Chapter 110

Equivalence Tests for One Proportion

Introduction

This module provides power analysis and sample size calculation for equivalence tests in one-sample designs in which the outcome is binary. The equivalence test is usually carried out using the Two One-Sided Tests (TOST) method. This procedure computes power and sample size for the TOST equivalence test method. Users may choose from among commonly used test statistics.

Approximate sample size formulas for equivalence tests of a single proportion are presented in Chow et al. (2008) page 86. However, only large sample (normal approximation) results are given there. Some sample size programs use only the normal approximation to the binomial distribution for power and sample size estimates. The normal approximation is accurate for large sample sizes and for proportions between 0.2 and 0.8, roughly. When the sample sizes are small, or the proportions are extreme (i.e., less than 0.2 or greater than 0.8) the binomial calculations are much more accurate.

Example

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

It is known that the new treatment will not have a response rate that is exactly the same as that of the standard treatment. After careful consideration, they decide that the margin of equivalence is plus or minus 10%. That is, if the response rate of the new treatment is between 60% and 80% it will be deemed equivalent to the standard treatment.

The developers must design an experiment to test the hypothesis that the response rate of the new treatment is within 10% of the standard (baseline) treatment. The statistical hypotheses to be tested are

$$H_0: |P - PB| \geq 0.1 \quad \text{versus} \quad H_1: |P - PB| < 0.1$$

Notice that when the null hypothesis is rejected the conclusion is that the response rate is between 0.6 and 0.8.

Binomial Model

A binomial variable should exhibit the following four properties:

1. The variable is binary --- it can take on one of two possible values.
2. The variable is observed a known number of times. Each observation or replication is called a Bernoulli trial. The number of replications is n. The number of times that the outcome of interest is observed is r. Thus, r takes on the possible values 0, 1, 2, ..., n.
3. The probability, P, that the outcome of interest occurs is constant for each trial.
4. The trials are independent. The outcome of one trial does not influence the outcome of the any other trial.

A binomial probability is calculated using the formula

$$b(r;n,P)=\binom{n}{r}P^{r}(1-P)^{n-r}$$

where

$$\binom{n}{r}=\frac{n!}{r!\,(n-r)!}$$

Hypothesis Testing

Parameterizations of the Proportions

In the discussion that follows, let P represent the proportion being investigated. That is, P is the actual probability of a *success* in a binomial experiment. Often, this proportion is a response rate, cure rate, or survival rate. Let PB represent the *baseline* proportion. In an equivalence trial, the baseline proportion is the response rate of the current (standard) treatment. Let $P0L$ represent the smallest value of P that still results in the conclusion that the new treatment is equivalent to the current treatment. Similarly, let $P0U$ represent the largest value of P that still results in the conclusion that the new treatment is equivalent to the current treatment. Note that PB will be between $P0L$ and $P0U$. The power of a test is computed at a specific value of the proportion, $P1$.

The statistical hypotheses that are tested are

$$H0{:}\,P\le P0L \text{or } H0{:}\,P\ge P0U \quad \text{versus} \quad H1{:}\,P0L<P<P0U$$

This unusual hypothesis test can be broken down into two, one-sided hypothesis tests (TOST) as follows

$$H0{:}\,P\le P0L \quad \text{versus} \quad H1{:}\,P>P0L$$

and

$$H0{:}\,P\ge P0U \quad \text{versus} \quad H1{:}\,P<P0U$$

If both of these one-sided tests are rejected at significance level α, then equivalence can be concluded at significance level α. Note that we do not conduct the individual tests at α/2.

There are three common methods of specifying the margin of equivalence. The most direct is to simply assign values for *P0L* and *P0U*. However, it is often more meaningful to identify *PB* and then specify *P0L* and *P0U* implicitly by giving a difference, ratio, or odds ratio. Mathematically, the definitions of these parameterizations are

Parameter	**Computation**	**Hypotheses**
Difference	$d0 = \|P0L - PB\| = P0U - PB$	$H0{:}\, d \geq d0 \quad \text{vs} \quad H1{:}\, d < d0$
Ratio	$r0 = \frac{1}{P0L/PB} = P0U/PB$	$H0{:}\, r \geq r0 \quad \text{vs} \quad H1{:}\, r < r0$
Odds Ratio	$o0 = \frac{1}{OddsL/OddsB} = \frac{OddsU}{OddsB}$	$H0{:}\, o \geq o0 \quad \text{vs} \quad H1{:}\, o < o0$

where

Difference $d = |P - PB|$

Ratio $r = \begin{cases} P/PB & \text{if } P > PB \\ PB/P & \text{if } P < PB \end{cases}$

Odds Ratio $o = \begin{cases} Odds/OddsB & \text{if } P > PB \\ OddsB/Odds & \text{if } P < PB \end{cases}$

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 use.

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. That is, there is usually little interest in a treatment that only decreases the probability from 0.400 to 0.399. 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. As a rule of thumb, the difference is best suited for those cases in which.

Equivalence Test using a Difference

The following example might be instructive. Suppose 60% of patients respond to the current treatment method (PB = 0.60). If the response rate of the new treatment is no less than five percentage points different (d0 = 0.05) from the existing treatment, it will be considered to be equivalent. Substituting these figures into the statistical hypotheses gives

$$H0{:}\, d \geq 0.05 \quad \text{versus} \quad H1{:}\, d < 0.05$$

where d = |P – PB|.

The resulting joint hypotheses are

$$H0: P \leq 0.55 \quad \text{versus} \quad H1: P > 0.55.$$

and

$$H0: P \geq 0.65 \quad \text{versus} \quad H1: P < 0.65.$$

In this example, when both null hypotheses are rejected, the concluded alternative is that the response rate is between 55% and 65%.

Ratio

The ratio r0 = PE / PB denotes the relative change in the probability of the response. Testing equivalence uses the hypotheses

$$H0: r \leq r0 \quad \text{versus} \quad H1: r > r0$$

where r = P / PB if P > PB or r = PB / P if P < PB.

Equivalence Test using a Ratio

The following example might help to understand the concept of *equivalence* as defined by the ratio. Suppose that 60% of patients (PB = 0.60) respond to the current treatment method. If a new treatment changes the response rate by no more than 10% (r0 = 1.1), it will be considered to be equivalent to the standard treatment. Substituting these figures into the statistical hypotheses gives

$$H0: r \geq 1.1 \quad \text{versus} \quad H1: r < 1.1$$

The relationship P0 = (r0)(PB) gives the two, one-sided, hypotheses

$$H0: P \leq 0.54 \quad \text{versus} \quad H1: P > 0.54$$

$$H0: P \geq 0.66 \quad \text{versus} \quad H1: P < 0.66$$

In this example, when the null hypothesis is rejected, the concluded alternative is that the response rate is between 54% and 66%.

Odds Ratio

The odds ratio, o0 = (P0 / (1 – P0)) / (PB / (1 – PB)), gives the relative change in the odds of the response. Testing equivalence use the same formulation, namely

$$H0: o \leq o0 \quad \text{versus} \quad H1: o > o0$$

where o = Odds / OddsB if P > PB or o = OddsB / Odds if P < PB.

Test Statistics

Many different test statistics have been proposed for equivalence tests of a single proportion. Most of these were proposed before computers or hand calculators were widely available. Although these legacy methods are still presented in textbooks, their power and accuracy should be compared against modern exact methods before they are adopted for serious research. To make this comparison easy, the power and significance of several tests of a single proportion are available in this procedure.

Exact Test

The test statistic is *r*, the number of successes in *n* trials. This test should be the standard against which other test statistics are judged. The significance level and power are computed by enumerating the possible values of *r*, computing the probability of each value, and then computing the corresponding value of the test statistic. Hence the values that are reported in the output for these tests are exact, not approximate.

Z-Tests

Several *z* statistics have been proposed that use the central limit theorem. This theorem states that for large sample sizes, the distribution of the *z* statistic is approximately normal. All of these tests take the following form:

$$z = \frac{p - P0L}{s} \quad \text{and} \quad z = \frac{p - P0U}{s}$$

Although these *z* tests were developed because the distribution of *z* is approximately normal in large samples, the actual significance level and power can be computed exactly using the binomial distribution.

We include four *z* tests which are based on two methods for computing *s* and whether a continuity correction is applied.

Z-Test using S(P0)

This test statistic uses the value of P0 to compute *s*.

$$z_1 = \frac{p - P0L}{\sqrt{\frac{(P0L)\left(1 - (P0L)\right)}{n}}} \quad \text{and} \quad z_1 = \frac{p - P0U}{\sqrt{\frac{(P0U)\left(1 - (P0U)\right)}{n}}}$$

Z-Test using S(P0) with Continuity Correction

This test statistic is similar to the one above except that a continuity correction is applied to make the normal distribution more closely approximate the binomial distribution.

$$z_2 = \frac{(p - P0L) + c}{\sqrt{\frac{(P0L)\left(1 - (P0L)\right)}{n}}} \quad \text{and} \quad z_2 = \frac{(p - P0U) + c}{\sqrt{\frac{(P0U)\left(1 - (P0U)\right)}{n}}}$$

where

$$c = \begin{cases} \frac{-1}{2n} & \text{if } p > P0 \\ \frac{1}{2n} & \text{if } p < P0 \\ 0 & \text{if } |p - P0| < \frac{1}{2n} \end{cases}$$

Z-Test using S(Phat)

This test statistic uses the value of *p* to compute *s*.

$$z_3 = \frac{p - P0L}{\sqrt{\frac{p(1-p)}{n}}} \quad \text{and} \quad z_3 = \frac{p - P0U}{\sqrt{\frac{p(1-p)}{n}}}$$

Z-Test using S(Phat) with Continuity Correction

This test statistic is similar to the one above except that a continuity correction is applied to make the normal distribution more closely approximate the binomial distribution.

$$z_4 = \frac{(p - P0L) + c}{\sqrt{\frac{p(1-p)}{n}}} \quad \text{and} \quad z_4 = \frac{(p - P0U) + c}{\sqrt{\frac{p(1-p)}{n}}}$$

where

$$c = \begin{cases} \frac{-1}{2n} & \text{if } p > P0 \\ \frac{1}{2n} & \text{if } p < P0 \\ 0 & \text{if } |p - P0| < \frac{1}{2n} \end{cases}$$

Power Calculation

Normal Approximation Method

Power may be calculated for one-sample proportions equivalence tests using the normal approximation to the binomial distribution. This section provides the power calculation formulas for the various test statistics available in this procedure. In the equations that follow, $\Phi()$ represents the standard normal cumulative distribution function, and z_α represents the value that leaves α in the upper tail of the standard normal distribution. All power values are evaluated at $P = P1$.

Exact Test

Power for the equivalence test is calculated as

$$Power_{\mathrm{ET}} = \Phi\left(\frac{\sqrt{n}(P0U - P1) - z_\alpha\sqrt{P0U(1-P0U)}}{\sqrt{P1(1-P1)}}\right) - \Phi\left(\frac{\sqrt{n}(P0L - P1) + z_\alpha\sqrt{P0L(1-P0L)}}{\sqrt{P1(1-P1)}}\right)$$

Z Test using S(P0)

Power for the equivalence test is calculated as

$$Power_{ZS(P0)} = \Phi\left(\frac{\sqrt{n}(P0U - P1) - z_\alpha\sqrt{P0U(1-P0U)}}{\sqrt{P1(1-P1)}}\right) - \Phi\left(\frac{\sqrt{n}(P0L - P1) + z_\alpha\sqrt{P0L(1-P0L)}}{\sqrt{P1(1-P1)}}\right)$$

Z Test using S(P0) with Continuity Correction

Power for the equivalence test is calculated as

$$\begin{aligned} Power_{ZS(P0)CC} = {} & \Phi\left(\frac{\sqrt{n}(P0U - P1) - z_\alpha\sqrt{P0U(1-P0U)} - c_2}{\sqrt{P1(1-P1)}}\right) \\ & - \Phi\left(\frac{\sqrt{n}(P0L - P1) + z_\alpha\sqrt{P0L(1-P0L)} + c_1}{\sqrt{P1(1-P1)}}\right) \end{aligned}$$

where $c_1 = 1/2\sqrt{n}$ if $|P1 - P0L| < 1/2n$ otherwise $c_1 = 0$ and $c_2 = 1/2\sqrt{n}$ if $|P1 - P0U| < 1/2n$ otherwise $c_2 = 0$.

Z Test using S(Phat)

Power for the equivalence test is calculated as

$$Power_{ZS(P1)} = \Phi\left(\frac{\sqrt{n}(P0U - P1) - z_\alpha\sqrt{P1(1-P1)}}{\sqrt{P1(1-P1)}}\right) - \Phi\left(\frac{\sqrt{n}(P0L - P1) + z_\alpha\sqrt{P1(1-P1)}}{\sqrt{P1(1-P1)}}\right)$$

Z Test using S(Phat) with Continuity Correction

Power for the equivalence test is calculated as

$$Power_{ZS(P1)CC} = \Phi\left(\frac{\sqrt{n}(P0U - P1) - z_\alpha\sqrt{P1(1-P1)} - c_2}{\sqrt{P1(1-P1)}}\right) - \Phi\left(\frac{\sqrt{n}(P0L - P1) + z_\alpha\sqrt{P1(1-P1)} + c_1}{\sqrt{P1(1-P1)}}\right)$$

where $c_1 = 1/2\sqrt{n}$ if $|P1 - P0L| < 1/2n$ otherwise $c_1 = 0$ and $c_2 = 1/2\sqrt{n}$ if $|P1 - P0U| < 1/2n$ otherwise $c_2 = 0$.

Steps to Calculate Power using Binomial Enumeration of All Possible Outcomes

Historically, power and sample size calculations for a one-sample proportion test have been based on normal approximations to the binomial. However, with the speed of modern computers using the normal approximation is unnecessary, especially for small samples. Rather, the significance level and power can be computed using complete enumeration of all possible values of *x*, the number of successes in a sample of size *n*.

This is done as follows.

1. The critical value of the test is computed using standard techniques.
2. For each possible value of *x*, the values of the two one-sided test statistics (*z*-test or exact test) are computed along with their associated probability of occurrence.
3. The significance level and power are computed by summing the probabilities of occurrence for all values of the test statistics that are greater than (or less than) the critical values. Each probability of occurrence is calculated using *P0L* and *P0U* for the significance level and *P1* for the power.

Other variables such as the sample size are then found using an efficient search algorithm. Although this method is not as elegant as a closed-form solution, it is completely accurate.

Examples of Power Calculation for the Exact Test using Binomial Enumeration

Suppose the baseline proportion, PB, is 0.50, the sample size is 10, and the target alpha level is 0.05. A typical value for the equivalence difference is 0.05. However, because the example is for a small sample size, the equivalence difference will be set to 0.4 (which is, of course, a very unrealistic figure) for illustrative purposes. Calculate the power of this design to detect equivalence if the actual difference between the proportions is 0.10.

The first step is to find the rejection region under the null hypothesis. In this example, the null hypothesis is $H0: P \leq 0.1$ or $H0: P > 0.9$and the alternative hypothesis is $H1: 0.1 < P < 0.9$ $H_1: 0.1 < P < 0.9$. This composite hypothesis breaks down into the following two, one-sided, simple hypotheses

1. $H0: P \leq 0.1$ versus $H1: P > 0.1$
2. $H0: P \geq 0.9$ versus $H1: P < 0.9$

The rejection regions for both tests are determined from the following table of cumulative binomial probabilities for *N* = 10. The first column of probabilities is for *r* greater than or equal to *R* while the second two columns of probabilities are for *r* less than or equal to *R*.

Table of Binomial Probabilities for N = 10 and P = 0.1, 0.9, and 0.6

R	Pr(r ≥ R | P=0.1)	Reject Test1	Pr(r ≤ R |P=0.9)	Reject Test2	Reject Both	Pr(r ≤ R|P=0.6)
0	1.0000	No	0.0000	Yes	No	0.0001
1	0.6513	No	0.0000	Yes	No	0.0017
2	0.2639	No	0.0000	Yes	No	0.0123
3	0.0702	No	0.0000	Yes	No	**0.0548**
4	**0.0128**	Yes	0.0001	Yes	Yes	0.1662
5	0.0016	Yes	0.0016	Yes	Yes	0.3669
6	0.0001	Yes	**0.0128**	Yes	Yes	**0.6177**
7	0.0000	Yes	0.0702	No	No	0.8327
8	0.0000	Yes	0.2639	No	No	0.9536
9	0.0000	Yes	0.6513	No	No	0.9940
10	0.0000	Yes	1.0000	No	No	1.0000

The second column gives the value of alpha for the first test ($H0: P \leq 0.1$ versus $H1: P > 0.1$). The rejection region for this test is all values of *R* greater than or equal to 4. The fourth column gives the values of alpha for the second test. The rejection region for the second test is all values of *R* less than or equal to 6. The rejection region for both tests is those values of *R* values that result in rejection of both individual tests. These are the *R* values 4, 5, and 6. The power is computed using the final column of the table which gives cumulative binomial probabilities for *P* = 0.5 + 0.1 = 0.6. The power is probability for the cases 4, 5, and 6. It is calculated as 0.6177 – 0.0548 = 0.5629.
It is informative to consider what happens when the equivalence difference is reduced from 0.4 to 0.2. The following table gives the appropriate cumulative binomial probabilities for this case.

Table of Binomial Probabilities for N = 10 and P = 0.3, 0.7, and 0.6

R	Pr(r ≥ R |P=0.3)	Reject Test1	Pr(r ≤ R |P=0.7)	Reject Test2	Reject Both	Pr(r ≤ R|P=0.6)
0	1.0000	No	0.0000	Yes	No	0.0001
1	0.9718	No	0.0001	Yes	No	0.0017
2	0.8507	No	0.0016	Yes	No	0.0123
3	0.6172	No	0.0106	Yes	No	0.0548
4	0.3504	No	**0.0473**	Yes	No	0.1662
5	0.1503	No	0.1503	No	No	0.3669
6	**0.0473**	Yes	0.3504	No	No	0.6177
7	0.0106	Yes	0.6172	No	No	0.8327
8	0.0016	Yes	0.8507	No	No	0.9536
9	0.0001	Yes	0.9718	No	No	0.9940
10	0.0000	Yes	1.0000	No	No	1.0000

The second column gives the value of alpha for the first test. The rejection region for this test is all values of *R* greater than or equal to 6. The fourth column gives the values of alpha for the second test. The rejection region for the second test is all values of *R* less than or equal to 4. The rejection region for both tests together is empty! There is no *R* for which both tests will be rejected. Hence, the alpha level and the power will both be 0.0.

Examples of Power Calculation for the Z S(P0) Test using Binomial Enumeration

The following example illustrates how to calculate the power of an approximate *z* test. There are several *z* tests to choose from. We will use the following test.

$$z = \frac{p - P0}{\sqrt{P0(1 - P0)/n}}$$

Calculating the rejection region for the *z* test is based on a table of normal probabilities. For the target alpha level of 0.05, the critical value is 1.6449. That is, the first hypothesis test that $H0: P \leq 0.1$ versus $H1: P > 0.1$ is rejected if the resulting calculated *z* value is greater than 1.6449. Similarly, the second hypothesis test that $H0: P \geq 0.9$ versus $H1: P < 0.9$ is rejected when the calculated *z* value is less than -1.6449. The rejection regions for both tests are shown in the following table of binomial probabilities for *N* = 10.

Table Showing Both One-Sided Z Tests for N = 10 and P = 0.1, 0.9, and 0.6

R	Z for P = 0.1	Reject Test1	Z for P = 0.9	Reject Test2	Reject Both	Pr(r ≤ R|P=0.6)
0	-1.0541	No	-9.4868	Yes	No	0.0001
1	0.0000	No	-8.4327	Yes	No	0.0017
2	**1.0541**	No	-7.3786	Yes	No	**0.0123**
3	2.1082	Yes	-6.3246	Yes	Yes	0.0548
4	3.1623	Yes	-5.2705	Yes	Yes	0.1662
5	4.2164	Yes	-4.2164	Yes	Yes	0.3669
6	5.2705	Yes	-3.1623	Yes	Yes	0.6177
7	6.3246	Yes	-2.1082	Yes	Yes	**0.8327**
8	7.3786	Yes	-1.0541	No	No	0.9536
9	8.4327	Yes	0.0000	No	No	0.9940
10	9.4868	Yes	1.0541	No	No	1.0000

Note that the null hypothesis is rejected for the equivalence test when *R* is 3, 4, 5, 6, and 7. The power is the probability of these values calculated using *P* = 0.60. It is calculated as 0.8327 – 0.0123 = 0.8204. Notice that this is much larger than 0.5629 which was the power for the exact test. The reason for this discrepancy is that the approximate test is actually testing at a larger alpha than the target of 0.05. The actual alpha is the maximum of the two individual alphas. From the first table, we can see that the actual alpha for the first test is Pr(r ≥ 3|P=0.1) = 0.0702. Similarly, the actual alpha for the second test is Pr(r ≤ 7|P=0.9) = 0.0702. Hence the alpha level is 0.0702. The actual alpha of the exact test was 0.0128.

Example 1 – Finding the Power

Suppose 50% of patients with a certain type of cancer survive two years using the current treatment. The current treatment is expensive and has several severe side effects. A new treatment has fewer side effects and is less expensive. An equivalence trial is to be conducted to show that the two-year survival rate of the new treatment is the same as the current treatment. After serious consideration, the margin of equivalence is set at 5%. What power will be achieved by sample sizes of 50, 100, 200, 300, 500, or 800 and a significance level of 0.05? For comparative purposes, also calculate the power for margin of equivalence of 10%. Assume that the true survival rate of the new treatment is the same as that of the current (baseline) treatment.

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**
Test Type.. **Exact Test**
Alpha... **0.05**
n (Sample Size) **50 100 200 300 500 800**
Input Type... **Differences**
PB (Baseline Proportion) **0.50**
d0 (Equivalence Difference)........................... **0.05 0.10**
d1 (Actual Difference) **0**

Output

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

Numeric Reports

Numeric Results

Solve For: Power
Alternative Hypothesis: Equivalence (H0: P ≤ P0L or P ≥ P0U vs. H1: P0L < P < P0U)
Test Type: Exact Test

Power*	Sample Size n	Baseline Proportion PB	Equivalence Proportion Lower P0L	Equivalence Proportion Upper P0U	Difference Equivalence d0	Difference Actual d1	Alpha	Reject H0 if R1 ≤ R ≤ R2 R1|R2
0.00000	50	0.5	0.45	0.55	0.05	0	0.05	29|21
0.00000	50	0.5	0.40	0.60	0.10	0	0.05	27|23
0.00000	100	0.5	0.45	0.55	0.05	0	0.05	54|46
0.30226	100	0.5	0.40	0.60	0.10	0	0.05	49|51
0.00000	200	0.5	0.45	0.55	0.05	0	0.05	103|97
0.77632	200	0.5	0.40	0.60	0.10	0	0.05	92|108
0.07604	300	0.5	0.45	0.55	0.05	0	0.05	150|150
0.93604	300	0.5	0.40	0.60	0.10	0	0.05	135|165
0.45113	500	0.5	0.45	0.55	0.05	0	0.05	244|256
0.99577	500	0.5	0.40	0.60	0.10	0	0.05	219|281
0.76667	800	0.5	0.45	0.55	0.05	0	0.05	384|416
0.99995	800	0.5	0.40	0.60	0.10	0	0.05	344|456

* Power was computed using the normal approximation method.

Power	The probability of concluding equivalence when the proportions are equivalent.
n	The size of the sample drawn from the population.
PB	The baseline or standard value of the proportion. This is the value under the current treatment.
P0L, P0U	The limits between which an equivalent proportion must fall.
d0	The smallest absolute difference that is still considered equivalent. d0 = |P0 - PB|.
d1	The value of the difference at which the power is calculated. d1 = P1 - PB.
Alpha	The probability of rejecting a true null hypothesis.
Reject H0 If	The critical value(s) for the test.

Summary Statements

A single-group design will be used to test the equivalence of a single proportion, with an equivalence difference (d0) of 0.05 (H0: P ≤ 0.45 or P ≥ 0.55 versus H1: 0.45 < P < 0.55). The comparison will be made using two one-sided, one-sample exact tests, with an overall Type I error rate (α) of 0.05. The baseline proportion (PB) is assumed to be 0.5. To detect a difference (d1) of 0 with a sample size of 50, the power is 0.

Dropout-Inflated Sample Size

Dropout Rate	Sample Size n	Dropout-Inflated Enrollment Sample Size n'	Expected Number of Dropouts D
20%	50	63	13
20%	100	125	25
20%	200	250	50
20%	300	375	75
20%	500	625	125
20%	800	1000	200

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.
n The evaluable sample size at which power is computed (as entered by the user). If n subjects are evaluated out of the n' subjects that are enrolled in the study, the design will achieve the stated power.
n' The total number of subjects that should be enrolled in the study in order to obtain n evaluable subjects, based on the assumed dropout rate. n' is calculated by inflating n using the formula n' = n / (1 - DR), with n' 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.)
D The expected number of dropouts. D = n' - n.

Dropout Summary Statements

Anticipating a 20% dropout rate, 63 subjects should be enrolled to obtain a final sample size of 50 subjects.

References

Blackwelder, W.C. 1998. 'Equivalence Trials.' In Encyclopedia of Biostatistics, John Wiley and Sons. New York. Volume 2, 1367-1372.
Chow, S.C. and Liu, J.P. 1999. Design and Analysis of Bioavailability and Bioequivalence Studies. Marcel Dekker. New York.
Chow, S.C., Shao, J., and Wang, H. 2008. Sample Size Calculations in Clinical Research, Second Edition. Chapman & Hall/CRC. Boca Raton, Florida.
Fleiss, J. L., Levin, B., and Paik, M.C. 2003. Statistical Methods for Rates and Proportions. Third Edition. John Wiley & Sons. New York.

This report shows the values of each of the parameters, one scenario per row. Because of the discrete nature of the binomial distribution, the target alpha is usually different than the actual alpha. Hence, the actual alpha is also shown.

Power

Power is the probability of concluding equivalence when the treatment is indeed equivalent.

n

This is the sample size.

Baseline Proportion

The baseline proportion, PB, is the response rate that is achieved by the current (standard) treatment.

Equivalence Difference (or Proportion, Ratio, or Odds Ratio)

The equivalence difference is the maximum difference from the baseline proportion, PB, that is still considered as unimportant or trivial. This value is used to calculate P0.

Lower and Upper Equivalence Proportions

If the true proportion is between these two limits, the treatment is considered to be equivalent to the baseline proportion. These are the bounds of equivalence.

Actual Difference (or Proportion, Ratio, or Odds Ratio)

The actual difference is the difference between the actual proportion, P1, and the baseline proportion, PB.

Alpha

This is the target (set in the design) value of the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected. That is, this is the probability of concluding equivalence when in fact the new treatment is not equivalent. Because of the discreteness of the binomial distribution from which this value is calculated, the target value is seldom actually achieved.

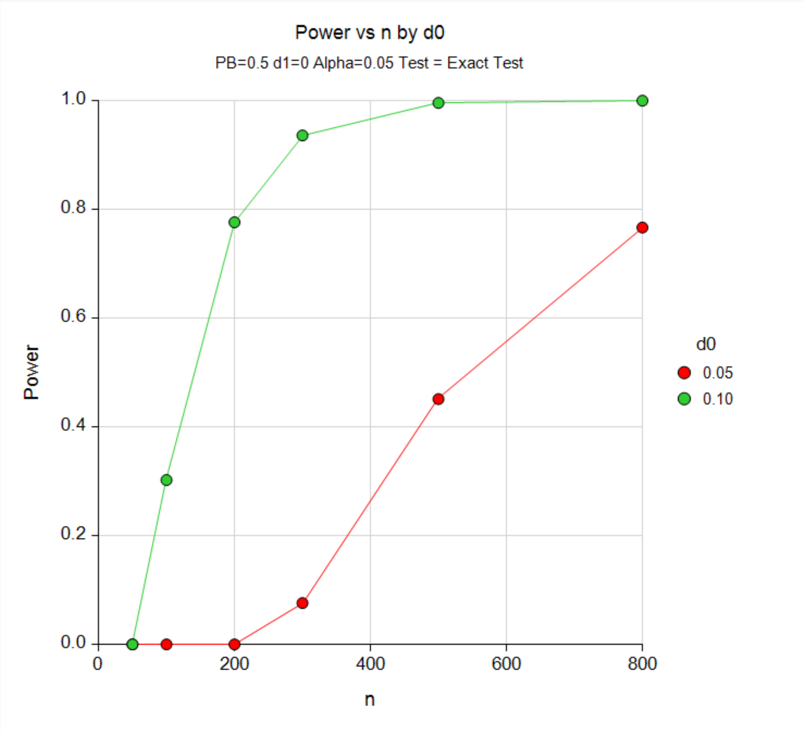
Reject H0 if R1 ≤ R ≤ R2

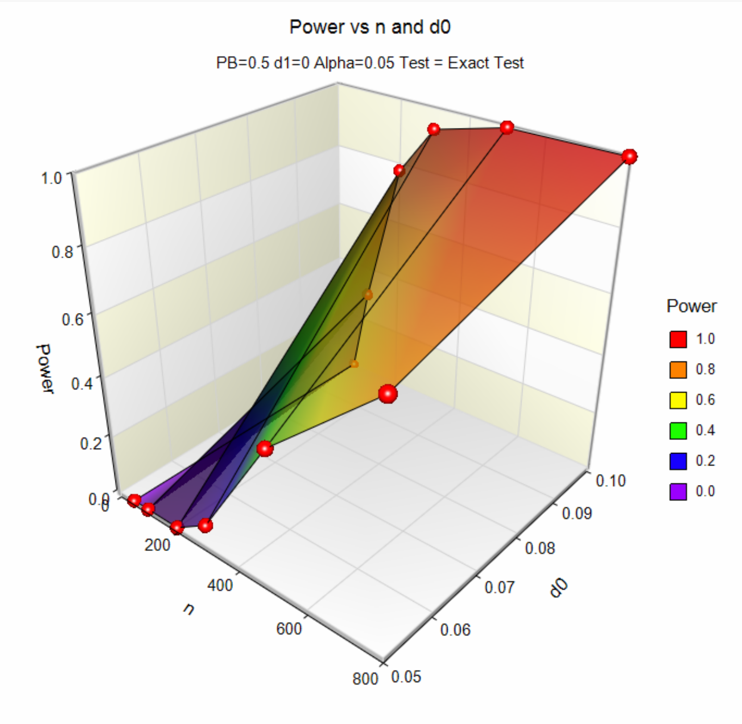
This value provides the bounds between which equivalence is concluded. For example, if *n* is 50, then a value here of 29|31 means that the null hypothesis of non-equivalence is rejected when the number of items with the characteristic of interest is 29, 30, or 31.

When the second number is less than the first as it is in the first line (29|21), the design can never reject the null hypothesis. These designs should never be used.

Plots Section

Plots





These plots show the relationship between power, sample size, and the trivial difference. Note that 80% power is achieved with a sample size of about 210 when the trivial difference is 0.10 and over 800 when the trivial difference is 0.05.

Example 2 – Finding the Sample Size

Continuing from Example 1, suppose you want to find the exact sample size necessary to achieve 90% power when the trivial difference is 0.05. Assume that an exact test will be used.

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**
Test Type.. **Exact Test**
Power... **0.90**
Alpha.. **0.05**
Input Type... **Differences**
PB (Baseline Proportion) **0.50**
d0 (Equivalence Difference)........................... **0.05**
d1 (Actual Difference) **0**

Output

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

Numeric Results

Solve For: Sample Size
Alternative Hypothesis: Equivalence (H0: P ≤ P0L or P ≥ P0U vs. H1: P0L < P < P0U)
Test Type: Exact Test

			Equivalence Proportion		Difference			
Power*	Sample Size n	Baseline Proportion PB	Lower P0L	Upper P0U	Equivalence d0	Actual d1	Alpha	Reject H0 if R1 ≤ R ≤ R2 R1|R2
0.90006	1077	0.5	0.45	0.55	0.05	0	0.05	513|564

* Power was computed using the normal approximation method.

This report shows that a sample size of 1077 will be necessary to achieve the design requirements.

Example 3 – Comparing Test Statistics

Continuing Example 1, suppose the researchers want to investigate which of the five test statistics to use. This is an important question since choosing the wrong test statistic can increase sample size and reduce power. The differences in the characteristics of test statistics are most noticeable in small samples. Hence, the investigation done here is for sample sizes of 20 to 200 in steps of 20. The trivial difference will be set to 0.10. All other settings are as given in Example 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**
Max n for Binomial Enumeration	**10000**
Test Type	**Exact Test**
Alpha	**0.05**
n (Sample Size)	**20 to 200 by 20**
Input Type	**Differences**
PB (Baseline Proportion)	**0.50**
d0 (Equivalence Difference)	**0.10**
d1 (Actual Difference)	**0**

Reports Tab

Show Numeric Reports	**Not checked**
Show Comparative Reports	**Checked**

Plots Tab

Show Plots	**Not checked**
Show Comparative Plots	**Checked**

Output

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

Power Comparison of Five Different Equivalence Tests for One Proportion

Alternative Hypothesis: Equivalence (H0: P ≤ P0L or P ≥ P0U vs. H1: P0L < P < P0U)

		Difference			Power				
Sample Size n	Baseline Proportion PB	Equivalence d0	Actual d1	Target Alpha	Exact Test	Z-Test S(P0)	Z-Test S(P0)C	Z-Test S(P)	Z-Test S(P)C
20	0.5	0.1	0	0.05	0.00000	0.00000	0.00000	0.00000	0.00000
40	0.5	0.1	0	0.05	0.00000	0.00000	0.00000	0.00000	0.00000
60	0.5	0.1	0	0.05	0.00000	0.00000	0.00000	0.00000	0.00000
80	0.5	0.1	0	0.05	0.08893	0.08893	0.08893	0.08893	0.08893
100	0.5	0.1	0	0.05	0.23565	0.23565	0.23565	0.23565	0.23565
120	0.5	0.1	0	0.05	0.35174	0.47701	0.35174	0.47701	0.35174
140	0.5	0.1	0	0.05	0.44573	0.55301	0.44573	0.55301	0.44573
160	0.5	0.1	0	0.05	0.61543	0.61543	0.61543	0.61543	0.61543
180	0.5	0.1	0	0.05	0.66742	0.73650	0.66742	0.66742	0.66742
200	0.5	0.1	0	0.05	0.77075	0.77075	0.77075	0.77075	0.71118

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

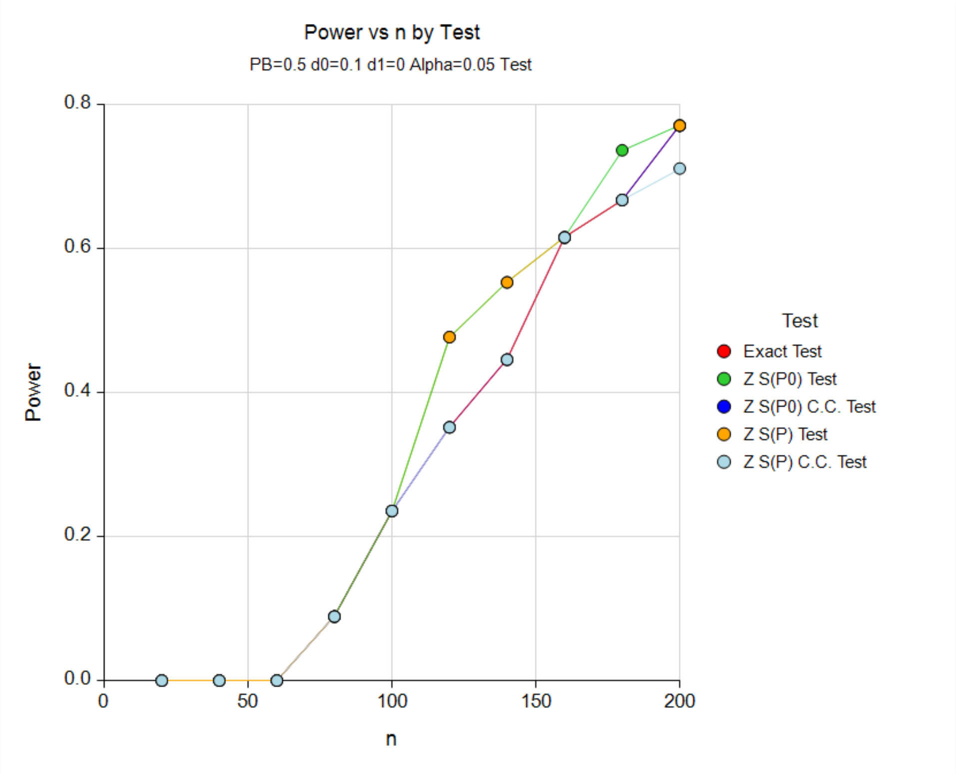
Actual Alpha Comparison of Five Different Equivalence Tests for One Proportion

Alternative Hypothesis: Equivalence (H0: P ≤ P0L or P ≥ P0U vs. H1: P0L < P < P0U)

		Difference		Alpha					
Sample Size n	Baseline Proportion PB	Equivalence d0	Actual d1	Target	Exact Test	Z-Test S(P0)	Z-Test S(P0)C	Z-Test S(P)	Z-Test S(P)C
20	0.5	0.1	0	0.05	0.0000	0.0565	0.0210	0.0565	0.0210
40	0.5	0.1	0	0.05	0.0000	0.0392	0.0392	0.0392	0.0392
60	0.5	0.1	0	0.05	0.0000	0.0445	0.0445	0.0445	0.0445
80	0.5	0.1	0	0.05	0.0445	0.0445	0.0445	0.0445	0.0445
100	0.5	0.1	0	0.05	0.0423	0.0423	0.0423	0.0423	0.0423
120	0.5	0.1	0	0.05	0.0392	0.0575	0.0392	0.0575	0.0392
140	0.5	0.1	0	0.05	0.0358	0.0514	0.0358	0.0514	0.0358
160	0.5	0.1	0	0.05	0.0459	0.0459	0.0459	0.0459	0.0459
180	0.5	0.1	0	0.05	0.0408	0.0558	0.0408	0.0408	0.0408
200	0.5	0.1	0	0.05	0.0492	0.0492	0.0492	0.0492	0.0363

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

Plots



The first report shows the power for each test statistic. The second report shows the actual alpha achieved by the design.

An examination of the first report shows that once non-zero powers are obtained, they are often different for at least one of the tests. Also notice that the exact test always has the minimum power in each row. This would lead us discard this test statistic. However, consider the second report which shows the actual alpha level (the target was 0.05) for each test. By inspecting corresponding entries in both tables, we see that whenever a test statistic achieves a better power than the exact test, it also yields an actual alpha level larger than the target alpha.

For example, look at the powers for *n* = 120. The *z* test using s(P0) has an unusually large power = 0.4770. This is a much larger power than the exact test's value of 0.3517. However, note that the actual alpha for this test is 0.0575 which is larger than the target alpha of 0.05 and the exact test's alpha of 0.0392.

We conclude that indeed, the exact test is consistently the best test since it always achieves a significance level that is less than the target value.

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 .... **Binomial Enumeration**
Max n for Binomial Enumeration.................... **10000**
Higher Proportions Are .... **Better**
Test Type.. **Z-test using S(P0)**
Alpha.. **0.05**
n (Sample Size) .... **20 to 200 by 20**
Input Type... **Differences**
PB (Baseline Proportion) **0.50**
d0 (Non-Inferiority Difference)........................ **-0.10**
d1 (Actual Difference) **0**

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 Equivalence of One Proportion using the Exact Test

Alternative Hypothesis: Equivalence (H0: P ≤ P0L or P ≥ P0U vs. H1: P0L < P < P0U)

		Difference		Normal Approximation		Binomial Enumeration	
Sample Size n	Baseline Proportion PB	Equivalence d0	Actual d1	Power	Alpha	Power	Alpha
20	0.5	0.1	0	0.00000	0.05	0.00000	0.0000
40	0.5	0.1	0	0.00000	0.05	0.00000	0.0000
60	0.5	0.1	0	0.00000	0.05	0.00000	0.0000
80	0.5	0.1	0	0.14068	0.05	0.08893	0.0445
100	0.5	0.1	0	0.30226	0.05	0.23565	0.0423
120	0.5	0.1	0	0.43759	0.05	0.35174	0.0392
140	0.5	0.1	0	0.54964	0.05	0.44573	0.0358
160	0.5	0.1	0	0.64149	0.05	0.61543	0.0459
180	0.5	0.1	0	0.71613	0.05	0.66742	0.0408
200	0.5	0.1	0	0.77632	0.05	0.77075	0.0492

Notice that the approximate power values consistently overestimate the power, particularly for small sample sizes. As the sample size increases, the approximate power values come nearer to the power values from binomial enumeration.

Example 5 – Finding Power after an Experiment

Researchers are testing a generic drug to determine if it is equivalent to the name-brand alternative. Equivalence is declared if the success rate of the generic brand is no more than 10% from that of the name-brand drug. Suppose that the name-brand drug is known to have a success rate of 60%. In a study of 500 individuals, they find that 265, or 53%, are successfully treated using the generic brand. An equivalence test (exact test) with alpha = 0.05 failed to declare that the two drugs are equivalent. The researchers would now like to compute the power for actual differences ranging from 0 to 9%.

Note that the power is not calculated solely at the difference observed in the study, 7%. It is more informative to study a range of values with practical significance.

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**
Max n for Binomial Enumeration	**10000**
Test Type	**Exact Test**
Alpha	**0.05**
n (Sample Size)	**500**
Input Type	**Differences**
PB (Baseline Proportion)	**0.60**
d0 (Equivalence Difference)	**0.10**
d1 (Actual Difference)	**0.0 to 0.09 by 0.01**

Output

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

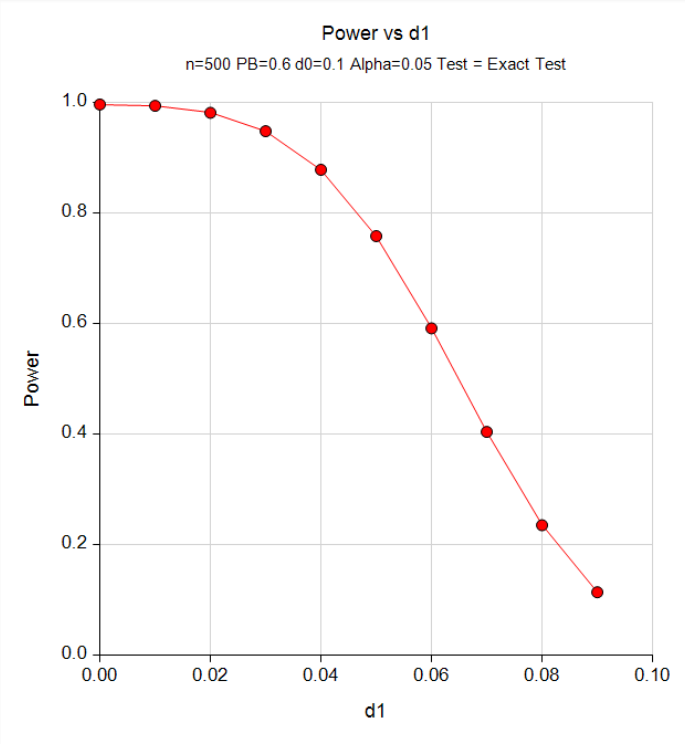
Numeric Results

Solve For: Power
Alternative Hypothesis: Equivalence (H0: P ≤ P0L or P ≥ P0U vs. H1: P0L < P < P0U)
Test Type: Exact Test

	Sample	Baseline	Equivalence Proportion		Difference		Alpha		Reject H0 if
Power*	Size n	Proportion PB	Lower P0L	Upper P0U	Equivalence d0	Actual d1	Target	Actual*	R1 ≤ R ≤ R2 R1|R2
0.99649	500	0.6	0.5	0.7	0.1	0.00	0.05	0.0489	269|332
0.99404	500	0.6	0.5	0.7	0.1	0.01	0.05	0.0489	269|332
0.98146	500	0.6	0.5	0.7	0.1	0.02	0.05	0.0489	269|332
0.94824	500	0.6	0.5	0.7	0.1	0.03	0.05	0.0489	269|332
0.87825	500	0.6	0.5	0.7	0.1	0.04	0.05	0.0489	269|332
0.75828	500	0.6	0.5	0.7	0.1	0.05	0.05	0.0489	269|332
0.59143	500	0.6	0.5	0.7	0.1	0.06	0.05	0.0489	269|332
0.40407	500	0.6	0.5	0.7	0.1	0.07	0.05	0.0489	269|332
0.23522	500	0.6	0.5	0.7	0.1	0.08	0.05	0.0489	269|332
0.11389	500	0.6	0.5	0.7	0.1	0.09	0.05	0.0489	269|332

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

Plots



The range in power is quite large. The power is relatively high and constant if the true difference is less than or equal to 4%, but it decreases rapidly as the differences increase from there.

Example 6 – Finding the Sample Size using Ratios

Researchers would like to compare a new treatment to an existing standard treatment. The new treatment will be deemed equivalent to the standard treatment if the response rate is changed by no more than 20%, hence, r = 1.20. It is known that 60% of patients respond to the standard treatment. If the researchers use the exact test and a significance level of 0.05, how large of a sample must they take to achieve 90% power if the actual ratio is 1.0?

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 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**
Test Type	**Exact Test**
Power	**0.90**
Alpha	**0.05**
Input Type	**Ratios**
PB (Baseline Proportion)	**0.60**
r0 (Equivalence Ratio)	**1.2**
r1 (Actual Ratio)	**1.0**

Output

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

Numeric Results

Solve For: Sample Size
Alternative Hypothesis: Equivalence (H0: P ≤ P0L or P ≥ P0U vs. H1: P0L < P < P0U)
Test Type: Exact Test

Power*	Sample Size n	Baseline Proportion PB	Equivalence Proportion Lower P0L	Equivalence Proportion Upper P0U	Ratio Equivalence r0	Ratio Actual r1	Alpha	Reject H0 if R1 ≤ R ≤ R2 R1|R2
0.90019	224	0.6	0.5	0.72	1.2	1	0.05	125|149

* Power was computed using the normal approximation method.

They must sample 224 individuals to achieve just over 90% power for an actual ratio of 1.0 and equivalence ratio of 1.20.

Example 7 – Validation using Chow, Shao, and Wang (2008)

Chow, Shao, and Wang (2008) page 88 gives the result of a sample size calculation for the z-test with S(Phat). They calculate a sample size of 52 when alpha = 0.05, power = 0.80, PB = 0.60, equivalence difference = 0.20, and actual difference = 0.0.

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 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**
Test Type.. **Z-test using S(Phat)**
Power... **0.80**
Alpha.. **0.05**
Input Type... **Differences**
PB (Baseline Proportion) **0.60**
d0 (Equivalence Difference).......................... **0.20**
d1 (Actual Difference) **0**

Output

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

Numeric Results

Solve For: Sample Size
Alternative Hypothesis: Equivalence (H0: P ≤ P0L or P ≥ P0U vs. H1: P0L < P < P0U)
Test Type: Z-Test with S(Phat)

			Equivalence Proportion		Difference			
Power*	Sample Size n	Baseline Proportion PB	Lower P0L	Upper P0U	Equivalence d0	Actual d1	Alpha	Reject H0 if |Z| >
0.80608	52	0.6	0.4	0.8	0.2	0	0.05	1.6449

* Power was computed using the normal approximation method.

PASS has also calculated the sample size to be 52.