

Vanbelle, Sophie; Albert, Adelin

Agreement between two independent groups of raters. (English) Zbl 1272.62135
Psychometrika 74, No. 3, 477-491 (2009).

Summary: We propose a coefficient of agreement to assess the degree of concordance between two independent groups of raters classifying items on a nominal scale. This coefficient, defined on a population-based model, extends the classical Cohen's kappa coefficient for quantifying agreement between two raters. Weighted and intraclass versions of the coefficient are also given and their sampling variance is determined by the Jackknife method. The method is illustrated on medical education data which motivated the research.

MSC:

62P15 Applications of statistics to psychology

Cited in 15 Documents

Keywords:

agreement; nominal scale; kappa coefficient

Software:

bootstrap

Full Text: [DOI](#)

References:

- [1] Barnhart, H.X., & Williamson, J.M. (2002). Weighted least squares approach for comparing correlated kappa. *Biometrics*, 58, 1012–1019. · [Zbl 1210.62142](#) · [doi:10.1111/j.0006-341X.2002.01012.x](#)
- [2] Bland, A.C., Kreiter, C.D., & Gordon, J.A. (2005). The psychometric properties of five scoring methods applied to the Script Concordance Test. *Academic Medicine*, 80, 395–399. · [doi:10.1097/00001888-200504000-00019](#)
- [3] Charlin, B., Gagnon, R., Sibert, L., & Van der Vleuten, C. (2002). Le test de concordance de script: un instrument d'évaluation du raisonnement clinique. *Pédagogie Médicale*, 3, 135–144. · [doi:10.1051/pmed:2002022](#)
- [4] Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37–46. · [doi:10.1177/001316446002000104](#)
- [5] Cohen, J. (1968). Weighted kappa: nominal scale agreement with provision for scaled disagreement of partial credit. *Psychological Bulletin*, 70, 213–220. · [doi:10.1037/h0026256](#)
- [6] Efron, B., & Tibshirani, R.J. (1993). An introduction to the bootstrap. New York: Chapman and Hall. · [Zbl 0835.62038](#)
- [7] Feigin, P.D., & Alvo, M. (1986). Intergroup diversity and concordance for ranking data: an approach via metrics for permutations. *The Annals of Statistics*, 14, 691–707. · [Zbl 0604.62042](#) · [doi:10.1214/aos/1176349947](#)
- [8] Fleiss, J.L. (1981). *Statistical methods for rates and proportions* (2nd ed.) New York: Wiley. · [Zbl 0544.62002](#)
- [9] 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. · [doi:10.1177/001316447303300309](#)
- [10] Hollander, M., & Sethuraman, J. (1978). Testing for agreement between two groups of judges. *Biometrika*, 65, 403–411. · [Zbl 0389.62031](#) · [doi:10.1093/biomet/65.2.403](#)
- [11] Kraemer, H.C. (1979). Ramifications of a population model for $\{\kappa\}$ as a coefficient of reliability. *Psychometrika*, 44, 461–472. · [Zbl 0425.62088](#) · [doi:10.1007/BF02296208](#)
- [12] Kraemer, H.C. (1981). Intergroup concordance: definition and estimation. *Biometrika*, 68, 641–646. · [doi:10.1093/biomet/68.3.641](#)
- [13] Kraemer, H.C., Vyjeyanthi, S.P., & Noda, A. (2004). Agreement statistics. In D'Agostino, R.B. (Ed.), *Tutorial in Biostatistics* (vol. 1, pp. 85–105). New York: Wiley.
- [14] Lipsitz, S.R., Williamson, J., Klar, N., Ibrahim, J., & Parzen, M. (2001). A simple method for estimating a regression model for $\{\kappa\}$ between a pair of raters. *Journal of the Royal Statistical Society Series A*, 164, 449–465. · [Zbl 1002.62523](#)
- [15] Raine, R., Sanderson, C., Hutchings, A., Carter, S., Larking, K., & Black, N. (2004). An experimental study of determinants of group judgments in clinical guideline development. *Lancet*, 364, 429–437. · [doi:10.1016/S0140-6736\(04\)16766-4](#)
- [16] Schouten, H.J.A. (1982). Measuring pairwise interobserver agreement when all subjects are judged by the same observers. *Statistica Neerlandica*, 36, 45–61. · [Zbl 0499.62095](#) · [doi:10.1111/j.1467-9574.1982.tb00774.x](#)

- [17] Schucany, W.R., \& Frawley, W.H. (1973). A rank test for two group concordance. *Psychometrika*, 38, 249–258. · [Zbl 0281.62098](#) · [doi:10.1007/BF02291117](#)
- [18] van Hoeij, M.J., Haarhuis, J.C., Wierstra, R.F., \& van Beukelen, P. (2004). Developing a classification tool based on Bloom's taxonomy to assess the cognitive level of short essay questions. *Journal of Veterinary Medical Education*, 31, 261–267. · [doi:10.3138/jvme.31.3.261](#)
- [19] Vanbelle, S., Massart, V., Giet, G., \& Albert, A. (2007). Test de concordance de script: un nouveau mode d'établissement des scores limitant l'effet du hasard. *Pédagogie Médicale*, 8, 71–81. · [doi:10.1051/pmed:2007002](#)
- [20] Vanbelle, S., \& Albert, A. (2009). Agreement between an isolated rater and a group of raters. *Statistica Neerlandica*, 63, 82–100. · [Zbl 1272.62135](#) · [doi:10.1111/j.1467-9574.2008.00412.x](#)
- [21] Williamson, J.M., Lipsitz, S.R., \& Manatunga, A.K. (2000). Modeling kappa for measuring dependent categorical agreement data. *Biostatistics*, 1, 191–202. · [Zbl 0959.62110](#) · [doi:10.1093/biostatistics/1.2.191](#)

This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.