

EFFECT OF MISSING VALUES ON INTRACLAS KAPPA STATISTIC FOR RATERS AGREEMENT MEASUREMENT

A. O. Adejumo

Department of Statistics, University of Ilorin, Ilorin, Nigeria.

ao123adejumo@yahoo.co.uk

ABSTRACT: In any experiment involving measurement, counting, treatments or diagnosis, irrespective of the level of monitoring, attention, interest and care, there is the likelihood for some elements of missing observations. Intraclass Kappa was introduced as an improvement on Cohen kappa statistic. It is also mainly for the measurement of overall level of agreement that exists between two raters. Subjects are being classified into categories by raters, interviewers or observers in all life or social science researches. In this research work, the effects of these missing observations were examined with different missing patterns from the square table of cross-classified ratings of raters on the Intraclass kappa statistic. A conventional missing at random (CMAR) was assumed, which is a special case of missing at random (MAR) mechanism criteria and it was observed that missingness improved or worsened the strength of agreement as the missing percentage increases depending on the missing pattern under consideration.

Keywords: Agreement, intraclass, conventional missing at random (CMAR), raters.

1.0 INTRODUCTION

Square contingency table can be used to display joint ratings of two raters. It is established in this paper that during the process of collecting and classifying these ratings on the subjects into classes of categories, there are possibilities of missing values. Categorical data are data that the response variable is classified into either nominal or ordinal categories. For nominal data, which is our focus area in this paper, a large numbers of estimation and testing procedures like the Cohen Kappa (Cohen, 1960), the intraclass Kappa (Bleck and Kraemer, 1988), and many other improved methods on kappa statistic, (Shoukri, 2004). On the side of ordinal data, most of the medical diagnoses data often involve responses taken on an ordinal scale; many of which are very subjective. As it has been pointed out by some authors on ordinal data, an intermediate category will often be subjective to more misclassification than an extreme category because there are two directions in which to err away from the extremes. Therefore a modified kappa statistic called weighted Kappa (Cohen, 1968) which accounts for severity of discordance or size of discrepancy is better suited for ordinal data.

Nonresponse in a sample survey or drop-out in clinical trials may be two of many examples of missing data one could imagine. Missing value problem in statistical analysis has been in existence over decades. For instance, suppose that in a medical screening study, where a pathologist may examine a specific number of patients for the level of seriousness of an ailment and later another pathologist is asked to examine the same set of patients. Missing of values are likely to occur in this kind of study because not all these patients will be available for the second test.

This paper examines the behavior of Intraclass kappa statistic on raters agreement measure where there are missing observations with different patterns. In the next two sections we present missing value and description of the intraclass kappa statistic with and without missing observation. Various empirical examples are provided in section 4 and results and conclusion are in section 5.

2.0 The concepts of missing values

Some authors have recommended various standard statistical methods for handling missing data which are presented in form of a rectangular array, that is, to analyze a matrix

$$X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & * & \dots & x_{2p} \\ x_{31} & x_{32} & \dots & * \\ \vdots & \vdots & \dots & \vdots \\ x_{n1} & * & \dots & x_{np} \end{pmatrix} \dots \dots \dots 1$$

The rows of this data matrix can be modelled as independent, identically distributed (iid) draws from some multivariate probability distribution. The missing values, denoted by * may occur in any pattern. The probability density function of the complete data may be written as

$$P(X | \theta) = \prod_{i=1}^n f(x_i | \theta), \dots \dots \dots 2$$

where f is the density or probability function for a single row, and θ is a vector of unknown parameters. Under missing value analysis, data on any scales can be observed.

In categorical data analysis, visualizing the structure of the data set with respect to the missing values may be the first way to get an impression of how to handle the problem. These patterns may give an impression to of the extent the data are missing. If X is assumed to be missing for large values of y , the values can be ordered and a missing data pattern may describe this behavior. However, this technique may be swamped with a high level of dependencies. A way to overcome this defect consists of defining the so called missing values pattern (Little and Rubins, 1987; Rubin, 1976 and Toutenburg *et al.*, 2002) for details missing values pattern.

2.1: Missing values mechanisms

Let the observed part of X be represented by X_{obs} , and the missing counterpart by X_{mis} . The next focus is whether the missing data mechanism can be *ignored* or not. It is possible to make an assumption that the mechanism is ignorable or should be included in the statistical model. Including the missing data mechanism implies including the distribution of an indicator variable R indicating if a component of the data matrix $Z = (Z_{obs}, Z_{mis})$ is observed or missing. The random variable R indicating the missingness within the data matrix Z is defined as

$$r_{ij} = \begin{cases} 1, & \text{if } z_{ij} \text{ observed} \\ 0, & \text{if } z_{ij} \text{ missing} \end{cases} \dots \dots \dots 3$$

for all $i=1, 2, \dots, n, j=1, 2, \dots, p+1$

The ignorability criteria of the missing data mechanism depends on whether statistical inference is based on the density $f(R, Z_{obs} | \theta, \phi)$ or on the simpler density $f(Z_{obs} | \theta)$ which is ignoring the missing mechanism, where θ is the parameter of the density of Z_{obs}, Z_{mis} and ϕ is the unknown parameter of the missing mechanism. Therefore, the classification of missing data mechanisms is thus based on the density $f(R | Z_{obs}, Z_{mis}, \phi)$ (Schafer, 1997; Toutenburg and Nittner, 2002; Toutenburg *et al.*, 2002)

In this research work, a conventional missing at random (CMAR) missing criteria mechanism is assumed for the missingness in such region of the square table (Adejumo *et al.*, 2004). These special parts may be observations in cells along the diagonal, lower diagonal, upper diagonal, off diagonal of the square table and so on. CMAR is assumed because of the special attention given to such specific region of the table under consideration before assuming further missing at random (MAR) in such region rather than considering the complete parts of the table together.

3.2: Missing in the ratings of raters

Missing observations can also be observed in the raw results for some of the subjects involved in the experiment or in the trials. For example, consider an experiment on the interpretation of Sputum cytology slides for lung by two observers; some of these slides may not be available for the second observer due to reasons like fault from the first observer or other reason. In addition to this if two neurologists are asked to examine certain number of patients for the classification of multiple sclerosis, the first neurologist might have successfully classified patients for multiple sclerosis, but before getting to the next neurologist, some of the patients might have died or not available again. These two cases are referred to as missing values in such experiments (see Tables 7 and 8 for the complete tables of such experiments). To this effect, Table 1 describes the pattern of how missing observations can occur in the raw ratings of two raters on some sets of subjects.

Table 1: Missing pattern for ratings of two raters

Subject	Rater 1		Rater 2		status
	resp	rating	resp	rating	
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

From the Table the rating column depends on the categorical scales agreed upon by the raters and also the combinations of these two ratings determine the form of the cross-classified square contingency table for agreement. A square contingency table can be used to display the joint ratings of these raters. In this case the categories used by the two raters must be the same. The $\{resp\}$ stands for the response status for i th subject with j th rater which we have defined in equation 3 as the random variable R indicating the missingness within the data matrix Z , defined in this case for two raters ($j=1,2$) as

$$r_{ij} = \begin{cases} 1, & \text{if } z_{ij} \text{ observed} \\ 0, & \text{if } z_{ij} \text{ missing} \end{cases} \dots\dots\dots 4$$

for all $i=1,2,\dots,n$, $j=1,2,\dots,p+1$,
such that the matrix for R base on Table 1 can be of the form

$$R = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 1 \\ 1 & 1 \\ 1 & 0 \\ 1 & 1 \\ 0 & 0 \\ . & . \\ . & . \\ . & . \\ 1 & 1 \end{pmatrix} \dots\dots\dots 5$$

Also each subject combined response status S_i is defined as

$$S_i = \begin{cases} \text{obs, if } r_{i1} = r_{i2} = 1 \text{ as observed} \\ \text{mis, if } (r_{i1} = 0 \text{ and } r_{i2} = 1 \text{ or } r_{i1} = 1 \text{ and } r_{i2} = 0) \text{ as missing} \\ \text{mis*}, \text{ if } r_{i1} = r_{i2} = 0 \text{ as totally missing } \forall i = 1, 2, \dots, n. \end{cases}$$

2.3: Missing pattern in the ratings of raters

The missing pattern depends on the nature of missingness as we have presented in Table 1. Take for instance if the categorical scale for the raters is numbered 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, then if we assume that this missing response is also 1, this implies that the missing is in the diagonal cell, but if it is assumed that it is not 1, that means it will be any of numbers 2 to 5, then the missing of that particular subject is in the off-diagonal cells. Generally, if there are many responses, says 'a,b,c,...', that are missing, all may fall along the diagonal or in the off-diagonal cells and sometimes there may be the combination of the two missing patterns, that is, off and along the diagonal combined. Examples of these tables with different patterns of missing for 2×2 contingency tables are as given in Tables 3 to 5.

Table 2: Complete cross-classified table for two raters

Category Rater 1	Rater 2				Total
	1	2	...	l	
1	n_{11}	n_{12}	...	n_{1l}	n_{1+}
2	n_{21}	n_{22}	...	n_{2l}	n_{2+}
...					
...					
l	n_{l1}	n_{l2}	...	n_{ll}	n_{l+}
total	n_{+1}	n_{+2}	...	n_{+l}	n_{++}

Now assume that we have a 2×2 contingency table for the ratings of two raters in Tables 3 to 5.

Table 3: 2×2 table for two raters with missing along the diagonal

Category rater 1	rater 2		Total	Missing
	1	2		
1	n_{11}^*	n_{12}	n_{1+}	a_1
2	n_{21}	n_{22}^*	n_{2+}	a_2
total	n_{+1}	n_{+2}	n_{++}	
missing	a_1	a_2		A

Table 4: 2×2 table for two raters with missing off the diagonal

Category rater 1	rater 2		Total	Missing
	1	2		
1	n_{11}	n_{12}^*	n_{1+}	b_1
2	n_{21}^*	n_{22}	n_{2+}	b_2
total	n_{+1}	n_{+2}	n_{++}	
missing	b_2	b_1		b

Table 5: 2 x 2 table for two raters with missing in both along and off the diagonal

Category rater 1	rater 2		total	Missing
	1	2		
1	n^{***}_{11}	n^{***}_{12}	n_{1+}	a_1+b_1
2	n^{***}_{21}	n^{***}_{22}	n_{2+}	a_2+b_2
total	n_{+1}	n_{+2}	n_{++}	
missing	$a'_1+b'_2$	$a'_2+b'_1$		$a'+b'$

3.0: Intraclass kappa

Intraclass kappa was defined for data consisting of blind dichotomous ratings on each of n subjects by two fixed raters. It is assumed that the ratings on a subject are interchangeable; that is in the population of subjects, the two ratings for each subject have a distribution that is invariant under permutations of the raters to ensure that there is no rater bias Bloch and Kraemer (1989). Barnhart and Williamson (2002) and others. Let X_{ij} denote the rating for the i th subject by the j th rater, $i=1,2,\dots,n$, $j=1,2$. And for

each subject i , let $\pi_i = P(X_{ij} = 1)$, be the probability that the rating is a success. Over the

population of subjects, let $E(\hat{\pi}_i) = \Pi$, $\Pi' = (1 - \Pi)$ and $\text{var}(\hat{\pi}_i) = \sigma_{\pi}^2$. The intraclass kappa as defined by Bloch and Kraemer (1989) is then

$$k_I = \frac{\sigma_{\pi}^2}{\Pi\Pi'} \dots\dots\dots 7$$

If Π - the log-likelihood function is given by

$$L(\Pi, k_I | n_{11}, n_{12}, n_{21}, n_{22}) = n_{11} \log(\pi^2 + k_I \Pi \Pi') + (n_{12} + n_{21}) \log\{\Pi \Pi' (1 - k_I)\} + n_{22} \log(\pi'^2 + k_I \Pi \Pi').$$

The maximum likelihood estimators $\hat{\pi}$ and $\hat{k}_I$ for Π and k_I are obtained as

$$\hat{\pi} = \frac{2n_{11} + n_{12} + n_{21}}{2n} \dots\dots\dots 8$$

and

$$\hat{k}_I = \frac{4(n_{11}n_{22} - n_{12}n_{21}) - (n_{12} - n_{21})^2}{(2n_{11} + n_{12} + n_{21})(2n_{22} + n_{12} + n_{21})} \dots\dots\dots 9$$

with the estimated standard error for $\hat{k}_I$ given as

$$SE(\hat{k}_I) = \left\{ \frac{1 - \hat{k}_I}{n} \left[(1 - \hat{k}_I)(1 - 2\hat{k}_I) + \frac{\hat{k}_I(2 - \hat{k}_I)}{2\hat{\pi}(1 - \hat{\pi})} \right] \right\}^{1/2} \dots\dots\dots 10$$

This is only restricted to a 2 x 2 table.

Barnhart and Williamson (2002) considered intraclass kappa for measuring agreement between two readings for a categorical response with 1 categories if the two readings are replicated measurements. It assumes no bias because the probability of a positive rating is the same for the two readings due to replication, and it is given as

$$\hat{k}_m = \frac{\sum_{i=1}^I \pi_{ii} - \sum_{i=1}^I ((\pi_{i+} + \pi_{+i})/2)^2}{1 - \sum_{i=1}^I ((\pi_{i+} + \pi_{+i})/2)^2} \quad (11)$$

with variance for the estimated value as

$$\hat{\text{var}}_o(\hat{k}_m) = \frac{\pi_{\text{exp}} + (\pi_{\text{exp}})^2 - \pi_{\text{exp}} (\sum_{i=1}^I ((\pi_{i+} + \pi_{+i})/2)^2)}{n(1 - \pi_{\text{exp}})^2} \quad (12)$$

where

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

We also assumed that

$$\frac{\hat{k}_m}{\sqrt{\hat{\text{var}}_o(\hat{k}_m)}} \quad (13)$$

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

$$100(1 - \alpha)\% = \hat{k}_m \pm z_{1-\frac{\alpha}{2}} SE(\hat{k}_m) \text{ can be obtained for } \hat{k}_m.$$

Also Cohen's kappa was introduced for measuring nominal scale chance-corrected agreement by Cohen (1960). Cohen's kappa is 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 estimated as

$$\pi_o = \sum_{i=1}^I \pi_{ii} \quad (14)$$

and under the baseline constraints of complete independence between ratings by the two observers, which is the expected agreement proportion estimated as

$$\pi_e = \sum_{i=1}^I \pi_{i+} \pi_{+i} \quad (15)$$

The Kappa statistic can now be estimated by

$$\hat{k}_c = \frac{\hat{\pi}_o - \hat{\pi}_e}{1 - \hat{\pi}_e} \quad (16)$$

where $\hat{\pi}_o$ and $\hat{\pi}_e$ are as defined above.

3.1: Intraclass kappa statistic with missing observations

In a given 1×1 contingency table as in Table 2, it is assumed that the missingness mechanism criteria is a conventional missing at random (CMAR) with special missing patterns in the cells as given in Tables 3, 4, and 5.

Now the joint probability distribution table for raters 1 and 2 for a 2×2 table will be as in Table 6.

Table 6: 2×2 probability table

Category rater 1	rater 2		Total
	1	2	
1	π_{11}	π_{12}	π_{1+}
2	π_{21}	π_{22}	π_{2+}
Total	π_{+1}	π_{+2}	π_{++}

Intraclass kappa statistic was given as

$$k_{in} = \frac{\sum_{i=1}^I \pi_{ii} - \sum_{i=1}^I ((\pi_{i+} + \pi_{+i})/2)^2}{1 - \sum_{i=1}^I ((\pi_{i+} + \pi_{+i})/2)^2}$$

This can also be expressed in terms of counts data rather than proportion as follows:

$$k_{in} = \frac{\sum_{i=1}^I \frac{n_{ii}}{n_{++}} - \sum_{i=1}^I ((\frac{n_{i+}}{n_{++}} + \frac{n_{+i}}{n_{++}})/2)^2}{1 - \sum_{i=1}^I ((\frac{n_{i+}}{n_{++}} + \frac{n_{+i}}{n_{++}})/2)^2}$$

So, if 'a' missing observations

$$k_{in} = \left(\frac{\frac{n_{11}}{n_{++}} + \frac{n_{22}}{n_{++}} - ((\frac{n_{1+}}{n_{++}} + \frac{n_{2+}}{n_{++}} + \frac{n_{+1}}{n_{++}} + \frac{n_{+2}}{n_{++}})/2)^2}{1 - ((\frac{n_{1+}}{n_{++}} + \frac{n_{2+}}{n_{++}} + \frac{n_{+1}}{n_{++}} + \frac{n_{+2}}{n_{++}})/2)^2} \right)$$

Now, it is assumed that the missing mechanism is a conventional missing at random (CMAR) with special missing patterns as in Tables 3 and 4 in some of the cells of the table. However, for simplicity in checking the effect of this missingness on the statistic under consideration, we substitute n_{11} with $(n_{11} - a_1)$, n_{1+} with $(n_{1+} - a_1)$, n_{+1} with $(n_{+1} - a_1)$ and n_{++} with $(n_{++} - a_1)$ for 'a' missing observations along the diagonal as in Table 3, such that,

$$\begin{aligned} \pi_{11}^* &= \frac{(n_{11} - a_1)}{(n_{++} - a_1)} \neq \frac{n_{11}}{n_{++}} \quad \text{if } a_1 \neq 0, \\ \pi_{1+}^* &= \frac{(n_{11} - a_1) + n_{12}}{(n_{++} - a_1)} \neq \frac{(n_{11} + n_{12})}{n_{++}} \quad \text{if } a_1 \neq 0, \\ \pi_{+1}^* &= \frac{(n_{11} - a_1) + n_{21}}{(n_{++} - a_1)} \neq \frac{(n_{11} + n_{21})}{n_{++}} \quad \text{if } a_1 \neq 0, \end{aligned}$$

And for 'b' missing observations in the off diagonal of the table as in Table 4 by substituting n_{12} with $(n_{12} - b_1)$, n_{1+} with $(n_{1+} - b_1)$, n_{+2} with $(n_{+2} - b_1)$ and n_{++} with $(n_{++} - b_1)$ such that

$$\begin{aligned} \pi_{12}^{**} &= \frac{(n_{12} - b_1)}{(n_{++} - b_1)} \neq \frac{n_{12}}{n_{++}} \quad \text{if } b_1 \neq 0, \\ \pi_{1+}^{**} &= \frac{n_{11} + (n_{12} - b_1)}{(n_{++} - b_1)} \neq \frac{(n_{11} + n_{12})}{n_{++}} \quad \text{if } b_1 \neq 0, \\ \pi_{+2}^{**} &= \frac{(n_{12} - b_1) + n_{22}}{(n_{++} - b_1)} \neq \frac{(n_{12} + n_{22})}{n_{++}} \quad \text{if } b_1 \neq 0, \end{aligned}$$

Also for 'a' and 'b' respectively being the combinations of missing observations along and off the diagonal of the table as in Table 5, that is, missing from the entire parts of the table, we substitute n_{11} with $(n_{11} - a_1)$, n_{12} with $(n_{12} - b_1)$, n_{1+} with $(n_{1+} - (a_1 + b_1))$, n_{+1} with $(n_{+1} - a_1)$, n_{+2} with $(n_{+2} - b_1)$ and n_{++} with $(n_{++} - (a_1 + b_1))$, such that,

$$\begin{aligned}\pi_{11} &= \frac{(n_{11} - a_1)}{(n_{++} - (a_1 + b_1))} \neq \frac{n_{11}}{n_{++}} \quad \text{if } a_1, b_1 \neq 0, \\ \pi_{12} &= \frac{(n_{12} - b_1)}{(n_{++} - (a_1 + b_1))} \neq \frac{n_{12}}{n_{++}} \quad \text{if } a_1, b_1 \neq 0, \\ \pi_{1+} &= \frac{(n_{11} - a_1) + (n_{12} - b_1)}{(n_{++} - (a_1 + b_1))} \neq \frac{n_{11} + n_{12}}{n_{++}} \quad \text{if } a_1, b_1 \neq 0, \\ \pi_{+1} &= \frac{(n_{11} - a_1) + n_{21}}{(n_{++} - (a_1 + b_1))} \neq \frac{n_{11} + n_{21}}{n_{++}} \quad \text{if } a_1, b_1 \neq 0, \\ \pi_{+2} &= \frac{(n_{12} - b_1) + n_{22}}{(n_{++} - (a_1 + b_1))} \neq \frac{n_{12} + n_{22}}{n_{++}} \quad \text{if } a_1, b_1 \neq 0,\end{aligned}$$

Case 1

Mathematically the effect of missing counts from the diagonal cells on the Intraclass Kappa statistic is shown in Table 3 as follows

$$\begin{aligned}k_m &= \left(\frac{\pi_{11}^m + \pi_{22}^m - ((\pi_{1+}^m + \pi_{2+}^m + \pi_{+1}^m + \pi_{+2}^m)/2)^2}{1 - ((\pi_{1+}^m + \pi_{2+}^m + \pi_{+1}^m + \pi_{+2}^m)/2)^2} \right) \\ &= \left(\frac{\frac{n_{11}^m}{n_{++}^m} + \frac{n_{22}^m}{n_{++}^m} - ((\frac{n_{1+}^m}{n_{++}^m} + \frac{n_{2+}^m}{n_{++}^m} + \frac{n_{+1}^m}{n_{++}^m} + \frac{n_{+2}^m}{n_{++}^m})/2)^2}{1 - ((\frac{n_{1+}^m}{n_{++}^m} + \frac{n_{2+}^m}{n_{++}^m} + \frac{n_{+1}^m}{n_{++}^m} + \frac{n_{+2}^m}{n_{++}^m})/2)^2} \right) \\ &= \left(\frac{\frac{n_{11} - a_1}{(n_{++} - a)} + \frac{n_{22} - a_2}{(n_{++} - a)} - ((\frac{n_{1+} - a_1}{(n_{++} - a)} + \frac{n_{2+} - a_2}{(n_{++} - a)} + \frac{n_{+1} - a_1}{(n_{++} - a)} + \frac{n_{+2} - a_2}{(n_{++} - a)})/2)^2}{1 - ((\frac{n_{1+} - a_1}{(n_{++} - a)} + \frac{n_{2+} - a_2}{(n_{++} - a)} + \frac{n_{+1} - a_1}{(n_{++} - a)} + \frac{n_{+2} - a_2}{(n_{++} - a)})/2)^2} \right) \\ &= \left(\frac{\frac{(n_{11} - a_1) + (n_{22} - a_2)}{(n_{++} - a)} - ((\frac{n_{1+} - a_1}{(n_{++} - a)} + \frac{n_{2+} - a_2}{(n_{++} - a)} + \frac{n_{+1} - a_1}{(n_{++} - a)} + \frac{n_{+2} - a_2}{(n_{++} - a)})/2)^2}{1 - ((\frac{n_{1+} - a_1}{(n_{++} - a)} + \frac{n_{2+} - a_2}{(n_{++} - a)} + \frac{n_{+1} - a_1}{(n_{++} - a)} + \frac{n_{+2} - a_2}{(n_{++} - a)})/2)^2} \right) \\ &= \left(\frac{\frac{(n_{11} - a_1) + (n_{22} - a_2)}{(n_{++} - a)} - ((\frac{n_{1+} - a_1}{(n_{++} - a)} + \frac{n_{2+} - a_2}{(n_{++} - a)} + \frac{n_{+1} - a_1}{(n_{++} - a)} + \frac{n_{+2} - a_2}{(n_{++} - a)})/2)^2}{1 - ((\frac{n_{1+} - a_1}{(n_{++} - a)} + \frac{n_{2+} - a_2}{(n_{++} - a)} + \frac{n_{+1} - a_1}{(n_{++} - a)} + \frac{n_{+2} - a_2}{(n_{++} - a)})/2)^2} \right) \\ &= \left(\frac{\frac{(n_{11} - a_1) + (n_{22} - a_2)}{(n_{++} - a)} - ((\frac{n_{1+} - a_1}{(n_{++} - a)} + \frac{n_{2+} - a_2}{(n_{++} - a)} + \frac{n_{+1} - a_1}{(n_{++} - a)} + \frac{n_{+2} - a_2}{(n_{++} - a)})/2)^2}{1 - ((\frac{n_{1+} - a_1}{(n_{++} - a)} + \frac{n_{2+} - a_2}{(n_{++} - a)} + \frac{n_{+1} - a_1}{(n_{++} - a)} + \frac{n_{+2} - a_2}{(n_{++} - a)})/2)^2} \right) \\ &= \left(\frac{4\{(n_{++} - a)((n_{11} - a_1) + (n_{22} - a_2))\} - \{(n_{1+} - a_1) + (n_{2+} - a_2) + (n_{+1} - a_1) + (n_{+2} - a_2)\}^2}{4\{(n_{++} - a)\}^2 - \{(n_{1+} - a_1) + (n_{2+} - a_2) + (n_{+1} - a_1) + (n_{+2} - a_2)\}^2} \right)\end{aligned}$$

$$\neq \left(\frac{4\{n_{++}(n_{11} + n_{22})\} - \{n_{1+} + n_{2+} + n_{+1} + n_{+2}\}^2}{4n_{++}^2 - \{n_{1+} + n_{2+} + n_{+1} + n_{+2}\}^2} \right) \quad \text{if } a_1, a_2 \neq 0.$$

Case 2

Also the effect when there are missing counts from the off diagonal cells for the Intraclass Kappa statistic, as in Table 4 is given by

$$k_m = \left(\frac{\pi_{11} + \pi_{22} - ((\pi_{1+}^m + \pi_{2+}^m + \pi_{+1}^m + \pi_{+2}^m)/2)^2}{1 - ((\pi_{1+}^m + \pi_{2+}^m + \pi_{+1}^m + \pi_{+2}^m)/2)^2} \right) \\ \left(\frac{\frac{n_{11}}{n_{++}} + \frac{n_{22}}{n_{++}} - ((\frac{n_{1+}^m}{n_{++}} + \frac{n_{2+}^m}{n_{++}} + \frac{n_{+1}^m}{n_{++}} + \frac{n_{+2}^m}{n_{++}})/2)^2}{1 - ((\frac{n_{1+}^m}{n_{++}} + \frac{n_{2+}^m}{n_{++}} + \frac{n_{+1}^m}{n_{++}} + \frac{n_{+2}^m}{n_{++}})/2)^2} \right) \\ \left(\frac{\frac{n_{11}}{(n_{++}-b)} + \frac{n_{22}}{(n_{++}-b)} - ((\frac{n_{1+}-b_1}{(n_{++}-b)} + \frac{n_{2+}-b_2}{(n_{++}-b)} + \frac{n_{+1}-b_2}{(n_{++}-b)} + \frac{n_{+2}-b_1}{(n_{++}-b)})/2)^2}{1 - ((\frac{n_{1+}-b_1}{(n_{++}-b)} + \frac{n_{2+}-b_2}{(n_{++}-b)} + \frac{n_{+1}-b_2}{(n_{++}-b)} + \frac{n_{+2}-b_1}{(n_{++}-b)})/2)^2} \right) \\ \left(\frac{\frac{(n_{11} + n_{22})}{(n_{++}-b)} - ((\frac{(n_{1+}-b_1) + (n_{2+}-b_2) + (n_{+1}-b_2) + (n_{+2}-b_1)}{(n_{++}-b)})/2)^2}{1 - ((\frac{(n_{1+}-b_1) + (n_{2+}-b_2) + (n_{+1}-b_2) + (n_{+2}-b_1)}{(n_{++}-b)})/2)^2} \right) \\ \left(\frac{\frac{(n_{11} + n_{22})}{(n_{++}-b)} - ((\frac{(n_{1+}-b_1) + (n_{2+}-b_2) + (n_{+1}-b_2) + (n_{+2}-b_1)}{2(n_{++}-b)})^2}{1 - ((\frac{(n_{1+}-b_1) + (n_{2+}-b_2) + (n_{+1}-b_2) + (n_{+2}-b_1)}{2(n_{++}-b)})^2} \right) \\ \left(\frac{4\{(n_{++}-b)(n_{11} + n_{22})\} - \{(n_{1+}-b_1) + (n_{2+}-b_2) + (n_{+1}-b_2) + (n_{+2}-b_1)\}^2}{4\{(n_{++}-b)\}^2 - \{(n_{1+}-b_1) + (n_{2+}-b_2) + (n_{+1}-b_2) + (n_{+2}-b_1)\}^2} \right) \\ \left(\frac{4\{n_{++}(n_{11} + n_{22})\} - \{n_{1+} + n_{2+} + n_{+1} + n_{+2}\}^2}{4n_{++}^2 - \{n_{1+} + n_{2+} + n_{+1} + n_{+2}\}^2} \right) \quad \text{if } b_1, b_2 \neq 0.$$

Case 3

For the case 3, when there are missing observations from the diagonal as well as off diagonal cells, the effect is mathematically on intraclass Kappa, as in Table 5 as follows.

$$k_m = \left(\frac{\pi_{11} + \pi_{22} - ((\pi_{1+}^m + \pi_{2+}^m + \pi_{+1}^m + \pi_{+2}^m)/2)^2}{1 - ((\pi_{1+}^m + \pi_{2+}^m + \pi_{+1}^m + \pi_{+2}^m)/2)^2} \right) \\ \left(\frac{\frac{n_{11}}{n_{++}} + \frac{n_{22}}{n_{++}} - ((\frac{n_{1+}^m}{n_{++}} + \frac{n_{2+}^m}{n_{++}} + \frac{n_{+1}^m}{n_{++}} + \frac{n_{+2}^m}{n_{++}})/2)^2}{1 - ((\frac{n_{1+}^m}{n_{++}} + \frac{n_{2+}^m}{n_{++}} + \frac{n_{+1}^m}{n_{++}} + \frac{n_{+2}^m}{n_{++}})/2)^2} \right)$$

$$\begin{aligned}
 &= \left[\frac{\frac{n_{11} - a_1}{(n_{++} - A)} + \frac{n_{22} - a_2}{(n_{++} - A)} - \left(\frac{n_{1+} - A_{11}}{(n_{++} - A)} + \frac{n_{2+} - A_{22}}{(n_{++} - A)} + \frac{n_{+1} - A_{11}}{(n_{++} - A)} + \frac{n_{+2} - A_{21}}{(n_{++} - A)} \right) / 2}{1 - \left(\frac{n_{1+} - A_{11}}{(n_{++} - A)} + \frac{n_{2+} - A_{22}}{(n_{++} - A)} + \frac{n_{+1} - A_{12}}{(n_{++} - A)} + \frac{n_{+2} - A_{21}}{(n_{++} - A)} \right) / 2} \right] \\
 &= \left[\frac{\frac{(n_{11} - a_1) + (n_{22} - a_2)}{(n_{++} - (a + b))} - \left(\frac{G}{(n_{++} - (a + b))} \right) / 2}{1 - \left(\frac{G}{(n_{++} - (a + b))} \right) / 2} \right] \\
 &= \left[\frac{\frac{(n_{11} - a_1) + (n_{22} - a_2)}{(n_{++} - (a + b))} - \left(\frac{G}{2(n_{++} - (a + b))} \right)^2}{1 - \left(\frac{(n_{1+} - (a_1 + b_1)) + G}{2(n_{++} - (a + b))} \right)^2} \right] \\
 &= \left[\frac{4\{(n_{++} - (a + b))(n_{11} - a_1) + (n_{22} - a_2)\} - \{G\}^2}{4\{(n_{++} - (a + b))\}^2 - \{G\}^2} \right] \\
 &= \left[\frac{4\{n_{++}(n_{11} + n_{22})\} - \{n_{1+} + n_{2+} + n_{+1} + n_{+2}\}^2}{4n_{++}^2 - \{n_{1+} + n_{2+} + n_{+1} + n_{+2}\}^2} \right] \quad \text{If } a_1, a_2, b_1, b_2 \neq 0.
 \end{aligned}$$

where

$$A_{11} = (a_1 + b_1), \quad A_{22} = (a_2 + b_2),$$

$$A_{12} = (a_1 + b_2), \quad A_{21} = (a_2 + b_1),$$

$$A = (a + b),$$

$$G = (n_{1+} - (a_1 + b_1)) + (n_{2+} - (a_2 + b_2)) + (n_{+1} - (a_1 + b_2)) + (n_{+2} - (a_2 + b_1)).$$

3.2: Case-by-case formulations

The effects of the missing observation with different missing patterns for a given table of ratings of raters on n_{++} subjects on intraclass kappa statistic is examined by considering the following cases with certain percentages missing.

On each of the following cases we consider 0%, 5% and 10% missing observations of the total counts n_{++} from their respective region.

- Case 1:- When there are certain percentages missing along the diagonal.
- Case 2:- When there are certain percentages missing off the diagonal.
- Case 3:- When there are certain percentages missing in the entire table (along and off diagonal combined).

For each of the cases we obtain the estimate for the statistic under consideration, with its standard error.

Z-value based on equation 13. We create a bound for this statistic $\hat{k}_m$ by obtaining the 95% confidence interval of size

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

4.0: Empirical examples

4.1: Example 1. Consider the data on cross-classification of two independent interpretations of Sputum cytology slides for lung by two observers taken from Becker (1990) based on the following categorical scale: 1= Negative, 2= Ambiguous cells, 3= suspect, 4= positive, and 5= technically unsatisfactory.

Table 7: Cross-classification of two independent interpretations of Sputum cytology slides for lung of 100 patients

Category First Observer	Second observer					Total
	1	2	3	4	5	
1	26	19	1	0	7	53
2	2	11	5	3	4	25
3	0	1	6	6	0	13
4	0	0	0	4	1	5
5	1	1	0	0	2	4
Total	29	32	12	13	14	100

4.2: Example 2

Consider also another data arising from the diagnosis of multiple sclerosis reported by two neurologists on 69 patients of New Orleans, using the following categorical scales for diagnostic classes: 1=certain multiple sclerosis, 2=probable multiple sclerosis, 3=possible multiple sclerosis, 4=doubtful, unlikely, or definitely not multiple sclerosis (Hann *et al.*, 1999).

Table 8: Cross-classification table of multiple sclerosis assessment for 69 patients

Category Neurologist 1	Neurologist 2				Total
	1	2	3	4	
1	5	3	0	0	8
2	3	11	4	0	18
3	2	13	3	4	22
4	1	2	4	14	21
Total	11	29	11	18	69

5.0: RESULTS AND CONCLUSION

5.1: Results

From the tables in the Appendix, under case 1, it is observed that the higher the missing value percentages the worse the strength of agreement become. But for cases 2 and 3 the higher the percentages of missing observations the better the strength of agreement between the two raters become.

5.2: Conclusions

It is observed that increment in the percentages of missingness along the diagonal worsen the strength of the agreement between the two raters. However, increments in the percentages of missingness in the off diagonal part of the table of the ratings for two raters improve the strength of the agreement between them. This implies that missingness from both upper and lower diagonal cells of the square table of the raters ratings over boosts the interclass kappa. To this effect, modelling the structure of agreement will be a better way of monitoring the effects of missing observation on the agreement measure.

REFERENCES

- Adejumo, A.O., Heumann, C. and Toutenburg, H. (2004). On the effects of missing values on comparison of Symmetry model and Quasi-symmetry model for raters agreement method. SFB386-Discussion Paper, Ludwig-Maximilians- University, Muenchen.
- Barnhart, H. X., and Williamson, J. M. (2002). Weighted Least squares approach for computing correlated kappa. *Biometrics*, 58, 1012-1019.
- Basu, S., Basu, A. and Raychaudhuri, A. (1999). Measuring agreement between two raters for ordinal response: a model based approach. *The Statistician*, 48(3), 339-348.
- Becker, M. P. (1990). Quasisymmetric models for the analysis of square contingency tables. *J. Roy Statist. Soc. Series B*, 52, 369-378.
- Bloch, D. A. and Kraemer, H. C. (1989). 2 x 2 kappa coefficients: Measures of agreement in association. *Biometrics*, 45, 269-287.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Edu. and Psych. Meas.*, 20, 37-46.
- Cohen, J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement on partial credit. *Psych.Bull.*, 70, 213-220.
- Little, R.J.A. and Rubin, D.B. (1987). *Statistical analysis with missing Data*. New York, Wiley.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63, 581-592.
- Schafer J. L. (1997). *Analysis of incomplete multivariate data*. Chapman and Hall.
- Shoukri, M. M. (2004). *Measures of interobserver agreement*. Chapman and Hall.
- Toutenburg, H., Heumann, C., Nittner, T. and Scheid, S. (2002). Parametric and Nonparametric Regression with Missing X'S- A review. *J. Iranian Statist. Socie.*, 1 (1,2), 79-109.
- Toutenburg, H. and Nittner, T. (2002). Linear regression models with incomplete categorical covariates. *Computational Statist.*, 17 (2), 215-232.

APPENDIX

Table 9: Intraclass kappa statistic estimates for Cross-classification of two independent interpretations of Sputum cytology slides for lung of 100 patients

Cases	%missing	Intraclass Kappa ($\hat{k}_m$)	Stand. error	z-value	95% CI for $\hat{k}_m$	
					Lower bound	Upper bound
complete	0%	0.2905	0.0832	3.4902	0.1274	0.4537
Case 1	5%	0.2581	0.0840	3.0731	0.0935	0.4227
	10%	0.2221	0.0848	2.6186	0.0559	0.3883
Case 2	5%	0.3263	0.0854	3.8193	0.1588	0.4940
	10%	0.3660	0.0878	4.1686	0.1939	0.5381
Case 3	5%	0.2970	0.0836	3.5523	0.1331	0.4608
	10%	0.2938	0.0862	3.4074	0.1248	0.4628

Table 10: Intraclass kappa statistic estimates for Cross-classification table of multiple sclerosis assessment for 69 patients

Cases	%missing	Intraclass		z-value	95% CI for $\hat{k}_m$	
		Kappa ($\hat{k}_m$)	Stand. error		Lower bound	Upper bound
complete	0%	0.2833	0.0971	2.9189	0.0931	0.4736
Case 1	5%	0.2437	0.0985	2.4746	0.0507	0.4367
	10%	0.1980	0.1004	1.9720	0.0012	0.3947
Case 2	5%	0.3238	0.1000	3.2386	0.1278	0.5198
	10%	0.3688	0.1035	3.5633	0.1659	0.5717
Case 3	5%	0.2840	0.0991	2.8653	0.0897	0.4782
	10%	0.2854	0.1010	2.8253	0.0874	0.4833