Hacettepe Journal of Mathematics and Statistics
Volume 39 (1) (2010), 97 – 107

AN ALTERNATIVE AGREEMENT STATISTICS WITH LINEAR WEIGHT BETWEEN ORDINAL CATEGORICAL MEASUREMENTS

Mustafa Semiz*† and Bengü Ocak‡

Received 17 : 02 : 2009 : Accepted 16 : 11 : 2009

Abstract

Accurate and precise measurement is an important issue in any study and in any scientific area. Weighted kappa, proposed by J. Cohen (*Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit*, Psychological Bulletin **70**, 213–220, 1968) is the most common and widely preferred coefficient for measuring agreement between two ordinally measured categorical variables. This article presents an alternative agreement coefficient between ordinal categorical measurements. The proposed coefficient takes values between 0 and 1. Therefore, the interpretation and the calculation of the proposed coefficient are also very simple. An SPSS Syntax program for the proposed coefficient and the weighted kappa is presented.

Keywords: Weighted kappa, Rater agreement, Proportion of exact agreement, Gamma statistics.

2000 AMS Classification: 62 F 10.

*Department of Statistics, Faculty of Sciences, Selcuk University, Konya, Turkey.
E-mail: `msemiz@selcuk.edu.tr`
†Corresponding Author.
‡Institute of Science and Technology, Selcuk University, Konya, Turkey.
E-mail: `benguocak@gmail.com`

1. Introduction

Agreement coefficients are needed to instrument or assay validation, method comparisons, statistical process control, goodness of fit, individual bioequivalance or the acceptability of a new or generic process, methodology, and formulation in many areas. Agreement between two different methods, graders or raters of ordered categorical measures is an important subject in any field of science.

The most common agreement measures for categorical nominal and ordinal outcomes are Cohen's kappa and the weighted kappa [5, 9]. If the outcome variable is ordered categorical, the weighted kappa is one of the most commonly used measure of agreement [6].

Assume that two raters assign each of n measures to one of I different categories. Let π_{ij} $(= n_{ij}/n)$ denote the $(i,j)^{\text{th}}$ cell joint probability of two ratings with $\pi_{i+} = \sum_{j=1}^{I} \pi_{ij}$ and $\pi_{+j} = \sum_{i=1}^{I} \pi_{ij}$. Let n_{ij} denote the frequency with which the first and second rater assigned targets to categories i and j, respectively. The weighted kappa is written as follows:

$$\kappa_w = \frac{\sum\limits_{i=1}^{I}\sum\limits_{j=1}^{J} w_{ij}\pi_{ij} - \sum\limits_{i=1}^{I}\sum\limits_{j=1}^{J} w_{ij}\pi_{i+}\pi_{+j}}{1 - \sum\limits_{i=1}^{I}\sum\limits_{j=1}^{J} w_{ij}\pi_{i+}\pi_{+j}},$$

where w_{ij} are the disagreement weights. Naturally, $w_{ij} = 0$ is selected for cells for which raters agree, and $w_{ij} > 0$ if $i \neq j$, that is, raters show disagreement. The weighting scheme based on table scores can be either a quadratic weight, $w_{ij} = 1 - (i-j)^2/(I-1)^2$, or a linear weight $w_{ij} = 1- \mid i-j \mid /(I-1)$ [9], [4].

The following are standards for strength of agreement for the kappa coefficient [12]:

≤ 0: poor, 0.01 - 0.20 : slight, 0.21 - 0.40 : fair, 0.41 - 0.60 : moderate,
0.61 - 0.80 : substantial, 0.81 - 1.00 : almost perfect.

Though kappa also has limitations, it is very important because it is the most widely used measure of interjudge reliability across the scientific literature. Kappa explicitly recognizes the likelihood of chance agreement between judges, and removes it from consideration [16]. Also, Brennan and Prediger [3] provide useful technical reviews of the problems and the limitations of kappa.

As an example, the following data shows high agreement, the weighted kappa can calculate negative and near zero values as seen in Table 1.

Table 1. 10 units rated twice

		Second rating		
		$X_1 = 1$	$X_2 = 2$	$X_3 = 3$
First rating	$Y_1 = 1$	1	0	1
	$Y_2 = 2$	0	7	0
	$Y_3 = 3$	1	0	0

In this case, the weighted kappa is 0.20 with linear weighting. However, the proportion of exact agreement $(\sum_{i=1}^{3} \pi_{ii})$ is 0.80. Therefore, it is not easy to make an inference on these kinds of result.

In addition, the value of kappa is affected by factors such as the weighting applied and the number of categories in the measurement scale. The larger the number of scale categories, the greater the potential for disagreement [15]. Dunn (1989) suggested that interpretation of kappa is assisted by also reporting the maximum value it could attain for the set of data concerned.

2. The similarity measure (s_l) for ordinal categorical agreement

Let us assume that a point of observation with pairs of samples (y_j, x_j), $j = 1, 2, \ldots, n$, is coming from two ratings from level I of ordinally scaled categories ($X_1 = Y_1 < X_2 = Y_2 < \cdots < X_I = Y_I$). As seen in Table 1., n data points are located in a $I \times I$ contingency table. For measuring the agreement with a similarity coefficient, the total disagreement of n points with linear distances ($\sum_{j=1}^{n} |y_j - x_j|$), and the possible maximum linear disagreement ($|Y_I - Y_1| = |X_I - X_1|$) are taken into consideration. Therefore, the similarity measure with linear weight is as proposed below:

$$\begin{aligned} s_l &= 1 - \frac{d_o}{d_t} \\ &= 1 - \frac{\sum_{j=1}^{n} |y_j - x_j|}{n|Y_I - Y_1|} \\ &= 1 - \frac{\sum_{j=1}^{n} |y_j - x_j|}{n|X_I - X_1|}, \end{aligned}$$

where d_o is the total observed disagreements with n points of measurement, and d_t is the total possible maximum disagreement with n points. The similarity measure for agreement in Table 1 is

$$\begin{aligned} s_l &= 1 - \frac{d_o}{d_t} \\ &= 1 - \frac{\sum_{j=1}^{n} |y_j - x_j|}{n|Y_I - Y_1|} \\ &= 1 - \frac{1 \times |1-1| + 1 \times |1-3| + 7|2-2| + 1 \times |3-1|}{10|3-1|} = 0.80. \end{aligned}$$

The uniform distribution of linear disagreements between measurements in a $I \times I$ contingency table depends upon the number of levels in the categorical variables, and the values of those levels. For our example, Table 2 shows the distributions of linear disagreements.

Table 2. The uniform distribution of linear disagreements $(i, \imath' = 1, 2, 3)$

		Second rating		
$\mid Y_i - X_{\imath'} \mid$		$X_1 = 1$	$X_2 = 2$	$X_3 = 3$
First rating	$Y_1 = 1$	0	1	2
	$Y_2 = 2$	1	0	1
	$Y_3 = 3$	2	1	0

In this case, s_l can also be shown in matrix form:

$$s_l = 1 - \frac{d' f}{n[\max(d'_k)]},$$

where d' is the $1 \times k$ vector of k distinct linear disagreement values, f is the $k \times 1$ vector of observed frequencies of the k distinct linear disagreement values, $\max(d')$ is the maximum value in the vector d', and as usual we identify the 1×1 matrix $d'f$ with its single entry.

Therefore, for our example, three linear disagreements ($d_k = 0, 1, 2$ for $k = 1, 2, 3$) are present, the expected frequencies under the uniform distributions are $F_k = 3, 4, 2$ for $k = 1, 2, 3$ in Table 2, and the observed frequencies of these disagreement values are $f_k = 8, 0, 2$ for $k = 1, 2, 3$, respectively. In matrix form, s_l is as follows:

$$\begin{aligned} s_l &= 1 - \frac{d'f}{n\left[\max(d'_k)\right]} \\ &= 1 - \frac{\begin{pmatrix} 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 8 \\ 0 \\ 2 \end{pmatrix}}{10 \times 2} \\ &= 1 - \frac{4}{20} = 0.800 \end{aligned}$$

where $\sum_{l=1}^{k} f_l = n$.

The distribution of the frequencies of k distinct linear disagreements $F' = [F_1 \ \cdots \ F_k]$ in an $I \times I$ table have multinomial distribution for $\sum_{l=1}^{k} F_l = I \times I = N$. If N points are measured then the distribution of the frequencies of k distinct linear disagreements is a multinomial distribution as follows:

$$f(F_1 = f_1, \ldots, F_k = f_k) = \frac{N!}{f_1! \times \cdots \times f_k!} \pi_1^{f_1} \times \cdots \times \pi_k^{f_k},$$
$$\sum_{l=1}^{k} \pi_l = 1 \text{ and } \sum_{l=1}^{k} f_l = N,$$

where π_l is the expected ratio of the frequency of the l^{th} linear disagreement ($l = 1, \ldots, k$). Therefore, for Table 2, the random vector $F' = [F_1 \ \cdots \ F_k]$ has the mean vector and covariance matrix

$$\mu = \begin{pmatrix} N\pi_1 \\ N\pi_2 \\ \vdots \\ N\pi_k \end{pmatrix} \text{ and } \Sigma = \begin{pmatrix} N\pi_1(1-\pi_1) & -N\pi_1\pi_2 & \cdots & -N\pi_1\pi_k \\ -N\pi_2\pi_1 & N\pi_2(1-\pi_2) & \cdots & -N\pi_2\pi_k \\ \vdots & \vdots & \ddots & \vdots \\ -N\pi_k\pi_1 & -N\pi_k\pi_2 & \cdots & N\pi_k(1-\pi_k) \end{pmatrix},$$

respectively. If n points are measured, then the observed distribution of the frequencies of k distinct linear disagreements is a multinomial distribution as follows:

$$\widehat{\mu} = \begin{pmatrix} np_1 \\ np_2 \\ \vdots \\ np_k \end{pmatrix} \text{ and } \widehat{\Sigma} = \begin{pmatrix} np_1(1-p_1) & -np_1p_2 & \cdots & -np_1p_k \\ -np_2p_1 & np_2(1-p_2) & \cdots & -np_2p_k \\ \vdots & \vdots & \ddots & \vdots \\ -np_kp_1 & -np_kp_2 & \cdots & np_k(1-p_k) \end{pmatrix}.$$

Hence, the expected value and the variance of s_l are respectively

$$\text{(2.1)} \qquad \mathrm{E}\left[s_l\right] = 1 - \frac{d'\mu}{n\left[\max(d'_k)\right]} = 1 - \frac{\sum\limits_{l=1}^{k} d_l \pi_l}{\left[\max(d'_k)\right]} = S_l$$

and

$$\text{(2.2)} \qquad \mathrm{V}\left[s_l\right] = \frac{d'\Sigma d}{n^2\left[\max\left(d'_k\right)\right]^2}.$$

As seen in (2.1), s_l is an unbiased estimator of S_l. In addition, s_l is also a consistent statistics. By equation (2.2),

$$
\begin{aligned}
\mathrm{V}\left[s_l\right] &= \frac{d'\Sigma d}{n^2\left[\max\left(d'_k\right)\right]^2} \\
&= \frac{n\left[\sum_{j=1}^{k} d_j^2\pi_j(1-\pi_j) - \sum_{j=1}^{k}\sum_{\substack{i=1\\ i\neq j}}^{k} d_j d_i \pi_j \pi_i\right]}{n^2\left[\max\left(d'_k\right)\right]^2} \\
&= \frac{\left[\sum_{j=1}^{k} d_j^2\pi_j(1-\pi_j) - \sum_{j=1}^{k}\sum_{\substack{i=1\\ i\neq j}}^{k} d_j d_i \pi_j \pi_i\right]}{n\left[\max\left(d'_k\right)\right]^2} \xrightarrow{n\to\infty} 0.
\end{aligned}
$$

For the example in Table 1, the expected value of s_l is

$$
\begin{aligned}
\mathrm{E}\left[s_l\right] &= 1 - \frac{d'\mu}{n\left[\max\left(d'_k\right)\right]} \\
&= 1 - \frac{\begin{pmatrix}0 & 1 & 2\end{pmatrix}\begin{pmatrix}10\left(\frac{3}{9}\right) & 10\left(\frac{4}{9}\right) & 10\left(\frac{2}{9}\right)\end{pmatrix}'}{10\,[2]} \\
&= 1 - \frac{40/9 + 40/9}{20} \\
&= 0.556
\end{aligned}
$$

and the variance of s_l is

$$
\begin{aligned}
\mathrm{V}\left[s_l\right] &= \frac{d'\Sigma d}{n^2\left[\max\left(d'_k\right)\right]^2} \\
&= \frac{\begin{pmatrix}0 & 1 & 2\end{pmatrix}\begin{pmatrix} 10\left(\frac{3}{9}\right)\left(\frac{6}{9}\right) & -10\left(\frac{3}{9}\right)\left(\frac{4}{9}\right) & -10\left(\frac{3}{9}\right)\left(\frac{2}{9}\right) \\ -10\left(\frac{4}{9}\right)\left(\frac{3}{9}\right) & 10\left(\frac{4}{9}\right)\left(\frac{5}{9}\right) & -10\left(\frac{4}{9}\right)\left(\frac{2}{9}\right) \\ -10\left(\frac{2}{9}\right)\left(\frac{3}{9}\right) & -10\left(\frac{2}{9}\right)\left(\frac{4}{9}\right) & 10\left(\frac{2}{9}\right)\left(\frac{7}{9}\right)\end{pmatrix}\begin{pmatrix}0\\1\\2\end{pmatrix}}{10^2 \times 2^2} \\
&= 0.0135.
\end{aligned}
$$

Consequently, the standard error of s_l is 0.116.

Similarly, the s_l statistics and its estimated variance may be calculated as follows:

$$
\begin{aligned}
s_l &= 1 - \frac{d'\hat{\mu}}{n\left[\max\left(d'_k\right)\right]} \\
&= 1 - \frac{\begin{pmatrix}0 & 1 & 2\end{pmatrix}\begin{pmatrix}10\left(\frac{8}{10}\right) & 10\left(\frac{0}{10}\right) & 10\left(\frac{2}{10}\right)\end{pmatrix}'}{10 \times 2} \\
&= 1 - \frac{4}{20} = 0.80,
\end{aligned}
$$

and the estimate of the variance of s_l is

$$
\begin{aligned}
\widehat{V}\left[s_l\right] &= \frac{d'\hat{\Sigma} d}{n^2\left[\max\left(d'_k\right)\right]^2} \\
&= \frac{\begin{pmatrix}0 & 1 & 2\end{pmatrix}\begin{pmatrix} 10\left(\frac{8}{10}\right)\left(\frac{2}{10}\right) & -10\left(\frac{8}{10}\right)\left(\frac{0}{10}\right) & -10\left(\frac{8}{10}\right)\left(\frac{2}{10}\right) \\ -10\left(\frac{0}{10}\right)\left(\frac{8}{10}\right) & 10\left(\frac{0}{10}\right)\left(\frac{10}{10}\right) & -10\left(\frac{0}{10}\right)\left(\frac{2}{10}\right) \\ -10\left(\frac{2}{10}\right)\left(\frac{8}{10}\right) & -10\left(\frac{2}{10}\right)\left(\frac{0}{10}\right) & 10\left(\frac{2}{10}\right)\left(\frac{8}{10}\right)\end{pmatrix}\begin{pmatrix}0\\1\\2\end{pmatrix}}{10^2 \times 2^2} \\
&= 0.016.
\end{aligned}
$$

Also, the estimate of the standard error of s_l is 0.126.

3. A simulation study

In order to show the properties and the distribution of the similarity measurement, and to make comparisons between the similarity measurements and the weighted kappa coefficients, a Monte Carlo simulation was performed for 3 different cases, which are uniform, twice weighted main diagonal and twice weighted reverse diagonal distributions in 3×3 and 4×4 cross tables. The tables were randomly generated with sample sizes of 10, 30 and 50 for each one of the 3 distributions. Finally, the number of repetitions performed for each of the settings are given in Table 3 and Table 4.

Table 3. Monte Carlo simulation with 10000 replications in a 3×3 cross table

			Expected value		Estimation		MSE	
Distribution in table	n	Proportion of exact agreement	Weighted kappa	s_l	Weighted kappa	s_l	Weighted kappa	s_l
Case I $\pi_{ij} = \frac{1}{9}$	10	0.333	0.000	0.556	0.002	0.556	0.055	0.013
	30	0.333	0.000	0.556	-0.001	0.555	0.020	0.005
	50	0.333	0.000	0.556	0.001	0.556	0.012	0.003
Case II	10	0.500	0.250	0.667	0.227	0.665	0.116	0.026
$\pi_{ii} = \frac{2}{12}$	30	0.500	0.250	0.667	0.244	0.666	0.081	0.017
otherwise $\pi_{ij} = \frac{1}{12}$	50	0.500	0.250	0.667	0.247	0.667	0.074	0.015
Case III	10	0.333	-0.125	0.500	-0.105	0.501	0.068	0.020
$\pi_{13}, \pi_{22}, \pi_{31} = \frac{2}{12}$	30	0.333	-0.125	0.500	-0.119	0.500	0.035	0.009
otherwise $\pi_{ij} = \frac{1}{12}$	50	0.333	-0.125	0.500	-0.121	0.500	0.027	0.000

Table 4. Monte Carlo simulation with 10000 replications in a 4×4 cross table

			Expected value		Estimation		MSE	
Distribution in table	n	Proportion of exact agreement	Weighted kappa	s_l	Weighted kappa	s_l	Weighted kappa	s_l
Case I $\pi_{ij} = \frac{1}{16}$	10	0.250	0.000	0.583	-0.004	0.582	0.047	0.011
	30	0.250	0.000	0.583	0.001	0.583	0.017	0.004
	50	0.250	0.000	0.583	-0.001	0.583	0.010	0.002
Case II	10	0.400	0.200	0.667	0.183	0.666	0.089	0.018
$\pi_{ii} = \frac{2}{20}$	30	0.400	0.200	0.667	0.193	0.666	0.058	0.011
otherwise $\pi_{ij} = \frac{1}{20}$	50	0.400	0.200	0.667	0.197	0.667	0.051	0.009
Case III	10	0.200	-0.120	0.533	-0.106	0.533	0.054	0.014
$\pi_{14}, \pi_{23}, \pi_{32}, \pi_{41} = \frac{2}{20}$	30	0.200	-0.120	0.533	-0.115	0.533	0.029	0.006
otherwise $\pi_{ij} = \frac{1}{20}$	50	0.200	-0.120	0.533	-0.117	0.533	0.023	0.005

In both tables it can be seen that s_l is both an unbiased and a consistent estimator of S_l. The estimate value of s_l converges to the expected value as the sample size increases in each case. For sample size 50, the estimated values are almost the same as the expected values of the similarity measurement. In addition, the MSE value of s_l decreases rapidly as the sample size increases in each case. However, for a good estimation weighted kappa needs much larger sample sizes, and its MSE values are much greater than the MSE values of s_l. Also, s_l is more consistent than the weighted kappa statistics. As seen in

Figure 1 and Figure 2, the histograms of the values of s_l almost fit the normal curve. The distributions of s_l are normally distributed according to the Kolmogorov-Smirnov test.

Figure 1. Case II in a 3×3 cross table with sample size 30

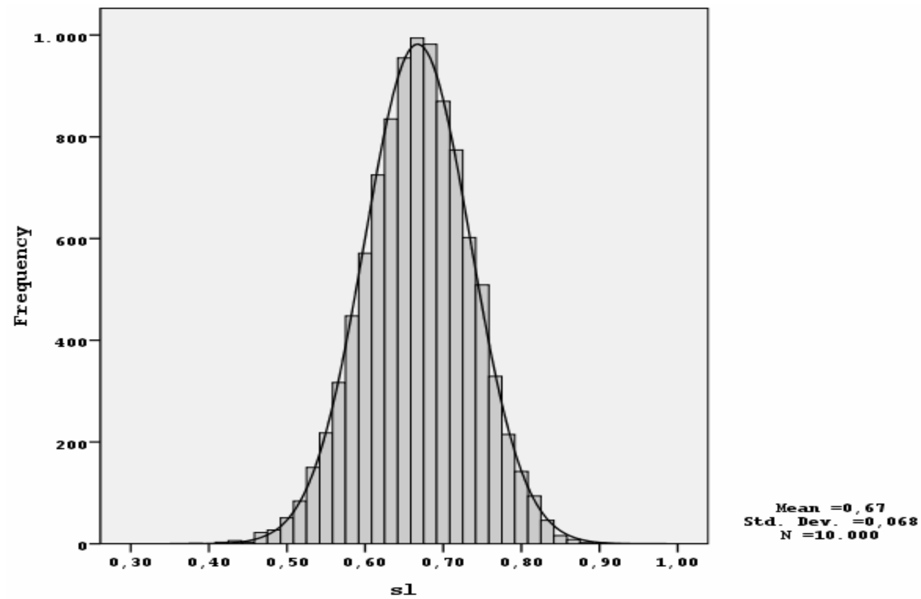
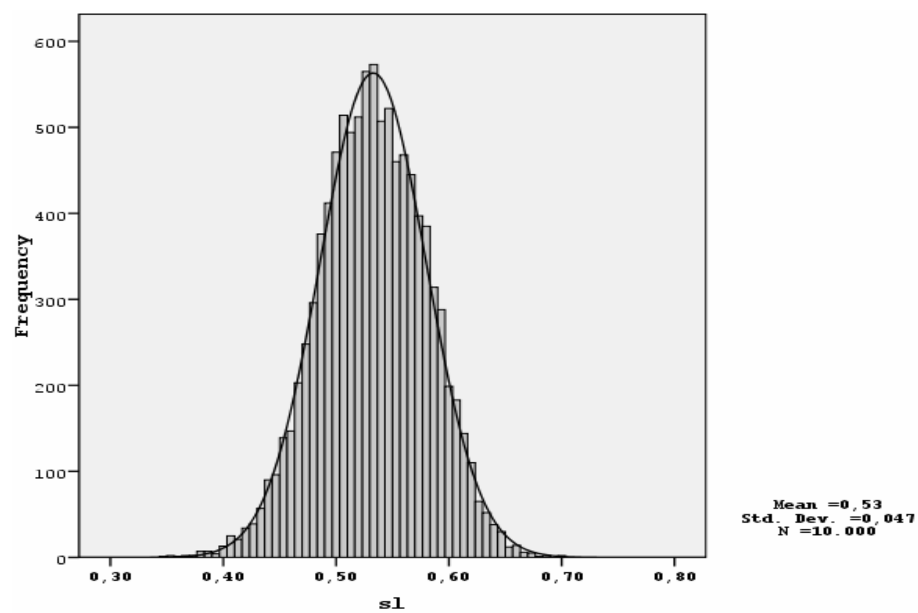


Figure 2. Case III in a 4×4 cross table with sample size 50



4. Example

This example evaluates the efficiency of a new E/F-speed film, Insight, for the determination of approximal carious lesion depths, compared with Ultraspeed. Radiographs of 80 extracted human molars and premolars were taken with both films under standardized conditions. The presence and absence of caries and the depth of the lesions were determined by three observers using a predetermined scale. The actual status of each surface was determined histologically. Differences between the observers' agreement levels were not significant [10].

Table 5. Histology agreement on approximal carious lesions using Insight films

	Actual Status Scores (Histology)						
Insight Film Scores	**0**	**1**	**2**	**3**	**4**	**5**	**Total**
0	54	15	4	9	1	0	**83**
1	5	9	0	7	5	0	**26**
2	1	6	1	4	4	0	**16**
3	3	0	0	12	12	0	**27**
4	0	0	1	4	21	3	**29**
5	0	0	0	0	20	30	**50**
Total	**63**	**30**	**6**	**36**	**63**	**33**	**231**

Table 6. Histology agreement on approximal carious lesions using Ultraspeed films

	Actual Status Scores (Histology)						
Ultraspeed Film Scores	**0**	**1**	**2**	**3**	**4**	**5**	**Total**
0	54	15	3	2	0	0	**74**
1	8	11	0	9	0	0	**28**
2	1	4	1	7	3	0	**16**
3	0	0	1	11	16	1	**29**
4	0	0	1	7	26	6	**40**
5	0	0	0	0	18	26	**44**
Total	63	30	6	36	63	33	**231**

Table 7. Relationship and agreement statistics for film and histology scores

	Insight film scores and Histology	**Ultraspeed** film scores and Histology
Gamma	0.883	0.922
Exact agreement proportion	0.549	0.558
Weighted kappa (linear)	**0.690**	**0.751**
Similarity measurement (s_l)	**0.863**	**0.893**
Expected value of s_l	0.611	0.611
Standard deviation of s_l	0.018	0.018
95% confidence interval for S_l	(0.827, 0.899)	(0.857, 0.929)

In Table 7, the Gamma values show the linear relationship between the two ordinal categorical variables [1]. It is not an agreement coefficient.

The other statistics are related with agreements. In the two data sets, the difference between the proportions of exact agreement is approximately 1%. The difference between the similarity measurements is 3%. On the other hand, the difference between the weighted kappa values is 6%. The similarity measurements show the greatest harmony with the proportion of exact agreement. The 95% confidence interval for S_l is determined by the equation $s_l \pm 1.96\text{Std.Dev.}(s_l)$.

The agreement levels of Insight and Ultraspeed for true depth diagnosis are at an almost perfect level [12].

5. Discussion

The proposed agreement coefficient, called the *linear similarity measurement* (S_l) is easily calculated. Its expected value and variance provided. It is shown that the estimate of the linear similarity measurement is an unbiased and consistent estimator. Since it is defined between zero and one, it provides for an easier interpretation than weighted kappa. It is also sensitive to the levels of the ordinal categorical variables and the number of levels of the ordinal categorical variables.

In future, the test statistics can be developed for the testing of two or more linear similarity measurements, might also be extended to multivariate cases.

Appendix

SPSS matrix language for the weighted kappa and the similarity coefficient.

Step 1: Enter the levels of ordinal categories in first row in SPSS Data Editor,
Step 2: Enter the contingency table following rows in SPSS Data Editor,
Step 3: RUN > ALL the program in SPSS Syntax Editor.

```
matrix.
get table /missing=omit.
compute I=ncol(table).
compute piart=make(I,1,0).
compute partj=piart.
compute sd=make(I,I,0).
compute lw=sd.
compute x=make(I*I,1,0).
compute fx=x.
compute ss=x.
compute pij=x.
compute hsay=0.
compute aggratio=0.
compute a=0.
compute b=0.
loop j=1 to I.
loop k=1 to I.
   compute sd(j,k)=abs(table(1,j)-table(1,k)).
end loop.
end loop.          /*print sd.: shows the linear disagreements in each cell */
compute data=table(1:I+1,1:I).    /*print data.: shows the observation table */
compute n=msum(data).
compute indis=0.
```

```
compute pij=data/n.        /*print pij.: shows the (ij) th cell probabilities */
compute piart=rsum(data)/n.
compute partj=csum(data)/n.
loop j=1 to I.
loop k=1 to I.
   compute lw(j,k)=1-(abs(j-k))/(I-1).
   compute a=a+lw(j,k)*pij(j,k).
   compute b=b+lw(j,k)*piart(j)*partj(k).
end loop.
compute aggratio=aggratio+pij(j,j).
end loop.        /*print aggratio: shows the agreeement ratio */
print aggratio.
compute lwkappa=(a-b)/(1-b).
print lwkappa.        /*print lwkappa.: shows the linear weighted kappa */
loop k=1 to table(1,I)**2.
compute kk=k-1.
compute say=0.
loop j=1 to I.
loop l=1 to I.
do if (sd(j,l)=kk).
compute say=say+1.
end if.
end loop.
end loop.
   do if (say > 0).
compute indis=indis+1.
compute fx(indis)=say.
compute ss(indis)=kk.
end if.
end loop.
compute p=make(indis,1,0).
compute s=make(1,indis,0).
compute varx=make(indis,indis,0).
loop j=1 to indis.
   compute p(j)=fx(j)/msum(fx).
   compute s(j)=ss(j).
end loop.
compute meansx=s*n*p.
loop j=1 to indis.
loop k=1 to indis.
do if (j=k).
   compute varx(j,k)=n*p(j)*(1-p(j)).
else.
compute varx(j,k)=-n*p(j)*p(k).
end if.
end loop.
end loop.
compute varsx=s*varx*t(s).
compute m=n*mmax(s).
compute Sl=1-(msum(sd&*data))/m).
print Sl /forma=f5.3.      /*print Sl:shows the linear similarity coefficient */
```

```
compute meanSl=1-meansx/m.
compute varSl=varsx/m**2.
compute sdevSl=varSl**0.5.
print meanSl /format=f5.3.    /*print meanSl: shows the expected value of Sl */
print varSl /format=f6.7.    /*print varSl: shows the variance of Sl */
print sdevSl /format=f6.7. /*print sdevSl: shows the standard deviation of Sl */
end matrix.
```

References

[1] Agresti, A. *Categorical Data Analysis* (John Wiley and Sons, Inc., New York, 1990).

[2] Bland, J. M. and Altman, D. *Statistical methods for assessing agreement between two methods of clinical measurement*, Lancet **8**, 307–310, 1986.

[3] Brennan, R. L. and Prediger, D. J. *Coefficient kappa: Some uses, misuses and alternatives*, Educational and Psychological Measurement **41**, 687–699, 1981.

[4] Cicchetti, D. V. and Allison, T. *A new procedure for assessing reliability of scoring EEG sleep recordings*, American Journal of EEG Technology **11**, 101–109, 1971.

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

[6] Cohen, J. *Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit*, Psychological Bulletin **70**, 213–220, 1968.

[7] Dunn, G. (1989). *Design and Analysis of Reliability Studies: The Statistical Evaluation of Measurements Errors* (Edward Arnold, London, 1989).

[8] Fay, M. P. *Measuring agreement between two statistics with applications to age standardization*, J. R. Statist. Soc. A **169** (1), 81–96, 2006.

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

[10] Güngör, K., Erten, H., Akarslan, Z. Z., Çelik, I. and Semiz, M. Approximal carious lesion depth assessment with Insight and Ultraspeed films, Operative Dentistry **30** (1), 58–62, 2005.

[11] Janson, H. and Olsson, U. *A measure of agreement for interval or nominal multivariate observations by different sets of judges*, Educational and Psychological Measurement **64** (1), 62–70, 2004.

[12] Landis, J. R. and Koch G. G. *The measurement of observer agreement for categorical data*, Biometrics **33**, 159–174, 1977.

[13] Lin, L., Hedayat, A. S., Sinha, B. and Yang, M. *Statistical methods in assessing agreement: Models, issues, and tools*, Journal of the American Statistical Association **97** (457), 257–270, 2002.

[14] Lin, L. I-K. *Total deviation index for measuring individual agreement with applications in laboratory performance and bioequivalence*, Statistics in Medicine **19**, 255–270, 2000.

[15] Maclura, M. and Willet, W. C. *Misinterpretation and misuse of kappa statistics*, American Journal of Epidemiology **126**, 161–169, 1987.

[16] Perreault, W. D. and Leigh, L. E. *Reliability of nominal data based on qualitative judgments*, Journal of Marketing Research **XXVI**, 135–148, 1989.

[17] Sim, J. and Wright, C. C. *The Kappa statistics in reliability studies: Use, interpretation, and sample size requirements*, Physical Therapy **85** (3), 257–268, 2005.

[18] Schuster, C. *A note on the interpretation of weighted kappa and its relations to other rater agreement statistics for metric scales*, Educational and Psychological Measurement **64** (2), 243–253, 2004.