

Chapter 216

Confidence Intervals for the Difference Between Two Proportions

Introduction

This routine calculates the group sample sizes necessary to achieve a specified interval width of the difference between two independent proportions.

Caution: These procedures assume that the proportions obtained from future samples will be the same as the proportions that are specified. If the sample proportions are different from those specified when running these procedures, the interval width may be narrower or wider than specified.

Technical Details

A background of the comparison of two proportions is given, followed by details of the confidence interval methods available in this procedure.

Comparing Two Proportions

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 a population and that the responses from one subject to the next are independent of one another.

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

	Success	Failure	Total
Population 1	<i>a</i>	<i>c</i>	<i>m</i>
Population 2	<i>b</i>	<i>d</i>	<i>n</i>
Totals	<i>s</i>	<i>f</i>	<i>N</i>

Confidence Intervals for the Difference Between Two Proportions

The following alternative notation is sometimes used:

	Success	Failure	Total
Population 1	x_{11}	x_{12}	n_1
Population 2	x_{21}	x_{22}	n_2
Totals	m_1	m_2	N

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

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

When analyzing studies such as these, you usually want to compare the two binomial probabilities p_1 and p_2 . The most direct methods of comparing these quantities are 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

<u>Parameter</u>	<u>Computation</u>
Difference	$\delta = p_1 - p_2$
Risk Ratio	$\phi = p_1 / p_2$
Odds Ratio	$\psi = \frac{p_1 / q_1}{p_2 / q_2} = \frac{p_1 q_2}{p_2 q_1}$

The choice of which of these measures is used might at seem arbitrary, but it is important. Not only is their interpretation different, but, for small sample sizes, the coverage probabilities may be different. This procedure focuses on the difference. Other procedures are available in **PASS** for computing confidence intervals for the ratio and odds ratio.

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 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 which 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 fits the situation.

Confidence Intervals for the Difference

Many methods have been devised for computing confidence intervals for the difference between two proportions $\delta = p_1 - p_2$. Seven of these methods are available in the Confidence Intervals for Two Proportions [Proportions] using Proportions and Confidence Intervals for Two Proportions [Differences] procedures. The seven confidence interval methods are

1. Score (Farrington and Manning)
2. Score (Miettinen and Nurminen)
3. Score with Correction for Skewness (Gart and Nam)
4. Score (Wilson)
5. Score with Continuity Correction (Wilson)
6. Chi-Square with Continuity Correction (Yates)
7. Chi-Square (Pearson)

Newcombe (1998b) conducted a comparative evaluation of eleven confidence interval methods. He recommended that the modified Wilson score method be used instead of the Pearson Chi-Square or the Yate's Corrected Chi-Square. Beal (1987) found that the Score methods performed very well. The lower L and upper U limits of these intervals are computed as follows. Note that, unless otherwise stated, $z = |z_{\alpha/2}|$ is the appropriate percentile from the standard normal distribution.

Farrington and Manning's Score

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 test statistic formula 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 as

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

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

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

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

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

Confidence Intervals for the Difference Between Two Proportions

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

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

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

$$L_3 = N$$

Farrington and Manning (1990) proposed inverting their score test to find the confidence interval. The lower limit is found by solving

$$z_{FMD} = |z_{\alpha/2}|$$

and the upper limit is the solution of

$$z_{FMD} = -|z_{\alpha/2}|$$

Miettinen and Nurminen's Score

Miettinen and Nurminen (1985) 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. A correction factor of $N/(N-1)$ is applied to make the variance estimate less biased. The significance level of the test statistic is based on the asymptotic normality of the score statistic.

The formula for computing this test statistic is

$$z_{MND} = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0}{\sqrt{\left(\frac{\tilde{p}_1 \tilde{q}_1}{n_1} + \frac{\tilde{p}_2 \tilde{q}_2}{n_2}\right) \left(\frac{N}{N-1}\right)}}$$

where

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

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

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

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

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

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

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

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

$$L_3 = N$$

Confidence Intervals for the Difference Between Two Proportions

Miettinen and Nurminen (1985) proposed inverting their score test to find the confidence interval. The lower limit is found by solving

$$z_{MND} = |z_{\alpha/2}|$$

and the upper limit is the solution of

$$z_{MND} = -|z_{\alpha/2}|$$

Gart and Nam's Score

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

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

where

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

Gart and Nam (1988) proposed inverting their score test to find the confidence interval. The lower limit is found by solving

$$z_{GND} = |z_{\alpha/2}|$$

and the upper limit is the solution of

$$z_{GND} = -|z_{\alpha/2}|$$

Wilson's Score as Modified by Newcombe (with and without Continuity Correction)

For details, see Newcombe (1998b), page 876.

$$L = \hat{p}_1 - \hat{p}_2 - B$$

$$U = \hat{p}_1 - \hat{p}_2 + C$$

where

$$B = z \sqrt{\frac{l_1(1-l_1)}{m} + \frac{u_2(1-u_2)}{n}}$$

$$C = z \sqrt{\frac{u_1(1-u_1)}{m} + \frac{l_2(1-l_2)}{n}}$$

and l_1 and u_1 are the roots of

$$|p_1 - \hat{p}_1| - z \sqrt{\frac{p_1(1-p_1)}{m}} = 0$$

and l_2 and u_2 are the roots of

$$|p_2 - \hat{p}_2| - z \sqrt{\frac{p_2(1-p_2)}{n}} = 0$$

Confidence Intervals for the Difference Between Two Proportions

Yate's Chi-Square with Continuity Correction

For details, see Newcombe (1998b), page 875.

$$L = \hat{p}_1 - \hat{p}_2 - z \sqrt{\left(\frac{\hat{p}_1(1-\hat{p}_1)}{m} + \frac{\hat{p}_2(1-\hat{p}_2)}{n} \right)} - \frac{1}{2} \left(\frac{1}{m} + \frac{1}{n} \right)$$

$$U = \hat{p}_1 - \hat{p}_2 + z \sqrt{\left(\frac{\hat{p}_1(1-\hat{p}_1)}{m} + \frac{\hat{p}_2(1-\hat{p}_2)}{n} \right)} + \frac{1}{2} \left(\frac{1}{m} + \frac{1}{n} \right)$$

Pearson's Chi-Square

For details, see Newcombe (1998b), page 875.

$$L = \hat{p}_1 - \hat{p}_2 - z \sqrt{\left(\frac{\hat{p}_1(1-\hat{p}_1)}{m} + \frac{\hat{p}_2(1-\hat{p}_2)}{n} \right)}$$

$$U = \hat{p}_1 - \hat{p}_2 + z \sqrt{\left(\frac{\hat{p}_1(1-\hat{p}_1)}{m} + \frac{\hat{p}_2(1-\hat{p}_2)}{n} \right)}$$

For each of the seven methods, one-sided intervals may be obtained by replacing $\alpha/2$ by α .

For two-sided intervals, the distance from the difference in sample proportions to each of the limits may be different. Thus, instead of specifying the distance to the limits we specify the width of the interval, W .

The basic equation for determining sample size for a two-sided interval when W has been specified is

$$W = U - L$$

For one-sided intervals, the distance from the variance ratio to limit, D , is specified.

The basic equation for determining sample size for a one-sided upper limit when D has been specified is

$$D = U - (\hat{p}_1 - \hat{p}_2)$$

The basic equation for determining sample size for a one-sided lower limit when D has been specified is

$$D = (\hat{p}_1 - \hat{p}_2) - L$$

Each of these equations can be solved for any of the unknown quantities in terms of the others.

Confidence Level

The confidence level, $1 - \alpha$, has the following interpretation. If thousands of random samples of size n_1 and n_2 are drawn from populations 1 and 2, respectively, and a confidence interval for the true difference/ratio/odds ratio of proportions is calculated for each pair of samples, the proportion of those intervals that will include the true difference/ratio/odds ratio of proportions is $1 - \alpha$.

Procedure Options

This section describes the options that are specific to this procedure. These are located on the Design tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Design Tab

The Design tab contains the parameters associated with this calculation such as the proportions or differences, sample sizes, confidence level, and interval width.

Solve For

Solve For

This option specifies the parameter to be solved for from the other parameters.

Confidence Interval Method

Confidence Interval Formula

Specify the formula to be in used in calculation of confidence intervals.

- **Score (Farrington & Manning)**
This formula is based on inverting Farrington and Manning's score test.
 - **Score (Miettinen & Nurminen)**
This formula is based on inverting Miettinen and Nurminen's score test.
 - **Score w/ Skewness (Gart & Nam)**
This formula is based on inverting Gart and Nam's score test, with a correction for skewness.
 - **Score (Wilson)**
This formula is based on the Wilson score method for a single proportion, without continuity correction.
 - **Score (Wilson C.C.)**
This formula is based on the Wilson score method for a single proportion, with continuity correction.
 - **Chi-Square C.C. (Yates)**
This is the commonly used simple asymptotic method, with continuity correction.
 - **Chi-Square (Pearson)**
This is the commonly used simple asymptotic method, without continuity correction.
-

One-Sided or Two-Sided Interval

Interval Type

Specify whether the interval to be used will be a two-sided confidence interval, an interval that has only an upper limit, or an interval that has only a lower limit.

Confidence Intervals for the Difference Between Two Proportions

Confidence

Confidence Level (1 – Alpha)

The confidence level, $1 - \alpha$, has the following interpretation. If thousands of random samples of size n_1 and n_2 are drawn from populations 1 and 2, respectively, and a confidence interval for the true difference/ratio/odds ratio of proportions is calculated for each pair of samples, the proportion of those intervals that will include the true difference/ratio/odds ratio of proportions is $1 - \alpha$.

Often, the values 0.95 or 0.99 are used. You can enter single values or a range of values such as 0.90, 0.95 or 0.90 to 0.99 by 0.01.

Sample Size (When Solving for Sample Size)

Group Allocation

Select the option that describes the constraints on $N1$ or $N2$ or both.

The options are

- **Equal ($N1 = N2$)**
This selection is used when you wish to have equal sample sizes in each group. Since you are solving for both sample sizes at once, no additional sample size parameters need to be entered.
- **Enter $N1$, solve for $N2$**
Select this option when you wish to fix $N1$ at some value (or values), and then solve only for $N2$. Please note that for some values of $N1$, there may not be a value of $N2$ that is large enough to obtain the desired power.
- **Enter $N2$, solve for $N1$**
Select this option when you wish to fix $N2$ at some value (or values), and then solve only for $N1$. Please note that for some values of $N2$, there may not be a value of $N1$ that is large enough to obtain the desired power.
- **Enter $R = N2/N1$, solve for $N1$ and $N2$**
For this choice, you set a value for the ratio of $N2$ to $N1$, and then PASS determines the needed $N1$ and $N2$, with this ratio, to obtain the desired power. An equivalent representation of the ratio, R , is

$$N2 = R * N1.$$
- **Enter percentage in Group 1, solve for $N1$ and $N2$**
For this choice, you set a value for the percentage of the total sample size that is in Group 1, and then PASS determines the needed $N1$ and $N2$ with this percentage to obtain the desired power.

$N1$ (Sample Size, Group 1)

This option is displayed if Group Allocation = "Enter $N1$, solve for $N2$ "

$N1$ is the number of items or individuals sampled from the Group 1 population.

$N1$ must be ≥ 2 . You can enter a single value or a series of values.

$N2$ (Sample Size, Group 2)

This option is displayed if Group Allocation = "Enter $N2$, solve for $N1$ "

$N2$ is the number of items or individuals sampled from the Group 2 population.

$N2$ must be ≥ 2 . You can enter a single value or a series of values.

Confidence Intervals for the Difference Between Two Proportions

R (Group Sample Size Ratio)

This option is displayed only if Group Allocation = "Enter R = N2/N1, solve for N1 and N2."

R is the ratio of N2 to N1. That is,

$$R = N2 / N1.$$

Use this value to fix the ratio of N2 to N1 while solving for N1 and N2. Only sample size combinations with this ratio are considered.

N2 is related to N1 by the formula:

$$N2 = [R \times N1],$$

where the value $[Y]$ is the next integer $\geq Y$.

For example, setting $R = 2.0$ results in a Group 2 sample size that is double the sample size in Group 1 (e.g., $N1 = 10$ and $N2 = 20$, or $N1 = 50$ and $N2 = 100$).

R must be greater than 0. If $R < 1$, then N2 will be less than N1; if $R > 1$, then N2 will be greater than N1. You can enter a single or a series of values.

Percent in Group 1

This option is displayed only if Group Allocation = "Enter percentage in Group 1, solve for N1 and N2."

Use this value to fix the percentage of the total sample size allocated to Group 1 while solving for N1 and N2. Only sample size combinations with this Group 1 percentage are considered. Small variations from the specified percentage may occur due to the discrete nature of sample sizes.

The Percent in Group 1 must be greater than 0 and less than 100. You can enter a single or a series of values.

Sample Size (When Not Solving for Sample Size)

Group Allocation

Select the option that describes how individuals in the study will be allocated to Group 1 and to Group 2.

The options are

- **Equal (N1 = N2)**
This selection is used when you wish to have equal sample sizes in each group. A single per group sample size will be entered.
- **Enter N1 and N2 individually**
This choice permits you to enter different values for N1 and N2.
- **Enter N1 and R, where N2 = R * N1**
Choose this option to specify a value (or values) for N1, and obtain N2 as a ratio (multiple) of N1.
- **Enter total sample size and percentage in Group 1**
Choose this option to specify a value (or values) for the total sample size (N), obtain N1 as a percentage of N, and then N2 as $N - N1$.

Confidence Intervals for the Difference Between Two Proportions

Sample Size Per Group

This option is displayed only if Group Allocation = "Equal ($N1 = N2$)."

The Sample Size Per Group is the number of items or individuals sampled from each of the Group 1 and Group 2 populations. Since the sample sizes are the same in each group, this value is the value for $N1$, and also the value for $N2$.

The Sample Size Per Group must be ≥ 2 . You can enter a single value or a series of values.

N1 (Sample Size, Group 1)

*This option is displayed if Group Allocation = "Enter $N1$ and $N2$ individually" or "Enter $N1$ and R , where $N2 = R * N1$."*

$N1$ is the number of items or individuals sampled from the Group 1 population.

$N1$ must be ≥ 2 . You can enter a single value or a series of values.

N2 (Sample Size, Group 2)

This option is displayed only if Group Allocation = "Enter $N1$ and $N2$ individually."

$N2$ is the number of items or individuals sampled from the Group 2 population.

$N2$ must be ≥ 2 . You can enter a single value or a series of values.

R (Group Sample Size Ratio)

*This option is displayed only if Group Allocation = "Enter $N1$ and R , where $N2 = R * N1$."*

R is the ratio of $N2$ to $N1$. That is,

$$R = N2/N1$$

Use this value to obtain $N2$ as a multiple (or proportion) of $N1$.

$N2$ is calculated from $N1$ using the formula:

$$N2 = [R \times N1],$$

where the value $[Y]$ is the next integer $\geq Y$.

For example, setting $R = 2.0$ results in a Group 2 sample size that is double the sample size in Group 1.

R must be greater than 0. If $R < 1$, then $N2$ will be less than $N1$; if $R > 1$, then $N2$ will be greater than $N1$. You can enter a single value or a series of values.

Total Sample Size (N)

This option is displayed only if Group Allocation = "Enter total sample size and percentage in Group 1."

This is the total sample size, or the sum of the two group sample sizes. This value, along with the percentage of the total sample size in Group 1, implicitly defines $N1$ and $N2$.

The total sample size must be greater than one, but practically, must be greater than 3, since each group sample size needs to be at least 2.

You can enter a single value or a series of values.

Percent in Group 1

This option is displayed only if Group Allocation = "Enter total sample size and percentage in Group 1."

This value fixes the percentage of the total sample size allocated to Group 1. Small variations from the specified percentage may occur due to the discrete nature of sample sizes.

The Percent in Group 1 must be greater than 0 and less than 100. You can enter a single value or a series of values.

Confidence Intervals for the Difference Between Two Proportions

Precision

Confidence Interval Width (Two-Sided)

This is the distance from the lower confidence limit to the upper confidence limit.

You can enter a single value or a list of values. The value(s) must be greater than zero.

Distance from Diff to Limit (One-Sided)

This is the distance from the difference in sample proportions to the lower or upper limit of the confidence interval, depending on whether the Interval Type is set to Lower Limit or Upper Limit.

You can enter a single value or a list of values. The value(s) must be greater than zero.

Proportions (Difference = $P1 - P2$)

Input Type

Indicate what type of values to enter to specify the difference. Regardless of the entry type chosen, the calculations are the same. This option is simply given for convenience in specifying the difference.

P1 – P2 (Difference in Sample Proportions)

This option is displayed only if Input Type = "Differences"

Enter an estimate of the difference between sample proportion 1 and sample proportion 2. The sample size and width calculations assume that the value entered here is the difference estimate that is obtained from the sample. If the sample difference is different from the one specified here, the width may be narrower or wider than specified.

The value(s) must be between -1 and 1, and such that $P1 = \text{Difference} + P2$ is between 0.0001 and 0.9999.

You can enter a range of values such as *.1 .2 .3* or *.1 to .5 by .1*.

P1 (Proportion Group 1)

This option is displayed only if Input Type = "Proportions"

Enter an estimate of the proportion for group 1. The sample size and width calculations assume that the value entered here is the proportion estimate that is obtained from the sample. If the sample proportion is different from the one specified here, the width may be narrower or wider than specified.

The value(s) must be between 0.0001 and 0.9999.

You can enter a range of values such as *.1 .2 .3* or *.1 to .5 by .1*.

P2 (Proportion Group 2)

Enter an estimate of the proportion for group 2. The sample size and width calculations assume that the value entered here is the proportion estimate that is obtained from the sample. If the sample proportion is different from the one specified here, the width may be narrower or wider than specified.

The value(s) must be between 0.0001 and 0.9999.

You can enter a range of values such as *.1 .2 .3* or *.1 to .5 by .1*.

Confidence Intervals for the Difference Between Two Proportions

Example 1 – Calculating Sample Size using Differences

Suppose a study is planned in which the researcher wishes to construct a two-sided 95% confidence interval for the difference in proportions such that the width of the interval is no wider than 0.1. The confidence interval method to be used is the Yates chi-square simple asymptotic method with continuity correction. The confidence level is set at 0.95, but 0.99 is included for comparative purposes. The difference estimate to be used is 0.05, and the estimate for proportion 2 is 0.3. Instead of examining only the interval width of 0.1, a series of widths from 0.05 to 0.3 will also be considered.

The goal is to determine the necessary sample size.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Confidence Intervals for the Difference Between Two Proportions** procedure window by expanding **Proportions**, then **Two Independent Proportions**, then clicking on **Confidence Interval**, and then clicking on **Confidence Intervals for the Difference Between Two Proportions**. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Sample Size
Confidence Interval Formula	Chi-Square C.C. (Yates)
Interval Type	Two-Sided
Confidence Level	0.95 0.99
Group Allocation	Equal (N1 = N2)
Confidence Interval Width (Two-Sided)	0.05 to 0.30 by 0.05
Input Type	Differences
P1 – P2 (Difference in Sample Proportions)	0.05
P2	0.3

Annotated Output

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

Numeric Results

Numeric Results for Two-Sided Confidence Intervals for the Difference in Proportions
Confidence Interval Method: Chi-Square - Simple Asymptotic with Continuity Correction (Yates)

Confidence Level	N1	N2	N	Target Width	Actual Width	P1	P2	P1 - P2	Lower Limit	Upper Limit
0.950	2769	2769	5538	0.050	0.050	0.35	0.30	0.05	0.03	0.07
0.950	712	712	1424	0.100	0.100	0.35	0.30	0.05	0.00	0.10
0.950	325	325	650	0.150	0.150	0.35	0.30	0.05	-0.02	0.12
0.950	188	188	376	0.200	0.200	0.35	0.30	0.05	-0.05	0.15
0.950	124	124	248	0.250	0.249	0.35	0.30	0.05	-0.07	0.17
0.950	88	88	176	0.300	0.299	0.35	0.30	0.05	-0.10	0.20
0.990	4725	4725	9450	0.050	0.050	0.35	0.30	0.05	0.03	0.07
0.990	1201	1201	2402	0.100	0.100	0.35	0.30	0.05	0.00	0.10
0.990	543	543	1086	0.150	0.150	0.35	0.30	0.05	-0.02	0.12
0.990	310	310	620	0.200	0.200	0.35	0.30	0.05	-0.05	0.15
0.990	202	202	404	0.250	0.250	0.35	0.30	0.05	-0.07	0.17
0.990	143	143	286	0.300	0.299	0.35	0.30	0.05	-0.10	0.20

Confidence Intervals for the Difference Between Two Proportions

References

Newcombe, R. G. 1998. 'Interval Estimation for the Difference Between Independent Proportions: Comparison of Eleven Methods.' *Statistics in Medicine*, 17, pp. 873-890.
 Fleiss, J. L., Levin, B., Paik, M.C. 2003. *Statistical Methods for Rates and Proportions*. Third Edition. John Wiley & Sons. New York.

Report Definitions

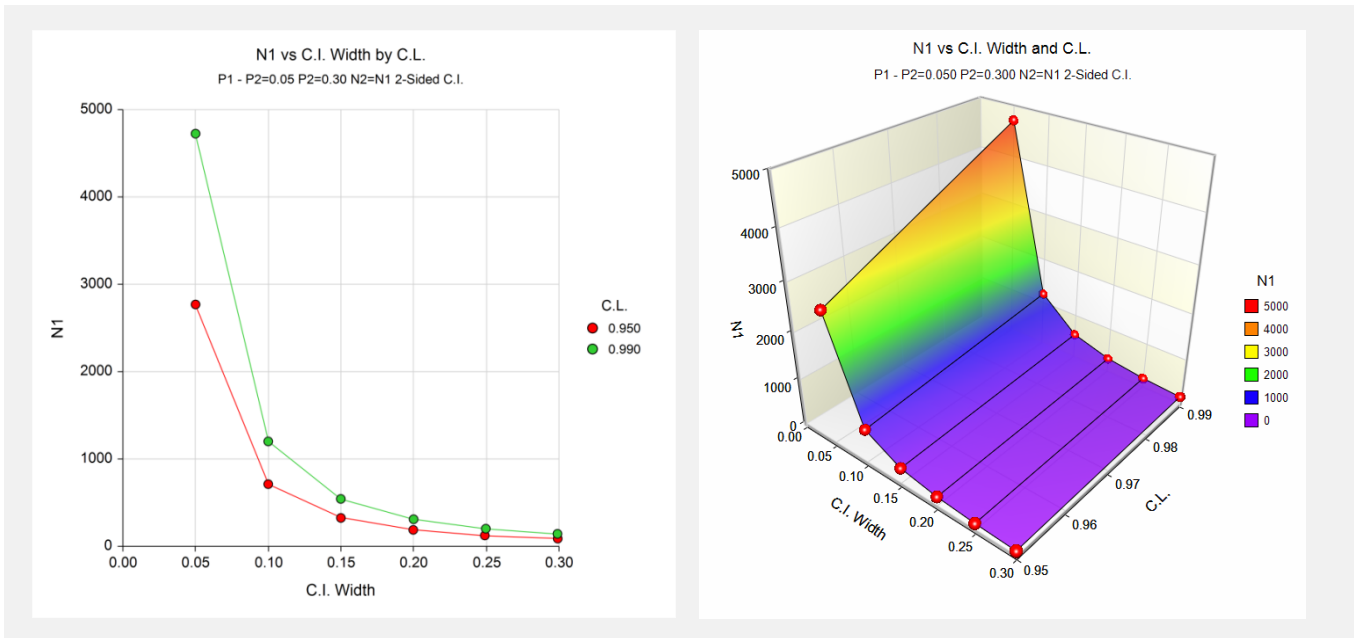
Confidence level is the proportion of confidence intervals (constructed with this same confidence level, sample size, etc.) that would contain the true difference in population proportions.
 N1 and N2 are the number of items sampled from each population.
 N is the total sample size, N1 + N2.
 Target Width is the value of the width that is entered into the procedure.
 Actual Width is the value of the width that is obtained from the procedure.
 P1 and P2 are the assumed sample proportions for sample size calculations.
 P1 - P2 is the difference between sample proportions at which sample size calculations are made.
 Lower Limit and Upper Limit are the lower and upper limits of the confidence interval for the true difference in proportions (Population Proportion 1 - Population Proportion 2).

Summary Statements

Group sample sizes of 2769 and 2769 produce a two-sided 95% confidence interval for the difference in population proportions with a width that is equal to 0.050 when the estimated sample proportion 1 is 0.35, the estimated sample proportion 2 is 0.30, and the difference in sample proportions is 0.05.

This report shows the calculated sample sizes for each of the scenarios.

Plots Section



These plots show the group sample size versus the confidence interval width for the two confidence levels.

Confidence Intervals for the Difference Between Two Proportions

Example 2 – Calculating Sample Size using Proportions

Suppose a study is planned in which the researcher wishes to construct a two-sided 95% confidence interval for the difference in proportions such that the width of the interval is no wider than 0.1. The confidence interval method to be used is the Yates chi-square simple asymptotic method with continuity correction. The confidence level is set at 0.95, but 0.99 is included for comparative purposes. The proportion estimates to be used are 0.6 for Group 1, and 0.4 for Group 2. Instead of examining only the interval width of 0.1, a series of widths from 0.05 to 0.3 will also be considered.

The goal is to determine the necessary sample size.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Confidence Intervals for the Difference Between Two Proportions** procedure window by expanding **Proportions**, then **Two Independent Proportions**, then clicking on **Confidence Interval**, and then clicking on **Confidence Intervals for the Difference Between Two Proportions**. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Sample Size
Confidence Interval Formula	Chi-Square C.C. (Yates)
Interval Type	Two-Sided
Confidence Level	0.95 0.99
Group Allocation	Equal (N1 = N2)
Confidence Interval Width (Two-Sided)	0.05 to 0.30 by 0.05
Input Type	Proportions
P1	0.6
P2	0.4

Output

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

Numeric Results

Numeric Results for Two-Sided Confidence Intervals for the Difference in Proportions										
Confidence Interval Method: Chi-Square - Simple Asymptotic with Continuity Correction (Yates)										
Confidence Level	N1	N2	N	Target Width	Actual Width	P1	P2	P1 - P2	Lower Limit	Upper Limit
0.950	3030	3030	6060	0.050	0.050	0.60	0.40	0.20	0.18	0.22
0.950	778	778	1556	0.100	0.100	0.60	0.40	0.20	0.15	0.25
0.950	354	354	708	0.150	0.150	0.60	0.40	0.20	0.13	0.27
0.950	204	204	408	0.200	0.200	0.60	0.40	0.20	0.10	0.30
0.950	134	134	268	0.250	0.250	0.60	0.40	0.20	0.08	0.32
0.950	95	95	190	0.300	0.300	0.60	0.40	0.20	0.05	0.35
0.990	5176	5176	10352	0.050	0.050	0.60	0.40	0.20	0.18	0.22
0.990	1314	1314	2628	0.100	0.100	0.60	0.40	0.20	0.15	0.25
0.990	593	593	1186	0.150	0.150	0.60	0.40	0.20	0.13	0.27
0.990	339	339	678	0.200	0.200	0.60	0.40	0.20	0.10	0.30
0.990	220	220	440	0.250	0.250	0.60	0.40	0.20	0.08	0.32
0.990	155	155	310	0.300	0.300	0.60	0.40	0.20	0.05	0.35

This report shows the calculated sample sizes for each of the scenarios.

Example 3 – Validation using Newcombe (1998b)

Newcombe (1998b) page 877 gives an example of a calculation for a confidence interval for the difference in proportions when the confidence level is 95%, the sample proportions are 0.9 and 0.3, and the interval width is 0.6790 for the Chi-Square (Pearson) method, 0.8395 for the Chi-Square C.C. (Yates) method, 0.67064 for the Score (Miettinen and Nurminen) method, 0.6385 for the Score (Wilson) method, and 0.7374 for the Score C.C. (Wilson) method. The necessary sample size in each case is 10 per group.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Confidence Intervals for the Difference Between Two Proportions** procedure window by expanding **Proportions**, then **Two Independent Proportions**, then clicking on **Confidence Interval**, and then clicking on **Confidence Intervals for the Difference Between Two Proportions**. You may then make the appropriate entries as listed below, or open **Example 3(a-e)** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Sample Size
Confidence Interval Formula	Varies [Chi-Square (Pearson), Chi-Square C.C. (Yates), Score (Miettinen & Nurminen), Score (Wilson), Score C.C. (Wilson)]
Interval Type	Two-Sided
Confidence Level	0.95
Group Allocation	Equal (N1 = N2)
Confidence Interval Width (Two-Sided)	Varies (0.6790, 0.8395, 0.67064, 0.6385, 0.7374)
Input Type	Proportions
P1	0.9
P2	0.3

Output

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

Chi-Square (Pearson)

Confidence Level	N1	N2	N	Target Width	Actual Width	P1	P2	P1 - P2	Lower Limit	Upper Limit
0.950	10	10	20	0.6790	0.6790	0.9000	0.3000	0.6000	0.2605	0.9395

PASS also calculates the necessary sample size to be 10 per group.

Chi-Square C.C. (Yates)

Confidence Level	N1	N2	N	Target Width	Actual Width	P1	P2	P1 - P2	Lower Limit	Upper Limit
0.950	10	10	20	0.8395	0.8395	0.9000	0.3000	0.6000	0.1605	1.0000

PASS also calculates the necessary sample size to be 10 per group.

Confidence Intervals for the Difference Between Two Proportions**Score (Miettinen & Nurminen)**

Confidence Level	N1	N2	N	Target Width	Actual Width	P1	P2	P1 - P2	Lower Limit	Upper Limit
0.950	10	10	20	0.6706	0.6706	0.9000	0.3000	0.6000	0.1700	0.8406

PASS also calculates the necessary sample size to be 10 per group.

Score (Wilson)

Confidence Level	N1	N2	N	Target Width	Actual Width	P1	P2	P1 - P2	Lower Limit	Upper Limit
0.950	10	10	20	0.6385	0.6385	0.9000	0.3000	0.6000	0.1705	0.8090

PASS also calculates the necessary sample size to be 10 per group.

Score C.C. (Wilson)

Confidence Level	N1	N2	N	Target Width	Actual Width	P1	P2	P1 - P2	Lower Limit	Upper Limit
0.950	10	10	20	0.7374	0.7374	0.9000	0.3000	0.6000	0.1013	0.8387

PASS also calculates the necessary sample size to be 10 per group.

Confidence Intervals for the Difference Between Two Proportions

Example 4 – Validation using Gart and Nam (1990)

Gart and Nam (1990) page 640 give an example of a calculation for a confidence interval for the difference in proportions when the confidence level is 95%, the sample proportions are 0.28 and 0.08, and the interval width is 0.4281 for the Score (Gart and Nam) method. The necessary sample size in each case is 25 per group.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Confidence Intervals for the Difference Between Two Proportions** procedure window by expanding **Proportions**, then **Two Independent Proportions**, then clicking on **Confidence Interval**, and then clicking on **Confidence Intervals for the Difference Between Two Proportions**. You may then make the appropriate entries as listed below, or open **Example 4** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Sample Size
Confidence Interval Formula	Score w/Skewness (Gart & Nam)
Interval Type	Two-Sided
Confidence Level	0.95
Group Allocation	Equal (N1 = N2)
Confidence Interval Width (Two-Sided)	0.4281
Input Type	Proportions
P1	0.28
P2	0.08

Output

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

Numeric Results

Confidence Level	N1	N2	N	Target Width	Actual Width	P1	P2	P1 - P2	Lower Limit	Upper Limit
0.950	25	25	50	0.4281	0.4281	0.2800	0.0800	0.2000	-0.0143	0.4137

PASS also calculates the necessary sample size to be 25 per group.