Albatineh, Ahmed N.; **Khan, Hafiz M. R.**; **Niewiadomska-Bugaj, Magdalena**

On the equivalence of some indices of similarity: implication for binary presence/absence data. (English) Zbl 1335.62090

Aust. N. Z. J. Stat. 54, No. 2, 189-198 (2012).

Summary: Cohen's kappa, a special case of the weighted kappa, is a chance-corrected index used extensively to quantify inter-rater agreement in validation and reliability studies. In this paper, it is shown that in inter-rater agreement for 2×2 tables, for two raters having the same number of opposite ratings, the weighted kappa, Cohen's kappa, Peirce, Yule, Maxwell and Pilliner and Fleiss indices are identical. This implies that the weights in the weighted kappa are less important under such assumptions. Equivalently, it is shown that for two partitions of the same data set, resulting from two clustering algorithms having the same number of clusters with equal cluster sizes, these similarity indices are identical. Hence, an important characterisation is formulated relating equal numbers of clusters with the same cluster sizes to the presence/absence of a trait in a reliability study. Two numerical examples that exemplify the implication of this relationship are presented.

MSC:

62H30 Classification and discrimination; cluster analysis (statistical aspects)
62H20 Measures of association (correlation, canonical correlation, etc.)
62H17 Contingency tables

Keywords:

Cohen's kappa; equivalent indices; Fleiss index; matching counts matrix; Maxwell & Pilliner index; Peirce index; weighted kappa; Yule index

Full Text: DOI

References:

[1] Albatineh, Correcting Jaccard and other similarity indices for chance agreement in cluster analysis, Adv. Data Anal. Classif. 5 pp 179– (2011a) · Zbl 1274.62414 · doi:10.1007/s11634-011-0090-y

[2] Albatineh, MCS: A method for finding the number of clusters, J. Classification 28 pp 184– (2011b) · Zbl 1271.62130 · doi:10.1007/s00357-010-9069-1

[3] Albatineh, Means and variances for a family of similarity indices used in cluster analysis, J. Statist. Plann. Inference 140 pp 2828– (2010) · Zbl 1191.62111 · doi:10.1016/j.jspi.2010.03.005

[4] Albatineh, On similarity indices and correction for chance agreement, J. Classification 23 pp 301– (2006) · Zbl 1336.62168 · doi:10.1007/s00357-006-0017-z

[5] Cohen, A coefficient of agreement for nominal scales, Educational and Psychological Measurement 20 pp 37– (1960)· doi:10.1177/001316446002001

[6] Fleiss, Measuring agreement between two judges on the presence or absence of a trait, Biometrics 31 pp 651– (1975)· doi:10.2307/2529549

[7] Fowlkes, A method for comparing two hierarchical clusterings, J. Amer. Statist. Assoc. 78 pp 553– (1983) · Zbl 0545.62042 · doi:10.1080/01621459.1983.10478008

[8] Gower, Metric and Euclidean properties of dissimilarity coefficients, J. Classification 3 pp 5– (1986) · Zbl 0592.62048 · doi:10.1007/BF01896809

[9] Guttman, Prediction of Personal Adjustment (1941)

[10] Goodman, Measures of association for cross classifications, J. Amer. Statist. Assoc. 49 pp 732– (1954) · Zbl 0056.12801

[11] Hubálek, Coefficients of association and similarity based on binary (presence/absence) data: an evaluation, Biological Review 57 pp 669– (1982)· doi:10.1111/j.1469-185X.1982.tb00376.x

[12] Hubert, Comparing partitions, J. Classification 2 pp 193– (1985) · Zbl 0587.62128 · doi:10.1007/BF01908075

[13] Jaccard, The distribution of the flora of the alpine zone, New Phytologist 11 pp 37– (1912)· doi:10.1111/j.1469-8137.1912.tb05611.x

[14] Jain, Algorithms for Clustering Data (1988) · Zbl 0665.62061

[15] Maxwell, Deriving coefficients of reliability and agreement for ratings, British J. Math. Statist. Psych. 21 pp 105– (1968)· doi:10.1111/j.2044-8317.1968.tb00401.x

[16] Milligan, A study of the comparability of external criteria for hierarchical cluster analysis, Multivariate Behavioral Research 21 pp 441– (1986)· doi:10.1207/s15327906mbr2104_5

[17] Milligan, The Effect of cluster size, dimensionality, and the number of clusters on recovery of true cluster structure, IEEE Trans. Pattern Anal. Machine Intell. PAMI-5 pp 40– (1983)· doi:10.1109/TPAMI.1983.4767342

[18] Morey, The measurement of classification agreement: an adjustment to Rand statistic for chance agreement, Educ. Psychol. Meas. 44 pp 33– (1984)· doi:10.1177/0013164484441003

[19] Peirce, The numerical measure of the success of predictions, Science 4 pp 453– (1884)· doi:10.1126/science.ns-4.93.453-a

[20] Rand, Objective criteria for the evaluation of clustering methods, J. Amer. Statist. Assoc. 66 pp 846– (1971)· doi:10.1080/01621459.1971.10482356

[21] Rosner, Fundamentals of biostatistics (2006)

[22] Saxena, The effect of cluster size, dimensionality, and number of clusters on recovery of true cluster structure through Chernoff-type faces, The Statistician 40 pp 415– (1991)· doi:10.2307/2348731

[23] Spitzer, Quantification of agreement in psychiatric diagnosis, Arch. General Psychiatry 17 pp 83– (1967)· doi:10.1001/archpsyc.1967.01730250085012

[24] Yule, On the methods of measuring association between two attributes, J. R. Stat. Soc. 75 pp 579– (1912)· doi:10.2307/2340126