Paper 382-2008

Generalized McNemar's Test for Homogeneity of the Marginal Distributions

Xuezheng Sun[1], Zhao Yang[2]

[1] Department of Epidemiology and Biostatistics
University of South Carolina, Columbia, SC 29208, USA

[2] Premier Research Group plc.
1800 Parkway Place, Suite 820, Marietta, GA 30067, USA

SUMMARY

In the matched-pairs data, McNemar's test (McNemar, 1947) can be applied only to the case in which there are two possible categories for the outcome. In practice, however, it is possible that the outcomes are classified into multiple categories. Under this situation, the test statistic proposed by Stuart (1955) and Maxwell (1970) is useful, it is actually the generalization of the McNemar's test, commonly referred to as generalized McNemar's or Stuart-Maxwell test. There is no public available SAS program to calculate this statistic, the author has developed a SAS macro (the code is detailed in appendix) to perform this test and briefly describes how to use the macro. Examples using the developed SAS macro are also included.

Keywords: Matched-pairs data; Marginal Homogeneity; McNemar's test; generalized McNemar's test (Stuart-Maxwell test); SAS/IML; SAS/Macro.

INTRODUCTION

Matched-pairs data commonly occur in studies with repeated measurement of subjects, such as longitudinal studies that observe subjects over time, e.g. in a 2×2 crossover clinical trial study, we may concern with the changes of interested endpoint between 2 treatment period; or symptom ratings by two raters on the same sample of subjects. Because of the matching, the responses in the two samples are statistically dependent. For matched pairs with a categorical response, a two-way contingency table with the same row and column categories summarizes the data, under this situation, the contingency table is also called square table.

Now, let Y and Z denote the two categorical response variable with r categories for matched-pairs data. Classifications of subjects on both variables have r^2 possible combinations, an example of square table is shown in Table 1 (The number, $n_{i,j}, i, j = 1, 2, \cdots, r$, in the parenthesis is the frequency).

Table 1: Cross-classification of square table.

Y	Z					Total
	1	2	$\cdots$	$r-1$	r	
1	$\pi_{1,1}(n_{1,1})$	$\pi_{1,2}(n_{1,2})$	$\cdots$	$\pi_{1,r-1}(n_{1,r-1})$	$\pi_{1,r}(n_{1,r})$	$\pi_{1,+}(n_{1,+})$
2	$\pi_{2,1}(n_{2,1})$	$\pi_{2,2}(n_{2,2})$	$\cdots$	$\pi_{2,r-1}(n_{2,r-1})$	$\pi_{2,r}(n_{2,r})$	$\pi_{2,+}(n_{2,+})$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$r-1$	$\pi_{r-1,1}(n_{r-1,1})$	$\pi_{r-1,2}(n_{r-1,2})$	$\cdots$	$\pi_{r-1,r-1}(n_{r-1,r-1})$	$\pi_{r-1,r}(n_{r-1,r})$	$\pi_{r-1,+}(n_{r-1,+})$
r	$\pi_{r,1}(n_{r,1})$	$\pi_{r,2}(n_{r,2})$	$\cdots$	$\pi_{r,r-1}(n_{r,r-1})$	$\pi_{r,r}(n_{r,r})$	$\pi_{r,+}(n_{r,+})$
Total	$\pi_{+,1}(n_{+,1})$	$\pi_{+,1}(n_{+,2})$	$\cdots$	$\pi_{+,r-1}(n_{+,r-1})$	$\pi_{+,r}(n_{+,r})$	$\pi_{+,+}(n_{+,+})$

In Table 1, let $\pi_{i,j}, i, j = 1, 2, \cdots, r$ denote the probability that (Y, Z) occurs in the cell in row i and column j. The probability distribution $\{\pi_{ij}\}$ is the joint distribution of Y and Z. The marginal distributions are the row and column totals that result from summing the joint probabilities. We denote these by $\{\pi_{i,+}\}$ for the row variable and $\{\pi_{+,j}\}$ for the column variable, where the subscript "+" denotes the sum over that index; that is

$$\pi_{i,+} = \sum_{j=1}^{r} \pi_{i,j}, \quad \pi_{+,j} = \sum_{i=1}^{r} \pi_{i,j}, \quad \text{and} \quad \sum_{i=1}^{r} \pi_{i,+} = \sum_{j=1}^{r} \pi_{+,j} = \sum_{i=1}^{r}\sum_{j=1}^{r} \pi_{i,j} = 1$$

Therefore, the marginal distributions provide single-variable information.

TEST OF MARGINAL HOMOGENEITY

Marginal homogeneity refers to equality (lack of significant difference) between one or more of the row marginal proportions and the corresponding column proportion(s). Formally, using the notation in Table 1, we are actually testing

$$\text{P}(Y = s) = \text{P}(Z = s) \Rightarrow \pi_{s,+} = \pi_{+,s}, \text{ for } s = 1, 2, \cdots, r \tag{1}$$

Letting $d_s = \pi_{+,s} - \pi_{s,+}$ and letting $\mathbf{d} = (d_1, \cdots, d_{r-1})$, we may see that it is redundant to include d_r in the vector $\mathbf{d}$, since we have $\sum_{i=1}^{r} \pi_{i,+} = 1$ (or $\sum_{j=1}^{r} \pi_{+,j} = 1$) and $\sum_{s=1}^{r} d_s = 0$, hence, generally there are $r - 1$ degree of freedom is the statistics used for testing marginal homogeneity is distributed as a χ^2 distribution. Therefore, marginal homogeneity occurs when the row totals are equal to the column totals, in the medical research, a common interpretation to the "marginal homogeneity" would mean there was no effect of the treatment.

(a) McNemar's Test

In statistics, McNemar's test is a non-parametric method used on nominal data to determine whether the row and column marginal frequencies are equal. It is named after Q. McNemar, who introduced it in 1947. It is applied to 2×2 contingency tables with a dichotomous categories with matched pairs of subjects.

Since there are only two categories, $r = 2$, using the test hypothesis in (1), i.e. $\pi_{1,+} = \pi_{+,1}$, and data structure in Table 1, the McNemar's test statistic is given as

$$Z = \frac{(n_{21} - n_{12})^2}{n_{21} + n_{12}} \sim \chi_1^2 \tag{2}$$

where the McNemar statistic is a chi-squared statistic with 1 degree of freedom. In the small sample situation, continuity-correction will be applied to the above formula, (7). An interesting observation when interpreting McNemar's test is that the elements of the main diagonal contribute no information whatsoever to the decision if pre- or post-treatment condition is more favorable.

The McNemar chi-square test is the procedure of choice assessing marginal homogeneity for repeated dichotomous measures. The research question of interest is whether or not is it reasonable to describe the marginal response rates for, say, a favorable one as equivalent (i.e., homogeneous).

(b) Generalized McNemar/Stuart-Maxwell Test

The generalization of McNemar's test to square tables larger than 2×2 is often referred to as the generalized McNemar or Stuart-Maxwell test (Stuart, 1955; Maxwell, 1970). To test the series of hypothesis (1), we are actually interested in the mean vector and variance-covariance matrix of vector $\mathbf{d}$. Clearly, under the marginal homogeneity, we have $\text{E}(\mathbf{d}) = 0$. And the variance-covariance matrix of vector $\sqrt{N}\mathbf{d}$, $\mathbf{V}$, is $(r - 1) \times (r - 1)$ dimension, the typical elements of $\widehat{\mathbf{V}}$ are

$$\widehat{\nu}_{st} = -(\pi_{st} + \pi_{ts}), \text{ for } t \neq s \text{ and } t, s = 1, \cdots, r - 1 \tag{3}$$
$$\widehat{\nu}_{ss} = \pi_{s+} + \pi_{+s} - 2\pi_{ss}, \text{ for } s = 1, \cdots, r - 1 \tag{4}$$

where $N = n_{+,+}$, the total sample size. Stuart (1955) proposed the statistic

$$Z_0 = N\mathbf{d}'\widehat{\mathbf{V}}^{-1}\mathbf{d} = N\mathbf{d}'\left(N\widehat{\mathbf{V}}\right)^{-1}N\mathbf{d} \sim \chi_{r-1}^2 \tag{5}$$

Since we have the relationship $N(\pi_{st} + \pi_{ts}) = n_{st} + n_{ts}$; $N(\pi_{s+} + \pi_{+s} - 2\pi_{ss}) = n_{s+} + n_{+s} - 2n_{ss}$ hold, the calculation of Z_0 can then be based on the frequency in the $r \times r$ square table. Z_0 is the so-called generalized McNemar's test statistic or Stuart-Maxwell test statistic. And Z_0 is distributed as a chi-squared distribution with $r - 1$ degree of freedom. When $r = 2$, the generalized McNemar's test statistic (or Stuart-Maxwell test statistic) Z_0 will be reduced to McNemar's statistic in (7). For $r = 3$, we also have a closed-form formula (Walker, 2002) to calculate Z_0. After defining the following items, we then can get the Z_0,

$$d_1 = (n_{12} + n_{13}) - (n_{21} + n_{31})$$
$$d_2 = (n_{21} + n_{23}) - (n_{12} + n_{32})$$
$$d_3 = (n_{31} + n_{32}) - (n_{13} + n_{23})$$
$$\overline{n}_{ij} = \frac{n_{ij} + n_{ji}}{2}, \text{ for } i \neq j$$
$$Z_0 = \frac{\overline{n}_{23}d_1^2 + \overline{n}_{13}d_2^2 + \overline{n}_{12}d_3^2}{2(\overline{n}_{12}\overline{n}_{23} + \overline{n}_{12}\overline{n}_{13} + \overline{n}_{13}\overline{n}_{23})} \tag{6}$$

(c) Bhapkar's test

Bhapkar (1966) tested marginal homogeneity by exploiting the asymptotic normality of marginal proportion, and this test is also called Bhapkar's test. The idea of constructing test statistic is similar to the one of generalized McNemar's test statistic, and the major difference lies in the calculation of elements in variance-covariance matrix $\widehat{\mathbf{V}}$, the typical elements are

$$\widehat{\nu}_{st} = -(\pi_{st} + \pi_{ts}) - (\pi_{+s} - \pi_{s+})(\pi_{+t} - \pi_{t+}), \text{ for } t \neq s \text{ and } t, s = 1, \cdots, r-1 \tag{7}$$
$$\widehat{\nu}_{ss} = \pi_{s+} + \pi_{+s} - 2\pi_{ss} - (\pi_{+s} - \pi_{s+})^2, \text{ for } s = 1, \cdots, r-1 \tag{8}$$

The Bhapkar's test statistic (1966) is also

$$Z_1 = N\mathbf{d}'\widehat{\mathbf{V}}^{-1}\mathbf{d} = N\mathbf{d}'\left(N^2\widehat{\mathbf{V}}\right)^{-1}N^2\mathbf{d} \sim \chi_{r-1}^2 \tag{9}$$

And Ireland et al (1969) noticed the relationship between the generalized McNemar's test statistic and the Bhapkar's test statistic as

$$Z_1 = \frac{Z_0}{1 - Z_0/N} \tag{10}$$

Although the Bhapkar and Stuart-Maxwell tests are asymptotically equivalent (Keefe, 1982). Generally, the Bhapkar (1966) test is a more powerful alternative to the Stuart-Maxwell test. With a large N, both will produce the same chi-squared value. As the Bhapkar test is more powerful, it is preferred.

MACRO DEVELOPMENT FOR GENERALIZED McNemar/Stuart-Maxwell TEST

The SAS system provides the easily-accessed calculation for McNemar's test (using option `AGREE` in `TABLE` statement of SAS/STAT procedure `FREQ`) and Bhapkar's test (using the `REPEATED` statement in the SAS/STAT procedure `CATMOD`), however, there are no public available SAS code to perform the calculation and test using generalized McNemar/Stuart-Maxwell test statistic. Some sample SAS code for McNemar's test and Bhapkar's test will be presented in the following section as a comparison among the test results. In this section, a brief summary to the developed macro %`gMcNemar` was presented, the SAS code is detailed in the Appendix.

The developed macro %`gMcNemar` has four parameters

```
%gMcNemar(  DSIN   = ,     /* INPUT DATASET*/
            ROWV   = ,     /* ROW VARAIBLE NAME */
            COLV   = ,     /* COLUMN VARIABLE NAME */
            COUNT  = );    /* CELL COUNT VARIABLE NAME OF R X R SQUARE TABLE*/
```

The parameter `COUNT` corresponds to the n_{ij} in Table 1. And the current version of macro can only be applied to the compiled dataset, however, only some simple modifications are needed to make the macro applicable to the raw dataset, e.g. the raw data of Table 18.2 in Walker's book (2002, p.295)

The output of the macro includes the generalized McNemar's test statistic `GMN`, the degree of freedom of χ^2 distributed generalized McNemar's test statistic `DF` and the corresponding $p-$value `PROBCHI`. Also, the 95% and 99% quantile of χ^2 distribution with degree of freedom `DF` are provided as `QCHI95` and `QCHI99`, respectively.

CASE STUDIES

This part includes three examples of applying the developed macro. To make sure the macro works correctly, all three examples are taken from Stuart (1955) and Walker (2002), in which there are given calculated results. The calculated statistics from the developed macro for the following examples are consistent with the given results in Stuart (1955) and Walker (2002).

(a) Unaided Distance Vision

The data, provided in Table 2, is taken from the example of Stuart (1955). The computation was shown via a step-by-step procedure, and the generalized McNemar's statistic was calculated as 11.96 (Stuart, 1955).

Table 2: 7477 Women Aged 30-39; Unaided Distance Vision.

Right Eye	Left Eye				Total
	Highest Grade	Second Grade	Third Grade	Lowest Grade	
Highest Grade	1520	266	124	66	1976
Second Grade	234	1512	432	78	2256
Third Grade	117	362	1772	205	2456
Lowest Grade	36	82	179	492	789
Total	1907	2222	2507	841	7477

The following code first creates the dataset, then call the developed macro for calculation. We may notice from the output that the resulting generalized McNemar's test statistic is 11.95657, same as the one in Stuart's paper (1955), and the statistic is approximately distributed as χ^2 with degree of freedom 3, the corresponding $p-$value is 0.008. Meanwhile, the 95% and 99% quantile of χ^2 with degree of freedom 3 are also provided in the output as 7.815 and 11.345, respectively.

```
data Vision;
    input r c count @@;
    cards;
    1 1 1520 1 2 266 1 3 124 1 4 66 2 1 234 2 2 1512 2 3 432 2 4 78 3 1 117
    3 2 362 3 3 1772 3 4 205 4 1 36 4 2 82 4 3 179 4 4 492
    ;
%gMcNemar(dsin = Vision, rowv = r, colv = c, count = count);
Output:
    GMN        DF   PROBCHI
11.95657        3 0.0075334
  QCHI95    QCHI99
7.8147279 11.344867
```

To make a comparison, the SAS code and its corresponding output using `PROC CATMOD` to calculate Bhapkar's test statistic was also provided in the following,

```
proc catmod data = Vision;
    weight count;
    response marginals;
    model R * C = _response_ / oneway;
    repeated vis 2 / _response_= vis;
run;
Output:
      The CATMOD Procedure
      Analysis of Variance
```

4

```
Source          DF   Chi-Square    Pr > ChiSq
-------------------------------------------
Intercept        3     78744.17       <.0001
vis              3        11.98       0.0075
```

Clearly, the results (statistic and $p-$value) of Bhapkar's test are equivalent to the generalized McNemar/Stuart-Maxwell test. And the analysis of variance table in Bhapkar's test output shows that there is not marginal homogeneity between left-eye vision and right-eye vision with $p = 0.0075 < 0.05$. In other words, the distribution of the quality of right-eye vision differs significantly from the quality of left-eye vision in the same subjects.

(b) Bilirubin Abnormalities Following Drug Treatment

The data of this example are from Example 18.1 of Walker (2002). There are 86 patients were treated with an experimental drug for 3 months. Pre- and post-study clinical laboratory results showed abnormally high total bilirubin values (above the upper limit of the normal range) as indicated in Table 3. And the research question is "is there evidence of a change in the pre- to post-treatment rates of abnormalities?".

Table 3: Bilirubin Abnormality frequency Summary Following Drug Treatment.

	Post-Treatment		
Pre-Treatment	NO	YES	Total
NO	60	14	74
YES	6	6	12
Total	66	20	86

Clearly, the standard McNemar's test can be applied to the data in Table 3. However, the developed macro can also be used to calculate the standard McNemar's test statistic, in the following code, after creating the analysis dataset, the application of the macro provide a consistent result with the value provided by Walker (2002).

```
data Abnormal;
    input r c count @@;
    cards;
    1 1 60 1 2 14 2 1 6 2 2 6
    ;
%gMcNemar(dsin = Abnormal, rowv = r, colv = c, count = count);
Output:
  GMN        DF   PROBCHI
   3.2         1 0.0736383
 QCHI95    QCHI99
3.8414588 6.6348966
```

The following SAS code is used to calculate the standard McNemar's test statistic, and the results is exact the same as the one calculated from the developed macro.

```
proc freq data = Abnormal;
    weight count;
    tables r*c/agree;
run;
    Output:
  McNemar's Test
-----------------------
 Statistic (S)    3.2000
 DF                    1
 Pr > S           0.0736
```

To make another comparison, the SAS code and its corresponding output using `PROC CATMOD` to calculate Bhapkar's test statistic was also provided in the following. And we may notice from the output, that the Bhapkar's test statistic was calculated as 3.32, and the generalized McNemar's test statistic is 3.2, there is some minor difference between them. However, the conclusion for the data is consistent.

```
proc catmod data = Abnormal;
    weight count;
    response marginals;
    model R * C = _response_ / oneway;
    repeated abnor 2 / _response_= abnor;
run;
Output:
                The CATMOD Procedure
        Analysis of Weighted Least Squares Estimates
                             Standard       Chi-
Effect          Parameter   Estimate    Error      Square   Pr > ChiSq
------------------------------------------------------------------------
Intercept          1         0.8140    0.0329      610.72     <.0001
abnor              2         0.0465    0.0255        3.32     0.0683
```

All the test statistics and their p–value show that there is marginal homogeneity between pre-treatment and post-treatment with $p > 0.05$, i.e. the distribution of the abnormality of pre-treatment does not differ from the abnormality of post-treatment in the same subjects.

(c) Symptom Frequency Before and After Treatment

The data of this example are from Example 18.2 of Walker (2002). Patients characterized their craving of certain high-fat food products before and two weeks after an experimental diet therapy as 'never', 'occasional', or 'frequent', and the data are summarized in Table 4. The research question is "Does the diet appear to have an effect on the frequency of these cravings?"

Table 4: Cell frequency for Symptom Frequency Before and After Treatment.

	Two Weeks			
Pre-study	Never	Occasional	Frequent	Total
Never	14	6	4	24
Occasional	9	17	2	28
Frequent	6	12	8	26
Total	21	33	34	88

Table 4 provides a 3×3 square table, and there is a step-by-step hand calculation to the statistic (Walker, 2002). In the following code, after creating the analysis dataset, the application of the macro provide a consistent result with the value provided by Walker (2002).

```
data Symptom;
    input r c count @@;
    cards;
    1 1 14 1 2 6 1 3 4 2 1 9 2 2 17
    2 3 2 3 1 6 3 2 12 3 3 8
    ;
%gMcNemar(dsin = Symptom, rowv = r, colv = c, count = count);
Output:
  GMN        DF    PROBCHI
    6          2  0.0497871
  QCHI95     QCHI99
5.9914645  9.2103404
```

To make another comparison, the SAS code and its corresponding output using `PROC CATMOD` to calculate Bhapkar's test statistic was also provided in the following. And we may notice from the output, that the Bhapkar's test statistic was calculated as 6.5, and the generalized McNemar's test statistic is 6. We also notice that the Bhapkar's test indicates significant difference ($p = 0.0388$), and the generalized McNemar/Stuart-Maxwell test only suggest marginal significant difference ($p = 0.0498$)

```
proc catmod data=Symptom;
    weight count;
```

```
        response marginals;
        model R * C = _response_ / oneway;
        repeated symp 2 / _response_= symp;
    run;
    Output:
         The CATMOD Procedure
         Analysis of Variance
 Source        DF   Chi-Square   Pr > ChiSq
 -------------------------------------------
Intercept       2      379.45       <.0001
symp            2        6.50       0.0388
```

DISCUSSION

Matched-pair study can be frequently encountered in the medical researching, e.g. a researcher may attempt to determine if a drug has an effect on a particular disease, requiring the same subjects to be included in the before- and after measurements (matched pairs). The results presented in this paper is the most appropriate for nominal response (all the test statistic has degree of freedom as $r - 1$), and for the ordered categories, when r is large and the dependence between classifications is strong, ordinal tests (with degree of freedom as 1) can be much more powerful (Agresti, 2002).

The SAS system implements the computation of McNemar's test statistics as an option (`AGREE`) in the FREQ procedure. A major limitation of this implementation is that the McNemar's test statistics can only be evaluated dichotomous categories. Extensions to the case of multiple categories due to Stuart (1955) have not been implemented in the SAS system. A SAS macro called %gMcNemar was developed in this paper and its application was illustrated. The macro has simple logic structure, and can be modified straightforwardly according to the objective of the user.

Though Bhapkar's test is uniformly powerful than the generalized McNemar's test, the developed macro can be used to calculate the generalized McNemar's test statistic, if needed in practice. Meanwhile, there is another convenient way to calculate the generalized McNemar's test statistic using the relationship in (10), i.e. having calculated the Bhapkar's test statistic using `PROC CATMOD`, we then can use (11) to calculate generalized McNemar's test statistic.

$$Z_0 = \frac{N Z_1}{N + Z_1} \tag{11}$$

The following code for Example **(c) Symptom Frequency Before and After Treatment** is used to show the above idea.

```
    ODS OUTPUT ANOVA = OUTVAR;
    ODS OUTPUT POPPROFILES = SAMPLEN;
    PROC CATMOD DATA=SYMPTOM;
        WEIGHT COUNT;
        RESPONSE MARGINALS;
        MODEL R * C = _RESPONSE_ / ONEWAY;
        REPEATED SYMP 2 / _RESPONSE_= SYMP;
    ODS EXCLUDE ANOVA;
    ODS EXCLUDE POPPROFILES;
    PROC SQL NOPRINT;
        SELECT SAMPLESIZE INTO: NOBS
        FROM SAMPLEN;
    QUIT;
    DATA OUTVAR (KEEP = DF CHISQ PROBCHISQ S_M_STA P_S_M_ST);
        SET OUTVAR (WHERE = (SOURCE NOT IN ("Intercept", "Residual")));
        S_M_STA = CHISQ*&NOBS/(CHISQ + &NOBS);
        P_S_M_ST = 1 - PROBCHI(S_M_STA, DF);
    PROC PRINT DATA = OUTVAR NOOBS;
    RUN;

    Output:
```

```
                  Prob
DF      ChiSq     ChiSq    S_M_STA     P_S_M_ST
 2       6.50    0.0388        6       0.049787
```

The value of variable `ChiSq` is the Bhapkar's test statistic, and its $p-$value is in the variable `ProbChiSq`, also `S_M_STA` is the generalized McNemar's test statistic, and `P_S_M_ST` is the corresponding $p-$value. We may notice that the results from this program are consistent with these shown in Example **(c) Symptom Frequency Before and After Treatment** in the previous section.

References

[1] Agresti A. (2002). *Categorical Data Analysis*, 2nd edition. John Wiley & Sons, Inc., New Jersey.

[2] Bhapkar V.P. (1966). A note on the equivalence of two test criteria for hypotheses in categorical data. *Journal of the American Statistical Association*, **61**: 228 – 235.

[3] Ireland C.T., Ku H.H., and Kullback S. (1969). Symmetry and marginal homogeneity of an $r \times r$ contingency table. *Journal of the American Statistical Association*, **64**: 1323 – 1341.

[4] Keefe T.J. (1982). On the relationship between two tests for homogeneity of the marginal distributions in a two-way classification. *Biometrika*, **69**: 683 – 684.

[5] Maxwell A.E. (1970). Comparing the classification of subjects by two independent judges. *British Journal of Psychiatry*, **116**: 651 – 655.

[6] McNemar Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, **12**: 153 – 157.

[7] Stuart A. (1955). A Test for Homogeneity of the Marginal Distributions in a Two-Way Classification. *Biometrika*, **42**: 412 – 416.

[8] Walker G.A. (2002). *Common Statistical Methods for Clinical Research with SAS Examples*, 2nd edition. SAS Institute Inc., Cary, North Carolina.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Xuezheng Sun
Department of Epidemiology & Biostatistics,
University of South Carolina
Columbia, SC 29208
803-777-8907
E-mail: `sun4@mailbox.sc.edu`

Zhao Yang
Premier Research Group plc.
1800 Parkway Place, Suite 820
Marietta, GA 30067, USA
678-279-4848
E-mail: `tonyyangsxz@gmail.edu`

APPENDIX

```
%MACRO gMcNemar(DSIN    = ,     /* INPUT DATASET*/
                ROWV    = ,     /* ROW VARAIBLE NAME */
                COLV    = ,     /* COLUMN VARIABLE NAME */
                COUNT   = );    /* CELL COUNT VARIABLE NAME OF R X R SQUARE TABLE*/
    PROC FREQ DATA = &DSIN NOPRINT;
        WEIGHT &COUNT;
        TABLES &ROWV * &COLV /OUT = FREQ (DROP = PERCENT);
    PROC SUMMARY DATA = FREQ;
        CLASS &ROWV  &COLV;
        FREQ COUNT;
        OUTPUT OUT = FREQ1 (WHERE = (_TYPE_ > 0 ));
    %GLOBAL LEVEL;
    PROC SQL NOPRINT;
        SELECT MAX(&ROWV) INTO: LEVEL
        FROM FREQ1;
    QUIT;
    DATA    TEMP1 (DROP = &ROWV RENAME = (RC = CT) )
            TEMP2 (DROP = &COLV RENAME = (RC = RT) ) TEMP3;
    SET FREQ1 (DROP = _TYPE_ RENAME = (_FREQ_ = RC) );
        IF &ROWV =  .   AND &COLV NE .  THEN OUTPUT TEMP1;
        IF &ROWV NE .   AND &COLV =  .  THEN OUTPUT TEMP2;
        IF &ROWV NE .   AND &COLV NE .  THEN OUTPUT TEMP3;
    PROC SORT DATA = TEMP1;
        BY &COLV;
    PROC SORT DATA = TEMP3;
        BY &COLV;
    DATA TEMP13;
        MERGE TEMP3 TEMP1;
        BY &COLV;
    PROC SORT DATA = TEMP2;
        BY &ROWV;
    PROC SORT DATA = TEMP13;
        BY &ROWV;
    DATA TEMP123;
        MERGE TEMP13 TEMP2;
        BY &ROWV;
        IF &ROWV = &COLV THEN VIJ = RT + CT - 2*RC;
        DIFF = &ROWV - &COLV;
        IF DIFF > 0 THEN SEQ = COMPRESS(&ROWV||&COLV);
        ELSE SEQ = COMPRESS(&COLV||&ROWV);
    PROC SORT DATA = TEMP123;
        BY SEQ;
    PROC SQL NOPRINT  ;
        CREATE TABLE TEMP4 AS
            SELECT SEQ, (-1)*SUM(RC) AS VIJ
            FROM TEMP123
            WHERE DIFF NE 0
            GROUP BY SEQ
            ORDER BY SEQ;
    QUIT;
    DATA TEMP4 (DROP = I);
        SET TEMP4;
        DO I = 1 TO 2;
            OUTPUT;
        END;
    DATA TEMP1234 (KEEP = &ROWV &COLV VIJ);
        MERGE TEMP123 TEMP4;
```

```
            BY SEQ;
            IF &ROWV < &LEVEL AND &COLV < &LEVEL;
        PROC SORT DATA = TEMP1234;
            BY &COLV &ROWV ;
        PROC SORT DATA = TEMP1 OUT = TEMP1_1(RENAME = (&COLV = &ROWV));
            BY &COLV;
        DATA RC;
            MERGE TEMP2 TEMP1_1;
            BY &ROWV;
            D = RT - CT;
            IF &ROWV < &LEVEL;
        PROC IML;
            USE TEMP1234;
            READ ALL VAR{VIJ} INTO A;
            X = J(&LEVEL - 1,&LEVEL - 1,0); /* CREATE A COLUMN MATRIX OF 1'S */
            %DO I = 1 %TO (&LEVEL - 1);
                X[,&I] = A[(&LEVEL - 1)*&I - (&LEVEL - 2) : (&LEVEL - 1)*&I, 1];
            %END;
            INVX = INV(X);
            CLOSE TEMP1234;
            USE RC;
            READ ALL VAR{D} INTO DIJ;
            GMN = DIJ' * INVX * DIJ;
            DF = &LEVEL - 1;
            QCHI95 = CINV(0.95,(&LEVEL - 1));
            QCHI99 = CINV(0.99,(&LEVEL - 1));
            PROBCHI = 1 - PROBCHI(GMN, (&LEVEL - 1));
            PRINT GMN DF PROBCHI, QCHI95 QCHI99;
        PROC DATASETS NOLIST KILL LIBRARY = WORK MEMTYPE = ALL;
        QUIT;
    %MEND;
```