Chapter 207

Non-Unity Null Tests for the Odds Ratio of Two Proportions

Introduction

This module computes power and sample size for hypothesis tests of the odds ratio of two independent proportions where the null-hypothesized value is not equal to one. The *non-offset* case is available in another procedure. This procedure compares the power achieved by each of several test statistics.

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

Technical Details

Suppose you have two populations from which dichotomous (binary) responses will be recorded. The probability (or risk) of obtaining the event of interest in population 1 (the treatment group) is p_1 and in population 2 (the control group) is p_2. The corresponding failure proportions are given by $q_1 = 1 - p_1$ and $q_2 = 1 - p_2$.

An assumption is made that the responses from each group follow a binomial distribution. This means that the event probability, p_i, is the same for all subjects within the group and that the response from one subject is independent of that of any other subject.

Random samples of *m* and *n* individuals are obtained from these two populations. The data from these samples can be displayed in a 2-by-2 contingency table as follows

Group	Success	Failure	Total
Treatment	*a*	*c*	*m*
Control	*b*	*d*	*n*
Total	*s*	*f*	*N*

The following alternative notation is sometimes used.

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

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

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

Comparing Two Proportions

When analyzing studies such as this, you usually want to compare the two binomial probabilities, p_1 and p_2. The most direct method of comparing these quantities is to calculate their difference or their ratio. If the binomial probability is expressed in terms of odds rather than probability, another measure is the odds ratio. Mathematically, these comparison parameters are

Parameter	**Computation**
Difference	$\delta = p_1 - p_2$
Risk Ratio	$\phi = p_1 / p_2$
Odds Ratio	$\psi = \frac{p_1/(1-p_1)}{p_2/(1-p_2)} = \frac{p_1 q_2}{p_2 q_1}$

The choice of which of these measures is used might seem arbitrary, but it is not. Not only will the interpretation be different, but, for small sample sizes, the powers of tests based on different parameters will be different. The non-null case is commonly used in equivalence and non-inferiority testing.

Odds Ratio

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

The odds of an event are calculated by dividing the event risk by the non-event risk. Thus, in our case of two populations, the odds are

$$o_1 = \frac{p_1}{1-p_1} \quad \text{and} \quad o_1 = \frac{p_2}{1-p_2}$$

For example, if p_1 is 0.60, the odds are 0.60/0.40 = 1.5. Rather than represent the odds as a decimal amount, it is re-scaled into whole numbers. Instead of saying the odds are 1.5-to-1, we say they are 3-to-2.

Another way to compare proportions is to compute the ratio of their odds. The odds ratio of two events is

$$\psi = \frac{o_1}{o_2} = \frac{p_1 \,/\, (1-p_1)}{p_2 \,/\, (1-p_2)} = \frac{p_1 q_2}{p_2 q_1}$$

Although the odds ratio is more complicated to interpret than the risk ratio, it is often the parameter of choice. One reason for this is the fact that the odds ratio can be accurately estimated from case-control studies, while the risk ratio cannot. Also, the odds ratio is the basis of logistic regression (used to study the influence of risk factors). Furthermore, the odds ratio is the natural parameter in the conditional likelihood of the two-group, binomial-response design. Finally, when the baseline event rates are rare, the odds ratio provides a close approximation to the risk ratio since, in this case, $1 - p_1 \approx 1 - p_2$, so that

$$\psi = \frac{o_1}{o_2} = \frac{p_1/(1-p_1)}{p_2/(1-p_2)} \approx \frac{p_1}{p_2} = \phi$$

Hypothesis Tests

Although several statistical tests are available for testing the inequality of two proportions, only a few can be generalized to the non-null case. No single test is the champion in every situation, so one should compare the powers of the various tests to determine which to use.

Odds Ratio

The odds ratio, $\psi = [p_1 \,/\, (1 - p_1)] \,/\, [p_2 \,/\, (1 - p_2)]$, is sometimes used as the comparison because of its statistical properties and because some convenient experimental designs only allow it to be estimated. Three sets of statistical hypotheses can be formulated:

1. $H_0\colon \psi = \psi_0$ versus $H_0\colon \psi \neq \psi_0$; this is often called the *two-tailed test*.
2. $H_0\colon \psi \leq \psi_0$ versus $H_0\colon \psi > \psi_0$; this is often called the *upper-tailed test*.
3. $H_0\colon \psi \geq \psi_0$ versus $H_0\colon \psi < \psi_0$; this is often called the *lower-tailed test*.

Power Calculation

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

1. Find the critical value (or values in the case of a two-sided test) using the standard normal distribution. The critical value, $z_{critical}$, is that value of *z* that leaves exactly the target value of alpha in the appropriate tail of the normal distribution. For example, for an upper-tailed test with a target alpha of 0.05, the critical value is 1.645.
2. Compute the value of the test statistic, z_t, for every combination of x_{11} andx_{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 $p_{1.1}$ (p_1 under the alternative) 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 $p_{1.0}$ (p_1 under the null) 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 is used. The large sample approximation is made by replacing the values of $\hat{p}_1$ and $\hat{p}_2$ in the *z* values with the corresponding values of p_1 and p_2 under the alternative hypothesis and then computing the results based on the normal distribution. Note that in large samples, the results for the odds ratio have not (to our knowledge) been published. In this case, we substitute the calculations based on the ratio.

Test Statistics

Two test statistics have been proposed for testing whether the odds ratio is different from a specified value. The main difference between the test statistics is in the formula used to compute the standard error used in the denominator. These tests are both

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 you should use. The answer is simple: you should use the test statistic that you will use to analyze your data. You may choose a method because it is a standard in your industry, because it seems to have better statistical properties, or because your statistical package calculates it. Whatever your reasons for selecting a certain test statistic, you should use the same test statistic during power or sample calculations.

Miettinen and Nurminen's Likelihood Score Test

Miettinen and Nurminen (1985) proposed a test statistic for testing whether the odds ratio is equal to a specified value, ψ_0. Because the approach they used with the difference and ratio does not easily extend to the odds ratio, they used a score statistic approach for the odds ratio. The regular MLE's are $\hat{p}_1$ and $\hat{p}_2$. The constrained MLE's are $\tilde{p}_1$ and $\tilde{p}_2$. These estimates are constrained so that $\tilde{\psi} = \psi_0$. A correction factor of N/(N-1) is applied to make the variance estimate less biased. The significance level of the test statistic is based on the asymptotic normality of the score statistic.

The formula for computing the test statistic is

$$z_{MNO} = \frac{\frac{(\hat{p}_1 - \tilde{p}_1)}{\tilde{p}_1\tilde{q}_1} - \frac{(\hat{p}_2 - \tilde{p}_2)}{\tilde{p}_2\tilde{q}_2}}{\sqrt{\left(\frac{1}{n_1\tilde{p}_1\tilde{q}_1} + \frac{1}{n_2\tilde{p}_2\tilde{q}_2}\right)\left(\frac{N}{N-1}\right)}}$$

where

$$\tilde{p}_1 = \frac{\tilde{p}_2\psi_0}{1 + \tilde{p}_2(\psi_0 - 1)}$$

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

$$A = n_2(\psi_0 - 1),$$

$$B = n_1\psi_0 + n_2 - m_1(\psi_0 - 1),$$

$$C = -m_1$$

Farrington and Manning's Likelihood Score Test of the Odds Ratio

Farrington and Manning (1990) indicate that the Miettinen and Nurminen statistic may be modified by removing the factor *N*/(*N*-1).

The formula for computing this test statistic is

$$z_{FMO} = \frac{\frac{(\hat{p}_1 - \tilde{p}_1)}{\tilde{p}_1\tilde{q}_1} - \frac{(\hat{p}_2 - \tilde{p}_2)}{\tilde{p}_2\tilde{q}_2}}{\sqrt{\left(\frac{1}{n_1\tilde{p}_1\tilde{q}_1} + \frac{1}{n_2\tilde{p}_2\tilde{q}_2}\right)}}$$

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

Example 1 – Finding Power

A study is being designed to determine the effectiveness of a new treatment. The standard treatment has a success rate of 65%. They would like to show that the odds ratio of success rates for the new treatment to the old treatment is at least 1.4.

The researchers plan to use the Farrington and Manning likelihood score test statistic to analyze the data that will be (or has been) obtained. They want to study the power of the Farrington and Manning test at group sample sizes ranging from 50 to 200 for detecting an odds ratio of 1.4 when the actual odds ratio ranges from 2 to 3. 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**
Alternative Hypothesis	**One-Sided (H1: OR > OR0)**
Test Type	**Likelihood Score (Farr. & Mann.)**
Alpha	**0.025**
Group Allocation	**Equal (N1 = N2)**
Sample Size Per Group	**50 100 150 200**
OR0 (Odds Ratio|H0 = O1.0/O2)	**1.4**
OR1 (Odds Ratio|H1 = O1.1/O2)	**2 2.5 3**
P2 (Group 2 Proportion)	**0.65**

Output

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

Numeric Reports

Numeric Results

Solve For: Power
Test Statistic: Farrington & Manning Likelihood Score Test
Hypotheses: H0: OR ≤ OR0 vs. H1: OR > OR0

	Sample Size			Proportions			Odds Ratio		
Power*	**N1**	**N2**	**N**	**P1|H0 P1.0**	**P1|H1 P1.1**	**Reference P2**	**OR|H0 OR0**	**OR|H1 OR1**	**Alpha**
0.12420	50	50	100	0.7222	0.7879	0.65	1.4	2.0	0.025
0.20182	100	100	200	0.7222	0.7879	0.65	1.4	2.0	0.025
0.27751	150	150	300	0.7222	0.7879	0.65	1.4	2.0	0.025
0.35055	200	200	400	0.7222	0.7879	0.65	1.4	2.0	0.025
0.24109	50	50	100	0.7222	0.8228	0.65	1.4	2.5	0.025
0.41585	100	100	200	0.7222	0.8228	0.65	1.4	2.5	0.025
0.56501	150	150	300	0.7222	0.8228	0.65	1.4	2.5	0.025
0.68469	200	200	400	0.7222	0.8228	0.65	1.4	2.5	0.025
0.35467	50	50	100	0.7222	0.8478	0.65	1.4	3.0	0.025
0.59377	100	100	200	0.7222	0.8478	0.65	1.4	3.0	0.025
0.75970	150	150	300	0.7222	0.8478	0.65	1.4	3.0	0.025
0.86432	200	200	400	0.7222	0.8478	0.65	1.4	3.0	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 proportion for group 1 under the null hypothesis.
P1.1	The proportion for group 1 under the alternative hypothesis at which power and sample size calculations are made.
P2	The proportion for group 2, which is the standard, reference, or control group.
OR0	The odds ratio (O1/O2) under the null hypothesis.
OR1	The odds ratio (O1/O2) under the alternative hypothesis used for power and sample size calculations.
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 different from the Group 2 (reference) proportion (P2) by a margin, with a non-unity null odds ratio of 1.4 (H0: OR ≤ 1.4 versus H1: OR > 1.4). 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 0.65. To detect an odds ratio (O1 / O2) of the group proportions of 2 (or P1 of 0.7879) with sample sizes of 50 for the treatment group and 50 for the reference group, the power is 0.1242.

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%	50	50	100	63	63	126	13	13	26
20%	100	100	200	125	125	250	25	25	50
20%	150	150	300	188	188	376	38	38	76
20%	200	200	400	250	250	500	50	50	100

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, 63 subjects should be enrolled in Group 1, and 63 in Group 2, to obtain final group sample sizes of 50 and 50, respectively.

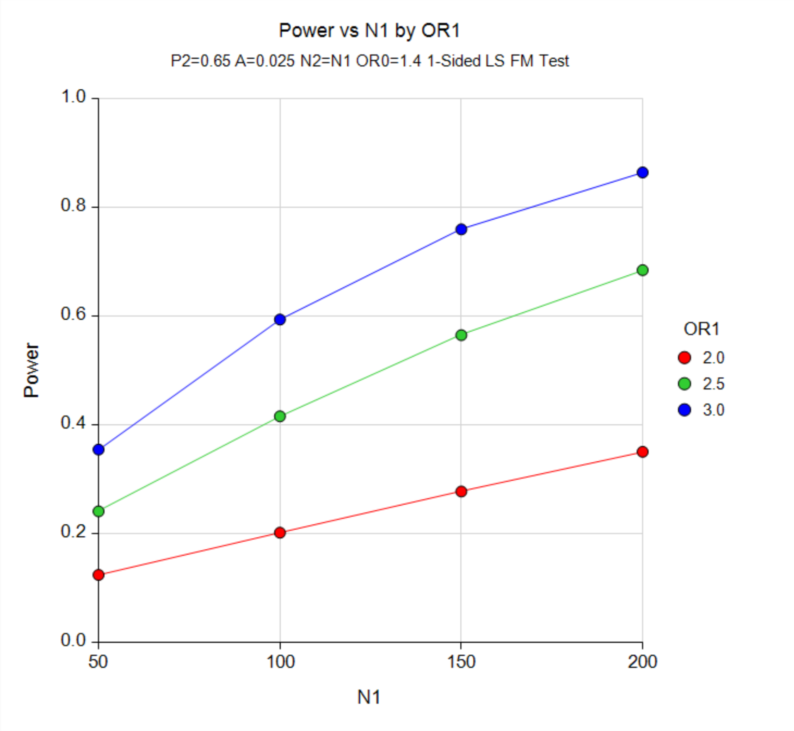
References

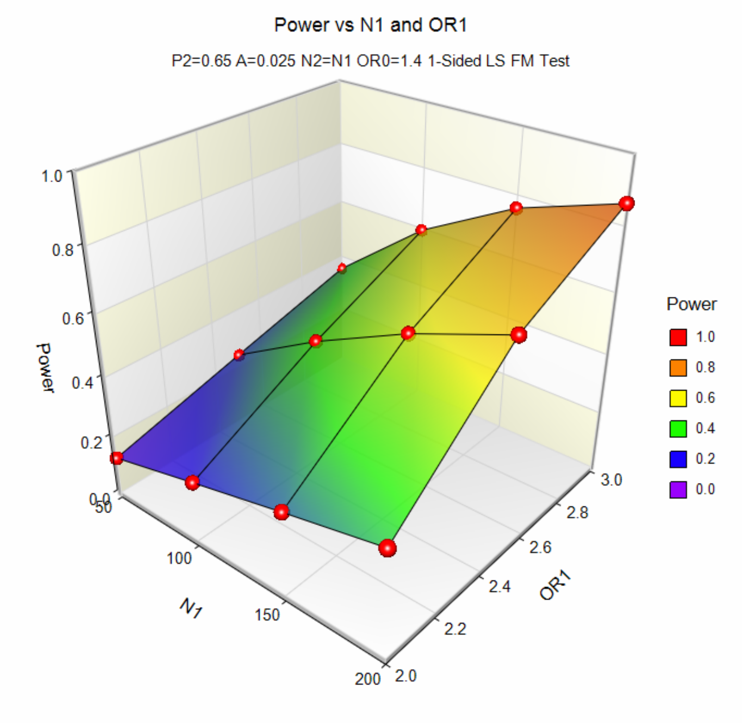
Chow, S.C., Shao, J., and Wang, H. 2008. Sample Size Calculations in Clinical Research, Second Edition. Chapman & Hall/CRC. Boca Raton, Florida.
Farrington, C. P. and Manning, G. 1990. 'Test Statistics and Sample Size Formulae for Comparative Binomial Trials with Null Hypothesis of Non-Zero Risk Difference 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 charts. These charts give us a quick look at the sample size that will be required for various values of OR1.

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 OR1 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**
Alternative Hypothesis .... **One-Sided (H1: OR > OR0)**
Test Type.. **Likelihood Score (Farr. & Mann.)**
Power... **0.80**
Alpha.. **0.025**
Group Allocation .... **Equal (N1 = N2)**
OR0 (Odds Ratio|H0 = O1.0/O2) **1.4**
OR1 (Odds Ratio|H1 = O1.1/O2) **2 2.5 3**
P2 (Group 2 Proportion)................................ **0.65**

Output

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

Numeric Results

Solve For: Sample Size
Test Statistic: Farrington & Manning Likelihood Score Test
Hypotheses: H0: OR ≤ OR0 vs. H1: OR > OR0

Power		Sample Size			Proportions			Odds Ratio		
Target	Actual*	N1	N2	N	P1|H0 P1.0	P1|H1 P1.1	Reference P2	OR|H0 OR0	OR|H1 OR1	Alpha
0.8	0.80022	645	645	1290	0.7222	0.7879	0.65	1.4	2.0	0.025
0.8	0.80057	266	266	532	0.7222	0.8228	0.65	1.4	2.5	0.025
0.8	0.80122	167	167	334	0.7222	0.8478	0.65	1.4	3.0	0.025

* Power was computed using the normal approximation method.

The required sample size will depend a great deal on the value of OR1. Any effort spent determining an accurate value for OR1 will be worthwhile.

Example 3 – Comparing the Power of the Two Test Statistics

Continuing with Example 2, the researchers want to determine which of the eight possible test statistics to adopt by using the comparative reports and charts that **PASS** produces. They decide to compare the powers from binomial enumeration and actual alphas for various sample sizes between 600 and 800 when OR1 is 2.0.

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**
Alternative Hypothesis ....**One-Sided (H1: OR > OR0)**
Test Type..**Likelihood Score (Farr. & Mann.)**
Alpha...**0.025**
Group Allocation ....**Equal (N1 = N2)**
Sample Size Per Group **600 700 800**
OR0 (Odds Ratio|H0 = O1.0/O2)...........................**1.4**
OR1 (Odds Ratio|H1 = O1.1/O2)...........................**2**
P2 (Group 2 Proportion)...**0.65**

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 Two Different Tests

Hypotheses: H0: OR ≤ OR0 vs. H1: OR > OR0

Sample Size							Power	
N1	N2	N	P2	OR0	OR1	Target Alpha	F.M. Score	M.N. Score
600	600	1200	0.65	1.4	2	0.025	0.7805	0.7805
700	700	1400	0.65	1.4	2	0.025	0.8404	0.8402
800	800	1600	0.65	1.4	2	0.025	0.8849	0.8849

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

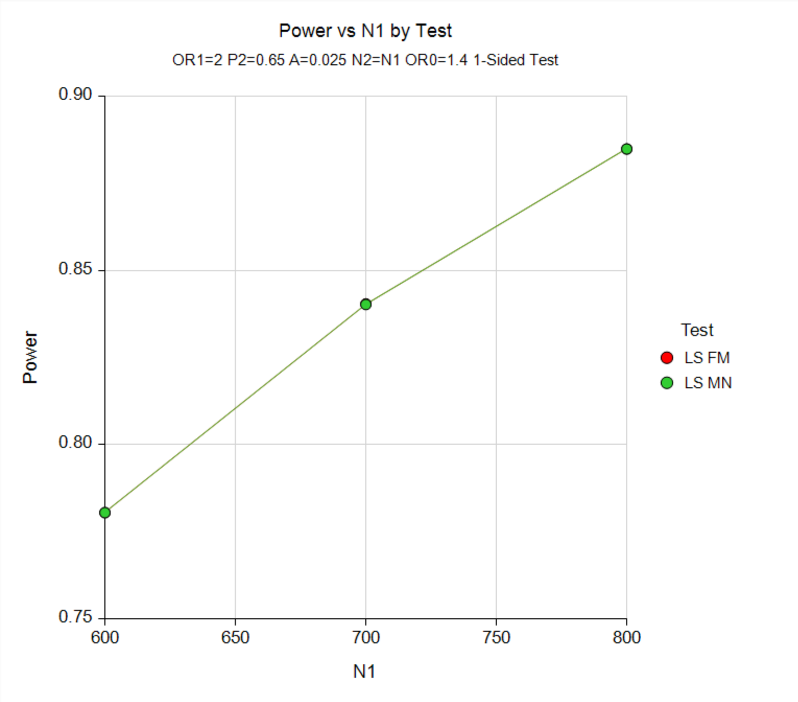
Actual Alpha Comparison of Two Different Tests

Hypotheses: H0: OR ≤ OR0 vs. H1: OR > OR0

Sample Size						Alpha		
N1	N2	N	P2	OR0	OR1	Target	F.M. Score	M.N. Score
600	600	1200	0.65	1.4	2	0.025	0.0250	0.0250
700	700	1400	0.65	1.4	2	0.025	0.0250	0.0249
800	800	1600	0.65	1.4	2	0.025	0.0249	0.0249

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

Plots



Both test statistics have almost identical power and actual alpha values for all sample sizes studied.

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**
Alternative Hypothesis .... **One-Sided (H1: OR > OR0)**
Test Type .... **Likelihood Score (Farr. & Mann.)**
Alpha.. **0.025**
Group Allocation .... **Equal (N1 = N2)**
Sample Size Per Group **600 700 800**
OR0 (Odds Ratio|H0 = O1.0/O2) **1.4**
OR1 (Odds Ratio|H1 = O1.1/O2) **2**
P2 (Group 2 Proportion)................................ **0.65**

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: OR ≤ OR0 vs. H1: OR > OR0

Sample Size						Normal Approximation		Binomial Enumeration	
N1	N2	N	P2	OR0	OR1	Power	Alpha	Power	Alpha
600	600	1200	0.65	1.4	2	0.77161	0.025	0.78049	0.0250
700	700	1400	0.65	1.4	2	0.83097	0.025	0.84041	0.0250
800	800	1600	0.65	1.4	2	0.87637	0.025	0.88489	0.0249

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

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

We could not find a validation example for a test for the odds ratio of two proportions with a non-unity null hypothesis. The calculations are basically the same as those for a test of the ratio of two proportions with a non-unity null hypothesis, which has been validated using Blackwelder (1993). We refer you to Example 5 of Chapter 206, “Non-Unity Null Tests for the Ratio of Two Proportions,” for a validation example.