Stability of Measure Comparisons for avg-idf on msmarco-subsample (p=0.5)