1.5 2 2.5* or *1.5 to 2 by 0.1*. The items in the list may be separated with commas or blanks.

Data Tab (Proportions)

Effect Size – Proportions

P1 (Group 1 Proportion|H1)

Enter a value for the proportion of 'successes' in group 1 under the alternative hypothesis. If you know the expected odds ratio for the two treatment groups, click on the 'P' button to the right to calculate the corresponding proportions. You may enter a single value or a range of values such as *0.1 0.2 0.3* or *0.1 to 0.5 by 0.1*. The items in the list may be separated with commas or blanks.

Example 1 – Determining Power

A study is being planned to determine the efficacy of a prophylactic treatment for the common cold. The study will follow a treatment group and placebo control group through the winter to determine if there is an overall difference between the two treatment groups in the proportion of patients who get sick. Subjects will take the treatment (or placebo) once daily throughout the duration of the study. The study will be conducted from September to April with scheduled, monthly visits (beginning in October) to determine the patient's disease status (present or absent). Therefore, a total of seven responses will be observed for each patient. Previous studies have indicated a baseline disease rate of 60% for the common cold. The researchers want to be able to detect a treatment to control odds ratio of 0.5 (an odds ratio of 0.5 corresponds to a treatment group proportion of 0.4285714). A compound-symmetry covariance pattern with autocorrelation of 0.5 is assumed to be adequate. The test will be conducted using a mixed models analysis with an alpha level of 0.05.

What power does the study achieve over a range of possible sample sizes?

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Two Proportions in a Repeated Measures Design [Odds Ratios]** or **Tests for Two Proportions in a Repeated Measures Design [Proportions]** procedure window by expanding **Proportions**, then **Two Independent Proportions**, then clicking on **Repeated Measures**, and then clicking on **Tests for Two Proportions in a Repeated Measures Design [Odds Ratios]** or **Tests for Two Proportions in a Repeated Measures Design [Proportions]**. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find	**Power and Beta**
Power	*Ignored since this is the Find setting*
Alpha	**0.05**
N1	**10 to 100 by 10**
N2	**Use R**
R	**1.0**
OR1 (*If using Odds Ratios*)	**0.5**
P1 (*If using Proportions*)	**0.4285714**
P2	**0.6**
M	**7**
Covariance Type	**Compound Symmetry**
Rho	**0.5**
Test Statistic Based on	**Log(OR): logit(P1)-logit(P2)**
Alternative Hypothesis	**Two-Sided**

Annotated Output

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

Numeric Results

Numeric Results
Test Statistic Based on Log(OR): logit(P1) - logit(P2).
Two-Sided Test. Null Hypothesis: OR = 1. Alternative Hypothesis: OR <> 1.
Covariance Type = Compound Symmetry

Power	Group 1 Sample Size (N1)	Group 2 Sample Size (N2)	Time Points (M)	Group 1 Prop. (P1)	Group 2 Prop. (P2)	Odds Ratio (OR1)	Auto Corr. (Rho)	Alpha	Beta
0.17843	10	10	7	0.429	0.600	0.500	0.500	0.050	0.82157
0.30742	20	20	7	0.429	0.600	0.500	0.500	0.050	0.69258
0.42768	30	30	7	0.429	0.600	0.500	0.500	0.050	0.57232
0.53515	40	40	7	0.429	0.600	0.500	0.500	0.050	0.46485
0.62800	50	50	7	0.429	0.600	0.500	0.500	0.050	0.37200
0.70610	60	60	7	0.429	0.600	0.500	0.500	0.050	0.29390
0.77040	70	70	7	0.429	0.600	0.500	0.500	0.050	0.22960
0.82241	80	80	7	0.429	0.600	0.500	0.500	0.050	0.17759
0.86386	90	90	7	0.429	0.600	0.500	0.500	0.050	0.13614
0.89646	100	100	7	0.429	0.600	0.500	0.500	0.050	0.10354

References
Brown, H., Prescott, R., 2006. Applied Mixed Models in Medicine. 2nd ed. John Wiley & Sons Ltd. Chichester, West Sussex, England.
Liu, H. and Wu, T., 2005. 'Sample Size Calculation and Power Analysis of Time-Averaged Difference.' Journal of Modern Applied Statistical Methods, Vol. 4, No. 2, pages 434-445.
Diggle, P.J., Liang, K.Y., and Zeger, S.L., 1994. Analysis of Longitudinal Data. Oxford University Press. New York, New York. Chapter 2.

Report Definitions
Power is the probability of rejecting a false null hypothesis. It should be close to one.
N1 and N2 are the number of subjects in groups 1 and 2, respectively.
M is the number of time points (repeated measurements) at which each subject is observed.
P1 and P2 are the proportions from groups 1 and 2, respectively.
OR1 is the odds ratio ((P1/(1-P1))/(P2/(1-P2))) to be detected.
Rho is the correlation between observations on the same subject.
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 10 and 10 achieve 18% power to detect an odds ratio of 0.500 in a design with 7 repeated measurements having a Compound Symmetry covariance structure when the proportion from group 2 is 0.600, the correlation between observations on the same subject is 0.500, and the alpha level is 0.050.

This report gives the power for each value of the other parameters.

Power

This is the computed power for detecting the time-averaged difference between the two group means.

Group 1 Sample Size (N1)

The value of *N1* is the number of subjects in group 1.

Group 2 Sample Size (N2)

The value of *N2* is the number of subjects in group 2.

Time Points (M)

This is the number of repeated measurements taken.

Group 1 Prop (P1) & Group 2 Prop (P2)

These are the proportions of successes in groups 1 and 2, respectively.

Odds Ratio (OR1)

This is the value of the odds ratio under the alternative hypothesis.

Autocorr. (Rho)

This is the correlation between observations from the same subject.

Alpha

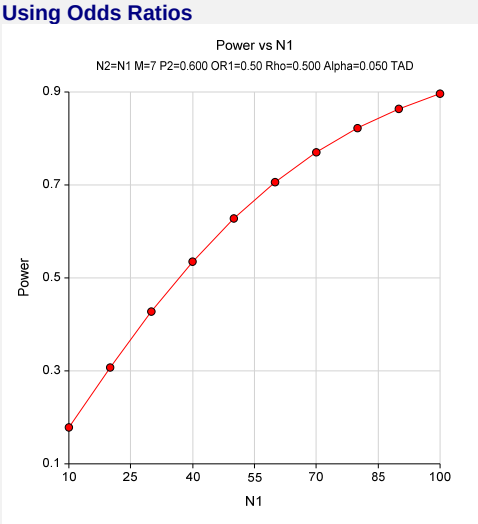
Alpha is the significance level of the test.

Beta

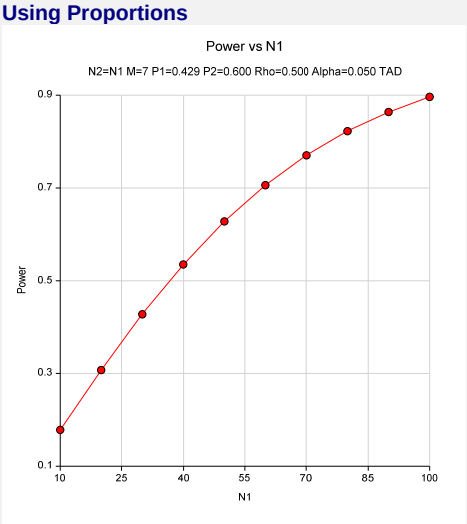
Beta is the probability of failing to reject the null hypothesis when the alternative hypothesis is true.

Plots Section

Using Odds Ratios



Using Proportions



The chart shows the relationship between power and *N1* when the other parameters in the design are held constant.

Example 2 – Finding the Sample Size

Continuing with Example 1, the researchers want to determine the exact sample size necessary to achieve at least 80% power.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Two Proportions in a Repeated Measures Design [Odds Ratios]** or **Tests for Two Proportions in a Repeated Measures Design [Proportions]** procedure window by expanding **Proportions**, then **Two Independent Proportions**, then clicking on **Repeated Measures**, and then clicking on **Tests for Two Proportions in a Repeated Measures Design [Odds Ratios]** or **Tests for Two Proportions in a Repeated Measures Design [Proportions]**. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find	**N1**
Power	**0.8**
Alpha	**0.05**
N1	*Ignored since this is the Find setting*
N2	**Use R**
R	**1.0**
OR1 (*If using Odds Ratios*)	**0.5**
P1 (*If using Proportions*)	**0.4285714**
P2	**0.6**
M	**7**
Covariance Type	**Compound Symmetry**
Rho	**0.5**
Test Statistic Based on	**Log(OR): logit(P1)-logit(P2)**
Alternative Hypothesis	**Two-Sided**

Output

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

Numeric Results

Power	Group 1 Sample Size (N1)	Group 2 Sample Size (N2)	Time Points (M)	Group 1 Prop. (P1)	Group 2 Prop. (P2)	Odds Ratio (OR1)	Auto Corr. (Rho)	Alpha	Beta
0.80297	76	76	7	0.429	0.600	0.500	0.500	0.050	0.19703

A group sample size of 76 is required to achieve at least 80% power.

Example 3 – Varying the Odds Ratio

Continuing with Examples 1 and 2, the researchers want to evaluate the impact on power of varying the odds ratio from 0.4 to 0.8. 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

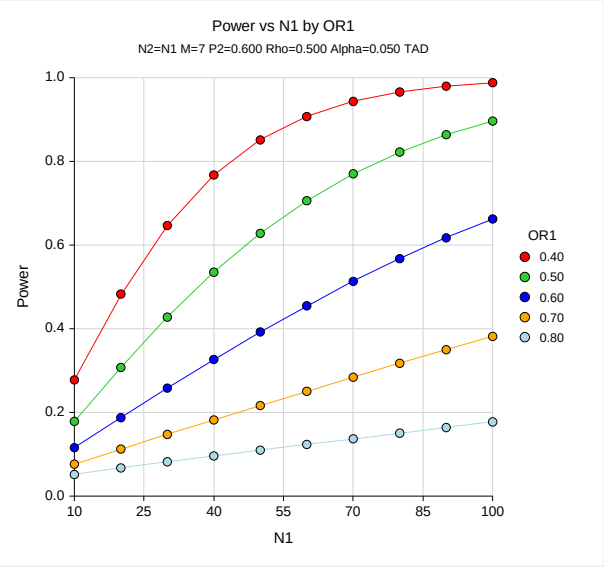
This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Two Proportions in a Repeated Measures Design [Odds Ratios]** procedure window by expanding **Proportions**, then **Two Independent Proportions**, then clicking on **Repeated Measures**, and then clicking on **Tests for Two Proportions in a Repeated Measures Design [Odds Ratios]**. You may then make the appropriate entries as listed below, or open **Example 3** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find	**Power and Beta**
Power	*Ignored since this is the Find setting*
Alpha	**0.05**
N1	**10 to 100 by 10**
N2	**Use R**
R	**1.0**
OR1	**0.4 to 0.8 by 0.1**
P2	**0.6**
M	**7**
Covariance Type	**Compound Symmetry**
Rho	**0.5**
Test Statistic Based on	**Log(OR): logit(P1)-logit(P2)**
Alternative Hypothesis	**Two-Sided**

Output

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

Plots Section



This chart shows how the power depends on odds ratio, OR1, as well as the group sample size N1.

Example 4 – Varying the Proportions

Continuing with Examples 1 and 2, the researchers want to evaluate the impact on power of varying the group 1 proportion from 0.2 to 0.5. 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

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Two Proportions in a Repeated Measures Design [Proportions]** procedure window by expanding **Proportions**, then **Two Independent Proportions**, then clicking on **Repeated Measures**, and then clicking on **Tests for Two Proportions in a Repeated Measures Design [Proportions]**. You may then make the appropriate entries as listed below, or open **Example 4** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find	**Power and Beta**
Power	*Ignored since this is the Find setting*
Alpha	**0.05**
N1	**10 to 100 by 10**

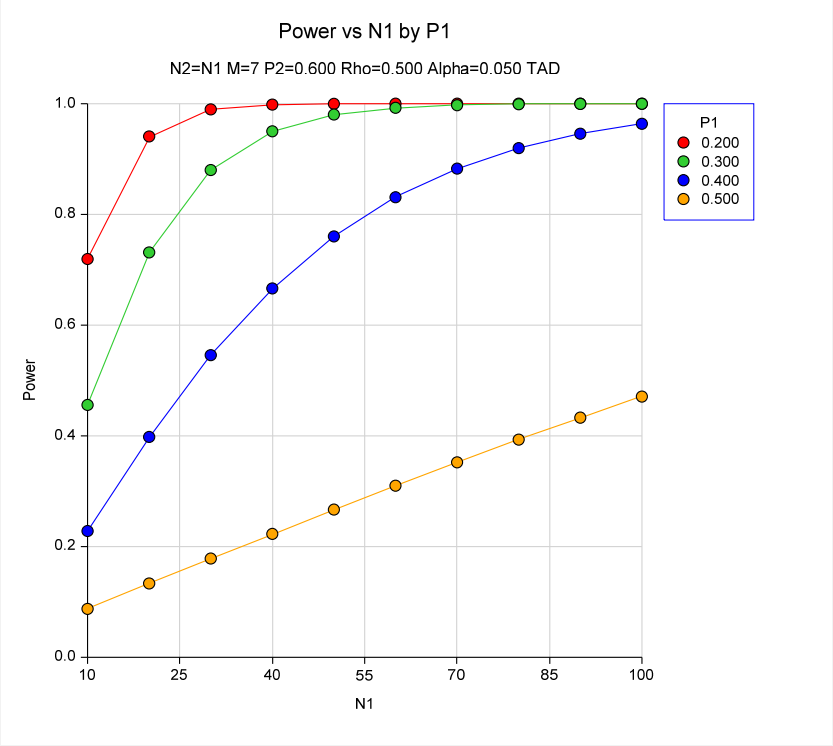
Data Tab (continued)

N2 ..... **Use R**
R **1.0**
P1 ..... **0.2 to 0.5 by 0.1**
P2 ..... **0.6**
M **7**
Covariance Type ..... **Compound Symmetry**
Rho **0.5**
Test Statistic Based on ..... **Log(OR): logit(P1)-logit(P2)**
Alternative Hypothesis ..... **Two-Sided**

Output

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

Plots Section



This chart shows how the power depends on the proportion, P1, as well as the group sample size N1.

Example 5 – Impact of the Number of Repeated Measurements

Continuing with Example 2, the researchers want to study the impact on the sample size if they changing the number of measurements made on each individual. Their experimental protocol calls for seven measurements. They want to see the impact of taking twice that many measurements.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Two Proportions in a Repeated Measures Design [Odds Ratios]** or **Tests for Two Proportions in a Repeated Measures Design [Proportions]** procedure window by expanding **Proportions**, then **Two Independent Proportions**, then clicking on **Repeated Measures**, and then clicking on **Tests for Two Proportions in a Repeated Measures Design [Odds Ratios]** or **Tests for Two Proportions in a Repeated Measures Design [Proportions]**. You may then make the appropriate entries as listed below, or open **Example 5** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find	**N1**
Power	**0.8**
Alpha	**0.05**
N1	*Ignored since this is the Find setting*
N2	**Use R**
R	**1.0**
OR1 (*If using Odds Ratios*)	**0.5**
P1 (*If using Proportions*)	**0.4285714**
P2	**0.6**
M	**7 14**
Covariance Type	**Compound Symmetry**
Rho	**0.5**
Test Statistic Based on	**Log(OR): logit(P1)-logit(P2)**
Alternative Hypothesis	**Two-Sided**

Output

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

Numeric Results

Power	Group 1 Sample Size (N1)	Group 2 Sample Size (N2)	Time Points (M)	Group 1 Prop. (P1)	Group 2 Prop. (P2)	Odds Ratio (OR1)	Auto Corr. (Rho)	Alpha	Beta
0.80297	76	76	7	0.429	0.600	0.500	0.500	0.050	0.19703
0.80161	71	71	14	0.429	0.600	0.500	0.500	0.050	0.19839

Doubling the number of repeated measurements per individual decreases the group sample size by 5. This reduction in sample size may not justify the additional seven trips to the clinic for each subject.

Example 6 – Validation using Diggle et al. (1994)

Diggle et al. (1994) pages 31 and 32 present an example of calculating the sample size for a TAD study. They calculate the group sample sizes for the cases where the difference in proportions (P1-P2) ranges from 0.1 to 0.3, ρ ranges from 0.2 to 0.8, alpha = 0.05, $p_2 = 0.5$, $M = 3$, and power = 0.8. Note that Diggle et al (1994) uses a one-sided test and the test statistic based on the difference in proportions.

To calculate the sample sizes using the odds ratio specification, we must first convert the differences to odds ratios using the formula:

$$OR = \frac{p_1/(1-p_1)}{p_2/(1-p_2)}$$

Differences of 0.1, 0.2, and 0.3 with P2 = 0.5 correspond to odds ratios of 1.5, 2.333, and 4.0, respectively.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Two Proportions in a Repeated Measures Design [Odds Ratios]** or **Tests for Two Proportions in a Repeated Measures Design [Proportions]** procedure window by expanding **Proportions**, then **Two Independent Proportions**, then clicking on **Repeated Measures**, and then clicking on **Tests for Two Proportions in a Repeated Measures Design [Odds Ratios]** or **Tests for Two Proportions in a Repeated Measures Design [Proportions]**. You may then make the appropriate entries as listed below, or open **Example 6** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find	**N1**
Power	**0.8**
Alpha	**0.05**
N1	*Ignored since this is the Find setting*
N2	**Use R**
R	**1.0**
OR1 (*If using Odds Ratios*)	**1.5 2.333 4**
P1 (*If using Proportions*)	**0.6 0.7 0.8**
P2	**0.5**
M	**3**
Covariance Type	**Compound Symmetry**
Rho	**0.2 0.5 0.8**
Test Statistic Based on	**Difference: P1-P2**
Alternative Hypothesis	**One-Sided**

Output

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

Numeric Results

Power	Group 1 Sample Size (N1)	Group 2 Sample Size (N2)	Time Points (M)	Group 1 Prop. (P1)	Group 2 Prop. (P2)	Odds Ratio (OR1)	Auto Corr. (Rho)	Alpha	Beta
0.80164	143	143	3	0.600	0.500	1.500	0.200	0.050	0.19836
0.80116	204	204	3	0.600	0.500	1.500	0.500	0.050	0.19884
0.80089	265	265	3	0.600	0.500	1.500	0.800	0.050	0.19911
0.80870	35	35	3	0.700	0.500	2.333	0.200	0.050	0.19130
0.80163	49	49	3	0.700	0.500	2.333	0.500	0.050	0.19837
0.80329	64	64	3	0.700	0.500	2.333	0.800	0.050	0.19671
0.82213	15	15	3	0.800	0.500	4.000	0.200	0.050	0.17787
0.81509	21	21	3	0.800	0.500	4.000	0.500	0.050	0.18491
0.81120	27	27	3	0.800	0.500	4.000	0.800	0.050	0.18880

The sample sizes calculated by ***PASS*** match the results of Diggle et al. (1994) exactly.

Example 7 – Validation using Brown and Prescott (2006)

Brown and Prescott (2006) page 270 presents an example of calculating the sample size for a future study. They calculate the group sample size to be 85 for a future study involving four post-treatment visits to detect a doubling of the odds ratio (i.e., OR1 = 2) at the 5% significance level with 80% power. They assume an autocorrelation of 0.5, and an expected rate of positives ((P1+P2)/2) of 0.4.

We can calculate the corresponding values of P1, P2, and OR for use in ***PASS*** by solving the following system of equations for P1 and P2:

$$\frac{P1+P2}{2}=0.4 \quad \text{and} \quad \frac{P1/(1-P1)}{P2/(1-P2)}=OR=2.0$$

The solution to these equations occurs when P1 = 0.482255312124 and P2 = 0.317744687876. The decimal places are kept to make the solution exact.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Two Proportions in a Repeated Measures Design [Odds Ratios]** or **Tests for Two Proportions in a Repeated Measures Design [Proportions]** procedure window by expanding **Proportions**, then **Two Independent Proportions**, then clicking on **Repeated Measures**, and then clicking on **Tests for Two Proportions in a Repeated Measures Design [Odds Ratios]** or **Tests for Two Proportions in a Repeated Measures Design [Proportions]**. You may then make the appropriate entries as listed below, or open **Example 7** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find	**N1**
Power	**0.8**

Data Tab (continued)

Alpha **0.05**
N1 ..... *Ignored since this is the Find setting*
N2 ..... **Use R**
R **1.0**
OR1 (*If using Odds Ratios*) **2.0**
P1 (*If using Proportions*) **0.482255312124**
P2 ..... **0.317744687876**
M **4**
Covariance Type ..... **Compound Symmetry**
Rho **0.5**
Test Statistic Based on ..... **Log(OR): logit(P1)-logit(P2)**
Alternative Hypothesis ..... **Two-Sided**

Output

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

Numeric Results

Power	Group 1 Sample Size (N1)	Group 2 Sample Size (N2)	Time Points (M)	Group 1 Prop. (P1)	Group 2 Prop. (P2)	Odds Ratio (OR1)	Auto Corr. (Rho)	Alpha	Beta
0.80080	86	86	4	0.482	0.318	2.000	0.500	0.050	0.19920

The sample size of 86 calculated by ***PASS*** matches the results of Brown and Prescott (2006). The slight difference is due to rounding. Calculation of the sample size presented by Brown and Prescott (2006) on page 270 results in a value of 85.025337, which they round down to 85. Note that the numerical formula has a typographical error: the denominator term should be $(4 \times .693^2)$, not $(4 \times .693)^2$ (see the formula on page 269 of Brown and Prescott (2006)).

Chapter 205

Tests for Two Proportions with Non-Zero Null Hypothesis

Introduction

This module computes power and sample size for hypothesis tests of the difference, ratio, or odds ratio of two independent proportions. The word 'offset' in the chapter title indicates that the difference being tested in the null hypothesis is not zero (or that the ratio is not one). The *non-offset* case is available in another procedure. This procedure compares the power achieved by each of several test statistics.

The power calculations assume that independent, random samples are drawn from two populations.

Four Procedures Documented Here

There are four procedures in the menus that use the program module described in this chapter. These procedures are identical except for the type of parameterization. The parameterization can be in terms of proportions, differences in proportions, ratios of proportions, and odds ratios. Each of these options is listed separately on the menus.

Technical Details

Suppose you have two populations from which dichotomous (binary) responses will be recorded. The probability (or risk) of obtaining the event of interest in population 1 (the treatment group) is p_1 and in population 2 (the control group) is p_2. The corresponding failure proportions are given by $q_1 = 1 - p_1$ and $q_2 = 1 - p_2$.

An assumption is made that the responses from each group follow a binomial distribution. This means that the event probability, p_i, is the same for all subjects within the group and that the response from one subject is independent of that of any other subject.

Random samples of *m* and *n* individuals are obtained from these two populations. The data from these samples can be displayed in a 2-by-2 contingency table as follows

Group	Success	Failure	Total
Treatment	*a*	*c*	*m*
Control	*b*	*d*	*n*
Total	*s*	*f*	*N*

The following alternative notation is sometimes used.

Group	Success	Failure	Total
Treatment	x_{11}	x_{12}	n_1
Control	x_{21}	x_{22}	n_2
Total	m_1	m_2	N

The binomial proportions, p_1 and p_2, are estimated from these data using the formulae

$$\hat{p}_1 = \frac{a}{m} = \frac{x_{11}}{n_1} \text{ and } \hat{p}_2 = \frac{b}{n} = \frac{x_{21}}{n_2}$$

Comparing Two Proportions

When analyzing studies such as this, you usually want to compare the two binomial probabilities, p_1 and p_2. The most direct method of comparing these quantities is to calculate their difference or their ratio. If the binomial probability is expressed in terms of odds rather than probability, another measure is the odds ratio. Mathematically, these comparison parameters are

Parameter	**Computation**
Difference	$\delta = p_1 - p_2$
Risk Ratio	$\phi = p_1 / p_2$
Odds Ratio	$\psi = \dfrac{p_1 / (1 - p_1)}{p_2 / (1 - p_2)} = \dfrac{p_1 q_2}{p_2 q_1}$

The choice of which of these measures is used might seem arbitrary, but it is not. Not only will the interpretation be different, but, for small sample sizes, the powers of tests based on different parameters will be different. The non-null case is commonly used in equivalence and non-inferiority testing.

Difference

The (risk) difference, $\delta = p_1 - p_2$,is perhaps the most direct method of comparison between the two event probabilities. This parameter is easy to interpret and communicate. It gives the absolute impact of the treatment. However, there are subtle difficulties that can arise with its interpretation.

One interpretation difficulty occurs when the event of interest is rare. If a difference of 0.001 were reported for an event with a baseline probability of 0.40, we would probably dismiss this as being of little importance. That is, there usually little interest in a treatment that decreases the probability from 0.400 to 0.399. However, if the baseline probability of a disease was 0.002 and 0.001 was the decrease in the disease probability, this would represent a reduction of 50%. Thus we see that interpretation depends on the baseline probability of the event.

A similar situation occurs when the amount of possible difference is considered. Consider two events, one with a baseline event rate of 0.40 and the other with a rate of 0.02. What is the maximum decrease that can occur? Obviously, the first event rate can be decreased by an absolute amount of 0.40, while the second can only be decreased by a maximum of 0.02.

So, although creating the simple difference is a useful method of comparison, care must be taken that it is appropriate for the situation.

Ratio

The (risk) ratio, $\phi = p_1 / p_2$, gives the relative change in the disease risk due to the application of the treatment. This parameter is also direct and easy to interpret. To compare this with the difference, consider a treatment that reduces the risk of disease from 0.1437 to 0.0793. One should consider which single number is more enlightening, the fact that the absolute risk of disease has been decreased by 0.0644, or the fact that risk of disease in the treatment group is only 55.18% of that in the control group. In many cases, the percentage (risk ratio) communicates the impact of the treatment better than the absolute change.

Perhaps the biggest drawback to this parameter is that it cannot be calculated in one of the most common experimental designs, the case-control study. Another drawback is that the odds ratio occurs directly in the likelihood equations and as a parameter in logistic regression.

Odds Ratio

Chances are usually communicated as long-term proportions or probabilities. In betting, chances are often given as odds. For example, the odds of a horse winning a race might be set at 10-to-1 or 3-to-2. Odds can easily be translated into probability. An odds of 3-to-2 means that the event is expected to occur three out of five times. That is, an odds of 3-to-2 (1.5) translates to a probability of winning of 0.60.

The odds of an event are calculated by dividing the event risk by the non-event risk. Thus, in our case of two populations, the odds are

$$o_1 = \frac{p_1}{1 - p_1} \text{ and } o_2 = \frac{p_2}{1 - p_2}$$

For example, if p_1 is 0.60, the odds are 0.60/0.40 = 1.5. Rather than represent the odds as a decimal amount, it is re-scaled into whole numbers. Instead of saying the odds are 1.5-to-1, we say they are 3-to-2.

Another way to compare proportions is to compute the ratio of their odds. The odds ratio of two events is

$$\psi = \frac{o_1}{o_2} = \frac{p_1 / (1 - p_1)}{p_2 / (1 - p_2)} = \frac{p_1 q_2}{p_2 q_1}$$

Although the odds ratio is more complicated to interpret than the risk ratio, it is often the parameter of choice. One reason for this is the fact that the odds ratio can be accurately estimated from case-control studies, while the risk ratio cannot. Also, the odds ratio is the basis of logistic regression (used to study the influence of risk factors). Furthermore, the odds ratio is the natural parameter in the conditional likelihood of the two-group, binomial-response design. Finally, when the baseline event rates are rare, the odds ratio provides a close approximation to the risk ratio since, in this case, $1 - p_1 \approx 1 - p_2$, so that

$$\psi = \frac{o_1}{o_2} = \frac{p_1 / (1 - p_1)}{p_2 / (1 - p_2)} \approx \frac{p_1}{p_2} = \phi$$

Hypothesis Tests

Although several statistical tests are available for testing the inequality of two proportions, only a few can be generalized to the non-null case. No single test is the champion in every situation, so one should compare the powers of the various tests to determine which to use.

Difference

The (risk) difference, $\delta = p_1 - p_2$, is perhaps the most direct method for comparing two proportions. Three sets of statistical hypotheses can be formulated:

1. $H_0: p_1 - p_2 = \delta_0$ versus $H_1: p_1 - p_2 \neq \delta_0$; this is often called the *two-tailed test*.
2. $H_0: p_1 - p_2 \leq \delta_0$ versus $H_1: p_1 - p_2 > \delta_0$; this is often called the *upper-tailed test*.
3. $H_0: p_1 - p_2 \geq \delta_0$ versus $H_1: p_1 - p_2 < \delta_0$; this is often called the *lower-tailed test*.

Ratio

The (risk) ratio, $\phi = p_1 / p_2$, is often preferred as a comparison parameter because it expresses the difference as a percentage rather than an amount. Three sets of statistical hypotheses can be formulated:

1. $H_0: p_1 / p_2 = \phi_0$ versus $H_1: p_1 / p_2 \neq \phi_0$; this is often called the *two-tailed test*.
2. $H_0: p_1 / p_2 \leq \phi_0$ versus $H_1: p_1 / p_2 > \phi_0$; this is often called the *upper-tailed test*.
3. $H_0: p_1 / p_2 \geq \phi_0$ versus $H_1: p_1 / p_2 < \phi_0$; this is often called the *lower-tailed test*.

Odds Ratio

The odds ratio, $\psi = \left[p_1 / (1 - p_1) \right] / \left[p_2 / (1 - p_2) \right]$, is sometimes used as the comparison because of its statistical properties and because some convenient experimental designs only allow it to be estimated. Three sets of statistical hypotheses can be formulated:

1. $H_0: \psi = \psi_0$ versus $H_1: \psi \neq \psi_0$; this is often called the *two-tailed test*.
2. $H_0: \psi \leq \psi_0$ versus $H_1: \psi > \psi_0$; this is often called the *upper-tailed test*.
3. $H_0: \psi \geq \psi_0$ versus $H_1: \psi < \psi_0$; this is often called the *lower-tailed test*.

Power Calculation

The power for a test statistic that is based on the normal approximation can be computed exactly using two binomial distributions. The following steps are taken to compute the power of such a test.

1. Find the critical value (or values in the case of a two-sided test) using the standard normal distribution. The critical value, $z_{critical}$, is that value of *z* that leaves exactly the target value of alpha in the appropriate tail of the normal distribution. For example, for an upper-tailed test with a target alpha of 0.05, the critical value is 1.645.
2. Compute the value of the test statistic, z_t, for every combination of x_{11} and x_{21}. Note that x_{11} ranges from 0 to n_1, and x_{21} ranges from 0 to n_2. A small value (around 0.0001) can be added to the zero cell counts to avoid numerical problems that occur when the cell value is zero.
3. If $z_t > z_{critical}$, the combination is in the rejection region. Call all combinations of x_{11} and x_{21} that lead to a rejection the set *A*.
4. Compute the power for given values of p_1 and p_2 as

$$1 - \beta = \sum_A \binom{n_1}{x_{11}} p_1^{x_{11}} q_1^{n_1 - x_{11}} \binom{n_2}{x_{21}} p_2^{x_{21}} q_2^{n_2 - x_{21}}$$

5. Compute the actual value of alpha achieved by the design by substituting p_2 for p_1 to obtain

$$\alpha^* = \sum_A \binom{n_1}{x_{11}} \binom{n_2}{x_{21}} p_2^{x_{11} + x_{21}} q_2^{n_1 + n_2 - x_{11} - x_{21}}$$

Asymptotic Approximations

When the values of n_1 and n_2 are large (say over 200), these formulas often take a long time to evaluate. In this case, a large sample approximation is used. The large sample approximation is made by replacing the values of $\hat{p}_1$ and $\hat{p}_2$ in the *z* values with the corresponding values of p_1 and p_2 under the alternative hypothesis and then computing the results based on the normal distribution. Note that in large samples, the Farrington and Manning statistic is substituted for the Gart and Nam statistic. Also, for large samples, the results for the odds ratio have not (to our knowledge) been published. In this case, we substitute the calculations based on the ratio.

Test Statistics

Several test statistics have been proposed for testing whether the difference, ratio, or odds ratio are different from a specified value. The main difference among the several test statistics is in the formula used to compute the standard error used in the denominator. These tests are based on the following *z*-test

$$z_t = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0 - c}{\hat{\sigma}}$$

The constant, *c*, represents a continuity correction that is applied in some cases. When the continuity correction is not used, *c* is zero. In power calculations, the values of $\hat{p}_1$ and $\hat{p}_2$ are not known. The corresponding values of p_1 and p_2 under the alternative hypothesis are reasonable substitutes.

Following is a list of the test statistics available in ***PASS***. The availability of several test statistics begs the question of which test statistic you should use. The answer is simple: you should use the test statistic that you will use to analyze your data. You may choose a method because it is a standard in your industry, because it seems to have better statistical properties, or because your statistical package calculates it. Whatever your reasons for selecting a certain test statistic, you should use the same test statistic during power or sample calculations.

Z Test (Pooled)

This test was first proposed by Karl Pearson in 1900. Although this test is usually expressed directly as a chi-square statistic, it is expressed here as a *z* statistic so that it can be more easily used for one-sided hypothesis testing. The proportions are pooled (averaged) in computing the standard error. The formula for the test statistic is

$$z_t = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0}{\hat{\sigma}_1}$$

where

$$\hat{\sigma}_1 = \sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

$$\bar{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$$

Z Test (Unpooled)

This test statistic does not pool the two proportions in computing the standard error.

$$z_t = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0}{\hat{\sigma}_2}$$

where

$$\hat{\sigma}_2 = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

Z Test with Continuity Correction (Pooled)

This test is the same as Z Test (Pooled), except that a continuity correction is used. Recall that in the null case, the continuity correction makes the results closer to those of Fisher's Exact test.

$$z_t = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0 + \frac{F}{2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}{\hat{\sigma}_1}$$

$$\hat{\sigma}_1 = \sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

$$\bar{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$$

where F is -1 for lower-tailed, 1 for upper-tailed, and both -1 and 1 for two-sided hypotheses.

Z Test with Continuity Correction (Unpooled)

This test is the same as the Z Test (Unpooled), except that a continuity correction is used. Recall that in the null case, the continuity correction makes the results closer to those of Fisher's Exact test.

$$z_t = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0 - \frac{F}{2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}{\hat{\sigma}_2}$$

$$\hat{\sigma}_2 = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

where F is -1 for lower-tailed, 1 for upper-tailed, and both -1 and 1 for two-sided hypotheses.

T-Test of Difference

Based on a detailed, comparative study of the behavior of several tests, D'Agostino (1988) and Upton (1982) proposed using the usual two-sample t-test for testing whether the two proportions are equal. One substitutes a '1' for a success and a '0' for a failure in the usual, two-sample *t*-test formula.

Miettinen and Nurminen's Likelihood Score Test of the Difference

Miettinen and Nurminen (1985) proposed a test statistic for testing whether the difference is equal to a specified, non-zero, value, δ_0. The regular MLE's, $\hat{p}_1$ and $\hat{p}_2$, are used in the numerator of the score statistic while MLE's $\tilde{p}_1$ and $\tilde{p}_2$, constrained so that $\tilde{p}_1 - \tilde{p}_2 = \delta_0$, are used in the denominator. A correction factor of $N/(N\text{-}1)$ is applied to make the variance estimate less biased. The significance level of the test statistic is based on the asymptotic normality of the score statistic.

The formula for computing this test statistic is

$$z_{MND} = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0}{\hat{\sigma}_{MND}}$$

where

$$\hat{\sigma}_{MND} = \sqrt{\left(\frac{\tilde{p}_1\tilde{q}_1}{n_1} + \frac{\tilde{p}_2\tilde{q}_2}{n_2}\right)\left(\frac{N}{N-1}\right)}$$

$$\tilde{p}_1 = \tilde{p}_2 + \delta_0$$

$$\tilde{p}_2 = 2B\cos(A) - \frac{L_2}{3L_3}$$

$$A = \frac{1}{3}\left[\pi + \cos^{-1}\left(\frac{C}{B^3}\right)\right]$$

$$B = \text{sign}(C)\sqrt{\frac{L_2^2}{9L_3} - \frac{L_1}{3L_3}}$$

$$C = \frac{L_2^3}{27L_3^3} - \frac{L_1L_2}{6L_3^2} + \frac{L_0}{2L_3}$$

$$L_0 = x_{21}\delta_0(1-\delta_0)$$

$$L_1 = \left[N_2\delta_0 - N - 2x_{21}\right]\delta_0 + M_1$$

$$L_2 = (N + N_2)\delta_0 - N - M_1$$

$$L_3 = N$$

Miettinen and Nurminen's Likelihood Score Test of the Ratio

Miettinen and Nurminen (1985) proposed a test statistic for testing whether the ratio is equal to a specified value, ϕ_0. The regular MLE's, $\hat{p}_1$ and $\hat{p}_2$, are used in the numerator of the score statistic while MLE's $\tilde{p}_1$ and $\tilde{p}_2$, constrained so that $\tilde{p}_1 / \tilde{p}_2 = \phi_0$, are used in the denominator. A correction factor of N/(N-1) is applied to make the variance estimate less biased. The significance level of the test statistic is based on the asymptotic normality of the score statistic.

The formula for computing the test statistic is

$$z_{MNR} = \frac{\hat{p}_1 / \hat{p}_2 - \phi_0}{\sqrt{\left(\frac{\tilde{p}_1\tilde{q}_1}{n_1} + \phi_0^2\frac{\tilde{p}_2\tilde{q}_2}{n_2}\right)\left(\frac{N}{N-1}\right)}}$$

where

$$\tilde{p}_1 = \tilde{p}_2\phi_0$$

$$\tilde{p}_2 = \frac{-B - \sqrt{B^2 - 4AC}}{2A}$$

$$A = N\phi_0$$

$$B = -\left[N_1\phi_0 + x_{11} + N_2 + x_{21}\phi_0\right]$$

$$C = M_1$$

Miettinen and Nurminen's Likelihood Score Test of the Odds Ratio

Miettinen and Nurminen (1985) proposed a test statistic for testing whether the odds ratio is equal to a specified value, ψ_0. Because the approach they used with the difference and ratio does not easily extend to the odds ratio, they used a score statistic approach for the odds ratio. The regular MLE's are $\hat{p}_1$ and $\hat{p}_2$. The constrained MLE's are $\tilde{p}_1$ and $\tilde{p}_2$. These estimates are constrained so that $\tilde{\psi} = \psi_0$. A correction factor of N/(N-1) is applied to make the variance estimate less biased. The significance level of the test statistic is based on the asymptotic normality of the score statistic.

The formula for computing the test statistic is

$$z_{MNO} = \frac{\frac{\left(\hat{p}_1 - \tilde{p}_1\right)}{\tilde{p}_1\tilde{q}_1} - \frac{\left(\hat{p}_2 - \tilde{p}_2\right)}{\tilde{p}_2\tilde{q}_2}}{\sqrt{\left(\frac{1}{N_2\tilde{p}_1\tilde{q}_1} + \frac{1}{N_2\tilde{p}_2\tilde{q}_2}\right)\left(\frac{N}{N-1}\right)}}$$

where

$$\tilde{p}_1 = \frac{\tilde{p}_2\psi_0}{1 + \tilde{p}_2\left(\psi_0 - 1\right)}, \quad \tilde{p}_2 = \frac{-B + \sqrt{B^2 - 4AC}}{2A},$$

$$A = N_2\left(\psi_0 - 1\right), \quad B = N_1\psi_0 + N_2 - M_1\left(\psi_0 - 1\right), \quad C = -M_1$$

Farrington and Manning's Likelihood Score Test of the Difference

Farrington and Manning (1990) proposed a test statistic for testing whether the difference is equal to a specified value, δ_0. The regular MLE's, $\hat{p}_1$ and $\hat{p}_2$, are used in the numerator of the score statistic while MLE's $\tilde{p}_1$ and $\tilde{p}_2$, constrained so that $\tilde{p}_1 - \tilde{p}_2 = \delta_0$, are used in the denominator. The significance level of the test statistic is based on the asymptotic normality of the score statistic.

The formula for computing the test statistic is

$$z_{FMD} = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0}{\sqrt{\left(\frac{\tilde{p}_1\tilde{q}_1}{n_1} + \frac{\tilde{p}_2\tilde{q}_2}{n_2}\right)}}$$

where the estimates, $\tilde{p}_1$ and $\tilde{p}_2$, are computed as in the corresponding test of Miettinen and Nurminen (1985) given above.

Farrington and Manning's Likelihood Score Test of the Ratio

Farrington and Manning (1990) proposed a test statistic for testing whether the ratio is equal to a specified value, ϕ_0. The regular MLE's, $\hat{p}_1$ and $\hat{p}_2$, are used in the numerator of the score statistic while MLE's $\tilde{p}_1$ and $\tilde{p}_2$, constrained so that $\tilde{p}_1 / \tilde{p}_2 = \phi_0$, are used in the denominator. A correction factor of N/(N-1) is applied to increase the variance estimate. The significance level of the test statistic is based on the asymptotic normality of the score statistic.

The formula for computing the test statistic is

$$z_{FMR} = \frac{\hat{p}_1 / \hat{p}_2 - \phi_0}{\sqrt{\left(\frac{\tilde{p}_1\tilde{q}_1}{n_1} + \phi_0^2\frac{\tilde{p}_2\tilde{q}_2}{n_2}\right)}}$$

where the estimates, $\tilde{p}_1$ and $\tilde{p}_2$, are computed as in the corresponding test of Miettinen and Nurminen (1985) given above.

Farrington and Manning's Likelihood Score Test of the Odds Ratio

Farrington and Manning (1990) indicate that the Miettinen and Nurminen statistic may be modified by removing the factor *N*/(*N*-1).

The formula for computing this test statistic is

$$z_{FMO} = \frac{\frac{\left(\hat{p}_1 - \tilde{p}_1\right)}{\tilde{p}_1\tilde{q}_1} - \frac{\left(\hat{p}_2 - \tilde{p}_2\right)}{\tilde{p}_2\tilde{q}_2}}{\sqrt{\left(\frac{1}{N_2\tilde{p}_1\tilde{q}_1} + \frac{1}{N_2\tilde{p}_2\tilde{q}_2}\right)}}$$

where the estimates, $\tilde{p}_1$ and $\tilde{p}_2$, are computed as in the corresponding test of Miettinen and Nurminen (1985) given above.

Gart and Nam's Likelihood Score Test of the Difference

Gart and Nam (1990), page 638, proposed a modification to the Farrington and Manning (1988) difference test that corrected for skewness. Let $z_{FMD}(\delta)$ stand for the Farrington and Manning difference test statistic described above. The skewness corrected test statistic, z_{GND}, is the appropriate solution to the quadratic equation

$$\left(-\tilde{\gamma}\right)z_{GND}^2 + \left(-1\right)z_{GND} + \left(z_{FMD}(\delta) + \tilde{\gamma}\right) = 0$$

where

$$\tilde{\gamma} = \frac{\tilde{V}^{3/2}(\delta)}{6}\left(\frac{\tilde{p}_1\tilde{q}_1(\tilde{q}_1 - \tilde{p}_1)}{n_1^2} - \frac{\tilde{p}_2\tilde{q}_2(\tilde{q}_2 - \tilde{p}_2)}{n_2^2}\right)$$

Gart and Nam's Likelihood Score Test of the Ratio

Gart and Nam (1988), page 329, proposed a modification to the Farrington and Manning (1988) ratio test that corrected for skewness. Let $z_{FMR}(\phi)$ stand for the Farrington and Manning ratio test statistic described above. The skewness corrected test statistic, z_{GNR}, is the appropriate solution to the quadratic equation

$$(-\tilde{\varphi})z_{GNR}^2 + (-1)z_{GNR} + (z_{FMR}(\phi) + \tilde{\varphi}) = 0$$

where

$$\tilde{\varphi} = \frac{1}{6\tilde{u}^{3/2}}\left(\frac{\tilde{q}_1(\tilde{q}_1 - \tilde{p}_1)}{n_1^2\tilde{p}_1^2} - \frac{\tilde{q}_2(\tilde{q}_2 - \tilde{p}_2)}{n_2^2\tilde{p}_2^2}\right)$$

$$\tilde{u} = \frac{\tilde{q}_1}{n_1\tilde{p}_1} + \frac{\tilde{q}_2}{n_2\tilde{p}_2}$$

Procedure Options

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

Data Tab (Common Options)

The Data tab contains the parameters associated with this test such as the proportions, sample sizes, alpha, and power. This chapter covers four procedures, each of which has different options. This section documents options that are common to all four procedures. Later, unique options for each procedure will be documented.

Solve For

Find (Solve For)

This option specifies the parameter to be solved for using the other parameters. The parameters that may be selected are *P1.1*, *Alpha*, *Power and Beta*, *N1*, and *N2*. Under most situations, you will select either *Power and Beta* or *N1*.

Select N1 when you want to calculate the sample size needed to achieve a given power and alpha level.

Select *Power and Beta* when you want to calculate the power of an experiment.

Error Rates

Power or Beta

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

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

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

Alpha (Significance Level)

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

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

Sample Size

N1 (Sample Size Group 1)

Enter a value (or range of values) for the sample size of this group. 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.

Effect Size – Control (Group 2)

P2 (Control Group Proportion)

Specify the value of P2, the control, baseline, or standard group's proportion. The null hypothesis is that the two proportions differ by a specified amount. Since P2 is a proportion, these values must be 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 0.1*.

Test

Test Type

Specify which test statistic is used in searching and reporting.

Note that *C.C.* is an abbreviation for *Continuity Correction*. This refers to the adding or subtracting 1/(2n) to (or from) the numerator of the z-value to bring the normal approximation closer to the binomial distribution.

Data Tab (Proportions)

This section documents options that are used when the parameterization is in terms of the values of the two proportions, P1 and P2. P1.0 is the value of the P1 assumed by the null hypothesis and P1.1 is the value of P1 at which the power is calculated.

Effect Size – Treatment (Group 1)

P1.0 (Group 1 Proportion|H0)

This option specifies the value of the group 1 proportion given the null hypothesis. The power calculations assume that P1.0 is the value of the P1 under the null hypothesis. In this non-null case, the value of P1.0 is not equal to P2 as it is in the null case.

You may enter a range of values such as *0.03 0.05 0.10* or *0.01 to 0.05 by 0.01*.

Proportions must be between zero and one. They cannot take on the values zero or one.

P1.1 (Group 1 Proportion|H1)

This is the value of P1 under the alternative hypothesis. It is written P1.1. The power calculations assume that this is the actual value of this proportion.

You may enter a range of values such as *0.1 0.2 0.3* or *0.1 to 0.9 by 0.1*.

Note that values must be between zero and one.

Test

H1 (Alternative Hypothesis)

This option specifies whether a one-sided or two-sided hypothesis is analyzed.

One-Sided (H1: P1<P2) refers to a one-sided test in which the alternative hypothesis is of the form H1: P1<P2.

One-Sided (H1:D1>D0) refers to a one-sided test in which the alternative hypothesis is of the form H1: P1>P2.

Two-Sided refers to a two-sided test in which the alternative hypothesis is of the type H1: P1<>P2+D0. Here '<>' means unequal.

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

Data Tab (Differences)

This section documents options that are used when the parameterization is in terms of the difference, P1 – P2. P1.0 is the value of P1 assumed by the null hypothesis and P1.1 is the value of P1 at which the power is calculated. Once P2, D0, and D1 are given, the values of P1.1 and P1.0 can be calculated.

Effect Size – Differences

D0 (Difference|H0 = P1.0 – P2)

This option specifies the difference between the two proportions given in the null hypothesis, H0. This difference is used with P2 to calculate the value of P1.0 using the formula: P1.0 = P2 + D0. Note that P1.0 here means the value of P1 under H0.

Differences must be between -1 and 1. They cannot take on the values -1, 0, or 1.

The power calculations use P1.0 as the value of the proportion in group 1 (the experimental or treatment group) under the null hypothesis. In this non-null case, the value of P1.0 is not equal to P2 as it is in the null case.

You may enter a range of values such as *0.03 0.05 0.10* or *0.01 to 0.05 by 0.01*.

D1 (Difference|H0 = P1.1 – P2)

This option specifies the difference between the P1.1 and P2. This difference is used with P2 to calculate the value of P1.1 using the formula: P1.1 = D1 + P2. Note that P1.1 here means the value of P1 under H1. Differences must be between -1 and 1. They cannot take on the values -1 or 1.

The power calculations assume that P1.1 is the actual value of the proportion in group 1 (experimental or treatment group).

This option is only used if you are specifying *Differences*.

You may enter a range of values such as *0.03 0.05 0.10* or *0.01 to 0.05 by 0.01*.

Test

H1 (Alternative Hypothesis)

This option specifies whether a one-sided or two-sided hypothesis is analyzed.

One-Sided (H1:D1<D0) refers to a one-sided test in which the alternative hypothesis is of the form H1: D1<D0.

One-Sided (H1:D1>D0) refers to a one-sided test in which the alternative hypothesis is of the form H1: D1>D0.

Two-Sided refers to a two-sided test in which the alternative hypothesis is of the type H1: D1<>D0. Here '<>' means unequal.

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

Data Tab (Ratios)

This section documents options that are used when the parameterization is in terms of the ratio, P1 / P2. P1.0 is the value of P1 assumed by the null hypothesis and P1.1 is the value of P1 at which the power is calculated. Once P2, R0, and R1 are given, the values of P1.0 and P1.1 can be calculated.

Effect Size – Ratios

R0 (Ratio|H0 = P1.0 / P2)

This option specifies the ratio between the group 1 proportion under the null hypothesis P1.0 and P2. This ratio is used with P2 to calculate the value of P1.0 using the formula: P1.0 = R0 x P2. The power calculations assume that P1.0 is the value of the P1 under the null hypothesis. In this non-null case, the value of P1.0 is not equal to P2 as it is in the null case.

You may enter a range of values such as *0.5 0.6 0.7 0.8* or *1.25 to 2.0 by 0.25*.

Ratios must greater than zero.

R1 (Ratio|H1 = P1.1 / P2)

This option specifies the ratio of P1.1 and P2, where P1.1 is the proportion in group 1 under the alternative hypothesis. This ratio is used with P2 to calculate the value of P1.1 using the formula: P1.1 = R1 x P2.The power calculations assume that P1.1 is the actual value of the proportion in group 1 (experimental or treatment group).

You may enter a range of values such as *0.5 0.6 0.7 0.8* or *1.25 to 2.0 by 0.25*.

Ratios must greater than zero. They cannot take on the value of one.

Test

H1 (Alternative Hypothesis)

This option specifies whether a one-sided or two-sided hypothesis is analyzed.

One-Sided (H1:R1<R0) refers to a one-sided test in which the alternative hypothesis is of the form H1: R1<R0.

One-Sided (H1:R1>R0) refers to a one-sided test in which the alternative hypothesis is of the form H1: R1>R0.

Two-Sided refers to a two-sided test in which the alternative hypothesis is of the type H1: R1<>R0. Here '<>' means unequal.

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

Data Tab (Odds Ratios)

This section documents options that are used when the parameterization is in terms of the odds ratios, O1.1 / O2 and O1.0 / O2. Note that the odds are defined as O2 = P2 / (1 – P2), O1.0 = P1.0 / (1 – P1.0), etc. P1.0 is the value of P1 assumed by the null hypothesis and P1.1 is the value of P1 at which the power is calculated. Once P2, OR0, and OR1 are given, the values of P1.1 and P1.0 can be calculated.

Effect Size – Odds Ratios

OR0 (Odds Ratio|H0 = O1.0 / O2)

This option specifies the odds ratio between the group 1 proportion under the null hypothesis and the proportion in group 2. This value is used with P2 to calculate the value of P1.0. The power calculations assume that P1.0 is the value of the P1 under the null hypothesis. In this non-null case, the value of P1.0 is not equal to P2 as it is in the null case.

You may enter a range of values such as *0.5 0.6 0.7 0.8* or *1.25 to 2.0 by 0.25*.

Odds ratios must greater than zero.

OR1 (Odds Ratio|H1 = O1.1 / O2)

This option specifies the odds ratio of the two proportions P1.1 and P2. This odds ratio is used with P2 to calculate the value of P1.1. The power calculations assume that P1.1 is the actual value of the proportion in group 1, which is the experimental, or treatment, group.

You may enter a range of values such as *0.5 0.6 0.7 0.8* or *1.25 to 2.0 by 0.25*. Odds ratios must greater than zero.

Test

H1 (Alternative Hypothesis)

This option specifies whether a one-sided or two-sided hypothesis is analyzed.

One-Sided (H1: OR1<OR0) refers to a one-sided test in which the alternative hypothesis is of the form H1: OR1<OR0.

One-Sided (H1: OR1>OR0) refers to a one-sided test in which the alternative hypothesis is of the form H1: OR1>OR0.

Two-Sided refers to a two-sided test in which the alternative hypothesis is of the type H1: OR1<>OR0. Here '<>' means unequal.

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

Options Tab

The Options tab contains various limits and options.

Maximum Iterations

Maximum Iterations Before Search Termination

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 not reported. A value of at least 500 is recommended.

Zero Counts

Zero Count Adjustment Method

Zero cell counts often cause calculation problems. To compensate for this, a small value (called the Zero Count Adjustment Value) can be added either to all cells or to all cells with zero counts. This option specifies whether you want to use the adjustment and which type of adjustment you want to use. We recommend that you use the option 'Add to zero cells only'.

Zero cell values often do not occur in practice. However, since power calculations are based on total enumeration, they will occur in power and sample size estimation.

Adding a small value is controversial, but can be necessary for computational considerations. Statisticians have recommended adding various fractions to zero counts.. We have found that adding 0.0001 seems to work well.

Zero Count Adjustment Value

Zero cell counts cause many calculation problems when computing power or sample size. To compensate for this, a small value may be added either to all cells or to all zero cells. This value indicates the amount that is added. We have found that 0.0001 works well.

Be warned that the value of the ratio and the odds ratio will be affected by the amount specified here!

Exact Test Options

Maximum N1 or N2 for Exact Calculations

When either N1 or N2 is above this amount, power calculations are based on the normal approximation to the binomial. In this case, the actual value of alpha is not calculated. Currently, for three-gigahertz computers, a value near 200 is reasonable. As computers increase in speed, this number may be increased.

Example 1 – Finding Power

A study is being designed to study the effectiveness of a new treatment. Historically, the standard treatment has enjoyed a 60% cure rate. The new treatment reduces the seriousness of certain side effects that occur with the standard treatment. Thus, the new treatment will be adopted even if it is slightly less effective than the standard treatment. The researchers will recommend adoption of the new treatment if it has a cure rate of at least 55%.

The researchers plan to use the Farrington and Manning likelihood score test statistic to analyze the data. They want to study the power of the one-sided Farrington and Manning test at group sample sizes ranging from 50 to 2000 for detecting a difference significantly greater than -0.05 when the actual cure rate of the new treatment ranges from 57% to 70%. The significance level will be 0.05.

Setup

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

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power	*Ignored since this is the Find setting*
Alpha	**0.05**
N1 (Sample Size Group 1)	**50 100 250 500 1000 1500 2000**
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
D0 (Difference|H0 = P1.0 – P2)	**-0.05**
D1 (Difference|H1 = P1.1 – P2)	**-0.03 0.00 0.05 0.10**
P2 (Control Group Proportion)	**0.6**
H1 (Alternative Hypothesis)	**One-Sided (H1:D1>D0)**
Test Type	**Likelihood Score (Farr. & Mann.)**
Options Tab	
Maximum N1 or N2 Exact	**300**

Annotated Output

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

Numeric Results

Numeric Results of Tests Based on the Difference: P1 – P2
H0: P1-P2<=D0. H1: P1-P2=D1>D0. Test Statistic: Score test (Farrington & Manning)

Power	Sample Size Grp 1 N1	Sample Size Grp 2 N2	Prop Grp 2 or Control P2	Prop|H0 Grp 1 or Trtmnt P1.0	Prop|H1 Grp 1 or Trtmnt P1.1	Diff if H0 D0	Diff if H1 D1	Target Alpha	Actual Alpha	Beta
0.0778	50	50	0.6000	0.5500	0.5700	-0.0500	-0.0300	0.0500	0.0527	0.9222
0.0865	100	100	0.6000	0.5500	0.5700	-0.0500	-0.0300	0.0500	0.0499	0.9135
0.1189	250	250	0.6000	0.5500	0.5700	-0.0500	-0.0300	0.0500	0.0516	0.8811
0.1583	500	500	0.6000	0.5500	0.5700	-0.0500	-0.0300	0.0500		0.8417
0.2310	1000	1000	0.6000	0.5500	0.5700	-0.0500	-0.0300	0.0500		0.7690
0.2976	1500	1500	0.6000	0.5500	0.5700	-0.0500	-0.0300	0.0500		0.7024
0.3596	2000	2000	0.6000	0.5500	0.5700	-0.0500	-0.0300	0.0500		0.6404

Report continues …

Note: exact results based on the binomial were only calculated when both N1 and N2 were less than 300.

Report Definitions
'H0' as an abbreviation for the NULL hypothesis. This is the hypothesis being evaluated by the statistical test.
'H1' as an abbreviation for the ALTERNATIVE hypothesis. This hypothesis gives the 'true' parameter values.
'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 corresponding populations.
'P2' is the proportion for group two. This is the standard, reference, baseline, or control group.
'P1.0' is the proportion for group one (treatment group) assuming the null hypothesis (H0).
'P1.1' is the proportion for group one (treatment group) assuming the alternative hypothesis (H1).
'D0: Diff|H0' is the difference P1 – P2 assuming the null hypothesis (H0).
'D1: Diff|H1' is the difference P1 – P2 assuming the alternative hypothesis (H1).
'Target Alpha' is the probability of rejecting a true null hypothesis that was desired.
'Actual Alpha' is the value of alpha that is actually achieved.
'Beta' is the probability of accepting a false H0. Beta = 1 - Power.

Summary Statements
Group sample sizes of 50 in group one and 50 in group two achieve 8% power to detect a difference between the group proportions of -0.0300. The proportion in group two is 0.6000. The proportion in group one is assumed to be 0.5500 under the null hypothesis and 0.5700 under the alternative hypothesis. The test statistic used is the one-sided Score test (Farrington & Manning). The significance level of the test was targeted at 0.0500. The significance level actually achieved by this design is 0.0527.

This report shows the values of each of the parameters, one scenario per row. Note that the actual alpha value is blank for sample sizes greater than 300, which was the limit set for exact computation.

Most of the report columns have obvious interpretations. Those that may not be obvious are presented here.

Prop Grp 2 or Control P2

This is the value of P2, the proportion responding positively in the control group.

Prop|H0 Grp 1 or Trtmnt P1.0

This is the value of P1.0, the proportion responding positively in the treatment group as specified by the null hypothesis. The difference between this value and P2 is the value specified by the null hypothesis.

Prop|H1 Grp 1 or Trtmnt P1.1

This is the value of P1.1, the proportion responding positively in the treatment group as specified by the alternative hypothesis. The difference between this value and P2 is the value specified by the alternative hypothesis.

Diff if H0 D0

This is the value of D0, the difference between proportions under the null hypothesis.

Diff if H1 D1

This is the value of D1, the difference between proportions under the alternative hypothesis.

Target Alpha

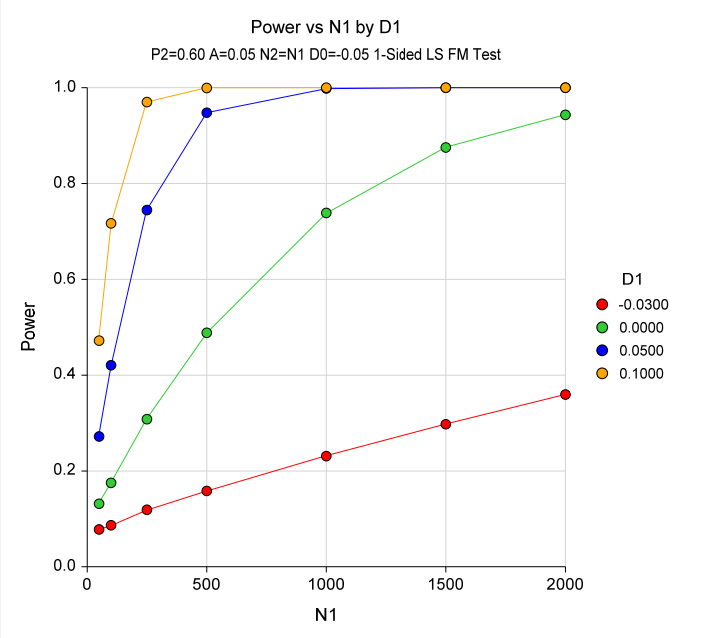
This is the value of alpha that was targeted by the design. Note that the target alpha is not usually achieved exactly.

Actual Alpha

This is the value of alpha that was actually achieved by this design. Note that since the limit on exact calculations was set to 300, and since this value is calculated exactly, it is not shown for values of N1 greater than 300.

The difference between the Target Alpha and the Actual Alpha is caused by the discrete nature of the binomial distribution and the use of the normal approximation to the binomial in determining the critical value of the test statistic.

Plots Section



The values from the table are displayed in the above chart. This chart gives us a quick look at the sample size that will be required for various values of D1.

Example 2 – Finding the Sample Size

Continuing with the scenario given in Example 1, the researchers want to determine the sample size needed to achieve 80% power for each value of D1. To cut down on the runtime, they decide to look at approximate values whenever N1 is greater than 100.

Setup

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

Option	Value
Data Tab	
Find (Solve For)	**N1**
Power	**0.80**
Alpha	**0.05**
N1 (Sample Size Group 1)	*Ignored since this is the Find setting*
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
D0 (Difference|H0 = P1.0 – P2)	**-0.05**
D1 (Difference|H1 = P1.1 – P2)	**-0.03 0.00 0.05 0.10**
P2 (Control Group Proportion)	**0.6**
H1 (Alternative Hypothesis)	**One-Sided (H1:D1>D0)**
Test Type	**Likelihood Score (Farr. & Mann.)**
Options Tab	
Maximum N1 or N2 Exact	**100**

Output

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

Numeric Results

Numeric Results of Tests Based on the Difference: P1 – P2
H0: P1-P2<=D0. H1: P1-P2=D1>D0. Test Statistic: Score test (Farrington & Manning)

Power	Sample Size Grp 1 N1	Sample Size Grp 2 N2	Prop Grp 2 or Control P2	Prop|H0 Grp 1 or Trtmnt P1.0	Prop|H1 Grp 1 or Trtmnt P1.1	Diff if H0 D0	Diff if H1 D1	Target Alpha	Actual Alpha	Beta
0.8000	7491	7491	0.6000	0.5500	0.5700	-0.0500	-0.0300	0.0500		0.2000
0.8002	1186	1186	0.6000	0.5500	0.6000	-0.0500	0.0000	0.0500		0.1998
0.8008	290	290	0.6000	0.5500	0.6500	-0.0500	0.0500	0.0500		0.1992
0.8011	125	125	0.6000	0.5500	0.7000	-0.0500	0.1000	0.0500		0.1989

The required sample size will depend a great deal on the value of D1. The researchers should spend time determining the most accurate value for D1.

Example 3 – Comparing the Power of Several Test Statistics

Continuing with Example 1, the researchers want to determine which of the eight possible test statistics to adopt by using the comparative reports and charts that ***PASS*** produces. They decide to compare the powers and actual alphas for various sample sizes between 50 and 200 when D1 is 0.05.

Setup

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

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power	*Ignored since this is the Find setting*
Alpha	**0.05**
N1 (Sample Size Group 1)	**50 100 150 200**
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
D0 (Difference|H0 = P1.0 – P2)	**-0.05**
D1 (Difference|H1 = P1.1 – P2)	**0.05**
P2 (Control Group Proportion)	**0.6**
H1 (Alternative Hypothesis)	**One-Sided (H1:D1>D0)**
Test Type	**Likelihood Score (Farr. & Mann.)**
Reports Tab	
Show Numeric Report	**Not checked**
Show Comparative Reports	**Checked**
Show Definitions	**Not checked**
Show Plots	**Not checked**
Show Comparative Plots	**Checked**
Number of Summary Statements	**0**
Options Tab	
Maximum N1 or N2 Exact	**300**

Output

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

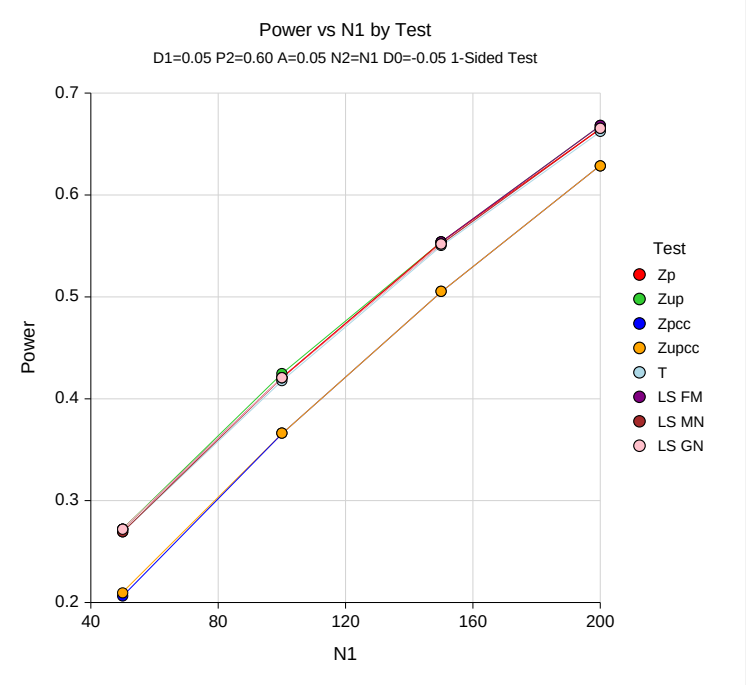
Numeric Results and Plots

Power Comparison of Tests Based on the Difference: P1 – P2
H0: P1-P2<=D0. H1: P1-P2=D1>D0.

N1/N2	P2	P1.1	Target Alpha	Z(P) Test Power	Z(UnP) Test Power	Z(P) CC Test Power	Z(UnP) CC Test Power	T Test Power	F.M. Score Power	M.N. Score Power	G.N. Score Power
50/50	0.6000	0.6500	0.0500	0.2720	0.2720	0.2064	0.2096	0.2694	0.2720	0.2694	0.2720
100/100	0.6000	0.6500	0.0500	0.4207	0.4248	0.3663	0.3663	0.4178	0.4207	0.4207	0.4207
150/150	0.6000	0.6500	0.0500	0.5540	0.5540	0.5054	0.5054	0.5519	0.5540	0.5519	0.5519
200/200	0.6000	0.6500	0.0500	0.6654	0.6683	0.6286	0.6286	0.6624	0.6683	0.6654	0.6654

Actual Alpha Comparison of Tests Based on the Difference: P1 – P2
H0: P1-P2<=D0. H1: P1-P2=D1>D0.

N1/N2	P2	P1.1	Target Alpha	Z(P) Test Alpha	Z(UnP) Test Alpha	Z(P) CC Test Alpha	Z(UnP) CC Test Alpha	T Test Alpha	F.M. Score Alpha	M.N. Score Alpha	G.N. Score Alpha
50/50	0.6000	0.6500	0.0500	0.0527	0.0527	0.0342	0.0343	0.0526	0.0527	0.0526	0.0527
100/100	0.6000	0.6500	0.0500	0.0499	0.0500	0.0369	0.0369	0.0499	0.0499	0.0499	0.0499
150/150	0.6000	0.6500	0.0500	0.0509	0.0509	0.0398	0.0398	0.0509	0.0509	0.0509	0.0509
200/200	0.6000	0.6500	0.0500	0.0479	0.0482	0.0387	0.0387	0.0477	0.0482	0.0479	0.0479



It is interesting to note that the powers of the continuity-corrected test statistics are consistently lower than the other tests. This occurs because the actual alpha achieved by these tests is lower than for the other tests.

Example 4 – Validation using Machin et al. with Equal Sample Sizes

Machin et al. (1997), page 106, present a sample size study in which P2 = 0.5, D0 = -0.2, D1=0, one-sided alpha = 0.1, and beta = 0.2. Using the Farrington and Manning test statistic, they found the sample size to be 55 in each group.

Setup

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

Option	Value
Data Tab	
Find (Solve For)	**N1**
Power	**0.80**
Alpha	**0.10**
N1 (Sample Size Group 1)	*Ignored since this is the Find setting*
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
D0 (Difference|H0 = P1.0 – P2)	**-0.2**
D1 (Difference|H1 = P1.1 – P2)	**0**
P2 (Control Group Proportion)	**0.50**
H1 (Alternative Hypothesis)	**One-Sided (H1:D1>D0)**
Test Type	**Likelihood Score (Farr. & Mann.)**
Options Tab	
Maximum N1 or N2 Exact	**2 (Set low for a rapid search.)**

Output

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

Numeric Results

Numeric Results of Tests Based on the Difference: P1 – P2
H0: P1-P2<=D0. H1: P1-P2=D1>D0. Test Statistic: Score test (Farrington & Manning)

Power	Sample Size Grp 1 N1	Sample Size Grp 2 N2	Prop Grp 2 or Control P2	Prop|H0 Grp 1 or Trtmnt P1.0	Prop|H1 Grp 1 or Trtmnt P1.1	Diff if H0 D0	Diff if H1 D1	Target Alpha	Actual Alpha	Beta
0.8001	55	55	0.5000	0.3000	0.5000	-0.2000	0.0000	0.1000		0.1999

PASS found the required sample size to be 55, which corresponds to the results of Machin et al.

Example 5 – Validation using Farrington and Manning

Farrington and Manning (1990), page 1451, present a sample size study in which P2 = 0.05, D0 = 0.2, D1=0.35, one-sided alpha = 0.05, and beta = 0.20. Using the Farrington and Manning test statistic, they found the sample size to be 80 in each group. They mention that the true power is 0.813.

Setup

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

Option	Value
Data Tab	
Find (Solve For)	**N1**
Power	**0.80**
Alpha	**0.05**
N1 (Sample Size Group 1)	*Ignored since this is the Find setting*
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
D0 (Difference|H0 = P1.0 – P2)	**0.2**
D1 (Difference|H1 = P1.1 – P2)	**0.35**
P2 (Control Group Proportion)	**0.05**
H1 (Alternative Hypothesis)	**One-Sided (H1:D1>D0)**
Test Type	**Likelihood Score (Farr. & Mann.)**
Options Tab	
Maximum N1 or N2 Exact	**2 (Set low for a rapid search.)**

Output

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

Numeric Results

Numeric Results of Tests Based on the Difference: P1 – P2
H0: P1-P2<=D0. H1: P1-P2=D1>D0. Test Statistic: Score test (Farrington & Manning)

Power	Sample Size Grp 1 N1	Sample Size Grp 2 N2	Prop Grp 2 or Control P2	Prop|H0 Grp 1 or Trtmnt P1.0	Prop|H1 Grp 1 or Trtmnt P1.1	Diff if H0 D0	Diff if H1 D1	Target Alpha	Actual Alpha	Beta
0.8007	80	80	0.0500	0.2500	0.4000	0.2000	0.3500	0.0500		0.1993

PASS also calculated the required sample size to be 80.

Next, to calculate the exact power for this sample size, we make the following changes to the template.

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
N1 (Sample Size Group 1)	**80**
Options Tab	
Maximum N1 or N2 Exact	**300 (Set >80 to force exact calculation.)**

Numeric Results

Numeric Results of Tests Based on the Difference: P1 – P2
H0: P2-P1<=D0. H1: P2-P1=D1>D0. Test Statistic: Score test (Farrington & Manning)

Power	Sample Size Grp 1 N1	Sample Size Grp 2 N2	Prop Grp 2 or Control P2	Prop|H0 Grp 1 or Trtmnt P1.0	Prop|H1 Grp 1 or Trtmnt P1.1	Diff if H0 D0	Diff if H1 D1	Target Alpha	Actual Alpha	Beta
0.8132	80	80	0.0500	0.2500	0.4000	0.2000	0.3500	0.0500	0.0553	0.1868

PASS also calculated the exact power to be 0.813.

Example 6 – Validation of Risk Ratio Calculations using Blackwelder

Blackwelder (1993), page 695, presents a table of power values for several scenarios using the risk ratio. The second line of the table presents the results for the following scenario: P2 = 0.04, R0 = 0.3, R1=0.1, N1=N2=1044, one-sided alpha = 0.05, and beta = 0.20. Using the Farrington and Manning likelihood-score test statistic, he found the exact power to be 0.812, the exact alpha to be 0.044, and, using the asymptotic formula, the approximate power to be 0.794.

Setup

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

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power	*Ignored since this is the Find setting*
Alpha	**0.05**
N1 (Sample Size Group 1)	**1044**
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
R0 (Ratio|H0 = P1.0/P2)	**0.3**
R1 (Ratio|H1 = P1.1/P2)	**0.1**
P2 (Control Group Proportion)	**0.04**
H1 (Alternative Hypothesis)	**One-Sided (H1:R1<R0)**
Test Type	**Likelihood Score (Farr. & Mann.)**
Options Tab	
Maximum N1 or N2 Exact	**2000 (Set high for exact results.)**

Annotated Output

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

Numeric Results

Numeric Results of Tests Based on the Ratio: P1 / P2
H0: P1/P2>=R0. H1: P1/P2=R1<R0. Test Statistic: Score test (Farrington & Manning)

Power	Sample Size Grp 1 N1	Sample Size Grp 2 N2	Prop Grp 2 or Control P2	Prop|H0 Grp 1 or Trtmnt P1.0	Prop|H1 Grp 1 or Trtmnt P1.1	Ratio if H0 R0	Ratio if H1 R1	Target Alpha	Actual Alpha	Beta
0.8118	1044	1044	0.0400	0.0120	0.0040	0.300	0.100	0.0500	0.0444	0.1882

PASS also calculated the exact power to be 0.812 and the actual alpha to be 0.044, after rounding.

Next, to calculate the asymptotic power, we make the following changes to the template.

Option	Value
Options Tab	
Maximum N1 or N2 Exact	**2 (Set < 1044 to force asymptotic calculation.)**

Numeric Results

Numeric Results of Tests Based on the Difference: P1 – P2
H0: P1/P2>=R0. H1: P1/P2=R1<R0. Test Statistic: Score test (Farrington & Manning)

Power	Sample Size Grp 1 N1	Sample Size Grp 2 N2	Prop Grp 2 or Control P2	Prop|H0 Grp 1 or Trtmnt P1.0	Prop|H1 Grp 1 or Trtmnt P1.1	Ratio if H0 R0	Ratio if H1 R1	Target Alpha	Actual Alpha	Beta
0.7937	1044	1044	0.0400	0.0120	0.0040	0.300	0.100	0.0500		0.2063

PASS also calculated the asymptotic power to be 0.794.

Example 7 – Finding the Power after Completing an Experiment

Researchers are studying the effectiveness of a new treatment for cancer. Historically, the standard treatment has enjoyed a 52% cure rate. The new experimental treatment is believed to be better, but it costs much more to administer. After weighing cost versus effectiveness, the researchers decided that they will adopt the new treatment if the cure rate is at least 59%. They conduct a study in which 200 patients are given the new treatment, and 200 are given the standard regimen. They find that 66% are cured by the new treatment, while 52% are cured by the standard treatment. The Farrington and Manning likelihood score test, however, indicates that the results are not statistically significant for alpha = 0.05. They now desire to compute the power for a range of alternative values.

Note that a range of alternatives is used in computing the power instead of the actual difference from the study. The power should be computed at values representing practically significant differences from the null value.

Setup

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

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power	*Ignored since this is the Find setting*
Alpha	**0.05**
N1 (Sample Size Group 1)	**200**
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
D0 (Difference|H0 = P1.0 – P2)	**0.07**
D1 (Difference|H1 = P1.1 – P2)	**0.08 to 0.20 by 0.02**
P2 (Control Group Proportion)	**0.52**
H1 (Alternative Hypothesis)	**One-Sided (H1:D1>D0)**
Test Type	**Likelihood Score (Farr. & Mann.)**
Options Tab	
Maximum N1 or N2 Exact	**200**

Output

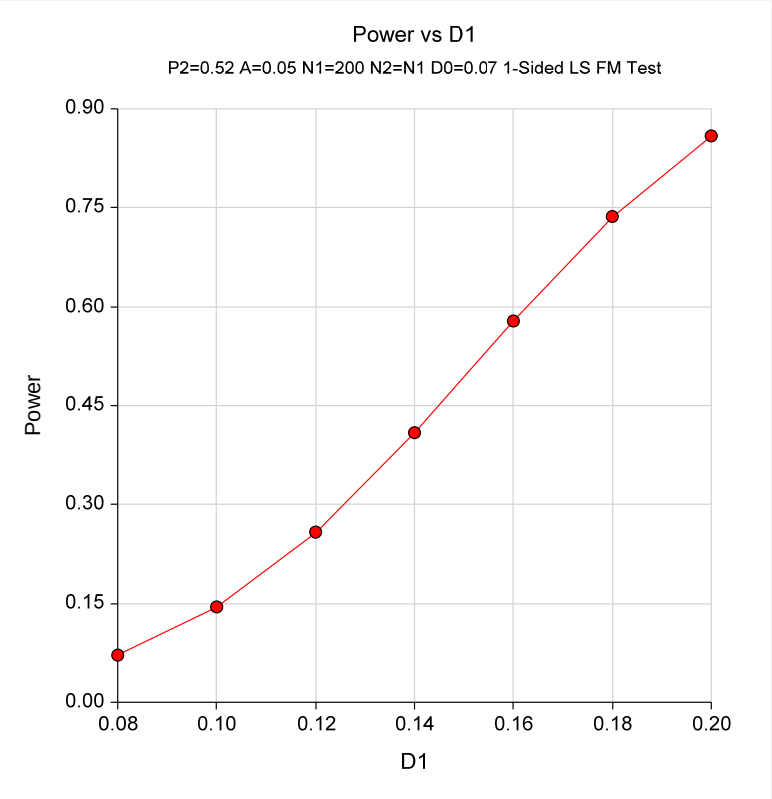
Click the Run button to perform the calculations and generate the following output.

Numeric Results and Plots

Numeric Results of Tests Based on the Difference: P1 - P2
H0: P1-P2<=D0. H1: P1-P2=D1>D0. Test Statistic: Score test (Farrington & Manning)

Power	Sample Size Grp 1 N1	Sample Size Grp 2 N2	Prop Grp 2 or Control P2	Prop|H0 Grp 1 or Trtmnt P1.0	Prop|H1 Grp 1 or Trtmnt P1.1	Diff if H0 D0	Diff if H1 D1	Target Alpha	Actual Alpha	Beta
0.0715	200	200	0.5200	0.5900	0.6000	0.0700	0.0800	0.0500	0.0479	0.9285
0.1446	200	200	0.5200	0.5900	0.6200	0.0700	0.1000	0.0500	0.0479	0.8554
0.2581	200	200	0.5200	0.5900	0.6400	0.0700	0.1200	0.0500	0.0479	0.7419
0.4089	200	200	0.5200	0.5900	0.6600	0.0700	0.1400	0.0500	0.0479	0.5911
0.5783	200	200	0.5200	0.5900	0.6800	0.0700	0.1600	0.0500	0.0479	0.4217
0.7368	200	200	0.5200	0.5900	0.7000	0.0700	0.1800	0.0500	0.0479	0.2632
0.8591	200	200	0.5200	0.5900	0.7200	0.0700	0.2000	0.0500	0.0479	0.1409

Note: exact results based on the binomial were only calculated when both N1 and N2 were less than 200.



The power depends a great deal on the value of D1 for this sample size. It is evident that the power is quite low for the majority of alternative values studied.

Example 8 – Finding the Sample Size using Ratios

A study is being designed to determine the effectiveness of a new treatment. Researchers would like to know how large of a sample is needed for comparison of the two treatments. The standard treatment has a success rate of 65%. The researchers will adopt the new treatment, which has fewer side effects, if the success rate is at least 90% of the rate for the standard treatment, i.e. P1 = 0.9 x P2 or P1/P2 = 0.9. They would like to calculate the sample sizes necessary to achieve 80%, 85%, 90%, and 95% power for the case where the true ratio between the two proportions is 1.1 and alpha = 0.05.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Two Proportions (Non-Zero Null Hypothesis) [Ratios]** procedure window by expanding **Proportions**, then **Two Independent Proportions**, then clicking on **Test (Non-Zero Null)**, and then clicking on **Tests for Two Proportions (Non-Zero Null Hypothesis) [Ratios]**. You may then make the appropriate entries as listed below, or open **Example 8** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**N1**
Power	**0.8 0.85 0.9 0.95**
Alpha	**0.05**
N1 (Sample Size Group 1)	*Ignored since this is the Find setting*
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
R0 (Ratio|H0 = P1.0/P2)	**0.9**
R1 (Ratio|H1 = P1.1/P2)	**1.1**
P2 (Control Group Proportion)	**0.65**
H1 (Alternative Hypothesis)	**One-Sided (H1:R1>R0)**
Test Type	**Likelihood Score (Farr. & Mann.)**
Options Tab	
Maximum N1 or N2 Exact	**300**

Output

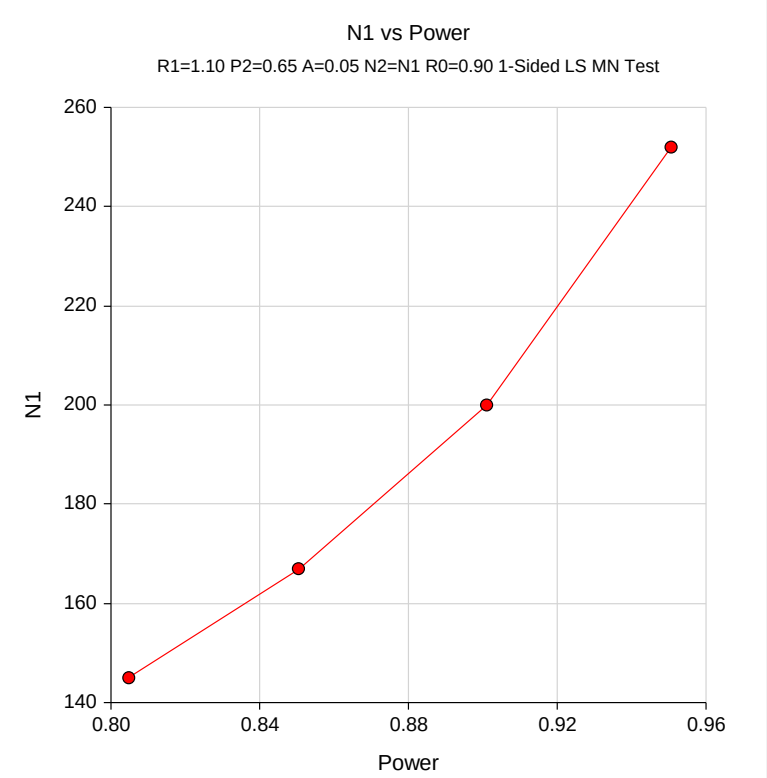
Click the Run button to perform the calculations and generate the following output.

Numeric Results

Numeric Results of Tests Based on the Ratio: P1 / P2
H0: P1/P2<=R0. H1: P1/P2=R1>R0. Test Statistic: Score test (Farrington & Manning)

Power	Sample Size Grp 1 N1	Sample Size Grp 2 N2	Prop Grp 2 or Control P2	Prop|H0 Grp 1 or Trtmnt P1.0	Prop|H1 Grp 1 or Trtmnt P1.1	Ratio if H0 R0	Ratio if H1 R1	Target Alpha	Actual Alpha	Beta
0.9506	252	252	0.6500	0.5850	0.7150	0.900	1.100	0.0500	0.0503	0.0494
0.9013	199	199	0.6500	0.5850	0.7150	0.900	1.100	0.0500	0.0508	0.0987
0.8504	167	167	0.6500	0.5850	0.7150	0.900	1.100	0.0500	0.0507	0.1496
0.8048	145	145	0.6500	0.5850	0.7150	0.900	1.100	0.0500	0.0499	0.1952

Note: exact results based on the binomial were only calculated when both N1 and N2 were less than 300.



Necessary sample sizes range from 145 for 80% power to 252 for 95% power for detecting a ratio of 1.1.

Chapter 210

Non-Inferiority Tests for Two Proportions

Introduction

This module provides power analysis and sample size calculation for non-inferiority and superiority tests in two-sample designs in which the outcome is binary. Users may choose from among eight popular test statistics commonly used for running the hypothesis test.

The power calculations assume that independent, random samples are drawn from two populations.

Four Procedures Documented Here

There are four procedures in the menus that use the program module described in this chapter. These procedures are identical except for the type of parameterization. The parameterization can be in terms of proportions, differences in proportions, ratios of proportions, and odds ratios. Each of these options is listed separately on the menus.

Example

A non-inferiority test example will set the stage for the discussion of the terminology that follows. Suppose that the current treatment for a disease works 70% of the time. Unfortunately, this treatment is expensive and occasionally exhibits serious side-effects. A promising new treatment has been developed to the point where it can be tested. One of the first questions that must be answered is whether the new treatment is as good as the current treatment. In other words, do at least 70% of treated subjects respond to the new treatment?

Because of the many benefits of the new treatment, clinicians are willing to adopt the new treatment even if it is slightly less effective than the current treatment. They must determine, however, how much less effective the new treatment can be and still be adopted. Should it be adopted if 69% respond? 68%? 65%? 60%? There is a percentage below 70% at which the difference between the two treatments is no longer considered ignorable. After thoughtful discussion with several clinicians, it was decided that if a response of at least 63% were achieved, the new treatment would be adopted. The difference between these two percentages is called the *margin of equivalence*. The margin of equivalence in this example is 7%.

The developers must design an experiment to test the hypothesis that the response rate of the new treatment is at least 0.63. The statistical hypothesis to be tested is

$$H_0: p_1 - p_2 \le -0.07 \text{ versus } H_1: p_1 - p_2 > -0.07$$

Notice that when the null hypothesis is rejected, the conclusion is that the response rate is at least 0.63. Note that even though the response rate of the current treatment is 0.70, the hypothesis test is about a response rate of 0.63. Also notice that a rejection of the null hypothesis results in the conclusion of interest.

Technical Details

The details of sample size calculation for the two-sample design for binary outcomes are presented in the chapter "Two Proportions Non-Null Case," and they will not be duplicated here. Instead, this chapter only discusses those changes necessary for non-inferiority and superiority tests.

Approximate sample size formulas for non-inferiority tests of two proportions are presented in Chow et al. (2003), page 90. Only large sample (normal approximation) results are given there. The results available in this module use exact calculations based on the enumeration of all possible values in the binomial distribution.

Suppose you have two populations from which dichotomous (binary) responses will be recorded. Assume without loss of generality that the higher proportions are better. The probability (or risk) of cure in population 1 (the treatment group) is p_1 and in population 2 (the reference group) is p_2. Random samples of n_1 and n_2 individuals are obtained from these two populations. The data from these samples can be displayed in a 2-by-2 contingency table as follows

Group	Success	Failure	Total
Treatment	x_{11}	x_{12}	n_1
Control	x_{21}	x_{22}	n_2
Totals	m_1	m_2	N

The binomial proportions, p_1 and p_2, are estimated from these data using the formulae

$$\hat{p}_1 = \frac{a}{m} = \frac{x_{11}}{n_1} \text{ and } \hat{p}_2 = \frac{b}{n} = \frac{x_{21}}{n_2}$$

Let $p_{1.0}$ represent the group 1 proportion tested by the null hypothesis, H_0. The power of a test is computed at a specific value of the proportion which we will call $p_{1.1}$. Let δ represent the smallest difference (margin of equivalence) between the two proportions that still results in the conclusion that the new treatment is not inferior to the current treatment. For a non-inferiority test, $\delta < 0$. The set of statistical hypotheses that are tested is

$$\mathrm{H}_0: p_{1.0} - p_2 \le \delta \text{ versus } \mathrm{H}_1: p_{1.0} - p_2 > \delta$$

which can be rearranged to give

$$\mathrm{H}_0: p_{1.0} \le p_2 + \delta \text{ versus } \mathrm{H}_1: p_{1.0} > p_2 + \delta$$

There are three common methods of specifying the margin of equivalence. The most direct is to simply give values for p_2 and $p_{1.0}$. However, it is often more meaningful to give p_2 and then specify $p_{1.0}$ implicitly by specifying the difference, ratio, or odds ratio. Mathematically, the definitions of these parameterizations are

Parameter	Computation	Hypotheses
Difference	$\delta = p_{1.0} - p_2$	$H_0: p_{1.0} - p_2 \le \delta_0$ vs. $H_1: p_{1.0} - p_2 > \delta_0, \quad \delta_0 < 0$
Ratio	$\phi = p_{1.0} / p_2$	$H_0: p_1 / p_2 \le \phi_0$ vs. $H_1: p_1 / p_2 > \phi_0, \quad \phi_0 < 1$
Odds Ratio	$\psi = Odds_{1.0} / Odds_2$	$H_0: o_{1.0} / o_2 \le \psi_0$ versus $H_1: o_{1.0} / o_2 > \psi_0, \quad \psi_0 < 1$

Difference

The difference is perhaps the most direct method of comparison between two proportions. It is easy to interpret and communicate. It gives the absolute impact of the treatment. However, there are subtle difficulties that can arise with its interpretation.

One difficulty arises when the event of interest is rare. If a difference of 0.001 occurs when the baseline probability is 0.40, it would be dismissed as being trivial. However, if the baseline probably of a disease is 0.002, a 0.001 decrease would represent a reduction of 50%. Thus interpretation of the difference depends on the baseline probability of the event.

Note that if $\delta < 0$, the procedure is called a *non-inferiority test* while if $\delta > 0$ the procedure is called a *superiority test*.

Non-Inferiority using a Difference

The following example might help you understand the concept of a *non-inferiority* test. Suppose 60% of patients respond to the current treatment method $(p_2 = 0.60)$. If the response rate of the new treatment is no less than 5 percentage points worse $(\delta = -0.05)$ than the existing treatment, it will be considered to be noninferior. Substituting these figures into the statistical hypotheses gives

$$H_0: \delta \le -0.05 \text{ versus } H_1: \delta > -0.05$$

Using the relationship

$$p_{1.0} = p_2 + \delta$$

gives

$$H_0: p_{1.0} \le 0.55 \text{ versus } H_1: p_{1.0} > 0.55$$

In this example, when the null hypothesis is rejected, the concluded alternative is that the response rate is at least 55%, which means that the new treatment is not inferior to the current treatment.

Superiority using a Difference

The following example is intended to help you understand the concept of a *superiority* test. Suppose 60% of patients respond to the current treatment method $(p_2 = 0.60)$. If the response rate of the new treatment is at least 10 percentage points better $(\delta = 0.10)$, it will be considered to be superior to the existing treatment. Substituting these figures into the statistical hypotheses gives

$$H_0: \delta \le 0.10 \text{ versus } H_1: \delta > 0.10$$

Using the relationship

$$p_{1.0} = p_2 + \delta$$

gives

$$H_0: p_{1.0} \le 0.70 \text{ versus } H_1: p_{1.0} > 0.70$$

In this example, when the null hypothesis is rejected, the concluded alternative is that the response rate is at least 0.70. That is, the conclusion of superiority is that the new treatment's response rate is at least 0.10 more than that of the existing treatment.

Ratio

The ratio, $\phi = p_{1.0} / p_2$, gives the relative change in the probability of the response. Testing non-inferiority and superiority use the formulation

$$H_0: p_{1.0} / p_2 \le \phi_0 \text{ versus } H_1: p_{1.0} / p_2 > \phi_0$$

The only subtlety is that for non-inferiority tests $\phi_0 < 1$, while for superiority tests $\phi_0 > 1$.

Non-Inferiority using a Ratio

The following example might help you understand the concept of *non-inferiority* as defined by the ratio. Suppose that 60% of patients $(p_2 = 0.60)$ respond to the current treatment method. If a new treatment decreases the response rate by no more than 10% $(\phi_0 = 0.90)$, it will be considered to be noninferior to the standard treatment. Substituting these figures into the statistical hypotheses gives

$$H_0: \phi \le 0.90 \text{ versus } H_1: \phi > 0.90$$

Using the relationship

$$p_{1.0} = \phi_0 p_2$$

gives

$$H_0: p_{1.0} \le 0.54 \text{ versus } H_1: p_{1.0} > 0.54$$

In this example, when the null hypothesis is rejected, the concluded alternative is that the response rate is at least 54%. That is, the conclusion of non-inferiority is that the new treatment's response rate is no worse than 10% less than that of the standard treatment.

Odds Ratio

The odds ratio, $\psi = \left(p_{1.0} / \left(1 - p_{1.0}\right)\right) / \left(p_2 / \left(1 - p_2\right)\right)$, gives the relative change in the odds of the response. Testing non-inferiority and superiority use the same formulation

$$H_0: \psi \le \psi_0 \text{ versus } H_1: \psi > \psi_0$$

The only difference is that for non-inferiority tests $\psi_0 < 1$, while for superiority tests $\psi_0 > 1$.

A Note on Setting the Significance Level, Alpha

Setting the significance level has always been somewhat arbitrary. For planning purposes, the standard has become to set alpha to 0.05 for two-sided tests. Almost universally, when someone states that a result is statistically significant, they mean statistically significant at the 0.05 level.

Although 0.05 may be the standard for two-sided tests, it is not always the standard for one-sided tests, such as non-inferiority tests. Statisticians often recommend that the alpha level for one-sided tests be set at 0.025 since this is the amount put in each tail of a two-sided test.

Power Calculation

The power for a test statistic that is based on the normal approximation can be computed exactly using two binomial distributions. The following steps are taken to compute the power of these tests.

1. Find the critical value using the standard normal distribution. The critical value, $z_{critical}$, is that value of *z* that leaves exactly the target value of alpha in the appropriate tail of the normal distribution.
2. Compute the value of the test statistic, z_t, for every combination of x_{11} and x_{21}. Note that x_{11} ranges from 0 to n_1, and x_{21} ranges from 0 to n_2. A small value (around 0.0001) can be added to the zero-cell counts to avoid numerical problems that occur when the cell value is zero.
3. If $z_t > z_{critical}$, the combination is in the rejection region. Call all combinations of x_{11} and x_{21} that lead to a rejection the set *A*.
4. Compute the power for given values of $p_{1.1}$ and p_2 as

$$1 - \beta = \sum_A \binom{n_1}{x_{11}} p_{1.1}^{x_{11}} q_{1.1}^{n_1 - x_{11}} \binom{n_2}{x_{21}} p_2^{x_{21}} q_2^{n_2 - x_{21}}$$

5. Compute the actual value of alpha achieved by the design by substituting p_2 for $p_{1.1}$ to obtain

$$\alpha^* = \sum_A \binom{n_1}{x_{11}} \binom{n_2}{x_{21}} p_2^{x_{11} + x_{21}} q_2^{n_1 + n_2 - x_{11} - x_{21}}$$

Asymptotic Approximations

When the values of n_1 and n_2 are large (say over 200), these formulas often take a long time to evaluate. In this case, a large sample approximation can be used. The large sample approximation is made by replacing the values of $\hat{p}_1$ and $\hat{p}_2$ in the *z* statistic with the corresponding values of $p_{1.1}$ and p_2, and then computing the results based on the normal distribution. Note that in large samples, the Farrington and Manning statistic is substituted for the Gart and Nam statistic. Also, for large samples, the results for the odds ratio have not (to our knowledge) been published. In this case, we substitute the calculations based on the ratio formulation.

Test Statistics

Several test statistics have been proposed for testing whether the difference, ratio, or odds ratio are different from a specified value. The main difference among the several test statistics is in the formula used to compute the standard error used in the denominator. These tests are based on the following *z*-test

$$z_t = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0 - c}{\hat{\sigma}}$$

The constant, *c*, represents a continuity correction that is applied in some cases. When the continuity correction is not used, *c* is zero. In power calculations, the values of $\hat{p}_1$ and $\hat{p}_2$ are not known. The corresponding values of $p_{1.1}$ and p_2 may be reasonable substitutes.

Following is a list of the test statistics available in ***PASS***. The availability of several test statistics begs the question of which test statistic one should use. The answer is simple: one should use the test statistic that will be used to analyze the data. You may choose a method because it is a standard in your industry, because it seems to have better statistical properties, or because your statistical package calculates it. Whatever your reasons for selecting a certain test statistic, you should use the same test statistic when doing the analysis after the data have been collected.

Z Test (Pooled)

This test was first proposed by Karl Pearson in 1900. Although this test is usually expressed directly as a chi-square statistic, it is expressed here as a *z* statistic so that it can be more easily used for one-sided hypothesis testing. The proportions are pooled (averaged) in computing the standard error. The formula for the test statistic is

$$z_t = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0}{\hat{\sigma}_1}$$

where

$$\hat{\sigma}_1 = \sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

$$\bar{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$$

Z Test (Unpooled)

This test statistic does not pool the two proportions in computing the standard error.

$$z_t = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0}{\hat{\sigma}_2}$$

where

$$\hat{\sigma}_2 = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

Z Test with Continuity Correction (Pooled)

This test is the same as Z Test (Pooled), except that a continuity correction is used. Remember that in the null case, the continuity correction makes the results closer to those of Fisher's Exact test.

$$z_t = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0 + \frac{F}{2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}{\hat{\sigma}_1}$$

$$\hat{\sigma}_1 = \sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

$$\bar{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$$

where F is -1 for lower-tailed hypotheses and 1 for upper-tailed hypotheses.

Z Test with Continuity Correction (Unpooled)

This test is the same as the Z Test (Unpooled), except that a continuity correction is used. Remember that in the null case, the continuity correction makes the results closer to those of Fisher's Exact test.

$$z_t = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0 - \frac{F}{2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}{\hat{\sigma}_2}$$

$$\hat{\sigma}_2 = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

where F is -1 for lower-tailed hypotheses and 1 for upper-tailed hypotheses.

T-Test of Difference

Because of a detailed, comparative study of the behavior of several tests, D'Agostino (1988) and Upton (1982) proposed using the usual two-sample t-test for testing whether the two proportions

are equal. One substitutes a '1' for a success and a '0' for a failure in the usual, two-sample *t*-test formula.

Miettinen and Nurminen's Likelihood Score Test of the Difference

Miettinen and Nurminen (1985) proposed a test statistic for testing whether the difference is equal to a specified, non-zero, value, δ_0. The regular MLE's, $\hat{p}_1$ and $\hat{p}_2$, are used in the numerator of the score statistic while MLE's $\tilde{p}_1$ and $\tilde{p}_2$, constrained so that $\tilde{p}_1 - \tilde{p}_2 = \delta_0$, are used in the denominator. A correction factor of *N*/(*N*-1) is applied to make the variance estimate less biased. The significance level of the test statistic is based on the asymptotic normality of the score statistic. The formula for computing this test statistic is

$$z_{MND} = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0}{\hat{\sigma}_{MND}}$$

where

$$\hat{\sigma}_{MND} = \sqrt{\left(\frac{\tilde{p}_1\tilde{q}_1}{n_1} + \frac{\tilde{p}_2\tilde{q}_2}{n_2}\right)\left(\frac{N}{N-1}\right)}$$

$$\tilde{p}_1 = \tilde{p}_2 + \delta_0$$

$$\tilde{p}_1 = 2B\cos(A) - \frac{L_2}{3L_3}$$

$$A = \frac{1}{3}\left[\pi + \cos^{-1}\left(\frac{C}{B^3}\right)\right]$$

$$B = \text{sign}(C)\sqrt{\frac{L_2^2}{9L_3^2} - \frac{L_1}{3L_3}}$$

$$C = \frac{L_2^3}{27L_3^3} - \frac{L_1L_2}{6L_3^2} + \frac{L_0}{2L_3}$$

$$L_0 = x_{21}\delta_0(1-\delta_0)$$

$$L_1 = \left[N_2\delta_0 - N - 2x_{21}\right]\delta_0 + M_1$$

$$L_2 = (N + N_2)\delta_0 - N - M_1$$

$$L_3 = N$$

Miettinen and Nurminen's Likelihood Score Test of the Ratio

Miettinen and Nurminen (1985) proposed a test statistic for testing whether the ratio is equal to a specified value ϕ_0. The regular MLE's, $\hat{p}_1$ and $\hat{p}_2$, are used in the numerator of the score statistic while MLE's $\tilde{p}_1$ and $\tilde{p}_2$, constrained so that $\tilde{p}_1 / \tilde{p}_2 = \phi_0$, are used in the denominator. A correction factor of N/(N-1) is applied to make the variance estimate less biased. The significance level of the test statistic is based on the asymptotic normality of the score statistic.

The formula for computing the test statistic is

$$z_{MNR} = \frac{\hat{p}_1 / \hat{p}_2 - \phi_0}{\sqrt{\left(\frac{\tilde{p}_1\tilde{q}_1}{n_1} + \phi_0^2\frac{\tilde{p}_2\tilde{q}_2}{n_2}\right)\left(\frac{N}{N-1}\right)}}$$

where

$$\tilde{p}_1 = \tilde{p}_2\phi_0$$

$$\tilde{p}_2 = \frac{-B - \sqrt{B^2 - 4AC}}{2A}$$

$$A = N\phi_0$$

$$B = -\left[N_1\phi_0 + x_{11} + N_2 + x_{21}\phi_0\right]$$

$$C = M_1$$

Miettinen and Nurminen's Likelihood Score Test of the Odds Ratio

Miettinen and Nurminen (1985) proposed a test statistic for testing whether the odds ratio is equal to a specified value, ψ_0. Because the approach they used with the difference and ratio does not easily extend to the odds ratio, they used a score statistic approach for the odds ratio. The regular MLE's are $\hat{p}_1$ and $\hat{p}_2$. The constrained MLE's are $\tilde{p}_1$ and $\tilde{p}_2$. These estimates are constrained so that $\tilde{\psi} = \psi_0$. A correction factor of N/(N-1) is applied to make the variance estimate less biased. The significance level of the test statistic is based on the asymptotic normality of the score statistic.

The formula for computing the test statistic is

$$z_{MNO} = \frac{\frac{(\hat{p}_1 - \tilde{p}_1)}{\tilde{p}_1\tilde{q}_1} - \frac{(\hat{p}_2 - \tilde{p}_2)}{\tilde{p}_2\tilde{q}_2}}{\sqrt{\left(\frac{1}{N_2\tilde{p}_1\tilde{q}_1} + \frac{1}{N_2\tilde{p}_2\tilde{q}_2}\right)\left(\frac{N}{N-1}\right)}}$$

where

$$\tilde{p}_1 = \frac{\tilde{p}_2\psi_0}{1 + \tilde{p}_2(\psi_0 - 1)}$$

$$\tilde{p}_2 = \frac{-B + \sqrt{B^2 - 4AC}}{2A}$$

$$A = N_2(\psi_0 - 1)$$

$$B = N_1\psi_0 + N_2 - M_1(\psi_0 - 1)$$

$$C = -M_1$$

Farrington and Manning's Likelihood Score Test of the Difference

Farrington and Manning (1990) proposed a test statistic for testing whether the difference is equal to a specified value δ_0. The regular MLE's, $\hat{p}_1$ and $\hat{p}_2$, are used in the numerator of the score statistic while MLE's $\tilde{p}_1$ and $\tilde{p}_2$, constrained so that $\tilde{p}_1 - \tilde{p}_2 = \delta_0$, are used in the denominator. The significance level of the test statistic is based on the asymptotic normality of the score statistic.

The formula for computing the test statistic is

$$z_{FMD} = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0}{\sqrt{\left(\frac{\tilde{p}_1\tilde{q}_1}{n_1} + \frac{\tilde{p}_2\tilde{q}_2}{n_2}\right)}}$$

where the estimates $\tilde{p}_1$ and $\tilde{p}_2$ are computed as in the corresponding test of Miettinen and Nurminen (1985) given above.

Farrington and Manning's Likelihood Score Test of the Ratio

Farrington and Manning (1990) proposed a test statistic for testing whether the ratio is equal to a specified value ϕ_0. The regular MLE's, $\hat{p}_1$ and $\hat{p}_2$, are used in the numerator of the score statistic while MLE's $\tilde{p}_1$ and $\tilde{p}_2$, constrained so that $\tilde{p}_1 / \tilde{p}_2 = \phi_0$, are used in the denominator. A correction factor of N/(N-1) is applied to increase the variance estimate. The significance level of the test statistic is based on the asymptotic normality of the score statistic.

The formula for computing the test statistic is

$$z_{FMR} = \frac{\hat{p}_1 / \hat{p}_2 - \phi_0}{\sqrt{\left(\frac{\tilde{p}_1\tilde{q}_1}{n_1} + \phi_0^2\frac{\tilde{p}_2\tilde{q}_2}{n_2}\right)}}$$

where the estimates $\tilde{p}_1$ and $\tilde{p}_2$ are computed as in the corresponding test of Miettinen and Nurminen (1985) given above.

Farrington and Manning's Likelihood Score Test of the Odds Ratio

Farrington and Manning (1990) indicate that the Miettinen and Nurminen statistic may be modified by removing the factor $N/(N\text{-}1)$.

The formula for computing this test statistic is

$$z_{FMO} = \frac{\frac{\left(\hat{p}_1 - \tilde{p}_1\right)}{\tilde{p}_1\tilde{q}_1} - \frac{\left(\hat{p}_2 - \tilde{p}_2\right)}{\tilde{p}_2\tilde{q}_2}}{\sqrt{\left(\frac{1}{N_2\tilde{p}_1\tilde{q}_1} + \frac{1}{N_2\tilde{p}_2\tilde{q}_2}\right)}}$$

where the estimates $\tilde{p}_1$ and $\tilde{p}_2$ are computed as in the corresponding test of Miettinen and Nurminen (1985) given above.

Gart and Nam's Likelihood Score Test of the Difference

Gart and Nam (1990), page 638, proposed a modification to the Farrington and Manning (1988) difference test that corrects for skewness. Let $z_{FMD}(\delta)$ stand for the Farrington and Manning difference test statistic described above. The skewness corrected test statistic, z_{GND}, is the appropriate solution to the quadratic equation

$$(-\tilde{\gamma})z_{GND}^2 + (-1)z_{GND} + \left(z_{FMD}(\delta) + \tilde{\gamma}\right) = 0$$

where

$$\tilde{\gamma} = \frac{\tilde{V}^{3/2}(\delta)}{6}\left(\frac{\tilde{p}_1\tilde{q}_1(\tilde{q}_1 - \tilde{p}_1)}{n_1^2} - \frac{\tilde{p}_2\tilde{q}_2(\tilde{q}_2 - \tilde{p}_2)}{n_2^2}\right)$$

Gart and Nam's Likelihood Score Test of the Ratio

Gart and Nam (1988), page 329, proposed a modification to the Farrington and Manning (1988) ratio test that corrects for skewness. Let $z_{FMR}(\phi)$ stand for the Farrington and Manning ratio test statistic described above. The skewness corrected test statistic, z_{GNR}, is the appropriate solution to the quadratic equation

$$(-\tilde{\varphi})z_{GNR}^2 + (-1)z_{GNR} + \left(z_{FMR}(\phi) + \tilde{\varphi}\right) = 0$$

where

$$\tilde{\varphi} = \frac{1}{6\tilde{u}^{3/2}}\left(\frac{\tilde{q}_1(\tilde{q}_1 - \tilde{p}_1)}{n_1^2\tilde{p}_1^2} - \frac{\tilde{q}_2(\tilde{q}_2 - \tilde{p}_2)}{n_2^2\tilde{p}_2^2}\right)$$

$$\tilde{u} = \frac{\tilde{q}_1}{n_1\tilde{p}_1} + \frac{\tilde{q}_2}{n_2\tilde{p}_2}$$

Procedure Options

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

Data Tab (Common Options)

The Data tab contains the parameters associated with this test such as the proportions, sample sizes, alpha, and power. This chapter covers four procedures, each of which has different options. This section documents options that are common to all four procedures. Later, unique options for each procedure will be documented.

Solve For

Find (Solve For)

This option specifies the parameter to be solved for using the other parameters. The parameters that may be selected are *P1.1*, *Alpha*, *Power and Beta*, *N1*, and *N2*. Under most situations, you will select either *Power and Beta* or *N1*.

Select N1 when you want to calculate the sample size needed to achieve a given power and alpha level.

Select *Power and Beta* when you want to calculate the power of an experiment.

Error Rates

Power or Beta

This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected.

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

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

Alpha (Significance Level)

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected.

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

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

Sample Size

N1 (Sample Size Group 1)

Enter a value (or range of values) for the sample size of this group. 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.

Effect Size – Reference (Group 2)

P2 (Reference Group Proportion)

Specify the value of p_2 , the reference, baseline, or control group's proportion. The null hypothesis is that the two proportions differ by no more than a specified amount. Since P2 is a proportion, these values must be between 0 and 1.

You may enter a range of values such as *0.1 0.2 0.3* or *0.1 to 0.9 by 0.1*.

Test

Higher Proportions Are

This option specifies whether proportions represent successes (better) or failures (worse).

- **Better (Successes)**

  When proportions represent successes, higher proportions are better. A noninferior treatment is one whose proportion is at least almost as high as that of the reference group.

  For testing non-inferiority, D0 is negative, R0 is less than 1, and OR0 is less than 1. For testing superiority, D0 is positive, R0 is greater than 1, and OR0 is greater than 1.

- **Worse (Failures)**

  When proportions represent failures, lower proportions are better. A noninferior treatment is one whose proportion is at most almost as low as that of the reference group.

  For testing non-inferiority, D0 is positive, R0 is greater than 1, and OR0 is greater than 1. For testing superiority, D0 is negative, R0 is less than 1, and OR0 is less than 1.

Test Type

Specify which test statistic is used in searching and reporting. Although the pooled *z*-test is commonly shown in elementary statistics books, the likelihood score test is arguably the best choice.

Note that *C.C.* is an abbreviation for *Continuity Correction*. This refers to the adding or subtracting 1/(2n) to (or from) the numerator of the z-value to bring the normal approximation closer to the binomial distribution.

Data Tab (Proportions)

This section documents options that are used when the parameterization is in terms of the values of the two proportions, P1 and P2. P1.0 is the value of the P1 assumed by the null hypothesis and P1.1 is the value of P1 at which the power is calculated.

Effect Size – Treatment (Group 1)

P1.0 (Equivalence Proportion)

This option allows you to specify the value P1.0 directly. This is that value of treatment group's proportion above which the treatment group is considered noninferior to the reference group.

When *Higher Proportions Are* is set to *Better*, the trivial proportion is the smallest value of P1 for which the treatment group is declared noninferior to the reference group. In this case, P1.0 should be less than P2 for non-inferiority tests and greater than P2 for superiority tests. The reverse is the case when *Higher Proportions Are* is set to *Worse*.

Proportions must be between 0 and 1. They cannot take on the values 0 or 1. This value should not be set to exactly the value of P2.

You may enter a range of values such as *0.03 0.05 0.10* or *0.01 to 0.05 by 0.01*.

P1.1 (Actual Proportion)

This option specifies the value of P1.1 which is the value of the treatment proportion at which the power is to be calculated. Proportions must be between 0 and 1. They cannot take on the values 0 or 1.

You may enter a range of values such as *0.03 0.05 0.10* or *0.01 to 0.05 by 0.01*.

Data Tab (Differences)

This section documents options that are used when the parameterization is in terms of the difference, P1 – P2. P1.0 is the value of P1 assumed by the null hypothesis and P1.1 is the value of P1 at which the power is calculated. Once P2, D0, and D1 are given, the values of P1.1 and P1.0 can be calculated.

Effect Size – Differences

D0 (Equivalence Difference)

This option specifies the trivial difference (often called the *margin of error*) between P1.0 (the value of P1 under H0) and P2. This difference is used with P2 to calculate the value of P1.0 using the formula: P1.0 = P2 + D0.

When *Higher Proportions Are* is set to *Better*, the trivial difference is that amount by which P1 can be less than P2 and still have the treatment group declared noninferior to the reference group. In this case, D0 should be negative for non-inferiority tests and positive for superiority tests.

The reverse is the case when *Higher Proportions Are* is set to *worse*.

You may enter a range of values such as *-.03 -.05 -.10* or *-.05 to -.01 by .01*. Differences must be between -1 and 1. D0 cannot take on the values -1, 0, or 1.

D1 (Actual Difference)

This option specifies the actual difference between P1.1 (the actual value of P1) and P2. This is the value of the difference at which the power is calculated. In non-inferiority trials, this difference is often set to 0.

The power calculations assume that P1.1 is the actual value of the proportion in group 1 (experimental or treatment group). This difference is used with P2 to calculate the value of P1 using the formula: P1.1 = D1 + P2.

You may enter a range of values such as *-.05 0 .5* or *-.05 to .05 by .02*. Actual differences must be between -1 and 1. They cannot take on the values -1 or 1.

Data Tab (Ratios)

This section documents options that are used when the parameterization is in terms of the ratio, P1 / P2. P1.0 is the value of P1 assumed by the null hypothesis and P1.1 is the value of P1 at which the power is calculated. Once P2, R0, and R1 are given, the values of P1.0 and P1.1 can be calculated.

Effect Size – Ratios

R0 (Equivalence Ratio)

This option specifies the trivial ratio (also called the Relative Margin of Equivalence) between P1.0 and P2. The power calculations assume that P1.0 is the value of the P1 under the null hypothesis. This value is used with P2 to calculate the value of P1.0 using the formula: P1.0 = R0 x P2.

When *Higher Proportions Are* is set to *Better*, the trivial ratio is the relative amount by which P1 can be less than P2 and still have the treatment group declared noninferior to the reference group. In this case, R0 should be less than one for non-inferiority tests and greater than 1 for superiority tests. The reverse is the case when *Higher Proportions Are* is set to *Worse*.

Ratios must be positive. R0 cannot take on the value of 1. You may enter a range of values such as *0.95 .97 .99* or *.91 to .99 by .02*.

R1 (Actual Ratio)

This option specifies the ratio of P1.1 and P2, where P1.1 is the actual proportion in the treatment group. The power calculations assume that P1.1 is the actual value of the proportion in group 1. This difference is used with P2 to calculate the value of P1 using the formula: P1.1 = R1 x P2. In non-inferiority trials, this ratio is often set to 1.

Ratios must be positive. You may enter a range of values such as *0.95 1 1.05* or *0.9 to 1.9 by 0.02*.

Data Tab (Odds Ratios)

This section documents options that are used when the parameterization is in terms of the odds ratios, O1.1 / O2 and O1.0 / O2. Note that the odds are defined as O2 = P2 / (1 – P2), O1.0 = P1.0 / (1 – P1.0), etc. P1.0 is the value of P1 assumed by the null hypothesis and P1.1 is the value of P1 at which the power is calculated. Once P2, OR0, and OR1 are given, the values of P1.1 and P1.0 can be calculated.

Effect Size – Odds Ratios

OR0 (Equivalence Odds Ratio)

This option specifies the trivial odds ratio between P1.0 and P2. The power calculations assume that P1.0 is the value of the P1 under the null hypothesis. OR0 is used with P2 to calculate the value of P1.0.

When *Higher Proportions Are* is set to *Better*, the trivial odds ratio implicitly gives the amount by which P1 can be less than P2 and still have the treatment group declared noninferior to the reference group. In this case, OR0 should be less than 1 for non-inferiority tests and greater than 1 for superiority tests. The reverse is the case when *Higher Proportions Are* is set to *Worse*.

Odds ratios must be positive. OR0 cannot take on the value of 1.

You may enter a range of values such as *0.95 0.97 0.99* or *0.91 to 0.99 by 0.02*.

OR1 (Actual Odds Ratio)

This option specifies the odds ratio of P1.1 and P2, where P1.1 is the actual proportion in the treatment group. The power calculations assume that P1.1 is the actual value of the proportion in group 1. This value is used with P2 to calculate the value of P1. In non-inferiority trials, this odds ratio is often set to 1.

Odds ratios must be positive. You may enter a range of values such as *0.95 1 1.05* or *0.9 to 1 by 0.02*.

Options Tab

The Options tab contains various limits and options.

Maximum Iterations

Maximum Iterations Before Search Termination

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 not reported. A value of at least 500 is recommended.

Zero Counts

Zero Count Adjustment Method

Zero cell counts cause many calculation problems. To compensate for this, a small value (called the Zero Count Adjustment Value) can be added either to all cells or to all cells with zero counts. This option specifies whether you want to use the adjustment and which type of adjustment you want to use. We recommend that you use the option ‘Add to zero cells only.’

Zero cell values often do not occur in practice. However, since power calculations are based on total enumeration, they will occur in power and sample size estimation.

Adding a small value is controversial, but can be necessary for computational considerations. Statisticians have recommended adding various fractions to zero counts. We have found that adding 0.0001 seems to work well.

Zero Count Adjustment Value

Zero cell counts cause many calculation problems when computing power or sample size. To compensate for this, a small value may be added either to all cells or to all zero cells. This is the amount that is added. We have found that 0.0001 works well.

Be warned that the value of the ratio and the odds ratio will be affected by the amount specified here!

Exact Test Options

Maximum N1 or N2 for Exact Calculations

When either N1 or N2 is above this amount, power calculations are based on the normal approximation to the binomial. In this case, the actual value of alpha is not calculated. Currently, for three-gigahertz computers, a value near 200 is reasonable. As computers get faster, this number may be increased.

Example 1 – Finding Power

A study is being designed to establish the non-inferiority of a new treatment compared to the current treatment. Historically, the current treatment has enjoyed a 60% cure rate. The new treatment reduces the seriousness of certain side effects that occur with the current treatment. Thus, the new treatment will be adopted even if it is slightly less effective than the current treatment. The researchers will recommend adoption of the new treatment if it has a cure rate of at least 55%.

The researchers plan to use the Farrington and Manning likelihood score test statistic to analyze the data that will be (or has been) obtained. They want to study the power of the Farrington and Manning test at group sample sizes ranging from 50 to 500 for detecting a difference of -0.05 when the actual cure rate of the new treatment ranges from 57% to 70%. The significance level will be 0.025.

Setup

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

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power	*Ignored since this is the Find setting*
Alpha	**0.025**
N1 (Sample Size Group 1)	**50 to 500 by 50**
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
D0 (Non-Inferiority Difference)	**-0.05**
D1 (Actual Difference)	**-0.03 0.00 0.05 0.10**
P2 (Reference Group Proportion)	**0.6**
Test Type	**Likelihood Score (Farr. & Mann.)**
Higher Proportions Are	**Better**
Options Tab	
Maximum N1 or N2 Exact	**300**

Annotated Output

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

Numeric Results

Numeric Results of Non-Inferiority Tests Based on the Difference: P1 – P2
H0: P1-P2<=D0. H1: P1-P2=D1>D0. Test Statistic: Score test (Farrington & Manning)

Power	Sample Size Grp 1 N1	Sample Size Grp 2 N2	Grp 2 Prop P2	Equiv. Grp 1 Prop P1.0	Actual Grp 1 Prop P1.1	Equiv. Margin Diff D0	Actual Margin Diff D1	Target Alpha	Actual Alpha	Beta
0.0380	50	50	0.6000	0.5500	0.5700	-0.0500	-0.0300	0.0250	0.0236	0.9620
0.0494	100	100	0.6000	0.5500	0.5700	-0.0500	-0.0300	0.0250	0.0267	0.9506
0.0525	150	150	0.6000	0.5500	0.5700	-0.0500	-0.0300	0.0250	0.0241	0.9475
0.0588	200	200	0.6000	0.5500	0.5700	-0.0500	-0.0300	0.0250	0.0244	0.9412
0.0650	250	250	0.6000	0.5500	0.5700	-0.0500	-0.0300	0.0250	0.0241	0.9350
0.0735	300	300	0.6000	0.5500	0.5700	-0.0500	-0.0300	0.0250	0.0261	0.9265
0.0776	350	350	0.6000	0.5500	0.5700	-0.0500	-0.0300	0.0250		0.9224
0.0832	400	400	0.6000	0.5500	0.5700	-0.0500	-0.0300	0.0250		0.9168
0.0886	450	450	0.6000	0.5500	0.5700	-0.0500	-0.0300	0.0250		0.9114

Report continues …

Note: exact results based on the binomial were only calculated when both N1 and N2 were less than 300.

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 corresponding groups.
'P2' is the response rate for group two which is the standard, reference, baseline, or control group.
'P1.0' is the smallest treatment-group response rate that still yields a non-inferiority conclusion.
'P1.1' is the treatment-group response rate at which the power is calculated.
'D0' is the non-inferiority margin. It is the difference P1-P2 assuming H0.
'D1' is the actual difference, P1-P2, at which the power is calculated.
'Target Alpha' is the probability of rejecting a true null hypothesis that was desired.
'Actual Alpha' is the value of alpha that is actually achieved.
'Beta' is the probability of accepting a false H0. Beta = 1 - Power.
'Grp 1' refers to Group 1 which is the treatment or experimental group.
'Grp 2' refers to Group 2 which is the reference, standard, or control group.
' Equiv.' refers to a small amount that is not of practical importance.
'Actual' refers to the true value at which the power is computed.

Summary Statements
Sample sizes of 50 in group one and 50 in group two achieve 4% power to detect a non-inferiority margin difference between the group proportions of -0.0500. The reference group proportion is 0.6000. The treatment group proportion is assumed to be 0.5500 under the null hypothesis of inferiority. The power was computed at for the case when the actual treatment group proportion is 0.5700. The test statistic used is the one-sided Score test (Farrington & Manning). The significance level of the test was targeted at 0.0250. The significance level actually achieved by this design is 0.0236.

This report shows the values of each of the parameters, one scenario per row. Note that the actual alpha value is blank for sample sizes greater than 300, which was the limit set for exact computation.

Most of the report columns have obvious interpretations. Those that may not be obvious are presented here. Note that the discussion below assumes that higher response rates are better and that non-inferiority testing (rather than superiority testing) is planned.

Prop Grp 2 P2

This is the value of P2, the response rate in the control group.

Equiv. Grp 1 Prop P1.0

This is the value of P1.0, the response rate of the treatment group, as specified by the null hypothesis of inferiority. Values of P1 less than this amount are considered different from P2. Values of P1 greater than this are considered noninferior to the reference group. The difference between this value and P2 is the value of the null hypothesis.

Actual Grp 1 Prop P1.1

This is the value of P1.1, the response rate of the treatment group, at which the power is computed. This is the value of P1 under the alternative hypothesis. The difference between this value and P2 is the value of the alternative hypothesis.

Equiv. Margin Diff D0

This is the value of D0, the difference between the two group proportions under the null hypothesis. This value is often called the *margin of non-inferiority*.

Actual Margin Diff D1

This is the value of D1, the difference between the two group proportions at which the power is computed. This is the value of the difference under the alternative hypothesis.

Target Alpha

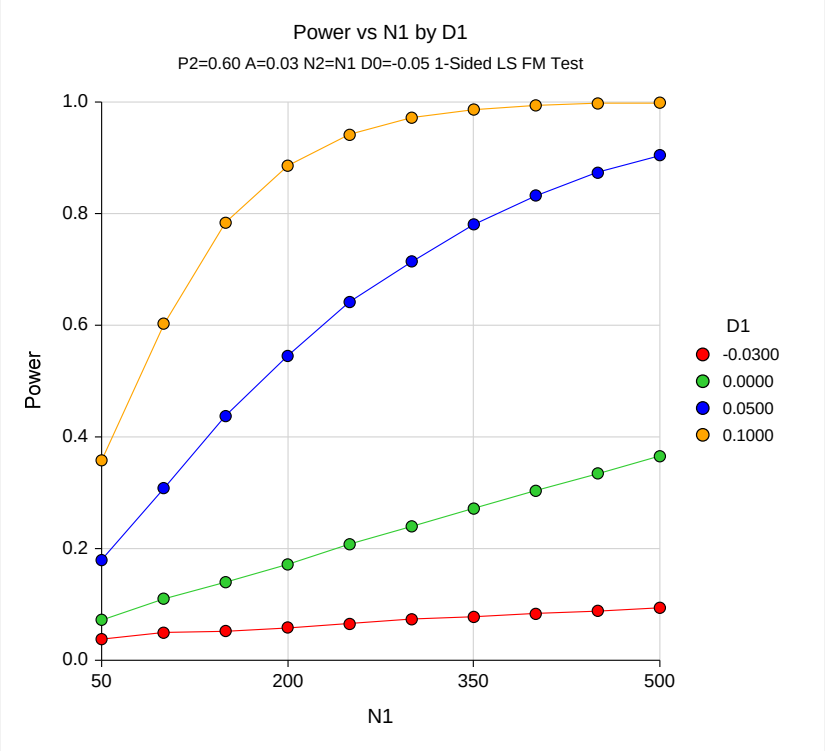
This is the value of alpha that was targeted by the design. Note that the target alpha is not usually achieved exactly. For one-sided tests, this value should usually be 0.025.

Actual Alpha

This is the value of alpha that was actually achieved by this design. Note that since the limit on exact calculations was set to 300, and since this value is calculated exactly, it is not shown for values of N1 greater than 300.

The difference between the Target Alpha and the Actual Alpha is caused by the discrete nature of the binomial distribution and the use of the normal approximation to the binomial in determining the critical value of the test statistic.

Plots Section



The values from the table are displayed in the above chart. This chart gives us a quick look at the sample size that will be required for various values of D1.

Example 2 – Finding the Sample Size

Continuing with the scenario given in Example 1, the researchers want to determine the sample size necessary for each value of D1 to achieve a power of 0.80. To cut down on the runtime, they decide to look at approximate values whenever N1 is greater than 100.

Setup

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

Option	Value
Data Tab	
Find (Solve For)	**N1**
Power	**0.8**
Alpha	**0.025**
N1 (Sample Size Group 1)	*Ignored since this is the Find setting*
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
D0 (Non-Inferiority Difference)	**-0.05**
D1 (Actual Difference)	**-0.03 0.00 0.05 0.10**
P2 (Reference Group Proportion)	**0.6**
Test Type	**Likelihood Score (Farr. & Mann.)**
Higher Proportions Are	**Better**
Options Tab	
Maximum N1 or N2 Exact	**100**

Output

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

Numeric Results

Numeric Results of Non-Inferiority Tests Based on the Difference: P1 – P2
H0: P1-P2<=D0. H1: P1-P2=D1>D0. Test Statistic: Score test (Farrington & Manning)

Power	Sample Size Grp 1 N1	Sample Size Grp 2 N2	Prop Grp 2 P2	Equiv. Grp 1 Prop P1.0	Actual Grp 1 Prop P1.1	Equiv. Margin Diff D0	Actual Margin Diff D1	Target Alpha	Actual Alpha	Beta
0.8000	9509	9509	0.6000	0.5500	0.5700	-0.0500	-0.0300	0.0250		0.2000
0.8001	1505	1505	0.6000	0.5500	0.6000	-0.0500	0.0000	0.0250		0.1999
0.8008	368	368	0.6000	0.5500	0.6500	-0.0500	0.0500	0.0250		0.1992
0.8019	159	159	0.6000	0.5500	0.7000	-0.0500	0.1000	0.0250		0.1981

The required sample size will depend a great deal on the value of D1. Any effort spent determining an accurate value for D1 will be worthwhile.

Example 3 – Comparing the Power of Several Test Statistics

Continuing with Example 1, the researchers want to determine which of the eight possible test statistics to adopt by using the comparative reports and charts that ***PASS*** produces. They decide to compare the powers and actual alphas for various sample sizes between 50 and 200 when D1 is 0.1.

Setup

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

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power	*Ignored since this is the Find setting*
Alpha	**0.025**
N1 (Sample Size Group 1)	**50 100 150 200**
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
D0 (Non-Inferiority Difference)	**-0.05**
D1 (Actual Difference)	**0.10**
P2 (Reference Group Proportion)	**0.6**
Test Type	**Likelihood Score (Farr. & Mann.)**
Higher Proportions Are	**Better**
Reports Tab	
Show Numeric Report	**Not checked**
Show Comparative Reports	**Checked**
Show Definitions	**Not checked**
Show Plots	**Not checked**
Show Comparative Plots	**Checked**
Show Summary Statements	**Not checked**
Options Tab	
Maximum N1 or N2 Exact	**300**

Output

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

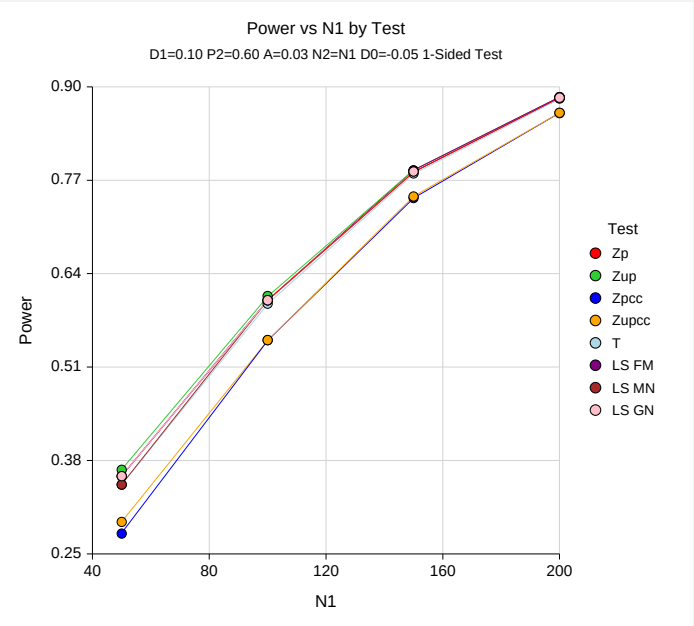
Numeric Results and Plots

Power Comparison of Non-Inferiority Tests Based on the Difference: P1 – P2
H0: P1-P2<=D0. H1: P1-P2=D1>D0.

N1/N2	P2	P1	Target Alpha	Z(P) Test Power	Z(UnP) Test Power	Z(P) CC Test Power	Z(UnP) CC Test Power	T Test Power	F.M. Score Power	M.N. Score Power	G.N. Score Power
50/50	0.6000	0.7000	0.0250	0.3581	0.3670	0.2782	0.2945	0.3464	0.3581	0.3464	0.3581
100/100	0.6000	0.7000	0.0250	0.6030	0.6088	0.5474	0.5475	0.5982	0.6030	0.6030	0.6030
150/150	0.6000	0.7000	0.0250	0.7821	0.7837	0.7453	0.7474	0.7821	0.7837	0.7821	0.7821
200/200	0.6000	0.7000	0.0250	0.8849	0.8857	0.8635	0.8638	0.8849	0.8857	0.8849	0.8849

Actual Alpha Comparison of Non-Inferiority Tests Based on the Difference: P1 – P2
H0: P2-P1<=D0. H1: P2-P1=D1>D0.

N1/N2	P1	P2	Target Alpha	Z(P) Test Alpha	Z(UnP) Test Alpha	Z(P) CC Test Alpha	Z(UnP) CC Test Alpha	T Test Alpha	F.M. Score Alpha	M.N. Score Alpha	G.N. Score Alpha
50/50	0.6000	0.7000	0.0250	0.0236	0.0253	0.0140	0.0161	0.0225	0.0236	0.0225	0.0236
100/100	0.6000	0.7000	0.0250	0.0267	0.0267	0.0190	0.0190	0.0266	0.0267	0.0267	0.0267
150/150	0.6000	0.7000	0.0250	0.0239	0.0241	0.0181	0.0183	0.0239	0.0241	0.0239	0.0239
200/200	0.6000	0.7000	0.0250	0.0243	0.0244	0.0191	0.0191	0.0243	0.0244	0.0243	0.0243



It is interesting to note that the powers of the continuity-corrected test statistics are consistently lower than the other tests. This occurs because the actual alpha achieved by these tests is lower than for the other tests. An interesting finding of this example is that the regular *t*-test performed about as well as the *z*-test.

Example 4 – Validation using Machin with Equal Sample Sizes

Machin et al. (1997), page 106, present a sample size study in which P2 = 0.5, D0 = -0.2, D1=0, one-sided alpha = 0.1, and beta = 0.2. Using the Farrington and Manning test statistic, they found the sample size to be 55 in each group.

Setup

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

Option	Value
Data Tab	
Find (Solve For)	**N1**
Power	**0.8**
Alpha	**0.1**
N1 (Sample Size Group 1)	*Ignored since this is the Find setting*
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
D0 (Non-Inferiority Difference)	**-0.2**
D1 (Actual Difference)	**0.0**
P2 (Reference Group Proportion)	**0.5**
Test Type	**Likelihood Score (Farr. & Mann.)**
Higher Proportions Are	**Better**
Options Tab	
Maximum N1 or N2 Exact	**2 (Set low for a rapid search.)**

Output

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

Numeric Results

Numeric Results of Non-Inferiority Tests Based on the Difference: P1 – P2
H0: P1-P2<=D0. H1: P1-P2=D1>D0. Test Statistic: Score test (Farrington & Manning)

Power	Sample Size Grp 1 N1	Sample Size Grp 2 N2	Prop Grp 2 P2	Equiv. Grp 1 Prop P1.0	Actual Grp 1 Prop P1.1	Equiv. Margin Diff D0	Actual Margin Diff D1	Target Alpha	Actual Alpha	Beta
0.8001	55	55	0.5000	0.3000	0.5000	-0.2000	0.0000	0.1000		0.1999

PASS found the required sample size to be 55 which corresponds to Machin.

Example 5 – Validation of a Superiority Test using Farrington and Manning

Farrington and Manning (1990), page 1451, present a sample size study for a superiority test in which P2 = 0.05, D0 = 0.2, D1=0.35, one-sided alpha = 0.05, and beta = 0.20. Using the Farrington and Manning test statistic, they found the sample size to be 80 in each group. They mention that the true power is 0.813.

Setup

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

Option	Value
Data Tab	
Find (Solve For)	**N1**
Power	**0.80**
Alpha	**0.05**
N1 (Sample Size Group 1)	*Ignored since this is the Find setting*
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
D0 (Non-Inferiority Difference)	**0.2**
D1 (Actual Difference)	**0.35**
P2 (Reference Group Proportion)	**0.05**
Test Type	**Likelihood Score (Farr. & Mann.)**
Higher Proportions Are	**Better**
Options Tab	
Maximum N1 or N2 Exact	**2 (Set low for a rapid search.)**

Output

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

Numeric Results

Numeric Results of Non-Inferiority Tests Based on the Difference: P1 – P2
H0: P1-P2<=D0. H1: P1-P2=D1>D0. Test Statistic: Score test (Farrington & Manning)

Power	Sample Size Grp 1 N1	Sample Size Grp 2 N2	Prop Grp 2 P2	Equiv. Grp 1 Prop P1.0	Actual Grp 1 Prop P1.1	Equiv. Margin Diff D0	Actual Margin Diff D1	Target Alpha	Actual Alpha	Beta
0.8007	80	80	0.0500	0.2500	0.4000	0.2000	0.3500	0.0500		0.1993

PASS also calculated the required sample size to be 80.

Next, to calculate the exact power for this sample size, we make the following changes to the template.

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
N1 (Sample Size Group 1)	**80**
Options Tab	
Maximum N1 or N2 Exact	**200 (**Set >80 to force exact calculation.**)**

Numeric Results

Numeric Results of Non-Inferiority Tests Based on the Difference: P1 – P2
H0: P1-P2<=D0. H1: P1-P2=D1>D0. Test Statistic: Score test (Farrington & Manning)

Power	Sample Size Grp 1 N1	Sample Size Grp 2 N2	Prop Grp 2 P2	Equiv. Grp 1 Prop P1.0	Actual Grp 1 Prop P1.1	Equiv. Margin Diff D0	Actual Margin Diff D1	Target Alpha	Actual Alpha	Beta
0.8132	80	80	0.0500	0.2500	0.4000	0.2000	0.3500	0.0500	0.0553	0.1993

PASS also calculated the exact power to be 0.813.

Example 6 – Validation of Risk Ratio Calculations using Blackwelder

Blackwelder (1993), page 695, presents a table of power values for several scenarios using the risk ratio. The second line of the table presents the results for the following scenario: P2 = 0.04, R0 = 0.3, R1=0.1, N1=N2=1044, one-sided alpha = 0.05, and beta = 0.20. Using the Farrington and Manning likelihood-score test statistic, he found the exact power to be 0.812, the exact alpha to be 0.044, and, using the asymptotic formula, the approximate power to be 0.794.

Setup

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

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power	*Ignored since this is the Find setting*
Alpha	**0.05**
N1 (Sample Size Group 1)	**1044**
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
R0 (Non-Inferiority Ratio)	**0.3**
R1 (Actual Ratio)	**0.1**
P2 (Reference Group Proportion)	**0.04**
Test Type	**Likelihood Score (Farr. & Mann.)**
Higher Proportions Are	**Worse**
Options Tab	
Maximum N1 or N2 Exact	**2000 (Set high for exact results.)**

Output

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

Numeric Results

Numeric Results of Non-Inferiority Tests Based on the Difference: P1 / P2
H0: P1/P2>=R0. H1: P1/P2=R1<R0. Test Statistic: Score test (Farrington & Manning)

Power	Sample Size Grp 1 N1	Sample Size Grp 2 N2	Prop Grp 2 P2	Equiv. Grp 1 Prop P1.0	Actual Grp 1 Prop P1.1	Equiv. Margin Ratio R0	Actual Margin Ratio R1	Target Alpha	Actual Alpha	Beta
0.8118	1044	1044	0.0400	0.0120	0.0040	0.300	0.100	0.0500	0.0444	0.1882

PASS also calculated the power to be 0.812 and the actual alpha to be 0.044, within rounding.

Next, to calculate the asymptotic power, we make the following changes to the template.

Option	Value
Options Tab	
Maximum N1 or N2 Exact	**2 (Set < 1044 to force asymptotic calculation.)**

Numeric Results

Numeric Results of Non-Inferiority Tests Based on the Difference: P1 / P2
H0: P1/P2>=R0. H1: P1/P2=R1<R0. Test Statistic: Score test (Farrington & Manning)

Power	Sample Size Grp 1 N1	Sample Size Grp 2 N2	Prop Grp 2 P2	Equiv. Grp 1 Prop P1.0	Actual Grp 1 Prop P1.1	Equiv. Margin Ratio R0	Actual Margin Ratio R1	Target Alpha	Actual Alpha	Beta
0.7937	1044	1044	0.0400	0.0120	0.0040	0.300	0.100	0.0500		0.2063

PASS also calculated the power to be 0.794.

Example 7 – Finding Power following an Experiment

In an effort to show a new treatment non-inferior to the current standard, researchers randomly assigned 80 subjects to each treatment. The new treatment was to be considered non-inferior if the odds ratio (treatment to standard) was at least 0.80. Using the Farrington and Manning Likelihood Score test, non-inferiority could not be concluded. The researchers now want to see the power of the test. The control proportion was 0.625.

Setup

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

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power	*Ignored since this is the Find setting*
Alpha	**0.05**
N1 (Sample Size Group 1)	**80**
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
OR0 (Non-Inferiority Odds Ratio)	**0.80**
OR1 (Actual Odds Ratio)	**1.0**
P2 (Reference Group Proportion)	**0.625**
Test Type	**Likelihood Score (Farr. & Mann.)**
Higher Proportions Are	**Better**

Output

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

Numeric Results

Numeric Results for Non-Inferiority Tests Based on the Odds Ratio: O1 / O2
H0: O1/O2<=OR0. H1: O1/O2=OR1>OR0. Test Statistic: Score test (Farrington & Manning)

Power	Sample Size Grp 1 N1	Sample Size Grp 2 N2	Grp 2 Prop P2	Equiv. Grp 1 Prop P1.0	Actual Grp 1 Prop P1.1	Equiv. Margin O.R. OR0	Actual Margin O.R. OR1	Target Alpha	Actual Alpha	Beta
0.1845	80	80	0.6250	0.5714	0.6250	0.800	1.000	0.0500	0.0571	0.8155

The power of a test with 80 receiving each treatment is only 0.1801.

Example 8 – Finding True Proportion Difference

Researchers have developed a new treatment with minimal side effects compared to the standard treatment. The researchers are limited by the number of subjects (140 per group) they can use to show the new treatment is non-inferior. The new treatment will be deemed non-inferior if it is at least 0.10 below the success rate of the standard treatment. The standard treatment has a success rate of about 0.75. The researchers want to know how much more successful the new treatment must be (in truth) to yield a test which has 90% power. The test statistic used will be the pooled Z test.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Non-Inferiority Tests for Two Proportions [Differences]** procedure window by expanding **Proportions**, then **Two Independent Proportions**, then clicking on **Non-Inferiority**, and then clicking on **Non-Inferiority Tests for Two Proportions [Differences]**. You may then make the appropriate entries as listed below, or open **Example 8** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**P1.1 (Search>P1.0)**
Power	**0.90**
Alpha	**0.05**
N1 (Sample Size Group 1)	**140**
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
D0 (Non-Inferiority Difference)	**-0.10**
D1 (Actual Difference)	*Ignored since this is the Find setting*
P2 (Reference Group Proportion)	**0.75**
Test Type	**Z Test (Pooled)**
Higher Proportions Are	**Better**
Options Tab	
Maximum N1 or N2 Exact	**500 (Set high for exact results.)**

Output

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

Numeric Results

Numeric Results for Non-Inferiority Tests Based on the Difference: P1 - P2
H0: P1-P2<=D0. H1: P1-P2=D1>D0. Test Statistic: Z test (pooled)

Power	Sample Size Grp 1 N1	Sample Size Grp 2 N2	Grp 2 Prop P2	Equiv. Grp 1 Prop P1.0	Actual Grp 1 Prop P1.1	Equiv. Margin Diff D0	Actual Margin Diff D1	Target Alpha	Actual Alpha	Beta
0.9000	140	140	0.7500	0.6500	0.7961	-0.1000	0.0461	0.0500	0.0505	0.1000

With 140 subjects in each group, the new treatment must have a success rate 0.0461 higher than the current treatment (or about 0.7961) to have 90% power in the test of non-inferiority.

Chapter 215

Equivalence Tests for Two Proportions

Introduction

This module provides power analysis and sample size calculation for equivalence tests in two-sample designs in which the outcome is binary. Users may choose from among eight popular test statistics commonly used for running the hypothesis test.

The power calculations assume that independent, random samples are drawn from two populations.

Four Procedures Documented Here

There are four procedures in the menus that use the program module described in this chapter. These procedures are identical except for the type of parameterization. The parameterization can be in terms of proportions, differences in proportions, ratios of proportions, and odds ratios. Each of these options is listed separately on the menus.

Example

An equivalence test example will set the stage for the discussion of the terminology that follows. Suppose that the response rate of the standard treatment of a disease is 0.70. Unfortunately, this treatment is expensive and occasionally exhibits serious side-effects. A promising new treatment has been developed to the point where it can be tested. One of the first questions that must be answered is whether the new treatment is therapeutically equivalent to the standard treatment.

Because of the many benefits of the new treatment, clinicians are willing to adopt the new treatment even if its effectiveness is slightly different from the standard. After thoughtful discussion with several clinicians, it is decided that if the response rate of the new treatment is between 0.63 and 0.77, the new treatment would be adopted. The *margin of equivalence* is 0.07.

The developers must design an experiment to test the hypothesis that the response rate of the new treatment does not differ from that of the standard treatment by more than 0.07. The statistical hypothesis to be tested is

$$H_0: |p_1 - p_2| \geq 0.07 \text{ versus } H_1: |p_1 - p_2| < 0.07$$

Technical Details

The details of sample size calculation for the two-sample design for binary outcomes are presented in the chapter entitled "Two Proportion Non-Null Case," and they will not be duplicated here. Instead, this chapter only discusses those changes necessary for equivalence tests.

Approximate sample size formulas for equivalence tests of two proportions are presented in Chow et al. (2003), page 88. Only large sample (normal approximation) results are given there. The results available in this module use exact calculations based on the enumeration of all possible values in the binomial distribution.

Suppose you have two populations from which dichotomous (binary) responses will be recorded. Assume without loss of generality that higher proportions are better. The probability (or risk) of cure in group 1 (the treatment group) is p_1 and in group 2 (the reference group) is p_2. Random samples of n_1 and n_2 individuals are obtained from these two groups. The data from these samples can be displayed in a 2-by-2 contingency table as follows

Group	Success	Failure	Total
Treatment	a	c	m
Control	b	d	n
Totals	s	f	N

The following alternative notation is also used.

Group	Success	Failure	Total
Treatment	x_{11}	x_{12}	n_1
Control	x_{21}	x_{22}	n_2
Totals	m_1	m_2	N

The binomial proportions p_1 and p_2 are estimated from these data using the formulae

$$\hat{p}_1 = \frac{a}{m} = \frac{x_{11}}{n_1} \text{ and } \hat{p}_2 = \frac{b}{n} = \frac{x_{21}}{n_2}$$

Let $p_{1.0}$ represent the group 1 proportion tested by the null hypothesis H_0. The power of a test is computed at a specific value of the proportion which we will call $p_{1.1}$. Let δ represent the smallest difference (margin of equivalence) between the two proportions that still results in the conclusion that the new treatment is equivalent to the current treatment. The set of statistical hypotheses that are tested is

$$\mathrm{H}_0: |p_{1.0} - p_2| \geq \delta \text{ versus } \mathrm{H}_1: |p_{1.0} - p_2| < \delta$$

These hypotheses can be rearranged to give

$$\mathrm{H}_0: p_{1.0} - p_2 \leq \delta_L \text{ or } p_{1.0} - p_2 \geq \delta_U \text{ versus } \mathrm{H}_1: \delta_L \leq p_{1.0} - p_2 \leq \delta_U$$

This composite hypothesis can be reduced to two one-sided hypotheses as follows

$$\mathrm{H}_{0L}: p_{1.0} - p_2 \leq \delta_L \text{ versus } \mathrm{H}_{1L}: \delta_L \leq p_{1.0} - p_2$$

$$\mathrm{H}_{0U}: p_{1.0} - p_2 \geq \delta_U \text{ versus } \mathrm{H}_{1U}: \delta_U \geq p_{1.0} - p_2$$

There are three common methods of specifying the margin of equivalence. The most direct is to simply give values for p_2 and $p_{1.0}$. However, it is often more meaningful to give p_2 and then specify $p_{1.0}$ implicitly by reporting the difference, ratio, or odds ratio. Mathematically, the definitions of these parameterizations are

Parameter	Computation	Alternative Hypotheses
Difference	$\delta = p_{1.0} - p_2$	$H_1: \delta_L \le p_{1.0} - p_2 \le \delta_U$
Ratio	$\phi = p_{1.0} / p_2$	$H_1: \phi_L \le p_{1.0} / p_2 \le \phi_U$
Odds Ratio	$\psi = Odds_{1.0} / Odds_2$	$H_1: \psi_L \le o_{1.0} / o_2 \le \psi_U$

Difference

The difference is perhaps the most direct method of comparison between two proportions. It is easy to interpret and communicate. It gives the absolute impact of the treatment. However, there are subtle difficulties that can arise with its interpretation.

One difficulty arises when the event of interest is rare. If a difference of 0.001 occurs when the baseline probability is 0.40, it would be dismissed as being trivial. However, if the baseline probability of a disease is 0.002, a 0.001 decrease would represent a reduction of 50%. Thus interpretation of the difference depends on the baseline probability of the event.

Note that $\delta_L < 0$ and $\delta_U > 0$. Usually, $\delta_L = -\delta_U$.

Equivalence using a Difference

The following example might help you understand the concept of an *equivalence* test. Suppose 60% of patients respond to the current treatment method $(p_2 = 0.60)$. If the response rate of the new treatment is no less than five percentage points better or worse than the existing treatment, it will be considered to be equivalent. Substituting these figures into the statistical hypotheses gives

$$H_0: p_{1.0} - p_2 \le -0.05 \text{ or } p_{1.0} - p_2 \ge 0.05 \text{ versus } H_1: -0.05 \le p_{1.0} - p_2 \le 0.05$$

Using the relationship

$$p_{1.0} = p_2 + \delta$$

gives

$$H_0: p_{1.0} \le 0.55 \text{ or } p_{1.0} \ge 0.65 \text{ versus } H_1: 0.55 \le p_{1.0} \le 0.65$$

In this example, when the null hypothesis is rejected, the concluded alternative is that the response rate is between 0.55 and 0.65.

Ratio

The ratio, $\phi = p_{1.0} / p_2$, gives the relative change in the probability of the response. Testing equivalence uses the formulation

$$H_0: p_{1.0} / p_2 \le \phi_L \text{ or } p_{1.0} / p_2 \ge \phi_U \text{ versus } H_1: \phi_L \le p_{1.0} / p_2 \le \phi_U$$

The only subtlety is that for equivalence tests $\phi_L < 1$ and $\phi_U > 1$. Usually, $\phi_L = 1 / \phi_U$.

Equivalence using a Ratio

The following example might help you understand the concept of *equivalence* as defined by the ratio. Suppose that 60% of patients $(p_2 = 0.60)$ respond to the current treatment method. If the response rate of a new treatment is within 10% of 0.60, it will be considered to be equivalent to the standard treatment. Substituting these figures into the statistical hypotheses gives

$$\mathrm{H}_0: p_{1.0} / p_2 \le 0.9 \text{ or } p_{1.0} / p_2 \ge 1.1 \text{ versus } \mathrm{H}_1: 0.9 \le p_{1.0} / p_2 \le 1.1$$

Using the relationship

$$p_{1.0} = \phi_0 p_2$$

gives

$$\mathrm{H}_0: p_{1.0} \le 0.54 \text{ or } p_{1.0} \ge 0.66 \text{ versus } \mathrm{H}_1: 0.54 \le p_{1.0} \le 0.66$$

Odds Ratio

The odds ratio, $\psi = \left(p_{1.0} / \left(1 - p_{1.0}\right)\right) / \left(p_2 / \left(1 - p_2\right)\right)$, gives the relative change in the odds (o) of the response. Testing equivalence use the formulation

$$\mathrm{H}_0: o_{1.0} / o_2 \le \psi_L \text{ or } o_{1.0} / o_2 \ge \psi_U \text{ versus } \mathrm{H}_1: \psi_L \le o_{1.0} / o_2 \le \psi_U$$

The only subtlety is that for equivalence tests $\psi_L < 1$ and $\psi_U > 1$. Usually, $\psi_L = 1 / \psi_U$.

Power Calculation

The power for a test statistic that is based on the normal approximation can be computed exactly using two binomial distributions. The following steps are taken to compute the power of these tests.

1. Find the critical values using the standard normal distribution. The critical values z_L and z_U are chosen as that value of z that leaves exactly the target value of alpha in the appropriate tail of the normal distribution.
2. Compute the value of the test statistic z_t for every combination of x_{11} and x_{21}. Note that x_{11} ranges from 0 to n_1, and x_{21} ranges from 0 to n_2. A small value (around 0.0001) can be added to the zero-cell counts to avoid numerical problems that occur when the cell value is zero.
3. If $z_t > z_L$ and $z_t < z_U$, the combination is in the rejection region. Call all combinations of x_{11} and x_{21} that lead to a rejection the set A.
4. Compute the power for given values of $p_{1.1}$ and p_2 as

$$1 - \beta = \sum_A \binom{n_1}{x_{11}} p_{1.1}^{x_{11}} q_{1.1}^{n_1 - x_{11}} \binom{n_2}{x_{21}} p_2^{x_{21}} q_2^{n_2 - x_{21}}$$

5. Compute the actual value of alpha achieved by the design by substituting $p_{1.0L}$ and $p_{1.0U}$ for $p_{1.1}$ to obtain

$$\alpha_L = \sum_A \binom{n_1}{x_{11}} p_{1.0L}^{x_{11}} q_{1.0L}^{n_1 - x_{11}} \binom{n_2}{x_{21}} p_2^{x_{21}} q_2^{n_2 - x_{21}}$$

and

$$\alpha_U = \sum_A \binom{n_1}{x_{11}} p_{1.0U}^{x_{11}} q_{1.0U}^{n_1 - x_{11}} \binom{n_2}{x_{21}} p_2^{x_{21}} q_2^{n_2 - x_{21}}$$

The value of alpha is then computed as the maximum of α_L and α_U .

Asymptotic Approximations

When the values of n_1 and n_2 are large (say over 200), these formulas take a long time to evaluate. In this case, a large sample approximation can be used. The large sample approximation is made by replacing the values of $\hat{p}_1$ and $\hat{p}_2$ in the *z* statistic with the corresponding values of $p_{1.1}$ and p_2 and then computing the results based on the normal distribution. Note that in large samples, the Farrington and Manning statistic is substituted for the Gart and Nam statistic. Also, for large samples, the results for the odds ratio have not (to our knowledge) been published. In this case, we substitute the calculations which are based on the ratio hypotheses.

Test Statistics

Several test statistics have been proposed for testing whether the difference, ratio, or odds ratio are different from a specified value. The main difference among the several test statistics is in the formula used to compute the standard error used in the denominator. These tests are based on the following *z*-test

$$z_t = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0 - c}{\hat{\sigma}}$$

The constant, *c*, represents a continuity correction that is applied in some cases. When the continuity correction is not used, *c* is zero. In power calculations, the values of $\hat{p}_1$ and $\hat{p}_2$ are not known. The corresponding values of $p_{1.1}$ and p_2 can be reasonable substitutes.

Following is a list of the test statistics available in ***PASS***. The availability of several test statistics begs the question of which test statistic one should use. The answer is simple: one should use the test statistic that will be used to analyze the data. You may choose a method because it is a standard in your industry, because it seems to have better statistical properties, or because your statistical package calculates it. Whatever your reasons for selecting a certain test statistic, you should use the same test statistic when doing the analysis after the data have been collected.

Z Test (Pooled)

This test was first proposed by Karl Pearson in 1900. Although this test is usually expressed directly as a chi-square statistic, it is expressed here as a *z* statistic so that it can be more easily used for one-sided hypothesis testing. The proportions are pooled (averaged) in computing the standard error. The formula for the test statistic is

$$z_t = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0}{\hat{\sigma}_1}$$

where

$$\hat{\sigma}_1 = \sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

$$\bar{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1+n_2}$$

Z Test (Unpooled)

This test statistic does not pool the two proportions in computing the standard error.

$$z_t = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0}{\hat{\sigma}_2}$$

where

$$\hat{\sigma}_2 = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

Z Test with Continuity Correction (Pooled)

This test is the same as Z Test (Pooled), except that a continuity correction is used. Remember that in the null case, the continuity correction makes the results closer to those of Fisher's Exact test.

$$z_t = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0 + \frac{F}{2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}{\hat{\sigma}_1}$$

$$\hat{\sigma}_1 = \sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

$$\bar{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1+n_2}$$

where F is -1 for lower-tailed hypotheses and 1 for upper-tailed hypotheses.

Z Test with Continuity Correction (Unpooled)

This test is the same as the Z Test (Unpooled), except that a continuity correction is used. Remember that in the null case, the continuity correction makes the results closer to those of Fisher's Exact test.

$$z_t = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0 - \frac{F}{2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}{\hat{\sigma}_2}$$

$$\hat{\sigma}_2 = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

where F is -1 for lower-tailed hypotheses and 1 for upper-tailed hypotheses.

T-Test of Difference

Because of a detailed, comparative study of the behavior of several tests, D'Agostino (1988) and Upton (1982) proposed using the usual two-sample t-test for testing whether the two proportions are equal. One substitutes a '1' for a success and a '0' for a failure in the usual, two-sample *t*-test formula.

Miettinen and Nurminen's Likelihood Score Test of the Difference

Miettinen and Nurminen (1985) proposed a test statistic for testing whether the difference is equal to a specified, non-zero, value, δ_0. The regular MLE's, $\hat{p}_1$ and $\hat{p}_2$, are used in the numerator of the score statistic while MLE's $\tilde{p}_1$ and $\tilde{p}_2$, constrained so that $\tilde{p}_1 - \tilde{p}_2 = \delta_0$, are used in the denominator. A correction factor of $N/(N\text{-}1)$ is applied to make the variance estimate less biased. The significance level of the test statistic is based on the asymptotic normality of the score statistic.

The formula for computing this test statistic is

$$z_{MND} = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0}{\hat{\sigma}_{MND}}$$

where

$$\hat{\sigma}_{MND} = \sqrt{\left(\frac{\tilde{p}_1\tilde{q}_1}{n_1} + \frac{\tilde{p}_2\tilde{q}_2}{n_2}\right)\left(\frac{N}{N-1}\right)}$$

$$\tilde{p}_1 = \tilde{p}_2 + \delta_0$$

$$\tilde{p}_1 = 2B\cos(A) - \frac{L_2}{3L_3}$$

$$A = \frac{1}{3}\left[\pi + \cos^{-1}\left(\frac{C}{B^3}\right)\right]$$

$$B = \text{sign}(C)\sqrt{\frac{L_2^2}{9L_3} - \frac{L_1}{3L_3}}$$

$$C = \frac{L_2^3}{27L_3^3} - \frac{L_1L_2}{6L_3^2} + \frac{L_0}{2L_3}$$

$$L_0 = x_{21}\delta_0(1-\delta_0)$$

$$L_1 = \left[N_2\delta_0 - N - 2x_{21}\right]\delta_0 + M_1$$

$$L_2 = \left(N + N_2\right)\delta_0 - N - M_1$$

$$L_3 = N$$

Miettinen and Nurminen's Likelihood Score Test of the Ratio

Miettinen and Nurminen (1985) proposed a test statistic for testing whether the ratio is equal to a specified value, ϕ_0. The regular MLE's, $\hat{p}_1$ and $\hat{p}_2$, are used in the numerator of the score statistic while MLE's $\tilde{p}_1$ and $\tilde{p}_2$, constrained so that $\tilde{p}_1 / \tilde{p}_2 = \phi_0$, are used in the denominator. A correction factor of N/(N-1) is applied to make the variance estimate less biased. The significance level of the test statistic is based on the asymptotic normality of the score statistic.

The formula for computing the test statistic is

$$z_{MNR} = \frac{\hat{p}_1 / \hat{p}_2 - \phi_0}{\sqrt{\left(\frac{\tilde{p}_1\tilde{q}_1}{n_1} + \phi_0^2\frac{\tilde{p}_2\tilde{q}_2}{n_2}\right)\left(\frac{N}{N-1}\right)}}$$

where

$$\tilde{p}_1 = \tilde{p}_2\phi_0$$

$$\tilde{p}_2 = \frac{-B - \sqrt{B^2 - 4AC}}{2A}$$

$$A = N\phi_0$$

$$B = -\left[N_1\phi_0 + x_{11} + N_2 + x_{21}\phi_0\right]$$

$$C = M_1$$

Miettinen and Nurminen's Likelihood Score Test of the Odds Ratio

Miettinen and Nurminen (1985) proposed a test statistic for testing whether the odds ratio is equal to a specified value, ψ_0. Because the approach they used with the difference and ratio does not easily extend to the odds ratio, they used a score statistic approach for the odds ratio. The regular MLE's are $\hat{p}_1$ and $\hat{p}_2$. The constrained MLE's are $\tilde{p}_1$ and $\tilde{p}_2$. These estimates are constrained so that $\tilde{\psi} = \psi_0$. A correction factor of N/(N-1) is applied to make the variance estimate less biased. The significance level of the test statistic is based on the asymptotic normality of the score statistic. The formula for computing the test statistic is

$$z_{MNO} = \frac{\frac{\left(\hat{p}_1 - \tilde{p}_1\right)}{\tilde{p}_1\tilde{q}_1} - \frac{\left(\hat{p}_2 - \tilde{p}_2\right)}{\tilde{p}_2\tilde{q}_2}}{\sqrt{\left(\frac{1}{N_2\tilde{p}_1\tilde{q}_1} + \frac{1}{N_2\tilde{p}_2\tilde{q}_2}\right)\left(\frac{N}{N-1}\right)}}$$

where

$$\tilde{p}_1 = \frac{\tilde{p}_2\psi_0}{1+\tilde{p}_2(\psi_0-1)}$$

$$\tilde{p}_2 = \frac{-B+\sqrt{B^2-4AC}}{2A}$$

$$A = N_2(\psi_0-1)$$

$$B = N_1\psi_0 + N_2 - M_1(\psi_0-1)$$

$$C = -M_1$$

Farrington and Manning's Likelihood Score Test of the Difference

Farrington and Manning (1990) proposed a test statistic for testing whether the difference is equal to a specified value, δ_0. The regular MLE's, $\hat{p}_1$ and $\hat{p}_2$, are used in the numerator of the score statistic while MLE's $\tilde{p}_1$ and $\tilde{p}_2$, constrained so that $\tilde{p}_1 - \tilde{p}_2 = \delta_0$, are used in the denominator. The significance level of the test statistic is based on the asymptotic normality of the score statistic.

The formula for computing the test statistic is

$$z_{FMD} = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0}{\sqrt{\left(\frac{\tilde{p}_1\tilde{q}_1}{n_1} + \frac{\tilde{p}_2\tilde{q}_2}{n_2}\right)}}$$

where the estimates $\tilde{p}_1$ and $\tilde{p}_2$ are computed as in the corresponding test of Miettinen and Nurminen (1985) given above.

Farrington and Manning's Likelihood Score Test of the Ratio

Farrington and Manning (1990) proposed a test statistic for testing whether the ratio is equal to a specified value, ϕ_0. The regular MLE's, $\hat{p}_1$ and $\hat{p}_2$, are used in the numerator of the score statistic while MLE's $\tilde{p}_1$ and $\tilde{p}_2$, constrained so that $\tilde{p}_1 / \tilde{p}_2 = \phi_0$, are used in the denominator. A correction factor of N/(N-1) is applied to increase the variance estimate. The significance level of the test statistic is based on the asymptotic normality of the score statistic.

The formula for computing the test statistic is

$$z_{FMR} = \frac{\hat{p}_1 / \hat{p}_2 - \phi_0}{\sqrt{\left(\frac{\tilde{p}_1\tilde{q}_1}{n_1} + \phi_0^2\frac{\tilde{p}_2\tilde{q}_2}{n_2}\right)}}$$

where the estimates $\tilde{p}_1$ and $\tilde{p}_2$ are computed as in the corresponding test of Miettinen and Nurminen (1985) given above.

Farrington and Manning's Likelihood Score Test of the Odds Ratio

Farrington and Manning (1990) indicate that the Miettinen and Nurminen statistic may be modified by removing the factor $N/(N\text{-}1)$.

The formula for computing this test statistic is

$$z_{FMO} = \frac{\frac{\left(\hat{p}_1 - \tilde{p}_1\right)}{\tilde{p}_1\tilde{q}_1} - \frac{\left(\hat{p}_2 - \tilde{p}_2\right)}{\tilde{p}_2\tilde{q}_2}}{\sqrt{\left(\frac{1}{N_2\tilde{p}_1\tilde{q}_1} + \frac{1}{N_2\tilde{p}_2\tilde{q}_2}\right)}}$$

where the estimates $\tilde{p}_1$ and $\tilde{p}_2$ are computed as in the corresponding test of Miettinen and Nurminen (1985) given above.

Gart and Nam's Likelihood Score Test of the Difference

Gart and Nam (1990), page 638, proposed a modification to the Farrington and Manning (1988) difference test that corrects for skewness. Let $z_{FMD}(\delta)$ stand for the Farrington and Manning difference test statistic described above. The skewness-corrected test statistic, z_{GND}, is the appropriate solution to the quadratic equation

$$\left(-\tilde{\gamma}\right)z_{GND}^2 + \left(-1\right)z_{GND} + \left(z_{FMD}(\delta) + \tilde{\gamma}\right) = 0$$

where

$$\tilde{\gamma} = \frac{\tilde{V}^{3/2}(\delta)}{6}\left(\frac{\tilde{p}_1\tilde{q}_1(\tilde{q}_1 - \tilde{p}_1)}{n_1^2} - \frac{\tilde{p}_2\tilde{q}_2(\tilde{q}_2 - \tilde{p}_2)}{n_2^2}\right)$$

Gart and Nam's Likelihood Score Test of the Ratio

Gart and Nam (1988), page 329, proposed a modification to the Farrington and Manning (1988) ratio test that corrects for skewness. Let $z_{FMR}(\phi)$ stand for the Farrington and Manning ratio test statistic described above. The skewness-corrected test statistic, z_{GNR}, is the appropriate solution to the quadratic equation

$$\left(-\tilde{\varphi}\right)z_{GNR}^2 + \left(-1\right)z_{GNR} + \left(z_{FMR}(\phi) + \tilde{\varphi}\right) = 0$$

where

$$\tilde{\varphi} = \frac{1}{6\tilde{u}^{3/2}}\left(\frac{\tilde{q}_1(\tilde{q}_1 - \tilde{p}_1)}{n_1^2\tilde{p}_1^2} - \frac{\tilde{q}_2(\tilde{q}_2 - \tilde{p}_2)}{n_2^2\tilde{p}_2^2}\right)$$

$$\tilde{u} = \frac{\tilde{q}_1}{n_1\tilde{p}_1} + \frac{\tilde{q}_2}{n_2\tilde{p}_2}$$

Procedure Options

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

Data Tab (Common Options)

The Data tab contains the parameters associated with this test such as the proportions, sample sizes, alpha, and power. This chapter covers four procedures, each of which has different options. This section documents options that are common to all four procedures. Later, unique options for each procedure will be documented.

Solve For

Find (Solve For)

This option specifies the parameter to be solved for using the other parameters. The parameters that may be selected are *P1.1*, *Alpha*, *Power and Beta*, *N1*, and *N2*. Under most situations, you will select either *Power and Beta* or *N1*.

Select N1 when you want to calculate the sample size needed to achieve a given power and alpha level.

Select *Power and Beta* when you want to calculate the power of an experiment.

Error Rates

Power or Beta

This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected. In this procedure, a type-II error occurs when you fail to reject the null hypothesis of unequal proportions when in fact they are equivalent.

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

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

Alpha (Significance Level)

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected. In this procedure, a type-I error occurs when you reject the null hypothesis of unequal proportions when in fact they are not equal.

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

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

Sample Size

N1 (Sample Size Group 1)

Enter a value (or range of values) for the sample size of this group. 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.

Effect Size – Reference (Group 2)

P2 (Reference Group Proportion)

Specify the value of P2, the reference, baseline, or control group's proportion. The null hypothesis is that the two proportions differ by no more than a specified amount. Since P2 is a proportion, these values must be between 0 and 1.

You may enter a range of values such as *0.1 0.2 0.3* or *0.1 to 0.9 by 0.1*.

Test

Test Type

Specify which test statistic is used in searching and reporting. Although the pooled *z*-test is commonly shown in elementary statistics books, the likelihood score test is arguably the best choice.

Note that *C.C.* is an abbreviation for *Continuity Correction*. This refers to the adding or subtracting 1/(2n) to (or from) the numerator of the z-value to bring the normal approximation closer to the binomial distribution.

Data Tab (Proportion)

This section documents options that are used when the parameterization is in terms of the values of the two proportions, P1 and P2. P1.0 is the value of the P1 assumed by the null hypothesis and P1.1 is the value of P1 at which the power is calculated.

Effect Size – Equivalence Proportions

P1.0U & P1.0L (Upper & Lower Equivalence Proportion)

Specify the *margin of equivalence* directly by giving the upper and lower bounds of P1.0. The two groups are assumed to be equivalent when P1.0 is between these values. Thus, P1.0U should be greater than P2 and P1.0L should be less than P2.

Note that the values of P1.0U and P1.0L are used in pairs. Thus, the first values of P1.0U and P1.0L are used together, then the second values of each are used, and so on.

You may enter a range of values such as *0.03 0.05 0.10* or *0.01 to 0.05 by 0.01*.

Proportions must be between 0 and 1. They cannot take on the values 0 or 1. These values should surround P2.

Effect Size – Actual Proportion

P1.1 (Actual Proportion)

This option specifies the value of P1.1, which is the value of the treatment proportion at which the power is to be calculated. Proportions must be between 0 and 1. They cannot take on the values 0 or 1.

You may enter a range of values such as *0.03 0.05 0.10* or *0.01 to 0.05 by 0.01*.

Data Tab (Difference)

This section documents options that are used when the parameterization is in terms of the difference, P1 – P2. P1.0 is the value of P1 assumed by the null hypothesis and P1.1 is the value of P1 at which the power is calculated. Once P2, D0, and D1 are given, the values of P1.1 and P1.0 can be calculated.

Effect Size – Equivalence Differences

D0.U & D0.L (Upper & Lower Equivalence Difference)

Specify the *margin of equivalence* by specifying the largest distance above (D0.U) and below (D0.L) P2 which will still result in the conclusion of equivalence. As long as the actual difference is between these two values, the difference is not considered to be large enough to be of practical importance.

The values of D0.U must be positive and the values of D0.L must be negative. D0.L can be set to '-D0.U,' which is usually what is desired.

The power calculations assume that P1.0 is the value of P1 under the null hypothesis. This value is used with P2 to calculate the value of P1.0 using the formula: P1.0U = D0.U + P2.

You may enter a range of values for D0.U such as *.03 .05 .10* or *.05 to .20 by .05*.

Note that if you enter values for D0.L (other than '-D0.U'), they are used in pairs with the values of D0.U. Thus, the first values of D0.U and D0.L are used together, then the second values of each are used, and so on.

RANGE:

D0.L must be between -1 and 0. D0.U must be between 0 and 1. Neither can take on the values -1, 0, or 1.

Effect Size – Actual Difference

D1 (Actual Difference)

This option specifies the actual difference between P1.1 (the actual value of P1) and P2. This is the value of the difference at which the power is calculated. In equivalence trials, this difference is often set to 0.

The power calculations assume that P1.1 is the actual value of the proportion in group 1 (experimental or treatment group). This difference is used with P2 to calculate the true value of P1 using the formula: P1.1 = D1 + P2.

You may enter a range of values such as *-.05 0 .5* or *-.05 to .05 by .02*. Actual differences must be between -1 and 1. They cannot take on the values -1 or 1.

Data Tab (Ratio)

This section documents options that are used when the parameterization is in terms of the ratio, P1 / P2. P1.0 is the value of P1 assumed by the null hypothesis and P1.1 is the value of P1 at which the power is calculated. Once P2, R0, and R1 are given, the values of P1.0 and P1.1 can be calculated.

Effect Size – Equivalence Ratios

R0.U & R0.L (Upper & Lower Equivalence Ratio)

Specify the *margin of equivalence* by specifying the largest ratio (P1/P2) above (R0.U) and below (R0.L) 1 which will still result in the conclusion of equivalence. As long as the actual ratio is between these two values, the difference between the proportions is not considered to be large enough to be of practical importance.

The values of R0.U must be greater than 1 and the values of R0.L must be less than 1. R0.L can be set to ‘1/R0.U,’ which is often desired.

The power calculations assume that P1.0 is the value of P1 under the null hypothesis. This value is used with P2 to calculate the value of P1.0 using the formula: P1.0U = R0.U x P2.

You may enter a range of values for R0.U such as *1.1 1.5 1.8* or *1.1 to 2.1 by 0.2*.

Note that if you enter values for R0.L (other than '1/R0.U'), they are used in pairs with the values of R0.U. Thus, the first values of R0.U and R0.L are used together, then the second values of each are used, and so on.

R0.L must be between 0 and 1. R0.U must be greater than 1. Neither can take on the value 1.

Effect Size – Actual Ratio

R1 (Actual Ratio)

This option specifies the ratio of P1.1 and P2, where P1.1 is the actual proportion in the treatment group. The power calculations assume that P1.1 is the actual value of the proportion in group 1. This difference is used with P2 to calculate the value of P1 using the formula: P1.1 = R1 x P2. In equivalence trials, this ratio is often set to 1.

Ratios must be positive. You may enter a range of values such as *0.95 1 1.05* or *0.9 to 1.9 by 0.02*.

Data Tab (Odds Ratio)

This section documents options that are used when the parameterization is in terms of the odds ratios, O1.1 / O2 and O1.0 / O2. Note that the odds are defined as O2 = P2 / (1 – P2), O1.0 = P1.0 / (1 – P1.0), etc. P1.0 is the value of P1 assumed by the null hypothesis and P1.1 is the value of P1 at which the power is calculated. Once P2, OR0, and OR1 are given, the values of P1.1 and P1.0 can be calculated.

Effect Size – Equivalence Odds Ratios

OR0.U & OR0.L (Upper & Lower Equivalence Odds Ratio)

Specify the *margin of equivalence* by specifying the largest odds ratio above (OR0.U) and below (OR0.L) 1 which will still result in the conclusion of equivalence. As long as the actual odds ratio is between these two values, the difference between the proportions is not large enough to be of practical importance.

The values of OR0.U must be greater than 1 and the values of OR0.L must be less than 1. OR0.L can be set to '1/OR0.U,' which is often desired.

The power calculations assume that P1.0 is the value of the P1 under the null hypothesis. This value is used with P2 to calculate the value of P1.0.

You may enter a range of values for OR0.U such as *1.1 1.5 1.8* or *1.1 to 2.1 by 0.2*.

Note that if you enter values for OR0.L (other than '1/OR0.U'), they are used in pairs with the values of OR0.U. Thus, the first values of OR0.U and OR0.L are used together, next the second values of each are used, and so on.

OR0.L must be between 0 and 1. OR0.U must be greater than 1. Neither can take on the value 1.

Effect Size – Actual Odds Ratio

OR1 (Actual Odds Ratio)

This option specifies the odds ratio of P1.1 and P2, where P1.1 is the actual proportion in the treatment group. The power calculations assume that P1.1 is the actual value of the proportion in group 1. This value is used with P2 to calculate the value of P1. In equivalence trials, this odds ratio is often set to 1.

Odds ratios must be positive. You may enter a range of values such as *0.95 1 1.05* or *0.9 to 1.9 by 0.02*.

Options Tab

The Options tab contains various limits and options.

Maximum Iterations

Maximum Iterations Before Search Termination

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 not reported. A value of at least 500 is recommended.

Zero Counts

Zero Count Adjustment Method

Zero cell counts cause many calculation problems. To compensate for this, a small value (called the Zero Count Adjustment Value) can be added either to all cells or to all cells with zero counts. This option specifies whether you want to use the adjustment and which type of adjustment you want to use. We recommend that you use the option 'Add to zero cells only.'

Zero cell values often do not occur in practice. However, since power calculations are based on total enumeration, they will occur in power and sample size estimation.

Adding a small value is controversial, but can be necessary for computational considerations. Statisticians have recommended adding various fractions to zero counts. We have found that adding 0.0001 seems to work well.

Zero Count Adjustment Value

Zero cell counts cause many calculation problems when computing power or sample size. To compensate for this, a small value may be added either to all cells or to all zero cells. This value indicates the amount that is added. We have found that 0.0001 works well.

Be warned that the value of the ratio and the odds ratio will be affected by the amount specified here!

Exact Test Options

Maximum N1 or N2 for Exact Calculations

When either N1 or N2 is above this amount, power calculations are based on the normal approximation to the binomial. When the normal approximation to the binomial is used, the actual value of alpha is not calculated. Currently, for three-gigahertz computers, a value near 200 is reasonable. As computers increase in speed, this number may be increased.

Example 1 – Finding Power

A study is being designed to establish the equivalence of a new treatment compared to the current treatment. Historically, the current treatment has enjoyed a 50% cure rate. The new treatment reduces the seriousness of certain side effects that occur with the current treatment. Thus, the new treatment will be adopted even if it is slightly less effective than the current treatment. The researchers will recommend adoption of the new treatment if its cure rate is within 15% of the standard treatment.

The researchers plan to use the Farrington and Manning likelihood score test statistic to analyze the data. They want to study the power of the Farrington and Manning test at group sample sizes ranging from 50 to 500 for detecting a difference inside 15% when the actual cure rate of the new treatment ranges from 50% to 60%. The significance level will be 0.05.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Equivalence Tests for Two Proportions [Differences]** procedure window by expanding **Proportions**, then **Two Independent Proportions**, then clicking on **Equivalence**, and then clicking on **Equivalence Tests for Two Proportions [Differences]**. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power	*Ignored since this is the Find setting*
Alpha	**0.05**
N1 (Sample Size Group 1)	**50 to 500 by 50**
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
D0.U (Upper Equivalence Difference)	**0.15**
D0.L (Lower Equivalence Difference)	**-D0.U**
D1 (Actual Difference)	**0.00 0.05 0.10**
P2 (Reference Group Proportion)	**0.5**
Test Type	**Likelihood Score (Farr. & Mann.)**
Options Tab	
Maximum N1 or N2 Exact	**100**

Annotated Output

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

Numeric Results

Numeric Results for Equivalence Tests Based on the Difference: P1 - P2
H0: P1-P2<=D0.L or P1-P2>=D0.U. H1: D0.L<P1-P2=D1<D0.U.
Test Statistic: Score test (Farrington & Manning)

Power	Sample Size Grp 1 N1	Sample Size Grp 2 N2	Prop Grp 2 P2	Lower Equiv. Grp 1 Prop P1.0L	Upper Equiv. Grp 1 Prop P1.0U	Lower Equiv. Margin Diff D0.L	Upper Equiv. Margin Diff D0.U	Actual Margin Diff D1	Target Alpha	Actual Alpha
0.0000	50	50	0.5000	0.3500	0.6500	-0.1500	0.1500	0.0000	0.0500	0.0515
0.3793	100	100	0.5000	0.3500	0.6500	-0.1500	0.1500	0.0000	0.0500	0.0489
0.6689	150	150	0.5000	0.3500	0.6500	-0.1500	0.1500	0.0000	0.0500	
0.8305	200	200	0.5000	0.3500	0.6500	-0.1500	0.1500	0.0000	0.0500	
0.9160	250	250	0.5000	0.3500	0.6500	-0.1500	0.1500	0.0000	0.0500	
0.9594	300	300	0.5000	0.3500	0.6500	-0.1500	0.1500	0.0000	0.0500	
0.9808	350	350	0.5000	0.3500	0.6500	-0.1500	0.1500	0.0000	0.0500	
0.9911	400	400	0.5000	0.3500	0.6500	-0.1500	0.1500	0.0000	0.0500	

Report continues …

Note: exact results based on the binomial were only calculated when both N1 and N2 were less than 100.

Report Definitions
'Power' is the probability of rejecting a false null hypothesis. It should be close to one.
'Power' is the probability of concluding equivalence when equivalence is correct.
'Beta' is the probability of accepting a false H0. Beta = 1 - Power.
'N1 and N2' are the sizes of the samples drawn from the corresponding groups.
'P2' is the response rate for group two which is the standard, reference, baseline, or control group.
'P1.0L' is the smallest treatment-group response rate that still yields an equivalence conclusion.
'P1.0U' is the largest treatment-group response rate that still yields an equivalence conclusion.
'D0.L' is the lowest difference that still results in the conclusion of equivalence.
'D0.U' is the highest difference that still results in the conclusion of equivalence.
'D1' is the actual difference, P1-P2, at which the power is calculated.
'Target Alpha' is the probability of rejecting a true null hypothesis that was desired.
'Actual Alpha' is the value of alpha that is actually achieved. Only available for exact results.
'Grp 1' refers to Group 1 which is the treatment or experimental group.
'Grp 2' refers to Group 2 which is the reference, standard, or control group.
' Equiv.' refers to a small amount that is not of practical importance.
'Actual' refers to the true value at which the power is computed.

Summary Statements
Sample sizes of 50 in the treatment group and 50 in the reference group achieve 0% power to detect equivalence. The margin of equivalence, given in terms of the difference, extends from -0.1500 to 0.1500. The actual difference is 0.0000. The reference group proportion is 0.5000. The calculations assume that two, one-sided likelihood score (Farrington & Manning) tests are used. Although the significance level is targeted at 0.0500, the level actually achieved is 0.0515.

This report shows the values of each of the parameters, one scenario per row. Note that the actual alpha value is blank for sample sizes greater than 100, which was the limit set for exact computation.

Most of the report columns have obvious interpretations. Those that may not be obvious are presented here.

Prop Grp 2 P2

This is the value of P2, the response rate in the control group.

Lower & Upper Equiv. Grp 1 Prop: P1.0L & P1.0U

These are the margin of equivalence for the response rate of the treatment group as specified by the null hypothesis of non-equivalence. Values of P1 inside these limits are considered equivalent to P2.

Lower & Upper Equiv. Margin Diff: D0.L & D0.U

These set the margin of equivalence for the difference in response rates. Values of the difference outside these limits are considered *non-equivalent*.

Actual Margin Diff D1

This is the value of D1, the difference between the two group proportions at which the power is computed. This is the value of the difference under the alternative hypothesis.

Target Alpha

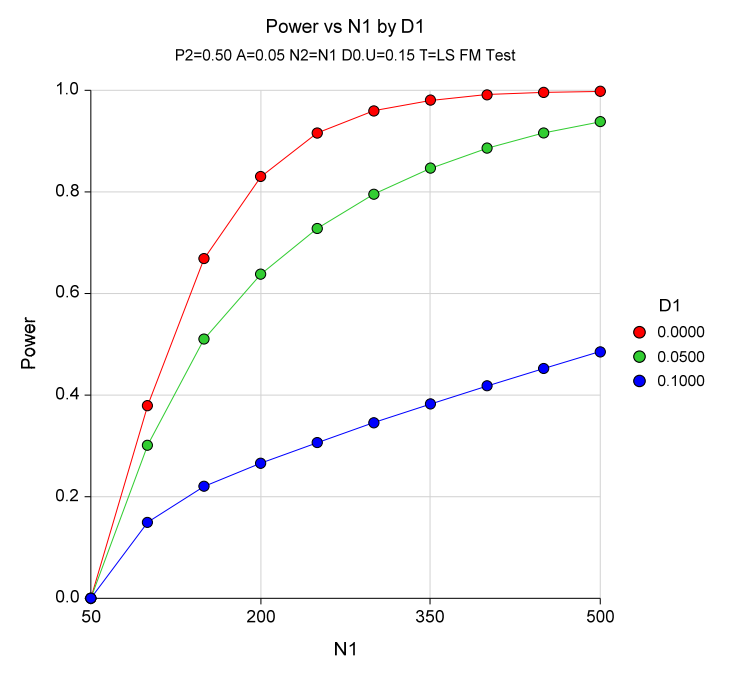
This is the value of alpha that was targeted by the design. Note that the target alpha is not usually achieved exactly. For two-sided tests, this value will usually be 0.05.

Actual Alpha

This is the value of alpha that was actually achieved by this design. Note that since the limit on exact calculations was set to 100, and since this value is calculated exactly, it is not shown for values of N1 greater than 100.

The difference between the Target Alpha and the Actual Alpha is caused by the discrete nature of the binomial distribution and the use of the normal approximation to the binomial in determining the critical value of the test statistic.

Plots Section



The values from the table are displayed in the above chart. This chart gives us a quick look at the sample size that will be required for various values of D1.

Example 2 – Finding the Sample Size

Continuing with the scenario given in Example 1, the researchers want to determine the sample size necessary for each value of D1 to achieve a power of 0.80. To cut down on the runtime, they decide to look at approximate values whenever N1 is greater than 100.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Equivalence Tests for Two Proportions [Differences]** procedure window by expanding **Proportions**, then **Two Independent Proportions**, then clicking on **Equivalence**, and then clicking on **Equivalence Tests for Two Proportions [Differences]**. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**N1**
Power	**0.80**
Alpha	**0.05**
N1 (Sample Size Group 1)	*Ignored since this is the Find setting.*
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
D0.U (Upper Equivalence Difference)	**0.15**
D0.L (Lower Equivalence Difference)	**-D0.U**
D1 (Actual Difference)	**0.00 0.05 0.10**
P2 (Reference Group Proportion)	**0.5**
Test Type	**Likelihood Score (Farr. & Mann.)**
Options Tab	
Maximum N1 or N2 Exact	**100**

Output

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

Numeric Results

Numeric Results for Equivalence Tests Based on the Difference: P1 - P2
H0: P1-P2<=D0.L or P1-P2>=D0.U. H1: D0.L<P1-P2=D1<D0.U.
Test Statistic: Score test (Farrington & Manning)

Power	Sample Size Grp 1 N1	Sample Size Grp 2 N2	Prop Grp 2 P2	Lower Equiv. Grp 1 Prop P1.0L	Upper Equiv. Grp 1 Prop P1.0U	Lower Equiv. Margin Diff D0.L	Upper Equiv. Margin Diff D0.U	Actual Margin Diff D1	Target Alpha	Actual Alpha
0.8003	188	188	0.5000	0.3500	0.6500	-0.1500	0.1500	0.0000	0.0500	
0.8001	304	304	0.5000	0.3500	0.6500	-0.1500	0.1500	0.0500	0.0500	
0.8001	1202	1202	0.5000	0.3500	0.6500	-0.1500	0.1500	0.1000	0.0500	

The required sample size will depend a great deal on the value of D1. Any effort spent determining an accurate value for D1 will be worthwhile.

Example 3 – Comparing the Power of Several Test Statistics

Continuing with Example 1, the researchers want to determine which of the eight possible test statistics to adopt by using the comparative reports and charts that ***PASS*** produces. They decide to compare the powers and actual alphas for various sample sizes between 50 and 200 when D1 is 0.1.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Equivalence Tests for Two Proportions [Differences]** procedure window by expanding **Proportions**, then **Two Independent Proportions**, then clicking on **Equivalence**, and then clicking on **Equivalence Tests for Two Proportions [Differences]**. You may then make the appropriate entries as listed below, or open **Example 3** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power	*Ignored since this is the Find setting*
Alpha	**0.05**
N1 (Sample Size Group 1)	**50 to 200 by 50**
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
D0.U (Upper Equivalence Difference)	**0.15**
D0.L (Lower Equivalence Difference)	**-D0.U**
D1 (Actual Difference)	**0.10**
P2 (Reference Group Proportion)	**0.5**
Test Type	**Likelihood Score (Farr. & Mann.)**
Reports Tab	
Show Numeric Report	**Not checked**
Show Comparative Reports	**Checked**
Show Definitions	**Not checked**
Show Plots	**Not checked**
Show Comparative Plots	**Checked**
Number of Summary Statements	**0**
Options Tab	
Maximum N1 or N2 Exact	**300**

Output

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

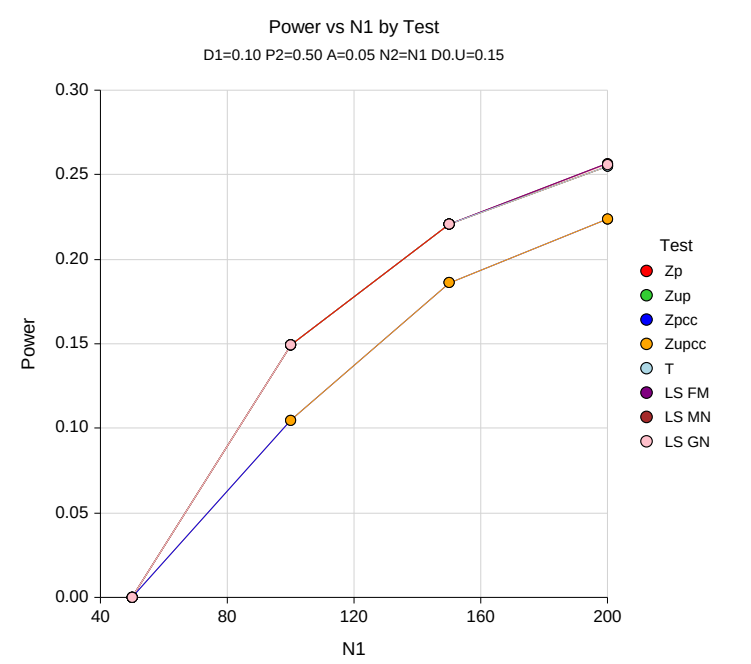
Numeric Results and Plots

Power Comparison of Equivalence Tests Based on the Difference: P1 - P2
H0: P1-P2<=D0.L or P1-P2>=D0.U. H1: D0.L<P1-P2=D1<D0.U.

N1/N2	P2	Upper Equiv. Margin D0.U	Target Alpha	Z(P) Test Power	Z(UnP) Test Power	Z(P) CC Test Power	Z(UnP) CC Test Power	T Test Power	F.M. Score Power	M.N. Score Power	G.N. Score Power
50/50	0.5000	0.1500	0.0500	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
100/100	0.5000	0.1500	0.0500	0.1494	0.1494	0.1047	0.1047	0.1493	0.1495	0.1494	0.1494
150/150	0.5000	0.1500	0.0500	0.2208	0.2208	0.1863	0.1863	0.2208	0.2208	0.2208	0.2208
200/200	0.5000	0.1500	0.0500	0.2552	0.2553	0.2238	0.2239	0.2552	0.2566	0.2566	0.2560

Actual Alpha Comparison of Equivalence Tests Based on the Difference: P1 - P2
H0: P1-P2<=D0.L or P1-P2>=D0.U. H1: D0.L<P1-P2=D1<D0.U.

N1/N2	P2	Upper Equiv. Margin D0.U	Target Alpha	Z(P) Test Alpha	Z(UnP) Test Alpha	Z(P) CC Test Alpha	Z(UnP) CC Test Alpha	T Test Alpha	F.M. Score Alpha	M.N. Score Alpha	G.N. Score Alpha
50/50	0.5000	0.1500	0.0500	0.0515	0.0515	0.0334	0.0334	0.0514	0.0515	0.0515	0.0515
100/100	0.5000	0.1500	0.0500	0.0486	0.0486	0.0358	0.0358	0.0485	0.0489	0.0487	0.0487
150/150	0.5000	0.1500	0.0500	0.0495	0.0495	0.0386	0.0386	0.0495	0.0495	0.0495	0.0495
200/200	0.5000	0.1500	0.0500	0.0465	0.0468	0.0376	0.0378	0.0465	0.0488	0.0488	0.0481



It is interesting to note that the powers of the continuity-corrected test statistics are consistently lower than the other tests. This occurs because the actual alpha achieved by these tests is lower than for the other tests. An interesting finding of this example is that the regular *t*-test performed about as well as the *z*-test.

Example 4 – Validation using Chow with Equal Sample Sizes

Chow et al. (2003), page 91, present a sample size study in which P2 = 0.75, D0.U = 0.2, D0.L = -0.2, D1 = 0.05, alpha = 0.05, and beta = 0.2. Using the pooled Z test statistic, they found the sample size to be 96 in each group.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Equivalence Tests for Two Proportions [Differences]** procedure window by expanding **Proportions**, then **Two Independent Proportions**, then clicking on **Equivalence**, and then clicking on **Equivalence Tests for Two Proportions [Differences]**. You may then make the appropriate entries as listed below, or open **Example 4** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**N1**
Power	**0.80**
Alpha	**0.05**
N1 (Sample Size Group 1)	*Ignored since this is the Find setting.*
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
D0.U (Upper Equivalence Difference)	**0.2**
D0.L (Lower Equivalence Difference)	**-D0.U**
D1 (Actual Difference)	**0.05**
P2 (Reference Group Proportion)	**0.75**
Test Type	**Z Test (Pooled)**
Options Tab	
Maximum N1 or N2 Exact	**2 (Set low for a rapid search.)**

Output

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

Numeric Results

Numeric Results for Equivalence Tests Based on the Difference: P1 - P2
H0: P1-P2<=D0.L or P1-P2>=D0.U. H1: D0.L<P1-P2=D1<D0.U.
Test Statistic: Z test (pooled)

Power	Sample Size Grp 1 N1	Sample Size Grp 2 N2	Prop Grp 2 P2	Lower Equiv. Grp 1 Prop P1.0L	Upper Equiv. Grp 1 Prop P1.0U	Lower Equiv. Margin Diff D0.L	Upper Equiv. Margin Diff D0.U	Actual Margin Diff D1	Target Alpha	Actual Alpha
0.8028	98	98	0.7500	0.5500	0.9500	-0.2000	0.2000	0.0500	0.0500	

PASS found the required sample size to be 98 which is slightly larger than the 96 that Chow obtained. This is mainly due to the rounding to two decimal places that Chow did in this example.

We used the exact option in ***PASS*** and obtained N1 = 99. Thus, ***PASS*** was indeed closer than was Chow.

Example 5 – Validation using Tuber-Bitter with Equal Sample Sizes

Tuber-Bitter et al. (2000), page 1271, present a sample size study in which P2 = 0.1; D0.U = 0.01, 0.02, 0.03; D0.L = -D0.U; D1 = 0.0; alpha = 0.05; and beta = 0.1. Using the pooled Z test statistic, they found the sample sizes to be 19484, 4871, and 2165 in each group.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Equivalence Tests for Two Proportions [Differences]** procedure window by expanding **Proportions**, then **Two Independent Proportions**, then clicking on **Equivalence**, and then clicking on **Equivalence Tests for Two Proportions [Differences]**. You may then make the appropriate entries as listed below, or open **Example 5** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**N1**
Power	**0.90**
Alpha	**0.05**
N1 (Sample Size Group 1)	*Ignored since this is the Find setting.*
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
Specify Treatment Proportion using	**Differences (P1-P2)**
D0.U (Upper Equivalence Difference)	**.01 .02 .03**
D0.L (Lower Equivalence Difference)	**-D0.U**
D1 (Actual Difference)	**0.0**
P2 (Reference Group Proportion)	**0.1**
Test Type	**Z Test (Pooled)**
Options Tab	
Maximum N1 or N2 Exact	**2 (Set low for a rapid search.)**

Output

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

Numeric Results

Numeric Results for Equivalence Tests Based on the Difference: P1 - P2
H0: P1-P2<=D0.L or P1-P2>=D0.U. H1: D0.L<P1-P2=D1<D0.U.
Test Statistic: Z test (pooled)

Power	Sample Size Grp 1 N1	Sample Size Grp 2 N2	Prop Grp 2 P2	Lower Trivial Grp 1 Prop P1.0L	Upper Trivial Grp 1 Prop P1.0U	Lower Trivial Margin Diff D0.L	Upper Trivial Margin Diff D0.U	Actual Margin Diff D1	Target Alpha	Actual Alpha
0.9000	19480	19480	0.1000	0.0900	0.1100	-0.0100	0.0100	0.0000	0.0500	
0.9000	4870	4870	0.1000	0.0800	0.1200	-0.0200	0.0200	0.0000	0.0500	
0.9001	2165	2165	0.1000	0.0700	0.1300	-0.0300	0.0300	0.0000	0.0500	

PASS found the required sample sizes to within rounding error of Tuber-Bitter.

Example 6 – Computing the Power after Completing an Experiment

Researchers are testing a generic drug to determine if it is equivalent to the name-brand alternative. Equivalence is declared if the success rate of the generic brand is no more than 5% from that of the name-brand drug. In a study with 1000 individuals in each group, they find that 774, or 77.4%, are successfully treated using the name-brand drug, and 700, or 70%, respond to the generic drug. An equivalence test (exact test) with alpha = 0.05 failed to declare that the two drugs are equivalent. The researchers would now like to compute the power for actual differences ranging from 0 to 4%. Suppose that the true value for the response rate for the name-brand drug is 77%.

Note that the power is not calculated at the difference observed in the study, 77.4%. In fact, the difference observed in the study is larger than the proposed equivalence difference, 5%. It would make no sense to perform a power calculation for a difference larger than the equivalence difference. It is more informative to study a range of values smaller than or equal to the equivalence difference.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Equivalence Tests for Two Proportions [Differences]** procedure window by expanding **Proportions**, then **Two Independent Proportions**, then clicking on **Equivalence**, and then clicking on **Equivalence Tests for Two Proportions [Differences]**. You may then make the appropriate entries as listed below, or open **Example 6** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power	*Ignored since this is the Find setting*
Alpha	**0.05**
N1 (Sample Size Group 1)	**1000**
N2 (Sample Size Group 2)	**Use R**
R	**1.0**
D0.U (Upper Equivalence Difference)	**0.05**
D0.L (Lower Equivalence Difference)	**-D0.U**
D1 (Actual Difference)	**0.00 to 0.04 by 0.01**
P2 (Reference Group Proportion)	**0.77**
Test Type	**Likelihood Score (Farr. & Mann.)**
Options Tab	
Maximum N1 or N2 Exact	**100**

Output

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

Numeric Results

Numeric Results for Equivalence Tests Based on the Difference: P1 - P2
H0: P1-P2<=D0.L or P1-P2>=D0.U. H1: D0.L<P1-P2=D1<D0.U.
Test Statistic: Score test (Farrington & Manning)

Power	Sample Size Grp 1 N1	Sample Size Grp 2 N2	Prop Grp 2 P2	Lower Trivial Grp 1 Prop P1.0L	Upper Trivial Grp 1 Prop P1.0U	Lower Trivial Margin Diff D0.L	Upper Trivial Margin Diff D0.U	Actual Margin Diff D1	Target Alpha	Actual Alpha
0.6875	1000	1000	0.7700	0.7200	0.8200	-0.0500	0.0500	0.0000	0.0500	
0.6313	1000	1000	0.7700	0.7200	0.8200	-0.0500	0.0500	0.0100	0.0500	
0.4731	1000	1000	0.7700	0.7200	0.8200	-0.0500	0.0500	0.0200	0.0500	
0.2857	1000	1000	0.7700	0.7200	0.8200	-0.0500	0.0500	0.0300	0.0500	
0.1362	1000	1000	0.7700	0.7200	0.8200	-0.0500	0.0500	0.0400	0.0500	

Note: exact results based on the binomial were only calculated when both N1 and N2 were less than 100.

The power of the test ranges from 68.75% if the true difference is actually 0.0% to 13.62% if the true difference is 4%.

Example 7 – Finding the Sample Size using Proportions

A study is being designed to prove the equivalence of a new drug to the current standard. The current drug is effective in 85% of cases. The new drug, however, is cheaper to produce. The new drug will be deemed equivalent to the standard if its success rate is between 78% and 92%. What sample sizes are necessary to obtain 80% or 90% power for actual success rates ranging from 80% to 90%? The researchers will test at a significance level of 0.05 using the Farrington and Manning likelihood score test.

Setup

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

Option	Value
Data Tab	
Find (Solve For)	**N1**
Power	**0.80 0.90**
Alpha	**0.05**
N1 (Sample Size Group 1)	*Ignored since this is the Find setting*
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
P1.0U (Upper Equivalence Prop)	**0.92**
P1.0L (Lower Equivalence Prop)	**0.78**
P1.1 (Actual Proportion)	**0.80 to 0.90 by 0.02**
P2 (Reference Group Proportion)	**0.85**
Test Type	**Likelihood Score (Farr. & Mann.)**
Options Tab	
Maximum N1 or N2 Exact	**100**

Output

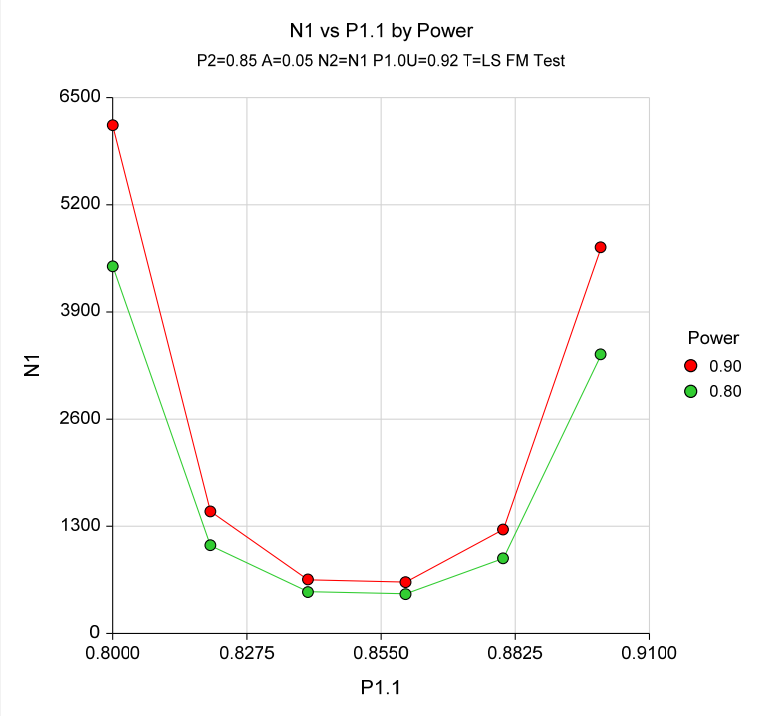
Click the Run button to perform the calculations and generate the following output.

Numeric Results and Plots

Numeric Results for Equivalence Tests Based on the Difference: P1 - P2
H0: P1-P2<=D0.L or P1-P2>=D0.U. H1: D0.L<P1-P2=D1<D0.U.
Test Statistic: Score test (Farrington & Manning)

Power	Sample Size Grp 1 N1	Sample Size Grp 2 N2	Prop Grp 2 P2	Lower Equiv. Grp 1 Prop P1.0L	Upper Equiv. Grp 1 Prop P1.0U	Lower Equiv. Margin Diff D0.L	Upper Equiv. Margin Diff D0.U	Actual Margin Diff D1	Target Alpha	Actual Alpha
0.9000	6166	6166	0.8500	0.7800	0.9200	-0.0700	0.0700	-0.0500	0.0500	
0.8001	4453	4453	0.8500	0.7800	0.9200	-0.0700	0.0700	-0.0500	0.0500	
0.9000	1480	1480	0.8500	0.7800	0.9200	-0.0700	0.0700	-0.0300	0.0500	
0.8002	1070	1070	0.8500	0.7800	0.9200	-0.0700	0.0700	-0.0300	0.0500	
0.9001	655	655	0.8500	0.7800	0.9200	-0.0700	0.0700	-0.0100	0.0500	
0.8008	503	503	0.8500	0.7800	0.9200	-0.0700	0.0700	-0.0100	0.0500	
0.9004	622	622	0.8500	0.7800	0.9200	-0.0700	0.0700	0.0100	0.0500	
0.8004	477	477	0.8500	0.7800	0.9200	-0.0700	0.0700	0.0100	0.0500	
0.9002	1261	1261	0.8500	0.7800	0.9200	-0.0700	0.0700	0.0300	0.0500	
0.8002	912	912	0.8500	0.7800	0.9200	-0.0700	0.0700	0.0300	0.0500	
0.9000	4685	4685	0.8500	0.7800	0.9200	-0.0700	0.0700	0.0500	0.0500	
0.8000	3386	3386	0.8500	0.7800	0.9200	-0.0700	0.0700	0.0500	0.0500	

Note: exact results based on the binomial were only calculated when both N1 and N2 were less than 100.



It is evident from these results that the sample sizes required to achieve 80% and 90% power depend a great deal on the actual value of the success rate, P1.1.

Chapter 216

Confidence Intervals for Two Proportions

Introduction

This routine calculates the group sample sizes necessary to achieve a specified interval width of the difference, ratio, or odds ratio of two independent proportions.

Caution: These procedures assume that the proportions obtained from future samples will be the same as the proportions that are specified. If the sample proportions are different from those specified when running these procedures, the interval width may be narrower or wider than specified.

Four Procedures Documented Here

There are four procedures in the menus described in this chapter. These procedures are very similar except for the type of parameterization. The parameterization can be in terms of proportions, differences in proportions, ratios of proportions, and odds ratios.

Technical Details

A background of the comparison of two proportions is given, followed by details of the confidence interval methods available in this procedure.

Comparing Two Proportions

Suppose you have two populations from which dichotomous (binary) responses will be recorded. The probability (or risk) of obtaining the event of interest in population 1 (the treatment group) is p_1 and in population 2 (the control group) is p_2. The corresponding failure proportions are given by $q_1 = 1 - p_1$ and $q_2 = 1 - p_2$.

The assumption is made that the responses from each group follow a binomial distribution. This means that the event probability p_i is the same for all subjects within a population and that the responses from one subject to the next are independent of one another.

Random samples of m and n individuals are obtained from these two populations. The data from these samples can be displayed in a 2-by-2 contingency table as follows

	Success	Failure	Total
Population 1	*a*	*c*	*m*
Population 2	*b*	*d*	*n*
Totals	*s*	*f*	*N*

The following alternative notation is sometimes used:

	Success	Failure	Total
Population 1	x_{11}	x_{12}	n_1
Population 2	x_{21}	x_{22}	n_2
Totals	m_1	m_2	*N*

The binomial proportions p_1 and p_2 are estimated from these data using the formulae

$$\hat{p}_1 = \frac{a}{m} = \frac{x_{11}}{n_1} \text{ and } \hat{p}_2 = \frac{b}{n} = \frac{x_{21}}{n_2}$$

When analyzing studies such as these, you usually want to compare the two binomial probabilities p_1 and p_2. The most direct methods of comparing these quantities are to calculate their difference or their ratio. If the binomial probability is expressed in terms of odds rather than probability, another measure is the odds ratio. Mathematically, these comparison parameters are

Parameter	Computation
Difference	$\delta = p_1 - p_2$
Risk Ratio	$\phi = p_1 / p_2$
Odds Ratio	$\psi = \frac{p_1 / q_1}{p_2 / q_2} = \frac{p_1 q_2}{p_2 q_1}$

The choice of which of these measures is used might at seem arbitrary, but it is important. Not only is their interpretation different, but, for small sample sizes, the coverage probabilities may be different.

Difference

The (risk) difference $\delta = p_1 - p_2$ is perhaps the most direct method of comparison between the two event probabilities. This parameter is easy to interpret and communicate. It gives the absolute impact of the treatment. However, there are subtle difficulties that can arise with its interpretation.

One interpretation difficulty occurs when the event of interest is rare. If a difference of 0.001 were reported for an event with a baseline probability of 0.40, we would probability dismiss this as being of little importance. That is, there usually little interest in a treatment that decreases the probability from 0.400 to 0.399. However, if the baseline probably of a disease was 0.002 and

0.001 was the decrease in the disease probability, this would represent a reduction of 50%. Thus we see that interpretation depends on the baseline probability of the event.

A similar situation occurs when the amount of possible difference is considered. Consider two events, one with a baseline event rate of 0.40 and the other with a rate of 0.02. What is the maximum decrease that can occur? Obviously, the first event rate can be decreased by an absolute amount of 0.40 which the second can only be decreased by a maximum of 0.02.

So, although creating the simple difference is a useful method of comparison, care must be taken that it fits the situation.

Ratio

The (risk) ratio $\phi = p_1 / p_2$ gives the relative change in the disease risk due to the application of the treatment. This parameter is also direct and easy to interpret. To compare this with the difference, consider a treatment that reduces the risk of disease for 0.1437 to 0.0793. Which single number is most enlightening, the fact that the absolute risk of disease has been decreased by 0.0644, or the fact that risk of disease in the treatment group is only 55.18% of that in the control group? In many cases, the percentage (risk ratio) communicates the impact of the treatment better than the absolute change.

Perhaps the biggest drawback to this parameter is that it cannot be calculated in one of the most common experimental designs: the case-control study.

Odds Ratio

Chances are usually communicated as long-term proportions or probabilities. In betting, chances are often given as odds. For example, the odds of a horse winning a race might be set at 10-to-1 or 3-to-2. How do you translate from odds to probability? An odds of 3-to-2 means that the event will occur three out of five times. That is, an odds of 3-to-2 (1.5) translates to a probability of winning of 0.60.

The odds of an event are calculated by dividing the event risk by the non-event risk. Thus, in our case of two populations, the odds are

$$o_1 = \frac{p_1}{1 - p_1} \text{ and } o_2 = \frac{p_2}{1 - p_2}$$

For example, if p_1 is 0.60, the odds are 0.60/0.4 = 1.5. Rather than represent the odds as a decimal amount, it is re-scaled into whole numbers. Thus, instead of saying the odds are 1.5-to-1, we say they are 3-to-2.

Another way to compare proportions is to compute the ratio of their odds. The odds ratio of two events is

$$\psi = \frac{o_1}{o_2}$$

$$= \frac{\dfrac{p_1}{1 - p_1}}{\dfrac{p_2}{1 - p_2}}$$

Although the odds ratio is more complicated to interpret than the risk ratio, it is often the parameter of choice. Reasons for this include the fact that the odds ratio can be accurately estimated from case-control studies, while the risk ratio cannot. Also, the odds ratio is the basis of logistic regression (used to study the influence of risk factors). Furthermore, the odds ratio is the natural parameter in the conditional likelihood of the two-group, binomial-response design. Finally, when the baseline event-rates are rare, the odds ratio provides a close approximation to the risk ratio since, in this case, $1-p_1 \approx 1-p_2$, so that

$$\psi = \frac{\frac{p_1}{1-p_1}}{\frac{p_2}{1-p_2}} \approx \frac{p_1}{p_2} = \phi$$

Confidence Intervals for the Difference

Many methods have been devised for computing confidence intervals for the difference between two proportions $\delta = p_1 - p_2$. Seven of these methods are available in the Confidence Intervals for Two Proportions [Proportions] using Proportions and Confidence Intervals for Two Proportions [Differences] procedures. The seven confidence interval methods are

1. Score (Farrington and Manning)
2. Score (Miettinen and Nurminen)
3. Score with Correction for Skewness (Gart and Nam)
4. Score (Wilson)
5. Score with Continuity Correction (Wilson)
6. Chi-Square with Continuity Correction (Yates)
7. Chi-Square (Pearson)

Newcombe (1998b) conducted a comparative evaluation of eleven confidence interval methods. He recommended that the modified Wilson score method be used instead of the Pearson Chi-Square or the Yate's Corrected Chi-Square. Beal (1987) found that the Score methods performed very well. The lower *L* and upper *U* limits of these intervals are computed as follows. Note that, unless otherwise stated, $z = |z_{\alpha/2}|$ is the appropriate percentile from the standard normal distribution.

C.I. for Difference: Farrington and Manning's Score

Farrington and Manning (1990) proposed a test statistic for testing whether the difference is equal to a specified value δ_0. The regular MLE's $\hat{p}_1$ and $\hat{p}_2$ are used in the numerator of the score statistic while MLE's $\tilde{p}_1$ and $\tilde{p}_2$ constrained so that $\tilde{p}_1 - \tilde{p}_2 = \delta_0$ are used in the denominator. The significance level of the test statistic is based on the asymptotic normality of the score statistic.

The test statistic formula is

$$z_{FMD} = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0}{\sqrt{\left(\frac{\tilde{p}_1\tilde{q}_1}{n_1} + \frac{\tilde{p}_2\tilde{q}_2}{n_2}\right)}}$$

where the estimates $\tilde{p}_1$ and $\tilde{p}_2$ are computed as in the corresponding test of Miettinen and Nurminen (1985) given as

$$\tilde{p}_1 = \tilde{p}_2 + \delta_0$$

$$\tilde{p}_2 = 2B\cos(A) - \frac{L_2}{3L_3}$$

$$A = \frac{1}{3}\left[\pi + \cos^{-1}\left(\frac{C}{B^3}\right)\right]$$

$$B = \text{sign}(C)\sqrt{\frac{L_2^2}{9L_3^2} - \frac{L_1}{3L_3}}$$

$$C = \frac{L_2^3}{27L_3^3} - \frac{L_1L_2}{6L_3^2} + \frac{L_0}{2L_3}$$

$$L_0 = x_{21}\delta_0(1-\delta_0)$$

$$L_1 = [N_2\delta_0 - N - 2x_{21}]\delta_0 + M_1$$

$$L_2 = (N + N_2)\delta_0 - N - M_1$$

$$L_3 = N$$

Farrington and Manning (1990) proposed inverting their score test to find the confidence interval. The lower limit is found by solving

$$z_{FMD} = |z_{\alpha/2}|$$

and the upper limit is the solution of

$$z_{FMD} = -|z_{\alpha/2}|$$

C.I. for Difference: Miettinen and Nurminen's Score

Miettinen and Nurminen (1985) proposed a test statistic for testing whether the difference is equal to a specified value δ_0. The regular MLE's $\hat{p}_1$ and $\hat{p}_2$ are used in the numerator of the score statistic while MLE's $\tilde{p}_1$ and $\tilde{p}_2$ constrained so that $\tilde{p}_1 - \tilde{p}_2 = \delta_0$ are used in the denominator. A correction factor of $N/(N\text{-}1)$ is applied to make the variance estimate less biased. The significance level of the test statistic is based on the asymptotic normality of the score statistic.

The formula for computing this test statistic is

$$z_{MND} = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0}{\sqrt{\left(\frac{\tilde{p}_1\tilde{q}_1}{n_1} + \frac{\tilde{p}_2\tilde{q}_2}{n_2}\right)\left(\frac{N}{N-1}\right)}}$$

where

$$\tilde{p}_1 = \tilde{p}_2 + \delta_0$$

$$\tilde{p}_2 = 2B\cos(A) - \frac{L_2}{3L_3}$$

$$A = \frac{1}{3}\left[\pi + \cos^{-1}\left(\frac{C}{B^3}\right)\right]$$

$$B = \text{sign}(C)\sqrt{\frac{L_2^2}{9L_3^2} - \frac{L_1}{3L_3}}$$

$$C = \frac{L_2^3}{27L_3^3} - \frac{L_1L_2}{6L_3^2} + \frac{L_0}{2L_3}$$

$$L_0 = x_{21}\delta_0(1-\delta_0)$$

$$L_1 = [N_2\delta_0 - N - 2x_{21}]\delta_0 + M_1$$

$$L_2 = (N + N_2)\delta_0 - N - M_1$$

$$L_3 = N$$

Miettinen and Nurminen (1985) proposed inverting their score test to find the confidence interval. The lower limit is found by solving

$$z_{MND} = |z_{\alpha/2}|$$

and the upper limit is the solution of

$$z_{MND} = -|z_{\alpha/2}|$$

C.I. for Difference: Gart and Nam's Score

Gart and Nam (1990) page 638 proposed a modification to the Farrington and Manning (1990) difference test that corrected for skewness. Let $z_{FM}(\delta)$ stand for the Farrington and Manning difference test statistic described above. The skewness corrected test statistic z_{GN} is the appropriate solution to the quadratic equation

$$(-\tilde{\gamma})z_{GND}^2 + (-1)z_{GND} + (z_{FMD}(\delta) + \tilde{\gamma}) = 0$$

where

$$\tilde{\gamma} = \frac{\tilde{V}^{3/2}(\delta)}{6}\left(\frac{\tilde{p}_1\tilde{q}_1(\tilde{q}_1 - \tilde{p}_1)}{n_1^2} - \frac{\tilde{p}_2\tilde{q}_2(\tilde{q}_2 - \tilde{p}_2)}{n_2^2}\right)$$

Gart and Nam (1988) proposed inverting their score test to find the confidence interval. The lower limit is found by solving

$$z_{GND} = |z_{\alpha/2}|$$

and the upper limit is the solution of

$$z_{GND} = -|z_{\alpha/2}|$$

C.I. for Difference: Wilson's Score as Modified by Newcombe (with and without Continuity Correction)

For details, see Newcombe (1998b), page 876.

$$L = \hat{p}_1 - \hat{p}_2 - B$$

$$U = \hat{p}_1 - \hat{p}_2 + C$$

where

$$B = z\sqrt{\frac{l_1(1-l_1)}{m} + \frac{u_2(1-u_2)}{n}}$$

$$C = z\sqrt{\frac{u_1(1-u_1)}{m} + \frac{l_2(1-l_2)}{n}}$$

and l_1 and u_1 are the roots of

$$|p_1 - \hat{p}_1| - z\sqrt{\frac{p_1(1-p_1)}{m}} = 0$$

and l_2 and u_2 are the roots of

$$|p_2 - \hat{p}_2| - z\sqrt{\frac{p_2(1-p_2)}{n}} = 0$$

C.I. for Difference: Yate's Chi-Square with Continuity Correction

For details, see Newcombe (1998b), page 875.

$$L = \hat{p}_1 - \hat{p}_2 - z\left[\sqrt{\left(\frac{\hat{p}_1(1-\hat{p}_1)}{m} + \frac{\hat{p}_2(1-\hat{p}_2)}{n}\right)} + \frac{1}{2}\left(\frac{1}{m} + \frac{1}{n}\right)\right]$$

$$U = \hat{p}_1 - \hat{p}_2 + z\left[\sqrt{\left(\frac{\hat{p}_1(1-\hat{p}_1)}{m} + \frac{\hat{p}_2(1-\hat{p}_2)}{n}\right)} + \frac{1}{2}\left(\frac{1}{m} + \frac{1}{n}\right)\right]$$

C.I. for Difference: Pearson's Chi-Square

For details, see Newcombe (1998b), page 875.

$$L = \hat{p}_1 - \hat{p}_2 - z\sqrt{\left(\frac{\hat{p}_1(1-\hat{p}_1)}{m} + \frac{\hat{p}_2(1-\hat{p}_2)}{n}\right)}$$

$$U = \hat{p}_1 - \hat{p}_2 + z\sqrt{\left(\frac{\hat{p}_1(1-\hat{p}_1)}{m} + \frac{\hat{p}_2(1-\hat{p}_2)}{n}\right)}$$

For each of the seven methods, one-sided intervals may be obtained by replacing $\alpha/2$ by α.

For two-sided intervals, the distance from the difference in sample proportions to each of the limits may be different. Thus, instead of specifying the distance to the limits we specify the width of the interval, *W*.

The basic equation for determining sample size for a two-sided interval when *W* has been specified is

$$W = U - L$$

For one-sided intervals, the distance from the variance ratio to limit, *D*, is specified.

The basic equation for determining sample size for a one-sided upper limit when *D* has been specified is

$$D = U - \left(\hat{p}_1 - \hat{p}_2\right)$$

The basic equation for determining sample size for a one-sided lower limit when *D* has been specified is

$$D = \left(\hat{p}_1 - \hat{p}_2\right) - L$$

Each of these equations can be solved for any of the unknown quantities in terms of the others.

Confidence Intervals for the Ratio (Relative Risk)

Many methods have been devised for computing confidence intervals for the ratio (relative risk) of two proportions $\phi = p_1 / p_2$. Six of these methods are available in the Confidence Intervals for Two Proportions [Ratios] procedure. The six confidence interval methods are

1. Score (Farrington and Manning)
2. Score (Miettinen and Nurminen)
3. Score with Correction for Skewness (Gart and Nam)
4. Logarithm (Katz)
5. Logarithm + 1/2 (Walter)
6. Fleiss

C.I. for Ratio: Farrington and Manning's Score

Farrington and Manning (1990) proposed a test statistic for testing whether the ratio is equal to a specified value ϕ_0. The regular MLE's $\hat{p}_1$ and $\hat{p}_2$ are used in the numerator of the score statistic while MLE's $\tilde{p}_1$ and $\tilde{p}_2$ constrained so that $\tilde{p}_1 / \tilde{p}_2 = \phi_0$ are used in the denominator. A correction factor of N/(N-1) is applied to increase the variance estimate. The significance level of the test statistic is based on the asymptotic normality of the score statistic.

Here is the formula for computing the test

$$z_{FMR} = \frac{\hat{p}_1 / \hat{p}_2 - \phi_0}{\sqrt{\left(\frac{\tilde{p}_1 \tilde{q}_1}{n_1} + \phi_0^2 \frac{\tilde{p}_2 \tilde{q}_2}{n_2} \right)}}$$

where

$$\tilde{p}_1 = \tilde{p}_2 \phi_0$$

$$\tilde{p}_2 = \frac{-B - \sqrt{B^2 - 4AC}}{2A}$$

$$A = N\phi_0$$

$$B = -\left[N_1 \phi_0 + x_{11} + N_2 + x_{21} \phi_0 \right]$$

$$C = M_1$$

as in the test of Miettinen and Nurminen (1985).

Farrington and Manning (1990) proposed inverting their score test to find the confidence interval. The lower limit is found by solving

$$z_{FMR} = \left| z_{\alpha/2} \right|$$

and the upper limit is the solution of

$$z_{FMR} = -\left| z_{\alpha/2} \right|$$

C.I. for Ratio: Miettinen and Nurminen's Score

Miettinen and Nurminen (1985) proposed a test statistic for testing whether the ratio is equal to a specified value ϕ_0. The regular MLE's $\hat{p}_1$ and $\hat{p}_2$ are used in the numerator of the score statistic while MLE's $\tilde{p}_1$ and $\tilde{p}_2$ constrained so that $\tilde{p}_1 / \tilde{p}_2 = \phi_0$ are used in the denominator. A correction factor of N/(N-1) is applied to make the variance estimate less biased. The significance level of the test statistic is based on the asymptotic normality of the score statistic.

Here is the formula for computing the test

$$z_{MNR} = \frac{\hat{p}_1 / \hat{p}_2 - \phi_0}{\sqrt{\left(\frac{\tilde{p}_1\tilde{q}_1}{n_1} + \phi_0^2 \frac{\tilde{p}_2\tilde{q}_2}{n_2}\right)\left(\frac{N}{N-1}\right)}}$$

where

$$\tilde{p}_1 = \tilde{p}_2\phi_0$$

$$\tilde{p}_2 = \frac{-B - \sqrt{B^2 - 4AC}}{2A}$$

$$A = N\phi_0$$

$$B = -\left[N_1\phi_0 + x_{11} + N_2 + x_{21}\phi_0\right]$$

$$C = M_1$$

Miettinen and Nurminen (1985) proposed inverting their score test to find the confidence interval. The lower limit is found by solving

$$z_{MNR} = \left|z_{\alpha/2}\right|$$

and the upper limit is the solution of

$$z_{MNR} = -\left|z_{\alpha/2}\right|$$

C.I. for Ratio: Gart and Nam's Score

Gart and Nam (1988) page 329 proposed a modification to the Farrington and Manning (1988) ratio test that corrected for skewness. Let $z_{FM}(\phi)$ stand for the Farrington and Manning ratio test statistic described above. The skewness corrected test statistic z_{GN} is the appropriate solution to the quadratic equation

$$\left(-\tilde{\varphi}\right)z_{GNR}^2 + \left(-1\right)z_{GNR} + \left(z_{FMR}(\phi) + \tilde{\varphi}\right) = 0$$

where

$$\tilde{\varphi} = \frac{1}{6\tilde{u}^{3/2}}\left(\frac{\tilde{q}_1(\tilde{q}_1 - \tilde{p}_1)}{n_1^2\tilde{p}_1^2} - \frac{\tilde{q}_2(\tilde{q}_2 - \tilde{p}_2)}{n_2^2\tilde{p}_2^2}\right)$$

$$\tilde{u} = \frac{\tilde{q}_1}{n_1\tilde{p}_1} + \frac{\tilde{q}_2}{n_2\tilde{p}_2}$$

Gart and Nam (1988) proposed inverting their score test to find the confidence interval. The lower limit is found by solving

$$z_{GNR} = |z_{\alpha/2}|$$

and the upper limit is the solution of

$$z_{GNR} = -|z_{\alpha/2}|$$

C.I. for Ratio: Logarithm (Katz)

This was one of the first methods proposed for computing confidence intervals for risk ratios.

For details, see Gart and Nam (1988), page 324.

$$L = \hat{\phi}\exp\left(-z\sqrt{\frac{\hat{q}_1}{n\hat{p}_1}+\frac{\hat{q}_2}{n\hat{p}_2}}\right)$$

$$U = \hat{\phi}\exp\left(z\sqrt{\frac{\hat{q}_1}{n\hat{p}_1}+\frac{\hat{q}_2}{n\hat{p}_2}}\right)$$

where

$$\hat{\phi} = \frac{\hat{p}_1}{\hat{p}_2}$$

C.I. for Ratio: Logarithm (Walters)

For details, see Gart and Nam (1988), page 324.

$$L = \hat{\phi}\exp\left(-z\sqrt{\hat{u}}\right)$$

$$U = \hat{\phi}\exp\left(z\sqrt{\hat{u}}\right)$$

where

$$\hat{\phi} = \exp\left(\ln\left(\frac{a+\frac{1}{2}}{m+\frac{1}{2}}\right)-\ln\left(\frac{b+\frac{1}{2}}{n+\frac{1}{2}}\right)\right)$$

$$\hat{u} = \frac{1}{a+\frac{1}{2}}-\frac{1}{m+\frac{1}{2}}+\frac{1}{b+\frac{1}{2}}-\frac{1}{n+\frac{1}{2}}$$

$$\tilde{q}_2 = 1-\tilde{p}_2$$

$$V = \left(\phi^2\left(\frac{\tilde{q}_1}{m\tilde{p}_1}+\frac{\tilde{q}_2}{n\tilde{p}_2}\right)\right)^{-1}$$

$$\tilde{p}_1 = \phi\tilde{p}_2$$

$$\tilde{q}_1 = 1-\tilde{p}_1$$

$$\tilde{q}_2 = 1-\tilde{p}_2$$

$$\tilde{\mu}_3 = v^{3/2}\left(\frac{\tilde{q}_1(\tilde{q}_1 - \tilde{p}_1)}{(m\tilde{p}_1)^2} - \frac{\tilde{q}_2(\tilde{q}_2 - \tilde{p}_2)}{(n\tilde{p}_2)^2}\right)$$

$$v = \left(\frac{\tilde{q}_1}{m\tilde{p}_1} + \frac{\tilde{q}_2}{n\tilde{p}_2}\right)^{-1}$$

C.I. for Odds Ratio and Relative Risk: Iterated Method of Fleiss

Fleiss (1981) presents an improved confidence interval for the odds ratio and relative risk. This method forms the confidence interval as all those value of the odds ratio which would not be rejected by a chi-square hypothesis test. Fleiss gives the following details about how to construct this confidence interval. To compute the lower limit, do the following.

1. For a trial value of ψ, compute the quantities *X*, *Y*, *W*, *F*, *U*, and *V* using the formulas

$$X = \psi(m+s)+(n-s)$$

$$Y = \sqrt{X^2 - 4ms\psi(\psi-1)}$$

$$A = \frac{X - Y}{2(\psi-1)}$$

$$B = s - A$$

$$C = m - A$$

$$D = f - m + A$$

$$W = \frac{1}{A} + \frac{1}{B} + \frac{1}{C} + \frac{1}{D}$$

$$F = \left(a - A - \tfrac{1}{2}\right)^2 W - z^2_{\alpha/2}$$

$$T = \frac{1}{2(\psi-1)^2}\left(Y - n - \frac{\psi-1}{Y}\left[X(m+s) - 2ms(2\psi-1)\right]\right)$$

$$U = \frac{1}{B^2} + \frac{1}{C^2} - \frac{1}{A^2} - \frac{1}{D^2}$$

$$V = T\left[\left(a - A - \tfrac{1}{2}\right)^2 U - 2W\left(a - A - \tfrac{1}{2}\right)\right]$$

   Finally, use the updating equation below to calculate a new value for the odds ratio using the updating equation

$$\psi^{(k+1)} = \psi^{(k)} - \frac{F}{V}$$

2. Continue iterating until the value of *F* is arbitrarily close to zero.

The upper limit is found by substituting $+\frac{1}{2}$ for $-\frac{1}{2}$ in the formulas for *F* and *V*.

Confidence limits for the *relative risk* can be calculated using the expected counts *A*, *B*, *C*, and *D* from the last iteration of the above procedure. The lower limit of the relative risk

$$\phi_{lower} = \frac{A_{lower} n}{B_{lower} m}$$

$$\phi_{upper} = \frac{A_{upper} n}{B_{upper} m}$$

Confidence Intervals for the Odds Ratio

Many methods have been devised for computing confidence intervals for the odds ratio of two proportions

$$\psi = \frac{\frac{p_1}{1-p_1}}{\frac{p_2}{1-p_2}}$$

Eight of these methods are available in the Confidence Intervals for Two Proportions [Odds Ratios] procedure. The eight confidence interval methods are

1. Exact (Conditional)
2. Score (Farrington and Manning)
3. Score (Miettinen and Nurminen)
4. Fleiss
5. Logarithm
6. Mantel-Haenszel
7. Simple
8. Simple + 1/2

C.I. for Odds Ratio: Conditional Exact

The conditional exact confidence interval of the odds ratio is calculated using the noncentral hypergeometric distribution as given in Sahai and Khurshid (1995). That is, a $100(1-\alpha)\%$ confidence interval is found by searching for ψ_L and ψ_U such that

$$\frac{\sum_{k=x}^{k_2} \binom{n_1}{k}\binom{n_2}{m_1-k}(\psi_L)^k}{\sum_{k=k_1}^{k_2} \binom{n_1}{k}\binom{n_2}{m_1-k}(\psi_L)^k} = \frac{\alpha}{2}$$

and

$$\frac{\sum_{k=k_1}^{x}\binom{n_1}{k}\binom{n_2}{m_1-k}(\psi_U)^k}{\sum_{k=k_1}^{k_2}\binom{n_1}{k}\binom{n_2}{m_1-k}(\psi_U)^k}=\frac{\alpha}{2}$$

where

$$k_1=\max(0,m_1-n_1) \text{ and } k_2=\min(n_1,m_1)$$

Farrington and Manning's Test of the Odds Ratio

Farrington and Manning (1990) developed a test statistic similar to that of Miettinen and Nurminen but with the factor $N/(N\text{-}1)$ removed.

The formula for computing this test statistic is

$$z_{FMO}=\frac{\frac{(\hat{p}_1-\tilde{p}_1)}{\tilde{p}_1\tilde{q}_1}-\frac{(\hat{p}_2-\tilde{p}_2)}{\tilde{p}_2\tilde{q}_2}}{\sqrt{\left(\frac{1}{N_2\tilde{p}_1\tilde{q}_1}+\frac{1}{N_2\tilde{p}_2\tilde{q}_2}\right)}}$$

where the estimates $\tilde{p}_1$ and $\tilde{p}_2$ are computed as in the corresponding test of Miettinen and Nurminen (1985) as

$$\tilde{p}_1=\frac{\tilde{p}_2\psi_0}{1+\tilde{p}_2(\psi_0-1)}$$

$$\tilde{p}_2=\frac{-B+\sqrt{B^2-4AC}}{2A}$$

$$A=N_2(\psi_0-1)$$

$$B=N_1\psi_0+N_2-M_1(\psi_0-1)$$

$$C=-M_1$$

Farrington and Manning (1990) proposed inverting their score test to find the confidence interval. The lower limit is found by solving

$$z_{FMO}=|z_{\alpha/2}|$$

and the upper limit is the solution of

$$z_{FMO}=-|z_{\alpha/2}|$$

C.I. for Odds Ratio: Miettinen and Nurminen

Miettinen and Nurminen (1985) proposed a test statistic for testing whether the odds ratio is equal to a specified value ψ_0. Because the approach they used with the difference and ratio does not easily extend to the odds ratio, they used a score statistic approach for the odds ratio. The regular MLE's are $\hat{p}_1$ and $\hat{p}_2$. The constrained MLE's are $\tilde{p}_1$ and $\tilde{p}_2$, These estimates are constrained so that $\tilde{\psi} = \psi_0$. A correction factor of N/(N-1) is applied to make the variance estimate less biased. The significance level of the test statistic is based on the asymptotic normality of the score statistic.

The formula for computing the test statistic is

$$z_{MNO} = \frac{\frac{(\hat{p}_1 - \tilde{p}_1)}{\tilde{p}_1\tilde{q}_1} - \frac{(\hat{p}_2 - \tilde{p}_2)}{\tilde{p}_2\tilde{q}_2}}{\sqrt{\left(\frac{1}{N_2\tilde{p}_1\tilde{q}_1} + \frac{1}{N_2\tilde{p}_2\tilde{q}_2}\right)\left(\frac{N}{N-1}\right)}}$$

where

$$\tilde{p}_1 = \frac{\tilde{p}_2\psi_0}{1 + \tilde{p}_2(\psi_0 - 1)}$$

$$\tilde{p}_2 = \frac{-B + \sqrt{B^2 - 4AC}}{2A}$$

$$A = N_2(\psi_0 - 1)$$

$$B = N_1\psi_0 + N_2 - M_1(\psi_0 - 1)$$

$$C = -M_1$$

Miettinen and Nurminen (1985) proposed inverting their score test to find the confidence interval. The lower limit is found by solving

$$z_{MNO} = |z_{\alpha/2}|$$

and the upper limit is the solution of

$$z_{MNO} = -|z_{\alpha/2}|$$

C.I. for Odds Ratio: Iterated Method of Fleiss

Fleiss (1981) presents an improve confidence interval for the odds ratio. This method forms the confidence interval as all those value of the odds ratio which would not be rejected by a chi-square hypothesis test. Fleiss gives the following details about how to construct this confidence interval. To compute the lower limit, do the following.

1. For a trial value of ψ, compute the quantities *X*, *Y*, *W*, *F*, *U*, and *V* using the formulas

$$X = \psi(m+s)+(n-s)$$

$$Y = \sqrt{X^2 - 4ms\psi(\psi-1)}$$

$$A = \frac{X-Y}{2(\psi-1)}$$

$$B = s - A$$

$$C = m - A$$

$$D = f - m + A$$

$$W = \frac{1}{A}+\frac{1}{B}+\frac{1}{C}+\frac{1}{D}$$

$$F = \left(a - A - \tfrac{1}{2}\right)^2 W - z^2_{\alpha/2}$$

$$T = \frac{1}{2(\psi-1)^2}\left(Y - n_{\cdot\cdot} - \frac{\psi-1}{Y}\left[X(m+s) - 2ms(2\psi-1)\right]\right)$$

$$U = \frac{1}{B^2}+\frac{1}{C^2}-\frac{1}{A^2}-\frac{1}{D^2}$$

$$V = T\left[\left(a - A - \tfrac{1}{2}\right)^2 U - 2W\left(a - A - \tfrac{1}{2}\right)\right]$$

   Finally, use the updating equation below to calculate a new value for the odds ratio using the updating equation

$$\psi^{(k+1)} = \psi^{(k)} - \frac{F}{V}$$

2. Continue iterating until the value of *F* is arbitrarily close to zero.

The upper limit is found by substituting $+\frac{1}{2}$ for $-\frac{1}{2}$ in the formulas for *F* and *V*.

Confidence limits for the *relative risk* can be calculated using the expected counts *A*, *B*, *C*, and *D* from the last iteration of the above procedure. The lower limit of the relative risk

$$\phi_{lower} = \frac{A_{lower}n}{B_{lower}m}$$

$$\phi_{upper} = \frac{A_{upper}n}{B_{upper}m}$$

C.I. for Odds Ratio: Mantel-Haenszel

The common estimate of the logarithm of the odds ratio is used to create this estimator. That is

$$\ln(\hat{\psi}) = \ln\left(\frac{ad}{bc}\right)$$

The standard error of this estimator is estimated using the Robins, Breslow, Greenland (1986) estimator which performs well in most situations. The standard error is given by

$$\text{se}(\ln(\hat{\psi})) = \sqrt{\frac{A}{2C} + \frac{AD + BC}{2CD} + \frac{B}{2D}}$$

where

$$A = \frac{a + d}{N}$$

$$B = \frac{b + c}{N}$$

$$C = \frac{ad}{N}$$

$$D = \frac{bc}{N}$$

The confidence limits are calculated as

$$\hat{\psi}_{lower} = \exp(\ln(\hat{\psi}) - z_{1-\alpha/2} se(\ln(\hat{\psi})))$$

$$\hat{\psi}_{upper} = \exp\left(\ln(\hat{\psi}) + z_{1-\alpha/2} se\left(\ln(\hat{\psi})\right)\right)$$

C.I. for Odds Ratio: Simple, Simple + ½, and Logarithm

The simple estimate of the odds ratio uses the formula

$$\hat{\psi} = \frac{\hat{p}_1\hat{q}_2}{\hat{p}_2\hat{q}_1}$$

$$= \frac{ad}{bc}$$

The standard error of this estimator is estimated by

$$se(\hat{\psi}) = \hat{\psi}\sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$$

Problems occur if any one of the quantities *a*, *b*, *c*, or *d* are zero. To correct this problem, many authors recommend adding one-half to each cell count so that a zero cannot occur. Now, the formulas become

$$\hat{\psi}' = \frac{(a + 0.5)(d + 0.5)}{(b + 0.5)(c + 0.5)}$$

and

$$se(\hat{\psi}') = \hat{\psi}'\sqrt{\frac{1}{a+0.5}+\frac{1}{b+0.5}+\frac{1}{c+0.5}+\frac{1}{d+0.5}}$$

The distribution of these direct estimates of the odds ratio do not converge to normality as fast as does their logarithm, so the logarithm of the odds ratio is used to form confidence intervals. The formula for the standard error of the log odds ratio is

$$L' = \ln(\hat{\psi}')$$

and

$$se(L') = \sqrt{\frac{1}{a+0.5}+\frac{1}{b+0.5}+\frac{1}{c+0.5}+\frac{1}{d+0.5}}$$

A $100(1-\alpha)\%$ confidence interval for the log odds ratio is formed using the standard normal distribution as follows

$$\hat{\psi}_{lower} = \exp(L' - z_{1-\alpha/2}se(L'))$$
$$\hat{\psi}_{upper} = \exp(L' + z_{1-\alpha/2}se(L'))$$

See Fleiss et al (2003) for more details.

Confidence Level

The confidence level, $1-\alpha$, has the following interpretation. If thousands of random samples of size n_1 and n_2 are drawn from populations 1 and 2, respectively, and a confidence interval for the true difference/ratio/odds ratio of proportions is calculated for each pair of samples, the proportion of those intervals that will include the true difference/ratio/odds ratio of proportions is $1-\alpha$.

Procedure Options

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

Data Tab (Common Options)

This chapter covers four procedures, each of which has different options. This section documents options that are common to all four procedures. Following this section, the unique options for each procedure (proportions, differences, ratios, and odds ratios) will be documented.

Solve For

Find (Solve For)

This option specifies the parameter to be solved for from the other parameters.

Confidence

Confidence Level

The confidence level, $1 - \alpha$, has the following interpretation. If thousands of random samples of size n_1 and n_2 are drawn from populations 1 and 2, respectively, and a confidence interval for the true difference/ratio/odds ratio of proportions is calculated for each pair of samples, the proportion of those intervals that will include the true difference/ratio/odds ratio of proportions is $1 - \alpha$.

Often, the values 0.95 or 0.99 are used. You can enter single values or a range of values such as *0.90, 0.95 or 0.90 to 0.99 by 0.01*.

Sample Size

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 the operator [*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*.

One-Sided or Two-Sided Interval

Interval Type

Specify whether the interval to be used will be a two-sided confidence interval, an interval that has only an upper limit, or an interval that has only a lower limit.

Data Tab (Proportions)

This section documents options that are used when the parameterization is in terms of the values of the two sample proportions, P1 and P2. The corresponding procedure is Confidence Intervals for the Difference between Two Proportions using Proportions.

Precision

Confidence Interval Width (Two-Sided)

This is the distance from the lower confidence limit to the upper confidence limit.

You can enter a single value or a list of values. The value(s) must be greater than zero.

Distance from Diff to Limit (One-Sided)

This is the distance from the difference in sample proportions to the lower or upper limit of the confidence interval, depending on whether the Interval Type is set to Lower Limit or Upper Limit.

You can enter a single value or a list of values. The value(s) must be greater than zero.

Proportions (Difference = P1 – P2)

P1 (Proportion Group 1)

Enter an estimate of the proportion for group 1. The sample size and width calculations assume that the value entered here is the proportion estimate that is obtained from the sample. If the sample proportion is different from the one specified here, the width may be narrower or wider than specified.

The value(s) must be between 0.0001 and 0.9999.

You can enter a range of values such as *.1 .2 .3* or *.1 to .5 by .1*.

P2 (Proportion Group 2)

Enter an estimate of the proportion for group 2. The sample size and width calculations assume that the value entered here is the proportion estimate that is obtained from the sample. If the sample proportion is different from the one specified here, the width may be narrower or wider than specified.

The value(s) must be between 0.0001 and 0.9999.

You can enter a range of values such as *.1 .2 .3* or *.1 to .5 by .1*.

Confidence Interval Method

Confidence Interval Formula

Specify the formula to be in used in calculation of confidence intervals.

- **Score (Farrington & Manning)**
  This formula is based on inverting Farrington and Manning's score test.

- **Score (Miettinen & Nurminen)**
  This formula is based on inverting Miettinen and Nurminen's score test.

- **Score w/ Skewness (Gart & Nam)**
  This formula is based on inverting Gart and Nam's score test, with a correction for skewness.

- **Score (Wilson)**
  This formula is based one the Wilson score method for a single proportion, without continuity correction.

- **Score (Wilson C.C.)**
  This formula is based one the Wilson score method for a single proportion, with continuity correction.

- **Chi-Square C.C. (Yates)**
  This is the commonly used simple asymptotic method, with continuity correction.

- **Chi-Square (Pearson)**
  This is the commonly used simple asymptotic method, without continuity correction.

Data Tab (Differences)

This section documents options that are used when the parameterization is in terms of the difference in sample proportions and the value of the second sample proportion, P2. The corresponding procedure is Confidence Intervals for the Difference between Two Proportions using Differences.

Precision

Confidence Interval Width (Two-Sided)

This is the distance from the lower confidence limit to the upper confidence limit.

You can enter a single value or a list of values. The value(s) must be greater than zero.

Distance from Diff to Limit (One-Sided)

This is the distance from the difference in sample proportions to the lower or upper limit of the confidence interval, depending on whether the Interval Type is set to Lower Limit or Upper Limit.

You can enter a single value or a list of values. The value(s) must be greater than zero.

Proportions (Difference = P1 – P2)

Difference in Sample Proportions

Enter an estimate of the difference between sample proportion 1 and sample proportion 2. The sample size and width calculations assume that the value entered here is the difference estimate that is obtained from the sample. If the sample difference is different from the one specified here, the width may be narrower or wider than specified.

The value(s) must be between -1 and 1, and such that P1 = Difference + P2 is between 0.0001 and 0.9999.

You can enter a range of values such as *.1 .2 .3* or *.1 to .5 by .1*.

P2 (Proportion Group 2)

Enter an estimate of the proportion for group 2. The sample size and width calculations assume that the value entered here is the proportion estimate that is obtained from the sample. If the sample proportion is different from the one specified here, the width may be narrower or wider than specified.

The value(s) must be between 0.0001 and 0.9999.

You can enter a range of values such as *.1 .2 .3* or *.1 to .5 by .1*.

Confidence Interval Method

Confidence Interval Formula

Specify the formula to be in used in calculation of confidence intervals.

- **Score (Farrington & Manning)**
  This formula is based on inverting Farrington and Manning's score test.
- **Score (Miettinen & Nurminen)**
  This formula is based on inverting Miettinen and Nurminen's score test.
- **Score w/ Skewness (Gart & Nam)**
  This formula is based on inverting Gart and Nam's score test, with a correction for skewness.
- **Score (Wilson)**
  This formula is based one the Wilson score method for a single proportion, without continuity correction.
- **Score (Wilson C.C.)**
  This formula is based one the Wilson score method for a single proportion, with continuity correction.
- **Chi-Square C.C. (Yates)**
  This is the commonly used simple asymptotic method, with continuity correction.
- **Chi-Square (Pearson)**
  This is the commonly used simple asymptotic method, without continuity correction.

Data Tab (Ratios)

This section documents options that are used when the parameterization is in terms of the ratio of sample proportions and the value of the second sample proportion, P2. The corresponding procedure is Confidence Intervals for the Difference between Two Proportions using Ratios.

Precision

Confidence Interval Width (Two-Sided)

This is the distance from the lower confidence limit to the upper confidence limit.

You can enter a single value or a list of values. The value(s) must be greater than zero.

Distance from Ratio to Limit (One-Sided)

This is the distance from the ratio of sample proportions to the lower or upper limit of the confidence interval, depending on whether the Interval Type is set to Lower Limit or Upper Limit.

You can enter a single value or a list of values. The value(s) must be greater than zero.

Proportions (Ratio = P1/P2)

Ratio of Sample Proportions

Enter an estimate of the ratio of sample proportion 1 to sample proportion 2. The sample size and width calculations assume that the value entered here is the ratio estimate that is obtained from the samples. If the sample ratio is different from the one specified here, the width may be narrower or wider than specified.

The value(s) must be greater than 0, and such that P1 = Ratio * P2 is between 0.0001 and 0.9999.

You can enter a range of values such as *.7 .8 .9* or *.5 to .9 by .1*.

P2 (Proportion Group 2)

Enter an estimate of the proportion for group 2. The sample size and width calculations assume that the value entered here is the proportion estimate that is obtained from the sample. If the sample proportion is different from the one specified here, the width may be narrower or wider than specified.

The value(s) must be between 0.0001 and 0.9999.

You can enter a range of values such as *.1 .2 .3* or *.1 to .5 by .1*.

Confidence Interval Method

Confidence Interval Formula

Specify the formula to be in used in calculation of confidence intervals.

- **Score (Farrington & Manning)**
  This formula is based on inverting Farrington and Manning's score test.
- **Score (Miettinen & Nurminen)**
  This formula is based on inverting Miettinen and Nurminen's score test.

- **Score w/ Skewness (Gart & Nam)**
  This formula is based on inverting Gart and Nam's score test, with a correction for skewness.

- **Logarithm (Katz)**
  This formula is based on the asymptotic normality of log(P1/P2).

- **Logarithm + 1/2 (Walter)**
  This formula is based on the asymptotic normality of log(P1/P2), but 1/2 is used as an adjustment.

- **Fleiss**
  This is an iterative method that was developed for the odds ratio and adapted to the proportion ratio.

Data Tab (Odds Ratios)

This section documents options that are used when the parameterization is in terms of the odds ratio and the value of the second sample proportion, P2. The corresponding procedure is Confidence Intervals for the Difference between Two Proportions using Odds Ratios.

Precision

Confidence Interval Width (Two-Sided)

This is the distance from the lower confidence limit to the upper confidence limit.

You can enter a single value or a list of values. The value(s) must be greater than zero.

Distance from OR to Limit (One-Sided)

This is the distance from the odds ratio to the lower or upper limit of the confidence interval, depending on whether the Interval Type is set to Lower Limit or Upper Limit.

You can enter a single value or a list of values. The value(s) must be greater than zero.

Proportions (OR = O1/O2)

Odds Ratio

Enter an estimate of the sample odds ratio (O1/O2). The sample size and width calculations assume that the value entered here is the odds ratio estimate that is obtained from the samples. If the sample odds ratio is different from the one specified here, the width may be narrower or wider than specified.

The value(s) must be greater than 0.

You can enter a range of values such as *.7 .8 .9* or *.5 to .9 by .1*.

P2 (Proportion Group 2)

Enter an estimate of the proportion for group 2. The sample size and width calculations assume that the value entered here is the proportion estimate that is obtained from the sample. If the sample proportion is different from the one specified here, the width may be narrower or wider than specified.

The value(s) must be between 0.0001 and 0.9999.

You can enter a range of values such as *.1 .2 .3* or *.1 to .5 by .1*.

Confidence Interval Method

Confidence Interval Formula

Specify the formula to be in used in calculation of confidence intervals.

- **Exact (Conditional)**
  This conditional exact confidence interval formula is calculated using the non-central hypergeometric distribution.
- **Score (Farrington & Manning)**
  This formula is based on inverting Farrington and Manning's score test.
- **Score (Miettinen & Nurminen)**
  This formula is based on inverting Miettinen and Nurminen's score test.
- **Fleiss**
  This iterative method forms the confidence interval as all those value of the odds ratio which would not be rejected by a chi-square hypothesis test.
- **Logarithm**
  This formula is similar to SIMPLE + 1/2, but with the logarithm of the odds ratio.
- **Mantel- Haenszel**
  This formula is based on the Mantel-Haenszel formula for the odds ratio.
- **Simple**
  This uses the simple odds ratio formula and large sample standard error estimate.
- **Simple + 1/2**
  This uses the simple odds ratio formula and large sample standard error estimate, but with 1/2 added to frequencies as a bias reduction device.

Iterations Tab

This tab sets an option used in the iterative procedures.

Maximum Iterations

Maximum Iterations Before Search Termination

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.

Example 1 – Calculating Sample Size using Proportions

Suppose a study is planned in which the researcher wishes to construct a two-sided 95% confidence interval for the difference in proportions such that the width of the interval is no wider than 0.1. The confidence interval method to be used is the Yates chi-square simple asymptotic method with continuity correction. The confidence level is set at 0.95, but 0.99 is included for comparative purposes. The proportion estimates to be used are 0.6 for Group 1, and 0.4 for Group 2. Instead of examining only the interval width of 0.1, a series of widths from 0.05 to 0.3 will also be considered.

The goal is to determine the necessary sample size.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Confidence Intervals for Two Proportions [Proportions]** procedure window by expanding **Proportions**, then **Two Independent Proportions**, then clicking on **Confidence Intervals**, and then clicking on **Confidence Intervals for Two Proportions [Proportions]**. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**N1**
Confidence Level	**0.95 0.99**
N1 (Sample Size Group 1)	*Ignored since this is the Find setting*
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
Confidence Interval Width (Two-Sided)	**0.05 to 0.30 by 0.05**
P1	**0.6**
P2	**0.4**
Confidence Interval Formula	**Chi-Square C.C. (Yates)**
Interval Type	**Two-Sided**

Annotated Output

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

Numeric Results

Numeric Results for Two-Sided Confidence Intervals for the Difference in Proportions
Confidence Interval Method: Chi-Square - Simple Asymptotic with Continuity Correction (Yates)

Confidence Level	N1	N2	Allocation Ratio	Target Width	Actual Width	P1	P2	P1 - P2	Lower Limit	Upper Limit
0.950	3030	3030	1.000	0.050	0.050	0.60	0.40	0.20	0.18	0.22
0.950	778	778	1.000	0.100	0.100	0.60	0.40	0.20	0.15	0.25
0.950	354	354	1.000	0.150	0.150	0.60	0.40	0.20	0.13	0.27
0.950	204	204	1.000	0.200	0.200	0.60	0.40	0.20	0.10	0.30
0.950	134	134	1.000	0.250	0.250	0.60	0.40	0.20	0.08	0.32
0.950	95	95	1.000	0.300	0.300	0.60	0.40	0.20	0.05	0.35

0.990	5176	5176	1.000	0.050	0.050	0.60	0.40	0.20	0.18	0.22
0.990	1314	1314	1.000	0.100	0.100	0.60	0.40	0.20	0.15	0.25
0.990	593	593	1.000	0.150	0.150	0.60	0.40	0.20	0.13	0.27
0.990	339	339	1.000	0.200	0.200	0.60	0.40	0.20	0.10	0.30
0.990	220	220	1.000	0.250	0.250	0.60	0.40	0.20	0.08	0.32
0.990	155	155	1.000	0.300	0.300	0.60	0.40	0.20	0.05	0.35

References

Newcombe, R. G. 1998. 'Interval Estimation for the Difference Between Independent Proportions: Comparison of Eleven Methods.' Statistics in Medicine, 17, pp. 873-890.

Fleiss, J. L., Levin, B., Paik, M.C. 2003. Statistical Methods for Rates and Proportions. Third Edition. John Wiley & Sons. New York.

Report Definitions

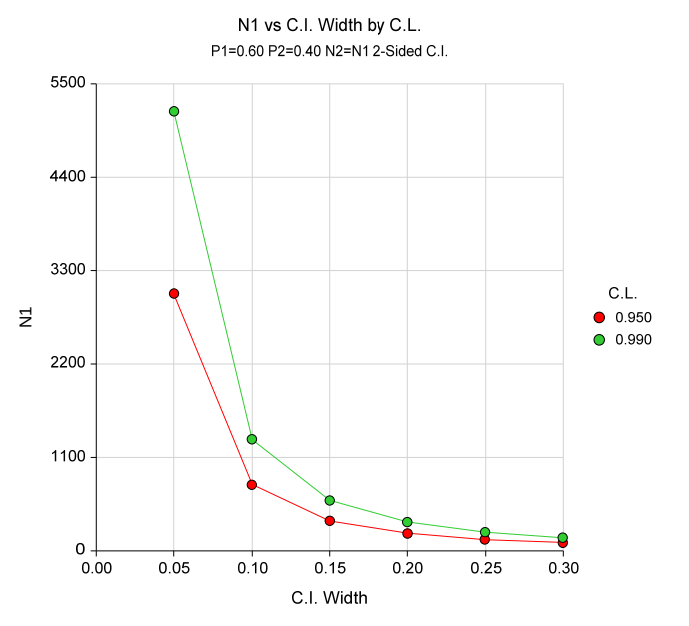
Confidence level is the proportion of confidence intervals (constructed with this same confidence level, sample size, etc.) that would contain the true difference in proportions.
N1 and N2 are the sample sizes drawn from the two populations.
Allocation Ratio is the ratio of the sample sizes, N2/N1.
Width is the distance from the lower limit to the upper limit.
Target Width is the value of the width that is entered into the procedure.
Actual Width is the value of the width that is obtained from the procedure.
P1 and P2 are the assumed sample proportions upon which the width calculations are based.
P1 - P2 is the difference in sample proportions.
Lower Limit and Upper Limit are the lower and upper limits of the confidence interval for the true difference in proportions (Population Proportion 1 - Population Proportion 2).

Summary Statements

Group sample sizes of 3030 and 3030 produce a two-sided 95% confidence interval for the difference in population proportions with a width that is equal to 0.050 when the estimated sample proportion 1 is 0.60 and the estimated sample proportion 2 is 0.40.

This report shows the calculated sample sizes for each of the scenarios.

Plots Section



This plot shows the group sample size versus the confidence interval width for the two confidence levels.

Example 2 – Validation (Proportions and Differences) using Newcombe

Newcombe (1998b) page 877 gives an example of a calculation for a confidence interval for the difference in proportions when the confidence level is 95%, the sample proportions are 0.9 and 0.3, and the interval width is 0.6790 for the Chi-Square (Pearson) method, 0.8395 for the Chi-Square C.C. (Yates) method, 0.67064 for the Score (Miettinen and Nurminen) method, 0.6385 for the Score (Wilson) method, and 0.7374 for the Score C.C. (Wilson) method. The necessary sample size in each case is 10 per group.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Confidence Intervals for Two Proportions [Proportions]** procedure window by expanding **Proportions**, then **Two Independent Proportions**, then clicking on **Confidence Intervals**, and then clicking on **Confidence Intervals for Two Proportions [Proportions]**. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**N1**
Confidence Level	**0.95**
N1 (Sample Size Group 1)	*Ignored since this is the Find setting*
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
Confidence Interval Width (Two-Sided)	**Varies (0.6790, 0.8395, 0.67064, 0.6385, 0.7374)**
P1	**0.9**
P2	**0.3**
Confidence Interval Formula	**Varies [Chi-Square (Pearson), Chi-Square C.C. (Yates), Score (Miettinen & Nurminen), Score (Wilson), Score C.C. (Wilson)]**
Interval Type	**Two-Sided**

Output

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

Chi-Square (Pearson)

Confidence Level	N1	N2	Allocation Ratio	Target Width	Actual Width	P1	P2	P1 - P2	Lower Limit	Upper Limit
0.950	10	10	1.000	0.6790	0.6790	0.9000	0.3000	0.6000	0.2605	0.9395

PASS also calculated the necessary sample size to be 10 per group.

Chi-Square C.C. (Yates)

Confidence Level	N1	N2	Allocation Ratio	Target Width	Actual Width	P1	P2	P1 - P2	Lower Limit	Upper Limit
0.950	10	10	1.000	0.8395	0.8395	0.9000	0.3000	0.6000	0.1605	1.0000

PASS also calculated the necessary sample size to be 10 per group.

Score (Miettinen & Nurminen)

Confidence Level	N1	N2	Allocation Ratio	Target Width	Actual Width	P1	P2	P1 - P2	Lower Limit	Upper Limit
0.950	10	10	1.000	0.6706	0.6706	0.9000	0.3000	0.6000	0.1700	0.8406

PASS also calculated the necessary sample size to be 10 per group.

Score (Wilson)

Confidence Level	N1	N2	Allocation Ratio	Target Width	Actual Width	P1	P2	P1 - P2	Lower Limit	Upper Limit
0.950	10	10	1.000	0.6385	0.6385	0.9000	0.3000	0.6000	0.1705	0.8090

PASS also calculated the necessary sample size to be 10 per group.

Score C.C. (Wilson)

Confidence Level	N1	N2	Allocation Ratio	Target Width	Actual Width	P1	P2	P1 - P2	Lower Limit	Upper Limit
0.950	10	10	1.000	0.7374	0.7374	0.9000	0.3000	0.6000	0.1013	0.8387

PASS also calculated the necessary sample size to be 10 per group.

Example 3 – Validation (Proportions and Differences) using Gart and Nam

Gart and Nam (1990) page 640 give an example of a calculation for a confidence interval for the difference in proportions when the confidence level is 95%, the sample proportions are 0.28 and 0.08, and the interval width is 0.4281 for the Score (Gart and Nam) method. The necessary sample size in each case is 25 per group.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Confidence Intervals for Two Proportions [Proportions]** procedure window by expanding **Proportions**, then **Two Independent Proportions**, then clicking on **Confidence Intervals**, and then clicking on **Confidence Intervals for Two Proportions [Proportions]**. You may then make the appropriate entries as listed below, or open **Example 3** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**N1**
Confidence Level	**0.95**
N1 (Sample Size Group 1)	*Ignored since this is the Find setting*
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
Confidence Interval Width (Two-Sided)	**0.4281**
P1	**0.28**
P2	**0.08**
Confidence Interval Formula	**Score w/Skewness (Gart & Nam)**
Interval Type	**Two-Sided**

Output

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

Numeric Results

Confidence Level	N1	N2	Allocation Ratio	Target Width	Actual Width	P1	P2	P1 - P2	Lower Limit	Upper Limit
0.950	25	25	1.000	0.4281	0.4281	0.2800	0.0800	0.2000	-0.0143	0.4137

PASS also calculated the necessary sample size to be 25 per group.

Example 4 – Calculating Sample Size using Differences

Suppose a study is planned in which the researcher wishes to construct a two-sided 95% confidence interval for the difference in proportions such that the width of the interval is no wider than 0.1. The confidence interval method to be used is the Yates chi-square simple asymptotic method with continuity correction. The confidence level is set at 0.95, but 0.99 is included for comparative purposes. The difference estimate to be used is 0.05, and the estimate for proportion 2 is 0.3. Instead of examining only the interval width of 0.1, a series of widths from 0.05 to 0.3 will also be considered.

The goal is to determine the necessary sample size.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Confidence Intervals for Two Proportions [Differences]** procedure window by expanding **Proportions**, then **Two Independent Proportions**, then clicking on **Confidence Intervals**, and then clicking on **Confidence Intervals for Two Proportions [Differences]**. You may then make the appropriate entries as listed below, or open **Example 4** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**N1**
Confidence Level	**0.95 0.99**
N1 (Sample Size Group 1)	*Ignored since this is the Find setting*
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
Confidence Interval Width (Two-Sided)	**0.05 to 0.30 by 0.05**
Difference in Sample Proportions	**0.05**
P2	**0.3**
Confidence Interval Formula	**Chi-Square C.C. (Yates)**
Interval Type	**Two-Sided**

Annotated Output

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

Numeric Results

Numeric Results for Two-Sided Confidence Intervals for the Difference in Proportions
Confidence Interval Method: Chi-Square - Simple Asymptotic with Continuity Correction (Yates)

Confidence Level	N1	N2	Allocation Ratio	Target Width	Actual Width	P1	P2	P1 - P2	Lower Limit	Upper Limit
0.950	2769	2769	1.000	0.050	0.050	0.35	0.30	0.05	0.03	0.07
0.950	712	712	1.000	0.100	0.100	0.35	0.30	0.05	0.00	0.10
0.950	325	325	1.000	0.150	0.150	0.35	0.30	0.05	-0.02	0.12
0.950	188	188	1.000	0.200	0.200	0.35	0.30	0.05	-0.05	0.15
0.950	124	124	1.000	0.250	0.249	0.35	0.30	0.05	-0.07	0.17
0.950	88	88	1.000	0.300	0.299	0.35	0.30	0.05	-0.10	0.20

0.990	4725	4725	1.000	0.050	0.050	0.35	0.30	0.05	0.03	0.07
0.990	1201	1201	1.000	0.100	0.100	0.35	0.30	0.05	0.00	0.10
0.990	543	543	1.000	0.150	0.150	0.35	0.30	0.05	-0.02	0.12
0.990	310	310	1.000	0.200	0.200	0.35	0.30	0.05	-0.05	0.15
0.990	202	202	1.000	0.250	0.250	0.35	0.30	0.05	-0.07	0.17
0.990	143	143	1.000	0.300	0.299	0.35	0.30	0.05	-0.10	0.20

References

Newcombe, R. G. 1998. 'Interval Estimation for the Difference Between Independent Proportions: Comparison of Eleven Methods.' Statistics in Medicine, 17, pp. 873-890.

Fleiss, J. L., Levin, B., Paik, M.C. 2003. Statistical Methods for Rates and Proportions. Third Edition. John Wiley & Sons. New York.

Report Definitions

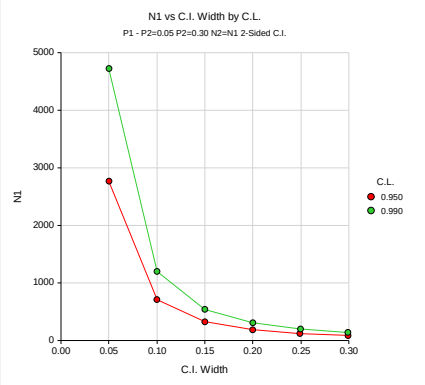
Confidence level is the proportion of confidence intervals (constructed with this same confidence level, sample size, etc.) that would contain the true difference in proportions.
N1 and N2 are the sample sizes drawn from the two populations.
Allocation Ratio is the ratio of the sample sizes, N2/N1.
Width is the distance from the lower limit to the upper limit.
Target Width is the value of the width that is entered into the procedure.
Actual Width is the value of the width that is obtained from the procedure.
P1 and P2 are the assumed sample proportions upon which the width calculations are based.
P1 - P2 is the difference in sample proportions.
Lower Limit and Upper Limit are the lower and upper limits of the confidence interval for the true difference in proportions (Population Proportion 1 - Population Proportion 2).

Summary Statements

Group sample sizes of 2769 and 2769 produce a two-sided 95% confidence interval for the difference in population proportions with a width that is equal to 0.050 when the estimated sample proportion 1 is 0.35, the estimated sample proportion 2 is 0.30, and the difference in sample proportions is 0.05.

This report shows the calculated sample sizes for each of the scenarios.

Plots Section



This plot shows the group sample size versus the confidence interval width for the two confidence levels.

Validation using Differences

The validation for the procedure Confidence Intervals for the Difference between Two Proportions using Differences is shown in Examples 2 and 3, which is the validation for the proportion specification.

Example 5 – Calculating Sample Size using Ratios

Suppose a study is planned in which the researcher wishes to construct a two-sided 95% confidence interval for the ratio of proportions such that the width of the interval is no wider than 0.2. The confidence interval method to be used is the Logarithm (Katz) method. The confidence level is set at 0.95, but 0.99 is included for comparative purposes. The ratio estimate to be used is 1.2, and the estimate for proportion 2 is 0.6. Instead of examining only the interval width of 0.2, a series of widths from 0.1 to 0.3 will also be considered.

The goal is to determine the necessary sample size.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Confidence Intervals for Two Proportions [Ratios]** procedure window by expanding **Proportions**, then **Two Independent Proportions**, then clicking on **Confidence Intervals**, and then clicking on **Confidence Intervals for Two Proportions [Ratios]**. You may then make the appropriate entries as listed below, or open **Example 5** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**N1**
Confidence Level	**0.95 0.99**
N1 (Sample Size Group 1)	*Ignored since this is the Find setting*
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
Confidence Interval Width (Two-Sided)	**0.10 to 0.30 by 0.05**
Ratio of Sample Proportions	**1.2**
P2	**0.6**
Confidence Interval Formula	**Logarithm (Katz)**
Interval Type	**Two-Sided**

Annotated Output

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

Numeric Results

Numeric Results for Two-Sided Confidence Intervals for the Ratio of Proportions
Confidence Interval Method: Logarithm (Katz)

Confidence Level	N1	N2	Allocation Ratio	Target Width	Actual Width	P1	P2	P1/P2	Lower Limit	Upper Limit
0.950	2337	2337	1.000	0.100	0.100	0.72	0.60	1.20	1.15	1.25
0.950	1040	1040	1.000	0.150	0.150	0.72	0.60	1.20	1.13	1.28
0.950	586	586	1.000	0.200	0.200	0.72	0.60	1.20	1.10	1.30
0.950	376	376	1.000	0.250	0.250	0.72	0.60	1.20	1.08	1.33
0.950	261	261	1.000	0.300	0.300	0.72	0.60	1.20	1.06	1.36

0.990	4037	4037	1.000	0.100	0.100	0.72	0.60	1.20	1.15	1.25
0.990	1796	1796	1.000	0.150	0.150	0.72	0.60	1.20	1.13	1.28
0.990	1011	1011	1.000	0.200	0.200	0.72	0.60	1.20	1.10	1.30
0.990	648	648	1.000	0.250	0.250	0.72	0.60	1.20	1.08	1.33
0.990	451	451	1.000	0.300	0.300	0.72	0.60	1.20	1.06	1.36

References

Gart, John J. and Nam, Jun-mo. 1988. 'Approximate Interval Estimation of the Ratio of Binomial Parameters: A Review and Corrections for Skewness.' Biometrics, Volume 44, 323-338.
Koopman, P. A. R. 1984. 'Confidence Intervals for the Ratio of Two Binomial Proportions.' Biometrics, Volume 40, Issue 2, 513-517.
Katz, D., Baptista, J., Azen, S. P., and Pike, M. C. 1978. 'Obtaining Confidence Intervals for the Risk Ratio in Cohort Studies.' Biometrics, Volume 34, 469-474.

Report Definitions

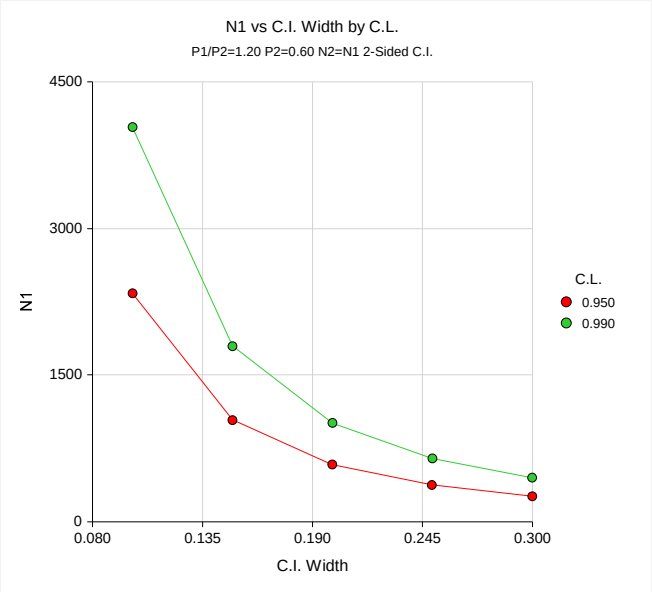
Confidence level is the proportion of confidence intervals (constructed with this same confidence level, sample size, etc.) that would contain the true ratio of proportions.
N1 and N2 are the sample sizes drawn from the two populations.
Allocation Ratio is the ratio of the sample sizes, N2/N1.
Width is the distance from the lower limit to the upper limit.
Target Width is the value of the width that is entered into the procedure.
Actual Width is the value of the width that is obtained from the procedure.
P1 and P2 are the assumed sample proportions upon which the width calculations are based.
P1/P2 is the ratio of sample proportions.
Lower Limit and Upper Limit are the lower and upper limits of the confidence interval for the true ratio of proportions (Population Proportion 1 / Population Proportion 2).

Summary Statements

Group sample sizes of 2337 and 2337 produce a two-sided 95% confidence interval for the ratio of population proportions with a width that is equal to 0.100 when the estimated sample proportion 1 is 0.72, the estimated sample proportion 2 is 0.60, and the ratio of the sample proportions is 1.20.

This report shows the calculated sample sizes for each of the scenarios.

Plots Section



This plot shows the group sample size versus the confidence interval width for the two confidence levels.

Example 6 – Validation (Ratios) using Gart and Nam

Gart and Nam (1988) page 331 give an example (Example 2) of a calculation for a confidence interval for the ratio of proportions when the confidence level is 95%, the sample proportion ratio is 2 and the sample proportion 2 is 0.3, the sample size for group 2 is 20, and the interval width is 3.437 for the Logarithm + 1/2 (Walter) method, 3.751 for the Score (Farrington and Manning) method, and 4.133 for the Score w/Skewness (Gart and Nam) method. The necessary sample size for group 1 in each case is 10.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Confidence Intervals for Two Proportions [Ratios]** procedure window by expanding **Proportions**, then **Two Independent Proportions**, then clicking on **Confidence Intervals**, and then clicking on **Confidence Intervals for Two Proportions [Ratios]**. You may then make the appropriate entries as listed below, or open **Example 6** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**N1**
Confidence Level	**0.95**
N1 (Sample Size Group 1)	*Ignored since this is the Find setting*
N2 (Sample Size Group 2)	**20**
R (Sample Allocation Ratio)	*Ignored*
Confidence Interval Width (Two-Sided)	**Varies (3.437, 3.751, 4.133)**
Ratio of Sample Proportions	**2**
P2	**0.3**
Confidence Interval Formula	**Varies [Logarithm + 1/2 (Walter), Score (Farrington and Manning), Score w/Skewness (Gart and Nam)]**
Interval Type	**Two-Sided**

Output

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

Logarithm + 1/2 (Walter)

Confidence Level	N1	N2	Allocation Ratio	Target Width	Actual Width	P1	P2	P1/P2	Lower Limit	Upper Limit
0.950	10	20	2.000	3.437	3.431	0.60	0.30	2.00	0.88	4.31

PASS also calculated the necessary sample size for group 1 to be 10.

Score (Farrington and Manning)

Confidence Level	N1	N2	Allocation Ratio	Target Width	Actual Width	P1	P2	P1/P2	Lower Limit	Upper Limit
0.950	10	20	2.000	3.751	3.751	0.60	0.30	2.00	0.84	4.59

PASS also calculated the necessary sample size for group 1 to be 10.

Score w/Skewness (Gart and Nam)

Confidence Level	N1	N2	Allocation Ratio	Target Width	Actual Width	P1	P2	P1/P2	Lower Limit	Upper Limit
0.950	10	20	2.000	4.133	4.132	0.60	0.30	2.00	0.82	4.95

PASS also calculated the necessary sample size for group 1 to be 10.

Example 7 – Validation (Ratios) using Katz et al

Katz et al (1978) pages 472-473 give an example of a calculation for a lower limit confidence interval for the ratio of proportions when the confidence level is 97.5%, the sample proportion ratio is 1.596078 and the sample proportion 2 is 0.153153, the sample size for group 2 is 111, and the distance from the ratio to the limit is 0.6223 for the Logarithm (Katz) method. The necessary sample size for group 1 is 225.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Confidence Intervals for Two Proportions [Ratios]** procedure window by expanding **Proportions**, then **Two Independent Proportions**, then clicking on **Confidence Intervals**, and then clicking on **Confidence Intervals for Two Proportions [Ratios]**. You may then make the appropriate entries as listed below, or open **Example 7** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**N1**
Confidence Level	**0.975**
N1 (Sample Size Group 1)	*Ignored since this is the Find setting*
N2 (Sample Size Group 2)	**111**
R (Sample Allocation Ratio)	*Ignored*
Interval Type	**Lower Limit**
Distance to from Ratio to Limit	**0.6223**
Ratio of Sample Proportions	**1.596078**
P2	**0.153153**
Confidence Interval Formula	**Logarithm (Katz)**

Output

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

Logarithm (Katz)

Confidence Level	N1	N2	Allocation Ratio	Target Width	Actual Width	P1	P2	P1/P2	Lower Limit	Upper Limit
0.975	225	111	0.493	0.622	0.622	0.24	0.15	1.60	0.97	Inf

PASS also calculated the necessary sample size for group 1 to be 225.

Example 8 – Calculating Sample Size using Odds Ratios

Suppose a study is planned in which the researcher wishes to construct a two-sided 95% confidence interval for the odds ratio such that the width of the interval is no wider than 0.5. The confidence interval method to be used is the Logarithm method. The confidence level is set at 0.95, but 0.99 is included for comparative purposes. The odds ratio estimate to be used is 1.5, and the estimate for proportion 2 is 0.4. Instead of examining only the interval width of 0.5, a series of widths from 0.1 to 1.0 will also be considered.

The goal is to determine the necessary sample size.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Confidence Intervals for Two Proportions [Odds Ratios]** procedure window by expanding **Proportions**, then **Two Independent Proportions**, then clicking on **Confidence Intervals**, and then clicking on **Confidence Intervals for Two Proportions [Odds Ratios]**. You may then make the appropriate entries as listed below, or open **Example 8** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**N1**
Confidence Level	**0.95 0.99**
N1 (Sample Size Group 1)	*Ignored since this is the Find setting*
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
Confidence Interval Width (Two-Sided)	**0.1 to 1.0 by 0.1**
Odds Ratio	**1.5**
P2	**0.4**
Confidence Interval Formula	**Logarithm**
Interval Type	**Two-Sided**

Annotated Output

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

Numeric Results

Numeric Results for Two-Sided Confidence Intervals for the Odds Ratio
Confidence Interval Method: Logarithm

Confidence Level	N1	N2	Allocation Ratio	Target Width	Actual Width	P1	P2	Odds Ratio O1/O2	Lower Limit	Upper Limit
0.950	28244	28244	1.000	0.100	0.100	0.50	0.40	1.50	1.45	1.55
0.950	7068	7068	1.000	0.200	0.200	0.50	0.40	1.50	1.40	1.60
0.950	3146	3146	1.000	0.300	0.300	0.50	0.40	1.50	1.36	1.66
0.950	1774	1774	1.000	0.400	0.400	0.50	0.40	1.50	1.31	1.71
0.950	1138	1138	1.000	0.500	0.500	0.50	0.40	1.50	1.27	1.77
0.950	793	793	1.000	0.600	0.600	0.50	0.40	1.50	1.23	1.83
0.950	585	585	1.000	0.700	0.700	0.50	0.40	1.50	1.19	1.89
0.950	450	450	1.000	0.800	0.800	0.50	0.40	1.50	1.15	1.95
0.950	358	358	1.000	0.900	0.899	0.50	0.40	1.50	1.11	2.01
0.950	291	291	1.000	1.000	1.000	0.50	0.40	1.50	1.08	2.08
0.990	48783	48783	1.000	0.100	0.100	0.50	0.40	1.50	1.45	1.55
0.990	12208	12208	1.000	0.200	0.200	0.50	0.40	1.50	1.40	1.60
0.990	5435	5435	1.000	0.300	0.300	0.50	0.40	1.50	1.36	1.66
0.990	3065	3065	1.000	0.400	0.400	0.50	0.40	1.50	1.31	1.71
0.990	1967	1967	1.000	0.500	0.500	0.50	0.40	1.50	1.27	1.77
0.990	1371	1371	1.000	0.600	0.600	0.50	0.40	1.50	1.23	1.83
0.990	1012	1012	1.000	0.700	0.700	0.50	0.40	1.50	1.19	1.89
0.990	778	778	1.000	0.800	0.800	0.50	0.40	1.50	1.15	1.95
0.990	618	618	1.000	0.900	0.900	0.50	0.40	1.50	1.12	2.02
0.990	504	504	1.000	1.000	0.999	0.50	0.40	1.50	1.08	2.08

References

Fleiss, J. L., Levin, B., Paik, M.C. 2003. Statistical Methods for Rates and Proportions. Third Edition. John Wiley & Sons. New York.

Report Definitions

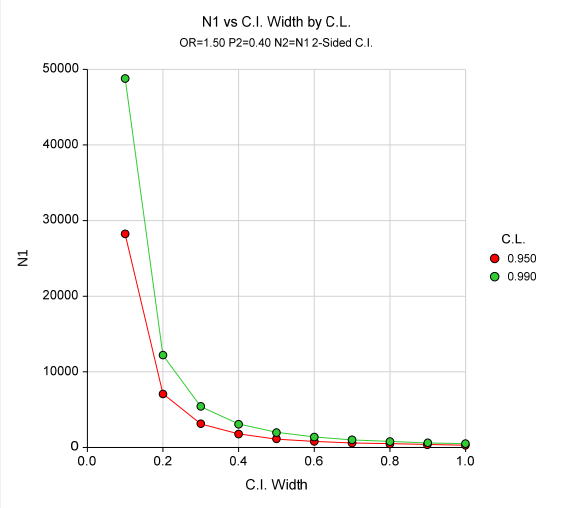
Confidence level is the proportion of confidence intervals (constructed with this same confidence level, sample size, etc.) that would contain the true odds ratio.
N1 and N2 are the sample sizes drawn from the two populations.
Allocation Ratio is the ratio of the sample sizes, N2/N1.
Width is the distance from the lower limit to the upper limit.
Target Width is the value of the width that is entered into the procedure.
Actual Width is the value of the width that is obtained from the procedure.
P1 and P2 are the assumed sample proportions upon which the width calculations are based.
Odds Ratio O1/O2 is the sample odds ratio.
Lower Limit and Upper Limit are the lower and upper limits of the confidence interval for the true odds ratio (Population Odds 1 / Population Odds 2).

Summary Statements

Group sample sizes of 28244 and 28244 produce a two-sided 95% confidence interval for the population odds ratio with a width that is equal to 0.100 when the estimated sample proportion 1 is 0.50, the estimated sample proportion 2 is 0.40, and the sample odds ratio is 1.50.

This report shows the calculated sample sizes for each of the scenarios.

Plots Section



This plot shows the group sample size versus the confidence interval width for the two confidence levels.

Example 9 – Validation (Odds Ratios) using Fleiss et al

Fleiss et al (2003) pages 117, 119 give an example of a calculation for a confidence interval for the odds ratio when the confidence level is 95%, the sample odds ratio is 2.25 and the sample proportion 2 is 0.1, the sample size for group 2 is 150, and the interval width is 4.387 for the Logarithm method, and 4.980 for the Fleiss method. The necessary sample size for group 1 in each case is 50.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Confidence Intervals for Two Proportions [Odds Ratios]** procedure window by expanding **Proportions**, then **Two Independent Proportions**, then clicking on **Confidence Intervals**, and then clicking on **Confidence Intervals for Two Proportions [Odds Ratios]**. You may then make the appropriate entries as listed below, or open **Example 9** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**N1**
Confidence Level	**0.95**
N1 (Sample Size Group 1)	*Ignored since this is the Find setting*
N2 (Sample Size Group 2)	**150**
R (Sample Allocation Ratio)	*Ignored*
Confidence Interval Width (Two-Sided)	**Varies (4.387, 4.980)**
Odds Ratio	**2.25**
P2	**0.1**

Data Tab (continued)

Confidence Interval Formula ..... **Varies (Logarithm, Fleiss)**
Interval Type ..... **Two-Sided**

Output

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

Logarithm

Confidence Level	N1	N2	Allocation Ratio	Target Width	Actual Width	P1	P2	Odds Ratio O1/O2	Lower Limit	Upper Limit
0.950	50	150	3.000	4.387	4.387	0.20	0.10	2.25	0.96	5.35

PASS also calculated the necessary sample size for group 1 to be 50.

Fleiss

Confidence Level	N1	N2	Allocation Ratio	Target Width	Actual Width	P1	P2	Odds Ratio O1/O2	Lower Limit	Upper Limit
0.950	50	150	3.000	4.980	4.980	0.20	0.10	2.25	0.86	5.84

PASS also calculated the necessary sample size for group 1 to be 50.

Chapter 220

Group-Sequential Tests for Two Proportions

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 group proportions. Other modules perform similar analyses for the comparison of means and survival functions. The program allows you to vary the number and times of interim tests, the type of alpha spending function, and the 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

Suppose the means of two samples of *N1* and *N2* individuals will be compared at various stages of a trial using the z_k statistic:

$$z_k = \frac{\hat{p}_{1k} - \hat{p}_{2k}}{\sqrt{\hat{p}_{1k}(1-\hat{p}_{1k}) + \hat{p}_{2k}(1-\hat{p}_{2k})}}$$

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*). This formula computes the standard z-test that is assumed to be normally distributed.

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 being 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:

$$\alpha(0) = 0$$
$$\alpha(1) = \alpha$$

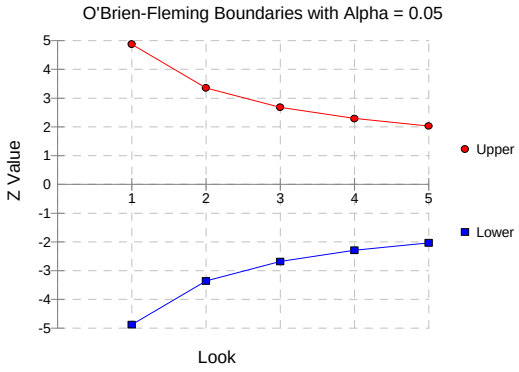
The last characteristic guarantees a fixed α level when the trial is complete. That is,

$$\Pr\left(|z_1| \geq b_1 \text{ or } |z_2| \geq b_2 \text{ or } \cdots \text{ or } |z_k| \geq b_k\right) = \alpha(\tau)$$

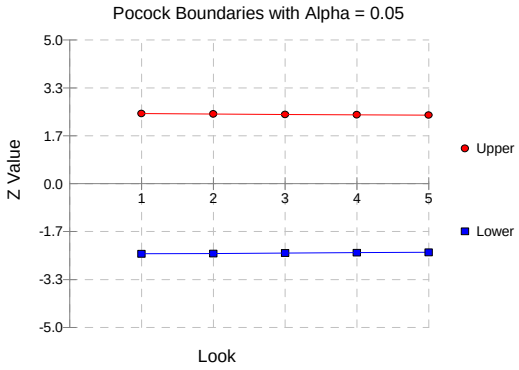
This methodology is very flexible since neither the times nor the number of analyses must be specified in advance. Only the functional form of $\alpha(\tau)$ must be specified.

PASS provides five popular spending functions plus the ability to enter and analyze your own boundaries. These are calculated as follows:

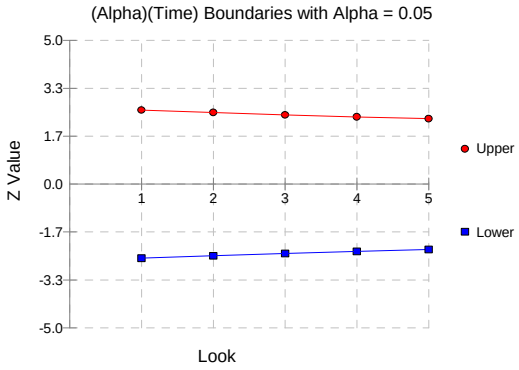
1. O'Brien-Fleming $2-2\Phi\left(\frac{Z_{\alpha/2}}{\sqrt{\tau}}\right)$



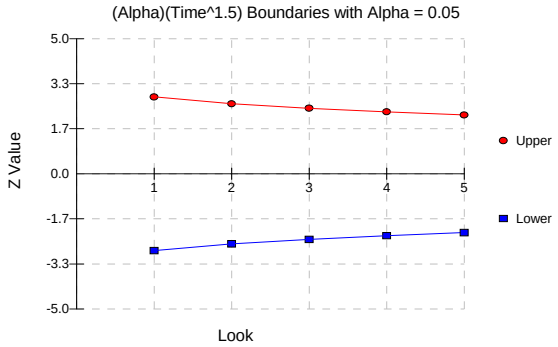
2. Pocock $\alpha\ln\left(1+\left(e-1\right)\tau\right)$



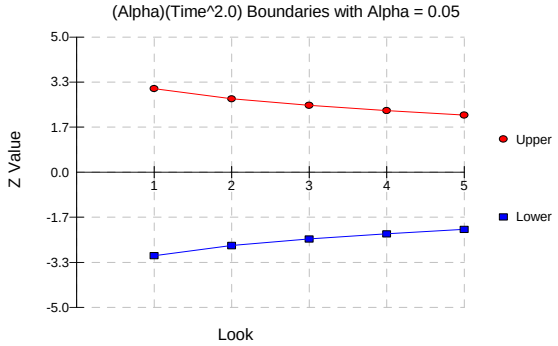
3. Alpha * time $\alpha\tau$



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 proportions 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.

Theory

A detailed account of the methodology is contained in Lan & DeMets (1983), DeMets & Lan (1984), Lan & Zucker (1993), and DeMets & Lan (1994). 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 parameter is related to the parameters of the z-test through the equation

$$\theta = \frac{p_1 - p_2}{\sqrt{\frac{\bar{p}(1-\bar{p})}{N1} + \frac{\bar{p}(1-\bar{p})}{N2}}}$$

where

$$\bar{p} = \frac{N1 p_1 + N2 p_2}{N1 + N2}$$

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 sample size.

Procedure Tabs

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

Data Tab

The Data tab contains the parameters associated with the z-test such as the proportions, sample sizes, alpha, and beta.

Solve For

Find (Solve For)

This option specifies the parameter to be solved for from the other parameters. The parameters that may be selected are *P1*, *P2*, *Alpha*, *Power and Beta*, *N1* or *N2*. Under most situations, you will select either *Power and Beta* or *N1*.

Select *N1* when you want to calculate the sample size needed to achieve a given power and alpha level.

Select *Power and Beta* when you want to calculate the power of an experiment.

Error Rates

Power or Beta

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

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

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

Alpha

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

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

Sample Size

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*.

Effect Size

P1 (Proportion in Group 1)

Enter value(s) for the response proportion in the first group under both hypotheses and the response proportion of the second group under the null hypothesis of equal proportions. The values must be 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*.

P2 (Proportion in Group 2)

Enter value(s) for the response proportion of the second group under the alternative hypothesis. You may enter a range of values such as *0.1, 0.2, 0.3* or *0.1 to 0.9 by 0.2.*

Look Details

This box 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.

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.

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.

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.

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.

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 (Spending = User)

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.

Test

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.

Continuity Correction

Specify whether to use the Continuity Correction. This option applies an adjustment to the sample sizes that is recommend by Fleiss(1981) page 45 to make the alpha and beta values more accurate. The formula for the adjustment is

$$N1_{new} = \frac{N1_{old}}{4}\left(1+\sqrt{1+\frac{2(R+1)}{R(N1_{old})|P1-P2|}}\right)^2$$

Options Tab

The Options tab controls the convergence of the various iterative algorithms used in the calculations.

Maximum Iterations

Maximum Iterations Before Search Termination

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).

Maximum Iterations (Lan-Demets algorithm)

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.

Tolerance

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 can try increasing this value.

Example 1 – Finding the Sample Size

A clinical trial is to be conducted over a two-year period to compare the proportion response of a new treatment to that of the current treatment. The current response proportion is 0.53. Although the researchers do not know the true proportion of patients that will survive with the new treatment, they would like to examine the power that is achieved if the proportion under the new treatment is 0.63. In order to compare the sample size requirements for different effect sizes, it is also of interest to compute the sample size at response rates of 0.60, 0.65, 0.70, and 0.75.

Testing will be done at the 0.05 significance level and the power should be set to 0.90. 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 per arm and two-sided hypothesis tests.

Setup

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

Option	Value
Data Tab	
Find (Solve For)	**N1**
Power	**0.90**
Alpha	**0.05**
N1 (Sample Size Group 1)	*Ignored since this is the Find setting*
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
P1 (Proportion in Group 1)	**0.53**
P2 (Proportion in Group 2)	**0.60 0.63 0.65 0.70 0.75**
Number of Looks	**4**
Spending Function	**O'Brien-Fleming**
Max Time	**2**
Times	**Equally Spaced**
Alternative Hypothesis	**Two-Sided**
Continuity Correction	**Checked**

Annotated Output

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

Numeric Results

Numeric Results for Two-Sided Test of Proportions. Continuity Correction Applied.

Power	N1	N2	Alpha	Beta	P1	P2
0.900168	1102	1102	0.050000	0.099832	0.53	0.60
0.900930	542	542	0.050000	0.099070	0.53	0.63
0.900408	376	376	0.050000	0.099592	0.53	0.65
0.901093	187	187	0.050000	0.098907	0.53	0.70
0.903126	111	111	0.050000	0.096874	0.53	0.75

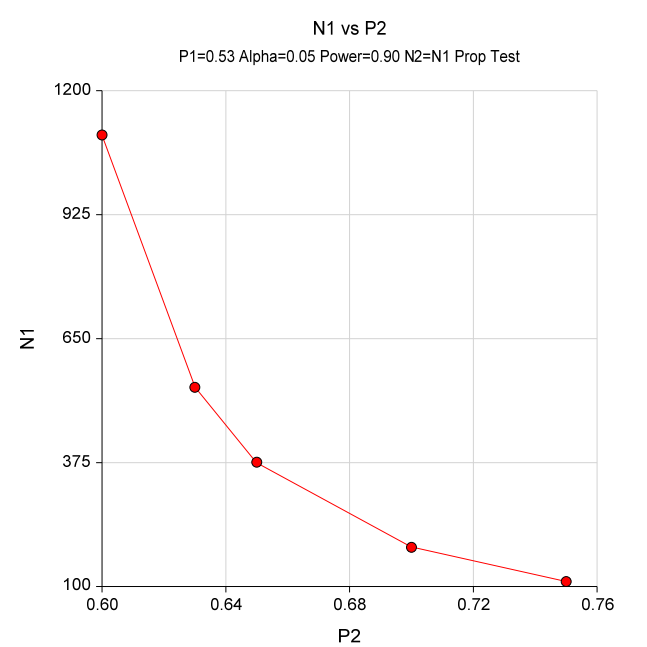
Report Definitions
Power is the probability of rejecting a false null hypothesis. Power should be close to one.
N1 and N2 are the number of items sampled from groups 1 and 2.
Alpha is the probability of rejecting a true null hypothesis in at least one of the sequential tests.
Beta is the probability of accepting a false null hypothesis at the conclusion of all tests.
P1 is the value of both proportions under the null hypothesis.
P2 is the proportion in group two under the alternative hypothesis.

Summary Statements
Sample sizes of 1102 and 1102 achieve 90% power to detect a difference of 0.07 between the group proportions of 0.53 and 0.60 at a significance level (alpha) of 0.0500 using a two-sided z-test with continuity correction. These results assume that 4 sequential tests are made using the O'Brien-Fleming spending function to determine the test boundaries.

This report shows the values of each of the parameters, one scenario per row. Note that 542 participants in each arm of the study are required to meet the 90% power requirement when the proportion is 0.63.

The values from this table are in the chart below. Note that this plot actually occurs further down in the report.

Plots Section



This plot shows that a large increase in sample size is necessary when the detectable proportion in group two is less than 0.63.

Details Section

Details when Spending = O'Brien-Fleming, N1 = 542, N2 =542, P1 = 0.53, P2 = 0.63, Continuity Correction.

Look	Time	Lower Bndry	Upper Bndry	Nominal Alpha	Inc Alpha	Total Alpha	Inc Power	Total Power
1	0.50	-4.33263	4.33263	0.000015	0.000015	0.000015	0.003525	0.003525
2	1.00	-2.96311	2.96311	0.003045	0.003036	0.003051	0.255573	0.259098
3	1.50	-2.35902	2.35902	0.018323	0.016248	0.019299	0.427801	0.686899
4	2.00	-2.01406	2.01406	0.044003	0.030701	0.050000	0.214031	0.900930
Drift	3.27640							

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.

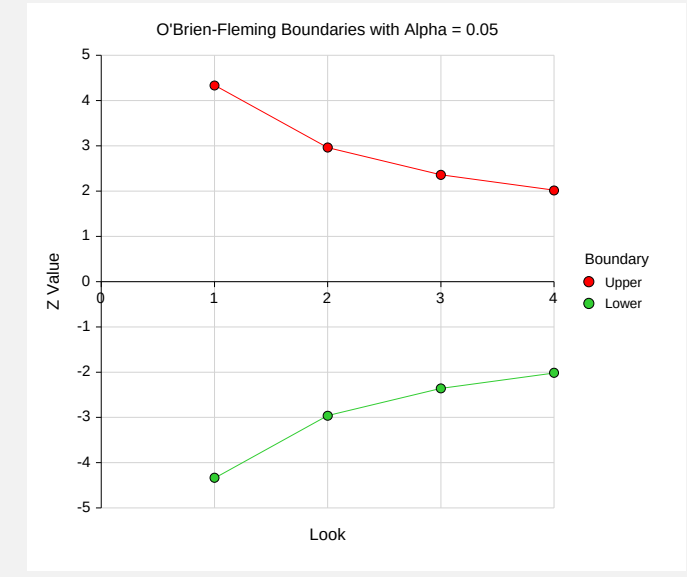
Total Power

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.

Example 2 – Finding the Power

Continuing the scenario began in Example1, the researcher wishes to calculate the power of the design at sample sizes 200, 400, 600, 800, 1000. 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

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

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power	*Ignored since this is the Find setting*
Alpha	**0.01 0.05 0.10**
N1 (Sample Size Group 1)	**200 to 1000 by 200**
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
P1 (Proportion in Group 1)	**0.53**
P2 (Proportion in Group 2)	**0.63**
Number of Looks	**4**
Spending Function	**O'Brien-Fleming**
Max Time	**2**
Times	**Equally Spaced**
Alternative Hypothesis	**Two-Sided**
Continuity Correction	**Checked**

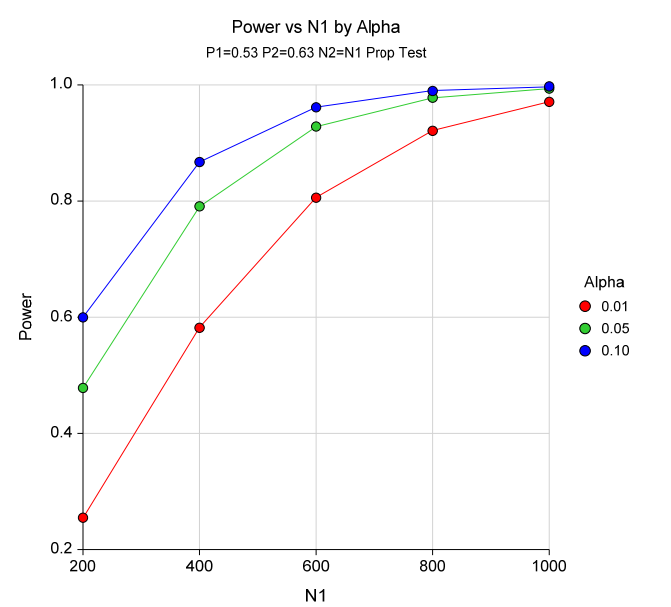
Output

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

Numeric Results and Plots

Numeric Results for Two-Sided Test of Proportions. Continuity Correction Applied.

Power	N1	N2	Alpha	Beta	P1	P2
0.254797	200	200	0.010000	0.745203	0.53	0.63
0.581920	400	400	0.010000	0.418080	0.53	0.63
0.805679	600	600	0.010000	0.194321	0.53	0.63
0.920951	800	800	0.010000	0.079049	0.53	0.63
0.970898	1000	1000	0.010000	0.029102	0.53	0.63
0.478378	200	200	0.050000	0.521622	0.53	0.63
0.790815	400	400	0.050000	0.209185	0.53	0.63
0.928267	600	600	0.050000	0.071733	0.53	0.63
0.977859	800	800	0.050000	0.022141	0.53	0.63
0.993673	1000	1000	0.050000	0.006327	0.53	0.63
0.599827	200	200	0.100000	0.400173	0.53	0.63
0.867330	400	400	0.100000	0.132670	0.53	0.63
0.961331	600	600	0.100000	0.038669	0.53	0.63
0.989659	800	800	0.100000	0.010341	0.53	0.63
0.997396	1000	1000	0.100000	0.002604	0.53	0.63



These data show the power for various sample sizes and alphas. It is interesting to note that once the sample size is greater than 700, the value of alpha makes little difference on the value of power.

Example 3 – 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, N1 to 500, 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

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

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power	*Ignored since this is the Find setting*
Alpha	**0.05**
N1 (Sample Size Group 1)	**500**
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
P1 (Proportion in Group 1)	**0.53**
P2 (Proportion in Group 2)	**0.63**
Number of Looks	**1 (Also run with 2, 3, 4, 6, 8, 10, and 20)**
Spending Function	**O'Brien-Fleming**
Max Time	**2**
Times	**Equally Spaced**
Alternative Hypothesis	**Two-Sided**
Continuity Correction	**Not checked**

Output

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

Numeric Results

Numeric Results for Two-Sided Test of Proportions

Power	N1	N2	Alpha	Beta	P1	P2	Looks
0.893174	500	500	0.050000	0.106826	0.53	0.63	1
0.892118	500	500	0.050000	0.107882	0.53	0.63	2
0.889620	500	500	0.050000	0.110380	0.53	0.63	3
0.887691	500	500	0.050000	0.112309	0.53	0.63	4
0.885125	500	500	0.050000	0.114875	0.53	0.63	6
0.883535	500	500	0.050000	0.116465	0.53	0.63	8
0.882456	500	500	0.050001	0.117544	0.53	0.63	10
0.879929	500	500	0.050001	0.120071	0.53	0.63	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 0.893174. When twenty interim looks are made, the power falls to 0.879929—a very small change.

Example 4 – 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

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

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power	*Ignored since this is the Find setting*
Alpha	**0.05 (will be calculated from boundaries)**
N1 (Sample Size Group 1)	**500**
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
P1 (Proportion in Group 1)	**0.53**
P2 (Proportion in Group 2)	**0.63**
Number of Looks	**5**
Spending Function	**User Supplied**
Max Time	**2**
Times	**Equally Spaced**
Upper Boundaries	**3.5 3.5 3.0 2.5 2.0**
Lower Boundaries	**Symmetric**
Alternative Hypothesis	**Two-Sided**
Continuity Correction	**Not checked**

Output

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

Numeric Results

Numeric Results for Two-Sided Test of Proportions

Power	N1	N2	Alpha	Beta	P1	P2
0.887792	500	500	0.048157	0.112208	0.53	0.63

Details when Spending = User Supplied, N1 = 500, N2 =500, P1 = 0.53, P2 = 0.63

Look	Time	Lower Bndry	Upper Bndry	Nominal Alpha	Inc Alpha	Total Alpha	Inc Power	Total Power
1	0.40	-3.50000	3.50000	0.000465	0.000465	0.000465	0.019352	0.019352
2	0.80	-3.50000	3.50000	0.000465	0.000408	0.000874	0.058108	0.077460
3	1.20	-3.00000	3.00000	0.002700	0.002410	0.003284	0.230567	0.308026
4	1.60	-2.50000	2.50000	0.012419	0.010331	0.013615	0.339341	0.647367
5	2.00	-2.00000	2.00000	0.045500	0.034542	0.048157	0.240425	0.887792
Drift	3.20355							

The power for this design is about 0.89. 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.

Example 5 – 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 = 5, alpha = 0.05, beta = 0.10, P1 = 0.1100, P2 = 0.0825. They compute a drift of 3.28 and a sample size of 2381.78 per group. 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

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

Option	Value
Data Tab	
Find (Solve For)	**N1**
Power	**0.90**
Alpha	**0.05**
N1 (Sample Size Group 1)	*Ignored since this is the Find setting*
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
P1 (Proportion in Group 1)	**0.1100**
P2 (Proportion in Group 2)	**0.0825**
Number of Looks	**5**
Spending Function	**O'Brien-Fleming**
Max Time	**1**
Times	**Equally Spaced**
Alternative Hypothesis	**Two-Sided**
Continuity Correction	**Not checked**

Output

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

Numeric Results

Numeric Results for Two-Sided Test of Proportions

Power	N1	N2	Alpha	Beta	P1	P2
0.900105	2474	2474	0.050000	0.099895	0.1100	0.0825

Details when Spending = O'Brien-Fleming, N1 = 2468, N2 =2468, P1 = 0.1100, P2 = 0.0825

Look	Time	Lower Bndry	Upper Bndry	Nominal Alpha	Inc Alpha	Total Alpha	Inc Power	Total Power
1	0.20	-4.87688	4.87688	0.000001	0.000001	0.000001	0.000324	0.000324
2	0.40	-3.35695	3.35695	0.000788	0.000787	0.000788	0.099454	0.099778
3	0.60	-2.68026	2.68026	0.007357	0.006828	0.007616	0.346699	0.446477
4	0.80	-2.28979	2.28979	0.022034	0.016807	0.024424	0.299644	0.746120
5	1.00	-2.03100	2.03100	0.042255	0.025576	0.050000	0.153985	0.900105
Drift	3.27939							

The difference in the sample sizes (2474 versus 2382) is due to rounding errors in the Reboussin article. Reboussin rounds from four-digits to three-digits, which caused a large difference. ***PASS*** uses more accurate routines.

To see that the results are equal to within rounding error, we will compute the sample size using Reboussin's results, but with more decimal places in the intermediate steps. They had

$$n_K = \frac{2(0.096)(0.904)(3.28)^2}{(0.028)^2} = 2381.78$$

When we compute this without rounding, we get

$$n_K = \frac{2(0.09625)(0.90375)(3.27939)^2}{(0.0275)^2} = 2474.00$$

A sample size of 2474 is the result obtained in ***PASS***.

Chapter 221

Group-Sequential Tests for Two Proportions (Simulation)

Introduction

This procedure can be used to determine power, sample size and/or boundaries for group sequential tests comparing the proportions of two groups. The tests that can be simulated in this procedure are the common two-sample Z-test with or without pooled standard error and with or without continuity correction, the Mantel-Haenszel test, the T-test, and Fisher's Exact test. For two-sided tests, significance (efficacy) boundaries can be generated. For one-sided tests, significance and futility boundaries can be produced. The spacing of the looks can be equal or custom specified. Boundaries can be computed based on popular alpha- and beta-spending functions (O'Brien-Fleming, Pocock, Hwang-Shih-DeCani Gamma family, linear) or custom spending functions. Boundaries can also be input directly to verify alpha- and/or beta-spending properties. Futility boundaries can be binding or non-binding. Maximum and average (expected) sample sizes are reported as well as the alpha and/or beta spent and incremental power at each look. Corresponding P-Value boundaries are also given for each boundary statistic. Plots of boundaries are also produced.

Four Procedures Documented Here

There are four procedures in the menus that use the program module described in this chapter. These procedures are identical except for the type of parameterization. The parameterization can be in terms of proportions, differences in proportions, ratios of proportions, and odds ratios. Each of these options is listed separately on the menus. Whether the parameterization is in terms of proportions, differences, risk ratios, or odds ratios, the hypothesis tests concerning the differences are those that are simulated.

Technical Details

This section outlines many of the technical details of the techniques used in this procedure including the simulation summary, the test statistic details, and the use of spending functions.

An excellent text for the background and details of many group-sequential methods is Jennison and Turnbull (2000).

Simulation Procedure

In this procedure, a large number of simulations are used to calculate boundaries and power using the following steps

1. Based on the specified proportions, random samples of size N1 and N2 are generated under the null distribution and under the alternative distribution. These are simulated samples as though the final look is reached.
2. For each sample, test statistics for each look are produced. For example, if N1 and N2 are 100 and there are 5 equally spaced looks, test statistics are generated from the random samples at N1 = N2 = 20, N1 = N2 = 40, N1 = N2 = 60, N1 = N2 = 80, and N1 = N2 = 100 for both null and alternative samples.
3. To generate the first significance boundary, the null distribution statistics of the first look (e.g., at N1 = N2 = 20) are ordered and the percent of alpha to be spent at the first look is determined (using either the alpha-spending function or the input value). The statistic for which the percent of statistics above (or below, as the case may be) that value is equal to the percent of alpha to be spent at the first look is the boundary statistic. It is seen here how important a large number of simulations is to the precision of the boundary estimates.
4. All null distribution samples that are outside the first significance boundary at the first look are removed from consideration for the second look. If binding futility boundaries are also being computed, all null distribution samples with statistics that are outside the first futility boundary are also removed from consideration for the second look. If non-binding futility boundaries are being computed, null distribution samples with statistics outside the first futility boundary are not removed.
5. To generate the second significance boundary, the remaining null distribution statistics of the second look (e.g., at N1 = N2 = 40) are ordered and the percent of alpha to be spent at the second look is determined (again, using either the alpha-spending function or the input value). The percent of alpha to be spent at the second look is multiplied by the total number of simulations to determine the number of the statistic that is to be the second boundary statistic. The statistic for which that number of statistics is above it (or below, as the case may be) is the second boundary statistic. For example, suppose there are initially 1000 simulated samples, with 10 removed at the first look (from, say, alpha spent at Look 1 equal to 0.01), leaving 990 samples considered for the second look. Suppose further that the alpha to be spent at the second look is 0.02. This is multiplied by 1000 to give 20. The 990 still-considered statistics are ordered and the 970^{th} (20 in from 990) statistic is the second boundary.
6. All null distribution samples that are outside the second significance boundary and the second futility boundary, if binding, at the second look are removed from consideration for the third look (e.g., leaving 970 statistics computed at N1 = N2 = 60 to be considered at the third look). Steps 4 and 5 are repeated until the final look is reached.

Futility boundaries are computed in a similar manner using the desired beta-spending function or custom beta-spending values and the alternative hypothesis simulated statistics at each look. For both binding and non-binding futility boundaries, samples for which alternative hypothesis statistics are outside either the significance or futility boundaries of the previous look are excluded from current and future looks.

Because the final futility and significance boundaries are required to be the same, futility boundaries are computed beginning at a small value of beta (e.g., 0.0001) and incrementing beta by that amount until the futility and significance boundaries meet.

When boundaries are entered directly, this procedure uses the null hypothesis and alternative hypothesis simulations to determine the number of test statistics that are outside the boundaries at each look. The cumulative proportion of alternative hypothesis statistics that are outside the significance boundaries is the overall power of the study.

Small Sample Considerations

When the sample size is small, say 200 or fewer per group, the discrete nature of the number of possible differences in proportions in the sampling distribution comes into play. This has led to a large number of proposed tests for comparing two proportions (or testing the 2 by 2 table of counts). For example, Upton (1982) considers twenty-two alternative tests for comparing two proportions. Sweeping statements about the power of one test over another are impossible to make, because the size of the Type I error depends upon the proportions used. At some proportions, some tests are overly conservative while others are not, while at other proportions the reverse may be true.

This simulation procedure, however, is based primarily on the ordering of the sample statistics in the simulation. The boundaries are determined by the spending function alphas. Thus, if a test used happens to be conservative in the single-look traditional sense, the boundaries chosen in the simulation results of this procedure will generally remove the conservative nature of the test. This makes comparisons to the one-look case surprising in many cases.

Definitions

Suppose you have two populations from which dichotomous (binary) responses will be recorded. The probability (or risk) of obtaining the event of interest in population 1 (the treatment group) is p_1 and in population 2 (the control group) is p_2. The corresponding failure proportions are given by $q_1 = 1 - p_1$ and $q_2 = 1 - p_2$.

The assumption is made that the responses from each group follow a binomial distribution. This means that the event probability, p_i, is the same for all subjects within the group and that the response from one subject is independent of that of any other subject.

Random samples of *m* and *n* individuals are obtained from these two populations. The data from these samples can be displayed in a 2-by-2 contingency table as follows

Group	Success	Failure	Total
Treatment	*a*	*c*	*m*
Control	*b*	*d*	*n*
Total	*s*	*f*	*N*

The following alternative notation is also used.

Group	Success	Failure	Total
Treatment	x_{11}	x_{12}	n_1
Control	x_{21}	x_{22}	n_2
Total	m_1	m_2	N

The binomial proportions p_1 and p_2 are estimated from these data using the formulae

$$\hat{p}_1 = \frac{a}{m} = \frac{x_{11}}{n_1} \text{ and } \hat{p}_2 = \frac{b}{n} = \frac{x_{21}}{n_2}$$

Comparing Two Proportions

When analyzing studies such as this, one usually wants to compare the two binomial probabilities, p_1 and p_2. Common measures for comparing these quantities are the difference and the ratio. If the binomial probabilities are expressed in terms of odds rather than probabilities, another common measure is the odds ratio. Mathematically, these comparison parameters are

Parameter	**Computation**
Difference	$\delta = p_1 - p_2$
Risk Ratio	$\phi = p_1 / p_2$
Odds Ratio	$\psi = \frac{p_1 / (1 - p_1)}{p_2 / (1 - p_2)} = \frac{p_1 q_2}{p_2 q_1}$

The (risk) difference, $\delta = p_1 - p_2$, is perhaps the most direct measure for comparing two proportions. Three sets of statistical hypotheses can be formulated:

1. $H_0: p_1 - p_2 = 0$ versus $H_1: p_1 - p_2 \neq 0$; this is often called the *two-tailed test*.
2. $H_0: p_1 - p_2 \leq 0$ versus $H_1: p_1 - p_2 > 0$; this is often called the *upper-tailed test*.
3. $H_0: p_1 - p_2 \geq 0$ versus $H_1: p_1 - p_2 < 0$; this is often called the *lower-tailed test*.

In this procedure, whether the parameterization is in terms of proportions, differences, risk ratios, or odds ratios, the hypothesis tests concerning the differences are those that are simulated.

Test Statistics

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

Z Test (Pooled and Unpooled)

This test statistic was first proposed by Karl Pearson in 1900. Although this test can be expressed as a Chi-Square statistic, it is expressed here as a *z* so that it can be used for one-sided hypothesis testing.

Both *pooled* and *unpooled* versions of this test have been discussed in the statistical literature. The pooling refers to the way in which the standard error is estimated. In the pooled version, the two proportions are averaged, and only one proportion is used to estimate the standard error. In the unpooled version, the two proportions are used separately.

The formula for the test statistic is

$$z_t = \frac{\hat{p}_1 - \hat{p}_2}{\hat{\sigma}_D}$$

Pooled Version

$$\hat{\sigma}_D = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

$$\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$$

Unpooled Version

$$\hat{\sigma}_D = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

Continuity Correction

Frank Yates is credited with proposing a correction to the Pearson Chi-Square test for the lack of continuity in the binomial distribution. However, the correction was in common use when he proposed it in 1922.

The continuity corrected z-test is

$$z = \frac{\left(\hat{p}_1 - \hat{p}_2\right) + \frac{F}{2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}{\hat{\sigma}_D}$$

where F is -1 for upper-tailed, 1 for lower-tailed, and either -1 or 1 for two-sided hypotheses, depending on whether the numerator difference is positive or negative.

Conditional Mantel Haenszel Test

The conditional Mantel Haenszel test, see Lachin (2000) page 40, is based on the *index frequency*, x_{11}, from the 2x2 table. The formula for the z-statistic is

$$z = \frac{x_{11} - E(x_{11})}{\sqrt{V_c(x_{11})}}$$

where

$$E(x_{11}) = \frac{n_1 m_1}{N}$$

$$V_c(x_{11}) = \frac{n_1 n_2 m_1 m_2}{N^2(N-1)}$$

T-Test

Based on a study of the behavior of several tests, D'Agostino (1988) and Upton (1982) proposed using the usual two-sample *t*-test for testing whether two proportions are equal. One substitutes a '1' for a success and a '0' for a failure in the usual, two-sample *t*-test formula. The test statistic is computed as

$$t_{N-2} = (ad - bc)\left(\frac{N-2}{N(nac + mbd)}\right)^{\frac{1}{2}}$$

which can be compared to the *t* distribution with *N*-2 degrees of freedom.

Fisher's Exact Test

The most useful reference we found for power analysis of Fisher's Exact test was in the StatXact 5 (2001) documentation. The material present here is summarized from Section 26.3 (pages 866 – 870) of the StatXact-5 documentation. In this case, the test statistic is

$$T = -\ln\left[\frac{\binom{n_1}{x_1}\binom{n_2}{x_2}}{\binom{N}{m}}\right]$$

The null distribution of T is based on the hypergeometric distribution. It is given by

$$\Pr(T \ge t | m, H_0) = \sum_{A(m)}\left[\frac{\binom{n_1}{x_1}\binom{n_2}{x_2}}{\binom{N}{m}}\right]$$

where

$$A(m) = \{\text{all pairs } x_1, x_2 \text{ such that } x_1 + x_2 = m, \text{ given } T \ge t\}$$

Conditional on *m*, the critical value, t_α, is the smallest value of *t* such that

$$\Pr(T \ge t_\alpha | m, H_0) \le \alpha$$

The power is defined as

$$1 - \beta = \sum_{m=0}^{N} P(m) \Pr(T \ge t_\alpha | m, H_1)$$

where

$$\Pr(T \ge t_\alpha | m, H_1) = \sum_{A(m, T \ge t_\alpha)}\left[\frac{b(x_1, n_1, p_1)b(x_2, n_2, p_2)}{\sum_{A(m)} b(x_1, n_1, p_1)b(x_2, n_2, p_2)}\right]$$

$$\begin{aligned} P(m) &= \Pr(x_1 + x_2 = m | H_1) \\ &= b(x_1, n_1, p_1)b(x_2, n_2, p_2) \end{aligned}$$

$$b(x, n, p) = \binom{n}{x} p^x (1-p)^{n-x}$$

Spending Functions

Spending functions can be used in this procedure to specify the proportion of alpha or beta that is spent at each look without having to specify the proportion directly.

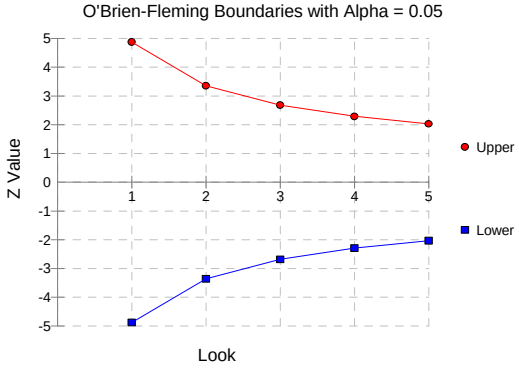
Spending functions have the characteristics that they are increasing and that

$$\alpha(0) = 0$$
$$\alpha(1) = \alpha$$

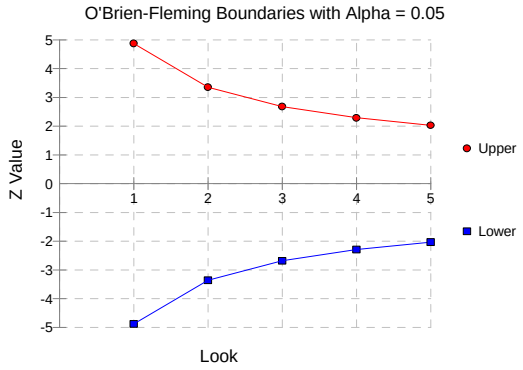
The last characteristic guarantees a fixed α level when the trial is complete. This methodology is very flexible since neither the times nor the number of analyses must be specified in advance. Only the functional form of $\alpha(\tau)$ must be specified.

PASS provides several popular spending functions plus the ability to enter and analyze your own percents of alpha or beta spent. These are calculated as follows (beta may be substituted for alpha for beta-spending functions):

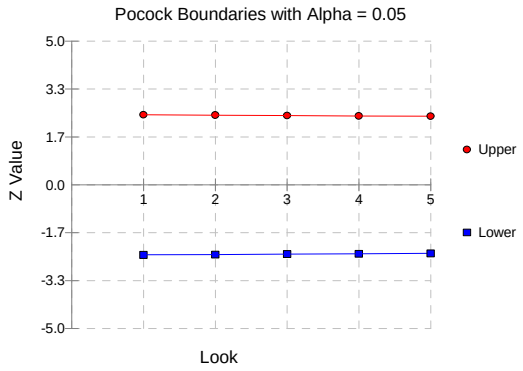
1. Hwang-Shih-DeCani (gamma family) $\alpha\left[\dfrac{1-e^{-\gamma t}}{1-e^{-\gamma}}\right], \gamma \neq 0$; $\alpha t, \gamma = 0$



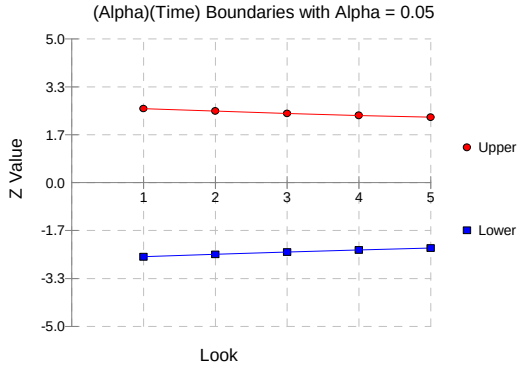
2. O'Brien-Fleming Analog $2 - 2\Phi\left(\frac{Z_{\alpha/2}}{\sqrt{t}}\right)$



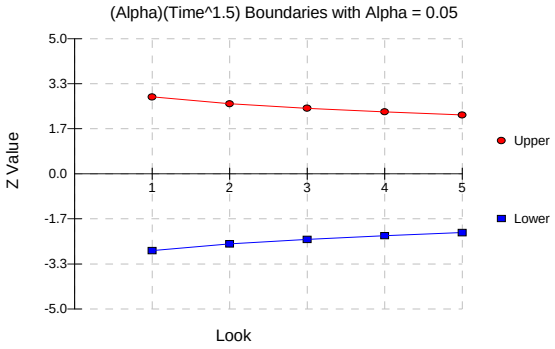
3. Pocock Analog $\alpha \cdot \ln(1 + (e-1)t)$



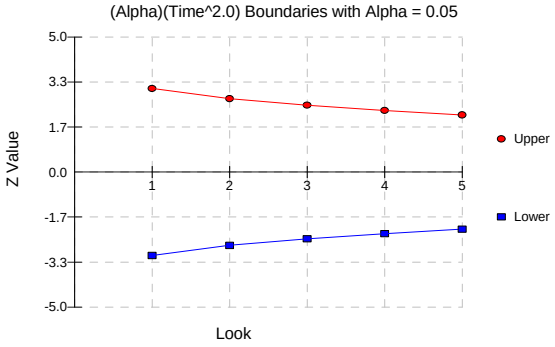
4. Alpha * time $\alpha \cdot t$



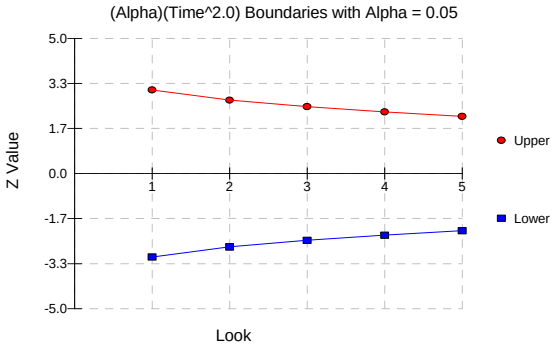
5. Alpha * time^1.5 $\alpha \cdot t^{3/2}$



6. Alpha * time^2 $\alpha \cdot t^{2}$



7. Alpha * time^C $\alpha \cdot t^{C}$



8. User Supplied Percents

A custom set of percents of alpha to be spent at each look may be input directly.

The O'Brien-Fleming Analog spends very little alpha or beta at the beginning and much more at the final looks. The Pocock Analog and (Alpha or Beta)(Time) spending functions spend alpha or beta more evenly across the looks. The Hwang-Shih-DeCani (C) (gamma family) spending functions and (Alpha or Beta)(Time^C) spending functions are flexible spending functions that can be used to spend more alpha or beta early or late or evenly, depending on the choice of C.

Procedure Options

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

Data Tab

The Data tab contains most of the parameters and options for the general setup of the procedure.

Solve For

Find (Solve For)

Solve for either power, sample size, or enter the boundaries directly and solve for power and alpha. When solving for power or sample size, the look and boundary details are specified on the "Looks & Boundaries" tab and the "Enter Boundaries" tab is ignored. When entering the boundaries directly and solving for power and alpha, the boundaries are input on the "Enter Boundaries" tab and the "Looks & Boundaries" tab is ignored.

When solving for power or N1, the early-stopping boundaries are also calculated. High accuracy for early-stopping boundaries requires a very large number of simulations (Recommended 100,000 to 10,000,000).

The parameter selected here is the parameter displayed on the vertical axis of the plot.

Because this is a simulation based procedure, the search for the sample size may take several minutes or hours to complete. You may find it quicker and more informative to solve for Power for a range of sample sizes.

Error Rates

Power or Beta

Power is the probability of rejecting the null hypothesis when it is false. Power is equal to 1-Beta, so specifying power implicitly specifies beta.

Beta is the probability obtaining a false negative on the statistical test. That is, it is the probability of accepting a false null hypothesis.

In the context of simulated group sequential trials, the power is the proportion of the alternative hypothesis simulations that cross any one of the significance (efficacy) boundaries.

The valid range is between 0 to 1.

Different disciplines and protocols have different standards for setting power. A common choice is 0.90, but 0.80 is also popular.

You can enter a range of values such as *0.70 0.80 0.90* or *0.70 to 0.90 by 0.1*.

Alpha (Significance Level)

Alpha is the probability of obtaining a false positive on the statistical test. That is, it is the probability of rejecting a true null hypothesis.

The null hypothesis is usually that the parameters (the means, proportions, etc.) are all equal.

In the context of simulated group sequential trials, alpha is the proportion of the null hypothesis simulations that cross any one of the significance (efficacy) boundaries.

Since Alpha is a probability, it is bounded by 0 and 1. Commonly, it is between 0.001 and 0.250.

Alpha is often set to 0.05 for two-sided tests and to 0.025 for one-sided tests.

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

Sample Size

N1 (Sample Size Group 1)

Enter a value for the sample size, N1. This is the number of subjects in the first group of the study at the final look.

You may enter a range such as 10 to 100 by 10 or a list of values separated by commas or blanks.

You might try entering the same number two or three times to get an idea of the variability in your results. For example, you could enter "10 10 10".

N2 (Sample Size Group 2)

Enter a value (or range of values) for the sample size of group 2. This is the number of subjects in the second group of the study at the final look.

- **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 the operator [*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*.

R = N2/N1

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*.

Effect Size

P1 (Treatment Group Proportion|H1) – Proportions

P1 is the proportion used in the simulations for Group 1 under H1.

The range is 0 to 1, but not zero or one. If the alternative hypothesis is one-sided, this value should reflect the direction of the one-sided test.

When P1 is close to zero or one, a Zero Count Adjustment may be needed.

You can enter a list of values such as *0.4 0.5 0.6* or *0.3 to 0.7 by 0.05*.

D1 (Difference|H1 = P1 – P2) – Differences

D1 is used to calculate P1, the proportion used in the simulations for Group 1 under H1.

P1 = P2 + D1

The range is -1 to 1, but not -1 or 1. This value should only result in values of P1 that are between 0 and 1. If the alternative hypothesis is one-sided, this value should reflect the direction of the one-sided test.

When the resulting P1 is close to zero or one, a Zero Count Adjustment may be needed.

You can enter a list of values such as *0.4 0.5 0.6* or *0.3 to 0.7 by 0.05*.

R1 (Ratio|H1 = P1 / P2) – Ratios

R1 is used to calculate P1, the proportion used in the simulations for Group 1 under H1.

P1 = R1 * P2

Although P1 is specified in terms of the ratio of P1 to P2, the simulated tests use the difference between P1 and P2.

The range is greater than zero, but typically not 1. This value should only result in values of P1 that are between 0 and 1. If the alternative hypothesis is one-sided, this value should reflect the direction of the one-sided test.

When the resulting P1 is close to zero or one, a Zero Count Adjustment may be needed.

You can enter a list of values such as *1.1 1.2 1.3* or *1.1 to 1.4 by .05*.

OR1 (Ratio|H1 = O1 / O2) – Odds Ratios

OR1 is used to calculate P1, the proportion used in the simulations for Group 1 under H1.

OR1 = [P1/(1-P1)]/[P2/(1-P2)] and P1 = [OR1 * P2] / [1 - P2 + OR1 * P2]

Although P1 is specified in terms of the oddes ratio of P1 to P2, the simulated tests use the difference between P1 and P2.

The range is greater than zero, but typically not 1. This value should only result in values of P1 that are between 0 and 1. If the alternative hypothesis is one-sided, this value should reflect the direction of the one-sided test.

When the resulting P1 is close to zero or one, a Zero Count Adjustment may be needed.

You can enter a list of values such as *1.1 1.2 1.3* or *1.1 to 1.4 by .05*.

P2 (Control Group Proportion)

P2 is the proportion used in the simulations for Group 1 and Group 2 under H0, and for Group 2 under H1.

The range is 0 to 1, but not zero or one.

When P2 is close to zero or one, a Zero Count Adjustment may be needed.

You can enter a list of values such as *0.4 0.5 0.6* or *0.3 to 0.7 by 0.05*.

Test and Simulations

Alternative Hypothesis

Specify the alternative hypothesis of the test. Since the null hypothesis is the opposite, specifing the alternative is the only hypothesis needed.

When a two-sided alternative is selected, futility boundaries are not permitted. If you wish to do a two-sided test with futility boundaries you may consider two one-sided tests.

Care need be taken in the choice of direction of the test relative to the choice of P1 and P2. For example, for a one-sided test to show that the true proportion for group 1 (control) is less than the true proportion for group 2 (treatment), P1 should be smaller than P2 and the alternative hypothesis chosen should be One-Sided (Prop1 < Prop2).

Test Type

Specify which test statistic is to be simulated and reported on.

The Z-test has the difference of the proportions in the numerator and a standard error of the difference in the denominator.

Pooled and unpooled refer to the method of estimating the standard error used in the denominator.

C. C. stands for continuity correction.

For details and formulation of the Z, Mantel-Haenszel, T, and Fisher's Exact tests, see the help manual.

Simulations

Specify the number of Monte Carlo iterations, M.

The following table gives an estimate of the precision that is achieved for various simulation sizes when the power is near 0.50 and 0.95. The values in the table are the "Precision" amounts that are added and subtracted to form a 95% confidence interval.

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

However, when solving for Power or N1, the simulations are used to calculate the look boundaries. To obtain precise boundary estimates, the number of simulations needs to be high. However, this consideration competes with the length of time to complete the simulation. When solving for power, a large number of simulations (100,000 or 1,000,000) will finish in several minutes. When solving for N1, perhaps 10,000 simulations can be run for each iteration. Then, a final run with the resulting N1 solving for power can be run with more simulations.

Looks & Boundaries Tab

The Data tab contains most of the parameters and options for the general setup of the procedure.

Looks and Boundaries

Specification of Looks and Boundaries

Choose whether spending functions will be used to divide alpha and beta for each look (Simple Specification), or whether the percents of alpha and beta to be spent at each look will be specified directly (Custom Specification).

Under Simple Specification, the looks are automatically considered to be equally spaced. Under Custom Specification, the looks may be equally spaced or custom defined based on the percent of accumulated information.

Looks and Boundaries – Simple Specification

Number of Equally Spaced Looks

Select the total number of looks that will be used if the study is not stopped early for the crossing of a boundary.

Alpha Spending Function

Specify the type of alpha spending function to use.

The O'Brien-Fleming Analog spends very little alpha at the beginning and much more at the final looks. The Pocock Analog and (Alpha)(Time) spending functions spend alpha more evenly across the looks. The Hwang-Shih-DeCani (C) (sometimes called the gamma family) spending functions and (Alpha)(Time^C) spending functions are flexible spending functions that can be used to spend more alpha early or late or evenly, depending on the choice of C.

C (Alpha Spending)

C is used to define the Hwang-Shih-DeCani (C) or (Alpha)(Time^C) spending functions.

For the Hwang-Shih-DeCani (C) spending function, negative values of C spend more of alpha at later looks, values near 0 spend alpha evenly, and positive values of C spend more of alpha at earlier looks.

For the (Alpha)(Time^C) spending function, only positive values for C are permitted. Values of C near zero spend more of alpha at earlier looks, values near 1 spend alpha evenly, and larger values of C spend more of alpha at later looks.

Type of Futility Boundary

This option determines whether or not futility boundaries will be created, and if so, whether they are binding or non-binding.

Futility boundaries are boundaries such that, if crossed at a given look, stop the study in favor of H0.

Binding futility boundaries are computed in concert with significance boundaries. They are called binding because they require the stopping of a trial if they are crossed. If the trial is not stopped, the probability of a false positive will exceed alpha.

When non-binding futility boundaries are computed, the significance boundaries are first computed, ignoring the futility boundaries. The futility boundaries are then computed. These futility boundaries are non-binding because continuing the trial after they are crossed will not affect the overall probability of a false positive declaration.

Number of Skipped Futility Looks

In some trials it may be desirable to wait a number of looks before examining the trial for futility. This option allows the beta to begin being spent after a specified number of looks.

The Number of Skipped Futility Looks should be less than the number of looks.

Beta Spending Function

Specify the type of beta spending function to use.

The O'Brien-Fleming Analog spends very little beta at the beginning and much more at the final looks. The Pocock Analog and (Beta)(Time) spending functions spend beta more evenly across the looks. The Hwang-Shih-DeCani (C) (sometimes called the gamma family) spending functions and (Beta)(Time^C) spending functions are flexible spending functions that can be used to spend more beta early or late or evenly, depending on the choice of C.

C (Beta Spending)

C is used to define the Hwang-Shih-DeCani (C) or (Beta)(Time^C) spending functions.

For the Hwang-Shih-DeCani (C) spending function, negative values of C spend more of beta at later looks, values near 0 spend beta evenly, and positive values of C spend more of beta at earlier looks.

For the (Beta)(Time^C) spending function, only positive values for C are permitted. Values of C near zero spend more of beta at earlier looks, values near 1 spend beta evenly, and larger values of C spend more of beta at later looks.

Looks and Boundaries – Custom Specification

Number of Looks

This is the total number of looks of either type (significance or futility or both).

Equally Spaced

If this box is checked, the Accumulated Information boxes are ignored and the accumulated information is evenly spaced.

Type of Futility Boundary

This option determines whether or not futility boundaries will be created, and if so, whether they are binding or non-binding.

Futility boundaries are boundaries such that, if crossed at a given look, stop the study in favor of H0.

Binding futility boundaries are computed in concert with significance boundaries. They are called binding because they require the stopping of a trial if they are crossed. If the trial is not stopped, the probability of a false positive will exceed alpha.

When Non-binding futility boundaries are computed, the significance boundaries are first computed, ignoring the futility boundaries. The futility boundaries are then computed. These futility boundaries are non-binding because continuing the trial after they are crossed will not affect the overall probability of a false positive declaration.

Accumulated Information

The accumulated information at each look defines the proportion or percent of the sample size that is used at that look.

These values are accumulated information values so they must be increasing.

Proportions, percents, or sample sizes may be entered. All proportions, percents, or sample sizes will be divided by the value at the final look to create an accumulated information proportion for each look.

Percent of Alpha Spent

This is the percent of the total alpha that is spent at the corresponding look. It is not the cumulative value.

Percents, proportions, or alphas may be entered here. Each of the values is divided by the sum of the values to obtain the proportion of alpha that is used at the corresponding look.

Percent of Beta Spent

This is the percent of the total beta (1-power) that is spent at the corresponding look. It is not the cumulative value.

Percents, proportions, or betas may be entered here. Each of the values is divided by the sum of the values to obtain the proportion of beta that is used at the corresponding look.

Enter Boundaries Tab

The Data tab contains most of the parameters and options for the general setup of the procedure.

Looks and Boundaries

Number of Looks

This is the total number of looks of either type (significance or futility or both).

Equally Spaced

If this box is checked, the Accumulated Information boxes are ignored and the accumulated information is evenly spaced.

Types of Boundaries

This option determines whether or not futility boundaries will be entered.

Futility boundaries are boundaries such that, if crossed at a given look, stop the study in favor of H0.

Accumulated Information

The accumulated information at each look defines the proportion or percent of the sample size that is used at that look.

These values are accumulated information values so they must be increasing.

Proportions, percents, or sample sizes may be entered. All proportions, percents, or sample sizes will be divided by the value at the final look to create an accumulated information proportion for each look.

Significance Boundary

Enter the value of the significance boundary corresponding to the chosen test statistic. These are sometimes called efficacy boundaries.

Futility Boundary

Enter the value of the futility boundary corresponding to the chosen test statistic.

Options Tab

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

Maximum Iterations

Maximum N1 Before Search Termination

Specify the maximum N1 before the search for N1 is aborted.

Since simulations for large sample sizes are very computationally intensive and hence time-consuming, this value can be used to stop searches when N1 is larger than reasonable sample sizes for the study.

This applies only when "Find (Solve For)" is set to N1.

The procedure uses a binary search when searching for N1. If a value for N1 is tried that exceeds this value, and the power is not reached, a warning message will be shown on the output indicating the desired power was not reached.

We recommend a value of at least 20000.

Matching Boundaries at Final Look

Beta Search Increment

For each simulation, when futility bounds are computed, the appropriate beta is found by searching from 0 to 1 by this increment. Smaller increments are more refined, but the search takes longer.

We recommend 0.001 or 0.0001.

Zero Count Adjustment

Zero Count Adjustment Method

Zero cell counts cause many calculation problems. To compensate for this, a small value (called the Zero Count Adjustment Value) may be added either to all cells or to all cells with zero counts. This option specifies whether you want to use the adjustment and which type of adjustment you want to use.

Adding a small value is controversial, but may be necessary. Some statisticians recommend adding 0.5 while others recommend 0.25. We have found that adding values as small as 0.0001 seems to work well.

Zero Count Adjustment Value

Zero cell counts cause many calculation problems. To compensate for this, a small value may be added either to all cells or to all zero cells. This is the amount that is added.

Some statisticians recommend that the value of 0.5 be added to all cells (both zero and non-zero). Others recommend 0.25. We have found that even a value as small as 0.0001 works well.

Example 1 – Power and Output

A clinical trial is to be conducted over a two-year period to compare the proportion response of a new treatment to that of the current treatment. The current response proportion is 0.56. Although the researchers do not know the true proportion of patients that will survive with the new treatment, they would like to examine the power that is achieved if the proportion under the new treatment is 0.63. The sample size at the final look is to be 1000 per group. Testing will be done at the 0.05 significance level. A total of five tests are going to be performed on the data as they are obtained. The O'Brien-Fleming (Analog) boundaries will be used. .

Find the power and test boundaries assuming equal sample sizes per arm and two-sided hypothesis tests.

Setup

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

Option	Value
Data Tab	
Find (Solve For)	**Power**
Power	*Ignored since this is the Find setting*
Alpha	**0.05**
N1 (Sample Size Group 1)	**1000**
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
P1 (Proportion in Group 1|H1)	**0.56**
P2 (Proportion in Group 2)	**0.63**
Alternative Hypothesis	**Two-Sided**
Test Type	**Z-Test (Pooled)**
Simulations	**100000**
Looks and Boundaries Tab	
Specification of Looks and Boundaries	**Simple**
Number of Equally Spaced Looks	**5**
Alpha Spending Function	**O'Brien-Fleming Analog**

Output

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

Numeric Results and Plots

Scenario 1 Numeric Results for Group Sequential Testing Proportion Difference = 0.
Hypotheses: H0: Proportion 1 = Proportion 2; H1: Proportion 1 <> Proportion 2
Test Statistic: Z-Test (Pooled)
Zero Adjustment Method: Add 0 to each zero
Alpha-Spending Function: O'Brien-Fleming Analog
Beta-Spending Function: None
Futility Boundary Type: None
Number of Looks: 5
Simulations: 100000

Numeric Summary for Scenario 1

Power			Alpha				
Value	95% LCL	95% UCL	Target	Actual	95% LCL	95% UCL	Beta
0.885	0.883	0.887	0.050	0.051	0.050	0.052	0.115

		Average Sample Size							
		Given H0		Given H1					
N1	N2	Grp1	Grp2	Grp1	Grp2	Diff0	Diff1	P1|H1	P2
1000	1000	993	993	757	757	0.0	-0.1	0.6	0.6

References

Jennison, C.; Turnbull, B.W. 2000. Group Sequential Methods with Applications to Clinical Trials. Chapman & Hall. Boca Raton, FL.
Matsumoto, M. and Nishimura,T. 1998. 'Mersenne twister: A 623-dimensionally equidistributed uniform pseudorandom number generator.' ACM Trans. On Modeling and Computer Simulations.

Report Definitions

Power is the probability of rejecting a false null hypothesis at one of the looks. It is the total proportion of alternative hypothesis simulations that are outside the significance boundaries.
Power 95% LCL and UCL are the lower and upper confidence limits for the power estimate. The width of the interval is based on the number of simulations.
Target Alpha is the user-specified probability of rejecting a true null hypothesis. It is the total alpha spent.
Alpha or Actual Alpha is the alpha level that was actually achieved by the experiment. It is the total proportion of the null hypothesis simulations that are outside the significance boundaries.
Alpha 95% LCL and UCL are the lower and upper confidence limits for the actual alpha estimate. The width of the interval is based on the number of simulations.
Beta is the probability of accepting a false null hypothesis. It is the total proportion of alternative hypothesis simulations that do not cross the significance boundaries.
N1 and N2 are the sample sizes of each group if the study reaches the final look.
Average Sample Size Given H0 Grp1 and Grp2 are the average or expected sample sizes of each group if H0 is true. These are based on the proportion of null hypothesis simulations that cross the significance or futility boundaries at each look.
Average Sample Size Given H1 Grp1 and Grp2 are the average or expected sample sizes of each group if H1 is true. These are based on the proportion of alternative hypothesis simulations that cross the significance or futility boundaries at each look.
Diff0 is the proportion difference between groups (Grp1 - Grp2) assuming the null hypothesis, H0.
Diff1 is the proportion difference between groups (Grp1 - Grp2) assuming the alternative hypothesis, H1.
P1|H1 is the proportion used in the simulations for Group 1 under H1.
P2 is the proportion used in the simulations for Group 1 and Group 2 under H0, and for Group 2 under H1.

Summary Statements

Group sequential trials with sample sizes of 1000 and 1000 at the final look achieve 89% power to detect a difference of -0.1 between a treatment group proportion of 0.6 and a control group proportion of 0.6 at the 0.101 significance level (alpha) using a two-sided Z-Test (Pooled).

221-20 Group-Sequential Tests for Two Proportions (Simulation)

Accumulated Information Details for Scenario 1

Look	Accumulated Information Percent	Accumulated Sample Size Group 1	Accumulated Sample Size Group 2	Accumulated Sample Size Total
1	20.0	200	200	400
2	40.0	400	400	800
3	60.0	600	600	1200
4	80.0	800	800	1600
5	100.0	1000	1000	2000

Accumulated Information Details Definitions

Look is the number of the look.
Accumulated Information Percent is the percent of the sample size accumulated up to the corresponding look.
Accumulated Sample Size Group 1 is total number of individuals in group 1 at the corresponding look.
Accumulated Sample Size Group 2 is total number of individuals in group 2 at the corresponding look.
Accumulated Sample Size Total is total number of individuals in the study (group 1 + group 2) at the corresponding look.

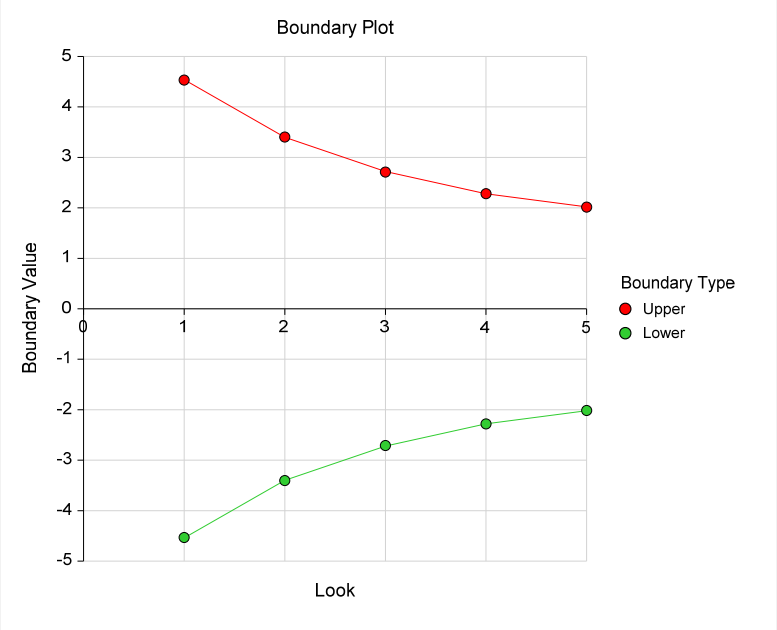
Boundaries for Scenario 1

Look	-- Significance Boundary -- Z-Value Scale	P-Value Scale
1	+/- 4.418	0.000
2	+/- 3.364	0.001
3	+/- 2.716	0.007
4	+/- 2.290	0.022
5	+/- 2.022	0.043

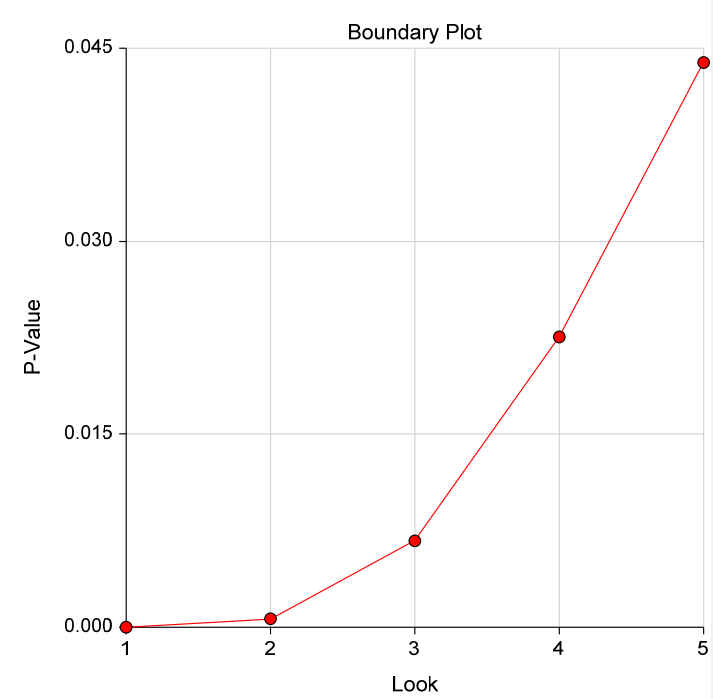
Boundaries Definitions

Look is the number of the look.
Significance Boundary Z-Value Scale is the value such that statistics outside this boundary at the corresponding look indicate termination of the study and rejection of the null hypothesis. They are sometimes called efficacy boundaries.
Significance Boundary P-Value Scale is the value such that P-Values outside this boundary at the corresponding look indicate termination of the study and rejection of the null hypothesis. This P-Value corresponds to the Z-Value Boundary and is sometimes called the nominal alpha.

Boundary Plot



Boundary Plot - P-Value



Significance Boundaries with 95% Simulation Confidence Intervals for Scenario 1

	--- Z-Value Boundary ---			--- P-Value Boundary ---		
Look	Value	95% LCL	95% UCL	Value	95% LCL	95% UCL
1	+/- 4.418			0.000		
2	+/- 3.364	-3.488	-3.302	0.001	0.000	0.001
3	+/- 2.716	-2.753	-2.677	0.007	0.006	0.007
4	+/- 2.290	-2.318	-2.272	0.022	0.020	0.023
5	+/- 2.022	-2.038	-2.000	0.043	0.042	0.045

Significance Boundary Confidence Limit Definitions

Look is the number of the look.

Z-Value Boundary Value is the value such that statistics outside this boundary at the corresponding look indicate termination of the study and rejection of the null hypothesis. They are sometimes called efficacy boundaries.

P-Value Boundary Value is the value such that P-Values outside this boundary at the corresponding look indicate termination of the study and rejection of the null hypothesis. This P-Value corresponds to the Z-Value Boundary and is sometimes called the nominal alpha.

95% LCL and UCL are the lower and upper confidence limits for the boundary at the given look. The width of the interval is based on the number of simulations.

Alpha-Spending and Null Hypothesis Simulation Details for Scenario 1

	--- Signif. Boundary---		--- Target ---		--- Actual ---		Proportion	Cum.
	Z-Value	P-Value	Spending Function	Cum. Spending Function		Cum.	H1 Sims Outside	H1 Sims Outside
Look	Scale	Scale	Alpha	Alpha	Alpha Spent	Alpha Spent	Signif. Boundary	Signif. Boundary
1	+/- 4.418	0.000	0.000	0.000	0.000	0.000	0.001	0.001
2	+/- 3.364	0.001	0.001	0.001	0.001	0.001	0.087	0.089
3	+/- 2.716	0.007	0.007	0.008	0.007	0.007	0.316	0.405
4	+/- 2.290	0.022	0.017	0.024	0.017	0.025	0.318	0.723
5	+/- 2.022	0.043	0.026	0.050	0.026	0.051	0.162	0.885

Alpha-Spending Details Definitions

Look is the number of the look.

Significance Boundary Z-Value Scale is the value such that statistics outside this boundary at the corresponding look indicate termination of the study and rejection of the null hypothesis. They are sometimes called efficacy boundaries.

Significance Boundary P-Value Scale is the value such that P-Values outside this boundary at the corresponding look indicate termination of the study and rejection of the null hypothesis. This P-Value corresponds to the Significance Z-Value Boundary and is sometimes called the nominal alpha.

Spending Function Alpha is the intended portion of alpha allocated to the particular look based on the alpha-spending function.

Cumulative Spending Function Alpha is the intended accumulated alpha allocated to the particular look. It is the sum of the Spending Function Alpha up to the corresponding look.

Alpha Spent is the proportion of the null hypothesis simulations resulting in statistics outside the Significance Boundary at this look.

Cumulative Alpha Spent is the proportion of the null hypothesis simulations resulting in Significance Boundary termination up to and including this look. It is the sum of the Alpha Spent up to the corresponding look.

Proportion H1 Sims Outside Significance Boundary is the proportion of the alternative hypothesis simulations resulting in statistics outside the Significance Boundary at this look. It may be thought of as the incremental power.

Cumulative H1 Sims Outside Significance Boundary is the proportion of the alternative hypothesis simulations resulting in Significance Boundary termination up to and including this look. It is the sum of the Proportion H1 Sims Outside Significance Boundary up to the corresponding look.

Run Time: 69.77 seconds.

The values obtained from any given run of this example will vary slightly due to the variation in simulations.

Example 2 – Power for One-Sided Test with Futility Boundaries

Suppose researchers would like to compare two treatments with a one-sided test at each look. Further, suppose they would like to terminate the study early when it can be deemed highly unlikely that the new treatment is better than the standard. Suppose the control group proportion is 0.46. The researchers wish to know the power of the test if the treatment group proportion is 0.53. The sample size at the final look is to be 1000 per group. Testing will be done at the 0.05 significance level. A total of five tests are going to be performed on the data as they are obtained. The O'Brien-Fleming (Analog) boundaries will be used for both significance and futility boundaries.

Find the power and test boundaries assuming equal sample sizes per arm and one-sided hypothesis tests.

Setup

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

Option	Value
Data Tab	
Find (Solve For)	**Power**
Power	*Ignored since this is the Find setting*
Alpha	**0.05**
N1 (Sample Size Group 1)	**1000**
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
P1 (Proportion in Group 1|H1)	**0.53**
P2 (Proportion in Group 2)	**0.46**
Alternative Hypothesis	**One-Sided (Prop1 > Prop2)**
Test Type	**Z-Test (Pooled)**
Simulations	**100000**
Looks and Boundaries Tab	
Specification of Looks and Boundaries	**Simple**
Number of Equally Spaced Looks	**5**
Alpha Spending Function	**O'Brien-Fleming Analog**
Type of Futility Boundary	**Non-Binding**
Number of Skipped Futility Looks	**0**
Beta Spending Function	**O'Brien-Fleming Analog**

Output

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

Numeric Results and Plots

Scenario 1 Numeric Results for Group Sequential Testing Proportion Difference = 0.
Hypotheses: H0: Proportion 1 = Proportion 2; H1: Proportion 1 > Proportion 2
Test Statistic: Z-Test (Pooled)
Zero Adjustment Method: Add 0 to each zero
Alpha-Spending Function: O'Brien-Fleming Analog
Beta-Spending Function: O'Brien-Fleming Analog
Futility Boundary Type: Non-Binding
Number of Looks: 5
Simulations: 100000

221-24 Group-Sequential Tests for Two Proportions (Simulation)

Numeric Summary for Scenario 1

Power			Alpha				
Value	95% LCL	95% UCL	Target	Actual	95% LCL	95% UCL	Beta
0.904	0.902	0.905	0.050	0.045	0.044	0.047	0.096

		Average Sample Size							
		Given H0		Given H1					
N1	N2	Grp1	Grp2	Grp1	Grp2	Diff0	Diff1	P1|H1	P2
1000	1000	600	600	675	675	0.0	0.1	0.5	0.5

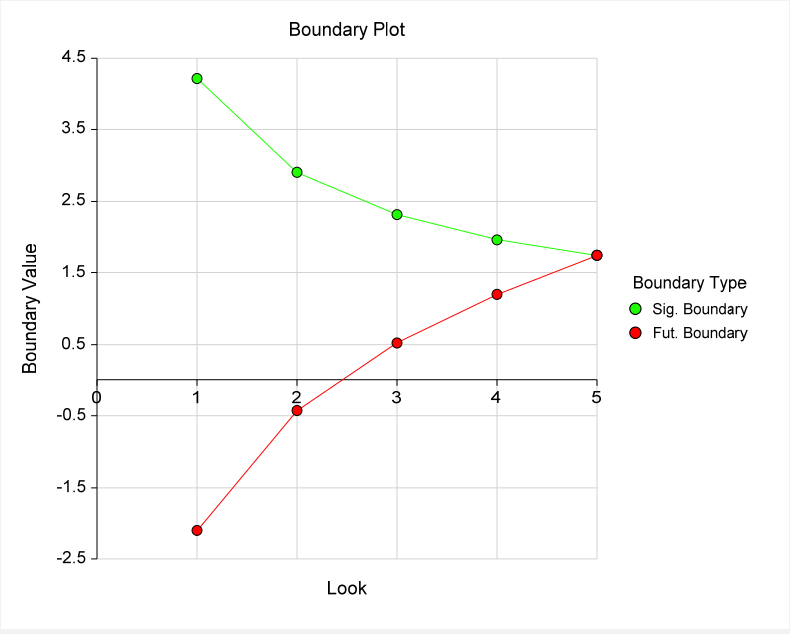
Accumulated Information Details for Scenario 1

	Accumulated Information	Accumulated Sample Size		
Look	Percent	Group 1	Group 2	Total
1	20.0	200	200	400
2	40.0	400	400	800
3	60.0	600	600	1200
4	80.0	800	800	1600
5	100.0	1000	1000	2000

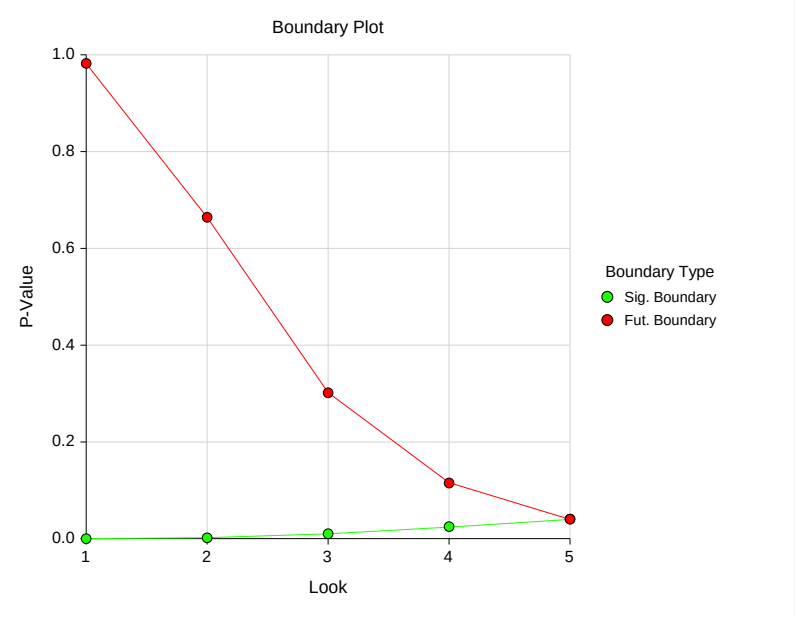
Boundaries for Scenario 1

	Significance Boundary		Futility Boundary	
	Z-Value	P-Value	Z-Value	P-Value
Look	Scale	Scale	Scale	Scale
1	4.020	0.000	-2.100	0.982
2	2.902	0.002	-0.424	0.664
3	2.312	0.010	0.520	0.302
4	1.958	0.025	1.200	0.115
5	1.746	0.040	1.746	0.040

Boundary Plot



Boundary Plot - P-Value



Significance Boundaries with 95% Simulation Confidence Intervals for Scenario 1

	Z-Value Boundary			P-Value Boundary		
Look	Value	95% LCL	95% UCL	Value	95% LCL	95% UCL
1	4.020			0.000		
2	2.902	2.841	2.946	0.002	0.002	0.002
3	2.312	2.263	2.317	0.010	0.010	0.012
4	1.958	1.955	1.964	0.025	0.025	0.025
5	1.746	1.707	1.749	0.040	0.040	0.044

Futility Boundaries with 95% Simulation Confidence Intervals for Scenario 1

	Z-Value Boundary			P-Value Boundary		
Look	Value	95% LCL	95% UCL	Value	95% LCL	95% UCL
1	-2.100	-2.300	-2.000	0.982		
2	-0.424	-0.425	-0.424	0.664	0.664	0.664
3	0.520	0.520	0.577	0.302	0.282	0.302
4	1.200	1.200	1.200	0.115	0.115	0.115
5	1.746	1.745	1.789	0.040	0.037	0.041

Alpha-Spending and Null Hypothesis Simulation Details for Scenario 1

	Signif. Boundary		Target		Actual		Proportion H0 Sims Outside	Cum. H0 Sims Outside
Look	Z-Value Scale	P-Value Scale	Spending Function Alpha	Cum. Spending Function Alpha	Alpha Spent	Cum. Alpha Spent	Futility Boundary	Futility Boundary
1	4.020	0.000	0.000	0.000	0.000	0.000	0.019	0.019
2	2.902	0.002	0.002	0.002	0.002	0.002	0.329	0.348
3	2.312	0.010	0.009	0.011	0.009	0.011	0.352	0.700
4	1.958	0.025	0.017	0.028	0.017	0.028	0.189	0.890
5	1.746	0.040	0.022	0.050	0.017	0.045	0.065	0.955

Beta-Spending and Alternative Hypothesis Simulation Details for Scenario 1

	-- Futility Boundary--		--------- Target ---------		--------- Actual ----------		Proportion H1 Sims Outside	Cum. H1 Sims Outside
Look	Z-Value Scale	P-Value Scale	Spending Function Beta	Cum. Spending Function Beta	Beta Spent	Cum. Beta Spent	Signif. Boundary	Signif. Boundary
1	-2.100	0.982	0.000	0.000	0.000	0.000	0.004	0.004
2	-0.424	0.664	0.008	0.009	0.008	0.009	0.168	0.172
3	0.520	0.302	0.023	0.032	0.023	0.032	0.369	0.541
4	1.200	0.115	0.031	0.064	0.031	0.063	0.263	0.803
5	1.746	0.040	0.033	0.097	0.033	0.096	0.100	0.904

Run Time: 70.39 seconds.

The values obtained from any given run of this example will vary slightly due to the variation in simulations.

Example 3 – Enter Boundaries

With a set-up similar to Example 2, suppose we wish to investigate the properties of a set of significance (3, 3, 3, 2, 2) and futility (-2, -1, 0, 0, 0) boundaries.

Setup

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

Option	Value
Data Tab	
Find (Solve For)	**Alpha and Power (Enter Boundaries)**
Power	*Ignored since this is the Find setting*
Alpha	*Ignored since this is the Find setting*
N1 (Sample Size Group 1)	**1000**
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
P1 (Proportion in Group 1|H1)	**0.53**
P2 (Proportion in Group 2)	**0.46**
Alternative Hypothesis	**One-Sided (Prop1 > Prop2)**
Test Type	**Z-Test (Pooled)**
Simulations	**100000**
Enter Boundaries Tab	
Number of Looks	**5**
Types of Boundaries	**Significance and Futility Boundaries**
Significance Boundary	**3 3 3 2 2 (for looks 1 through 5)**
Futility Boundary	**-2 -1 0 0 0 (for looks 1 through 5)**

Output

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

Numeric Results and Plots

Scenario 1 Numeric Results for Group Sequential Testing Proportion Difference = 0.
Hypotheses: H0: Proportion 1 = Proportion 2; H1: Proportion 1 > Proportion 2
Test Statistic: Z-Test (Pooled)
Zero Adjustment Method: Add 0 to each zero
Type of Boundaries: Significance and Futility Boundaries
Number of Looks: 5
Simulations: 100000

Numeric Summary for Scenario 1

Power			Alpha			
Value	95% LCL	95% UCL	Value	95% LCL	95% UCL	Beta
0.890	0.888	0.892	0.035	0.034	0.036	0.110

		Average Sample Size							
		Given H0		Given H1					
N1	N2	Grp1	Grp2	Grp1	Grp2	Diff0	Diff1	P1|H1	P2
1000	1000	403	403	625	625	0.0	0.1	0.5	0.5

Accumulated Information Details for Scenario 1

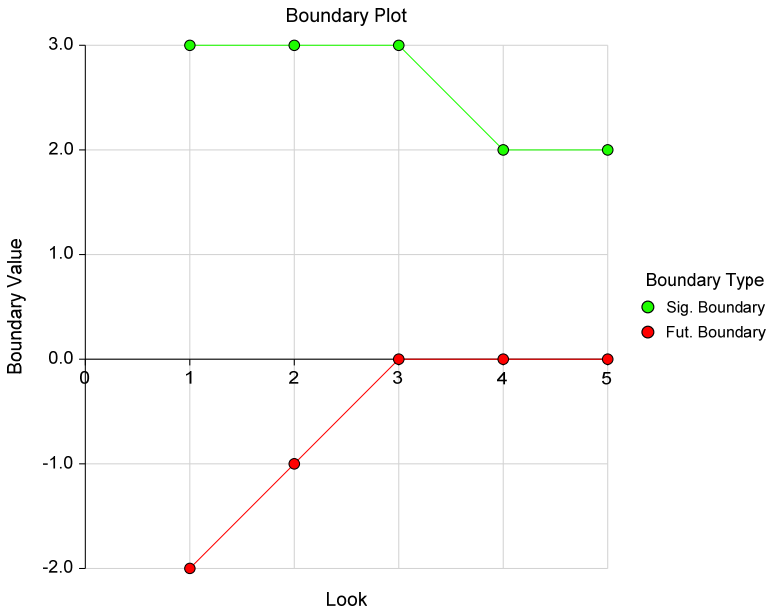
	Accumulated Information	Accumulated Sample Size		
Look	Percent	Group 1	Group 2	Total
1	20.0	200	200	400
2	40.0	400	400	800
3	60.0	600	600	1200
4	80.0	800	800	1600
5	100.0	1000	1000	2000

Boundaries for Scenario 1

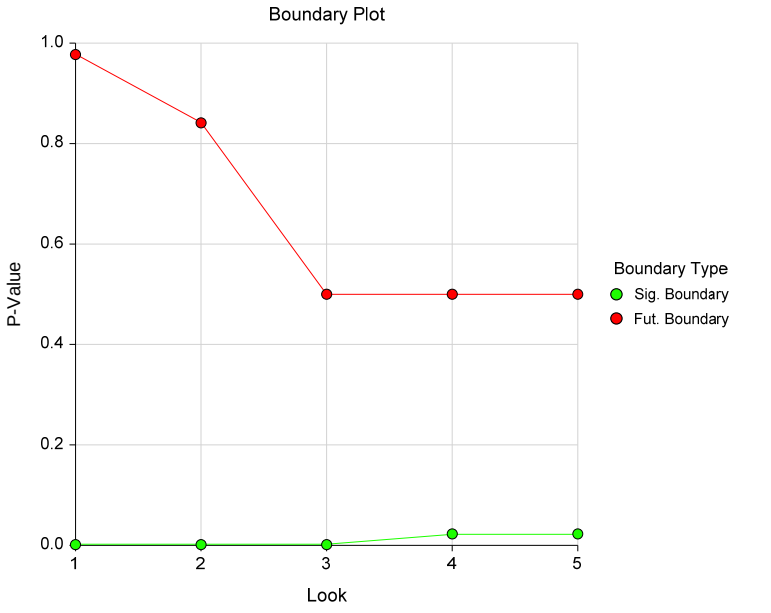
	Significance Boundary		Futility Boundary	
	Z-Value	P-Value	Z-Value	P-Value
Look	Scale	Scale	Scale	Scale
1	3.000	0.001	-2.000	0.977
2	3.000	0.001	-1.000	0.841
3	3.000	0.001	0.000	0.500
4	2.000	0.023	0.000	0.500
5	2.000	0.023	0.000	0.500

221-28 Group-Sequential Tests for Two Proportions (Simulation)

Boundary Plot



Boundary Plot - P-Value



Alpha-Spending and Null Hypothesis Simulation Details for Scenario 1

Look	--- Signif. Boundary--- Z-Value Scale	P-Value Scale	Alpha Spent	Cum. Alpha Spent	Proportion H0 Sims Outside Futility Boundary	Cum. H0 Sims Outside Futility Boundary
1	3.000	0.001	0.002	0.002	0.025	0.025
2	3.000	0.001	0.001	0.003	0.132	0.157
3	3.000	0.001	0.001	0.004	0.336	0.493
4	2.000	0.023	0.022	0.026	0.082	0.575
5	2.000	0.023	0.009	0.035	0.050	0.625

Beta-Spending and Alternative Hypothesis Simulation Details for Scenario 1

Look	-- Futility Boundary-- Z-Value Scale	P-Value Scale	Beta Spent	Cum. Beta Spent	Proportion H1 Sims Outside Signif. Boundary	Cum. H1 Sims Outside Signif. Boundary
1	-2.000	0.977	0.000	0.000	0.059	0.059
2	-1.000	0.841	0.001	0.002	0.111	0.170
3	0.000	0.500	0.006	0.008	0.153	0.323
4	0.000	0.500	0.001	0.009	0.477	0.800
5	0.000	0.500	0.000	0.009	0.091	0.890

Run Time: 62.89 seconds.

The values obtained from any given run of this example will vary slightly due to the variation in simulations.

Example 4 – 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 = 5, alpha = 0.05, beta = 0.10, P1 = 0.1100, P2 = 0.0825. They compute a drift of 3.28 and a sample size of 2381.78 per group. The upper boundaries are: 4.8769, 3.3569, 2.6803, 2.2898, 2.0310.

Setup

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

Option	Value
Data Tab	
Find (Solve For)	**N1**
Power	**0.90**
Alpha	**0.05**
N1 (Sample Size Group 1)	*Ignored since this is the Find setting*
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
P1 (Proportion in Group 1|H1)	**0.1100**
P2 (Proportion in Group 2)	**0.0825**
Alternative Hypothesis	**Two-Sided**
Test Type	**Z-Test (Pooled)**
Simulations	**100000**
Looks and Boundaries Tab	
Specification of Looks and Boundaries	**Simple**
Number of Equally Spaced Looks	**5**
Alpha Spending Function	**O'Brien-Fleming Analog**

Output

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

Numeric Results and Plots

Scenario 1 Numeric Results for Group Sequential Testing Proportion Difference = 0.
Hypotheses: H0: Proportion 1 = Proportion 2; H1: Proportion 1 <> Proportion 2
Test Statistic: Z-Test (Pooled)
Zero Adjustment Method: Add 0 to each zero
Alpha-Spending Function: O'Brien-Fleming Analog
Beta-Spending Function: None
Futility Boundary Type: None
Number of Looks: 5
Simulations: 100000

Numeric Summary for Scenario 1

Power			Alpha				
Value	95% LCL	95% UCL	Target	Actual	95% LCL	95% UCL	Beta
0.900	0.899	0.902	0.050	0.097	0.095	0.099	0.100

		Average Sample Size							
		Given H0		Given H1					
N1	N2	Grp1	Grp2	Grp1	Grp2	Diff0	Diff1	P1|H1	P2
2487	2487	2456	2456	1830	1830	0.0	0.0	0.1	0.1

The values obtained from any given run of this example will vary slightly due to the variation in simulations.

The sample size generated in this run is 2487. As discussed in Example 5 of Chapter 220, the correct sample size is 2474. Each run of this simulation procedure may produce a slightly different value for the sample size, especially if the number of simulations is lower. Larger numbers of simulations are required to obtain the increased precision for the boundary values. A run with N1 and N2 equal to 2474 and 1000000 simulations gives the following boundaries.

Boundaries for Scenario 1

Look	-- Significance Boundary -- Z-Value Scale	P-Value Scale
1	+/- 4.154	0.000
2	+/- 3.334	0.001
3	+/- 2.685	0.007
4	+/- 2.313	0.021
5	+/- 2.058	0.040

Run Time: 19.48 minutes.

These values are slightly off at the second or third decimal place, showing that very large simulation numbers are needed to obtain accurate boundaries.

Example 5 – Validation Using Simulation

With a set-up similar to Example 1, we examine the power and alpha generated by the set of two-sided significance boundaries (+/- 4.418, +/- 3.364, +/- 2.716, +/- 2.290, +/- 2.022).

Setup

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

Option	Value
Data Tab	
Find (Solve For)	**Alpha and Power (Enter Boundaries)**
Power	*Ignored since this is the Find setting*
Alpha	*Ignored since this is the Find setting*
N1 (Sample Size Group 1)	**1000**
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
P1 (Proportion in Group 1|H1)	**0.56**
P2 (Proportion in Group 2)	**0.63**
Alternative Hypothesis	**Two-Sided**
Test Type	**Z-Test (Pooled)**
Simulations	**100000**
Enter Boundaries Tab	
Number of Looks	**5**
Types of Boundaries	**Significance Boundaries**
Significance Boundary	**4.418, 3.364, 2.716, 2.290, 2.022 (looks 1 through 5)**

Output

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

Numeric Results and Plots

Scenario 1 Numeric Results for Group Sequential Testing Proportion Difference = 0.
Hypotheses: H0: Proportion 1 = Proportion 2; H1: Proportion 1 <> Proportion 2
Test Statistic: Z-Test (Pooled)
Zero Adjustment Method: Add 0 to each zero
Type of Boundaries: Significance Boundaries Only
Number of Looks: 5
Simulations: 100000

Numeric Summary for Scenario 1

Power			Alpha			
Value	95% LCL	95% UCL	Value	95% LCL	95% UCL	Beta
0.886	0.884	0.888	0.051	0.049	0.052	0.114

		Average Sample Size							
		Given H0		Given H1					
N1	N2	Grp1	Grp2	Grp1	Grp2	Diff0	Diff1	P1|H1	P2
1000	1000	993	993	756	756	0.0	-0.1	0.6	0.6

Run Time: 62.61 seconds.

The values obtained from any given run of this example will vary slightly due to the variation in simulations. The power and alpha generated with these boundaries are very close to the values of Example 1.

Chapter 222

Group-Sequential Non-Zero Null Tests for Two Proportions (Simulation)

Introduction

This procedure can be used to determine power, sample size and/or boundaries for group sequential non-zero (or non-unity, for ratios and odds ratios) null tests comparing the proportions of two groups. The tests that can be simulated in this procedure are the common two-sample Z-test with or without pooled standard error and with or without continuity correction, the T-test, and three score tests. Significance and futility boundaries can be produced. The spacing of the looks can be equal or custom specified. Boundaries can be computed based on popular alpha- and beta-spending functions (O'Brien-Fleming, Pocock, Hwang-Shih-DeCani Gamma family, linear) or custom spending functions. Boundaries can also be input directly to verify alpha- and/or beta-spending properties. Futility boundaries can be binding or non-binding. Maximum and average (expected) sample sizes are reported as well as the alpha and/or beta spent and incremental power at each look. Corresponding P-Value boundaries are given for each boundary statistic. Plots of boundaries are also produced.

Four Procedures Documented Here

There are four procedures that use the program module described in this chapter. The Proportions and Differences procedures are identical except for the type of parameterization. The Ratios and Odds Ratios procedures use only likelihood score tests that are specific to testing the ratio of two proportions and the odds ratio of two proportions, respectively. Each procedure is listed separately on the menus.

Technical Details

This section outlines many of the technical details of the techniques used in this procedure including the simulation summary, the test statistic details, and the use of spending functions.

An excellent text for the background and details of many group-sequential methods is Jennison and Turnbull (2000).

Simulation Procedure

In this procedure, a large number of simulations are used to calculate boundaries and power using the following steps

1. Based on the specified proportions, random samples of size N1 and N2 are generated under the null distribution and under the alternative distribution. These are simulated samples as though the final look is reached.
2. For each sample, test statistics for each look are produced. For example, if N1 and N2 are 100 and there are 5 equally spaced looks, test statistics are generated from the random samples at N1 = N2 = 20, N1 = N2 = 40, N1 = N2 = 60, N1 = N2 = 80, and N1 = N2 = 100 for both null and alternative samples.
3. To generate the first significance boundary, the null distribution statistics of the first look (e.g., at N1 = N2 = 20) are ordered and the percent of alpha to be spent at the first look is determined (using either the alpha-spending function or the input value). The statistic for which the percent of statistics above (or below, as the case may be) that value is equal to the percent of alpha to be spent at the first look is the boundary statistic. It is seen here how important a large number of simulations is to the precision of the boundary estimates.
4. All null distribution samples that are outside the first significance boundary at the first look are removed from consideration for the second look. If binding futility boundaries are also being computed, all null distribution samples with statistics that are outside the first futility boundary are also removed from consideration for the second look. If non-binding futility boundaries are being computed, null distribution samples with statistics outside the first futility boundary are not removed.
5. To generate the second significance boundary, the remaining null distribution statistics of the second look (e.g., at N1 = N2 = 40) are ordered and the percent of alpha to be spent at the second look is determined (again, using either the alpha-spending function or the input value). The percent of alpha to be spent at the second look is multiplied by the total number of simulations to determine the number of the statistic that is to be the second boundary statistic. The statistic for which that number of statistics is above it (or below, as the case may be) is the second boundary statistic. For example, suppose there are initially 1000 simulated samples, with 10 removed at the first look (from, say, alpha spent at Look 1 equal to 0.01), leaving 990 samples considered for the second look. Suppose further that the alpha to be spent at the second look is 0.02. This is multiplied by 1000 to give 20. The 990 still-considered statistics are ordered and the 970^{th} (20 in from 990) statistic is the second boundary.
6. All null distribution samples that are outside the second significance boundary and the second futility boundary, if binding, at the second look are removed from consideration for the third look (e.g., leaving 970 statistics computed at N1 = N2 = 60 to be considered at the third look). Steps 4 and 5 are repeated until the final look is reached.

Futility boundaries are computed in a similar manner using the desired beta-spending function or custom beta-spending values and the alternative hypothesis simulated statistics at each look. For both binding and non-binding futility boundaries, samples for which alternative hypothesis statistics are outside either the significance or futility boundaries of the previous look are excluded from current and future looks.

Because the final futility and significance boundaries are required to be the same, futility boundaries are computed beginning at a small value of beta (e.g., 0.0001) and incrementing beta by that amount until the futility and significance boundaries meet.

When boundaries are entered directly, this procedure uses the null hypothesis and alternative hypothesis simulations to determine the number of test statistics that are outside the boundaries at each look. The cumulative proportion of alternative hypothesis statistics that are outside the significance boundaries is the overall power of the study.

Small Sample Considerations

When the sample size is small, say 200 or fewer per group, the discrete nature of the number of possible differences in proportions in the sampling distribution comes into play. This has led to a large number of proposed tests for comparing two proportions (or testing the 2 by 2 table of counts). For example, Upton (1982) considers twenty-two alternative tests for comparing two proportions. Sweeping statements about the power of one test over another are impossible to make, because the size of the Type I error depends upon the proportions used. At some proportions, some tests are overly conservative while others are not, while at other proportions the reverse may be true.

This simulation procedure, however, is based primarily on the ordering of the sample statistics in the simulation. The boundaries are determined by the spending function alphas. Thus, if a test used happens to be conservative in the single-look traditional sense, the boundaries chosen in the simulation results of this procedure will generally remove the conservative nature of the test. This makes comparisons to the one-look case surprising in many cases.

Definitions

Suppose you have two populations from which dichotomous (binary) responses will be recorded. The probability (or risk) of obtaining the event of interest in population 1 (the treatment group) is p_1 and in population 2 (the control group) is p_2. The corresponding failure proportions are given by $q_1 = 1 - p_1$ and $q_2 = 1 - p_2$.

The assumption is made that the responses from each group follow a binomial distribution. This means that the event probability, p_i, is the same for all subjects within the group and that the response from one subject is independent of that of any other subject.

Random samples of *m* and *n* individuals are obtained from these two populations. The data from these samples can be displayed in a 2-by-2 contingency table as follows

Group	Success	Failure	Total
Treatment	*a*	*c*	*m*
Control	*b*	*d*	*n*
Total	*s*	*f*	*N*

The following alternative notation is also used.

Group	Success	Failure	Total
Treatment	x_{11}	x_{12}	n_1
Control	x_{21}	x_{22}	n_2
Total	m_1	m_2	N

The binomial proportions p_1 and p_2 are estimated from these data using the formulae

$$\hat{p}_1 = \frac{a}{m} = \frac{x_{11}}{n_1} \text{ and } \hat{p}_2 = \frac{b}{n} = \frac{x_{21}}{n_2}$$

Comparing Two Proportions

Let $p_{1.0}$ represent the group 1 proportion tested by the null hypothesis, H_0. The power of a test is computed at a specific value of the proportion which we will call $p_{1.1}$. Let δ represent the smallest difference (margin of equivalence) between the two proportions that still results in the conclusion that the new treatment is not inferior to the current treatment. For a superiority test, $\delta > 0$**.** The set of statistical hypotheses that are tested is

$$\mathrm{H}_0: p_{1.0} - p_2 \le \delta \text{ versus } \mathrm{H}_1: p_{1.0} - p_2 > \delta$$

which can be rearranged to give

$$\mathrm{H}_0: p_{1.0} \le p_2 + \delta \text{ versus } \mathrm{H}_1: p_{1.0} > p_2 + \delta$$

There are multiple methods of specifying the margin of superiority. The most direct is to simply give values for p_2 and $p_{1.0}$. However, it is often more meaningful to give p_2 and then specify $p_{1.0}$ implicitly by specifying the difference, ratio, or odds ratio. Mathematically, assuming higher proportions are better, the definitions of these parameterizations are

Parameter	**Computation**	**Hypotheses**
Difference	$\delta = p_{1.0} - p_2$	$\mathrm{H}_0: p_{1.0} - p_2 \le \delta_0$ vs. $\mathrm{H}_1: p_{1.0} - p_2 > \delta_0, \quad \delta_0 < 0$
Ratio	$\phi = p_{1.0} / p_2$	$\mathrm{H}_0: p_1 / p_2 \le \phi_0$ vs. $\mathrm{H}_1: p_1 / p_2 > \phi_0, \quad \phi_0 < 1$
Odds Ratio	$\psi = Odds_{1.0} / Odds_2$	$\mathrm{H}_0: o_{1.0} / o_2 \le \psi_0$ versus $\mathrm{H}_1: o_{1.0} / o_2 > \psi_0, \quad \psi_0 < 1$

Difference

The difference is perhaps the most direct method of comparison between two proportions. It is easy to interpret and communicate. It gives the absolute impact of the treatment. However, there are subtle difficulties that can arise with its interpretation.

One difficulty arises when the event of interest is rare. If a difference of 0.001 occurs when the baseline probability is 0.40, it would be dismissed as being trivial. However, if the baseline probably of a disease is 0.002, a 0.001 decrease would represent a reduction of 50%. Thus interpretation of the difference depends on the baseline probability of the event.

The following example is used to convey the concept of a *superiority* test. Suppose 60% of patients respond to the current treatment method $(p_2 = 0.60)$. If the response rate of the new treatment is no less than 5 percentage points better $(\delta = 0.05)$ than the existing treatment, it will be considered to be superior. Substituting these figures into the statistical hypotheses gives

$$H_0 : \delta \leq 0.05 \text{ versus } H_1 : \delta > 0.05$$

In this example, when the null hypothesis is rejected, the concluded alternative is that the response rate is at least 65%, which means that the new treatment is superior to the current treatment.

Ratio

The ratio, $\phi = p_{1.0} / p_2$, gives the relative change in the probability of the response. Testing superiority uses the formulation

$$H_0: p_{1.0} / p_2 \leq \phi_0 \text{ versus } H_1: p_{1.0} / p_2 > \phi_0$$

when higher proportions are better.

The following example is used to convey the concept of *superiority* as defined by the ratio. Suppose that 60% of patients $(p_2 = 0.60)$ respond to the current treatment method. If a new treatment increases the response rate by no less than 10% $(\phi_0 = 1.10)$, it will be considered non-inferior to the standard treatment. Substituting these figures into the statistical hypotheses gives

$$H_0 : \phi \leq 1.10 \text{ versus } H_1 : \phi > 1.10$$

In this example, when the null hypothesis is rejected, the concluded alternative is that the response rate is at least 66%. That is, the conclusion of superiority is that the new treatment's response rate is no worse than 10% more than that of the standard treatment.

Odds Ratio

The odds ratio, $\psi = \left(p_{1.0} / \left(1 - p_{1.0}\right)\right) / \left(p_2 / \left(1 - p_2\right)\right)$, gives the relative change in the odds of the response. Testing non-inferiority uses the formulation

$$H_0: \psi \leq \psi_0 \text{ versus } H_1: \psi > \psi_0$$

when higher proportions are better.

Test Statistics

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

Z Test (Pooled and Unpooled)

This test statistic was first proposed by Karl Pearson in 1900. Although this test can be expressed as a Chi-Square statistic, it is expressed here as a *z* so that it can be used for one-sided hypothesis testing.

Both *pooled* and *unpooled* versions of this test have been discussed in the statistical literature. The pooling refers to the way in which the standard error is estimated. In the pooled version, the

two proportions are averaged, and only one proportion is used to estimate the standard error. In the unpooled version, the two proportions are used separately.

The formula for the test statistic is

$$z_t = \frac{\hat{p}_1 - \hat{p}_2}{\hat{\sigma}_D}$$

Pooled Version

$$\hat{\sigma}_D = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

$$\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$$

Unpooled Version

$$\hat{\sigma}_D = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

Continuity Correction

Frank Yates is credited with proposing a correction to the Pearson Chi-Square test for the lack of continuity in the binomial distribution. However, the correction was in common use when he proposed it in 1922.

The continuity corrected z-test is

$$z = \frac{\left(\hat{p}_1 - \hat{p}_2\right) + \frac{F}{2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}{\hat{\sigma}_D}$$

where F is -1 for upper-tailed, 1 for lower-tailed, and either -1 or 1 for two-sided hypotheses, depending on whether the numerator difference is positive or negative.

T-Test

Based on a study of the behavior of several tests, D'Agostino (1988) and Upton (1982) proposed using the usual two-sample *t*-test for testing whether two proportions are equal. One substitutes a '1' for a success and a '0' for a failure in the usual, two-sample *t*-test formula. The test statistic is computed as

$$t_{N-2} = (ad - bc)\left(\frac{N-2}{N(nac + mbd)}\right)^{\frac{1}{2}}$$

which can be compared to the *t* distribution with N-2 degrees of freedom.

Miettinen and Nurminen's Likelihood Score Test of the Difference

Miettinen and Nurminen (1985) proposed a test statistic for testing whether the difference is equal to a specified, non-zero, value, δ_0. The regular MLE's, $\hat{p}_1$ and $\hat{p}_2$, are used in the numerator of the score statistic while MLE's $\tilde{p}_1$ and $\tilde{p}_2$, constrained so that $\tilde{p}_1 - \tilde{p}_2 = \delta_0$, are used in the

denominator. A correction factor of $N/(N\text{-}1)$ is applied to make the variance estimate less biased. The significance level of the test statistic is based on the asymptotic normality of the score statistic. The formula for computing this test statistic is

$$z_{MND} = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0}{\hat{\sigma}_{MND}}$$

where

$$\hat{\sigma}_{MND} = \sqrt{\left(\frac{\tilde{p}_1\tilde{q}_1}{n_1} + \frac{\tilde{p}_2\tilde{q}_2}{n_2}\right)\left(\frac{N}{N-1}\right)}$$

$$\tilde{p}_1 = \tilde{p}_2 + \delta_0$$

$$\tilde{p}_1 = 2B\cos(A) - \frac{L_2}{3L_3}$$

$$A = \frac{1}{3}\left[\pi + \cos^{-1}\left(\frac{C}{B^3}\right)\right]$$

$$B = \text{sign}(C)\sqrt{\frac{L_2^2}{9L_3} - \frac{L_1}{3L_3}}$$

$$C = \frac{L_2^3}{27L_3^3} - \frac{L_1L_2}{6L_3^2} + \frac{L_0}{2L_3}$$

$$L_0 = x_{21}\delta_0(1-\delta_0)$$

$$L_1 = \left[N_2\delta_0 - N - 2x_{21}\right]\delta_0 + M_1$$

$$L_2 = (N + N_2)\delta_0 - N - M_1$$

$$L_3 = N$$

Miettinen and Nurminen's Likelihood Score Test of the Ratio

Miettinen and Nurminen (1985) proposed a test statistic for testing whether the ratio is equal to a specified value ϕ_0. The regular MLE's, $\hat{p}_1$ and $\hat{p}_2$, are used in the numerator of the score statistic while MLE's $\tilde{p}_1$ and $\tilde{p}_2$, constrained so that $\tilde{p}_1 / \tilde{p}_2 = \phi_0$, are used in the denominator. A correction factor of N/(N-1) is applied to make the variance estimate less biased. The significance level of the test statistic is based on the asymptotic normality of the score statistic.

The formula for computing the test statistic is

$$z_{MNR} = \frac{\hat{p}_1 / \hat{p}_2 - \phi_0}{\sqrt{\left(\frac{\tilde{p}_1\tilde{q}_1}{n_1} + \phi_0^2\frac{\tilde{p}_2\tilde{q}_2}{n_2}\right)\left(\frac{N}{N-1}\right)}}$$

where

$$\tilde{p}_1 = \tilde{p}_2\phi_0$$

$$\tilde{p}_2 = \frac{-B - \sqrt{B^2 - 4AC}}{2A}$$

$$A = N\phi_0$$

$$B = -\left[N_1\phi_0 + x_{11} + N_2 + x_{21}\phi_0\right]$$

$$C = M_1$$

Miettinen and Nurminen's Likelihood Score Test of the Odds Ratio

Miettinen and Nurminen (1985) proposed a test statistic for testing whether the odds ratio is equal to a specified value, ψ_0. Because the approach they used with the difference and ratio does not easily extend to the odds ratio, they used a score statistic approach for the odds ratio. The regular MLE's are $\hat{p}_1$ and $\hat{p}_2$. The constrained MLE's are $\tilde{p}_1$ and $\tilde{p}_2$. These estimates are constrained so that $\tilde{\psi} = \psi_0$. A correction factor of N/(N-1) is applied to make the variance estimate less biased. The significance level of the test statistic is based on the asymptotic normality of the score statistic.

The formula for computing the test statistic is

$$z_{MNO} = \frac{\frac{\left(\hat{p}_1 - \tilde{p}_1\right)}{\tilde{p}_1\tilde{q}_1} - \frac{\left(\hat{p}_2 - \tilde{p}_2\right)}{\tilde{p}_2\tilde{q}_2}}{\sqrt{\left(\frac{1}{N_2\tilde{p}_1\tilde{q}_1} + \frac{1}{N_2\tilde{p}_2\tilde{q}_2}\right)\left(\frac{N}{N-1}\right)}}$$

where

$$\tilde{p}_1 = \frac{\tilde{p}_2\psi_0}{1 + \tilde{p}_2\left(\psi_0 - 1\right)}$$

$$\tilde{p}_2 = \frac{-B + \sqrt{B^2 - 4AC}}{2A}$$

$$A = N_2\left(\psi_0 - 1\right)$$

$$B = N_1\psi_0 + N_2 - M_1\left(\psi_0 - 1\right)$$

$$C = -M_1$$

Farrington and Manning's Likelihood Score Test of the Difference

Farrington and Manning (1990) proposed a test statistic for testing whether the difference is equal to a specified value δ_0. The regular MLE's, $\hat{p}_1$ and $\hat{p}_2$, are used in the numerator of the score statistic while MLE's $\tilde{p}_1$ and $\tilde{p}_2$, constrained so that $\tilde{p}_1 - \tilde{p}_2 = \delta_0$, are used in the denominator. The significance level of the test statistic is based on the asymptotic normality of the score statistic.

The formula for computing the test statistic is

$$z_{FMD} = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0}{\sqrt{\left(\frac{\tilde{p}_1\tilde{q}_1}{n_1} + \frac{\tilde{p}_2\tilde{q}_2}{n_2}\right)}}$$

where the estimates $\tilde{p}_1$ and $\tilde{p}_2$ are computed as in the corresponding test of Miettinen and Nurminen (1985) given above.

Farrington and Manning's Likelihood Score Test of the Ratio

Farrington and Manning (1990) proposed a test statistic for testing whether the ratio is equal to a specified value ϕ_0. The regular MLE's, $\hat{p}_1$ and $\hat{p}_2$, are used in the numerator of the score statistic while MLE's $\tilde{p}_1$ and $\tilde{p}_2$, constrained so that $\tilde{p}_1 / \tilde{p}_2 = \phi_0$, are used in the denominator. A correction factor of N/(N-1) is applied to increase the variance estimate. The significance level of the test statistic is based on the asymptotic normality of the score statistic.

The formula for computing the test statistic is

$$z_{FMR} = \frac{\hat{p}_1 / \hat{p}_2 - \phi_0}{\sqrt{\left(\frac{\tilde{p}_1\tilde{q}_1}{n_1} + \phi_0^2\frac{\tilde{p}_2\tilde{q}_2}{n_2}\right)}}$$

where the estimates $\tilde{p}_1$ and $\tilde{p}_2$ are computed as in the corresponding test of Miettinen and Nurminen (1985) given above.

Farrington and Manning's Likelihood Score Test of the Odds Ratio

Farrington and Manning (1990) indicate that the Miettinen and Nurminen statistic may be modified by removing the factor $N/(N\text{-}1)$.

The formula for computing this test statistic is

$$z_{FMO} = \frac{\frac{\left(\hat{p}_1 - \tilde{p}_1\right)}{\tilde{p}_1\tilde{q}_1} - \frac{\left(\hat{p}_2 - \tilde{p}_2\right)}{\tilde{p}_2\tilde{q}_2}}{\sqrt{\left(\frac{1}{N_2\tilde{p}_1\tilde{q}_1} + \frac{1}{N_2\tilde{p}_2\tilde{q}_2}\right)}}$$

where the estimates $\tilde{p}_1$ and $\tilde{p}_2$ are computed as in the corresponding test of Miettinen and Nurminen (1985) given above.

Gart and Nam's Likelihood Score Test of the Difference

Gart and Nam (1990), page 638, proposed a modification to the Farrington and Manning (1988) difference test that corrects for skewness. Let $z_{FMD}(\delta)$ stand for the Farrington and Manning difference test statistic described above. The skewness corrected test statistic, z_{GND}, is the appropriate solution to the quadratic equation

$$\left(-\tilde{\gamma}\right)z_{GND}^2 + \left(-1\right)z_{GND} + \left(z_{FMD}(\delta) + \tilde{\gamma}\right) = 0$$

where

$$\tilde{\gamma} = \frac{\tilde{V}^{3/2}(\delta)}{6}\left(\frac{\tilde{p}_1\tilde{q}_1(\tilde{q}_1 - \tilde{p}_1)}{n_1^2} - \frac{\tilde{p}_2\tilde{q}_2(\tilde{q}_2 - \tilde{p}_2)}{n_2^2}\right)$$

Gart and Nam's Likelihood Score Test of the Ratio

Gart and Nam (1988), page 329, proposed a modification to the Farrington and Manning (1988) ratio test that corrects for skewness. Let $z_{FMR}(\phi)$ stand for the Farrington and Manning ratio test statistic described above. The skewness corrected test statistic, z_{GNR}, is the appropriate solution to the quadratic equation

$$(-\tilde{\varphi})z_{GNR}^2 + (-1)z_{GNR} + (z_{FMR}(\phi) + \tilde{\varphi}) = 0$$

where

$$\tilde{\varphi} = \frac{1}{6\tilde{u}^{3/2}}\left(\frac{\tilde{q}_1(\tilde{q}_1 - \tilde{p}_1)}{n_1^2\tilde{p}_1^2} - \frac{\tilde{q}_2(\tilde{q}_2 - \tilde{p}_2)}{n_2^2\tilde{p}_2^2}\right)$$

$$\tilde{u} = \frac{\tilde{q}_1}{n_1\tilde{p}_1} + \frac{\tilde{q}_2}{n_2\tilde{p}_2}$$

Spending Functions

Spending functions can be used in this procedure to specify the proportion of alpha or beta that is spent at each look without having to specify the proportion directly.

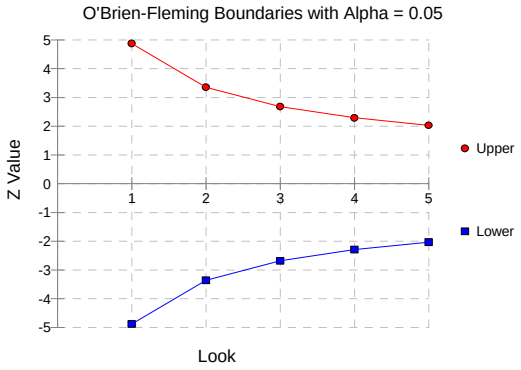
Spending functions have the characteristics that they are increasing and that

$$\alpha(0) = 0$$
$$\alpha(1) = \alpha$$

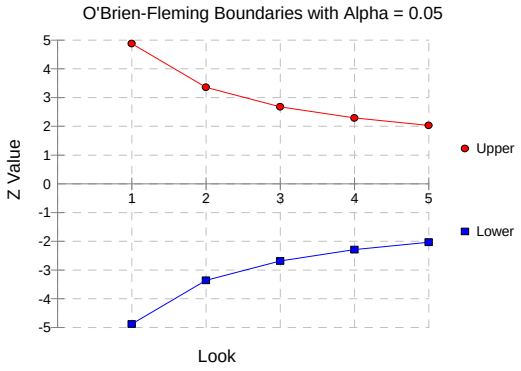
The last characteristic guarantees a fixed α level when the trial is complete. This methodology is very flexible since neither the times nor the number of analyses must be specified in advance. Only the functional form of $\alpha(\tau)$ must be specified.

PASS provides several popular spending functions plus the ability to enter and analyze your own percents of alpha or beta spent. These are calculated as follows (beta may be substituted for alpha for beta-spending functions):

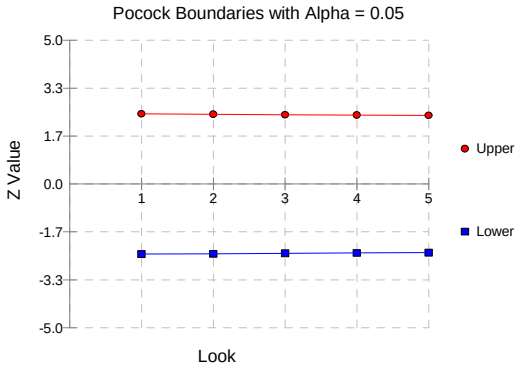
1. Hwang-Shih-DeCani (gamma family) $\alpha\left[\frac{1-e^{-\gamma t}}{1-e^{-\gamma}}\right], \gamma \neq 0\,; \quad \alpha t, \gamma = 0$



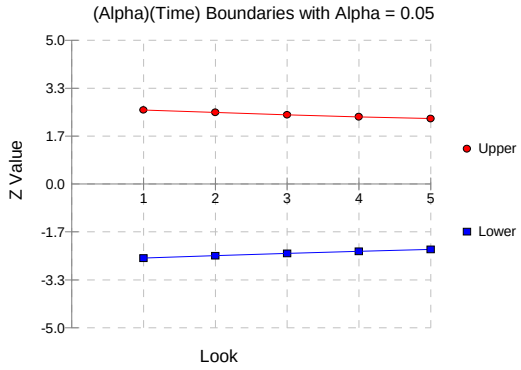
2. O'Brien-Fleming Analog $2 - 2\Phi\left(\frac{Z_{\alpha/2}}{\sqrt{t}}\right)$



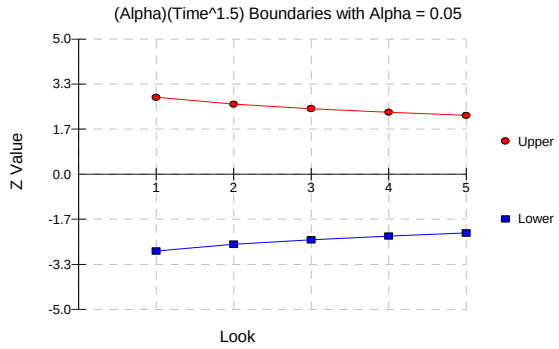
3. Pocock Analog $\alpha \cdot \ln\left(1 + (e-1)t\right)$



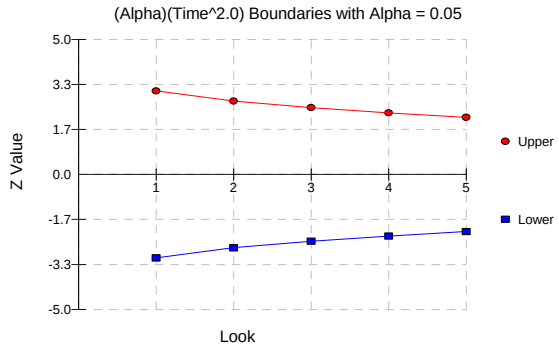
4. Alpha * time $\alpha \cdot t$



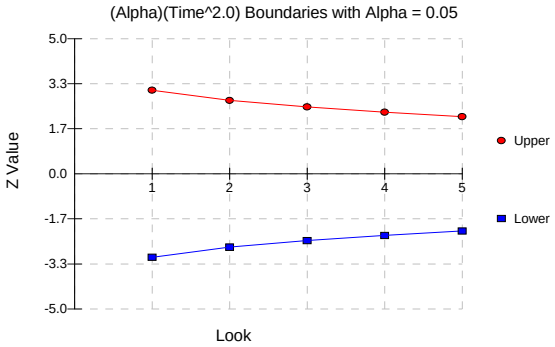
5. Alpha * time^1.5 $\alpha \cdot t^{3/2}$



6. Alpha * time^2 $\alpha \cdot t^2$



7. Alpha * time^C $\alpha \cdot t^C$



8. User Supplied Percents

A custom set of percents of alpha to be spent at each look may be input directly.

The O'Brien-Fleming Analog spends very little alpha or beta at the beginning and much more at the final looks. The Pocock Analog and (Alpha or Beta)(Time) spending functions spend alpha or beta more evenly across the looks. The Hwang-Shih-DeCani (C) (gamma family) spending functions and (Alpha or Beta)(Time^C) spending functions are flexible spending functions that can be used to spend more alpha or beta early or late or evenly, depending on the choice of C.

Procedure Options

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

Data Tab

The Data tab contains most of the parameters and options for the general setup of the procedure.

Solve For

Find (Solve For)

Solve for either power, sample size, or enter the boundaries directly and solve for power and alpha. When solving for power or sample size, the look and boundary details are specified on the "Looks & Boundaries" tab and the "Enter Boundaries" tab is ignored. When entering the boundaries directly and solving for power and alpha, the boundaries are input on the "Enter Boundaries" tab and the "Looks & Boundaries" tab is ignored.

When solving for power or N1, the early-stopping boundaries are also calculated. High accuracy for early-stopping boundaries requires a very large number of simulations (Recommended 100,000 to 10,000,000).

The parameter selected here is the parameter displayed on the vertical axis of the plot.

Because this is a simulation based procedure, the search for the sample size may take several minutes or hours to complete. You may find it quicker and more informative to solve for Power for a range of sample sizes.

Error Rates

Power or Beta

Power is the probability of rejecting the null hypothesis when it is false. Power is equal to 1-Beta, so specifying power implicitly specifies beta.

Beta is the probability obtaining a false negative on the statistical test. That is, it is the probability of accepting a false null hypothesis.

In the context of simulated group sequential trials, the power is the proportion of the alternative hypothesis simulations that cross any one of the significance (efficacy) boundaries.

The valid range is between 0 to 1.

Different disciplines and protocols have different standards for setting power. A common choice is 0.90, but 0.80 is also popular.

You can enter a range of values such as *0.70 0.80 0.90* or *0.70 to 0.90 by 0.1*.

Alpha (Significance Level)

Alpha is the probability of obtaining a false positive on the statistical test. That is, it is the probability of rejecting a true null hypothesis.

The null hypothesis is usually that the parameters (the means, proportions, etc.) are all equal.

In the context of simulated group sequential trials, alpha is the proportion of the null hypothesis simulations that cross any one of the significance (efficacy) boundaries.

Since Alpha is a probability, it is bounded by 0 and 1. Commonly, it is between 0.001 and 0.250.

Alpha is often set to 0.05 for two-sided tests and to 0.025 for one-sided tests.

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

Sample Size

N1 (Sample Size Group 1)

Enter a value for the sample size, N1. This is the number of subjects in the first group of the study at the final look.

You may enter a range such as 10 to 100 by 10 or a list of values separated by commas or blanks.

You might try entering the same number two or three times to get an idea of the variability in your results. For example, you could enter "10 10 10".

N2 (Sample Size Group 2)

Enter a value (or range of values) for the sample size of group 2. This is the number of subjects in the second group of the study at the final look.

- **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 the operator [*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*.

R = N2/N1

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*.

Effect Size

P1.0 (Treatment Proportion|H0) – Proportions

Specify the value P1.0 directly.

When higher proportions are "Better", P1.0 is the treatment group proportion such that proportions greater than P1.0 are considered superior to the reference group proportion. When higher proportions are "Better", P1.0 should be greater than P2.

When higher proportions are "Worse", P1.0 is the treatment group proportion such that proportions less than P1.0 are considered superior to the reference group proportion. When higher proportions are "Worse", P1.0 should be less than P2.

When P1.0 is close to zero or one, a Zero Count Adjustment may be needed.

You can enter a list of values such as *0.4 0.5 0.6* or *0.3 to 0.7 by 0.05*.

P1.1 (Treatment Proportion|H1) – Proportions

This is the value of the treatment proportion (P1) at which the power calculations are made.

When higher proportions are "Better", P1.1 should be greater than P1.0.

When higher proportions are "Worse", P1.1 should be less than P1.0.

When P1.1 is close to zero or one, a Zero Count Adjustment may be needed.

You can enter a list of values such as *0.4 0.5 0.6* or *0.3 to 0.7 by 0.05*.

D0 (Difference|H0 = P1.0 – P2) – Differences

Specify the difference between P1.0 and P2 that will be considered superior.

When higher proportions are "Better", the superiority difference is that amount that P1 need be greater than P2 to be declared superior to the reference group. When higher proportions are "Better", D0 should be positive.

When higher proportions are "Worse", the superiority difference is that amount that P1 need be less than P2 to be declared superior to the reference group. When higher proportions are "Worse", D0 should be negative.

The power calculations assume that P1.0 is the value of the P1 under the null hypothesis. This value is used with P2 to calculate the value of P1.0 using the formula: P1.0 = D0 + P2

You may enter a range of values such as *-.03 -.05 -.10* or *-.05 to -.01 by .01*.

Differences must be between -1 and 1. D0 cannot take on the values -1, 0, or 1. D0 should be selected such that P1.0 is between zero and one.

D1 (Difference|H1 = P1.1 – P2) – Differences

Specify the actual difference between P1.1 (the actual value of P1) and P2. This is the value at which the power is calculated.

When higher proportions are "Better", D1 should be positive and greater than D0.

When higher proportions are "Worse", D1 should be negative and more extreme than D0.

The power calculations assume that P1.1 is the actual value of the proportion in group 1 (experimental or treatment group). This difference is used with P2 to calculate the value of P1 using the formula: P1.1 = D1 + P2

You may enter a range of values such as -*.05 0 .5* or -*.05 to .05 by .02*.

Actual differences must be between -1 and 1. They cannot take on the values -1 or 1.

R0 (Ratio|H0 = P1.0 / P2) – Ratios

Specify the superiority ratio of P1.0 to P2.

When higher proportions are "Better", the superiority ratio is smallest ratio of P1 to P2 for which P1 will still be considered superior. When higher proportions are "Better", R0 should be greater than one.

When higher proportions are "Worse", the superiority ratio is largest ratio of P1 to P2 for which P1 will still be considered superior. When higher proportions are "Worse", R0 should be less than one.

The power calculations assume that P1.0 is the value of P1 under the null hypothesis. This value is used with P2 to calculate the value of P1.0 using the formula: P1.0 = R0 x P2

You may enter a range of values such as *.95 .97 .99* or *.91 to .99 by .02*.

Ratios must be positive. R0 cannot take on the value of one.

R1 (Ratio |H1 = P1.1 / P2) – Ratios

Specify the actual ratio between P1.1 (the actual value of P1) and P2. This is the value at which the power is calculated.

When higher proportions are "Better", R1 should be greater than R0.

When higher proportions are "Worse", R1 should be less than R0.

The power calculations assume that P1.1 is the actual value of the proportion in group 1 (experimental or treatment group). This ratio is used with P2 to calculate the value of P1 using the formula: P1.1 = R1 x P2

You may enter a range of values such as *.95 1 1.05* or *.9 to 1 by .02*.

Ratios must be positive.

OR0 (Odds Ratio|H0 = O1.0 / O2) – Odds Ratios

Specify the superiority odds ratio of O1.0 to O2.

When higher proportions are "Better", the superiority odds ratio is smallest odds ratio of P1 to P2 for which P1 will still be considered superior. When higher proportions are "Better", OR0 should be greater than one.

When higher proportions are "Worse", the superiority odds ratio is largest odds ratio of P1 to P2 for which P1 will still be considered superior. When higher proportions are "Worse", OR0 should be less than one.

The power calculations assume that P1.0 is the value of P1 under the null hypothesis. OR0 is used with P2 to calculate the value of P1.0.

You may enter a range of values such as *.95 .97 .99* or *.91 to .99 by .02*.

Odds ratios must be positive. OR0 cannot take on the value of one.

OR1 (Odds Ratio|H1 = O1.1 / O2) – Odds Ratios

Specify the actual odds ratio between P1.1 (the actual value of P1) and P2. This is the value at which the power is calculated.

When higher proportions are "Better", OR1 should be greater than OR0.

When higher proportions are "Worse", OR1 should be less than OR0.

The power calculations assume that P1.1 is the actual value of the proportion in group 1 (experimental or treatment group). This odds ratio is used with P2 to calculate the value of P1.1 using the formula: P1.1 = R1 x P2

You may enter a range of values such as.95 1 1.05 or .9 to 1 by .02.

Ratios must be positive.

P2 (Control Group Proportion)

Enter a value for the proportion in the control (baseline, standard, or reference) group, P2.

When P2 is close to zero or one, a Zero Count Adjustment may be needed.

Values must be between 0 and 1.

You can enter a list of values such as *0.4 0.5 0.6* or *0.3 to 0.7 by 0.05*.

Test and Simulations

Higher Proportions Are

Use this option to specify the direction of the test.

If Higher Proportions are "Better", the alternative hypothesis is H1: P1 - P2 > Superiority Difference, H1: P1 - P2 > D0, H1: P1 / P2 > R0, or H1: O1 / O2 > OR0.

If Higher Proportions are "Worse", the alternative hypothesis is H1: P1 - P2 < Superiority Difference, H1: P1 - P2 < D0, H1: P1 / P2 < R0, or H1: O1 / O2 < OR0.

Test Type

Specify which test statistic is to be simulated and reported on.

For details and formulation of the tests, see the section in this manual above.

Simulations

Specify the number of Monte Carlo iterations, M.

The following table gives an estimate of the precision that is achieved for various simulation sizes when the power is near 0.50 and 0.95. The values in the table are the "Precision" amounts that are added and subtracted to form a 95% confidence interval.

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

However, when solving for Power or N1, the simulations are used to calculate the look boundaries. To obtain precise boundary estimates, the number of simulations needs to be high. However, this consideration competes with the length of time to complete the simulation. When solving for power, a large number of simulations (100,000 or 1,000,000) will finish in several minutes. When solving for N1, perhaps 10,000 simulations can be run for each iteration. Then, a final run with the resulting N1 solving for power can be run with more simulations.

Looks & Boundaries Tab

The Data tab contains most of the parameters and options for the general setup of the procedure.

Looks and Boundaries

Specification of Looks and Boundaries

Choose whether spending functions will be used to divide alpha and beta for each look (Simple Specification), or whether the percents of alpha and beta to be spent at each look will be specified directly (Custom Specification).

Under Simple Specification, the looks are automatically considered to be equally spaced. Under Custom Specification, the looks may be equally spaced or custom defined based on the percent of accumulated information.

Looks and Boundaries – Simple Specification

Number of Equally Spaced Looks

Select the total number of looks that will be used if the study is not stopped early for the crossing of a boundary.

Alpha Spending Function

Specify the type of alpha spending function to use.

The O'Brien-Fleming Analog spends very little alpha at the beginning and much more at the final looks. The Pocock Analog and (Alpha)(Time) spending functions spend alpha more evenly across the looks. The Hwang-Shih-DeCani (C) (sometimes called the gamma family) spending functions and (Alpha)(Time^C) spending functions are flexible spending functions that can be used to spend more alpha early or late or evenly, depending on the choice of C.

C (Alpha Spending)

C is used to define the Hwang-Shih-DeCani (C) or (Alpha)(Time^C) spending functions.

For the Hwang-Shih-DeCani (C) spending function, negative values of C spend more of alpha at later looks, values near 0 spend alpha evenly, and positive values of C spend more of alpha at earlier looks.

For the (Alpha)(Time^C) spending function, only positive values for C are permitted. Values of C near zero spend more of alpha at earlier looks, values near 1 spend alpha evenly, and larger values of C spend more of alpha at later looks.

Type of Futility Boundary

This option determines whether or not futility boundaries will be created, and if so, whether they are binding or non-binding.

Futility boundaries are boundaries such that, if crossed at a given look, stop the study in favor of H0.

Binding futility boundaries are computed in concert with significance boundaries. They are called binding because they require the stopping of a trial if they are crossed. If the trial is not stopped, the probability of a false positive will exceed alpha.

When non-binding futility boundaries are computed, the significance boundaries are first computed, ignoring the futility boundaries. The futility boundaries are then computed. These futility boundaries are non-binding because continuing the trial after they are crossed will not affect the overall probability of a false positive declaration.

Number of Skipped Futility Looks

In some trials it may be desirable to wait a number of looks before examining the trial for futility. This option allows the beta to begin being spent after a specified number of looks.

The Number of Skipped Futility Looks should be less than the number of looks.

Beta Spending Function

Specify the type of beta spending function to use.

The O'Brien-Fleming Analog spends very little beta at the beginning and much more at the final looks. The Pocock Analog and (Beta)(Time) spending functions spend beta more evenly across the looks. The Hwang-Shih-DeCani (C) (sometimes called the gamma family) spending functions and (Beta)(Time^C) spending functions are flexible spending functions that can be used to spend more beta early or late or evenly, depending on the choice of C.

C (Beta Spending)

C is used to define the Hwang-Shih-DeCani (C) or (Beta)(Time^C) spending functions.

For the Hwang-Shih-DeCani (C) spending function, negative values of C spend more of beta at later looks, values near 0 spend beta evenly, and positive values of C spend more of beta at earlier looks.

For the (Beta)(Time^C) spending function, only positive values for C are permitted. Values of C near zero spend more of beta at earlier looks, values near 1 spend beta evenly, and larger values of C spend more of beta at later looks.

Looks and Boundaries – Custom Specification

Number of Looks

This is the total number of looks of either type (significance or futility or both).

Equally Spaced

If this box is checked, the Accumulated Information boxes are ignored and the accumulated information is evenly spaced.

Type of Futility Boundary

This option determines whether or not futility boundaries will be created, and if so, whether they are binding or non-binding.

Futility boundaries are boundaries such that, if crossed at a given look, stop the study in favor of H0.

Binding futility boundaries are computed in concert with significance boundaries. They are called binding because they require the stopping of a trial if they are crossed. If the trial is not stopped, the probability of a false positive will exceed alpha.

When Non-binding futility boundaries are computed, the significance boundaries are first computed, ignoring the futility boundaries. The futility boundaries are then computed. These futility boundaries are non-binding because continuing the trial after they are crossed will not affect the overall probability of a false positive declaration.

Accumulated Information

The accumulated information at each look defines the proportion or percent of the sample size that is used at that look.

These values are accumulated information values so they must be increasing.

Proportions, percents, or sample sizes may be entered. All proportions, percents, or sample sizes will be divided by the value at the final look to create an accumulated information proportion for each look.

Percent of Alpha Spent

This is the percent of the total alpha that is spent at the corresponding look. It is not the cumulative value.

Percents, proportions, or alphas may be entered here. Each of the values is divided by the sum of the values to obtain the proportion of alpha that is used at the corresponding look.

Percent of Beta Spent

This is the percent of the total beta (1-power) that is spent at the corresponding look. It is not the cumulative value.

Percents, proportions, or betas may be entered here. Each of the values is divided by the sum of the values to obtain the proportion of beta that is used at the corresponding look.

Enter Boundaries Tab

The Data tab contains most of the parameters and options for the general setup of the procedure.

Looks and Boundaries

Number of Looks

This is the total number of looks of either type (significance or futility or both).

Equally Spaced

If this box is checked, the Accumulated Information boxes are ignored and the accumulated information is evenly spaced.

Types of Boundaries

This option determines whether or not futility boundaries will be entered.

Futility boundaries are boundaries such that, if crossed at a given look, stop the study in favor of H0.

Accumulated Information

The accumulated information at each look defines the proportion or percent of the sample size that is used at that look.

These values are accumulated information values so they must be increasing.

Proportions, percents, or sample sizes may be entered. All proportions, percents, or sample sizes will be divided by the value at the final look to create an accumulated information proportion for each look.

Significance Boundary

Enter the value of the significance boundary corresponding to the chosen test statistic. These are sometimes called efficacy boundaries.

Futility Boundary

Enter the value of the futility boundary corresponding to the chosen test statistic.

Options Tab

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

Maximum Iterations

Maximum N1 Before Search Termination

Specify the maximum N1 before the search for N1 is aborted.

Since simulations for large sample sizes are very computationally intensive and hence time-consuming, this value can be used to stop searches when N1 is larger than reasonable sample sizes for the study.

This applies only when "Find (Solve For)" is set to N1.

The procedure uses a binary search when searching for N1. If a value for N1 is tried that exceeds this value, and the power is not reached, a warning message will be shown on the output indicating the desired power was not reached.

We recommend a value of at least 20000.

Matching Boundaries at Final Look

Beta Search Increment

For each simulation, when futility bounds are computed, the appropriate beta is found by searching from 0 to 1 by this increment. Smaller increments are more refined, but the search takes longer.

We recommend 0.001 or 0.0001.

Zero Count Adjustment

Zero Count Adjustment Method

Zero cell counts cause many calculation problems. To compensate for this, a small value (called the Zero Count Adjustment Value) may be added either to all cells or to all cells with zero counts. This option specifies whether you want to use the adjustment and which type of adjustment you want to use.

Adding a small value is controversial, but may be necessary. Some statisticians recommend adding 0.5 while others recommend 0.25. We have found that adding values as small as 0.0001 seems to work well.

Zero Count Adjustment Value

Zero cell counts cause many calculation problems. To compensate for this, a small value may be added either to all cells or to all zero cells. This is the amount that is added.

Some statisticians recommend that the value of 0.5 be added to all cells (both zero and non-zero). Others recommend 0.25. We have found that even a value as small as 0.0001 works well.

Example 1 – Power and Output

A clinical trial is to be conducted over a two-year period to compare the proportion response of a new treatment to that of the current treatment. The current response proportion is 0.58. The researchers would like to show that the new treatment is at least 0.05 better than the standard treatment. Although the researchers do not know the true proportion of patients that will survive with the new treatment, they would like to examine the power that is achieved if the proportion under the new treatment is 0.68. The sample size at the final look is to be 1000 per group. Testing will be done at the 0.05 significance level. A total of five tests are going to be performed on the data as they are obtained. The O'Brien-Fleming (Analog) boundaries will be used. .

Find the power and test boundaries assuming equal sample sizes per arm.

Setup

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

Option	Value
Data Tab	
Find (Solve For)	**Power**
Power	*Ignored since this is the Find setting*
Alpha	**0.05**
N1 (Sample Size Group 1)	**1000**
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
P1.0 (Treatment Proportion|H0)	**0.58**
P1.1 (Treatment Proportion|H1)	**0.63**
P2 (Proportion in Group 2)	**0.68**
Higher Proportions Are	**Better**
Test Type	**Z-Test (Pooled)**
Simulations	**100000**
Looks and Boundaries Tab	
Specification of Looks and Boundaries	**Simple**
Number of Equally Spaced Looks	**5**
Alpha Spending Function	**O'Brien-Fleming Analog**

Output

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

Numeric Results and Plots

Scenario 1 Numeric Results for Group Sequential Testing Proportion Difference = 0.
Hypotheses: H0: Proportion 1 - Proportion 2 = D0; H1: Proportion 1 - Proportion 2 > D0
Test Statistic: Z-Test (Pooled)
Zero Adjustment Method: None
Alpha-Spending Function: O'Brien-Fleming Analog
Beta-Spending Function: None
Futility Boundary Type: None
Number of Looks: 5
Simulations: 100000

Numeric Summary for Scenario 1

Power			Alpha				
Value	**95% LCL**	**95% UCL**	**Target**	**Actual**	**95% LCL**	**95% UCL**	**Beta**
0.738	0.736	0.741	0.050	0.050	0.048	0.051	0.262

		Average Sample Size								
		Given H0		Given H1						
N1	**N2**	**Grp1**	**Grp2**	**Grp1**	**Grp2**	**D0**	**D1**	**P1.0**	**P1.1**	**P2**
1000	1000	992	992	811	811	0.1	0.1	0.6	0.7	0.6

222-24 Group-Sequential Non-Zero Null Tests for Two Proportions (Simulation)

Report Definitions

Power is the probability of rejecting a false null hypothesis at one of the looks. It is the total proportion of alternative hypothesis simulations that are outside the significance boundaries.
Power 95% LCL and UCL are the lower and upper confidence limits for the power estimate. The width of the interval is based on the number of simulations.
Target Alpha is the user-specified probability of rejecting a true null hypothesis. It is the total alpha spent.
Alpha or Actual Alpha is the alpha level that was actually achieved by the experiment. It is the total proportion of the null hypothesis simulations that are outside the significance boundaries.
Alpha 95% LCL and UCL are the lower and upper confidence limits for the actual alpha estimate. The width of the interval is based on the number of simulations.
Beta is the probability of accepting a false null hypothesis. It is the total proportion of alternative hypothesis simulations that do not cross the significance boundaries.
N1 and N2 are the sample sizes of each group if the study reaches the final look.
Average Sample Size Given H0 Grp1 and Grp2 are the average or expected sample sizes of each group if H0 is true. These are based on the proportion of null hypothesis simulations that cross the significance or futility boundaries at each look.
Average Sample Size Given H1 Grp1 and Grp2 are the average or expected sample sizes of each group if H1 is true. These are based on the proportion of alternative hypothesis simulations that cross the significance or futility boundaries at each look.
D0, or superiority difference, is the proportion difference between groups (Grp1 - Grp2) assuming the null hypothesis, H0.
D1 is the proportion difference between groups (Grp1 - Grp2) assuming the alternative hypothesis, H1.
P1.0 is the proportion used in the simulations for Group 1 under H0.
P1.1 is the proportion used in the simulations for Group 1 under H1.
P2 is the proportion used in the simulations for Group 2 under H0 and H1.

Summary Statements

Group sequential trials with sample sizes of 1000 and 1000 at the final look achieve 74% power to detect a difference of 0.1 between a treatment group proportion of 0.7 and a control group proportion of 0.6 with a superiority difference of 0.1 at the 0.050 significance level (alpha) using a one-sided Z-Test (Pooled).

Accumulated Information Details for Scenario 1

	Accumulated Information	Accumulated Sample Size		
Look	Percent	Group 1	Group 2	Total
1	20.0	200	200	400
2	40.0	400	400	800
3	60.0	600	600	1200
4	80.0	800	800	1600
5	100.0	1000	1000	2000

Accumulated Information Details Definitions

Look is the number of the look.
Accumulated Information Percent is the percent of the sample size accumulated up to the corresponding look.
Accumulated Sample Size Group 1 is total number of individuals in group 1 at the corresponding look.
Accumulated Sample Size Group 2 is total number of individuals in group 2 at the corresponding look.
Accumulated Sample Size Total is total number of individuals in the study (group 1 + group 2) at the corresponding look.

Boundaries for Scenario 1

	Significance Boundary	
	Z-Value	P-Value
Look	Scale	Scale
1	4.120	0.000
2	2.893	0.002
3	2.296	0.011
4	1.950	0.026
5	1.734	0.041

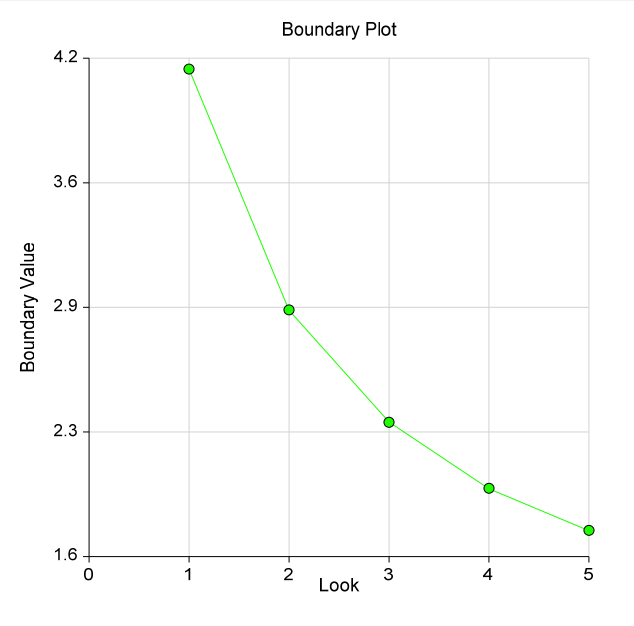
Boundaries Definitions

Look is the number of the look.

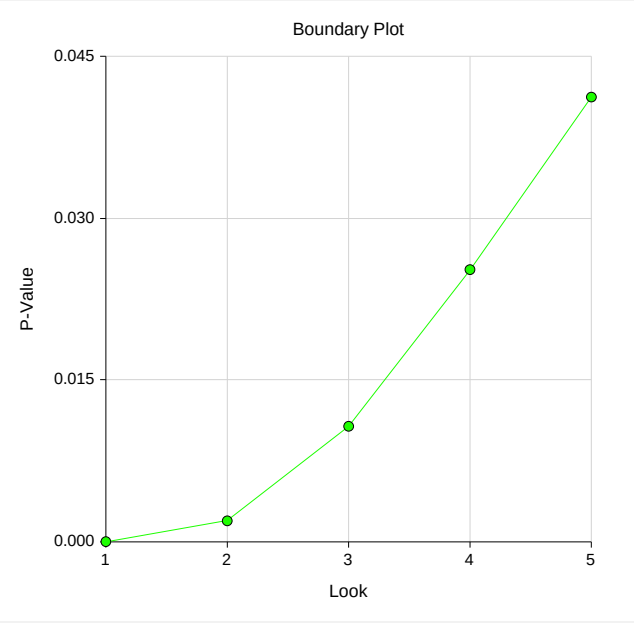
Significance Boundary Z-Value Scale is the value such that statistics outside this boundary at the corresponding look indicate termination of the study and rejection of the null hypothesis. They are sometimes called efficacy boundaries.

Significance Boundary P-Value Scale is the value such that P-Values outside this boundary at the corresponding look indicate termination of the study and rejection of the null hypothesis. This P-Value corresponds to the Z-Value Boundary and is sometimes called the nominal alpha.

Boundary Plot



Boundary Plot - P-Value



Significance Boundaries with 95% Simulation Confidence Intervals for Scenario 1

	Z-Value Boundary			P-Value Boundary		
Look	Value	95% LCL	95% UCL	Value	95% LCL	95% UCL
1	4.120			0.000		
2	2.893	2.858	2.951	0.002	0.002	0.002
3	2.296	2.274	2.307	0.011	0.011	0.011
4	1.950	1.941	1.984	0.026	0.024	0.026
5	1.734	1.725	1.740	0.041	0.041	0.042

Significance Boundary Confidence Limit Definitions

Look is the number of the look.

Z-Value Boundary Value is the value such that statistics outside this boundary at the corresponding look indicate termination of the study and rejection of the null hypothesis. They are sometimes called efficacy boundaries.

P-Value Boundary Value is the value such that P-Values outside this boundary at the corresponding look indicate termination of the study and rejection of the null hypothesis. This P-Value corresponds to the Z-Value Boundary and is sometimes called the nominal alpha.

95% LCL and UCL are the lower and upper confidence limits for the boundary at the given look. The width of the interval is based on the number of simulations.

Alpha-Spending and Null Hypothesis Simulation Details for Scenario 1

	--- Signif. Boundary---		Target		Actual		Proportion	Cum.
				Cum.			H1 Sims	H1 Sims
	Z-Value	P-Value	Spending Function	Spending Function		Cum. Alpha	Outside Signif.	Outside Signif.
Look	Scale	Scale	Alpha	Alpha	Alpha Spent	Spent	Boundary	Boundary
1	4.120	0.000	0.000	0.000	0.000	0.000	0.001	0.001
2	2.893	0.002	0.002	0.002	0.002	0.002	0.075	0.076
3	2.296	0.011	0.009	0.011	0.009	0.011	0.235	0.310
4	1.950	0.026	0.017	0.028	0.017	0.028	0.253	0.563
5	1.734	0.041	0.022	0.050	0.021	0.050	0.176	0.740

Alpha-Spending Details Definitions

Look is the number of the look.

Significance Boundary Z-Value Scale is the value such that statistics outside this boundary at the corresponding look indicate termination of the study and rejection of the null hypothesis. They are sometimes called efficacy boundaries.

Significance Boundary P-Value Scale is the value such that P-Values outside this boundary at the corresponding look indicate termination of the study and rejection of the null hypothesis. This P-Value corresponds to the Significance Z-Value Boundary and is sometimes called the nominal alpha.

Spending Function Alpha is the intended portion of alpha allocated to the particular look based on the alpha-spending function.

Cumulative Spending Function Alpha is the intended accumulated alpha allocated to the particular look. It is the sum of the Spending Function Alpha up to the corresponding look.

Alpha Spent is the proportion of the null hypothesis simulations resulting in statistics outside the Significance Boundary at this look.

Cumulative Alpha Spent is the proportion of the null hypothesis simulations resulting in Significance Boundary termination up to and including this look. It is the sum of the Alpha Spent up to the corresponding look.

Proportion H1 Sims Outside Significance Boundary is the proportion of the alternative hypothesis simulations resulting in statistics outside the Significance Boundary at this look. It may be thought of as the incremental power.

Cumulative H1 Sims Outside Significance Boundary is the proportion of the alternative hypothesis simulations resulting in Significance Boundary termination up to and including this look. It is the sum of the Proportion H1 Sims Outside Significance Boundary up to the corresponding look.

Run Time: 62.05 seconds.

The values obtained from any given run of this example will vary slightly due to the variation in simulations.

Example 2 – Power with Futility Boundaries

Continuing with Example 1, suppose that the researchers would also like to terminate the study early if there is indication that the treatment is not better than the standard by 0.05.

Setup

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

Option	Value
Data Tab	
Find (Solve For)	**Power**
Power	*Ignored since this is the Find setting*
Alpha	**0.05**
N1 (Sample Size Group 1)	**1000**
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
P1.0 (Treatment Proportion|H0)	**0.63**
P1.1 (Treatment Proportion|H1)	**0.68**
P2 (Proportion in Group 2)	**0.58**
Higher Proportions Are	**Better**
Test Type	**Z-Test (Pooled)**
Simulations	**100000**
Looks and Boundaries Tab	
Specification of Looks and Boundaries	**Simple**
Number of Equally Spaced Looks	**5**
Alpha Spending Function	**O'Brien-Fleming Analog**
Type of Futility Boundary	**Non-Binding**
Number of Skipped Futility Looks	**0**
Beta Spending Function	**O'Brien-Fleming Analog**

Output

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

Numeric Results and Plots

Scenario 1 Numeric Results for Group Sequential Testing Proportion Difference = 0.
Hypotheses: H0: Proportion 1 - Proportion 2 = D0; H1: Proportion 1 - Proportion 2 > D0
Test Statistic: Z-Test (Pooled)
Zero Adjustment Method: None
Alpha-Spending Function: O'Brien-Fleming Analog
Beta-Spending Function: None
Futility Boundary Type: None
Number of Looks: 5
Simulations: 100000

Numeric Summary for Scenario 1

Power			Alpha				
Value	95% LCL	95% UCL	Target	Actual	95% LCL	95% UCL	Beta
0.654	0.651	0.657	0.050	0.039	0.038	0.040	0.346

		Average Sample Size								
		Given H0		Given H1						
N1	N2	Grp1	Grp2	Grp1	Grp2	D0	D1	P1.0	P1.1	P2
1000	1000	464	464	676	676	0.1	0.1	0.6	0.7	0.6

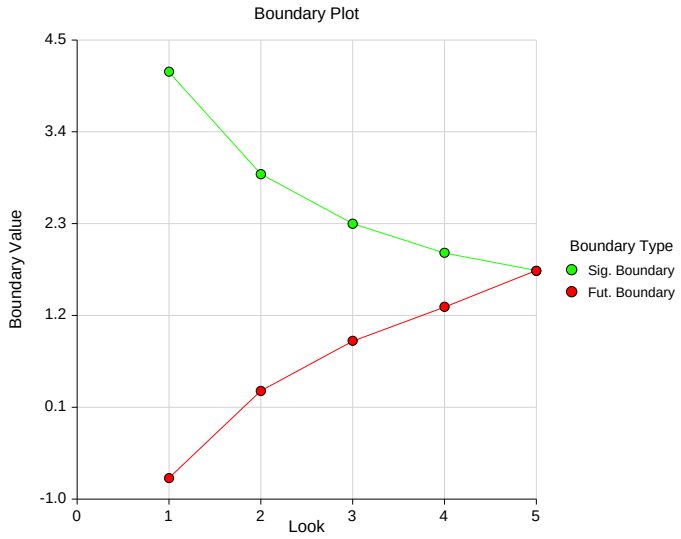
Accumulated Information Details for Scenario 1

	Accumulated Information	Accumulated Sample Size		
Look	Percent	Group 1	Group 2	Total
1	20.0	200	200	400
2	40.0	400	400	800
3	60.0	600	600	1200
4	80.0	800	800	1600
5	100.0	1000	1000	2000

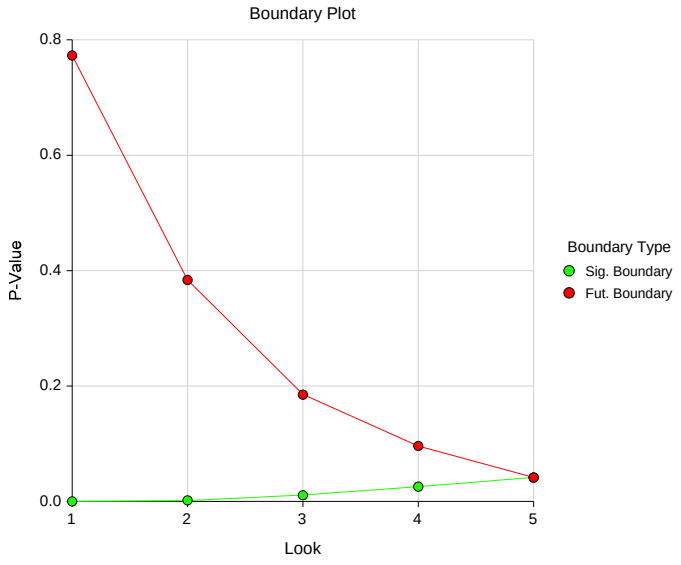
Boundaries for Scenario 1

	Significance Boundary		Futility Boundary	
	Z-Value	P-Value	Z-Value	P-Value
Look	Scale	Scale	Scale	Scale
1	3.762	0.000	-0.735	0.769
2	2.890	0.002	0.361	0.359
3	2.302	0.011	0.898	0.185
4	1.957	0.025	1.318	0.094
5	1.736	0.041	1.736	0.041

Boundary Plot



Boundary Plot - P-Value



Significance Boundaries with 95% Simulation Confidence Intervals for Scenario 1

	--------- Z-Value Boundary ---------			--------- P-Value Boundary ---------		
Look	Value	95% LCL	95% UCL	Value	95% LCL	95% UCL
1	3.762			0.000		
2	2.890	2.846	2.941	0.002	0.002	0.002
3	2.302	2.284	2.315	0.011	0.010	0.011
4	1.957	1.945	1.986	0.025	0.024	0.026
5	1.736	1.729	1.741	0.041	0.041	0.042

Futility Boundaries with 95% Simulation Confidence Intervals for Scenario 1

	Z-Value Boundary			P-Value Boundary		
Look	Value	95% LCL	95% UCL	Value	95% LCL	95% UCL
1	-0.735	-0.746	-0.730	0.769		
2	0.361	0.297	0.363	0.359	0.358	0.383
3	0.898	0.896	0.902	0.185	0.184	0.185
4	1.318	1.305	1.338	0.094	0.090	0.096
5	1.736	1.723	1.752	0.041	0.040	0.042

Alpha-Spending and Null Hypothesis Simulation Details for Scenario 1

	Signif. Boundary		Target		Actual		Proportion H0 Sims Outside Futility Boundary	Cum. H0 Sims Outside Futility Boundary
Look	Z-Value Scale	P-Value Scale	Spending Function Alpha	Cum. Spending Function Alpha	Alpha Spent	Cum. Alpha Spent		
1	3.762	0.000	0.000	0.000	0.000	0.000	0.222	0.222
2	2.890	0.002	0.002	0.002	0.002	0.002	0.426	0.649
3	2.302	0.011	0.009	0.011	0.009	0.011	0.194	0.843
4	1.957	0.025	0.017	0.028	0.015	0.027	0.082	0.925
5	1.736	0.041	0.022	0.050	0.012	0.039	0.036	0.961

Beta-Spending and Alternative Hypothesis Simulation Details for Scenario 1

	Futility Boundary		Target		Actual		Proportion H1 Sims Outside Signif. Boundary	Cum. H1 Sims Outside Signif. Boundary
Look	Z-Value Scale	P-Value Scale	Spending Function Beta	Cum. Spending Function Beta	Beta Spent	Cum. Beta Spent		
1	-0.735	0.769	0.035	0.035	0.035	0.035	0.003	0.003
2	0.361	0.359	0.102	0.137	0.101	0.137	0.074	0.077
3	0.898	0.185	0.088	0.225	0.087	0.224	0.230	0.307
4	1.318	0.094	0.068	0.293	0.068	0.292	0.237	0.543
5	1.736	0.041	0.054	0.347	0.054	0.346	0.111	0.654

Run Time: 93.05 seconds.

The values obtained from any given run of this example will vary slightly due to the variation in simulations.

Example 3 – Enter Boundaries

With a set-up similar to Example 2, suppose we wish to investigate the properties of a set of significance (3, 3, 3, 2, 1) and futility (-2, -1, 0, 0, 1) boundaries.

Setup

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

Option	Value
Data Tab	
Find (Solve For)	**Alpha and Power (Enter Boundaries)**
Power	*Ignored since this is the Find setting*
Alpha	*Ignored since this is the Find setting*
N1 (Sample Size Group 1)	**1000**
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
P1.0 (Treatment Proportion|H0)	**0.63**
P1.1 (Treatment Proportion|H1)	**0.68**
P2 (Proportion in Group 2)	**0.58**
Higher Proportions Are	**Better**
Test Type	**Z-Test (Pooled)**
Simulations	**100000**
Enter Boundaries Tab	
Number of Looks	**5**
Types of Boundaries	**Significance and Futility Boundaries**
Significance Boundary	**3 3 3 2 1 (for looks 1 through 5)**
Futility Boundary	**-2 -1 0 0 1 (for looks 1 through 5)**

Output

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

Numeric Results and Plots

Scenario 1 Numeric Results for Group Sequential Testing Proportion Difference = 0.
Hypotheses: H0: Proportion 1 = Proportion 2; H1: Proportion 1 > Proportion 2
Test Statistic: Z-Test (Pooled)
Zero Adjustment Method: Add 0 to each zero
Type of Boundaries: Significance and Futility Boundaries
Number of Looks: 5
Simulations: 100000

Numeric Summary for Scenario 1

Power			Alpha			
Value	95% LCL	95% UCL	Value	95% LCL	95% UCL	Beta
0.897	0.895	0.899	0.153	0.151	0.155	0.103

		Average Sample Size								
		Given H0		Given H1						
N1	N2	Grp1	Grp2	Grp1	Grp2	D0	D1	P1.0	P1.1	P2
1000	1000	738	738	829	829	0.1	0.1	0.6	0.7	0.6

Accumulated Information Details for Scenario 1

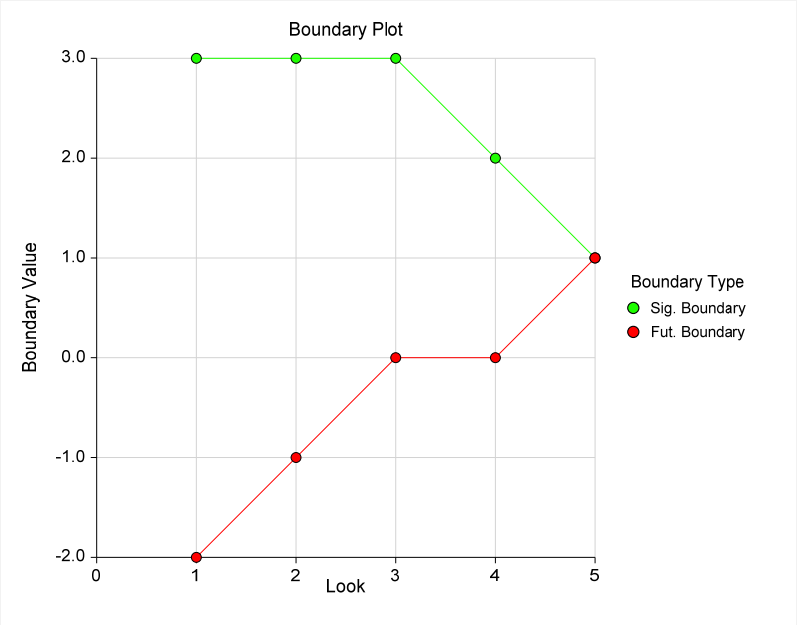
	Accumulated Information	Accumulated Sample Size		
Look	Percent	Group 1	Group 2	Total
1	20.0	200	200	400
2	40.0	400	400	800
3	60.0	600	600	1200
4	80.0	800	800	1600
5	100.0	1000	1000	2000

222-32 Group-Sequential Non-Zero Null Tests for Two Proportions (Simulation)

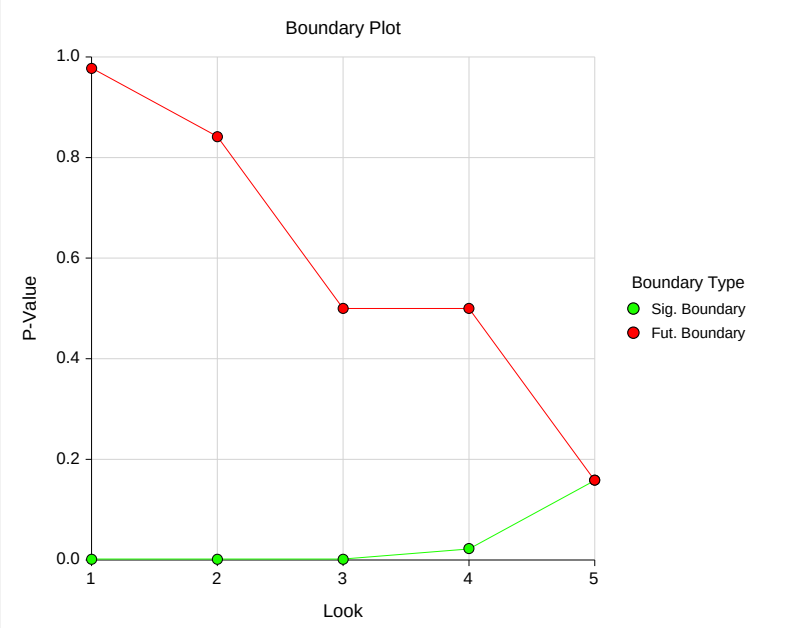
Boundaries for Scenario 1

	-- Significance Boundary --		----- Futility Boundary -----	
	Z-Value	P-Value	Z-Value	P-Value
Look	Scale	Scale	Scale	Scale
1	3.000	0.001	-2.000	0.977
2	3.000	0.001	-1.000	0.841
3	3.000	0.001	0.000	0.500
4	2.000	0.023	0.000	0.500
5	1.000	0.159	1.000	0.159

Boundary Plot



Boundary Plot - P-Value



Alpha-Spending and Null Hypothesis Simulation Details for Scenario 1

Look	--- Signif. Boundary--- Z-Value Scale	P-Value Scale	Alpha Spent	Cum. Alpha Spent	Proportion H0 Sims Outside Futility Boundary	Cum. H0 Sims Outside Futility Boundary
1	3.000	0.001	0.001	0.001	0.022	0.022
2	3.000	0.001	0.001	0.003	0.144	0.167
3	3.000	0.001	0.001	0.003	0.338	0.505
4	2.000	0.023	0.020	0.024	0.082	0.586
5	1.000	0.159	0.129	0.153	0.261	0.847

Beta-Spending and Alternative Hypothesis Simulation Details for Scenario 1

Look	-- Futility Boundary-- Z-Value Scale	P-Value Scale	Beta Spent	Cum. Beta Spent	Proportion H1 Sims Outside Signif. Boundary	Cum. H1 Sims Outside Signif. Boundary
1	-2.000	0.977	0.001	0.001	0.024	0.024
2	-1.000	0.841	0.006	0.008	0.048	0.072
3	0.000	0.500	0.030	0.038	0.064	0.136
4	0.000	0.500	0.005	0.043	0.397	0.533
5	1.000	0.159	0.060	0.103	0.364	0.897

Run Time: 62.89 seconds.

The values obtained from any given run of this example will vary slightly due to the variation in simulations.

Example 4 – Validation using Simulation

With a set-up similar to Example 2, we examine the power and alpha generated by the set of significance (3.762, 2.890, 2.302, 1.957, 1.736) and futility (-0.735, 0.361, 0.898, 1.318, 1.736) boundaries.

Setup

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

Option	Value
Data Tab	
Find (Solve For)	**Alpha and Power (Enter Boundaries)**
Power	*Ignored since this is the Find setting*
Alpha	*Ignored since this is the Find setting*
N1 (Sample Size Group 1)	**1000**
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
P1.0 (Treatment Proportion|H0)	**0.63**
P1.1 (Treatment Proportion|H1)	**0.68**
P2 (Proportion in Group 2)	**0.58**
Higher Proportions Are	**Better**
Test Type	**Z-Test (Pooled)**
Simulations	**100000**
Enter Boundaries Tab	
Number of Looks	**5**
Types of Boundaries	**Significance and Futility Boundaries**
Significance Boundary	**3.762, 2.890, 2.302, 1.957, 1.736**
Futility Boundary	**-0.735, 0.361, 0.898, 1.318, 1.736**

Output

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

Numeric Results and Plots

Numeric Summary for Scenario 1

Power			Alpha			
Value	95% LCL	95% UCL	Value	95% LCL	95% UCL	Beta
0.654	0.651	0.656	0.039	0.038	0.041	0.346

		Average Sample Size								
		Given H0		Given H1						
N1	N2	Grp1	Grp2	Grp1	Grp2	D0	D1	P1.0	P1.1	P2
1000	1000	464	464	676	676	0.1	0.1	0.6	0.7	0.6

The values obtained from any given run of this example will vary slightly due to the variation in simulations. The power and alpha generated with these boundaries are very close to the values of Example 2.

Chapter 223

Group-Sequential Non-Inferiority Tests for Two Proportions (Simulation)

Introduction

This procedure can be used to determine power, sample size and/or boundaries for group sequential non-inferiority tests comparing the proportions of two groups. The tests that can be simulated in this procedure are the common two-sample Z-test with or without pooled standard error and with or without continuity correction, the T-test, and three score tests. Significance and futility boundaries can be produced. The spacing of the looks can be equal or custom specified. Boundaries can be computed based on popular alpha- and beta-spending functions (O'Brien-Fleming, Pocock, Hwang-Shih-DeCani Gamma family, linear) or custom spending functions. Boundaries can also be input directly to verify alpha- and/or beta-spending properties. Futility boundaries can be binding or non-binding. Maximum and average (expected) sample sizes are reported as well as the alpha and/or beta spent and incremental power at each look. Corresponding P-Value boundaries are given for each boundary statistic. Plots of boundaries are also produced.

Four Procedures Documented Here

There are four procedures that use the program module described in this chapter. The Proportions and Differences procedures are identical except for the type of parameterization. The Ratios and Odds Ratios procedures use only likelihood score tests that are specific to testing the ratio of two proportions and the odds ratio of two proportions, respectively. Each procedure is listed separately on the menus.

Technical Details

This section outlines many of the technical details of the techniques used in this procedure including the simulation summary, the test statistic details, and the use of spending functions.

An excellent text for the background and details of many group-sequential methods is Jennison and Turnbull (2000).

Simulation Procedure

In this procedure, a large number of simulations are used to calculate boundaries and power using the following steps

1. Based on the specified proportions, random samples of size N1 and N2 are generated under the null distribution and under the alternative distribution. These are simulated samples as though the final look is reached.
2. For each sample, test statistics for each look are produced. For example, if N1 and N2 are 100 and there are 5 equally spaced looks, test statistics are generated from the random samples at N1 = N2 = 20, N1 = N2 = 40, N1 = N2 = 60, N1 = N2 = 80, and N1 = N2 = 100 for both null and alternative samples.
3. To generate the first significance boundary, the null distribution statistics of the first look (e.g., at N1 = N2 = 20) are ordered and the percent of alpha to be spent at the first look is determined (using either the alpha-spending function or the input value). The statistic for which the percent of statistics above (or below, as the case may be) that value is equal to the percent of alpha to be spent at the first look is the boundary statistic. It is seen here how important a large number of simulations is to the precision of the boundary estimates.
4. All null distribution samples that are outside the first significance boundary at the first look are removed from consideration for the second look. If binding futility boundaries are also being computed, all null distribution samples with statistics that are outside the first futility boundary are also removed from consideration for the second look. If non-binding futility boundaries are being computed, null distribution samples with statistics outside the first futility boundary are not removed.
5. To generate the second significance boundary, the remaining null distribution statistics of the second look (e.g., at N1 = N2 = 40) are ordered and the percent of alpha to be spent at the second look is determined (again, using either the alpha-spending function or the input value). The percent of alpha to be spent at the second look is multiplied by the total number of simulations to determine the number of the statistic that is to be the second boundary statistic. The statistic for which that number of statistics is above it (or below, as the case may be) is the second boundary statistic. For example, suppose there are initially 1000 simulated samples, with 10 removed at the first look (from, say, alpha spent at Look 1 equal to 0.01), leaving 990 samples considered for the second look. Suppose further that the alpha to be spent at the second look is 0.02. This is multiplied by 1000 to give 20. The 990 still-considered statistics are ordered and the 970^{th} (20 in from 990) statistic is the second boundary.
6. All null distribution samples that are outside the second significance boundary and the second futility boundary, if binding, at the second look are removed from consideration for the third look (e.g., leaving 970 statistics computed at N1 = N2 = 60 to be considered at the third look). Steps 4 and 5 are repeated until the final look is reached.

Futility boundaries are computed in a similar manner using the desired beta-spending function or custom beta-spending values and the alternative hypothesis simulated statistics at each look. For both binding and non-binding futility boundaries, samples for which alternative hypothesis statistics are outside either the significance or futility boundaries of the previous look are excluded from current and future looks.

Because the final futility and significance boundaries are required to be the same, futility boundaries are computed beginning at a small value of beta (e.g., 0.0001) and incrementing beta by that amount until the futility and significance boundaries meet.

When boundaries are entered directly, this procedure uses the null hypothesis and alternative hypothesis simulations to determine the number of test statistics that are outside the boundaries at each look. The cumulative proportion of alternative hypothesis statistics that are outside the significance boundaries is the overall power of the study.

Small Sample Considerations

When the sample size is small, say 200 or fewer per group, the discrete nature of the number of possible differences in proportions in the sampling distribution comes into play. This has led to a large number of proposed tests for comparing two proportions (or testing the 2 by 2 table of counts). For example, Upton (1982) considers twenty-two alternative tests for comparing two proportions. Sweeping statements about the power of one test over another are impossible to make, because the size of the Type I error depends upon the proportions used. At some proportions, some tests are overly conservative while others are not, while at other proportions the reverse may be true.

This simulation procedure, however, is based primarily on the ordering of the sample statistics in the simulation. The boundaries are determined by the spending function alphas. Thus, if a test used happens to be conservative in the single-look traditional sense, the boundaries chosen in the simulation results of this procedure will generally remove the conservative nature of the test. This makes comparisons to the one-look case surprising in many cases.

Definitions

Suppose you have two populations from which dichotomous (binary) responses will be recorded. The probability (or risk) of obtaining the event of interest in population 1 (the treatment group) is p_1 and in population 2 (the control group) is p_2. The corresponding failure proportions are given by $q_1 = 1 - p_1$ and $q_2 = 1 - p_2$.

The assumption is made that the responses from each group follow a binomial distribution. This means that the event probability, p_i, is the same for all subjects within the group and that the response from one subject is independent of that of any other subject.

Random samples of *m* and *n* individuals are obtained from these two populations. The data from these samples can be displayed in a 2-by-2 contingency table as follows

Group	Success	Failure	Total
Treatment	*a*	*c*	*m*
Control	*b*	*d*	*n*
Total	*s*	*f*	*N*

The following alternative notation is also used.

Group	Success	Failure	Total
Treatment	x_{11}	x_{12}	n_1
Control	x_{21}	x_{22}	n_2
Total	m_1	m_2	N

The binomial proportions p_1 and p_2 are estimated from these data using the formulae

$$\hat{p}_1 = \frac{a}{m} = \frac{x_{11}}{n_1} \text{ and } \hat{p}_2 = \frac{b}{n} = \frac{x_{21}}{n_2}$$

Comparing Two Proportions

Let $p_{1.0}$ represent the group 1 proportion tested by the null hypothesis, H_0. The power of a test is computed at a specific value of the proportion which we will call $p_{1.1}$. Let δ represent the smallest difference (margin of non-inferiority) between the two proportions that still results in the conclusion that the new treatment is not inferior to the current treatment. For a non-inferiority test, $\delta < 0$. The set of statistical hypotheses that are tested is

$$\mathrm{H}_0: p_{1.0} - p_2 \le \delta \text{ versus } \mathrm{H}_1: p_{1.0} - p_2 > \delta$$

which can be rearranged to give

$$\mathrm{H}_0: p_{1.0} \le p_2 + \delta \text{ versus } \mathrm{H}_1: p_{1.0} > p_2 + \delta$$

There are multiple methods of specifying the margin of non-inferiority. The most direct is to simply give values for p_2 and $p_{1.0}$. However, it is often more meaningful to give p_2 and then specify $p_{1.0}$ implicitly by specifying the difference, ratio, or odds ratio. Mathematically, assuming higher proportions are better, the definitions of these parameterizations are

Parameter	**Computation**	**Hypotheses**
Difference	$\delta = p_{1.0} - p_2$	$\mathrm{H}_0: p_{1.0} - p_2 \le \delta_0$ vs. $\mathrm{H}_1: p_{1.0} - p_2 > \delta_0, \quad \delta_0 < 0$
Ratio	$\phi = p_{1.0} / p_2$	$\mathrm{H}_0: p_1 / p_2 \le \phi_0$ vs. $\mathrm{H}_1: p_1 / p_2 > \phi_0, \quad \phi_0 < 1$
Odds Ratio	$\psi = Odds_{1.0} / Odds_2$	$\mathrm{H}_0: o_{1.0} / o_2 \le \psi_0$ versus $\mathrm{H}_1: o_{1.0} / o_2 > \psi_0, \quad \psi_0 < 1$

Difference

The difference is perhaps the most direct method of comparison between two proportions. It is easy to interpret and communicate. It gives the absolute impact of the treatment. However, there are subtle difficulties that can arise with its interpretation.

One difficulty arises when the event of interest is rare. If a difference of 0.001 occurs when the baseline probability is 0.40, it would be dismissed as being trivial. However, if the baseline probably of a disease is 0.002, a 0.001 decrease would represent a reduction of 50%. Thus interpretation of the difference depends on the baseline probability of the event.

The following example is used to convey the concept of a *non-inferiority* test. Suppose 60% of patients respond to the current treatment method $\left(p_2 = 0.60\right)$. If the response rate of the new treatment is no less than 5 percentage points worse $\left(\delta = -0.05\right)$ than the existing treatment, it will be considered to be non-inferior. Substituting these figures into the statistical hypotheses gives

$$H_0: \delta \le -0.05 \text{ versus } H_1: \delta > -0.05$$

In this example, when the null hypothesis is rejected, the concluded alternative is that the response rate is at least 55%, which means that the new treatment is not inferior to the current treatment.

Ratio

The ratio, $\phi = p_{1.0} / p_2$, gives the relative change in the probability of the response. Testing non-inferiority uses the formulation

$$H_0: p_{1.0} / p_2 \le \phi_0 \text{ versus } H_1: p_{1.0} / p_2 > \phi_0$$

when higher proportions are better.

The following example is used to convey the concept of *non-inferiority* as defined by the ratio. Suppose that 60% of patients $\left(p_2 = 0.60\right)$ respond to the current treatment method. If a new treatment decreases the response rate by no more than 10% $\left(\phi_0 = 0.90\right)$, it will be considered non-inferior to the standard treatment. Substituting these figures into the statistical hypotheses gives

$$H_0: \phi \le 0.90 \text{ versus } H_1: \phi > 0.90$$

Using the relationship

$$p_{1.0} = \phi_0 p_2$$

gives

$$H_0: p_{1.0} \le 0.54 \text{ versus } H_1: p_{1.0} > 0.54$$

In this example, when the null hypothesis is rejected, the concluded alternative is that the response rate is at least 54%. That is, the conclusion of non-inferiority is that the new treatment's response rate is no worse than 10% less than that of the standard treatment.

Odds Ratio

The odds ratio, $\psi = \left(p_{1.0} / \left(1 - p_{1.0}\right)\right) / \left(p_2 / \left(1 - p_2\right)\right)$, gives the relative change in the odds of the response. Testing non-inferiority uses the formulation

$$H_0: \psi \le \psi_0 \text{ versus } H_1: \psi > \psi_0$$

when higher proportions are better.

Test Statistics

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

Z Test (Pooled and Unpooled)

This test statistic was first proposed by Karl Pearson in 1900. Although this test can be expressed as a Chi-Square statistic, it is expressed here as a *z* so that it can be used for one-sided hypothesis testing.

Both *pooled* and *unpooled* versions of this test have been discussed in the statistical literature. The pooling refers to the way in which the standard error is estimated. In the pooled version, the two proportions are averaged, and only one proportion is used to estimate the standard error. In the unpooled version, the two proportions are used separately.

The formula for the test statistic is

$$z_t = \frac{\hat{p}_1 - \hat{p}_2}{\hat{\sigma}_D}$$

Pooled Version

$$\hat{\sigma}_D = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

$$\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$$

Unpooled Version

$$\hat{\sigma}_D = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

Continuity Correction

Frank Yates is credited with proposing a correction to the Pearson Chi-Square test for the lack of continuity in the binomial distribution. However, the correction was in common use when he proposed it in 1922.

The continuity corrected z-test is

$$z = \frac{\left(\hat{p}_1 - \hat{p}_2\right) + \frac{F}{2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}{\hat{\sigma}_D}$$

where *F* is -1 for upper-tailed, 1 for lower-tailed, and either -1 or 1 for two-sided hypotheses, depending on whether the numerator difference is positive or negative.

T-Test

Based on a study of the behavior of several tests, D'Agostino (1988) and Upton (1982) proposed using the usual two-sample *t*-test for testing whether two proportions are equal. One substitutes a '1' for a success and a '0' for a failure in the usual, two-sample *t*-test formula. The test statistic is computed as

$$t_{N-2} = (ad - bc)\left(\frac{N-2}{N(nac + mbd)}\right)^{\frac{1}{2}}$$

which can be compared to the *t* distribution with *N*-2 degrees of freedom.

Miettinen and Nurminen's Likelihood Score Test of the Difference

Miettinen and Nurminen (1985) proposed a test statistic for testing whether the difference is equal to a specified, non-zero, value, δ_0. The regular MLE's, $\hat{p}_1$ and $\hat{p}_2$, are used in the numerator of the score statistic while MLE's $\tilde{p}_1$ and $\tilde{p}_2$, constrained so that $\tilde{p}_1 - \tilde{p}_2 = \delta_0$, are used in the denominator. A correction factor of *N*/(*N*-1) is applied to make the variance estimate less biased. The significance level of the test statistic is based on the asymptotic normality of the score statistic. The formula for computing this test statistic is

$$z_{MND} = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0}{\hat{\sigma}_{MND}}$$

where

$$\hat{\sigma}_{MND} = \sqrt{\left(\frac{\tilde{p}_1\tilde{q}_1}{n_1} + \frac{\tilde{p}_2\tilde{q}_2}{n_2}\right)\left(\frac{N}{N-1}\right)}$$

$$\tilde{p}_1 = \tilde{p}_2 + \delta_0$$

$$\tilde{p}_1 = 2B\cos(A) - \frac{L_2}{3L_3}$$

$$A = \frac{1}{3}\left[\pi + \cos^{-1}\left(\frac{C}{B^3}\right)\right]$$

$$B = \text{sign}(C)\sqrt{\frac{L_2^2}{9L_3} - \frac{L_1}{3L_3}}$$

$$C = \frac{L_2^3}{27L_3^3} - \frac{L_1L_2}{6L_3^2} + \frac{L_0}{2L_3}$$

$$L_0 = x_{21}\delta_0(1-\delta_0)$$

$$L_1 = \left[N_2\delta_0 - N - 2x_{21}\right]\delta_0 + M_1$$

$$L_2 = (N + N_2)\delta_0 - N - M_1$$

$$L_3 = N$$

Miettinen and Nurminen's Likelihood Score Test of the Ratio

Miettinen and Nurminen (1985) proposed a test statistic for testing whether the ratio is equal to a specified value ϕ_0. The regular MLE's, $\hat{p}_1$ and $\hat{p}_2$, are used in the numerator of the score statistic while MLE's $\tilde{p}_1$ and $\tilde{p}_2$, constrained so that $\tilde{p}_1 / \tilde{p}_2 = \phi_0$, are used in the denominator. A correction factor of N/(N-1) is applied to make the variance estimate less biased. The significance level of the test statistic is based on the asymptotic normality of the score statistic.

The formula for computing the test statistic is

$$z_{MNR} = \frac{\hat{p}_1 / \hat{p}_2 - \phi_0}{\sqrt{\left(\frac{\tilde{p}_1\tilde{q}_1}{n_1} + \phi_0^2\frac{\tilde{p}_2\tilde{q}_2}{n_2}\right)\left(\frac{N}{N-1}\right)}}$$

where

$$\tilde{p}_1 = \tilde{p}_2\phi_0$$

$$\tilde{p}_2 = \frac{-B - \sqrt{B^2 - 4AC}}{2A}$$

$$A = N\phi_0$$

$$B = -\left[N_1\phi_0 + x_{11} + N_2 + x_{21}\phi_0\right]$$

$$C = M_1$$

Miettinen and Nurminen's Likelihood Score Test of the Odds Ratio

Miettinen and Nurminen (1985) proposed a test statistic for testing whether the odds ratio is equal to a specified value, ψ_0. Because the approach they used with the difference and ratio does not easily extend to the odds ratio, they used a score statistic approach for the odds ratio. The regular MLE's are $\hat{p}_1$ and $\hat{p}_2$. The constrained MLE's are $\tilde{p}_1$ and $\tilde{p}_2$. These estimates are constrained so that $\tilde{\psi} = \psi_0$. A correction factor of N/(N-1) is applied to make the variance estimate less biased. The significance level of the test statistic is based on the asymptotic normality of the score statistic.

The formula for computing the test statistic is

$$z_{MNO} = \frac{\frac{\left(\hat{p}_1 - \tilde{p}_1\right)}{\tilde{p}_1\tilde{q}_1} - \frac{\left(\hat{p}_2 - \tilde{p}_2\right)}{\tilde{p}_2\tilde{q}_2}}{\sqrt{\left(\frac{1}{N_2\tilde{p}_1\tilde{q}_1} + \frac{1}{N_2\tilde{p}_2\tilde{q}_2}\right)\left(\frac{N}{N-1}\right)}}$$

where

$$\tilde{p}_1 = \frac{\tilde{p}_2\psi_0}{1 + \tilde{p}_2\left(\psi_0 - 1\right)}$$

$$\tilde{p}_2 = \frac{-B + \sqrt{B^2 - 4AC}}{2A}$$

$$A = N_2\left(\psi_0 - 1\right)$$

$$B = N_1\psi_0 + N_2 - M_1\left(\psi_0 - 1\right)$$

$$C = -M_1$$

Farrington and Manning's Likelihood Score Test of the Difference

Farrington and Manning (1990) proposed a test statistic for testing whether the difference is equal to a specified value δ_0. The regular MLE's, $\hat{p}_1$ and $\hat{p}_2$, are used in the numerator of the score statistic while MLE's $\tilde{p}_1$ and $\tilde{p}_2$, constrained so that $\tilde{p}_1 - \tilde{p}_2 = \delta_0$, are used in the denominator. The significance level of the test statistic is based on the asymptotic normality of the score statistic.

The formula for computing the test statistic is

$$z_{FMD} = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0}{\sqrt{\left(\frac{\tilde{p}_1\tilde{q}_1}{n_1} + \frac{\tilde{p}_2\tilde{q}_2}{n_2}\right)}}$$

where the estimates $\tilde{p}_1$ and $\tilde{p}_2$ are computed as in the corresponding test of Miettinen and Nurminen (1985) given above.

Farrington and Manning's Likelihood Score Test of the Ratio

Farrington and Manning (1990) proposed a test statistic for testing whether the ratio is equal to a specified value ϕ_0. The regular MLE's, $\hat{p}_1$ and $\hat{p}_2$, are used in the numerator of the score statistic while MLE's $\tilde{p}_1$ and $\tilde{p}_2$, constrained so that $\tilde{p}_1 / \tilde{p}_2 = \phi_0$, are used in the denominator. A correction factor of N/(N-1) is applied to increase the variance estimate. The significance level of the test statistic is based on the asymptotic normality of the score statistic.

The formula for computing the test statistic is

$$z_{FMR} = \frac{\hat{p}_1 / \hat{p}_2 - \phi_0}{\sqrt{\left(\frac{\tilde{p}_1\tilde{q}_1}{n_1} + \phi_0^2\frac{\tilde{p}_2\tilde{q}_2}{n_2}\right)}}$$

where the estimates $\tilde{p}_1$ and $\tilde{p}_2$ are computed as in the corresponding test of Miettinen and Nurminen (1985) given above.

Farrington and Manning's Likelihood Score Test of the Odds Ratio

Farrington and Manning (1990) indicate that the Miettinen and Nurminen statistic may be modified by removing the factor $N/(N\text{-}1)$.

The formula for computing this test statistic is

$$z_{FMO} = \frac{\frac{\left(\hat{p}_1 - \tilde{p}_1\right)}{\tilde{p}_1\tilde{q}_1} - \frac{\left(\hat{p}_2 - \tilde{p}_2\right)}{\tilde{p}_2\tilde{q}_2}}{\sqrt{\left(\frac{1}{N_2\tilde{p}_1\tilde{q}_1} + \frac{1}{N_2\tilde{p}_2\tilde{q}_2}\right)}}$$

where the estimates $\tilde{p}_1$ and $\tilde{p}_2$ are computed as in the corresponding test of Miettinen and Nurminen (1985) given above.

Gart and Nam's Likelihood Score Test of the Difference

Gart and Nam (1990), page 638, proposed a modification to the Farrington and Manning (1988) difference test that corrects for skewness. Let $z_{FMD}(\delta)$ stand for the Farrington and Manning difference test statistic described above. The skewness corrected test statistic, z_{GND}, is the appropriate solution to the quadratic equation

$$(-\tilde{\gamma})z_{GND}^2 + (-1)z_{GND} + (z_{FMD}(\delta) + \tilde{\gamma}) = 0$$

where

$$\tilde{\gamma} = \frac{\tilde{V}^{3/2}(\delta)}{6}\left(\frac{\tilde{p}_1\tilde{q}_1(\tilde{q}_1 - \tilde{p}_1)}{n_1^2} - \frac{\tilde{p}_2\tilde{q}_2(\tilde{q}_2 - \tilde{p}_2)}{n_2^2}\right)$$

Gart and Nam's Likelihood Score Test of the Ratio

Gart and Nam (1988), page 329, proposed a modification to the Farrington and Manning (1988) ratio test that corrects for skewness. Let $z_{FMR}(\phi)$ stand for the Farrington and Manning ratio test statistic described above. The skewness corrected test statistic, z_{GNR}, is the appropriate solution to the quadratic equation

$$(-\tilde{\varphi})z_{GNR}^2 + (-1)z_{GNR} + (z_{FMR}(\phi) + \tilde{\varphi}) = 0$$

where

$$\tilde{\varphi} = \frac{1}{6\tilde{u}^{3/2}}\left(\frac{\tilde{q}_1(\tilde{q}_1 - \tilde{p}_1)}{n_1^2\tilde{p}_1^2} - \frac{\tilde{q}_2(\tilde{q}_2 - \tilde{p}_2)}{n_2^2\tilde{p}_2^2}\right)$$

$$\tilde{u} = \frac{\tilde{q}_1}{n_1\tilde{p}_1} + \frac{\tilde{q}_2}{n_2\tilde{p}_2}$$

Spending Functions

Spending functions can be used in this procedure to specify the proportion of alpha or beta that is spent at each look without having to specify the proportion directly.

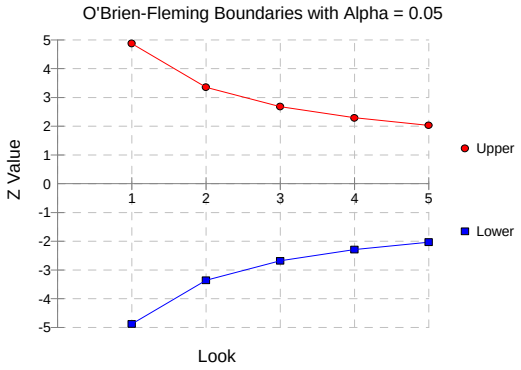
Spending functions have the characteristics that they are increasing and that

$$\alpha(0) = 0$$
$$\alpha(1) = \alpha$$

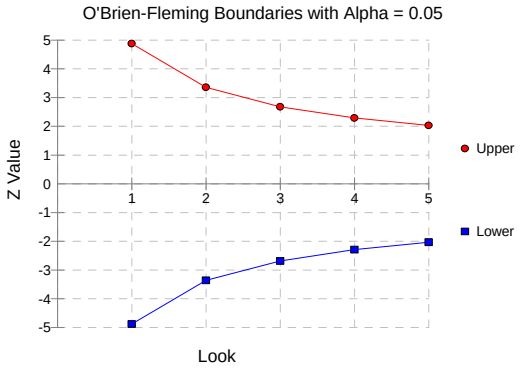
The last characteristic guarantees a fixed α level when the trial is complete. This methodology is very flexible since neither the times nor the number of analyses must be specified in advance. Only the functional form of $\alpha(\tau)$ must be specified.

PASS provides several popular spending functions plus the ability to enter and analyze your own percents of alpha or beta spent. These are calculated as follows (beta may be substituted for alpha for beta-spending functions):

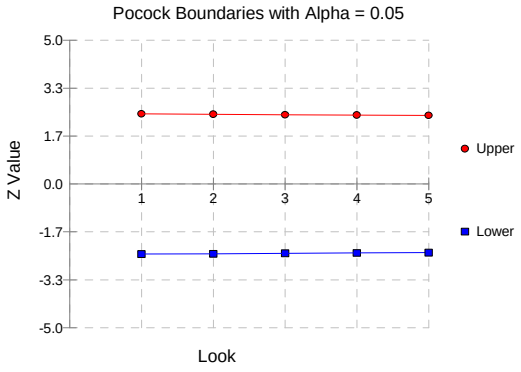
1. Hwang-Shih-DeCani (gamma family) $\alpha\left[\frac{1-e^{-\gamma t}}{1-e^{-\gamma}}\right], \gamma \neq 0\,;\ \ \alpha t, \gamma = 0$



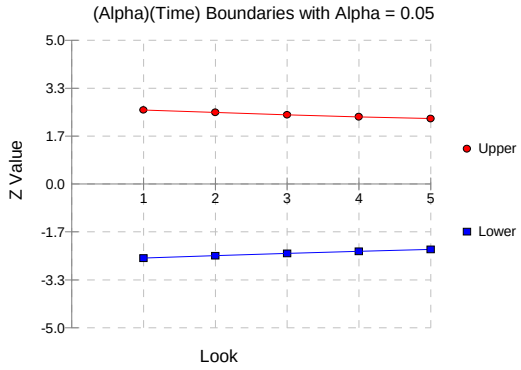
2. O'Brien-Fleming Analog $2 - 2\Phi\left(\frac{Z_{\alpha/2}}{\sqrt{t}}\right)$



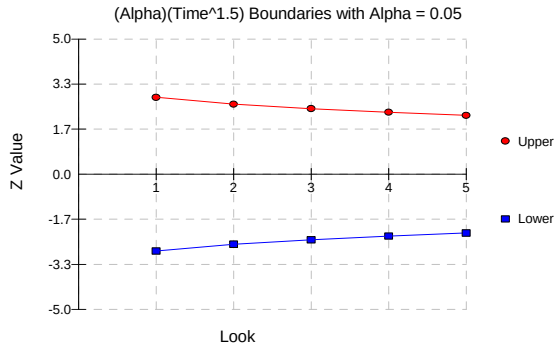
3. Pocock Analog $\alpha \cdot \ln\left(1 + (e-1)t\right)$



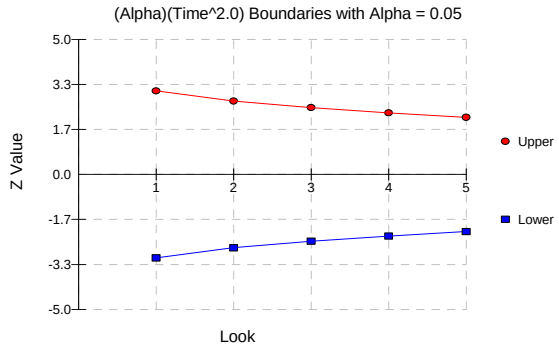
4. Alpha * time $\alpha \cdot t$



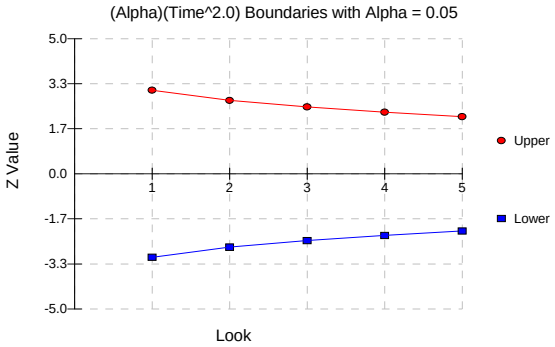
5. Alpha * time^1.5 $\alpha \cdot t^{3/2}$



6. Alpha * time^2 $\alpha \cdot t^2$



7. **Alpha * time^C** $\alpha \cdot t^C$



8. **User Supplied Percents**

   A custom set of percents of alpha to be spent at each look may be input directly.

The O'Brien-Fleming Analog spends very little alpha or beta at the beginning and much more at the final looks. The Pocock Analog and (Alpha or Beta)(Time) spending functions spend alpha or beta more evenly across the looks. The Hwang-Shih-DeCani (C) (gamma family) spending functions and (Alpha or Beta)(Time^C) spending functions are flexible spending functions that can be used to spend more alpha or beta early or late or evenly, depending on the choice of C.

Procedure Options

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

Data Tab

The Data tab contains most of the parameters and options for the general setup of the procedure.

Solve For

Find (Solve For)

Solve for either power, sample size, or enter the boundaries directly and solve for power and alpha. When solving for power or sample size, the look and boundary details are specified on the "Looks & Boundaries" tab and the "Enter Boundaries" tab is ignored. When entering the boundaries directly and solving for power and alpha, the boundaries are input on the "Enter Boundaries" tab and the "Looks & Boundaries" tab is ignored.

When solving for power or N1, the early-stopping boundaries are also calculated. High accuracy for early-stopping boundaries requires a very large number of simulations (Recommended 100,000 to 10,000,000).

The parameter selected here is the parameter displayed on the vertical axis of the plot.

Because this is a simulation based procedure, the search for the sample size may take several minutes or hours to complete. You may find it quicker and more informative to solve for Power for a range of sample sizes.

Error Rates

Power or Beta

Power is the probability of rejecting the null hypothesis when it is false. Power is equal to 1-Beta, so specifying power implicitly specifies beta.

Beta is the probability obtaining a false negative on the statistical test. That is, it is the probability of accepting a false null hypothesis.

In the context of simulated group sequential trials, the power is the proportion of the alternative hypothesis simulations that cross any one of the significance (efficacy) boundaries.

The valid range is between 0 to 1.

Different disciplines and protocols have different standards for setting power. A common choice is 0.90, but 0.80 is also popular.

You can enter a range of values such as *0.70 0.80 0.90* or *0.70 to 0.90 by 0.1*.

Alpha (Significance Level)

Alpha is the probability of obtaining a false positive on the statistical test. That is, it is the probability of rejecting a true null hypothesis.

The null hypothesis is usually that the parameters (the means, proportions, etc.) are all equal.

In the context of simulated group sequential trials, alpha is the proportion of the null hypothesis simulations that cross any one of the significance (efficacy) boundaries.

Since Alpha is a probability, it is bounded by 0 and 1. Commonly, it is between 0.001 and 0.250.

Alpha is often set to 0.05 for two-sided tests and to 0.025 for one-sided tests.

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

Sample Size

N1 (Sample Size Group 1)

Enter a value for the sample size, N1. This is the number of subjects in the first group of the study at the final look.

You may enter a range such as 10 to 100 by 10 or a list of values separated by commas or blanks.

You might try entering the same number two or three times to get an idea of the variability in your results. For example, you could enter "10 10 10".

N2 (Sample Size Group 2)

Enter a value (or range of values) for the sample size of group 2. This is the number of subjects in the second group of the study at the final look.

- **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 the operator $[Y]$ 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*.

R = N2/N1

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*.

Effect Size

P1.0 (Non-Inferiority Proportion) – Proportions

Specify the value P1.0 directly.

When higher proportions are "Better", P1.0 is the treatment group proportion such that proportions greater than P1.0 are considered non-inferior to the reference group proportion. When higher proportions are "Better", P1.0 should be less than P2.

When higher proportions are "Worse", P1.0 is the treatment group proportion such that proportions less than P1.0 are considered non-inferior to the reference group proportion. When higher proportions are "Worse", P1.0 should be greater than P2.

When P1.0 is close to zero or one, a Zero Count Adjustment may be needed.

You can enter a list of values such as *0.4 0.5 0.6* or *0.3 to 0.7 by 0.05*.

P1.1 (Actual Proportion) – Proportions

This is the value of the treatment proportion (P1) at which the power calculations are made.

When P1.1 is close to zero or one, a Zero Count Adjustment may be needed.

You can enter a list of values such as *0.4 0.5 0.6* or *0.3 to 0.7 by 0.05*.

D0 (Non-Inferiority Difference) – Differences

Specify the difference between P1.0 and P2 that will be considered non-inferior.

When higher proportions are "Better", the non-inferiority difference is that amount that P1 can be less than P2 and still have the treatment group declared non-inferior to the reference group. When higher proportions are "Better", D0 should be negative.

When higher proportions are "Worse", the non-inferiority difference is that amount that P1 can be greater than P2 and still have the treatment group declared non-inferior to the reference group. When higher proportions are "Worse", D0 should be positive.

The power calculations assume that P1.0 is the value of the P1 under the null hypothesis. This value is used with P2 to calculate the value of P1.0 using the formula: P1.0 = D0 + P2

You may enter a range of values such as *-.03 -.05 -.10* or *-.05 to -.01 by .01*.

Differences must be between -1 and 1. D0 cannot take on the values -1, 0, or 1. D0 should be selected such that P1.0 is between zero and one.

D1 (Actual Difference) – Differences

Specify the actual difference between P1.1 (the actual value of P1) and P2. This is the value at which the power is calculated.

In non-inferiority and equivalence trials, this difference is often set to zero.

The power calculations assume that P1.1 is the actual value of the proportion in group 1 (experimental or treatment group). This difference is used with P2 to calculate the value of P1 using the formula: P1.1 = D1 + P2

You may enter a range of values such as -*.05 0 .5* or -*.05 to .05 by .02*.

Actual differences must be between -1 and 1. They cannot take on the values -1 or 1.

R0 (Non-Inferiority Ratio) – Ratios

Specify the non-inferiority ratio of P1.0 to P2.

When higher proportions are "Better", the non-inferiority ratio is smallest ratio of P1 to P2 for which P1 will still be considered non-inferior. When higher proportions are "Better", R0 should be less than one.

When higher proportions are "Worse", the non-inferiority ratio is largest ratio of P1 to P2 for which P1 will still be considered non-inferior. When higher proportions are "Worse", R0 should be greater than one.

The power calculations assume that P1.0 is the value of P1 under the null hypothesis. This value is used with P2 to calculate the value of P1.0 using the formula: P1.0 = R0 x P2

You may enter a range of values such as *.95 .97 .99* or *.91 to .99 by .02*.

Ratios must be positive. R0 cannot take on the value of one.

R1 (Actual Ratio) – Ratios

Specify the actual ratio between P1.1 (the actual value of P1) and P2. This is the value at which the power is calculated.

In non-inferiority trials, this ratio is often set to one.

The power calculations assume that P1.1 is the actual value of the proportion in group 1 (experimental or treatment group). This ratio is used with P2 to calculate the value of P1 using the formula: P1.1 = R1 x P2

You may enter a range of values such as *.95 1 1.05* or *.9 to 1 by .02*.

Ratios must be positive.

OR0 (Non-Inferiority Odds Ratio) – Odds Ratios

Specify the non-feriority odds ratio between P1.0 and P2.

When higher proportions are "Better", the non-inferiority odds ratio is smallest odds ratio of P1 to P2 for which P1 will still be considered non-inferior. When higher proportions are "Better", OR0 should be less than one.

When higher proportions are "Worse", the non-inferiority odds ratio is largest odds ratio of P1 to P2 for which P1 will still be considered non-inferior. When higher proportions are "Worse", OR0 should be greater than one.

The power calculations assume that P1.0 is the value of P1 under the null hypothesis. OR0 is used with P2 to calculate the value of P1.0.

You may enter a range of values such as *.95 .97 .99* or *.91 to .99 by .02*.

Odds ratios must be positive. OR0 cannot take on the value of one.

OR1 (Actual Odds Ratio) – Odds Ratios

Specify the actual odds ratio between P1.1 (the actual value of P1) and P2. This is the value at which the power is calculated.

In non-inferiority trials, this value is often set to one.

The power calculations assume that P1.1 is the actual value of the proportion in group 1 (experimental or treatment group). This odds ratio is used with P2 to calculate the value of P1.1 using the formula: P1.1 = R1 x P2

You may enter a range of values such as.95 1 1.05 or .9 to 1 by .02.

Ratios must be positive.

P2 (Control Group Proportion)

Enter a value for the proportion in the control (baseline, standard, or reference) group, P2.

When P2 is close to zero or one, a Zero Count Adjustment may be needed.

Values must be between 0 and 1.

You can enter a list of values such as *0.4 0.5 0.6* or *0.3 to 0.7 by 0.05*.

Test and Simulations

Higher Proportions Are

Use this option to specify the direction of the test.

If Higher Proportions are "Better", the alternative hypothesis is H1: P1-P2 > NIM, H1: P1/P2 > NIR, or H1: O1/O2 > NIOR.

If Higher Proportions are "Worse", the alternative hypothesis is H1: P1-P2 < NIM, H1: P1/P2 < NIR, or H1: O1/O2 < NIOR.

Test Type

Specify which test statistic is to be simulated and reported on.

For details and formulation of the tests, see the section in this manual above.

Simulations

Specify the number of Monte Carlo iterations, M.

The following table gives an estimate of the precision that is achieved for various simulation sizes when the power is near 0.50 and 0.95. The values in the table are the "Precision" amounts that are added and subtracted to form a 95% confidence interval.

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

However, when solving for Power or N1, the simulations are used to calculate the look boundaries. To obtain precise boundary estimates, the number of simulations needs to be high. However, this consideration competes with the length of time to complete the simulation. When solving for power, a large number of simulations (100,000 or 1,000,000) will finish in several minutes. When solving for N1, perhaps 10,000 simulations can be run for each iteration. Then, a final run with the resulting N1 solving for power can be run with more simulations.

Looks & Boundaries Tab

The Data tab contains most of the parameters and options for the general setup of the procedure.

Looks and Boundaries

Specification of Looks and Boundaries

Choose whether spending functions will be used to divide alpha and beta for each look (Simple Specification), or whether the percents of alpha and beta to be spent at each look will be specified directly (Custom Specification).

Under Simple Specification, the looks are automatically considered to be equally spaced. Under Custom Specification, the looks may be equally spaced or custom defined based on the percent of accumulated information.

Looks and Boundaries – Simple Specification

Number of Equally Spaced Looks

Select the total number of looks that will be used if the study is not stopped early for the crossing of a boundary.

Alpha Spending Function

Specify the type of alpha spending function to use.

The O'Brien-Fleming Analog spends very little alpha at the beginning and much more at the final looks. The Pocock Analog and (Alpha)(Time) spending functions spend alpha more evenly across the looks. The Hwang-Shih-DeCani (C) (sometimes called the gamma family) spending functions and (Alpha)(Time^C) spending functions are flexible spending functions that can be used to spend more alpha early or late or evenly, depending on the choice of C.

C (Alpha Spending)

C is used to define the Hwang-Shih-DeCani (C) or (Alpha)(Time^C) spending functions.

For the Hwang-Shih-DeCani (C) spending function, negative values of C spend more of alpha at later looks, values near 0 spend alpha evenly, and positive values of C spend more of alpha at earlier looks.

For the (Alpha)(Time^C) spending function, only positive values for C are permitted. Values of C near zero spend more of alpha at earlier looks, values near 1 spend alpha evenly, and larger values of C spend more of alpha at later looks.

Type of Futility Boundary

This option determines whether or not futility boundaries will be created, and if so, whether they are binding or non-binding.

Futility boundaries are boundaries such that, if crossed at a given look, stop the study in favor of H0.

Binding futility boundaries are computed in concert with significance boundaries. They are called binding because they require the stopping of a trial if they are crossed. If the trial is not stopped, the probability of a false positive will exceed alpha.

When non-binding futility boundaries are computed, the significance boundaries are first computed, ignoring the futility boundaries. The futility boundaries are then computed. These futility boundaries are non-binding because continuing the trial after they are crossed will not affect the overall probability of a false positive declaration.

Number of Skipped Futility Looks

In some trials it may be desirable to wait a number of looks before examining the trial for futility. This option allows the beta to begin being spent after a specified number of looks.

The Number of Skipped Futility Looks should be less than the number of looks.

Beta Spending Function

Specify the type of beta spending function to use.

The O'Brien-Fleming Analog spends very little beta at the beginning and much more at the final looks. The Pocock Analog and (Beta)(Time) spending functions spend beta more evenly across the looks. The Hwang-Shih-DeCani (C) (sometimes called the gamma family) spending functions and (Beta)(Time^C) spending functions are flexible spending functions that can be used to spend more beta early or late or evenly, depending on the choice of C.

C (Beta Spending)

C is used to define the Hwang-Shih-DeCani (C) or (Beta)(Time^C) spending functions.

For the Hwang-Shih-DeCani (C) spending function, negative values of C spend more of beta at later looks, values near 0 spend beta evenly, and positive values of C spend more of beta at earlier looks.

For the (Beta)(Time^C) spending function, only positive values for C are permitted. Values of C near zero spend more of beta at earlier looks, values near 1 spend beta evenly, and larger values of C spend more of beta at later looks.

Looks and Boundaries – Custom Specification

Number of Looks

This is the total number of looks of either type (significance or futility or both).

Equally Spaced

If this box is checked, the Accumulated Information boxes are ignored and the accumulated information is evenly spaced.

Type of Futility Boundary

This option determines whether or not futility boundaries will be created, and if so, whether they are binding or non-binding.

Futility boundaries are boundaries such that, if crossed at a given look, stop the study in favor of H0.

Binding futility boundaries are computed in concert with significance boundaries. They are called binding because they require the stopping of a trial if they are crossed. If the trial is not stopped, the probability of a false positive will exceed alpha.

When Non-binding futility boundaries are computed, the significance boundaries are first computed, ignoring the futility boundaries. The futility boundaries are then computed. These futility boundaries are non-binding because continuing the trial after they are crossed will not affect the overall probability of a false positive declaration.

Accumulated Information

The accumulated information at each look defines the proportion or percent of the sample size that is used at that look.

These values are accumulated information values so they must be increasing.

Proportions, percents, or sample sizes may be entered. All proportions, percents, or sample sizes will be divided by the value at the final look to create an accumulated information proportion for each look.

Percent of Alpha Spent

This is the percent of the total alpha that is spent at the corresponding look. It is not the cumulative value.

Percents, proportions, or alphas may be entered here. Each of the values is divided by the sum of the values to obtain the proportion of alpha that is used at the corresponding look.

Percent of Beta Spent

This is the percent of the total beta (1-power) that is spent at the corresponding look. It is not the cumulative value.

Percents, proportions, or betas may be entered here. Each of the values is divided by the sum of the values to obtain the proportion of beta that is used at the corresponding look.

Enter Boundaries Tab

The Data tab contains most of the parameters and options for the general setup of the procedure.

Looks and Boundaries

Number of Looks

This is the total number of looks of either type (significance or futility or both).

Equally Spaced

If this box is checked, the Accumulated Information boxes are ignored and the accumulated information is evenly spaced.

Types of Boundaries

This option determines whether or not futility boundaries will be entered.

Futility boundaries are boundaries such that, if crossed at a given look, stop the study in favor of H0.

Accumulated Information

The accumulated information at each look defines the proportion or percent of the sample size that is used at that look.

These values are accumulated information values so they must be increasing.

Proportions, percents, or sample sizes may be entered. All proportions, percents, or sample sizes will be divided by the value at the final look to create an accumulated information proportion for each look.

Significance Boundary

Enter the value of the significance boundary corresponding to the chosen test statistic. These are sometimes called efficacy boundaries.

Futility Boundary

Enter the value of the futility boundary corresponding to the chosen test statistic.

Options Tab

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

Maximum Iterations

Maximum N1 Before Search Termination

Specify the maximum N1 before the search for N1 is aborted.

Since simulations for large sample sizes are very computationally intensive and hence time-consuming, this value can be used to stop searches when N1 is larger than reasonable sample sizes for the study.

This applies only when "Find (Solve For)" is set to N1.

The procedure uses a binary search when searching for N1. If a value for N1 is tried that exceeds this value, and the power is not reached, a warning message will be shown on the output indicating the desired power was not reached.

We recommend a value of at least 20000.

Matching Boundaries at Final Look

Beta Search Increment

For each simulation, when futility bounds are computed, the appropriate beta is found by searching from 0 to 1 by this increment. Smaller increments are more refined, but the search takes longer.

We recommend 0.001 or 0.0001.

Zero Count Adjustment

Zero Count Adjustment Method

Zero cell counts cause many calculation problems. To compensate for this, a small value (called the Zero Count Adjustment Value) may be added either to all cells or to all cells with zero counts. This option specifies whether you want to use the adjustment and which type of adjustment you want to use.

Adding a small value is controversial, but may be necessary. Some statisticians recommend adding 0.5 while others recommend 0.25. We have found that adding values as small as 0.0001 seems to work well.

Zero Count Adjustment Value

Zero cell counts cause many calculation problems. To compensate for this, a small value may be added either to all cells or to all zero cells. This is the amount that is added.

Some statisticians recommend that the value of 0.5 be added to all cells (both zero and non-zero). Others recommend 0.25. We have found that even a value as small as 0.0001 works well.

Example 1 – Power and Output

A clinical trial is to be conducted over a two-year period to compare the proportion response of a new treatment to that of the current treatment. The current response proportion is 0.58. The researchers would like to show that the new treatment is no worse than 0.05 worse than the standard treatment. Although the researchers do not know the true proportion of patients that will survive with the new treatment, they would like to examine the power that is achieved if the proportion under the new treatment is also 0.58. The sample size at the final look is to be 1000 per group. Testing will be done at the 0.05 significance level. A total of five tests are going to be performed on the data as they are obtained. The O'Brien-Fleming (Analog) boundaries will be used. .

Find the power and test boundaries assuming equal sample sizes per arm.

Setup

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

Option	Value
Data Tab	
Find (Solve For)	**Power**
Power	*Ignored since this is the Find setting*
Alpha	**0.05**
N1 (Sample Size Group 1)	**1000**
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
P1.0 (Non-Inferiority Proportion)	**0.53**
P1.1 (Actual Proportion)	**0.58**
P2 (Proportion in Group 2)	**0.58**
Higher Proportions Are	**Better**
Test Type	**Z-Test (Pooled)**
Simulations	**100000**
Looks and Boundaries Tab	
Specification of Looks and Boundaries	**Simple**
Number of Equally Spaced Looks	**5**
Alpha Spending Function	**O'Brien-Fleming Analog**

Output

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

Numeric Results and Plots

Scenario 1 Numeric Results for Group Sequential Testing Proportion Difference = NIM.
Hypotheses: H0: Proportion 1 - Proportion 2 = NIM; H1: Proportion 1 - Proportion 2 > NIM
Test Statistic: Z-Test (Pooled)
Zero Adjustment Method: None
Alpha-Spending Function: O'Brien-Fleming Analog
Beta-Spending Function: None
Futility Boundary Type: None
Number of Looks: 5
Simulations: 100000

Numeric Summary for Scenario 1

Power			Alpha				
Value	95% LCL	95% UCL	Target	Actual	95% LCL	95% UCL	Beta
0.730	0.727	0.733	0.050	0.050	0.049	0.051	0.270

		Average Sample Size								
		Given H0		Given H1						
N1	N2	Grp1	Grp2	Grp1	Grp2	D0	D1	P1.0	P1.1	P2
1000	1000	992	992	819	819	0.0	0.0	0.5	0.6	0.6

223-24 Group-Sequential Non-Inferiority Tests for Two Proportions (Simulation)

Report Definitions

Power is the probability of rejecting a false null hypothesis at one of the looks. It is the total proportion of alternative hypothesis simulations that are outside the significance boundaries.
Power 95% LCL and UCL are the lower and upper confidence limits for the power estimate. The width of the interval is based on the number of simulations.
Target Alpha is the user-specified probability of rejecting a true null hypothesis. It is the total alpha spent.
Alpha or Actual Alpha is the alpha level that was actually achieved by the experiment. It is the total proportion of the null hypothesis simulations that are outside the significance boundaries.
Alpha 95% LCL and UCL are the lower and upper confidence limits for the actual alpha estimate. The width of the interval is based on the number of simulations.
Beta is the probability of accepting a false null hypothesis. It is the total proportion of alternative hypothesis simulations that do not cross the significance boundaries.
N1 and N2 are the sample sizes of each group if the study reaches the final look.
Average Sample Size Given H0 Grp1 and Grp2 are the average or expected sample sizes of each group if H0 is true. These are based on the proportion of null hypothesis simulations that cross the significance or futility boundaries at each look.
Average Sample Size Given H1 Grp1 and Grp2 are the average or expected sample sizes of each group if H1 is true. These are based on the proportion of alternative hypothesis simulations that cross the significance or futility boundaries at each look.
D0, or non-inferiority difference, is the proportion difference between groups (Grp1 - Grp2) assuming the null hypothesis, H0.
D1 is the proportion difference between groups (Grp1 - Grp2) assuming the alternative hypothesis, H1.
P1.0 is the proportion used in the simulations for Group 1 under H0.
P1.1 is the proportion used in the simulations for Group 1 under H1.
P2 is the proportion used in the simulations for Group 2 under H0 and H1.

Summary Statements

Group sequential trials with sample sizes of 1000 and 1000 at the final look achieve 72% power to detect a difference of 0.0 between a treatment group proportion of 0.6 and a control group proportion of 0.6 with a non-inferiority margin of 0.0 at the 0.050 significance level (alpha) using a one-sided Z-Test (Pooled).

Accumulated Information Details for Scenario 1

Look	Accumulated Information Percent	Accumulated Sample Size Group 1	Accumulated Sample Size Group 2	Accumulated Sample Size Total
1	20.0	200	200	400
2	40.0	400	400	800
3	60.0	600	600	1200
4	80.0	800	800	1600
5	100.0	1000	1000	2000

Accumulated Information Details Definitions

Look is the number of the look.
Accumulated Information Percent is the percent of the sample size accumulated up to the corresponding look.
Accumulated Sample Size Group 1 is total number of individuals in group 1 at the corresponding look.
Accumulated Sample Size Group 2 is total number of individuals in group 2 at the corresponding look.
Accumulated Sample Size Total is total number of individuals in the study (group 1 + group 2) at the corresponding look.

Boundaries for Scenario 1

Look	-- Significance Boundary -- Z-Value Scale	P-Value Scale
1	4.127	0.000
2	2.919	0.002
3	2.324	0.010
4	1.963	0.025
5	1.716	0.043

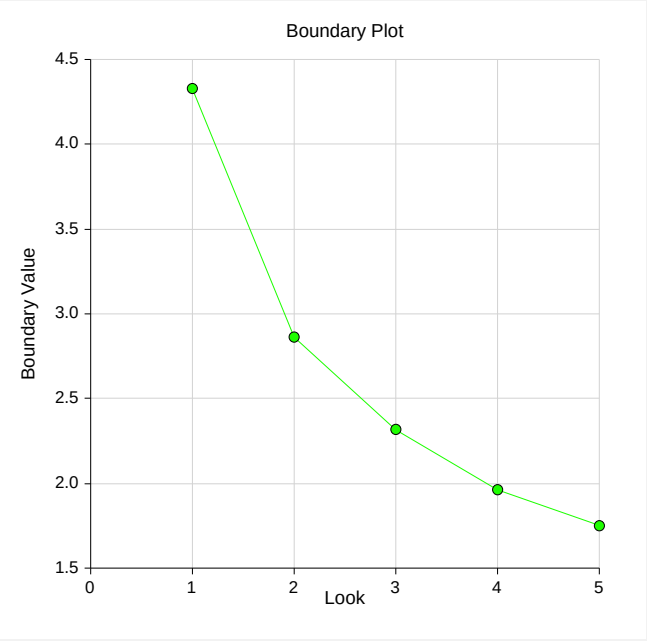
Boundaries Definitions

Look is the number of the look.

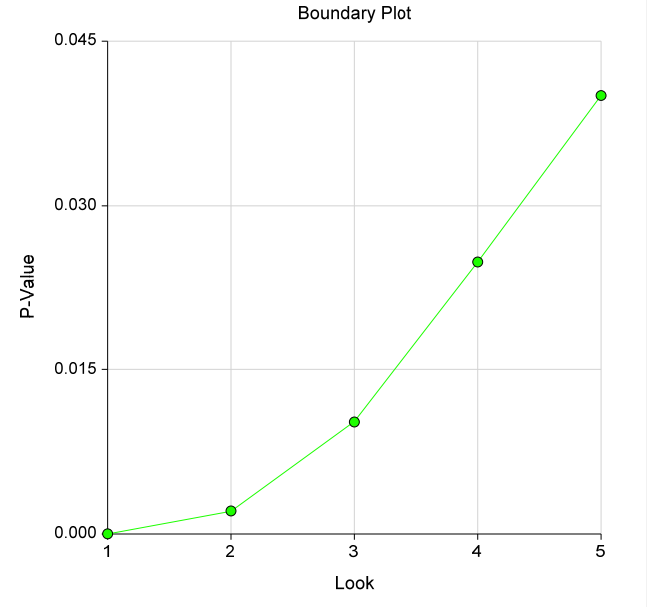
Significance Boundary Z-Value Scale is the value such that statistics outside this boundary at the corresponding look indicate termination of the study and rejection of the null hypothesis. They are sometimes called efficacy boundaries.

Significance Boundary P-Value Scale is the value such that P-Values outside this boundary at the corresponding look indicate termination of the study and rejection of the null hypothesis. This P-Value corresponds to the Z-Value Boundary and is sometimes called the nominal alpha.

Boundary Plot



Boundary Plot - P-Value



Significance Boundaries with 95% Simulation Confidence Intervals for Scenario 1

	Z-Value Boundary			P-Value Boundary		
Look	Value	95% LCL	95% UCL	Value	95% LCL	95% UCL
1	4.127			0.000		
2	2.919	2.873	2.982	0.002	0.001	0.002
3	2.324	2.316	2.333	0.010	0.010	0.010
4	1.963	1.958	1.969	0.025	0.024	0.025
5	1.716	1.711	1.752	0.043	0.040	0.044

Significance Boundary Confidence Limit Definitions

Look is the number of the look.

Z-Value Boundary Value is the value such that statistics outside this boundary at the corresponding look indicate termination of the study and rejection of the null hypothesis. They are sometimes called efficacy boundaries.

P-Value Boundary Value is the value such that P-Values outside this boundary at the corresponding look indicate termination of the study and rejection of the null hypothesis. This P-Value corresponds to the Z-Value Boundary and is sometimes called the nominal alpha.

95% LCL and UCL are the lower and upper confidence limits for the boundary at the given look. The width of the interval is based on the number of simulations.

Alpha-Spending and Null Hypothesis Simulation Details for Scenario 1

	--- Signif. Boundary---		Target		Actual		Proportion	Cum.
				Cum.			H1 Sims	H1 Sims
	Z-Value	P-Value	Spending Function	Spending Function	Alpha	Cum. Alpha	Outside Signif.	Outside Signif.
Look	Scale	Scale	Alpha	Alpha	Spent	Spent	Boundary	Boundary
1	4.127	0.000	0.000	0.000	0.000	0.000	0.001	0.001
2	2.919	0.002	0.002	0.002	0.002	0.002	0.069	0.070
3	2.324	0.010	0.009	0.011	0.009	0.011	0.221	0.291
4	1.963	0.025	0.017	0.028	0.017	0.028	0.252	0.543
5	1.716	0.043	0.022	0.050	0.022	0.050	0.187	0.730

Alpha-Spending Details Definitions

Look is the number of the look.

Significance Boundary Z-Value Scale is the value such that statistics outside this boundary at the corresponding look indicate termination of the study and rejection of the null hypothesis. They are sometimes called efficacy boundaries.

Significance Boundary P-Value Scale is the value such that P-Values outside this boundary at the corresponding look indicate termination of the study and rejection of the null hypothesis. This P-Value corresponds to the Significance Z-Value Boundary and is sometimes called the nominal alpha.

Spending Function Alpha is the intended portion of alpha allocated to the particular look based on the alpha-spending function.

Cumulative Spending Function Alpha is the intended accumulated alpha allocated to the particular look. It is the sum of the Spending Function Alpha up to the corresponding look.

Alpha Spent is the proportion of the null hypothesis simulations resulting in statistics outside the Significance Boundary at this look.

Cumulative Alpha Spent is the proportion of the null hypothesis simulations resulting in Significance Boundary termination up to and including this look. It is the sum of the Alpha Spent up to the corresponding look.

Proportion H1 Sims Outside Significance Boundary is the proportion of the alternative hypothesis simulations resulting in statistics outside the Significance Boundary at this look. It may be thought of as the incremental power.

Cumulative H1 Sims Outside Significance Boundary is the proportion of the alternative hypothesis simulations resulting in Significance Boundary termination up to and including this look. It is the sum of the Proportion H1 Sims Outside Significance Boundary up to the corresponding look.

Run Time: 61.06 seconds.

The values obtained from any given run of this example will vary slightly due to the variation in simulations.

Example 2 – Power with Futility Boundaries

Continuing with Example 1, suppose that the researchers would also like to terminate the study early if there is indication that the treatment is not non-inferior.

Setup

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

Option	Value
Data Tab	
Find (Solve For)	**Power**
Power	*Ignored since this is the Find setting*
Alpha	**0.05**
N1 (Sample Size Group 1)	**1000**
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
P1.0 (Non-Inferiority Proportion)	**0.53**
P1.1 (Actual Proportion)	**0.58**
P2 (Proportion in Group 2)	**0.58**
Higher Proportions Are	**Better**
Test Type	**Z-Test (Pooled)**
Simulations	**100000**
Looks and Boundaries Tab	
Specification of Looks and Boundaries	**Simple**
Number of Equally Spaced Looks	**5**
Alpha Spending Function	**O'Brien-Fleming Analog**
Type of Futility Boundary	**Non-Binding**
Number of Skipped Futility Looks	**0**
Beta Spending Function	**O'Brien-Fleming Analog**

Output

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

Numeric Results and Plots

Scenario 1 Numeric Results for Group Sequential Testing Proportion Difference = NIM.
Hypotheses: H0: Proportion 1 - Proportion 2 = NIM; H1: Proportion 1 - Proportion 2 > NIM
Test Statistic: Z-Test (Pooled)
Zero Adjustment Method: None
Alpha-Spending Function: O'Brien-Fleming Analog
Beta-Spending Function: O'Brien-Fleming Analog
Futility Boundary Type: Non-binding
Number of Looks: 5
Simulations: 100000

Numeric Summary for Scenario 1

Power			Alpha				
Value	95% LCL	95% UCL	Target	Actual	95% LCL	95% UCL	Beta
0.636	0.633	0.639	0.050	0.038	0.037	0.040	0.364

		Average Sample Size								
		Given H0		Given H1						
N1	N2	Grp1	Grp2	Grp1	Grp2	D0	D1	P1.0	P1.1	P2
1000	1000	459	459	674	674	0.0	0.0	0.5	0.6	0.6

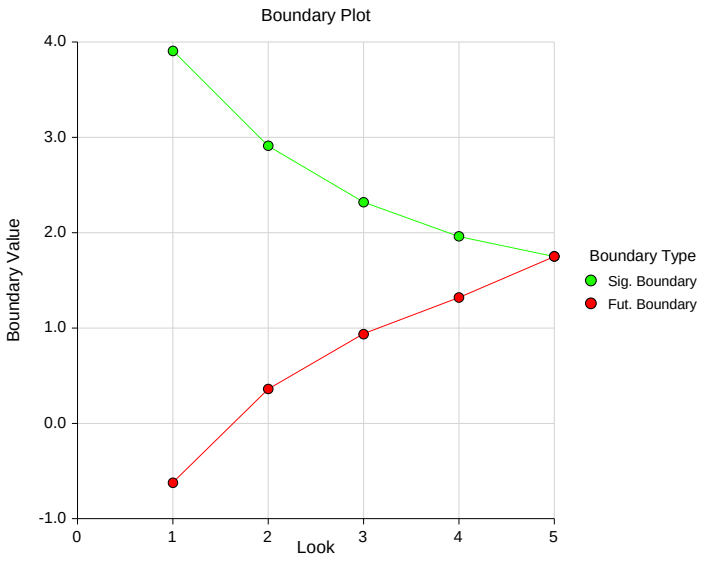
Accumulated Information Details for Scenario 1

	Accumulated Information	Accumulated Sample Size		
Look	Percent	Group 1	Group 2	Total
1	20.0	200	200	400
2	40.0	400	400	800
3	60.0	600	600	1200
4	80.0	800	800	1600
5	100.0	1000	1000	2000

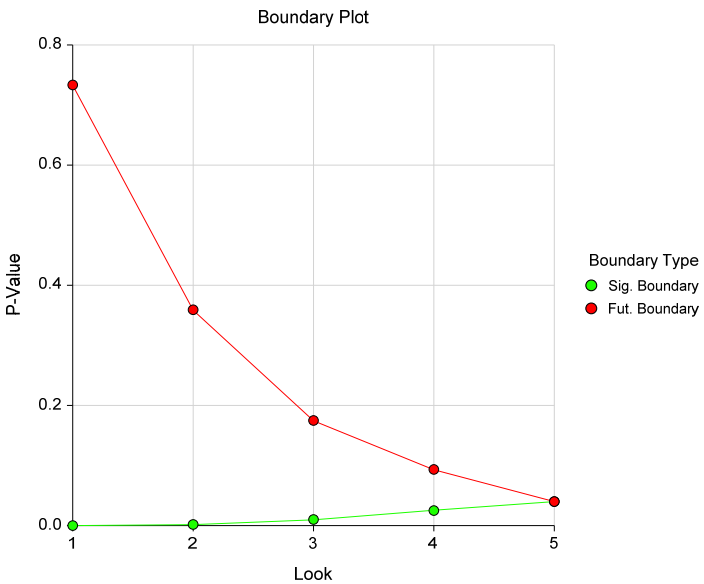
Boundaries for Scenario 1

	Significance Boundary		Futility Boundary	
	Z-Value	P-Value	Z-Value	P-Value
Look	Scale	Scale	Scale	Scale
1	4.127	0.000	-0.709	0.761
2	2.914	0.002	0.358	0.360
3	2.320	0.010	0.929	0.176
4	1.964	0.025	1.314	0.094
5	1.718	0.043	1.718	0.043

Boundary Plot



Boundary Plot - P-Value



Significance Boundaries with 95% Simulation Confidence Intervals for Scenario 1

	--------- Z-Value Boundary ---------			--------- P-Value Boundary ---------		
Look	Value	95% LCL	95% UCL	Value	95% LCL	95% UCL
1	4.127			0.000		
2	2.914	2.857	2.979	0.002	0.001	0.002
3	2.320	2.276	2.328	0.010	0.010	0.011
4	1.964	1.961	1.971	0.025	0.024	0.025
5	1.718	1.712	1.752	0.043	0.040	0.043

Futility Boundaries with 95% Simulation Confidence Intervals for Scenario 1

	Z-Value Boundary			P-Value Boundary		
Look	Value	95% LCL	95% UCL	Value	95% LCL	95% UCL
1	-0.709	-0.711	-0.706	0.761		
2	0.358	0.357	0.358	0.360	0.360	0.361
3	0.929	0.883	0.932	0.176	0.176	0.189
4	1.314	1.312	1.316	0.094	0.094	0.095
5	1.718	1.715	1.722	0.043	0.043	0.043

Alpha-Spending and Null Hypothesis Simulation Details for Scenario 1

	Signif. Boundary		Target		Actual		Proportion H0 Sims Outside	Cum. H0 Sims Outside
Look	Z-Value Scale	P-Value Scale	Spending Function Alpha	Cum. Spending Function Alpha	Alpha Spent	Cum. Alpha Spent	Futility Boundary	Futility Boundary
1	4.127	0.000	0.000	0.000	0.000	0.000	0.230	0.230
2	2.914	0.002	0.002	0.002	0.002	0.002	0.430	0.660
3	2.320	0.010	0.009	0.011	0.009	0.011	0.188	0.848
4	1.964	0.025	0.017	0.028	0.015	0.027	0.081	0.929
5	1.718	0.043	0.022	0.050	0.012	0.038	0.033	0.962

Beta-Spending and Alternative Hypothesis Simulation Details for Scenario 1

	Futility Boundary		Target		Actual		Proportion H1 Sims Outside	Cum. H1 Sims Outside
Look	Z-Value Scale	P-Value Scale	Spending Function Beta	Cum. Spending Function Beta	Beta Spent	Cum. Beta Spent	Signif. Boundary	Signif. Boundary
1	-0.709	0.761	0.043	0.043	0.042	0.042	0.001	0.001
2	0.358	0.360	0.109	0.152	0.109	0.151	0.070	0.071
3	0.929	0.176	0.090	0.242	0.090	0.242	0.219	0.290
4	1.314	0.094	0.069	0.311	0.069	0.310	0.231	0.521
5	1.718	0.043	0.054	0.365	0.054	0.364	0.115	0.636

Run Time: 93.05 seconds.

The values obtained from any given run of this example will vary slightly due to the variation in simulations.

Example 3 – Enter Boundaries

With a set-up similar to Example 2, suppose we wish to investigate the properties of a set of significance (3, 3, 3, 2, 1) and futility (-2, -1, 0, 0, 1) boundaries.

Setup

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

Option	Value
Data Tab	
Find (Solve For)	**Alpha and Power (Enter Boundaries)**
Power	*Ignored since this is the Find setting*
Alpha	*Ignored since this is the Find setting*
N1 (Sample Size Group 1)	**1000**
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
P1.0 (Non-Inferiority Proportion)	**0.53**
P1.1 (Actual Proportion)	**0.58**
P2 (Proportion in Group 2)	**0.58**
Higher Proportions Are	**Better**
Test Type	**Z-Test (Pooled)**
Simulations	**100000**
Enter Boundaries Tab	
Number of Looks	**5**
Types of Boundaries	**Significance and Futility Boundaries**
Significance Boundary	**3 3 3 2 1 (for looks 1 through 5)**
Futility Boundary	**-2 -1 0 0 1 (for looks 1 through 5)**

Output

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

Numeric Results and Plots

Scenario 1 Numeric Results for Group Sequential Testing Proportion Difference = 0.
Hypotheses: H0: Proportion 1 = Proportion 2; H1: Proportion 1 > Proportion 2
Test Statistic: Z-Test (Pooled)
Zero Adjustment Method: Add 0 to each zero
Type of Boundaries: Significance and Futility Boundaries
Number of Looks: 5
Simulations: 100000

Numeric Summary for Scenario 1

Power			Alpha			
Value	95% LCL	95% UCL	Value	95% LCL	95% UCL	Beta
0.885	0.883	0.887	0.146	0.144	0.149	0.115

		Average Sample Size								
		Given H0		Given H1						
N1	N2	Grp1	Grp2	Grp1	Grp2	D0	D1	P1.0	P1.1	P2
1000	1000	736	736	833	833	0.0	0.0	0.5	0.6	0.6

Accumulated Information Details for Scenario 1

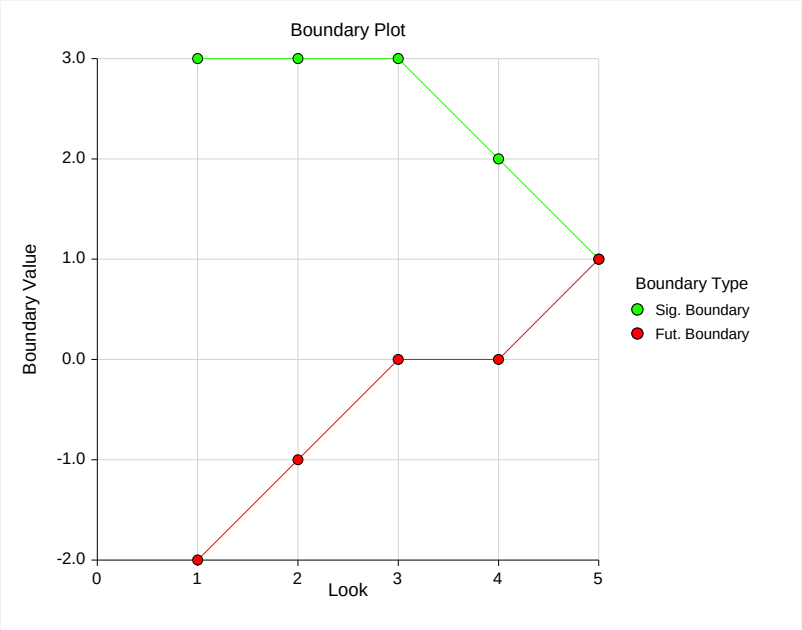
	Accumulated Information	Accumulated Sample Size		
Look	Percent	Group 1	Group 2	Total
1	20.0	200	200	400
2	40.0	400	400	800
3	60.0	600	600	1200
4	80.0	800	800	1600
5	100.0	1000	1000	2000

223-32 Group-Sequential Non-Inferiority Tests for Two Proportions (Simulation)

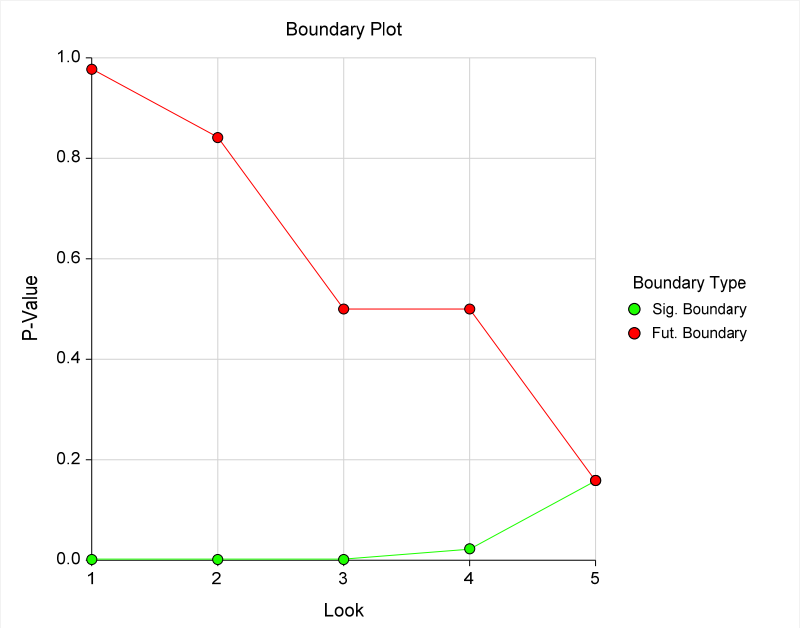
Boundaries for Scenario 1

	-- Significance Boundary --		----- Futility Boundary -----	
	Z-Value	P-Value	Z-Value	P-Value
Look	Scale	Scale	Scale	Scale
1	3.000	0.001	-2.000	0.977
2	3.000	0.001	-1.000	0.841
3	3.000	0.001	0.000	0.500
4	2.000	0.023	0.000	0.500
5	1.000	0.159	1.000	0.159

Boundary Plot



Boundary Plot - P-Value



Alpha-Spending and Null Hypothesis Simulation Details for Scenario 1

Look	--- Signif. Boundary--- Z-Value Scale	P-Value Scale	Alpha Spent	Cum. Alpha Spent	Proportion H0 Sims Outside Futility Boundary	Cum. H0 Sims Outside Futility Boundary
1	3.000	0.001	0.002	0.002	0.024	0.024
2	3.000	0.001	0.001	0.003	0.135	0.159
3	3.000	0.001	0.001	0.004	0.351	0.510
4	2.000	0.023	0.021	0.025	0.082	0.592
5	1.000	0.159	0.121	0.146	0.262	0.854

Beta-Spending and Alternative Hypothesis Simulation Details for Scenario 1

Look	-- Futility Boundary-- Z-Value Scale	P-Value Scale	Beta Spent	Cum. Beta Spent	Proportion H1 Sims Outside Signif. Boundary	Cum. H1 Sims Outside Signif. Boundary
1	-2.000	0.977	0.001	0.001	0.024	0.024
2	-1.000	0.841	0.007	0.008	0.046	0.069
3	0.000	0.500	0.034	0.042	0.061	0.130
4	0.000	0.500	0.006	0.048	0.382	0.512
5	1.000	0.159	0.067	0.115	0.372	0.885

Run Time: 61.00 seconds.

The values obtained from any given run of this example will vary slightly due to the variation in simulations.

Example 4 – Validation using Simulation

With a set-up similar to Example 2, we examine the power and alpha generated by the set of significance (4.127, 2.914, 2.320, 1.964, 1.718) and futility -0.709, 0.358, 0.929, 1.314, 1.718) boundaries.

Setup

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

Option	Value
Data Tab	
Find (Solve For)	**Alpha and Power (Enter Boundaries)**
Power	*Ignored since this is the Find setting*
Alpha	*Ignored since this is the Find setting*
N1 (Sample Size Group 1)	**1000**
N2 (Sample Size Group 2)	**Use R**
R (Sample Allocation Ratio)	**1.0**
P1.0 (Non-Inferiority Proportion)	**0.53**
P1.1 (Actual Proportion)	**0.58**
P2 (Proportion in Group 2)	**0.58**
Higher Proportions Are	**Better**
Test Type	**Z-Test (Pooled)**
Simulations	**100000**
Enter Boundaries Tab	
Number of Looks	**5**
Types of Boundaries	**Significance and Futility Boundaries**
Significance Boundary	**4.127, 2.914, 2.320, 1.964, 1.718**
Futility Boundary	**-0.709, 0.358, 0.929, 1.314, 1.718**

Output

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

Numeric Results and Plots

Numeric Summary for Scenario 1

Power			Alpha			
Value	95% LCL	95% UCL	Value	95% LCL	95% UCL	Beta
0.636	0.633	0.639	0.039	0.037	0.040	0.364

		Average Sample Size								
		Given H0		Given H1						
N1	N2	Grp1	Grp2	Grp1	Grp2	D0	D1	P1.0	P1.1	P2
1000	1000	459	459	673	673	0.0	0.0	0.5	0.6	0.6

The values obtained from any given run of this example will vary slightly due to the variation in simulations. The power and alpha generated with these boundaries are very close to the values of Example 2.

Chapter 225

Inequality Tests for Two Proportions in a Stratified Design (Cochran-Mantel-Haenszel Test)

Introduction

In a stratified design, the subjects are selected from two or more strata which are formed from important covariates such as gender, income level, or marital status. The number of subjects in each of the two groups in each strata is set (fixed) by the design. A separate 2-by-2 table is formed for each stratum. Although response rates may vary among strata, hypotheses about the overall odds ratio can be tested the Cochran-Mantel-Haenszel test. This module allows you to determine power and sample size for such a study.

Technical Details

This procedure is based on the results of Woolson, Bean, and Rojas (1986) which were extended to include a continuity correction by Nam (1992). For more details, consult those articles or chapter 4 in Lachin (2000). We will now briefly summarize these results.

Suppose you are interested in comparing the disease response rates of two groups (treatment and control). Further suppose that response rate is known to be related to another covariate (such as age, race, or gender). It is often desirable to remove the covariate's impact from the comparison of the two proportions. This is accomplished by stratifying on the covariate and forming

hypotheses about a common odds ratio across all strata. Data from such a stratified design may be analyzed by the Cochran-Mantel-Haenszel test.

There are two versions of the Cochran-Mantel-Haenszel test: one that is continuity corrected and one that is not. The continuity-corrected test is more commonly used.

The computation of the test statistic is as follows. Suppose there are *J* strata. The result of each 2-by-2 table may be summarized as follows.

	Groups		
	Group 1	**Group 2**	
Response	**Treatment**	**Control**	**Total**
Yes	x_{1j}	x_{2j}	$x_{.j}$
No	$n_{1j} - x_{1j}$	$n_{2j} - x_{2j}$	$N_j - x_{.j}$
Total	n_{1j}	n_{2j}	N_j

where $j = 1, 2, \ldots, J$ and $N = \sum_{j=1}^{J} N_j$.

The parameters of interest are the success proportions p_{1j} and p_{2j}. These parameters are estimated by

$$\hat{p}_{1j} = \frac{x_{1j}}{n_{1j}} \text{ and } \hat{p}_{2j} = \frac{x_{2j}}{n_{2j}}$$

The odds of response in each of the two groups in each strata is given by

$$o_{1j} = \frac{p_{1j}}{1 - p_{1j}} \text{ and } o_{2j} = \frac{p_{2j}}{1 - p_{2j}}$$

The strata odds ratio ψ_j is calculated using the equation

$$\psi_j = \frac{o_{1j}}{o_{2j}} = \frac{\left(\frac{p_{1j}}{1 - p_{1j}}\right)}{\left(\frac{p_{2j}}{1 - p_{2j}}\right)}$$

In the sequel, it is assumed that the strata odds ratios are all equal. That is, it is assumed that $\psi_1 = \psi_2 = \cdots \psi_J = \psi$. Solving this relationship for p_{1j} in terms of ψ and p_{2j} gives

$$p_{1j} = \frac{\psi\, p_{2j}}{1 - p_{2j} + \psi\, p_{2j}}$$

If values for the odds ratio under the null hypothesis (ψ_0), under the alternative hypothesis (ψ_1), and p_{2j} are specified, values for p_{1j} under the null hypothesis (p_{1j0}) and the alternative hypothesis (p_{1j1}) can be calculated as follows

$$p_{1j0} = \frac{\psi_0 p_{2j}}{1 - p_{2j} + \psi_0 p_{2j}}, \quad j = 1, 2, \cdots, J$$

$$p_{1j1} = \frac{\psi_1 p_{2j}}{1 - p_{2j} + \psi_1 p_{2j}}, \quad j = 1, 2, \cdots, J$$

Assuming a common odds ratio across all strata of ψ (that is, assuming $\psi_1 = \psi_2 = \cdots \psi_J = \psi$), hypotheses of the form $H_0: \psi \le \psi_0$ versus $H_1: \psi > \psi_0$ may be tested using Cochran's U statistic (Woolson et al. 1986, page 928)

$$U_G = \sum_{j=1}^{J} w_j \left\{ \left(\hat{p}_{1j} - \hat{p}_{2j} \right) - \left(p_{1j0} - p_{2j} \right) \right\}, \quad \text{where } w_j = \frac{n_{1j} n_{2j}}{N_j}$$

Note that when $\psi_0 = 1$, U_G reduces to

$$U_0 = \sum_{j=1}^{J} w_j \left(\hat{p}_{1j} - \hat{p}_{2j} \right).$$

The value U_0 is commonly used to form the Cochran-Mantel-Haenszel statistic. U_G is an extension of this statistic which allows $\psi_0 \ne 1$.

The calculation of the asymptotically normal test statistic, z_c, may or may not include a continuity correction factor depending on whether the parameter *cc* is set to 1/2 or 0. The formula for z_{CMH} is

$$z_{CMH} = \frac{U_G - cc}{\sqrt{v_0(U_G)}}$$

where

$$v_0(U_G) = \begin{cases} \sum_{j=1}^{J} w_j^2 \left\{ \frac{\hat{p}_{1j}\left(1 - \hat{p}_{1j}\right)}{n_{1j}} + \frac{\hat{p}_{2j}\left(1 - \hat{p}_{2j}\right)}{n_{2j}} \right\} & \text{if } \psi_0 \ne 1 \\ \sum_{j=1}^{J} w_j \hat{p}_j \left(1 - \hat{p}_j\right) & \text{if } \psi_0 = 1 \end{cases}$$

$$\hat{p}_j = \frac{x_{.j}}{N_j}$$

The name *Cochran-Mantel-Haenszel test* actually refers to two tests: the Cochran test and the Mantel-Haenszel test. The difference is between these test is that Cochran's test uses $v_0(U_G)$ to estimate the unconditional variance assuming that the group sample sizes are fixed, while the Mantel-Haenszel test replaces $v_0(U_G)$ with an estimate of the conditional variance of U

assuming that both row and column marginals are fixed. Asymptotically the two variances are equivalent, so the test is often called the Cochran-Mantel-Haenszel statistic.

Power Calculations

The asymptotic power of z_{CMH} for testing a one-sided hypothesis of the form $H_0: \psi \le \psi_0$ versus $H_1: \psi > \psi_0$ is

$$Power = 1 - \Phi\left(\frac{z_{1-\alpha}\sqrt{V_0(U_G)} - E(U_G) + cc}{\sqrt{V_1(U_G)}}\right)$$

where

$$E(U_G) = \sum_{j=1}^{J} w_j \left\{\left(p_{1j1} - p_{2j}\right) - \left(p_{1j0} - p_{2j}\right)\right\}$$

$$V_0(U_G) = \begin{cases} \sum_{j=1}^{J} w_j^2 \left\{\frac{p_{1j0}(1-p_{1j0})}{n_{1j}} + \frac{p_{2j}(1-p_{2j})}{n_{2j}}\right\} & \text{if } \psi_0 \ne 1 \\ \sum_{j=1}^{J} w_j \bar{p}_j (1-\bar{p}_j) & \text{if } \psi_0 = 1 \end{cases}$$

$$\bar{p}_j = p_{1j1}\left(\frac{n_{1j}}{N_j}\right) + p_{2j}\left(\frac{n_{2j}}{N_j}\right)$$

$$V_1(U_G) = \sum_{j=1}^{J} w_j^2 \left\{\frac{p_{1j1}(1-p_{1j1})}{n_{1j}} + \frac{p_{2j}(1-p_{2j})}{n_{2j}}\right\}$$

Note that Woolson et al. (1986) and Nam (1992) give results for the usual case when $\psi_0 = 1$. The above results are our extension to the important case when $\psi_0 \ne 1$. We could not find published results for this case, so we have made this extension. When published results become available, we will adopt those results. If you have $\psi_0 \ne 1$, you must use U_G, rather than U_0, in the calculation of the test statistic.

Similar calculations may also be made for testing the other one-sided hypothesis $H_0: \psi \ge \psi_0$ versus $H_1: \psi < \psi_0$ and the two-sided hypothesis $H_0: \psi = \psi_0$ versus $H_1: \psi \ne \psi_0$.

Procedure Options

This section describes the options that are specific to this procedure. These are located on the Data 1 and Data 2 tabs. For more information about the options of other tabs, go to the Procedure Window chapter.

Data 1 and Data 2 Tabs

The Data tabs contain most of the parameters and options of interest for this procedure.

Solve For

Find (Solve For)

This option specifies the parameter to be solved for using the other parameters. The parameters that may be selected are *OR1*, *Alpha*, *Power and Beta*, or *N*. In most cases, you will select either *Power and Beta* or *N*.

Select *N* when you want to calculate the sample size needed to achieve a given power and alpha level.

Select *Power and Beta* when you want to calculate the power.

Error Rates

Power or Beta

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

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

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

Alpha (Significance Level)

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

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

Sample Size

M (Sample Size Multiplier)

M and the values of R1 and R2 are used to calculate the group sample sizes within each strata using the formulas N1 = M x R1 and N2 = M x R2. The total sample size, N, is found by summing N1 and N2 across all strata. Note that fractional values for N, N1, and N2 will usually result. In practice these values are rounded up to the next integer value.

One or more values, separated by blanks or commas, may be entered. A separate analysis is performed for each value.

Using M as the Group Size

To use M as the sample size in each group, the values of R1 and R2 must each be set to one.

Using M as the Strata Size

To use M as the sample size in each strata, the values of R1 and R2 must sum to one within each strata. For example, suppose M = 30 and R1 = R2 = 0.5. The values of N1 and N2, the group sample sizes within a stratum, will be 0.5 x 30 = 15. Thus, the total sample size within the strata is 15 + 15 = 30.

Using M as Total Sample Size

To use M as the total sample size across all strata, the values of R1 and R2 must sum to one across all values. Note that the resulting value of N may not exactly equal M because of rounding.

For example, suppose there are three strata with R1 = 0.1, 0.2, and 0.2 and R2 = 0.1, 0.3, and 0.1. (Note that these values sum to one.) If M were 100, then the values of N1 would be 10, 20, and 20 and the values of N2 would be 10, 30, and 10. These sum to 100, the value of M.

Effect Size

OR1 (Odds Ratio|H1)

This option specifies the odds ratio of the two proportions P1 and P2 at which the power is to be computed. This odds ratio is used to specify the size of the difference between the two proportions at which the power is calculated.

You may enter a range of values such as *0.5 0.6 0.7 0.8* or *1.25 to 2.0 by 0.25*.

Odds ratios must greater than zero.

OR0 (Odds Ratio|H0)

Specify the odds ratio under the null hypothesis, H0. For each strata, this value is used with the value of Pr(Success) to calculate the probability of obtaining a success in group one (the treatment group) assuming the null hypothesis. In the standard Cochran-Mantel-Haenszel test, this value is assumed to be (and should be entered as) one. If you enter a value other than one, your data analysis should use the more general test statistic.

Note that OR0 must be greater than zero and cannot be equal to OR1.

Test

Alternative Hypothesis (H1)

This option specifies whether a one-sided or two-sided hypothesis is analyzed.

One-Sided (H1: OR1 < OR0) refers to a one-sided test in which the alternative hypothesis is of the form H1: *OR1 < OR0*.

One-Sided (H1: OR1 > OR0) refers to a one-sided test in which the alternative hypothesis is of the form H1: *OR1 > OR0*.

Two-Sided refers to a two-sided test in which the alternative hypothesis is of the type *H1: OR1 <> OR0*. Here '<>' means 'is not equal to' or 'is less than or greater than'.

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

Continuity Correction

Specify whether to use the Continuity Correction. When selected, a continuity correction is made that is recommend by Fleiss et al. (2003) to make the alpha and beta values achieved by the test more accurate.

Strata Information

Strata

This option specifies the number of strata specified on this line. Usually, you will enter a '1' to specify a single stratum, or you will enter a '0' to ignore this line. However, this option lets you specify several strata at once.

The total number of strata is equal to the sum of these values.

R1 = N1 / M, R2 = N2 / M

R1 and R2 are used to obtain the sample sizes in groups 1 (treatment) and 2 (control) within a strata using the formulas N1 = R1 x M and N2 = R2 x M. The only limitation on R1 and R2 is that they are positive (non-zero) values. See the comments under M for more information.

Note that only a single value may be entered for this parameter—you cannot enter several values.

Pr(Success)

This is the baseline probability of a successful response. This value is used with OR1 to calculate the probability of a success in group 1 (the treatment or numerator group).

Since this value is a probability, it must be between zero and one.

Note that only one value may be entered here.

Example 1 – Finding Power

Nam (1992) discusses a case-control study investigating the possible association between chlorinated water and colon cancer among males in Iowa. Since age is known to affect colon cancer rates, the population is stratified into four age groups with weights of 10%, 40%, 35%, and 15%. An equal number of cases and controls will be selected in each age-group. Prior studies had shown the probability of chlorinated water exposure among non-cancer subjects was 0.75, 0.70, 0.65, and 0.60, respectively, among the four age groups. The significance level is set to 0.05. The investigators want to consider various total sample sizes from 50 to 500. They also want to consider odds ratios of 2 and 3.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Two Proportions in a Stratified Design (Cochran/Mantel-Haenszel Test)** procedure window by expanding **Proportions**, then clicking on **Stratified**, and then clicking on **Tests for Two Proportions in a Stratified Design (Cochran/Mantel-Haenszel Test)**. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power	*Ignored since this is the Find setting*
Alpha	**0.05**
M (Sample Size Multiplier)	**50 to 500 by 50**
OR0 (Odds Ratio | H0)	**1**
OR1 (Odds Ratio | H1)	**2 3**
H1 (Alternative Hypothesis)	**One-Sided (H1:OR1>OR0)**
Continuity Correction	**Checked**
Strata(1)	**1**
R1(1)	**0.05 (half of 10%)**
R2(1)	**R1**
Pr(Success)(1)	**0.75**
Strata(2)	**1**
R1(2)	**0.20 (half of 40%)**
R2(2)	**R1**
Pr(Success)(2)	**0.70**
Strata(3)	**1**
R1(3)	**0.175 (half of 35%)**
R2(3)	**R1**
Pr(Success)(3)	**0.65**
Strata(4)	**1**
R1(4)	**0.075 (half of 15%)**
R2(4)	**R1**
Pr(Success)(4)	**0.60**

Annotated Output

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

Numeric Results

Numeric Results of Cochran-Mantel-Haenszel Test of an Odds Ratio
H0: OR1=OR0. H1: OR1>OR0. Test: Continuity-Corrected Z-Test.

Power	Total Sample Size (N)	Sample Size Multiplier (M)	Sample Size of Group 1 (N1)	Sample Size of Group 2 (N2)	H0 Odds Ratio (OR0)	Actual Odds Ratio (OR1)	Signif. Level Alpha	Beta
0.1783	50	50.000	25	25	1.000	2.000	0.0500	0.8217
0.3505	100	100.000	50	50	1.000	2.000	0.0500	0.6495
0.4992	150	150.000	75	75	1.000	2.000	0.0500	0.5008
0.6215	200	200.000	100	100	1.000	2.000	0.0500	0.3785
0.7186	250	250.000	125	125	1.000	2.000	0.0500	0.2814
0.7937	300	300.000	150	150	1.000	2.000	0.0500	0.2063
0.8506	350	350.000	175	175	1.000	2.000	0.0500	0.1494
0.8929	400	400.000	200	200	1.000	2.000	0.0500	0.1071
0.9239	450	450.000	225	225	1.000	2.000	0.0500	0.0761
0.9464	500	500.000	250	250	1.000	2.000	0.0500	0.0536
0.3356	50	50.000	25	25	1.000	3.000	0.0500	0.6644
0.6337	100	100.000	50	50	1.000	3.000	0.0500	0.3663
0.8151	150	150.000	75	75	1.000	3.000	0.0500	0.1849
0.9121	200	200.000	100	100	1.000	3.000	0.0500	0.0879
0.9601	250	250.000	125	125	1.000	3.000	0.0500	0.0399
0.9825	300	300.000	150	150	1.000	3.000	0.0500	0.0175
0.9925	350	350.000	175	175	1.000	3.000	0.0500	0.0075
0.9969	400	400.000	200	200	1.000	3.000	0.0500	0.0031
0.9987	450	450.000	225	225	1.000	3.000	0.0500	0.0013
0.9995	500	500.000	250	250	1.000	3.000	0.0500	0.0005

Report Definitions
'Power' is the probability of rejecting a false null hypothesis. It should be close to one.
'N' is the total sample size summed across all groups and strata.
'M' is the factor by which the values of R1 and R2 are multiplied.
'N1 and N2' are the sample sizes from groups 1 and 2 summed across all strata.
'OR0' is the odds ratio [P1/(1-P1)] / [P2/(1-P2)] assuming the null hypothesis (H0).
'OR1' is the value of the odds ratio at which the power is computed.
'Alpha' is the probability of rejecting a true null hypothesis.
'Beta' is the probability of accepting a false null hypothesis.
In a treatment vs. control design, the treatment group is 1 and the control group is 2.

Summary Statements
A stratified design, which divides the sample among 4 strata, is analyzed using the one-sided, Cochran-Mantel-Haenszel test. Sample sizes, summed across all strata, of 25 in group 1 (treatment group) and 25 in group 2 (control group) achieve 18% power to reject the odds ratio set by the null hypothesis of 1.000 when the odds ratio is actually 2.000. The significance level of the test was set at 0.0500.

Sample Sizes: N, N1, and N2

The value of N is the sum of N1 and N2. The values of N1 and N2 are found by summing the individual strata-group sample sizes. These are found by multiplying R1 and R2 by M.

Note that this multiplication will usually result in fractional sample sizes across the strata. As a practical matter, we recommend rounding each fractional value up to the next integer when implementing a given design.

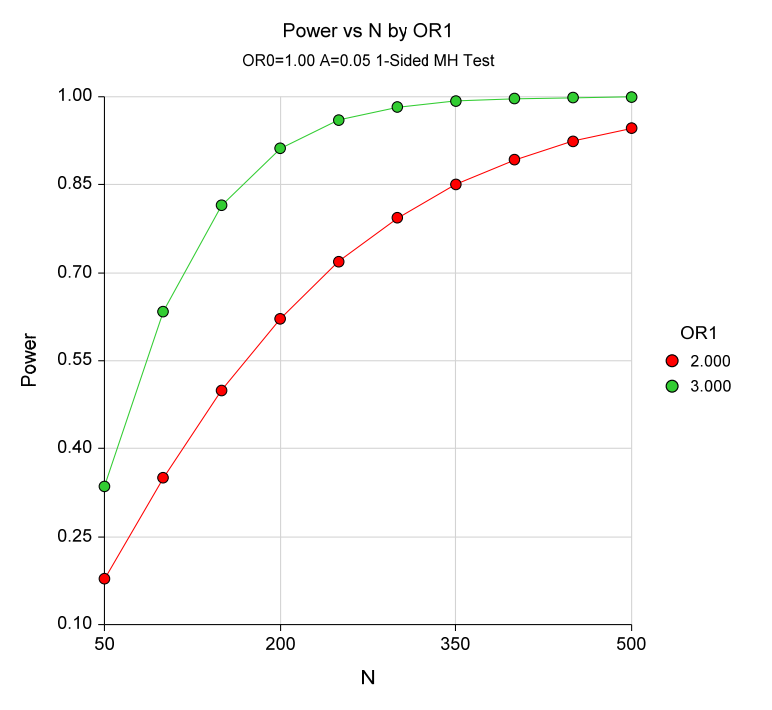
Strata-Detail Report

Strata-Detail Report

Number of Strata	Proportion of Total Sample in each Strata	Proportion of this Strata in Group 1	Proportion of this Strata in Group 2	Group 1 Multiplier (R1)	Group 2 Multiplier (R2)	Strata Probability of Success
1	0.1000	0.5000	0.5000	0.050	0.050	0.7500
1	0.4000	0.5000	0.5000	0.200	0.200	0.7000
1	0.3500	0.5000	0.5000	0.175	0.175	0.6500
1	0.1500	0.5000	0.5000	0.075	0.075	0.6000

This report shows the values of the individual, strata-level parameters that were used. These parameters are the same for all rows of the Numerical Results Report (shown above), so they are only displayed once.

Plots Section



The values from the Numerical Results Report are displayed in this scatter plot. This chart provides a quick view of the power that is achieved for various sample sizes.

Example 2 – Validation using Nam

To validate the procedure, we will compare ***PASS's*** results to those on page 392 of Nam (1992). Most of the settings in this example are the same as those of Example 1, except that the power is 90% and the odds ratio is 3. Nam (1992) found the necessary sample sizes to be 192 for the corrected test and 171 for the uncorrected test.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Two Proportions in a Stratified Design (Cochran/Mantel-Haenszel Test)** procedure window by expanding **Proportions**, then clicking on **Stratified**, and then clicking on **Tests for Two Proportions in a Stratified Design (Cochran/Mantel-Haenszel Test)**. You may then make the appropriate entries as listed below, or open **Example2a** or **Example2b** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**N (Sample Size)**
Power	**0.90**
Alpha	**0.05**
M (Sample Size Multiplier)	*Ignored since this is the Find setting*
OR0 (Odds Ratio | H0)	**1**
OR1 (Odds Ratio | H1)	**3**
H1 (Alternative Hypothesis)	**One-Sided (H1:OR1>OR0)**
Continuity Correction	**Checked**
Strata(1)	**1**
R1(1)	**0.05 (half of 10%)**
R2(1)	**R1**
Pr(Success)(1)	**0.75**
Strata(2)	**1**
R1(2)	**0.20 (half of 40%)**
R2(2)	**R1**
Pr(Success)(2)	**0.70**
Strata(3)	**1**
R1(3)	**0.175 (half of 35%)**
R2(3)	**R1**
Pr(Success)(3)	**0.65**
Strata(4)	**1**
R1(4)	**0.075 (half of 15%)**
R2(4)	**R1**
Pr(Success)(4)	**0.60**

Output

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

Numeric Results

Numeric Results of Cochran-Mantel-Haenszel Test of an Odds Ratio
H0: OR1=OR0. H1: OR1>OR0. Test: Continuity-Corrected Z-Test.

Power	Total Sample Size (N)	Sample Size Multiplier (M)	Sample Size of Group 1 (N1)	Sample Size of Group 2 (N2)	H0 Odds Ratio (OR0)	Actual Odds Ratio (OR1)	Signif. Level Alpha	Beta
0.9000	192	191.538	96	96	1.000	3.000	0.0500	0.1000

The value of 192 agrees exactly with that of Nam (1992).

If you uncheck the Continuity Correction option and rerun the analysis, you will get the following results.

Numeric Results – No Continuity Correction

Numeric Results of Cochran-Mantel-Haenszel Test of an Odds Ratio
H0: OR1=OR0. H1: OR1>OR0. Test: Uncorrected Z-Test.

Power	Total Sample Size (N)	Sample Size Multiplier (M)	Sample Size of Group 1 (N1)	Sample Size of Group 2 (N2)	H0 Odds Ratio (OR0)	Actual Odds Ratio (OR1)	Signif. Level Alpha	Beta
0.9000	171	170.741	85	85	1.000	3.000	0.0500	0.1000

The value of 171 agrees exactly with that of Nam (1992).

Example 3 – Finding Power of a Completed Experiment

Suppose you want to find the power for a completed experiment in which the individual strata sample sizes are known. In this example there are three strata with success probabilities 0.72, 0.66, and 0.69. The sample sizes for the treatment group in each stratum are 102, 113, and 97. The sample sizes for the control group in each stratum are 98, 110, and 114. The experiment was designed to detect an odds ratio of at least 1.5 with alpha equal to 0.05 for a one-sided test.

To calculate the power in this situation, we set M to 1 and enter the sample sizes directly into R1 and R2.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Two Proportions in a Stratified Design (Cochran/Mantel-Haenszel Test)** procedure window by expanding **Proportions**, then clicking on **Stratified**, and then clicking on **Tests for Two Proportions in a Stratified Design (Cochran/Mantel-Haenszel Test)**. You may then make the appropriate entries as listed below, or open **Example 3** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power	*Ignored since this is the Find setting*
Alpha	**0.05**
M (Sample Size Multiplier)	**1**
OR0 (Odds Ratio | H0)	**1**
OR1 (Odds Ratio | H1)	**1.5**
H1 (Alternative Hypothesis)	**One-Sided (H1:OR1>OR0)**
Continuity Correction	**Checked**
Strata(1)	**1**
R1(1)	**102**
R2(1)	**98**
Pr(Success)(1)	**0.72**
Strata(2)	**1**
R1(2)	**113**
R2(2)	**110**
Pr(Success)(2)	**0.66**
Strata(3)	**1**
R1(3)	**97**
R2(3)	**114**
Pr(Success)(3)	**0.69**

Output

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

Numeric Results

Numeric Results of Cochran-Mantel-Haenszel Test of an Odds Ratio
H0: OR1=OR0. H1: OR1>OR0. Test: Continuity-Corrected Z-Test.

Power	Total Sample Size (N)	Sample Size Multiplier (M)	Sample Size of Group 1 (N1)	Sample Size of Group 2 (N2)	H0 Odds Ratio (OR0)	Actual Odds Ratio (OR1)	Signif. Level Alpha	Beta
0.6980	634	1.000	312	322	1.000	1.500	0.0500	0.3020

The power to detect an odds ratio of 1.5 is only 0.6980 in this experiment.

Chapter 230

Inequality Tests for Two Proportions in a Cluster-Randomized Design

Introduction

A *cluster (group) randomized design* is one in which whole units, or clusters, of subjects are randomized to the groups rather than the individual subjects in those clusters. However, the conclusions of the study concern individual subjects rather than the clusters. Examples of clusters are families, school classes, neighborhoods, and hospital wards.

Cluster-randomized designs are often adopted when there is a high risk of contamination if cluster members were randomized individually. For example, it may be difficult for an instructor to use two methods of teaching individuals in the same class. The price of randomizing by clusters is a loss of efficiency--the number of subjects needed to obtain a certain level of precision in a cluster-randomized trial is usually much larger than the number needed when the subjects are randomized individually. Hence, the standard methods of sample size estimation cannot be used.

Three Procedures Documented Here

There are three procedures in the menus that use the program module described in this chapter. These procedures are identical except for the type of parameterization. The parameterization can be in terms of proportions, differences in proportions, or ratios of proportions. Each of these options is listed separately on the menus.

Technical Details

Our formulation comes from Donner and Klar (2000). Denote a binary observation by Y_{gkm} where g = 1 or 2 is the group, k = 1, 2, …, K_g is a cluster within group g, and m = 1, 2, …, M_g is an individual in cluster k of group g. The results that follow assume an equal number of individuals per cluster. When the number of subjects from cluster to cluster are about the same, the power and sample size values should be fairly accurate. In these cases, the average number of subjects per cluster can be used.

The statistical hypothesis that is tested concerns the difference between the two group proportions, p_1 and p_2. When necessary, we assume that group 1 is the treatment group and group 2 is the control group. With a simple modification, all of the large-sample sample size formulas that are listed in the module for testing two proportions can be used here. When the individual subjects are randomly assigned to one of the two groups, the variance of the sample proportion is

$$\sigma_{S,g}^2 = \frac{p_g\left(1-p_g\right)}{n_g}$$

When the randomization is by clusters of subjects, the variance of the sample proportion is

$$\begin{aligned}\sigma_{C,g}^2 &= \frac{p_g\left(1-p_g\right)\left(1+\left(m_g-1\right)\rho\right)}{k_g m_g} \\ &= \sigma_{S,g}^2\left[1+\left(m_g-1\right)\rho\right] \\ &= F_{g,\rho}\sigma_{S,g}^2\end{aligned}$$

The factor $\left[1+\left(m_g-1\right)\rho\right]$ is called the *inflation factor*. The Greek letter ρ is used to represent the *intracluster correlation coefficient (ICC)*. This correlation may be thought of as the simple correlation between any two subjects within the same cluster. If we stipulate that ρ is positive, it may also be interpreted as the proportion of total variability that is attributable to differences between clusters. This value is critical to the sample size calculation.

All of the asymptotic formulas that were used in comparing two proportions may be used with cluster-randomized designs as well, as long as an adjustment is made for the inflation factor. The basic form of the z-test becomes

$$z = \frac{\left|\hat{D}-\delta_0\right|}{\hat{\sigma}_{\hat{D}}\left(\delta_0\right)}$$

where

$$\hat{D} = \hat{p}_1 - \hat{p}_2$$

$$\delta_0 = p_1 - p_2 | H_0$$

$$\hat{\sigma}_{\hat{D}}\left(\delta_0\right) = \sqrt{\frac{\tilde{p}_1\left(1-\tilde{p}_1\right)F_{1,\rho}}{n_1} + \frac{\tilde{p}_2\left(1-\tilde{p}_2\right)F_{2,\rho}}{n_2}}$$

The quantities $\tilde{p}_1$ and $\tilde{p}_2$ are the maximum likelihood estimates constrained by $\tilde{p}_1 - \tilde{p}_2 = \delta_0$.

Power Calculations

A large sample approximation may be used that is most accurate when the values of n_1 and n_2 are large. The large approximation is made by replacing the values of $\hat{p}_1$ and $\hat{p}_2$ in the *z* statistic with the corresponding values of p_1 and p_2 under the alternative hypothesis, and then computing the results based on the normal distribution.

Note that in this case, exact calculations are not possible.

Procedure Options

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

Data Tab (Common Options)

The Data tab contains the parameters associated with this test such as the proportions, sample sizes, alpha, and power. This chapter covers three procedures, each of which has different options. This section documents options that are common to all three procedures. Later, unique options for each procedure will be documented.

Solve For

Find (Solve For)

This option specifies the parameter to be solved for using the other parameters. The parameters that may be selected are *P1.1*, *Alpha*, *Power and Beta*, *K1*, and *M1*. Under most situations, you will select either *Power and Beta* or *K1*.

Select *K1* when you want to calculate the sample size needed to achieve a given power and alpha level.

Select *Power and Beta* when you want to calculate the power of an experiment.

Error Rates

Power or Beta

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

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

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

Alpha (Significance Level)

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

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

Sample Size – Treatment (Group 1)

K1 (Clusters Group 1)

Enter a value (or range of values) for the number of clusters in this group. You may enter a range of values such as *10 to 20 by 2*. The sample size for this group is equal to the number of clusters times the number of subjects per cluster.

M1 (Items Group 1)

This is the average number of items (subjects) per cluster in group one. This value must be a positive number that is at least one. You can use a list of values such as *100 150 200*.

Sample Size – Control (Group 2)

K2 (Clusters Group 2)

This is the number of clusters in group two. The sample size for this group is equal to the number of clusters times the number of subjects per cluster. This value must be a positive number.

If you simply want a multiple of the value for group one, you would enter the multiple followed by *K1*, with no blanks. If you want to use *K1* directly, you do not have to premultiply by *1*. For example, all of the following are valid entries:*10 K1 2K1 0.5K1*.

You can use a list of values such as *10 20 30* or *K1 2K1 3K1*.

M2 (Items Group 2)

This is the average number of items (subjects) per cluster in group two. This value must be at least one.

If you simply want a multiple of the value for group one, you would enter the multiple followed by *M1*, with no blanks. If you want to use *M1* directly, you do not have to premultiply by *1*. For example, all of the following are valid entries: *10 M1 2M1 0.5M1*.

You can use a list of values such as *10 20 30* or *M1 2M1 3M1*.

Effect Size – Control (Group 2)

P2 (Control Group Proportion)

Specify the value of p_2, the control, baseline, or standard group's proportion. The null hypothesis is that the two proportions differ by a specified amount (See *Specify Group 1 Proportion using* below).

Since p_2 is a proportion, these values must be 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 0.1*.

Effect Size – Intracluster Correlation

ICC (Intracluster Correlation)

Enter a value (or range of values) for the intracluster correlation. This correlation may be thought of as the simple correlation between any two observations in the same cluster. It may also be thought of as the proportion of total variance in the observations that can be attributed to difference between clusters.

Although the actual range for this value is from zero to one, typical values range from 0.002 to 0.05.

Test

Test Type

Specify which test statistic is used in searching and reporting. We recommend the likelihood score test.

Data Tab (Proportions)

This section documents options that are used when the parameterization is in terms of the values of the two proportions, P1 and P2. P1.0 is the value of the P1 assumed by the null hypothesis and P1.1 is the value of P1 at which the power is calculated.

Effect Size – Treatment (Group 1)

P1.0 (Group 1 Proportion |H0)

This option specifies the value of the group 1 proportion given the null hypothesis. The power calculations assume that P1.0 is the value of P2 under the null hypothesis. In this non-null case, the value of P1.0 is not equal to P2 as it is in the null case.

You may enter a range of values such as *0.03 0.05 0.10* or *0.01 to 0.05 by 0.01*.

Proportions must be between zero and one. They cannot take on the values zero or one.

P1.1 (Group 1 Proportion |H1)

This is the value of P1 under the alternative hypothesis. It is written P1.1. The power calculations assume that this is the actual value of the proportion.

You may enter a range of values such as *0.1 0.2 0.3* or *0.1 to 0.9 by 0.1*.

Note that values must be between zero and one.

Test

Alternative Hypothesis (H1)

This option specifies whether a one-sided or two-sided hypothesis is analyzed.

One-Sided (H1: P1<P2) refers to a one-sided test in which the alternative hypothesis is of the form H1: P1<P2.

One-Sided (H1:D1>D0) refers to a one-sided test in which the alternative hypothesis is of the form H1: P1>P2.

Two-Sided refers to a two-sided test in which the alternative hypothesis is of the type H1: P1<>P2+D0. Here '<>' means unequal.

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

Data Tab (Differences)

This section documents options that are used when the parameterization is in terms of the difference, P1 – P2. P1.0 is the value of P1 assumed by the null hypothesis and P1.1 is the value of P1 at which the power is calculated. Once P2, D0, and D1 are given, the values of P1.1 and P1.0 can be calculated.

Effect Size – Differences

D0 (Difference|H0 = P1.0 – P2)

This option specifies the difference between the two proportions given in the null hypothesis, H0. This difference is used with P2 to calculate the value of P1.0 using the formula: P1.0 = P2 + D0. Note that P1.0 here means the value of P1 under H0.

Differences must be between -1 and 1. They cannot take on the values -1, 0, or 1.

The power calculations use P1.0 as the value of the proportion in group 2 (the experimental or treatment group) under the null hypothesis. In the non-null case, the value of P1.0 is not equal to P2 as it is in the null case.

You may enter a range of values such as *0.03 0.05 0.10* or *0.01 to 0.05 by 0.01*.

D1 (Difference|H1 = P1.1 – P2)

This option specifies the difference between P1.1 and P2. This difference is used with P2 to calculate the value of P1.1 using the formula: P1.1 = D1 + P2. Note that P1.1 here means the value of P1 under H1. Differences must be between -1 and 1. They cannot take on the values -1 or 1.

The power calculations assume that P1.1 is the actual value of the proportion in group 2 (experimental or treatment group).

You may enter a range of values such as *0.03 0.05 0.10* or *0.01 to 0.05 by 0.01*.

Test

Alternative Hypothesis (H1)

This option specifies whether a one-sided or two-sided hypothesis is analyzed.

One-Sided (H1:D1<D0) refers to a one-sided test in which the alternative hypothesis is of the form H1: D1<D0.

One-Sided (H1:D1>D0) refers to a one-sided test in which the alternative hypothesis is of the form H1: D1>D0.

Two-Sided refers to a two-sided test in which the alternative hypothesis is of the type H1: D1<>D0. Here '<>' means unequal.

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

Data Tab (Ratios)

This section documents options that are used when the parameterization is in terms of the ratio, P1 / P2. P1.0 is the value of P1 assumed by the null hypothesis and P1.1 is the value of P1 at which the power is calculated. Once P2, R0, and R1 are given, the values of P1.0 and P1.1 can be calculated.

Effect Size – Ratios

R0 (Ratio|H0 = P1.0 / P2)

This option specifies the ratio between the group 1 proportion under the null hypothesis, P1.0, and P2. This ratio is used with P2 to calculate the value of P1.0 using the formula: P1.0 = R0 x P2. The power calculations assume that P1.0 is the value of the P1 under the null hypothesis. In this non-null case, the value of P1.0 is not equal to P2 as it is in the null case.

You may enter a range of values such as *0.5 0.6 0.7 0.8* or *1.25 to 2.0 by 0.25*.

Ratios must be greater than zero.

R1 (Ratio|H1 = P1.1 / P2)

This option specifies the ratio of P1.1 and P2, where P1.1 is the proportion in group 1 under the alternative hypothesis. This ratio is used with P2 to calculate the value of P1.1 using the formula: P1.1 = R1 x P2.The power calculations assume that P1.1 is the actual value of the proportion in group 1 (experimental or treatment group).

You may enter a range of values such as *0.5 0.6 0.7 0.8* or *1.25 to 2.0 by 0.25*.

Ratios must be greater than zero. They cannot take on the value of one.

Test

Alternative Hypothesis (H1)

This option specifies whether a one-sided or two-sided hypothesis is analyzed.

One-Sided (H1:R1<R0) refers to a one-sided test in which the alternative hypothesis is of the form H1: R1<R0.

One-Sided (H1:R1>R0) refers to a one-sided test in which the alternative hypothesis is of the form H1: R1>R0.

Two-Sided refers to a two-sided test in which the alternative hypothesis is of the type H1: R1<>R0. Here '<>' means unequal.

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

Iterations/Zeroes Tab

The Iterations/Zeroes tab contains various limits and options.

Maximum Iterations

Maximum Iterations Before Search Termination

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 not reported. A value of at least 500 is recommended.

Zero Counts

Zero Count Adjustment Method

Zero cell counts often cause calculation problems. To compensate for this, a small value (called the Zero Count Adjustment Value) can be added either to all cells or to all cells with zero counts. This option specifies whether you want to use the adjustment and which type of adjustment you want to use. We recommend that you use the option *Add to zero cells only*.

Zero cell values often do not occur in practice. However, since power calculations are based on total enumeration, they will occur in power and sample size estimation.

Adding a small value is controversial, but can be necessary for computational considerations. Statisticians have recommended adding various fractions to zero counts. We have found that adding 0.0001 seems to work well.

Zero Count Adjustment Value

Zero cell counts cause many calculation problems when computing power or sample size. To compensate for this, a small value may be added either to all cells or to all zero cells. This is the amount that is added. We have found that 0.0001 works well.

Be warned that the values of the ratio and the odds ratio will be affected by the amount specified here!

Example 1 – Finding Power

Two competing physical therapy treatments have been available for several years but have not yet been compared as to their effectiveness. The comparison of the two treatments is complicated by the sampling method that will be used. Instead of randomly assigning individuals to treatments, the researchers will randomly select two groups of physical therapists. The first group will be selected from those who use treatment 1. The second group will be selected from those who use treatment 2. The researchers will then follow up on the success or failure of the treatment for multiple patients of each physical therapist. The success rate of treatment 2 is known to be about 0.44. The researchers want to examine effect of the number of physical therapists used in each group and the number of patients for each physical therapist on the power of the test. They wish to determine the power if the treatments are at least 0.07 apart in proportion. They plan to use the Farrington and Manning likelihood score test statistic to analyze the data. Based on similar studies, the intracluster correlation is estimated to be 0.02.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Two Proportions in a Cluster-Randomized Design [Differences]** procedure window by expanding **Proportions**, then **Two Proportions – Cluster Randomized**, then clicking on **Test (Inequality)**, and then clicking on **Tests for Two Proportions in a Cluster-Randomized Design [Differences]**. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power	*Ignored since this is the Find setting*
Alpha	**0.05**
K1 (Clusters Per Group 1)	**10 15 20 25**
M1 (Items Per Cluster in Group 1)	**10 to 50 by 10**
K2 (Clusters Per Group 2)	**K1**
M2 (Items Per Cluster in Group 2)	**M1**
D0 (Difference|H0 = P1.0 – P2)	**0.0**
D1 (Difference|H1 = P1.1 – P2)	**0.07**
P2 (Group 2 Proportion)	**0.44**
ICC (Intracluster Correlation)	**0.02**
H1 (Alternative Hypothesis)	**Two-Sided**
Test Type	**Likelihood Score (Farr. & Mann.)**
Axes/Legend/Grid Tab	
Horizontal Axis Parameter	**K1**

Annotated Output

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

Numeric Results

Numeric Results of Tests Based on the Difference: P1 - P2
H0: P1-P2=D0. H1: P1-P2=D1<>D0. Test Statistic: Score test (Farrington & Manning)

Power	Group 1 Clusters/ Items K1/M1	Group 2 Clusters/ Items K2/M2	Intra-Cluster Corr. ICC	Prop Grp 2 Control P2	Prop|H0 Grp 1 or Trtmnt P1.0	Prop|H1 Grp 1 or Trtmnt P1.1	Diff if H0 D0	Diff if H1 D1	Alpha	Beta
0.1491	10/10	10/10	0.0200	0.4400	0.4400	0.5100	0.0000	0.0700	0.0500	0.8509
0.2219	10/20	10/20	0.0200	0.4400	0.4400	0.5100	0.0000	0.0700	0.0500	0.7781
0.2763	10/30	10/30	0.0200	0.4400	0.4400	0.5100	0.0000	0.0700	0.0500	0.7237
0.3179	10/40	10/40	0.0200	0.4400	0.4400	0.5100	0.0000	0.0700	0.0500	0.6821
0.3506	10/50	10/50	0.0200	0.4400	0.4400	0.5100	0.0000	0.0700	0.0500	0.6494
0.2013	15/10	15/10	0.0200	0.4400	0.4400	0.5100	0.0000	0.0700	0.0500	0.7987
0.3096	15/20	15/20	0.0200	0.4400	0.4400	0.5100	0.0000	0.0700	0.0500	0.6904
0.3871	15/30	15/30	0.0200	0.4400	0.4400	0.5100	0.0000	0.0700	0.0500	0.6129
0.4452	15/40	15/40	0.0200	0.4400	0.4400	0.5100	0.0000	0.0700	0.0500	0.5548
0.4879	15/50	15/50	0.0200	0.4400	0.4400	0.5100	0.0000	0.0700	0.0500	0.5121
0.2522	20/10	20/10	0.0200	0.4400	0.4400	0.5100	0.0000	0.0700	0.0500	0.7478
0.3927	20/20	20/20	0.0200	0.4400	0.4400	0.5100	0.0000	0.0700	0.0500	0.6073
0.4889	20/30	20/30	0.0200	0.4400	0.4400	0.5100	0.0000	0.0700	0.0500	0.5111
0.5569	20/40	20/40	0.0200	0.4400	0.4400	0.5100	0.0000	0.0700	0.0500	0.4431
0.6066	20/50	20/50	0.0200	0.4400	0.4400	0.5100	0.0000	0.0700	0.0500	0.3934
0.3025	25/10	25/10	0.0200	0.4400	0.4400	0.5100	0.0000	0.0700	0.0500	0.6975
0.4715	25/20	25/20	0.0200	0.4400	0.4400	0.5100	0.0000	0.0700	0.0500	0.5285
0.5796	25/30	25/30	0.0200	0.4400	0.4400	0.5100	0.0000	0.0700	0.0500	0.4204
0.6520	25/40	25/40	0.0200	0.4400	0.4400	0.5100	0.0000	0.0700	0.0500	0.3480
0.7030	25/50	25/50	0.0200	0.4400	0.4400	0.5100	0.0000	0.0700	0.0500	0.2970

Report Definitions
H0 is an abbreviation for the NULL hypothesis. This is the hypothesis being evaluated by the statistical test.
H1 is an abbreviation for the ALTERNATIVE hypothesis. This hypothesis gives the 'true' parameter values.
Power is the probability of rejecting a false null hypothesis. It should be close to one.
K1 & K2 are the number of clusters in groups 1 & 2, respectively.
M1 & M2 are the average number of items (subjects) per cluster in groups 1 & 2, respectively.
ICC is the intracluster correlation.
P2 is the proportion for group two. This is the standard, reference, baseline, or control group.
P1.0 is the proportion for group one (treatment group) assuming the null hypothesis (H0).
P1.1 is the proportion for group one (treatment group) assuming the alternative hypothesis (H1).
D0: Diff|H0 is the difference P1 - P2 assuming the null hypothesis (H0).
D1: Diff|H1 is the difference P1 - P2 assuming the alternative hypothesis (H1).
Alpha is the probability of rejecting a true null hypothesis.
Beta is the probability of accepting a false H0. Beta = 1 - Power.

Summary Statements
Sample sizes of 100 in group one and 100 in group two, which were obtained by sampling 10 clusters with 10 subjects each in group one and 10 clusters with 10 subjects each in group two, achieve 15% power to detect a difference between the group proportions of 0.0700. The group two proportion is 0.4400. The group one proportion is assumed to be 0.4400 under the null hypothesis and 0.5100 under the alternative hypothesis. The test statistic used is the two-sided Score test (Farrington & Manning). The significance level of the test was 0.0500.

Group 1 Clusters/Items: K1/M1

This line gives the value of K1, the number of clusters in group 1, followed by M1, the number of items per cluster in this group. The total number of items sampled in group 1 is N1 = K1 x M1.

Group 2 Clusters/Items: K2/M2

This line gives the value of K2, the number of clusters in group 2, followed by M2, the number of items per cluster in this group. The total number of items sampled in group 2 is N2 = K2 x M2.

Intracluster Corr.: ICC

This is the value of the intracluster correlation coefficient, ICC.

Prop Grp 2 or Control: P2

This is the value of P2, the proportion responding positively in the control group.

Prop|H0 Grp 1 or Trtmnt: P1.0

This is the value of P1.0, the proportion responding positively in the treatment group as specified by the null hypothesis. The difference between this value and P2 is the value used in the null hypothesis.

Prop|H1 Grp 1 or Trtmnt: P1.1

This is the value of P1.1, the proportion responding positively in the treatment group as specified by the alternative hypothesis. The difference between this value and P2 is the value used in the alternative hypothesis.

Diff if H0: D0

This is the value of D0, the difference between proportions under the null hypothesis.

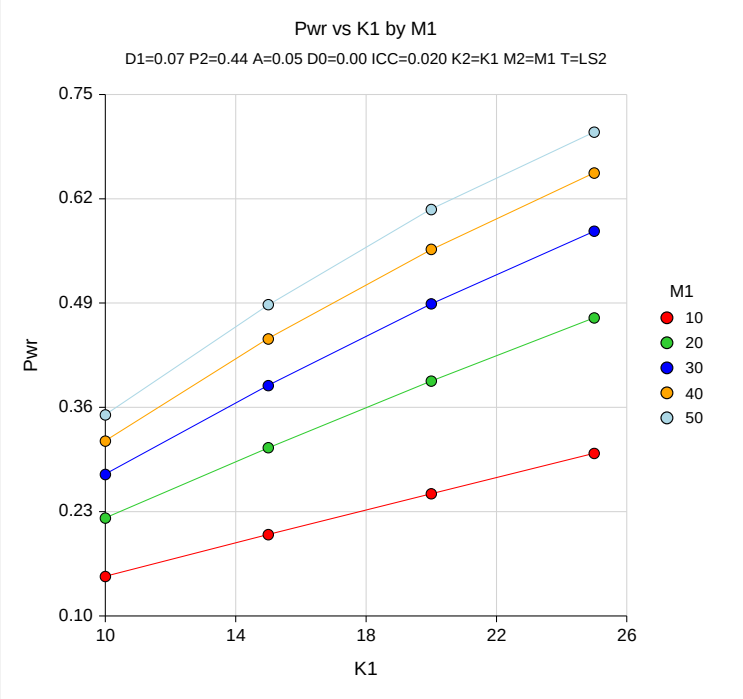
Alpha

This is the value of alpha (significance level) that was targeted by the design.

Beta

This is the value of beta, which is the probability of not rejecting a false null hypothesis.

Plots Section



The values from the table are displayed on the above chart. This chart gives us a quick look at the power that is achieved for various combinations of cluster size and numbers of clusters.

Example 2 – Finding Sample Size

Continuing with Example1, the maximum number of therapists the researchers hope to use is 25 for each treatment. They decide to determine how many patients each therapist would have to treat to achieve 90% power if the maximum number of therapists is used.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Two Proportions in a Cluster-Randomized Design [Differences]** procedure window by expanding **Proportions**, then **Two Proportions – Cluster Randomized**, then clicking on **Test (Inequality)**, and then clicking on **Tests for Two Proportions in a Cluster-Randomized Design [Differences]**. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**M1**
Power	**0.90**
Alpha	**0.05**
K1 (Clusters Per Group 1)	**25**
M1 (Items Per Cluster in Group 1)	*Ignored since this is the Find setting*
K2 (Clusters Per Group 2)	**K1**
M2 (Items Per Cluster in Group 2)	**M1**
D0 (Difference|H0 = P1.0 – P2)	**0.0**
D1 (Difference|H1 = P1.1 – P2)	**0.07**
P2 (Group 2 Proportion)	**0.44**
ICC (Intracluster Correlation)	**0.02**
H1 (Alternative Hypothesis)	**Two-Sided**
Test Type	**Likelihood Score (Farr. & Mann.)**

Output

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

Numeric Results

Numeric Results of Tests Based on the Difference: P1 - P2
H0: P1-P2=D0. H1: P1-P2=D1<>D0. Test Statistic: Score test (Farrington & Manning)

Power	Group 1 Clusters/ Items K1/M1	Group 2 Clusters/ Items K2/M2	Intra-Cluster Corr. ICC	Prop Grp 2 Control P2	Prop|H0 Grp 1 or Trtmnt P1.0	Prop|H1 Grp 1 or Trtmnt P1.1	Diff if H0 D0	Diff if H1 D1	Alpha	Beta
0.9002	25/286	25/286	0.0200	0.4400	0.4400	0.5100	0.0000	0.0700	0.0500	0.0998

To achieve 90% power, each therapist would need to be evaluated on 286 patients.

Example 3 – Finding Power (Non-Inferiority)

A study is being designed to study the effectiveness of a new treatment. Historically, the standard treatment has enjoyed a 60% cure rate. The new treatment reduces the seriousness of certain side effects that occur with the standard treatment. Thus, the new treatment will be adopted even if it is slightly less effective than the standard treatment. The researchers will recommend adoption of the new treatment if it has a cure rate of at least 55%.

The researchers will recruit patients from various hospitals. All patients at a particular hospital will receive the same treatment. They anticipate an average of 100 patients per hospital.

The researchers plan to use the Farrington and Manning likelihood score test statistic to analyze the data. They want to study the power of the one-sided Farrington and Manning test at group cluster sizes ranging from 2 to 10 for detecting a difference of -0.05 when the actual cure rate of the new treatment ranges from 60% to 66%. The significance level will be 0.05. Based on similar studies, they estimate the intracluster correlation to be 0.002.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Two Proportions in a Cluster-Randomized Design [Differences]** procedure window by expanding **Proportions**, then **Two Proportions – Cluster Randomized**, then clicking on **Test (Inequality)**, and then clicking on **Tests for Two Proportions in a Cluster-Randomized Design [Differences]**. You may then make the appropriate entries as listed below, or open **Example 3** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power	*Ignored since this is the Find setting*
Alpha	**0.05**
K1 (Clusters Per Group 1)	**2 4 6 8 10**
M1 (Items Per Cluster in Group 1)	**100**
K2 (Clusters Per Group 2)	**K1**
M2 (Items Per Cluster in Group 2)	**M1**
D0 (Difference|H0 = P1.0 – P2)	**-0.05**
D1 (Difference|H1 = P1.1 – P2)	**0 0.02 0.04 0.06**
P2 (Group 2 Proportion)	**0.6**
ICC (Intracluster Correlation)	**0.002**
H1 (Alternative Hypothesis)	**One-Sided (H1:D1>D0)**
Test Type	**Likelihood Score (Farr. & Mann.)**

Output

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

Numeric Results

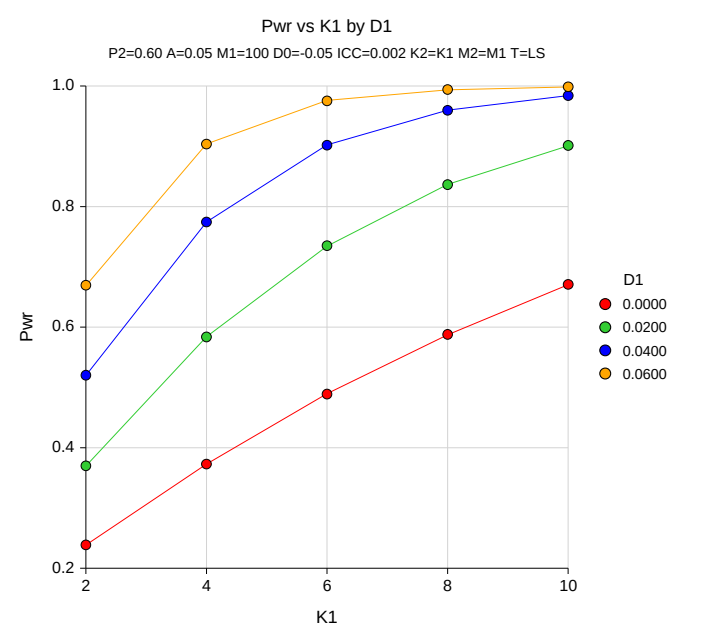
Numeric Results of Tests Based on the Difference: P1 – P2
H0: P1-P2<=D0. H1: P1-P2=D1>D0. Test Statistic: Score test (Farrington & Manning)

Power	Group 1 Clusters/ Items K1/M1	Group 2 Clusters/ Items K2/M2	Intra-Cluster Corr. ICC	Prop Grp 2 Control P2	Prop|H0 Grp 1 or Trtmnt P1.0	Prop|H1 Grp 1 or Trtmnt P1.1	Diff if H0 D0	Diff if H1 D1	Alpha	Beta
0.2387	2/100	2/100	0.0020	0.6000	0.5500	0.6000	-0.0500	0.0000	0.0500	0.7613
0.3729	4/100	4/100	0.0020	0.6000	0.5500	0.6000	-0.0500	0.0000	0.0500	0.6271
0.4889	6/100	6/100	0.0020	0.6000	0.5500	0.6000	-0.0500	0.0000	0.0500	0.5111
0.5879	8/100	8/100	0.0020	0.6000	0.5500	0.6000	-0.0500	0.0000	0.0500	0.4121
0.6709	10/100	10/100	0.0020	0.6000	0.5500	0.6000	-0.0500	0.0000	0.0500	0.3291
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.

Summary Statements
Sample sizes of 200 in group one and 200 in group two, which were obtained by sampling 2 clusters with 100 subjects each in group one and 2 clusters with 100 subjects each in group two, achieve 24% power to detect a difference between the group proportions of 0.0000. The group two proportion is 0.6000. The group one proportion is assumed to be 0.5500 under the null hypothesis and 0.6000 under the alternative hypothesis. The test statistic used is the one-sided Score test (Farrington & Manning). The significance level of the test was 0.0500.

This report shows the values of each of the parameters, one scenario per row. Most of the report columns have obvious interpretations. Those that may not be obvious are presented here.

Plots Section



The values from the table are displayed on the above chart. This chart gives us a quick look at the sample size that will be required for various values of D1.

Example 4 – Finding the Sample Size (Non-Inferiority)

Continuing with the scenario given in Example 3, the researchers want to determine the number of clusters necessary for each value of D1 when the target power is set to 0.80.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Two Proportions in a Cluster-Randomized Design [Differences]** procedure window by expanding **Proportions**, then **Two Proportions – Cluster Randomized**, then clicking on **Test (Inequality)**, and then clicking on **Tests for Two Proportions in a Cluster-Randomized Design [Differences]**. You may then make the appropriate entries as listed below, or open **Example 4** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**K1**
Power	**0.80**
Alpha	**0.05**
K1 (Clusters Per Group 1)	*Ignored since this is the Find setting*
M1 (Items Per Cluster in Group 1)	**100**
K2 (Clusters Per Group 2)	**K1**
M2 (Items Per Cluster in Group 2)	**M1**
D0 (Difference|H0 = P1.0 – P2)	**-0.05**
D1 (Difference|H1 = P1.1 – P2)	**0 0.02 0.04 0.06**
P2 (Group 2 Proportion)	**0.6**
ICC (Intracluster Correlation)	**0.002**
H1 (Alternative Hypothesis)	**One-Sided (H1:D1>D0)**
Test Type	**Likelihood Score (Farr. & Mann.)**

Output

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

Numeric Results

Numeric Results of Tests Based on the Difference: P1 – P2
H0: P1-P2<=D0. H1: P1-P2=D1>D0. Test Statistic: Score test (Farrington & Manning)

Power	Group 1 Clusters/ Items K1/M1	Group 2 Clusters/ Items K2/M2	Intra-Cluster Corr. ICC	Prop Grp 2 Control P2	Prop|H0 Grp 1 or Trtmnt P1.0	Prop|H1 Grp 1 or Trtmnt P1.1	Diff if H0 D0	Diff if H1 D1	Alpha	Beta
0.8190	15/100	15/100	0.0020	0.6000	0.5500	0.6000	-0.0500	0.0000	0.0500	0.1810
0.8364	8/100	8/100	0.0020	0.6000	0.5500	0.6200	-0.0500	0.0200	0.0500	0.1636
0.8503	5/100	5/100	0.0020	0.6000	0.5500	0.6400	-0.0500	0.0400	0.0500	0.1497
0.8186	3/100	3/100	0.0020	0.6000	0.5500	0.6600	-0.0500	0.0600	0.0500	0.1814

The required sample size depends a great deal on the value of D1. The researchers should spend time determining the most appropriate value for D1.

Example 5 – Investigating the Impact of the Intracluster Correlation

Continuing with the scenario given in Example 4, the researchers, having decided that the most appropriate value of D1 is 0.02, now want to investigate the effect of the intracluster correlation on the sample size. From values found in other studies, they believe the ICC will be somewhere between 0.001 and 0.009.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Two Proportions in a Cluster-Randomized Design [Differences]** procedure window by expanding **Proportions**, then **Two Proportions – Cluster Randomized**, then clicking on **Test (Inequality)**, and then clicking on **Tests for Two Proportions in a Cluster-Randomized Design [Differences]**. You may then make the appropriate entries as listed below, or open **Example 5** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**K1**
Power	**0.80**
Alpha	**0.05**
K1 (Clusters Per Group 1)	*Ignored since this is the Find setting*
M1 (Items Per Cluster in Group 1)	**100**
K2 (Clusters Per Group 2)	**K1**
M2 (Items Per Cluster in Group 2)	**M1**
D0 (Difference|H0 = P1.0 – P2)	**-0.05**
D1 (Difference|H1 = P1.1 – P2)	**0.02**
P2 (Group 2 Proportion)	**0.6**
ICC (Intracluster Correlation)	**0.001 to 0.009 by 0.002**
H1 (Alternative Hypothesis)	**One-Sided (H1:D1>D0)**
Test Type	**Likelihood Score (Farr. & Mann.)**

Output

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

Numeric Results

Numeric Results of Tests Based on the Difference: P1 – P2
H0: P1-P2<=D0. H1: P1-P2=D1>D0. Test Statistic: Score test (Farrington & Manning)

Power	Group 1 Clusters/ Items K1/M1	Group 2 Clusters/ Items K2/M2	Intra-Cluster Corr. ICC	Prop Grp 2 Control P2	Prop|H0 Grp 1 or Trtmnt P1.0	Prop|H1 Grp 1 or Trtmnt P1.1	Diff if H0 D0	Diff if H1 D1	Alpha	Beta
0.8207	7/100	7/100	0.0010	0.6000	0.5500	0.6200	-0.0500	0.0200	0.0500	0.1793
0.8099	8/100	8/100	0.0030	0.6000	0.5500	0.6200	-0.0500	0.0200	0.0500	0.1901
0.8020	9/100	9/100	0.0050	0.6000	0.5500	0.6200	-0.0500	0.0200	0.0500	0.1980
0.8275	11/100	11/100	0.0070	0.6000	0.5500	0.6200	-0.0500	0.0200	0.0500	0.1725
0.8197	12/100	12/100	0.0090	0.6000	0.5500	0.6200	-0.0500	0.0200	0.0500	0.1803

This chart shows that the necessary sample size almost doubles when the ICC is changed from 0.001 to 0.009. The researchers decide to obtain a narrower range for the value of ICC.

Example 6 – Validation using Donner and Klar

Donner and Klar (2000), page 63, present a sample size study in which P2 = 0.06, D1 = -0.02, D0=0, ICC = 0.01, M1=M2=100, two-sided alpha = 0.05, and beta = 0.20. Using the pooled z test statistic, they found the number of subjects to be 3698 in each group, which they round off to 38 clusters per group.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Two Proportions in a Cluster-Randomized Design [Differences]** procedure window by expanding **Proportions**, then **Two Proportions – Cluster Randomized**, then clicking on **Test (Inequality)**, and then clicking on **Tests for Two Proportions in a Cluster-Randomized Design [Differences]**. You may then make the appropriate entries as listed below, or open **Example 6** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**K1**
Power	**0.80**
Alpha	**0.05**
K1 (Clusters Per Group 1)	*Ignored since this is the Find setting*
M1 (Items Per Cluster in Group 1)	**100**
K2 (Clusters Per Group 2)	**K1**
M2 (Items Per Cluster in Group 2)	**M1**
D0 (Difference|H0 = P1.0 – P2)	**0.0**
D1 (Difference|H1 = P1.1 – P2)	**-0.02**
P2 (Group 2 Proportion)	**0.06**
ICC (Intracluster Correlation)	**0.01**
H1 (Alternative Hypothesis)	**Two-Sided**
Test Type	**Z test (Pooled)**

Output

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

Numeric Results

Numeric Results of Tests Based on the Difference: P1 – P2
H0: P1-P2=D0. H1: P1-P2=D1<>D0. Test Statistic: Z test (pooled)

Power	Group 1 Clusters/ Items K1/M1	Group 2 Clusters/ Items K2/M2	Intra-Cluster Corr. ICC	Prop Grp 2 Control P2	Prop|H0 Grp 1 or Trtmnt P1.0	Prop|H1 Grp 1 or Trtmnt P1.1	Diff if H0 D0	Diff if H1 D1	Alpha	Beta
0.8097	38/100	38/100	0.0100	0.0600	0.0600	0.0400	0.0000	-0.0200	0.0500	0.1903

PASS has also found the required sample size to be 38 clusters.

Chapter 233

Tests for Two Proportions with Non-Zero Null Hypothesis in a Cluster-Randomized Design

Introduction

This module provides power analysis and sample size calculation for inequality tests with a non-zero null hypothesis in two-sample, cluster-randomized designs in which the outcome is binary.

Three Procedures Documented Here

There are four procedures in the menus that use the program module described in this chapter. These procedures are identical except for the type of parameterization. The parameterization can be in terms of proportions, differences in proportions, or ratios of proportions. Each of these options is listed separately on the menus.

Technical Details

The methods contained in this module are identical to those discussed in the chapter “Inequality Tests for Two Proportions in a Cluster-Randomized Design.” The input and output has simply been reformatted in a manner that is convenient for non-zero null hypothesis testing. A complete review of non-zero null testing is given in the chapter “Inequality Tests for Two Proportions with Non-Zero Null Hypothesis.” We refer you to these two chapters for complete technical details on the methods used in this module.

Procedure Options

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

Data Tab (Common Options)

The Data tab contains the parameters associated with this test such as the proportions, sample sizes, alpha, and beta. This chapter covers three procedures, each of which has different options. This section documents options that are common to all three procedures. Later, unique options for each procedure will be documented.

Solve For

Find (Solve For)

This option specifies the parameter to be solved for using the other parameters. The parameters that may be selected are *P1.1, Alpha, Power and Beta, K1, M1,* or *ICC*. Under most situations, you will select either *Power and Beta* or *K1*.

Select *K1* when you want to calculate the sample size needed to achieve a given power and alpha level.

Select *Power and Beta* when you want to calculate the power of an experiment that has already been run.

Error Rates

Power or Beta

This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected.

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

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

Alpha (Significance Level)

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected.

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

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

Sample Size – Treatment (Group 1)

K1 (Clusters Group 1)

Enter a value (or range of values) for the number of clusters in group one. You may enter a range of values such as *10 to 20 by 2*. The sample size for this group is equal to the number of clusters times the number of subjects per cluster.

M1 (Items Group 1)

This is the average number of items (subjects) per cluster in group one. This value must be a positive number that is at least 1. You can use a list of values such as *100 150 200*.

Sample Size – Control (Group 2)

K2 (Clusters Group 2)

This is the number of clusters in group two. The sample size for this group is equal to the number of clusters times the number of subjects per cluster. This value must be a positive number.

If you simply want a multiple of the value for group one, you would enter the multiple followed by *K1*, with no blanks. If you want to use *K1* directly, you do not have to pre-multiply by 1. For example, all of the following are valid entries:*10 K1 2K1 0.5K1*.

You can use a list of values such as *10 20 30* or *K1 2K1 3K1*.

M2 (Items per Cluster in Group 2)

This is the number of items (subjects) per cluster in group two. This value must be a positive number.

If you simply want a multiple of the value for group one, you would enter the multiple followed by *M1*, with no blanks. If you want to use *M1* directly, you do not have to pre-multiply by *1*. For example, all of the following are valid entries: *10 M1 2M1 0.5M1*.

You can use a list of values such as *10 20 30* or *M1 2M1 3M1*.

Effect Size – Control (Group 2)

P2 (Group 2 Proportion)

Specify the value of p_2, the control, baseline, or standard group's proportion. The null hypothesis is that the two proportions differ by a specified amount (See *Group 1 Proportion* below).

Since p_2 is a proportion, these values must be between 0 and 1.

You may enter a range of values such as *0.1 0.2 0.3* or *0.1 to 0.9 by 0.1*.

Effect Size – Intracluster Correlation

ICC (Intracluster Correlation)

Enter a value (or range of values) for the intracluster correlation. This correlation may be thought of as the simple correlation between any two observations in the same cluster. It may also be thought of as the proportion of total variance in the observations that can be attributed to difference between clusters.

Although the actual range for this value is between 0 to 1, typical values range from 0.002 to 0.05.

Test

Higher Proportions Are

This option specifies whether proportions represent successes (better) or failures (worse).

- **Better (Successes)**
  If Higher Proportions are "Better", the alternative hypothesis is H1: P1>P2.
- **Worse (Failures)**
  If Higher Proportions are "Worse", the alternative hypothesis is H1: P1<P2.

Test Type

Specify which test statistic is used in searching and reporting. We recommend the likelihood score test.

Data Tab (Proportions)

This section documents options that are used when the parameterization is in terms of the values of the two proportions, P1 and P2. P1.0 is the value of the P1 assumed by the null hypothesis and P1.1 is the value of P1 at which the power is calculated.

Effect Size – Treatment (Group 1)

P1.0 (Superiority Group 1 Proportion)

This option allows you to specify the value P1.0 directly.

When higher proportions are "Better", P1.0 is the treatment group proportion such that proportions greater than P1.0 are considered superior to the reference group proportion. When higher proportions are "Better", P1.0 should be greater than P2.

When higher proportions are "Worse", P1.0 is the treatment group proportion such that proportions less than P1.0 are considered superior to the reference group proportion. When higher proportions are "Worse", P1.0 should be less than P2.

Proportions must be between 0 and 1. They cannot take on the values 0 or 1. This value should not be set to exactly the value of P2.

You may enter a range of values such as *0.03 0.05 0.10* or *0.01 to 0.05 by 0.01*.

P1.1 (Actual Group 1 Proportion)

This is the value of the treatment proportion (P1) at which the power calculations are made.

When higher proportions are "Better", P1.1 should be greater than P1.0.

When higher proportions are "Worse", P1.1 should be less than P1.0.

Proportions must be between 0 and 1. They cannot take on the values 0 or 1.

You may enter a range of values such as *0.03 0.05 0.10* or *0.01 to 0.05 by 0.01*.

Data Tab (Differences)

This section documents options that are used when the parameterization is in terms of the difference, P1 – P2. P1.0 is the value of P1 assumed by the null hypothesis and P1.1 is the value of P1 at which the power is calculated. Once P2, D0, and D1 are given, the values of P1.1 and P1.0 can be calculated.

Effect Size – Differences

D0 (Superiority Difference)

Specify the difference between P1.0 and P2 that will be considered superior.

When higher proportions are "Better", the superiority difference is that amount that P1 need be greater than P2 to be declared superior to the reference group. When higher proportions are "Better", D0 should be positive.

When higher proportions are "Worse", the superiority difference is that amount that P1 need be less than P2 to be declared superior to the reference group. When higher proportions are "Worse", D0 should be negative.

The power calculations assume that P1.0 is the value of the P1 under the null hypothesis. This value is used with P2 to calculate the value of P1.0 using the formula: P1.0 = D0 + P2.

You may enter a range of values such as *-.03 -.05 -.10* or *-.05 to -.01 by .01*. Differences must be between -1 and 1. D0 cannot take on the values -1, 0, or 1.

D1 (Actual Difference)

Specify the actual difference between P1.1 (the actual value of P1) and P2. This is the value at which the power is calculated.

When higher proportions are "Better", D1 should be positive and greater than D0.

When higher proportions are "Worse", D1 should be negative and more extreme than D0.

The power calculations assume that P1.1 is the actual value of the proportion in group 1 (experimental or treatment group). This difference is used with P2 to calculate the value of P1 using the formula: P1.1 = D1 + P2.

You may enter a range of values such as *-.05 0 .5* or *-.05 to .05 by .02*. Actual differences must be between -1 and 1. They cannot take on the values -1 or 1.

Data Tab (Ratios)

This section documents options that are used when the parameterization is in terms of the ratio, P1 / P2. P1.0 is the value of P1 assumed by the null hypothesis and P1.1 is the value of P1 at which the power is calculated. Once P2, R0, and R1 are given, the values of P1.0 and P1.1 can be calculated.

Effect Size – Ratios

R0 (Superiority Ratio)

Specify the superiority ratio of P1.0 to P2.

When higher proportions are "Better", the superiority ratio is smallest ratio of P1 to P2 for which P1 will still be considered superior. When higher proportions are "Better", R0 should be greater than one.

When higher proportions are "Worse", the superiority ratio is largest ratio of P1 to P2 for which P1 will still be considered superior. When higher proportions are "Worse", R0 should be less than one.

The power calculations assume that P1.0 is the value of P1 under the null hypothesis. This value is used with P2 to calculate the value of P1.0 using the formula: P1.0 = R0 x P2.

Ratios must be positive. R0 cannot take on the value of 1.

You may enter a range of values such as *0.95 .97 .99* or *.91 to .99 by .02*.

R1 (Actual Ratio)

Specify the actual ratio between P1.1 (the actual value of P1) and P2. This is the value at which the power is calculated.

When higher proportions are "Better", R1 should be greater than R0.

When higher proportions are "Worse", R1 should be less than R0.

The power calculations assume that P1.1 is the actual value of the proportion in group 1 (experimental or treatment group). This ratio is used with P2 to calculate the value of P1 using the formula: P1.1 = R1 x P2.

Ratios must be positive. You may enter a range of values such as *0.95 1 1.05* or *0.9 to 1.9 by 0.02*.

Iterations/Zeroes Tab

The Iterations/Zeroes tab contains various limits and options.

Maximum Iterations

Maximum Iterations Before Search Termination

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 not reported. A value of at least 500 is recommended.

Zero Counts

Zero Count Adjustment Method

Zero cell counts often cause calculation problems. To compensate for this, a small value (called the Zero Count Adjustment Value) can be added either to all cells or to all cells with zero counts. This option specifies whether you want to use the adjustment and which type of adjustment you want to use. We recommend that you use the option *Add to zero cells only*.

Zero cell values often do not occur in practice. However, since power calculations are based on total enumeration, they will occur in power and sample size estimation.

Adding a small value is controversial, but can be necessary for computational considerations. Statisticians have recommended adding various fractions to zero counts. We have found that adding 0.0001 seems to work well.

Zero Count Adjustment Value

Zero cell counts cause many calculation problems when computing power or sample size. To compensate for this, a small value may be added either to all cells or to all zero cells. This value indicates the amount that is added. We have found that 0.0001 works well.

Be warned that the value of the ratio and the odds ratio will be affected by the amount specified here!

Example 1 – Finding Power

A study is being designed to study the effectiveness of a new treatment. Historically, the standard treatment has enjoyed a 60% cure rate. The new treatment is more expensive and has more side effects than the standard treatment. Thus, the new treatment will be adopted if it is at least 8% better than the standard treatment. %.

The researchers will recruit patients from various hospitals. All patients at a particular hospital will receive the same treatment. They anticipate an average of 100 patients per hospital. Based on similar studies, they estimate the intracluster correlation to be 0.002.

The researchers plan to use the Farrington and Manning likelihood score test statistic to analyze the data. They want to study the power of the one-sided Farrington and Manning test at group cluster sizes ranging from 2 to 10 for detecting a difference of 0.08 when the actual cure rate of the new treatment ranges from 70% to 76%. The significance level will be 0.05.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Two Proportions in a Cluster-Randomized Design (Non-Zero Null) [Differences]** procedure window by expanding **Proportions**, then **Two Proportions – Cluster Randomized**, then clicking on **Test (Non-Zero Null)**, and then clicking on **Tests for Two Proportions in a Cluster-Randomized Design (Non-Zero Null) [Differences]**. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power	*Ignored since this is the Find setting*
Alpha	**0.05**
K1 (Clusters Per Group 1)	**2 4 6 8 10**
M1 (Items Per Cluster in Group 1)	**100**
K2 (Clusters Per Group 2)	**K1**
M2 (Items Per Cluster in Group 2)	**M1**
D0 (Superiority Difference)	**0.08**
D1 (Actual Difference)	**0.10 to 0.16 by 0.01**
P2 (Group 2 Proportion)	**0.6**
ICC (Intracluster Correlation)	**0.002**
Higher Proportions Are	**Better**
Test Type	**Likelihood Score (Farr. & Mann.)**

Annotated Output

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

Numeric Results

Numeric Results for Non-Zero Null (Superiority) Tests Based on the Difference: P1 - P2
H0: P1-P2<=D0. H1: P1-P2=D1>D0. Test Statistic: Score test (Farrington & Manning)

Power	Group 1 Clusters/ Items K1/M1	Group 2 Clusters/ Items K2/M2	Intra-Cluster Corr. ICC	Grp 2 Prop P2	Super. Grp 1 Prop P1.0	Actual Grp 1 Prop P1.1	Super. Margin Diff D0	Actual Margin Diff D1	Alpha	Beta
0.1034	2/100	2/100	0.0020	0.6000	0.6800	0.7000	0.0800	0.1000	0.0500	0.8966
0.1351	4/100	4/100	0.0020	0.6000	0.6800	0.7000	0.0800	0.1000	0.0500	0.8649
0.1635	6/100	6/100	0.0020	0.6000	0.6800	0.7000	0.0800	0.1000	0.0500	0.8365
0.1902	8/100	8/100	0.0020	0.6000	0.6800	0.7000	0.0800	0.1000	0.0500	0.8098
0.2159	10/100	10/100	0.0020	0.6000	0.6800	0.7000	0.0800	0.1000	0.0500	0.7841
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.

Summary Statements
Sample sizes of 200 in group one and 200 in group two, which were obtained by sampling 2 clusters with 100 subjects each in group one and 2 clusters with 100 subjects each in group two, achieve 10% power to detect a superiority margin difference between the group proportions of 0.0800. The group two proportion is 0.6000. The group one proportion is assumed to be 0.6800 under the null hypothesis and 0.7000 under the alternative hypothesis. The test statistic used is the one-sided Z test (unpooled). The significance level of the test was 0.0500.

This report shows the values of each of the parameters, one scenario per row. Most of the report columns have obvious interpretations. Those that may not be obvious are presented here.

Group 1 Clusters/Items: K1/M1

This line gives the value of K1, the number of clusters in group 1, followed by M1, the number of items per cluster in this group. The total number of items sampled in group 1 is N1 = K1 x M1.

Group 2 Clusters/Items: K2/M2

This line gives the value of K2, the number of clusters in group 2, followed by M2, the number of items per cluster in this group. The total number of items sampled in group 2 is N2 = K2 x M2.

Intracluster Corr.: ICC

This is the value of the intracluster correlation coefficient, ICC.

Prop Grp 2: P2

This is the value of P2, the proportion responding positively in the control group.

Super. Grp 1 Prop P1.0

This is the value of P1.0, the response rate of the treatment group, as specified by the null hypothesis of superiority. Values of P1 greater than this amount are considered different from P2. The difference between this value and P2 is the value of the null hypothesis.

Actual Grp 1 Prop P1.1

This is the value of P1.1, the response rate of the treatment group, at which the power is computed. This is the value of P1 under the alternative hypothesis. The difference between this value and P2 is the value of the alternative hypothesis.

Super. Margin Diff D0

This is the value of D0, the difference between the two group proportions under the null hypothesis. This value is often called the *margin of superiority*.

Actual Margin Diff D1

This is the value of D1, the difference between the two group proportions at which the power is computed. This is the value of the difference under the alternative hypothesis.

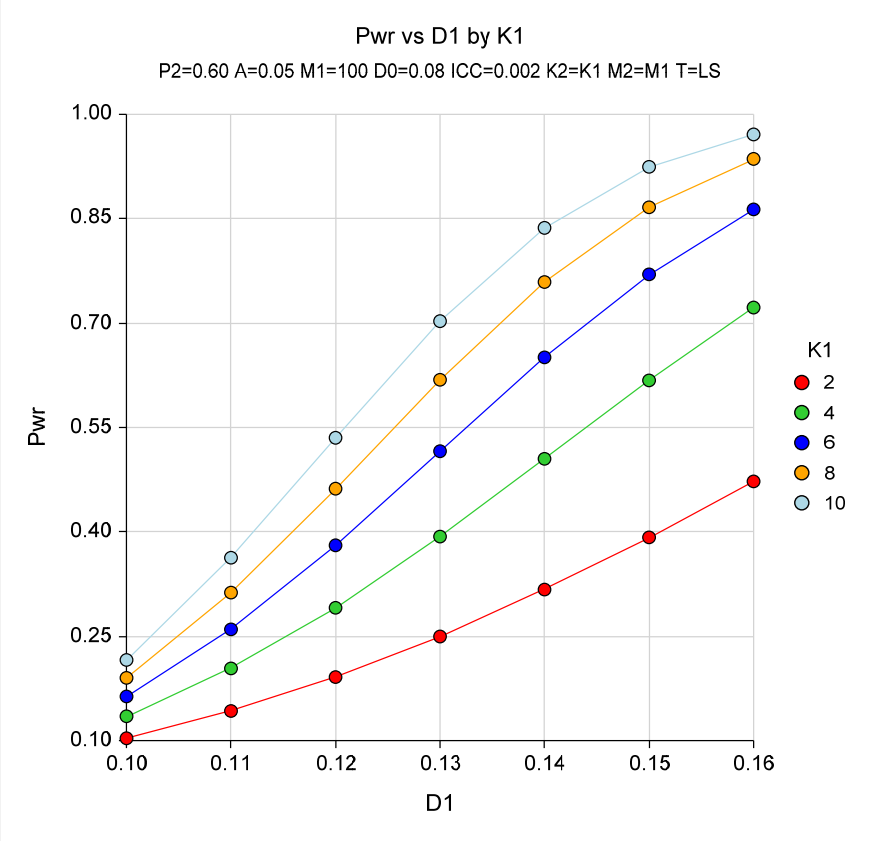
Alpha

This is the value of alpha (significance level) that was targeted by the design.

Beta

This is the value of beta, which is the probability of not rejecting a false null hypothesis.

Plots Section



The values from the table are displayed on the above chart. This chart gives us a quick look at the sample sizes that will be required for various values of D1.

Example 2 – Finding the Sample Size (Number of Clusters)

Continuing with the scenario given in Example 1, the researchers want to determine the number of clusters necessary for each value of D1 when the target power is set to 0.80.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Two Proportions in a Cluster-Randomized Design (Non-Zero Null) [Differences]** procedure window by expanding **Proportions**, then **Two Proportions – Cluster Randomized**, then clicking on **Test (Non-Zero Null)**, and then clicking on **Tests for Two Proportions in a Cluster-Randomized Design (Non-Zero Null) [Differences]**. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**K1**
Power	**0.80**
Alpha	**0.05**
K1 (Clusters Per Group 1)	*Ignored since this is the Find setting*
M1 (Items Per Cluster in Group 1)	**100**
K2 (Clusters Per Group 2)	**K1**
M2 (Items Per Cluster in Group 2)	**M1**
D0 (Superiority Difference)	**0.08**
D1 (Actual Difference)	**0.10 to 0.16 by 0.01**
P2 (Group 2 Proportion)	**0.6**
ICC (Intracluster Correlation)	**0.002**
Higher Proportions Are	**Better**
Test Type	**Likelihood Score (Farr. & Mann.)**

Output

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

Numeric Results

Numeric Results for Non-Zero Null (Superiority) Tests Based on the Difference: P1 - P2
H0: P1-P2<=D0. H1: P1-P2=D1>D0. Test Statistic: Score test (Farrington & Manning)

Power	Group 1 Clusters/ Items K1/M1	Group 2 Clusters/ Items K2/M2	Intra-Cluster Corr. ICC	Grp 2 Prop P2	Super. Grp 1 Prop P1.0	Actual Grp 1 Prop P1.1	Super. Margin Diff D0	Actual Margin Diff D1	Alpha	Beta
0.8021	84/100	84/100	0.0020	0.6000	0.6800	0.7000	0.0800	0.1000	0.0500	0.1979
0.8017	37/100	37/100	0.0020	0.6000	0.6800	0.7100	0.0800	0.1100	0.0500	0.1983
0.8077	21/100	21/100	0.0020	0.6000	0.6800	0.7200	0.0800	0.1200	0.0500	0.1923
0.8246	14/100	14/100	0.0020	0.6000	0.6800	0.7300	0.0800	0.1300	0.0500	0.1754
0.8016	9/100	9/100	0.0020	0.6000	0.6800	0.7400	0.0800	0.1400	0.0500	0.1984
0.8244	7/100	7/100	0.0020	0.6000	0.6800	0.7500	0.0800	0.1500	0.0500	0.1756
0.8043	5/100	5/100	0.0020	0.6000	0.6800	0.7600	0.0800	0.1600	0.0500	0.1957

The required sample size depends a great deal on the value of D1. The researchers should spend time determining the most appropriate value for D1.

Example 3 – Validation

These procedures use the same mechanics as the Non-Inferiority Tests for Two Proportions in a Cluster-Randomized Design procedures. We refer the user to Example 3 of Chapter 235 for the validation.

Chapter 235

Non-Inferiority Tests for Two Proportions in a Cluster-Randomized Design

Introduction

This module provides power analysis and sample size calculation for non-inferiority tests in two-sample, cluster-randomized designs in which the outcome is binary.

Three Procedures Documented Here

There are four procedures in the menus that use the program module described in this chapter. These procedures are identical except for the type of parameterization. The parameterization can be in terms of proportions, differences in proportions, or ratios of proportions. Each of these options is listed separately on the menus.

Technical Details

The methods contained in this module are identical to those discussed in the chapter “Inequality Tests for Two Proportions in a Cluster-Randomized Design.” The input and output has simply been reformatted in a manner that is convenient for non-inferiority testing. A complete review of

non-inferiority testing is given in the chapter “Non-Inferiority Tests for Two Proportions.” We refer you to these two chapters for complete technical details on the methods used in this module.

Procedure Options

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

Data Tab (Common Options)

The Data tab contains the parameters associated with this test such as the proportions, sample sizes, alpha, and beta. This chapter covers three procedures, each of which has different options. This section documents options that are common to all three procedures. Later, unique options for each procedure will be documented.

Solve For

Find (Solve For)

This option specifies the parameter to be solved for using the other parameters. The parameters that may be selected are *P1.1, Alpha, Power and Beta, K1, M1,* or *ICC*. Under most situations, you will select either *Power and Beta* or *K1*.

Select *K1* when you want to calculate the sample size needed to achieve a given power and alpha level.

Select *Power and Beta* when you want to calculate the power of an experiment that has already been run.

Error Rates

Power or Beta

This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected.

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

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

Alpha (Significance Level)

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected.

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

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

Sample Size – Treatment (Group 1)

K1 (Clusters Group 1)

Enter a value (or range of values) for the number of clusters in group one. You may enter a range of values such as *10 to 20 by 2*. The sample size for this group is equal to the number of clusters times the number of subjects per cluster.

M1 (Items Group 1)

This is the average number of items (subjects) per cluster in group one. This value must be a positive number that is at least 1. You can use a list of values such as *100 150 200*.

Sample Size – Control (Group 2)

K2 (Clusters Group 2)

This is the number of clusters in group two. The sample size for this group is equal to the number of clusters times the number of subjects per cluster. This value must be a positive number.

If you simply want a multiple of the value for group one, you would enter the multiple followed by *K1*, with no blanks. If you want to use *K1* directly, you do not have to pre-multiply by 1. For example, all of the following are valid entries:*10 K1 2K1 0.5K1*.

You can use a list of values such as *10 20 30* or *K1 2K1 3K1*.

M2 (Items per Cluster in Group 2)

This is the number of items (subjects) per cluster in group two. This value must be a positive number.

If you simply want a multiple of the value for group one, you would enter the multiple followed by *M1*, with no blanks. If you want to use *M1* directly, you do not have to pre-multiply by *1*. For example, all of the following are valid entries: *10 M1 2M1 0.5M1*.

You can use a list of values such as *10 20 30* or *M1 2M1 3M1*.

Effect Size – Control (Group 2)

P2 (Group 2 Proportion)

Specify the value of p_2, the control, baseline, or standard group's proportion. The null hypothesis is that the two proportions differ by a specified amount (See *Specify Group 1 Proportion using* below).

Since p_2 is a proportion, these values must be between 0 and 1.

You may enter a range of values such as *0.1 0.2 0.3* or *0.1 to 0.9 by 0.1*.

Effect Size – Intracluster Correlation

ICC (Intracluster Correlation)

Enter a value (or range of values) for the intracluster correlation. This correlation may be thought of as the simple correlation between any two observations in the same cluster. It may also be

thought of as the proportion of total variance in the observations that can be attributed to difference between clusters.

Although the actual range for this value is between 0 to 1, typical values range from 0.002 to 0.05.

Test

Higher Proportions Are

This option specifies whether proportions represent successes (better) or failures (worse).

- **Better (Successes)**
  If Higher Proportions are "Better", the alternative hypothesis is H1: P1 > P2.
- **Worse (Failures)**
  If Higher Proportions are "Worse", the alternative hypothesis is H1: P1 < P2.

Test Type

Specify which test statistic is used in searching and reporting. We recommend the likelihood score test.

Data Tab (Proportions)

This section documents options that are used when the parameterization is in terms of the values of the two proportions, P1 and P2. P1.0 is the value of the P1 assumed by the null hypothesis and P1.1 is the value of P1 at which the power is calculated.

Effect Size – Treatment (Group 1)

P1.0 (Non-Inferiority Group 1 Proportion)

Specify the value P1.0 directly.

When higher proportions are "Better", P1.0 is the treatment group proportion such that proportions greater than P1.0 are considered non-inferior to the reference group proportion. When higher proportions are "Better", P1.0 should be less than P2.

When higher proportions are "Worse", P1.0 is the treatment group proportion such that proportions less than P1.0 are considered non-inferior to the reference group proportion. When higher proportions are "Worse", P1.0 should be greater than P2.

Proportions must be between 0 and 1. They cannot take on the values 0 or 1. This value should not be set to exactly the value of P2.

You may enter a range of values such as *0.03 0.05 0.10* or *0.01 to 0.05 by 0.01*.

P1.1 (Actual Group 1 Proportion)

This option specifies the value of P1.1, the value of the treatment proportion at which the power is to be calculated. It is only used for *Proportions*. Proportions must be between 0 and 1. They cannot take on the values 0 or 1.

You may enter a range of values such as *0.03 0.05 0.10* or *0.01 to 0.05 by 0.01*.

Data Tab (Differences)

This section documents options that are used when the parameterization is in terms of the difference, P1 – P2. P1.0 is the value of P1 assumed by the null hypothesis and P1.1 is the value of P1 at which the power is calculated. Once P2, D0, and D1 are given, the values of P1.1 and P1.0 can be calculated.

Effect Size – Differences

D0 (Non-Inferiority Difference)

Specify the difference between P1.0 and P2 that will be considered non-inferior.

When higher proportions are "Better", the non-inferiority difference is that amount that P1 can be less than P2 and still have the treatment group declared non-inferior to the reference group. When higher proportions are "Better", D0 should be negative.

When higher proportions are "Worse", the non-inferiority difference is that amount that P1 can be greater than P2 and still have the treatment group declared non-inferior to the reference group. When higher proportions are "Worse", D0 should be positive.

The power calculations assume that P1.0 is the value of the P1 under the null hypothesis. This value is used with P2 to calculate the value of P1.0 using the formula: P1.0 = D0 + P2.

You may enter a range of values such as *-.03 -.05 -.10* or *-.05 to -.01 by .01*. Differences must be between -1 and 1. D0 cannot take on the values -1, 0, or 1.

D1 (Actual Difference)

Specify the actual difference between P1.1 (the actual value of P1) and P2. This is the value at which the power is calculated.

In non-inferiority and equivalence trials, this difference is often set to zero.

The power calculations assume that P1.1 is the actual value of the proportion in group 1 (experimental or treatment group). This difference is used with P2 to calculate the value of P1 using the formula: P1.1 = D1 + P2.

You may enter a range of values such as *-.05 0 .5* or *-.05 to .05 by .02*. Actual differences must be between -1 and 1. They cannot take on the values -1 or 1.

This option is only used for *Differences*.

Data Tab (Ratios)

This section documents options that are used when the parameterization is in terms of the ratio, P1 / P2. P1.0 is the value of P1 assumed by the null hypothesis and P1.1 is the value of P1 at which the power is calculated. Once P2, R0, and R1 are given, the values of P1.0 and P1.1 can be calculated.

Effect Size – Ratios

R0 (Equivalence Ratio)

Specify the non-inferiority ratio of P1.0 to P2.

When higher proportions are "Better", the non-inferiority ratio is smallest ratio of P1 to P2 for which P1 will still be considered non-inferior. When higher proportions are "Better", R0 should be less than one.

When higher proportions are "Worse", the non-inferiority ratio is largest ratio of P1 to P2 for which P1 will still be considered non-inferior. When higher proportions are "Worse", R0 should be greater than one.

The power calculations assume that P1.0 is the value of P1 under the null hypothesis. This value is used with P2 to calculate the value of P1.0 using the formula: P1.0 = R0 x P2.

Ratios must be positive. R0 cannot take on the value of 1.

You may enter a range of values such as *0.95 .97 .99* or *.91 to .99 by .02*.

R1 (Actual Ratio)

Specify the actual ratio between P1.1 (the actual value of P1) and P2. This is the value at which the power is calculated.

In non-inferiority trials, this ratio is often set to one.

The power calculations assume that P1.1 is the actual value of the proportion in group 1 (experimental or treatment group). This ratio is used with P2 to calculate the value of P1 using the formula: P1.1 = R1 x P2.

Ratios must be positive. You may enter a range of values such as *0.95 1 1.05* or *0.9 to 1.9 by 0.02*.

Iterations/Zeroes Tab

The Iterations/Zeroes tab contains various limits and options.

Maximum Iterations

Maximum Iterations Before Search Termination

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 not reported. A value of at least 500 is recommended.

Zero Counts

Zero Count Adjustment Method

Zero cell counts often cause calculation problems. To compensate for this, a small value (called the Zero Count Adjustment Value) can be added either to all cells or to all cells with zero counts. This option specifies whether you want to use the adjustment and which type of adjustment you want to use. We recommend that you use the option *Add to zero cells only*.

Zero cell values often do not occur in practice. However, since power calculations are based on total enumeration, they will occur in power and sample size estimation.

Adding a small value is controversial, but can be necessary for computational considerations. Statisticians have recommended adding various fractions to zero counts. We have found that adding 0.0001 seems to work well.

Zero Count Adjustment Value

Zero cell counts cause many calculation problems when computing power or sample size. To compensate for this, a small value may be added either to all cells or to all zero cells. This value indicates the amount that is added. We have found that 0.0001 works well.

Be warned that the value of the ratio and the odds ratio will be affected by the amount specified here!

Example 1 – Finding Power

A study is being designed to study the effectiveness of a new treatment. Historically, the standard treatment has enjoyed a 60% cure rate. The new treatment reduces the seriousness of certain side effects that occur with the standard treatment. Thus, the new treatment will be adopted even if it is slightly less effective than the standard treatment. The researchers will recommend adoption of the new treatment if it has a cure rate of at least 55%. That is, the margin of inferiority is -5%.

The researchers will recruit patients from various hospitals. All patients at a particular hospital will receive the same treatment. They anticipate an average of 100 patients per hospital. Based on similar studies, they estimate the intracluster correlation to be 0.002.

The researchers plan to use the Farrington and Manning likelihood score test statistic to analyze the data. They want to study the power of the one-sided Farrington and Manning test at group cluster sizes ranging from 2 to 10 for detecting a difference of -0.05 when the actual cure rate of the new treatment ranges from 60% to 66%. The significance level will be 0.05.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Non-Inferiority Tests for Two Proportions in a Cluster-Randomized Design [Differences]** procedure window by expanding **Proportions**, then **Two Proportions – Cluster Randomized**, then clicking on **Non-Inferiority**, and then clicking on **Non-Inferiority Tests for Two Proportions in a Cluster-Randomized Design [Differences]**. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power	*Ignored since this is the Find setting*
Alpha	**0.05**
K1 (Clusters Per Group 1)	**2 4 6 8 10**
M1 (Items Per Cluster in Group 1)	**100**
K2 (Clusters Per Group 2)	**K1**
M2 (Items Per Cluster in Group 2)	**M1**
D0 (Equivalence Difference)	**-0.05**
D1 (Actual Difference)	**0 .02 .04 .06**
P2 (Group 2 Proportion)	**0.6**
ICC (Intracluster Correlation)	**0.002**
Higher Proportions Are	**Better**
Test Type	**Likelihood Score (Farr. & Mann.)**

Annotated Output

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

Numeric Results

Numeric Results for Non-Inferiority Tests Based on the Difference: P1 - P2
H0: P1-P2<=D0. H1: P1-P2=D1>D0. Test Statistic: Score test (Farrington & Manning)

Power	Group 1 Clusters/ Items K1/M1	Group 2 Clusters/ Items K2/M2	Intra-Cluster Corr. ICC	Grp 2 Prop P2	Non-Inf. Grp 1 Prop P1.0	Actual Grp 1 Prop P1.1	Non-Inf. Margin Diff D0	Actual Margin Diff D1	Alpha	Beta
0.2387	2/100	2/100	0.0020	0.6000	0.5500	0.6000	-0.0500	0.0000	0.0500	0.7613
0.3729	4/100	4/100	0.0020	0.6000	0.5500	0.6000	-0.0500	0.0000	0.0500	0.6271
0.4889	6/100	6/100	0.0020	0.6000	0.5500	0.6000	-0.0500	0.0000	0.0500	0.5111
0.5879	8/100	8/100	0.0020	0.6000	0.5500	0.6000	-0.0500	0.0000	0.0500	0.4121
0.6709	10/100	10/100	0.0020	0.6000	0.5500	0.6000	-0.0500	0.0000	0.0500	0.3291
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.

Summary Statements
Sample sizes of 200 in group one and 200 in group two, which were obtained by sampling 2 clusters with 100 subjects each in group one and 2 clusters with 100 subjects each in group two, achieve 24% power to detect a non-inferiority margin difference between the group proportions of -0.0500. The group two proportion is 0.6000. The group one proportion is assumed to be 0.5500 under the null hypothesis and 0.6000 under the alternative hypothesis. The test statistic used is the one-sided Score test (Farrington & Manning). The significance level of the test was 0.0500.

This report shows the values of each of the parameters, one scenario per row. Most of the report columns have obvious interpretations. Those that may not be obvious are presented here.

Group 1 Clusters/Items: K1/M1

This line gives the value of K1, the number of clusters in group 1, followed by M1, the number of items per cluster in this group. The total number of items sampled in group 1 is N1 = K1 x M1.

Group 2 Clusters/Items: K2/M2

This line gives the value of K2, the number of clusters in group 2, followed by M2, the number of items per cluster in this group. The total number of items sampled in group 2 is N2 = K2 x M2.

Intracluster Corr.: ICC

This is the value of the intracluster correlation coefficient, ICC.

Prop Grp 2: P2

This is the value of P2, the proportion responding positively in the control group.

Non-Inf. Grp 1 Prop P1.0

This is the value of P1.0, the response rate of the treatment group, as specified by the null hypothesis of inferiority. Values of P1 less than this amount are considered different from P2. Values of P1 greater than this are considered non-inferior to the reference group. The difference between this value and P2 is the value of the null hypothesis.

Actual Grp 1 Prop P1.1

This is the value of P1.1, the response rate of the treatment group, at which the power is computed. This is the value of P1 under the alternative hypothesis. The difference between this value and P2 is the value of the alternative hypothesis.

Non-Inf. Margin Diff D0

This is the value of D0, the difference between the two group proportions under the null hypothesis. This value is often called the *margin of non-inferiority*.

Actual Margin Diff D1

This is the value of D1, the difference between the two group proportions at which the power is computed. This is the value of the difference under the alternative hypothesis.

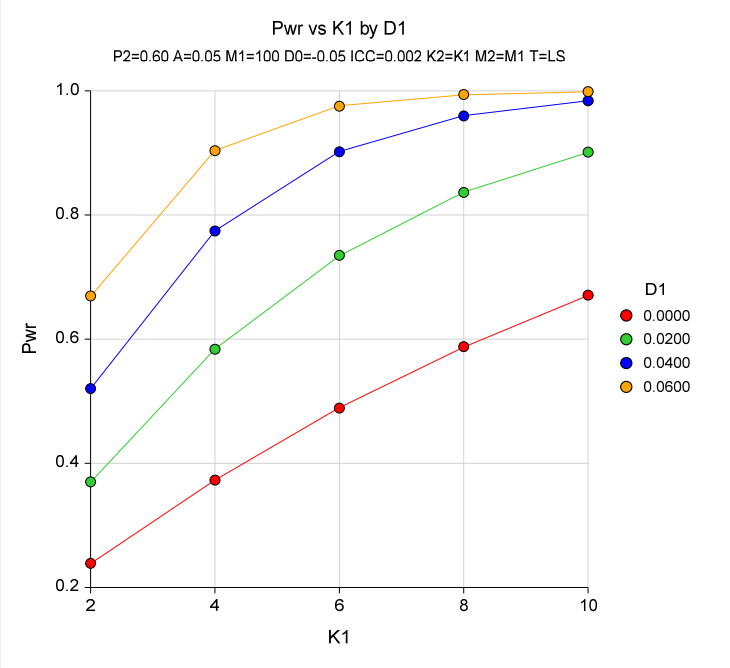
Alpha

This is the value of alpha (significance level) that was targeted by the design.

Beta

This is the value of beta, which is the probability of not rejecting a false null hypothesis.

Plots Section



The values from the table are displayed on the above chart. This chart gives us a quick look at the sample sizes that will be required for various values of D1.

Example 2 – Finding the Sample Size (Number of Clusters)

Continuing with the scenario given in Example 1, the researchers want to determine the number of clusters necessary for each value of D1 when the target power is set to 0.80.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Non-Inferiority Tests for Two Proportions in a Cluster-Randomized Design [Differences]** procedure window by expanding **Proportions**, then **Two Proportions – Cluster Randomized**, then clicking on **Non-Inferiority**, and then clicking on **Non-Inferiority Tests for Two Proportions in a Cluster-Randomized Design [Differences]**. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**K1**
Power	**0.80**
Alpha	**0.05**
K1 (Clusters Per Group 1)	*Ignored since this is the Find setting*
M1 (Items Per Cluster in Group 1)	**100**
K2 (Clusters Per Group 2)	**K1**
M2 (Items Per Cluster in Group 2)	**M1**
D0 (Equivalence Difference)	**-0.05**
D1 (Actual Difference)	**0 .02 .04 .06**
P2 (Group 2 Proportion)	**0.6**
ICC (Intracluster Correlation)	**0.002**
Higher Proportions Are	**Better**
Test Type	**Likelihood Score (Farr. & Mann.)**

Output

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

Numeric Results

Numeric Results for Non-Inferiority Tests Based on the Difference: P1 - P2
H0: P1-P2<=D0. H1: P1-P2=D1>D0. Test Statistic: Score test (Farrington & Manning)

Power	Group 1 Clusters/ Items K1/M1	Group 2 Clusters/ Items K2/M2	Intra-Cluster Corr. ICC	Grp 2 Prop P2	Non-Inf. Grp 1 Prop P1.0	Actual Grp 1 Prop P1.1	Non-Inf. Margin Diff D0	Actual Margin Diff D1	Alpha	Beta
0.8190	15/100	15/100	0.0020	0.6000	0.5500	0.6000	-0.0500	0.0000	0.0500	0.1810
0.8364	8/100	8/100	0.0020	0.6000	0.5500	0.6200	-0.0500	0.0200	0.0500	0.1636
0.8503	5/100	5/100	0.0020	0.6000	0.5500	0.6400	-0.0500	0.0400	0.0500	0.1497
0.8186	3/100	3/100	0.0020	0.6000	0.5500	0.6600	-0.0500	0.0600	0.0500	0.1814

The required sample size depends a great deal on the value of D1. The researchers should spend time determining the most appropriate value for D1.

Example 3 – Validation

We could not find an example of this type of analysis in the literature. Therefore, we will validate the procedure by comparing the results to those given in Example 3 in the chapter “Inequality Tests of Two Proportions in a Cluster-Randomized Design,” since both modules should give identical results. Validation can be accomplished by running Example 1 in this chapter and Example 3 in that chapter. If you do this, you will see that both procedures give the same results.

Example 4 – Finding Power after an Experiment

A group of researchers want to show that a new, less expensive treatment works at least as well as the current treatment. They believe, in fact, that the new treatment is about 0.10 higher in proportion of success. One hundred patients at each of 10 randomly chosen hospitals were given

the current treatment. One hundred patients at each of 10 randomly chosen hospitals were given the new treatment. It was agreed before the experiment that the new treatment needed to be no less than 0.05 in proportion of success below the current treatment to be considered non-inferior. The proportion of patients responding to the current treatment was 821/1000 = 0.821. The proportion of patients responding to the new treatment was 819/1000 = 0.819. This result did not show significant non-inferiority at the 0.05 level. The researchers want to know the power of their non-inferiority test. They decide to use the intracluster correlation coefficient estimated from the data, which was 0.0068. Although the observed difference in proportions is 0.819 – 0.821 = -0.002, the trivial difference is still -0.05. This value is used in the power calculation.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Non-Inferiority Tests for Two Proportions in a Cluster-Randomized Design [Differences]** procedure window by expanding **Proportions**, then **Two Proportions – Cluster Randomized**, then clicking on **Non-Inferiority**, and then clicking on **Non-Inferiority Tests for Two Proportions in a Cluster-Randomized Design [Differences]**. You may then make the appropriate entries as listed below, or open **Example 4** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power	*Ignored since this is the Find setting*
Alpha	**0.05**
K1 (Clusters Per Group 1)	**10**
M1 (Items Per Cluster in Group 1)	**100**
K2 (Clusters Per Group 2)	**K1**
M2 (Items Per Cluster in Group 2)	**M1**
D0 (Equivalence Difference)	**-0.05**
D1 (Actual Difference)	**0.0 0.10**
P2 (Group 2 Proportion)	**0.821**
ICC (Intracluster Correlation)	**0.0068**
Higher Proportions Are	**Better**
Test Type	**Likelihood Score (Farr. & Mann.)**

Output

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

Numeric Results

Numeric Results for Non-Inferiority Tests Based on the Difference: P1 - P2
H0: P1-P2<=D0. H1: P1-P2=D1>D0. Test Statistic: Score test (Farrington & Manning)

Power	Group 1 Clusters/ Items K1/M1	Group 2 Clusters/ Items K2/M2	Intra-Cluster Corr. ICC	Grp 2 Prop P2	Non-Inf. Grp 1 Prop P1.0	Actual Grp 1 Prop P1.1	Non-Inf. Margin Diff D0	Actual Margin Diff D1	Alpha	Beta
0.7272	10/100	10/100	0.0068	0.8210	0.7710	0.8210	-0.0500	0.0000	0.0500	0.2728
1.0000	10/100	10/100	0.0068	0.8210	0.7710	0.9210	-0.0500	0.1000	0.0500	0.0000

If indeed the new treatment were 0.10 higher in proportion of success, the power for showing non-inferiority would be 1.0000. If the true proportions are the same, the power would be 0.7272.

Example 5 – Finding Sample Size (Individuals within Clusters)

An agency would like to show the proportion of success of a new treatment is no less than that of the current treatment. Thirty doctors are available for the study. Fifteen will be randomly chosen to be trained to administer the new treatment. The remaining fifteen will continue to administer the current treatment. The new treatment will be considered non-inferior if the proportion of success is at least 90% of the current treatment success. The agency would like to know the number of patients that need to be treated by each doctor to achieve 80% power for the non-inferiority test. Various values for the intracluster correlation coefficient will be use since its true value is unknown. It is expected that the two treatments will have a success rate near 0.65. Alpha is set at 0.05.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Non-Inferiority Tests for Two Proportions in a Cluster-Randomized Design [Ratios]** procedure window by expanding **Proportions**, then **Two Proportions – Cluster Randomized**, then clicking on **Non-Inferiority**, and then clicking on **Non-Inferiority Tests for Two Proportions in a Cluster-Randomized Design [Ratios]**. You may then make the appropriate entries as listed below, or open **Example 5** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**M1**
Power	**0.80**
Alpha	**0.05**
K1 (Clusters Per Group 1)	**15**
M1 (Items Per Cluster in Group 1)	*Ignored since this is the find setting*
K2 (Clusters Per Group 2)	**K1**
M2 (Items Per Cluster in Group 2)	**M1**
R0 (Equivalence Ratio)	**0.90**
R1 (Actual Ratio)	**1.0**
P2 (Group 2 Proportion)	**0.65**
ICC (Intracluster Correlation)	**0.001 to 0.01 by 0.001**
Higher Proportions Are	**Better**
Test Type	**Likelihood Score (Farr. & Mann.)**

Output

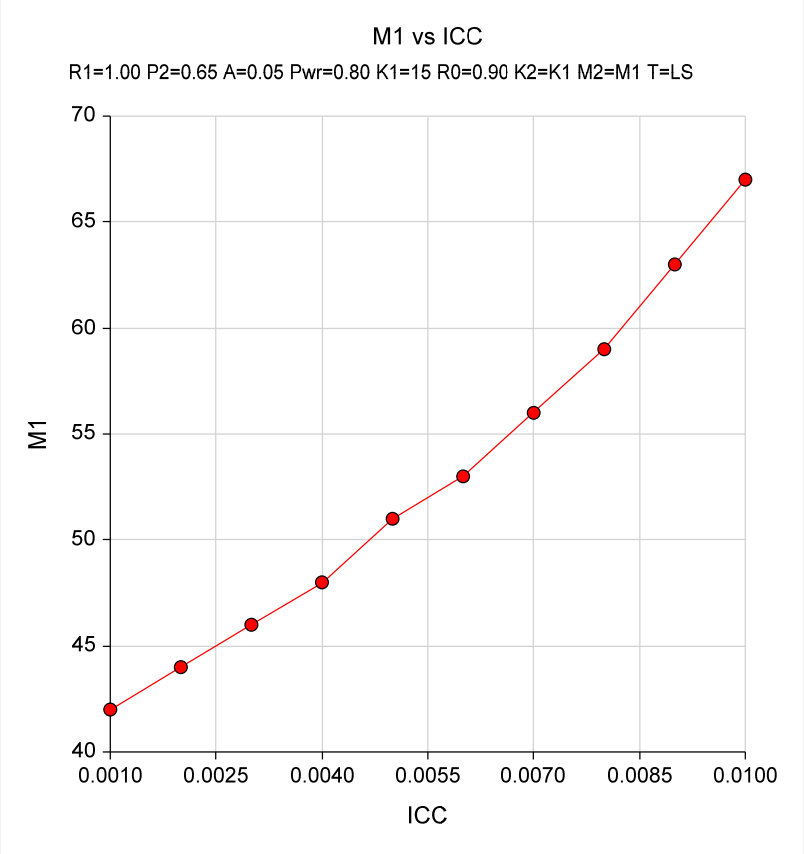
Click the Run button to perform the calculations and generate the following output.

Numeric Results

Numeric Results for Non-Inferiority Tests Based on the Ratio: P1 / P2
H0: P1/P2<=R0. H1: P1/P2=R1>R0. Test Statistic: Score test (Farrington & Manning)

Power	Group 1 Clusters/ Items K1/M1	Group 2 Clusters/ Items K2/M2	Intra-Cluster Corr. ICC	Grp 2 Prop P2	Trivial Grp 1 Prop P1.0	Actual Grp 1 Prop P1.1	Trivial Margin Ratio R0	Actual Margin Ratio R1	Alpha	Beta
0.8011	15/42	15/42	0.0010	0.6500	0.5850	0.6500	0.900	1.000	0.0500	0.1989
0.8023	15/44	15/44	0.0020	0.6500	0.5850	0.6500	0.900	1.000	0.0500	0.1977
0.8023	15/46	15/46	0.0030	0.6500	0.5850	0.6500	0.900	1.000	0.0500	0.1977
0.8017	15/48	15/48	0.0040	0.6500	0.5850	0.6500	0.900	1.000	0.0500	0.1983
0.8045	15/51	15/51	0.0050	0.6500	0.5850	0.6500	0.900	1.000	0.0500	0.1955
0.8011	15/53	15/53	0.0060	0.6500	0.5850	0.6500	0.900	1.000	0.0500	0.1989
0.8017	15/56	15/56	0.0070	0.6500	0.5850	0.6500	0.900	1.000	0.0500	0.1983
0.8005	15/59	15/59	0.0080	0.6500	0.5850	0.6500	0.900	1.000	0.0500	0.1995
0.8017	15/63	15/63	0.0090	0.6500	0.5850	0.6500	0.900	1.000	0.0500	0.1983
0.8011	15/67	15/67	0.0100	0.6500	0.5850	0.6500	0.900	1.000	0.0500	0.1989

Plots Section



The number of patients that should be seen by each doctor ranges from 42 to 67, depending on the intracluster correlation coefficient.

Chapter 240

Equivalence Tests for Two Proportions in a Cluster-Randomized Design

Introduction

This module provides power analysis and sample size calculation for equivalence tests in two-sample, cluster-randomized designs in which the outcome is binary.

Three Procedures Documented Here

There are three procedures in the menus that use the program module described in this chapter. These procedures are identical except for the type of parameterization. The parameterization can be in terms of proportions, differences in proportions, or ratios of proportions. Each of these options is listed separately on the menus.

Technical Details

The methods contained in this module are identical to those discussed in the chapter entitled “Inequality Tests for Two Proportions in a Cluster-Randomized Design.” The input and output has simply been reformatted in a manner that is convenient for equivalence testing. A complete review of equivalence testing is given in the chapter “Equivalence Tests for Two Proportions.”

We refer you to these two chapters for complete technical details on the methods used in this module.

Procedure Options

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

Data Tab (Common Options)

The Data tab contains the parameters associated with this test such as the proportions, sample sizes, alpha, and power. This chapter covers three procedures, each of which has different options. This section documents options that are common to all three procedures. Later, unique options for each procedure will be documented.

Solve For

Find (Solve For)

This option specifies the parameter to be solved for from the other parameters. The parameters that may be selected are *P1.1, Alpha, Power and Beta, K1, M1,* or *ICC*. Under most situations, you will select either *Power and Beta* or *K1*.

Select *K1* when you want to calculate the sample size needed to achieve a given power and alpha level.

Select *Power and Beta* when you want to calculate the power of an experiment.

Error Rates

Power or Beta

This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected.

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

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

Alpha (Significance Level)

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected.

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

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

Sample Size – Treatment (Group 1)

K1 (Clusters Group 1)

Enter a value (or range of values) for the number of clusters in this group. You may enter a range of values such as *10 to 20 by 2*. The sample size for this group is equal to the number of clusters times the number of subjects per cluster.

M1 (Items Group 1)

This is the average number of items (subjects) per cluster in group one. This value must be a positive number that is at least 1. You can use a list of values such as *100 150 200*.

Sample Size – Control (Group 2)

K2 (Clusters Group 2)

This is the number of clusters in group two. The sample size for this group is equal to the number of clusters times the number of subjects per cluster. This value must be a positive number.

If you simply want a multiple of the value for group one, you would enter the multiple followed by *K1*, with no blanks. If you want to use *K1* directly, you do not have to pre-multiply by 1. For example, all of the following are valid entries: *10 K1 2K1 0.5K1*.

You can use a list of values such as *10 20 30* or *K1 2K1 3K1*.

M2 (Items per Cluster in Group 2)

This is the number of items (subjects) per cluster in group two. This value must be a positive number.

If you simply want a multiple of the value for group one, you would enter the multiple followed by *M1*, with no blanks. If you want to use *M1* directly, you do not have to pre-multiply by 1. For example, all of the following are valid entries: *10 M1 2M1 0.5M1*.

You can use a list of values such as *10 20 30* or *M1 2M1 3M1*.

Effect Size – Control (Group 2)

P2 (Control Group Proportion)

Specify the value of P2, the control, baseline, or standard group's proportion. The null hypothesis is that the two proportions differ by a specified amount (See *Specify Group 1 Proportion using* below).

Since P2 is a proportion, these values must be between 0 and 1.

You may enter a range of values such as *0.1 0.2 0.3* or *0.1 to 0.9 by 0.1*.

Effect Size – Intracluster Correlation

ICC (Intracluster Correlation)

Enter a value (or range of values) for the intracluster correlation. This correlation may be thought of as the simple correlation between any two observations in the same cluster. It may also be

thought of as the proportion of total variance in the observations that can be attributed to difference between clusters.

Although the actual range for this value is from 0 to 1, typical values range from 0.002 to 0.05.

Test

Test Type

Specify which test statistic is used in searching and reporting. We recommend the likelihood score test.

Data Tab (Proportion)

This section documents options that are used when the parameterization is in terms of the values of the two proportions, P1 and P2. P1.0 is the value of the P1 assumed by the null hypothesis and P1.1 is the value of P1 at which the power is calculated.

Effect Size – Equivalence Proportions (Group 1)

P1.0U & P1.0L (Upper & Lower Equivalence Proportions)

Specify the *margin of equivalence* directly by giving the upper and lower bounds of P1.0. The two groups are assumed to be equivalent when P1.0 is between these values. Thus, P1.0U should be greater than P2 and P1.0L should be less than P2.

This option is only used for *Proportions*.

Note that the values of P1.0U and P1.0L are used in pairs. Thus, the first values of P1.0U and P1.0L are used together, and then the second values of each are used, and so on.

You may enter a range of values such as *0.03 0.05 0.10* or *0.01 to 0.05 by 0.01*.

Proportions must be between 0 and 1. They cannot take on the values 0 or 1. These values should surround P2.

Effect Size – Actual Proportion (Group 1)

P1.1 (Actual Proportion)

This option specifies the value of P1.1, which is the value of the treatment proportion at which the power is to be calculated. It is only used for *Proportions*. Proportions must be between 0 and 1. They cannot take on the values 0 or 1.

You may enter a range of values such as *0.03 0.05 0.10* or *0.01 to 0.05 by 0.01*.

Data Tab (Difference)

This section documents options that are used when the parameterization is in terms of the difference, P1 – P2. P1.0 is the value of P1 assumed by the null hypothesis and P1.1 is the value of P1 at which the power is calculated. Once P2, D0, and D1 are given, the values of P1.1 and P1.0 can be calculated.

Effect Size – Equivalence Differences

D0.U & D0.L (Upper & Lower Equivalence Difference)

Specify the *margin of equivalence* by specifying the largest distance above (D0.U) and below (D0.L) P2 which will still result in the conclusion of equivalence. As long as the actual difference is between these two values, the difference is not large enough to be of practical importance.

The values of D0.U must be positive and the values of D0.L must be negative. D0.L can be set to '-D0.U,' which is usually what is desired.

The power calculations assume that P1.0 is the value of the P1 under the null hypothesis. This value is used with P2 to calculate the value of P1.0 using the formula: P1.0U = D0.U + P2.

This option is only used for *Differences*.

You may enter a range of values for D0.U such as *.03 .05 .10* or *.05 to .20 by .05*.

Note that if you enter values for D0.L (other than '-D0.U'), they are used in pairs with the values of D0.U. Thus, the first values of D0.U and D0.L are used together, then the second values of each are used, and so on.

RANGE:

D0.L must be between -1 and 0. D0.U must be between 0 and 1. Neither can take on the values -1, 0, or 1.

Effect Size – Actual Difference

D1 (Actual Difference)

This option specifies the actual difference between P1.1 (the actual value of P1) and P2. This is the value of the difference at which the power is calculated. In equivalence trials, this difference is often set to zero.

The power calculations assume that P1.1 is the actual value of the proportion in group 1 (experimental or treatment group). This difference is used with P2 to calculate the true value of P1 using the formula: P1.1 = D1 + P2.

You may enter a range of values such as *-.05 0 .5* or *-.05 to .05 by .02*. Actual differences must be between -1 and 1. They cannot take on the values -1 or 1.

This option is only used for *Differences*.

Data Tab (Ratio)

This section documents options that are used when the parameterization is in terms of the ratio, P1 / P2. P1.0 is the value of P1 assumed by the null hypothesis and P1.1 is the value of P1 at which the power is calculated. Once P2, R0, and R1 are given, the values of P1.0 and P1.1 can be calculated.

Effect Size – Equivalence Ratios

R0.U & R0.L (Upper & Lower Equivalence Ratio)

Specify the *margin of equivalence* by specifying the largest ratio (P1/P2) above (R0.U), and below (R0.L), which will still result in the conclusion of equivalence. As long as the actual ratio is between these two values, the difference between the proportions is not said to be large enough to be of practical importance.

The values of R0.U must be greater than 1 and the values of R0.L must be less than 1. R0.L can be set to '1/R0.U', which is most often desired.

The power calculations assume that P1.0 is the value of the P1 under the null hypothesis. This value is used with P2 to calculate the value of P1.0 using the formula: P1.0U = R0.U x P2.

This option is only used for *Ratios*.

You may enter a range of values for R0.U such as *1.1 1.5 1.8* or *1.1 to 2.1 by 0.2*.

Note that if you enter values for R0.L (other than '1/R0.U'), they are used in pairs with the values of R0.U. Thus, the first values of R0.U and R0.L are used together, then the second values of each are used, and so on.

R0.L must be between 0 and 1. R0.U must be greater than 1. Neither can take on the value 1.

Effect Size – Actual Ratio

R1 (Actual Ratio)

This option specifies the ratio of P1.1 and P2, where P1.1 is the actual proportion in the treatment group. The power calculations assume that P1.1 is the actual value of the proportion in group 1. This difference is used with P2 to calculate the value of P1 using the formula: P1.1 = R1 x P2. In equivalence trials, this ratio is often set to 1.

This option is only used for *Ratios*.

Ratios must be positive. You may enter a range of values such as *0.95 1 1.05* or *0.9 to 1.9 by 0.02*.

Iterations/Zeroes Tab

The Iterations/Zeroes tab contains various limits and options.

Maximum Iterations

Maximum Iterations Before Search Termination

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 not reported. A value of at least 500 is recommended.

Zero Counts

Zero Count Adjustment Method

Zero cell counts often cause calculation problems. To compensate for this, a small value (called the Zero Count Adjustment Value) can be added either to all cells or to all cells with zero counts. This option specifies whether you want to use the adjustment and which type of adjustment you want to use. We recommend that you use the option *Add to zero cells only*.

Zero cell values often do not occur in practice. However, since power calculations are based on total enumeration, they will occur in power and sample size estimation.

Adding a small value is controversial, but can be necessary for computational considerations. Statisticians have recommended adding various fractions to zero counts. We have found that adding 0.0001 seems to work well.

Zero Count Adjustment Value

Zero cell counts cause many calculation problems when computing power or sample size. To compensate for this, a small value may be added either to all cells or to all zero cells. This value indicates the amount that is added. We have found that 0.0001 works well.

Be warned that the value of the ratio and the odds ratio will be affected by the amount specified here!

Example 1 – Finding Power

A study is being designed to establish the equivalence of a new treatment compared to the current treatment. Historically, the standard treatment has enjoyed a 60% cure rate. The new treatment reduces the seriousness of certain side effects that occur with the standard treatment. Thus, the new treatment will be adopted even if it is slightly less effective than the standard treatment. The researchers will recommend adoption of the new treatment if its cure rate is within 0.15 of the standard treatment.

The researchers will recruit patients from various hospitals. All patients at a particular hospital will receive the same treatment. They anticipate enlisting an average of 50 patients per hospital. Based on similar studies, they estimate the intracluster correlation to be 0.002.

The researchers plan to use the Farrington and Manning likelihood score test statistic to analyze the data. They want to study the power of the two, one-sided tests proposed by Farrington and Manning when the number of clusters per groups ranges from 2 to 10. They want to investigate the behavior of this test when the actual cure rate of the new treatment ranges from 60% to 66%. The significance level will be 0.05.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Equivalence Tests for Two Proportions in a Cluster-Randomized Design [Differences]** procedure window by expanding **Proportions**, then **Two Proportions – Cluster Randomized**, then clicking on **Equivalence**, and then clicking on **Equivalence Tests for Two Proportions in a Cluster-Randomized Design [Differences]**. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power	*Ignored since this is the Find setting*
Alpha	**0.05**
K1 (Clusters Per Group 1)	**2 4 6 8 10**
M1 (Items Per Cluster in Group 1)	**50**
K2 (Clusters Per Group 2)	**K1**
M2 (Items Per Cluster in Group 2)	**M1**
D0.U (Upper Equivalence Difference)	**0.15**
D0.L (Lower Equivalence Difference)	**-D0.U**
D1 (Difference|H1 = P1.1 – P2)	**0 .03 .06**
P2 (Group 2 Proportion)	**0.6**
ICC (Intracluster Correlation)	**0.002**
Test Type	**Likelihood Score (Farr. & Mann.)**

Annotated Output

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

Numeric Results

Numeric Results for Equivalence Tests Based on the Difference: P1 - P2
H0: P1-P2<=D0L or P1-P2>=D0U. H1: D0L<P1-P2=D1<D0U.
Test Statistic: Score tests

Power	Group 1 Clusters/ Items K1/M1	Group 2 Clusters/ Items K2/M2	Intra-Cluster Corr. ICC	Prop Grp 2 P2	Lower Equiv. Grp 1 Prop P1.0L	Upper Equiv. Grp 1 Prop P1.0U	Lower Equiv. Margin Diff D0.L	Upper Equiv. Margin Diff D0.U	Actual Margin Diff D1	Alpha	Beta
0.3459	2/50	2/50	0.0020	0.6000	0.4500	0.7500	-0.1500	0.1500	0.0000	0.0500	0.6541
0.8065	4/50	4/50	0.0020	0.6000	0.4500	0.7500	-0.1500	0.1500	0.0000	0.0500	0.1935
0.9494	6/50	6/50	0.0020	0.6000	0.4500	0.7500	-0.1500	0.1500	0.0000	0.0500	0.0506
0.9879	8/50	8/50	0.0020	0.6000	0.4500	0.7500	-0.1500	0.1500	0.0000	0.0500	0.0121
0.9973	10/50	10/50	0.0020	0.6000	0.4500	0.7500	-0.1500	0.1500	0.0000	0.0500	0.0027
0.3279	2/50	2/50	0.0020	0.6000	0.4500	0.7500	-0.1500	0.1500	0.0300	0.0500	0.6721
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.

Summary Statements
Sample sizes of 100 in group one and 100 in group two, which were obtained by sampling 2 clusters with 50 subjects each in group one and 2 clusters with 50 subjects each in group two, achieve 33% power to detect equivalence. The margin of equivalence, given in terms of the difference between the proportions, extends from -0.1500 to 0.1500. The actual difference between the proportions is 0.0000. The group two proportion is 0.6000. The calculations assume that two, one-sided z tests (unpooled) were used. The significance level of the test was 0.0500.

This report shows the values of each of the parameters, one scenario per row. Most of the report columns have obvious interpretations. Those that may not be obvious are presented here.

Group 1 Clusters/Items: K1/M1

This line gives the value of K1, the number of clusters in group 1, followed by M1, the number of items per cluster in this group. The total number of items sampled in group 1 is N1 = K1 x M1.

Group 2 Clusters/Items: K2/M2

This line gives the value of K2, the number of clusters in group 2, followed by M2, the number of items per cluster in this group. The total number of items sampled in group 2 is N2 = K2 x M2.

Intracluster Corr.: ICC

This is the value of the intracluster correlation coefficient, ICC.

Prop Grp 2: P2

This is the value of P2, the proportion responding positively in the control group.

Lower & Upper Equiv. Grp 1 Prop: P1.0L & P1.0U

These are the margin of equivalence for the response rate of the treatment group, as specified by the null hypothesis of non-equivalence. Values of P1 inside these limits are considered equivalent to P2.

Lower & Upper Equiv. Margin Diff: D0.L & D0.U

These set the margin of equivalence for the different in response rates. Values of the difference outside these limits are considered *non-equivalent*.

Actual Margin Diff D1

This is the value of D1, the difference between the two group proportions at which the power is computed. This is the value of the difference under the alternative hypothesis.

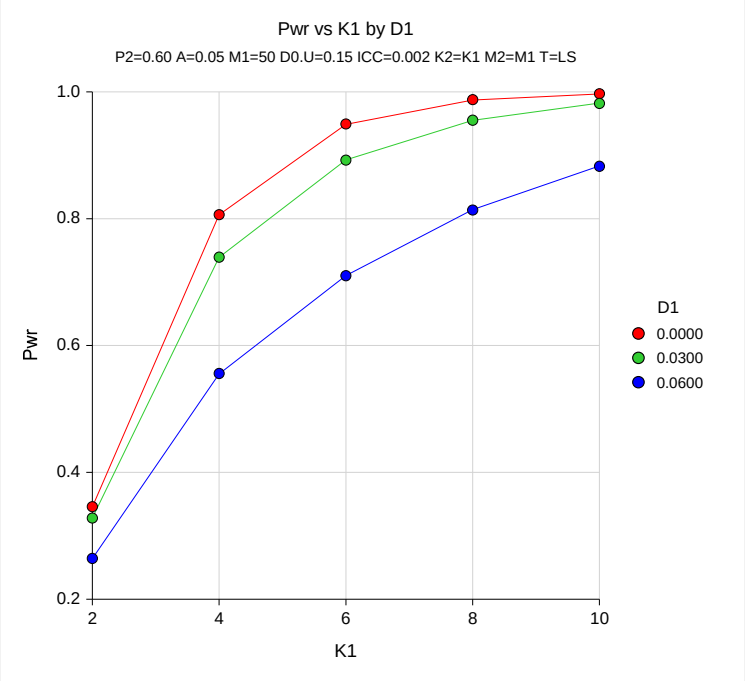
Alpha

This is the value of alpha (significance level) that was targeted by the design.

Beta

This is the value of beta, which is the probability of not rejecting a false null hypothesis.

Plots Section



The values from the table are displayed on the above chart. This chart gives us a quick look at the sample size that will be required for various values of D1.

Example 2 – Finding the Sample Size (Number of Clusters)

Continuing with the scenario given in Example 1, the researchers want to determine the number of clusters necessary for each value of D1 when the target power is set to 0.80.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Equivalence Tests for Two Proportions in a Cluster-Randomized Design [Differences]** procedure window by expanding **Proportions**, then **Two Proportions – Cluster Randomized**, then clicking on **Equivalence**, and then clicking on **Equivalence Tests for Two Proportions in a Cluster-Randomized Design [Differences]**. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**K1**
Power	**0.80**
Alpha	**0.05**
K1 (Clusters Per Group 1)	*Ignored since this is the Find setting*
M1 (Items Per Cluster in Group 1)	**50**
K2 (Clusters Per Group 2)	**K1**
M2 (Items Per Cluster in Group 2)	**M1**
D0.U (Upper Equivalence Difference)	**0.15**
D0.L (Lower Equivalence Difference)	**-D0.U**
D1 (Difference|H1 = P1.1 – P2)	**0 .03 .06**
P2 (Group 2 Proportion)	**0.6**
ICC (Intracluster Correlation)	**0.002**
Test Type	**Likelihood Score (Farr. & Mann.)**

Output

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

Numeric Results

Numeric Results for Equivalence Tests Based on the Difference: P1 - P2
H0: P1-P2<=D0L or P1-P2>=D0U. H1: D0L<P1-P2=D1<D0U.
Test Statistic: Score tests

Power	Group 1 Clusters/ Items K1/M1	Group 2 Clusters/ Items K2/M2	Intra-Cluster Corr. ICC	Prop Grp 2 P2	Lower Equiv. Grp 1 Prop P1.0L	Upper Equiv. Grp 1 Prop P1.0U	Lower Equiv. Margin Diff D0.L	Upper Equiv. Margin Diff D0.U	Actual Margin Diff D1	Alpha	Beta
0.8065	4/50	4/50	0.0020	0.6000	0.4500	0.7500	-0.1500	0.1500	0.0000	0.0500	0.1935
0.8324	5/50	5/50	0.0020	0.6000	0.4500	0.7500	-0.1500	0.1500	0.0300	0.0500	0.1676
0.8137	8/50	8/50	0.0020	0.6000	0.4500	0.7500	-0.1500	0.1500	0.0600	0.0500	0.1863

The required sample size depends a great deal on the value of D1. The researchers should spend time determining the most appropriate value for D1.

Example 3 – Validation

We could not find an example of this type of analysis in the literature. Therefore, we will validate the procedure by comparing the results to those given in the chapter entitled “Equivalence Tests for Two Proportions,” since both modules should give identical results for the same sample sizes when the ICC is set to zero. We ran the case when N1 = N2 = 200, P2 = 0.6, D0.U = 0.15, D1 = 0, and Alpha = 0.05. In this module, set M1 = 1 and set K1 = 200. Both program modules calculated the power to be 0.8482 in this case.

Example 4 – Finding Power after an Experiment

Individuals promoting a new, more expensive treatment claim that it achieves better results than the current treatment without citing statistical evidence. A group of researchers attempted to show the claim was false through a study involving 12 hospitals. Two hundred patients at each of 6 randomly chosen hospitals were given the current treatment. Two hundred patients at each of the remaining 6 hospitals were given the new treatment. It was agreed before the experiment that if a difference of less than 0.05 in proportion of success could be shown, the two treatments would be deemed equivalent. The proportion of patients responding properly to the current treatment was 540/1200 = 0.450. The proportion of patients responding properly to the new treatment was 570/1200 = 0.475. This result did not show significant equivalence at the 0.05 level. The researchers want to know the power of their equivalence test. They decide to use the intracluster correlation coefficient estimated from the data, which was 0.0043. Although the observed difference in proportions is 0.475 – 0.450 = 0.025, the equivalence difference is still 0.05. This value is used in the power calculation.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Equivalence Tests for Two Proportions in a Cluster-Randomized Design [Differences]** procedure window by expanding **Proportions**, then **Two Proportions – Cluster Randomized**, then clicking on **Equivalence**, and then clicking on **Equivalence Tests for Two Proportions in a Cluster-Randomized Design [Differences]**. You may then make the appropriate entries as listed below, or open **Example 4** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power	*Ignored since this is the Find setting*
Alpha	**0.05**
K1 (Clusters Per Group 1)	**6**
M1 (Items Per Cluster in Group 1)	**200**
K2 (Clusters Per Group 2)	**K1**
M2 (Items Per Cluster in Group 2)	**M1**
D0.U (Upper Equivalence Difference)	**0.05**
D0.L (Lower Equivalence Difference)	**-D0.U**
D1 (Difference|H1 = P1.1 – P2)	**0.0**
P2 (Group 2 Proportion)	**0.45**
ICC (Intracluster Correlation)	**0.0043**
Test Type	**Likelihood Score (Farr. & Mann.)**

Output

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

Numeric Results

Numeric Results for Equivalence Tests Based on the Difference: P1 - P2
H0: P1-P2<=D0.L or P1-P2>=D0.U. H1: D0.L<P1-P2=D1<D0.U.
Test Statistic: Score tests

Power	Group 1 Clusters/ Items K1/M1	Group 2 Clusters/ Items K2/M2	Intra-Cluster Corr. ICC	Prop Grp 2 P2	Lower Equiv. Grp 1 Prop P1.0L	Upper Equiv. Grp 1 Prop P1.0U	Lower Equiv. Margin Diff D0.L	Upper Equiv. Margin Diff D0.U	Actual Margin Diff D1	Alpha	Beta
0.1309	6/200	6/200	0.0043	0.4500	0.4000	0.5000	-0.0500	0.0500	0.0000	0.0500	0.8691

The power of the test of equivalence was only 0.1309.

Example 5 – Finding Sample Size (Individuals within Clusters)

An agency would like to show the proportion of success is the same for two treatments. Eight doctors are available for the study. Four will be randomly chosen to be trained to administer treatment 1. The remaining four will administer treatment 2. The treatments will be considered equivalent if the the proportion of success of treatment 1 is within 0.10 of treatment 2 success. The agency would like to know the number of patients that need to be treated by each doctor to achieve 80% power for the equivalence test. Various values for the intracluster correlation coefficient will be use since its true value is unknown. It is expected that the two treatments will have a success rate near 0.70. Alpha is set at 0.05.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Equivalence Tests for Two Proportions in a Cluster-Randomized Design [Differences]** procedure window by expanding **Proportions**, then **Two Proportions – Cluster Randomized**, then clicking on **Equivalence**, and then clicking on **Equivalence Tests for Two Proportions in a Cluster-Randomized Design [Differences]**. You may then make the appropriate entries as listed below, or open **Example 5** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**M1**
Power	**0.80**
Alpha	**0.05**
K1 (Clusters Per Group 1)	**4**
M1 (Items Per Cluster in Group 1)	*Ignored since this is the find setting*
K2 (Clusters Per Group 2)	**K1**
M2 (Items Per Cluster in Group 2)	**M1**
D0.U (Upper Equivalence Difference)	**0.10**
D0.L (Lower Equivalence Difference)	**-D0.U**
D1 (Difference|H1 = P1.1 – P2)	**0.0**
P2 (Group 2 Proportion)	**0.70**
ICC (Intracluster Correlation)	**0.001 to 0.01 by 0.001**
Test Type	**Likelihood Score (Farr. & Mann.)**

Output

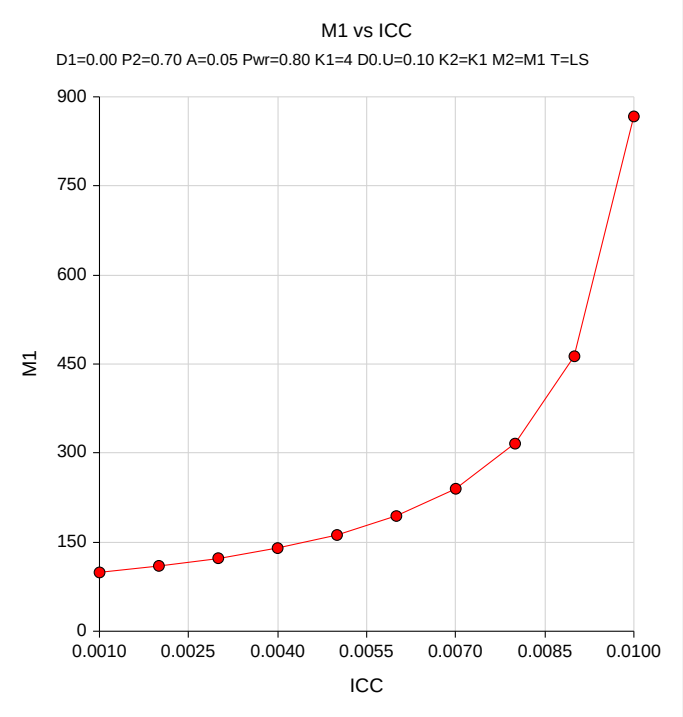
Click the Run button to perform the calculations and generate the following output.

Numeric Results

Numeric Results for Equivalence Tests Based on the Difference: P1 - P2
H0: P1-P2<=D0.L or P1-P2>=D0.U. H1: D0.L<P1-P2=D1<D0.U.
Test Statistic: Score tests

Power	Group 1 Clusters/ Items K1/M1	Group 2 Clusters/ Items K2/M2	Intra-Cluster Corr. ICC	Prop Grp 2 P2	Lower Equiv. Grp 1 Prop P1.0L	Upper Equiv. Grp 1 Prop P1.0U	Lower Equiv. Margin Diff D0.L	Upper Equiv. Margin Diff D0.U	Actual Margin Diff D1	Alpha	Beta
0.8028	4/99	4/99	0.0010	0.7000	0.6000	0.8000	-0.1000	0.1000	0.0000	0.0500	0.1972
0.8042	4/110	4/110	0.0020	0.7000	0.6000	0.8000	-0.1000	0.1000	0.0000	0.0500	0.1958
0.8028	4/123	4/123	0.0030	0.7000	0.6000	0.8000	-0.1000	0.1000	0.0000	0.0500	0.1972
0.8014	4/140	4/140	0.0040	0.7000	0.6000	0.8000	-0.1000	0.1000	0.0000	0.0500	0.1986
0.8014	4/162	4/162	0.0050	0.7000	0.6000	0.8000	-0.1000	0.1000	0.0000	0.0500	0.1986
0.8014	4/194	4/194	0.0060	0.7000	0.6000	0.8000	-0.1000	0.1000	0.0000	0.0500	0.1986
0.8014	4/240	4/240	0.0070	0.7000	0.6000	0.8000	-0.1000	0.1000	0.0000	0.0500	0.1986
0.8014	4/316	4/316	0.0080	0.7000	0.6000	0.8000	-0.1000	0.1000	0.0000	0.0500	0.1986
0.8014	4/463	4/463	0.0090	0.7000	0.6000	0.8000	-0.1000	0.1000	0.0000	0.0500	0.1986
0.8014	4/867	4/867	0.0100	0.7000	0.6000	0.8000	-0.1000	0.1000	0.0000	0.0500	0.1986

Plots Section



The number of patients needed to be treated by each doctor ranges from 99 to 867 depending on the value of the intracluster correlation coefficient.

Chapter 250

Chi-Square Tests for Multiple Proportions

Introduction

The *Chi-square test* is often used to test whether sets of frequencies or proportions follow certain patterns. The two most common instances are tests of goodness of fit using multinomial tables and tests of independence in contingency tables.

The *Chi-square goodness of fit* test is used to test whether the distribution of a set of data follows a particular pattern. For example, the goodness-of-fit Chi-square may be used to test whether a set of values follow the normal distribution or whether the proportions of Democrats, Republicans, and other parties are equal to a certain set of values, say 0.4, 0.4, and 0.2.

The *Chi-square test for independence* in a contingency table is the most common Chi-square 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. Hence, many surveys are analyzed with Chi-square tests.

The following table is an example of data arranged in a two-way contingency table. The rows of the table represent the stated political party of a respondent. The columns represent the respondent's answer to a question about whether they favor a certain proposition. The body of the table represents the number of individuals that fall into each cell (category). Note that the opinions of 311 individuals are recorded in this table.

(Count)	Favor Proposition A	
Political Party	Yes	No
Democrats	86	21
Republican	54	59
Others	34	57

The table below presents the row percentages for each category.

(Row Percentage)	Favor Proposition A	
Political Party	Yes	No
Democrats	80.4	19.6
Republican	47.8	52.2
Others	37.4	62.6

The Chi-square statistic tests whether the percentage of *Yes* responses remains constant across the three political parties. Notice that 80% of the Democrats said *Yes*, while only 37% of those in the Other category chose *Yes*. The Chi-square value for the above table is 41.71, which is statistically significant. Obviously, there is quite a shift in response pattern on this item across political parties.

Effect Size

We begin by defining what we will call the *effect size*. For each cell of a table containing *m* cells, suppose there are two proportions considered: one specified by a null hypothesis and the other specified by an alternative hypothesis. Often, the proportions specified by the alternative hypothesis are those occurring in the data. Define p_{0i} to be the proportion in cell *i* under the null hypothesis and p_{1i} to be the proportion in cell *i* under the alternative hypothesis. The effect size, *w*, is calculated using the formula

$$w = \sqrt{\sum_{i=1}^{m} \frac{(p_{1i} - p_{0i})^2}{p_{0i}}}.$$

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

$$\chi^2 = Nw^2$$

Note that when you are dealing with a contingency table, the cell index, *i*, is often replaced by two indices, one representing columns and the other representing rows.

The effect size, *w*, was used by Cohen (1988) because it does not depend on the sample size. He sets a small value of *w* at 0.1, a medium value at 0.3, and a large value at 0.5. Although these are rather arbitrary settings, they are useful for planning purposes.

Chi-Square Effect Size Estimator

PASS provides a special module to aid in finding an appropriate value for *w* called the Chi-Square Effect Size Estimator. This module may be loaded by pressing the *CS* button near the W (Effect Size) box or from the menus by selecting *PASS* and then *Other*.

You will find that as values are typed into the body of the table, the value of the effect size (shown at the bottom in the box labeled Effect Size - W) is also changed. Using this utility program, you can quickly determine the impact of table configurations on the value of *w*.

For example, suppose the cell proportions under the null and alternative hypotheses are as follows:

Cell	1	2	3	4	
p_{0i}	0.25	0.25	0.25	0.25	(Null: Equal distribution across the four cells.)
p_{1i}	0.40	0.20	0.20	0.20	(Alternative: cell 1 has twice the probability as the rest.)

To calculate *w*, first create the differences (ignoring the signs since the differences will be squared

|Diff|	0.15	0.05	0.05	0.05

Next, square the differences

Diff^2	0.0225	0.0025	0.0025	0.0025

Divide by the null

X^2	0.09	0.01	0.01	0.01

When these are summed, the result is 0.12. Taking the square of 0.12 gives the value of *w* as 0.3464.

As an experiment, load the Chi-Square Effect Size Estimator and enter these values on the Multinomial Test window. Enter the p_{0i} values in the column labeled Data Values and the p_{1i} values in1 the column marked Hypothesized Proportions. Check that the value of *w* is 0.3464. Next, change the Data Values to 4,2,2,2 and the Hypothesized Proportions to 1,1,1,1. Check that the value of *w* is still the same.

Calculating the Power

The power is calculated as follows:

1. Find x_α such that $1-\chi^2\left(x_\alpha|df\right)=\alpha$, where $\chi^2\left(x_\alpha|df\right)$ is the area to the left of *x* under a Chi-square distribution with *df* degrees of freedom.
2. Power = $1-\chi'^2_{df,\lambda}$, where $\chi'^2_{k,\lambda}$ is the left-tail area of the noncentral Chi-square distribution with *k* degrees of freedom and noncentrality parameter λ. Note that $\lambda = Nw^2$.

Procedure Options

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

Data Tab

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

Solve For

Find (Solve For)

This option specifies the parameter to be solved for from the other parameters. The parameters that may be selected are *W*, *DF*, *N*, *Alpha*, and *Power and Beta*. Under most situations, you will select either *Power and 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 *Power and Beta* when you want to calculate the power of an experiment that has already been run.

Error Rates

Power or Beta

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

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

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

Alpha (Significance Level)

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected. For this procedure, a type-I error occurs when you reject the null hypothesis of equal row (or column) proportions 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.

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

Sample Size

N (Sample Size)

This option specifies the number of individuals whose responses are recorded in the table. This number should be greater than or equal to the number of cells in the table.

Effect Size

W (Effect Size)

This is the value of *w*, the effect size. If you have Chi-square values that you want to analyze, use the following formula to transform them to *w's*:

$$w = \sqrt{\frac{\chi^2}{N}}$$

Remember that a small value of *w* is 0.1, a medium value is 0.3, and a large value is 0.5.

Test

DF (Degrees of Freedom)

This options specifies the degrees of freedom of the Chi-square test. For a test of independence in a contingency table, the degrees of freedom is *(R-1)(C-1)* where *R* is the number of rows and *C* is the number of columns. For example, for a 3-by-4 table, DF = (3-1)(4-1) = 6.

In a goodness of fit test, the degrees of freedom is the number of cells minus one. You may have to further adjust it for every distributional parameter that is estimated from the data. For example, suppose a Chi-square goodness-of-fit will be used to test the adequacy of the normality assumption on a set of 300 observations. Two parameters, the mean and variance, are estimated from the data. Suppose the data are categorized into six categories. DF = 6 - 2 - 1 = 3.

Example 1 – Finding the Power for an Existing Contingency Table

This example will compute the power of the Chi-square test of independence of the data in the contingency table that was discussed at the beginning of this chapter. If you would like to follow along, load the Chi-Square Effect Size Estimator window, select the Contingency Table tab, enter *86, 54, 34* in the first column and *21, 59, 57* in the second column. The results are Chi-square = 41.708829, *DF* = 2, *N* = 311, and *W* = 0.366213.

We will compute the power when alpha = 0.01, 0.05, and 0.10. For evaluation purposes, we will compute the power when *N* = 20, 50, 100, and 200 as well as at 311.

Setup

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

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power	*Ignored since this is the Find setting.*
Alpha	**0.01 0.05 0.10**
N (Sample Size)	**20 50 100 200 311**
W (Effect Size)	**0.366213**
DF (Degrees of Freedom)	**2**

Annotated Output

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

Numeric Results

Numeric Results

Power	N	W	Chi-Square	DF	Alpha	Beta
0.12127	20	0.3662	2.6822	2	0.01000	0.87873
0.29104	20	0.3662	2.6822	2	0.05000	0.70896
0.41007	20	0.3662	2.6822	2	0.10000	0.58993
0.39621	50	0.3662	6.7056	2	0.01000	0.60379
0.63538	50	0.3662	6.7056	2	0.05000	0.36462
0.74622	50	0.3662	6.7056	2	0.10000	0.25378
0.78214	100	0.3662	13.4112	2	0.01000	0.21786
0.91678	100	0.3662	13.4112	2	0.05000	0.08322
0.95512	100	0.3662	13.4112	2	0.10000	0.04488
0.98840	200	0.3662	26.8224	2	0.01000	0.01160
0.99795	200	0.3662	26.8224	2	0.05000	0.00205
0.99927	200	0.3662	26.8224	2	0.10000	0.00073
0.99980	311	0.3662	41.7088	2	0.01000	0.00020
0.99998	311	0.3662	41.7088	2	0.05000	0.00002
1.00000	311	0.3662	41.7088	2	0.10000	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. To conserve resources, it should be small.
W is the effect size--a measure of the magnitude of the Chi-Square that is to be detected.
DF is the degrees of freedom of the Chi-Square distribution.
Alpha is the probability of rejecting a true null hypothesis.
Beta is the probability of accepting a false null hypothesis.

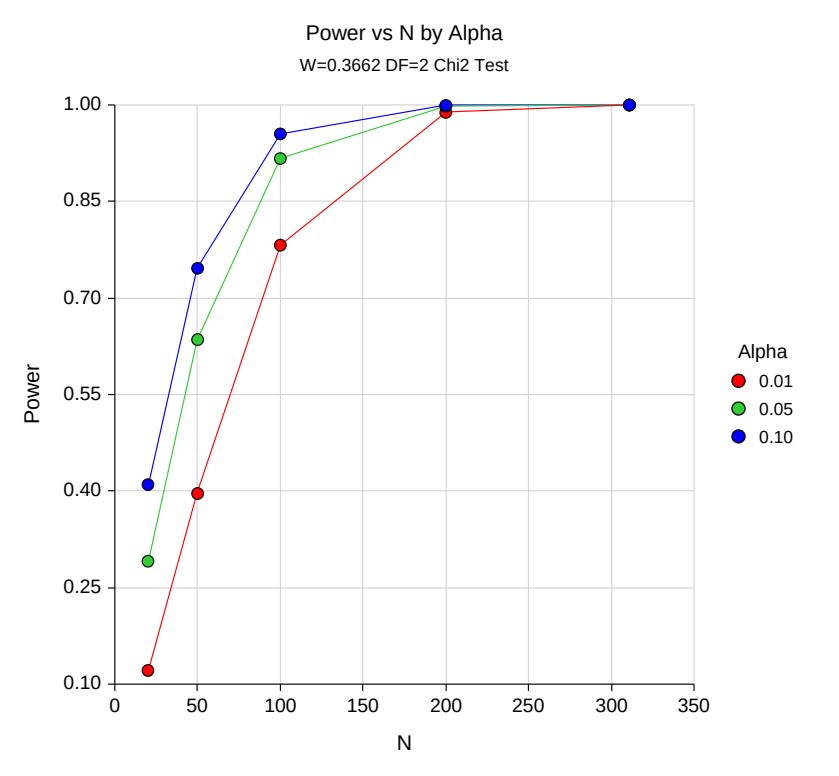
Summary Statements
A sample size of 20 achieves 12% power to detect an effect size (W) of 0.3662 using a 2 degrees of freedom Chi-Square Test with a significance level (alpha) of 0.01000.

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.

Note that in this particular example, a reasonable power of about 0.80 is reached for all values of alpha once the sample size is greater than 100.

The values from this table are plotted in the chart below.

Plots Section



This plot shows the relationship between sample size, power, and alpha.

Example 2 – Finding the Sample Size

A survey is being planned that will contain several questions with three possible answers: agree, neutral, disagree. The researchers are planning to analyze the questionnaires using Chi-square tests of independence in two-way contingency tables. How many respondents are needed to detect small (w = 0.1), medium (w = 0.3), or large (w = 0.5) effects if all hypothesis testing will be done at the 0.05 significance level?

Since the researchers are planning for 3-by-3 tables, DF = (3 - 1)(3 - 1) = 4.

Setup

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

Option	Value
Data Tab	
Find (Solve For)	**N**
Power	**0.80 0.90**
Alpha	**0.05**
N (Sample Size)	*Ignored since this is the Find setting.*
W (Effect Size)	**0.1 0.3 0.5**
DF (Degrees of Freedom)	**4**

Output

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

Numeric Results

Numeric Results

Power	N	W	Chi-Square	DF	Alpha	Beta
0.90010	1541	0.1000	15.4100	4	0.05000	0.09990
0.80018	1194	0.1000	11.9400	4	0.05000	0.19982
0.90157	172	0.3000	15.4800	4	0.05000	0.09843
0.80130	133	0.3000	11.9700	4	0.05000	0.19870
0.90198	62	0.5000	15.5000	4	0.05000	0.09802
0.80243	48	0.5000	12.0000	4	0.05000	0.19757

This report shows that for 80% power, 1194 (about 1200) respondents are needed to detect small effects, 133 respondents are needed to detect medium effects, and 48 respondents are needed to detect large effects.

Example 3 – Validation using Cohen

Cohen (1988) page 251 presents an example in which W = 0.30 and 0.40, N = 140, alpha = 0.01, and DF = 2. He gives the power as 0.75 for W = 0.3 and 0.97 for W = 0.4.

Setup

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

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power	*Ignored since this is the Find setting.*
Alpha	**0.01**
N (Sample Size)	**140**
W (Effect Size)	**0.3 0.4**
DF (Degrees of Freedom)	**2**

Output

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

Numeric Results

Numeric Results

Power	N	W	Chi-Square	DF	Alpha	Beta
0.74841	140	0.3000	12.6000	2	0.01000	0.25159
0.96641	140	0.4000	22.4000	2	0.01000	0.03359

PASS matches Cohen's power values of 0.75 and 0.97.

Example 4 – Finding the Sample Size for a Normality Goodness-of-Fit Test

A researcher is planning a study to determine if the distribution of scores on a certain test is normal. He plans to divide the test scores from his sample into five intervals of equal probability under the normal distribution using the sample mean and sample variance. After experimenting with the Chi-Square Effect Size Estimator, he decides that he must be able to detect a departure from normality of $w = 0.20$. He sets his significance level at 0.10 so that he will be lenient in his rejection of normality. He decides to focus on a power of 0.80. How large of a sample size will the researcher need?

The value of DF = 5 - 2 - 1 = 2, since there are five intervals and two parameters, mean and variance, are used.

Setup

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

Option	Value
Data Tab	
Find (Solve For)	**N**
Power	**0.80**
Alpha	**0.10**
N (Sample Size)	*Ignored since this is the Find setting.*
W (Effect Size)	**0.20**
DF (Degrees of Freedom)	**2**

Output

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

Numeric Results

Numeric Results

Power	N	W	Chi-Square	DF	Alpha	Beta
0.80046	193	0.2000	7.7200	2	0.10000	0.19954

This report shows that for 80% power, 193 observations are needed.

Chapter 253

Tests for Two Ordered Categorical Variables

Introduction

This module computes power and sample size for tests of ordered categorical data such as Likert scale data. Assuming proportional odds, such data can be analyzed directly with a z-test, or using logistic regression or the Mann-Whitney test. The power and sample size formulae presented here are consistent with any of these analysis methods. The results used here were presented in a paper by Whitehead (1993). They are also mentioned in the book by Julious (2010) and Machin et al (1997).

Ordered categorical data often results from surveys such as a quality of life (QoL) survey in which responses are categories such as *very good, good, moderate, poor*. When there are only two categories, an analysis using two proportions should be used. When there are more than two responses, and those responses can be ordered, the techniques described in this chapter can be used.

Technical Details

Suppose a variable has *K* possible responses $C_1,...,C_K$. Further suppose that these categories can be ordered so that C_i is more desirable than C_j if $i < j$. Hence C_1 is the best outcome and C_K is the worst. This procedure compares the results from two groups which will be called control (C) and experimental (E). The number of respondents falling within the i^{th} category of the control group is labeled N_{i1}. The total number in the control group is N_1 and in the experimental group is N_2. The total sample size of the study is $N = N_1 + N_2$.

Let p_{iE} denote the probability that an individual in the experimental group gives response C_i, and let Q_{iE} be the probability of an outcome of C_i or better. Thus $Q_{iE} = \sum_{j=1}^{i} p_{jE}$. Define p_{iC} and Q_{iC} similarly for the control group. Define the log-odds ratio for a particular category as

$$\theta_i = \log\left\{\frac{\dfrac{Q_{iE}}{1-Q_{iE}}}{\dfrac{Q_{iC}}{1-Q_{iC}}}\right\} \quad i = 1,\ldots,K-1.$$

This measures the advantage of the experimental group over the control group. A positive θ_i indicates that the experimental treatment is better than the control treatment.

Proportional Odds

The proportional odds model assumes that all of these log-odds ratios are equal to a common value θ. That is, the *proportional odds* assumption is that $\theta_1 = \cdots = \theta_{K-1} = \theta$. Thus the whole pattern of response differences can be summarized a single parameter.

The formulae to follow use the fact that efficient score *Y* is asymptotically normally distributed when θ is small and *n* is large. The test statistic and power formulae are as follows.

$$Y = \frac{1}{N+1}\sum_{i=1}^{k} N_{i1}\left(L_{iC} - U_{iC}\right)$$

$$Z = \frac{Y - \mu_Y}{\sigma_Y}$$

$$L_{iC} = \sum_{j=1}^{i-1} N_{j1}, \quad i = 2,...,K$$

$$U_{iC} = \sum_{j=i+1}^{K} N_{j1}, \quad i = 1,...,K-1$$

$$L_{1C} = U_{KC} = 0$$

$$\mu_Y = \theta V$$

$$\sigma_Y^2 = V$$

$$V = \frac{N_1 N_2 N}{3(N+1)^2}\left\{1 - \sum_{i=1}^{K}\left(\frac{N_{i1}+N_{i2}}{N}\right)^3\right\}$$

$$\approx \frac{N_1 N_2 N}{3(N+1)^2}\left\{1 - \sum_{i=1}^{K}\left(\frac{p_{iE}+p_{iC}}{2}\right)^3\right\}$$

The null hypothesis $H_0 : \theta = 0$ (the two treatments are equivalent) can be tested against the alternative $H_a : \theta \neq 0$ by computing Z and rejecting if Z is greater than $z_{\alpha/2}$. That is,

$$P\left(Z > z_{\alpha} \mid \theta = 0\right) = \alpha / 2$$

The power is the probability of rejecting a false null hypothesis, thus the power for a specified value θ_R is

$$\begin{aligned} Power &= P\left(Z > z_{\alpha/2} \mid \theta = \theta_R\right) \\ &= 1 - \Phi\left(z_{\alpha/2} - \theta_R\sqrt{V}\right) \end{aligned}$$

If a one-sided test is needed, replace $\alpha / 2$ with α.

Procedure Options

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

Data Tab

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

Solve For

Find (Solve For)

This option specifies the parameter to be solved for from the other parameters. The parameters that may be selected are *Theta*, *Power and Beta,* or *n (Group Sample Size)*.

Error Rates

Power or Beta

This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected. In this procedure, a type-II error occurs when you fail to reject the null hypothesis of $\theta = 0$ when in fact $\theta > 0$.

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

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

Alpha (Significance Level)

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected. In this procedure, a type-I error occurs when you reject the null hypothesis that $\theta = 0$ when in fact it is.

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.

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

Note that when you are analyzing a two-sided test, you should enter alpha, not alpha/2.

Sample Size

N1 (Control Group Sample Size)

This option specifies one or more values of the control group sample size. This value must be positive and greater than zero. Note that you may enter a list of values using the syntax *50,100,150,200,250* or *50 to 250 by 50*.

N2 (Experimental Group Sample Size)

Enter a value (or range of values) for the sample size of the experimental group. Enter ***Use R*** if you want N2 to be calculated using the formula: N2 = R(N1) where R is the Sample Allocation Ratio.

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. That is, R = N2/N1. 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). Note that setting R = 1.0 forces N2 = N1.

Effect Size

Control Group Category Proportions

This option contains the p_{iC}, the proportions in each category of the control group. The number of categories is implied by the number of items in the list.

For example, an entry of "20 30 40 10" implies that there are four categories and that 20% are category 1, 30% are category 2, 40% are category 3, and 10% are category 4.

The first category is the assumed to be the best outcome and the last category is the worst. Also, the categories are assumed to be ordered so that each is no better than the previous category. For example, you might have categories *Very Good, Good, Neutral, Poor, Very Poor*.

Note that the values entered here should to one. If they do not, they are scaled so that they do. For example, the values "20 30 40 10" (which sum to 100) are scaled to "0.2 0.3 0.4 0.1" (which sum to one).

You can enter any list of positive, non-zero numbers. For example, if you wanted to indicate that you have five, equally-likely categories, you could enter "1 1 1 1 1."

θ (Log Odds Ratio of Treatment Group)

Enter one or more values of θ, the logarithm of the odds. This value represents the effect size to be detected by the study. Note that the log odds ratio is assumed to be constant for all categories.

Examples

0.25 0.5 0.75 1.0

0.8 to 1.4 by 0.2

Test Type

Null Hypothesis

Indicate whether the hypothesis is one-sided or two-sided. If *two-sided* is selected, alpha is automatically divided by 2, so you do not need to half alpha.

Example 1 – Finding the Power

Suppose a clinical trial is planned to compare the response, made by a doctor, to certain treatment. The subjects are divided into two groups: those that will receive the current treatment and those that will receive an experimental treatment. Three months after the administration of the treatment, the doctor rates the response as *very good, good, moderate,* or *poor.* Historically, the responses have been about 20% *very good*, 50% *good*, 20% *moderate*, and 10% *poor*.

The researchers want to consider a range of possible value of θ from 0.5 to 2.0. They want to look at the power achieved by sample sizes from 30 to 150 per group. They want to set alpha to 0.05 and analyze the results with a two-sided test.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Test Comparing Two Ordered Categorical Variables** procedure window by expanding **Proportions**, then clicking on **Ordered Categorical**, and then clicking on **Test Comparing Two Ordered Categorical Variables**. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power	*Ignored since this is the Find setting*
Alpha	**0.05**
N1(Control Group Sample Size)	**30 to 150 by 10**
N2 (Experimental Group Sample Size)	**Use R**
R (Sample Allocation Ratio)	**1**
Control Group Category Proportions	**20 50 20 10**
θ (Log Odds Ratio of Treatment Group)	**0.5 to 2.0 by 0.5**
Null Hypothesis	**Two-Sided**

Annotated Output

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

Numeric Results

Numeric Results for Two-Sample, Ordered-Categorical Test
Null Hypothesis: θ = 0 Alternative Hypothesis: θ ≠ 0

Power	N1	N2	Log Odds Ratio θ	Alpha	Beta	Control Cat. 1 Prob. PC(1)	Exp'l Cat. 1 Prob. PE(1)
0.1726	30	30	0.5000	0.0500	0.8274	0.2000	0.2919
0.5310	30	30	1.0000	0.0500	0.4690	0.2000	0.4046
0.8653	30	30	1.5000	0.0500	0.1347	0.2000	0.5284
0.9828	30	30	2.0000	0.0500	0.0172	0.2000	0.6488
0.2172	40	40	0.5000	0.0500	0.7828	0.2000	0.2919
0.6564	40	40	1.0000	0.0500	0.3436	0.2000	0.4046
0.9444	40	40	1.5000	0.0500	0.0556	0.2000	0.5284
0.9972	40	40	2.0000	0.0500	0.0028	0.2000	0.6488
0.2613	50	50	0.5000	0.0500	0.7387	0.2000	0.2919
0.7543	50	50	1.0000	0.0500	0.2457	0.2000	0.4046
0.9784	50	50	1.5000	0.0500	0.0216	0.2000	0.5284
0.9996	50	50	2.0000	0.0500	0.0004	0.2000	0.6488
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.

References
Whitehead, John. 1993. 'Sample Size Calculations for Ordered Categorical Data.' Statistics in Medicine, 12, 2257-2271.
Julious, Steven A. 2009. Sample Sizes for Clinical Trials. Chapman & Hall/CRC. Boca Raton, Fl.
Machin, D., Campbell, M., Fayers, P., and Pinol, A. 1997. Sample Size Tables for Clinical Studies, 2nd Edition. Blackwell Science. Malden, Mass.

Report Definitions
Power is the probability of rejecting a false null hypothesis. It should be close to one.
Power is the probability of rejecting a false null hypothesis. It should be close to one.
N1 is the number of subjects in the control group.
N2 is the number of subjects in the experimental group.
θ is log odds ratio = log[odds(experimental)/odds(control)].
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.
Control Cat. 1 Prob is the probability of the first (best) category of the control group.
Exp'l Cat. 1 Prob is the probability of the first (best) category of the experimental group.

Summary Statements
Samples of 30 subjects in the control group and 30 subjects in experimental group achieve 17% power to detect a change in the log odds ratio (θ) of 0.5000 when the significance level (alpha) is 0.0500 using a two-sided test.

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 of rejecting a false null hypothesis.

N1, N2

This is the number of subjects in the control group and experimental group, respectively.

Log Odds Ratio

This is the log of the odds ratio. It is calculated using the formula

$$\theta = \log\left\{\frac{\dfrac{Q_{iE}}{1-Q_{iE}}}{\dfrac{Q_{iC}}{1-Q_{iC}}}\right\} \quad i = 1,\ldots,K-1.$$

Alpha

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

Beta

This is the probability of accepting a false null hypothesis.

Control Cat. 1 Prob. PC(1)

This is the value of p_{1C}, the probability of a control group subject responding in category 1.

Exp'l Cat. 1 Prob. PC(1)

This is the value of p_{1E}, the probability of an experimental group subject responding in category 1.

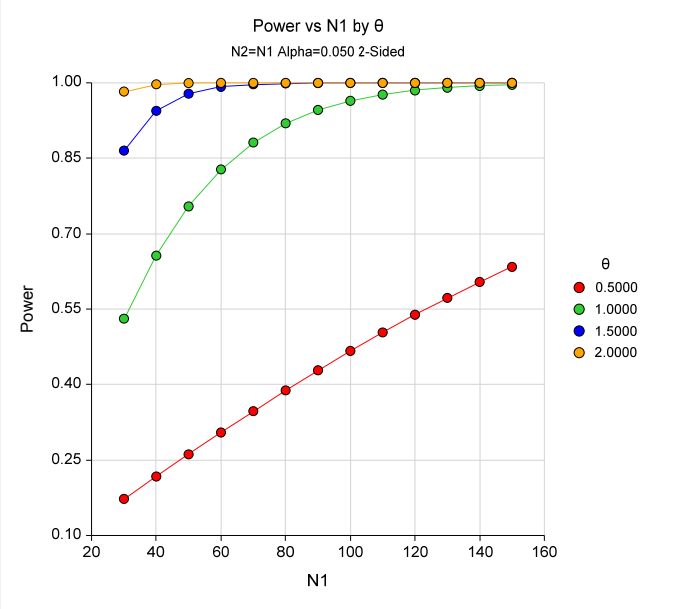
Category Probability Distribution

Category Probability Distribution

θ	Pr(1)	Pr(2)	Pr(3)	Pr(4)
0.000000	0.2000	0.5000	0.2000	0.1000
0.500000	0.2919	0.5018	0.1432	0.0631
1.000000	0.4046	0.4592	0.0969	0.0393
1.500000	0.5284	0.3843	0.0631	0.0242
2.000000	0.6488	0.2964	0.0400	0.0148

This report shows the individual response probabilities. The first row contains the results for the control group where θ is zero. The next row gives the values of p_{iE} for each value of θ. The goal of the report is to let you study the impact on the p_{iE} of each value of θ. You can make this assessment by watching how much these values change over the corresponding value in the first row.

Plots Section



This plot gives a visual presentation to the results in the Numeric Report.

Example 2 – Finding the Sample Size

Continuing with Example 1, the researchers want to find the sample size necessary to achieve 90% power when θ is 0.9.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Test Comparing Two Ordered Categorical Variables** procedure window by expanding **Proportions**, then clicking on **Ordered Categorical**, and then clicking on **Test Comparing Two Ordered Categorical Variables**. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**N1**
Power	**0.9**
Alpha	**0.05**
N1(Control Group Sample Size)	*Ignored since this is the Find setting*
N2 (Experimental Group Sample Size)	**Use R**
R (Sample Allocation Ratio)	**1**
Control Group Category Proportions	**20 50 20 10**
θ (Log Odds Ratio of Treatment Group)	**0.9**
Null Hypothesis	**Two-Sided**

Annotated Output

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

Numeric Results

Numeric Results for Two-Sample, Ordered-Categorical Test
Null Hypothesis: θ = 0 Alternative Hypothesis: θ ≠ 0

Power	N1	N2	Log Odds Ratio θ	Alpha	Beta	Control Cat. 1 Prob. PC(1)	Exp'l Cat. 1 Prob. PE(1)
0.9000	92	92	0.9000	0.0500	0.1000	0.2000	0.3808

The required sample size is 92 in each group.

Example 3 – Validation using Whitehead (1993)

Whitehead (1993) has an example in which he calculates the sample size to be 94 when θ is 0.887, alpha is 0.05, power is 90%, the control group proportions are 0.2, 0.5, 0.2, and 0.1.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Test Comparing Two Ordered Categorical Variables** procedure window by expanding **Proportions**, then clicking on **Ordered Categorical**, and then clicking on **Test Comparing Two Ordered Categorical Variables**. You may then make the appropriate entries as listed below, or open **Example 3** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**N1**
Power	**0.9**
Alpha	**0.05**
N1(Control Group Sample Size)	*Ignored since this is the Find setting*
N2 (Experimental Group Sample Size)	**Use R**
R (Sample Allocation Ratio)	**1**
Control Group Category Proportions	**20 50 20 10**
θ (Log Odds Ratio of Treatment Group)	**0.887**
Null Hypothesis	**Two-Sided**

Annotated Output

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

Numeric Results

Numeric Results for Two-Sample, Ordered-Categorical Test
Null Hypothesis: θ = 0 Alternative Hypothesis: θ ≠ 0

Power	N1	N2	Log Odds Ratio θ	Alpha	Beta	Control Cat. 1 Prob. PC(1)	Exp'l Cat. 1 Prob. PE(1)
0.9000	94	94	0.8870	0.0500	0.1000	0.2000	0.3777

PASS has matched the required sample size of 94 per group.

Chapter 255

Cochran-Armitage Test for Trend in Proportions

Introduction

This module computes power and sample size for the Cochran-Armitage test for a linear trend in proportions based on the results in Nam (1987). Asymptotic and exact power calculations for the uncorrected and continuity-corrected tests are available. The results assume that the proportions follow a linear trend on the logistic scale, with *X* being the covariate (or dose) variable, and that random samples are drawn from *k* separate populations.

Technical Details

Suppose we have *k* independent binomial variates y_i, with response probabilities p_i based on samples of size n_i at covariate (or dose) levels x_i, for *i* = 1, 2, …, *k*, where $x_1 < x_2 < ... < x_k$. Define the following:

$$N = \sum_{i=1}^{k} n_i$$

$$\overline{p} = \frac{1}{N}\sum_{i=1}^{k} y_i$$

$$\overline{q} = 1 - \overline{p}$$

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{k} n_i x_i$$

If we assume that the probability of response follows a linear trend on the logistic scale, then

$$p_i = \frac{\exp(\alpha + \beta x_i)}{1 + \exp(\alpha + \beta x_i)}.$$

Hypothesis Tests

The Cochran-Armitage test can be used to test the following hypotheses:

One-Sided (Increasing Trend)	$H_0 : p_1 = p_2 = ... = p_k$	vs. $H_1 : p_1 < p_2 < ... < p_k$
One-Sided (Decreasing Trend)	$H_0 : p_1 = p_2 = ... = p_k$	vs. $H_1 : p_1 > p_2 > ... > p_k$
Two-Sided	$H_0 : p_1 = p_2 = ... = p_k$	vs. $H_1 : p_1 < p_2 < ... < p_k$ or $p_1 > p_2 > ... > p_k$

One-Sided Test of Increasing Linear Trend in Proportions

Continuity-Corrected Test

Nam (1987) presents the following continuity-corrected asymptotic test statistic for detecting an increasing linear trend in proportions

$$z_{c.c.} = \frac{\sum_{i=1}^{k} y_i (x_i - \bar{x}) - \frac{\Delta}{2}}{\sqrt{\overline{pq}\left[\sum_{i=1}^{k} n_i (x_i - \bar{x})^2\right]}}.$$

The factor $\Delta / 2$ is the continuity correction. If the covariates x_i are equally-spaced then

$$\Delta = x_{i+1} - x_i \text{ for all } i < k$$

or the interval between adjacent covariates. ***PASS*** computes Δ for unequally-spaced covariates as

$$\Delta = \frac{1}{k-1} \sum_{i=1}^{k-1} (x_{i+1} - x_i).$$

For the case of unequally-spaced covariates, Nam (1987) has this to say, “For unequally spaced doses, no constant correction is adequate for all outcomes.” Therefore, we caution against the use of the continuity-corrected test statistic in the case of unequally-spaced covariates.

The test rejects H_0 if $z_{c.c.} \geq z_{1-\alpha}$, where $z_{1-\alpha}$ is the value that leaves $1 - \alpha$ in the upper tail of the standard normal distribution.

Uncorrected Test

The uncorrected test statistic is equivalent to the corrected test statistic except that $\Delta = 0$,

$$z = \frac{\sum_{i=1}^{k} y_i (x_i - \bar{x})}{\sqrt{\overline{pq}\left[\sum_{i=1}^{k} n_i (x_i - \bar{x})^2\right]}}.$$

The test rejects H_0 if $z \geq z_{1-\alpha}$, where $z_{1-\alpha}$ is the value that leaves $1 - \alpha$ in the upper tail of the standard normal distribution.

One-Sided Test of Decreasing Linear Trend in Proportions

Continuity-Corrected Test

Nam (1987) presents a continuity-corrected asymptotic test statistic for detecting an increasing linear trend in proportions. The continuity-corrected test statistic for a decreasing trend is the same as that for an increasing trend, except that $\Delta/2$ is added in the numerator instead of subtracted

$$z_{c.c.} = \frac{\sum_{i=1}^{k} y_i(x_i - \bar{x}) + \frac{\Delta}{2}}{\sqrt{\bar{p}\bar{q}\left[\sum_{i=1}^{k} n_i(x_i - \bar{x})^2\right]}}.$$

The factor Δ is defined the same as in the case of a test for increasing trend, and the caution about the use of the continuity-corrected test statistic in the case of unequally-spaced covariates also applies here.

The test rejects H_0 if $z_{c.c.} \le z_\alpha$, where z_α is the value that leaves α in the lower tail of the standard normal distribution.

Uncorrected Test

The uncorrected test statistic is equivalent to the corrected test statistic except that $\Delta = 0$,

$$z = \frac{\sum_{i=1}^{k} y_i(x_i - \bar{x})}{\sqrt{\bar{p}\bar{q}\left[\sum_{i=1}^{k} n_i(x_i - \bar{x})^2\right]}}.$$

The test rejects H_0 if $z \le z_\alpha$, where z_α is the value that leaves α in the lower tail of the standard normal distribution.

Two-Sided Test for Linear Trend in Proportions

Continuity-Corrected Test

Nam (1987) presents a continuity-corrected asymptotic test statistic for detecting an increasing linear trend in proportions. A two-sided test statistic utilizes a combination of the upper- and lower-tailed test statistics.

$$z_{c.c.U} = \frac{\sum_{i=1}^{k} y_i(x_i - \bar{x}) - \frac{\Delta}{2}}{\sqrt{\bar{p}\bar{q}\left[\sum_{i=1}^{k} n_i(x_i - \bar{x})^2\right]}} \quad \text{and} \quad z_{c.c.L} = \frac{\sum_{i=1}^{k} y_i(x_i - \bar{x}) + \frac{\Delta}{2}}{\sqrt{\bar{p}\bar{q}\left[\sum_{i=1}^{k} n_i(x_i - \bar{x})^2\right]}}.$$

The factor Δ is defined the same as in the case of a test for increasing trend, and the caution about the use of the continuity-corrected test statistic in the case of unequally-spaced covariates also applies here.

The test rejects H_0 if $z_{c.c.U} \ge z_{1-\alpha/2}$ or if $z_{c.c.L} \le z_{\alpha/2}$.

Uncorrected Test

The uncorrected test statistic is the same as the corrected test statistic except that $\Delta = 0$, which reduces the upper- and lower-tailed statistics to a single test statistic

$$z = \frac{\sum_{i=1}^{k} y_i (x_i - \bar{x})}{\sqrt{\overline{pq}\left[\sum_{i=1}^{k} n_i (x_i - \bar{x})^2\right]}}.$$

The test rejects H_0 if $|z| \geq z_{1-\alpha/2}$.

Exact Power Calculations

The power for the previous test statistics that are based on the normal approximation can be computed exactly using the binomial distribution. The following steps are taken to compute exact power.

One-Sided Test of Increasing Linear Trend in Proportions

1. Find the critical value using the standard normal distribution. The critical value, $z_{critical}$, is that value of *z* that leaves exactly the target value of alpha in the upper tail of the normal distribution. For example, for an upper-tailed test (increasing trend) with a target alpha of 0.05, the critical value is 1.645.
2. Compute the value of the test statistic, z_t, for every **y**, where $\mathbf{y} = (y_1, y_2, ..., y_k)$. Note that y_1 ranges from 0 to n_1, y_2 ranges from 0 to n_2, and so on. The test statistic z_t can be either the corrected or uncorrected test statistic.
3. If $z_t \geq z_{critical}$, the combination is in the rejection region. Call all **y** that lead to a rejection the set *A*.
4. Compute the power for given values of $\mathbf{p} = (p_1, p_2, ..., p_k)$ as

$$1 - \beta = \sum_{A} \left\{ \prod_{i=1}^{k} \left[\binom{n_i}{y_i} p_i^{y_i} (1 - p_i)^{n_i - y_i} \right] \right\}$$

When the values of n_i are large (say over 50) or *k* is large (say over 5), these formulas may take a little time to evaluate. In this case, a large sample approximation may be used.

One-Sided Test of Decreasing Linear Trend in Proportions

1. Find the critical value using the standard normal distribution. The critical value, $z_{critical}$, is that value of *z* that leaves exactly the target value of alpha in the upper tail of the normal distribution. For example, for a lower-tailed test (decreasing trend) with a target alpha of 0.05, the critical value is 1.645.
2. Compute the value of the test statistic, z_t, for every **y**, where $\mathbf{y} = (y_1, y_2, ..., y_k)$. Note that y_1 ranges from 0 to n_1, y_2 ranges from 0 to n_2, and so on. The test statistic z_t can be either the corrected or uncorrected test statistic.

3. If $z_t \le -z_{critical}$, the combination is in the rejection region. Call all **y** that lead to a rejection the set A.
4. Compute the power for given values of $\mathbf{p} = (p_1, p_2, ..., p_k)$ as

$$1 - \beta = \sum_A \left\{ \prod_{i=1}^{k} \left[\binom{n_i}{y_i} p_i^{y_i} (1 - p_i)^{n_i - y_i} \right] \right\}$$

When the values of n_i are large (say over 50) or k is large (say over 5), these formulas may take a little time to evaluate. In this case, a large sample approximation may be used.

Two-Sided Test of Linear Trend in Proportions

1. Find the critical value using the standard normal distribution. The critical value, $z_{critical}$, is that value of z that leaves exactly alpha/2 in the upper tail of the normal distribution. For example, for a two-sided test with a target alpha of 0.05, the critical value is 1.96.
2. Compute the value of the test statistics, z_U and z_L, for every **y**, where $\mathbf{y} = (y_1, y_2, ..., y_k)$. Note that y_1 ranges from 0 to n_1, y_2 ranges from 0 to n_2, and so on. The test statistics z_U and z_L can be either the corrected or uncorrected test statistics. In the case of the uncorrected test, $z_U = z_L$.
3. If $z_U \ge z_{critical}$ or $z_L \le -z_{critical}$, the combination is in the rejection region. Call all **y** that lead to a rejection the set A.
4. Compute the power for given values of $\mathbf{p} = (p_1, p_2, ..., p_k)$ as

$$1 - \beta = \sum_A \left\{ \prod_{i=1}^{k} \left[\binom{n_i}{y_i} p_i^{y_i} (1 - p_i)^{n_i - y_i} \right] \right\}$$

When the values of n_i are large (say over 50) or k is large (say over 5), these formulas may take a little time to evaluate. In this case, a large sample approximation may be used.

Approximate Power Calculation

The power for the Cochran-Armitage test can be computed quickly using the normal approximation to the binomial distribution. The following steps are taken to compute approximate power.

One-Sided Test of Increasing Linear Trend in Proportions

1. Find the critical value using the standard normal distribution. The critical value, $z_{critical}$, is that value of z that leaves exactly the target value of alpha in the upper tail of the normal distribution. For example, for an upper-tailed test with a target alpha of 0.05, the critical value is 1.645.
2. For a one-sided test of the alternative hypothesis that p_i is a monotone increasing function of x_i, compute the power for given values of $\mathbf{p} = (p_1, p_2, ..., p_k)$ as

$$1-\beta = \Pr\left(z \geq z_{critical} \mid H_1\right) = 1-\Phi\left(u_U\right)$$

where $\Phi()$ is the cumulative normal distribution and

$$u_U = \frac{-\left[\sum_{i=1}^{k} n_i p_i \left(x_i - \bar{x}\right) - \frac{\Delta}{2}\right] + z_{critical}\sqrt{p(1-p)\sum_{i=1}^{k} n_i \left(x_i - \bar{x}\right)^2}}{\sqrt{\sum_{i=1}^{k} n_i p_i \left(1-p_i\right)\left(x_i - \bar{x}\right)^2}},$$

where

$$p = \frac{1}{N}\sum_{i=1}^{k} n_i p_i .$$

The power for the uncorrected test is computed with $\Delta = 0$.

One-Sided Test of Decreasing Linear Trend in Proportions

1. Find the critical value using the standard normal distribution. The critical value, $z_{critical}$, is that value of z that leaves exactly the target value of alpha in the upper tail of the normal distribution. For example, for a lower-tailed test with a target alpha of 0.05, the critical value is 1.645.

2. For a one-sided test of the alternative hypothesis that p_i is a monotone decreasing function of x_i, compute the power for given values of $\mathbf{p} = (p_1, p_2, ..., p_k)$ as

$$1-\beta = \Pr\left(z \leq -z_{critical} \mid H_1\right) = \Phi\left(u_L\right)$$

where $\Phi()$ is the cumulative normal distribution and

$$u_L = \frac{-\left[\sum_{i=1}^{k} n_i p_i \left(x_i - \bar{x}\right) + \frac{\Delta}{2}\right] - z_{critical}\sqrt{p(1-p)\sum_{i=1}^{k} n_i \left(x_i - \bar{x}\right)^2}}{\sqrt{\sum_{i=1}^{k} n_i p_i \left(1-p_i\right)\left(x_i - \bar{x}\right)^2}},$$

where

$$p = \frac{1}{N}\sum_{i=1}^{k} n_i p_i .$$

The power for the uncorrected test is computed with $\Delta = 0$.

Two-Sided Test of Linear Trend in Proportions

1. Find the critical value using the standard normal distribution. The critical value, $z_{critical}$, is that value of z that leaves exactly alpha/2 in the upper tail of the normal distribution. For example, for a two-tailed test with a target alpha of 0.05, the critical value is 1.96.

2. For a two-sided test of the alternative hypothesis that p_i is a monotone decreasing or increasing function of x_i, compute the power for given values of $\mathbf{p} = (p_1, p_2, ..., p_k)$ as

$$\begin{aligned} 1-\beta &= \Pr(z_U \geq z_{critical} \mid H_1) + \Pr(z_L \leq -z_{critical} \mid H_1) \\ &= 1 - \Phi(u_U) + \Phi(u_L) \end{aligned},$$

   where $\Phi()$ is the cumulative normal distribution and u_U and u_L are as defined previously. The power for the uncorrected test is computed with $\Delta = 0$.

Procedure Options

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

Data Tab

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

Solve For

Find (Solve For)

This option specifies the parameter to be solved for from the other parameters. The parameters that may be selected are *Power and Beta* or *n (Group Sample Size)*.

Error Rates

Power or Beta

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

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

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

Alpha (Significance Level)

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

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

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

Sample Size / Groups – Sample Size Multiplier

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 $m_1, m_2, \ldots, m_k$ and this value is represented by n, the group sample sizes $n_1, n_2, \ldots, n_k$ are calculated as follows:

$n_1=[n(m_1)]$

$n_2=[n(m_2)]$

$n_3=[n(m_3)]$

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.

Sample Size / Groups – Groups

k (Number of Groups)

This is the number of groups being compared. Thus, it is the number of proportions and X values (or covariates). It must be greater than or equal to two.

The Cochran-Armitage method tests for a trend in proportions among these groups.

Note that the number of values used in the P (Proportions), X Values, and Group Sample Size Pattern boxes are all controlled by this number.

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 $n_1 = n_2 = n_3 = n$. That is, all sample sizes are equal to *n*.

Effect Size – Proportions

P (Proportions)

Specify two or more proportions. These are the alternative proportions for the Cochran-Armitage test of trend. The proportions should be strictly increasing or decreasing (depending on the alternative hypothesis) and all values should be greater than zero and less than one. The number of proportions entered should equal the value of *k*, the number of groups. If the number of proportions entered is less than *k*, the last proportion is repeated. If the number is greater than *k*, the extra proportions are ignored.

Several sets of proportions can be entered by using the ***PASS*** spreadsheet. To launch the spreadsheet, click on the "Spreadsheet" button above the box. To select columns from the spreadsheet, click on the button with the arrow pointing down to the right. Specify the column (or columns) to be used by beginning your entry with an equals sign, e.g. enter *=C1-C3*.

List Input

Specify a single set of proportions as a list. For example, with three groups you might enter *0.1 0.2 0.3*.

Spreadsheet Column Input

Specify more than one set of proportions using the column input syntax

=[column 1] [column 2] etc.

For example, if you have three proportion sets stored in the spreadsheet in columns C1, C2, and C3, you would enter *=C1 C2 C3* in the P (Proportions) box.

Each column in the spreadsheet corresponds to a single set of proportions. Missing cells are not allowed. The number of proportions entered in each column should equal the value of *k*. If the number of proportions entered is less than *k*, the last proportion is repeated. If the number is greater than *k*, the extra proportions are ignored.

Effect Size – X's (Covariate or Dose Values)

Equally-Spaced X Values

Check this box if the *x*'s (covariates or doses) are equally spaced. It is not necessary to specify the individual *x*'s if they are equally spaced, e.g. *x* value sets of 0, 1, 2 and 10, 20, 30 yield the same results.

If the covariates or doses are not equally spaced (e.g. *x*'s = 1, 3, 7), you should enter the individual values in the box below after unchecking this option. The continuity correction factor $\Delta/2$ and the power calculation then depend on the actual values entered.

X Values

Enter a list of *x* values if the covariates are unequally spaced (e.g. *x*'s = 1, 3, 7). The values should be strictly increasing. The continuity correction factor $\Delta/2$ and the power calculation then depend on the actual values entered here. The factor Δ is calculated as the average difference between adjacent *x* values.

If the "Equally-Spaced X Values" option is checked, these values are ignored.

Test

Test Type

Specify which type of test will be used in all searching and reporting.

The continuity correction refers to adding or subtracting $\Delta/2$ from the numerator of the *z*-value to bring the normal approximation closer to the binomial distribution. The factor Δ is calculated as the average difference between adjacent *x* values.

If the *x*'s are equally-spaced, Δ is equal to the difference between adjacent *x*'s

$$\Delta = x_{i+1} - x_i \text{ for all } i < k.$$

In the case of unequally-spaced *x*'s, Nam (1987) states, "For unequally spaced doses, no constant correction is adequate for all outcomes." Therefore, we recommend using the continuity-corrected test in the case of equally-spaced *x*'s, but caution against its use in the case of unequally-spaced *x*'s.

Alternative Hypothesis (H1)

Specify the alternative hypothesis of the test. Since the null hypothesis is the opposite, specifying the alternative is all that is needed. The alternative hypothesis determines how the alternative proportions, P (Proportions), should be entered. Usually, the two-sided option is selected.

For a one-sided alternative hypothesis test of increasing trend, the proportions should be strictly increasing, e.g. *0.1 0.2 0.3*.

For a one-sided alternative hypothesis test of decreasing trend, the proportions should be strictly decreasing, e.g. *0.5 0.4 0.3*.

For a two-sided test, the proportions can be either increasing or decreasing.

Exact Power Calculation

Maximum Group Sample Size for Exact Power Calculations

When all group sample sizes are less than or equal to this amount and the product of all group sample sizes is less than the Max Group Size Product for Exact Power Calculations, exact power calculations using the binomial distribution are made. Otherwise, the normal approximation to the binomial is used for calculating power.

Large values for this option can greatly increase the time required to calculate the power, especially when searching for sample size or when k is large. For larger sample sizes, the power based on the normal approximation is very close to the exact power. We recommend keeping this value less than 50.

Maximum Group Sample Size Product for Exact Power Calculations

When the product of all group sample sizes is less than this amount and all individual group sample sizes are less than the Maximum Group Sample Size for Exact Power Calculations, exact power calculations using the binomial distribution are made. Otherwise, the normal approximation to the binomial is used for calculating power.

This option is used to reduce computing time and to avoid running out of memory in the case of large sample sizes and/or large k. Raising this value will increase computing time.

Large values for this option can greatly increase the time required to calculate the power, especially when searching for sample size or when k is large. For larger sample sizes, the power based normal approximation is very close to the exact power. We recommend keeping this value less than 10 million when solving for power and beta and less than 1 million when solving for sample size.

Example 1 – Finding the Power

An experiment is being designed to determine if there exists a dose-response relationship for a particular drug. Researchers will administer the drug at three dose levels: control (no drug), low, and high. The low dose is exactly half of the high dose, so the dosage structure is equally spaced. They expect to find response proportions of 0.05, 0.15, and 0.25 corresponding to the three doses, control, low, and high, respectively. A two-sided test with an alpha level of 0.05 will be used along with the continuity-corrected Cochran-Armitage test. They wish to compute the power for conducting the study with equal-sized groups ranging from 30 to 70 subjects in size.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Cochran-Armitage Test for Trend in Proportions** procedure window by expanding **Proportions**, then clicking on **Trend**, and then clicking on **Cochran-Armitage Test for Trend in Proportions**. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power	*Ignored since this is the Find setting*
Alpha	**0.05**
n (Sample Size Multiplier)	**30 to 70 by 5**
k (Number of Groups)	**3**
Group Sample Size Pattern	**Equal**
P (Proportions)	**0.05 0.15 0.25**
Equally-Spaced X Values	**Checked**
Test Type	**Z test with continuity correction**
Alternative Hypothesis (H1)	**Two-Sided**
Max Grp Sample Size for Exact Power	**20**
Max Grp Size Product for Exact Power	**1000000**

Annotated Output

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

Numeric Results

Numeric Results
Test Type = Two-sided Z test with continuity correction. Correction Factor = 0.5.
H0: P1 = P2 = ... = Pk. H1: Increasing or Decreasing Trend.
Equally-Spaced X Values.

Power	Average n	k	Total N	Alpha	Beta	Proportions
0.51187	30.00	3	90	0.05000	0.48813	0.05, 0.15, 0.25
0.58893	35.00	3	105	0.05000	0.41107	0.05, 0.15, 0.25
0.65710	40.00	3	120	0.05000	0.34290	0.05, 0.15, 0.25
0.71640	45.00	3	135	0.05000	0.28360	0.05, 0.15, 0.25
0.76724	50.00	3	150	0.05000	0.23276	0.05, 0.15, 0.25
0.81029	55.00	3	165	0.05000	0.18971	0.05, 0.15, 0.25
0.84635	60.00	3	180	0.05000	0.15365	0.05, 0.15, 0.25
0.87629	65.00	3	195	0.05000	0.12371	0.05, 0.15, 0.25
0.90093	70.00	3	210	0.05000	0.09907	0.05, 0.15, 0.25

References
Nam, J. 1987. 'A Simple Approximation for Calculating Sample Sizes for Detecting Linear Trend in Proportions'. Biometrics 43, 701-705.

Report Definitions
Power is the probability of rejecting a false null hypothesis. It should be close to one.
Average n is the average group sample size.
k is the number of groups.
Total N is the total sample size of all groups combined.
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.
Proportions lists the set of proportions used. The number of proportions is equal to k.

Summary Statements
In a Cochran-Armitage test for trend in proportions, sample sizes of 30, 30, and 30 are obtained from 3 groups with equally-spaced X values and proportions equal to 0.05, 0.15, and 0.25, respectively. The total sample of 90 subjects achieves 51% power to detect a linear trend using a two-sided Z test with continuity correction and a significance level of 0.05000.

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.

Proportions

The alternative proportions used to calculate the power.

Detailed Results Report

Details when Power = 0.51187 and Alpha = 0.05000

Group	Sample Size (Ni)	Percent Ni of Total N	Proportion (Pi)
1	30	33.33	0.05
2	30	33.33	0.15
3	30	33.33	0.25
ALL	90	100.00	

Details when Power = 0.58893 and Alpha = 0.05000

Group	Sample Size (Ni)	Percent Ni of Total N	Proportion (Pi)
1	35	33.33	0.05
2	35	33.33	0.15
3	35	33.33	0.25
ALL	105	100.00	

(More Reports Follow)

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 total sample size for the scenario.

Ni

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

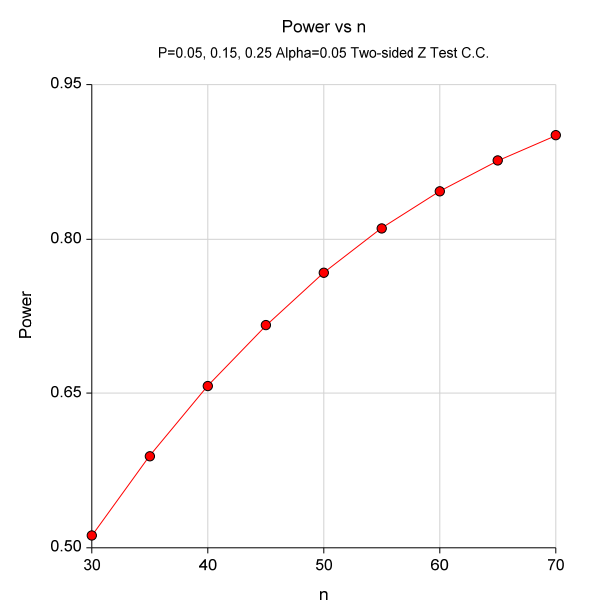
Percent Ni of Total Ni

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

Pi

This is the value of the hypothesized proportion.

Plots 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.

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.

Exact Power Calculation

You can calculate the exact power for this scenario by setting the maximum group sample size for exact power calculations to *70*. You can do this yourself or load the completed template **Example1b** from the Template tab on the procedure window.

Numeric Results
Test Type = Two-sided Z test with continuity correction. Correction Factor = 0.5.
H0: P1 = P2 = ... = Pk. H1: Increasing or Decreasing Trend.
Equally-Spaced X Values.

Power	Average n	k	Total N	Alpha	Beta	Proportions
0.51173*	30.00	3	90	0.05000	0.48827	0.05, 0.15, 0.25
0.60387*	35.00	3	105	0.05000	0.39613	0.05, 0.15, 0.25
0.67534*	40.00	3	120	0.05000	0.32466	0.05, 0.15, 0.25
0.74067*	45.00	3	135	0.05000	0.25933	0.05, 0.15, 0.25
0.78352*	50.00	3	150	0.05000	0.21648	0.05, 0.15, 0.25
0.83170*	55.00	3	165	0.05000	0.16830	0.05, 0.15, 0.25
0.86462*	60.00	3	180	0.05000	0.13538	0.05, 0.15, 0.25
0.89489*	65.00	3	195	0.05000	0.10511	0.05, 0.15, 0.25
0.91511*	70.00	3	210	0.05000	0.08489	0.05, 0.15, 0.25

* Values in this row are based on exact power calculations. Exact power was calculated for scenarios in which the largest group sample size is less than or equal to 70 and the product of all group samples sizes is less than or equal to 1000000.

This report indicates that all power values were calculated exactly based on the binomial distribution. The approximate power values calculated earlier are very close to these values.

Example 2 – Finding the Sample Size

Continuing with the last example, we will determine how large the sample size would need to be to have the power at least 0.95 with an alpha level of 0.05.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Cochran-Armitage Test for Trend in Proportions** procedure window by expanding **Proportions**, then clicking on **Trend**, and then clicking on **Cochran-Armitage Test for Trend in Proportions**. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**n (Group Sample Size)**
Power	**0.95**
Alpha	**0.05**
n (Sample Size Multiplier)	*Ignored since this is the Find setting*
k (Number of Groups)	**3**
Group Sample Size Pattern	**Equal**
P (Proportions)	**0.05 0.15 0.25**
Equally-Spaced X Values	**Checked**
Test Type	**Z test with continuity correction**
Alternative Hypothesis (H1)	**Two-Sided**
Max Grp Sample Size for Exact Power	**20**
Max Grp Size Product for Exact Power	**1000000**

Output

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

Numeric Results

Numeric Results
Test Type = Two-sided Z test with continuity correction. Correction Factor = 0.5.
H0: P1 = P2 = ... = Pk. H1: Increasing or Decreasing Trend.
Equally-Spaced X Values.

	Average		Total			
Power	**n**	**k**	**N**	**Alpha**	**Beta**	**Proportions**
0.95054	85.00	3	255	0.05000	0.04946	0.05, 0.15, 0.25

The required sample size is 85 per group or 255 subjects.

Example 3 – Calculating Power with Unequal Group Sample Sizes

Continuing with the last example, consider the impact of allowing the group sample sizes to be unequal. Suppose we have twice as many control subjects receiving no drug as subjects at the low and high dose levels. What is the power for group sample sizes of 120, 60, and 60?

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Cochran-Armitage Test for Trend in Proportions** procedure window by expanding **Proportions**, then clicking on **Trend**, and then clicking on **Cochran-Armitage Test for Trend in Proportions**. You may then make the appropriate entries as listed below, or open **Example 3** by going to the **File** menu and choosing **Open Example Template**.

Pay particular attention to how the sample size parameters were changed. The sample size multiplier, n, was set to 1 so that it is essentially ignored. The Group Sample Size Pattern contains the three sample sizes.

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power	*Ignored since this is the Find setting*
Alpha	**0.05**
n (Sample Size Multiplier)	**1**
k (Number of Groups)	**3**
Group Sample Size Pattern	**120 60 60**
P (Proportions)	**0.05 0.15 0.25**
Equally-Spaced X Values	**Checked**
Test Type	**Z test with continuity correction**
Alternative Hypothesis (H1)	**Two-Sided**
Max Grp Sample Size for Exact Power	**20**
Max Grp Size Product for Exact Power	**1000000**

Output

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

Numeric Results

Numeric Results
Test Type = Two-sided Z test with continuity correction. Correction Factor = 0.5.
H0: P1 = P2 = ... = Pk. H1: Increasing or Decreasing Trend.
Equally-Spaced X Values.

Power	Average n	k	Total N	Alpha	Beta	Proportions
0.95196	80.00	3	240	0.05000	0.04804	0.05, 0.15, 0.25

Details when Power = 0.95196 and Alpha = 0.05000

Group	Sample Size (Ni)	Percent Ni of Total N	Proportion (Pi)
1	120	50.00	0.05
2	60	25.00	0.15
3	60	25.00	0.25
ALL	240	100.00	

Group sample sizes of 120, 60, and 60 yield just over 95% power. The total sample size of 240 for 95% power for this scenario is actually less than the total of 255 from Example 2, where equal group sample sizes were used.

Example 4 – Calculating Power with Unequally-Spaced X Values

Continuing with Example 1, consider the impact of using unequally-spaced dose levels: 0, 2, and 5. Because the doses are not equally spaced, we will use the uncorrected *z* test for power calculations.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Cochran-Armitage Test for Trend in Proportions** procedure window by expanding **Proportions**, then clicking on **Trend**, and then clicking on **Cochran-Armitage Test for Trend in Proportions**. You may then make the appropriate entries as listed below, or open **Example 4** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power	*Ignored since this is the Find setting*
Alpha	**0.05**
n (Sample Size Multiplier)	**30 to 70 by 5**
k (Number of Groups)	**3**
Group Sample Size Pattern	**Equal**
P (Proportions)	**0.05 0.15 0.25**
Equally-Spaced X Values	**Unchecked**
X Values	**0 2 5**
Test Type	**Z test**
Alternative Hypothesis (H1)	**Two-Sided**
Max Grp Sample Size for Exact Power	**20**
Max Grp Size Product for Exact Power	**1000000**

Output

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

Numeric Results

Numeric Results
Test Type = Two-sided Z test.
H0: P1 = P2 = ... = Pk. H1: Increasing or Decreasing Trend.
X Values = 0.0, 2.0, 5.0.

Power	Average n	k	Total N	Alpha	Beta	Proportions
0.57754	30.00	3	90	0.05000	0.42246	0.05, 0.15, 0.25
0.64383	35.00	3	105	0.05000	0.35617	0.05, 0.15, 0.25
0.70190	40.00	3	120	0.05000	0.29810	0.05, 0.15, 0.25
0.75214	45.00	3	135	0.05000	0.24786	0.05, 0.15, 0.25
0.79514	50.00	3	150	0.05000	0.20486	0.05, 0.15, 0.25
0.83161	55.00	3	165	0.05000	0.16839	0.05, 0.15, 0.25
0.86229	60.00	3	180	0.05000	0.13771	0.05, 0.15, 0.25
0.88790	65.00	3	195	0.05000	0.11210	0.05, 0.15, 0.25
0.90915	70.00	3	210	0.05000	0.09085	0.05, 0.15, 0.25

The power values are quite different from those calculated with the continuity-corrected *z* test when the dose-spacing is equal. Of course, the covariate spacing you use will likely depend on more factors than the achievable power.

Example 5 – Validation of Sample Size Calculations with Approximate Power using Nam

Nam (1987) page 703 presents a table of calculated sample sizes with three equally-spaced doses and equal group sample sizes using the one-sided continuity-corrected *z* test for an increasing trend in proportions. Sample size is calculated for various proportion sets, alpha levels of 0.05 and 0.025, and power values of 0.5, 0.7, and 0.9. The table based on approximate power calculations is given below.

Alternative Proportion			Specified Nominal Power					
			.50		.70		.90	
p_0	p_1	p_2	α = .025	α = .05	α = .025	α = .05	α = .025	α = .05
.05	.10	.15	79	58	120	94	197	162
.10	.15	.20	108	79	167	129	276	226
.20	.25	.30	154	111	241	186	402	329
.30	.35	.40	185	133	290	223	486	398
.05	.15	.25	29	22	44	34	70	58
.10	.20	.30	36	26	54	42	87	72
.20	.30	.40	45	33	69	54	113	93
.30	.40	.50	51	37	78	61	129	106
.05	.25	.45	11	9	16	13	25	21
.10	.30	.50	12	9	18	14	28	23
.20	.40	.60	14	10	20	16	32	26
.30	.50	.70	14	11	21	17	33	28

This example will replicate these results.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Cochran-Armitage Test for Trend in Proportions** procedure window by expanding **Proportions**, then clicking on **Trend**, and then clicking on **Cochran-Armitage Test for Trend in Proportions**. You may then make the appropriate entries as listed below, or open **Example 5** by going to the **File** menu and choosing **Open Example Template**. Check to see that the values have been entered into the spreadsheet by clicking the spreadsheet button to the right of P (Proportions).

Option	Value
Data Tab	
Find (Solve For)	**n (Group Sample Size)**
Power	**0.5 0.7 0.9**
Alpha	**0.025 0.05**
n (Sample Size Multiplier)	*Ignored since this is the Find setting*
k (Number of Groups)	**3**
Group Sample Size Pattern	**Equal**
P (Proportions)	**=C1-C12**
Equally-Spaced X Values	**Checked**
Test Type	**Z test with continuity correction**
Alternative Hypothesis (H1)	**One-Sided (Increasing Trend)**
Max Grp Sample Size for Exact Power	**0**
Max Grp Size Product for Exact Power	**1000000**

Output

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

Numeric Results

Numeric Results
Test Type = One-sided Z test with continuity correction. Correction Factor = 0.5.
H0: P1 = P2 = ... = Pk. H1: P1 < P2 < ... < Pk.
Equally-Spaced X Values.

Power	Average n	k	Total N	Alpha	Beta	Proportions
0.90012	197.00	3	591	0.02500	0.09988	0.05, 0.10, 0.15
0.70301	121.00	3	363	0.02500	0.29699	0.05, 0.10, 0.15
0.50098	79.00	3	237	0.02500	0.49902	0.05, 0.10, 0.15
0.90150	163.00	3	489	0.05000	0.09850	0.05, 0.10, 0.15
0.70061	94.00	3	282	0.05000	0.29939	0.05, 0.10, 0.15
0.50493	59.00	3	177	0.05000	0.49507	0.05, 0.10, 0.15
0.90025	276.00	3	828	0.02500	0.09975	0.10, 0.15, 0.20
0.70115	167.00	3	501	0.02500	0.29885	0.10, 0.15, 0.20
0.50110	108.00	3	324	0.02500	0.49890	0.10, 0.15, 0.20
0.90073	227.00	3	681	0.05000	0.09927	0.10, 0.15, 0.20
0.70244	130.00	3	390	0.05000	0.29756	0.10, 0.15, 0.20
0.50156	79.00	3	237	0.05000	0.49844	0.10, 0.15, 0.20
0.90008	402.00	3	1206	0.02500	0.09992	0.20, 0.25, 0.30

0.70057	241.00	3	723	0.02500	0.29943	0.20, 0.25, 0.30
0.50029	154.00	3	462	0.02500	0.49971	0.20, 0.25, 0.30
0.90065	330.00	3	990	0.05000	0.09935	0.20, 0.25, 0.30
0.70052	186.00	3	558	0.05000	0.29948	0.20, 0.25, 0.30
0.50249	112.00	3	336	0.05000	0.49751	0.20, 0.25, 0.30
0.90003	486.00	3	1458	0.02500	0.09997	0.30, 0.35, 0.40
0.70141	291.00	3	873	0.02500	0.29859	0.30, 0.35, 0.40
0.50078	185.00	3	555	0.02500	0.49922	0.30, 0.35, 0.40
0.90019	398.00	3	1194	0.05000	0.09981	0.30, 0.35, 0.40
0.70102	224.00	3	672	0.05000	0.29898	0.30, 0.35, 0.40
0.50023	133.00	3	399	0.05000	0.49977	0.30, 0.35, 0.40
0.90093	70.00	3	210	0.02500	0.09907	0.05, 0.15, 0.25
0.70523	44.00	3	132	0.02500	0.29477	0.05, 0.15, 0.25
0.51187	30.00	3	90	0.02500	0.48813	0.05, 0.15, 0.25
0.90203	58.00	3	174	0.05000	0.09797	0.05, 0.15, 0.25
0.70988	35.00	3	105	0.05000	0.29012	0.05, 0.15, 0.25
0.50072	22.00	3	66	0.05000	0.49928	0.05, 0.15, 0.25
0.90039	87.00	3	261	0.02500	0.09961	0.10, 0.20, 0.30
0.70361	54.00	3	162	0.02500	0.29639	0.10, 0.20, 0.30
0.50579	36.00	3	108	0.02500	0.49421	0.10, 0.20, 0.30
0.90182	72.00	3	216	0.05000	0.09818	0.10, 0.20, 0.30
0.70083	42.00	3	126	0.05000	0.29917	0.10, 0.20, 0.30
0.50913	27.00	3	81	0.05000	0.49087	0.10, 0.20, 0.30
0.90216	114.00	3	342	0.02500	0.09784	0.20, 0.30, 0.40
0.70640	70.00	3	210	0.02500	0.29360	0.20, 0.30, 0.40
0.50791	46.00	3	138	0.02500	0.49209	0.20, 0.30, 0.40
0.90010	93.00	3	279	0.05000	0.09990	0.20, 0.30, 0.40
0.70220	54.00	3	162	0.05000	0.29780	0.20, 0.30, 0.40
0.50912	34.00	3	102	0.05000	0.49088	0.20, 0.30, 0.40
0.90012	129.00	3	387	0.02500	0.09988	0.30, 0.40, 0.50
0.70399	79.00	3	237	0.02500	0.29601	0.30, 0.40, 0.50
0.50022	51.00	3	153	0.02500	0.49978	0.30, 0.40, 0.50
0.90046	106.00	3	318	0.05000	0.09954	0.30, 0.40, 0.50
0.70142	61.00	3	183	0.05000	0.29858	0.30, 0.40, 0.50
0.50717	38.00	3	114	0.05000	0.49283	0.30, 0.40, 0.50
0.90119	25.00	3	75	0.02500	0.09881	0.05, 0.25, 0.45
0.72689	17.00	3	51	0.02500	0.27311	0.05, 0.25, 0.45
0.53014	12.00	3	36	0.02500	0.46986	0.05, 0.25, 0.45
0.90649	21.00	3	63	0.05000	0.09351	0.05, 0.25, 0.45
0.70730	13.00	3	39	0.05000	0.29270	0.05, 0.25, 0.45
0.51914	9.00	3	27	0.05000	0.48086	0.05, 0.25, 0.45
0.90139	28.00	3	84	0.02500	0.09861	0.10, 0.30, 0.50
0.70192	18.00	3	54	0.02500	0.29808	0.10, 0.30, 0.50
0.52280	13.00	3	39	0.02500	0.47720	0.10, 0.30, 0.50
0.90025	23.00	3	69	0.05000	0.09975	0.10, 0.30, 0.50
0.72817	15.00	3	45	0.05000	0.27183	0.10, 0.30, 0.50
0.52790	10.00	3	30	0.05000	0.47210	0.10, 0.30, 0.50
0.90163	32.00	3	96	0.02500	0.09837	0.20, 0.40, 0.60
0.71861	21.00	3	63	0.02500	0.28139	0.20, 0.40, 0.60
0.50324	14.00	3	42	0.02500	0.49676	0.20, 0.40, 0.60
0.90863	27.00	3	81	0.05000	0.09137	0.20, 0.40, 0.60
0.70206	16.00	3	48	0.05000	0.29794	0.20, 0.40, 0.60
0.52289	11.00	3	33	0.05000	0.47711	0.20, 0.40, 0.60
0.90793	34.00	3	102	0.02500	0.09207	0.30, 0.50, 0.70
0.72308	22.00	3	66	0.02500	0.27692	0.30, 0.50, 0.70
0.52104	15.00	3	45	0.02500	0.47896	0.30, 0.50, 0.70
0.90750	28.00	3	84	0.05000	0.09250	0.30, 0.50, 0.70
0.71329	17.00	3	51	0.05000	0.28671	0.30, 0.50, 0.70
0.50788	11.00	3	33	0.05000	0.49212	0.30, 0.50, 0.70

The sample sizes calculated by ***PASS*** match those of Nam (1987). In many cases, ***PASS*** reports a sample size that is one greater than that reported Nam (1987). This difference is due to rounding. Nam (1987) rounds some power values up when they are actually slightly lower than the nominal value. ***PASS*** does not round power values up when computing the sample size. All sample sizes result in at least the nominal power.

Example 6 – Validation of Exact Power Calculations using Nam

Nam (1987) page 703 presents a table of calculated sample sizes with three equally-spaced doses and equal group sample sizes using the one-sided continuity-corrected *z* test for an increasing trend in proportions. Sample size is calculated for various proportion sets, alpha levels of 0.05 and 0.025, and power values of 0.5, 0.7, and 0.9. The table of calculated sample sizes is given in Example 5. Nam (1987) further calculates the exact power for scenarios in which the resulting sample size is less than or equal to 50. The results are given below.

Alternative Proportion			Specified Nominal Power					
			.50		.70		.90	
p_0	p_1	p_2	α = .025	α = .05	α = .025	α = .05	α = .025	α = .05
.05	.15	.25	29 (.51)	22 (.51)	44 (.73)	34 (.71)		
.10	.20	.30	36 (.52)	26 (.50)		42 (.71)		
.20	.30	.40	45 (.50)	33 (.51)				
.30	.40	.50		37 (.49)				
.05	.25	.45	11 (.50)	9 (.57)	16 (.71)	13 (.71)	25 (.92)	21 (.91)
.10	.30	.50	12 (.50)	9 (.54)	18 (.72)	14 (.71)	28 (.91)	23 (.91)
.20	.40	.60	**14 (.53)**	10 (.47)	20 (.71)	16 (.69)	32 (.90)	26 (.89)
.30	.50	.70	**14 (.53)**	11 (.50)	21 (.69)	17 (.69)	33 (.90)	28 (.91)

This example will replicate the results in bold type.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Cochran-Armitage Test for Trend in Proportions** procedure window by expanding **Proportions**, then clicking on **Trend**, and then clicking on **Cochran-Armitage Test for Trend in Proportions**. You may then make the appropriate entries as listed below, or open **Example 6** by going to the **File** menu and choosing **Open Example Template**. Check to see that the values have been entered into the spreadsheet by clicking the spreadsheet button to the right of P (Proportions).

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power	*Ignored since this is the Find setting*
Alpha	**0.025**
n (Sample Size Multiplier)	**14**
k (Number of Groups)	**3**
Group Sample Size Pattern	**Equal**
P (Proportions)	**=C11-C12**
Equally-Spaced X Values	**Checked**
Test Type	**Z test with continuity correction**
Alternative Hypothesis (H1)	**One-Sided (Increasing Trend)**

Data Tab (continued)
Max Grp Sample Size for Exact Power...**50**
Max Grp Size Product for Exact Power ..**1000000**

Output

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

Numeric Results

Numeric Results
Test Type = One-sided Z test with continuity correction. Correction Factor = 0.5.
H0: P1 = P2 = ... = Pk. H1: P1 < P2 < ... < Pk.
Equally-Spaced X Values.

Power	Average n	k	Total N	Alpha	Beta	Proportions
0.53000*	14.00	3	42	0.02500	0.47000	0.20, 0.40, 0.60
0.52761*	14.00	3	42	0.02500	0.47239	0.30, 0.50, 0.70

* Values in this row are based on exact power calculations. Exact power was calculated for scenarios in which the largest group sample size is less than or equal to 50 and the product of all group samples sizes is less than or equal to 1000000.

The exact power values calculated by ***PASS*** match those calculated in Nam (1987) exactly if you round to two decimal places. Group sample sizes of *14* results in power of *0.53* for both scenarios. If you replicate the other scenarios in the table, you will find that the ***PASS*** results for exact power match Nam (1987) after rounding to two decimal places.

255-24 Cochran-Armitage Test for Trend in Proportions

Chapter 260

Inequality Tests for One ROC Curve

Introduction

Receiver operating characteristic (ROC) curves are used to assess the accuracy of a diagnostic test. The technique is used when you have a criterion variable which will be used to make a yes or no decision based on the value of this variable. The area under the ROC curve (AUC) is a popular summary index of an ROC curve.

This module computes power and sample size when a new diagnostic test is compared to an existing (gold) standard. Two approaches are available: the approach of Hanley and McNeil (1982) is used when the criterion variable is continuous and the approach of Obuchowski and McClish (1997) is used when the criterion variable is a discrete rating scale.

Technical Details

In the following, we suppose that we have two groups of patients, those with a condition of interest (the positive group) and those without it (the negative group). This classification may be known from extensive diagnosis or based on the value of another diagnostic test. The diagnostic test of interest is performed on each patient and the resulting test value is recorded. At each specified cutoff value of the criterion variable, the true positive rate (TPR) and the false positive rate (FPR) are calculated. A plot of the TPR versus the FPR allows you study the consequences of using various cutoff values. This plot is called the *ROC curve*.

It should be noted that TPR is similar to the statistical power of the diagnostic test at a particular cutoff value of the criterion variable. Similarly, FPR is an estimate of the probability that the diagnostic test results in a type I (alpha) error. Thus the ROC curve may be interpreted as a plot of the diagnostic test's power versus it's significance level at various possible criterion cutoff values.

Users of ROC curves have developed special names for TPR and FPR. They call TPR the *sensitivity* of the test and 1 - FPR the *specificity* of the test. Statisticians will be more familiar with using the word *power* instead of sensitivity and the phrase '1 - *alpha*' instead of specificity.

An ROC curve may be summarized by the area under it (AUC). This area has an additional interpretation. Suppose that a rater is asked to study two subjects, one that is actually disease

positive and one that is disease negative. The AUC is equal to the probability that the rater will give the disease positive subject a higher score than the disease negative subject. That is, the AUC is the probability that the rater will correctly order the two subjects as to which is more likely to have the disease.

Several methods of computing the AUC have been proposed. One method uses the trapezoidal rule to calculate the AUC directly. Another method, called the *binormal model*, computes the area by fitting two normal distributions to the data.

The Binormal Model

Let X denote the distribution of the criterion variable for negative (normal) patients and Y denote the distribution of the criterion variable for positive (diseased) patients. It is assumed that

$$X \sim \mathrm{N}\left(\mu_-, \sigma_-^2\right)$$

and

$$Y \sim \mathrm{N}\left(\mu_+, \sigma_+^2\right)$$

For a particular cutoff value of the criterion variable, c, the true positive rate is given by

$$\begin{aligned} TPR(c) &= \mathrm{P}(Y > c) \\ &= 1 - \Phi\left(\frac{c - \mu_+}{\sigma_+}\right) \\ &= \Phi\left(\frac{\mu_+ - c}{\sigma_+}\right) \end{aligned}$$

where $\Phi(z)$ is the cumulative normal distribution.

Similarly, the false positive rate is given by

$$\begin{aligned} FPR(c) &= \mathrm{P}(X > c) \\ &= 1 - \Phi\left(\frac{c - \mu_-}{\sigma_-}\right) \\ &= \Phi\left(\frac{\mu_- - c}{\sigma_-}\right) \end{aligned}$$

The ROC curve is thus the curve traced out by the functions

$$\left[FPR(c), TPR(c)\right] = \left[\Phi\left(\frac{\mu_- - c}{\sigma_-}\right), \Phi\left(\frac{\mu_+ - c}{\sigma_+}\right)\right]$$

The area under the ROC curve, AUC, is defined as

$$\begin{aligned}\theta &= \int_{-\infty}^{\infty} TPR(c)FPR'(c)\mathrm{d}c \\ &= \int_{-\infty}^{\infty} \Phi\left(\frac{\mu_+ - c}{\sigma_+}\right)\phi\left(\frac{\mu_- - c}{\sigma_-}\right)\left(-\frac{1}{\sigma_-}\right)\mathrm{d}c \\ &= \int_{-\infty}^{\infty} \Phi(A + Bv)\phi(v)\mathrm{d}v \\ &= \Phi\left(\frac{A}{\sqrt{1+B^2}}\right)\end{aligned}$$

where

$$c = \mu_- - v\sigma_-$$

$$A = \frac{|\mu_+ - \mu_-|}{\sigma_+}$$

$$B = \frac{\sigma_-}{\sigma_+}$$

Maximum likelihood estimates of *A* and *B* can be computed and used to compute AUC. The variances and covariance of these MLE's can be estimated from Fisher's information matrix.

Define $\Delta = \theta_0 - \theta_1$ to be the difference in the accuracies (AUC's) of two diagnostic tests. A hypothesis test of whether the two AUC's are different amounts to testing whether $\Delta = 0$. The test statistic for this test is

$$Z = \frac{\hat{\theta} - \theta_1}{\sqrt{\mathrm{var}_0\left(\hat{\theta}\right)}}$$

where $\mathrm{var}_0\left(\hat{\theta}\right)$ is the variance of $\hat{\theta}$ under the null hypothesis of equality. The above test statistic gives the following formulae for computing sample size or power

$$N_+ = \frac{\left(z_\alpha\sqrt{V(\theta_0)} + z_\beta\sqrt{V(\theta_1)}\right)^2}{(\theta_1 - \theta_0)^2}$$

$$\beta = \Phi\left(\frac{|\theta_1 - \theta_0|\sqrt{N_+} - z_\alpha\sqrt{V(\theta_0)}}{\sqrt{V(\theta_1)}}\right)$$

Rating Data

For a criterion variable yielding a discrete rating, Obuchowski (1998) recommends

$$V(\theta) = f^2\left(1+\frac{B^2}{R}+\frac{A^2}{2}\right)+g^2B^2\left(\frac{1+R}{2R}\right)$$

where

$$f = \frac{E_1}{\sqrt{2\pi\left(1+B^2\right)}}$$

$$g = -\frac{ABE_1}{\sqrt{2\pi\left(1+B^2\right)^3}}$$

$$E_1 = \exp\left(-\frac{A^2}{2+2B^2}\right)$$

The value of A can be found as

$$A = \Phi^{-1}(\theta)\sqrt{1+B^2}$$

For the most conservative results, Obuchowski (1998) recommends setting $B=1$, so that

$$A = \Phi^{-1}(\theta)\sqrt{2}$$

Continuous Data

For a criterion variable yielding a continuous result, Obuchowski (1998) suggests that the following formula of Hanley and McNeil (1983) is more appropriate

$$V(\theta) = \frac{\theta}{R(2-\theta)}+\frac{2\theta^2}{1+\theta}-\theta^2\left(\frac{1+R}{R}\right)$$

Procedure Options

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

Data Tab

Solve For

Find (Solve For)

This option specifies the parameter to be solved for from the other parameters. Under most situations, you will select either *Power and 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 *Power and Beta* when you want to calculate the power of an experiment that has already been run.

Error Rates

Power or Beta

This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected.

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

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

Alpha (Significance Level)

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected.

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

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

Sample Size

N+ (Size of Positive Group)

Specify the number of patients, that is, the sample size, in the positive (abnormal or diseased) group. Note that these values are ignored when you are solving for *N+*. You may enter a range of values such as *10 to 100 by 10*.

N- (Size of Negative Group)

Specify the number of patients, that is, the sample size, in the negative (normal) group. Enter *Use R* to base *N*- on the value of *N*+. You may enter a range of values such as *10 to 100 by 10.*

- **Use R**

  When *Use R* is entered here, *N*- is calculated using the formula

  $$N- = [R(N+)]$$

  where *R* is the Sample Allocation Ratio and the operator [*Y*] is the first integer greater than or equal to *Y*. For example, if you want *N*+ = *N*-, enter *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 *N*- is set to *Use R*.

When used, *N*- is calculated from *N*+ using the formula: *N*- = [*R*(*N*+)] where [*Y*] is the next integer greater than or equal to *Y*. Note that setting *R* = 1.0 forces *N*- = *N*+.

Effect Size – Area Under the Curve

AUC0 (Area Under Curve|H0)

Specify one or more values of the AUC for the diagnostic test. The range of values is from 0.5 (indicative of a test useless in diagnosis) to 1.0 (indicative of a test that is perfect in diagnosis).

Since the AUC may include a portion of the ROC curve that is not of interest because the FPR values are unrealistic, you may be interested in only a portion of the area. In this case, you can specify a range of FPR values for which the area is to be calculated. Unfortunately, the definition of the area becomes more difficult. When analyzing the whole ROC curve, the area is known to be between 0.50 and 1.0. Following the suggestion of Obuchowski and McClish (1997), the following transformation is applied so that the values of AUC remain between 0.5 and 1.0.

$$AUC' = \frac{1}{2}\left(1 + \frac{AUC - \min}{\max - \min}\right)$$

where

$$\max = FPR2 - FPR1$$

$$\min = \frac{\max}{2}\left(FPR2 + FPR1\right)$$

Thus, when a partial range is entered for FPR1 and FPR2, the values entered here are assumed to be AUC' and are translated to AUC using the above formulas.

AUC1 (Area Under Curve|H1)

Specify one or more values of AUC under the alternative hypothesis. The range of values is from 0.5 (indicative of a test useless in diagnosis) to 1.0 (indicative of a test that is perfect in diagnosis). Note that, as discussed above, this is the value of AUC' when a partial area is being analyzed.

Effect Size – False Positive Rate Limits

Lower FPR

This option specifies the lower (left) limit of the false positive rate (FPR) for which the area is to be computed. If the area under the whole ROC curve is wanted, set this value to 0.0. If the partial area is wanted, set this value to the desired left limit.

Note that the range of possible values is 0.0 <= Lower FPR < Upper FPR <= 1.0

Upper FPR

This option specifies the upper (right) limit of the false positive rate (FPR). If the area under the whole ROC curve is wanted, set this value to 1.0. If the partial area is wanted, set this value to the desired right limit.

Note that the range of possible values is 0.0 <= Lower FPR < Upper FPR <= 1.0

Effect Size – Type of Data

Type of Data

Specify the type of data that will be collected from the tests. The formulas for the variance are determined by this option. Possible types are:

- **Continuous**

  The test results are from a continuum of possible values. The Hanley and McNeil (1983) variance formulas are used. Note that this option does not allow a partial range of FPR values to be analyzed.

- **Discrete**

  The test results are from a small set of rating values such as 1, 2, 3, 4, 5. The Obuchowski & McClish (1997) variance formulas are used.

B (SD Ratio = SD-/SD+)

B is the ratio of the standard deviation of the negative group to the positive group (SD-/SD+) for the diagnostic test. That is, assuming the binormal model

$$B = \frac{\sigma_-}{\sigma_+}$$

Note that this parameter is ignored for continuous data.

Although *B* can be any positive number, typical values are between 0.3 and 3.0. Obuchowski suggests that if the value of *B* is not known, a value of 1.0 is used since this will result in a conservative (extra large) sample size. She reports that in her experience, typical values are much less than 1.0, often near 0.3.

Test

Alternative Hypothesis

Specify whether the test is *one-sided* or *two-sided*. When a two-sided test is selected, the value of alpha is divided by two.

Note that most researchers assume that, unless stated otherwise, all statistical tests are two-sided. If you use a one-sided test, you should clearly state and justify this in all reports.

Example 1 – Calculating Power

An investigator wants to study the accuracy of a diagnostic test which yields measurements on a rating scale from 1 to 5. Historically, such tests have had an AUC of 0.80. The investigator wants to investigate three alternative AUC values: 0.825, 0.850, and 0.900. A two-sided test is planned with a significance level of 0.05. Since no other information is available, B is set to 1.0. The investigator would like to achieve a power of 90% in the study. Patients without the disease under study are about twice as frequent as patients with the disease. The investigator wants to see results for a sample size of up to 6000 patients.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for One ROC Curve** procedure window by clicking on **ROC Curves**, and then clicking on **Tests for One ROC Curve**. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power	*Ignored since this is the Find Setting*
Alpha	**0.05**
N+ (Size of Positive Group)	**20 50 100 250 500 1000 2000**
N- (Size of Negative Group)	**Use R**
R (Sample Allocation Ratio)	**2**
AUC0 (Area Under Curve|H0)	**0.80**
AUC1 (Area Under Curve|H1)	**0.825 0.85 0.90**
Lower FPR	**0.00**
Upper FPR	**1.00**
Type of Data	**Discrete (Ratings)**
B (SD Ratio = SD-/SD+)	**1.0**
Alternative Hypothesis	**Two-Sided Test**

Annotated Output

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

Numeric Report

Numeric Results for Testing AUC0 = AUC1 with Discrete (Rating) Data
Test Type = Two-Sided. FPR1 = 0.0. FPR2 = 1.0. B = 1.000. Allocation Ratio = 2.000.

Power	N+	N-	AUC0'	AUC1'	Diff'	AUC0	AUC1	Diff	Alpha	Beta
0.0481	20	40	0.8000	0.8250	0.0250	0.8000	0.8250	0.0250	0.0500	0.9519
0.0739	50	100	0.8000	0.8250	0.0250	0.8000	0.8250	0.0250	0.0500	0.9261
0.1146	100	200	0.8000	0.8250	0.0250	0.8000	0.8250	0.0250	0.0500	0.8854
0.2365	250	500	0.8000	0.8250	0.0250	0.8000	0.8250	0.0250	0.0500	0.7635
0.4321	500	1000	0.8000	0.8250	0.0250	0.8000	0.8250	0.0250	0.0500	0.5679
0.7264	1000	2000	0.8000	0.8250	0.0250	0.8000	0.8250	0.0250	0.0500	0.2736
0.9550	2000	4000	0.8000	0.8250	0.0250	0.8000	0.8250	0.0250	0.0500	0.0450
0.0870	20	40	0.8000	0.8500	0.0500	0.8000	0.8500	0.0500	0.0500	0.9130
0.1834	50	100	0.8000	0.8500	0.0500	0.8000	0.8500	0.0500	0.0500	0.8166
0.3491	100	200	0.8000	0.8500	0.0500	0.8000	0.8500	0.0500	0.0500	0.6509
0.7369	250	500	0.8000	0.8500	0.0500	0.8000	0.8500	0.0500	0.0500	0.2631
0.9629	500	1000	0.8000	0.8500	0.0500	0.8000	0.8500	0.0500	0.0500	0.0371
0.9997	1000	2000	0.8000	0.8500	0.0500	0.8000	0.8500	0.0500	0.0500	0.0003
1.0000	2000	4000	0.8000	0.8500	0.0500	0.8000	0.8500	0.0500	0.0500	0.0000
0.2489	20	40	0.8000	0.9000	0.1000	0.8000	0.9000	0.1000	0.0500	0.7511
0.6563	50	100	0.8000	0.9000	0.1000	0.8000	0.9000	0.1000	0.0500	0.3437
0.9474	100	200	0.8000	0.9000	0.1000	0.8000	0.9000	0.1000	0.0500	0.0526
1.0000	250	500	0.8000	0.9000	0.1000	0.8000	0.9000	0.1000	0.0500	0.0000
1.0000	500	1000	0.8000	0.9000	0.1000	0.8000	0.9000	0.1000	0.0500	0.0000
1.0000	1000	2000	0.8000	0.9000	0.1000	0.8000	0.9000	0.1000	0.0500	0.0000
1.0000	2000	4000	0.8000	0.9000	0.1000	0.8000	0.9000	0.1000	0.0500	0.0000

Report Definitions
Power is the probability of rejecting a false null hypothesis.
N+ is the sample size from the positive (diseased) population.
N- is the sample size from the negative (non-diseased) population.
Alloc Ratio is the Sample Allocation Ratio (R = N- / N+).
AUC0' is the adjusted area under the ROC curve under the null hypothesis.
AUC1' is the adjusted area under the ROC curve under the alternative hypothesis.
Diff' is AUC1' - AUC0'. This is the adjusted difference to be detected.
AUC0 is the actual area under the ROC curve under the null hypothesis.
AUC1 is the actual area under the ROC curve under the alternative hypothesis.
Diff is AUC1 - AUC0. This is the difference to be detected.
Alpha is the probability of rejecting a true null hypothesis.
Beta is the probability of accepting a false null hypothesis.
FPR1, FPR2 are the lower and upper bounds on the false positive rates.
B is the ratio of the standard deviations of the negative and positive groups.

Summary Statements
A sample of 20 from the positive group and 40 from the negative group achieve 5% power to detect a difference of 0.0250 between the area under the ROC curve (AUC) under the null hypothesis of 0.8000 and an AUC under the alternative hypothesis of 0.8250 using a two-sided z-test at a significance level of 0.0500. The data are discrete (rating scale) responses. The AUC is computed between false positive rates of 0.000 and 1.000. The ratio of the standard deviation of the responses in the negative group to the standard deviation of the responses in the positive group is 1.000.

This report shows the power for each of the sample sizes. Most of the definitions are standard. However, a special explanation must be given for AUC and AUC'.

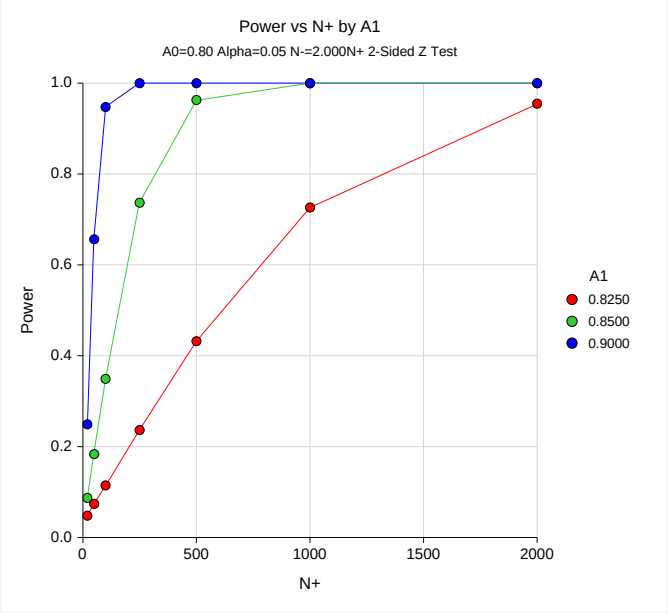
AUC'

This is the adjusted area under the curve. A rescaling, discussed earlier, has been applied so that the minimum area is 0.5 and the maximum area is 1.0.

AUC

This is the actual area under the curve. This value will equal the adjusted area when the FPR range is set from 0.0 to 1.0. Otherwise, these values will be different.

Plot Section



This plot shows the power versus the sample size for the three values of AUC1.

Example 2 – Calculating Sample Size

Continuing on with Example1, the investigator wants to know the exact sample size needed for each of the three values of AUC2. The investigator wants to look at the Numeric Report. The panel from Example1 is modified as follows.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for One ROC Curve** procedure window by clicking on **ROC Curves**, and then clicking on **Tests for One ROC Curve**. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**N+**
Reports Tab	
Show Definitions	**Unchecked**
Show Plots	**Unchecked**
Show Summary Statements	**Unchecked**

Output

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

Numeric Report

Numeric Results for Testing AUC0 = AUC1 with Discrete (Rating) Data
Test Type = Two-Sided. FPR1 = 0.0. FPR2 = 1.0. B = 1.000. Allocation Ratio = 2.000.

Power	N+	N-	AUC0'	AUC1'	Diff'	AUC0	AUC1	Diff	Alpha	Beta
0.9001	1582	3164	0.8000	0.8250	0.0250	0.8000	0.8250	0.0250	0.0500	0.0999
0.9007	381	762	0.8000	0.8500	0.0500	0.8000	0.8500	0.0500	0.0500	0.0993
0.9024	85	170	0.8000	0.9000	0.1000	0.8000	0.9000	0.1000	0.0500	0.0976

This report shows the sample size needed to achieve 90% power for each value of AUC1.

Example 3 – Partial Area under Curve

Continuing on with Example 2, the investigator knows that FPR values between 0.0 and 0.20 are the only values of interest. Hence, he wants to investigate the sample size needed when the FPR range is confined to this range.

The panel from Example 2 is modified as follows.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for One ROC Curve** procedure window by clicking on **ROC Curves**, and then clicking on **Tests for One ROC Curve**. You may then make the appropriate entries as listed below, or open **Example 3** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Upper FPR	**0.20**

Output

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

Numeric Results

Numeric Results for Testing AUC0 = AUC1 with Discrete (Rating) Data
Test Type = Two-Sided. FPR1 = 0.0. FPR2 = 0.200. B = 1.000. Allocation Ratio = 2.000.

Power	N+	N-	AUC0'	AUC1'	Diff'	AUC0	AUC1	Diff	Alpha	Beta
0.9001	2663	5326	0.8000	0.8250	0.0250	0.1280	0.1370	0.0090	0.0500	0.0999
0.9002	645	1290	0.8000	0.8500	0.0500	0.1280	0.1460	0.0180	0.0500	0.0998
0.9013	144	288	0.8000	0.9000	0.1000	0.1280	0.1640	0.0360	0.0500	0.0987

Note that the necessary sample size has almost doubled.

Example 4 – Validation using Obuchowski

The formulas used in this module were given in Obuchowski and McClish (1997). On page 1538, they provide an example which will be duplicated here. The study investigated the accuracy of MRI for detecting abnormalities in patients with symptomatic knees. In order to do this, they wanted to know the sample size that would be needed to construct a 95% confidence interval so that the length of the confidence interval is no more than 0.10.

The measure of diagnostic accuracy is the AUC from an FPR of 0.0 to an FPR of 1.0. The allocation ratio is 1.5. B = 1.0. The value of A is found to be 1.2. This translates to an AUC0 of 0.7995. The value of AUC1 = AUC0 + 0.10 / 2, where 0.10 is the maximum length of the confidence interval. A two-tailed confidence interval is envisioned in which alpha is 0.05. In order to find the sample size of a confidence interval, the power is set to 50%. In their article, they found N+ = 161 and N- = 242.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for One ROC Curve** procedure window by clicking on **ROC Curves**, and then clicking on **Tests for One ROC Curve**. You may then make the appropriate entries as listed below, or open **Example 4** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**N+**
Power	**0.50**
Alpha	**0.05**
N+ (Size of Positive Group)	*Ignored since this is the Find Setting*
N- (Size of Negative Group)	**Use R**
R (Sample Allocation Ratio)	**1.5**
AUC0 (Area Under Curve|H0)	**0.7995**
AUC1 (Area Under Curve|H1)	**0.8495**
Lower FPR	**0.00**
Upper FPR	**1.00**
Type of Data	**Discrete (Ratings)**
B (SD Ratio = SD-/SD+)	**1**
Alternative Hypothesis	**Two-Sided Test**

Output

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

Numeric Results

Numeric Results for Testing AUC0 = AUC1 with Discrete (Rating) Data
Test Type = Two-Sided. FPR1 = 0.0. FPR2 = 0.200. B1 = 1.000. B2 = 1.000. Allocation Ratio = 2.000.

Power	N+	N-	AUC1'	AUC2'	Diff'	AUC1	AUC2	Diff	Alpha	Beta
0.5026	162	243	0.7995	0.8495	0.0500	0.7995	0.8495	0.0500	0.0500	0.4974

Note that the sample sizes of 162 and 243 are within one of the results of Obuchowski. The difference occurs because their values of 161 and 242 produce a power that is slightly less than 0.5, so PASS increased the sample size slightly.

Chapter 265

Inequality Tests for Two ROC Curves

Introduction

Receiver operating characteristic (ROC) curves are used to summarize the accuracy of diagnostic tests. The technique is used when a criterion variable is available which is used to make a yes or no decision. The area under the ROC curve (AUC) is a popular summary index of an ROC curve.

This module computes power and sample size for comparing the AUC's of two diagnostic tests obtained from the same patients. The methodology of Obuchowski and McClish (1997) is used when the criterion variable yields a discrete value. The methodology of Hanley and McNeil (1983) is used when the criterion variable yields a continuous value.

Technical Details

In the following, we suppose that we have two groups of patients, those with a condition of interest (the disease) and those without it. A patient's classification may be known from extensive diagnosis or based on the value of another diagnostic test. The diagnostic tests of interest are performed on each patient and the resulting test values are recorded. At each specified cutoff value of the criterion variable, the true positive rate (TPR) and the false positive rate (FPR) are calculated. An ROC curve is generating by plotting TPR versus FPR. The plot allows the consequences of using various cutoff values to be evaluated. The area under the ROC curve, either for the whole or partial range, is often used as a summary measure of the accuracy of the test.

It should be noted that TPR is similar to the statistical power of the diagnostic test at a particular cutoff value of the criterion variable. Similarly, FPR is an estimate of the probability that the diagnostic test results in a type I (alpha) error. Thus the ROC curve may be interpreted as a plot of the diagnostic test's power versus it's significance level at various possible criterion cutoff values.

Users of ROC curves have developed special names for TPR and FPR. They call TPR the *sensitivity* of the test and 1 - FPR the *specificity* of the test. Statisticians will be more familiar with using the word *power* instead of sensitivity and the phrase '1 - *alpha*' instead of specificity.

An ROC curve may be summarized by the area under it (AUC). This area has an additional interpretation. Suppose that a rater is asked to study two subjects, one that is actually disease positive and one that is disease negative. The AUC is equal to the probability that the rater will give the disease positive subject a higher score than the disease negative subject. That is, the AUC is the probability that the rater will correctly order the two subjects as to which is more likely to have the disease.

Several methods of computing the AUC have been proposed. One method uses the trapezoidal rule to calculate the AUC directly. Another method, called the *binormal model*, computes the area by fitting two normal distributions to the data.

The Binormal Model

Let *X* denote the distribution of the criterion variable for normal (non-diseased) patients and *Y* denote the distribution of the criterion variable for abnormal (diseased) patients. It is assumed that

$$X \sim \mathrm{N}\left(\mu_-, \sigma_-^2\right)$$

and

$$Y \sim \mathrm{N}\left(\mu_+, \sigma_+^2\right)$$

The partial area under the ROC curve, *AUC*, is defined as

$$\theta_i = \int_{c_1}^{c_2} \Phi\left(A_i + B_i v\right)\phi\left(v\right)\mathrm{d}v$$

where $\Phi\left(z\right)$ is the cumulative normal distribution, $c_j = \Phi^{-1}\left(FPR_j\right)$, and

$$A_i = \frac{\mu_{i+} - \mu_{i-}}{\sigma_{i+}}$$

$$B_i = \frac{\sigma_{i-}}{\sigma_{i+}}$$

Note that for the full range area under the curve, $c_1 = -\infty$ and $c_2 = \infty$.

Maximum likelihood estimates of *A* and *B* can be computed. The variances and covariance of these MLE's can be estimated from Fisher's information matrix.

Define $\Delta = \theta_1 - \theta_2$ to be the difference in the accuracies (AUC's) of the two tests. A test of whether the two AUC's are different amounts to testing whether $\Delta = 0$. The test statistic for this test is

$$Z = \frac{\hat{\Delta} - 0}{\sqrt{\mathrm{var}_0\left(\hat{\Delta}\right)}}$$

where $\text{var}_0\left(\hat{\Delta}\right)$ is the variance of $\hat{\Delta}$ under the null hypothesis of equality. The above test statistic results in the following formula for computing sample size

$$N_+ = \frac{\left(z_\alpha\sqrt{V_0\left(\hat{\Delta}\right)} + z_\beta\sqrt{V_{Alt}\left(\hat{\Delta}\right)}\right)^2}{\Delta^2}$$

Rating Data

When the criterion values are discrete rating values, Obuchowski and McClish (1997) showed that the variances could be calculated using

$$V_0\left(\hat{\Delta}\right) = V\left(\hat{\theta}_1\right) + V\left(\hat{\theta}_1\right) - 2C\left(\hat{\theta}_1, \hat{\theta}_1\right)$$

$$V_{Alt}\left(\hat{\Delta}\right) = V\left(\hat{\theta}_1\right) + V\left(\hat{\theta}_2\right) - 2C\left(\hat{\theta}_1, \hat{\theta}_2\right)$$

where

$$V\left(\hat{\theta}_i\right) = f_i^2\left(1 + \frac{B_i^2}{R} + \frac{A_i^2}{2}\right) + g_i^2\left(B_i^2\left(\frac{1+R}{2R}\right)\right)$$

$$C\left(\hat{\theta}_1, \hat{\theta}_2\right) = f_1 f_2\left(r_+ + r_- \frac{B_1 B_2}{R} + r_+^2 \frac{A_1 A_2}{2}\right) + \frac{g_1 g_2 B_1 B_2\left(r_-^2 + R r_+^2\right)}{2R} + \frac{f_1 g_2 A_1 B_2 r_+^2}{2} + \frac{f_2 g_1 A_2 B_1 r_+^2}{2}$$

$$f_i = \frac{E_{1i}E_{3i}}{\sqrt{2\pi E_{2i}}}$$

$$g_i = \frac{E_{1i}E_{4i}}{\sqrt{2\pi E_{2i}}} - \frac{A_i B_i E_{1i} E_{3i}}{\sqrt{2\pi E_{2i}^3}}$$

$$E_{1i} = \exp\left(-\frac{A_i^2}{2 + 2B_i^2}\right)$$

$$E_{2i} = 1 + B_i^2$$

$$E_{3i} = \Phi\left(c_2\right) - \Phi\left(c_1\right)$$

$$E_{4i} = \exp\left(-\frac{c_1^2}{2}\right) - \exp\left(-\frac{c_2^2}{2}\right)$$

$$c_j = \frac{\Phi^{-1}\left(FPR_j\right) + \frac{A_j B_j}{1 + B_j^2}}{\sqrt{1 + B_j^2}}$$

$$R = \frac{N_-}{N_+}$$

$$A_i = B_i \Phi^{-1}\left(TNR_i\right) - \Phi^{-1}\left(FPR_i\right)$$

r_- and r_+ are the correlations between the results of the two diagnostics tests for normal and abnormal patients, respectively. For the most conservative results, set $B_i = 1$.

Continuous Data

When the criterion values are continuous, Obuchowski (1998) suggests that the following formulas of Hanley and McNeil (1983) are more appropriate. Note that these formulas cannot be used for evaluating the AUC for a partial range.

$$V\left(\hat{\Delta}\right) = V\left(\hat{\theta}_1\right) + V\left(\hat{\theta}_2\right) - 2C\left(\hat{\theta}_1, \hat{\theta}_2\right)$$

where

$$V\left(\hat{\theta}_i\right) = \frac{\theta_i}{R\left(2 - \theta_i\right)} + \frac{2\theta_i^2}{1 + \theta_i} - \theta_i^2\left(\frac{1 + R}{R}\right)$$

$$C\left(\hat{\theta}_1, \hat{\theta}_2\right) = 2r\sqrt{V\left(\theta_1\right)V\left(\theta_2\right)}$$

and *r* is derived from a special table provided by Hanley and McNeil (1983).

Procedure Options

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

Data Tab

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

Solve For

Find (Solve For)

This option specifies the parameter to be solved for from the other parameters. Under most situations, you will select either *Power and 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 *Power and Beta* when you want to calculate the power of an experiment that has already been run.

Error Rates

Power or Beta

This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected.

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

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

Alpha (Significance Level)

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected.

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

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

Sample Size

N+ (Size of Positive Group)

Specify the number of patients, that is, the sample size, in the positive (abnormal or diseased) group. Note that these values are ignored when you are solving for *N*+. You may enter a range of values such as *10 to 100 by 10*.

N- (Size of Negative Group)

Specify the number of patients, that is, the sample size, in the negative (normal) group. Enter *Use R* to base *N*- on the value of *N*+. You may enter a range of values such as *10 to 100 by 10.*

- **Use R**

  When *Use R* is entered here, *N*- is calculated using the formula

  $$N- = [R(N+)]$$

  where *R* is the Sample Allocation Ratio and the operator [*Y*] is the first integer greater than or equal to *Y*. For example, if you want *N*+ = *N*-, enter *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 *N*- is set to *Use R*.

When used, *N*- is calculated from *N*+ using the formula: *N*- = [*R*(*N*+)] where [*Y*] is the next integer greater than or equal to *Y*. Note that setting *R* = 1.0 forces *N*- = *N*+.

Test

Alternative Hypothesis

Specify whether the test is *one-sided* or *two-sided*. When a two-sided test is selected, the value of alpha is divided by two.

Note that most researchers assume that, unless stated otherwise, all statistical tests are two-sided. If you use a one-sided test, you should clearly state and justify this in all reports.

Effect Size – Area Under the Curve

AUC1 (Area Under Curve 1)

Specify one or more values of the AUC for diagnostic test 1. The range of values is from 0.5 (indicative of a test useless in diagnosis) to 1.0 (indicative of a test that is perfect in diagnosis).

Since the AUC may include a portion of the ROC curve that is not of interest because the FPR values are unrealistic, you may be interested in only a portion of the area. In this case, you can specify a range of FPR values for which the area is to be calculated. Unfortunately, the definition of the area becomes more difficult. When analyzing the whole ROC curve, the area is known to be between 0.50 and 1.0. Following the suggestion of Obuchowski and McClish (1997), the following transformation is applied so that the values of AUC remain between 0.5 and 1.0.

$$AUC' = \frac{1}{2}\left(1 + \frac{AUC - \min}{\max - \min}\right)$$

where

$$\max = FPR2 - FPR1$$

$$\min = \frac{\max}{2}\left(FPR2 + FPR1\right)$$

Thus, when a partial range is entered for FPR1 and FPR2, the values entered here are assumed to be AUC' and are translated to AUC using the above formulas.

AUC2 (Area Under Curve 2)

Specify one or more values of the AUC for diagnostic test 2. The range of values is from 0.5 (indicative of a test useless in diagnosis) to 1.0 (indicative of a test that is perfect in diagnosis). Note that, as discussed above, this is the value of AUC' when a partial area is being analyzed.

Effect Size – False Positive Rate Limits

Lower FPR

This option specifies the lower (left) limit of the false positive rate (FPR) for which the area is to be computed. If the area under the whole ROC curve is wanted, set this value to 0.0. If the partial area is wanted, set this value to the desired left limit.

Note that the range of possible values is 0.0 <= Lower FPR < Upper FPR <= 1.0

Upper FPR

This option specifies the upper (right) limit of the false positive rate (FPR). If the area under the whole ROC curve is wanted, set this value to 1.0. If the partial area is wanted, set this value to the desired right limit.

Note that the range of possible values is 0.0 <= Lower FPR < Upper FPR <= 1.0

Effect Size – Correlations

Correlation+

This is the correlation between the two diagnostic-test scores for the positive group. Although correlations can range between -1 and 1, typical values are from 0.3 to 0.6.

Note that if you want to analyze a design in which a separate set of patients receive each diagnostic test, this may be done by setting this correlation value to 0.

Correlation-

This is the correlation between the two diagnostic-test scores for the negative group. Although correlations can range between -1 and 1, typical values are from 0.3 to 0.6.

Note that if you want to analyze a design in which a separate set of patients receive each diagnostic test, this may be done by setting this correlation value to 0.

Effect Size – Type of Data

Type of Data

Specify the type of data that will be collected from the tests. The formulas for the variance are determined by this option. Possible types are:

- **Continuous**

  The test results are from a continuum of possible values. The Hanley and McNeil (1983) variance formulas are used. Note that this option does not allow a partial range of FPR values to be analyzed.

- **Discrete (Ratings)**

  The test results are from a small set of rating values such as 1, 2, 3, 4, 5. The Obuchowski & McClish (1997) variance formulas are used.

B1 (SD Ratio)

*B*1 is the ratio of the standard deviation of the negative group to the positive group (SD-/SD+) for diagnostic test 1. That is, assuming the binormal model

$$B1 = \frac{\sigma_{1-}}{\sigma_{1+}}$$

Note that this parameter is ignored for continuous data.

Although *B*1 can be any positive number, typical values are between 0.3 and 3.0. Obuchowski suggests that if the value of *B*1 is not known, a value of 1.0 is used since this will result in a conservative (extra large) sample size. She reports that in her experience, typical values are much less than 1.0, often near 0.3.

B2 (SD Ratio)

*B*2 is the ratio of the standard deviation of the negative group to the positive group (SD-/SP+) for diagnostic test 2. That is, assuming the binormal model

$$B2 = \frac{\sigma_{2-}}{\sigma_{2+}}$$

Note that this parameter is ignored for continuous data.

Although *B*2 can be any positive number, typical values are between 0.3 and 3.0. Obuchowski suggests that if the value of *B*2 is not known, a value of 1.0 is used since this will result in a conservative (extra large) sample size. She reports that in her experience, typical values are much less than 1.0, often near 0.3.

Example 1 – Calculating Power

An investigator wants to compare the accuracy of two diagnostic tests which yield measurements on a rating scale from 1 to 5. Historically, such tests have had an AUC of 0.80. The investigator wants to investigate three alternative AUC values: 0.825, 0.850, and 0.900. A two-sided test is planned with a significance level of 0.05. Historically, both the positive and negative correlations between the responses on two such tests have been close to 0.60. Since no other information is available, B1 and B2 are both set to 1.0. The investigator would like to achieve a power of 90% in the study. Patients without the disease under study are about twice as frequent as patients with the disease. The investigator wants to see results for a sample size of up to 6000 patients.

Setup

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

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power	*Ignored since this is the Find Setting*
Alpha	**0.05**
N+ (Size of Positive Group)	**20 50 100 250 500 1000 2000**
N- (Size of Negative Group)	**Use R**
R (Sample Allocation Ratio)	**2**
AUC1 (Area Under Curve 1)	**0.80**
AUC2 (Area Under Curve 2)	**0.825 0.85 0.9**
Lower FPR	**0.00**
Upper FPR	**1.00**
Correlation+	**0.6**
Correlation-	**0.6**
Type of Data	**Discrete (Ratings)**

Data Tab (continued)

B1 (SD Ratio) **1**
B2 (SD Ratio) **1**
Alternative Hypothesis ..... **Two-Sided Test**

Annotated Output

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

Numeric Report

Numeric Results for Testing AUC1 = AUC2 with Discrete (Rating) Data
Test Type = Two-Sided. FPR1 = 0.0. FPR2 = 1.0. B1 = 1.000. B2 = 1.000. Allocation Ratio = 2.000.

Power	N+	N-	AUC1'	AUC2'	Diff'	AUC1	AUC2	Diff	Alpha	Beta
0.0501	20	40	0.8000	0.8250	0.0250	0.8000	0.8250	0.0250	0.0500	0.9499
0.0733	50	100	0.8000	0.8250	0.0250	0.8000	0.8250	0.0250	0.0500	0.9267
0.1084	100	200	0.8000	0.8250	0.0250	0.8000	0.8250	0.0250	0.0500	0.8916
0.2104	250	500	0.8000	0.8250	0.0250	0.8000	0.8250	0.0250	0.0500	0.7896
0.3744	500	1000	0.8000	0.8250	0.0250	0.8000	0.8250	0.0250	0.0500	0.6256
0.6426	1000	2000	0.8000	0.8250	0.0250	0.8000	0.8250	0.0250	0.0500	0.3574
0.9090	2000	4000	0.8000	0.8250	0.0250	0.8000	0.8250	0.0250	0.0500	0.0910
0.0920	20	40	0.8000	0.8500	0.0500	0.8000	0.8500	0.0500	0.0500	0.9080
0.1737	50	100	0.8000	0.8500	0.0500	0.8000	0.8500	0.0500	0.0500	0.8263
0.3083	100	200	0.8000	0.8500	0.0500	0.8000	0.8500	0.0500	0.0500	0.6917
0.6442	250	500	0.8000	0.8500	0.0500	0.8000	0.8500	0.0500	0.0500	0.3558
0.9116	500	1000	0.8000	0.8500	0.0500	0.8000	0.8500	0.0500	0.0500	0.0884
0.9969	1000	2000	0.8000	0.8500	0.0500	0.8000	0.8500	0.0500	0.0500	0.0031
1.0000	2000	4000	0.8000	0.8500	0.0500	0.8000	0.8500	0.0500	0.0500	0.0000
0.2470	20	40	0.8000	0.9000	0.1000	0.8000	0.9000	0.1000	0.0500	0.7530
0.5494	50	100	0.8000	0.9000	0.1000	0.8000	0.9000	0.1000	0.0500	0.4506
0.8496	100	200	0.8000	0.9000	0.1000	0.8000	0.9000	0.1000	0.0500	0.1504
0.9978	250	500	0.8000	0.9000	0.1000	0.8000	0.9000	0.1000	0.0500	0.0022
1.0000	500	1000	0.8000	0.9000	0.1000	0.8000	0.9000	0.1000	0.0500	0.0000
1.0000	1000	2000	0.8000	0.9000	0.1000	0.8000	0.9000	0.1000	0.0500	0.0000
1.0000	2000	4000	0.8000	0.9000	0.1000	0.8000	0.9000	0.1000	0.0500	0.0000

Report Definitions
Power is the probability of rejecting a false null hypothesis.
N+ is the sample size from the positive (diseased) population.
N- is the sample size from the negative (non-diseased) population.
Alloc Ratio is the Sample Allocation Ratio (R = N- / N+).
AUC1' is the adjusted area under the ROC curve for diagnostic test 1.
AUC2' is the adjusted area under the ROC curve for diagnostic test 2.
Diff' is AUC2' - AUC1'. This is the adjusted difference to be detected.
AUC1 is the actual area under the ROC curve for diagnostic test 1.
AUC2 is the actual area under the ROC curve for diagnostic test 2.
Diff is AUC2 - AUC1. This is the difference to be detected.
Alpha is the probability of rejecting a true null hypothesis.
Beta is the probability of accepting a false null hypothesis.
FPR1, FPR2 are the lower and upper bounds on the false positive rates.
B1 and B2 are the ratios of the standard deviations of the negative and positive groups for each test.

Summary Statements
A sample of 20 from the positive group and 40 from the negative group achieve 5% power to detect a difference of 0.0250 between a diagnostic test with an area under the ROC curve (AUC) of 0.8000 and another diagnostic test with an AUC of 0.8250 using a two-sided z-test at a significance level of 0.0500. The data are discrete (rating scale) responses. The AUC is computed between false positive rates of 0.000 and 1.000. The ratio of the standard deviation of the responses in the negative group to the standard deviation of the responses in the positive group for diagnostic test 1 is 1.000 and for diagnostic test 2 is 1.000. The correlation between the two diagnostic tests is assumed to be 0.600 for the positive group and 0.600 for the negative group.

This report shows the power for each of the sample sizes. Most of the definitions are standard. However, a special explanation must be given for AUC and AUC'.

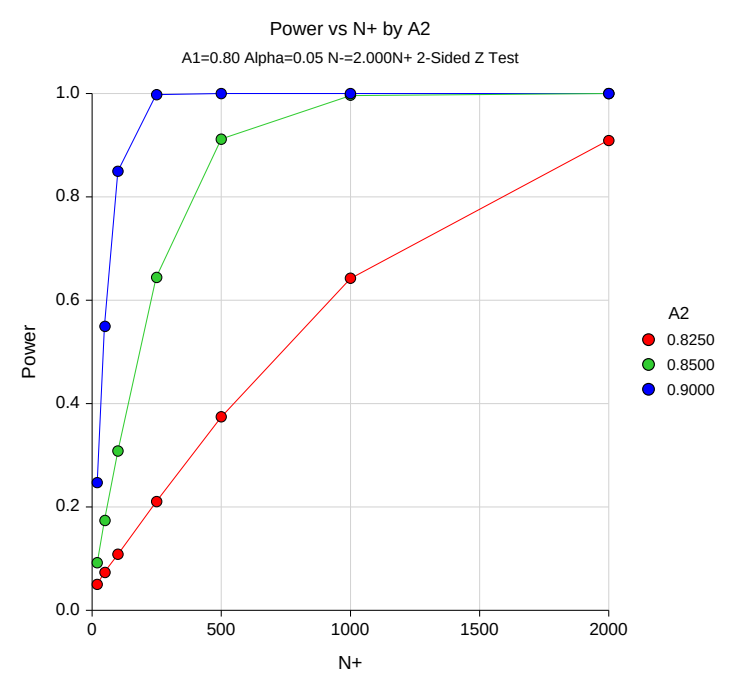
AUC'

This is the adjusted area under the curve. A rescaling, discussed earlier, has been applied so that the minimum area is 0.5 and the maximum area is 1.0.

AUC

This is the actual area under the curve. This value will equal the adjusted area when the FPR range is set from 0.0 to 1.0. Otherwise, these values will be different.

Plots Section



This plot shows the power versus the sample size for the three values of AUC1.

Example 2 – Calculating Sample Size

Continuing Example 1, the investigator wants to know the exact sample size needed for each of the three values of AUC2. The investigator wants to look at the Numeric Report. The panel from Example1 is modified as follows.

Setup

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

Option	Value
Data Tab	
Find (Solve For)	**N+**
Reports Tab	
Show Definitions	**Unchecked**
Show Plots	**Unchecked**
Show Summary Statements	**Unchecked**

Output

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

Numeric Results

Numeric Results for Testing AUC1 = AUC2 with Discrete (Rating) Data

Test Type = Two-Sided. FPR1 = 0.0. FPR2 = 1.0. B1 = 1.000. B2 = 1.000. Allocation Ratio = 2.000.

Power	N+	N-	AUC1'	AUC2'	Diff'	AUC1	AUC2	Diff	Alpha	Beta
0.9001	1937	3874	0.8000	0.8250	0.0250	0.8000	0.8250	0.0250	0.0500	0.0999
0.9002	480	960	0.8000	0.8500	0.0500	0.8000	0.8500	0.0500	0.0500	0.0998
0.9012	117	234	0.8000	0.9000	0.1000	0.8000	0.9000	0.1000	0.0500	0.0988

This report shows the sample size needed to achieve 90% power for each value of AUC2.

Example 3 – Partial Area Under Curve

Continuing Example 2, the investigator knows that FPR values between 0.0 and 0.20 are the only values of interest. Hence, he wants to investigate the sample size needed when the FPR range is confined to this range.

The panel from Example 2 is modified as follows.

Setup

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

Option	Value
Data Tab	
Upper FPR	**0.20**

Output

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

Numeric Results

Numeric Results for Testing AUC1 = AUC2 with Discrete (Rating) Data
Test Type = Two-Sided. FPR1 = 0.0. FPR2 = 0.200. B1 = 1.000. B2 = 1.000. Allocation Ratio = 2.000.

Power	N+	N-	AUC1'	AUC2'	Diff'	AUC1	AUC2	Diff	Alpha	Beta
0.9000	4095	8190	0.8000	0.8250	0.0250	0.1280	0.1370	0.0090	0.0500	0.1000
0.9002	1012	2024	0.8000	0.8500	0.0500	0.1280	0.1460	0.0180	0.0500	0.0998
0.9001	242	484	0.8000	0.9000	0.1000	0.1280	0.1640	0.0360	0.0500	0.0999

Note that the necessary sample size has more than doubled.

Example 4 – Validation using Obuchowski

The formulas used in this module were given in Obuchowski and McClish (1997). On pages 1538 - 1540, they provide an example which will be duplicated here. The study compared an automated classification system with an expert mammographer in their ability to find malignant breast lesions. The measure of diagnostic accuracy is the AUC from an FPR of 0.0 to an FPR of 0.2. The allocation ratio is 2. B1 = B2 = 1.0. Correlation+ = Correlation- = 0.6. The values of A1 and A2 are found to be 2.6 and 1.9. These translate to adjusted AUC's of 0.922222 and 0.819444. A two-tailed test is envisioned in which alpha is 0.05. A power of 80% is desired. In their article, they found N+ = 109 and N- = 218.

Setup

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

Option	Value
Data Tab	
Find (Solve For)	**N+**
Power	**0.80**
Alpha	**0.05**
N+ (Size of Positive Group)	*Ignored since this is the Find Setting*
N- (Size of Negative Group)	**Use R**
R (Sample Allocation Ratio)	**2**
AUC1 (Area Under Curve 1)	**0.922222**
AUC2 (Area Under Curve 2)	**0.819444**
Lower FPR	**0.00**
Upper FPR	**0.20**
Correlation+	**0.6**
Correlation-	**0.6**
Type of Data	**Discrete (Ratings)**
B1 (SD Ratio)	**1**
B2 (SD Ratio)	**1**
Alternative Hypothesis	**Two-Sided Test**

Output

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

Numeric Results

Numeric Results for Testing AUC1 = AUC2 with Discrete (Rating) Data
Test Type = Two-Sided. FPR1 = 0.0. FPR2 = 0.200. B1 = 1.000. B2 = 1.000. Allocation Ratio = 2.000.

Power	N+	N-	AUC1'	AUC2'	Diff'	AUC1	AUC2	Diff	Alpha	Beta
0.8027	109	218	0.9222	0.8194	-0.1028	0.1720	0.1350	-0.0370	0.0500	0.1973

Note that the sample sizes of 109 and 218 match exactly with the results of Obuchowski.

Example 5 – Validation using Hanley

The formulas for continuous data were given in Hanley and McNeil (1982). On page 34 of their article they provide a table of sample sizes calculated using their formulas. We will duplicate their results for AUC1 = 0.70 and AUC2 = 0.75. Using a one-sided test of significance with alpha = 0.05 and a sample allocation ratio of 1.0, they found the number of subjects for both the positive and negative groups to be 652, 897, and 1131 for statistical powers of 80%, 90%, and 95%, respectively.

When using Hanley and McNeil's formulation, the values of B1, B2, FPR1, and FPR2 are ignored. Also, in this case, the correlations are set to 0.0.

Setup

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

Option	Value
Data Tab	
Find (Solve For)	**N+**
Power	**0.8 0.9 0.95**
Alpha	**0.05**
N+ (Size of Positive Group)	*Ignored since this is the Find Setting*
N- (Size of Negative Group)	**Use R**
R (Sample Allocation Ratio)	**1**
AUC1 (Area Under Curve 1)	**0.7**
AUC2 (Area Under Curve 2)	**0.75**
Lower FPR	**0.00**
Upper FPR	**1.00**
Correlation+	**0.0**
Correlation-	**0.0**
Type of Data	**Continuous**
Alternative Hypothesis	**One-Sided Test**

Output

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

Numeric Results 1

Numeric Results for Testing AUC1 = AUC2 with Continuous Data
Test Type = One-Sided. FPR1 = 0.0. FPR2 = 1.0. B1 = 1.000. B2 = 1.000. Allocation Ratio = 1.000.

Power	N+	N-	AUC1'	AUC2'	Diff'	AUC1	AUC2	Diff	Alpha	Beta
0.9501	1129	1129	0.7000	0.7500	0.0500	0.7000	0.7500	0.0500	0.0500	0.0499
0.9001	897	897	0.7000	0.7500	0.0500	0.7000	0.7500	0.0500	0.0500	0.0999
0.8003	652	652	0.7000	0.7500	0.0500	0.7000	0.7500	0.0500	0.0500	0.1997

Note that the sample sizes of 897 and 652 match exactly with the results of Hanley and McNeil. The 1129 is two less than their 1131. This difference may be due to refinements in computing the normal probability distribution used in ***PASS***. You can compare these sample sizes by calculating their power.

Numeric Results 2

Numeric Results for Testing AUC1 = AUC2 with Continuous Data
Test Type = One-Sided. FPR1 = 0.0. FPR2 = 1.0. B1 = 1.000. B2 = 1.000. Allocation Ratio = 1.000.

Power	N+	N-	AUC1'	AUC2'	Diff'	AUC1	AUC2	Diff	Alpha	Beta
0.9499	1128	1128	0.7000	0.7500	0.0500	0.7000	0.7500	0.0500	0.0500	0.0501
0.9501	1129	1129	0.7000	0.7500	0.0500	0.7000	0.7500	0.0500	0.0500	0.0499
0.9502	1130	1130	0.7000	0.7500	0.0500	0.7000	0.7500	0.0500	0.0500	0.0498
0.9504	1131	1131	0.7000	0.7500	0.0500	0.7000	0.7500	0.0500	0.0500	0.0496
0.9505	1132	1132	0.7000	0.7500	0.0500	0.7000	0.7500	0.0500	0.0500	0.0495

Note that the power for 1129 is 0.9501 while the power for 1131 is 0.9505. This is only a slight difference and explains why this value showed up in their table.

Chapter 270

Sensitivity and Specificity Test for One Group

Introduction

The power analysis of a diagnostic test is often based on the sensitivity and specificity of the test. In such a test, the outcome of the diagnostic screening test is compared to the gold standard. In the common case-control study, the gold standard must be known before. In a prospective study, the gold standard is determined subsequent to the study, so the case-control framework is not appropriate. In this procedure, the power analysis and sample size requirements of such a design are considered.

In a prospective study, a group of *n* subjects is obtained. Some of the subjects have the disease (condition of interest) and some do not. Each subject is given the diagnostic test for the disease. Subsequently, a gold standard test is used to obtain the true presence or absence of the disease. The gold standard may be a more expensive test, or it may be following the subject to determine if the disease status becomes more apparent.

The measures of diagnostic accuracy are sensitivity and specificity. *Sensitivity* (Se) is the probability that the diagnostic test is positive for the disease, given that the subject actually has the disease. *Specificity* (Sp) is the probability that the diagnostic test is negative, given that the subject does not have the disease. Mathematically,

Sensitivity = Pr(+test|disease)

Specificity = Pr(-test|no disease)

Li and Fine (2004) present sample size methodology for testing sensitivity and specificity using a prospective design. Their methodology will be used here. Other useful references are Obuchowski and Zhou (2002), Machin, Campbell, Tan, and Tan (2008), and Zhou, Obuchowski, and McClish (2002).

Prospective Study Design

In a prospective study, a group of *n* subjects is split into two groups: those with the disease of interest and those without it. Suppose a particular sample has n_1 with the disease and n_2 without the disease. A diagnostic test is administered to each subject (usually before the disease status is determined) and its output is recorded. The diagnostic test outcome is either positive or negative for the disease. Suppose that in the n_1 subjects with the disease, s_1 have a positive test outcome and s_2 have a negative outcome. Similarly, in the n_2 subjects without the disease r_1 have positive outcomes and r_2 have negative outcomes. Sensitivity is estimated by s_1 / n_1 and specificity is estimated by r_2 / n_2. A useful diagnostic test has high values of both *Se* and *Sp*.

Conditional on the values of n_1 and n_2, s_1 is $Binomial(n_1, Se)$. Thus, a one-sided test of the statistical hypothesis $H_0 : Se = Se_0$ *versus* $H_1 : Se = Se_1 > Se_0$ can be carried out using a binomial test. Hence, the power analysis is based on the binomial distribution conditional on the value of n_1. Similarly, the hypothesis test $H_0 : Sp = Sp_0$ *versus* $H_1 : Sp = Sp_1 > Sp_0$ is also based on the binomial distribution.

Binomial Model

A binomial variable should exhibit the following four properties:

1. The variable is binary --- it can take on one of two possible values.
2. The variable is observed a known number of times. Each observation or replication is called a Bernoulli trial. The number of replications is *n*. The number of times that the outcome of interest is observed is *r*. Thus *r* takes on the possible values 0, 1, 2, ..., *n*.
3. The probability, *P*, that the outcome of interest occurs is constant for each trial.
4. The trials are independent. The outcome of one trial does not influence the outcome of the any other trial.

The binomial probability is calculated using the formula

$$b(x \mid n, \pi) = \binom{n}{x} \pi^x (1-\pi)^{n-x} \text{ where } \binom{n}{r} = \frac{n!}{r!(n-r)!}.$$

Using the binomial probability formula, the sample size necessary to meet both a significance level and a power requirement may be found by solving the following to equations simultaneously:

Significance Level $B(s_1 > s_\alpha \mid n_1, Se_0) = \alpha$

$$B(s_1 > s_\alpha \mid n_1, Se_1) = 1 - \beta$$

where

$$B(s_1 > s_\alpha \mid n_1, \pi) = \sum_{x=s_\alpha+1}^{n_1} b(x \mid n_1, Se_0)$$

A similar calculation can be made for the specificity. Further details of this procedure are given in the Tests for One Proportion chapter.

Note that these formulas give n_1, not n. To obtain n, n_1 is inflated by the disease prevalence P to obtain $n = n_1 / P$. This is called Method 0 in the paper by Li and Fine (2004).

Procedure Options

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

Data Tab

The Data tab contains the parameters associated with this test such as the sensitivities, specificities, sample sizes, alphas, and powers.

Solve For

Find (Solve For)

This option specifies the parameter to be solved for using the other parameters. The parameters that may be selected are *Se1*, *Sp1*, *Power and Beta*, and *n*. Under most situations, you will select either *Power and Beta* or *n*.

Since this procedure actually results in two tests, you can search for *n* based on the power of the sensitivity test or the power of the specificity test.

If you want to search for *Se1* or *Sp1*, you must indicate whether to search above or below the values of *Se0* or *Sp0*, respectively.

Error Rates

Power (1-Beta)

This option specifies one or more values for the desired power. Power is the probability of rejecting a false null hypothesis, and is equal to 1- Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected.

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

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

Alpha (Significance Level)

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

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha and this is still the most common choice today.

Note that because of the discrete nature of the binomial distribution, the alpha level rarely will be achieved exactly.

A single value may be entered here or a range of values such as *0.05 to 0.2 by 0.05* may be entered.

Sample Size

n (Sample Size)

This option specifies the total number of observations in the sample. Values must be integers greater than one.

You may enter a single value or a range of values such as *10, 50, 100* or *10 to 100 by 10*.

P (Prevalence)

Specify one or more values for the disease prevalence: the anticipated proportion of the population of interest that has the disease. Because this is a proportion all values must be between zero and one.

You may enter a single value or a range of values such as *0.1, 0.2, 0.3*.

Effect Size

Se0 (Null Sensitivity)

This is the value of the sensitivity under the null (H0) hypothesis. The sensitivity estimated from the data is compared to this value by the statistical test. Usually, this is the base line or standard value.

Sensitivity = Pr(+Test|Disease).

Range

All sensitivities must be between zero and one.

You may enter a list of values such as *0.5, 0.6, 0.7* or *0.5 to 0.8 by 0.1*.

Se1 (Alternative Sensitivity)

This is the value of the sensitivity under the alternative (H1) hypothesis.

Sensitivity = Pr(+Test|Disease).

Detectable Difference

Se1 – Se0 is the size of the difference that the study will detect.

Range

All sensitivities must be between zero and one. Also, this value cannot be set equal to Se0.

You may enter a list of values such as *0.5, 0.6, 0.7* or *0.5 to 0.8 by 0.1*.

Sp0 (Null Specificity)

This is the value of the specificity under the null (H0) hypothesis. The specificity estimated from the data is compared to this value by the statistical test. Usually, this is the base line or standard value.

Specificity = Pr(-Test|Not Disease).

Range

All specificities must be between zero and one.

You may enter a list of values such as *0.5, 0.6, 0.7* or *0.5 to 0.8 by 0.1*.

Sp1 (Alternative Specificity)

This is the value of the specificity under the alternative (H1) hypothesis.

Specificity = Pr(+Test|Disease).

Detectable Difference

Sp1 – Sp0 is the size of the difference that the study will detect.

Range

All specificities must be between zero and one. Also, this value cannot be set equal to Se0.

You may enter a list of values such as *0.5, 0.6, 0.7* or *0.5 to 0.8 by 0.1*.

Test

H1 (Alternative Hypothesis)

Specify the alternative hypothesis of both the sensitivity and the specificity tests. Since the null hypothesis is equality (a difference of zero), the alternative is all that needs to be specified here.

Usually, the two-sided ('≠') option is selected.

Note that the "<" and ">" options are one-sided tests. When you choose one of these, you must make sure that the other settings (i.e. Solve For and Effect Size) are consistent with this choice.

Iterations Tab

The Iterations tab allows for specification of the maximum number of iterations to be used in searches.

Maximum Iterations

Maximum Iterations Before Search Termination

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 not reported.

Example 1 – Finding the Power

Suppose that diagnosing a certain type of cancer has required expensive and invasive test procedure. The sensitivity of this procedure is 71% and the specificity is 82%. A new diagnostic test has been developed that is much less expensive and invasive. The developers of the test want to design a prospective study to compare the old and new tests using a two-sided binomial test with a significance level of 0.05. They want to consider changes in sensitivity of 10%, 15%, 20%, and 25%. These changes translate to sensitivities of 79.20%, 81.65%, 85.20%, and 88.75%. The prevalence of the disease in the population of interest is 6%. The power will be determined for trials with sample sizes between 300 and 3000 incremented by 300. They want to consider a 10% increase in specificity which is 90.2%.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for One-Sample Sensitivity and Specificity** procedure window by clicking on **Diagnostic Tests (ROC)**, and then clicking on **Tests for One-Sample Sensitivity and Specificity**. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power or Beta	*Ignored since this is the Find setting*
Alpha	**0.05**
n (Sample Size)	**300 to 3000 by 300**
P (Prevalence)	**0.06**
Se0 (Null Sensitivity)	**0.71**
Se1 (Alternative Sensitivity)	**0.792 0.864 0.828 0.8875**
Sp0 (Null Specificity)	**0.82**
Sp1 (Alternative Specificity)	**0.902**
H1 (Alternative Hypothesis)	**H1: Se ≠ Se0, H1: Sp ≠ Sp0**
Reports Tab	
All Show Options	**Checked**
Plot Label Probabilities	**4**

Annotated Output

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

Numeric Results

Numeric Results for testing H0: Se = Se0 vs. H1: Se ≠ Se0 and H0: Sp = Sp0 vs. H1: Sp ≠ Sp0
Test Statistic: Binomial Test

			--- Sensitivity ----		--- Specificity ---		---------- Alpha ----------			
----- Power ------		Sample Size	H0	H1	H0	H1		Sens.	Spec.	Prevalence
Sens.	Spec.	N1 and N	Se0	Se1	Sp0	Sp1	Target	Actual	Actual	P
0.0874	0.9719	18 300	0.7100	0.7920	0.8200	0.9020	0.0500	0.0340	0.0360	0.0600
0.1051	0.9999	36 600	0.7100	0.7920	0.8200	0.9020	0.0500	0.0256	0.0425	0.0600
0.1813	1.0000	54 900	0.7100	0.7920	0.8200	0.9020	0.0500	0.0349	0.0488	0.0600
0.3429	1.0000	72 1200	0.7100	0.7920	0.8200	0.9020	0.0500	0.0385	0.0439	0.0600
0.3846	1.0000	90 1500	0.7100	0.7920	0.8200	0.9020	0.0500	0.0362	0.0481	0.0600
0.4183	1.0000	108 1800	0.7100	0.7920	0.8200	0.9020	0.0500	0.0333	0.0462	0.0600
0.5340	1.0000	126 2100	0.7100	0.7920	0.8200	0.9020	0.0500	0.0392	0.0463	0.0600
0.5526	1.0000	144 2400	0.7100	0.7920	0.8200	0.9020	0.0500	0.0342	0.0485	0.0600
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.

References

Obuchowski, N.A., Zhou, X.H. 2002. 'Prospective studies of diagnostic test accuracy when disease prevalence is low,' Biostatistics, Volume 3, No. 4, pages 477-492.
Li, J., Fine, J. 2004. 'On sample size for sensitivity and specificity in prospective diagnostic accuracy studies,' Statistics in Medicine, Volume 23, pages 2537-2550.
Machin, D., Campbell, M.J., Tan, S.B., Tan, S.H. 2008. Sample Size Tables for Clinical Studies, Third Edition. Wiley-Blackwell, Chichester, United Kingdom.
Zhou, X.H., Obuchowski, N.A., McClish, D.K. 2002. Statistical Methods in Diagnostic Medicine. Wiley-Interscience, New York.

Report Definitions

Sens. Power is the power of the sensitivity test. It is based on the N1 observations.
Spec. Power is the power of the specificity test. It is based on the N-N1 observations.
N is the total sample size of the study. It is equal to N1 + N2.
Se0 is the sensitivity under H0. The sensitivity is the proportion of diseased subjects that yield a positive test result.
Se1 is the sensitivity under H1. The sensitivity is the proportion of diseased subjects that yield a positive test result.
Sp0 is the specificity under H0. The specificity is the proportion of non-diseased subjects that yield a negative test result.
Sp1 is the specificity under H1. The specificity is the proportion of non-diseased subjects that yield a negative test result.
Target Alpha is the alpha (probability of rejecting H0 when H0 is true) that was desired.
Actual Sens. Alpha is the alpha that was actually achieved by the sensitivity test, calculated from the binomial distribution.
Actual Spec. Alpha is the alpha that was actually achieved by the specificity test, calculated from the binomial distribution.
P is proportion of the population that actually has the condition (disease) of interest, called the prevalence.

Summary Statements

A total sample size of 300 (which includes 18 subjects with the disease) achieves 9% power to detect a change in sensitivity from 0.71 to 0.792 using a two-sided binomial test and 97% power to detect a change in specificity from 0.82 to 0.902 using a two-sided binomial test. The target significance level is 0.05. The actual significance level achieved by the sensitivity test is 0.0340 and achieved by the specificity test is 0.0360. The prevalence of the disease is 0.06.

This report shows the values of each of the parameters, one scenario per row. Because of the discrete nature of the binomial distribution, the stated (Target) alpha is usually greater than the actual alpha. Hence, we also show the Actual Alpha along with the rejection region.

Sens. Power

This is the power of the sensitivity test. It is calculated from the binomial distribution using the N1 observations of the diseased subjects.

Spec. Power

This is the power of the specificity test. It is calculated from the binomial distribution using the N-N1 observations of the non-diseased subjects.

N

This is the total sample size of the study, n. It is equal to N1 + N2. The number of diseased subjects is N1 = NP. The number of non-diseased subjects is N2 = N(1-P).

Se0

This is the sensitivity under H0. The sensitivity is the proportion of diseased subjects that yield a positive test result.

Se1

This is the sensitivity under H1. The difference between Se1 and Se0 is the difference that is detected by the study.

Sp0

is the specificity under H0. The specificity is the proportion of non-diseased subjects that yield a negative test result.

Sp1

This is the specificity under H1. The difference between Sp1 and Sp0 is the difference that is detected by the study.

Target Alpha

This is the alpha (probability of rejecting H0 when H0 is true) that was desired. Because the binomial is a discrete distribution, the target value is seldom obtained. Rather, the actual value is lower than alpha.

Actual Sens. Alpha

This is the alpha that was actually achieved by the sensitivity test, calculated from the binomial distribution.

Actual Spec. Alpha

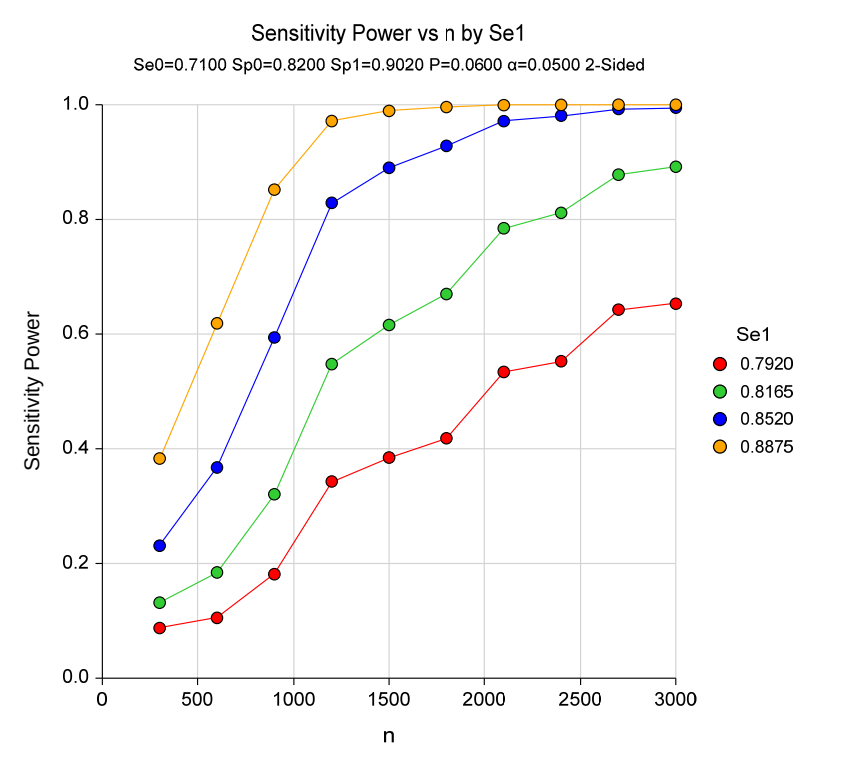
This is the alpha that was actually achieved by the specificity test, calculated from the binomial distribution.

P

This is proportion of the population that actually has the condition (disease) of interest, called the prevalence.

The values from this table are plotted in the charts below.

Plots Section



This plot shows the relationship between power, sample size, and Se1 in this example.

Example 2 – Finding the Sample Size

Continuing with Example 1, suppose you want to study the impact of various choices for Se1 on sample size. Using a significance level of 0.05 and 90% power, find the sample size when Se1 is 79.20%, 81.65%, 85.20%, and 88.75%. Assume that a two-tailed binomial test will be used.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for One-Sample Sensitivity and Specificity** procedure window by clicking on **Diagnostic Tests (ROC)**, and then clicking on **Tests for One-Sample Sensitivity and Specificity**. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**n (Sensitivity)**
Power or Beta	**0.90**
Alpha	**0.05**
n (Sample Size)	*Ignored since this is the Find setting*
P (Prevalence)	**0.06**
Se0 (Null Sensitivity)	**0.71**
Se1 (Alternative Sensitivity)	**0.792 0.864 0.828 0.8875**
Sp0 (Null Specificity)	**0.82**
Sp1 (Alternative Specificity)	**0.902**
H1 (Alternative Hypothesis)	**H1: Se ≠ Se0, H1: Sp ≠ Sp0**
Reports Tab	
All Show Options	**Checked**
Plot Label Probabilities	**4**

Annotated Output

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

Numeric Results

Numeric Results for testing H0: Se = Se0 vs. H1: Se ≠ Se0 and H0: Sp = Sp0 vs. H1: Sp ≠ Sp0
Test Statistic: Binomial Test

----- Power ------			--- Sensitivity ----		--- Specificity ---		---------- Alpha ----------			
								Sens.	Spec.	Preva-
Sens.	Spec.	Sample Size N1 and N	H0 Se0	H1 Se1	H0 Sp0	H1 Sp1	Target	Actual	Actual	lence P
0.9060	1.0000	299 4983	0.7100	0.7920	0.8200	0.9020	0.0500	0.0480	0.0480	0.0600
0.9054	1.0000	173 2883	0.7100	0.8165	0.8200	0.9020	0.0500	0.0441	0.0482	0.0600
0.9128	1.0000	93 1550	0.7100	0.8520	0.8200	0.9020	0.0500	0.0399	0.0479	0.0600
0.9153	1.0000	55 917	0.7100	0.8875	0.8200	0.9020	0.0500	0.0379	0.0413	0.0600

Example 3 – Validation using Li and Fine

Li and Fine (2004) page 2545 give the results of a power analysis indicate that if Se0 = 0.5, Se1 = .9, P = .01, alpha = 0.05 (one-sided), and power = 0.90, that *n* = 1100.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for One-Sample Sensitivity and Specificity** procedure window by clicking on **Diagnostic Tests (ROC)**, and then clicking on **Tests for One-Sample Sensitivity and Specificity**. You may then make the appropriate entries as listed below, or open **Example 3** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**n (Sensitivity)**
Power or Beta	**0.90**
Alpha	**0.05**
n (Sample Size)	*Ignored since this is the Find setting*
P (Prevalence)	**0.01**
Se0 (Null Sensitivity)	**0.5**
Se1 (Alternative Sensitivity)	**0.9**
Sp0 (Null Specificity)	**0.5**
Sp1 (Alternative Specificity)	**0.9**
H1 (Alternative Hypothesis)	**H1: Se > Se0, H1: Sp > Sp0**

Annotated Output

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

Numeric Results

Numeric Results for testing H0: Se = Se0 vs. H1: Se > Se0 and H0: Sp = Sp0 vs. H1: Sp > Sp0
Test Statistic: Binomial Test

----- Power ------			--- Sensitivity ----		--- Specificity ---		---------- Alpha ----------			
		Sample Size	H0	H1	H0	H1		Sens.	Spec.	Prevalence
Sens.	Spec.	N1 and N	Se0	Se1	Sp0	Sp1	Target	Actual	Actual	P
0.9104	1.0000	11 1100	0.5000	0.9000	0.5000	0.9000	0.0500	0.0327	0.0448	0.0100

PASS has also obtained an *n* of 1100.

It is interesting to note that a sample size of 1050 will also result in the identical power. This is because the amount of interest is N1 which is 11. If *n* is 1050, N1 = *n*P = 1050 x 0.01 = 10.5 ~ 11. Thus all values of *n* between 1050 and 1149 will result in the same power.

Also, the value of the power in the article is 0.904 while PASS has obtained 0.910. This difference arises because 0.904 is the unconditional power while 0.910 is the conditional power.

270-12 Sensitivity and Specificity Test for One Group

Chapter 275

Sensitivity Test for Two Groups

Introduction

This procedure gives power or required sample size for comparing two diagnostic tests when the outcome is sensitivity (or specificity). In this design, the outcome of each of two diagnostic screening tests is compared to a gold standard.

Specifically, a set of *n* subjects is randomly divided into two groups. In each group, a portion of the subjects have the disease (condition of interest) and a portion does not. Each subject is given the one of the diagnostic tests. Subsequently, a gold standard test is used to obtain the true presence or absence of the disease. The gold standard may be a more expensive test, difficult to determine, or require the sacrifice of the subject.

The measures of diagnostic accuracy are sensitivity and specificity. *Sensitivity* (Se) is the probability that the diagnostic test is positive for the disease, given that the subject actually has the disease. *Specificity* (Sp) is the probability that the diagnostic test is negative, given that the subject does not have the disease. Mathematically,

Sensitivity = Pr(+test|disease)

Specificity = Pr(-test|no disease)

Li and Fine (2004) present sample size methodology for testing sensitivity and specificity using a two-group, prospective design. Their methodology is used here. Other useful references are Obuchowski and Zhou (2002), Machin, Campbell, Tan, and Tan (2008), and Zhou, Obuchowski, and McClish (2002).

Prospective Study Design

In a two-group, prospective study, a group of *n* subjects is split into two groups: those that receive diagnostic test 1 and those that receive diagnostic test 2. Each of these groups is divided further into those with the disease of interest and those without it. Suppose that the k^{th} group (k = 1 or 2) has n_{k1} with the disease and n_{k2} without the disease. A diagnostic test is administered to each subject (usually before the disease status is determined) and its output is recorded. The diagnostic test outcome is either positive or negative for the disease. Suppose that in the n_{k1} subjects with the disease, s_{k1} have a positive test outcome and s_{k2} have a negative outcome. Similarly, in the n_{k2} subjects without the disease r_{k1} have positive outcomes and r_{k2} have negative outcomes. Sensitivity in each group is estimated by s_{k1}/n_{k1} and specificity is estimated by r_{k2}/n_{k2}. A useful diagnostic test has high values of both *Se* and *Sp*.

Conditional on the values of n_{k1} and n_{k2}, s_{k1} is $Binomial(n_{k1}, Se_k)$. Thus, a one-sided test of the statistical hypothesis $H_0 : Se_1 = Se_2$ *versus* $Se_1 < Se_2$ can be carried out using any of the two-sample proportion tests (see chapter 200 for more details on two-sample proportion tests). The power analysis of these tests follows the same pattern as other two-sample proportion tests, except that the disease prevalence in the two groups must be accounted for.

Specificity

This procedure computes the sample size for sensitivity. If you want a power analysis or sample size for specificity, you can use this procedure with the following minor adjustments.

1. Replace *Se* with *Sp* in all entries.
2. Replace the disease prevalence with 1 – prevalence. That is, the prevalence becomes the proportion without the disease.

Comparing Two Sensitivities

When analyzing the data from studies such as this, one usually compares the two binomial sensitivities, Se_1 and Se_2. Note that these values are estimated solely by the subjects with the disease. The data for those subjects without the disease is ignored. The data is displayed in a 2-by-2 contingency table as follows

	Test Outcome		
Group	**Positive**	**Negative**	**Total**
1	s_{11}	s_{12}	n_{11}
2	s_{21}	s_{22}	n_{21}

A popular test statistic for comparing the sensitivities is *Fisher's Exact Test* or the *Chi-square Test* with one degree of freedom.

Power Calculation

The power for a test statistic that is based on the normal approximation can be computed exactly using two binomial distributions. The following steps are taken to compute the power of such a test.

1. Find the critical value (or values in the case of a two-sided test) using the standard normal distribution. The critical value, $z_{critical}$, is that value of z that leaves exactly the target value of alpha in the appropriate tail of the normal distribution. For example, for an upper-tailed test with a target alpha of 0.05, the critical value is 1.645.
2. Compute the value of the test statistic, z_t, for every combination of s_{11} and s_{21}. A small value (around 0.0001) can be added to the zero cell counts to avoid numerical problems that occur when the cell value is zero.
3. If $z_t > z_{critical}$, the combination is in the rejection region. Call all combinations of s_{11} and s_{21} that lead to a rejection the set *A*.
4. Compute the power for given values of Se_1 and Se_2 as

$$1-\beta = \sum_A \binom{n_{11}}{s_{11}} Se_1^{s_{11}} (1-Se_1)^{n_{11}-s_{11}} \binom{n_{21}}{s_{21}} Se_2^{s_{21}} (1-Se_2)^{n_{21}-s_{21}}$$

5. Compute the actual value of alpha achieved by the design by substituting Se_1 for Se_2 in the above formula

$$\alpha^* = \sum_A \binom{n_{11}}{s_{11}} Se_1^{s_{11}} (1-Se_1)^{n_{11}-s_{11}} \binom{n_{21}}{s_{21}} Se_1^{s_{21}} (1-Se_1)^{n_{21}-s_{21}}$$

When the sample sizes are large (say over 200), these formulas may take a little time to evaluate. In this case, a large sample approximation may be used.

Test Statistics

Various test statistics are available. The formulas for their power are given in Chapter 200 and they are not repeated here. The test statistics are

Fisher's Exact Test

The most useful reference we found for power analysis of Fisher's Exact test was in the StatXact 5 (2001) documentation. The material presented here is summarized from Section 26.3 (pages 866 – 870) of the StatXact-5 documentation. In this case, the test statistic is

$$T = -\ln\left[\frac{\binom{n_1}{x_1}\binom{n_2}{x_2}}{\binom{N}{m}}\right]$$

Chi-Square Test (Pooled and Unpooled)

This test statistic was first proposed by Karl Pearson in 1900. Although this test is usually expressed directly as a Chi-Square statistic, it is expressed here as a *z* statistic so that it can be more easily used for one-sided hypothesis testing.

Both *pooled* and *unpooled* versions of this test have been discussed in the statistical literature. The pooling refers to the way in which the standard error is estimated. In the pooled version, the two proportions are averaged, and only one proportion is used to estimate the standard error. In the unpooled version, the two proportions are used separately.

The formula for the test statistic is

$$z_t = \frac{\hat{p}_1 - \hat{p}_2}{\hat{\sigma}_D}$$

Pooled Version

$$\hat{\sigma}_D = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

$$\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$$

Unpooled Version

$$\hat{\sigma}_D = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

Chi-Square Test with Continuity Correction

Frank Yates is credited with proposing a correction to the Pearson Chi-Square test for the lack of continuity in the binomial distribution. However, the correction was in common use when he proposed it in 1922.

Both *pooled* and *unpooled* versions of this test have been discussed in the statistical literature. The pooling refers to the way in which the standard error is estimated. In the pooled version, the two proportions are averaged, and only one proportion is used to estimate the standard error. In the unpooled version, the two proportions are used separately.

The continuity corrected z-test is

$$z = \frac{\left(\hat{p}_1 - \hat{p}_2\right) + \frac{F}{2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}{\hat{\sigma}_D}$$

where F is -1 for lower-tailed, 1 for upper-tailed, and both -1 and 1 for two-sided hypotheses.

Pooled Version

$$\hat{\sigma}_D = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

$$\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$$

Unpooled Version

$$\hat{\sigma}_D = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

Conditional Mantel Haenszel Test

The conditional Mantel Haenszel test, see Lachin (2000) page 40, is based on the *index frequency*, x_{11}, from the 2x2 table. The formula for the z-statistic is

$$z = \frac{x_{11} - E(x_{11})}{\sqrt{V_c(x_{11})}}$$

where

$$E(x_{11}) = \frac{n_1 m_1}{N}$$

$$V_c(x_{11}) = \frac{n_1 n_2 m_1 m_2}{N^2(N-1)}$$

Likelihood Ratio Test

In 1935, Wilks showed that the following quantity has a chi-square distribution with one degree of freedom. Using this test statistic to compare proportions is presented, among other places, in Upton (1982). The likelihood ratio test statistic is computed as

$$LR = 2\begin{bmatrix} a\ln(a) + b\ln(b) + c\ln(c) + d\ln(d) + \\ N\ln(N) - s\ln(s) - f\ln(f) - m\ln(m) - n\ln(n) \end{bmatrix}$$

Procedure Options

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

Data Tab

The Data tab contains the parameters associated with this test such as the sensitivities, specificities, sample sizes, alphas, and powers.

Solve For

Find (Solve For)

This option specifies the parameter to be solved for using the other parameters. The parameters that may be selected are *Se2*, *Power and Beta*, *N1*, and *N2*. Under most situations, you will select either *Power and Beta* or *N1*.

Error Rates

Power (1-Beta)

This option specifies one or more values for the desired power. Power is the probability of rejecting a false null hypothesis, and is equal to 1- Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected.

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

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

Alpha (Significance Level)

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

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha and this is still the most common choice today.

Note that because of the discrete nature of the binomial distribution, the alpha level rarely will be achieved exactly.

A single value may be entered here or a range of values such as *0.05 to 0.2 by 0.05* may be entered.

Sample Size

N1 (Sample Size Group 1)

This option specifies the total number of observations in the group 1. Note that in prospective designs, the effective (those with disease) sample size is found by multiplying N1 by the prevalence, P. For example, if N1 = 2000 and P = 0.1, the effective sample size for this group is 200.

Values must be integers greater than one. You may enter a single value or a range of values such as *10, 50, 100* or *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 also 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.

P (Prevalence)

Specify one or more values for the disease prevalence: the anticipated proportion of the population of interest that has the disease. Because this is a proportion all values must be between zero and one.

You may enter a single value or a range of values such as *0.1, 0.2, 0.3*.

Effect Size

Se1 (Sensitivity Groups 1 and 2|H0)

Enter the value of Se1 which is the sensitivity in both groups assumed by the null hypothesis, H0. The detectable difference detected by this design is Se1 - Se2.

Note: Sensitivity = Pr(+Test|Disease).

Range

All sensitivities must be between zero and one.

You may enter a list of values such as *0.5, 0.6, 0.7* or *0.5 to 0.8 by 0.1*.

Se2 (Sensitivity Group 2|H1)

Enter the value of Se2 which is the sensitivity in group 2 assumed by the alternative hypothesis, H1. The detectable difference detected by this design is Se1 - Se2.

Note: Sensitivity = Pr(+Test|Disease).

Range

All sensitivities must be between zero and one. Also, this value cannot be set equal to Se0.

You may enter a list of values such as *0.5, 0.6, 0.7* or *0.5 to 0.8 by 0.1*.

Test

Alternative Hypothesis (H1)

Specify the alternative hypothesis of the test. Usually, the two-sided ('≠') option is selected.

Note that the "<" and ">" options are one-sided tests. When you choose one of these, you must make sure that the other settings (i.e. Solve For and Effect Size) are consistent with this choice.

Test Type

Specify which test statistic will be used in searching and reporting.

Note that 'C.C.' is an abbreviation for *Continuity Correction*. This refers to the adding or subtracting 2/(N1+N2) to (or from) the numerator of the z-value to bring the normal approximation closer to the binomial distribution.

Iterations Tab

The Iterations tab allows for specification of the maximum number of iterations to be used in searches.

Maximum Iterations

Maximum Iterations Before Search Termination

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 not reported.

Zero Counts

Zero Count Adjustment Method

Zero cell counts often cause calculation problems. To compensate for this, a small value (called the Zero Count Adjustment Value) can be added either to all cells or to all cells with zero counts. This option specifies whether you want to use the adjustment and which type of adjustment you want to use. We recommend that you use the option *Add to zero cells only*.

Zero cell values often do not occur in practice. However, since power calculations are based on total enumeration, they will occur in power and sample size estimation.

Adding a small value is controversial, but can be necessary for computational considerations. Statisticians have recommended adding various fractions to zero counts. We have found that adding 0.0001 seems to work well.

Zero Count Adjustment Value

Zero cell counts cause many calculation problems when computing power or sample size. To compensate for this, a small value may be added either to all cells or to all zero cells. This value indicates the amount that is added. We have found that 0.0001 works well.

Be warned that the value of the ratio and the odds ratio will be affected by the amount specified here!

Exact Test Options

Maximum N1 or N2 for Exact Calculations

When either [N1(P)] or [N2(P)] is above this amount, power calculations are based on the normal approximation to the binomial. In this case, the actual value of alpha is not calculated. Currently, for three-gigahertz computers, a value near 200 is reasonable. As computers increase in speed, this number may be increased.

Example 1 – Finding the Power

Suppose that diagnosing a certain disease has used a certain diagnostic test which has a sensitivity of 71%. A new diagnostic test has been developed that is much less expensive and invasive. Researchers want to design a prospective study to compare the old and new tests using a two-sided Z Test with a significance level of 0.05. They want to consider changes in sensitivity of 10%, 15%, 20%, and 25%. These changes translate to sensitivities of 79.20%, 81.65%, 85.20%, and 88.75%. The prevalence of the disease in the population of interest is 20%. The power will be determined for trials with sample sizes between 300 and 3000 incremented by 300.

Setup

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

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power or Beta	*Ignored since this is the Find setting*
Alpha	**0.05**
n (Sample Size)	**300 to 3000 by 300**
P (Prevalence)	**0.2**
Se1 (Sensitivity Groups 1 and 2|H0)	**0.71**
Se2 (Sensitivity Group 2|H1)	**0.792 0.864 0.828 0.8875**
Alternative Hypothesis (H1)	**H1: Se1 ≠ Se2**
Test Type	**Z Test (Pooled)**
Reports Tab	
All Show Options	**Checked**
Plot Label Probabilities	**4**

Annotated Output

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

Numeric Results

Numeric Results for testing H0: Se1 = Se2 vs. H1: Se1 ≠ Se2
Test Statistic: Z test with pooled variance

	------- Sample Size -------			----- Sensitivity -----		Prev-	Target	Actual	
Power	N1	N2	N	Se1	Se2	alence	Alpha	Alpha	Beta
0.1840	300	300	600	0.7100	0.7920	0.2000	0.0500	0.0505	0.8160
0.3112	600	600	1200	0.7100	0.7920	0.2000	0.0500		0.6888
0.4358	900	900	1800	0.7100	0.7920	0.2000	0.0500		0.5642
0.5469	1200	1200	2400	0.7100	0.7920	0.2000	0.0500		0.4531
0.6421	1500	1500	3000	0.7100	0.7920	0.2000	0.0500		0.3579
0.7213	1800	1800	3600	0.7100	0.7920	0.2000	0.0500		0.2787
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.

References

Obuchowski, N.A., Zhou, X.H. 2002. 'Prospective studies of diagnostic test accuracy when disease prevalence is low,' Biostatistics, Volume 3, No. 4, pages 477-492.
Li, J., Fine, J. 2004. 'On sample size for sensitivity and specificity in prospective diagnostic accuracy studies,' Statistics in Medicine, Volume 23, pages 2537-2550.
Machin, D., Campbell, M.J., Tan, S.B., Tan, S.H. 2008. Sample Size Tables for Clinical Studies, Third Edition. Wiley-Blackwell, Chichester, United Kingdom.
Zhou, X.H., Obuchowski, N.A., McClish, D.K. 2002. Statistical Methods in Diagnostic Medicine. Wiley-Interscience, New York.
Chow, S.C.; Shao, J.; Wang, H. 2003. Sample Size Calculations in Clinical Research. Marcel Dekker. New York.
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.
Fleiss, J. L., Levin, B., Paik, M.C. 2003. Statistical Methods for Rates and Proportions. Third Edition. John Wiley & Sons. New York.
Lachin, John M. 2000. Biostatistical Methods. John Wiley & Sons. New York.

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 corresponding populations.
Se1 is the sensitivity of groups 1 and 2 assuming H0.
Se2 is the sensitivity for group 2 assuming H1.
Target Alpha' is the probability of rejecting a true null hypothesis that was desired.
Actual Alpha' is the value of alpha that is actually achieved.
Beta is the probability of accepting a false null hypothesis.
P is proportion of the population that actually has the condition (disease) of interest, called the prevalence.

Summary Statements

Group sample sizes of 300 in group 1 and 300 in group 2 achieve 18% power to detect the change in group 2 sensitivity from 0.7100 under the null hypothesis to 0.7920 under the alternative hypothesis. The sensitivity in group 1 is assumed to be 0.7100 in both cases. The test statistic used is the two-sided Z test with pooled variance. The significance level of the test was targeted at 0.0500. The significance level actually achieved by this design is 0.0505. The prevalence of the disease (or condition of interest) is 0.2000.

This report shows the values of each of the parameters, one scenario per row. Because of the discrete nature of the binomial distribution, the stated (Target) alpha is usually unequal to the actual alpha. Hence, we also show the Actual Alpha along with the rejection region. Note that the actual alpha is only computed when the effective sample size is greater than 100.

Power

This is the power of the test.

N1 and N2

These are the group sample sizes. The number of diseased subjects in each group is N1 x P.

Se1

This is the sensitivity of both groups assuming H0. The sensitivity is the proportion of diseased subjects that yield a positive test result.

Se2

This is the sensitivity of group 2 assuming H1. The difference between Se1 and Se2 is the difference that is detected by the study.

Target Alpha

This is the alpha (probability of rejecting H0 when H0 is true) that was desired. Because the binomial is a discrete distribution, the target value is seldom obtained.

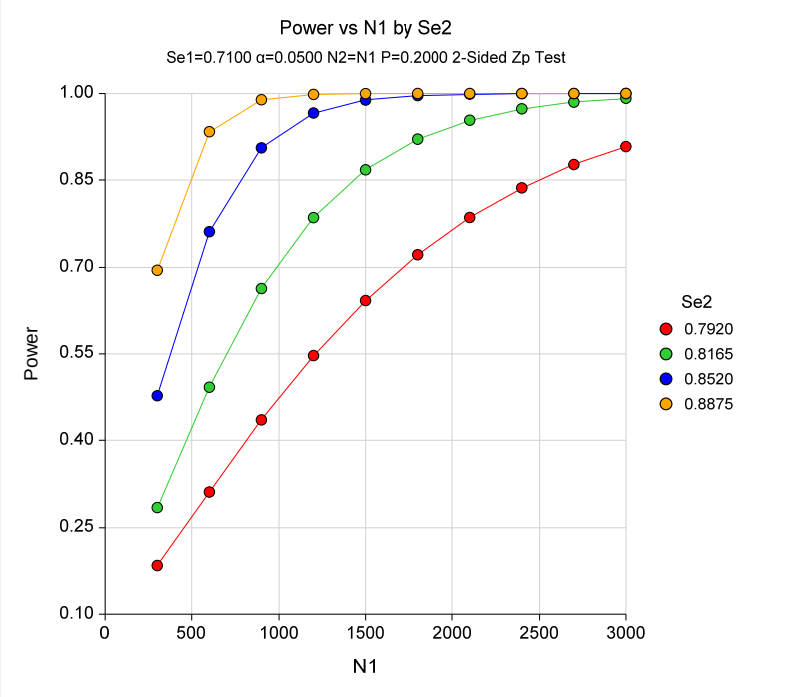
Actual Alpha

This is the alpha that was actually achieved by the test, calculated from the binomial distribution.

P

This is proportion of the population that actually has the condition (disease) of interest, called the prevalence.

Plots Section



Example 2 – Finding the Sample Size

Continuing with Example 1, suppose you want to study the impact of various choices for Se2 on sample size. Using a significance level of 0.05 and 90% power, find the sample size when Se2 is 79.20%, 81.65%, 85.20%, and 88.75%. Assume a two-tailed test is used.

Setup

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

Option	Value
Find (Solve For)	**Power and Beta**
Power or Beta	*Ignored since this is the Find setting*
Alpha	**0.05**
n (Sample Size)	**300 to 3000 by 300**
P (Prevalence)	**0.2**
Se1 (Sensitivity Groups 1 and 2|H0)	**0.71**
Se2 (Sensitivity Group 2|H1)	**0.792 0.864 0.828 0.8875**
Alternative Hypothesis (H1)	**H1: Se1 ≠ Se2**
Test Type	**Z Test (Pooled)**
Reports Tab	
All Show Options	**Checked**
Plot Label Probabilities	**4**

Annotated Output

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

Numeric Results

Numeric Results for testing H0: Se1 = Se2 vs. H1: Se1 ≠ Se2
Test Statistic: Z test with pooled variance

	------- Sample Size -------			----- Sensitivity -----		Prev-	Target	Actual	
Power	N1	N2	N	Se1	Se2	alence	Alpha	Alpha	Beta
0.9003	2915	2915	5830	0.7100	0.7920	0.2000	0.0500		0.0997
0.9002	1665	1665	3330	0.7100	0.8165	0.2000	0.0500		0.0998
0.9013	885	885	1770	0.7100	0.8520	0.2000	0.0500		0.0987
0.9024	530	530	1060	0.7100	0.8875	0.2000	0.0500		0.0976

Example 3 – Validation using Machin and Campbell (2008)

Machin, et al (2008) page 166-167 give the results of a sample size determination in which Se1 = 0.27, Se2 = 0.66, P = 0.25, alpha = 0.05 (two-sided), and power = 0.80. The resulting sample size is 98 per group or 196 total.

Setup

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

Option	Value
Data Tab	
Find (Solve For)	**N1**
Power or Beta	**0.80**
Alpha	**0.05**
N1	*Ignored since this is the Find setting*
P (Prevalence)	**0.25**
Se1 (Sensitivity Groups 1 and 2|H0)	**0.27**
Se2 (Sensitivity Group 2|H1)	**0.66**
Alternative Hypothesis (H1)	**H1: Se1 ≠ Se2**

Annotated Output

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

Numeric Results

Numeric Results for testing H0: Se1 = Se2 vs. H1: Se1 ≠ Se2
Test Statistic: Z test with pooled variance

	------- Sample Size -------			----- Sensitivity -----		Prev-	Target	Actual	
Power	N1	N2	N	Se1	Se2	alence	Alpha	Alpha	Beta
0.8170	96	96	192	0.2700	0.6600	0.2500	0.0500	0.0520	0.1830

PASS has also obtained an *N* of 192 which is slightly different from 196 in the book. This difference is probably the result of the exact binomial calculation used by PASS.

275-14 Sensitivity Test for Two Groups

Chapter 276

Tests for Paired Sensitivities

Introduction

This procedure gives power or required sample size for comparing two diagnostic tests when the outcome is sensitivity. In the design covered in this chapter, the sensitivities of two diagnostic tests are each performed on the same subject.

Specifically, a set of *N* subjects is selected and each is given the both of the diagnostic tests being considered. Subsequently, a gold standard test is also given and used to obtain the true presence or absence of the disease. The gold standard may be a more expensive test or require the sacrifice of the subject.

Measures of diagnostic accuracy are sensitivity and specificity. *Sensitivity* (Se) is the probability that the diagnostic test is positive for the disease, given that the subject actually has the disease. *Specificity* (Sp) is the probability that the diagnostic test is negative, given that the subject does not have the disease. Mathematically,

Sensitivity = Pr(+test|disease)

Specificity = Pr(-test|no disease)

Li and Fine (2004) present sample size methodology for testing sensitivity and specificity using a paired, prospective design. Their methodology is used here. Other useful references are Obuchowski and Zhou (2002), Machin, Campbell, Tan, and Tan (2008), and Zhou, Obuchowski, and McClish (2002).

Technical Details

In this, two diagnostic tests are administered to each of N subjects. Of those, N1 have the disease of interest. The N2 subjects without the disease are set aside and only the N1 diseased subjects are used in the analysis. The results can be displayed in the following 2-by-2 table.

	Test 2		
Test 1	**Yes**	**No**	**Total**
Yes	a	b	a+b
No	c	d	c+d
Total	a+c	b+d	N1

McNemar's Test

The McNemar test statistic is

$$\chi_1^2 = \frac{(b-c)^2}{b+c}$$

The null hypothesis that the sensitivities of the two tests are equal, that P(a+b) = P(a+c) is equivalent to the hypothesis that P(b) = P(c). The two-sided alternative hypothesis is that P(b) ≠ P(c).

Power Analysis

The sample size problem reduces to a study of how many Yes-No's and No-Yes's are needed. Once this has been determined, the overall sample size is found using the proportion of discordant pairs to inflate the sample size appropriately.

Some power analysis programs follow an approximate procedure. Since the McNemar statistic follows the binomial probability distribution for a fixed number of discordant pairs, they use formulas that use the normal approximation to the binomial and then adjust the sample size depending on the proportion of discordant pairs, D=P(b)+P(c). This is called the conditional procedure.

One such approximate formula is given by Machin, Campbell, Fayers, and Pinol (1997).

$$N1 = \frac{\left(z_{1-\alpha/s}(\psi+1) + z_{1-\beta}\sqrt{(\psi+1)^2 - (\psi-1)^2 D} \right)^2}{(\psi-1)^2 D}$$

where s is the number of sides to the test (one or two), $\psi = \frac{P(b)}{P(c)}$, $D = P(b) + P(c)$, and α and β are the usual type 1 and type 2 error rate probabilities.

However, Schork and Williams (1980) published a formula which provides the exact results for the unconditional case. This is the formulation that is used in ***PASS***.

$$Power = \sum_{x=r}^{N1} \sum_{y=0}^{I} \frac{N1!}{(N1-x)!\, y!\, (x-y)!} (1-D)^{N1-x} (D+\delta)^{y} (D-\delta)^{x-y}$$

where

$$D = P(b) + P(c)$$

$$\delta = P(b) - P(c)$$

N1 is total of all four cells

r is the smallest integer for which $\left(\frac{1}{2}\right)^{r} \leq \alpha$

I is the largest integer such that $\sum_{j=0}^{IR} \binom{x}{j} \left(\frac{1}{2}\right)^{x} \leq \alpha$

Finally, to obtain the estimate of N, we inflate N1 by the prevalence using N = N1/P, where P is the prevalence.

Specificity

This procedure computes the sample size for sensitivity. If you want a power analysis or sample size for specificity, you can use this procedure with the following minor adjustments.

1. Replace *Se* with *Sp* in all entries.
2. Replace the disease prevalence with 1 – prevalence. That is, the prevalence becomes the proportion without the disease.

Procedure Options

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

Data Tab

The Data tab contains the parameters associated with this test such as the sensitivities, specificities, sample sizes, alphas, and powers.

Solve For

Find (Solve For)

Select the parameter to be solved for in terms of the other parameters. Note that this parameter will be displayed on the vertical axis of the plot.

Possible choices for this parameter are

Se2 (Search > Se1)

Search for the Se2 (sensitivity of test 2 under H1) that results in the specified values of the other parameters. The binary search is conducted between Se1 and Min(1, Se1 + D).

Se2 (Search > Se1)

Search for the Se2 (sensitivity of test 2 under H1) that results in the specified values of the other parameters. The binary search is conducted between Max(0, Se1 - D) and Se1.

Power and Beta

Calculate the power based on the other parameters.

N

Search for the sample size needed to attain the values of the other parameters.

Error Rates

Power (1-Beta)

This option specifies one or more values for the desired power. Power is the probability of rejecting a false null hypothesis, and is equal to 1- Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected.

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

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

Alpha (Significance Level)

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

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha and this is still the most common choice today.

Note that because of the discrete nature of the binomial distribution, the alpha level rarely will be achieved exactly.

A single value may be entered here or a range of values such as *0.05 to 0.2 by 0.05* may be entered.

Sample Size

N (Number of Subjects)

Enter a value (or range of values) for the sample size, N. N is the total number of subjects (both diseased and non-diseased) in the study.

You may enter a range of values such as 10 to 100 by 10.

P (Prevalence)

Specify one or more values for the disease prevalence: the anticipated proportion of the population of interest that has the disease. Because this is a proportion all values must be between zero and one.

You may enter a single value or a range of values such as *0.1, 0.2, 0.3*.

Effect Size

Se1 (Sensitivity of Test 1)

Enter the value of Se1, the sensitivity of both tests assumed by the null hypothesis, H0. The difference detected by this design is Se1 - Se2. The values must be between 0 and 1 and cannot be equal to Se2.

Sensitivity = Pr(+Test|Disease).

You may enter a range of values such as *0.7, 0.8, 0.9* or *0.6 to 0.9 by 0.1*.

Se2 (Sensitivity of Test 2)

Enter the value of Se2 which is the sensitivity in test 2 assumed by the alternative hypothesis, H1. The difference detected by this design is Se1 - Se2. The values must be between 0 and 1 and cannot be equal to Se1.

Note: Sensitivity = Pr(+Test|Disease).

You can enter a range of values such as *0.75, 0.85, 0.95* or *0.55 to 0.95 by 0.1*.

Proportion Discordant (P10+P01)

This is the proportion of discordant pairs D. This value will be difficult to specify unless you have previous studies that give you some idea of what to expect. When you have no idea, Machin, Campbell, Fayers, and Pinol (1997) suggest the following strategy. Estimate values of *Se1* and *Se1*. Calculate the proportion of discordant pairs using the approximation

$$D = Se1(1 - Se2) + Se2(1 - Se1)$$

This approximation assumes that the two responses are independent within each subject, which will usually not be true. However, it may be the only way of determining a ball park value for this parameter.

Test

Alternative Hypothesis (H1)

Specify the alternative hypothesis of the test. Usually, the two-sided option is selected.

Options Tab

The Options tab allows the specification of options that control the searches.

Maximum Iterations

Maximum Iterations Before Search Termination

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 not reported.

Approximations

Use Approximations if N(P) is greater than

Below this value of N(P), the exact power calculation formula based on the binomial is used. Above this value of N(P), the approximate formula based on the normal approximation to the binomial is used. The exact formula suffers from numerical problems when N(P) is greater than 2000. On the other hand, the approximate formula tends to underestimate the N(P) necessary to achieve a certain beta value by about 5%.

You control which formula is used by setting this value.

Example 1 – Finding the Power

Suppose that diagnosing a certain disease has used a certain diagnostic test which has a sensitivity of 71%. A new diagnostic test has been developed that is much less expensive and invasive. Researchers want to design a paired, prospective study in which each subject will be given both the old and new diagnostic tests. The data will be analyzed using a two-sided McNemar's test with a significance level of 0.05. From previous studies, the percent discordant is estimated at 30%.

They want to consider changes in sensitivity of 10%, 15%, 20%, and 25%. These changes translate to sensitivities of 79.20%, 81.65%, 85.20%, and 88.75%. The prevalence of the disease in the population of interest is 20%. The power will be determined for trials with sample sizes between 300 and 2400 incremented by 300.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Paired Sensitivities** procedure window by clicking on **Diagnostic Tests (ROC)**, and then clicking on **Tests for Paired Sensitivities**. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**Power and Beta**
Power or Beta	*Ignored since this is the Find setting*
Alpha	**0.05**
N (Number of Subjects)	**300 to 2400 by 300**
P (Prevalence)	**0.2**
Se1 (Sensitivity Test 1)	**0.71**
Se2 (Sensitivity Test 2)	**0.792 0.864 0.828 0.8875**
D (Proportion Discordant)	**0.3**
Alternative Hypothesis	**Two-Sided**
Reports Tab	
All Show Options	**Checked**

Annotated Output

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

Numeric Results

Numeric Results for a Two-Sided, McNemar's Test

Power	N	Sensitivity of Test 1 Se1	Sensitivity of Test 2 Se2	Difference Se1-Se2	Proportion Discordant D	Prev-alence-P	Alpha	Beta
0.15137	300	0.710	0.792	-0.082	0.300	0.200	0.05000	0.84863
0.31319	600	0.710	0.792	-0.082	0.300	0.200	0.05000	0.68681
0.47378	900	0.710	0.792	-0.082	0.300	0.200	0.05000	0.52622
0.60339	1200	0.710	0.792	-0.082	0.300	0.200	0.05000	0.39661
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.

References

Obuchowski, p.n.A., Zhou, X.H. 2002. 'Prospective studies of diagnostic test accuracy when disease prevalence is low,' Biostatistics, Volume 3, No. 4, pages 477-492.

Li, j., Fine, j. 2004. 'On sample size for sensitivity and specificity in prospective diagnostic accuracy studies,' Statistics in Medicine, Volume 23, pages 2537-2550.

Machin, D., Campbell, M.j., Tan, S.B., Tan, S.H. 2008. Sample Size Tables for Clinical Studies, Third Edition. Wiley-Blackwell, Chichester, United Kingdom.

Zhou, X.H., Obuchowski, p.n.A., McClish, D.K. 2002. Statistical Methods in Diagnostic Medicine. Wiley-Interscience, New York.

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.

Report Definitions

Power is the probability of rejecting a false null hypothesis. It should be close to one.
N is the total number of subjects in the study.
Se1 is the sensitivity of test 1 assuming both H0 and H1.
Se2 is the sensitivity for test 2 assuming H1.
Difference is the Se1-Se2 assuming H1.
D, the proportion discordant, is proportion of diseased subjects that have unmatching test results.
Prevalence is proportion of diseased individuals in the population.
Alpha is the probability of rejecting a true null hypothesis, H0. It should be small.
Beta is the probability of accepting a false null hypothesis, H0. It should be small.

Summary Statements

A sample size of 300 subjects achieves 15% power to detect a difference of -0.082 between two diagnostic tests whose sensitivities are 0.710 and 0.792. This procedure uses a two-sided McNemar test with a significance level of 0.05000. The prevalence of disease in the population is 0.200. The proportion of discordant pairs is 0.300.

This report shows the values of each of the parameters, one scenario per row.

Power

This is the power of the test.

N

This is the total sample size. The number of diseased subjects is N(P).

Se1

This is the sensitivity of tests assuming H0. It is the sensitivity of test 1 under H1. The sensitivity is the proportion of diseased subjects that yield a positive test result.

Se2

This is the sensitivity of test 2 assuming H1.

Difference

This is the difference between the two sensitivities under H1. It is calculated as Se1 – Se2.

Proportion Discordant

D, the proportion discordant, is proportion of diseased subjects that have non-matching test results.

Prevalence

P, the prevalence, is proportion of the population that actually has the condition of interest (disease).

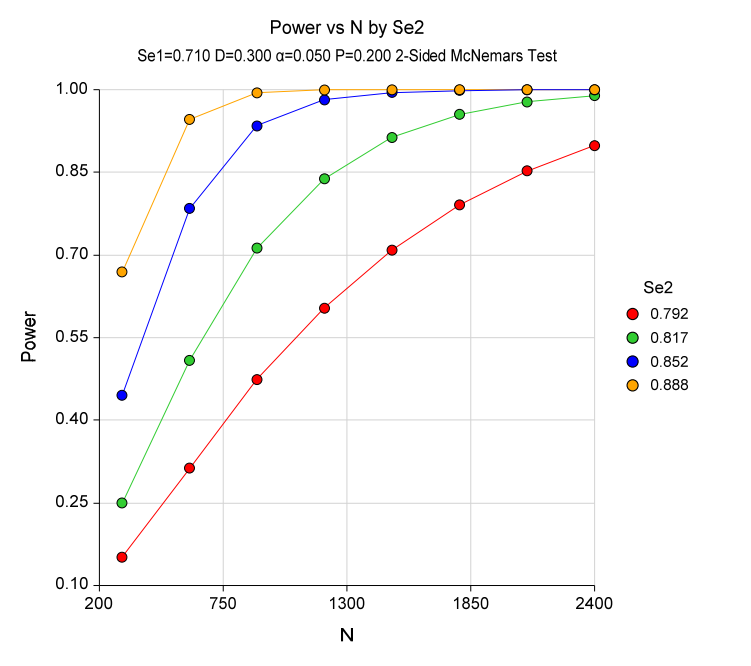
Alpha

This is the alpha (probability of rejecting H0 when H0 is true) that was desired.

Beta

Beta is the probability of accepting the null hypothesis when it is false.

Plots Section



Example 2 – Finding the Sample Size

Continuing with Example 1, suppose you want to study the impact of various choices for Se2 on sample size. Using a significance level of 0.05 and 90% power, find the sample size when Se2 is 79.20%, 81.65%, 85.20%, and 88.75%. Assume a two-tailed test is used.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Paired Sensitivities** procedure window by clicking on **Diagnostic Tests (ROC)**, and then clicking on **Tests for Paired Sensitivities**. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**N**
Power or Beta	**0.90**
Alpha	**0.05**
N (Number of Subjects)	*Ignored since this is the Find setting*
P (Prevalence)	**0.2**
Se1 (Sensitivity Test 1)	**0.71**
Se2 (Sensitivity Test 2)	**0.792 0.864 0.828 0.8875**
D (Proportion Discordant)	**0.3**
Alternative Hypothesis	**Two-Sided**

Annotated Output

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

Numeric Results

Numeric Results for a Two-Sided, McNemar's Test

Power	N	Sensitivity of Test 1 Se1	Sensitivity of Test 2 Se2	Difference Se1-Se2	Proportion Discordant D	Prev-alence- P	Alpha	Beta
0.90030	2415	0.710	0.792	-0.082	0.300	0.200	0.05000	0.09970
0.90097	1440	0.710	0.817	-0.107	0.300	0.200	0.05000	0.09903
0.90001	795	0.710	0.852	-0.142	0.300	0.200	0.05000	0.09999
0.90102	510	0.710	0.888	-0.178	0.300	0.200	0.05000	0.09898

It is interesting to note the reduction in sample size of the paired design over the two-sample design discussed in Chapter 275. A comparison of these sample sizes to those shows a reduction of over 50%.

Example 3 – Validation using Machin and Campbell (2008)

Machin, et al (2008) page 166-167 give the results of a sample size determination in which Se1 = 0.27, Se2 = 0.66, P = 0.25, alpha = 0.05 (two-sided), and power = 0.80. The resulting sample size is 80 using one formula and 100 using another. The formulas used in their book are based on a conditional argument that does not require the proportion discordant. We will compare our results to theirs using D = 0.4, 0.5, and 0.6.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Paired Sensitivities** procedure window by clicking on **Diagnostic Tests (ROC)**, and then clicking on **Tests for Paired Sensitivities**. You may then make the appropriate entries as listed below, or open **Example 3** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Data Tab	
Find (Solve For)	**N**
Power or Beta	**0.80**
Alpha	**0.05**
N (Number of Subjects)	*Ignored since this is the Find setting*
P (Prevalence)	**0.25**
Se1 (Sensitivity Test 1)	**0.66**
Se2 (Sensitivity Test 2)	**0.27**
D (Proportion Discordant)	**0.4 0.5 0.6**
Alternative Hypothesis	**Two-Sided**

Annotated Output

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

Numeric Results

Numeric Results for a Two-Sided, McNemar's Test

Power	N	Sensitivity of Test 1 Se1	Sensitivity of Test 2 Se2	Difference Se1-Se2	Proportion Discordant D	Prev-alence-P	Alpha	Beta
0.83196	80	0.660	0.270	0.390	0.400	0.250	0.05000	0.16804
0.80961	104	0.660	0.270	0.390	0.500	0.250	0.05000	0.19039
0.81101	128	0.660	0.270	0.390	0.600	0.250	0.05000	0.18899

PASS has also obtained an N of 80 for the case when D = 0.4. However, the sample size increases if D is 0.5 or 0.6.

References

A

A'Hern, R. P. A. 2001. "Sample size tables for exact single-stage phase II designs." *Statistics in Medicine*, Volume 20, pages 859-866.

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.

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.

B

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.

Barnard, G.A. 1947. "Significance tests for 2 x 2 tables." *Biometrika* 34:123-138.

Bartholomew, D.J. 1963. "The Sampling Distribution of an Estimate Arising in Life Testing." *Technometrics*, Volume 5 No. 3, 361-374.

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.

Benjamini, Y. and Hochberg, Y. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society, Series B (Methodological)*, Vol. 57, No. 1, 289-300.

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.

Bonett, Douglas. 2002. "Sample Size Requirements for Testing and Estimating Coefficient Alpha." *Journal of Educational and Behavioral Statistics*, Vol. 27, pages 335-340.

Borm, G.F., Fransen, J., and Lemmens, W.A.J.G. 2007. "A simple sample size formula for analysis of covariance in randomized clinical trials." *Journal of Clinical Epidemiology*, Vol. 60, pages 1234-1238.

Box, G.E.P., Hunter, S. and Hunter, J.S.. 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.

Brown, H., and Prescott, R. 2006. *Applied Mixed Models in Medicine*. 2nd ed. John Wiley & Sons Ltd. Chichester, West Sussex, England.

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.

C

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.

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.

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.

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.

Chow, S.-C.; Shao, J.; Wang, H. 2008. *Sample Size Calculations in Clinical Research, Second Edition*. Chapman & Hall/CRC. Boca Raton, Florida.

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.

Cook, R. D. and **Weisburg, S.** 1999. Applied Regression Including Computing and Graphics. John Wiley and Sons, Inc.

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.

D

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.

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.

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.

Demidenko, E. 2004. *Mixed Models – Theory and Applications*. John Wiley & Sons. Hoboken, New Jersey.

Desu, M. M. and Raghavarao, D. 1990. *Sample Size Methodology*. Academic Press. New York. (Presents many useful results for determining sample sizes.)

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.

Diggle, P.J., Liang, K.Y., and Zeger, S.L. 1994. *Analysis of Longitudinal Data*. Oxford University Press. New York, New York.

Dixon, W. J. and Tukey, J. W. 1968. "Approximate behavior of the distribution of Winsorized t," *Technometrics*, Volume 10, pages 83-98.

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.

Draghici, S. 2003. *Data Analysis Tools for DNA Microarrays*. Chapman & Hall/CRC. London. This is an excellent overview of most areas of Microarray analysis.

Dudoit, S., Yang, Y.H., Callow, M.J., and Speed, T.P. 2002. "Statistical Methods for Identifying Differentially Expressed Genes in Replicated cDNA Experiments," *Statistica Sinica*, Volume 12, pages 111-139.

Dunn, O. J. 1961. "Multiple comparisons among means," *Journal of the American Statistical Association*, Volume 56, pages 52-64.

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.

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.

E

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. 1971. “Forcing a Sequential Experiment to be Balanced.” *Biometrika*. Volume 58, pages 403-417.

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.

F

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., and Salih, F.A. 1987. “Statistical inference for coefficient alpha.” *Applied Psychological Measurement*, Vol. 11, pages 93-103.

Feldt, L.S. and 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.

Fisher, R. A. 1921. “On the probable error of a coefficient of correlation deduced from a small sample.” *Metron*, i (4), 1-32.

Flack, V. F., Afifi, A. A., Lachenbruch, P. A., and Schouten, H. J. A. 1988. "Sample Size Determinations for the Two Rater Kappa Statistic." *Psychometrika*, Volume 53, No. 3, pages 321-325.

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, 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.

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.

Fleming, T. R. 1982. "One-sample multiple testing procedure for Phase II clinical trials." *Biometrics*, Volume 38, pages 143-151.

Freedman, L.S. 1982. "Tables of the Number of Patients Required in Clinical Trials using the Logrank Test." *Statistics in Medicine*, 1:121-129.

G

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.

Gibbons, J. 1985. *Nonparametric Methods for Quantitative Analysis (2nd Edition)*. American Sciences Press. New York.

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.

Greenwood, J. A. and **Sandomire, M. M.** 1950. "Sample Size Required for Estimating the Standard Deviation as a Per Cent of its True Value", *Journal of the American Statistical Association*, Vol. 45, No. 250, pp. 257-260.

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.

Gu, K., Ng, H.K.T., Tang, M.L., and Schucany, W. 2008. "Testing the Ratio of Two Poisson Rates," *Biometrical Journal*, Volume 50, Number 2, pages 283-298.

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.

H

Hahn, G. J. and Meeker, W.Q. 1991. *Statistical Intervals*. John Wiley & Sons. New York.

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.

Harris, M., Horvitz, D. J., and **Mood, A. M.** 1948. "On the Determination of Sample Sizes in Designing Experiments", Journal of the American Statistical Association, Volume 43, No. 243, pp. 391-402.

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.

Hoaglin, Mosteller, and Tukey. 1985. *Exploring Data Tables, Trends, and Shapes*. John Wiley. New York.

Hochberg, Y. and Tamhane, A. C. 1987. *Multiple Comparison Procedures*. John Wiley & Sons. New York.

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.

Huffman, Michael. 1984. "An Improved Approximate Two-Sample Poisson Test," *Applied Statistics,* Volume 33, Number 2, pages 224-226.

J

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.

Julious, Steven A. 2004. "Tutorial in Biostatistics. Sample sizes for clinical trials with Normal data." *Statistics in Medicine,* 23:1921-1986.

Jung, Sin-Ho. 2005. "Sample size for FDR-control in microarray data analysis" *Bioinformatics*, 21(14):3097-3104.

Jung, Sin-Ho; Kang, Sun J.; McCall, Linda M.; Blumenstein, Brent. 2005**.** "Sample Sizes Computation for Two-Sample Noninferiority Log-Rank Test", *J. of Biopharmaceutical Statistics*, Volume 15, pages 969-979.

Juran, J.M. 1979. *Quality Control Handbook*. McGraw-Hill. New York.

K

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.

Katz, D., Baptista, J., Azen, S. P., and Pike, M. C. 1978. "Obtaining Confidence Intervals for the Risk Ratio in Cohort Studies," *Biometrics*, 34, pages 469-474.

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.

Kenward, M. G. and Roger, J. H. 1997. "Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood," *Biometrics*, 53, pages 983-997.

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.

Keppel, Geoffrey. 1991. *Design and Analysis: A Researcher's Handbook. Third Edition.* Prentice Hall. Englewood Cliffs, New Jersey.

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.

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.

Kupper, L. L. and **Hafner, K. B.** 1989. 'How Appropriate are Popular Sample Size Formulas?', The American Statistician, Volume 43, No. 2, pp. 101-105.

L

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.

Lakatos, Edward. 1988**.** "Sample Sizes Based on the Log-Rank Statistic in Complex Clinical Trials", *Biometrics*, Volume 44, March, pages 229-241.

Lakatos, Edward. 2002**.** "Designing Complex Group Sequential Survival Trials", *Statistics in Medicine,* Volume 21, pages 1969-1989.

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.

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, M.-L. T. 2004. *Analysis of Microarray Gene Expression Data*. Kluwer Academic Publishers. Norwell, Massachusetts.

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.

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.

Li, J. and Fine, J. 2004. "On sample size for sensitivity and specificity in prospective diagnostic accuracy studies" *Statistics in Medicine,* 23, pages 2537-2550.

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.

Littell, R. C. et al. 2006. *SAS for Mixed Models – Second Edition.* SAS Institute Inc., Cary, North Carolina.

Liu, 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.

Liu, H. and Wu, T. 2005. "Sample Size Calculation and Power Analysis of Time-Averaged Difference," *Journal of Modern Applied Statistical Methods*, Vol. 4, No. 2, pages 434-445.

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.

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.

M

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.

Machin, D., Campbell, M., Tan, S.B., and Tan, S.H. 2008. *Sample Size Tables for Clinical Studies, 3nd Edition*. Wiley-Blackwell. Chichester, United Kingdom.

Marubini, E. and **Valsecchi, M.G.** 1996. *Analysing Survival Data from Clinical Trials and Observational Studies*. John Wiley: New York, New York.

Matsumoto, M. and Nishimura,T. 1998. "Mersenne twister: A 623-dimensionally equidistributed uniform pseudorandom number generator" *ACM Trans. On Modeling and Computer Simulations*.

McClish, D.K. 1989. "Analyzing a Portion of the ROC Curve." *Medical Decision Making*, 9: 190-195

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.

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.

Moore, D. S. and McCabe, G. P. 1999. *Introduction to the Practice of Statistics*. W. H. Freeman and Company. New York.

Moura, Eduardo C. 1991. *How To Determine Sample Size And Estimate Failure Rate in Life Testing*. ASQC Quality Press. Milwaukee, Wisconsin.

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.

Muller, 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.

Muller, 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.

Muller, K. E. and Stewart, P.W. 2006. *Linear Model Theory: Univariate, Multivariate, and Mixed Models*. John Wiley & Sons Inc. Hoboken, New Jersey.

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.

N

Nam, Jun-mo. 1987. "A Simple Approximation for Calculating Sample Sizes for Detecting Linear Trend in Proportions," *Biometrics*, Volume 43, pages 701-705.

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.

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.

Ng, H.K.T, Gu, K., Tang, M.L. 2007. "A comparative study of tests for the difference of two Poisson means." *Computational Statistics & Data Analysis*, Volume 51, 3085-3099.

O

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.

Obuchowski, N. and Zhou, X.H. 2002. “Prospective studies of diagnostic test accuracy when disease prevalence is low.” *Biostatistics*, 3, pages 477-492.

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).

Ostle, B. and Malone, L. C. 1988. *Statistics in Research. Fourth Edition*. Iowa State Press. Ames, Iowa. A comprehension book on statistical methods.

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.

P

Parmar, M.K.B. and Machin, D. 1995. *Survival Analysis*. John Wiley and Sons. New York.

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.

Pearson, E.S. and Hartley, H.O. 1976. *Biometrika Tables For Statistics, Volume 1*. Biometrika Trust. London.

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.

Piantadosi, S. 2005. *Clinical Trials – A Methodological Perspective*. John Wiley & Sons. New Jersey.

Pocock, S.J. 1977. "Group sequential methods in the design and analysis of clinical trials." *Biometrika*, 64, pages 191-199.

Pocock, S.J. 1983. *Clinical Trials – A Practical Approach*. John Wiley & Sons. New York.

Price, K., Storn R., and Lampinen, J. 2005. *Differential Evolution – A Practical Approach to Global Optimization.* Springer. Berlin, Germany.

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.

R

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.

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.

Rosenberger, W.F., and Lachin, J.M. 2002. *Randomization in Clinical Trials – Theory and Practice*. John Wiley & Sons. New York.

S

Sachs, Lothar. 1984. *Applied Statistics: A Handbook of Techniques*. Springer-Verlag. New York, New York.

Sahai, Hardeo & Khurshid, Anwer. 1995. *Statistics in Epidemiology*. CRC Press. Boca Raton, Florida.

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.

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.

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.

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.

Smith, R.L. 1984. “Sequential Treatment Allocation using Biased Coin Designs.” *Journal of the Royal Statistical Society B*. Volume 46, pages 519-543.

Stekel, D. 2003. *Microarray Bioinformatics.* Cambridge University Press. Cambridge, United Kingdom.

Statxact 5. 2001. *Statistical Software for exact nonparametric inference, user manual.* Cytel Software Corporation. Cambridge, Massachusetts.

Swets, John A. 1996. *Signal Detection Theory and ROC Analysis in Psychology and Diagnostics - Collected Papers.* Lawrence Erlbaum Associates. Mahway, New Jersey.

T

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.

Thode, Henry C. 2002. *Testing for Normality*. Marcel Dekker, Inc. New York.

Thompson, Simon G. 1998. *Encyclopedia of Biostatistics, Volume 4*. John Wiley & Sons. New York. Article on Meta-Analysis on pages 2570-2579.

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.

U

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.

W

Walter, S.D., Eliasziw, M., and Donner, A. 1998. "Sample Size and Optimal Designs For Reliability Studies.” *Statistics in Medicine*, 17, 101-110.

Wei, L.J., and Lachin, J.M. 1988. “Properties of the Urn Randomization in Clinical Trials.” *Controlled Clinical Trials*. Volume 9, pages 345-364.

Welch, B.L. 1938. "The significance of the difference between two means when the population variances are unequal.” *Biometrika*, 29, 350-362.

Westlake, W.J. 1981. “Bioequivalence testing—a need to rethink,” *Biometrics*, Volume 37, pages 591-593.

Whitehead, John. 1993. “Sample Size Calculations for Ordered Categorical Data,” *Statistics in Medicine*, Volume 12, pages 2257-2271.

Whittemore, Alice. 1981. “Sample Size for Logistic Regression with Small Response Probability,” *Journal of the American Statistical Association*, Volume 76, pages 27-32.

Williams, D.A. 1971. “A Test for Differences between Treatment Means When Several Dose Levels are Compared with a Zero Dose Control,” *Biometrics*, Volume 27, No. 1 (March), pages 103-117.

Williams, D.A. 1972. “The Comparison of Several Dose Levels with a Zero Dose Control,” *Biometrics*, Volume 28, No. 2 (June), pages 519-531.

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.

Wolfinger, R., Tobias, R. and Sall, J. 1994. "Computing Gaussian likelihoods and their derivatives for general linear mixed models,” *SIAM Journal of Scientific Computing*, 15, no.6, pages 1294-1310.

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.

Y

Yateman, Nigel A. and Skene, Allan M. 1992. "Sample Sizes for Proportional Hazards Survival Studies with Arbitrary Patient Entry and Loss to Follow-Up Distributions." *Statistics in Medicine*, 11:1103-1113.

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.

Z

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

Index entries are of the form “chapter-page”. A list of chapters is given in the Table of Contents.

A

Abbreviations, 4-10
Absolute deviation
  standard deviation estimator, 905-3
Accrual time
  logrank, 705-6
A-efficiency
  D-optimal designs, 888-13
Agreement between two raters - kappa, 811-1
Alias
  two-level designs, 881-2
Alignment
  histogram, 945-4
  scatter plot, 940-2
All-contrasts power
  simulation, 590-4
All-pair power
  multiple comparisons - simulation, 580-4
  simulation, 585-4
Alpha, 6-3
  adjusting, 8-5
Alpha - simulation
  multiple comparisons, 580-1, 585-2
  multiple contrasts, 590-1
Alpha spending function
  survival - group sequential, 710-1, 711-1
  two means - group sequential, 475-1
  two proportions - group sequential, 220-1
Alternative hypothesis, 6-1, 8-1
Analysis of covariance, 551-1
Analysis of variance
  fixed effects ANOVA, 560-1
  one-way, 550-1
  one-way - simulation, 555-1
  randomized block, 565-1
  repeated measures, 570-1
ANCOVA, 551-1
  assumptions, 551-2
  examples, 551-10
  validation, 551-14
Anderson-Darling test, 670-2
ANOVA
  factorial, 560-1
  fixed effects, 560-1
  multiple comparisons, 575-2
  multiple comparisons - simulation, 580-2
  multiple contrasts - simulation, 590-2
  one-way, 550-1
  one-way - simulation, 555-1
  randomized block, 565-1
  repeated measures, 570-1
  three-way, 560-1
  two-way, 560-1
ANOVA - fixed effects
  examples, 560-13
  validation, 560-19
ANOVA - one-way
  assumptions, 550-2
  examples, 550-10
  validation, 550-19
ANOVA - one-way - simulation
  examples, 555-10
  validation, 555-16
ANOVA - randomized block
  examples, 565-9
  validation, 565-11
ANOVA - repeated measures
  examples, 570-29
  validation, 570-49
Any-contrast power
  simulation, 590-4
Any-pair power
  multiple comparisons - simulation, 580-4
  simulation, 585-4
Assigning subjects to groups, 880-1
AUC
  ROC curve (one), 260-3
  ROC curves (two), 265-1, 265-6
Autocorrelation
  repeated measures ANOVA, 570-25
Average absolute deviation
  standard deviation estimator, 905-3
Axis
  histogram, 945-3
  parameters, 4-10
  scatter plot, 940-1
Axis line
  histogram, 945-5
  scatter plot, 940-3

B

Background color
  histogram, 945-11
  scatter plot, 940-11
Balaam's design, 528-1, 530-1, 533-1, 535-1, 540-1, 545-1
Balance, 880-1
Balanced incomplete block designs, 883-1
  examples, 883-3
Bars
  histogram, 945-1
  scatter plot, 940-11
Beta
  calculating, 6-4
Beta distribution
  probablility calculator, 915-1
  simulation, 920-3
Between standard deviation
  repeated measures ANOVA, 570-14
BIB designs, 883-1
Bimodal data
  simulation, 920-22
Bin width
  histogram, 945-2
Binomial distribution
  probablility calculator, 915-2
  simulation, 920-5
Binomial model
  one proportion, 100-2, 270-2
Binomial probabilities, 6-2
Binomial test
  one proportion, 100-4
Binormal model
  ROC curve (one), 260-2
  ROC curves (two), 265-2
Bins
  histogram, 945-1
Bioequivalence
  two proportions, 215-1
Bivariate normal distribution
  probablility calculator, 915-2
Blackwelder
  risk ratio, 205-26
Blocking
  two-level designs, 881-2
Bonferroni adjustment
  t-test - one group, 610-5
  t-test - two groups, 615-4
Bonferroni test
  multiple contrasts - simulation, 590-3
Bootstrap test
  simulation, 410-3
Bootstrap test - paired means
  simulation, 490-4
Bootstrap test - paired means - equivalence
  simulation, 495-5
Border plots
  histogram, 945-10
Bottom bar
  output window, 5-10
Boundaries
  histogram, 945-3
  scatter plot, 940-1
Box plot
  histogram, 945-11
Box-Behnken designs, 885-1
Brownian motion
  survival - group sequential, 710-5
  two means - group sequential, 475-4
  two proportions - group sequential, 220-4
Buttons
  output window, 5-8
  PASS home window, 3-3
  procedure window, 4-5

C

Candidate points
  D-optimal designs, 888-14
Carryover effects
  repeated measures ANOVA, 570-3
Case-control
  matched, 155-1
Cauchy distribution
  simulation, 920-5
Central-composite designs, 885-1
Chen
  three-stage phase II trials, 130-1
Chi-square
  estimator, 900-1
  one variance, 650-1
Chi-square distribution
  probablility calculator, 915-2
Chi-Square estimator
  examples, 900-2
Chi-square test, 250-1
  examples, 250-6
  two proportions, 200-5, 223-6
  validation, 250-9
Circularity
  repeated measures ANOVA, 570-13
Clinical trial
  three-stage, 130-1
Clinical trials
  one proportion, 120-1
Clopper-Pearson confidence interval, 115-2
Cluster randomized - equivalence
  two proportions, 240-1
Cluster randomized - non-inferiority
  two proportions, 235-1
Cluster randomized - non-zero null
  two proportions, 233-1
Cluster randomized design
  two means, 480-1
  two proportions, 230-1

Cochran-Armitage test, 255-1
examples, 255-12
validation, 255-19
Cochran-Mantel-Haenszel, 225-1
Coefficient alpha
one, 815-1
two, 820-1
Coefficient alpha (one)
examples, 815-4
validation, 815-7
Coefficient alphas (two)
examples, 820-6
validation, 820-10
Coefficient of variation
cross-over, 505-2
cross-over - equivalence, 525-3
cross-over - higher-order - equivalence, 545-5
cross-over - higher-order - non- zero null, 533-4
cross-over - higher-order - non-inferiority, 535-4
cross-over - non-inferiority, 515-3
cross-over - non-unity null, 513-3
mean ratio, 445-2
mean ratio - equivalence, 470-3
mean ratio - non- zero null, 453-2
mean ratio - non-inferiority, 455-2
standard deviation estimator, 905-9
Cohort study
post-marketing surveillance, 135-2
Color
histogram, 945-11
scatter plot, 940-11
Comments
histogram, 945-10
scatter plot, 940-10
Comparisons
ANCOVA, 551-1
multiple comparisons, 575-1
one-way ANOVA, 550-1
Comparisonwise error rate
simulation, 580-1, 585-2, 590-1
Complete randomization, 880-3
Compound symmetry
repeated measures ANOVA, 570-13
Confidence interval
one proportion, 115-1
paired means, 496-1
Confidence intervals
one correlation, 801-1
one mean, 420-1
one mean - tolerance probability, 421-1
one standard deviation, 640-1
one standard deviation - relative error, 642-1
one standard deviation - tolerance probability, 641-1
one variance, 651-1
one variance - relative error, 653-1
one variance - tolerance probability, 652-1
paired means - tolerance probability, 497-1
slope - simple linear regression, 856-1
two means, 471-1
two means - tolerance probability, 472-1
two proportions - difference, 216-1
two proportions - odds ratio, 216-1
two proportions - ratio, 216-1
variance ratio, 656-1
variance ratio - relative error, 657-1
Confounding
two-level designs, 881-2
Connecting points
scatter plot, 940-10
Constant distribution
simulation, 920-6
Consumer’s risk
exponential mean (one), 405-2
Contaminated normal simulation, 920-19
Contingency table, 250-1
chi-square estimator, 900-1
Continuity correction
one proportion, 100-5
two proportions - group sequential, 220-8
two proportions - offset, 205-7
Contrast matrix
repeated measures ANOVA, 570-10
Contrasts
ANCOVA, 551-1
MANOVA, 605-4
mixed models, 201-6, 431-7
one-way ANOVA, 550-1
Contrasts - multiple
simulation, 590-1
Control chart
Cusum, 290-1
EWMA, 290-1
FIR, 290-1
one mean - simulation, 290-1
R Chart, 295-1
S Chart, 295-1
S Chart with Probability Limits, 295-1
Shewhart, 290-1
variability - simulation, 295-1
Control charts
formulas, 290-2, 295-2
Control charts - means
examples, 290-10
validation, 290-14
Control charts - variability
examples, 295-8
validation, 295-11
Controlled variables
multiple regression, 865-4
Corners
histogram, 945-4
scatter plot, 940-1
Correlated proportions
matched case-control, 155-1
McNemar test, 150-1
Correlation
intraclass, 810-1
Correlation (one)
confidence interval, 801-1
examples, 800-4
validation, 800-7

Correlation (one) - confidence interval
examples, 801-4
validation, 801-7
Correlation coefficient, 800-1
Correlation coefficient distribution
probablility calculator, 915-3
Correlation test
one, 800-1
two, 805-1
Correlations (two)
examples, 805-5
validation, 805-8
Covariance matrix
repeated measures ANOVA, 570-13
Covariance patterns
AR(1), 201-11, 431-4
banded, 201-11, 431-4
compound symmetry, 201-11, 431-4
simple, 201-12
simple, 431-5
Cox regression, 850-1
examples, 850-4
validation, 850-6
Creating data
simulation, 920-1
Creating randomization lists, 880-1
Cronbach's alpha
one, 815-1
two, 820-1
Crossed factors
design generator, 889-1
Cross-over
higher-order - equivalence, 540-1, 545-1
higher-order - non-inferiority, 530-1
higher-order - non-zero null, 528-1
ratio - equivalence, 525-1
ratio - higher-order - non-inferiority, 535-1
ratio - higher-order - non-zero null, 533-1
ratio - non-inferiority, 515-1
ratio - non-unity null, 513-1
repeated measures ANOVA, 570-44
two means, 500-1
two means - equivalence, 520-1
two means - non-inferiority, 510-1
two means - non-zero null, 508-1
two means - ratio, 505-1
Cross-over - higher-order - equivalence
examples, 540-9
validation, 540-12
Cross-over - higher-order - non- zero null
examples, 528-8
validation, 528-11
Cross-over - higher-order - non-inferiority
examples, 530-8
validation, 530-11
Cross-over - ratio - equivalence
examples, 525-7
validation, 525-9
Cross-over - ratio - higher-order - equivalence
examples, 545-9
validation, 545-12
Cross-over - ratio - higher-order - non- zero null
examples, 533-9
validation, 533-12
Cross-over - ratio - higher-order - non-inferiority
examples, 535-9
validation, 535-12
Cross-over - ratio - non-inferiority
examples, 515-6
validation, 515-8
Cross-over - ratio - non-unity null
examples, 513-6
Cross-over (two means)
examples, 500-7
validation, 500-10
Cross-over (two means) - equivalence
examples, 520-7
validation, 520-10
Cross-over (two means) - non- unity null
validation, 513-8
Cross-over (two means) - non- zero null
examples, 508-7
validation, 508-10
Cross-over (two means) - non-inferiority
examples, 510-7
validation, 510-10
Cross-over (two means) - ratio
examples, 505-6
validation, 505-8
Crossover analysis
mixed models, 571-1
Cumulative scale
histogram, 945-4
CUSUM chart, 290-4

D

D'Agostino kurtosis, 670-2
D'Agostino omnibus test, 670-3
D'Agostino skewness test, 670-4
Data
histogram, 945-11
scatter plot, 940-11
simulation of, 920-1
Data bars
scatter plot, 940-11
Data entry, 925-1
Data simulation
examples, 920-18
Data simulator, 920-1
Data tab, 4-6
Data ticks
scatter plot, 940-10
Decimals
histogram, 945-6
scatter plot, 940-4
Default template, 4-2
D-efficiency
D-optimal designs, 888-12

Density trace
histogram, 945-3
Design generator, 889-1
examples, 889-3
Design of experiments
randomization lists, 880-1
Designs
Box-Behnken, 885-1
central-composite, 885-1
design generator, 889-1
factorial, 881-2
fractional factorial, 882-1
Plackett-Burman, 886-1
response surface, 885-1
screening, 886-1
Taguchi, 887-1
two-level factorial, 881-1, 889-1
Determinant
D-optimal designs, 888-13
Determinant analysis
D-optimal designs, 888-11
Diagnostic testing
ROC curve (one), 260-1
Difference
proportions, 7-2
Difference data
simulation, 920-25
Discordant pairs
McNemar test, 150-2
Distributions
combining, 920-13
mixing, 920-13
simulation, 920-1
Documentation
printing, 1-7
Donner and Klar
cluster randomized design, 230-2
D-optimal designs, 888-1
examples, 888-8
Dot plot
histogram, 945-11
Downloading updates, 1-2
Drift
survival - group sequential, 710-5
two means - group sequential, 475-4
two proportions - group sequential, 220-4
Dunn’s test
multiple contrasts - simulation, 590-3
Dunnett’s test
multiple comparisons, 575-3
Dunnett's test
simulation, 585-1, 585-3
Dunn's test
simulation, 590-1
Duplicates
D-optimal designs, 888-4

E

Edit menu
output window, 5-5
spreadsheet, 925-2
Effect size, 6-5
ANOVA, 560-2, 560-7
chi-square estimator, 900-1
chi-square test, 250-2
multiple regression, 865-2
one-way ANOVA, 550-3
randomized block ANOVA, 565-2
Efficacy boundaries, 221-1, 222-1, 223-1, 476-1, 477-1, 478-1
Efron’s biased coin randomization, 880-3
Entering procedure options, 4-1
Entering your serial number, 1-2
Equivalence
correlated proportions, 165-1
means (two), 460-1
means (two) - ratio, 470-1
means (two) - simulation, 465-1
one proportion, 110-1
two proportions, 215-1
Equivalence - two correlated proportions
examples, 165-11
validation, 165-14
Equivalence hypothesis, 6-9
Equivalence limits - paired means
simulation, 495-9
Equivalence test data
simulating, 495-3
Error rates - simulation
multiple comparison, 580-1, 585-2
multiple contrasts, 590-1
Errors, 6-2
EWMA chart limits, 290-5
Exact binomial test
one proportion, 100-4
Example data
histogram, 945-11
scatter plot, 940-11
Examples
ANCOVA, 551-10
balanced incomplete block designs, 883-3
Chi-square estimator, 900-2
Chi-square test, 250-6
Cochran-Armitage test, 255-12
control charts - means, 290-10
control charts - variability, 295-8
correlation (one) - confidence interval, 801-4
Cox regression, 850-4
cross-over - higher-order - equivalence, 540-9
cross-over - higher-order - non- zero null, 528-8
cross-over - higher-order - non-inferiority, 530-8
cross-over - ratio - equivalence, 525-7
cross-over - ratio - higher-order - equivalence, 545-9
cross-over - ratio - higher-order - non- zero null, 533-9

cross-over - ratio - higher-order - non-inferiority, 535-9
cross-over - ratio - non-inferiority, 515-6
cross-over - ratio - non-unity null, 513-6
cross-over (two means), 500-7
cross-over (two means) - equivalence, 520-7
cross-over (two means) - non- zero null, 508-7
cross-over (two means) - non-inferiority, 510-7
cross-over (two means) - ratio, 505-6
data simulation, 920-18
design generator, 889-3
D-optimal designs, 888-8
equivalence - two correlated proportions, 165-11
exponential mean (one), 405-8
exponential means (two), 435-5
fixed effects ANOVA, 560-13
fractional factorial designs, 882-3
intraclass correlation, 810-4
kappa, 811-6
Latin square designs, 884-4
linear regression, 855-5
logistic regression, 860-9
logrank tests - non-inferiority, 706-7
logrank tests (Lakatos), 715-15
Mann-Whitney test, 430-16
MANOVA, 605-12
many proportions - trend, 255-12
matched case-control - proportions, 155-6
McNemar test - two correlated proportions, 150-7
mean ratio - non- zero null, 453-6
mean ratio - non-inferiority, 455-6
microarray one-sample or paired t-test, 610-12
microarray two-sample t-test, 615-12
minimum effective dos, 595-4
mixed models, 571-36
multiple comparisons, 575-13
multiple comparisons - simulation, 580-13
multiple comparisons - vs control - simulation, 585-12
multiple contrasts - simulation, 590-14
multiple regression, 865-6
non-inferiority - two correlated proportions, 160-10
normality - simulation, 670-8
odds ratio estimator, 910-2
one coefficient alpha, 815-4
one correlation, 800-4
one mean - confidence interval, 420-5
one mean - tolerance - confidence interval, 421-6
one proportion, 100-13
one proportion - confidence interval, 115-6
one proportion - equivalence, 110-11
one proportion - non-inferiority, 105-10
one proportion - non-zero null, 103-10
one sensitivity, 270-6
one specificity, 270-6
one standard deviation - confidence interval, 640-5
one standard deviation - relative error - confidence interval, 642-4
one standard deviation - tolerance - confidence interval, 641-6
one variance - confidence interval, 651-4
one variance - relative error - confidence interval, 653-4
one variance - tolerance - confidence interval, 652-6
one-way ANOVA, 550-10
one-way ANOVA - simulation, 555-10
paired means - confidence interval, 496-5
paired means - tolerance - confidence interval, 497-6
paired sensitivities, 276-7
Poisson mean (one), 412-4
Poisson means (two), 437-7
Poisson regression, 870-7
post-marketing surveillance, 135-6
proportions estimator, 910-2
randomization lists, 880-9
randomized block ANOVA, 565-9
ratio of two means, 445-6
ratio of two means - equivalence, 470-7
regression - confidence interval, 856-7
repeated measures - two means, 431-13
repeated measures - two proportions, 201-16
repeated measures ANOVA, 570-29
response surface designs, 885-3
ROC curve (one), 260-8
ROC curves (two), 265-8
screening designs, 886-3
single-stage phase II trials, 120-3
standard deviation estimator, 905-6
survival - group sequential, 710-11, 711-27
survival - logrank, 700-6
survival - logrank - advanced, 705-8
Taguchi designs, 887-4
three-stage phase II trials, 130-8
t-test (one mean), 400-8
t-test (one mean) - non- zero null, 413-8
t-test (one mean) - non-inferiority, 415-8
t-test (one mean) - simulation, 410-10
t-test (paired means) - equivalence - simulation, 495-11
t-test (paired means) - simulation, 490-11
t-test (paired), 400-13, 485-8
t-test (two means), 430-10
t-test (two means) - simulation, 440-12
two coefficient alphas, 820-6
two correlated proportions - equivalence, 165-11
two correlated proportions - non-inferiority, 160-10
two correlations, 805-5
two means - cluster randomized, 480-4
two means - confidence interval, 471-5
two means - equivalence, 460-5
two means - equivalence - simulation, 465-12
two means - group sequential, 475-10, 476-17, 477-15
two means - non- zero null, 448-10
two means - non-inferiority, 450-10
two means – non-inferiority – group sequential, 478-17
two means - ratio, 445-6

two means - ratio - equivalence, 470-7
two means - tolerance - confidence interval, 472-6
two ordered categorical variables, 253-5
two proportions - cluster - equivalence, 240-8
two proportions - cluster - non-inferiority, 235-7
two proportions - cluster - non-zero null, 233-7
two proportions - cluster randomized, 230-9
two proportions - confidence interval, 216-26
two proportions - equivalence, 215-17
two proportions - group sequential, 220-9, 221-18
two proportions - inequality, 200-14
two proportions - non-inferiority, 210-17
two proportions – non-inferiority – group sequential, 223-22
two proportions – non-zero null – group sequential, 222-22
two proportions - offset, 205-18
two proportions - stratified design, 225-8
two sensitivities, 275-9
two-level designs, 881-6
two-stage phase II trials, 125-8
variance (one), 650-4
variance ratio - confidence interval, 656-5
variance ratio - relative error - confidence interval, 657-4
variances (two), 655-4
Williams' test, 595-4
Exiting PASS, 4-3
Experiment (run)
two-level designs, 881-2
Experimental design, 881-1
two-level designs, 881-2
Experimental error
two-level designs, 881-2
Experimentwise error rate
simulation, 580-1, 585-2, 590-1
Exponential
logrank, 705-3
log-rank, 715-2
Exponential distribution
simulation, 920-6
Exponential mean (one), 405-1
examples, 405-8
validation, 405-10
Exponential means (two), 435-1
examples, 435-5
validation, 435-7
Exponential test
simulation, 410-4
Exposure
Poisson regression, 870-1
Exposure probability
matched case-control, 155-4

F

F distribution
probablility calculator, 915-3
simulation, 920-7
Factor scaling
D-optimal designs, 888-2
Factorial ANOVA, 560-1
Factorial designs
two-level designs, 881-1, 881-2
False discovery rate adjustment
t-test
two groups, 615-4
t-test - one group, 610-6
Familywise error rate
simulation, 580-1, 585-2, 590-1
Farrington-Manning confidence interval
two proportions, 216-5
Farrington-Manning test
two proportions - equivalence, 215-9
two proportions - non-inferiority, 210-10
two proportions - offset, 205-9
Fast initial response, 290-4
File menu
output window, 5-2
procedure window, 4-3
spreadsheet, 925-2
Finite population correction
t test, 400-8, 485-8
Finite population size
one mean - confidence interval, 420-2
FIR, 290-4
Fisher's exact test
two proportions, 200-4
Fisher-z transformation
two correlations, 805-1
Fixed effects ANOVA
examples, 560-13
validation, 560-19
Fixed effects models
mixed models, 571-1
Fixed factor
ANOVA, 560-5
Fleiss Confidence intervals
two proportions, 216-12, 216-16
Fleming algorithm
single-stage phase II trials, 120-1
Fleming-Harrington tests
Group Sequential, 711-5
Follow-up
logrank, 705-4
Font
histogram, 945-4
scatter plot, 940-2
Forced points
D-optimal designs, 888-5
FPR
ROC curve (one), 260-7
ROC curves (two), 265-1
Fractional factorial designs, 882-1
examples, 882-3
Freedman
logrank, 700-1
Frequency polygon
histogram, 945-3

F-test
Geisser-Greenhouse, 570-4
simulation, 555-1
two variances, 655-1
Futility boundaries, 221-1, 222-1, 223-1, 476-1, 477-1, 478-1

G

Gallery, 5-2
Games-Howell - simulation
multiple comparison, 580-1
Games-Howell test
simulation, 580-3
Gamma data
simulation, 920-23
Gamma distribution
probablility calculator, 915-4
simulation, 920-7
Gart-Nam confidence interval
two proportions, 216-7
Gart-Nam test
two proportions - equivalence, 215-10
two proportions - non-inferiority, 210-11
two proportions - offset, 205-10
Gehan test
Group Sequential, 711-5
Geisser-Greenhouse
repeated measures ANOVA, 570-1
Geisser-Greenhouse F-test
repeated measures ANOVA, 570-4
General linear multivariate model
MANOVA, 605-2
repeated measures ANOVA, 570-3
Generating data, 920-1
Goodness of fit
chi-square estimator, 900-1
Chi-square test, 250-1
Graeco-Latin square designs, 884-1
Grid lines
histogram, 945-7
scatter plot, 940-5
Group sequential test
log-rank, 710-1, 711-1
survival, 710-1, 711-1
two means, 475-1, 476-1, 477-1
two means – non-inferiority, 478-1
two means – non-inferiority (simulation), 478-1
two proportions, 220-1, 221-1
two proportions – non-inferiority, 223-1
two proportions – non-inferiority (simulation), 223-1
two proportions – non-zero null, 222-1
two proportions – non-zero null (simulation), 222-1
two proportions (simulation), 221-1

H

Hazard rate
Cox regression, 850-1
Hazard rate conversion, 903-1
Hazard rate parameterization
logrank, 711-2, 715-2
Hazard rates
logrank, 700-2
logrank - advanced, 705-2
time dependent, 715-22
Hazard ratio
group sequential, 710-2
Help menu
output window, 5-7
procedure window, 4-5
spreadsheet, 925-3
Help system, 1-3
contents window, 1-5
index window, 1-4
navigating, 1-4
printing documentation, 1-7
search window, 1-6
Heterogeneous variances
mixed models, 571-47, 571-49
Histogram
alignment, 945-4
axis, 945-3
axis line, 945-5
background color, 945-11
bars, 945-1
bin width, 945-2
bins, 945-1
border plots, 945-10
boundaries, 945-3
box plot, 945-11
color, 945-11
comments, 945-10
corners, 945-4
cumulative scale, 945-4
data, 945-11
decimals, 945-6
density trace, 945-3
dot plot, 945-11
example data, 945-11
font, 945-4
frequency polygon, 945-3
grid lines, 945-7
interior lines, 945-1
labels, 945-4
layout, 945-4
legend, 945-9
lines, 945-10
major ticks, 945-5
maximum, 945-3
minimum, 945-3
minor ticks, 945-6
normal density, 945-3
notes, 945-10
number bins, 945-2

painting order, 945-11
scale, 945-4
scaling, 945-2
settings, 945-1
size, 945-11
text, 945-10
ticks, 945-5
titles, 945-8
Histograms, 945-1
Hotelling's T2, 600-1
validation, 600-11
Hotelling's T2 distribution
probablility calculator, 915-4
Hotelling-Lawley trace, 605-1
MANOVA, 605-3
repeated measures ANOVA, 570-1, 570-7
Hypergeometric distribution
probablility calculator, 915-4
Hypergeometric model
one proportion, 100-2
Hypothesis
equivalence, 6-9
inequality, 6-7
introduction, 6-1
means, 8-1
non-inferiority, 6-7
superiority, 6-8
types, 6-6
Hypothesis testing
introduction, 6-1

I

Icons
output window, 5-8
procedure window, 4-5
Incidence rate
post-marketing surveillance, 135-2
Inclusion points
D-optimal designs, 888-6
Independence test, 250-1
Inequality
one proportion, 100-1
two means ratio, 445-1
two proportions, 200-1
two proportions - offset, 205-1
Inequality hypothesis, 6-7
Installation, 1-1
Interaction
two-level designs, 881-3
Intercept
linear regression, 855-1
Interim analysis
survival, 710-1, 711-1
three-stage trial, 130-1
two means, 475-1, 476-1, 477-1
two means – non-inferiority, 478-1
two proportions, 220-1, 221-1
two proportions – non-inferiority, 223-1
two proportions – non-zero null, 222-1
Interior lines
histogram, 945-1
Intraclass correlation, 810-1
examples, 810-4
validation, 810-6
Intracluster correlation
cluster randomization - two means, 480-3
cluster randomized - non-inferiority, 235-4
cluster randomized - non-zero null, 233-4
cluster randomized design, 230-2
Introduction to power analysis, 6-1
Iterations
maximum, 4-7
Iterations tab, 4-7

K

Kappa
examples, 811-6
validation, 811-11
Kappa test for rater agreement, 811-1
Kenward and Roger method
mixed models, 571-11
Kolmogorov-Smirnov test, 670-2
Kruskal-Wallis
multiple comparisons - simulation, 580-1
simulation, 555-1
Kruskal-Wallis test
multiple comparisons - simulation, 580-3, 585-1
simulation, 585-3

L

Labels
histogram, 945-4
scatter plot, 940-2
Lachin and Foulkes
logrank test, 705-1
Lakatos
logrank, 715-1
Lan-DeMets
survival - group sequential, 710-1, 711-1
two means - group sequential, 475-1
two proportions - group sequential, 220-1
Latin square
ANOVA, 560-17
Latin square designs, 884-1
examples, 884-4
Layout
histogram, 945-4
scatter plot, 940-2
Legend
histogram, 945-9
parameter, 4-10
scatter plot, 940-8

Likelihood ratio test
two proportions, 200-7
Likert-scale
simulation, 410-21, 920-8
simulation, 920-21
Lilliefors' critical values
normality tests, 670-2
Linear model
ANOVA, 560-2
repeated measures ANOVA, 570-3
Linear regression, 855-1
confidence interval, 856-1
examples, 855-5
validation, 855-7
Lines
histogram, 945-10
scatter plot, 940-10
Loading a procedure, 2-3
Log base
scatter plot, 940-2
Log transformation
cross-over, 505-2
cross-over - equivalence, 525-2
cross-over - higher-order - equivalence, 545-4
cross-over - higher-order - non- zero null, 533-4
cross-over - higher-order - non-inferiority, 535-4
cross-over - non-inferiority, 515-2
cross-over - non-unity null, 513-2
mean ratio, 445-2
mean ratio - equivalence, 470-2
mean ratio - non- zero null, 453-2
mean ratio - non-inferiority, 455-2
Logarithm
scatter plot, 940-1
Logistic regression, 860-1
examples, 860-9
validation, 860-14
Logit
logistic regression, 860-2
Logrank
hazard rate parameterization, 711-2, 715-2
median survival time parameterization, 711-2, 715-2
mortality parameterization, 711-2, 715-2
proportion surviving parameterization, 711-2, 715-2
Log-rank
group sequential test, 710-1, 711-1
Logrank procedure comparison, 715-3
Logrank test, 700-1, 715-1
Lachin and Foulkes, 705-1
non-inferiority, 706-1
Log-rank test
Group Sequential, 711-5
Logrank tests - non-inferiority
examples, 706-7
validation, 706-11
Logrank tests (Lakatos)
examples, 715-15
validation, 715-21
Longitudinal data models
mixed models, 571-2
Longitudinal models, 571-1

M

Macros, 930-1
commands, 930-6
examples, 930-18
syntax, 930-2
Major ticks
histogram, 945-5
scatter plot, 940-4
Mann-Whitney test, 430-1
assumptions, 430-4
examples, 430-16
group-sequential - simulation, 476-4, 477-4, 478-5
non- zero null, 448-5
non-inferiority, 450-5
simulation, 440-5
Mann-Whitney test - equivalence
simulation, 465-5
Mann-Whitney test - simulation
equivalence, 465-1
MANOVA, 605-1
assumptions, 605-1
examples, 605-12
validation, 605-17
Mantel Haenszel test
two proportions, 200-7
Mantel-Haenszel, 225-1
Many proportions - trend
examples, 255-12
validation, 255-19
Martinez-Iglewicz test, 670-3
Matched case-control, 155-1
post-marketing surveillance, 135-2
Matched case-control - proportions
examples, 155-6
validation, 155-10
Maximum
histogram, 945-3
scatter plot, 940-1
Maximum iterations, 4-7
Maximum likelihood
mixed models, 571-8
McNemar test, 150-1
McNemar test - two correlated proportions
examples, 150-7
validation, 150-9
Mean (one)
confidence interval, 420-1
control chart - simulation, 290-1
exponential, 405-1
Poisson, 412-1
simulation, 410-1
Mean (one) - confidence interval
examples, 420-5
validation, 420-7

Mean (one) - tolerance - confidence interval
examples, 421-6
validation, 421-8
Mean (one) - tolerance probability
confidence interval, 421-1
Mean absolute deviation
standard deviation estimator, 905-3
Mean ratio
equivalence, 470-1
inequality, 445-1
Means
introduction, 8-1
one-way - simulation, 555-1
Means - ratio - equivalence
cross-over, 525-1
Means - ratio - non- zero null - higher-order
cross-over, 533-1
Means - ratio - non-inferiority
cross-over, 515-1
Means - ratio - non-inferiority - higher-order
cross-over, 535-1
Means - ratio - non-unity null
cross-over, 513-1
Means (paired)
confidence interval, 496-1
simulation, 490-1
Means (paired) - confidence interval
examples, 496-5
validation, 496-6
Means (paired) - equivalence
simulation, 495-1
Means (paired) - tolerance - confidence interval
examples, 497-6
validation, 497-8
Means (paired) - tolerance probability
confidence interval, 497-1
Means (two)
cluster randomized design, 480-1
confidence interval, 471-1
cross-over, 500-1
equivalence, 460-1
exponential, 435-1
group sequential test, 475-1, 476-1, 477-1
interim analysis, 475-1, 476-1, 477-1
interim analysis – non-inferiority, 478-1
non-inferiority, 450-1
non-zero null, 448-1
Poisson, 437-1
ratio, 445-1
simulation, 440-1
T-test, 430-1
t-test - equivalence, 460-1
T-test - non-inferiority, 450-1
T-test - non-zero null, 448-1
Means (two) - cluster randomized design
examples, 480-4
validation, 480-6
Means (two) - confidence interval
examples, 471-5
validation, 471-7
Means (two) - equivalence
cross-over, 520-1
examples, 460-5
validation, 460-7
Means (two) - equivalence - higher-order
cross-over, 540-1, 545-1
Means (two) - equivalence - simulation
examples, 465-12
validation, 465-19
Means (two) - group sequential
examples, 475-10, 476-17, 477-15
validation, 475-17, 476-28, 477-26
Means (two) – non- inferiority – group sequential
examples, 478-17
validation, 478-28
Means (two) - non- zero null
cross-over, 508-1
examples, 448-10
validation, 448-13
Means (two) - non- zero null - higher-order
cross-over, 528-1
Means (two) - non-inferiority
cross-over, 510-1
examples, 450-10
validation, 450-13
Means (two) - non-inferiority - higher-order
cross-over, 530-1
Means (two) - ratio
cross-over, 505-1
equivalence, 470-1
non- zero null, 453-1
non-inferiority, 455-1
Means (two) - ratio - non- zero null
examples, 453-6
validation, 453-8
Means (two) - ratio - non-inferiority
examples, 455-6
validation, 455-8
Means (two) - simulation
equivalence, 465-1
t-test - equivalence, 465-1
Means (two) - tolerance - confidence interval
examples, 472-6
validation, 472-8
Means (two) - tolerance probability
confidence interval, 472-1
Means matrix
MANOVA, 605-4
repeated measures ANOVA, 570-7
Measurement error
randomized block ANOVA, 565-2
Median survival time conversion, 903-1
Median survival time parameterization
logrank, 711-2, 715-2
Menus
output window, 5-2
PASS home window, 3-2
procedure window, 4-3
spreadsheet, 925-2

Microarray data
one sample t-test, 610-1
paired T-test, 610-1
two sample t-test, 615-1
Microarray one-sample or paired t-test
examples, 610-12
validation, 610-19
Microarray two-sample t-test
examples, 615-12
validation, 615-19
Miettinen-Nurminen confidence interval
two proportions, 216-6
Miettinen-Nurminen test
two proportions - equivalence, 215-7
two proportions - non-inferiority, 210-8
two proportions - offset, 205-7
Minimum
histogram, 945-3
scatter plot, 940-1
Minimum detectable difference
mixed models, 571-14
multiple comparisons, 575-11
one-way ANOVA, 550-17
t-test, 400-12
two-sample t test, 430-14
Minimum effective dose
examples, 595-4
validation, 595-7
Minimum Effective Dose
Williams' Test, 595-1
Minimum system requirements, 1-1
Minor ticks
histogram, 945-6
scatter plot, 940-4
Mixed model
defined, 571-2
Mixed models, 571-1
differential evolution, 571-12
examples, 571-36
F test, 571-11
Fisher scoring, 571-12
fixed effects, 571-1
G matrix, 571-5
heterogeneous variances, 571-47, 571-49
Kenward and Roger method, 571-11
L matrix, 571-9
likelihood formulas, 571-8
maximum likelihood, 571-8
minimum detectable difference, 571-14
MIVQUE, 571-12
Newton-Raphson, 571-12
pairwise contrasts, 201-6, 431-7
R matrix, 571-5
restricted maximum likelihood, 571-9
simulation steps, 571-17
types, 571-1
validation, 571-42
Mixture data
simulation, 920-24
Mixture design
D-optimal designs, 888-20
Monte Carlo, 8-3
Monte Carlo simulation, 920-1
Mortality conversion, 903-1
Mortality parameterization
logrank, 711-2, 715-2
MTBF
exponential mean (one), 405-1
Multinomial
chi-square estimator, 900-3
Multinomial distribution
simulation, 920-8
Multiple comparisons, 575-1
Dunnett's test - simulation, 585-1
examples, 575-13
Games-Howell - simulation, 580-1
pair-wise - simulation, 580-1
validation, 575-19
vs control - simulation, 585-1
with a control, 575-2
with best, 575-5
Multiple comparisons - simulation
examples, 580-13
power, 580-4
validation, 580-22
Multiple comparisons - vs control - simulation
examples, 585-12
validation, 585-21
Multiple contrasts
simulation, 590-1
Multiple contrasts - simulation
examples, 590-14
validation, 590-24
Multiple regression, 865-1
examples, 865-6
validation, 865-10
Multiple testing adjustment
t-test - two groups, 615-2
t-test -one group, 610-4
Multivariate analysis of variance, 605-1

N

Navigating the help system, 1-4
Navigation pane
output window, 5-9
Negative binomial distribution
probablility calculator, 915-5
Nested factors
design generator, 889-1
New template, 4-3
Noncentrality
one-way ANOVA, 550-3
Noncentrality parameter
one-way ANOVA, 550-4
Non-inferiority
correlated proportions, 160-1
logrank, 705-13, 706-1
means (two), 450-1
means (two) - ratio, 455-1

one proportion, 105-1
paired t-test, 415-1
survival, 706-1
two proportions, 210-1
Non-inferiority - two correlated proportions
examples, 160-10
validation, 160-13
Non-inferiority hypothesis, 6-7
Non-Inferiority test (two means)
simulation, 440-17
Nonparametric
t-test, 400-7, 485-7
Nonparametric tests
Wilcoxon test, 400-1
Non-zero null
means (two), 448-1
means (two) - ratio, 453-1
one proportion, 103-1
paired t-test, 413-1
Normal density
histogram, 945-3
Normal distribution
probablility calculator, 915-5
simulation, 920-9, 920-18
Normality - simulation
examples, 670-8
validation, 670-11
Normality tests, 670-1
Anderson-Darling test, 670-2
D'Agostino kurtosis, 670-2
D'Agostino omnibus, 670-3
D'Agostino skewness, 670-4
Kolmogorov-Smirnov, 670-2
Lilliefors' critical values, 670-2
Martinez-Iglewicz, 670-3
range, 670-4
Shapiro-Wilk, 670-4
Notes
histogram, 945-10
scatter plot, 940-10
Nuisance parameter, 8-2
Nuisance parameters, 6-6
Null hypothesis, 6-1, 8-1
Number bins
histogram, 945-2

O

O'Brien-Fleming
survival - group sequential, 710-1, 711-1
two means - group sequential, 475-1
two proportions - group sequential, 220-1
Odds ratio
logistic regression, 860-2
matched case-control, 155-1
McNemar test, 150-3
one proportion, 100-6
proportions, 7-3
two proportions, 200-3
Odds ratio estimator, 910-1
examples, 910-2
One proportion
equivalence, 110-1
inequality, 100-1
non-inferiority, 105-1
non-zero null, 103-1
superiority, 103-1
One-sample t-test
microarray data, 610-1
One-way ANOVA, 550-1
examples, 550-10
simulation, 555-1
validation, 550-19
One-way ANOVA - simulation
examples, 555-10
validation, 555-16
Open example template, 4-3
Open template, 4-3
Operating system, 1-1
Ordered categorical variables (two), 253-1
examples, 253-5
validation, 253-10
Orthogonal arrays, 887-1
Orthogonal sets of Latin squares, 884-2
Outliers
multiple comparisons - simulation, 580-23
one-way ANOVA - simulation, 555-18
Outliers (two means)
simulation, 440-19
Output, 2-5
Output Gallery, 5-2
Output window, 5-1
bottom bar, 5-10
edit menu, 5-5
file menu, 5-2
help menu, 5-7
navigation pane, 5-9
toolbar, 5-8
view menu, 5-4
window menu, 5-7

P

P value, 6-4
Painting order
histogram, 945-11
scatter plot, 940-11
Paired design
microarray data, 610-1
Paired designs
non- zero null, 413-1
non-inferiority, 415-1
Paired distributions
simulating, 490-2
Paired means
simulation, 490-1
Paired means - equivalence
simulation, 495-1

Paired proportions
  equivalence, 165-1
  non-inferiority, 160-1
Paired t-test, 400-1, 485-1
  assumptions, 400-3, 485-3
  microarray data, 610-1
  non- zero null, 413-1
  non-inferiority, 415-1
  superiority, 413-1
Pairwise comparisons
  multiple comparisons, 575-7
Pair-wise comparisons
  simulation, 580-1
Pairwise contrasts
  mixed models, 201-6
Pairwise contrasts
  mixed models, 431-7
Panel, 4-1
Parameters
  abbreviations, 4-10
  axis, 4-10
  entering, 4-6
  legend, 4-10
PASS help system, 1-3
PASS home window, 3-1
  Category Tree, 3-4
  Find, 3-5
  Menus, 3-2
  Procedure list, 3-5
  Search, 3-5
  toolbar, 3-3
Patient entry
  logrank, 705-3
Peto-Peto test
  Group Sequential, 711-5
Phase I trials
  definition, 120-1
Phase II clinical trials
  single-stage one proportion, 120-1
  three-stage one proportion, 130-1
  two-stage one proportion, 125-1
Phase II trials
  definition, 120-1
Pillai-Bartlett trace, 605-1
  MANOVA, 605-3
  repeated measures ANOVA, 570-1, 570-6
Plackett-Burman designs, 886-1
Planned comparisons
  ANCOVA, 551-1
  one-way ANOVA, 550-1
Plot setup tab, 4-9
Pocock
  survival - group sequential, 710-1, 711-1
  two means - group sequential, 475-1
  two proportions - group sequential, 220-1
Poisson distribution
  probablility calculator, 915-5
  simulation, 920-9
Poisson incidence
  post-marketing surveillance, 135-1
Poisson mean (one), 412-1
  examples, 412-4
  validation, 412-9
Poisson means (two), 437-1
  examples, 437-7
  validation, 437-10
Poisson regression, 870-1
  examples, 870-7
  validation, 870-9
Population size
  t-test, 400-8, 485-8
Post-marketing surveillance, 135-1
  examples, 135-6
  validation, 135-8
Power
  calculating, 6-4
  introduction, 6-1
  means, 8-1
  multiple comparisons - simulation, 580-4
Prevalence
  correlated proportions, 160-2, 165-2
Printing documentation, 1-7
Printing output, 5-4, 5-8
Probability calculator, 915-1
  Beta distribution, 915-1
  Binomial distribution, 915-2
  Bivariate normal distribution, 915-2
  Chi-square distribution, 915-2
  Correlation coefficient distribution, 915-3
  F distribution, 915-3
  Gamma distribution, 915-4
  Hotelling's T2 distribution, 915-4
  Hypergeometric distribution, 915-4
  Negative binomial distribution, 915-5
  Normal distribution, 915-5
  Poisson distribution, 915-5
  Student's t distribution, 915-6
  Studentized range distribution, 915-6
  Weibull distribution, 915-6
Procedure menus
  procedure window, 4-4
Procedure options, 4-1
Procedure window, 4-1
  file menu, 4-3
  help menu, 4-5
  procedure menus, 4-4
  run menu, 4-4
  tabs, 4-6
  toolbar, 4-5
  tools menu, 4-4
  view menu, 4-4
  window menu, 4-4
Producer's risk
  exponential mean (one), 405-2
Program updates, 1-2
Proportion (one)
  binomial model, 100-2, 270-2
  confidence interval, 115-1
  continuity correction, 100-5
  equivalence, 110-1
  exact binomial test, 100-4
  examples, 100-13

hypergeometric model, 100-2
inequality, 100-1
non-inferiority, 105-1
odds ratio, 100-6
saw-tooth power function, 100-16
single-stage phase II trials, 120-1
superiority, 103-1
three-stage phase II trials, 130-1
two-stage phase II trials, 125-1
validation, 100-20
Z test, 100-4
Proportion (one) - confidence interval
examples, 115-6
validation, 115-9
Proportion (one) - equivalence
examples, 110-11
validation, 110-17
Z test, 110-6
Proportion (one) - non-inferiority
examples, 105-10
validation, 105-16
Proportion (one) - non-zero null
examples, 103-10
validation, 103-14
Proportion surviving parameterization
logrank, 711-2, 715-2
Proportion trend test, 255-1
Proportional hazards regression, 850-1
Proportions
comparing, 7-1
difference, 7-2
interpretation, 7-3
introduction, 7-1
logistic regression, 860-1
odds ratio, 7-3
odds ratio estimator, 910-1
paired (equivalence), 165-1
paired (non-inferiority), 160-1
ratio, 7-2
Proportions (many)
trend, 255-1
Proportions (many) - trend
examples, 255-12
validation, 255-19
Proportions (multiple)
Chi-square test, 250-1
Proportions (two)
Chi-square test, 200-5, 223-6
cluster randomized - equivalence, 240-1
cluster randomized - non-inferiority, 235-1
cluster randomized – non-zero null, 233-1
cluster randomized design, 230-1
confidence intervals, 216-1
equivalence, 215-1
Fisher's exact, 200-4
group sequential test, 220-1, 221-1
group sequential test – non-inferiority, 223-1, 478-1
group sequential test – non-zero null, 222-1
independent, 200-1
independent - equivalence, 215-1
independent - non-inferiority & superiority, 210-1
independent - offset, 205-1
inequality, 200-1
inequality - offset, 205-1
interim analysis, 220-1, 221-1
interim analysis – non-inferiority, 223-1
interim analysis – non-zero null, 222-1
matched case control, 155-1
McNemar test, 150-1
non-inferiority & superiority, 210-1
stratified, 225-1
Proportions (two) - cluster - equivalence
examples, 240-8
validation, 240-12
Proportions (two) - cluster - non-inferiority
examples, 235-7
validation, 235-11
Proportions (two) - cluster - non-zero null
examples, 233-7
validation, 233-11
Proportions (two) - cluster randomized design
examples, 230-9
validation, 230-17
Proportions (two) - confidence interval
examples, 216-26
validation, 216-28
Proportions (two) - equivalence
examples, 215-17
validation, 215-23
Proportions (two) - group sequential
examples, 220-9, 221-18
validation, 220-17, 221-29
Proportions (two) - inequality
examples, 200-14
validation, 200-19
Proportions (two) – non- inferiority – group sequential
examples, 223-22
validation, 223-33
Proportions (two) - non-inferiority
examples, 210-17
validation, 210-24
Proportions (two) – non-zero null – group sequential
examples, 222-22
validation, 222-33
Proportions (two) - offset
examples, 205-18
validation, 205-24
Proportions (two) - stratified design
examples, 225-8
validation, 225-11
Proportions (two) correlated
equivalence, 165-1
non-inferiority, 160-1
Proportions estimator, 910-1
examples, 910-2

Q

Qualitative factors
  D-optimal designs, 888-5
Quitting PASS, 4-3

R

Random assignment, 880-1
Random effects models, 571-1
  mixed models, 571-2
Random factor
  ANOVA, 560-5
Random numbers, 920-1
  multiple comparisons - simulation, 580-5
  simulation, 585-5, 590-5
Random sorting, 880-3
Random sorting using max % deviation, 880-5
Randomization
  complete, 880-3
  Efron's biased coin, 880-3
  Latin square designs, 884-2
  random sorting, 880-3
  random sorting using max % deviation, 880-5
  Smith, 880-4
  Wei's urn, 880-4
Randomization lists, 880-1
  examples, 880-9
Randomized block ANOVA, 565-1
  examples, 565-9
  validation, 565-11
Randomized block design
  design generator, 889-5
Range test, 670-4
Rater agreement - kappa, 811-1
Rating data
  ROC curve (one), 260-4
  ROC curves (two), 265-3
Ratio
  proportions, 7-2
  two means, 445-1
  two means - equivalence, 470-1
Ratio (two means)
  non- zero null, 453-1
  non-inferiority, 455-1
Ratio of two means
  examples, 445-6
  validation, 445-8
Ratio of two means - equivalence
  examples, 470-7
  validation, 470-9
Ratio -two means
  cross-over, 505-1
Regression
  confidence interval, 856-1
  Cox, 850-1
  linear, 855-1
  logistic, 860-1
  multiple, 865-1
  Poisson, 870-1
Regression - confidence interval
  examples, 856-7
  validation, 856-10
Rejection region, 6-3
Repeated measures
  mixed models, 571-1
  two means, 431-1
  two proportions, 201-1
Repeated measures - two means
  examples, 431-13
  validation, 431-19
Repeated measures - two proportions
  examples, 201-16
  validation, 201-24
Repeated measures ANOVA, 570-1
  examples, 570-29
  validation, 570-49
Repeated measures design
  design generator, 889-6
Replication
  two-level designs, 881-3
Reports tab, 4-8
Requirements, 1-1
Resetting a template, 4-3
Response surface designs, 885-1
  examples, 885-3
Restricted maximum likelihood
  mixed models, 571-9
Risk ratio
  Blackwelder, 205-26
ROC curve (one), 260-1
  examples, 260-8
  validation, 260-12
ROC curves (two), 265-1
  examples, 265-8
  validation, 265-13
R-squared
  logistic regression, 860-7
  multiple regression, 865-1
RTF files, 5-1
Ruler, 5-4
Run menu
  procedure window, 4-4
Running PASS, 2-1

S

Sample size
  introduction, 6-1
Save template, 4-3
Saw-tooth power function
  one proportion, 100-16
Scale
  histogram, 945-4
  scatter plot, 940-1
Scaling
  histogram, 945-2

Scaling factors
D-optimal designs, 888-2
Scatter plot
alignment, 940-2
axis, 940-1
axis line, 940-3
background color, 940-11
bars, 940-11
boundaries, 940-1
color, 940-11
comments, 940-10
connecting points, 940-10
corners, 940-1
data, 940-11
data bars, 940-11
data ticks, 940-10
decimals, 940-4
example data, 940-11
font, 940-2
grid lines, 940-5
labels, 940-2
layout, 940-2
legend, 940-8
lines, 940-10
log base, 940-2
logarithm, 940-1
major ticks, 940-4
maximum, 940-1
minimum, 940-1
minor ticks, 940-4
notes, 940-10
painting order, 940-11
scale, 940-1
settings, 940-1
size, 940-11
symbols, 940-6, 940-10
text, 940-10
ticks, 940-3, 940-10
titles, 940-7
trend lines, 940-10
Score test
two proportions - equivalence, 215-7
two proportions - non-inferiority, 210-8
two proportions - offset, 205-7
Screening designs, 886-1
examples, 886-3
Search
PASS home window, 3-5
Sensitivities (paired)
examples, 276-7
validation, 276-11
Sensitivities (two)
examples, 275-9
validation, 275-13
Sensitivity
correlated proportions, 160-2, 165-2
one-sample, 270-1
paired, 276-1
ROC curve (one), 260-1
two-sample, 275-1
Sensitivity (one)
examples, 270-6
validation, 270-11
Serial number, 1-2
Settings
histogram, 945-1
Shapiro-Wilk test, 670-4
Sign test
simulation, 410-3
Sign test - paired means
simulation, 490-4
Sign test - paired means - equivalence
simulation, 495-5
Significance boundaries, 221-1, 222-1, 223-1, 476-1, 477-1, 478-1
Significance level, 6-3
adjusting, 8-5
Significance level - simulation
multiple comparisons, 580-1, 585-2
multiple contrasts, 590-1
Simon
two-stage phase II trials, 125-1
Simple linear regression
confidence interval, 856-1
Simulation, 8-3, 920-1
Beta distribution, 920-3
Binomial distribution, 920-5
Cauchy distribution, 920-5
Constant distribution, 920-6
contaminated normal, 920-19
Exponential distribution, 920-6
F distribution, 920-7
Gamma distribution, 920-7
Likert-scale, 920-8, 920-21
Multinomial distribution, 920-8
multiple comparisons, 580-1, 580-5, 585-1, 585-4
multiple contrasts, 590-1, 590-4
Normal distribution, 920-9, 920-18
normality tests, 670-1
one mean, 410-1
one mean – control chart, 290-1
one-way ANOVA, 555-1
paired means, 490-1
paired means - equivalence, 495-1
Poisson distribution, 920-9
random number generation, 580-5, 585-5, 590-5
size, 8-3
skewed distribution, 920-10
Student's T distribution, 920-10
syntax, 920-13
T distribution, 920-10
Tukey's lambda distribution, 920-10
two means, 440-1
two means - equivalence, 465-1
Uniform distribution, 920-11
variability – control chart, 295-1
Weibull distribution, 920-12
Simulation steps
mixed models, 571-17
Single-stage design
one proportion, 120-1

Single-stage phase II trials
examples, 120-3
validation, 120-3
Size
histogram, 945-11
scatter plot, 940-11
Skewed data
one-way ANOVA - simulation, 555-20
simulation, 410-12
Skewed data (two means)
simulation, 440-20
Skewed distribution
simulation, 920-10
Skewness test, 670-4
Slope
linear regression, 855-1
Slope - simple linear regression
confidence interval, 856-1
Smith's randomization, 880-4
Specificity
correlated proportions, 160-2, 165-2
one-sample, 270-1
ROC curve (one), 260-1
Specificity (one)
examples, 270-6
validation, 270-11
Spending functions
survival - group sequential, 710-3
two means - group sequential, 475-2, 476-5, 477-4
two proportions - group sequential, 220-2, 221-7
two proportions - group sequential – non-zero null, 222-10, 223-10
Split plot analysis
mixed models, 571-1
Spreadsheet, 925-1
edit menu, 925-2
file menu, 925-2
help menu, 925-3
menus, 925-2
Standard deviation, 8-2
estimator, 905-1
interpretation, 905-1
Standard deviation (one)
confidence interval, 640-1
Standard deviation (one) - confidence interval
examples, 640-5
validation, 640-7
Standard deviation (one) - relative error
confidence interval, 642-1
Standard deviation (one) - relative error - confidence interval
examples, 642-4
validation, 642-6
Standard deviation (one) - tolerance - confidence interval
examples, 641-6
validation, 641-8
Standard deviation (one) - tolerance probability
confidence interval, 641-1
Standard deviation estimator
examples, 905-6
Standard deviation of means estimator, 912-1
Standard deviation test
one, 650-1
two, 655-1
Starting PASS, 1-2, 2-1
Stratified designs
two proportions, 225-1
Student's t distribution
probablility calculator, 915-6
Studentized range distribution
probablility calculator, 915-6
Student's T distribution
simulation, 920-10
Superiority
one proportion, 103-1
paired t-test, 413-1
two proportions, 210-1
Superiority hypothesis, 6-8
Survival
logrank, 700-1, 715-1
logrank - Lachin and Foulkes, 705-1
non-inferiority, 706-1
Survival - group sequential
examples, 710-11, 711-27
validation, 710-18, 711-40
Survival - logrank
examples, 700-6
validation, 700-9
Survival - logrank - Lachin and Foulkes
examples, 705-8
validation, 705-12
Survival Parameter Conversion Tool, 903-1
Symbols
scatter plot, 940-6, 940-10
Syntax
macros, 930-2
System requirements, 1-1

T

T distribution
simulation, 920-10
T test
two proportions - equivalence, 215-7
two proportions - non-inferiority, 210-8
T2
Hotelling's, 600-1
Tabs
data, 4-6
iterations, 4-7
plot setup, 4-9
reports, 4-8
Tabs on the procedure window, 4-6
Taguchi designs, 887-1
examples, 887-4
Tarone-Ware test
Group Sequential, 711-5
Technical support, 1-9

Templates, 4-1
automatic, 4-2
creating a new, 4-3
default, 4-2
loading, 4-3
loading examples, 4-3
opening, 4-3
opening examples, 4-3
saving, 4-3
Test statistics, 6-6
Text
histogram, 945-10
scatter plot, 940-10
Three-stage design
one proportion, 130-1
Three-stage phase II trials
examples, 130-8
validation, 130-8
Ticks
histogram, 945-5
scatter plot, 940-3, 940-10
Time averaged difference
binary data, 201-1
normal data, 431-1
power for, 201-2, 431-2
two means, 431-1
two proportions, 201-1
Titles
histogram, 945-8
scatter plot, 940-7
Toolbar
output window, 5-8
PASS home window, 3-3
procedure window, 4-5
Tools menu
procedure window, 4-4
Trend in proportions, 255-1
Trend lines
scatter plot, 940-10
Trimmed t-test (two means)
simulation, 440-4
Trimmed t-test (two means) - equivalence
simulation, 465-4
T-test
cross-over, 500-3
microarray data, 610-1, 615-1
one group - multiple testing adjustment, 610-4
one mean, 400-1
one mean - simulation, 410-1
paired, 400-1, 485-1
paired - equivalence - simulation, 495-1
paired - simulation, 490-1
two groups - multiple testing adjustment, 615-2
two means, 430-1
two means - equivalence, 460-1
two means - non-inferiority, 450-1
two means - non-zero null, 448-1
two means - simulation, 440-1
two proportions, 200-8, 223-6
T-test - equivalence
cross-over, 520-3
T-test - simulation
two means - equivalence, 465-1
T-test (one mean)
assumptions, 400-3
examples, 400-8
non-inferiority, 415-1
non-zero null, 413-1
superiority, 413-1
validation, 400-17
T-test (one mean) - non- zero null
examples, 413-8
validation, 413-10
T-test (one mean) - non-inferiority
examples, 415-8
validation, 415-11
T-test (one mean) - simulation
examples, 410-10
validation, 410-16
T-test (paired means) - equivalence - simulation
examples, 495-11
validation, 495-17
T-test (paired means) - simulation
examples, 490-11
validation, 490-16
T-test (paired)
examples, 400-13, 485-8
validation, 485-10
T-test (two means)
assumptions, 430-3
examples, 430-10
validation, 430-18
T-test (two means) - simulation
examples, 440-12
validation, 440-16
Tukey-Kramer
simulation, 580-1
Tukey-Kramer test
multiple comparisons, 575-7
simulation, 580-3
Tukey's lambda distribution
simulation, 920-10
Two correlated proportions - equivalence
examples, 165-11
validation, 165-14
Two correlated proportions - non-inferiority
examples, 160-10
validation, 160-13
Two means
cluster randomized design, 480-1
cross-over, 500-1
Two means - cluster randomized design
examples, 480-4
validation, 480-6
Two means - equivalence
cross-over, 520-1
examples, 460-5
validation, 460-7
Two means - equivalence - higher-order
cross-over, 540-1, 545-1

Two means - equivalence - simulation
examples, 465-12
validation, 465-19
Two means - group sequential
examples, 475-10, 476-17, 477-15
validation, 475-17, 476-28, 477-26
Two means – non- inferiority – group sequential
examples, 478-17
validation, 478-28
Two means - non- zero null
cross-over, 508-1
examples, 448-10
validation, 448-13
Two means - non- zero null - higher-order
cross-over, 528-1
Two means - non-inferiority
cross-over, 510-1
examples, 450-10
validation, 450-13
Two means - non-inferiority - higher-order
cross-over, 530-1
Two means - ratio
cross-over, 505-1
examples, 445-6
validation, 445-8
Two means - ratio - equivalence
cross-over, 525-1
examples, 470-7
validation, 470-9
Two means - ratio - non- zero null - higher-order
cross-over, 533-1
Two means - ratio - non-inferiority
cross-over, 515-1
Two means - ratio - non-inferiority - higher-order
cross-over, 535-1
Two means - ratio - non-unity null
cross-over, 513-1
Two means (ratio) - non- zero null
examples, 453-6
validation, 453-8
Two means (ratio) - non-inferiority
examples, 455-6
validation, 455-8
Two ordered categorical variables, 253-1
examples, 253-5
validation, 253-10
Two proportions
cluster randomized - equivalence, 240-1
cluster randomized - non-inferiority, 235-1
cluster randomized - non-zero null, 233-1
cluster randomized design, 230-1
equivalence, 215-1
inequality, 200-1
inequality - offset, 205-1
matched case-control, 155-1
McNemar test, 150-1
non-inferiority & superiority, 210-1
Two proportions - cluster - equivalence
examples, 240-8
validation, 240-12
Two proportions - cluster - non-inferiority
examples, 235-7
validation, 235-11
Two proportions - cluster - non-zero null
examples, 233-7
validation, 233-11
Two proportions - cluster randomized design
examples, 230-9
validation, 230-17
Two proportions - equivalence
examples, 215-17
validation, 215-23
Two proportions - group sequential
examples, 220-9, 221-18
validation, 220-17, 221-29
Two proportions - inequality
examples, 200-14
validation, 200-19
Two proportions – non- inferiority – group sequential
examples, 223-22
validation, 223-33
Two proportions - non-inferiority
examples, 210-17
validation, 210-24
Two proportions – non-zero null – group sequential
examples, 222-22
validation, 222-33
Two proportions - offset
examples, 205-18
validation, 205-24
Two proportions - stratified design
examples, 225-8
validation, 225-11
Two-channel arrays, 610-1
Two-level designs, 881-1
examples, 881-6
Two-level factorial designs, 881-1
Two-sample t-test, 430-1
equivalence, 460-1
microarray data, 615-1
non-inferiority, 450-1
non-zero null, 448-1
Two-sample t-test - simulation
equivalence, 465-1
Two-stage design
one proportion, 125-1
Two-stage phase II trials
examples, 125-8
validation, 125-10
Type-I error, 6-2
Type-II error, 6-2

U

Uniform distribution
simulation, 920-11
Updates, 1-2

V

Validation
ANCOVA, 551-14
Chi-square test, 250-9
Cochran-Armitage test, 255-19
control charts - means, 290-14
control charts - variability, 295-11
correlation (one) - confidence interval, 801-7
Cox regression, 850-6
cross-over - higher-order - equivalence, 540-12
cross-over - higher-order - non- zero null, 528-11
cross-over - higher-order - non-inferiority, 530-11
cross-over - ratio - equivalence, 525-9
cross-over - ratio - higher-order - equivalence, 545-12
cross-over - ratio - higher-order - non- zero null, 533-12
cross-over - ratio - higher-order - non-inferiority, 535-12
cross-over - ratio - non-inferiority, 515-8
cross-over (two means), 500-10
cross-over (two means) - equivalence, 520-10
cross-over (two means) - non- unity null, 513-8
cross-over (two means) - non- zero null, 508-10
cross-over (two means) - non-inferiority, 510-10
cross-over (two means) - ratio, 505-8
equivalence - two correlated proportions, 165-14
exponential mean (one), 405-10
exponential means (two), 435-7
fixed effects ANOVA, 560-19
Hotelling's T2, 600-11
intraclass correlation, 810-6
kappa, 811-11
linear regression, 855-7
logistic regression, 860-14
logrank tests - non-inferiority, 706-11
logrank tests (Lakatos), 715-21
MANOVA, 605-17
many proportions - trend, 255-19
matched case-control - proportions, 155-10
McNemar test - two correlated proportions, 150-9
mean ratio - non- zero null, 453-8
mean ratio - non-inferiority, 455-8
microarray one-sample or paired t-test, 610-19
microarray two-sample t-test, 615-19
minimum effective dos, 595-7
mixed models, 571-42
multiple comparisons, 575-19
multiple comparisons - simulation, 580-22
multiple comparisons - vs control - simulation, 585-21
multiple contrasts - simulation, 590-24
multiple regression, 865-10
non-inferiority - two correlated proportions, 160-13
normality - simulation, 670-11
one coefficient alpha, 815-7
one correlation, 800-7
one mean - confidence interval, 420-7
one mean - tolerance - confidence interval, 421-8
one proportion, 100-20
one proportion - confidence interval, 115-9
one proportion - equivalence, 110-17
one proportion - non-inferiority, 105-16
one proportion - non-zero null, 103-14
one sensitivity, 270-11
one specificity, 270-11
one standard deviation - confidence interval, 640-7
one standard deviation - relative error - confidence interval, 642-6
one standard deviation - tolerance - confidence interval, 641-8
one variance - confidence interval, 651-6
one variance - relative error - confidence interval, 653-6
one variance - tolerance - confidence interval, 652-8
one-way ANOVA, 550-19
one-way ANOVA - simulation, 555-16
paired means - confidence interval, 496-6
paired means - tolerance - confidence interval, 497-8
paired sensitivities, 276-11
Poisson mean (one), 412-9
Poisson means (two), 437-10
Poisson regression, 870-9
post-marketing surveillance, 135-8
randomized block ANOVA, 565-11
ratio of two means, 445-8
ratio of two means - equivalence, 470-9
regression - confidence interval, 856-10
repeated measures - two means, 431-19
repeated measures - two proportions, 201-24
repeated measures ANOVA, 570-49
ROC curve (one), 260-12
ROC curves (two), 265-13
single-stage phase II trials, 120-3
survival - group sequential, 710-18, 711-40
survival - logrank, 700-9
survival - logrank - Lachin and Foulkes, 705-12
three-stage phase II trials, 130-8
t-test (one mean), 400-17
t-test (one mean) - non- zero null, 413-10
t-test (one mean) - non-inferiority, 415-11
t-test (one mean) - simulation, 410-16
t-test (paired means) - equivalence - simulation, 495-17
t-test (paired means) - simulation, 490-16
t-test (paired), 485-10
t-test (two means), 430-18
t-test (two means) - simulation, 440-16
two coefficient alphas, 820-10
two correlated proportions - equivalence, 165-14
two correlated proportions - non-inferiority, 160-13
two correlations, 805-8
two means - cluster randomized, 480-6
two means - confidence interval, 471-7
two means - equivalence, 460-7
two means - equivalence - simulation, 465-19

two means - group sequential, 475-17, 476-28, 477-26
two means - non- zero null, 448-13
two means - non-inferiority, 450-13
two means – non-inferiority – group sequential, 478-28
two means - ratio, 445-8
two means - ratio - equivalence, 470-9
two means - tolerance - confidence interval, 472-8
two ordered categorical variables, 253-10
two proportions - cluster - equivalence, 240-12
two proportions - cluster - non-inferiority, 235-11
two proportions - cluster - non-zero null, 233-11
two proportions - cluster randomized, 230-17
two proportions - confidence interval, 216-28
two proportions - equivalence, 215-23
two proportions - group sequential, 220-17, 221-29
two proportions - inequality, 200-19
two proportions – non- inferiority – group sequential, 223-33
two proportions - non-inferiority, 210-24
two proportions – non-zero null – group sequential, 222-33
two proportions - offset, 205-24
two proportions - stratified design, 225-11
two sensitivities, 275-13
two-stage phase II trials, 125-10
variance (one), 650-7
variance ratio - confidence interval, 656-8
variance ratio - relative error - confidence interval, 657-6
variances (two), 655-8
Williams' test, 595-7
Variance (one)
confidence interval, 651-1
examples, 650-4
validation, 650-7
Variance (one) - confidence interval
examples, 651-4
validation, 651-6
Variance (one) - relative error
confidence interval, 653-1
Variance (one) - relative error - confidence interval
examples, 653-4
validation, 653-6
Variance (one) - tolerance - confidence interval
examples, 652-6
validation, 652-8
Variance (one) - tolerance probability
confidence interval, 652-1
Variance ratio
confidence interval, 656-1
Variance ratio - confidence interval
examples, 656-5
validation, 656-8
Variance ratio - relative error
confidence interval, 657-1
Variance ratio - relative error - confidence interval
examples, 657-4
validation, 657-6
Variance test
one, 650-1
two, 655-1
Variances (two)
examples, 655-4
validation, 655-8
View menu
output window, 5-4
procedure window, 4-4
Viewing output, 2-5

W

Walter's confidence intervals
two proportions, 216-11
Wei's urn randomization, 880-4
Weibull distribution
probablility calculator, 915-6
simulation, 920-12
Welch test
multiple contrasts - simulation, 590-4
Welch's test - simulation
equivalence, 465-1
Welch's t-test
non- zero null, 448-4
non-inferiority, 450-4
simulation, 440-3
Wilcoxon test, 400-1, 400-7, 413-7, 415-7, 448-9, 450-9, 485-7
assumptions, 400-3, 485-3
non- zero null, 413-1
non-inferiority, 415-1
simulation, 410-3
superiority, 413-1
Wilcoxon test - paired means
simulation, 490-4
Wilcoxon test - paired means - equivalence
simulation, 495-4
Wilks' Lambda, 605-1
MANOVA, 605-2
repeated measures ANOVA, 570-1, 570-6
Williams test
examples, 595-4
validation, 595-7
Williams' test, 595-1
Wilson score limits
one proportion, 115-2
Wilson's score confidence interval
two proportions, 216-7
Window menu
output window, 5-7
procedure window, 4-4
Winsorized test (two means) - equivalence
simulation, 465-4
Within standard deviation
repeated measures ANOVA, 570-15
Within-subjects design
repeated measures ANOVA, 570-3

Z

Z test
one proportion, 100-4
one proportion - equivalence, 110-6
two proportions - equivalence, 215-5
two proportions - non-inferiority, 210-6
two proportions - offset, 205-6