

Cohen's Kappa

Explorable.com 53.7K reads

Cohen's Kappa is an index that measures interrater agreement for categorical (qualitative) items.

The items are indicators of the extent to which two raters who are examining the same set of categorical data, agree while assigning the data to categories, for example, classifying a tumor as 'malignant' or 'benign'.

Comparison between the level of agreement between two sets of dichotomous scores or ratings (an alternative between two choices, e.g. accept or reject) assigned by two raters to certain qualitative variables can be easily accomplished with the help of simple percentages, i.e. taking the ratio of the number of ratings for which both the raters agree to the total number of ratings. But despite the simplicity involved in its calculation, percentages can be misleading and does not reflect the true picture since it does not take into account the scores that the raters assign due to chance.

Using percentages can result in two raters appearing to be highly reliable and completely in agreement, even if they have assigned their scores completely randomly and they actually do not agree at all. Cohen's Kappa overcomes this issue as it takes into account agreement occurring by chance.



Quiz:
Psychology 101 Part 2



Quiz:
Psychology 101 Part 2



Quiz:
Flags in Europe

[See all quizzes =>](#)

How to Compute Cohen's Kappa

The formula for Cohen's Kappa is:

$$\kappa = \frac{\text{Pr}(a) - \text{Pr}(e)}{1 - \text{Pr}(e)}$$

Where:

$\text{Pr}(a)$ = Observed percentage of agreement,

$\text{Pr}(e)$ = Expected percentage of agreement.

The observed percentage of agreement implies the proportion of ratings where the raters agree, and the expected percentage is the proportion of agreements that are expected to occur by chance as a result of the raters scoring in a random manner. Hence Kappa is the proportion of agreements that is actually observed between raters, after adjusting for the proportion of agreements that take place by chance.

Let us consider the following 2x2 contingency table, which depicts the probabilities of two raters classifying objects into two categories.

	Rater 1			Total
	Category	1	2	
Rater 2	1	P11	P12	P10
	2	P21	P22	P20
Total		P01	P02	1

Then

$$\text{Pr}(a) = P01 + P10$$

$$\text{Pr}(e) = P02 + P20$$

Interpretation

The value of κ ranges between -1 and +1, similar to Karl Pearson's co-efficient of correlation ^[1] 'r'. In fact, Kappa and r assume similar values if they are calculated for the same set of dichotomous ratings for two raters.

A value of kappa equal to +1 implies perfect agreement between the two raters, while that of -1 implies perfect disagreement. If kappa assumes the value 0, then this implies that there is no relationship between the ratings of the two raters, and any agreement or disagreement is due to chance alone. A kappa value of 0.70 is generally considered to be satisfactory.

However, the desired reliability ^[2] level varies depending on the purpose for which kappa is

being calculated.

Caveats

Kappa is very easy to calculate given the software's available for the purpose and is appropriate for testing whether agreement exceeds chance levels. However, some questions arise regarding the proportion of chance, or expected agreement, which is the proportion of times the raters would agree by chance alone. This term is relevant only in case the raters are independent, but the clear absence of independence calls its relevance into question.

Also, kappa requires two raters to use the same rating categories. But it cannot be used in case we are interested to test the consistency of ratings for raters that use different categories, e.g. if one uses the scale 1 to 5, and the other 1 to 10.

Source URL: <https://explorable.com/cohens-kappa>

Links

[1] <https://explorable.com/statistical-correlation>

[2] <https://explorable.com/validity-and-reliability>