ICASTOR Journal of Mathematical Sciences
Vol. 4, No. 1 (2010) 49 – 67

A DATA BASED APPROACH OF IMPUTING MISSING RATINGS IN RATERS AGREEMENT MEASUREMENT FOR TWO RATERS

A. O. Adejumo

Department of Statistics
University of Ilorin
Ilorin, Nigeria

ABSTRACT

Subjects are being classified into categories by raters, interviewers or observers in almost all life or social science researches. However, some of these subjects' ratings or responses may be missed or misplaced by any of the raters involved. We refer to this situation as missing ratings. In a situation where only two raters are involved in the experiment, there are possibilities of missing ratings for either or both of the raters. Likelihood approaches were used to a great advantage in estimating the missing values. Moreover, during the course of imputing these missing ratings into their appropriate level or category, it is expected that certain percentage of the missing ratings at one categorical level of a rater will be the same as the others whose ratings are known. We therefore proposed a data-based approach called Percentage Missing Allocation Proportional to Size (PMAPS). This method can be seen as a multiple way of imputing the missing ratings in a cross-classified table of ratings for two raters.

KEYWORDS: EM-algorithm, Kappa-like statistics, PMAPS.

INTRODUCTION

For agreement measurements, square contingency tables can be used to display joint ratings of two raters. In nearly all the researches that involve ratings, measurements or diagnosis of subjects by various raters, researchers are already aware that the most important measurement error or bias is the rater involved in such studies. Incomplete cross-classifications occur by chance, by design or via anomalies inherent in the data. The simplest way to analyze incomplete contingency tables of this type is to delete the observations for which any of the variables is missing. However, this practice may adversely affect both the accuracy and the precision of the results because the missing data may contain pieces of information that cannot be extracted from the remaining completely cross-classified data. To this effect, all the available pieces of information have to be incorporated into the analysis. This process of estimating the missing values or redistributing the missing values into the complete table has statistical advantages because it strengthens inferences by reducing loss of information, bias and variance of estimates (Chen and Fienberg, 1974).

Correspondence: Department of Statistics, University of Ilorin, P. M. B. 1515, Ilorin, Nigeria.
E-mail: aodejumo@unilorin.edu.ng

Many authors have proposed different means of dealing with incomplete cross-classified tables. Woolson and Clarke (1984) proposed single-sample methods, which view the fully and partially classified data as a single multinomial sample. This approach corresponds to the common situation in which the occurrence of missing data cannot be predicted in advance. Hocking and Oxspring (1971, 1974) also proposed another method called multiple-sample, which also views the fully and partially classified data as independent multinomial samples. This approach is appropriate when data are missing by design, usually due to practical constraints. These two approaches yield the same maximum likelihood estimators of the cell probabilities if the probabilities of missing classifications are homogeneous (Haber, Chen, and Williamson, 1991). Another method that restricted the parameterisation of general interest in the two-dimensional case was presented by Chen and Fienberg (1974). The approach viewed both completely and partially cross-classified data as a single sample, and considered both multinomial and Poisson sampling schemes. Lipsitz *et al.* (1998) also obtained the maximum likelihood estimates in an incomplete contingency table by using a Poisson generalized linear model. Weighted least squares analysis was also applied to the problem of missing values in a contingency table by using multiple-sample and single-sample approaches respectively by Koch *et al.* (1972) and Woolson and Clarke (1984). A framework for characterizing the estimation of parameters in incomplete data problem was given by Rubin (1974), by decomposing the original estimation problem into a product of complete data factors, (using standard complete table) and missing data factors, (using special incomplete data techniques). Factorization of the likelihood is used to a great advantage in the analysis of contingency tables containing both complete and partially cross-classified data. Maximum likelihood estimates (MLE) are obtained for each factor of the likelihood function, and evaluating the function at the MLEs of the individual factors to obtain the MLEs of the entire likelihood function (Hocking and Oxspring, 1974; Chen and Fienberg, 1974; Little and Rubin, 2002). Shih (1987) also considered maximum likelihood estimates and likelihood ratio test for square tables when there are missing data. Expectation maximization (EM) algorithm has been used by numerous authors as a convenient way to maximize the observed data likelihood via the complete data likelihood (Dempster, Laird and Rubin, 1977; Shih, 1987; Baker and Laird, 1988; Chambers and Welsh, 1993 and others). The main advantages of the EM algorithm are its generality and stability and, given a method to analyze complete data, its ease of implementation. However, its drawbacks are its typically slow rate of convergence and its lack of the direct provision of a measure of precision for the estimators. Louis (1982), Meilijson (1989), Meng and Rubin (1991), Baker (1992), and Meng and Van Dyk (1997) have proposed methods of overcoming these limitations of EM algorithm. Combinations of the EM algorithm and Newton-Raphson iterations, which require explicit modelling of the expected cell counts, are considered by Baker (1994) to reduce sensitivity to poor starting values and to speed up convergence. Meng and Rubin (1993) and Liu and Rubin (1994) proposed extensions of EM called Expectation-Conditional Maximization (ECM) and Expectation/ Conditional Maximization Either (ECME) respectively.

In this research, rather than the usual likelihood approach, we present a data based method of imputing the missing ratings called percentage missing allocation proportional to size

(PMAPS). We also assume ignorability criteria for the missingness mechanism. In section 1 we present the missing ratings mechanism. In section 2 we present the likelihood based and the proposed data based approaches. Kappa-like statistics are given in section 3. We present empirical studies using real life data, results and discussion in section 4.

1. MISSING RATINGS

1.1. MISSING VALUES IN THE RATINGS OF RATERS

As said in the introduction, any square contingency table can be used to display joint ratings of two raters since the two raters must work with the same categorical scales, so that the resulting contingency table will then be $I \times I$, if the there are I ratings categories for each rater. Missing observations can be observed in the raw results for some of the subjects involved in the experiment. For example, if a certain number of culture plates of bacteria samples are collected and examined by one rater, but before getting to the next rater, some of the plates break, are misplaced, are wrongly handled by the Laboratory attendant or are wrongly labelled or identified, we can regard such as missing values in that experiment. To this effect, Table 1 describes the process of how missing observations can occur in the raw ratings of two raters on some sets of subjects (Adejumo, 2005).

	Rater 1 (X_{i_1})		Rater 2 (X_{i_2})		
subject (i)	resp (r_{ij})	rating (Zij)	resp (r_{ij})	rating (Zij)	status (S_{ij})
1	1	1	1	2	obs
2	1	1	1	3	obs
3	0	-	1	2	mis
4	1	5	1	1	obs
5	1	3	0	-	mis
6	1	2	1	3	obs
7	0	-	0	-	mis*
⋮	⋮	⋮	⋮	⋮	⋮
n	1	4	1	3	obs

Table 1. Missing pattern for ratings of two raters.

Table 1 shows a typical result of two raters. The $\{resp\}$ column indicates the values of $\{r_{ij}\}$ while $\{rating\}$ column indicates the corresponding values $\{Zij\}$ of the rating of the *ith* subject by the *jth* rater. We define in this case for two raters ($j = 1;\ 2$) as

$$r_{ij} = \begin{cases} 1, & \text{if } z_{ij} \textit{ observed} \\ 0, & \text{if } z_{ij} \textit{ missing}, \quad \forall i = 1.2. \dots, n,\ \ j = 1,2 \end{cases} \tag{1}$$

where z_{ij}constitutes the design matrix for the table. Also the matrix for $R = [R_{i_1}; R_{i_2}]$ based on Table 1 can be of the form

$$R = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 1 \\ 1 & 1 \\ 1 & 0 \\ 1 & 1 \\ 0 & 0 \\ \vdots & \vdots \\ 1 & 1 \end{pmatrix} \tag{2}$$

$$Z = \begin{pmatrix} 1 & 2 \\ 1 & 3 \\ - & 2 \\ 5 & 1 \\ 3 & - \\ 2 & 3 \\ - & - \\ \vdots & \vdots \\ 4 & 3 \end{pmatrix} \tag{3}$$

Each *ith* subject combined response status S_i is defined as

$$S_i = \begin{cases} obs, & \text{if } r_{i_1} = r_{i_2} = 1 \text{ as obsered,} \\ mis, & \text{if } (r_{i_1} = 0 \text{ } and \text{ } r_{i_2} = 1) \text{ or } (r_{i_1} = 1 \text{ } and \text{ } r_{i_2} = 0) \text{ as missing} \\ mis *^1 & \text{if } r_{i_1} = r_{i_2} = 0 \text{ as totally missing, } \forall i = 1, 2, \ldots, n. \end{cases}$$

The dimension of the resulting cross-classified table with missing ratings on both sides of the raters will be $(I + 1) \times (I + 1)$, with the $(I + 1)th$ categories on both sides representing their respective marginal missing ratings. This can be represented by the following indicator function:

$$R_{i_h} = \begin{cases} 1, & if\ X_{i_h} \in \{1, 2, \ldots, I\} \\ 0, & if\ X_{i_h} \in (I + 1), \quad h = 1\ or\ 2 \end{cases} \tag{4}$$

1.2. MISSING RATINGS PATTERN

The missing pattern depends on the nature of missingness as presented in Table 1. There are two possible missing patterns which can occur between two raters in their resulting cross-classified ratings' table. These are: two-way with either one sided missing pattern or two sided missing patterns.

1.2.1. TWO-WAY WITH ONE SIDED MISSING PATTERN

This is a situation where only one of the two raters is observed incompletely. For example, in an experiment, if rater 1 rates all the n subjects and rater 2 only rates t of the

n subjects, then $\{n-t\}$ subjects' ratings are missing through rater 2. Tables 2 and 3 show respectively the situations where only ratings of rater 2 and rater 1 are observed incompletely.

		Rater 2					missing
	category	1	2	$\cdots$	I	total	t_{j+}
Rater 1	1	n_{11}	n_{12}	$\cdots$	n_{1I}	n_{1+}	t_{1+}
	2	n_{21}	n_{22}	$\cdots$	n_{2I}	n_{2+}	t_{2+}
	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
	I	n_{I1}	n_{I2}	$\cdots$	n_{II}	n_{I+}	t_{I+}
	total	n_{+1}	n_{+2}	$\cdots$	n_{+I}	n_{++}	

Table 2. Cross-classified table for two raters with missing ratings from only rater 2

		Rater 2				
	category	1	2	$\cdots$	I	total
Rater 1	1	n_{11}	n_{12}	$\cdots$	n_{1I}	n_{1+}
	2	n_{21}	n_{22}	$\cdots$	n_{2I}	n_{2+}
	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
	I	n_{I1}	n_{I2}	$\cdots$	n_{II}	n_{I+}
	total	n_{+1}	n_{+2}	$\cdots$	n_{+I}	n_{++}
missing	w_{+k}	w_{+1}	w_{+2}	$\cdots$	w_{+I}	

Table 3. Cross-classified table for two raters with missing ratings from only rater 1

1.2.2. TWO-WAY WITH TWO SIDED MISSING PATTERNS

This is a situation where some of the subjects are observed incompletely and by the two raters. For instance, rater 1 may observe some of the subjects which rater 2 may not be able to observe or vice versa. But in a situation where both of them are unable to observe some subjects, such subjects have to be removed completely from the analysis based on the ignorability criteria rules. Table 4 gives the description of a two-way cross-classified table of ratings with missing ratings from both sides.

		Rater 2					missing
	category	1	2	$\cdots$	I	total	t_{j+}
Rater 1	1	n_{11}	n_{12}	$\cdots$	n_{1I}	n_{1+}	t_{1+}
	2	n_{21}	n_{22}	$\cdots$	n_{2I}	n_{2+}	t_{2+}
	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
	I	n_{I1}	n_{I2}	$\cdots$	n_{II}	n_{I+}	t_{I+}
	total	n_{+1}	n_{+2}	$\cdots$	n_{+I}	n_{++}	
missing	w_{+k}	w_{+1}	w_{+2}	$\cdots$	w_{+I}		

Table 4. Cross-classified table for two raters with missing ratings on both sides

1.3. MISSING RATINGS MECHANISMS

The missing ratings mechanism will be based on the work of Little and Rubin (1987, 2002). Let us consider a cross-classified table of ratings for N subjects or individuals by two different raters, (Rater 1 (X_{i_1}) and Rater 2 (X_{i_2})), in which some of the subjects or individuals are not observed completely by either or both of the raters. As said in subsection 1.1, let R_{i_1} and R_{i_2} be the indicator functions for the missing ratings for X_{i_1} and X_{i_2} respectively. The classification of missing ratings mechanisms depends on the joint probability

$$f(x_{i_1}, x_{i_2}, r_{i_1}, r_{i_2} | \Pi, \phi) = f(r_{i_1}, r_{i_2} | x_{i_1}, x_{i_2}, \phi) f(x_{i_1}, x_{i_2} | \Pi),$$

where $f(r_{i_1}, r_{i_2} | x_{i_1}, x_{i_2}, \phi)$ is the missing ratings mechanism with parameter vector ϕ. This classification depends on whether $f(r_{i_1}, r_{i_2} | x_{i_1}, x_{i_2}, \phi)$ is independent of both X_{i_1} and X_{i_2}, which is referred to as *Missing completely at random* (MCAR). If $f(r_{i_1}, r_{i_2} | x_{i_1}, x_{i_2}, \phi)$ depends on the observed ratings but not on the missing ratings, it is referred to as *Missing at random* or if $f(r_{i_1}, r_{i_2} | x_{i_1}, x_{i_2}, \phi)$ depends on both the observed ratings as well as the missing ratings, it is referred to as *Missing not completely at random* (MNAR). MCAR is clearly a special case of MAR, although the two (MCAR, MAR) mechanisms are referred to as *ignorable* in the sense that inference does not depend on it. MNAR missing mechanism is sometimes said to be *non-ignorable* (Toutenburg, 2002).

2. IMPUTATION OF MISSING RATINGS

Our major objective in this section is to derive means of redistributing the missing ratings in $(I+1)th$ category into the main $I \times I$ cross-classified table of ratings using a likelihood based approach and our proposed data based approach.

2.1. MAXIMUM LIKELIHOOD ESTIMATION IN INCOMPLETE RATINGS

Let $\pi_{jk} = P(X_{i_1} = j, X_{i_2} = k), k = 1,2, \ldots, I$ be the joint probability of rater 1 and rater 2 for the complete data case. Also let $\pi_{j+} = pr[X_{i_1} = j]$, and $\pi_{+k} = pr[X_{i_2} = k]$ be the marginal probabilities. Since the sample size is fixed, we assume a multinomial probability distribution for this table, thus the joint probability distribution function of X_{i_1}, X_{i_2} is given as

$$f\{X_{i_1}, X_{i_2} | \Pi\} = \prod_{j=1}^{I} \prod_{k=1}^{I} \pi_{jk}^{I[x_{i_1}=j, x_{i_2}=k]},$$

where $I[.]$ is an indicator function. As said, we assumed ignorability criteria for the missingness. As stated in equation 4, let R_{i_1} and R_{i_2} be the indicator functions for the missing ratings of X_{i_1} and X_{i_1} respectively. The complete data for subject i is $(R_{i_1}, R_{i_2}, X_{i_1}, X_{i_2})$ with joint distribution

$$f(x_{i_1}, x_{i_2}, r_{i_1}, r_{i_2} | \Pi, \phi) = f(r_{i_1}, r_{i_2} | x_{i_1}, x_{i_2}, \phi) f(x_{i_1}, x_{i_2} | \Pi). \tag{5}$$

As said in section 1.3, $f(r_{i_1}, r_{i_2} | x_{i_1}, x_{i_2}, \phi)$ is the missing rating mechanism with parameter ϕ.

Our target is to estimate Π using likelihood method, which will determine the size of each cell in the final complete table. To this effect, when either X_{i_1} or X_{i_2} is missing, the contribution of a single element or subject i to the likelihood under the MAR assumption is the sum over either X_{i_1} or X_{i_2}, that is

$$\sum_{x_{i_h}} f(r_{i_1}, r_{i_2} | x_{i_1}, x_{i_2}, \phi) f(x_{i_1}, x_{i_2} | \Pi)$$

where $h = 1$ or 2.

According to Little and Rubin (2002), the full likelihood can be written as

$$L(\phi, \Pi) = L_1(\phi, \Pi) L_2(\phi, \Pi) L_3(\phi, \Pi) \tag{6}$$

each of which represents different missing ratings' pattern, where

$$L_1(\phi, \Pi) = \prod_{i=1}^{N} \left[f(r_{i_1} = 1, r_{i_2} = 1 | x_{i_1}, x_{i_2}, \phi) f(x_{i_1}, x_{i_2} | \Pi) \right]^{r_{i_1} r_{i_2}} \tag{7}$$

$$L_2(\phi, \Pi) = \prod_{i=1}^{N} \left[\sum_{x_{i_1}} f(r_{i_1} = 0, r_{i_2} = 1 | x_{i_1}, x_{i_2}, \phi) f(x_{i_1}, x_{i_2} | \Pi) \right]^{(1 - r_{i_1}) r_{i_2}} \tag{8}$$

$$L_3(\phi, \Pi) = \prod_{i=1}^{N} \left[\sum_{x_{i_2}} f(r_{i_1} = 1, r_{i_2} = 0 | x_{i_1}, x_{i_2}, \phi) f(x_{i_1}, x_{i_2} | \Pi) \right]^{r_{i_1} (1 - r_{i_2})} \tag{9}$$

Furthermore, if the missing values are missing at random (MAR), the missing probability is independent of such observed value (Toutenburg, 2002), then

$$f(r_{i_1} = 0, r_{i_2} = 1 | x_{i_1}, x_{i_2}, \phi) = f(r_{i_1} = 0, r_{i_2} = 1 | x_{i_2}, \phi)$$

which implies that

$$\begin{aligned} L_2(\phi, \Pi) &= \prod_{i=1}^{N} \left[f(r_{i_1} = 0, r_{i_2} = 1 | x_{i_2}, \phi) \sum_{x_{i_2}} f(x_{i_1}, x_{i_2} | \Pi) \right]^{(1 - r_{i_1}) r_{i_2}} \\ &= \prod_{i=1}^{N} \left[f(r_{i_1} = 0, r_{i_2} = 1 | x_{i_2}, \phi) f(x_{i_2} | \Pi) \right]^{(1 - r_{i_1}) r_{i_2}} \end{aligned} \tag{10}$$

where $f(x_{i_2} | \Pi) = \prod_{k=1}^{I} \pi_{+k}^{I[x_{i_2} = k]}$, which is the marginal distribution of X_{i_2}. $L_3(\phi, \Pi)$, will also become

$$L_3(\phi,\Pi) = \prod_{i=1}^{N}\left[f(r_{i_1}=1, r_{i_2}=0|x_{i_1},\phi)\sum_{x_{i_2}} f(x_{i_1},x_{i_2}|\Pi)\right]^{r_{i_1}(1-r_{i_2})}$$
$$= \prod_{i=1}^{N}[f(r_{i_1}=1, r_{i_2}=0|x_{i_1},\phi)f(x_{i_1}|\Pi)]^{r_{i_1}(1-r_{i_2})} \quad (11)$$

where $f(x_{i_1}|\Pi) = \prod_{j=1}^{I}\pi_{j+}^{I[x_{i_1}=k]}$ is also the marginal of X_{i_1}. By selection model approach (Little, 1993), $L(\phi,\Pi)$ in equation (6) can be factored into two components under MAR, which implies

$L(\phi,\Pi) = L(\phi)L(\Pi)$,

where $L(\phi)$ is a function of ϕ which is

$$L(\phi) = \prod_{i=1}^{N}\{[f(r_{i_1}=1, r_{i_2}=1|x_{i_1},x_{i_2},\phi)]^{r_{i_1}r_{i_2}}\}$$
$$\times\left\{[f(r_{i_1}=0, r_{i_2}=1|x_{i_2},\phi)]^{(1-r_{i_1})r_{i_2}}\right\}$$
$$\times\left\{[f(r_{i_1}=1, r_{i_2}=0|x_{i_1},\phi)]^{r_{i_1}(1-r_{i_2})}\right\} \quad (12)$$

Also $L(\Pi)$, which is a function of Π, is

$$L(\Pi) = \prod_{i=1}^{N}\left[f(x_{i_1},x_{i_2}|\Pi)^{r_{i_1}r_{i_2}} f(x_{i_1}|\Pi)^{r_{i_1}(1-r_{i_2})}\right]$$
$$\times\left[f(x_{i_2}|\Pi)^{(1-r_{i_1})r_{i_2}}\right] \quad (13)$$

from which the MLE of ϕ and Π can be obtained. To estimate Π, we simplify equation (13) further by using the notation in Table 4

$$n_{jk} = \sum_{i=1}^{N} r_{i_1}r_{i_2}I[X_{i_1}=j, X_{i_2}=k]$$

$$t_{j+} = \sum_{i=1}^{N} r_{i_1}(1-r_{i_2})I[X_{i_1}=j]$$

$$w_{+k} = \sum_{i=1}^{N}(1-r_{i_1})r_{i_2}I[X_{i_2}=k]$$

to be the numbers of subjects which are observed completely by the two raters, that is, by only rater 1 and by only rater 2 respectively. Therefore the likelihood for the cell probabilities can then be written as

$$L(\Pi) = \left\{\prod_{j=1}^{I}\prod_{k=1}^{I}\pi_{jk}^{n_{jk}}\right\}\left\{\prod_{j=1}^{I}\pi_{j+}^{t_{j+}}\right\}\left\{\prod_{k=1}^{I}\pi_{+k}^{w_{+k}}\right\}. \tag{14}$$

where the first term is for the complete case on subjects with $(r_{i_1} = R_{i_1} = 1,\ r_{i_2} = R_{i_2} = 1)$, which give a multinomial likelihood with sample $\{n_{++}\}$ and $(I^2 - 1)$ nonredundant probabilities $\{\pi_{11}, \pi_{12}, \dots, \pi_{I^2-1}\}$, the second term is for when subjects are observed only by X_{i_1} , which give a multinomial likelihood with sample size $\{t_{++}\}$ and $(I-1)$ nonredundant probabilities $\{\pi_{1+}, \pi_{2+}, \dots, \pi_{(I-1)+}\}$, and the third term is for when subjects are observed only by X_{i_2}, which give a multinomial likelihood with sample size $\{w_{++}\}$ and $(I-1)$ nonredundant probabilities $\{\pi_{+1}, \pi_{+2}, \dots, \pi_{+(I-1)}\}$.

Then the observed incomplete ratings likelihood function in (14) can also be presented as

$$L(\Pi) = \prod_{j=1}^{I}\prod_{k=1}^{I}\left\{\pi_{jk}^{n_{jk}}\pi_{j+}^{t_{j+}}\pi_{+k}^{w_{+k}}\right\} \tag{15}$$

Under the complete data case, if the cell counts $\{n_{jk}\}$ are independent Poisson random variables with mean $\{m_{jk} = n_{++}\pi_{jk}\}$ and cell probabilities $\left\{\pi_{jk}^{*} = \frac{m_{jk}}{\sum_j \sum_k m_{jk}}\right\}$, and if the missing mechanism is ignorable, as we have assumed in this paper, likelihood inferences for $\{\pi_{jk}^{*}\}$ are the same as for those $\{\pi_{jk}\}$ under the multinomial model (Bishop *et. al.*, 1975). The only difference is the way they are expressed. Multinomial is expressed in terms of the cell probabilities $\{\pi_{jk}\}$ while Poisson model is expressed in terms of the expected cell counts $\{m_{jk} = n_{++}\pi_{jk}\}$. Thus, one can use any of the two models and still arrive at the target. Agresti (1990) provided the multinomial likelihood approximation to Poisson likelihood model only with some constrains on the sample sizes and probabilities.

We used EM algorithm to maximize the likelihood equation in (14).

2.2. DATA BASED APPROACH IN INCOMPLETE RATINGS CASE

The major objective in this paper is to propose another means of redistributing the missing ratings in $(I + 1)th$ categories into the main $I \times I$ cross-classified table of ratings using the data based approach. For this we assume that certain percentages of the missing ratings are to be reallocated in the diagonal or off-diagonal section of the table. Consider a cross-classified table of ratings with missing ratings on both sides of the table as in Table 5 n_{jk} is the cell counts for the completely observed ratings, while t_{j+} and w_{+k} are the marginal partially observed ratings by raters 1 and 2 respectively. $j = k = 1, 2, \dots I$ is for the completely observed categories and $j = k = I + 1$ is for the missing categories at the margin, although the missing pattern depends on the nature of missingness between the two raters as we have presented in Tables 2 to 4. Now, considering the missing pattern of the form in Table 5, let the rating scale be from 1 to 5. If rater 1's response to one subject is 1, and the rater 2's response to that same subject is missing, and we assume that this missing response is also 1, it implies that the missing is in the diagonal cell. But if we assume it is not 1, it will then be any of letters 2 to 5 and the missing of that particular subject will be in the off-diagonal cell. Generally, if there are many responses, 'a,b,c ...', that are missing, all or a certain percentage may fall along the diagonal or in the off-diagonal cells and sometimes both off and along the diagonal combined. We may take a certain ψ percentage of the missing values at the margins to be in their respective diagonal and the remaining $(1-\psi)$ to be in the off-diagonal cells. This gives a combination of the two (along and off-diagonal).

		Rater 2					$(I+1)$
	category	1	2	...	I	total	t_{j+}
Rater 1	1	n_{11}	n_{12}	...	n_{1I}	n_{1+}	t_{1+}
	2	n_{21}	n_{22}	...	n_{2I}	n_{2+}	t_{2+}
	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
	I	n_{I1}	n_{I2}	...	n_{II}	n_{I+}	t_{I+}
	total	n_{+1}	n_{+2}	...	n_{+I}	n_{++}	
$(I+1)$	w_{+k}	w_{+1}	w_{+2}	...	w_{+I}		

Table 5. Cross-classified table for two raters with missing ratings on both sides

To achieve this, we propose a method called percentage missing allocation proportional to size (PMAPS). In this method we still assume the missing at random (MAR) missingness mechanism criteria. Let $n^*_{jk}, j = k = 1, 2, \dots I$, represent the new sets of cell counts based on (PMAPS) such that:

$$n^*_{jk} = \begin{cases} n_{jk} + \psi t_{j+} + \psi w_{+k}, & \text{if } j = k \\ n_{jk} + (1-\psi)t_{j+} + \{\pi_{j|k}\} + (1-\psi)w_{+k}\{\pi_{k|j}\}, & \text{if } j \neq k \end{cases} \tag{16}$$

where ψ is the percentage (%) to be specified,

$\pi_{j|k} = \frac{n_{jk}}{\sum_{k,k\neq j} n_{jk}}$ and $\pi_{k|j} = \frac{n_{jk}}{\sum_{j,j\neq k} n_{jk}}$ are the cell sizes (proportions) based on their respective marginal total minus the diagonal cell count without missing values,

t_{j+} and w_{+k} are the marginal total missing ratings respectively for the two raters.

PMAPS estimates effectively distribute $\frac{n_{jk}}{\sum_{k,k\neq j} n_{jk}}$ and $\frac{n_{jk}}{\sum_{j,j\neq k} n_{jk}}$ of the unclassified ratings t_{j+} and w_{+k} into the off-diagonal $(j,k)th$ cell based on the specified percentage $(1-\psi)$ with the rest ψ in the diagonal $(j,j)th$ cell.

Also the corresponding cell probabilities π^*_{jk} based on PMAPS for (16) will be

$$\pi^*_{jk} = \begin{cases} n*^{-1}\left[\, n_{jk} + \psi t_{j+} + \psi w_{+k}\right], & \text{if } j = k \\ n*^{-1}\left[\, n_{jk} + (1-\psi)t_{j+} + \{\pi_{j|k}\} + (1-\psi)w_{+k}\{\pi_{k|j}\}\right], & \text{if } j \neq k \end{cases} \quad (17)$$

where $n* = \sum_j \sum_k n^*_{jk}$ is the total estimated with PMAPS, which is equivalent to $\{n = \sum_j \sum_k n_{jk} + \sum_j t_{j+} + \sum_k w_{+k}\}$ total subjects or observations in the experiment or trials that are observed at least by one rater.

The large sample variance for this estimate is

$$Var(\pi^*_{jk}) = \frac{\pi^*_{jk}(1-\pi^*_{jk})}{n*} \quad (18)$$

Then $100(1-\alpha)\%$ confidence interval for each π^*_{jk} will be

$$\pi^*_{jk} \pm z_1 - \frac{\alpha}{2}\sqrt{\frac{\pi^*_{jk}(1-\pi^*_{jk})}{n*}} \quad (19)$$

Case-by-case formulations for missing structures with PMAPS

In order to examine further the effects of allocating certain percentages of the missing observation to the diagonal or off-diagonal cells of a given table of ratings of n_{++} subjects, we allowed certain percentages (0%, 50%, 75% and 100%) of the missing values from various marginal total missing values (t_{j+} and w_{+k}) to be distributed to the diagonal cells.

3. SOME KAPPA-LIKE STATISTICS FOR AGREEMENT MEASUREMENT

We considered two Kappa-like statistics (Cohen Kappa statistic and Intraclass kappa statistic) that are freely used to measure agreement without any attached weight (Shoukri, 2004). We used these statistics to obtain the measure of agreement that exists between raters for different imputed matrix of $\theta = (\pi_{jk}; j = k = 1,2,\ldots,I)$.

Cohen (1960) proposed a standardized coefficient of raw agreement for nominal scales in terms of the proportion of the subjects classified into the same category by the two observers, which is called Cohen Kappa statistic, estimated as

$$\hat{k}_c = \frac{\hat{\pi}_o - \hat{\pi}_e}{1 - \hat{\pi}_e} \quad (20)$$

where

$$\hat{\pi}_o = \sum_{i=1}^{I} \pi_{ii} \quad (21)$$

and

$$\hat{\pi}_e = \sum_{i=1}^{I} \pi_{i+}\pi_{+i}. \tag{22}$$

To determine whether $\hat{k}$ differs significantly from zero, one could use the asymptotic variance formulae given by Fleiss *et al.* (1969) for the general $I \times I$ tables. Under the hypothesis of only chance agreement, the estimated large-sample variance of $\hat{k}$ is given by

$$\widehat{var_o}(\hat{k}_c) = \frac{\pi_e + \pi_e^2 - \sum_{i=1}^{I} \pi_{i+}\pi_{+i}(\pi_{i+}\pi_{+i})}{n(1-\pi_e)^2} \tag{23}$$

Assuming that

$$\frac{\hat{k}}{\sqrt{\widehat{var_o}(\hat{k})}} \tag{24}$$

follows a normal distribution, one can test the hypothesis of chance agreement by reference to the standard normal distribution, and the confidence interval (CI) of size

$$100(1-\alpha)\% = \hat{k} \pm Z_{1-\frac{\alpha}{2}} SE(\hat{k})$$

can be obtained for $\hat{k}$, where SE is the standard error.

The intraclass kappa by Barnhart and Williamson (2002) Kappa for measuring agreement (for $I \times I$ tables) between two readings for a categorical response with I categories when the two readings are replicated measurements assume no bias because the probability of a positive rating is the same for the two readings due to replication. It is given as

$$\hat{k}_{In} = \frac{\sum_{i=1}^{I} \pi_{ii} - \sum_{i=1}^{I}((\pi_{i+} + \pi_{+i})/2)^2}{1 - \sum_{i=1}^{I}((\pi_{i+}\pi_{+i})/2)^2} \tag{25}$$

with variance for the estimated value as

$$\widehat{var_o}(\hat{k}_{In}) = \frac{\pi_{exp} + (\pi_{exp})^2 - \pi_{exp}(\sum_{i=1}^{I}((\pi_{i+}\pi_{+i})/2)^2)}{n(1-\pi_{exp})^2} \tag{26}$$

where

$$\pi_{exp} = \sum_{i=1}^{I}((\pi_{i+} + \pi_{+i})/2)^2$$

We also assumed that

$$\frac{\hat{k}_{In}}{\sqrt{\widehat{var_o}(\hat{k}_{In})}} \tag{27}$$

follows a normal distribution. Also the confidence interval (CI) of size

$$100(1-\alpha)\% = \hat{k}_{In} \pm Z_{1-\frac{\alpha}{2}} SE(\hat{k}_{In})$$

can be obtained for $\hat{k}_{In}$.

4. EMPIRICAL STUDIES, RESULTS AND DISCUSSION

4.1. EMPIRICAL STUDIES

An algorithm was developed for the approach discussed in sections 3 and 4 based on the underlying assumptions and sufficient statistics stated there. In order to justify the implementation of this method, consider the data from Adejumo (2005) in Tables 6 to 8 which were High Blood Pressure (HBP) tests conducted for a sample of in-patients in University of Ilorin Teaching Hospital (UITH), Ilorin, Nigeria in October, 2004 by three Nurses, taken independently at different times of their working hours of the day on each of 365 patients.

Tables 6 to 8 show the cross-classification of the rating (including missing ratings) for the Nurses.

		Nurse 2						missing
	category	1	2	3	4	5	total	t_{j+}
Nurse 1	1	32	24	11	1	0	68	4
	2	0	10	26	15	0	51	3
	3	0	2	24	30	15	71	7
	4	0	0	3	18	37	58	6
	5	0	0	0	3	57	60	7
	total	32	36	64	67	109	306	
missing	w_{+k}	4	6	5	3	11		1

Table 6. Cross-classified table for the high blood pressure readings of some in-patients by nurse 1 and nurse 2 with missing values on both sides (N12)

		Nurse 3						missing
	category	1	2	3	4	5	total	t_{j+}
Nurse 1	1	68	0	1	0	0	69	5
	2	37	15	2	0	0	54	2
	3	22	33	16	2	0	73	8
	4	1	15	25	11	3	55	5
	5	0	0	10	30	19	59	7
	total	128	63	54	43	22	310	
missing	w_{+k}	10	6	5	2	4		1

Table 7. Cross-classified table for the high blood pressure readings of some in-patients by nurse 1 and nurse 3 with missing values on both sides (N13)

	category	Nurse 3 1	2	3	4	5	total	missing t_{j+}
Nurse 2	1	32	0	0	0	0	32	2
	2	31	1	0	0	0	32	3
	3	46	20	1	0	0	67	7
	4	14	38	13	4	0	69	8
	5	0	4	45	41	22	112	9
	total	123	63	59	45	22	312	
missing	w_{+k}	8	5	4	3	3		1

Table 8. Cross-classified table for the high blood pressure readings of some in-patients by nurse 2 and nurse 3 with missing values on both sides (N23)

To use PMAPS method to impute the cell proportions $\theta = (\pi_{jk}; j = k = 1,2,\dots,I)$, we have the following steps.

1. Fix the ψ, which is the percentage of the missing ratings to be distributed to the diagonal cells. The second step is to construct two contingency tables of proportion obtained with $\pi_{j|k} = \frac{n_{jk}}{\sum_{k,k\neq j} n_{jk}}$ and $\pi_{k|j} = \frac{n_{jk}}{\sum_{j,j\neq k} n_{jk}}$, which are the cell proportion base on their respective marginal totals excluding the diagonal cells.

2. Estimate $\{n_{jk}^{*}, j = k = 1,2,\dots I\}$ by substituting ψ, $\pi_{j|k}$ and $\pi_{k|j}$ in the following equation

$$n_{jk}^{*} = \begin{cases} n_{jk} + \psi t_{j+} + \psi w_{+k}, & \text{if } j = k \\ n_{jk} + (1-\psi)t_{j+}\{\pi_{j|k}\} + (1-\psi)w_{+k}\{\pi_{k|j}\}, & \text{if } j \neq k \end{cases} \tag{28}$$

3. Calculate the estimates for the parameter $\hat{\theta} = \{\hat{\pi}_{jk}; j,k = 1,2,\dots,I\}$ by using the following equation

$$\pi_{jk}^{*} = \begin{cases} n*^{-1}\left[\, n_{jk} + \psi t_{j+} + \psi w_{+k}\right], & \text{if } j = k \\ n*^{-1}\left[\, n_{jk} + (1-\psi)t_{j+}\{\pi_{j|k}\} + (1-\psi)w_{+k}\{\pi_{k|j}\}\right], & \text{if } j \neq k \end{cases} \tag{29}$$

In order to compare our result with the Maximum Likelihood approach (section 2.1), we applied the EM algorithm to obtain the MLE of θ for the observed incomplete ratings likelihood function (15). The algorithm began with some initial estimates of the parameter. We replaced the indicator function involving missing rating in the sufficient statistics $\left\{u_{jk} = \sum_{i=1}^{N} I_{(x_{i_1}=j,\, x_{i_2}=k)}\right\}$ with their conditional expectations at each iteration, given the observed data and the current parameter estimates at E-step. We then obtained new parameter estimates from these estimated sufficient statistics as if they had come from a complete sample (M-step). We continued the iteration until it converged. One can start specifically by putting $\hat{\theta}_1 = \{\hat{\pi}_{jk}; j,k = 1,2,\dots,I\}$ as the current estimate of $\theta = \{\pi_{jk}; j,k = 1,2,\dots,I\}$. We took the initial estimates from the $\pi_{jk} = \left\{\frac{n_{jk}}{n_{.}}; \forall j,k = 1,2,\dots,I\right\}$. The E-step and the M-step were as follows:

E-step: Fill in the indicator functions involving missing data for $i = (n+1)\ to\ (n+t)$ and $i = (n+t+1)\ to\ (n+t+w)$ respectively with

$$\begin{aligned} I_{(X_{i_1}=j,X_{i_2}=k),1} &= E_{\hat{\theta}_1} I_{(X_{i_2}=k|X_{i_1}=j)} I_{(X_{i_1}=j)} \\ &= \left(\frac{\hat{\pi}_{jk},1}{\hat{\pi}_{j+},1}\right) I_{(X_{i_2}=k|X_{i_1}=j)} I_{(X_{i_1}=j)} \end{aligned} \tag{30}$$

and

$$\begin{aligned} I_{(X_{i_1}=j,X_{i_2}=k),1} &= E_{\hat{\theta}_1} I_{(X_{i_1}=j|X_{i_2}=k)} I_{(X_{i_2}=k)} \\ &= \left(\frac{\hat{\pi}_{jk},1}{\hat{\pi}_{+k},1}\right) I_{(X_{i_1}=j|X_{i_2}=k)} I_{(X_{i_2}=k)} \end{aligned} \tag{31}$$

Then, for all $j = k = 1,2,\dots,I$ replace the sufficient statistics with

$$\begin{aligned} \hat{n}_{jk,1} &= \sum_{i=1}^{n} I_{(X_{i_1}=j,X_{i_2}=k),1} + \sum_{i=n+1}^{n+t} I_{(X_{i_1}=j,X_{i_2}=k),1} + \sum_{i=n+t+1}^{n+t+k} I_{(X_{i_1}=j,X_{i_2}=k),1} \\ &= n_{jk} + t_{j+}\left(\frac{\hat{\pi}_{jk},1}{\hat{\pi}_{j+},1}\right) + w_{+k}\left(\frac{\hat{\pi}_{jk},1}{\hat{\pi}_{+k},1}\right) \end{aligned} \tag{32}$$

M-step: Compute the new estimates

$$\hat{\theta}_{(m)} = \{\hat{\pi}_{jk,m}; j,k = 1,2,\dots,I\}, \tag{33}$$

where $\left\{\hat{\pi}_{jk,m} = \frac{\hat{n}_{jk,1}}{N}\right\}$. Continue the circle until it converges. Thus, equation (33) provides the required MLEs $\hat{\theta} = \{\hat{\pi}_{jk}; j,k = 1,2,\dots,I\}$ for the likelihood in equation (14).

These two approaches (likelihood and data based) were compared using empirical study based on the assumptions and sufficient statistics stated under each, using the data in Tables 6 to 8 as test cases.

4.2. RESULTS

4.2.1. LIKELIHOOD BASED VIA EM ALGORITHM

As said, our interest was to estimate the cell probabilities for the complete table. For EM algorithm we iteratively computed the cell proportions $\theta = \{\pi_{jk}; j, k = 1,2, \ldots, I\}$ until it converged based on the algorithm stated in section 4.1. The first step was to replace missing values by estimated values. This is done by supplying the starting value for the first iteration of the algorithm using the $(\hat{\pi}_{jk}; j = k = 1,2, \ldots, I)$ calculated from the partially complete table. The second step was to substitute $(\hat{\pi}_{jk}; j = k = 1,2, \ldots, I)$ in the equation

$$\hat{u}_{jk,1} = n_{jk} + t_j\left(\frac{\hat{\pi}_{jk}, 1}{\hat{\pi}_{j+}, 1}\right) + w_k\left(\frac{\hat{\pi}_{jk}, 1}{\hat{\pi}_{+k}, 1}\right) \tag{34}$$

The third step was to calculate $\hat{\theta}_m = \{\hat{\pi}_{jk,m}; j, k = 1,2, \ldots, I\}$, using

$$\hat{\theta}_{(m)} = n^{-1}\left\{ n_{jk} + t_j\left(\frac{\hat{\pi}_{jk}, 1}{\hat{\pi}_{j+}, 1}\right) + w_k\left(\frac{\hat{\pi}_{jk}, 1}{\hat{\pi}_{+k}, 1}\right), \right\} \tag{35}$$

where $\{n = \sum_j \sum_k n_{jk} + \sum_j t_j + \sum_k w_{+k}\}$. The process was continued until it converged. This could be done with either *R* or *Splus* statistical programming packages.

We applied this on the data, so for Nurses 1 & 2; 1 & 3; and 2 & 3, the estimates for the parameter $\hat{\theta} = \{\hat{\pi}_{jk}; j, k = 1,2, \ldots, I\}$ obtained at 6th iteration for Nurses 1 & 2, 1 & 3 and 2 & 3 respectively were:

$$\hat{\pi}_{12_{jk}} = \begin{pmatrix} 0.104 & 3.74\text{e-}07 & 1.50\text{e-}06 & 1.47\text{e-}06 & 2.04\text{e-}06 \\ 0.0809 & 0.0338 & 0.0070 & 3.62\text{e-}06 & 4.78\text{e-}06 \\ 0.0342 & 0.0811 & 0.078 & 0.010 & 8.38\text{e-}07 \\ 0.00302 & 0.0453 & 0.0941 & 0.0566 & 0.0095 \\ 3.98\text{e-}07 & 3.88\text{e-}07 & 0.0495 & 0.1225 & 0.1908 \end{pmatrix} \tag{36}$$

$$\hat{\pi}_{13_{jk}} = \begin{pmatrix} 0.215 & 0.113 & 0.0719 & 0.0032 & 1.28\text{e-}06 \\ 4.04\text{e-}07 & 0.0466 & 0.109 & 0.0489 & 1.39\text{e-}06 \\ 0.0032 & 0.0062 & 0.0529 & 0.0813 & 0.0333 \\ 0.00\text{e-}00 & 0.00\text{e-}00 & 0.0063 & 0.0342 & 0.0956 \\ 1.68\text{e-}6 & 5.099\text{e-}07 & 4.34\text{e-}06 & 0.011 & 0.0684 \end{pmatrix} \tag{37}$$

$$\hat{\pi}_{23_{jk}} = \begin{pmatrix} 0.099 & 0.0987 & 0.148 & 0.0454 & 3.36\text{e-}07 \\ 0.00\text{e-}00 & 0.00322 & 0.0651 & 0.125 & 0.0127 \\ 0.00\text{e-}00 & 2.16\text{e-}07 & 0.00322 & 0.0423 & 0.1418 \\ 0.00\text{e-}00 & 0.00\text{e-}00 & 4.07\text{e-}07 & 0.0130 & 0.129 \\ 3.50\text{e } 07 & 8.42\text{e } 07 & 1.76\text{e } 06 & 2.30\text{e } 06 & 0.0738 \end{pmatrix} \tag{38}$$

The corresponding cell counts for each of these equations based on the 364 in-patients that were observed at least by one of the nurses was obtained by multiplying the estimates of parameter $\hat{\theta} = \{\hat{\pi}_{jk}; j, k = 1,2, \dots, I\}$ in (36) to (38) by 364.

4.2.2. DATA BASED WITH PMAPS

In the case of the proposed PMAPS method we also computed the cell proportions $\theta = \{\pi_{jk}; j, k = 1,2, \dots, I\}$ following the steps stated in section 4.1. Under this method, we considered two structures as stated in subsection 3.2. We assigned to be 0%, 50%, 75% and 100%, keeping the estimate of Kappa statistic $\hat{k}_c$ given in equation 20 from the completely observed part of the table constant.

To illustrate this, we used the same sets of data used under likelihood approach. For each of the data set, we obtained the estimates of the parameter $\hat{\theta} = \{\hat{\pi}_{jk}; j, k = 1,2, \dots, I\}$ for cases when ψ = 0%, 50%, 75%, and 100%. We created bounds for each estimate of $\hat{\theta} = \{\hat{\pi}_{jk}; j, k = 1,2, \dots, I\}$ by using equations 18 and 19. We obtained the 95% confidence interval for each of the Kappa-like statistics estimates for the imputed matrices under the two approaches which are in Tables 9 and 10.

4.3. DISCUSSION

Likelihood approach has been used to a great advantage, the proposed data based approach called PMAPS is better way of estimating the missing ratings. We observed that both approaches have the same pattern based on the imputed matrices, as well as the summary tables of 95% confidence interval for their respective estimates of Kappa-like statistics in Tables 9 and 10. The agreement estimates obtained from these tables under PMAPS approach distribute around that of EM, depending on the level of ψ. The advantages this approach is having over that of EM are numerous; imputing multiple estimates for the missing rating can be achieved, rather than a single estimate produced by the EM approach. The proposed approach is very easy to apply compare with the EM approach. This proposed approach can as well be improved on by generating random numbers between 0 and 1 for ψ or by narrowing down the intervals for the percentages of preassigned ψ, for instance, an interval of 2% or 5%, so as to have more chances of imputing the missing ratings.

			95% CI for k				95% CI for k_{I_n}	
Table	k_c	S.E	lower	upper	k_{I_n}	S.E	lower	upper
N_{12}	0.3279	0.02556	0.2778	0.3780	0.3187	0.03365	0.2528	0.3847
N_{12}	0.2673	0.02544	0.2174	0.3172	0.2543	0.03453	0.1866	0.3219
N_{12}	0.03997	0.02147	-0.002123	0.08205	-0.0156	0.03270	-0.07974	0.04846

Table 9. 95% confidence bounds for the two Kappa-like statistics estimates for the imputed table by EM

				95% CI for k				95% CI for k_{l_w}	
ψ	Table	k_c	S.E	lower	upper	k_{l_w}	S.E	lower	upper
0%	N_{12}	0.2408	0.02537	0.1911	0.2905	0.2291	0.03353	0.1634	0.2948
	N_{12}	0.1911	0.02536	0.1414	0.2408	0.1765	0.03423	0.1094	0.2436
	N_{12}	0.0148	0.02108	-0.0265	0.05614	-0.0459	0.03259	-0.1099	0.01788
50%	N_{12}	0.3297	0.02589	0.2793	0.38001	0.3217	0.03360	0.25579	0.3875
	N_{12}	0.2811	0.0256	0.2309	0.3313	0.2700	0.03438	0.2026	0.3374
	N_{12}	0.0896	0.02195	0.04658	0.1326	0.04256	0.03262	-0.02138	0.1065
75%	N_{12}	0.3780	0.02578	0.3275	0.4285	0.3712	0.03361	0.3053	0.4370
	N_{12}	0.3261	0.0257	0.2757	0.3766	0.3165	0.03447	0.2490	0.3841
	N_{12}	0.1274	0.02235	0.0836	0.1712	0.08635	0.03264	0.02238	0.15033
100%	N_{12}	0.4248	0.02591	0.3740	0.4756	0.41926	0.03370	0.3532	0.4853
	N_{12}	0.3720	0.02585	0.3213	0.4227	0.3638	0.03456	0.2960	0.43149
	N_{12}	0.1655	0.02274	0.12090	0.2100	0.1299	0.03266	0.06585	0.1939

Table 10. 95% confidence bounds for the two Kappa-like statistics estimates for the imputed table by PMAPS

REFERENCES

1. Adejumo, A. O. Modelling Generalized (loglinear) models on raters Agreement measures with complete and missing values cases; Ph D dissertation in Institut für Statistik, Ludwig-Maximilians-Universität, München, Germany, 2005.
2. Agresti, A. Categorical data analysis; Wiley, 1990.
3. Baker, S. G. A simple method for computing the observed information matrix when using EM algorithm with categorical data; J. Comput. Graph. Statist.; 1, 1992, pp. 63-73.
4. Baker, S. G. Composite linear models for incomplete multinomial data; Statist. Med.; 13, 1994, pp. 609-622.
5. Baker, S. G. and Laird, N. M. Regression analysis for categorical variables with outcome subject to nonignorable non-response; J. Amer. Statist. Asso.; 83, 1988, pp. 62-69.
6. Barnhart, H. X. and Williamson, J. M. Weighted Least squares approach for comparing correlated kappa; Biometrics; 58, 2002, pp. 1012-1019.
7. Bishop, Y. V. V., Fienberg, S. E. and Holland, P. W. Discrete multivariate analysis; Cambridge, MA: MIT Press, 1975.
8. Chamber, R. L. and Welsh, A. H. Loglinear models for survey data with non-ignorable non-response; J. Roy. Statist. Soc. Series B; 55, 1993, pp. 157-170.
9. Chen, T. and Fienberg, S. E. Two-dimensional contingency tables with both completely and partially cross-classified data; Biometrics; 30, 1974, pp. 629-642.
10. Cohen, J. A coefficient of agreement for nominal scales; Edu. and Psych. Meas.; 20, 1960, pp. 37-46.

11. Dempster, A. P., Laird, N. M. and Rubin, D. B. Maximum likelihood from incomplete data via the EM algorithm; J. of Roy Statist. Socie. Series B; 39, 1977, pp. 1 - 38.
12. Fleiss, J. L., Cohen, J. and Everitt, B. S. Large sample standard errors of kappa and weighted kappa; Psych. Bull.; 72, 1969, pp. 323-327.
13. Haber, M., Chen, C.C., Williamson, G. D. Analysis of repeated categorical responses from fully and partially cross-classified data; Comm. in Statist.; 20, 1991, pp. 3293-3313.
14. Hocking, R. R., and Oxspring, H. H. Maximum likelihood estimation with incomplete multinomial data; J. Amer. Statist. Assoc.; 66, 1971, pp. 65-70.
15. Hocking, R. R. and Oxspring, H. H. The Analysis of partially categorized contingency data; Biometrics; 66, 1974, pp. 65-70.
16. Koch, G. G., Imrey, P. B. and Reinfurt, D. W. Linear model analysis of categorical data with incomplete response vectors; Biometrics; 28, 1972, pp. 663-692.
17. Lipsitz, S. R., Parzen, M. and Molenberghs, G. Obtaining the maximum likelihood estimates in incomplete $(r \times c)$ contingency tables using a Poisson generalized linear model; J. Comp. and Graphics Statist.; 7(3), 1998, pp. 356-376.
18. Little, R. J. A. Pattern-mixture models for multivariate incomplete data; J. Amer. Statist. Asso.; 88, 1993, pp. 125-138.
19. Little, R, J. A., Bubin, D. B. Statistical analysis with missing Data. New York: Wiley, 1987.
20. Little, R, J. A. and Bubin, D. B. Statistical analysis with missing Data. 2nd Edit., New York: Wiley, 2002.
21. Liu, C. H. and Rubin, D. B. The ECME algorithm: a simple extension of EM and ECM with fast monotone convergence; Biometrika; 81, 1994, pp. 633-648.
22. Louis, T. A. Finding the observed information matrix when using the EM algorithm; J. Roy. Statist. Soc. Series B; 44, 1982, pp. 226-233.
23. Meilijson, I. A fast improvement to the EM algorithm on its own terms; J. Roy. Statist. Soc. Series B; 51, 1989, pp. 127-138.
24. Meng, X. L. and Van Dyk, D. A. The EM algorithm-an old folk song sung to a fast new tune (with discussion); J. Roy. Statist. Soc. Series B; 59, 1997, pp. 511-567.
25. Meng, X. L. and Rubin, D. B. Using EM to obtain asymptotic variance-covariance matrices: the SEM algorithm; J. Amer. Statist. Asso.; 86, 1991, pp. 899-909.
26. Meng, X. L. and Rubin, D. B. Maximum likelihood estimation via ECM algorithm: A general framework; Biometrika; 80, 1993, pp. 267-278.
27. Rubin, D. B. Characterizing the estimation of parameters in incomplete-data problems; J. Amer. Statist. Assoc.; 69, 1974, pp. 467-474.
28. Shih, W. J. Maximum likelihood estimation and likelihood ratio test for square tables with missing data; Statist. Med.; 6, 1987, pp. 91-97.
29. Shoukri, M. M. Measures of interobserver agreement; Chapman and Hall, 2004.
30. Toutenburg, H. Statistical Analysis of designed experiments. 2nd edition, Springer, 2002.
31. Woolson, R. F. and Clarke, W. R. Analysis of categorical incomplete longitudinal data; J. Roy. Statist. Soc. Series A; 147, 1984, pp. 87-99.