STATISTICAL SOLUTIONS

nTerim

Version 4.0

User Manual

nTerim 4.0
User Manual

Statistical Solutions Ltd.

4500 Airport Business Park,
Cork,
Ireland
Web: www.statsols.com
Email: sales@statsols.com
Tel: +353 21 4839100
Fax: +353 21 4840026

One International Place,
100 Oliver Street, Suite 1400,
Boston, MA 02110
Web: www.statsols.com
Email: sales@statsols.com
Tel: (617) 535 7677
Fax: (617) 535 7717

Printed in the Republic of Ireland
No part of this manual may be reproduced, stored in a retrieval system, transmitted, translated into any other language, or distributed in any form by any means without prior permission of Statistical Solutions Ltd.

Statistical Solutions Ltd. nTerim® License Agreement

IMPORTANT: READ BEFORE PROCEEDING WITH INSTALLATION

THIS DOCUMENT SETS FORTH THE TERMS AND CONDITIONS OF THE LICENSE AND THE LIMITED WARRANTY FOR nTerim®. PROCEEDING WITH THIS INSTALLATION CONSTITUTES YOUR ACCEPTANCE OF THIS LICENSE AGREEMENT WITH RESPECT TO ALL ACCOMPANYING nTerim® SOFTWARE RECEIVED BY YOU. IF YOU DO NOT ACCEPT THIS AGREEMENT, YOU MAY RETURN THIS SOFTWARE UNDAMAGED WITHIN 10 DAYS OF RECEIPT AND YOUR MONEY WILL BE REFUNDED.

1. GRANT OF LICENSE.

In consideration of payment of the license fee, which is part of the price you paid for this product, Statistical Solutions Ltd., as LICENSOR, grants to you, the LICENSEE, a non-exclusive right to use this copy of nTerim ("SOFTWARE") on a single COMPUTER (i.e., with a single CPU) at a single location. THIS LICENSE SHALL NOT APPLY TO AND DOES NOT PERMIT THE ELECTRONIC TRANSFER OF THE SOFTWARE FROM ONE COMPUTER TO ANOTHER unless a Network Addendum to the Agreement is executed by Licensee and returned to LICENSOR. Licensor reserves all rights not expressly granted to LICENSEE. LICENSOR also agrees to provide free maintenance of the SOFTWARE for sixty (60) days.

2. TRIAL PERIOD.

LICENSEE shall have sixty (60) days, commencing on day of receipt by LICENSEE, in which to return the SOFTWARE provided hereunder and shall be entitled to receive a full refund. All refunds are contingent upon receipt of LICENSOR, in undamaged condition, of all materials provided hereunder.

3. OWNERSHIP OF SOFTWARE.

LICENSOR retains title to and ownership of the SOFTWARE. This LICENSE is not a sale of the original SOFTWARE or any copy.

4. COPY RESTRICTIONS.

This SOFTWARE and the accompanying written materials are copyrighted. Unauthorised copying of the SOFTWARE, including SOFTWARE which has been modified, merged, or included with other software, or of the written materials is expressly forbidden. You may be held legally responsible for any copyright infringement that is caused or encouraged by your failure to abide by the terms of the LICENSE. Subject to these restrictions, you may make one (1) copy of the SOFTWARE solely for backup purposes. You may reproduce and include the copyright notice on the backup copy.

5. USE RESTRICTIONS.

As the LICENSEE, you may physically transfer the SOFTWARE from one computer to another provided that the SOFTWARE is used on only one computer at a time. You may not translate, reverse engineer, decompile or disassemble the software. You may not distribute copies of the SOFTWARE or accompanying written materials to others.

6. TRANSFER RESTRICTIONS.

This SOFTWARE is licensed only to you, the LICENSEE and may not be transferred to anyone without the prior written consent of LICENSOR. Any authorised transferee of the SOFTWARE shall be bound by the terms and conditions of this Agreement.

7. TERMINATION.

This LICENSE is effective until terminated. This LICENSE will terminate automatically without notice from LICENSOR if you fail to comply with any provision of this LICENSE. Upon termination, you shall destroy the written materials and all copies of the SOFTWARE, including modified copies, if any, and shall notify LICENSOR of same.

8. GOVERNING LAW, MISCELLANEOUS.
This Agreement is governed by the laws of Ireland. If any of the provisions or portions thereof of this License Agreement are invalid under any applicable statute or rule of law, they are to that extent to be deemed omitted.

9. DECISION OF ARBITRATORS.
At the option of the LICENSOR, any dispute or controversy shall be finally resolved in accordance with the rules of the International Chamber of Commerce. The Arbitration shall be conducted in Ireland with 3 Arbitrators (unless Licensor opts for 1 Arbitrator).

10. LIMITED WARRANTY AND REMEDY.
LICENSOR warrants that for ninety (90) days after purchase of SOFTWARE by the LICENSEE, SOFTWARE shall reasonably perform in accordance with the accompanying documentation (manuals and computer aided instructions). LICENSEE'S SOLE AND EXCLUSIVE REMEDY for a failure of SOFTWARE to perform in accordance with accompanying documentation is for LICENSEE to return SOFTWARE to LICENSOR and to notify LICENSOR in writing of the non-performance within ninety (90) days of purchase of the media. LICENSOR's sole obligation within a reasonable time after receiving notification of non-performance shall be, at LICENSOR's option to (i) correct demonstrable and significant program errors, (ii) provide the LICENSEE with functionally equivalent software, or (iii) provide or authorise a refund.

11. DISCLAIMER OF OTHER WARRANTIES.
THE LICENSOR'S LIMITED WARRANTY SET FORTH IN PARAGRAPH 10 IS IN LIEU OF ANY OTHER WARRANTIES. LICENSOR DISCLAIMS ANY IMPLIED WARRANTY INCLUDING, BUT NOT LIMITED TO ANY IMPLIED WARRANTY OF FITNESS FOR A PARTICULAR PURPOSE OR MERCHANTABILITY. LICENSOR DISCLAIMS ALL OTHER WARRANTIES INCLUDING, BUT NOT LIMITED TO , THOSE MADE BY DISTRIBUTORS AND RETAILERS OF LICENSOR'S SOFTWARE. THIS LICENSE AGREEMENT GIVES YOU SPECIFIC LEGAL RIGHTS. YOU MAY HAVE OTHER RIGHTS, WHICH VARY FROM STATE TO STATE.

12. LICENSEE RESPONSIBILITY FOR SELECTION OF SOFTWARE.
LICENSEE is solely responsible for selecting of SOFTWARE to achieve the LICENSEE'S intended results or for particular applications.

13. DISCLAIMER.
IN NO EVENT SHALL LICENSOR OR ITS SUPPLIERS BE LIABLE TO LICENSEE FOR ANY SPECIAL INDIRECT, INCIDENTAL OR CONSEQUENTIAL DAMAGES IN ANY WAY RELATING TO THE USE OR ARISING OUT OT THE USE OF SOFTWARE, EVEN IF LICENSOR HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. LICENSOR'S LIABILITY SHALL IN NO EVENT EXCEED THE TOTAL AMOUNT OF THE PURCHASE PRICE/LICENSEE FEE ACTUALLY PAID BY THE LICENSEE FOR THE USE OF SOFTWARE.

Acknowledgements
We would like to sincerely thank all those who made the production of Statistical Solutions software package nQuery + nTerim 4.0 possible.

The Statistical Solutions Team:
Andrew Grannell, Ronan Fitzpatrick, Steven Keady, Brendan Nyhan, Diana Scriven, Caroline Costello, Helen Murphy, Ian McDonald, Shane Thornhill, Aisling Leen, Kevin Sievewright, Mary Byrne.

Special Thanks to:
Eoghan Murphy, Brian Sullivan, Denis Moore and Niall Fitzgerald.

Contents

Chapter 1 1
Systems Guide 1
1.1 System Requirements 2
1.2 Validation 2
1.3 Support 2
Chapter 2 4
Getting Started Guide 4
2.1 Starting nTerim 5
2.2 Home Window 5
2.3 Menu Bar 6
2.4 Opening a New Design 8
2.5 Selecting an nQuery Advisor Design Table through nTerim 13
2.6 Using the Assistant Tables 14
2.7 Plotting 15
2.8 Help and Support 18
Chapter 3 19
Group Sequential / Interim Design 19
3.1 Two Means 20
3.1.1. Introduction 20
3.1.2. Methodology 22
3.1.3. Examples 24
3.2 Two Proportions 32
3.2.1. Introduction 32
3.2.2. Methodology 34
3.2.3. Examples 37
3.3 Survival 44
3.3.1. Introduction 44
3.3.2. Methodology 46
3.3.3. Examples 48
Chapter 4 57
Fixed Term Design 57
4.1 One-Way Repeated Measures Contrast (Constant Correlation) 58
4.1.1. Introduction 58
4.1.2. Methodology 59

4.1.3. Examples 60
4.2 Repeated Measures Design for Two Means 72
4.2.1. Introduction 72
4.2.2. Methodology 73
4.2.3. Examples 74
4.3 Repeated Measure for Two Proportions 88
4.3.1. Introduction 88
4.3.2. Methodology 89
4.3.3. Examples 90
4.4 One-Way Analysis of Variance (ANOVA) 101
4.4.1. Introduction 101
4.4.2. Methodology 102
4.4.3. Examples 103
4.5 Analysis of Covariance (ANCOVA) 110
4.5.1. Introduction 110
4.5.2. Methodology 111
4.5.3. Examples 113
4.6. Multivariate Analysis of Variance (MANOVA) 123
4.6.1. Introduction 123
4.6.2. Methodology 124
4.6.3. Examples 128
4.7. One Poisson Mean 140
4.7.1. Introduction 140
4.7.2 Methodology 141
4.7.3 Examples 143
4.8. Two Poisson Means 148
4.8.1 Introduction 148
4.8.2 Methodology 149
4.8.3 Examples 153
4.9. Two Negative Binomial Rates 158
4.9.1 Introduction 158
4.9.2 Methodology 159
4.9.3 Examples 162
4.10. Two Incidence Rates 168
4.10.1 Introduction 168

4.10.2 Methodology....169
4.10.3 Examples....170
4.11. Cox Regression....175
4.11.1 Introduction....175
4.11.2 Methodology....176
4.11.3 Example....178
4.12. Poisson Regression....183
4.12.1 Introduction....183
4.12.2 Methodology....184
4.12.3 Examples....187
4.13. Probit Regression....199
4.13.1 Introduction....199
4.13.2 Methodology....200
4.13.3 Examples....202
4.14. Higher-Order Cross-Over Design for Two Means – Superiority by a Margin– Using Differences....211
4.14.1 Introduction....211
4.14.2 Methodology....211
4.14.3 Examples....214
4.15. Higher-Order Cross-Over Design for Two Means – Non–Inferiority - Using Differences....218
4.15.1 Introduction....218
4.15.2 Methodology....218
4.15.3 Examples....221
4.16. Higher-Order Cross-Over Design for Two Means – Equivalence - Using Differences....225
4.16.1 Introduction....225
4.16.2 Methodology....225
4.16.3 Examples....228
4.17. Higher-Order Cross-Over Design for Two Means – Superiority by a Margin– Using Ratios....232
4.17.1 Introduction....232
4.17.2 Methodology....232
4.17.3 Examples....236
4.18. Higher-Order Cross-Over Design for Two Means – Non–Inferiority - Using Ratios....240
4.18.1 Introduction....240

4.18.2 Methodology....240
4.18.3 Examples....244
4.19. Higher-Order Cross-Over Design for Two Means – Equivalence - Using Ratios....248
4.19.1 Introduction....248
4.19.2 Methodology....248
4.19.3 Examples....251
4.20. Two Exponential Means....255
4.20.1 Introduction....255
4.20.2 Methodology....255
4.20.3 Examples....256
4.21. One Sample Z-test....260
4.21.1 Introduction....260
4.21.2 Methodology....260
4.21.3 Examples....262
4.22. Paired Samples Z-test....266
4.22.1 Introduction....266
4.22.2 Methodology....266
4.22.3 Examples....268
4.23. Two Sample Z-test....272
4.23.1 Introduction....272
4.23.2 Methodology....272
4.23.3 Examples....274
4.24. One Variance....278
4.24.1 Introduction....278
4.24.2 Methodology....278
4.24.3 Examples....279
4.25. Two Variances....283
4.25.1 Introduction....283
4.25.2 Methodology....283
4.25.3 Examples....285
4.26. One ROC Curve....289
4.26.1 Introduction....289
4.26.2 Methodology....289
4.26.3 Examples....292
4.27. Two ROC Curves....296

4.27.1 Introduction 296
4.27.2 Methodology....296
4.27.3 Examples 299
4.28. Two Correlations....304
4.28.1 Introduction 304
4.28.2 Methodology....304
4.28.3 Examples 305
4.29. One Sensitivity 310
4.29.1 Introduction 310
4.29.2 Methodology....310
4.29.3 Examples 311
4.30. One Specificity 315
4.30.1 Introduction 315
4.30.2 Methodology....315
4.30.3 Examples 316
Chapter 5....320
Cluster Randomized Trials 320
5.1. CRT Two Means – Completely Randomized 321
5.1.1 Introduction 321
5.1.2 Methodology....322
5.1.3 Examples 323
5.2. CRT Two Proportions Inequality – Completely Randomized....328
5.2.1 Introduction 328
5.2.2 Methodology....329
5.2.3 Examples 332
5.3. CRT Two Proportions Equivalence – Completely Randomized....340
5.3.1 Introduction 340
5.3.2 Methodology....341
5.3.3 Examples 344
5.4. CRT Two Proportions Non-Inferiority – Completely Randomized....349
5.4.1 Introduction 349
5.4.2 Methodology....350
5.4.3 Examples 353
5.5. CRT Two Proportions Superiority – Completely Randomized 359
5.5.1 Introduction 359

5.5.2 Methodology 360
5.5.3 Examples 363
5.6 CRT Two Incidence Rates – Completely Randomized 369
5.6.1 Introduction 369
5.6.2 Methodology 369
5.6.3 Examples 371
5.7 CRT Two Means – Matched-Pair 375
5.7.1 Introduction 375
5.7.2 Methodology 375
5.7.3 Examples 377
5.8 CRT Two Proportions Inequality – Matched-Pair 381
5.8.1 Introduction 381
5.8.2 Methodology 381
5.8.3 Examples 383
5.9 CRT Two Incidence Rates – Matched-Pair 387
5.9.1 Introduction 387
5.9.2 Methodology 387
5.9.3 Examples 389
Chapter 6 394
References 394

Chapter 1

Systems Guide

1.1 System Requirements

As with most software packages, there are a set of requirements on the various aspects of the users machine in order to achieve full functionality. For nTerim 4.0, the set of system requirements are listed in full below.

Operating System:	Windows 8 or later Windows 7 Windows Vista Windows Server 2012 (R2 or later) Windows Server 2012 Windows Server 2008 (R2 or later) Windows Server 2008 Windows Server 2003
Processor:	Either 32-bit or 64-bit processor Minimum of 450MHz processor
Hard Disc:	250MB for the nTerim software package
RAM:	512MB
Additional Software:	Microsoft .NET Framework 4

Note: Administrative privileges to the end users machine will be required for installation process only.

1.2 Validation

The calculations contained within this software package have been widely and exhaustively tested. Various steps of each calculation along with the results have been verified using many text books and published journal articles. Furthermore, the calculations contained within this software package have been compared to, and verified against various additional sources when possible.

1.3 Support

For issues pertaining to the methodology and calculations of each test in nTerim, there is a brief outline of how each test is calculated in the Methodology section of each test chapter of the manual. There are accompanying references for each test throughout the text and can be located in the References section of the manual. If further clarification is required, please contact our support statisticians by email at support@statsols.com.

If there are any issues with any aspect of the installation process, there are three approaches you can take: (i) you can check the system requirements outline in Section 1.1 of this manual, (ii) look up the installation help and FAQ's on our website: http://www.statsols.com/products/nquery-advisor-nterim/, and (iii) you can email us for technical help at support@statsols.com.

In order to help us address your questions in the best way possible, the more information you can provide us with, the better. If it is a technical question about one of our test tables, screen shots of the completed tables of issues you are having are very helpful.

In order to address any installation issues or technical questions relating to the users machines, the more information provided about the type of machine in question can speed up the process by a great deal. Screen shots of installation issues are very helpful to us in solving any issue you may have.

Chapter 2

Getting Started Guide

This chapter is a guide to help users get acquainted with the layout and various aspects of the interface of nQuery Advisor + nTerim 4.0. This chapter aims at getting the user a firm understanding of how to approach study design using nTerim, in a quick and easy way. Every aspect of the nTerim interface will be presented in this chapter, from the home window to the various plotting menus and side tables.

2.1 Starting nTerim

There are two main ways to open nTerim on your desktop. By double-clicking on the desktop icon, nTerim will be automatically launched. Alternatively, if you chose not to have a desktop shortcut to nTerim, you can find it by clicking on the Windows Start button, and then select “All Programs”. A list of all the programs on the user’s machine will be listed in alphabetical order. You can locate nTerim under the title “nQuery Advisor + nTerim 4.0”. Click on this folder, and then select “nQuery Advisor + nTerim 4.0” to launch the program.

2.2 Home Window

Once the user has launched nTerim, the home window will appear as illustrated below in Figure 2.2.1. From the home window, there are several options open to the user depending on what they want to do. The user can open a new fixed term or interim design table, open a previous design that was saved before, access the manual or access the Statistical Solutions support website for help or guidance.

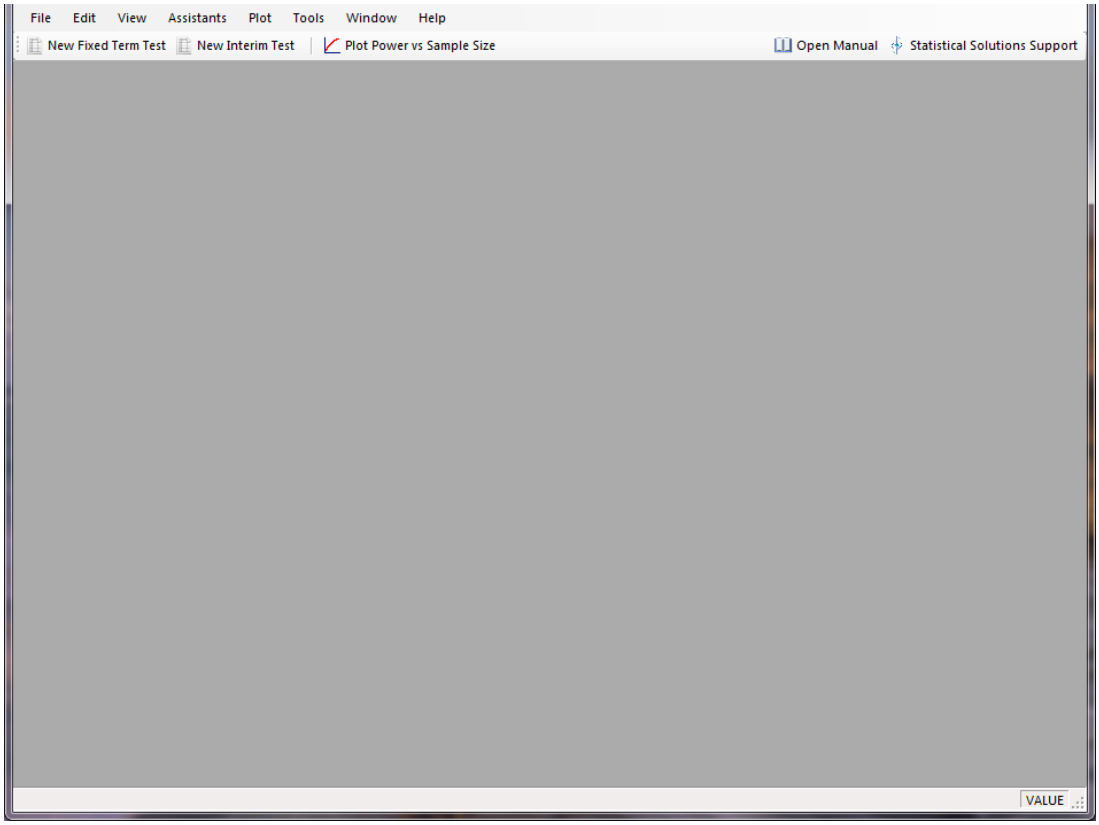


Figure 2.2.1. Home Window

2.3 Menu Bar

The first aspect of the interface we will review is the menu bar and all the options available. There are eight options on the menu bar: File, Edit, View, Assistants, Plot, Tools, Window and Help. These are highlighted in Figure 2.3.1 below.

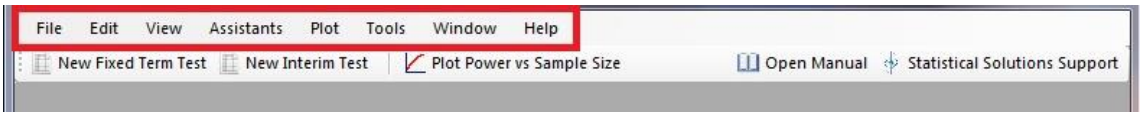


Figure 2.3.1. Menu Bar

The **File** menu allows the user to open a new or previously saved design table, as well as enabling the user to save a design and allowing the user to exit nTerim whenever they wish. Design tables can be saved as .nia format, which is the Statistical Solutions file format for nTerim.

The **Edit** menu enables to user to fill a design table using the Fill Right option. This is where the user, when defining multiple columns, enters certain information into a column and can copy this information across the remaining empty columns.

The **View** menu is initially unavailable until the user opens a design table. Once a table has been opened, several options appear enabling the user to view various plots, and toggle between various assistant tables, help/guides cards and notes.

The **Assistants** menu is initially unavailable until the user opens a design table. Once a table has been opened, the menu enables the user to open and toggle between various side tables depending on the design table. Another side table located under the Assistants menu is the Specify Multiple Factor table. This table enables the user to specify a range of designs (or columns in a table).

The **Plot** menu is initially unavailable until the user opens a design table. Once a table has been opened, the user can use this menu to create certain plots such as Power vs. Sample Size plots, Boundaries Plots and Spending Function Plots.

The **Tools** menu allows the user to define certain settings before running any analysis such as defining the minimum cell count and outlining various assumptions in relation to group proportions and means. This also enables the user to save design tables and Looks tables as images for transporting.

The **Window** menu is initially unavailable until the user opens a design table. Once a table has been opened, the menu enables the user to toggle between the various tables and plots they may be working on during their session.

The **Help** menu gives access to the nTerim manual and supplies the nTerim version information and license agreement.

Below is a complete list of menu options from the menu bar:

File –> New
Open
Fresh Table
Save
Save As
Close Test
Exit

Edit –> Fill Right
Clear Table
Clear Column
Clear Selection

View –> [Option not available until a test window is opened]
Looks
Specify Multiple Factor Table
Covariance Matrix (MANOVA design table only)
Boundaries Graph
Power vs. Sample Size Plot

Boundaries Plot
Spending Function Plot
Output
Help
Notes

Assistants –> Specify Multiple Factor Table
Compute Effect Size
Randomization
Survival Parameter Converter
Distribution Function
Windows Calculator

Plot –> Power vs. Sample Size Plot
Spending Function Plot
Boundaries Plot

Tools –> Print Main Table to Clipboard
Print Looks Table to Clipboard
Settings

Windows –> [Close All if no test window open]
Close All
{List of Open Windows}

Help –> Help
About
Manual

2.4 Opening a New Design

The next aspect of the interface we will review is opening a new design, both Fixed term and Interim. There are two ways in which the user can open a new design in nTerim; (i) by clicking on the File > Open option or (ii) using the shortcut buttons highlighted in Figure 2.4.1 below.

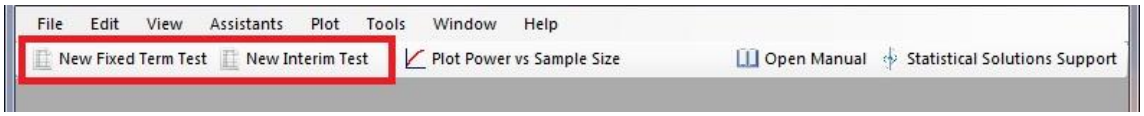


Figure 2.4.1. New Design Tabs

Using either of the steps outlined above, the user will then be presented with the "Study Goal and Design" window, as shown in Figure 2.4.2 below. In relation to selecting the term of their designs, the user must select either Fixed or Interim. The user will then be presented with a list of options to the type of design they require.

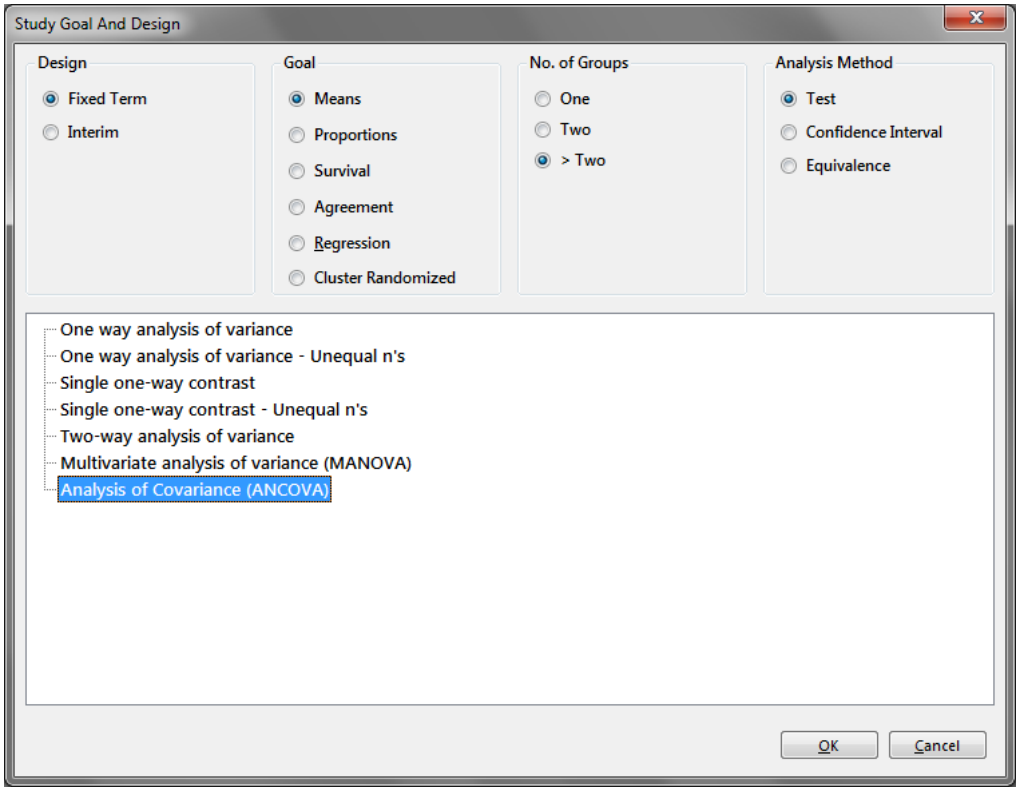


Figure 2.4.2. Open New Fixed Term Design

The options for Fixed term designs are presented in Figure 2.4.2 For example, If you want to choose the Analysis of Covariance (ANCOVA) table, you must first select "Means" as the Goal, ">Two" as the No. of Groups and "Test" as the Analysis Method. You can then select Analysis of Covariance (ANCOVA) from the list of tests. Once you click "OK", the design table will be launched.

In this example, the Analysis of Covariance (ANCOVA) table was selected. A screen shot of this design table is given in Figure 2.4.3.

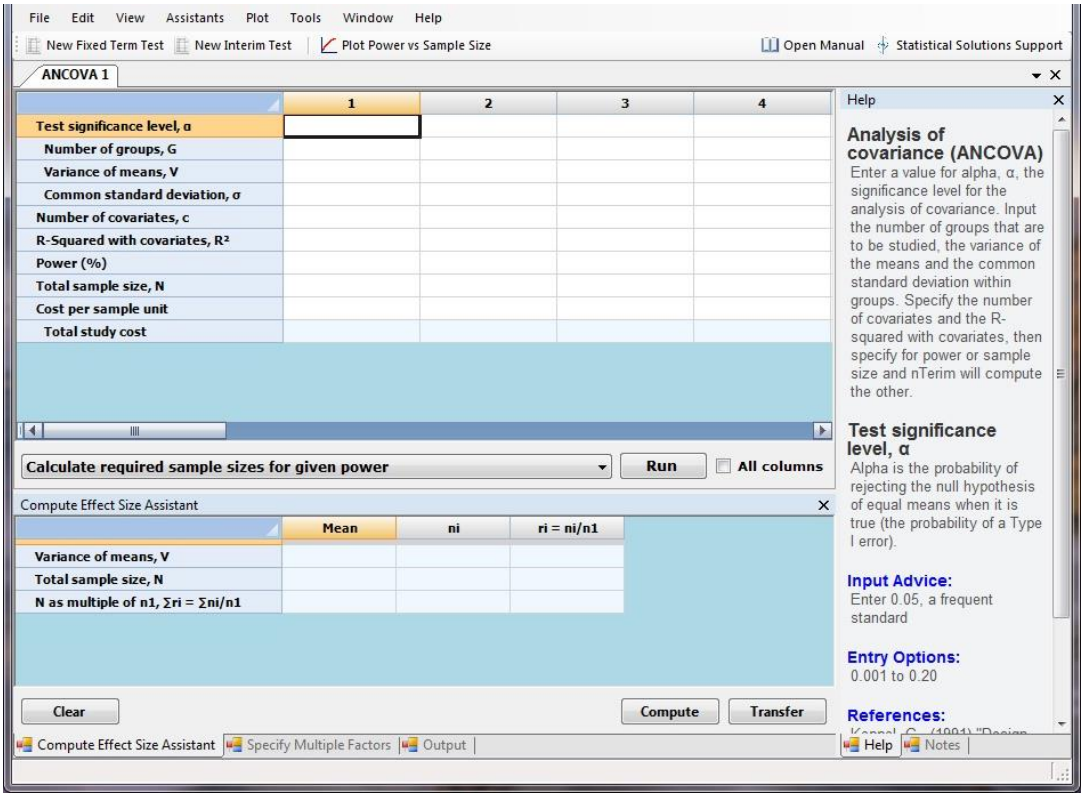


Figure 2.4.3. Example of Fixed Term Design Interface

As it can be seen from Figure 2.4.3, the Fixed term design window is split into three main sections: (i) the test table, (ii) Assistant Tables & Output and (iii) Help/Guide Cards. The main table represents the test table. In this example it is an ANCOVA table. Values for various parameters can be entered by the user. For some tests, additional values need to be calculated. This is provided for by using the Assistants tables, found at the bottom half of the interface. Additional calculations can be done and the appropriate values can be transferred from the Assistants tables to the main test table.

Once all the appropriate information has been entered in the test table, the user must select the appropriate calculation to run, i.e. whether you want to solve for power given a specified sample size, or solve for sample size given a specified power. The user can select the appropriate calculation to run from the drop-down menu between the main test table and the Assistants table. Once the appropriate test is selected, the user must click on “Run” to run the analysis.

If multiple columns have been specified by the user, there is an option to run the calculation for all the columns. This is achieved by simply ticking the “All columns” box beside the “Run” button before clicking “Run”. This will tell nTerim to concurrently run the calculations for all columns. Then, by simply clicking on a column, the output statement will be presented.

Similarly to opening a Fixed Term test, if the user clicks on the “New Interim Test” button below the menu bar, the “Study Goal and Design” menu window will appear with the list of interim designs available in nTerim. This “Study Goal and Design” window is presented below in Figure 2.4.4.

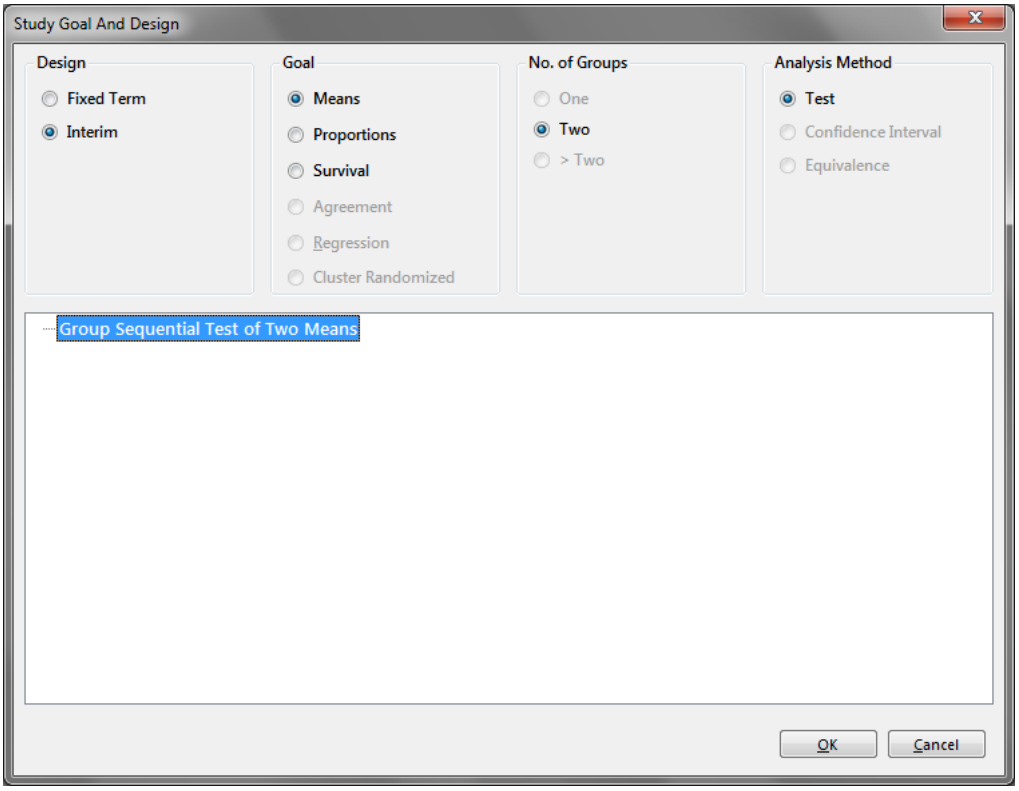


Figure 2.4.4. Open New Interim Design

The options for Interim term designs are presented in Figure 2.4.4. For example, if you want to choose the Group Sequential Test of Two Means table, you must first select “Means” as the Goal, “>Two” as the No. of Groups and “Test” as the Analysis Method. You can then select Group Sequential Test of Two Means from the list of tests. Once you click “OK”, the design table will be launched.

As it can be seen from Figure 2.4.5, the Interim term design window is split into four main sections: (i) the test table, (ii) Looks Table & Output (iii) Boundary Graph and (iv) Help/Guide Cards. The main table represents the test table. In this example it is a Group Sequential Test of Two Means table. The top half of the main test table is for various parameters to be entered by the user. The bottom half is for the user to define parameters relating to the interim design such as number of looks, spending function to be used, and so on.

Once all the appropriate information has been entered in the test table, the user must select the appropriate calculation to run, i.e. whether you want to solve for power given a specified sample size, or solve for sample size given a specified power. The user can select the appropriate calculation to run from the drop-down menu between the main test table and the Looks table. Once the appropriate test is selected, the user must click on “Run” to run the analysis.

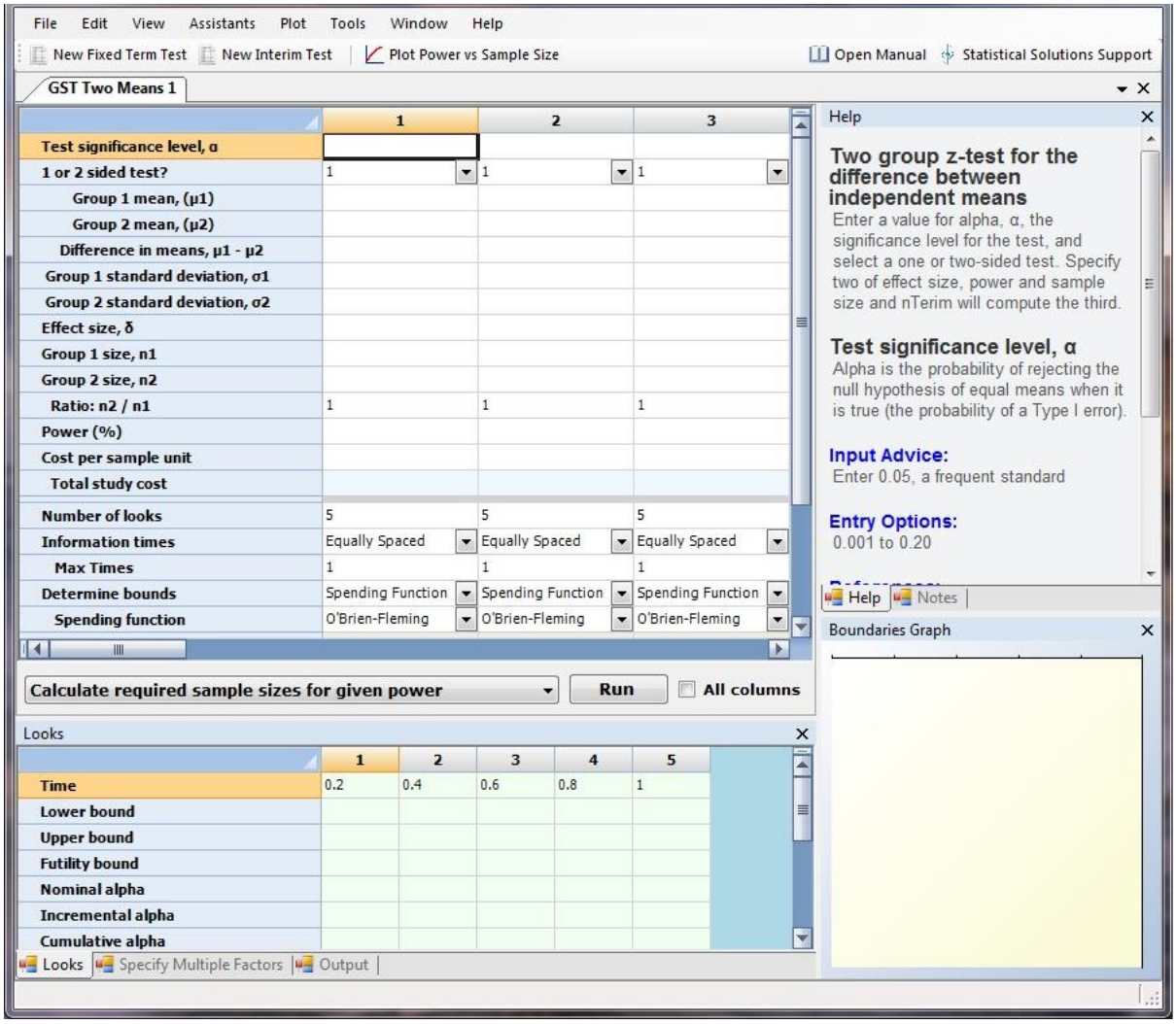


Figure 2.4.5. Example of an Interim Design Window

If multiple columns have been specified by the user, there is an option to run the calculation for all the columns. This is achieved by simply ticking the “All columns” box beside the “Run” button before clicking “Run”. This will tell nTerim to concurrently run the calculations for all columns. Then, by simply clicking on a column, the output statement will be presented as well as the boundary graph for each column in the bottom right hand corner of the interface.

2.5 Selecting an nQuery Advisor Design Table through nTerim

A new feature added to nTerim 4.0 is the ability to open an nQuery design table through nTerim. This enables the user to seamlessly transition between nTerim and nQuery. By opening the "Study Goal and Design" window using the options outlined in the previous section (Section 2.4), the user has the full range of design tables available in both nTerim and nQuery at their disposal.

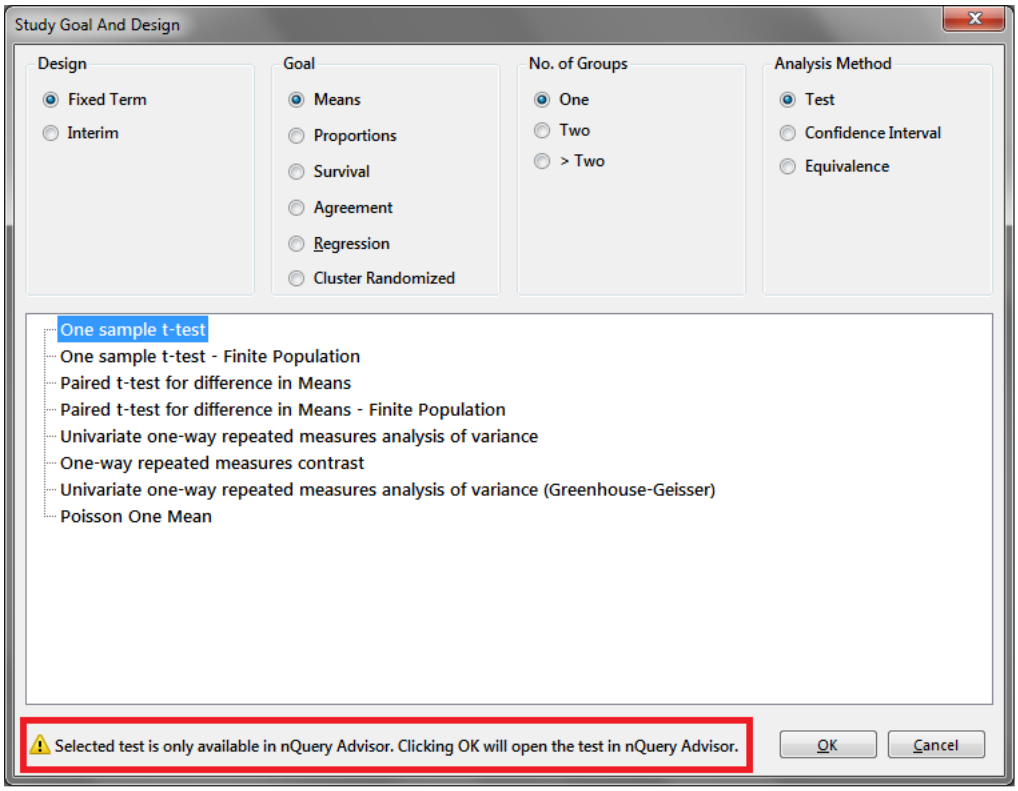


Figure 2.5.1. Study Goal and Design Window

As shown in the "Study Goal and Design" window in Figure 2.5.1 above, the user has selected a One sample t test. This test is available in nQuery, therefore a message has appeared at the bottom of the "Study Goal and Design" window stating "Selected test is only available in nQuery Advisor. Clicking OK will open the test in nQuery Advisor".

This message is highlighted in the red box in Figure 2.5.1. Once the user clicks "OK", this will prompt nQuery to open the specified test.

2.6 Using the Assistant Tables

The Assistants tables are a new feature added to nTerim to aid the user in calculating various additional components of certain study designs. These tables are only associated with certain design tables. With nTerim, we know which Assistant table is associated with each test so they automatically pop up once a design table is opened.

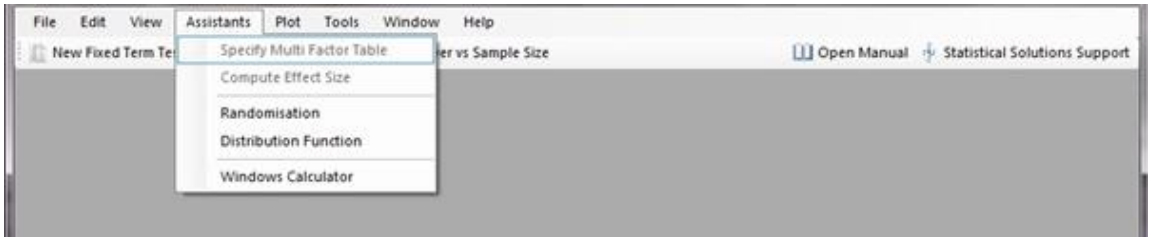


Figure 2.6.1. Assistants Menu Options

The full list of Assistants tables is given in the menu bar, as shown in Figure 2.6.1, including; Compute Effect Size and Specify Multi Factor table.

A very common Assistant table that is regularly required is the "Compute Effect Size" table. Once the appropriate information is entered, nTerim will calculate the values required for the main test table. Once the user is happy with the values entered and calculated, they can click "Transfer" and the required values from the Assistant table will be transferred up to the main design table. An example of the "Compute Effect Size" assistant table is shown below in Figure 2.6.2.

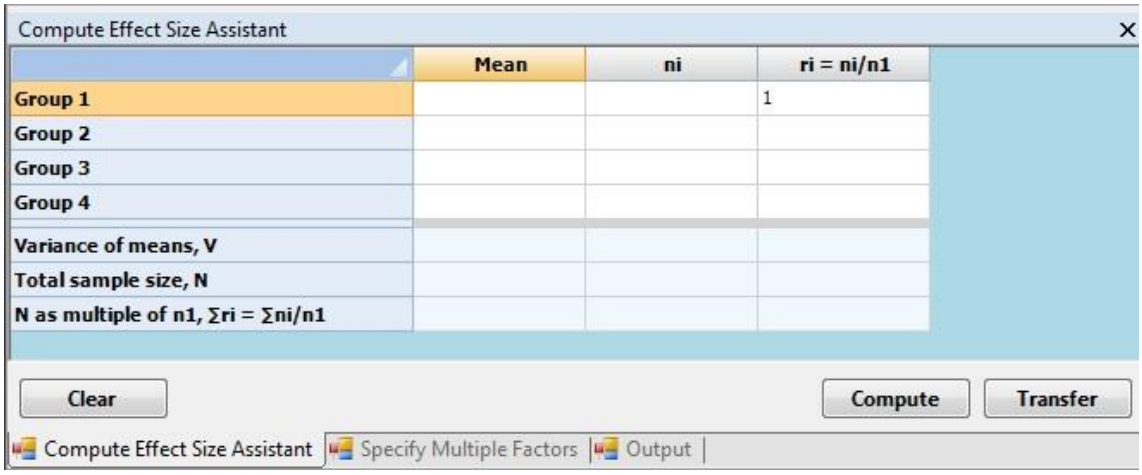


Figure 2.6.2. Example of Effect Size Assistant Table

The "Specify Covariance Matrix" assistant table, where available, can be utilised by the user to manually define the covariance matrix they wish to use in their study design.

The “Specify Multi Factor” assistant table is used to define a range values to be filled in across several columns in the test design table. Once the user fills in this table with the range of values they require, by clicking “Run”, nTerim will fill out the required number of columns to satisfy the outlined range of parameters.

2.7 Plotting

A plotting menu has been introduced to nTerim 4.0 for all the additional graphing features that have been added. Additional features have been added to the Power vs. Sample Size and Boundary plots including multiple plotting capabilities, highlighting various boundary functions of interest and scrolling features to enable users to pin-point exact values. The plotting menu bar is displayed in Figure 2.7.1 below.

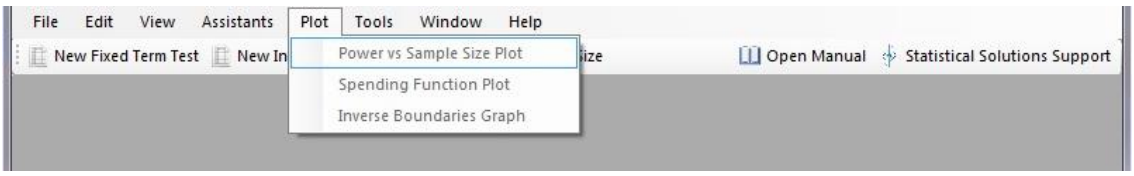


Figure 2.7.1. Plot Menu Options

In relation to Interim designs, a boundary plot is automatically displayed after running the calculations. This is always displayed on the bottom right hand corner of the nTerim window. An example of an O’Brien-Fleming boundary is given in Figure 2.7.2 below.

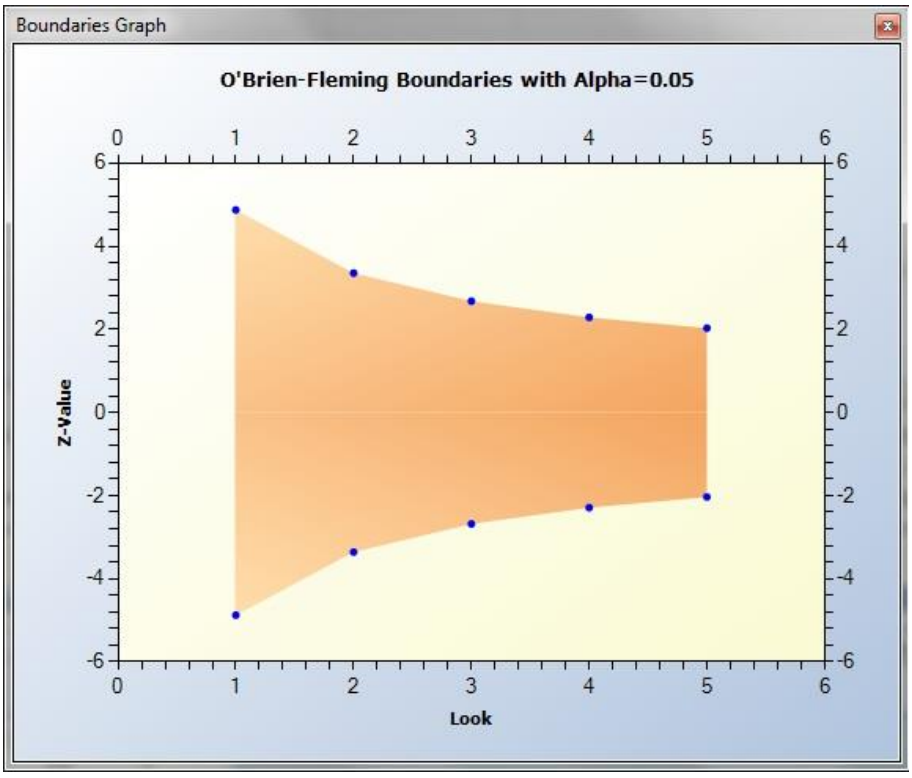


Figure 2.7.2. Example of a Boundary Plot

In relation to Power vs. Sample Size plots, there is also a shortcut button provided in the tool bar, just below the menu bar as highlighted in Figure 2.7.3 below. In order to use this

function, the user must highlight the columns which they would like to compare and then click on the “Plot Power vs. Sample Size” button.

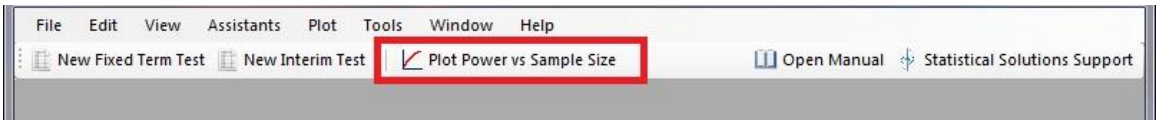


Figure 2.7.3. Power vs. Sample Size Plot Shortcut Tab

An example of the new Power vs. Sample Size plot is displayed in Figure 2.7.4 below. This plot shows three columns being compared. The legend on the right side of the window can be altered to label each line appropriately.

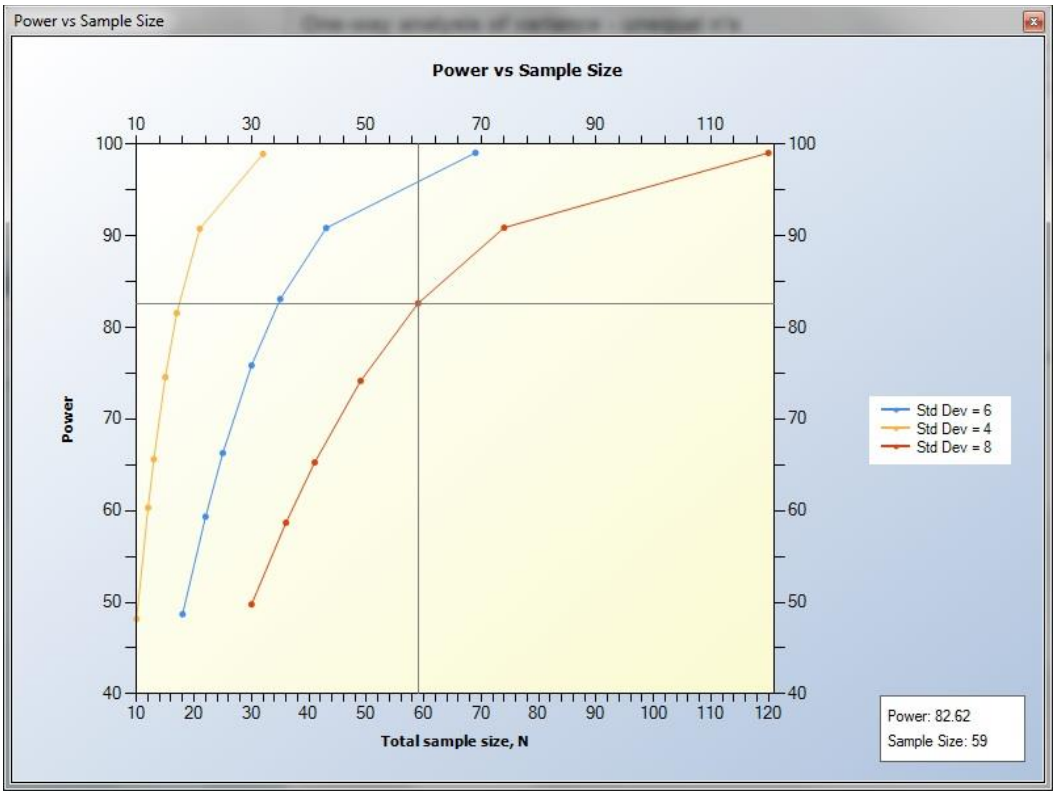


Figure 2.7.4. Power vs. Sample Size Plot

A crosshair is provided to enable the user to pin-point exact values for power and sample size at various points on each line. These exact values are given in the box in the bottom right hand corner of the plot window.

In order to save a plot in nTerim, simply right click anywhere on the plot window and a list of options will be presented as illustrated in Figure 2.7.5. The options include “Save Image”, “Print”, “Print Preview” and “Page Setup”. Select “Save Image” from this list to save the plot.

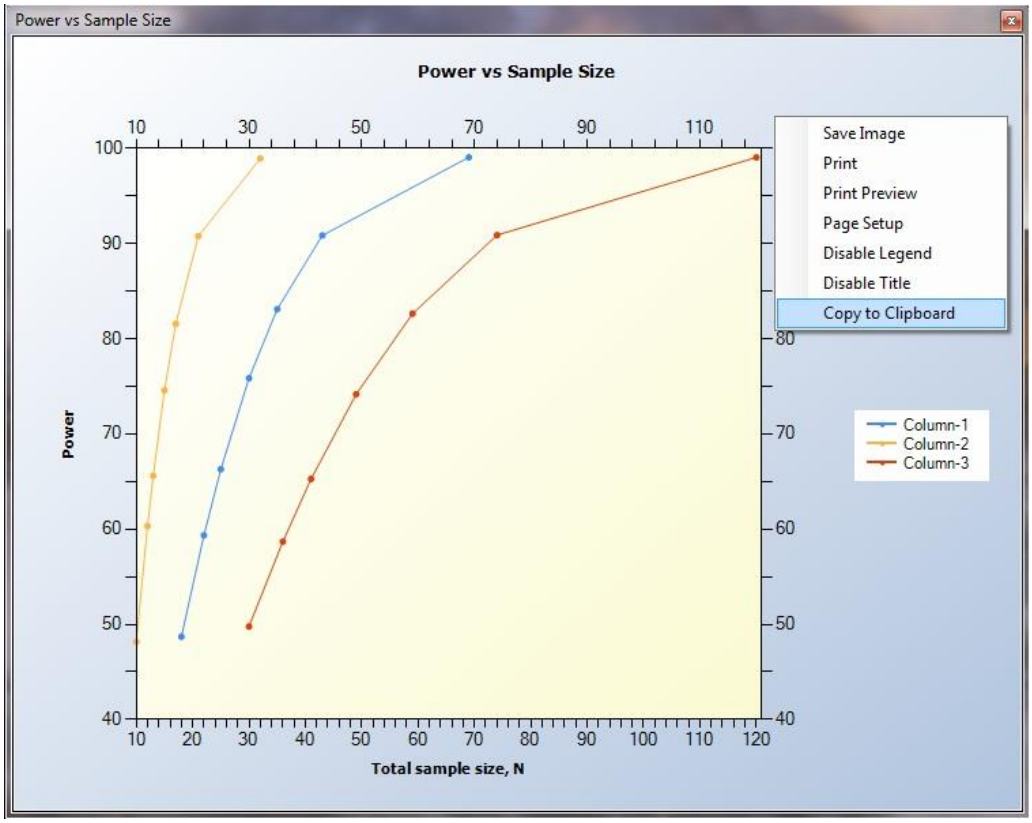


Figure 2.7.5. Saving a plot

A separate window will appear prompting the user to select the folder in which they would like to save the plot. Once the user has chosen the folder to save the plot in, they can select what format to save in. The format options available to save a plot are in a .JPEG or .PNG format. Once the location and format have been selected by the user, simply click “Save” to save the plot.

This image can now be imported to many Microsoft applications such as MS Word for reporting or MS Powerpoint for presentation purposes.

2.8 Help and Support

For issues pertaining to the methodology and calculations of each test in nTerim, there is a brief outline of how each test is calculated in the Methodology section of each test chapter of the manual. There are accompanying references for each test throughout the text and these can be located in the References section of the manual.

In the nTerim window there are two useful shortcuts that have been added to the tool bar. The first shortcut is the "Open Manual" button which has been added to help the user find the appropriate chapter of the manual much easier. If the user is working in a particular design window, for example the MANOVA window, and the user clicks on the "Open Manual" button, a PDF of the MANOVA chapter in the manual will automatically open, providing the user with the background and technical information on MANOVA as well as examples in nTerim.

The second shortcut is the "Statistical Solutions Support" button. If further clarification on any aspect of nTerim is required, please contact our support statisticians by clicking on this button. This shortcut takes the user to the Statistical Solutions support website where queries can be entered and sent directly to our support team.

These support shortcuts are highlighted in the nTerim tool bar in Figure 2.8.1 below.

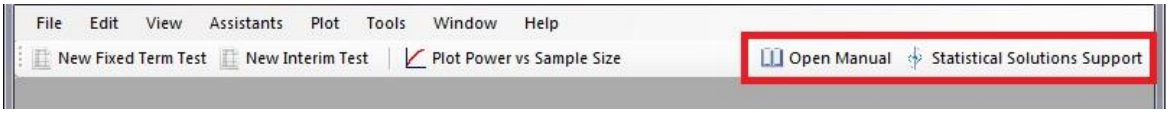


Figure 2.8.1. Manual and Support Shortcut Tabs

If there are any issues with any aspect of the installation process, there are three approaches you can take: (i) you can check the system requirements outline in Section 1.1 of this manual, (ii) look up the installation help and FAQ's on our website: http://www.statsols.com/products/nquery-advisor-nterim/, and (iii) you can email us for technical help at support@statsols.com.

In order to help us address your questions in the best way possible, the more information you can provide us with, the better. If it is a technical question about one of our test tables, screen shots of the completed tables of issues you are having are very helpful.

In order to address any installation issues or technical questions relating to the users machines, the more information provided about the type of machine in question can speed up the process by a great deal. Screen shots of installation issues are very helpful to us in solving any issue you may have.

Chapter 3

Group Sequential / Interim Design

3.1 Two Means

3.1.1. Introduction

nTerim 4.0 is designed for the calculation of Power and Sample Size for both Fixed Period and Group Sequential design. In relation to Group Sequential designs, calculations are performed using the Lan-DeMets alpha spending function approach (DeMets & Lan, 1984; DeMets & Lan, 1994) for estimating boundary values. Using this approach, boundary values can be estimated for O'Brien-Fleming (O'Brien & Fleming, 1979), Pocock (Pocock, 1977), Hwang-Shih-DeCani (Hwang, Shih & DeCani, 1990) and the Power family of spending functions. Calculations follow the approach of Reboussin et al (1992) and Jennison & Turnbull (2000). Calculations can be performed for studies that involve comparisons of means, comparisons of proportions and survival studies as well as early stopping for Futility.

Group Sequential Designs

Group Sequential designs differ from Fixed Period designs in that the data from the trial is analyzed at one or more stages prior to the conclusion of the trial. As a result the alpha value applied at each analysis or `look' must be adjusted to preserve the overall Type 1 error. The alpha values used at each look are calculated based upon the spending function chosen, the number of looks to be taken during the course of the study as well as the overall Type 1 error rate. For a full introduction to group sequential methods see Jennison & Turnbull (2000) and Chow et al (2008).

Spending Function

There are four alpha and beta spending functions available to the user in nTerim 4.0 as well as an option to manually input boundary values. As standard all alpha spending functions have the properties that $\alpha(0) = 0$ and $\alpha(1) = \alpha$. Similarly, all beta spending functions have the properties that $\beta(0) = 0$ and $\beta(1) = \beta$. Functionally the alpha and beta spending functions are the same. In Table 3.1.1 we list the alpha spending functions available in nTerim 4.0.

Table 3.1.1. Spending Function Equations

O'Brien-Fleming	$\alpha(\tau) = 2\left(1 - \Phi\left(\frac{z_{\alpha/2}}{\sqrt{\tau}}\right)\right)$
Pocock	$\alpha(\tau) = \alpha ln(1 + (e - 1)\tau)$
Power	$\alpha(\tau) = \alpha\tau^{\Phi}, \Phi > 0$
Hwang-Shih-DeCani	$\alpha(\tau) = \alpha\left[\frac{(1 - e^{-\Phi\tau})}{(1 - e^{-\Phi})}\right], \Phi \neq 0$

The parameter τ represents the time elapsed in the trial. This can either be as a proportion of the overall time elapsed or a proportion of the sample size enrolled.

The common element among most of the different spending functions is to use lower error values for the earlier looks. By doing this it means that the results of any analysis will only be considered significant in an early stage if it gives an extreme result.

Boundaries

The boundaries in nTerim 4.0 represent the critical values at each look. These boundaries are constructed using the alpha and beta spending functions. Users in nTerim 4.0 are given the option to generate boundaries for early rejection of the null hypothesis, H_0, using the alpha spending function, or to generate boundaries for early rejection of either the null or alternative hypothesis, H_0 or H_1, using a combination of both the alpha and beta spending functions. The notion of using an alpha spending function approach to generate stopping boundaries for early rejection of H_0 was first proposed by Lan and DeMets (1983), we refer to such boundaries in nTerim 4.0 as efficacy boundaries. Building on the work of Lan and DeMets, Pampallona, Tsiatis, and Kim (1995, 2001) later put forward the concept of using a beta spending approach to construct boundaries for early rejection of H_1, we refer to these boundaries in nTerim as futility boundaries.

Essentially, if a test statistic crosses an efficacy boundary then it can be concluded that the experimental treatment shows a statistically significant effect, the trial can be stopped with rejection of the null hypothesis. If the test statistic crosses a futility boundary then this indicates with high probability that an effect will not be found, that the trial can be terminated by rejecting the alternative hypothesis.

In the case where the user wishes to generate boundaries for early rejection of either the null or alternative hypothesis, H_o or H_1, they are given two options; either to have the boundaries binding, or non-binding. With binding boundaries, if the test statistic crosses the futility boundary, the test must be stopped, otherwise the type-1 error may become inflated. The reason for this is that there is an interaction between the efficacy and futility boundaries in their calculation that could cause the efficacy boundary to shift. In the case of non-binding boundaries; the efficacy boundaries are calculated as normal, that is, as if the futility boundaries did not exist. This eliminates the danger of inflating the type-1 error when the futility boundary is overruled. The downside of the non-binding case is that it may increase the required sample size relative to the binding case.

The boundaries calculated in nTerim 4.0 follow the procedures outlined by Reboussin et al (1992) and Jennison & Turnbull (2000).

3.1.2. Methodology

The variables are defined as

Symbol	Description
α	Probability of Type I error
β	Probability of Type II error
$1-\beta$	Power of the Test
μ_1, μ_2	Group Means
σ_1, σ_2	Group Standard Deviations
N_1, N_2	Group Sample Sizes
R	Ratio of N_1 to N_2
Θ	Drift Parameter
K	Number of Time-points (Looks)

Calculate Sample Sizes for a given Power

Using the number of time-points (K), number of sides, type of spending function, the hypothesis to be rejected, the type 1 error, α, and the power $(1-\beta)$, the drift parameter Θ can be obtained using the algorithms by Reboussin et al (1992) and Jennison & Turnbull (2000). The test statistic is defined as;

$$\Theta = \frac{\mu_1 - \mu_2}{\sqrt{\frac{\sigma_1^2}{N_1} - \frac{\sigma_2^2}{N_2}}} \qquad \{3.1.1\}$$

The user supplies the means (μ_1, μ_2), and either R, N_1 or N_2. Since $R = \frac{N_2}{N_1}$ it follows that a value of $R = 1$ indicates equal sample sizes. The approach to solving this problem is dependent on what information the user supplies. Given any two of R, N_1 or N_2, the unknown is obtained by solving Equation {3.1.1}.

Calculate Attainable Power with the given Sample Sizes

Given α, N_1, group means (μ_1, μ_2), group standard deviations (σ_1, σ_2), R (or N_2), time-points and type of spending function. The requirement is to obtain the power. The steps are:

- Obtain Θ by solving Equation {3.1.1} (given that $N_1, R,\ \mu_1,\ \mu_2,\ \sigma_1$ and σ_2 are known),
- Obtain power using the algorithm by Reboussin et al (1992) and Jennison & Turnbull (2000)

Calculate Means given all other information

Given α, N_1, group standard deviations (σ_1, σ_2), R (or N_2), power $(1-\beta)$, time-points and type of spending function. The requirement is to obtain either μ_1 or μ_2, given the other. The steps are:

- Obtain Θ by solving Equation {3.1.1} (given that $N_1, R,\ \mu_1,\ \mu_2,\ \sigma_1$ and σ_2 are known),
- Equation {3.1.1} can be expressed as a quadratic in μ_1 or μ_2. The roots give the unknown μ.

By default, nTerim assumes that $\mu_1 < \mu_2$ and will select the appropriate root.

3.1.3. Examples

Example 1: O'Brien-Fleming Spending Function
This example is adopted from Reboussin et al (1992) using the O'Brien-Fleming spending function.

1. Open nTerim through the Start Menu or by double clicking on the nTerim desktop icon. Then click on “New Interim Test” from the tool bar at the top of the window. A “Study Goal and Design” window will appear as shown below. Select the options as mapped out in Figure 3.1.1, then Click “OK”.

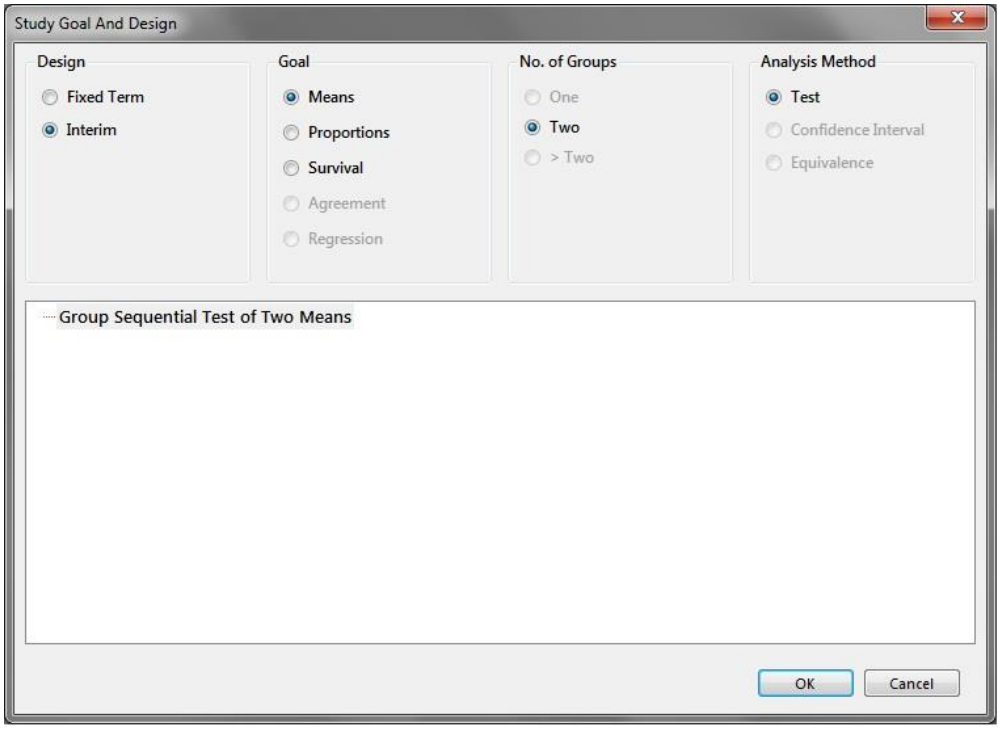


Figure 3.1.1. Study Goal and Design Window

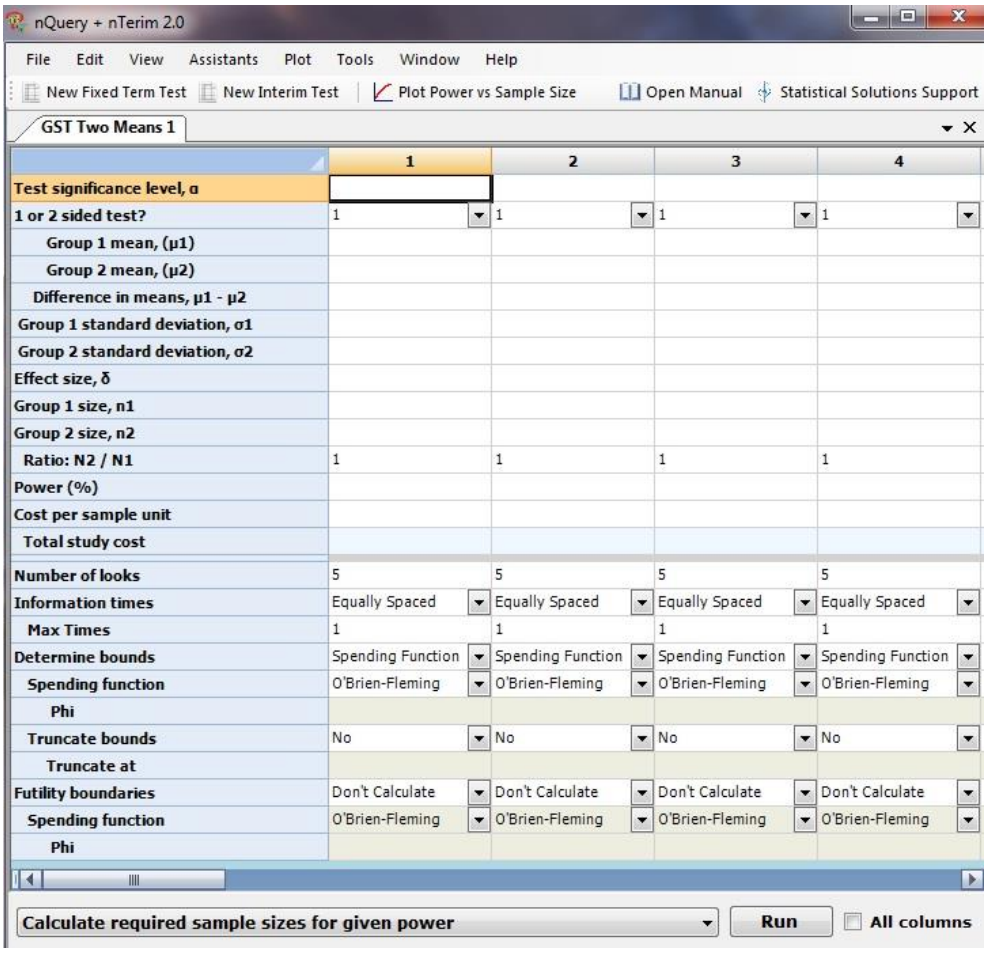
2. Now you have opened the test table, as illustrated in Figure 3.1.2, you can begin entering values.

3. Enter 0.05 for alpha, 2 sided, 220 for Group 1 mean, 200 for Group 2 mean. The difference in means is calculated as 20.

4. Enter 30 for Standard Deviation for Group 1 and Group 2. We are interested in solving for sample size given 90% power so enter 90 in the Power row.

5. This study planned for 4 interim analyses. Including the final analysis this requires Number of Looks to be 5.

6. The looks will be equally spaced and the O'Brien-Fleming spending function is to be used. There will be no truncation of bounds.

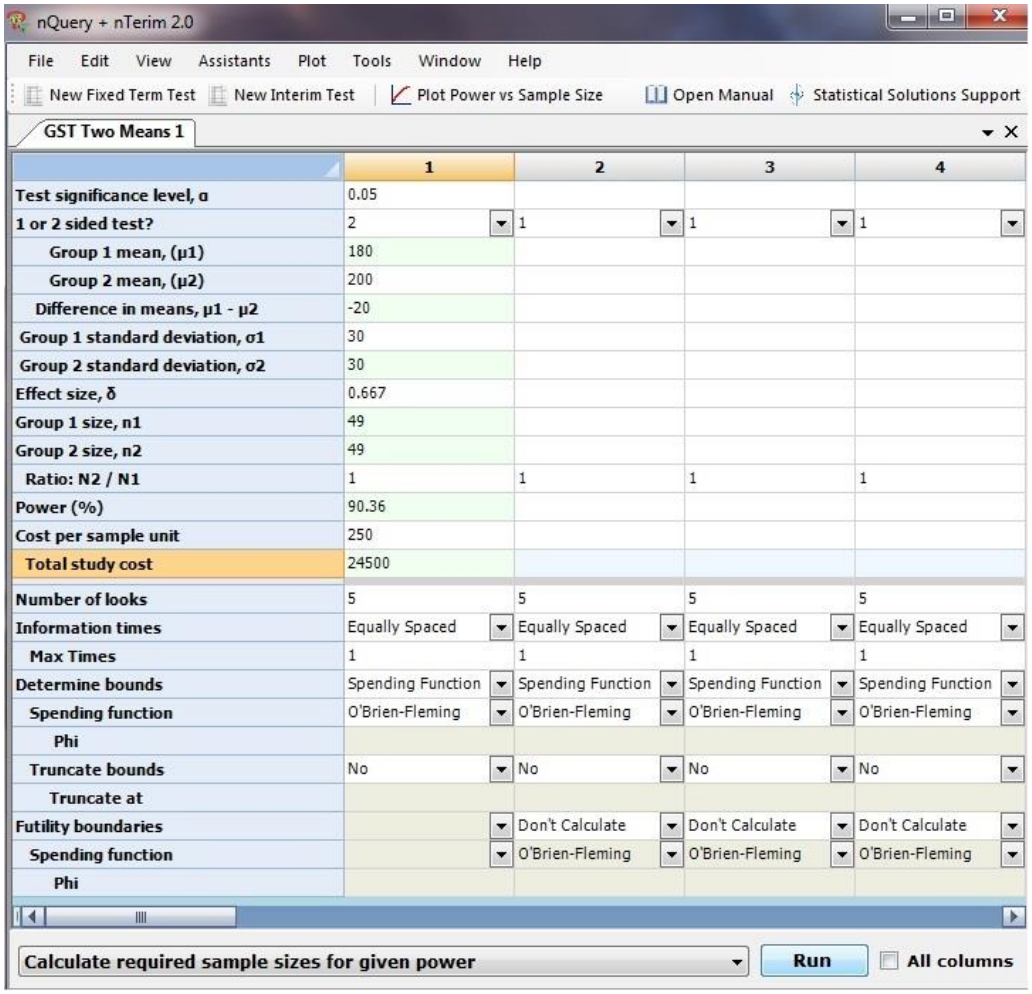


	1	2	3	4
Test significance level, α				
1 or 2 sided test?	1	1	1	1
Group 1 mean, (μ1)				
Group 2 mean, (μ2)				
Difference in means, μ1 - μ2				
Group 1 standard deviation, σ1				
Group 2 standard deviation, σ2				
Effect size, δ				
Group 1 size, n1				
Group 2 size, n2				
Ratio: N2 / N1	1	1	1	1
Power (%)				
Cost per sample unit				
Total study cost				
Number of looks	5	5	5	5
Information times	Equally Spaced	Equally Spaced	Equally Spaced	Equally Spaced
Max Times	1	1	1	1
Determine bounds	Spending Function	Spending Function	Spending Function	Spending Function
Spending function	O'Brien-Fleming	O'Brien-Fleming	O'Brien-Fleming	O'Brien-Fleming
Phi				
Truncate bounds	No	No	No	No
Truncate at				
Futility boundaries	Don't Calculate	Don't Calculate	Don't Calculate	Don't Calculate
Spending function	O'Brien-Fleming	O'Brien-Fleming	O'Brien-Fleming	O'Brien-Fleming
Phi				

Figure 3.1.2. Two Means Test Table

7. It is estimated that the cost per unit is roughly $250, so enter 250 in the Cost per sample unit row.

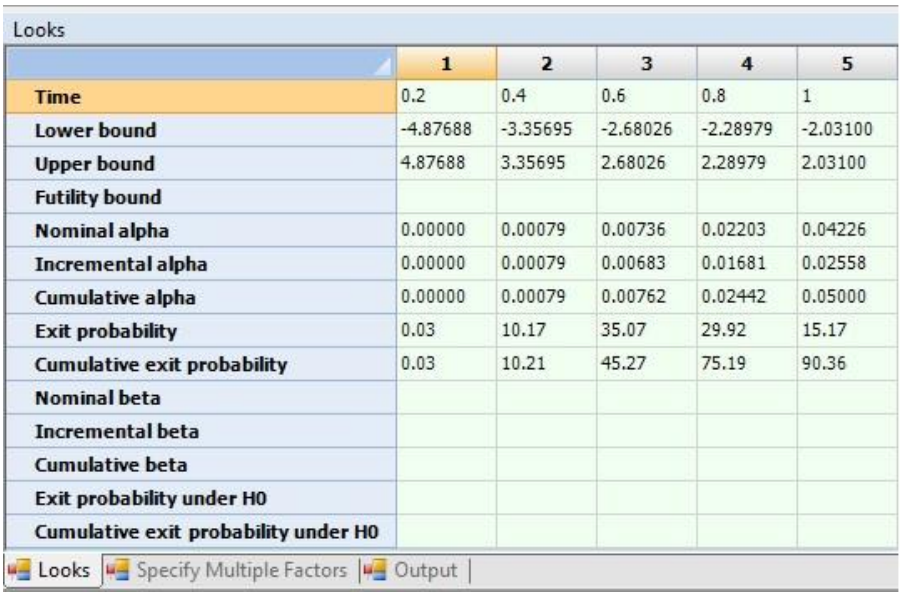
8. Once all the values have been entered, select “Calculate required sample size for given power” from the drop-down menu and click “Run”.



	1	2	3	4
Test significance level, α	0.05			
1 or 2 sided test?	2	1	1	1
Group 1 mean, (μ1)	180			
Group 2 mean, (μ2)	200			
Difference in means, μ1 - μ2	-20			
Group 1 standard deviation, σ1	30			
Group 2 standard deviation, σ2	30			
Effect size, δ	0.667			
Group 1 size, n1	49			
Group 2 size, n2	49			
Ratio: N2 / N1	1	1	1	1
Power (%)	90.36			
Cost per sample unit	250			
Total study cost	24500			
Number of looks	5	5	5	5
Information times	Equally Spaced	Equally Spaced	Equally Spaced	Equally Spaced
Max Times	1	1	1	1
Determine bounds	Spending Function	Spending Function	Spending Function	Spending Function
Spending function	O'Brien-Fleming	O'Brien-Fleming	O'Brien-Fleming	O'Brien-Fleming
Phi				
Truncate bounds	No	No	No	No
Truncate at				
Futility boundaries		Don't Calculate	Don't Calculate	Don't Calculate
Spending function		O'Brien-Fleming	O'Brien-Fleming	O'Brien-Fleming
Phi				

Figure 3.1.3. Completed Two Means Test Table

9. The boundaries calculated are shown in Figure 3.1.4.



	1	2	3	4	5
Time	0.2	0.4	0.6	0.8	1
Lower bound	-4.87688	-3.35695	-2.68026	-2.28979	-2.03100
Upper bound	4.87688	3.35695	2.68026	2.28979	2.03100
Futility bound					
Nominal alpha	0.00000	0.00079	0.00736	0.02203	0.04226
Incremental alpha	0.00000	0.00079	0.00683	0.01681	0.02558
Cumulative alpha	0.00000	0.00079	0.00762	0.02442	0.05000
Exit probability	0.03	10.17	35.07	29.92	15.17
Cumulative exit probability	0.03	10.21	45.27	75.19	90.36
Nominal beta					
Incremental beta					
Cumulative beta					
Exit probability under H0					
Cumulative exit probability under H0					

Figure 3.1.4. Boundary Table for Two Means Test

10. Finally, the boundaries calculated in the table in Figure 3.1.4 are automatically plotted as illustrated in Figure 3.1.5.

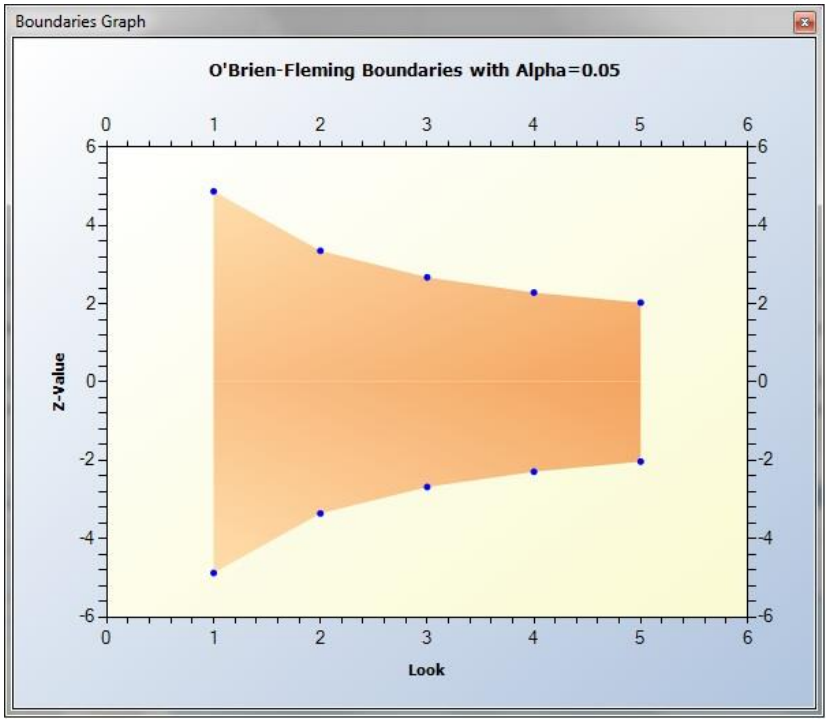


Figure 3.1.5. Boundary Plot for Two Means Test

By clicking on the Output tab at the bottom of the screen you can see a statement giving details of the calculation:

"Sample sizes of 49 in group 1 and 49 in group 2 are required to achieve 90.36% power to detect a difference in means of 20 (the difference between group 1 mean, $\mu 1$, of 220 and group 2 mean, $\mu 2$, of 200) assuming that the common standard deviation is 30 using a 2-sided z-test with 0.05 significance level. These results assume that 5 sequential tests are made and the O'Brien-Fleming spending function is used to determine the test boundaries. Drift: 3.29983"

Example 2: Pocock Spending Function and Unequal N's

This example is taken from Reboussin et al (1992) using the Pocock spending function.

1. Open nTerim through the Start Menu or by double clicking on the nTerim desktop icon. Then click on "New Interim Test" from the menu bar at the top of the window. A "Study Goal and Design" window will appear as shown below. Select the options as mapped out in Figure 3.1.6, then Click "OK".

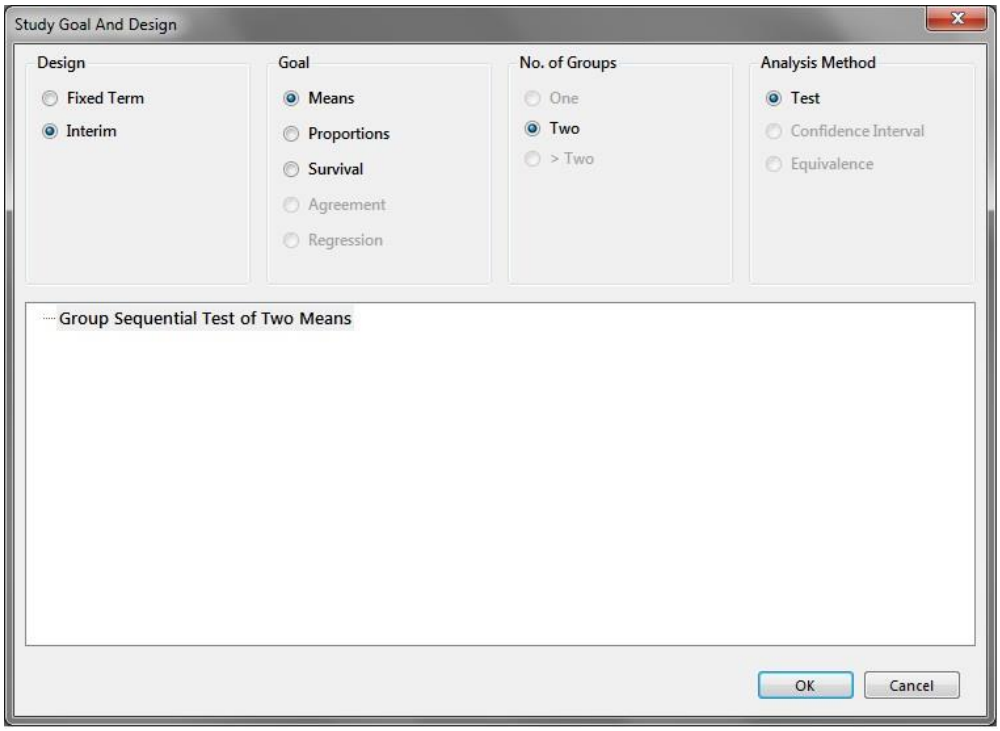
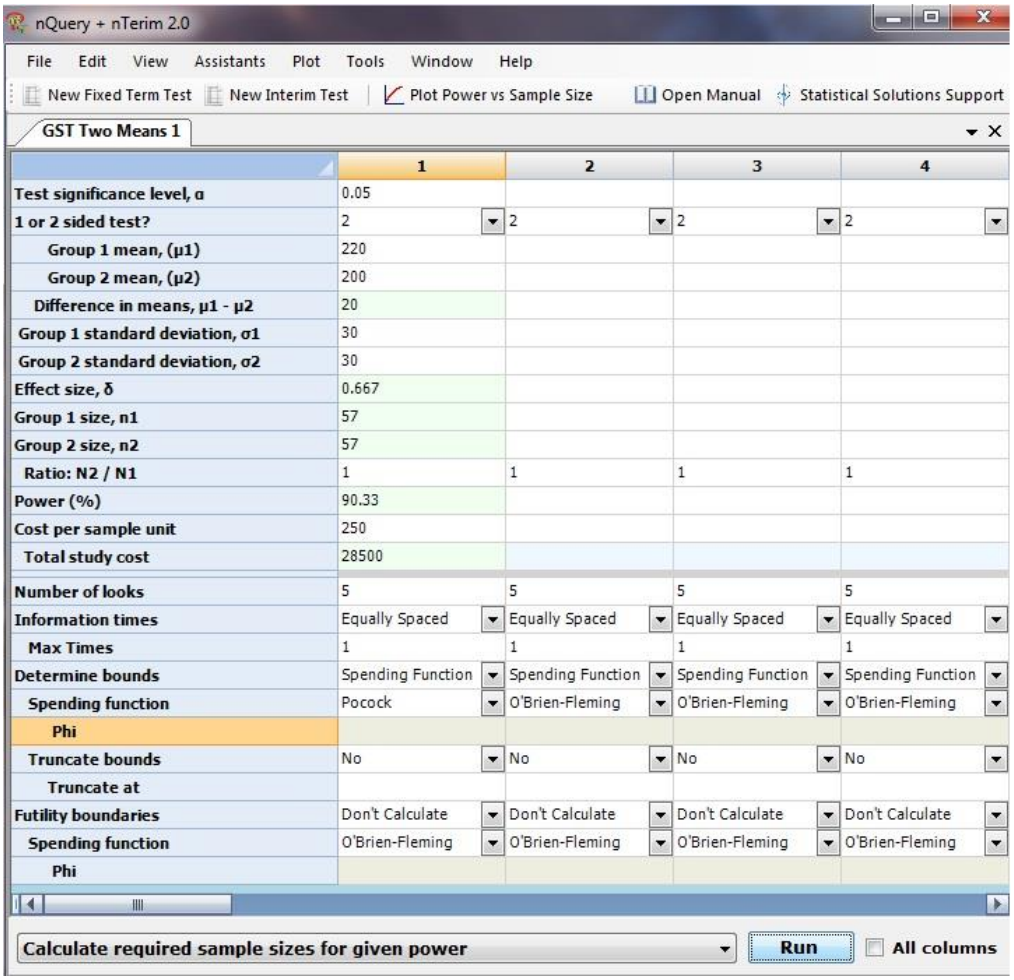


Figure 3.1.6. Study Goal and Design Window

2. Setup the table as in the Example 1.

3. We will again use 5 looks but this time change the Spending Function to Pocock in the dropdown box.



	1	2	3	4
Test significance level, α	0.05			
1 or 2 sided test?	2	2	2	2
Group 1 mean, (μ1)	220			
Group 2 mean, (μ2)	200			
Difference in means, μ1 - μ2	20			
Group 1 standard deviation, σ1	30			
Group 2 standard deviation, σ2	30			
Effect size, δ	0.667			
Group 1 size, n1	57			
Group 2 size, n2	57			
Ratio: N2 / N1	1	1	1	1
Power (%)	90.33			
Cost per sample unit	250			
Total study cost	28500			
Number of looks	5	5	5	5
Information times	Equally Spaced	Equally Spaced	Equally Spaced	Equally Spaced
Max Times	1	1	1	1
Determine bounds	Spending Function	Spending Function	Spending Function	Spending Function
Spending function	Pocock	O'Brien-Fleming	O'Brien-Fleming	O'Brien-Fleming
Phi				
Truncate bounds	No	No	No	No
Truncate at				
Futility boundaries	Don't Calculate	Don't Calculate	Don't Calculate	Don't Calculate
Spending function	O'Brien-Fleming	O'Brien-Fleming	O'Brien-Fleming	O'Brien-Fleming
Phi				

Figure 3.1.7. Complete Two Means Test Table

4. Select Run and the sample size along with the boundary values will be calculated.

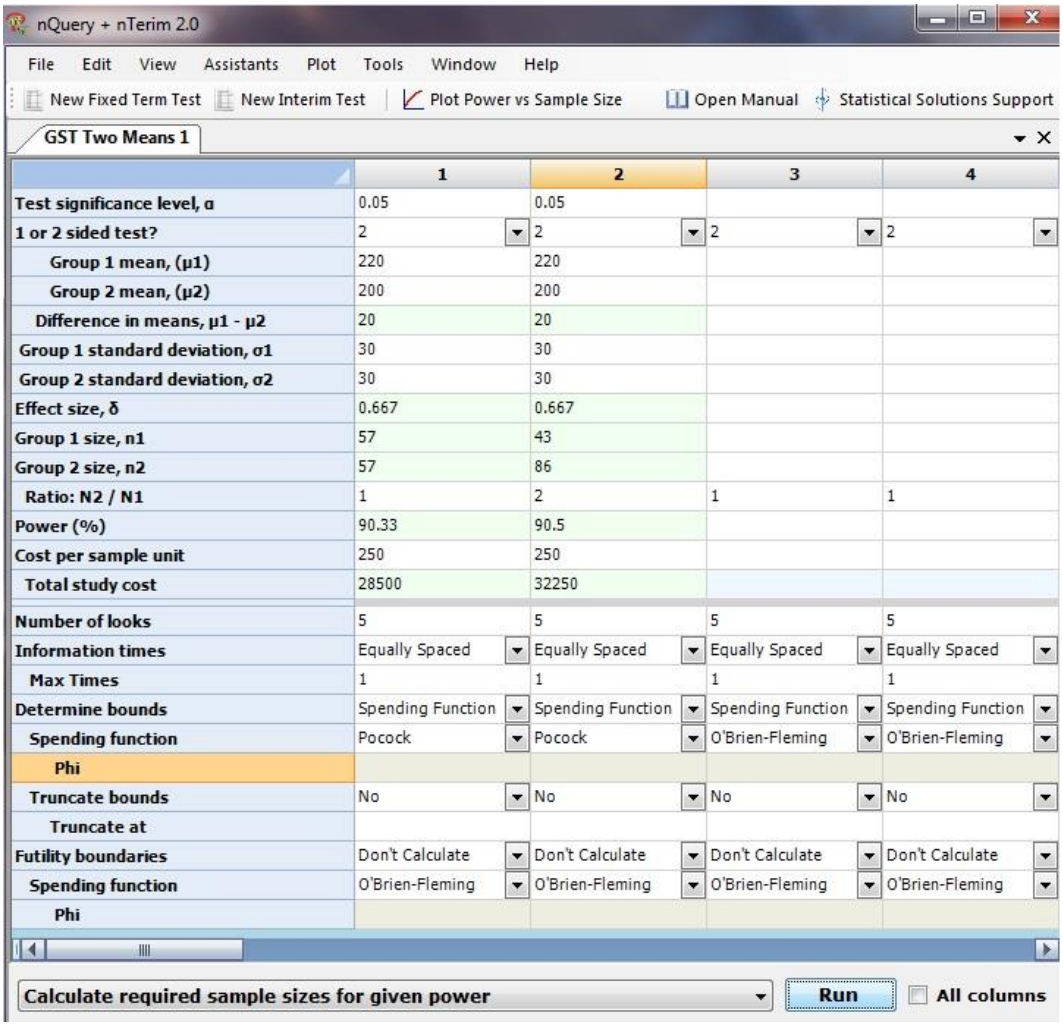
5. The boundaries that are calculated will be automatically plotted.

Clicking on the Output tab at the bottom of the screen you can see a statement giving details of the calculation:

"Sample sizes of 57 in group 1 and 57 in group 2 are required to achieve 90.33% power to detect a difference in means of 20 (the difference between group 1 mean, $\mu 1$, of 220 and group 2 mean, $\mu 2$, of 200) assuming that the common standard deviation is 30 using a 2-sided z-test with 0.05 significance level. These results assume that 5 sequential tests are made and the Pocock spending function is used to determine the test boundaries. Drift: 3.55903"

6. In the main table, in Column 2, enter the same parameter values again except enter a value of 2 for the Ratio parameter. Don't forget to change the spending function to Pocock.

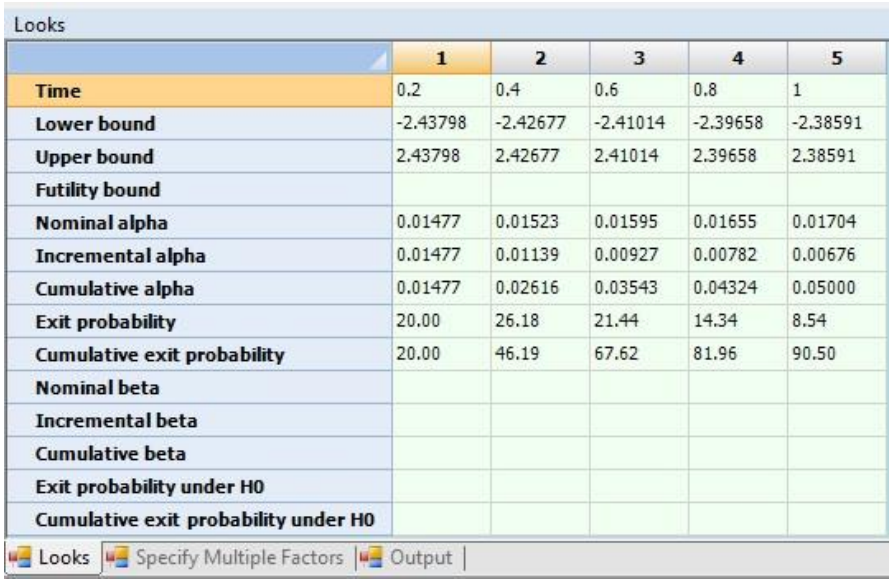
7. Select Run and the sample size will be re-calculated as shown in Figure 3.1.8 below.



	1	2	3	4
Test significance level, α	0.05	0.05		
1 or 2 sided test?	2	2	2	2
Group 1 mean, (μ1)	220	220		
Group 2 mean, (μ2)	200	200		
Difference in means, μ1 - μ2	20	20		
Group 1 standard deviation, σ1	30	30		
Group 2 standard deviation, σ2	30	30		
Effect size, δ	0.667	0.667		
Group 1 size, n1	57	43		
Group 2 size, n2	57	86		
Ratio: N2 / N1	1	2	1	1
Power (%)	90.33	90.5		
Cost per sample unit	250	250		
Total study cost	28500	32250		
Number of looks	5	5	5	5
Information times	Equally Spaced	Equally Spaced	Equally Spaced	Equally Spaced
Max Times	1	1	1	1
Determine bounds	Spending Function	Spending Function	Spending Function	Spending Function
Spending function	Pocock	Pocock	O'Brien-Fleming	O'Brien-Fleming
Phi				
Truncate bounds	No	No	No	No
Truncate at				
Futility boundaries	Don't Calculate	Don't Calculate	Don't Calculate	Don't Calculate
Spending function	O'Brien-Fleming	O'Brien-Fleming	O'Brien-Fleming	O'Brien-Fleming
Phi				

Figure 3.1.8. Comparison of two separate Means Tests

8. Also the boundary values will be recalculated and boundary plot will automatically be plotted as shown in Figure 3.1.9 and 3.1.10 below.



	1	2	3	4	5
Time	0.2	0.4	0.6	0.8	1
Lower bound	-2.43798	-2.42677	-2.41014	-2.39658	-2.38591
Upper bound	2.43798	2.42677	2.41014	2.39658	2.38591
Futility bound					
Nominal alpha	0.01477	0.01523	0.01595	0.01655	0.01704
Incremental alpha	0.01477	0.01139	0.00927	0.00782	0.00676
Cumulative alpha	0.01477	0.02616	0.03543	0.04324	0.05000
Exit probability	20.00	26.18	21.44	14.34	8.54
Cumulative exit probability	20.00	46.19	67.62	81.96	90.50
Nominal beta					
Incremental beta					
Cumulative beta					
Exit probability under H0					
Cumulative exit probability under H0					

Figure 3.1.9. Boundary Table for Column 2

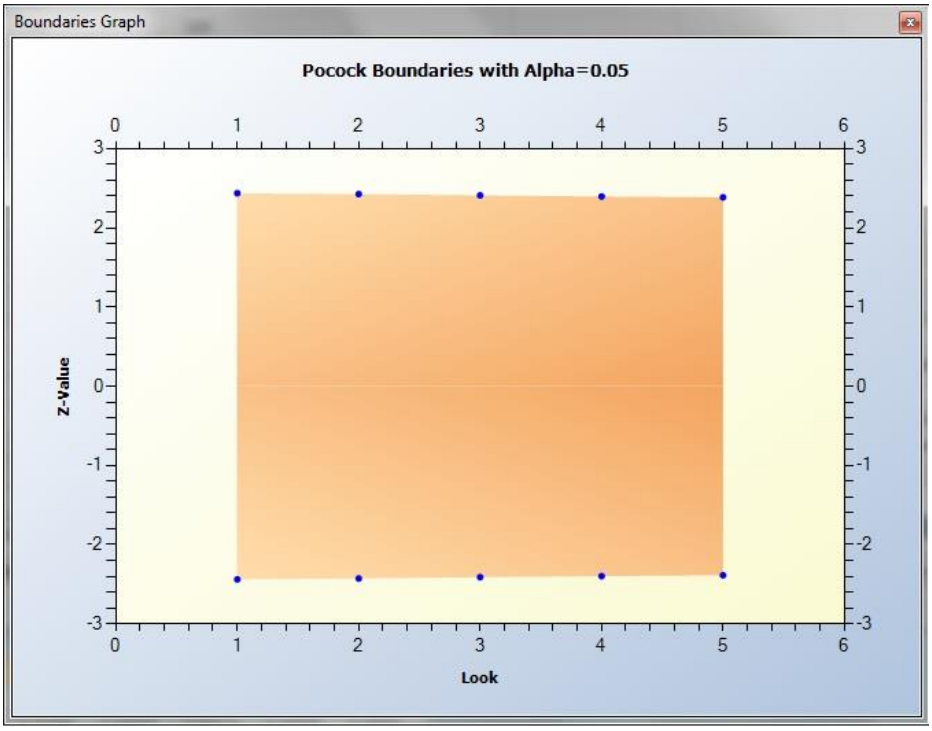


Figure 3.1.10. Boundary Plot for Column 2

Likewise, by clicking on the Output tab at the bottom of the screen you can see a statement giving details of the calculation:

Output Statement – Column 2:
"Sample sizes of 43 in group 1 and 86 in group 2 are required to achieve 90.5% power to detect a difference in means of 20 (the difference between group 1 mean, $\mu 1$, of 220 and group 2 mean, $\mu 2$, of 200) assuming that the common standard deviation is 30 using a 2-sided z-test with 0.05 significance level. These results assume that 5 sequential tests are made and the Pocock spending function is used to determine the test boundaries. Drift: 3.56942"

3.2 Two Proportions

3.2.1. Introduction

nTerim 4.0 is designed for the calculation of Power and Sample Size for both Fixed Period and Group Sequential design. In relation to Group Sequential designs, calculations are performed using the Lan-DeMets alpha spending function approach (DeMets & Lan, 1984; DeMets & Lan, 1994) for estimating boundary values. Using this approach, boundary values can be estimated for O'Brien-Fleming (O'Brien & Fleming, 1979), Pocock (Pocock, 1977), Hwang-Shih-DeCani (Hwang, Shih & DeCani, 1990) and the Power family of spending functions. Calculations follow the approach of Reboussin et al (1992) and Jennsion & Turnbull (2000). Calculations can be performed for studies that involve comparisons of means, comparisons of proportions and survival studies as well as early stopping for Futility.

Group Sequential Designs

Group Sequential designs differ from Fixed Period designs in that the data from the trial is analyzed at one or more stages prior to the conclusion of the trial. As a result the alpha value applied at each analysis or `look' must be adjusted to preserve the overall Type 1 error. The alpha values used at each look are calculated based upon the spending function chosen, the number of looks to be taken during the course of the study as well as the overall Type 1 error rate. For a full introduction to group sequential methods see Jennison & Turnbull (2000) and Chow et al (2008).

Spending Function

There are four alpha and beta spending functions available to the user in nTerim 4.0 as well as an option to manually input boundary values. As standard all alpha spending functions have the properties that $\alpha(0) = 0$ and $\alpha(1) = \alpha$. Similarly, all beta spending functions have the properties that $\beta(0) = 0$ and $\beta(1) = \beta$. Functionally the alpha and beta spending functions are the same. In Table 3.2.1 we list the alpha spending functions available in nTerim 4.0.

Table 3.2.1. Spending Function Equations

O'Brien-Fleming	$\alpha(\tau) = 2\left(1 - \Phi\left(\frac{z_{\alpha/2}}{\sqrt{\tau}}\right)\right)$
Pocock	$\alpha(\tau) = \alpha ln(1 + (e - 1)\tau)$
Power	$\alpha(\tau) = \alpha\tau^{\Phi}, \Phi > 0$
Hwang-Shih-DeCani	$\alpha(\tau) = \alpha\left[\frac{(1 - e^{-\Phi\tau})}{(1 - e^{-\Phi})}\right], \Phi \neq 0$

The parameter τ represents the time elapsed in the trial. This can either be as a proportion of the overall time elapsed or a proportion of the sample size enrolled.

The common element among most of the different spending functions is to use lower error values for the earlier looks. By doing this it means that the results of any analysis will only be considered significant in an early stage if it gives an extreme result.

Boundaries

The boundaries in nTerim 4.0 represent the critical values at each look. These boundaries are constructed using the alpha and beta spending functions. Users in nTerim 4.0 are given the option to generate boundaries for early rejection of the null hypothesis, H_0, using the alpha spending function, or to generate boundaries for early rejection of either the null or alternative hypothesis, H_0 or H_1, using a combination of both the alpha and beta spending functions. The notion of using an alpha spending function approach to generate stopping boundaries for early rejection of H_0 was first proposed by Lan and DeMets (1983), we refer to such boundaries in nTerim 4.0 as efficacy boundaries. Building on the work of Lan and DeMets, Pampallona, Tsiatis, and Kim (1995, 2001) later put forward the concept of using a beta spending approach to construct boundaries for early rejection of H_1, we refer to these boundaries in nTerim as futility boundaries.

Essentially, if a test statistic crosses an efficacy boundary then it can be concluded that the experimental treatment shows a statistically significant effect, the trial can be stopped with rejection of the null hypothesis. If the test statistic crosses a futility boundary then this indicates with high probability that an effect will not be found, that the trial can be terminated by rejecting the alternative hypothesis.

In the case where the user wishes to generate boundaries for early rejection of either the null or alternative hypothesis, H_o or H_1, they are given two options; either to have the boundaries binding, or non-binding. With binding boundaries, if the test statistic crosses the futility boundary, the test must be stopped, otherwise the type-1 error may become inflated. The reason for this is that there is an interaction between the efficacy and futility boundaries in their calculation that could cause the efficacy boundary to shift. In the case of non-binding boundaries; the efficacy boundaries are calculated as normal, that is, as if the futility boundaries did not exist. This eliminates the danger of inflating the type-1 error when the futility boundary is overruled. The downside of the non-binding case is that it may increase the required sample size relative to the binding case.

The boundaries calculated in nTerim 4.0 follow the procedures outlined by Reboussin et al (1992) and Jennison & Turnbull (2000).

3.2.2. Methodology

The variables are defined as

Symbol	Description
α	Probability of Type I error
β	Probability of Type II error
$1-\beta$	Power of the Test
p_1, p_2	Group Means
σ_1, σ_2	Group Standard Deviations
N_1, N_2	Group Sample Sizes
R	Ratio of N_1 to N_2
Θ	Drift Parameter
K	Number of Time-points (Looks)
α^*	Spending Function (O'Brien-Fleming, Pocock, etc.)

Calculate Sample Sizes for a given Power

Using the number of time-points (K), number of sides, type of spending function, the hypothesis to be rejected, the type 1 error, α, and power $(1-\beta)$, the drift parameter Θ can be obtained using algorithms by Reboussin et al (1992) and Jennison & Turnbull (2000). The test statistic is defined as

$$\Theta = \frac{|p_1 - p_2|}{\sqrt{\frac{\bar{p}(1-\bar{p})}{N_1} + \frac{\bar{p}(1-\bar{p})}{N_2}}} \qquad \{3.2.1\}$$

where $\bar{p} = \frac{N_1 p_1 + N_2 p_2}{N_1 + N_2}$. The user supplies the proportions (p_1, p_2), and either R, N_1 or N_2. Since $R = \frac{N_2}{N_1}$ it follows that a value of $R = 1$ indicates equal sample sizes and that $\bar{p} = \frac{p_1 + R * p_2}{1+R}$. The approach to solving this problem is dependent on what information the user supplies. For the case of continuity correction the formula can be written as:

$$\Theta = \frac{|p_1 - p_2| - \frac{1}{2N_1}(\frac{R+1}{R})}{\sqrt{\frac{\bar{p}(1-\bar{p})(R+1)}{N_1 R}}} \qquad \{3.2.2\}$$

as per Fleiss (1981). The validity of this formula relies on the assumption of minimum expected cell count being above a pre-specified threshold. As a rule of thumb, the normal approximation to the binomial will hold if the following conditions are met:

$p_1 N_1 > T$
$(1-p_1)N_1 > T$
$p_2 N_2 > T$
$(1-p_2)N_2 > T$

$\{3.2.3\}$

where T is a predefined threshold.

User supplies R only
The requirement is to obtain N_1 and N_2. Using that $N_2 = R \times N_1$ the result from Equation {3.2.2} obtained is:

$$N_1 = \frac{\Theta^2(R\bar{p}(1-\bar{p}) + \bar{p}(1-\bar{p})}{R(p_1 - p_2)^2} \qquad \{3.2.4\}$$

The steps involved are:

- Obtain Θ
- Solve Equation {3.2.4} for N_1 and $N_2 = R \times N_1$

User supplies R only and selects Continuity Correction
If the user has selected to use the continuity correction then apply the formula from Fleiss et al (1980).

$$N_{1cc} = \frac{N_1}{4}\left(1 + \sqrt{1 + \frac{2(R+1)}{R(N_1)|p_1 - p_2|}}\right)^2 \qquad \{3.2.5\}$$

to obtain N_{1cc}. It follows that N_{2cc} is then $R \times N_{1cc}$. If the user has NOT selected to use continuity correction then $N_1 = N_1$ and $N_2 = R \times N_1$.

User specifies N_1 only or N_2 only
When the user specifies N_1, then Equation {3.2.1} can be re-expressed as a quadratic in N_2 from which two roots are obtained, one less than and one greater than N_1. Similarly, if N_2 is specified the roots gives the values of N_1.

Calculate Attainable Power with the given Sample Sizes
Given α, N_1, proportions (p_1, p_2), R $(or\ N_2)$, time-points and type of spending function, the requirement is to obtain the power.

If the user has NOT selected to use continuity correction
The steps are:

- Obtain Θ by solving Equation {3.2.1} (given that $N_1, R, p_1, p_2, \bar{p}$ are known)
- Obtain power using the algorithm by Reboussin et al (1992) and Jennison & Turnbull (2000)

If the user has selected to use continuity correction
The steps are:

- Obtain Θ by solving Equation {3.2.2} (given that $N_1, R, p_1, p_2, \bar{p}$ are known)
- Obtain power using the algorithm by Reboussin et al (1992) and Jennison & Turnbull (2000)

Calculate missing proportion given N_1, N_2, α, power and the other proportion.

Calculate p_1 given p_2
In order to solve for p_1 given p_2 and all other information Equation {3.2.1} can be re-expressed as a quadratic with respect to p_2 the roots of which give p_1. Similarly if $p_1$is specified the roots give the values of p_2.

Calculate p_1 given $p_2$with Continuity Correction
In order to solve for p_1 given p_2 and all other information Equation {3.2.2} can be re-expressed as a quadratic with respect to p_2 the roots of which give p_1. Similarly if $p_1$is specified the roots give the values of p_2.

3.2.3. Examples

Example 1: Pocock Spending Function
This example is adopted from Reboussin et al (1992) using Pocock spending function.

1. Open nTerim through the Start Menu or by double clicking on the nTerim desktop icon. Then click on "New Interim Test" from the menu bar at the top of the window. A "Study Goal and Design" window will appear as shown below. Select the options as mapped out in Figure 3.2.1, then Click "OK".

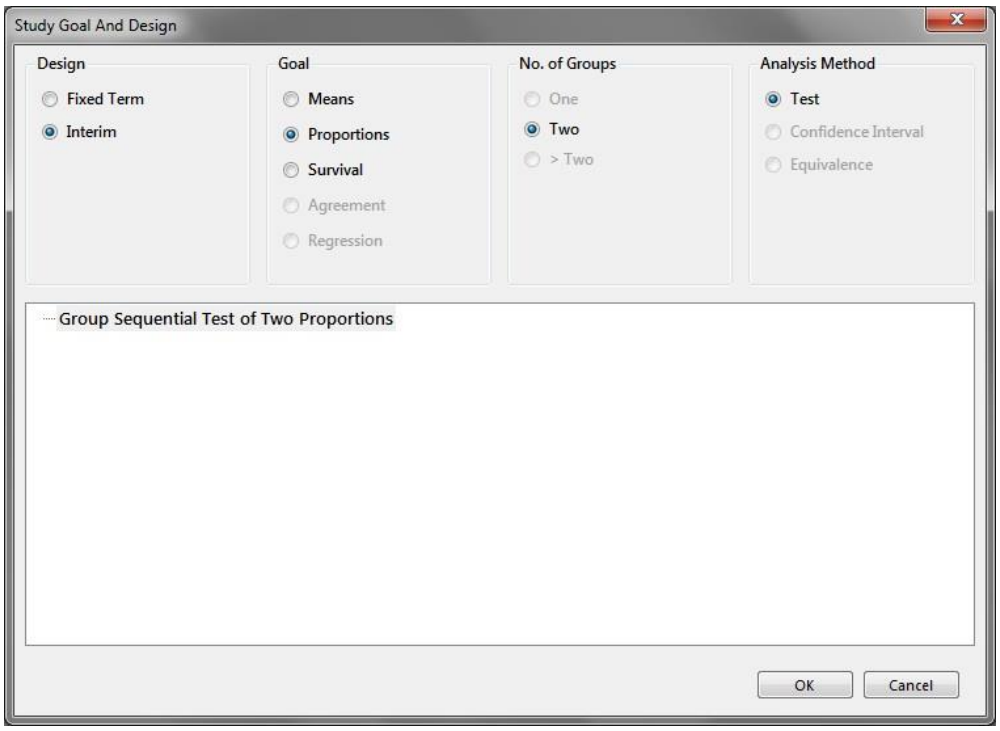


Figure 3.2.1. Study Goal and Design Window

2. Now you have opened the test table, as illustrated in Figure 3.2.2, you can begin entering values.

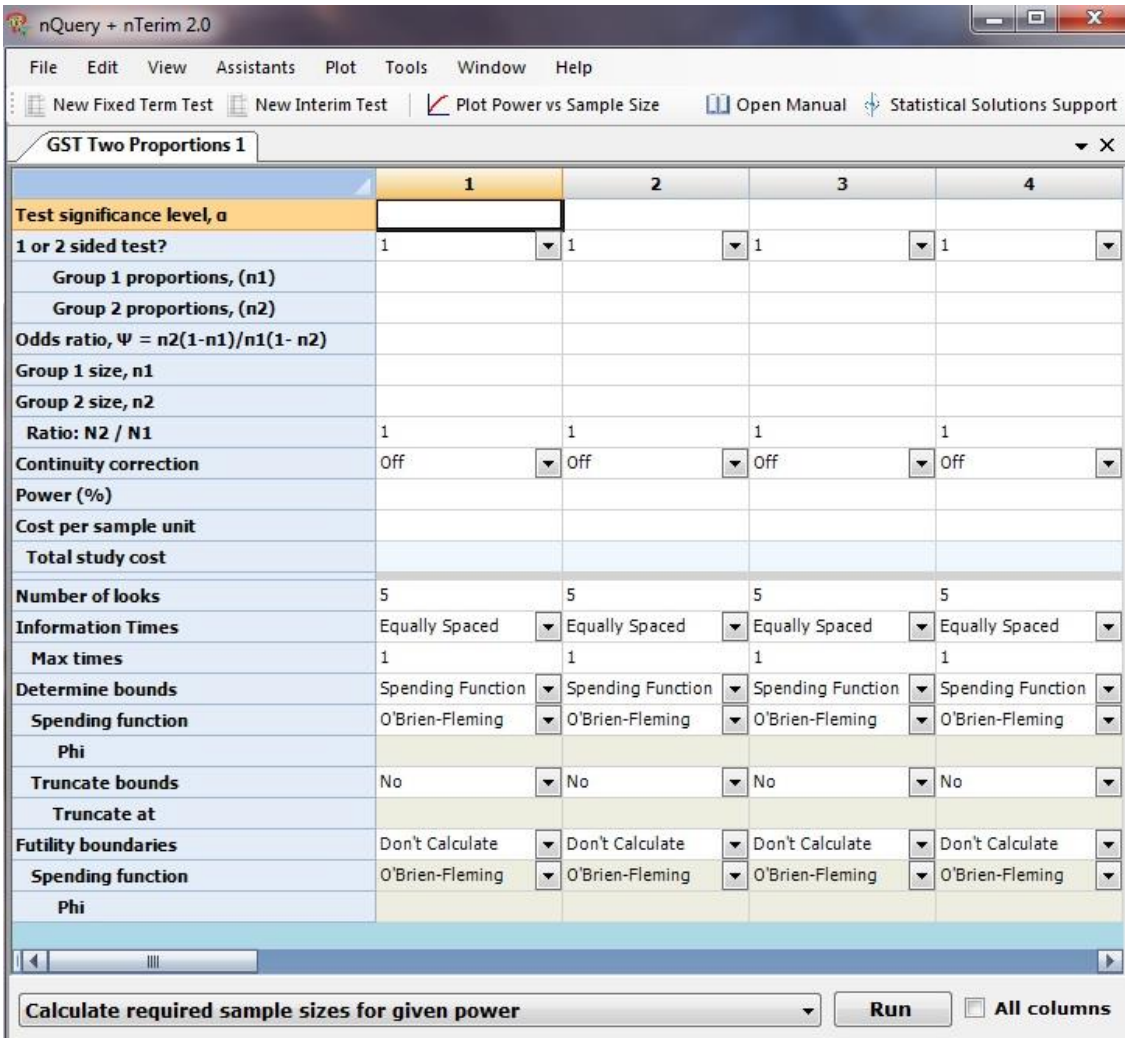
3. Enter 0.05 for alpha, 2 sided, 0.4 for Group 1 proportion, 0.6 for Group 2 proportion. The odds ratio is calculated as 2.25.

4. Select Off for the Continuity Correction. We are interested in solving for sample size given 90% power so enter 90 in the Power row.

5. This study planned for 4 interim analyses. Including the final analysis this requires Number of Looks to be 5.

6. The looks will be equally spaced and the Pocock spending function is to be used. There will be no truncation of bounds.

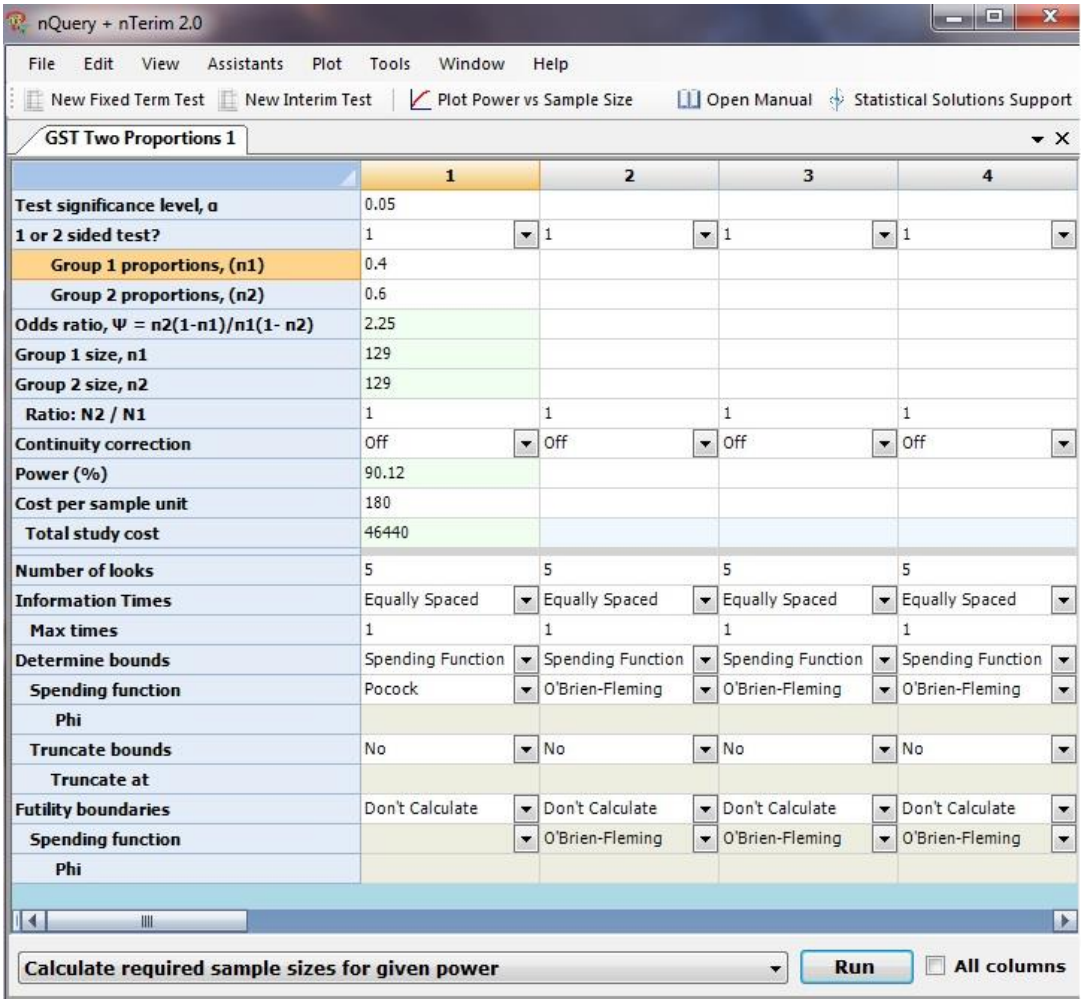
7. It is estimated that the cost per unit is roughly $180, so enter 180 in the Cost per sample unit row.



	1	2	3	4
Test significance level, α				
1 or 2 sided test?	1	1	1	1
Group 1 proportions, (n1)				
Group 2 proportions, (n2)				
Odds ratio, Ψ = n2(1-n1)/n1(1- n2)				
Group 1 size, n1				
Group 2 size, n2				
Ratio: N2 / N1	1	1	1	1
Continuity correction	Off	Off	Off	Off
Power (%)				
Cost per sample unit				
Total study cost				
Number of looks	5	5	5	5
Information Times	Equally Spaced	Equally Spaced	Equally Spaced	Equally Spaced
Max times	1	1	1	1
Determine bounds	Spending Function	Spending Function	Spending Function	Spending Function
Spending function	O'Brien-Fleming	O'Brien-Fleming	O'Brien-Fleming	O'Brien-Fleming
Phi				
Truncate bounds	No	No	No	No
Truncate at				
Futility boundaries	Don't Calculate	Don't Calculate	Don't Calculate	Don't Calculate
Spending function	O'Brien-Fleming	O'Brien-Fleming	O'Brien-Fleming	O'Brien-Fleming
Phi				

Figure 3.2.2. Two Proportions Test Table

8. Once all the values have been entered, select “Calculate required sample size for given power” from the drop-down menu and click “Run”.



	1	2	3	4
Test significance level, α	0.05			
1 or 2 sided test?	1	1	1	1
Group 1 proportions, (n1)	0.4			
Group 2 proportions, (n2)	0.6			
Odds ratio, Ψ = n2(1-n1)/n1(1- n2)	2.25			
Group 1 size, n1	129			
Group 2 size, n2	129			
Ratio: N2 / N1	1	1	1	1
Continuity correction	Off	Off	Off	Off
Power (%)	90.12			
Cost per sample unit	180			
Total study cost	46440			
Number of looks	5	5	5	5
Information Times	Equally Spaced	Equally Spaced	Equally Spaced	Equally Spaced
Max times	1	1	1	1
Determine bounds	Spending Function	Spending Function	Spending Function	Spending Function
Spending function	Pocock	O'Brien-Fleming	O'Brien-Fleming	O'Brien-Fleming
Phi				
Truncate bounds	No	No	No	No
Truncate at				
Futility boundaries	Don't Calculate	Don't Calculate	Don't Calculate	Don't Calculate
Spending function		O'Brien-Fleming	O'Brien-Fleming	O'Brien-Fleming
Phi				

Figure 3.2.3. Completed Two Proportions Test Table

9. The boundaries calculated are shown in Figure 3.2.4.

Looks

	1	2	3	4	5
Time	0.2	0.4	0.6	0.8	1
Lower bound	-8.00000	-8.00000	-8.00000	-8.00000	-8.00000
Upper bound	2.17621	2.14371	2.11322	2.08952	2.07091
Futility bound					
Nominal alpha	0.01477	0.01603	0.01729	0.01833	0.01918
Incremental alpha	0.01477	0.01139	0.00927	0.00782	0.00676
Cumulative alpha	0.01477	0.02616	0.03543	0.04324	0.05000
Exit probability	22.98	25.92	20.00	13.21	8.00
Cumulative exit probability	22.98	48.90	68.90	82.11	90.12
Nominal beta					
Incremental beta					
Cumulative beta					
Exit probability under H0					
Cumulative exit probability under H0					

Looks | Specify Multiple Factors | Output

Figure 3.2.4. Boundary Table for Pocock Spending Function

10. Finally, the boundaries calculated in the table in Figure 3.2.4 are automatically plotted as illustrated in Figure 3.2.5.

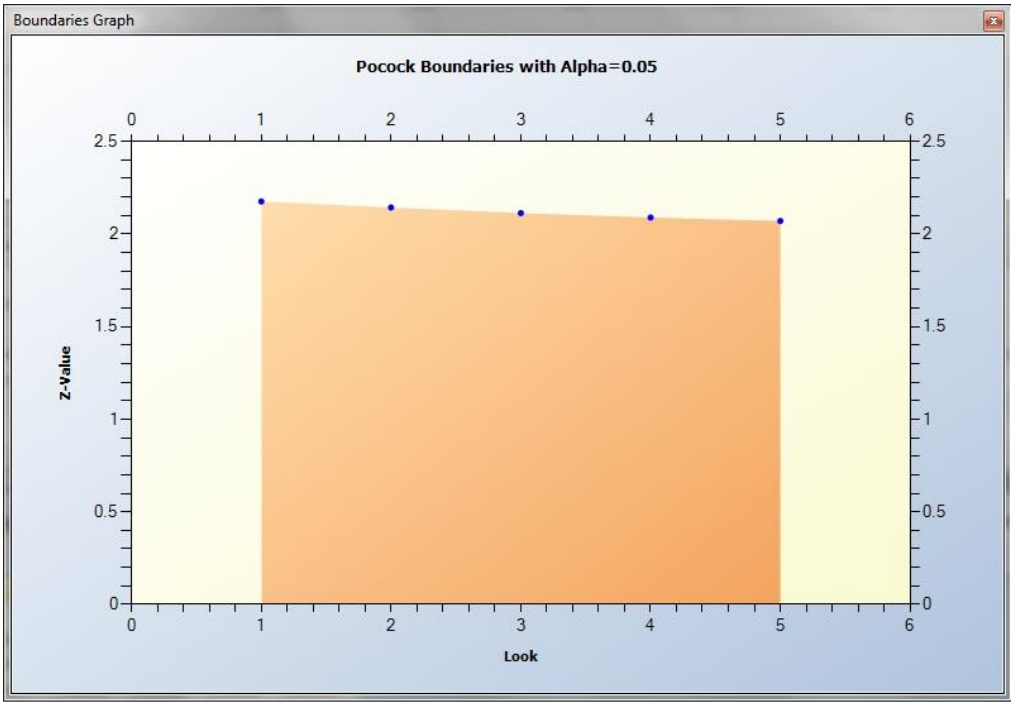


Figure 3.2.5. Boundary Plot for Two Proportions (one-sided) Test

By clicking on the Output tab at the bottom of the screen you can see a statement giving details of the calculation:

"Sample sizes of at least 2237 in group 1 and 2237 in group 2 are required to achieve 80.09% power to detect an odds ratio of 0.72752 (for proportions of 0.11 in group 1 and 0.083 in group 2) using a 2-sided continuity corrected χ^2 test with 0.05 significance level. These results assume that 4 sequential tests are made and the Pocock spending function is used to determine the test boundaries."

Example 2: Power Family spending function with truncated bounds

This example is an adaptation from Reboussin et al (1992) using Power Family spending function with truncated bounds.

1. Open nTerim through the Start Menu or by double clicking on the nTerim desktop icon. Then click on "New Interim Test" from the menu bar at the top of the window. A "Study Goal and Design" window will appear as shown below. Select the options as mapped out in Figure 3.2.6, then Click "OK".

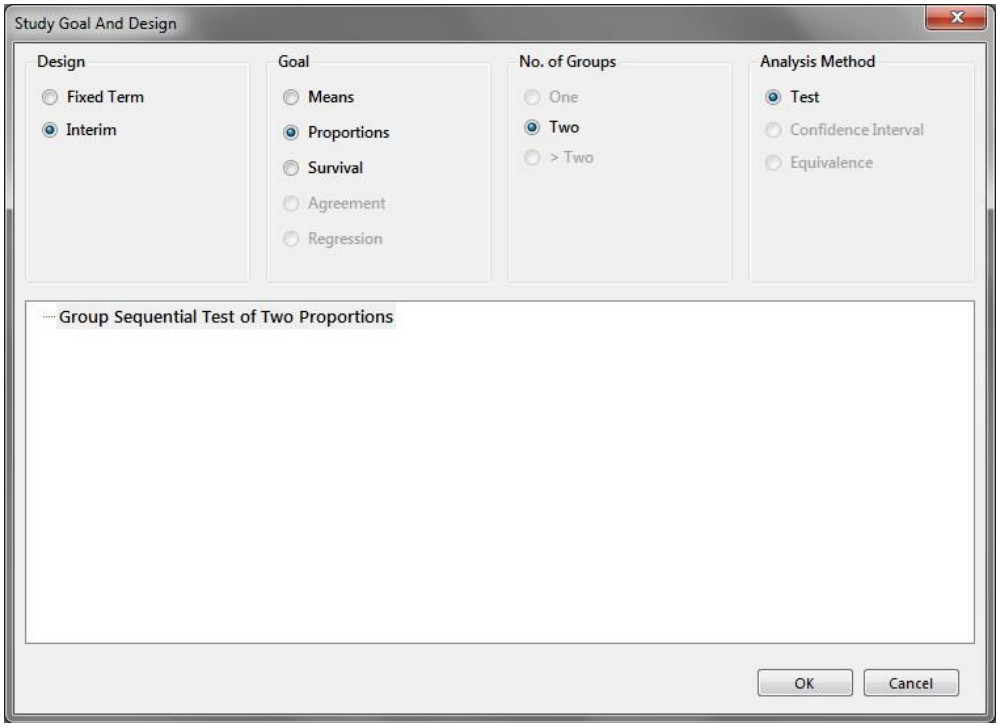


Figure 3.2.6. Study Goal and Design Window

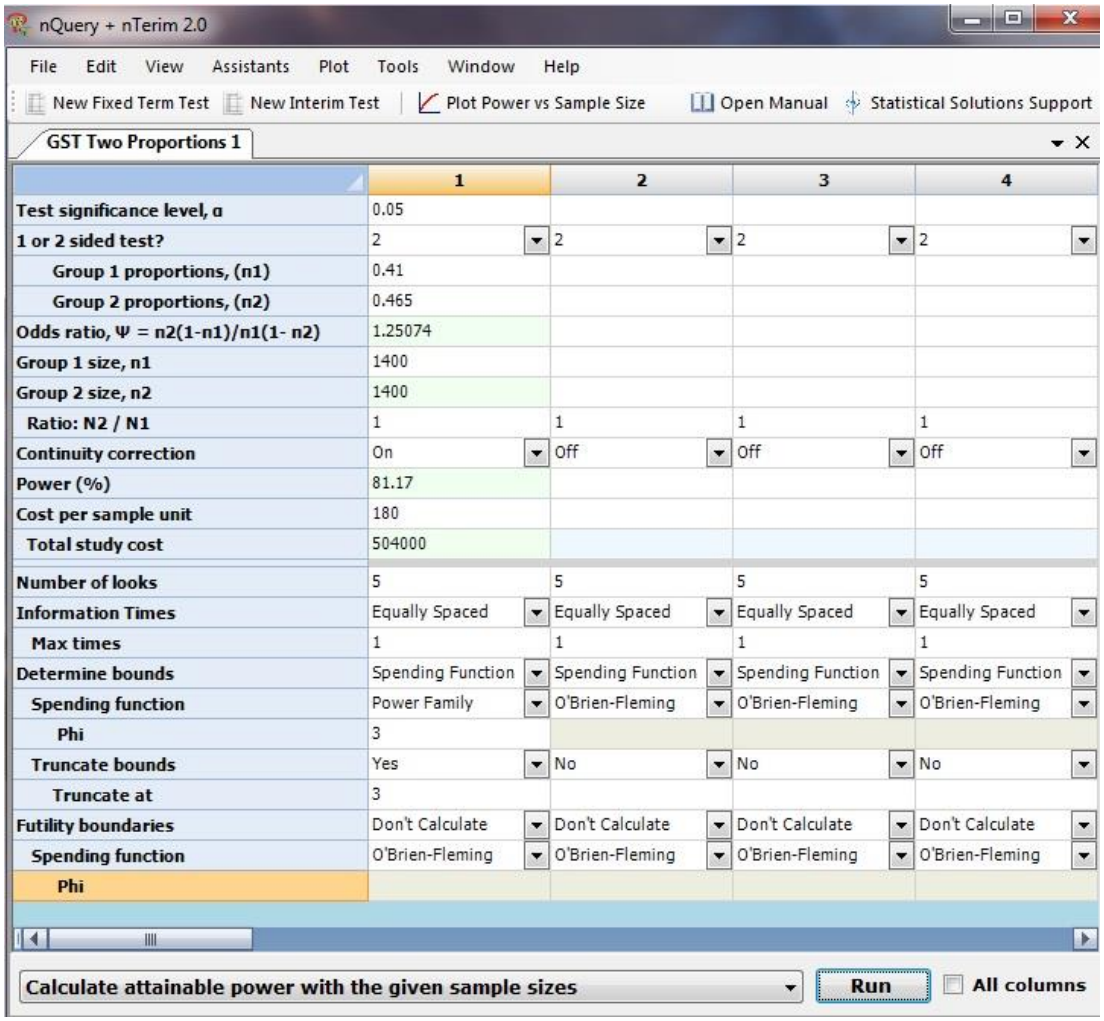
2. Enter 0.05 for alpha, 2 sided, 0.41 for Group 1 proportion, 0.465 for Group 2 proportion. The odds ratio is calculated as 1.25074.

3. Select On for the Continuity Correction. We are interested in solving for power given a sample size of 1400 per group so enter 1400 in the Group 1 size row.

4. This study planned for 4 interim analyses. Including the final analysis this requires Number of Looks to be 5.
5. The looks will be equally spaced and the Power Family spending function is to be used. Enter 3 for Phi.

6. For this example we want to truncate the boundaries so as not to be over-conservative. Enter Yes for truncate bounds and then enter 3 for the value to truncate at.

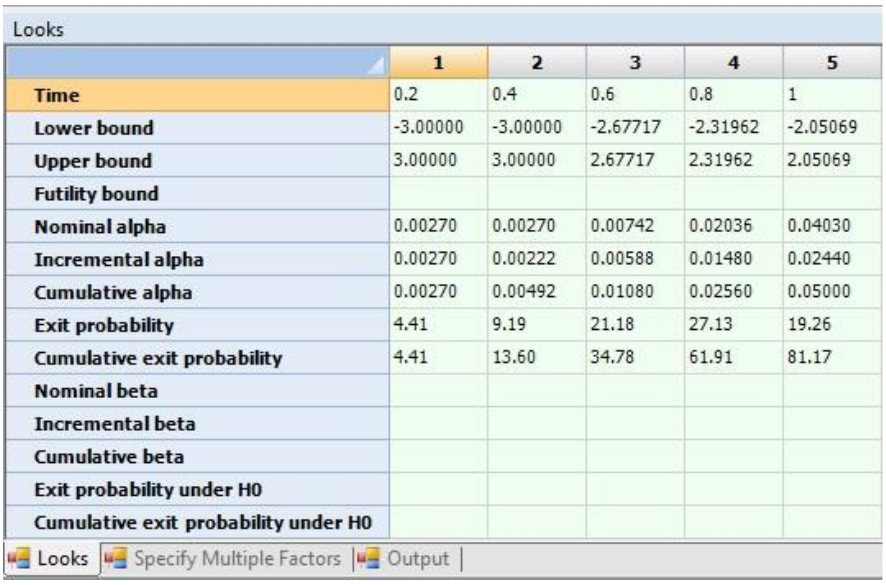
7. Select "Calculate the attainable power with the given sample sizes" from the drop-down menu and then click "Run".



	1	2	3	4
Test significance level, α	0.05			
1 or 2 sided test?	2	2	2	2
Group 1 proportions, (n1)	0.41			
Group 2 proportions, (n2)	0.465			
Odds ratio, Ψ = n2(1-n1)/n1(1- n2)	1.25074			
Group 1 size, n1	1400			
Group 2 size, n2	1400			
Ratio: N2 / N1	1	1	1	1
Continuity correction	On	Off	Off	Off
Power (%)	81.17			
Cost per sample unit	180			
Total study cost	504000			
Number of looks	5	5	5	5
Information Times	Equally Spaced	Equally Spaced	Equally Spaced	Equally Spaced
Max times	1	1	1	1
Determine bounds	Spending Function	Spending Function	Spending Function	Spending Function
Spending function	Power Family	O'Brien-Fleming	O'Brien-Fleming	O'Brien-Fleming
Phi	3			
Truncate bounds	Yes	No	No	No
Truncate at	3			
Futility boundaries	Don't Calculate	Don't Calculate	Don't Calculate	Don't Calculate
Spending function	O'Brien-Fleming	O'Brien-Fleming	O'Brien-Fleming	O'Brien-Fleming
Phi				

Figure 3.2.7. Completed Two Proportions Test using Power Family Spending Function

8. Also the boundary values will be recalculated and boundary plot will automatically be plotted as shown in Figure 3.2.8 and 3.2.9 below.



	1	2	3	4	5
Time	0.2	0.4	0.6	0.8	1
Lower bound	-3.00000	-3.00000	-2.67717	-2.31962	-2.05069
Upper bound	3.00000	3.00000	2.67717	2.31962	2.05069
Futility bound					
Nominal alpha	0.00270	0.00270	0.00742	0.02036	0.04030
Incremental alpha	0.00270	0.00222	0.00588	0.01480	0.02440
Cumulative alpha	0.00270	0.00492	0.01080	0.02560	0.05000
Exit probability	4.41	9.19	21.18	27.13	19.26
Cumulative exit probability	4.41	13.60	34.78	61.91	81.17
Nominal beta					
Incremental beta					
Cumulative beta					
Exit probability under H0					
Cumulative exit probability under H0					

Figure 3.2.8. Boundary Table for Power Family Spending Function

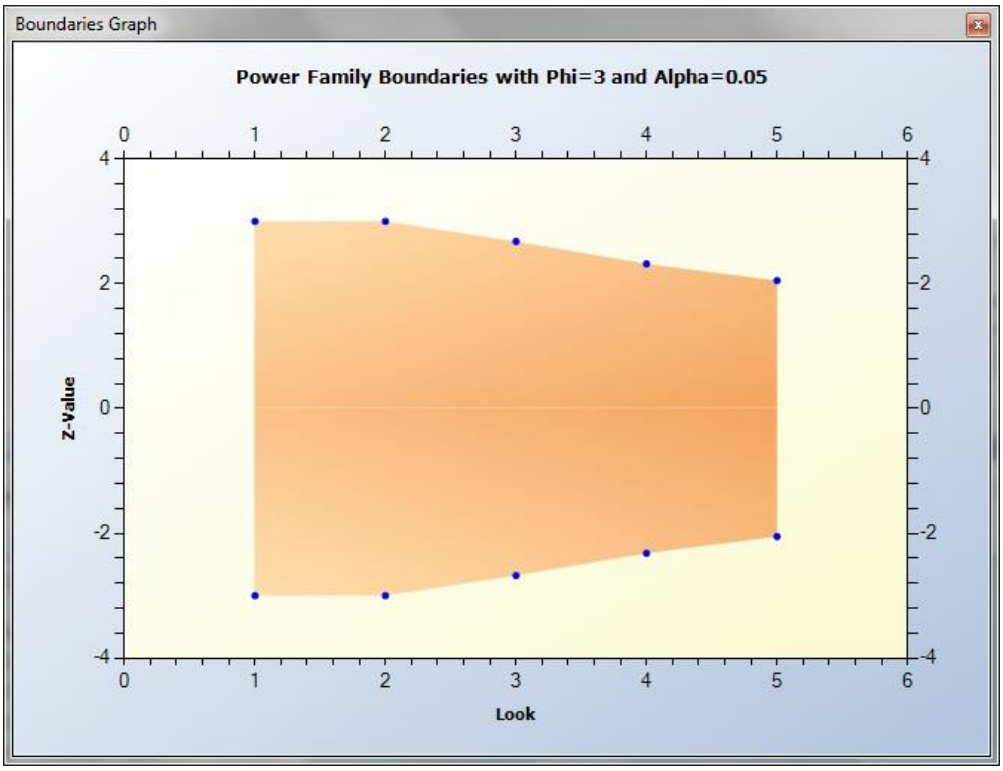


Figure 3.2.9. Boundary Plot for Power Family Spending Function

Finally, by clicking on the Output tab at the bottom of the screen you can see a statement giving details of the calculation:

"Sample sizes of at least 1400 in group 1 and 1400 in group 2 are required to achieve 81.17% power to detect an odds ratio of 1.25074 (for proportions of 0.41 in group 1 and 0.465 in group 2) using a 2-sided continuity corrected χ^2 test with 0.05 significance level. These results assume that 5 sequential tests are made and the Power Family spending function is used to determine the test boundaries."

3.3 Survival

3.3.1. Introduction

nTerim 4.0 is designed for the calculation of Power and Sample Size for both Fixed Period and Group Sequential design. In relation to Group Sequential designs, calculations are performed using the Lan-DeMets alpha spending function approach (DeMets & Lan, 1984; DeMets & Lan, 1994) for estimating boundary values. Using this approach, boundary values can be estimated for O'Brien-Fleming (O'Brien & Fleming, 1979), Pocock (Pocock, 1977), Hwang-Shih-DeCani (Hwang, Shih & DeCani, 1990) and the Power family of spending functions. Calculations follow the approach of Reboussin et al (1992) and Jennison & Turnbull (2000). Calculations can be performed for studies that involve comparisons of means, comparisons of proportions and survival studies as well as early stopping for Futility.

Group Sequential Designs

Group Sequential designs differ from Fixed Period designs in that the data from the trial is analyzed at one or more stages prior to the conclusion of the trial. As a result the alpha value applied at each analysis or `look' must be adjusted to preserve the overall Type 1 error. The alpha values used at each look are calculated based upon the spending function chosen, the number of looks to be taken during the course of the study as well as the overall Type 1 error rate. For a full introduction to group sequential methods see Jennison & Turnbull (2000) and Chow et al (2008).

Spending Function

There are four alpha and beta spending functions available to the user in nTerim 4.0 as well as an option to manually input boundary values. As standard all alpha spending functions have the properties that $\alpha(0) = 0$ and $\alpha(1) = \alpha$. Similarly, all beta spending functions have the properties that $\beta(0) = 0$ and $\beta(1) = \beta$. Functionally the alpha and beta spending functions are the same. In Table 3.3.1 we list the alpha spending functions available in nTerim 4.0.

Table 3.3.1. Spending Function Equations

O'Brien-Fleming	$\alpha(\tau) = 2\left(1 - \Phi\left(\frac{z_{\alpha/2}}{\sqrt{\tau}}\right)\right)$
Pocock	$\alpha(\tau) = \alpha ln(1 + (e - 1)\tau)$
Power	$\alpha(\tau) = \alpha\tau^{\Phi}, \Phi > 0$
Hwang-Shih-DeCani	$\alpha(\tau) = \alpha\left[\frac{(1 - e^{-\Phi\tau})}{(1 - e^{-\Phi})}\right], \Phi \neq 0$

The parameter τ represents the time elapsed in the trial. This can either be as a proportion of the overall time elapsed or a proportion of the sample size enrolled.

The common element among most of the different spending functions is to use lower error values for the earlier looks. By doing this it means that the results of any analysis will only be considered significant in an early stage if it gives an extreme result.

Boundaries

The boundaries in nTerim 4.0 represent the critical values at each look. These boundaries are constructed using the alpha and beta spending functions. Users in nTerim 4.0 are given the option to generate boundaries for early rejection of the null hypothesis, H_0, using the alpha spending function, or to generate boundaries for early rejection of either the null or alternative hypothesis, H_0 or H_1, using a combination of both the alpha and beta spending functions. The notion of using an alpha spending function approach to generate stopping boundaries for early rejection of H_0 was first proposed by Lan and DeMets (1983), we refer to such boundaries in nTerim 4.0 as efficacy boundaries. Building on the work of Lan and DeMets, Pampallona, Tsiatis, and Kim (1995, 2001) later put forward the concept of using a beta spending approach to construct boundaries for early rejection of H_1, we refer to these boundaries in nTerim as futility boundaries.

Essentially, if a test statistic crosses an efficacy boundary then it can be concluded that the experimental treatment shows a statistically significant effect, the trial can be stopped with rejection of the null hypothesis. If the test statistic crosses a futility boundary then this indicates with high probability that an effect will not be found, that the trial can be terminated by rejecting the alternative hypothesis.

In the case where the user wishes to generate boundaries for early rejection of either the null or alternative hypothesis, H_o or H_1, they are given two options; either to have the boundaries binding, or non-binding. With binding boundaries, if the test statistic crosses the futility boundary, the test must be stopped, otherwise the type-1 error may become inflated. The reason for this is that there is an interaction between the efficacy and futility boundaries in their calculation that could cause the efficacy boundary to shift. In the case of non-binding boundaries; the efficacy boundaries are calculated as normal, that is, as if the futility boundaries did not exist. This eliminates the danger of inflating the type-1 error when the futility boundary is overruled. The downside of the non-binding case is that it may increase the required sample size relative to the binding case.

The boundaries calculated in nTerim 4.0 follow the procedures outlined by Reboussin et al (1992) and Jennison & Turnbull (2000).

3.3.2. Methodology

Sequential Log-Rank test of survival in to groups, the variables are defined as:

Symbol	Description
α	Probability of Type I error
β	Probability of Type II error
$1-\beta$	Power of the Test
s_1, s_2	Group Survival Proportions
d	Number of Events
N	Sample Size
R	Ratio of N_1 to N_2
Θ	Drift Parameter
K	Number of Time-points (Looks)

Calculate Sample Size for a given Power

Using the number of time-points (K), number of sides, type of spending function, the hypothesis to be rejected, the type 1 error, α, and the power, $(1-\beta)$, the drift parameter Θ can be obtained using algorithms by Reboussin et al (1992) and Jennison & Turnbull (2000).

$$HR = log\left(\frac{s_2}{s_1}\right) \qquad \{3.3.1\}$$

For the Exponential Survival Curve, this is defined by the expression below.

$$\Theta = \frac{\log(HR)\sqrt{d_k}}{2} \qquad \{3.3.2\}$$

This can be solved for d_k, the required number of events using the equation below.

$$d_k = \frac{4\Theta^2}{[\log(HR)]^2} \qquad \{3.3.3\}$$

Then, to calculate the Proportional Hazards Curve, Equation {3.3.4} is employed.

$$\Theta = \frac{1-\mathrm{HR}\sqrt{d_k}}{1+HR} \qquad \{3.3.4\}$$

This can be solved for d_k, the required number of events using Equation {3.3.5}.

$$d_k = \left[\frac{(1+HR)\Theta}{(1-HR)}\right]^2 \qquad \{3.3.5\}$$

To calculate the sample size, N, the following formula is used.

$$N = \frac{2d_k}{2 - s_1 - s_2} \qquad \{3.3.6\}$$

Calculate Attainable Power with the given Sample Size

Given α, N, group survival proportions (s_1, s_2), number of time-points, K, number of sides, type of spending function, the hypothesis to be rejected, the requirement is to obtain the power.

For the Exponential Survival Curve, Equation {3.3.7} is used.

$$\Theta = \sqrt{\frac{N(2 - s_1 - s_2)\log^2(HR)}{8}} \qquad \{3.3.7\}$$

For the Proportional Hazards Curve, Equation {3.3.8} is used.

$$\Theta = \sqrt{\frac{N(2 - s_1 - s_2)(1 - HR)^2}{2(1 + HR)^2}} \qquad \{3.3.8\}$$

3.3.3. Examples

Example 1: O'Brien-Fleming Spending function – with Power vs. Sample Size Plot

1. Open nTerim through the Start Menu or by double clicking on the nTerim desktop icon. Then click on "New Interim Test" from the menu bar at the top of the window. A "Study Goal and Design" window will appear as shown below. Select the options as mapped out in Figure 3.3.1, then Click "OK".

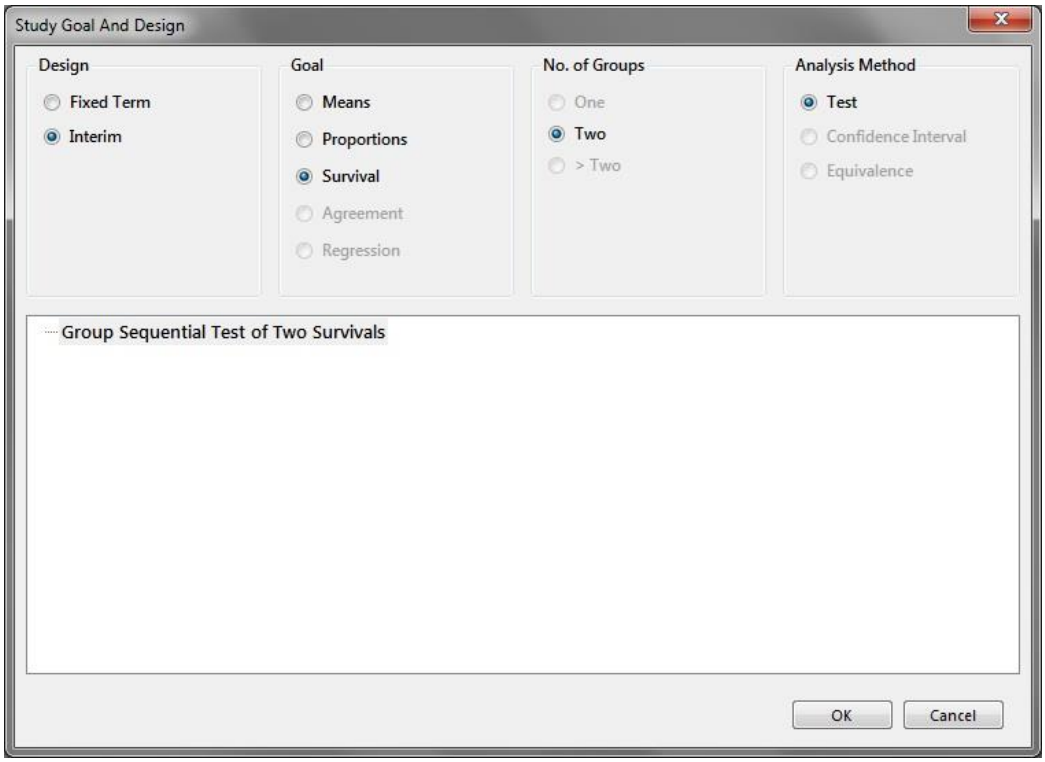


Figure 3.3.1. Study Goal and design Window

2. Enter 0.05 for alpha, 2 sided, 0.3 for Group 1 proportion (this is the proportion surviving until time t) and 0.45 for Group 2 proportion. The hazard ratio is calculated as 1.508.

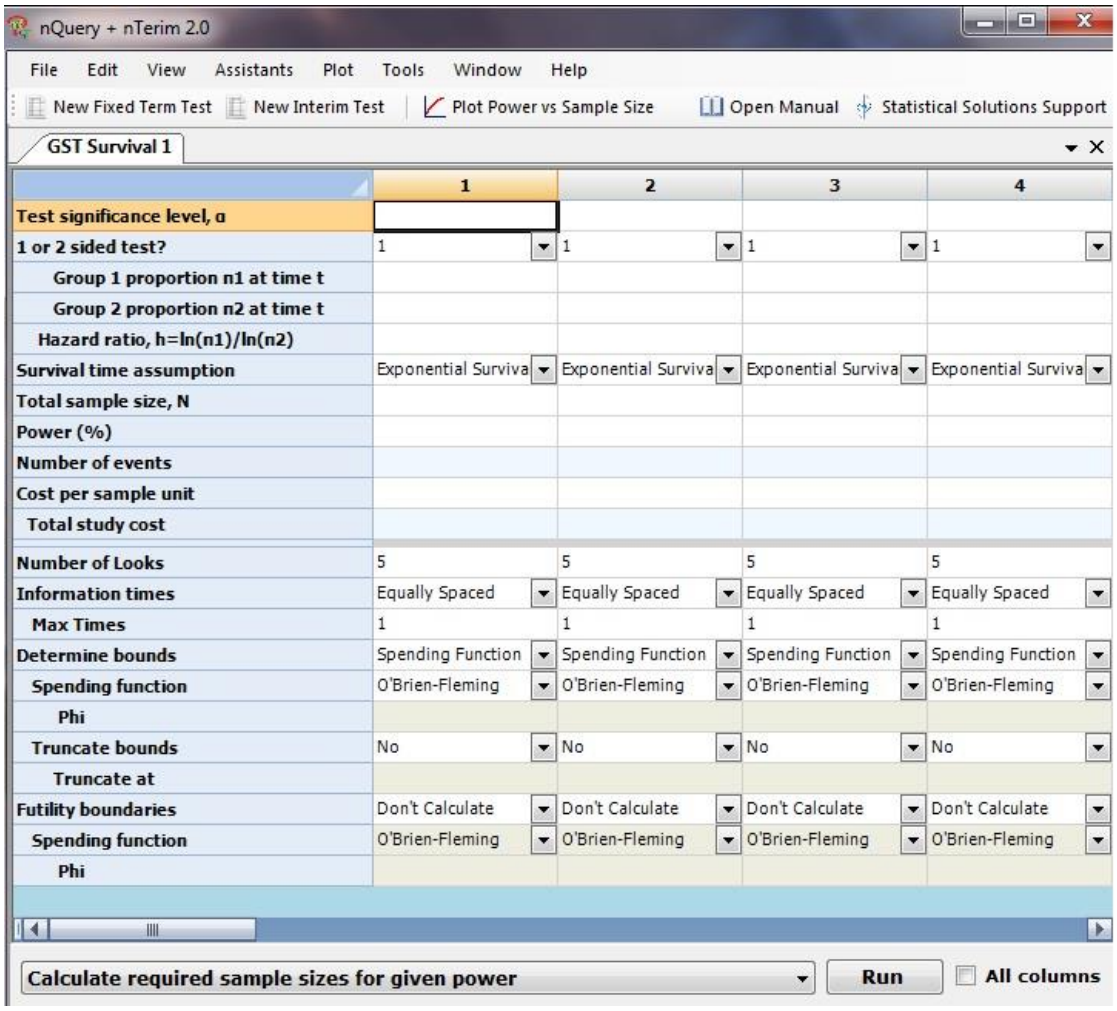
3. Select Exponential Survival for the Survival time assumption.

4. We are interested in solving for sample size given 90% power so enter 90 in the Power row.

5. This study planned for 4 interim analyses. Including the final analysis this requires Number of Looks to be 5.

6. The looks will be equally spaced and the O'Brien-Fleming spending function is to be used. There will be no truncation of bounds.

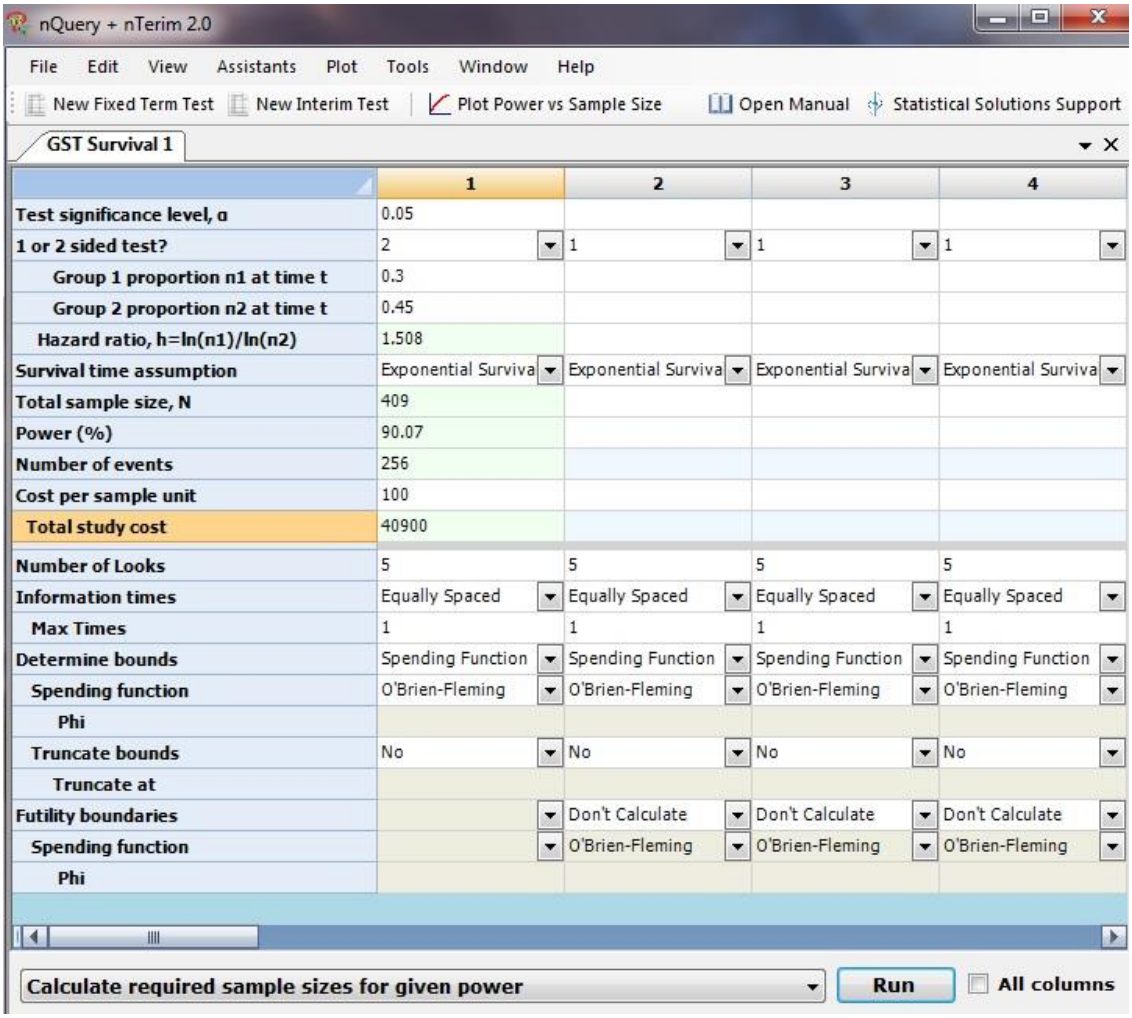
7. It is estimated that the cost per unit is roughly $100, so enter 100 in the Cost per sample unit row.



	1	2	3	4
Test significance level, α				
1 or 2 sided test?	1	1	1	1
Group 1 proportion n1 at time t				
Group 2 proportion n2 at time t				
Hazard ratio, h=ln(n1)/ln(n2)				
Survival time assumption	Exponential Surviva	Exponential Surviva	Exponential Surviva	Exponential Surviva
Total sample size, N				
Power (%)				
Number of events				
Cost per sample unit				
Total study cost				
Number of Looks	5	5	5	5
Information times	Equally Spaced	Equally Spaced	Equally Spaced	Equally Spaced
Max Times	1	1	1	1
Determine bounds	Spending Function	Spending Function	Spending Function	Spending Function
Spending function	O'Brien-Fleming	O'Brien-Fleming	O'Brien-Fleming	O'Brien-Fleming
Phi				
Truncate bounds	No	No	No	No
Truncate at				
Futility boundaries	Don't Calculate	Don't Calculate	Don't Calculate	Don't Calculate
Spending function	O'Brien-Fleming	O'Brien-Fleming	O'Brien-Fleming	O'Brien-Fleming
Phi				

Figure 3.3.2. Survival Test Table

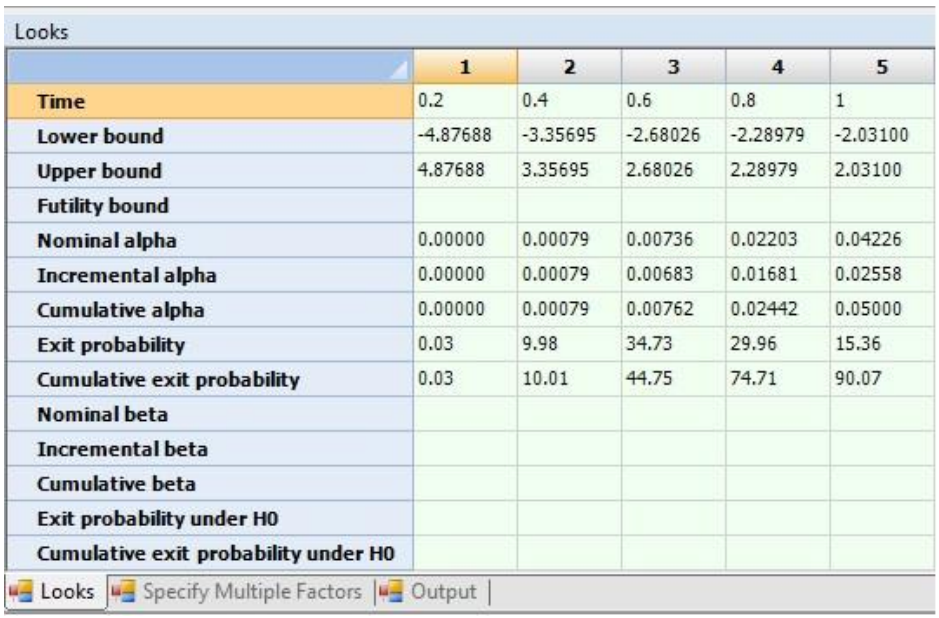
8. Once all values have been entered, select “Calculate required sample size for given power” from the drop-down menu and click “Run”.



	1	2	3	4
Test significance level, α	0.05			
1 or 2 sided test?	2	1	1	1
Group 1 proportion n1 at time t	0.3			
Group 2 proportion n2 at time t	0.45			
Hazard ratio, h=ln(n1)/ln(n2)	1.508			
Survival time assumption	Exponential Surviva	Exponential Surviva	Exponential Surviva	Exponential Surviva
Total sample size, N	409			
Power (%)	90.07			
Number of events	256			
Cost per sample unit	100			
Total study cost	40900			
Number of Looks	5	5	5	5
Information times	Equally Spaced	Equally Spaced	Equally Spaced	Equally Spaced
Max Times	1	1	1	1
Determine bounds	Spending Function	Spending Function	Spending Function	Spending Function
Spending function	O'Brien-Fleming	O'Brien-Fleming	O'Brien-Fleming	O'Brien-Fleming
Phi				
Truncate bounds	No	No	No	No
Truncate at				
Futility boundaries		Don't Calculate	Don't Calculate	Don't Calculate
Spending function		O'Brien-Fleming	O'Brien-Fleming	O'Brien-Fleming
Phi				

Figure 3.3.3. Complete Survival Table for One test

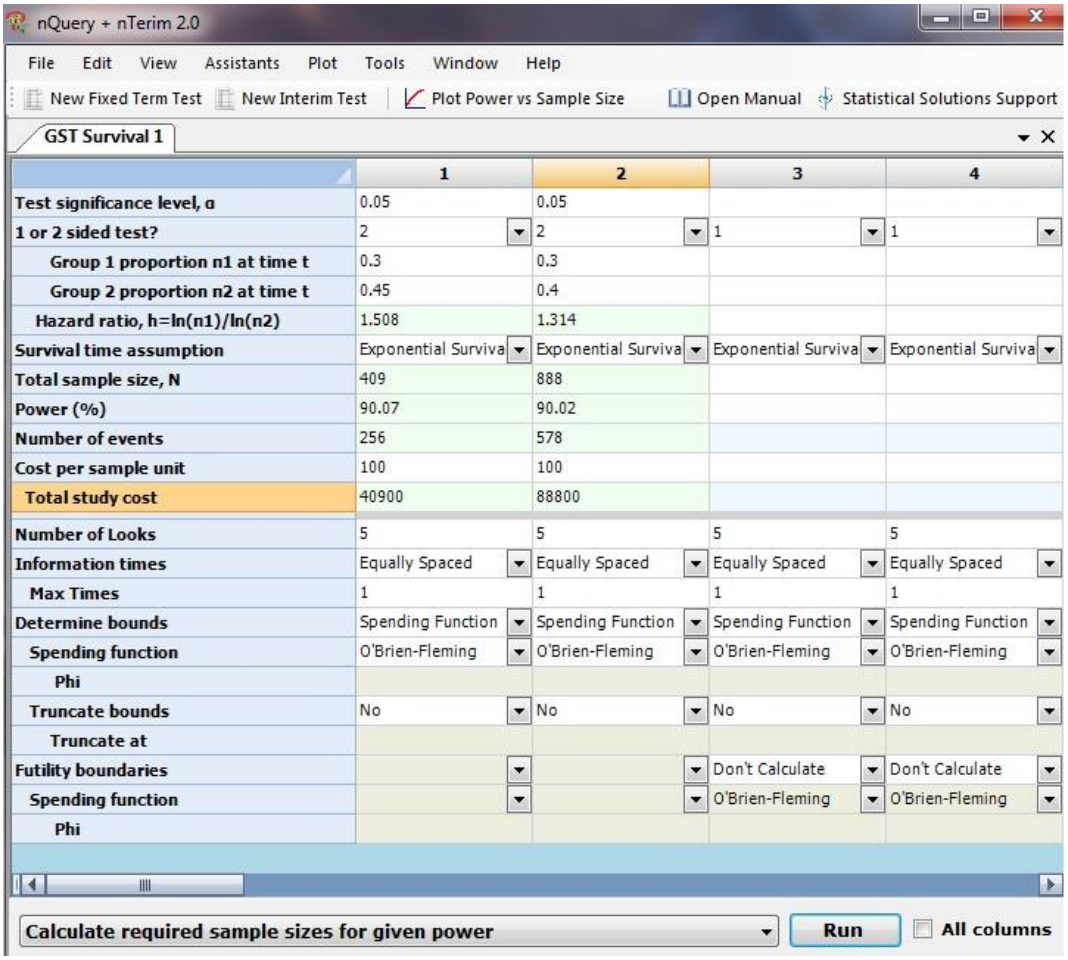
In addition to the sample size and cost output for Column 1, the boundary calculations are also presented as shown below.



	1	2	3	4	5
Time	0.2	0.4	0.6	0.8	1
Lower bound	-4.87688	-3.35695	-2.68026	-2.28979	-2.03100
Upper bound	4.87688	3.35695	2.68026	2.28979	2.03100
Futility bound					
Nominal alpha	0.00000	0.00079	0.00736	0.02203	0.04226
Incremental alpha	0.00000	0.00079	0.00683	0.01681	0.02558
Cumulative alpha	0.00000	0.00079	0.00762	0.02442	0.05000
Exit probability	0.03	9.98	34.73	29.96	15.36
Cumulative exit probability	0.03	10.01	44.75	74.71	90.07
Nominal beta					
Incremental beta					
Cumulative beta					
Exit probability under H0					
Cumulative exit probability under H0					

Figure 3.3.4. Boundary Table for Column 1

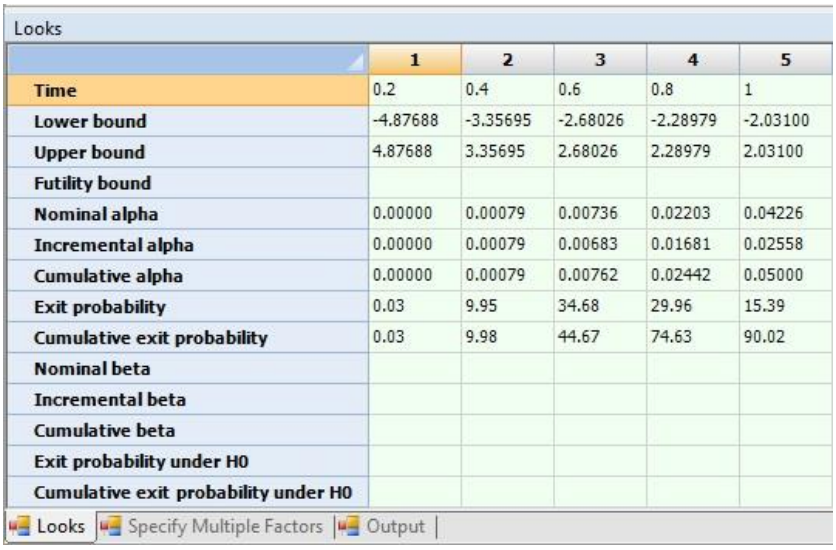
9. In the second column enter the same parameters as above but change the Group 2 proportion to 0.40. Select “Run”.



	1	2	3	4
Test significance level, α	0.05	0.05		
1 or 2 sided test?	2	2	1	1
Group 1 proportion n1 at time t	0.3	0.3		
Group 2 proportion n2 at time t	0.45	0.4		
Hazard ratio, h=ln(n1)/ln(n2)	1.508	1.314		
Survival time assumption	Exponential Surviva	Exponential Surviva	Exponential Surviva	Exponential Surviva
Total sample size, N	409	888		
Power (%)	90.07	90.02		
Number of events	256	578		
Cost per sample unit	100	100		
Total study cost	40900	88800		
Number of Looks	5	5	5	5
Information times	Equally Spaced	Equally Spaced	Equally Spaced	Equally Spaced
Max Times	1	1	1	1
Determine bounds	Spending Function	Spending Function	Spending Function	Spending Function
Spending function	O'Brien-Fleming	O'Brien-Fleming	O'Brien-Fleming	O'Brien-Fleming
Phi				
Truncate bounds	No	No	No	No
Truncate at				
Futility boundaries			Don't Calculate	Don't Calculate
Spending function			O'Brien-Fleming	O'Brien-Fleming
Phi				

Figure 3.3.5. Complete Survival Table for Two tests

In addition to the sample size and cost output for Column 2, the boundary calculations are also presented as shown below.



	1	2	3	4	5
Time	0.2	0.4	0.6	0.8	1
Lower bound	-4.87688	-3.35695	-2.68026	-2.28979	-2.03100
Upper bound	4.87688	3.35695	2.68026	2.28979	2.03100
Futility bound					
Nominal alpha	0.00000	0.00079	0.00736	0.02203	0.04226
Incremental alpha	0.00000	0.00079	0.00683	0.01681	0.02558
Cumulative alpha	0.00000	0.00079	0.00762	0.02442	0.05000
Exit probability	0.03	9.95	34.68	29.96	15.39
Cumulative exit probability	0.03	9.98	44.67	74.63	90.02
Nominal beta					
Incremental beta					
Cumulative beta					
Exit probability under H0					
Cumulative exit probability under H0					

Figure 3.3.6. Boundary Table for Column 2

Finally, in terms of output, the boundaries that were calculated as shown in Figure 3.3.4 and 3.3.6 were automatically plotted by nTerim, the boundary plot for Column 1 is given below.

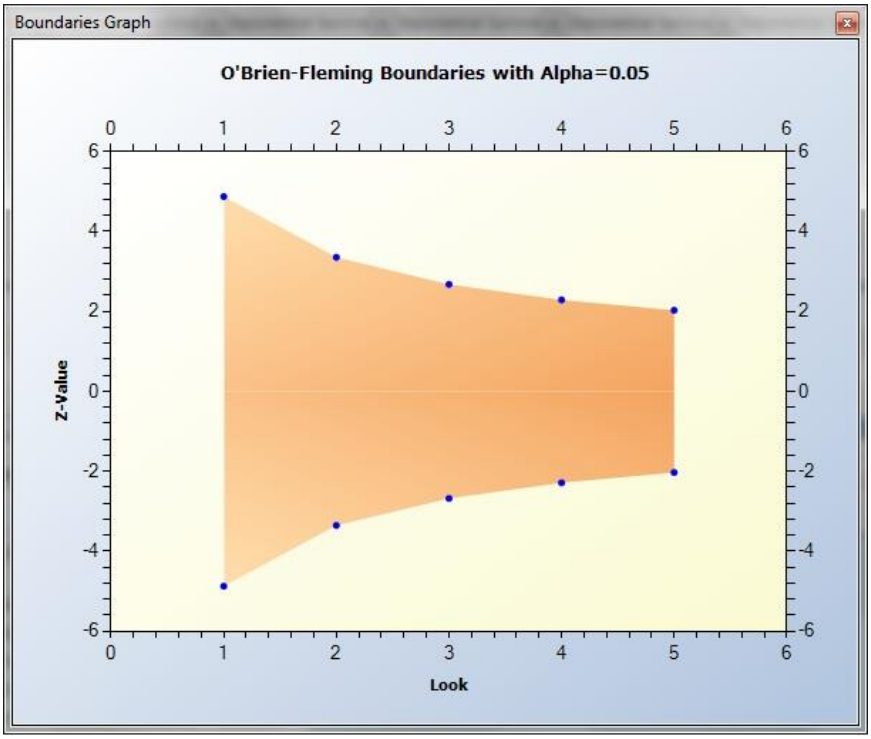


Figure 3.3.7. Boundary Plot for Column 1

10. Click on the column title for Column 1 and drag across to highlight both Columns 1 and 2.

11. Select Plot Power-Sample Size from the toolbar, (it may take a moment to generate the plot as multiple calculations are performed)

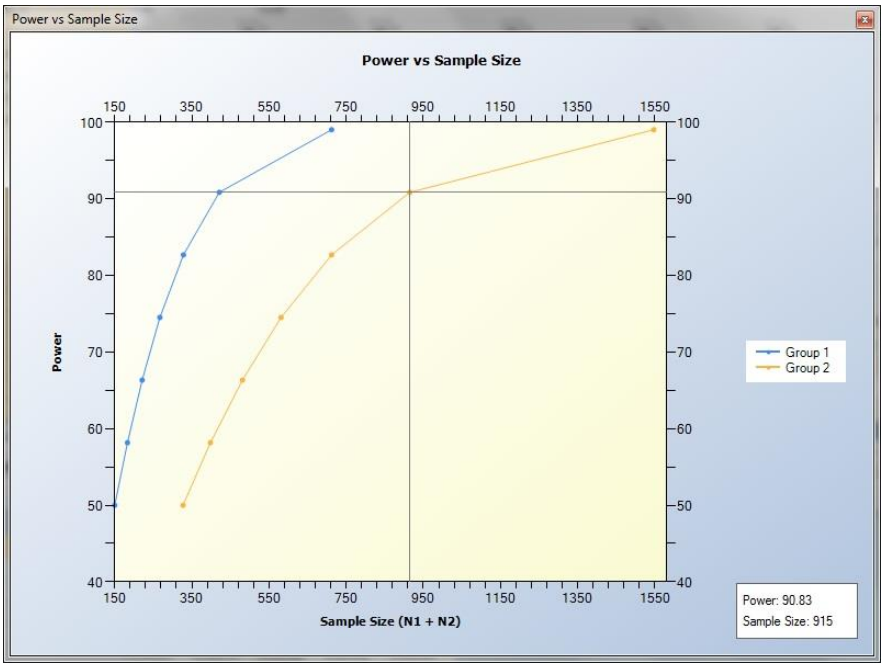


Figure 3.3.8. Power vs. Sample Size Plot

As it can be seen in Figure 3.3.8, an illustration of the comparison between Column 1 and Column 2 in relation to Power vs. Sample Size performance can be created. The cross on the graph illustrates how the user can identify what the sample size is for a corresponding power value for each column. In the bottom right corner of the plot indicated the exact values for Power and Sample Size for each identifier on the graph.

Finally, by clicking on the Output tab at the bottom of the screen you can see a statement giving details of the calculation:

Column 1 – Output Statement
"A total sample size of at least 409 (256 events) is required to achieve 90.07% power to detect a hazard ratio of 1.508 (for survival rates of 0.3 in group 1 and 0.45 in group 2), using a 2-sided log rank test with 0.05 significance level assuming that the survival times are exponential. These results assume that 5 sequential tests are made and the O Brien-Fleming spending function is used to determine the test boundaries."

Example 2: Pocock Spending Function – with Non-equally Spaced Looks

1. Open nTerim through the Start Menu or by double clicking on the nTerim desktop icon. Then click on "New Interim Test" from the menu bar at the top of the window. A "Study Goal and Design" window will appear as shown below. Select the options as mapped out in Figure 3.3.9, then Click "OK".

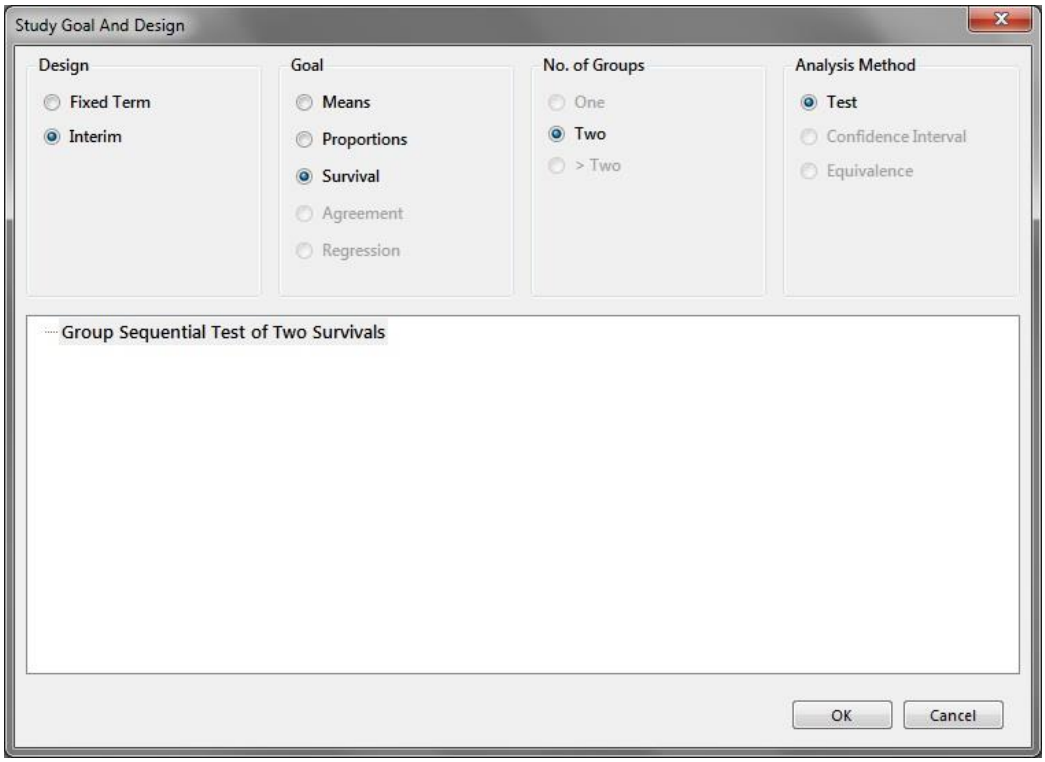


Figure 3.3.9. Study Goal and design Window

2. Enter 0.05 for alpha, 2 sided, 0.5 for Group 1 proportion, 0.4 for Group 2 proportion. The hazard ratio is calculated as 0.756.

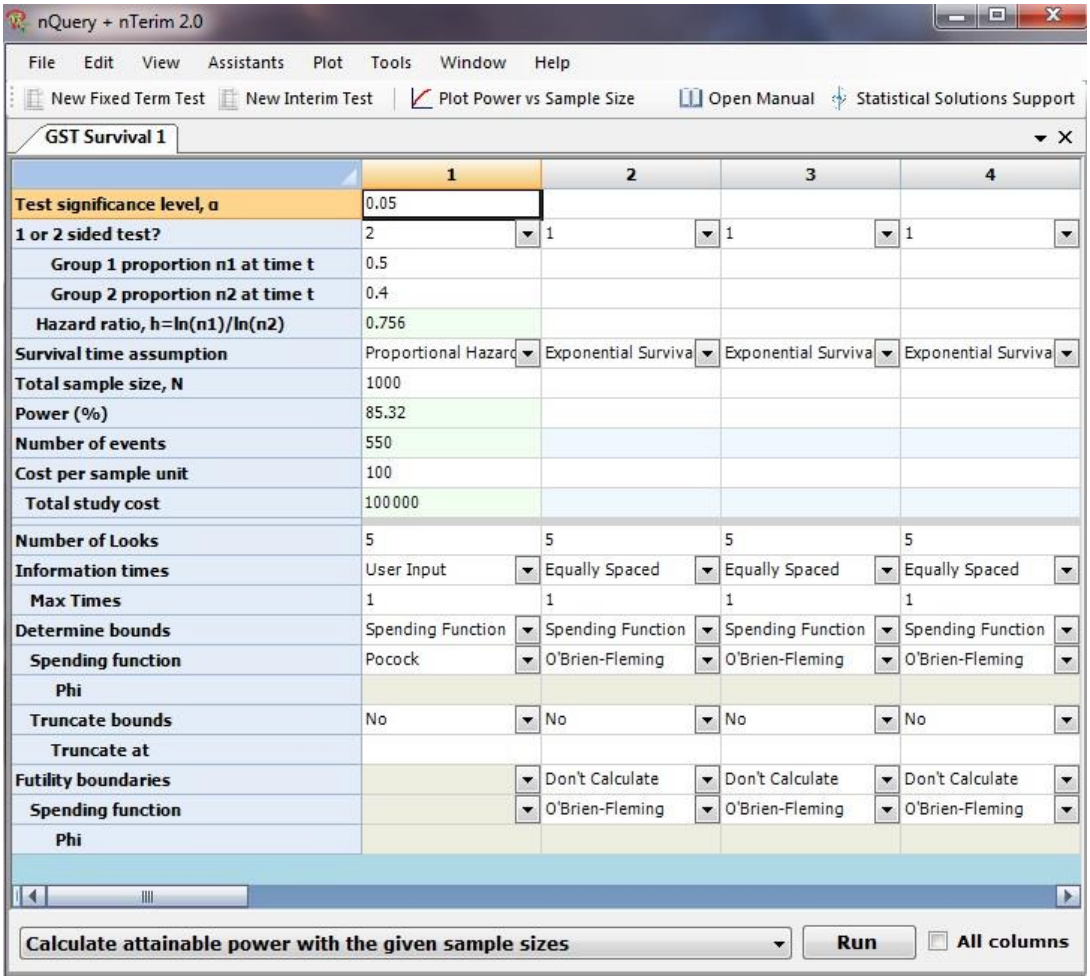
3. Select Proportional Hazards for the Survival Time Assumption. We are interested in solving for power given a sample size of 1000, so enter 1000 in the Total Sample Size row.

4. This study planned for 4 interim analyses. Including the final analysis this requires Number of Looks to be 5.

5. The Pocock spending function is to be used, however the looks will not be evenly spaced.

6. For Information Times, select User Input. Then in the Times row in the lower table enter the values 0.1, 0.2, 0.3, 0.6 and 1.

7. It is estimated that the cost per unit is roughly $100, so enter 100 in the Cost per sample unit row.

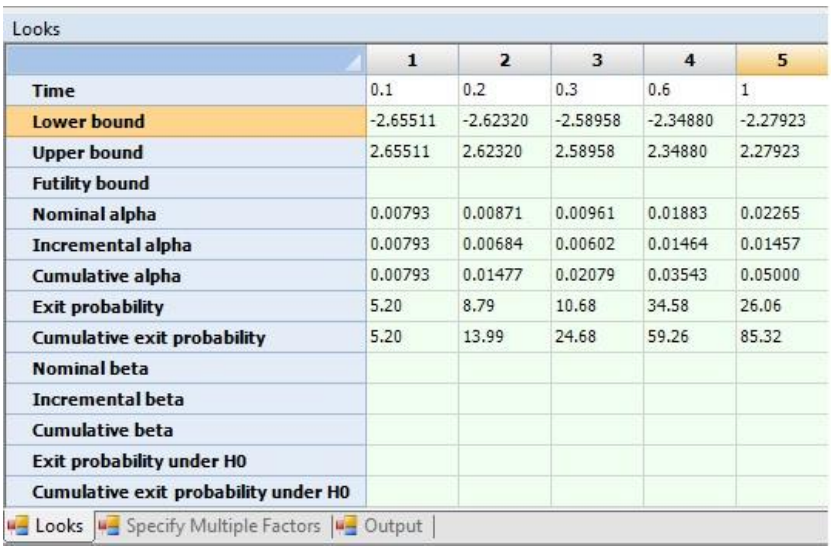


	1	2	3	4
Test significance level, α	0.05			
1 or 2 sided test?	2	1	1	1
Group 1 proportion n1 at time t	0.5			
Group 2 proportion n2 at time t	0.4			
Hazard ratio, h=ln(n1)/ln(n2)	0.756			
Survival time assumption	Proportional Hazard	Exponential Surviva	Exponential Surviva	Exponential Surviva
Total sample size, N	1000			
Power (%)	85.32			
Number of events	550			
Cost per sample unit	100			
Total study cost	100000			
Number of Looks	5	5	5	5
Information times	User Input	Equally Spaced	Equally Spaced	Equally Spaced
Max Times	1	1	1	1
Determine bounds	Spending Function	Spending Function	Spending Function	Spending Function
Spending function	Pocock	O'Brien-Fleming	O'Brien-Fleming	O'Brien-Fleming
Phi				
Truncate bounds	No	No	No	No
Truncate at				
Futility boundaries		Don't Calculate	Don't Calculate	Don't Calculate
Spending function		O'Brien-Fleming	O'Brien-Fleming	O'Brien-Fleming
Phi				

Figure 3.3.10. Complete Survival Table with Pocock Spending Function

8. Once all the values have been entered, select "Calculate the attainable power with the given sample sizes" from the drop-down menu and click "Run".

9. The boundaries calculated are shown in Figure 3.3.11.



Looks

	1	2	3	4	5
Time	0.1	0.2	0.3	0.6	1
Lower bound	-2.65511	-2.62320	-2.58958	-2.34880	-2.27923
Upper bound	2.65511	2.62320	2.58958	2.34880	2.27923
Futility bound					
Nominal alpha	0.00793	0.00871	0.00961	0.01883	0.02265
Incremental alpha	0.00793	0.00684	0.00602	0.01464	0.01457
Cumulative alpha	0.00793	0.01477	0.02079	0.03543	0.05000
Exit probability	5.20	8.79	10.68	34.58	26.06
Cumulative exit probability	5.20	13.99	24.68	59.26	85.32
Nominal beta					
Incremental beta					
Cumulative beta					
Exit probability under H0					
Cumulative exit probability under H0					

Looks | Specify Multiple Factors | Output

Figure 3.3.11. Boundary Table for Pocock Spending Function

10. Finally, the boundaries calculated in the table displayed in Figure 3.3.11 are automatically plotted as illustrated in Figure 3.3.12.

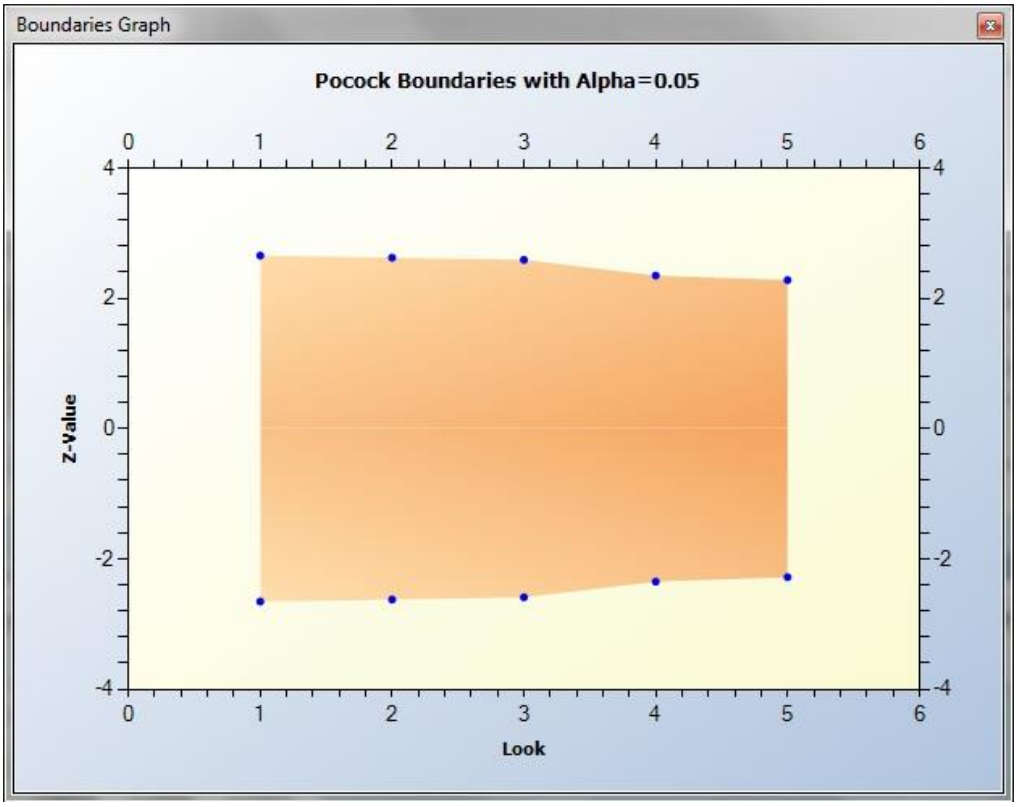


Figure 3.3.12. Boundary Plot for Proportional Hazard Survival Test

By clicking on the output tab at the bottom of the screen you can see a statement giving details of the calculation:

"A total sample size of at least 1000 (550 events) is required to achieve 85.32% power to detect a hazard ratio of 0.756 (for survival rates of 0.5 in group 1 and 0.4 in group 2), using a 2-sided log rank test with 0.05 significance level assuming that the hazards are proportional. These results assume that 5 sequential tests are made and the Pocock spending function is used to determine the test boundaries."

Chapter 4

Fixed Term Design

4.1 One-Way Repeated Measures Contrast (Constant Correlation)

4.1.1. Introduction

This table facilitates the calculation of power and sample size for a one-way repeated measures contrast design. Calculations are performed using the methods outlined by Overall and Doyle (1994).

A one-way repeated measures contrast is used to analyse specific planned contrasts in a repeated measures one-way analysis of variance (ANOVA) design. This is an experimental design in which multiple measurements are taken on a group of subjects over time or under different conditions. This design is the same as the one-way ANOVA but for related not independent groups. It can be viewed as an extension of the dependent t-test.

To give an example of such a design; consider a study of a three month intervention aimed at raising self-esteem in children. Self-esteem will be measured before, after one month, after two months, and after three months of the intervention. It is assumed that self-esteem will increase monotonically over time. Thus, for this study it may be of interest to test for a linear trend in self-esteem. The contrasts -3, -1, 1, 3 would be appropriate for such a study. Such planned contrasts are useful because they provide a more sharply focused analysis compared to overall tests. This usually makes tests of planned contrasts easier to interpret and more powerful.

4.1.2. Methodology

Power and sample size is calculated using central and non-central F-distributions and follows the procedures outlined by Overall, and Doyle (1994)

To calculate power and sample size the user must specify the test significance level, α, and the number of levels, M. The user must then enter values for the contrast, C, and the Scale, D. Alternatively, the user can enter the expected means at each level and the respective contrast coefficients using the compute effect size assistant. nTerim will then calculate the contrast and scale using the following formulas for contrast:

$$C = \sum_{i=1}^{M} c_i \mu_i \qquad \{4.1.1\}$$

and scale,

$$D = \sqrt{\sum_{i=1}^{M} c_i^2} \qquad \{4.1.2\}$$

Once the contrast and the scale have been entered, the user must input values for the common standard deviation, σ, and the between level correlation, ρ. The standard deviation at each level is assumed to be the same and the correlation between each pair of levels is assumed to be the same. Given these four values, nTerim will automatically calculate the effect size using the following formula:

$$\Delta = \frac{|C|}{D\sigma\sqrt{(1-\rho)}} \qquad \{4.1.3\}$$

In order to calculate power, a value for the total sample size, N, must be entered. nTerim then calculates the power of the design by first determining the critical value $F_{DF_1,\ DF_2,\ \alpha}$. Where, $DF_1 = 1$ is the numerator degrees of freedom, and $DF_2 = (M-1) * (N-1)$ is the denominator degrees of freedom.

The non-centrality parameter, λ, is then calculated using the equation:

$$\lambda = N\Delta^2 \qquad \{4.1.4\}$$

Using these two values, nTerim calculates the power of this design as the probability of being greater than $F_{DF_1,\ DF_2,\ \alpha}$ on a non-central F-distribution with non-centrality parameter λ.

In order to calculate sample size nTerim does not use a closed form equation. Instead a search algorithm is used. This search algorithm calculates power at various sample sizes until the desired power is reached.

4.1.3. Examples

Example 1: Examining the specific contrast between high and low doses of a new drug
This test can be incorporated when examining different levels within a certain variable. In this example we want to examine the contrast between high doses and low doses of a specific new drug.

The following steps outline the procedure for Example 1.

1. Open nTerim through the Start Menu or by double clicking on the nTerim desktop icon. Then click on "New Fixed Term Test" from the menu bar at the top of the window. A "Study Goal and Design" window will appear.

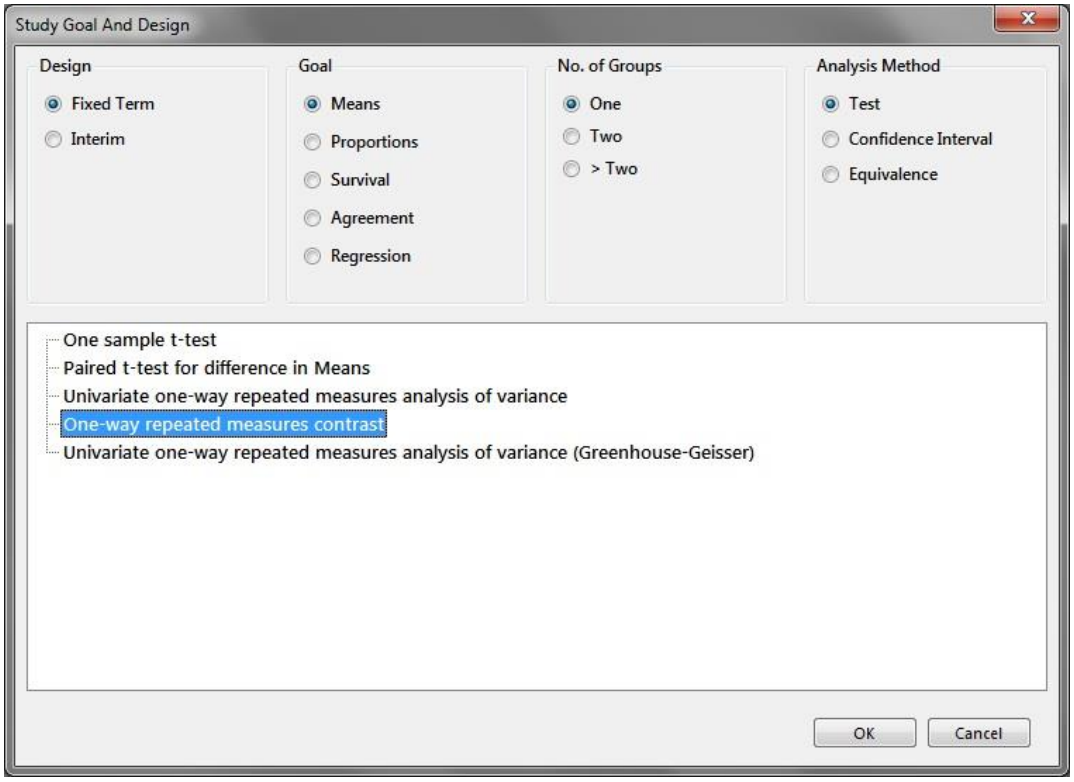


Figure 4.1.1. Study Goal and Design Window

2. Once the correct test has been selected, click "OK" and the test window will appear.

3. There are two main tables required for this test, the main test table illustrated in Figure 4.1.2 and the effect size assistant table shown in Figure 4.1.3.

4. Enter 0.05 for alpha, the desired significance level, and enter 3 for the number of levels, M, as shown in Figure 4.1.4.

5. Now you are required to complete the "Compute Effect Size Assistant" table in order to calculate values for the Contrast (C) and Scale (D) parameters.

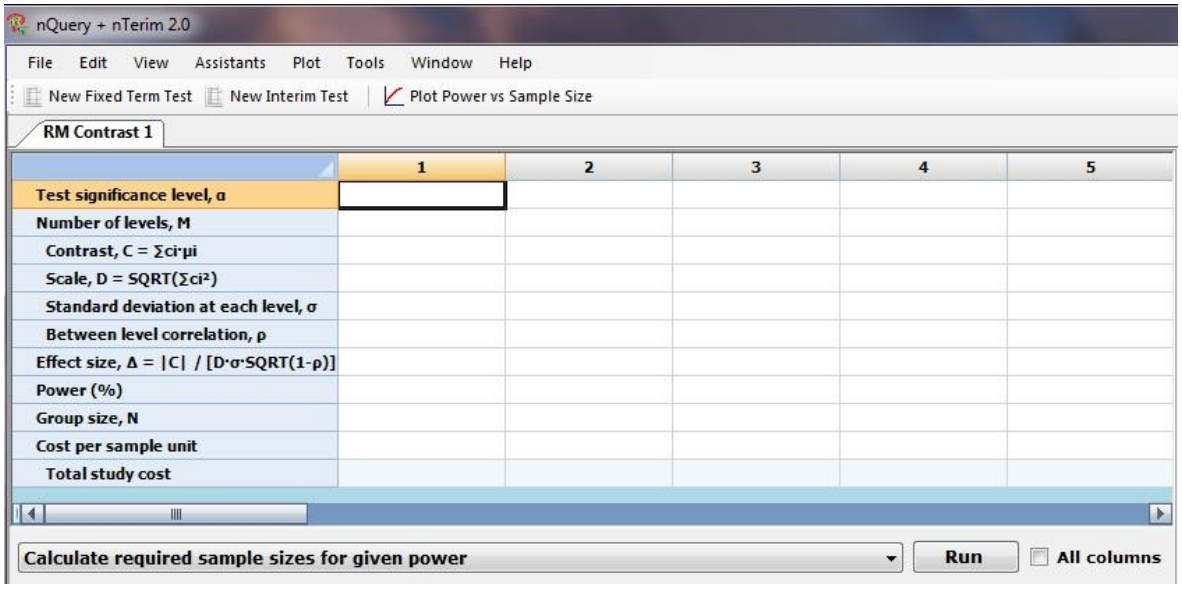


Figure 4.1.2. One-way Repeated Measures Contrast Test Table

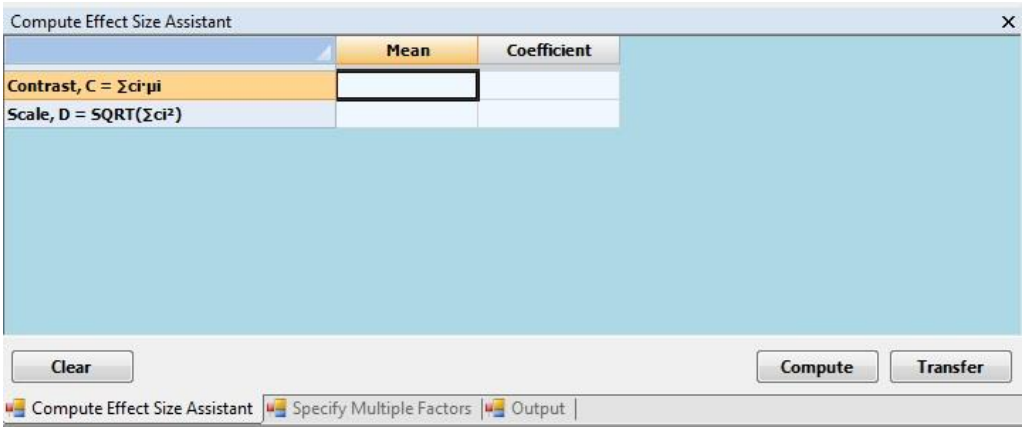


Figure 4.1.3. Compute Effect Size Assistant Table

6. Once you enter a value for the number of levels, M, the "Compute Effect Size Assistant" table automatically updates as shown in Figure 4.1.4.

7. In order to calculate a value for Effect Size, two parameters need to be calculated first, the Contrast (C) and Scale (D).

8. The mean for each level and the corresponding coefficient value need to be entered in the "Compute effect Size Assistant" table.

9. For the "Mean" values for each level, enter 12 for level 1, 12 for level 2 and 14 for level 3.

10. For the "Coefficient" values for each level, enter 0 for level 1, -1 for level 2 and 1 for level 3. The sum of these values must always equate to zero. This is illustrated in Figure 4.1.5 below.

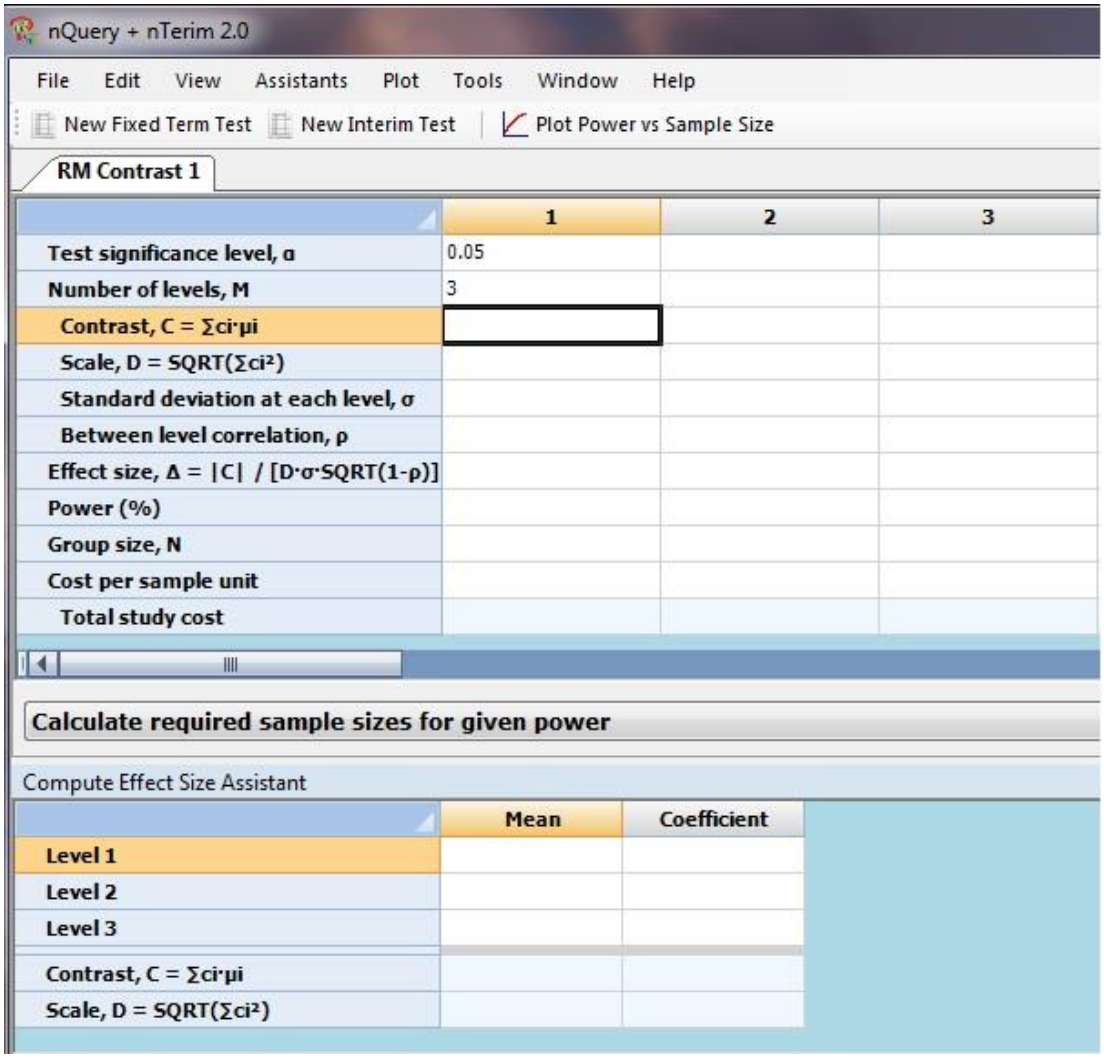


Figure 4.1.4. Automatically Updated Compute Effect Size Assistant Table

11. Once the table in Figure 4.1.5 is completed, and values for Contrast (C) and Scale (D) are computed, click on "Transfer" to automatically transfer these values to the main table.

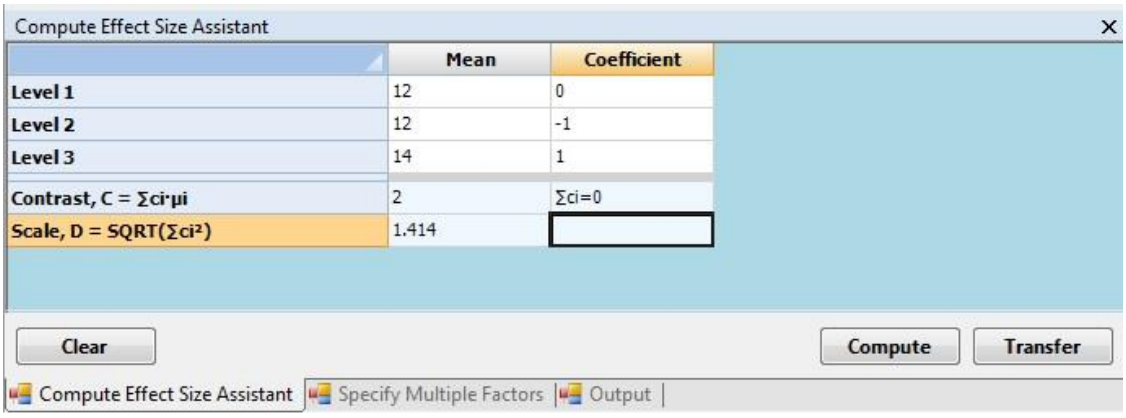


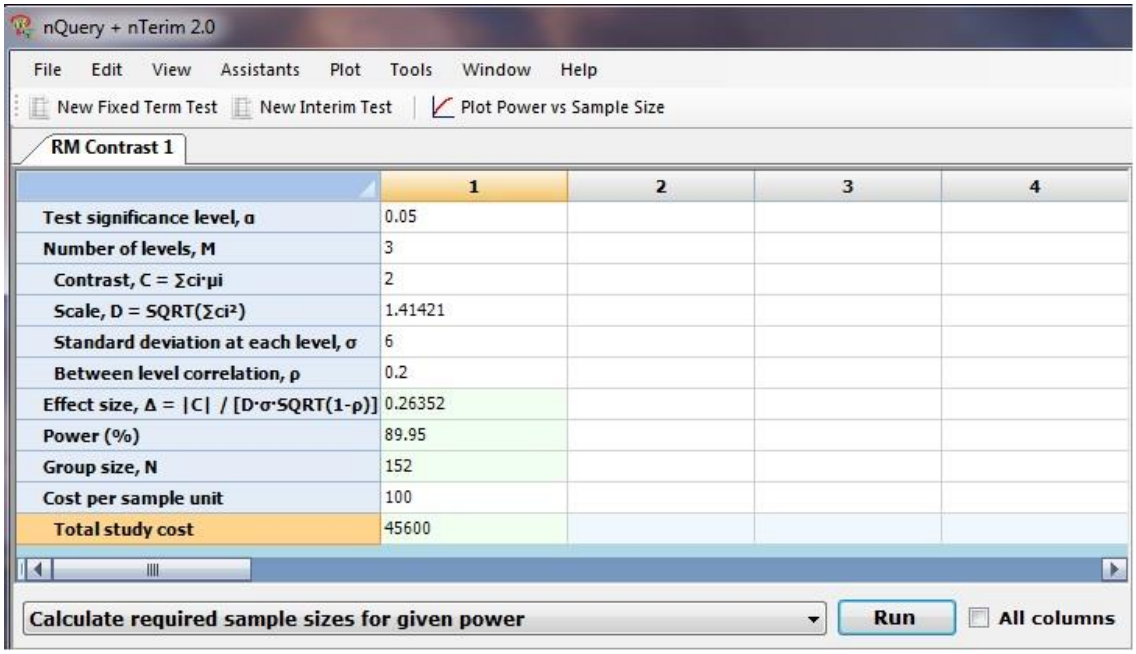
Figure 4.1.5. Completed Compute Effect Size Assistant Table

12. Now that values for Contrast (C) and Scale (D) have been computed we can continue with filling in the main table. For the Standard Deviation, enter a value of 6. For the between level correlation, enter a value of 0.2.

13. We want to calculate the sample size required obtain a power of 90%. Therefore enter 90 in the Power row.

14. It has been estimated that it will cost $100 per sample unit in this study. Therefore enter 100 in the "Cost per sample unit" row.

15. Select "Calculate required sample size for given power" from the drop-down menu below the main table and click "Run". This is displayed in Figure 4.1.6.



	1	2	3	4
Test significance level, α	0.05			
Number of levels, M	3			
Contrast, C = ∑ci·μi	2			
Scale, D = SQRT(∑ci²)	1.41421			
Standard deviation at each level, σ	6			
Between level correlation, ρ	0.2			
Effect size, Δ = |C| / [D·σ·SQRT(1-ρ)]	0.26352			
Power (%)	89.95			
Group size, N	152			
Cost per sample unit	100			
Total study cost	45600			

Figure 4.1.6. Completed One-way Repeated Measures Contrast Table

It can be seen from Figure 4.1.6 that a sample size of 152 per group (for each of the three groups, thus a total sample size, N, of 456) is required to obtain a power of 89.95%. Due to the cost per sample unit of $100, the overall cost of sample size required has amounted to $45,600.

By clicking on the Output tab at the bottom of the screen you can see a statement giving details of the calculation:

"When the group sample size (n) is 152, the test of a single contrast at the 0.05 level in a one way repeated measures analysis of variance with 3 levels will have 89.95% power to detect a contrast C = ∑ci·μi of 2, with a scale D = SQRT(∑ci²) of 1.41421, assuming a standard deviation at each level of 6 and a between level correlation of 0.2."

Example 2: Examining M Period Crossover Design

This design may require treatments to appear an equal number of times per each sequence. It can be assumed these sequences are chosen in order to prevent confounding from occurring between treatment and period effects. Therefore this is ensuring the design is balanced. In this example, we will investigate a three period, two treatment design of ABB and BAA.

The following steps outline the procedure for Example 2.

1. Open nTerim through the Start Menu or by double clicking on the nTerim desktop icon. Then click on "New Fixed Term Test" from the menu bar at the top of the window. A "Study Goal and Design" window will appear.

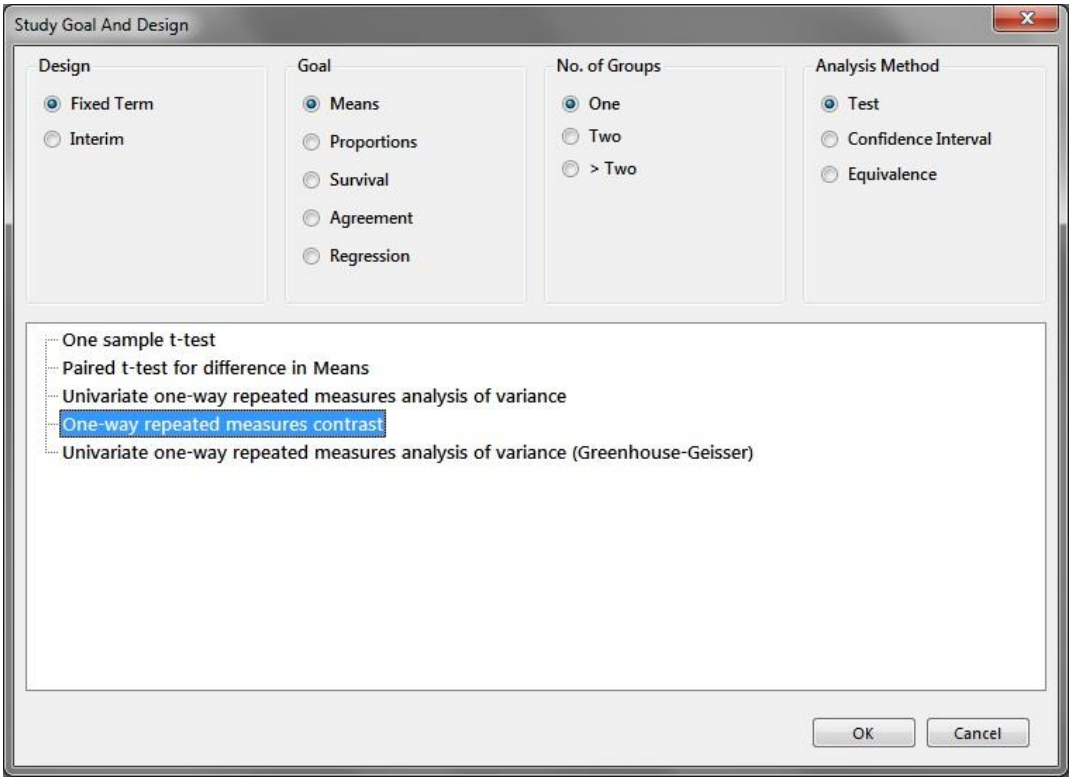


Figure 4.1.7. Study Goal and Design Window

2. Once the correct test has been selected, click "OK" and the test window will appear.

3. There are two main tables required for this test, the main test table illustrated in Figure 4.1.8 and the effect size assistant table shown in Figure 4.1.9.

4. Enter 0.05 for alpha, the desired significance level, and enter 3 for the number of levels, M, as shown in Figure 4.1.10.

5. Now you are required to complete the "Compute Effect Size Assistant" table in order to calculate values for the Contrast (C) and Scale (D) parameters.

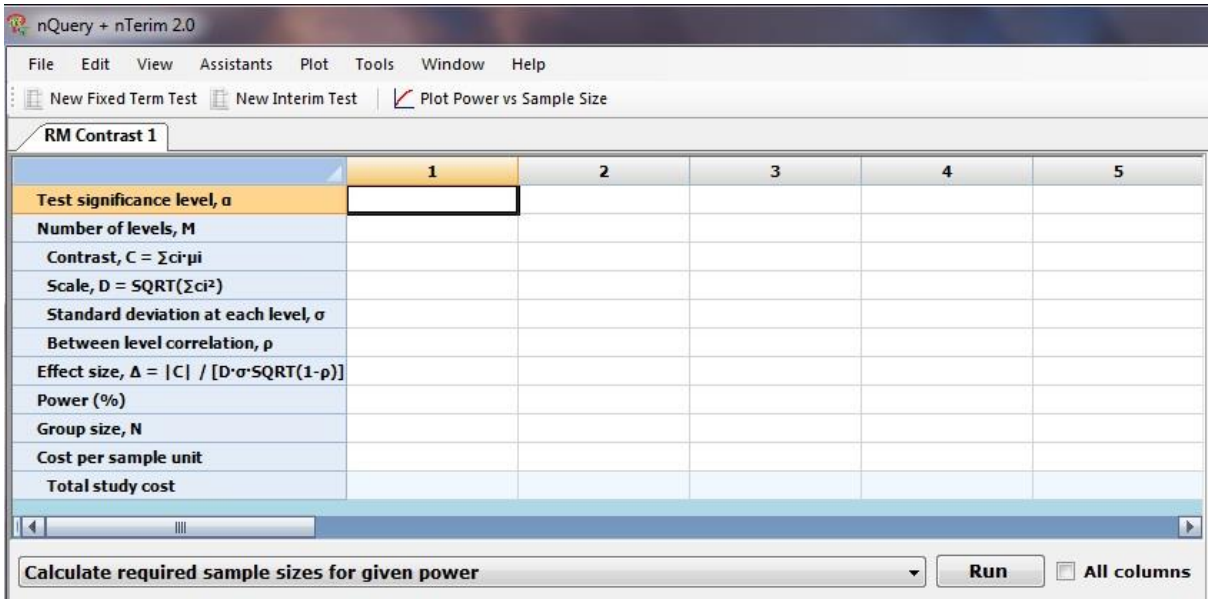


Figure 4.1.8. One-way Repeated Measures Contrast Test Table

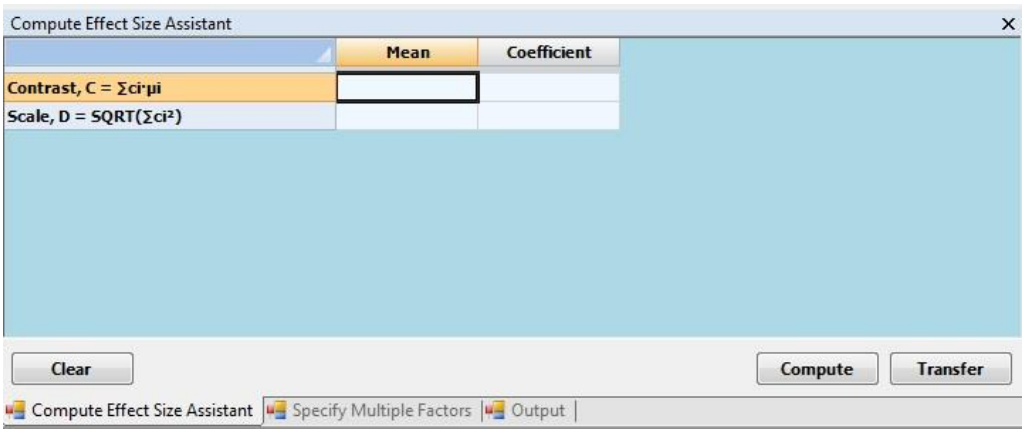


Figure 4.1.9. Compute Effect Size Assistant Table

6. Once you enter a value for the number of levels, M, the "Compute Effect Size Assistant" table automatically updates as shown in Figure 4.1.10.

7. In order to calculate a value for Effect Size, two parameters need to be calculated first, the Contrast (C) and Scale (D).

8. The mean for each level and the corresponding coefficient value need to be entered in the "Compute effect Size Assistant" table.

9. For the "Mean" values for each level, enter 6 for level 1, 3 for level 2 and 3 for level 3.

10. For the "Coefficient" values for each level, enter 2 for level 1, -1 for level 2 and -1 for level 3. The sum of these values must always equate to zero. This is illustrated in Figure 4.1.11 below.

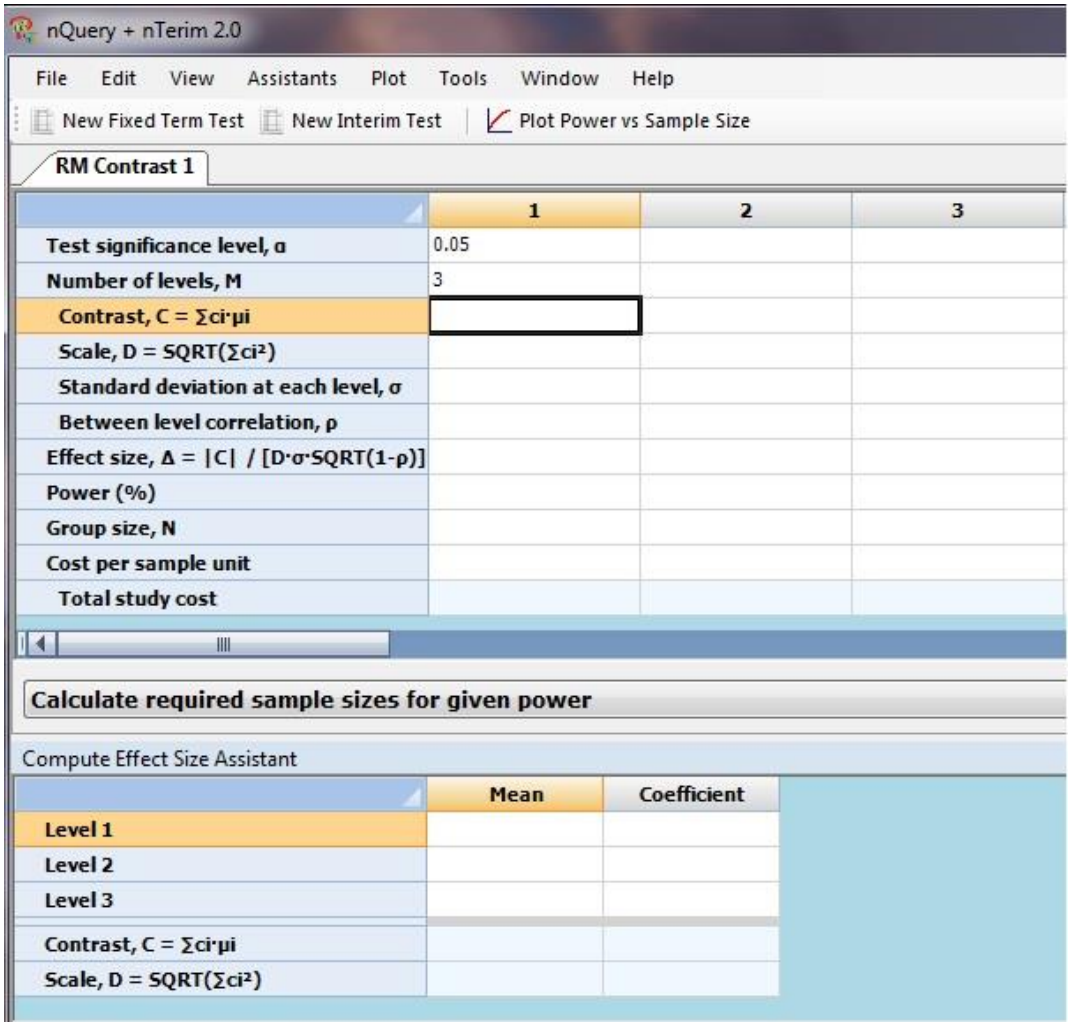


Figure 4.1.10. Automatically Updated Compute Effect Size Assistant Table

11. Once the table in Figure 4.1.11 is completed, and values for Contrast (C) and Scale (D) are computed, click on "Transfer" to automatically transfer these values to the main table.

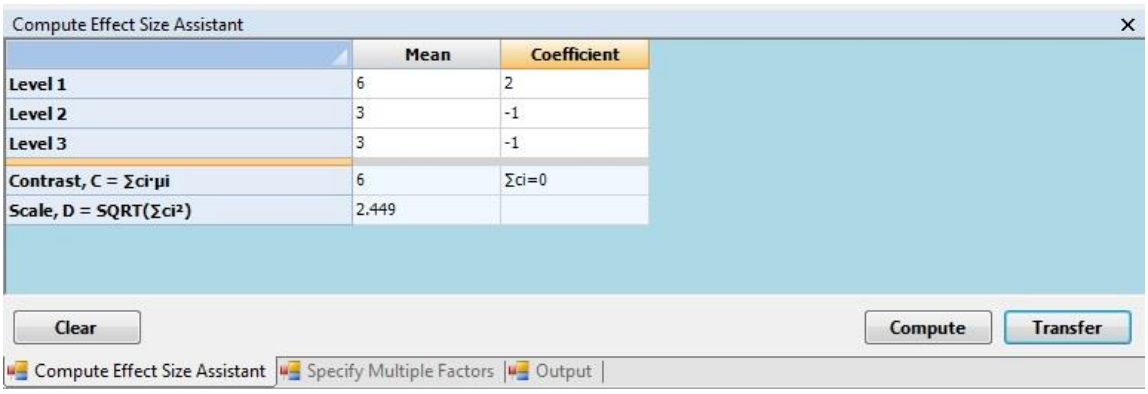


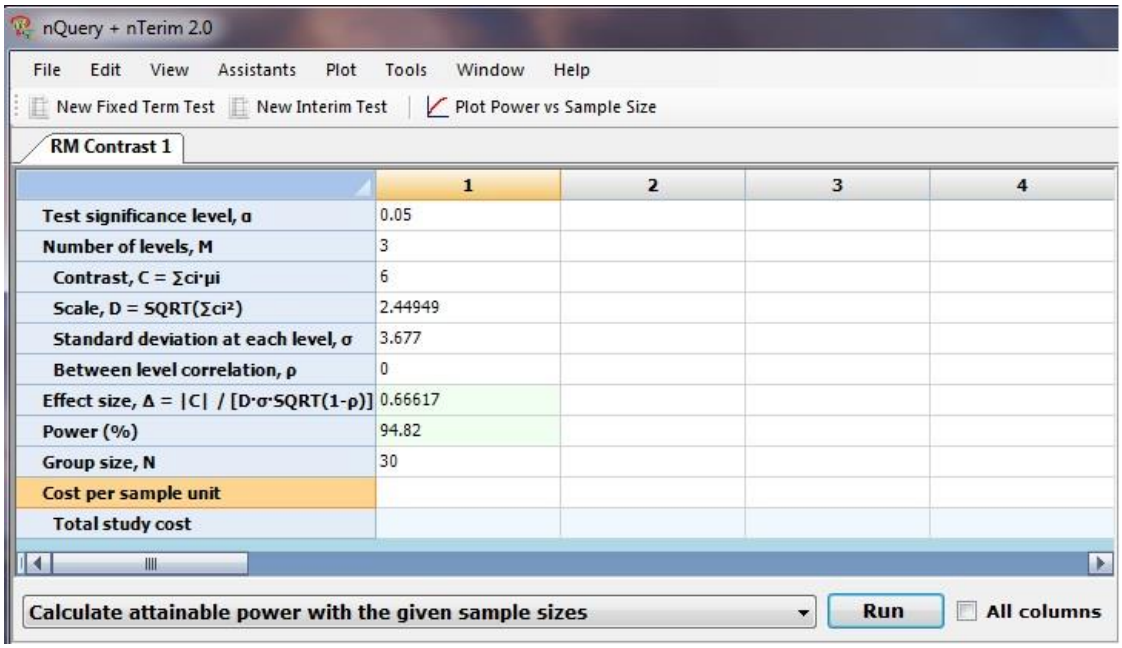
Figure 4.1.11. Completed Compute Effect Size Assistant Table

12. Now that values for Contrast (C) and Scale (D) have been computed we can continue with filling in the main table. For the Standard Deviation, enter a value of 3.677. For the between level correlation, enter a value of 0.

13. We want to calculate the attainable power given the sample size, therefore enter 30 in the "Group size, n" row.

14. The cost per sample unit cannot be estimate yet in this study so we will leave this row blank for this calculation. This value has no impact on the sample size or power calculation.

15. Select "Calculate attainable power with the given sample sizes" from the drop-down menu below the main table and click "Run". This is displayed in Figure 4.1.12.



	1	2	3	4
Test significance level, α	0.05			
Number of levels, M	3			
Contrast, C = ∑ci·μi	6			
Scale, D = SQRT(∑ci²)	2.44949			
Standard deviation at each level, σ	3.677			
Between level correlation, ρ	0			
Effect size, Δ = |C| / [D·σ·SQRT(1-ρ)]	0.66617			
Power (%)	94.82			
Group size, N	30			
Cost per sample unit				
Total study cost				

Figure 4.1.12. Completed One-way Repeated Measures Contrast Table

It can be seen from Figure 4.1.12 that a sample size of 30 per group (for each of the three groups, thus a total sample size, N, of 90) is required to obtain a power of 94.82%.

By clicking on the Output tab at the bottom of the screen you can see a statement giving details of the calculation:

"When the group sample size (n) is 30, the test of a single contrast at the 0.05 level in a one way repeated measures analysis of variance with 3 levels will have 94.82% power to detect a contrast C = ∑ci·μi of 6, with a scale D = SQRT(∑ci²) of 2.44949, assuming a standard deviation at each level of 3.677 and a between level correlation of 0."

Example 3: Investigating Self-Esteem Scores over time

In this example we will be examining self-esteem scores over time. For the researchers involved, they expect the self-esteem scores to increase monotonically over time. Therefore, the researchers would wish to test the linear contrast following the repeated measures ANOVA to assess what sample size is requires for the contrast to have 90% power.

The following steps outline the procedure for Example 3.

1. Open nTerim through the Start Menu or by double clicking on the nTerim desktop icon. Then click on "New Fixed Term Test" from the menu bar at the top of the window. A "Study Goal and Design" window will appear.

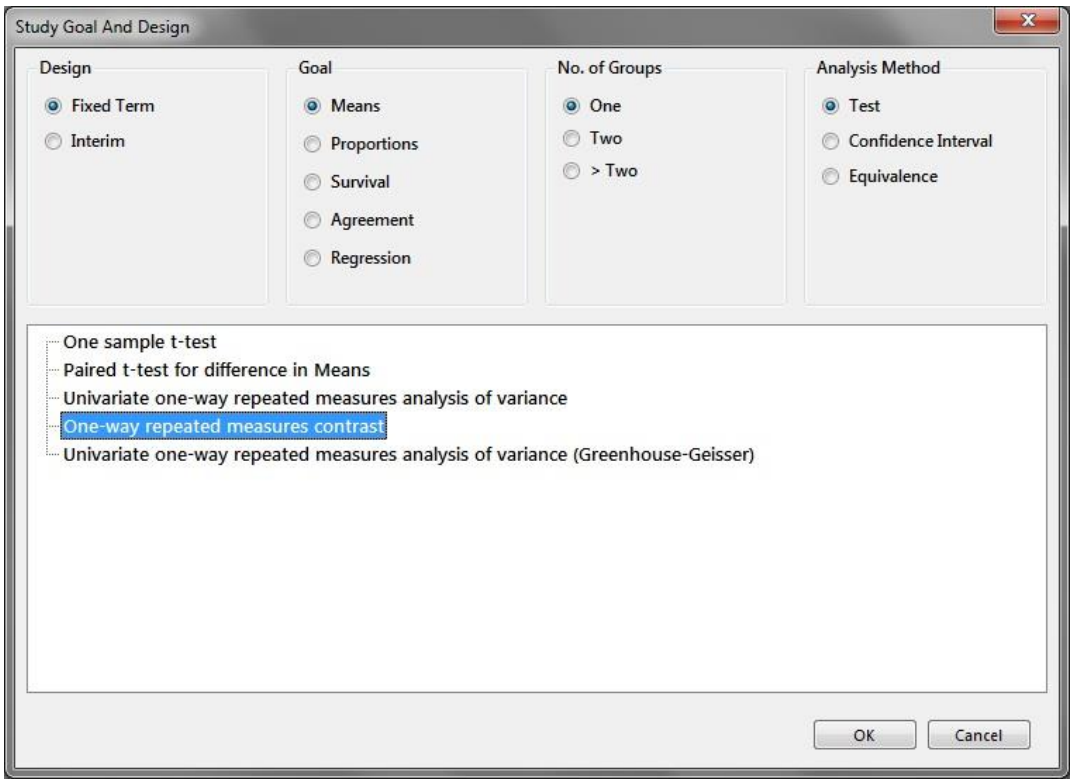


Figure 4.1.13. Study Goal and Design Window

2. Once the correct test has been selected, click "OK" and the test window will appear.

3. There are two main tables required for this test, the main test table illustrated in Figure 4.1.14 and the effect size assistant table shown in Figure 4.1.15.

4. Enter 0.05 for alpha, the desired significance level, and enter 4 for the number of levels, M, as shown in Figure 4.1.16.

5. Now you are required to complete the "Compute Effect Size Assistant" table in order to calculate values for the Contrast (C) and Scale (D) parameters.

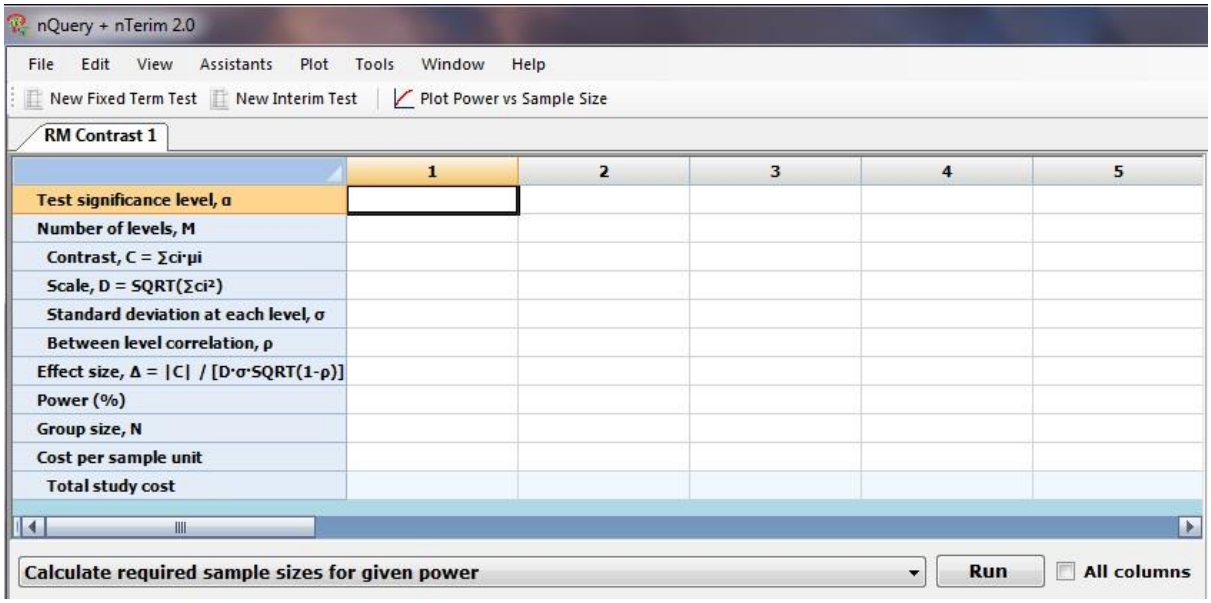


Figure 4.1.14. One-way Repeated Measures Contrast Test Table

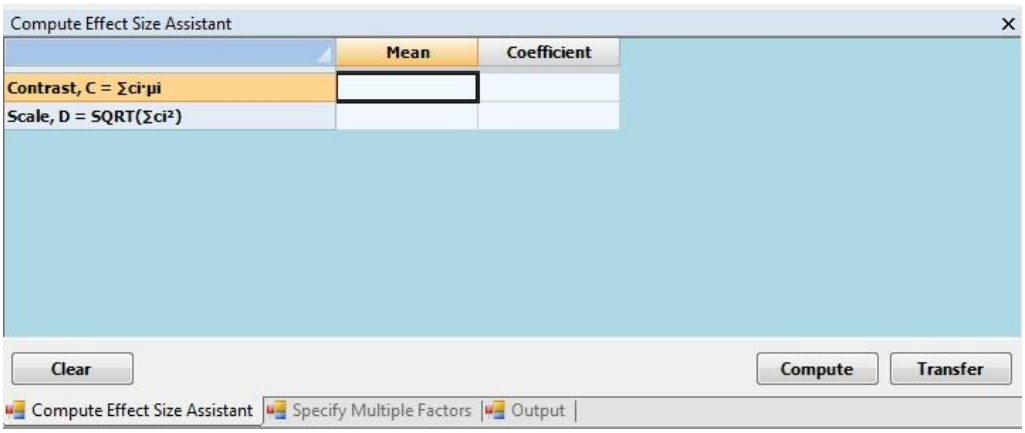


Figure 4.1.15. Compute Effect Size Assistant Table

6. Once you enter a value for the number of levels, M, the "Compute Effect Size Assistant" table automatically updates as shown in Figure 4.1.16.

7. In order to calculate a value for Effect Size, two parameters need to be calculated first, the Contrast (C) and Scale (D).

8. The mean for each level and the corresponding coefficient value need to be entered in the "Compute effect Size Assistant" table.

9. For the "Mean" values for each level, enter 55 for level 1, 56.5 for level 2, 58 for level 3 and 59.5 for level 4.

10. For the "Coefficient" values for each level, enter -3 for level 1, -1 for level 2, 1 for level 3 and 3 for level 4. The sum of these values must always equate to zero. This is illustrated in Figure 4.1.17 below.

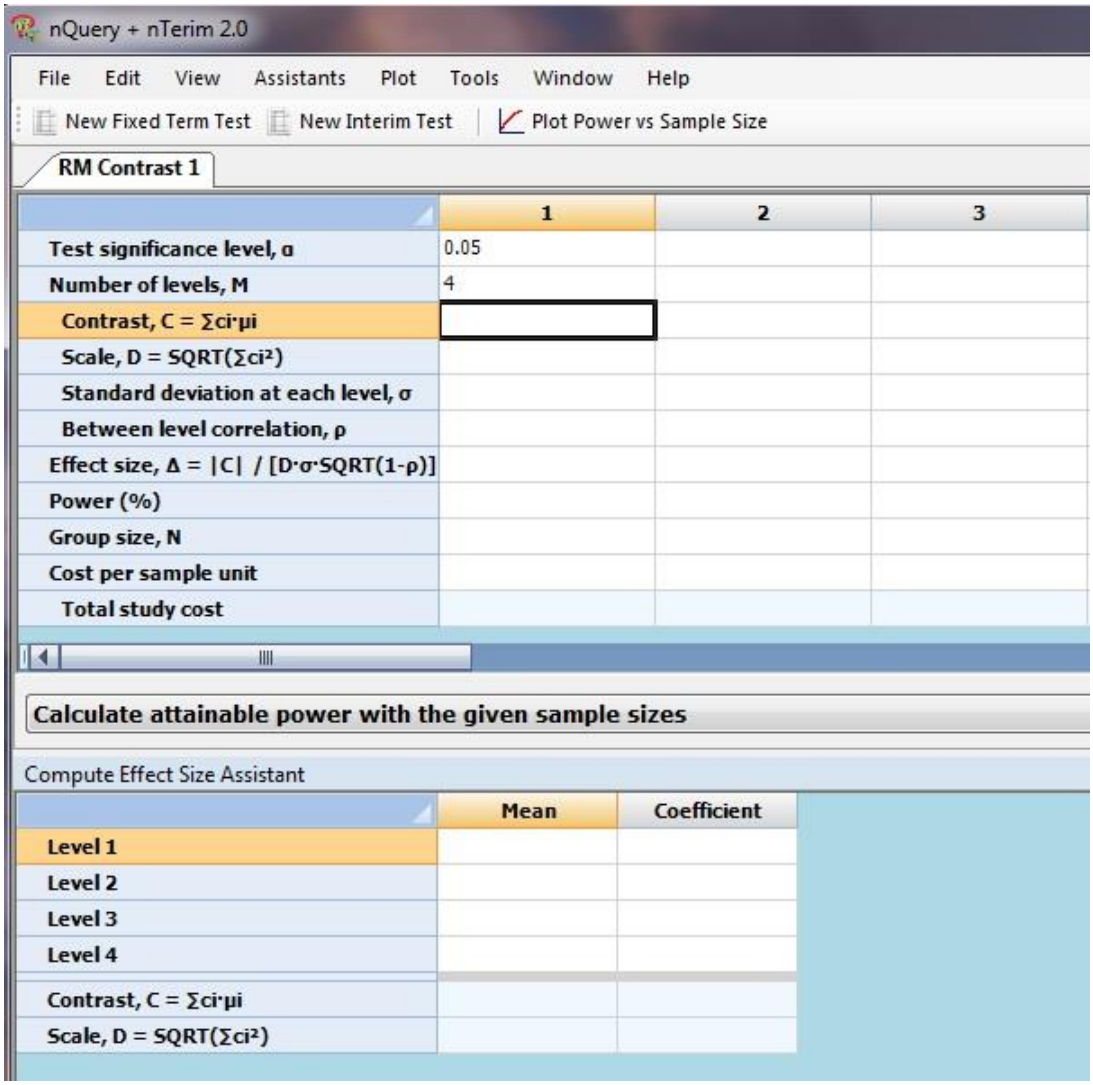


Figure 4.1.16. Automatically Updated Compute Effect Size Assistant Table

11. Once the table in Figure 4.1.17 is completed, and values for Contrast (C) and Scale (D) are computed, click on "Transfer" to automatically transfer these values to the main table.

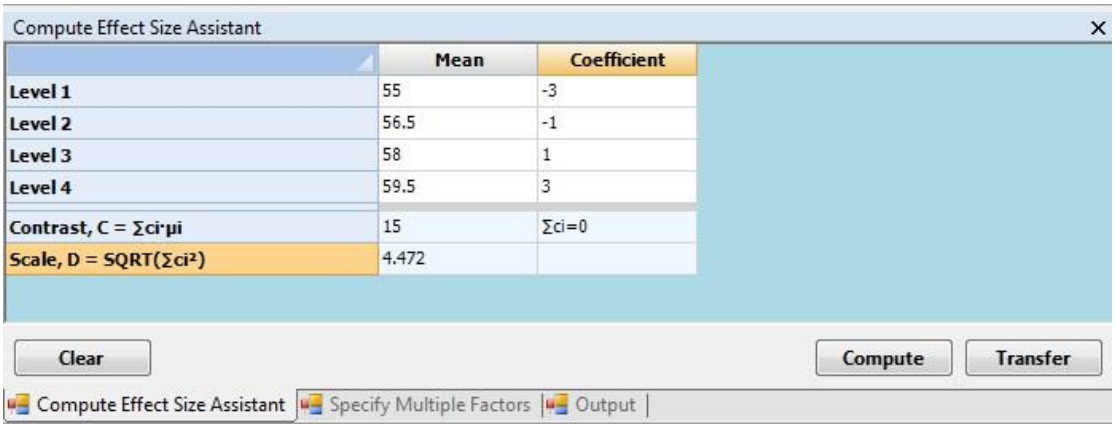


Figure 4.1.17. Completed Compute Effect Size Assistant Table

12. Now that values for Contrast (C) and Scale (D) have been computed we can continue with filling in the main table. For the Standard Deviation, enter a value of 10. For the between level correlation, enter a value of 0.7.

13. We want to calculate the sample size required obtain a power of 90%. Therefore enter 90 in the Power row.

14. The cost per sample unit cannot be estimate yet in this study so we will leave this row blank for this calculation. This value has no impact on the sample size or power calculation.

15. Select "Calculate required sample size for given power" from the drop-down menu below the main table and click "Run". This is displayed in Figure 4.1.18.

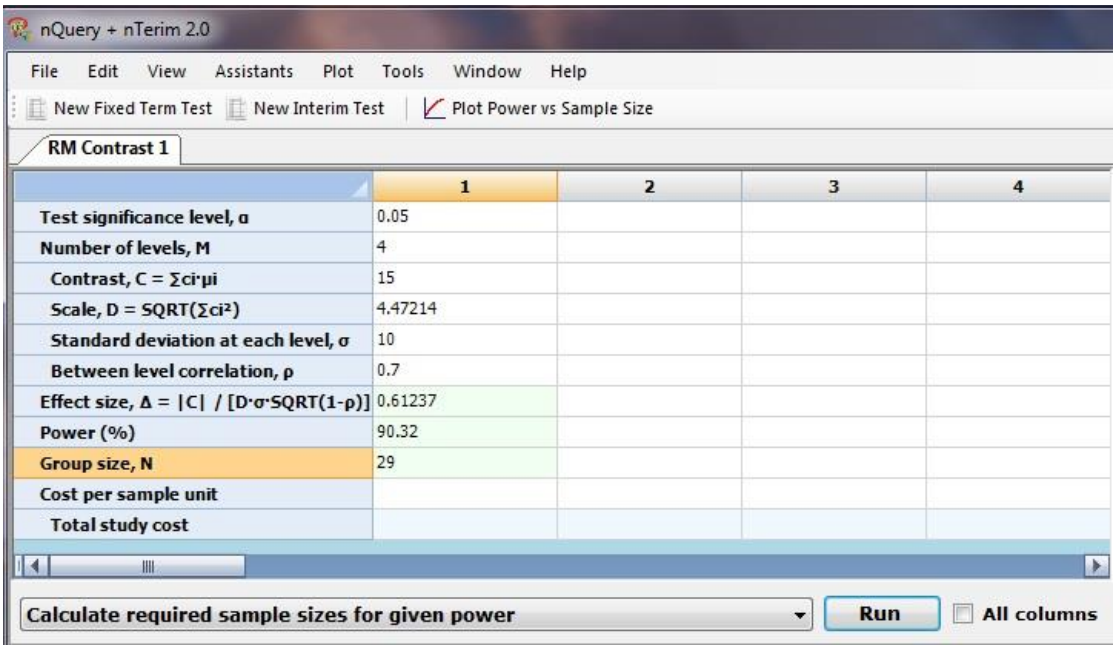


Figure 4.1.18. Completed One-way Repeated Measures Contrast Table

It can be seen from Figure 4.1.18 that a sample size of 29 per group (for each of the three groups, thus a total sample size, N, of 116) is required to obtain a power of 90.32%.

By clicking on the Output tab at the bottom of the screen you can see a statement giving details of the calculation:

"When the group sample size (n) is 29, the test of a single contrast at the 0.05 level in a one way repeated measures analysis of variance with 4 levels will have 90.32% power to detect a contrast C = ∑ci·μi of 15, with a scale D = SQRT(∑ci²) of 4.47214, assuming a standard deviation at each level of 10 and a between level correlation of 0.7."

4.2 Repeated Measures Design for Two Means

4.2.1. Introduction

A repeated measures design is an experimental design in which multiple measurements are taken on one or more groups of subjects over time or under different conditions. This type of design leads to a more precise estimate of an endpoint and can avoid the bias from a single measure. For example, an individual's blood pressure is known to be sensitive to many temporary factors such as amount of sleep had the night before, mood, excitement level, exercise, etc. If there is just a single measurement taken from each patient, then comparing the mean blood pressure between two groups could be invalid as there could be a large degree of variation in the single measures of blood pressure levels among patients. However, by obtaining multiple measurements from each individual and comparing the time averaged difference between the two groups, the precision of the experiment is increased.

This table facilitates the calculation of power and sample size for the time averaged difference between two means in a repeated measures design. Power and sample size is computed using the method outlined by Liu and Wu (2005).

4.2.2. Methodology

Power and sample size are calculated using standard normal distributions and follow the procedures outlined by Liu and Wu (2005).

To calculate power and sample size the user must first specify the test significance level, α, and choose between a one or a two sided test. The user must then enter a value for the number of levels, M. This value corresponds to the number of measurements that will be taken on each subject. Values must then be provided for the difference in means, d, the standard deviation at each level, σ, and the between level correlation, ρ. The difference in means that must be specified is the smallest meaningful time-averaged difference to be detected.

Given the above values, in order to calculate the power for this design the user must enter the expected sample size for each group, N_1 and N_2. nTerim then uses the total sample size, N, to calculate the power of the design using the following equation:

$$Power = 1 - \Phi\left(\frac{Z_\alpha - d\left(\sqrt{MN\big(\pi(1-\pi)\big)}\right)}{\sigma\big(\sqrt{1+\rho(M-1)}\big)}\right) \qquad \{4.2.1\}$$

where, $\Phi()$ is the standard normal density function, and

$$N = N_1 + N_2 \qquad \{4.2.2\}$$

$$\pi = \frac{N_1}{N} \qquad \{4.2.3\}$$

In order to calculate sample size for a given power the following formula is used:

$$N = \frac{(Z_\alpha + Z_\beta)^2(1+(M-1)\rho)(\sigma^2)}{Md^2\big(\pi(1-\pi)\big)} \qquad \{4.2.4\}$$

where, β is the probability of a type II error.

$$\beta = 1 - Power \qquad \{4.2.5\}$$

4.2.3. Examples

Example 1: Comparing the Difference in Sample Size due to change in Significance Level
In this example we are going to investigate how a difference in the level of significance for a study design can impact the sample size required to obtain a given power.

The following steps outline the procedure for Example 1.

1. Open nTerim through the Start Menu or by double clicking on the nTerim desktop icon. Then click on "New Fixed Term Test" from the menu bar at the top of the window. A "Study Goal and Design" window will appear.

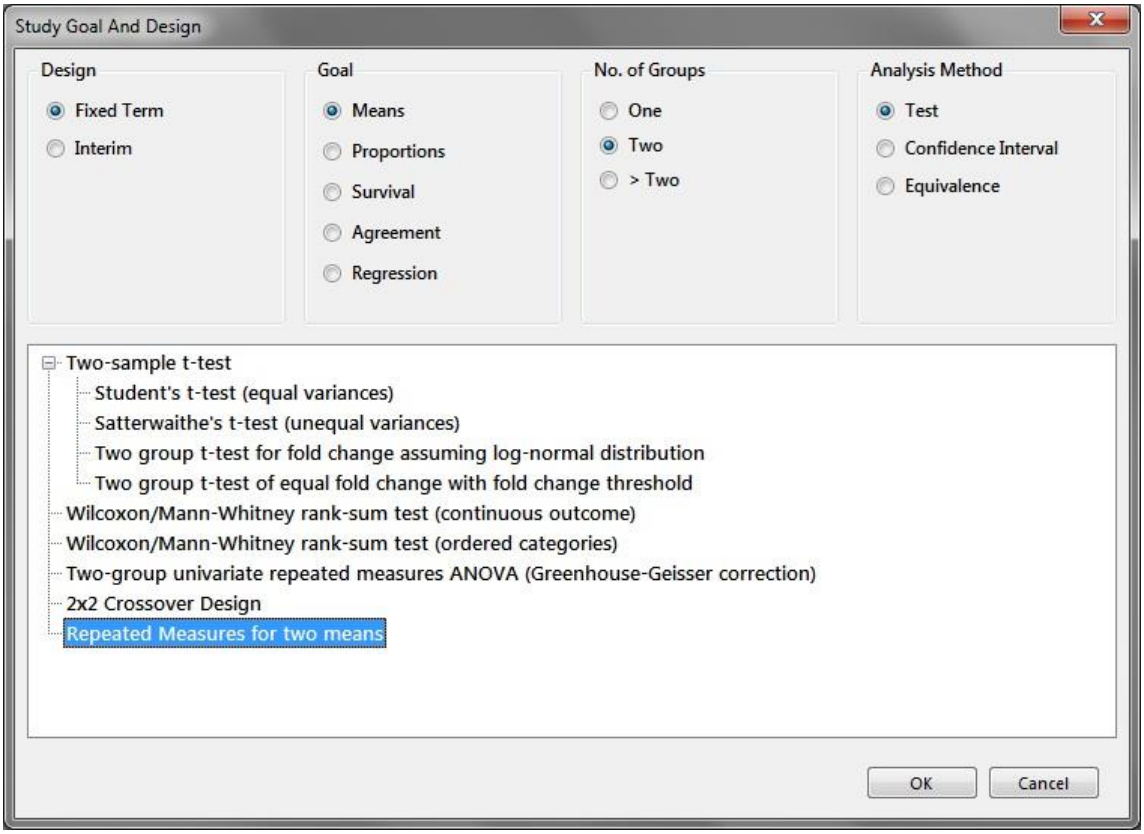


Figure 4.2.1. Study Goal and Design Window

2. Once the correct test has been selected, click "OK" and the test window will appear. This test table is illustrated in Figure 4.2.2.

3. Enter 0.05 for alpha, the desired significance level, and enter 4 for the number of levels, M, as shown in Figure 4.2.4.

4. Two sided test is the default setting in nTerim as well as a Ratio value of 1 for the group sizes.

5. In this example we will examine a study where the difference in means is 10 and the standard deviation at each level is 20. Therefore, enter a value of 10 in the "Difference in Means" row and a value of 20 in the "Standard deviation at each level" row.

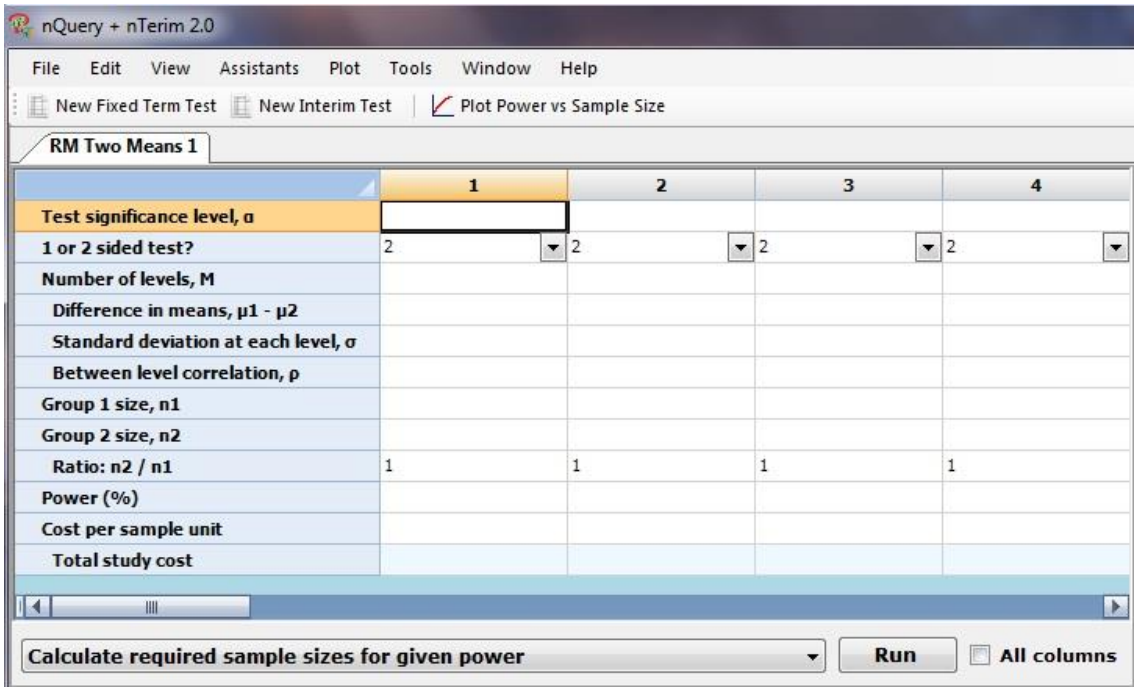


Figure 4.2.2. Repeated Measures for Two Means Test Table

6. We also know that the between level correlation is 0.5 so enter 0.5 into the "Between level correlation" row.

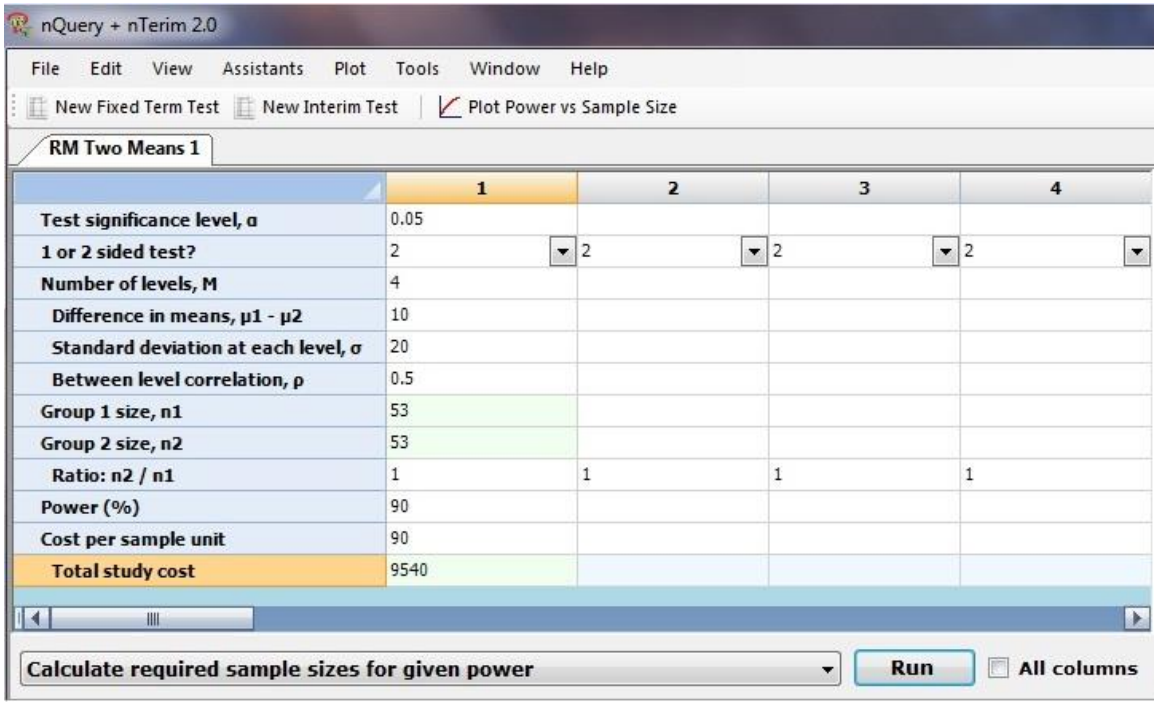


Figure 4.2.3. Completed Repeated Measures Design for Two Means

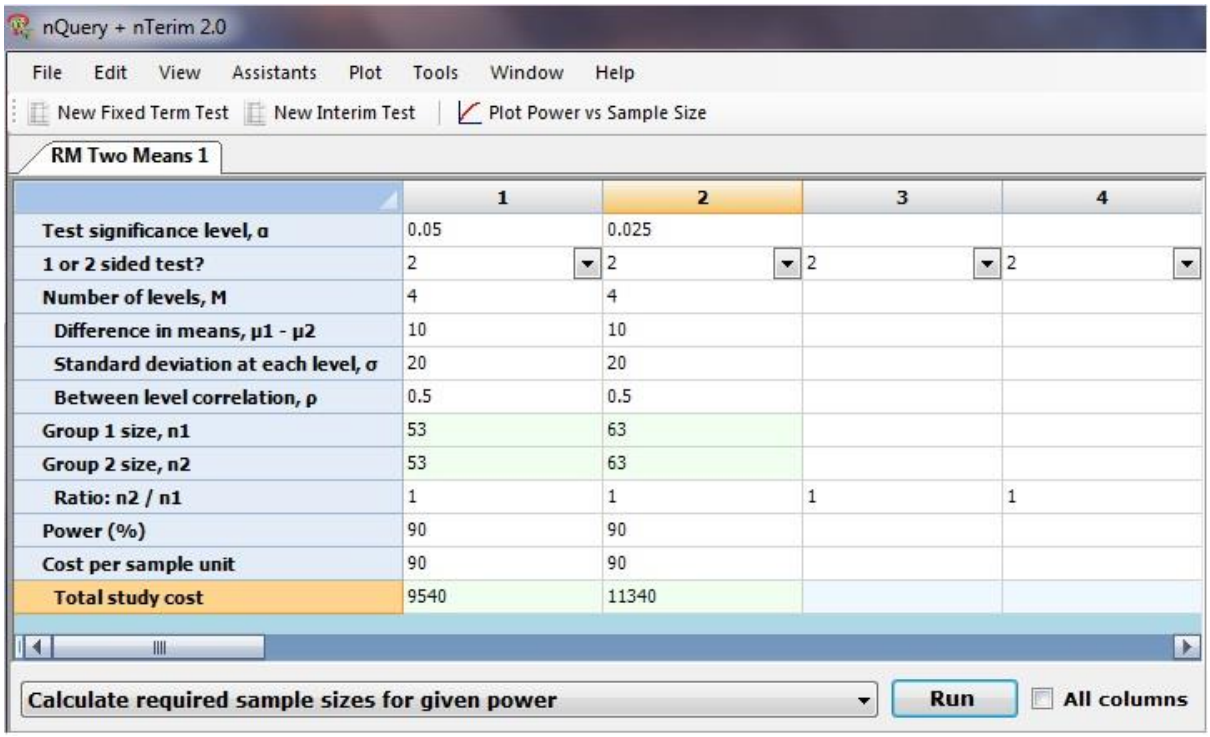
7. We want to calculate the required sample size for each group in order to obtain 90% power. To do this, enter 90 in the “Power (%)” row.

8. It has also been estimated that the associated cost per unit in this study will amount to $90. Therefore enter 90 in the “Cost per sample unit” row in order to calculate the Total study cost associated with the sample size.

9. Then select “Calculate required sample size for given power” from the drop-down menu below the main table and click “Run”. This is displayed in Figure 4.2.3 above.

By clicking on the Output tab at the bottom of the screen you can see a statement giving details of the calculation:

“When the sample size is 53 in group 1 and 53 in group 2, a test for the time averaged difference between two means in a repeated measures design with a 0.05 significance level will have 90% power to detect a difference in means of 10 in a design with 4 repeated measurements when the standard deviation is 20 and the between level correlation is 0.5.”



	1	2	3	4
Test significance level, α	0.05	0.025		
1 or 2 sided test?	2	2	2	2
Number of levels, M	4	4		
Difference in means, μ1 - μ2	10	10		
Standard deviation at each level, σ	20	20		
Between level correlation, ρ	0.5	0.5		
Group 1 size, n1	53	63		
Group 2 size, n2	53	63		
Ratio: n2 / n1	1	1	1	1
Power (%)	90	90		
Cost per sample unit	90	90		
Total study cost	9540	11340		

Figure 4.2.4. Re-run calculations to update Column 2

10. Now we are going to repeat the same study design example except we’re going to enforce a stricter level of significance. In the second column, enter 0.025 in the “Test Significance Level” row. Now we are looking for a 2.5% level of significance instead of a 5% level as in the first column.

11. We want to see the effects of changing the level of significance has on sample size, and perhaps the total study cost.

12. Enter the same information for number of levels, Difference in Means, standard deviation at each level, between level correlation, power and cost per sample unit.

13. Select “Calculate required sample size for given power” from the drop-down menu below the main table and click “Run”. This is displayed in Figure 4.2.4 above.

It can be seen from Figure 4.2.4 that sample size has increase be 20 (10 per group) and the estimated cost has increased by $1,800.

14. Another feature that enables us to compare designs side-by-side is by using the Power vs. Sample Size plot. Multiple columns can be plotted together by simply highlighting the desired columns and clicking on the “Plot Power vs Sample Size” button on the menu bar.

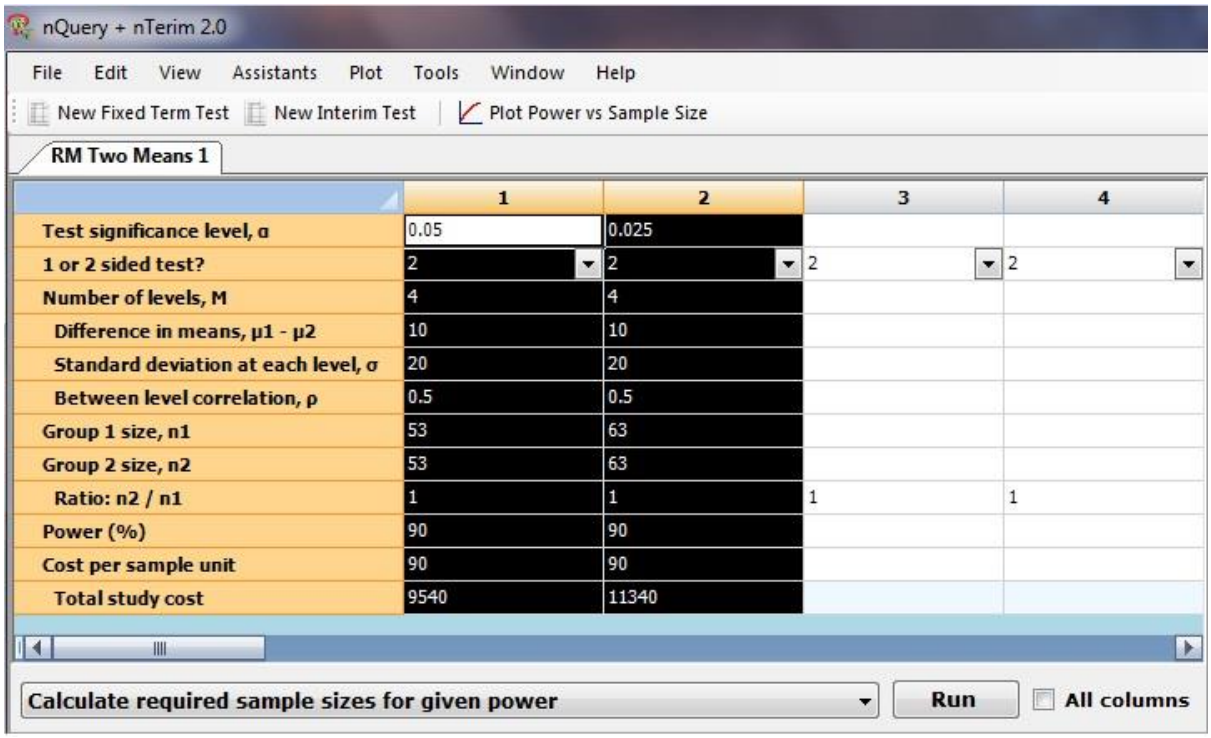


Figure 4.2.5. Highlight desired columns for plotting

15. To highlight the desired columns, click on the column title for Column 1 and drag across to Column 2 as illustrated in Figure 4.2.5.

16. Then click on the “Plot Power vs Sample Size” button on the menu bar. The multiple column plot is displayed in Figure 4.2.6.

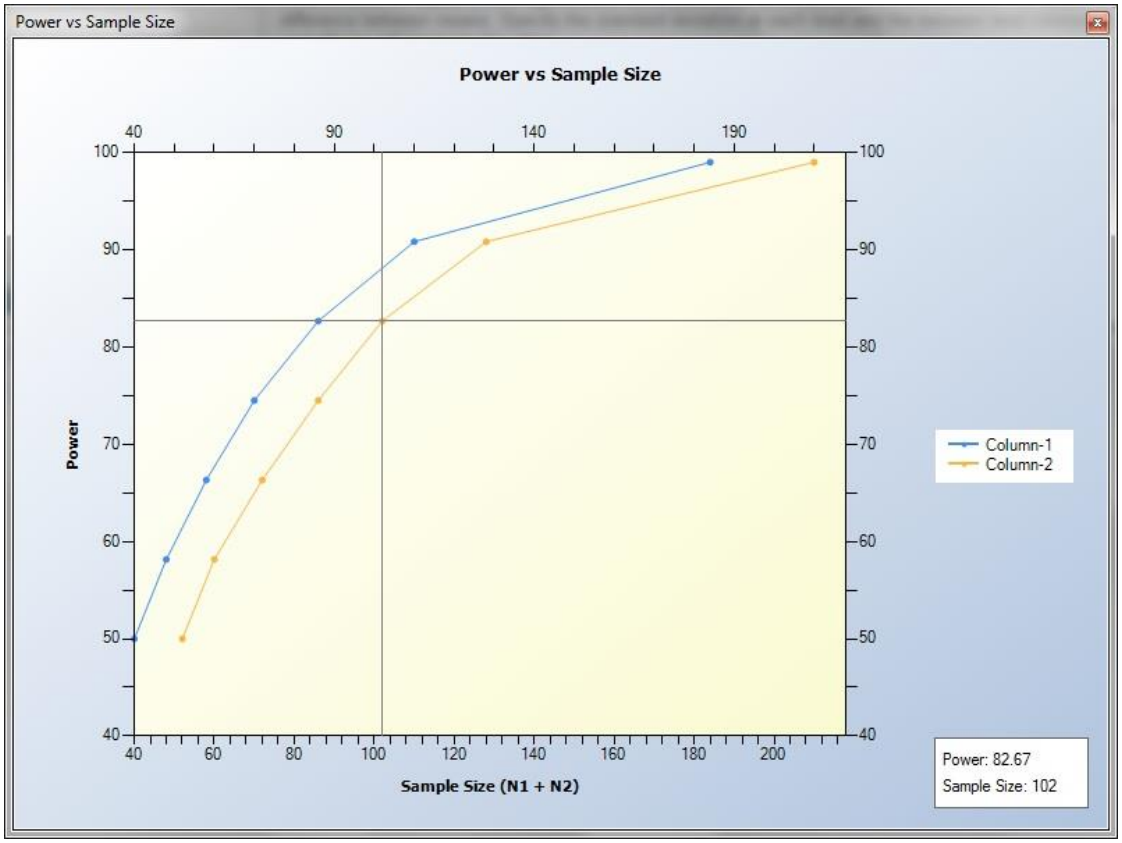


Figure 4.2.6. Power vs. Sample Size Plot

It can be seen from the legend on the left-hand side (legend can be altered manually) that the blue line represents Column 1 and the orange line represents Column 2. The cross on the graph illustrates how the user can identify what the sample size is for a corresponding power value for each column. In the bottom right corner of the plot indicated the exact values for Power and Sample Size for each identifier on the graph.

It can be seen in Figure 4.2.6 that Column 1 reaches an acceptable power level faster than the design in Column 2. The researcher can now make an assessment as to which design they would prefer to use.

Example 2: Differences in Power and Between Level Correlations

In this example we investigate how a change in Power and a change in Between Level Correlation has an effect on sample size.

The following steps outline the procedure for Example 2.

1. Open nTerim through the Start Menu or by double clicking on the nTerim desktop icon. Then click on "New Fixed Term Test" from the menu bar at the top of the window. A "Study Goal and Design" window will appear.

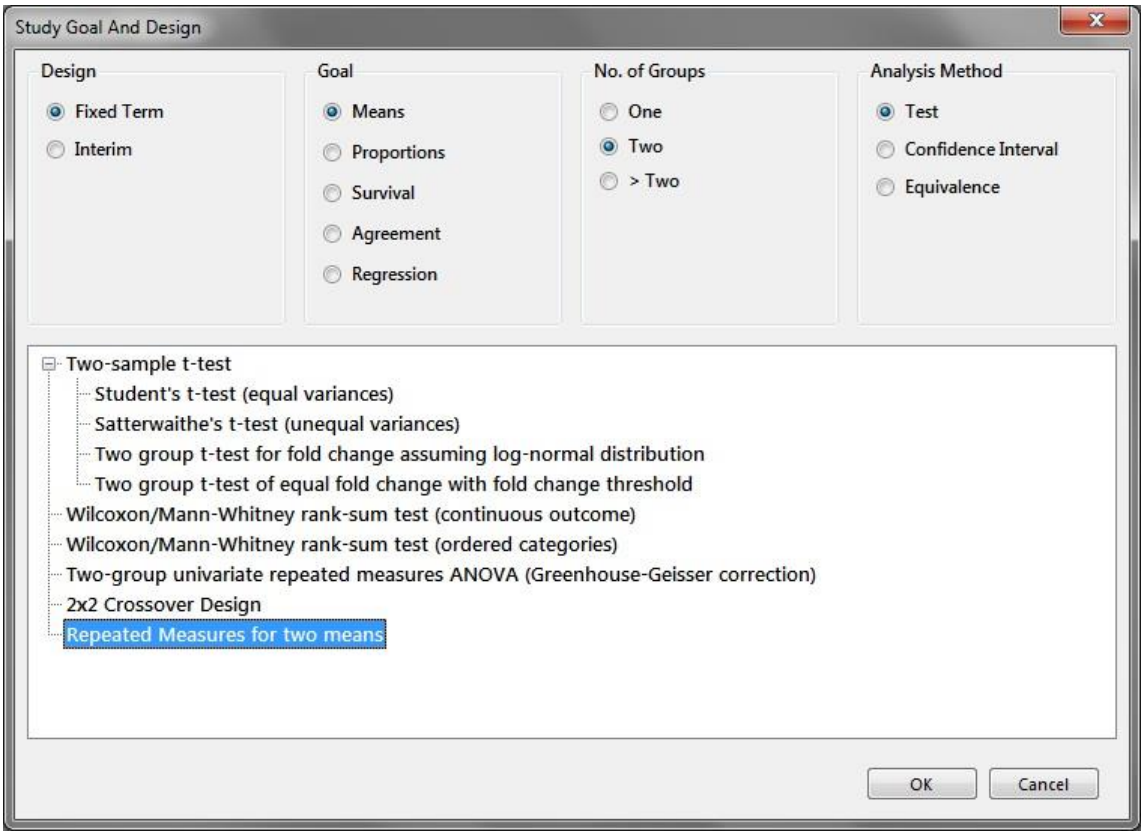


Figure 4.2.7. Study Goal and Design Window

2. Once the correct test has been selected, click "OK" and the test window will appear. This test table is illustrated in Figure 4.2.8.

3. Enter 0.05 for alpha, the desired significance level, and enter 4 for the number of levels, M, as shown in Figure 4.2.9.

4. Two sided test is the default setting in nTerim as well as a Ratio value of 1 for the group sizes.

5. In this example we will examine a study where the difference in means is 15 and the standard deviation at each level is 25. Therefore, enter a value of 10 in the "Difference in Means" row and a value of 20 in the "Standard deviation at each level" row as shown in Figure 4.2.9.

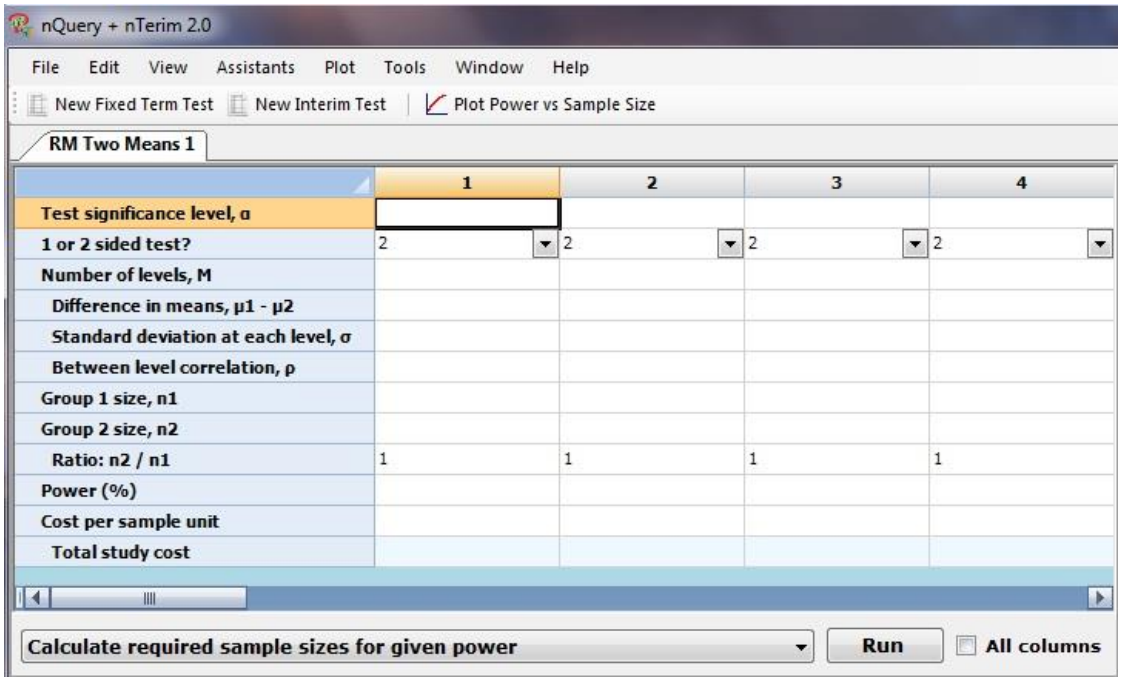


Figure 4.2.8. Repeated Measures for Two Means Test Table

6. The between level correlation is estimated as 0.4, so enter 0.4 in the "Between level correlation" row.

7. We want to calculate the required sample size to obtain a power of 90% so enter 90 on the "Power" row.

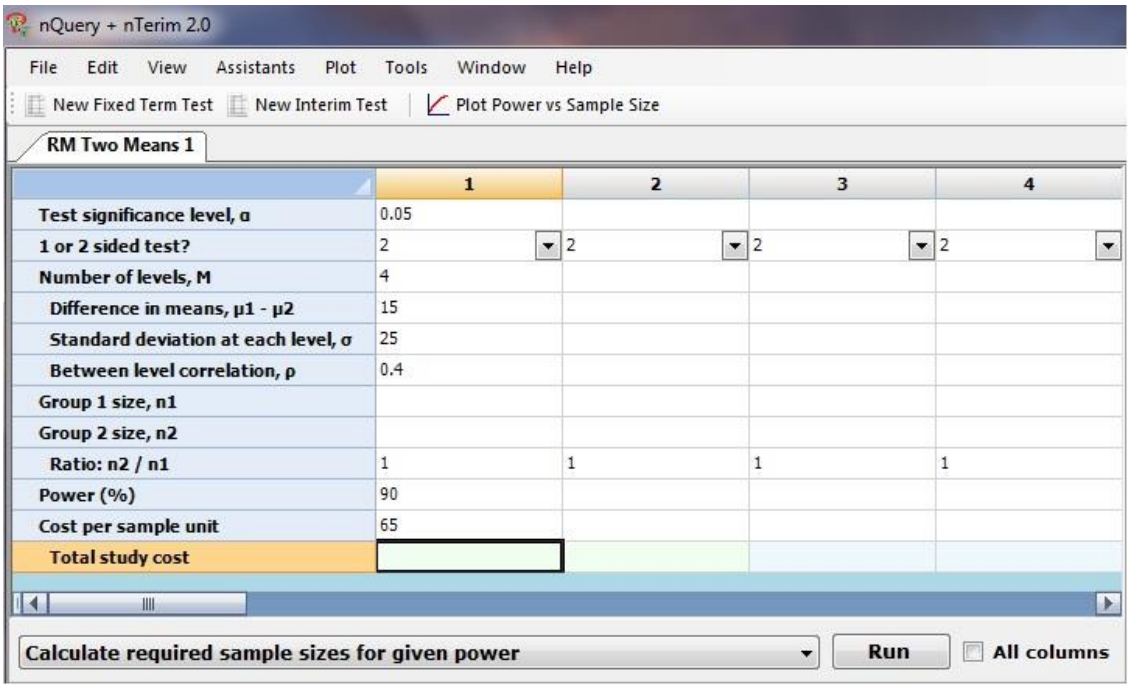


Figure 4.2.9. Design Entry for Two Means Repeated Measures Study

8. The cost per sample unit has been estimated as $65 in this particular study. Therefore to calculate the overall cost associated with the sample size, enter 65 in the “Cost per sample unit” row as shown in Figure 4.2.9.

9. As we want to try several different parameter values for both Power and between level correlation, we can use the “Fill Right” function to fill out multiple columns with the same information entered in Column 1.

10. Once all the parameter information has been entered, click on Edit and Fill Right as shown in Figure 4.2.10.

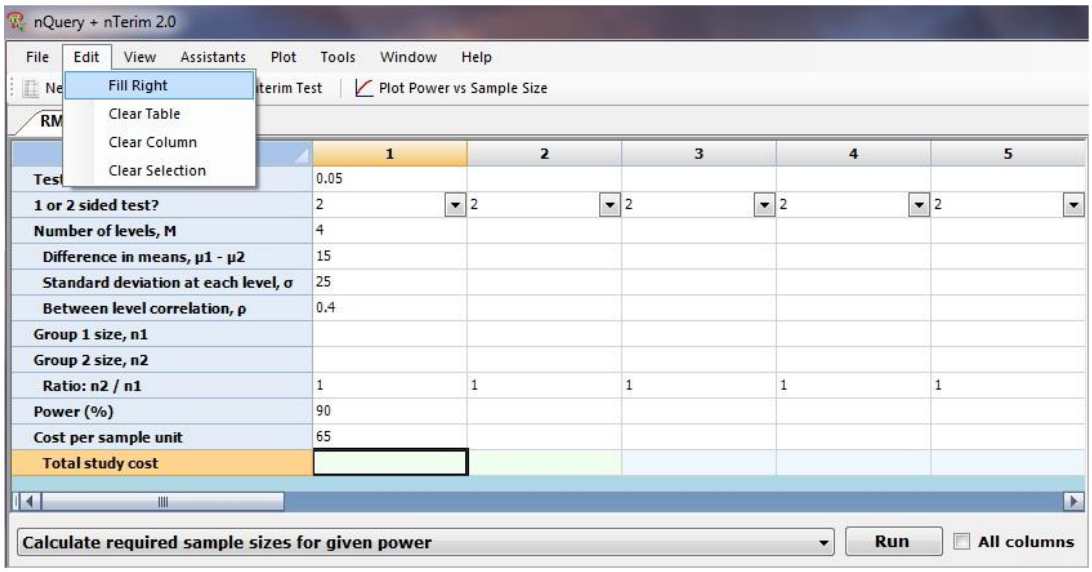


Figure 4.2.10. Fill Right function

11. As shown in Figure 4.2.11, all columns have been filled in with the same parameter information contained in Column 1. We want to alter the other columns (Columns 2 to 5) to see how the sample size is affected by various parameter changes.

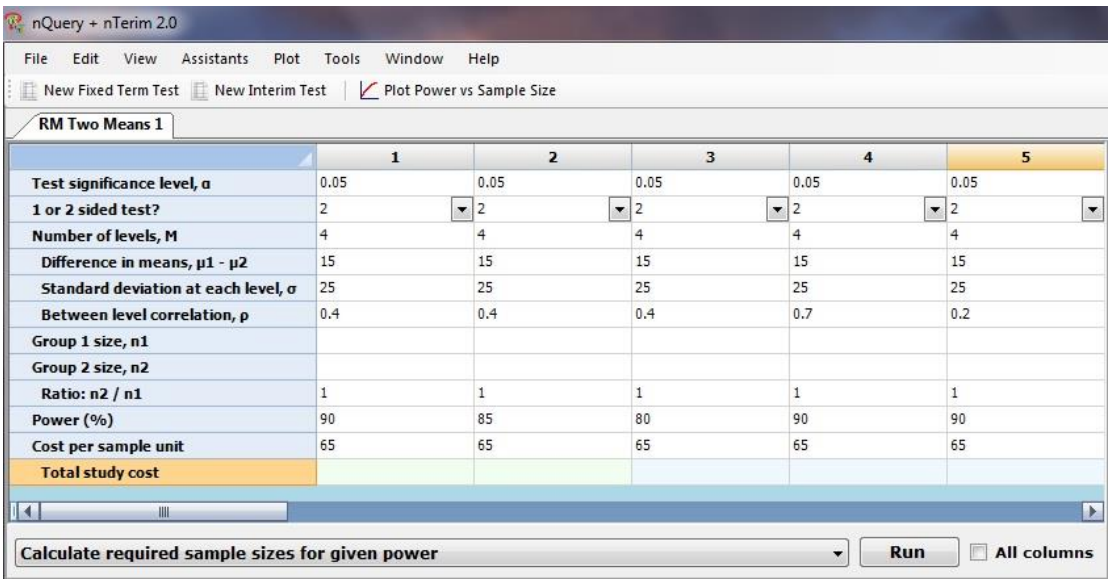


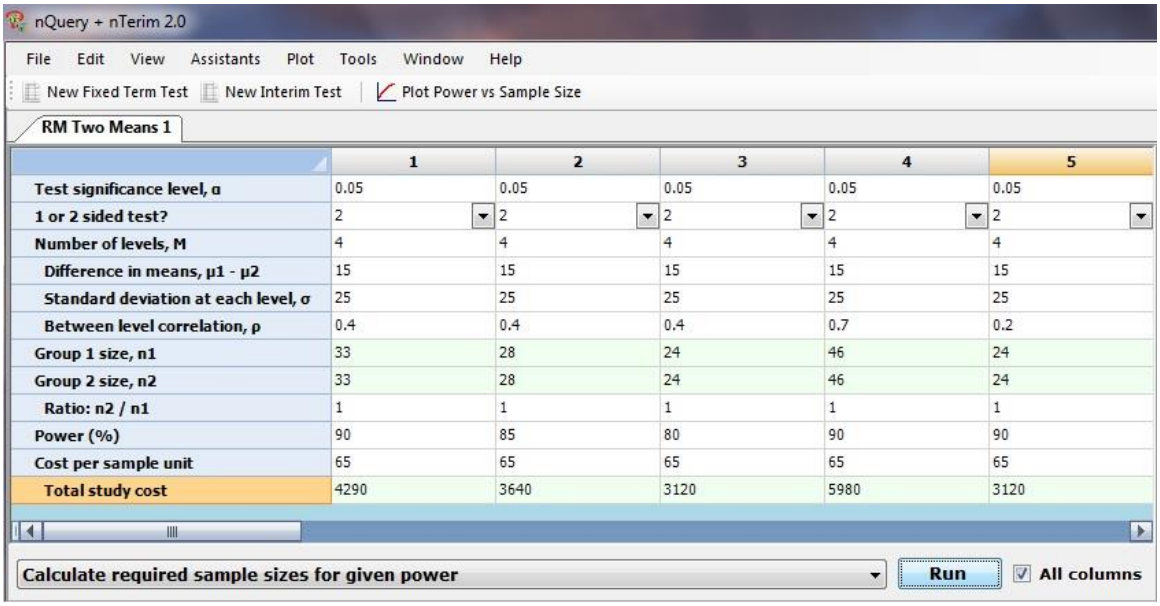
Figure 4.2.11. Altered columns for comparison

12. Firstly we want to investigate how the sample size will be affected by a change in Power. To do this we will enter 85 and 80 in the “Power” row for Columns 2 and 3 respectively as shown in Figure 4.2.11.

13. We also would like to examine how the sample size is affected by an increase or decrease in the between level correlation. Therefore we will change the between level correlation to 0.7 and 0.2 in Columns 4 and 5 respectively as shown in Figure 4.2.11.

14. As we want to calculated the required sample size to obtain the given power, select “Calculate required sample sizes for given power” from the drop-down menu below the test table.

15. As we want to run this calculation for multiple columns, tick the “All Columns” box beside the “Run” button as shown in Figure 4.2.12, then click Run.



	1	2	3	4	5
Test significance level, α	0.05	0.05	0.05	0.05	0.05
1 or 2 sided test?	2	2	2	2	2
Number of levels, M	4	4	4	4	4
Difference in means, μ1 - μ2	15	15	15	15	15
Standard deviation at each level, σ	25	25	25	25	25
Between level correlation, ρ	0.4	0.4	0.4	0.7	0.2
Group 1 size, n1	33	28	24	46	24
Group 2 size, n2	33	28	24	46	24
Ratio: n2 / n1	1	1	1	1	1
Power (%)	90	85	80	90	90
Cost per sample unit	65	65	65	65	65
Total study cost	4290	3640	3120	5980	3120

Figure 4.2.12. Completed multiple design Repeated Measures for Two Means Table

As it can be seen in Figure 4.2.12, there is a drop in sample size of 5 units per group if you reduce the power to 85% and a further drop of 4 units per group when reducing power to 80%. Depending on the different constraints on the study design, 80% power may be acceptable and would reduce costs by approximately 25% when compared with the same study design with 90% power.

When we examined the volatility in relation to the between level correlation (and keeping the power fixed at 90%), we can see that as the between level correlation increases, so does the sample size required. With a lower between level correlation, a lower sample size is required.

Example 3: Differences in Group Size Ratios

In this example we investigate how the sample size ratio between Group 1 and Group 2 affects the overall sample size required to obtain a given power.

The following steps outline the procedure for Example 3.

1. Open nTerim through the Start Menu or by double clicking on the nTerim desktop icon. Then click on "New Fixed Term Test" from the menu bar at the top of the window. A "Study Goal and Design" window will appear.

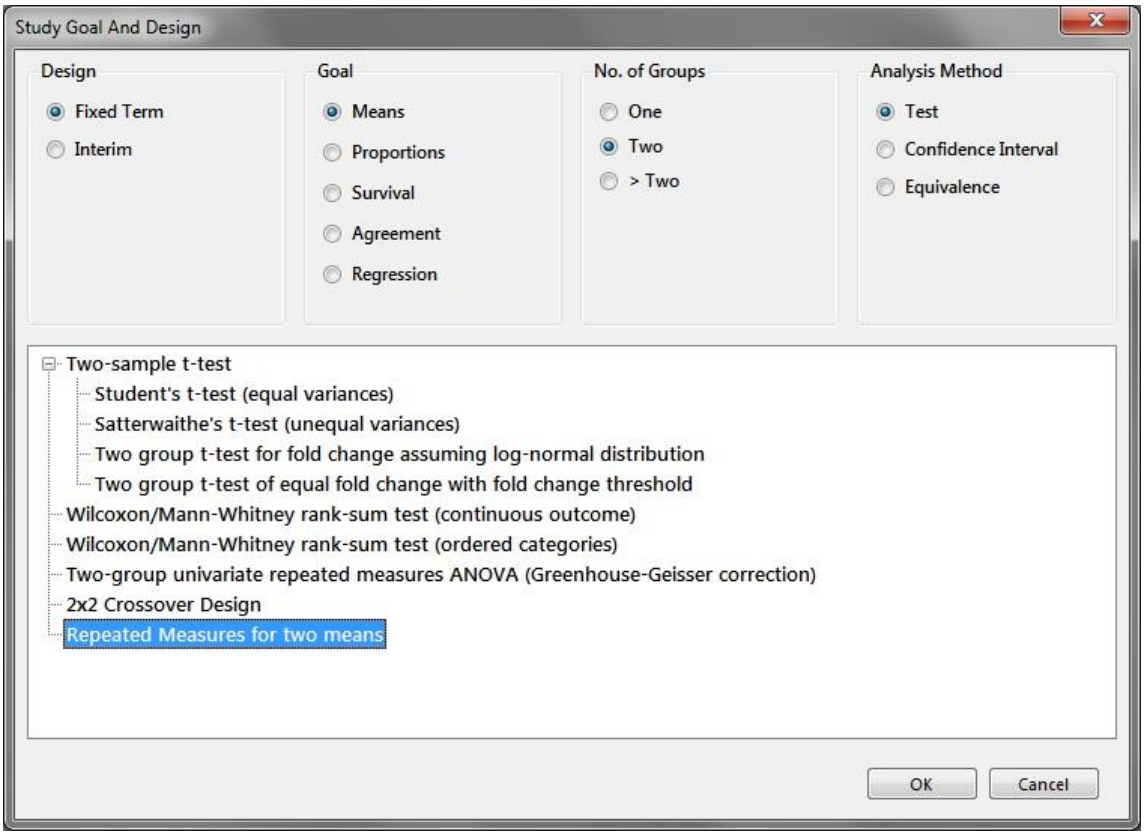


Figure 4.2.13. Study Goal and Design Window

2. Once the correct test has been selected, click "OK" and the test window will appear. This test table is illustrated in Figure 4.2.14.

3. Enter 0.05 for alpha, the desired significance level, and enter 5 for the number of levels, M, as shown in Figure 4.2.15.

4. Two sided test is the default setting in nTerim as well as a Ratio value of 1 for the group sizes.

5. In this example we will examine a study where the difference in means is 40 and the standard deviation at each level is 80. Therefore, enter a value of 40 in the "Difference in Means" row and a value of 80 in the "Standard deviation at each level" row as shown in Figure 4.2.15.

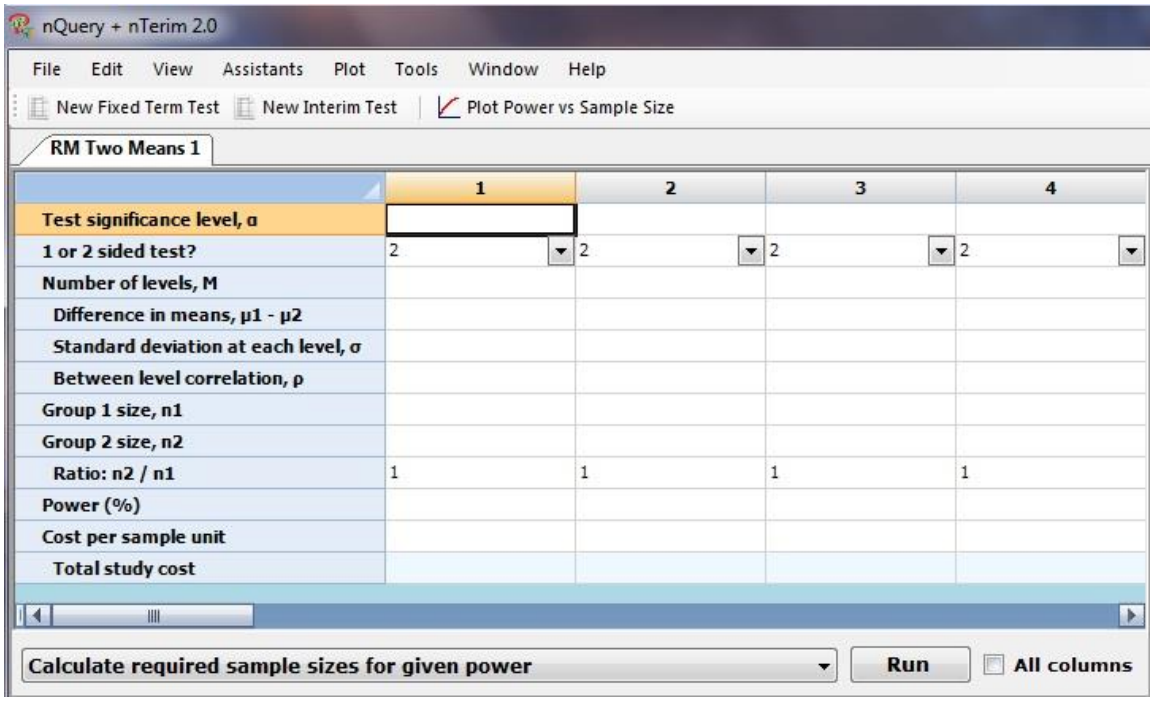


Figure 4.2.14. Repeated Measures for Two Means Test Table

6. The between level correlation is estimated as 0.5, so enter 0.5 in the "Between level correlation" row.

7. We want to calculate the required sample size to obtain a power of 85% so enter 85 in the "Power" row.

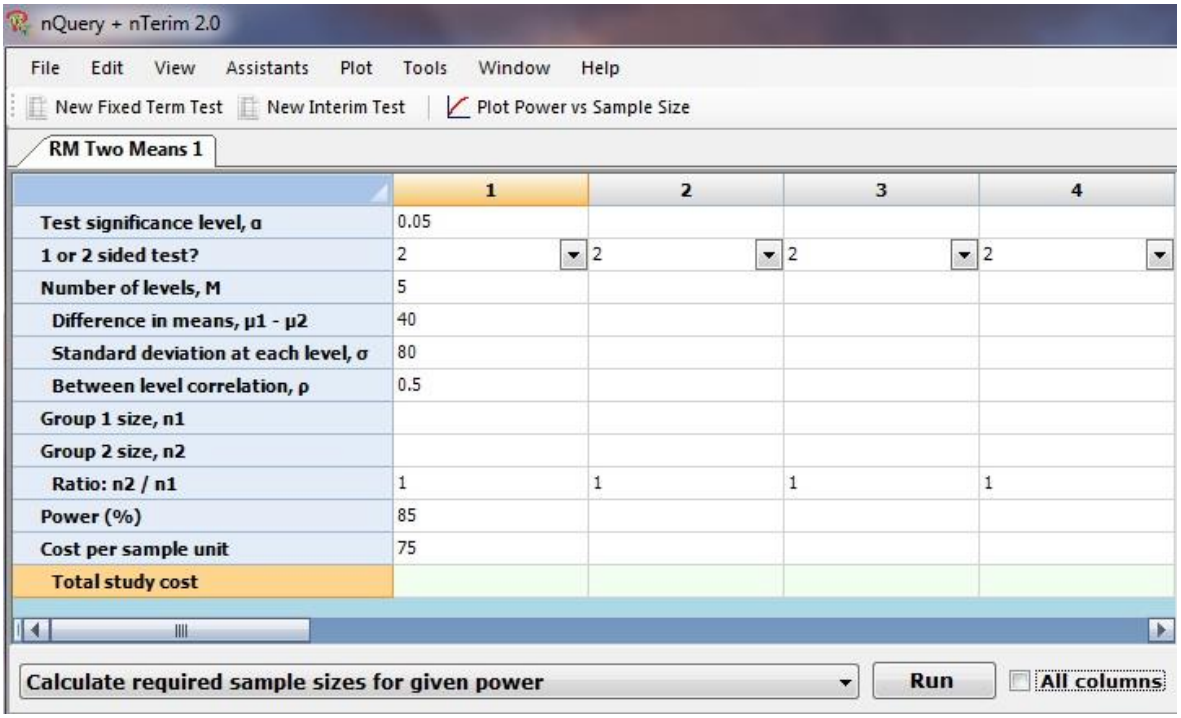


Figure 4.2.15. Design Entry for Two Means Repeated Measures Study

8. The cost per sample unit has been estimated as $75 in this particular study. Therefore to calculate the overall cost associated with the sample size, enter 75 in the “Cost per sample unit” row as shown in Figure 4.2.15.

9. As we want to try several different parameter values for sample size Ratio (R), we can use the “Fill Right” function to fill out multiple columns with the same information entered in Column 1.

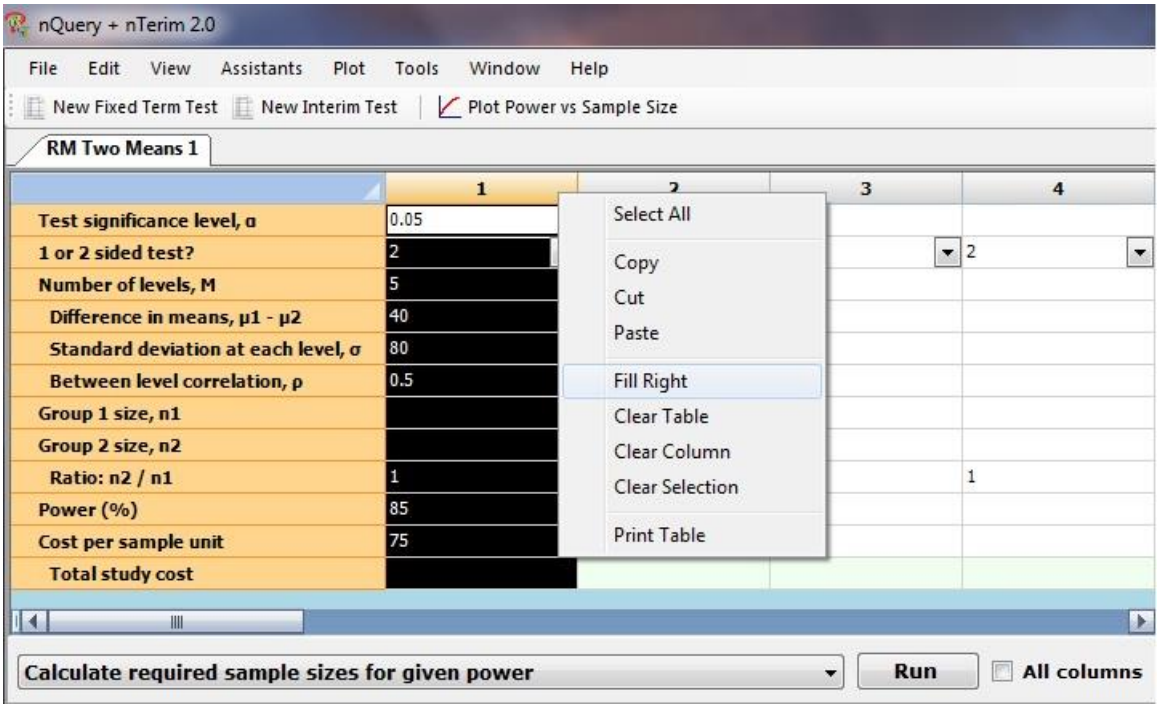


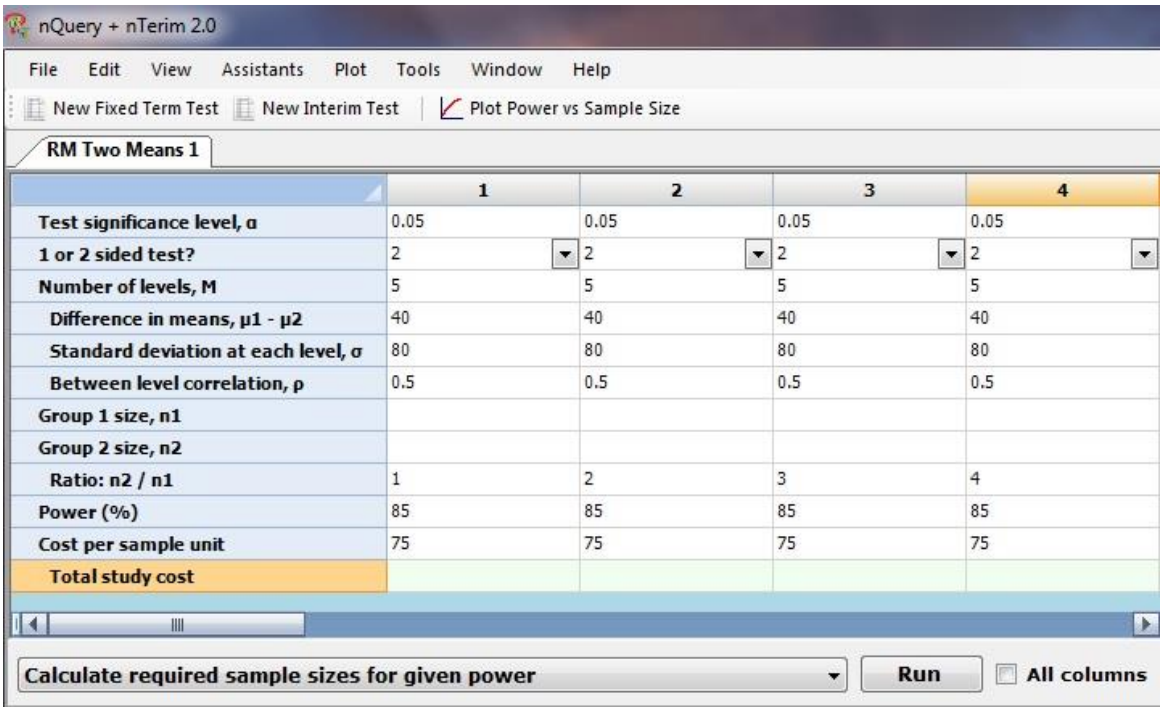
Figure 4.2.16. Fill Right Function Shortcut

10. Once all the parameter information has been entered, right click on the Column 1 heading and select “Fill Right” from the drop-down menu as shown in Figure 4.2.16.

11. As illustrated in Figure 4.2.17, all columns have been filled in with the same parameter information contained in Column 1. We want to alter the other columns (Columns 2 to 4) to see how the sample size is affected by various parameter changes.

12. In this example we want to investigate how the sample size will be affected by a change in the Ratio between the two groups’ sample sizes. To do this we will enter Ratio values of 2, 3, and 4 in Columns 2, 3 and 4 respectively.

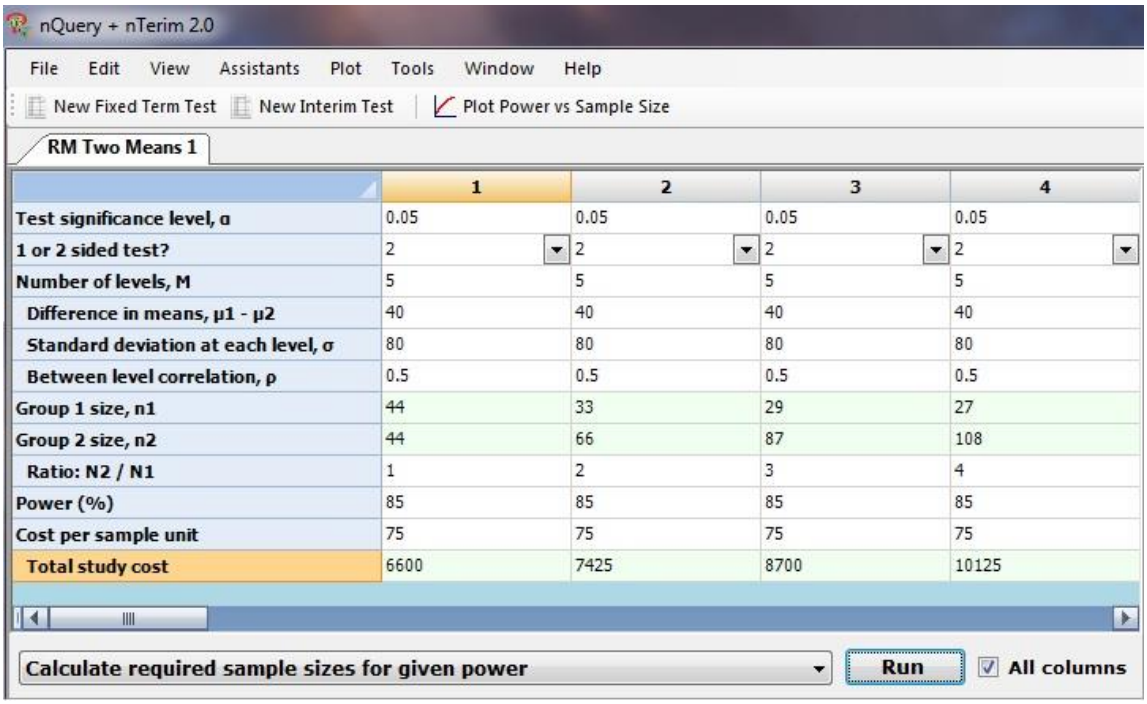
13. As we want to calculated the required sample size to obtain the given power, select "Calculate required sample sizes for given power" from the drop-down menu below the test table.



	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
1 or 2 sided test?	2	2	2	2
Number of levels, M	5	5	5	5
Difference in means, μ1 - μ2	40	40	40	40
Standard deviation at each level, σ	80	80	80	80
Between level correlation, ρ	0.5	0.5	0.5	0.5
Group 1 size, n1				
Group 2 size, n2				
Ratio: n2 / n1	1	2	3	4
Power (%)	85	85	85	85
Cost per sample unit	75	75	75	75
Total study cost				

Figure 4.2.17. Altered columns for comparison

14. As we want to run this calculation for multiple columns, tick the "All Columns" box beside the "Run" button as shown in Figure 4.2.17, then click Run.



	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
1 or 2 sided test?	2	2	2	2
Number of levels, M	5	5	5	5
Difference in means, μ1 - μ2	40	40	40	40
Standard deviation at each level, σ	80	80	80	80
Between level correlation, ρ	0.5	0.5	0.5	0.5
Group 1 size, n1	44	33	29	27
Group 2 size, n2	44	66	87	108
Ratio: N2 / N1	1	2	3	4
Power (%)	85	85	85	85
Cost per sample unit	75	75	75	75
Total study cost	6600	7425	8700	10125

Figure 4.2.18. Completed multiple design Repeated Measures for Two Means Table

15. Another feature that enables us to compare designs side-by-side is by using the Power vs. Sample Size plot. Multiple columns can be plotted together by simply highlighting the desired columns and clicking on the “Plot Power vs Sample Size” button on the menu bar.

16. To highlight the desired columns, click on the column title for Column 1 and drag across to Column 4.

17. Then click on the “Plot Power vs Sample Size” button on the menu bar. The multiple column plot is displayed in Figure 4.2.19.

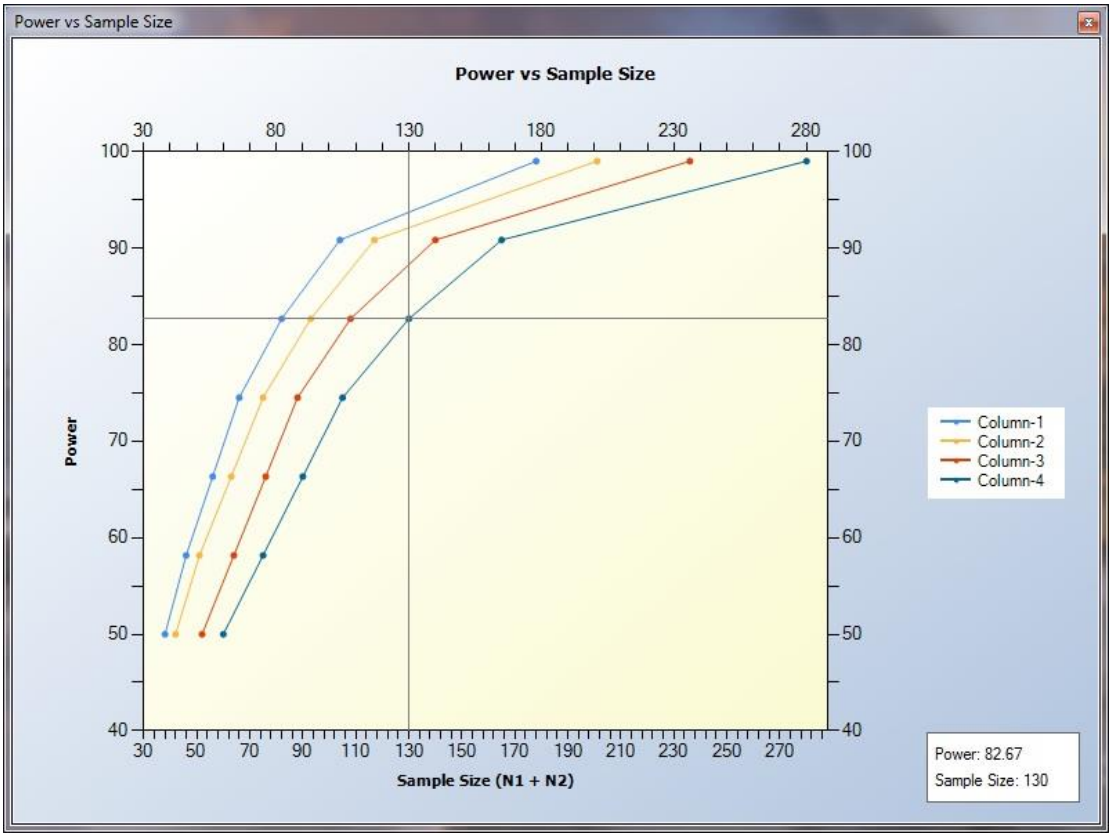


Figure 4.2.19. Power vs. Sample Size Plot

It can be seen from the legend on the left-hand side (legend can be altered manually) that the blue line represents Column 1, the orange line represents Column 2, the red line represents Column 3 and the navy line represents Column 4. The cross on the graph illustrates how the user can identify what the sample size is for a corresponding power value for each column. In the bottom right corner of the plot indicated the exact values for Power and Sample Size for each identifier on the graph.

It can be seen in Figure 4.2.19 that Column 1 reaches an acceptable power level faster than the design in Column 2, 3 or 4. The researcher can now make an assessment as to which design they would prefer to use.

4.3 Repeated Measure for Two Proportions

4.3.1. Introduction

This table facilitates the calculation of power and sample size for the time averaged difference between two proportions in a repeated measures design. Power and sample size is computed using the method outlined by Liu and Wu (2005).

A repeated measures design is an experimental design in which multiple measurements are taken on one or more groups of subjects over time or under different conditions. This type of design leads to a more precise estimate of an endpoint and can avoid the bias from a single measure. For example, an individual's blood pressure is known to be sensitive to many temporary factors such as amount of sleep had the night before, mood, excitement level, exercise, etc. If there is just a single measurement taken from each patient, then comparing the mean blood pressure between two groups could be invalid as there could be a large degree of variation in the single measures of blood pressure levels among patients. However, by obtaining multiple measurements from each individual and comparing the time averaged difference between the two groups, the precision of the experiment is increased.

4.3.2. Methodology

Power and sample size are calculated using standard normal distributions following procedures outlined in Liu and Wu (2005).

To calculate power and sample size the user must first specify the test significance level, α, and choose between a one or a two sided test. The user must then enter a value for the number of levels, M. This value corresponds to the number of measurements that will be taken on each subject. Values must then be provided for the between level correlation, ρ, and any two of group 1 proportions, p_1, group 2 proportions, p_2, and odds ratio, Ψ. Given two of $p_1, p_2, or\ \Psi$ nTerim will compute the other using the following equation:

$$\Psi = \frac{p_2(1-p_1)}{p_1(1-p_2)} \qquad \{4.3.1\}$$

Given the above values, in order to calculate the power for this design the user must enter the expected sample size for each group, N_1 and N_2. nTerim then uses the total sample size, N, to calculate the power of the design using the following equation:

The formula used to calculate power is:

$$Power = 1 - \Phi\left[\left(Z_\alpha\sqrt{\frac{(N_1p_1+N_2p_2)(N_1q_1+N_2q_2)}{N(N_1p_1q_1+N_2p_2q_2)}}\right) - \left(d\sqrt{\frac{MN(1-\pi-\pi^2)}{\big(1+\rho(M-1)\big)(\pi p_1q_1+(1-\pi)p_2q_2)}}\right)\right] \qquad \{4.3.2\}$$

where, $\Phi()$ is the standard normal density function, and

$$N = N_1 + N_2 \qquad \{4.3.3\}$$

$$\pi = \frac{N_1}{N} \qquad \{4.3.4\}$$

$$q_1 = 1 - p_1 \qquad \{4.3.5\}$$

$$q_2 = 1 - p_2 \qquad \{4.3.6\}$$

$$d = p_1 - p_2 \qquad \{4.3.7\}$$

In order to calculate sample size a value for power must be specified. nTerim does not use a closed form equation to calculate sample size. Instead a search algorithm is used. This search algorithm calculates power at various sample sizes until the desired power is reached.

4.3.3. Examples

Example 1: Investigate how Group Proportion affects Sample size for a given Power
In this example we examine how the group proportion affects sample size values for a given power.

The following steps outline the procedure for Example 1.

1. Open nTerim through the Start Menu or by double clicking on the nTerim desktop icon. Then click on "New Fixed Term Test" from the menu bar at the top of the window. A "Study Goal and Design" window will appear.

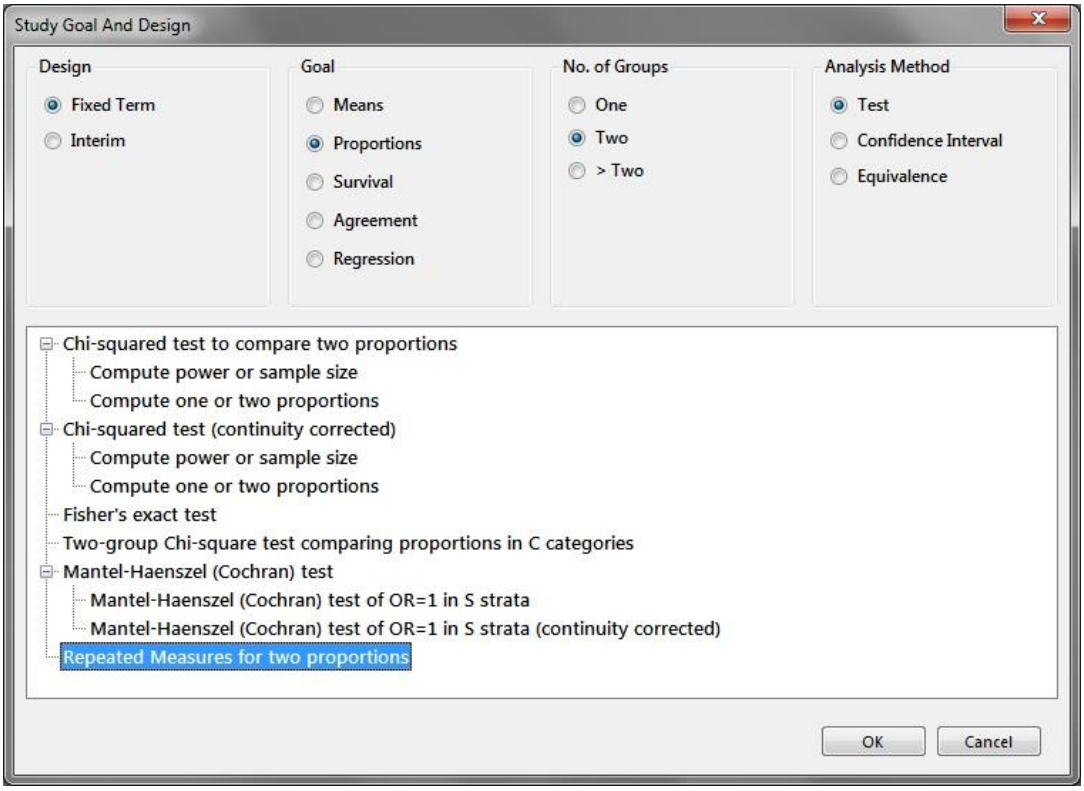


Figure 4.3.1. Study Goal and Design Window

2. Once the correct test has been selected, click "OK" and the test window will appear.

3. Enter 0.05 for alpha, the desired significance level, and enter 3 for the number of levels, M, as shown in Figure 4.3.3.

4. Two sided test is the default setting in nTerim as well as a Ratio value of 1 for the group sizes as shown in Figure 4.3.2.

5. In this example we will examine a study where the group 1 proportion is estimated as 0.45 and the group 2 proportion is estimated as 0.55. Enter 0.45 in the "Group 1 Proportion" row and enter 0.55 in the "Group 2 Proportion" row.

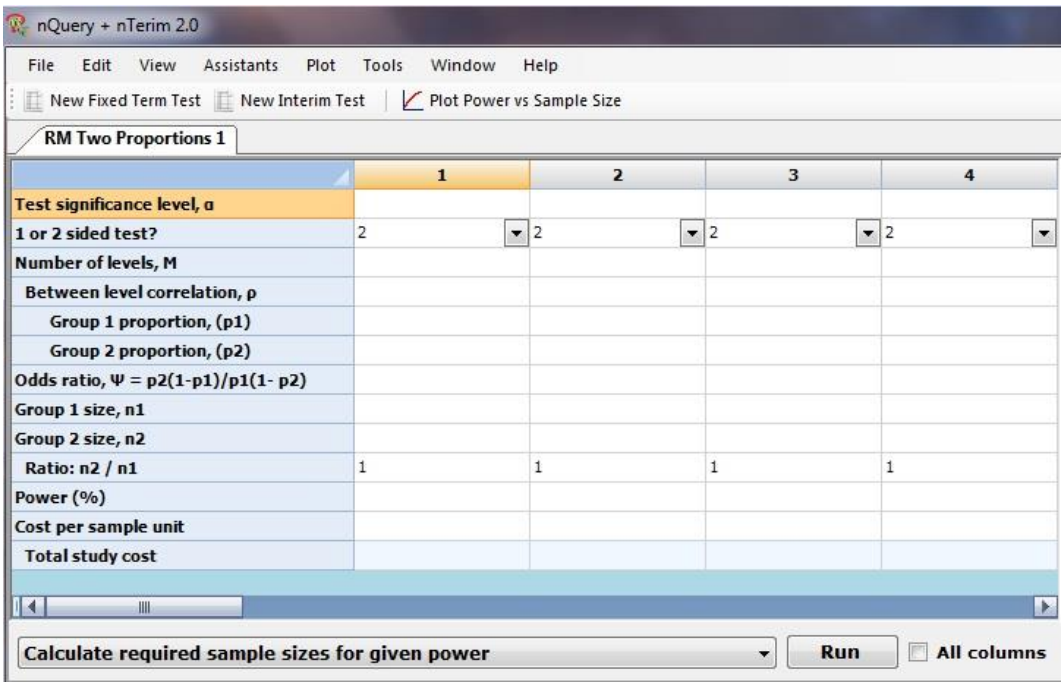


Figure 4.3.2. Repeated Measures for Two Proportions Test Table

6. We also know that the between level correlation is 0.5 so enter 0.5 into the "Between level correlation" row.

7. We want to calculate the required sample size for each group in order to obtain 90% power. To do this, enter 90 in the "Power (%)" row.

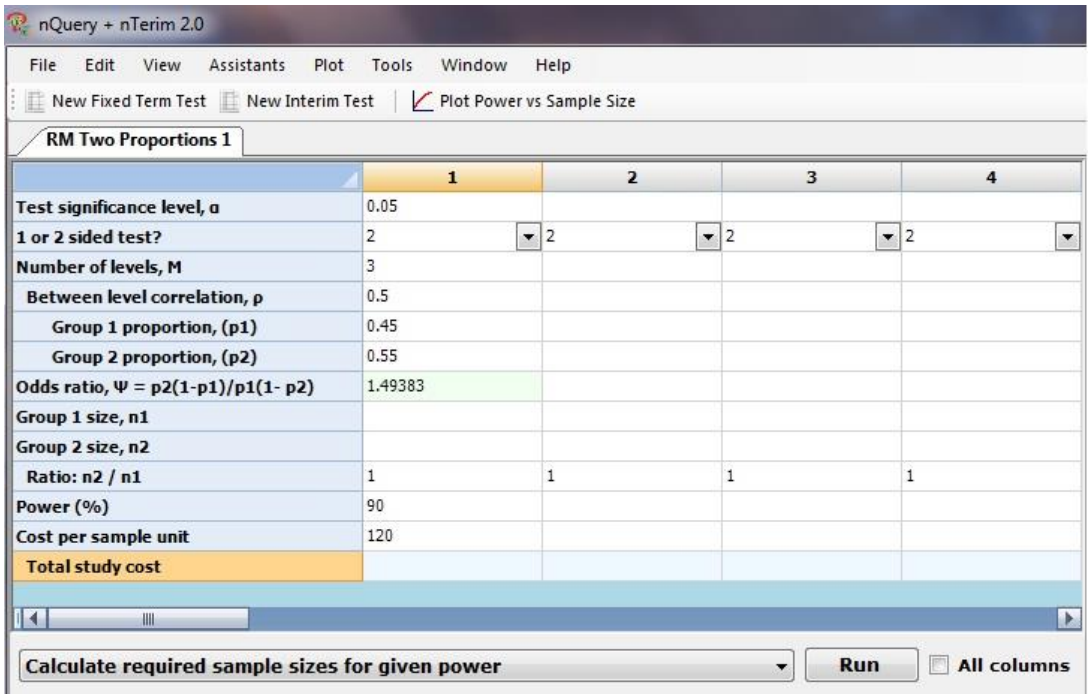
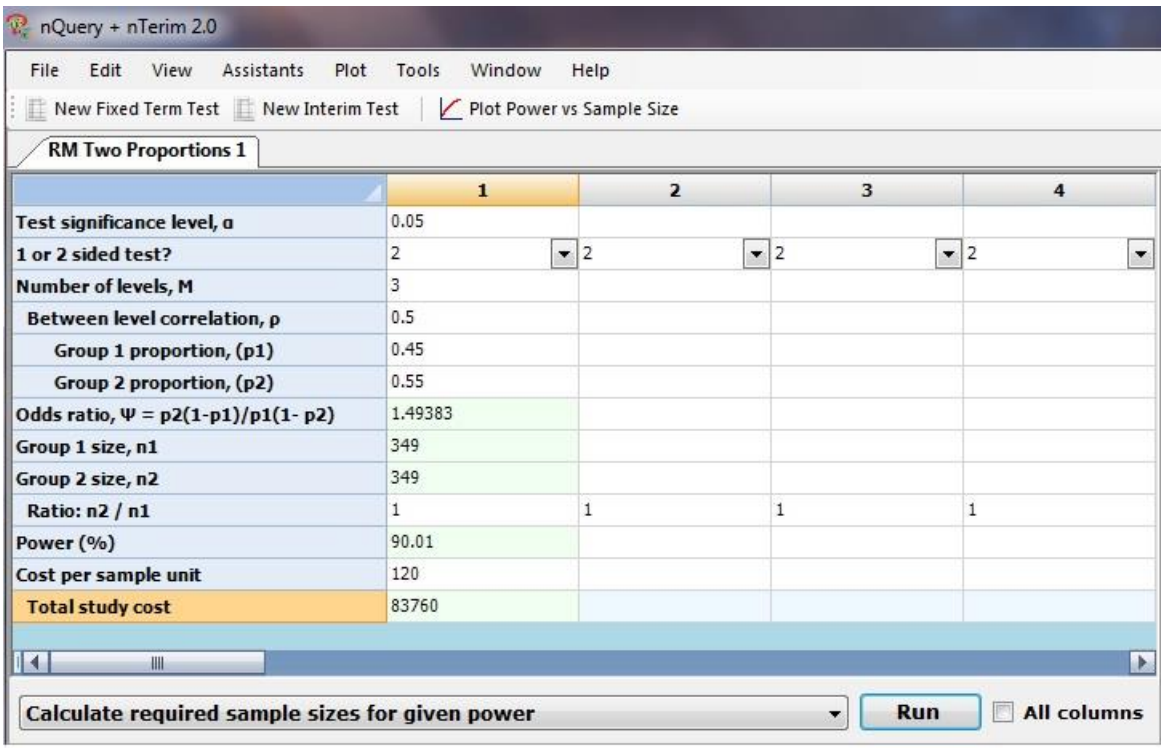


Figure 4.3.3. Design Entry for Two Proportions Repeated Measures Study

8. The cost per sample unit has been estimated as $120 in this particular study. Therefore to calculate the overall cost associated with the sample size, enter 120 in the “Cost per sample unit” row in order to calculate the total study cost associated with the sample size.

9. Then select “Calculate required sample size for given power” from the drop-down menu below the main table and click “Run”. This is displayed in Figure 4.3.4.



	1	2	3	4
Test significance level, α	0.05			
1 or 2 sided test?	2	2	2	2
Number of levels, M	3			
Between level correlation, ρ	0.5			
Group 1 proportion, (p1)	0.45			
Group 2 proportion, (p2)	0.55			
Odds ratio, Ψ = p2(1-p1)/p1(1- p2)	1.49383			
Group 1 size, n1	349			
Group 2 size, n2	349			
Ratio: n2 / n1	1	1	1	1
Power (%)	90.01			
Cost per sample unit	120			
Total study cost	83760			

Figure 4.3.4. Completed Repeated Measures Design for Two Proportions

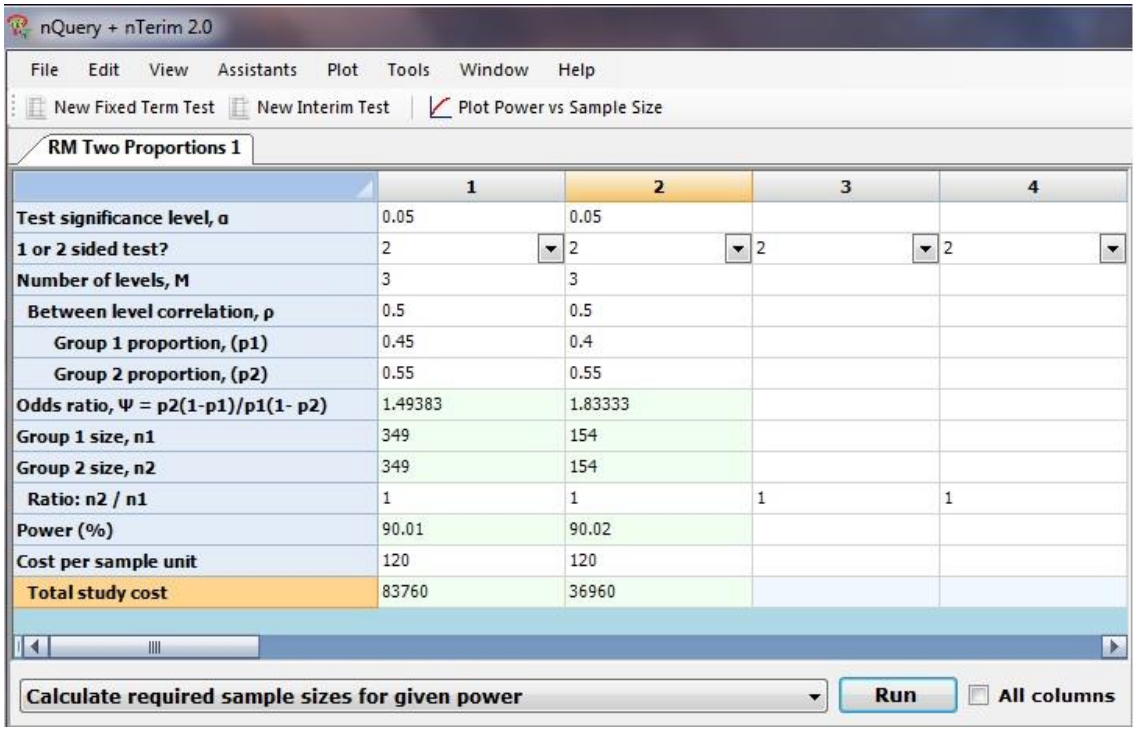
10. Now we are going to repeat this study design example except we’re going to explore how the sample size varies as we alter the proportion in both Group 1 and Group 2. Previously, in Column 1 we had a Group 1 proportion of 0.45 and Group 2 proportion of 0.55. Next we are going to proportions 0.40 and 0.55 for Group 1 and Group 2 respectively.

11. We want to see the effects of changing the group proportion levels has on sample size and perhaps total study cost.

12. In Column 2, enter the same information for level of significance, number of levels, between level correlation, Group 2 proportion, power and cost per sample unit.

13. Now enter 0.4 for Group 1 Proportion in the “Group 1 Proportions” row.

14. Select “Calculate required sample size for given power” from the drop-down menu below the main table and click “Run”. This is displayed in Figure 4.3.5.

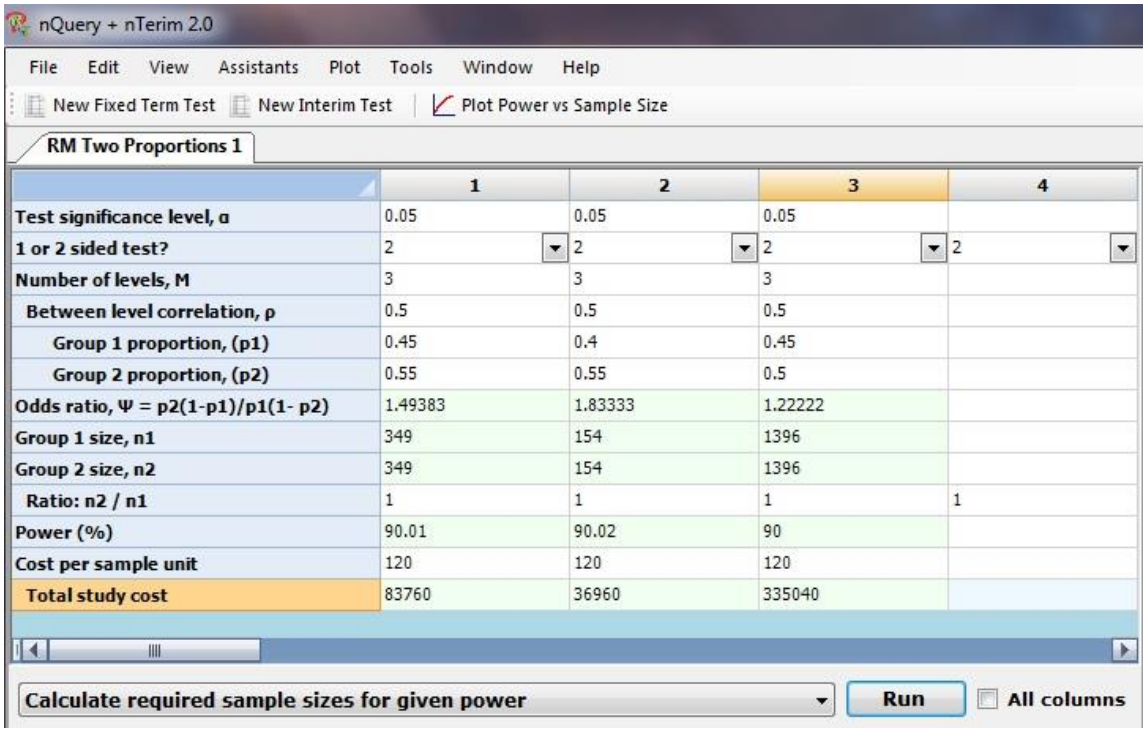


	1	2	3	4
Test significance level, α	0.05	0.05		
1 or 2 sided test?	2	2	2	2
Number of levels, M	3	3		
Between level correlation, ρ	0.5	0.5		
Group 1 proportion, (p1)	0.45	0.4		
Group 2 proportion, (p2)	0.55	0.55		
Odds ratio, Ψ = p2(1-p1)/p1(1- p2)	1.49383	1.83333		
Group 1 size, n1	349	154		
Group 2 size, n2	349	154		
Ratio: n2 / n1	1	1	1	1
Power (%)	90.01	90.02		
Cost per sample unit	120	120		
Total study cost	83760	36960		

Figure 4.3.5. Re-run calculation for Column 2

15. Figure 4.3.5 illustrates the impact of reducing Group 1 proportion. We would also like to see the effect of altering the Group 2 proportion.

16. Similar to step 12, enter the same information from Column 1 into Column 3. This time enter 0.45 for Group 1 proportion and 0.50 for the Group 2 proportion. This is displayed in Figure 4.3.6.



	1	2	3	4
Test significance level, α	0.05	0.05	0.05	
1 or 2 sided test?	2	2	2	2
Number of levels, M	3	3	3	
Between level correlation, ρ	0.5	0.5	0.5	
Group 1 proportion, (p1)	0.45	0.4	0.45	
Group 2 proportion, (p2)	0.55	0.55	0.5	
Odds ratio, Ψ = p2(1-p1)/p1(1- p2)	1.49383	1.83333	1.22222	
Group 1 size, n1	349	154	1396	
Group 2 size, n2	349	154	1396	
Ratio: n2 / n1	1	1	1	1
Power (%)	90.01	90.02	90	
Cost per sample unit	120	120	120	
Total study cost	83760	36960	335040	

Figure 4.3.6. Re-run calculation for Column 3

It can be seen from Figure 4.3.6 that when the Group 1 Proportion was reduced (Column 2), the difference between the two groups increased, the odds ratio in turn increased and the sample size was dramatically reduced. When the Group 2 Proportion was reduced (Column 3), the difference between the two groups reduced and the odds ratio in turn was reduced. The sample size was subsequently increased quite substantially. This all had an knock on effect on the total study cost associate with the sample size.

17. Another feature that enables us to compare designs side-by-side is by using the Power vs. Sample Size plot. Multiple columns can be plotted together by simply highlighting the desired columns and clicking on the "Plot Power vs Sample Size" button on the menu bar.

18. To highlight the desired columns, click on the column title for Column 1 and drag across to Column 3.

19. Then click on the "Plot Power vs. Sample Size" button on the menu bar. The multiple column plot is displayed in Figure 4.3.7 below.

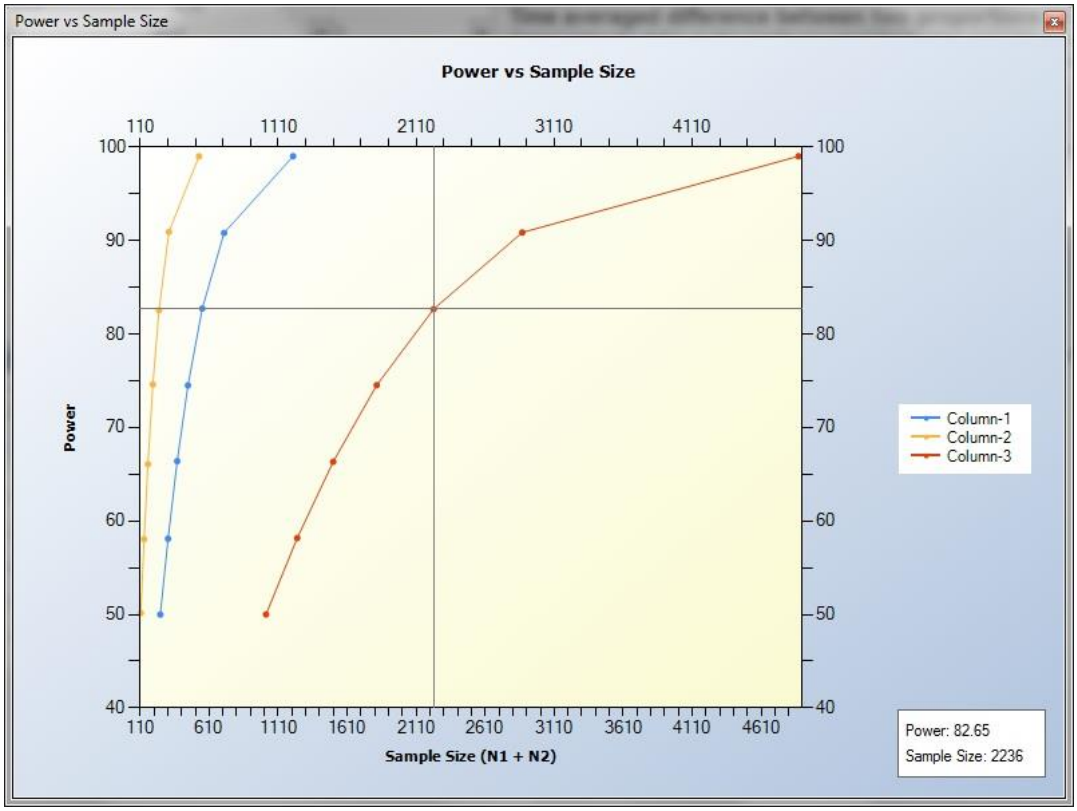


Figure 4.3.7. Power vs. Sample Size Plot

It can be seen from the legend on the left-hand side (legend can be altered manually) that the blue line represents Column 1, the orange line represents Column 2 and the red line represents Column 3. The cross on the graph illustrates how the user can identify what the sample size is for a corresponding power value for each column. In the bottom right corner of the plot indicated the exact values for Power and Sample Size for each identifier on the graph.

Example 2: Specifying and Comparing Multiple Designs

In this example we use the Multiple Factor table to specify multiple designs and then compare the designs appropriately.

The following steps outline the procedure for Example 2.

1. Open nTerim through the Start Menu or by double clicking on the nTerim desktop icon. Then click on "New Fixed Term Test" from the menu bar at the top of the window. A "Study Goal and Design" window will appear.

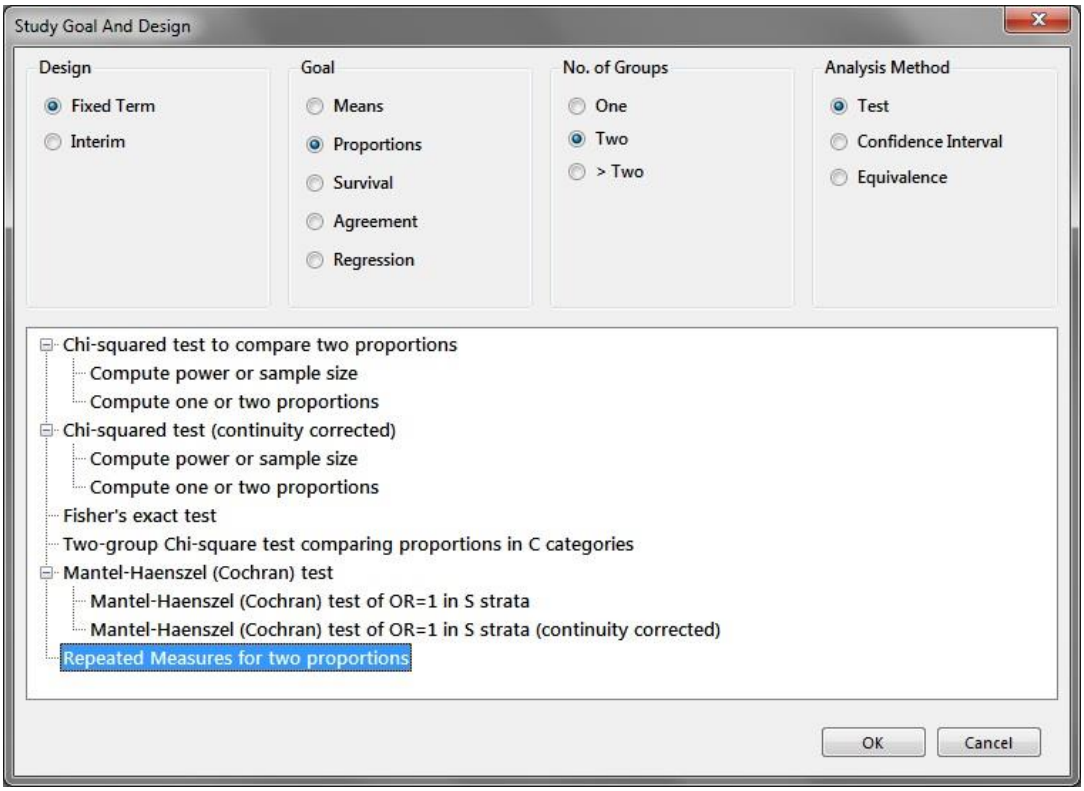


Figure 4.3.8. Study Goal and Design Window

2. Once the correct test has been selected, click "OK" and the test window will appear. This test table is illustrated in Figure 4.3.8.

3. An additional table that will be used in this example is the Specify Multiple Factors table displayed in Figure 4.3.9. This is used to generate multiple columns and designs by entering a range of values for particular parameters.

4. For this example it is known that the proportion of interest in Group 1 ranges from 0.45 to 0.55 and the proportion of interest in Group 2 ranges from 0.39 to 0.51. Therefore, we want to see what the required samples sizes would be at the extremes of these ranges. For example at the maximum proportion for Group 1 and the minimum proportion for Group 2.

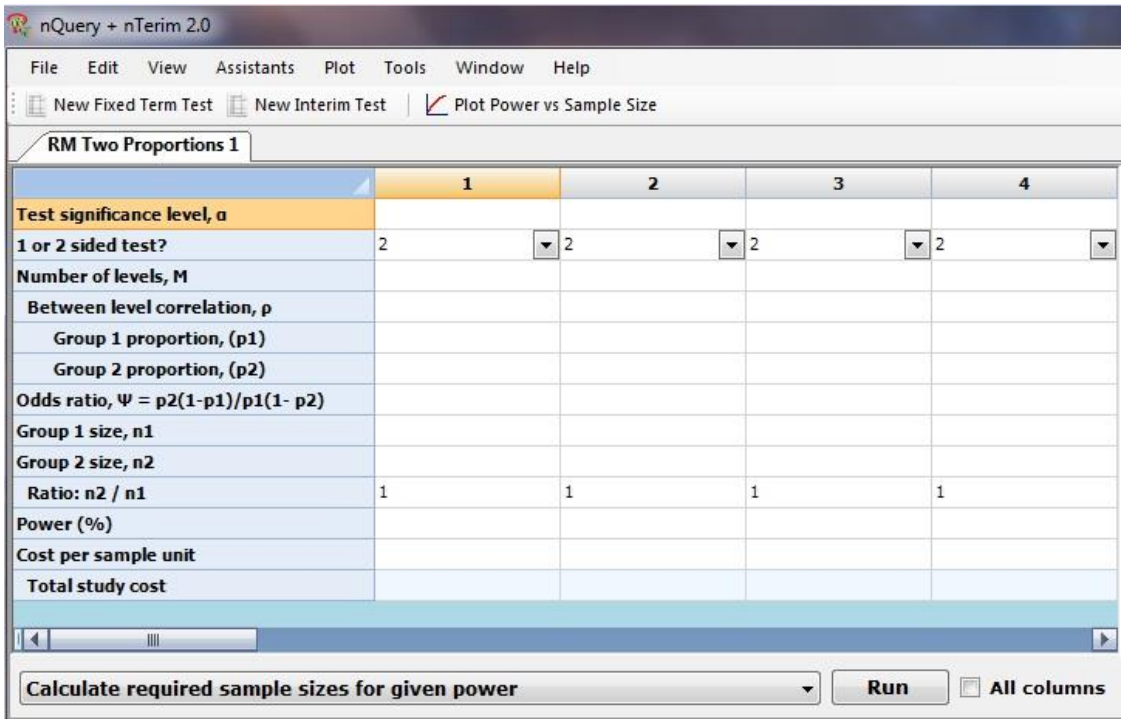


Figure 4.3.8. Repeated Measures for Two Proportions Test Table

5. By incorporating the Specify Multiple Factors table shown in Figure 4.3.9, the user can specify many designs (columns) by entering the desired parameter values and ranges in the provided boxes.

6. We just want to define a two-sided test design. Enter 2 in the "1 or 2 sided test" box. In this study we want 3 levels so enter 3 in the "Number of levels, M" box. We also know that the between level correlation is 0.4 so enter 0.4 in the "Between level correlation" box.

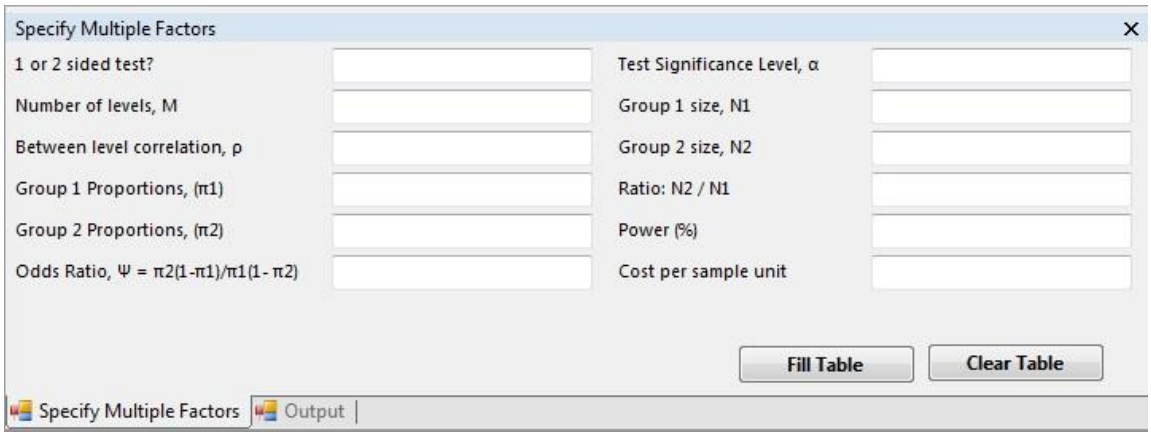


Figure 4.3.9. Specify Multiple Factors Table

7. We know that the Group 1 proportion ranges from 0.45 to 0.55 so enter 0.45 0.55 in the "Group 1 Proportions" box with a space separating the two numbers. We also know that the Group 2 proportion ranges from 0.39 to 0.51 so enter 0.39 0.51 in the "Group 2 Proportions" box. These entries are displayed in Figure 4.3.10 below.

8. We want a 5% level of significance so enter 0.05 in the “Test Significance Level” box. We want an equal sample size for each group so enter 1 in the “Ratio: N2/N1” box. We would like to obtain 90% power in this study design so enter 90 in the “Power (%)” box.

9. Finally, it has been projected that the cost per sample unit will be $100, therefore, enter 100 in the “Cost per sample unit” box.

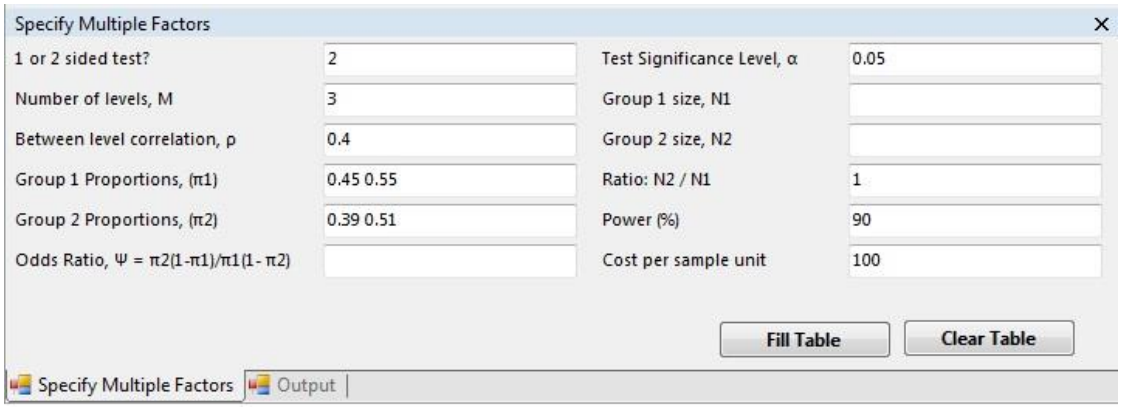
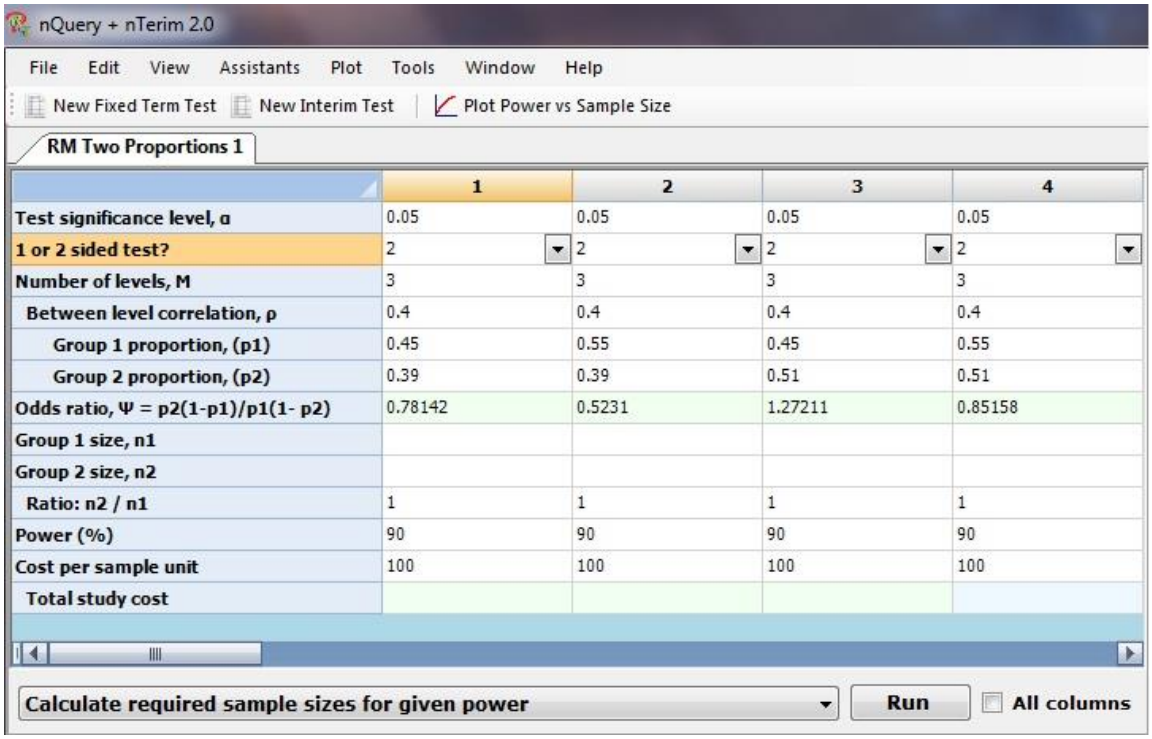


Figure 4.3.10. Completed Specify Multiple Factors Table

10. Once all the parameter values and ranges have been entered correctly, click on “Fill Table” at the bottom right side of the Specify Multiple Factors table.

11. This will automatically fill in the required amount of columns in the test table as illustrated in Figure 4.3.11. In this example we require four columns.

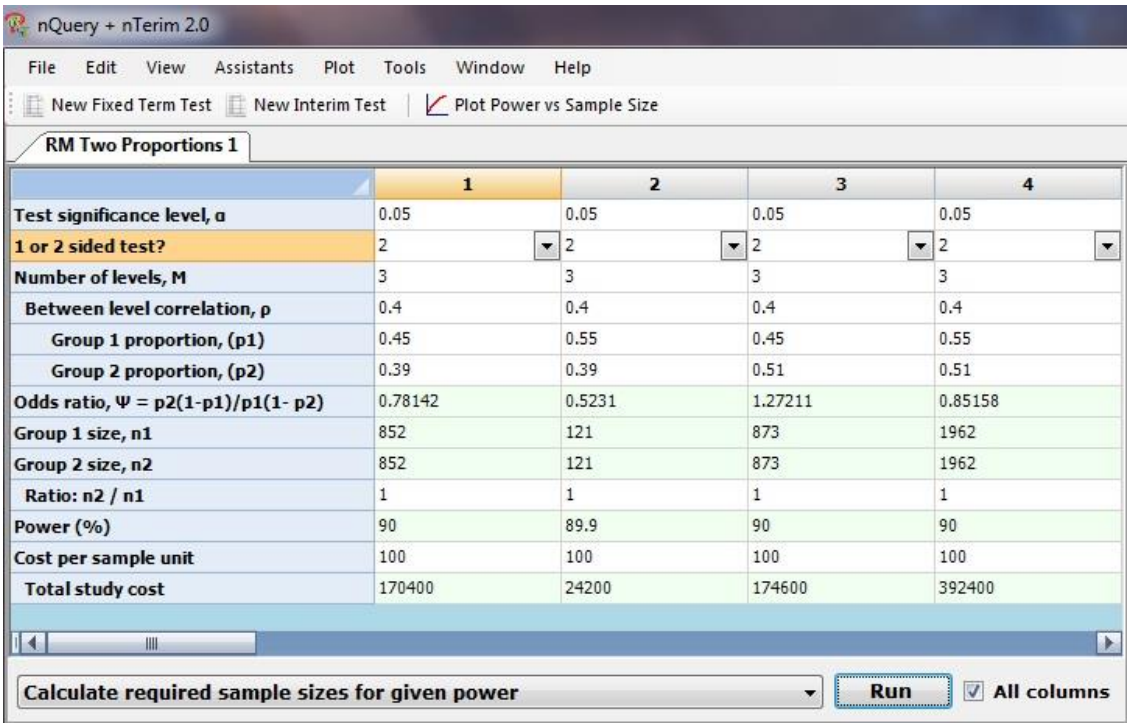


	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
1 or 2 sided test?	2	2	2	2
Number of levels, M	3	3	3	3
Between level correlation, ρ	0.4	0.4	0.4	0.4
Group 1 proportion, (p1)	0.45	0.55	0.45	0.55
Group 2 proportion, (p2)	0.39	0.39	0.51	0.51
Odds ratio, Ψ = p2(1-p1)/p1(1- p2)	0.78142	0.5231	1.27211	0.85158
Group 1 size, n1				
Group 2 size, n2				
Ratio: n2 / n1	1	1	1	1
Power (%)	90	90	90	90
Cost per sample unit	100	100	100	100
Total study cost				

Figure 4.3.11. Design Entry for Multiple columns

12. It can be seen from Figure 4.3.11 that different designs have been created for each combination of the proportions for both groups.

13. In order to calculate appropriate sample size calculations, tick the “All columns” box beside the run button, then select “Calculate required sample sizes for given power” from the drop-down menu below the main table and click “Run”.



	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
1 or 2 sided test?	2	2	2	2
Number of levels, M	3	3	3	3
Between level correlation, ρ	0.4	0.4	0.4	0.4
Group 1 proportion, (p1)	0.45	0.55	0.45	0.55
Group 2 proportion, (p2)	0.39	0.39	0.51	0.51
Odds ratio, Ψ = p2(1-p1)/p1(1- p2)	0.78142	0.5231	1.27211	0.85158
Group 1 size, n1	852	121	873	1962
Group 2 size, n2	852	121	873	1962
Ratio: n2 / n1	1	1	1	1
Power (%)	90	89.9	90	90
Cost per sample unit	100	100	100	100
Total study cost	170400	24200	174600	392400

Figure 4.3.12. Comparison of four Repeated Measures Designs

It can be seen in Figure 4.3.12 that all combinations of the minimum and maximum values for Group 1 and 2 proportions are created. This allows us to evaluate how the sample size varies as the values of the group proportions change.

We can see from Columns 1 and 2 that if we fix the Group 2 proportion at the minimum value of 0.39, and increase the Group 1 proportion, the required sample size decreases. We can also see from Columns 3 and 4 that if we fix the Group 2 proportion at the maximum value of 0.51 and increase the Group 1 proportion, the sample size also increases.

With this approach we are able to quantify how the sample size is affected by changes in both Group 1 and 2 proportions.

14. Another feature that enables us to compare designs side-by-side is by using the Power vs. Sample Size plot. Multiple columns can be plotted together by simply highlighting the desired columns and clicking on the “Plot Power vs Sample Size” button on the menu bar.

15. To highlight the desired columns, click on the column title for Column 1 and drag across to Column 4.

16. Then click on the “Plot Power vs Sample Size” button on the menu bar. The multiple column plot is displayed in Figure 4.3.13.

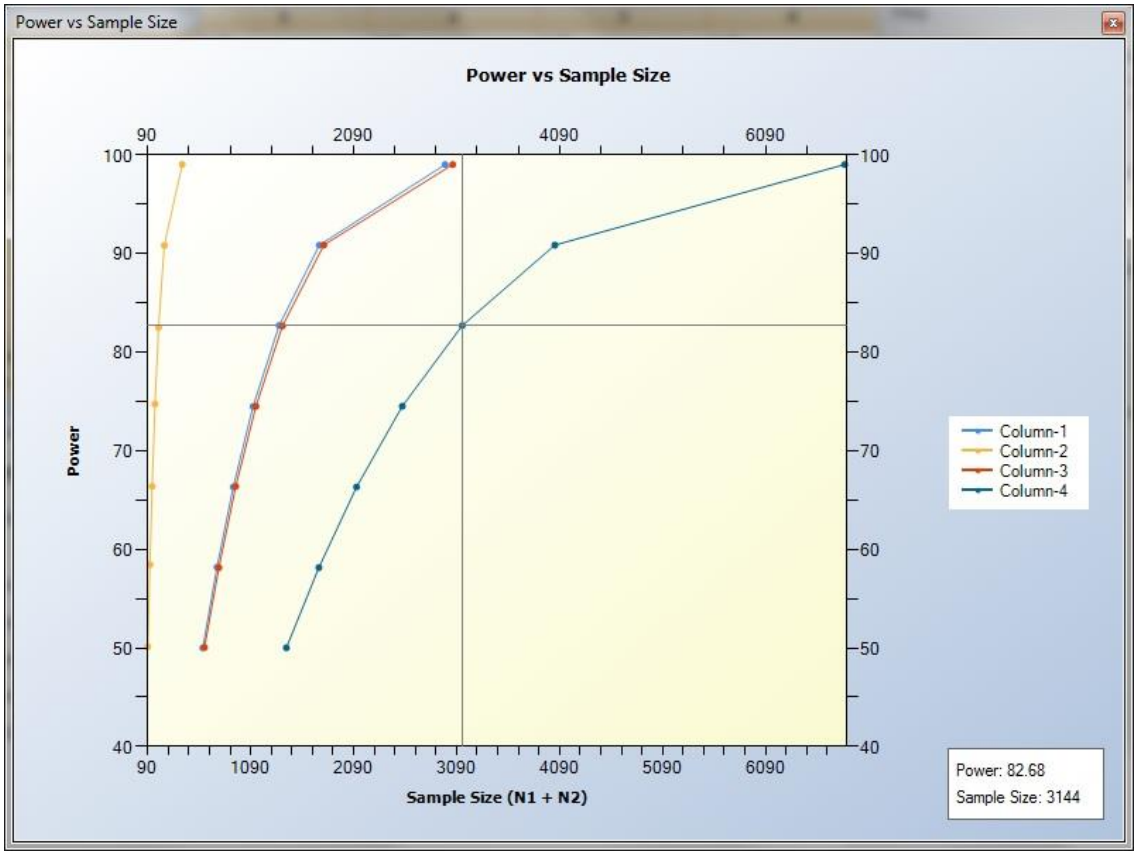


Figure 4.3.13. Power vs. Sample Size Plot

It can be seen from the legend on the left-hand side (legend can be altered manually) that the blue line represents Column 1, the orange line represents Column 2, the red line represents Column 3 and the navy line represents Column 4. The cross on the graph illustrates how the user can identify what the sample size is for a corresponding power value for each column. In the bottom right corner of the plot indicated the exact values for Power and Sample Size for each identifier on the graph.

17. Finally, by clicking on the Output tab at the bottom of the screen you can see a statement giving details of the calculation depending on which column you have clicked on.

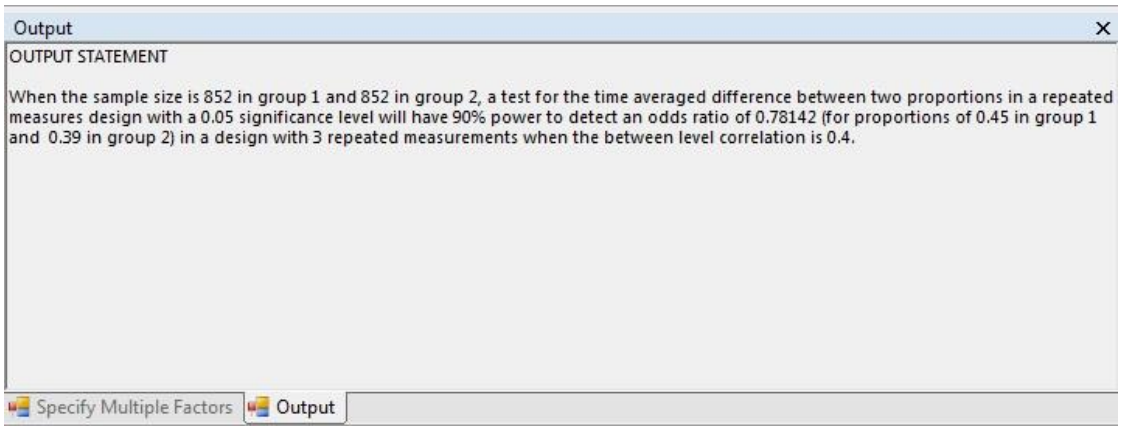


Figure 4.3.14. Output statement

The output statement in Figure 4.3.14 is for Column 1. This statement can be copied and pasted into any report.

4.4 One-Way Analysis of Variance (ANOVA)

4.4.1. Introduction

This table facilitates the calculation of power and sample size for a one-way analysis of variance (ANOVA) design. Calculations are performed using the methods outlined by O'Brien and Muller (1993).

A one-way ANOVA compares means from two or more groups in order to determine whether any of those means are significantly different from each other. Note if we were to compare just two means using the one-way ANOVA then this would be equivalent to a t-test for two independent means. In fact the one-way ANOVA can be viewed as being an extension of a two group t-test.

To give an example of a one-way ANOVA design; consider a study on cholesterol. Suppose we wanted to compare the reduction in cholesterol resulting from the use of a placebo, the current standard drug, and a new drug. The one-way ANOVA tests the null hypothesis that the mean reductions in cholesterol in all three groups are equal. The alternative hypothesis is that the mean reductions in cholesterol in the three groups are not all equal.

4.4.2. Methodology

Power and sample size are calculated using central and non-central F-distributions and follow the procedures outlined by O'Brien and Muller (1993)

To calculate power and sample size the user must specify the test significance level, α, and the number of groups G. The user must then enter a value for the variance of means, V. Alternatively, the user can enter the expected means in each group using the compute effect size assistant. nTerim will then calculate the variance of means using the formula:

$$V = \frac{\sum_{i=1}^{G} r_i(\mu_i - \bar{\mu})^2}{\sum_{i=1}^{G} r_i} \qquad \{4.4.1\}$$

where,

$$\bar{\mu} = \sum_{i=1}^{G} \frac{n_i \mu_i}{N} \qquad \{4.4.2\}$$

The compute effect size assistant also allows the user to enter the expected sample sizes in each group or the expected ratio to group 1 for each group, r_i. This is particularly useful when you expect unequal sample sizes per group.

Once the variance in means is calculated the user must input a value for the common standard deviation, σ. This is a measure of the variability between subjects within a group and is assumed to be the same for all groups. Given the common standard deviation and variance of means nTerim will automatically calculate the effect size using the formula:

$$\Delta^2 = \frac{V}{\sigma^2} \qquad \{4.4.3\}$$

In order to calculate power, a value for the total sample size, N, must be entered (remember this can also be read in from the effect size assistant). nTerim then calculates the power of the design by first determining the critical value $F_{DF_1,\ DF_2,\ \alpha}$. Where, $DF_1 = G - 1$ is the numerator degrees of freedom, and $DF_2 = N - G$ is the denominator degrees of freedom. The non-centrality parameter, λ, is then calculated using the equation:

$$\lambda = N\Delta^2 \qquad \{4.4.4\}$$

Using these two values, nTerim calculates the power of this design as the probability of being greater than $F_{DF_1,\ DF_2,\ \alpha}$ on a non-central F-distribution with non-centrality parameter λ.

In order to calculate sample size nTerim does not use a closed form equation. Instead a search algorithm is used. This search algorithm calculates power at various sample sizes until the desired power is reached.

4.4.3. Examples

Example 1: One-way ANOVA with unequal n's in a Blood Pressure Study
In this example we will compare the reduction in blood pressure resulting from the use of three potential treatments: (i) Placebo, (ii) current Standard Drug and (iii) New Drug. According to similar previous studies on the Standard Drug, we have approximated the reduction in blood pressure as roughly 12mmHg with a standard deviation of 6mmHg. Likewise, in previous studies, the Placebo has resulted in an estimated reduction of 5mmHg. This example will examine using a One-way Analysis of Variance with a 0.05 level of significance.

The following steps outline the procedure for Example 1.

1. Open nTerim through the Start Menu or by double clicking on the nTerim desktop icon. Then click on "New Fixed Term Test" from the menu bar at the top of the window. A "Study Goal and Design" window will appear.

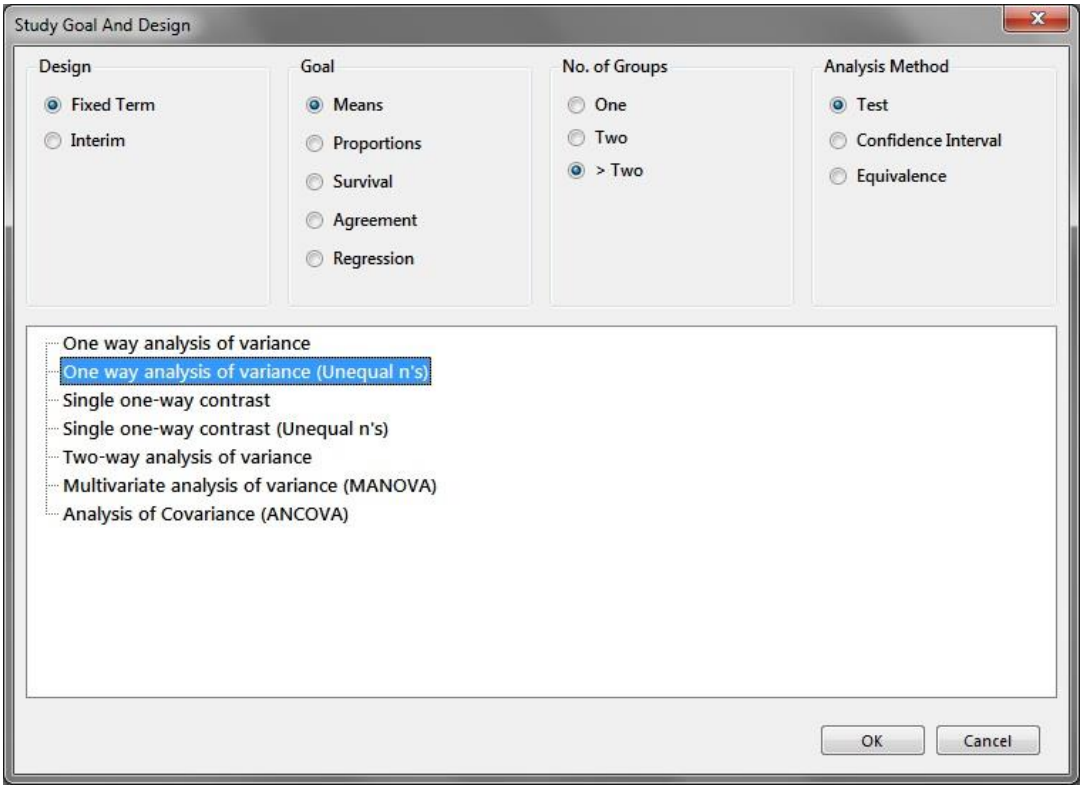


Figure 4.4.1. Study Goal and Design Window

2. Once the correct test has been selected, click "OK" and the test window will appear.

3. There are two main tables required for this test, the main test table illustrated in Figure 4.4.2 and the effect size assistant table shown in Figure 4.4.3.

4. Enter 0.05 for alpha, the desired significance level, and enter 3 for the number of groups, G, as shown in Figure 4.4.4.

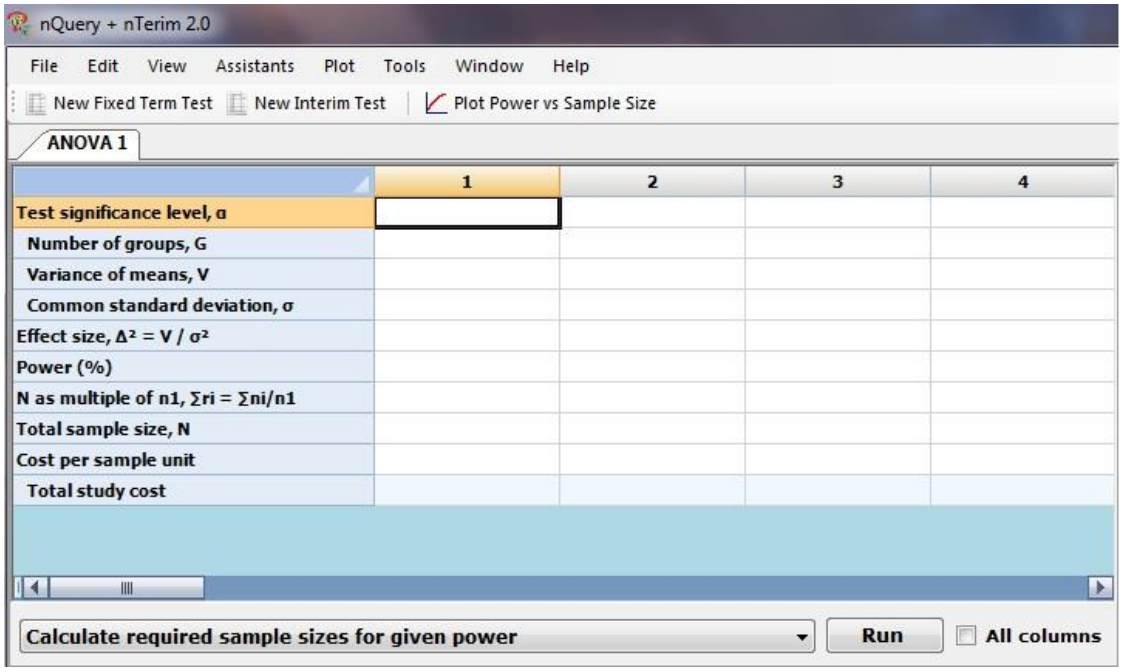


Figure 4.4.2. One-way Analysis of Variance Test Table

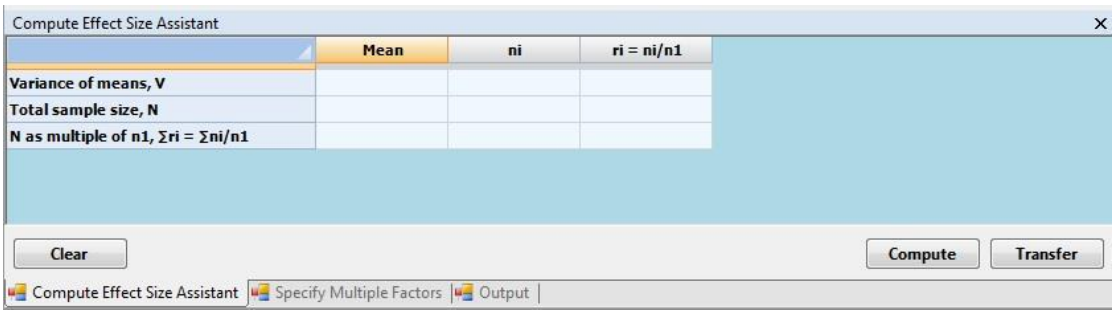


Figure 4.4.3. Compute Effect Size Assistant Window

5. Once you enter a value for the number of groups, G, the "Compute Effect Size Assistant" table automatically updates as shown in Figure 4.4.4.

6. In order to calculate a value for Effect Size, the Variance of Means (V) needs to be calculated first.

7. The mean for each level and the corresponding sample size need to be entered in the "Compute effect Size Assistant" table.

8. For the "Mean" values for each group, enter 5 for group 1, 12 for group 2 and 12 for group 3.

9. For the group sample size (n_i) values for each group, enter 20 for group 1, 12 for group 2 and 18 for group 3. As a result, the ratio (r_i) is calculated for each group as a proportion of group 1.

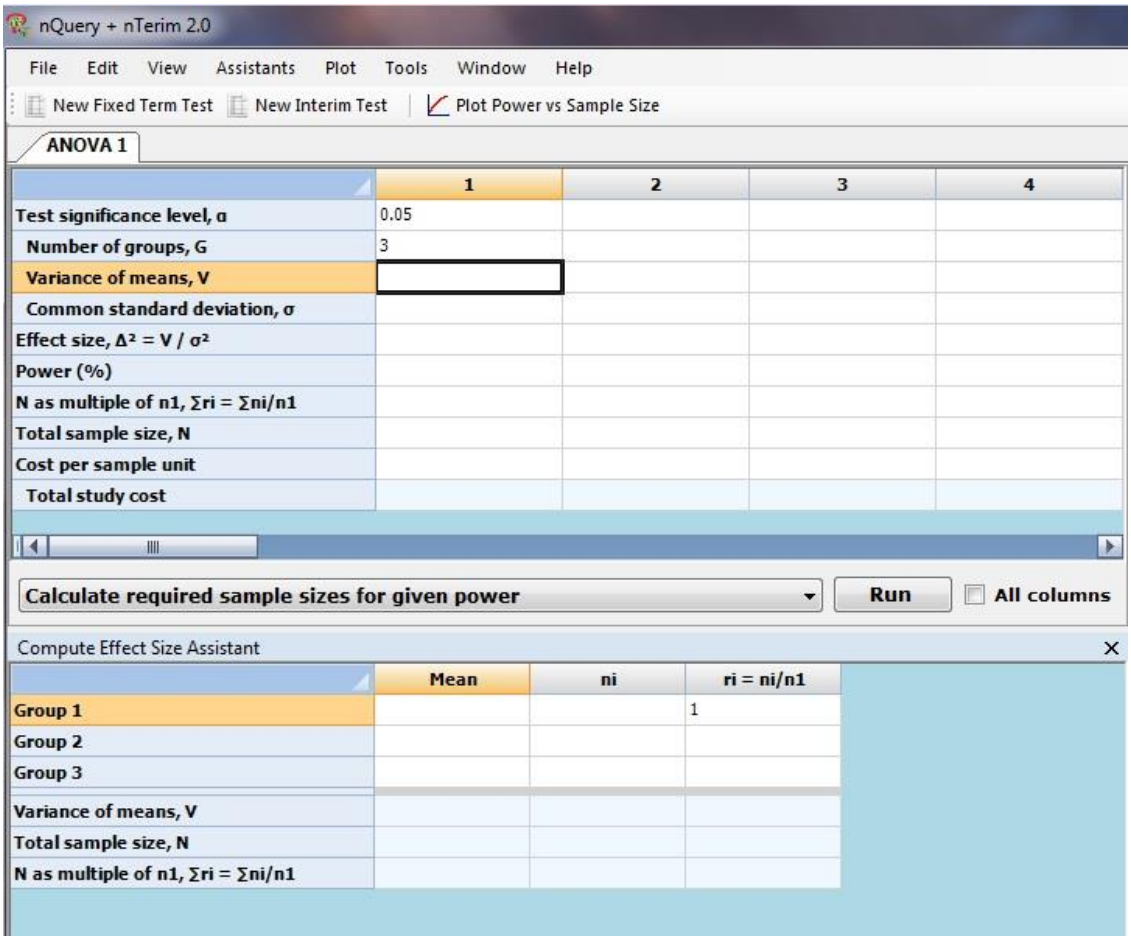


Figure 4.4.4. Automatically updated Compute effect size Assistant Table

10. Once the table in Figure 4.4.5 is completed, and values for Variance of Means (V) and total Sample Size (N) are computed, click on "Transfer" to automatically transfer these values to the main table.

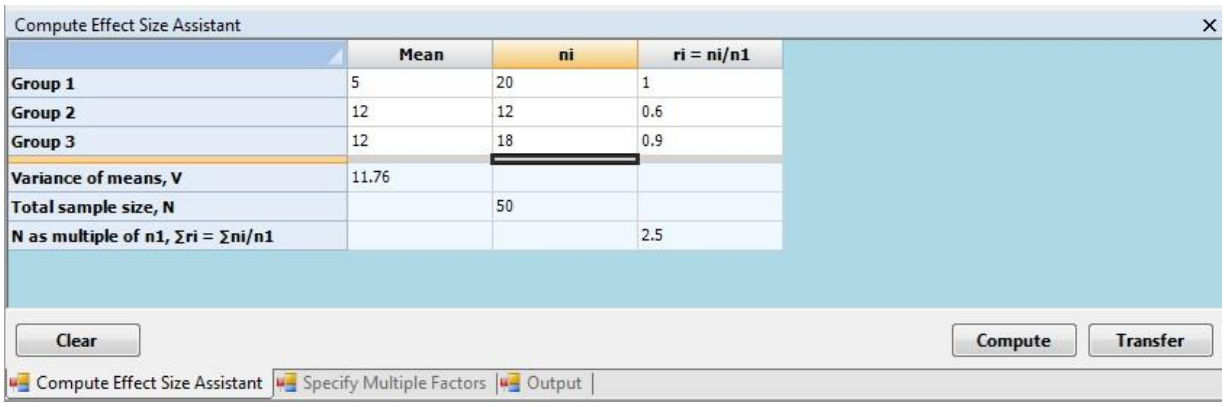


Figure 4.4.5. Completed Compute Effect Size Assistants Table

11. Now that values for Variance of Means (V) and total Sample Size (N) are computed we can continue with filling in the main table. For the Common Standard Deviation, enter a value of 6. Now the Effect Size is automatically calculated.

12. We want to calculate the attainable power given the sample size of 50.

13. It has been estimated that it will cost $85 per sample unit in this study. Therefore enter 85 in the "Cost per sample unit" row.

14. Select "Calculate attainable power with the given sample size" from the drop-down menu below the main table and click "Run". This is displayed in Figure 4.4.6.

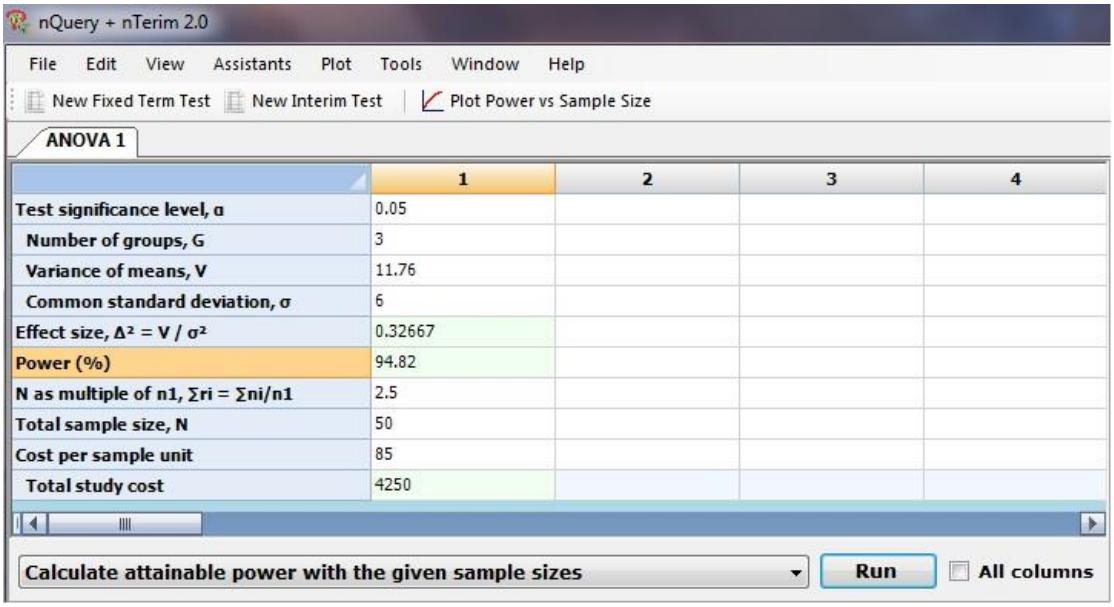


Figure 4.4.6. Completed One-Way Analysis of Variance Test Table

It can be seen from Figure 4.4.6 that a sample size of 50 is required to obtain a power of 94.82%. Due to the cost per sample unit of $85, the overall cost of sample size required has amounted to $4,250.

By clicking on the Output tab at the bottom of the screen you can see a statement giving details of the calculation:

"When the total sample size across the 3 groups is 50, distributed across the groups as specified, a one-way analysis of variance will have 94.82% power to detect at the 0.05 level a difference in means characterized by a Variance of means, $V = \sum r_i \cdot (\mu_i - \mu_m)^2 / (\sum r_i)$ of 11.76, assuming that the common standard deviation is 6."

In this example we can also perform sensitivity analysis to see how volatile this study is to slight changes in a particular parameter. For example, let us examine how the attainable power alters under slight changes in Standard Deviation.

1. Firstly, we much copy the information in Column 1 to Column 2. To do this, highlight Column 1 by clicking on the column title as shown in Figure 4.4.7. Then right click and select "Copy".

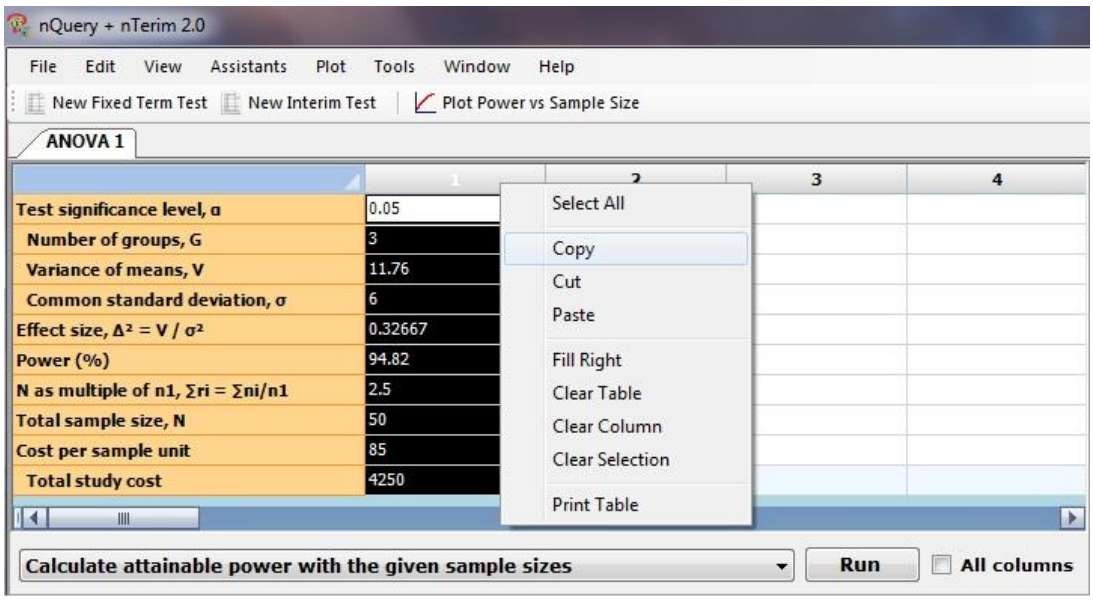


Figure 4.4.7. Copy Column 1

2. Then right click on the first cell in Column 2 and select "Paste" as illustrated in Figure 4.4.8 below.

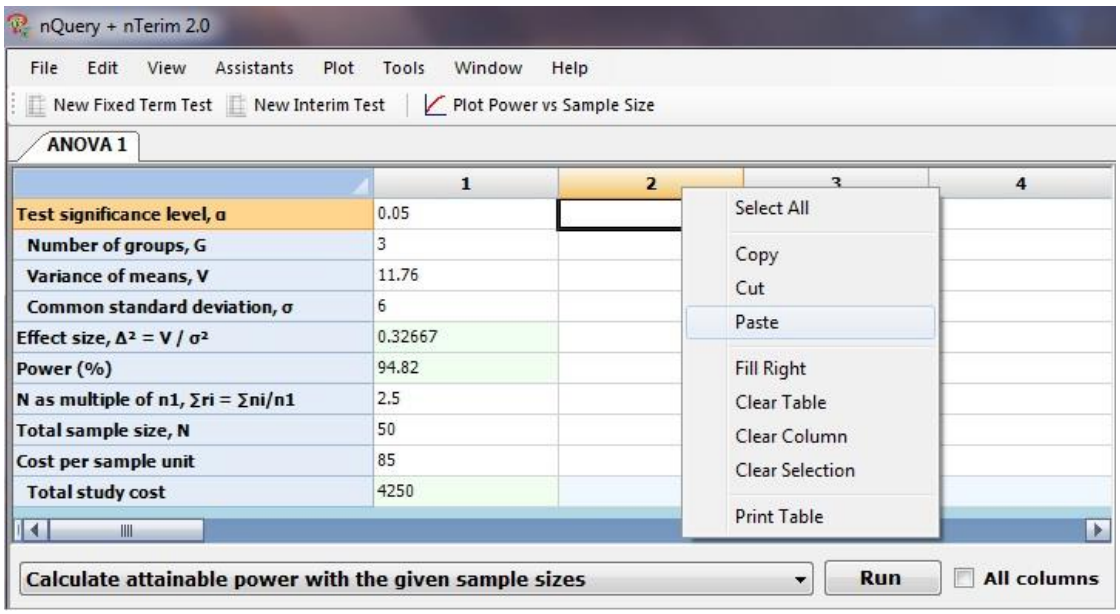
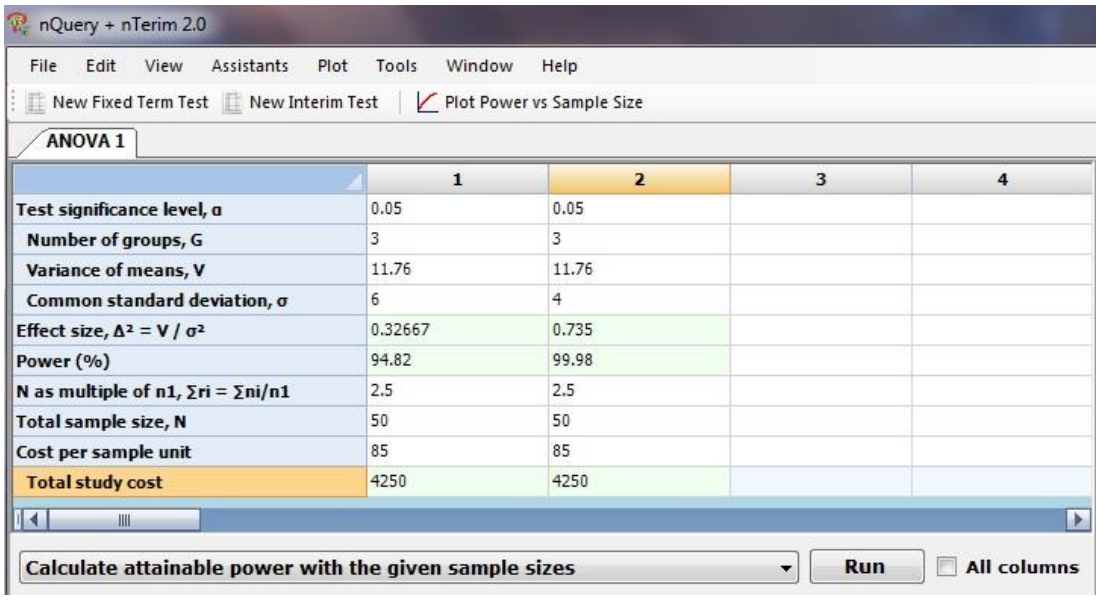


Figure 4.4.8. Paste contents of Column 1 into Column 2

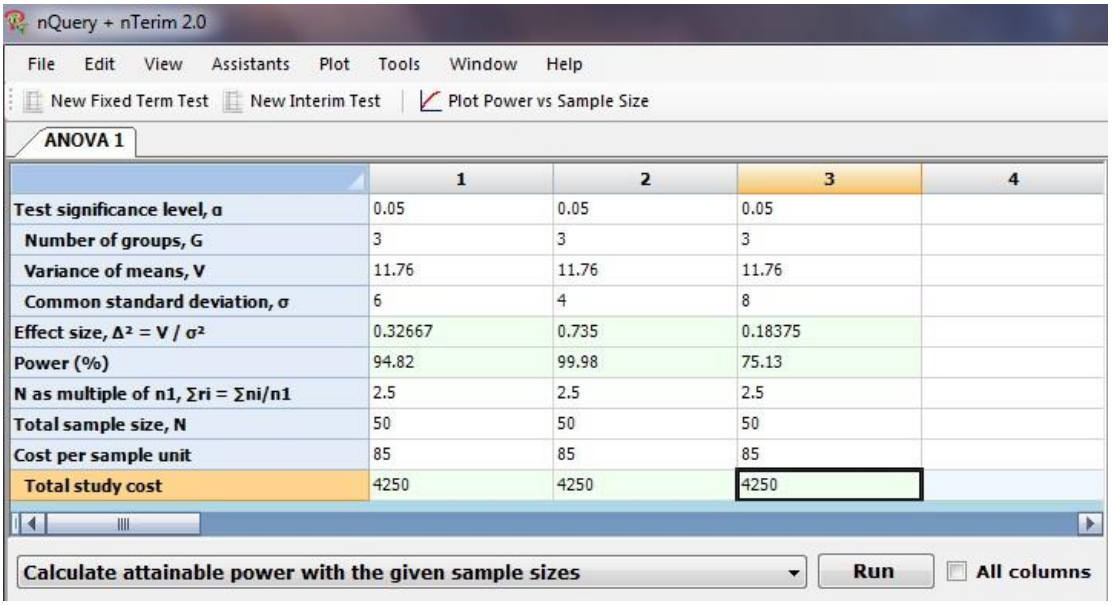
3. Once the contents of Column 1 have been copied over to Column 2, you can change the value of the Common Standard Deviation to 4 and click “Run”. This will update Column 2 to its new attainable value for power, as seen in Figure 4.4.9.



	1	2	3	4
Test significance level, α	0.05	0.05		
Number of groups, G	3	3		
Variance of means, V	11.76	11.76		
Common standard deviation, σ	6	4		
Effect size, $\Delta^2 = V / \sigma^2$	0.32667	0.735		
Power (%)	94.82	99.98		
N as multiple of n1, Σri = Σni/n1	2.5	2.5		
Total sample size, N	50	50		
Cost per sample unit	85	85		
Total study cost	4250	4250		

Figure 4.4.9. Re-run calculations to update Column 2

4. Repeat Steps 2 & 3 except paste the contents of Column 1 into Column 3, change the Common Standard Deviation to 8 and click “Run”. This is displayed in Figure 4.4.10.



	1	2	3	4
Test significance level, α	0.05	0.05	0.05	
Number of groups, G	3	3	3	
Variance of means, V	11.76	11.76	11.76	
Common standard deviation, σ	6	4	8	
Effect size, $\Delta^2 = V / \sigma^2$	0.32667	0.735	0.18375	
Power (%)	94.82	99.98	75.13	
N as multiple of n1, Σri = Σni/n1	2.5	2.5	2.5	
Total sample size, N	50	50	50	
Cost per sample unit	85	85	85	
Total study cost	4250	4250	4250	

Figure 4.4.10. Re-run calculations for Column 3

5. Now it can be seen from Figure 4.4.10 that there is a change in Effect Size and ultimately Power due to both increasing and decreasing the Common Standard Deviation. It’s easy to compare the implications of a slight increase or decrease in the Common Standard Deviation.

6. Another feature that enables us to compare designs side-by-side is by using the Power vs. Sample Size plot. Multiple columns can be plotted together by simply highlighting the desired columns and clicking on the “Plot Power vs Sample Size” button on the menu bar.

7. To highlight the desired columns, click on the column title for Column 1 and drag across to Column 3. Then click on the “Plot Power vs Sample Size” button on the menu bar. The multiple column plot is displayed in Figure 4.4.11.

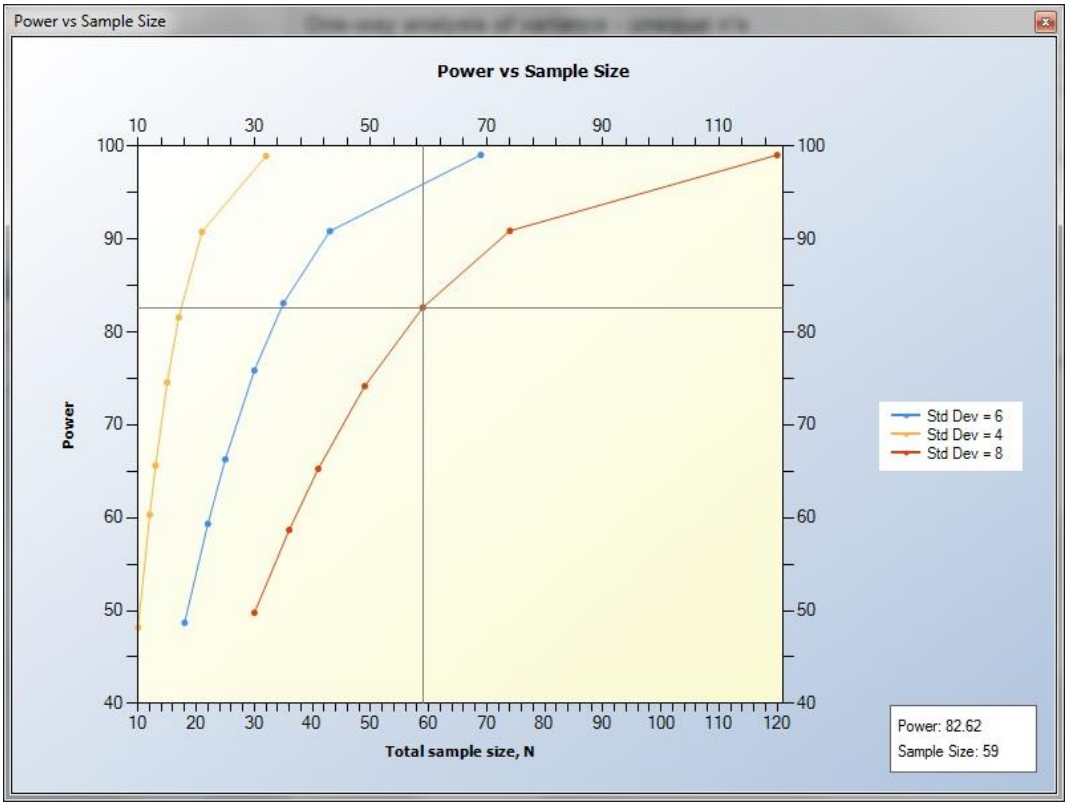


Figure 4.4.11. Multiple Column Power vs. Sample Size Plot

It can be seen from the legend on the left-hand side (legend can be altered manually) that the blue line represents Column 1, the orange line represents Column 2 and the red line represents Column 3. The cross on the graph illustrates how the user can identify what the sample size is for a corresponding power value for each column. In the bottom right corner of the plot indicated the exact values for Power and Sample Size for each identifier on the graph.

It can be seen in Figure 4.4.11 that Column 2 reaches an acceptable power level much faster than the other two designs as it has the lowest value for Common Standard Deviation. This plot also shows us how volatile this study design is to any change in Common Standard Deviation.

4.5 Analysis of Covariance (ANCOVA)

4.5.1. Introduction

This table facilitates the calculation of power and sample size for analysis of covariance (ANCOVA) designs. Calculations are performed using the approximations outlined by Muller and Barton (1989), and Muller, LaVange, Ramey and Ramey (1992).

An analysis of covariance (ANCOVA) design can be viewed as an extension of the one-way analysis of variance (ANOVA). In ANOVA, differences in means between two or more groups are tested on a single response variable. An ANCOVA, on the other hand, does the same analysis while adjusting for covariates. These covariates provide a way of statistically controlling the effect of variables one does not want to examine in a study. It is assumed that the inclusion of these covariates will increase the statistical power of a design. However, it must be noted that adding a covariate also reduces the degrees of freedom. Therefore, adding a covariate that accounts for very little variance in the response variable may actually reduce power.

To give an example of an ANCOVA design; consider a study where we are examining test scores among students. In this study it is found that boys and girls test scores for a particular subject differ. However, it is known that girls take more classes in the subject than boys. We can use ANCOVA to adjust the test scores based on the relationship between the number of classes taken and the test score. Thus, enabling us to determine whether boys and girls have different test scores while, adjusting for the number of classes taken.

4.5.2. Methodology

Power and sample size are calculated using central and non-central F-distributions and follow the procedures outlined by Keppel (1991). To calculate power and sample size the user must specify the test significance level, α, and the number of groups G. The user must then enter a value for the variance of means, V. Alternatively, the user can enter the expected means in each group using the compute effect size assistant. nTerim will then calculate the variance of means using the formula:

$$V = \frac{\sum_{i=1}^{G} r_i(\mu_i - \bar{\mu})^2}{\sum_{i=1}^{G} r_i} \qquad \{4.5.1\}$$

where,

$$\bar{\mu} = \sum_{i=1}^{G} \frac{n_i \mu_i}{N} \qquad \{4.5.2\}$$

The compute effect size assistant also allows the user to enter the expected sample sizes in each group or the expected ratio to group 1 for each group, r_i. This is particularly useful when you expect unequal sample sizes per group.

Once the variance in means is calculated the user must input a value for the common standard deviation, σ. This is a measure of the variability between subjects within a group and is assumed to be the same for all groups. The user must then also enter the number covariates, c, to be used in the study along with the average r-squared value between the response and the covariates, R^2.

In order to calculate power, a value for the total sample size, N, must be entered (remember this can also be read in from the effect size assistant). nTerim then calculates the power of the design by first determining the critical value $F_{G-1,\ N-G-c,\ \alpha}$.

The non-centrality parameter, λ, is then calculated using the equation:

$$\lambda = \bar{n} G \frac{V}{\sigma_\epsilon^2} \qquad \{4.5.3\}$$

where,

$$\bar{n} = \frac{N}{G} \qquad \{4.5.4\}$$

and,

$$\sigma_\epsilon^2 = (1 - \rho^2)\sigma^2 \qquad \{4.5.5\}$$

where σ_{ϵ}^{2} is the within-group variance after considering the covariates and ρ^{2} is the coefficient of multiple determination (estimated by R^{2}).

Using these two values, nTerim calculates the power of this design as the probability of being greater than $F_{G-1,\ N-G-c,\ \alpha}$ on a non-central F-distribution with non-centrality parameter λ.

In order to calculate sample size nTerim does not use a closed form equation. Instead a search algorithm is used. This search algorithm calculates power at various sample sizes until the desired power is reached.

4.5.3. Examples

Example 1: Calculating Attainable Power given Sample Size
In this example we are going to calculate the attainable power for a given sample size for an ANCOVA design.

The following steps outline the procedure for Example 1.

1. Open nTerim through the Start Menu or by double clicking on the nTerim desktop icon. Then click on "New Fixed Term Test" from the menu bar at the top of the window. A "Study Goal and Design" window will appear.

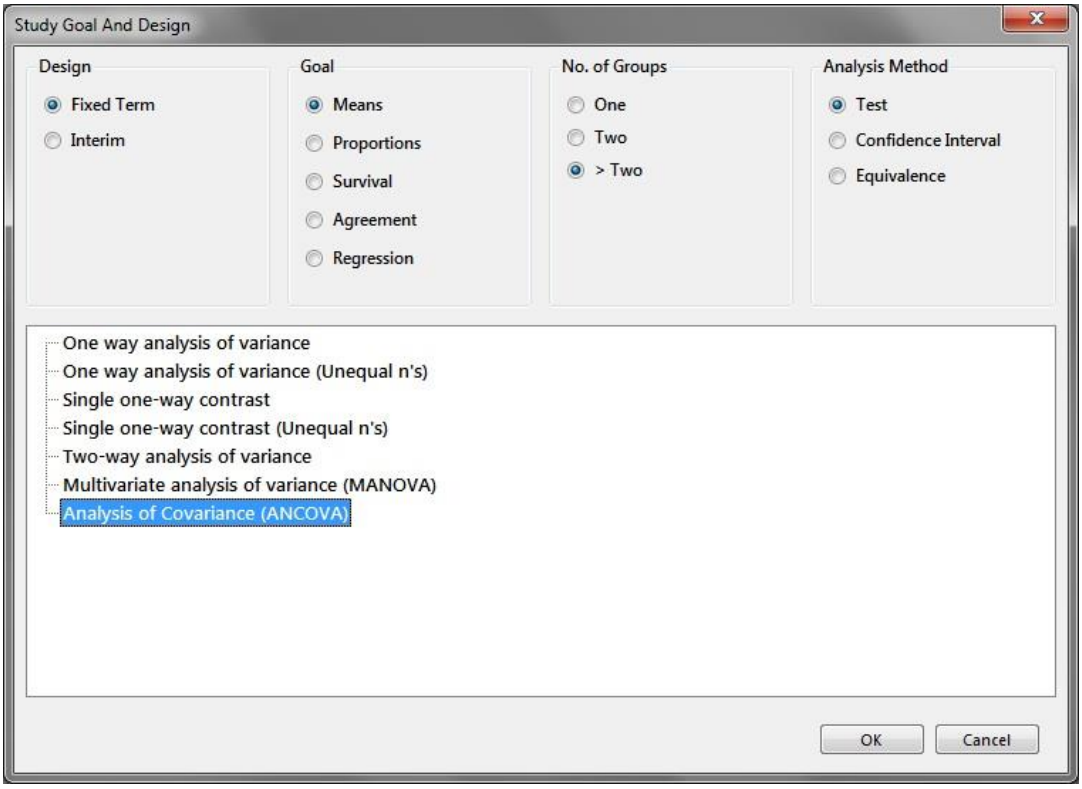


Figure 4.5.1. Study Goal and Design Window

2. Once the correct test has been selected, click "OK" and the test window will appear.

3. There are two main tables required for this test, the main test table illustrated in Figure 4.5.2 and the effect size assistant table shown in Figure 4.5.3.

4. Enter 0.05 for alpha, the desired significance level, and enter 4 for the number of groups, G, as shown in Figure 4.5.4.

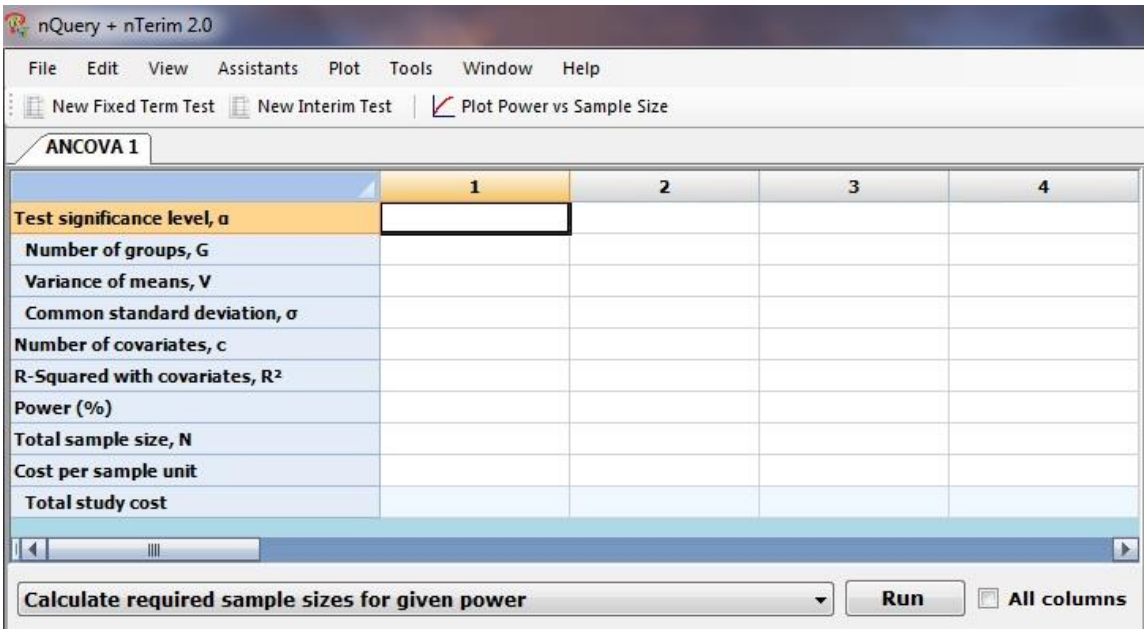


Figure 4.5.2. Analysis of Covariance Test Table

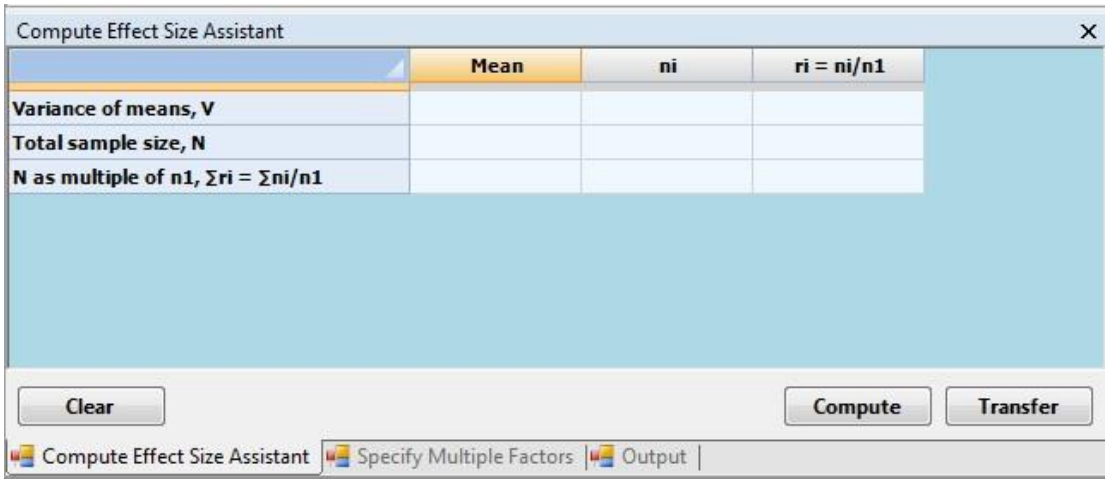


Figure 4.5.3. Compute Effect size Assistant Window

5. Once you enter a value for the number of groups, G, the "Compute Effect Size Assistant" table automatically updates as shown in Figure 4.5.4.

6. In order to calculate a value for Effect Size, the Variance of Means (V) needs to be calculated first.

7. The mean for each level and the corresponding sample size need to be entered in the "Compute Effect Size Assistant" table.

8. For the "Mean" values for each group, enter 15 for group 1, 20 for group 2, 25 for group 3 and 18 for group 4.

9. For the group sample size (n_i) values for each group, enter 30 for group 1, 45 for group 2, 45 for group 3 and 30 for group 4. As a result, the ratio (r_i) is calculated for each group as a proportion of group 1.

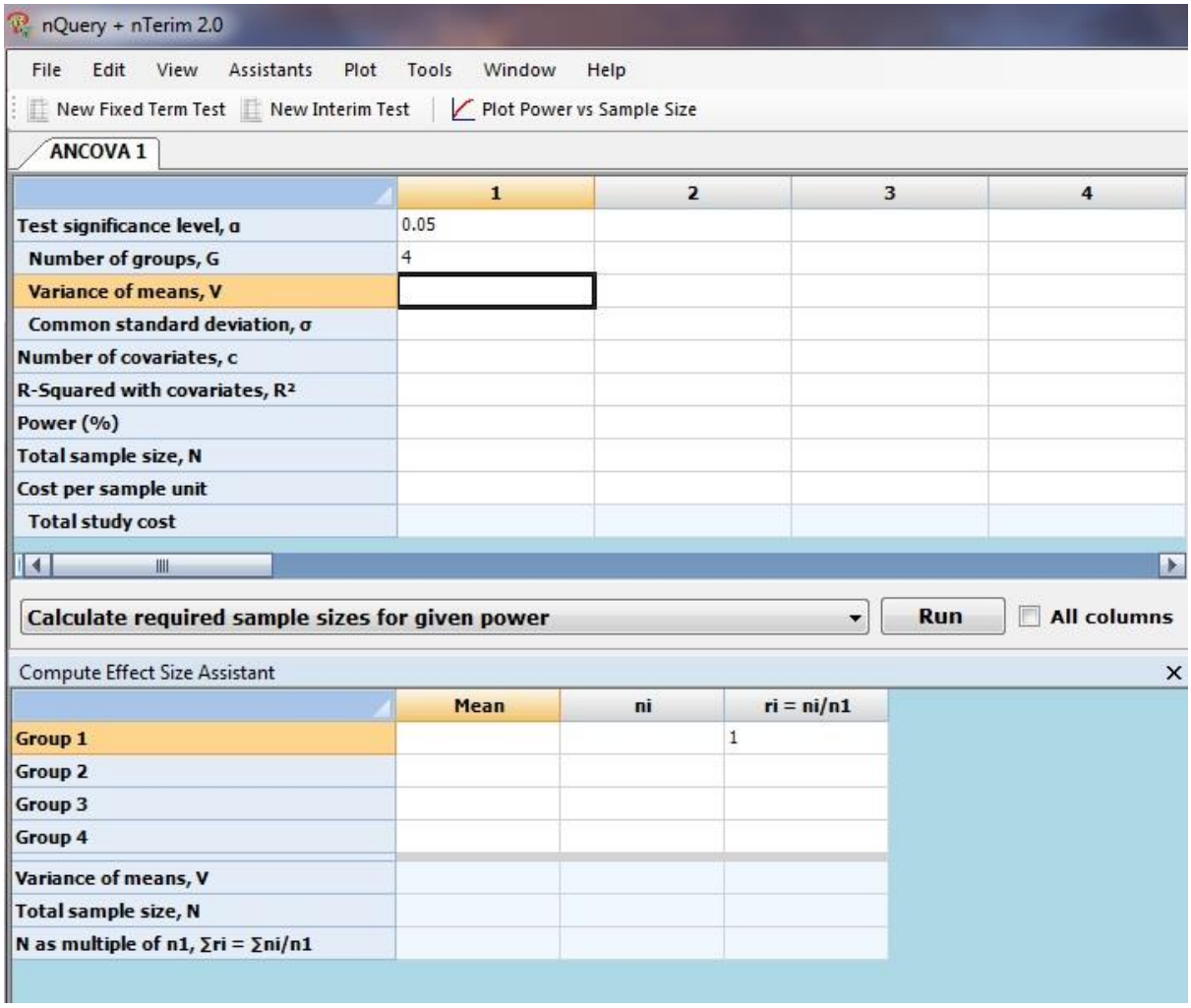


Figure 4.5.4. Automatically updated Compute effect size Assistant Window

10. Once the table illustrated in Figure 4.5.5 is completed, and the values for Variance of Means (V) and Total sample size (N) are computed, click on "Transfer" to automatically transfer these values to the main table.

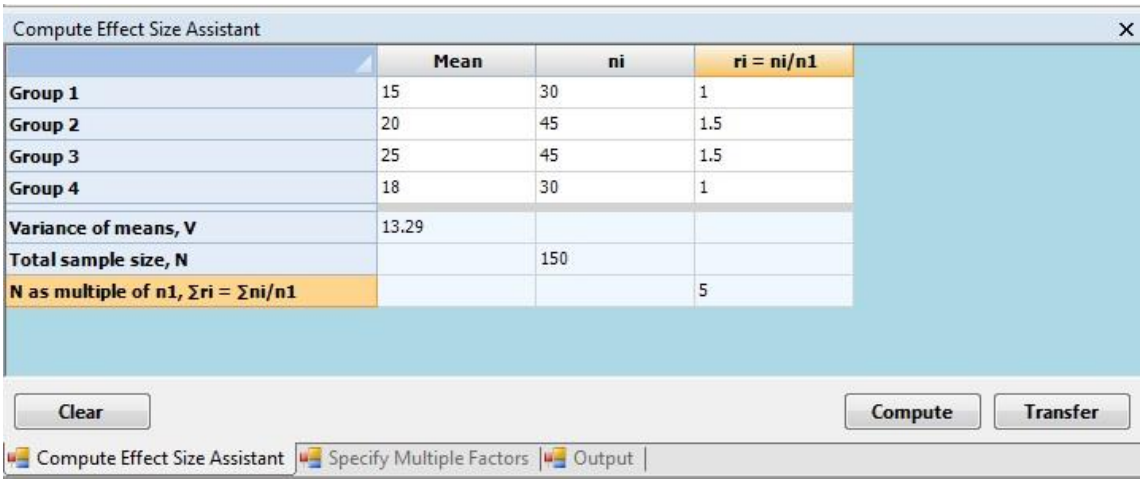


Figure 4.5.5. Completed Compute Effect size Assistant Window

11. Now that values for Variance of Means (V) and Total sample size (N) are computed we can continue with filling in the main table. For the Common Standard Deviation, enter a value of 25.

12. The number of covariates to be used in this study is set at 1, so enter the value 1 in the "Number of covariates" row. Also the R-Squared value has been estimated as 0.75 for this study design so enter 0.75 in the "R-Squared with covariates" row.

13. We want to calculate the attainable power give the sample size of 150.

14. It has been estimated that it will cost $100 per sample unit in this study. Therefore enter 100 in the "Cost per sample unit" row.

15. Select "Calculate attainable power with the given sample size" from the drop-down menu below the main table and click "Run". This is displayed in Figure 4.5.6.

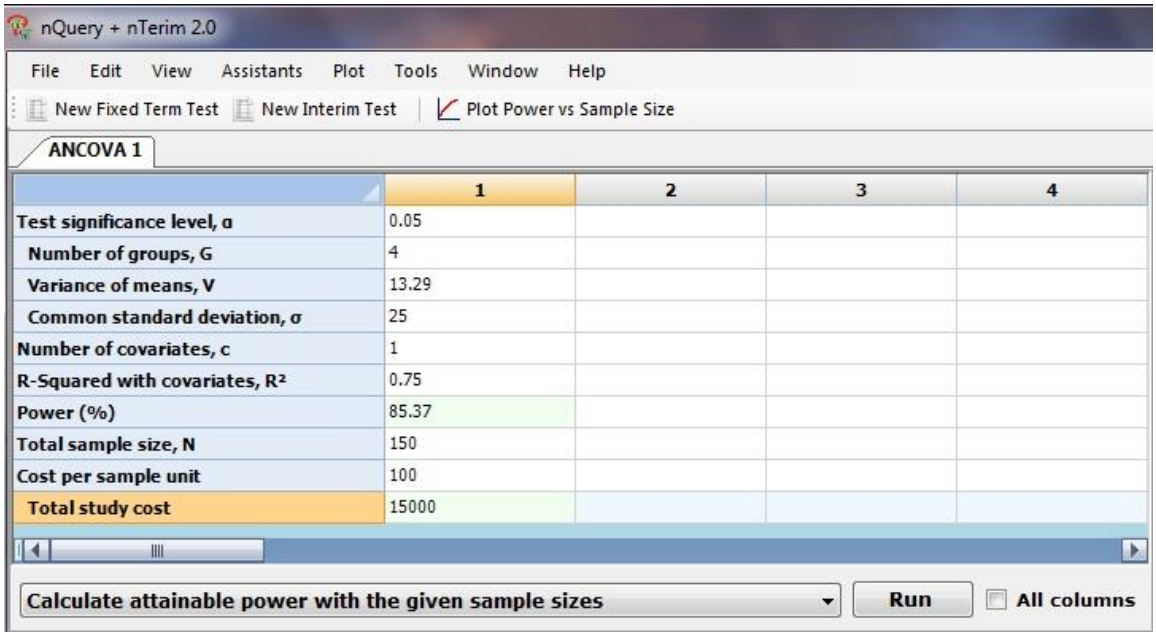


Figure 4.5.6. Completed ANCOVA Test Table

It can be seen from Figure 4.5.6 that a sample size of 150 is required to obtain a power of 85.37%. Due to the cost per sample unit of $100, the overall cost of sample size required has amounted to $15,000.

By clicking on the Output tab at the bottom of the screen you can see a statement giving details of the calculation:

"When the total sample size across the 4 groups is 150, distributed across the groups as specified, an analysis of covariance will have 85.37% power to detect at the 0.05 level a difference in means characterized by a Variance of means of 13.29, assuming that the common standard deviation is 25, and assuming the covariate(s) has an R-squared of 0.75."

Example 2: Investigating the effects of R-squared on attainable Power
In this example we will examine how the R-squared with covariates value has an impact on the attainable power given a certain sample size.

The following steps outline the procedure for Example 2.

1. Open nTerim through the Start Menu or by double clicking on the nTerim desktop icon. Then click on "New Fixed Term Test" from the menu bar at the top of the window. A "Study Goal and Design" window will appear.

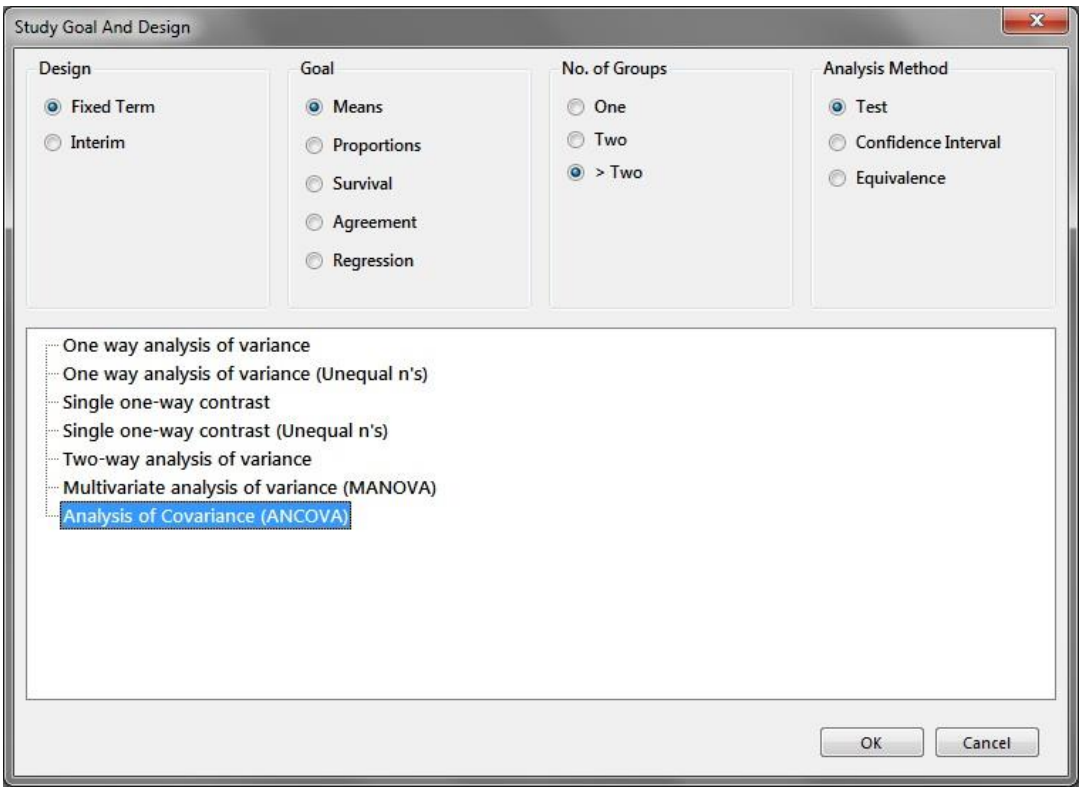


Figure 4.5.7. Study Goal and Design Window

2. Once the correct test has been selected, click "OK" and the test window will appear.

3. There are two main tables required for this test, the main test table illustrated in Figure 4.5.8 and the effect size assistant table shown in Figure 4.5.9.

4. Enter 0.05 for alpha, the desired significance level, and enter 3 for the number of groups, G, as shown in Figure 4.5.10.

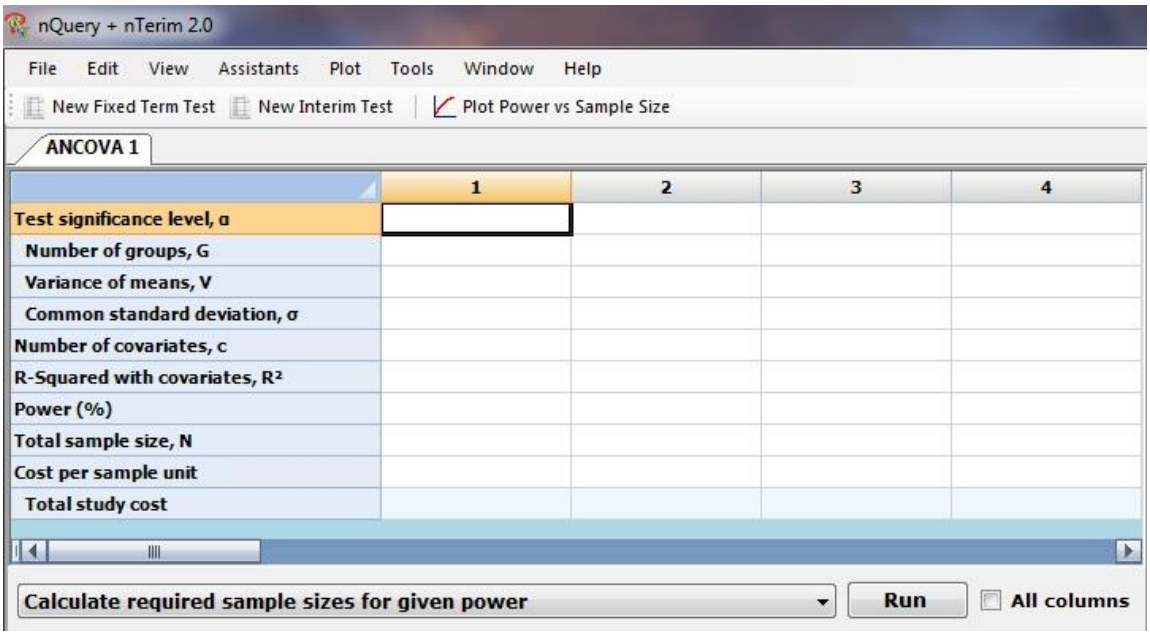


Figure 4.5.8. Analysis of Covariance Test Table

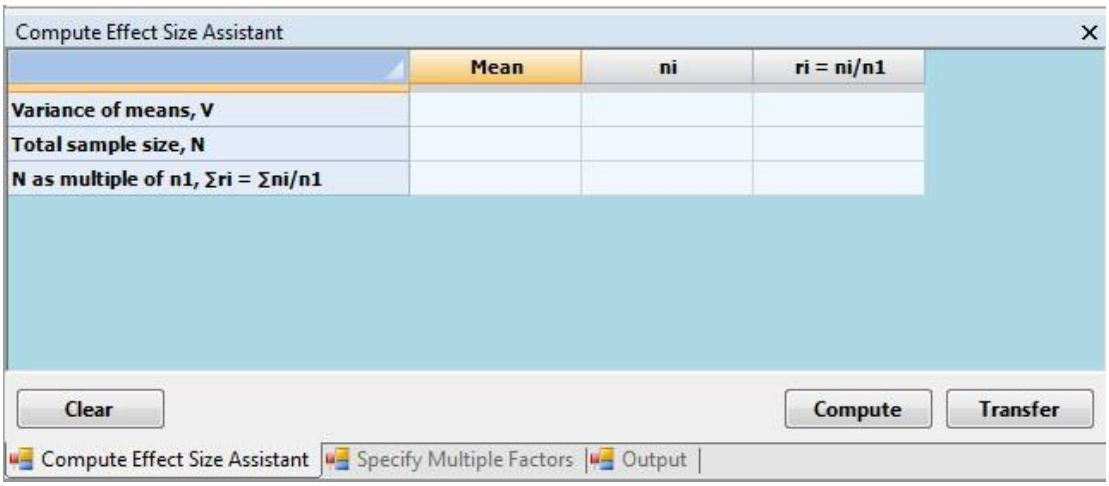


Figure 4.5.9. Compute Effect size Assistant Window

5. Once you enter a value for the number of groups, G, the "Compute Effect Size Assistant" table updates automatically as shown in Figure 4.5.10.

6. In order to calculate a value for Effect Size, the Variance of Means (V) needs to be calculated first.

7. The mean for each level and the corresponding sample size need to be entered in the "Compute Effect Size Assistant" table.

8. For the “Mean” values for each group, enter 31 for group 1, 41 for group 2 and 45 for group 3.

9. For the group sample size (n_i) values for each group, enter 40 for group 1, 45 for group 2 and 35 for group 3. As a result, the ratio (r_i) is calculated for each group as a proportion of group 1.

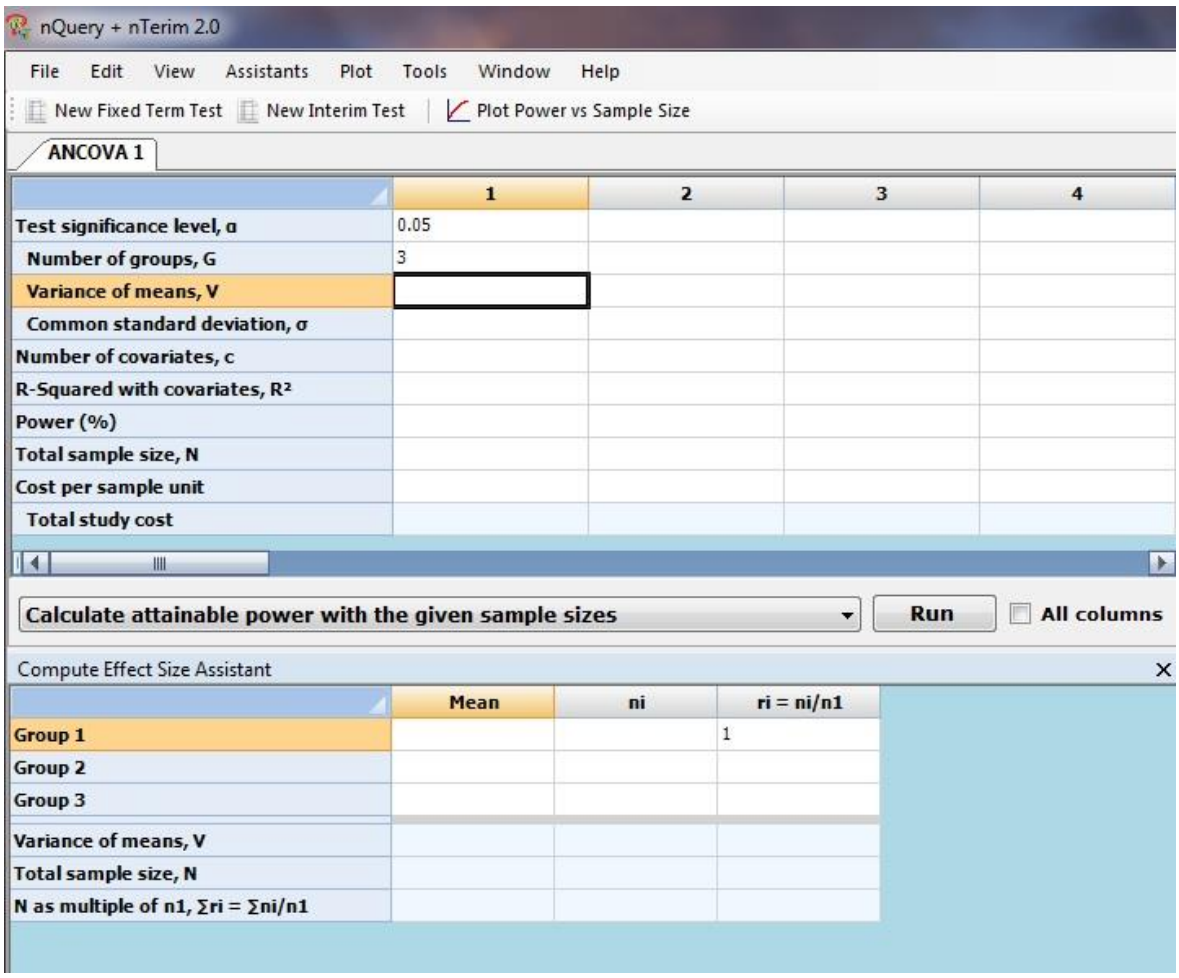


Figure 4.5.10. Automatically updated Compute effect size Assistant Window

10. Once the table in Figure 4.5.11 has been completed, the values for Variance of Means (V) and Total sample size (N) are computed, click on “Transfer” to automatically transfer these values to the main ANCOVA test table.

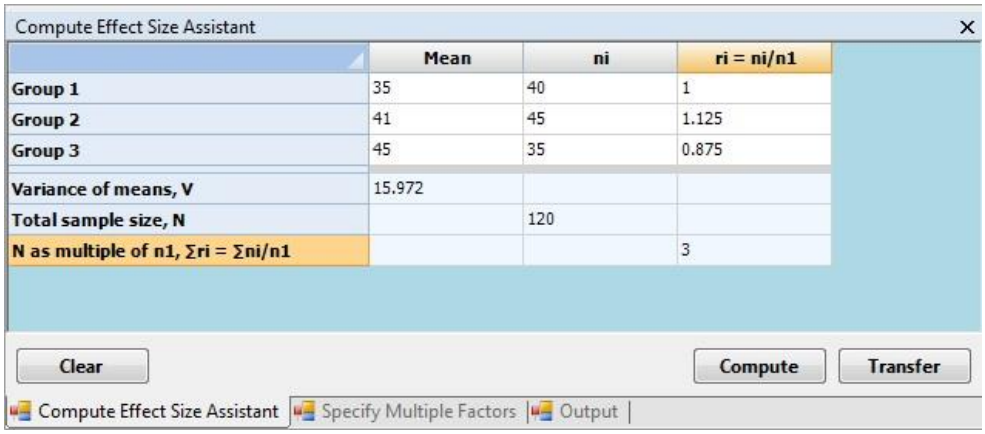


Figure 4.5.11. Completed Compute Effect size Assistant Window

11. Now that values for Variance of Means (V) and Total sample size (N) are computed we can continue with filling in the main table. For the Common Standard Deviation, enter a value of 30.

12. The number of covariates to be used in this study is set at 1, so enter the value 1 in the "Number of covariates" row. Also the R-Squared value has been estimated as 0.5 for this study design so enter 0.5 in the "R-Squared with covariates" row.

13. We want to calculate the attainable power give the sample size of 120.

14. It has been estimated that it will cost $80 per sample unit in this study. Therefore enter 80 in the "Cost per sample unit" row.

15. As we want to compare the effects that the R-Squared value has on the Power of the study, we will re-run this design for several values of R-Squared. To do this, right click on Column 1 as shown in Figure 4.5.12 and select "Fill Right". This will replicate the information in Column 1 across all the columns in this window.

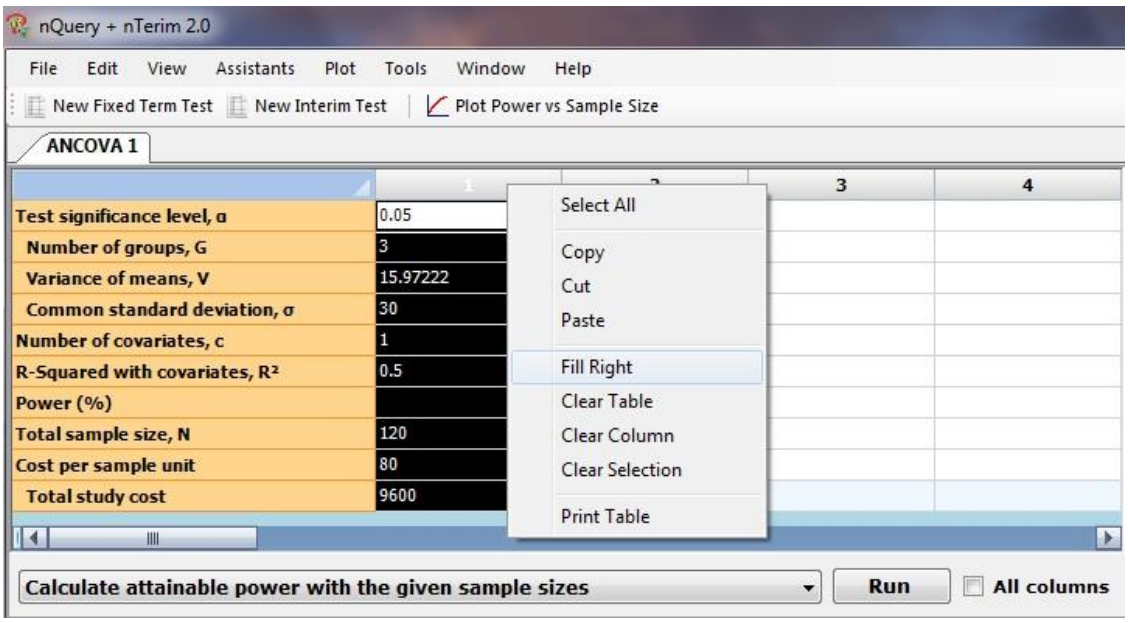
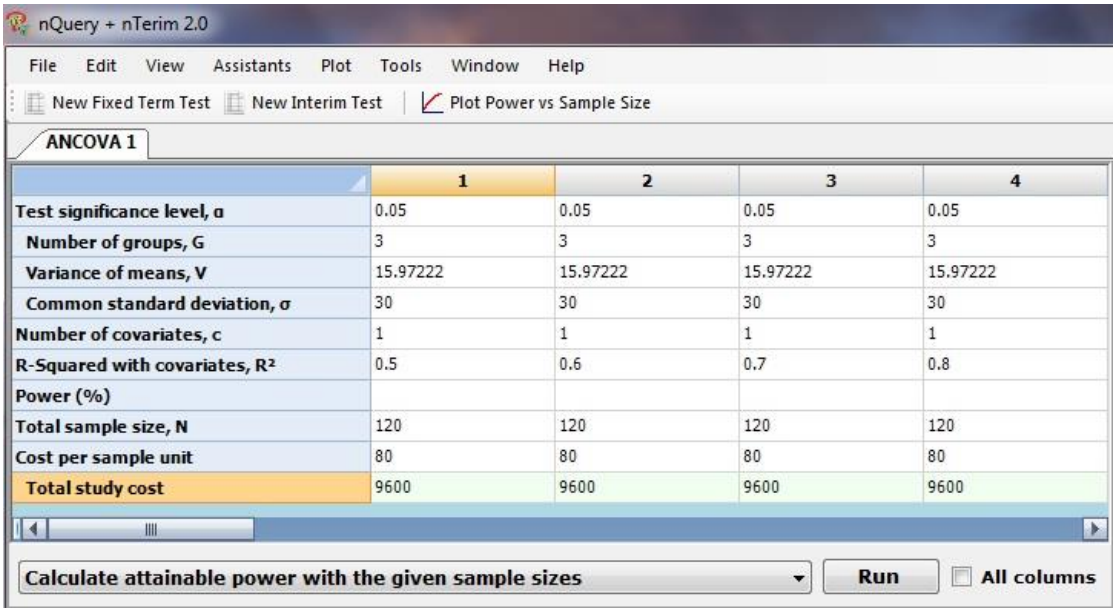


Figure 4.5.12. Fill Right Shortcut Feature

16. Now we want to change the R-Squared values in Columns 2, 3 and 4 to represent the remaining possible estimated R-Squared values for our study design. We would like to investigate R-Squared ranging from 0.5 (in Column 1) to 0.8 (in Column 4). To do this, enter 0.6 in the "R-Squared with covariates" row in Column 2, 0.7 in Column 3 and 0.8 in Column 4 as illustrated in Figure 4.5.13 below.

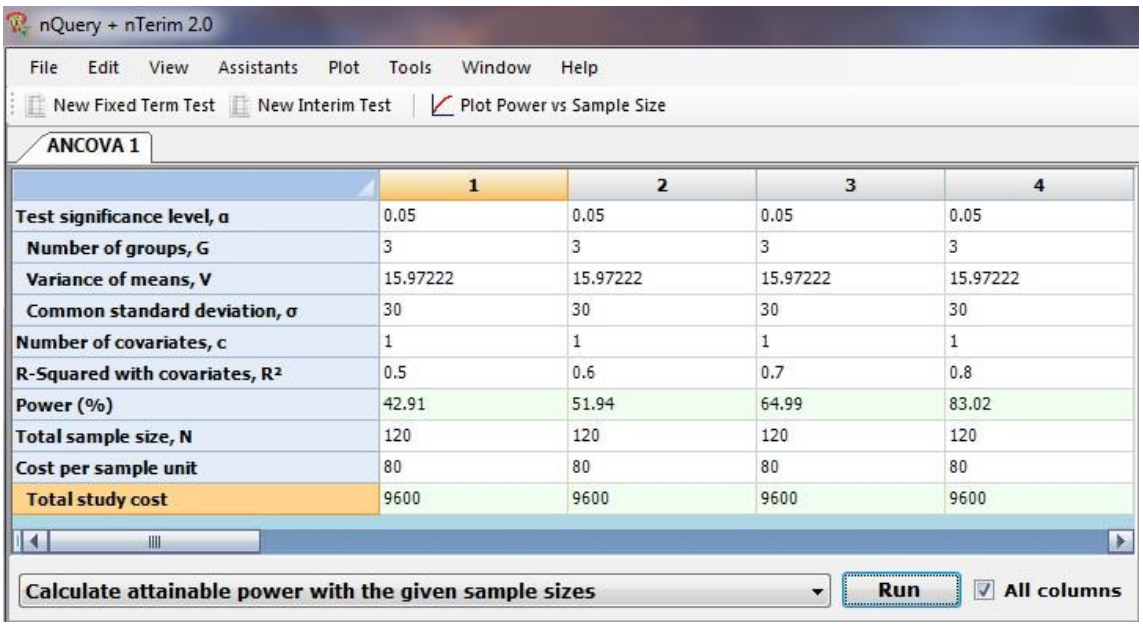


	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
Number of groups, G	3	3	3	3
Variance of means, V	15.97222	15.97222	15.97222	15.97222
Common standard deviation, σ	30	30	30	30
Number of covariates, c	1	1	1	1
R-Squared with covariates, R^2	0.5	0.6	0.7	0.8
Power (%)				
Total sample size, N	120	120	120	120
Cost per sample unit	80	80	80	80
Total study cost	9600	9600	9600	9600

Figure 4.5.13. Altered columns for R-Squared Comparison

17. Now that all the information in each column has been entered, we are ready to run the calculations. In order to calculate the power for all the columns together, tick the "All columns" box beside the "Run" button as shown in Figure 4.5.13.

18. Now select "Calculate attainable power given sample size" from the drop-down menu below the main table and click "Run".



	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
Number of groups, G	3	3	3	3
Variance of means, V	15.97222	15.97222	15.97222	15.97222
Common standard deviation, σ	30	30	30	30
Number of covariates, c	1	1	1	1
R-Squared with covariates, R^2	0.5	0.6	0.7	0.8
Power (%)	42.91	51.94	64.99	83.02
Total sample size, N	120	120	120	120
Cost per sample unit	80	80	80	80
Total study cost	9600	9600	9600	9600

Figure 4.5.14. Completed multiple design ANCOVA Table

As the results show in Figure 4.5.14, as the R-Squared value is increase from 0.5 up to 0.8, the corresponding power also increase dramatically, almost doubling from 42.91% to 83.02%. It can be seen from this approach that we would want an R-Squared value approximately equal to 0.8 to obtain a credible value for power.

19. Another feature that enables us to compare designs side-by-side is by using the Power vs. Sample Size plot. Multiple columns can be plotted together by simply highlighting the desired columns and clicking on the "Plot Power vs Sample Size" button on the menu bar.

20. To highlight the desired columns, click on the column title for Column 1 and drag across to Column 4. Then click on the "Plot Power vs Sample Size" button on the menu bar. The multiple column plot is displayed in Figure 4.5.15.

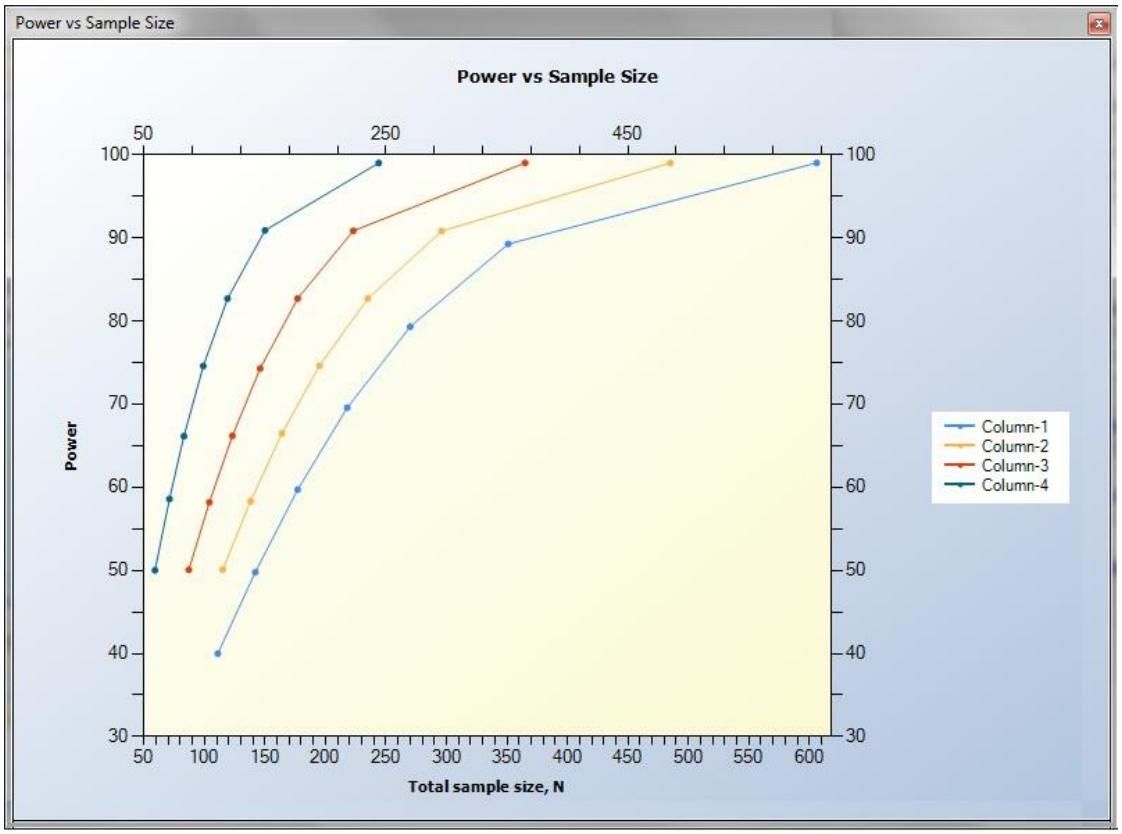


Figure 4.5.15. Power vs. Sample Size Plot

It can be seen from the legend on the left-hand side (legend can be altered manually) that the blue line represents Column 1, the orange line represents Column 2 and the red line represents Column 3. The cross on the graph illustrates how the user can identify what the sample size is for a corresponding power value for each column. In the bottom right corner of the plot indicated the exact values for Power and Sample Size for each identifier on the graph.

4.6. Multivariate Analysis of Variance (MANOVA)

4.6.1. Introduction

This table facilitates the calculation of power and sample size for multivariate analysis of variance (MANOVA) designs. In multivariate models there are several test statistics that can be used. In nTerim we provide the option for power and sample size calculations using three common test statistics; Wilks' likelihood ratio statistic, Pillai-Bartlett trace, and Hotelling-Lawley trace. Calculations are performed using the approximations outlined by Muller and Barton (1989), and Muller, LaVange, Ramey and Ramey (1992).

Multivariate analysis of variance (MANOVA) analysis is very similar to its univariate counterpart, analysis of variance (ANOVA). MANOVA can be described simply as an ANOVA with several response variables. In ANOVA, differences in means between two or more groups are tested on a single response variable. In MANOVA the number of response variables is increased to two or more. The purpose of MANOVA is to test for the difference in the vectors of means for two or more groups.

To give an example; we may be conducting a study where we are comparing two different treatments; a new treatment and a standard treatment, and we are interested in improvements in subjects scores for depression, life satisfaction and physical health. In this example, improvements in depression, life satisfaction and physical health are the response variables and our null hypothesis is that a subject's treatment has not effect on any of the three different ratings. As there are three response variables, MANOVA is used to this hypothesis.

4.6.2. Methodology

Power and sample size is calculated using central and non-central F-distributions and follows the procedures outlined by Muller and Barton (1989), and Muller, LaVange, Ramey and Ramey (1992).

To calculate power and sample size the user must first enter the number of response variables, p. The user must then specify the number of levels (categories) per factor in their design using the Factor Level Table assistant. Note if you wish to not use a factor in your design then you can simply leave the number of levels blank for that factor. Using this same table the alpha value and desired power per factor, and per factor interaction, must also be specified. Note if you are solving for power then you must leave the power fields blank.

Having specified the number of response variables and the number of levels per factor, the Means Matrix, M, becomes populated with empty cells that must be filled in by the user. The numbered rows of this matrix represent the response variables, p and the columns represent the factors, or to be more specific the number of groups that a subject can be classified in to, q. Where $q = Levels_A * Levels_B * Levels_C$

For example if you had a design with two response variables and 2 factors (Factor A and Factor B), each with two levels. This design would give a matrix with 2 rows and $q = 2 * 2 = 4$ columns.

$$M' = \begin{bmatrix} & A_1B_1 & A_1B_2 & A_2B_1 & A_2B_2 \\ p_1 & \mu_{11} & \mu_{12} & \mu_{13} & \mu_{14} \\ p_2 & \mu_{21} & \mu_{22} & \mu_{23} & \mu_{24} \end{bmatrix} \qquad \{4.6.1\}$$

Where, for example μ_{23} is the mean of the second response of subjects in the third group. Note the matrix is in this form for ease of user input. The transpose of this inputted matrix is used in the power calculations.

In the means matrix there is also a row labelled n. This row is used to specify the number of subjects per group. This row need only be specified when solving for power and it is anticipated that the sample size per group will be unequal.

The next step for the user is to input values for the standard deviation, σ, and the correlation, ρ. These two values are used by nTerim to calculate the covariance matrix Σ.

$$\Sigma = \begin{bmatrix} \sigma^2 & \sigma^2\rho & \dots & \sigma^2\rho \\ \sigma^2\rho & \sigma^2 & \dots & \sigma^2\rho \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^2\rho & \sigma^2\rho & \dots & \sigma^2 \end{bmatrix} \qquad \{4.6.2\}$$

Where Σ is a $p\ x\ p$ matrix.

Alternatively the user may manually specify the covariance matrix using the Covariance Matrix assistant. When values for standard deviation and the correlation are not entered and the covariance matrix has been filled out, nTerim will use the specified covariance matrix to compute power and sample size.

In order to calculate power, a value for the group size, n, must be entered. Entering this value in the main table assumes that group sizes are equal. If it is expected that the sample sizes in each group will be different, then, the expected sample size in each group must be specified in the Means Matrix.

nTerim gives the option of calculating power using one of three commonly used test statistics; Wilks' lambda, Pillai-Bartlett Trace, or Hotelling-Lawley trace.

In order to perform calculations using either of these three statistics nTerim first calculates the matrices Θ, H, E, and T using the following formulas:

$$\Theta = CM \qquad \{4.6.3\}$$

where C is a matrix of contrasts that nTerim automatically generates. This is an orthogonal matrix that is unique to each factor and factor interaction. M is the means matrix which has been inputted by the user.

$$H = \left(\widehat{\Theta} - \Theta_0\right)'[C(X'X)^{-}C']^{-1}(\widehat{\Theta} - \Theta_0) \qquad \{4.6.4\}$$

where Θ_0 is the matrix of hypothesised means, which is zero for this test, and X is the design matrix.

$$E = \hat{\Sigma}(N - r) \qquad \{4.6.5\}$$

where $\hat{\Sigma}$ is the covariance matrix.

$$T = H + E \qquad \{4.6.6\}$$

Wilks' Lambda
Using these matrices the test statistic for Wilks' lambda is calculated using the formula

$$W = |ET^{-1}| \qquad \{4.6.7\}$$

The transformation of this test statistic to an approximate F is given by

$$F_{df_1,df_2} = \frac{\eta/df_1}{(1-\eta)/df_2} \qquad \{4.6.8\}$$

where,

$$\eta = 1 - W^{\frac{1}{g}} \qquad \{4.6.9\}$$

$$g = \left(\frac{a^2p^2 - 4}{a^2 + p^2 - 5}\right)^{\frac{1}{2}} \qquad \{4.6.10\}$$

$$a = q - 1 \qquad \{4.6.11\}$$

$$df_1 = ap \qquad \{4.6.12\}$$

$$df_2 = g[(N - r) - (p - a + 1)/2] - (ap - 2)/2 \qquad \{4.6.13\}$$

Pillai-Bartlett Trace
The test statistic for Pillai-Bartlett trace is calculated using the formula

$$PBT = tr(HT^{-1}) \qquad \{4.6.14\}$$

The transformation of this test statistic to an approximate F is given by

$$F_{df_1,df_2} = \frac{\eta/df_1}{(1 - \eta)/df_2} \qquad \{4.6.15\}$$

$$\eta = \frac{PBT}{s} \qquad \{4.6.16\}$$

$$s = \min(a, p) \qquad \{4.6.17\}$$

$$a = q - 1 \qquad \{4.6.18\}$$

$$df_1 = ap \qquad \{4.6.19\}$$

$$df_2 = s[(N - r) - p + s] \qquad \{4.6.20\}$$

Hotelling-Lawley Trace
The test statistic for Pillai-Bartlett trace is calculated using the formula

$$HLT = tr(HE^{-1}) \qquad \{4.6.21\}$$

The transformation of this test statistic to an approximate F is given by

$$F_{df_1,df_2} = \frac{\eta/df_1}{(1 - \eta)/df_2} \qquad \{4.6.22\}$$

$$\eta = \frac{HLT/s}{1 + HLT/s} \qquad \{4.6.23\}$$

$$df_1 = ap \qquad \{4.6.24\}$$

$$df_2 = s[(N - r) - p - 1] + 2 \qquad \{4.6.25\}$$

Depending on which of these three statistics is chosen, nTerim then calculates the power of the design by first determining the critical value $F_{df_1,\ df_2,\ \alpha}$ and then the noncentrality parameter, λ. Where,

$$\lambda = \eta df_1 \qquad \{4.6.26\}$$

Using these two values, nTerim will calculate the power of this design as the probability of being greater than $F_{df_1,\ df_2,\ \alpha}$ on a non-central F-distribution with non-centrality parameter λ.

In order to calculate sample size, values for power must be specified in the Factor Level Table. nTerim does not use a closed form equation. Instead a search algorithm is used. This search algorithm calculates power at various sample sizes until the desired power is reached.

4.6.3. Examples

Example 1: Pillai – Bartlett Trace
In this example we will calculate the attainable power given a specified sample size using the Pillai – Bartlett trace method

The following steps outline the procedure for Example 1.

1. Open nTerim through the Start Menu or by double clicking on the nTerim desktop icon. Then click on "New Fixed Term Test" from the menu bar at the top of the window. A "Study Goal and Design" window will appear.

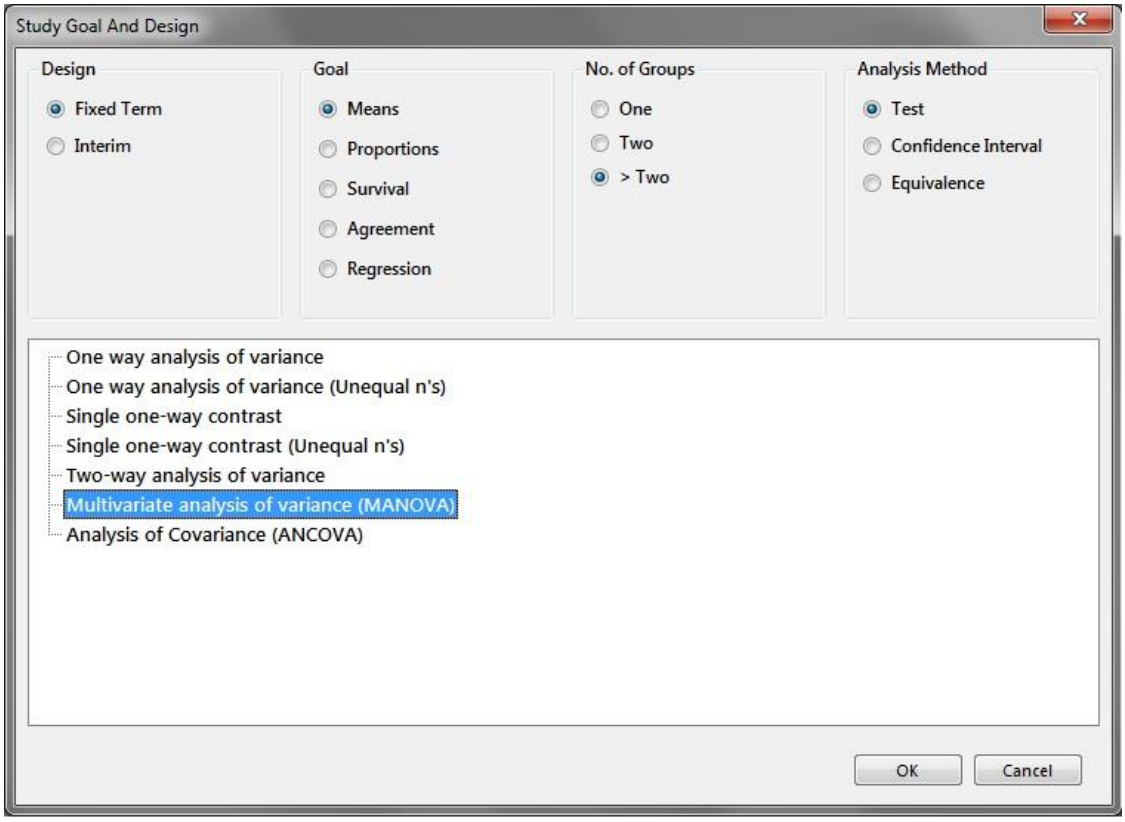


Figure 4.6.1. Study Goal and Design Window

2. Once the correct test has been selected, click "OK" and the test window will appear. This window is illustrated in Figure 4.6.2.

3. There are several tables required for this test including; the main test table shown in Figure 4.6.2, the Factor Level table illustrated in Figure 4.6.4, the Means Matrix assistant table presented in Figure 4.6.5, and the Covariance Matrix assistant table shown in Figure 4.6.7.

4. To begin we first need to specify the number of response variables to be used in the study. In this example we are using 2 so enter 2 in the "Number of response variables, p" row as shown in Figure 4.6.3.

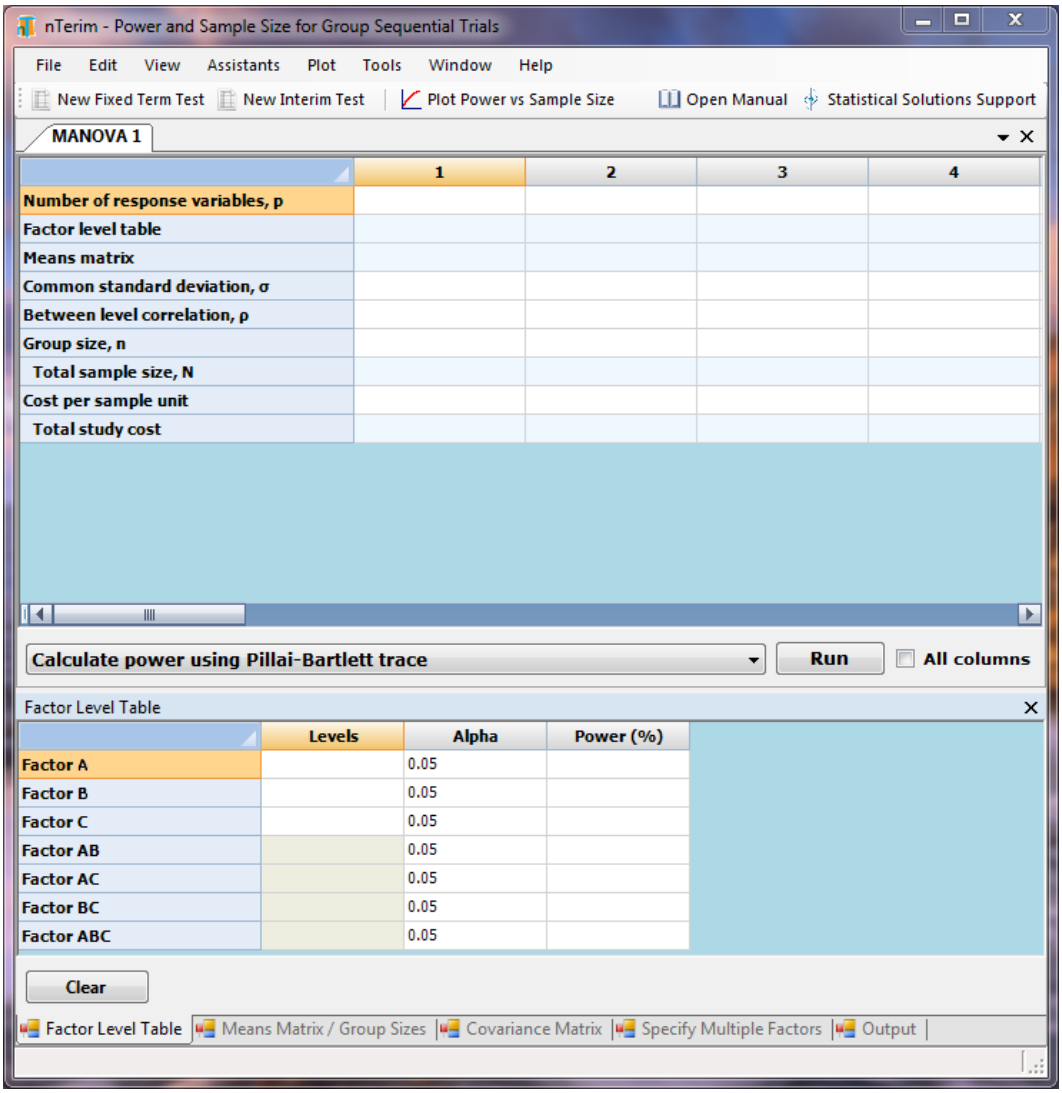


Figure 4.6.2. Multivariate Analysis of Variance Design Window

5. The next step in this process is to specify the number of levels per factor. This can be done using the Factor Level Assistant table illustrated in Figure 4.6.4.

6. In this example we are going to specify 4 levels for Factor A and 3 levels for Factor B. Seeing as we only highlighted two response variables in this example, we can leave Factor C empty.

7. We can also alter the default settings of 0.05 for the alpha value. This represents the significance level for each factor. In this example we will leave it at 0.05.

8. Finally, the as we are calculating attainable power, the Power (%) is where our output power values for each factor will appear, thus we leave this column empty.

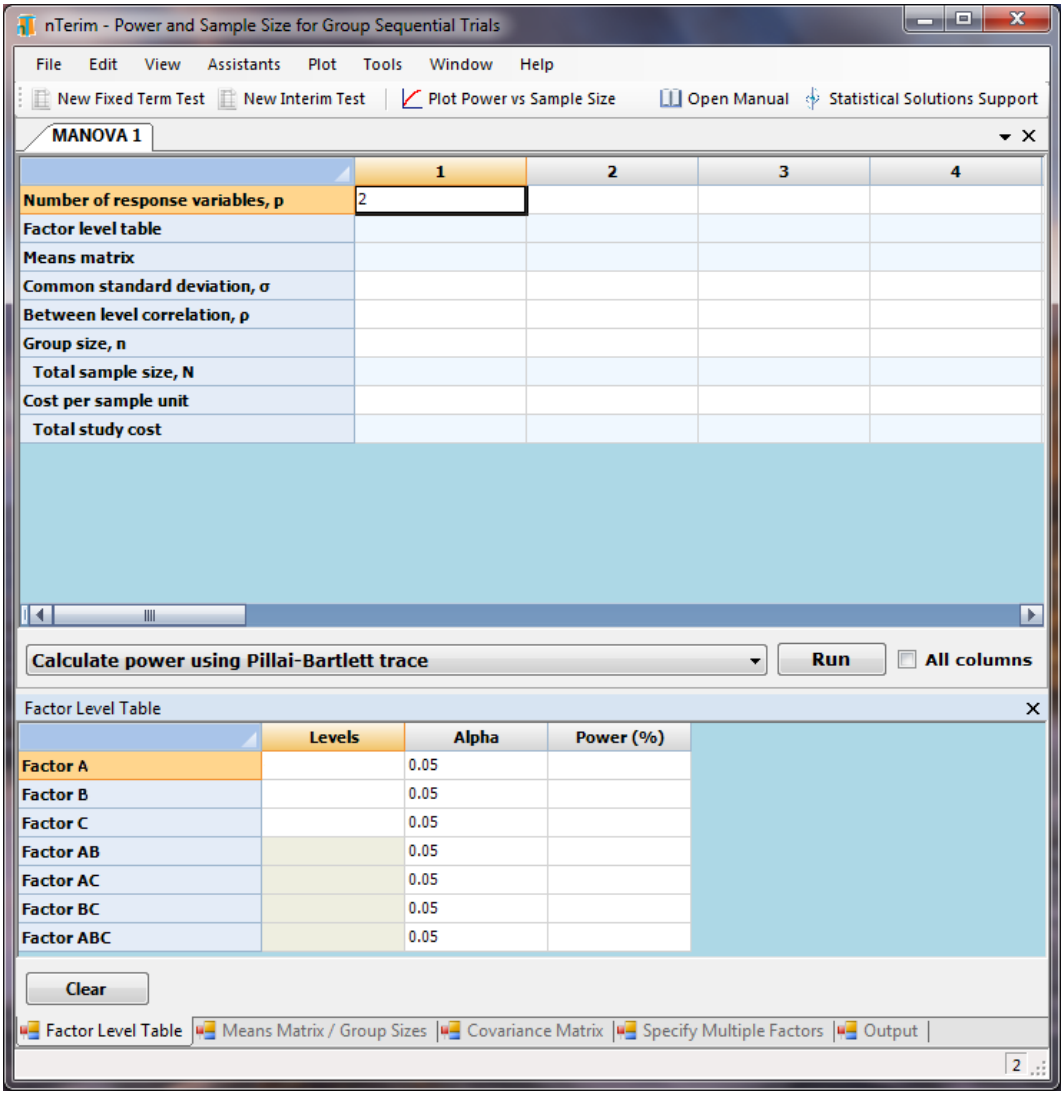


Figure 4.6.3. Enter Number of Response variables

9. Once the number of levels for each factor has been specified, the next step is to populate the Means Matrix. The Means Matrix is displayed in Figure 4.6.5.

10. Depending on the values entered into the Factor Level table, the size of the means matrix will be created.

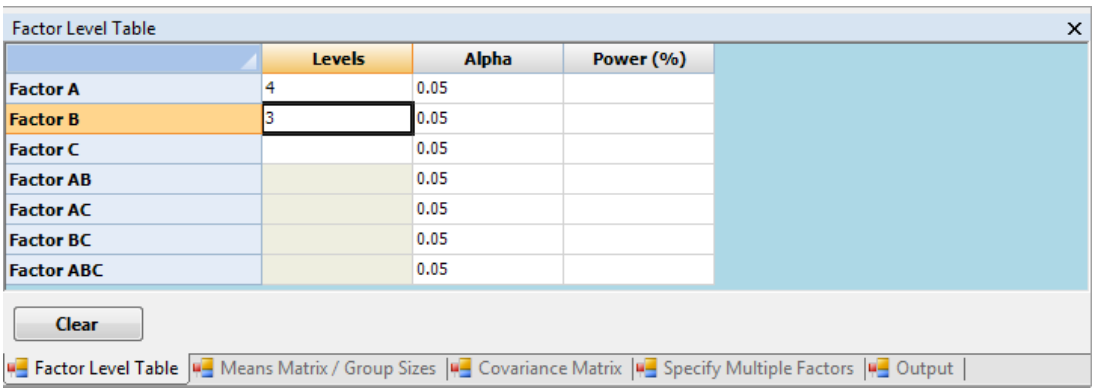


Figure 4.6.4. Factor Level Table

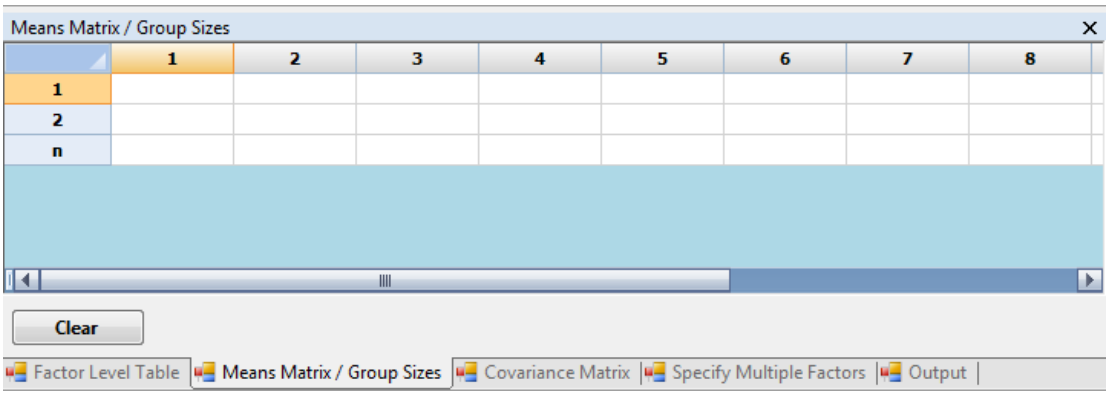


Figure 4.6.5. Means Matrix / Group Sizes Assistants Table

11. As we have defined 2 response variables, one with 4 levels and one with 3 levels, we will require a Means Matrix with 2 rows and 3x4 columns. There is an extra row included to enable the user to specify the individual level sample size (only needed if unequal sample sizes per level).

12. The next step is to fill in all the values for each part of the Means Matrix. In this example we will define the Means Matrix as below (first column of matrix are row names):

$$M' = \begin{bmatrix} 1 & 1 & 1 & 2 & 1 & 2 & 3 & 4 & 2 & 3 & 4 & 2 & 5 \\ 2 & 1 & 2 & 1 & 4 & 1 & 2 & 1 & 4 & 2 & 1 & 1 & 4 \\ n & 6 & 8 & 7 & 4 & 5 & 6 & 4 & 4 & 4 & 4 & 5 & 4 \end{bmatrix}$$

13. Enter this matrix in the Means Matrix Assistant table as illustrated in Figure 4.6.6.

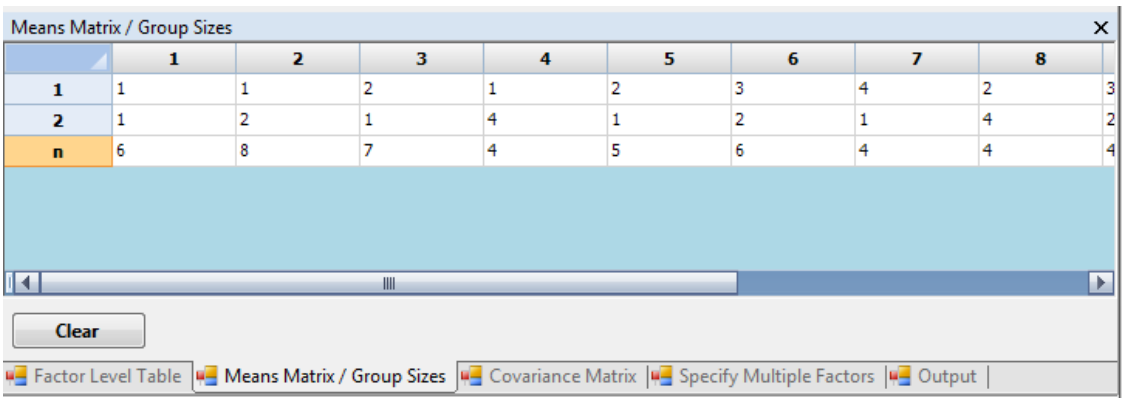


Figure 4.6.6. Completed Means Matrix and Group Sizes Assistants Table

14. The bottom row is summed to give the total sample size required and automatically entered into the main design table.

15. The next step in this MANOVA process is to specify the Covariance Matrix.

16. The user has the option to enter the standard deviation and correlation, thus nTerim will automatically calculate the Covariance Matrix but in this example we do not have this information so we will enter the Covariance Matrix directly.

17. Enter the Covariance Matrix as given below:

$$\Sigma = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}$$

18. The Covariance Matrix entered in nTerim is shown in Figure 4.6.7 below.

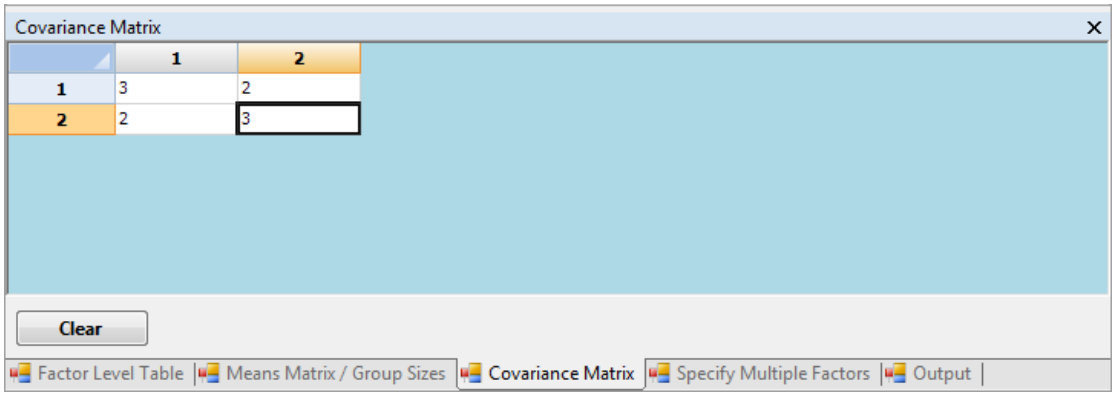


Figure 4.6.7. Covariance Matrix Assistant Table

19. Now we can entered all the information required to calculate attainable the Power given a specified sample size.

20. The final step is to select which method we want to use. In this case we want to use the Pillai – Bartlett Trace approach.

21. In order to do this simply select the "Calculate power using Pillai – Bartlett trace" and the click on "Run" as shown in Figure 4.6.8 below.

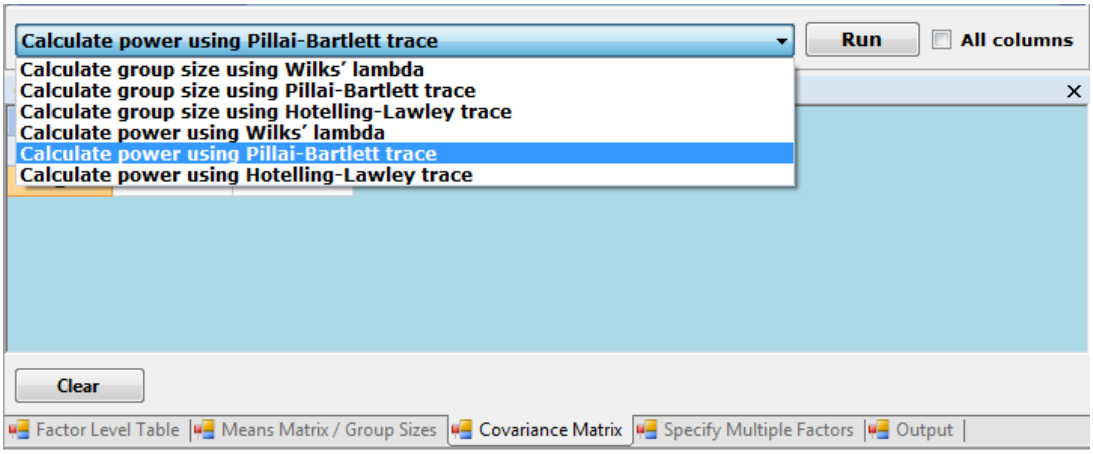
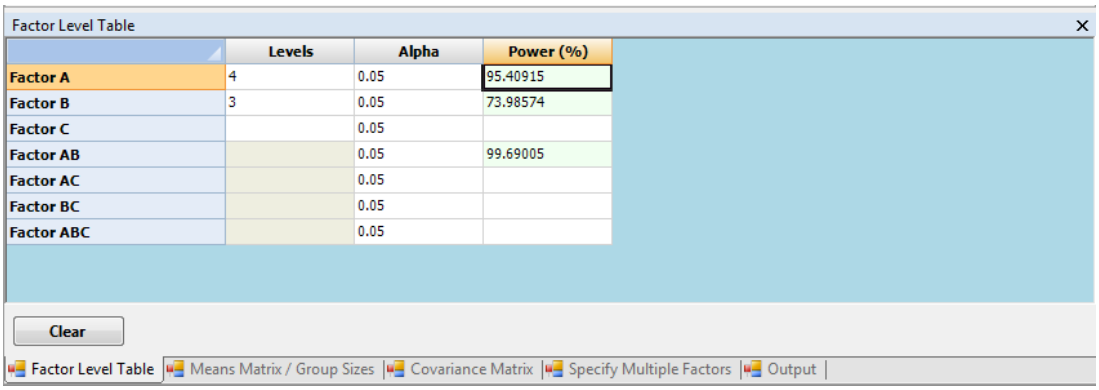


Figure 4.6.8. Selecting Type of Test to Run

22. In order to view the results for Power for each level, the power values are displayed in the Factor Level Assistants table as illustrated below in Figure 4.6.9.



	Levels	Alpha	Power (%)
Factor A	4	0.05	95.40915
Factor B	3	0.05	73.98574
Factor C		0.05	
Factor AB		0.05	99.69005
Factor AC		0.05	
Factor BC		0.05	
Factor ABC		0.05	

Figure 4.6.9. Output Power values calculated

23. Finally, the output statement can be obtained by clicking on the Output tab on the bottom of the nTerim window.

Output Statement:
"A multivariate analysis of variance design with 2 factors and 2 response variables has 12 groups. When the total sample size across the 12 groups is 61, distributed across the groups as specified, a multivariate analysis of variance will have:

95.41 power to test Factor A if a Pillai Bartlett Trace test statistic is used with 0.05 significance level.
73.99 power to test Factor B if a Pillai Bartlett Trace test statistic is used with 0.05 significance level.
99.69 power to test Factor AB if a Pillai Bartlett Trace test statistic is used with 0.05 significance level."

Example 2: Wilks' Lambda

In this example we will calculate the attainable power given a specified sample size using the Wilks' Lambda method

The following steps outline the procedure for Example 2.

1. Open nTerim through the Start Menu or by double clicking on the nTerim desktop icon. Then click on "New Fixed Term Test" from the menu bar at the top of the window. A "Study Goal and Design" window will appear.

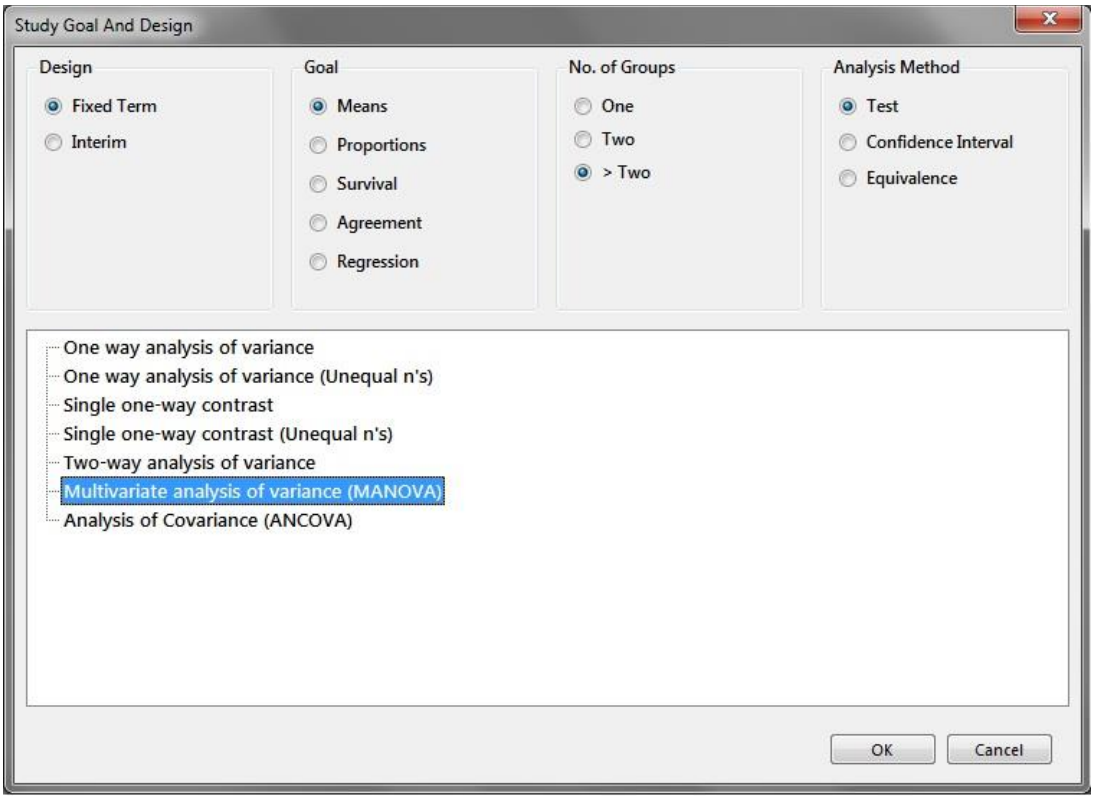


Figure 4.6.10. Study Goal and Design Window

2. Once the correct test has been selected, click "OK" and the test window will appear. This window is illustrated in Figure 4.6.11.

3. There are several tables required for this test including; the main test table shown in Figure 4.6.11, the Factor Level table illustrated in Figure 4.6.4, the Means Matrix assistant table presented in Figure 4.6.5, and the Covariance Matrix assistant table shown in Figure 4.6.7.

4. To begin we first need to specify the number of response variables to be used in the study. In this example we are using 3 so enter 3 in the "Number of response variables, p" row as shown in Figure 4.6.12.

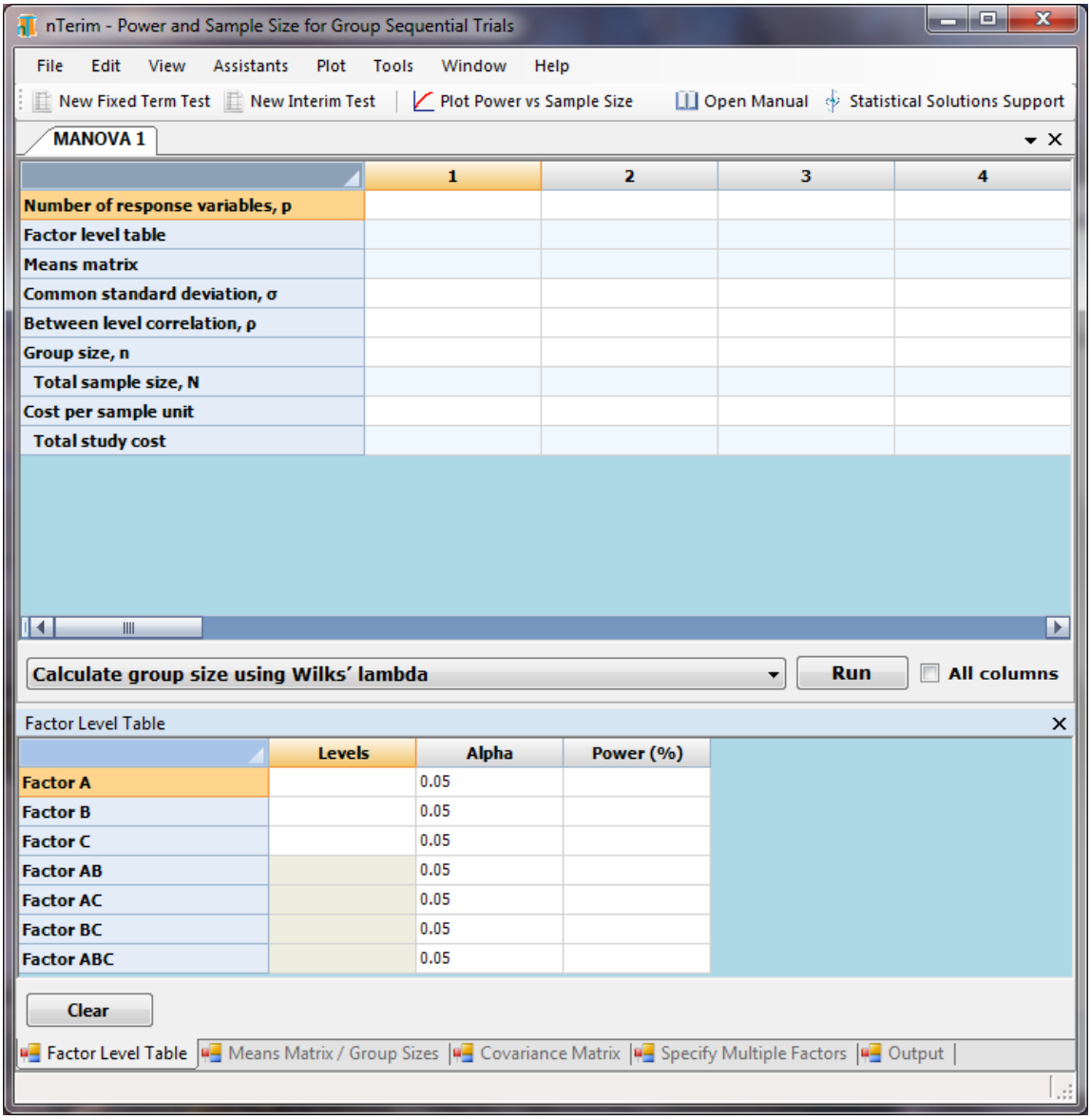


Figure 4.6.11. Multivariate Analysis of Variance Table

5. The next step in this process is to specify the number of levels per factor. This can be done using the Factor Level Assistant table illustrated in Figure 4.6.13.

6. In this example we are going to specify 3 levels for Factor A, 3 levels for Factor B and 3 levels for Factor C.

7. We can also alter the default settings of 0.05 for the alpha value. This represents the significance level for each factor. In this example we will leave it at 0.05.

8. Finally, the as we are calculating attainable power, the Power (%) is where our output power values for each factor will appear, thus we leave this column empty.

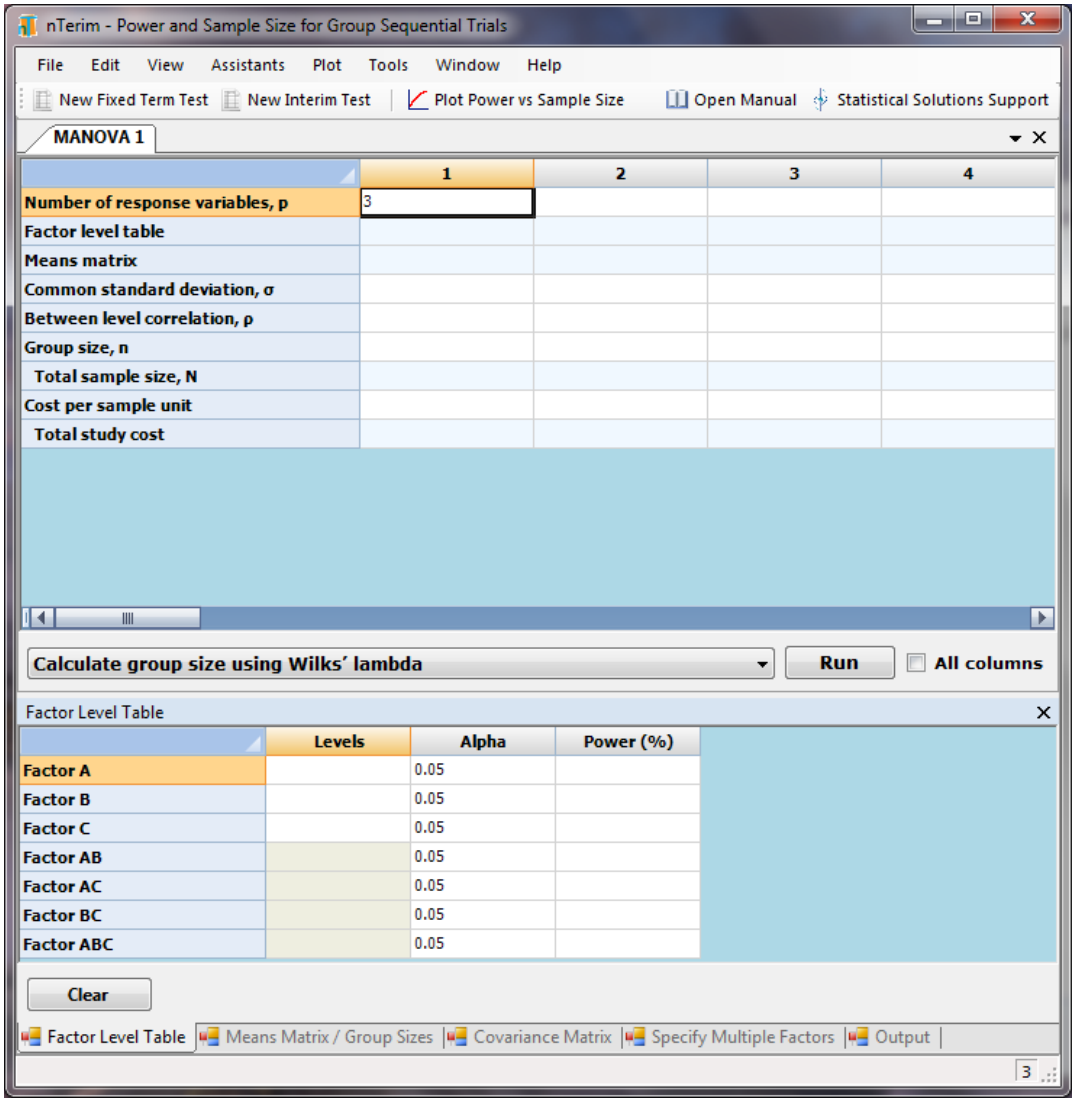


Figure 4.6.12. Enter Number of Response variables

9. Once the number of levels for each factor has been specified, the next step is to populate the Means Matrix. The Means Matrix is displayed in Figure 4.6.14.

10. Depending on the values entered into the Factor Level table, the size of the means matrix will be created.

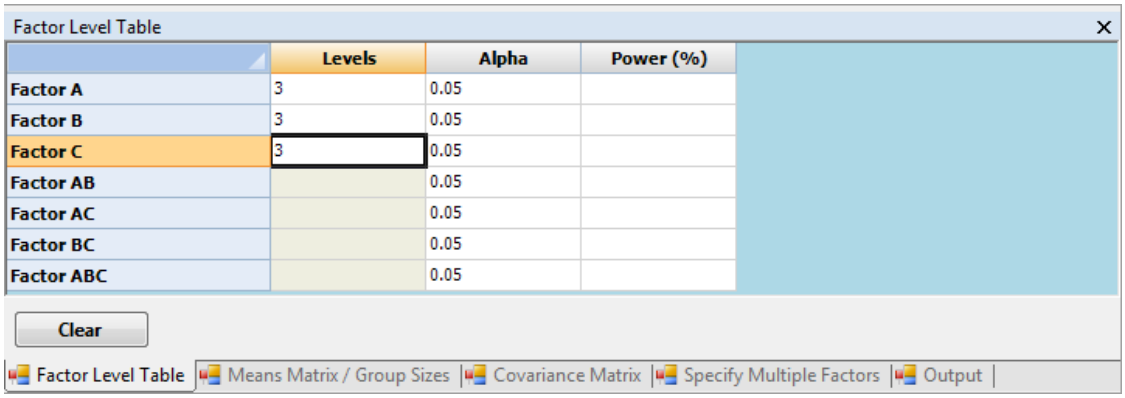


Figure 4.6.13. Factor Level Table

11. As we have defined 3 response variables, all with 3 levels each, we will require a Means Matrix with 3 rows and 3x3x3 columns. There is an extra row included to enable the user to specify the individual level sample size (only needed if unequal sample sizes per level).

12. The next step is to fill in all the values for each part of the Means Matrix. In this example we will define the Means Matrix as below (first column of matrix are row names):

$$M' = \begin{bmatrix} 1 & 1 & 1 & 2 & 1 & 2 & 3 & 4 & 2 & 3 & 4 & 2 & 5 & 1 & 1 & 2 & 1 & 2 & 3 & 4 & 2 & 3 & 4 & 2 & 1 & 1 & 2 \\ 2 & 1 & 2 & 1 & 4 & 1 & 2 & 1 & 4 & 2 & 1 & 1 & 4 & 1 & 2 & 1 & 4 & 1 & 2 & 1 & 1 & 4 & 1 & 2 & 1 \\ 3 & 6 & 8 & 7 & 4 & 5 & 6 & 4 & 4 & 4 & 4 & 5 & 4 & 6 & 8 & 7 & 4 & 5 & 6 & 4 & 5 & 4 & 6 & 8 & 7 \end{bmatrix}$$

13. Enter this matrix in the Means Matrix Assistant table as illustrated in Figure 4.6.14.

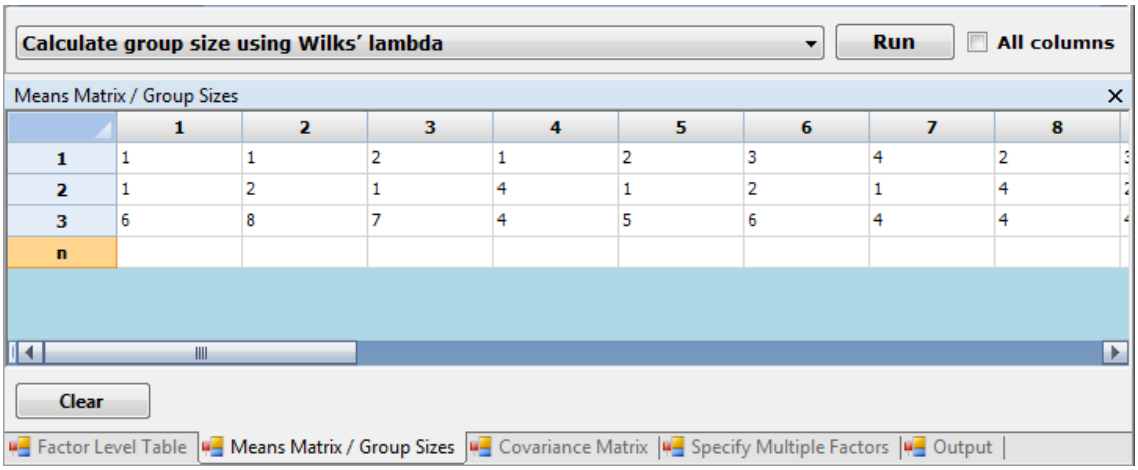


Figure 4.6.14. Completed Means Matrix / Group Sizes Assistant Table

14. The bottom row is summed to give the total sample size required and automatically entered into the main design table. In this case we are leaving the bottom row empty as we are going to specify the all groups have equal sample size. In this event, nTerim will automatically update this matrix once we have entered a value for Group Size in the main design table.

15. The next step is to create the Covariance Matrix. There are two ways of doing this in nTerim, one is to enter specify the matrix manually in the Covariance Matrix Assistant table and the other way is to enter values for common standard deviation and correlation so nTerim can create the matrix automatically.

16. In this example we are going to enter values for common standard deviation and correlation.

17. In the “Common standard deviation” row enter a value of 2. In the “Between level correlation” row, enter a value of 0.6. This is shown in Figure 4.6.15.

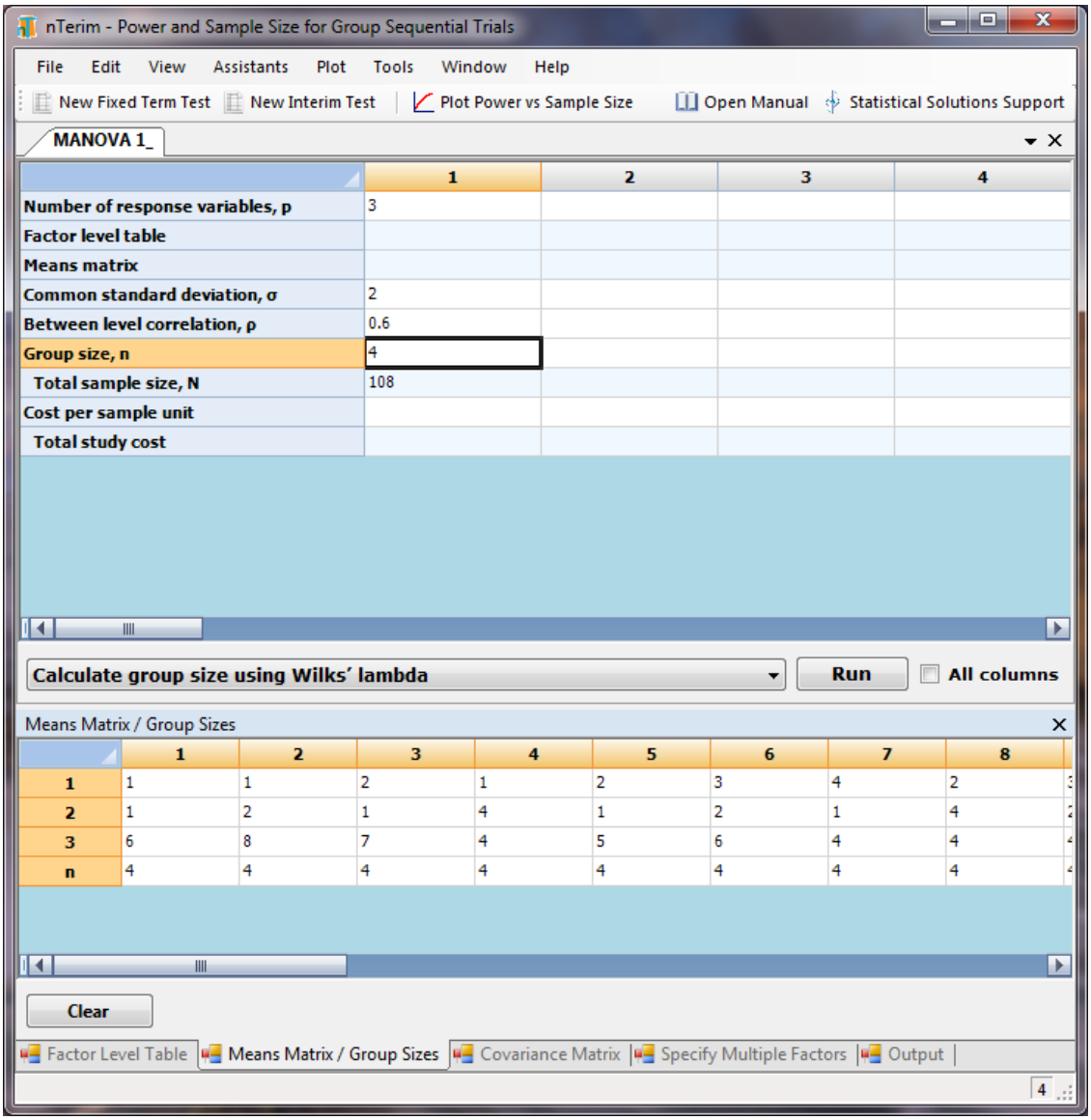


Figure 4.6.15. Completed MANOVA Table

18. The next step is to enter the Group Size and as the groups will have equal sizes in this example of 4, enter 4 in the “Group size, n” row. Notice that the Means Matrix in Figure 4.6.15 has now been updated with the sample size per group.

19. The total sample size is also automatically calculated and given in the “Total sample size, N” row.

20. The final step is to select which method we want to use. In this case we want to use the Wilks’ Lambda approach.

21. In order to do this simply select the “Calculate power using Wilks’ Lambda” and the click on “Run” as shown in Figure 4.6.16 below.

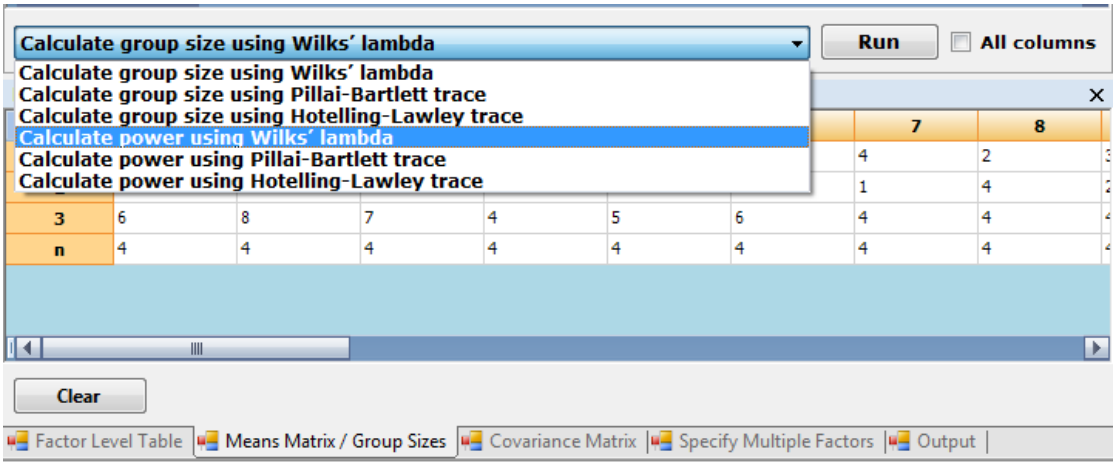


Figure 4.6.16. Selecting the Wilks’ Lambda option

22. In order to view the results for Power for each level, the power values are displayed in the Factor Level Assistants table as illustrated below in Figure 4.6.17.

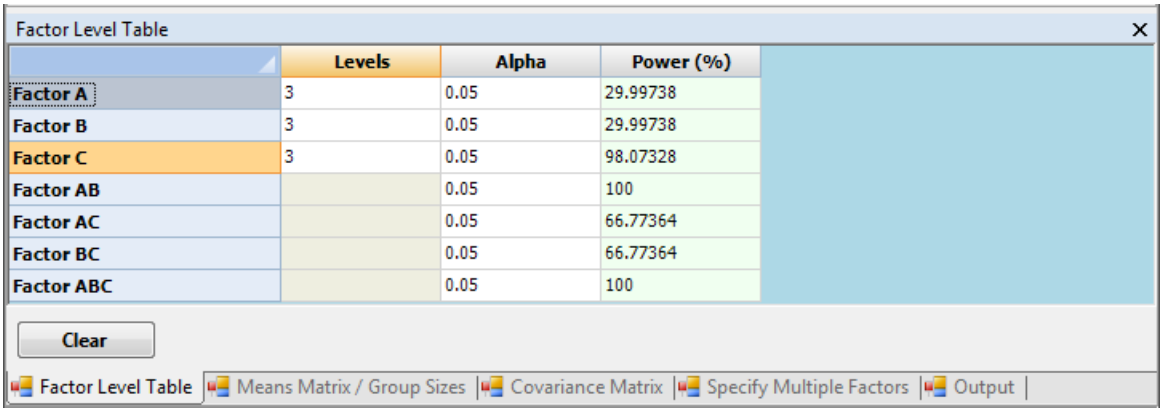


Figure 4.6.17. Output Power values calculated

23. Finally, the output statement can be obtained by clicking on the Output tab on the bottom of the nTerim window.

Output Statement:

“A multivariate analysis of variance design with 3 factors and 3 response variables has 27 groups. When the total sample size across the 27 groups is 108, distributed across the groups as specified, a multivariate analysis of variance will have:

30 power to test Factor A if a Wilks’ Lambda test statistic is used with 0.05 significance level.
30 power to test Factor B if a Wilks’ Lambda test statistic is used with 0.05 significance level.
98.07 power to test Factor C if a Wilks’ Lambda test statistic is used with 0.05 significance level.
100 power to test Factor AB if a Wilks’ Lambda test statistic is used with 0.05 significance level.
66.77 power to test Factor AC if a Wilks’ Lambda test statistic is used with 0.05 significance level.
66.77 power to test Factor BC if a Wilks’ Lambda test statistic is used with 0.05 significance level.
100 power to test Factor ABC if a Wilks’ Lambda test statistic is used with 0.05 significance level.”

4.7. One Poisson Mean

4.7.1. Introduction

Count data is obtained in a variety of clinical and commercial activities such as the number of accidents at a junction or number of occurrences of a disease in a year. The most common distribution used to model count data is the Poisson distribution. The one sample test is used to test hypotheses about the mean rate of a Poisson distributed sample against an alternative specified value.

This table facilitates the calculation of the power and sample size for hypothesis tests of the mean of single Poisson distributed sample. Power and sample size is computed using the method outlined by Guenther (1977).

4.7.2 Methodology

The test for hypotheses regarding λ, the Poisson mean, are taken from Ostle (1988). It is conducted in two steps.

1. Calculate the critical value (X_{crit}) such that the probability of rejecting H_0 is equal to alpha by finding the minimum value of X_{crit} which fulfils the relevant inequality:

$$f(\alpha) = \begin{cases} \alpha \leq 1 - \sum_{-\infty}^{x=X_{crit}} e^{-n\lambda_0} \frac{(n\lambda_0)^x}{x!}, & H_0: \lambda_0 \geq \lambda_1, H_a: \lambda_0 < \lambda_1 \\ \alpha \geq \sum_{-\infty}^{x=X_{crit}} e^{-n\lambda_0} \frac{(n\lambda_0)^x}{x!}, & H_0: \lambda_0 \leq \lambda_1, H_a: \lambda_0 > \lambda_1 \end{cases} \qquad \{4.7.1\}$$

2. Select n items and sum the total number of events. If the total number of events is greater than X_{crit} for the first equation or less than X_{crit} for the second equation, the null hypothesis is rejected.

The sample size calculation is taken from Guenther (1977). This table can be used to calculate the power, the sample size or the minimum detectable Poisson mean given all other terms in the table are specified.

To calculate power and sample size, the user must specify the test significance level (α), the Poisson mean under the null hypothesis (λ_0) and the Poisson mean under the alternative hypothesis (λ_1).

The sample size is calculated by using one of the following equations:

$$f(n) = \begin{cases} \frac{\chi_{2d;1-\beta}}{2\lambda_1} \leq n \leq \frac{\chi_{2d;\alpha}}{2\lambda_0}, & H_0: \lambda_0 \geq \lambda_1, H_a: \lambda_0 < \lambda_1 \\ \frac{\chi_{2d;\beta}}{2\lambda_1} \geq n \geq \frac{\chi_{2d;1-\alpha}}{2\lambda_0}, & H_0: \lambda_0 \leq \lambda_1, H_a: \lambda_0 > \lambda_1 \end{cases}, \; d = 1,2,3 \ldots \qquad \{4.7.2\}$$

where $\chi_{2d;p}$ is the inverse of the cumulative distribution of the chi-squared distribution evaluated at $2d$ degrees of freedom and probability p.

In this formula, the value of d is increased until the inequality above is fulfilled and the interval between the left hand side expression and right hand side expression contains at least one integer.

The power is calculated by finding X_{crit} as outlined above for the specified value of λ_0. This value of X_{crit} is used to calculate power as follows:

$$f(1-\beta) = \begin{cases} Power = 1 - \sum_{-\infty}^{x=X_{crit}} e^{-n\lambda_1} \frac{(n\lambda_1)^x}{x!}, & H_0: \lambda_0 \geq \lambda_1, H_a: \lambda_0 < \lambda_1 \\ Power = \sum_{-\infty}^{x=X_{crit}} e^{-n\lambda_1} \frac{(n\lambda_1)^x}{x!}, & H_0: \lambda_0 \leq \lambda_1, H_a: \lambda_0 > \lambda_1 \end{cases} \quad \{4.7.3\}$$

A closed form equation is not used to calculate the rate ratio. Instead a search algorithm is used. Firstly, X_{crit} is calculated as outlined above for the specified value of λ_0. The search algorithm then calculates power at various values of λ_1 until the desired power is reached.

4.7.3 Examples

Example 1: Validation example calculating required sample size for a given power
The following example is taken from Question 1.31 from page 29 of Guenther (1977) where a sample size calculation problem is followed by a sensitivity problem for the Alternative Hypothesis Rate variable.

The following steps outline the procedure for this example:

1. Open nQuery + nTerim 4.0 via the start menu or desktop shortcut. Click “New Fixed Term Test” from the top of the window. The “Study Goal and Design” window will appear as Figure 4.7.1 below.

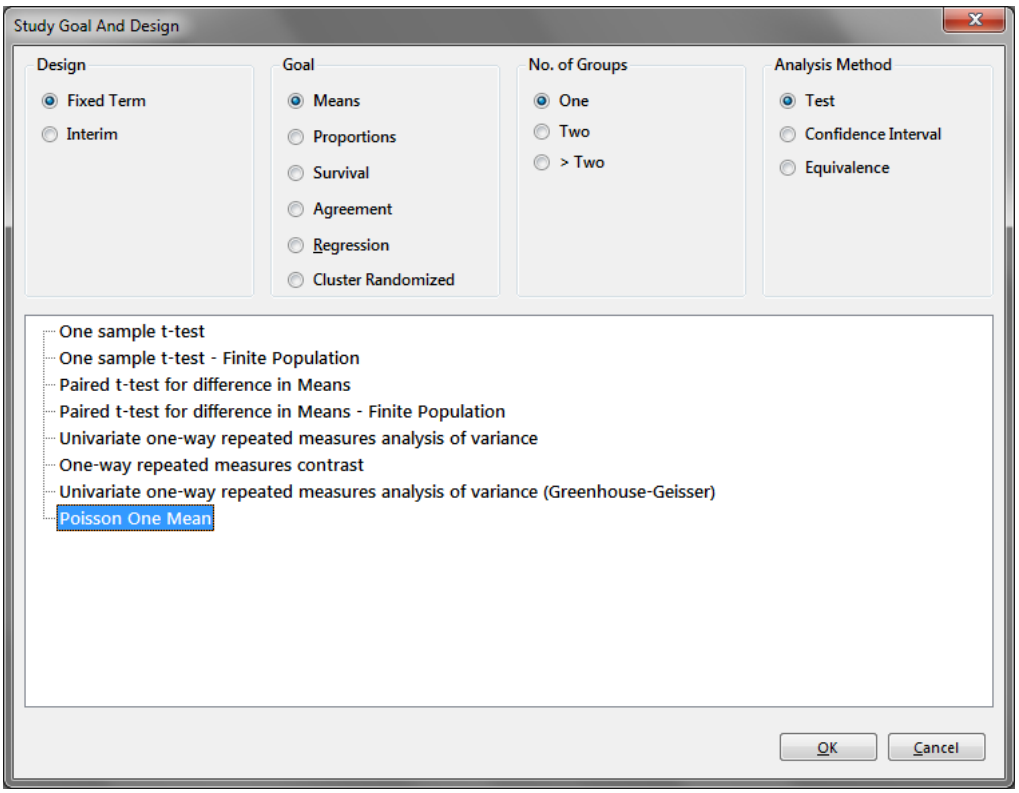


Figure 4.7.1. Study Goal and Design Window

2. In order to select the Poisson One Mean design table, navigate through the Study Goal and Design Window by selecting "Fixed Term" as the Design, "Means" as the Goal, "One" as the Number of Groups and "Test" as the Analysis Method. Then click "OK" and the test window will appear as per Figure 4.7.2.

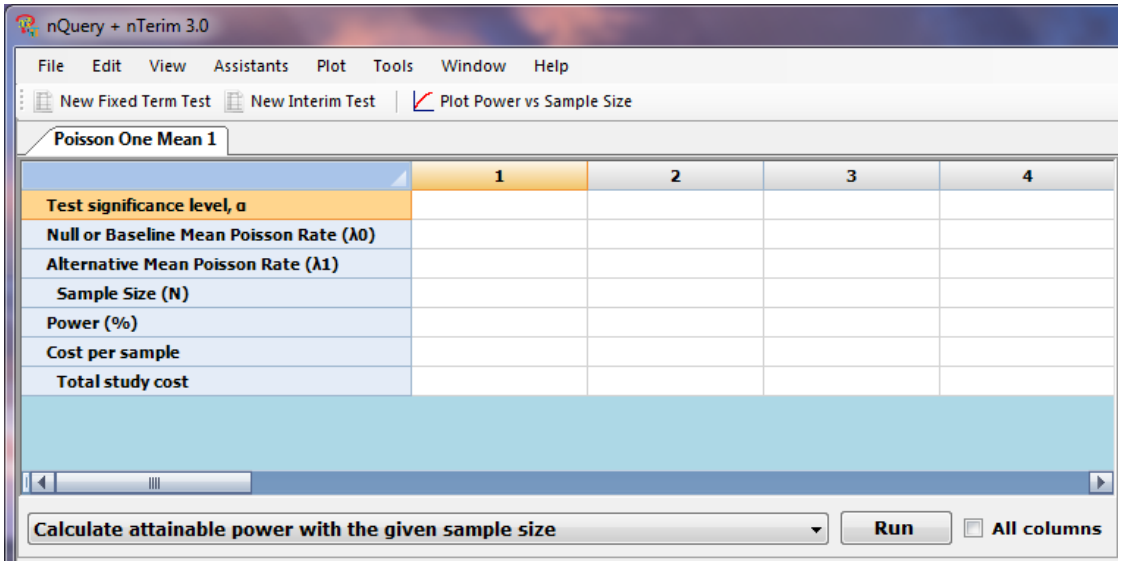


Figure 4.7.2. Poisson One Mean Test Table

The first calculation will be for Sample Size, to begin we enter the values as follows.

3. First enter 0.05 for the Test Significance level row. Next, enter 0.03 and 0.1 for the "Null or Baseline Mean Poisson Rate" and "Alternative Mean Poisson Rate" variable rows respectively. Finally, enter 90 in the Power row. The table will appear as per Figure 4.7.3.

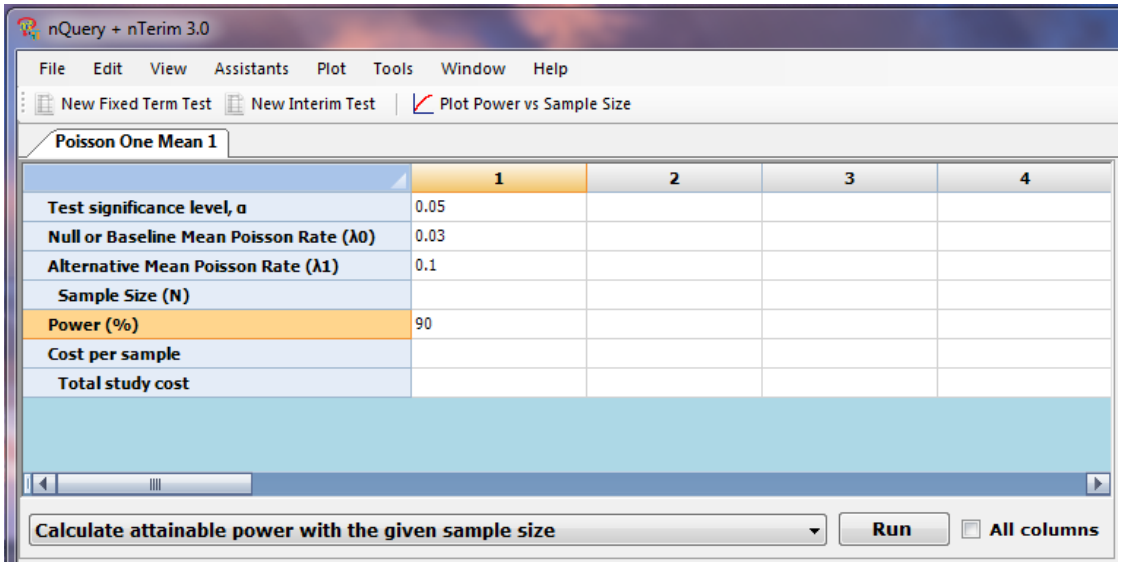


Figure 4.7.3. Values entered for Poisson One Mean study design

4. Select "Calculate required sample size for the given power" from the dropdown menu beside the "Run" button. Then click "Run".

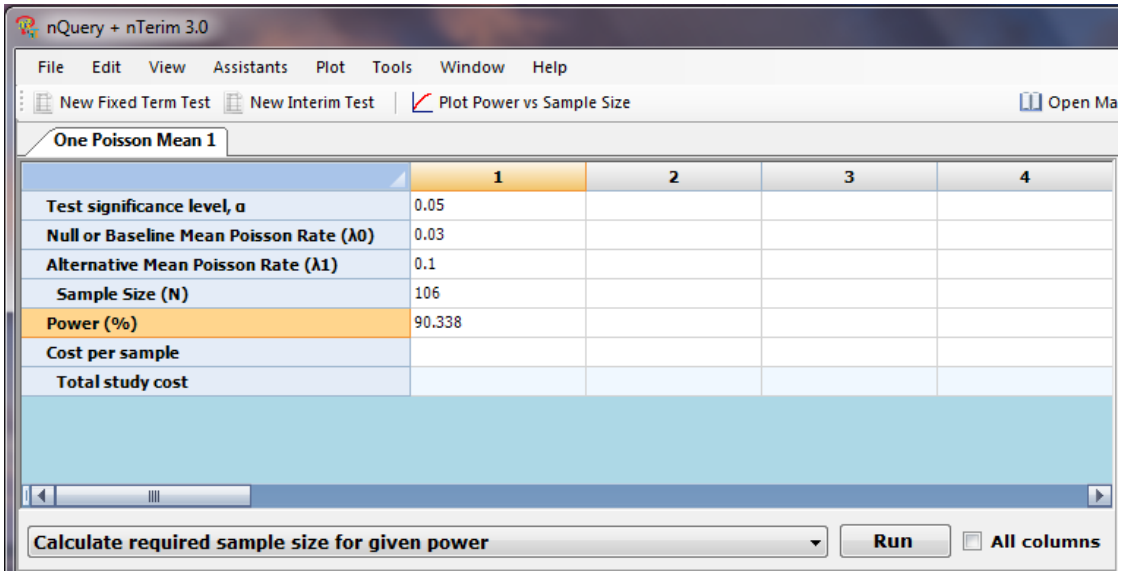


Figure 4.7.4.Completed Poisson One Mean study design

This will give a result of 106 as displayed in Figure 4.7.4 above.

Example 2: Validation example calculating required power for a given sample size
The next calculation is a sensitivity analysis for Power where the Alternative Mean Poisson Rate is varied while the sample size is fixed at 106.

5. To do this, copy the same values across to column 2 and 3. Then change the value for Alternative Mean Poisson Rate to 0.05 and 0.08 for column 2 and 3 respectively. This will give a table as per Figure 4.7.5.

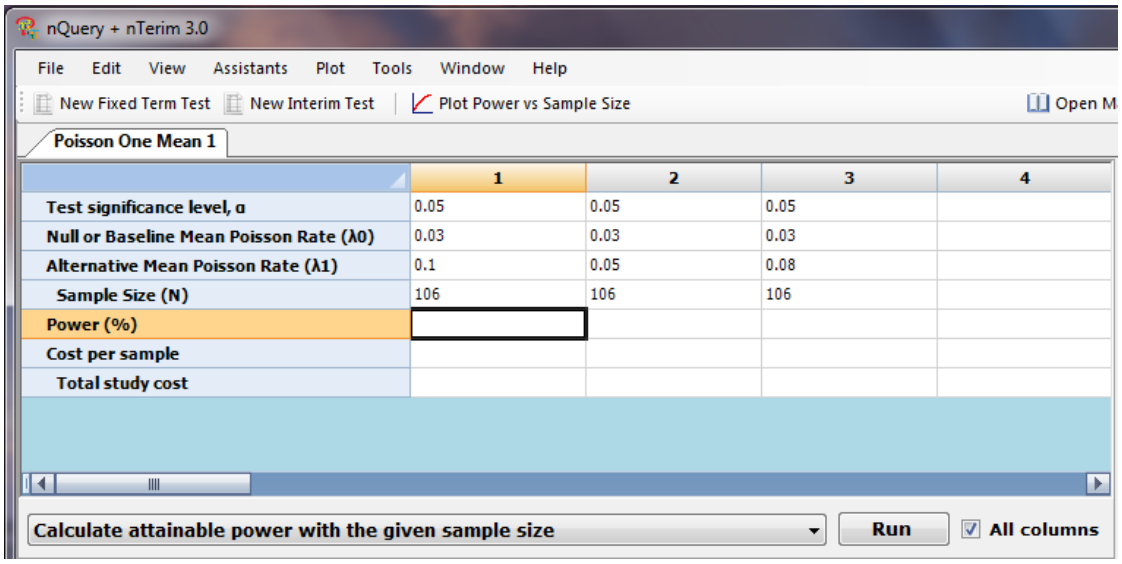
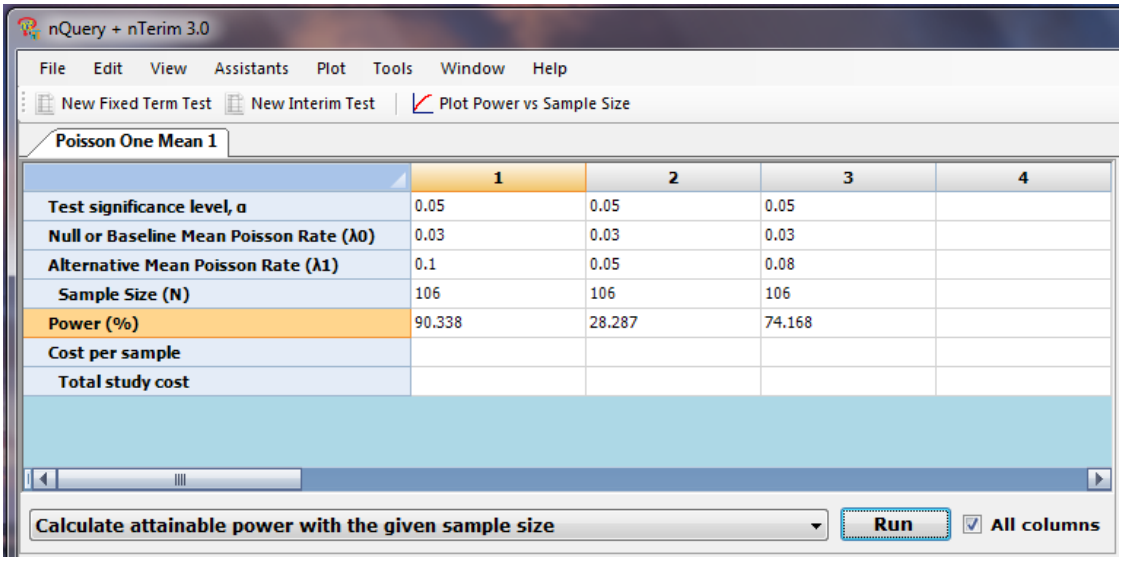


Figure 4.7.5.Sensitivity analysis around the Alternative Mean Poisson rate

6. Select “Calculate attainable power for the given sample size” from the dropdown menu beside the “Run” button. Then tick the box to run “All Columns”. Then click “Run”. This will give the resultant powers of 90.338, 28.287 and 74.168 for columns 1, 2 and 3 respectively as in Figure 4.7.6.



	1	2	3	4
Test significance level, α	0.05	0.05	0.05	
Null or Baseline Mean Poisson Rate (λ0)	0.03	0.03	0.03	
Alternative Mean Poisson Rate (λ1)	0.1	0.05	0.08	
Sample Size (N)	106	106	106	
Power (%)	90.338	28.287	74.168	
Cost per sample				
Total study cost				

Figure 4.7.6.Results from Sensitivity analysis

The effect of changing Alternative Mean Poisson Rate on the relationship between Power and Sample size can be explored further using the “Plot Power vs Sample Size” button at the top of the screen.

7. Select all three columns by clicking the 1 at the top of column 1. Then hold down Shift and click the 3 above columns. All three columns will now be highlighted.

8. Click the “Plot Power vs Sample Size” button at the top of the screen. This will give you a plot as displayed in Figure 4.7.7 which will show the relationship between power and sample size for each column. Right-click to add feature such as a legend to the graph and double click elements for user options and editing.

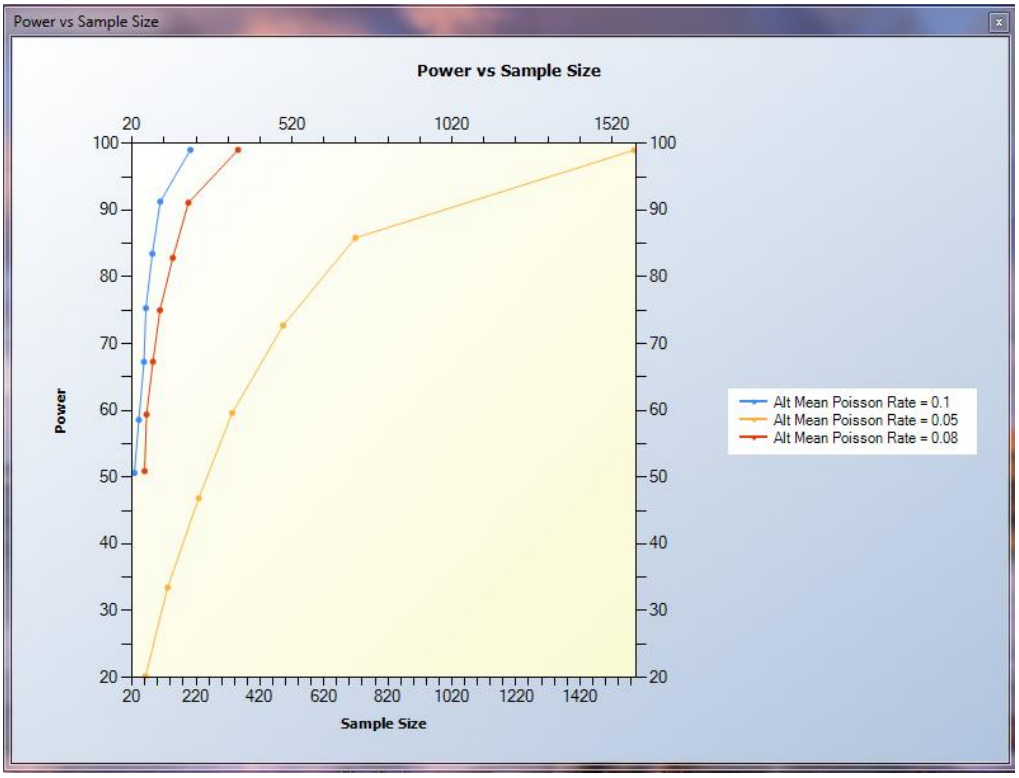


Figure 4.7.7. Power vs. Sample Size plot

9. Finally, by clicking on the desired study design column and going to the "Output" tab at the bottom of the screen, you can get an output statement for the results given by the current calculation as per Figure 4.7.8.

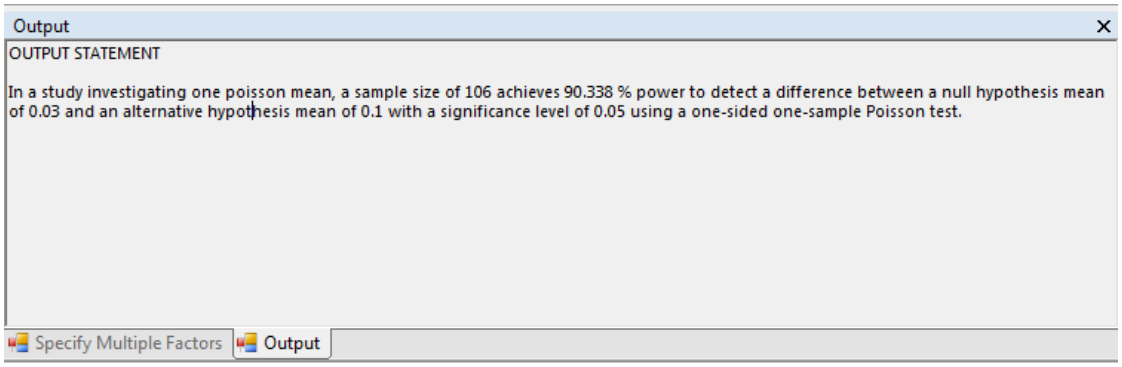


Figure 4.7.8. Study design Output statement

This Output statement can then be easily transferred directly from the output window into a report.

4.8. Two Poisson Means

4.8.1 Introduction

Count data is often obtained in a variety of clinical and commercial activities such as the number of accidents at a junction or number of occurrences of a disease in a year. The most common distribution used to model count data is the Poisson distribution. The two sample test is used to test hypotheses about the difference between two Poisson distributed samples.

This table facilitates the calculation of the power and sample size for hypothesis tests of the mean of two Poisson distributed samples. Power and sample size is computed using the method outlined by Gu et al. (2008)

4.8.2 Methodology

For this study design, the fixed time intervals (t_1, t_2), the sample sizes (N_1, N_2), the number of events that occurred (X_1, X_2) and the mean Poisson event rates (γ_1, γ_2) are used to specify the distribution of each sample.

As suggested by Gu et al. (2008), there are five test statistics which could be used to test statistical hypotheses comparing two Poisson mean rates.

The statistics proposed tested the null hypothesis of $H_0: \left(\frac{\gamma_2}{\gamma_1}\right) = R_0$ (usually R_0 = 1) against the specified alternative hypothesis of $H_1: \left(\frac{\gamma_2}{\gamma_1}\right) = R_1 > R_0$. These are given by the following equations:

1. Unconstrained Maximum Likelihood Estimate (MLE)
   The unconstrained maximum likelihood estimator is given by $\hat{\gamma}_i = X_i/t_i\,, i = 1, 2$. The null hypothesis can be re-arranged as follows to yield a new statistic:

   $$H_0: W = \gamma_1 - R_0\gamma_2 \qquad \{4.8.1\}$$

   The statistic W is asymptotically normally distributed with a mean of $\mu_W = \hat{\gamma}_1 - R\hat{\gamma}_2$ and variance of $\sigma_W^2 = \lambda_1/t_1 + R^2\lambda_2/t_2$. Thus the test statistic of μ_W/σ_W^2 can be used for tests of H_0.

   For the unconstrained maximum likelihood estimate this yields the following formulation:

   $$W_{MLE}(X_1, X_2) = \frac{X_1/t_1 - RX_2/t_2}{\sqrt{X_1/t_1^2 + X_2R^2/t_2^2}} = \frac{X_1 - X_2\rho}{\sqrt{X_1 + X_2\rho^2}} \qquad \{4.8.2\}$$

   where $\rho = Rt_1N_1/t_2N_2$.

2. Constrained Maximum Likelihood Estimate (CMLE)
   The constrained maximum likelihood estimators are given by the following:

   $$\tilde{\gamma}_1 = \frac{X_1 + X_2}{t_1(1 + 1/\rho)}, \tilde{\gamma}_2 = \frac{X_1 + X_2}{t_2(1 + \rho)} \qquad \{4.8.3\}$$

   Similar to the unconstrained maximum likelihood estimator, these can be used to generate the following statistic for hypothesis testing:

   $$W_{CMLE}(X_1, X_2) = \frac{X_1/t_1 - RX_2/t_2}{\sqrt{(X_1 + X_2)(R^2/d + R)/(R + d)}} = \frac{X_1 - X_2\rho}{\sqrt{X_1 + X_2\rho}} \qquad \{4.8.4\}$$

   where $d = t_2/t_1$.

3. Log Unconstrained Maximum Likelihood Estimate (ln(MLE))
   Testing the natural log for the unconstrained MLE can yield a null hypothesis of $H_0: U = \ln(\gamma_1/\gamma_2) - \ln R$. The variance of U is given by $\sigma_U^2 = 1/t_1\gamma_1 + 1/t_2\gamma_2$. Similar re-arrangement as for the MLE above yields the following test statistic:

$$W_{\ln(MLE)}(X_1, X_2) = \frac{\ln(X_1/X_2) + \ln d - \ln R}{\sqrt{1/X_1 + 1/X_2}} = \frac{\ln(X_1/X_2) - \ln \rho}{\sqrt{1/X_1 + 1/X_2}} \qquad \{4.8.5\}$$

4. Log Constrained Maximum Likelihood Estimate (ln(CMLE))
   Similar arguments for the above equation can be used to yield the following test statistic for the log of the constrained estimators:

$$W_{\ln(CMLE)}(X_1, X_2) = \frac{\ln(X_1/X_2) + \ln d - \ln R}{\sqrt{(2 + d/R + R/d)/(X_1 + X_2)}}$$
$$W_{\ln(CMLE)}(X_1, X_2) = \frac{\ln(X_1/X_2) - \ln \rho}{\sqrt{(2 + 1/\rho + \rho)/(X_1 + X_2)}} \qquad \{4.8.6\}$$

5. Variance Stabilized Estimate
   The variance stabilizing transformation of Huffman (1984) can accelerate the convergence to normality. This transformation uses the following test statistic:

$$W_{VS}(X_1, X_2) = \frac{2\left(\sqrt{X_1 + 3/8} - \sqrt{\rho(X_2 + 3/8)}\right)}{\sqrt{1 - \rho}} \qquad \{4.8.7\}$$

The sample size calculation for two Poisson rates is taken from Gu et al. (2008). This table can be used to calculate the power, the sample size and the rate ratio under the alternative hypothesis given all other terms in the table are specified.

To calculate power and sample size, the user must specify the test significance level (α), the ratio of the two samples Poisson rates under the null hypothesis (R_0), the ratio of the two samples Poisson rates under the alternative hypothesis (R_1), the fixed observation times for group 1 (t_1) and group 2 (t_1) , the mean Poisson rate in group 1 (γ_1), one of the group sample sizes (N_1 or N_2) or the sample size ratio (N_2/N_1) and the test statistic that is being used to test the difference in ratios.

The formulas for the power which correspond to each of the statistics outlined above are as follows:

1. Unconstrained Maximum Likelihood Estimate (MLE)

$$Power_{MLE} = \Phi\left(\frac{\sqrt{N_1\gamma_1 t_1}(1-c)}{\sqrt{c/\rho + c^2}} - z_{1-\alpha}\right) \qquad \{4.8.8\}$$

2. Constrained Maximum Likelihood Estimate (CMLE)

$$Power_{CMLE} = \Phi\left(\frac{\sqrt{N_1\gamma_1 t_1}(1-c)}{\sqrt{c/\rho + c^2}} - z_{1-\alpha}\sqrt{(c+\rho)/(1+c\rho)}\right) \qquad \{4.8.9\}$$

3. Log Unconstrained Maximum Likelihood Estimate (ln(MLE))

$$Power_{\ln(MLE)} = \Phi\left(\frac{\sqrt{N_1\gamma_1 t_1}|\ln c\,|}{\sqrt{c/\rho + 1}} - z_{1-\alpha}\right) \qquad \{4.8.10\}$$

4. Log Constrained Maximum Likelihood Estimate (ln(CMLE))

$$Power_{\ln(CMLE)} = \Phi\left(\frac{\sqrt{N_1\gamma_1 t_1}|\ln c\,|}{\sqrt{c/\rho + 1}} - z_{1-\alpha}\left[\frac{\sqrt{c(\rho^2+2\rho+1)}}{c+\rho}\right]\right) \qquad \{4.8.11\}$$

5. Variance Stabilized Estimate

$$Power_{VS} = \Phi\left(\frac{2\left(1-\sqrt{c}\right)\sqrt{N_1\gamma_1 t_1 + \frac{3}{8}} - z_{1-\alpha}\sqrt{c/\rho + c}}{\sqrt{c/\rho + 1}}\right) \qquad \{4.8.12\}$$

where $c = R_0/R_1$ and Φ is the inverse normal cumulative distribution.

To calculate the power when the test for the null hypothesis is $H_0: \left(\frac{\gamma_2}{\gamma_1}\right) = R_0$ versus the specified alternative hypothesis of $H_1: \left(\frac{\gamma_2}{\gamma_1}\right) = R_1 < R_0$, the power calculation is the same except z_α is used instead of $z_{1-\alpha}$ and the power is equal to one minus the normal cumulative distribution. For example, the power formula for the MLE becomes the following:

$$Power_{MLE} = 1 - \Phi\left(\frac{\sqrt{N_1\gamma_1 t_1}(1-c)}{\sqrt{c/\rho + c^2}} - z_\alpha\right) \qquad \{4.8.13\}$$

A simulation study by Gu et al. (2008) indicated that the variance stabilized statistic had the most reliable performance.

A closed form equation is not used to calculate the sample size or rate ratio. Instead a search algorithm is used. The search algorithm calculates power at various values for the sample size or rate ratio until the desired power is reached.

4.8.3 Examples

Example 1: Validation example calculating required sample size for a given power
The following example is taken from Table 6 of Gu et al (2008) where a sample size calculation problem is conducted comparing the 5 test statistics.

The following steps outline the procedure for this example:

1. Open nQuery + nTerim 4.0 via the start menu or desktop shortcut. Click “New Fixed Term Test” from the top of the window. The “Study Goal and Design” window will appear as Figure 4.8.1 below.

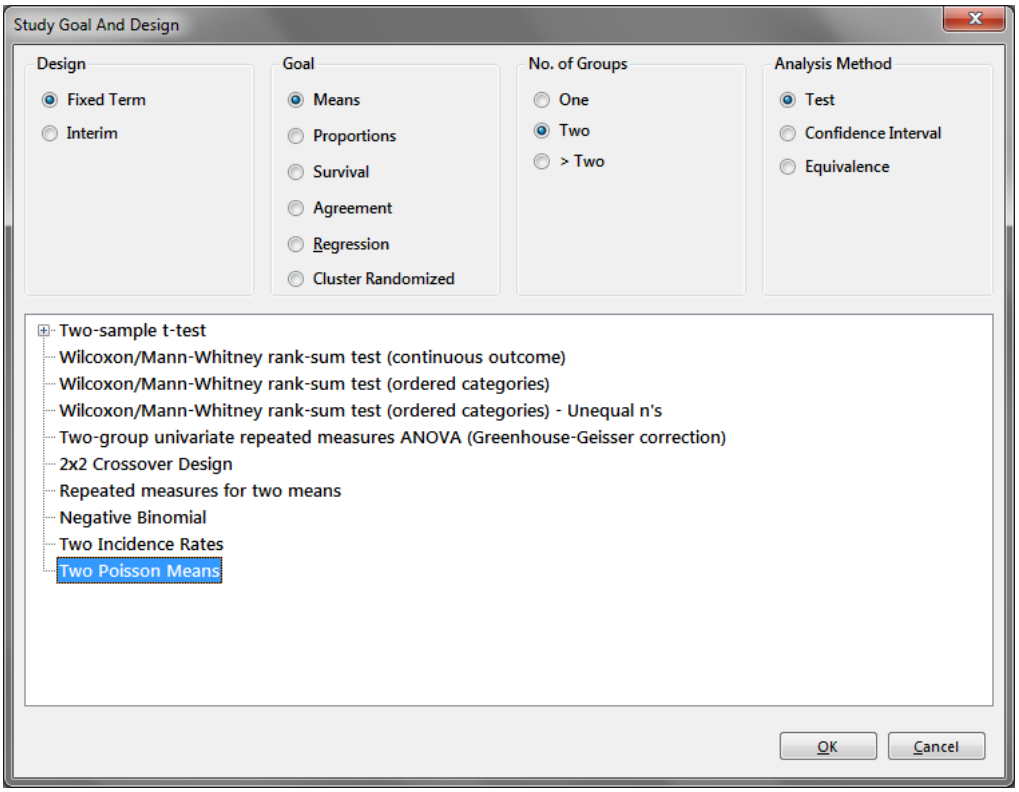


Figure 4.8.1. Study Goal and Design Window

2. In order to select the Two Poisson Means design table, navigate through the Study Goal and Design Window by selecting "Fixed Term" as the Design, "Means" as the Goal, "Two" as the Number of Groups and "Test" as the Analysis Method. Then click "OK" and the test window will appear as per Figure 4.8.2.

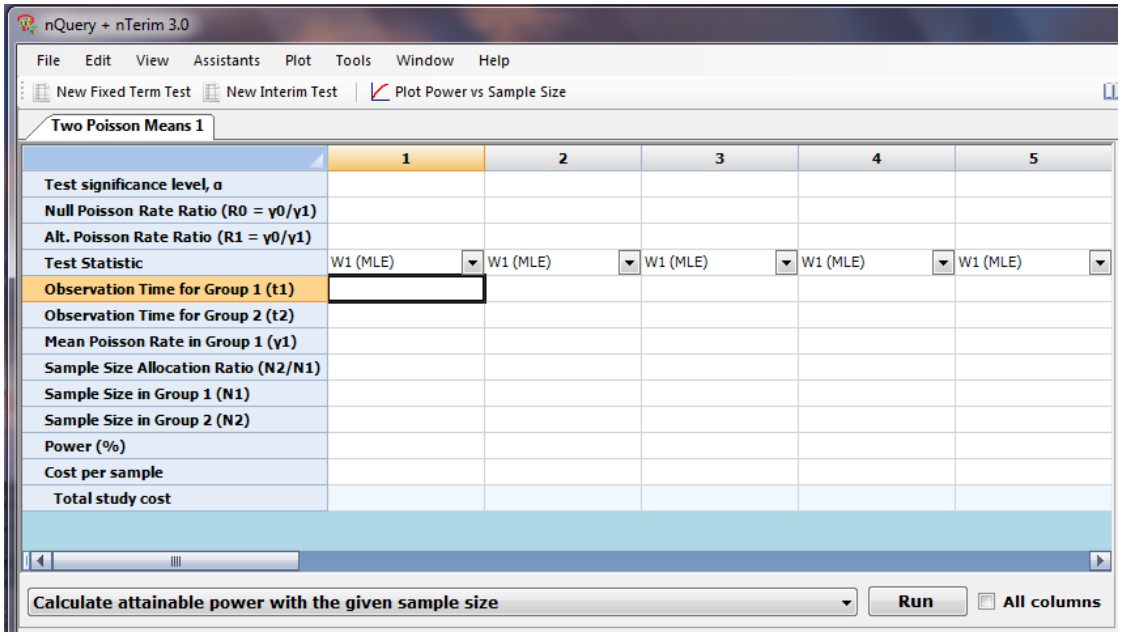


Figure 4.8.2. Two Poisson Means Test Table

The first calculation will be for Sample Size using the "W1 (MLE)" test statistic, to begin we enter the values as follows.

3. First enter 0.05 for the Test Significance level row, then enter 1 and 4 for the "Null Poisson Rate Ratio" and "Alternative Poisson Rate Ratio" variable rows respectively.

4. Next, enter two for both Observation time variables (t1, t2), 0.0005 for the Mean Poisson Rate in Group 1 and a Sample size allocation ratio of 0.5 and finally, enter 90 in the Power row. The table will appear as per Figure 4.8.3.

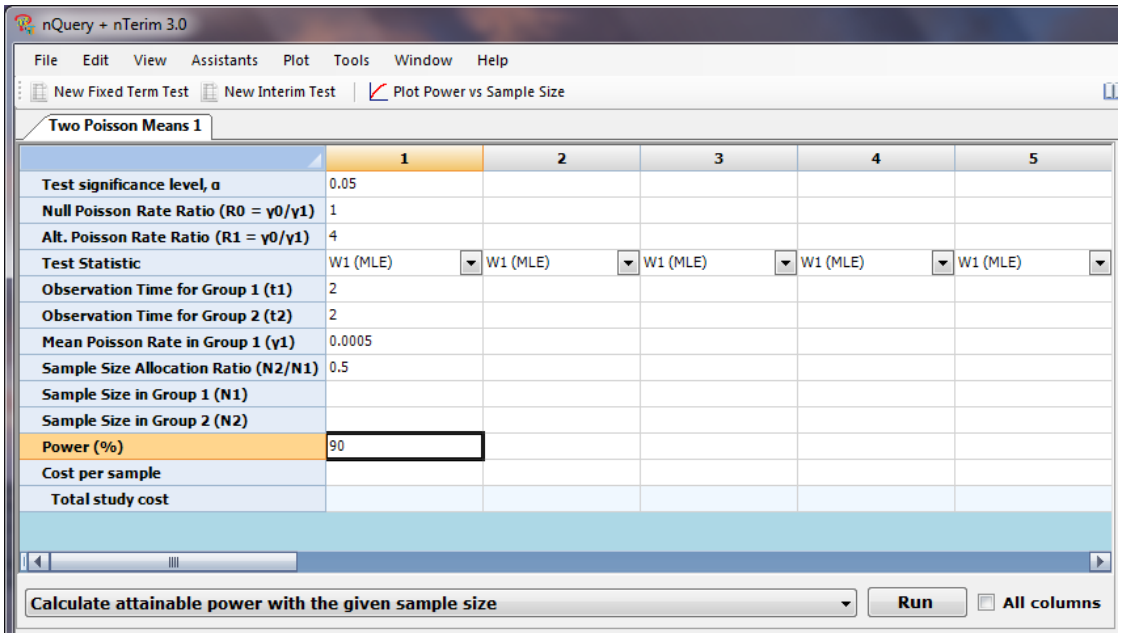


Figure 4.8.3. Values entered for Two Poisson Mean study design

5. Select “Calculate required Group 1 and 2 sample sizes for given power and sample size allocation” from the dropdown menu beside the “Run” button. Then click “Run”.

This will give a result of 8564 and 4282 for the Group 1 and 2 sample sizes respectively as in Figure 4.8.4. The result in Gu et al (2008) is 8527 for the sample size in Group 1 but the authors agreed this was the more accurate result due to their usage of two decimal place rounding.

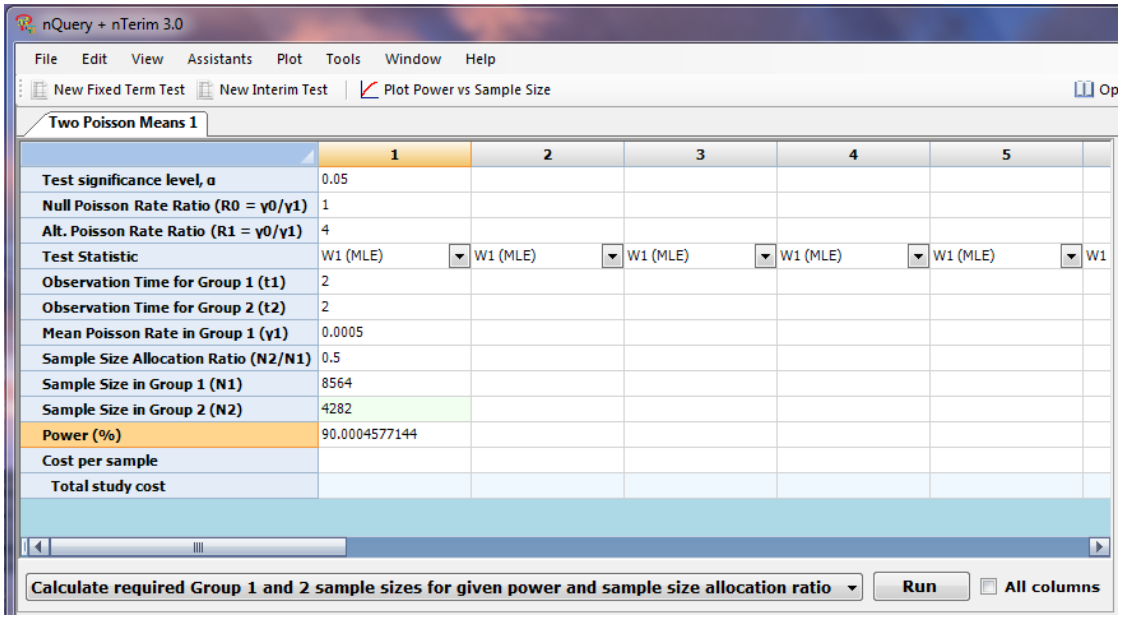
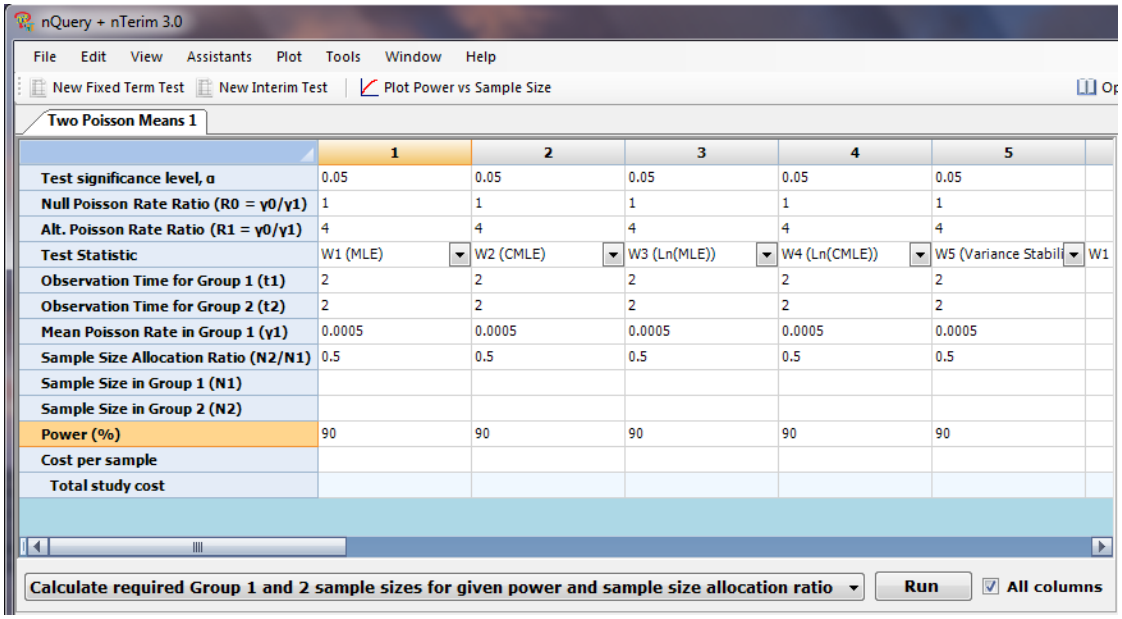


Figure 4.8.4. Completed Two Poisson Means study design

6. The next calculation is a sensitivity analysis for sample size where we change the Test Statistic to investigate the impact this has on the sample size estimate. To do this, replace the updated power with 90 and then copy the same values across to columns 2 to 5. Then, select from the Test Statistic dropdown menu W2 (CMLE) for column 2, W3 (ln(MLE)) for column 3, W4 (ln(CMLE)) for column 4 and W5 (Variance Stabilizing) for column 5. This will give a table as per Figure 4.8.5.

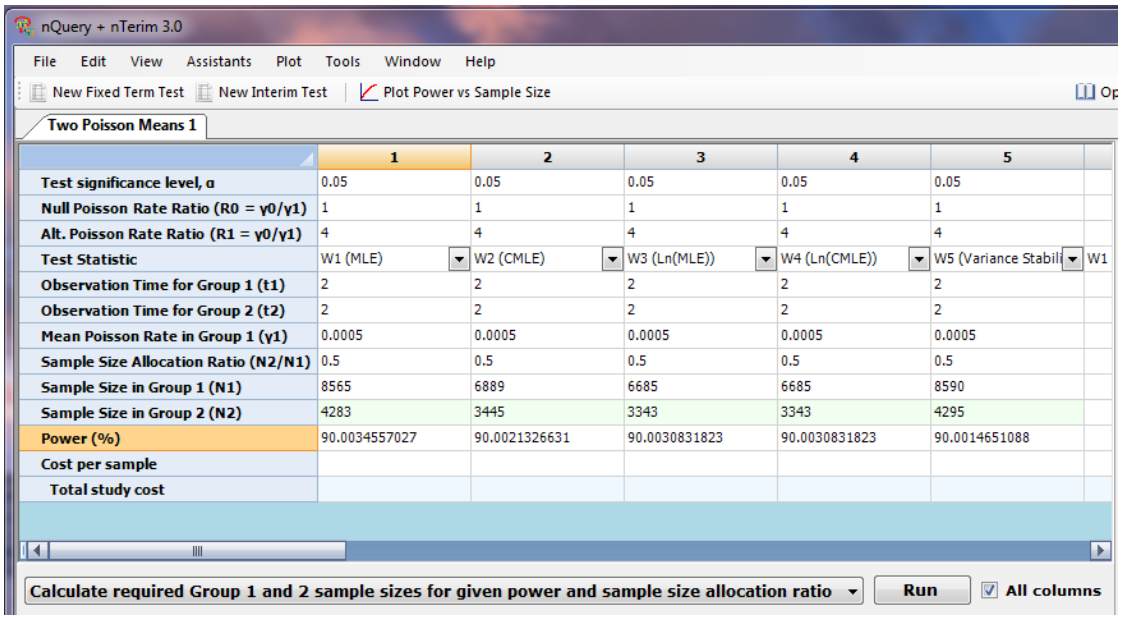


	1	2	3	4	5
Test significance level, α	0.05	0.05	0.05	0.05	0.05
Null Poisson Rate Ratio (R0 = γ0/γ1)	1	1	1	1	1
Alt. Poisson Rate Ratio (R1 = γ0/γ1)	4	4	4	4	4
Test Statistic	W1 (MLE)	W2 (CMLE)	W3 (Ln(MLE))	W4 (Ln(CMLE))	W5 (Variance Stabili
Observation Time for Group 1 (t1)	2	2	2	2	2
Observation Time for Group 2 (t2)	2	2	2	2	2
Mean Poisson Rate in Group 1 (γ1)	0.0005	0.0005	0.0005	0.0005	0.0005
Sample Size Allocation Ratio (N2/N1)	0.5	0.5	0.5	0.5	0.5
Sample Size in Group 1 (N1)					
Sample Size in Group 2 (N2)					
Power (%)	90	90	90	90	90
Cost per sample					
Total study cost					

Figure 4.8.5. Sensitivity analysis around the Test Statistic options

7. Select "Calculate required Group 1 and 2 sample sizes for given power and sample size allocation" from the dropdown menu beside the "Run" button. Then tick the box to run "All Columns". Then click "Run".

This will give the resultant Group 1 sample sizes of 8564 (W1), 6889 (W2), 6685 (W3), 6685 (W4) and 8590 (W5) for each statistic sequentially as in Figure 4.8.6. Similar to the W1 example above, these answers differ due to rounding.



	1	2	3	4	5
Test significance level, α	0.05	0.05	0.05	0.05	0.05
Null Poisson Rate Ratio (R0 = γ0/γ1)	1	1	1	1	1
Alt. Poisson Rate Ratio (R1 = γ0/γ1)	4	4	4	4	4
Test Statistic	W1 (MLE)	W2 (CMLE)	W3 (Ln(MLE))	W4 (Ln(CMLE))	W5 (Variance Stabili
Observation Time for Group 1 (t1)	2	2	2	2	2
Observation Time for Group 2 (t2)	2	2	2	2	2
Mean Poisson Rate in Group 1 (γ1)	0.0005	0.0005	0.0005	0.0005	0.0005
Sample Size Allocation Ratio (N2/N1)	0.5	0.5	0.5	0.5	0.5
Sample Size in Group 1 (N1)	8565	6889	6685	6685	8590
Sample Size in Group 2 (N2)	4283	3445	3343	3343	4295
Power (%)	90.0034557027	90.0021326631	90.0030831823	90.0030831823	90.0014651088
Cost per sample					
Total study cost					

Figure 4.8.6. Results from Sensitivity analysis

The effect of changing the test statistic on the relationship between Power and Sample size can be explored further using the "Plot Power vs Sample Size" button at the top of the screen. For this particular calculation, W3 and W4 give the same answer and thus only W3 was plotted in this example.

8. Select the first column by clicking the 1 at the top of column 1. Then hold down Ctrl and click the 2 at the top of column 2, the 3 at the top of column 3 and the 5 at the top of column 5. All four columns will now be highlighted.

9. Click the "Plot Power vs Sample Size" button at the top of the screen. This will give you a plot as displayed in Figure 4.8.7 which will show the relationship between power and sample size for each column. Right-click to add feature such as a legend to the graph and double click elements for user options and editing.

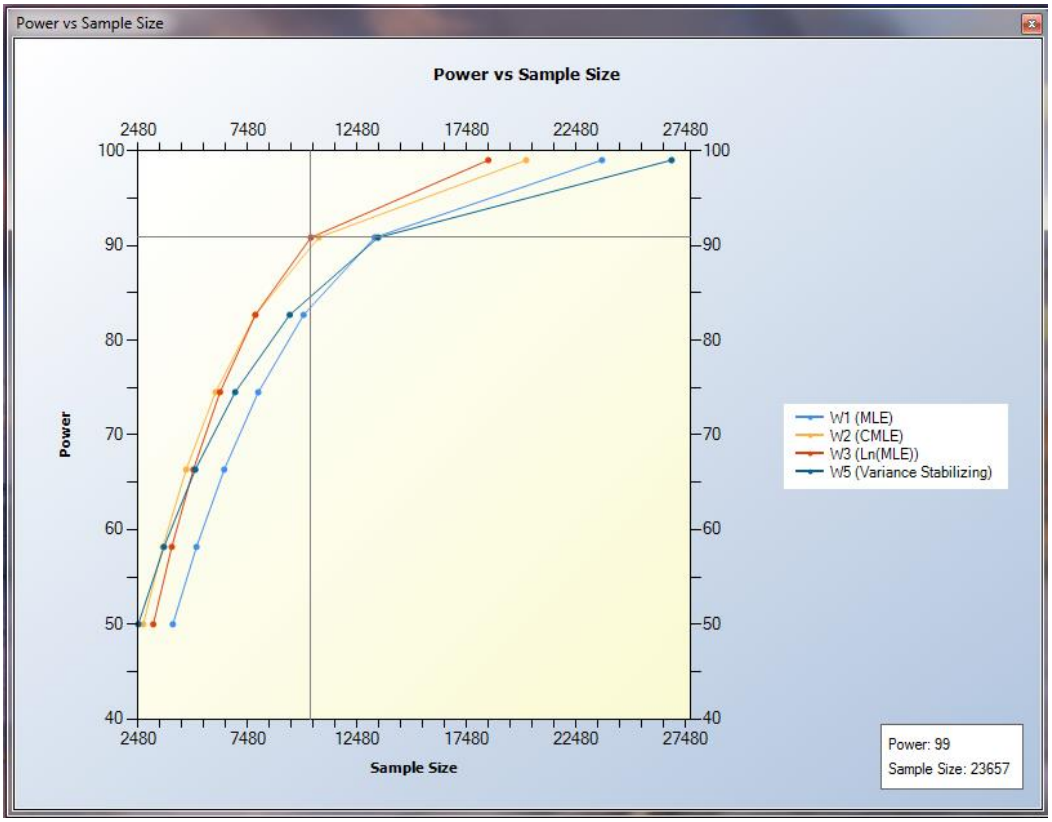


Figure 4.8.7. Power vs. Sample Size plot

10. Finally, by clicking on the desired study design column and going to the "Output" tab at the bottom of the screen, you can get an output statement for the results given by the current calculation as per Figure 4.8.8.

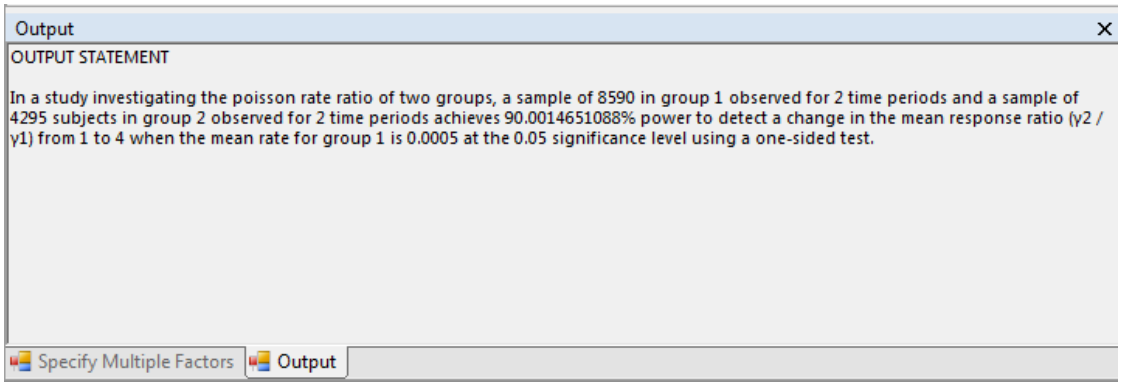


Figure 4.8.8. Study design Output statement

4.9. Two Negative Binomial Rates

4.9.1 Introduction

The Negative Binomial Model has been increasingly used to model count data in clinical trials. This technique allows for the modelling of heterogeneity in count data, has greater flexibility in modelling the mean-variance relationship over the Poisson model and is expected to be less biased by non-missing at random data. This flexibility and robustness has contributed to its increased use in the literature.

This table facilitates the calculation of the power and sample size for the difference between two negative binomial rates. Power and sample size is computed using the method outlined by Zhu and Lakkis (2014).

4.9.2 Methodology

Let y_{ij} equal the number of events in time t_{ij} for subject i in group j (treatment and control/placebo groups). Assuming y_{ij} follows a negative binomial distribution with mean μ_{ij} and dispersion parameter k, the mean for a negative binomial regression can be modelled as:

$$\log[\mu_{ij}] = \log t_{ij} + \beta_0 + \beta_1 X_{ij} \qquad \{4.9.1\}$$

where x_{ij} is 0 for subjects in Group 0 (control/placebo) and equal to 1 for subjects in Group 1 (treatment). Let r_0 and r_1 be the mean rate of events per unit time in Group 0 and 1 respectively. Then $r_0 = e^{\beta_0}$, $r_1 = e^{\beta_0+\beta_1}$ and thus $\frac{r_1}{r_0} = e^{\beta_1}$.

The sample size calculation for two negative binomial rates is taken from Zhu and Lakkis (2014). This table can be used to calculate the power, the sample size and the rate ratio given all other terms in the table are specified.

Calculations use a standard normal approximation as β_1 follows the normal distribution asymptotically, with a mean of $\log\left(\frac{r_1}{r_0}\right)$ and a variance equal to $\frac{1}{n_0}\left[\frac{1}{\mu_t}\left(\frac{1}{r_0}+\frac{1}{\theta r_1}\right)+\frac{(1+\theta)k}{\theta}\right]$. To calculate power and sample size, the user must specify the test significance level (α), the mean rate of the event for the control group (r_0), the ratio between the control group and treatment group mean event rates $\left(\frac{r_1}{r_0}\right)$, the average exposure time for each subject (μ_t), the common dispersion parameter of the two groups (k), the specification of how the null hypothesis variance is calculated ("Rates Variance" in table) and the sample size allocation ratio $\left(\theta = \frac{n_1}{n_0}\right)$. It is assumed that the dispersion parameter is the same for both groups and guidance on how to estimate the dispersion parameter is outlined in Zhu and Lakkis (2014).

The formula for the sample size for the null hypothesis of $\left(\frac{r_1}{r_0}\right) = 1$ (equivalent to $\beta_1 = 0$) versus the alternative hypothesis of $\left(\frac{r_1}{r_0}\right) \neq 1$ (equivalent to $\beta_1 \neq 0$) is given by the following equation:

$$n_0 \geq \frac{\left(z_{1-\frac{\alpha}{2}}\sqrt{V_0} + z_{1-\beta}\sqrt{V_1}\right)^2}{\left(\log\left(\frac{r_1}{r_0}\right)\right)^2}, \qquad n_1 = \theta n_0 \qquad \{4.9.2\}$$

where n_0 and n_1 are the sample sizes of group 0 and 1 respectively. V_0 and V_1 are the estimated variances of the normal approximation of β_1 under the null and alternative hypotheses respectively. These are defined below.

The power is calculated by re-arrangement of the above formula to give the following equation:

$$Power = \Phi\left(\frac{\sqrt{n_0}\left|\log\left(\frac{r_1}{r_0}\right)\right| - z_{1-\frac{\alpha}{2}}\sqrt{V_0}}{\sqrt{V_1}}\right) \qquad \{4.9.3\}$$

In the above equations V_0 and V_1 are defined as follows:

$$V_0 = \frac{1}{\mu_t}\left(\frac{1}{\tilde{r}_0} + \frac{1}{\theta\tilde{r}_1}\right) + \frac{(1+\theta)k}{\theta} \qquad \{4.9.4\}$$

$$V_1 = \frac{1}{\mu_t}\left(\frac{1}{r_0} + \frac{1}{\theta r_1}\right) + \frac{(1+\theta)k}{\theta} \qquad \{4.9.5\}$$

where $\tilde{r}_0$ and $\tilde{r}_1$ are the rates under the null hypothesis and r_1 and r_0 are the assumed true rates specified in the table i.e. their values under the alternative hypothesis. The values of $\tilde{r}_0$ and $\tilde{r}_1$, and by extension V_0, can be specified in three ways.

These are selected using "Rates Variance" option in the table. These are detailed below:

1. Reference Rate
   Under H_0, $r_0 = r_1$. Thus can set $\tilde{r}_0 = \tilde{r}_1 = r_0$ which gives the following:

$$V_{0(RR)} = \frac{(1+\theta)}{\theta\mu_t r_0} + \frac{(1+\theta)k}{\theta} \qquad \{4.9.6\}$$

2. True Rates
   The values of $\tilde{r}_0$ and $\tilde{r}_1$are set to their true values under the alternative hypothesis. This gives the following:

$$V_{0(TR)} = \frac{1}{\mu_t}\left(\frac{1}{r_0} + \frac{1}{\theta r_1}\right) + \frac{(1+\theta)k}{\theta} \qquad \{4.9.7\}$$

3. Maximum Likelihood
   Maximising the log-likelihood function under the null hypothesis restriction yields the following maximum likelihood estimate of the overall event rate:

$$\tilde{\tilde{r}} = \frac{n_0\mu_t r_0 + n_1\mu_t r_1}{n_0\mu_t + n_1\mu_t} = \frac{r_0 + \theta r_1}{1+\theta} \qquad \{4.9.8\}$$

   Setting $\tilde{r}_0 = \tilde{r}_1 = \tilde{\tilde{r}}$ gives the following variance formulation:

$$V_{0(ML)} = \frac{(1+\theta)^2}{\theta\mu_t(r_0 + \theta r_1)} + \frac{(1+\theta)k}{\theta} \qquad \{4.9.9\}$$

In Zhu and Lakkis (2014), simulation indicated the true rates and maximum likelihood methods gave the best estimates of the sample size for Wald and likelihood ratio tests for β_1.

A closed form equation is not used to calculate the rate ratio. Instead a search algorithm is used. This search algorithm calculates power at various values of the rate ratio until the desired power is reached.

4.9.3 Examples

Example 1: Validation example calculating required sample size for a given power
The following example is taken from Table I of Zhu and Lakkis (2014) where a sample size calculation problem is conducted to show the effect of changing the mean rate for the control group and the method for calculating the variance under the null hypothesis.

The following steps outline the procedure for this example:

1. Open nQuery + nTerim 4.0 via the start menu or desktop shortcut. Click “New Fixed Term Test” from the top of the window. The “Study Goal and Design” window will appear as Figure 4.9.1 below.

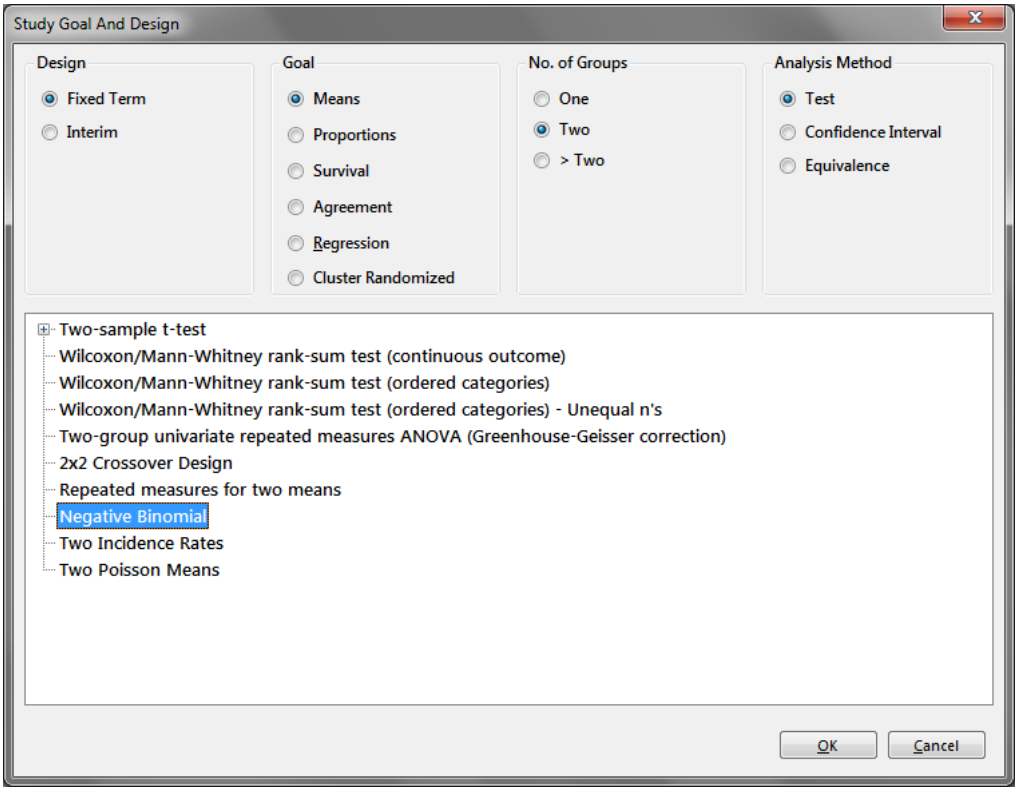


Figure 4.9.1. Study Goal and Design Window

2. In order to select the Negative Binomial design table, navigate through the Study Goal and Design Window by selecting "Fixed Term" as the Design, "Means" as the Goal, "Two" as the Number of Groups and "Test" as the Analysis Method. Then click "OK" and the test window will appear as per Figure 4.9.2.

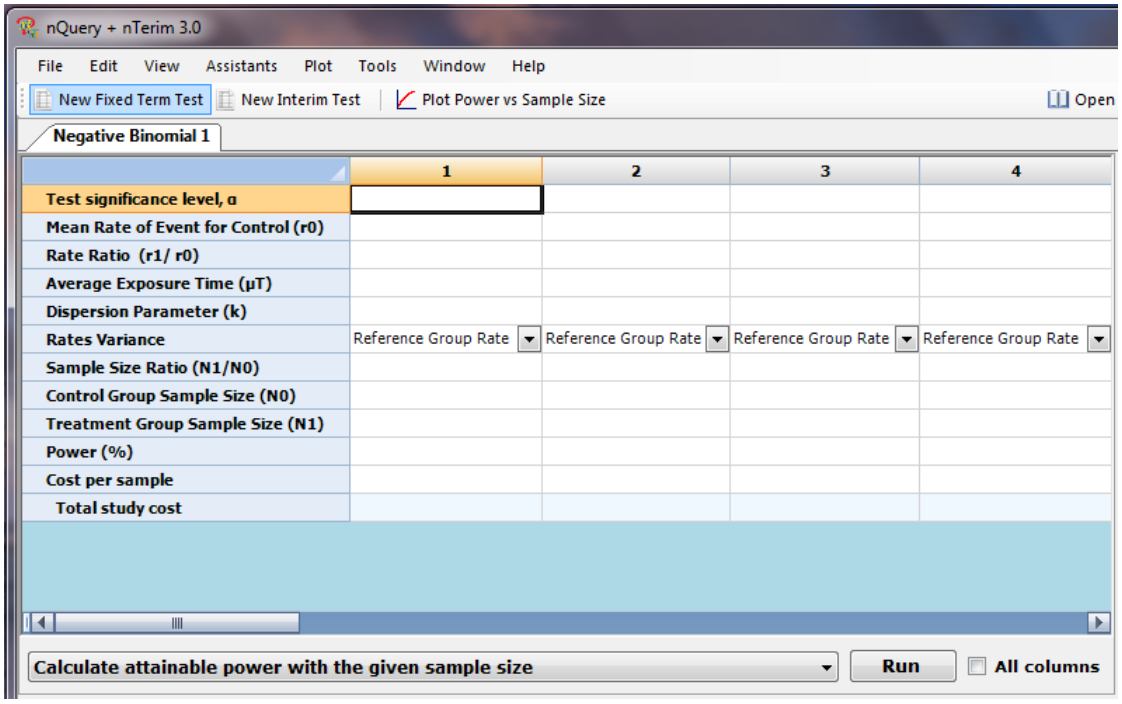


Figure 4.9.2. Two Negative Binomial Rates Test Table

The first calculation will be for Sample Size, to begin we enter the values as follows.

3. First enter 0.05 for the Test Significance level row, then enter 0.8 and 0.85 for the "Mean Rate of the Event for the Control" and "Rate Ratio" variable rows respectively.

4. Next, enter 0.75 for the "Average Exposure Time", 0.7 for the "Dispersion Parameter" and 1 for the "Sample Size Ratio". Finally, enter 80 in the Power row.

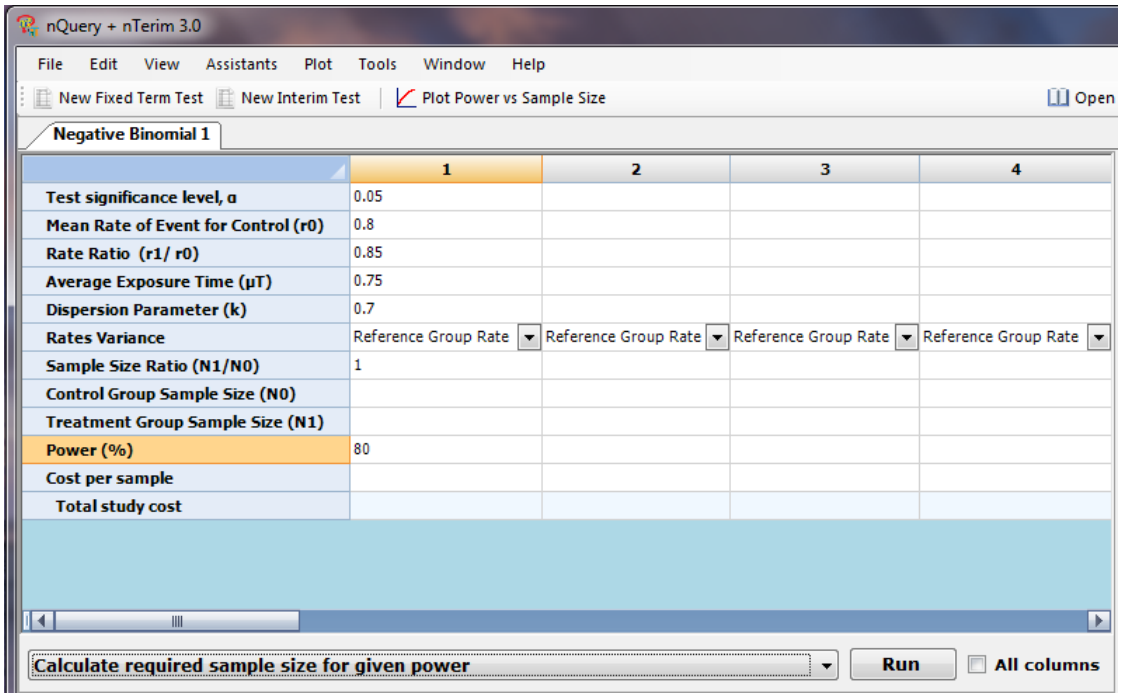
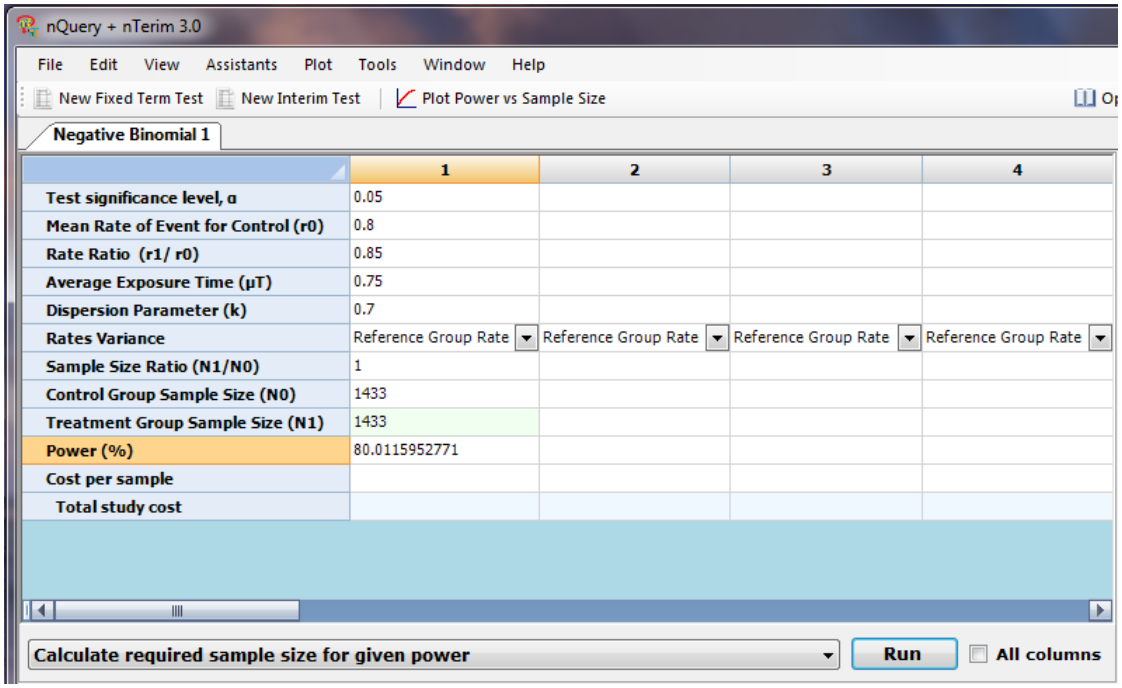


Figure 4.9.3. Values entered for Negative Binomial study design

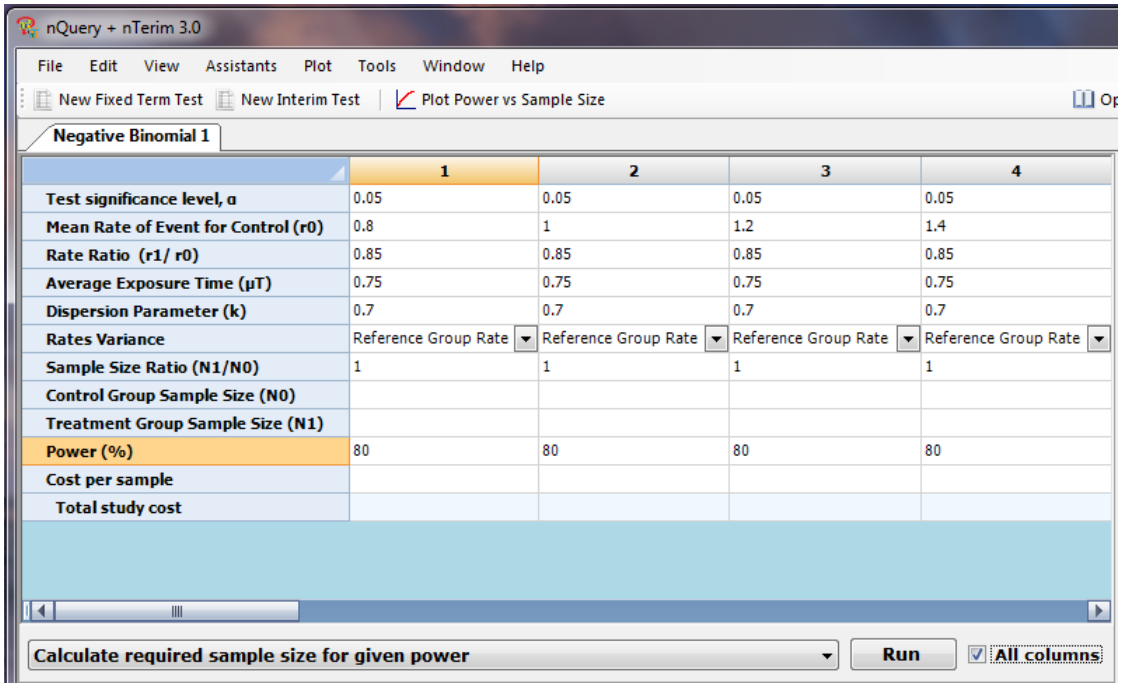
5. Select “Calculate required sample size for given power” from the dropdown menu beside the “Run” button. Then click “Run”.



	1	2	3	4
Test significance level, α	0.05			
Mean Rate of Event for Control (r0)	0.8			
Rate Ratio (r1/ r0)	0.85			
Average Exposure Time (μT)	0.75			
Dispersion Parameter (k)	0.7			
Rates Variance	Reference Group Rate	Reference Group Rate	Reference Group Rate	Reference Group Rate
Sample Size Ratio (N1/N0)	1			
Control Group Sample Size (N0)	1433			
Treatment Group Sample Size (N1)	1433			
Power (%)	80.0115952771			
Cost per sample				
Total study cost				

Figure 4.9.4. Completed Negative Binomial study design

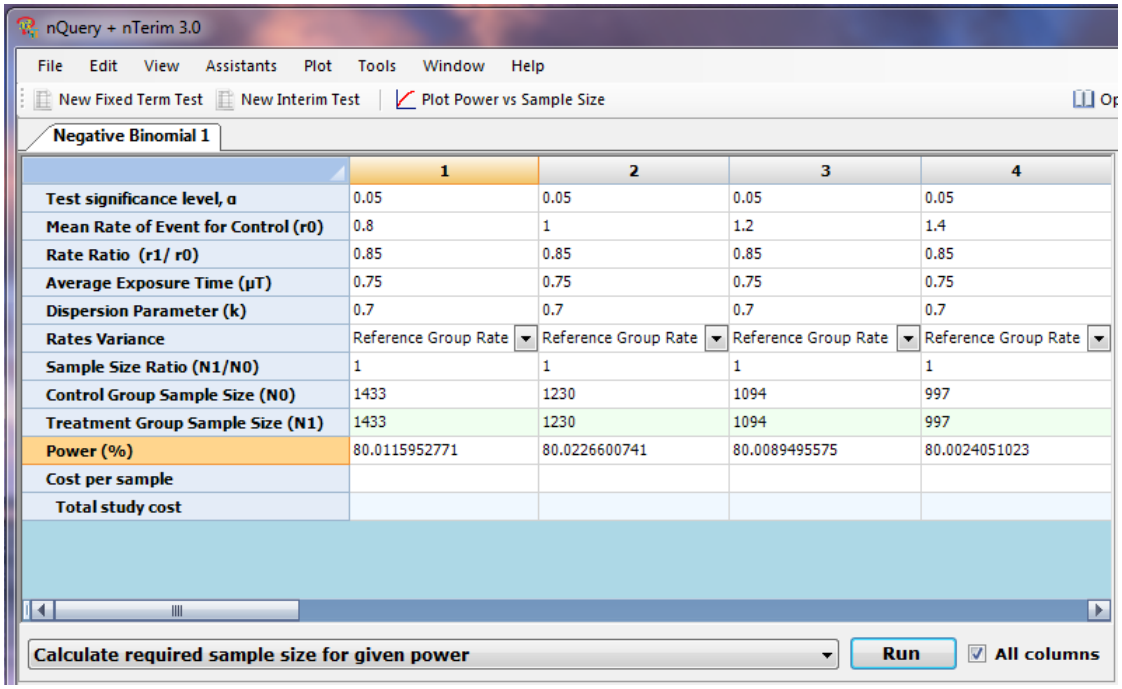
6. The next calculation is a sensitivity analysis for sample size where we change the mean rate of the event for the control to investigate the impact this has on the sample size estimate. To do this, copy the same values across to columns 2 to 4 and delete the values for sample size. Then, change the values for columns 2, 3 and 4 to 1, 1.2 and 1.4 respectively. This will give a table as per Figure 4.9.5.



	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
Mean Rate of Event for Control (r0)	0.8	1	1.2	1.4
Rate Ratio (r1/ r0)	0.85	0.85	0.85	0.85
Average Exposure Time (μT)	0.75	0.75	0.75	0.75
Dispersion Parameter (k)	0.7	0.7	0.7	0.7
Rates Variance	Reference Group Rate	Reference Group Rate	Reference Group Rate	Reference Group Rate
Sample Size Ratio (N1/N0)	1	1	1	1
Control Group Sample Size (N0)				
Treatment Group Sample Size (N1)				
Power (%)	80	80	80	80
Cost per sample				
Total study cost				

Figure 4.9.5. Sensitivity analysis around the Mean Rate of Event for Control

7. Select “Calculate required sample size for given power” from the dropdown menu beside the “Run” button. Then tick the box to run “All Columns”. Then click “Run”. This will give the resultant sample sizes of 1433, 1230, 1094 and 997 sequentially as in Figure 4.9.6.



	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
Mean Rate of Event for Control (r0)	0.8	1	1.2	1.4
Rate Ratio (r1/ r0)	0.85	0.85	0.85	0.85
Average Exposure Time (μT)	0.75	0.75	0.75	0.75
Dispersion Parameter (k)	0.7	0.7	0.7	0.7
Rates Variance	Reference Group Rate	Reference Group Rate	Reference Group Rate	Reference Group Rate
Sample Size Ratio (N1/N0)	1	1	1	1
Control Group Sample Size (N0)	1433	1230	1094	997
Treatment Group Sample Size (N1)	1433	1230	1094	997
Power (%)	80.0115952771	80.0226600741	80.0089495575	80.0024051023
Cost per sample				
Total study cost				

Figure 4.9.6. Results from Sensitivity analysis

The effect of changing the mean rate of the event for the control group on the relationship between Power and Sample size can be explored further using the “Plot Power vs Sample Size” button at the top of the screen.

8. Select the first column by clicking the 1 at the top of column 1. Then hold down the Shift key and click the 4 at the top of column 4. All four columns will now be highlighted.

9. Click the “Plot Power vs Sample Size” button at the top of the screen. This will give you a plot as displayed in Figure 4.9.7 which will show the relationship between power and sample size for each column. Right-click to add features such as a legend to the graph and double click elements for user options and editing.

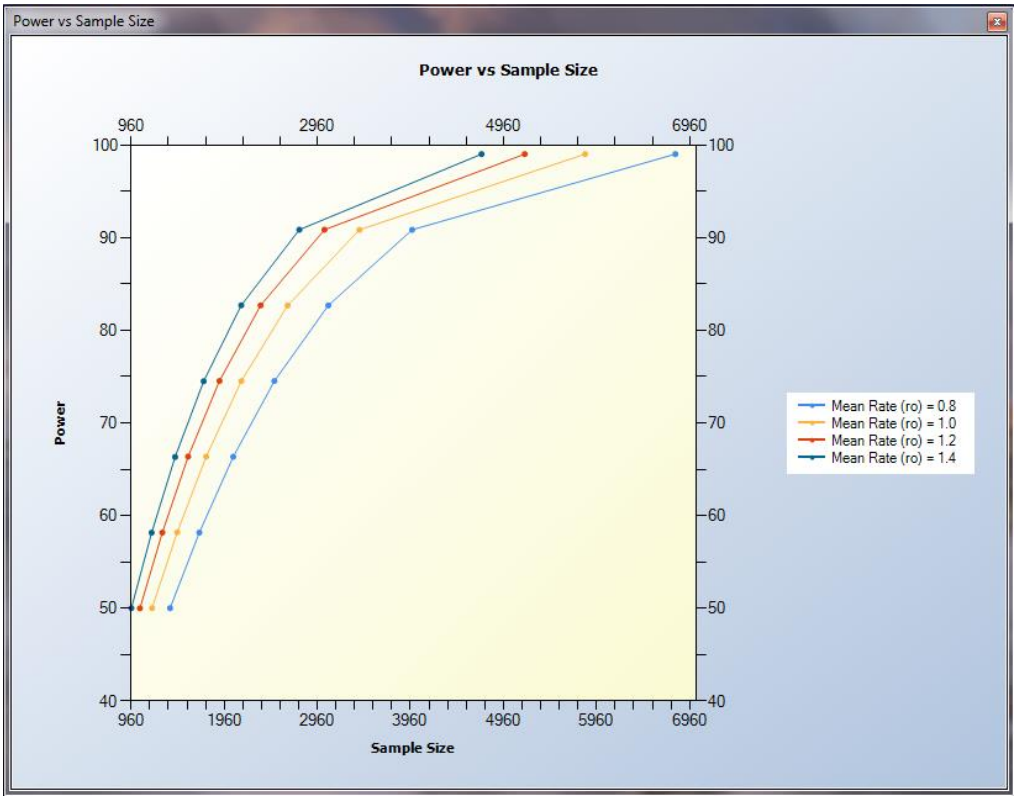


Figure 4.9.7. Power vs. Sample Size plot

10. Additionally, by clicking on the desire study design column and going to the "Output" tab at the bottom of the screen, you can get an output statement for the results given by the current calculation as per Figure 4.9.8.

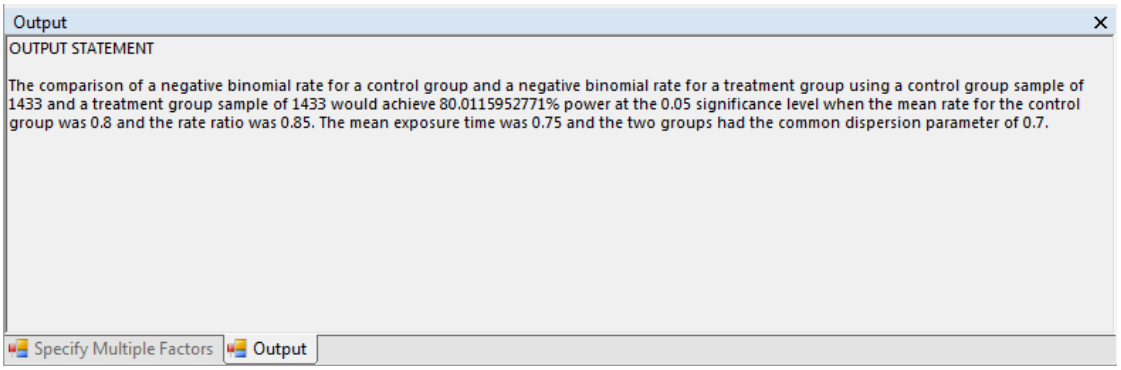
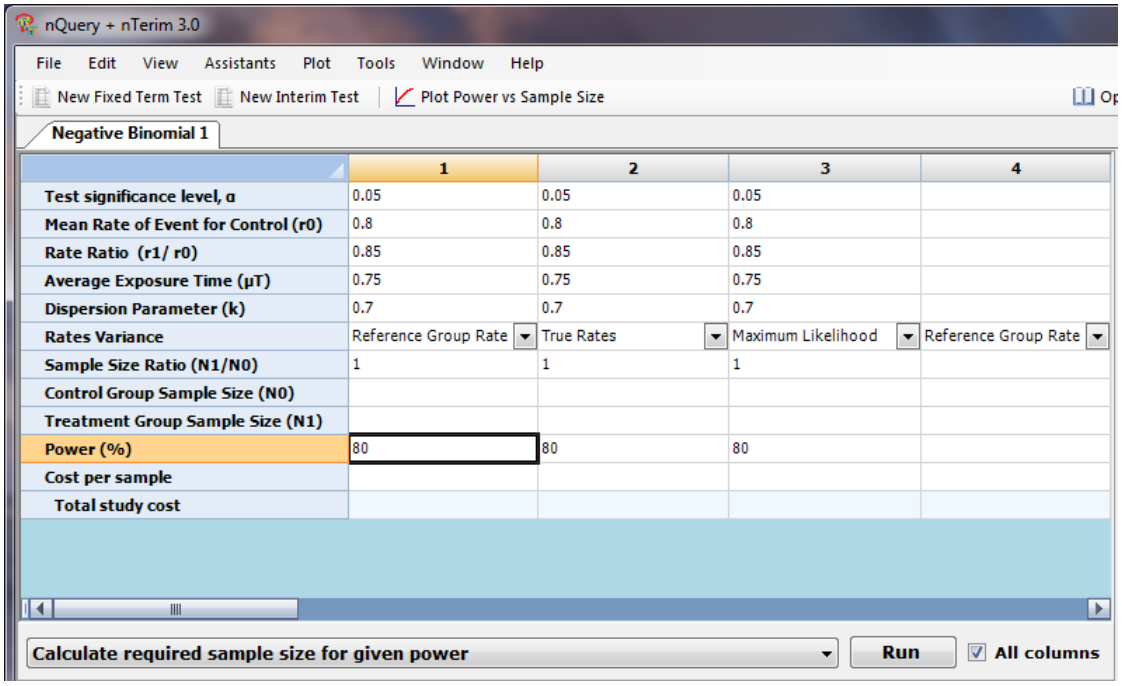


Figure 4.9.8. Study design Output statement

11. Finally, we investigate the effect on the choice of "Rates Variance" on sample size. To do this, delete the columns 2 to 4 and copy the column 1 values across to columns 2 and 3.

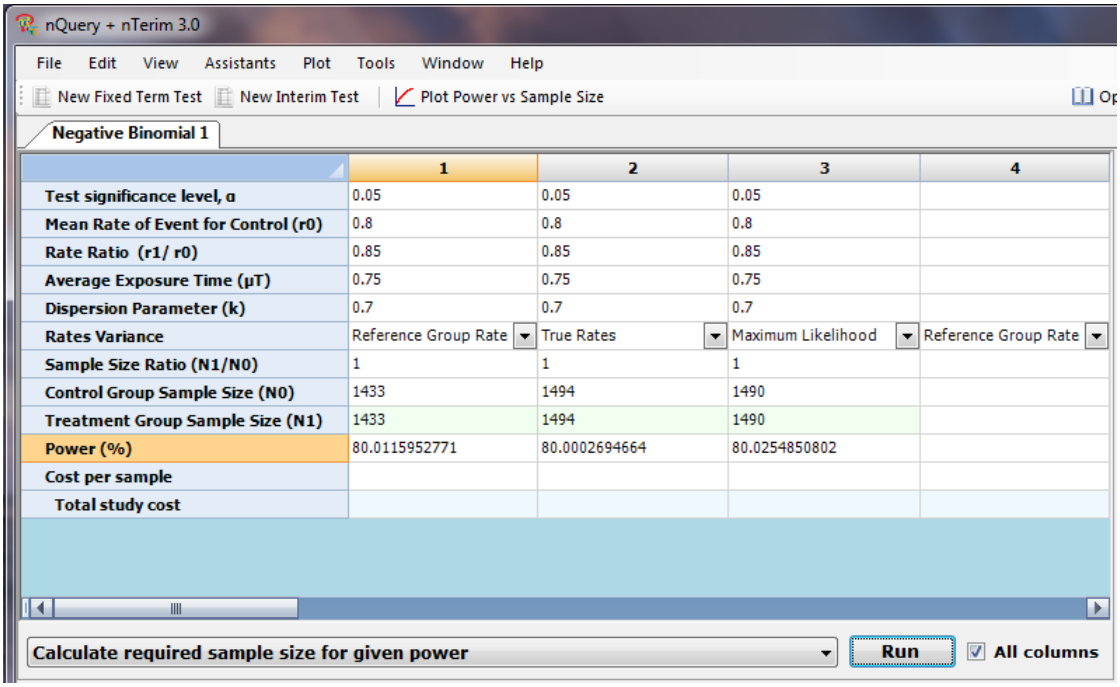
12. For Rates Variance, select "True Rates" and "Maximum Likelihood" for column 2 and 3 respectively. This will give a table as per Figure 4.9.9.



	1	2	3	4
Test significance level, α	0.05	0.05	0.05	
Mean Rate of Event for Control (r0)	0.8	0.8	0.8	
Rate Ratio (r1/ r0)	0.85	0.85	0.85	
Average Exposure Time (μT)	0.75	0.75	0.75	
Dispersion Parameter (k)	0.7	0.7	0.7	
Rates Variance	Reference Group Rate	True Rates	Maximum Likelihood	Reference Group Rate
Sample Size Ratio (N1/N0)	1	1	1	
Control Group Sample Size (N0)				
Treatment Group Sample Size (N1)				
Power (%)	80	80	80	
Cost per sample				
Total study cost				

Figure 4.9.9. Comparing Rates Variance options

13. Select "Calculate required sample size for given power" from the dropdown menu beside the "Run" button. Then tick the box to run "All Columns". Then click "Run".



	1	2	3	4
Test significance level, α	0.05	0.05	0.05	
Mean Rate of Event for Control (r0)	0.8	0.8	0.8	
Rate Ratio (r1/ r0)	0.85	0.85	0.85	
Average Exposure Time (μT)	0.75	0.75	0.75	
Dispersion Parameter (k)	0.7	0.7	0.7	
Rates Variance	Reference Group Rate	True Rates	Maximum Likelihood	Reference Group Rate
Sample Size Ratio (N1/N0)	1	1	1	
Control Group Sample Size (N0)	1433	1494	1490	
Treatment Group Sample Size (N1)	1433	1494	1490	
Power (%)	80.0115952771	80.0002694664	80.0254850802	
Cost per sample				
Total study cost				

Figure 4.9.10. Result from comparing Rate Variance options

This will give the resultant sample sizes of 1433, 1494 and 1490 sequentially as in Figure 4.9.10. A similar graph and output statement can be generated for this example as for the above example.

4.10. Two Incidence Rates

4.10.1 Introduction

Incidence data is often obtained in a variety of clinical and field research activities such as the number of accidents at a junction or number of occurrences of a disease in a year. Researchers often plan studies in terms of years of exposure rather than directly through sample size. This two sample test is used to test hypotheses about the difference between two incidence rates in terms of person-years of exposure.

This table facilitates the calculation of the power and sample size for hypothesis tests comparing two incidence rates. Power and sample size is computed using the method outlined by Smith and Morrow (1996).

4.10.2 Methodology

This table provides generic sample size and power calculations for studies which will be comparing two incidence rates. The tables for two Poisson means or two negative binomial means can be used instead if the researcher will know the analysis method and estimates of the additional associated parameters beforehand.

The sample size calculation for two incidence rates is taken from Smith and Morrow (1996). This table can be used to calculate the power, the sample size and the treatment group incidence rate given all other terms in the table are specified.

To calculate power and sample size, the user must specify the test significance level (α), whether to use a one or two sided test, the control group incidence rate (λ_1), the treatment group incidence rate (λ_2) and the difference in rates ($\lambda_1 - \lambda_2$).

The sample size formula uses the normal approximation to yield the following:

$$y = \frac{\left(z_{1-\beta} + z_{1-\alpha/2}\right)^2 (\lambda_1 + \lambda_2)}{(\lambda_1 - \lambda_2)^2} \qquad \{4.10.1\}$$

where y is the sample size per group expressed in person-years (or person-(units of time)). For the one-sided test, $z_{1-\alpha/2}$ is replaced with $z_{1-\alpha}$. Simple rearrangement of the above formula gives the following equation for the power:

$$Power = \Phi\left(\frac{\sqrt{y(\lambda_1 - \lambda_2)}}{(\lambda_1 + \lambda_2)} - z_{1-\alpha/2}\right) \qquad \{4.10.2\}$$

A closed form equation is not used to calculate the treatment group incidence rate. Instead a search algorithm is used. The search algorithm calculates power at various values for the treatment group incidence rate until the desired power is reached.

4.10.3 Examples

Example 1: Validation example calculating required sample size for a given power
The following example is taken from Table 3.2 on page 55 of Smith and Morrow (1996) where a sample size calculation problem is conducted to show the effect of changing the treatment group incidence rate and power.

The following steps outline the procedure for this example:

1. Open nQuery + nTerim 4.0 via the start menu or desktop shortcut. Click “New Fixed Term Test” from the top of the window. The “Study Goal and Design” window will appear as Figure 4.10.1 below.

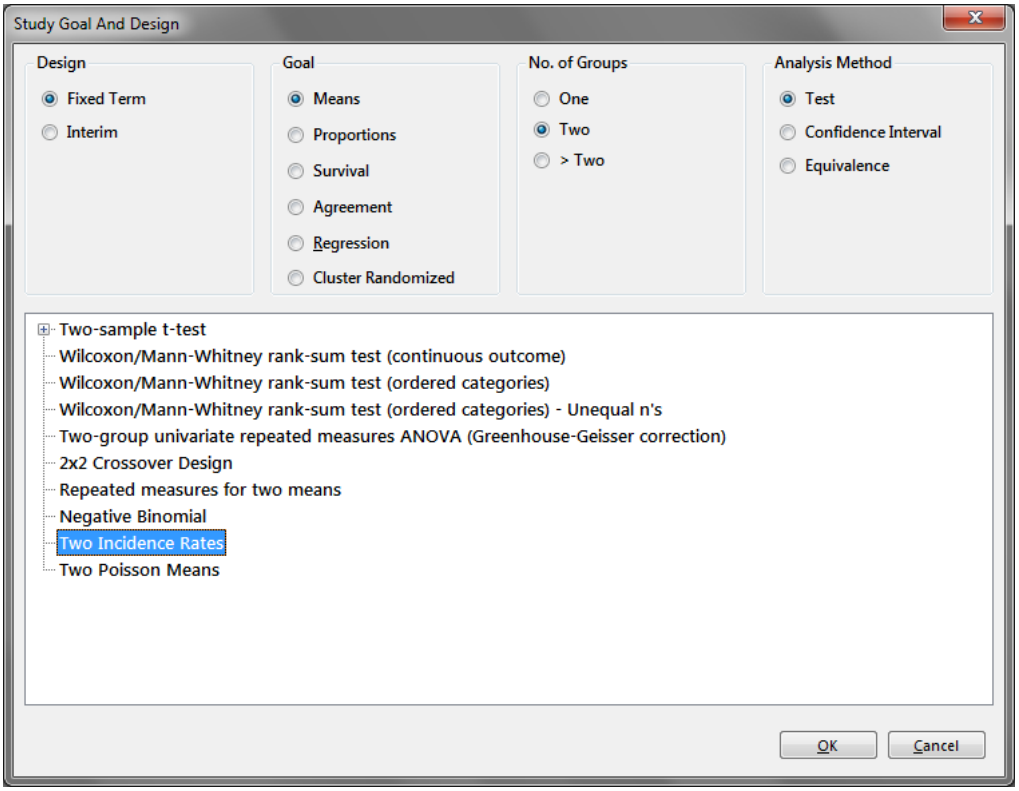


Figure 4.10.1. Study Goal and Design Window

2. In order to select the Two Incidence Rates design table, navigate through the Study Goal and Design Window by selecting "Fixed Term" as the Design, "Means" as the Goal, "Two" as the Number of Groups and "Test" as the Analysis Method. Then click "OK" and the test window will appear as per Figure 4.10.2.

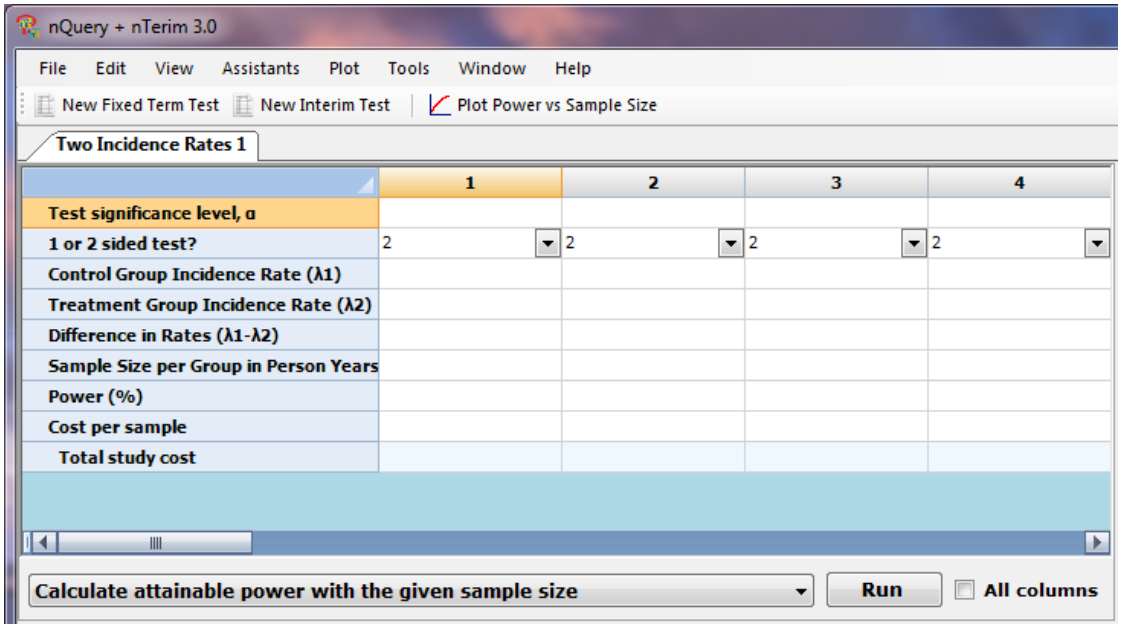


Figure 4.10.2. Two Incidence Rates Test Table

The first calculation will be for Sample Size, to begin we enter the values as follows.

3. First enter 0.05 for the Test Significance level row, then enter 1 and 0.8 for the "Control Group Incidence Rate" and "Treatment Group Incidence Rate" variable rows respectively. The "Difference in Rates" variable will update automatically to 0.2.
4. Then, enter 80 in the Power row, this table will appear as per Figure 4.10.3.

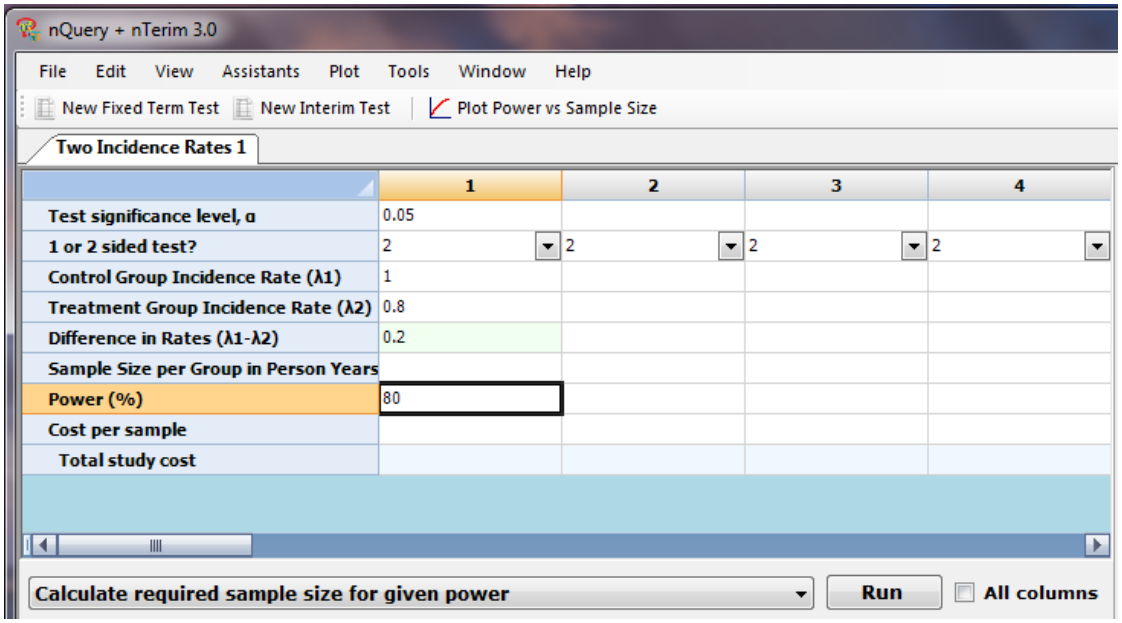
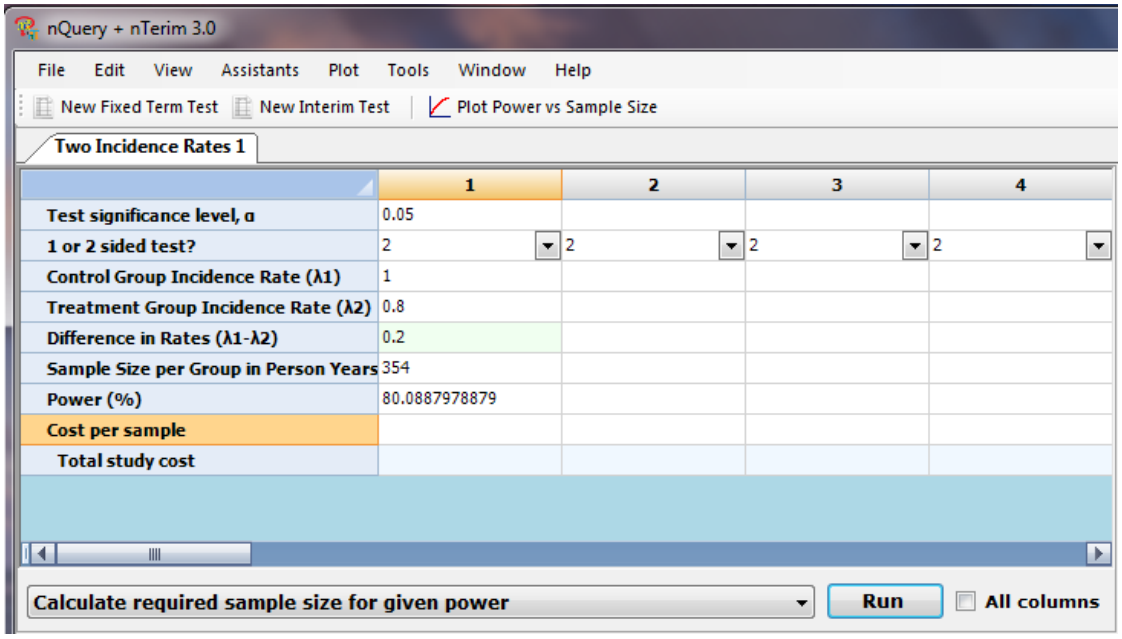


Figure 4.10.3. Values entered for Two Incidence Rates study design

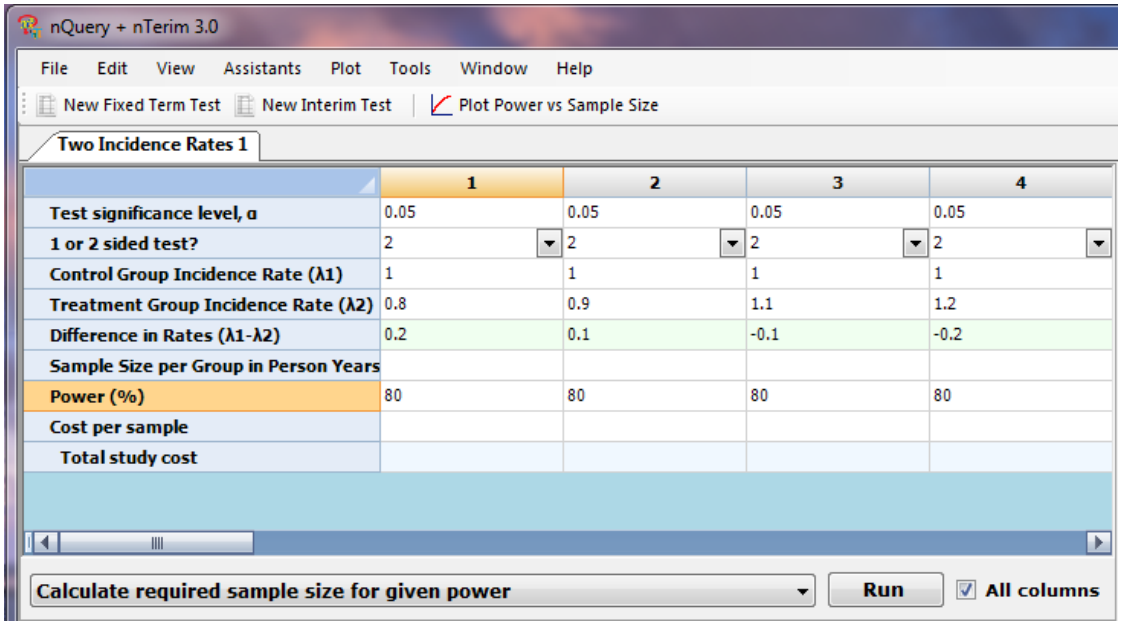
5. Select "Calculate required sample size for given power" from the dropdown menu beside the "Run" button. Then click "Run". This will give a result of 354 for the sample size as in Figure 4.10.4. The result in Smith and Morrow is 353 but this is due to two decimal place rounding in their calculations.



	1	2	3	4
Test significance level, α	0.05			
1 or 2 sided test?	2	2	2	2
Control Group Incidence Rate (λ1)	1			
Treatment Group Incidence Rate (λ2)	0.8			
Difference in Rates (λ1-λ2)	0.2			
Sample Size per Group in Person Years	354			
Power (%)	80.0887978879			
Cost per sample				
Total study cost				

Figure 4.10.4. Completed Two Incidence Rates study design

6. The next calculation is a sensitivity analysis for sample size where we change the treatment group incidence rate to investigate the impact this has on the sample size estimate. To do this, copy the same values across to columns 2 to 4 and delete the values for sample size. Then, change the values for columns 2 to 4 to 0.9, 1.1 and 1.2 respectively. This will give a table as per Figure 4.10.5.

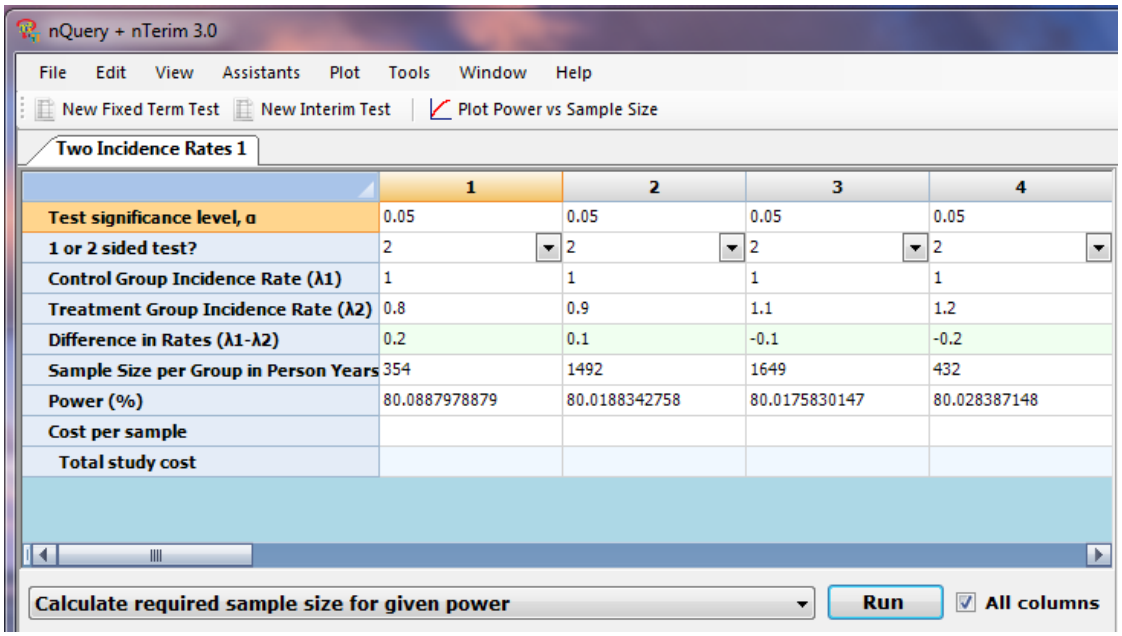


	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
1 or 2 sided test?	2	2	2	2
Control Group Incidence Rate (λ1)	1	1	1	1
Treatment Group Incidence Rate (λ2)	0.8	0.9	1.1	1.2
Difference in Rates (λ1-λ2)	0.2	0.1	-0.1	-0.2
Sample Size per Group in Person Years				
Power (%)	80	80	80	80
Cost per sample				
Total study cost				

Figure 4.10.5. Sensitivity analysis around the Treatment Group Incidence Rate

7. Select “Calculate required sample size for given power” from the dropdown menu beside the “Run” button. Then tick the box to run “All Columns”. Then click “Run”.

This will give the resultant sample sizes of 354, 1492, 1649 and 432 sequentially as in Figure 4.10.6. Similar to the example above; these answers differ from Smith and Morrow (1996) due to rounding.



	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
1 or 2 sided test?	2	2	2	2
Control Group Incidence Rate (λ1)	1	1	1	1
Treatment Group Incidence Rate (λ2)	0.8	0.9	1.1	1.2
Difference in Rates (λ1-λ2)	0.2	0.1	-0.1	-0.2
Sample Size per Group in Person Years	354	1492	1649	432
Power (%)	80.0887978879	80.0188342758	80.0175830147	80.028387148
Cost per sample				
Total study cost				

Figure 4.10.6. Results from Sensitivity analysis

The effect of changing the test statistic on the relationship between Power and Sample size can be explored further using the "Plot Power vs Sample Size" button at the top of the screen.

8. Select the first column by clicking the 1 at the top of column 1. Then hold down Shift and click the 4 at the top of column 4. All four columns will now be highlighted.

9. Click the "Plot Power vs Sample Size" button at the top of the screen. This will give you a plot as displayed in Figure 4.10.7 which will show the relationship between power and sample size for each column. Right-click to add features such as a legend to the graph and double click elements for user options and editing.

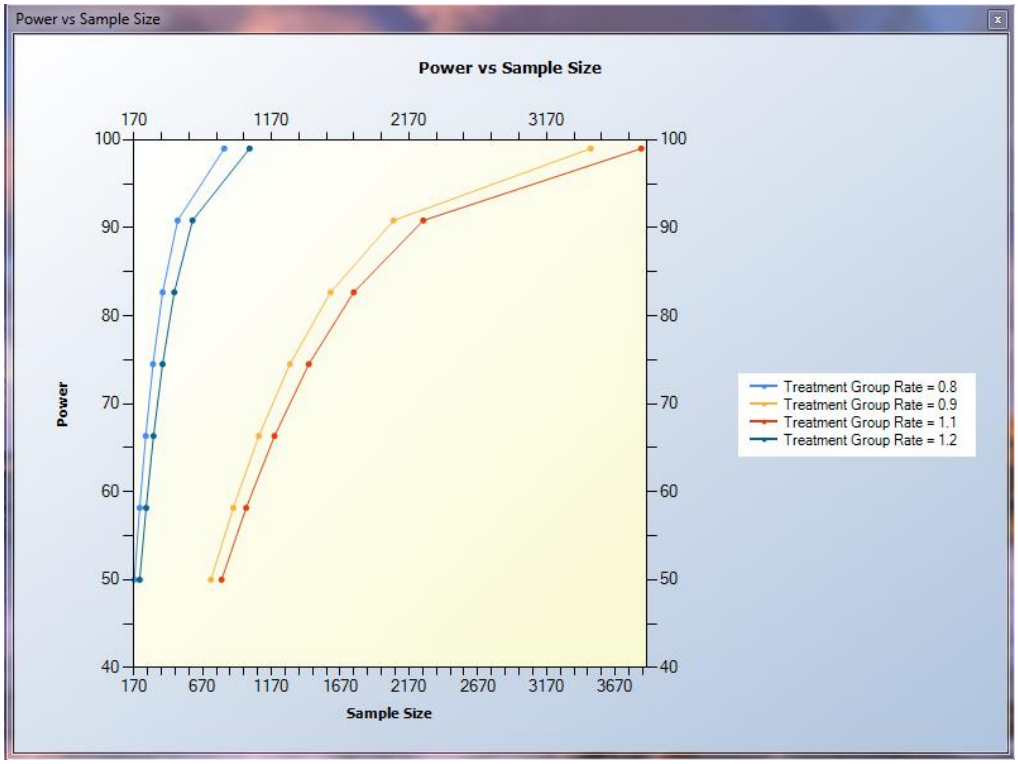


Figure 4.10.7. Power vs. Sample Size plot

10. Finally, by clicking on the desired study design column and going to the "Output" tab at the bottom of the screen, you can get an output statement for the results given by the current calculation as per Figure 4.10.8 below.

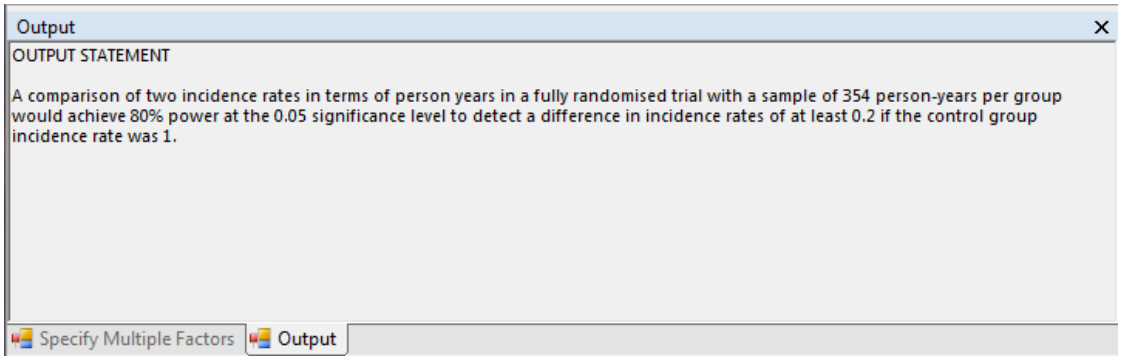


Figure 4.10.8. Study design Output statement

This Output statement can then be easily transferred directly from the output window into a report.

4.11. Cox Regression

4.11.1 Introduction

The Cox Proportional Hazards Regression model is a semi-parametric method used to measure the effect on the hazard ratio of independent variable(s) on a dependent variable in survival analysis. For example, taking a new therapy may halve the hazard rate for a cancer occurring. This model can be used to calculate the effect of a variable on the hazard rate without needing to specify the underlying hazard function while also allowing censoring. This flexibility has made the Cox Proportional Hazards model a widely used tool in survival analysis.

This table facilitates the calculation of the power and sample size for an independent variable, both in a single variable model and given its relationship with other variables in a multivariable model. Power and sample size is computed using the method outlined by Hsieh and Lavori (2000).

4.11.2 Methodology

The Cox Proportional Hazards model assumes that the hazard function, $\lambda(t)$, for the survival time (T), given the discrete or continuous predictors $X_1 + X_2 \dots X_K$, is described by the following regression equation:

$$\log\left[\frac{\lambda(t|X)}{\lambda_0(t)}\right] = \beta_1 X_1 + \beta_2 X_2 \dots \beta_K X_K \qquad \{4.11.1\}$$

where $\lambda_0(t)$ is the baseline hazard.

The sample size calculation for an independent variable in the Cox Proportional Hazards model is taken from Hsieh and Lavori (2000). This table can be used to calculate the power, the sample size, the test significance level or the log hazard ratio given all other terms in the table are specified.

Calculations use a standard normal approximation and assume the statistical hypothesis is being tested using the Wald statistic. To calculate power and sample size, the user must specify the test significance level (α), whether to use a 1 or 2 sided test, the overall event rate (P), the expected log hazard ratio for the independent variable given the other terms in the model (β_1), the standard deviation of the independent variables (σ) and the expected multiple correlation coefficient (R^2) between the independent variable and the other independent variables in the model. If there is only one independent variable in the model, then R^2 is set to zero.

The formula for the sample size in a 2 sided test for the null hypothesis of $\beta_1 = 0$ versus the alternative hypothesis of $\beta_1 = B$ for an independent variable is given by the following equation:

$$N = \frac{\left(z_{1-\frac{\alpha}{2}} + z_{1-\beta}\right)^2}{P(1-R^2)\sigma^2 B^2} \qquad \{4.11.2\}$$

The power is calculated by re-arrangement of the above formula to give the following equation:

$$Power = \Phi\left(\sqrt{P(1-R^2)\sigma^2 B^2 N} - z_{1-\frac{\alpha}{2}}\right) \qquad \{4.11.3\}$$

Similar re-arrangements yield the following equations for the test significance level and log hazard ratio:

$$\alpha = 2\left\{1 - \Phi\left(\sqrt{P(1-R^2)\sigma^2 B^2 N} - z_{1-\beta}\right)\right\} \qquad \{4.11.4\}$$

$$\beta_1 = \sqrt{\frac{\left(z_{1-\frac{\alpha}{2}} + z_{1-\beta}\right)^2}{P(1-R^2)\sigma^2 N}} \qquad \{4.11.5\}$$

For the one sided test, $z_{1-\alpha}$ would be used in place of $z_{1-\frac{\alpha}{2}}$.

4.11.3 Example

Example 1: Validation example calculating required sample size for a given power
The following examples are taken from Hsieh and Lavori (2000) where a sample size calculation problem is conducted for a multiple myeloma data set (page 557) and then a sensitivity analysis is conducted (Table 1, page 555) to show the effect of changing the event rate.

The following steps outline the procedure for this example:

1. Open nQuery + nTerim 4.0 via the start menu or desktop shortcut. Click “New Fixed Term Test” from the top of the window. The “Study Goal and Design” window will appear as displayed in Figure 4.11.1 below.

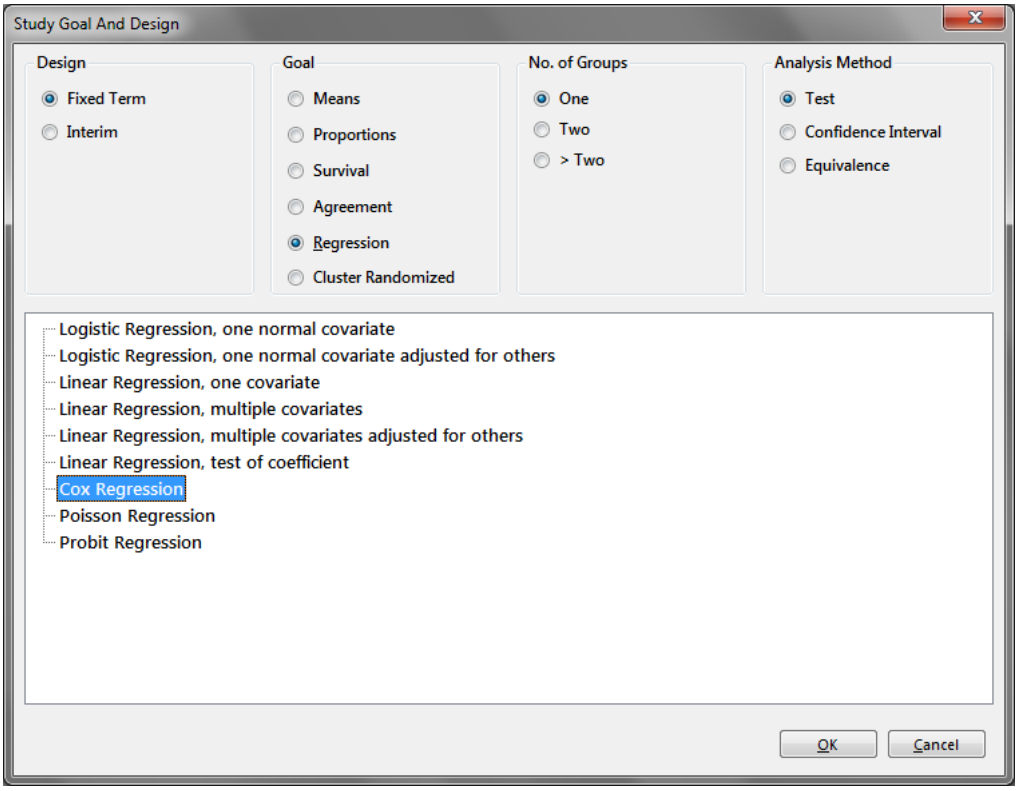


Figure 4.11.1. Study Goal and Design Window

2. In order to select the Cox Regression design table, navigate through the Study Goal and Design Window by selecting "Fixed Term" as the Design, "Regression" as the Goal, "One" as the Number of Groups and "Test" as the Analysis Method. Then click "OK" and the test window will appear as per Figure 4.11.2.

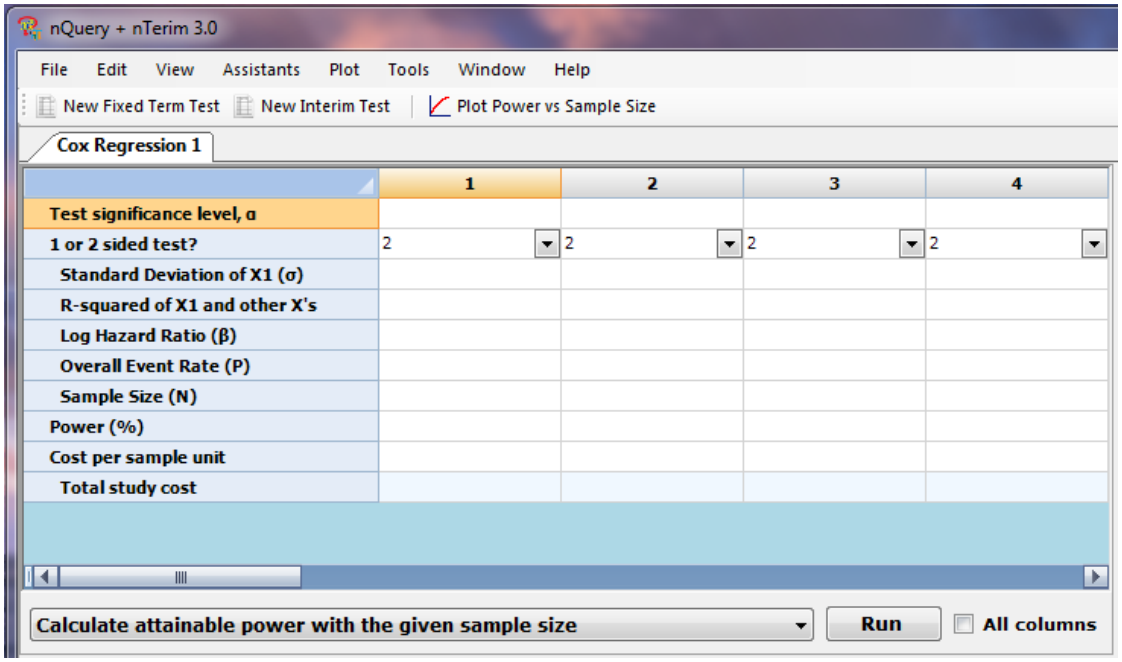


Figure 4.11.2. Cox Regression Test Table

The first calculation will be for Sample Size for the multiple myeloma data set example.

3. First enter 0.05 for the Test Significance level, then select 1 for the "1 or 2 sided test?" variable dropdown menu.

4. Enter 0.3126 for the "Standard Deviation of X1" variable, enter 0.1837 for the "R-squared of X1 and other X's" variable, enter 1 for the Log Hazard Ratio, enter 0.738 for the Overall Event Rate and finally, enter 80 in the Power row. The table will appear as per Figure 4.11.3.

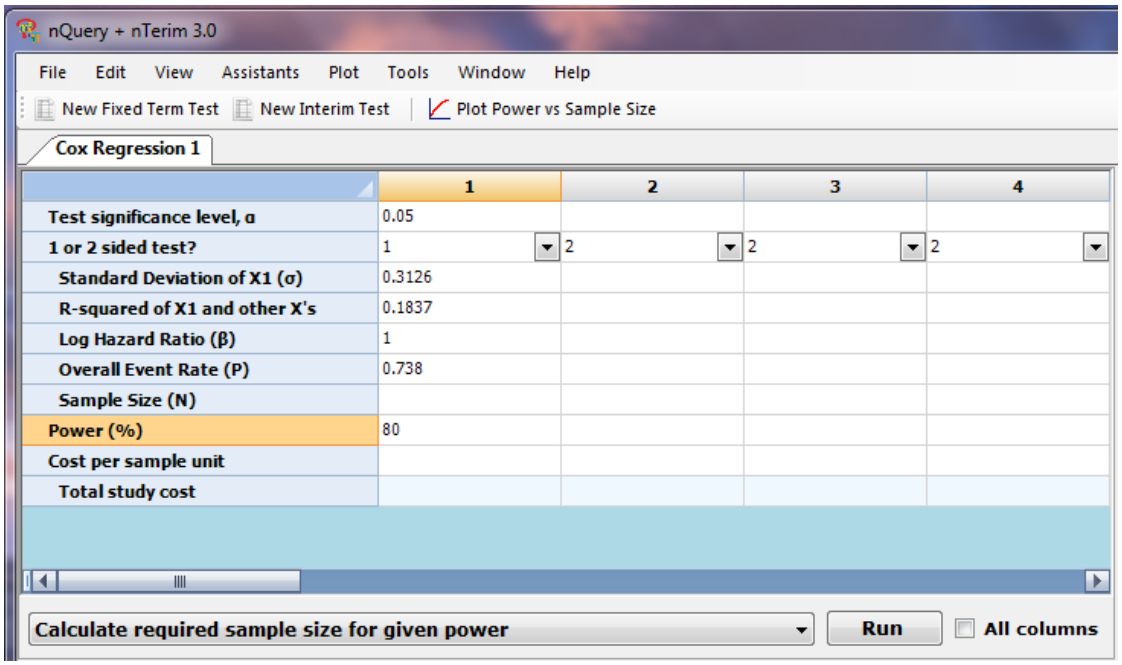
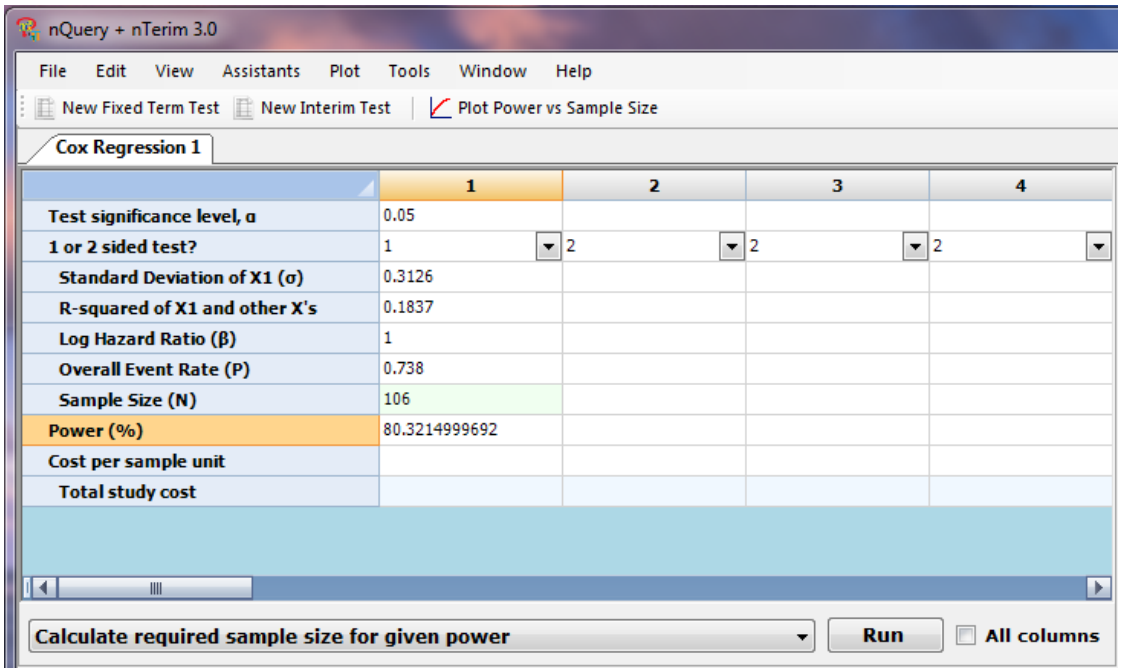


Figure 4.11.3. Values entered for Cox Regression design

5. Select “Calculate required sample size for given power” from the dropdown menu beside the “Run” button, then click “Run”. This will give a result of 106 for the sample size as in Figure 4.11.4. The result presented in the paper by Hsieh and Lavori is 107, however this is due to them rounding the unadjusted for R^2 sample size as an intermediate step. The nQuery + nTerim calculation treats the interim sample size as an unrounded figure.



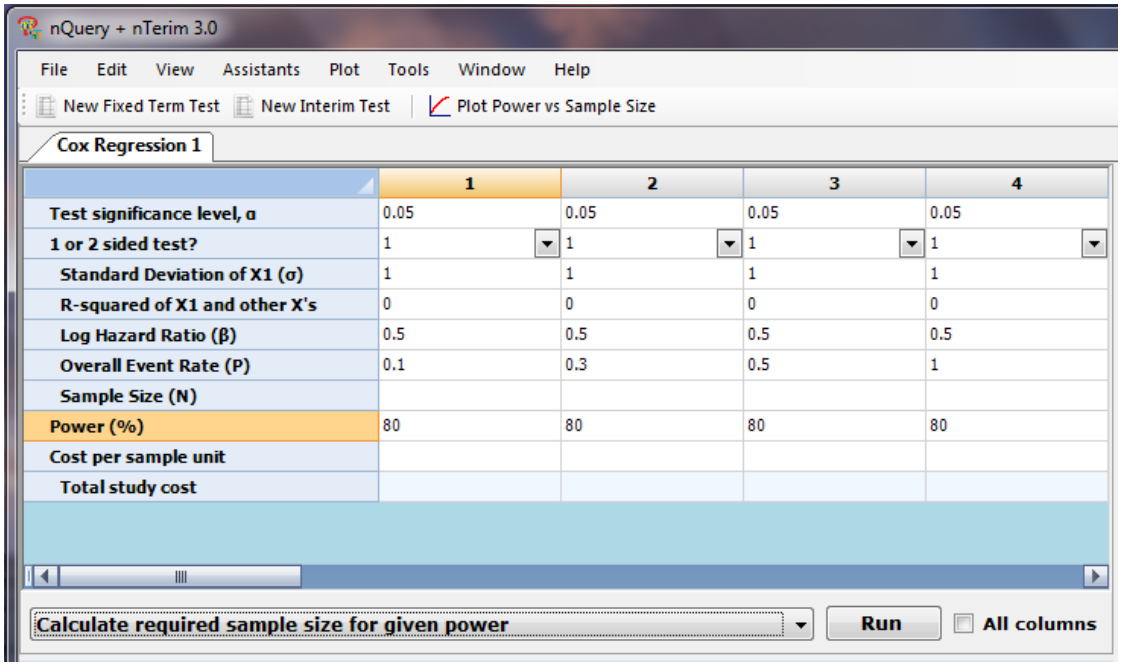
	1	2	3	4
Test significance level, α	0.05			
1 or 2 sided test?	1	2	2	2
Standard Deviation of X1 (σ)	0.3126			
R-squared of X1 and other X's	0.1837			
Log Hazard Ratio (β)	1			
Overall Event Rate (P)	0.738			
Sample Size (N)	106			
Power (%)	80.3214999692			
Cost per sample unit				
Total study cost				

Figure 4.11.4. Completed Cox Regression study design

The next calculation is a sensitivity analysis for sample size where we change the event rate to explore its effect on sample size. These values are taken from Table 1 (page 555) of Hsieh and Lavori, 2000.

6. In the first column, enter 0.05 for Test Significance, 1 for “1 or 2 sided test”, 1 for the standard deviation, zero for the R^2 value, 0.5 for the log hazard ratio, 0.1 for the “Overall Event Rate” and 80 for the power.

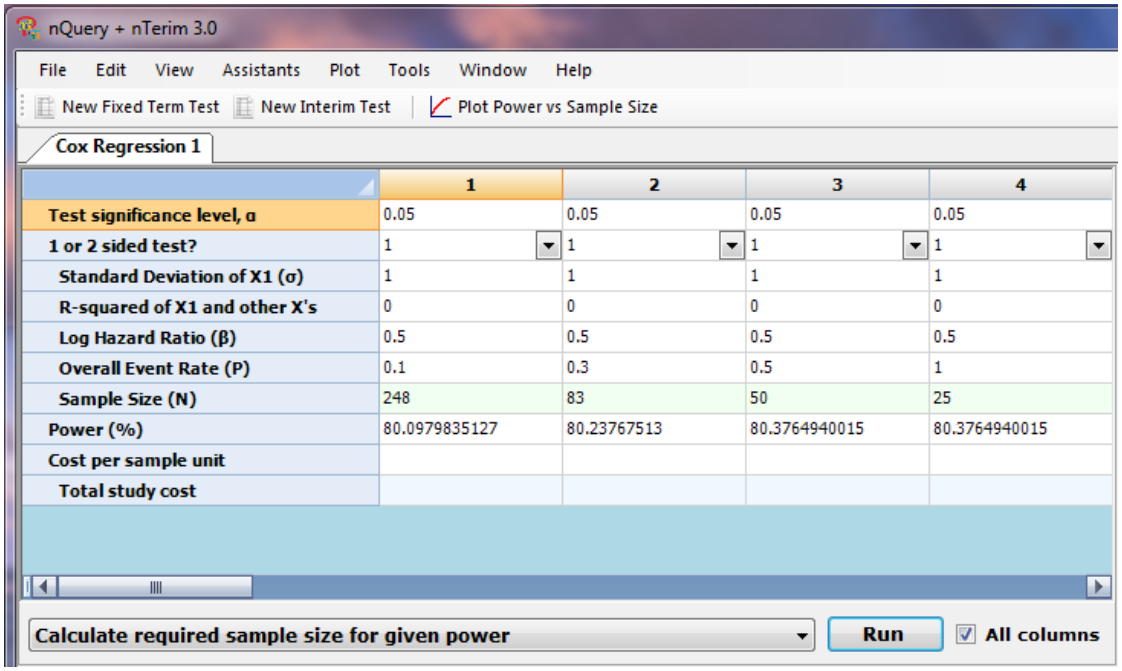
7. Then copy these values across to columns 2 to 4 and change the value for “overall Event Rate” for columns 2 to 4 to 0.3, 0.5 and 1 respectively. This will give a table as per Figure 4.11.5.



	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
1 or 2 sided test?	1	1	1	1
Standard Deviation of X1 (σ)	1	1	1	1
R-squared of X1 and other X's	0	0	0	0
Log Hazard Ratio (β)	0.5	0.5	0.5	0.5
Overall Event Rate (P)	0.1	0.3	0.5	1
Sample Size (N)				
Power (%)	80	80	80	80
Cost per sample unit				
Total study cost				

Figure 4.11.5. Sensitivity analysis on the Overall Event Rate

8. Select "Calculate required sample size for given power" from the dropdown menu beside the "Run" button. Then tick the box for "All Columns" and then click "Run". This will give the resultant sample sizes of 248, 83, 50 and 25 sequentially as in Figure 4.11.6. Similar to the example above; the answers for column 1 and 2 differ from those presented in Hsieh and Lavori (2000) due to intermediate rounding.



	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
1 or 2 sided test?	1	1	1	1
Standard Deviation of X1 (σ)	1	1	1	1
R-squared of X1 and other X's	0	0	0	0
Log Hazard Ratio (β)	0.5	0.5	0.5	0.5
Overall Event Rate (P)	0.1	0.3	0.5	1
Sample Size (N)	248	83	50	25
Power (%)	80.0979835127	80.23767513	80.3764940015	80.3764940015
Cost per sample unit				
Total study cost				

Figure 4.11.6. Results from Sensitivity analysis

9. The effect of changing the test statistic on the relationship between Power and Sample size can be explored further using the "Plot Power vs Sample Size" button at the top of the screen.

10. Select the first column by clicking the 1 at the top of column 1. Then hold down Shift and click the 4 at the top of column 4. All four columns will now be highlighted.

11. Click the "Plot Power vs Sample Size" button at the top of the screen. This will give you a plot of Power vs Sample Size as displayed Figure 4.11.7. This plot highlights the relationship between power and sample size for each column. Right-click to add features such as a legend to the graph and double click elements for user options and editing.

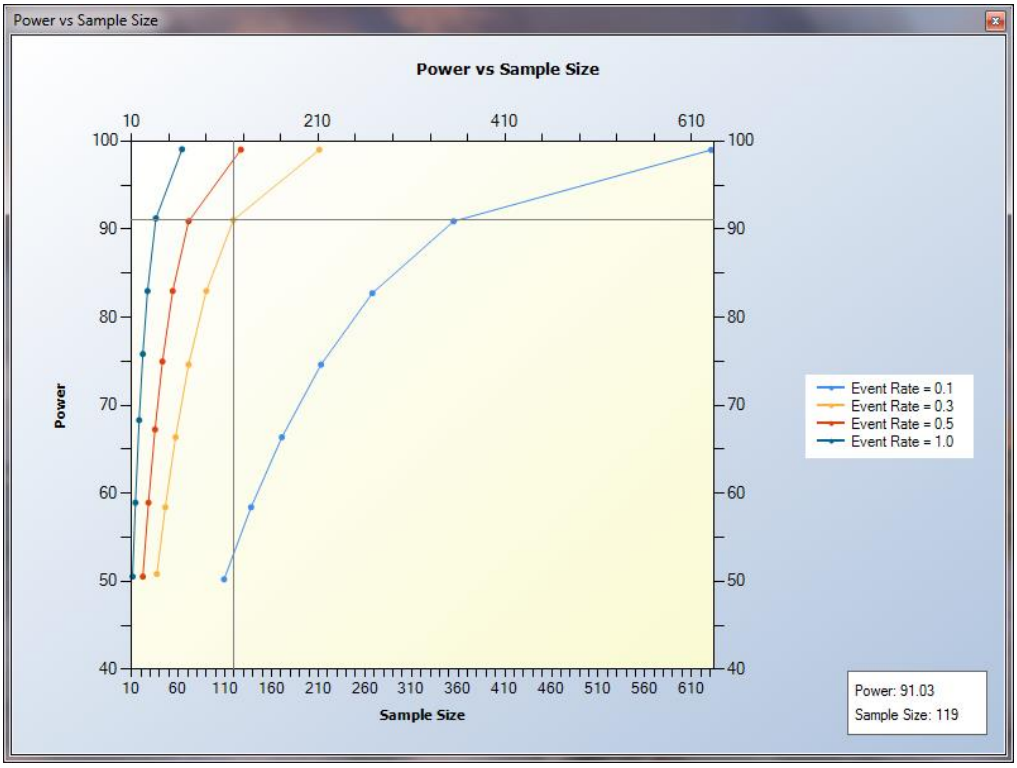


Figure 4.11.7. Power vs Sample Size Plot

12. Finally, by clicking on any of the columns and going to the "Output" tab at the bottom of the screen, you can get an output statement for the results given by the current calculation as per Figure 4.11.8.

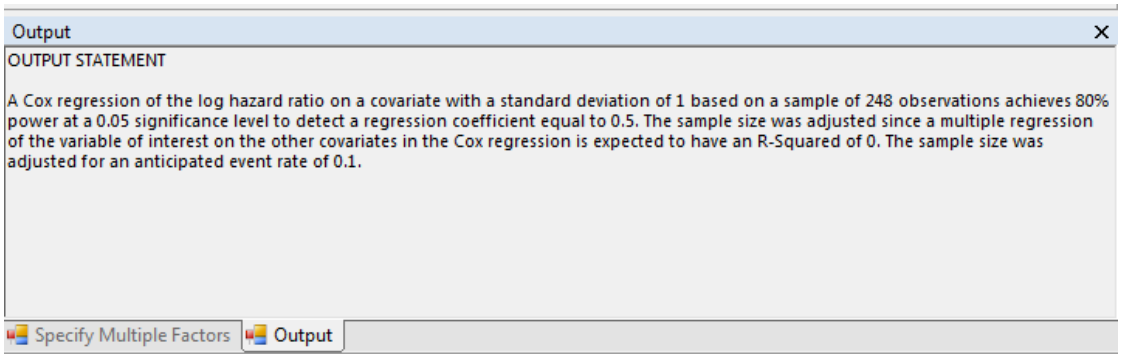


Figure 4.11.8. Study design Output Statement

4.12. Poisson Regression

4.12.1 Introduction

The Poisson Regression Model is a method used to analyse the relationship between a dependent variable which is a count and one or more independent variables. Count data is common in clinical and epidemiological studies and thus the Poisson Regression Model is widely used in these areas.

This table facilitates the calculation of the power and sample size for an independent variable, both in a single variable model and given its relationship with other variables in a multivariable model. Power and sample size is computed using the method outlined by Signorini (1991).

4.12.2 Methodology

The Poisson Regression Model assumes that the dependent variable (Y) follows the Poisson distribution. The Poisson distribution models the probability of y events using the following formula:

$$P(Y = y|\mu, t) = \frac{e^{-\mu t}(\mu t)^y}{y!} \qquad \{4.12.1\}$$

where μ is the mean number of events per unit time of the Poisson distribution and t is length of time of the study in the units of the mean.

In Poisson regression, we assume the mean, μ, is determined by k independent variables by the following relationship:

$$\mu = exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 \dots \beta_K X_K) \qquad \{4.12.2\}$$

where $\beta_1 + \beta_2 \dots \beta_K$ are the regression coefficients of the independent variables. For observation i, the regression model can be expressed as the following:

$$P(Y_i = y_i|\mu_i, t_i) = \frac{e^{-\mu_i t_i}(\mu_i t_i)^{y_i}}{y_i!} \qquad \{4.12.3\}$$

where $\mu_i = exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} \dots \beta_K X_{Ki})$.

The sample size calculation for an independent variable in Poisson Regression is taken from Signorini (1991), with an extension adapted from Hsieh and Lavori (2000) to account for the multivariable case. This table can be used to calculate the power, the sample size, the test significance level or the response rate ratio given all other terms in the table are specified.

Calculations use a standard normal approximation. To calculate power and sample size, the user must specify the test significance level (α), whether to use a 1 or 2 sided test, the baseline response rate (e^{β_0}), the response rate ratio (e^{b_1}/e^{β_0}), the mean exposure time (μ_t), the overdispersion parameter (φ), the expected multiple correlation coefficient (R^2) between the independent variable of interest (X_1) and the other independent variables in the model and the distribution, with the appropriate parameters specified, of the independent variable of interest. If there is only one independent variable in the model, then R^2 is set to zero.

The formula for the sample size in a 2 sided test for the null hypothesis of $\beta_1 = 0$ versus the alternative hypothesis of $\beta_1 = b_1$ for an independent variable is given by the following equation:

$$N = \varphi \left[\frac{\left(z_{1-\frac{\alpha}{2}} \sqrt{V_0} + z_{1-\beta} \sqrt{V_1} \right)^2}{\mu_t (1 - R^2) e^{\beta_0} {b_1}^2} \right] \qquad \{4.12.4\}$$

where V_0 and V_1 are the variances of the estimated regression coefficient for the independent variable under the null and alternative hypotheses respectively. These are defined below. For the one sided test, $z_{1-\alpha}$ would be used in place of $z_{1-\frac{\alpha}{2}}$.

The power is calculated by re-arrangement of the above formula to give the following equation:

$$Power = \Phi\left[\left(\left(\sqrt{\mu_t(1-R^2)e^{\beta_0}{b_1}^2N}\Big/\sqrt{\varphi}\right) - z_{1-\frac{\alpha}{2}}\sqrt{V_0}\right)\Big/\sqrt{V_1}\right] \qquad \{4.12.5\}$$

with a similar re-arrangement yielding the following equation for the test significance level:

$$\alpha = 2\left\{1 - \Phi\left[\left(\left(\sqrt{\mu_t(1-R^2)e^{\beta_0}{b_1}^2N}\Big/\sqrt{\varphi}\right) - z_{1-\beta}\sqrt{V_1}\right)\Big/\sqrt{V_0}\right]\right\} \qquad \{4.12.6\}$$

In the above equations, V_0 and V_1 are defined by the choice of distribution for the independent variable (X_1). These are outlined below:

1. Normal Distribution for X_1
   The calculation requires the mean μ and standard deviation σ of X_1. These are used to calculate V_0 and V_1 as follows:

$$V_0 = \frac{1}{\sigma^2} \qquad \{4.12.7\}$$

$$V_1 = \frac{1}{\sigma^2}e^{-\left(b_1\mu + \frac{b_1^2\sigma^2}{2}\right)} \qquad \{4.12.8\}$$

2. Exponential Distribution for X_1
   The calculation requires the exponential mean λ. These are used to calculate V_0 and V_1 as follows:

$$V_0 = \frac{1}{\lambda^{-2}} \qquad \{4.12.9\}$$

$$V_1 = \frac{|(\lambda - b_1)|^3}{\lambda} \qquad \{4.12.10\}$$

3. Uniform Distribution for X_1
   The calculation requires the minimum (A) and maximum (B) of the uniform distribution interval. These are used to calculate V_0 and V_1 as follows:

$$V_0 = \frac{12}{(B-A)^2} \qquad \{4.12.11\}$$

$$V_1 = \frac{w}{w(w_b) - w_a^2} \qquad \{4.12.12\}$$

where

$$w = \frac{e^{b_1 B} - e^{b_1 A}}{(D-C)b_1} \qquad \{4.12.13\}$$

$$w_a = \frac{e^{b_1 B}(b_1 B - 1) - e^{b_1 A}(b_1 A - 1)}{(B-A)b_1^2} \qquad \{4.12.14\}$$

$$w_b = \frac{e^{b_1 B}(2 - 2b_1 B + b_1^2 B^2) - e^{b_1 A}(2 - 2b_1 A + b_1^2 A^2)}{(B-A)b_1^3} \qquad \{4.12.15\}$$

4. Binomial Distribution for X_1
   The calculation requires the proportion of the variable p e.g. $p = 0.5$ if 50% of variable were treatment and 50% controls. These are used to calculate V_0 and V_1 as follows:

$$V_0 = \frac{1}{p(1-p)} \qquad \{4.12.16\}$$

$$V_1 = \frac{1}{(1-p)} + \frac{1}{pe^{b_1}} \qquad \{4.12.17\}$$

A closed form equation is not used to calculate the response rate ratio. Instead a search algorithm is used. This search algorithm calculates power at various values of the rate ratio until the desired power is reached.

4.12.3 Examples

Example 1: Validation example calculating required sample size for a given power
The following examples are taken from Signorini (1991) where a sample size calculation problem is conducted for a normally distributed independent variable, followed by a sensitivity analysis for the effect of changing response rate ratio and of differently distributed independent variables. Finally, an example of the calculation for the rate ratio is introduced.

The following steps outline the procedure for this example:

1. Open nQuery + nTerim 4.0 via the start menu or desktop shortcut. Click “New Fixed Term Test” from the top of the window. The “Study Goal and Design” window will appear as Figure 4.12.1 below.

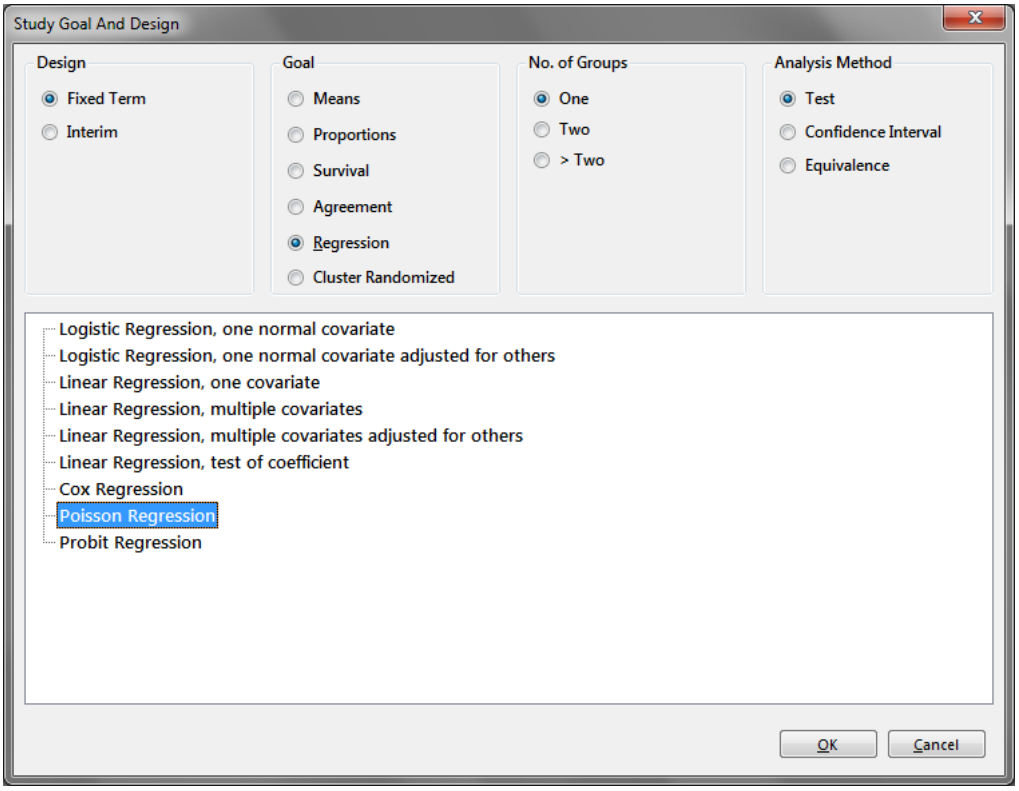


Figure 4.12.1. Study Goal and Design Window

2. In order to select the Poisson Regression design table, navigate through the Study Goal and Design Window by selecting "Fixed Term" as the Design, "Regression" as the Goal, "One" as the Number of Groups and "Test" as the Analysis Method. Then click "OK" and the test window will appear as per Figure 4.12.2.

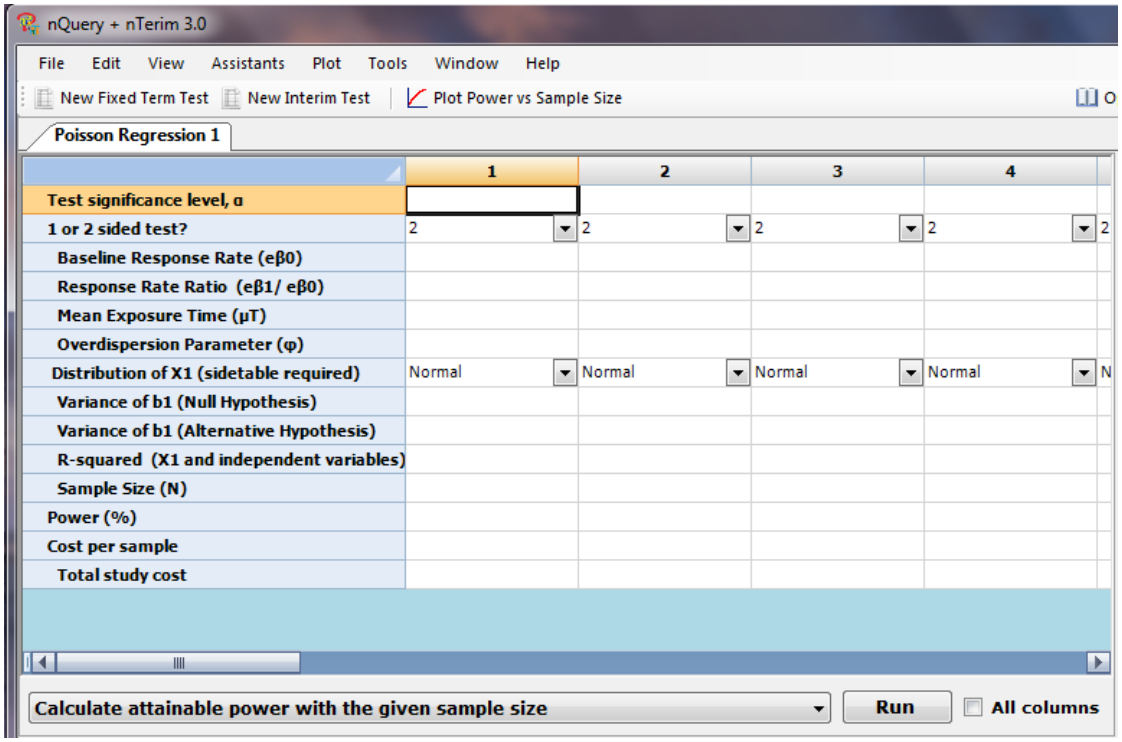


Figure 4.12.2. Poisson Regression Test Table

At the bottom will be the "Normal Side Table" to define a normally distributed independent variable. This will appear as in Figure 4.12.3.

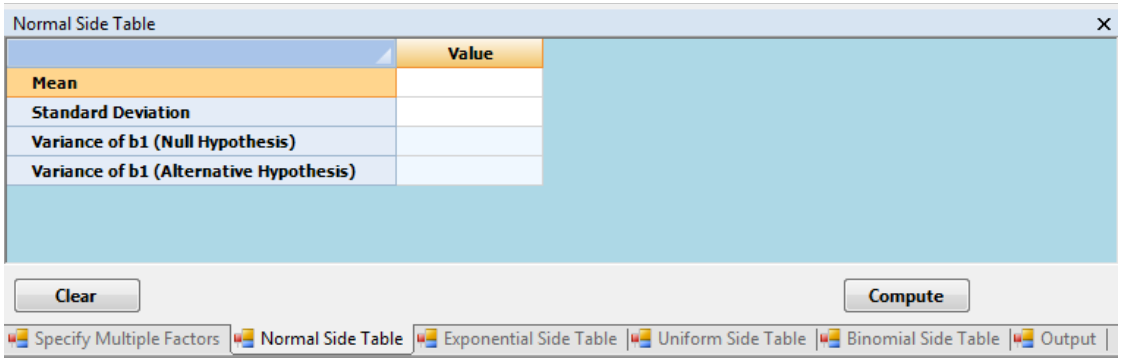


Figure 4.12.3. Normal Distribution Assistant table

There are three other side tables, one for an exponentially distributed variable, one for a uniform distributed variable and one for a binomial distributed variable. These can be selected in two manners. First using the dropdown menu for the "Distribution of X1" variable in the main table or secondly using the relevant tabs at the bottom of the side table. The other three sides tables will look as follows:

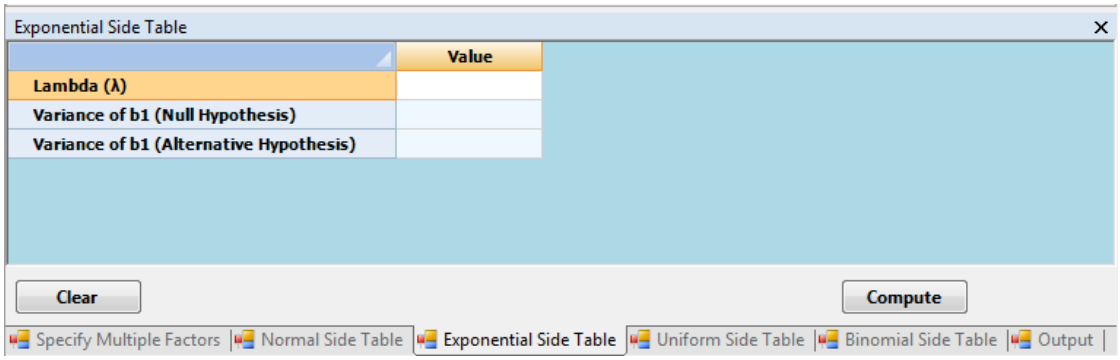


Figure 4.12.4. Exponential Distribution Assistant table

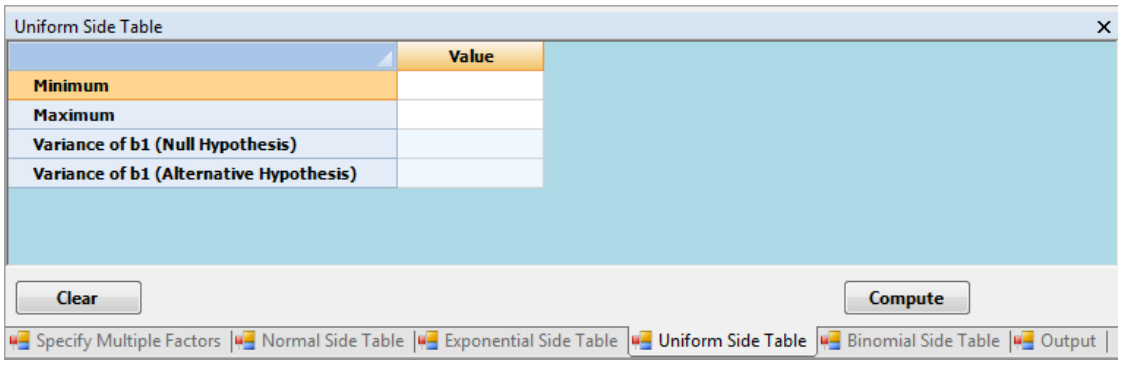


Figure 4.12.5. Uniform Distribution Assistant table

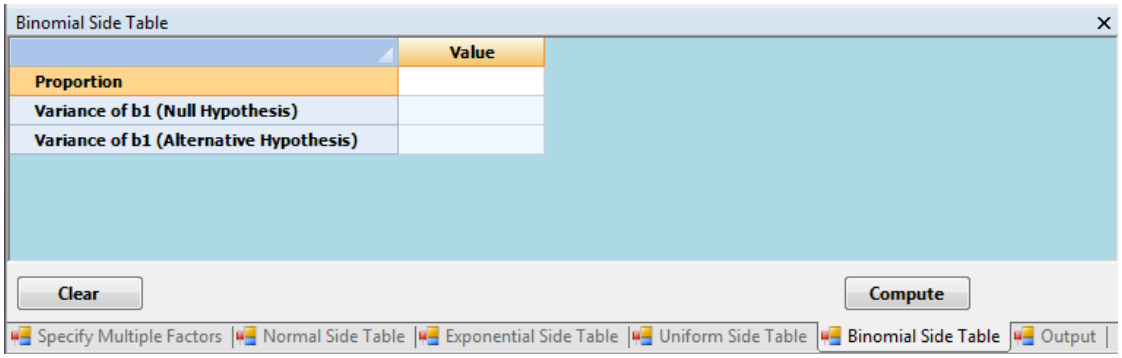


Figure 4.12.6. Binomial Distribution Assistant table

The first calculation will be for Sample Size for using a normally distributed variable from Table 2 of Signorini (1991).

3. First, enter 0.05 for the Test Significance level row, then select 1 for the “1 or 2 sided test?” variable dropdown menu, enter 1 and 0.5 for the “Baseline Response Rate” and “Response Rate Ratio” variables respectively, enter 1 for both the “Mean Exposure Time” and “Overdispersion Parameter” variables and enter 0 for “R-squared” parameter.

4. Finally, enter 95 in the Power row, the table will appear as per Figure 4.12.7.

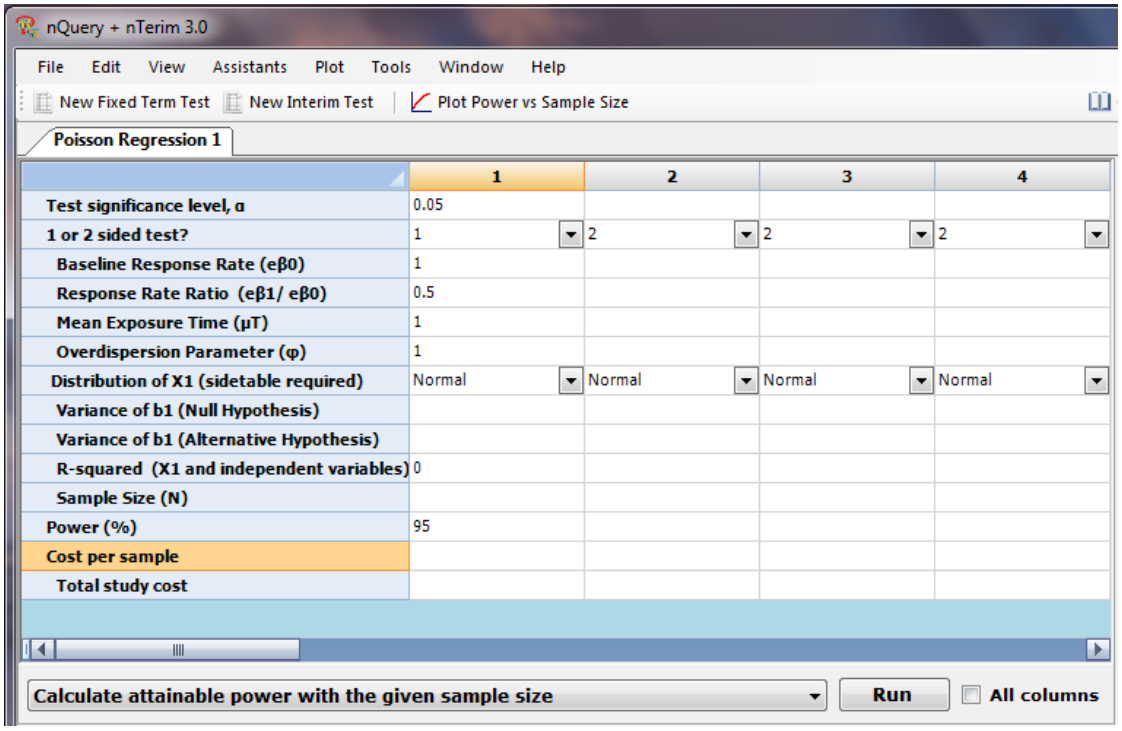


Figure 4.12.7. Values entered for Poisson Regression study design

5. Next, we fill in the side table for a normally distributed variable by entering 0 for the "Mean" and 1 for the "Standard Deviation" of the independent variable. This is displayed in Figure 4.12.8.

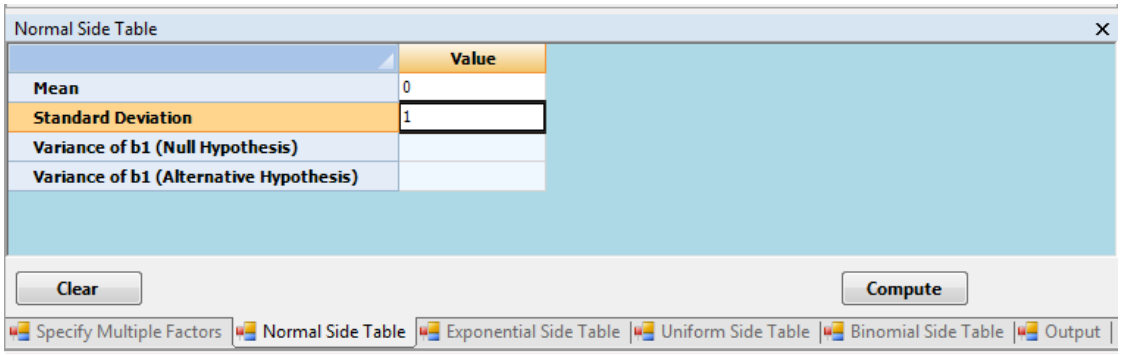


Figure 4.12.8. Values entered for the Normal Distribution side table

6. Next, click "Compute" on the side table and this will calculate and then transfer the "Variance of b1 (Null Hypothesis)" and "Variance of b1 (Alternative Hypothesis)" values into the main table as shown in Figure 4.12.9.

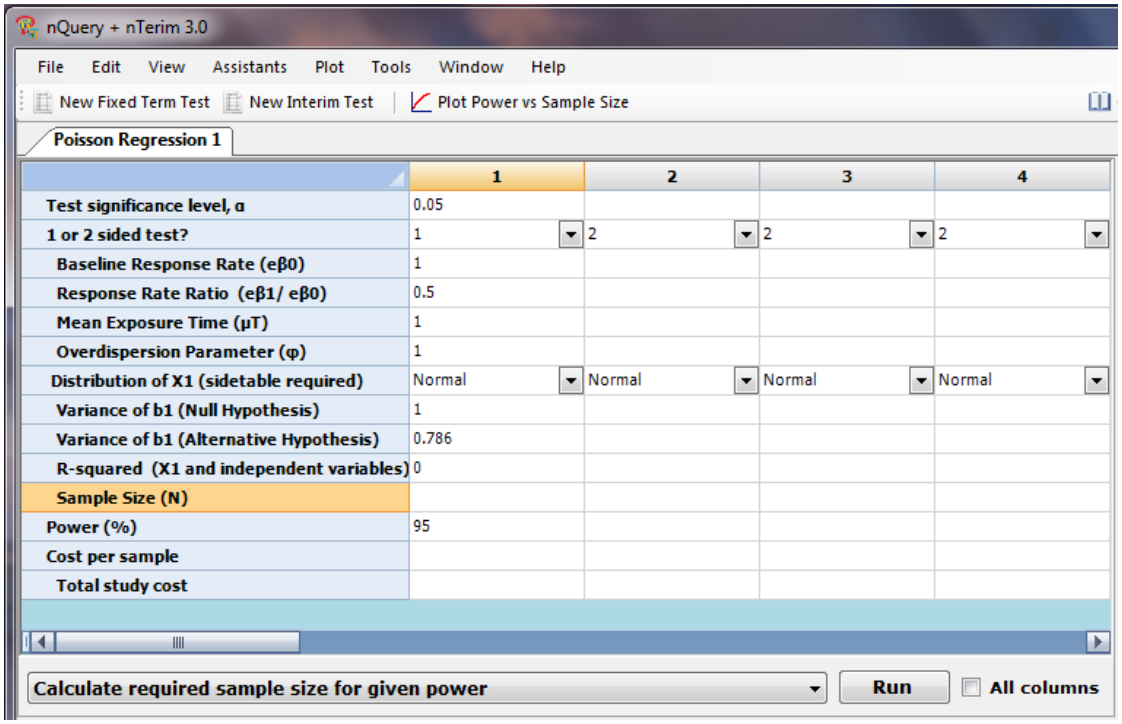


Figure 4.12.9. Values entered for Poisson Regression study design

7. Select “Calculate required sample size for given power” from the dropdown menu beside the “Run” button. Then click “Run”.

This will give a result of 21 for the sample size as in Figure 4.12.10. The result in Signorini (1991) is 20, however this is due to the two-decimal place rounding used in the paper.

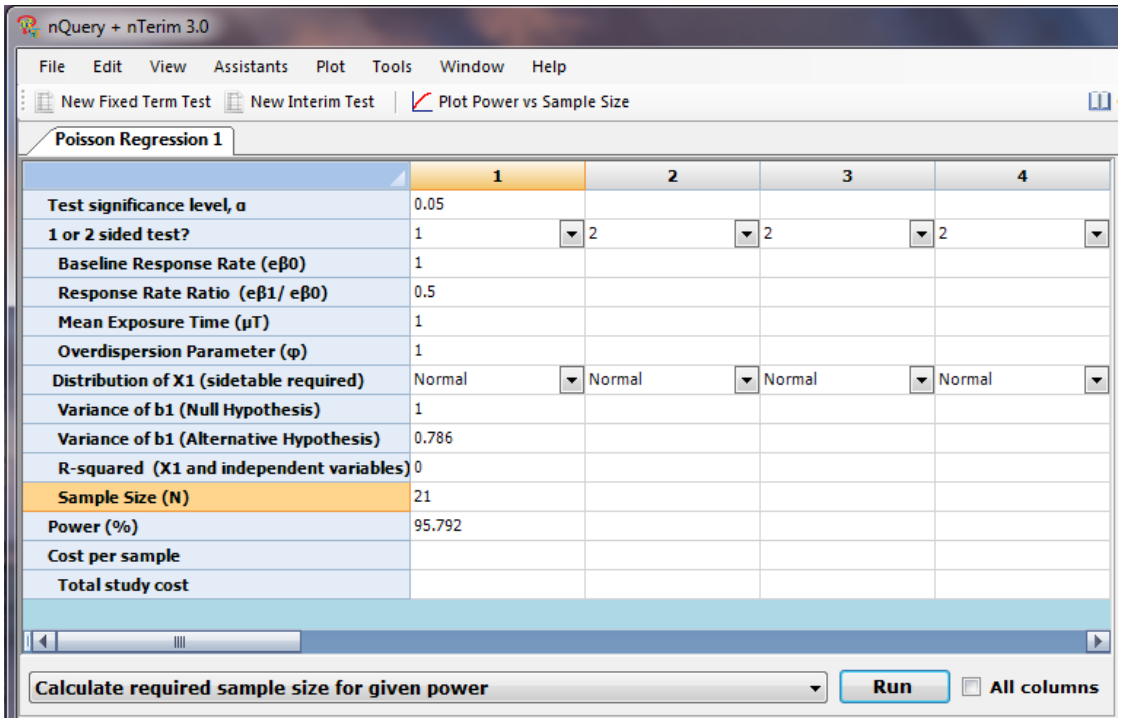
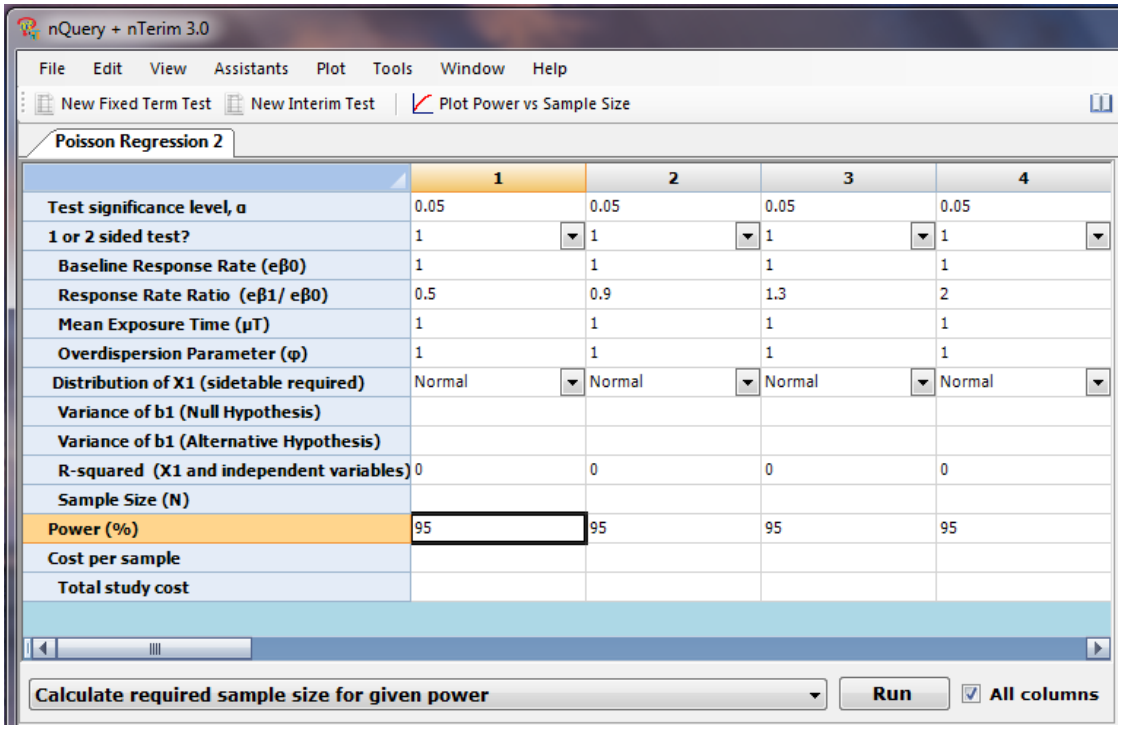


Figure 4.12.10. Completed Poisson Regression study design

The next calculation is a sensitivity analysis for sample size where we change the Response Rate Ratio and see its effect on sample size. These values are taken from Table 2 of Signorini (1991).

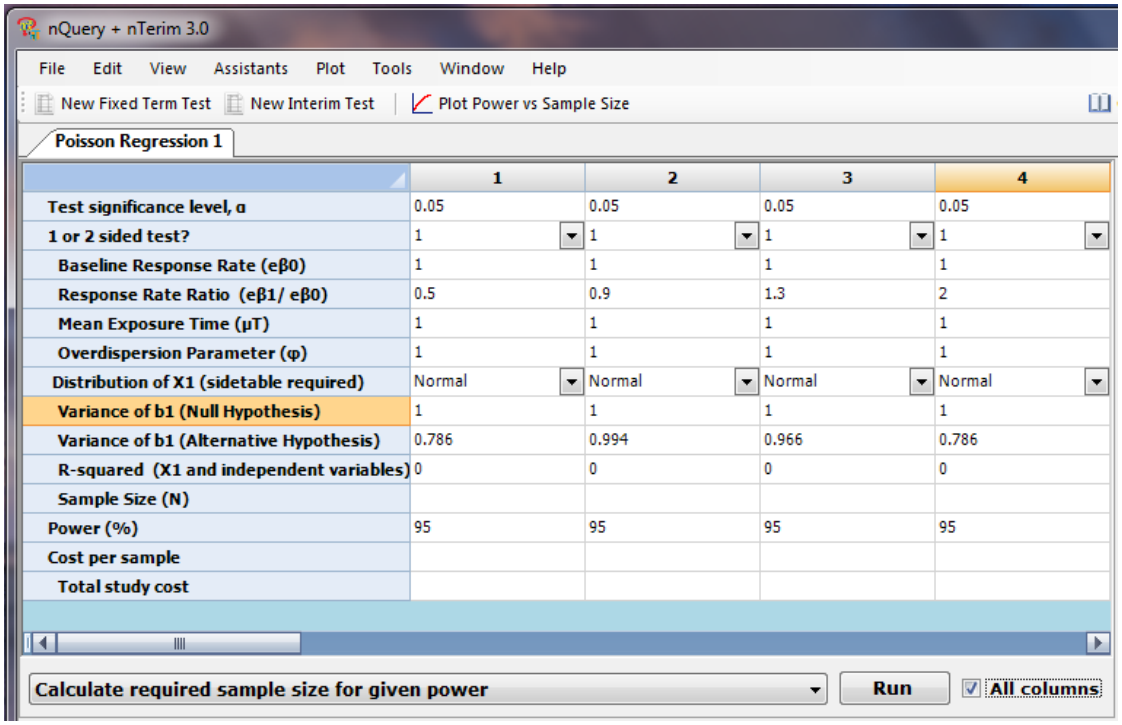
8. First, delete the two “Variance of b1” variables in the main table and re-enter 95 for Power in column 1. Then copy the values in column 1 across to columns 2, 3 and 4. Then, change the value for “Response Rate Ratio” in columns 2 to 4 from 0.5 to 0.9, 1.3 and 2 respectively. This will give a table as per Figure 4.12.11.



	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
1 or 2 sided test?	1	1	1	1
Baseline Response Rate (eβ0)	1	1	1	1
Response Rate Ratio (eβ1/ eβ0)	0.5	0.9	1.3	2
Mean Exposure Time (μT)	1	1	1	1
Overdispersion Parameter (φ)	1	1	1	1
Distribution of X1 (sidetable required)	Normal	Normal	Normal	Normal
Variance of b1 (Null Hypothesis)				
Variance of b1 (Alternative Hypothesis)				
R-squared (X1 and independent variables)	0	0	0	0
Sample Size (N)				
Power (%)	95	95	95	95
Cost per sample				
Total study cost				

Figure 4.12.11. Sensitivity analysis around the Response Rate Ratio

9. Using the Normal Distribution side table as before, recalculate the “Variance of b1 (Null Hypothesis)” and “Variance of b1 (Alternative Hypothesis)” variables for each column with a mean of zero and a standard deviation of one. This will give a table as per Figure 4.12.12.

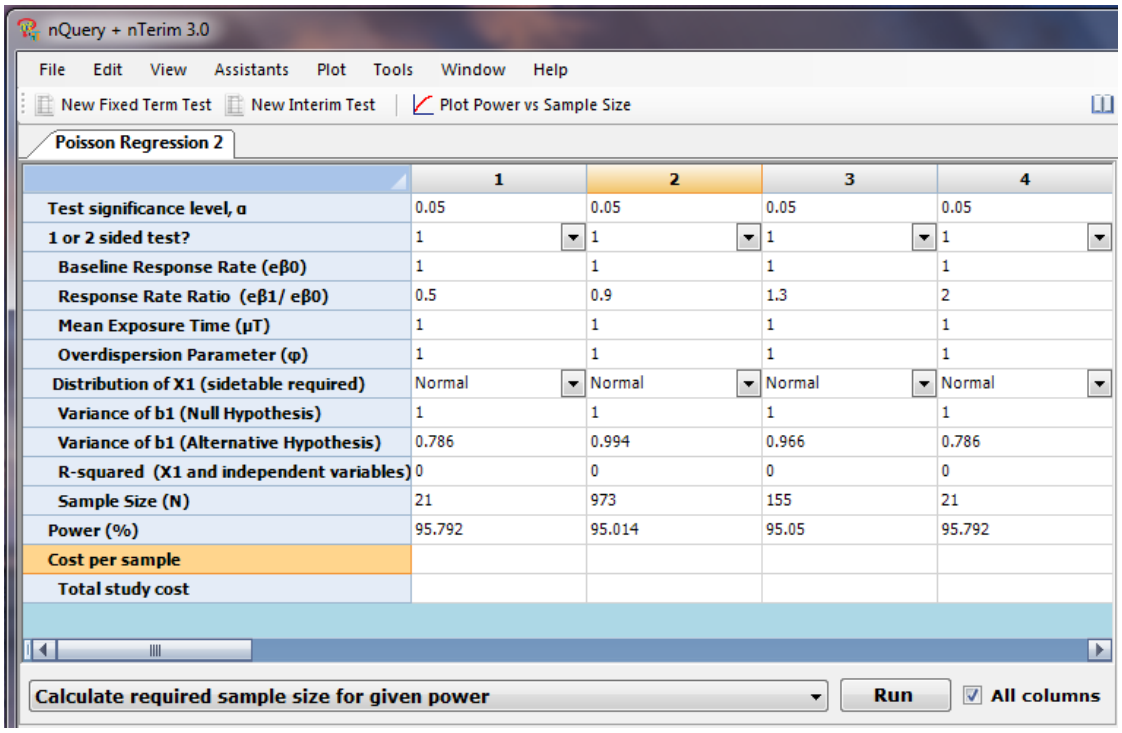


	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
1 or 2 sided test?	1	1	1	1
Baseline Response Rate (eβ0)	1	1	1	1
Response Rate Ratio (eβ1/ eβ0)	0.5	0.9	1.3	2
Mean Exposure Time (μT)	1	1	1	1
Overdispersion Parameter (φ)	1	1	1	1
Distribution of X1 (sidetable required)	Normal	Normal	Normal	Normal
Variance of b1 (Null Hypothesis)	1	1	1	1
Variance of b1 (Alternative Hypothesis)	0.786	0.994	0.966	0.786
R-squared (X1 and independent variables)	0	0	0	0
Sample Size (N)				
Power (%)	95	95	95	95
Cost per sample				
Total study cost				

Figure 4.12.12. Variance of b1 values entered for Sensitivity analysis

10. Select "Calculate required sample size for given power" from the dropdown menu beside the "Run" button. Next, tick the box to run "All Columns". Then click "Run".

This will give the resultant sample sizes of 21, 973, 155 and 21 sequentially as in Figure 4.12.13. Similar to the example above, the answers for column 2 to 4 differ from Signorini (1991) due to two-decimal place rounding.



	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
1 or 2 sided test?	1	1	1	1
Baseline Response Rate (eβ0)	1	1	1	1
Response Rate Ratio (eβ1/ eβ0)	0.5	0.9	1.3	2
Mean Exposure Time (μT)	1	1	1	1
Overdispersion Parameter (φ)	1	1	1	1
Distribution of X1 (sidetable required)	Normal	Normal	Normal	Normal
Variance of b1 (Null Hypothesis)	1	1	1	1
Variance of b1 (Alternative Hypothesis)	0.786	0.994	0.966	0.786
R-squared (X1 and independent variables)	0	0	0	0
Sample Size (N)	21	973	155	21
Power (%)	95.792	95.014	95.05	95.792
Cost per sample				
Total study cost				

Figure 4.12.13. Results from Sensitivity analysis

The effect of changing the response rate ratio on the relationship between Power and Sample size can be explored further using the “Plot Power vs Sample Size” button at the top of the screen.

11. Select the first column by clicking the 1 at the top of column 1. Then hold down the Shift key and click the 4 at the top of column 4. All four columns will now be highlighted.

12. Click the “Plot Power vs Sample Size” button at the top of the screen. This will give you a plot as displayed in Figure 4.12.14 which will show the relationship between power and sample size for each column. Right-click to add features such as a legend to the graph and double click elements for user options and editing.

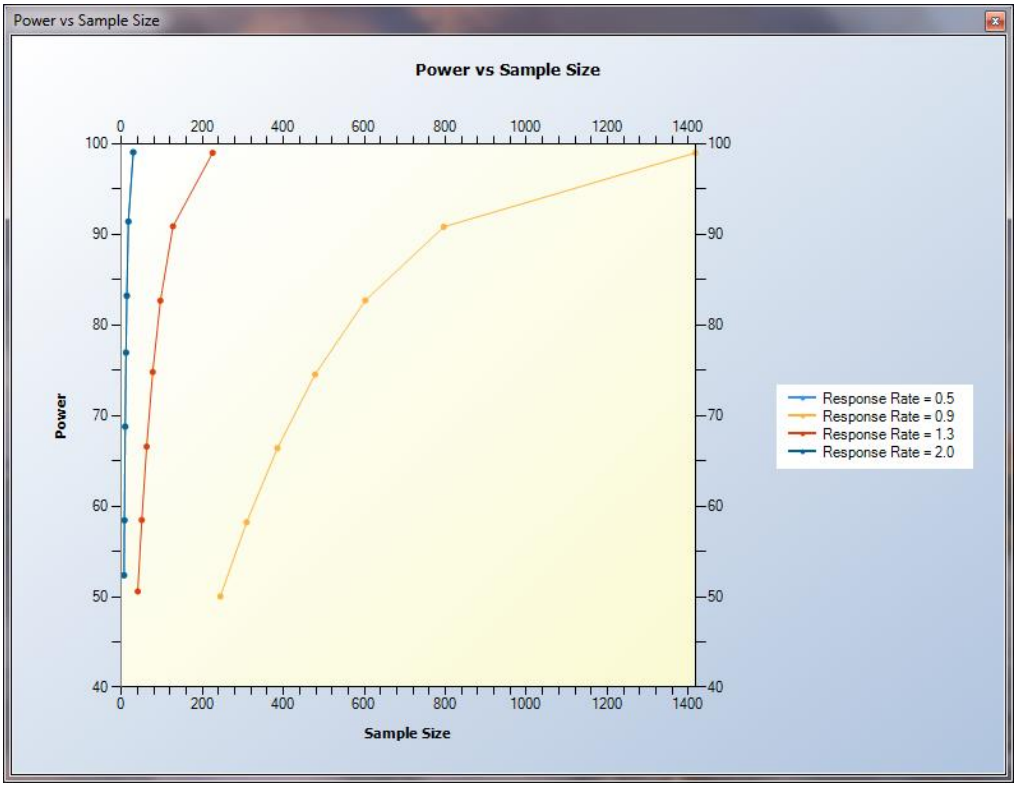


Figure 4.12.14. Power vs. Sample Size plot

13. Finally, by clicking on the desired study design column and going to the “Output” tab at the bottom of the screen, you can get an output statement for the results given by the current calculation as per Figure 4.12.15.

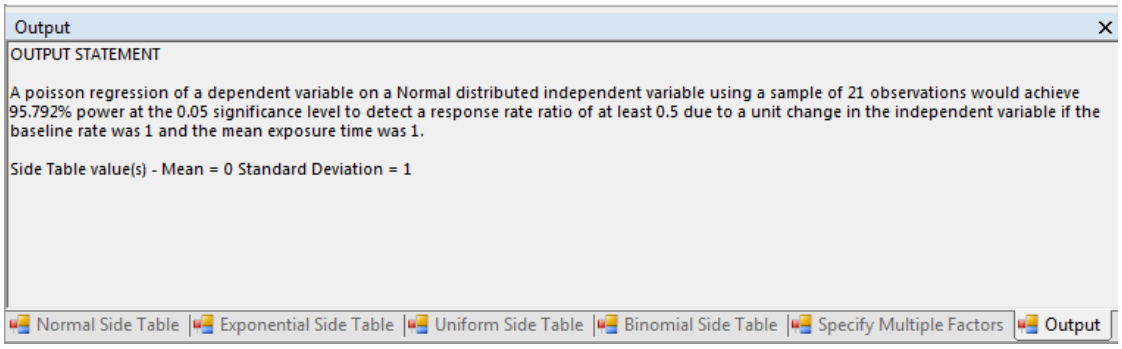


Figure 4.12.15. Study design Output statement

The next calculation is a sensitivity analysis for sample size where we change the distribution of the independent variable and see its effect on sample size. These values are taken from Table 2 of Signorini.

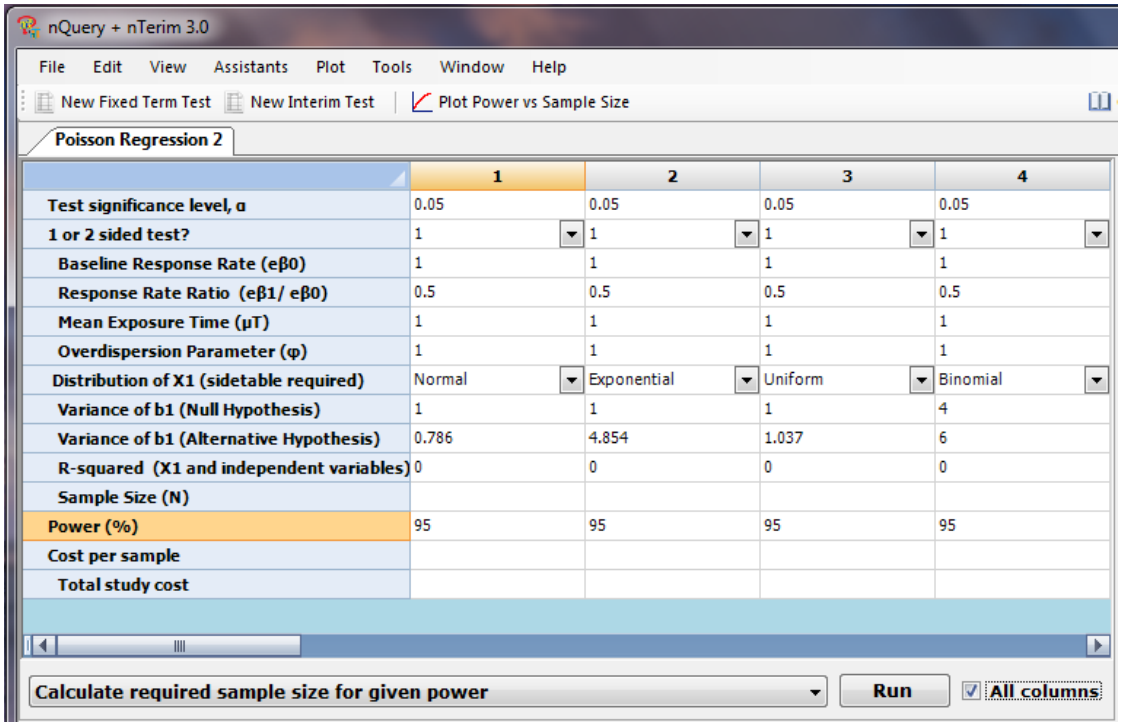
14. First, delete the two “Variance of b1” variables in the main table and the sample sizes for all four columns. Then, enter 0.5 for Response Rate Ratio and 95 for Power in all four columns. Then, select “Exponential”, “Uniform” and “Binomial” from the “Distribution of X1” variable for columns 2 to 4 respectively.

15. For the normally distributed column 1, enter a mean of 0 and standard deviation of 1 in the “Normal Side Table”, then click compute.

16. For the exponentially distributed column 2, enter a lambda of 1 in the “Exponential Side Table”, then click compute.

17. For the uniform distributed column 3, enter a minimum of -1.732 and maximum of 1.732 in the “Uniform Side Table”, then click compute.

18. For the binomial distributed column 4, enter a proportion of 0.5 in the “Binomial Side Table”, then click compute. This will give you a table that is displayed in Figure 4.12.16.

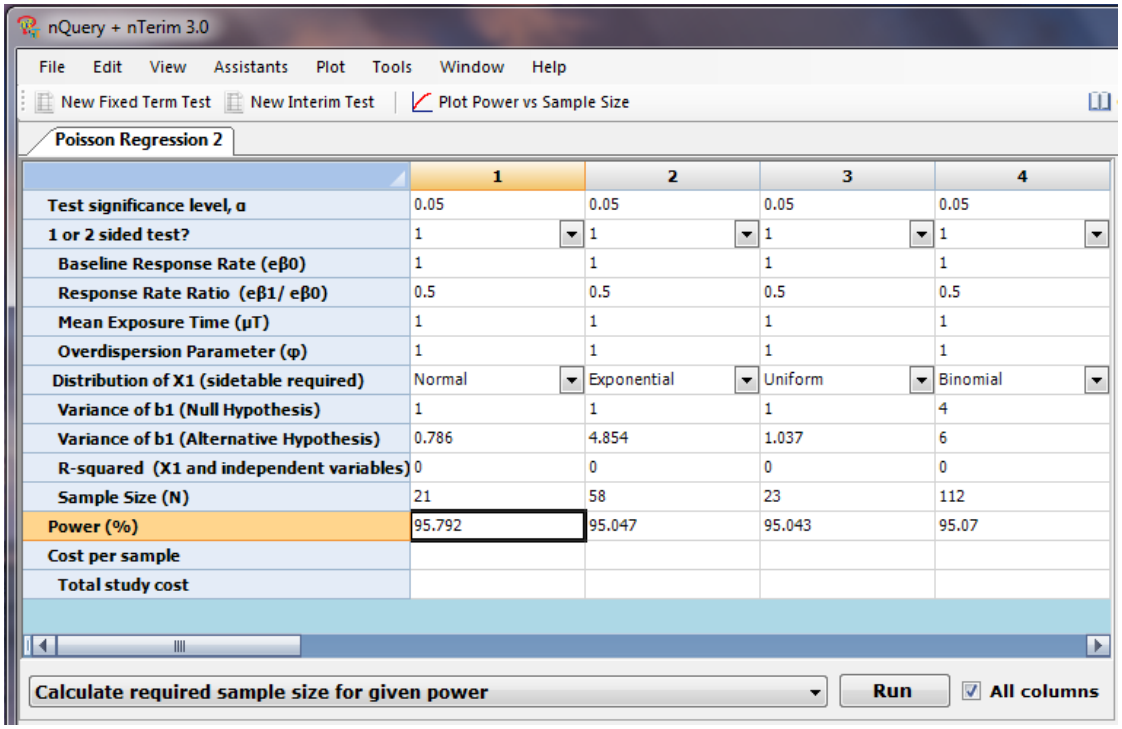


	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
1 or 2 sided test?	1	1	1	1
Baseline Response Rate (eβ0)	1	1	1	1
Response Rate Ratio (eβ1/ eβ0)	0.5	0.5	0.5	0.5
Mean Exposure Time (μT)	1	1	1	1
Overdispersion Parameter (φ)	1	1	1	1
Distribution of X1 (sidetable required)	Normal	Exponential	Uniform	Binomial
Variance of b1 (Null Hypothesis)	1	1	1	4
Variance of b1 (Alternative Hypothesis)	0.786	4.854	1.037	6
R-squared (X1 and independent variables)	0	0	0	0
Sample Size (N)				
Power (%)	95	95	95	95
Cost per sample				
Total study cost				

Figure 4.12.16. Sensitivity analysis around the Distribution of X1

19. Select "Calculate required sample size for given power" from the dropdown menu beside the "Run" button. Then tick the box to run "All Columns". Then click "Run".

This will give the resultant sample sizes of 21, 58, 325 and 112 sequentially as in Figure 4.12.17. Similar to the example above; the answers for column 2 to 4 differ from Signorini due to rounding.



	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
1 or 2 sided test?	1	1	1	1
Baseline Response Rate (eβ0)	1	1	1	1
Response Rate Ratio (eβ1/ eβ0)	0.5	0.5	0.5	0.5
Mean Exposure Time (μT)	1	1	1	1
Overdispersion Parameter (φ)	1	1	1	1
Distribution of X1 (sidetable required)	Normal	Exponential	Uniform	Binomial
Variance of b1 (Null Hypothesis)	1	1	1	4
Variance of b1 (Alternative Hypothesis)	0.786	4.854	1.037	6
R-squared (X1 and independent variables)	0	0	0	0
Sample Size (N)	21	58	23	112
Power (%)	95.792	95.047	95.043	95.07
Cost per sample				
Total study cost				

Figure 4.12.17. Results from second Sensitivity analysis

Example 2: Calculating required Response Rate Ratio for given power and sample size

Finally, an example is provided for calculating the response rate ratio given the sample size and power.

20. To do this, set sample size to 100 and power to 80 in column 1. Remove the two "Variance of b1" variable values from the main table and the response rate ratio value. Delete all other columns.
21. Next, select "Response Rate Ratio > 1 given sample size and power" from the dropdown menu beside the "Run" button.
22. Next, enter a mean of 0 and a standard deviation of 1 in the Normal Distribution side table. Then click "Compute".

Note that no figure is transferred into the main table, as the "Variance of b1 (Null Hypothesis)" and "Variance of b1 (Alternative Hypothesis)" statistics require the response rate ratio. These values will be displayed after the calculation is run.

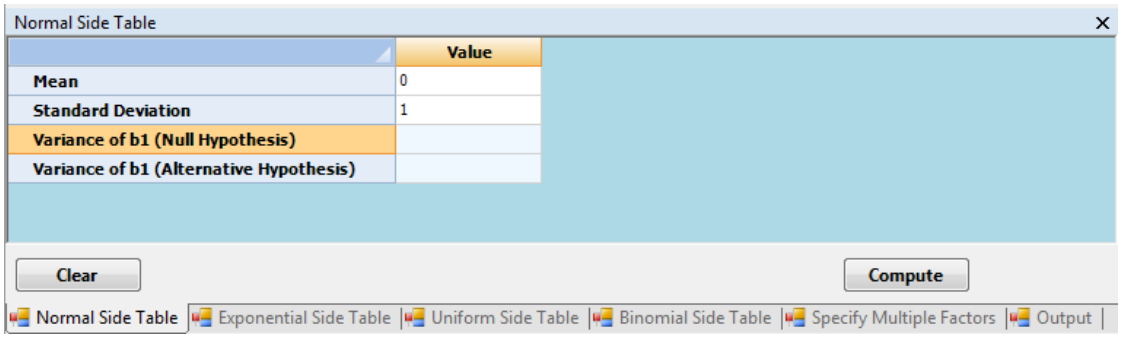


Figure 4.12.18. Normal Distribution Side Table

23. Click "Run". The response rate ratio will update to 1.281 as in Figure 4.12.19.

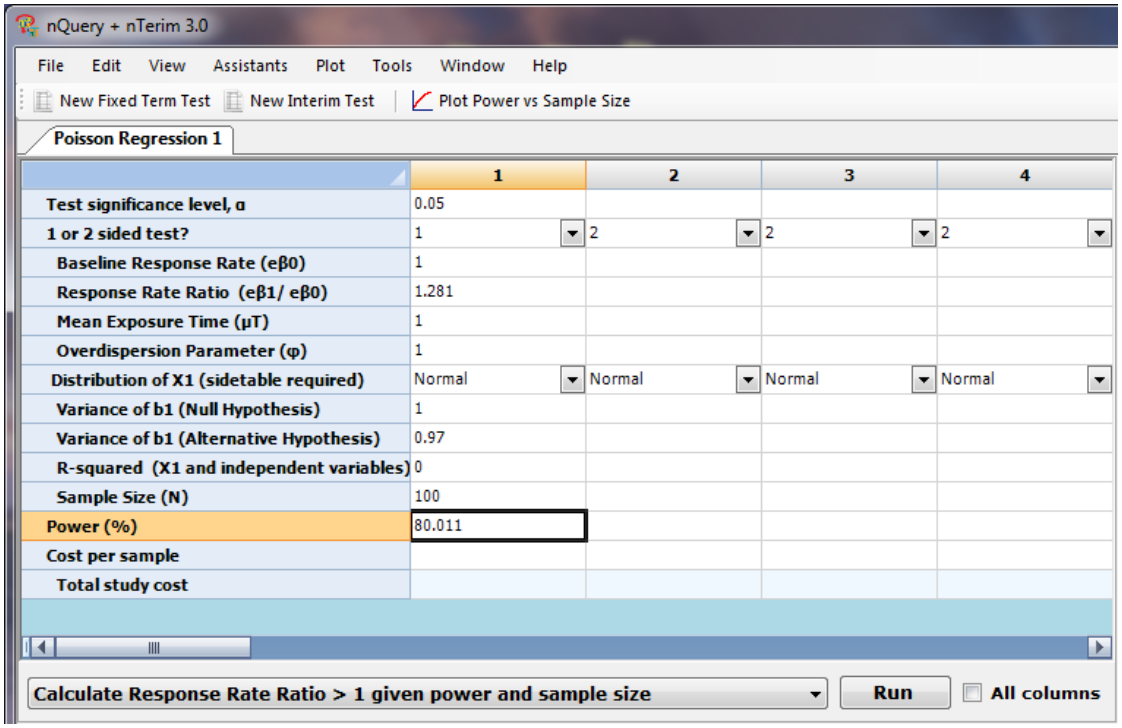


Figure 4.12.19. Completed Poisson Regression study design

Note that, for this example, no values are transferred into the main table, as the “Variance of b1 (Null Hypothesis)” and “Variance of b1 (Alternative Hypothesis)” statistics require the response rate ratio. **However, these figures will be saved in the system memory for calculation purposes for that column while it is selected.**

4.13. Probit Regression

4.13.1 Introduction

The Probit Regression Model is an analysis method often used to model the relationship between a dependent variable which is a proportion and one or more independent variables in clinical trials testing subject exposure e.g. to drugs, radiation, in lethal dose 50 (LD_{50}) trials. These are common in early stage animal trials for example.

This table facilitates the calculation of the power and sample size for an LD_{50} using probit analysis. Power and sample size is computed using the method outlined by Kodell et al. (2010).

4.13.2 Methodology

Probit analysis is often used in the study of relative potency between test and control treatments. The Probit Regression Model is given by the following equation:

$$Y = F^{-1}(P) = \beta_0 + \beta_1 \log_{10} D \qquad \{4.13.1\}$$

where $F^{-1}(P)$ is the cumulative normal distribution for the lethality proportions and D is dose of the substance of interest. For LD_{50} trials, we define $LD_{50}(T)$ as the dose which is lethal for 50% of the treatment group and $LD_{50}(C)$ as the dose which is lethal for 50% of the control group. The relative potency (ρ) is defined as follows to give the following formulae:

$$\rho = \frac{\mathrm{LD}_{50}(\mathrm{T})}{\mathrm{LD}_{50}(\mathrm{C})} \qquad \{4.13.2\}$$

$$\log_{10} \rho = \log_{10}(\mathrm{LD}_{50}(\mathrm{T})) - \log_{10}(\mathrm{LD}_{50}(\mathrm{C})) \qquad \{4.13.3\}$$

The mean and variance of $\log_{10} \rho$ are derived in Kodell et al. (2010) and are used to generate the following statistic for hypothesis testing:

$$\frac{\log_{10} \hat{\rho}}{\sqrt{Var(\log_{10} \hat{\rho})}} \qquad \{4.13.4\}$$

The sample size calculation for this Probit analysis is taken from Kodell et al. (2010). This table can be used to calculate the power, the sample size and the relative potency.

Calculations use a t-distribution approximation. To calculate power and sample size, the user must specify the test significance level (α), the number of dose levels, the sum of the probit weights, the slope of the Probit Regression Model (β_1) and the relative potency (ρ). The sum of the probit weights is calculated from the side table using the values of the target response proportions expected for each dose level. The slope of the probit regression model can also be optionally calculated from the side table if the user provides the number of doses for the control group associated with each target response proportion.

The formula for the sample size, using the appropriate simplifications and approximations outlined in Kodell et al. (2010), for the null hypothesis of $\rho = 1$ versus the alternative hypothesis of $\rho > 1$ (alternatively $\mathrm{H}_0: \log_{10} \rho = 0$ vs $\mathrm{H}_1: \log_{10} \rho > 0$) is as follows:

$$N = \left\lceil \frac{2\left(t_{f,1-\alpha} + t_{f,1-\beta}\right)^2}{(\log_{10}(\rho)\, \beta_1)^2 \sum_{i=1}^{g} w_i} \right\rceil \qquad \{4.13.5\}$$

where $t_{f,p}$ is the inverse cumulative t-distribution function with f degrees of freedom evaluated at probability p and $\sum_{i=1}^{g} w_i$ is the sum of the Probit weights. The values of f and w_i are evaluated as follows:

$$f = 2g - 3 \qquad \{4.13.6\}$$

$$w_i = \frac{\varphi\left(\Phi^{-1}(P_i)\right)^2}{P_i(1-P_i)}, \qquad i = 1,2 \ldots . g \qquad \{4.13.7\}$$

where g is the number of dose levels, $\varphi(x)$ is the density function of the normal distribution, Φ^{-1} is the cumulative distribution of the normal distribution and P_i is the target response proportion for each dose level.

The power is calculated by re-arrangement of the above formula to give the following equation:

$$Power = t_f^{-1}\left[\left(\sqrt{\frac{\left((\log_{10}(\rho)\,\beta_1)^2 \sum_{i=1}^{g} w_i\, N\right)}{2}}\right) - t_{f,1-\alpha}\right] \qquad \{4.13.8\}$$

where $t_f^{-1}(x)$ is the cumulative t-distribution function evaluated at f degrees of freedom. The relative potency is given by the following formulation:

$$\rho = 10^A \qquad \{4.13.9\}$$

where

$$A = \sqrt{\frac{2\left(t_{f,1-\alpha} + t_{f,1-\beta}\right)^2}{N(\beta_1)^2 \sum_{i=1}^{g} w_i}} \qquad \{4.13.10\}$$

4.13.3 Examples

Example 1: Validation example calculating required sample size for a given power
The following examples are taken from Kodell et al (2010) where a sample size calculation problem is conducted for a five dose level study, followed by two sensitivity analyses; one for the effect of power and relative potency on sample size and one for studies with a different number of dose levels. Finally, there will be a sample size calculation where the slope is calculated using the side table.

The following steps outline the procedure for this example:

1. Open nQuery + nTerim 4.0 via the start menu or desktop shortcut. Click “New Fixed Term Test” from the top of the window. The “Study Goal and Design” window will appear as Figure 4.13.1 below.

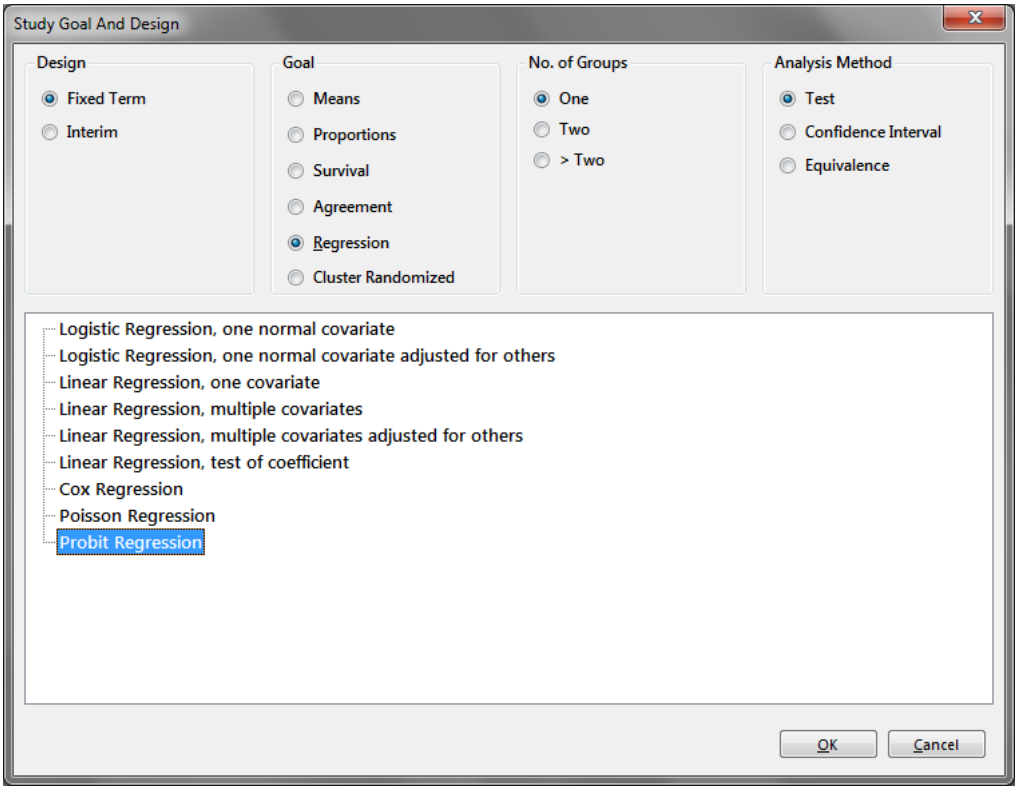


Figure 4.13.1. Study Goal and Design Window

2. In order to select the Probit Regression design table, navigate through the Study Goal and Design Window by selecting "Fixed Term" as the Design, "Regression" as the Goal, "One" as the Number of Groups and "Test" as the Analysis Method. Then click "OK" and the test window will appear as per Figure 4.13.2.

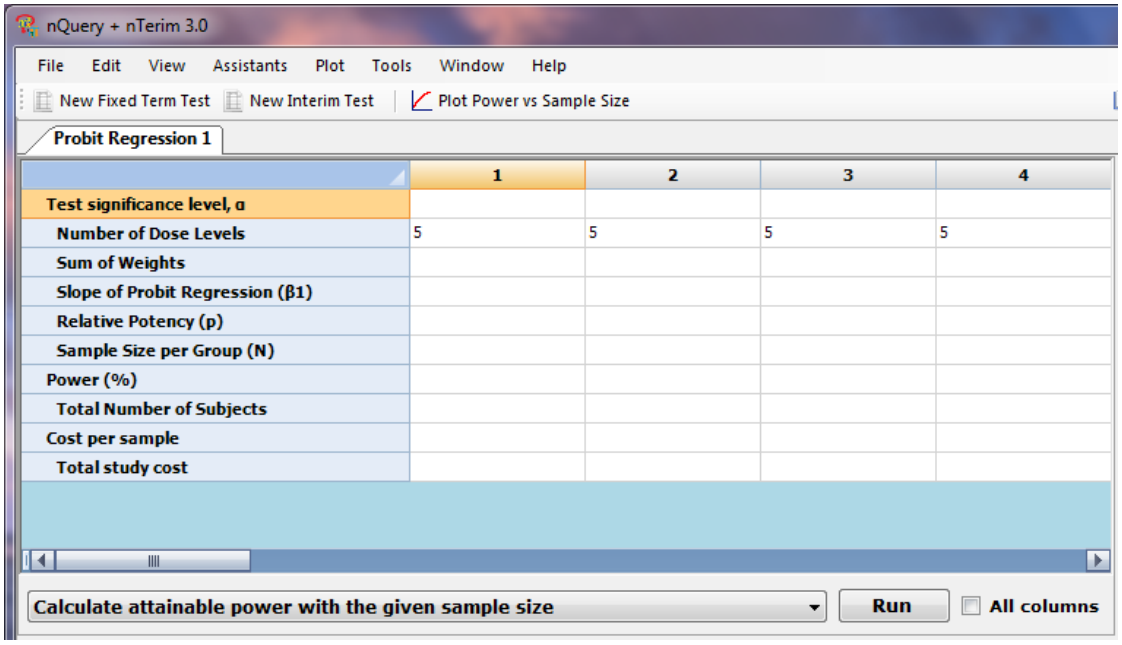


Figure 4.13.2. Probit Regression Test Table

At the bottom of the screen will be the "Probit Regression Side Table". This will appear as in Figure 4.13.3 below. The number of columns in the "Probit Regression Side Table" is defined by the "Number of Dose Levels" variable in the main table, where the default is set to 5.

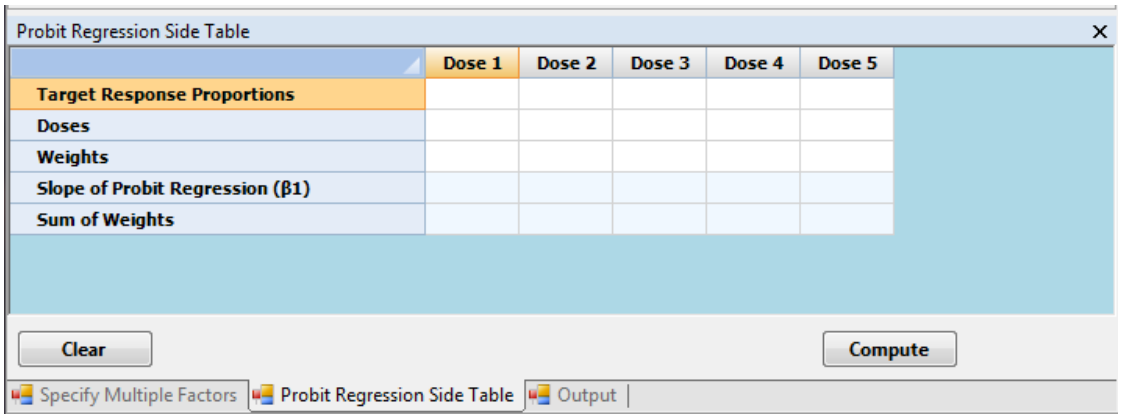


Figure 4.13.3. Probit Regression Assistant Table

The first calculation will be for Sample Size from Table 2 of Kodell et al (2010) for a five dose study design.

3. First enter 0.05 for the Test Significance level row, then enter 23.25 for "Slope of Probit Regression", and enter 1.1 for "Relative Potency".

4. Finally, enter 90 in the Power row and the table will appear as per Figure 4.13.4.

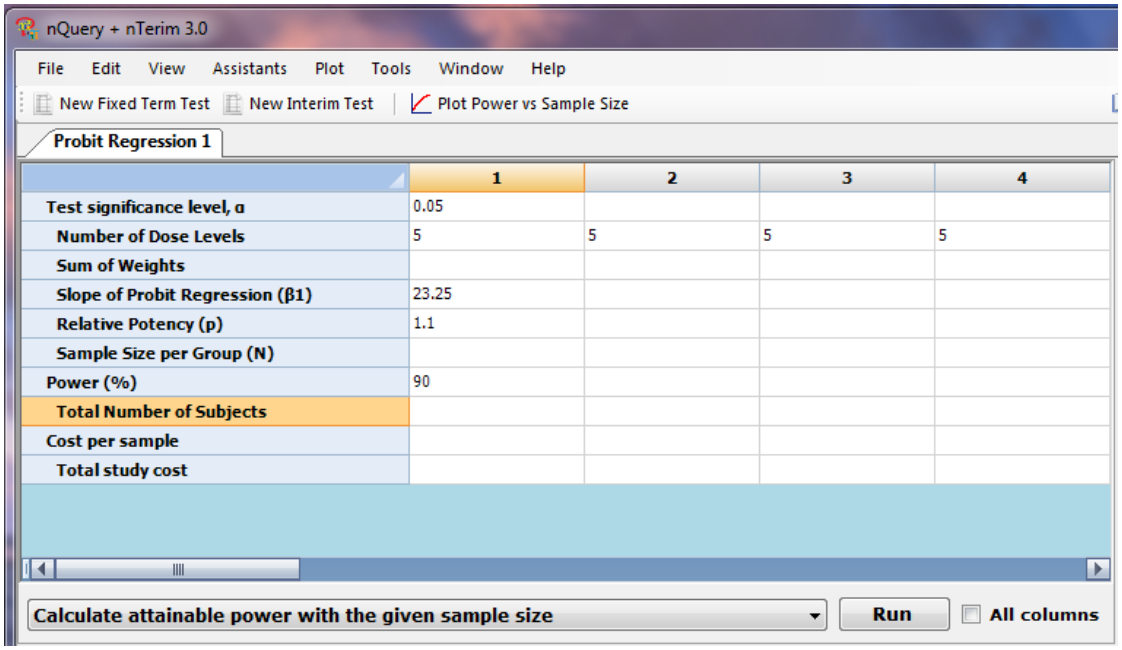


Figure 4.13.4. Values entered for Probit Regression study design

5. Next, we need to fill in the side table. To do this we will fill the Target Response Proportions row with 0.05, 0.275, 0.5 0.725, 0.95. This gives Figure 4.13.5.

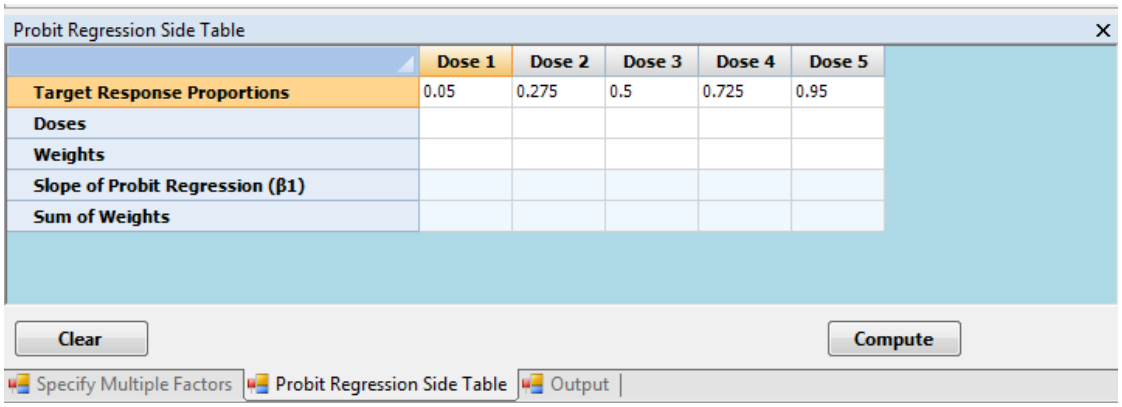
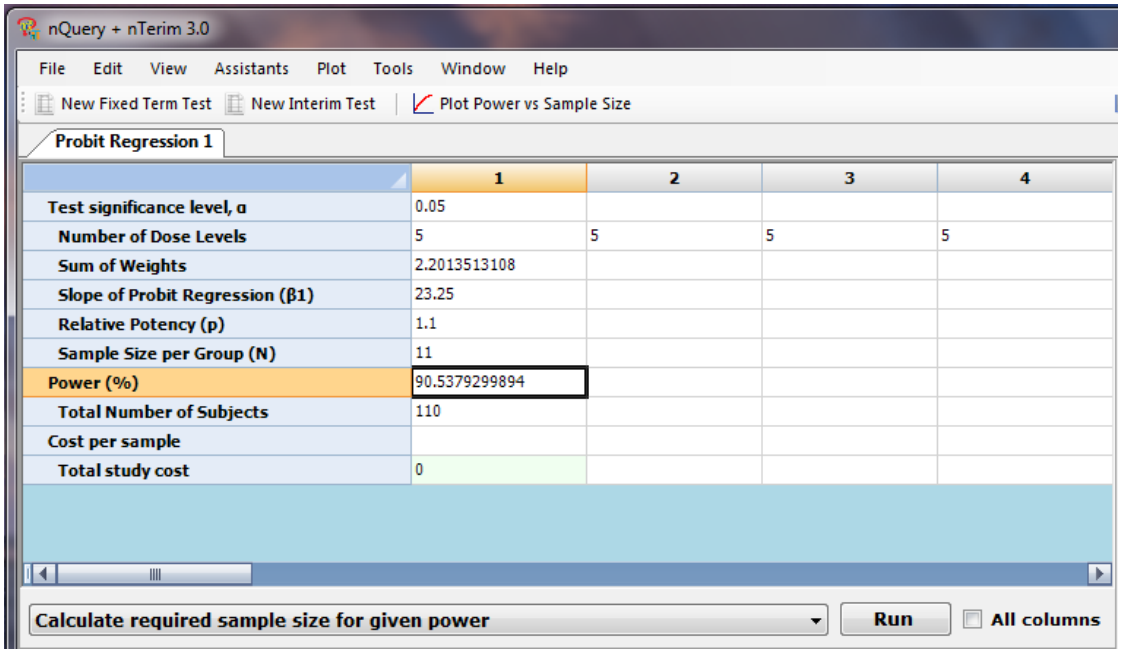


Figure 4.13.5. Values entered for Probit Regression Assistant table

6. Next, click the “Compute” button at the bottom of the Probit Regression side table and this will calculate and transfer the “Sum of Weights” value to the main table.

7. Select “Calculate required sample size for given power” from the dropdown menu beside the “Run” button. Then click “Run”. This will give a result of 11 as per Kodell et al (2010) for the sample size per group as in Figure 4.13.6.

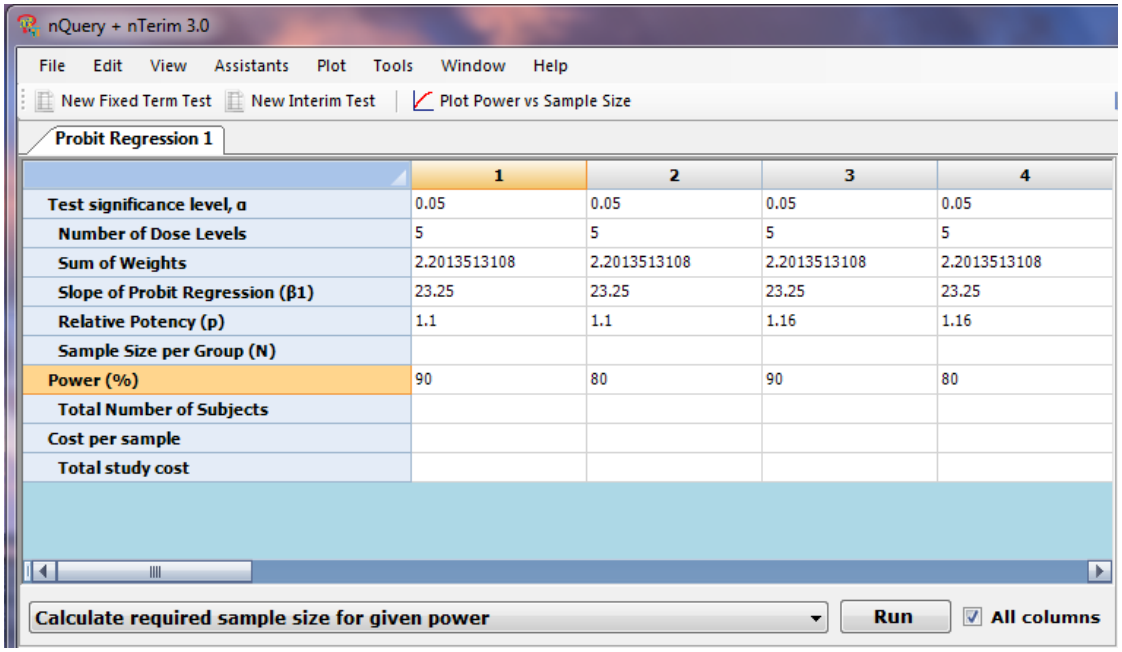


	1	2	3	4
Test significance level, α	0.05			
Number of Dose Levels	5	5	5	5
Sum of Weights	2.2013513108			
Slope of Probit Regression (β1)	23.25			
Relative Potency (p)	1.1			
Sample Size per Group (N)	11			
Power (%)	90.5379299894			
Total Number of Subjects	110			
Cost per sample				
Total study cost	0			

Figure 4.13.6. Completed Probit Regression study design

The next calculation is a sensitivity analysis for sample size where we change the power and the relative potency and see its effect on sample size. These values are taken from Table 2 of Kodell et al (2010).

8. First, enter 90 for Power in column 1 and delete the sample size values. Then copy the values in column 1 across to columns 2 to 4. Since the variables defined by the side table are the same for these 4 columns, the user can copy the previously calculated value rather than recalculating for each column separately. Then, change the value for Power to 80 in column 2 and 4 and change Relative Potency to 1.16 in column 3 and 4. This will give a table as per Figure 4.13.7.

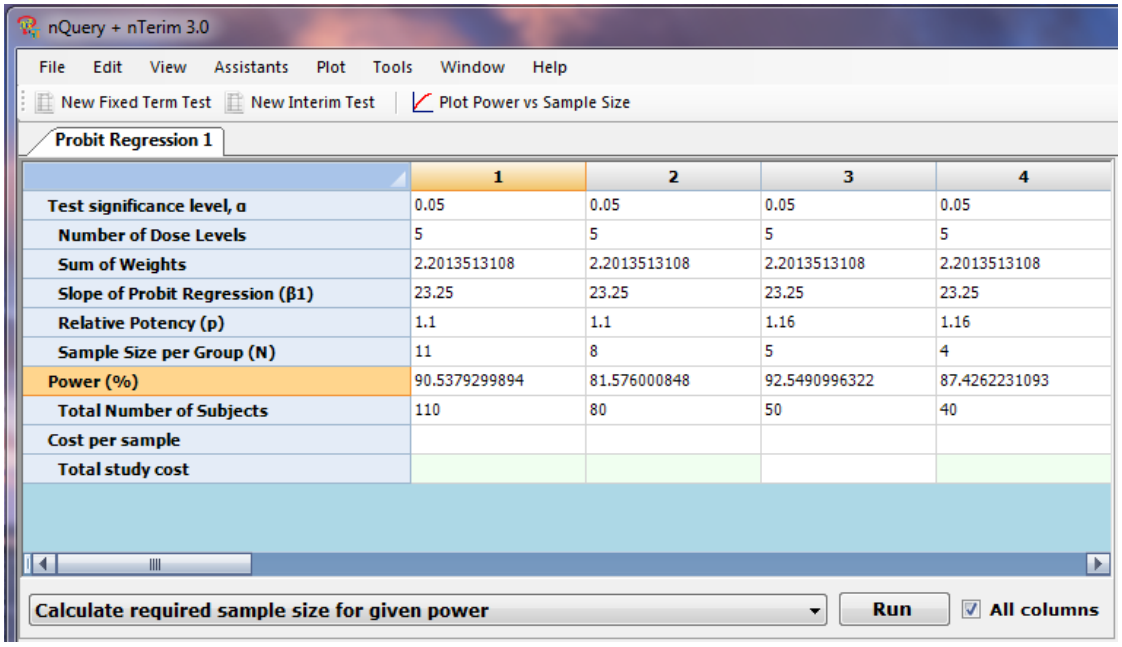


	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
Number of Dose Levels	5	5	5	5
Sum of Weights	2.2013513108	2.2013513108	2.2013513108	2.2013513108
Slope of Probit Regression (β1)	23.25	23.25	23.25	23.25
Relative Potency (p)	1.1	1.1	1.16	1.16
Sample Size per Group (N)				
Power (%)	90	80	90	80
Total Number of Subjects				
Cost per sample				
Total study cost				

Figure 4.13.7. Sensitivity analysis around Relative Potency and Power

9. Select “Calculate required sample size for given power” from the dropdown menu beside the “Run” button. Then tick the box to run “All Columns”. Then click “Run”.

This will give the resultant sample sizes of 11, 8, 5 and 4 per group, sequentially as per Kodell et al (2010) as in Figure 4.13.8 below.



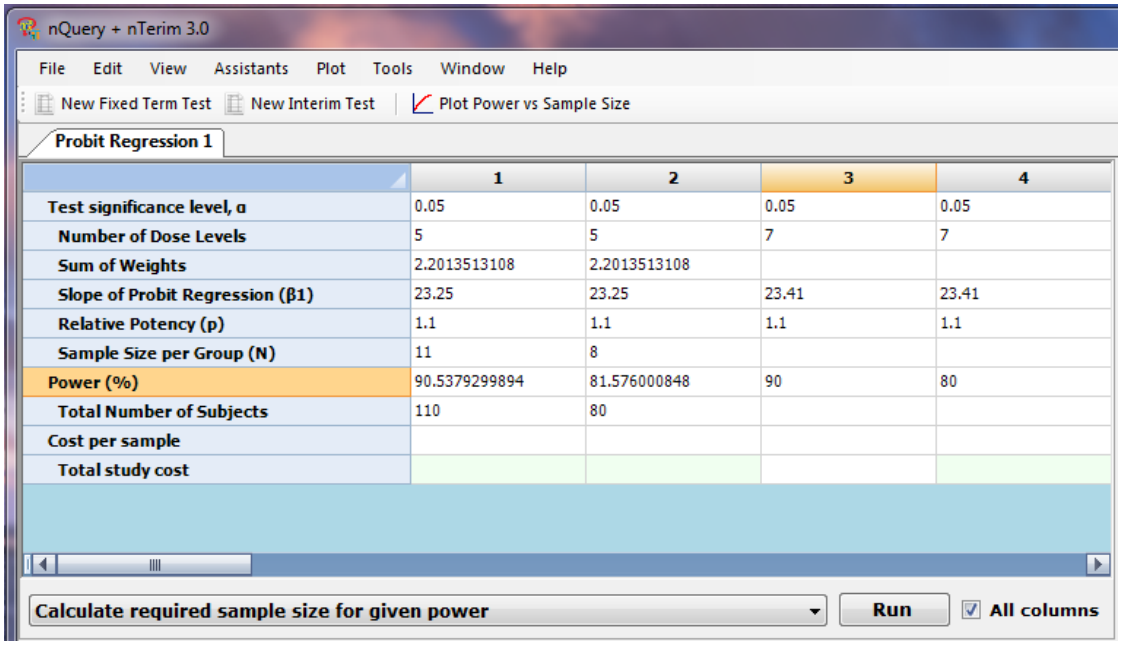
	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
Number of Dose Levels	5	5	5	5
Sum of Weights	2.2013513108	2.2013513108	2.2013513108	2.2013513108
Slope of Probit Regression (β1)	23.25	23.25	23.25	23.25
Relative Potency (p)	1.1	1.1	1.16	1.16
Sample Size per Group (N)	11	8	5	4
Power (%)	90.5379299894	81.576000848	92.5490996322	87.4262231093
Total Number of Subjects	110	80	50	40
Cost per sample				
Total study cost				

Figure 4.13.8. Results from Sensitivity analysis

The next calculation is a sensitivity analysis for sample size where we change “Number of Dose Levels”, where these values are taken from Table 2 of Kodell et al (2010).

10. For columns 3 and 4, delete the sum of weights and sample size values. Then, for the “Slope of Probit Regression”, replace 23.25 with 23.41. For “Relative Potency”, replace 1.16 with 1.1 and return the power values to 90 and 80.

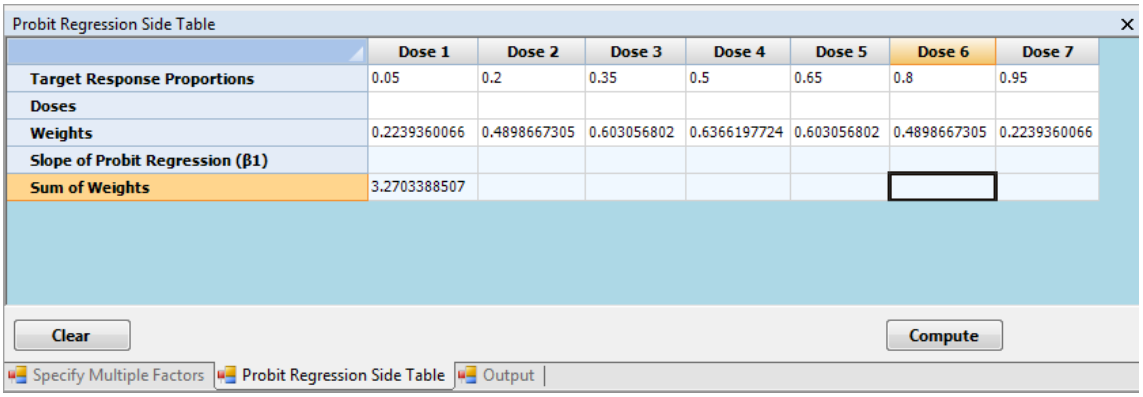
11. Finally, set “Number of Dose Levels” to 7 for both. This will give a table as displayed in Figure 4.13.9.



	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
Number of Dose Levels	5	5	7	7
Sum of Weights	2.2013513108	2.2013513108		
Slope of Probit Regression (β1)	23.25	23.25	23.41	23.41
Relative Potency (p)	1.1	1.1	1.1	1.1
Sample Size per Group (N)	11	8		
Power (%)	90.5379299894	81.576000848	90	80
Total Number of Subjects	110	80		
Cost per sample				
Total study cost				

Figure 4.13.9. Sensitivity analysis around the Number of Dose Levels

12. Next we need to complete the "Probit Regression Side Table" for both columns 3 and 4. Enter 0.05, 0.2, 0.35, 0.5, 0.65, 0.8, 0.95 in the Target Response Proportions and then click "Compute". This will give the results as shown in Figure 4.13.10.

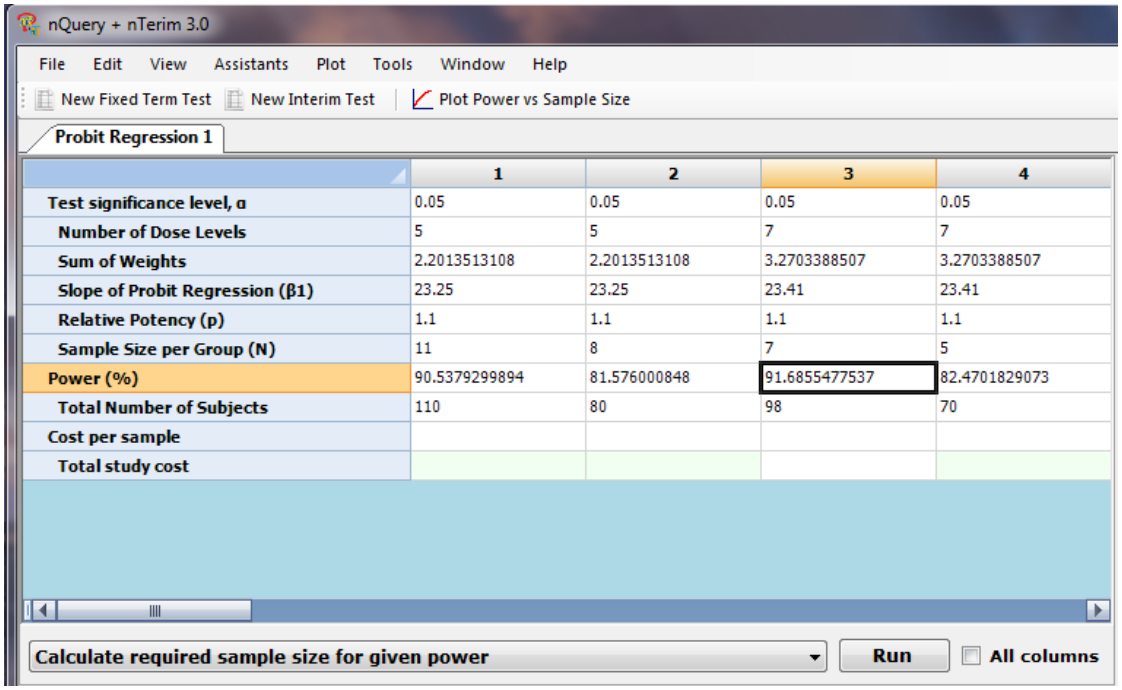


	Dose 1	Dose 2	Dose 3	Dose 4	Dose 5	Dose 6	Dose 7
Target Response Proportions	0.05	0.2	0.35	0.5	0.65	0.8	0.95
Doses							
Weights	0.2239360066	0.4898667305	0.603056802	0.6366197724	0.603056802	0.4898667305	0.2239360066
Slope of Probit Regression (β1)							
Sum of Weights	3.2703388507						

Figure 4.13.10. Completed Probit Regression Side Table for Sensitivity analysis

13. Select "Calculate required sample size for given power" from the dropdown menu beside the "Run" button. Then tick the box to run "All Columns". Then click "Run".

This will give the resultant sample sizes of 11, 8, 7 and 5 per group, sequentially as per Kodell et al (2010) and this is displayed in Figure 4.13.11.



	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
Number of Dose Levels	5	5	7	7
Sum of Weights	2.2013513108	2.2013513108	3.2703388507	3.2703388507
Slope of Probit Regression (β1)	23.25	23.25	23.41	23.41
Relative Potency (p)	1.1	1.1	1.1	1.1
Sample Size per Group (N)	11	8	7	5
Power (%)	90.5379299894	81.576000848	91.6855477537	82.4701829073
Total Number of Subjects	110	80	98	70
Cost per sample				
Total study cost				

Figure 4.13.11. Results from Sensitivity analysis

The effect of changing these values on the relationship between Power and Sample size can be explored further using the "Plot Power vs Sample Size" button at the top of the screen.

14. Select the first column by clicking the 1 at the top of column 1. Then hold down the Shift key and click the 4 at the top of column 4. All four columns will now be highlighted.

15. Click the "Plot Power vs Sample Size" button at the top of the screen. This will give you a plot as displayed in Figure 4.13.12 which will show the relationship between power and sample size for each column. Right-click to add features such as a legend to the graph and double click elements for user options and editing.

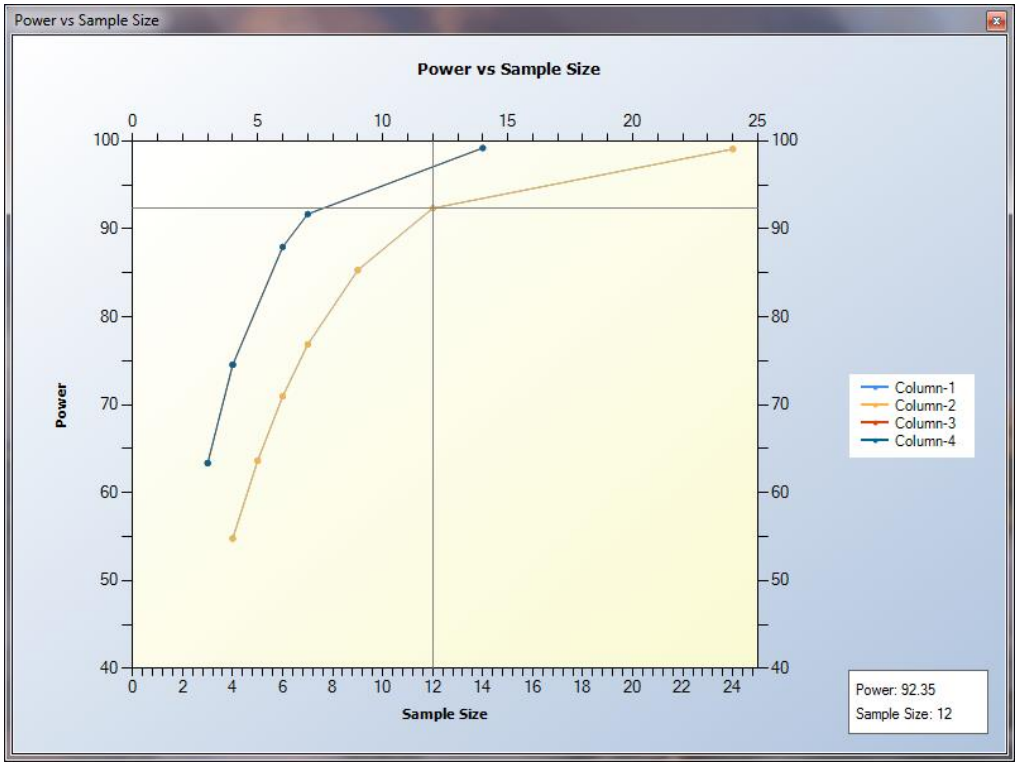


Figure 4.13.12. Power vs. Sample Size plot

16. Finally, by clicking on the desired study design column and going to the "Output" tab at the bottom of the screen, you can get an output statement for the results given by the current calculation as per Figure 4.13.13.

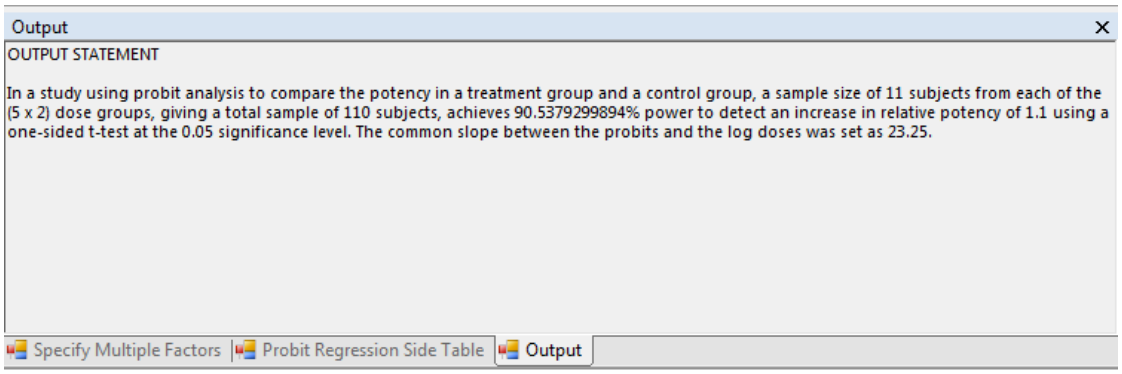


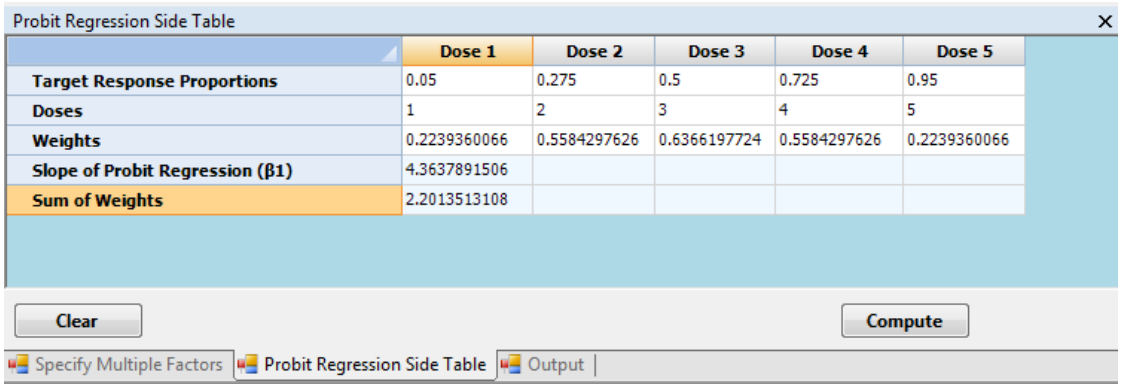
Figure 4.13.13. Study design Output statement

Finally, we will use the side table to calculate the regression slope. This will be done by specifying both the Target Response Proportions and the Number of Doses rows in the side table.

17. Delete columns 2 to 4. Then, in column 1, delete the slope, sum of weights and sample size values.

18. In the side table, enter the 0.05, 0.275, 0.5, 0.725 and 0.95 in the Target Response Proportions row. Enter 1 to 5 in the Number of Doses row.

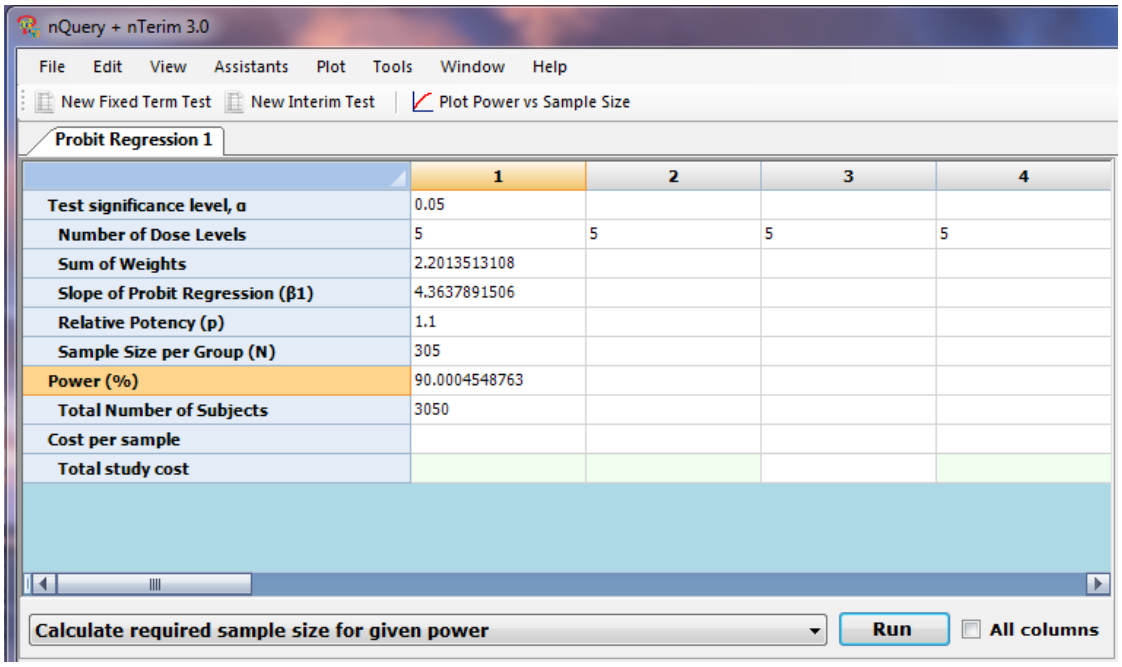
19. Click the "Compute" button and this will transfer a value 4.364 for the "Slope of Probit Regression" and 2.201 for the "Sum of Weights" into the main table.



	Dose 1	Dose 2	Dose 3	Dose 4	Dose 5
Target Response Proportions	0.05	0.275	0.5	0.725	0.95
Doses	1	2	3	4	5
Weights	0.2239360066	0.5584297626	0.6366197724	0.5584297626	0.2239360066
Slope of Probit Regression (β1)	4.3637891506				
Sum of Weights	2.2013513108				

Figure 4.13.14. Completed Probit Regression Side Table

20. Run the sample size calculation with the updated values for the slope and you get a "Sample Size per Group" of 305 as in Figure 4.13.15.



	1	2	3	4
Test significance level, α	0.05			
Number of Dose Levels	5	5	5	5
Sum of Weights	2.2013513108			
Slope of Probit Regression (β1)	4.3637891506			
Relative Potency (p)	1.1			
Sample Size per Group (N)	305			
Power (%)	90.0004548763			
Total Number of Subjects	3050			
Cost per sample				
Total study cost				

Figure 4.13.15. Completed Probit Regression study design

4.14. Higher-Order Cross-Over Design for Two Means – Superiority by a Margin– Using Differences

4.14.1 Introduction

In a cross-over design, each subject receives all treatments at least once with the objective of measuring study differences among the treatments. The term "cross-over" derives from the most common two treatments case, in which each subject crosses over from one treatment to another. The design assumes a washout period between treatments during which the response returns to its baseline. If this does not occur, there is said to be a carryover effect.

Cross-over designs are popular due to an expected increase in precision, since each subject effectively acts as their own control, and the concomitant reduction in study size. However, they do require more complex statistical analysis than parallel designs and suffer from complications and strong assumptions related to time effects, carry-over effects and patient attrition. These issues and the statistical methodology are discussed more thoroughly in Chow et al. (2008) and Chen et al. (1997).

In this table, we look beyond the most common 2x2 design to higher-order designs where there can be up to four sequences of treatments and/or measured time periods

This table facilitates the calculation of the power and sample size for four higher-order designs using the "additive model", namely the 2x3, 2x4, 4x2 and 4x4 designs (discussed below) for superiority by a margin hypotheses. Power and sample size is computed using the method outlined by Chen et al. (1997).

4.14.2 Methodology

Before discussing statistical methodology, the structure of the four additional cross-over designs is given below. A cross-over design consists of two parameters: the sequences i.e. the pattern of treatments given to a patient and periods i.e. the number of time periods of which the outcome variable is measured. In the numeric summary, the AxB design indicates A number of sequences and B number of periods. Note that all of the following designs are for a two treatment design. The designs are as follows:

Table 4.14.1: Details of Higher-Cross Over Designs

Design Name	Advantage	Design
Balaam Design (4x2)	Allows estimation of intra-subject variabilities	Sequence Period1 Period2 1 A A 2 B B 3 A B 4 B A
Two-Sequence Dual Design (2x3)	Allows estimation of intra-subject variabilities	Sequence Period1 Period2 Period3 1 A B B 2 B A A
Four-Period with Two Sequences Design	Allows estimation of effect multiple sequential doses	Sequence Period1 Period2 Period3 Period4 1 A B B A 2 B A A B
Four-Period Design with Four Sequences	Allows estimation of effect multiple sequential doses	Sequence Period1 Period2 Period3 Period4 1 A A B B 2 B B A A 3 A B B A 4 B A A B

The statistical methodology for cross-over designs is based on the ANOVA methodology and is outlined in detail in Chow and Liu (2000) and Chow et al. (2008). Note that the method outlined here is based on additive model (i.e. from an analysis of variance using untransformed data) as opposed to the multiplicative model (i.e. from an analysis of variance using log-transformed data). The multiplicative model case is dealt with the equivalent ratios table.

We now briefly discuss the superiority by a margin hypothesis and effect of whether higher values of the outcome variables are considered a positive or negative effect.

Superiority by a margin hypothesis tests are a subset of one-sided directional tests. In these tests, the researcher specifies an estimate for the expected true difference δ (i.e. $\mu_{Treat} - \mu_{Control}$) between treatments and a superiority margin, which defines the difference above which would indicate practical or clinical significance. Superiority by a margin tests a non-zero null test where the null hypothesis relates to the relationship between the true difference and margin. The direction of these hypotheses is defined by whether higher values of the outcome variable are considered better or not. This is outlined below:

If higher means are considered better:

$$H_0: \delta \leq |SM|, H_1: \delta > |SM| \qquad \{4.14.1\}$$

where $|SM|$ is the **absolute value** of the entered superiority margin. If higher means are considered worse:

$$H_0: \delta \geq -|SM|, H_1: \delta < -|SM| \qquad \{4.14.2\}$$

Note that it is the **negative of the absolute value** used in the worse means case.

The sample size and power calculations for higher-cross over design are derived from the work of Chen et al. (1997). Note that Chen only covers the case of assessing equivalence under a mean ratio hypothesis; however this methodology can be easily extended to non-inferiority/superiority margin type analysis and to the usage of differences instead of ratios. This table can be used to calculate the power, the total sample size, the superiority margin, the test significance level, the true difference and the within-standard error.

Calculations use a t-distribution approximation. To calculate power and total sample size the user must specify the test significance level (α), whether higher means are better or worse, the higher-order cross-over design of interest, the true difference ($\mu_{Treat} - \mu_{Cont}$), the superiority margin (SM) and the within standard error (S_w) (which is derived from the square root of the within mean square error (WMSE) from the ANOVA analysis for the Cross-Over design, either from previous studies or pilot studies).

The formulae for the power, derived from the methods from Chen et al. (1997), follow. When **higher means are better** the formula for power is as follows:

$$1 - \beta = T_v \left[\left(\frac{(\delta - |SM|)}{S_w \sqrt{b/n}} \right) - t_{V,1-\alpha} \right] \qquad \{4.14.3\}$$

where T_V is the cumulative t-distribution function with V degrees of freedom, $t_{V,1-\alpha}$ is the inverse cumulative t-distribution evaluated at $1 - \alpha$, b is a design parameter (both V and b are defined by design choice, see below) and n is **average number of subjects per sequence** i.e. $N_{Total}/\#\ sequences$. If **higher means are worse** then the formula becomes

$$1 - \beta = 1 - T_v \left[-\left(\frac{(-\delta - |SM|)}{S_w \sqrt{b/n}} \right) + t_{V,1-\alpha} \right] \qquad \{4.14.4\}$$

V and b are defined by the study design and are given by following equations:
$Balaam\ (4x2): V = 4n - 3, b = 2$
$Two - Sequence\ Dual\ (3x2): V = 4n - 4, b = 3/4$
$Four\ Periods\ \&\ Two\ Sequences\ (4x2): V = 6n - 5, b = 11/20$
$Four\ Periods\ and\ Four\ Sequences\ (4x4): V = 12n - 5, b = 1/4$

A closed form equation is not used to calculate the total sample, the superiority margin, the test significance levels, the within standard error or the smallest true detectable difference.

Instead a search algorithm is used. The search algorithm calculates power at various values for the relevant parameter until the desired power is reached.

Note also that the total sample size calculated is the “exact” sample size (i.e. the first value of N_{Total} which achieves the desired power). However, for an equal sample size per sequence you need to round to the nearest 2 for the 2x3 and 2x4 designs and to the nearest four for the 4x2 and 4x4 designs.

4.14.3 Examples

Example 1: Validation example calculating required sample size for a given power
This table uses the same mechanics as the non-inferiority table and thus the example used for validation there is sufficient for this table (see Chapter 4.15). The following is an example of a sample size calculation with this table followed by an illustration of the effect of different design types on sample size.

The following steps outline the procedure for this example:

1. Open nQuery + nTerim 4.0 via the start menu or desktop shortcut. Click “New Fixed Term Test” from the top of the window. The “Study Goal and Design” window will appear as per Figure 4.14.1 below.

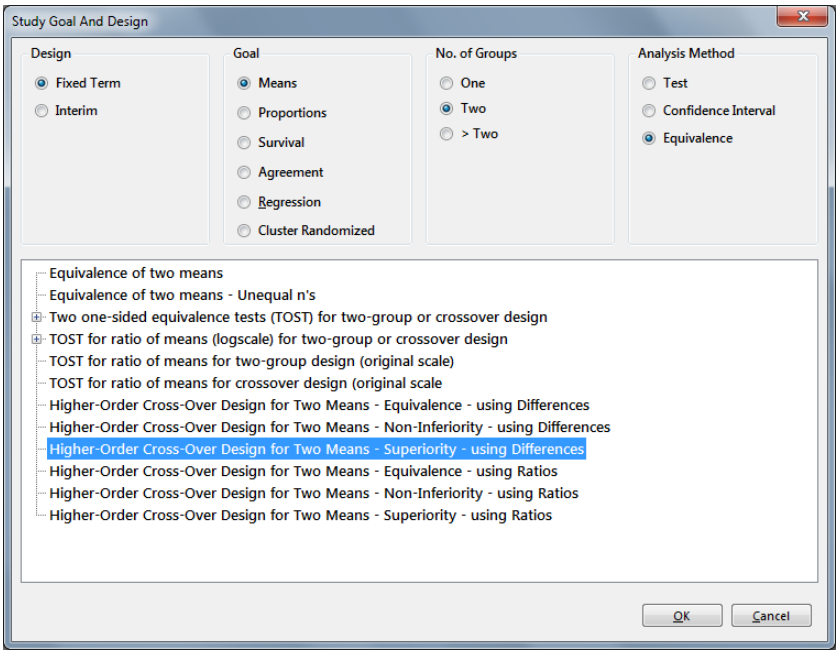


Figure 4.14.1. Study Goal and Design Window

2. In order to select the “Higher-Order Cross Over Design – Superiority by a Margin – Using Differences” design table, navigate through the Study Goal and Design Window by selecting "Fixed Term" as the Design, "Means" as the Goal, "Two" as the Number of Groups and "Equivalence" as the Analysis Method. Then click "OK" and the test window will appear as per Figure 4.14.2.

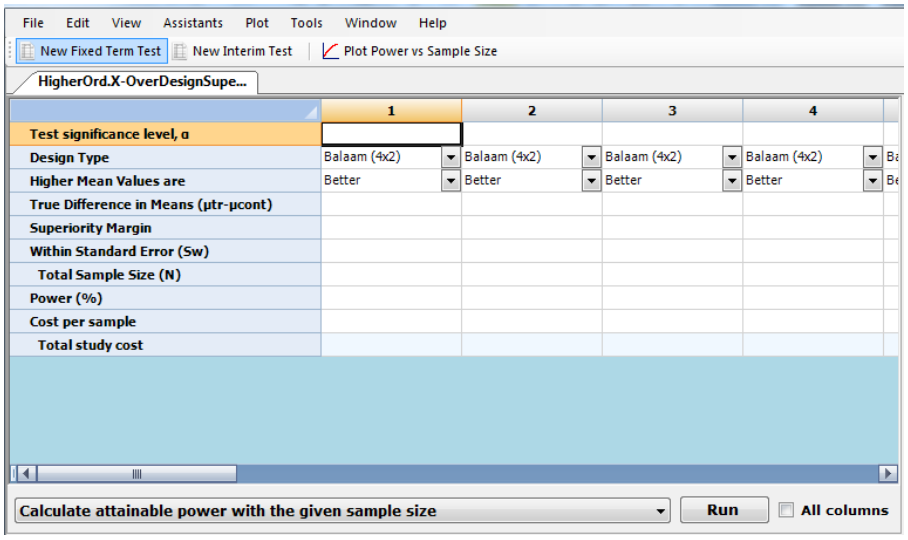


Figure 4.14.2. Higher-Order Cross-Over Design for Two Means – Superiority Test Table

The first calculation will be an example of total sample size for the Balaam Design (4x2).

3. First enter 0.025 for the Test Significance level row

4. Select "Balaam (4x2)" from the "Design Type" row and "Better" from the "Higher Means Values are" row.

5. Enter 1 for "True Difference in Means" and 0.5 for both the "Superiority Margin" and "Within Standard Error" rows.

6. Finally, enter 80 in the "Power" row and the table will appear as per Figure 4.14.3.

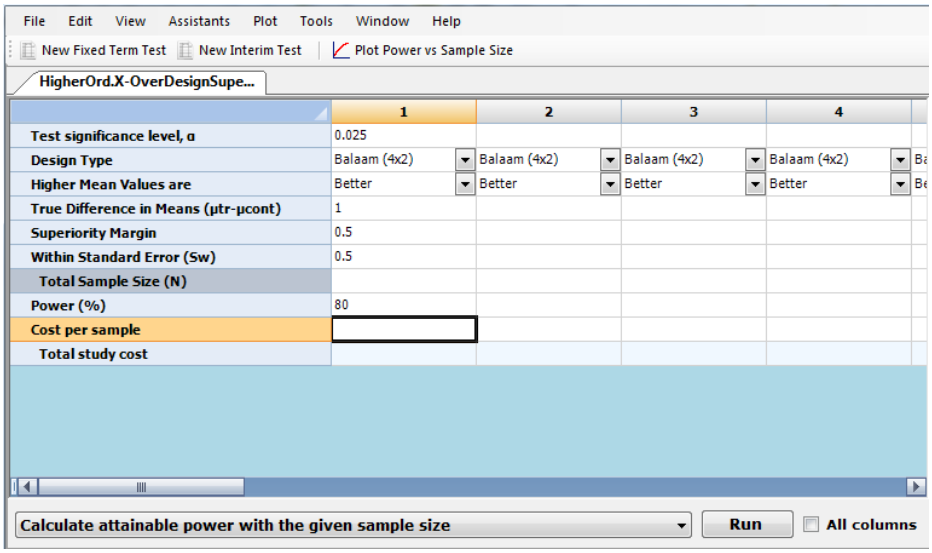


Figure 4.14.3. Values entered for Higher-Order Cross-Over Design – Superiority study design

7. Select "Calculate required sample size for given power" from the dropdown menu beside the "Run" button. Then click "Run". This will give a result of 65 as in Figure 4.14.4.

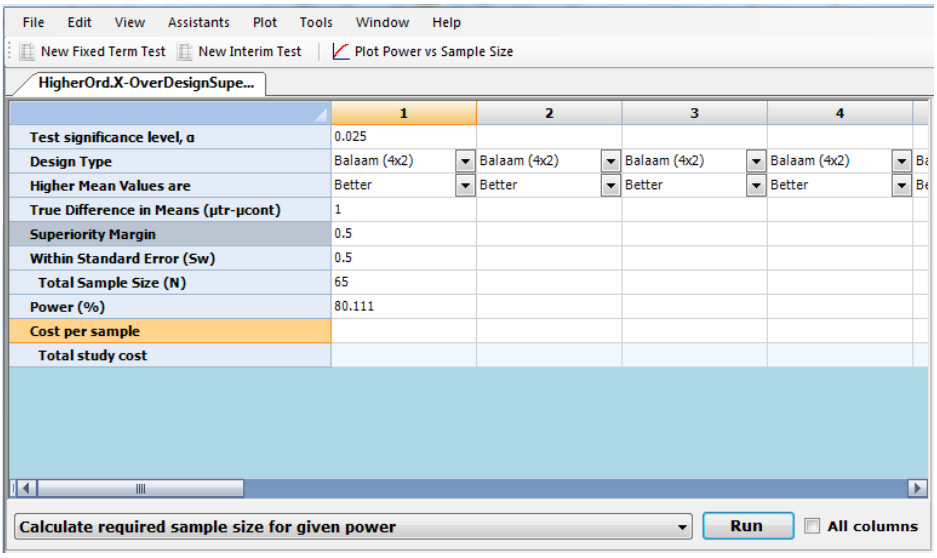


Figure 4.14.4. Completed Higher-Order Cross-Over Design – Superiority study design

The next calculation shows the effective equivalence between superiority margins when higher mean values are better or worse and effect of design type on sample size.

8. First, enter 80 for "Power" in column 1, delete the "Total Sample Size" value, select "Worse" in row three and change the "True Difference in Means" to -1. Then copy the values in column 1 across to columns 2 to 4. In columns 2 to 4, change design type to Two-Sequence Dual (2x3), Four Sequences & Two Periods (2x4) and Four Sequences and Periods (4x4) respectively. This will give a table as per Figure 4.14.5.

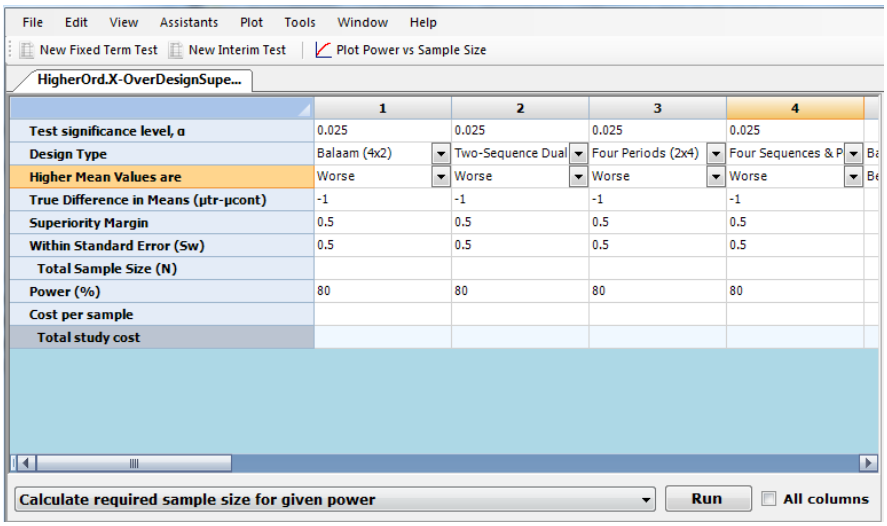
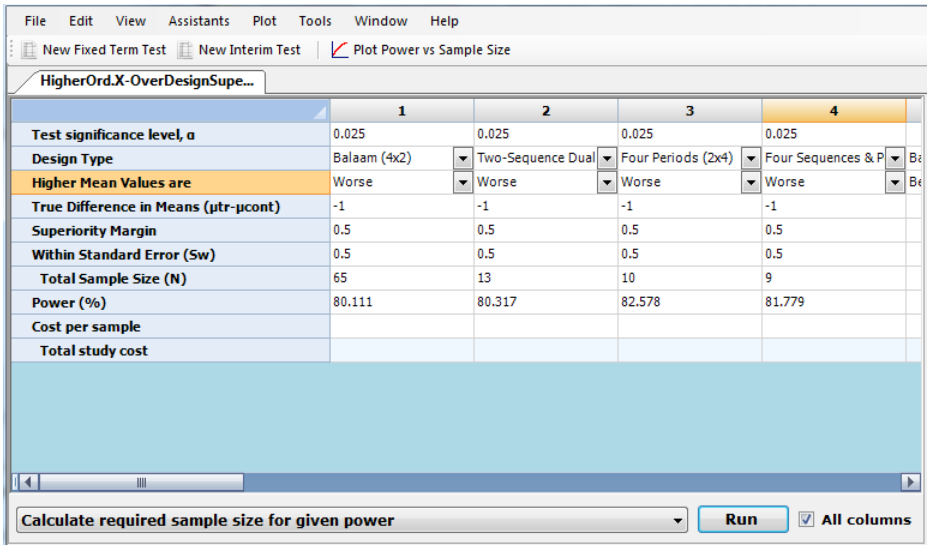


Figure 4.14.5. Analysis for Effect of Worse Means being Better and Effect of Design Type

9. Select "Calculate required sample size for given power" from the dropdown menu beside the "Run" button. Then tick the box to run "All Columns". Then click "Run".

This will give the resultant total sample sizes of 65, 13, 10 and 9 sequentially as in Figure 4.14.6 below.



	1	2	3	4
Test significance level, α	0.025	0.025	0.025	0.025
Design Type	Balaam (4x2)	Two-Sequence Dual	Four Periods (2x4)	Four Sequences & P
Higher Mean Values are	Worse	Worse	Worse	Worse
True Difference in Means (μtr-μcont)	-1	-1	-1	-1
Superiority Margin	0.5	0.5	0.5	0.5
Within Standard Error (Sw)	0.5	0.5	0.5	0.5
Total Sample Size (N)	65	13	10	9
Power (%)	80.111	80.317	82.578	81.779
Cost per sample				
Total study cost				

Figure 4.14.6. Results from Sensitivity analysis

10. Finally, by clicking on the desired study design column and going to the "Output" tab at the bottom of the screen, you can get an output statement for the results given by the current calculation as per Figure 4.14.7.

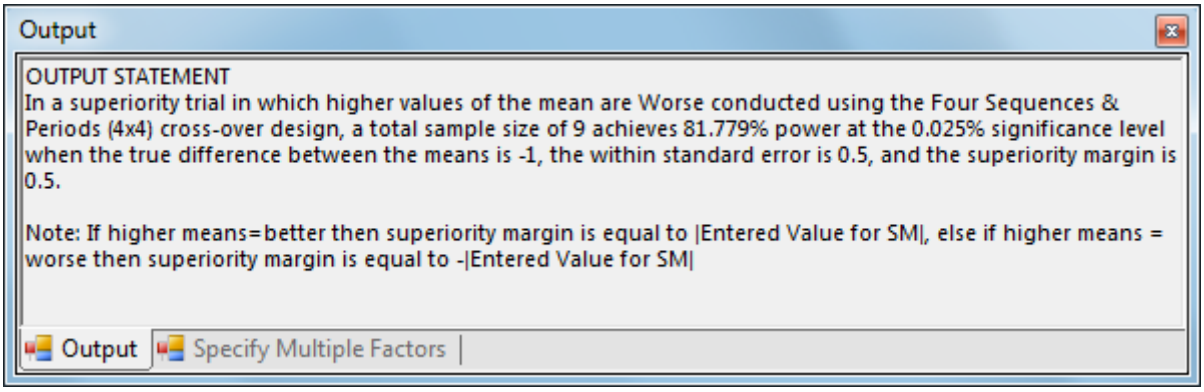


Figure 4.14.7. Study design Output statement

4.15. Higher-Order Cross-Over Design for Two Means – Non–Inferiority - Using Differences

4.15.1 Introduction

In a cross-over design each subject receives all treatments at least once with the objective of measuring study differences among the treatments. The term "cross-over" derives from the most common two treatments case, in which each subject crosses over from one treatment to another. The design assumes a washout period between treatments during which the response returns to its baseline. If this does not occur, there is said to be a carryover effect.

Cross-over designs are popular due to an expected increase in precision, since each subject effectively acts as their own control, and the concomitant reduction in study size. However, they do require more complex statistical analysis than parallel designs and suffer from complications and strong assumptions related to time effects, carry-over effects and patient attrition. These issues and the statistical methodology are discussed more thoroughly in Chow et al. (2008) and Chen et al. (1997).

In this table, we look beyond the most common 2x2 design to higher-order designs where there can be up to four sequences of treatments and/or measured time periods

This table facilitates the calculation of the power and sample size for four higher-order designs using the "additive model", namely the 2x3, 2x4, 4x2 and 4x4 designs (discussed below), for non-inferiority hypotheses. Power and sample size is computed using the method outlined by Chen et al. (1997).

4.15.2 Methodology

Before discussing statistical methodology, the structure of the four additional cross-over designs is given below. A cross-over design consists of two parameters: the sequences i.e. the pattern of treatments given to a patient and periods i.e. the number of time periods over which the outcome variable is measured. In the numeric summary, the AxB design indicates A number of sequences and B number of periods. Note that all of the following designs are for a two treatment design. The designs are as follows:

Table 4.15.1: Details of Higher-Cross Over Designs

Design Name	Advantage	Design
Balaam Design (4x2)	Allows estimation of intra-subject variabilities	Sequence Period1 Period2 1 A A 2 B B 3 A B 4 B A
Two-Sequence Dual Design (2x3)	Allows estimation of intra-subject variabilities	Sequence Period1 Period2 Period3 1 A B B 2 B A A
Four-Period with Two Sequences Design	Allows estimation of effect multiple sequential doses	Sequence Period1 Period2 Period3 Period4 1 A B B A 2 B A A B
Four-Period Design with Four Sequences	Allows estimation of effect multiple sequential doses	Sequence Period1 Period2 Period3 Period4 1 A A B B 2 B B A A 3 A B B A 4 B A A B

The statistical methodology for cross-over designs is based on the ANOVA methodology and is outlined in detail in Chow and Liu (2000) and Chow et al. (2008). Note that the method outlined here is based on the additive model (i.e. from an analysis of variance using untransformed data) as opposed to the multiplicative model (i.e. from an analysis of variance using log-transformed data). The multiplicative model case is dealt with the equivalent ratios table.

We now briefly discuss the non-inferiority hypothesis and effect of whether higher values of the outcome variables are considered a positive or negative effect.

Non-inferiority hypothesis tests are a subset of one-sided directional tests. In these tests, the researcher specifies an estimate for the expected true difference δ (i.e. $\mu_{Treat} - \mu_{Control}$) between treatments and a non-inferiority margin, which defines the difference above which would indicate the treatment is non-inferior to a control. Non-inferiority tests are a non-zero null test where the null hypothesis relates to the relationship between the true difference and margin. The direction of these hypotheses is defined by whether higher values of the outcome variable are considered better or not. This is outlined below:

If higher means are considered better:

$$H_0: \delta \leq -|NIM|, H_1: \delta > -|NIM| \qquad \{4.15.1\}$$

where $-|NIM|$ is the negative of the **absolute value** of the entered non-inferiority margin.

If higher means are considered worse:

$$H_0: \delta \geq |NIM|, H_1: \delta < |NIM| \qquad \{4.15.2\}$$

Note that it is **the absolute value** used in the worse means case.

The sample size and power calculations for higher-cross over design is derived from the work of Chen et al. (1997). Note that Chen only covers the case of assessing equivalence under a mean ratio hypothesis; however this methodology can be easily extended to non-inferiority/superiority margin type analysis and to the usage of differences instead of ratios. This table can be used to calculate the power, the total sample size, the non-inferiority margin, the test significance level, the true difference and the within-standard error.

Calculations use a t-distribution approximation. To calculate power and total sample size the user must specify the test significance level (α), whether higher means are better or worse, the higher-order cross-over design of interest, the true difference ($\mu_{Treat} - \mu_{Cont}$), the non-inferiority margin (NIM) and the within standard error (S_w) (which is derived from the square root of the within mean square error (WMSE) from the ANOVA analysis for the Cross-Over design, either from previous studies or pilot studies).

The formula for the power, derived from the methods from Chen et al. (1997), follow. When **higher means are better** then the formula for power is as follows:

$$1 - \beta = T_v\left[\left(\frac{(\delta - |NIM|)}{S_w\sqrt{b/n}}\right) - t_{V,1-\alpha}\right] \qquad \{4.15.3\}$$

where T_V is the cumulative t-distribution function with V degrees of freedom, $t_{V,1-\alpha}$ is the inverse cumulative t-distribution evaluated at $1-\alpha$, b is a design parameter (both V and b are defined by design choice, see below) and n is **average number of subjects per sequence** i.e. $N_{Total}/\#\ sequences$. If **higher means are worse** then the formula becomes

$$1 - \beta = 1 - T_v\left[-\left(\frac{(-\delta + |NIM|)}{S_w\sqrt{b/n}}\right) + t_{V,1-\alpha}\right] \qquad \{4.15.4\}$$

V and b are defined by the study design and are given by following equations:
$Balaam\ (4x2): V = 4n - 3, b = 2$
$Two - Sequence\ Dual\ (3x2): V = 4n - 4, b = 3/4$
$Four\ Periods\ \&\ Two\ Sequences\ (4x2): V = 6n - 5, b = 11/20$
$Four\ Periods\ and\ Four\ Sequences\ (4x4): V = 12n - 5, b = 1/4$

A closed form equation is not used to calculate the total sample size, the non-inferiority margin, the test significance levels, the within standard error or the smallest true detectable difference. Instead a search algorithm is used. The search algorithm calculates power at various values for the relevant parameter until the desired power is reached.

Note also that the total sample size calculated is the "exact" sample size (i.e. the first value of N_{Total} which achieves the desired power). However, for an equal sample size per sequence you need to round to the nearest 2 for the 2x3 and 2x4 designs and nearest four for the 4x2 and 4x4 designs.

4.15.3 Examples

Example 1: Validation example calculating required sample size for a given power
Chen et al. only consider an equivalence test and do not directly consider a one-sided non-inferiority hypothesis. However, using the validated equivalence table (see Chapter 4.16) we can confirm that the non-inferiority table gives the same results. We then conduct an analysis to show the effect of changing the design type and setting higher means as being worse.

The following steps outline the procedure for this example:

1. Open nQuery + nTerim 4.0 via the start menu or desktop shortcut. Click "New Fixed Term Test" from the top of the window. The "Study Goal and Design" window will appear as per Figure 4.15.1 below.

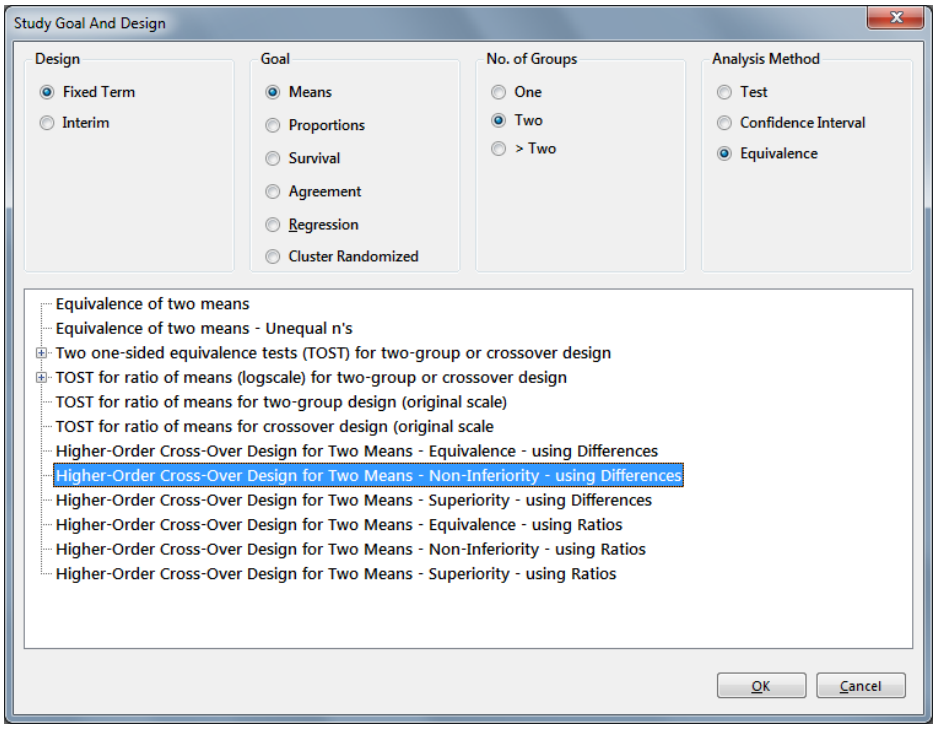


Figure 4.15.1. Study Goal and Design Window

2. In order to select the "Higher-Order Cross Over Design – Non-Inferiority – Using Differences" design table, navigate through the Study Goal and Design Window by selecting "Fixed Term" as the Design, "Means" as the Goal, "Two" as the Number of Groups and "Equivalence" as the Analysis Method. Then click "OK" and the test window will appear as per Figure 4.15.2.

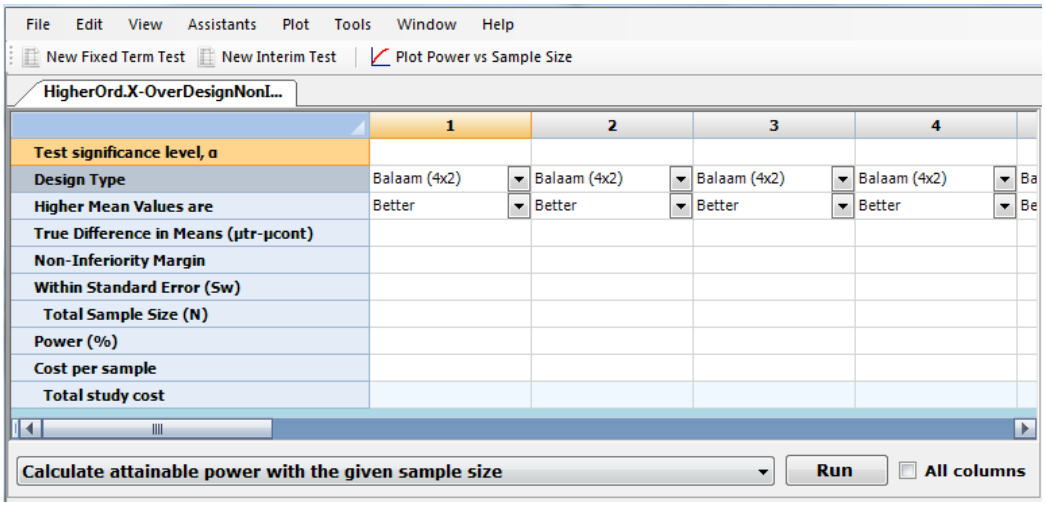


Figure 4.15.2. Higher-Order Cross-Over Design for Two Means – Non-Inferiority Test Table

The first calculation will be an example of total sample size for the Balaam Design (4x2).

3. First enter 0.025 for the "Test Significance level" row

4. Select "Balaam (4x2)" from the "Design Type" row and "Better" from the "Higher Mean Values are" row.

5. Enter -0.1 for "True Difference in Means" and 0.5 for both the "Non-inferiority Margin" and "Within Standard Error" rows.

6. Finally, enter 80 in the "Power" row and the table will appear as per Figure 4.15.3.

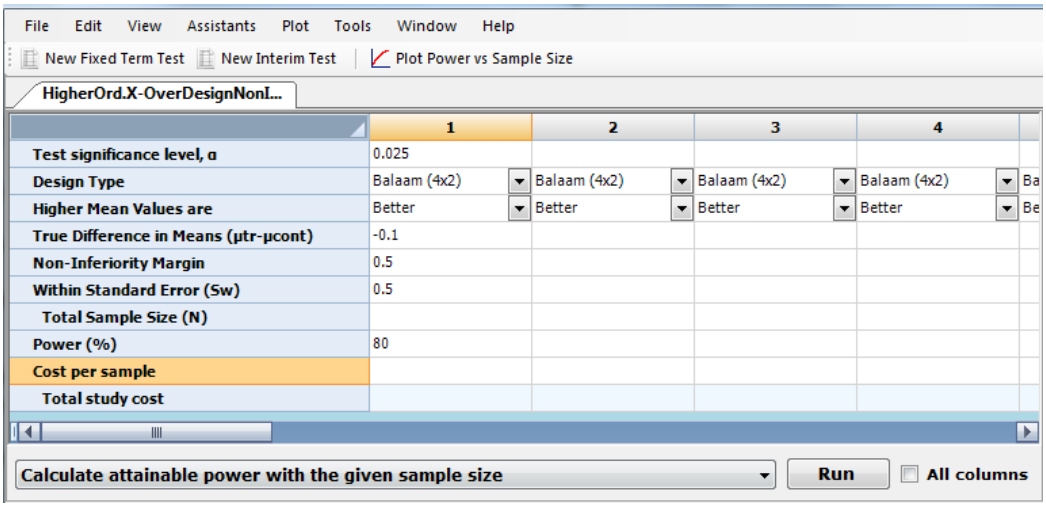


Figure 4.15.3. Values entered for Higher-Order Cross-Over Design – Non-Inferiority design

7. Select "Calculate required sample size for given power" from the dropdown menu beside the "Run" button. Then click "Run". This will give a result of 101 as in Figure 4.15.4.

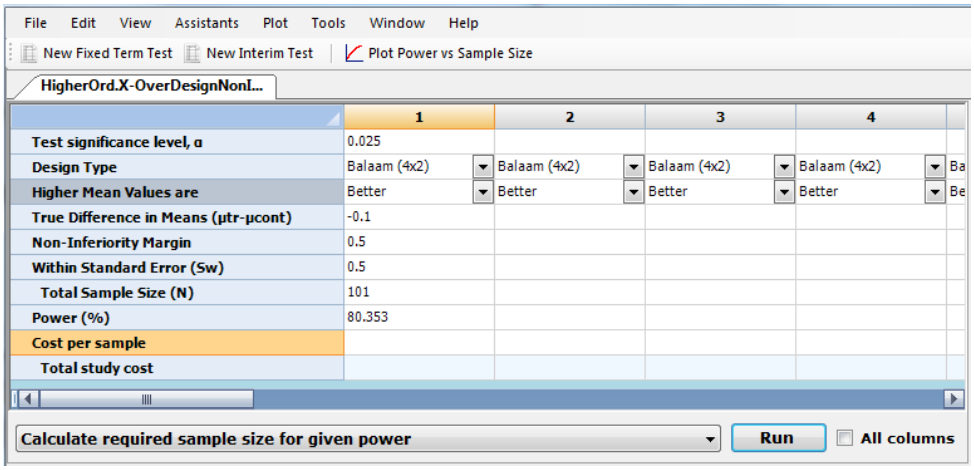


Figure 4.15.4. Completed Higher-Order Cross-Over Design – Non-Inferiority study design

8. To validate this example we open the "Higher-Order Cross-Over Design – Equivalence – using Differences" table using the same options as in step 2.

9. Enter the same values as from the non-inferiority table above except placing the non-inferiority margin value in the "Lower Equivalence Margin" row and making the "Upper Equivalence Margin" an arbitrarily high value (100 in this example) and then run the sample size calculation. This will give 101 as above as seen in Figure 4.15.5.

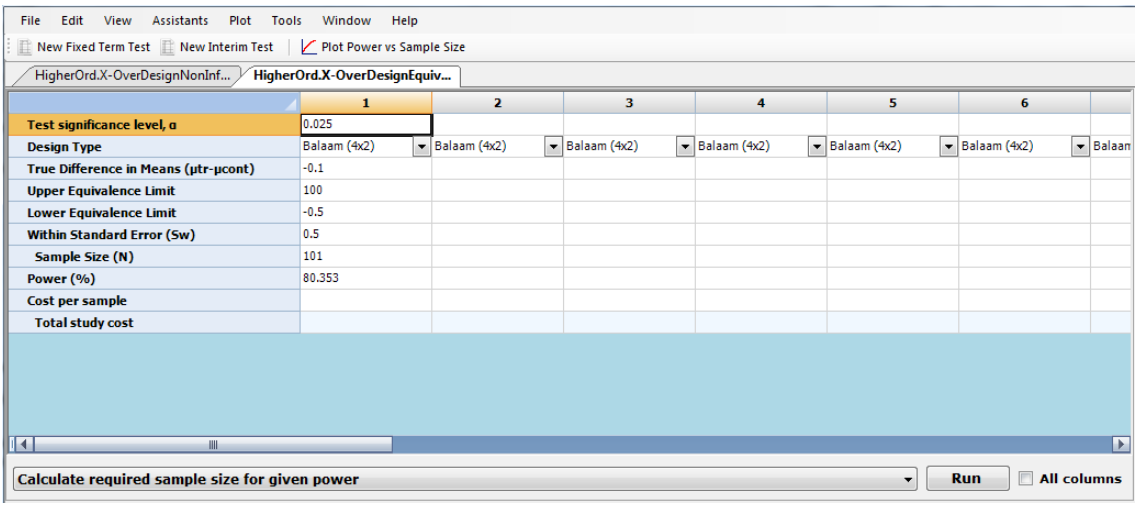


Figure 4.15.5. Validation of Non-inferiority table using Equivalence table

The next calculation shows the effective equivalence between non-inferiority margins when higher mean values are better or worse and the effect of design type on sample size.

10. First, return to the previous non-inferiority table using the tab just below the "New Fixed Term Test" button.

11. Now enter 80 for "Power" in column 1, delete the "Total Sample Size" value, select "Worse" in row three and change the "True Difference in Means" to 0.1. Then copy the values in column 1 across to columns 2 to 4. In columns 2 to 4, change design type to Two-Sequence Dual (2x3), Four Sequences & Two Periods (2x4) and Four Sequences and Periods (4x4) respectively. This will give a table as per Figure 4.15.6.

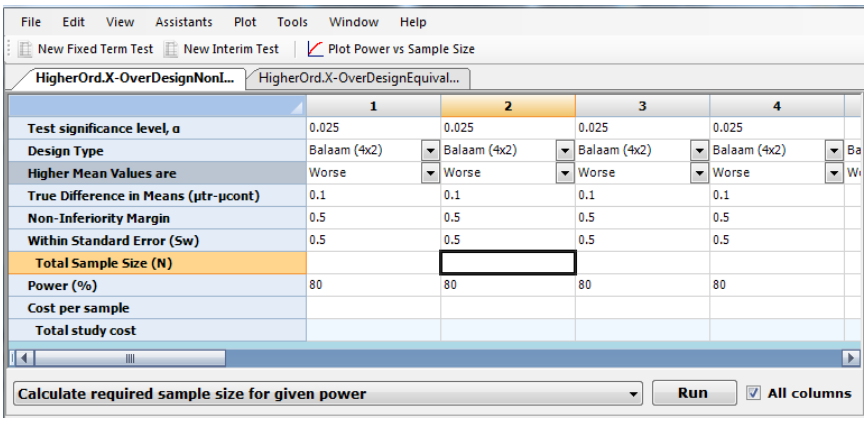


Figure 4.15.6. Analysis for Effect of Worse Means being Better and Effect of Design Type

12. Select "Calculate required sample size for given power" from the dropdown menu beside the "Run" button. Then tick the box to run "All Columns". Then click "Run".

This will give the resultant total sample sizes of 101, 20, 15 and 14 sequentially as in Figure 4.15.7 below. Note that we got the same value in column one as above, showing the equivalence in approaches.

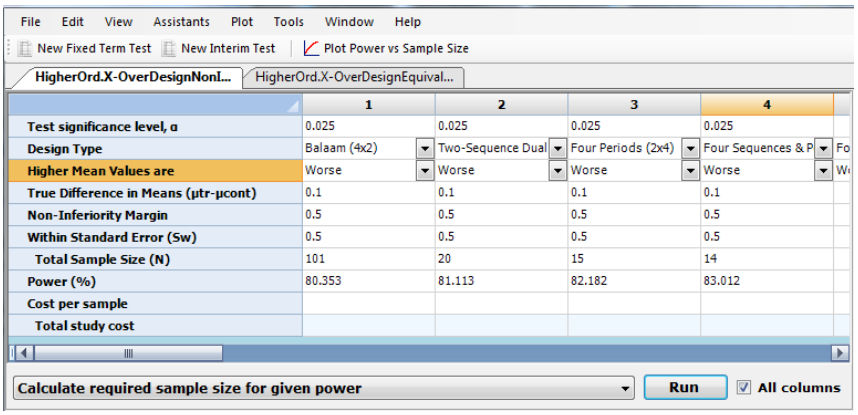


Figure 4.15.7. Results from Sensitivity analysis

13. Finally, by clicking on the desired study design column and going to the "Output" tab at the bottom of the screen, you can get an output statement for the results given by the current calculation as per Figure 4.15.8.

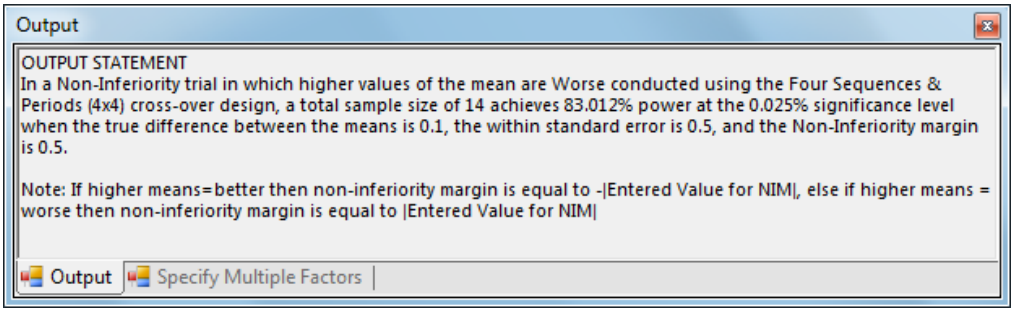


Figure 4.15.8. Study design Output statement

4.16. Higher-Order Cross-Over Design for Two Means – Equivalence - Using Differences

4.16.1 Introduction

In a cross-over design each subject receives all treatments at least once with the objective of measuring study differences among the treatments. The term "cross-over" derives from the most common two treatments case, in which each subject crosses over from one treatment to another. The design assumes a washout period between treatments during which the response returns to its baseline. If this does not occur, there is said to be a carryover effect.

Cross-over designs are popular due to an expected increase in precision, since each subject effectively acts as their own control, and the concomitant reduction in study size. However, they do require more complex statistical analysis than parallel designs and do suffer from complications and strong assumptions related to time effects, carry-over effects and patient attrition. These issues and the statistical methodology are discussed more thoroughly in Chow et al. (2008) and Chen et al. (1997).

In this table, we look beyond the most common 2x2 design to higher-order designs where there can be up to four sequences of treatments and/or measured time periods

This table facilitates the calculation of the power and sample size for four higher-order designs using the "additive model", namely the 2x3, 2x4, 4x2 and 4x4 designs (discussed below), for two one-sided tests (TOST) equivalence hypotheses. Power and sample size is computed using the method outlined by Chen et al. (1997).

4.16.2 Methodology

Before discussing statistical methodology, the structure of the four additional cross-over designs is given below. A cross-over design consists of two parameters: the sequences i.e. the pattern of treatments given to a patient and periods i.e. the number of time periods over which the outcome variable is measured. In the numeric summary, the AxB design indicates A number of sequences and B number of periods. Note that all of the following designs are for a two treatment design. The designs are as follows:

Table 4.16.1: Details of Higher-Cross Over Designs

Design Name	Advantage	Design
Balaam Design (4x2)	Allows estimation of intra-subject variabilities	Sequence Period1 Period2 1 A A 2 B B 3 A B 4 B A
Two-Sequence Dual Design (2x3)	Allows estimation of intra-subject variabilities	Sequence Period1 Period2 Period3 1 A B B 2 B A A
Four-Period with Two Sequences Design	Allows estimation of effect multiple sequential doses	Sequence Period1 Period2 Period3 Period4 1 A B B A 2 B A A B
Four-Period Design with Four Sequences	Allows estimation of effect multiple sequential doses	Sequence Period1 Period2 Period3 Period4 1 A A B B 2 B B A A 3 A B B A 4 B A A B

The statistical methodology for cross-over designs is based on ANOVA methodology and is outlined in detail in Chow and Liu (2000) and Chow et al. (2008). Note that the method outlined here is based on the additive model (i.e. from an analysis of variance using untransformed data) as opposed to the multiplicative model (i.e. from an analysis of variance using log-transformed data). The multiplicative model case is dealt with the equivalent ratios table.

We now briefly discuss the equivalence two one-sided tests (TOST) hypothesis.

Equivalence hypothesis tests are based on Schuirmann's two-one sided tests (TOST) methodology. In these tests, the researcher specifies an estimate for the expected true difference δ (i.e. $\mu_{Treat} - \mu_{Control}$) between treatments and two equivalence margins, a lower margin which is below zero and an upper margin which is above zero, which define the range over which the true difference would indicate the treatment and reference groups were equivalent i.e. if the true difference falls between the lower and upper equivalence margins that implies the two treatment groups are effectively equivalent on the outcome variable. This is outlined below.

The null hypothesis is given as follows:

$$H_0: \delta \leq LEL \text{ or } \delta \geq UEL, \text{where } LEL < 0, UEL > 0 \qquad \{4.16.1\}$$

where LEL is lower equivalence limit and UEL is the upper equivalence limit. The alternative hypothesis is as follows:

$$H_1: LEL < \delta < UEL \qquad \{4.16.2\}$$

Thus equivalence is powered based on the probability that the true difference falls between the two equivalence margins.

The sample size and power calculations for higher-cross over designs are derived from the work of Chen et al. (1997). Note that Chen only covers the case of assessing equivalence under a mean ratio hypothesis; however this methodology can be easily extended to the usage of differences instead of ratios. This table can be used to calculate the power, the total sample size, the upper equivalence limit margin, the lower equivalence limit margin, the test significance level, the true difference and the within-standard error.

Calculations use a t-distribution approximation. To calculate power and total sample size the user must specify the test significance level (α), the higher-order cross-over design of interest, the true difference ($\mu_{Treat} - \mu_{Cont}$), the upper equivalence limit (UEL), lower equivalence limit (LEL) and the within standard error (S_w) (which is derived from the square root of the within mean square error (WMSE) from the ANOVA analysis for the Cross-Over design, either from previous studies or pilot studies).

The formula for the power, derived from the methods from Chen et al. (1997), is as follows:

$$1 - \beta = T_v\left[\left(\frac{(-\delta + UEL)}{S_w\sqrt{b/n}}\right) - t_{V,1-\alpha}\right] - T_v\left[-\left(\frac{(\delta - LEL)}{S_w\sqrt{b/n}}\right) + t_{V,1-\alpha}\right] \qquad \{4.16.3\}$$

where T_V is the cumulative t-distribution function with V degrees of freedom, $t_{V,1-\alpha}$ is the inverse cumulative t-distribution evaluated at $1 - \alpha$, b is a design parameter (both V and b are defined by design choice, see below) and n is **average number of subjects per sequence** i.e. $N_{Total}/\#\ sequences$. V and b are defined by the study design and are given by following equations:

$Balaam\ (4x2): V = 4n - 3, b = 2$
$Two - Sequence\ Dual\ (3x2): V = 4n - 4, b = 3/4$
$Four\ Periods\ \&\ Two\ Sequences\ (4x2): V = 6n - 5, b = 11/20$
$Four\ Periods\ and\ Four\ Sequences\ (4x4): V = 12n - 5, b = 1/4$

The formulation used in Chen et. al., which can be shown to be equivalent to the above, was the following:

$$1 - \beta = T_v\left[\left(\frac{(-\theta + \nabla_U)}{CV\sqrt{b/n}}\right) - t_{V,1-\alpha}\right] - T_v\left[-\left(\frac{(\theta - \nabla_L)}{CV\sqrt{b/n}}\right) + t_{V,1-\alpha}\right] \qquad \{4.16.4\}$$

where $\theta = \delta/\mu_r$, μ_r is the reference mean, $CV = S_w/\mu_r$, $\nabla_U = UEL/\mu_r$ and $\nabla_L = LEL/\mu_r$. Thus if you were multiply the inside of each tail by μ_r/μ_r from 4.16.3, you would get back Chen's formulation. Thus if you entered the values where$\theta = \delta$, $UEL/LEL = \nabla_U/\nabla_L$ and

$S_w = CV$ into our table, you will get the same result as if using the direct difference measures.

A closed form equation is not used to calculate the total sample, the upper equivalence limit, the lower equivalence limit, the test significance levels, the within standard error or the smallest true detectable difference. Instead a search algorithm is used. The search algorithm calculates power at various values for the relevant parameter until the desired power is reached.

Note also that the total sample size calculated is the "exact" sample size (i.e. the first value of N_{Total} which achieves the desired power). However, for an equal sample size per sequence you need to round to the nearest 2 for the 2x3 and 2x4 designs and nearest four for the 4x2 and 4x4 designs.

4.16.3 Examples

Example 1: Validation example calculating required sample size for a given power
The following validation example is from Table 1 from Chen et al (1997). Note that, as mentioned above, Chen parameterises via the reference mean ratio but that by entering these values directly we are effectively converting them to the parameterisation explicitly used above.

The following steps outline the procedure for this example:

1. Open nQuery + nTerim 4.0 via the start menu or desktop shortcut. Click "New Fixed Term Test" from the top of the window. The "Study Goal and Design" window will appear as Figure 4.16.1 below.

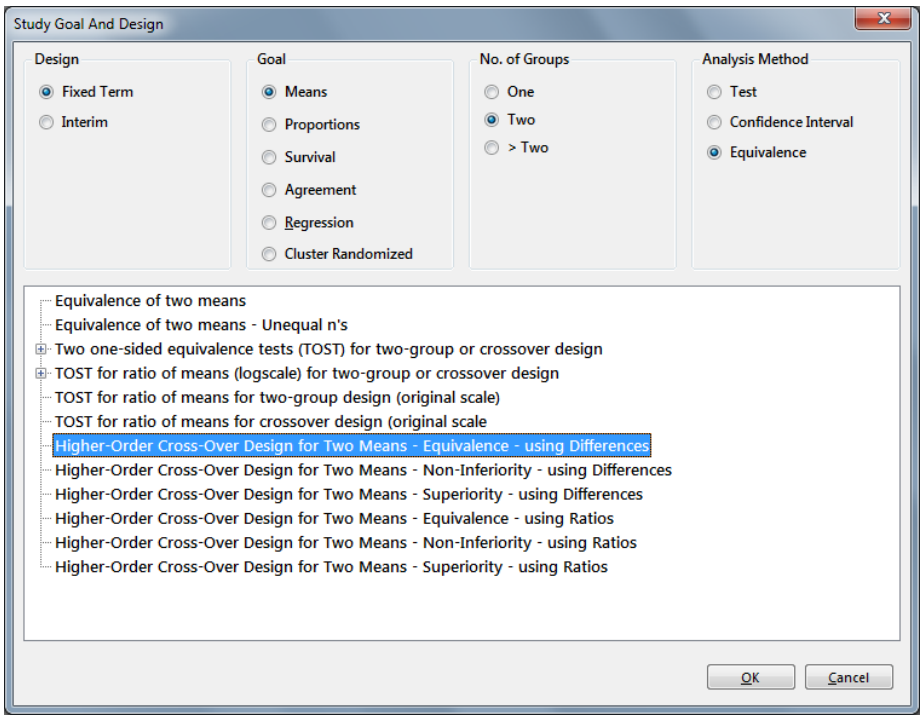


Figure 4.16.1. Study Goal and Design Window

2. In order to select the “Higher-Order Cross Over Design – Equivalence – Using Differences” design table, navigate through the Study Goal and Design Window by selecting "Fixed Term" as the Design, "Means" as the Goal, "Two" as the Number of Groups and "Equivalence" as the Analysis Method. Then click "OK" and the test window will appear as per Figure 4.16.2.

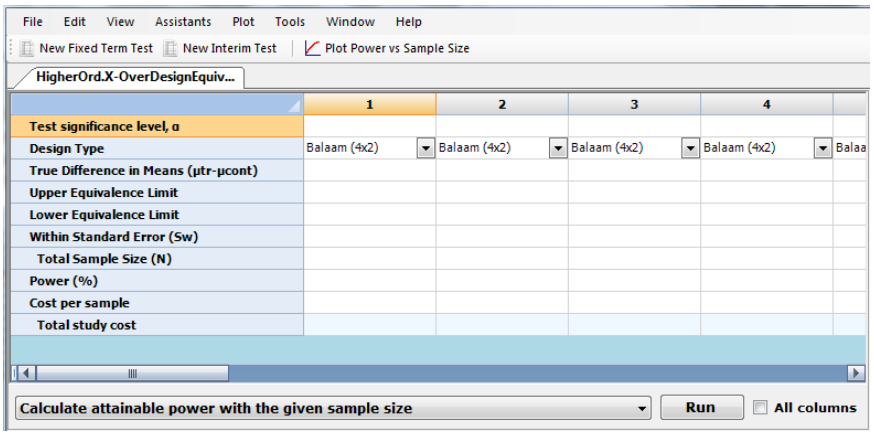


Figure 4.16.2. Higher-Order Cross-Over Design for Two Means – Equivalence Test Table

The first calculation will be an example of total sample size for the Balaam Design (4x2) taken from Table 1 from Chen et al.

3. First enter 0.05 for the “Test Significance level” row

4. Select “Balaam (4x2)” from the “Design Type” row and enter 0 for “True Difference in Means”.

5. Enter 0.2 for the “Upper Equivalence Limit”, -0.2 for the “Lower Equivalence Limit” and 0.1 for the “Within Standard Error” row.

6. Finally, enter 80 in the “Power” row and the table will appear as per Figure 4.16.3.

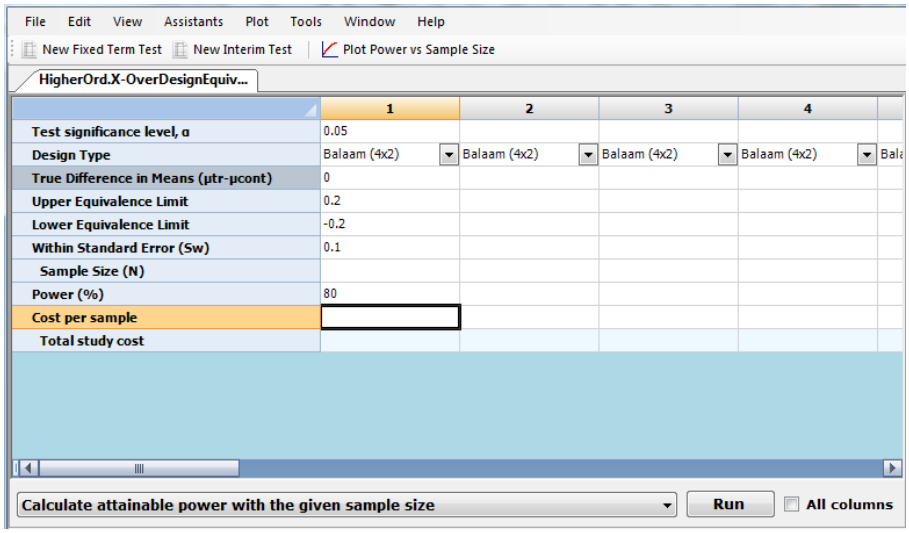
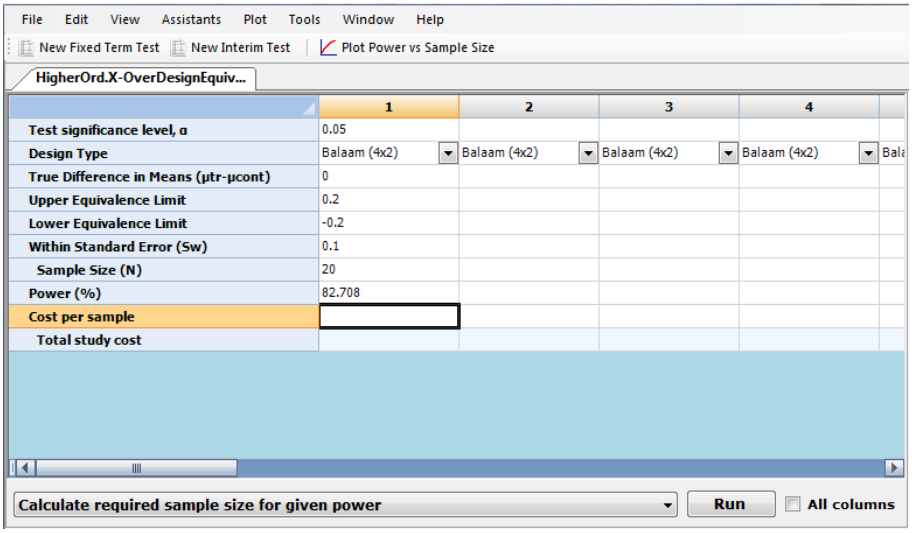


Figure 4.16.3. Values entered for Higher-Order Cross-Over Design – Equivalence design

7. Select “Calculate required sample size for given power” from the dropdown menu beside the “Run” button. Then click “Run”. This will give a result of 20 as in Figure 4.16.4. This agrees with the value from Chen et al.

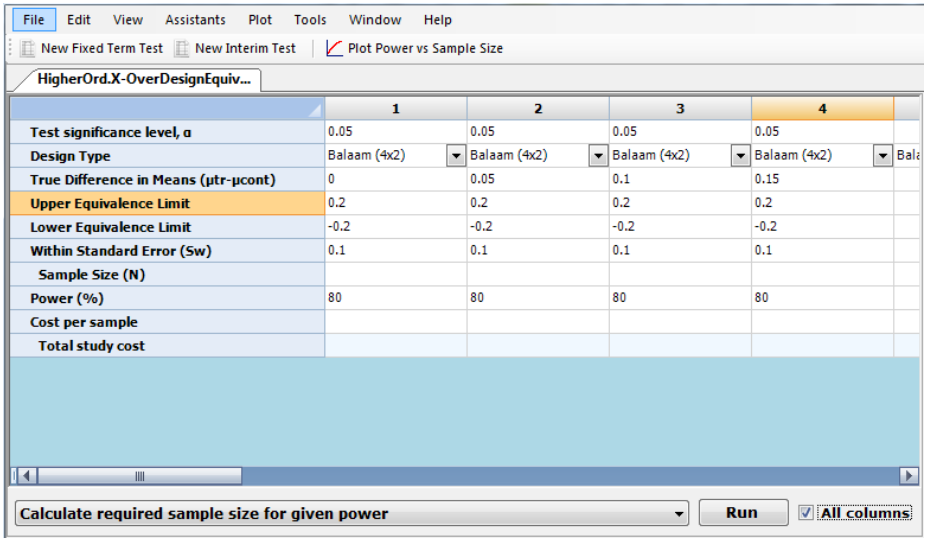


	1	2	3	4
Test significance level, α	0.05			
Design Type	Balaam (4x2)	Balaam (4x2)	Balaam (4x2)	Balaam (4x2)
True Difference in Means (μtr-μcont)	0			
Upper Equivalence Limit	0.2			
Lower Equivalence Limit	-0.2			
Within Standard Error (Sw)	0.1			
Sample Size (N)	20			
Power (%)	82.708			
Cost per sample				
Total study cost				

Figure 4.16.4. Completed Higher-Order Cross-Over Design – Equivalence study design

The next calculation shows the effect of changing the true difference in means by replicating the first row of Chen et al Table 1.

8. Enter 80 for “Power” in column 1 and delete the “Total Sample Size” value. Then copy the values in column 1 across to columns 2 to 4. In columns 2 to 4, replace the values of the “True Difference in Means” to 0.05, 0.1 and 0.15 respectively. This will give a table as per Figure 4.16.5.

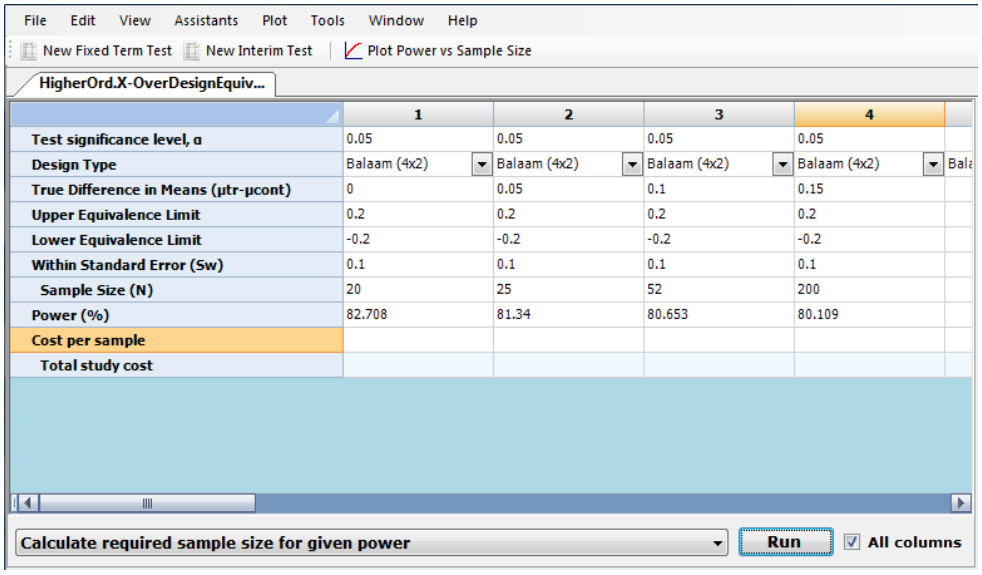


	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
Design Type	Balaam (4x2)	Balaam (4x2)	Balaam (4x2)	Balaam (4x2)
True Difference in Means (μtr-μcont)	0	0.05	0.1	0.15
Upper Equivalence Limit	0.2	0.2	0.2	0.2
Lower Equivalence Limit	-0.2	-0.2	-0.2	-0.2
Within Standard Error (Sw)	0.1	0.1	0.1	0.1
Sample Size (N)				
Power (%)	80	80	80	80
Cost per sample				
Total study cost				

Figure 4.16.5. Analysis for Effect of Changing the True Difference

9. Select “Calculate required sample size for given power” from the dropdown menu beside the “Run” button. Then tick the box to run “All Columns”. Then click “Run”.

This will give the resultant total sample sizes of 20, 25, 52 and 200 sequentially as in Figure 4.16.6 below. These agree with Chen et al, except column 2 which was 24 in the paper. There are some discrepancies with Chen et al over the entire table. However they used a one-sided approximation for sample size rather than an exact search procedure (pp. 757/758) so this is no unexpected.



	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
Design Type	Balaam (4x2)	Balaam (4x2)	Balaam (4x2)	Balaam (4x2)
True Difference in Means (μtr-μcont)	0	0.05	0.1	0.15
Upper Equivalence Limit	0.2	0.2	0.2	0.2
Lower Equivalence Limit	-0.2	-0.2	-0.2	-0.2
Within Standard Error (Sw)	0.1	0.1	0.1	0.1
Sample Size (N)	20	25	52	200
Power (%)	82.708	81.34	80.653	80.109
Cost per sample				
Total study cost				

Figure 4.16.6. Results from Sensitivity analysis

10. Finally, by clicking on the desired study design column and going to the "Output" tab at the bottom of the screen, you can get an output statement for the results given by the current calculation as per Figure 4.16.7.

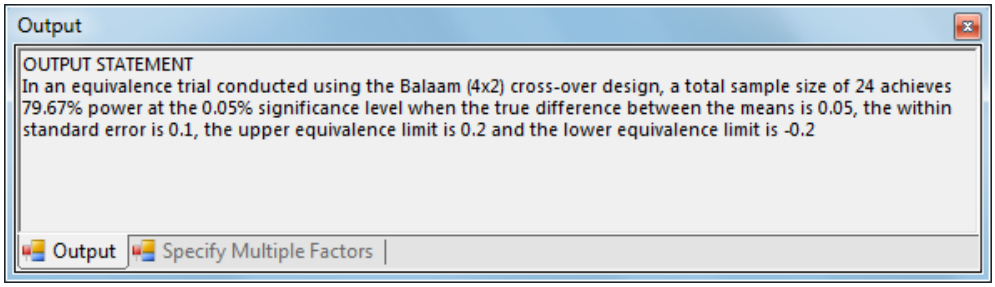


Figure 4.16.7. Study design Output statement

4.17. Higher-Order Cross-Over Design for Two Means – Superiority by a Margin– Using Ratios

4.17.1 Introduction

In a cross-over design each subject receives all treatments at least once with the objective of measuring study differences among the treatments. The term "cross-over" derives from the most common two treatments case, in which each subject crosses over from one treatment to another. The design assumes a washout period between treatments during which the response returns to its baseline. If this does not occur, there is said to be a carryover effect.

Cross-over designs are popular due to an expected increase in precision, since each subject effectively acts as their own control, and the concomitant reduction in study size. However, they do require more complex statistical analysis than parallel designs and do suffer from complications and strong assumptions related to time effects, carry-over effects and patient attrition. These issues and the statistical methodology are discussed more thoroughly in Chow et al. (2008) and Chen et al. (1997).

In this table, we look beyond the most common 2x2 design to higher-order designs where there can be up to four sequences of treatments and/or measured time periods

This table facilitates the calculation of the power and sample size for four higher-order designs using the "multiplicative model", namely the 2x3, 2x4, 4x2 and 4x4 designs (discussed below), for superiority by a margin hypotheses. Power and sample size is computed using the method outlined by Chen et al. (1997).

4.17.2 Methodology

Before discussing statistical methodology, the structure of the four additional cross-over designs is given below. A cross-over design consists of two parameters: the sequences i.e. the pattern of treatments given to a patient and periods i.e. the number of time periods over which the outcome variable is measured. In the numeric summary, the AxB design indicates A number of sequences and B number of periods. Note that all of the following designs are for a two treatment design. The designs are as follows:

Table 4.17.1: Details of Higher-Cross Over Designs

Design Name	Advantage	Design
Balaam Design (4x2)	Allows estimation of intra-subject variabilities	Sequence Period1 Period2 1 A A 2 B B 3 A B 4 B A
Two-Sequence Dual Design (2x3)	Allows estimation of intra-subject variabilities	Sequence Period1 Period2 Period3 1 A B B 2 B A A
Four-Period with Two Sequences Design	Allows estimation of effect multiple sequential doses	Sequence Period1 Period2 Period3 Period4 1 A B B A 2 B A A B
Four-Period Design with Four Sequences	Allows estimation of effect multiple sequential doses	Sequence Period1 Period2 Period3 Period4 1 A A B B 2 B B A A 3 A B B A 4 B A A B

The statistical methodology for cross-over designs is based on ANOVA methodology and is outlined in detail in Chow and Liu (2000) and Chow et al. (2008). Note that the method outlined here is based on multiplicative model (i.e. from an analysis of variance using log-transformed data) as opposed to the additive model (i.e. from an analysis of variance using untransformed data). The additive model case is dealt with the equivalent differences table.

We now briefly discuss the superiority by a margin hypothesis and the effect of whether higher values of the outcome variables are considered a positive or negative effect.

Superiority by a margin hypothesis tests are a subset of one-sided directional tests. In these tests, the researcher specifies an estimate for the expected true difference ratio φ (i.e. $\mu_{Treat}/\mu_{Control}$) between treatments and a superiority margin, which defines the distance above or below an effect size of no effect (i.e. $\varphi = 1$) which would indicate practical or clinical significance. Superiority by a margin tests are a non-zero null test where the null hypothesis relates to the relationship between the true ratio and the margin. The direction of these hypotheses is defined by whether higher values of the outcome variable are considered better or not.

Moreover, hypotheses expressed as ratios can be easily converted into differences on the log-scale. This offers the advantage of being able to use an unscaled percentage metric. Thus, data is often log-transformed and then analysed using the differences in logs but with the final conclusions being reported via the ratio. This is the process which constitutes the

multiplicative model of Chen et al. Thus, the hypotheses can stated in two ways and both these are provided below.

If higher means are considered better:

$$H_0: \varphi \leq 1 + |SM|, H_1: \varphi > 1 + |SM| \quad \{4.17.1\}$$

$$H_0: \ln(\mu_{Treat}) - \ln(\mu_{Treat}) \leq \ln(1 + |SM|)$$
$$H_1: \ln(\mu_{Treat}) - \ln(\mu_{Treat}) > \ln(1 + |SM|) \quad \{4.17.2\}$$

Where $\varphi = \mu_{Treat}/\mu_{Control}$, $|SM|$ is the **absolute distance above no effect/difference** (i.e. a ratio of one) that indicates the superiority margin ratio (i.e. a true ratio higher than $1 + |SM|$ would indicate superiority), μ_{Treat} is the true treatment mean and $\mu_{Control}$ is the true control mean.

If higher means are considered worse:

$$H_0: \varphi \geq 1 - |SM|, H_1: \varphi < 1 - |SM| \quad \{4.17.3\}$$

$$H_0: \ln(\mu_{Treat}) - \ln(\mu_{Treat}) \geq \ln(1 - |SM|)$$
$$H_1: \ln(\mu_{Treat}) - \ln(\mu_{Treat}) < \ln(1 - |SM|) \quad \{4.17.4\}$$

Note that $|SM|$ is now the distance below no effect (i.e. a ratio of one) that indicates the superiority margin ratio (i.e. a true ratio lower than $1 - |SM|$ would indicate superiority).

The sample size and power calculations for higher-cross over design is derived from the work of Chen et al. (1997). Note that Chen uses the coefficient of variation for the log-scale data and our table uses the coefficient of variation for the unscaled data (the coefficient of variation is equal to the within-standard error from the ANOVA analysis divided by the reference mean (see Chapter 4.16)). However, these can easily be derived from each other using the following relationships derived from Julious (2004):

$$X = \ln Y, Y = e^X \quad \{4.17.5\}$$

$$\mu_Y = e^{\mu_X + \frac{\sigma_X^2}{2}}, \sigma_Y^2 = \mu_Y^2\left(e^{\sigma_X^2} - 1\right) \quad \{4.17.6\}$$

$$=> COV_Y = \frac{\sqrt{\mu_Y^2\left(e^{\sigma_X^2} - 1\right)}}{\mu_Y} = \sqrt{\left(e^{\sigma_X^2} - 1\right)} \quad \{4.17.7\}$$

$$=> \sigma_X = \sqrt{\ln(COV_Y^2 + 1)}, \mu_X = \frac{\mu_Y}{\ln(COV_Y^2 + 1)} \quad \{4.17.8\}$$

Thus hypotheses stated in terms of X can be stated in terms of Y and vice-versa. Note that the sigma terms are equivalent to the within-standard error from the ANOVA on the unscaled σ_Y and log-scaled data σ_X.

This table can be used to calculate the power, the total sample size, the superiority margin, the test significance level, the true ratio and the coefficient of variation.

Calculations use a t-distribution approximation. To calculate power and total sample size the user must specify the test significance level (α), whether higher means are better or worse, the higher-order cross-over design of interest, the true ratio (φ or μ_{Treat}/μ_{Cont}), the superiority margin (SM) and the coefficient of variation (COV) (which is derived from either from previous studies or pilot studies based on the original data).

The formula for the power, derived from the methods from Chen et al. (1997), follow. When **higher means are better** the formula for power is as follows:

$$1-\beta = T_v\left[\left(\frac{(|\ln(\varphi)| - \ln(1+SM))}{S_w\sqrt{b/n}}\right) - t_{V,1-\alpha}\right] \qquad \{4.17.9\}$$

where T_V is the cumulative t-distribution function with V degrees of freedom, $t_{V,1-\alpha}$ is the inverse cumulative t-distribution evaluated at $1-\alpha$, b is a design parameter (both V and b are defined by design choice, see below), n is **average number of subjects per sequence** i.e. $N_{Total}/\#\ sequences$ and the square root of within-standard error S_w equals $\sqrt{\ln\left(COV_Y^2+1\right)}$.

If **higher means are worse** then the formula becomes

$$1-\beta = 1-T_v\left[-\left(\frac{(|\ln(\varphi)| + \ln(1-SM))}{S_w\sqrt{b/n}}\right) + t_{V,1-\alpha}\right] \qquad \{4.17.10\}$$

V and b are defined by the study design and are given by following equations:
$Balaam\ (4x2): V = 4n-3, b = 2$
$Two-Sequence\ Dual\ (3x2): V = 4n-4, b = 3/4$
$Four\ Periods\ \&\ Two\ Sequences\ (4x2): V = 6n-5, b = 11/20$
$Four\ Periods\ and\ Four\ Sequences\ (4x4): V = 12n-5, b = 1/4$

A closed form equation is not used to calculate the total sample, the superiority margin, the test significance levels, the within standard error or the smallest true ratio. Instead a search algorithm is used. The search algorithm calculates power at various values for the relevant parameter until the desired power is reached.

Note also that the total sample size calculated is the "exact" sample size (i.e. the first value of N_{Total} which achieves the desired power). However, for an equal sample size per sequence you need to round to the nearest 2 for the 2x3 and 2x4 designs and nearest four for the 4x2 and 4x4 designs.

4.17.3 Examples

Example 1: Validation example calculating required sample size for a given power
This table uses the same mechanics as the non-inferiority table and thus the example used for validation used there is sufficient for this table (see Chapter 4.18). The following is an example of sample size calculation with the table followed by an illustration of the effect of different design types on sample size.

The following steps outline the procedure for this example:

11. Open nQuery + nTerim 4.0 via the start menu or desktop shortcut. Click “New Fixed Term Test” from the top of the window. The “Study Goal and Design” window will appear as Figure 4.17.1 below.

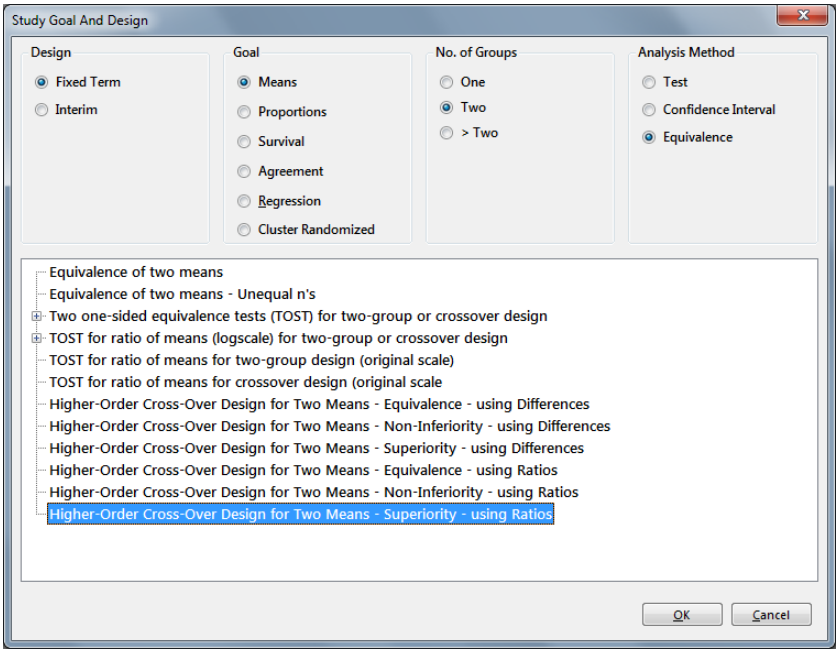


Figure 4.17.1. Study Goal and Design Window

12. In order to select the “Higher-Order Cross Over Design – Superiority by a Margin – Using Ratios” design table, navigate through the Study Goal and Design Window by selecting "Fixed Term" as the Design, "Means" as the Goal, "Two" as the Number of Groups and "Equivalence" as the Analysis Method. Then click "OK" and the test window will appear as per Figure 4.17.2.

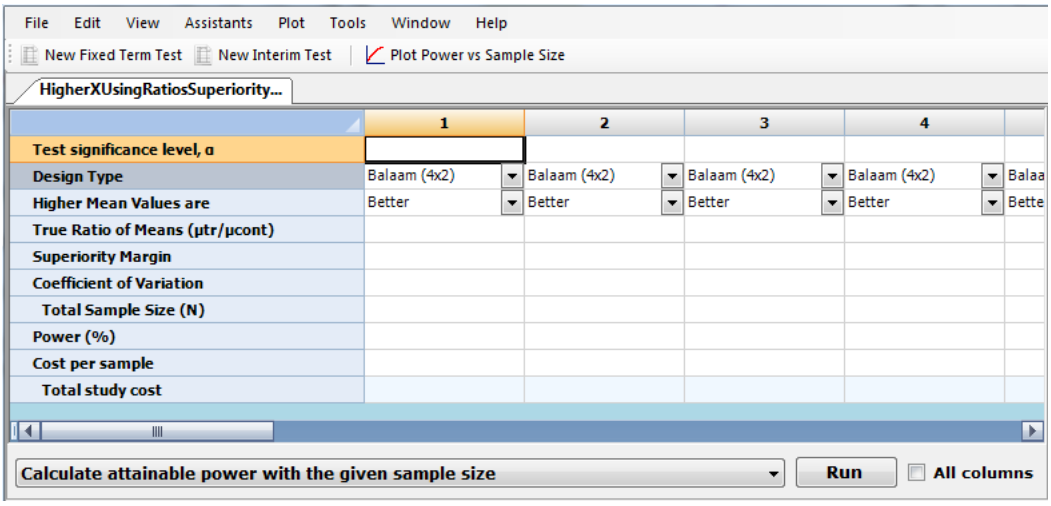


Figure 4.17.2. Higher-Order Cross-Over Design for Two Means – Superiority Test Table

The first calculation will be an example of total sample size for the Balaam Design (4x2).

13. First enter 0.025 for the "Test Significance level" row

14. Select "Balaam (4x2)" from the "Design Type" row and "Better" from the "Higher Mean Values are" row.

15. Enter 2 for "True Ratio of Means" and 0.5 for both the "Superiority Margin" and "Coefficient of Variation" rows.

16. Finally, enter 80 in the "Power" row and the table will appear as per Figure 4.17.3.

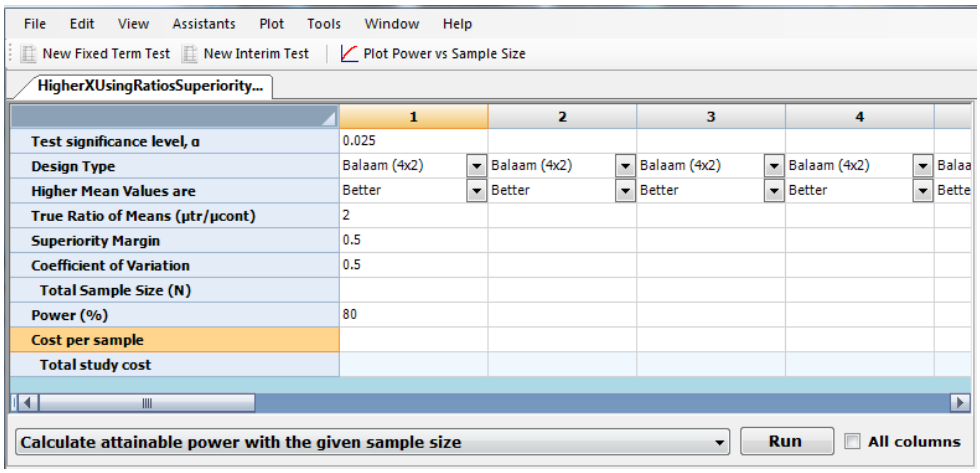


Figure 4.17.3. Values entered for Higher-Order Cross-Over Design – Superiority study design

17. Select "Calculate required sample size for given power" from the dropdown menu beside the "Run" button. Then click "Run". This will give a result of 172 as in Figure 4.17.4.

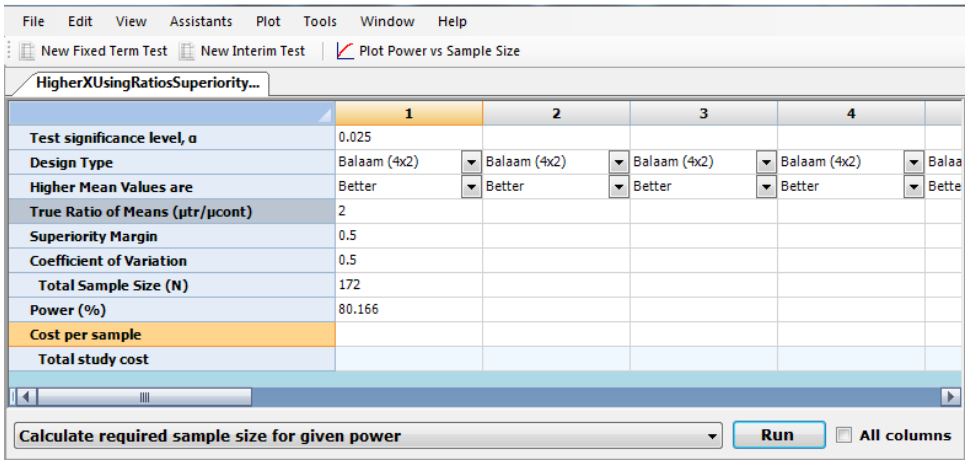


Figure 4.17.4. Completed Higher-Order Cross-Over Design – Superiority study design

The next calculation shows the effect of design type on sample size.

18. First, enter 80 for Power in column 1, delete the "Total Sample Size" value, select "Worse" in row three and change the "True Ratio of Means" to 0.25. Then copy the values in column 1 across to columns 2 to 4. In columns 2 to 4, change design type to Two-Sequence Dual (2x3), Four Sequences & Two Periods (2x4) and Four Sequences and Periods (4x4) respectively. This will give a table as per Figure 4.17.5.

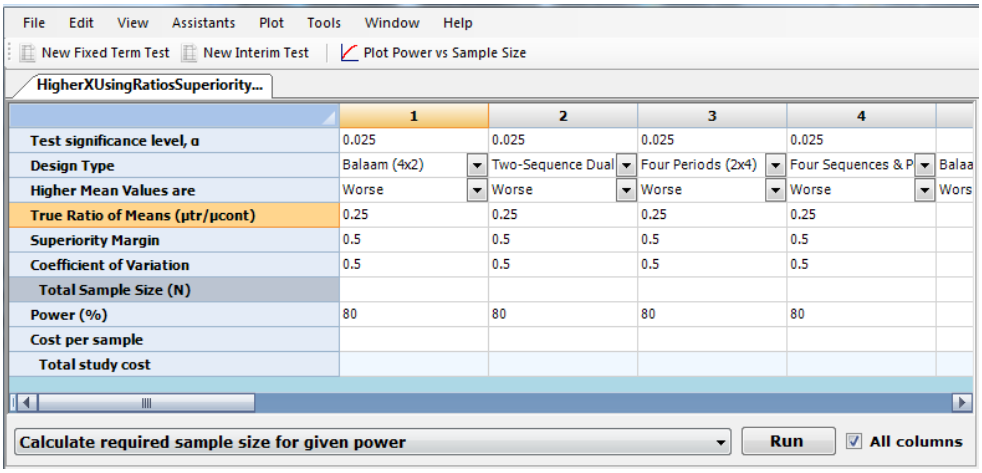
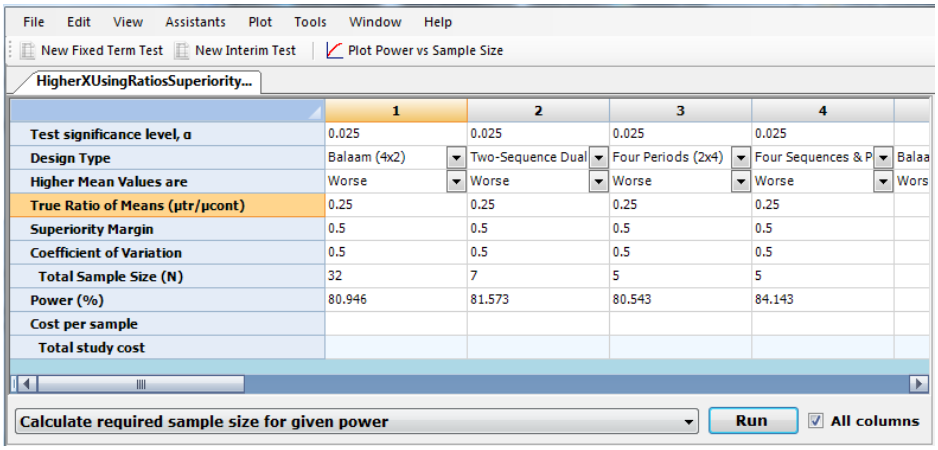


Figure 4.17.5. Analysis for Effect of Worse Means being better and Effect of Design Type

19. Select "Calculate required sample size for given power" from the dropdown menu beside the "Run" button. Then tick the box to run "All Columns". Then click "Run".

This will give the resultant total sample sizes of 32, 7, 5 and 5 sequentially as in Figure 4.17.6 below.



	1	2	3	4	
Test significance level, α	0.025	0.025	0.025	0.025	
Design Type	Balaam (4x2)	Two-Sequence Dual	Four Periods (2x4)	Four Sequences & P	Balaa
Higher Mean Values are	Worse	Worse	Worse	Worse	Wors
True Ratio of Means (μtr/μcont)	0.25	0.25	0.25	0.25	
Superiority Margin	0.5	0.5	0.5	0.5	
Coefficient of Variation	0.5	0.5	0.5	0.5	
Total Sample Size (N)	32	7	5	5	
Power (%)	80.946	81.573	80.543	84.143	
Cost per sample					
Total study cost					

Figure 4.17.6. Results from Sensitivity analysis

20. Finally, by clicking on the desired study design column and going to the “Output” tab at the bottom of the screen, you can get an output statement for the results given by the current calculation as per Figure 4.17.7.

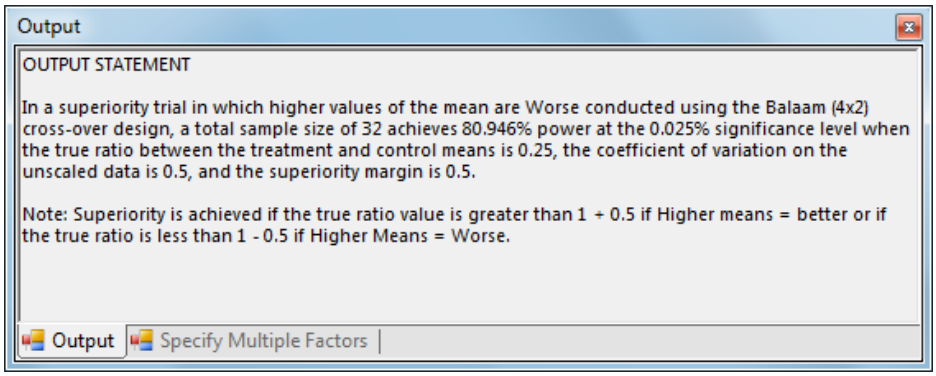


Figure 4.17.7. Study design Output statement

4.18. Higher-Order Cross-Over Design for Two Means – Non–Inferiority - Using Ratios

4.18.1 Introduction

In a cross-over design each subject receives all treatments at least once with the objective of measuring study differences among the treatments. The term "cross-over" derives from the most common two treatments case, in which each subject crosses over from one treatment to another. The design assumes a washout period between treatments during which the response returns to its baseline. If this does not occur, there is said to be a carryover effect.

Cross-over designs are popular due to an expected increase in precision, since each subject effectively acts as their own control, and the concomitant reduction in study size. However, they do require more complex statistical analysis than parallel designs and do suffer from complications and strong assumptions related to time effects, carry-over effects and patient attrition. These issues and statistical methodology are discussed more thoroughly in Chow et al. (2008) and Chen et al. (1997).

In this table, we look beyond the most common 2x2 design to higher-order designs where there can be up to four sequences of treatments and/or measured time periods

This table facilitates the calculation of the power and sample size for four higher-order designs using the multiplicative model, namely the 2x3, 2x4, 4x2 and 4x4 designs (discussed below), for non-inferiority hypotheses. Power and sample size is computed using the method outlined by Chen et al. (1997).

4.18.2 Methodology

Before discussing statistical methodology, the structure of the four additional cross-over designs is given below. A cross-over design consists of two parameters: the sequences i.e. the pattern of treatments given to a patient and periods i.e. the number of time periods of which the outcome variable is measured. In the numeric summary, the AxB design indicates A number of sequences and B number of periods. Note that all of the following designs are for a two treatment design. The designs are as follows:

Table 4.18.1: Details of Higher-Cross Over Designs

Design Name	Advantage	Design
Balaam Design (4x2)	Allows estimation of intra-subject variabilities	Sequence Period1 Period2 1 A A 2 B B 3 A B 4 B A
Two-Sequence Dual Design (2x3)	Allows estimation of intra-subject variabilities	Sequence Period1 Period2 Period3 1 A B B 2 B A A
Four-Period with Two Sequences Design	Allows estimation of effect multiple sequential doses	Sequence Period1 Period2 Period3 Period4 1 A B B A 2 B A A B
Four-Period Design with Four Sequences	Allows estimation of effect multiple sequential doses	Sequence Period1 Period2 Period3 Period4 1 A A B B 2 B B A A 3 A B B A 4 B A A B

The statistical methodology for cross-over designs is based on ANOVA methodology and is outlined in detail in Chow and Liu (2000) and Chow et al. (2008). Note that the method outlined here is based on multiplicative model (i.e. from an analysis of variance using log-transformed data) as opposed to the additive model (i.e. from an analysis of variance using untransformed data). The additive model case is dealt with the equivalent differences table.

We now briefly discuss the non-inferiority hypothesis and effect of whether higher values of the outcome variables are considered a positive or negative effect.

Non-inferiority hypothesis tests are a subset of one-sided directional tests. In these tests, the researcher specifies an estimate for the expected true ratio φ (i.e. $\mu_{Treat}/\mu_{Control}$) between treatments and a non-inferiority margin, which defines the distance above which the true ratio would indicate the treatment is non-inferior to the control. Non-inferiority tests are a non-zero null test where the null hypothesis relates to the relationship between the true ratio and the margin. The direction of these hypotheses is defined by whether higher values of the outcome variable are considered better or not.

Moreover, hypotheses expressed as ratios can be easily converted into differences on the log-scale. This offers the advantage of being able to use an unscaled percentage metric. Thus, data is often log-transformed and then analysed using the differences in logs but the final conclusions are drawn back in the ratio metric. This is the process which constitutes the

multiplicative model of Chen et al. Thus the hypotheses can stated in two ways and these are both provided below.

If higher means are considered better:

$$H_0: \varphi \leq 1 - |NIM|, H_1: \varphi > 1 - |NIM| \quad \{4.18.1\}$$

$$H_0: \ln(\mu_{Treat}) - \ln(\mu_{Treat}) \leq \ln(1 - |NIM|)$$
$$H_1: \ln(\mu_{Treat}) - \ln(\mu_{Treat}) > \ln(1 - |NIM|) \quad \{4.18.2\}$$

Where $\varphi = \mu_{Treat}/\mu_{Control}$, $|NIM|$ is the **distance below no effect/difference** (i.e. a ratio of one) that indicates the non-inferiority margin ratio (i.e. true ratios higher than $1 - |NIM|$ would be considered non-inferior), μ_{Treat} is the true treatment mean and $\mu_{Control}$ is the true control mean. If higher means are considered worse:

$$H_0: \varphi \geq 1 + |NIM|, H_1: \varphi < 1 + |NIM| \quad \{4.18.3\}$$

$$H_0: \ln(\mu_{Treat}) - \ln(\mu_{Treat}) \geq \ln(1 + |NIM|)$$
$$H_1: \ln(\mu_{Treat}) - \ln(\mu_{Treat}) < \ln(1 + |NIM|) \quad \{4.18.4\}$$

Note that $|NIM|$ is now the distance above no effect (i.e. ratio of one) that indicates the non-inferiority margin ratio (i.e. true ratios lower than $1 + |NIM|$ would be considered non-inferior).

The sample size and power calculations for higher-cross over design is derived from the work of Chen et al. (1997). Note that Chen uses the coefficient of variation for the log-scale data and our table uses the coefficient of variation for the unscaled data (the coefficient of variation is equal to the within-standard error from the ANOVA analysis divided by the reference mean (see Chapter 4.16)). However, these can easily be derived from each other using the following relationships derived from Julious (2004):

$$X = \ln Y, Y = e^X \quad \{4.18.5\}$$

$$\mu_Y = e^{\mu_X + \frac{\sigma_X^2}{2}}, \sigma_Y^2 = \mu_Y^2\left(e^{\sigma_X^2} - 1\right) \quad \{4.18.6\}$$

$$=> COV_Y = \sqrt{\mu_Y^2\left(e^{\sigma_X^2} - 1\right)} \Big/ \mu_Y = \sqrt{\left(e^{\sigma_X^2} - 1\right)} \quad \{4.18.7\}$$

$$=> \sigma_X = \sqrt{\ln(COV_Y^2 + 1)}, \mu_X = \frac{\mu_Y}{\ln(COV_Y^2 + 1)} \quad \{4.18.8\}$$

Thus hypotheses stated in terms of X can be stated in terms of Y and vice-versa. Note that the sigma terms are equivalent to the within-standard error from the ANOVA on the unscaled σ_Y and log-scaled data σ_X.

This table can be used to calculate the power, the total sample size, the non-inferiority margin, the test significance level, the true ratio and the coefficient of variation for the unscaled data.

Calculations use a t-distribution approximation. To calculate power and total sample size the user must specify the test significance level (α), whether higher means are better or worse, the higher-order cross-over design of interest, the true ratio (φ or μ_{Treat}/μ_{Cont}), the non-inferiority margin (NIM) and the coefficient of variation (COV) (which is derived from either from previous studies or pilot studies and is based on the unscaled (i.e. not log-transformed) data.

The formula for the power, derived from the methods from Chen et al. (1997), follow. When **higher means are better** then the formula for power is as follows:

$$1-\beta = T_v\left[\left(\frac{(-\ln\varphi + \ln(1-NIM))}{S_w\sqrt{b/n}}\right) - t_{V,1-\alpha}\right] \qquad \{4.18.9\}$$

where T_V is the cumulative t-distribution function with V degrees of freedom, $t_{V,1-\alpha}$ is the inverse cumulative t-distribution evaluated at $1-\alpha$, b is a design parameter (both V and b are defined by design choice, see below), n is **average number of subjects per sequence** i.e. $N_{Total}/\#\ sequences$ and the square root of within-standard error S_w equals $\sqrt{\ln\left(COV_Y^2+1\right)}$.

If **higher means are worse** then the formula becomes

$$1-\beta = 1-T_v\left[-\left(\frac{(\ln\varphi - \ln(1+NIM))}{S_w\sqrt{b/n}}\right) + t_{V,1-\alpha}\right] \qquad \{4.18.10\}$$

V and b are defined by the study design and are given by following equations:
$Balaam\ (4x2): V = 4n-3, b = 2$
$Two-Sequence\ Dual\ (3x2): V = 4n-4, b = 3/4$
$Four\ Periods\ \&\ Two\ Sequences\ (4x2): V = 6n-5, b = 11/20$
$Four\ Periods\ and\ Four\ Sequences\ (4x4): V = 12n-5, b = 1/4$

A closed form equation is not used to calculate the total sample, the non-inferiority margin, the test significance levels, the within standard error or the smallest true detectable ratio. Instead a search algorithm is used. The search algorithm calculates power at various values for the relevant parameter until the desired power is reached.

Note also that the total sample size calculated is the "exact" sample size (i.e. the first value of N_{Total} which achieves the desired power). However, for an equal sample size per sequence you need to round to the nearest 2 for the 2x3 and 2x4 designs and nearest four for the 4x2 and 4x4 designs.

4.18.3 Examples

Example 1: Validation example calculating required sample size for a given power
Chen et al. only consider an equivalence test and do not directly consider a one-sided non-inferiority hypothesis. However, using the validated equivalence table (see Chapter 4.19) we can confirm that the non-inferiority table gives the same results. We then conduct an analysis to show the effect of changing the design type.

The following steps outline the procedure for this example:

14. Open nQuery + nTerim 4.0 via the start menu or desktop shortcut. Click “New Fixed Term Test” from the top of the window. The “Study Goal and Design” window will appear as Figure 4.18.1 below.

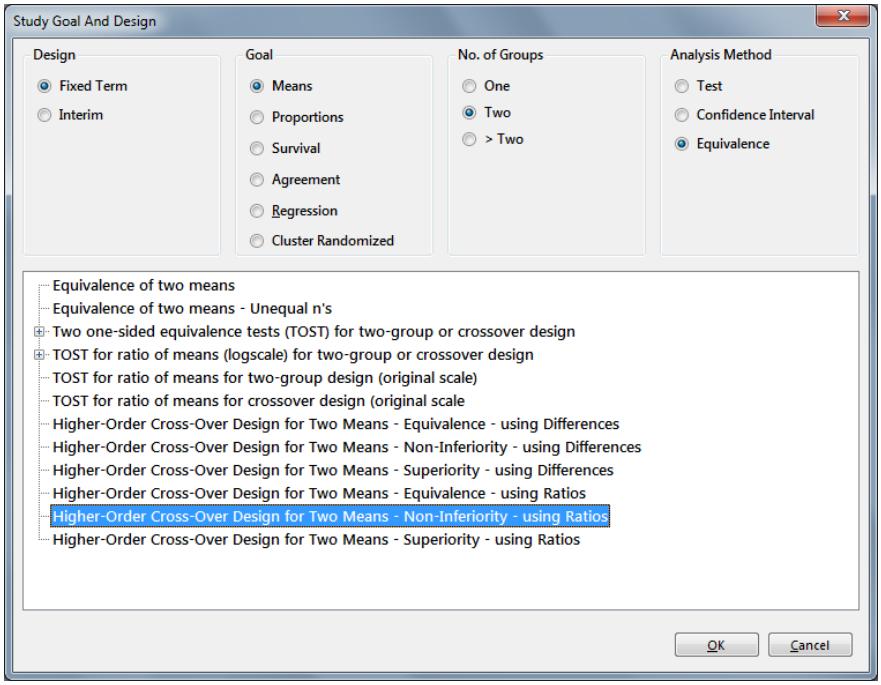


Figure 4.18.1. Study Goal and Design Window

15. In order to select the “Higher-Order Cross Over Design – Non-Inferiority – Using Ratios” design table, navigate through the Study Goal and Design Window by selecting "Fixed Term" as the Design, "Means" as the Goal, "Two" as the Number of Groups and "Equivalence" as the Analysis Method. Then click "OK" and the test window will appear as per Figure 4.18.2.

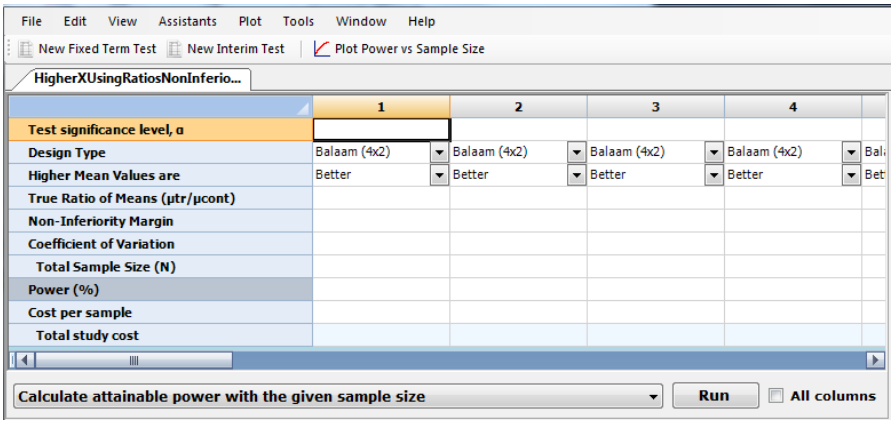


Figure 4.18.2. Higher-Order Cross-Over Design for Two Means – Non-Inferiority Test Table

The first calculation will be an example of total sample size for the Balaam Design (4x2).

16. First enter 0.025 for the "Test Significance level" row

17. Select "Balaam (4x2)" from the "Design Type" row and "Better" from the "Higher Mean Values are" row.

18. Enter 1 for "True Ratio of Means" and 0.5 for both the "Non-inferiority Margin" and "Within Standard Error" rows.

19. Finally, enter 80 in the "Power" row and the table will appear as per Figure 4.18.3.

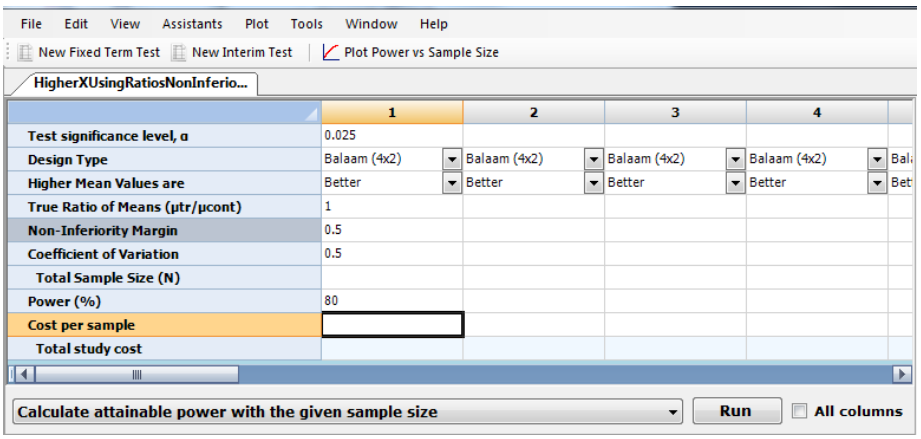


Figure 4.18.3. Values entered for Higher-Order Cross-Over Design – Non-Inferiority design

20. Select "Calculate required sample size for given power" from the dropdown menu beside the "Run" button. Then click "Run". This will give a result of 32 as in Figure 4.18.4.

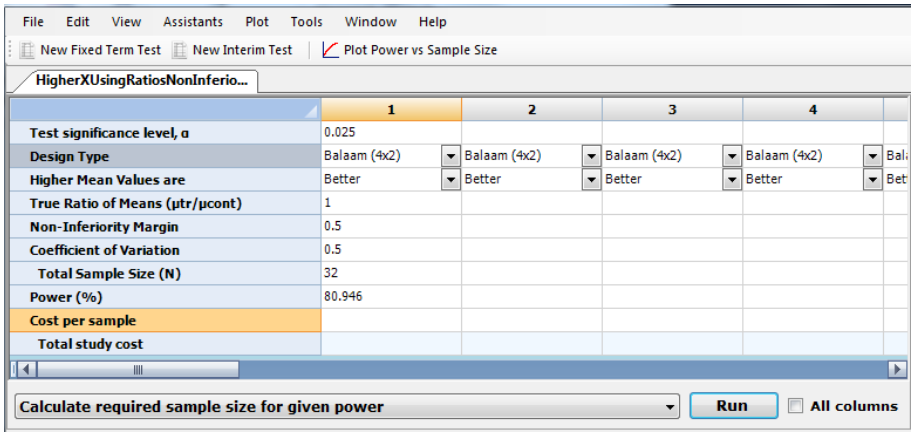


Figure 4.18.4. Completed Higher-Order Cross-Over Design – Non-Inferiority study design

21. To validate this example we open the "Higher-Order Cross-Over Design – Equivalence – using Ratios" using the same window as in step 2.

22. Enter the same values as from the non-inferiority table above except placing the one minus non-inferiority margin ($1 - |NIM|$) value (i.e. $1 - 0.5 = 0.5$) in the "Lower Equivalence Margin" row and making the upper equivalence margin an arbitrarily high value (100 in this example) and run the sample size calculation. This will give 32 as above as seen in Figure 4.18.5.

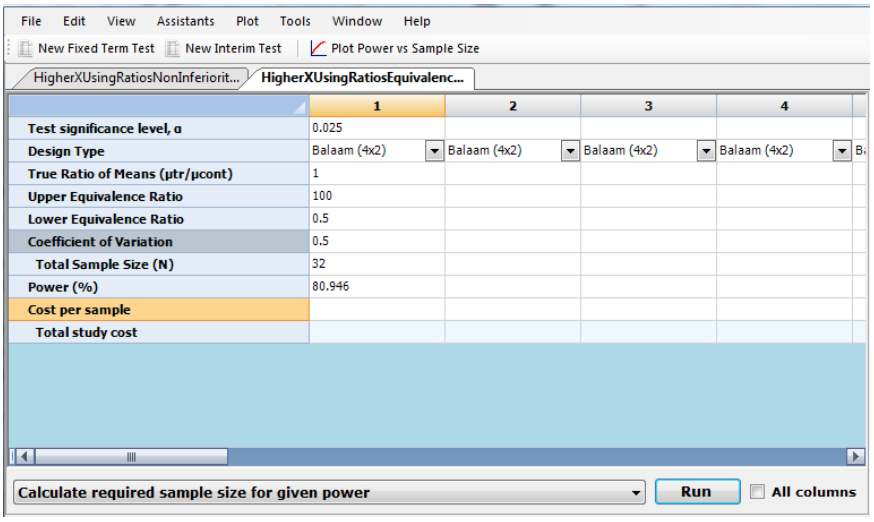


Figure 4.18.5. Validation of Non-inferiority table using Equivalence table

The next calculation shows the effect of design type on sample size.

23. First, return to non-inferiority table using the tab just below the "New Fixed Term Test" button.

24. Now enter 80 for Power in column 1, delete the "Total Sample Size" value. Then copy the values in column 1 across to columns 2 to 4. In columns 2 to 4, change design type to Two-Sequence Dual (2x3), Four Sequences & Two Periods (2x4) and Four Sequences and Periods (4x4) respectively. This will give a table as per Figure 4.18.6.

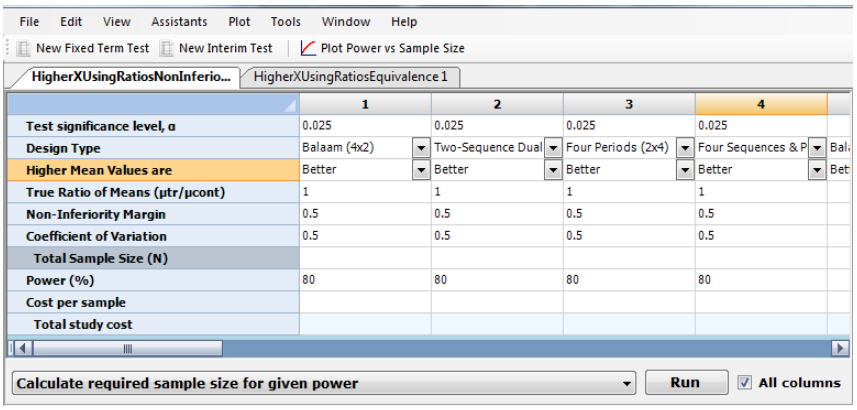


Figure 4.18.6. Analysis for Effect of Worse Means being Better and Effect of Design Type

25. Select "Calculate required sample size for given power" from the dropdown menu beside the "Run" button. Then tick the box to run "All Columns". Then click "Run".

This will give the resultant total sample sizes of 32, 7, 5 and 5 sequentially as in Figure 4.18.7 below.

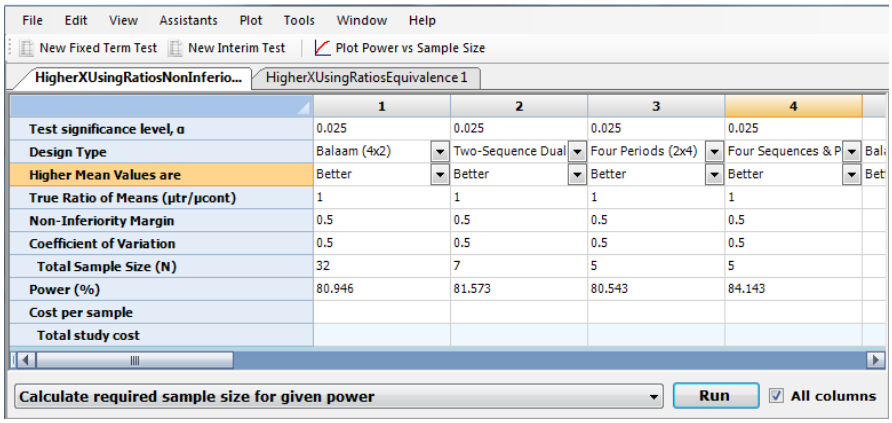


Figure 4.18.7. Results from Sensitivity analysis

26. Finally, by clicking on the desired study design column and going to the "Output" tab at the bottom of the screen, you can get an output statement for the results given by the current calculation as per Figure 4.18.8.

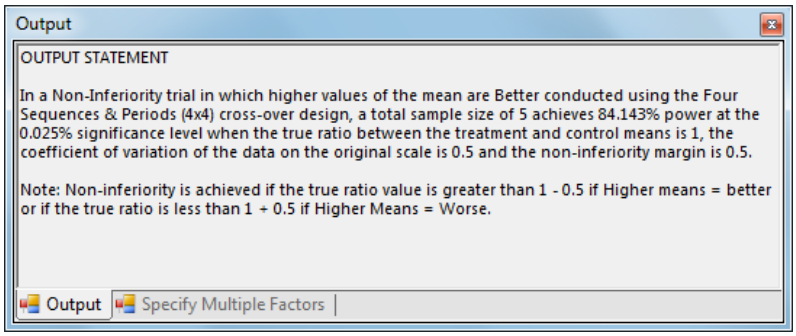


Figure 4.18.8. Study design Output statement

4.19. Higher-Order Cross-Over Design for Two Means – Equivalence - Using Ratios

4.19.1 Introduction

In a cross-over design each subject receives all treatments at least once with the objective of measuring study differences among the treatments. The term "cross-over" derives from the most common two treatments case, in which each subject crosses over from one treatment to another. The design assumes a washout period between treatments during which the response returns to its baseline. If this does not occur, there is said to be a carryover effect.

Cross-over designs are popular due to an expected increase in precision, since each subject effectively acts as their own control, and the concomitant reduction in study size. However, they do require more complex statistical analysis than parallel designs and do suffer from complications and strong assumptions related to time effects, carry-over effects and patient attrition. These issues and the statistical methodology are discussed more thoroughly in Chow et al. (2008) and Chen et al. (1997).

In this table, we look beyond the most common 2x2 design to higher-order designs where there can be up to four sequences of treatments and/or measured time periods

This table facilitates the calculation of the power and sample size for four higher-order designs using the multiplicative model, namely the 2x3, 2x4, 4x2 and 4x4 designs (discussed below), for two one-sided tests equivalence hypotheses. Power and sample size is computed using the method outlined by Chen et al. (1997).

4.19.2 Methodology

Before discussing statistical methodology, the structure of the four additional cross-over designs is given below. A cross-over design consists of two parameters: the sequences i.e. the pattern of treatments given to a patient and periods i.e. the number of time periods of which the outcome variable is measured. In the numeric summary, the AxB design indicates A number of sequences and B number of periods. Note that all of the following designs are for a two treatment design. The designs are as follows:

Table 4.19.1: Details of Higher-Cross Over Designs

Design Name	Advantage	Design
Balaam Design (4x2)	Allows estimation of intra-subject variabilities	Sequence Period1 Period2 1 A A 2 B B 3 A B 4 B A
Two-Sequence Dual Design (2x3)	Allows estimation of intra-subject variabilities	Sequence Period1 Period2 Period3 1 A B B 2 B A A
Four-Period with Two Sequences Design	Allows estimation of effect multiple sequential doses	Sequence Period1 Period2 Period3 Period4 1 A B B A 2 B A A B
Four-Period Design with Four Sequences	Allows estimation of effect multiple sequential doses	Sequence Period1 Period2 Period3 Period4 1 A A B B 2 B B A A 3 A B B A 4 B A A B

The statistical methodology for cross-over designs is based on ANOVA methodology and is outlined in detail in Chow and Liu (2000) and Chow et al. (2008). Note that this analysis is based on the multiplicative model from Chen et al. where the log-transformed data is analysed using the ANOVA model, as opposed to the additive model which analyses the unscaled data, which is covered in the equivalent table for the differences in means.

We now briefly discuss the two one-sided tests (TOST) equivalence hypothesis.

Equivalence hypothesis tests are based on Schuirmann's two-one sided tests (TOST) methodology. In these tests, the researcher specifies an estimate for the expected true ratio φ (i.e. $\mu_{Treat}/\mu_{Control}$) between treatments and two equivalence ratio margins, a lower ratio margin which is below 1 and an upper ratio margin which is above 1, which defines the range under which the true ratio would indicate the treatment and reference groups were equivalent i.e. if the true ratio falls between the lower and upper equivalence margins that implies the two treatment groups are equivalent on the outcome variable. This is outlined below:

The null hypothesis is given as follows:

$$H_0: \varphi \leq LEL \text{ or } \varphi \geq UEL, \text{where } LEL < 1, UEL > 1 \qquad \{4.19.1\}$$

where LEL is lower equivalence limit ratio and UEL is the upper equivalence limit ratio. The alternative hypothesis is as follows:

$$H_1: LEL < \varphi < UEL \qquad \{4.19.2\}$$

Thus equivalence is powered based on the probability that the true ratio falls between the two equivalence margins.

The sample size and power calculations for higher-cross over design is derived from the work of Chen et al. (1997). Note that Chen only covers the case of assessing equivalence under a mean ratio hypothesis using the coefficient of variance under the log-scaled data. However, the relationship between COV under the unscaled and log-scaled data is well understood and these can easily be derived from each other using the following relationships derived from Julious (2004):

$$X = \ln Y\,, Y = e^X \qquad \{4.19.3\}$$

$$\mu_Y = e^{\mu_X + \frac{\sigma_X^2}{2}}, \sigma_Y^2 = \mu_Y^2\left(e^{\sigma_X^2} - 1\right) \qquad \{4.19.4\}$$

$$=> COV_Y = \sqrt{\mu_Y^2\left(e^{\sigma_X^2} - 1\right)} \Big/ \mu_Y = \sqrt{\left(e^{\sigma_X^2} - 1\right)} \qquad \{4.19.5\}$$

$$=> \sigma_X = \sqrt{\ln(COV_Y^2 + 1)}\,, \mu_X = \frac{\mu_Y}{\ln(COV_Y^2 + 1)}$$

Thus hypotheses stated in terms of X can be stated in terms of Y and vice-versa. Note that the sigma terms are equivalent to the within-standard error from the ANOVA on the unscaled σ_Y and log-scaled data σ_X.

This table can be used to calculate the power, the total sample size, the upper equivalence limit ratio, the lower equivalence limit ratio, the test significance level, the true ratio and the coefficient of variation (of the unscaled data).

Calculations use a t-distribution approximation. To calculate power and total sample size the user must specify the test significance level (α), the higher-order cross-over design of interest, the true ratio ($\varphi = \mu_{Treat}/\mu_{Cont}$), the upper equivalence ratio limit (UEL), lower equivalence ratio limit (LEL) and the coefficient of variation (COV) under the unscaled data.

The formula for the power, derived from the methods from Chen et al. (1997), follows:

$$1 - \beta = T_v\left[\left(\frac{(-|\ln\varphi| + \ln UEL)}{CV_m\sqrt{b/n}}\right) - t_{V,1-\alpha}\right] - T_v\left[-\left(\frac{(|\ln\varphi| - \ln LEL)}{CV_m\sqrt{b/n}}\right) + t_{V,1-\alpha}\right] \qquad \{4.19.6\}$$

where T_V is the cumulative t-distribution function with V degrees of freedom, $t_{V,1-\alpha}$ is the inverse cumulative t-distribution evaluated at $1-\alpha$, b is a design parameter (both V and b are defined by design choice, see below), CV_m is the coefficient of variation of the logged data and is equal to $\sqrt{e^{MSE_{logged}}-1}$ and n is **average number of subjects per sequence** i.e. $N_{Total}/\#\ sequences$. V and b are defined by the study design and are given by following equations:

$$Balaam\ (4x2): V = 4n-3, b = 2$$
$$Two-Sequence\ Dual\ (3x2): V = 4n-4, b = 3/4$$
$$Four\ Periods\ \&\ Two\ Sequences\ (4x2): V = 6n-5, b = 11/20$$
$$Four\ Periods\ and\ Four\ Sequences\ (4x4): V = 12n-5, b = 1/4$$

A closed form equation is not used to calculate the total sample, the upper equivalence limit, the lower equivalence limit, the test significance levels, the within standard error or the smallest true detectable ratio. Instead a search algorithm is used. The search algorithm calculates power at various values for the relevant parameter until the desired power is reached.

Note also that the total sample size calculated is the "exact" sample size (i.e. the first value of N_{Total} which achieves the desired power). However, for an equal sample size per sequence you need to round to the nearest 2 for the 2x3 and 2x4 designs and nearest four for the 4x2 and 4x4 designs.

4.19.3 Examples

Example 1: Validation example calculating required sample size for a given power
The following validation example is from Table 5 from Chen et al. Note that, as mentioned above, Chen parameterises via the coefficient of variation under the log-scaled data but this is converted into our unscaled coefficient below. A sensitivity analysis is then used to show the effect of changing the true ratio of means.

The following steps outline the procedure for this example:

11. Open nQuery + nTerim 4.0 via the start menu or desktop shortcut. Click "New Fixed Term Test" from the top of the window. The "Study Goal and Design" window will appear as Figure 4.19.1 below.

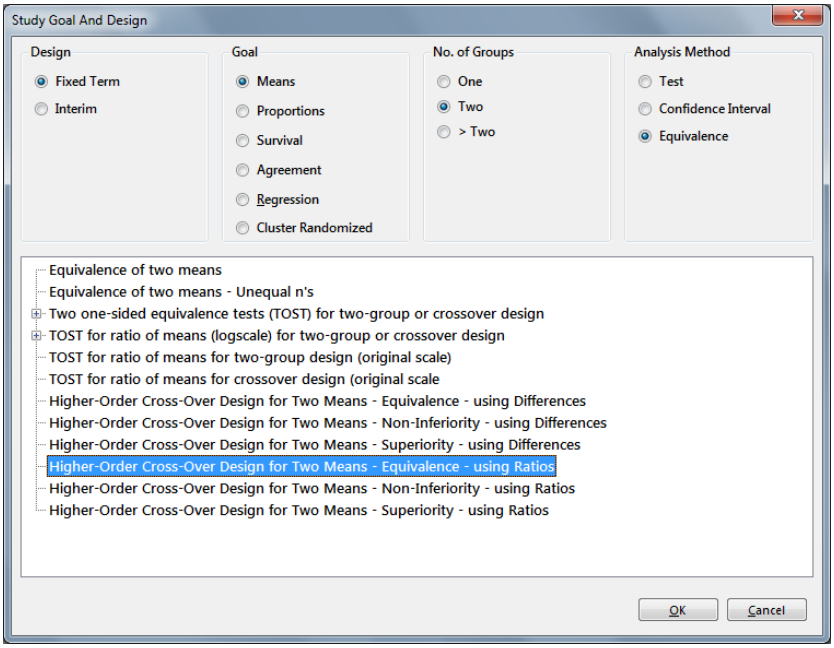


Figure 4.19.1. Study Goal and Design Window

12. In order to select the “Higher-Order Cross Over Design – Equivalence – Using Ratios” design table, navigate through the Study Goal and Design Window by selecting "Fixed Term" as the Design, "Means" as the Goal, "Two" as the Number of Groups and "Equivalence" as the Analysis Method. Then click "OK" and the test window will appear as per Figure 4.19.2.

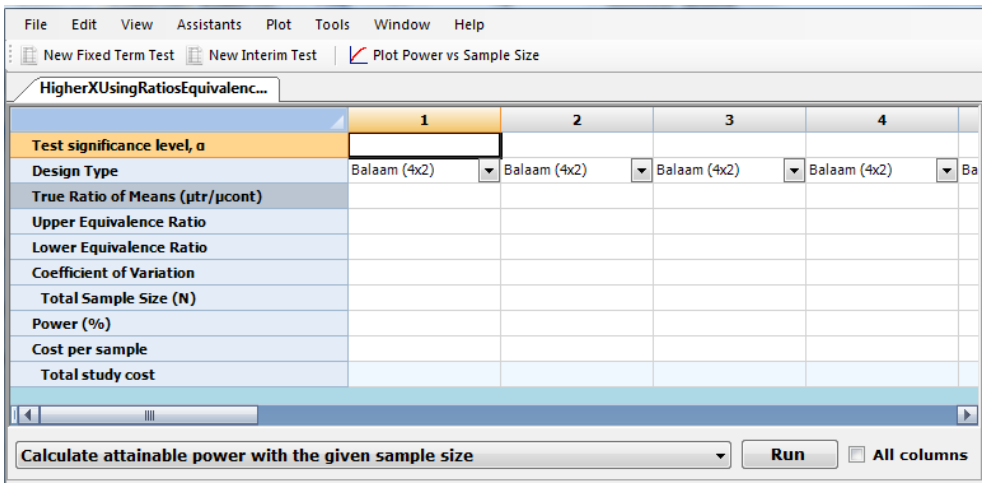


Figure 4.16.2. Higher-Order Cross-Over Design for Two Means – Equivalence Test Table

The first calculation will be an example of total sample size for the Balaam Design (4x2) taken from Table 1 from Chen et al.

13. First enter 0.05 for the “Test Significance level” row

14. Select “Balaam (4x2”) from the “Design Type” row and enter 1 for “True Ratio of Means”.

15. Enter 1.25 for the “Upper Equivalence Ratio”, 0.8 for the “Lower Equivalence Ratio” and 0.10025 for the Coefficient of Variation row (derived from $\sqrt{e^{0.1} - 1}$).

16. Finally, enter 80 in the Power row and the table will appear as per Figure 4.19.3.

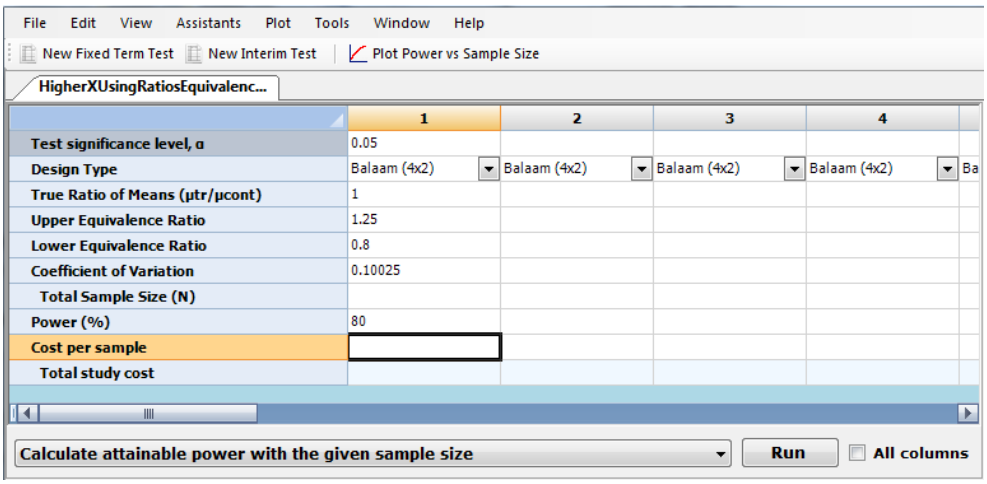


Figure 4.19.3. Values entered for Higher-Order Cross-Over Design – Equivalence design

17. Select “Calculate required sample size for given power” from the dropdown menu beside the “Run” button. Then click “Run”. This will give a result of 16 as in Figure 4.19.4. This agrees with the value from Table 5 of Chen et al.

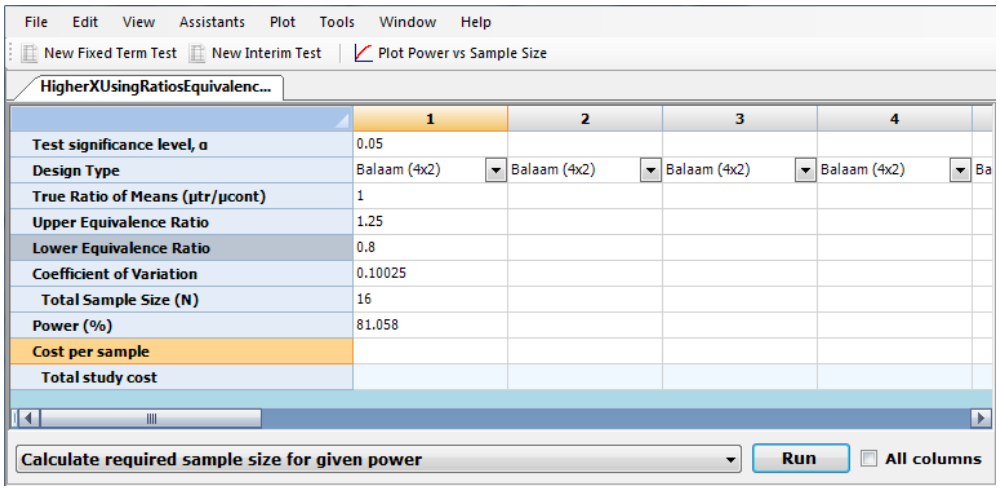
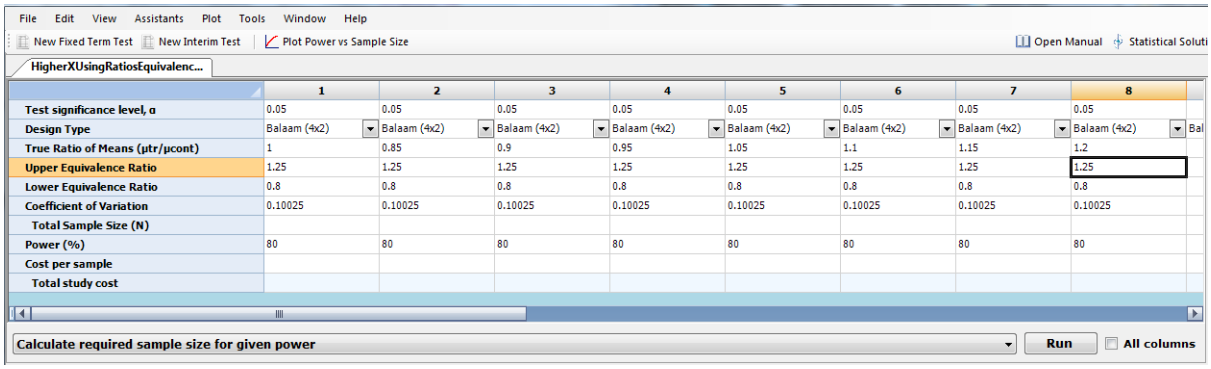


Figure 4.19.4. Completed Higher-Order Cross-Over Design – Equivalence study design

The next calculation shows the effect of changing the true ratio in means by replicating the first row of Table 5 of Chen et al.

18. Enter 80 for Power in column 1 and delete the “Total Sample Size” value. Then copy the values in column 1 across to columns 2 to 8. In columns 2 to 8, replace the values of true ratio to 0.85, 0.9, 0.95, 1.05, 1.1, 1.15 and 1.2 respectively. This will give a table as per Figure 4.19.5.

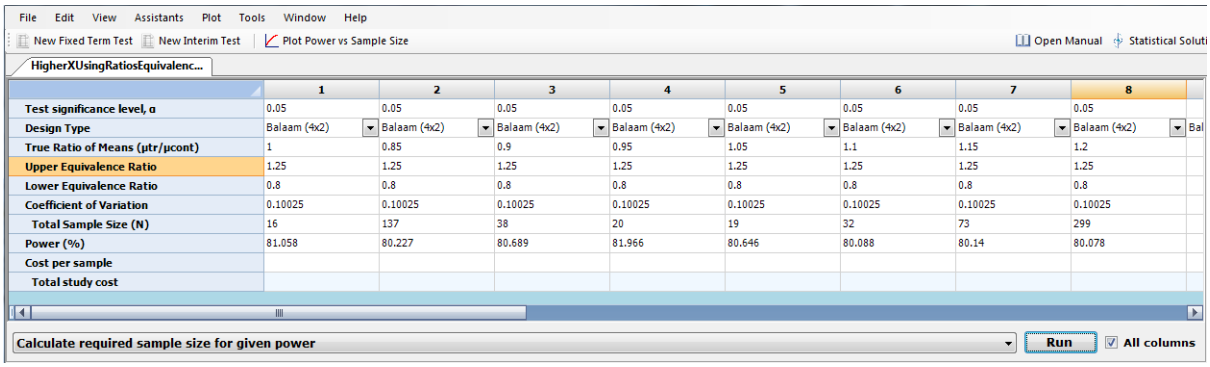


	1	2	3	4	5	6	7	8
Test significance level, α	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
Design Type	Balaam (4x2)	Balaam (4x2)	Balaam (4x2)	Balaam (4x2)	Balaam (4x2)	Balaam (4x2)	Balaam (4x2)	Balaam (4x2)
True Ratio of Means (μtr/μcont)	1	0.85	0.9	0.95	1.05	1.1	1.15	1.2
Upper Equivalence Ratio	1.25	1.25	1.25	1.25	1.25	1.25	1.25	1.25
Lower Equivalence Ratio	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
Coefficient of Variation	0.10025	0.10025	0.10025	0.10025	0.10025	0.10025	0.10025	0.10025
Total Sample Size (N)								
Power (%)	80	80	80	80	80	80	80	80
Cost per sample								
Total study cost								

Figure 4.19.5. Analysis for Effect of Changing the True Ratio

19. Select "Calculate required sample size for given power" from the dropdown menu beside the "Run" button. Then tick the box to run "All Columns". Then click "Run".

This will give the resultant total sample sizes of 16, 137, 38, 20, 19, 32, 73 and 299 sequentially as in Figure 4.19.6 below. There are some discrepancies with Chen et al, however they used a one-sided approximation for sample size rather than an exact search procedure (pp. 759) so this is not unexpected.



	1	2	3	4	5	6	7	8
Test significance level, α	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
Design Type	Balaam (4x2)	Balaam (4x2)	Balaam (4x2)	Balaam (4x2)	Balaam (4x2)	Balaam (4x2)	Balaam (4x2)	Balaam (4x2)
True Ratio of Means (μtr/μcont)	1	0.85	0.9	0.95	1.05	1.1	1.15	1.2
Upper Equivalence Ratio	1.25	1.25	1.25	1.25	1.25	1.25	1.25	1.25
Lower Equivalence Ratio	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
Coefficient of Variation	0.10025	0.10025	0.10025	0.10025	0.10025	0.10025	0.10025	0.10025
Total Sample Size (N)	16	137	38	20	19	32	73	299
Power (%)	81.058	80.227	80.689	81.966	80.646	80.088	80.14	80.078
Cost per sample								
Total study cost								

Figure 4.19.6. Results from Sensitivity analysis

20. Finally, by clicking on the desired study design column and going to the "Output" tab at the bottom of the screen, you can get an output statement for the results given by the current calculation as per Figure 4.19.7.

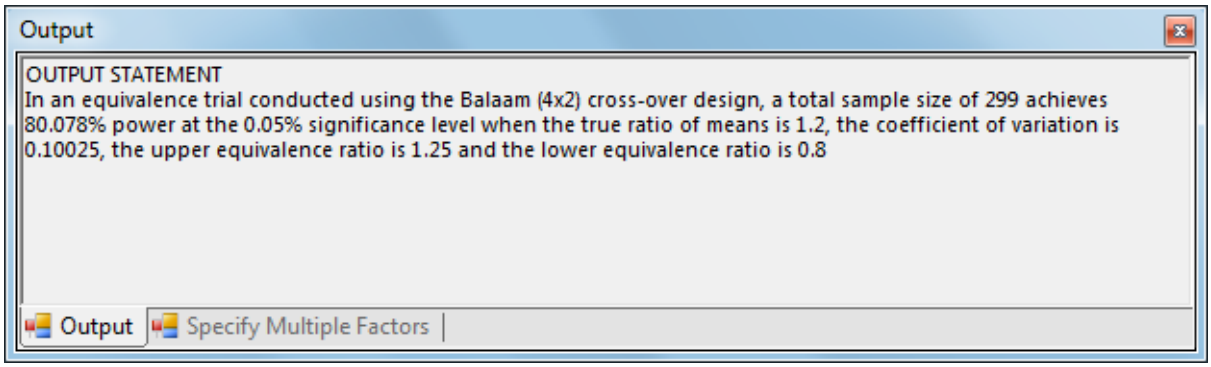


Figure 4.19.7. Study design Output statement

4.20. Two Exponential Means

4.20.1 Introduction

The exponential model is a widely used method to establish the reliability of a process and thus is widely used in medical and industrial activities. Thus, the comparison of two exponential mean parameters is a common test to compare reliability. The test used here assumes that the exponential mean parameters being used here are continuous, positive outcomes such as a mean lifetime.

This table facilitates the calculation of the power and sample size for hypothesis tests comparing two exponential means. Power and sample size is computed using the method outlined by Bain and Engelhardt (1991) and Desu and Raghavarao (1990).

4.20.2 Methodology

This table provides sample size and power calculations for studies which will be comparing two mean exponential lifetimes. The sample size calculation for two exponential means is taken from Desu and Raghavarao (1990). The test of interest is an F-test for the ratio of two exponential lifetimes which have the following distribution:

$$f_i(t) = \frac{1}{\theta_i} e^{-\frac{t}{\theta_i}}, where\ i = 1,2. \qquad \{4.20.1\}$$

Where θ is the mean time to failure or the mean time between failures. Thus at time t, the probability of a unit keeps running (i.e. its reliability) is given by $R_i(s) = e^{-\frac{t}{\theta_i}}$.

Note that the model is based on the items being sampled without replacement and where all items are allowed to fail. If all items do not fail then the effective sample size is the number of failures rather than the number of items. The mean lifetime estimator is the sum of the times at which failures occurred in a group divided by the number of failures in a group (where t is either the time of failure or the end of the study).

This table can be used to calculate the power, the sample size given a sample size ratio, the treatment sample size (N_2), the control sample size (N_1), the significance level and the control and treatment group exponential mean lifetimes given all other terms in the table are specified.

To calculate power and sample size, the user must specify the test significance level (α), whether to use a one or two sided test, the control group exponential mean lifetime (θ_1) and the treatment group exponential mean lifetime (θ_2).

The power formula uses the F-test of the ratio to yield the following formulae:

$$\begin{cases} 1-\beta = F_{2N_1,2N_2}\left[\frac{\theta_2}{\theta_1}F^{-1}_{2N_1,2N_2}(\alpha)\right], & \theta_1 < \theta_2, One-sided \\ 1-\beta = 1-F_{2N_1,2N_2}\left[\frac{\theta_2}{\theta_1}F^{-1}_{2N_1,2N_2}(1-\alpha)\right], & \theta_1 > \theta_2, One-sided \\ 1-\beta = 1-F_{2N_1,2N_2}\left[\frac{\theta_2}{\theta_1}F^{-1}_{2N_1,2N_2}(1-\alpha/2)\right] + F_{2N_1,2N_2}\left[\frac{\theta_2}{\theta_1}F^{-1}_{2N_1,2N_2}(\alpha/2)\right], Two-sided \end{cases} \quad \{4.20.2\}$$

where $F_{2N_1,2N_2}$ is the cumulative F-distribution with 2*N_1 and 2*N_2 degrees of freedom and $F^{-1}_{2N_1,2N_2}$ is the inverse cumulative F-distribution.

A closed form equation is not used to calculate the sample size calculations, the control mean, the treatment mean and the significance level. Instead a search algorithm is used. The search algorithm calculates power at various values for the relevant parameter until the desired power is reached.

4.20.3 Examples

Example 1: Validation example calculating required sample size for a given power
The following example is taken from Example 9.19 on page 207 of Matthews (2010) where a power calculation problem is conducted. Then we conduct a sensitivity analysis to show the effect of changing the control group exponential mean lifetime and power.

The following steps outline the procedure for this example:

1. Open nQuery + nTerim 4.0 via the start menu or desktop shortcut. Click “New Fixed Term Test” from the top of the window. The “Study Goal and Design” window will appear as Figure 4.20.1 below.

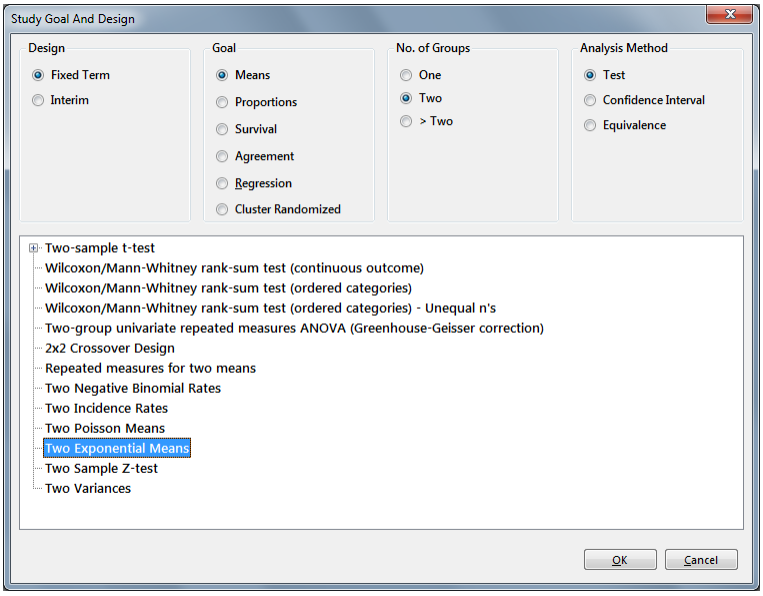


Figure 4.20.1. Study Goal and Design Window

2. In order to select the “Two Exponential Means” design table, navigate through the Study Goal and Design Window by selecting "Fixed Term" as the Design, "Means" as the Goal, "Two" as the Number of Groups and "Test" as the Analysis Method. Then click "OK" and the test window will appear as per Figure 4.20.2.

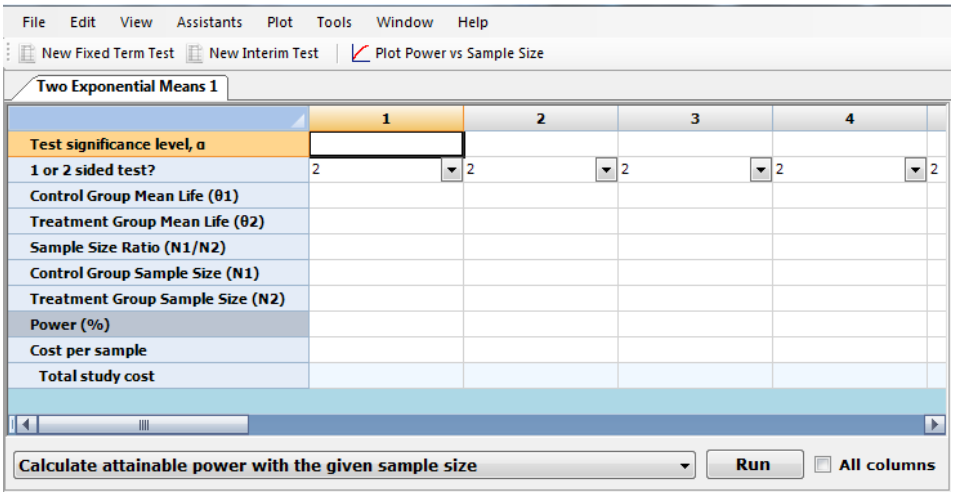


Figure 4.20.2. Two Exponential Means Test Table

The first calculation will be for power, to begin we enter the values as follows.

3. First enter 0.05 for the “Test Significance level” row and 1 for “1 or 2 sided test?”

4. Then enter 200 (hours) and 100 (hours) for the “Control Group Mean Life” and “Treatment Group Mean Life” variable rows respectively.

5. Then, enter 30 in both the “Control Group Sample Size” and “Treatment Group Sample Size” rows. The table will appear as per Figure 4.20.3.

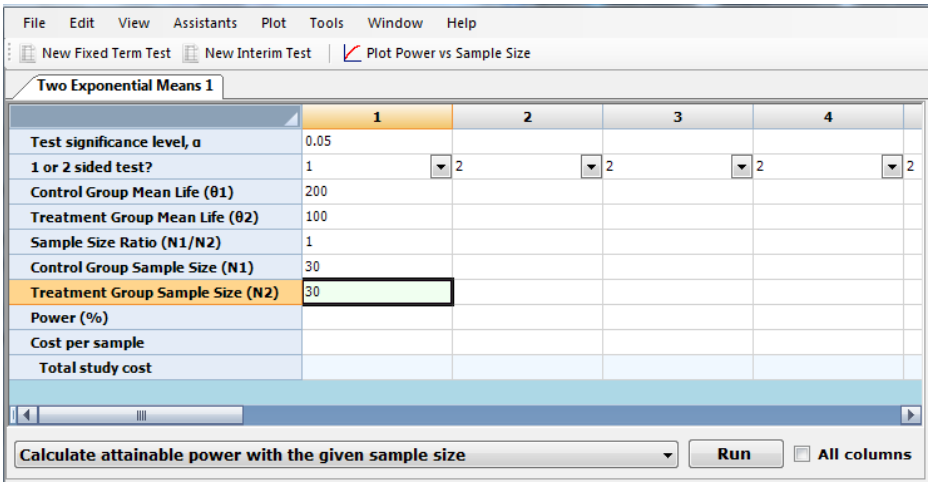
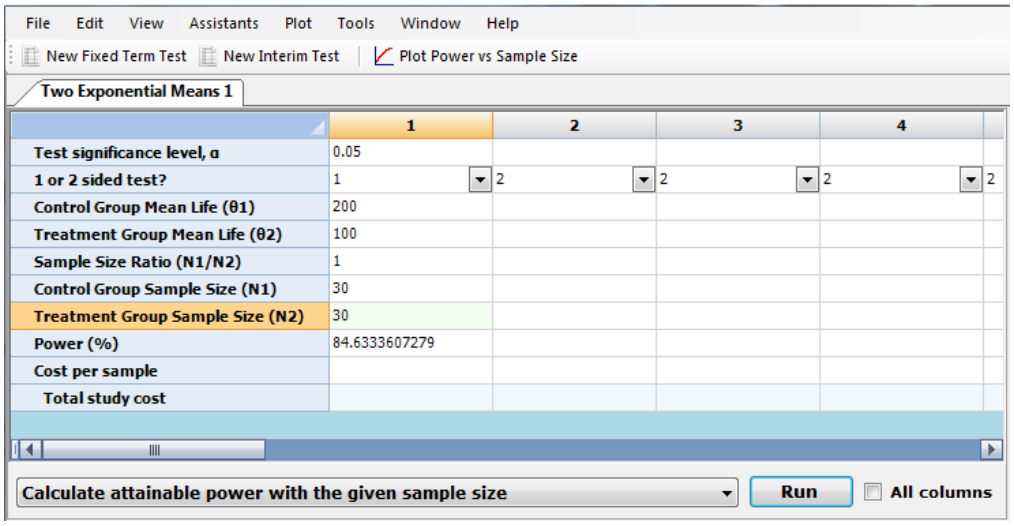


Figure 4.20.3. Values entered for Two Exponential Mean study design

6. Select “Calculate attainable power with the given sample size” from the dropdown menu beside the “Run” button. Then click “Run”. This will give a result of 84.633 for

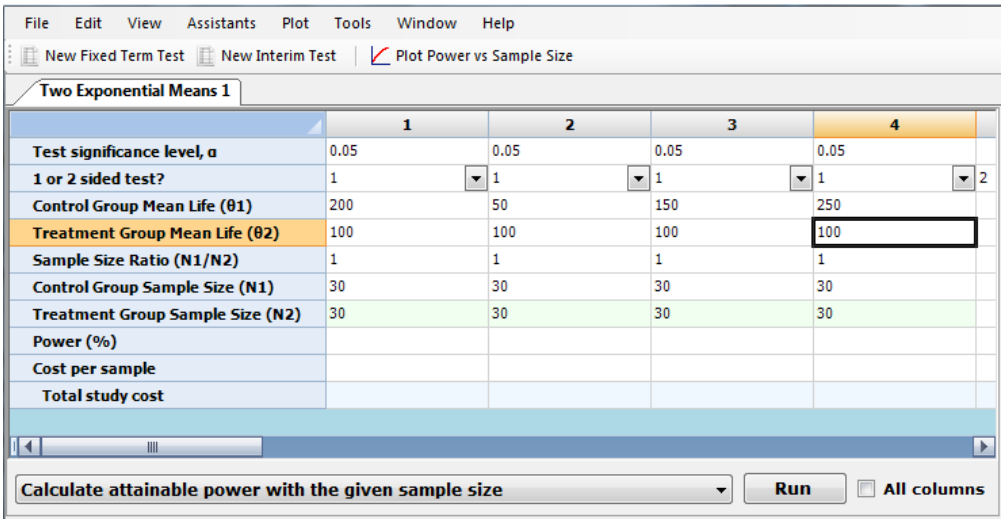
power as in Figure 4.20.4. This agrees with the value of 84.6% given in Matthews (2010).



	1	2	3	4
Test significance level, α	0.05			
1 or 2 sided test?	1	2	2	2
Control Group Mean Life (θ1)	200			
Treatment Group Mean Life (θ2)	100			
Sample Size Ratio (N1/N2)	1			
Control Group Sample Size (N1)	30			
Treatment Group Sample Size (N2)	30			
Power (%)	84.6333607279			
Cost per sample				
Total study cost				

Figure 4.20.4. Completed Two Exponential Means study design

7. The next calculation is a sensitivity analysis for power where we change the control group mean life to investigate the impact this has on the power estimate. To do this, copy the same values across to columns 2 to 4 and delete the values for power. Then, change the values for columns 2 to 4 to 50, 150 and 250 respectively. This will give a table as per Figure 4.20.5.

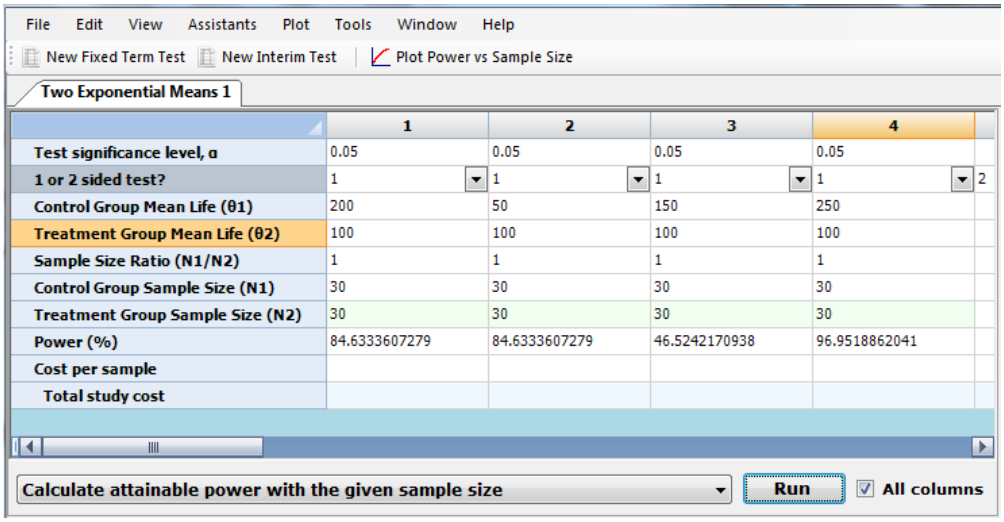


	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
1 or 2 sided test?	1	1	1	1
Control Group Mean Life (θ1)	200	50	150	250
Treatment Group Mean Life (θ2)	100	100	100	100
Sample Size Ratio (N1/N2)	1	1	1	1
Control Group Sample Size (N1)	30	30	30	30
Treatment Group Sample Size (N2)	30	30	30	30
Power (%)				
Cost per sample				
Total study cost				

Figure 4.20.5. Sensitivity analysis around the Control Group Mean Life

8. Select “Calculate attainable power with the given sample size” from the dropdown menu beside the “Run” button. Then tick the box to run “All Columns”. Then click “Run”.

This will give the resultant powers of 84.633, 84.633, 46.524 and 96.951 sequentially as in Figure 4.20.6. Note that the equivalent ratios above and below (200/100 & 50/100) the treatment mean life given the same answer.



	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
1 or 2 sided test?	1	1	1	1
Control Group Mean Life (θ1)	200	50	150	250
Treatment Group Mean Life (θ2)	100	100	100	100
Sample Size Ratio (N1/N2)	1	1	1	1
Control Group Sample Size (N1)	30	30	30	30
Treatment Group Sample Size (N2)	30	30	30	30
Power (%)	84.6333607279	84.6333607279	46.5242170938	96.9518862041
Cost per sample				
Total study cost				

Figure 4.20.6. Results from Sensitivity analysis

9. Finally, by clicking on the desired study design column and going to the "Output" tab at the bottom of the screen, you can get an output statement for the results given by the current calculation as per Figure 4.20.7 below.

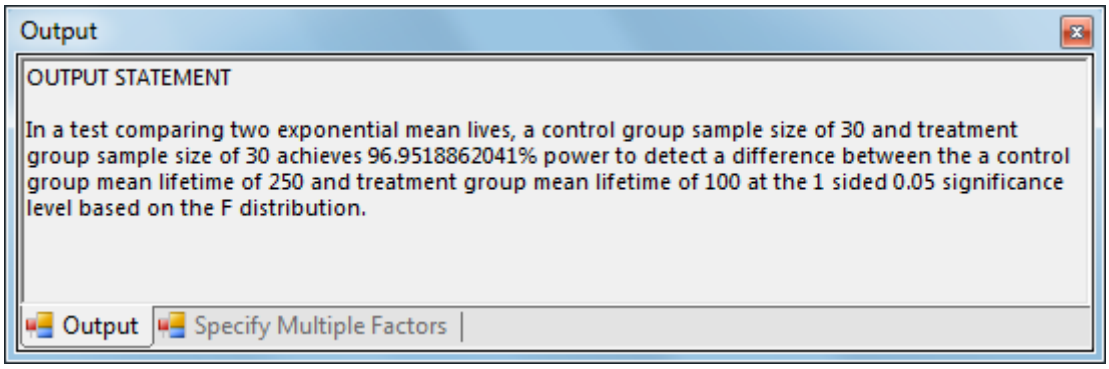


Figure 4.20.7. Study design Output statement

This Output statement can then be easily transferred directly from the output window into a report.

4.21. One Sample Z-test

4.21.1 Introduction

The Z-test is one of the most commonly used statistical tests in nearly all areas of statistical interest. One of its most common and simplest implementations is in the context of the analysis of continuous data which is normally or approximately normally distributed. One such use is the statistical comparison of a single sample mean to a pre-specified mean index (or one-sample case) to determine whether the sample mean is unequal to the index mean.

This table facilitates the calculation of the power and sample size for hypothesis tests comparing one sample mean to an index mean using the Z-test. Power and sample size calculations are available in most introductory sample size texts. Here we use Machin et al. (2009) and Chow et al. (2008).

4.21.2 Methodology

This table provides sample size and power calculations for studies which will be comparing one mean to a mean index using a Z-test. The sample size calculation for one mean is taken from Machin et al. (2009) and Chow et al. (2008). The test of interest is for the following Z-statistic:

$$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{N}} \qquad \{4.21.1\}$$

Where $\bar{x}$ is the measured or "true" sample mean under the alternative hypothesis, μ_0 is the index or null hypothesis mean (where the sample mean is expected to equal the index mean under the null hypothesis) and σ is the known standard deviation.

The Z-test is usually used instead of the more common t-test in two contexts: a) when the population variance is known or b) when the expected sample size is high (i.e. where the Z and t-tests converge). The usage of this sample size table as opposed to the t-test table depends on which of these tests will be used by the end user for their statistical analysis.

However, both tables require the assumptions that the data is continuous, the underlying data distribution is normal and that the sample used was randomly selected i.e. every possible participant are equally likely to be selected in a sample, be fulfilled. If these assumptions are violated then the usage of these tables is incorrect and the appropriate alternative parametric or non-parametric equivalent should be used.

This table can be used to calculate the power, the sample size, the significance level, the null hypothesis mean, the alternative hypothesis mean and the standard deviation given all other terms in the table are specified.

To calculate power and sample size, the user must specify the test significance level (α), whether to use a one or two sided test, the null hypothesis mean (μ_0), the alternative

hypothesis mean (μ_1 i.e. $\bar{x}$), a finite population adjustment (if any) and the standard deviation.

Note that if the finite population adjustment is set to zero or less than the sample size (if specified in the table) then the calculation will be conducted assuming no adjustment for population size (i.e. assuming infinite population)

The power formula uses the Z-test for the mean to yield the following formula:

$$\begin{cases} 1-\beta = \Phi\left[\frac{|\bar{x}-\mu_0|}{\sigma/N} - z_{1-\alpha}\right], \quad One-sided \\ 1-\beta = \Phi\left[\frac{|\bar{x}-\mu_0|}{\sigma/N} - z_{1-\alpha/2}\right] + \Phi\left[-\left(\frac{|\bar{x}-\mu_0|}{\sigma/N}\right) - z_{1-\alpha/2}\right], Two-sided \end{cases} \quad \{4.21.2\}$$

where Φ is the cumulative Z distribution evaluated at the calculated value and $z_{1-\alpha}$ is the inverse cumulative Z-distribution evaluated at the specified probability. If a finite population adjustment is used then the standard deviation is specified instead as the following:

$$\sigma_{adj} = \sqrt{\sigma^2\left(\frac{1-N}{N_{Population}}\right)} \quad \{4.21.3\}$$

Where $N_{Population}$ is the population specified in the Finite Population Adjustment row.

The other formulae use an approximation in the two-sided case where the second term in the two-sided case is ignored since this value will usually be very small. It can then be shown through rearrangement that the following equations hold for the following calculable parameters in the two-sided case:

$$N = \left(\frac{\left(z_{1-\frac{\alpha}{2}} - z_{1-\beta}\right)^2 \sigma^2}{(\bar{x}-\mu_0)^2}\right) \quad \{4.21.4\}$$

$$\bar{x} = \mu_0 \mp \left(\frac{\sigma\left(z_{1-\frac{\alpha}{2}} - z_{1-\beta}\right)}{\sqrt{N}}\right) \quad \{4.21.5\}$$

$$\mu_0 = \bar{x} \mp \left(\frac{\sigma\left(z_{1-\frac{\alpha}{2}} - z_{1-\beta}\right)}{\sqrt{N}}\right) \quad \{4.21.6\}$$

$$\begin{cases} \sigma = \dfrac{|\bar{x} - \mu_0|\sqrt{N}}{\left(z_{1-\frac{\alpha}{2}} - z_{1-\beta}\right)}, & No\ Pop.Adjust. \\ \sigma = \sqrt{\dfrac{\left[\dfrac{|\bar{x} - \mu_0|\sqrt{N}}{\left(z_{1-\frac{\alpha}{2}} - z_{1-\beta}\right)}\right]^2 N_{Population}}{1 - N}}, & With\ Pop.Adjust \end{cases} \qquad \{4.21.7\}$$

Where $z_{1-\frac{\alpha}{2}}$ becomes $z_{1-\alpha}$ for the one-sided case.

A closed form equation is not used to calculate the significance level. Instead a search algorithm is used. The search algorithm calculates power at various values for the relevant parameter until the desired power is reached.

4.21.3 Examples

Example 1: Validation example calculating required sample size for a given power
The following example is taken from page 56 of Chow et al. (2008) where a sample size calculation problem is conducted. Then we conduct a sensitivity analysis to show the effect of changing the standard deviation and having a finite population adjustment has on sample size.

The following steps outline the procedure for this example:

1. Open nQuery + nTerim 4.0 via the start menu or desktop shortcut. Click “New Fixed Term Test” from the top of the window. The “Study Goal and Design” window will appear as Figure 4.21.1 below.

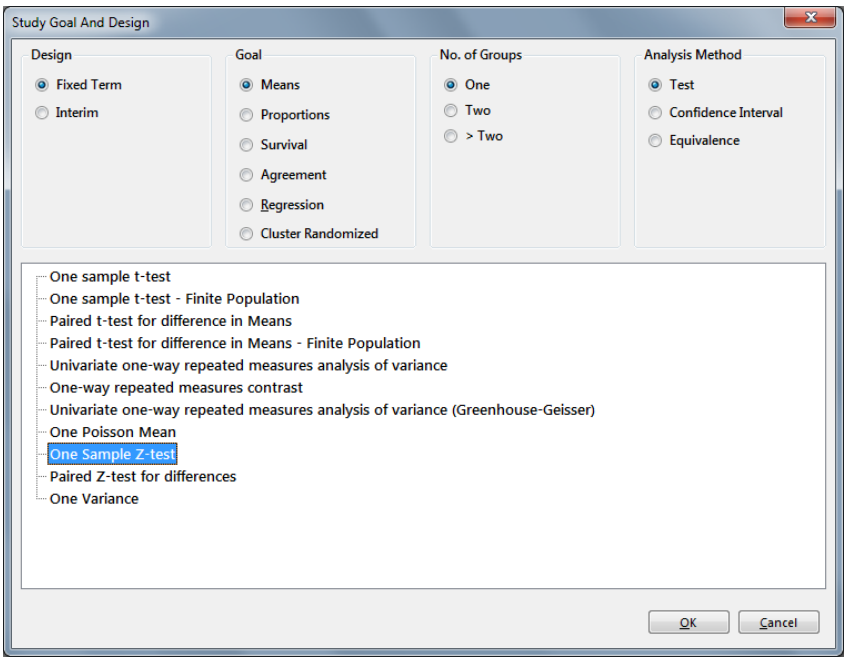


Figure 4.21.1. Study Goal and Design Window

2. In order to select the “One Sample Z Test” design table, navigate through the Study Goal and Design Window by selecting "Fixed Term" as the Design, "Means" as the Goal, "One" as the Number of Groups and "Test" as the Analysis Method. Then click "OK" and the test window will appear as per Figure 4.21.2.

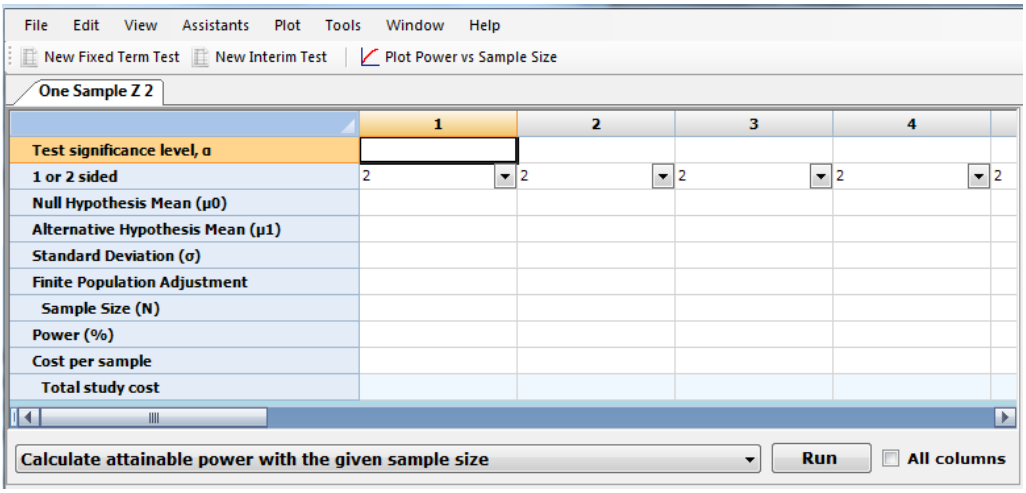


Figure 4.21.2. One Sample Z-test Table

The first calculation will be for sample size taken from Chow et al. We enter the values as follows:

3. First enter 0.05 for the “Test Significance level” row and 2 for “1 or 2 sided test?”

4. Then enter 1.5 and 2.0 for the “Null Hypothesis Mean” and “Alternative Hypothesis Mean” variable rows respectively.

5. Enter 1 for the “Standard Deviation” row and 0 in the “Finite Population Adjustment” row (equivalent to setting for the infinite population assumption case).

6. Finally, enter 80 in the “Power” row. The table will appear as per Figure 4.21.3.

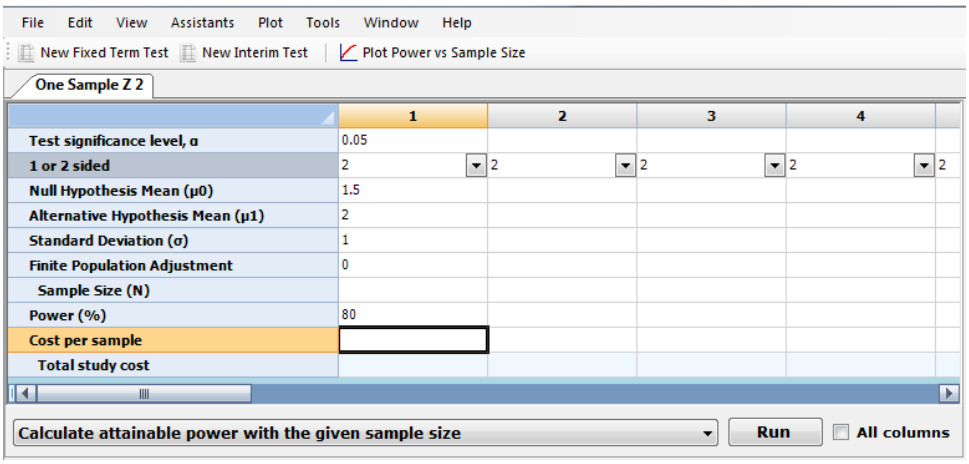
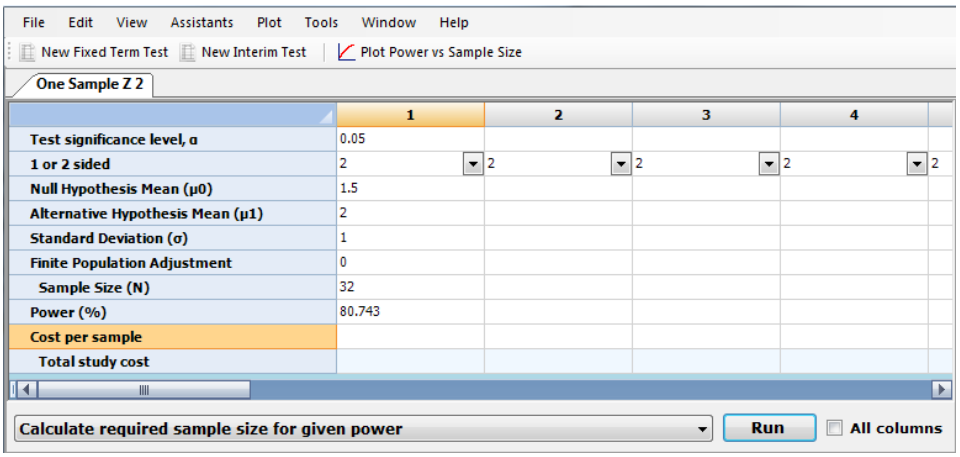


Figure 4.21.3. Values entered for One Sample Z-test study design

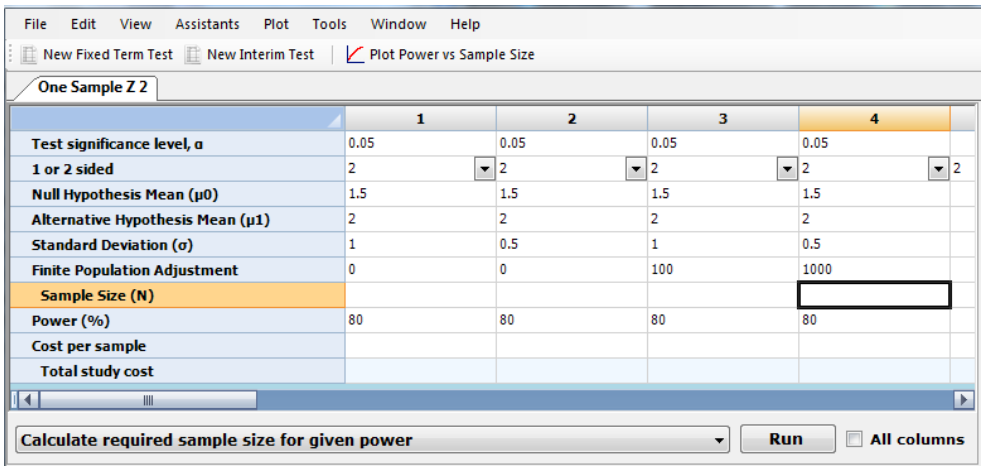
7. Select “Calculate required sample size for given power” from the dropdown menu beside the “Run” button. Then click “Run”. This will give a result of 32 for sample size as in Figure 4.21.4. This agrees with the value of 32 given in Chow et al.



	1	2	3	4
Test significance level, α	0.05			
1 or 2 sided	2	2	2	2
Null Hypothesis Mean (μ0)	1.5			
Alternative Hypothesis Mean (μ1)	2			
Standard Deviation (σ)	1			
Finite Population Adjustment	0			
Sample Size (N)	32			
Power (%)	80.743			
Cost per sample				
Total study cost				

Figure 4.21.4. Completed One Sample Z-test study design

8. The next calculation is a sensitivity analysis for sample size where we change the standard deviation and add a finite population adjustment to investigate the impact this has on the sample size estimate. To do this, copy the same values across to columns 2 to 4 where you delete the values for sample size and replace the value for power with 80. Then, change the values for columns 2 to 4 in the “Standard Deviation” row to 0.5, 1 and 0.5 respectively. Then, change the values for columns 2 to 4 in the “Finite Population Adjustment” row to 0, 100 and 1000 respectively. This will give a table as per Figure 4.21.5.



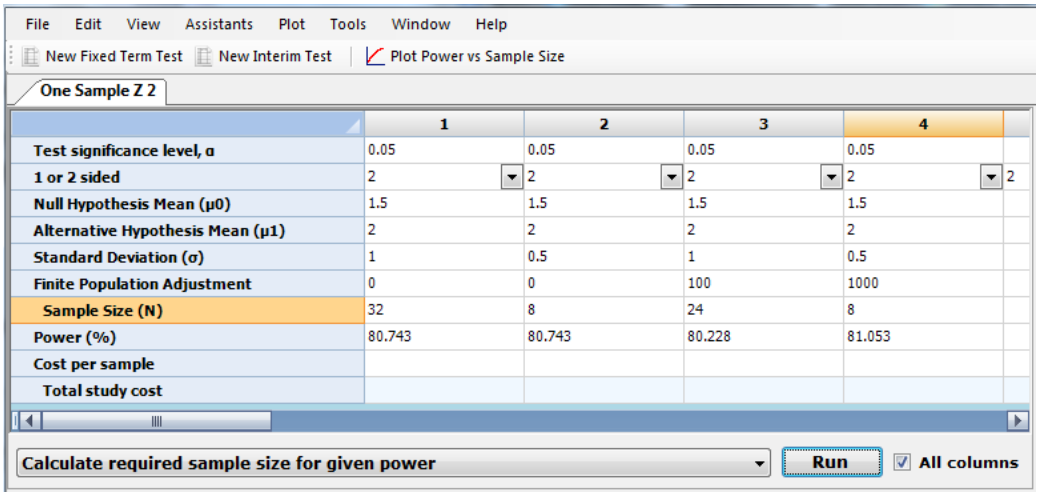
	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
1 or 2 sided	2	2	2	2
Null Hypothesis Mean (μ0)	1.5	1.5	1.5	1.5
Alternative Hypothesis Mean (μ1)	2	2	2	2
Standard Deviation (σ)	1	0.5	1	0.5
Finite Population Adjustment	0	0	100	1000
Sample Size (N)				
Power (%)	80	80	80	80
Cost per sample				
Total study cost				

Figure 4.21.5. Sensitivity analysis around the Standard Deviation and Finite Pop. Adjustment

9. Select “Calculate required sample size for given power” from the dropdown menu beside the “Run” button. Then tick the box to run “All Columns”. Then click “Run”.

This will give the resultant sample sizes of 32, 8, 24 and 8 sequentially as in Figure 4.21.6. Note that column 2 & 4 gave the same answer reflecting the larger increases in power per increase in N and the convergence between the infinite and finite cases as the assumed

finite population increases in value. Note though that the exact power has changed between them.



	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
1 or 2 sided	2	2	2	2
Null Hypothesis Mean (μ0)	1.5	1.5	1.5	1.5
Alternative Hypothesis Mean (μ1)	2	2	2	2
Standard Deviation (σ)	1	0.5	1	0.5
Finite Population Adjustment	0	0	100	1000
Sample Size (N)	32	8	24	8
Power (%)	80.743	80.743	80.228	81.053
Cost per sample				
Total study cost				

Figure 4.21.6. Results from Sensitivity analysis

10. Finally, by clicking on the desired study design column and going to the "Output" tab at the bottom of the screen, you can get an output statement for the results given by the current calculation as per Figure 4.21.7 below.

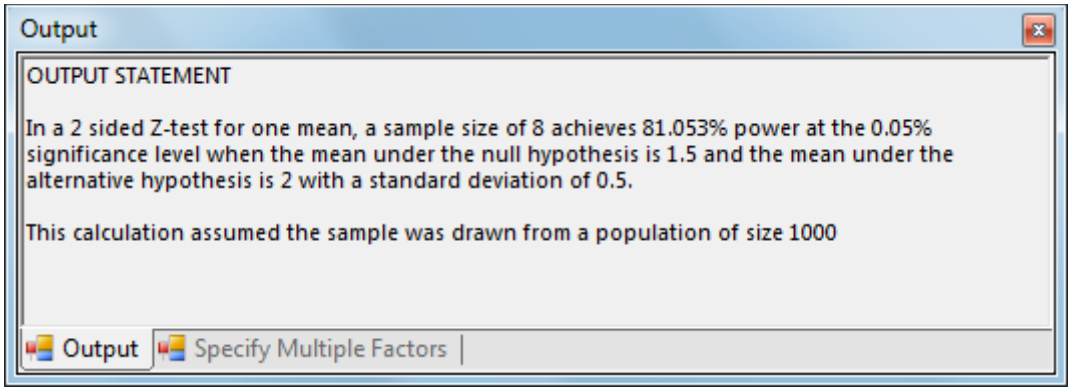


Figure 4.21.7. Study design Output statement

This Output statement can then be easily transferred directly from the output window into a report.

4.22. Paired Samples Z-test

4.22.1 Introduction

The Z-test is one of the most commonly used statistical tests in nearly all areas of statistical research. One of its most common and simplest implementations is in the context of the analysis of continuous data which is normally or approximately normally distributed. One such use is the statistical comparison of the mean of the differences between paired subjects to a pre-specified paired mean difference index (or paired sample case) to determine whether the mean of the paired differences is unequal to the index paired mean difference.

This table facilitates the calculation of the power and sample size for hypothesis tests comparing the differences between paired data values. Power and sample size calculations are available in most introductory sample size texts. Here we use Ryan (2013).

4.22.2 Methodology

This table provides sample size and power calculations for studies which will be comparing the mean of paired differences to an index mean difference. The sample size calculation for paired samples is taken from Ryan (2013). The test of interest is for the following Z-statistic:

$$Z = \frac{d_1 - d_0}{\sigma_d / \sqrt{N_{Pairs}}} \qquad \{4.22.1\}$$

Where d_1 is the measured or "true" mean of the differences between paired data values under the alternative hypothesis, d_0 is the index or null hypothesis mean of the paired differences (this is assumed to be zero in this table i.e. no difference between paired subjects under the null hypothesis), σ is the known standard deviation of the paired differences and N_{Pairs} is the number of pairs of subjects included in the study. Note that this calculation is effectively the same as from chapter 4.21 for the one mean case with the paired differences effectively considered as a de-facto single sample. Note that σ_d can be estimated using the following equation:

$$\sigma_d = \sqrt{{\sigma_1}^2 + {\sigma_2}^2 - 2\sigma_{12}} = \sqrt{{\sigma_1}^2 + {\sigma_2}^2 - 2\rho_{12}\sigma_1\sigma_1} \qquad \{4.22.2\}$$

Where σ_1 is the standard deviation of the samples from one end of the paired samples (e.g. before treatment on subject), σ_2 is the standard deviation on the other end (e.g. after treatment on subject), σ_{12} is the covariance between the paired observations and ρ_{12} is the correlation between the paired observations.

The Z-test is usually used instead of the more common t-test in two contexts: a) when the population variance is known or b) when the expected sample size is high (where the Z and t-tests converge). The usage of this sample size table as opposed to the t-test table depends on which of these tests will be used by the end user for their statistical analysis.

However, both tables require the assumptions that the data is continuous, the underlying data distribution of the paired differences is normal and that the sample pairs used were randomly selected i.e. every possible participant pair was equally likely to be selected in a sample, be fulfilled. If these assumptions are violated then the usage of these tables is incorrect and the appropriate alternative parametric or non-parametric equivalent should be used.

This table can be used to calculate the power, the sample size, the significance level, the mean of the paired differences and the standard deviation of the paired differences given all other terms in the table are specified.

To calculate power and sample size/number of pairs, the user must specify the test significance level (α), whether to use a one or two sided test, the mean of the paired differences (d_1), a finite population adjustment (if any) and the standard deviation of the paired differences (σ_d). Note that if the finite population adjustment is set to zero or less than the sample size (if specified) then the calculation will be conducted assuming no adjustment for population size (i.e. assuming an infinite population)

The power formula uses the Z-test of the mean of paired differences to yield the following formula assuming that $d_0 = 0$:

$$\begin{cases} 1-\beta = \Phi\left[\dfrac{|d_1|}{\sigma_d/N_{Pairs}} - z_{1-\alpha}\right], & One-sided \\ 1-\beta = \Phi\left[\dfrac{|d_1|}{\sigma_d/N_{Pairs}} - z_{1-\alpha/2}\right] + \Phi\left[-\left(\dfrac{|d_1|}{\sigma_d/N_{Pairs}}\right) - z_{1-\alpha/2}\right], & Two-sided \end{cases} \qquad \{4.22.3\}$$

where Φ is the cumulative Z distribution evaluated at the calculated value and $z_{1-\alpha}$ is the inverse cumulative Z-distribution evaluated at the specified probability. If a finite population adjustment is used then the standard deviation of the paired differences is specified as the following:

$$\sigma_{adj} = \sqrt{\sigma_d^2\left(\frac{1-N_{Pairs}}{N_{Population}}\right)} \qquad \{4.22.4\}$$

Where $N_{Population}$ is the population of pairs specified in the Finite Population Adjustment row.

The other formulae use an approximation in the two-sided case where the second term in the two-sided case is ignored (it can usually be assumed to negligible). It can then be shown through rearrangement that the following equations hold for the following calculable parameters in the two-sided case:

$$N_{Pairs} = \left(\frac{\left(z_{1-\frac{\alpha}{2}} - z_{1-\beta}\right)^2 \sigma_d^2}{(d_1)^2}\right) \qquad \{4.22.5\}$$

$$d_1 = \mp\left(\frac{\sigma_d\left(z_{1-\frac{\alpha}{2}} - z_{1-\beta}\right)}{\sqrt{N_{Pairs}}}\right) \qquad \{4.22.5\}$$

$$\begin{cases} \sigma_d = \dfrac{|d_1|\sqrt{N_{Pairs}}}{\left(z_{1-\frac{\alpha}{2}} - z_{1-\beta}\right)}, & No\ Pop.Adjust. \\ \sigma_d = \sqrt{\dfrac{\left[\dfrac{|d_1|\sqrt{N_{Pairs}}}{\left(z_{1-\frac{\alpha}{2}} - z_{1-\beta}\right)}\right]^2 N_{Population}}{1 - N_{Pairs}}}, & With\ Pop.Adjust \end{cases} \qquad \{4.22.6\}$$

Where $z_{1-\frac{\alpha}{2}}$ becomes $z_{1-\alpha}$ for the one-sided case.

A closed form equation is not used to calculate the significance level. Instead a search algorithm is used. The search algorithm calculates power at various values for the relevant parameter until the desired power is reached.

4.22.3 Examples

Example 1: Validation example calculating required sample size for a given power
The following example is taken from Example 3.5 on page 79 of Ryan (2013) where a sample size calculation problem is conducted. Then we conduct a sensitivity analysis to show the effect of changing the standard deviation of the paired differences has on the sample size.

The following steps outline the procedure for this example:

1. Open nQuery + nTerim 4.0 via the start menu or desktop shortcut. Click “New Fixed Term Test” from the top of the window. The “Study Goal and Design” window will appear as Figure 4.22.1 below.

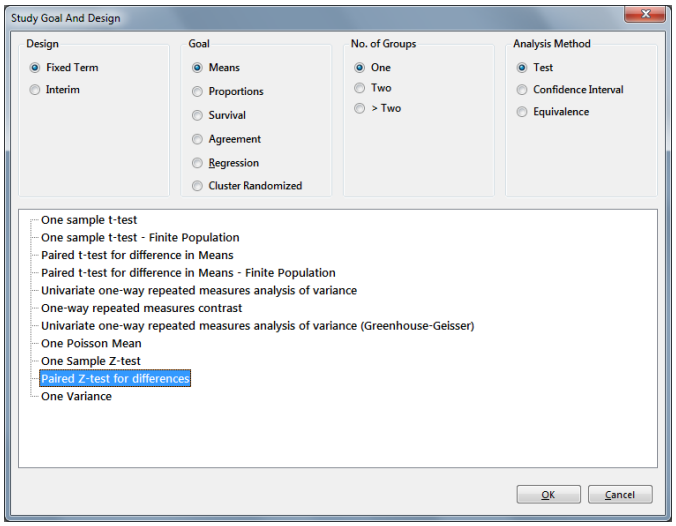


Figure 4.22.1. Study Goal and Design Window

2. In order to select the “Paired Sample Z Test” design table, navigate through the Study Goal and Design Window by selecting "Fixed Term" as the Design, "Means" as the Goal, "One" as the Number of Groups and "Test" as the Analysis Method. Then click "OK" and the test window will appear as per Figure 4.22.2.

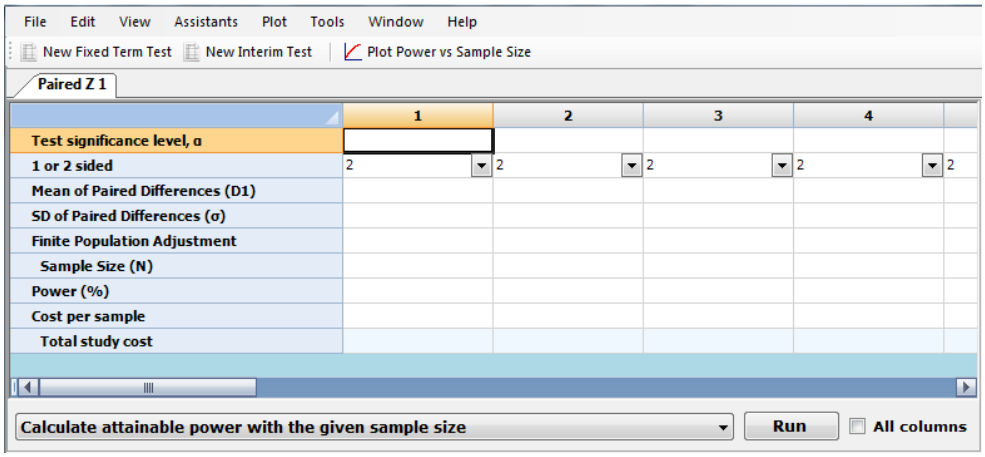


Figure 4.22.2. Paired Sample Z-test Table

The first calculation will be for sample size taken from Ryan. We enter the values as follows:

3. First enter 0.05 for the “Test Significance level” row and 1 for “1 or 2 sided test?”

4. Then enter 1 for the “Mean of Paired Differences” row

5. Enter 3.162 for the “Standard Deviation of Paired Differences” row and 0 in the “Finite Population Adjustment” row (equivalent to setting for infinite population assumption case).

6. Finally, enter 80 in the “Power” row. The table will appear as per Figure 4.22.3.

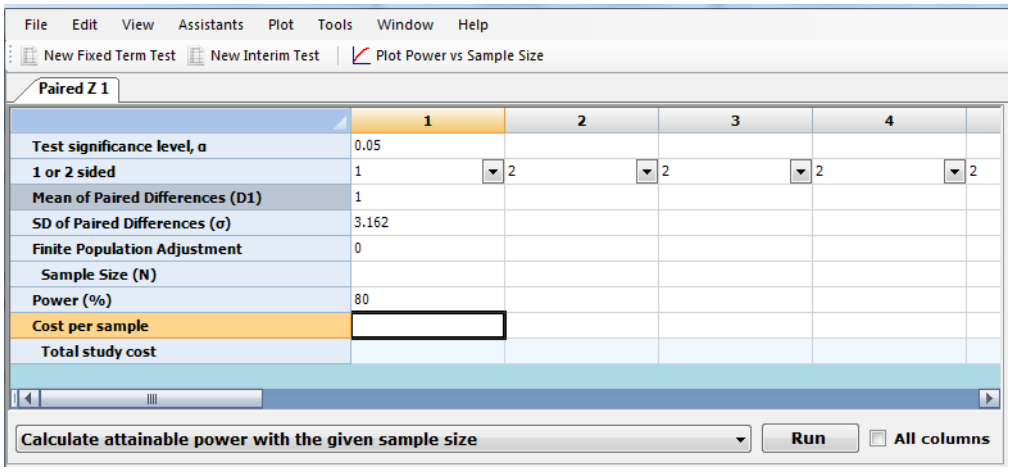
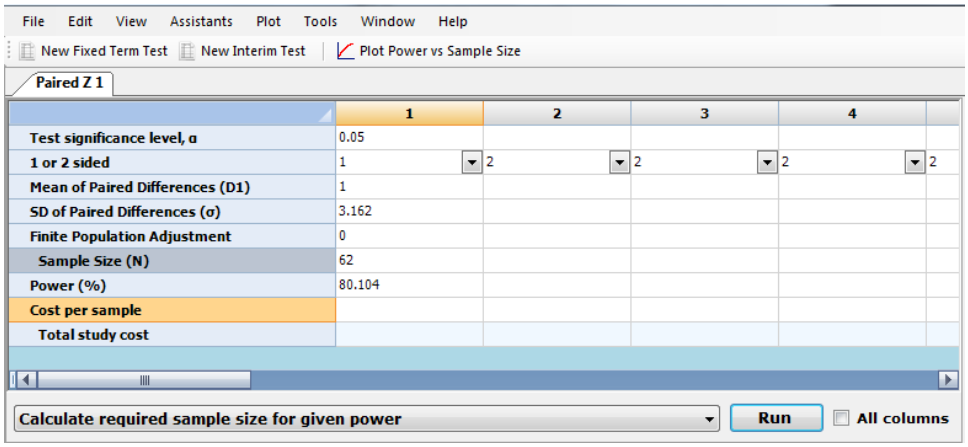


Figure 4.22.3. Values entered for Paired Sample Z-test study design

7. Select “Calculate required sample size for given power” from the dropdown menu beside the “Run” button. Then click “Run”. This will give a result of 62 for sample size

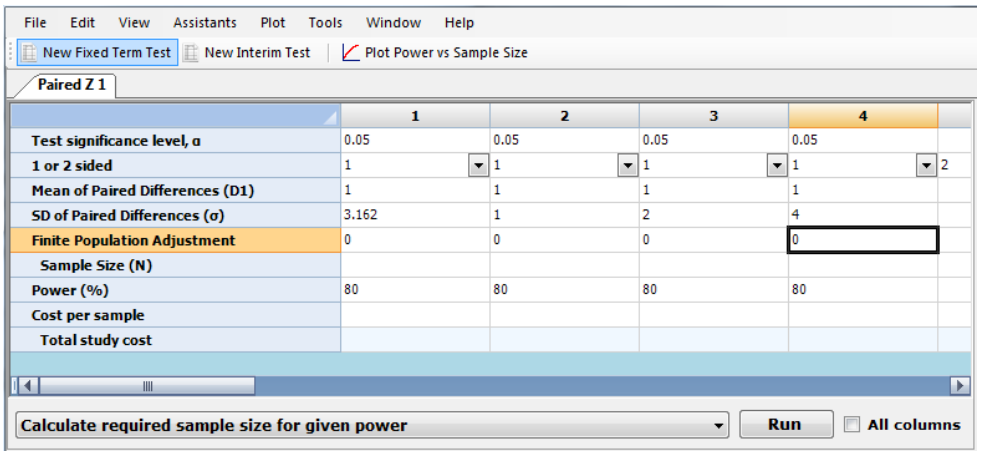
(i.e. number of pairs) as in Figure 4.22.4. This agrees with the value of 62 given in Ryan.



	1	2	3	4	
Test significance level, α	0.05				
1 or 2 sided	1	2	2	2	2
Mean of Paired Differences (D1)	1				
SD of Paired Differences (σ)	3.162				
Finite Population Adjustment	0				
Sample Size (N)	62				
Power (%)	80.104				
Cost per sample					
Total study cost					

Figure 4.22.4. Completed Paired Sample Z-test study design

8. The next calculation is a sensitivity analysis for sample size where we change the standard deviation to investigate the impact this has on the sample size estimate. To do this, copy the same values across to columns 2 to 4 where you delete the values for sample size and replace the values for power with 80. Then, change the values for columns 2 to 4 in the standard deviation row to 1, 2 and 4 respectively. This will give a table as per Figure 4.22.5.

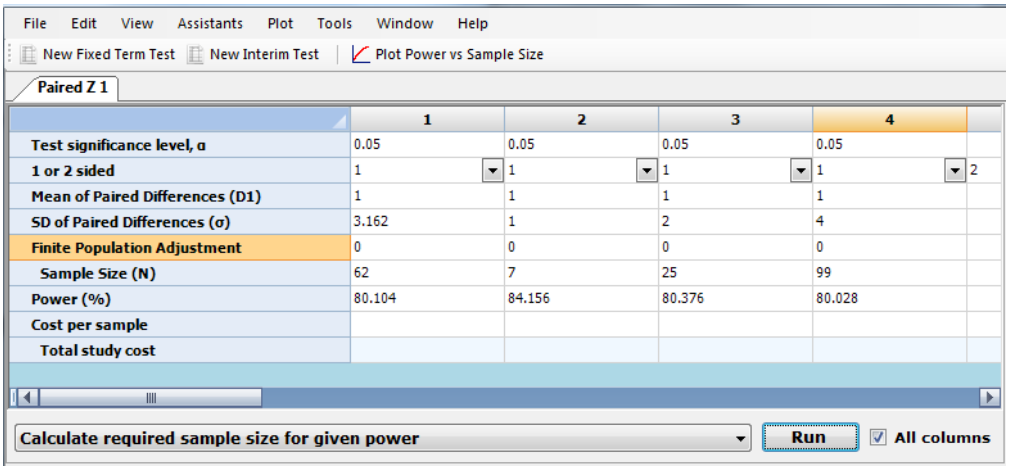


	1	2	3	4	
Test significance level, α	0.05	0.05	0.05	0.05	
1 or 2 sided	1	1	1	1	2
Mean of Paired Differences (D1)	1	1	1	1	
SD of Paired Differences (σ)	3.162	1	2	4	
Finite Population Adjustment	0	0	0	0	
Sample Size (N)					
Power (%)	80	80	80	80	
Cost per sample					
Total study cost					

Figure 4.22.5. Sensitivity analysis around the Standard Deviation

9. Select “Calculate required sample size for given power” from the dropdown menu beside the “Run” button. Then tick the box to run “All Columns”. Then click “Run”.

This will give the resultant sample sizes of 67, 7, 25 and 99 sequentially as in Figure 4.22.6.



	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
1 or 2 sided	1	1	1	1
Mean of Paired Differences (D1)	1	1	1	1
SD of Paired Differences (σ)	3.162	1	2	4
Finite Population Adjustment	0	0	0	0
Sample Size (N)	62	7	25	99
Power (%)	80.104	84.156	80.376	80.028
Cost per sample				
Total study cost				

Figure 4.22.6. Results from Sensitivity analysis

10. Finally, by clicking on the desired study design column and going to the "Output" tab at the bottom of the screen, you can get an output statement for the results given by the current calculation as per Figure 4.22.7 below.

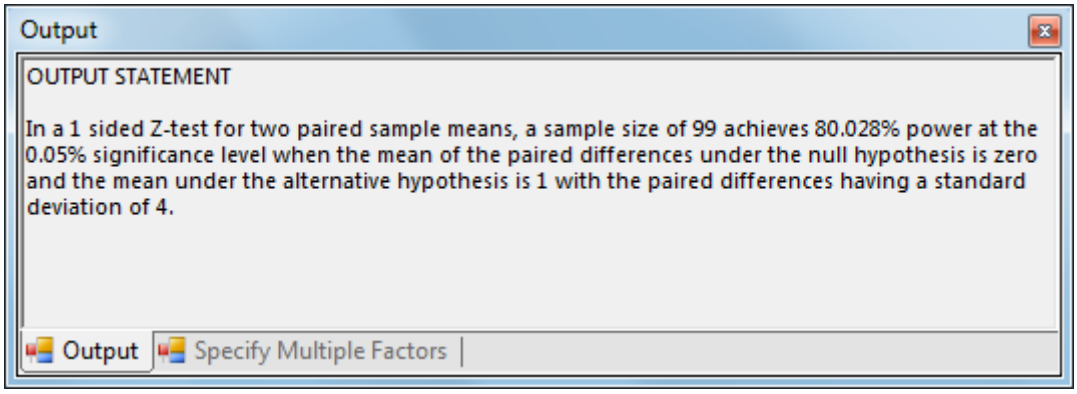


Figure 4.22.7. Study design Output statement

This Output statement can then be easily transferred directly from the output window into a report.

4.23. Two Sample Z-test

4.23.1 Introduction

The Z-test is one of the most commonly used statistical tests in nearly all areas of statistical research. One of its most common and simplest implementations is in the context of the analysis of continuous data which is normally or approximately normally distributed. One such use is the statistical comparison of two independent means and to test whether the difference between them is equal to zero.

This table facilitates the calculation of the power and sample size for hypothesis tests comparing the means of two independent samples using the Z-test. Power and sample size calculations are available in most introductory sample size texts. Here we use Chow et al. (2008).

4.23.2 Methodology

This table provides sample size and power calculations for studies which will be comparing the two independent means using the Z-test. The sample size calculation for independent samples is taken from Chow et al. (2008). The test of interest is for the following Z-statistic:

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{{\sigma_1}^2}{N_1} + \frac{{\sigma_2}^2}{N_2}}} \qquad \{4.23.1\}$$

Where $(\bar{x}_1 - \bar{x}_2) = \delta$ is the measured or "true" difference in means under the alternative hypothesis, $(\mu_1 - \mu_2) = \delta_0$ is the index or null hypothesis difference in means (this is assumed to be zero in this table i.e. no difference between means under the null hypothesis), σ_1 is the known standard deviation of the first group, σ_2 is the known standard deviation of the second group, N_1 is the sample size in the first group and N_2 is the sample size in the second group. Note that in our table's nomenclature that group one becomes the control group and group two becomes the treatment group.

Note that when the two standard deviations are the same that the denominator of equation 4.23.1 becomes the following:

$$SE = \sigma\sqrt{\frac{1}{N_1} + \frac{1}{N_2}} \qquad \{4.23.2\}$$

Where σ is the common standard deviation of the two groups.

The Z-test is usually used instead of the more common t-test in two contexts: a) when the population variance is known or b) when the expected sample size is high (where the Z and t-tests converge). The usage of this sample size table as opposed to the t-test table depends on which of these tests will be used by the end user for their statistical analysis.

However, both tables require the assumptions that the data is continuous, the underlying data distribution of both groups is normal, the variance is known and that both group samples used were randomly selected i.e. every possible participant was equally likely to be selected in each sample, be fulfilled. If these assumptions are violated then the usage of these tables is incorrect and the appropriate alternative parametric or non-parametric equivalent should be used.

This table can be used to calculate the power, the sample size given a sample size ratio, the control group sample size, the treatment group sample size, the significance level, the difference in means, the control group standard deviation and the treatment group standard deviation given all other terms in the table are specified.

To calculate power and sample size, the user must specify the test significance level (α), whether to use a one or two sided test, the difference in means under the alternative hypothesis (δ), the control standard deviation (σ_1) and the treatment standard deviation (σ_2).

The power formula uses the Z-test of the difference to yield the following formulae assuming that $\delta_0 = 0$:

$$\begin{cases} 1-\beta = \Phi\left[\dfrac{|\delta|}{\sqrt{\dfrac{{\sigma_1}^2}{N_1}+\dfrac{{\sigma_2}^2}{N_2}}} - z_{1-\alpha}\right], & One-sided \\ 1-\beta = \Phi\left[\dfrac{|\delta|}{\sqrt{\dfrac{{\sigma_1}^2}{N_1}+\dfrac{{\sigma_2}^2}{N_2}}} - z_{1-\frac{\alpha}{2}}\right] + \Phi\left[-\dfrac{|\delta|}{\sqrt{\dfrac{{\sigma_1}^2}{N_1}+\dfrac{{\sigma_2}^2}{N_2}}} - z_{1-\alpha/2}\right], & Two-sided \end{cases} \qquad \{4.23.3\}$$

where Φ is the cumulative Z distribution evaluated at the calculated value and $z_{1-\alpha}$ is the inverse cumulative Z-distribution evaluated at the specified probability.

The other formulae use an approximation in the two-sided case where the second term in the two-sided case is ignored (it can usually be assumed to negligible). It can then be shown through rearrangement that the following equations hold for the following calculable parameters in the two-sided case:

$$N_2 = \left(\frac{\left(z_{1-\frac{\alpha}{2}} - z_{1-\beta}\right)^2\left(\frac{{\sigma_1}^2}{k} + {\sigma_2}^2\right)}{(\delta)^2}\right), \quad N_1 = kN_2 \qquad \{4.23.4\}$$

$$N_1 = \frac{{\sigma_1}^2}{\left[\frac{\delta^2}{\left(z_{1-\frac{\alpha}{2}} - z_{1-\beta}\right)^2} - \frac{{\sigma_2}^2}{N_2}\right]} \qquad \{4.23.5\}$$

$$N_2 = \frac{{\sigma_2}^2}{\left[\frac{\delta^2}{\left(z_{1-\frac{\alpha}{2}} - z_{1-\beta}\right)^2} - \frac{{\sigma_1}^2}{N_1}\right]} \qquad \{4.23.6\}$$

$$\delta = \mp\left(\left(z_{1-\frac{\alpha}{2}} - z_{1-\beta}\right)\sqrt{\frac{{\sigma_1}^2}{N_1} + \frac{{\sigma_2}^2}{N_2}}\right) \qquad \{4.23.7\}$$

$$\sigma_1 = \sqrt{N_1\left(\frac{\delta^2}{\left(z_{1-\frac{\alpha}{2}} - z_{1-\beta}\right)^2} - \frac{{\sigma_2}^2}{N_2}\right)} \qquad \{4.23.8\}$$

$$\sigma_2 = \sqrt{N_2\left(\frac{\delta^2}{\left(z_{1-\frac{\alpha}{2}} - z_{1-\beta}\right)^2} - \frac{{\sigma_1}^2}{N_1}\right)} \qquad \{4.23.9\}$$

Where k is the sample is size ratio and $z_{1-\frac{\alpha}{2}}$ becomes $z_{1-\alpha}$ for the one-sided case.

A closed form equation is not used to calculate the significance level. Instead a search algorithm is used. The search algorithm calculates power at various values for the relevant parameter until the desired power is reached.

4.23.3 Examples

Example 1: Validation example calculating required sample size for a given power
The following example is taken from Example 3.2.4 on page 64 of Chow et al. (2008) where a sample size calculation problem is conducted. Then we conduct a sensitivity analysis to show the effect of changing the treatment standard deviation has on sample size.

The following steps outline the procedure for this example:

11. Open nQuery + nTerim 4.0 via the start menu or desktop shortcut. Click “New Fixed Term Test” from the top of the window. The “Study Goal and Design” window will appear as Figure 4.23.1 below.

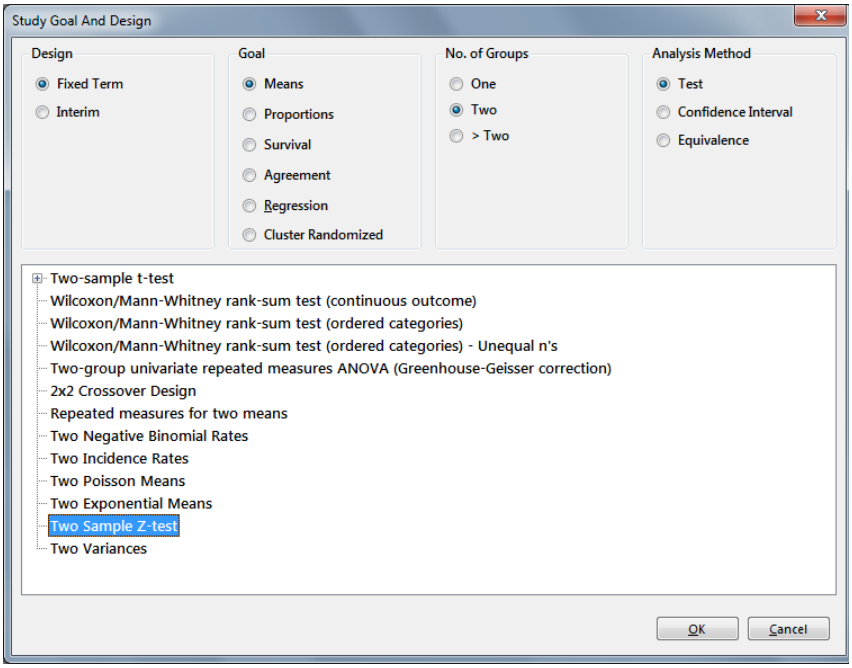


Figure 4.23.1. Study Goal and Design Window

12. In order to select the “Two Sample Z-Test” design table, navigate through the Study Goal and Design Window by selecting "Fixed Term" as the Design, "Means" as the Goal, "Two" as the Number of Groups and "Test" as the Analysis Method. Then click "OK" and the test window will appear as per Figure 4.23.2.

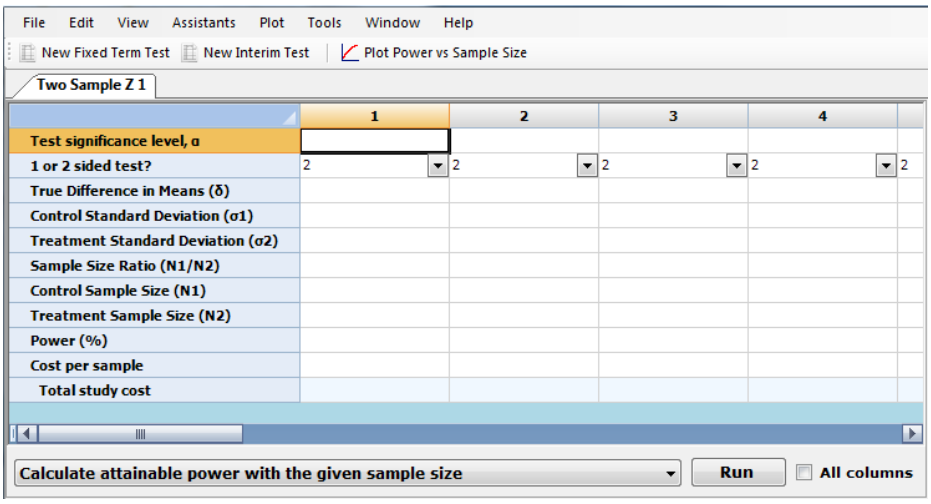


Figure 4.23.2. Two Sample Z-test Table

The first calculation will be for sample size taken from Chow et al. We enter the values as follows:

13. First enter 0.05 for the “Test Significance level” row and 1 for “1 or 2 sided test?”

14. Then enter 0.05 for the “True Difference in Means” row

15. Enter 0.1 for both the “Control Standard Deviation” and “Treatment Standard Deviation” rows and 1 in the “Sample Size Ratio” row (equivalent to setting for a balanced design).

16. Finally, enter 80 in the “Power” row. The table will appear as per Figure 4.23.3.

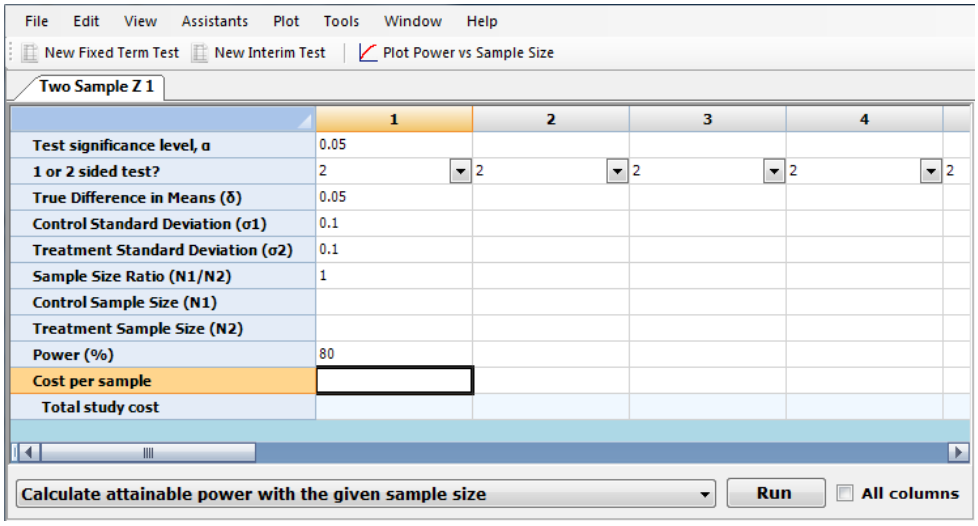


Figure 4.23.3. Values entered for Two Sample Z-test study design

17. Select “Calculate required N1 and N2 for given power and sample size ratio” from the dropdown menu beside the “Run” button. Then click “Run”. This will give a result of 63 for sample size as in Figure 4.23.4. This agrees with the value of 63 given in Chow et al.

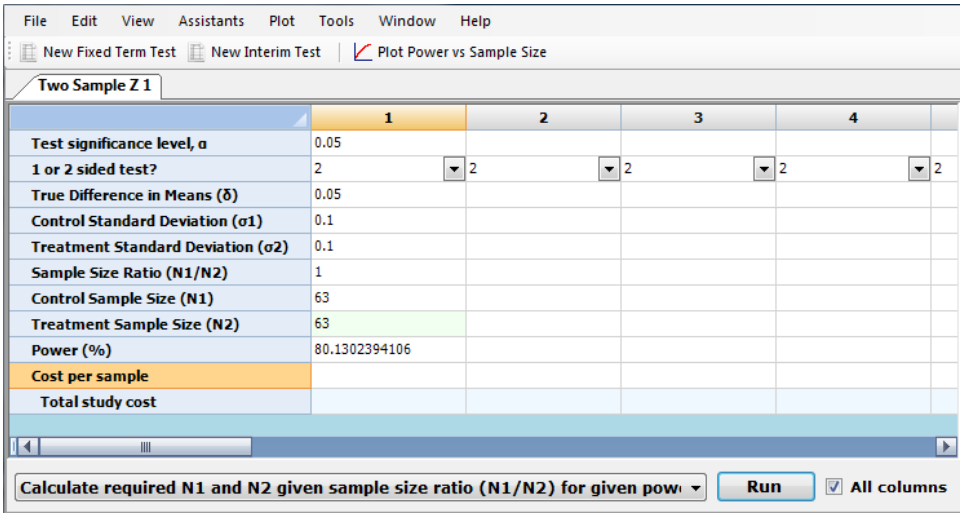


Figure 4.23.4. Completed Two Sample Z-test study design

18. The next calculation is a sensitivity analysis for sample size where we change the treatment standard deviation to investigate the impact this has on the sample size estimate. To do this, copy the same values across to columns 2 to 4 where you delete the values for sample size and replace the values for power with 80. Then, change the values for columns 2 to 4 in the “Treatment Standard Deviation” row to 0.01, 0.5 and 1 respectively. This will give a table as per Figure 4.23.5.

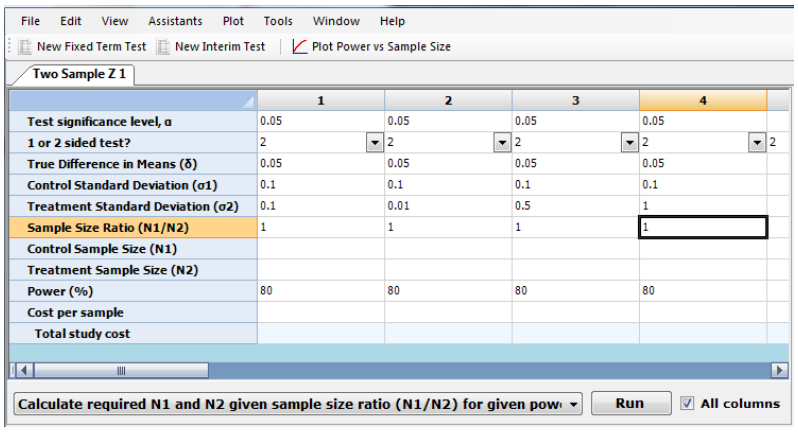


Figure 4.23.5. Sensitivity analysis around the Treatment Standard Deviation

19. Select "Calculate required sample size for given power" from the dropdown menu beside the "Run" button. Then tick the box to run "All Columns". Then click "Run".

This will give the resultant sample sizes of 63, 32, 817 and 3171 sequentially as in Figure 4.23.6.

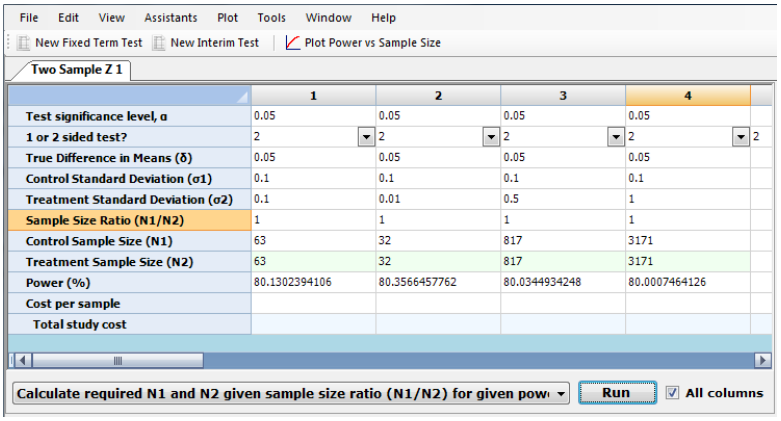


Figure 4.23.6. Results from Sensitivity analysis

20. Finally, by clicking on the desired study design column and going to the "Output" tab at the bottom of the screen, you can get an output statement for the results given by the current calculation as per Figure 4.23.7 below.

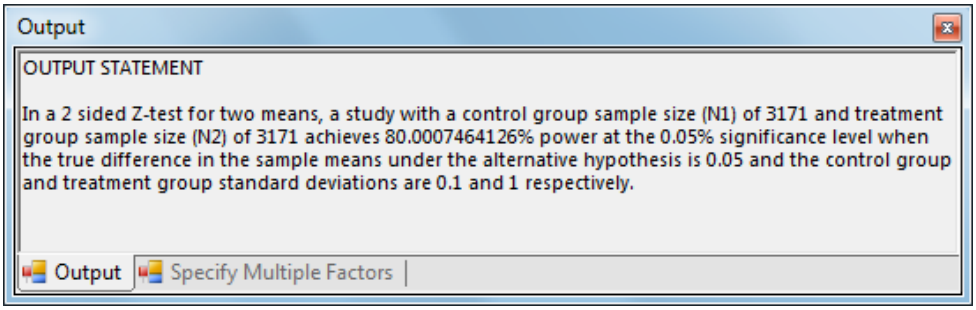


Figure 4.23.7. Study design Output statement

4.24. One Variance

4.24.1 Introduction

The variance is one of the most widely used measures of variability in statistical analysis. Thus researchers are sometimes interested in hypotheses which are related to the variance parameter. One such case is the statistical comparison of a single variance to a pre-specified variance index (or one-sample case) to determine whether the sample variance is equal to the index variance.

This table facilitates the calculation of the power and sample size for hypothesis tests comparing one variance to an index variance. Power and sample size calculations are taken from Ostle (1988).

4.24.2 Methodology

This table provides sample size and power calculations for studies which will be comparing one variance to a variance index. The sample size calculation for one variance is taken from Ostle (1988). The test of interest is for the following chi-squared statistics:

$$\sigma^2 = \frac{\sum_{i=1}^{N}(\mu - x_i)}{N} \qquad \{4.24.1\}$$

$$s^2 = \frac{\sum_{i=1}^{N}(\bar{x} - x_i)}{N-1} \qquad \{4.24.2\}$$

Where $\bar{x}$ is a sample mean, μ is the known population mean, σ is the known standard deviation, s is the sample estimate of the standard deviation, N is the sample size and x_i is the value of the outcome variable for the ith subject. Note that the degrees of freedom of the chi-squared statistic are equal to the denominator in each case. This test assumes that population data of interest is normally distributed and that the sample is randomly drawn.

In this table, we are assuming that the null and alternative variances are the same under the null hypothesis and the difference between them is measured by the ratio of the null variance and alternative variance.

This table can be used to calculate the power, the sample size, the significance level, the null hypothesis variance and the alternative hypothesis variance given all other terms in the table are specified.

To calculate power and sample size, the user must specify the test significance level (α), whether to use a one or two sided test, the null hypothesis variance (V_0), the alternative hypothesis variance (V_1) and whether the true mean is known or not. Note that if the true mean is known them equation 4.24.1 is the basis of the test's power and if it is not known then 4.24.2 is the basis of the test's power.

The power formula uses the chi-squared distribution of the ratio of the two variances to yield the following:

$$\begin{cases} 1-\beta = 1-\chi_{N-1}\left[\frac{s_0^2}{s_1^2}\chi_{N-1}^{-1}(1-\alpha)\right], & \sigma_0^2 < \sigma_1^2, One-sided \\ 1-\beta = \chi_{N-1}\left[\frac{s_0^2}{s_1^2}\chi_{N-1}^{-1}(\alpha)\right], & \sigma_0^2 > \sigma_1^2, One-sided \\ 1-\beta = 1-\chi_{N-1}\left[\frac{s_0^2}{s_1^2}\chi_{N-1}^{-1}(1-\alpha/2)\right] + \chi_{N-1}\left[\frac{s_0^2}{s_1^2}\chi_{N-1}^{-1}(\alpha/2)\right], Two-sided \end{cases} \quad \{4.24.3\}$$

Where χ_{N-1} is the cumulative chi-squared distribution with $N-1$ degrees of freedom evaluated at the calculated value and χ_{N-1}^{-1} is the inverse cumulative chi-squared distribution with $N-1$ degrees of freedom evaluated at the specified probability. If the mean is known, then the degrees of freedom becomes N instead $N-1$.

A closed form equation is not used to calculate the sample size, the significance level, the null hypothesis variance or the alternative hypothesis variance. Instead a search algorithm is used. The search algorithm calculates power at various values for the relevant parameter until the desired power is reached.

4.24.3 Examples

Example 1: Validation example calculating required sample size for a given power
The following example is taken from Appendix 23 of Ostle (1988) where we replicate a sample size calculation table. Then we conduct a sensitivity analysis to show the effect of changing the alternative hypothesis variance on power with fixed sample size.

The following steps outline the procedure for this example:

21. Open nQuery + nTerim 4.0 via the start menu or desktop shortcut. Click “New Fixed Term Test” from the top of the window. The “Study Goal and Design” window will appear as Figure 4.24.1 below.

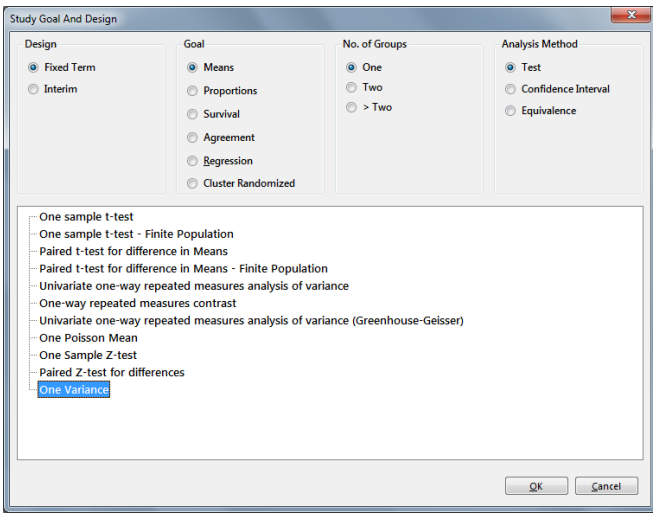


Figure 4.24.1. Study Goal and Design Window

22. In order to select the “One Variance” design table, navigate through the Study Goal and Design Window by selecting "Fixed Term" as the Design, "Means" as the Goal, "One" as the Number of Groups and "Test" as the Analysis Method. Then click "OK" and the test window will appear as per Figure 4.24.2.

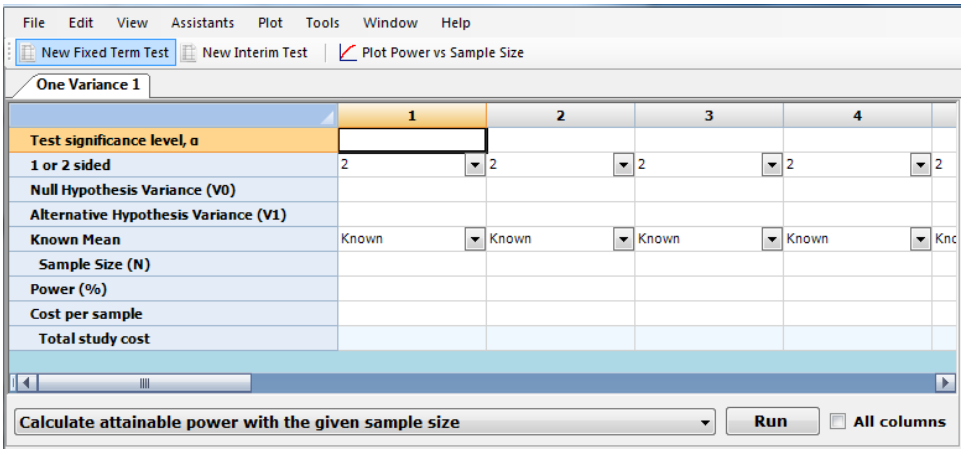


Figure 4.24.2. One Variance Table

The first calculation will be for sample size taken from penultimate ($v = 120$) row of Appendix 23 of Ostle. We enter the values as follows:

23. First enter 0.05 for the “Test Significance level” row and 1 for “1 or 2 sided test?”

24. Then enter 1 and 1.457 for the “Null Hypothesis Variance” and “Alternative Hypothesis Variance” variable rows respectively.

25. Finally, select “Known” for the “Known Mean” row and enter 90 in the “Power” row. The table will appear as per Figure 4.24.3.

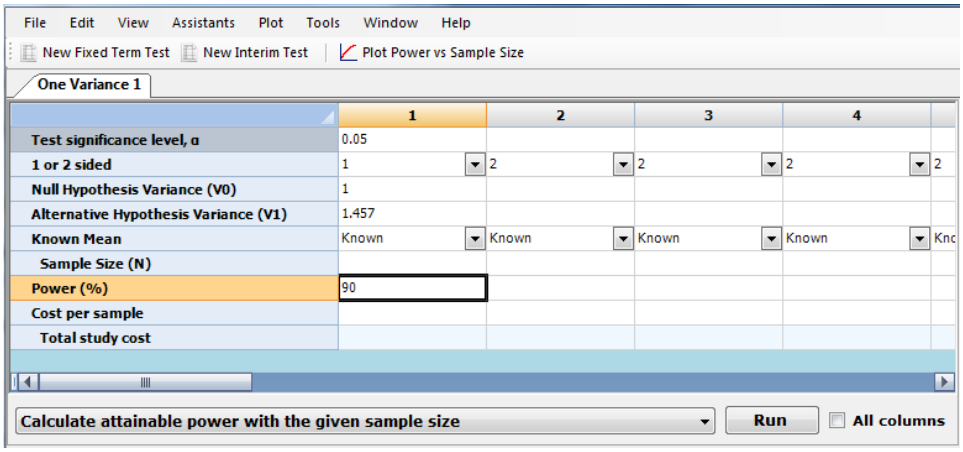
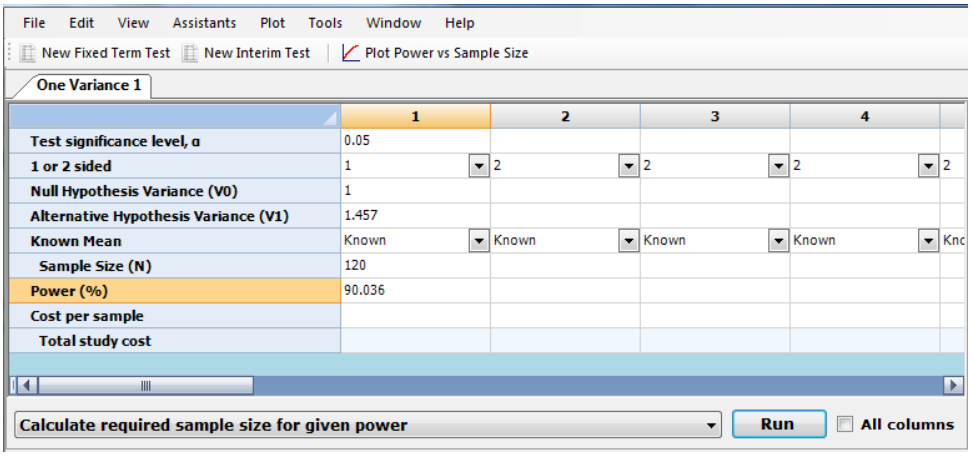


Figure 4.24.3. Values entered for One Variance study design

26. Select “Calculate required sample size for given power” from the dropdown menu beside the “Run” button. Then click “Run”. This will give a result of 120 for sample

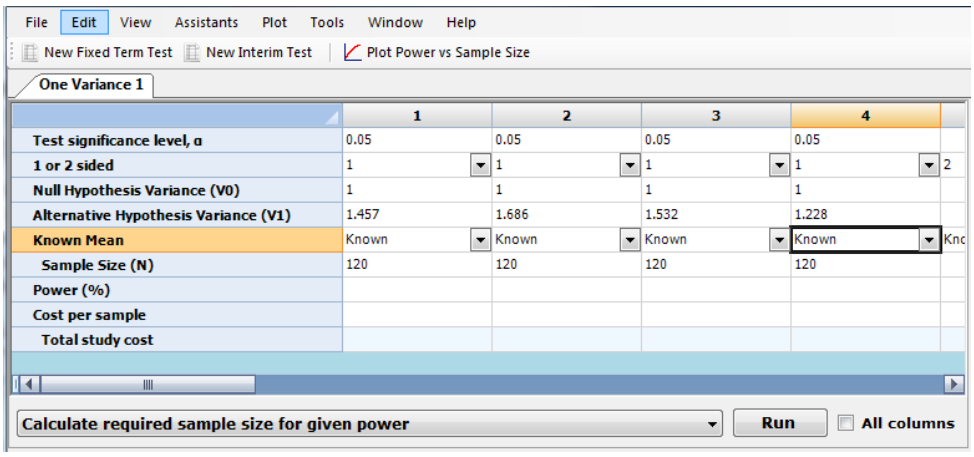
size as in Figure 4.24.4. This agrees with the value of 120 given for the degrees of freedom in Ostle (remember that degrees of freedom = N for Known Mean = "Yes")



	1	2	3	4	
Test significance level, α	0.05				
1 or 2 sided	1	2	2	2	2
Null Hypothesis Variance (V0)	1				
Alternative Hypothesis Variance (V1)	1.457				
Known Mean	Known	Known	Known	Known	Knc
Sample Size (N)	120				
Power (%)	90.036				
Cost per sample					
Total study cost					

Figure 4.24.4. Completed One Variance study design

27. The next calculation is a sensitivity analysis for power where we change the alternative hypothesis variance and fix sample size. To do this, copy the same values across to columns 2 to 4 where you delete the values for power. Then, change the values for columns 2 to 4 in the "Alternative Hypothesis Variance" row to 1.686, 1.532 and 1.228 respectively. This will give a table as per Figure 4.24.5.

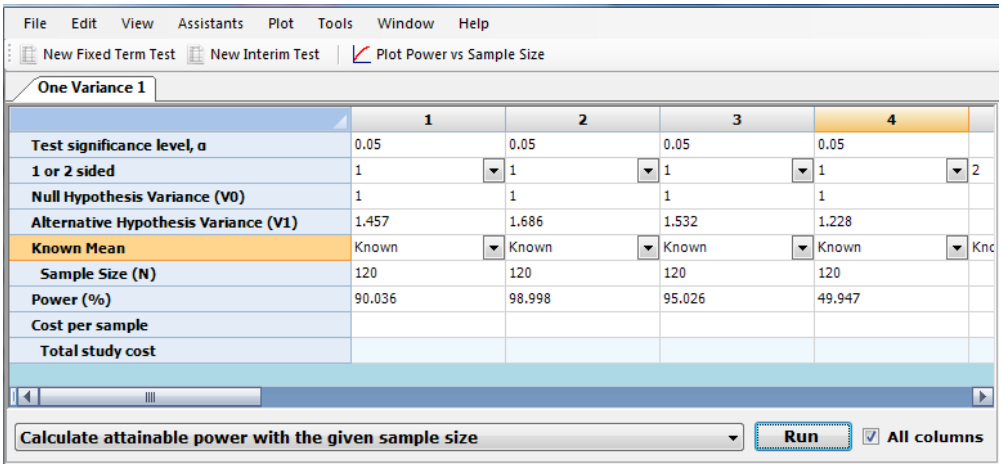


	1	2	3	4	
Test significance level, α	0.05	0.05	0.05	0.05	
1 or 2 sided	1	1	1	1	2
Null Hypothesis Variance (V0)	1	1	1	1	
Alternative Hypothesis Variance (V1)	1.457	1.686	1.532	1.228	
Known Mean	Known	Known	Known	Known	Knc
Sample Size (N)	120	120	120	120	
Power (%)					
Cost per sample					
Total study cost					

Figure 4.24.5. Sensitivity analysis around the Alternative Hypothesis Variance

28. Select "Calculate attainable power for given sample size" from the dropdown menu beside the "Run" button. Then tick the box to run "All Columns". Then click "Run".

This will give the resultant powers of 90.036, 98.988, 95.026 and 49.947 sequentially as in Figure 4.24.6. These correspond closely to given values in Ostle of 90%, 99%, 95% and 50% for these values (Note Ostle uses β instead of 1-β in Appendix 23)



	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
1 or 2 sided	1	1	1	1
Null Hypothesis Variance (V0)	1	1	1	1
Alternative Hypothesis Variance (V1)	1.457	1.686	1.532	1.228
Known Mean	Known	Known	Known	Known
Sample Size (N)	120	120	120	120
Power (%)	90.036	98.998	95.026	49.947
Cost per sample				
Total study cost				

Figure 4.24.6. Results from Sensitivity analysis

29. Finally, by clicking on the desired study design column and going to the "Output" tab at the bottom of the screen, you can get an output statement as per Figure 4.24.7.

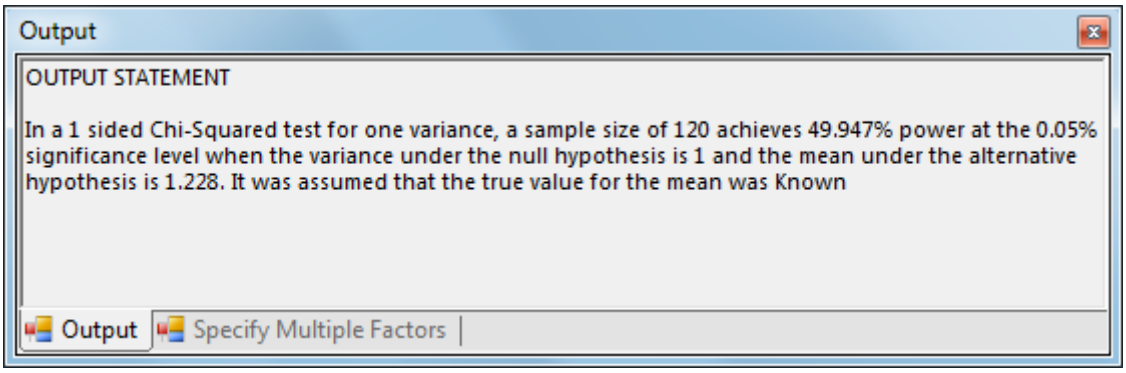


Figure 4.24.7. Study design Output statement

This Output statement can then be easily transferred directly from the output window into a report.

4.25. Two Variances

4.25.1 Introduction

The variance is one of the most widely used measures of variability in statistical analysis. Thus researchers are sometimes interested in hypotheses which are related to the variance parameter. One such case is the statistical comparison of two independent variances to determine whether the one variance is equal to another variance.

This table facilitates the calculation of the power and sample size for hypothesis tests comparing two variances using the F-test. Power and sample size calculations are taken from Ostle (1988).

4.25.2 Methodology

This table provides sample size and power calculations for studies which will be comparing two independent variances. The sample size calculation for two variances is taken from Ostle (1988). The test of interest is based on the following F-statistic:

$$F_{N1-1,N2-1} = \frac{s_1^2}{s_2^2} \qquad \{4.25.1\}$$

$$s_j^2 = \frac{\sum_{i=1}^{N}(\bar{x}_j - x_{ij})}{N_j - 1}, where\ j = 1,2. \qquad \{4.25.2\}$$

Where s_j is the sample standard deviation in group j, $\bar{x}_j$ is the measured sample mean in group j, N_j is the sample size in group j and x_{ij} are the observed values for the ith subject in group j.

Note that the degrees of freedom of the F-statistic are equal to the respective group denominators in each case minus one. This test assumes that both groups are independent, that each group's data is normally distributed and that the sample is randomly drawn from each independent group.

In this table, we are assuming that the group one is equivalent to the control group and that group two is equivalent to the treatment group. We are assuming the two variances are the same under the null hypothesis and the difference between them is measured by the ratio of the control variance and treatment variance.

This table can be used to calculate the power, the sample size, the significance level, the control variance and the treatment variance given all other terms in the table are specified.

To calculate power and sample size, the user must specify the test significance level (α), whether to use a one or two sided test, the control group variance (V_1) and the treatment group variance (V_2). Note that sample size can be calculated based on the sample size ratio, the treatment group sample size given the control group sample size or vice-versa.

The power formula uses the F-distribution of the ratio to yield the following formulae:

$$\begin{cases} 1-\beta = 1 - F_{N1-1,N2-1}\left[\frac{s_0^2}{s_1^2} F^{-1}_{N1-1,N2-1}(1-\alpha)\right], & s_0^2 < s_1^2, One-sided \\ 1-\beta = F_{N1-1,N2-1}\left[\frac{s_0^2}{s_1^2} F^{-1}_{N1-1,N2-1}(\alpha)\right], & s_0^2 > s_1^2, One-sided \\ 1-\beta = 1 - F_{N1-1,N2-1}\left[\frac{s_0^2}{s_1^2} F^{-1}_{N1-1,N2-1}(1-\alpha/2)\right] + F_{N1-1,N2-1}\left[\frac{s_0^2}{s_1^2} F^{-1}_{N1-1,N2-1}(\alpha/2)\right], & Two-sided \end{cases}$$

Where $F_{N1-1,N2-1}$ is the cumulative F-distribution with $N1-1$ and $N2-1$ degrees of freedom evaluated at the calculated value and $F^{-1}_{N1-1,N2-1}$ is the inverse cumulative F-distribution with $N1-1$ and $N2-1$ degrees of freedom evaluated at the specified probability.

A closed form equation is not used to calculate the sample size, the significance level, the control variance or the treatment variance. Instead a search algorithm is used. The search algorithm calculates power at various values for the relevant parameter until the desired power is reached.

4.25.3 Examples

Example 1: Validation example calculating required sample size for a given power
The following example is taken from Appendix 24 of Ostle (1988) where we replicate a sample size calculation table. Then we conduct a sensitivity analysis to show the effect of changing the treatment variance on power with fixed sample size.

The following steps outline the procedure for this example:

1. Open nQuery + nTerim 4.0 via the start menu or desktop shortcut. Click “New Fixed Term Test” from the top of the window. The “Study Goal and Design” window will appear as Figure 4.25.1 below.

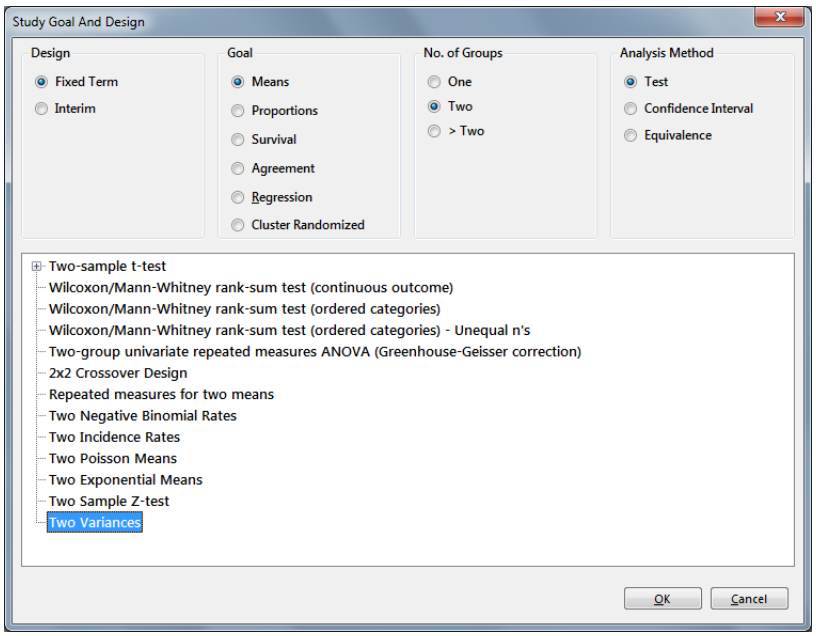


Figure 4.25.1. Study Goal and Design Window

2. In order to select the “Two Variances” design table, navigate through the Study Goal and Design Window by selecting "Fixed Term" as the Design, "Means" as the Goal, "Two" as the Number of Groups and "Test" as the Analysis Method. Then click "OK" and the test window will appear as per Figure 4.25.2.

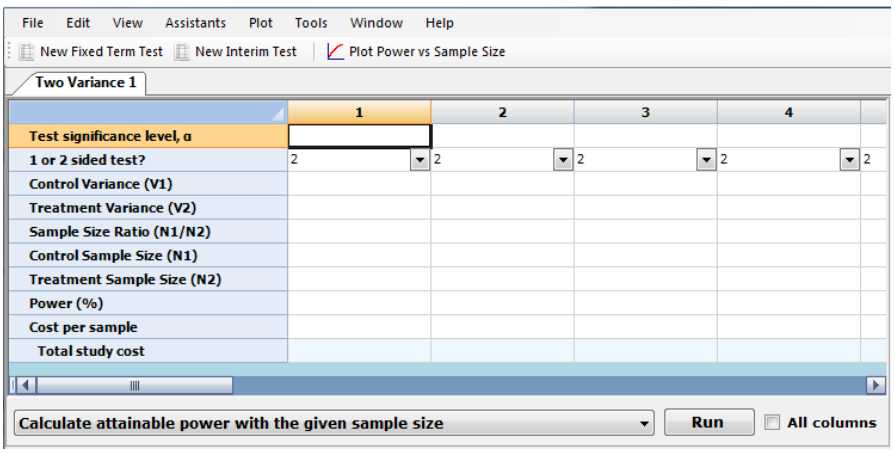


Figure 4.25.2. Two Variances Table

The first calculation will be for a sample size taken from penultimate (v=120) row of Appendix 24 of Ostle. We enter the values as follows:

3. First enter 0.05 for the “Test Significance level” row and 1 for “1 or 2 sided test?”

4. Then enter 1 and 1.71 for the “Control Variance” and “Treatment Variance” variable rows respectively.

5. Finally, enter 1 for the “Sample Size Ratio” row and enter 90 in the “Power” row. The table will appear as per Figure 4.25.3.

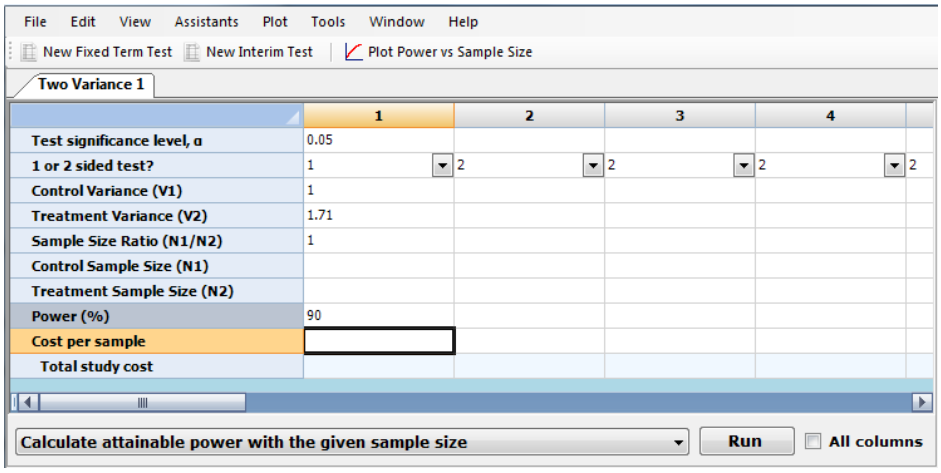


Figure 4.25.3. Values entered for Two Variances study design

6. Select “Calculate required N1 and N2 given sample size ratio (N1/N2) for given power” from the dropdown menu beside the “Run” button. Then click “Run”. This will give a result of 121 for both control and treatment sample size as in Figure 4.25.4. This agrees with the value of 120 given for the degrees of freedom in Ostle (remember that degrees of freedom = $N - 1$ for both groups i.e. $121 - 1 = 120$)

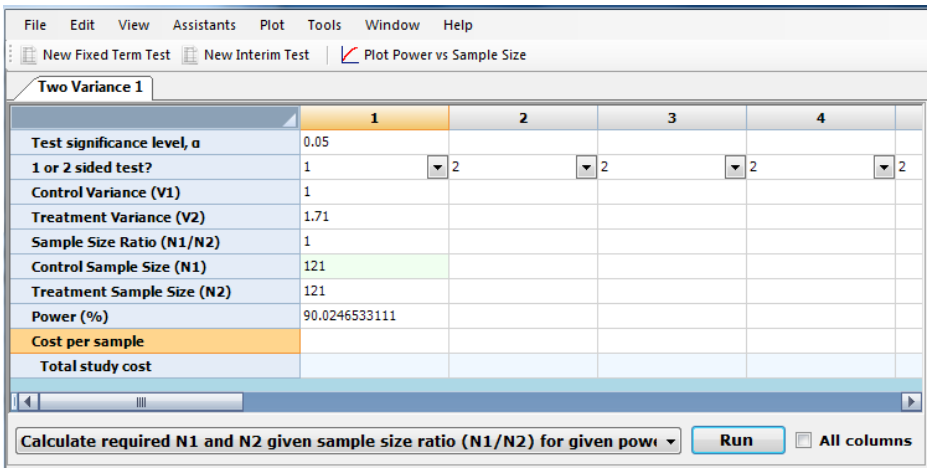
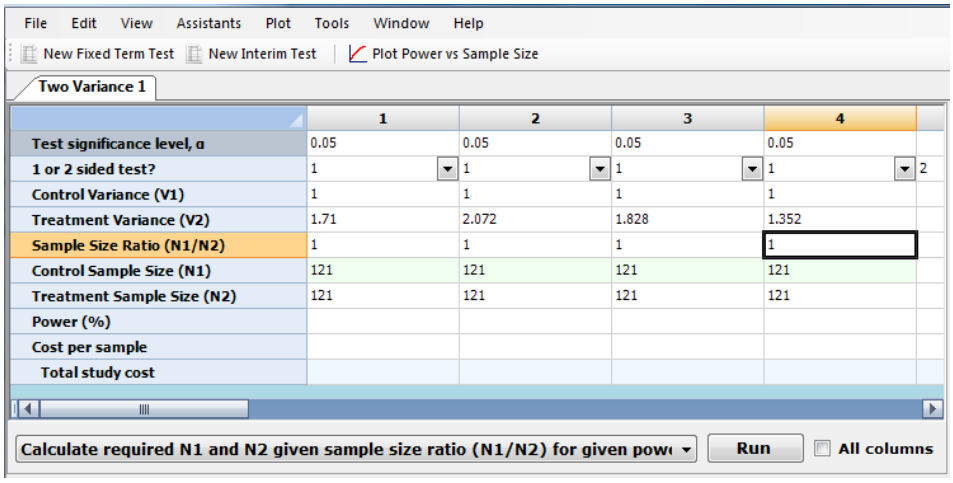


Figure 4.25.4. Completed Two Variances study design

7. The next calculation is a sensitivity analysis for power where we change the treatment variance and fix sample size. To do this, copy the same values across to columns 2 to 4 where you delete the values for power. Then, change the values for columns 2 to 4 in the "Treatment Variance" row to 2.072, 1.828 and 1.352 respectively. This will give a table as per Figure 4.25.5.

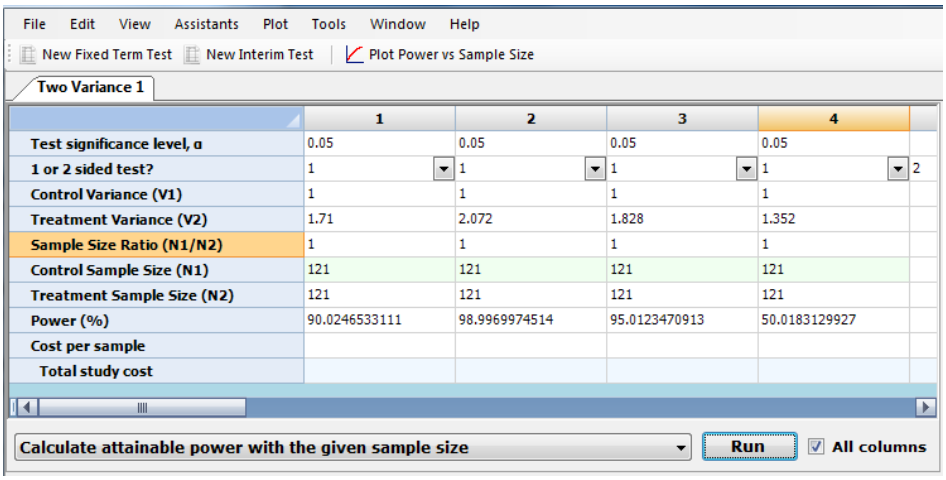


	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
1 or 2 sided test?	1	1	1	1
Control Variance (V1)	1	1	1	1
Treatment Variance (V2)	1.71	2.072	1.828	1.352
Sample Size Ratio (N1/N2)	1	1	1	1
Control Sample Size (N1)	121	121	121	121
Treatment Sample Size (N2)	121	121	121	121
Power (%)				
Cost per sample				
Total study cost				

Figure 4.25.5. Sensitivity analysis around the Treatment Variance

8. Select "Calculate attainable power for given sample size" from the dropdown menu beside the "Run" button. Then tick the box to run "All Columns". Then click "Run".

This will give the resultant powers of 90.025, 98.996, 95.012 and 50.018 sequentially as in Figure 4.25.6. These correspond closely to given values in Ostle of 90%, 99%, 95% and 50% for these values (Note Ostle uses β instead of 1-β in table)



	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
1 or 2 sided test?	1	1	1	1
Control Variance (V1)	1	1	1	1
Treatment Variance (V2)	1.71	2.072	1.828	1.352
Sample Size Ratio (N1/N2)	1	1	1	1
Control Sample Size (N1)	121	121	121	121
Treatment Sample Size (N2)	121	121	121	121
Power (%)	90.0246533111	98.9969974514	95.0123470913	50.0183129927
Cost per sample				
Total study cost				

Figure 4.25.6. Results from Sensitivity analysis

9. Finally, by clicking on the desired study design column and going to the "Output" tab at the bottom of the screen, you can get an output statement as per Figure 4.25.7.

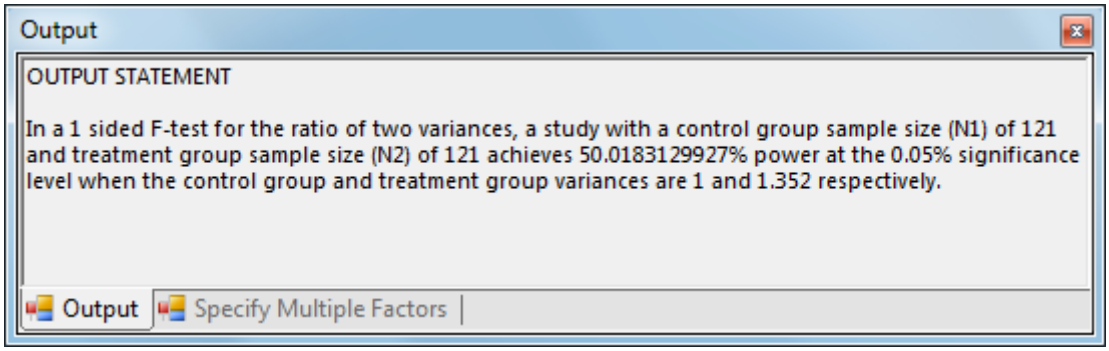


Figure 4.25.7. Study design Output statement

This Output statement can then be easily transferred directly into a report.

4.26. One ROC Curve

4.26.1 Introduction

Receiver Operating Curve (ROC) analysis is an increasingly used tool in medical diagnostics, industrial and data mining applications. The area under the curve (AUC) metric derived from ROC analysis has become a widely cited method to indicate and compare the accuracy of binary classification methods. Thus formalised statistical analysis for the AUC has become more common. One such application is the comparison of a study AUC to a pre-determined index AUC.

This table facilitates the calculation of the power and sample size for hypothesis tests comparing one AUC from an ROC curve to an index AUC. Power and sample size calculations are taken from Obuchowski and McClish (1997) and Hanley and McNeil (1983).

4.26.2 Methodology

This table provides sample size and power calculations for studies which will be comparing one AUC to an AUC index. The sample size calculation for one AUC is derived from the two correlated groups sample size methods outlined in Obuchowski and McClish (1997) and Hanley and McNeil (1983). The former deals with the case where the predictor of interest for classification is ordinal and the latter for when the predictor is continuous. Both of these methods derive from the binormal model of the ROC curve which is outlined below:

$$X \sim N(\mu_-, \sigma_-^2), \qquad Y \sim N(\mu_+, \sigma_+^2), \tag{4.26.1}$$

$$TPR(c) = P(Y > c) = 1 - \Phi\left(\frac{c - \mu_+}{\sigma_+}\right) = \Phi\left(\frac{-c + \mu_+}{\sigma_+}\right) \tag{4.26.2}$$

$$FPR(c) = P(X > c) = 1 - \Phi\left(\frac{c - \mu_-}{\sigma_-}\right) = \Phi\left(\frac{-c + \mu_-}{\sigma_-}\right) \tag{4.26.3}$$

Where X is the distribution of the negative subjects (i.e. without the condition/disease), Y is the distribution of the positive subjects (i.e. with the condition/disease). Both are assumed to be normally distributed with mean μ and standard deviation σ. The true positive rate ($TPR(c)$, also known as the sensitivity) and false positive rate ($FPR(c)$, also known as the specificity) can then be calculated using the normal cumulative distribution (Φ) of the positive and negative distributions respectively at a given classification cut-off point "c". The ROC Curve is then the mapping of the values of $1 - FPR$ versus TPR over increasing values of c. The area of the curve can thus be defined as the following:

$$\theta = \int_{-\infty}^{\infty} TPR(c) FPR'(c) dc \tag{4.26.4}$$

$$\theta = \int_{-\infty}^{\infty} \Phi\left(\frac{-c+\mu_+}{\sigma_+}\right)\Phi\left(\frac{-c+\mu_-}{\sigma_-}\right)\left(\frac{1}{\sigma_-}\right)dc \qquad \{4.26.5\}$$

$$\theta = \int_{-\infty}^{\infty} \Phi(A+Bv)\varphi(v)dv = \Phi\left(\frac{A}{\sqrt{1+B^2}}\right) \qquad \{4.26.6\}$$

Where $\theta = AUC$, $A = \frac{|\mu_+ - \mu_-|}{\sigma_+}$, $B = \frac{\sigma_-}{\sigma_+}$ and $c = \mu_- - v\sigma_-$

A and B can calculated using maximum likelihood methods. Based on this the test statistic for the test becomes the following Z-statistic:

$$Z = \left(\frac{\theta_1 - \theta_0}{V_0(\theta)}\right) \qquad \{4.26.7\}$$

Where $V_0(\theta)$ is the variance of the area under the curve under the null hypothesis.

In this table, we are assuming that the null and alternative hypothesis AUC are the same under the null hypothesis and the difference between them is measured by the difference of the null variance and alternative variance. It is assumed that alternative hypothesis AUC is the estimated or true value expected from your new/prospective diagnostic test.

This table can be used to calculate the power, the positive sample size, negative sample size, the significance level, the null hypothesis AUC and the alternative hypothesis AUC given all other terms in the table are specified.

To calculate power and sample size, the user must specify the test significance level (α), whether to use a one or two sided test, the null hypothesis AUC (AUC_0), the alternative hypothesis AUC (AUC_1), whether the classification predictor is continuous or discrete and the standard deviation ratio ($\frac{SD1}{SD2} i.e. B$). Note that the standard deviation ratio is only used when the predictor is discrete.

The power formula uses the Z-statistic approximation in equation 4.26.7 to derive the following expressions for sample size and power:

$$N_+ = \left(\frac{\left(z_{1-\alpha/2}\sqrt{V(\theta_0)} + z_{1-\beta}\sqrt{V(\theta_1)}\right)^2}{(\theta_1 - \theta_0)^2}\right), N_+ = N_+R \qquad \{4.26.8\}$$

$$1-\beta = \Phi\left(\frac{|\theta_1 - \theta_0| - z_{1-\frac{\alpha}{2}}\sqrt{V(\theta_0)}}{\sqrt{V(\theta_1)}}\right) + \Phi\left(\frac{-|\theta_1 - \theta_0| - z_{1-\frac{\alpha}{2}}\sqrt{V(\theta_0)}}{\sqrt{V(\theta_1)}}\right) \qquad \{4.26.9\}$$

Where Φ is the cumulative z-distribution evaluated at the calculated value, $z_{1-\frac{\alpha}{2}}$ is the inverse cumulative z-distribution evaluated at the specified probability and R is the ratio between the positive and negative sample sizes. In the one-sided case, the second term in 4.26.9 is not included and 1-α is used instead. The variance values are dependent on the whether the classification predictor is discrete or continuous. If the data is discrete the variances are calculated as following from Obuchowski and McClish:

$$V(\theta_i) = f^2\left(1 + \frac{B^2}{R} + \frac{A^2}{2}\right) + g^2B^2\left(\frac{1+R}{2R}\right) \qquad \{4.26.10\}$$

$$f = \left(\frac{E}{\sqrt{2\pi(1+B^2)}}\right) \qquad \{4.26.11\}$$

$$g = \left(\frac{ABE}{\sqrt{2\pi(1+B^2)^3}}\right) \qquad \{4.26.12\}$$

$$E = \exp\left(-\frac{A^2}{2+2B^2}\right) \qquad \{4.26.13\}$$

$$A = \Phi^{-1}(\theta)\sqrt{1+B^2} \qquad \{4.26.14\}$$

Where B is the standard deviation ratio and Φ^{-1} is the inverse normal cumulative distribution. For $V(\theta_0)$, the value used for θ in the calculation of A (equation 4.26.14) is the null hypothesis AUC and for $V(\theta_1)$ the alternative hypothesis AUC is used instead.

When the classification predictor is continuous then the following equation from Hanley and McNeil is used to calculate the variances:

$$V(\theta_i) = \left(\frac{\theta_i}{R(2-\theta_i)} + \frac{2\theta_i^2}{1+\theta_i}\right) - \theta_i^2\left(\frac{1+R}{R}\right), where\ i = 0,1. \qquad \{4.26.15\}$$

A closed form equation is not used to calculate the sample sizes (negative or positive) when the sample size ratio is not given, the significance level, the null hypothesis AUC or the alternative hypothesis AUC. Instead a search algorithm is used. The search algorithm calculates power at various values for the relevant parameter until the desired power is reached.

4.26.3 Examples

Example 1: Validation example calculating required sample size for a given power
The following example is taken from Obuchowski and McClish (1997) where we replicate a sample size calculation. Then we conduct a sensitivity analysis to show the effect of changing the alternative hypothesis AUC and the data type on sample size.

The following steps outline the procedure for this example:

1. Open nQuery + nTerim 4.0 via the start menu or desktop shortcut. Click “New Fixed Term Test” from the top of the window. The “Study Goal and Design” window will appear as Figure 4.26.1 below.

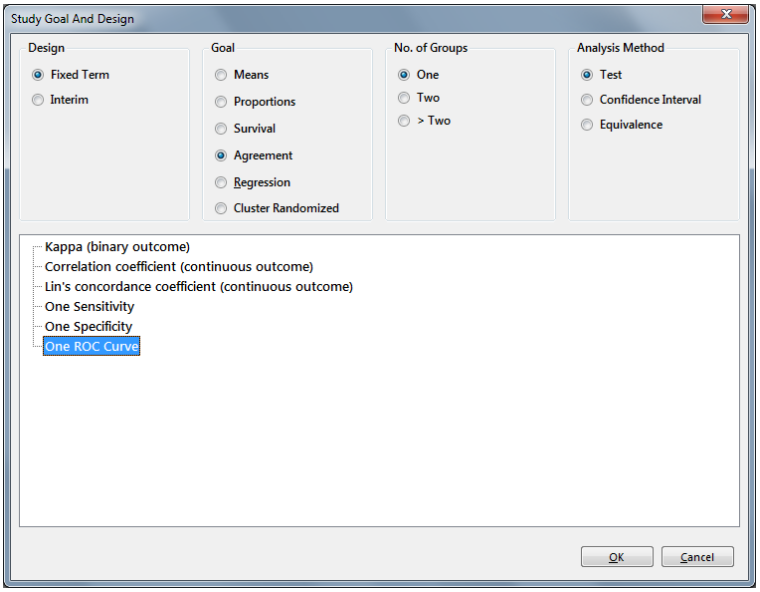


Figure 4.26.1. Study Goal and Design Window

2. In order to select the “One ROC Curve” design table, navigate through the Study Goal and Design Window by selecting "Fixed Term" as the Design, "Agreement" as the Goal, "One" as the Number of Groups and "Test" as the Analysis Method. Then click "OK" and the test window will appear as per Figure 4.26.2.

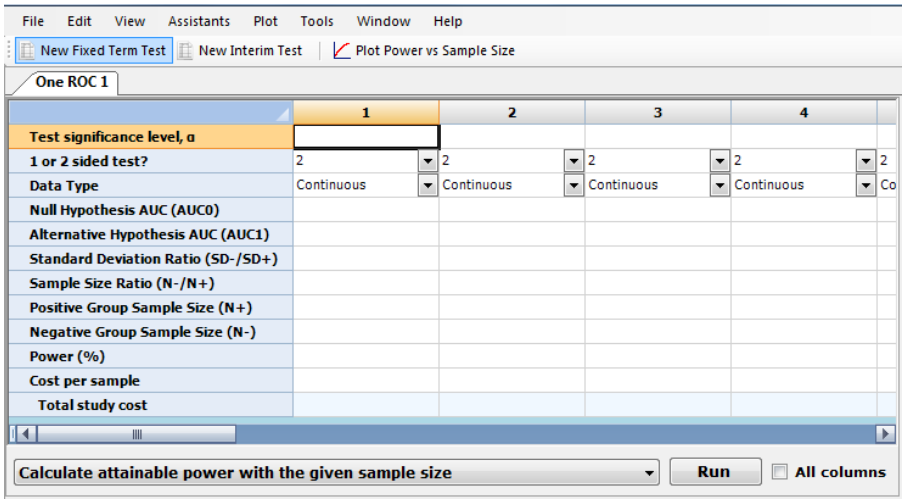


Figure 4.26.2. One ROC Curve Table

The first calculation will be for sample size taken from Obuchowski and McClish. Note that their example is a precision based method. The value of A was calculated as being 1.2. This can be shown to be equivalent to an AUC0 of 0.7995 using equations 4.26.6 or 4.26.14, assuming that the standard deviation ratio is one. Thus the value of AUC1 = AUC0 + 0.10 / 2 = 0.8495, where 0.10 was the maximum length of the confidence interval. This was two-tailed confidence interval with alpha of 0.05. To find the sample size of a confidence interval, the power is set to 50%. Thus we enter the values as follows:

3. First enter 0.05 for the "Test Significance level" row, 2 for "1 or 2 sided test?" and "Discrete" for the "Data Type" row.

4. Then enter 0.7995 and 0.8495 for the "Null Hypothesis AUC" and "Alternative Hypothesis AUC" variable rows respectively. Then enter 1 for the "Standard Deviation Ratio" row.

5. Finally, select 1.5 for the "Sample Size Ratio" row and enter 50 in the "Power" row. The table will appear as per Figure 4.26.3.

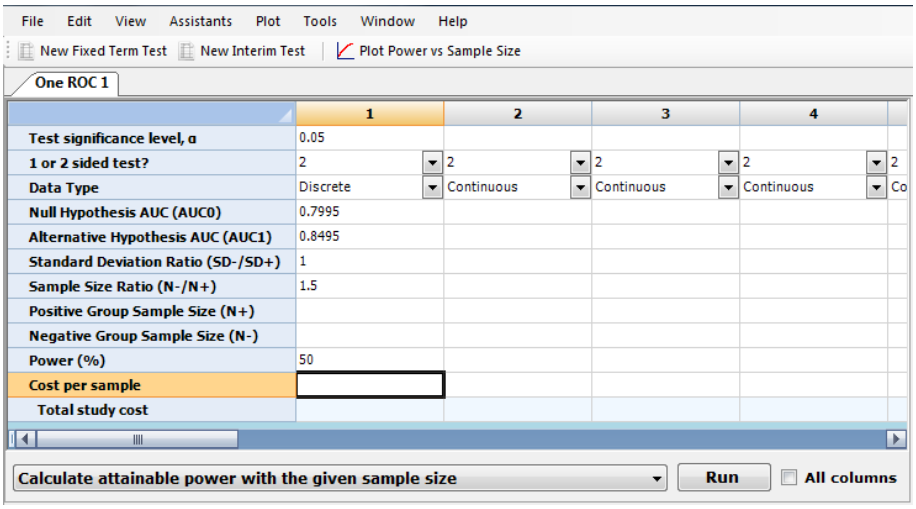
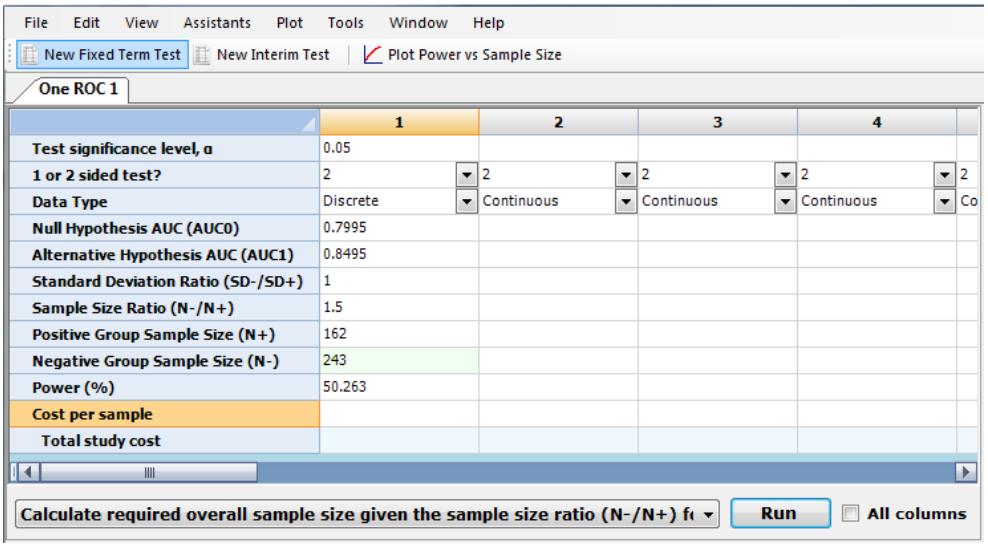


Figure 4.26.3. Values entered for One ROC Curve study design

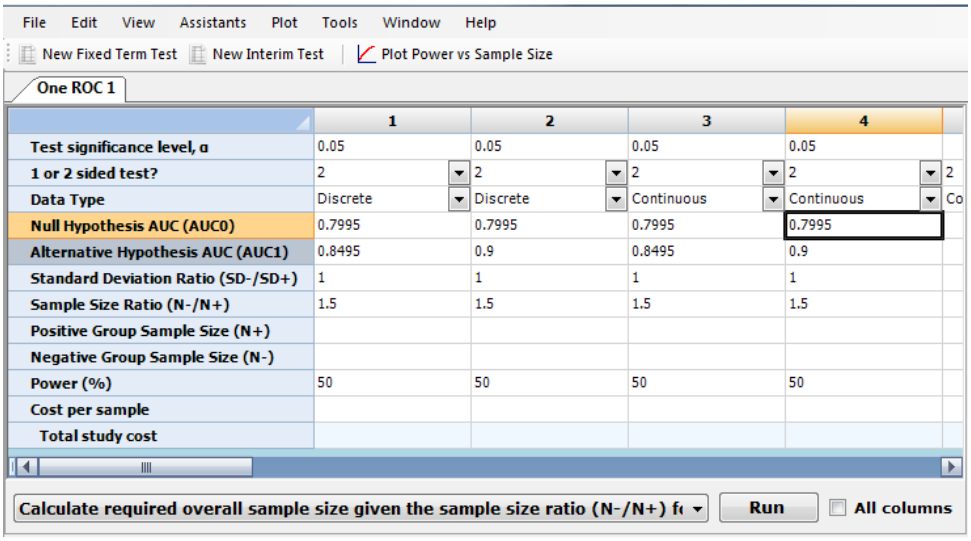
6. Select "Calculate required sample size given the sample size ratio for given power" from the dropdown menu beside the "Run" button. Then click "Run". This will give a result of 162 and 243 for the positive and negative sample sizes as in Figure 4.26.4. The values in Obuchowski and McClish were 161 and 242 but these were rounded up here to ensure the final sample size gave greater than 50% power.



	1	2	3	4	
Test significance level, α	0.05				
1 or 2 sided test?	2	2	2	2	2
Data Type	Discrete	Continuous	Continuous	Continuous	Co
Null Hypothesis AUC (AUC0)	0.7995				
Alternative Hypothesis AUC (AUC1)	0.8495				
Standard Deviation Ratio (SD-/SD+)	1				
Sample Size Ratio (N-/N+)	1.5				
Positive Group Sample Size (N+)	162				
Negative Group Sample Size (N-)	243				
Power (%)	50.263				
Cost per sample					
Total study cost					

Figure 4.26.4. Completed One ROC Curve study design

7. The next calculation is a sensitivity analysis for sample size where we change the alternative hypothesis AUC and change data type. To do this, copy the same values across to columns 2 to 4 where you reset the values for power and delete the sample size values. Then, change the values for columns 2 to 4 in the "Alternative Hypothesis Variance" row to 0.9, 0.8495 and 0.9 respectively. Then change "Data Type" to "Continuous" in columns 3 and 4. This will give a table as per Figure 4.26.5.

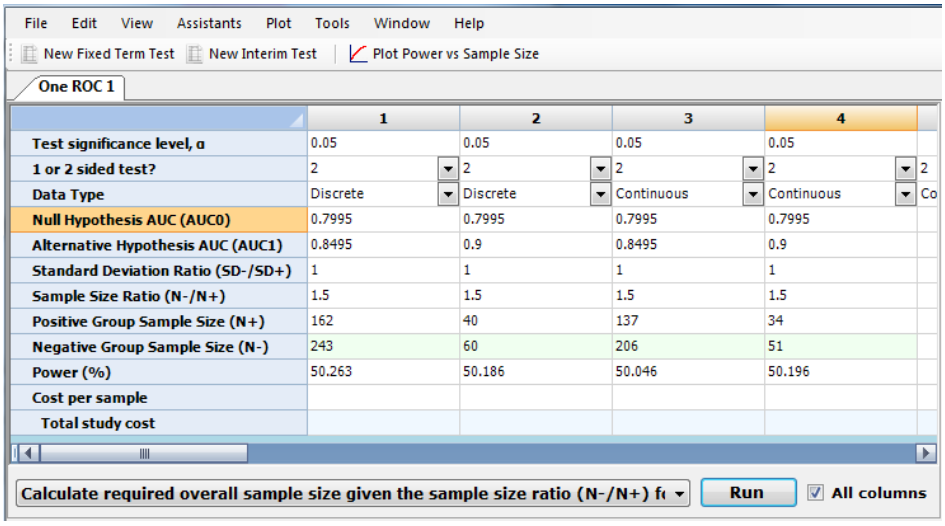


	1	2	3	4	
Test significance level, α	0.05	0.05	0.05	0.05	
1 or 2 sided test?	2	2	2	2	2
Data Type	Discrete	Discrete	Continuous	Continuous	Co
Null Hypothesis AUC (AUC0)	0.7995	0.7995	0.7995	0.7995	
Alternative Hypothesis AUC (AUC1)	0.8495	0.9	0.8495	0.9	
Standard Deviation Ratio (SD-/SD+)	1	1	1	1	
Sample Size Ratio (N-/N+)	1.5	1.5	1.5	1.5	
Positive Group Sample Size (N+)					
Negative Group Sample Size (N-)					
Power (%)	50	50	50	50	
Cost per sample					
Total study cost					

Figure 4.26.5. Sensitivity analysis around the Alternative Hypothesis Variance and Data Type

8. Select "Calculate required sample size given the sample size ratio for given power" from the dropdown menu beside the "Run" button. Then tick the box to run "All Columns". Then click "Run".

This will give the resultant sample sizes of 162/243, 40/60, 137/206 and 34/51 sequentially as in Figure 4.26.6.



	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
1 or 2 sided test?	2	2	2	2
Data Type	Discrete	Discrete	Continuous	Continuous
Null Hypothesis AUC (AUC0)	0.7995	0.7995	0.7995	0.7995
Alternative Hypothesis AUC (AUC1)	0.8495	0.9	0.8495	0.9
Standard Deviation Ratio (SD-/SD+)	1	1	1	1
Sample Size Ratio (N-/N+)	1.5	1.5	1.5	1.5
Positive Group Sample Size (N+)	162	40	137	34
Negative Group Sample Size (N-)	243	60	206	51
Power (%)	50.263	50.186	50.046	50.196
Cost per sample				
Total study cost				

Figure 4.26.6. Results from Sensitivity analysis

9. Finally, by clicking on the desired study design column and going to the "Output" tab at the bottom of the screen, you can get an output statement as per Figure 4.26.7.

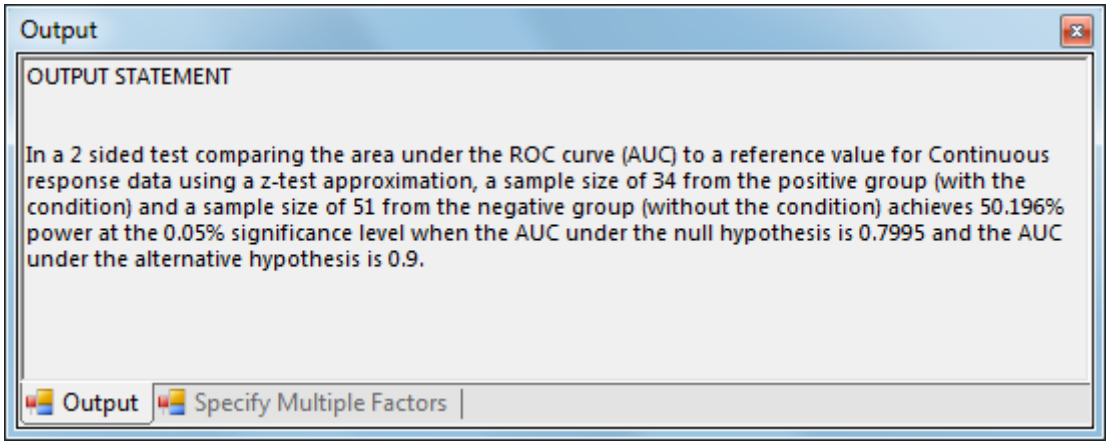


Figure 4.26.7. Study design Output statement

This Output statement can then be easily transferred directly from the output window into a report.

4.27. Two ROC Curves

4.27.1 Introduction

Receiver Operating Curve (ROC) analysis is an increasingly used tool in medical diagnostics, industrial and data mining applications. The area under the curve (AUC) metric derived from ROC analysis has become a widely cited method to indicate and compare the accuracy of binary classification methods. Thus formalised statistical analysis for the AUC has become more common. One such application is the comparison of the two paired AUC curves to determine whether they are unequal.

This table facilitates the calculation of the power and sample size for hypothesis tests comparing two AUC from two paired ROC curves (usually the ROC curves from two diagnostic tests applied to the same patients). Power and sample size calculations are taken from Obuchowski and McClish (1997) and Hanley and McNeil (1983).

4.27.2 Methodology

This table provides sample size and power calculations for studies which will be comparing two paired ROC curves. The sample size calculation for two AUC values is derived from the two sample size methods outlined in Obuchowski and McClish (1997) and Hanley and McNeil (1983). The former deals with the case where the predictor of interest for classification is ordinal and the latter for when the predictor is continuous. Both of these methods derive from the binormal model of the ROC curve outlined below:

$$X \sim N(\mu_-, \sigma_-^2), \qquad Y \sim N(\mu_+, \sigma_+^2), \qquad \{4.27.1\}$$

$$TPR(c) = P(Y > c) = 1 - \Phi\left(\frac{c - \mu_+}{\sigma_+}\right) = \Phi\left(\frac{-c + \mu_+}{\sigma_+}\right) \qquad \{4.27.2\}$$

$$FPR(c) = P(X > c) = 1 - \Phi\left(\frac{c - \mu_-}{\sigma_-}\right) = \Phi\left(\frac{-c + \mu_-}{\sigma_-}\right) \qquad \{4.27.3\}$$

Where X is the distribution of the negative subjects (i.e. without the condition/diseases), Y is the distribution of the positive subjects (i.e. with the condition/diseases). Both are assumed to be normally distributed with mean μ and standard deviation σ. The true positive rate ($TPR(c)$, also known as the sensitivity) and false positive rate ($FPR(c)$, also known as the specificity) can then be calculated using the normal cumulative distribution (Φ) of the positive and negative distributions respectively at a given classification cut-off point "c". The ROC Curve is then the mapping of the value of $1 - FPR$ versus TPR over increasing values of c. The area of the curve can then be defined as the following:

$$\theta = \int_{-\infty}^{\infty} TPR(c) FPR'(c) dc \qquad \{4.27.4\}$$

$$\theta = \int_{-\infty}^{\infty} \Phi\left(\frac{-c+\mu_+}{\sigma_+}\right)\Phi\left(\frac{-c+\mu_-}{\sigma_-}\right)\left(\frac{1}{\sigma_-}\right)dc \qquad \{4.27.5\}$$

$$\theta = \int_{-\infty}^{\infty} \Phi(A+Bv)\varphi(v)dv = \Phi\left(\frac{A}{\sqrt{1+B^2}}\right) \qquad \{4.27.6\}$$

Where $\theta = AUC$, $A = \frac{|\mu_+ - \mu_-|}{\sigma_+}$, $B = \frac{\sigma_-}{\sigma_+}$ and $c = \mu_- - v\sigma_-$

A and B can calculated using maximum likelihood methods. The values for AUC can then be compared. Based on this, the test statistic for the test of the difference between two paired ROC curves becomes the following Z-statistic:

$$Z = \left(\frac{(\theta_2 - \theta_1) - 0}{V_0(\theta_2 - \theta_1)}\right) \qquad \{4.27.7\}$$

Where $V_0(\theta_2 - \theta_1)$ is the variance of the difference in the areas under the curve under the null hypothesis. Note that $\theta_2 - \theta_1$ is referred to as Δ from here onwards.

In this table, we are assuming that the first and second AUC values are the same under the null hypothesis and their difference is that given under the alternative hypothesis. It is assumed in this table that AUC1 is the control area under the curve (i.e. of the standard test) and AUC2 is the treatment AUC (i.e. of the new/prospective test).

This table can be used to calculate the power, the positive sample size, negative sample size, the significance level, the null hypothesis AUC and the alternative hypothesis AUC given all other terms in the table are specified.

To calculate power and sample size, the user must specify the test significance level (α), whether to use a one or two sided test, the reference test AUC (AUC_1), the new test AUC (AUC_2), whether the classification predictor is continuous or discrete and the standard deviation ratio ($\frac{SD1}{SD2} i.e. B_i$) for each test and the correlation between the test results for the positive group and negative group. Note that the standard deviation ratios are only used when the predictor is discrete and that by setting both correlations to zero that you get the equivalent of a test for the two independent samples case (Hanley and McNeil (1982)).

The power formula uses the Z-statistic in equation 4.27.7 to derive the following expressions for sample size and power:

$$N_+ = \left(\frac{\left(z_{1-\alpha/2}\sqrt{V_0(\Delta)} + z_{1-\beta}\sqrt{V_1(\Delta)}\right)^2}{(\Delta)^2}\right), N_+ = N_+R \qquad \{4.27.8\}$$

$$1-\beta=\Phi\left(\frac{|\Delta|-z_{1-\frac{\alpha}{2}}\sqrt{V_0(\Delta)}}{\sqrt{V_1(\Delta)}}\right)+\Phi\left(\frac{-|\Delta|-z_{1-\frac{\alpha}{2}}\sqrt{V_0(\Delta)}}{\sqrt{V_1(\Delta)}}\right) \qquad \{4.27.9\}$$

Where Φ is the cumulative z-distribution evaluated at the calculated value, $z_{1-\frac{\alpha}{2}}$ is the inverse cumulative z-distribution evaluated at the specified probability and R is the ratio between the positive and negative sample sizes. In the one-sided case, the second term in 4.27.9 is not included and 1-α is used instead. The variance values are dependent on the whether the classification predictor is discrete or continuous.

If the data is discrete the variances are calculated as following from Obuchowski and McClish:

$$V_0(\Delta)=V(\theta_1)+V(\theta_1)-2C(\theta_1,\theta_1) \qquad \{4.26.10\}$$

$$V_1(\Delta)=V(\theta_1)+V(\theta_2)-2C(\theta_1,\theta_2) \qquad \{4.27.11\}$$

$$V(\theta_i)=f_i^{\,2}\left(1+\frac{B_i^{\,2}}{R}+\frac{A_i^{\,2}}{2}\right)+g_i^{\,2}B_i^{\,2}\left(\frac{1+R}{2R}\right) \qquad \{4.27.12\}$$

$$C(\theta_i,\theta_j)=f_if_j\left(r_++r_-+\frac{B_iB_j}{R}+\frac{A_iA_j}{2}\right)+g_iB_ig_jB_j\left(\frac{r_-^2+Rr_+^2}{2R}\right)+\frac{f_ig_jA_iB_jr_+^2}{2}+\frac{f_jg_iA_jB_ir_+^2}{2} \qquad \{4.27.13\}$$

$$f_i=\left(\frac{E_i}{\sqrt{2\pi(1+B_i^{\,2})}}\right) \qquad \{4.26.14\}$$

$$g_i=\left(\frac{A_iB_iE_i}{\sqrt{2\pi(1+B_i^{\,2})^3}}\right) \qquad \{4.27.15\}$$

$$E_i=\exp\left(-\frac{A_i^{\,2}}{2+2B_i^{\,2}}\right) \qquad \{4.27.16\}$$

$$A_i=\Phi^{-1}(\theta_i)\sqrt{1+B_i^{\,2}} \qquad \{4.27.17\}$$

Where B_i is the standard deviation ratio in the ith group and Φ^{-1} is the inverse normal cumulative distribution. For $V(\theta_1)$, the value used for A is the control AUC and for $V(\theta_2)$ the new test AUC is used.

When the classification predictor is continuous then the following equation from Hanley and McNeil is used to calculate the variances:

$$V_0(\Delta) = V(\theta_1) + V(\theta_1) - 2C(\theta_1, \theta_1) \qquad \{4.27.18\}$$

$$V_1(\Delta) = V(\theta_1) + V(\theta_2) - 2C(\theta_1, \theta_2) \qquad \{4.27.19\}$$

$$V(\theta_i) = \left(\frac{\theta_i}{R(2-\theta_i)} + \frac{2\theta_i^2}{1+\theta_i}\right) - \theta_i^2\left(\frac{1+R}{R}\right), where\ i = 1,2. \qquad \{4.27.20\}$$

$$C(\theta_i, \theta_j) = 2r\sqrt{V(\theta_i)V(\theta_j)} \qquad \{4.27.13\}$$

Where r is a special value derived from a table in Hanley and McNeil (1983). For values not covered in the table, non-linear extrapolation is used to calculate r based on the table values.

A closed form equation is not used to calculate the sample sizes (negative or positive) when the sample size ratio is not given, the significance level, the reference AUC or the new test AUC. Instead a search algorithm is used. The search algorithm calculates power at various values for the relevant parameter until the desired power is reached.

4.27.3 Examples

Example 1: Validation example calculating required sample size for a given power
The following example is taken from page 34 of Hanley and McNeil (1982) where we replicate power. Then we conduct a sensitivity analysis to show the effect of changing the power and the data type on sample size.

The following steps outline the procedure for this example:

1. Open nQuery + nTerim 4.0 via the start menu or desktop shortcut. Click “New Fixed Term Test” from the top of the window. The “Study Goal and Design” window will appear as Figure 4.27.1 below.

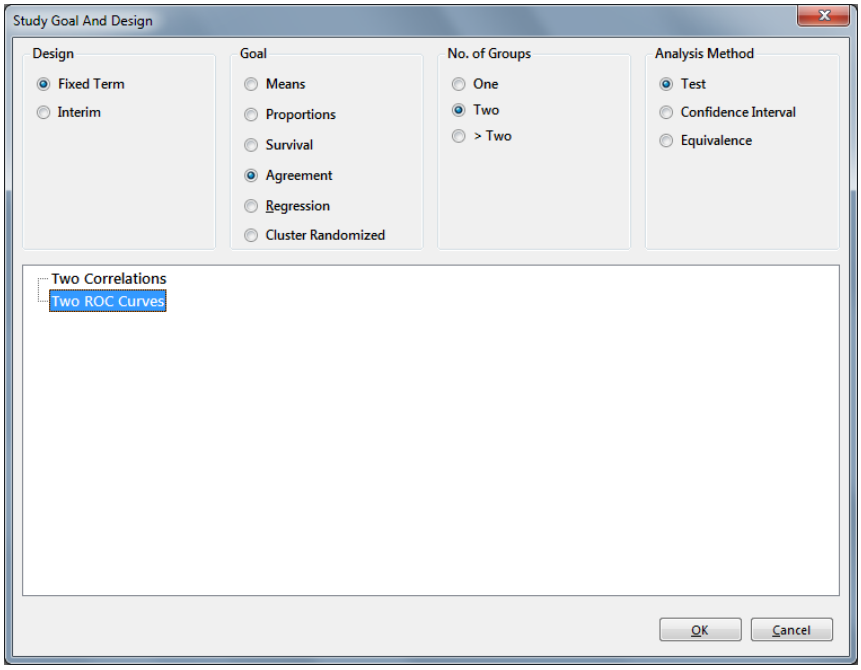


Figure 4.27.1. Study Goal and Design Window

2. In order to select the “Two ROC Curves” design table, navigate through the Study Goal and Design Window by selecting "Fixed Term" as the Design, "Agreement" as the Goal, "Two" as the Number of Groups and "Test" as the Analysis Method. Then click "OK" and the test window will appear as per Figure 4.27.2.

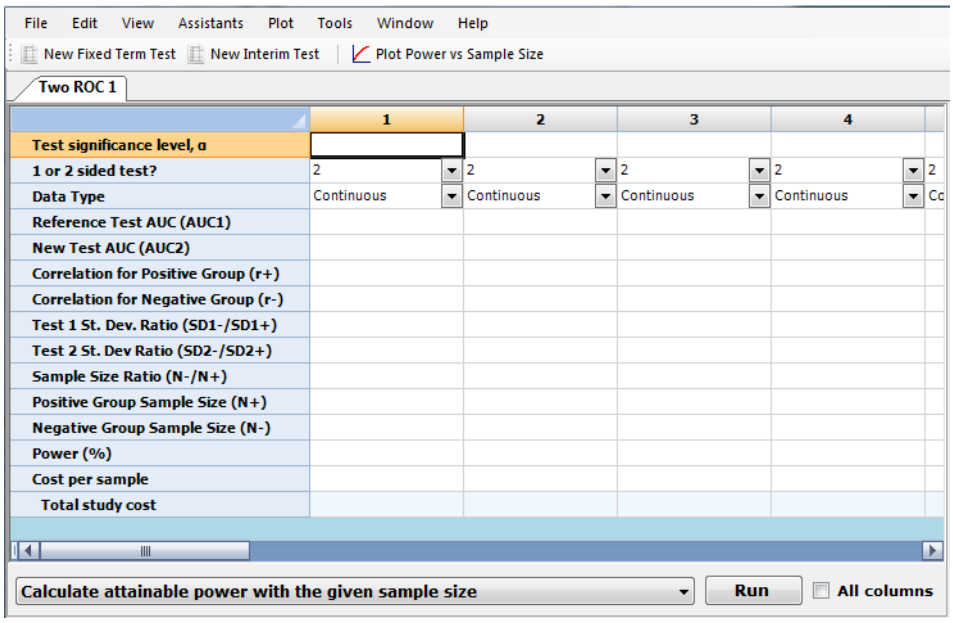


Figure 4.27.2. Two ROC Curve Table

The first calculation will be for sample size taken from Hanley and McNish (1982). Thus we enter the values as follows:

3. First enter 0.05 for the “Test Significance level” row, 1 for “1 or 2 sided test?” and “Continuous” for the “Data Type” row.

4. Then enter 0.7 and 0.75 for the “Null Hypothesis AUC” and “Alternative Hypothesis AUC” variable rows respectively.

5. Set both “Correlation for Positive Group” and “Correlation for Negative Group” to zero.

6. Then enter any value for both “Standard Deviation Ratio” row (not used in continuous data calculations but we will enter 1 for both in this example)

7. Finally, select 1 for the “Sample Size Ratio” row and enter 80 in the “Power” row. The table will appear as per Figure 4.27.3.

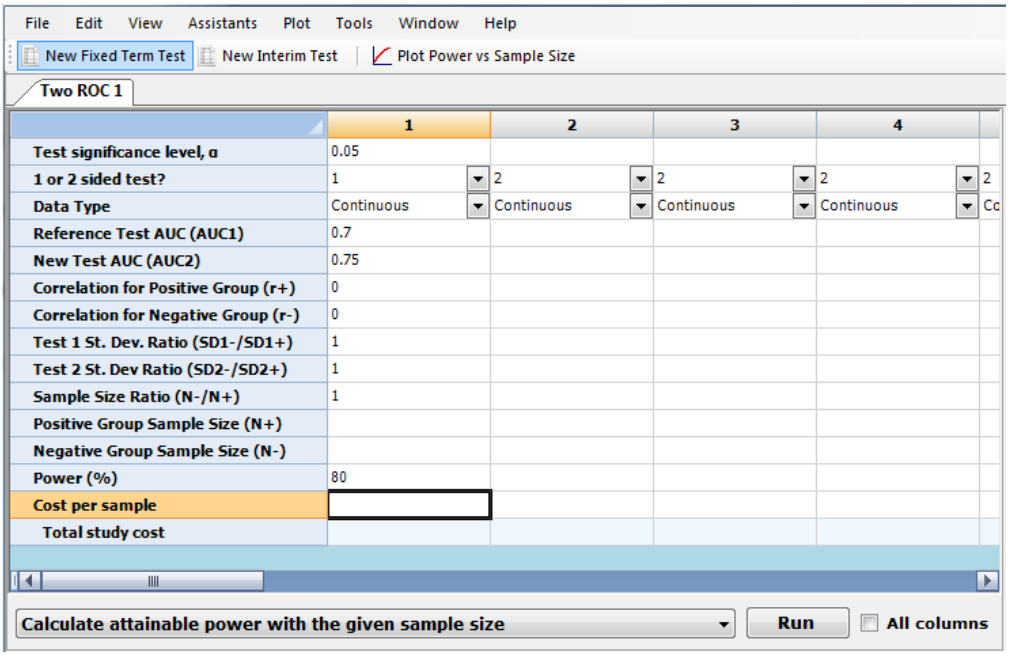
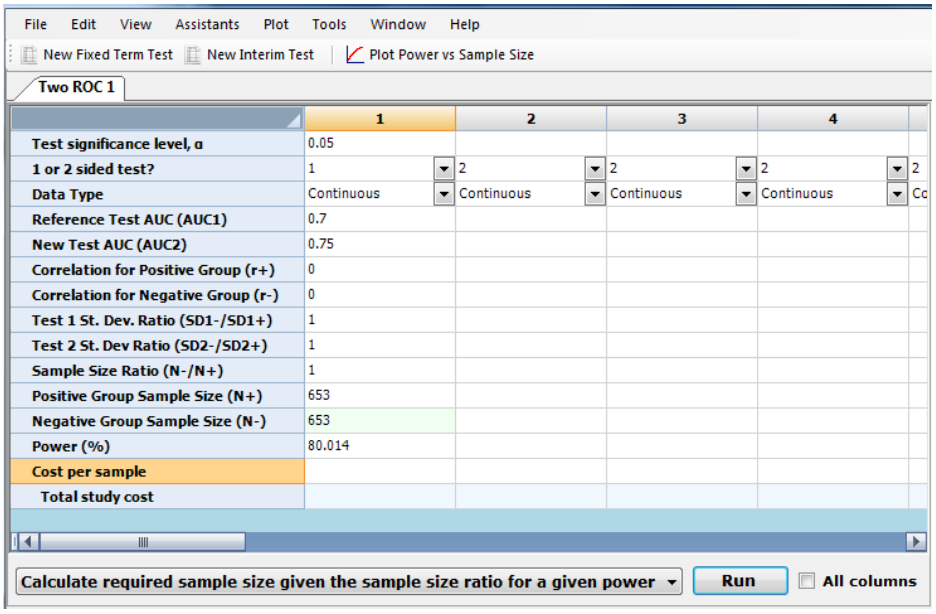


Figure 4.27.3. Values entered for Two ROC Curves study design

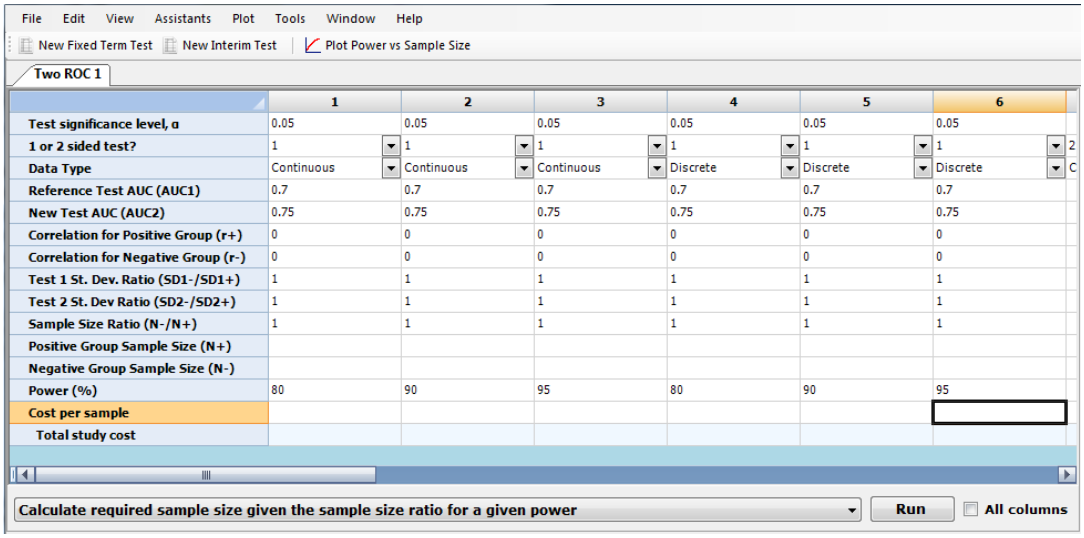
8. Select “Calculate required sample size given the sample size ratio for given power” from the dropdown menu beside the “Run” button. Then click “Run”. This will give a result of 653 for both the positive and negative sample sizes as in Figure 4.27.4. The value in Hanley and McNeil were 652 but this difference is likely a product of the extrapolation for r (see equation 4.27.13) which was not used in this paper (which dealt with independent ROC curves case), less precise estimates of cumulative normal statistics and rounding up to ensure a power greater than 80% rather than simply the value of N which was closest to 80% power.



	1	2	3	4
Test significance level, α	0.05			
1 or 2 sided test?	1	2	2	2
Data Type	Continuous	Continuous	Continuous	Continuous
Reference Test AUC (AUC1)	0.7			
New Test AUC (AUC2)	0.75			
Correlation for Positive Group (r+)	0			
Correlation for Negative Group (r-)	0			
Test 1 St. Dev. Ratio (SD1-/SD1+)	1			
Test 2 St. Dev Ratio (SD2-/SD2+)	1			
Sample Size Ratio (N-/N+)	1			
Positive Group Sample Size (N+)	653			
Negative Group Sample Size (N-)	653			
Power (%)	80.014			
Cost per sample				
Total study cost				

Figure 4.27.4. Completed Two ROC Curves study design

9. The next calculation is a sensitivity analysis for sample size where we change the power (which replicates Hanley and McNeil) and change data type. To do this, copy the same values across to columns 2 to 6 where you delete the sample size values. Then, change the values for columns 2 to 6 in the "Power" row to 0.9, 0.95, 0.8, 0.9 and 0.95 respectively. Then change "Data Type" to "Discrete" in columns 4 and 6. This will give a table as per Figure 4.27.5.



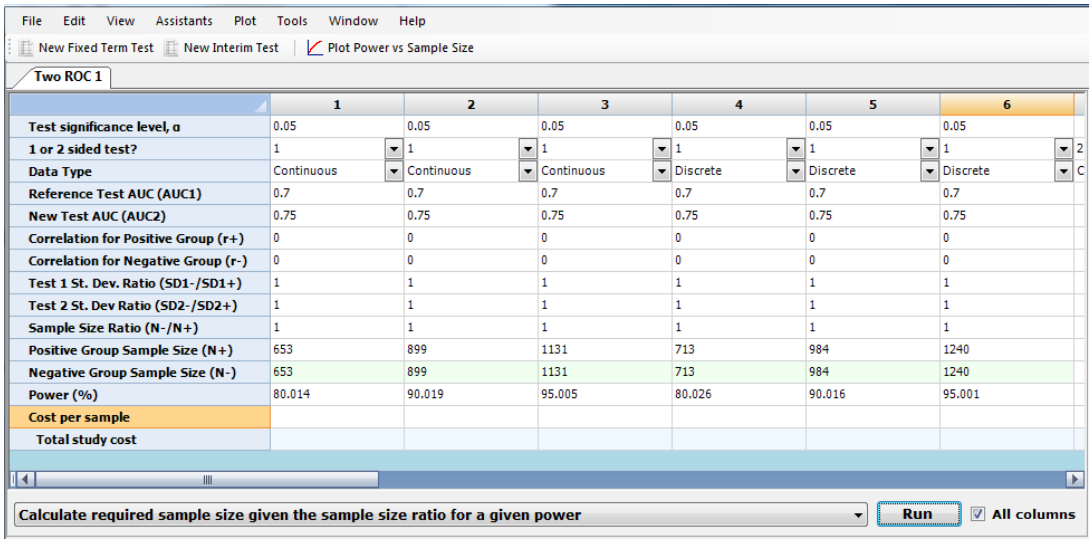
	1	2	3	4	5	6
Test significance level, α	0.05	0.05	0.05	0.05	0.05	0.05
1 or 2 sided test?	1	1	1	1	1	1
Data Type	Continuous	Continuous	Continuous	Discrete	Discrete	Discrete
Reference Test AUC (AUC1)	0.7	0.7	0.7	0.7	0.7	0.7
New Test AUC (AUC2)	0.75	0.75	0.75	0.75	0.75	0.75
Correlation for Positive Group (r+)	0	0	0	0	0	0
Correlation for Negative Group (r-)	0	0	0	0	0	0
Test 1 St. Dev. Ratio (SD1-/SD1+)	1	1	1	1	1	1
Test 2 St. Dev Ratio (SD2-/SD2+)	1	1	1	1	1	1
Sample Size Ratio (N-/N+)	1	1	1	1	1	1
Positive Group Sample Size (N+)						
Negative Group Sample Size (N-)						
Power (%)	80	90	95	80	90	95
Cost per sample						
Total study cost						

Figure 4.27.5. Sensitivity analysis around the Alternative Hypothesis Variance and Data Type

10. Select "Calculate required sample size given the sample size ratio for given power" from the dropdown menu beside the "Run" button. Then tick the box to run "All Columns". Then click "Run".

This will give the resultant sample sizes of 653, 899, 1131, 713, 984 and 1240 sequentially as in Figure 4.27.6. The value in column 2 is two off the value in Hanley and McNeil (N=897)

and the column 3 value is the same as Hanley and McNeil. The potential causes for these differences are discussed above.



	1	2	3	4	5	6
Test significance level, α	0.05	0.05	0.05	0.05	0.05	0.05
1 or 2 sided test?	1	1	1	1	1	1
Data Type	Continuous	Continuous	Continuous	Discrete	Discrete	Discrete
Reference Test AUC (AUC1)	0.7	0.7	0.7	0.7	0.7	0.7
New Test AUC (AUC2)	0.75	0.75	0.75	0.75	0.75	0.75
Correlation for Positive Group (r+)	0	0	0	0	0	0
Correlation for Negative Group (r-)	0	0	0	0	0	0
Test 1 St. Dev. Ratio (SD1-/SD1+)	1	1	1	1	1	1
Test 2 St. Dev Ratio (SD2-/SD2+)	1	1	1	1	1	1
Sample Size Ratio (N-/N+)	1	1	1	1	1	1
Positive Group Sample Size (N+)	653	899	1131	713	984	1240
Negative Group Sample Size (N-)	653	899	1131	713	984	1240
Power (%)	80.014	90.019	95.005	80.026	90.016	95.001
Cost per sample						
Total study cost						

Figure 4.27.6. Results from Sensitivity analysis

11. Finally, by clicking on the desired study design column and going to the "Output" tab at the bottom of the screen, you can get an output statement as per Figure 4.27.7.

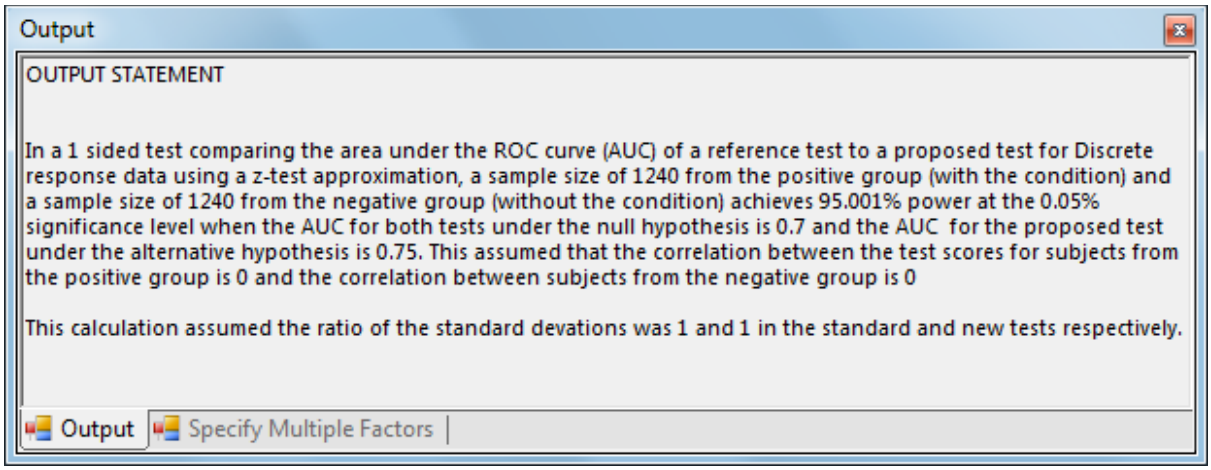


Figure 4.27.7. Study design Output statement

This Output statement can then be easily transferred directly from the output window into a report.

4.28. Two Correlations

4.28.1 Introduction

The correlation coefficient is one of the most widely used methods for the measuring the strength of association between two variables. For example, in regression analysis the correlation coefficient quantifies the slope of the relationship between two standardized variables and is thus an integral part of regression analysis. One statistical enquiry of interest is the comparison of two independent correlations and to test whether they are equal.

This table facilitates the calculation of the power and sample size for hypothesis tests comparing the two independent correlations. Power and sample size calculations are from Zar (2010).

4.28.2 Methodology

This table provides sample size and power calculations for studies which will be comparing two independent correlations. The sample size calculation for independent samples is taken from Zar (2010). The test of interest uses the following Z-statistic approximation for the correlation to conduct power analysis:

$$Z_i = \frac{1}{2}\ln\left(\frac{1+\rho_i}{1-\rho_i}\right), where\ i = 1,2. \qquad \{4.28.1\}$$

Where Z_i is the Z-statistic approximation for the correlation in group i and ρ_i is the true correlation coefficient in group i. Once these values are converted into Z-values then the sample size calculation becomes a version of the standard two-sample Z-test analysis. In this case, the standardised standard deviation is given by the following equation:

$$\sigma_{Z1-Z2} = \sqrt{\frac{1}{N_1-3} + \frac{1}{N_2-3}} \qquad \{4.28.2\}$$

Thus the Z-statistic for testing the differences between two correlations is given by the following equation:

$$Z_\Delta = \frac{(z_1 - z_2) - (Z_1 - Z_2)}{\sigma_{Z1-Z2}} \qquad \{4.28.3\}$$

Where z_i are the sample estimates of the z-statistic in 4.28.1 using the sample estimates of the correlation coefficients and $(Z_1 - Z_2)$ is the null hypothesis expectation of the differences in correlations, which is assumed to zero in this table. In the nomenclature of this table, group one will be referred to as the control group and group two as the treatment group.

This table can be used to calculate the power, the sample size given a sample size ratio, the control group sample size, the treatment group sample size, the significance level, the control group correlation and the treatment group correlation given all other terms in the table are specified.

To calculate power and sample size, the user must specify the test significance level (α), whether to use a one or two sided test, the control correlation (ρ_1) and the treatment correlation (ρ_2).

The power formula uses the Z-test of the difference in correlations to yield the following assuming that there is no difference under the null hypothesis:

$$\begin{cases} 1-\beta = \Phi\left[\frac{(Z_1 - Z_2)}{\sigma_{Z1-Z2}} - z_{1-\alpha}\right], & \rho_1 < \rho_2 \; One-sided \\ 1-\beta = \Phi\left[-\left(\frac{(Z_1 - Z_2)}{\sigma_{Z1-Z2}}\right) - z_{1-\alpha}\right], & \rho_1 > \rho_2 \; One-sided \\ 1-\beta = \Phi\left[\frac{(Z_1 - Z_2)}{\sigma_{Z1-Z2}} - z_{1-\frac{\alpha}{2}}\right] + \Phi\left[-\left(\frac{(Z_1 - Z_2)}{\sigma_{Z1-Z2}}\right) - z_{1-\alpha/2}\right], Two-sided \end{cases} \quad \{4.28.4\}$$

where Φ is the cumulative Z distribution evaluated at the calculated value and $z_{1-\alpha}$ is the inverse cumulative Z-distribution evaluated at the specified probability.

A closed form equation is not used to calculate the sample size, the significance level, the control correlation or the treatment correlation. Instead a search algorithm is used. The search algorithm calculates power at various values for the relevant parameter until the desired power is reached.

4.28.3 Examples

Example 1: Validation example calculating required sample size for a given power
The following example is taken from Example 19.7 on page 392 of Zar (2010) where a power calculation problem is conducted. Then we conduct a sensitivity analysis to show the effect of changing the treatment correlation has on power.

The following steps outline the procedure for this example:

1. Open nQuery + nTerim 4.0 via the start menu or desktop shortcut. Click “New Fixed Term Test” from the top of the window. The “Study Goal and Design” window will appear as Figure 4.28.1 below.

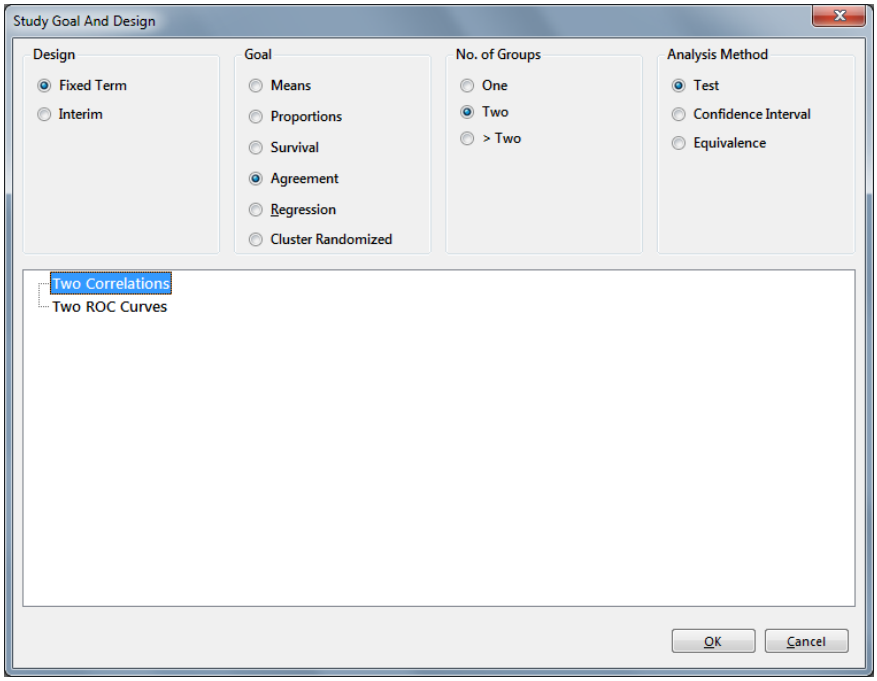


Figure 4.28.1. Study Goal and Design Window

2. In order to select the “Two Correlations” design table, navigate through the Study Goal and Design Window by selecting "Fixed Term" as the Design, "Agreement" as the Goal, "Two" as the Number of Groups and "Test" as the Analysis Method. Then click "OK" and the test window will appear as per Figure 4.28.2.

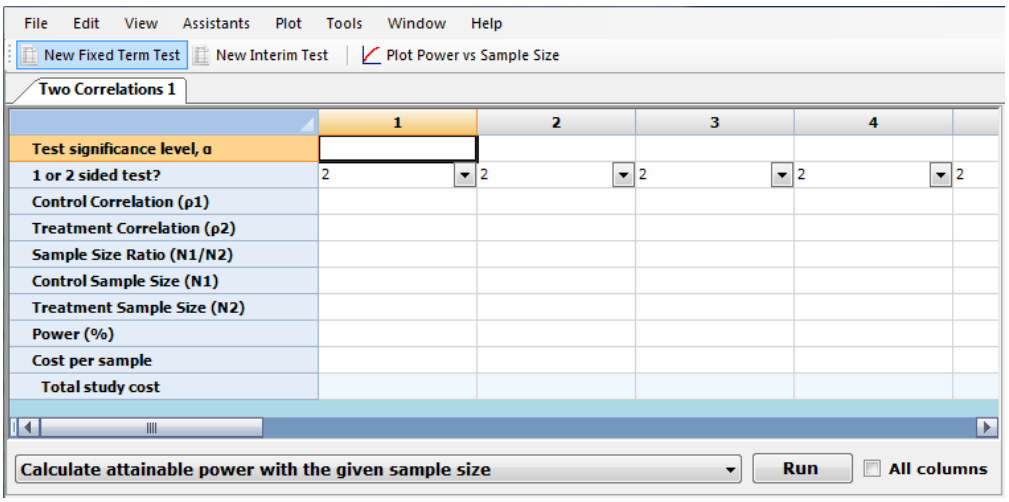


Figure 4.23.2. Two Correlations Table

The first calculation will be for power taken from Zar. We enter the values as follows:

3. First enter 0.05 for the “Test Significance level” row and 2 for “1 or 2 sided test?”

4. Enter 0.78 for the “Control Correlation” row and 0.84 for the “Treatment Correlation” row.

5. Finally, enter 98 in the “Control Sample Size” row and 95 in the “Treatment Sample size”. The table will appear as per Figure 4.28.3.

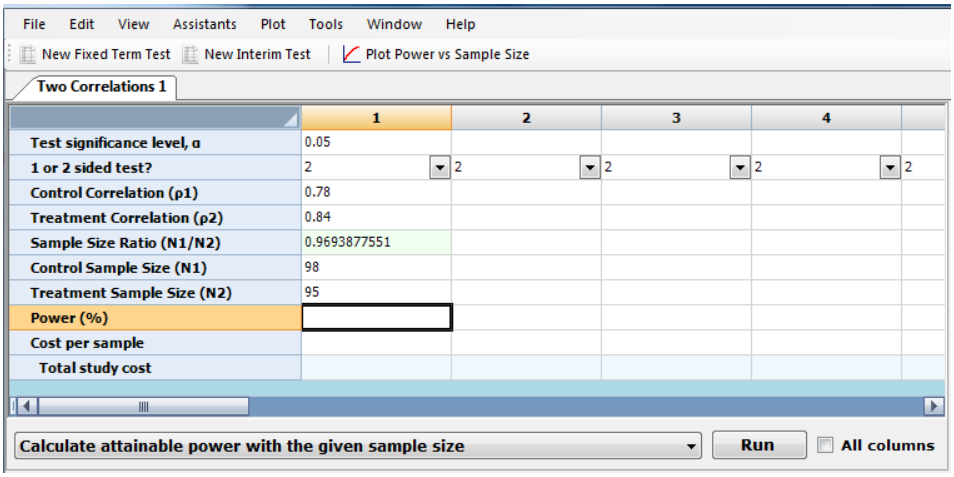


Figure 4.28.3. Values entered for Two Correlations study design

6. Select “Calculate attainable power with the given sample size” from the dropdown menu beside the “Run” button. Then click “Run”. This will give a result of 22.498 for power as in Figure 4.28.4. This agrees with the value of 22% given in Zar.

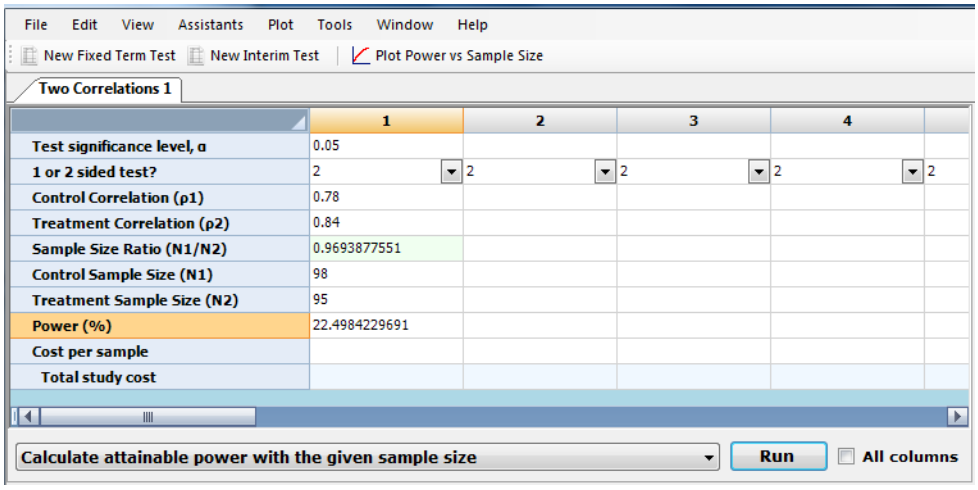
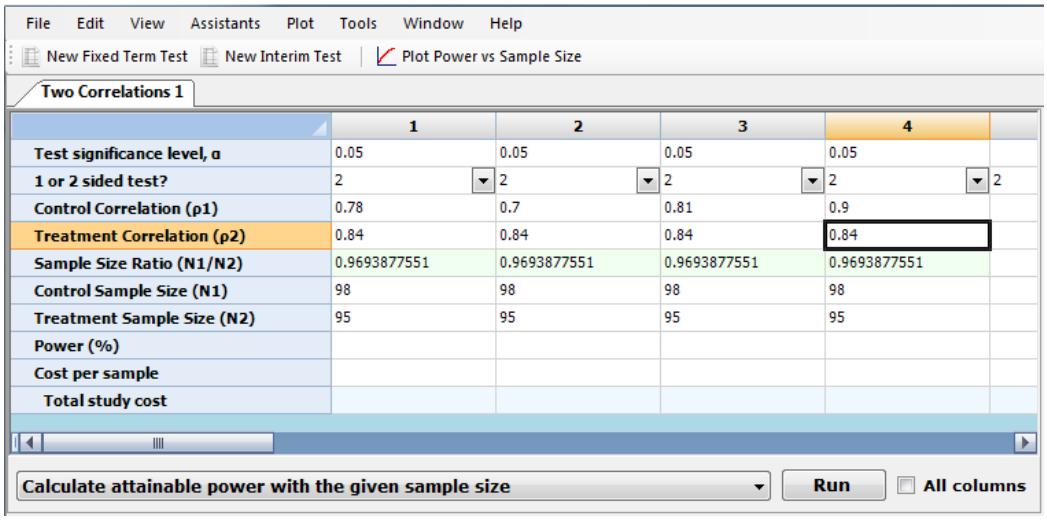


Figure 4.28.4. Completed Two Correlations study design

7. The next calculation is a sensitivity analysis for power where we change the treatment correlation to investigate the impact this has on the power estimate. To do this, copy the same values across to columns 2 to 4 where you delete the values for power. Then, change the values for columns 2 to 4 in the treatment correlation row to 0.7, 0.81 and 0.9 respectively. This will give a table as per Figure 4.28.5.

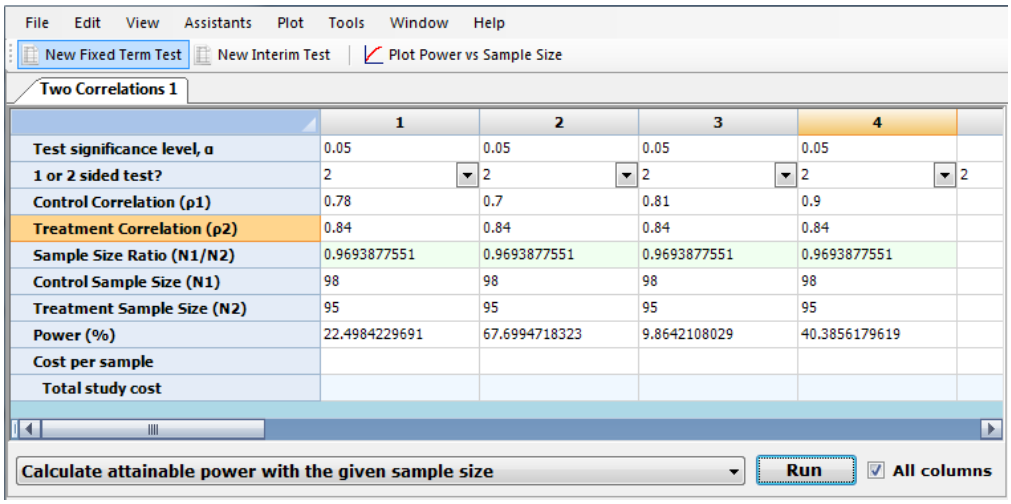


	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
1 or 2 sided test?	2	2	2	2
Control Correlation (ρ1)	0.78	0.7	0.81	0.9
Treatment Correlation (ρ2)	0.84	0.84	0.84	0.84
Sample Size Ratio (N1/N2)	0.9693877551	0.9693877551	0.9693877551	0.9693877551
Control Sample Size (N1)	98	98	98	98
Treatment Sample Size (N2)	95	95	95	95
Power (%)				
Cost per sample				
Total study cost				

Figure 4.28.5. Sensitivity analysis around the Treatment Correlation

8. Select "Calculate attainable power with the given sample size" from the dropdown menu beside the "Run" button. Then tick the box to run "All Columns". Then click "Run".

This will give the resultant powers of 22.498, 67.699, 9.864 and 40.386 sequentially as in Figure 4.28.6.



	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
1 or 2 sided test?	2	2	2	2
Control Correlation (ρ1)	0.78	0.7	0.81	0.9
Treatment Correlation (ρ2)	0.84	0.84	0.84	0.84
Sample Size Ratio (N1/N2)	0.9693877551	0.9693877551	0.9693877551	0.9693877551
Control Sample Size (N1)	98	98	98	98
Treatment Sample Size (N2)	95	95	95	95
Power (%)	22.4984229691	67.6994718323	9.8642108029	40.3856179619
Cost per sample				
Total study cost				

Figure 4.28.6. Results from Sensitivity analysis

9. Finally, by clicking on the desired study design column and going to the "Output" tab at the bottom of the screen, you can get an output statement for the results given by the current calculation as per Figure 4.28.7 below.

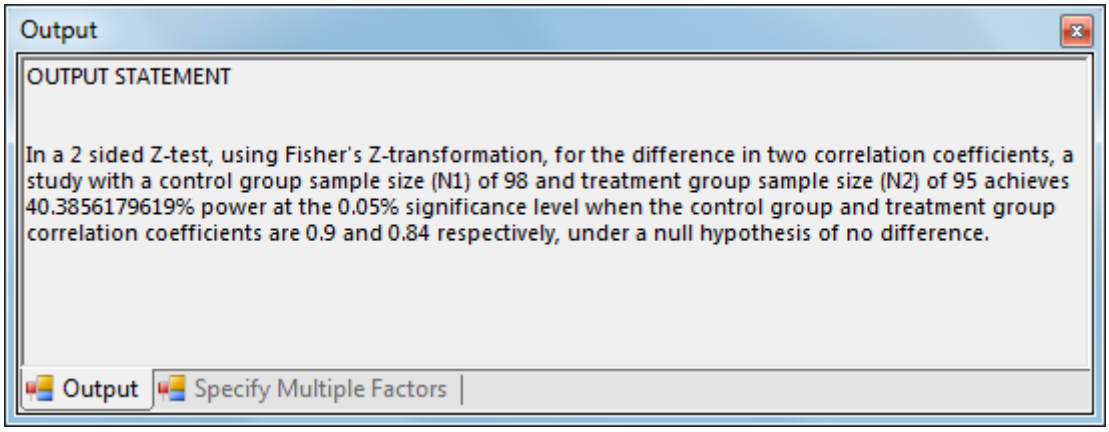


Figure 4.28.7. Study design Output statement

This Output statement can then be easily transferred directly from the output window into a report.

4.29. One Sensitivity

4.29.1 Introduction

The sensitivity is a widely used metric used to assess the accuracy of a classification method. Thus researchers are sometimes interested in hypotheses which are related to the sensitivity. One such use is the statistical comparison of a single sensitivity to a pre-specified sensitivity index (or one-sample case) to determine whether the sample sensitivity is equal to the index sensitivity.

This table facilitates the calculation of the power and sample size for hypothesis tests comparing one sensitivity to an index sensitivity. Power and sample size calculations are taken from Li and Fine (2004).

4.29.2 Methodology

This table provides sample size and power calculations for studies which will be comparing one sensitivity to a sensitivity index. The sample size calculation for one sensitivity is taken from Li and Fine (2004). The test of interest is based on the binomial test since the sensitivity is itself a binomial proportion. The sensitivity is defined as the probability that a given classification will be positive for a subject that is actually positive. The following table gives an example of how the sensitivity would be calculated for a diagnostic test:

Table 4.29.1: Confusion Matrix for Diagnostic Test

		Test Value for Disease		
		Positive Test	Negative Test	Prevalence
Actual Subject Value	Disease	A (TP)	B (FN)	A+B
	Not Disease	C (FP)	D (TN)	C+D
	Accuracy	A+C	B+D	A+B+C+D

Based on Table 4.29.1, the sensitivity would be equal to the following:

$$Sensitivity = \frac{A}{A+B} \qquad \{4.29.1\}$$

Where A is the number of subjects who had the disease (or True Positive Rate) and positive test and $A+B$ is equal to the number of subjects with the disease. Since this value is a proportion we can use the binomial distribution to conduct hypothesis tests for the sensitivity and thus the sample size and power are derived from exact methods for binomial proportions, which is defined based on the following binomial probability distribution:

$$b(x|n,p) = \binom{n}{x} p^x (1-p)^{n-x} \qquad \{4.29.2\}$$

In this table, we are assuming that the null and alternative sensitivity are the same under the null hypothesis and the difference between them is measured by the difference of the null hypothesis and alternative hypothesis sensitivities.

This table can be used to calculate the power, the sample size, the null hypothesis sensitivity and the alternative hypothesis sensitivity given all other terms in the table are specified.

To calculate power and sample size, the user must specify the test significance level (α), whether to use a one or two sided test, the null hypothesis sensitivity (Se_0), the alternative hypothesis sensitivity (Se_1) and the prevalence of the disease. Note that the sample size and power calculations are actually for the number of positive patients and that this is adjusted using method zero from Li and Fine in which the positive patient sample size is simply divided by the prevalence.

The power formula uses the exact binomial distribution to calculate power. This involves a two-step method to calculate power which is outlined below (Note that $N_{Positive} = N_{Total} \times Prevalence$)

Step 1

$$X_{Crit1} = B^{-1}\left(1 - \frac{alpha}{\#\ sided\ test}, N_{Positive}, Se_0\right) \qquad \{4.29.3\}$$
$$X_{Crit2} = B^{-1}\left(\frac{alpha}{\#\ sided\ test}, N_{Positive}, Se_0\right)$$

Step 2

$$\begin{cases} 1-\beta = 1 - B(X_{Crit1}, N_{Positive}, Se_1), & Se_0 < Se_1\ One-Sided \\ 1-\beta = B(X_{Crit2} - 1, N_{Positive}, Se_1), & Se_0 > Se_1\ One-Sided \\ 1-\beta = 1 - B(X_{Crit1}, N_{Positive}, Se_1) + B(X_{Crit2} - 1, N_{Positive}, Se_1), Two-sided \end{cases} \qquad \{4.29.4\}$$

Where $B(X, N, p)$ is the cumulative binomial distribution assessed at X number of positive trials from N trials with success probability p and $B^{-1}(\pi, N, p)$ is the inverse cumulative binomial distribution with N number of trials with success probability p for the specified cumulative distribution probability π.

A closed form equation is not used to calculate the sample size, the null hypothesis sensitivity or the alternative hypothesis sensitivity. Instead a search algorithm is used. The search algorithm calculates power at various values for the relevant parameter until the desired power is reached.

4.29.3 Examples

Example 1: Validation example calculating required sample size for a given power
The following example is taken from Table 1 of Li and Fine (2004) where we replicate a sample size calculation table. Then we conduct a sensitivity analysis to show the effect of changing the null hypothesis sensitivity on power with fixed sample size.

The following steps outline the procedure for this example:

1. Open nQuery + nTerim 4.0 via the start menu or desktop shortcut. Click “New Fixed Term Test” from the top of the window. The “Study Goal and Design” window will appear as Figure 4.29.1 below.

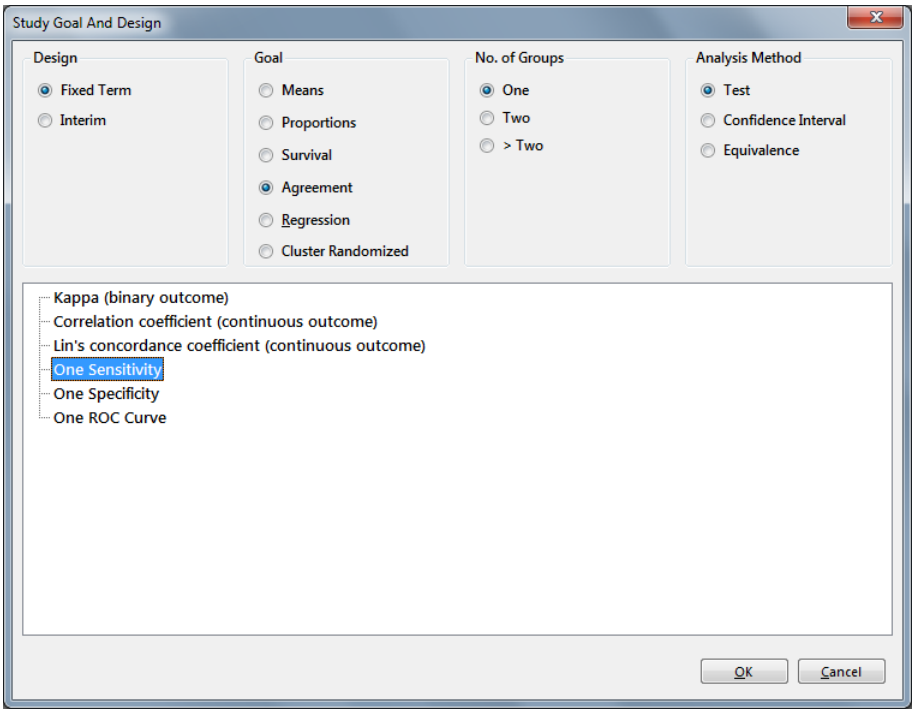


Figure 4.29.1. Study Goal and Design Window

2. In order to select the “One Sensitivity” design table, navigate through the Study Goal and Design Window by selecting "Fixed Term" as the Design, "Agreement" as the Goal, "One" as the Number of Groups and "Test" as the Analysis Method. Then click "OK" and the test window will appear as per Figure 4.29.2.

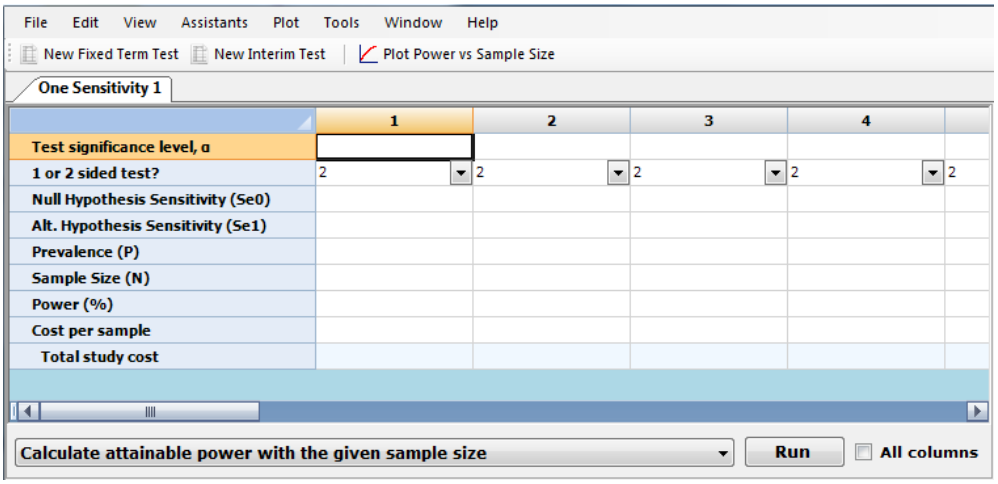


Figure 4.29.2. One Sensitivity Table

The first calculation will be for sample size taken from Table 1 of Li and Fine. We enter the values as follows:

3. First enter 0.05 for the “Test Significance level” row and 1 for “1 or 2 sided test?”

4. Then enter 0.5 and 0.9 for the “Null Hypothesis Sensitivity” and “Alternative Hypothesis Sensitivity” variable rows respectively.

5. Finally, enter 0.01 for the “Prevalence” row and enter 90 in the “Power” row. The table will appear as per Figure 4.29.3.

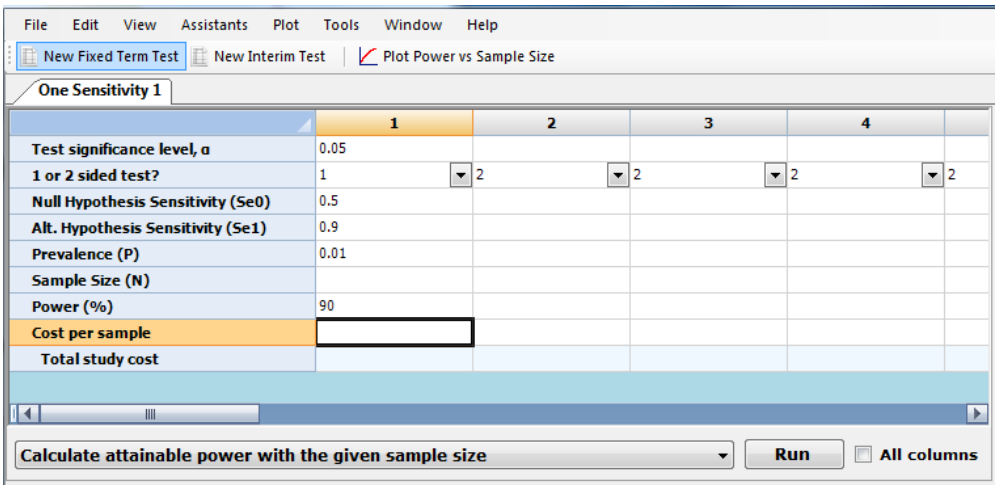


Figure 4.29.3. Values entered for One Sensitivity study design

6. Select “Calculate required sample size for given power” from the dropdown menu beside the “Run” button. Then click “Run”. This will give a result of 1100 for sample size as in Figure 4.29.4. This agrees with the value of 1100 given in Li and Fine.

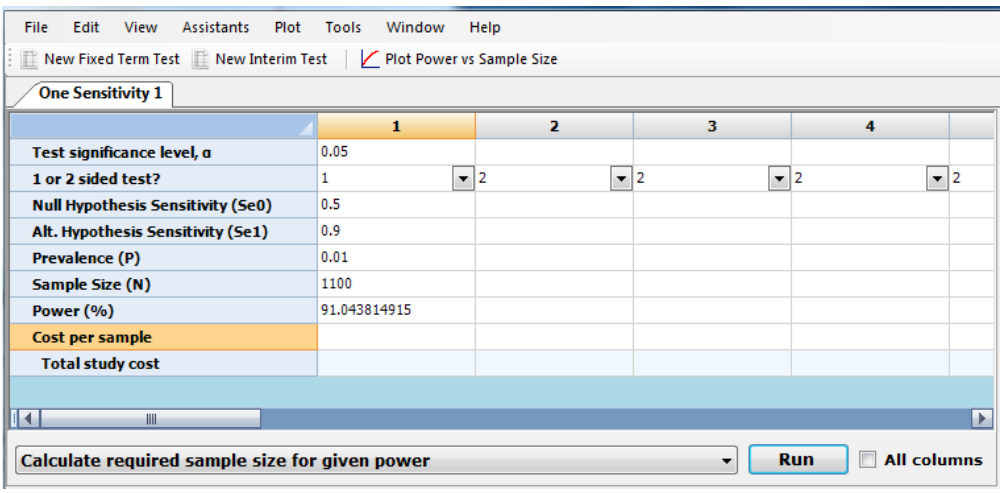
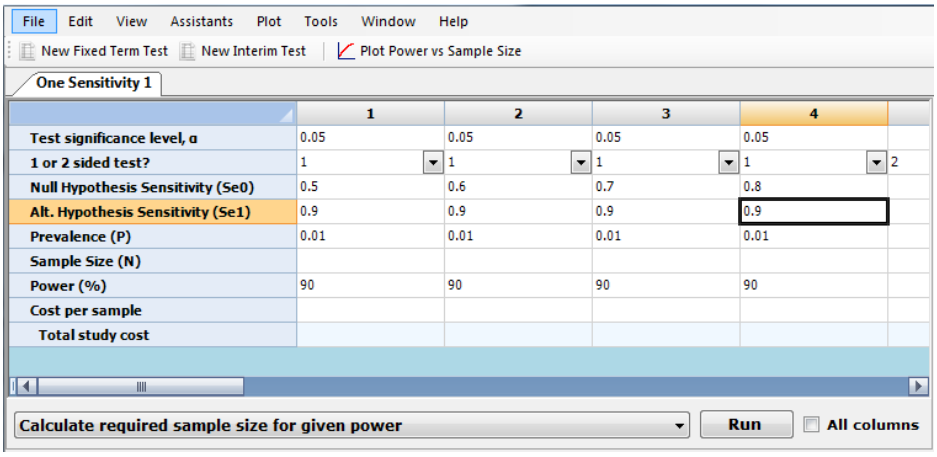


Figure 4.29.4. Completed One Sensitivity study design

7. The next calculation is a sensitivity analysis for sample size where we change the null hypothesis sensitivity. To do this, copy the same values across to columns 2 to 4 where with power reset to 90. Then, change the values for columns 2 to 4 in the “Alternative Hypothesis Sensitivity” row to 0.6, 0.7 and 0.8 respectively. This will give a table as per Figure 4.29.5.

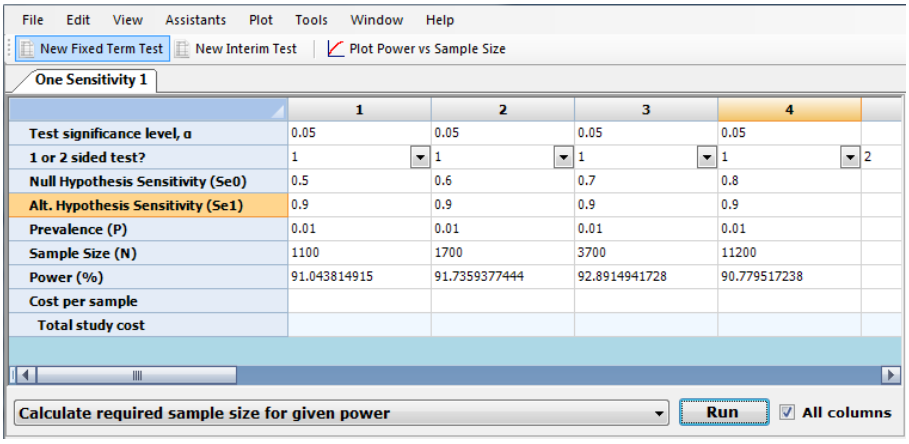


	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
1 or 2 sided test?	1	1	1	1
Null Hypothesis Sensitivity (Se0)	0.5	0.6	0.7	0.8
Alt. Hypothesis Sensitivity (Se1)	0.9	0.9	0.9	0.9
Prevalence (P)	0.01	0.01	0.01	0.01
Sample Size (N)				
Power (%)	90	90	90	90
Cost per sample				
Total study cost				

Figure 4.29.5. Sensitivity analysis around the Null Hypothesis Sensitivity

8. Select "Calculate attainable power for given sample size" from the dropdown menu beside the "Run" button. Then tick the box to run "All Columns". Then click "Run".

This will give the resultant powers of 1100, 1700, 3700 and 11200 sequentially as in Figure 4.29.6. These correspond to values given in Table 1 of Li and Fine.



	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
1 or 2 sided test?	1	1	1	1
Null Hypothesis Sensitivity (Se0)	0.5	0.6	0.7	0.8
Alt. Hypothesis Sensitivity (Se1)	0.9	0.9	0.9	0.9
Prevalence (P)	0.01	0.01	0.01	0.01
Sample Size (N)	1100	1700	3700	11200
Power (%)	91.043814915	91.7359377444	92.8914941728	90.779517238
Cost per sample				
Total study cost				

Figure 4.29.6. Results from Sensitivity analysis

9. Finally, by clicking on the desired study design column and going to the "Output" tab at the bottom of the screen, you can get an output statement as per Figure 4.29.7.

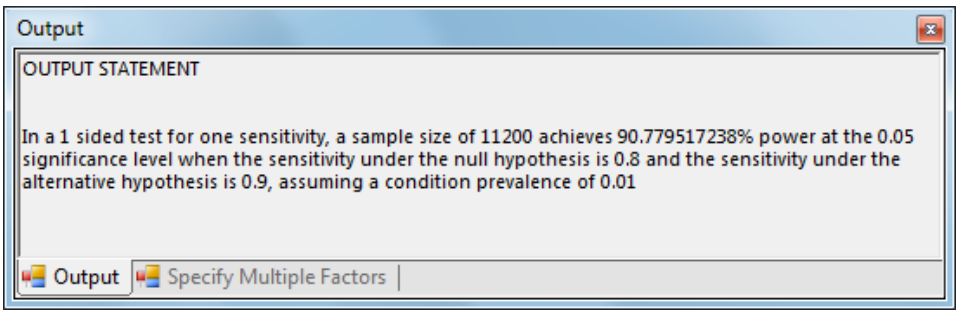


Figure 4.29.7. Study design Output statement

This Output statement can then be easily transferred directly from the output window into a report.

4.30. One Specificity

4.30.1 Introduction

The specificity is a widely used metric used to assess the accuracy of a classification method. Thus researchers are sometimes interested in hypotheses which are related to the specificity. One such use is the statistical comparison of a single specificity to a pre-specified specificity index (or one-sample case) to determine whether the sample specificity is equal to the index specificity.

This table facilitates the calculation of the power and sample size for hypothesis tests comparing one specificity to an index specificity. Power and sample size calculations are taken from Li and Fine (2004).

4.30.2 Methodology

This table provides sample size and power calculations for studies which will be comparing one specificity to a specificity index. The sample size calculation for one specificity is taken from Li and Fine (2004). The test of interest is based on the binomial test since the specificity is itself a binomial proportion. The specificity is defined as the probability that a given classification will be negative for a subject that is actually negative. The following table gives an example of how the specificity would be calculated for a diagnostic test:

Table 4.30.1: Confusion Matrix for Diagnostic Test

		Test Value for Disease		
		Positive Test	Negative Test	Prevalence
Actual Subject Value	Disease	A (TPR)	B (FNR)	A+B
	Not Disease	C (FPR)	D (TNR)	C+D
	Accuracy	A+C	B+D	A+B+C+D

Based on Table 4.30.1, the specificity would be equal to the following:

$$Specificity = \frac{D}{C + D} \qquad \{4.30.1\}$$

Where D is the number of subjects who did not had the disease (or True Negative Rate) and negative test and $C + D$ is equal to the number of subjects without the disease. Since this value is a proportion we can use the binomial distribution to conduct hypothesis tests for the specificity and thus the sample size and power are derived from exact methods for binomial proportions, which is defined based on the following binomial probability distribution:

$$b(x|n,p) = \binom{n}{x} p^x (1-p)^{n-x} \qquad \{4.30.2\}$$

In this table, we are assuming that the null and alternative specificity are the same under the null hypothesis and the difference between them is measured by the difference of the null and alternative specificities.

This table can be used to calculate the power, the sample size, the null hypothesis specificity and the alternative hypothesis specificity given all other terms in the table are specified.

To calculate power and sample size, the user must specify the test significance level (α), whether to use a one or two sided test, the null hypothesis specificity (Sp_0), the alternative hypothesis specificity (Sp_1) and the prevalence of the disease. Note that the sample size and power calculations are actually for the number of negative patients and that this is adjusted using method zero from Li and Fine in which the negative patient sample size is simply divided by the one minus the prevalence.

The power formula uses the exact binomial distribution to calculate power. This involves a two-step method to calculate power which is outlined below (Note that $N_{Negative} = N_{Total} \times (1 - Prevalence)$)

1 $$X_{Crit1} = B^{-1}\left(1 - \frac{alpha}{\#\ sided\ test}, N_{Negative}, 1 - Sp_0\right) \qquad \{4.30.3\}$$
$$X_{Crit2} = B^{-1}\left(\frac{alpha}{\#\ sided\ test}, N_{Negative}, 1 - Sp_0\right)$$

2 $$\begin{cases} 1 - \beta = 1 - B(X_{Crit1}, N_{Negative}, 1 - Sp_1), & Sp_0 > Sp_1\ One - Sided \\ 1 - \beta = B(X_{Crit2} - 1, N_{Negative}, 1 - Sp_1), & Sp_0 < Sp_1\ One - Sided \\ 1 - \beta = 1 - B(X_{Crit1}, N_{Negative}, 1 - Sp_1) + B(X_{Crit2} - 1, N_{Negative}, 1 - Sp_1), Two - sided \end{cases} \qquad \{4.30.4\}$$

Where $B(X, N, p)$ is the cumulative binomial distribution assessed at X number of negative trials from N trials with failure probability p and $B^{-1}(\pi, N, p)$ is the inverse cumulative binomial distribution with N number of trials with failure probability p for the specified cumulative distribution probability π.

A closed form equation is not used to calculate the sample size, the null hypothesis specificity or the alternative hypothesis specificity. Instead a search algorithm is used. The search algorithm calculates power at various values for the relevant parameter until the desired power is reached.

4.30.3 Examples

Example 1: Validation example calculating required sample size for a given power
The validation for specificity is based on the validation from Chapter 4.29. By taking one-minus the values provided for the sensitivities and one minus the prevalence, the two calculations can be shown to be equivalent. Then we conduct a sensitivity analysis to show the effect of changing the null hypothesis specificity on power with fixed sample size to replicate Example 1 from chapter 4.29.

The following steps outline the procedure for this example:

1. Open nQuery + nTerim 4.0 via the start menu or desktop shortcut. Click “New Fixed Term Test” from the top of the window. The “Study Goal and Design” window will appear as Figure 4.30.1 below.

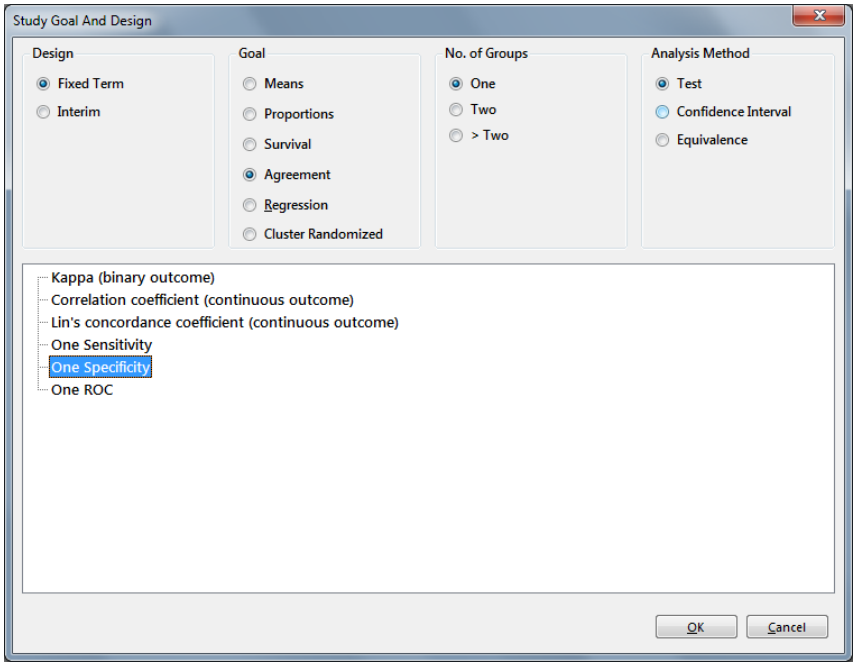


Figure 4.30.1. Study Goal and Design Window

2. In order to select the “One Specificity” design table, navigate through the Study Goal and Design Window by selecting "Fixed Term" as the Design, "Agreement" as the Goal, "One" as the Number of Groups and "Test" as the Analysis Method. Then click "OK" and the test window will appear as per Figure 4.30.2.

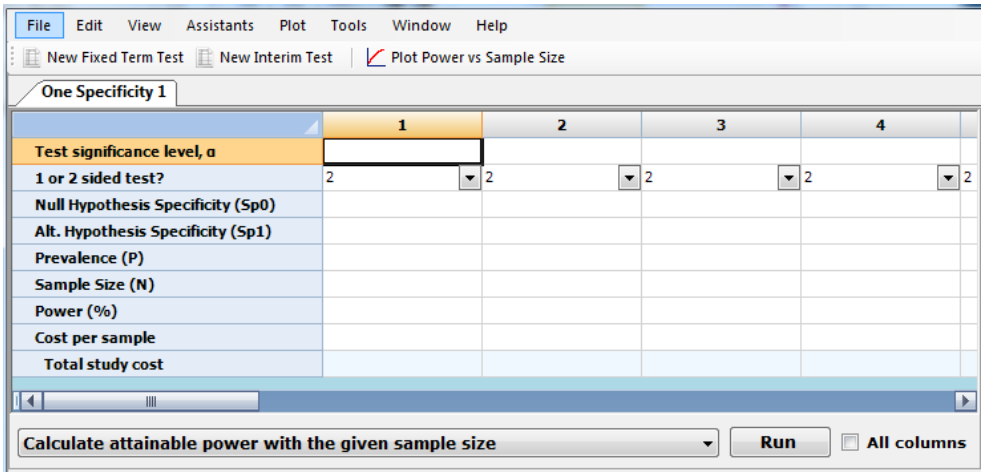


Figure 4.30.2. One Specificity Table

The first calculation will be from sample size taken from Table 1 of Li and Fine except using one minus the sensitivity values (0.5 and 0.9 in Chapter 4.29) and prevalence (0.01 in chapter 4.29). We enter the values as follows:

3. First enter 0.05 for the “Test Significance level” row and 1 for “1 or 2 sided test?”

4. Then enter 0.5 (i.e. $1 - 0.5$) and 0.1 (i.e. $1 - 0.9$) for the “Null Hypothesis Specificity” and “Alternative Hypothesis Specificity” variable rows respectively.

5. Finally, enter 0.99 (i.e. $1 - 0.01$) for the “Prevalence” row and enter 90 in the “Power” row. The table will appear as per Figure 4.30.3.

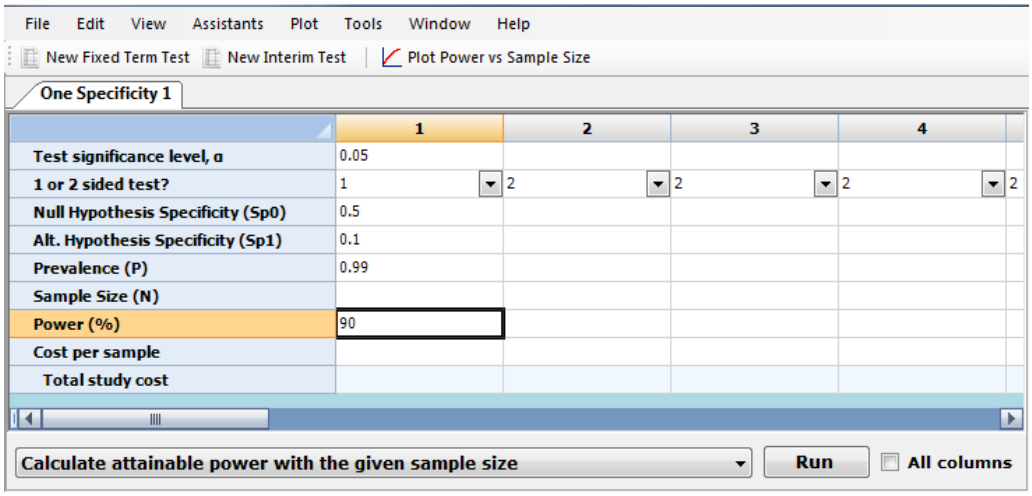


Figure 4.30.3. Values entered for One Specificity study design

6. Select “Calculate required sample size for given power” from the dropdown menu beside the “Run” button. Then click “Run”. This will give a result of 1100 for sample size as in Figure 4.30.4. This agrees with the value from Chapter 4.29.

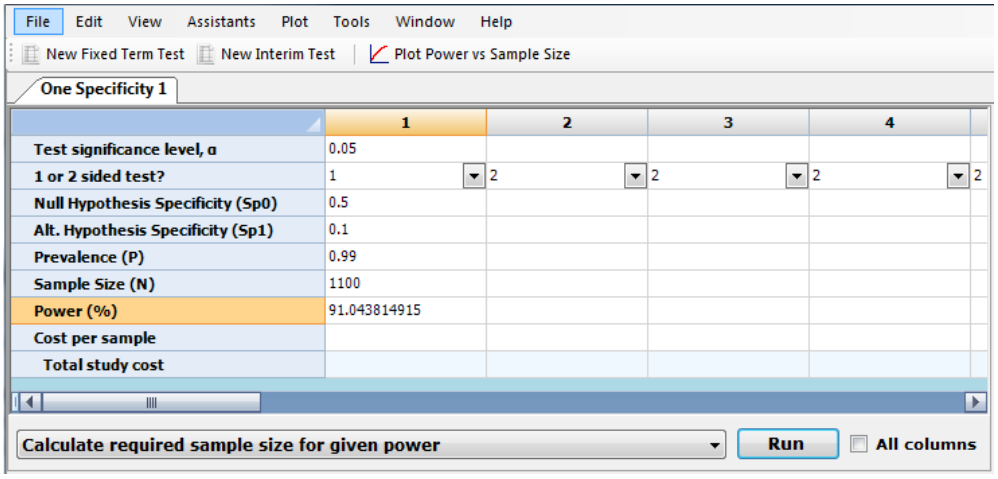


Figure 4.30.4. Completed One Specificity study design

7. The next calculation is a sensitivity analysis for sample size where we change the null hypothesis specificity. To do this, copy the same values across to columns 2 to 4 where with power reset to 90. Then, change the values for columns 2 to 4 in the “Alternative Hypothesis Specificity” row to 0.4, 0.3 and 0.2 respectively. This will give a table as per Figure 4.30.5.

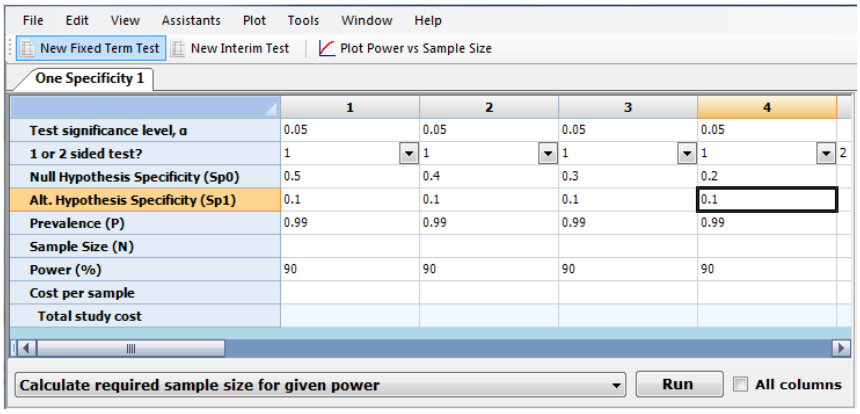


Figure 4.30.5. Sensitivity analysis around the Null Hypothesis Specificity

8. Select "Calculate attainable power for given sample size" from the dropdown menu beside the "Run" button. Then tick the box to run "All Columns". Then click "Run".

This will give the resultant powers of 1100, 1700, 3700 and 11200 sequentially as in Figure 4.30.6. These correspond to values given Chapter 4.29.

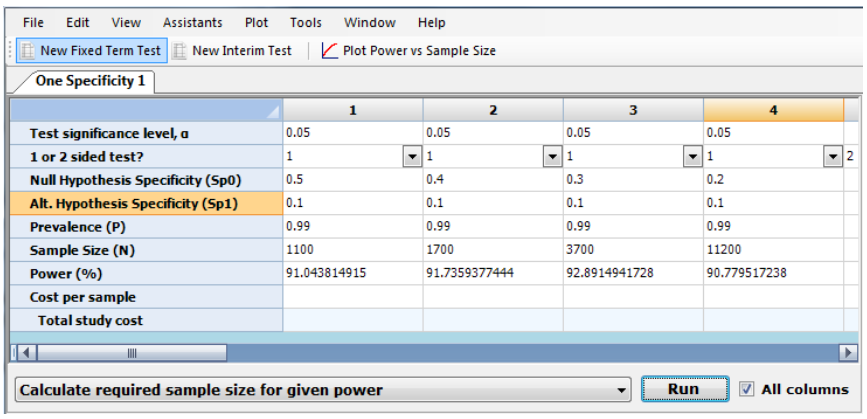


Figure 4.30.6. Results from Sensitivity analysis

9. Finally, by clicking on the desired study design column and going to the "Output" tab at the bottom of the screen, you can get an output statement as per Figure 4.30.7.

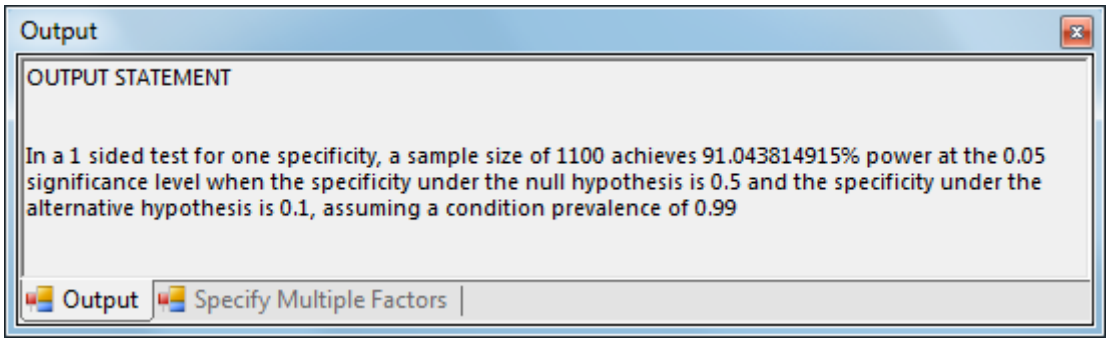


Figure 4.30.7. Study design Output statement

This Output statement can then be easily transferred directly from the output window into a report.

Chapter 5

Cluster Randomized Trials

5.1. CRT Two Means – Completely Randomized

5.1.1 Introduction

Continuous data is found in nearly every area of research interest. Procedures such as Z and t-tests are used to evaluate the differences between two continuous means. Clustered data is very common in a wide variety of academic, social policy and economic studies. This two sample test is used to test hypotheses about the difference between two means in a completely randomized cluster randomized trial.

This table facilitates the calculation of the power and sample size for hypothesis tests comparing means in a cluster randomized trial. Power and sample size is computed using the method outlined by Donner and Klar (1996).

5.1.2 Methodology

This table provides sample size and power calculations for studies which will be comparing means which use a completely randomized cluster randomization study design. A completely randomized design assigns clusters randomly to control and treatment groups. This table assumes a balanced study design.

The sample size calculation for means is taken from Donner and Klar (1996) and is an extension of the t-test for the comparison of two independent means. The extension uses the intracluster correlation (ICC), which is the ordinary product moment correlation between any two observations in the same cluster, to adjust for the effect of within cluster correlation. For means, it can be calculated as the proportion of variability explained by the between-cluster variation. It is assumed the ICC is the same in both groups.

This table can be used to calculate the power, the test significance level (α), the number of clusters per group, the sample size per cluster, the intracluster correlation, the standard deviation and the smallest detectable difference given all other terms in the table are specified.

To calculate power, the number of clusters or the sample size per cluster, the user must specify the test significance level (α), whether to use a one or two sided test, the difference in means ($|\mu_1 - \mu_2|$), the standard deviation (σ) and the intracluster correlation (ICC).

The formulae use the non-central t-distribution to calculate power. This distribution requires the appropriate degrees of freedom and a non-centrality parameter (NC). The formula for the power is given by the following equation:

$$Power = 1 - t_{2(m-1),NC}\left(t^{-1}_{2(m-1)}\left(1-\frac{\alpha}{2}\right)\right) + t_{2(m-1),NC}\left(t^{-1}_{2(m-1)}\left(1-\frac{\alpha}{2}\right)\right) \qquad \{5.1.1\}$$

where

$$NC = \frac{(|\mu_1 - \mu_2|/\sigma)}{\left[\left(2\left(1+\left(ICC(N-1)\right)\right)\right)/(mN)\right]^{1/2}} \qquad \{5.1.2\}$$

and m is the number of clusters per group and N is the sample size per cluster. For the one-sided test, $z_{1-\alpha/2}$ is replaced with $z_{1-\alpha}$.

A closed form equation is not used to calculate the number of clusters, the sample size per cluster, the test significance levels, the standard deviation or the smallest detectable difference. Instead a search algorithm is used. The search algorithm calculates power at various values for the relevant parameter until the desired power is reached.

5.1.3 Examples

Example 1: Validation example calculating required sample size for a given power
The following examples are taken from Donner and Klar (1996) where a power calculation problem is conducted and then a sensitivity analysis is conducted to show the effect of changing the sample size per cluster and difference between means. The following steps outline the procedure for this example:

1. Open nQuery + nTerim 4.0 via the start menu or desktop shortcut. Click “New Fixed Term Test” from the top of the window. The “Study Goal and Design” window will appear as Figure 5.1.1 below.

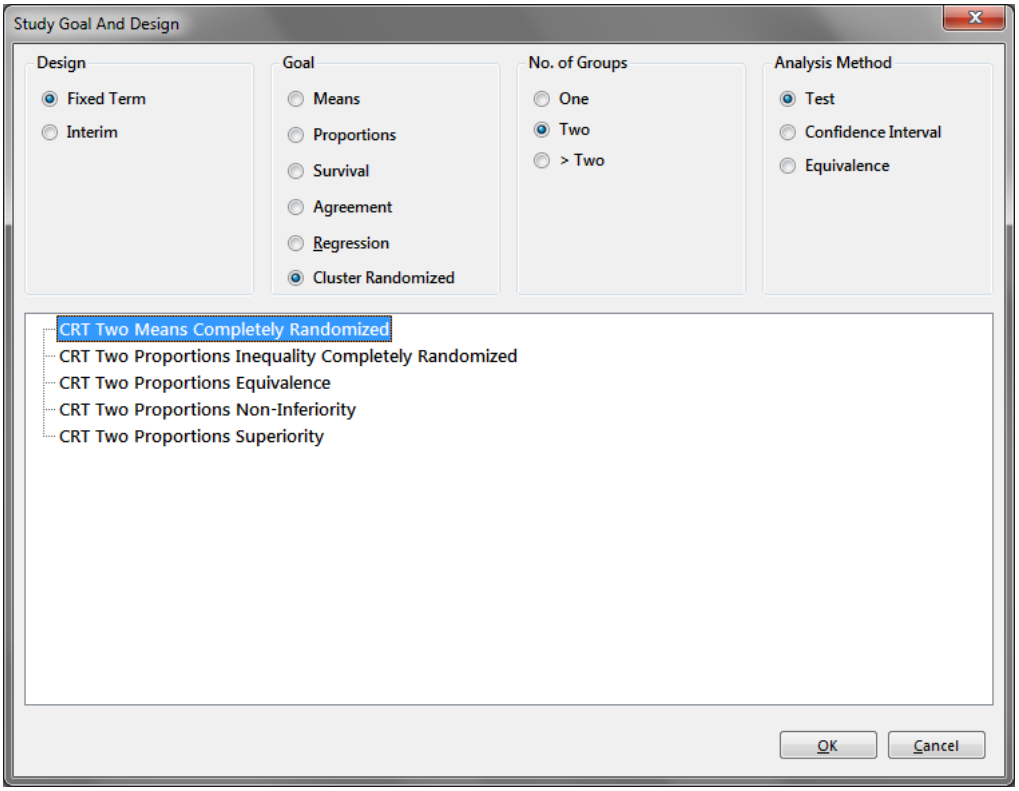


Figure 5.1.1. Study Goal and Design Window

2. In order to select the CRT Two Means Completely Randomized design table, navigate through the Study Goal and Design Window by selecting "Fixed Term" as the Design, "Cluster Randomized" as the Goal, "Two" as the Number of Groups and "Test" as the Analysis Method. Then click "OK" and the test window will appear as per Figure 5.1.2.

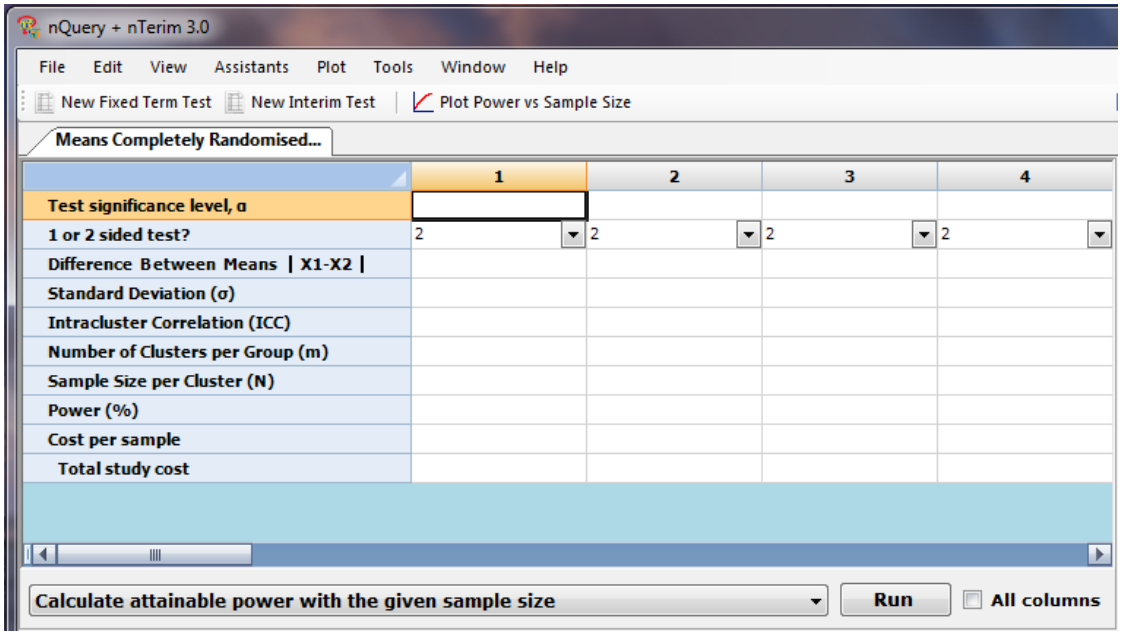


Figure 5.1.2. CRT Two Means Inequality Completely Randomized Test Table

The first calculation that will be performed is for power. This example is can be undertaken using the following steps.

3. First enter 0.05 for the Test Significance level row, then select 2 for the "1 or 2 sided test?" variable dropdown menu, enter 0.2 for the "Difference Between Means", enter 1 for the "Standard Deviation" variable and enter 0.01 for the "Intracluster Correlation".
4. Finally, enter 5 for the number of clusters per group and 100 for the sample size per cluster. The table will appear as per Figure 5.1.3.

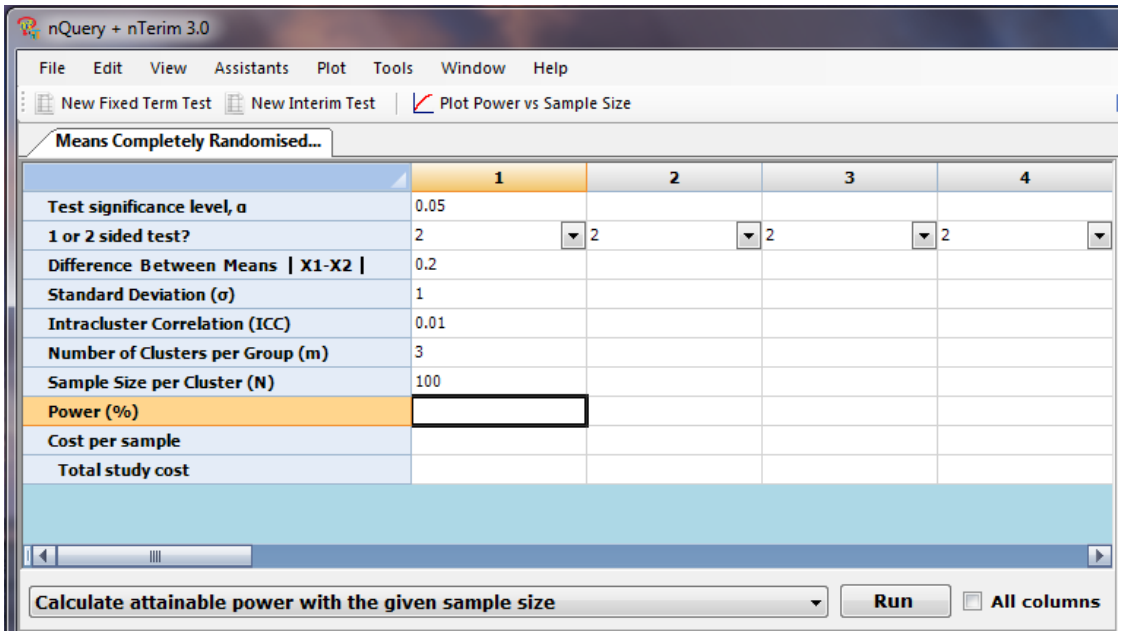
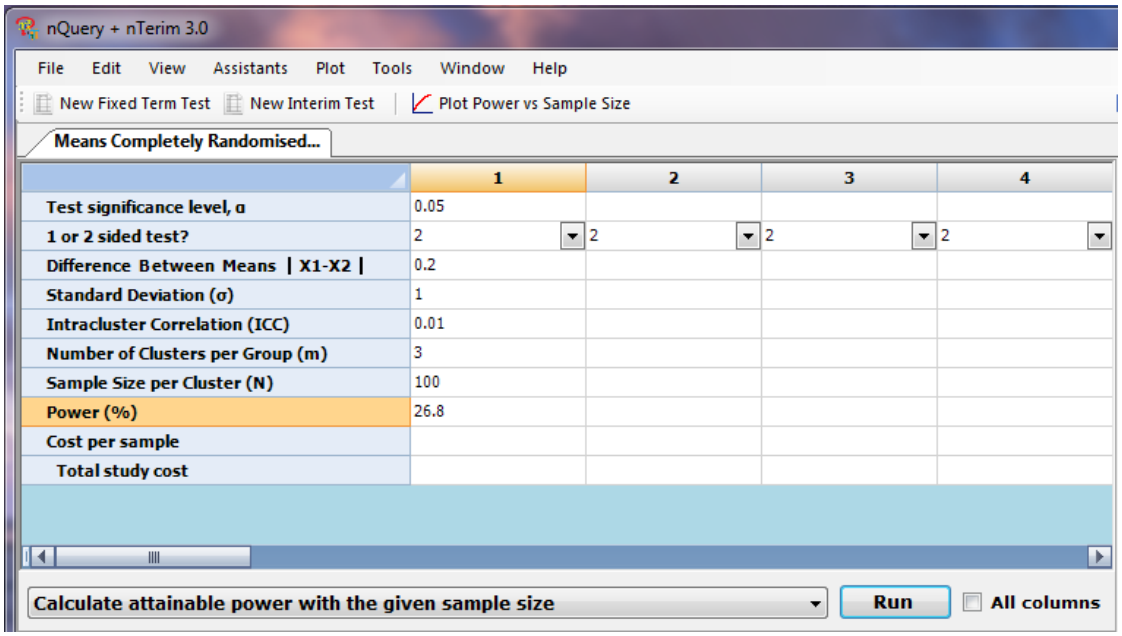


Figure 5.1.3. Values entered for CRT Two Means Completely Randomized study design

5. Select “Calculate attainable power with given sample size” from the dropdown menu beside the “Run” button. Then click “Run”. This will give a result of 26.8 for the power as in Figure 5.1.4.

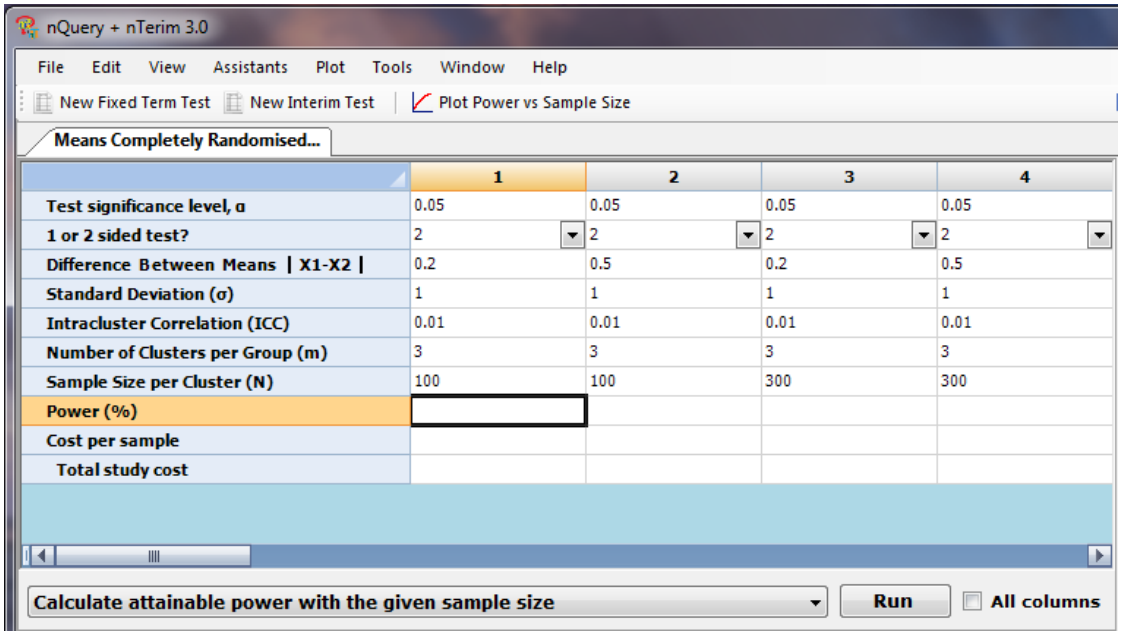


	1	2	3	4
Test significance level, α	0.05			
1 or 2 sided test?	2	2	2	2
Difference Between Means | X1-X2 |	0.2			
Standard Deviation (σ)	1			
Intracluster Correlation (ICC)	0.01			
Number of Clusters per Group (m)	3			
Sample Size per Cluster (N)	100			
Power (%)	26.8			
Cost per sample				
Total study cost				

Figure 5.1.4. Completed CRT Two Means Completely Randomized study design

The next calculation is a sensitivity analysis for power where we change the difference between means and the sample size per cluster and see their effect on power. These values are taken from Table 1 of Donner and Klar (1996).

6. Copy the first column above into columns 2 to 4. Then, change the value for “Difference between Means” to 0.5 in columns 2 and 4 and “Sample Size per Cluster” to 300 in columns 3 and 4. This will give a table as per Figure 5.1.5.

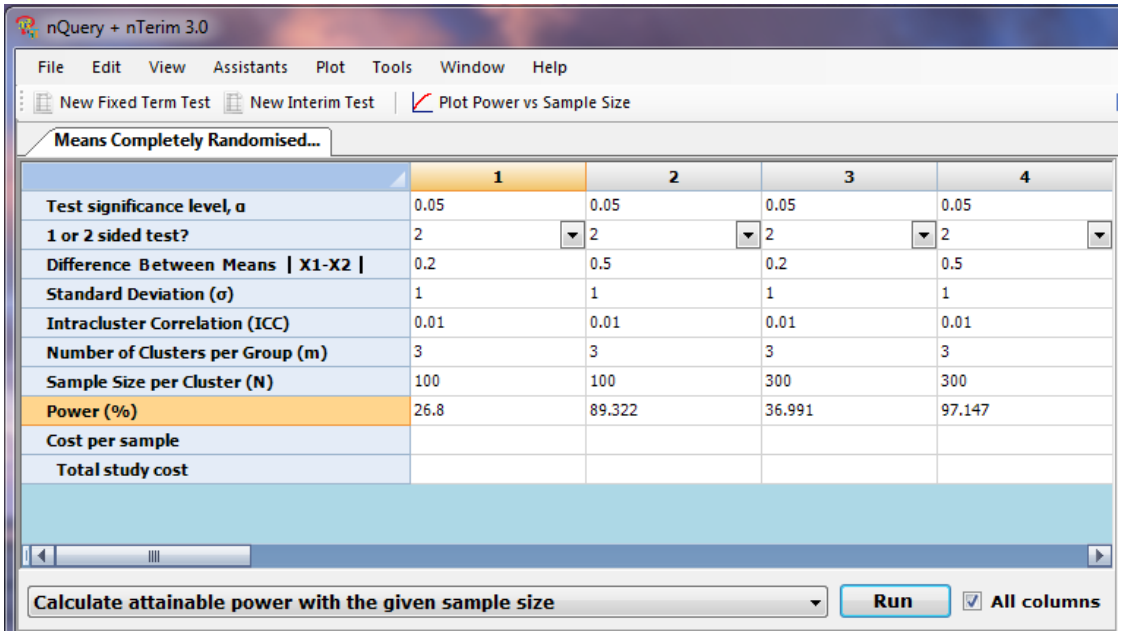


	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
1 or 2 sided test?	2	2	2	2
Difference Between Means | X1-X2 |	0.2	0.5	0.2	0.5
Standard Deviation (σ)	1	1	1	1
Intracluster Correlation (ICC)	0.01	0.01	0.01	0.01
Number of Clusters per Group (m)	3	3	3	3
Sample Size per Cluster (N)	100	100	300	300
Power (%)				
Cost per sample				
Total study cost				

Figure 5.1.5. Sensitivity analysis around Difference Between Means and Sample Size

7. Select “Calculate attainable power with given sample size” from the dropdown menu beside the “Run” button. Then tick the box to run “All Columns”. Then click “Run”.

This will give the resultant powers of 26.8%, 89.322%, 36.991% and 97.147% sequentially as in Figure 5.1.6.



	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
1 or 2 sided test?	2	2	2	2
Difference Between Means | X1-X2 |	0.2	0.5	0.2	0.5
Standard Deviation (σ)	1	1	1	1
Intracluster Correlation (ICC)	0.01	0.01	0.01	0.01
Number of Clusters per Group (m)	3	3	3	3
Sample Size per Cluster (N)	100	100	300	300
Power (%)	26.8	89.322	36.991	97.147
Cost per sample				
Total study cost				

Figure 5.1.6. Results from Sensitivity analysis

The effect of changing these parameters on the relationship between Power and Sample size can be explored further using the “Plot Power vs Sample Size” button at the top of the screen.

8. Select the first column by clicking the 1 at the top of column 1. Then hold down the Shift key and click the 4 at the top of column 4. All four columns will now be highlighted.

9. Click the “Plot Power vs Sample Size” button at the top of the screen. This will give you a plot as displayed in Figure 5.1.7 which will show the relationship between power and number of clusters for each column. Right-click to add features such as a legend to the graph and double click elements for user options and editing.

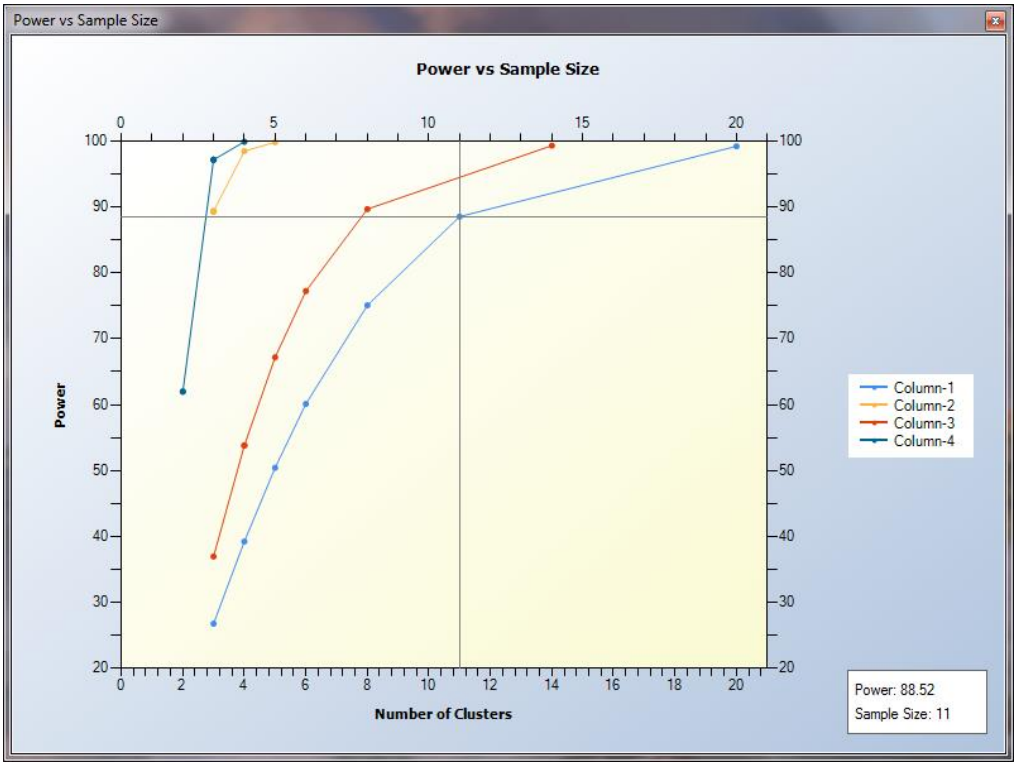


Figure 5.1.7. Power vs Number of Clusters plot

10. Finally, by clicking on the desired study design column and going to the "Output" tab at the bottom of the screen, you can get an output statement for the results given by the current calculation as per Figure 5.1.8.

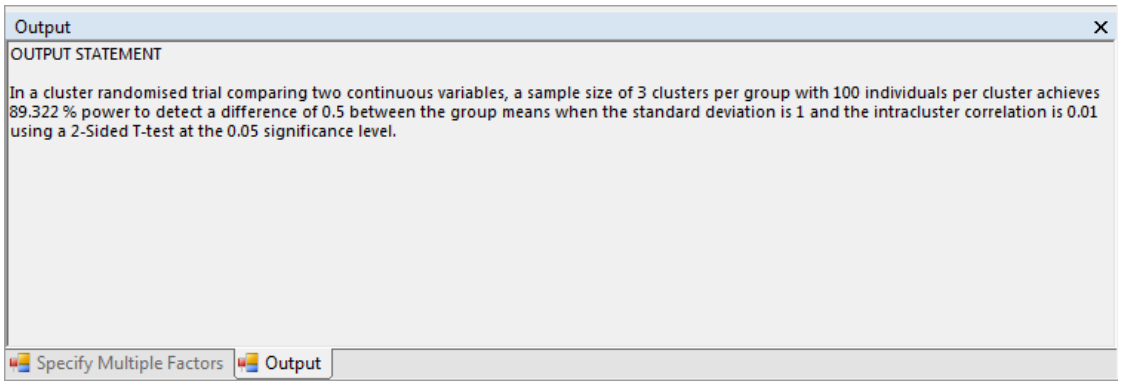


Figure 5.1.8. Study design Output statement

This Output statement can then be easily transferred directly from the output window into a report.

5.2. CRT Two Proportions Inequality – Completely Randomized

5.2.1 Introduction

Binary data is commonly studied in variety of different fields. Clustered data is very common in a wide variety of academic, social policy and economic studies. This two sample test is used to test hypotheses about the difference between two proportions in a completely randomized cluster randomized trial.

This table facilitates the calculation of the power and sample size for hypothesis tests comparing proportions in a completely randomized cluster randomized trial. Power and sample size is computed using the method outlined by Donner and Klar (2000).

5.2.2 Methodology

This table provides sample size and power calculations for studies which will be comparing proportions in a trial which uses a completely randomized cluster randomization study design. A completely randomized design assigns clusters randomly to control and treatment groups.

The sample size calculation for cluster randomized proportions is taken from Donner and Klar (2000) and is an extension of the methods used to compare two proportions in a fully randomized trial. The extension uses the intracluster correlation (ICC), which is the ordinary product moment correlation between any two observations in the same cluster, to adjust for the effect of within cluster correlation. It is assumed the ICC is the same in both groups.

This table can be used to calculate the power, the number of clusters in the treatment group, the sample size per cluster in the treatment group, the intracluster correlation and the smallest detectable difference given all other terms in the table are specified.

To calculate power, the number of clusters in the treatment or the sample size per cluster in the treatment group, the user must specify the test significance level (α), whether to use a one or two sided test, the control group proportion (p_2), which test type to be used, the format of the test statistic, the value of the test statistic under the null and alternative hypotheses, the intracluster correlation (ICC), the number of clusters in the control group (K_2) and the sample size per cluster in the control group (M_2).

The formulae use the normal approximation to calculate power. The formulae use the difference between the proportions under the null (Δ) and alternative (δ) hypotheses regardless of the format of statistic used. Proportions and ratios are converted to the relevant differences.

The formula for the power is given by the following equation:

$$Power = \Phi\left(\left[\frac{\delta - \Delta}{\sigma_{Unpool}}\right] - z_{1-\alpha/2}\left[\frac{\sigma_{Test\ Type}}{\sigma_{Unpool}}\right]\right) + \Phi\left(\left[\frac{-(\delta - \Delta)}{\sigma_{Unpool}}\right] - z_{1-\alpha/2}\left[\frac{\sigma_{Test\ Type}}{\sigma_{Unpool}}\right]\right) \quad \{5.2.1\}$$

where $\sigma_{Test\ Type}$ is the standard error defined by the test type being used and σ_{Unpool} is the unpooled standard error. The formula for the three standard error statistics is as follows:

1. Farrington and Manning Test Statistic (Likelihood Score)
   The Farrington and Manning test uses the constrained maximum likelihood estimator of the two proportions to calculate the standard error for the calculation of power and was proposed as method to test against a null hypothesis of a specified difference (Δ).

The standard error used by the Farrington and Manning test statistic is defined as follows:

$$\sigma_{FM} = \sqrt{\frac{p_{FM1}(1-p_{FM1})IF_1}{K_1M_1} + \frac{p_{FM2}(1-p_{FM2})IF_2}{K_2M_2}} \qquad \{5.2.2\}$$

where K_1 is the number of clusters in the treatment group, M_1 is the sample size per cluster in the treatment group, p_{FMi} is the maximum likelihood estimator for each group proportion and IF_i is the inflation factor for the effect of clustering in the treatment and control groups respectively. IF_i is defined as follows:

$$IF_i = \left(1 + \left(ICC(M_i - 1)\right)\right) \qquad i = 1,2 \qquad \{5.2.3\}$$

The constrained maximum likelihood (where $\tilde{p}_1 - \tilde{p}_2 = \Delta$) estimator of the two proportions is calculated using the following calculations:

$$t = \frac{(K_1M_1/IF_1)}{(K_2M_2/IF_2)} \qquad \{5.2.4\}$$

$$a = 1 + t \qquad \{5.2.5\}$$

$$b = -\left(1 + t + p_1 + tp_2 + \left(\Delta(t+2)\right)\right) \qquad \{5.2.6\}$$

$$c = \Delta^2 + \left(\Delta(2p_1 + t + 1)\right) + p_1 + tp_2 \qquad \{5.2.7\}$$

$$d = -p_1\Delta(1+\Delta) \qquad \{5.2.8\}$$

$$v = \left(\frac{b}{3a}\right)^3 - \left(\frac{bc}{6a^2}\right) + \left(\frac{d}{2a}\right) \qquad \{5.2.9\}$$

$$u = sign(v)\sqrt{\left(\frac{b}{3a}\right)^2 - \frac{c}{3a}} \qquad \{5.2.10\}$$

$$w = \frac{\left(\pi + \cos^{-1}\left(\frac{v}{u^3}\right)\right)}{3} \qquad \{5.2.11\}$$

$$p_{FM1} = 2u\cos(w) - b/3a \qquad \{5.2.12\}$$

$$p_{FM2} = p_{FM1} - \Delta \qquad \{5.2.13\}$$

2. Unpooled Test Statistic
   This test statistic uses the estimated group proportions to calculate the standard error. Its formula is as follows:

$$\sigma_{Unpool} = \sqrt{\frac{p_1(1-p_1)IF_1}{K_1M_1} + \frac{p_2(1-p_2)IF_2}{K_2M_2}} \qquad \{5.2.14\}$$

   where p_1 and p_2 are the estimated mean proportions for the two groups.

3. Pooled Test Statistic
   The pooled test statistic uses the weighted average of the two proportions to calculate the standard error. Its formula is as follows:

$$\sigma_{Pool} = \sqrt{\frac{\bar{p}(1-\bar{p})IF_1}{K_1M_1} + \frac{\bar{p}(1-\bar{p})IF_2}{K_2M_2}} \qquad \{5.2.15\}$$

   where

$$\bar{p} = \frac{K_1M_1p_1 + K_2M_2p_2}{K_1M_1 + K_2M_2} \qquad \{5.2.16\}$$

For the one-sided test, $z_{1-\alpha/2}$ is replaced with $z_{1-\alpha}$ and the following equations are used depending on the values of δ and Δ:

$$Power = \begin{cases} \Phi\left(\left[\frac{\delta-\Delta}{\sigma_{Unpool}}\right] - z_{1-\alpha}\left[\frac{\sigma_{Test\ Type}}{\sigma_{Unpool}}\right]\right), & \Delta < \delta \\ \Phi\left(\left[\frac{-(\delta-\Delta)}{\sigma_{Unpool}}\right] - z_{1-\alpha}\left[\frac{\sigma_{Test\ Type}}{\sigma_{Unpool}}\right]\right), & \Delta > \delta \end{cases} \qquad \{5.2.17\}$$

A closed form equation is not used to calculate the other parameters. Instead a search algorithm is used. The search algorithm calculates power at various values for the relevant parameter until the desired power is reached.

5.2.3 Examples

Example 1: Validation example calculating required number of clusters for a given power
The following examples are taken from Donner and Klar (2000) where a number of clusters calculation problem is conducted and then a sensitivity analysis is conducted to show the effect of changing control group number of clusters. The following steps outline the procedure for this example:

1. Open nQuery + nTerim 4.0 via the start menu or desktop shortcut. Click “New Fixed Term Test” from the top of the window. The “Study Goal and Design” window will appear as Figure 5.2.1 below.

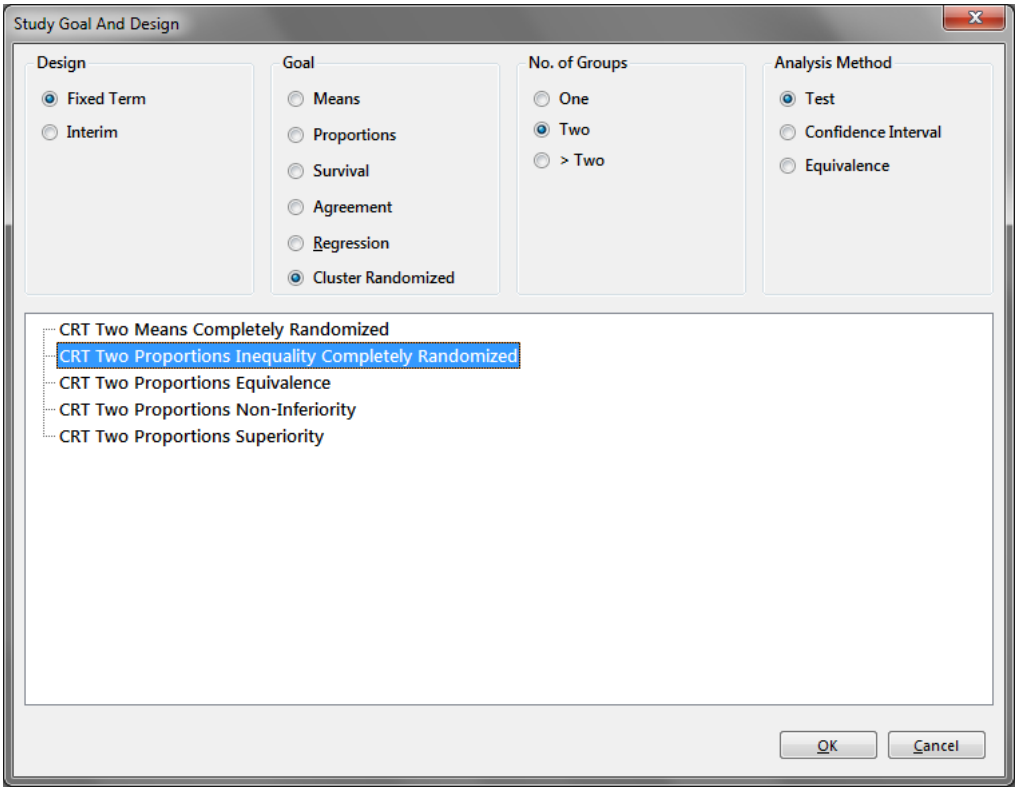


Figure 5.2.1. Study Goal and Design Window

2. In order to select the CRT Two Proportions Completely Randomized design table, navigate through the Study Goal and Design Window by selecting "Fixed Term" as the Design, "Cluster Randomized" as the Goal, "Two" as the Number of Groups and "Test" as the Analysis Method. Then click "OK" and the test window will appear as per Figure 5.2.2.

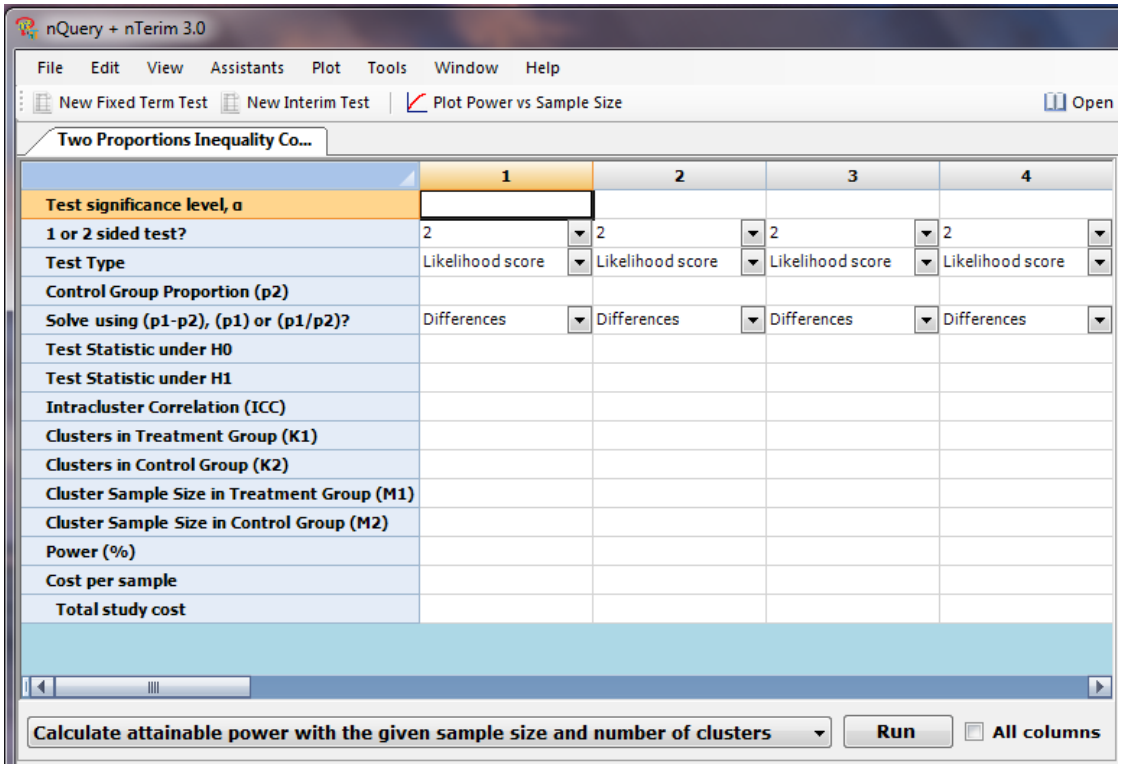


Figure 5.2.2. CRT Two Proportions Inequality Completely Randomized Test Table

The first calculation will be for the "Clusters in Treatment Group (K1)". This example is can be undertaken using the following steps.

3. First enter 0.05 for the Test Significance level row and select 2 for the "1 or 2 sided test?" variable dropdown menu, select "Unpooled" from the Test Type option and enter a control group proportion of 0.06.

4. Next, select "Proportions" from the "Solve Using" row, enter 0.06 and 0.04 for the value of the test statistic under H0 and H1 respectively, enter 0.01 for the intracluster correlation, enter 100 for both cluster sample size variables and enter 80 for Power.

5. Finally, enter K1 in the "Clusters in Control Group (K2)" row. This will solve so that K1 and K2 must be equal. Other ratios between K1 and K2 can be calculated by using arguments for K2 such as 2*K1 or K1*2 to have K2 be twice as large as K1 for example. The table will appear as in Figure 5.2.3.

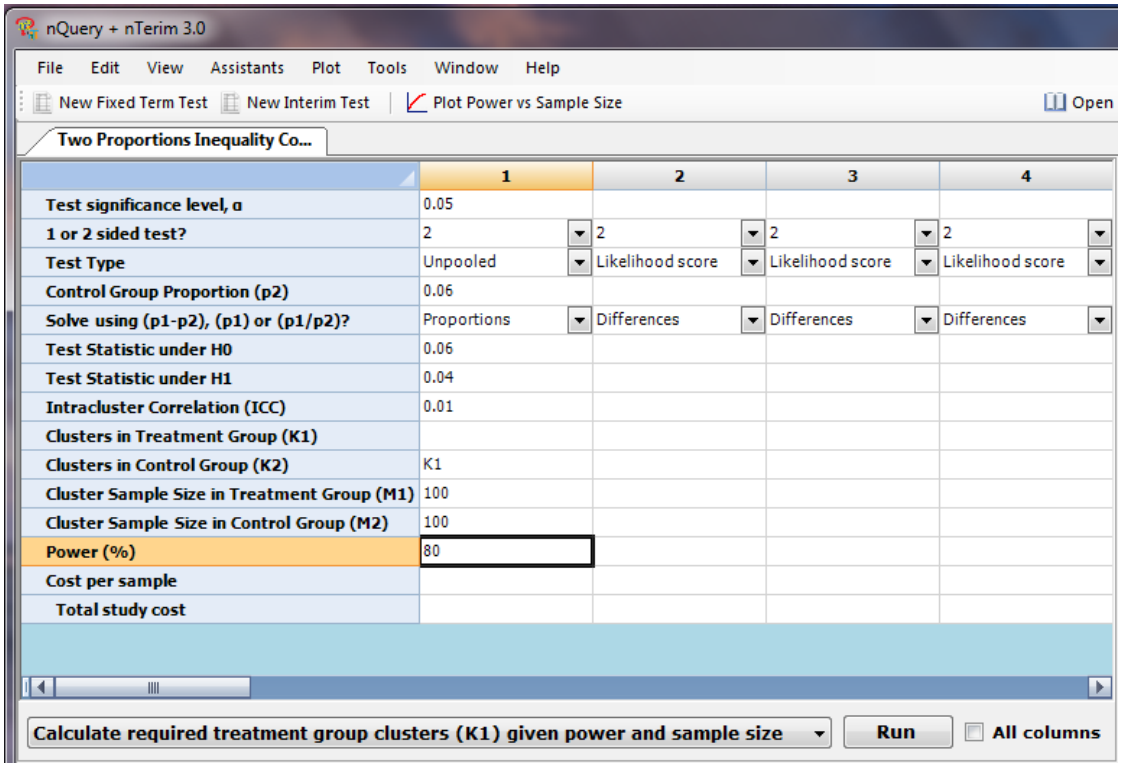


Figure 5.2.3. Values entered for CRT Two Proportions Completely Randomized study design

6. Select “Calculate required treatment group clusters (K1) given power and sample size” from the dropdown menu beside the “Run” button. Then click “Run”. This will give a result of 38 for K1 and K2 as in Figure 5.2.4.

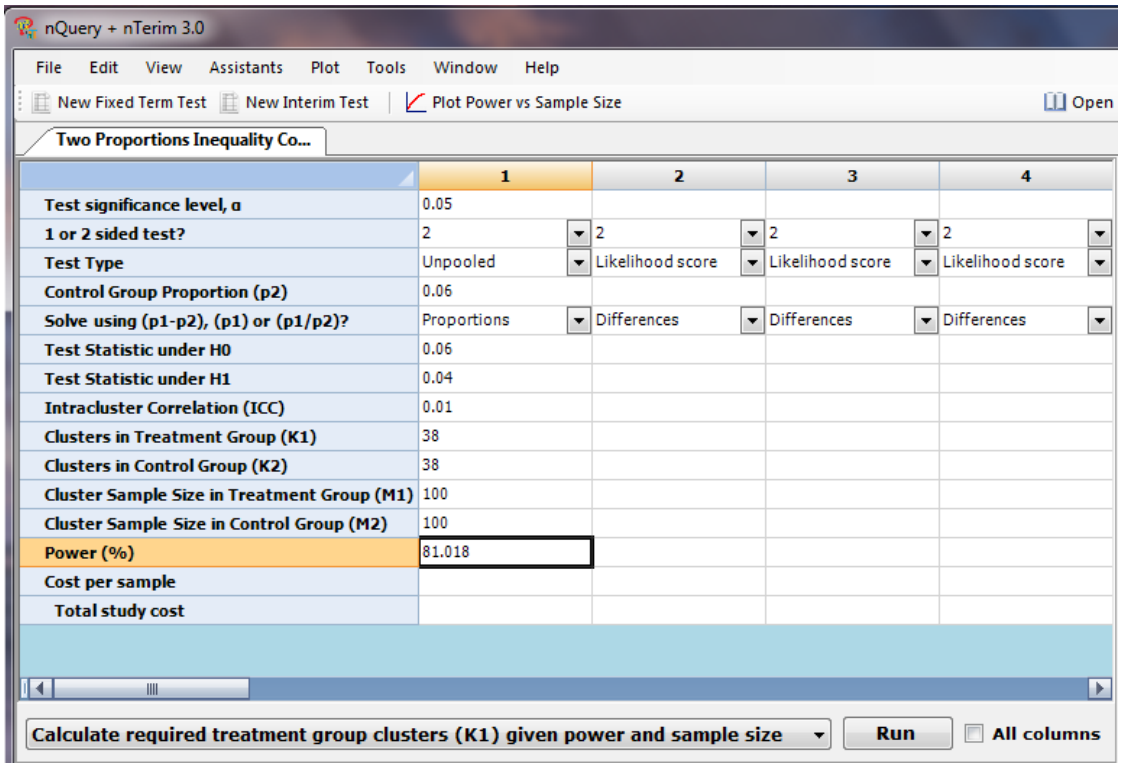
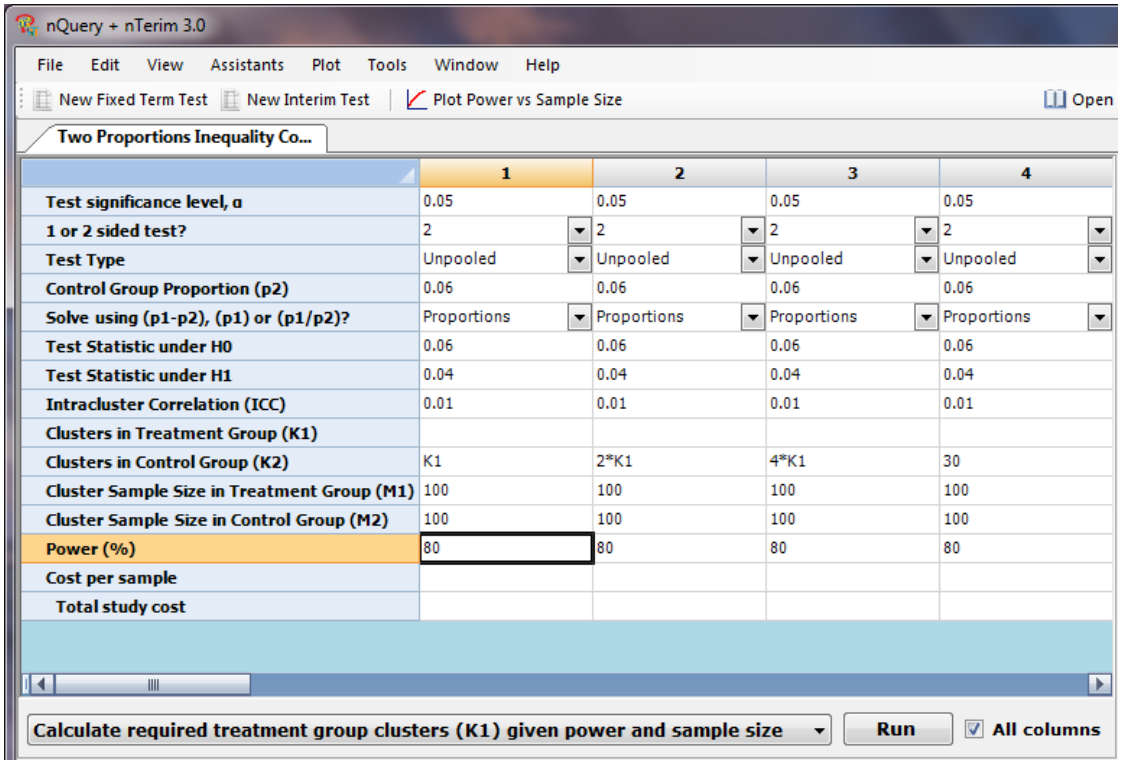


Figure 5.2.4. Completed CRT Two Proportions Completely Randomized study design

The next calculation is a sensitivity analysis for the treatment group number of clusters when the control group number of clusters is changed.

7. Delete the values for K1 and K2 in the first column then replace the updated power with 80. Enter “K1” in the control group clusters row. Then copy the first column into columns 2 to 4.

8. Other ratios, other than K1 being equal to K2, between K1 and K2 can be calculated by using arguments for K2, such as 2*K1 or K1*2 to have K2 be twice as large as K1 for example. Enter 2*K1 in column 2 for “Clusters in Control Group (K2)”, 4*K1 in column 3 and 30 in column four. This will give a table as per Figure 5.2.5.

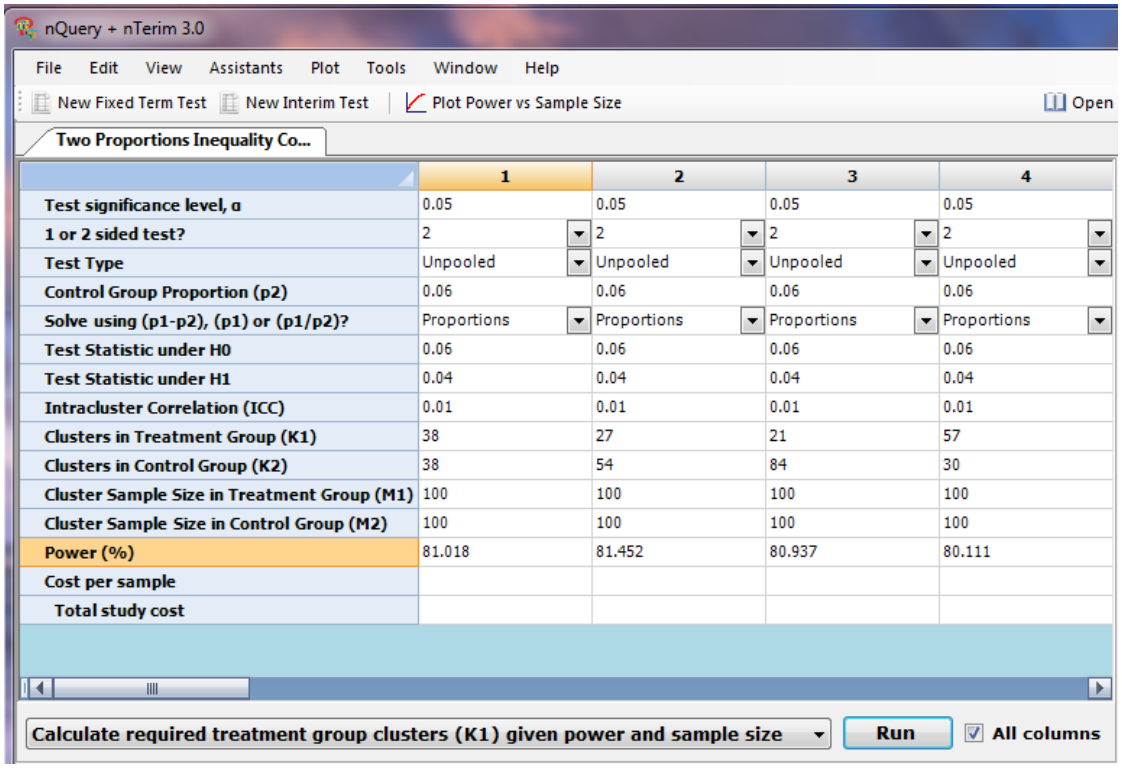


	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
1 or 2 sided test?	2	2	2	2
Test Type	Unpooled	Unpooled	Unpooled	Unpooled
Control Group Proportion (p2)	0.06	0.06	0.06	0.06
Solve using (p1-p2), (p1) or (p1/p2)?	Proportions	Proportions	Proportions	Proportions
Test Statistic under H0	0.06	0.06	0.06	0.06
Test Statistic under H1	0.04	0.04	0.04	0.04
Intracluster Correlation (ICC)	0.01	0.01	0.01	0.01
Clusters in Treatment Group (K1)				
Clusters in Control Group (K2)	K1	2*K1	4*K1	30
Cluster Sample Size in Treatment Group (M1)	100	100	100	100
Cluster Sample Size in Control Group (M2)	100	100	100	100
Power (%)	80	80	80	80
Cost per sample				
Total study cost				

Figure 5.2.5. Sensitivity analysis around the Control Group Number of Clusters

9. Select “Calculate required treatment group clusters (K1) given power and sample size” from the dropdown menu beside the “Run” button. Then tick the box to run “All Columns”. Then click “Run”.

This will give the resultant values of K1 of 38, 27, 21 and 57 sequentially, with the values of K2 updating automatically to reflect the desired ratio between K1 and K2, as in Figure 5.2.6.



	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
1 or 2 sided test?	2	2	2	2
Test Type	Unpooled	Unpooled	Unpooled	Unpooled
Control Group Proportion (p2)	0.06	0.06	0.06	0.06
Solve using (p1-p2), (p1) or (p1/p2)?	Proportions	Proportions	Proportions	Proportions
Test Statistic under H0	0.06	0.06	0.06	0.06
Test Statistic under H1	0.04	0.04	0.04	0.04
Intracluster Correlation (ICC)	0.01	0.01	0.01	0.01
Clusters in Treatment Group (K1)	38	27	21	57
Clusters in Control Group (K2)	38	54	84	30
Cluster Sample Size in Treatment Group (M1)	100	100	100	100
Cluster Sample Size in Control Group (M2)	100	100	100	100
Power (%)	81.018	81.452	80.937	80.111
Cost per sample				
Total study cost				

Figure 5.2.6. Results from Sensitivity analysis

The effect of changing these parameters on the relationship between Power and Sample size can be explored further using the “Plot Power vs Sample Size” button at the top of the screen.

10. Select the first column by clicking the 1 at the top of column 1. Then hold down Ctrl and click the 3 at the top of column 3 and the 4 at the top of column 4. All three columns will now be highlighted.

11. Click the “Plot Power vs Sample Size” button at the top of the screen. This will give you a Figure 5.2.7 which will show the relationship between power and **the number of clusters in the treatment group** for each column. Right-click to add features such as a legend to the graph and double click elements for user options and editing.

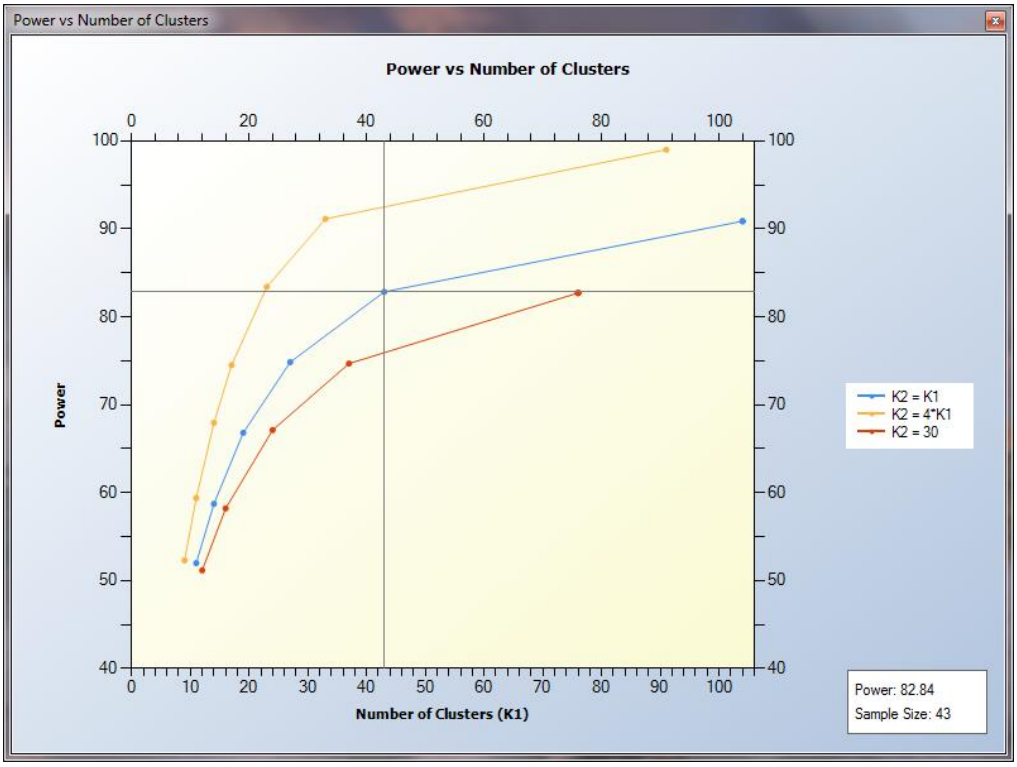


Figure 5.2.7. Power vs Number of Clusters plot

12. By clicking on the desired study design column and going to the "Output" tab at the bottom of the screen, you can get an output statement for the results given by the current calculation as per Figure 5.2.8.

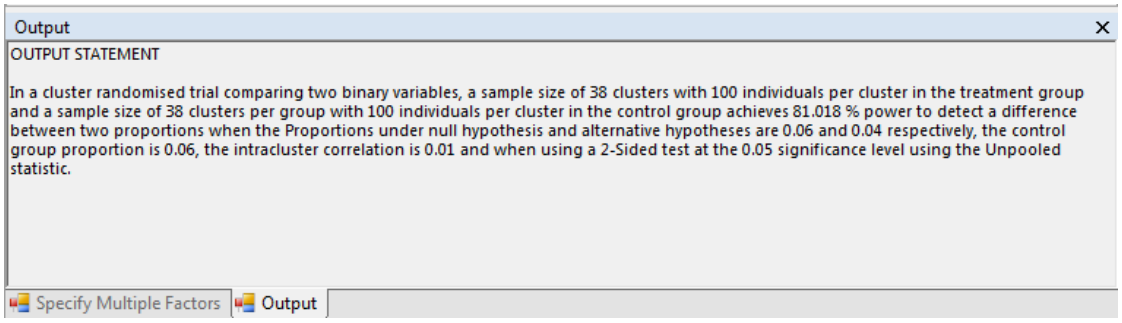


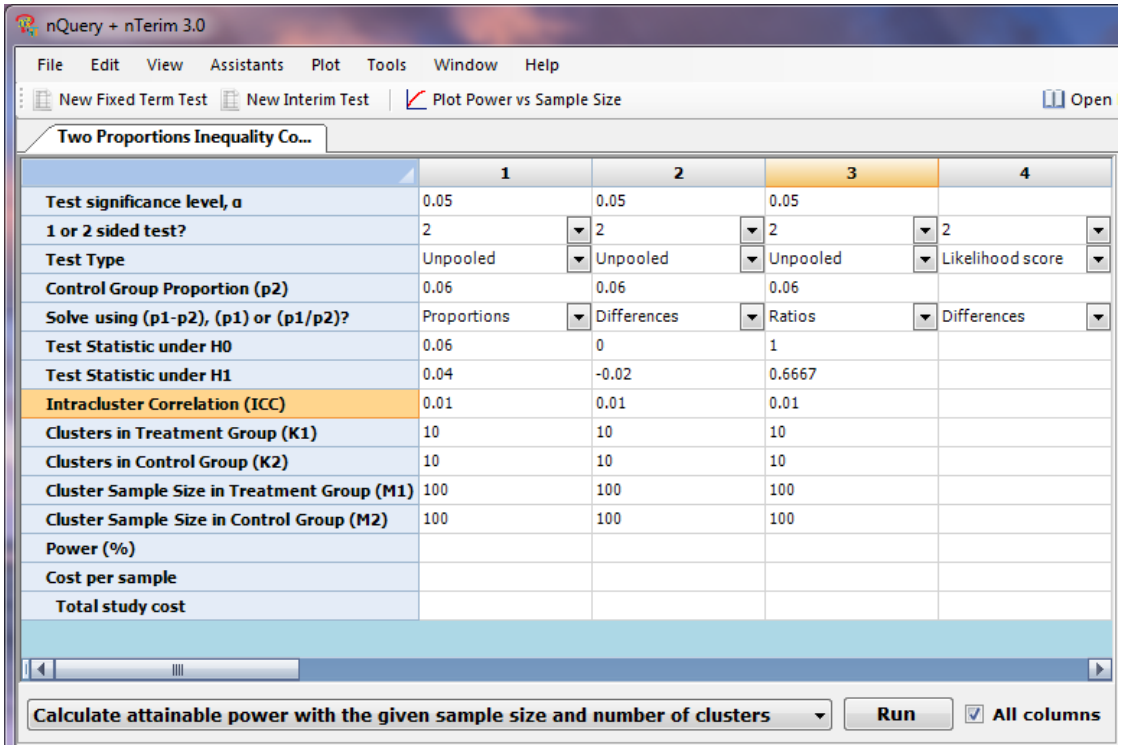
Figure 5.2.8. Study design Output statement

Example 2: Validation example calculating power using different Test Statistic formats
A calculation is conducted to show the equivalence of the test statistic approaches. The following steps outline the procedure for this example:

13. Delete columns 2 to 4 and delete the Power value in column 1. Then replace K1 and K2 with 10 in column 1. Next, in column 2, change the "Solve Using" switch to "Differences" and change it to "Ratios" in column 3. Copy and paste across the column 1 values to column 2 and 3 with the exception of the two "Test Statistic Under H0/H1" variables.

14. For column 2 ("Differences"), enter zero for "Test Statistic under H0" and -0.02 for "Test Statistic under H1".

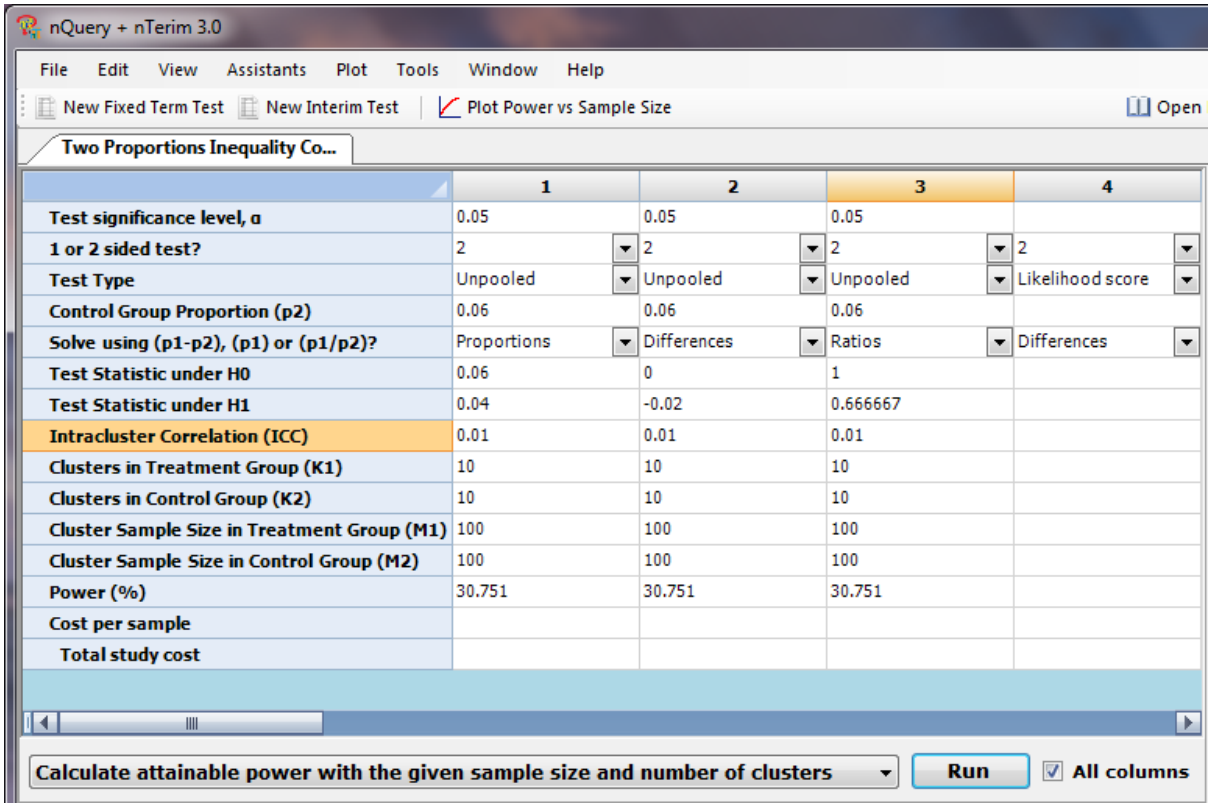
15. For column 3 ("Ratios"), enter 1 for "Test Statistic under H0" and 0.666667 for "Test Statistic under H1". **The values for column 2 and 3 are the values for these two formats which correspond to the values given for the first column using "Proportions".** This will give Figure 5.2.9.



	1	2	3	4
Test significance level, α	0.05	0.05	0.05	
1 or 2 sided test?	2	2	2	2
Test Type	Unpooled	Unpooled	Unpooled	Likelihood score
Control Group Proportion (p2)	0.06	0.06	0.06	
Solve using (p1-p2), (p1) or (p1/p2)?	Proportions	Differences	Ratios	Differences
Test Statistic under H0	0.06	0	1	
Test Statistic under H1	0.04	-0.02	0.6667	
Intracluster Correlation (ICC)	0.01	0.01	0.01	
Clusters in Treatment Group (K1)	10	10	10	
Clusters in Control Group (K2)	10	10	10	
Cluster Sample Size in Treatment Group (M1)	100	100	100	
Cluster Sample Size in Control Group (M2)	100	100	100	
Power (%)				
Cost per sample				
Total study cost				

Figure 5.2.9. Sensitivity analysis using different Test Statistics

16. Select "Calculate attainable power with the given sample size and number of clusters" from the dropdown menu beside the "Run" button. Then tick the box for "All Columns". Then click "Run". This will give 30.751% as per Figure 5.2.9.



	1	2	3	4
Test significance level, α	0.05	0.05	0.05	
1 or 2 sided test?	2	2	2	2
Test Type	Unpooled	Unpooled	Unpooled	Likelihood score
Control Group Proportion (p2)	0.06	0.06	0.06	
Solve using (p1-p2), (p1) or (p1/p2)?	Proportions	Differences	Ratios	Differences
Test Statistic under H0	0.06	0	1	
Test Statistic under H1	0.04	-0.02	0.666667	
Intracluster Correlation (ICC)	0.01	0.01	0.01	
Clusters in Treatment Group (K1)	10	10	10	
Clusters in Control Group (K2)	10	10	10	
Cluster Sample Size in Treatment Group (M1)	100	100	100	
Cluster Sample Size in Control Group (M2)	100	100	100	
Power (%)	30.751	30.751	30.751	
Cost per sample				
Total study cost				

Figure 5.2.9. Results from Sensitivity Analysis

5.3. CRT Two Proportions Equivalence – Completely Randomized

5.3.1 Introduction

Binary data is commonly studied in variety of different fields. Equivalence trials are commonly used to assess whether a treatment is equivalent to another treatment in the clinical setting e.g. comparing two competitor drugs. Clustered data is very common in a wide variety of academic, social policy and economic studies. This two sample test is used to test hypotheses about the equivalence of two treatments in a completely randomized cluster randomized trial.

This table facilitates the calculation of the power and sample size for equivalence hypothesis tests comparing proportions in a completely randomized cluster randomized trial. Power and sample size is computed using the method outlined by Donner and Klar (2000).

5.3.2 Methodology

This table provides sample size and power calculations for studies which will be conducting an equivalence trial between proportions in trials which use a completely randomized cluster randomization study design. Equivalence trials are those in which the researcher is testing that two treatments have an equivalent effect. A completely randomized design assigns clusters randomly to control and treatment groups.

The sample size calculation for cluster randomized proportions is taken from Donner and Klar (2000) and is an extension of the methods used for equivalence trials in a fully randomized trial, as outlined in Chow et al. (2008). The extension uses the intracluster correlation (ICC), which is the ordinary product moment correlation between any two observations in the same cluster, to adjust for the effect of within cluster correlation. It is assumed the ICC is the same in both groups.

This table can be used to calculate the power, the number of clusters in the treatment group, the sample size per cluster in the treatment group, the intracluster correlation and the smallest detectable difference given all other terms in the table are specified.

To calculate power, the number of clusters in the treatment group or the sample size per cluster in the treatment group, the user must specify the test significance level (α), the control group proportion (p_2), which test type is being used, the format of the test statistic, the upper equivalence margin, the lower equivalence margin, the actual expected difference (δ), the intracluster correlation (ICC), the number of clusters in the control group (K_2) and the sample size per cluster in the control group (M_2).

The formula uses the normal approximation to calculate power. The formulae use the difference between the proportions under the equivalence margins ($\Delta_{u/l}$) and the actual expected difference (δ) regardless of the format of statistic used. Proportions and ratios are converted to the relevant differences.

The equivalence trial is assumed to be using the two one-sided test (TOST) methodology for analysis. Thus, the formula for the power is given by the following equation:

$$Power = \Phi\left(\left[\frac{\delta - \Delta_l}{\sigma_{Unpool}}\right] - z_{1-\alpha}\left[\frac{\sigma_{Test\ Type}}{\sigma_{Unpool}}\right]\right) + \Phi\left(\left[\frac{-(\delta - \Delta_u)}{\sigma_{Unpool}}\right] - z_{1-\alpha}\left[\frac{\sigma_{Test\ Type}}{\sigma_{Unpool}}\right]\right) \quad \{5.3.1\}$$

where $\sigma_{Test\ Type}$ is the standard error defined by the test type being used and σ_{Unpool} is the unpooled standard error. The formula for the three standard error statistics is as follows:

1. Farrington and Manning Test Statistic (Likelihood Score)
The Farrington and Manning test uses the constrained maximum likelihood estimator of the two proportions to calculate the standard error for the calculation of power and was proposed as method to test against a null hypothesis of a specified difference (Δ).

The standard error used by the Farrington and Manning test statistic is defined as follows:

$$\sigma_{FM} = \sqrt{\frac{p_{FM1}(1-p_{FM1})IF_1}{K_1M_1} + \frac{p_{FM2}(1-p_{FM2})IF_2}{K_2M_2}} \qquad \{5.3.2\}$$

where K_1 is the number of clusters in the treatment group, M_1 is the sample size per cluster in the treatment group, p_{FMi} is the maximum likelihood estimator for each group proportion and IF_i is the inflation factor for the effect of clustering in the treatment and control groups respectively. IF_i is defined as follows:

$$IF_i = \left(1 + \left(ICC(M_i - 1)\right)\right) \qquad i = 1,2 \qquad \{5.3.3\}$$

The constrained maximum likelihood (where $\tilde{p}_1 - \tilde{p}_2 = \Delta$) estimator of the two proportions is calculated using the following calculations:

$$t = \frac{(K_1M_1/IF_1)}{(K_2M_2/IF_2)} \qquad \{5.3.4\}$$

$$a = 1 + t \qquad \{5.3.5\}$$

$$b = -\left(1 + t + p_1 + tp_2 + \left(\Delta(t+2)\right)\right) \qquad \{5.3.6\}$$

$$c = \Delta^2 + \left(\Delta(2p_1 + t + 1)\right) + p_1 + tp_2 \qquad \{5.3.7\}$$

$$d = -p_1\Delta(1+\Delta) \qquad \{5.3.8\}$$

$$v = \left(\frac{b}{3a}\right)^3 - \left(\frac{bc}{6a^2}\right) + \left(\frac{d}{2a}\right) \qquad \{5.3.9\}$$

$$u = sign(v)\sqrt{\left(\frac{b}{3a}\right)^2 - \frac{c}{3a}} \qquad \{5.3.10\}$$

$$w = \frac{\left(\pi + \cos^{-1}\left(\frac{v}{u^3}\right)\right)}{3} \qquad \{5.3.11\}$$

$$p_{FM1} = 2u\cos(w) - b/3a \qquad \{5.3.12\}$$

$$p_{FM2} = p_{FM1} - \Delta \qquad \{5.3.13\}$$

For equivalence trials, it is necessary to calculate the constrained maximum likelihood estimators for both equivalence margins separately if their absolute values are different.

2. Unpooled Test Statistic
   This test statistic uses the estimated group proportions to calculate the standard error. Its formula is as follows:

$$\sigma_{Unpool} = \sqrt{\frac{p_1(1-p_1)IF_1}{K_1M_1} + \frac{p_2(1-p_2)IF_2}{K_2M_2}} \qquad \{5.3.14\}$$

   where p_1 and p_2 are the estimated mean proportions for the two groups.

3. Pooled Test Statistic
   The pooled test statistic uses the weighted average of the two proportions to calculate the standard error. Its formula is as follows:

$$\sigma_{Pool} = \sqrt{\frac{\bar{p}(1-\bar{p})IF_1}{K_1M_1} + \frac{\bar{p}(1-\bar{p})IF_2}{K_2M_2}} \qquad \{5.3.15\}$$

   where

$$\bar{p} = \frac{K_1M_1p_1 + K_2M_2p_2}{K_1M_1 + K_2M_2} \qquad \{5.3.16\}$$

A closed form equation is not used to calculate the other parameters. Instead a search algorithm is used. The search algorithm calculates power at various values for the relevant parameter until the desired power is reached.

5.3.3 Examples

Example 1: Validation example calculating required number of clusters for a given power
The following example will look at a number of clusters calculation problem is conducted and then a sensitivity analysis is conducted to show the effect of changing control group number of clusters. The following steps outline the procedure for this example:

1. Open nQuery + nTerim 4.0 via the start menu or desktop shortcut. Click “New Fixed Term Test” from the top of the window. The “Study Goal and Design” window will appear as Figure 5.3.1 below.

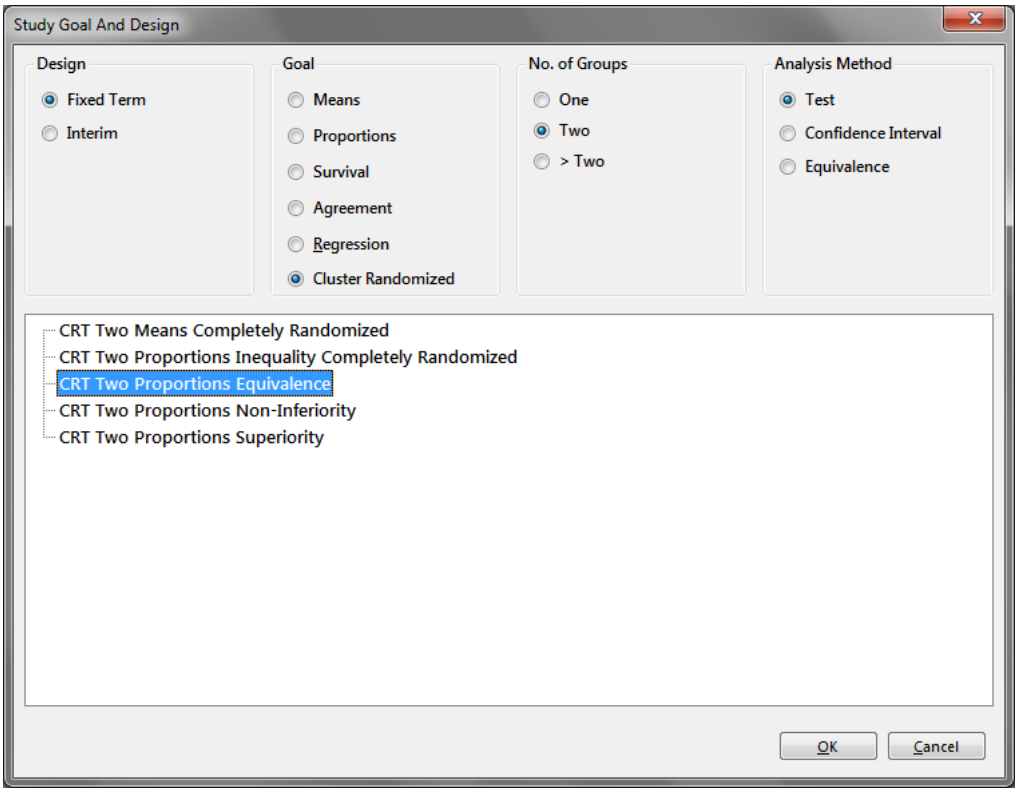


Figure 5.3.1. Study Goal and Design Window

2. In order to select the CRT Two Proportions Equivalence Completely Randomized design table, navigate through the Study Goal and Design Window by selecting "Fixed Term" as the Design, "Cluster Randomized" as the Goal, "Two" as the Number of Groups and "Test" as the Analysis Method. Then click "OK" and the test window will appear as per Figure 5.3.2.

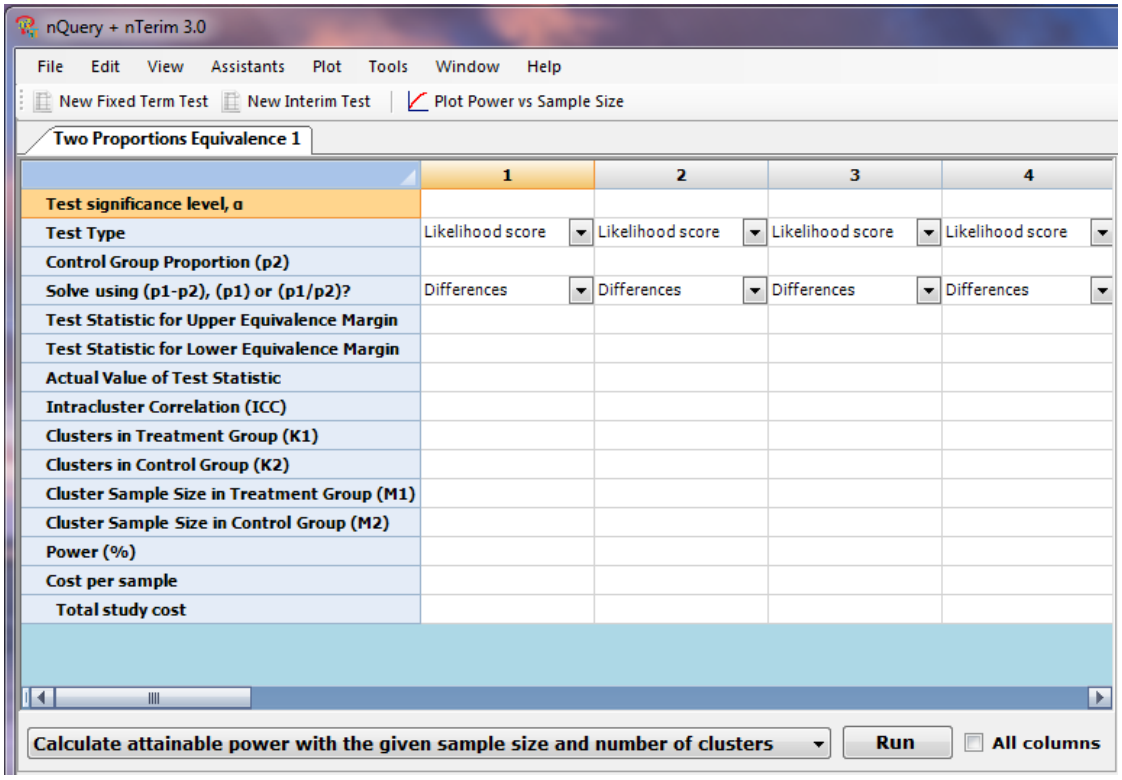


Figure 5.3.2. CRT Two Proportions Equivalence Completely Randomized Test Table

The first calculation will be for the "Clusters in Treatment Group (K1)". This example is can be undertaken using the following steps.

3. First enter 0.05 for the Test Significance level row. Next, select "Pooled" from the Test Type option, and enter a control group proportion of 0.5, then select "Differences" from the "Solve Using" dropdown option.

4. Enter 0.1 for the "Upper Equivalence Margin", -0.1 for the "Lower Equivalence Margin" and 0 for the "Actual Value of Test Statistic". Enter 0.001 for the intracluster correlation. Enter 100 for both cluster sample size variables and 80 for power.

5. Finally, enter K1 in the "Clusters in Control Group (K2)" row. This will solve so that K1 and K2 must be equal. Other ratios between K1 and K2 can be calculated by using arguments for K2 such as 2*K1 or K1*2 to have K2 be twice as large as K1 for example. The table will appear as in Figure 5.3.3.

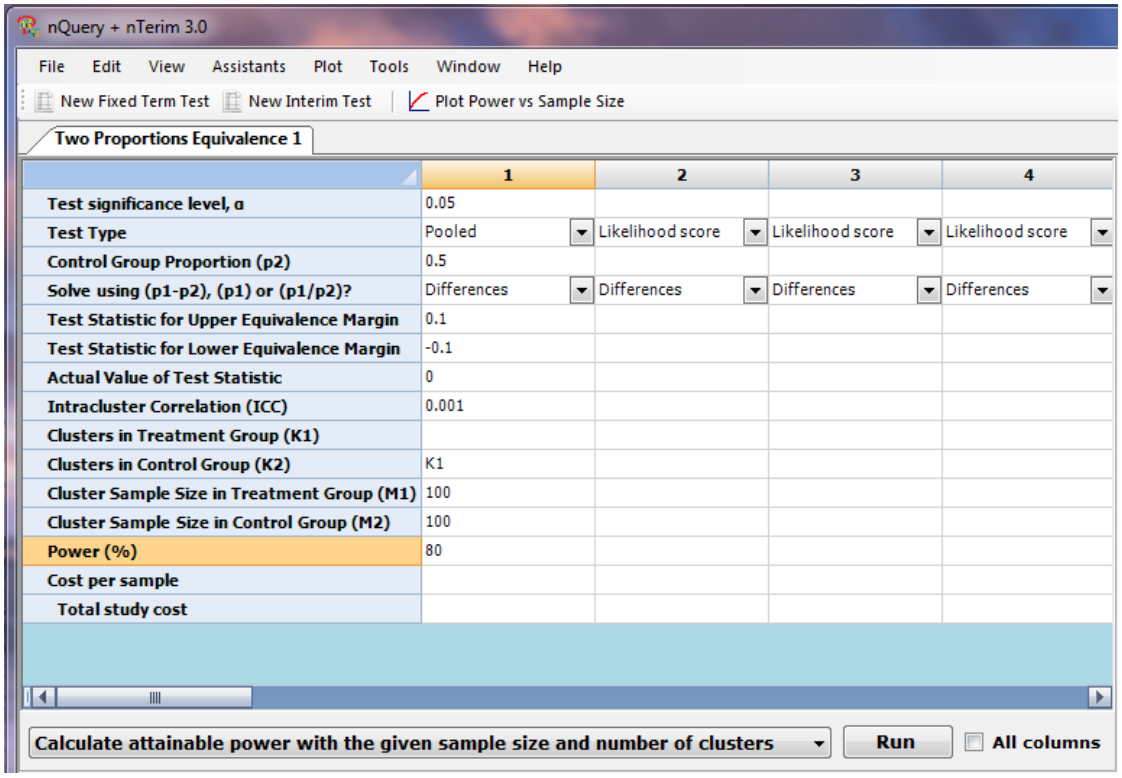


Figure 5.3.3. Values entered for CRT Two Proportions Equivalence study design

6. Select "Calculate required treatment group clusters (K1) given power and sample size" from the dropdown menu beside the "Run" button. Then click "Run". This will give a result of 5 for K1 and K2 as in Figure 5.3.4.

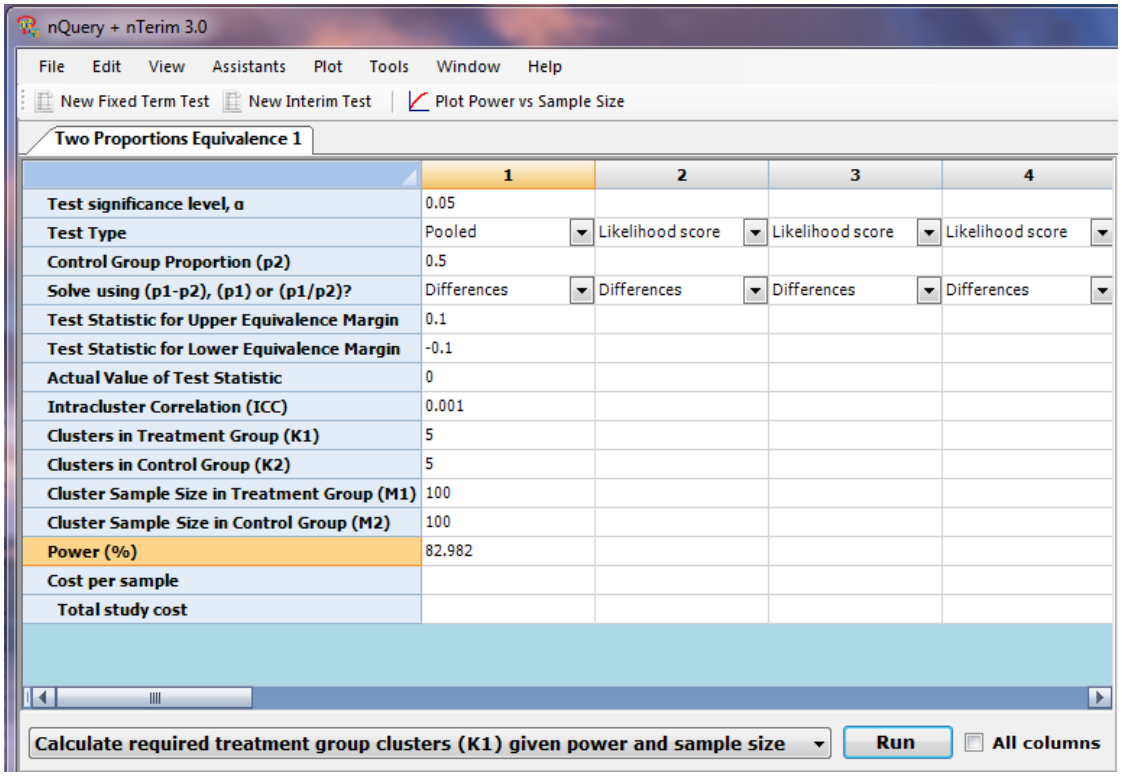
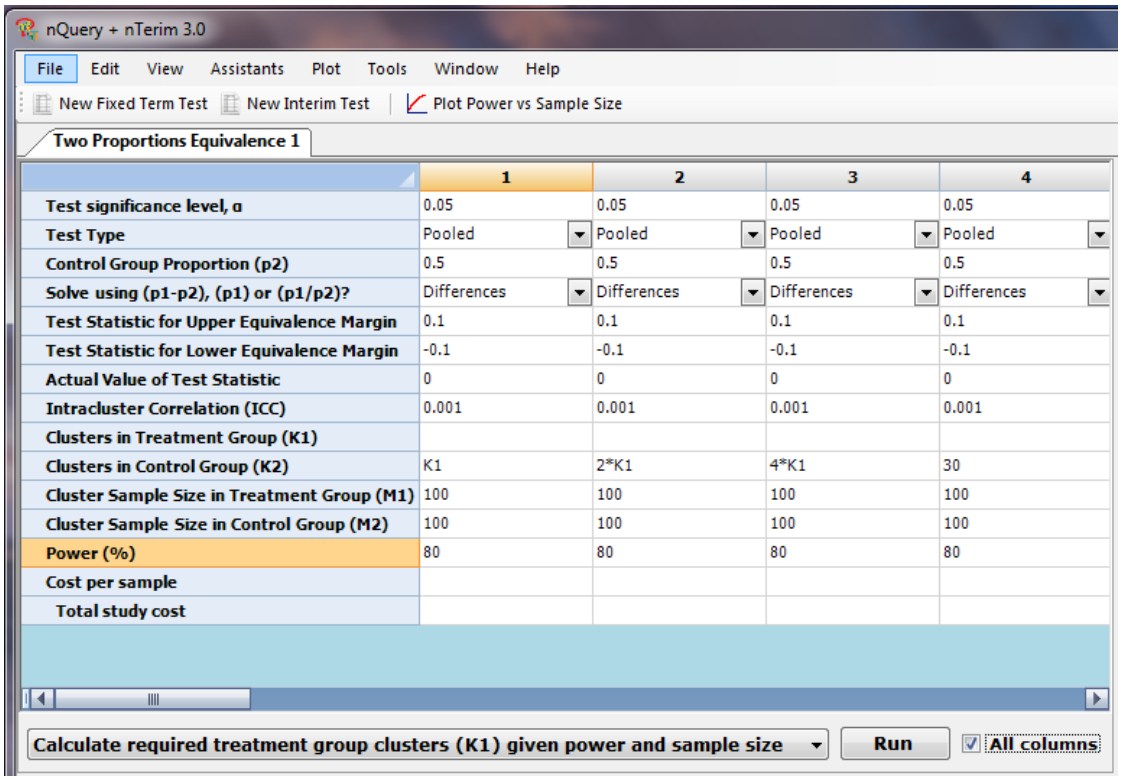


Figure 5.3.4. Completed CRT Two Proportions Equivalence study design

The next calculation is a sensitivity analysis for the treatment group number of clusters when the control group number of clusters is changed.

7. Delete the values for K1 and K2 in the first column then replace the updated power with 80 and enter “K1” in the control group clusters row. Then copy the first column into columns 2 to 4.

8. Enter K1*2 in column 2 for “Clusters in Control Group (K2)”, 4*K1 in column 3 and 30 in column four. This will give a table as per Figure 5.3.5.

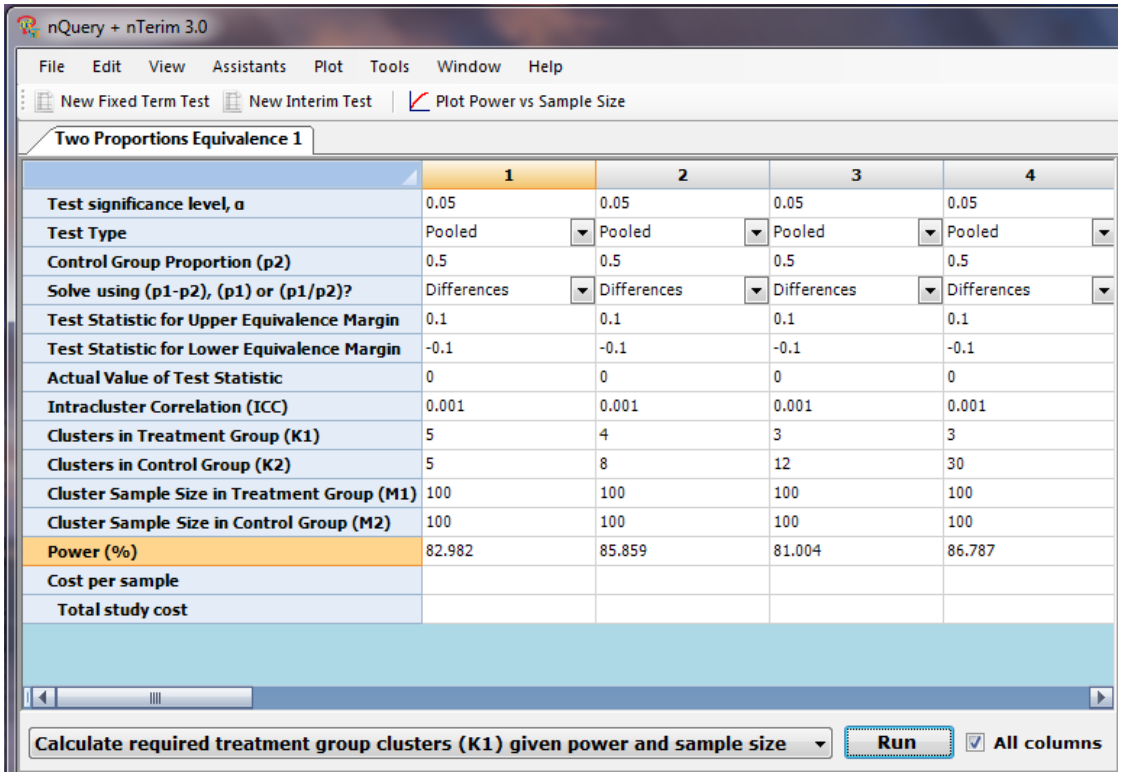


	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
Test Type	Pooled	Pooled	Pooled	Pooled
Control Group Proportion (p2)	0.5	0.5	0.5	0.5
Solve using (p1-p2), (p1) or (p1/p2)?	Differences	Differences	Differences	Differences
Test Statistic for Upper Equivalence Margin	0.1	0.1	0.1	0.1
Test Statistic for Lower Equivalence Margin	-0.1	-0.1	-0.1	-0.1
Actual Value of Test Statistic	0	0	0	0
Intracluster Correlation (ICC)	0.001	0.001	0.001	0.001
Clusters in Treatment Group (K1)				
Clusters in Control Group (K2)	K1	2*K1	4*K1	30
Cluster Sample Size in Treatment Group (M1)	100	100	100	100
Cluster Sample Size in Control Group (M2)	100	100	100	100
Power (%)	80	80	80	80
Cost per sample				
Total study cost				

Figure 5.3.5. Sensitivity analysis around the Control Group Number of Clusters

9. Select “Calculate required treatment group clusters (K1) given power and sample size” from the dropdown menu beside the “Run” button. Then tick the box to run “All Columns”. Then click “Run”.

This will give the resultant values of K1 of 5, 4, 3 and 3 sequentially, with the values of K2 updating automatically to reflect the desired ratio between K1 and K2, as in Figure 5.3.6.



	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
Test Type	Pooled	Pooled	Pooled	Pooled
Control Group Proportion (p2)	0.5	0.5	0.5	0.5
Solve using (p1-p2), (p1) or (p1/p2)?	Differences	Differences	Differences	Differences
Test Statistic for Upper Equivalence Margin	0.1	0.1	0.1	0.1
Test Statistic for Lower Equivalence Margin	-0.1	-0.1	-0.1	-0.1
Actual Value of Test Statistic	0	0	0	0
Intracluster Correlation (ICC)	0.001	0.001	0.001	0.001
Clusters in Treatment Group (K1)	5	4	3	3
Clusters in Control Group (K2)	5	8	12	30
Cluster Sample Size in Treatment Group (M1)	100	100	100	100
Cluster Sample Size in Control Group (M2)	100	100	100	100
Power (%)	82.982	85.859	81.004	86.787
Cost per sample				
Total study cost				

Figure 5.3.6. Results from Sensitivity analysis

10. By clicking on the desired study design column and going to the "Output" tab at the bottom of the screen, you can get an output statement for the results given by the current calculation as per Figure 5.3.7.

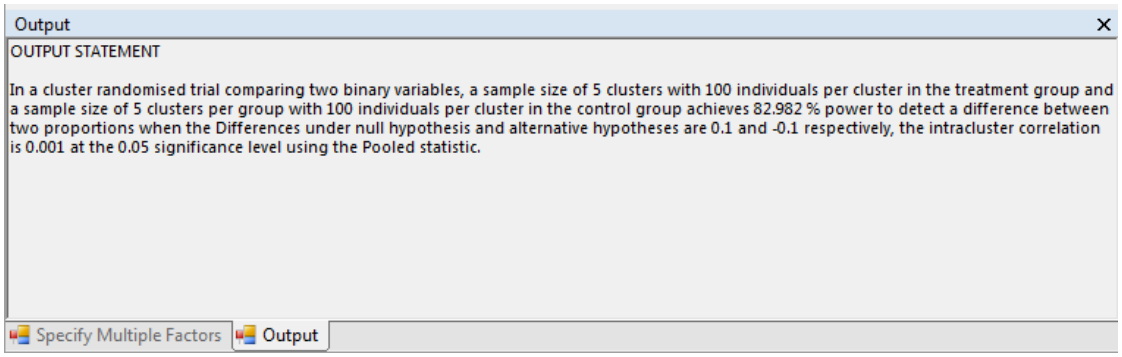


Figure 5.3.7. Study design Output statement

This Output statement can then be easily transferred directly from the output window into a report.

5.4. CRT Two Proportions Non-Inferiority – Completely Randomized

5.4.1 Introduction

Binary data is commonly studied in variety of different fields. Non-inferiority trials are commonly used to assess whether a new treatment is at least as effective as a pre-existing treatment in the clinical setting e.g. comparing a generic drug to its competitor. Clustered data is very common in a wide variety of academic, social policy and economic studies. This two sample test is used to test hypotheses about the non-inferiority of a treatment in a completely randomized cluster randomized trial.

This table facilitates the calculation of the power and sample size for non-inferiority hypothesis tests comparing proportions in a completely randomized cluster randomized trial. Power and sample size is computed using the method outlined by Donner and Klar (2000).

5.4.2 Methodology

This table provides sample size and power calculations for studies which will be conducting a non-inferiority trial between proportions in trials which use a completely randomized cluster randomization study design. Non-inferiority trials are those in which the researcher is testing that a treatment is no worse than the pre-existing control treatment. A completely randomized design assigns clusters randomly to control and treatment groups.

The sample size calculation for cluster randomized proportions is taken from Donner and Klar (2000) and is an extension of the methods used for non-inferiority trials in a fully randomized trial, as outlined in Chow et al. (2008). The extension uses the intracluster correlation (ICC), which is the ordinary product moment correlation between any two observations in the same cluster, to adjust for the effect of within cluster correlation. It is assumed the ICC is the same in both groups.

This table can be used to calculate the power, the number of clusters in the treatment group, the sample size per cluster in the treatment group, the intracluster correlation and the smallest detectable difference given all other terms in the table are specified.

To calculate power, the number of clusters in the treatment group or the sample size per cluster in the treatment group, the user must specify the test significance level (α), the control group proportion (p_2), which test type is being used, the format of the test statistic, whether higher values for the proportion are better or worse from the researcher's perspective, the non-inferiority margin, the actual expected difference, the intracluster correlation (ICC), the number of clusters in the control group (K_2) and the sample size per cluster in the control group (M_2).

The formula uses the normal approximation to calculate power. The formulae use the difference between the proportions under the non-inferiority margin (Δ) and the actual expected difference (δ) regardless of the format of statistic used. Proportions and ratios are converted to the relevant differences.

If higher values are considered better then the expected difference should be higher than the non-inferiority margin and if the lower values are considered better then the expected difference should be lower than the non-inferiority margin. The formulae for the power are thus given by the following equations:

$$Power = \begin{cases} \Phi\left(\left[\frac{\delta - \Delta}{\sigma_{Unpool}}\right] - z_{1-\alpha}\left[\frac{\sigma_{Test\ Type}}{\sigma_{Unpool}}\right]\right), & \Delta < \delta \\ \Phi\left(\left[\frac{-(\delta - \Delta)}{\sigma_{Unpool}}\right] - z_{1-\alpha}\left[\frac{\sigma_{Test\ Type}}{\sigma_{Unpool}}\right]\right), & \Delta > \delta \end{cases} \qquad \{5.4.1\}$$

where $\sigma_{Test\ Type}$ is the standard error defined by the test type being used and σ_{Unpool} is the unpooled standard error. The formula for the three standard error statistics is as follows:

1. Farrington and Manning Test Statistic (Likelihood Score)
The Farrington and Manning test uses the constrained maximum likelihood estimator of the two proportions to calculate the standard error for the calculation of power and was proposed as method to test against a null hypothesis of a specified difference (Δ).

The standard error used by the Farrington and Manning test statistic is defined as follows:

$$\sigma_{FM} = \sqrt{\frac{p_{FM1}(1-p_{FM1})IF_1}{K_1M_1} + \frac{p_{FM2}(1-p_{FM2})IF_2}{K_2M_2}} \qquad \{5.4.2\}$$

where K_1 is the number of clusters in the treatment group, M_1 is the sample size per cluster in the treatment group, p_{FMi} is the maximum likelihood estimator for each group proportion and IF_i is the inflation factor for the effect of clustering in the treatment and control groups respectively. IF_i is defined as follows:

$$IF_i = \left(1 + \left(ICC(M_i - 1)\right)\right) \qquad i = 1,2 \qquad \{5.4.3\}$$

The constrained maximum likelihood (where $\tilde{p}_1 - \tilde{p}_2 = \Delta$) estimator of the two proportions is calculated using the following calculations:

$$t = \frac{(K_1M_1/IF_1)}{(K_2M_2/IF_2)} \qquad \{5.4.4\}$$

$$a = 1 + t \qquad \{5.4.5\}$$

$$b = -\left(1 + t + p_1 + tp_2 + \left(\Delta(t+2)\right)\right) \qquad \{5.4.6\}$$

$$c = \Delta^2 + \left(\Delta(2p_1 + t + 1)\right) + p_1 + tp_2 \qquad \{5.4.7\}$$

$$d = -p_1\Delta(1+\Delta) \qquad \{5.4.8\}$$

$$v = \left(\frac{b}{3a}\right)^3 - \left(\frac{bc}{6a^2}\right) + \left(\frac{d}{2a}\right) \qquad \{5.4.9\}$$

$$u = sign(v)\sqrt{\left(\frac{b}{3a}\right)^2 - \frac{c}{3a}} \qquad \{5.4.10\}$$

$$w = \frac{\left(\pi + \cos^{-1}\left(\frac{v}{u^3}\right)\right)}{3} \qquad \{5.4.11\}$$

$$p_{FM1} = 2u\cos(w) - b/3a \qquad \{5.4.12\}$$

$$p_{FM2} = p_{FM1} - \Delta \qquad \{5.4.13\}$$

2. Unpooled Test Statistic
   This test statistic uses the estimated group proportions to calculate the standard error. Its formula is as follows:

$$\sigma_{Unpool} = \sqrt{\frac{p_1(1-p_1)IF_1}{K_1M_1} + \frac{p_2(1-p_2)IF_2}{K_2M_2}} \qquad \{5.4.14\}$$

   where p_1 and p_2 are the estimated mean proportions for the two groups.

3. Pooled Test Statistic
   The pooled test statistic uses the weighted average of the two proportions to calculate the standard error. Its formula is as follows:

$$\sigma_{Pool} = \sqrt{\frac{\bar{p}(1-\bar{p})IF_1}{K_1M_1} + \frac{\bar{p}(1-\bar{p})IF_2}{K_2M_2}} \qquad \{5.4.15\}$$

   where

$$\bar{p} = \frac{K_1M_1p_1 + K_2M_2p_2}{K_1M_1 + K_2M_2} \qquad \{5.4.16\}$$

A closed form equation is not used to calculate the other parameters. Instead a search algorithm is used. The search algorithm calculates power at various values for the relevant parameter until the desired power is reached.

5.4.3 Examples

Example 1: Validation example calculating required number of clusters for a given power
The following example will look at a number of clusters calculation problem is conducted and then a sensitivity analysis is conducted to show the effect of changing control group number of clusters. The following steps outline the procedure for this example:

1. Open nQuery + nTerim 4.0 via the start menu or desktop shortcut. Click “New Fixed Term Test” from the top of the window. The “Study Goal and Design” window will appear as Figure 5.4.1 below.

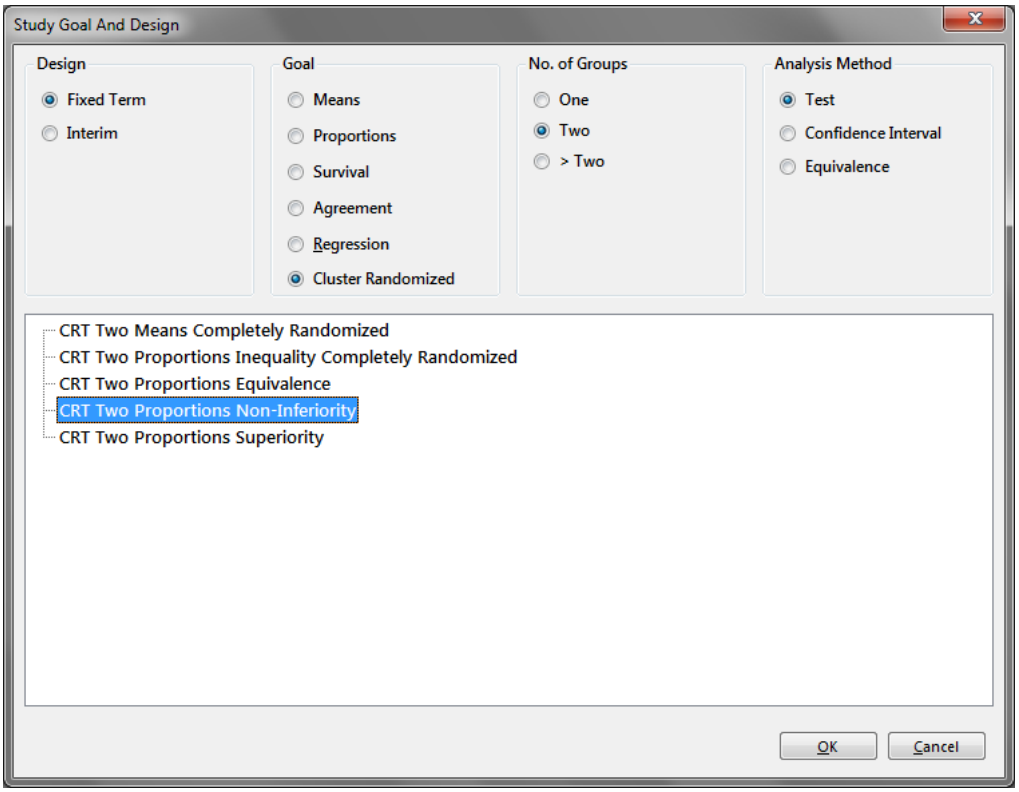


Figure 5.4.1. Study Goal and Design Window

2. In order to select the CRT Two Proportions Non-Inferiority Completely Randomized design table, navigate through the Study Goal and Design Window by selecting "Fixed Term" as the Design, "Cluster Randomized" as the Goal, "Two" as the Number of Groups and "Test" as the Analysis Method. Then click "OK" and the test window will appear as per Figure 5.4.2.

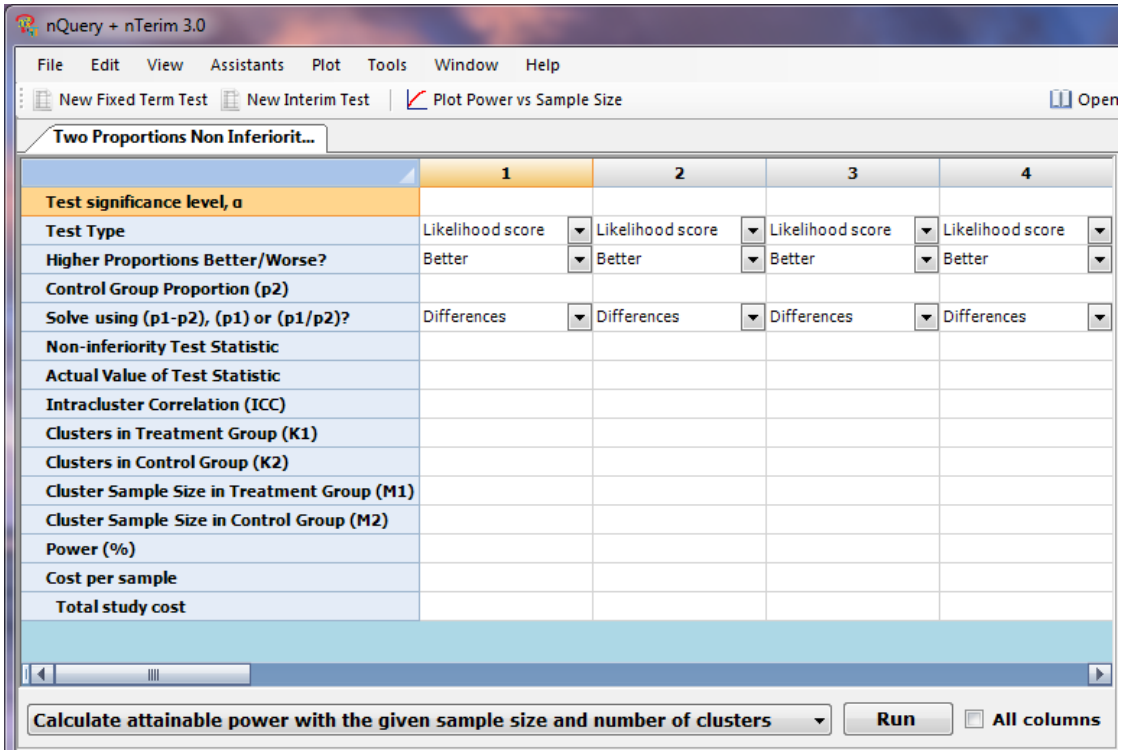


Figure 5.4.2. CRT Two Proportions Non-Inferiority Completely Randomized Test Table

The first calculation will be for the "Clusters in Treatment Group (K1)". This example is can be undertaken using the following steps.

3. First enter 0.05 for the Test Significance level row. Next, select "Unpooled" from the Test Type option, "Better" from the "Higher Proportions Better/Worse" option and enter a control group proportion of 0.5.

4. Next, select "Differences" from the "Solve Using" dropdown option and enter -0.1 for the "Non-Inferiority Test Statistic" and -0.05 for the "Actual Value of Test Statistic". Enter 0.001 for the intracluster correlation. Enter 100 for both cluster sample size variables and 80 for power.

5. Finally, enter K1 in the "Clusters in Control Group (K2)" row. This will solve so that K1 and K2 must be equal. Other ratios between K1 and K2 can be calculated by using arguments for K2 such as 2*K1 or K1*2 to have K2 be twice as large as K1 for example. The table will appear as in Figure 5.4.3.

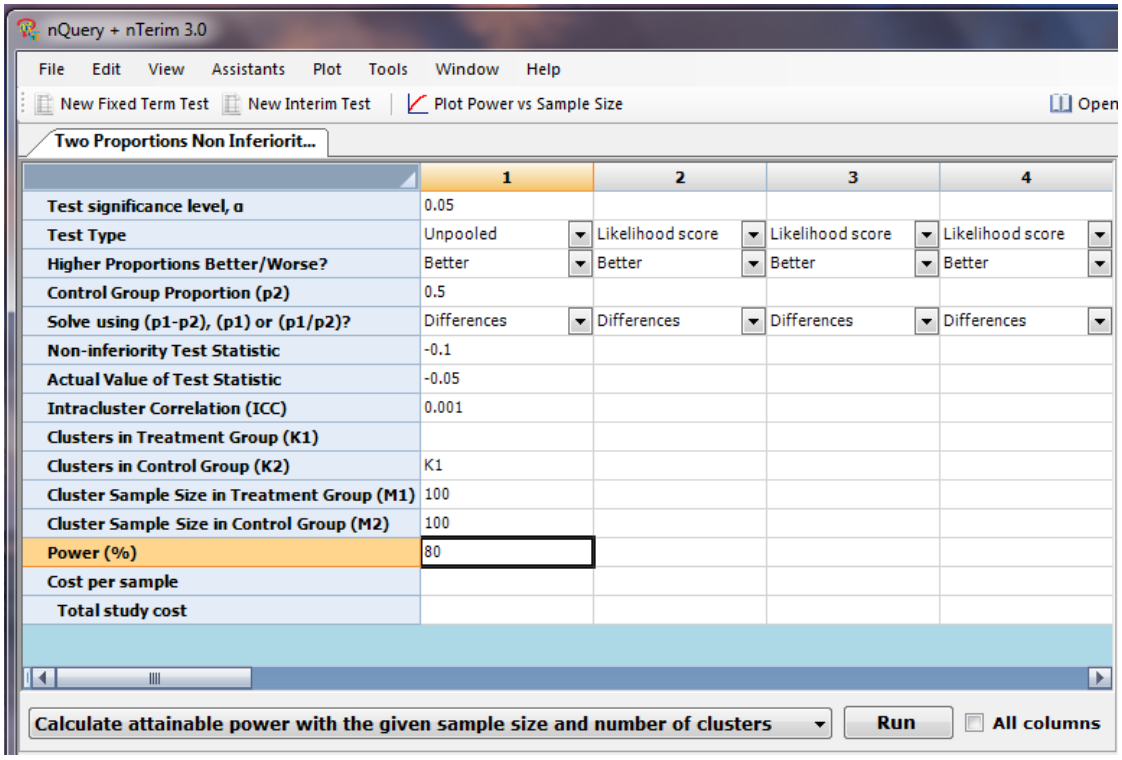


Figure 5.4.3. Values entered for CRT Two Proportions Non-Inferiority study design

6. Select “Calculate required treatment group clusters (K1) given power and sample size” from the dropdown menu beside the “Run” button. Then click “Run”. This will give a result of 14 for K1 and K2 as in Figure 5.4.4.

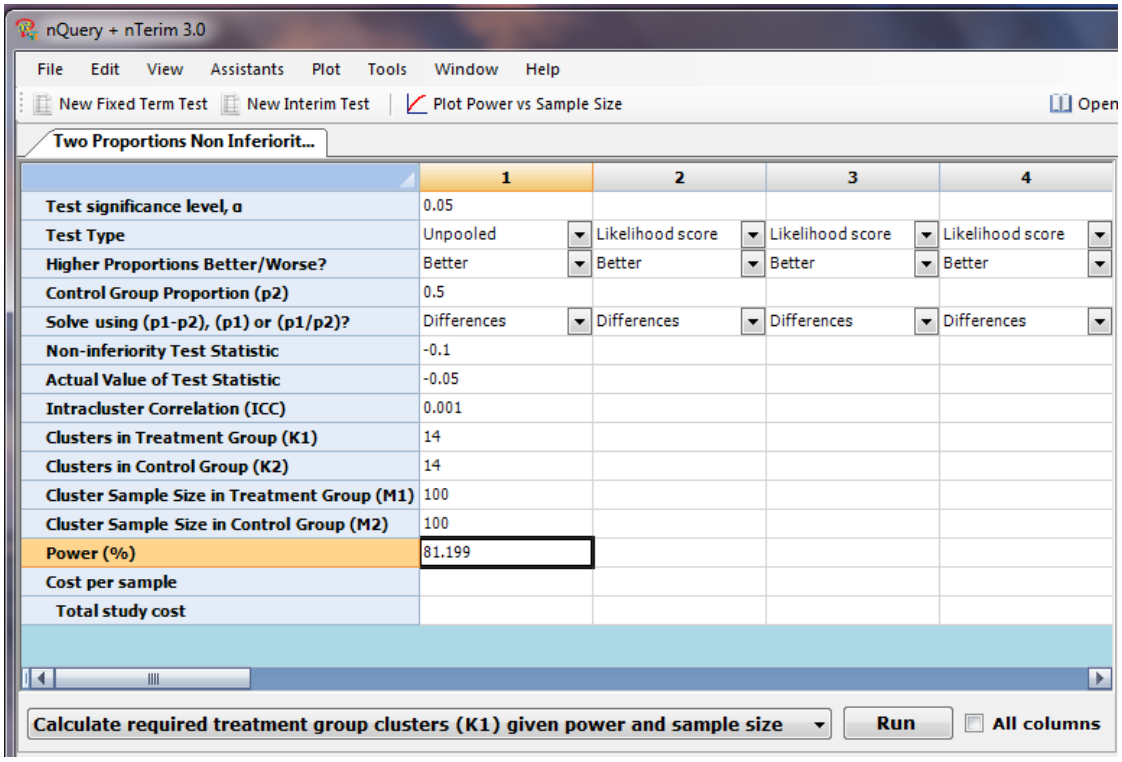
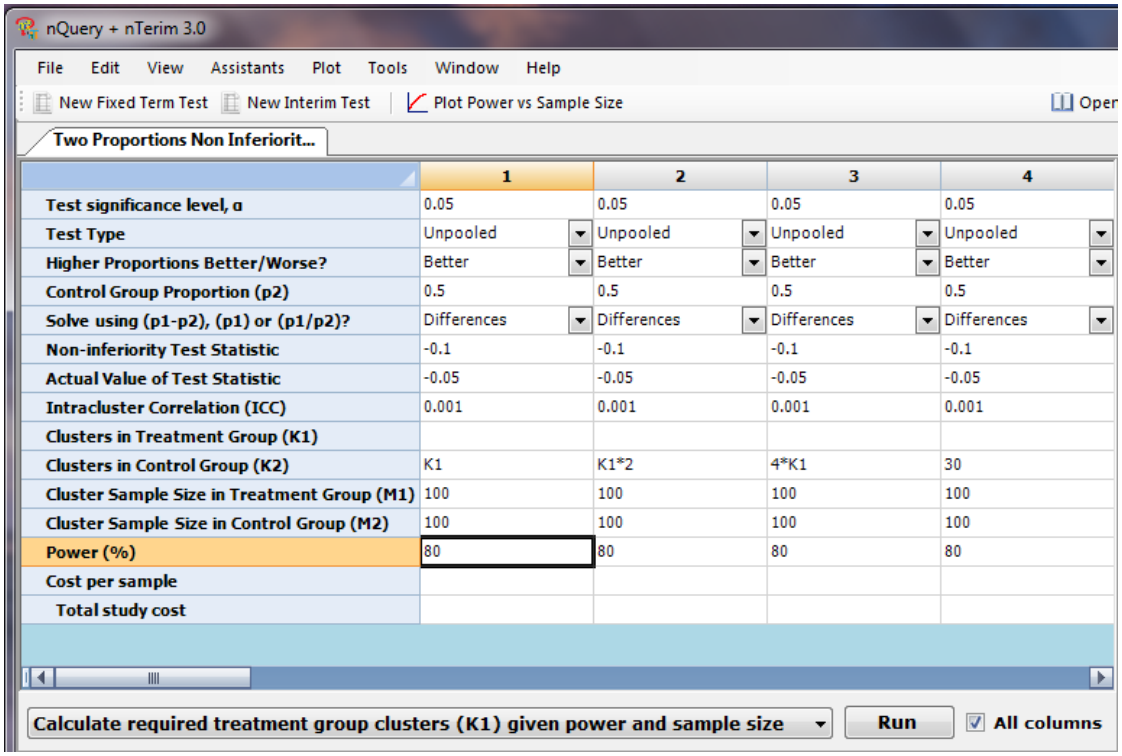


Figure 5.4.4. Completed CRT Two Proportions Non-Inferiority study design

The next calculation is a sensitivity analysis for the treatment group number of clusters when the control group number of clusters is changed.

7. Delete the values for K1 and K2 in the first column then replace the updated power with 80 and enter “K1” in the control group clusters row. Then copy the first column into columns 2 to 4. Other ratios, other than K1 and K2 being same, between K1 and K2 can be calculated by using arguments for K2 such as 2*K1 or K1*2 to have K2 be twice as large as K1 for example.

8. Enter K1*2 in column 2 for “Clusters in Control Group (K2)”, 4*K1 in column 3 and 30 in column four. This will give a table as per Figure 5.4.5.

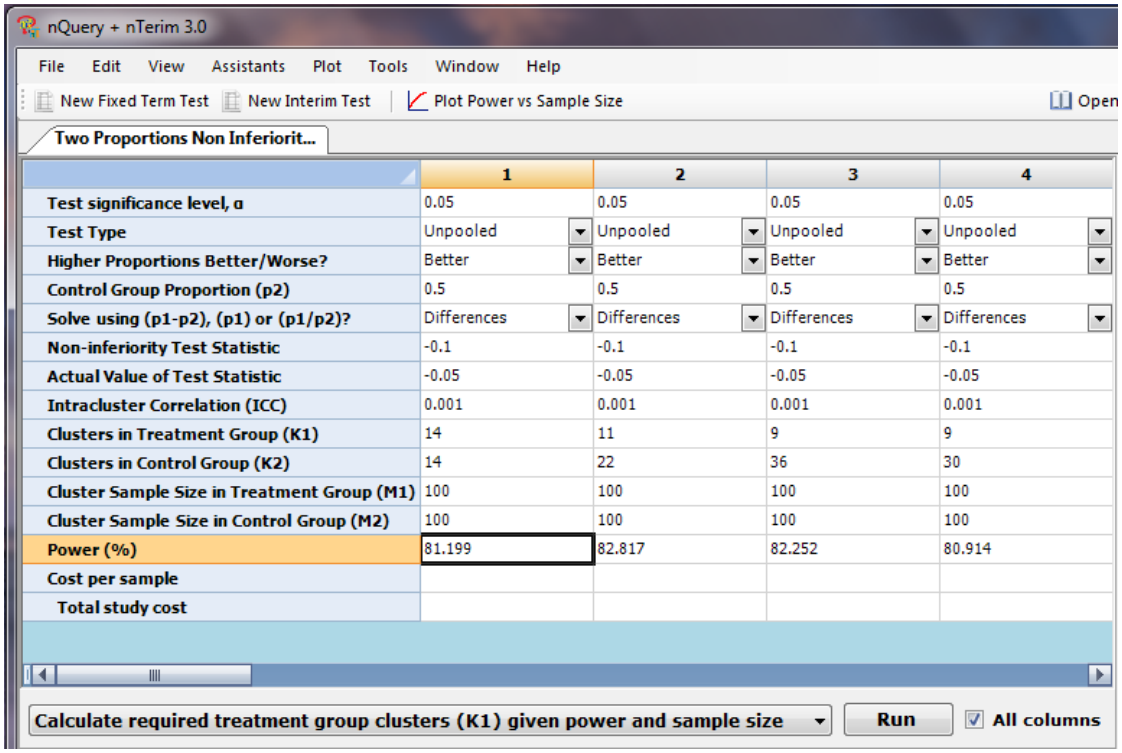


	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
Test Type	Unpooled	Unpooled	Unpooled	Unpooled
Higher Proportions Better/Worse?	Better	Better	Better	Better
Control Group Proportion (p2)	0.5	0.5	0.5	0.5
Solve using (p1-p2), (p1) or (p1/p2)?	Differences	Differences	Differences	Differences
Non-inferiority Test Statistic	-0.1	-0.1	-0.1	-0.1
Actual Value of Test Statistic	-0.05	-0.05	-0.05	-0.05
Intracluster Correlation (ICC)	0.001	0.001	0.001	0.001
Clusters in Treatment Group (K1)				
Clusters in Control Group (K2)	K1	K1*2	4*K1	30
Cluster Sample Size in Treatment Group (M1)	100	100	100	100
Cluster Sample Size in Control Group (M2)	100	100	100	100
Power (%)	80	80	80	80
Cost per sample				
Total study cost				

Figure 5.4.5. Sensitivity analysis around the Control Group Number of Clusters

9. Select “Calculate required treatment group clusters (K1) given power and sample size” from the dropdown menu beside the “Run” button. Then tick the box for “All Columns”. Then click “Run”.

This will give the resultant values of K1 of 14, 11, 9 and 9 sequentially, with the values of K2 updating automatically to reflect the desired ratio between K1 and K2, as in Figure 5.4.6.



	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
Test Type	Unpooled	Unpooled	Unpooled	Unpooled
Higher Proportions Better/Worse?	Better	Better	Better	Better
Control Group Proportion (p2)	0.5	0.5	0.5	0.5
Solve using (p1-p2), (p1) or (p1/p2)?	Differences	Differences	Differences	Differences
Non-inferiority Test Statistic	-0.1	-0.1	-0.1	-0.1
Actual Value of Test Statistic	-0.05	-0.05	-0.05	-0.05
Intracluster Correlation (ICC)	0.001	0.001	0.001	0.001
Clusters in Treatment Group (K1)	14	11	9	9
Clusters in Control Group (K2)	14	22	36	30
Cluster Sample Size in Treatment Group (M1)	100	100	100	100
Cluster Sample Size in Control Group (M2)	100	100	100	100
Power (%)	81.199	82.817	82.252	80.914
Cost per sample				
Total study cost				

Figure 5.4.6. Results from Sensitivity analysis

10. By clicking on the desired study design column and going to the "Output" tab at the bottom of the screen, you can get an output statement for the results given by the current calculation as per Figure 5.4.7.

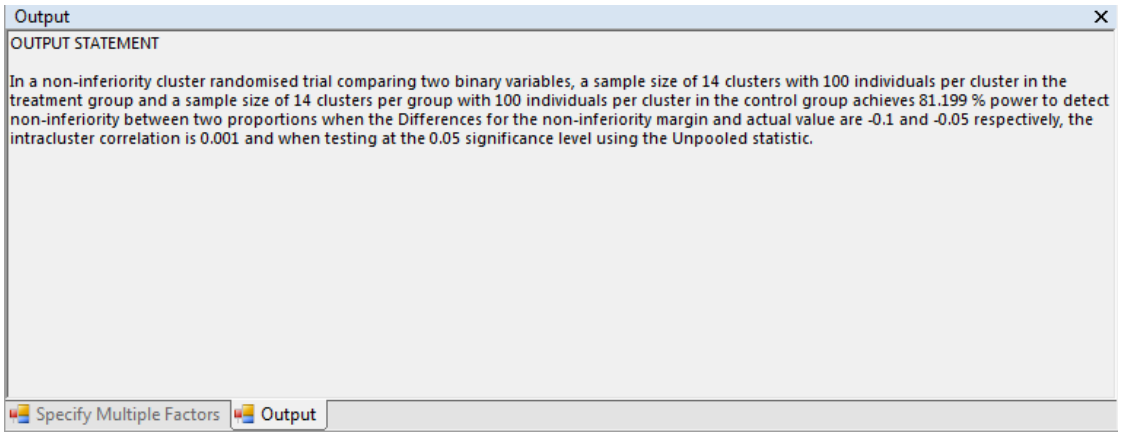


Figure 5.4.7. Study design Output statement

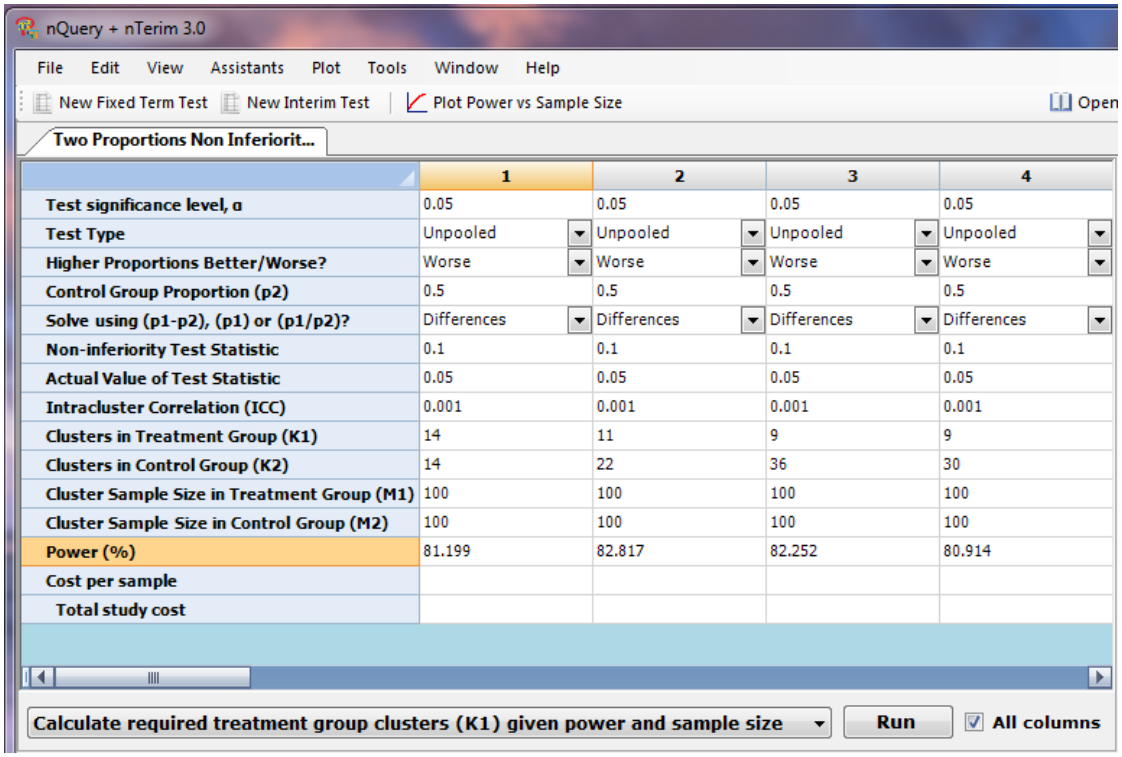
This Output statement can then be easily transferred directly from the output window into a report.

Example 2: Validation example calculating clusters when higher proportions are worse

A calculation is conducted to show the effect of selecting "Worse" for the "Higher Proportions are Better/Worse?" option. If "Worse" is selected then higher values for the proportion are considered worse from the study's perspective and thus positive differences would be used to test non-inferiority.

11. Return the table to its values before the sensitivity analysis. Then replace the -0.1 with 0.1 and -0.05 with 0.05 in the Non-Inferiority Test Statistic and Actual Value Test Statistic rows respectively. Set the Clusters in Control Group (K2) back to K1, K1*2, 4*K1 and 30 for columns 1 to 4 respectively. Then, set the power values back to 80.

12. Finally, select "Calculate required treatment group clusters (K1) given power and sample size" from the dropdown menu beside the "Run" button. Next, tick the box to run "All Columns". Then click "Run". This will give Figure 5.4.8.



	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
Test Type	Unpooled	Unpooled	Unpooled	Unpooled
Higher Proportions Better/Worse?	Worse	Worse	Worse	Worse
Control Group Proportion (p2)	0.5	0.5	0.5	0.5
Solve using (p1-p2), (p1) or (p1/p2)?	Differences	Differences	Differences	Differences
Non-inferiority Test Statistic	0.1	0.1	0.1	0.1
Actual Value of Test Statistic	0.05	0.05	0.05	0.05
Intracluster Correlation (ICC)	0.001	0.001	0.001	0.001
Clusters in Treatment Group (K1)	14	11	9	9
Clusters in Control Group (K2)	14	22	36	30
Cluster Sample Size in Treatment Group (M1)	100	100	100	100
Cluster Sample Size in Control Group (M2)	100	100	100	100
Power (%)	81.199	82.817	82.252	80.914
Cost per sample				
Total study cost				

Figure 5.4.8. Results from second Sensitivity analysis

This will give the same answers as for the above sensitivity calculation due to the control proportion being 0.5 (in which case these values lower and higher than 0.5 are symmetric in terms of the calculation.). If the control group proportion were not 0.5, we would have expected different values for the two calculations.

5.5. CRT Two Proportions Superiority – Completely Randomized

5.5.1 Introduction

Binary data is commonly studied in variety of different fields. Superiority trials are commonly used to assess whether a new treatment is better than a pre-existing treatment in the clinical setting e.g. comparing a new drug to its competitor. Clustered data is very common in a wide variety of academic, social policy and economic studies. This two sample test is used to test hypotheses about the superiority of a treatment in a completely randomized cluster randomized trial.

This table facilitates the calculation of the power and sample size for superiority hypothesis tests comparing proportions in a completely randomized cluster randomized trial. Power and sample size is computed using the method outlined by Donner and Klar (2000).

5.5.2 Methodology

This table provides sample size and power calculations for studies which will be conducting a superiority trial between proportions in trials which use a completely randomized cluster randomization study design. Superiority trials are those in which the researcher is testing that a treatment is better than the pre-existing control treatment. A completely randomized design assigns clusters randomly to control and treatment groups.

The sample size calculation for cluster randomized proportions is taken from Donner and Klar (2000) and is an extension of the methods used for superiority trials in a fully randomized trial, as outlined in Chow et al. (2008). The extension uses the intracluster correlation (ICC), which is the ordinary product moment correlation between any two observations in the same cluster, to adjust for the effect of within cluster correlation. It is assumed the ICC is the same in both groups.

This table can be used to calculate the power, the number of clusters in the treatment group, the sample size per cluster in the treatment group, the intracluster correlation and the smallest detectable difference given all other terms in the table are specified.

To calculate power, the number of clusters in the treatment group or the sample size per cluster in the treatment group, the user must specify the test significance level (α), the control group proportion (p_2), which test type is being used, the format of the test statistic, whether higher values for the proportion are better or worse from the researcher's perspective, the superiority margin, the actual expected difference, the intracluster correlation (ICC), the number of clusters in the control group (K_2) and the sample size per cluster in the control group (M_2).

The formula uses the normal approximation to calculate power. The formulae use the difference between the proportions under the superiority margin (Δ) and the actual expected difference (δ) regardless of the format of statistic used. Proportions and ratios are converted to the relevant differences.

If higher values are considered better then the expected difference should be higher than the superiority margin and if the lower values are considered better then the expected difference should be lower than the superiority margin. The formulae for the power are thus given by the following equations:

$$Power = \begin{cases} \Phi\left(\left[\frac{\delta - \Delta}{\sigma_{Unpool}}\right] - z_{1-\alpha}\left[\frac{\sigma_{Test\ Type}}{\sigma_{Unpool}}\right]\right), & \Delta < \delta \\ \Phi\left(\left[\frac{-(\delta - \Delta)}{\sigma_{Unpool}}\right] - z_{1-\alpha}\left[\frac{\sigma_{Test\ Type}}{\sigma_{Unpool}}\right]\right), & \Delta > \delta \end{cases} \qquad \{5.5.1\}$$

where $\sigma_{Test\ Type}$ is the standard error defined by the test type being used and σ_{Unpool} is the unpooled standard error. The formula for the three standard error statistics is as follows:

1. Farrington and Manning Test Statistic (Likelihood Score)
The Farrington and Manning test uses the constrained maximum likelihood estimator of the two proportions to calculate the standard error for the calculation of power and was proposed as method to test against a null hypothesis of a specified difference (Δ).

The standard error used by the Farrington and Manning test statistic is defined as follows:

$$\sigma_{FM} = \sqrt{\frac{p_{FM1}(1-p_{FM1})IF_1}{K_1M_1} + \frac{p_{FM2}(1-p_{FM2})IF_2}{K_2M_2}} \qquad \{5.5.2\}$$

where K_1 is the number of clusters in the treatment group, M_1 is the sample size per cluster in the treatment group, p_{FMi} is the maximum likelihood estimator for each group proportion and IF_i is the inflation factor for the effect of clustering in the treatment and control groups respectively. IF_i is defined as follows:

$$IF_i = \left(1 + \left(ICC(M_i - 1)\right)\right) \qquad i = 1,2 \qquad \{5.5.3\}$$

The constrained maximum likelihood (where $\tilde{p}_1 - \tilde{p}_2 = \Delta$) estimator of the two proportions is calculated using the following calculations:

$$t = \frac{(K_1M_1/IF_1)}{(K_2M_2/IF_2)} \qquad \{5.5.4\}$$

$$a = 1 + t \qquad \{5.5.5\}$$

$$b = -\left(1 + t + p_1 + tp_2 + \left(\Delta(t+2)\right)\right) \qquad \{5.5.6\}$$

$$c = \Delta^2 + \left(\Delta(2p_1 + t + 1)\right) + p_1 + tp_2 \qquad \{5.5.7\}$$

$$d = -p_1\Delta(1+\Delta) \qquad \{5.5.8\}$$

$$v = \left(\frac{b}{3a}\right)^3 - \left(\frac{bc}{6a^2}\right) + \left(\frac{d}{2a}\right) \qquad \{5.5.9\}$$

$$u = sign(v)\sqrt{\left(\frac{b}{3a}\right)^2 - \frac{c}{3a}} \qquad \{5.5.10\}$$

$$w = \frac{\left(\pi + \cos^{-1}\left(\frac{v}{u^3}\right)\right)}{3} \qquad \{5.5.11\}$$

$$p_{FM1} = 2u\cos(w) - b/3a \qquad \{5.5.12\}$$

$$p_{FM2} = p_{FM1} - \Delta \qquad \{5.5.13\}$$

2. Unpooled Test Statistic
   This test statistic uses the estimated group proportions to calculate the standard error. Its formula is as follows:

$$\sigma_{Unpool} = \sqrt{\frac{p_1(1-p_1)IF_1}{K_1M_1} + \frac{p_2(1-p_2)IF_2}{K_2M_2}} \qquad \{5.5.14\}$$

   where p_1 and p_2 are the estimated mean proportions for the two groups.

3. Pooled Test Statistic
   The pooled test statistic uses the weighted average of the two proportions to calculate the standard error. Its formula is as follows:

$$\sigma_{Pool} = \sqrt{\frac{\bar{p}(1-\bar{p})IF_1}{K_1M_1} + \frac{\bar{p}(1-\bar{p})IF_2}{K_2M_2}} \qquad \{5.5.15\}$$

   where

$$\bar{p} = \frac{K_1M_1p_1 + K_2M_2p_2}{K_1M_1 + K_2M_2} \qquad \{5.5.16\}$$

A closed form equation is not used to calculate the other parameters. Instead a search algorithm is used. The search algorithm calculates power at various values for the relevant parameter until the desired power is reached.

5.5.3 Examples

Example 1: Validation example calculating required number of clusters for a given power
The following example will look at a number of clusters calculation problem is conducted and then a sensitivity analysis is conducted to show the effect of changing control group number of clusters. The following steps outline the procedure for this example:

1. Open nQuery + nTerim 4.0 via the start menu or desktop shortcut. Click “New Fixed Term Test” from the top of the window. The “Study Goal and Design” window will appear as Figure 5.5.1 below.

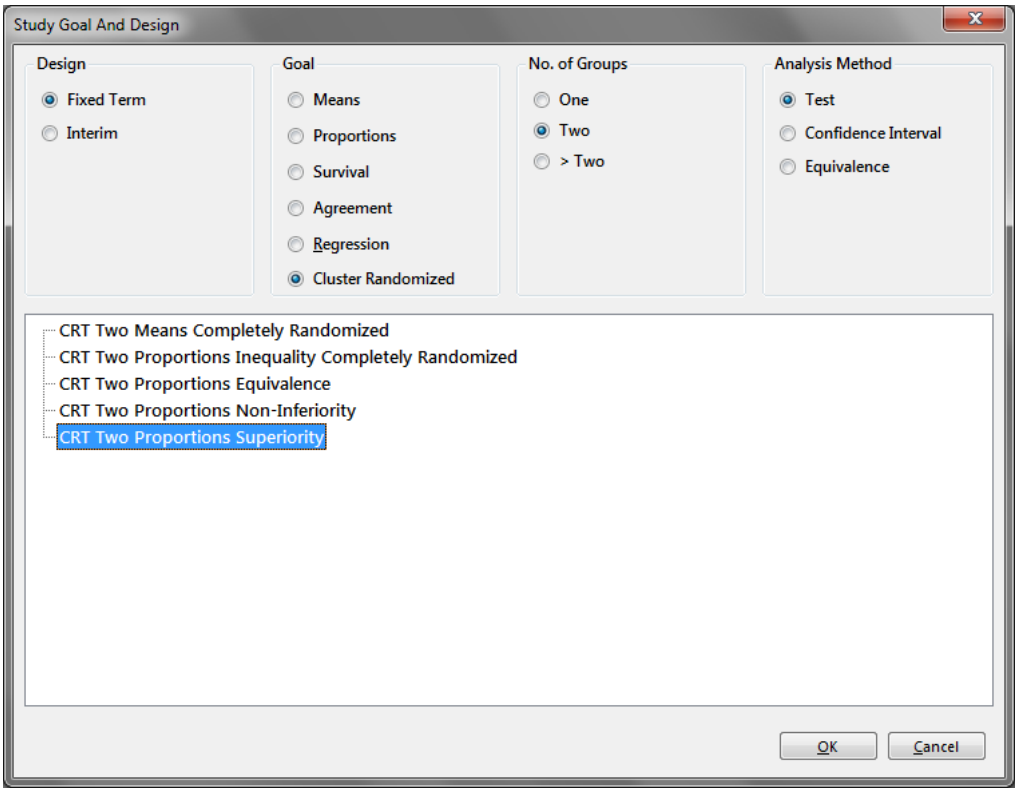


Figure 5.5.1. Study Goal and Design Window

2. In order to select the CRT Two Proportions Superiority Completely Randomized design table, navigate through the Study Goal and Design Window by selecting "Fixed Term" as the Design, "Cluster Randomized" as the Goal, "Two" as the Number of Groups and "Test" as the Analysis Method. Then click "OK" and the test window will appear as per Figure 5.5.2.

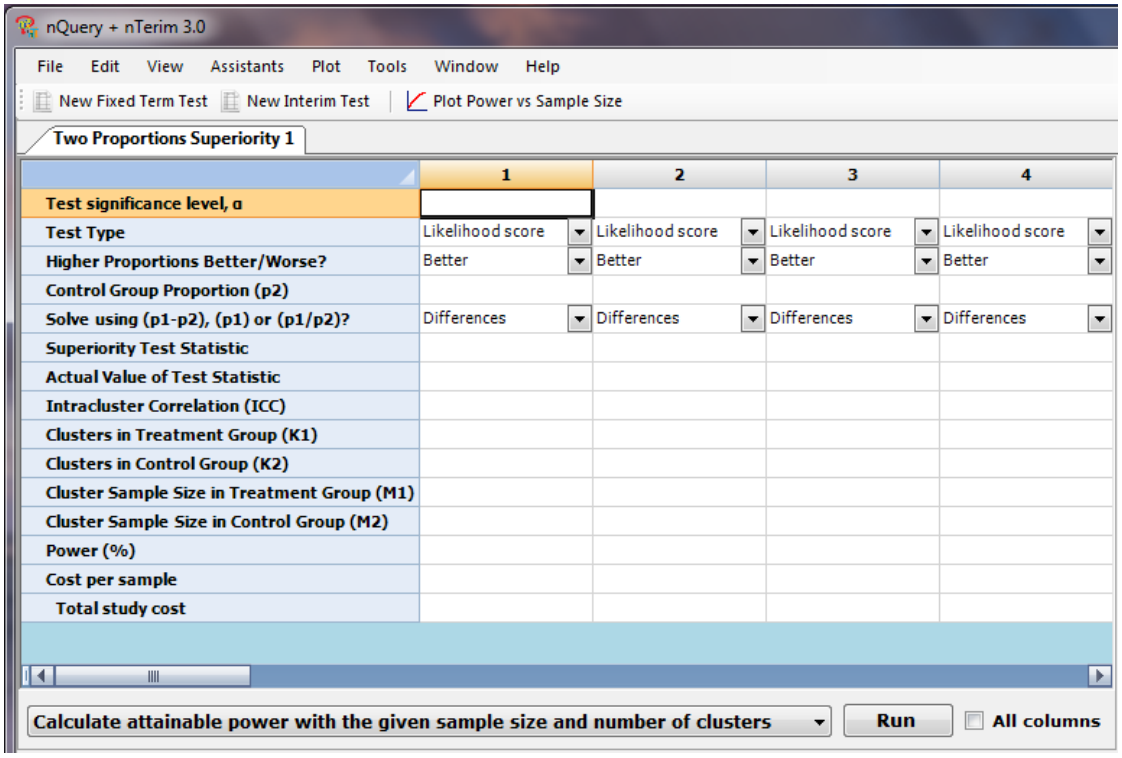


Figure 5.5.2. CRT Two Proportions Superiority Completely Randomized Test Table

The first calculation will be for the "Clusters in Treatment Group (K1)". This example is can be undertaken using the following steps.

3. First enter 0.05 for the Test Significance level row. Next, select "Likelihood Score" from the Test Type option, "Better" from the "Higher Proportions Better/Worse" option and enter a control group proportion of 0.5.

4. Next, select "Differences" from the "Solve Using" dropdown option and enter 0.1 for the "Superiority Test Statistic" and 0.15 for the "Actual Value of Test Statistic". Enter 0.001 for the intracluster correlation. Enter 100 for both cluster sample size variables and 80 for power.

5. Finally, enter K1 in the "Clusters in Control Group (K2)" row. This will solve so that K1 and K2 must be equal. Other ratios between K1 and K2 can be calculated by using arguments for K2 such as 2*K1 or K1*2 to have K2 be twice as large as K1 for example. The table will appear as in Figure 5.5.3.

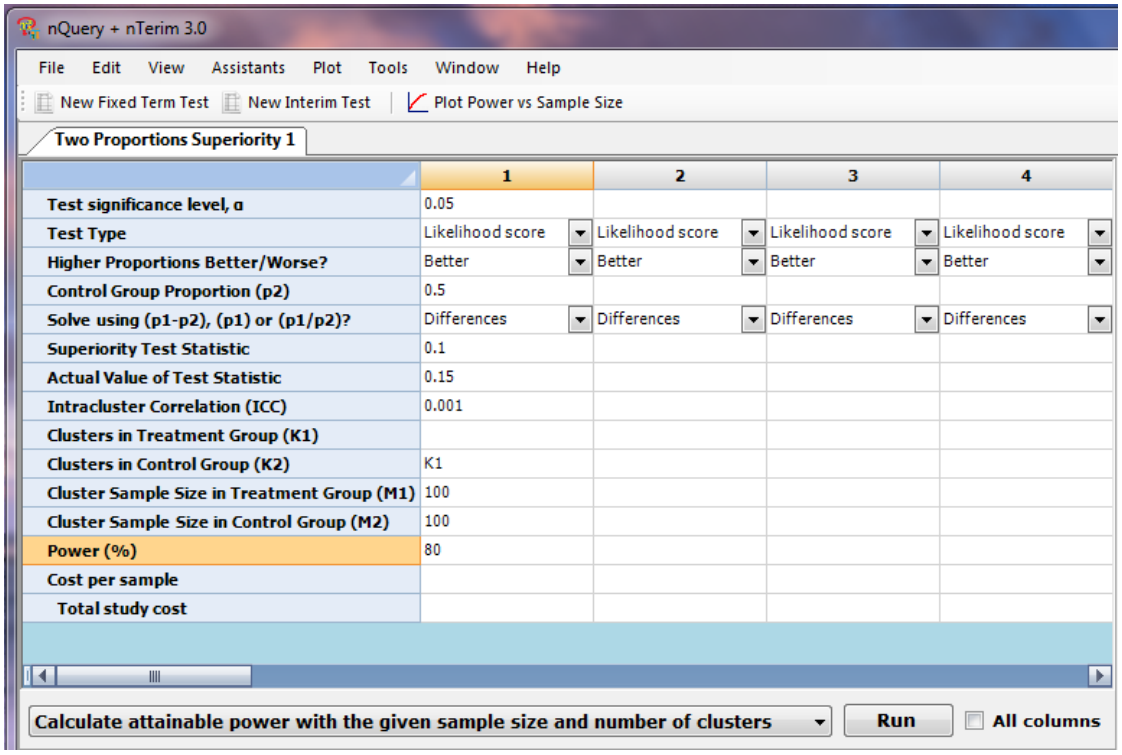


Figure 5.5.3. Values entered for CRT Two Proportions Superiority study design

6. Select "Calculate required treatment group clusters (K1) given power and sample size" from the dropdown menu beside the "Run" button. Then click "Run". This will give a result of 14 for K1 and K2 as in Figure 5.5.4.

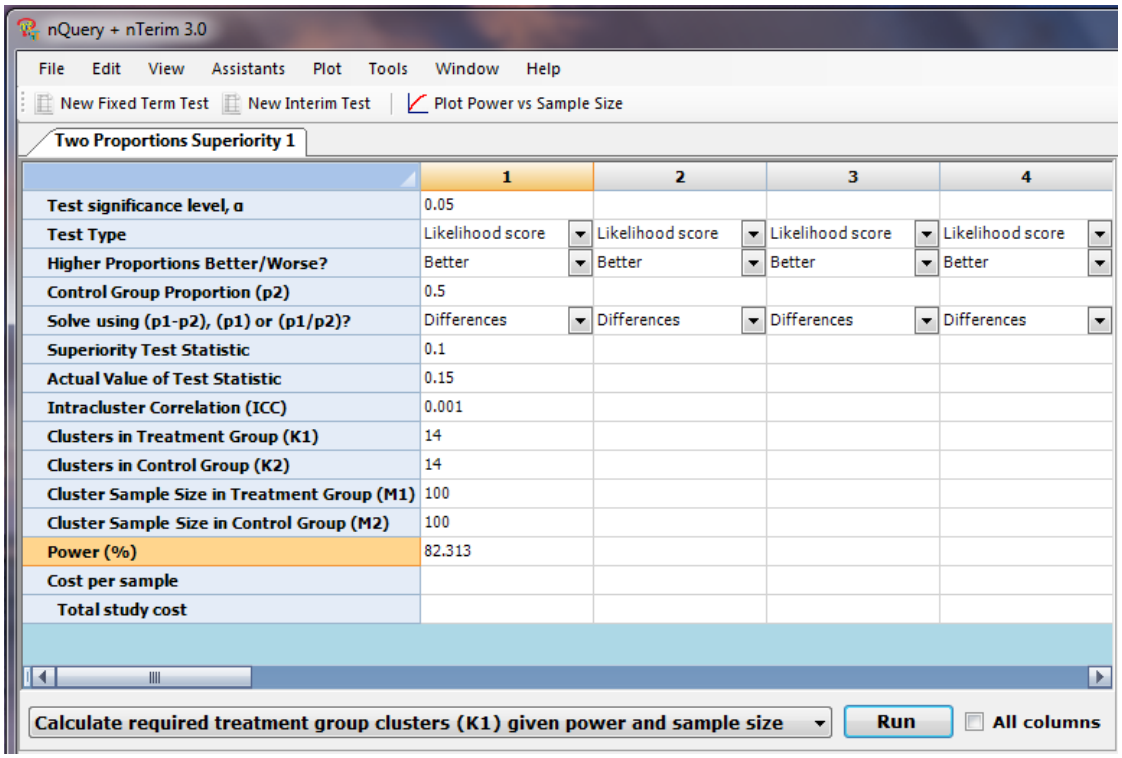
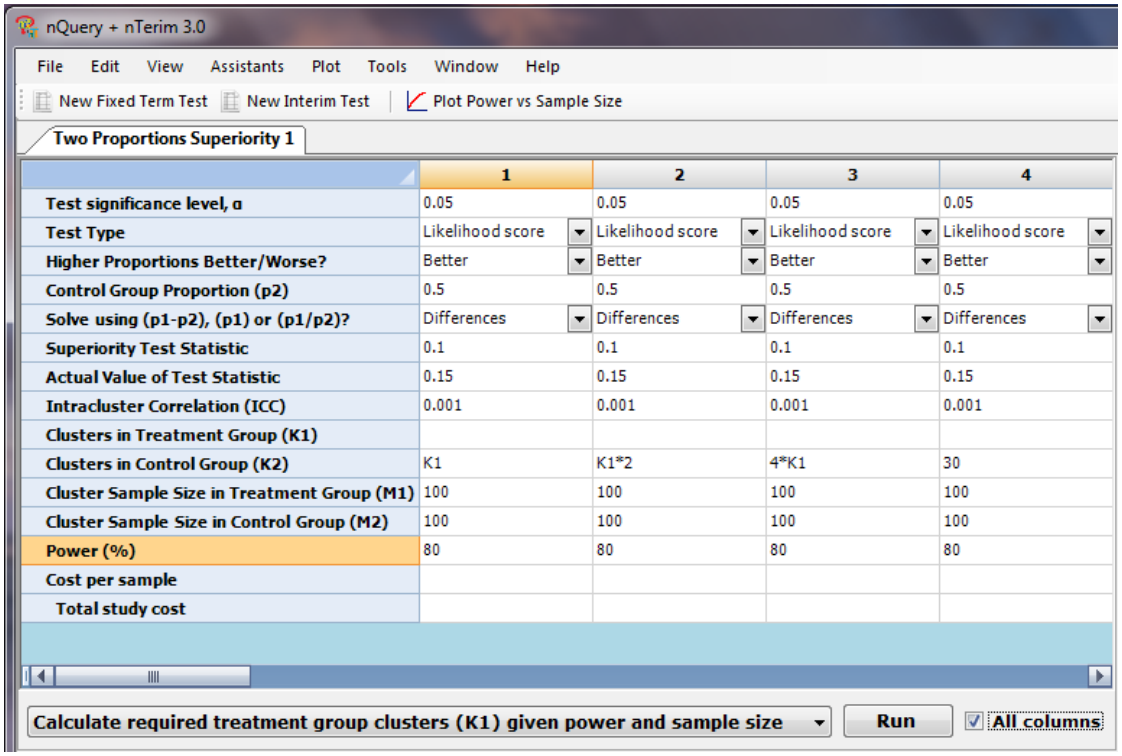


Figure 5.5.4. Completed CRT Two Proportions Superiority study design

The next calculation is a sensitivity analysis for the treatment group number of clusters when the control group number of clusters is changed.

7. Delete the values for K1 and K2 in the first column then replace the updated power with 80 and enter “K1” in the control group clusters row. Then copy the first column into columns 2 to 4. Other ratios, other than K1 and K2 being same, between K1 and K2 can be calculated by using arguments for K2 such as 2*K1 or K1*2 to have K2 be twice as large as K1 for example.

8. Enter K1*2 in column 2 for “Clusters in Control Group (K2)”, 4*K1 in column 3 and 30 in column four. This will give a table as per Figure 5.5.5.

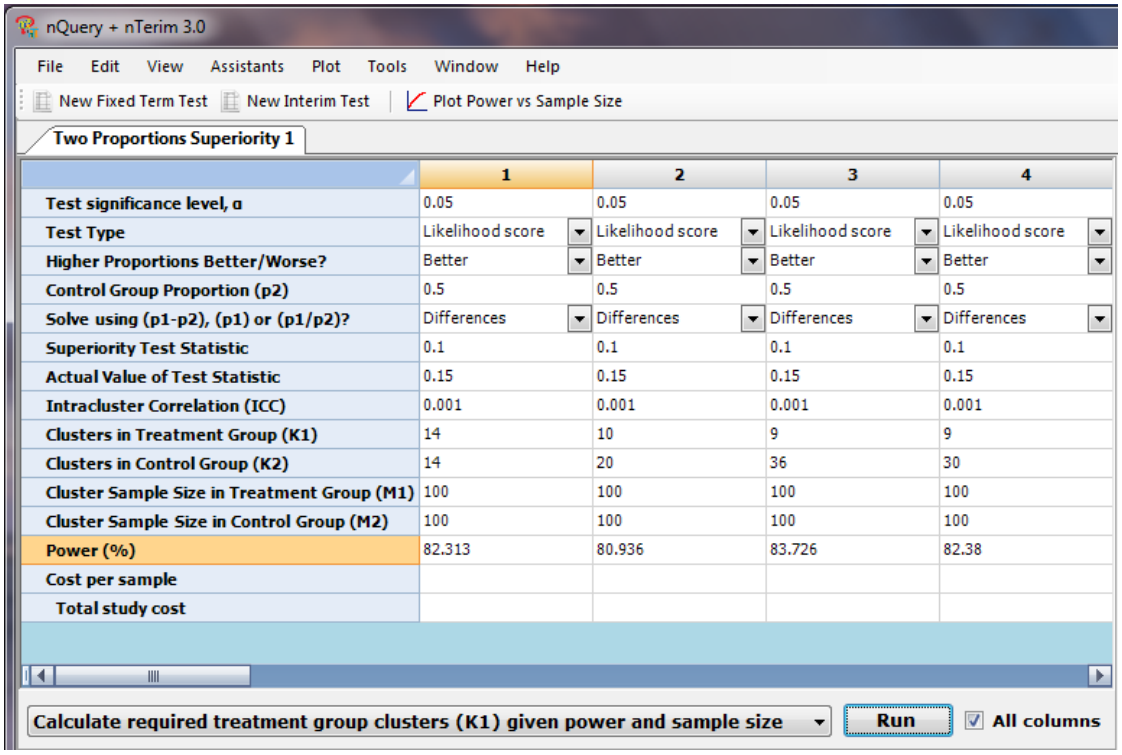


	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
Test Type	Likelihood score	Likelihood score	Likelihood score	Likelihood score
Higher Proportions Better/Worse?	Better	Better	Better	Better
Control Group Proportion (p2)	0.5	0.5	0.5	0.5
Solve using (p1-p2), (p1) or (p1/p2)?	Differences	Differences	Differences	Differences
Superiority Test Statistic	0.1	0.1	0.1	0.1
Actual Value of Test Statistic	0.15	0.15	0.15	0.15
Intracluster Correlation (ICC)	0.001	0.001	0.001	0.001
Clusters in Treatment Group (K1)				
Clusters in Control Group (K2)	K1	K1*2	4*K1	30
Cluster Sample Size in Treatment Group (M1)	100	100	100	100
Cluster Sample Size in Control Group (M2)	100	100	100	100
Power (%)	80	80	80	80
Cost per sample				
Total study cost				

Figure 5.5.5. Sensitivity analysis around the Control Group Number of Clusters

9. Select “Calculate required treatment group clusters (K1) given power and sample size” from the dropdown menu beside the “Run” button. Then tick the box for “All Columns”. Then click “Run”.

This will give the resultant values of K1 of 14, 10, 9 and 9 sequentially, with the values of K2 updating automatically to reflect the desired ratio between K1 and K2, as in Figure 5.5.6.



	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
Test Type	Likelihood score	Likelihood score	Likelihood score	Likelihood score
Higher Proportions Better/Worse?	Better	Better	Better	Better
Control Group Proportion (p2)	0.5	0.5	0.5	0.5
Solve using (p1-p2), (p1) or (p1/p2)?	Differences	Differences	Differences	Differences
Superiority Test Statistic	0.1	0.1	0.1	0.1
Actual Value of Test Statistic	0.15	0.15	0.15	0.15
Intracluster Correlation (ICC)	0.001	0.001	0.001	0.001
Clusters in Treatment Group (K1)	14	10	9	9
Clusters in Control Group (K2)	14	20	36	30
Cluster Sample Size in Treatment Group (M1)	100	100	100	100
Cluster Sample Size in Control Group (M2)	100	100	100	100
Power (%)	82.313	80.936	83.726	82.38
Cost per sample				
Total study cost				

Figure 5.5.6. Results from Sensitivity analysis

10. By clicking on the desired study design column and going to the "Output" tab at the bottom of the screen, you can get an output statement for the results given by the current calculation as per Figure 5.5.7.

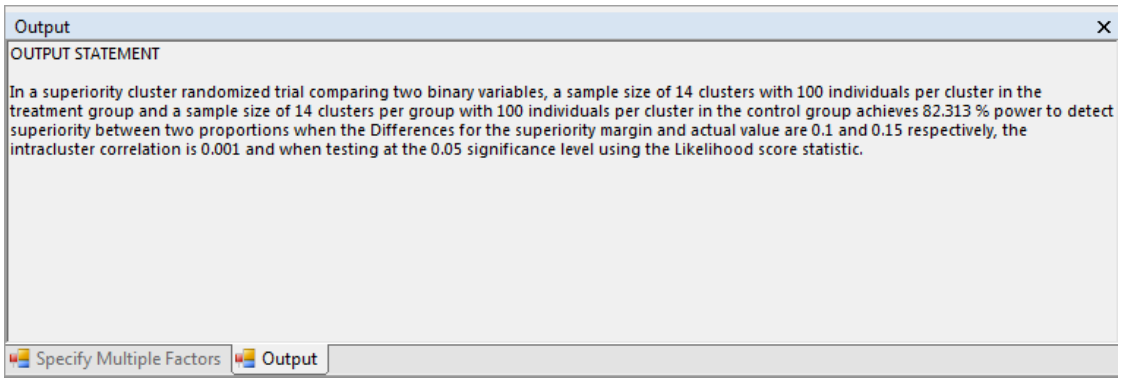


Figure 5.5.7. Study design Output statement

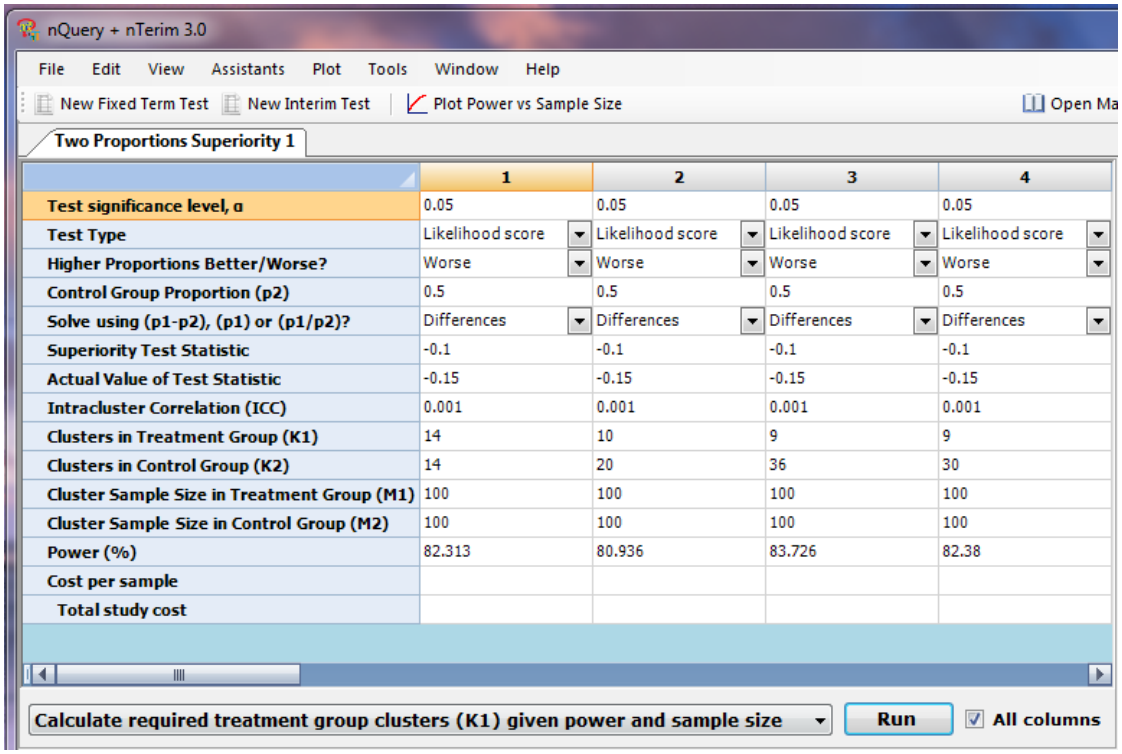
This Output statement can then be easily transferred directly from the output window into a report.

Example 2: Validation example calculating clusters when higher proportions are worse

A calculation is conducted to show the effect of selecting "Worse" for the "Higher Proportions are Better/Worse?" option. If "Worse" is selected then higher values for the proportion are considered worse from the study's perspective and thus positive differences would be used to test non-inferiority.

11. Return the table to its values before the sensitivity analysis. Then replace the 0.1 with -0.1 and 0.15 with -0.15 in the Superiority Test Statistic and Actual Value Test Statistic rows respectively. Set the Clusters in Control Group (K2) back to K1, K1*2, 4*K1 and 30 for columns 1 to 4 respectively. Then, set the power values back to 80.

12. Finally, select "Calculate required treatment group clusters (K1) given power and sample size" from the dropdown menu beside the "Run" button. Next, tick the box to run "All Columns". Then click "Run". This will give Figure 5.5.8.



	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
Test Type	Likelihood score	Likelihood score	Likelihood score	Likelihood score
Higher Proportions Better/Worse?	Worse	Worse	Worse	Worse
Control Group Proportion (p2)	0.5	0.5	0.5	0.5
Solve using (p1-p2), (p1) or (p1/p2)?	Differences	Differences	Differences	Differences
Superiority Test Statistic	-0.1	-0.1	-0.1	-0.1
Actual Value of Test Statistic	-0.15	-0.15	-0.15	-0.15
Intracluster Correlation (ICC)	0.001	0.001	0.001	0.001
Clusters in Treatment Group (K1)	14	10	9	9
Clusters in Control Group (K2)	14	20	36	30
Cluster Sample Size in Treatment Group (M1)	100	100	100	100
Cluster Sample Size in Control Group (M2)	100	100	100	100
Power (%)	82.313	80.936	83.726	82.38
Cost per sample				
Total study cost				

Figure 5.5.8. Results from second Sensitivity analysis

This will give the same answers as for the above sensitivity calculation due to the control proportion being 0.5 (in which case these values lower and higher than 0.5 are symmetric in terms of the calculation.). If the control group proportion were not 0.5, we would have expected different values for the two calculations.

5.6 CRT Two Incidence Rates – Completely Randomized

5.6.1 Introduction

Incidence data is often obtained in a variety of clinical and field research activities such as the number of accidents at a junction or number of occurrences of a disease in a year. Clustered data is very common in a wide variety of academic, social policy and economic studies. Researchers often plan studies which deal with years of exposure and sample size.

This table facilitates the calculation of the power and sample size for hypothesis tests comparing two incidence rates in a cluster randomised trial. Power and sample size is computed using the method outlined by Donner and Klar (2000) and Hayes and Bennett (1999).

5.6.2 Methodology

This table provides sample size and power calculations for studies which will be comparing two incidence rates which use a completely randomised cluster randomisation study design. A completely randomised design assigns clusters randomly to control and treatment groups. This table assumes a balanced study design.

The sample size calculation for two incidence rates is taken from Donner and Klar (2000) and is an extension of the method outlined by Smith and Morrow (1996). The extension uses the coefficient of variation, which is equal to the between cluster standard deviation divided by the mean rate, to adjust for the effect of within cluster correlation. It is assumed the coefficient of variation is the same in both groups.

This table can be used to calculate the power, the number of clusters per group, the sample size per cluster, the fixed observation period per subject, the coefficient of variation and the smallest detectable difference given all other terms in the table are specified.

To calculate power, the number of clusters or the sample size per cluster, the user must specify the test significance level (α), whether to use a one or two sided test, the control group incidence rate (λ_1), the treatment group incidence rate (λ_2), the difference in rates ($\lambda_1 - \lambda_2$), the fixed observation time per subject (t) and the coefficient of variation (CV).

The formulae use the normal approximation to produce estimates. The formula for the number of clusters needed per group is given by the following:

$$k = \frac{\left(z_{1-\beta} + z_{1-\alpha/2}\right)^2 (\lambda_1 + \lambda_2) IF}{yt(\lambda_1 - \lambda_2)^2} \qquad \{5.6.1\}$$

Where

$$IF = 1 + \frac{CV^2(\lambda_1^2 + \lambda_2^2)yt}{(\lambda_1 + \lambda_2)} \qquad \{5.6.2\}$$

where k is the number of clusters per group and y is the sample size per cluster. For the one-sided test, $z_{1-\alpha/2}$ is replaced with $z_{1-\alpha}$. Simple rearrangement of the above formula gives the following equation for the power:

Rearrangement can be used to yield the following formula for the sample size y:

$$y = \frac{\left[\frac{\left(z_{1-\beta} + z_{1-\alpha/2}\right)^2 (\lambda_1 + \lambda_2)}{k(\lambda_1 - \lambda_2)^2}\right]}{t\left(1 - \frac{\left(z_{1-\beta} + z_{1-\alpha/2}\right)^2 (\lambda_1 + \lambda_2) IF_m}{k(\lambda_1 - \lambda_2)^2}\right)} \qquad \{5.6.3\}$$

A similar rearrangement can be used to yield the following formula for the observation time per subject:

$$t = \frac{\left[\frac{\left(z_{1-\beta} + z_{1-\alpha/2}\right)^2 (\lambda_1 + \lambda_2)}{k(\lambda_1 - \lambda_2)^2}\right]}{y\left(1 - \frac{\left(z_{1-\beta} + z_{1-\alpha/2}\right)^2 (\lambda_1 + \lambda_2) IF_m}{k(\lambda_1 - \lambda_2)^2}\right)} \qquad \{5.6.4\}$$

where in both equations

$$IF_m = \frac{CV^2(\lambda_1^2 + \lambda_2^2)}{(\lambda_1 + \lambda_2)} \qquad \{5.6.5\}$$

The formula for the coefficient of variation can be shown to be equal to the following:

$$CV = \sqrt{\left(\frac{(\lambda_1 + \lambda_2)}{yt(\lambda_1^2 + \lambda_2^2)}\right)\left(\frac{ykt(\lambda_1 - \lambda_2)^2}{\left(z_{1-\beta} + z_{1-\alpha/2}\right)^2 (\lambda_1 + \lambda_2)} - 1\right)} \qquad \{5.6.6\}$$

A closed form equation is not used to calculate the smallest detectable difference. Instead a search algorithm is used. The search algorithm calculates power at various values for the treatment group incidence rate until the desired power is reached.

5.6.3 Examples

Example 1: Validation example calculating required number of clusters for a given power
The following example will look at a number of clusters calculation problem from Donner and Klar (2000) and then a sensitivity analysis is conducted to show the effect of changing the coefficient of variation on the number of clusters required. The following steps outline the procedure for this example:

1. Open nQuery + nTerim 4.0 via the start menu or desktop shortcut. Click “New Fixed Term Test” from the top of the window. The “Study Goal and Design” window will appear as Figure 5.5.1 below.

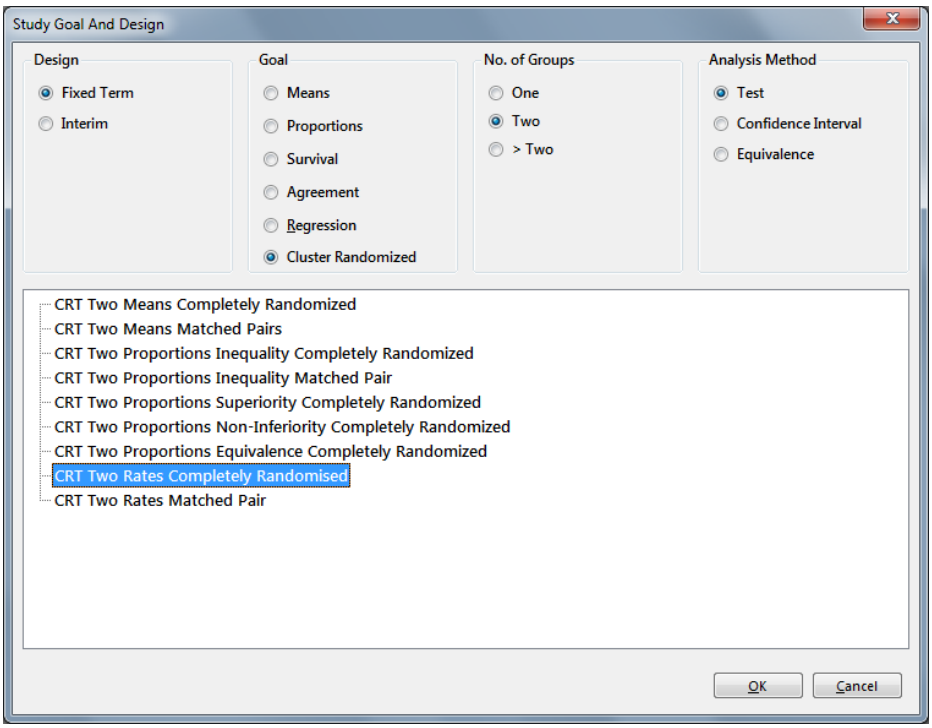


Figure 5.6.1. Study Goal and Design Window

2. In order to select the “CRT Two Incidence Rates Completely Randomized” design table, navigate through the Study Goal and Design Window by selecting "Fixed Term" as the Design, "Cluster Randomized" as the Goal, "Two" as the Number of Groups and "Test" as the Analysis Method. Then click "OK" and the test window will appear as per Figure 5.5.2.

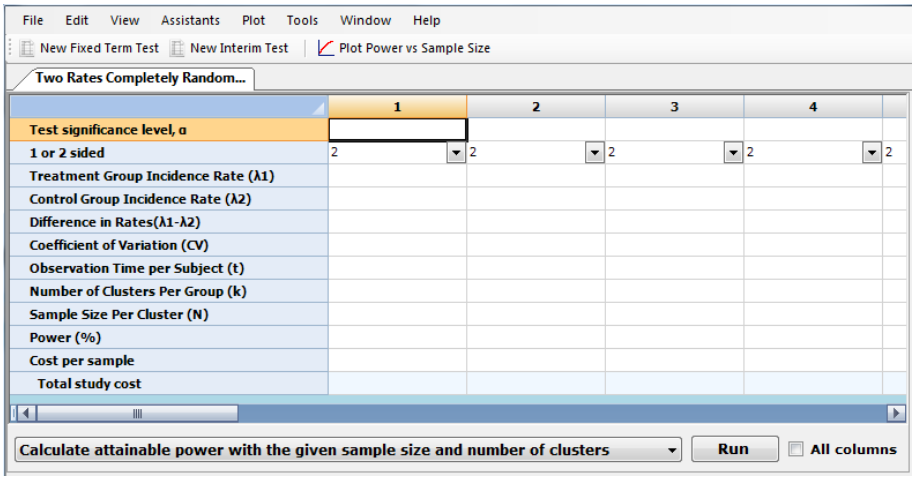


Figure 5.6.2. CRT Two Rates Completely Randomized Test Table

The first calculation will be for the "Calculate Required Number of Clusters for Given power and Sample Size" calculation using the example from Donner and Klar (2000). This example can be undertaken using the following steps.

3. First enter 0.05 for the "Test Significance level" row. Next, select 2 from the "1 or 2 sided" row.

4. Next, enter 0.005 and 0.01 for the "Treatment Group Incidence Rate (λ1)" and "Control Group Incidence (λ2)" rows respectively. The "Difference in Rates" row will populate automatically.

5. Next, enter 0.25 in the "Coefficient of Variation (CV)" row and 2 in the "Observation Time (t)" row.

6. Next, enter 1000 for the "Sample Size per Cluster (N)" row and 80 for the "Power (%)" row. This will give inputs as per Figure 5.6.3 below.

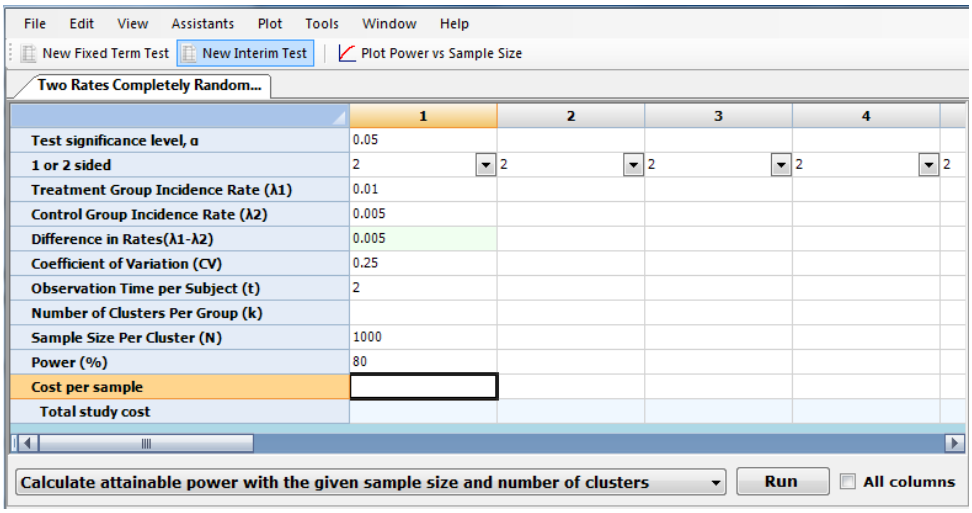
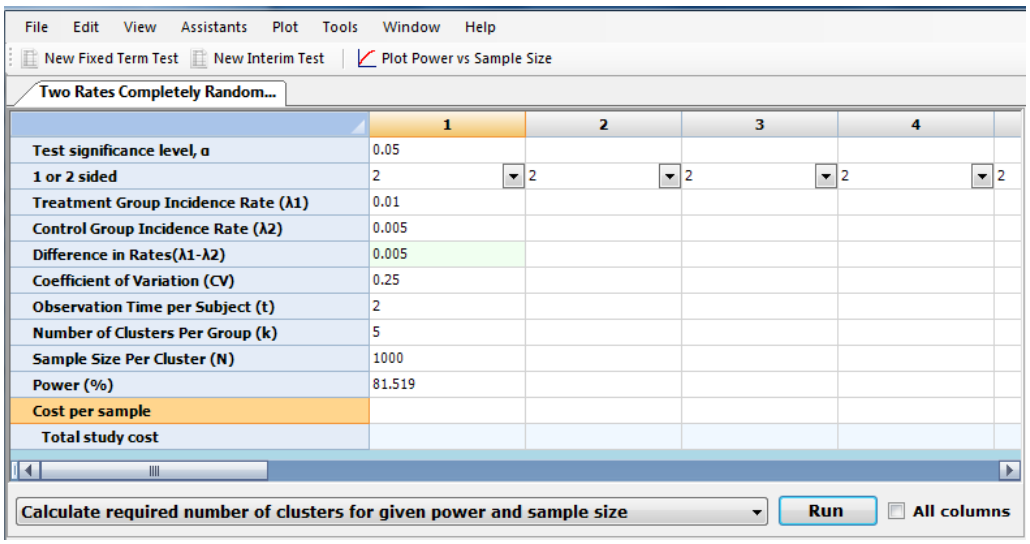


Figure 5.6.3. Values entered for CRT Two Rates Completely Randomized study design

7. Select “Calculate Required Number of Clusters for Given power and Sample Size” from the dropdown menu beside the “Run” button. Then click “Run”. This will give a result of 5 clusters per group for the number of clusters as in Figure 5.6.4. This exact value given in Donner and Klar is 4.8, which rounds to five clusters per group.



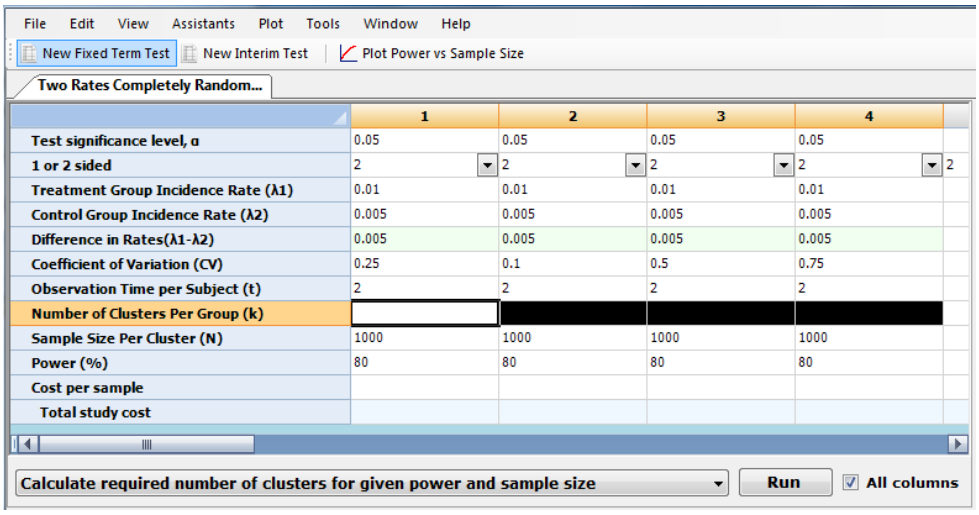
	1	2	3	4
Test significance level, α	0.05			
1 or 2 sided	2	2	2	2
Treatment Group Incidence Rate (λ1)	0.01			
Control Group Incidence Rate (λ2)	0.005			
Difference in Rates(λ1-λ2)	0.005			
Coefficient of Variation (CV)	0.25			
Observation Time per Subject (t)	2			
Number of Clusters Per Group (k)	5			
Sample Size Per Cluster (N)	1000			
Power (%)	81.519			
Cost per sample				
Total study cost				

Figure 5.6.4. Completed CRT Two Rates Completely Randomized study design

The next calculation is a sensitivity analysis for the treatment group number of clusters when the coefficient of variation is changed.

8. Delete the value for “k” in the first column then replace the updated power with 80. Then copy the first column into columns 2 to 4.

9. Enter 0.1 in column 2 for “Coefficient of Variation (CV)”, 0.5 in column 3 and 0.75 in column four. This will give a table as per Figure 5.6.5.

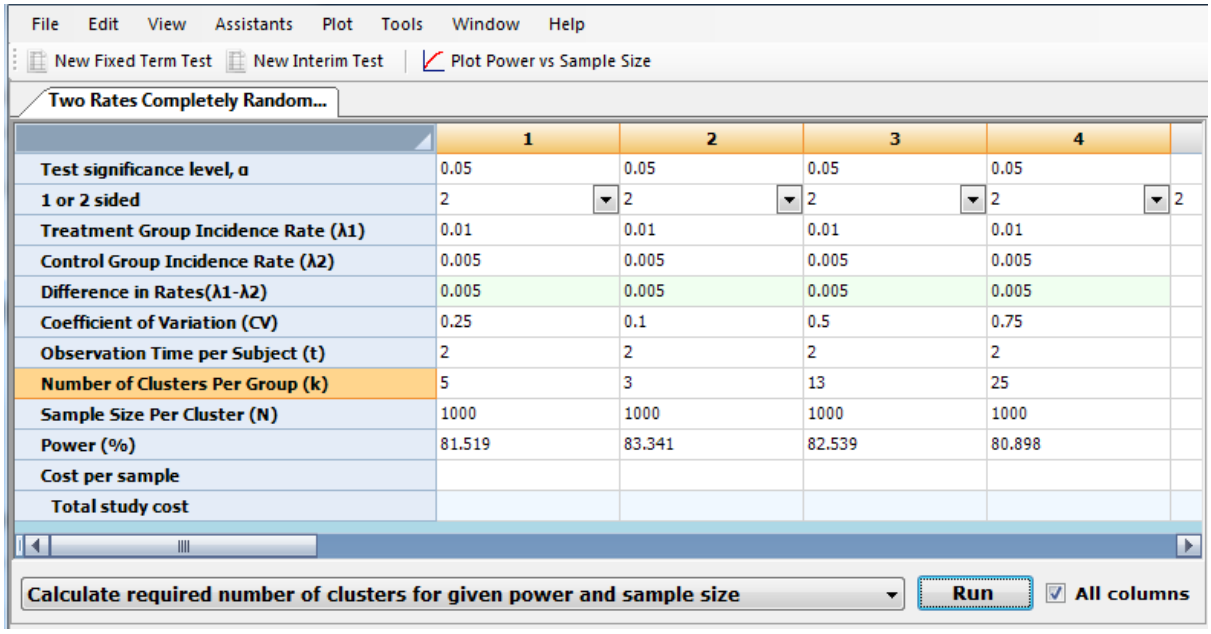


	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
1 or 2 sided	2	2	2	2
Treatment Group Incidence Rate (λ1)	0.01	0.01	0.01	0.01
Control Group Incidence Rate (λ2)	0.005	0.005	0.005	0.005
Difference in Rates(λ1-λ2)	0.005	0.005	0.005	0.005
Coefficient of Variation (CV)	0.25	0.1	0.5	0.75
Observation Time per Subject (t)	2	2	2	2
Number of Clusters Per Group (k)				
Sample Size Per Cluster (N)	1000	1000	1000	1000
Power (%)	80	80	80	80
Cost per sample				
Total study cost				

Figure 5.6.5. Sensitivity analysis for Coefficient of Variation

10. Select “Calculate Required Number of Clusters for Given power and Sample Size” from the dropdown menu beside the “Run” button. Then tick the box for “All Columns”. Then click “Run”.

This will give the resultant values for “k” of 5, 3, 13 and 25 sequentially as in Figure 5.6.6.



	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
1 or 2 sided	2	2	2	2
Treatment Group Incidence Rate (λ1)	0.01	0.01	0.01	0.01
Control Group Incidence Rate (λ2)	0.005	0.005	0.005	0.005
Difference in Rates(λ1-λ2)	0.005	0.005	0.005	0.005
Coefficient of Variation (CV)	0.25	0.1	0.5	0.75
Observation Time per Subject (t)	2	2	2	2
Number of Clusters Per Group (k)	5	3	13	25
Sample Size Per Cluster (N)	1000	1000	1000	1000
Power (%)	81.519	83.341	82.539	80.898
Cost per sample				
Total study cost				

Figure 5.6.6. Results from Sensitivity analysis

11. By clicking on the desired study design column and going to the “Output” tab at the bottom of the screen, you can get an output statement for the results given by the current calculation as per Figure 5.6.7.

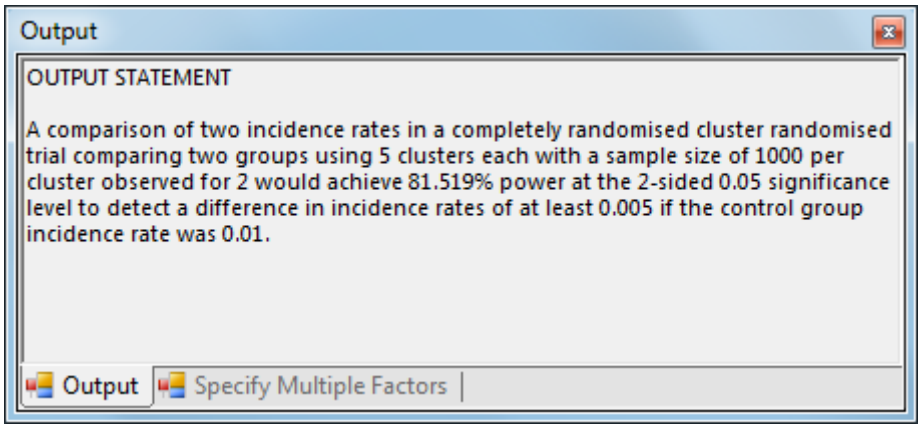


Figure 5.6.7. Study design Output statement

This Output statement can then be easily transferred directly from the output window into a report.

5.7 CRT Two Means – Matched-Pair

5.7.1 Introduction

Continuous data is found in nearly every area of research interest. Procedures such as Z and t-tests are used to evaluate the differences between two continuous means. Clustered data is very common in a wide variety of academic, social policy and economic studies. This two sample test is used to test hypotheses about the difference between two means in a matched-pair cluster randomised trial.

This table facilitates the calculation of the power and sample size for hypothesis tests comparing means in a matched-pair cluster randomised trial. Power and sample size is computed using the method outlined by Donner and Klar (2000).

5.7.2 Methodology

This table provides sample size and power calculations for studies which will be comparing means in a trials which uses a matched-pair cluster randomisation study design. A matched pair design assigns clusters into pairs depending on their homogeneity which is based on relevant information known for each cluster e.g. size of school, GDP of country. This can lead to increases in power when compared to the fully randomised design if the matching criteria is associated with the mean. This table assumes a balanced study design.

The sample size calculation for cluster randomised means is taken from Donner and Klar (2000). It uses the within-cluster and between-pair components of variability to account for the cluster randomisation and matched-pair design effects.

This table can be used to calculate the power, the test significance level (α), the number of paired clusters, the sample size per cluster, the within-cluster variance, the between-pair variance and the smallest detectable difference given all other terms in the table are specified.

To calculate power, the number of clusters or the sample size per cluster, the user must specify the test significance level (α), whether to use a one or two sided test, the difference in means ($|\mu_1 - \mu_2|$), the within-cluster variance (σ_W^2) and the between-cluster variance (σ_B^2).

The formulae use the normal distribution to calculate power. The formula for the number of cluster pairs needed is given by the following equation:

$$k = \frac{\left(z_{1-\beta} + z_{1-\alpha/2}\right)^2 V(d_j)}{|\mu_1 - \mu_2|^2} \qquad \{5.7.1\}$$

where

$$V(d_j) = 2\left(\frac{\sigma_W^2}{N} + \sigma_B^2\right) \qquad \{5.7.2\}$$

In the above formula, $V(d_j)$ denotes the variance of the paired differences, k is the number of clusters per group and N is the sample size per cluster. For the one-sided test, $z_{1-\alpha/2}$ is replaced with $z_{1-\alpha}$. Simple and similar rearrangements of the above formula give the following equations for the power, the test significance level and the expected absolute difference:

$$1 - \beta = \Phi\left(\frac{|\mu_1 - \mu_2|\sqrt{k}}{V(d_j)} - z_{1-\alpha/2}\right) \qquad \{5.7.3\}$$

$$\alpha = 2\left[1 - \Phi\left(\frac{|\mu_1 - \mu_2|\sqrt{k}}{V(d_j)} - z_{1-\beta}\right)\right] \qquad \{5.7.4\}$$

$$|\mu_1 - \mu_2| = \sqrt{\frac{\left(z_{1-\beta} + z_{1-\alpha/2}\right)^2 V(d_j)}{k}} \qquad \{5.7.5\}$$

Rearrangement of the above formulae can be used to yield the sample size per cluster as follows:

$$N = \frac{\sigma_W^2}{\left[\left(\frac{|\mu_1 - \mu_2|^2 k}{2\left(z_{1-\beta} + z_{1-\alpha/2}\right)^2}\right) - \sigma_B^2\right]} \qquad \{5.7.6\}$$

The within-cluster variance and between-cluster variance can be calculated using the following rearrangements:

$$\sigma_B^2 = \left[\frac{|\mu_1 - \mu_2|^2 k}{2\left(z_{1-\beta} + z_{1-\alpha/2}\right)^2}\right] - \frac{\sigma_W^2}{N} \qquad \{5.7.7\}$$

$$\sigma_W^2 = N\left[\left(\frac{|\mu_1 - \mu_2|^2 k}{2\left(z_{1-\beta} + z_{1-\alpha/2}\right)^2}\right) - \sigma_B^2\right] \qquad \{5.7.8\}$$

5.7.3 Examples

Example 1: Validation example calculating required number of clusters for a given power
Donner and Klar only go through a precision based example in their book. Thus, the following example will look at a matched cluster pairs calculation example from a from study design formulated by Sivakumar, Bhattiprolu, et al. (2006) based on Donner and Klar. Than a sensitivity analysis is conducted to show the effect of changing the between cluster variance on the number of paired clusters required. The following steps outline the procedure for this example:

1. Open nQuery + nTerim 4.0 via the start menu or desktop shortcut. Click “New Fixed Term Test” from the top of the window. The “Study Goal and Design” window will appear as Figure 5.7.1 below.

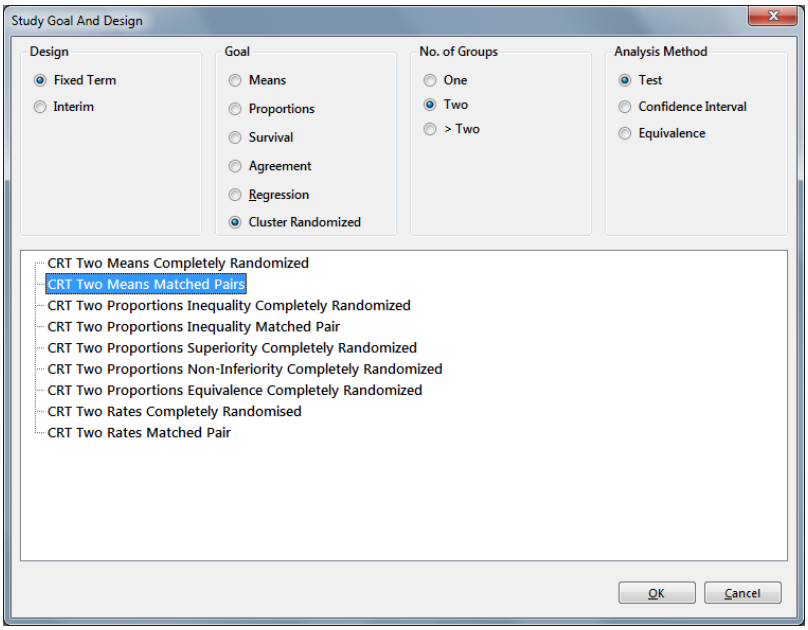


Figure 5.7.1. Study Goal and Design Window

2. In order to select the “CRT Two Mean Matched-Pairs” design table, navigate through the Study Goal and Design Window by selecting "Fixed Term" as the Design, "Cluster Randomized" as the Goal, "Two" as the Number of Groups and "Test" as the Analysis Method. Then click "OK" and the test window will appear as per Figure 5.7.1.

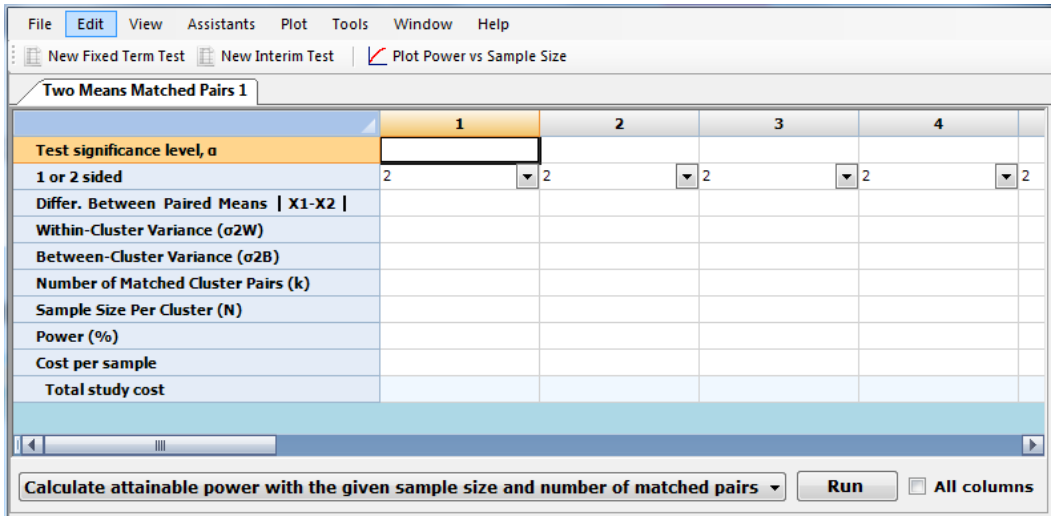


Figure 5.7.2. CRT Two Means Matched Pair Test Table

The first calculation will be for the "Calculate Required Number of Matched Pairs of Clusters for Given power and Sample Size" calculation using the example from Sivakumar, Bhattiprolu, et al. (2006) for study participants' weight. This example can be undertaken using the following steps.

3. First enter 0.05 for the Test Significance level row. Next, select "2" from the 1 or 2 sided option.

4. Next, enter 0.9 for the "Difference Between Paired Cluster Means".

5. Next, enter 3.883 in the "Within-Cluster Variance" row and 0.31 in the "Between-Cluster Variance" row.

6. Next, enter 38 for the "Sample Size per Cluster (N)" row and 80 for the "Power (%)" row. This will give inputs as per Figure 5.6.3 below.

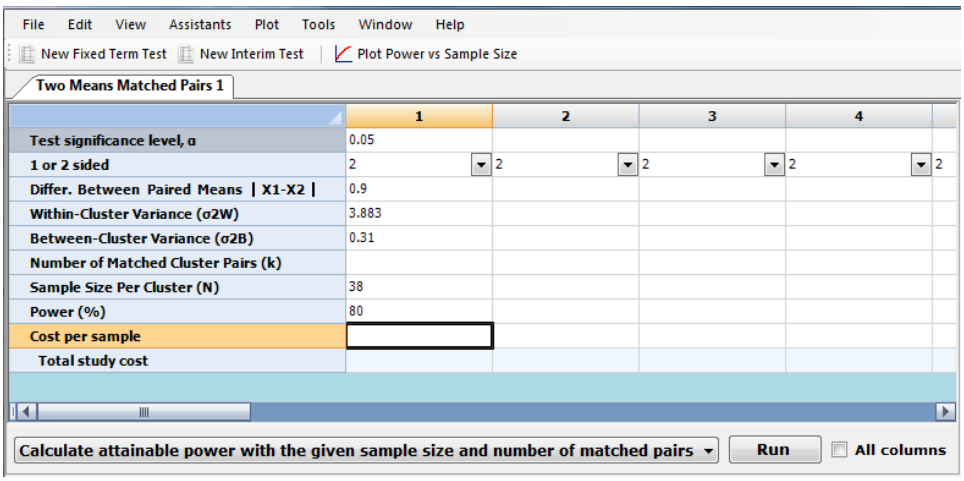
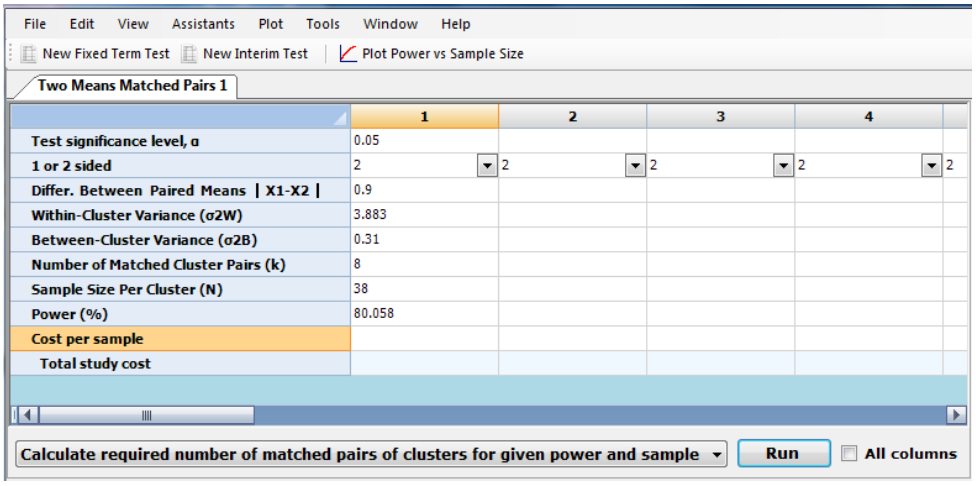


Figure 5.7.3. Values entered for CRT Two Means Matched Pair study design

7. Select "Calculate Required Number of Matched Pairs of Clusters for Given power and Sample Size" from the dropdown menu beside the "Run" button. Then click "Run". This

will give a result of 8 clusters per group for the number of paired clusters as in Figure 5.7.4. This agrees with the value of 8 pairs in Sivakumar, Bhattiprolu, et al.



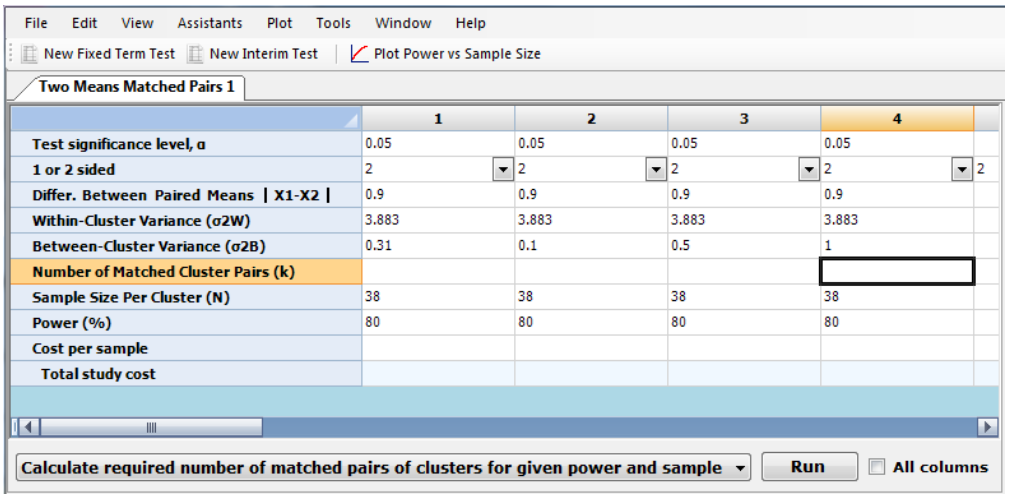
	1	2	3	4
Test significance level, α	0.05			
1 or 2 sided	2	2	2	2
Differ. Between Paired Means | X1-X2 |	0.9			
Within-Cluster Variance (σ2W)	3.883			
Between-Cluster Variance (σ2B)	0.31			
Number of Matched Cluster Pairs (k)	8			
Sample Size Per Cluster (N)	38			
Power (%)	80.058			
Cost per sample				
Total study cost				

Figure 5.7.4. Completed CRT Two Means Matched Pair study design

The next calculation is a sensitivity analysis for the number of paired clusters when the "Between-Cluster Variance" is changed.

8. Delete the value for "k" in the first column then replace the updated power with 80. Then copy the first column into columns 2 to 4.

9. Enter 0.1 in column 2 for "Coefficient of Variation (CV)", 0.5 in column 3 and 1 in column four. This will give a table as per Figure 5.7.5.

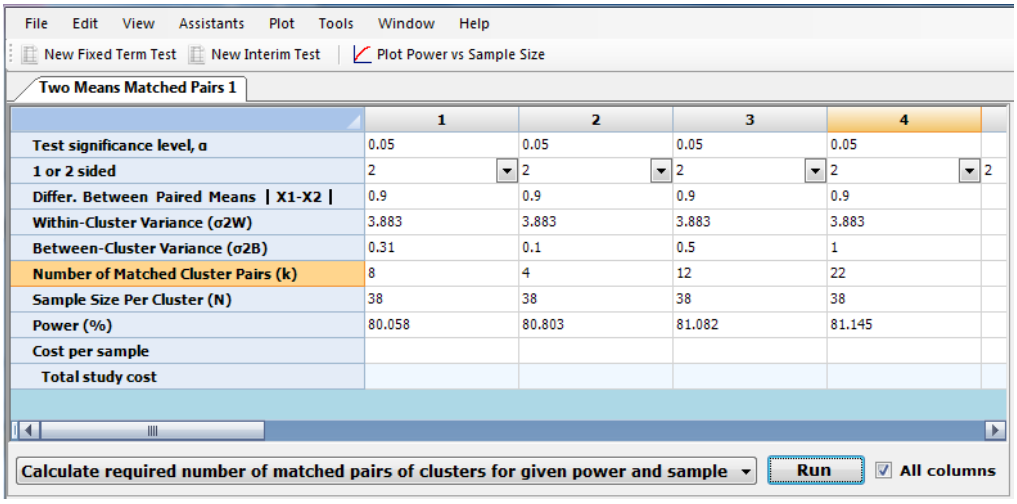


	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
1 or 2 sided	2	2	2	2
Differ. Between Paired Means | X1-X2 |	0.9	0.9	0.9	0.9
Within-Cluster Variance (σ2W)	3.883	3.883	3.883	3.883
Between-Cluster Variance (σ2B)	0.31	0.1	0.5	1
Number of Matched Cluster Pairs (k)				
Sample Size Per Cluster (N)	38	38	38	38
Power (%)	80	80	80	80
Cost per sample				
Total study cost				

Figure 5.7.5. Sensitivity analysis for Between-Cluster Variance

10. Select "Calculate Required Number of Matched Pairs of Clusters for Given power and Sample Size" from the dropdown menu beside the "Run" button. Then tick the box for "All Columns". Then click "Run".

This will give the resultant values for "k" of 8, 4, 12 and 22 sequentially as in Figure 5.7.6.



	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
1 or 2 sided	2	2	2	2
Differ. Between Paired Means | X1-X2 |	0.9	0.9	0.9	0.9
Within-Cluster Variance (σ2W)	3.883	3.883	3.883	3.883
Between-Cluster Variance (σ2B)	0.31	0.1	0.5	1
Number of Matched Cluster Pairs (k)	8	4	12	22
Sample Size Per Cluster (N)	38	38	38	38
Power (%)	80.058	80.803	81.082	81.145
Cost per sample				
Total study cost				

Figure 5.7.6. Results from Sensitivity analysis

11. By clicking on the desired study design column and going to the "Output" tab at the bottom of the screen, you can get an output statement for the results given by the current calculation as per Figure 5.7.7.

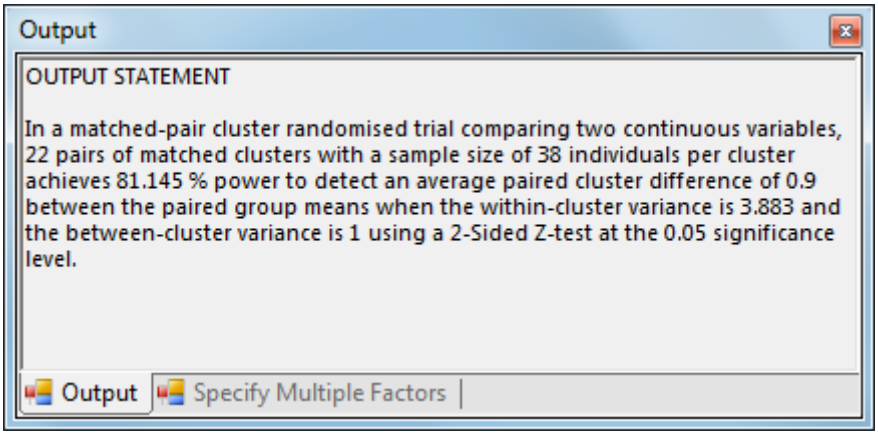


Figure 5.7.7. Study design Output statement

This Output statement can then be easily transferred directly from the output window into a report.

5.8 CRT Two Proportions Inequality – Matched-Pair

5.8.1 Introduction

Binary data is commonly studied in variety of different fields. Clustered data is very common in a wide variety of academic, social policy and economic studies. This two sample test is used to test hypotheses about the inequality of two treatments in a matched-pair cluster randomized trial.

This table facilitates the calculation of the power and sample size for inequality hypothesis tests comparing proportions in a matched-pair cluster randomized trial. Power and sample size is computed using the method outlined by Donner and Klar (2000).

5.8.2 Methodology

This table provides sample size and power calculations for studies which will be conducting an inequality trial between proportions in trials which use a matched pair cluster randomization study design. Inequality trials are those in which the researcher is testing that two treatments have different values under the alternative hypothesis. A matched pair design assigns clusters into pairs depending on their homogeneity which is based on relevant information known for each cluster e.g. size of school, GDP of country. This can lead to increases in power when compared to the fully randomised design if the matching criteria is associated with the prevalence. This table assumes a balanced study design.

The sample size calculation for two binary proportions is taken from Donner and Klar (2000) and is an extension of the method for the completely randomised design. It uses the between-cluster variance component to account for the effect of clustering and pairing. It measures the inherent variability between two clusters in a matched pair in the absence of an intervention effect.

This table can be used to calculate the power, the number of cluster pairs, the sample size per cluster in the treatment group, the between cluster variance and the effect size under the alternative hypothesis (given via a difference, treatment proportion or ratio) given all other terms in the table are specified. It assumed there is no difference in groups under the null hypothesis.

To calculate power, the number of cluster pairs or the sample size per cluster in the treatment group, the user must specify the test significance level (α), the control group proportion (p_2), which test type is being used, the format of the test statistic, the actual value of the chosen test statistic under the alternative hypothesis, the between cluster component of variance (σ^2_B) and the within cluster component of variance (σ^2_B).

The formula uses the normal approximation to calculate power. The formulae use the mean of paired mean differences between the proportions $(p_1 - p_2)$ under the alternative hypothesis as the effect size. Proportions and ratios are converted to the relevant differences using the control proportion and then power and sample size is calculated using

the difference methodology. It is assumed that power is being calculated under a null hypothesis of no difference.

The number of paired clusters is given by the following equation:

$$k = \frac{\left(z_{1-\beta} + z_{1-\alpha/2}\right)^2 Var(d_j)}{(p_1 - p_2)^2} \qquad \{5.8.1\}$$

where $Var(d_j)$ is the variance of the paired cluster mean differences and is estimated by the following equation:

$$Var(d_j) = \frac{p_1(1-p_1) + p_2(1-p_2)}{m} + 2\sigma_B^2 \qquad \{5.8.2\}$$

Where $p_1$is the treatment proportion, $p_2$is the control proportion, m is the sample size per cluster and σ_B^2 is the between cluster component of variance, which estimates how similar paired clusters are. In the one-sided case, $z_{1-\alpha/2}$ is replaced with $z_{1-\alpha}$.

By rearranging the following equation and taking account of the lower tail in the 2-sided case (this is ignored in the sample size approximations for convenience), the power becomes the following:

$$1 - \beta = \Phi\left(\left(\frac{|p_1 - p_2|\sqrt{k}}{\sqrt{Var(d_j)}}\right) - z_{1-\frac{a}{2}}\right) + \Phi\left(-\left(\frac{|p_1 - p_2|\sqrt{k}}{\sqrt{Var(d_j)}}\right) - z_{1-\frac{a}{2}}\right) \qquad \{5.8.3\}$$

Note that in the one-sided case, $z_{1-\alpha/2}$ is replaced with $z_{1-\alpha}$ and the second term in equation 5.8.3 is not calculated. Note that the power row in the table is automatically updated based on this equation after the other closed form equations are used.

Rearrangement of equation 5.8.1 and 5.8.2 can be used to generate the following estimates of the sample size per cluster and between cluster variance:

$$m = \frac{\left(z_{1-\beta} + z_{1-\alpha/2}\right)^2 [p_1(1-p_1) + p_2(1-p_2)]}{\left[(k(p_1 - p_2)^2) - (2\sigma_B^2\left(z_{1-\beta} + z_{1-\alpha/2}\right)^2)\right]} \qquad \{5.8.4\}$$

$$\sigma_B^2 = \frac{\left[(k(p_1 - p_2)^2) - \left(\frac{[p_1(1-p_1) + p_2(1-p_2)]\left(z_{1-\beta} + z_{1-\alpha/2}\right)^2}{m}\right)\right]}{2\left(z_{1-\beta} + z_{1-\alpha/2}\right)^2} \qquad \{5.8.5\}$$

A closed form equation is not used to calculate the value of the test statistic under the alternative hypothesis. Instead a search algorithm is used. The search algorithm calculates power at various values for the relevant parameter until the desired power is reached. Note that values are searched for as the appropriate difference and then converted into the correct statistic in the table.

5.8.3 Examples

Example 1: Validation example calculating required number of clusters for a given power
The following example will look at a number of clusters calculation problem from Donner and Klar (2000) and then a sensitivity analysis is conducted to show the effect of changing the between cluster variance on the number of cluster pairs required. The following steps outline the procedure for this example:

1. Open nQuery + nTerim 4.0 via the start menu or desktop shortcut. Click “New Fixed Term Test” from the top of the window. The “Study Goal and Design” window will appear as Figure 5.7.1 below.

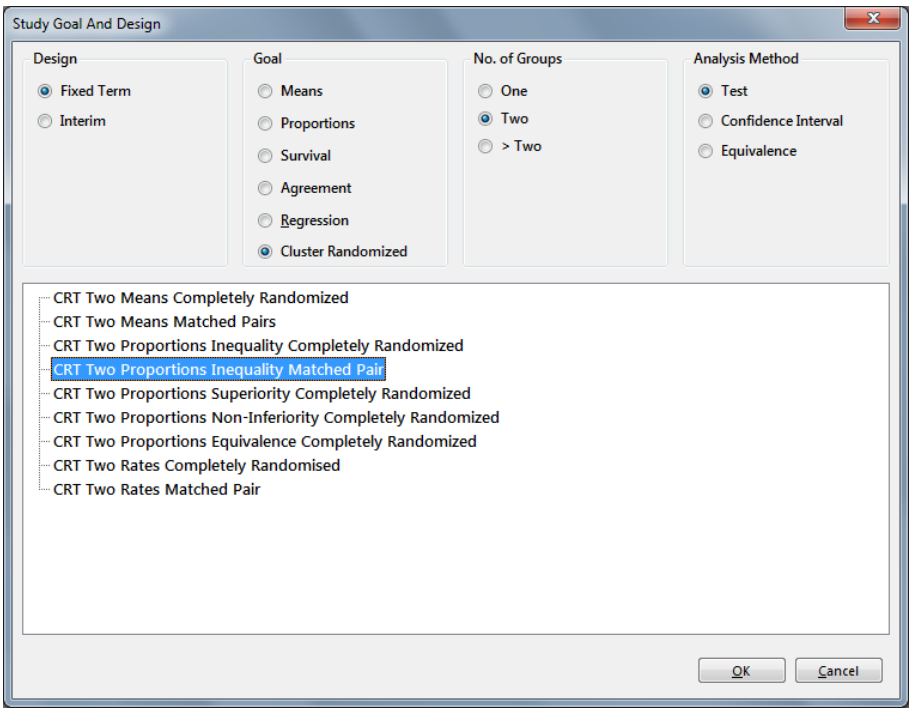


Figure 5.8.1. Study Goal and Design Window

2. In order to select the CRT Two Mean Matched-Pair design table, navigate through the Study Goal and Design Window by selecting "Fixed Term" as the Design, "Cluster Randomized" as the Goal, "Two" as the Number of Groups and "Test" as the Analysis Method. Then click "OK" and the test window will appear as per Figure 5.7.1.

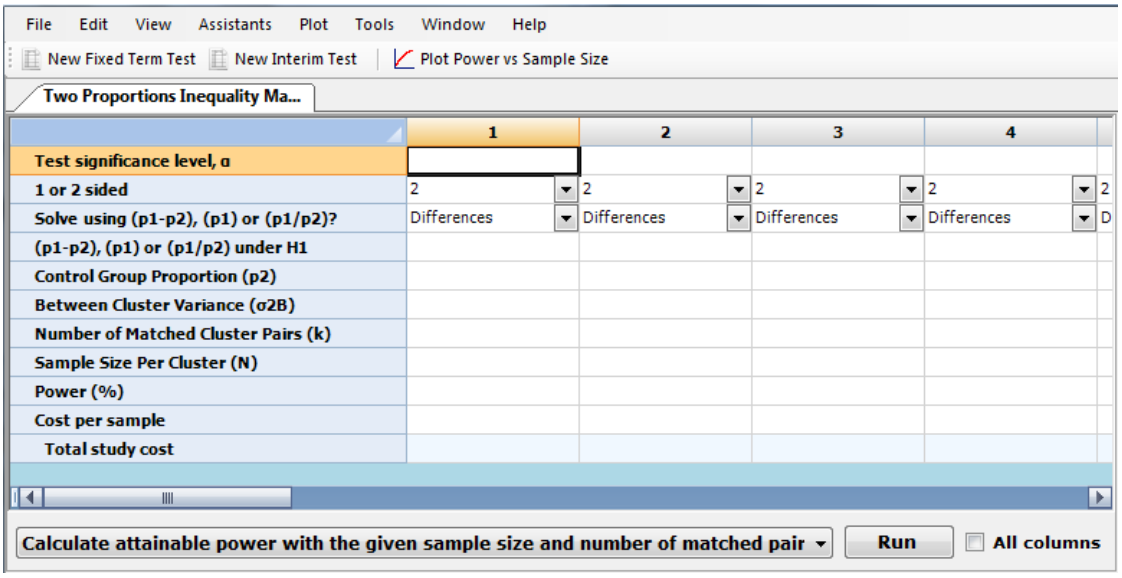


Figure 5.8.2. CRT Two Proportions Inequality Matched Pair Test Table

The first calculation will be for the "Calculate Required Number of Matched Pairs of Clusters for Given power and Sample Size" calculation using the example from Donner and Klar. This example can be undertaken using the following steps.

3. First enter 0.05 for the Test Significance level row. Next, select "1" from the 1 or 2 sided option and "Proportions" from Solve using (p1-p2), (p1) or (p1/p2).

4. Next, enter 0.25 for the "(p1-p2), (p1) or (p1/p2) under H1" row. With "Proportions" selected from row three, this will set the treatment proportion to be 0.25.

5. Next, enter 0.15 in the "Control Group Proportion (p2)" row and 0.00318 in the "Between Cluster Variance" row.

6. Next, enter 500 for the "Sample Size per Cluster (N)" row and 90 for the "Power (%)" row. This will give inputs as per Figure 5.7.3 below.

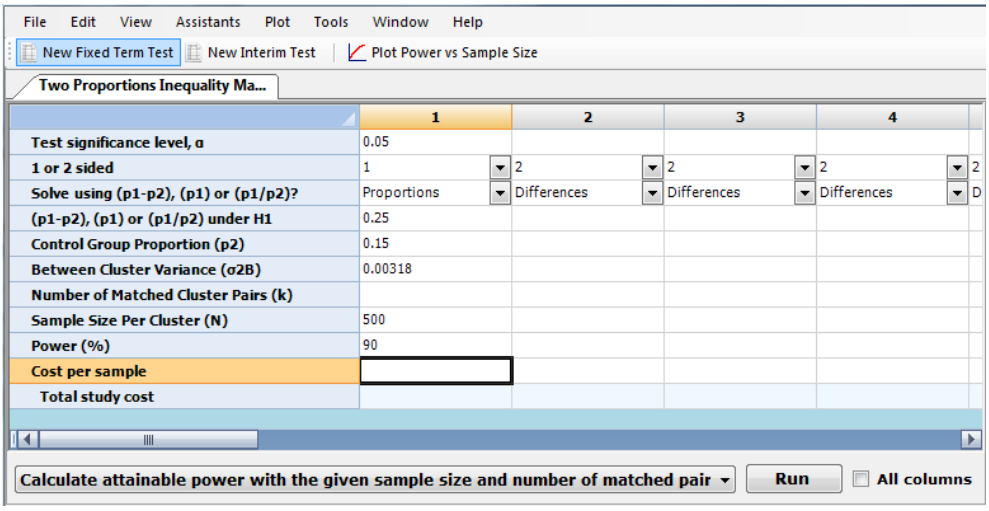
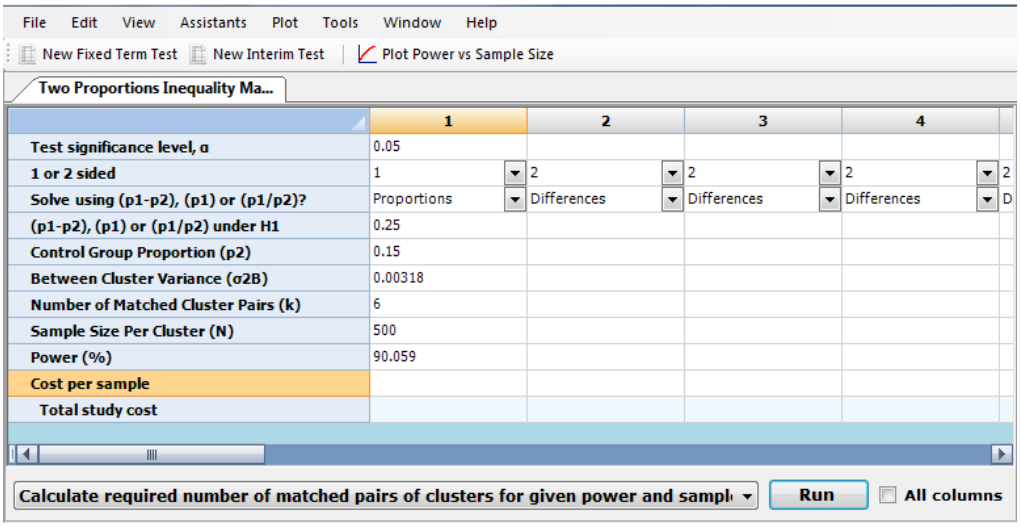


Figure 5.7.3. Values entered for Two Proportions Inequality Matched Pair study design

7. Select “Calculate Required Number of Matched Pairs of Clusters for Given power and Sample Size” from the dropdown menu beside the “Run” button. Then click “Run”. This will give a result of 6 pairs of clusters as per Figure 5.7.4. This agrees with the value of 6 pairs in Donner and Klar.



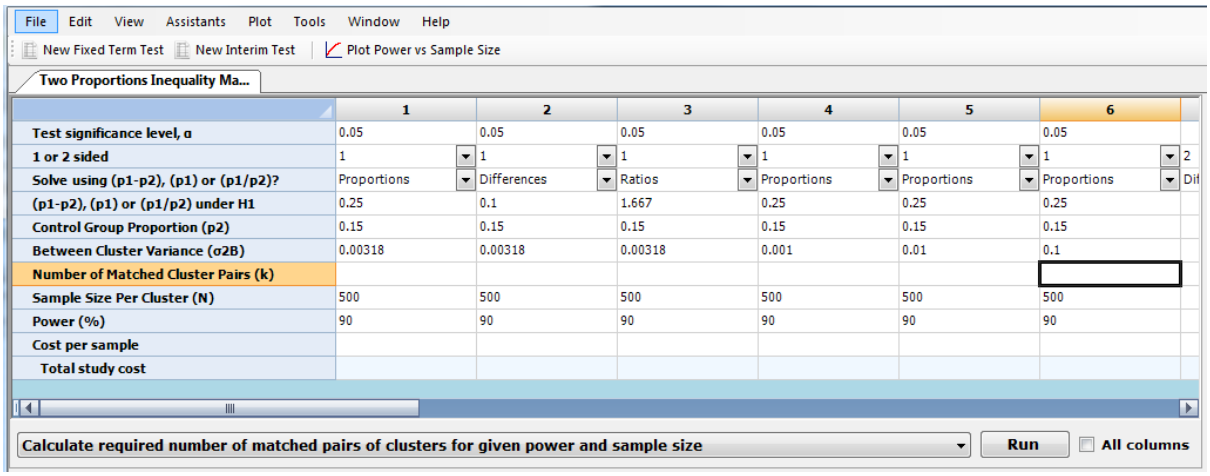
	1	2	3	4
Test significance level, α	0.05			
1 or 2 sided	1	2	2	2
Solve using (p1-p2), (p1) or (p1/p2)?	Proportions	Differences	Differences	Differences
(p1-p2), (p1) or (p1/p2) under H1	0.25			
Control Group Proportion (p2)	0.15			
Between Cluster Variance (σ2B)	0.00318			
Number of Matched Cluster Pairs (k)	6			
Sample Size Per Cluster (N)	500			
Power (%)	90.059			
Cost per sample				
Total study cost				

Figure 5.7.4. Completed Two Proportions Inequality Matched Pair study design

The next calculation is a sensitivity analysis for the number of paired clusters when the “Between-Cluster Variance” is changed and also a proof of concept of the equivalence of the difference, treatment proportion and ratio methods.

8. Delete the value for “k” in the first column then replace the updated power with 90. Then copy the first column into columns 2 to 6.

9. In column two, select “Differences” from row 3 and replace row four with 0.1 ($p1 - p2 = 0.25 - 0.15 = 0.1$). In column three, select “Ratios” from row 3 and replace row four with 1.667 ($p1/p2 = 0.25/0.15 = 1.667$). In columns four to six, replace the between cluster variance with 0.001, 0.01 and 0.1 respectively. This will give a table as per Figure 5.7.5.

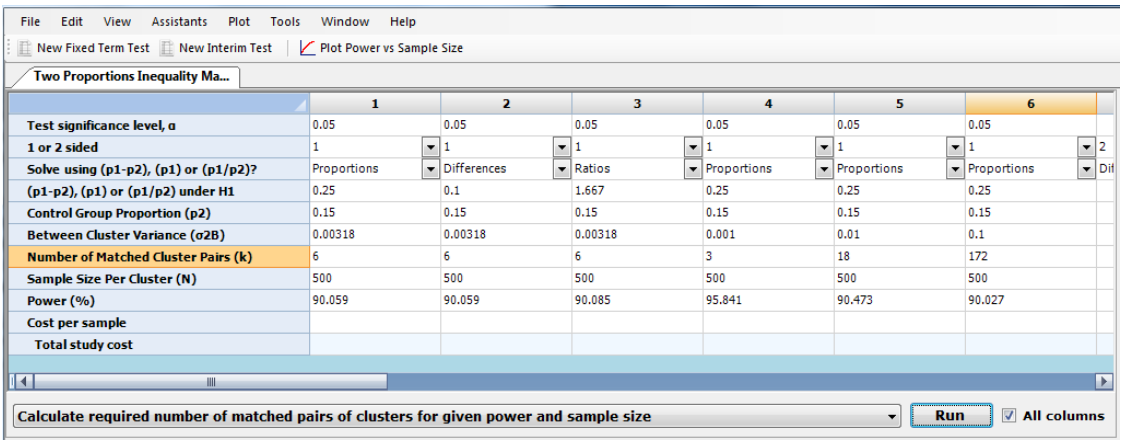


	1	2	3	4	5	6
Test significance level, α	0.05	0.05	0.05	0.05	0.05	0.05
1 or 2 sided	1	1	1	1	1	1
Solve using (p1-p2), (p1) or (p1/p2)?	Proportions	Differences	Ratios	Proportions	Proportions	Proportions
(p1-p2), (p1) or (p1/p2) under H1	0.25	0.1	1.667	0.25	0.25	0.25
Control Group Proportion (p2)	0.15	0.15	0.15	0.15	0.15	0.15
Between Cluster Variance (σ2B)	0.00318	0.00318	0.00318	0.001	0.01	0.1
Number of Matched Cluster Pairs (k)						
Sample Size Per Cluster (N)	500	500	500	500	500	500
Power (%)	90	90	90	90	90	90
Cost per sample						
Total study cost						

Figure 5.7.5. Sensitivity analysis for Between-Cluster Variance

10. Select “Calculate Required Number of Matched Pairs of Clusters for Given power and Sample Size” from the dropdown menu beside the “Run” button. Then tick the box for “All Columns”. Then click “Run”.

This will give the resultant values for “k” of 6, 6, 6, 3, 18 and 172 sequentially as in Figure 5.7.6. Note that the three first columns give effectively the same result (there is a slight difference in exact power due to rounding of (0.25/0.15)). Note also as the greater the variance between matched pairs (i.e. increasing dissimilarity between paired clusters), the greater the required number of pairs due to reduced power from the matching effect.



	1	2	3	4	5	6
Test significance level, α	0.05	0.05	0.05	0.05	0.05	0.05
1 or 2 sided	1	1	1	1	1	1
Solve using (p1-p2), (p1) or (p1/p2)?	Proportions	Differences	Ratios	Proportions	Proportions	Proportions
(p1-p2), (p1) or (p1/p2) under H1	0.25	0.1	1.667	0.25	0.25	0.25
Control Group Proportion (p2)	0.15	0.15	0.15	0.15	0.15	0.15
Between Cluster Variance (σ2B)	0.00318	0.00318	0.00318	0.001	0.01	0.1
Number of Matched Cluster Pairs (k)	6	6	6	3	18	172
Sample Size Per Cluster (N)	500	500	500	500	500	500
Power (%)	90.059	90.059	90.085	95.841	90.473	90.027
Cost per sample						
Total study cost						

Figure 5.7.6. Results from Sensitivity analysis

11. By clicking on the desired study design column and going to the “Output” tab at the bottom of the screen, you can get an output statement for the results given by the current calculation as per Figure 5.7.7.

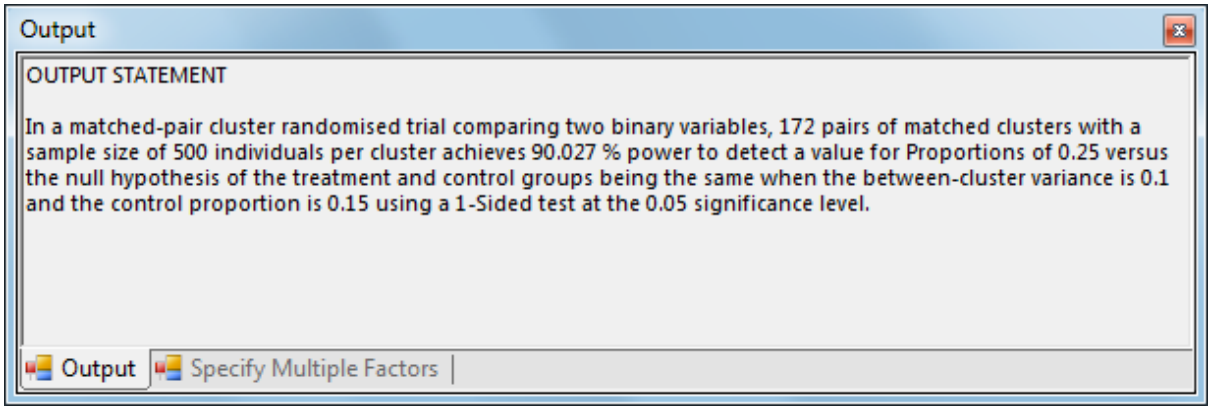


Figure 5.7.7. Study design Output statement

This Output statement can then be easily transferred directly from the output window into a report.

5.9 CRT Two Incidence Rates – Matched-Pair

5.9.1 Introduction

Incidence data is often obtained in a variety of clinical and field research activities such as the number of accidents at a junction or number of occurrences of a disease in a year. Clustered data is very common in a wide variety of academic, social policy and economic studies. Researchers often plan studies in terms of years of exposure rather than directly through sample size.

This table facilitates the calculation of the power and sample size for hypothesis tests comparing two incidence rates in a matched-pair cluster randomised trial. Power and sample size is computed using the method outlined by Donner and Klar (2000).

5.9.2 Methodology

This table provides sample size and power calculations for studies which will be comparing two incidence rates which use a matched-pair cluster randomisation study design. A matched pair design assigns clusters into pairs depending on their homogeneity which is based on relevant information known for each cluster e.g. size of school, GDP of country. This can lead to increases in power when compared to the fully randomised design if the matching criteria is associated with the rate. This table assumes a balanced study design.

The sample size calculation for two incidence rates is taken from Donner and Klar (2000) and is an extension of the method for the completely randomised design. It uses the average correlation between the true incidence rates in the control and treatment groups for the ith matched-pair of clusters to adjust the coefficient of variation, which is equal to the between cluster standard deviation divided by the mean rate, which is used to adjust for the effect of within cluster correlation. It is assumed the coefficient of variation is the same in both groups.

This table can be used to calculate the power, the number of matched cluster pairs, the sample size per cluster, the fixed observation period per subject, the coefficient of variation and the smallest detectable difference given all other terms in the table are specified.

To calculate power, the number of clusters or the sample size per cluster, the user must specify the test significance level (α), whether to use a one or two sided test, the control group incidence rate (λ_1), the treatment group incidence rate (λ_2), the difference in rates ($\lambda_1 - \lambda_2$), the fixed observation time per subject (t), the average correlation between matched pairs of cluster (ρ) and the coefficient of variation (CV).

The formulae use the normal approximation to produce estimates. The formula for the number of matched cluster pairs is given by the following:

$$k = \frac{\left(z_{1-\beta} + z_{1-\alpha/2}\right)^2 (\lambda_1 + \lambda_2) IF}{yt(\lambda_1 - \lambda_2)^2} \qquad \{5.9.1\}$$

where

$$IF = 1 + \frac{\left(CV(1-\rho)\right)^2(\lambda_1^2 + \lambda_2^2)yt}{(\lambda_1 + \lambda_2)} \qquad \{5.9.2\}$$

where k is the number of clusters pairs and y is the sample size per cluster. For the one-sided test, $z_{1-\alpha/2}$ is replaced with $z_{1-\alpha}$. Simple rearrangement of the above formula gives the following equation for the power:

$$1 - \beta = \Phi\left(\frac{\sqrt{ytk(\lambda_1 - \lambda_2)^2}}{\sqrt{(\lambda_1 + \lambda_2)IF}} - z_{1-\alpha/2}\right) \qquad \{5.9.3\}$$

Rearrangement can be used to yield the following formula for the sample size y:

$$y = \frac{\left[\frac{\left(z_{1-\beta} + z_{1-\alpha/2}\right)^2(\lambda_1 + \lambda_2)}{k(\lambda_1 - \lambda_2)^2}\right]}{t\left(1 - \frac{\left(z_{1-\beta} + z_{1-\alpha/2}\right)^2(\lambda_1 + \lambda_2)IF_m}{k(\lambda_1 - \lambda_2)^2}\right)} \qquad \{5.9.4\}$$

A similar rearrangement yields the following formula for the observation time per subject:

$$t = \frac{\left[\frac{\left(z_{1-\beta} + z_{1-\alpha/2}\right)^2(\lambda_1 + \lambda_2)}{k(\lambda_1 - \lambda_2)^2}\right]}{y\left(1 - \frac{\left(z_{1-\beta} + z_{1-\alpha/2}\right)^2(\lambda_1 + \lambda_2)IF_m}{k(\lambda_1 - \lambda_2)^2}\right)} \qquad \{5.9.5\}$$

Where

$$IF_m = \frac{\left(CV(1-\rho)\right)^2(\lambda_1^2 + \lambda_2^2)}{(\lambda_1 + \lambda_2)} \qquad \{5.9.6\}$$

The formula for the co-efficient of variation and correlation between matched pairs can be shown to be the following:

$$CV = \frac{\sqrt{\left(\frac{(\lambda_1+\lambda_2)}{yt(\lambda_1^2+\lambda_2^2)}\right)\left(\frac{ykt(\lambda_1-\lambda_2)^2}{\left(z_{1-\beta}+z_{1-\alpha/2}\right)^2(\lambda_1+\lambda_2)}-1\right)}}{(1-\rho)} \quad \{5.9.7\}$$

$$\rho = -\frac{\sqrt{\left(\frac{(\lambda_1+\lambda_2)}{yt(\lambda_1^2+\lambda_2^2)}\right)\left(\frac{ykt(\lambda_1-\lambda_2)^2}{\left(z_{1-\beta}+z_{1-\alpha/2}\right)^2(\lambda_1+\lambda_2)}-1\right)}}{CV}+1 \quad \{5.9.8\}$$

A closed form equation is not used to calculate the smallest detectable difference. Instead a search algorithm is used. The search algorithm calculates power at various values for the treatment group incidence rate until the desired power is reached.

5.9.3 Examples

Example 1: Validation example calculating required number of clusters for a given power
Donner and Klar (2000) do not provide a validation example thus the example below is based on a study design from Picado, Albert, et al. (2010). A sensitivity analysis is conducted to show the effect of changing the correlation between paired clusters on the number of clusters required. The following steps outline the procedure for this example:

1. Open nQuery + nTerim 4.0 via the start menu or desktop shortcut. Click "New Fixed Term Test" from the top of the window. The "Study Goal and Design" window will appear as Figure 5.9.1 below.

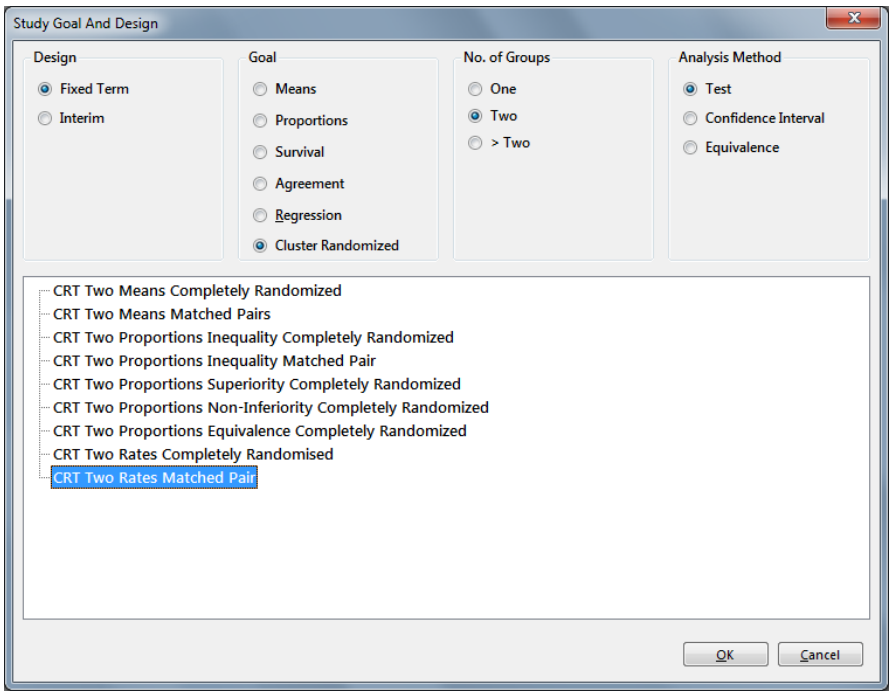


Figure 5.9.1. Study Goal and Design Window

2. In order to select the CRT Two Incidence Rates Completely Randomized design table, navigate through the Study Goal and Design Window by selecting "Fixed Term" as the Design, "Cluster Randomized" as the Goal, "Two" as the Number of Groups and "Test" as the Analysis Method. Then click "OK" and the test window will appear as per Figure 5.9.2.

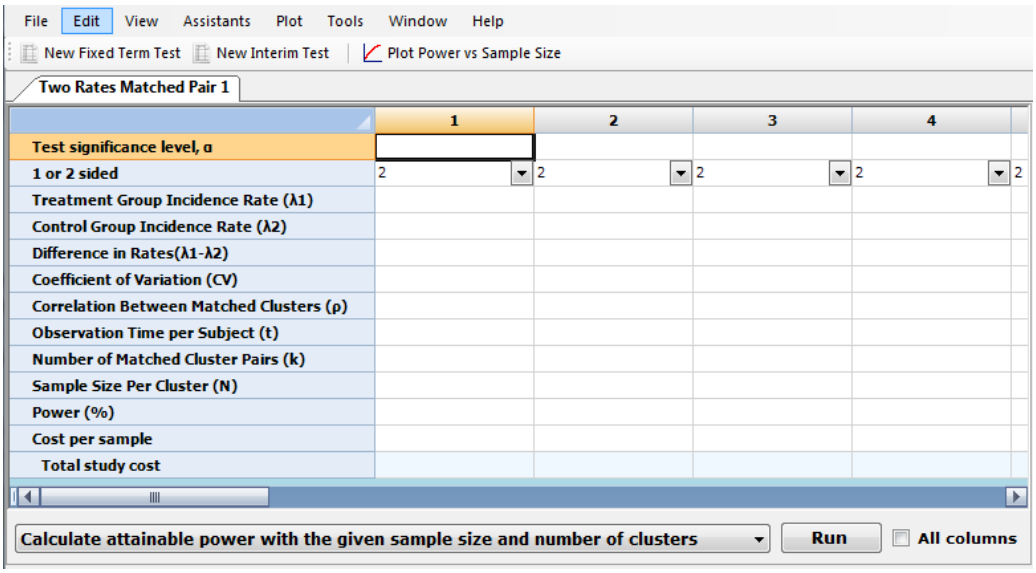


Figure 5.9.2. CRT Two Rates Matched Pair Test Table

The first calculation will be for the "Calculate Required Number of Clusters for Given power and Sample Size" calculation using the example from Picado, Albert, et al. (2010). This example can be undertaken using the following steps.

3. First enter 0.05 for the Test Significance level row. Next, select "2" from the 1 or 2 sided option.

4. Next, enter 0.01 and 0.02 for the "Treatment Group Incidence Rate (λ1)" and "Control Group Incidence (λ2)" rows respectively. The "Difference in Rates" row will populate automatically.

5. Next, enter 0.5 in the "Coefficient of Variation (CV)" row, 0.5 for the "Correlation between Matched Clusters (ρ)" and 2 in the "Observation Time (t)" row.

Note: Study design only specifies value of matched-pair adjusted CV which is equal to CV(1-ρ) and is 0.25 in the study design.*

6. Next, enter 500 for the "Sample Size per Cluster (N)" row and 90 for the "Power (%)" row. This will give inputs as per Figure 5.9.3 below.

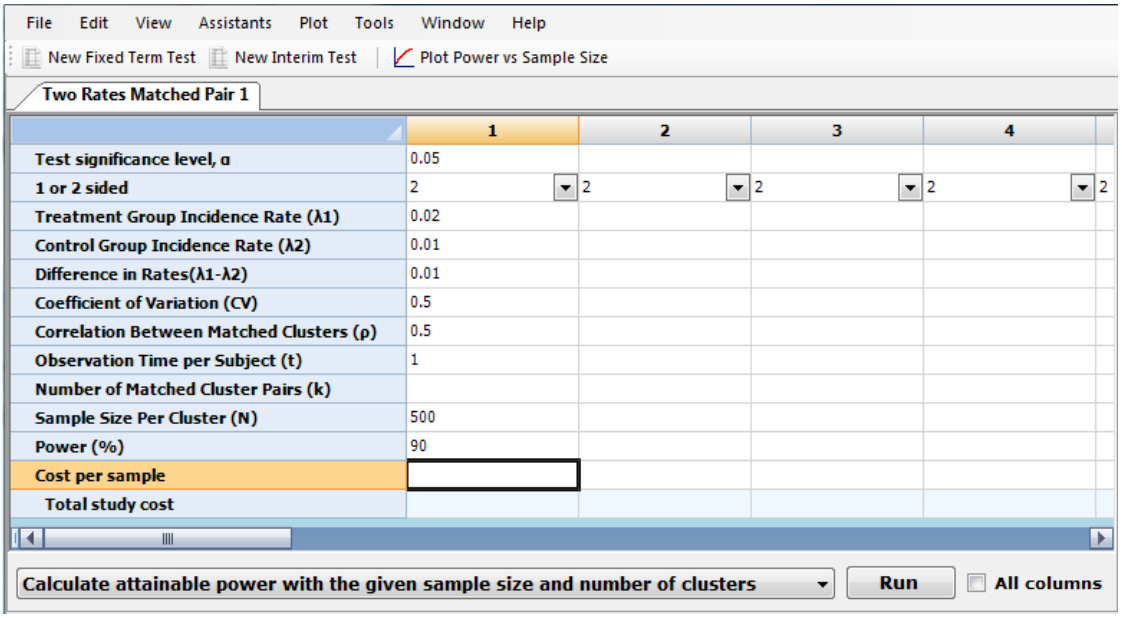


Figure 5.9.3. Values entered for CRT Two Rates Matched Pair study design

7. Select "Calculate Required Number Paired Clusters for Given power and Sample Size" from the dropdown menu beside the "Run" button. Then click "Run". This will give a result of 10 cluster pairs as in Figure 5.9.4. This agrees with the value from Picado, Albert, et al..

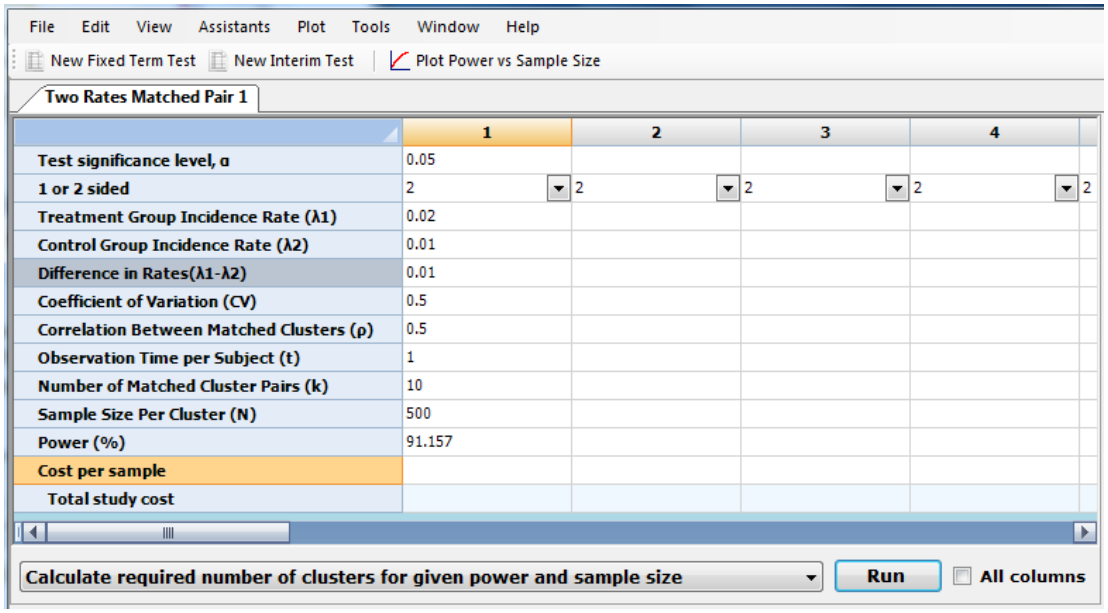
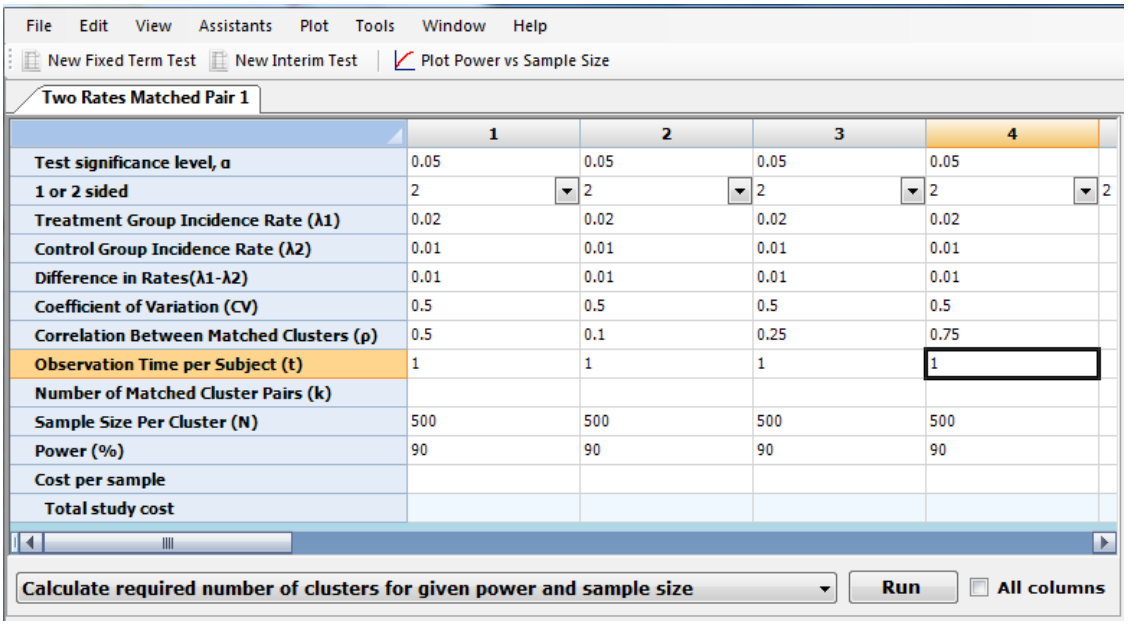


Figure 5.9.4. Completed CRT Two Rates Matched Pair study design

The next calculation is a sensitivity analysis for the number of paired clusters when the correlation between paired clusters is changed.

8. Delete the value for "k" in the first column then replace the updated power with 90. Then copy the first column into columns 2 to 4.

9. Enter 0.1 in column 2 for “Correlation between Matched Clusters”, 0.25 in column 3 and 0.75 in column four. This will give a table as per Figure 5.9.5.

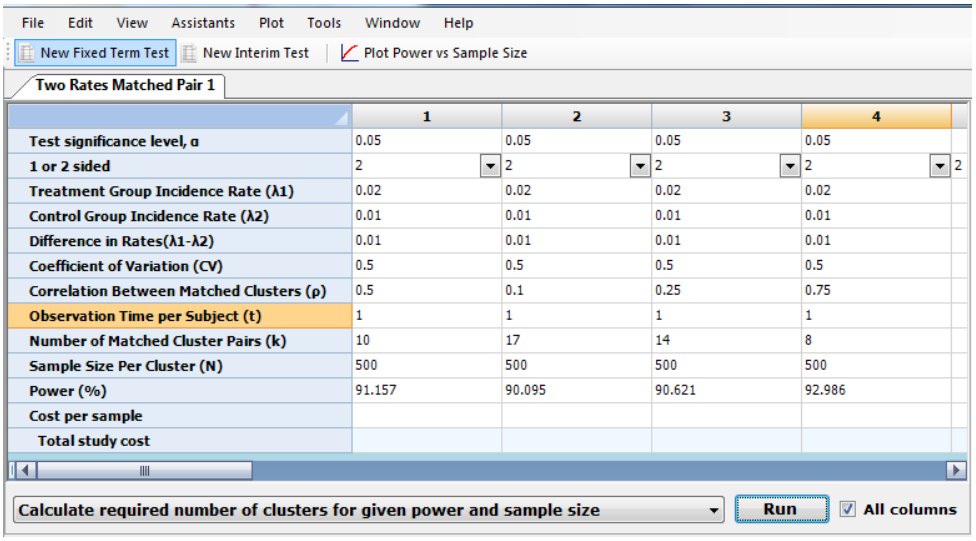


	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
1 or 2 sided	2	2	2	2
Treatment Group Incidence Rate (λ1)	0.02	0.02	0.02	0.02
Control Group Incidence Rate (λ2)	0.01	0.01	0.01	0.01
Difference in Rates(λ1-λ2)	0.01	0.01	0.01	0.01
Coefficient of Variation (CV)	0.5	0.5	0.5	0.5
Correlation Between Matched Clusters (ρ)	0.5	0.1	0.25	0.75
Observation Time per Subject (t)	1	1	1	1
Number of Matched Cluster Pairs (k)				
Sample Size Per Cluster (N)	500	500	500	500
Power (%)	90	90	90	90
Cost per sample				
Total study cost				

Figure 5.9.5. Sensitivity analysis for Correlation Between Matched Clusters

10. Select “Calculate Required Number of Clusters for Given power and Sample Size” from the dropdown menu beside the “Run” button. Then tick the box for “All Columns”. Then click “Run”.

This will give the resultant values for “k” of 10, 17, 14 and 8 sequentially as in Figure 5.9.6.



	1	2	3	4
Test significance level, α	0.05	0.05	0.05	0.05
1 or 2 sided	2	2	2	2
Treatment Group Incidence Rate (λ1)	0.02	0.02	0.02	0.02
Control Group Incidence Rate (λ2)	0.01	0.01	0.01	0.01
Difference in Rates(λ1-λ2)	0.01	0.01	0.01	0.01
Coefficient of Variation (CV)	0.5	0.5	0.5	0.5
Correlation Between Matched Clusters (ρ)	0.5	0.1	0.25	0.75
Observation Time per Subject (t)	1	1	1	1
Number of Matched Cluster Pairs (k)	10	17	14	8
Sample Size Per Cluster (N)	500	500	500	500
Power (%)	91.157	90.095	90.621	92.986
Cost per sample				
Total study cost				

Figure 5.9.6. Results from Sensitivity analysis

11. By clicking on the desired study design column and going to the “Output” tab at the bottom of the screen, you can get an output statement for the results given by the current calculation as per Figure 5.9.7.

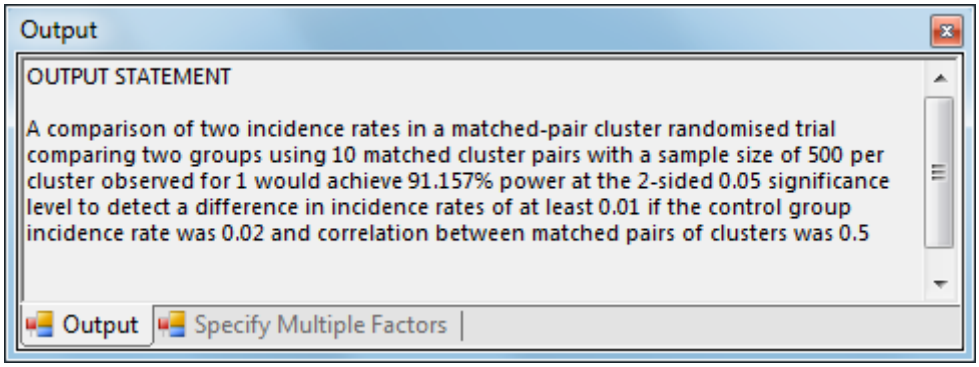


Figure 5.9.7. Study design Output statement

This Output statement can then be easily transferred directly from the output window into a report.

Chapter 6

References

Bain, L., & Englehardt, M., (1991). “Statistical Analysis of Reliability and Life-Testing Models: Theory and Methods”, CRC Press.

Chen, K. W., Chow, S. C., & Li, G., (1997) “A Note on Sample Size Determination for Bioequivalence Studies with Higher-order Crossover Designs”, Journal of Pharmacokinetics and Biopharmaceutics, 25(6), pp 753-765.

Chow, S. C., & Liu, J. P. (2000)., “Design and Analysis of Bioavailability and Bioequivalence Studies” CRC Press

Chow, S.C., Shao, J., and Wang, H., (2008) “Sample Size Calculations in Clinical Research”, Second Edition, Chapman & Hall

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, pp 2315-2338

DeMets, D.L., and Lan, K.K.G., (1994) “Interim Analysis: The Alpha Spending Function Approach”, Statistics in Medicine, 13, pp 1341-1352

Desu, M. M., & Raghavarao, D., (1990). “Sample Size Methodology”, Academic Press

Donner, A., & Klar, N., (1996) "Statistical Considerations in the Design and Analysis of Community Intervention Trials", Journal of Clinical Epidemiology, 49(4), pp 435-439

Donner, A., & Klar, N., (2000) “Design and Analysis of Cluster Randomization Trials in Health Research”, Arnold Publishers, London

Fleiss, J.L., Tytun, A., Ury, S.H.K., (1980) “A Simple Approximation for Calculating Sample Sizes for Comparing Independent Proportions”, Biometrics, 36, pp 343-346

Fleiss, J.L., (1981) “Statistical Methods for Rates and Proportions”, Second Edition, Wiley

Gu, K., Ng, H.K.T., Tang, M.L., & Schucany, W.R., (2008) “Testing the Ratio of Two Poisson Rates”, Biometrical Journal, 50(2), pp 283-298

Guenther, W.C., (1977) “Sampling Inspection in Statistical Quality Control”, Charles Griffin and Company Limited, pp 25–30

Hanley, J. A., & McNeil, B. J., (1982) “The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve”, Radiology, 143(1), pp 29-36.

Hanley, J. A., & McNeil, B. J., (1983) “A Method of Comparing the Areas under Receiver Operating Characteristic Curves Derived from the Same Cases”, Radiology, 148(3), pp 839-843

Hayes, R. J., & Bennett, S., (1999) “Simple Sample Size Calculation for Cluster-randomized Trials”, International Journal of Epidemiology, 28(2), pp 319-326.

Hsieh, F.Y., & Lavori, P.W., (2000) “Sample-size calculations for the Cox proportional hazards regression model with nonbinary covariates”, Controlled Clinical Trials, 21(6), pp 552-560

Huffman, M.D., (1984) “An Improved Approximate Two-sample Poisson Test”, Applied Statistics, 33(2), pp 224-226

Hwang, I,K., Shih, W.J., and deCani, J.S., (1990) “Group Sequential Designs using a Family Type I Error Probability Spending Functions”, Statistics in Medicine, 9, pp 1439-1445

Jennison, C., and Turnbull, B.W., (2000) “Group Sequential Methods with Applications to Clinical Trials”, Chapman & Hall

Julious, S. A. (2004)., “Sample Sizes for Clinical Trials with Normal Data” Statistics in Medicine, 23(12), pp 1921-1986.

Keppel, G., (1991) “Design and Analysis - A Researcher's Handbook”, Third Edition, Prentice Hall

Kodell, R.L., Lensing, S.Y., Landes, R.D., Kumar, K.S., & Hauer-Jensen, M., (2010) “Determination of Sample Sizes for Demonstrating Efficacy of Radiation Countermeasures”, Biometrics, 66(1), pp 239-248

Li, J., & Fine, J., (2004) “On Sample Size for Sensitivity and Specificity in Prospective Diagnostic Accuracy Studies”, Statistics in Medicine, 23(16), pp 2537-2550.

Liu, H.H., Wu, T.T., (2005) “Sample Size Calculation and Power Analysis for Time-Averaged Difference”, Journal of Modern Applied Statistical Methods, 4(2), pp 434-445

Machin, D., Campbell, M. J., Tan, S. B., & Tan, S. H. (2009)., “Sample Size Tables for Clinical Studies”, John Wiley & Sons

Mathews, P. (2010)., “Sample Dize Calculations: Practical Methods for Engineers and Scientists”, Mathews, Malnar and Bailey

Muller, K.E., and Barton, C.N., (1989) “Approximate Power for Repeated Measures ANOVA Lacking Sphericity”, Journal of the American Statistical Association, 84, pp 549–555 (with correction in volume 86 (1991), pp 255–256)

Muller, K.E., LaVange, L.M., 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, 87, pp 1209–1226

O’Brien, P.C., and Fleming, T.R., (1979) “A Multipe Testing Procedure for Clinical Trials”,

Biometrika, 35, pp 549-556

O'Brien, R.G., Muller, K.E., (1993) "Unified Power Analysis for t-tests through Multivariate Hypotheses", Edwards, L.K., (Ed.), Applied Analysis of Variance in Behavioral Science, Marcel Dekker, pp 297-344

Obuchowski, N. A., & McClish, D. K., (1997). "Sample Size Determination for Diagnostic Accuracy Studies involving Binormal ROC Curve Indices." Statistics in Medicine, 16(13), pp 1529-1542.

Ostle, B., Malone, L., (1988) "Statistics in Research: Basic Concepts and Techniques for Research Workers", Fourth Edition, Iowa State Press, Ames, Iowa

Overall, J.E., Doyle, S.R., (1994) "Estimating Sample Sizes for Repeated Measures Designs", Controlled Clinical Trials, 15, pp 100-123

Pampallona, S., Tsiatis, A.A., and Kim, K., (1995) "Spending functions for type I and type II error probabilities of group sequential trials", Technical report, Dept. of Biostatistics, Harvard School of Public Health, Boston

Pampallona, S., Tsiatis, A.A., and Kim, K., (2001) "Interim monitoring of group sequential trials using spending functions for the type I and type II error probabilities", Drug Information Journal, 35, pp 1113-1121

Picado, A., Singh, S. P., Rijal, S., Sundar, S., Ostyn, B., Chappuis, F., ... & Boelaert, M., (2010) "Longlasting Insecticidal Nets for Prevention of *Leishmania Donovani* Infection in India and Nepal: Paired Cluster Randomised Trial", BMJ, 341.

Pocock, S.J., (1977) "Group Sequential Methods in the Design and Analysis of Clinical Trials", Biometrika, 64, pp 191-199

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 Winconsin-Madison

Signorini, D.F., (1991) "Sample size for Poisson regression", Biometrika, 78(2), pp 446-450

Sivakumar, B., Vijayaraghavan, K., Vazir, S., Balakrishna, N., Shatrugna, V., Sarma, K. V. R., ... & Krishnaswamy, K., (2006) "Effect of Micronutrient Supplement on Health and Nutritional Status of Schoolchildren: Study Design" Nutrition, 22(1), S1-S7.

Smith, P. G., & Morrow, R.H., (1996) "Field Trials of Health Interventions in Developing Countries: A Toolbox", Second Edition, Macmillan, Malaysia

Ryan, Thomas P., (2013) "Sample size determination and power", John Wiley & Sons

Zar, J. H., (2010) "Biostatistical Analysis", 5th Edition, Prentice Hall

Zhu, H., & Lakkis, H., (2014) “Sample size calculation for comparing two negative binomial rates”, Statistics in Medicine, 33(3), pp 376-387

North/Central/South America & Canada
Statistical Solutions,
One International Place,
100 Oliver Street, Suite 1400,
Boston, MA 02110

Tel: (617) 535 7677
Fax: (617) 535 7717
Email: sales@statsols.com

Europe, Middle East, Africa & Asia
Statistical Solutions,
4500 Airport Business Park,
Cork,
Rep. of Ireland

Tel: +353 21 4839100
Fax: + 353 21 4840026
Email: sales@statsols.com