Kuwait J. Sci. **45 (1)** pp 89-99, 2018

Ridit and exponential type scores for estimating the kappa statistic

Ayfer E. Yilmaz[1*], Serpil Aktas[2]

[1] *Dept. of Statistics, Faculty of Science, Hacettepe University, Beytepe-Ankara, Turkey, Email: ezgiyilmaz@hacettepe.edu.tr*

[2] *Dept. of Statistics, Faculty of Science, Hacettepe University, Beytepe-Ankara, Turkey, Email: spxl@hacettepe.edu.tr*

* *Corresponding author: ezgiyilmaz@hacettepe.edu.tr*

Abstract

Cohen's kappa coefficient is a commonly used method for estimating interrater agreement for nominal and/or ordinal data; thus agreement is adjusted for that expected by chance. The weighted kappa statistic is used as an agreement index for ordinal data. The weights quantify the degree of discrepancy between the two categories. The choice of this particular set of weights affects the value of kappa. The common scores are Cicchetti-Allison and Fleiss-Cohen weights. In this article, we discuss the use of ridit type and exponential scores to compute kappa statistics in general.

Key words: Cohen's kappa; exponential scores; ordinal; ridit type scores; weights.

1. Introduction

Categorical variables, which have a measurement scale consisting of a set of categories, are important in many fields such as medical, social, and behavioral sciences. The tables that represent these variables are called contingency tables. The contingency tables in which the classificatory variables are related are referred square contingency tables. When working on these kinds of tables, analysis of agreement between row and column classifications is of interest. For nominal categories, kappa coefficient can be calculated or marginal homogeneity can be tested. In analyzing an ordered contingency table, we often study the association in such a table. To test the independence between variables, Pearson's Chi-squared test, the likelihood ratio test or the tests based on a divergence measure may be used (Saberi&Ganjali, 2013). Beside association, rater agreement can be analyzed between the ordinal variables of a square table. In the literature, different statistical methods have been proposed for analyzing rater agreement. Such as log-linear agreement models, agreement plus association models etc. (Tanner & Young, 1985; Agresti, 1988). In practice, because the coefficients of agreement summarize the degree of agreement between raters with a single number, researchers prefer using coefficient of agreements. Different statistics have been proposed in the literature, but the most popular statistic of rater agreement is the kappa statistics (Warrens, 2013).

The Cohen's kappa coefficient was suggested for analysis of agreement between the variables of square contingency tables (Cohen, 1960). Let n_{ij} denote the number of objects and n shows the total number of observers. The cell probabilities are p_{ij} and $p_{i.}$ indicates the ith row total probability, $p_{.j}$ indicates the jth column total probability of an RxR contingency table. Since the observed agreement P_0 and the proportion agreement expected by chance P_e are.

$$P_0 = \sum_{i=1}^{R} p_{ii} \tag{1}$$

and

$$P_e = \sum_{i=1}^{R} p_{i.}\, p_{.i} \tag{2}$$

then, the kappa coefficient $\hat{\kappa}$ is,

$$\hat{\kappa} = \frac{\mathrm{P_0} - \mathrm{P_e}}{1 - \mathrm{P_e}}. \tag{3}$$

Fleiss *et al.* (1969) suggested an approximate asymptotic expression for variance of $\hat{\kappa}$, which is given as following equation (Shoukri, 2004).

$$\widehat{Var}(\kappa) = \\ = \frac{1}{n(1-P_e)^2}\left(\sum_{i=1}^{R} p_{ii}\{1-(p_{i.}+p_{.i})(1-\hat{\kappa})\}^2 + (1-\hat{\kappa})^2 \sum_{i\neq j}^{R} p_{ij}(p_{i.}+p_{.i})^2 - \{\hat{\kappa} - P_e(1-\hat{\kappa})^2\}\right). \tag{4}$$

For ordinal responses, instead of kappa, weighted kappa coefficient was suggested by Cohen (1968). The coefficient allows each (i, j) cell to be weighted according to the degree of agreement between ith and jth categories (Shoukri, 2004). Since the observed agreement and the proportion agreement expected by chance are,

$$P_0 = \sum_{i=1}^{R}\sum_{j=1}^{R} w_{ij}p_{ij} \tag{5}$$

and

$$P_e = \sum_{i=1}^{R}\sum_{j=1}^{R} w_{ij}p_{i.}\,p_{.j} \tag{6}$$

then, the weighted kappa coefficient $\hat{\kappa}_w$ is,

$$\hat{\kappa}_w = \frac{P_0 - P_e}{1 - P_e} \tag{6}$$

where w_{ij} are the weight ranges $0 \leq w_{ij} \leq 1$. Let $\overline{w}_{i.}$ a weighted average of the weights in the *i*th row and $\overline{w}_{.j}$ a weighted average of the weights in the *j*th column.

$$\overline{w}_{i.} = \sum_{j=1}^{R} w_{ij}p_{.j}$$

$$\overline{w}_{.j} = \sum_{i=1}^{R} w_{ij}p_{i.} \tag{8}$$

Fleiss *et al*. (1969) suggested the approximate asymptotic expression for variance of $\hat{\kappa}_w$,

$$\widehat{Var}(\kappa_w) = \\ = \frac{\sum_{i=1}^{R}\sum_{j=1}^{R} p_{ij}\left[w_{ij} - (\overline{w}_{i.} + \overline{w}_{.j})(1-\hat{\kappa}_w)\right]^2}{n(1-P_e)^2} - \frac{\left(\hat{\kappa}_w - P_e(1-\hat{\kappa}_w)\right)^2}{n(1-P_e)^2}. \tag{9}$$

The weights indicate disagreement and are used to calculate the weighted kappa. Under this situation, selection of the appropriate weights is important. In this article, we discuss different weight matrices. Linear, quadratic, and the suggested weights are introduced in Section 2. The weights for the generated two-way square tables are discussed in Section 3, followed by Conclusion in Section 4.

2. Weights of weighted Kappa

In the presence of ordinal variables in square contingency tables, weighted kappa coefficient should be considered. Different values can be assigned for each cell of the table as the weights. Thus, the levels of agreement can change depending on the values of weights. Popular weights for weighted kappa are the linear and the quadratic weights shown in Equations (10) and (11), respectively (Cicchetti & Allison, 1971; Fleiss & Cohen, 1973).

Linear and quadratic weights of Cicchetti & Allison (1971) and Fleiss & Cohen (1973) are given in the following equations, respectively.

• Linear weights:

$$w_{ij} = 1 - |i-j|/R - 1 \tag{10}$$

• Quadratic weights:

$$w_{ij} = 1 - (i-j)^2/(R-1)^2. \tag{11}$$

It has been seen in the literature that the value of the quadratically weighted kappa is higher than the value of the linearly weighted kappa (Warrens, 2012). This result implies that the level of the agreement depends on assigned weights. This is one of the disadvantages of weighted kappa. Besides, the weighted kappa coefficient with linear and quadratic weights are used for continuous-ordinal scale data. However, in practice, as many scales are measured as ordinal. We use the information that comes from the frequency distribution of the variables to overcome this problem and discuss the ridit type and exponential score values to calculate weights.

2.1Ridit type scores for weights

Ridit type scores were suggested by Bross (1958). Cumulative probabilities are used to calculate the ridit scores. Let X be the row variable and Y be the column variable of a square contingency table. The *i*th ridit score of X; r_i^X and the *i*th ridit score of Y; r_j^Y are shown in Equation (12).

$$r_i^X = \frac{F_{i-1}^X + F_i^X}{2}$$

$$r_j^Y = \frac{F_{j-1}^Y + F_j^Y}{2} \tag{12}$$

where $\mathrm{F}_{\mathrm{i}}^{\mathrm{X}} = \sum_{k\leq i} p_{k.}$ and $\mathrm{F}_{\mathrm{j}}^{\mathrm{Y}} = \sum_{k\leq j} p_{.k}$ is the cumulative distribution function (cdf) of X and Y, respectively. Iki *et al*. (2009) adapted ridit type scores for square contingency tables. The score value is shown in Equation (13).

$$u_{ij}^r = \frac{r_i^X + r_j^Y}{2}. \tag{13}$$

The linear and quadratic weights of weighted kappa coefficient can be calculated in terms of ridit type scores as shown in Equations (14) and (15).

• Linear weights:

$$w_{ij} = 1 - \frac{\left|r_i^X - r_j^Y\right|}{u_{ij}^r (R-1)} \tag{14}$$

• Quadratic weights:

$$w_{ij} = 1 - \frac{\left(r_i^X - r_j^Y\right)^2}{(u_{ij}^r)^2 (R-1)^2}. \tag{15}$$

2.2Exponential scores for weights

Bagheban & Zayeri (2010) suggested the exponential scores when the baseline characteristic of categories changing by a geometric progression.

$$\begin{aligned} u_i &= i^a \quad for \quad i = 1,2,\ldots,R, \\ v_j &= j^b \quad for \quad j = 1,2,\ldots,C. \end{aligned} \tag{16}$$

a and b in Equation (16) are called the power parameters $(a, b > 0)$. In log-linear models, one can use the exponential scores to obtain the model with the best goodness-of-fit statistic (Bagheban & Zayeri, 2010). Different values of power parameter are applied to the model and one of them with the best goodness-of-fit statistic is selected. In weights case, we proposed a method to directly calculate the power parameters. The calculation of power parameters are shown in Equation (17).

$$a = \left[\prod_{i=1}^{R-1} \alpha_i\right]^{R-1} \tag{17a}$$

$$b = \left[\prod_{j=1}^{R-1} \beta_j\right]^{R-1} \tag{17b}$$

where $\alpha_i = F_{i+1}^X / F_i^X$ and $\beta_j = F_{j+1}^Y / F_j^Y$ where $i, j = 1,2,\ldots,R.$ .Then, the exponential score value of the *i*th row and *j*th column of a square contingency table is,

$$u_{ij}^e = \frac{i^a + j^b}{2}. \tag{18}$$

The linear and quadratic weights of weighted kappa coefficient can be calculated in terms of exponential scores are shown in Equations (19) and (20).

• Linear weights:

$$w_{ij} = 1 - \frac{|i^a - j^b|}{u_{ij}^e (R-1)} \tag{19}$$

• Quadratic weights:

$$w_{ij} = 1 - \frac{(i^a - j^b)^2}{(u_{ij}^e)^2 (R-1)^2}. \tag{20}$$

The interpretation of kappa statistic shown in Table 1 can be made to the strength of agreement (Landis & Koch, 1977).

Table 1. Interpretation of kappa statistics

Kappa Statistics	Strength of Agreement
0.00>	Poor
0.00-0.20	Slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-1.00	Almost perfect

3. Numerical examples

In this section, 3x3, 4x4, and 5x5 square contingency tables are generated randomly from multinomial distribution.

The levels of agreement between raters are taken as: slight (0.0), moderate (0.5), and almost perfect (0.9). The number of observers (*n*) is considered as 100 (Yang, 2007). In the Appendix, we give the codes for calculation of weighted kappa coefficient and its standard error for ridit linear and ridit exponential weights. All analyses were performed in MATLAB R2015a.

Example 1:

A generated 3x3 contingency table with slight $(\kappa \cong 0.0)$ agreement is,

	Rater 2			
Rater 1	1	2	3	Totals
1	9	28	8	45
2	3	5	4	12
3	3	30	10	43
Totals	15	63	22	100

The calculated weights for Example (1) are given in the following matrices. All tables which have the same number of categories have the same linear and quadratic weight matrices. Here, the linear weight and quadratic weight matrices are symmetric, and the exponential weight matrices are approximately symmetric. By contrast, the ridit weight matrices are unstructured.

$$w_l = \begin{bmatrix} 1.000 & 0.500 & 0.000 \\ 0.500 & 1.000 & 0.500 \\ 0.000 & 0.500 & 1.000 \end{bmatrix}$$

$$w_q = \begin{bmatrix} 1.000 & 0.750 & 0.000 \\ 0.750 & 1.000 & 0.750 \\ 0.000 & 07.50 & 1.000 \end{bmatrix}$$

$$w_{e_l} = \begin{bmatrix} 1.000 & 0.603 & 0.428 \\ 0.674 & 0.919 & 0.685 \\ 0.509 & 0.883 & 0.872 \end{bmatrix}$$

$$w_{e_q} = \begin{bmatrix} 1.000 & 0.843 & 0.662 \\ 0.894 & 0.994 & 0.901 \\ 0.759 & 0.986 & 0.984 \end{bmatrix}$$

$$w_{r_l} = \begin{bmatrix} 0.500 & 0.732 & 0.692 \\ 0.417 & 0.844 & 0.803 \\ 0.429 & 0.827 & 0.786 \end{bmatrix}$$

$$w_{r_q} = \begin{bmatrix} 0.750 & 0.928 & 0.905 \\ 0.660 & 0.976 & 0.961 \\ 0.673 & 0.970 & 0.954 \end{bmatrix}$$

For Example (1), the estimation values, standard errors, and absolute deviation, which is the absolute value of the difference between observed and calculated agreement are calculated and given in Table 2. The results in Table 2 show that the weighted kappa with linear ridit weights has the lowest standard error and absolute deviation.

Example 2:

A generated 3x3 contingency table with slight $(\kappa \cong 0.5)$ agreement is,

	Rater 2			
Rater 1	1	2	3	Totals
1	21	12	4	37
2	2	20	7	29
3	5	2	27	34
Totals	28	34	38	100

The calculated weights for Example (2) are given in the following matrices.

$$w_{e_l} = \begin{bmatrix} 1.000 & 0.617 & 0.435 \\ 0.679 & 0.929 & 0.702 \\ 0.517 & 0.878 & 0.887 \end{bmatrix}$$

$$w_{e_q} = \begin{bmatrix} 1.000 & 0.853 & 0.681 \\ 0.897 & 0.995 & 0.911 \\ 0.767 & 0.985 & 0.987 \end{bmatrix}$$

$$w_{r_l} = \begin{bmatrix} 0.862 & 0.747 & 0.679 \\ 0.596 & 0.969 & 0.957 \\ 0.615 & 0.992 & 0.933 \end{bmatrix}$$

$$w_{r_q} = \begin{bmatrix} 0.981 & 0.936 & 0.897 \\ 0.837 & 0.999 & 0.998 \\ 0.852 & 1.000 & 0.996 \end{bmatrix}$$

For Example (2), the estimation values, standard errors, and absolute deviation, which is the absolute value of the difference between observed and calculated agreement are calculated and given in Table 3. The results in Table 3 show that the weighted kappa with quadratic exponential and quadratic ridit weights have the lowest standard error and absolute deviation.

Example 3:

A generated 3x3 contingency table with slight $(\kappa \cong 0.9)$ agreement is,

	Rater 2			
Rater 1	1	2	3	Totals
1	21	1	2	24
2	2	40	2	44
3	1	1	30	32
Totals	24	42	34	100

The calculated weights for Example (3) are given in the following matrices.

$$w_{e_l} = \begin{bmatrix} 1.000 & 0.609 & 0.426 \\ 0.620 & 0.988 & 0.752 \\ 0.439 & 0.782 & 0.980 \end{bmatrix}$$

$$w_{e_q} = \begin{bmatrix} 1.000 & 0.847 & 0.670 \\ 0.856 & 1.000 & 0.938 \\ 0.685 & 0.952 & 1.000 \end{bmatrix}$$

$$w_{r_l} = \begin{bmatrix} 1.000 & 0.533 & 0.480 \\ 0.522 & 0.985 & 0.944 \\ 0.480 & 0.930 & 1.000 \end{bmatrix}$$

$$w_{r_q} = \begin{bmatrix} 1.000 & 0.782 & 0.730 \\ 0.771 & 1.000 & 0.997 \\ 0.730 & 0.995 & 1.000 \end{bmatrix}$$

For Example (3), the estimation values, standard errors, and absolute deviation, which is the absolute value of the difference between observed and calculated agreement are calculated and given in Table 4. The results in Table 4 show that the weighted kappa coefficient with linear weights and linear exponential weights have the lowest standard error, but the absolute deviation of the weighted kappa coefficient calculated by using the linear weights is lower.

Example 4:

A generated 4x4 contingency table with slight $(\kappa \cong 0.0)$ agreement is,

	Rater 2				
Rater 1	1	2	3	4	Totals
1	2	2	10	20	34
2	11	5	15	14	45
3	5	2	5	0	12
4	0	0	2	7	9
Totals	18	9	32	41	100

The calculated weights for Example (4) are given in the following matrices. All 4x4 tables have the same linear and quadratic weight matrices.

$$w_l = \begin{bmatrix} 1.000 & 0.667 & 0.333 & 0.000 \\ 0.667 & 1.000 & 0.667 & 0.333 \\ 0.333 & 0.667 & 1.000 & 0.667 \\ 0.000 & 0.333 & 0.667 & 1.000 \end{bmatrix}$$

$$w_q = \begin{bmatrix} 1.000 & 0.889 & 0.556 & 0.000 \\ 0.889 & 1.000 & 0.889 & 0.556 \\ 0.556 & 0.889 & 1.000 & 0.889 \\ 0.000 & 0.556 & 0.889 & 1.000 \end{bmatrix}$$

$$w_{e_l} = \begin{bmatrix} 1.000 & 0.755 & 0.636 & 0.568 \\ 0.757 & 0.997 & 0.849 & 0.752 \\ 0.640 & 0.857 & 0.995 & 0.889 \\ 0.571 & 0.760 & 0.900 & 0.994 \end{bmatrix}$$

$$w_{e_q} = \begin{bmatrix} 1.000 & 0.940 & 0.868 & 0.813 \\ 0.941 & 1.000 & 0.977 & 0.939 \\ 0.870 & 0.979 & 1.000 & 0.988 \\ 0.816 & 0.942 & 0.990 & 1.000 \end{bmatrix}$$

$$w_{r_l} = \begin{bmatrix} 0.795 & 0.923 & 0.938 & 0.757 \\ 0.581 & 0.673 & 0.789 & 0.974 \\ 0.653 & 0.762 & 0.891 & 0.918 \\ 0.949 & 0.917 & 0.785 & 0.631 \end{bmatrix}$$

$$w_{r_q} = \begin{bmatrix} 0.958 & 0.994 & 0.996 & 0.941 \\ 0.824 & 0.893 & 0.955 & 0.999 \\ 0.880 & 0.943 & 0.988 & 0.993 \\ 0.997 & 0.993 & 0.954 & 0.964 \end{bmatrix}$$

For Example (4), the estimation values, standard errors, and absolute deviation, which is the absolute value of the difference between observed and calculated agreement are calculated and given in Table 5. The results in Table 5 show that the weighted kappa with quadratic and linear exponential weights have similar standard errors, but the weighted kappa coefficient with linear exponential weights has lower absolute deviation.

Example 5:

A generated 4x4 contingency table with slight $(\kappa \cong 0.5)$ agreement is,

	Rater 2				
Rater 1	1	2	3	4	Totals
1	10	5	2	1	18
2	5	16	2	5	28
3	4	2	17	7	30
4	2	4	2	16	24
Totals	21	27	23	29	100

The calculated weights for Example (5) are given in the following matrices.

$$w_{e_l} = \begin{bmatrix} 1.000 & 0.755 & 0.636 & 0.568 \\ 0.757 & 0.997 & 0.849 & 0.752 \\ 0.640 & 0.857 & 0.995 & 0.889 \\ 0.571 & 0.760 & 0.900 & 0.994 \end{bmatrix}$$

$$w_{e_q} = \begin{bmatrix} 1.000 & 0.940 & 0.868 & 0.813 \\ 0.941 & 1.000 & 0.977 & 0.939 \\ 0.870 & 0.979 & 1.000 & 0.988 \\ 0.816 & 0.942 & 0.990 & 1.000 \end{bmatrix}$$

$$w_{r_l} = \begin{bmatrix} 0.949 & 0.697 & 0.686 & 0.676 \\ 0.751 & 0.986 & 0.972 & 0.959 \\ 0.688 & 0.937 & 0.951 & 0.964 \\ 0.707 & 0.961 & 0.974 & 0.987 \end{bmatrix}$$

$$w_{r_q} = \begin{bmatrix} 0.997 & 0.908 & 0.902 & 0.895 \\ 0.938 & 1.000 & 0.999 & 0.998 \\ 0.903 & 0.996 & 0.998 & 0.999 \\ 0.914 & 0.998 & 0.999 & 1.000 \end{bmatrix}$$

For Example (5), the estimation values, standard errors, and absolute deviation, which is the absolute value of the difference between observed and calculated agreement are calculated and given in Table 6. The results in Table 6 show that the weighted kappa coefficients with linear and linear exponential weights have similar standard errors, but the absolute deviation of kappa with linear exponential weights is lower.

Example 6:

A generated 4x4 contingency table with slight $(\kappa \cong 0.9)$ agreement is,

	Rater 2				
Rater 1	1	2	3	4	Totals
1	15	2	1	0	18
2	2	23	1	1	27
3	2	4	23	1	30
4	0	1	2	22	25
Totals	19	30	27	24	100

The calculated weights for Example (6) are given in the following matrices.

$$w_{e_l} = \begin{bmatrix} 1.000 & 0.761 & 0.645 & 0.577 \\ 0.754 & 0.992 & 0.864 & 0.768 \\ 0.636 & 0.844 & 0.987 & 0.910 \\ 0.567 & 0.747 & 0.882 & 0.984 \end{bmatrix}$$

$$w_{e_q} = \begin{bmatrix} 1.000 & 0.943 & 0.874 & 0.821 \\ 0.940 & 1.000 & 0.981 & 0.946 \\ 0.867 & 0.976 & 1.000 & 0.992 \\ 0.813 & 0.936 & 0.986 & 1.000 \end{bmatrix}$$

$$w_{r_l} = \begin{bmatrix} 0.982 & 0.692 & 0.653 & 0.681 \\ 0.729 & 0.972 & 0.922 & 0.958 \\ 0.667 & 0.950 & 1.000 & 0.963 \\ 0.676 & 0.962 & 0.988 & 0.975 \end{bmatrix}$$

$$w_{r_q} = \begin{bmatrix} 0.997 & 0.908 & 0.902 & 0.895 \\ 0.938 & 1.000 & 0.999 & 0.998 \\ 0.903 & 0.996 & 0.998 & 0.999 \\ 0.914 & 0.998 & 0.999 & 1.000 \end{bmatrix}$$

The matrix of quadratic ridit weights converges to an identity matrix. Thus, the weighted kappa coefficient with quadratic ridit weights converges to unweighted kappa statistic.

For Example (6), the estimation values, standard errors, and absolute deviation, which is the absolute value of the difference between observed and calculated agreement are calculated and given in Table 7. The results in Table 7 show that the weighted kappa coefficient with quadratic weights has lowest absolute deviation and standard error.

Example 7:

A generated 5x5 contingency table with slight $(\kappa \cong 0.0)$ agreement is,

	Rater 2					
Rater 1	1	2	3	4	5	Totals
1	3	2	0	4	2	11
2	5	7	0	8	10	30
3	0	1	1	0	0	2
4	3	5	2	10	6	26
5	6	10	1	10	4	31
Totals	17	25	4	32	22	100

The calculated weights for Example (7) are given in the following matrices. All 5x5 tables have the same linear and quadratic weight matrices.

$$w_l = \begin{bmatrix} 1.000 & 0.750 & 0.500 & 0.250 & 0.000 \\ 0.750 & 1.000 & 0.750 & 0.500 & 0.250 \\ 0.500 & 0.750 & 1.000 & 0.750 & 0.500 \\ 0.250 & 0.500 & 0.750 & 1.000 & 0.750 \\ 0.000 & 0.250 & 0.500 & 0.750 & 1.000 \end{bmatrix}$$

$$w_q = \begin{bmatrix} 1.000 & 0.938 & 0.750 & 0.250 & 0.000 \\ 0.938 & 1.000 & 0.938 & 0.750 & 0.438 \\ 0.750 & 0.938 & 1.000 & 0.938 & 0.750 \\ 0.438 & 0.750 & 0.938 & 1.000 & 0.938 \\ 0.000 & 0.438 & 0.750 & 0.938 & 1.000 \end{bmatrix}$$

$$w_{e_l} = \begin{bmatrix} 1.000 & 0.823 & 0.737 & 0.686 & 0.652 \\ 0.789 & 0.960 & 0.932 & 0.859 & 0.806 \\ 0.694 & 0.835 & 0.937 & 0.986 & 0.927 \\ 0.642 & 0.758 & 0.849 & 0.921 & 0.980 \\ 0.611 & 0.707 & 0.768 & 0.853 & 0.909 \end{bmatrix}$$

$$w_{e_q} = \begin{bmatrix} 1.000 & 0.969 & 0.931 & 0.901 & 0.879 \\ 0.956 & 0.998 & 0.995 & 0.980 & 0.962 \\ 0.906 & 0.973 & 0.996 & 1.000 & 0.995 \\ 0.872 & 0.941 & 0.977 & 0.994 & 1.000 \\ 0.848 & 0.914 & 0.954 & 0.978 & 0.992 \end{bmatrix}$$

$$w_{r_l} = \begin{bmatrix} 0.893 & 0.708 & 0.775 & 0.734 & 0.669 \\ 0.793 & 0.994 & 0.914 & 0.968 & 0.932 \\ 0.847 & 0.932 & 0.975 & 0.971 & 0.872 \\ 0.878 & 0.900 & 0.991 & 0.938 & 0.841 \\ 0.730 & 0.924 & 0.837 & 0.887 & 0.986 \end{bmatrix}$$

$$w_{r_q} = \begin{bmatrix} 0.989 & 0.914 & 0.949 & 0.929 & 0.891 \\ 0.957 & 1.000 & 0.993 & 0.999 & 0.995 \\ 0.977 & 0.995 & 0.999 & 0.999 & 0.984 \\ 0.985 & 0.990 & 1.000 & 0.996 & 0.975 \\ 0.927 & 0.994 & 0.973 & 0.987 & 1.000 \end{bmatrix}$$

For Example (7), the estimation values, standard errors, and absolute deviation, which is the absolute value of the difference between observed and calculated agreement are calculated and given in Table 8.

The results in Table 8 show that the weighted kappa coefficient with linear exponential and linear ridit weights give better results. The absolute deviation and standard errors of these kappas' are similar.

Table 2. The results of weighted kappa for Example (1)

	Classical		**Exponential**		**Ridit**	
Results	**Linear**	**Quadratic**	**Linear**	**Quadratic**	**Linear**	**Quadratic**
Estimation	0.0603	0.1299	0.0721	0.1188	0.0124	0.0289
St.Error	0.0621	0.0846	0.0529	0.0733	**0.0118**	0.0257
Abs.Deviation	0.0603	0.1299	0.0721	0.1188	**0.0124**	0.0289

Table 3. The results of weighted kappa for Example (2)

	Classical		**Exponential**		**Ridit**	
Results	**Linear**	**Quadratic**	**Linear**	**Quadratic**	**Linear**	**Quadratic**
Estimation	0.5467	0.5712	0.3766	0.4879	0.2870	0.4692
St.Error	0.0717	0.0832	0.0616	0.0772	**0.0546**	0.0769
Abs.Deviation	0.0467	0.0712	0.1234	**0.0121**	0.2130	0.0308

Table 4. The results of weighted kappa for Example (3)

	Classical		**Exponential**		**Ridit**	
Results	**Linear**	**Quadratic**	**Linear**	**Quadratic**	**Linear**	**Quadratic**
Estimation	0.8513	0.8399	0.8064	0.8353	0.8123	0.8328
St.Error	**0.0500**	0.0628	**0.0491**	0.0615	0.0588	0.0653
Abs.Deviation	**0.0487**	0.0601	0.0936	0.0647	0.0879	0.0672

Table 5. The results of weighted kappa for Example (4)

	Classical		Exponential		Ridit	
Results	**Linear**	**Quadratic**	**Linear**	**Quadratic**	**Linear**	**Quadratic**
Estimation	-0.0294	-0.0801	-0.0289	-0.0335	-0.2159	-0.3372
St.Error	0.0468	0.0594	**0.0166**	**0.0198**	0.0521	0.0730
Abs.Deviation	**0.0294**	0.0801	**0.0289**	0.0335	0.2159	0.3372

Table 6. The results of weighted kappa for Example (5)

	Classical		Exponential		Ridit	
Results	**Linear**	**Quadratic**	**Linear**	**Quadratic**	**Linear**	**Quadratic**
Estimation	0.4849	0.5259	0.4546	0.4927	0.3091	0.4036
St.Error	**0.0697**	0.0841	**0.0698**	0.0835	0.0782	0.1059
Abs.Deviation	0.0151	0.0259	0.0454	**0.0073**	0.1909	0.0964

Table 7. The results of weighted kappa for Example (6)

	Classical		Exponential		Ridit	
Results	**Linear**	**Quadratic**	**Linear**	**Quadratic**	**Linear**	**Quadratic**
Estimation	0.8121	0.8549	0.7489	0.8182	0.6590	0.7664
St.Error	0.0454	0.0426	0.0487	0.0525	0.0658	0.0788
Abs.Deviation	0.0879	**0.0451**	0.1511	0.0818	0.2410	0.1336

Table 8. The results of weighted kappa for Example (7)

	Classical		Exponential		Ridit	
Results	**Linear**	**Quadratic**	**Linear**	**Quadratic**	**Linear**	**Quadratic**
Estimation	-0.0333	-0.0665	-0.0069	-0.0106	0.0083	0.0300
St.Error	0.0756	0.1016	**0.0538**	0.0809	**0.0537**	0.0950
Abs.Deviation	0.0333	0.0665	**0.0069**	0.0106	**0.0083**	0.0300

Example 8:

A generated 5x5 contingency table with slight $(\kappa \cong 0.5)$ agreement is,

	Rater 2					
Rater 1	1	2	3	4	5	Totals
1	27	9	1	0	7	44
2	2	14	5	3	1	25
3	0	1	6	6	0	13
4	0	0	0	4	1	5
5	1	1	0	0	11	13
Totals	30	25	12	13	20	100

The calculated weights for Example (8) are given in the following matrices.

$$w_{e_l} = \begin{bmatrix} 1.000 & 0.847 & 0.768 & 0.720 & 0.687 \\ 0.901 & 0.942 & 0.854 & 0.796 & 0.755 \\ 0.846 & 0.999 & 0.909 & 0.847 & 0.803 \\ 0.809 & 0.957 & 0.950 & 0.886 & 0.839 \\ 0.782 & 0.925 & 0.982 & 0.917 & 0.869 \end{bmatrix}$$

$$w_{e_q} = \begin{bmatrix} 1.000 & 0.976 & 0.946 & 0.921 & 0.902 \\ 0.990 & 0.997 & 0.979 & 0.958 & 0.940 \\ 0.976 & 1.000 & 0.992 & 0.977 & 0.961 \\ 0.964 & 0.998 & 0.998 & 0.987 & 0.974 \\ 0.952 & 0.994 & 1.000 & 0.993 & 0.983 \end{bmatrix}$$

$$w_{r_l} = \begin{bmatrix} 0.905 & 0.944 & 0.957 & 0.862 & 0.929 \\ 0.803 & 0.994 & 0.849 & 0.766 & 0.824 \\ 0.941 & 0.909 & 0.993 & 0.897 & 0.965 \\ 0.875 & 0.747 & 0.827 & 0.919 & 0.853 \\ 0.875 & 0.747 & 0.927 & 0.919 & 0.853 \end{bmatrix}$$

$$w_{r_q} = \begin{bmatrix} 0.991 & 0.997 & 0.998 & 0.981 & 0.995 \\ 0.961 & 0.997 & 0.977 & 0.945 & 0.969 \\ 0.997 & 0.992 & 1.000 & 0.989 & 0.999 \\ 0.984 & 0.936 & 0.970 & 0.993 & 0.978 \\ 0.984 & 0.936 & 0.970 & 0.993 & 0.978 \end{bmatrix}$$

For Example (8), the estimation values, standard errors, and absolute deviation, which is the absolute value of the difference between observed and calculated agreement are calculated and given in Table 9. The results in Table 9 show that the weighted kappa with quadratic exponential weights has lowest absolute deviation.

Example 9:

A generated 5x5 contingency table with slight $(\kappa \cong 0.9)$ agreement is,

	Rater 2					
Rater 1	1	2	3	4	5	Totals
1	7	2	0	0	2	11
2	1	17	2	0	1	21
3	0	1	13	0	0	14
4	3	1	2	19	1	26
5	2	1	1	2	22	28
Totals	13	22	18	21	26	100

The calculated weights for Example (9) are given in the following matrices.

$$w_{e_l} = \begin{bmatrix} 1.000 & 0.805 & 0.713 & 0.661 & 0.629 \\ 0.794 & 0.987 & 0.894 & 0.816 & 0.761 \\ 0.700 & 0.963 & 0.980 & 0.935 & 0.871 \\ 0.648 & 0.784 & 0.891 & 0.974 & 0.959 \\ 0.616 & 0.730 & 0.826 & 0.905 & 0.970 \end{bmatrix}$$

$$w_{e_q} = \begin{bmatrix} 1.000 & 0.962 & 0.918 & 0.885 & 0.862 \\ 0.958 & 1.000 & 0.989 & 0.966 & 0.943 \\ 0.910 & 0.981 & 1.000 & 0.996 & 0.983 \\ 0.876 & 0.953 & 0.988 & 0.999 & 0.998 \\ 0.852 & 0.927 & 0.970 & 0.991 & 0.999 \end{bmatrix}$$

$$w_{r_l} = \begin{bmatrix} 0.958 & 0.739 & 0.716 & 0.720 & 0.690 \\ 0.789 & 0.978 & 0.944 & 0.951 & 0.905 \\ 0.771 & 1.000 & 0.967 & 0.973 & 0.927 \\ 0.745 & 0.967 & 1.000 & 0.994 & 0.960 \\ 0.694 & 0.893 & 0.926 & 0.919 & 0.965 \end{bmatrix}$$

$$w_{r_q} = \begin{bmatrix} 0.998 & 0.932 & 0.919 & 0.922 & 0.904 \\ 0.955 & 0.999 & 0.997 & 0.998 & 0.991 \\ 0.947 & 1.000 & 0.999 & 0.999 & 0.995 \\ 0.935 & 0.999 & 1.000 & 1.000 & 0.998 \\ 0.906 & 0.989 & 0.994 & 0.993 & 0.999 \end{bmatrix}$$

For Example (9), the estimation values, standard errors, and absolute deviation, which is the absolute value of the difference between observed and calculated agreement are calculated and given in Table 10. The results in Table 10 show that the weighted kappa with linear weights has lowest absolute deviation. The deviations of Example (9) are higher than the deviations of Example (7) and (8).

4. Conclusions

The agreement between objects rated independently by two raters (or twice by the same rater) is investigated with the agreement coefficients. There are different agreement coefficients for different scale types. For ordinal categories, weighted kappa coefficient (Cohen, 1968) is the most used agreement coefficient. The weighted kappa coefficient is a generalization of the simple kappa coefficient that uses weights to measure the relative difference between the row and column categories. Popular weights for weighted kappa are the linear and the quadratic weights and standard statistical packages mostly use these weights. The quadratically and linearly weighted kappas are used for continuous-ordinal scale data. However, in practice, many scales are dichotomous ordinal. In this case, we suggested using of the ridit type and exponential scores to compute the kappa statistic.

According to the two lowest absolute deviations (italic) and the lowest standard errors (bold) of Examples (1)-(9), the best method to calculate the weights are summarized in Table 11.

Results confirm that, if there is a slight agreement between raters, ridit weights for 3x3 and linear exponential weights for 4x4 and 5x5 tables give better results. For the 3x3 and 5x5 tables the weighted kappa, which is calculated from the quadratic weights is tend to give a higher agreement than the observed agreement. When there is a moderate agreement, using the quadratic exponential weights for all the table dimensions have lower absolute deviation. Linear ridit weights for 3x3 and 5x5 tables, and Cicchetti & Allison (1971)'s linear weights for 4x4 table have lower standard errors. Differently from the slight and moderate agreement, in case of almost perfect agreement, the weighted kappa coefficients with classical linear and quadratic weights have lower absolute deviation and tend to give a higher agreement. Linear exponential weights for 3x3 and 5x5 tables, and Fleiss & Cohen (1973)'s quadratic weights for 4x4 table have lower standard errors.The results also show that ridit type scores are insufficient for $R \geq 4$.

Note that these results are based upon the several numerical examples. The assignment of weights for a weighted kappa is a subjective issue.

Table 9. The results of weighted kappa for Example (8)

	Classical		Exponential		Ridit	
Results	**Linear**	**Quadratic**	**Linear**	**Quadratic**	**Linear**	**Quadratic**
Estimation	0.5695	0.5806	0.4256	0.5248	0.1353	0.2741
St.Error	0.0684	0.0917	0.0528	0.0729	**0.0355**	0.0651
Abs.Deviation	0.0695	0.0806	0.0744	**0.0248**	0.3647	0.2258

Table 10. The results of weighted kappa for Example (9)

	Classical		Exponential		Ridit	
Results	**Linear**	**Quadratic**	**Linear**	**Quadratic**	**Linear**	**Quadratic**
Estimation	0.7040	0.6654	0.5889	0.6319	0.4374	0.5251
St.Error	0.0640	0.0897	**0.0597**	0.0862	0.0726	0.1160
Abs.Deviation	**0.1960**	0.2346	0.3111	0.2681	0.4626	0.3749

Table 11. The summary of weights in terms of dimension of table and level of agreement with regard to the absolute deviations (italic) and standard errors (bold)

κ	3×3	4×4	5×5
0.0	*Ridit$_L$;Ridit$_Q$* **Ridit$_L$;Ridit$_Q$**	*Exponential$_L$; Classic$_L$* **Exponential$_L$; Exponential$_Q$**	*Exponential$_L$; Ridit$_L$* **Ridit$_L$; Exponential$_L$**
0.5	*Exponential$_Q$; Ridit$_Q$* **Ridit$_L$; Exponential$_L$**	*Exponential$_Q$; Classic$_L$* **Classic$_L$; Exponential$_L$**	*Exponential$_Q$; Classic$_L$* **Ridit$_L$; Exponential$_L$**
0.9	*Classic$_L$; Classic$_Q$* **Exponential$_L$; Classic$_L$**	*Classic$_Q$; Classic$_L$* **Classic$_Q$; Classic$_L$**	*Classic$_L$; Classic$_Q$* **Exponential$_L$; Classic$_L$**

L:Linear, Q:Quadratic

Appendix

```
f=[9 28 8;3 5 4;3 30 10]; %frequencies matrix
R=3;  %number of categories
n=sum(f(:)); %sample size
F_y=sum(f,1)/n;
F_x=sum(f,2)/n';
p=f/n;

%Calculation of weighted kappa with ridit linear and ridit quadratic weights
r_x(1)=n*F_x(1)/2;
r_y(1)=n*F_y(1)/2;
for i=2:R
r_x(i)=n*(F_x(i-1)+F_x(i))/2;
r_y(i)=n*(F_y(i-1)+F_y(i))/2;
end
Pro_l=0; Pre_l=0;Pro_q=0; Pre_q=0;
for i=1:R
for j=1:R
wr_l(i,j)=1-abs(r_x(i)-r_y(j))/((r_x(i)+r_y(j))/2*(R-1));
Pro_l=Pro_l+wr_l(i,j)*p(i,j);
Pre_l=Pre_l+wr_l(i,j)*F_x(i)*F_y(j);
wr_q(i,j)=1-(r_x(i)-r_y(j))^2/((r_x(i)+r_y(j))/2*(R-1))^2;
Pro_q=Pro_q+wr_q(i,j)*p(i,j);
Pre_q=Pre_q+wr_q(i,j)*F_x(i)*F_y(j);
end
end
wr_l1=zeros(1,R);wr_l2=zeros(1,R);wr_q1=zeros(1,R);wr_q2=zeros(1,R);
for i=1:R
for j=1:R
wr_l1(i)=wr_l1(i)+wr_l(i,j)*F_y(j);
wr_l2(j)=wr_l2(j)+wr_l(i,j)*F_x(i);
wr_q1(i)=wr_q1(i)+wr_q(i,j)*F_y(j);
wr_q2(j)=wr_q2(j)+wr_q(i,j)*F_x(i);
end
end
Kw_rl=(Pro_l-Pre_l)/(1-Pre_l);%Weighted kappa coefficient with ridit linear weights
Kw_rq=(Pro_q-Pre_q)/(1-Pre_q);%Weighted kappa coefficient with ridit linear weights
Mr_l=0;Mr_q=0;
for i=1:R
for j=1:R
Mr_l=Mr_l+p(i,j)*[wr_l(i,j)-(wr_l1(i)+wr_l2(j))*(1-Kw_rl)]^2;
Mr_q=Mr_q+p(i,j)*[wr_q(i,j)-(wr_q1(i)+wr_q2(j))*(1-Kw_rq)]^2;
end
end
Nr_l=Mr_l-(Kw_rl-Pre_l*(1-Kw_rl))^2;
Nr_q=Mr_q-(Kw_rq-Pre_q*(1-Kw_rq))^2;
V_wrl=Nr_l/(n*(1-Pre_l)^2); %The variance of weighted kappa coefficient with ridit linear weights
V_wrq=Nr_q/(n*(1-Pre_q)^2); %The variance of weighted kappa coefficient with ridit quadratic weights
Kappaw_r=[Kw_rl sqrt(V_wrl);Kw_rq sqrt(V_wrq)]'
```

References

Agresti, A. (1988). A model for agreement between ratings on an ordinal scale. Biometrics, **44**(2):539–548.

Bagheban, A.A. & Zayeri, F. (2010). A generalization of the uniform association model for assessing rater agreement in ordinal scales. Journal of Applied Statistics, **37**(8): 1265–1273, DOI: 10.108002664760903012666/.

Bross, I.D.J. (1958). How to us eridit analysis. Biometrics, **14**:18–38.

Cicchetti, D. & Allison, T. (1971). A new procedure for assessing reliability of scoring eeg sleep recordings. American Journal EEG Technology, **11**:101–109.

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, **20**(1):37-46.

Cohen, J. (1968). Weighted Kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. Psychological Bulletin, **70**(4): 213 -220.

Fleiss, J.L. & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measure of reliability. Educational and Psychological Measurement, **33**: 613–619.

Fleiss, J.L., Cohen, J. & Everitt, B.S. (1969). Large sample standard errors of kappa and weighted kappa. Psychological Bulletin, **72**:323–327.

Iki, K., Tahata, K. & Tomizawa, S. (2009). Ridit score type quasi-symmetry and decomposition of symmetry for square contingency tables with ordered categories. Austrian Journal of Statistics, **38**(3): 183–192.

Landis, J.R. & Koch, G.G. (1977). The measurement of observed agreement for categorical data. Biometrics, **33**(1):159–174.

Saberi, Z. & Ganjali, M.A.(2013). A conditional Bayesian approach for testing independence in two-way contingency tables. Kuwait Journal of Science,**40**(2):103- 113.

Shoukri, M.M. (2004). Measures of Interobserver Agreement. Chapman & Hall/CRC, Florida.

Tanner, M.A. & Young M.A. (1985). Modeling agreement among raters. Journal of the American Statistical Association, **80**(389):175– 180.

Warrens, M. J. (2012). Cohen's quadratically weighted kappa is higher than linearly weighted kappa for tridiagonal agreement tables. Statistical Methodology, **9**:440 -444, DOI: 10.1016/j.stamet. 2011.08.006.

Warrens, M. J. (2013). Weighted kappas for 3×3 tables. Journal of Probability and Statistics, DOI: 10.1155325831/2013/.

Yang, J. (2007). Measure of agreement for categorical data. Ph.D. thesis, The Pennsylvania State University, The Graduate School, Department of Statistics, USA.

Submitted: 19/10/2015
Revised : 15/02/2016
Accepted : 07/03/2016

سجلات من نوع ريديت
والنوع الأسي لتقدير إحصاء كابا

ايفر يلماز[1،*] ، سيربيل أكتاز[2]

[1]قسم الإحصاء، كلية العلوم، جامعة هاستيب، بايتيب – أنقرة، تركيا

ezgiyilmaz@hacettepe.edu.tr

[2]قسم الإحصاء، كلية العلوم، جامعة هاستيب، بايتيب – أنقرة، تركيا

spxl@hacettepe.edu.tr

* ezgiyilmaz@hacettepe.edu.tr

خلاصة

يُستخدم معامل كابا لكوهين بكثرة لتقدير درجة توافق التقييم للبيانات الأسمية و / أو بيانات الرُتب. وبذلك يتم تعديل درجة التوافق آخذين في الاعتبار التوافق المتوقع نتيجة الصدفة. يُستخدم إحصاء كابا الموزون كمؤشر في بيانات الرُتب. تقيس الأوزان درجة الاختلافات بين القطاعين. تتأثر قيمة كابا باختيار الأوزان. الأوزان المُستخدمة بكثرة هي أوزان Cicchetti-Allison وأوزان Fleiss-Cohen. في هذا البحث نناقش استخدام أوزان من نوع ريديت ومن النوع الأسي عند حساب إحصاء كابا.