http://www.pjbs.org

PJBS

ISSN 1028-8880

Pakistan Journal of Biological Sciences

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Pakistan Journal of Biological Sciences 9 (15): 2798-2801, 2006
ISSN 1028-8880

Calculation of Power in Chi-Square and Likelihood Ratio Chi-Square Statistics by a Special SAS Macro

[1]Taner Özdemir, [2]Sıddık Keskin and [3]Bahattin Çak
[1]Biometry Genetics Unit, Department of Animal Science, Agricultural Faculty,
[2]Biostatistics Department, Faculty of Medicine,
[3]Department of Animal Science, Faculty of Veterinary Medicine,
University of Yüzüncü Yıl, 65080 Van-Turkey

Abstract: The goal of this study was relatively analyzed as to power in Chi-Square and Likelihood Ratio Chi-Square Statistics by using SAS special macro which is presented in Appendix. For the aim, data sets regarding questionnaire responses of 107 refugees were utilized. Contrary to other data sets (had power values with high-level), sample size for only data set 3 having power values with low-moderate level for both statistics were artificially increased from backward to forward and optimum samples sizes for Ch-Square and other were determined as 280 and 170, respectively. As a result, it was concluded that power of Chi-Square and Likelihood Ratio Chi-Square Statistics changed to some factors: the size of sample and combinations of all cells' frequencies of contingency table. Besides, it is possible that researchers can determine sample size which is suitable for each data set by means of special SAS macro in appendix. Moreover, ones should not forget that power concept in any statistic technique means reliability.

Key words: Chi-Square, likelihood ratio Chi-Square, suitable sample size

INTRODUCTION

Chi-Square and Likelihood Ratio Chi-Square statistics have been widely used as criteria of independence and goodness of fit in contingency tables as well as multivariate analysis techniques. Likelihood Ratio Chi-square (which is called also as G statistics) has asymptotically Chi-squared approximation along with was preferred in small samples (Duzguneo *et al.*, 1983; Sokal and Rohlf, 1981; Everitt, 1992). Besides, it was reported that G Statistic were more suitable than other when observed frequencies were less than five (Everitt, 1992; Agresti, 2002). However, the best choice between Chi-Square and Likelihood Ratio Chi-Square statistics based on the size of sample, probability of type error I and power values. Moreover, it should be forgotten that non-significant results for both statistics does not guarantee independence. On the other hands, if power values for both are too-low (for example, a power of 20-40%), the experiment that researcher carried out is not sensitive enough to determine dependent. The power analysis can help researchers decide whether non-significant values of both will be reliable. The most important question for a researcher is How many observations should we survey to ensure statistics having a power of 80-90%.

Therefore, in present study, questionnaire responses of 107 refugees (Özdemir, 2001) were to evaluate and to discuss by using special SAS macro regarding Chi-Square and Likelihood Ratio Chi-Square statistics (SAS, 1998).

MATERIALS AND METHODS

Materials: Questionnaire responses of 107 refugees on their cases of psychological and social life were used for research material. The information various data sets as contingency tables were showed in Table 1.

Table 1: Three contingency tables on 107 refugees

Data set 1: Contingency Table 1 of psychological case by sex		
	Sex (X)	
N = 107		
Psychological problem (Y)	Female (1)	Male (0)
Presence (1)	22	54
Absence (0)	0	31
Data set 2: Contingency Table 2 of psychological case by turning to their country		
	(X)	
N = 107		
Psychological problem (Y)	I want (1)	I don't want (0)
Presence (1)	69	7
Absence (0)	31	0
Data set 3: Contingency Table 3 of marital status by sex		
	Sex	
N = 107		
Material status (Y)	Male (1)	Female (0)
Married (1)	47	20
Single (0)	38	2

Corresponding Author: Taner Özdemır, Biometry Genetics Unit, Department of Animal Science, Faculty of Agriculture, University of Yüzüncü Yıl, 65080, Van-Turkey

Data sets were performed by using special SAS macro. (http://ftp.sas.com/techsup/download/stat/powerrxc.html).

Methods: The notation of Chi-Square (1) and Likelihood Ratio Chi-Square statistics (2) are given below (Everitt, 1992; Agresti, 2002; Eyduran and Özdemir, 2005):

$$\chi^2 = \sum \frac{(f - f_i)^2}{f_i} \quad (1)$$

$$G = 2\sum f.\ln\left(\frac{f}{f_i}\right) \quad (2)$$

Where, f, observed frequency and f_i, expected frequency.

Power theory for Chi-Square and G statistics: Assume that H_0 is the same to model M for a contingency table. Let π_i indicate the true probability in ith cell and Let $\pi_1(M)$ represent the value to which the Maximum likelihood (ML) estimate $\hat{\pi}_i$ for model M converges, where $\sum\pi_i = \sum\pi_i(M) = 1$. For multinomial sample of size n, the non-centrality parameter for Chi-Square (3) can be uttered as follows:

$$\lambda = n\sum_i \frac{[\pi_i - \pi_i(M)]^2}{\pi_i(M)} \quad (3)$$

Expression 3 is the similar form as Chi-Square statistics, with for the sample proportion p_i and $\pi_i(M)$ in place of $\hat{\pi}_i$. The non-centrality parameter for Likelihood Ratio Chi-Square Statistics (4) can be written in this manner:

$$\lambda = 2n\sum_i \pi_i \log\frac{\pi_i}{\pi_i(M)} \quad (4)$$

RESULTS AND DISCUSSION

The values and power values of Chi-square and Likelihood Ratio Chi-Square statistics calculated for Table 1 are showned in Table 2. As seen in Table 2, the statistic and power values of Likelihood Ratio Chi-Square for contingency Table 1 (Data Set 1) were much larger than those of Chi-Square. Because, power of Likelihood Ratio Chi-Square statistic were more advantageous than that of other when observed frequencies were less than five (Duzguneo *et al*., 1983; Everitt, 1992; Agresti, 2002; Sokal and Rohlf, 1981). It could be suggested, therefore, sample size for data set 1 was enough and the data set was more reliable.

Considered corresponding values for contingency Table 2, both statistics were close on each other. Power of Likelihood Ratio Chi-Square statistic were less

Table 2: The values, power values and contingency coefficient of G and Chi-Square statistics in each data set alpha = 0.05

	LR Chi-Square statistic value	LR Chi statistic probability	Chi-Square statistic value	Chi-Square statistic probability	LR Chi statistic power value	Chi-Square power value
Data set 1	17.2735	<0.0001	11.2963	0.0008	0.98596	0.91940
Data set 2	11.1619	0.0008	9.4701	0.0021	0.91636	0.86809
Data set 3[a]	4.9866	0.0255	3.0551	0.0805	0.60762	0.41613

[a]Warning: 50% of the cells have expected counts less than 5. Chi-Square may not be a valid test

Table 3: The power values of Chi-Square and Likelihood Ratio Chi-Square statistics obtained by artificially increasing sample size for data set 3 (alpha = 0.05)

Sample size	Chi-Square statistic power value	Likelihood ratio Chi-Square statistic power values	Sample size	Chi-Square statistic power value	Likelihood ratio Chi-Square statistic power values
20	0.11755	0.16170	210	0.68749	0.87868
30	0.15242	0.21925	220	0.70759	0.89289
40	0.18762	0.27649	230	0.72664	0.90558
50	0.22290	0.33258	240	0.74467	0.91688
60	0.25803	0.38690	250	0.76170	0.92694
70	0.29283	0.43897	260	0.77777	0.93587
80	0.32712	0.48845	270	0.79292	0.94378
90	0.36076	0.53511	280	0.80718	0.95077
100	0.39363	0.57882	290	0.82058	0.95695
110	0.42564	0.61952	300	0.83317	0.96239
120	0.45670	0.65722	310	0.84498	0.96719
130	0.48675	0.69197	320	0.85605	0.97140
140	0.51574	0.72386	330	0.86642	0.97510
150	0.54365	0.75301	340	0.87612	0.97834
160	0.57043	0.77956	350	0.88518	0.98118
170	0.59609	0.80365	360	0.89364	0.98366
180	0.62062	0.82545	370	0.90153	0.98583
190	0.64402	0.84512	380	0.90889	0.98772
200	0.66630	0.86281	390	0.91575	0.98937

advantageous for contingency Table 2 than that of other. The findings were consistent with those reported be other authors when observed frequencies of each cell in it were less than five (Duzguneo *et al*., 1983; Everitt, 1992; Agresti, 2002; Sokal and Rohlf, 1981). However, probability of Likelihood Ratio Chi-Square Statistics was only significant.

On condition that contingency coefficient for the contingency Table 3 was fixed for databases when we artificially increased 20 to 390 by 10 by using special SAS macro mentioned above in order to determine or obtain sufficient sample size for the third contingency table, minimum sample should be 280D for Chi-square and 170 for other (Table 3).

However, if sample size were 390, the power values of Chi-square and G statistics would be achieved to 91.575 and 98.937% (Table 3). SAS Macro regarding power test for data set 3 and others is given in Appendix.

Appendix: Special SAS macro which were used for databases were downloaded from the web site http://ftp.sas.com/ techsup/download/stat/powerrxc.html

```
data aa;
        do row =1 to 2; do col=0,1;
            input freq @@;
            do I= 1 to freq;
                drop i freq;
                output;
            end;
        end; end;
        cards;
47  20
38  2
        ;
        %powerRxC (row = row, col = col, nrange = 20 to 400 by 10)
%macro powerRxC(
                data = _last_,          /*  input data set                          */
                row =,                  /*  the row variable_ REQUIRED              */
                col =,                  /*  the column variable_ REQUIRED           */
                count =,                /*  the variable of frequency counts,
                                            if the input data are cell counts
                                            of a table                              */
                level = 0.05,           /*  the level of the test                   */
                nrange =                /*  the sample size or range of sample
                                            sizes for which power is desired.
                                            If not specified, the actual sample
                                            size is used. Examples:
                                                nrange=20 to 200 by 20
                                                nrange=%str(20,50,100,140)
                                                nrange=%str(20, 50 to 100 by 10)
                                            Note that %STR( ) should be used when
                                            commas appear in your range
                                            specification.                          */
            ) ;
options nonotes;
%let lastds=&syslast;
%if %bquote(&row)=%then %do;
   %put ERROR: The ROW=argument must be specified.;
   %goto exit;
%end;
%if %bquote(&col)=%then %do;
   %put ERROR: The COL=argument must be specified.;
   %goto exit;
%end;
proc freq data=&data;
   %if %bquote(&count) ne %then weight &count%str(;);
   tables &row * &col / chisq out = _ cells;
   output ou t= _chi pchi lrchi;
   run;
data_power;
   merge_cells_chi;
   if_n_=2 then stop;
   sampsize=100*count/percent;
   do n= %if %bquote(&nrange)=%then sampsize; %else &nrange;
      powerp=1_probchi(cinv(1_&level, df_pchi), d_pchi, _pchi_*n/sampsize);
      powerlr=1_probchi(cinv(1_&level, df_lrchi), df_lrchi, _lrchi_*n/sampsize);
      keep n powerp powerlr; output;
   end;
```

```
   run;
proc print noobs split='/';
      label powerp="Power of/Pearson/Chi_square"
            powerlr="Power/of L.R. /Chi_square";
   title "Approximate Power of Chi_square Tests for Independence";
   title2 "Test Level=&level";
   run;
title;
%exit:;
options notes_last_=&lastds;
title;
%mend powerRxC;
```

CONCLUSIONS

The goal of this study was comparatively analyzed as to power in Chi-Square and Likelihood Ratio Chi-Square Statistics by using SAS special macro in appendix section. Power for two statistics based on sample size, the case on whether observed frequencies in each cells of contingency table are less than five. Sample sizes for the initial two contingency tables (data set 1 and 2) were sufficient in place of two statistics because corresponding power values were larger than 86%.

However, sample size for the contingency Table 3 insufficient. Provided that contingency coefficient of two statistics in third contingency table was fixed, to obtain a reliable result (or power value of 80%), sample size for G and Chi-Square statistics in the third contingency table should be minimum 170 and 280 optimum sample size calculated by SAS macro which is presented in appendix, respectively.

As a result, to make very appropriate decision on the best choice between Chi-Square and Likelihood Ratio Chi-Square Statistics should be performed power analysis. Researchers should not forget that power concept in any statistic technique means reliability.

ACKNOWLEDGMENT

The authors wish to thank Ecevit EYDURAN research assistant for technical assistance.

REFERENCES

Agresti, A., 2002. Categorical Data Analysis. 2nd Edn., Wiley, New York.

Duzguneo, O., T. Kesici and F. Gurbuz, 1983. Statistics Methods I. 1st Edn., University of Ankara Publishings of Agriculture Faculty, Ankara, Turkey, Pages: 229.

Everitt, B.S., 1992. The Analysis of Contingency Tables. 2nd Edn., Chapman and Hall, London

Eyduran, E. and T. Özdemir, 2005. Examining Chi-Square, Likelihood Ratio Chi-Square and Independent Ratios in 2×2 Tables: Power of Test. International Congress on Information Technology in Agriculture, Food and Environment-itafe'05. Proceedings. Adana-turkey

Eyduran, E., T. Özdemir and M. Küçük, 2005. Chi-Square and G Test in Animal Science. J. Fac Vet. Med. UYY., pp: 1-3.

http://ftp.sas. com/techsup/download/stat/powerrxc.html

Özdemir, T., 2001. The Problems of Asylum Seeker Waiting for Refugees or Accepted as Refugees by Union National High Commissary Refugees in Van, in Turkey. Master Thesis, Van-turkey.

SAS, 1998. SAS Institute, Inc. Cary, NC, USA.

Sokal, R.R. and F.J. Rohlf, 1981. Biometry. The Principles and Practice of Statistics in Biological Research. W.H. Freeman and Company. New York.