Chapter 200

Tests for Two Proportions

Introduction

This module computes power and sample size for hypothesis tests of the difference, ratio, or odds ratio of two independent proportions. The test statistics analyzed by this procedure assume that the difference between the two proportions is zero or their ratio is one under the null hypothesis. The *non-null* (*offset*) *case* is discussed in another procedure. This procedure computes and compares the power achieved by each of several test statistics that have been proposed.

For example, suppose you want to compare two methods for treating cancer. Your experimental design might be as follows. Select a sample of patients and randomly assign half to one method and half to the other. After five years, determine the proportion surviving in each group and test whether the difference in the proportions is significantly different from zero.

The power calculations assume that random samples are drawn from two separate 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 $q_2 = 1 - p_2$.

The 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 also 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

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

Comparing Two Proportions

When analyzing studies such as this, one usually wants to compare the two binomial probabilities, p_1 and p_2. Common measures for comparing these quantities are the difference and the ratio. If the binomial probabilities are expressed in terms of odds rather than probabilities, another common 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 tests analyzed by this routine are for the *null case*. This refers to the values of the above parameters under the null hypothesis. In the *null case*, the difference is zero and the ratios are one under the null hypothesis. In the *non-null case*, discussed in another chapter, the difference is some value other than zero and the ratios are some value other than one. The non-null case often appears in equivalence and non-inferiority testing.

Hypothesis Tests

Several statistical tests have been developed for testing the inequality of two proportions. For large samples, the powers of the various tests are about the same. However, for small samples, the differences in the powers can be quite large. Hence, it is important to base the power analysis on the test statistic that will be used to analyze the data. If you have not selected a test statistic, you may wish to determine which one offers the best power in your situation. No single test is the champion in every situation, so you must 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 measure for comparing two proportions. Three sets of statistical hypotheses can be formulated:

1. $H_0: p_1 - p_2 = 0$ versus $H_1: p_1 - p_2 \neq 0$; this is often called the *two-tailed test*.
2. $H_0: p_1 - p_2 \leq 0$ versus $H_1: p_1 - p_2 > 0$; this is often called the *upper-tailed test*.
3. $H_0: p_1 - p_2 \geq 0$ versus $H_1: p_1 - p_2 < 0$; this is often called the *lower-tailed test*.

The traditional approach for testing these hypotheses has been to use the Pearson chi-square test for large samples, the Yates chi-square for intermediate sample sizes, and the Fisher Exact test for small samples.

Recently, some authors have begun questioning this solution. For example, based on exact enumeration, Upton (1982) and D'Agostino (1988) conclude that the Fisher Exact test and Yates test should never be used.

Ratio

The (risk) ratio, $\phi = p_1 / p_2$, is often preferred to the difference when the baseline proportion is small (less than 0.1) or large (greater than 0.9) because it expresses the difference as a percentage rather than an amount. In this null case, the null hypothesized ratio of proportions, ϕ_0, is one. Three sets of statistical hypotheses can be formulated:

1. $H_0: p_1 / p_2 = \phi_0$ versus $H_1: p_1 / p_2 \neq \phi_0$; this is often called the *two-tailed test*.
2. $H_0: p_1 / p_2 \leq \phi_0$ versus $H_1: p_1 / p_2 > \phi_0$; this is often called the *upper-tailed test*.
3. $H_0: p_1 / p_2 \geq \phi_0$ versus $H_1: p_1 / p_2 < \phi_0$; this is often called the *lower-tailed test*.

Odds Ratio

The odds ratio, $\psi = \frac{o_1}{o_2} = \frac{p_1/(1-p_1)}{p_2/(1-p_2)} = \frac{p_1 q_2}{p_2 q_1}$, is sometimes used to compare the two proportions because of its statistical properties and because some experimental designs require its use. In this null case, the null hypothesized odds ratio, $\psi_0\, \psi_0$, is one. Three sets of statistical hypotheses can be formulated:

1. $H_0: \psi = \psi_0$ versus $H_1: \psi \neq \psi_0$; this is often called the *two-tailed test*.
2. $H_0: \psi \leq \psi_0$ versus $H_1: \psi > \psi_0$; this is often called the *upper-tailed test*.
3. $H_0: \psi \geq \psi_0$ versus $H_1: \psi < \psi_0$; 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} 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 and p_2 as

$$1 - \beta = \sum_{A} \binom{n_1}{x_{11}} p_1^{x_{11}} q_1^{n_1 - x_{11}} \binom{n_2}{x_{21}} p_2^{x_{21}} q_2^{n_2 - x_{21}}.$$

5. Compute the actual value of alpha achieved by the design by substituting p_2 for p_1 to obtain

$$\alpha^* = \sum_A \binom{n_1}{x_{11}} p_2^{x_{11}} q_2^{n_1 - x_{11}} \binom{n_2}{x_{21}} p_2^{x_{21}} q_2^{n_2 - x_{21}}$$

$$= \sum_A \binom{n_1}{x_{11}} \binom{n_2}{x_{21}} p_2^{x_{11}+x_{21}} q_2^{n_1+n_2-x_{11}-x_{21}}.$$

When the values of n_1 and n_2 are large (say over 200), these formulas may take a little time to evaluate. In this case, a large sample approximation may be used.

Test Statistics

The various test statistics that are available in this routine are listed next.

Fisher's Exact Test

The most useful reference we found for power analysis of Fisher's Exact test was in the StatXact 5 (2001) documentation. The material present here is summarized from Section 26.3 (pages 866 – 870) of the StatXact-5 documentation. In this case, the test statistic is

$$T = -\ln\left[\frac{\binom{n_1}{x_1}\binom{n_2}{x_2}}{\binom{N}{m}}\right]$$

The null distribution of T is based on the hypergeometric distribution. It is given by

$$\Pr(T \geq t | m, H_0) = \sum_{A(m)} \left[\frac{\binom{n_1}{x_1}\binom{n_2}{x_2}}{\binom{N}{m}}\right]$$

where

$$A(m) = \{\text{all pairs } x_1, x_2 \text{ such that } x_1 + x_2 = m, \text{given } T \geq t\}$$

Conditional on *m*, the critical value, t_α, is the smallest value of *t* such that

$$\Pr(T \geq t_\alpha | m, H_0) \leq \alpha$$

The power is defined as

$$1 - \beta = \sum_{m=0}^{N} P(m) \Pr(T \geq t_\alpha | m, H_1)$$

where

$$\Pr(T \geq t_\alpha | m, H_1) = \sum_{A(m, T \geq t_\alpha)} \left[\frac{b(x_1, n_1, p_1) b(x_2, n_2, p_2)}{\sum_{A(m)} b(x_1, n_1, p_1) b(x_2, n_2, p_2)} \right]$$

$$\begin{aligned} P(m) &= \Pr(x_1 + x_2 = m | H_1) \\ &= \mathrm{b}(x_1, n_1, p_1) b(x_2, n_2, p_2) \end{aligned}$$

$$b(x, n, p) = \binom{n}{x} p^x (1-p)^{n-x}$$

When the normal approximation is used to compute power, the result is based on the pooled, continuity corrected Z test.

Z Test (or Chi-Square Test) (Pooled and Unpooled)

This test statistic 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.

Both *pooled* and *unpooled* versions of this test have been discussed in statistical literature. The pooling refers to the way in which the standard error is estimated. In the pooled version, the two proportions are averaged, and only one proportion is used to estimate the standard error. In the unpooled version, the two proportions are used separately.

The formula for the test statistic is

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

Pooled Version

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

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

Unpooled Version

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

Power

The power of this test is computed using the enumeration procedure described above. For large sample sizes, the following approximation is used as presented in Chow et al. (2008).

1. Find the critical value (or values in the case of a two-sided test) using the standard normal distribution. The critical value is that value of z that leaves exactly the target value of alpha in the tail.
2. Use the normal approximation to binomial distribution to compute binomial probabilities, compute the power for the pooled and unpooled tests, respectively, using

Pooled: $1-\beta = \Pr\left(Z < \frac{z_\alpha \sigma_{D,p} + (p_1 - p_2)}{\sigma_{D,u}}\right)$ **Unpooled:** $1-\beta = \Pr\left(Z < \frac{z_\alpha \sigma_{D,u} + (p_1 - p_2)}{\sigma_{D,u}}\right)$

where

$$\sigma_{D,u} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}} \quad \text{(unpooled standard error)}$$

$$\sigma_{D,p} = \sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \quad \text{(pooled standard error)}$$

with $\bar{p} = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$ and $\bar{q} = 1 - \bar{p}$

Z Test (or Chi-Square Test) with Continuity Correction (Pooled and Unpooled)

Frank Yates is credited with proposing a correction to the Pearson Chi-Square test for the lack of continuity in the binomial distribution. However, the correction was in common use when he proposed it in 1922. Although this test is often 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.

Both *pooled* and *unpooled* versions of this test have been discussed in statistical literature. The pooling refers to the way in which the standard error is estimated. In the pooled version, the two proportions are averaged, and only one proportion is used to estimate the standard error. In the unpooled version, the two proportions are used separately.

The continuity corrected z-test is

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

where *F* is -1 for lower-tailed, 1 for upper-tailed, and both -1 and 1 for two-sided hypotheses.

Pooled Version

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

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

Unpooled Version

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

Power

The power of this test is computed using the enumeration procedure described for the *z*-test above. For large samples, approximate results based on the normal approximation to the binomial are used.

Conditional Mantel-Haenszel Test

The conditional Mantel-Haenszel test, see Lachin (2000) page 40, is based on the *index frequency*, x_{11}, from the 2x2 table. The formula for the z-statistic is

$$z = \frac{x_{11} - E(x_{11})}{\sqrt{V_c(x_{11})}}$$

where

$$E(x_{11}) = \frac{n_1 m_1}{N}$$

$$V_c(x_{11}) = \frac{n_1 n_2 m_1 m_2}{N^2(N-1)}$$

Power

The power of this test is computed using the enumeration procedure described above.

Likelihood Ratio Test

In 1935, Wilks showed that the following quantity has a chi-square distribution with one degree of freedom. Using this test statistic to compare proportions is presented, among other places, in Upton (1982). The likelihood ratio test statistic is computed as

$$LR = 2\begin{bmatrix} a\ln(a) + b\ln(b) + c\ln(c) + d\ln(d) + \\ N\ln(N) - s\ln(s) - f\ln(f) - m\ln(m) - n\ln(n) \end{bmatrix}$$

Power

The power of this test is computed using the enumeration procedure described above. When large sample results are needed, the results for the *z* test are used.

T-Test

Based on a study of the behavior of several tests, D'Agostino (1988) and Upton (1982) proposed using the usual two-sample *t*-test for testing whether 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. The test statistic is computed as

$$t_{N-2} = (ad - bc)\left(\frac{N-2}{N(nac + mbd)}\right)^{\frac{1}{2}}$$

which can be compared to the *t* distribution with *N*-2 degrees of freedom.

Power

The power of this test is computed using the enumeration procedure described above, except that the *t* tables are used instead of the standard normal tables.

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. Researchers want to compute the power of the two-sided *z*-test at group sample sizes ranging from 50 to 650 for detecting differences of 0.05 and 0.10 in the cure rate at the 0.05 significance level.

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	**Two-Sided**
Test Type	**Z-Test (Pooled)**
Alpha	**0.05**
Group Allocation	**Equal (N1 = N2)**
Sample Size Per Group	**50 to 650 by 100**
Input Type	**Differences**
D1 (Difference|H1 = P1–P2)	**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
Groups: 1 = Treatment, 2 = Control
Test Type: Z-Test with Pooled Variance
Hypotheses: H0: P1 - P2 = 0 vs. H1: P1 - P2 ≠ 0

	Sample Size			Proportions		Difference	
Power*	N1	N2	N	P1	P2	δ1	Alpha
0.08073	50	50	100	0.65	0.6	0.05	0.05
0.14513	150	150	300	0.65	0.6	0.05	0.05
0.21093	250	250	500	0.65	0.6	0.05	0.05
0.27652	350	350	700	0.65	0.6	0.05	0.05
0.34064	450	450	900	0.65	0.6	0.05	0.05
0.40234	550	550	1100	0.65	0.6	0.05	0.05
0.46095	650	650	1300	0.65	0.6	0.05	0.05
0.18089	50	50	100	0.70	0.6	0.10	0.05
0.44240	150	150	300	0.70	0.6	0.10	0.05
0.65033	250	250	500	0.70	0.6	0.10	0.05
0.79333	350	350	700	0.70	0.6	0.10	0.05
0.88326	450	450	900	0.70	0.6	0.10	0.05
0.93640	550	550	1100	0.70	0.6	0.10	0.05
0.96636	650	650	1300	0.70	0.6	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 at which power and sample size calculations are made. This is the treatment or experimental group.
P2	The proportion for Group 2. This is the standard, reference, or control group.
δ1	The difference assumed for power and sample size calculations. δ1 = P1 - 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 (control) proportion (P2) (H0: P1 - P2 = 0 versus H1: P1 - P2 ≠ 0). The comparison will be made using a two-sided, two-sample Z-Test with pooled variance, with a Type I error rate (α) of 0.05. The control group proportion (P2) is assumed to be 0.6. To detect a proportion difference (P1 - P2) of 0.05 (or P1 of 0.65) with sample sizes of 50 for Group 1 (treatment) and 50 for Group 2 (control), the power is 0.08073.

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%	150	150	300	188	188	376	38	38	76
20%	250	250	500	313	313	626	63	63	126
20%	350	350	700	438	438	876	88	88	176
20%	450	450	900	563	563	1126	113	113	226
20%	550	550	1100	688	688	1376	138	138	276
20%	650	650	1300	813	813	1626	163	163	326

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.

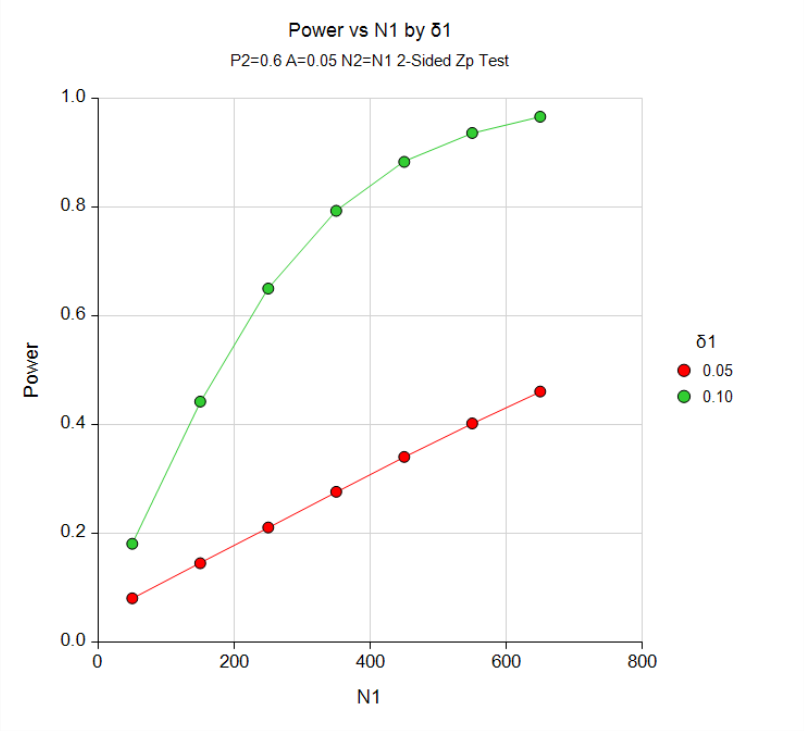
References

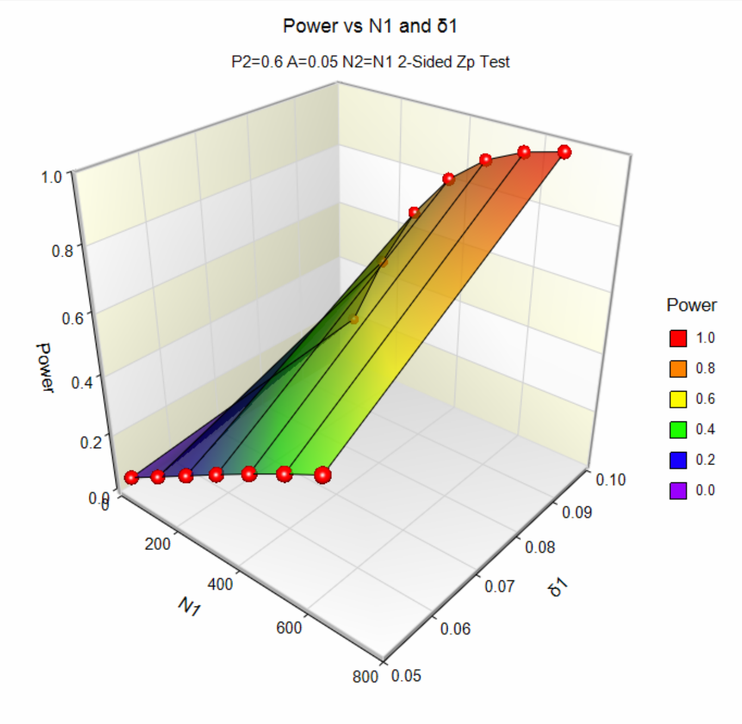
Chow, S.C., Shao, J., and Wang, H. 2008. Sample Size Calculations in Clinical Research, Second Edition. Chapman & Hall/CRC. Boca Raton, Florida.
D'Agostino, R.B., Chase, W., and Belanger, A. 1988. 'The Appropriateness of Some Common Procedures for Testing the Equality of Two Independent Binomial Populations', The American Statistician, August 1988, Volume 42 Number 3, pages 198-202.
Fleiss, J. L., Levin, B., and Paik, M.C. 2003. Statistical Methods for Rates and Proportions. Third Edition. John Wiley & Sons. New York.
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.
Ryan, Thomas P. 2013. Sample Size Determination and Power. John Wiley & Sons. Hoboken, New Jersey.

This report shows the values of each of the parameters, one scenario per row. The values from this table are displayed in the plots below.

Plots Section

Plots





The values from the table are displayed on the above charts.

Example 2 – Finding the Sample Size

A clinical trial is being designed to test effectiveness of new drug in reducing mortality. Suppose the current cure rate during the first year is 0.44. The sample size should be large enough to detect a difference in the cure rate of 0.10. Assuming the test statistic is a two-sided *z*-test with a significance level of 0.05, what sample size will be necessary to achieve 90% power?

In this example we'll show you how to set up the calculation by inputting proportions, differences, ratios, and odds ratios. In all cases, you'll see that the sample sizes are exactly the same. The only difference is in the way the effect size is specified.

If P2 = 0.44 and D1 = 0.10, then P1 = D1 + P2 = 0.54.

Setup (Proportions)

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 2a** 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 .... **Two-Sided**
Test Type.. **Z-Test (Pooled)**
Power... **0.90**
Alpha.. **0.05**
Group Allocation .... **Equal (N1 = N2)**
Input Type... **Proportions**
P1 (Group 1 Proportion|H1) **0.54**
P2 (Group 2 Proportion)................................ **0.44**

Output (Proportions)

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

Numeric Results

Solve For: Sample Size
Groups: 1 = Treatment, 2 = Control
Test Type: Z-Test with Pooled Variance
Hypotheses: H0: P1 - P2 = 0 vs. H1: P1 - P2 ≠ 0

Power		Sample Size			Proportions		Difference	
Target	Actual*	N1	N2	N	P1	P2	δ1	Alpha
0.9	0.9005	524	524	1048	0.54	0.44	0.1	0.05

* Power was computed using the normal approximation method.

The required sample size is 524 per group. These results use the large sample approximation. As an exercise, change the Power Calculation Method to "Binomial Enumeration". When this is done, the sample size is 521—not much of a difference from the 524 that was found by approximate methods. The actual alpha is 0.0493 which is very close to the target of 0.05.

Setup (Differences)

The setup for differences is exactly the same as that for proportions except for the following two inputs on the Design tab. You may then make the appropriate entries as listed below, or open **Example 2b** by going to the **File** menu and choosing **Open Example Template**.

Design Tab

Input Type .... **Differences**
D1 (Difference|H1 = P1–P2) **0.10**

Output (Differences)

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

Numeric Results

Solve For: Sample Size
Groups: 1 = Treatment, 2 = Control
Test Type: Z-Test with Pooled Variance
Hypotheses: H0: P1 - P2 = 0 vs. H1: P1 - P2 ≠ 0

Power		Sample Size			Proportions		Difference	
Target	Actual*	N1	N2	N	P1	P2	δ1	Alpha
0.9	0.9005	524	524	1048	0.54	0.44	0.1	0.05

* Power was computed using the normal approximation method.

This report shows the same sample size when the effect size is specified using the difference.

Setup (Ratios)

The setup for differences is exactly the same as that for proportions except for the following two inputs on the Design tab. You may then make the appropriate entries as listed below, or open **Example 2c** by going to the **File** menu and choosing **Open Example Template**.

If P1 = 0.54 and P2 = 0.44, then R1 = P1/P2 = 0.54/0.44 = 1.227272727.

Design Tab

Input Type..**Ratios**
R1 (Ratio|H1 = P1/P2)....................................**1.227272727**

Output (Ratios)

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

Numeric Results

Solve For: Sample Size
Groups: 1 = Treatment, 2 = Control
Test Type: Z-Test with Pooled Variance
Hypotheses: H0: P1 / P2 = 1 vs. H1: P1 / P2 ≠ 1

Power		Sample Size			Proportions		Ratio	
Target	Actual*	N1	N2	N	P1	P2	R1	Alpha
0.9	0.9005	524	524	1048	0.54	0.44	1.227	0.05

* Power was computed using the normal approximation method.

This report shows the same sample size when the effect size is specified using the ratio.

Setup (Odds Ratios)

The setup for differences is exactly the same as that for proportions except for the following two inputs on the Design tab. You may then make the appropriate entries as listed below, or open **Example 2d** by going to the **File** menu and choosing **Open Example Template**.

If P1 = 0.54 and P2 = 0.44, then OR1 = O1/O2 = [0.54/(1 – 0.54)]/[0.44(1 – 0.44)] = 1.494071146

Design Tab

Input Type..**Odds Ratios**
OR1 (Odds Ratio|H1 = O1/O2).....................**1.494071146**

Output (Odds Ratios)

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

Numeric Results

Solve For: Sample Size
Groups: 1 = Treatment, 2 = Control
Test Type: Z-Test with Pooled Variance
Hypotheses: H0: OR = 1 vs. H1: OR ≠ 1

Power		Sample Size			Proportions		Odds Ratio	
Target	Actual*	N1	N2	N	P1	P2	OR1	Alpha
0.9	0.9005	524	524	1048	0.54	0.44	1.494	0.05

* Power was computed using the normal approximation method.

This report shows the same sample size when the effect size is specified using the odds ratio.

Example 3 – Comparing the Power of Several Test Statistics

Researchers want to determine which of the eight test statistics to adopt using the comparative reports and charts that **PASS** produces. They want to detect a difference of 0.20 when the response rate of the control group is 0.30. The significance level is 0.05. They want to study sample sizes from 10 to 100.

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**
Max N1 or N2 for Binomial Enumeration........**10000**
Zero Count Adjustment Method ....**Add to zero cells only**
Zero Count Adjustment Value.......................**0.0001**
Alternative Hypothesis ....**Two-Sided**
Test Type...**Z-Test (Pooled)**
Alpha...**0.05**
Group Allocation ....**Equal (N1 = N2)**
Sample Size Per Group ....**10 to 100 by 10**
Input Type..**Differences**
D1 (Difference|H1 = P1–P2) **0.2**
P2 (Group 2 Proportion)................................**0.3**

Reports Tab

Show Comparative Reports ....**Checked**
Calculate Exact Test Results ....**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 for Two Proportions

Groups: 1 = Treatment, 2 = Control
Hypotheses: H0: P1 - P2 = 0 vs. H1: P1 - P2 ≠ 0

Sample Size			Proportions			Power							
N1	N2	N	P1	P2	Target Alpha	Exact Test	Z(Pooled) Test	Z(UnPooled) Test	Z(Pooled) CC Test	Z(UnPooled) CC Test	Mantel Haenszel	Likelihood Ratio	T Test
10	10	20	0.5	0.3	0.05	0.0547	0.1275	0.2215	0.0547	0.1215	0.1275	0.1629	0.1275
20	20	40	0.5	0.3	0.05	0.1632	0.2452	0.3167	0.1419	0.2067	0.2452	0.2452	0.2452
30	30	60	0.5	0.3	0.05	0.2594	0.3511	0.3604	0.2594	0.2708	0.3511	0.3604	0.3511
40	40	80	0.5	0.3	0.05	0.3683	0.4581	0.4612	0.3683	0.3728	0.4581	0.4612	0.4581
50	50	100	0.5	0.3	0.05	0.4635	0.5455	0.5481	0.4635	0.4671	0.5455	0.5455	0.5455
60	60	120	0.5	0.3	0.05	0.5424	0.6177	0.6214	0.5424	0.5501	0.6157	0.6177	0.6157
70	70	140	0.5	0.3	0.05	0.6138	0.6771	0.6815	0.6101	0.6195	0.6771	0.6771	0.6771
80	80	160	0.5	0.3	0.05	0.6773	0.7310	0.7435	0.6773	0.6917	0.7310	0.7368	0.7310
90	90	180	0.5	0.3	0.05	0.7485	0.7930	0.8036	0.7485	0.7589	0.7882	0.7969	0.7930
100	100	200	0.5	0.3	0.05	0.7924	0.8320	0.8328	0.7924	0.7942	0.8316	0.8320	0.8316

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

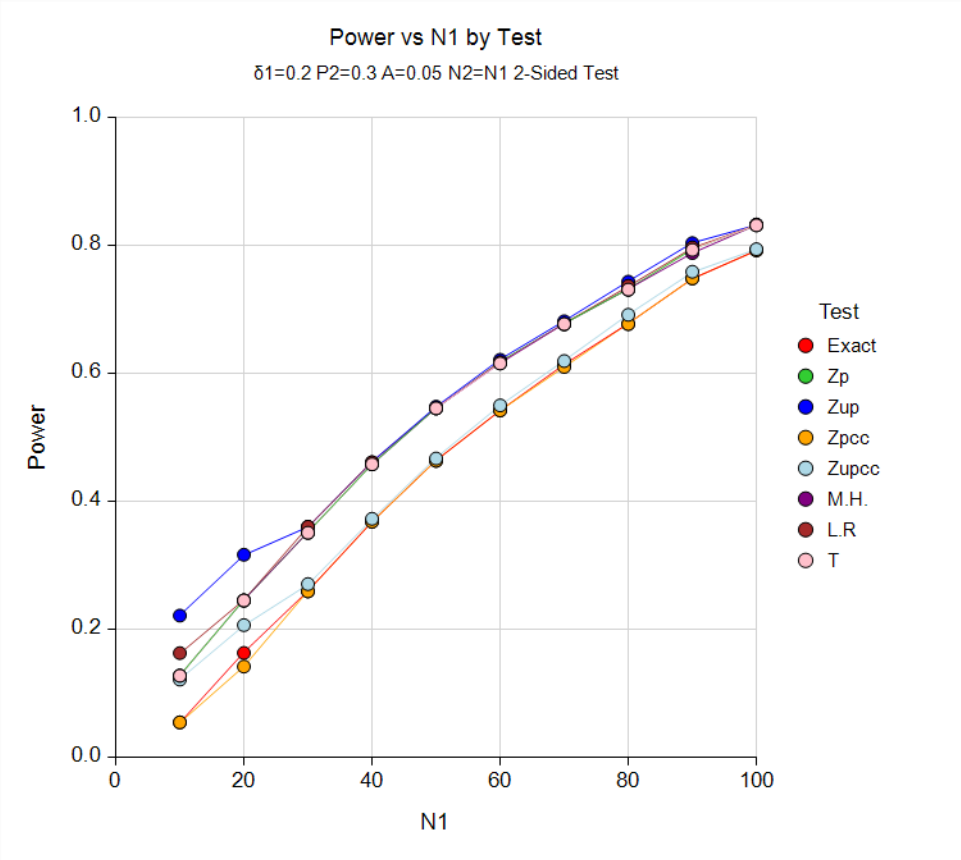
Alpha Comparison of Eight Different Tests for Two Proportions

Groups: 1 = Treatment, 2 = Control
Hypotheses: H0: P1 - P2 = 0 vs. H1: P1 - P2 ≠ 0

Sample Size			Proportions		Alpha								
N1	N2	N	P1	P2	Target	Exact Test	Z(Pooled) Test	Z(UnPooled) Test	Z(Pooled) CC Test	Z(UnPooled) CC Test	Mantel Haenszel	Likelihood Ratio	T Test
10	10	20	0.5	0.3	0.05	0.0119	0.0371	0.0949	0.0119	0.0258	0.0371	0.0771	0.0371
20	20	40	0.5	0.3	0.05	0.0248	0.0533	0.0686	0.0214	0.0267	0.0533	0.0534	0.0533
30	30	60	0.5	0.3	0.05	0.0261	0.0487	0.0583	0.0261	0.0321	0.0487	0.0583	0.0487
40	40	80	0.5	0.3	0.05	0.0282	0.0484	0.0541	0.0276	0.0317	0.0484	0.0541	0.0484
50	50	100	0.5	0.3	0.05	0.0307	0.0498	0.0554	0.0307	0.0334	0.0498	0.0498	0.0498
60	60	120	0.5	0.3	0.05	0.0308	0.0525	0.0552	0.0308	0.0353	0.0483	0.0525	0.0491
70	70	140	0.5	0.3	0.05	0.0330	0.0516	0.0549	0.0318	0.0348	0.0516	0.0516	0.0516
80	80	160	0.5	0.3	0.05	0.0331	0.0513	0.0518	0.0331	0.0350	0.0493	0.0516	0.0493
90	90	180	0.5	0.3	0.05	0.0344	0.0497	0.0525	0.0344	0.0365	0.0497	0.0500	0.0497
100	100	200	0.5	0.3	0.05	0.0348	0.0510	0.0529	0.0348	0.0373	0.0494	0.0517	0.0494

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

Plots



It is interesting to note that the power of Fisher's Exact Test and the *z*-test with continuity correction are consistently lower than the other tests. This occurs because the actual alpha achieved by these tests is much lower than that of the other tests. An interesting finding of this short study was that the regular *t*-test performed better than the more popular *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**
Alternative Hypothesis	**Two-Sided**
Test Type	**Z-Test (Pooled)**
Alpha	**0.05**
Group Allocation	**Equal (N1 = N2)**
Sample Size Per Group	**10 to 100 by 10**
Input Type	**Differences**
D1 (Difference|H1 = P1–P2)	**0.2**
P2 (Group 2 Proportion)	**0.3**

Reports Tab

Show Power Detail Report	**Checked**

Output

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

Power Detail Report for Testing Two Proportions

Groups: 1 = Treatment, 2 = Control
Test Type: Z-Test with Pooled Variance
Hypotheses: H0: P1 - P2 = 0 vs. H1: P1 - P2 ≠ 0

Sample Size			Proportions		Difference	Normal Approximation		Binomial Enumeration	
N1	N2	N	P1	P2	δ1	Power	Alpha	Power	Alpha
10	10	20	0.5	0.3	0.2	0.14407	0.05	0.12752	0.0371
20	20	40	0.5	0.3	0.2	0.24764	0.05	0.24517	0.0533
30	30	60	0.5	0.3	0.2	0.34954	0.05	0.35106	0.0487
40	40	80	0.5	0.3	0.2	0.44553	0.05	0.45805	0.0484
50	50	100	0.5	0.3	0.2	0.53311	0.05	0.54554	0.0498
60	60	120	0.5	0.3	0.2	0.61105	0.05	0.61769	0.0525
70	70	140	0.5	0.3	0.2	0.67906	0.05	0.67713	0.0516
80	80	160	0.5	0.3	0.2	0.73742	0.05	0.73103	0.0513
90	90	180	0.5	0.3	0.2	0.78681	0.05	0.79302	0.0497
100	100	200	0.5	0.3	0.2	0.82811	0.05	0.83201	0.0510

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

Example 5 – Determining the Power after Completing an Experiment

A study has just been completed aimed at determining the effectiveness of a new treatment for cancer. Because of the cost of administering the new treatment, they would adopt the new treatment only if the difference between the proportion cured by the new treatment and that cured by the standard treatment is at least 0.10. The researchers enrolled 200 cancer patients in the study (100 for each treatment) and found that 51% were cured by the standard treatment, while 62% were cured by the new treatment. These results, however, showed no statistically significant difference based on the pooled z-test with continuity correction and alpha = 0.05. Therefore, the researchers want to compute the power of this test for detecting a difference of 0.10 for standard treatment proportions ranging from 0.40 to 0.60.

Note that the power was not exclusively computed at the observed sample proportion for the standard treatment group, 0.51. It is more informative to compute the power for a range of likely values suggested by historical evidence.

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	**Normal Approximation**
Alternative Hypothesis	**Two-Sided**
Test Type	**Z-Test C.C. (Pooled)**
Alpha	**0.05**
Group Allocation	**Equal (N1 = N2)**
Sample Size Per Group	**100**
Input Type	**Differences**
D1 (Difference|H1 = P1–P2)	**0.10**
P2 (Group 2 Proportion)	**0.40 to 0.60 by 0.04**

Output

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

Numeric Results

Numeric Results

Solve For: Power
Groups: 1 = Treatment, 2 = Control
Test Type: Continuity Corrected Z-Test with Pooled Variance
Hypotheses: H0: P1 - P2 = 0 vs. H1: P1 - P2 ≠ 0

	Sample Size			Proportions		Difference	
Power*	N1	N2	N	P1	P2	δ1	Alpha
0.24712	100	100	200	0.50	0.40	0.1	0.05
0.24518	100	100	200	0.54	0.44	0.1	0.05
0.24582	100	100	200	0.58	0.48	0.1	0.05
0.24909	100	100	200	0.62	0.52	0.1	0.05
0.25523	100	100	200	0.66	0.56	0.1	0.05
0.26477	100	100	200	0.70	0.60	0.1	0.05

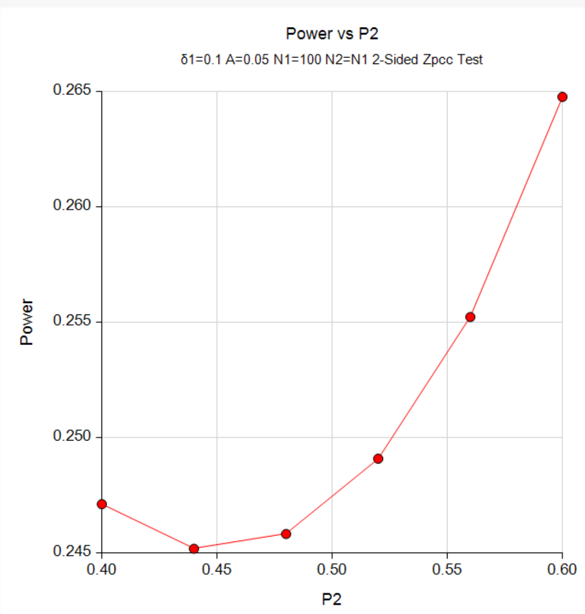
* Power was computed using the normal approximation method.

This report shows the values of each of the parameters, one scenario per row. The power over the entire range of the likely standard treatment proportions is relatively constant at about 0.25 to 0.26.

The values from this table are displayed in the plot below.

Plots Section

Plots



It is evident from these results that the test performed by the researchers had very low power to detect a difference of 0.10 with the sample size used. The power is only about 0.25 or 0.26 for a large range of standard treatment proportions.

Example 6 – Finding the Sample Size using Ratios

Researchers would like to design an experiment to compare the infection rate of a rare disease among two populations. More specifically, they would like to determine how many subjects they need to sample from each population to determine if the disease rate in population 1 is at least three times that of population 2 with 80% power. Suppose that the researchers are confident from previous studies that the infection rate in population 2 is 0.025. The researchers plan to use the likelihood ratio test and alpha = 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 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	**Two-Sided**
Test Type	**Likelihood Ratio Test**
Power	**0.80**
Alpha	**0.05**
Group Allocation	**Equal (N1 = N2)**
Input Type	**Ratios**
R1 (Ratio|H1 = P1/P2)	**3**
P2 (Group 2 Proportion)	**0.025**

Output

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

Numeric Results

Solve For: Sample Size
Groups: 1 = Treatment, 2 = Control
Test Type: Likelihood Ratio Test
Hypotheses: H0: P1 / P2 = 1 vs. H1: P1 / P2 ≠ 1

Power		Sample Size			Proportions		Ratio	
Target	Actual*	N1	N2	N	P1	P2	R1	Alpha
0.8	0.80122	298	298	596	0.075	0.025	3	0.05

* Power was computed using the normal approximation method.

The researchers must sample 298 individuals from each population to achieve 80% power to detect a ratio of 3.0.

Example 7 – Validation of Sample Size Calculation for the Pooled Z-Test using Ryan (2013)

Ryan (2013), page 117, presents an example using the Z-Test with pooled variance in which P2 = 0.55, D1 = 0.1, and alpha = 0.05. Assuming a one-sided test and equal sample allocation, Ryan (2013) finds the necessary sample sizes to be 296 in each group to detect the difference of 0.1 with 80% power.

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** 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**
Test Type...**Z-Test (Pooled)**
Power..**0.80**
Alpha...**0.05**
Group Allocation ....**Equal (N1 = N2)**
Input Type..**Differences**
D1 (Difference|H1 = P1–P2) **0.10**
P2 (Group 2 Proportion)................................**0.55**

Output

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

Numeric Results

Solve For: Sample Size
Groups: 1 = Treatment, 2 = Control
Test Type: Z-Test with Pooled Variance
Hypotheses: H0: P1 - P2 ≤ 0 vs. H1: P1 - P2 > 0

Power		Sample Size			Proportions		Difference	
Target	Actual*	N1	N2	N	P1	P2	δ1	Alpha
0.8	0.80034	296	296	592	0.65	0.55	0.1	0.05

* Power was computed using the normal approximation method.

PASS also found the required sample size to be 296 in each group.

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 a two-sided unpooled Z-test. When P2 = 0.65, D1 = 0.2, power = 0.8, and alpha = 0.05, Chow, Shao, and Wang (2008) reports a required sample size of 70.

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 .... **Two-Sided**
Test Type.. **Z-Test (Unpooled)**
Power.. **0.80**
Alpha... **0.05**
Group Allocation .... **Equal (N1 = N2)**
Input Type... **Differences**
D1 (Difference|H1 = P1–P2) **0.20**
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
Groups: 1 = Treatment, 2 = Control
Test Type: Z-Test with Unpooled Variance
Hypotheses: H0: P1 - P2 = 0 vs. H1: P1 - P2 ≠ 0

Power		Sample Size			Proportions		Difference	
Target	Actual*	N1	N2	N	P1	P2	δ1	Alpha
0.8	0.80191	70	70	140	0.85	0.65	0.2	0.05

* Power was computed using the normal approximation method.

PASS also found the required sample size to be 70 in each group.

Example 9 – Validation of Sample Size Calculation for the Continuity Corrected Z-Test using Pooled Variances with Equal Sample Sizes using Fleiss, Levin, and Paik (2003)

Fleiss, Levin, and Paik (2003), page 74, presents a sample size study in which P1 = 0.7, P2 = 0.6, alpha = 0.01, and beta = 0.05 or 0.25. Assuming two-sided testing and equal sample allocation, Fleiss finds the necessary sample sizes to be 827 in each group for 95% power and 499 in each group for 75% power. The calculations of Fleiss, Levin, and Paik (2003) included an adjustment for continuity correction. This continuity correction is not necessary here when exact calculations are made. However, when the sample size is large enough so that approximate calculations are used, the continuity correction must be applied to obtain the same results. This is done by setting the Test Statistic to "Z Test C.C.". Note that this adjustment is used here to keep our results consistent with those of Fleiss, Levin, and Paik (2003). In practice, this adjustment is not recommended because it reduces the power and the actual alpha of the test procedure.

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 9** 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	**Two-Sided**
Test Type	**Z-Test C.C. (Pooled)**
Power	**0.95 0.75**
Alpha	**0.01**
Group Allocation	**Equal (N1 = N2)**
Input Type	**Proportions**
P1 (Group 1 Proportion|H1)	**0.70**
P2 (Group 2 Proportion)	**0.60**

Output

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

Numeric Results

Solve For: Sample Size
Groups: 1 = Treatment, 2 = Control
Test Type: Continuity Corrected Z-Test with Pooled Variance
Hypotheses: H0: P1 - P2 = 0 vs. H1: P1 - P2 ≠ 0

Power		Sample Size			Proportions		Difference	
Target	Actual*	N1	N2	N	P1	P2	δ1	Alpha
0.75	0.75066	500	500	1000	0.7	0.6	0.1	0.01
0.95	0.95001	827	827	1654	0.7	0.6	0.1	0.01

* Power was computed using the normal approximation method.

PASS found the required sample sizes to be 500 and 827 which correspond to Fleiss, Levin, and Paik (2003) with a slight difference due to rounding.

Example 10 – Validation of Sample Size Calculation for the Continuity Corrected Z-Test using Pooled Variances with Unequal Sample Sizes using Fleiss, Levin, and Paik (2003)

Fleiss, Levin, and Paik (2003), pages 76-77, presents a sample size study in which P1 = 0.25, P2 = 0.40, alpha = 0.01, and beta = 0.05. Assuming two-sided testing with half as many in the second group as the first, Fleiss, Levin, and Paik (2003) finds the sample sizes to be 530 in the first group and 265 in the second.

Note that half as many in the second group is achieved by setting *R* to 0.5.

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 10** 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 ....**Two-Sided**
Test Type..**Z-Test C.C. (Pooled)**
Power...**0.95**
Alpha..**0.01**
Group Allocation ....**Enter R = N2/N1, solve for N1 and N2**
R (Sample Allocation Ratio)..........................**0.5**
Input Type..**Proportions**
P1 (Group 1 Proportion|H1)..........................**0.25**

Output

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

Numeric Results

Solve For: Sample Size
Groups: 1 = Treatment, 2 = Control
Test Type: Continuity Corrected Z-Test with Pooled Variance
Hypotheses: H0: P1 - P2 = 0 vs. H1: P1 - P2 ≠ 0

Power		Sample Size			R (N2 / N1)		Proportions		Difference	
Target	Actual*	N1	N2	N	Target	Actual	P1	P2	δ1	Alpha
0.95	0.95066	531	266	797	0.5	0.5	0.25	0.4	-0.15	0.01

* Power was computed using the normal approximation method.

PASS found the required sample sizes to be 531 and 266 which nearly corresponds to the results in Fleiss, Levin, and Paik (2003). Fleiss, Levin, and Paik (2003) computed 530 instead of 531. The number 531 is correct because the power for 530 is slightly less than the required 0.95.