

Research Article

On Marginal Dependencies of the 2×2 Kappa

Matthijs J. Warrens

Methodology and Statistics Unit, 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 19 July 2014; Accepted 1 November 2014; Published 19 November 2014

Academic Editor: Lynn Kuo

Copyright © 2014 Matthijs J. Warrens. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Cohen's kappa is a standard tool for the analysis of agreement in a 2×2 reliability study. Researchers are frequently only interested in the kappa-value of a sample. Various authors have observed that if two pairs of raters have the same amount of observed agreement, the pair whose marginal distributions are more similar to each other may have a lower kappa-value than the pair with more divergent marginal distributions. Here we present exact formulations of some of these properties. The results provide a better understanding of the 2×2 kappa for situations where it is used as a sample statistic.

1. Introduction

Results from experimental studies and research studies can often be summarized in a 2×2 table [1]. An example is a reliability study in which two observers rate the same sample of subjects on the presence/absence of a trait or an ability [2, 3]. In this example the four cells of the 2×2 table are the proportion of times the observers agreed on the presence of the trait, the proportion of times a trait was present according to the first observer but absent according to the second observer, the proportion of times a trait was absent according to the first observer but present according to the second observer, and the proportion of times the observers agreed on the absence of the trait.

To assess the quality of the ratings, the agreement between the ratings is taken as an indicator of the quality of the category definitions and the observers' ability to apply them. A standard tool for estimating agreement in a 2×2 reliability study is Cohen's kappa [4–8]. Its value is 1 when there is perfect agreement, 0 when agreement is equal to that expected under independence, and negative when agreement is less than expected by chance.

Several authors have presented population models for Cohen's kappa [2, 7]. Under these models kappa can be interpreted as an association coefficient. However, kappa is also frequently used as a sample statistic [4, 8–11], for example, when calculating kappa for a sample of subjects is one step in a series of research steps. In this case, researchers

are merely interested in the agreement in the sample not that of a population.

As a sample statistic, kappa is known to be marginal or prevalence dependent since it takes the marginal totals with which raters use the rating categories into account [12–14]. The value of kappa depends on the prevalence of the condition being diagnosed. Values of kappa can be quite low if a condition is quite common or very rare. Various authors have shown that if two pairs of observers have the same amount of observed agreement, the pair whose marginal distributions are more similar to each other may have a lower kappa-value than the pair with more divergent marginal distributions [4, 9, 15, 16]. Since observers with similar marginal distributions usually have a higher amount of agreement expected to occur by chance, a fixed amount of observed agreement will lead to a lower kappa-value due to the definition of the statistic [14].

The marginal dependence of Cohen's kappa has almost exclusively been demonstrated and described by means of examples of 2×2 tables [4, 8–11, 16]. However, for better understanding of the behavior of the 2×2 kappa as a sample statistic, it is desirable to have exact formulations of its marginal dependence. Such formulations are presented in this paper. The paper is organized as follows. Section 2 is used to introduce notation and define Cohen's kappa. In Section 3 several concepts for a 2×2 table are presented. The main results are presented in Section 4. The results show that the 2×2 kappa may exhibit several different forms of

TABLE 1: Bivariate table of relative frequencies corresponding to two observers making binary (1, 0) decisions on a sample of subjects.

Observer 1	Observer 2		Total
	1	0	
1	a	b	p_1
0	c	d	q_1
Total	p_2	q_2	1

marginal dependence. The results do not necessarily suggest that Cohen's kappa should be discarded as an agreement measure. Instead, the exact formulations provide a better understanding of the 2×2 statistic. Section 5 contains several conclusions.

2. Notation and Kappa

In this section we introduce notation and define the 2×2 kappa coefficient. Suppose two fixed observers independently rate the same set of subjects using the same two categories 1 and 0. For example, 1 = presence and 0 = absence of a trait. For a sample of subjects, let $a, b, c,$ and d denote, respectively, the proportion classified in category 1 by both observers, the proportion classified by the first observer in category 1 and by the second observer in category 0, the proportion classified by the first observer in category 0 and by the second observer in category 1, and the proportion classified in category 0 by both observers.

A general 2×2 table with observed proportions, denoted by P , is presented in Table 1. The row and column totals are the marginal totals that result from summing the relative frequencies. We denote these by p_1 and q_1 for rater 1 and by p_2 and q_2 for rater 2. They reflect how often the observers used the categories.

Using the notation presented in Table 1, the proportion of observed agreement is given by $a + d$, while the proportion of expected agreement is given by $p_1 p_2 + q_1 q_2$. Furthermore, Cohen's kappa can be defined as

$$\kappa := \frac{a + d - p_1 p_2 - q_1 q_2}{1 - p_1 p_2 - q_1 q_2}. \quad (1)$$

Cohen's kappa is a chance-corrected coefficient [17]. The value of kappa is 1 when perfect agreement between the two observers occurs, 0 when agreement is equal to that expected under independence ($a + d = p_1 p_2 + q_1 q_2$), and negative when agreement is less than expected by chance.

3. Concepts for a 2×2 Table

In this section we introduce several concepts for a 2×2 table P . The 2-tuples (p_1, q_1) and (p_2, q_2) contain the marginal distributions.

Definition 1. Two tuples (p_1, q_1) and (p_2, q_2) are said to be

- (i) *similarly arranged* if both increase (i.e., $p_1 < q_1$ and $p_2 < q_2$) or both decrease (i.e., $p_1 > q_1$ and $p_2 > q_2$);

- (ii) *oppositely arranged* if one increases and the other decreases.

Furthermore, a tuple (p, q) is said to be *balanced* if $p = q = 1/2$.

In the following definition we use the concepts from Definition 1 to define terminology that will be used to formalize the marginal dependencies of kappa.

Definition 2. A 2×2 table P is said to be

- (i) *strongly marginal symmetric* if $p_1 = p_2 \neq 1/2$ (and hence $q_1 = q_2$);
- (ii) *weakly marginal symmetric* if (p_1, q_1) and (p_2, q_2) are similarly arranged;
- (iii) *balanced* if (p_1, q_1) or (p_2, q_2) is balanced;
- (iv) *marginal asymmetric* if (p_1, q_1) and (p_2, q_2) are oppositely arranged.

Note that strong marginal symmetry implies weak marginal symmetry. Furthermore, strong marginal symmetry coincides with the usual definition of a symmetric matrix.

The following result relates some of the concepts in Definition 2 to the value of the chance-expected agreement $p_1 p_2 + q_1 q_2$. Lemma 3 is used in the proof of Theorem 5.

Lemma 3. For a 2×2 table P with marginal distributions (p_1, q_1) and (p_2, q_2) the following equivalences hold.

- (1) P is weakly marginal symmetric $\Leftrightarrow p_1 p_2 + q_1 q_2 > 1/2$;
- (2) P is balanced $\Leftrightarrow p_1 p_2 + q_1 q_2 = 1/2$;
- (3) P is marginal symmetric $\Leftrightarrow p_1 p_2 + q_1 q_2 < 1/2$.

Proof. We prove equivalence 1. The other equivalences follow from using similar arguments.

(\Rightarrow) If P is weakly marginal symmetric, (p_1, q_1) and (p_2, q_2) are similarly arranged and we have $(p_1 - q_1)(p_2 - q_2) > 0$, or equivalently

$$p_1 p_2 + q_1 q_2 > p_1 q_2 + p_2 q_1. \quad (2)$$

Adding $p_1 p_2 + q_1 q_2$ to both sides of inequality (2) and dividing the result by 2, we obtain

$$\begin{aligned} p_1 p_2 + q_1 q_2 &> \frac{p_1 p_2 + q_1 q_2 + p_1 q_2 + p_2 q_1}{2} \\ &= \frac{(p_1 + q_1)(p_2 + q_2)}{2} = \frac{1}{2}. \end{aligned} \quad (3)$$

(\Leftarrow) If $p_1 p_2 + q_1 q_2 > 1/2$ then $(p_1 - q_1)(p_2 - q_2) > 0$ must hold using the same steps as in (\Rightarrow) backwards. Hence, (p_1, q_1) and (p_2, q_2) are similarly arranged, and it follows that P is weakly marginal symmetric. \square

4. Main Results

In this section we present several marginal dependencies of Cohen's kappa (Theorems 5, 7, and 10). The following lemma will be used repeatedly.

Lemma 4. *Coefficient (1) is strictly decreasing in $p_1p_2 + q_1q_2$.*

Proof. The first order partial derivative

$$\frac{\partial \kappa}{\partial (p_1p_2 + q_1q_2)} = \frac{a + d - 1}{(1 - p_1p_2 - q_1q_2)^2} < 0 \quad (4)$$

is negative for $a + d < 1$. \square

The following result is a slightly stronger version of a theorem in Warrens [16] for a rating scale with two categories. Theorem 5 shows that, for a fixed value of the proportion of observed agreement $a + d$, 2×2 tables that possess weak marginal symmetry produce lower values of kappa than tables that are marginal asymmetric.

Theorem 5. *Let P_1 , P_2 , and P_3 be 2×2 tables with the same value of $a + d < 1$ that are, respectively, weakly marginal symmetric, balanced, and marginal asymmetric. Furthermore, let κ_1 , κ_2 , and κ_3 denote the associated values of kappa. Then $\kappa_1 < \kappa_2 < \kappa_3$.*

Proof. Lemma 4 shows that κ is strictly decreasing in $p_1p_2 + q_1q_2$. The result then follows from application of Lemma 3. \square

Example 6 illustrates Theorem 5.

Example 6. Consider the three hypothetical 2×2 tables in Table 2.

Each table has the same proportion of observed agreement $a + d = .80$. Table 2(a) is strongly marginal symmetric, Table 2(b) is balanced, and Table 2(c) is marginal asymmetric. We have the double inequality $\kappa_1 < \kappa_2 < \kappa_3$, which illustrates Theorem 5.

Theorem 5 also considers 2×2 tables that have asymmetric marginals. However, asymmetric tables may not be relevant in practice. If the classifications are hard to make researchers will often make use of expert observers. Furthermore, novice observers usually receive some training before the actual classifications have to be made. Asymmetric tables are therefore rarely encountered in practice.

Theorem 7 shows that 2×2 tables that are symmetric may have lower kappa-values than tables with unequal marginal distributions.

Theorem 7. *Let P_1 be a weakly marginal symmetric table with marginals (p_1, q_1) and (p_2, q_2) and value κ_1 and P_2 strongly marginal symmetric with marginals (p_3, q_3) and value κ_2 . Furthermore, suppose that P_1 and P_2 have the same proportion of observed agreement $a + d < 1$. If*

$$\max\{p_1, q_1\}, \max\{p_2, q_2\} < \max\{p_3, q_3\}, \quad (5)$$

then $\kappa_1 > \kappa_2$.

TABLE 2: Three hypothetical 2×2 tables.

(a)				
	1	0	Total	
1	.60	.10	.70	$a + d = .80$
0	.10	.20	.30	$p_1p_2 + q_1q_2 = .58$
Total	.70	.30	1	$\kappa_1 = .52$
(b)				
	1	0	Total	
1	.45	.05	.50	$a + d = .80$
0	.15	.35	.50	$p_1p_2 + q_1q_2 = .50$
Total	.60	.40	1	$\kappa_2 = .60$
(c)				
	1	0	Total	
1	.40	.20	.60	$a + d = .80$
0	.00	.40	.40	$p_1p_2 + q_1q_2 = .48$
Total	.40	.60	1	$\kappa_3 = .62$

Proof. Due to the symmetries of the proportion of expected agreement $p_1p_2 + q_1q_2$, we may assume, without loss of generality, that $p_1 > q_1$, $p_2 > q_2$, and $p_3 > q_3$. It then follows from inequality (5) that $p_3 > p_1, p_2$. Furthermore, since $p_3 > q_3$, we have $p_3 > 1/2$, or $2p_3 > 1$, and since $p_1 > q_1$, we also have $2p_1 > 1$.

Since $2p_3 - 1 > 0$, multiplying both sides of inequality $p_3 > p_1$ by $2p_3 - 1$ yields

$$2p_3^2 - p_3 > 2p_1p_3 - p_1. \quad (6)$$

Adding $1 - p_3$ to both sides of inequality (6) we obtain

$$p_3^2 + 1 - 2p_3 + p_3^2 > p_1p_3 + 1 - p_1 - p_3 + p_1p_3, \quad (7)$$

or equivalently,

$$p_3^2 + (1 - p_3)^2 > p_1p_3 + (1 - p_1)(1 - p_3). \quad (8)$$

Next, since $2p_1 - 1 > 0$, multiplying both sides of the inequality $p_3 > p_2$ by $2p_1 - 1$ we obtain

$$2p_1p_3 - p_3 > 2p_1p_2 - p_2. \quad (9)$$

Adding $1 - p_1$ to both sides of inequality (9) we obtain

$$p_1p_3 + 1 - p_1 - p_3 + p_1p_3 > p_1p_2 + 1 - p_1 - p_2 + p_1p_2, \quad (10)$$

or equivalently,

$$p_1p_3 + (1 - p_1)(1 - p_3) > p_1p_2 + (1 - p_1)(1 - p_2). \quad (11)$$

Combining (8) and (11) we obtain the inequality

$$p_3^2 + (1 - p_3)^2 > p_1p_2 + (1 - p_1)(1 - p_2). \quad (12)$$

The inequality $\kappa_1 > \kappa_2$ then follows from inequality (12) and application of Lemma 4. \square

TABLE 3: Two hypothetical 2×2 tables.

(a)				
	1	0	Total	
1	.60	.05	.65	$a + d = .80$
0	.15	.20	.35	$p_1 p_2 + q_1 q_2 = .58$
Total	.75	.25	1	$\kappa_1 = .53$

(b)				
	1	0	Total	
1	.70	.10	.80	$a + d = .80$
0	.10	.10	.20	$p_3 p_3 + q_3 q_3 = .68$
Total	.80	.20	1	$\kappa_2 = .38$

Example 8 illustrates Theorem 7.

Example 8. Consider the two hypothetical 2×2 tables in Table 3.

Both tables have the same proportion of observed agreement $a + d = .80$. Table 3(a) is weakly marginal symmetric, whereas Table 3(b) is strongly marginal symmetric. Since

$$\begin{aligned} \max \{p_1, q_1\} &= \max \{.65, .35\} = .65, \\ \max \{p_2, q_2\} &= \max \{.75, .25\} = .75, \\ \max \{p_3, q_3\} &= \max \{.80, .20\} = .80, \end{aligned} \tag{13}$$

the inequalities in (5) hold. Furthermore, we have $\kappa_1 > \kappa_2$, which illustrates Theorem 7.

Using similar arguments as in the proof of Theorem 7 we may obtain the following result.

Theorem 9. *Let P_1 be a weakly marginal symmetric table with marginals (p_1, q_1) and (p_2, q_2) and κ -value κ_1 and P_2 strongly marginal symmetric with marginals (p_3, q_3) and κ -value κ_2 . Furthermore, suppose that P_1 and P_2 have the same proportion of observed agreement $a + d < 1$. If*

$$\max \{p_1, q_1\}, \max \{p_2, q_2\} > \max \{p_3, q_3\}, \tag{14}$$

then $\kappa_1 < \kappa_2$.

If we have $p_1 > q_1$, $p_2 > q_2$, and $p_3 > q_3$ then Theorems 7 and 9 cover the cases that $p_1, p_2 > p_3$ and $p_1, p_2 < p_3$. The cases $p_1 > p_3 > p_2$ and $p_2 > p_3 > p_1$ turn out to be more complicated.

Another marginal dependence of kappa is presented in Theorem 10. The theorem shows that, for a constant value of the proportion of observed agreement $a + d$, 2×2 tables that exhibit weak marginal symmetry may produce higher kappa-values than tables with strong marginal symmetry. Theorem 10 is similar to Theorem 7, but conditions (5) and (15) are different, unrelated requirements.

Theorem 10. *Let P_1 be a weakly marginal symmetric table with marginals (p_1, q_1) and (p_2, q_2) and value κ_1 and P_2 strongly marginal symmetric with marginals (p_3, q_3) and value*

κ_2 . Furthermore, suppose that P_1 and P_2 have the same proportion of observed agreement $a + d < 1$. If

$$\sqrt{\max \{p_1 p_2, q_1 q_2\}} < \max \{p_3, q_3\}, \tag{15}$$

then $\kappa_1 > \kappa_2$.

Proof. Due to the symmetries of the proportion of expected agreement $p_1 p_2 + q_1 q_2$, we may assume, without loss of generality, that $p_1 p_2 > q_1 q_2$ and $p_3 > q_3$. It then follows from inequality (15) that $p_3 > \sqrt{p_1 p_2}$. Furthermore, since P_1 is weakly marginal symmetric we must have $p_1 > q_1$ and $p_2 > q_2$. It follows that $p_1, p_2 > 1/2$, and thus $\sqrt{p_1 p_2} > 1/2$.

For $p \in (0, 1)$ the function $p \mapsto p(1 - p)$ is concave with a maximum at $p = 1/2$. Since $p_3 > \sqrt{p_1 p_2} > 1/2$, we have

$$\sqrt{p_1 p_2} (1 - \sqrt{p_1 p_2}) > p_3 (1 - p_3), \tag{16}$$

or equivalently,

$$2p_1 p_2 - 2\sqrt{p_1 p_2} < 2p_3^2 - 2p_3. \tag{17}$$

Next, since a square is nonnegative we have $(\sqrt{p_1} - \sqrt{p_2})^2 \geq 0$, or

$$-p_1 - p_2 \leq -2\sqrt{p_1 p_2}. \tag{18}$$

Adding $2p_1 p_2$ to both sides of inequality (18) we obtain

$$2p_1 p_2 - p_1 - p_2 \leq 2p_1 p_2 - 2\sqrt{p_1 p_2}. \tag{19}$$

Combining inequalities (17) and (19) we obtain

$$2p_1 p_2 - p_1 - p_2 < 2p_3^2 - 2p_3. \tag{20}$$

Adding 1 to both sides of inequality (20) we obtain

$$p_1 p_2 + 1 - p_1 - p_2 + p_1 p_2 < p_3^2 + 1 - 2p_3 + p_3^2, \tag{21}$$

or equivalently,

$$p_1 p_2 + (1 - p_1)(1 - p_2) < p_3^2 + (1 - p_3)^2. \tag{22}$$

The inequality $\kappa_1 > \kappa_2$ then follows from inequality (22) and application of Lemma 4. \square

Example 11 illustrates Theorem 10 and a corollary of Theorem 10.

Example 11. Consider the four hypothetical 2×2 tables in Table 4.

Each table has the same proportion of observed agreement $a + d = .80$. Tables 4(a) and 4(d) are strongly marginal symmetric, whereas Tables 4(b) and 4(c) are weakly marginal symmetric. For the largest marginals of Tables 4(a) and 4(b), we have $(.70)(.70) > (.60)(.80)$ and $\kappa_1 < \kappa_2$, which illustrates Theorem 10.

For many 2×2 tables from the literature the converse of Theorem 10 also holds. However, Tables 4(a) and 4(c) provide us with a counterexample that illustrates that the

TABLE 4: Four hypothetical 2×2 tables.

(a)				
	1	0	Total	
1	.60	.10	.70	$a + d = .80$
0	.10	.20	.30	$p_1 p_2 + q_1 q_2 = .58$
Total	.70	.30	1	$\kappa_1 = .52$

(b)				
	1	0	Total	
1	.60	.00	.60	$a + d = .80$
0	.20	.20	.40	$p_1 p_2 + q_1 q_2 = .56$
Total	.80	.20	1	$\kappa_2 = .55$

(c)				
	1	0	Total	
1	.61	.01	.62	$a + d = .80$
0	.19	.19	.38	$p_1 p_2 + q_1 q_2 = .57$
Total	.80	.20	1	$\kappa_3 = .53$

(d)				
	1	0	Total	
1	.50	.10	.60	$a + d = .80$
0	.10	.30	.40	$p_1 p_2 + q_1 q_2 = .52$
Total	.60	.40	1	$\kappa_4 = .58$

converse does not hold in general. We have $\kappa_1 < \kappa_3$, but not $(.70)(.70) > (.62)(.80) = .50$.

Finally, Tables 4(a) and 4(d) illustrate a special application of Theorem 10. If two 2×2 tables are strongly marginal symmetric and have the same proportion of observed agreement $a + d$, then the table with the most skewed (unbalanced) marginals (Table 4(a)) has the lowest value of kappa. This is illustrated by the fact that $\kappa_1 < \kappa_4$.

5. Conclusions

Cohen’s kappa is presently a standard tool for the analysis of agreement in a 2×2 reliability study. The statistic is frequently used as a sample statistic. Various authors have observed in this context that if two pairs of raters have the same amount of observed agreement, the pair whose marginal distributions are more similar to each other may have a lower kappa-value than the pair with more divergent marginal distributions. These properties of Cohen’s kappa have almost exclusively been demonstrated and described by means of examples of 2×2 tables [4, 8–11, 16]. In this paper we presented exact formulations and proved several marginal dependencies of this type (Theorems 5, 7, and 10). In general, they show that, for 2×2 tables with the same value of observed agreement, tables with marginal distributions that are more similar have lower associated kappa-values than tables with marginal distributions that are less similar. Each result was illustrated by an example with hypothetical 2×2 tables. The results provide a better understanding of the 2×2 kappa when it is used as a sample statistic.

Theorem 5 considers 2×2 tables that have asymmetric marginals. Although several authors have provided examples with asymmetric marginals, asymmetric tables may not be relevant in practice. If the classifications are hard to make researchers will often make use of expert observers. Furthermore, novice observers usually receive some training before the actual classifications have to be made. Asymmetric tables are therefore rarely encountered in practice.

Vach [14] emphasizes that kappa should not simply be interpreted as a measure of agreement but that Cohen’s kappa expresses the degree to which observed agreement exceeds the agreement that was expected by chance. The marginal dependencies are a direct consequence of the definition of kappa and its aim to adjust the observed agreement with respect to the expected amount of agreement under chance conditions [14, p. 659]. It is not a reason for discarding Cohen’s kappa.

Conflict of Interests

The author declares that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

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

References

- [1] M. J. Warrens, “On association coefficients for 2×2 tables and properties that do not depend on the marginal distributions,” *Psychometrika*, vol. 73, no. 4, pp. 777–789, 2008.
- [2] D. A. Bloch and H. C. Kraemer, “ 2×2 kappa coefficients: measures of agreement or association,” *Biometrics*, vol. 45, no. 1, pp. 269–287, 1989.
- [3] J. L. Fleiss, “Measuring agreement between two judges on the presence or absence of a trait,” *Biometrics*, vol. 31, no. 3, pp. 651–659, 1975.
- [4] R. L. Brennan and D. J. Prediger, “Coefficient kappa: some uses, misuses, and alternatives,” *Educational and Psychological Measurement*, vol. 41, pp. 687–699, 1981.
- [5] J. Cohen, “A coefficient of agreement for nominal scales,” *Educational and Psychological Measurement*, vol. 20, pp. 37–46, 1960.
- [6] L. M. Hsu and R. Field, “Interrater agreement measures: comments on kappa, Cohen’s kappa, Scott’s π , and Aickin’s α ,” *Understanding Statistics*, vol. 2, no. 3, pp. 205–219, 2003.
- [7] H. C. Kraemer, “Ramifications of a population model for κ as a coefficient of reliability,” *Psychometrika*, vol. 44, no. 4, pp. 461–472, 1979.
- [8] J. Sim and C. C. Wright, “The kappa statistic in reliability studies: use, interpretation, and sample size requirements,” *Physical Therapy*, vol. 85, no. 3, pp. 257–268, 2005.
- [9] A. R. Feinstein and D. V. Cicchetti, “High agreement but low kappa: I. the problems of two paradoxes,” *Journal of Clinical Epidemiology*, vol. 43, no. 6, pp. 543–549, 1990.

- [10] C. A. Lantz and E. Nebenzahl, "Behavior and interpretation of the κ statistic: resolution of the two paradoxes," *Journal of Clinical Epidemiology*, vol. 49, no. 4, pp. 431–434, 1996.
- [11] 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.
- [12] A. Von Eye and M. Von Eye, "On the marginal dependency of Cohen's κ ," *European Psychologist*, vol. 13, no. 4, pp. 305–315, 2008.
- [13] W. D. Thompson and S. D. Walter, "A reappraisal of the kappa coefficient," *Journal of Clinical Epidemiology*, vol. 41, no. 10, pp. 949–958, 1988.
- [14] W. Vach, "The dependence of Cohen's kappa on the prevalence does not matter," *Journal of Clinical Epidemiology*, vol. 58, no. 7, pp. 655–661, 2005.
- [15] T. Byrt, J. Bishop, and J. B. Carlin, "Bias, prevalence and kappa," *Journal of Clinical Epidemiology*, vol. 46, no. 5, pp. 423–429, 1993.
- [16] M. J. Warrens, "A formal proof of a paradox associated with Cohen's kappa," *Journal of Classification*, vol. 27, no. 3, pp. 322–332, 2010.
- [17] M. J. Warrens, "On similarity coefficients for 2×2 tables and correction for chance," *Psychometrika*, vol. 73, no. 3, pp. 487–502, 2008.