AGREEMENT AMONG MULTIPLE RATERS ON AN ORDINAL SCALE: COMPARISON OF MEASUREMENTS FOR AGREEMENT.

Sertdemir Y[1], Burgut R[1], Unal I[1], and Gunasti S[2].

[1]*Cukurova University School of Medicine Department of Biostatistics*
[2]*Cukurova University School of Medicine Department of Dermatology*

Agreement (reliability) among raters is an important issue in medical research. Such methods like the Kappa statistic used for the measure of agreement between 2 or more raters on a nominal scale are well known. For normally distributed continuous data, the Intra Class Correlation Coefficient (ICC) or the Gauge R&R estimator are very useful and informative. For ordinal scaled data, however, the ICC method is inappropriate as ordinal scale is not a distance metric. For bounded ordinal data, Mast and Wieriengen (2004) proposed the modified ICC which is based on a latent variable model and is a variant of the ICC. Moreover, there are other methods such as “Kendall’s tau”, “Spearman’s rho”, “Krippendorff’s alpha” used to measure agreement for ordinal data.
In our study we will discuss the properties of methods of agreement based on simulated ordered data and a subset of real data from a study in Dermatology where lesions were rated on a scale from 0 to 6. In that study, three different groups of raters evaluated border irregularity in 108 (54+54) clinical images of pigmented skin lesions using clinical ABCD criteria before and after a special training program. Simulation results and estimates for measurements from different methods for groups of raters will be given and discussed.