Chapter 519

Two Correlated Proportions – Non-Inferiority, Superiority, and Equivalence Tests

Introduction

This chapter documents three closely related procedures: non-inferiority tests, superiority (by a margin) tests, and equivalence tests. These procedures compute asymptotic 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.

This procedure computes confidence intervals and hypothesis tests for the comparison of the marginal frequencies of two factors (each with two levels) based on a 2-by-2 table of n pairs. Confidence limits can be obtained for the marginal probability difference, ratio, or odds ratio. Non-inferiority, superiority (by a margin) tests, and equivalence tests are available for the marginal probability difference and ratio.

Experimental Design

A typical design for this scenario involves N pairs of individuals where a dichotomous measurement of one factor is measured on one of the individuals of the pair (case), and a second dichotomous measurement based on a second factor is measured on the second individual of the pair (control). Or similarly N individuals are measured twice, once for each of two dichotomous factors.

Comparing Two Correlated Proportions

Suppose you have two dichotomous measurements Y_1 and Y_2 on each of N subjects (where in many cases the 'subject' may be a pair of matched individuals). The proportions P_1 and P_2 represent the success probabilities. That is,

$$P_1 = \Pr(Y_1 = 1)$$
$$P_2 = \Pr(Y_2 = 1)$$

The data from this design can be summarized in the following 2-by-2 table:

	$Y_2 = 1$ (Yes, Present)	$Y_2 = 0$ (No, Absent)	Total
$Y_1 = 1$ (Yes, Present)	A	B	A + B
$Y_1 = 0$ (No, Absent)	C	D	C + D
Total	A + C	B + D	N

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

$$p_1 = \frac{A + B}{N} \quad \text{and} \quad p_2 = \frac{A + C}{N}$$

Three quantities which allow these proportions to be compared are

Quantity	**Notation**
Difference	$\Delta = P_1 - P_2$
Risk Ratio	$\phi = P_1 / P_2$
Odds Ratio	$\psi = \dfrac{O_1}{O_2}$

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

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

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

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

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

In the case of two correlated proportions, the odds ratio is calculated as

$$\psi = \frac{B}{C}$$

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

Several methods for computing confidence intervals for proportion difference, proportion ratio, and odds ratio have been proposed. We now show the methods that are available in **NCSS**.

Difference

Four methods are available for computing a confidence interval of the difference between the two proportions $\Delta = P_1 - P_2$. The lower (L) and upper (U) limits of these intervals are computed as follows. Note that $z = |z_{\alpha/2}|$ is the appropriate percentile from the standard normal distribution.

Newcombe (1998) conducted a comparative evaluation of ten 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.

Nam's Score

For details, see Nam (1997) or Tango (1998). The lower limit is the solution of

$$L = \inf\left\{\Delta_0 : \frac{\hat{\Delta} - \Delta_0}{\tilde{\sigma}_{\Delta_0}} < z\right\}$$

and the upper limit is the solution of

$$U = \sup\left\{\Delta_0 : \frac{\hat{\Delta} - \Delta_0}{\tilde{\sigma}_{\Delta_0}} > -z\right\}$$

where $\tilde{\sigma}_{\Delta_0}$ is given by

$$\tilde{\sigma}_\Delta = \frac{\tilde{p}_{21} + \tilde{p}_{12} - \Delta^2}{n}$$

with

$$\tilde{p}_{21} = \left\{\frac{-e + \sqrt{e^2 - 8f}}{4}\right\}$$

$$\tilde{p}_{12} = \tilde{p}_{21} - \Delta$$

$$e = -\hat{\Delta}(1 - \Delta) - 2(\hat{p}_{21} + \Delta)$$

$$f = \Delta(1 + \Delta)\hat{p}_{21}$$

Wilson's Score as modified by Newcombe

For further details, see Newcombe (1998c), page 2639. This is Newcombe's method 10.

$$L = \hat{\Delta} - \delta$$

$$U = \hat{\Delta} + \varepsilon$$

where

$$\delta = \sqrt{f_2^2 - 2\hat{\phi}f_2 g_3 + g_3^2}$$

$$\varepsilon = \sqrt{g_2^2 - 2\hat{\phi}g_2 f_3 + f_3^2}$$

$$f_2 = \frac{(A + B)}{N} - l_2$$

$$g_2 = u_2 - \frac{(A + B)}{N}$$

$$f_3 = \frac{(A + C)}{N} - l_3$$

$$g_3 = u_3 - \frac{(A + C)}{N}$$

and l_2 and u_2 are the roots of

$$\left| x - \frac{A + B}{N} \right| = z \sqrt{\frac{x(1 - x)}{N}}$$

and l_3 and u_3 are the roots of

$$\left| x - \frac{A + C}{N} \right| = z \sqrt{\frac{x(1 - x)}{N}}$$

$$\hat{\phi} = \begin{cases} \dfrac{\max{(AD - BC - N/2, 0)}}{\sqrt{(A + B)(C + D)(A + C)(B + D)}} & \text{if } AD > BC \\ \dfrac{AD - BC}{\sqrt{(A + B)(C + D)(A + C)(B + D)}} & \text{otherwise} \end{cases}$$

Note that if the denominator of $\hat{\phi}$ is zero, $\hat{\phi}$ is set to zero.

Wald Z Method

For further details, see Newcombe (1998c), page 2638.

$$L = \hat{\Delta} - z s_W$$

$$U = \hat{\Delta} + z s_W$$

where

$$\hat{\Delta} = p_1 - p_2 = (B - C)/N$$

$$s_W^2 = \frac{(A + D)(B + C) + 4BC}{N^3}$$

Wald Z Method with Continuity Correction

For details, see Newcombe (1998c), page 2638.

$$L = \hat{\Delta} - z s_W - \frac{1}{N}$$

$$U = \hat{\Delta} + z s_W + \frac{1}{N}$$

Ratio

Two methods are available for computing a confidence interval of the risk ratio $\phi = P_1 \, / \, P_2$. Note that $z = \left| z_{\alpha/2} \right|$ is the appropriate percentile from the standard normal distribution.

Nam and Blackwelder (2002) present two methods for computing confidence intervals for the risk ratio. These are presented here. Note that the score method is recommended.

Nam and Blackwelder Score

For details, see Nam and Blackwelder (2002), page 691. The lower limit is the solution of

$$z(\phi) = \left| z_{\alpha/2} \right|$$

and the upper limit is the solution of

$$z(\phi) = -\left| z_{\alpha/2} \right|$$

where

$$z(\phi) = \frac{\sqrt{N}(p_1 - \phi p_2)}{\sqrt{\phi(\tilde{p}_{12} + \tilde{p}_{21})}}$$

and

$$\tilde{p}_{12} = \frac{-p_1 + \phi^2(p_2 + 2p_{12}) + \sqrt{(p_1 - \phi p_2)^2 + 4\phi^2 p_{12} p_{12}}}{2\phi(\phi + 1)}$$

$$\tilde{p}_{21} = \phi\tilde{p}_{12} - (\phi - 1)(1 - p_{22})$$

Nam and Blackwelder Wald Z

For details, see Nam and Blackwelder (2002), page 692. The lower limit is the solution of

$$z_W(\phi) = |z_{\alpha/2}|$$

and the upper limit is the solution of

$$z_W(\phi) = -|z_{\alpha/2}|$$

where

$$z_W(\phi) = \frac{\sqrt{N}(\hat{p}_1 - \phi\hat{p}_2)}{\sqrt{\phi(\hat{p}_{12} + \hat{p}_{21})}}$$

Hypothesis Tests

Several statistical tests are available for testing hypotheses about two correlated proportions. Some tests are based on the *difference* in the proportions while others are based on the *ratio* of proportions. In this section, some of these distinctions are explained.

Types of Hypothesis Tests

The tests used in these procedures are 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. We suggest a minimum of 25 pairs of data points.

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.

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 ≤ 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 ≥ 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 ≤ 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 ≥ 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.

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.

Test Statistics

This section documents the various test statistics that are available.

Difference

These modules test statistical hypotheses about the difference in the two proportions:

1. $H_{0L}: P_1 - P_2 \geq \Delta$ versus $H_{aL}: P_1 - P_2 < \Delta$; this is a *one-tailed test*.

2. $H_{0U}: P_1 - P_2 \leq \Delta$ versus $H_{aU}: P_1 - P_2 > \Delta$; this is a *one-tailed test*.

Nam Test

Liu *et al.* (2002) recommend a likelihood score test which was originally proposed by Nam (1997). The tests are calculated as

$$z_L = \frac{\hat{\Delta} + \Delta}{\tilde{\sigma}_L} \text{ and } z_U = \frac{\hat{\Delta} - \Delta}{\tilde{\sigma}_U}$$

where

$$\tilde{\sigma}_L = \sigma_{-\Delta}$$

$$\tilde{\sigma}_U = \sigma_{\Delta}$$

and

$$\sigma_D = \frac{\tilde{p}_{21} + \tilde{p}_{12} - D^2}{N}$$

$$\tilde{p}_{21} = \left\{ \frac{-e + \sqrt{e^2 - 8f}}{4} \right\}$$

$$\tilde{p}_{12} = \tilde{p}_{21} - D$$

$$e = -\hat{\Delta}(1 - D) - 2(p_{21} + D)$$

$$f = D(1 + D)p_{21}$$

Wald Test

Liu *et al.* (2002) present a pair of large-sample, Wald-type z tests for testing the two one-tailed hypothesis about the difference $p_1 - p_2 = \Delta$. These are calculated as

$$z_L = \frac{\hat{\Delta} + \Delta - \frac{1}{2N}}{\hat{\sigma}} \text{ and } z_U = \frac{\hat{\Delta} - \Delta + \frac{1}{2N}}{\hat{\sigma}}$$

where

$$\hat{\sigma}^2 = \frac{p_{21} + p_{12} - \hat{\Delta}^2}{N}$$

$$\hat{\Delta} = p_1 - p_2$$

Ratio

These modules test three statistical hypotheses about the difference in the two proportions:

1. $H_{0L}: P_1/P_2 \geq \phi$ versus $H_{aL}: P_1/P_2 > \phi$; this is a *one-tailed test*.

2. $H_{0U}: P_1/P_2 \geq \phi$ versus $H_{aU}: P_1/P_2 < \phi$; this is a *one-tailed test*.

Nam Test

For details, see Nam and Blackwelder (2002), page 691. The test statistic for testing a specific value of ϕ is

$$z(\phi) = \frac{\sqrt{N}(p_1 - \phi p_2)}{\sqrt{\phi(\tilde{p}_{12} + \tilde{p}_{21})}}$$

where

$$\tilde{p}_{12} = \frac{-p_1 + \phi^2(p_2 + 2p_{12}) + \sqrt{(p_1 - \phi p_2)^2 + 4\phi^2 p_{12} p_{12}}}{2\phi(\phi + 1)}$$

$$\tilde{p}_{21} = \phi \tilde{p}_{12} - (\phi - 1)(1 - p_{22})$$

Equivalence Tests

Equivalence tests are hypothesis tests in which the alternative hypothesis, not the null hypothesis, is equality. For example, suppose an accurate diagnostic test has serious side effects, so a replacement test is sought. In this case, researchers are not interested in showing that the two tests are different, but rather that they are the same.

These tests are often divided into two categories: *equivalence* (two-sided) tests and *non-inferiority* (one-sided) tests. Here, the term *equivalence tests* means that we want to show that two tests are equivalent—that is, their accuracy is about the same. This requires a two-sided hypothesis test. On the other hand, *non-inferiority tests* are used when we want to show that a new (experimental) test is no worse than the existing (reference or gold-standard) test. This requires a one-sided hypothesis test.

In the discussion to follow, two ways of expressing difference are considered: the difference and the ratio. The simple difference between two proportions is perhaps the most straight forward way of expressing that these two proportions are different. A difference of zero means that the two proportions are equal. Unfortunately, this method does not work well near zero or one. For example, suppose a diagnostic test achieves 95% accuracy and we wish to establish that a new test is within 7 percentage points of the original test. The acceptable range is from 95% - 7% = 88% to 95% + 7% = 102%. Of course, 102% is impossible.

A second method of setting up the hypotheses that does not suffer from this problem is to consider the ratio of the two proportions. A ratio of one indicates that the two proportions are equal. Using the ratio to define the hypotheses is less prone to impossible values.

Equivalence Based on the Difference

The equivalence between two proportions may be tested using the following hypothesis which is based on the difference:

$$H_0: P_1 - P_2 \leq \Delta_L \ \text{ or } \ H_0: P_1 - P_2 \geq \Delta_U \quad \text{versus} \quad H_a: \Delta_L < P_1 - P_2 < \Delta_U$$

where Δ_L and Δ_U are the established equivalence limits.

This hypothesis is often tested at the alpha significance level by determining if $100(1 - 2\alpha)\%$ confidence limits are both between Δ_L and Δ_U. A second method to test this hypothesis is to separate it into two sets of one-sided hypothesis tests (TOST).

First One-Sided Test

$H_{0L}: P_1 - P_2 \leq \Delta_L$ versus $H_{aL}: P_1 - P_2 > \Delta_L$.

Second One-Sided Test

$H_{0U}: P_1 - P_2 \geq \Delta_U$ versus $H_{aU}: P_1 - P_2 < \Delta_U$.

Nam Test

Both the two, one-sided hypotheses method and the confidence interval test are tested using the score test proposed by Nam (1997). These tests and the confidence interval formulas were presented above.

Equivalence Based on the Ratio

The equivalence between two proportions may be tested using the following hypothesis which is based on the ratio:

$$H_0 : \frac{P_1}{P_2} \leq \phi_L \;\; \text{or} \;\; H_0 : \frac{P_1}{P_2} \geq \phi_U \;\;\; \text{versus} \;\;\; H_a : \phi_L < P_1/P_2 < \phi_U$$

where the ϕ_L and ϕ_U are the established equivalence limits.

This hypothesis is often tested at the alpha significance level by determining if $100(1 - 2\alpha)\%$ confidence limits are both between ϕ_L and ϕ_U. A second method to test this hypothesis is to separate it into two sets of one-sided hypotheses (TOST).

First One-Sided Test

$H_{0L} : P_1/P_2 \leq \phi_L \; versus \, H_{aL} : P_1/P_2 > \phi_L.$

Second One-Sided Test

$H_{0U} : P_1/P_2 \geq \phi_U \; versus \, H_{aU} : P_1/P_2 < \phi_U.$

Nam Test

Both the two, one-sided hypotheses method and the confidence interval test are tested using the score test proposed by Nam and Blackwelder (2002). These tests and the confidence interval where presented above.

Data Structure

This procedure can summarize data from a database or summarized count values can be entered directly into the procedure panel.

Example 1 – Non-Inferiority Test of Two Correlated Proportions

A study was made of 100 matched pairs of individuals. One individual from each pair was given a new treatment and the other individual in the pair was given a standard treatment. The choice was made by the flip of a coin. The response of each individual is either recovery or no recovery. The researchers wish to show that the new treatment is **no worse** than the standard treatment. The results were tabulated into the following table. Note that the table is arranged so that the new treatment is shown as the rows of the table and the standard treatment is shown as the columns of the table.

	Standard Treatment	
New Treatment	Recovery	No Recovery
Recovery	53	23
No Recovery	15	9

The researchers wish to show that the proportion recovering with the new treatment is no less than 0.05 below the proportion recovering with the standard treatment. The value for alpha to be used is 0.025.

Setup

To run this example, complete the following steps:

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

- Find and open the **Two Correlated 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 ....**Enter Table of Counts**	
Variable 1 Heading ....**New**	
Variable 1 Label of 1st Value..........................**Recovery**	
Variable 1 Label of 2nd Value.........................**No Recovery**	
Variable 2 Heading ....**Standard**	
Variable 2 Label of 1st Value..........................**Recovery**	
Variable 2 Label of 2nd Value.........................**No Recovery**	
N11 ....**53**	
N12 ....**23**	
N21 ....**15**	
N22 ....**9**	

Summary Reports Tab

Alpha for Tests...**0.025**	
Counts and Proportions ....**Checked**	
Proportions Analysis ....**Checked**	

Difference Reports Tab

Lower Proportions are ....**Worse (H0: P1 - P2 ≤ NIM vs. Ha: P1 - P2 > NIM, with NIM < 0)**
Non-Inferiority Margin (NIM) ....**-0.05**
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

	Standard		
New	**Recovery Count**	**No Recovery Count**	**Total Count**
Recovery	53	23	76
No Recovery	15	9	24
Total	68	32	100

Proportions

p1 = (76/100) = 0.7600
p2 = (68/100) = 0.6800

Proportions Analysis

Statistic	Value
Variable 1 Event Rate (p1)	0.7600
Variable 2 Event Rate (p2)	0.6800
Proportion Matching	0.6200
Proportion Not Matching	0.3800
Absolute Risk Difference |p1 - p2|	0.0800
Number Needed to Treat 1/|p1 - p2|	12.5000
Relative Risk Reduction |p1 - p2|/p2	0.1176
Relative Risk p1/p2	1.1176
Odds Ratio o1/o2	1.5333

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

Non-Inferiority Test Report

Upper Non-Inferiority Tests of the Difference (P1 - P2)

H0: P1 - P2 ≤ -0.05 vs. Ha: P1 - P2 > -0.05

Test Statistic Name	p1	p2	Difference p1 - p2	Test Statistic Value	Prob Level	Reject H0 at α = 0.025?
Nam Score*	0.7600	0.6800	0.0800	2.088	0.0184	Yes

* It is recommended that N be at least 25 when using the Nam Score test.

This report provides the Nam Score non-inferiority test. The p-value of the test is the Prob Level.

Plots Section

Plots Section
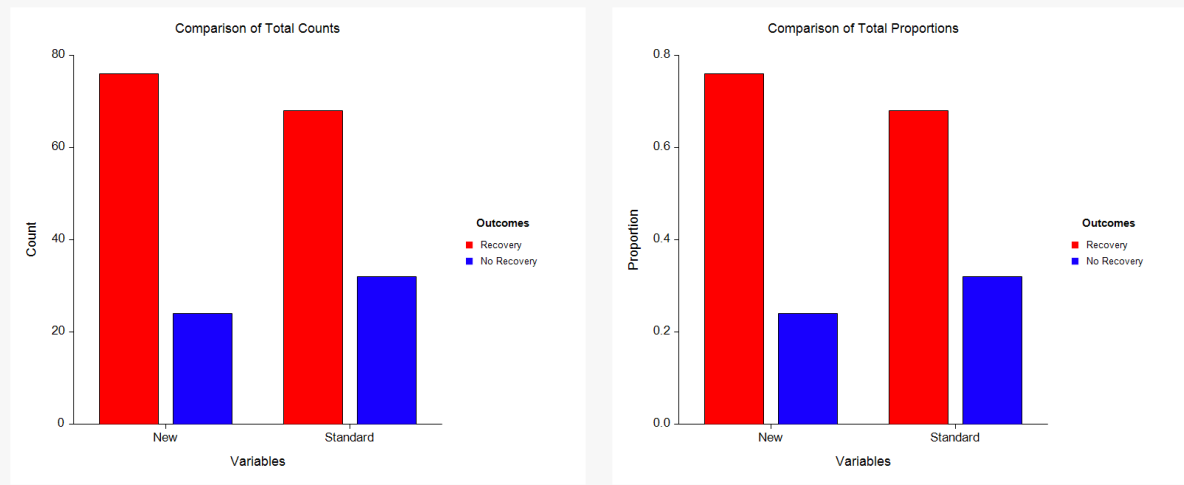
These reports show the marginal totals and proportions of the two variables.

Example 2 – Superiority Test of Two Correlated Proportions

A study was made of 100 matched pairs of individuals. One individual from each pair was given a new treatment and the other individual in the pair was given a standard treatment. The choice was made by the flip of a coin. The response of each individual is either recovery or no recovery. The researchers wish to show that the new treatment is **definitely better** than the standard treatment. The results were tabulated into the following table. Note that the table is arranged so that the new treatment is shown as the rows of the table and the standard treatment is shown as the columns of the table.

	Standard Treatment	
New Treatment	Recovery	No Recovery
Recovery	50	26
No Recovery	9	15

The researchers wish to show that the proportion recovering with the new treatment is at least 0.05 above the proportion recovering with the standard treatment. The value for alpha to be used is 0.025.

Setup

To run this example, complete the following steps:

1 Specify the Two Correlated Proportions – Superiority by a Margin Tests procedure options
- Find and open the **Two Correlated 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 ....**Enter Table of Counts**
Variable 1 Heading ....**New**
Variable 1 Label of 1st Value..........................**Recovery**
Variable 1 Label of 2nd Value.........................**No Recovery**
Variable 2 Heading ....**Standard**
Variable 2 Label of 1st Value..........................**Recovery**
Variable 2 Label of 2nd Value.........................**No Recovery**
N11 ....**50**
N12 ....**26**
N21 ....**9**
N22 ....**15**

Summary Reports Tab
_____

Alpha for Tests...**0.025**
Lower Proportions are ....**Worse (H0: P1 - P2 ≤ SM vs. Ha: P1 - P2 > SM, with SM > 0)**
Superiority Margin (SM).................................**0.05**
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
_____

	Standard		
New	**Recovery Count**	**No Recovery Count**	**Total Count**
Recovery	50	26	76
No Recovery	9	15	24
Total	59	41	100

Proportions
_____

p1 = (76/100) = 0.7600
p2 = (59/100) = 0.5900

_____

Proportions Analysis
_____

Statistic	Value
Variable 1 Event Rate (p1)	0.7600
Variable 2 Event Rate (p2)	0.5900
Proportion Matching	0.6500
Proportion Not Matching	0.3500
Absolute Risk Difference |p1 - p2|	0.1700
Number Needed to Treat 1/|p1 - p2|	5.8824
Relative Risk Reduction |p1 - p2|/p2	0.2881
Relative Risk p1/p2	1.2881
Odds Ratio o1/o2	2.8889

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

Superiority Test Report

Upper Superiority Tests of the Difference (P1 - P2)

H0: P1 - P2 ≤ 0.05 vs. Ha: P1 - P2 > 0.05

Test Statistic Name	p1	p2	Difference p1 - p2	Test Statistic Value	Prob Level	Reject H0 at α = 0.025?
Nam Score*	0.7600	0.5900	0.1700	2.071	0.0192	Yes

* It is recommended that N be at least 25 when using the Nam Score test.

This report provides the Nam Score superiority test. The p-value of the test is the Prob Level.

Example 3 – Equivalence Test of Two Correlated Proportions

A study was made of 200 matched pairs of individuals. One individual from each pair was given a new treatment and the other individual in the pair was given a standard treatment. The choice was made by the flip of a coin. The response of each individual is either recovery or no recovery. The researchers wish to show that the new treatment is **equivalent** to the standard treatment. The results were tabulated into the following table. Note that the table is arranged so that the new treatment is shown as the rows of the table and the standard treatment is shown as the columns of the table.

	Standard Treatment	
New Treatment	Recovery	No Recovery
Recovery	106	36
No Recovery	40	18

The researchers wish to show that the proportion recovering with the new treatment is no less than 0.1 above or below the proportion recovering with the standard treatment. The value for alpha to be used is 0.05.

Setup

To run this example, complete the following steps:

1   **Specify the Two Correlated Proportions – Equivalence Tests procedure options**

- Find and open the **Two Correlated 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 ....**Enter Table of Counts**
Variable 1 Heading ....**New**
Variable 1 Label of 1st Value..........................**Recovery**
Variable 1 Label of 2nd Value.........................**No Recovery**
Variable 2 Heading ....**Standard**
Variable 2 Label of 1st Value..........................**Recovery**
Variable 2 Label of 2nd Value.........................**No Recovery**
N11 ....**106**
N12 ....**36**
N21 ....**40**
N22 ....**18**

Summary Reports Tab
_____

Alpha for Tests...**0.05**
Counts and Proportions ....**Checked**
Proportions Analysis ....**Checked**

Difference Reports Tab
_____

Upper Equivalence Bound **0.10**
Lower Equivalence Bound ....**-0.10**
Equivalence Test (Difference).........................**Checked**

2   Run the procedure

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

Counts and Proportions Sections

Counts and Proportions
_____

	Standard		
New	**Recovery Count**	**No Recovery Count**	**Total Count**
Recovery	106	36	142
No Recovery	40	18	58
Total	146	54	200

Proportions
_____

p1 = (142/200) = 0.7100
p2 = (146/200) = 0.7300
_____
_____

Proportions Analysis
_____

Statistic	**Value**
Variable 1 Event Rate (p1)	0.7100
Variable 2 Event Rate (p2)	0.7300
Proportion Matching	0.6200
Proportion Not Matching	0.3800
Absolute Risk Difference |p1 - p2|	0.0200
Number Needed to Treat 1/|p1 - p2|	50.0000
Relative Risk Reduction |p1 - p2|/p2	0.0274
Relative Risk p1/p2	0.9726
Odds Ratio o1/o2	0.9000

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

Equivalence Test Report

Equivalence Test of the Difference (P1 - P2)

H0: P1 - P2 ≤ -0.1 or P1 - P2 ≥ 0.1 vs. Ha: Equivalence

Test Name	Difference p1 - p2	Lower Test Statistic	Lower Prob Level	Upper Test Statistic	Upper Prob Level	TOST Prob Level	Reject H0 at α = 0.05?
Nam Score*	-0.0200	1.829	0.0337	-2.722	0.0032	0.0337	Yes

* It is recommended that N be at least 25 when using the Nam Score test.

This report provides the Nam Score equivalence test. The p-value of the test is the TOST Prob Level.