Chapter 205

Non-Zero Null Tests for the Difference Between Two Proportions

Introduction

This module computes power and sample size for hypothesis tests of the difference between two independent proportions where the null-hypothesized value is non-zero. 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	$\psi = \dfrac{p_1/(1-p_1)}{p_2/(1-p_2)} = \dfrac{p_1 q_2}{p_2 q_1}$

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

Difference

The (risk) difference, $\delta = p_1 - p_2$, is perhaps the most direct method of comparison between the two event probabilities. This parameter is easy to interpret and communicate. It gives the absolute impact of the treatment. However, there are subtle difficulties that can arise with its interpretation.

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

A similar situation occurs when the amount of possible difference is considered. Consider two events, one with a baseline event rate of 0.40 and the other with a rate of 0.02. What is the maximum decrease that can occur? Obviously, the first event rate can be decreased by an absolute amount of 0.40, while the second can only be decreased by a maximum of 0.02.

So, although creating the simple difference is a useful method of comparison, care must be taken that it is appropriate for the situation.

Hypothesis Tests

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

Difference

The (risk) difference, $\delta = p_1 - p_2$, is perhaps the most direct method for comparing two proportions. Three sets of statistical hypotheses can be formulated:

1. $H_0: p_1 - p_2 = \delta_0$ versus $H_1: p_1 - p_2 \neq \delta_0$; this is often called the *two-tailed test*.

2. $H_0: p_1 - p_2 \leq \delta_0$ versus $H_1: p_1 - p_2 > \delta_0$; this is often called the *upper-tailed test*.

3. $H_0: p_1 - p_2 \geq \delta_0$ versus $H_1: p_1 - p_2 < \delta_0$; this is often called the *lower-tailed test*.

Power Calculation

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

1. Find the critical value (or values in the case of a two-sided test) using the standard normal distribution. The critical value, $z_{critical}$, is that value of z that leaves exactly the target value of alpha in the appropriate tail of the normal distribution. For example, for an upper-tailed test with a target alpha of 0.05, the critical value is 1.645.

2. Compute the value of the test statistic, z_t, for every combination of x_{11} and x_{21}. Note that x_{11} ranges from 0 to n_1, and x_{21} ranges from 0 to n_2. A small value (around 0.0001) can be added to the zero cell counts to avoid numerical problems that occur when the cell value is zero.

3. If $z_t > z_{critical}$, the combination is in the rejection region. Call all combinations of x_{11} and x_{21} that lead to a rejection the set *A*.

4. Compute the power for given values of $p_{1.1}$ (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 Farrington and Manning statistic is substituted for the Gart and Nam statistic. Also, for large samples, the results for the odds ratio have not (to our knowledge) been published. In this case, we substitute the calculations based on the ratio.

Test Statistics

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

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

The constant, c, represents a continuity correction that is applied in some cases. When the continuity correction is not used, c is zero. In power calculations, the values of $\hat{p}_1$ and $\hat{p}_2$ are not known. The corresponding values of p_1 and p_2 under the alternative hypothesis are reasonable substitutes.

Following is a list of the test statistics available in **PASS**. The availability of several test statistics begs the question of which test statistic you should use. The answer is simple: <u>you should use the test statistic that you will use to analyze your data</u>. You may choose a method because it is a standard in your industry, because it seems to have better statistical properties, or because your statistical package calculates it. Whatever your reasons for selecting a certain test statistic, you should use the same test statistic during power or sample calculations.

Z Test (Pooled)

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

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

where

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

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

Z Test (Unpooled)

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

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

where

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

Z Test with Continuity Correction (Pooled)

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

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

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, 1 for upper-tailed, and both -1 and 1 for two-sided hypotheses.

Z Test with Continuity Correction (Unpooled)

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

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

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, 1 for upper-tailed, and both -1 and 1 for two-sided hypotheses.

T-Test

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

Miettinen and Nurminen's Likelihood Score Test

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

The formula for computing this test statistic is

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

where

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

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

$$\tilde{p}_2 = 2B\cos(A) - \frac{L_2}{3L_3}$$

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

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

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

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

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

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

$$L_3 = N$$

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 study the effectiveness of a new treatment. Historically, the standard treatment has enjoyed a 60% cure rate. The new treatment reduces the seriousness of certain side effects that occur with the standard treatment. Thus, the new treatment will be adopted even if it is slightly less effective than the standard treatment. The researchers will recommend adoption of the new treatment if it has a cure rate of at least 55%.

The researchers plan to use the Farrington and Manning likelihood score test statistic to analyze the data. They want to study the power of the one-sided Farrington and Manning test at group sample sizes ranging from 50 to 2000 for detecting a difference significantly greater than -0.05 when the actual cure rate of the new treatment ranges from 57% to 70%. The significance level will be 0.05.

Setup

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: P1 - P2 > δ0)**
Test Type...**Likelihood Score (Farr. & Mann.)**
Alpha..**0.05**
Group Allocation ....**Equal (N1 = N2)**
Sample Size Per Group **50 100 250 500 1000 1500 2000**
Input Type..**Differences**
δ0 (Difference|H0 = P1.0 - P2).......................**-0.05**
δ1 (Difference|H1 = P1.1 - P2).......................**-0.03 0.00 0.05 0.10**
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
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**	**P1|H0** **P1.0**	**P1|H1** **P1.1**	**Reference** **P2**	**Diff|H0** **δ0**	**Diff|H1** **δ1**	**Alpha**
0.07486	50	50	100	0.55	0.57	0.6	-0.05	-0.03	0.05
0.08748	100	100	200	0.55	0.57	0.6	-0.05	-0.03	0.05
0.11711	250	250	500	0.55	0.57	0.6	-0.05	-0.03	0.05
0.15829	500	500	1000	0.55	0.57	0.6	-0.05	-0.03	0.05
0.23101	1000	1000	2000	0.55	0.57	0.6	-0.05	-0.03	0.05
0.29755	1500	1500	3000	0.55	0.57	0.6	-0.05	-0.03	0.05
0.35965	2000	2000	4000	0.55	0.57	0.6	-0.05	-0.03	0.05
0.12866	50	50	100	0.55	0.60	0.6	-0.05	0.00	0.05
0.17843	100	100	200	0.55	0.60	0.6	-0.05	0.00	0.05
0.30784	250	250	500	0.55	0.60	0.6	-0.05	0.00	0.05
0.48830	500	500	1000	0.55	0.60	0.6	-0.05	0.00	0.05
0.73862	1000	1000	2000	0.55	0.60	0.6	-0.05	0.00	0.05
0.87534	1500	1500	3000	0.55	0.60	0.6	-0.05	0.00	0.05
0.94345	2000	2000	4000	0.55	0.60	0.6	-0.05	0.00	0.05
0.27032	50	50	100	0.55	0.65	0.6	-0.05	0.05	0.05
0.42722	100	100	200	0.55	0.65	0.6	-0.05	0.05	0.05
0.74745	250	250	500	0.55	0.65	0.6	-0.05	0.05	0.05
0.94785	500	500	1000	0.55	0.65	0.6	-0.05	0.05	0.05
0.99855	1000	1000	2000	0.55	0.65	0.6	-0.05	0.05	0.05
0.99997	1500	1500	3000	0.55	0.65	0.6	-0.05	0.05	0.05
1.00000	2000	2000	4000	0.55	0.65	0.6	-0.05	0.05	0.05
0.47083	50	50	100	0.55	0.70	0.6	-0.05	0.10	0.05
0.71963	100	100	200	0.55	0.70	0.6	-0.05	0.10	0.05
0.97003	250	250	500	0.55	0.70	0.6	-0.05	0.10	0.05
0.99959	500	500	1000	0.55	0.70	0.6	-0.05	0.10	0.05
1.00000	1000	1000	2000	0.55	0.70	0.6	-0.05	0.10	0.05
1.00000	1500	1500	3000	0.55	0.70	0.6	-0.05	0.10	0.05
1.00000	2000	2000	4000	0.55	0.70	0.6	-0.05	0.10	0.05

_____

* 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.
δ0	The difference in proportions under the null hypothesis, H0. δ0 = P1.0 - P2.
δ1	The difference in proportions under 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 different from the Group 2 (reference) proportion (P2) by a margin, with a non-zero null margin of -0.05 (H0: P1 - P2 ≤ -0.05 versus H1: P1 - P2 > -0.05). The comparison will be made using a one-sided, two-sample Score test (Farrington & Manning) with a Type I error rate (α) of 0.05. The reference group proportion is assumed to be 0.6. To detect a proportion difference (P1 - P2) of -0.03 (or P1 of 0.57) with sample sizes of 50 for the treatment group and 50 for the reference group, the power is 0.07486.
_____

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%	250	250	500	313	313	626	63	63	126
20%	500	500	1000	625	625	1250	125	125	250
20%	1000	1000	2000	1250	1250	2500	250	250	500
20%	1500	1500	3000	1875	1875	3750	375	375	750
20%	2000	2000	4000	2500	2500	5000	500	500	1000

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
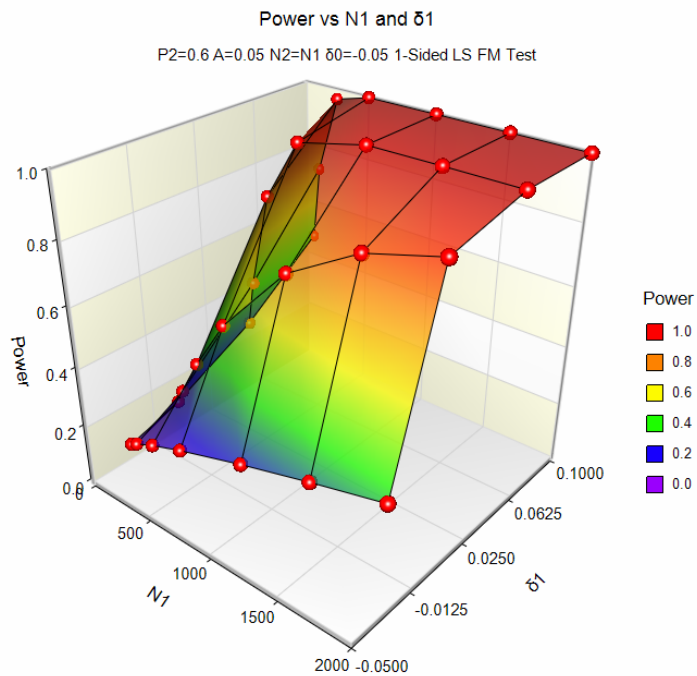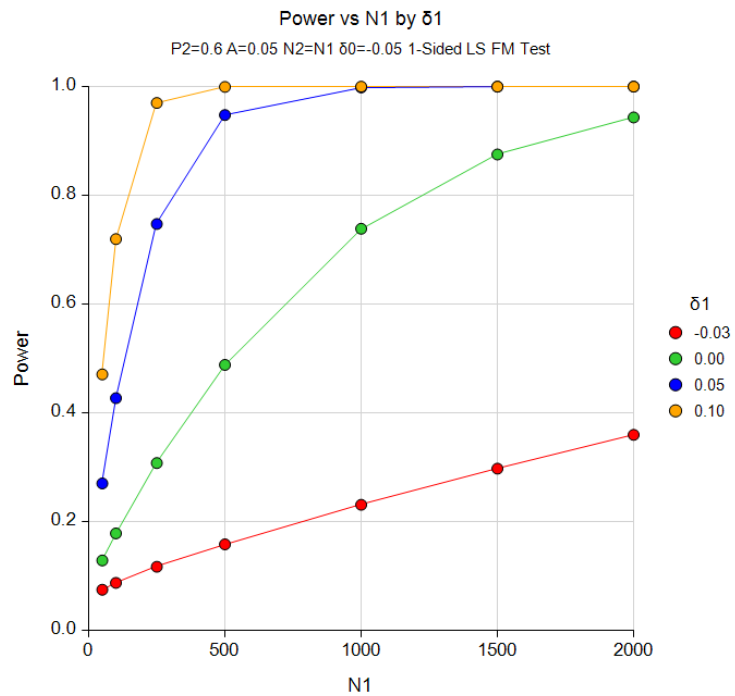
The values from the table are displayed in the above charts. These charts give us a quick look at the sample size that will be required for various values of δ1.

Example 2 – Finding the Sample Size

Continuing with the scenario given in Example 1, the researchers want to determine the sample size needed to achieve 80% power for each value of $\Delta 1$.

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: P1 - P2 > Δ0)**
Test Type	**Likelihood Score (Farr. & Mann.)**
Power	**0.80**
Alpha	**0.05**
Group Allocation	**Equal (N1 = N2)**
Input Type	**Differences**
Δ0 (Difference|H0 = P1.0 - P2)	**-0.05**
Δ1 (Difference|H1 = P1.1 - P2)	**-0.03 0.00 0.05 0.10**
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
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	P1|H0 P1.0	P1|H1 P1.1	Reference P2	Diff|H0 δ0	Diff|H1 δ1	Alpha
0.8	0.80003	7491	7491	14982	0.55	0.57	0.6	-0.05	-0.03	0.05
0.8	0.80019	1186	1186	2372	0.55	0.60	0.6	-0.05	0.00	0.05
0.8	0.80084	290	290	580	0.55	0.65	0.6	-0.05	0.05	0.05
0.8	0.80113	125	125	250	0.55	0.70	0.6	-0.05	0.10	0.05

* Power was computed using the normal approximation method.

The required sample size will depend a great deal on the value of $\delta 1$. The researchers should spend time determining the most accurate value for $\delta 1$.

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 0.1.

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: P1 - P2 > δ0)**
Test Type	**Likelihood Score (Farr. & Mann.)**
Alpha	**0.05**
Group Allocation	**Equal (N1 = N2)**
Sample Size Per Group	**50 100 150 200**
Input Type	**Differences**
δ0 (Difference|H0 = P1.0 - P2)	**-0.05**
δ1 (Difference|H1 = P1.1 - P2)	**0.05**
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
50	50	100	0.6	-0.05	0.05	0.05	0.2720	0.2720	0.2064	0.2096	0.2694	0.2720	0.2694	0.2720
100	100	200	0.6	-0.05	0.05	0.05	0.4207	0.4248	0.3663	0.3663	0.4178	0.4207	0.4207	0.4207
150	150	300	0.6	-0.05	0.05	0.05	0.5540	0.5540	0.5054	0.5054	0.5504	0.5540	0.5519	0.5519
200	200	400	0.6	-0.05	0.05	0.05	0.6654	0.6683	0.6286	0.6286	0.6624	0.6683	0.6654	0.6654

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
50	50	100	0.6	-0.05	0.05	0.05	0.0527	0.0527	0.0342	0.0343	0.0526	0.0527	0.0526	0.0527
100	100	200	0.6	-0.05	0.05	0.05	0.0499	0.0500	0.0369	0.0369	0.0499	0.0499	0.0499	0.0499
150	150	300	0.6	-0.05	0.05	0.05	0.0509	0.0509	0.0398	0.0398	0.0509	0.0509	0.0509	0.0509
200	200	400	0.6	-0.05	0.05	0.05	0.0479	0.0482	0.0387	0.0387	0.0477	0.0482	0.0479	0.0479

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

Plots

Power vs N1 by Test
δ1=0.05 P2=0.6 A=0.05 N2=N1 δ0=-0.05 1-Sided Test
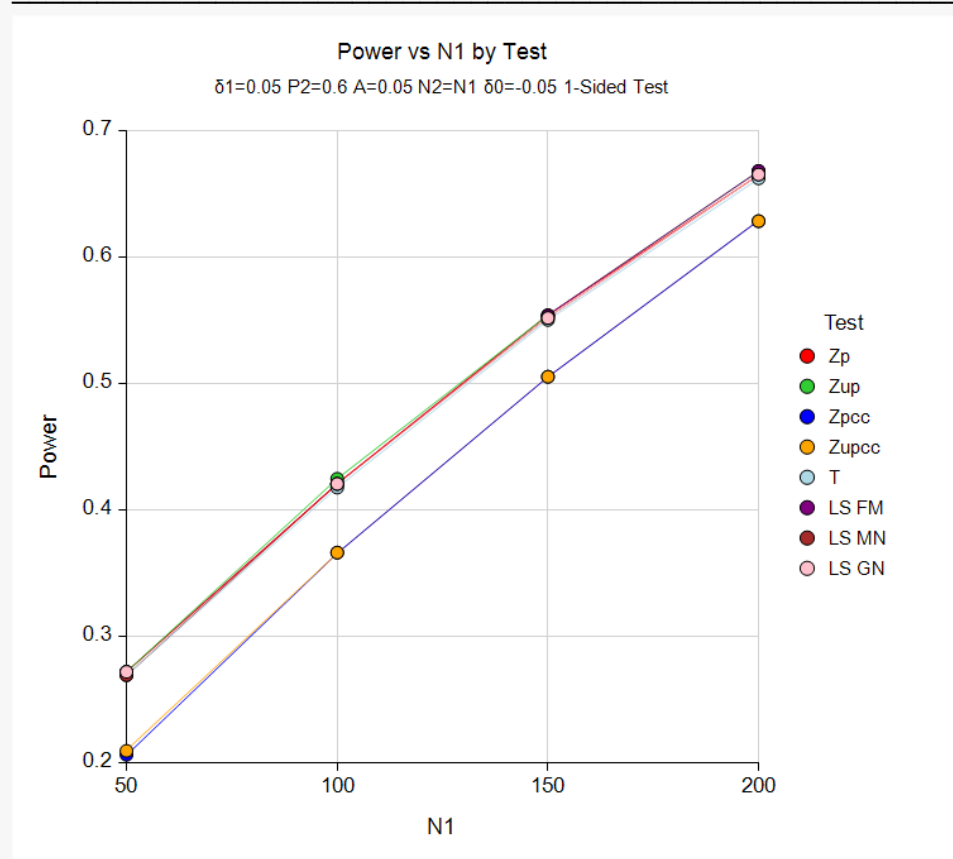
It is interesting to note that the powers of the continuity-corrected test statistics are consistently lower than the other tests. This occurs because the actual alpha achieved by these tests is lower than for the other tests.

Example 4 – 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: P1 - P2 > δ0)**

Test Type..**Likelihood Score (Farr. & Mann.)**

Alpha...**0.05**

Group Allocation ....**Equal (N1 = N2)**

Sample Size Per Group **50 100 150 200**

Input Type...**Differences**

δ0 (Difference|H0 = P1.0 - P2).......................**-0.05**

δ1 (Difference|H1 = P1.1 - P2).......................**0.05**

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
50	50	100	0.6	-0.05	0.05	0.27032	0.05	0.27200	0.0527
100	100	200	0.6	-0.05	0.05	0.42722	0.05	0.42069	0.0499
150	150	300	0.6	-0.05	0.05	0.55774	0.05	0.55405	0.0509
200	200	400	0.6	-0.05	0.05	0.66361	0.05	0.66826	0.0482

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

Example 5 – Finding the Power after Completing an Experiment

Researchers are studying the effectiveness of a new treatment for cancer. Historically, the standard treatment has enjoyed a 52% cure rate. The new experimental treatment is believed to be better, but it costs much more to administer. After weighing cost versus effectiveness, the researchers decided that they will adopt the new treatment if the cure rate is at least 59%. They conduct a study in which 200 patients are given the new treatment, and 200 are given the standard regimen. They find that 66% are cured by the new treatment, while 52% are cured by the standard treatment. The Farrington and Manning likelihood score test, however, indicates that the results are not statistically significant for alpha = 0.05. They now desire to compute the power for a range of alternative values.

Note that a range of alternatives is used in computing the power instead of the actual difference from the study. The power should be computed at values representing practically significant differences from the null value.

Setup

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 .... **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: P1 - P2 > δ0)**
Test Type .... **Likelihood Score (Farr. & Mann.)**
Alpha **0.05**
Group Allocation .... **Equal (N1 = N2)**
Sample Size Per Group **200**
Input Type .... **Differences**
δ0 (Difference|H0 = P1.0 - P2) **0.07**
δ1 (Difference|H1 = P1.1 - P2) .... **0.08 to 0.20 by 0.02**
P2 (Group 2 Proportion) **0.52**

Output

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

Numeric Results
_____

Solve For: Power
Test Statistic: Farrington & Manning Likelihood Score Test
Hypotheses: H0: P1 - P2 ≤ δ0   vs.   H1: P1 - P2 > δ0
_____

	Sample Size			Proportions			Difference		Alpha	
				P1|H0	P1|H1	Reference	Diff|H0	Diff|H1		
Power*	N1	N2	N	P1.0	P1.1	P2	δ0	δ1	Target	Actual*
0.07152	200	200	400	0.59	0.60	0.52	0.07	0.08	0.05	0.0479
0.14459	200	200	400	0.59	0.62	0.52	0.07	0.10	0.05	0.0479
0.25814	200	200	400	0.59	0.64	0.52	0.07	0.12	0.05	0.0479
0.40895	200	200	400	0.59	0.66	0.52	0.07	0.14	0.05	0.0479
0.57829	200	200	400	0.59	0.68	0.52	0.07	0.16	0.05	0.0479
0.73684	200	200	400	0.59	0.70	0.52	0.07	0.18	0.05	0.0479
0.85910	200	200	400	0.59	0.72	0.52	0.07	0.20	0.05	0.0479
_____

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

Plots
_____
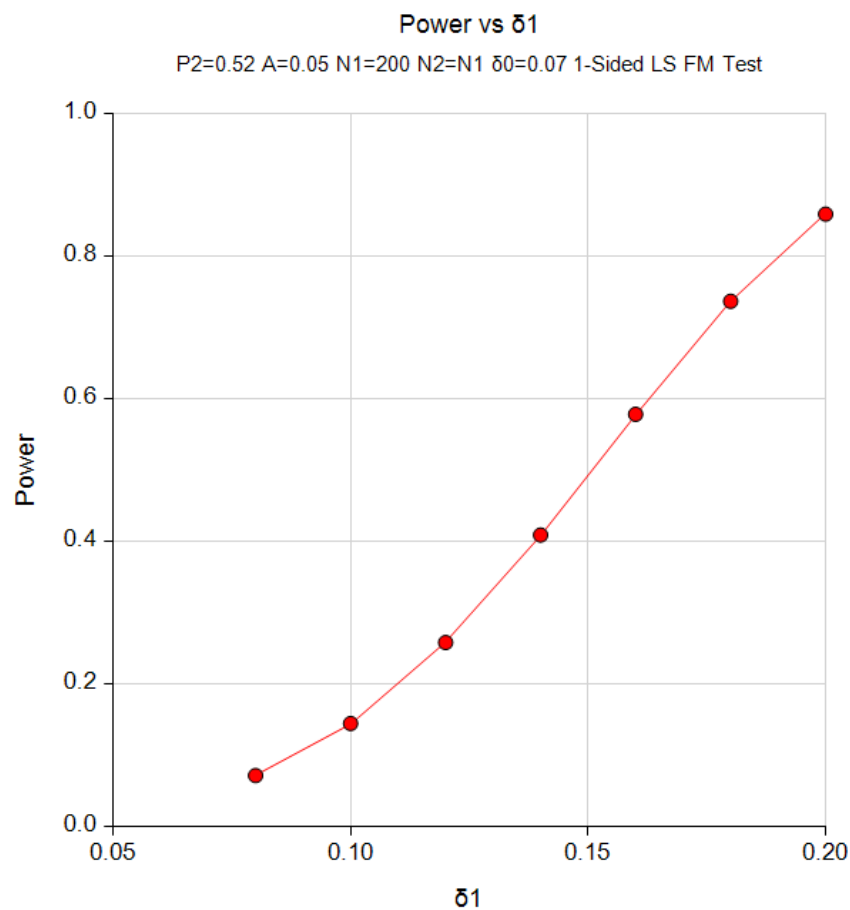
The power depends a great deal on the value of δ1 for this sample size. It is evident that the power is quite low for the majority of alternative values studied.

Example 6 – Validation of Sample Size Calculation for the Farrington and Manning Test using Machin et al. (1997)

Machin et al. (1997), page 106, present a sample size study in which P2 = 0.5, δ0 = -0.2, δ1=0, one-sided alpha = 0.1, and beta = 0.2. Using the Farrington and Manning test statistic, they found the sample size to be 55 in each group.

Setup

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** 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: P1 - P2 > δ0)**
Test Type..**Likelihood Score (Farr. & Mann.)**
Power..**0.80**
Alpha...**0.10**
Group Allocation ....**Equal (N1 = N2)**
Input Type...**Differences**
δ0 (Difference|H0 = P1.0 - P2)......................**-0.2**
δ1 (Difference|H1 = P1.1 - P2).....................**0**
P2 (Group 2 Proportion)................................**0.5**

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: P1 - P2 ≤ δ0   vs.   H1: P1 - P2 > δ0
_____

Power		Sample Size			Proportions			Difference		
					P1|H0	P1|H1	Reference	Diff|H0	Diff|H1	
Target	Actual*	N1	N2	N	P1.0	P1.1	P2	δ0	δ1	Alpha
0.8	0.80009	55	55	110	0.3	0.5	0.5	-0.2	0	0.1

_____

* Power was computed using the normal approximation method.

PASS found the required sample size to be 55, which corresponds exactly to the results of Machin et al. (1997).

Example 7 – 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, $\delta0$ = 0.2, $\delta1$=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 7(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**
Alternative Hypothesis ....**One-Sided (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 (Difference|H0 = P1.0 – P2)......................**0.2**
δ1 (Difference|H1 = P1.1 – P2)......................**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
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	P1|H0 P1.0	P1|H1 P1.1	Reference P2	Diff|H0 δ0	Diff|H1 δ1	Alpha
0.8	0.80068	80	80	160	0.25	0.4	0.05	0.2	0.35	0.05

* Power was computed using the normal approximation method.

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
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	P1|H0 P1.0	P1|H1 P1.1	Reference P2	Diff|H0 δ0	Diff|H1 δ1	Target	Actual*
0.8132	80	80	160	0.25	0.4	0.05	0.2	0.35	0.05	0.0553

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

PASS also calculated the true power to be 0.813.

Example 8 – Validation of Sample Size Calculation for the Unpooled Z-Test using Chow, Shao, and Wang (2008)

Chow, Shao, and Wang (2008) page 92 gives the results of a sample size calculation for an unpooled Z-test for non-inferiority. This procedure can be used for the same test. When P1.0 = 0.55 (from δ = -0.1), P1.1 =0.85, P2 = 0.65, power = 0.8, and alpha = 0.05, Chow, Shao, and Wang (2008) reports a required sample size of 25.

Setup

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

Design Tab

Solve For ....**Sample Size**
Power Calculation Method............................**Normal Approximation**
Alternative Hypothesis ....**One-Sided (H1: δ1 > δ0)**
Test Type..**Z-Test (Unpooled)**
Power...**0.80**
Alpha..**0.05**
Group Allocation ....**Equal (N1 = N2)**
Input Type...**Proportions**
P1.0 (Group 1 Proportion|H0) **0.55**
P1.1 (Group 1 Proportion|H1) **0.85**
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: Z-Test with Unpooled Variance
Hypotheses: H0: P1 - P2 ≤ δ0   vs.   H1: P1 - P2 > δ0
_____

| Power | | Sample Size | | | Proportions | | | Difference | | |
Target	Actual*	N1	N2	N	P1|H0 P1.0	P1|H1 P1.1	Reference P2	Diff|H0 δ0	Diff|H1 δ1	Alpha
0.8	0.80858	25	25	50	0.55	0.85	0.65	-0.1	0.2	0.05

* Power was computed using the normal approximation method.

PASS also found the required sample size to be 25.

Example 9 – Validation of Sample Size Calculation for the Unpooled Z-Test using Julius and Campbell (2012)

Julius and Campbell (2012) presents Table XIII gives the results of sample size calculations for an unpooled Z-test for non-inferiority for P2 between 0.7 and 0.9, $|\delta 0|$ between 0.05 and 0.20 and $\delta 1$ between -0.05 and 0.05. Sample sizes are calculated for 90% power and alpha = 0.025. This example will replicate all values of $\delta 1$ for P1 = 0.70 and $|\delta 0|$ = 0.20 in the table.

The sample sizes reported in the table for $\delta 1$ between -0.05 and 0.05 are 205, 179, 157, 139, 124, 111, 100, 90, 81, 74, and 67.

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 8** 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: P1 - P2 > δ0)**
Test Type..**Z-Test (Unpooled)**
Power...**0.90**
Alpha..**0.025**
Group Allocation ....**Equal (N1 = N2)**
Input Type...**Differences**
δ0 (Difference|H0 = P1.0 - P2).....................**-0.2**
δ1 (Difference|H1 = P1.1 - P2).......................**-0.05 to 0.05 by 0.01**
P2 (Group 2 Proportion)...............................**0.70**

Output

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

Numeric Results
_____

Solve For: Sample Size
Test Statistic: Z-Test with Unpooled Variance
Hypotheses: H0: P1 - P2 ≤ δ0   vs.   H1: P1 - P2 > δ0
_____

Power		Sample Size			Proportions			Difference		
Target	Actual*	N1	N2	N	P1|H0 P1.0	P1|H1 P1.1	Reference P2	Diff|H0 δ0	Diff|H1 δ1	Alpha
0.9	0.90096	205	205	410	0.5	0.65	0.7	-0.2	-0.05	0.025
0.9	0.90111	179	179	358	0.5	0.66	0.7	-0.2	-0.04	0.025
0.9	0.90047	157	157	314	0.5	0.67	0.7	-0.2	-0.03	0.025
0.9	0.90067	139	139	278	0.5	0.68	0.7	-0.2	-0.02	0.025
0.9	0.90142	124	124	248	0.5	0.69	0.7	-0.2	-0.01	0.025
0.9	0.90172	111	111	222	0.5	0.70	0.7	-0.2	0.00	0.025
0.9	0.90257	100	100	200	0.5	0.71	0.7	-0.2	0.01	0.025
0.9	0.90203	90	90	180	0.5	0.72	0.7	-0.2	0.02	0.025
0.9	0.90049	81	81	162	0.5	0.73	0.7	-0.2	0.03	0.025
0.9	0.90228	74	74	148	0.5	0.74	0.7	-0.2	0.04	0.025
0.9	0.90073	67	67	134	0.5	0.75	0.7	-0.2	0.05	0.025

* Power was computed using the normal approximation method.

The sample sizes from **PASS** match Table XIII of Julius and Campbell (2012) exactly.

We should point out that the values reported in Table XIII for P1 – P2 = -0.04 where |δ0| = 0.05 (45845, 41537, 36178, etc.) are incorrect for all P1 given. If you calculate the table values using formula (30) of Julius and Campbell (2012) or using **PASS**, you'll find that each sample size in the table is 200 more than the correct value. All other values in Table XIII are correct.