

MODELLING QUASI-INDEPENDENCE AS A MEASURE OF AGREEMENT FOR DISAGREEMENT BETWEEN RATERS

A. O. Adejumo

Department of Statistics
University of Ilorin
Ilorin, Nigeria

ABSTRACT

One of the major statistical issue in raters agreement studies is concerned with the extent to which observers classify individual subjects into the same category on the measurement scale. There are various testing procedures, modification of statistics and modelling for this purpose. Subjects are being classified into categories by raters, virtually in all life or social science researches, such that in the arising cross-classified table, the raw agreements for the categorical scales are the subjects classified into the same category by the two raters, that is, subject along the diagonal, while disagreement are those classified differently by the two raters. In this paper we examined what happen on agreement measurement from the off-diagonal elements of the table. In order to achieve this goal, a parametric model that excludes the diagonal elements and add some agreement component beyond expectation under independence which is quasi-independence (QI) model will be appropriate for this on nominal scale. We observed that a better fit and very great improvement would be achieved if quasi-independence model should be used in comparison with the baseline independence model for agreement on raters disagreement.

KEYWORDS AND PHRASES: Agreement, disagreement, Independence, nominal scale, Quasi-independence, Independence, agreements, nominal scales, raters.

INTRODUCTION

Items that are classified equally by the two raters are being referred to as raw agreement, but others being classified differently by the same raters are disagreement. A lot of work has been done on agreement, however, in this paper we examined what happens to the level of agreement that can be achieved from these items that fall under disagreement. To achieve this, a parametric model which will exclude the diagonal elements and add some agreement component beyond expectation under independent will be needed. Quasi independence model (QI) is a special loglinear model which was initially proposed for any square table like the one that could existed between two raters in an experiment on certain number of subjects. This model added a parameter which is for agreement beyond chance to the baseline statistical independence model (SIND) to arrive at its own model.

In any research that involves ratings, measurements or diagnosis of subjects by various raters; observers; Neurologists or Pathologists, the researchers are already aware that the most important measurement of error or bias are the raters involved in such studies.

Modelling was introduced into raters agreement as a remedy to various problems and shortcomings 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 (Goodman and Kruskal, 1954). Due to this problem of using a single value to represent agreement, some authors like Tanner and Young (1985), Agresti (1988), Agresti and Lang (1993), Banerjee *et al.* (1999), Shoukri (2004), Adejumo (2005; 2010) presented in their papers different methods and procedures for modelling raters agreement.

The main objective of this paper is to model quasi independence model on the disagreement of the raters by assuming that the raters disagree. So to model the independence between the ratings of the two observers disagreement on nominal scale, we need a parametric model that includes some components that has to do with agreement beyond chance in its model or that excludes the diagonal cells in the cross-classified table which is quasi-independence. However, for an ordinal scale, given that two raters disagree, it is unrealistic to expect independence ratings, that is, we cannot model quasi-independence for data on ordinal scales. To this effect, we shall model quasi-independence to data on nominal scales only in comparison with the baseline independence model for raters disagreement's agreement.

We shall use generalized linear models (GLMs) via loglinear model to structure the pattern of agreement for these models under consideration. In sections 1 and 2, we have the models and their properties, and modelling and estimation of parameters respectively. Section 3 contains some working empirical examples and results. Finally, conclusions are drawn in the last section.

1. MODELS AND THEIR PROPERTIES

Quasi independence was originated by Goodman, 1968. A more general definition of quasi independence, for some fixed set of cells, is given as

$$\pi_{ij} = \alpha_i \beta_j \quad \text{for } i \neq j \quad (1)$$

where $\{\alpha_i\}$ and $\{\beta_j\}$ are constants (Agresti, 1990). This implies that a square contingency table satisfies quasi independence when the variables are independent, given that the raters responses differ from one another, that is the row response is differs from the column response. In order to reduce errors or biases, an initial reliability study between the raters or observers is necessary to assess the variability level in their measurements criteria. Based on this reliability test that is already in place, it is highly expected that their ratings will be more uniform and consistence. In the arising cross-classified table, any subject along the diagonal n_{ii} are refers to as raw agreement, while those that are off-diagonal $n_{ij} \forall i \neq j$ constitute disagreement.

MODELLING QUASI-INDEPENDENCE AS A MEASURE OF AGREEMENT
FOR DISAGREEMENT BETWEEN RATERS

One of the usual ways of evaluating agreement between raters in a two-way or square table is by comparing the cell observed responses $\{n_{ij}\}$ with their corresponding expected cell estimates $\left\{\frac{n_{i+}n_{+j}}{n}\right\}$ obtained by the loglinear model of independence given as

$$\log(m_{ij}) = \mu + \lambda_i^{R_1} + \lambda_j^{R_2} \quad (2)$$

Model (2) will provide a baseline by showing us the degree of agreement expected if there is no association existed between the ratings. It is also expected that model (2) will fit very poorly even if there is a mild agreement.

To this effect, a more complex loglinear model that add components that relate to agreement beyond that expected under independence is required. One useful generalization of independence is the quasi independence model given as

$$\log(m_{ij}) = \mu + \lambda_i^{R_1} + \lambda_j^{R_2} + \delta_i I(i = j) \quad (3)$$

for $I(i = j)$ is an indicator function and $\delta_i I(i = j)$ is defined as

$$\delta_i I(i = j) = \begin{cases} \delta_1 & \text{if } i = j = 1 \\ \delta_2 & \text{if } i = j = 2 \\ \vdots & \vdots \\ \delta_l & \text{if } i = j = l \\ 0 & \text{if } i \neq j \end{cases} \quad (4)$$

and $\sum_i \lambda_i = 0$, and $\sum_j \lambda_j = 0$ for $i = j = 1, 2, \dots, l$. This model add to the independence model (2) the parameter δ_1 for cell in row 1 and column 1, δ_2 for cell in row 2 and column 2, and so on, which are indicator variables that represent the main diagonal cells. When $\delta_i > 0$, more agreements regarding outcome i occur than would be expected under independence.

The likelihood equations for QI are

$$\begin{aligned} \hat{m}_{i+} &= n_{i+} \\ \hat{m}_{+j} &= n_{+j} \\ \text{and } \hat{m}_{ii} &= n_{ii} \text{ for } i = 1, 2, \dots, l \\ \hat{m}_{ij} \text{ for } i \neq j &\text{ has to be obtained by iterative method.} \end{aligned}$$

The goodness of fit statistics for the model is:

Likelihood ratio statistic

$$G^2 = \sum_{i \neq j} n_{ij} \log \left(\frac{n_{ij}}{\hat{m}_{ij}} \right) \quad (5)$$

The degrees of freedom (df) for the residuals of this statistic is given as $(I^2 - 3I + 1)$.

Quasi-independence is a special case to baseline statistical independence model given in equation (2) when $\delta_i = 0$ for all $i = 1, 2, \dots, l$ (See Caussinus, 1965; Altham, 1975; Agresti, 1996 for more details).

2. MODELLING AND ESTIMATION OF PARAMETERS

Based on generalized linear models (GLMs), the influence of the explanatory variables is linear, such that

$$\eta_i = x'_{ij}\beta, \quad (6)$$

where η_i is called linear predictor. x_{ij} contains $n \times p$ covariates, where n is the number of cells and p is the number of parameters involve in the model. The connection between the model mean μ_i and the linear predictor η_i is generalized from $\mu_i = \eta_i$ to

$$g(\mu_i) = \eta_i = x'_{ij}\beta, \quad (7)$$

where g is a link function and the inverse of g is called the inverse link function (See McCullagh and Nelder, 1989; Fahrmeir and Tutz, 2001 for details).

The assumption that y_i is normal is generalized to the assumption that y_i has a specified exponential family distribution given as

$$f(y|\theta, \phi) = \exp\left(\frac{(y_i\theta_i - b(\theta_i))}{\phi} + c(y_i, \phi)\right), \quad (8)$$

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

The log-likelihood for the model, maximized over the regression coefficients, is

$$\log L = \sum_{i=1}^n \log(p(\mu_i, \phi, y_i)) \quad (9)$$

where $p(\cdot)$ is the probability or probability-density function corresponding to the family, which in this case will be Poisson. The corresponding log-likelihood for a Poisson model will then be

$$l = \sum_{i=1}^n (y_i \log(\mu_i) - \mu_i) - \log(y_i!) \quad (10)$$

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

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 (10).

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

So the link function g of the Poisson model above is a log function which leads us to loglinear model of the form

$$\log(m_{ij...l}) = g(\mu) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{j2} + \dots + \beta_k x_{lk}, \quad i, j, \dots, l = 1, \dots, m \quad (11)$$

which can also be written in a better form as

$$\log(m_{ij...l}) = g(\mu) = \beta + \beta_1^{R_1} + \beta_j^{R_2} + \dots + \beta_l^{R_k}, \quad i, j, \dots, l = 1, \dots, m \quad (12)$$

MODELLING QUASI-INDEPENDENCE AS A MEASURE OF AGREEMENT
FOR DISAGREEMENT BETWEEN RATERS

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 ($k \geq 2$) into m nominal categories, the loglinear model for testing the baseline statistical independence among k raters or observers is given as

$$\log(m_{ij...l}) = \mu + \lambda_i^{R_1} + \lambda_j^{R_2} + \dots + \lambda_l^{R_k}, \quad i, j, \dots, l = 1, \dots, m \quad (13)$$

where $m_{ij...l}$ is the expected cell counts in the $ij \dots l$ th cell of the joint K -dimensional cross classification of the ratings, μ is the overall effect, λ^{R_l} is the effect due to categorization by the k th rater in the c th category ($k = 1, \dots, K$; $c = 1, \dots, m$), and $\sum_{i=1}^m \lambda_i^{R_1} = \dots = \sum_{l=1}^m \lambda_l^{R_k} = 0$

(See Tanner and Young (1985) which refers to Equation (13) as chance agreement and Agresti (1988)).

Quasi-independence model for $k > 2$ raters will then be

$$\log(m_{ij...l}) = \mu + \lambda_i^{R_1} + \lambda_j^{R_2} + \dots + \lambda_l^{R_k} + \delta_{il}(i = j = \dots = l), \quad i, j, \dots, l = 1, \dots, m \quad (14)$$

where

$$\delta_{il}(i = j = \dots = l) = \begin{cases} \delta_1 & \text{if } i = j = \dots = 1 \\ \delta_2 & \text{if } i = j = \dots = 2 \\ \vdots & \vdots \\ \delta_l & \text{if } i = j = \dots = l \\ 0 & \text{if otherwise} \end{cases} \quad (15)$$

Tanner and Young (1985) refers to Equation (14) as independence model with agreement beyond chance with a nonhomogeneous agreement effect. $\delta_{il}(i = j = \dots = l)$ are indicator variables that represent the main diagonal cells.

For the purpose of this paper, we shall restrict ourself to $k = 2$ raters, so the baseline statistical independence and quasi-independence models for this size of k will then be

$$\log(m_{ij}) = \mu + \lambda_i^{R_1} + \lambda_j^{R_2}, \quad i, j = 1, \dots, m \quad (16)$$

and

$$\log(m_{ij}) = \mu + \lambda_i^{R_1} + \lambda_j^{R_2} + \delta_{il}(i = j), \quad i, j = 1, \dots, m \quad (17)$$

where $\delta_{il}(i = j)$ is defined as

$$\delta_{il}(i = j) = \begin{cases} \delta_1 & \text{if } i = j = 1 \\ \delta_2 & \text{if } i = j = 2 \\ \vdots & \vdots \\ \delta_l & \text{if } i = j = l \\ 0 & \text{if otherwise} \end{cases} \quad (18)$$

Given the likelihood function in (10) and substituting each of (16) and (17) for statistical independence and quasi-independence respectively, we then adopt Fisher scoring algorithm for the maximization of β s. The basic step of Fisher scoring iteration is

$$\beta^{(t+1)} = \beta^{(t)} - \left\{ E \left(\frac{\partial^2 l}{\partial \beta \partial \beta'} \right) \right\}^{-1} \frac{\partial l}{\partial \beta} \quad (19)$$

which can also be written as

$$\beta^{(t+1)} = \beta^{(t)} + \left(-E \left(l''(\beta^{(t)}) \right) \right)^{-1} l'(\beta^{(t)}) \quad (20)$$

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 step, except that Hessian of l is replaced by it's expectation. Fisher scoring simplifies to

$$\beta^{(k+1)} = (X'WX)^{-1}X'WZ \quad (21)$$

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

$$W_{ii} = (G'(\mu_i)^2 b''(\theta_i))^{-1} \quad (22)$$

and

$$Z_i = x_i\beta + (Y_i - \mu_i)G'(\mu_i) \quad (23)$$

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

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

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

Fisher scoring may be written as well as

$$\beta^{(t+1)} = \beta^t + (X'WX)^{-1}X'WZ_i^*$$

where

$$Z_i^* = \eta_i + \left(\frac{(Y_i - \mu_i)}{\mu_i} \right)$$

Both equations (22) and (23) use $\beta = \beta^{(k)}$ and derived 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 and the $G'(\mu)$ factor in Z_i is squared in W_{ii} . Fisher scoring may also be written as

MODELLING QUASI-INDEPENDENCE AS A MEASURE OF AGREEMENT
FOR DISAGREEMENT BETWEEN RATERS

$$\beta^{(k+1)} = \beta^{(k)} + (X'WX)^{-1}X'WZ_i^* \quad (24)$$

where

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

The iteration (24) is known as "iteration reweighted least squares", or IRLS.

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

This Fisher scoring Iterative procedure is used to formulate algorithms under GLMs, which can be executed with any of the R-program, or the Splus.

3. EMPIRICAL EXAMPLES

In this section we present some empirical examples which categorical scales are assumed or are nominal.

Example 1.

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: NE=negative; ASH=Atypical squamous hyperplasia; CIS=carcinoma in situ; SCI=squamous carcinoma with early stromal invasion; IC=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 D as follows.

Category	Pathologist D					
Pathologist A	NE	ASH	CIS	SCI	IC	Total
NE	25	1	0	0	0	26
ASH	11	15	0	0	0	26
CIS	1	22	13	2	0	38
SCI	1	8	9	4	0	22
IC	0	2	1	2	1	6
Total	38	48	23	8	1	118

Table 1. Cross-classification of pathologists A and D on carcinoma
in situ of the uterine cervix of 118 slides

Coefficients	Value	Std. Error	t-value	p-value
Intercept	2.125	0.2370	8.968	< 0.0001
$\lambda_2^{R_1}$	0.2336	0.2171	1.076	0.281849
$\lambda_3^{R_1}$	-0.5021	0.2640	-1.902	0.0572
$\lambda_4^{R_1}$	-1.558	0.3386	-4.010	< 0.0001
$\lambda_5^{R_1}$	-3.637	0.9905	-3.671	< 0.0001
$\lambda_2^{R_2}$	-7638e-08	0.2770	-2.76e-05	0.999978
$\lambda_3^{R_2}$	0.3795	0.2542	1.493	0.1355
$\lambda_4^{R_2}$	-0.1871	0.2894	-0.577	0.563735
$\lambda_5^{R_2}$	-1.466	0.4526	-3.240	< 0.0001
goodness of fit statistics				
	G^2	117.34		
	df	16		

Table 2. Parameter estimates under baseline statistical independence model on pathologists A and D on carcinoma in situ of the uterine cervix of 118 slides

Coefficients	Value	Std. Error	t-value	p-value
Intercept	-186493	1.05047	-1.585	0.11288
$\lambda_2^{R_1}$	1.30302	0.35265	3.695	< 0.0001
$\lambda_3^{R_1}$	0.19077	0.43704	0.437	0.66247
$\lambda_4^{R_1}$	-0.82955	0.57613	-1.613	0.10665
$\lambda_5^{R_1}$	-9.43122	19.24252	-0.490	0.62405
$\lambda_2^{R_2}$	3.10543	1.08625	2.912	0.0036
$\lambda_3^{R_2}$	3.25946	1.02426	3.182	0.0015
$\lambda_4^{R_2}$	2.78195	1.02926	2.703	0.0089
$\lambda_5^{R_2}$	1.43615	1.08682	1.310	0.19033
δ_1	4.88381	1.06934	4.567	< 0.0001
δ_2	-0.03547	0.50764	-0.070	0.094330
δ_3	0.77966	0.52493	1.485	0.13747
δ_4	1.19883	0.76709	1.563	0.11809
δ_5	9.66000	19.27223	0.501	0.61620
goodness of fit statistics				
	G^2	41.51		
	df	11		

Table 3. Parameter estimates under Quasi-independence model on pathologists A and D on carcinoma in situ of the uterine cervix of 118 slides

MODELLING QUASI-INDEPENDENCE AS A MEASURE OF AGREEMENT
FOR DISAGREEMENT BETWEEN RATERS

Example 2.

Consider a set of data on the accuracy of certain items collected on subjects enrolled in a cancer registry (Brownson *et al.*, 1989). Table 4 is a joint occupation classification of cancer registry subjects by registry protocol and telephone interview. The following categorical scales are used: MP=Managerial/professional, TSA=Technical/sales/ administrative support, SO=Service occupations, FFF= Farming/forestry/fishing, PCR=Precision production/craft/repair, OFL=Operators/fabricators/labourers, AF=Armed forces.

Category Registry class	Interview Class							Total
	MP	TSA	SO	FFF	PCR	OFL	AF	
MP	80	5	1	0	3	1	0	70
TSA	2	67	0	1	1	2	0	73
SO	7	2	67	1	3	6	0	86
FFF	0	0	0	27	1	1	0	29
PCR	2	1	0	0	64	7	0	74
OFL	0	2	2	1	5	82	0	92
AF	1	0	0	0	0	2	13	18
Total	72	77	70	30	77	101	13	440

Table 4. Joint occupation classification of cancer registry subjects
by registry protocol and telephone interview

Coefficients	Value	Std. Error	t-value	p-value
Intercept	2.43839	0.16094	15.151	< 0.0001
$\lambda_1^{R_1}$	0.06714	0.16394	0.410	0.6821
$\lambda_2^{R_1}$	-0.02817	0.16785	-0.168	0.8687
$\lambda_3^{R_1}$	-0.87547	0.21731	-4.029	< 0.0001
$\lambda_4^{R_1}$	0.06714	0.16394	0.410	0.6821
$\lambda_5^{R_1}$	0.33845	0.15424	2.194	0.0282
$\lambda_6^{R_1}$	-1.71171	0.30081	-5.690	< 0.0001
$\lambda_7^{R_1}$	0.04196	0.16727	0.251	0.8019
$\lambda_1^{R_2}$	0.20585	0.16097	1.279	0.2010
$\lambda_2^{R_2}$	-0.68120	0.22082	-3.991	< 0.0001
$\lambda_3^{R_2}$	0.05557	0.16672	0.333	0.7389
$\lambda_4^{R_2}$	0.27329	0.15859	1.723	0.0848
$\lambda_5^{R_2}$	-1.47591	0.27708	-5.327	< 0.0001
goodness of fit statistics				
G^2	1123.10			
df	36			

Table 5. Parameter estimates under baseline statistical independence model on occupation
classification of cancer registry subjects by registry protocol and telephone interview

Coefficients	Value	Std. Error	t-value	p-value
Intercept	0.9157	0.4589	1.995	0.0460
$\lambda_{R_1}^1$	-0.2687	0.4355	-0.617	0.5372
$\lambda_{R_2}^1$	-1.2499	0.6550	-1.908	0.0564
$\lambda_{R_3}^1$	-1.5482	0.8489	-2.383	0.0172
$\lambda_{R_4}^1$	0.0848	0.4119	0.206	0.8370
$\lambda_{R_5}^1$	0.5007	0.3848	1.302	0.1929
$\lambda_{R_6}^1$	-9.9330	23.3523	-0.425	0.6706
$\lambda_{R_7}^1$	-0.5680	0.5247	-1.083	0.2790
$\lambda_{R_2}^2$	0.4778	0.3985	1.199	0.2305
$\lambda_{R_3}^2$	-1.7889	0.7780	-2.299	0.0215
$\lambda_{R_4}^2$	0.0224	0.4602	0.049	0.9613
$\lambda_{R_5}^2$	0.1772	0.46741	0.379	0.7046
$\lambda_{R_6}^2$	-1.4270	0.6816	-2.157	0.0310
δ_1	3.1788	0.4767	6.668	< 0.0001
δ_2	4.1257	0.5522	7.471	< 0.0001
δ_3	4.0810	0.6541	6.209	< 0.0001
δ_4	5.7153	0.9428	6.062	< 0.0001
δ_5	3.1361	0.4697	6.677	< 0.0001
δ_6	2.8131	0.448	6.327	< 0.0001
δ_7	13.0083	23.3596	0.557	0.5776
goodness of fit statistics				
	G^2	32.19		
	df	29		

Table 6. Parameter estimates under Quasi-independence model on occupation classification of cancer registry subjects by registry protocol and telephone interview

4. RESULTS

We present the summary of the working examples results in the previous section showing their sample size, Likelihood ratio (G^2) and degrees of freedom under each mother. Also we present another table showing the differences in Likelihood ratio (G^2) as well as the degrees of freedom between the two models, so as to see the extent to which the quasi-independence has improved the results.

From the results of the two examples, we can observe that the matrices of the fitted values along the diagonal for quasi-independence are the same as we have in their respective observed data, which indicated that, quasi-independence model ignore or exclude the diagonal elements, which are refer to as raw agreement, by adding some components that are directly related to the agreement beyond that which is expected under independence to

MODELLING QUASI-INDEPENDENCE AS A MEASURE OF AGREEMENT
FOR DISAGREEMENT BETWEEN RATERS

the model. QI focus on the off-diagonal elements to test for the independence between the ratings of the two raters disagreement. In Table 7, we observed that Likelihood ratio (G^2) for the baseline statistical independence is far greater than that obtained under quasi-independence. Also from the table of differences in Table 8, we can notice the amount of improvement that quasi-independence has put by including components that are related to agreement beyond expectation under independence parameters in its model, which number was indicated in the table by the degrees of freedom column df_d as well as the estimates of the parameters δ_i s. Also δ_i s with * are significant.

Table	n	Statistical independence		Quasi Independence	
		G^2_I	df_I	G^2_Q	df_Q
1	118	117.34	16	41.51	11
2	440	1123.10	36	32.19	29

Table 7. Summary results of the examples in section 4 under baseline statistical independence model and quasi-independence model.

table	Difference (I-Q)		Estimate of parameter delta						
	G^2_d	df_d	δ_1	δ_2	δ_3	δ_4	δ_5	δ_6	δ_7
1	75.83	5	4.88*	-0.035	0.78	1.20	9.66		
2	1090.91	7	3.18*	4.13*	4.06*	5.72*	3.14*	2.81*	13.00

Table 8. Table of difference for Likelihood ratio (G^2) between baseline statistical independence and quasi-independence model with degrees of freedom and the components of agreement beyond independence parameter δ_i

CONCLUSIONS

Whenever there is disagreement between raters on nominal scales, quasi-independence model is the best for modelling the structure of agreement between their ratings. The matrices for the expected values under quasi-independence model for all the data in section (4) show that all the subjects classified into the same category by the two raters contribute nothing to the realization of the Likelihood ratio(G^2) and this in turns, improve its value over that of baseline statistical independence model by adding components that relate to agreement beyond that expected under independence to the model. Hence, quasiindependence model is good for modelling the structures of agreement on disagreement of two raters' ratings over a given nominal scales.

REFERENCES

1. Adejumo, A. O. *Modelling Generalized (loglinear) Models on raters Agreement*; Peter Lang, Frankfurt am Main, Germany, 2005.
2. Adejumo, A. O. A data based approach of imputing missing ratings in raters agreement measurement for two raters; *ICASTOR Journal of Mathematical Sciences*; Vol. 4 (1), 2010, pp. 49 - 67.
3. Agresti, A. A model for agreement between ratings on an ordinal scales; *Biometrics*; 44, 1988, pp. 539-548.
4. Agresti, A. *Categorical Data Analysis*; Wiley, New York. 1990.
5. Agresti, A. *An introduction to Categorical Data Analysis*; Wiley, New York. 1996.
6. Agresti, A., Lang, J. B. Quasi-symmetry latent class models, with application to rater agreement; *Biometrics*; 49, 1993, pp. 131-139.
7. Altham, P. M. E. Quasi-independence triangle contingency tables; *Biometrics*; 31, 1975, pp. 233-238.
8. Banerjee, M., Capozzoli, M., Mcsweney, L., Sinha, D. Beyond Kappa: A review of interrater agreement measure; *The Cana. J. of Statist.*; 27(1), 1999, pp. 3-23.
9. Brownson, R. C., David, J. R., Chang, J. C., DiLorenzo, T. M., Keefe, T. J., Bagby, J. R., Jr. A study of the accuracy of cancer risk factor information reported to a central registry compare with that obtained by interview; *Americ. J. Epidemiology*; 45, 1989, 269-287.
10. Caussinus, H. Contribution a l'analyse statistique des tableaux de correlation; *Ann. Fac. Sci. Univ. Toulouse*; 29, 1965, pp. 77-182.
11. Fahrmeir, L., Tutz, G. *Multivariate statistical modelling based on generalized linear models*; Springer, 2001.
12. Goodman L. A. The analysis of cross-classified data: independence, quasiindependence, and interactions in contingency tables with or without missing entries; *J. Amer. Statist. Assoc.*; 63, 1968, pp. 1091-1131.
13. Goodman L. A., Kruskal, W. H. Measuring of association for cross classifications; *J. Amer. Statist. Assoc.*; 49, 1954, pp. 732-768.
14. Holmquist, N. S., McMahon, C. A., Williams, O. D. Variability in classification of carcinoma in situ of the uterine cervix; *Archives of Pathology*; 84, 1967, pp. 334-345.
15. McCullagh, P., Nelder, J. A. *Generalized linear models*, 2nd Edit.; Chapman and Hall, 1989.
16. Shoukri, M. M. *Measures of interobserver agreement*; Chapman and Hall, 2004.
17. Tanner, M. A., Young, M. A. Modelling agreement among raters; *J. Amer. Statist. Assoc.*; 80, 1985, pp. 175-180.