Conditional symmetry model as a better alternative to Symmetry Model for rater agreement measure

A. O. Adejumo, C. Heumann and H. Toutenburg
Department of Statistics
Ludwig-Maximilians-Universität, München
Ludwigstr. 33, D-80539 Munich,
Germany

26th May 2004

Abstract

In almost all life or social science researches, subjects are classified into categories by raters, interviewers or observers. Many approaches have been proposed by various authors for analyzing the data or the results obtained from these raters. Symmetry and conditional symmetry models are models designed for square tables like the one arising from the raters results. Conditional symmetry model which possessed an extra parameter for the off-diagonal cells is a special case to symmetry. In this research work, we examined the effect of the extra parameter introduced by conditional symmetry model over that of symmetry on structure of agreement as well as their fittings. Generalized linear model (GLM) approach was used to model the loglinear model forms of these models with empirical examples. We observed that conditional symmetry based on it extra parameter gave a tremendous improvement to the significant level of the test statistics over that of its symmetry model counterpart, hence conditional symmetry model is better for raters agreement modelling which require symmetric table.

keywords: Agreement, symmetry, conditional symmetry, raters, loglinear.

1 Introduction

Symmetry and conditional symmetry models were originally proposed for any square table. For agreement measurement square contingency table can be used to display joint ratings of two raters. Testing symmetry model is an important preliminary analysis for other analysis which require symmetric table. Conditional symmetry model is a special case to symmetry that introduced a special parameter to preserve the off-diagonal cells. The main objective of this paper is to observe the effect of this extra parameter in conditional symmetry on the structure of agreement as well as the goodness of fit statistics over that of symmetry model.

In nearly all the researches that involves ratings, measurements or diagnosis of subjects by various raters, observers or Pathologists, the researchers are already awared that the most important measurement of error or bias is the raters involved in such studies (Fletcher and Oldham, 1962). At the initial stage of the experiment reliability test or studies has to be conducted among the raters, interviewers or observers involved to assess the level of raters variability in the measurement procedure to be used in data acquisition. The data arising from such studies can be of any form like, quantitative or qualitative, depending on the measuring scales involved in such studies. In a situation where the data arising from such studies are quantitative, tests for interobserver bias and measures of interobserver agreement are usually obtained from standard analysis of variance (ANOVA) mixed models or random effects model. Hypothesis testing of observers effects are used to investigate differences in the mean response among observers and estimates of intraclass correlation coefficients are used to measure interobserver reliability (Chinchinlli et al. 1996). Improvement have been made on standard ANOVA model as proposed by Crowder and Hand (1990) and Toutenburg (2002). However, when the arising data is qualitative in nature, that is categorical data in which the response variable is classified into either nominal or ordinal categories. For nominal data, as reviewed by landis and koch (1975a, 1975b), Banerjee et al. (1999) and Shoukri (2004), a large numbers of estimation and testing procedures like the Cohen Kappa (Cohen, 1960), the weighted Kappa (Cohen, 1968), the intraclass Kappa (Block and kraemer, 1989), Tetrachoric correlation coefficient, Weighted least squares (WLS) method for correlated kappa (Barnhart and Williamson 2002) and many other improved methods on kappa statistic.

Later, some authors such as Darroch and McCloud (1986); Tanner and Young (1985a) pointed out the shortcomings of kappa liked statistics, these include, loss of information from summarizing the table by a single number; sensitivity of value to the form of the marginal distributions; and subsequence dangers in comparing values of kappa between two tables. To this effect, some methods that involve modelling loglinear and latent class models for raters agreement measures are proposed by various authors such as Tanner and Young (1985a), Agresti (1992); Graham (1995); and others. On the side of ordinal data, most of the medical diagnoses data often involve responses taken on an ordinal scale and many of which are very subjective. As it is pointed out by some authors with ordinal data, an intermediate category will often be subjective to more misclassification than an extreme category because there are two directions in which to err away from the extremes. Therefore a modified kappa statistic which accounts for severity of discordance or size of discrepancy is better suited for ordinal data like the weighted kappa (Cohen, 1968) that offers such modification. Also for modelling ordinal data with loglinear and latent class models, see Tanner and Young (1985b), Agresti (1988); Agresti and Lang (1993); Barnhart and Williamson 2002 and others for more details. A square table can be used to display joint ratings of two raters or observers. Two matters are usually considered for this type of table. Firstly, one can analyze differences in the marginal distributions. For ordered response categories, there is usually interest in whether classifications by one rater tend to be higher than those by the other rater. Secondly, one can analyze the extent of subject-wise agreement between raters, which involves investigating the frequency of main-diagonal occurrence within the joint distribution of the ratings. As we have said our major aim is to examine the effect of the extra parameter incorporated in conditional symmetry on the assessment of the structure of raters agreement on their ratings compare with the ordinary symmetry model. We shall have the algorithms for modelling these models written in accordance with generalized linear model (GLM) techniques for modelling loglinear model. These algorithms can be executed with any of Splus or SPSS program. In sections 2 and 3 we present modelling raters agreement and description for GLM respectively and we have empirical examples in section 4. And lastly in section 5 we have summary of results and conclusion.

2 Modelling raters agreement measure

Tanner and Young (1985a) proposed a modelling structure of agreement, by considering loglinear models to express agreement in terms of components, such as chance agreement and beyond chance agreement. Using the loglinear model approach one can display patterns of agreement among several observers, or compare patterns of agreement when subjects are stratified by values of a covariate. Assuming there are n subjects who are related by the same k raters $(k > 2)$ into m nominal categories, Tanner and Young (1985a) express chance agreement, or statistical independence of the ratings, using the following loglinear model representation:

$$\log(v_{ij\ldots l}) = u + u_i^{R_1} + u_j^{R_2} + \ldots + u_l^{R_k}, \quad i, j, \ldots, l = 1, \ldots, m \tag{2.1}$$

where $vij \ldots l$ is the expected cell counts in the $ij \ldots lth$ cell of the joint k-dimensional cross classification of the ratings, u is the overall effect, u^{R_i} is he effect due to categorization by the kth rater in the cth category $(k = 1, \ldots, K; c = 1, \ldots, m)$, and $\sum_{i=1}^{m} u_i^{R_1} = \ldots = \sum_{l=1}^{m} u_l^{R_k} = 0$. They also presented a useful generalization of the independence model that incorporated agreement beyond chance in the following fashion:

$$\begin{aligned} \log(v_{ij\ldots l}) &= u + u_i^{R_1} + u_j^{R_2} + \ldots + u_l^{R_k} + \delta_{ij\ldots l}, \\ & \quad i, j, \ldots, l = 1, \ldots, m. \end{aligned} \tag{2.2}$$

The additional term $\delta_{ij\ldots l}$ represents agreement beyond chance for the $ij\ldots lth$ cell.

Agresti (1988) also proposed a model in form of loglinear model for agreement plus linear-by-linear association, which is the combination of Tanner and Young (1985a)' s model and the uniform association model of Goodman (1979) for bivariate cross-classifications of ordinal variables. The model is

$$\begin{aligned} \log(v_{ij}) &= u + u_i^{R_1} + u_j^{R_2} + \beta\lambda_i\lambda_j + \delta_{ij}, \\ & \quad i, j, \ldots, l = 1, \ldots, m \end{aligned} \tag{2.3}$$

where $\lambda_1 < \ldots < \lambda_m$ are fixed scores assigned to the response categories, and the u's and v's are as defined in equation (2.1).

Following the same structure of loglinear model as presented above by Tanner and Young (1985a); Agresti (1988) in their respective works, we shall also use the loglinear models forms of symmetry and conditional symmetry models to model the structure of agreement by presenting the structures and properties of the models. These two models are originally designed for any square tables. Also due to the fact that testing symmetry model is an important preliminary analysis for other analysis which require symmetric table, we shall observe how conditional symmetry model stands as a better alternative to ordinary symmetry model.

For a given m×m contingency table, let π_{ij} be the probability of cell i, j. There exist symmetry if

$$\pi_{ij} = \pi_{ji} \tag{2.4}$$

whenever $i \neq j$. Let v_{ij} be the expected value of the cell i, j as defined before, such that

$$v_{ij} = n\pi_{i,j} \tag{2.5}$$

In a 2× 2 table, the hypothesis of marginal homogeneity ($n_{i+} = n_{+i}$) is equivalent to the hypothesis of symmetry ($n_{ij} = n_{ji}$). However for more general square (m×m) tables, symmetry is a much stronger hypothesis for agreement than marginal homogeneity.

Symmetry model as loglinear model (Agresti, 1990) is

$$\log(v_{ij}) = u + u_i^{R_1} + u_j^{R_2} + u_{ij}^{R_1R_2}, \quad i, j, = 1, \ldots, m \tag{2.6}$$

where

$$u_{ij}^{R_1R_2} = u_{ji}^{R_1R_2}, \ \sum_{j=1}^{m} u_j^{R_1} = 0,$$

and

$$\sum_{j=0}^{m} u_{ij}^{R_1R_2} = 0 \ for \ i = 1, 2, \ldots, m.$$

There are no superscripts on the main or marginal effect terms because they are the same for rows and columns, that is, $u_i^{R_1} = u_j^{R_2}$ when $i = j$. In other words, the row and column margins are equal, that is, $v_{i+} = v_{+i}$ The likelihood equations are

$$\widehat{v}_{ij} + \widehat{v}_{ji} = n_{ij} + n_{ji} \quad for\ all\ i\ and\ j.$$

The kernel of the log-likelihood is

$$\sum_{ij} \log(v_{ij}) = n_{++}u + \sum_{i}(n_{i+} + n_{+i})u_i + \sum_{ij}\{\frac{n_{ij} + n_{ji}}{2}\}u_{ij}. \tag{2.7}$$

Maximizing this equation yields the following expected cell values

$$\widehat{v}_{ij} = \begin{cases} \frac{n_{ij}+n_{ji}}{2} & \text{If} \quad i \neq j, \\ n_{ii} & \text{if} \quad i = j, \end{cases} \tag{2.8}$$

The goodness of fit statistics for symmetry model are: Pearson's chi-square statistic is

$$\chi^2 = \sum_{i>j} \frac{(n_{ij} - n_{ji})^2}{n_{ij} + n_{ji}}, \tag{2.9}$$

Likelihood ratio statistic is

$$G^2 = \sum_{i \neq j} n_{ij} \log\left(\frac{2n_{ij}}{n_{ij} + n_{ji}}\right). \tag{2.10}$$

The degrees of freedom (df) for the residuals of these statistics can be obtained as (number of cells) minus (number of non-redundant parameters) or (number of off-diagonal cells) minus (number of unique parameters), mathematically

$$\begin{aligned} df &= m(m-1) - \frac{m(m-1)}{2} \\ &= \frac{m(m-1)}{2}. \end{aligned}$$

The two statistics have asymptotic χ^2 distribution with the above degrees of freedom under the null hypothesis of independency in the symmetry model.

Also consider conditional symmetry model which was proposed by McCullagh (1978) and has an additional parameter over symmetry, is given as

$$\pi_{ij} = \begin{cases} \theta\phi_{ji} & \text{if} \quad i < j, \\ \phi_{ii} & \text{if} \quad i = j, \\ (2-\theta)\phi_{ij} & \text{if} \quad i > j, \end{cases} \tag{2.11}$$

with $\phi_{ij} = \phi_{ji}$ and $\sum_{i=1}^{m}\sum_{j=1}^{m} \phi_{ij} = 1$ for $i = j = 1, 2, \ldots, m$.

Conditional symmetry model is a palindronic invariant and not a permutation invariant. However, the reverse permutation applied to conditional symmetry model (2.11) yields

$$\pi_{ij} = \begin{cases} (2-\theta)\phi_{m-i+1,m-j+1} & \text{if} \quad i < j, \\ \phi_{m-i+1,m-i+1} & \text{if} \quad i = j, \\ \theta\phi_{m-i+1,m-j+1} & \text{if} \quad i > j, \end{cases} \tag{2.12}$$

with $\phi_{m-i+1,m-j+1} = \phi_{m-j+1,m-i+1} \sum_{i=1}^{m}\sum_{j=1}^{m}\phi_{m-i+1,m-j+1} = 1$ for $i = j = 1,2,\ldots,m$.

The loglinear version of conditional symmetry model (Agresti 1990) is defined as

$$\log(v_{ij}) = u + u_i^{R_1} + u_j^{R_2} + u_{ij}^{R_1R_2} + \tau I(i<j), \tag{2.13}$$

where $u_{ij}^{R_1R_2} = u_{ji}^{R_1R_2}$, $\sum_{j=1}^{m} u_j^{R_j} = 0$, $\sum_{j=1}^{m} u_{ij}^{R_1R_2} = 0$ for $i = 1,2,\ldots,m$ and $I(i<j)$ is the indicator function defined as

$$I(i<j) = \begin{cases} 1 & \text{if} \quad i<j, \\ 0 & \text{if} \quad i \geq j. \end{cases} \tag{2.14}$$

Based on this additional parameter, the model is mainly for ordered classification when symmetry may not hold, often either

$$\pi_{ij} > \pi_{ji} \forall i < j,$$

or

$$\pi_{ij} < \pi_{ji} \forall i < j.$$

The model (2.11) implies that for all $i<j$, if R_1 denote the row number and R_2 the column number of an observation made according to distribution $\{\pi_{ij}\}$, then the conditional interpretation of model (2.13) is

$$P(R_1 = i, R_2 = j / R_1 < R_2) = P(R_1 = j, R_2 = i / R_1 > R_2) = \phi_{ij}.$$

This means the cells probabilities above the main diagonal are mirror image of the cells probabilities below it. Based on this property, the model is called Conditional symmetry model.

The likelihood equations for the model are

$$\widehat{v}_{ij} + \widehat{v}_{ji} = n_{ij} + n_{ji} \quad for\ all\ i\ and\ j.$$

and

$$\sum_{i<j} \widehat{v}_{ij} = \sum_{i<j} n_{ij}$$

or

$$\sum_{i>j} \widehat{v}_{ij} = \sum_{i>j} n_{ij}.$$

The solution of this equations that satisfies the model is

$$\widehat{\tau} = \log\left\{\frac{\sum_{i<j} n_{ij}}{\sum_{i>j} n_{ij}}\right\}$$

$$\widehat{v}_{ij} = \begin{cases} \frac{\exp(\hat{\tau})(n_{ij}+n_{ji})}{\exp(\hat{\tau})+1} & \text{If} \quad i<j, \\ \\ \frac{(n_{ij}+n_{ji})}{\exp(\hat{\tau})+1} & \text{if} \quad i>j \\ \\ n_{ii} & \text{if} \quad i=j,\ for\ i=j=1,2,\ldots,m. \end{cases} \tag{2.15}$$

The residual degrees of freedom (df) is

$$df = \frac{(m+1)(m-2)}{2},$$

which is one less than that of the symmetry because of the additional parameter $\tau I(i < j)$.

Symmetry model is a special case to Conditional symmetry if $\tau = 0$.

3 Fittings of the models

Generalized linear model (GLM) procedure would be used to fit these models. GLM has three important components, the random, the systematic and the link components, for detail see McCullagh and Nelder (1989). Poisson sampling is mostly assumed when fitting GLM to categorical data with $m > 2$. The log likelihood function is

$$l(\theta, \phi) = \sum_{i=1}^{n} \left(\frac{y_i\theta_i - b(\theta_i))}{\phi} + c(y_i, \phi) \right) \tag{3.1}$$

where θ subsumes all of the θ_i. It could also be written as a function of β and ϕ because (given the x_i), β determines all the θ_i. The main way of maximizing β is by maximizing (3.1). The fact that $G(\mu_i) = x_i\beta$ suggests a crude approximation estimate: regress $G(y_i)$ on x_i, perhaps modifying y_i in order to avoid violating range restrictions (such as taking log(0)), and accounting for the differing variances of the observations.

Fisher scoring iteration is the widely use technique for maximizing the GLM likelihood over β. The basic step is

$$\beta^{(k+1)} = \beta^k - \left(E\left(\frac{\partial^2 l}{\partial\beta\partial\beta'} \right) \right)^{-1} \frac{\partial l}{\partial\beta} \tag{3.2}$$

which can also be written as,

$$\beta^{(k+1)} = \beta^k + \left(-E\left(l''(\beta^{(k)}) \right) \right)^{-1} l'(\beta^{(k)}) \tag{3.3}$$

where l is the loglikelihood function for the entire sample $y_1, \ldots, y_N$ and the expectations are taken with $\beta = \beta^{(k)}$. This is the same as Newton step, except that Hessian of l is replaced by it's expectation. Fisher scoring simplifies to

$$\beta^{(k+1)} = (X'WX)^{-1}X'WZ \tag{3.4}$$

where W is a diagonal matrix with

$$W_{ii} = (G'(\mu_i)^2 b''(\theta_i))^{-1} \tag{3.5}$$

and

$$Z_i = (Y_i - \mu_i)G'(\mu_i) + x_i\beta \tag{3.6}$$

Both equations (3.5) and (3.6) use $\beta = \beta^{(k)}$ and derived values of $\theta_i^{(k)}$ and $\mu_i^{(k)}$. The iteration (3.4) is known as "iteration reweighted least squares", or IRLS. The weights W_{ii} have the usual interpretation as reciprocal variances: $b''(\theta_i)$ is proportional to the variance of Y_i and the $G'(\mu)$ factor in Z_i is squared in W_{ii}. Fisher scoring may also be written as

$$\beta^{(k+1)} = \beta^{(k)} + (X'WX)^{-1}X'WZ^* \tag{3.7}$$

where

$$Z_i^* = (Y_i - \mu_i)G'(\mu_i).$$

We need to describe the structure of variables involve in the modification of symmetry model to obtain their estimates as stated in the model. To this effect, we need to create variable that takes on a unique value for each diagonal cells and a unique value of each pair of cells. So for example if $I = 5$, the symmetry variables, which are also parameters in the model, we shall have the following,

$$u = \begin{cases} u_1 & \text{if} \quad i = j = 1 \\ u_2 & \text{if} \quad i = j = 2 \\ u_3 & \text{if} \quad i = j = 3 \\ u_4 & \text{if} \quad i = j = 4 \\ u_5 & \text{if} \quad i = j = 5 \\ u_{13} & \text{if} \quad (i,j) = (1,2)\ (2,1) \\ u_{14} & \text{if} \quad (i,j) = (1,3)\ (3,1) \\ \vdots & \qquad \vdots \\ u_{45} & \text{if} \quad (i,j) = (4,5)\ (5,4) \end{cases} \tag{3.8}$$

where u_1 is the intercept. All these variables are treated as nominal variables.

In the case of conditional symmetry model in equation (2.13) which add parameter $\tau I(i < j)$ to the symmetry model to derive its own model, the $\tau I(i < j)$ which is defined by the indicator variable in equation (2.14) above can be expressed as a matrix of dummies as follows

$$\tau I(i < j) = \tau = \left\{ \begin{array}{ccccc} 0 & 1 & 1 & \dots & 1 \\ 0 & 0 & 1 & \dots & 1 \\ 0 & 0 & 0 & \dots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 \end{array} \right\}$$

For example also, if $I = 5$ we have,

$$\tau I(i < j) = \tau = \left\{ \begin{array}{ccccc} 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array} \right\}$$

The Fisher scoring iteration procedure stated above was used to fit each of these models under generalized linear model (GLM) procedure. In other to justify further the effect of the extra parameter τ in (2.13) above, we considered some reasonable and appropriated data sets on this subject.

4 EMPIRICAL EXAMPLES

4.1 Example 1

Consider an opinion poll data, as presented by Agresti (1996), in which 475 people were taken their opinion on early teens (age 14-16) having sex relations and on a man and a woman having sex relations before marriage. The response categories are 1 = *always wrong*, 2 = *almost always wrong*, 3 = *wrong only sometime*, 4 = *not wrong at all*.

Table 4.1: Opinions about teenage sex and premarital sex.

Category	Premarital sex				
Teen sex	1	2	3	4	Total
1	141	34	72	109	356
2	4	5	23	38	70
3	1	0	9	23	33
4	0	0	1	15	16
Total	146	39	105	185	475

$$\begin{aligned} Null\ Deviance &= 765.6051 \textit{ on } 15 \textit{ df} \\ Residual\ Deviance(G^2) &= 378.3651 \textit{ on } 6 \textit{ df} \\ Pearson\ X^2 &= 282.9057 \textit{ on } 6 \textit{ df} \\ Number\ of\ iterations &= 5. \end{aligned}$$

Table 4.2: Parameter estimates under symmetry model on sex opinion poll

Coefficients	Value	Std. Error	t-value
intercept	2.98751172	0.07385859	40.449077
u_2	-1.00216046	0.09138922	-10.965850
u_3	-0.11642906	0.04949709	-2.352241
u_4	0.04200758	0.03000732	1.399911
u_{12}	-0.45254801	0.09022581	-5.015727
u_{13}	-0.16288049	0.03863115	-4.216300
u_{14}	-0.04461578	0.02658803	-1.678040
u_{23}	-0.12686364	0.04289622	-2.957455
u_{24}	-0.06670704	0.02446886	-2.726202
u_{34}	-0.03105128	0.02694036	-1.152593

Matrix of fitted values under symmetry model

$$\begin{pmatrix} 141.0 & 19.0 & 36.5 & 54.5 \\ 19.0 & 5.0 & 11.5 & 19.0 \\ 36.5 & 11.5 & 9.0 & 12.0 \\ 54.5 & 19.0 & 12.0 & 15.0 \end{pmatrix}$$

Table 4.3: Parameter estimates under conditional symmetry model on sex opinion poll

Coefficients	Value	Std. Error	t-value
intercept	3.391479105	0.07401941	45.8187817
u_2	-0.665520967	0.09147869	-7.2751474
u_3	-0.004215895	0.04951628	-0.0851416
u_4	0.098114163	0.03001611	3.2687163
u_{12}	-0.553539856	0.09023404	-6.1344905
u_{13}	-0.117995220	0.03863623	-3.0540047
u_{14}	-0.012554880	0.02659188	-0.4721321
u_{23}	-0.186977832	0.04290237	-4.3582168
u_{24}	-0.038653753	0.02447126	-1.5795571
u_{34}	-0.075936545	0.02694581	-2.8181208
τ	-3.908683758	0.41215595	-9.4835067

$$
\begin{aligned}
Null\ Deviance &= 765.6051\ on\ 15\ df \\
Residual\ Deviance(G^2) &= 14.56913\ on\ 5\ df \\
Pearson\ X^2 &= 18.58348\ on\ 5\ df \\
Number\ of\ iterations &= 5.
\end{aligned}
$$

Matrix of fitted values under conditional symmetry model

$$
\begin{pmatrix}
141.0 & 37.2525 & 71.5639 & 6.8557 \\
0.7475 & 5.0 & 22.5475 & 37.2525 \\
1.4361 & 0.4525 & 9.0 & 23.5279 \\
2.1443 & 0.7475 & 0.4721 & 15.0
\end{pmatrix}
$$

4.2 Example 2.

Consider the data arising from the study reported in Holmquist et al. (1967) that investigated the variability in the classification of carcinoma in situ of the uterine cervix in which seven pathologists were requested to separately evaluate and classify 118 slides into one of the the following five categorical scales based on the most involved lesion: 1=negative; 2=Atypical squamous hyperplasia; 3=carcinoma in situ; 4=squamous carcinoma with early stromal invasion; 5=invasive carcinoma. These pathologists are labelled with letters A, B, C, D, E, F and G. We have the cross-classification for pathologists A and C as follows.

Table 4.4: Cross-classification of pathologists A and G on carcinoma in situ of the uterine cervix of 118 slides

Category	Pathologist G					
Pathologist A	1	2	3	4	5	Total
1	24	2	0	0	0	26
2	7	13	6	0	0	26
3	1	4	32	1	0	38
4	0	1	20	1	0	22
5	0	0	3	1	2	6
Total	32	20	61	3	2	118

Table 4.5: Parameter estimates under symmetry model of pathologists A and G on carcinoma in situ of the uterine cervix

Coefficients	Value	Std. Error	t-value
intercept	-0.84509893	4.0159979	-0.21043311
u_2	-0.83698822	0.1954340	-4.28271570
u_3	-1.01140426	0.3396394	-2.97787652
u_4	-2.52985570	8.6890690	-0.29115383
u_5	-1.51791342	7.1650111	-0.21185081
u_{12}	0.88050952	1.6393824	0.53709832
u_{13}	0.49243374	1.1714119	0.42037626
u_{14}	0.08150217	0.8867777	0.09190823
u_{15}	-0.83623324	3.9215344	-0.21324134
u_{23}	0.55656315	0.6692979	0.83156273
u_{24}	0.35406434	0.5477780	0.64636466
u_{25}	0.13289443	0.4587153	0.28971003
u_{34}	0.08125951	0.3936103	0.20644661
u_{35}	0.02014050	0.3385415	0.05949196
u_{45}	0.10987472	0.2906851	0.37798541

$$\begin{aligned}
Null\ Deviance &= 266.6158\ on\ 24\ df\\
Residual\ Deviance(G^2) &= 32.73577\ on\ 10\ df\\
Pearson\ X^2 &= 26.36917\ on\ 10\ df\\
Number\ of\ iterations &= 7.
\end{aligned}$$

Matrix of fitted values under symmetry model

$$\begin{pmatrix}
24.0 & 4.5 & 0.5 & 0.000152 & 0.000152\\
4.5 & 13.0 & 5.0 & 0.5 & 0.000152\\
0.5 & 5.0 & 32.0 & 10.5 & 1.5\\
0.000152 & 0.5 & 10.5 & 1.0 & 0.5\\
0.000152 & 0.000152 & 1.5 & 0.5 & 2.0
\end{pmatrix}$$

Table 4.6: Parameter estimates under conditional symmetry model of pathologists A and G on carcinoma in situ of the uterine cervix

Coefficients	Value	Std. Error	t-value
intercept	-1.64520863	6.2144122	-0.26474083
u_2	-1.30612304	0.2460425	-5.30852506
u_3	-1.16778254	0.3432746	-3.40189051
u_4	-2.82629059	13.4439360	-0.21022791
u_5	-1.69577435	11.0860291	-0.15296499
u_{12}	1.06381101	2.5359397	0.41949382
u_{13}	0.48932485	1.8114921	0.27012254
u_{14}	0.07917050	1.3639482	0.05804509
u_{15}	-0.93504486	6.0676230	-0.15410398
u_{23}	0.65863911	1.0354410	0.63609527
u_{24}	0.35228379	0.8471149	0.41586304
u_{25}	0.13141065	0.7073756	0.18577211
u_{34}	0.15217859	0.6023274	0.25265096
u_{35}	0.01390902	0.5168706	0.02691006
u_{45}	0.16702541	0.4463700	0.37418602
τ	1.41369334	0.3716652	3.80367418

$$\begin{aligned}
Null\ Deviance &= 266.6158\ \mathit{on}\ 24\ \mathit{df} \\
Residual\ Deviance(G^2) &= 14.44221\ \mathit{on}\ 9\ \mathit{df} \\
Pearson\ X^2 &= 14.81361\ \mathit{on}\ 9\ \mathit{df} \\
Number\ of\ iterations &= 8.
\end{aligned}$$

Matrix of fitted values under conditional symmetry model

$$\begin{pmatrix}
24.0 & 1.7609 & 0.1957 & 0.0000249 & 0.0000249 \\
7.2391 & 13.0 & 1.9565 & 0.1957 & 0.0000249 \\
0.8043 & 8.0435 & 32.0 & 4.1087 & 0.5870 \\
0.0001023 & 0.8043 & 16.8913 & 1.0 & 0.1957 \\
0.0001023 & 0.0001023 & 2.4130 & 0.8043 & 2.0
\end{pmatrix}$$

5 Result and Conclusion

5.1 Result

We present the table of the of the summary results in the previous section by including columns for differences between the respective goodness of fits statistics of symmetry and conditional symmetry models with their degrees of freedom.

Table 5.1: Summary results of the examples in section 4 with columns for differences between SM and CS.

		symmetry			conditional symmetry			difference (SM-CS)		
table	n	G^2_{SM}	χ^2_{SM}	df	G^2_{CS}	χ^2_{CS}	df	G^2_d	χ^2_d	df_d
1	475	378.3651	282.9057	6	14.5691	18.5835	5	363.796	264.3222	1
2	118	32.7358	26.3692	10	14.4422	14.8136	9	18.2936	11.8256	1

From the results in Table 5.1 and figures A.1 to A.4, one can observed that there was a great improvement on both statistics, Likelihood ratio (G^2) and Pearson's chi-square (χ^2), as well as the fits of the data, when the extra parameter τ was included in the model for modelling the ratings of raters in the previous section § 4 for conditional symmetry model, irrespective of the sample sizes or the number of categories involved. The columns of difference also in Table 5.1, showed the major contribution of the parameter τ being included in conditional symmetry model over that of ordinary symmetry model. We also tested these with two other different data on cervical ectopy ratings of two different methods, Visual assessment and Computerized planimetry, rated by two different raters as given by Barnhart and Williamson (2002) and we have the same conclusion.

5.2 Conclusion

In order to preserve the totals in each triangular array but not the marginal distributions in the resulted two-way cross-classified table of the ratings of the raters, we need to fit conditional symmetry model to the data. From the result above we observed that inclusion of the extra parameter τ by the conditional symmetry had a great improvement on the goodness of fits statistics for the cross-classified tables. Conditional symmetry is not only testing the pattern of association, but also whether the distribution between categories is the same whichever rater is better. Therefore, introducing the extra parameter to preserve off-diagonal cells thus leads to significant improvement in fit of the model over the symmetry model for raters agreement modelling which require symmetric table.

6 References

[**1**] Agresti, A. (1988). A model for agreement between ratings on an ordinal scales, *Biometrics*, 44, 539- 548.

[**2**] Agresti, A. (1990). *Categorical Data Analysis*, Wiley, New York.

[**3**] Agresti, A. (1992). Modelling patterns of agreement and disagreement. *Statist. Methods Med. Res.*, 1, 201-218.

[**4**] Agresti, A. (1996). *An introduction to Categorical Data Analysis*, Wiley, New York.

[**5**] Agresti, A. and Lang, J. B. (1993). Quasi-symmetry latent class models, with application to rater agreement. *Biometrics*,49, 131-139.

[**6**] Banerjee, M., Capozzoli, M., Mcsweeney, L., and Sinha, D. (1999). Beyond Kappa: A review of interrater agreement measure. *The Cana. J. of. Statist.*, 27(1), 03-23.

[**7**] Barnhart, H. X, and Williamson, J. M. (2002). Weighted least-squares approach for comparing correlated kappa. *Biometrics*, 58,1012-1019.

[**8**] Block, D. A., and Kraemar, H. C. (1989). 2 x 2 kappa coefficients: Measures of agreement or association. *Biometrics*, 45, 269-287.

[**9**] Chinchilli, V. M., Martel, J. K., Kumanyika, S., and Lloyd, T. (1996). A weighted concordance correlation coefficient for repeated measurement designs. *Biometrics*, 52, 341-353.

[**10**] Cohen, J. (1960). A coefficient of agreement for nominal scales. *Edu. and Psych. Meas.*, 20, 37-46.

[**11**] Cohen, J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psych. Bull.*, 70, 213-220.

[**12**] Crowder, M. J., and Hand, D. J. (1990). *Analysis of repeated measures*, Chapman and Hall.

[**13**] Darroch, J. N., and McCloud, P. I. (1986). Category distinguishability and observer agreement. *Australian J. Statist.*, 28, 371-388.

[**14**] Fletcher, C. M., and Oldham, P. D. (1962). *Diagnosis in group resaerch.* Chapter 2 of Med. Surv and Clini. Trials, ed. L. J. Witts. 2nd edit. Oxford univ. Press, London.

[**15**] Graham, P. (1995). Modeling covariate effects in observer agreement studies: The case of nominal scale agreement. *Statist. Med.*, 14, 299-310.

[**16**] Maclure, M., and Willett, W. C. (1987). Misinterpretation and misuse of the kappa statistic. *Amer. J. Epidemiol.*, 126, 161- 169.

[17] McCullagh, P. (1978). A class of parametric models for the analysis of square contingency tables with ordered categories. *Biometrika*, 65(2), 413-418.

[18] McCullagh, P., and Nelder, J. A. (1989). *Generalized linear models*, 2nd Edit. Chapman and Hall.

[19] Shoukri, M. M. (2004). *Measures of interobserver agreement*. Chapman and Hall.

[20] Tanner, M. A., and Young, M. A. (1985a). Modelling agreement among raters. *J. Amer. Statist. Assoc.*, 80, 175-180.

[21] Tanner, M. A., and Young, M. A. (1985b). Modelling ordinal scale agreement . *Psych. Bull.*, 98, 408-415.

[22] Toutenburg, H. (2002). *Statistical analysis of designed experiments*. Springer.

A Appendix

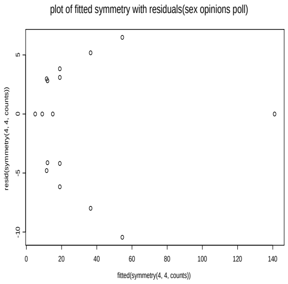

Figure A.1: Fitted symmetry with Residual for sex opinions poll

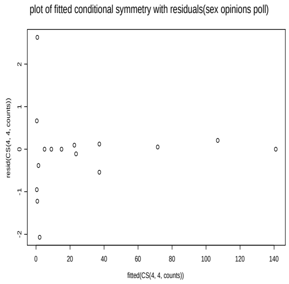

Figure A.2: Fitted conditional symmetry with Residual for sex opinions poll

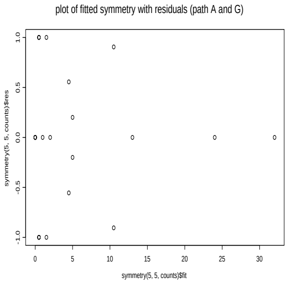

Figure A.3: Fitted symmetry with Residual for path A and G

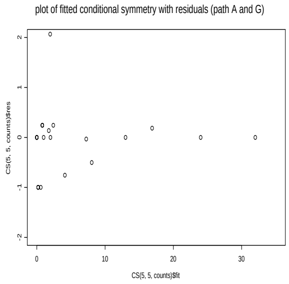

Figure A.4: fitted Conditional Symmetry with residuals for path A and G