

On Quasi-Association Modeling of Agreement between Raters when Ratings are in Ordered Categories

*Adejumo A. O; Sanni, O. O. M; Job, O; Abidoye, A. O

Department of Statistics
University of Ilorin,
Ilorin

*ao123adejumo@yahoo.co.uk.

Abstract

Whenever the categorical scale in the arising cross-classified table from the ratings of two raters or observers is ordered, quasi-independence model rarely fits well. Also, even after conditioning on the event that raters' responses differ, there is often a monotone pattern to the association that remains. One of the major statistical issues in raters' agreement studies is concerned with the extent to which observers classify individual subjects into the same category on the measurement scale. In order to assess the levels of agreement in the ordered categories of the arising cross-classified table, a parametric model like quasi-association (QA) that includes the pattern or level of agreement parameter which permits linear-by-linear association off the main diagonal in its model is required. In addition, when our interest is to identify category(ies) that is of importance to or favoured by the raters out of the entire ordered categories, quasi-association will be suitable. In this paper we model quasi-association for raters' agreement by considering different combination of categories involved in the measurement by step-wise selection procedure using both the overall model of quasi-association and uniform association (UA) models as the bound or baseline. We observed that inclusion of any level of agreement that is significant in the model improved the fit of the model. In addition, inclusion of level of agreement parameter by quasi-association improved the fit. Also when our goal is to identify the most favoured categories out of the existing categories, quasi-association is appropriate.

Keywords: Quasi-association, Quasi-independence, Uniform association, agreement, ordered category, raters.

1.0 Introduction

Modeling was introduced into raters' agreement as a remedy to various problems encountered with some of the kappa like statistics initially proposed for the measurement. Early approaches to this subject have focused on the observed proportion of agreement, suggesting that chance agreement can be ignored. Many authors like Cohen (1960, 1968); Fleiss (1971); Davies and Fleiss (1982); Banerjee et al. (1999); Barnhart and Williamson (2002) and Shoukri (2004) have described various testing procedures and various modifications of kappa-like statistics for raters' agreement measures. Such includes Cohen kappa coefficient, weighted kappa coefficient, intraclass kappa statistic and Weighted least squares (WLS) method for correlated kappa.

Authors like Tanner and Young (1985a,b); Agresti (1988); Block and Kraemer (1989); Oden (1991); Barlow et al. (1991); Shoukri et al. (1995); Barlow (1996); Donner et al. (1996); Banerjee et al. (1999); Adejumo, (2005) presented different

methods and procedures for modeling raters' agreement due to the shortcomings of the kappa-like statistics.

In order to reduce errors, an initial reliability study between the raters or observers is necessary to assess the variability level in their measurement criteria. The assessment of the reliability of rating system has been approached from the perspective of inter-rater agreement. Based on this reliability test that is already in place, it is highly expected that ratings of the raters will be more uniform and correlated. In order to assess the effect and the contribution of the level of agreement parameter of each of the ordered categories, a Quasi-Association model which is a parametric model is needed. It includes some components that have to do with agreement beyond independence in its model and permits linear-by-linear association off the main diagonal in the cross-classified table. In this paper, we model quasi-association to data with ordered categories so as to assess the contribution of the level of each and combination of agreement parameters to the model by step-wise procedures. We also consider uniform association model.

Consequently, we use Generalized Linear Model (GLMs) through loglinear model to structure the pattern of agreement for the models under consideration. In Section 2, we develop the models and their properties while Section 3 contains the modeling and estimation of parameters. Some empirical examples and results as well as conclusion are provided in sections 4 and 5.

2. Models and their properties.

2.1 Uniform association model (UA).

Uniform association (UA) model is also referred to as *linear-by-linear association model* (Birch, 1965; Haberman, 1974; Goodman, 1979a and Agresti, 1988, 1990, 1996). UA is an ordinal loglinear model for two-way tables. The model requires assigning fixed scores $\{u_i, i=1 \text{ to } I\}$ to the rows and $\{v_j, j=1 \text{ to } J\}$ to the columns in order to reflect the category ordering, $u_1 < u_2 < \dots < u_I$ and $v_1 < v_2 < \dots < v_J$. UA model is given as

$$\log(m_{ij}) = \mu + \lambda_i^{R_1} + \lambda_j^{R_2} + \beta u_i v_j \quad (2.1)$$

$\sum_i \lambda_i^{R_1} = 0, \sum_j \lambda_j^{R_2} = 0$, for $i = j = 1, 2, \dots, I$ with R_k representing rater k , $k=1, 2$. This model provides us with the degree of agreement expected when there exists an association between the ratings.

However, independence model which is defined as

$$\log(m_{ij}) = \mu + \lambda_i^{R_1} + \lambda_j^{R_2} \quad (2.2)$$

provides a baseline by showing the degree of agreement expected if there is no association existing between the ratings of the two raters. Model 2.2 is a special case of UA when $\beta=0$. The definition of the scores affects the interpretation of β , which describes the strength of association. When $\beta>0$, there is a tendency for R_2 to increase as R_1 increases. Expected frequencies are larger than expected, under independence, in cells of the table where R_1 and R_2 are both either high or low. On the other hand, when $\beta<0$, there is the tendency for R_2 to decrease as R_1 increases, and for the expected frequencies to be relatively larger in cells of the table where R_1 is high and R_2 is low or where R_1 is low and R_2 is high. In a situation where scores are of equal intervals, we have

$$u_2 - u_1 = u_3 - u_2 = \dots = u_I - u_{I-1} \text{ and } v_2 - v_1 = v_3 - v_2 = \dots = v_J - v_{J-1}.$$

The unit-spaced scores $\{u_i = i\}$ and $\{v_j = j\}$ or their zero-sum equivalents $\{u_i = i - \frac{I+1}{2}\}$ and $\{v_j = j - \frac{I+1}{2}\}$ are simple equal-interval scores.

$\sum_i \beta u_i v_j = \sum_j \beta u_i v_j = 0$ is true for the latter scores.

For the coding $\sum_i u_i = \sum_j v_j = 0$, $\beta u_i v_j$ is regarded as a deviation of $\log(m_{ij})$ from the independence since UA model (2.1) is equivalent to independence model (2.2) plus $\beta u_i v_j$. The deviation is linear in the R_2 scores at a fixed level of R_1 and linear in the R_1 scores at a fixed level of R_2 . The linear-by-linear deviation implies that the model has its greatest departures from independence in the corners of the table.

The kernel of the log-likelihood for uniform association model is

$$L(m) = n\mu + \sum_i n_{i+} \lambda_i^{R_1} + \sum_j n_{+j} \lambda_j^{R_2} + \beta \sum_i \sum_j u_i v_j n_{ij} - \sum_i \sum_j \exp[\mu + \lambda_i^{R_1} + \lambda_j^{R_2} + \beta u_i v_j]. \quad (2.3)$$

By differentiating $L(m)$ with respect to $\lambda_i^{R_1}$, $\lambda_j^{R_2}$ and β , and setting the three partial derivatives equal to zero, we obtain the following likelihood equations:

$$\begin{aligned} \hat{m}_{i+} &= n_{i+} \text{ for } i = 1, 2, \dots, I, \\ \hat{m}_{+j} &= n_{+j} \text{ for } j = 1, 2, \dots, I, \\ \sum_i \sum_j u_i v_j \hat{m}_{ij} &= \sum_i \sum_j u_i v_j n_{ij}, \\ \sum_i \sum_j u_i v_j \hat{\pi}_{ij} &= \sum_i \sum_j u_i v_j p_{ij}, \end{aligned}$$

where $p_{ij} = \frac{n_{ij}}{n}$ and $\hat{\pi}_{ij} = \frac{\hat{m}_{ij}}{n}$

The residual degrees of freedom for UA, since it has an additional parameter $\beta u_i v_j$ over independence model (2.2), is $I(I-2)$.

2.2 Quasi-association model

Quasi association model was also initiated by Goodman (1979a, 1979b) as a better model when we have ordered categories. Often, there is a monotone pattern to the association that remains even after conditioning on the event that the raters' responses differ.

The QA model is given as

$$\log(m_{ij}) = \mu + \lambda_i^{R_1} + \lambda_j^{R_2} + \beta u_i v_j + \delta_i I(i = j), \quad (2.4)$$

$\sum_i \lambda_i^{R_1} = 0, \sum_j \lambda_j^{R_2} = 0$, for $i = j = 1, 2, \dots, I$ where $\{u_i\}$ and $\{v_j\}$ are scores that may either be specified apriori, that is, $\{u_i\}$ and $\{v_j\}$ may be fixed set of ordered scores or estimated from the data. This model provides a parameter for patterns of agreement in ordered categories involved in addition to showing the degree of agreement expected if there is association existing between the ratings of the two raters. The model permits linear-by-linear association off the main diagonal. It is a parsimonious quasi symmetry model. $I(i=j)$ is an indicator function and

$$\delta_i I(i=j) = \begin{cases} \delta_1 & \text{if } i=j=1 \\ \delta_2 & \text{if } i=j=2 \\ \vdots & \vdots \\ \delta_I & \text{if } i=j=I \\ 0 & \text{if otherwise} \end{cases} \quad (2.5)$$

So $\delta_i I(i=j)$ for $i=1,2,\dots,I$ are indicator variables that represent the main diagonal cells which we also refer to as pattern of agreement parameters. The likelihood equations for QA are

$$\hat{m}_{i+} = n_{i+},$$

$$\hat{m}_{+j} = n_{+j}$$

$$\hat{m}_{ii} = n_{ii} \text{ for } i=1,2,\dots,I,$$

$$\sum_i \sum_j u_i v_j \hat{m}_{ij} = \sum_i \sum_j u_i v_j n_{ij},$$

and $\hat{m}_{ij}$ for $i \neq j$ has to be obtained by iterative method.

QA is a special case to quasi independence (QI) model (Goodman, 1968; Caussinus, 1965; Fienberg, 1972; Altham, 1975; and Agresti, 1996) since

$$\log(m_{ij}) = \mu + \lambda_i^{R_1} + \lambda_j^{R_2} + \delta_i I(i=j), \quad (2.6)$$

in which $\beta=0$. It is nested between quasi symmetry and quasi independence. Quasi-symmetry (QS) (Adejumo et al., 2006), as a loglinear model, is defined as

$$\log(m_{ij}) = \mu + \lambda_i^{R_1} + \lambda_j^{R_2} + \lambda_{ij}^{R_1 R_2} \quad (2.7)$$

The residual degrees of freedom for QA (2.4), since it has I additional parameters from $\delta_i I(i=j)$ for $i=1,2,\dots,I$ over UA (2.1), is $(I^2 - 3I)$.

The model when $\beta=0$ and $\delta_i=0$ for all i is the model of statistical independence in (2.2), while the model where only $\beta=0$ is a quasi independence in (2.6) and the model when only $\delta_i=0$ for all i is uniform association model (UA) in (2.1) (Adejumo, 2005).

3.0 Modeling and estimation of parameters

Recall the quasi-association model given in equation (2.4)

$$\log(m_{ij}) = \mu + \lambda_i^{R_1} + \lambda_j^{R_2} + \beta u_i v_j + \delta_i I(i=j).$$

In order to assess the effect of "agreement pattern" parameter of the ordered categories in $\delta_i I(i=j)$ of (2.5), step-wise method of selecting the categories either by forward elimination or backward elimination is implemented.

Different combinations of these ordered categories are be added to uniform association model (2.1) to create different forms of quasi-association models Adejumo, (2005).

According to this rule, the following list of equations are considered depending on the size of category I , ($i=1,2,\dots,I$) which is considered by dropping certain numbers of categories δ_i at a time and using at least one in the model and when eventually there is none, a uniform association model is attained.

Each of these models in Table 3.1 and uniform association model are considered under the GLM assumptions for the estimation of their parameters and fittings.

Table 3.1: Quasi-association models selection for l ordered categories.

Notation	Model
$QA(\delta_1, \delta_2, \dots, \delta_l)$	$UA + \delta_1 + \delta_2 + \dots + \delta_l$
$QA(\delta_2, \delta_3, \dots, \delta_l)$	$UA + \delta_2 + \delta_3 + \dots + \delta_l$
$QA(\delta_1, \delta_3, \dots, \delta_l)$	$UA + \delta_1 + \delta_3 + \dots + \delta_l$
$QA(\delta_1, \delta_4, \dots, \delta_l)$	$UA + \delta_1 + \delta_4 + \dots + \delta_l$
$\vdots$	$\vdots$
$QA(\delta_{l-1})$	$UA + \delta_{l-1}$
$QA(\delta_l)$	$UA + \delta_l$

Table 3.2: Quasi-association models selection for $l=5$ ordered categories.

Notation	Model	Dropped $\delta_i(s)$
$QA(\delta_1, \delta_2, \delta_3, \delta_4, \delta_5)$	$UA + \delta_1 + \delta_2 + \delta_3 + \delta_4 + \delta_5$	None
$QA(\delta_2, \delta_3, \delta_4, \delta_5)$	$UA + \delta_2 + \delta_3 + \delta_4 + \delta_5$	δ_1
$QA(\delta_1, \delta_3, \delta_4, \delta_5)$	$UA + \delta_1 + \delta_3 + \delta_4 + \delta_5$	δ_2
$QA(\delta_1, \delta_2, \delta_4, \delta_5)$	$UA + \delta_1 + \delta_2 + \delta_4 + \delta_5$	δ_3
$QA(\delta_1, \delta_2, \delta_3, \delta_5)$	$UA + \delta_1 + \delta_2 + \delta_3 + \delta_5$	δ_4
$QA(\delta_1, \delta_2, \delta_3, \delta_4)$	$UA + \delta_1 + \delta_2 + \delta_3 + \delta_4$	δ_5
$QA(\delta_1, \delta_2, \delta_3)$	$UA + \delta_1 + \delta_2 + \delta_3$	δ_4, δ_5
$QA(\delta_1, \delta_2, \delta_4)$	$UA + \delta_1 + \delta_2 + \delta_4$	δ_3, δ_5
$QA(\delta_1, \delta_2, \delta_5)$	$UA + \delta_1 + \delta_2 + \delta_5$	δ_3, δ_4
$QA(\delta_1, \delta_3, \delta_4)$	$UA + \delta_1 + \delta_3 + \delta_4$	δ_2, δ_5
$QA(\delta_1, \delta_3, \delta_5)$	$UA + \delta_1 + \delta_3 + \delta_5$	δ_2, δ_4
$QA(\delta_2, \delta_3, \delta_4)$	$UA + \delta_2 + \delta_3 + \delta_4$	δ_1, δ_5
$QA(\delta_2, \delta_3, \delta_5)$	$UA + \delta_2 + \delta_3 + \delta_5$	δ_1, δ_4
$QA(\delta_3, \delta_4, \delta_5)$	$UA + \delta_3 + \delta_4 + \delta_5$	δ_1, δ_2
$\vdots$	$\vdots$	$\vdots$
$QA(\delta_5)$	$UA + \delta_5$	$\delta_1, \delta_2, \delta_3, \delta_4$

According to McCulagh and Nelder (1989) and Fahrmeir and Tutz (2001), the connection between the model mean μ_i and the linear predictor η_i is generalized from

$$\mu_i = \eta_i = x_{ij}'\beta \quad (3.1)$$

to

$$g(\mu_i) = \eta_{ij} = x_{ij}'\beta \quad (3.2)$$

where g is a link function and the inverse of g is called the inverse link function. The assumption that $y_i = x_{ji}$ is normal, and is generalized to the assumption that y_i has a specified exponential family distribution given as

$$f(y | \theta, \phi) = \exp\left(\frac{(y_i\theta_i - b(\theta_i))}{\phi} + c(y_i, \phi)\right), \quad (3.3)$$

where θ controls the distribution of y and ϕ is the scale parameter.

Based on this assumption, the log-likelihood for a Poisson model will be

$$l = \sum_{i=1}^n (y_i \log(\mu_i) - \mu_i) - \log(y_i!) \quad (3.4)$$

So $\theta_i = \log(\mu_i)$, $\phi = 1$, and $b(\theta_i) = \mu_i = \exp(\theta_i)$.

It could also be written as a function of β and ϕ because (given x_i), β determines all the θ_i . The main way of maximizing β is by maximizing (3.4)

So the link function g of the Poisson model above is a log function which leads to loglinear model (Tanner and Young, 1985a; and Agresti, 1988) of the forms under consideration, through which one can display patterns of agreement between two raters.

Given the likelihood function in (3.4) and substituting each of the models for quasi-association in Table 3.1 and model (2.1) for uniform association, we adopt Fisher scoring algorithm for the maximization of β s. The basic step of Fisher scoring iteration is

$$\beta^{(t+1)} = \beta^{(t)} - \left\{ E \left(\frac{\partial^2 l}{\partial \beta \partial \beta'} \right) \right\}^{-1} \frac{\partial l}{\partial \beta} \quad (3.5)$$

which can also be written as,

$$\beta^{(t+1)} = \beta^{(t)} + (-E(l''(\beta^{(t)}))^{-1} l'(\beta^{(t)})) \quad (3.6)$$

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

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

where X is the matrix of the covariates, W is the weight which is a diagonal matrix with

$$W_{ii} = (G'(\mu_i)^2 b''(\theta_i))^{-1} \quad (3.8)$$

and

$$Z_i = x_i \beta + (Y_i - \mu_i) G'(\mu_i) \quad (3.9)$$

is the working variate.

Because $\eta = \log(\mu)$, it follows that

$$G'(\mu_i) = \frac{\partial \eta_i}{\partial \mu_i} = \frac{1}{\mu_i},$$

so that

$$W = \text{Diag}(\mu_i)$$

$$Z_i = \eta_i + \left(\frac{y_i - \mu_i}{\mu_i} \right)$$

Both equations (3.8) and (3.9) use $\beta = \beta^{(k)}$ to derive values of $\theta_i^{(k)}$ and $\mu_i^{(k)}$. The weights W_{ii} have the usual interpretation as reciprocal of variances, $b''(\theta_i)$ is proportional to the variance of Y_i .

Fisher scoring may also be written as

$$\beta^{(k+1)} = \beta^{(k)} + (X'WX)^{-1} X'WZ_i^* \quad (3.10)$$

where

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

We carry out this process of maximizing β , using X as covariate and the z_i 's as weights and continue the process until a converged value of β is obtained.

This Fisher scoring iterative procedure is used to formulate algorithms based on the underlying assumptions in each of the models in line with (GLM) procedures. This can be executed with R or Splus statistical programs.

4.0 Empirical examples

We shall use the data from Holmquist et al. (1967) on 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 examined two pairs of these pathologists.

Example 1. Consider the data arising from the study between pathologists D and F that investigated the variability in the classification of carcinoma in situ of the uterine cervix of 118 slides.

Table 4.1: Cross-classification of pathologists D and F on carcinoma in situ of the uterine cervix of 118 slides

Category	Pathologist F					
Pathologist D	1	2	3	4	5	Total
1	35	3	0	0	0	38
2	25	17	5	1	0	48
3	2	9	11	0	1	23
4	0	2	4	0	2	8
5	0	0	0	0	1	1
Total	62	31	20	1	4	118

Under Uniform association model

Coefficients:	Estimate	Std. Error	z value	Pr(> z)
μ	2.69346	0.20419	13.191	< 2e-16 ***
λ_1^A	-0.71022	0.30078	-2.361	0.018215 *
λ_2^A	-2.62408	0.56915	-4.611	4.02e-06 ***
λ_3^A	-0.22309	0.41168	-0.542	0.587884
λ_4^A	-4.72655	1.13762	-4.155	3.26e-05 ***
λ_1^B	-1.07985	0.28366	-3.807	0.000141 ***
λ_2^B	-4.58815	1.02316	-4.484	7.32e-06 ***
λ_3^B	-1.35404	0.49837	-2.717	0.006588 **
λ_4^B	-3.33910	0.74115	-4.505	6.63e-06 ***
β	0.13012	0.05119	2.542	0.011021 *

*, ** and *** indicate significance at 5%, 1% and 0.1% respectively.

Number of Fisher Scoring iterations: 5

$G^2 = 78.366$, $df = 15$

Under Quasi-association model

Coefficients:	Estimate	Std. Error	z value	Pr(> z)
μ	3.01927	0.64788	4.660	3.16e-06 ***
λ_1^A	-0.77031	0.83166	-0.926	0.3543
λ_2^A	-1.85401	1.05384	-1.759	0.0785 .
λ_3^A	0.58083	1.04760	0.554	0.5793
λ_4^A	-2.52485	1.52537	-1.655	0.0979 .
λ_1^B	-0.88986	0.71386	-1.247	0.2126
λ_2^B	-10.53435	23.733	-0.444	0.6571
λ_3^B	-1.64173	1.14304	-1.436	0.1509
λ_4^B	-0.63746	1.28327	-0.497	0.6194
δ_1	3.83576	0.88333	4.342	1.41e-05 ***
δ_2	-0.04614	0.77539	-0.060	0.9526
δ_3	1.59849	0.65523	2.440	0.0147 *
δ_4	-5.66143	63.53724	-0.089	0.9290
δ_5	10.62840	23.74684	0.448	0.6545
β	-0.13992	0.12114	-1.155	0.2481

*, ** and *** indicate significance at 5%, 1% and 0.1% respectively.
 Number of Fisher Scoring iterations: 7
 $G^2 = 25.712$, $df = 10$

Example 2. Consider the data from the same study in Example 1 for pathologists D and F.

Table 4.2: Cross-classification of pathologists D and F on carcinoma in situ of the uterine cervix of 118 slides

Category	Pathologist G					Total
	1	2	3	4	5	
1	26	1	0	0	0	27
2	3	8	1	0	0	12
3	3	11	54	1	0	69
4	0	0	6	1	0	7
5	0	0	0	1	2	3
Total	32	20	61	3	2	118

Under Uniform association model

Coefficients:	Estimate	Std. Error	z value	Pr(> z)
μ	0.87488	0.33523	2.610	0.00906 **
λ_1^A	-0.03277	0.31647	-0.104	0.91753
λ_2^A	-4.83768	0.85657	-5.648	1.63e-08 ***
λ_3^A	-3.82977	0.84722	-4.520	6.17e-06 ***
λ_4^A	-8.34421	1.41070	-5.915	3.32e-09 ***
λ_1^B	0.67731	0.35553	1.905	0.05677 .
λ_2^B	-3.78865	0.75805	-4.998	5.80e-07 ***

λ_3^B	-3.29665	0.84442	-3.904	9.46e-05 ***
λ_4^B	-6.69899	1.30784	-5.122	3.02e-07 ***
β	0.56412	0.10332	5.460	4.76e-08***

*, ** and *** indicate significance at 5%, 1% and 0.1% respectively.

Number of Fisher Scoring iterations: 6

$$G^2 = 98.128, df = 15$$

Under Quasi-association model

Coefficients:	Estimate	Std. Error	z value	Pr(> z)
μ	0.55645	0.78345	0.710	0.4775
λ_1^A	0.53936	0.83633	0.645	0.5190
λ_2^A	-9.80096	24.02838	-0.408	0.6834
λ_3^A	-0.57752	1.20555	-0.479	0.6319
λ_4^A	-1.51029	1.48583	-1.016	0.3094
λ_1^B	1.72273	0.90329	1.907	0.0565 .
λ_2^B	-1.60360	1.37434	-1.167	0.2433
λ_3^B	-1.40237	1.51526	-0.926	0.3547
λ_4^B	0.35902	1.47954	0.243	0.8083
δ_1	5.18749	1.29860	3.995	6.48e-05 ***
δ_2	1.55461	0.92606	1.679	0.0932 .
δ_3	1.29693	0.68852	1.884	0.05006 *
δ_4	1.38537	1.98920	0.696	0.4861
δ_5	11.82585	24.04732	0.492	0.6229
β	-0.03162	0.14995	-0.211	0.8330

*, ** and *** indicate significance at 5%, 1% and 0.1% respectively.

Number of Fisher Scoring iterations: 7

$$G^2 = 22.257, df = 10$$

5. Results and Discussion

5.1 Results

Table 5.1: Summary results of the examples in section 4 under Uniform association model and quasi-association model with their difference.

	Uniform association		Quasi association		Difference	
Example	G_{UA}^2	df_{UA}	G_{QA}^2	df_{QA}	G_d^2	df_d
1.	78.318	15	25.712	10	52.606	5
2.	98.128	15	22.257	10	75.871	5

Table 5.2: Quasi-Association model selection with different combinations of δ_i s in Table 3.2

Model notation	G_{DF}^2	G_{BG}^2	df
QA(δ_1)	40.586*	53.557*	14
QA(δ_2)	78.357	91.73	14

QA(δ_3)	63.574	71.653	14
QA(δ_4)	78.108	98.121	14
QA(δ_5)	71.406	82.888	14
QA(δ_1, δ_2)	38.255	42.374	13
QA(δ_1, δ_3)	32.113*	37.560*	13
QA(δ_1, δ_4)	40.554	53.431	13
QA(δ_1, δ_5)	33.935	38.648	13
QA(δ_2, δ_3)	62.989	71.185	13
QA(δ_2, δ_4)	78.103	91.706	13
QA(δ_2, δ_5)	71.404	76.413	13
QA(δ_3, δ_4)	63.315	71.512	13
QA(δ_3, δ_5)	56.766	58.032	13
QA(δ_4, δ_5)	71.146	82.880	13
QA($\delta_1, \delta_2, \delta_3$)	32.101	35.133	12
QA($\delta_1, \delta_2, \delta_4$)	38.247	40.919	12
QA($\delta_1, \delta_2, \delta_5$)	31.809	27.425	12
QA($\delta_1, \delta_3, \delta_4$)	32.062	37.540	12
QA($\delta_1, \delta_3, \delta_5$)	25.764*	25.254*	12
QA($\delta_1, \delta_4, \delta_5$)	33.905	38.433	12
QA($\delta_2, \delta_3, \delta_4$)	62.692	71.007	12
QA($\delta_2, \delta_3, \delta_5$)	56.052	57.564	12
QA($\delta_2, \delta_4, \delta_5$)	71.146	76.386	12
QA($\delta_3, \delta_4, \delta_5$)	56.505	57.895	12
QA($\delta_1, \delta_2, \delta_3, \delta_4$)	32.053	34.746	11
QA($\delta_1, \delta_2, \delta_3, \delta_5$)	25.762	22.767	11
QA($\delta_1, \delta_3, \delta_4, \delta_5$)	25.715*	25.228*	11
QA($\delta_2, \delta_3, \delta_4, \delta_5$)	55.750	57.423	11
QA($\delta_1, \delta_2, \delta_3, \delta_4, \delta_5$)	25.712	22.257	10

*The quasi association model in the specific class size of δ_i s with the best fits.

We observe, from Table 5.1, that the values of the Likelihood ratio G^2 for Uniform Association were greater than that of Quasi-Association model for each of the examples. The column of difference in the table shows the level of improvement in the fit, obtained by including the pattern of agreement parameters, for Likelihood ratio G_a^2 with degrees of freedom df_a . We also observed that the pattern of agreement parameter for category 1 which was labeled as "Negative", δ_1 was highly significant in the two examples, which implied that category 1 was observed to be most common, follow by 3 "carcinoma in situ", δ_3 which is also significant in the two examples. We can also confirmed this in Table 5.2, where we have QA(δ_1) in the first group, follow by QA(δ_1, δ_3) in the next group with the best fits and so on. Whenever there were no "pattern of agreement" in the model, the strength of association parameter (β) was significant as we had in the two examples, however, when the "pattern of agreement" parameters were included in the model, strength of association parameter was no more significant.

5.2 Discussion

Inclusion of the pattern of agreement parameters δ_i for the ordered categories in addition to the strength of association parameter β by Quasi-Association improved the fit of the model compared to Uniform Association. This is more so when the pattern of agreement parameter δ_i is significant. In order to identify the most common categories out of the I ordered categories considered by all the raters, Quasi-Association is a more appropriate model to be used.

References

1. Adejumo, A. O. (2005). Modelling Generalized Linear (Log-linear) Models for Raters agreement Measures; Peter Lang, Frankfurt am Main, Germany.
2. Adejumo, A. O, Heumann, C and Toutenburg, H (2006), "Comparison of symmetry and Conditional Symmetry Models as tools for raters agreement Measures, *Journal of Nigerian Statistical Association*, 18: 77 – 90.
3. Agresti, A. (1988). A model for agreement between ratings on an ordinal scales, *Biometrics*, 44, 539- 548.
4. Agresti, A. (1990). *Categorical Data Analysis*, Wiley, New York.
5. Agresti, A. (1996). *An introduction to Categorical Data Analysis*, Wiley, New York.
6. Altham, P. M. E. (1975). Quasi independence triangular contingency tables. *Biometrics*, 31, 233-238.
7. 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.
8. Barlow, W. (1996). Measurement of interrater agreement with adjustment for covariates. *Biometrics*, 52, 695-702.
9. Barlow, W., Lai, M. Y., and Azen, S. P. (1991). A comparison of methods for calculating a stratified kappa. *Statist. Med.*, 10, 1465-1472.
10. Barnhart, H. X, and Williamson, J. M. (2002). Weighted least-squares approach for comparing correlated kappa. *Biometrics*, 58, 1012-1019.
11. Becker, M. P. (1990). Quasisymmetric models for the analysis of square contingency tables. *J. Roy. Statist. Soc. Series B*, 52, 369-378.
12. Birch, M. W. (1965). The detection of partial association II: The general case. *J. oy. Statist. Soc.*, B, 27, 111-124.
13. Block, D. A., and Kraemar, H. C. (1989). 2 x 2 kappa coefficients: Measures of agreement or association. *Biometrics*, 45, 269-287.
14. Caussinus, H. (1965). Contribution l'analyse statistique des tableaux de correlation. *Ann. Fac. Sci. Univ. Toulouse*, 29, 77-182.
15. Cohen, J. (1960). A coefficient of agreement for nominal scales. *Edu. and Psych. Meas.*, 20, 37-46.
16. Cohen, J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psych. Bull.*, 70, 213-220.
17. Darroch, J. N., and McCloud, P. I. (1986). Category distinguishability and observer agreement. *Australian J. Statist.*, 28, 371-388.
18. Davies, M. and Fleiss, J. L. (1982). Measuring agreement for multinomial data. *Biometrics*, 38, 1047-1051.
19. Donner, A., Eliasziw, M. and Klar, N. (1996). Testing homogeneity of kappa statistics. *Biometrics*, 52, 176-183.
20. Fahrmeir, L., and Tutz, G. (2001). *Multivariate statistical modelling based on generalized linear models*. Springer.

21. **Fienberg, S. E. (1972).** The analysis of incomplete multi-way contingency tables. *Biometrics*, 28, 177-202.
22. **Fleiss, J. L. (1971).** Measuring nominal scale agreement among many raters. *Psych. Bull.*, 76, 378-382.
23. **Goodman L. A. (1968).** The analysis of cross-classified data: independence, quasi-independence, and interactions in contingency tables with or without missing entries. *J. Amer. Statist. Assoc.*, 63, 1091-1131.
24. **Goodman L. A. (1979a).** Simple models for the analysis of association in cross classifications having ordered categories. *J. Amer. Statist. Assoc.*, 74, 537-552.
25. **Goodman L. A. (1979b).** Multiplicative models for square contingency tables with ordered categories. *Biometrika*, 66, 413-418.
26. **Haberman, S. J. (1974).** Loglinear models for frequency tables with ordered classifications. *Biometrics*, 36, 589-600.
27. **Holmquist, N. S., McMahon, C. A. and Williams, O. D. (1967).** Variability in classification of carcinoma in situ of the uterine cervix. *Archives of Pathology*, 84, 334-345.
28. **McCullagh, P., and Nelder, J. A. (1989).** *Generalized linear models*, 2nd Edit. Chapman and Hall.
29. **Oden, N. L. (1991).** Estimating kappa from binocular data. *Statist. Med.*, 10, 1303-1311.
30. **Shoukri, M. M. (2004).** *Measures of interobserver agreement*. Chapman and Hall.
31. **Shoukri, M. M., Martin, S. W., and Mian, I. U. H. (1995).** Maximum likelihood estimation of the kappa coefficient from models of matched binary responses. *Statist. Med.* 14, 83-99.
32. **Tanner, M. A., and Young, M. A. (1985a).** Modelling agreement among raters. *J. Amer. Statist. Assoc.*, 80, 175-180.
33. **Tanner, M. A., and Young, M. A. (1985b).** Modelling ordinal scale agreement. *Psych. Bull.*, 98, 408-415.