

Chapter 151

Tests for Multiple Correlated Proportions (McNemar-Bowker Test of Symmetry)

Introduction

McNemar's test for correlated proportions requires that there be only 2 possible categories for each outcome. It is possible, however, that in some cases the outcomes could be classified into more than 2 categories (i.e. k categories). The paired data that results from this type of experiment can be summarized in a $k \times k$ contingency table of counts or proportions and can be analyzed using the McNemar-Bowker Test. This procedure calculates power and sample size for the McNemar-Bowker Test of symmetry in a paired experiment.

The sample size calculations in the procedure are based on the formulas presented in Fagerland, Lydersen, and Laake (2017) on pages 536 and 537 and given in Chow, Shao, Wang, & Lokhnygina (2018) on pages 141-143.

A Note about the McNemar-Bowker and Stuart-Maxwell Tests

Chow, Shao, Wang, and Lokhnygina (2018) incorrectly refers to the McNemar-Bowker Test as the Stuart-Maxwell Test. The Stuart-Maxwell Test is used for testing marginal homogeneity, utilizes a different and more complicated test statistic, and has no known closed-form sample size calculation method. The McNemar-Bowker Test (the test for which power is computed in this procedure) is used for testing paired table symmetry. While it's true that symmetry implies marginal homogeneity (see Fagerland, Lydersen, and Laake (2017)), the power and sample size calculations for the McNemar-Bowker Test do not necessarily extend to the Stuart-Maxwell Test.

Technical Details

Consider N subjects in a paired experiment where the outcome has k possible categories. The results can be summarized in a $k \times k$ contingency table of counts as

		Treatment 2				
		1	2	...	k	Total
Treatment 1						
1		n_{11}	n_{12}	...	n_{1k}	$n_{1\cdot}$
2		n_{21}	n_{22}	...	n_{2k}	$n_{2\cdot}$
\vdots		\vdots	\vdots	\ddots	\vdots	\vdots
k		n_{k1}	n_{k2}	...	n_{kk}	$n_{k\cdot}$
Total		$n_{\cdot 1}$	$n_{\cdot 2}$...	$n_{\cdot k}$	N

This table of counts can be summarized as proportions by dividing each cell by N , resulting in the following table of proportions, p_{ij} :

		Treatment 2				
		1	2	...	k	Total
Treatment 1						
1		p_{11}	p_{12}	...	p_{1k}	$p_{1\cdot}$
2		p_{21}	p_{22}	...	p_{2k}	$p_{2\cdot}$
\vdots		\vdots	\vdots	\ddots	\vdots	\vdots
k		p_{k1}	p_{k2}	...	p_{kk}	$p_{k\cdot}$
Total		$p_{\cdot 1}$	$p_{\cdot 2}$...	$p_{\cdot k}$	1

Test Statistic

The hypotheses tested by the McNemar-Bowker test are

$$H_0: p_{ij} = p_{ji} \text{ for all } i \neq j \quad \text{vs} \quad H_A: p_{ij} \neq p_{ji} \text{ for some } i \neq j.$$

The McNemar-Bowker test statistic used to test these hypotheses is

$$T_{MB} = \sum_{i < j} \frac{(n_{ij} - n_{ji})^2}{n_{ij} + n_{ji}}.$$

For a significance level α , the test rejects the null hypothesis if $T_{MB} > \chi_{1-\alpha, k(k-1)/2}^2$, where $\chi_{1-\alpha, k(k-1)/2}^2$ is the upper $1 - \alpha$ percentile of the standard chi-square distribution with $k(k-1)/2$ degrees of freedom.

Power Calculation

According to Chow, Shao, Wang, & Lohknygina (2018) page 141, the power for the McNemar-Bowker test is

$$1 - \chi_{k(k-1)/2}^2(\chi_{1-\alpha, k(k-1)/2}^2 | \delta)$$

where $\chi_{k(k-1)/2}^2(X|NCP)$ is the non-central chi-square cumulative distribution function with $k(k-1)/2$ degrees of freedom and non-centrality parameter NCP evaluated at X , and $\chi_{1-\alpha, k(k-1)/2}^2$ is the upper $1 - \alpha$ percentile of the standard chi-square distribution with $k(k-1)/2$ degrees of freedom. The non-centrality parameter δ is equal to

$$\delta = T_{MB} = N \sum_{i < j} \frac{(p_{ij} - p_{ji})^2}{p_{ij} + p_{ji}}$$

In the **PASS**, the summed quantity $\sum_{i < j} \frac{(p_{ij} - p_{ji})^2}{p_{ij} + p_{ji}}$ is referred to as the discordant proportion ratio sum (DPRS) such that

$$\delta = N \times DPRS.$$

The sample size is determined using a binary search of possible values for N .

Procedure Options

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

Design Tab

The Design tab contains most of the parameters and options that you will be concerned with.

Solve For

Solve For

This option specifies the parameter to be calculated from the values of the other parameters. Under most conditions, you would select either *Power* or *Sample Size*.

Select *Sample Size* when you want to determine the sample size needed to achieve a given power and alpha level.

Select *Power* when you want to calculate the power of an experiment that has already been run.

Select *Effect Size (DPRS)* when you want to calculate the minimum effect size that can be detected.

Power and Alpha

Power

This option specifies one or more values for power. Power is the probability of rejecting a false null hypothesis and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected. In this procedure, a type-II error occurs when you fail to reject the null hypothesis of equal means when in fact the means are different.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as *0.8 to 0.95 by 0.05* may be entered.

Tests for Multiple Correlated Proportions (McNemar-Bowker Test of Symmetry)

Alpha

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected. In this procedure, a type-I error occurs when you reject the null hypothesis of equal means when in fact the means are equal.

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as *0.01 0.05 0.10* or *0.01 to 0.10 by 0.01*.

Sample Size

N (Number of Pairs)

Enter a value (or range of values) for the sample size, N , the number of pairs in the study.

You can enter single value such as *100* or a range of values such as *100 200 300* or *100 to 300 by 50*.

Effect Size

k (Number of Categories)

This is the number of possible categories for the outcomes. The paired results for the McNemar-Bowker test are summarized in a $k \times k$ table of counts or proportions. If $k = 2$, then the McNemar-Bowker test reduces to McNemar's test.

Chi-Square Test Degrees of Freedom

The degrees of freedom (df) of the chi-square test is equal to $k(k-1)/2$.

Range

$k \geq 2$.

Effect Size Input Type

Indicate how to specify the discordant proportion ratio sum (DPRS). DPRS is the effect size and for a table with k categories is equal to

$$\text{DPRS} = \text{Sum}_{(i < j)} [(P_{ij} - P_{ji})^2 / (P_{ij} + P_{ji})]$$

where $i, j = 1, \dots, k$. The choices are

- **Enter DPRS (Discordant Prop. Ratio Sum) Directly**
Enter one or more DPRS values. This option allows you to study multiple effect sizes together in a single run.
- **$k \times k$ Table of Counts or Proportions (Spreadsheet)**
This option uses a $k \times k$ table of counts or proportions from the input spreadsheet to calculate DPRS. The values in the table are scaled so that the calculated proportions sum to 1. If you enter counts, individual cell proportions are computed from the counts.
You must specify the k columns in the spreadsheet that contain the $k \times k$ table. This option allows you to study only one effect size at a time.

Tests for Multiple Correlated Proportions (McNemar-Bowker Test of Symmetry)**DPRS (Discordant Proportion Ratio Sum)**

In a matched-pairs study with k possible categories, we define DPRS as

$$\text{DPRS} = \text{Sum}_{(i < j)} [(P_{ij} - P_{ji})^2 / (P_{ij} + P_{ji})]$$

with $i, j = 1, \dots, k$. The test statistic for the McNemar-Bowker test is equal to

$$T_{SM} = N \times \text{DPRS},$$

where N is the total number of paired observations.

You can enter a single value such as *0.1* or a series of values such as *0.1 0.2 0.3* or *0.1 to 0.3 by 0.1* in the range $\text{DPRS} > 0$.

Columns Containing the $k \times k$ Table

Specify the k columns in the spreadsheet that contain the $k \times k$ table of counts or proportions from which the discordant proportion ratio sum (DPRS) will be calculated. DPRS is the effect size and is equal to

$$\text{DPRS} = \text{Sum}_{(i < j)} [(P_{ij} - P_{ji})^2 / (P_{ij} + P_{ji})]$$

Press the Spreadsheet Columns icon (directly to the right) to select the columns and enter the table values.

Press the Input Spreadsheet icon (to the right and slightly up) to view/edit the spreadsheet. Also note that you can view the spreadsheet by selecting “Tools”, then “Input Spreadsheet”, from the menus

Tests for Multiple Correlated Proportions (McNemar-Bowker Test of Symmetry)

Example 1 – Power Analysis

This example will show you how to compute the power of the McNemar-Bowker test with $k = 4$ response categories for sample sizes between 60 and 200 to detect a discordant proportion ratio sum (DPRS) of 0.1. We'll assume a significance level of 0.05.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Multiple Correlated Proportions (McNemar-Bowker Test of Symmetry)** procedure window by expanding **Proportions**, then **Contingency Table (Chi-Square Tests)**, and then clicking on **Tests for Multiple Correlated Proportions (McNemar-Bowker Test of Symmetry)**. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Power
Alpha.....	0.05
N (Number of Pairs).....	60 to 200 by 20
k (Number of Categories)	4
Effect Size Input Type.....	Enter DPRS (Discordant Prop. Ratio Sum) Directly
DPRS (Discordant Prop. Ratio Sum).....	0.1

Annotated Output

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

Numeric Results

Numeric Results for the McNemar-Bowker Test with $k = 4$ Categories

Power*	Pairs N	Discordant Proportion Ratio Sum	
		DPRS	Alpha
0.40283	60	0.100	0.050
0.53065	80	0.100	0.050
0.64385	100	0.100	0.050
0.73803	120	0.100	0.050
0.81256	140	0.100	0.050
0.86917	160	0.100	0.050
0.91070	180	0.100	0.050
0.94026	200	0.100	0.050

* Power was computed using the non-central chi-square distribution with $k(k-1)/2 = 6$ degrees of freedom.

References

Fagerland, M.W., Lydersen, S., and Laake, P. 2017. *Statistical Analysis of Contingency Tables*. Taylor & Francis/CRC. Boca Raton, Florida.
 Chow, S.C., Shao, J., Wang, H., and Lokhnygina, Y. 2018. *Sample Size Calculations in Clinical Research*, Third Edition. Taylor & Francis/CRC. Boca Raton, Florida.

Report Definitions

Power is the probability of rejecting a false null hypothesis. It should be close to one.

N is the number of pairs in the sample.

DPRS (the discordant proportion ratio sum) is the effect size and is equal to $\sum_{(i < j)} [(P_{ij} - P_{ji})^2 / (P_{ij} + P_{ji})]$.

Alpha is the probability of rejecting a true null hypothesis. It should be small.

Tests for Multiple Correlated Proportions (McNemar-Bowker Test of Symmetry)

Summary Statements

A sample size of 60 pairs achieves 40.283% power to detect a ratio sum of 0.100 for discordant proportions from 4 categories using the McNemar-Bowker test of symmetry with a significance level of 0.050.

Dropout-Inflated Sample Size

Dropout Rate	Sample Size	Dropout-Inflated Enrollment Sample Size	Expected Number of Dropouts
	N	N'	D
20%	60	75	15
20%	80	100	20
20%	100	125	25
20%	120	150	30
20%	140	175	35
20%	160	200	40
20%	180	225	45
20%	200	250	50

Definitions

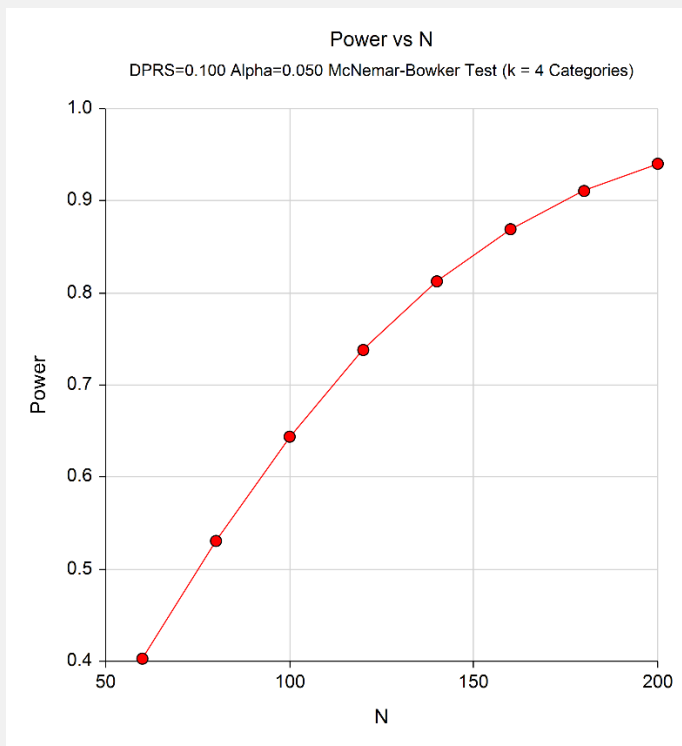
Dropout Rate (DR) is the percentage of subjects (or items) that are expected to be lost at random during the course of the study and for whom no response data will be collected (i.e. will be treated as "missing").

N is the evaluable sample size at which power is computed (as entered by the user). If N subjects are evaluated out of the N' subjects that are enrolled in the study, the design will achieve the stated power.

N' is the total number of subjects that should be enrolled in the study in order to end up with N evaluable subjects, based on the assumed dropout rate. N' is calculated by inflating N using the formula $N' = N / (1 - DR)$, with N' always rounded up. (See Julious, S.A. (2010) pages 52-53, or Chow, S.C., Shao, J., and Wang, H. (2008) pages 39-40.)

D is the expected number of dropouts. $D = N' - N$.

Chart Section



This report shows the values of each of the parameters, one scenario per row. This plot shows the relationship between sample size and power. We see that a sample size of around 135 pairs is required to detect a DPRS of 0.1 with 80% power.

Example 2 – Calculating Sample Size (Validation using Chow, Shao, Wang, & Lokhnygina (2018))

On pages 142-143, Chow, Shao, Wang, & Lokhnygina (2018) presents an example of finding the sample size required for a McNemar-Bowker test with $k = 3$ categories. Note that the authors incorrectly refer to the McNemar-Bowker test as the Stuart-Maxwell test.

A pilot study was conducted to test the effects of a treatment compound on the monocytes in the respondents' blood. Though the number of monocytes was recorded for each respondent, the results were categorized into 3 ranges: below, within, and above normal range. The results are summarized in the following 3×3 table.

Pretreatment	Posttreatment			Total
	Below	Normal	Above	
Below	3	4	4	11
Normal	2	3	3	8
Above	1	2	3	6
Total	6	9	10	25

From this table it appears that the treatment increases the number of monocytes in the blood since the upper off-diagonal counts are all larger than their corresponding lower off-diagonal counts (see colored count pairs). The DPRS for this pilot study was calculated to be 0.107. To verify this the researchers are planning a larger study for which they want to know how many pairs to sample to detect this same magnitude of DPRS with 80% power using a significance level of 0.05. They calculate the required sample size to be 102 pairs.

PASS will calculate the DPRS value for you if you simply enter the 3×3 table counts in the spreadsheet and reference the columns.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Multiple Correlated Proportions (McNemar-Bowker Test of Symmetry)** procedure window by expanding **Proportions**, then **Contingency Table (Chi-Square Tests)**, and then clicking on **Tests for Multiple Correlated Proportions (McNemar-Bowker Test of Symmetry)**. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Sample Size
Power.....	0.80
Alpha.....	0.05
k (Number of Categories)	3
Effect Size Input Type.....	k × k Table of Counts or Proportions (Spreadsheet)
Columns Containing the k × k Table	C1-C3

Tests for Multiple Correlated Proportions (McNemar-Bowker Test of Symmetry)

Input Spreadsheet Data

Row	C1	C2	C3
1	3	4	4
2	2	3	3
3	1	2	3

Output

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

Numeric Results
Numeric Results for the McNemar-Bowker Test with k = 3 Categories

Power*	Pairs N	Discordant Proportion Ratio Sum DPRS	Alpha
0.80335	103	0.107	0.050

* Power was computed using the non-central chi-square distribution with $k(k-1)/2 = 3$ degrees of freedom.

3 x 3 Table of Proportions used to Calculate DPRS

Row	C1	C2	C3	Total
1	0.120	0.160	0.160	0.440
2	0.080	0.120	0.120	0.320
3	0.040	0.080	0.120	0.240
Total	0.240	0.360	0.400	1.000

The results from **PASS** match the results from Chow, Shao, Wang, & Lokhnygina (2018), with a slight difference because the authors compute the sample size based on rounded values, while **PASS** always uses full precision. If you compute the power for 102 pairs, you'll see that the actual power is 79.909%, which is slightly less than the required 80%. A sample size of 103 pairs is needed to achieve 80% power.