

Implementing Multiple Comparisons on Pearson Chi-square Test for an RxC Contingency Table in SAS®

Man Jin, Forest Research Institute; Binhuan Wang, New York University School of Medicine

ABSTRACT

This paper illustrates a permutation method for implementing multiple comparisons on Pearson's Chi-square test for an RxC contingency table, using the SAS procedure FREQ and a newly developed SAS macro called CHISQ_MC. This method is analogous to the Tukey-type multiple comparison method for one-way analysis of variance.

INTRODUCTION

A common statistical analysis for count data is to compare the frequency distributions of a multinomial outcome of interest for two or more groups. Assume that there are R groups and the categorical outcome variable has C nominal categories. The data for this analysis can be displayed in a contingency table with R rows and C columns. In SAS, the analysis can be performed using the FREQ procedure with /CHISQ or /Pearson option. The output provides the result of using Pearson's Chi-squared test for testing against the null hypothesis that the frequency distributions are the same across R groups.

If the Chi-square test results in a p-value which is smaller than some significance level, say 0.05, then there is strong evidence to state that there is difference between at least two of the groups. In the case when there are only two groups being compared (2xC table), it is easy to conclude that the frequency distributions are different between the two groups. However, in the case where there are three or more groups, there is no follow-up test available in SAS to directly identify the significant differences between pairs of groups.

In this paper, we propose a Tukey-type multiple comparison method for RxC tables based on permutation. The method is analogous to the Tukey-type multiple comparison method for one-way analysis of variance (ANOVA; Tukey, 1953, by option TUKEY in PROC GLM), and methods for Rx2 contingency tables (Elliott and Reisch, 2006, by SAS macro CompProp; Brown and Fears, 1981, by option Fisher in PROC MULTTEST).

As pointed out by SAS Usage Note 22265 "Performing multiple comparisons or tests on subtables following a significant Pearson chi-square" (<http://support.sas.com/kb/22/565.html>), there are no multiple comparison methods directly available in PROC FREQ, but there are a number of approaches to consider, including the following three approaches.

First, correspondence analysis is often used as a way to visualize the non-independence in a table. However, no statistical tests of comparisons are performed. Second, use the WHERE statement to conduct Pearson's Chi-square test on each pair of groups and then use PROC MULTTEST to adjust pairwise p-values. However, there is no SAS macro can implement these steps directly and the methods available are less powerful than the Tukey-type adjustment. Third, multiple comparisons can be done by fitting a Poisson loglinear model with PROC GENMOD, and then constructing custom CONTRAST statements to test pair comparisons. However, the method is model based and is less powerful than the Tukey-type adjustment.

METHOD

To illustrate the newly proposed method, we consider the same example used in the aforementioned SAS Usage Note 22265. The SAS output of the data Melanoma are printed in Table 1, where Type is the group variable (R=4) and Site the outcome variable (C=3). The significant chi-square test, produced by the CHISQ option in PROC FREQ, indicates lack of independence between TYPE and SITE.

First, we develop a macro, which is called PAIRWISE_CHISQ and used as a subroutine in main macro CHISQ_MC, to calculate Chi-square test statistics and p-values for all pairs of groups. Because there are 4 groups, there are $4 \times 3 / 2 = 6$ pairs and consequently there are 6 comparisons. The output of PAIRWISE_CHISQ is shown in Table 2, where the p-values are unadjusted p-values. The most conservative multiple comparison is Bonferroni correction

(e.g., Abdi, 2007), which use significant level $0.05 \times 6 = 0.3$ instead of 0.05 to claim a significance, leading to family-wise-error-rate (FWER) $\alpha = 0.05$.

Obs	Type	site	Count
1	Hutchinson's	Head	22
2	Hutchinson's	Trunk	2
3	Hutchinson's	Extremities	10
4	Superficial	Head	16
5	Superficial	Trunk	54
6	Superficial	Extremities	115
7	Nodular	Head	19
8	Nodular	Trunk	33
9	Nodular	Extremities	73
10	Indeterminate	Head	11
11	Indeterminate	Trunk	17
12	Indeterminate	Extremities	28

Table 1: Melanoma Data. Overall Chi-square Test Statistic=65.81 (df= 6; $p < 0.0001$).

Obs	PAIR_COMPARED	CHI_STAT	DF	P_VALUE
1	Hutchinson's vs Indeterminate	19.8430	2	0.00005
2	Hutchinson's vs Nodular	34.8196	2	0.00000
3	Hutchinson's vs Superficial	63.5139	2	0.00000
4	Indeterminate vs Nodular	1.1688	2	0.55743
5	Indeterminate vs Superficial	5.7295	2	0.05700
6	Nodular vs Superficial	3.2167	2	0.20022

Table 2: Output of Macro PAIRWISE_CHISQ. There are 6 Comparisons.

To describe the Tukey-type adjustment, denote pairwise Chi-squared test statistics as T_1, \dots, T_m , where m is the number of pairs. Under the null hypothesis that there is no difference across groups, $T_i \sim \chi^2_{C-1}$, $i = 1, \dots, m$. Motivated by the idea of Tukey's honest significance test (HSD), we consider the distribution of $Q = \max_{1 \leq i \leq m} T_i$ under the null hypothesis. Denote the upper $\alpha \times 100\%$ percentile of the null distribution of Q as Q_α and use it as critical value, then the family-wise-error-rate is controlled as $\text{Prob}(\cup_{i=1}^m \{T_i > Q_\alpha\}) = \text{Prob}(Q > Q_\alpha) = \alpha$.

In ANOVA, individual test statistic T_i is T-test statistic and Q follows the studentize range distribution under the null hypothesis. However, in the current case where individual test statistic T_i is Chi-square test statistic and the null distribution of Q is unknown. Here we develop a permutation procedure to approximate the null distribution of Q .

After we ungroup the grouped data Melanoma in Table 1, we randomly permute the group assignment. For each random permutation j , we execute macro PAIRWISE_CHISQ, obtain the corresponding pairwise Chi-square test statistics as in Table 2, and calculate the maximum, say Q_j^* , of these Chi-square test statistics. If we repeat such random permutation by many times, say $J=1000$, we can approximate the p-value for T_i by $\#\{Q_j^* \geq T_i, j = 1, \dots, J\}/J$. We call such p-values Tukey-type adjusted p-values.

These steps are written into our main macro CHISQ_MC. This macro has 6 arguments: (1) grouped dataset; (2) group variable; (3) outcome variable; (4) frequency variable; (5) number of permutations; and (6) seed for random permutations. Table 3 is obtained by running:

```
%Chisq_MC(data=melanoma, group=type, outcome=site, count=count, num_perm=5000, seed=100);
```

In output such as Table 3, the original pairwise p-values are shown in the 5th column. The Bonferroni adjusted p-values, which are the multiplications of the original p-values by the number of pairs (and truncated by 1), are shown in the 6th column. The Tukey-type adjusted p-values, which are based on random permutation, are shown in the last column. We add the seed argument in the macro just for reproducibility. Both Bonferroni adjusted p-values and Tukey adjusted p-values can be compared with some α , say $\alpha=0.05$, to control the FWER as α .

Obs	PAIR_COMPARED	CHI_STAT	DF	P_VALUE	P_BONFERRONI	P_TUKEY
1	Hutchinson's vs Indeterminate	19.8430	2	0.00005	0.00029	0.0000
2	Hutchinson's vs Nodular	34.8196	2	0.00000	0.00000	0.0000
3	Hutchinson's vs Superficial	63.5139	2	0.00000	0.00000	0.0000
4	Indeterminate vs Nodular	1.1688	2	0.55743	1.00000	0.9442
5	Indeterminate vs Superficial	5.7295	2	0.05700	0.34199	0.2198
6	Nodular vs Superficial	3.2167	2	0.20022	1.00000	0.5906

Table 3: Final Output of Macro Chisq_MC. The Last Column is Based on 5,000 Permutations.

MACROS

The code for the macro PAIRWISE_CHISQ:

```
%macro Pairwise_Chisq(data=, group=, outcome=, count=);
  proc freq data=&data noprint;
    table &group/out=out_group;
  run;
  data groupID;
    set out_group;
    GID=_n_;
  run;
  data _null_;
    set groupID nobs=nobs;
    call symput('num_group', compress(put(nobs, 11.)));
    stop;
  run;
  data ChiTest; *store pairwise Chisquare statistics to this;
    stop;
  run;
  %Do group1=1 %to &num_group-1; *for each pair obtain chisquare stat;
    %Do group2=&group1+1 %to &num_group;
      data group_pair;
        set groupID;
        if GID EQ &group1 then call symput('name_group1', type);
        if GID EQ &group2 then call symput('name_group2', type);
      run;
      proc freq data=&data noprint; output out=CTout PCHI;
        tables &group*&outcome /chisq cellchi2 nopercnt;
        where &group in ("&name_group1", "&name_group2");
        weight &count;
      run;
```

```

        data CTout1;
            set CTout;
            PAIR_COMPARED="&name_group1"||" vs "||"&name_group2";
        run;
        data ChiTest;
            set ChiTest CTout1;
        run;
    %End;
%End;
data ChiTest; *rename variables for final results;
    set ChiTest;
    CHI_STAT=_PCHI_; DF=DF_PCHI; P_VALUE=P_PCHI;
    keep pair_compared Chi_stat DF p_value;
run;
%mend Pairwise_Chisq;

```

The code for macro CHISQ_MC:

```

%macro Chisq_MC (data=, group=, outcome=, count=, num_perm=, seed=);
    options nosource nonotes;
    data data_ungrouped; *convert grouped data to ungrouped data;
        set &data;
        keep &group &outcome;
        do i=1 to &count;
            output;
        end;
run;
*** excute submacro pairwise_chisq on the original data;
%pairwise_chisq(data=&data, group=&group, outcome=&outcome, count=&count);
data ChiTest_original; *save the pairwise chisquare test on original data;
    set ChiTest;
run;
data ChiMax_all; *store max chisq stats from all permutations
    stop;
run;
*** repeart permutations;
%Do perm=1 %to &num_perm;
    data outcome_var; *start to ermute the group assignment;
        set data_ungrouped;
        keep &outcome;
    run;
    data group_var;
        set data_ungrouped;
        random_number=ranuni(&seed);
        call symput('seed', &seed+1);
        keep &group random_number;
    run;
    proc sort data=group_var; *randomly permute group assignment;
        by random_number;
    run;
    data data_permuted; *obtain a randomly permuted data;
        merge group_var outcome_var;
    run;
    proc freq data=data_permuted noprint; *group ungrouped permuted data;
        table &group*&outcome/out=data_permuted_grouped;
    run;
    *execute the submacro pairwise_chisq on permuted data;
    %pairwise_chisq(data=data_permuted_grouped, group=&group,
        outcome=&outcome, count=count);
    Data ChiTest_perm; *save the pairwise chisq stat on a permuted data;
        set ChiTest;
    run;
    proc iml; *calculate the max of pairwise chisq on a permuted data;
        use ChiTest_perm; read all;
        Chi_max=max(Chi_stat);

```

```

        create ChiMax_perm var {Chi_max &perm}; append;
    quit;
    data ChiMax_all;
        set ChiMax_all ChiMax_perm;
    run;
%End; *end of all permutations;
proc iml;
    use ChiTest_original; read all;
    use ChiMax_all; read all;
    n_pair=nrow(p_value);*number of pairs compared;
    *initinals for Bonferroni and Tukey ajusted p-values;
    p_Bonferroni=p_value; p_Tukey=p_value;
    *for each pair calculate adjusted p-values;
    do pair=1 to n_pair;
        p_Bonferroni[pair]=min(p_value[pair]*n_pair,1);
        p_Tukey[pair]=sum(Chi_max>Chi_stat[pair])/nrow(Chi_max);
    end;
    create p_adjusted var{p_Bonferroni p_Tukey}; append;
quit;
data output_final; *THIS IS FINAL OUTPUT;
    merge ChiTest_original p_adjusted;
run;
proc print data=output_final; *PRINT FINAL OUTPUT;
    title "RESULTS of MACRO Chisq_MC";
run;
options source notes;
%mend Chisq_MC;

```

CONCLUSION

The proposed multiple comparison method for an RxC contingency table analysis provides a post hoc test when the overall Chi-square test is significant. The proposed macro CHISQ_MC makes the interpretation of results easier and clearer. The proposed method can also be applied to arbitrary comparisons other than pairwise, and to other test statistics other than Chisq-test.

REFERENCES

- Abdi, H. (2007). Bonferroni and Šidák corrections for multiple comparisons. In N.J. Salkind (ed.). *Encyclopedia of Measurement and Statistics*. Thousand Oaks, CA: Sage.
- Brown, C.C. and Fears, T.R. (1981), Exact significance levels for multiple binomial testing with application to carcinogenicity screens," *Biometrics* 37: 763–774.
- Elliott, A.C. and Reisch, J.S. (2006). Implementing a multiple comparison test for proportions in a 2xC crosstabulation in SAS®. *Proceedings of the SAS User's Group International 31, San Francisco, CA, March 26–29, 2006*. Paper 204-31.
- Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemotherapy Reports* 50: 163–70.
- Šidák, Z. (1967). Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association* 62: 626-633.
- Tukey, J.W. (1953). The problem of multiple comparisons. *Dittoed manuscript of 396 pages, Department of Statistics, Princeton University*.

CONTACT INFORMATION

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

Man Jin
 Forest Research Institute
 Harborside Financial Center, PLAZA 5
 Jersey City, NJ 07311
 Phone: (201) 427-8568
 Email: man.jin@frx.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.
Other brand and product names are trademarks of their respective companies.