Chapter 516

Two Proportions – Non-Inferiority, Superiority, Equivalence, and Two-Sided Tests vs a Margin

Introduction

This chapter documents four closely related procedures: non-inferiority tests, superiority (by a margin) tests, equivalence tests, and two-sided tests versus a margin. These procedures compute both asymptotic and exact confidence intervals and hypothesis tests for the difference, ratio, and odds ratio of two proportions.

Non-Inferiority Tests

Non-inferiority tests are one-sided hypothesis tests in which the null and alternative hypotheses are arranged to test whether one group is almost as good (not much worse) than the other group. So, we might be interested in showing that a new, less-expensive method of treatment is no worse than the current method of treatment.

Superiority Tests

Superiority tests are one-sided hypothesis tests in which the null and alternative hypotheses are arranged to test whether one group is better than the other group by more than a stated margin. So, we might be interested in showing that a new method of treatment is better by more than a clinically insignificant amount.

Equivalence Tests

Equivalence tests are used to show conclusively that two methods (i.e., drugs) are equivalent.

The conventional method of testing equivalence hypotheses is to perform two, one-sided tests (TOST) of hypotheses. The null hypothesis of non-equivalence is rejected in favor of the alternative hypothesis of equivalence if both one-sided tests are rejected. Unlike the common two-sided tests, however, the type I error rate is set directly at the nominal level (usually 0.05)—it is not split in half. So, to perform the test, two, one-sided tests are conducted at the significance level α. If both are rejected, the alternative hypothesis is concluded at the significance level α. Note that the p-value of the test is the maximum of the p-values of the two tests.

Two-Sided Tests versus a Margin

These are two-sided tests in which the null hypotheses is that the two proportions differ by given, non-zero amount called the margin.

Comparing Two Proportions

In recent decades, a number of notation systems have been used to present the results of a study for comparing two proportions. For the purposes of the technical details of this chapter, we will use the following notation:

	Event	Non-Event	Total
Group 1	x_{11}	x_{12}	n_1
Group 2	x_{21}	x_{22}	n_2
Totals	m_1	m_2	N

In this table, the label *Event* is used, but might instead be *Success*, *Attribute of Interest, Positive Response, Disease*, *Yes*, or something else.

The binomial proportions P_1 and P_2 are estimated from the data using the formulae

$$p_1 = \frac{x_{11}}{n_1} \quad \text{and} \quad p_2 = \frac{x_{21}}{n_2}$$

Three common comparison parameters of two proportions are the proportion difference, proportion (risk) ratio, and the odds ratio:

Parameter	**Notation**
Difference	$\Delta = P_1 - P_2$
Risk Ratio	$\phi = P_1 \,/\, P_2$
Odds Ratio	$\psi = \frac{O_1}{O_2} = \frac{P_1/(1-P_1)}{P_2/(1-P_2)}$

Although these three parameters are (non-linear) functions of each other, the choice of which is to be used should not be taken lightly. The associated tests and confidence intervals of each of these parameters can vary widely in power and coverage probability.

Difference

The proportion (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 is little interest in a treatment that decreases the probability from 0.400 to 0.399. However, if the baseline probably of a disease was 0.002 and 0.001 was the decrease in the disease probability, this would represent a reduction of 50%. Thus, we see that interpretation depends on the baseline probability of the event.

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

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

Ratio

The proportion (risk) ratio $\phi = p_1 \,/\, p_2$ gives the relative change in risk in a treatment group (group 1) compared to a control group (group 2). This parameter is also direct and easy to interpret. To compare this with the difference, consider a treatment that reduces the risk of disease from 0.1437 to 0.0793. Which single number is most enlightening, the fact that the absolute risk of disease has been decreased by 0.0644, or the fact that risk of disease in the treatment group is only 55.18% of that in the control group? In many cases, the percentage (100 x risk ratio) communicates the impact of the treatment better than the absolute change.

Perhaps the biggest drawback of this parameter is that it cannot be calculated in one of the most common experimental designs: the case-control study. Another drawback, when compared to the odds ratio, is that the odds ratio occurs naturally in the likelihood equations and as a parameter in logistic regression, while the proportion ratio does not.

Odds Ratio

Chances are usually communicated as long-term proportions or probabilities. In betting, chances are often given as odds. For example, the odds of a horse winning a race might be set at 10-to-1 or 3-to-2. How do you translate from odds to probability? An odds of 3-to-2 means that the event will occur three out of five times. That is, an odds of 3-to-2 (1.5) translates to a probability of winning of 0.60.

The odds of an event are calculated by dividing the event risk by the non-event risk. Thus, in our case of two populations, the odds are

$$O_1 = \frac{P_1}{1 - P_1} \text{ and } O_2 = \frac{P_2}{1 - P_2}$$

For example, if P_1 is 0.60, the odds are 0.60/0.40 = 1.5. In some cases, rather than representing the odds as a decimal amount, it is re-scaled into whole numbers. Thus, instead of saying the odds are 1.5-to-1, we may equivalently say they are 3-to-2.

In this context, the comparison of proportions may be done by comparing the odds through the ratio of the odds. The odds ratio of two events is

$$\psi = \frac{O_1}{O_2}$$

$$= \frac{\frac{P_1}{1-P_1}}{\frac{P_2}{1-P_2}}$$

Until one is accustomed to working with odds, the odds ratio is usually more difficult to interpret than the proportion (risk) ratio, but it is still the parameter of choice for many researchers. Reasons for this include the fact that the odds ratio can be accurately estimated from case-control studies, while the risk ratio cannot. Also, the odds ratio is the basis of logistic regression (used to study the influence of risk factors). Furthermore, the odds ratio is the natural parameter in the conditional likelihood of the two-group, binomial-response design. Finally, when the baseline event-rates are rare, the odds ratio provides a close approximation to the risk ratio since, in this case, $1 - P_1 \approx 1 - P_2$, so that

$$\psi = \frac{\frac{P_1}{1-P_1}}{\frac{P_2}{1-P_2}} \approx \frac{P_1}{P_2} = \phi$$

One benefit of the log of the odds ratio is its desirable statistical properties, such as its continuous range from negative infinity to positive infinity.

Confidence Intervals

Both large sample and exact confidence intervals may be computed for the difference, the ratio, and the odds ratio.

Confidence Intervals for the Difference

Several methods are available for computing a confidence interval of the difference between two proportions $\delta = P_1 - P_2$. Newcombe (1998) 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.

Cells with Zero Counts

Extreme cases in which some cells are zero require special approaches with some of the tests given below. We have found that a simple solution that works well is to change the zeros to a small positive number such as 0.01. This produces the same results as other techniques of which we are aware.

C.I. for Difference: Wald Z with Continuity Correction

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

$$L = p_1 - p_2 - z\left[\sqrt{\left(\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}\right)} + \frac{1}{2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right]$$

$$U = p_1 - p_2 + z\left[\sqrt{\left(\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}\right)} + \frac{1}{2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right]$$

C.I. for Difference: Wald Z

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

$$L = p_1 - p_2 - z\sqrt{\left(\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}\right)}$$

$$U = p_1 - p_2 + z\sqrt{\left(\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}\right)}$$

C.I. for Difference: Wilson's Score as modified by Newcombe

For details, see Newcombe (1998), page 876

$$L = p_1 - \hat{p}_2 - B$$

$$U = p_1 - \hat{p}_2 + C$$

where

$$B = z\sqrt{\frac{l_1(1-l_1)}{n_1} + \frac{u_2(1-u_2)}{n_2}}$$

$$C = z\sqrt{\frac{u_1(1-u_1)}{n_1} + \frac{l_2(1-l_2)}{n_2}}$$

and l_1 and u_1 are the roots of

$$|P_1 - p_1| - z\sqrt{\frac{P_1(1-P_1)}{n_1}} = 0$$

and l_2 and u_2 are the roots of

$$|P_2 - p_2| - z\sqrt{\frac{P_2(1-P_2)}{n_2}} = 0$$

C.I. for Difference: Miettinen-Nurminen Score

Miettinen and Nurminen (1985) proposed a test statistic for testing whether the odds ratio is equal to a specified value ψ_0. Because the approach they used with the difference and ratio does not easily extend to the odds ratio, they used a score statistic approach for the odds ratio. The regular MLE's are $\hat{p}_1$ and $\hat{p}_2$. The constrained MLE's are $\tilde{p}_1$ and $\tilde{p}_2$, These estimates are constrained so that $\tilde{\psi} = \psi_0$. 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 the test statistic is

$$z_{MNO} = \frac{\frac{(\hat{p}_1 - \tilde{p}_1)}{\tilde{p}_1\tilde{q}_1} - \frac{(\hat{p}_2 - \tilde{p}_2)}{\tilde{p}_2\tilde{q}_2}}{\sqrt{\left(\frac{1}{N_2\tilde{p}_1\tilde{q}_1} + \frac{1}{N_2\tilde{p}_2\tilde{q}_2}\right)\left(\frac{N}{N-1}\right)}}$$

where

$$\tilde{p}_1 = \frac{\tilde{p}_2\psi_0}{1 + \tilde{p}_2(\psi_0 - 1)}$$

$$\tilde{p}_2 = \frac{-B + \sqrt{B^2 - 4AC}}{2A}$$

$$A = N_2(\psi_0 - 1)$$

$$B = N_1\psi_0 + N_2 - M_1(\psi_0 - 1)$$

$$C = -M_1$$

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

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

C.I. for Difference: Farrington-Manning 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 formula for computing the test is

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

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

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

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

C.I. for Difference: Gart-Nam Score

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

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

where

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

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

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

C.I. for Difference: Chen's Quasi-Exact Method

Chen (2002) proposed a quasi-exact method for generating confidence intervals. This method produces intervals that are close to unconditional exact intervals that are available in specialized software like StatXact, but do not require as much time to compute. Chen's method inverts a hypothesis test based on Farrington and Manning's method. That is, the confidence interval is found by finding those values at which the hypothesis test that the difference is a given, non-zero value become significant. However, instead of searching for the maximum significance level of all possible values of the nuisance parameter as the exact tests do, Chen proposed using the significance level at the constrained maximum likelihood estimate of p_2 as given by Farrington and Manning. This simplification results in a huge reduction in computation with only a minor reduction in accuracy. Also, it allows much larger sample sizes to be analyzed.

Note on Exact Methods

A word of caution should be raised about the phrase *exact tests* or *exact confidence intervals*. Many users assume that methods that are based on exact methods are always better than other, non-exact methods. After all, 'exact' sounds better than 'approximate'. However, tests and confidence intervals based on exact methods are not necessarily better. In fact, some prominent statisticians are of the opinion that they are actually worse (see Agresti and Coull (1998) for one example). *Exact* simply means that they are based on exact distributional calculations. They may be, however, conservative in terms of their coverage probabilities (the probability that the confidence interval includes the true value). That is, they are wider than they need to be because they are based on worst case scenarios.

Confidence Intervals for the Ratio

C.I. for Ratio: Miettinen-Nurminen Score

Miettinen and Nurminen (1985) proposed a test statistic for testing whether the ratio is equal to a specified value ϕ_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 $\tilde{p}_1 / \tilde{p}_2 = \phi_0$ 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.

Here is the formula for computing the test

$$z_{MNR} = \frac{\hat{p}_1 / \hat{p}_2 - \phi_0}{\sqrt{\left(\frac{\tilde{p}_1\tilde{q}_1}{n_1} + \phi_0^2\frac{\tilde{p}_2\tilde{q}_2}{n_2}\right)\left(\frac{N}{N-1}\right)}}$$

where

$$\tilde{p}_1 = \tilde{p}_2\phi_0$$

$$\tilde{p}_2 = \frac{-B - \sqrt{B^2 - 4AC}}{2A}$$

$$A = N\phi_0$$

$$B = -[N_1\phi_0 + x_{11} + N_2 + x_{21}\phi_0]$$

$$C = M_1$$

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

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

and the upper limit is the solution of

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

C.I. for Ratio: Farrington-Manning Score

Farrington and Manning (1990) proposed a test statistic for testing whether the ratio is equal to a specified value ϕ_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 $\tilde{p}_1 / \tilde{p}_2 = \phi_0$ are used in the denominator. A correction factor of N/(N-1) is applied to increase the variance estimate. The significance level of the test statistic is based on the asymptotic normality of the score statistic.

Here is the formula for computing the test

$$z_{FMR} = \frac{\hat{p}_1/\hat{p}_2 - \phi_0}{\sqrt{\left(\frac{\tilde{p}_1\tilde{q}_1}{n_1} + \phi_0^2\frac{\tilde{p}_2\tilde{q}_2}{n_2}\right)}}$$

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

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

$$z_{FMR} = \left|z_{\alpha/2}\right|$$

and the upper limit is the solution of

$$z_{FMR} = -\left|z_{\alpha/2}\right|$$

C.I. for Ratio: Gart-Nam Score

Gart and Nam (1988) page 329 proposed a modification to the Farrington and Manning (1988) ratio test that corrected for skewness. Let $z_{FMR}(\phi)$ stand for the Farrington and Manning ratio test statistic described above. The skewness corrected test statistic z_{GNR} is the appropriate solution to the quadratic equation

$$(-\tilde{\varphi})z_{GNR}^2 + (-1)z_{GNR} + (z_{FMR}(\phi) + \tilde{\varphi}) = 0$$

where

$$\tilde{\varphi} = \frac{1}{6\tilde{u}^{3/2}}\left(\frac{\tilde{q}_1(\tilde{q}_1 - \tilde{p}_1)}{n_1^2\tilde{p}_1^2} - \frac{\tilde{q}_2(\tilde{q}_2 - \tilde{p}_2)}{n_2^2\tilde{p}_2^2}\right)$$

$$\tilde{u} = \frac{\tilde{q}_1}{n_1\tilde{p}_1} + \frac{\tilde{q}_2}{n_2\tilde{p}_2}$$

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

$$z_{GNR} = \left|z_{\alpha/2}\right|$$

and the upper limit is the solution of

$$z_{GNR} = -\left|z_{\alpha/2}\right|$$

C.I. for Ratio: Katz Logarithm

This was one of the first methods proposed for computing confidence intervals for risk ratios.

For details, see Gart and Nam (1988), page 324.

$$L = \hat{\phi} \exp\left(-z\sqrt{\frac{q_1}{np_1} + \frac{q_2}{np_2}}\right)$$

$$U = \hat{\phi} \exp\left(z\sqrt{\frac{q_1}{np_1} + \frac{q_2}{np_2}}\right)$$

where

$$\hat{\phi} = \frac{p_1}{p_2}$$

C.I. for Ratio: Walters Logarithm + 1/2

For details, see Gart and Nam (1988), page 324.

$$L = \hat{\phi} \exp\left(-z\sqrt{\hat{u}}\right)$$

$$U = \hat{\phi} \exp\left(z\sqrt{\hat{u}}\right)$$

where

$$\hat{\phi} = \exp\left(\ln\left(\frac{x_{11} + \frac{1}{2}}{n_1 + \frac{1}{2}}\right) - \ln\left(\frac{x_{21} + \frac{1}{2}}{n_2 + \frac{1}{2}}\right)\right)$$

$$\hat{u} = \frac{1}{x_{11} + \frac{1}{2}} - \frac{1}{n_1 + \frac{1}{2}} + \frac{1}{x_{21} + \frac{1}{2}} - \frac{1}{n_2 + \frac{1}{2}}$$

$$\tilde{q}_2 = 1 - \tilde{p}_2$$

$$V = \left(\phi^2\left(\frac{\tilde{q}_1}{n_1\tilde{p}_1} + \frac{\tilde{q}_2}{n_2\tilde{p}_2}\right)\right)^{-1}$$

$$\tilde{p}_1 = \phi\tilde{p}_2$$

$$\tilde{q}_1 = 1 - \tilde{p}_1$$

$$\tilde{q}_2 = 1 - \tilde{p}_2$$

$$\tilde{\mu}_3 = v^{3/2}\left(\frac{\tilde{q}_1(\tilde{q}_1 - \tilde{p}_1)}{(n_1\tilde{p}_1)^2} - \frac{\tilde{q}_2(\tilde{q}_2 - \tilde{p}_2)}{(n_2\tilde{p}_2)^2}\right)$$

$$v = \left(\frac{\tilde{q}_1}{n_1\tilde{p}_1} + \frac{\tilde{q}_2}{n_2\tilde{p}_2}\right)^{-1}$$

C.I. for Ratio: Chen's Quasi-Exact Method

Chen (2002) proposed a quasi-exact method for generating confidence intervals. This method produces intervals that are close to unconditional exact intervals that are available in specialized software like StatXact, but do not require as much time to compute. Chen's method inverts a hypothesis test based on Farrington and Manning's method. That is, the confidence interval is found by finding those values at which the hypothesis test that the difference is a given, non-zero value become significant. However, instead of searching for the maximum significance level of all possible values of the nuisance parameter as the exact tests do, Chen proposed using the significance level at the constrained maximum likelihood estimate of p_2 as given by Farrington and Manning. This simplification results in a huge reduction in computation with only a minor reduction in accuracy. Also, it allows much larger sample sizes to be analyzed.

Confidence Intervals for the Odds Ratio

The odds ratio is a commonly used measure of treatment effect when comparing two binomial proportions. It is the ratio of the odds of the event in group one divided by the odds of the event in group two. The results given below are found in Fleiss (1981).

Symbolically, the odds ratio is defined as

$$\psi = \frac{\frac{P_1}{1 - P_1}}{\frac{P_2}{1 - P_2}}$$

C.I. for Odds Ratio: Simple Technique

The simple estimate of the odds ratio uses the formula

$$\hat{\psi} = \frac{\frac{p_1}{1 - p_1}}{\frac{p_2}{1 - p_2}} = \frac{x_{11}x_{22}}{x_{21}x_{12}}$$

The standard error of this estimator is estimated by

$$se(\hat{\psi}) = \hat{\psi}\sqrt{\frac{1}{x_{11}} + \frac{1}{x_{21}} + \frac{1}{x_{12}} + \frac{1}{x_{22}}}$$

Problems occur if any one of the quantities x_{11}, x_{21}, x_{12}, or x_{22} are zero. To correct this problem, many authors recommend adding one-half to each cell count so that a zero cannot occur. Now, the formulas become

$$\hat{\psi}' = \frac{(x_{11} + 0.5)(x_{22} + 0.5)}{(x_{21} + 0.5)(x_{12} + 0.5)}$$

and

$$se(\hat{\psi}') = \hat{\psi}' \sqrt{\frac{1}{x_{11} + 0.5} + \frac{1}{x_{21} + 0.5} + \frac{1}{x_{12} + 0.5} + \frac{1}{x_{22} + 0.5}}$$

The distribution of these direct estimates of the odds ratio do not converge to normality as fast as does their logarithm, so the logarithm of the odds ratio is used to form confidence intervals. The formula for the standard error of the log odds ratio is

$$L' = \ln(\hat{\psi}')$$

and

$$se(L') = \sqrt{\frac{1}{x_{11} + 0.5} + \frac{1}{x_{21} + 0.5} + \frac{1}{x_{12} + 0.5} + \frac{1}{x_{22} + 0.5}}$$

A $100(1 - \alpha)\%$ confidence interval for the log odds ratio is formed using the standard normal distribution as follows

$$\hat{\psi}_{lower} = \exp\left(L' - z_{1-\alpha/2} se(L')\right)$$

$$\hat{\psi}_{upper} = \exp\left(L' + z_{1-\alpha/2} se(L')\right)$$

C.I. for Odds Ratio: Iterated Method of Fleiss

Fleiss (1981) presents an improved confidence interval for the odds ratio. This method forms the confidence interval as all those values of the odds ratio which would not be rejected by a chi-square hypothesis test. Fleiss gives the following details about how to construct this confidence interval. To compute the lower limit, do the following.

1. For a trial value of ψ, compute the quantities *X, Y, W, F, U,* and *V* using the formulas

$$X = \psi(n_1 + m_1) + (n_2 - m_1)$$

$$Y = \sqrt{X^2 - 4n_1 m_1 \psi(\psi - 1)}$$

$$W = \frac{1}{A} + \frac{1}{B} + \frac{1}{C} + \frac{1}{D}$$

$$F = \left(x_{11} - A - \tfrac{1}{2}\right)^2 W - z_{\alpha/2}^2$$

$$U = \frac{1}{B^2} + \frac{1}{C^2} - \frac{1}{A^2} - \frac{1}{D^2}$$

$$V = T\left[\left(x_{11} - A - \tfrac{1}{2}\right)^2 U - 2W\left(x_{11} - A - \tfrac{1}{2}\right)\right]$$

where

$$A = \frac{X - Y}{2(\psi - 1)}$$

$$B = m_1 - A$$

$$C = n_1 - A$$

$$D = f - n_1 + A$$

$$T = \frac{1}{2(\psi - 1)^2}\left(Y - N - \frac{\psi - 1}{Y}[X(n_1 + m_1) - 2n_1 m_1(2\psi - 1)]\right)$$

Finally, use the updating equation below to calculate a new value for the odds ratio using the updating equation

$$\psi^{(k+1)} = \psi^{(k)} - \frac{F}{V}$$

2. Continue iterating until the value of *F* is arbitrarily close to zero.

The upper limit is found by substituting $+\frac{1}{2}$ for $-\frac{1}{2}$ in the formulas for *F* and *V*.

Confidence limits for the *relative risk* can be calculated using the expected counts *A, B, C,* and *D* from the last iteration of the above procedure. The lower limit of the relative risk

$$\phi_{lower} = \frac{A_{lower} n_2}{B_{lower} n_1}$$

$$\phi_{upper} = \frac{A_{upper} n_2}{B_{upper} n_1}$$

C.I. for Odds Ratio: Mantel-Haenszel

The common estimate of the logarithm of the odds ratio is used to create this estimator. That is

$$\ln(\hat{\psi}) = \ln\left(\frac{x_{11}x_{22}}{x_{21}x_{12}}\right)$$

The standard error of this estimator is estimated using the Robins, Breslow, Greenland (1986) estimator which performs well in most situations. The standard error is given by

$$se(\ln(\hat{\psi})) = \sqrt{\frac{A}{2C} + \frac{AD + BC}{2CD} + \frac{B}{2D}}$$

where

$$A = x_{11} + x_{22}$$

$$B = x_{21} + x_{12}$$

$$C = x_{11}x_{22}$$

$$D = x_{21}x_{12}$$

The confidence limits are calculated as

$$\hat{\psi}_{lower} = \exp\left(\ln(\hat{\psi}) - z_{1-\alpha/2} se(\ln(\hat{\psi}))\right)$$

$$\hat{\psi}_{upper} = \exp\left(\ln(\hat{\psi}) + z_{1-\alpha/2} se(\ln(\hat{\psi}))\right)$$

C.I. for Odds Ratio: Miettinen-Nurminen Score

Miettinen and Nurminen (1985) proposed a test statistic for testing whether the odds ratio is equal to a specified value ψ_0. Because the approach they used with the difference and ratio does not easily extend to the odds ratio, they used a score statistic approach for the odds ratio. The regular MLE's are $\hat{p}_1$ and $\hat{p}_2$. The constrained MLE's are $\tilde{p}_1$ and $\tilde{p}_2$. These estimates are constrained so that $\tilde{\psi} = \psi_0$. 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 the test statistic is

$$z_{MNO} = \frac{\frac{(\hat{p}_1 - \tilde{p}_1)}{\tilde{p}_1\tilde{q}_1} - \frac{(\hat{p}_2 - \tilde{p}_2)}{\tilde{p}_2\tilde{q}_2}}{\sqrt{\left(\frac{1}{N_2\tilde{p}_1\tilde{q}_1} + \frac{1}{N_2\tilde{p}_2\tilde{q}_2}\right)\left(\frac{N}{N-1}\right)}}$$

where

$$\tilde{p}_1 = \frac{\tilde{p}_2\psi_0}{1 + \tilde{p}_2(\psi_0 - 1)}$$

$$\tilde{p}_2 = \frac{-B + \sqrt{B^2 - 4AC}}{2A}$$

$$A = N_2(\psi_0 - 1)$$

$$B = N_1\psi_0 + N_2 - M_1(\psi_0 - 1)$$

$$C = -M_1$$

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

$$z_{MNO} = \left|z_{\alpha/2}\right|$$

and the upper limit is the solution of

$$z_{MNO} = -\left|z_{\frac{\alpha}{2}}\right|$$

C.I. for Odds Ratio: Farrington-Manning Score

Farrington and Manning (1990) indicate that the Miettinen and Nurminen statistic may be modified by removing the factor *N*/(*N*-1).

The formula for computing this test statistic is

$$z_{FMO} = \frac{\frac{(\hat{p}_1 - \tilde{p}_1)}{\hat{p}_1\hat{q}_1} - \frac{(\hat{p}_2 - \tilde{p}_2)}{\hat{p}_2\hat{q}_2}}{\sqrt{\left(\frac{1}{N_2\tilde{p}_1\tilde{q}_1} + \frac{1}{N_2\tilde{p}_2\tilde{q}_2}\right)}}$$

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

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

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

and the upper limit is the solution of

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

C.I. for Odds Ratio: Conditional Exact

The conditional exact confidence interval of the odds ratio is calculated using the noncentral hypergeometric distribution as given in Sahai and Khurshid (1995). That is, a $100(1-\alpha)\%$ confidence interval is found by searching for ψ_L and ψ_U such that

$$\frac{\sum_{k=x}^{k_2}\binom{n_1}{k}\binom{n_2}{m_1-k}(\psi_L)^k}{\sum_{k=k_1}^{k_2}\binom{n_1}{k}\binom{n_2}{m_1-k}(\psi_L)^k} = \frac{\alpha}{2}$$

$$\frac{\sum_{k=k_1}^{x}\binom{n_1}{k}\binom{n_2}{m_1-k}(\psi_U)^k}{\sum_{k=k_1}^{k_2}\binom{n_1}{k}\binom{n_2}{m_1-k}(\psi_U)^k} = \frac{\alpha}{2}$$

where

$$k_1 = \max(0, m_1 - n_1) \text{ and } k_2 = \min(n_1, m_1)$$

Hypothesis Tests

A wide variety of statistical tests are available for testing hypotheses about two proportions. Some tests are based on the *difference* in proportions, others are based on the *ratio* of proportions, and still others are based on the *odds ratio*. Some tests are *conditional*, while others are *unconditional*. Some tests are said to be *large sample*, while others are said to be *exact*. In this section, some of these distinctions are explained.

Types of Hypothesis Tests

Hypothesis tests concerning two proportions can be separated into three categories: large sample, conditional exact, and unconditional exact.

Large Sample Tests

Large sample (or asymptotic) tests are based on the central limit theorem (CLT) which states that for large samples, the distribution of many of these test statistics approach the normal distribution. Hence, significance levels can be computed using the normal distribution which has been extensively tabulated and can now be easily computed.

A difficult determination when deciding whether to use a large sample test is whether or not the sample is large enough for the CLT to properly take effect.

Exact Tests in General

Because of the inaccuracy of applying a large sample procedure to a small sample study, another class of tests has been devised called *exact tests*. The significance levels of these tests are calculated from their exact distribution, usually by considering either the binomial or the hypergeometric distribution. No appeal is made to the CLT. Because these tests are computationally intensive, they have increased in popularity with increase of the computational abilities of computers.

Even with the availability of modern computers, approximate large sample techniques cannot be abandoned completely in favor of exact tests, due to the assumptions required by the exact tests.

The distribution of the proportions in a 2-by-2 table involves two parameters: p_1 and $p_1 + \delta$ in the case of the difference and p_1 and p_1 / ϕ in the case of the ratio. The hypothesis only involves one parameter, the difference, or the ratio. The other parameter, p_1, is called a *nuisance parameter* because it is not part of the hypothesis of interest. That is, the hypothesis that $\delta = 0$ or $\phi = 1$ does not involve p_1. In order to test hypotheses about the parameter of interest, the nuisance parameter must be eliminated. This may be accomplished either by conditional methods or unconditional methods.

Conditional Exact Test

The nuisance parameter can be eliminated by conditioning on a sufficient statistic. Fisher's exact test is an example of this. The conditioning occurs by considering only those tables in which the row and column totals remain the same as for the data. This removes the nuisance parameter p_1 from the distribution formula. This has drawn criticism because most experimental designs do not fix both the row and column totals. Others have argued that since the significance level is preserved unconditionally, the test is valid.

Unconditional Exact Test

The unconditional exact test approach is to remove the nuisance parameter by computing the significance level at all possible values of the nuisance parameter and choosing the largest (worst case). That is, find the value of p_1 which gives the maximum significance level (least significant) for the hypothesis test. That is, these tests find an upper bound for the significance level.

The problem with the unconditional approach is that the upper bound may occur at a value of p_1 that is far from the true value. For example, suppose the true value of p_1 is 0.711 where the significance level is 0.032. However, suppose the maximum significance level of 0.213 occurs at p_1= 0.148. Hence, near the actual value of the nuisance value, the results are statistically significant, but the results of the exact test are not! Of course, in a particular study, we do not know the true value of the nuisance parameter. The message is that although these tests are called exact tests, they are not! They are approximate tests computed using exact distributions. Hence, one cannot say broadly that exact tests are always better than the large-sample test counterparts.

Non-Inferiority Tests

Non-inferiority tests are one-sided hypothesis tests in which the null and alternative hypotheses are arranged to test whether one group is almost as good (not much worse) than the other group. So, we might be interested in showing that a new, less-expensive method of treatment is no worse than the current method of treatment.

The statistical hypotheses used are as follows:

Differences

Assume that P1 and P2 are the event proportions of an experimental treatment and a control treatment, respectively. NIM is the non-inferiority margin.

If Lower Proportions are Worse

H0: P1 – P2 ≤ NIM Ha: P1 – P2 > NIM NIM < 0

NIM is the smallest amount that the new treatment's proportion is less than the current treatment's proportion and we still are willing to consider it not inferior.

If Lower Proportions are Better

H0: P1 – P2 ≥ NIM Ha: P1 – P2 < NIM NIM > 0

NIM is the largest amount that the new treatment's proportion is greater than the current treatment's proportion and we still are willing to consider it not inferior.

Ratios

Assume that P1 and P2 are the event proportions of an experimental treatment and a control treatment, respectively. NIR is the non-inferiority ratio.

If Lower Proportions are Worse

H0: P1 / P2 ≤ NIR Ha: P1 / P2 > NIR NIR < 1

NIR is the smallest that the proportion ratio can be and we still are willing to consider the experimental treatment not inferior to the control treatment.

If Lower Proportions are Better

H0: P1 / P2 ≥ NIR Ha: P1 / P2 < NIR NIR > 1

NIR is the largest that the proportion ratio can be and we still are willing to consider the experimental treatment not inferior to the control treatment.

Odds Ratios

Assume that O1 and O2 are the event odds of an experimental treatment and a control treatment, respectively. NIO is the non-inferiority odds ratio.

If Lower Proportions are Worse

H0: O1 / O2 ≤ NIO Ha: O1 / O2 > NIO NIO < 1

NIO is the smallest that the odds ratio can be and we still are willing to consider the experimental treatment not inferior to the control treatment.

If Lower Proportions are Better

H0: O1 / O2 ≥ NIO Ha: O1 / O2 < NIO NIO > 1

NIO is the largest that the odds ratio can be and we still are willing to consider the experimental treatment not inferior to the control treatment.

Superiority Tests

Superiority tests are one-sided hypothesis tests in which the null and alternative hypotheses are arranged to test whether one group is better than the other group by more than a stated margin. So, we might be interested in showing that a new method of treatment is better by more than a clinically insignificant amount.

The statistical hypotheses used are as follows:

Differences

Assume that P1 and P2 are the event proportions of an experimental treatment and a control treatment, respectively. SM is the superiority margin.

If Lower Proportions are Worse

H0: $P1 - P2 \leq SM$ Ha: $P1 - P2 > SM$ $SM > 0$

NIM is the smallest amount that the new treatment's proportion is greater than the current treatment's proportion and we still are willing to consider it superior.

If Lower Proportions are Better

H0: $P1 - P2 \geq SM$ Ha: $P1 - P2 < SM$ $SM < 0$

NIM is the largest amount that the new treatment's proportion is less than the current treatment's proportion and we still are willing to consider it superior.

Ratios

Assume that P1 and P2 are the event proportions of an experimental treatment and a control treatment, respectively. SR is the superiority ratio.

If Lower Proportions are Worse

H0: $P1 / P2 \leq SR$ Ha: $P1 / P2 > SR$ $SR > 1$

SR is the smallest that the proportion ratio can be and we still are willing to consider the experimental treatment superior to the control treatment.

If Lower Proportions are Better

H0: $P1 / P2 \geq SR$ Ha: $P1 / P2 < SR$ $SR < 1$

SR is the largest that the proportion ratio can be and we still are willing to consider the experimental treatment superior to the control treatment.

Odds Ratios

Assume that O1 and O2 are the event odds an experimental treatment and a control treatment, respectively. SOR is the superior odds ratio.

If Lower Proportions are Worse

H0: O1 / O2 ≤ SOR Ha: O1 / O2 > SOR SOR > 1

SOR is the smallest that the odds ratio can be and we still are willing to consider the experimental treatment superior to the control treatment.

If Lower Proportions are Better

H0: O1 / O2 ≥ SOR Ha: O1 / O2 < SOR SOR < 1

SOR is the largest that the odds ratio can be and we still are willing to consider the experimental treatment superior to the control treatment.

Equivalence Tests

Equivalence tests are used to show conclusively that two methods (i.e. drugs) are equivalent.

The conventional method of testing equivalence hypotheses is to perform two, one-sided tests (TOST) of hypotheses. The null hypothesis of non-equivalence is rejected in favor of the alternative hypothesis of equivalence if both one-sided tests are rejected. Unlike the common two-sided tests, however, the type I error rate is set directly at the nominal level (usually 0.05)—it is not split in half. So, to perform the test, two, one-sided tests are conducted at the significance level α. If both are rejected, the alternative hypothesis is concluded at the significance level α. Note that the p-value of the test is the maximum of the p-values of the two tests.

Differences

Assume that P1 and P2 are the event proportions of an experimental treatment and a control treatment, respectively.

H0: P1 - P2 ≤ LM or P1 - P2 ≥ UM Ha: LM < P1 - P2 < UM (Equivalence) LM < 0, UM > 0

LM and UM are the lower and upper margins of equivalence, respectively. If the proportion difference is between these bounds, we are willing to say that the two treatments are equivalent.

Ratios

Assume that P1 and P2 are the event proportions of an experimental treatment and a control treatment, respectively.

H0: P1 / P2 ≤ LR or P1 / P2 ≥ UR Ha: LR < P1 / P2 < UR (Equivalence) LR < 1, UR > 1

LR and UR are the lower and upper margins of equivalence, respectively. If the proportion ratio is between these bounds, we are willing to say that the two treatments are equivalent.

Odds Ratios

Assume that O1 and O2 are the event odds an experimental treatment and a control treatment, respectively.

H0: O1 / O2 ≤ LO or O1 / O2 ≥ UO Ha: LO < O1 / O2 < UO (Equivalence) LO < 1, UO > 1

LO and UO are the lower and upper margins of equivalence, respectively. If the odds ratio is between these bounds, we are willing to say that the two treatments are equivalent.

Two-Sided Tests versus a Margin

These are two-sided tests in which the null hypotheses is that the two proportions differ by given, non-zero amount called the margin.

The statistical hypotheses used are as follows:

Differences

Assume that P1 and P2 are the event proportions of an experimental treatment and a control treatment, respectively. DM is the hypothesized difference margin.

H0: P1 – P2 = DM Ha: P1 – P2 ≠ DM DM ≠ 0

Ratios

Assume that P1 and P2 are the event proportions of an experimental treatment and a control treatment, respectively. RV is the hypothesize value of P1 / P2.

H0: P1 / P2 = RV Ha: P1 / P2 ≠ RV RV ≠ 1

Odds Ratios

Assume that O1 and O2 are the event odds of an experimental treatment and a control treatment, respectively. ORV is the hypothesize value of O1 / O2.

H0: P1 / P2 = ORV Ha: P1 / P2 ≠ ORV ORV ≠ 1

Hypothesis Test Technical Details

The sections that follow give formulaic details of the hypothesis tests associated with this procedure.

Notation for Hypothesis Test Statistics

The following notation is used in the formulas for the test statistics.

	Event	Non-Event	Total	Sample Proportion
Group 1	x_{11}	x_{12}	n_1	$p_1 = x_{11}/n_1$
Group 2	x_{21}	x_{22}	n_2	$p_2 = x_{21}/n_2$
Total	m_1	m_2	N	$p = m_1/N$

Non-Inferiority, Superiority, Inequality Margin, and Equivalence Test Statistics

Non-inferiority tests, superiority tests, equivalence tests, and two-sided inequality tests versus a margin all use one of the three versions of the score test presented next.

Large-Sample (Asymptotic) Tests

The tests in this section are valid for large samples.

Miettinen-Nurminen Large-Sample Score Test of the Difference

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

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

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

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

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

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

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

$$L_1 = [N_2\delta_0 - N - 2x_{21}]\delta_0 + M_1$$

$$L_2 = (N+N_2)\delta_0 - N - M_1$$

$$L_3 = N$$

Miettinen-Nurminen Large-Sample Score Test of the Ratio

Miettinen and Nurminen (1985) proposed a test statistic for testing whether the ratio is equal to a specified value ϕ_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 $\tilde{p}_1 / \tilde{p}_2 = \phi_0$ 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.

Here is the formula for computing the test

$$z_{MNR} = \frac{\hat{p}_1 / \hat{p}_2 - \phi_0}{\sqrt{\left(\frac{\tilde{p}_1\tilde{q}_1}{n_1} + \phi_0^2\frac{\tilde{p}_2\tilde{q}_2}{n_2}\right)\left(\frac{N}{N-1}\right)}}$$

where

$$\tilde{p}_1 = \tilde{p}_2\phi_0$$

$$\tilde{p}_2 = \frac{-B - \sqrt{B^2 - 4AC}}{2A}$$

$$A = N\phi_0$$

$$B = -[N_1\phi_0 + x_{11} + N_2 + x_{21}\phi_0]$$

$$C = M_1$$

Miettinen-Nurminen Large-Sample Score Test of the Odds Ratio

Miettinen and Nurminen (1985) proposed a test statistic for testing whether the odds ratio is equal to a specified valueψ_0. Because the approach they used with the difference and ratio does not easily extend to the odds ratio, they used a score statistic approach for the odds ratio. The regular MLE's are $\hat{p}_1$ and $\hat{p}_2$. The constrained MLE's are $\tilde{p}_1$ and $\tilde{p}_2$, These estimates are constrained so that $\tilde{\psi} = \psi_0$. 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 the test statistic is

$$z_{MNO} = \frac{\frac{(\hat{p}_1 - \tilde{p}_1)}{\tilde{p}_1\tilde{q}_1} - \frac{(\hat{p}_2 - \tilde{p}_2)}{\tilde{p}_2\tilde{q}_2}}{\sqrt{\left(\frac{1}{N_2\tilde{p}_1\tilde{q}_1} + \frac{1}{N_2\tilde{p}_2\tilde{q}_2}\right)\left(\frac{N}{N-1}\right)}}$$

where

$$\tilde{p}_1 = \frac{\tilde{p}_2\psi_0}{1 + \tilde{p}_2(\psi_0 - 1)}$$

$$\tilde{p}_2 = \frac{-B + \sqrt{B^2 - 4AC}}{2A}$$

$$A = N_2(\psi_0 - 1)$$

$$B = N_1\psi_0 + N_2 - M_1(\psi_0 - 1)$$

$$C = -M_1$$

Farrington-Manning Large-Sample Score Test of the Difference

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.

Here is the formula for computing the test

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

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

Farrington-Manning Large-Sample Score Test of the Ratio

Farrington and Manning (1990) proposed a test statistic for testing whether the ratio is equal to a specified value ϕ_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 $\tilde{p}_1 / \tilde{p}_2 = \phi_0$ are used in the denominator. A correction factor of N/(N-1) is applied to increase the variance estimate. The significance level of the test statistic is based on the asymptotic normality of the score statistic.

Here is the formula for computing the test

$$z_{FMR} = \frac{\hat{p}_1 / \hat{p}_2 - \phi_0}{\sqrt{\left(\frac{\tilde{p}_1\tilde{q}_1}{n_1} + \phi_0^2\frac{\tilde{p}_2\tilde{q}_2}{n_2}\right)}}$$

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

Farrington-Manning Large-Sample Score Test of the Odds Ratio

Farrington and Manning (1990) indicate that the Miettinen and Nurminen statistic may be modified by removing the factor *N*/(*N*-1).

The formula for computing this test statistic is

$$z_{FMO} = \frac{\frac{(\hat{p}_1 - \tilde{p}_1)}{\tilde{p}_1\tilde{q}_1} - \frac{(\hat{p}_2 - \tilde{p}_2)}{\tilde{p}_2\tilde{q}_2}}{\sqrt{\left(\frac{1}{N_2\tilde{p}_1\tilde{q}_1} + \frac{1}{N_2\tilde{p}_2\tilde{q}_2}\right)}}$$

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

Gart-Nam Large-Sample Score Test of the Difference

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

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

where

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

Gart-Nam Large-Sample Score Test of the Ratio

Gart and Nam (1988) page 329 proposed a modification to the Farrington and Manning (1988) ratio test that corrected for skewness. Let $z_{FMR}(\phi)$ stand for the Farrington and Manning ratio test statistic described above. The skewness corrected test statistic z_{GNR} is the appropriate solution to the quadratic equation

$$(-\tilde{\varphi})z_{GNR}^2 + (-1)z_{GNR} + (z_{FMR}(\phi) + \tilde{\varphi}) = 0$$

where

$$\tilde{\varphi} = \frac{1}{6\tilde{u}^{3/2}}\left(\frac{\tilde{q}_1(\tilde{q}_1 - \tilde{p}_1)}{n_1^2\tilde{p}_1^2} - \frac{\tilde{q}_2(\tilde{q}_2 - \tilde{p}_2)}{n_2^2\tilde{p}_2^2}\right)$$

$$\tilde{u} = \frac{\tilde{q}_1}{n_1\tilde{p}_1} + \frac{\tilde{q}_2}{n_2\tilde{p}_2}$$

General Form of the Exact Tests in NCSS

All of the exact tests follow the same pattern. We will present the general procedure here, and then give the specifics for each test.

Specify the Null and Alternative Hypotheses

The first step is to select a method to compare the proportions and determine if the test is to be one-, or two-, sided. These may be written in general as

$$H_0: h_j(P_1, P_2) = \theta_0$$
$$H_1: h_j(P_1, P_2) \neq \theta_0$$

where '≠' (for two-sided tests) could be replaced with '<' or '>' for a one-sided test and the index *j* is defined as

$$h_1(P_1, P_2) = P_1 - P_2.$$

$$h_2(P_1, P_2) = P_1/P_2$$

$$h_3(P_1, P_2) = \frac{P_1/(1 - P_1)}{P_2/(1 - P_2)}$$

Specify the Reference Set

The next step is to specify the *reference set* of possible tables to compare the observed table against. Two reference sets are usually considered. Define Ω as the complete set of tables that are possible by selecting n_1 observations from one group and n_2 observations from another group. Define Γ as the subset from Ω for which $x_{11} + x_{21} = m_1$. Tests using Ω are unconditional tests while tests using Γ are conditional tests.

Specify the Test Statistic

The next step is to select the test statistic. In most cases, the score statistic is used which has the general form

$$D(x) = \frac{h_j(p_1, p_2) - \theta_0}{\sqrt{\tilde{V}_{h_j}(\theta_0)}}$$

where *x* represents a table with elements $x_{11}, x_{12}, x_{21}, x_{22}$ and $\tilde{V}_{h_j}(\theta_0)$ is the estimated variance of the score numerator with the constraint that the null hypothesis is true.

Select the Probability Distribution

The probability distribution an unconditional test based on the score statistic is

$$f_{p_1,p_2}(x) = \binom{n_1}{x_{11}}\binom{n_2}{x_{21}} p_1^{x_{11}}(1-p_1)^{x_{12}} p_2^{x_{21}}(1-p_2)^{x_{22}}$$

The probability distribution of a conditional test based on the score statistic is

$$f_{\psi}(x) = \frac{\binom{n_1}{x_{11}}\binom{n_2}{x_{21}}\psi^{x_{11}}}{\sum_{x\in\Gamma}\binom{n_1}{x_{11}}\binom{n_2}{x_{21}}\psi^{x_{11}}}$$

Calculate the Significance Level

The significance level (rejection probability) is found by summing the probabilities of all tables that for which the computed test statistic is at least as favorable to the alternative hypothesis as is the observed table. This may be written as

$$p(y) = \sum_{I\left(D(y),D(x)\right)} f_{p_1,p_2}(x)$$

where $I\left(D(y), D(x)\right)$ is an indicator function.

Maximize the Significance Level

The final step is to find the maximum value (supremum) of the significance level over all possible values of the nuisance parameter. This may be written as

$$p_{supp_2}(x) = \sup_{0<p_2<1}\left(\sum_{I\left(D(y),D(x)\right)} f_{p_1,p_2}(x)\right)$$

Note that the choice of either p_1 or p_2 as the nuisance parameter is arbitrary.

Exact Tests in NCSS

Miettinen-Nurminen Unconditional Exact Test of the Difference

Miettinen and Nurminen (1985) proposed an unconditional exact test for testing whether the difference between two proportions is a specified value δ_0. When $\delta_0 = 0$, this test reduces to Barnard's test. Here are the details of this test:

Null Hypothesis: $p_1 - p_2 = \delta_0$

Hypothesis Types: Both one-sided and two-sided

Reference Set: Ω

Test Statistic: $D(y) = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0}{\sqrt{\left(\frac{\tilde{p}_1\tilde{q}_1}{n_1} + \frac{\tilde{p}_2\tilde{q}_2}{n_2}\right)\left(\frac{N}{N-1}\right)}}$

where $\tilde{p}_1$ and $\tilde{p}_2$ are constrained MLE's discussed below.

Two-Sided Test: $I\big(D(y), D(x)\big) = |D(x)| \geq |D(y)|$

Lower One-Sided Test: $I\big(D(y), D(x)\big) = D(x) \leq D(y)$

Upper One-Sided Test: $I\big(D(y), D(x)\big) = D(x) \geq D(y)$

Farrington-Manning Unconditional Exact Test of the Difference

Farrington and Manning (1990) proposed an unconditional exact test for testing whether the difference is a specified value δ_0. This test was also discussed by Gart and Nam (1990). This test is only slightly different from the test of Miettinen and Nurminen (1985). Here are the details of this test:

Null Hypothesis: $p_1 - p_2 = \delta_0$

Hypothesis Types: Both one-sided and two-sided

Reference Set: Ω

Test Statistic: $D(y) = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0}{\sqrt{\left(\frac{\tilde{p}_1\tilde{q}_1}{n_1} + \frac{\tilde{p}_2\tilde{q}_2}{n_2}\right)}}$

where $\tilde{p}_1$ and $\tilde{p}_2$ are constrained MLE's discussed below.

Two-Sided Test: $I\big(D(y), D(x)\big) = |D(x)| \geq |D(y)|$

Lower One-Sided Test: $I\big(D(y), D(x)\big) = D(x) \leq D(y)$

Upper One-Sided Test: $I\big(D(y), D(x)\big) = D(x) \geq D(y)$

Miettinen-Nurminen Unconditional Exact Test of the Ratio

Miettinen and Nurminen (1985) proposed an unconditional exact test for testing whether the ratio between two proportions is a specified value ϕ_0. When $\phi_0 = 1$, this test reduces to Barnard's test. Here are the details of this test:

Null Hypothesis: $p_1 / p_2 = \phi_0$

Hypothesis Types: Both one-sided and two-sided

Reference Set: Ω

Test Statistic: $D(y) = \frac{\hat{p}_1 - \hat{p}_2 - \phi_0}{\sqrt{\left(\frac{\tilde{p}_1\tilde{q}_1}{n_1} + \phi_0^2\frac{\tilde{p}_2\tilde{q}_2}{n_2}\right)\left(\frac{N}{N-1}\right)}}$

Two-Sided Test: $I\big(D(y), D(x)\big) = |D(x)| \geq |D(y)|$

Lower One-Sided Test: $I\big(D(y), D(x)\big) = D(x) \leq D(y)$

Upper One-Sided Test: $I\big(D(y), D(x)\big) = D(x) \geq D(y)$

Farrington-Manning Unconditional Exact Test of the Ratio

Farrington and Manning (1990) proposed an unconditional exact test for testing whether the ratio is a specified value ϕ_0. This test was also discussed by Gart and Nam (1988). This test is only slightly different from the test of Miettinen and Nurminen (1985). Here are the details of this test:

Null Hypothesis: $p_1 / p_2 = \phi_0$

Hypothesis Types: Both one-sided and two-sided

Reference Set: Ω

Test Statistic: $D(y) = \frac{\hat{p}_1 - \hat{p}_2 - \phi_0}{\sqrt{\left(\frac{\tilde{p}_1\tilde{q}_1}{n_1} + \phi_0^2\frac{\tilde{p}_2\tilde{q}_2}{n_2}\right)}}$

Two-Sided Test: $I\big(D(y), D(x)\big) = |D(x)| \geq |D(y)|$

Lower One-Sided Test: $I\big(D(y), D(x)\big) = D(x) \leq D(y)$

Upper One-Sided Test: $I\big(D(y), D(x)\big) = D(x) \geq D(y)$

Constrained MLE's

The Miettinen and Nurminen (1985) and Farrington and Manning (1990) tests given above require maximum likelihood estimates that are constrained to follow the null hypothesis that $p_1 - p_2 = \delta_0$. The constrained maximum likelihood estimate for $\tilde{p}_2$ when considering the difference is the appropriate solution of the cubic equation

$$N\tilde{p}_2^3 + [(N + n_2)\delta_0 - (N + m_2)]\tilde{p}_2^2 + [m_2 - \delta_0(N + 2y_{21}) + n_2\delta_0^2]\tilde{p}_2 + y_{21}\delta_0(1 - \delta_0) = 0$$

The value for $\tilde{p}_1$ is found using the constraint

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

The constrained maximum likelihood estimate of $\tilde{p}_2$ when considering the ratio with the constraint that $p_1 / p_2 = \phi_0$ is the appropriate solution of the quadratic equation

$$N\tilde{p}_2^2 - [\phi_0(n_1 + y_{11}) + y_{21} + n_1]\tilde{p}_2 + m_2 = 0$$

The value for $\tilde{p}_1$ is found using the constraint

$$\tilde{p}_1 = \tilde{p}_2\phi_0$$

Data Structure

This procedure can summarize data from a database or summarized count values can be entered directly into the procedure panel in one of two ways: group sample sizes and group 'successes', or group 'successes' and group 'non-successes'.

Example 1 – Non-Inferiority Test of Two Proportions

This section presents an example of a non-inferiority test of the ratio of two proportions. The decision is made that a new treatment will be considered non-inferior to the current (standard) treatment if the positive response rate of the new treatment is at least 80% of the standard treatment.

In this example, 14 of 26 receiving the standard treatment responded positively and 16 of 23 receiving the experimental treatment responded positively.

Setup

To run this example, complete the following steps:

1 Specify the Two Proportions – Non-Inferiority Tests procedure options

- Find and open the **Two Proportions – Non-Inferiority Tests** procedure using the menus or the Procedure Navigator.
- The settings 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.

Data Tab	
Type of Data Input	**Summary Table of Counts: Enter Row Totals and First Column**
Group Heading	**Treatment**
Group Label of 1st Value	**Experimental**
Group Label of 2nd Value	**Standard**
Outcome Heading	**Response**
Outcome Label of 1st Value	**Positive**
Outcome Label of 2nd Value	**Negative**
Group 1 Total	**23**
Group 1, Outcome 1 Value	**16**
Group 2 Total	**26**
Group 2, Outcome 1 Value	**14**

Summary Reports Tab	
Counts and Proportions	**Checked**
Proportions Analysis	**Checked**

Ratio Reports Tab	
Gart-Nam Score Confidence Interval	**Checked**
Lower Proportions are	**Worse (H0: P1/P2 ≤ NIR vs. Ha: P1/P2 > NIR, with NIR < 1)**
Non-Inferiority Ratio (NIR)	**0.8**
Gart-Nam Score Non-Inferiority Test	**Checked**

2 Run the procedure

- Click the **Run** button to perform the calculations and generate the output.

Counts and Proportions Sections

Counts and Proportions

	Response			
Treatment	Positive Count	Negative Count	Total Count	Proportion*
Experimental	16	7	23	p1 = 0.6957
Standard	14	12	26	p2 = 0.5385

* Proportion = Positive / Total

Proportions Analysis

Statistic	Value
Group 1 Event Rate (p1)	0.6957
Group 2 Event Rate (p2)	0.5385
Absolute Risk Difference |p1 - p2|	0.1572
Number Needed to Treat 1/|p1 - p2|	6.36
Relative Risk Reduction |p1 - p2|/p2	0.29
Relative Risk p1/p2	1.29
Odds Ratio o1/o2	1.96

These reports document the values that were input and give various summaries of these values.

Confidence Interval

Confidence Intervals of the Ratio (P1/P2)

Confidence Interval Name	p1	p2	Ratio p1/p2	Lower 95% C.L. of P1/P2	Upper 95% C.L. of P1/P2	Confidence Interval Width
Gart-Nam Score	0.6957	0.5385	1.29	0.82	2.10	1.28

This report provides a large sample confidence interval of the ratio based on Gart-Nam Score formula shown earlier in this chapter.

Non-Inferiority Test

Upper Non-Inferiority Tests of the Ratio (P1/P2)

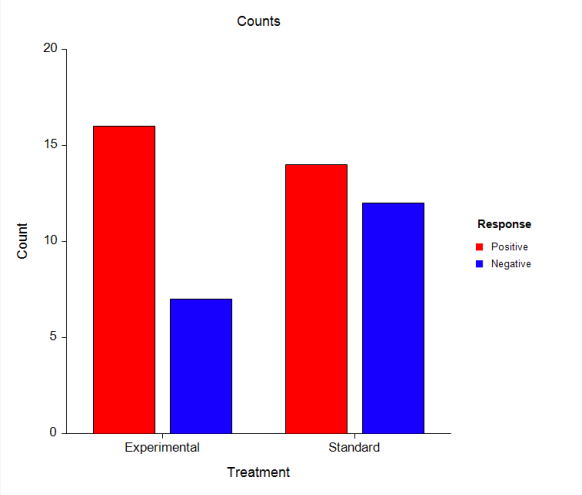
H0: P1/P2 ≤ 0.80 vs. Ha: P1/P2 > 0.80

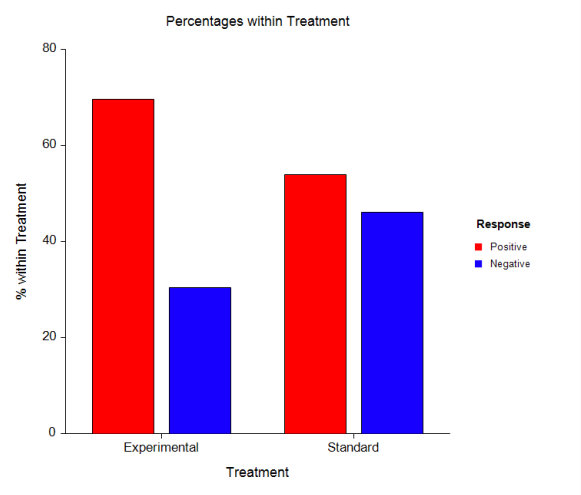
Test Statistic Name	p1	p2	Ratio p1/p2	Test Statistic Value	Prob Level	Reject H0 at α = 0.05?
Gart-Nam Score	0.6957	0.5385	1.29	2.057	0.0198	Yes

This report provides the large-sample Gart-Nam Score test. The p-value of the test is 0.0198. It indicates that the experimental treatment is non-inferior to the standard treatment.

Plots

Plots Section





These bar charts show the count and row percentages of the data.

Example 2 – Superiority Test of Two Proportions

This section presents an example of a superiority test of the ratio of two proportions. The decision is made that a new treatment will be considered superior to the current (standard) treatment if the positive response rate of the new treatment is at least 110% of the standard treatment.

In this example, 14 of 26 receiving the standard treatment responded positively and 19 of 23 receiving the experimental treatment responded positively.

Setup

To run this example, complete the following steps:

1 Specify the Two Proportions – Superiority by a Margin Tests procedure options

- Find and open the **Two Proportions – Superiority by a Margin Tests** procedure using the menus or the Procedure Navigator.
- The settings 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.

Data Tab	
Type of Data Input	**Summary Table of Counts: Enter Row Totals and First Column**
Group Heading	**Treatment**
Group Label of 1st Value	**Experimental**
Group Label of 2nd Value	**Standard**
Outcome Heading	**Response**
Outcome Label of 1st Value	**Positive**
Outcome Label of 2nd Value	**Negative**
Group 1 Total	**23**
Group 1, Outcome 1 Value	**19**
Group 2 Total	**26**
Group 2, Outcome 1 Value	**14**

Summary Reports Tab	
Counts and Proportions	**Checked**
Proportions Analysis	**Checked**

Ratio Reports Tab	
Gart-Nam Score Confidence Interval	**Checked**
Lower Proportions are	**Worse (H0: P1/P2 ≤ SR vs. Ha: P1/P2 > SR, with SR > 1)**
Superiority Ratio (SR)	**1.1**
Gart-Nam Score Superiority Test	**Checked**

2 Run the procedure

- Click the **Run** button to perform the calculations and generate the output.

Counts and Proportions Sections

Counts and Proportions

	Response			
Treatment	Positive Count	Negative Count	Total Count	Proportion*
Experimental	19	4	23	p1 = 0.8261
Standard	14	12	26	p2 = 0.5385

* Proportion = Positive / Total

Proportions Analysis

Statistic	Value
Group 1 Event Rate (p1)	0.8261
Group 2 Event Rate (p2)	0.5385
Absolute Risk Difference |p1 - p2|	0.2876
Number Needed to Treat 1/|p1 - p2|	3.48
Relative Risk Reduction |p1 - p2|/p2	0.53
Relative Risk p1/p2	1.53
Odds Ratio o1/o2	4.07

These reports document the values that were input and give various summaries of these values.

Confidence Interval

Confidence Intervals of the Ratio (P1/P2)

Confidence Interval Name	p1	p2	Ratio p1/p2	Lower 95% C.L. of P1/P2	Upper 95% C.L. of P1/P2	Confidence Interval Width
Gart-Nam Score	0.8261	0.5385	1.53	1.04	2.43	1.39

This report provides a large sample confidence interval of the ratio based on Gart-Nam Score formula shown earlier in this chapter.

Superiority Test

Upper Superiority Tests of the Ratio (P1/P2)

H0: P1/P2 ≤ 1.10 vs. Ha: P1/P2 > 1.10

Test Statistic Name	p1	p2	Ratio p1/p2	Test Statistic Value	Prob Level	Reject H0 at α = 0.05?
Gart-Nam Score	0.8261	0.5385	1.53	1.687	0.0458	Yes

This report provides the large-sample Gart-Nam Score test. The p-value of the test is 0.0458. It indicates that the experimental treatment is superior to the standard treatment.

Example 3 – Equivalence Test of Two Proportions

This section presents an example of an equivalence test of the ratio. To run an equivalence test, both upper and lower bounds of equivalence must be set. In this example, the lower bound will be set at 0.80 and the upper bound is set to 1.20. That is, if the positive response rate of the experimental group is at least 80% of the control group, and at most 120% of the control group, the experimental group is concluded to be 'equivalent' to the control group.

In this example, 265 of 337 receiving the standard treatment responded positively and 221 of 315 receiving the experimental treatment responded positively.

Setup

To run this example, complete the following steps:

1 Specify the Two Proportions – Equivalence Tests procedure options

- Find and open the **Two Proportions – Equivalence Tests** procedure using the menus or the Procedure Navigator.
- The settings 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.

Data Tab	
Type of Data Input	**Summary Table of Counts: Enter Row Totals and First Column**
Group Heading	**Treatment**
Group Label of 1st Value	**Experimental**
Group Label of 2nd Value	**Control**
Outcome Heading	**Response**
Outcome Label of 1st Value	**Positive**
Outcome Label of 2nd Value	**Negative**
Group 1 Total	**315**
Group 1, Outcome 1 Value	**221**
Group 2 Total	**337**
Group 2, Outcome 1 Value	**265**

Reports Tab	
Counts and Proportions	**Checked**
Proportions Analysis	**Checked**

Plots Tab	
Gart-Nam Score Confidence Interval	**Checked**
Upper Equivalence Bound	**1.2**
Lower Equivalence Bound	**0.8**
Gart-Nam Score Equivalence Test	**Checked**

2 Run the procedure

- Click the **Run** button to perform the calculations and generate the output.

Counts and Proportions Sections

Counts and Proportions

	Response			
Treatment	Positive Count	Negative Count	Total Count	Proportion*
Experimental	221	94	315	p1 = 0.7016
Control	265	72	337	p2 = 0.7864

* Proportion = Positive / Total

Proportions Analysis

Statistic	Value
Group 1 Event Rate (p1)	0.7016
Group 2 Event Rate (p2)	0.7864
Absolute Risk Difference |p1 - p2|	0.0848
Number Needed to Treat 1/|p1 - p2|	11.80
Relative Risk Reduction |p1 - p2|/p2	0.11
Relative Risk p1/p2	0.89
Odds Ratio o1/o2	0.64

These reports document the values that were input and give various summaries of these values.

Confidence Interval

Confidence Intervals of the Ratio (P1/P2)

Confidence Interval Name	p1	p2	Ratio p1/p2	Lower 95% C.L. of P1/P2	Upper 95% C.L. of P1/P2	Confidence Interval Width
Gart-Nam Score	0.7016	0.7864	0.89	0.81	0.98	0.16

This report provides a large sample confidence interval of the ratio based on Gart-Nam Score formula shown earlier in this chapter.

Equivalence Test

Equivalence Tests of the Ratio (P1/P2)

H0: P1/P2 ≤ 0.80 or P1/P2 ≥ 1.20 vs. Ha: Equivalence

Test Name	Ratio p1/p2	Lower Test Statistic	Lower Prob Level	Upper Test Statistic	Upper Prob Level	TOST Prob Level	Reject H0 at α = 0.05?
Gart-Nam Score	0.89	2.285	0.0112	-6.271	0.0000	0.0112	Yes

This report gives the p-value for each of the two one-sided tests (TOST). The larger of the two individual p-values is the equivalence test p-value of 0.0112, which indicates strong evidence in favor of the alternative hypothesis of equivalence.

Example 4 – Two-Sided Test of Two Proportions versus a Margin

This section presents an example of two-sided of a difference versus a non-zero margin. In this example, the Margin will be set at -0.05.

In this example, 265 of 337 receiving the standard treatment responded positively and 221 of 315 receiving the experimental treatment responded positively.

Setup

To run this example, complete the following steps:

1 Specify the Two Proportions – Two-Sided Tests vs. a Margin procedure options

- Find and open the **Two Proportions – Two-Sided Tests vs. a Margin** procedure using the menus or the Procedure Navigator.
- The settings 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.

Data Tab	
Type of Data Input	**Summary Table of Counts: Enter Row Totals and First Column**
Group Heading	**Treatment**
Group Label of 1st Value	**Experimental**
Group Label of 2nd Value	**Standard**
Outcome Heading	**Response**
Outcome Label of 1st Value	**Positive**
Outcome Label of 2nd Value	**Negative**
Group 1 Total	**315**
Group 1, Outcome 1 Value	**221**
Group 2 Total	**337**
Group 2, Outcome 1 Value	**265**
Summary Reports Tab	
Counts and Proportions	**Checked**
Proportions Analysis	**Checked**
Difference Reports Tab	
Gart-Nam Score Confidence Interval	**Checked**
Difference Margin (DM)	**-0.05**
Gart-Nam Score Test	**Checked**

2 Run the procedure

- Click the **Run** button to perform the calculations and generate the output.

Counts and Proportions Sections

Counts and Proportions

	Response			
Treatment	Positive Count	Negative Count	Total Count	Proportion*
Experimental	221	94	315	p1 = 0.7016
Standard	265	72	337	p2 = 0.7864

* Proportion = Positive / Total

Proportions Analysis

Statistic	Value
Group 1 Event Rate (p1)	0.7016
Group 2 Event Rate (p2)	0.7864
Absolute Risk Difference |p1 - p2|	0.0848
Number Needed to Treat 1/|p1 - p2|	11.80
Relative Risk Reduction |p1 - p2|/p2	0.11
Relative Risk p1/p2	0.89
Odds Ratio o1/o2	0.64

These reports document the values that were input and give various summaries of these values.

Confidence Interval

Confidence Intervals of the Difference (P1 - P2)

Confidence Interval Name	p1	p2	Difference p1 - p2	Lower 95% C.L. of P1 - P2	Upper 95% C.L. of P1 - P2	Confidence Interval Width
Gart-Nam Score	0.7016	0.7864	-0.0848	-0.1517	-0.0178	0.1338

This report provides a large sample confidence interval of the ratio based on Gart-Nam Score formula shown earlier in this chapter.

Two-Sided Tests of the Difference

Two-Sided Tests of the Difference (P1 - P2) Compared to a Hypothesized Value

H0: P1 - P2 = -0.0500 vs. Ha: P1 - P2 ≠ -0.0500

Test Statistic Name	p1	p2	Difference p1 - p2	Test Statistic Value	Prob Level	Reject H0 at α = 0.05?
Gart-Nam Score	0.7016	0.7864	-0.0848	-1.019	0.3081	No

This report gives the p-value as 0.3081. Since this is not less than 0.05, the null hypothesis cannot be rejected.