Hindawi Publishing Corporation
Journal of Probability and Statistics
Volume 2013, Article ID 325831, 9 pages
http://dx.doi.org/10.1155/2013/325831

Research Article

Weighted Kappas for 3×3 Tables

Matthijs J. Warrens

Unit of Methodology and Statistics, Institute of Psychology, Leiden University, P.O. Box 9555, 2300 RB Leiden, The Netherlands

Correspondence should be addressed to Matthijs J. Warrens; warrens@fsw.leidenuniv.nl

Received 10 April 2013; Accepted 26 May 2013

Academic Editor: Ricardas Zitikis

Weighted kappa is a widely used statistic for summarizing inter-rater agreement on a categorical scale. For rating scales with three categories, there are seven versions of weighted kappa. It is shown analytically how these weighted kappas are related. Several conditional equalities and inequalities between the weighted kappas are derived. The analytical analysis indicates that the weighted kappas are measuring the same thing but to a different extent. One cannot, therefore, use the same magnitude guidelines for all weighted kappas.

1. Introduction

In biomedical, behavioral, and engineering research, it is frequently required that a group of objects is rated on a categorical scale by two observers. Examples are the following: clinicians that classify the extent of disease in patients; pathologists that rate the severity of lesions from scans; and experts that classify production faults. Analysis of the agreement between the two observers can be used to assess the reliability of the rating system. High agreement would indicate consensus in the diagnosis and interchangeability of the observers. Various authors have proposed statistical methodology for analyzing agreement. For example, for modeling patterns of agreement, the loglinear models proposed in Tanner and Young [1] and Agresti [2, 3] can be used. However, in practice researchers are frequently only interested in a single number that quantifies the degree of agreement between the raters [4, 5]. Various statistics have been proposed in the literature [6, 7], but the most popular statistic for summarizing rater agreement is the weighted kappa introduced by Cohen [8].

Weighted kappa allows the use of weighting schemes to describe the closeness of agreement between categories. Each weighting scheme defines a different version or special case of weighted kappa. Different weighting schemes have been proposed for the various scale types. In this paper, we only consider scales of three categories. This is the smallest number of categories for which we can distinguish three types of categorical scales, namely, nominal scales, continuous-ordinal scales, and dichotomous-ordinal scales [9]. A dichotomous-ordinal scale contains a point of "absence" and two points of "presence", for example, no disability, moderate disability, or severe disability. A continuous-ordinal scale does not have a point of "absence". The scale can be described by three categories of "presence", for example, low, moderate, or high. Identity weights are used when the categories are nominal [10]. In this case, weighted kappa becomes the unweighted kappa introduced by Cohen [11], also known as Cohen's kappa. Linear weights [12, 13] or quadratic weights [14, 15] can be used when the categories are continuous ordinal. The modified linear weights introduced in Cicchetti [9] are suitable if the categories are dichotomous ordinal.

Although weighted kappa has been used in thousands of research applications [16], it has also been criticized by various authors [17–19]. Most of the criticism has focused on a particular version of weighted kappa, namely, Cohen's kappa for nominal categories. Weighted kappa and unweighted kappa correct for rater agreement due to chance alone using the marginal distributions. For example, in the context of latent class models, de Mast [18] and de Mast and van Wieringen [6] argued that the premise that chance measurements have the distribution defined by the marginal distributions cannot be defended. It is, therefore, difficult to interpret the value of Cohen's kappa, and it makes the question of how large or how small the value should be arbitrary. Using signal detection theory, Uebersax [19] showed that different agreement studies with different marginal distributions can produce the same value of Cohen's kappa. Again, this makes the value difficult to interpret. Alternative statistics for summarizing inter-rater agreement are discussed in, for example, de Mast [18] and Perreault and Leigh [20].

Although the choice for a specific version of weighted kappa usually depends on the type of categorical scale at hand, it frequently occurs that weighted kappas corresponding to different weighting schemes are applied to the same data. For example, Cohen's kappa for nominal scales [11] is also frequently applied when the categories are continuous ordinal. When different weighted kappas are applied to the same data, they usually produce different values [5, 21]. For understanding the behavior of weighted kappa and its dependence on the weighting scheme, it is useful to compare the different versions of weighted kappa analytically [21]. For example, if the agreement table is tridiagonal, then the value of the quadratically weighted kappa exceeds the value of the linearly weighted kappa, which, in turn, is higher than the value of unweighted kappa [22, 23]. An agreement table is tridiagonal if it has nonzero elements only on the main diagonal and on the two diagonals directly adjacent to the main diagonal. These analytic results explain orderings of the weighted kappas that are observed in practice.

In this paper, we consider scales that consist of three categories and compare the values of seven special cases of weighted kappa. There are several reasons why the case of three categories is an interesting topic of investigation. First of all, various scales that are used in practice consist of three categories only. Examples can be found in Anderson et al. [24] and Martin et al. [25]. Furthermore, the case of three categories is the smallest case where symmetrically weighted kappas in general have different values, since all weighted kappas with symmetric weighting schemes coincide with two categories. Finally, as it turns out, with three categories we may derive several strong analytic results, which do not generalize to the case of four or more categories. The seven weighted kappas belong to two parameter families. For each parameter family, it is shown that there are only two possible orderings of its members. Hence, despite the fact that the paper is limited to weighted kappas for three categories, we present various interesting and useful results that deepen our understanding of the application of weighted kappa.

The paper is organized as follows. In Section 2 we introduce notation and define four versions of weighted kappa. In Section 3, we introduce the three category reliabilities of a 3×3 agreement table as special cases of weighted kappa. The two parameter families are defined in Section 4. In Section 5, we present several results on inequalities between the seven weighted kappas. In Section 6, we consider the case that all special cases of weighted kappa coincide. Section 7 contains a discussion.

2. Weighted Kappas

Suppose that two raters, each, independently classify the same set of objects (individuals, observations) into the same set of three categories that are defined in advance. For a population of n objects, let π_{ij} for $i, j \in \{1, 2, 3\}$ denote the proportion classified into category i by the first observer and into category j by the second observer. Table 1 presents an abstract version of a 3×3 population agreement table of proportions. The marginal totals $\pi_{1+}, \pi_{2+}, \pi_{3+}$ and $\pi_{+1}, \pi_{+2}, \pi_{+3}$ indicate how often raters 1 and 2 used the categories 1, 2, and 3. Four examples of 3×3 agreement tables from the literature with frequencies are presented in Table 2. The marginal totals of the tables are in bold. For each table, the last column of Table 2 contains the corresponding estimates of seven weighted kappas. Between brackets behind each point estimate is the associated 95% confidence interval. Definitions of the weighted kappas are presented below.

TABLE 1: Notation for a 3 × 3 agreement table with proportions.

		Rater 2			
		1	2	3	Total
Rater 1	1	π_{11}	π_{12}	π_{13}	π_{1+}
	2	π_{21}	π_{22}	π_{23}	π_{2+}
	3	π_{31}	π_{32}	π_{33}	π_{3+}
	Total	π_{+1}	π_{+2}	π_{+3}	1

Recall that weighted kappa allows the use of weighting schemes to describe the closeness of agreement between categories. For each cell probability π_{ij}, we may specify a weight. A weighting scheme is called symmetric if for all i, j cell probabilities π_{ij} and π_{ji} are assigned the same weight. The weighting schemes can be formulated from either a similarity or a dissimilarity perspective. Definitions of weighted kappa in terms of similarity scaling can be found in Warrens [13, 22]. For notational convenience, we will define the weights in terms of dissimilarity scaling here. For the elements on the agreement diagonal, there is no disagreement. The diagonal elements are, therefore, assigned zero weight [8, page 215]. The other six weights are non-negative real numbers w_i for $i \in \{1, 2, \ldots, 6\}$. The inequality $w_i > 0$ indicates that there is some disagreement between the assignments by the raters. Categories that are more similar are assigned smaller weights. For example, ordinal scale categories that are one unit apart in the natural ordering are assigned smaller weights than categories that are more units apart.

Table 3 presents one general and seven specific weighting schemes from the literature. The identity weighting scheme for nominal categories was introduced in Cohen [11]. The top table in Table 2 is an example of a nominal scale. The quadratic weighting scheme for continuous-ordinal categories was introduced in Cohen [8]. The quadratically weighted kappa is the most popular version of weighted kappa [4, 5, 15]. The linear weighting scheme for continuous-ordinal categories was introduced in Cicchetti and Allison [29] and Cicchetti [30]. The second table in Table 2 is an example of a continuous-ordinal scale. The dichotomous-ordinal weighting scheme was introduced in Cicchetti [9]. The two bottom tables in Table 2 are examples of dichotomous-ordinal scales. All weighting schemes in Table 3, except the general symmetric and the quadratic, are special cases of the weighting scheme with additive weights introduced in Warrens [31].

Table 2: Four examples of 3 × 3 agreement tables from the literature with corresponding values of weighted kappas.

Category labels	3 × 3 table				Kappas	
					Estimates	95% CI
Psychotic	106	10	4	**120**	$\widehat{\kappa} = .429$	(.323–.534)
Neurotic	22	28	10	**60**	$\widehat{\kappa}_\ell = .492$	(.393–.592)
Personality disorder	2	12	6	**20**	$\widehat{\kappa}_q = .567$	(.458–.676)
	130	**50**	**20**	**200**	$\widehat{\kappa}_c = .536$	(.434–.637)
Spitzer et al. [26]					$\widehat{\kappa}_1 = .596$	(.481–.710)
Personality types					$\widehat{\kappa}_2 = .325$	(.182–.468)
					$\widehat{\kappa}_3 = .222$	(.024–.420)
No atopy	136	12	1	**149**	$\widehat{\kappa} = .730$	(.645–.815)
Atopy, no neurodermatitis	8	59	4	**71**	$\widehat{\kappa}_\ell = .737$	(.652–.822)
Neurodermatitis	2	4	6	**12**	$\widehat{\kappa}_q = .748$	(.651–.845)
	146	**75**	**11**	**232**	$\widehat{\kappa}_c = .759$	(.678–.840)
Simonoff [27]					$\widehat{\kappa}_1 = .786$	(.703–.869)
Stability of atopic disease					$\widehat{\kappa}_2 = .720$	(.624–.817)
					$\widehat{\kappa}_3 = .497$	(.240–.754)
Negative	1360	63	8	**1431**	$\widehat{\kappa} = .675$	(.632–.719)
Low positive	61	66	13	**140**	$\widehat{\kappa}_\ell = .761$	(.725–.798)
High positive	10	16	137	**163**	$\widehat{\kappa}_q = .830$	(.798–.862)
	1431	**145**	**158**	**1734**	$\widehat{\kappa}_c = .744$	(.705–.782)
Castle et al. [28]					$\widehat{\kappa}_1 = .716$	(.672–.760)
Results of hybrid capture testing					$\widehat{\kappa}_2 = .415$	(.339–.491)
					$\widehat{\kappa}_3 = .839$	(.794–.884)
Good recovery	36	4	1	**41**	$\widehat{\kappa} = .689$	(.549–.828)
Moderate disability	5	20	4	**29**	$\widehat{\kappa}_\ell = .735$	(.610–.861)
Severe disability	0	1	9	**10**	$\widehat{\kappa}_q = .788$	(.667–.910)
	41	**25**	**14**	**80**	$\widehat{\kappa}_c = .741$	(.614–.868)
Anderson et al. [24]					$\widehat{\kappa}_1 = .750$	(.605–.895)
Glasgow outcome scale scores					$\widehat{\kappa}_2 = .610$	(.427–.793)
					$\widehat{\kappa}_3 = .707$	(.489–.925)

In this paper, we only consider weighted kappas with symmetric weighting schemes. For notational convenience, we define the following six coefficients:

$$\begin{aligned} a_1 &= \pi_{23} + \pi_{32}, & b_1 &= \pi_{2+}\pi_{+3} + \pi_{3+}\pi_{+2}, \\ a_2 &= \pi_{13} + \pi_{31}, & b_2 &= \pi_{1+}\pi_{+3} + \pi_{3+}\pi_{+1}, \\ a_3 &= \pi_{12} + \pi_{21}, & b_3 &= \pi_{1+}\pi_{+2} + \pi_{2+}\pi_{+1}. \end{aligned} \tag{1}$$

To avoid pathological cases, we assume that $b_1, b_2, b_3 > 0$. The coefficients a_1, a_2, and a_3 reflect raw disagreement between the raters, whereas b_1, b_2, and b_3 reflect chance-expected disagreement. The general formula of weighted kappa for 3×3 tables with symmetric weights will be denoted by κ_w. In terms of the coefficients a_1, a_2, a_3 and b_1, b_2, b_3, this weighted kappa is defined as

$$\kappa_w = 1 - \frac{w_1 a_1 + w_2 a_2 + w_3 a_3}{w_1 b_1 + w_2 b_2 + w_3 b_3}. \tag{2}$$

The value of κ_w lies between 1 and $-\infty$. The numerator $w_1 a_1 + w_2 a_2 + w_3 a_3$ of the fraction in (2) reflects raw weighted disagreement. It is a weighted sum of the cell probabilities π_{ij} that are not on the main diagonal of the 3 × 3 table, and it quantifies the disagreement between the raters. The denominator $w_1 b_1 + w_2 b_2 + w_3 b_3$ of the fraction in (2) reflects weighted disagreement under chance. It is a weighted sum of the products $\pi_{i+}\pi_{+j}$ for $i \neq j$. High values of $w_1 a_1 + w_2 a_2 + w_3 a_3$ correspond to high disagreement. If $w_1 a_1 + w_2 a_2 + w_3 a_3 = 0$, then we have $\kappa_w = 1$, and there is perfect agreement between the observers. Furthermore, we have $\kappa_w = 0$ if the raw weighted disagreement is equal to the weighted disagreement under chance.

Special cases of κ_w are obtained by using the specific weighting schemes in Table 3 in the general formula (2). Unweighted kappa, linearly weighted kappa, quadratically weighted kappa, and Cicchetti's weighted kappa are, respectively, defined as

$$\begin{aligned} \kappa &= 1 - \frac{a_1 + a_2 + a_3}{b_1 + b_2 + b_3}, & \kappa_\ell &= 1 - \frac{a_1 + 2a_2 + a_3}{b_1 + 2b_2 + b_3}, \\ \kappa_q &= 1 - \frac{a_1 + 4a_2 + a_3}{b_1 + 4b_2 + b_3}, & \kappa_c &= 1 - \frac{a_1 + 3a_2 + 2a_3}{b_1 + 3b_2 + 2b_3}. \end{aligned} \tag{3}$$

Table 3: Eight weighting schemes for 3 × 3 tables.

Name	Source	Scale type	Symbol	Scheme		
General symmetric	[8]		κ_w	0	w_3	w_2
				w_3	0	w_1
				w_2	w_1	0
Identity	[11]	Nominal	κ	0	1	1
				1	0	1
				1	1	0
Linear	[29, 30]	Continuous-ordinal	κ_ℓ	0	1	2
				1	0	1
				2	1	0
Quadratic	[8]	Continuous-ordinal	κ_q	0	1	4
				1	0	1
				4	1	0
	[9, 31]	Dichotomous-ordinal	κ_c	0	2	3
				2	0	1
				3	1	0
Reliability category 1		Dichotomous	κ_1	0	1	1
				1	0	0
				1	0	0
Reliability category 2		Dichotomous	κ_2	0	1	0
				1	0	1
				0	1	0
Reliability category 3		Dichotomous	κ_3	0	0	1
				0	0	1
				1	1	0

Assuming a multinominal sampling model with the total numbers of objects n fixed, the maximum likelihood estimate of the cell probability π_{ij} for $i, j \in \{1, 2, 3\}$ is given by $\widehat{\pi}_{ij} = n_{ij}/n$, where n_{ij} is the observed frequency. Note that the a_1, a_2, a_3 and b_1, b_2, b_3 are functions of the cell probabilities π_{ij}. The maximum likelihood estimate $\widehat{\kappa}_w$ of κ_w in (2) is obtained by replacing the cell probabilities π_{ij} by $\widehat{\pi}_{ij}$ [32]. The last column of Table 2 contains the estimates of the weighted kappas for each of the four 3 × 3 tables. For example, for the top table of Table 2, we have $\widehat{\kappa} = .429$, $\widehat{\kappa}_\ell = .492$, $\widehat{\kappa}_q = .567$, and $\widehat{\kappa}_c = .536$. Between brackets behind the kappa estimates are the 95% confidence intervals. These were obtained using the asymptotic variance of weighted kappa derived in Fleiss et al. [33].

3. Category Reliabilities

With a categorical scale, it is sometimes desirable to combine some of the categories [34], for example, when two categories are easily confused, and then calculate weighted kappa for the collapsed table. If we combine two of the three categories, the 3 × 3 table collapses into a 2 × 2 table. For a 2 × 2 table, all weighted kappas with symmetric weighting schemes coincide. Since we have three categories, there are three possible ways to combine two categories. The three κ-values of the collapsed 2 × 2 tables are given by

$$\kappa_1 = 1 - \frac{a_2 + a_3}{b_2 + b_3}, \qquad \kappa_2 = 1 - \frac{a_1 + a_3}{b_1 + b_3}, \qquad \kappa_3 = 1 - \frac{a_1 + a_2}{b_1 + b_2}. \tag{4}$$

These three kappas are obtained by using the three bottom weighting schemes in Table 3 in the general formula (2). The last column of Table 2 contains the estimates of these weighted kappas for each of the four 3 × 3 tables.

Weighted kappa κ_i for $i \in \{1, 2, 3\}$ corresponds to the 2 × 2 table that is obtained by combining the two categories other than category i. The 2 × 2 table reflects how often the two raters agreed on the category i and on the category "all others". Weighted kappa κ_i for $i \in \{1, 2, 3\}$, hence, summarizes the agreement or reliability between the raters on the single category i, and it is, therefore, also called the category reliability of i [10]. It quantifies how good category i can be distinguished from the other two categories. For example, for the second table of Table 2, we have $\widehat{\kappa}_1 = .786$, $\widehat{\kappa}_2 = .720$, and $\widehat{\kappa}_3 = .497$. The substantially lower value of κ_3 indicates that the third category is not well distinguished from the other two categories.

Unweighted kappa κ and linearly weighted kappa κ_ℓ are weighted averages of the category reliabilities. Unweighted kappa is a weighted average of κ_1, κ_2, and κ_3, where the weights are the denominators of the category reliabilities [10]:

$$\frac{(b_2+b_3)\kappa_1+(b_1+b_3)\kappa_2+(b_1+b_2)\kappa_3}{(b_2+b_3)+(b_1+b_3)+(b_1+b_2)}=\kappa. \quad (5)$$

Since κ is a weighted average of the category reliabilities, the κ-value always lies between the values of κ_1, κ_2, and κ_3. This property can be verified for all four tables of Table 2. Therefore, when combining two categories, the κ-value can go either up or down, depending on which two categories are combined [34]. The value of κ is a good summary statistic of the category reliabilities if the values of κ_1, κ_2, and κ_3 are (approximately) identical. Table 2 shows that this is not the case in general. With an ordinal scale, it only makes sense to combine categories that are adjacent in the ordering. We should, therefore, ignore κ_2 with ordered categories, since this statistic corresponds to the 2×2 table that is obtained by merging the two categories that are furthest apart. Furthermore, note that for the two bottom 3×3 tables of Table 2 the first category is the "absence" category. If the scale is dichotomous ordinal and category 1 is the "absence" category, then κ_1 is the κ-value of the 2×2 table that corresponds to "absence" versus "presence" of the characteristic.

The statistic κ_ℓ is a weighted average of κ_1 and κ_3, where the weights are the denominators of the category reliabilities [13, 35]:

$$\frac{(b_2+b_3)\kappa_1+(b_1+b_2)\kappa_3}{(b_2+b_3)+(b_1+b_2)}=\kappa_\ell. \quad (6)$$

Since κ_ℓ is a weighted average of the category reliabilities κ_1 and κ_3, the κ_ℓ-value always lies between the values of κ_1 and κ_3. This property can be verified for all four tables of Table 2. Unlike κ_q, statistic κ_ℓ can be considered an extension of κ to ordinal scales that preserves the "weighted average" property [13, 35]. The value of κ_ℓ is a good summary statistic of κ_1 and κ_3 if the two weighted kappas are (approximately) identical. This is the case for the two bottom tables of Table 2.

The statistic κ_c is also a weighted average of κ_1 and κ_3, where the weights are $2(b_2+b_3)$ and (b_1+b_2):

$$\frac{2(b_2+b_3)\kappa_1+(b_1+b_2)\kappa_3}{2(b_2+b_3)+(b_1+b_2)}=\kappa_c. \quad (7)$$

A proof can be found in Warrens [31].

4. Families of Weighted Kappas

In this section, we show that the seven weighted kappas introduced in Sections 2 and 3 are special cases of two families. Let $r \geq 0$ be a real number. Inspection of the formulas κ_2, κ, κ_ℓ, and κ_q shows that they only differ on how the coefficients a_2 and b_2 are weighted. The first family is, therefore, given by

$$\lambda_r = 1-\frac{a_1+ra_2+a_3}{b_1+rb_2+b_3}. \quad (8)$$

For $r = 0, 1, 2, 4$, we have, respectively, the special cases κ_2, κ, κ_ℓ, and κ_q.

Recall that κ_ℓ and κ_c are weighted averages of the category reliabilities κ_1 and κ_3. This motivates the following definition. Let $s \in [0, 1]$. Then the second family is defined as

$$\mu_s = \frac{(1-s)(b_2+b_3)\kappa_1+s(b_1+b_2)\kappa_3}{(1-s)(b_2+b_3)+s(b_1+b_2)}. \quad (9)$$

The family μ_s consists of the weighted averages of κ_1 and κ_3 where the weights are multiples of (b_2+b_3) and (b_1+b_2). For $s = 0, 1/3, 1/2, 1$, we have, respectively, the special cases κ_1, κ_c, κ_ℓ, and κ_3. Note that κ_ℓ belongs to both λ_r and μ_s.

The following proposition presents a formula for the family in (9) that will be used in Theorem 6 below.

Proposition 1. *The family in* (9) *is equivalent to*

$$\mu_s = 1-\frac{sa_1+a_2+(1-s)a_3}{sb_1+b_2+(1-s)b_3}. \quad (10)$$

Proof. Since κ_1 and κ_3 are equal to, respectively,

$$\kappa_1 = \frac{b_2+b_3-(a_2+a_3)}{b_2+b_3}, \qquad \kappa_3 = \frac{b_1+b_2-(a_1+a_2)}{b_1+b_2}, \quad (11)$$

we can write (9) as

$$\begin{aligned}\mu_s &= \frac{(1-s)(b_2+b_3-a_2-a_3)+s(b_1+b_2-a_1-a_2)}{(1-s)(b_2+b_3)+s(b_1+b_2)}\\ &= 1-\frac{(1-s)(a_2+a_3)+s(a_1+a_2)}{(1-s)(b_2+b_3)+s(b_1+b_2)},\end{aligned} \quad (12)$$

which is identical to the expression in (10). □

5. Inequalities

In this section, we present inequalities between the seven weighted kappas. We will use the following lemma repeatedly.

Lemma 2. *Let $u, v \geq 0$ and $r, w, z > 0$. Then one has the following:*

$$\begin{aligned}&(i) \quad \frac{u}{w}<\frac{v}{z} \Longleftrightarrow \frac{u}{w}<\frac{ru+v}{rw+z};\\ &(ii) \quad \frac{u}{w}=\frac{v}{z} \Longleftrightarrow \frac{u}{w}=\frac{ru+v}{rw+z};\\ &(iii) \quad \frac{u}{w}>\frac{v}{z} \Longleftrightarrow \frac{u}{w}>\frac{ru+v}{rw+z}.\end{aligned} \quad (13)$$

Proof. Since w and z are positive numbers, we have $u/w < v/z$, or $uz < vw$. Adding ruw to both sides, we obtain $u(rw + z) < w(ru + v)$, or $u/w < (ru + v)/(rw + z)$. □

Theorem 3 classifies the orderings of the special cases of the family λ_r in (8).

Theorem 3. *For $r < r'$ one has the following:*

$$\begin{aligned}
&(i) \quad \lambda_r < \lambda_{r'} \Longleftrightarrow \frac{a_1 + a_3}{b_1 + b_3} > \frac{a_2}{b_2};\\
&(ii) \quad \lambda_r = \lambda_{r'} \Longleftrightarrow \frac{a_1 + a_3}{b_1 + b_3} = \frac{a_2}{b_2};\\
&(iii) \quad \lambda_r > \lambda_{r'} \Longleftrightarrow \frac{a_1 + a_3}{b_1 + b_3} < \frac{a_2}{b_2}.
\end{aligned} \tag{14}$$

Proof. The inequality $\lambda_r < \lambda_{r'}$ is equivalent to

$$\frac{a_1 + ra_2 + a_3}{b_1 + rb_2 + b_3} > \frac{a_1 + r'a_2 + a_3}{b_1 + r'b_2 + b_3}. \tag{15}$$

Since $r < r'$, it follows from Lemma 2 that inequality (15) is equivalent to

$$\frac{a_1 + ra_2 + a_3}{b_1 + rb_2 + b_3} > \frac{(r' - r) a_2}{(r' - r) b_2} = \frac{a_2}{b_2}. \tag{16}$$

Applying Lemma 2 for a second time, we find that inequality (16) is equivalent to

$$\frac{a_1 + a_3}{b_1 + b_3} > \frac{a_2}{b_2}. \tag{17}$$

This completes the proof. □

Theorem 3 shows that, in practice, we only observe one of two orderings of κ_2, κ, κ_ℓ, and κ_q. In most cases, we have $\kappa_2 < \kappa < \kappa_\ell < \kappa_q$. For example, in Table 2 all 3 × 3 tables exhibit this ordering. For all these 3 × 3 tables, it holds that $(a_1 + a_3)/(b_1 + b_3) > a_2/b_2$. Furthermore, if the 3 × 3 table would be tridiagonal [22, 23], we would have $a_2 = 0$, and the inequality $(a_1 + a_3)/(b_1 + b_3) > a_2/b_2$ would also hold. The other possibility is that we have $\kappa_2 > \kappa > \kappa_\ell > \kappa_q$. The only example from the literature where we found this ordering is the 3 × 3 table presented in Cohen [11]. The table in Cohen satisfies the condition in (iii) of Theorem 3. We conclude that, with ordinal scales, we almost always have the ordering $\kappa_2 < \kappa < \kappa_\ell < \kappa_q$. The equality condition in Theorem 3 is discussed in Section 6.

Theorem 4 classifies the orderings of the special cases of the family μ_s in (9).

Theorem 4. *For $s < s'$, one has the following:*

$$\begin{aligned}
&(i) \quad \mu_s < \mu_{s'} \Longleftrightarrow \frac{a_1 + a_2}{b_1 + b_2} > \frac{a_2 + a_3}{b_2 + b_3};\\
&(ii) \quad \mu_s = \mu_{s'} \Longleftrightarrow \frac{a_1 + a_2}{b_1 + b_2} = \frac{a_2 + a_3}{b_2 + b_3};\\
&(iii) \quad \mu_s > \mu_{s'} \Longleftrightarrow \frac{a_1 + a_2}{b_1 + b_2} < \frac{a_2 + a_3}{b_2 + b_3}.
\end{aligned} \tag{18}$$

Proof. The special cases of μ_s are weighted averages of κ_1 and κ_3. For $s < s'$, we have $\mu_s < \mu_{s'}$ if and only if $\kappa_1 < \kappa_3$; that is, a statistic that gives more weight to κ_3 will be higher if the κ_3-value exceeds the κ_1-value. Furthermore, we have $\kappa_1 < \kappa_3 \Leftrightarrow$

$$\frac{a_1 + a_2}{b_1 + b_2} > \frac{a_2 + a_3}{b_2 + b_3}. \tag{19}$$

This completes the proof. □

Theorem 4 shows that, in practice, we only observe one of two orderings of $\kappa_3, \kappa_\ell, \kappa_c$, and κ_1. We either have the ordering $\kappa_3 < \kappa_\ell < \kappa_c < \kappa_1$, which is the case in the first, second, and fourth 3 × 3 tables of Table 2, or we have $\kappa_3 > \kappa_\ell > \kappa_c > \kappa_1$, which is the case in the third 3 × 3 table in Table 2.

Proposition 5 follows from Theorems 3 and 4 and the fact that κ is a weighted average of κ_1, κ_2, and κ_3 [10].

Proposition 5. *Consider the following:*

$$\begin{aligned}
&(i) \quad \kappa < \kappa_3 < \kappa_1 \Longleftrightarrow \kappa_2 < \kappa < \kappa_3 < \kappa_\ell < \kappa_c < \kappa_1;\\
&(ii) \quad \kappa < \kappa_1 < \kappa_3 \Longleftrightarrow \kappa_2 < \kappa < \kappa_1 < \kappa_c < \kappa_\ell < \kappa_3, \kappa_q;\\
&(iii) \quad \kappa_3 < \kappa_1 < \kappa \Longleftrightarrow \kappa_3, \kappa_q < \kappa_\ell < \kappa_c < \kappa_1 < \kappa < \kappa_2;\\
&(iv) \quad \kappa_1 < \kappa_3 < \kappa \Longleftrightarrow \kappa_1 < \kappa_c < \kappa_\ell < \kappa_3 < \kappa < \kappa_2.
\end{aligned} \tag{20}$$

Proposition 5 shows that we have an almost complete picture of how the seven weighted kappas are ordered just by comparing the values of κ, κ_1, and κ_3. The double inequality $\kappa < \kappa_3 < \kappa_1$ holds for the fourth 3×3 table of Table 2, whereas the inequality $\kappa < \kappa_1 < \kappa_3$ holds for the third 3 × 3 table of Table 2. Both tables have a dichotomous-ordinal scale. Recall that κ_c corresponds to a weighting scheme specifically formulated for dichotomous-ordinal scales. It turns out that the κ_c-value can be both lower and higher than the κ_ℓ-value with dichotomous-ordinal scales. Which statistic is higher depends on the data. Furthermore, κ tends to be smaller than κ_1 and κ_3. The condition $\kappa < \kappa_1, \kappa_3$ can be interpreted as an increase in the κ-value if we combine the middle category of the 3-category scale with one of the outer categories. This way of merging categories makes sense if the categories are ordered.

6. Equalities

Apart from the equality conditions in (ii) of Theorems 3 and 4, we only considered inequalities between the weighted kappas in the previous section. Unless there is perfect agreement, the values of the weighted kappas are usually different. Table 4 contains three hypothetical agreement tables that we have constructed to illustrate that the three equality conditions in Theorems 3, 4, and 6 (below) are not identical. For the top table in Table 4, we have $(a_1 + a_3)/(b_1 + b_3) = a_2/b_2$, which is equivalent to the equality $\kappa_2 = \kappa = \kappa_\ell = \kappa_q$ (Theorem 3). Although all weighted kappas of the family λ_r coincide, the kappas not belonging to this family produce different values. For the middle table in Table 4 we have $(a_1 + a_2)/(b_1 + b_2) = (a_2 + a_3)/(b_2 + b_3)$, which is equivalent to the equality $\kappa_3 = \kappa_\ell = \kappa_c = \kappa_1$ (Theorem 4). Although all weighted kappas of

Table 4: Three hypothetical 3 × 3 agreement tables with corresponding values of weighted kappas.

Categories	3 × 3 table				Kappas	
1	4	1	0	**5**	$\widehat{\kappa} = .617$	$\widehat{\kappa}_1 = .475$
2	1	2	0	**3**	$\widehat{\kappa}_\ell = .617$	$\widehat{\kappa}_2 = .617$
3	3	0	12	**15**	$\widehat{\kappa}_q = .617$	$\widehat{\kappa}_3 = .736$
	8	**3**	**12**	**23**	$\widehat{\kappa}_c = .572$	
1	6	0	1	**7**	$\widehat{\kappa} = .581$	$\widehat{\kappa}_1 = .635$
2	3	6	0	**9**	$\widehat{\kappa}_\ell = .635$	$\widehat{\kappa}_2 = .479$
3	0	3	6	**9**	$\widehat{\kappa}_q = .668$	$\widehat{\kappa}_3 = .635$
	9	**9**	**7**	**25**	$\widehat{\kappa}_c = .635$	
1	11	1	0	**12**	$\widehat{\kappa} = .603$	$\widehat{\kappa}_1 = .603$
2	2	5	0	**7**	$\widehat{\kappa}_\ell = .603$	$\widehat{\kappa}_2 = .603$
3	2	1	3	**6**	$\widehat{\kappa}_q = .603$	$\widehat{\kappa}_3 = .603$
	15	**7**	**3**	**25**	$\widehat{\kappa}_c = .603$	

the family μ_s coincide, the kappas that do not belong to this family produce different values.

For the bottom table in Table 4, we have the stronger condition $a_1/b_1 = a_2/b_2 = a_3/b_3$. Theorem 6 (below) shows that this condition is equivalent to the case that all weighted kappas, that is, all special cases of (2), coincide.

Theorem 6. *The following conditions are equivalent:*

$$\begin{aligned} &(i) \quad \frac{a_1}{b_1} = \frac{a_2}{b_2} = \frac{a_3}{b_3} = c \geq 0; \\ &(ii) \quad \kappa_w = 1 - c; \\ &(iii) \quad \lambda_r = \lambda_t = \mu_s \quad \text{for } r \neq t,\ s \neq \frac{1}{2}; \\ &(iv) \quad \lambda_r = \mu_s = \mu_t \quad \text{for } r \neq 2,\ s \neq t. \end{aligned} \tag{21}$$

Proof. In words, (ii) means that all special cases of (2) are identical. Therefore, (ii) ⇒ (iii), (iv). We first show that (i) ⇒ (ii). It then suffices to show that (iii), (iv) ⇒ (i).

If (i) holds, we have

$$\frac{a_2}{b_2} = \frac{c_1 a_1}{c_1 b_1}, \qquad \frac{a_3}{b_3} = \frac{c_2 a_1}{c_2 b_1} \tag{22}$$

for certain $c_1, c_2 > 0$. Hence,

$$\begin{aligned} \kappa_w &= 1 - \frac{w_1 a_1 + w_2 c_1 a_1 + w_3 c_2 a_1}{w_1 b_1 + w_2 c_1 b_1 + w_3 c_2 b_1} \\ &= 1 - \frac{a_1 (w_1 + w_2 c_1 + w_3 c_2)}{b_1 (w_1 + w_2 c_1 + w_3 c_2)} \\ &= 1 - \frac{a_1}{b_1} = 1 - c. \end{aligned} \tag{23}$$

Thus, all special cases of weighted kappa in (2) coincide if (i) is valid.

Next, we show that (iii), (iv) ⇒ (i). Consider condition (iii) first. If two special cases of λ_r are identical, it follows from Theorem 3 that all of them are identical. Hence, we have $\kappa_2 = \mu_s$ for a certain $s \in [0, 1]$ with $s \neq 1/2$. Using formula (10), we have $\kappa_2 = \mu_s \Leftrightarrow$

$$\frac{a_1 + a_3}{b_1 + b_3} = \frac{s a_1 + a_2 + (1 - s) a_3}{s b_1 + b_2 + (1 - s) b_3}. \tag{24}$$

Combining (24) with $a_2/b_2 = (a_1 + a_3)/(b_1 + b_3)$ (Theorem 3), we obtain

$$\frac{a_2}{b_2} = \frac{a_1 + a_3}{b_1 + b_3} = \frac{s a_1 + a_2 + (1 - s) a_3}{s b_1 + b_2 + (1 - s) b_3}. \tag{25}$$

Applying Lemma 2 to the outer ratios of (25), we obtain

$$\frac{a_2}{b_2} = \frac{a_1 + a_3}{b_1 + b_3} = \frac{s a_1 + (1 - s) a_3}{s b_1 + (1 - s) b_3}. \tag{26}$$

First, suppose that $s < 1/2$. Applying Lemma 2 to the right-hand side equality of (26), we obtain

$$\frac{a_2}{b_2} = \frac{a_1 + a_3}{b_1 + b_3} = \frac{(1 - 2s) a_3}{(1 - 2s) b_3} = \frac{a_3}{b_3}, \tag{27}$$

or $a_2/b_2 = a_3/b_3$. Applying Lemma 2 to the second and fourth term of the triple equality (27), we obtain $a_1/b_1 = a_3/b_3$. Thus, we have $a_1/b_1 = a_2/b_2 = a_3/b_3$, which completes the proof for $s < 1/2$. Next, suppose that $s > 1/2$. Applying Lemma 2 to the right-hand side equality of (26), we obtain

$$\frac{a_2}{b_2} = \frac{a_1 + a_3}{b_1 + b_3} = \frac{(2s - 1) a_1}{(2s - 1) b_1} = \frac{a_1}{b_1}, \tag{28}$$

or $a_1/b_1 = a_2/b_2$. Applying Lemma 2 to the second and fourth terms of the triple equality (28), we obtain $a_1/b_1 = a_3/b_3$. Thus, we also have $a_1/b_1 = a_2/b_2 = a_3/b_3$ for $s > 1/2$, which completes the proof for condition (iii).

Next, consider condition (iv). If two special cases of μ_s are identical, it follows from Theorem 4 that all of them are identical. Hence, we have $\kappa_1 = \kappa_3 = \lambda_r$ for a certain $r \geq 0$ and $r \neq 2$. We have $\kappa_3 = \kappa_1 = \lambda_r \Leftrightarrow$

$$\frac{a_1 + a_2}{b_1 + b_2} = \frac{a_2 + a_3}{b_2 + b_3} = \frac{a_1 + r a_2 + a_3}{b_1 + r b_2 + b_3}. \tag{29}$$

First, suppose that $r > 2$. Applying Lemma 2 to the outer ratios of (29), we obtain

$$\frac{a_1 + a_2}{b_1 + b_2} = \frac{a_2 + a_3}{b_2 + b_3} = \frac{(r-1)\,a_2 + a_3}{(r-1)\,b_2 + b_3}. \tag{30}$$

Applying Lemma 2 to the right-hand side equality of (30) gives

$$\frac{a_1 + a_2}{b_1 + b_2} = \frac{a_2 + a_3}{b_2 + b_3} = \frac{(r-2)\,a_2}{(r-2)\,b_2} = \frac{a_2}{b_2}. \tag{31}$$

Applying Lemma 2 to the outer ratios of (31), we obtain $a_1/b_1 = a_2/b_2$, while applying Lemma 2 to the second and fourth terms of the triple equality (31), we obtain $a_3/b_3 = a_2/b_2$. Thus, we have $a_1/b_1 = a_2/b_2 = a_3/b_3$.

Finally, if $r < 2$, then consider the equality $\kappa_1 = \kappa_3 = \kappa_\ell = \lambda_r \Leftrightarrow$

$$\begin{aligned}\frac{a_2 + a_3}{b_2 + b_3} = \frac{a_1 + a_2}{b_1 + b_2} &= \frac{a_1 + 2a_2 + a_3}{b_1 + 2b_2 + b_3} \\ &= \frac{(2/r)\,a_1 + 2a_2 + (2/r)\,a_3}{(2/r)\,b_1 + 2b_2 + (2/r)\,b_3}.\end{aligned} \tag{32}$$

Since $2/r > 1$, applying Lemma 2 to the right-hand side equality of (32) gives

$$\begin{aligned}\frac{a_1 + a_2}{b_1 + b_2} = \frac{a_2 + a_3}{b_2 + b_3} &= \frac{a_1 + 2a_2 + a_3}{b_1 + 2b_2 + b_3} \\ &= \frac{(2/r-1)\,a_1 + (2/r-1)\,a_3}{(2/r-1)\,b_1 + (2/r-1)\,b_3} \\ &= \frac{a_1 + a_3}{b_1 + b_3}.\end{aligned} \tag{33}$$

However,

$$\frac{a_1 + a_2}{b_1 + b_2} = \frac{a_2 + a_3}{b_2 + b_3} = \frac{a_1 + a_3}{b_1 + b_3} \tag{34}$$

is equivalent to $\kappa_1 = \kappa_2 = \kappa_3$. Since κ is a weighted average of κ_1, κ_2, and κ_3, we must have $\kappa = \kappa_2$. But then condition (iii) holds, and we have already shown that (iii) ⇒ (i). This completes the proof for condition (iv). □

Theorem 6 shows that all weighted kappas for 3 × 3 tables are identical if we have the double inequality $a_1/b_1 = a_2/b_2 = a_3/b_3$. If this condition holds, the equalities $(a_1+a_3)/(b_1+b_3) = a_2/b_2$ and $(a_1 + a_2)/(b_1 + b_2) = (a_2 + a_3)/(b_2 + b_3)$ also hold. Theorem 6 also shows that if any two special cases of the family λ_r are equal to a member of the family μ_s other than κ_ℓ, then all weighted kappas coincide. Furthermore, if any two special cases of the family μ_s are identical to a member of the family λ_r other than κ_ℓ, then all weighted kappas must be identical.

7. Discussion

Since it frequently happens that different versions of the weighted kappa are applied to the same contingency data, regardless of the scale type of the categories, it is useful to compare the various versions analytically. For rating scales with three categories, we may define seven special cases of weighted kappa. The seven weighted kappas belong to two different parameter families. Only the weighted kappa with linear weights belongs to both families. For both families, it was shown that there are only two possible orderings of its members (Theorems 3 and 4). We conclude that with ordinal scales consisting of three categories, quadratically weighted kappa usually produces higher values than linearly weighted kappa, which in turn has higher values than unweighted kappa.

Since there are only a few possible orderings of the weighted kappas, it appears that the kappas are measuring the same thing, but to a different extent. Various authors have presented magnitude values for evaluating the values of kappa statistics [36–38]. For example, an estimated value of 0.80 generally indicates good or excellent agreement. There is general consensus in the literature that uncritical use of these guidelines leads to questionable decisions in practice. If the weighted kappas are measuring the same thing, but some kappas produce substantially higher values than others, then the same guidelines cannot be applied to all weighted kappas. However, using the same guidelines for different kappas appears to be common practice. If one wants to work with magnitude guidelines, then it seems reasonable to use stricter criteria for the quadratically weighted kappa than for unweighted kappa, since the former statistic generally produces higher values.

The quadratically and linearly weighted kappas were formulated for continuous-ordinal scale data. However, in practice, many scales are dichotomous ordinal (see, e.g., Anderson et al. [24] and Martin et al. [25]). In this case, the application of the weighted kappa proposed by Cicchetti [9] or the additively weighted kappa introduced in Warrens [31] is perhaps more appropriate. Unfortunately, Cicchetti's weighted kappa has been largely ignored in the application of kappa statistics. In most applications, the quadratically weighted kappa is used [4, 5]. The observation that the quadratically weighted kappa tends to produce the highest value for many data may partly explain this popularity. As pointed out by one of the reviewers, to determine whether Cicchetti's weighted kappa has real advantages, the various weighted kappas need to be compared on the quality and efficiency of prediction. This is a possible topic for future work.

Acknowledgments

The author thanks four anonymous reviewers for their helpful comments and valuable suggestions on an earlier version of this paper. This research is part of Project 451-11-026 funded by the Netherlands Organisation for Scientific Research.

References

[1] M. A. Tanner and M. A. Young, "Modeling ordinal scale disagreement," *Psychological Bulletin*, vol. 98, no. 2, pp. 408–415, 1985.

[2] A. Agresti, "A model for agreement between ratings on an ordinal scale," *Biometrics*, vol. 44, no. 2, pp. 539–548, 1988.

[3] A. Agresti, *Analysis of Ordinal Categorical Data*, John Wiley & Sons, Hoboken, NJ, USA, 2nd edition, 2010.

[4] P. Graham and R. Jackson, "The analysis of ordinal agreement data: beyond weighted kappa," *Journal of Clinical Epidemiology*, vol. 46, no. 9, pp. 1055–1062, 1993.

[5] M. Maclure and W. C. Willett, "Misinterpretation and misuse of the Kappa statistic," *American Journal of Epidemiology*, vol. 126, no. 2, pp. 161–169, 1987.

[6] J. de Mast and W. N. van Wieringen, "Measurement system analysis for categorical measurements: agreement and kappa-type indices," *Journal of Quality Technology*, vol. 39, no. 3, pp. 191–202, 2007.

[7] M. J. Warrens, "Inequalities between kappa and kappa-like statistics for $k \times k$ tables," *Psychometrika*, vol. 75, no. 1, pp. 176–185, 2010.

[8] J. Cohen, "Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit," *Psychological Bulletin*, vol. 70, no. 4, pp. 213–220, 1968.

[9] D. V. Cicchetti, "Assessing inter rater reliability for rating scales: resolving some basic issues," *British Journal of Psychiatry*, vol. 129, no. 11, pp. 452–456, 1976.

[10] M. J. Warrens, "Cohen's kappa is a weighted average," *Statistical Methodology*, vol. 8, no. 6, pp. 473–484, 2011.

[11] J. Cohen, "A coefficient of agreement for nominal scales," *Educational and Psychological Measurement*, vol. 20, pp. 37–46, 1960.

[12] S. Vanbelle and A. Albert, "A note on the linearly weighted kappa coefficient for ordinal scales," *Statistical Methodology*, vol. 6, no. 2, pp. 157–163, 2009.

[13] M. J. Warrens, "Cohen's linearly weighted kappa is a weighted average of 2×2 kappas," *Psychometrika*, vol. 76, no. 3, pp. 471–486, 2011.

[14] J. L. Fleiss and J. Cohen, "The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability," *Educational and Psychological Measurement*, vol. 33, pp. 613–619, 1973.

[15] M. J. Warrens, "Some paradoxical results for the quadratically weighted kappa," *Psychometrika*, vol. 77, no. 2, pp. 315–323, 2012.

[16] L. M. Hsu and R. Field, "Interrater agreement measures: comments on kappa$_n$, Cohen's kappa, Scott's π and Aickin's α," *Understanding Statistics*, vol. 2, pp. 205–219, 2003.

[17] E. Bashkansky, T. Gadrich, and D. Knani, "Some metrological aspects of the comparison between two ordinal measuring systems," *Accreditation and Quality Assurance*, vol. 16, no. 2, pp. 63–72, 2011.

[18] J. de Mast, "Agreement and kappa-type indices," *The American Statistician*, vol. 61, no. 2, pp. 148–153, 2007.

[19] J. S. Uebersax, "Diversity of decision-making models and the measurement of interrater agreement," *Psychological Bulletin*, vol. 101, no. 1, pp. 140–146, 1987.

[20] W. D. Perreault and L. E. Leigh, "Reliability of nominal data based on qualitative judgments," *Journal of Marketing Research*, vol. 26, pp. 135–148, 1989.

[21] M. J. Warrens, "Conditional inequalities between Cohen's kappa and weighted kappas," *Statistical Methodology*, vol. 10, pp. 14–22, 2013.

[22] M. J. Warrens, "Weighted kappa is higher than Cohen's kappa for tridiagonal agreement tables," *Statistical Methodology*, vol. 8, no. 2, pp. 268–272, 2011.

[23] M. J. Warrens, "Cohen's quadratically weighted kappa is higher than linearly weighted kappa for tridiagonal agreement tables," *Statistical Methodology*, vol. 9, no. 3, pp. 440–444, 2012.

[24] S. I. Anderson, A. M. Housley, P. A. Jones, J. Slattery, and J. D. Miller, "Glasgow outcome scale: an inter-rater reliability study," *Brain Injury*, vol. 7, no. 4, pp. 309–317, 1993.

[25] C. S. Martin, N. K. Pollock, O. G. Bukstein, and K. G. Lynch, "Inter-rater reliability of the SCID alcohol and substance use disorders section among adolescents," *Drug and Alcohol Dependence*, vol. 59, no. 2, pp. 173–176, 2000.

[26] R. L. Spitzer, J. Cohen, J. L. Fleiss, and J. Endicott, "Quantification of agreement in psychiatric diagnosis. A new approach," *Archives of General Psychiatry*, vol. 17, no. 1, pp. 83–87, 1967.

[27] J. S. Simonoff, *Analyzing Categorical Data*, Springer, New York, NY, USA, 2003.

[28] P. E. Castle, A. T. Lorincz, I. Mielzynska-Lohnas et al., "Results of human papillomavirus DNA testing with the hybrid capture 2 assay are reproducible," *Journal of Clinical Microbiology*, vol. 40, no. 3, pp. 1088–1090, 2002.

[29] D. Cicchetti and T. Allison, "A new procedure for assessing reliability of scoring EEG sleep recordings," *The American Journal of EEG Technology*, vol. 11, pp. 101–110, 1971.

[30] D. V. Cicchetti, "A new measure of agreement between rank ordered variables," in *Proceedings of the Annual Convention of the American Psychological Association*, vol. 7, pp. 17–18, 1972.

[31] M. J. Warrens, "Cohen's weighted kappa with additive weights," *Advances in Data Analysis and Classification*, vol. 7, pp. 41–55, 2013.

[32] Y. M. M. Bishop, S. E. Fienberg, and P. W. Holland, *Discrete Multivariate Analysis: Theory and Practice*, The MIT Press, Cambridge, Mass, USA, 1975.

[33] J. L. Fleiss, J. Cohen, and B. S. Everitt, "Large sample standard errors of kappa and weighted kappa," *Psychological Bulletin*, vol. 72, no. 5, pp. 323–327, 1969.

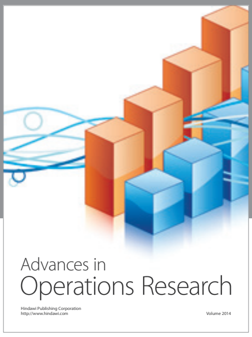
[34] M. J. Warrens, "Cohen's kappa can always be increased and decreased by combining categories," *Statistical Methodology*, vol. 7, no. 6, pp. 673–677, 2010.

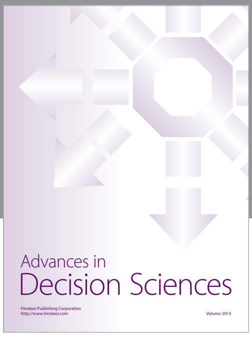
[35] M. J. Warrens, "Cohen's linearly weighted kappa is a weighted average," *Advances in Data Analysis and Classification*, vol. 6, no. 1, pp. 67–79, 2012.

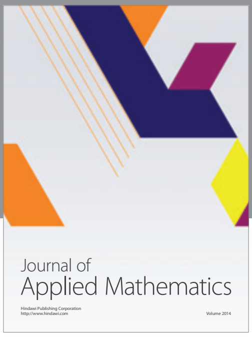
[36] D. V. Cicchetti and S. A. Sparrow, "Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior," *American Journal of Mental Deficiency*, vol. 86, no. 2, pp. 127–137, 1981.

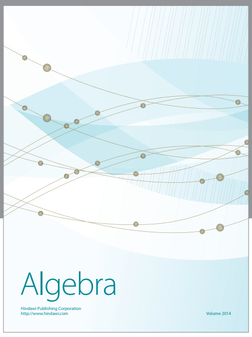
[37] P. E. Crewson, "Reader agreement studies," *American Journal of Roentgenology*, vol. 184, no. 5, pp. 1391–1397, 2005.

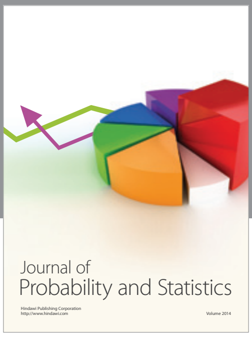
[38] J. R. Landis and G. G. Koch, "A one-way components of variance model for categorical data," *Biometrics*, vol. 33, no. 4, pp. 159–174, 1977.


Advances in
Operations Research

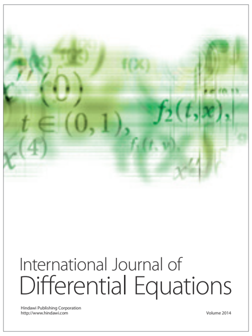

Advances in
Decision Sciences


Journal of
Applied Mathematics

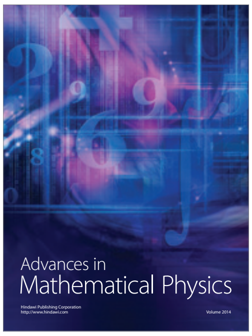

Algebra

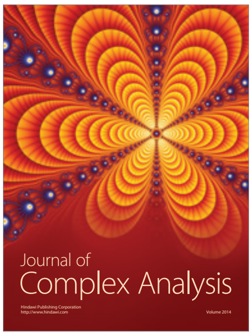

Journal of
Probability and Statistics

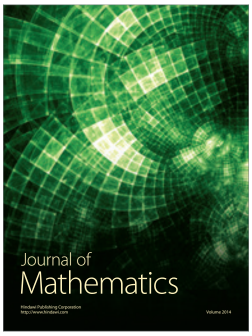
The Scientific
World Journal

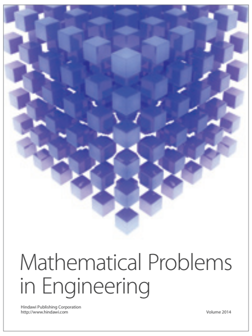

International Journal of
Differential Equations

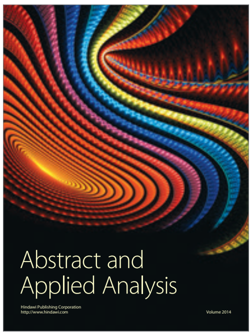
International Journal of
Combinatorics

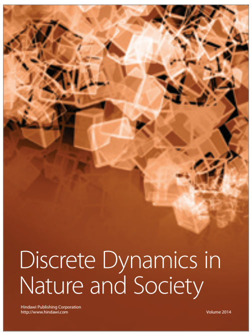

Advances in
Mathematical Physics

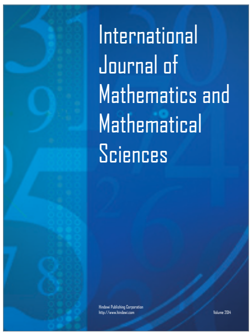

Journal of
Complex Analysis


Journal of
Mathematics


Mathematical Problems
in Engineering

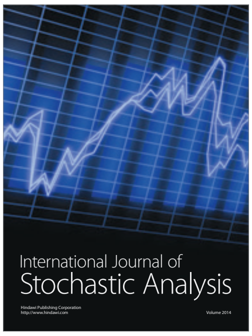

Abstract and
Applied Analysis


Discrete Dynamics in
Nature and Society


International
Journal of
Mathematics and
Mathematical
Sciences

Journal of
Discrete Mathematics

Journal of
Function Spaces


International Journal of
Stochastic Analysis

Journal of
Optimization