Chapter 195

Superiority by a Margin Tests for the Difference Between Two Proportions

Introduction

This module provides power analysis and sample size calculation for superiority by a margin tests of the difference in two-sample designs in which the outcome is binary. Users may choose from among eight popular test statistics commonly used for running the hypothesis test.

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

Example

A superiority by a margin 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. A promising new treatment has been developed to the point where it can be tested. The researchers wish to show that the new treatment is better than the current treatment by at least some amount. In other words, does a clinically significant higher number of treated subjects respond to the new treatment?

Clinicians want to demonstrate the new treatment is superior to the current treatment. They must determine, however, how much more effective the new treatment must be to be adopted. Should it be adopted if 71% respond? 72%? 75%? 80%? There is a percentage above 70% at which the difference between the two treatments is no longer considered ignorable. After thoughtful discussion with several clinicians, it was decided that if a response of at least 77% were achieved, the new treatment would be adopted. The difference between these two percentages is called the *margin of superiority*. The margin of superiority in this example is 7%.

The developers must design an experiment to test the hypothesis that the response rate of the new treatment is at least 0.77. The statistical hypothesis to be tested is

$$H_0: p_1 - p_2 \leq 0.07 \quad \text{versus} \quad H_1: p_1 - p_2 > 0.07$$

Notice that when the null hypothesis is rejected, the conclusion is that the response rate is at least 0.77. Note that even though the response rate of the current treatment is 0.70, the hypothesis test is about a response rate of 0.77. 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 superiority by a margin tests.

Approximate sample size formulas for superiority by a margin tests of the difference between two proportions are presented in Chow et al. (2008), page 90. Only large sample (normal approximation) results are given there. It is also possible to calculate power based on the enumeration of all possible values in the binomial distribution. Both options are available in this procedure.

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

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

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

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

Let $p_{1.0}$ represent the group 1 proportion tested by the null hypothesis, H_0. The power of a test is computed at a specific value of the proportion which we will call $p_{1.1}$. Let δ_0 represent the smallest difference (margin of superiority) between the two proportions that results in the conclusion that the new treatment is superior to the current treatment. For a superiority by a margin test, $\delta_0 > 0$. The set of statistical hypotheses that are tested is

$$H_0: p_1 - p_2 \leq \delta_0 \quad \text{versus} \quad H_1: p_1 - p_2 > \delta_0$$

which can be rearranged to give

$$H_0: p_1 \leq p_2 + \delta_0 \quad \text{versus} \quad H_1: p_1 > p_2 + \delta_0$$

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

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

Difference

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

One difficulty arises when the event of interest is rare. If a difference of 0.001 occurs when the baseline probability is 0.40, it would be dismissed as being trivial. However, if the baseline probability of a disease is 0.002, a 0.001 decrease would represent a reduction of 50%. Thus, interpretation of the difference depends on the baseline probability of the event.

Superiority by a Margin

The following example is intended to help you understand the concept of a *superiority by a margin* test. Suppose 60% of patients respond to the current treatment method ($p_2 = 0.60$). If the response rate of the new treatment is at least 10 percentage points better ($\delta_0 = 0.10$), it will be considered to be superior to the existing treatment. Substituting these figures into the statistical hypotheses gives

$$H_0: p_1 - p_2 \leq 0.10 \quad \text{versus} \quad H_1: p_1 - p_2 > 0.10$$

In this example, when the null hypothesis is rejected, the concluded alternative is that the new treatment response rate is at least 0.10 more than that of the existing 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 superiority by a margin 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 $\hat{p}_2$ in the *z* statistic with the corresponding values of $p_{1.1}$ and p_2, and then computing the results based on the normal distribution. Note that in large samples, the Farrington and Manning statistic is substituted for the Gart and Nam statistic.

Test Statistics

Several test statistics have been proposed for testing whether the difference 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

$$z_t = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0 - c}{\hat{\sigma}}$$

The constant, *c,* represents a continuity correction that is applied in some cases. When the continuity correction is not used, *c* is zero. In power calculations, the values of $\hat{p}_1$ and $\hat{p}_2$ are not known. The corresponding values of $p_{1.1}$ and p_2 may be reasonable substitutes.

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

Z Test (Pooled)

This test was first proposed by Karl Pearson in 1900. Although this test is usually expressed directly as a chi-square statistic, it is expressed here as a *z* statistic so that it can be more easily used for one-sided hypothesis testing. The proportions are pooled (averaged) in computing the standard error. The formula for the test statistic is

$$z_t = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0}{\hat{\sigma}_1}$$

where

$$\hat{\sigma}_1 = \sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

$$\bar{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$$

Z Test (Unpooled)

This test statistic does not pool the two proportions in computing the standard error.

$$z_t = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0}{\hat{\sigma}_2}$$

where

$$\hat{\sigma}_2 = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

Z Test with Continuity Correction (Pooled)

This test is the same as Z Test (Pooled), except that a continuity correction is used. Remember that in the null case, the continuity correction makes the results closer to those of Fisher's Exact test.

$$z_t = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0 + \frac{F}{2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}{\hat{\sigma}_1}$$

where

$$\hat{\sigma}_1 = \sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

$$\bar{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$$

where F is -1 for lower-tailed hypotheses and 1 for upper-tailed hypotheses.

Z Test with Continuity Correction (Unpooled)

This test is the same as the Z Test (Unpooled), except that a continuity correction is used. Remember that in the null case, the continuity correction makes the results closer to those of Fisher's Exact test.

$$z_t = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0 - \frac{F}{2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}{\hat{\sigma}_2}$$

where

$$\hat{\sigma}_2 = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

where *F* is -1 for lower-tailed hypotheses and 1 for upper-tailed hypotheses.

T-Test

Because of a detailed, comparative study of the behavior of several tests, D'Agostino (1988) and Upton (1982) proposed using the usual two-sample t-test for testing whether the two proportions are equal. One substitutes a '1' for a success and a '0' for a failure in the usual, two-sample *t*-test formula.

Miettinen and Nurminen's Likelihood Score Test

Miettinen and Nurminen (1985) proposed a test statistic for testing whether the difference is equal to a specified, non-zero, value, δ_0. The regular MLE's, $\hat{p}_1$ and $\hat{p}_2$, are used in the numerator of the score statistic while MLE's $\tilde{p}_1$ and $\tilde{p}_2$, constrained so that $\tilde{p}_1 - \tilde{p}_2 = \delta_0$, are used in the denominator. A correction factor of *N*/(*N*-1) is applied to make the variance estimate less biased. The significance level of the test statistic is based on the asymptotic normality of the score statistic. The formula for computing this test statistic is

$$z_{MND} = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0}{\hat{\sigma}_{MND}}$$

where

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

$$\tilde{p}_1 = \tilde{p}_2 + \delta_0$$

$$\tilde{p}_1 = 2B\cos(A) - \frac{L_2}{3L_3}$$

$$A = \frac{1}{3}\left[\pi + \cos^{-1}\left(\frac{C}{B^3}\right)\right]$$

$$B = \text{sign}(C)\sqrt{\frac{L_2^2}{9L_3^2} - \frac{L_1}{3L_3}}$$

$$C = \frac{L_2^3}{27L_3^3} - \frac{L_1 L_2}{6L_3^2} + \frac{L_0}{2L_3}$$

$$L_0 = x_{21}\delta_0(1 - \delta_0)$$

$$L_1 = [n_2\delta_0 - N - 2x_{21}]\delta_0 + m_1$$

$$L_2 = (N + n_2)\delta_0 - N - m_1$$

$$L_3 = N$$

Farrington and Manning's Likelihood Score Test

Farrington and Manning (1990) proposed a test statistic for testing whether the difference is equal to a specified value, δ_0. The regular MLE's, $\hat{p}_1$ and $\hat{p}_2$, are used in the numerator of the score statistic while MLE's $\tilde{p}_1$ and $\tilde{p}_2$, constrained so that $\tilde{p}_1 - \tilde{p}_2 = \delta_0$, are used in the denominator. The significance level of the test statistic is based on the asymptotic normality of the score statistic.

The formula for computing the test statistic is

$$z_{FMD} = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0}{\sqrt{\left(\frac{\tilde{p}_1\tilde{q}_1}{n_1} + \frac{\tilde{p}_2\tilde{q}_2}{n_2}\right)}}$$

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

Gart and Nam's Likelihood Score Test

Gart and Nam (1990), page 638, proposed a modification to the Farrington and Manning (1988) difference test that corrects for skewness. Let $z_{FMD}(\delta)$ stand for the Farrington and Manning difference test statistic described above. The skewness-corrected test statistic, z_{GND}, is the appropriate solution to the quadratic equation

$$(-\tilde{\gamma})z_{GND}^2 + (-1)z_{GND} + (z_{FMD}(\delta) + \tilde{\gamma}) = 0$$

where

$$\tilde{\gamma} = \frac{\tilde{V}^{3/2}(\delta)}{6}\left(\frac{\tilde{p}_1\tilde{q}_1(\tilde{q}_1 - \tilde{p}_1)}{n_1^2} - \frac{\tilde{p}_2\tilde{q}_2(\tilde{q}_2 - \tilde{p}_2)}{n_2^2}\right)$$

Example 1 – Finding Power

A study is being designed to establish the superiority of a new treatment compared to the current treatment. Historically, the current treatment has enjoyed a 60% cure rate. The new treatment is hoped to perform better than the current treatment. Thus, the new treatment will be adopted if it is more effective than the current treatment by a clinically significant amount. The researchers will recommend adoption of the new treatment if it has a cure rate of at least 70%.

The researchers plan to use the Farrington and Manning likelihood score test statistic to analyze the data that will be (or has been) obtained. They want to study the power of the Farrington and Manning test at group sample sizes ranging from 50 to 500 when the actual cure rate of the new treatment ranges from 71% to 80%. 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	**Better (H1: P1 - P2 > δ0)**
Test Type	**Likelihood Score (Farr. & Mann.)**
Alpha	**0.025**
Group Allocation	**Equal (N1 = N2)**
Sample Size Per Group	**50 to 500 by 50**
Input Type	**Differences**
δ0 (Superiority Difference)	**0.1**
δ1 (Actual Difference)	**0.11 0.14 0.17 0.20**
P2 (Group 2 Proportion)	**0.6**

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 ≤ δ0 vs. H1: P1 - P2 > δ0

	Sample Size			Proportions			Difference		
Power*	N1	N2	N	Superiority P1.0	Actual P1.1	Reference P2	Superiority δ0	Actual δ1	Alpha
0.03173	50	50	100	0.7	0.71	0.6	0.1	0.11	0.025
0.03499	100	100	200	0.7	0.71	0.6	0.1	0.11	0.025
0.03767	150	150	300	0.7	0.71	0.6	0.1	0.11	0.025
0.04006	200	200	400	0.7	0.71	0.6	0.1	0.11	0.025
0.04226	250	250	500	0.7	0.71	0.6	0.1	0.11	0.025
0.04434	300	300	600	0.7	0.71	0.6	0.1	0.11	0.025
0.04632	350	350	700	0.7	0.71	0.6	0.1	0.11	0.025
0.04823	400	400	800	0.7	0.71	0.6	0.1	0.11	0.025
0.05008	450	450	900	0.7	0.71	0.6	0.1	0.11	0.025
0.05188	500	500	1000	0.7	0.71	0.6	0.1	0.11	0.025
0.06199	50	50	100	0.7	0.74	0.6	0.1	0.14	0.025
0.08690	100	100	200	0.7	0.74	0.6	0.1	0.14	0.025
0.11059	150	150	300	0.7	0.74	0.6	0.1	0.14	0.025
0.13390	200	200	400	0.7	0.74	0.6	0.1	0.14	0.025
0.15707	250	250	500	0.7	0.74	0.6	0.1	0.14	0.025
0.18015	300	300	600	0.7	0.74	0.6	0.1	0.14	0.025
0.20317	350	350	700	0.7	0.74	0.6	0.1	0.14	0.025
0.22609	400	400	800	0.7	0.74	0.6	0.1	0.14	0.025
0.24889	450	450	900	0.7	0.74	0.6	0.1	0.14	0.025
0.27153	500	500	1000	0.7	0.74	0.6	0.1	0.14	0.025
0.11309	50	50	100	0.7	0.77	0.6	0.1	0.17	0.025
0.18599	100	100	200	0.7	0.77	0.6	0.1	0.17	0.025
0.25811	150	150	300	0.7	0.77	0.6	0.1	0.17	0.025
0.32858	200	200	400	0.7	0.77	0.6	0.1	0.17	0.025
0.39631	250	250	500	0.7	0.77	0.6	0.1	0.17	0.025
0.46044	300	300	600	0.7	0.77	0.6	0.1	0.17	0.025
0.52037	350	350	700	0.7	0.77	0.6	0.1	0.17	0.025
0.57577	400	400	800	0.7	0.77	0.6	0.1	0.17	0.025
0.62650	450	450	900	0.7	0.77	0.6	0.1	0.17	0.025
0.67255	500	500	1000	0.7	0.77	0.6	0.1	0.17	0.025
0.19253	50	50	100	0.7	0.80	0.6	0.1	0.20	0.025
0.34256	100	100	200	0.7	0.80	0.6	0.1	0.20	0.025
0.48000	150	150	300	0.7	0.80	0.6	0.1	0.20	0.025
0.59849	200	200	400	0.7	0.80	0.6	0.1	0.20	0.025
0.69615	250	250	500	0.7	0.80	0.6	0.1	0.20	0.025
0.77397	300	300	600	0.7	0.80	0.6	0.1	0.20	0.025
0.83433	350	350	700	0.7	0.80	0.6	0.1	0.20	0.025
0.88013	400	400	800	0.7	0.80	0.6	0.1	0.20	0.025
0.91426	450	450	900	0.7	0.80	0.6	0.1	0.20	0.025
0.93930	500	500	1000	0.7	0.80	0.6	0.1	0.20	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 smallest group 1 proportion that still yields a superiority conclusion. P1.0 = P1|H0.
P1.1	The proportion for group 1 used for the alternative hypothesis, H1. P1.1 = P1|H1.
P2	The proportion for group 2, which is the standard, reference, or control group.
δ0	The superiority difference (or margin) used for the null hypothesis, H0. δ0 = P1.0 - P2.
δ1	The proportion difference used for the alternative hypothesis, H1. δ1 = P1.1 - P2.
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 superior to the Group 2 (reference) proportion (P2) by a margin, with a superiority margin of 0.1 (H0: P1 - P2 ≤ 0.1 versus H1: P1 - P2 > 0.1). 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 0.6. To detect a proportion difference (P1 - P2) of 0.11 (or P1 of 0.71) with sample sizes of 50 for the treatment group and 50 for the reference group, the power is 0.03173.

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
20%	250	250	500	313	313	626	63	63	126
20%	300	300	600	375	375	750	75	75	150
20%	350	350	700	438	438	876	88	88	176
20%	400	400	800	500	500	1000	100	100	200
20%	450	450	900	563	563	1126	113	113	226
20%	500	500	1000	625	625	1250	125	125	250

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.

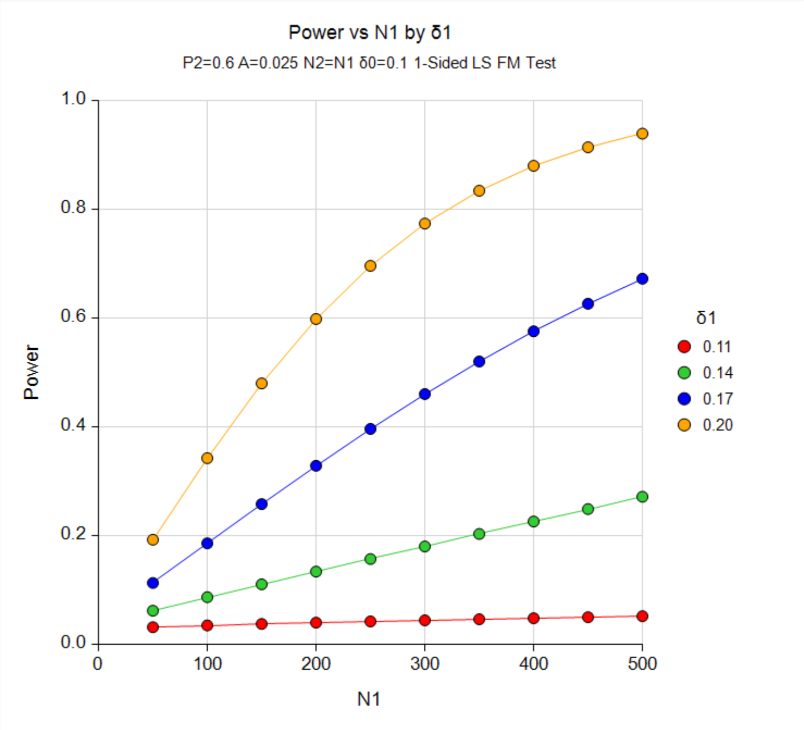
References

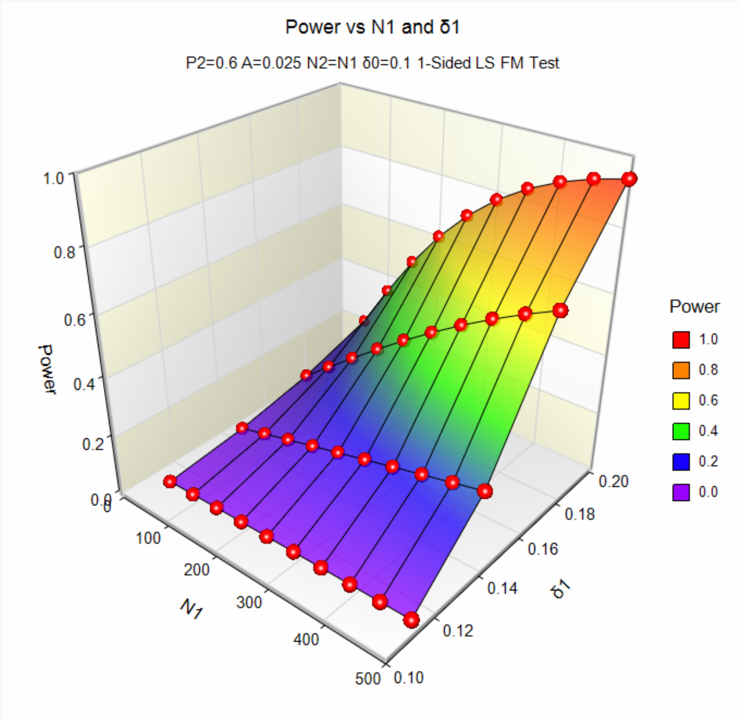
Chow, S.C., Shao, J., and Wang, H. 2008. Sample Size Calculations in Clinical Research, Second Edition. Chapman & Hall/CRC. Boca Raton, Florida.
Farrington, C. P. and Manning, G. 1990. 'Test Statistics and Sample Size Formulae for Comparative Binomial Trials with Null Hypothesis of Non-Zero Risk Difference 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 chart. These charts give us a quick look at the sample size that will be required for various values of δ1.

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 δ1 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	**Better (H1: P1 - P2 > δ0)**
Test Type	**Likelihood Score (Farr. & Mann.)**
Power	**0.80**
Alpha	**0.025**
Group Allocation	**Equal (N1 = N2)**
Input Type	**Differences**
δ0 (Superiority Difference)	**0.1**
δ1 (Actual Difference)	**0.11 0.14 0.17 0.20**
P2 (Group 2 Proportion)	**0.6**

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 ≤ δ0 vs. H1: P1 - P2 > δ0

Power		Sample Size			Proportions			Difference		
Target	Actual*	N1	N2	N	Superiority P1.0	Actual P1.1	Reference P2	Superiority δ0	Actual δ1	Alpha
0.8	0.80000	35044	35044	70088	0.7	0.71	0.6	0.1	0.11	0.025
0.8	0.80001	2134	2134	4268	0.7	0.74	0.6	0.1	0.14	0.025
0.8	0.80052	677	677	1354	0.7	0.77	0.6	0.1	0.17	0.025
0.8	0.80005	320	320	640	0.7	0.80	0.6	0.1	0.20	0.025

The required sample size will depend a great deal on the value of δ1. Any effort spent determining an accurate value for δ1 will be worthwhile.

Example 3 – Comparing the Power of Several 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 50 and 200 when δ1 is 0.15.

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 ....**Better (H1: P1 - P2 > δ0)**
Test Type..**Likelihood Score (Farr. & Mann.)**
Alpha..**0.025**
Group Allocation ....**Equal (N1 = N2)**
Sample Size Per Group **200 250 300 350**
Input Type..**Differences**
δ0 (Superiority Difference).....................................**0.1**
δ1 (Actual Difference) **0.2**
P2 (Group 2 Proportion)...**0.6**

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 Eight Different Tests

Hypotheses: H0: P1 - P2 ≤ δ0 vs. H1: P1 - P2 > δ0

Sample Size							Power							
N1	N2	N	P2	δ0	δ1	Target Alpha	Z(P) Test	Z(UnP) Test	Z(P) CC Test	Z(UnP) CC Test	T Test	F.M. Score	M.N. Score	G.N. Score
200	200	400	0.6	0.1	0.2	0.025	0.5930	0.6110	0.5470	0.5690	0.6052	0.6012	0.6012	0.6023
250	250	500	0.6	0.1	0.2	0.025	0.6909	0.7050	0.6532	0.6708	0.7023	0.6974	0.6974	0.7000
300	300	600	0.6	0.1	0.2	0.025	0.7685	0.7805	0.7409	0.7534	0.7786	0.7751	0.7751	0.7767
350	350	700	0.6	0.1	0.2	0.025	0.8315	0.8388	0.8085	0.8177	0.8386	0.8355	0.8355	0.8360

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

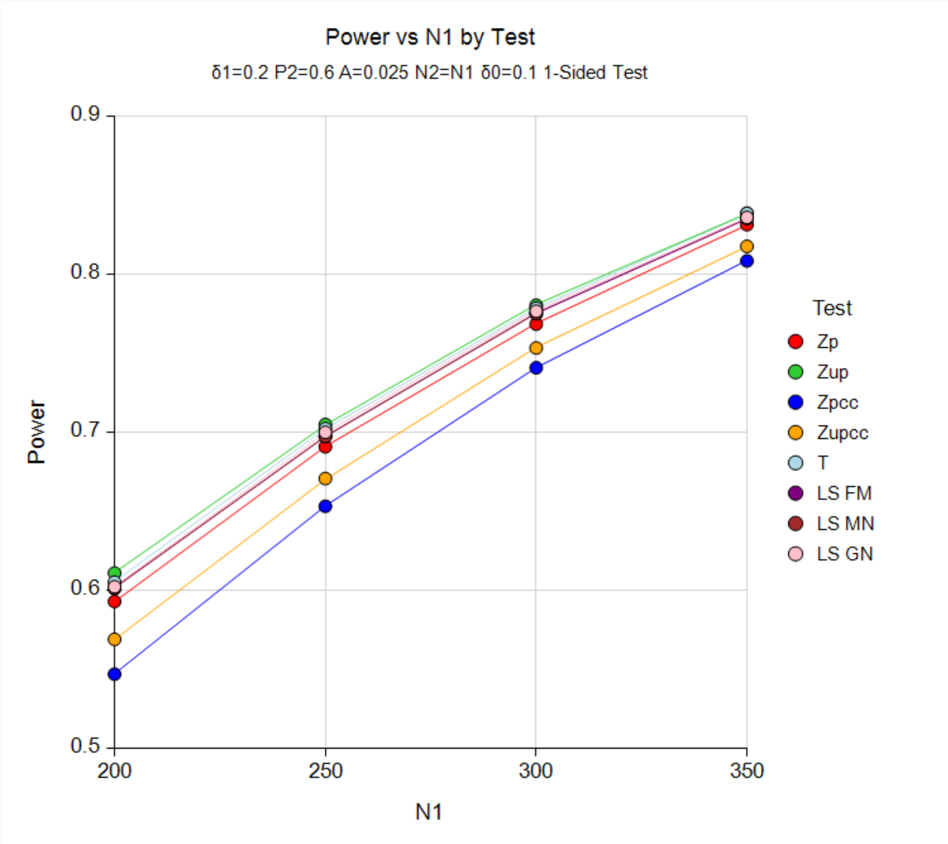
Actual Alpha Comparison of Eight Different Tests

Hypotheses: H0: P1 - P2 ≤ δ0 vs. H1: P1 - P2 > δ0

Sample Size						Alpha								
N1	N2	N	P2	δ0	δ1	Target	Z(P) Test	Z(UnP) Test	Z(P) CC Test	Z(UnP) CC Test	T Test	F.M. Score	M.N. Score	G.N. Score
200	200	400	0.6	0.1	0.2	0.025	0.0243	0.0262	0.0189	0.0205	0.0256	0.0252	0.0252	0.0253
250	250	500	0.6	0.1	0.2	0.025	0.0242	0.0264	0.0191	0.0211	0.0260	0.0253	0.0250	0.0253
300	300	600	0.6	0.1	0.2	0.025	0.0241	0.0262	0.0197	0.0214	0.0259	0.0251	0.0251	0.0253
350	350	700	0.6	0.1	0.2	0.025	0.0244	0.0258	0.0202	0.0213	0.0256	0.0251	0.0251	0.0252

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

Plots



It is interesting to note that the powers of the continuity-corrected test statistics are consistently lower than the other tests. This occurs because the actual alpha achieved by these tests is lower than for the other tests. An interesting finding of this example is that the regular *t*-test performed about as well as the *z*-test.

Example 4 – 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	**Better (H1: P1 - P2 > δ0)**
Test Type	**Likelihood Score (Farr. & Mann.)**
Alpha	**0.025**
Group Allocation	**Equal (N1 = N2)**
Sample Size Per Group	**200 250 300 350**
Input Type	**Differences**
δ0 (Superiority Difference)	**0.1**
δ1 (Actual Difference)	**0.2**
P2 (Group 2 Proportion)	**0.6**

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 ≤ δ0 vs. H1: P1 - P2 > δ0

Sample Size						Normal Approximation		Binomial Enumeration	
N1	N2	N	P2	δ0	δ1	Power	Alpha	Power	Alpha
200	200	400	0.6	0.1	0.2	0.59849	0.025	0.60124	0.025
250	250	500	0.6	0.1	0.2	0.69615	0.025	0.69744	0.025
300	300	600	0.6	0.1	0.2	0.77397	0.025	0.77512	0.025
350	350	700	0.6	0.1	0.2	0.83433	0.025	0.83554	0.025

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

Example 5 – Finding the True Proportion Difference

Researchers have developed a new treatment with minimal side effects compared to the standard treatment. The researchers are limited by the number of subjects (140 per group) they can use to show the new treatment is superior. The new treatment will be deemed superior if it is at least 0.10 above the success rate of the standard treatment. The standard treatment has a success rate of about 0.65. The researchers want to know how much more successful the new treatment must be (in truth) to yield a test which has 90% power. The test statistic used will be the pooled Z test.

Setup

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** 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 .... **Effect Size (δ1, P1.1)**
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 .... **Better (H1: P1 - P2 > δ0)**
Test Type .... **Z-Test (Pooled)**
Power.. **0.90**
Alpha... **0.05**
Group Allocation .... **Equal (N1 = N2)**
Sample Size Per Group **140**
Input Type .... **Differences**
δ0 (Superiority by a Margin Difference) .... **-0.10**
P2 (Group 2 Proportion).. **0.75**

Output

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

Numeric Results

Solve For: Effect Size (δ1, P1.1)
Groups: 1 = Treatment, 2 = Reference
Test Statistic: Z-Test with Pooled Variance
Hypotheses: H0: P1 - P2 ≤ δ0 vs. H1: P1 - P2 > δ0

	Sample Size			Proportions			Difference		Alpha	
Power*	N1	N2	N	Superiority P1.0	Actual P1.1	Reference P2	Superiority δ0	Actual δ1	Target	Actual*†
0.9	140	140	280	0.75	0.906	0.65	0.1	0.256	0.025	0.024

* Power and actual alpha were computed using binomial enumeration of all possible outcomes.
† Warning: When solving for effect size with power computed using binomial enumeration, the target alpha level is not guaranteed. Actual alpha may be greater than target alpha in some cases.

With 140 subjects in each group, the new treatment must have a success rate 0.2560 higher than the current treatment (or about 0.9060) to have 90% power in the test of superiority.

Example 6 – Validation of Sample Size Calculation using Farrington and Manning (1990)

Farrington and Manning (1990), page 1451, present a sample size study in which P2 = 0.05, δ0 = 0.2, δ1=0.35, one-sided alpha = 0.05, and beta = 0.20. Using the Farrington and Manning test statistic, they found the sample size to be 80 in each group. They mention that the true power is 0.813.

Setup

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 6(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 .... **Sample Size**
Power Calculation Method .... **Normal Approximation**
Higher Proportions Are .... **Better (H1: P1 - P2 > δ0)**
Test Type .... **Likelihood Score (Farr. & Mann.)**
Power... **0.80**
Alpha.. **0.05**
Group Allocation .... **Equal (N1 = N2)**
Input Type .... **Differences**
δ0 (Superiority Difference) **0.2**
δ1 (Actual Difference) **0.35**
P2 (Group 2 Proportion)................................ **0.05**

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 ≤ δ0 vs. H1: P1 - P2 > δ0

Power		Sample Size			Proportions			Difference		
Target	Actual*	N1	N2	N	Superiority P1.0	Actual P1.1	Reference P2	Superiority δ0	Actual δ1	Alpha
0.8	0.80068	80	80	160	0.25	0.4	0.05	0.2	0.35	0.05

PASS also calculated the required sample size to be 80.

Next, to calculate the true power based on binomial enumeration for this sample size, we make the following changes to the template.

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**
Sample Size Per Group **80**

Numeric Results

Numeric Results

Solve For: Power
Groups: 1 = Treatment, 2 = Reference
Test Statistic: Farrington & Manning Likelihood Score Test
Hypotheses: H0: P1 - P2 ≤ δ0 vs. H1: P1 - P2 > δ0

	Sample Size			Proportions			Difference		Alpha	
Power*	N1	N2	N	Superiority P1.0	Actual P1.1	Reference P2	Superiority δ0	Actual δ1	Target	Actual*
0.8132	80	80	160	0.25	0.4	0.05	0.2	0.35	0.05	0.055

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

PASS also calculated the true power to be 0.813.