

What is Cohen's Kappa Statistic and how is it used in healthcare?

Authored by
stats writer

December 8, 2025

RECOMMENDED CITATION

stats writer (2025). *What is Cohen's Kappa Statistic and how is it used in healthcare?*.
PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=106726>

Cohen's Kappa Statistic is a measure of inter-rater reliability that is often used to assess the accuracy and consistency of healthcare professionals. It is calculated by comparing the actual agreement between raters to the agreement expected by chance alone, and is typically used to assess the reliability of diagnostic tests, coding accuracy, or other types of agreement between two or more raters. By quantifying the level of agreement between healthcare professionals, it allows healthcare organizations to identify areas in need of improvement and ensure accuracy of their data.

Cohen's Kappa Statistic is used to measure the level of agreement between two raters or judges who each classify items into mutually exclusive categories.

The formula for Cohen's kappa is calculated as:

$$k = (po - pe) / (1 - pe)$$

where:

po: Relative observed agreement among raters

pe: Hypothetical probability of chance agreement

Rather than just calculating the percentage of items that the raters agree on, Cohen's Kappa attempts to account for the fact that the raters may happen to agree on some items purely by chance.

How to Interpret Cohen's Kappa

Cohen's Kappa always ranges between 0 and 1, with 0 indicating no agreement between the two raters and 1 indicating perfect agreement between the two raters.

The following table summarizes how to interpret different values for Cohen's Kappa:

Cohen's Kappa	Interpretation
0	No agreement
0.10 - 0.20	Slight agreement
0.21 - 0.40	Fair agreement
0.41 - 0.60	Moderate agreement
0.61 - 0.80	Substantial agreement
0.81 - 0.99	Near perfect agreement
1	Perfect agreement

The following step-by-step example shows how to calculate Cohen's Kappa by hand.

Calculating Cohen's Kappa: Step-by-Step Example

Suppose two museum curators are asked to rate 70 paintings on whether they're good enough to be hung in a new exhibit.

The following 2×2 table shows the results of the ratings:

		Rater 2	
		Yes	No
Rater 1	Yes	25	10
	No	15	20

Step 1: Calculate relative agreement (po) between raters.

First, we'll calculate the relative agreement between the raters. This is simply the proportion of total ratings that the raters both said "Yes" or both said "No" on.

$$p_o = (\text{Both said Yes} + \text{Both said No}) / (\text{Total Ratings})$$

$$p_o = (25 + 20) / (70) = \mathbf{0.6429}$$

Step 2: Calculate the hypothetical probability of chance agreement (pe) between raters.

Next, we'll calculate the probability that the raters could have agreed purely by chance.

This is calculated as the total number of times that Rater 1 said "Yes" divided by the total number

of responses, multiplied by the total number of times that Rater 2 said "Yes" divided by the total number of responses, added to the total number of times that Rater 1 said "No" multiplied by the total number of times that Rater 2 said "No."

For our example, this is calculated as:

$$P(\text{"Yes"}) = ((25+10)/70) * ((25+15)/70) = 0.285714$$

$$P(\text{"No"}) = ((15+20)/70) * ((10+20)/70) = 0.214285$$

$$pe = 0.285714 + 0.214285 = \mathbf{0.5}$$

Step 3: Calculate Cohen's Kappa

Lastly, we'll use p_o and p_e to calculate Cohen's Kappa:

$$k = (p_o - p_e) / (1 - p_e)$$

$$k = (0.6429 - 0.5) / (1 - 0.5)$$

$$k = 0.2857$$

Cohen's Kappa turns out to be **0.2857**. Based on the table from earlier, we would say that the two raters only had a "fair" level of agreement.

You can use this to automatically calculate Cohen's Kappa for any two raters.