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* × *k* contingency table of counts as

	Treatment 2				
Treatment 1	1	2	⋯	k	Total
1	n_{11}	n_{12}	⋯	n_{1k}	$n_{1\cdot}$
2	n_{21}	n_{22}	⋯	n_{2k}	$n_{2\cdot}$
⋮	⋮	⋮	⋱	⋮	⋮
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				
Treatment 1	1	2	⋯	k	Total
1	p_{11}	p_{12}	⋯	p_{1k}	$p_{1\cdot}$
2	p_{21}	p_{22}	⋯	p_{2k}	$p_{2\cdot}$
⋮	⋮	⋮	⋱	⋮	⋮
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

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

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

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

For a significance level α, the test rejects the null hypothesis if $T_{MB} > \chi^2_{1-\alpha,\, k(k-1)/2}$, where $\chi^2_{1-\alpha,\, k(k-1)/2}$ 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, & Lokhnygina (2018) page 141, the power for the McNemar-Bowker test is

$$1 - \chi^2_{k(k-1)/2}\left(\chi^2_{1-\alpha,\, k(k-1)/2} \middle| \delta\right)$$

where $\chi^2_{k(k-1)/2}(X|NCP)$ 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 $\chi^2_{1-\alpha,\, k(k-1)/2}$ 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

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

In the **PASS**, the summed quantity $\sum_{i<j} \frac{\left(p_{ij}-p_{ji}\right)^2}{p_{ij}+p_{ji}}$ 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*.

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

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 1** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

Solve For .... **Power**
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**

Output

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

Numeric Results

Solve For: Power
Number of Categories: 4

Power*	Number of Pairs N	Discordant Proportion Ratio Sum DPRS	Alpha
0.40283	60	0.1	0.05
0.53065	80	0.1	0.05
0.64385	100	0.1	0.05
0.73803	120	0.1	0.05
0.81256	140	0.1	0.05
0.86917	160	0.1	0.05
0.91070	180	0.1	0.05
0.94026	200	0.1	0.05

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

Power The probability of rejecting a false null hypothesis when the alternative hypothesis is true.
N The number of pairs in the sample.
DPRS Discordant Proportion Ratio Sum is the effect size. DPRS = $\Sigma(i < j)\ [(Pij - Pji)^2/(Pij + Pji)]$.
Alpha The probability of rejecting a true null hypothesis.

Summary Statements

A paired design will be used to test the proportion symmetry across response categories. The comparison will be made using a paired-sample McNemar-Bowker Test of Symmetry, with a Type I error rate (α) of 0.05. The number of response categories is 4. To detect a ratio sum for discordant proportions of 0.1 with a sample size of 60 pairs, the power is 0.40283.

Dropout-Inflated Sample Size

Dropout Rate	Sample Size N	Dropout-Inflated Enrollment Sample Size N'	Expected Number of Dropouts 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

Dropout Rate The percentage of subjects (or items) that are expected to be lost at random during the course of the study and for whom no response data will be collected (i.e., will be treated as "missing"). Abbreviated as DR.
N The evaluable sample size at which power is computed (as entered by the user). If N subjects are evaluated out of the N' subjects that are enrolled in the study, the design will achieve the stated power.
N' The total number of subjects that should be enrolled in the study in order to obtain N evaluable subjects, based on the assumed dropout rate. N' is calculated by inflating N using the formula N' = N / (1 - DR), with N' always rounded up. (See Julious, S.A. (2010) pages 52-53, or Chow, S.C., Shao, J., Wang, H., and Lokhnygina, Y. (2018) pages 32-33.)
D The expected number of dropouts. D = N' - N.

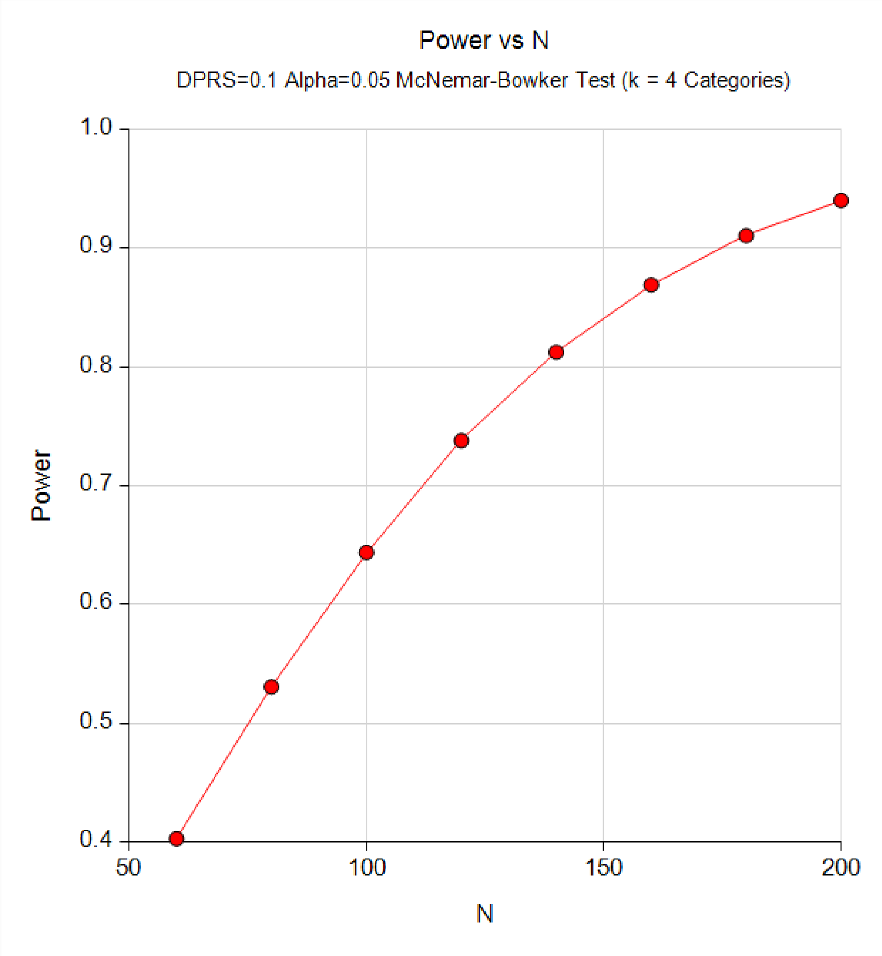
Dropout Summary Statements

Anticipating a 20% dropout rate, 75 subjects should be enrolled to obtain a final sample size of 60 subjects.

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.

Plots



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.

	Posttreatment			
Pretreatment	Below	Normal	Above	Total
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

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 2** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

Solve For .... **Sample Size**
Power.. **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**

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

Solve For: Sample Size
Number of Categories: 3

Power*	Number of Pairs N	Discordant Proportion Ratio Sum DPRS	Alpha
0.80335	103	0.107	0.05

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

3 × 3 Table of Proportions used to Calculate DPRS

Row	C1	C2	C3	Total
1	0.12	0.16	0.16	0.44
2	0.08	0.12	0.12	0.32
3	0.04	0.08	0.12	0.24
Total	0.24	0.36	0.40	1.00

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.