Chapter 211

Non-Inferiority Tests for the Ratio of Two Proportions

Introduction

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

The power calculations assume that independent, random samples are drawn from two populations.

Example

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

Because of the many benefits of the new treatment, clinicians are willing to adopt the new treatment even if it is slightly less effective than the current treatment. They must determine, however, how much less effective the new treatment can be and still be adopted. Should it be adopted if 69% respond? 68%? 65%? 60%? There is a percentage below 70% at which the ratio of the two treatments is no longer considered ignorable. After thoughtful discussion with several clinicians, it was decided that if the response rate ratio is no less than 0.9, the new treatment would be adopted. This ratio is called the *margin of non-inferiority*. The margin of non-inferiority in this example is 0.9.

The developers must design an experiment to test the hypothesis that the response rate ratio of the new treatment to the standard is at least 0.9. The statistical hypothesis to be tested is

$$H_0: p_1/p_2 \leq 0.9 \quad \text{versus} \quad H_1: p_1/p_2 > 0.9$$

Notice that when the null hypothesis is rejected, the conclusion is that the response rate ratio is at least 0.9. Note that even though the response rate of the current treatment is 0.70, the hypothesis test is about a response rate ratio of 0.9. Also notice that a rejection of the null hypothesis results in the conclusion of interest.

Technical Details

The details of sample size calculation for the two-sample design for binary outcomes are presented in the chapter "Tests for Two Proportions," and they will not be duplicated here. Instead, this chapter only discusses those changes necessary for non-inferiority tests.

This procedure has the capability for calculating power based on large sample (normal approximation) results and based on the enumeration of all possible values in the binomial distribution.

Suppose you have two populations from which dichotomous (binary) responses will be recorded. Assume without loss of generality that the higher proportions are better. The probability (or risk) of cure in population 1 (the treatment group) is p_1 and in population 2 (the reference group) is p_2. Random samples of n_1 and n_2 individuals are obtained from these two populations. The data from these samples can be displayed in a 2-by-2 contingency table as follows

Group	Success	Failure	Total
Treatment	x_{11}	x_{12}	n_1
Control	x_{21}	x_{22}	n_2
Totals	m_1	m_2	N

The binomial proportions, p_1 and p_2, are estimated from these data using the formulae

$$\hat{p}_1 = \frac{a}{m} = \frac{x_{11}}{n_1} \quad \text{and} \quad \hat{p}_2 = \frac{b}{n} = \frac{x_{21}}{n_2}$$

Let $p_{1.0}$ represent the group 1 proportion tested by the null hypothesis, H_0. The power of a test is computed at a specific value of the proportion which we will call $p_{1.1}$. Let ϕ_0 represent the smallest ratio (margin of non-inferiority) between the two proportions that still results in the conclusion that the new treatment is not inferior to the current treatment. For a non-inferiority test, $\phi_0 < 1$ The set of statistical hypotheses that are tested is

$$H_0: p_1/p_2 \le \phi_0 \quad \text{versus} \quad H_1: p_1/p_2 > \phi_0$$

which can be rearranged to give

$$H_0: p_1 \le p_2\phi_0 \quad \text{versus} \quad H_1: p_1 > p_2\phi_0$$

There are three common methods of specifying the margin of non-inferiority. The most direct is to simply give values for p_2 and $p_{1.0}$. However, it is often more meaningful to give p_2 and then specify $p_{1.0}$ implicitly by specifying the difference, ratio, or odds ratio. Mathematically, the definitions of these parameterizations are

Parameter	Computation	Hypotheses
Difference	$\delta_0 = p_{1.0} - p_2$	$H_0: p_1 - p_2 \le \delta_0$ versus $H_1: p_1 - p_2 > \delta_0$
Ratio	$\phi_0 = p_{1.0}/p_2$	$H_0: p_1/p_2 \le \phi_0$ versus $H_1: p_1/p_2 > \phi_0$
Odds Ratio	$\psi_0 = O_{1.0}/O_2$	$H_0: O_1/O_2 \le \psi_0$ versus $H_1: O_1/O_2 > \psi_0$

Ratio

The ratio, $\phi = p_1/p_2$, gives the relative change in the probability of the response. Testing non-inferiority uses the formulation

$$H_0\colon p_1/p_2 \le \phi_0 \quad \text{versus} \quad H_1\colon p_1/p_2 > \phi_0$$

or equivalently

$$H_0\colon \phi \le \phi_0 \quad \text{versus} \quad H_1\colon \phi > \phi_0.$$

For non-inferiority tests with higher proportions better, $\phi_0 < 1$. For non-inferiority tests with higher proportions worse, $\phi_0 > 1$.

Non-Inferiority

The following example might help you understand the concept of *non-inferiority* as defined by the ratio. Suppose that 60% of patients respond to the current treatment method ($p_2 = 0.60$). If a new treatment decreases the response rate by no more than 10% ($\phi_0 = 0.90$), it will be considered to be non-inferior to the standard treatment. Substituting these figures into the statistical hypotheses gives

$$H_0\colon \phi \le 0.90 \quad \text{versus} \quad H_1\colon \phi > 0.90.$$

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

A Note on Setting the Significance Level, Alpha

Setting the significance level has always been somewhat arbitrary. For planning purposes, the standard has become to set alpha to 0.05 for two-sided tests. Almost universally, when someone states that a result is statistically significant, they mean statistically significant at the 0.05 level.

Although 0.05 may be the standard for two-sided tests, it is not always the standard for one-sided tests, such as non-inferiority tests. Statisticians often recommend that the alpha level for one-sided tests be set at 0.025 since this is the amount put in each tail of a two-sided test.

Power Calculation

The power for a test statistic that is based on the normal approximation can be computed exactly using two binomial distributions. The following steps are taken to compute the power of these tests.

1. Find the critical value using the standard normal distribution. The critical value, $z_{critical}$, is that value of *z* that leaves exactly the target value of alpha in the appropriate tail of the normal distribution.
2. Compute the value of the test statistic, z_t, for every combination of x_{11} and x_{21}. Note that x_{11} ranges from 0 to n_1, and x_{21} ranges from 0 to n_2. A small value (around 0.0001) can be added to the zero-cell counts to avoid numerical problems that occur when the cell value is zero.
3. If $z_t > z_{critical}$, the combination is in the rejection region. Call all combinations of x_{11} and x_{21} that lead to a rejection the set *A*.

4. Compute the power for given values of $p_{1.1}$ and p_2 as

$$1-\beta=\sum_{A}\binom{n_1}{x_{11}}p_{1.1}^{x_{11}}q_{1.1}^{n_1-x_{11}}\binom{n_2}{x_{21}}p_2^{x_{21}}q_2^{n_2-x_{21}}.$$

5. Compute the actual value of alpha achieved by the design by substituting $p_{1.0}$ for $p_{1.1}$ to obtain

$$\alpha^*=\sum_{A}\binom{n_1}{x_{11}}p_{1.0}^{x_{11}}q_{1.0}^{n_1-x_{11}}\binom{n_2}{x_{21}}p_2^{x_{21}}q_2^{n_2-x_{21}}.$$

Asymptotic Approximations

When the values of n_1 and n_2 are large (say over 200), these formulas often take a long time to evaluate. In this case, a large sample approximation can be used. The large sample approximation is made by replacing the values of $\hat{p}_1$ and $\hat{p}_1$ in the *z* statistic with the corresponding values of $p_{1.1}$ and p_2, and then computing the results based on the normal distribution. Note that in large samples, the Farrington and Manning statistic is substituted for the Gart and Nam statistic.

Test Statistics

Three test statistics have been proposed for testing whether the ratio is different from a specified value. The main difference among the several test statistics is in the formula used to compute the standard error used in the denominator. These tests are based on the following *z*-test

$$z_t=\frac{\hat{p}_1/\hat{p}_2-\phi_0}{\hat{\sigma}}$$

In power calculations, the values of $\hat{p}_1$ and $\hat{p}_2$ are not known. The corresponding values of $p_{1.1}$ and p_2 may be reasonable substitutes.

Following is a list of the test statistics available in **PASS**. The availability of several test statistics begs the question of which test statistic one should use. The answer is simple: one should use the test statistic that will be used to analyze the data. You may choose a method because it is a standard in your industry, because it seems to have better statistical properties, or because your statistical package calculates it. Whatever your reasons for selecting a certain test statistic, you should use the same test statistic when doing the analysis after the data have been collected.

Miettinen and Nurminen's Likelihood Score Test

Miettinen and Nurminen (1985) proposed a test statistic for testing whether the ratio is equal to a specified value ϕ_0. The regular MLE's, $\hat{p}_1$ and $\hat{p}_2$, are used in the numerator of the score statistic while MLE's $\tilde{p}_1$ and $\tilde{p}_2$, constrained so that $\tilde{p}_1\,/\,\tilde{p}_2=\phi_0$, are used in the denominator. A correction factor of *N*/(*N*-1) is applied to make the variance estimate less biased. The significance level of the test statistic is based on the asymptotic normality of the score statistic.

The formula for computing the test statistic is

$$z_{MNR}=\frac{\hat{p}_1\,/\,\hat{p}_2-\phi_0}{\sqrt{\left(\frac{\tilde{p}_1\tilde{q}_1}{n_1}+\phi_0^2\frac{\tilde{p}_2\tilde{q}_2}{n_2}\right)\left(\frac{N}{N-1}\right)}}$$

where

$$\tilde{p}_1 = \tilde{p}_2\phi_0$$

$$\tilde{p}_2 = \frac{-B - \sqrt{B^2 - 4AC}}{2A}$$

$$A = N\phi_0$$

$$B = -[n_1\phi_0 + x_{11} + n_2 + x_{21}\phi_0]$$

$$C = m_1$$

Farrington and Manning's Likelihood Score Test

Farrington and Manning (1990) proposed a test statistic for testing whether the ratio is equal to a specified value ϕ_0. The regular MLE's, $\hat{p}_1$ and $\hat{p}_2$, are used in the numerator of the score statistic while MLE's $\tilde{p}_1$ and $\tilde{p}_2$, constrained so that $\tilde{p}_1/\tilde{p}_2 = \phi_0$, are used in the denominator. The significance level of the test statistic is based on the asymptotic normality of the score statistic.

The formula for computing the test statistic is

$$z_{FMR} = \frac{\hat{p}_1 \,/\, \hat{p}_2 - \phi_0}{\sqrt{\left(\frac{\tilde{p}_1\tilde{q}_1}{n_1} + \phi_0^2\frac{\tilde{p}_2\tilde{q}_2}{n_2}\right)}}$$

where the estimates $\tilde{p}_1$ and $\tilde{p}_2$ are computed as in the corresponding test of Miettinen and Nurminen (1985) given above.

Gart and Nam's Likelihood Score Test

Gart and Nam (1988), page 329, proposed a modification to the Farrington and Manning (1988) ratio test that corrects for skewness. Let $z_{FMR}(\phi)$ stand for the Farrington and Manning ratio test statistic described above. The skewness corrected test statistic, z_{GNR}, is the appropriate solution to the quadratic equation

$$(-\tilde{\varphi})z_{GNR}^2 + (-1)z_{GNR} + (z_{FMR}(\phi) + \tilde{\varphi}) = 0$$

where

$$\tilde{\varphi} = \frac{1}{6\tilde{u}^{3/2}}\left(\frac{\tilde{q}_1(\tilde{q}_1 - \tilde{p}_1)}{n_1^2\tilde{p}_1^2} - \frac{\tilde{q}_2(\tilde{q}_2 - \tilde{p}_2)}{n_2^2\tilde{p}_2^2}\right)$$

$$\tilde{u} = \frac{\tilde{q}_1}{n_1\tilde{p}_1} + \frac{\tilde{q}_2}{n_2\tilde{p}_2}$$

Example 1 – Finding Power

A study is being designed to establish the non-inferiority of a new treatment compared to the current treatment. Historically, under the current treatment only 6% experience side effects. The new treatment is much less expensive and has been proven to be as effective. Thus, the new treatment will be adopted even if the rate of side effects is slightly worse than the current treatment. The researchers will recommend adoption of the new treatment if the rate ratio for side effects is less than 2.0.

The researchers plan to use the Farrington and Manning likelihood score test statistic to analyze the data that will be (or has been) obtained. They want to study the power of the Farrington and Manning test at group sample sizes ranging from 200 to 1000 when the actual ratio ranges from 1 to 1.5. The significance level will be 0.025.

Setup

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

Design Tab

Solve For	**Power**
Power Calculation Method	**Normal Approximation**
Higher Proportions Are	**Worse (H1: P1/P2 < R0)**
Test Type	**Likelihood Score (Farr. & Mann.)**
Alpha	**0.025**
Group Allocation	**Equal (N1 = N2)**
Sample Size Per Group	**200 to 1000 by 200**
R0 (Non-Inferiority Ratio)	**2.0**
R1 (Actual Ratio)	**1 1.25 1.5**
P2 (Group 2 Proportion)	**0.06**

Output

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

Numeric Reports

Numeric Results

Solve For: Power
Groups: 1 = Treatment, 2 = Reference
Test Statistic: Farrington & Manning Likelihood Score Test
Hypotheses: H0: P1 / P2 ≥ R0 vs. H1: P1 / P2 < R0

	Sample Size			Proportions			Ratio		
Power*	N1	N2	N	Non-Inferiority P1.0	Actual P1.1	Reference P2	Non-Inferiority R0	Actual R1	Alpha
0.43819	200	200	400	0.12	0.060	0.06	2	1.00	0.025
0.69368	400	400	800	0.12	0.060	0.06	2	1.00	0.025
0.84475	600	600	1200	0.12	0.060	0.06	2	1.00	0.025
0.92539	800	800	1600	0.12	0.060	0.06	2	1.00	0.025
0.96558	1000	1000	2000	0.12	0.060	0.06	2	1.00	0.025
0.26051	200	200	400	0.12	0.075	0.06	2	1.25	0.025
0.43785	400	400	800	0.12	0.075	0.06	2	1.25	0.025
0.58551	600	600	1200	0.12	0.075	0.06	2	1.25	0.025
0.70194	800	800	1600	0.12	0.075	0.06	2	1.25	0.025
0.79005	1000	1000	2000	0.12	0.075	0.06	2	1.25	0.025
0.13521	200	200	400	0.12	0.090	0.06	2	1.50	0.025
0.21618	400	400	800	0.12	0.090	0.06	2	1.50	0.025
0.29391	600	600	1200	0.12	0.090	0.06	2	1.50	0.025
0.36806	800	800	1600	0.12	0.090	0.06	2	1.50	0.025
0.43787	1000	1000	2000	0.12	0.090	0.06	2	1.50	0.025

* Power was computed using the normal approximation method.

Power	The probability of rejecting a false null hypothesis when the alternative hypothesis is true.
N1 and N2	The number of items sampled from each population.
N	The total sample size. N = N1 + N2.
P1	The proportion for group 1, which is the treatment or experimental group.
P1.0	The largest group 1 proportion that still yields a non-inferiority conclusion. P1.0 = P1|H0.
P1.1	The proportion for group 1 under the alternative hypothesis at which power and sample size calculations are made. P1.1 = P1|H1.
P2	The proportion for group 2, which is the standard, reference, or control group.
R0	The non-inferiority ratio, P1 / P2, assuming H0.
R1	The non-inferiority ratio, P1 / P2, assuming H1.
Alpha	The probability of rejecting a true null hypothesis.

Summary Statements

A parallel, two-group design will be used to test whether the Group 1 (treatment) proportion (P1) is non-inferior to the Group 2 (reference) proportion (P2), with a non-inferiority ratio of 2 (H0: P1 / P2 ≥ 2 versus H1: P1 / P2 < 2). The comparison will be made using a one-sided, two-sample Score test (Farrington & Manning) with a Type I error rate (α) of 0.025. The reference group proportion is assumed to be 0.06. To detect a proportion ratio (P1 / P2) of 1 (or P1 of 0.06) with sample sizes of 200 for Group 1 (treatment) and 200 for Group 2 (reference), the power is 0.43819.

Dropout-Inflated Sample Size

	Sample Size			Dropout-Inflated Enrollment Sample Size			Expected Number of Dropouts		
Dropout Rate	**N1**	**N2**	**N**	**N1'**	**N2'**	**N'**	**D1**	**D2**	**D**
20%	200	200	400	250	250	500	50	50	100
20%	400	400	800	500	500	1000	100	100	200
20%	600	600	1200	750	750	1500	150	150	300
20%	800	800	1600	1000	1000	2000	200	200	400
20%	1000	1000	2000	1250	1250	2500	250	250	500

Dropout Rate The percentage of subjects (or items) that are expected to be lost at random during the course of the study and for whom no response data will be collected (i.e., will be treated as "missing"). Abbreviated as DR.
N1, N2, and N The evaluable sample sizes at which power is computed (as entered by the user). If N1 and N2 subjects are evaluated out of the N1' and N2' subjects that are enrolled in the study, the design will achieve the stated power.
N1', N2', and N' The number of subjects that should be enrolled in the study in order to obtain N1, N2, and N evaluable subjects, based on the assumed dropout rate. N1' and N2' are calculated by inflating N1 and N2 using the formulas N1' = N1 / (1 - DR) and N2' = N2 / (1 - DR), with N1' and N2' always rounded up. (See Julious, S.A. (2010) pages 52-53, or Chow, S.C., Shao, J., Wang, H., and Lokhnygina, Y. (2018) pages 32-33.)
D1, D2, and D The expected number of dropouts. D1 = N1' - N1, D2 = N2' - N2, and D = D1 + D2.

Dropout Summary Statements

Anticipating a 20% dropout rate, 250 subjects should be enrolled in Group 1, and 250 in Group 2, to obtain final group sample sizes of 200 and 200, respectively.

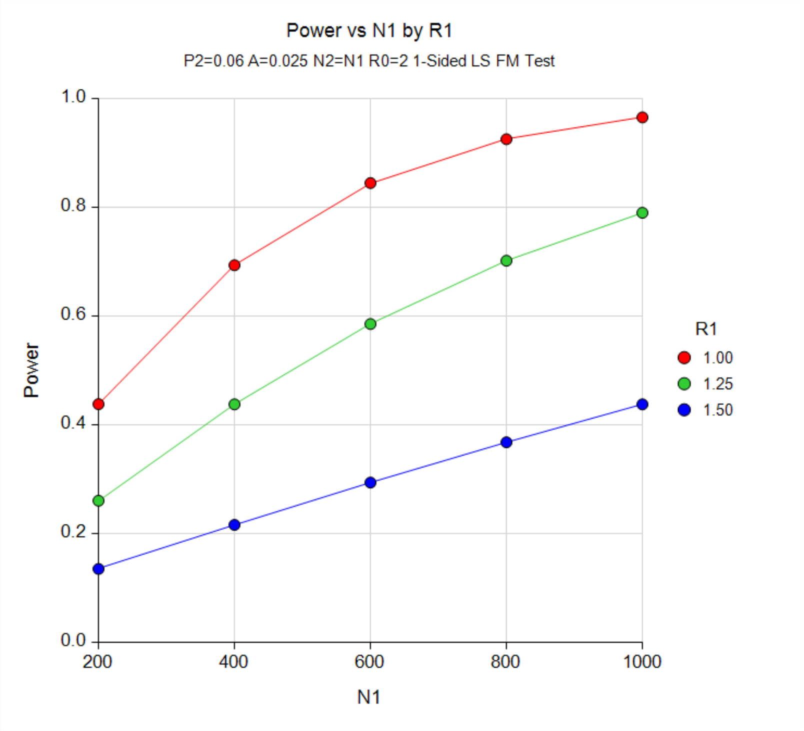
References

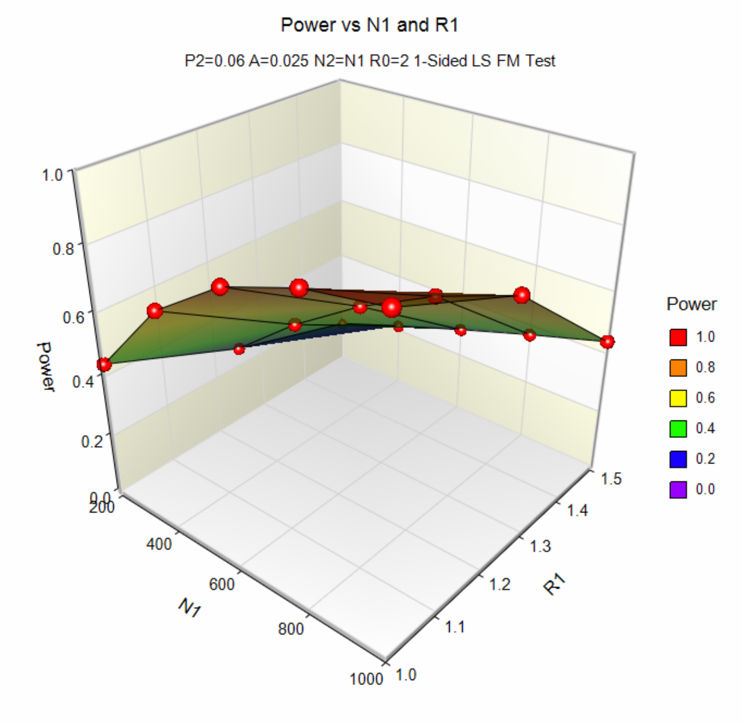
Chow, S.C., Shao, J., and Wang, H. 2008. Sample Size Calculations in Clinical Research, Second Edition. Chapman & Hall/CRC. Boca Raton, Florida.
Farrington, C. P. and Manning, G. 1990. 'Test Statistics and Sample Size Formulae for Comparative Binomial Trials with Null Hypothesis of Non-Zero Risk Difference or Non-Unity Relative Risk.' Statistics in Medicine, Vol. 9, pages 1447-1454.
Fleiss, J. L., Levin, B., Paik, M.C. 2003. Statistical Methods for Rates and Proportions. Third Edition. John Wiley & Sons. New York.
Gart, John J. and Nam, Jun-mo. 1988. 'Approximate Interval Estimation of the Ratio in Binomial Parameters: A Review and Corrections for Skewness.' Biometrics, Volume 44, Issue 2, 323-338.
Gart, John J. and Nam, Jun-mo. 1990. 'Approximate Interval Estimation of the Difference in Binomial Parameters: Correction for Skewness and Extension to Multiple Tables.' Biometrics, Volume 46, Issue 3, 637-643.
Julious, S. A. and Campbell, M. J. 2012. 'Tutorial in biostatistics: sample sizes for parallel group clinical trials with binary data.' Statistics in Medicine, 31:2904-2936.
Lachin, John M. 2000. Biostatistical Methods. John Wiley & Sons. New York.
Machin, D., Campbell, M., Fayers, P., and Pinol, A. 1997. Sample Size Tables for Clinical Studies, 2nd Edition. Blackwell Science. Malden, Mass.
Miettinen, O.S. and Nurminen, M. 1985. 'Comparative analysis of two rates.' Statistics in Medicine 4: 213-226.

This report shows the values of each of the parameters, one scenario per row.

Plots Section

Plots





The values from the table are displayed in the above chart. These charts give us a quick look at the sample size that will be required for various values of R1.

Example 2 – Finding the Sample Size

Continuing with the scenario given in Example 1, the researchers want to determine the sample size necessary for each value of R1 to achieve a power of 0.80.

Setup

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

Design Tab

Solve For ....**Sample Size**
Power Calculation Method ....**Normal Approximation**
Higher Proportions Are ....**Worse (H1: P1/P2 < R0)**
Test Type...**Likelihood Score (Farr. & Mann.)**
Power...**0.80**
Alpha..**0.025**
Group Allocation ....**Equal (N1 = N2)**
R0 (Non-Inferiority Ratio)..............................**2.0**
R1 (Actual Ratio) **1 1.25 1.5**
P2 (Group 2 Proportion)................................**0.06**

Output

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

Numeric Results

Solve For: Sample Size
Groups: 1 = Treatment, 2 = Reference
Test Statistic: Farrington & Manning Likelihood Score Test
Hypotheses: H0: P1 / P2 ≥ R0 vs. H1: P1 / P2 < R0

Power		Sample Size			Proportions			Ratio		
Target	Actual*	N1	N2	N	Non-Inferiority P1.0	Actual P1.1	Reference P2	Non-Inferiority R0	Actual R1	Alpha
0.8	0.80033	528	528	1056	0.12	0.060	0.06	2	1.00	0.025
0.8	0.80003	1027	1027	2054	0.12	0.075	0.06	2	1.25	0.025
0.8	0.80015	2508	2508	5016	0.12	0.090	0.06	2	1.50	0.025

* Power was computed using the normal approximation method.

The required sample size will depend a great deal on the value of R1. Any effort spent determining an accurate value for R1 will be worthwhile.

Example 3 – Comparing the Power of the Three Test Statistics

Continuing with Example 2, the researchers want to determine which of the three possible test statistics to adopt by using the comparative reports and charts that **PASS** produces. They decide to compare the powers from binomial enumeration and actual alphas for various sample sizes between 1000 and 1200 when R1 is 1.25.

Setup

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

Design Tab

Solve For	**Power**
Power Calculation Method	**Binomial Enumeration**
Maximum N1 or N2 for Binomial Enumeration	**5000**
Zero Count Adjustment Method	**Add to zero cells only**
Zero Count Adjustment Value	**0.0001**
Higher Proportions Are	**Worse (H1: P1/P2 < R0)**
Test Type	**Likelihood Score (Farr. & Mann.)**
Alpha	**0.025**
Group Allocation	**Equal (N1 = N2)**
Sample Size Per Group	**1000 1100 1200**
R0 (Non-Inferiority Ratio)	**2.0**
R1 (Actual Ratio)	**1.25**
P2 (Group 2 Proportion)	**0.06**

Reports Tab

Show Comparative Reports	**Checked**

Comparative Plots Tab

Show Comparative Plots	**Checked**

Output

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

Power Comparison of Three Different Tests

Hypotheses: H0: P1 / P2 ≥ R0 vs. H1: P1 / P2 < R0

Sample Size							Power		
N1	N2	N	P2	R0	R1	Target Alpha	F.M. Score	M.N. Score	G.N. Score
1000	1000	2000	0.06	2	1.25	0.025	0.7923	0.7923	0.7847
1100	1100	2200	0.06	2	1.25	0.025	0.8276	0.8276	0.8220
1200	1200	2400	0.06	2	1.25	0.025	0.8578	0.8578	0.8527

Note: Power was computed using binomial enumeration of all possible outcomes.

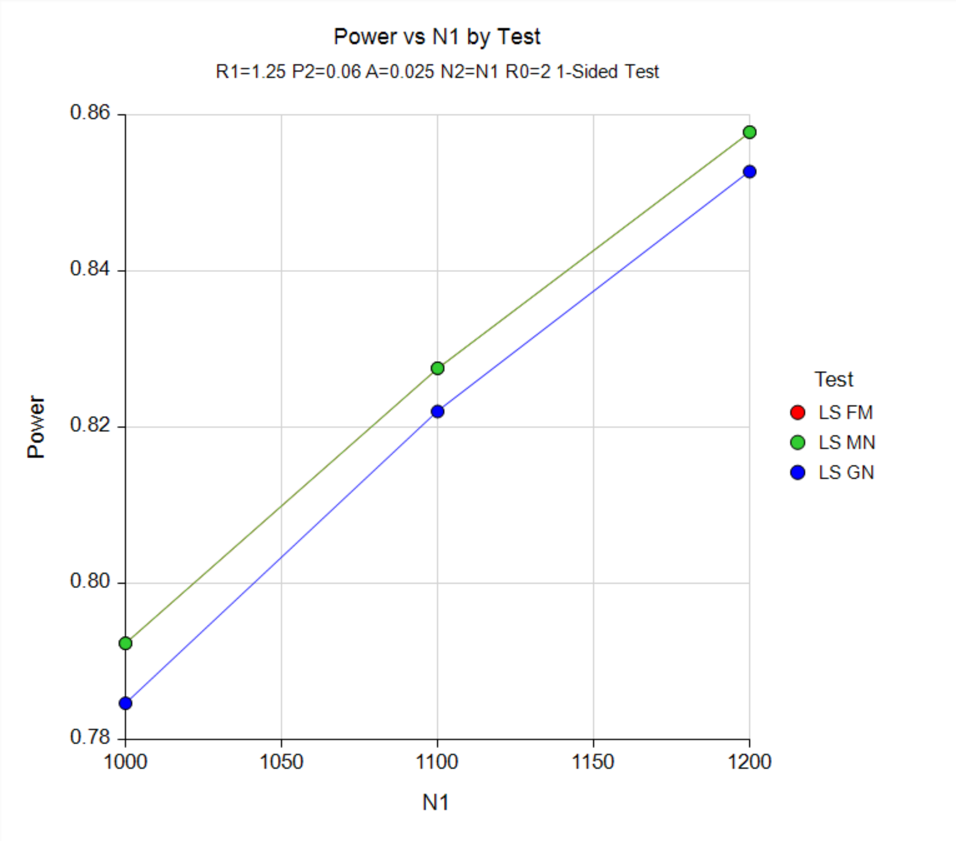
Actual Alpha Comparison of Three Different Tests

Hypotheses: H0: P1 / P2 ≥ R0 vs. H1: P1 / P2 < R0

Sample Size						Alpha			
N1	N2	N	P2	R0	R1	Target	F.M. Score	M.N. Score	G.N. Score
1000	1000	2000	0.06	2	1.25	0.025	0.0264	0.0264	0.0248
1100	1100	2200	0.06	2	1.25	0.025	0.0265	0.0262	0.0250
1200	1200	2400	0.06	2	1.25	0.025	0.0262	0.0262	0.0250

Note: Actual alpha was computed using binomial enumeration of all possible outcomes.

Plots



It is interesting to note that the powers of the Gart & Nam test statistics are consistently lower than the other tests. Notice, however, that the actual alpha levels for the Gart & Nam tests are consistently lower than the other two tests and achieve the target alpha level.

Example 4 – Comparing Power Calculation Methods

Continuing with Example 3, let's see how the results compare if we were to use approximate power calculations instead of power calculations based on binomial enumeration.

Setup

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

Design Tab

Solve For ....**Power**
Power Calculation Method............................**Normal Approximation**
Higher Proportions Are ....**Worse (H1: P1/P2 < R0)**
Test Type...**Likelihood Score (Farr. & Mann.)**
Alpha...**0.025**
Group Allocation ....**Equal (N1 = N2)**
Sample Size Per Group **1000 1100 1200**
R0 (Non-Inferiority Ratio)..............................**2.0**
R1 (Actual Ratio) **1.25**
P2 (Group 2 Proportion)................................**0.06**

Reports Tab

Show Power Detail Report............................**Checked**

Output

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

Power Detail Report

Test Statistic: Farrington & Manning Likelihood Score Test
Hypotheses: H0: P1 / P2 ≥ R0 vs. H1: P1 / P2 < R0

Sample Size						Normal Approximation		Binomial Enumeration	
N1	N2	N	P2	R0	R1	Power	Alpha	Power	Alpha
1000	1000	2000	0.06	2	1.25	0.79005	0.025	0.79234	0.0264
1100	1100	2200	0.06	2	1.25	0.82497	0.025	0.82757	0.0265
1200	1200	2400	0.06	2	1.25	0.85467	0.025	0.85780	0.0262

Notice that the approximate power values are close to the binomial enumeration values for all sample sizes.

Example 5 – Validation of Power Calculations using Blackwelder (1993)

Blackwelder (1993), page 695, presents a table of power values for several scenarios using the risk ratio. The second line of the table presents the results for the following scenario: P2 = 0.04, R0 = 0.3, R1=0.1, N1=N2=1044, one-sided alpha = 0.05, and beta = 0.20. Using the Farrington and Manning likelihood-score test statistic, he found the binomial enumeration power to be 0.812, the actual alpha to be 0.044, and, using the asymptotic formula, the approximate power to be 0.794.

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 5(a or b)** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

Solve For .... **Power**
Power Calculation Method .... **Binomial Enumeration**
Maximum N1 or N2 for Binomial Enumeration **5000**
Zero Count Adjustment Method .... **Add to zero cells only**
Zero Count Adjustment Value **0.0001**
Higher Proportions Are .... **Worse (H1: P1/P2 < R0)**
Test Type .... **Likelihood Score (Farr. & Mann.)**
Alpha **0.05**
Group Allocation .... **Equal (N1 = N2)**
Sample Size Per Group **1044**
R0 (Non-Inferiority Ratio) **0.3**
R1 (Actual Ratio) **0.1**
P2 (Group 2 Proportion) **0.04**

Output

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

Numeric Results

Solve For: Power
Groups: 1 = Treatment, 2 = Reference
Test Statistic: Farrington & Manning Likelihood Score Test
Hypotheses: H0: P1 / P2 ≥ R0 vs. H1: P1 / P2 < R0

	Sample Size			Proportions			Ratio		Alpha	
Power*	N1	N2	N	Non-Inferiority P1.0	Actual P1.1	Reference P2	Non-Inferiority R0	Actual R1	Target	Actual*
0.81178	1044	1044	2088	0.012	0.004	0.04	0.3	0.1	0.05	0.0444

* Power and actual alpha were computed using binomial enumeration of all possible outcomes.

PASS calculated the power to be 0.81178 and the actual alpha to be 0.0444, which round to Blackwelder's values.

Next, to calculate the asymptotic power, we make the following changes to the template:

Design Tab

Power Calculation Method .... **Normal Approximation**

Numeric Results

Numeric Results

Solve For: Power
Groups: 1 = Treatment, 2 = Reference
Test Statistic: Farrington & Manning Likelihood Score Test
Hypotheses: H0: P1 / P2 ≥ R0 vs. H1: P1 / P2 < R0

	Sample Size			Proportions			Ratio		
Power*	N1	N2	N	Non-Inferiority P1.0	Actual P1.1	Reference P2	Non-Inferiority R0	Actual R1	Alpha
0.79373	1044	1044	2088	0.012	0.004	0.04	0.3	0.1	0.05

* Power was computed using the normal approximation method.

PASS also calculated the power to be 0.794.