Learn How to Calculate Cohen’s Kappa in Excel: A Step-by-Step Guide

Authored by

Mohammed looti

October 31, 2025

The measurement of inter-rater reliability is a cornerstone of robust statistical analysis, especially in fields like psychology, medicine, and quality control. Among the various metrics available, Cohen's Kappa stands out as a powerful statistic used to quantify the level of agreement between two independent raters or judges who classify items into specific, mutually exclusive categories. Unlike simple percentage agreement, Kappa adjusts for the agreement that could occur purely by chance, providing a more accurate and conservative measure of reliability. Understanding how to calculate and interpret this metric is essential for drawing reliable conclusions from qualitative data.

This guide provides a comprehensive breakdown of the Kappa statistic and offers a clear, step-by-step methodology for calculating **Cohen's Kappa** directly within Microsoft Excel, making complex statistical analysis accessible without specialized software.

Deconstructing the Kappa Formula

The core purpose of the Kappa statistic is to differentiate between genuine consensus and accidental agreement. It achieves this by contrasting the observed agreement with the expected chance agreement. The fundamental formula for calculating **Cohen's Kappa** is mathematically defined as:

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

This compact formula relies on two critical components that must be accurately calculated from the rating data: the observed agreement and the chance agreement.

The variables within the formula represent the following key statistical concepts:

po: This is the Relative Observed Agreement among raters. It is simply the proportion of items for which both raters assigned the exact same category.
pe: This represents the Hypothetical Probability of Chance Agreement. This is the probability that the raters would agree simply because of random chance, given the marginal distributions of their individual ratings.

By subtracting the expected chance agreement (pe) from the observed agreement (po) in the numerator, we isolate the non-chance agreement. The denominator normalizes this value, ensuring the result is constrained between 0 and 1, where 1 signifies perfect reliability.

Interpreting Kappa Values: The Scale of Agreement

The resulting value for Cohen's Kappa (k) always falls within the range of 0 to 1, though negative values are technically possible if agreement is less than chance (indicating systemic disagreement). A value of 0 suggests that the observed agreement is no better than what would be expected if the raters were simply guessing randomly. Conversely, a value of 1 signifies **perfect**

agreement between the two raters on every single item, after accounting for chance.

Since Kappa is a standardized measure, its interpretation is typically guided by established benchmarks. These guidelines help researchers translate the numerical score into a qualitative assessment of the reliability achieved. While slight variations exist in these standards, the following table provides a commonly accepted framework for interpreting the strength of the agreement:

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

A Kappa value close to 0.8 or higher is generally considered excellent, suitable for high-stakes research or clinical decisions, whereas values below 0.4 often suggest poor or questionable reliability, necessitating improved training or clearer classification criteria.

Practical Example: Curating an Art Exhibit

To illustrate the calculation process, consider a scenario involving two expert art museum curators. They are tasked with evaluating a collection of 70 unique paintings. Their assignment is to independently rate each painting on a binomial scale: whether it is "Good Enough" (Yes) or "Not Good Enough" (No) to be included in a prestigious new exhibit. This setup perfectly aligns with the requirements for calculating **Cohen's Kappa**, as we have two raters classifying items into mutually exclusive categories.

The combined results of their independent ratings are summarized in a standard 2x2 contingency table, showing the frequency of agreement and disagreement across the 70 paintings. This matrix forms the foundation for all subsequent calculations required in Excel.

	A	B	C	D	E	F	G
1			Rater 2		Total		
2			Yes	No			
3	Rater 1	Yes	25	10	35		
4		No	15	20	35		
5	Total		40	30	70		
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							

From this table, we can immediately identify the agreements: 30 instances where both said "Yes" and 15 instances where both said "No." The total number of agreements is 45 (30 + 15). The total number of disagreements is 25 (10 + 15). This raw data now allows us to proceed with the calculation of the necessary probabilities.

Step-by-Step Calculation in Microsoft Excel

Calculating Cohen's Kappa in Excel requires a structured approach to compute po, pe, and finally, the k value. The following steps, mirrored by the accompanying screenshot, demonstrate how to efficiently set up the spreadsheet and apply the necessary statistical formulas to the 2x2 matrix.

First, determine the **Observed Agreement (po)**. This value measures the proportion of all ratings where the two curators were in agreement. It is calculated by summing the diagonal values (where agreement occurs) and dividing by the total number of observations (N=70). In Excel terms, this calculation is: (Agreement Yes/Yes + Agreement No/No) / Total.

Second, calculate the **Expected Chance Agreement (pe)**. This requires calculating the marginal totals (row and column sums) first. The probability of chance agreement is the sum of the products of the marginal probabilities for each category. Specifically, it is the probability that Rater 1 said "Yes" multiplied by the probability that Rater 2 said "Yes," added to the probability that Rater 1 said "No" multiplied by the probability that Rater 2 said "No." This calculation accounts for the individual biases or distributions of ratings used by each curator.

The subsequent screenshot illustrates the implementation of these formulas in Excel, clearly showing how the cell references lead to the final Kappa result:

	A	B	C	D	E	F
1			**Rater 2**		Total	
2			**Yes**	**No**		
3	**Rater 1**	**Yes**	25	10	35	
4		**No**	15	20	35	
5	Total		40	30	70	
6						
7						
8	p_0	0.6429	=(C3+D4)/SUM(C3:D4)			
9	p_e	0.5	=MMULT(C5:D5, E3:E4)/E5^2			
10	**k**	0.2857	=(B8-B9)/(1-B9)			
11						
12						
13						
14						
15						
16						

Analyzing the Final Kappa Result

Once the intermediate probabilities are established, the final value for **Cohen's Kappa** (k) can be determined using the main formula. In this specific example, the calculations yield the following intermediate results:

The **po** value, representing the relative observed agreement between the curators, is the proportion of total ratings where they both concurred ("Yes" or "No"). This calculation results in a value of **0.6429** (45 agreements / 70 total paintings).

The **pe** value, which models the hypothetical chance agreement based on marginal totals, measures the probability that the raters could have agreed purely by chance. In this instance, the calculation of the expected chance agreement yields a result of **0.5**. This figure is crucial as it acts as the baseline for comparison.

The final **k** value is then calculated by substituting these values into the Kappa formula:

k = (po - pe) / (1 - pe)
k = (0.6429 - 0.5) / (1 - 0.5)

k = **0.2857**

The resulting Cohen's Kappa value of **0.2857** indicates a modest level of agreement between the two art curators. Referring back to the standard interpretation table, this value falls squarely within the "fair" level of agreement (0.21-0.40). This suggests that while there is some consensus beyond random chance, the reliability of their joint rating decisions is moderate, potentially warranting a review of the rating criteria or further discussion between the experts.

Additional Resources for Inter-Rater Reliability

For those seeking deeper statistical understanding or alternative measures of reliability, the following tutorials offer additional resources and comparative analyses related to **Cohen's Kappa** and related inter-rater metrics: