

Inter-rater reliability

In statistics, inter-rater reliability (also called by various similar names, such as inter-rater agreement, inter-rater concordance, interobserver reliability, and so on) is the degree of agreement among raters. It is a score of how much homogeneity, or consensus, there is in the ratings given by various judges. In contrast, intra-rater reliability is a score of the consistency in ratings given by the same person across multiple instances. Inter-rater and intra-rater reliability are aspects of test validity. Assessments of them are useful in refining the tools given to human judges, for example by determining if a particular scale is appropriate for measuring a particular variable. If various raters do not agree, either the scale is defective or the raters need to be re-trained.

There are a number of statistics that can be used to determine inter-rater reliability. Different statistics are appropriate for different types of measurement. Some options are: joint-probability of agreement, Cohen's kappa, Scott's pi and the related Fleiss' kappa, inter-rater correlation, concordance correlation coefficient and intra-class correlation.

From:

<https://neurosurgerywiki.com/wiki/> - **Neurosurgery Wiki**

Permanent link:

https://neurosurgerywiki.com/wiki/doku.php?id=inter-rater_reliability

Last update: **2024/06/07 02:58**

